

ÉCOLE DOCTORALE Santé, Sciences, Technologies

GICC – UMR CNRS 6239

THÈSE

présentée par :

Fabien PALAZZOLI

soutenue le : 27 mai 2011

pour obtenir le grade de : **Docteur de l'université François - Rabelais**

Discipline/ Spécialité : Sciences de la Vie

**Exploitation de l'information brevets dans
un laboratoire de recherche public :
identification de niches de développement
technologique en bioproduction et en thérapie génique**

THÈSE dirigée par :

Monsieur BIGOT Yves

Directeur de Recherche CNRS, Université F. Rabelais de Tours

Monsieur TESTU François-Xavier

Professeur des Universités, Université F. Rabelais de Tours

RAPPORTEURS :

Monsieur MERMOD Nicolas

Directeur de l'Institut de Biotechnologie, Université de Lausanne

Monsieur PITARD Bruno

Directeur de Recherche CNRS, INSERM, Nantes

JURY :

Madame EQUINET Leïla

Responsable de l'antenne Région Centre, INPI

Madame GUIVARC'H Nathalie

Professeur des Universités, Université F. Rabelais de Tours

Monsieur DOU Henri

Directeur d'Atelis, ESCM Tours-Poitiers

Monsieur MERMOD Nicolas

Directeur de l'Institut de Biotechnologie, Université de Lausanne

Monsieur PITARD Bruno

Directeur de Recherche CNRS, INSERM, Nantes

Monsieur BIGOT Yves

Directeur de Recherche CNRS, Université F. Rabelais de Tours

Monsieur TESTU François-Xavier

Professeur des Universités, Université F. Rabelais de Tours

Si le Prince éclairé et le Général avisé défont l'ennemi chaque fois qu'ils passent à l'action, si leurs réalisations surpassent celles du commun, c'est grâce à l'information préalable - Sun Tzu, L'art de la guerre.

Remerciements

Après avoir réalisé plusieurs paysages brevets et réseaux de collaboration, c'est à mon tour de prendre un peu de recul pour définir mon propre réseau. En effet, une thèse est tout sauf une aventure solitaire et je souhaite par conséquent remercier toutes les personnes qui m'ont encouragé et qui ont contribué à ce projet. Les travaux de cette thèse qui associe deux mondes différents mais liés, les biotechnologies et la Propriété Intellectuelle, ont été réalisés au GICC sous la direction d'Yves Bigot et de François-Xavier Testu. Je tiens à les remercier de m'avoir fait confiance et de m'avoir conseillé tout au long de ces trois ans, tout en me laissant une certaine autonomie. Yves, j'ai énormément appris à vos côtés et je vous suis infiniment reconnaissant de tout le temps que vous m'avez consacré. Vos conseils, votre indéfectible soutien, votre implication énergétique et votre PATIENCE ont beaucoup compté au cours des années passées au labo. Je vous souhaite, ainsi qu'à votre future équipe, mes meilleurs vœux de réussite pour vos (nombreux) projets.

Je remercie Nicolas Mermod d'avoir accepté de faire partie du jury et de nous avoir accueillis chaleureusement avec Igor Fisch à Lausanne. Faire une présentation d'un paysage brevets devant l'un des leaders technologiques du domaine étudié a été pour moi une consécration.

Je voudrais aussi remercier Henri Dou, pionnier de l'Intelligence Économique en France, pour me faire l'honneur de faire partie de mon jury et pour nous avoir donné l'opportunité de présenter notre démarche de labo de Biotech à un congrès sur l'Intelligence Économique! 非常感谢您! J'espère sincèrement que ses efforts pour promouvoir l'Intelligence Économique en France, ainsi que ceux des autres experts français, ne seront pas vains.

Je remercie Leïla Equinet, Bruno Pitard et Daniel Leroy (j'aurais pu faire un diagramme à bulles pour les remerciements) d'avoir accepté d'être membres du jury de thèse et de m'avoir transmis leurs encouragements et leurs conseils.

Je tiens également à remercier Nathalie Guivarc'h pour avoir accepté d'être membre du jury. D'une certaine manière, c'est grâce à vous et aux autres responsables du Master Biotech & Droit, Nadine Imbault et Olivier Pichon, que cette thèse a pu voir le jour.

Je ne peux bien entendu que remercier Sylvain Goiran, Delphine Cardi, Ugo Streichenberger, Eugénie Mérigeault, Saliou Sarr, Romain Liège et Frank Merly de FIST. Sans votre dynamisme, votre enthousiasme, votre disponibilité et votre savoir-faire, je n'en serai pas là aujourd'hui. Notre fructueuse collaboration a été déterminante pour mes deux dernières années de thèse. C'est donc tout naturellement que je vous dédie l'article *Nature Biotechnology* qui est, à mon sens, l'aboutissement de mes travaux de thèse. Cet article est la preuve que, même quand on travaille sur les brevets, on peut obtenir des publications dans des journaux prestigieux...

Et c'est avec un grand plaisir et une grande fierté que je vous ai rejoints début 2011 !

Je n'oublierai certainement pas Angélique Dacheux du LFB qui m'a fait bénéficier de son expérience d'Ingénieur Brevets et qui a toujours essayé d'être disponible malgré son emploi du temps. Angélique, je te souhaite mes vœux de réussite pour ton examen de mandataire européen !

Je tiens également à remercier le service Partenariat et Valorisation du CNRS de la DR08 à Orléans, Christophe Terrasse, Camile Seze-Goismier, Alexandra Cheri, pour leur soutien et leurs encouragements. J'en profite pour remercier les « anciens » du service Céline Rozier, Léontina Ritan et Benoît-Jules Youbicier-Simo qui sont partis vers d'autres horizons professionnels. J'espère avoir le plaisir de retravailler avec vous un jour. J'ai également une pensée pour Ginette Ollivier...

Ma reconnaissance s'adresse aussi à Eric Levacher de l'IMT de Tours, pour avoir accepté un chapitre sur la Propriété Industrielle dans le livre Bio3 : Biotechnologies – Bioproduction – Biomédicaments (un futur best-seller sans aucun doute) et pour m'avoir donné l'occasion de présenter l'importance de l'information brevets lors du show de lancement de l'ouvrage.

Je tiens maintenant à remercier tous les membres du GICC qui ont contribué à mon intégration et à mon épanouissement au sein ou en dehors du laboratoire. Merci à tous les anciens et actuels thésards et post-docs, merci à Nico (que l'on admire tous pour ce qu'il a fait en Côte d'Ivoire et bravo pour ta petite fille), Jérémy (un « expert » à la française), Dina (l'une des plus ferventes supportrices de Benfica, qui excusera à l'avance les victoires de

Lyon contre son équipe favorite), Sami (le seul à avoir vu que Jacky Chan jouait dans l'équipe du Tours FC) et tous les autres.

Merci à toute l'équipe 2, à Guillaume (shhhhhhhhhhhh! Les expériences avec la carboglace ont été... détonantes! J'espère que tu continueras à t'épanouir dans ton job), à Jeanne (à qui on aura fait les 400 coups avec les gars comme lui glisser du papier bulles sous les roues de son fauteuil, tu me manqueras Jeannette!), Sophie (la reine du dancefloor québécois, j'espère sincèrement que tu auras bientôt un poste!). Solenne et Florence je vous remercie pour vos nombreux conseils, votre soutien quotidien et l'affection que vous m'avez apportée tout au long de ces trois années, sans oublier toutes nos discussions passionnantes sur les romans *Heroic Fantasy* et la mode! Bonne continuation et bon épanouissement dans votre vie!

Je remercie aussi les financeurs des travaux de thèse : la Commission Européenne, le CNRS, le Ministère de l'Éducation Nationale, de la Recherche et de la Technologie, et le Groupement de Recherche CNRS 2157, sans qui cette formidable expérience n'aurait pu être possible. Et je ne parle pas que des voyages en Belgique (Brugge et ce repas dans un certain restaurant), au Canada (quel pays chaleureux!), en Suisse (où nous avons eu l'occasion de manger une bonne fondue), à Nice (avec une virée à Monaco)...

Je ne peux oublier tous mes proches Alan et Lise, Thomas, Charlotte et leurs adorables enfants, Yannick, Pierre (from London), Aurélien (from Tokyo), Emilie et Yannick, Julian (from Sidney) et tous les autres que je n'ai pas cités.

Je remercie également toute la famille Bonnet pour toute l'affection qu'ils m'apportent depuis plusieurs années. Anne-Laure, Pierre-Jean, Nathalie, Sylvie, Patrick, Christine, Sandra, Christophe et leur petit garçon, Simone, Jacqueline, votre soutien sincère pendant les moments difficiles, notamment en 2008, m'a beaucoup touché et m'a permis de surmonter ces épreuves. Enfin, je ne remercierai jamais assez toute ma famille, ma petite sœur, mon père et ma mère qui est partie trop tôt, d'avoir toujours eu confiance en moi. Marjorie, j'espère que tu décrocheras ton CDI et que tu continueras à « remonter la pente » comme tu le fais actuellement avec succès. Papa, saches que je te soutiendrai toujours. Je suis fier de tous les deux. Maman, je te dédie cette thèse.

Résumé

Dans un monde où la course à l'innovation est de plus en plus rapide, il est important pour une entreprise innovante ou un laboratoire de recherche public de mettre en place une stratégie de protection et de valorisation de ses inventions qui soit performante. La protection des résultats par des brevets revêt une importance capitale pour le développement industriel des biotechnologies qui forment un secteur innovant et prometteur, et où la R&D exige des investissements financiers considérables. Au-delà de cet intérêt fondamental, les brevets sont aussi une source de premier plan en matière d'informations technologiques, juridiques et stratégiques, pouvant être exploitées à travers des paysages brevets. Ces études constituent un outil privilégié d'aide à la décision en matière de stratégie de R&D puisqu'elles permettent de définir les axes de recherche des concurrents et les niches de développement technologique libres de droits de Propriété Intellectuelle.

Dans cette optique, mes travaux de thèse ont consisté en l'élaboration et l'analyse de plusieurs paysages brevets sur des technologies utilisées en bioproduction et en thérapie génique : les transposons, les domaines de liaison à l'ADN de type doigt de zinc et les éléments de contrôle de la chromatine. Ces études ont permis l'identification de niches de développement technologique qui ont été intégrées dans les projets de recherche du laboratoire. Cependant, l'information brevets comme indicateur concurrentiel et technologique est largement méconnue dans les laboratoires de recherche publics pourtant eux-aussi des acteurs de l'innovation. J'ai donc montré par mes travaux que les laboratoires de recherche publics ont tout autant intérêt à miser sur l'information brevets et la veille stratégique pour rester compétitifs et innover. En effet, dans un contexte international où la compétition scientifique et économique s'intensifie, il est fort probable que les acteurs de l'innovation qui réussiront seront ceux qui utiliseront l'information disponible avec le plus de pertinence.

Mots clés : information brevets, paysages brevets, veille stratégique, biotechnologies, bioproduction, thérapie génique.

Résumé en anglais

In a world where the innovation race is increasing fast, it is of economic importance for an innovative company or a public research laboratory to develop a strategy for the protection and enhancement of its inventions is efficient. Protection of results through patents is critical for the industrial development of biotechnology which are an innovative and promising sector where R&D requires considerable financial investments. Beyond this fundamental interest, patents are also a source of information on technological, legal and strategic, which can be exploited through patent landscapes. These studies are a key tool for decision support in R&D since they allow to identify research strategies of competitors and technological niches free from of Intellectual Property rights.

In this context, my thesis work has consisted in the development and analysis of several landscapes patents on different technologies used in bioproduction and gene therapy: transposons, zinc finger DNA-binding domains and chromatin control elements. These studies allowed the identification of niches of technological development that have been then incorporated within the laboratory's research projects. However, patent information as a competitive and technology indicator is largely unknown in the public research laboratories, yet they are also actors of innovation. I have therefore showed in my PhD thesis that public research laboratories have as much interest to capitalize on the patent information and intelligence to stay competitive and innovate. Indeed, in an international context where the scientific and economic competition intensifies, it is likely that the innovation actors that succeed will be those who will use the information available with the most relevance.

Keywords: patent information, patent landscape, strategic survey, biotechnology, bioproduction, gene therapy.

Table des matières

Remerciements	4
Résumé.....	7
Résumé en anglais.....	8
Table des matières	9
Liste des tableaux.....	14
Liste des figures.....	15
Liste des annexes.....	17
Liste des abréviations.....	18
1. Introduction.....	22
1.1. Origine des travaux de thèse	23
1.2. Le brevet, outil de protection et outil d'information.....	27
1.2.1. Le brevet, outil de protection : généralités et cas des biomédicaments.....	27
1.2.2. Le brevet, outil d'information - Article 1 : <i>Les brevets : une source d'informations stratégiques pour les acteurs privés et académiques</i>	33
1.3. Du médicament au biomédicament : évolution du paysage des produits thérapeutiques et importance des brevets	42
1.3.1. Impact des biotechnologies sur le marché des produits thérapeutiques	42
1.3.2. Les biomédicaments : généralités.....	45
1.3.3. Perspectives du secteur des biomédicaments et nouvelle évolution de la situation de la Propriété Intellectuelle.....	50
1.4. Collaboration avec la société France Innovation Scientifique et Transfert	54
1.4.1. Présentation de l'UMR 6239 CNRS Génétique-Immunothérapie-Chimie & Cancer et de la Cellule StratégICC	54

1.4.2.	Présentation de la société partenaire FIST	56
1.4.3.	Présentation de la Direction Marketing Brevets & Licences.....	58
1.4.4.	Contexte et historique de la collaboration	58
1.4.5.	Modalités de la collaboration	60
2.	Matériel et méthodes	63
2.1.	Présentation de la méthode.....	64
2.2.	Formulation de la stratégie de recherche.....	70
2.2.1.	Identification de mots clés à partir de la littérature bibliographique	70
2.2.2.	Identification de mots clés à partir du web	71
2.2.3.	Identification de mots clés à partir de la littérature brevets.....	72
2.2.4.	Généralités sur le système de classification des brevets.....	74
2.2.5.	Identification de symboles de classification à l'aide de mots clés.....	78
2.2.6.	Association de mots clés et de symboles de classification	80
2.2.7.	Utilisation d'autres critères de recherche.....	81
2.3.	Choix de la base de données de brevets.....	82
2.3.1.	La base de données esp@cenet.....	83
2.3.2.	La base de données PATENTSCOPE.....	89
2.3.3.	La base de données FamPat de Questel et l'outil de recherche QPAT.....	93
2.4.	Formatage et analyses des documents brevets extraits.....	98
2.4.1.	Présentation de l'outil d'analyse Intellixir	98
2.4.2.	Formatage manuel des données.....	100
2.4.3.	Analyses proposées par le système Intellixir et exemples de représentations graphiques.....	101
2.5.	Recherche d'informations complémentaires.....	109

2.5.1.	Informations économiques et industrielles	109
2.5.2.	Informations liées à la littérature bibliographique	111
3.	Résultats	114
3.1.	Article 2 : <i>Sustained transgene expression using non-viral enzymatic systems for stable chromosomal integration</i>	115
3.1.1.	Contexte et objectif de l'article.....	115
3.1.2.	Résumé de l'article	115
3.1.3.	Conclusion de l'article	116
3.1.4.	Conséquences pour le choix des études et la suite de mes travaux	117
3.1.5.	Remarque sur la méthodologie et les outils employés.....	118
3.1.6.	Article 2	118
3.2.	Article 3 : <i>Transposon tools: worldwide landscape of intellectual property and technological developments</i>	143
3.2.1.	Contexte et objectif de l'article.....	143
3.2.2.	Résumé de l'article	143
3.2.3.	Conclusion de l'article	144
3.2.4.	Remarque sur la méthodologie et les outils employés.....	146
3.2.5.	De l'importance de connaître les paysages brevets avant de financer des projets de R&D : exemple du projet INTHER	146
3.2.6.	Article 3	148
3.3.	Article 4 : <i>Zinc finger monopoly : quelles sont les règles du jeu ?</i>	197
3.3.1.	Contexte et objectif de l'article.....	197
3.3.2.	Résumé de l'article	198
3.3.3.	Conclusion de l'article	199

3.3.4.	Conséquences pour la suite de mes travaux de thèse et le projet SyntheGeneDelivery.....	200
3.3.5.	Remarque sur la méthodologie et les outils employés.....	200
3.3.6.	Article 4	201
3.4.	Article 5 : <i>Characterization of monomeric protein domains binding specifically to a highly conserved 100-bp target within rDNA genes</i>	207
3.4.1.	Contexte et objectif de l'article.....	207
3.4.2.	Résumé et conclusion de l'article	207
3.4.3.	Article 5	209
3.5.	Article 6 : <i>Patent landscape of chromatin control elements: position effects in the pharmaceutical bioproduction</i>	277
3.5.1.	Contexte et objectif de l'article.....	277
3.5.2.	Résumé de l'article	278
3.5.3.	Conclusion sur l'article	279
3.5.4.	Article 6	280
4.	Discussion.....	282
4.1.	Importance de la veille pour les paysages brevets	298
4.1.1.	Exemple des transposons.....	298
4.1.2.	Exemple des éléments de contrôle de la chromatine.....	302
4.2.	Stratégies de PI propres à certains déposants.....	302
4.2.1.	Contournement des brevets considérés comme bloquants.....	302
4.2.2.	Stratégies de PI de certains acteurs privés profitant du système des brevets 303	
4.2.3.	Stratégie originale exploitée par des acteurs académiques : l' <i>open innovation/source</i>	308

4.2.4.	Déviance et manipulation de la notion d' <i>open innovation/source</i>	311
4.3.	Problème d'accès à l'information asiatique.....	313
4.3.1.	Accès à l'information brevets chinoise	313
4.3.2.	Solution envisagée et actions proposées.....	316
5.	Conclusion :	321
	Bibliographie	328
	Annexe 1.....	344
	Annexe 2.....	374
	Annexe 3.....	376
	Annexe 4.....	378
	Annexe 5.....	380
	Annexe 6.....	381
	Annexe 7.....	382

Liste des tableaux

Tableau 1 : Liste des opérateurs à disposition des utilisateurs de QPAT pour la construction de requêtes complexes	96
Tableau 2 : Description des principaux symboles CIB des brevets de la base créée dans Intellixir.....	107
Tableau 3 : Résultats issus du poster <i>Transposon tools: worldwide patent landscape and patent exploitation by key actors</i> , présenté lors du congrès international <i>Mobile DNA</i> à Montréal du 24 au 28 avril 2010.....	145

Liste des figures

Figure 1 : Étapes majeures de la vie d'un brevet en procédures nationale française et internationale.....	28
Figure 2 : Capture d'écran d'esp@cenet présentant la demande de brevet internationale WO2004078981	35
Figure 3 : Utilisation des brevets et de l'information brevets dans les entreprises technologiques en Europe et aux États-Unis	36
Figure 4 : Schéma de développement d'une molécule thérapeutique, du dépôt de brevet à son exploitation commerciale	42
Figure 5 : Schéma d'un procédé standard de bioproduction	47
Figure 6 : Pourcentage de croissance et chiffre d'affaires de biomédicaments. La couleur indique les classes de biomédicaments	50
Figure 7 : Hypothèse haute du nombre de molécules biologiques qui sont ou seront commercialisées entre 2006 et 2012.....	51
Figure 8 : Pourcentage des brevets en fonction de la classe des biomédicaments	53
Figure 9 : Organigramme technique de la Cellule StratéGICC au 1er janvier 2009.....	56
Figure 10 : Schéma présentant la collaboration et les caractéristiques des partenaires.....	60
Figure 11 : Schéma des étapes successives de recherche d'informations et d'analyse pour la réalisation d'un paysage brevets	64
Figure 12 : Symbole complet de classement représentant un symbole CIB	76
Figure 13 : Représentation graphique simplifiée du système de la CIB	76
Figure 14 : Illustration d'une recherche de symbole ECLA à partir de mots clés	79

Figure 15 : Représentation schématique de la constitution d'une requête concernant une étude sur les nanoparticules utilisées en bio-imagerie	80
Figure 16 : Représentation schématique de la structure d'esp@cenet, le réseau européen de bases de données de brevets.....	84
Figure 17 : Copie d'écran de la recherche avancée d'esp@cenet.....	86
Figure 18 : Exemples de graphiques proposés par l'outil d'analyse statistique de PATENTSCOPE	90
Figure 19 : Exemple de traduction du terme anglais « transposase » en allemand, français et japonais.	91
Figure 20 : Capture d'écran de l'interface de recherche avancée de QPAT, comprenant le détail des différents choix proposés pour certains critères de recherche	95
Figure 21 : Évolution des dépôts de demandes de brevets entre 1987 et 2008.....	101
Figure 22 : Distribution géographique de la protection accordée par les brevets.....	102
Figure 23 : Réseau de collaboration entre les inventeurs présents dans la base de données de brevets créée dans Intellixir	103
Figure 24 : Top 15 des inventeurs considérés comme experts par Intellixir	104
Figure 25 : Top 15 des déposants de brevets	105
Figure 26 : Nuage des mots clés ou concepts présents dans les documents brevets portant sur les éléments de contrôle de la chromatine.....	106
Figure 27 : Taille du portefeuille de brevets couvrant les systèmes d'éléments de contrôle de la chromatine pour chaque déposant.....	108
Figure 28 : Réseau de collaboration des auteurs d'articles pour la requête « Matrix Attachment Regions », créé par le site GoPubMed.....	112

Liste des annexes

Annexe 1 : Le rôle des brevets dans la protection des biomédicaments et de leurs systèmes de production. Chapitre 6 de l'ouvrage Bio3 : Biotechnologies – Bioproduction – Biomédicaments.....	344
Annexe 2: Poster <i>Transposon tools: worldwide patent landscape and patent exploitation by key actors</i>	374
Annexe 3 : Poster <i>Patent landscape analysis “Nanoparticles for bio-imaging”</i>	376
Annexe 4 : Communications orales réalisées	378
Annexe 5 : Article BIOTECH.INFO n°437	380
Annexe 6: <i>New concept of technology classification</i> - WIPO.....	381
Annexe 7 : Recherche d'informations en chinois sur Baidu et sur le portail d'informations CNKI	382

Liste des abréviations

AAP	Appels À Projets
ADN	Acide DésoxyriboNucléique
AMM	Autorisation de Mise sur le Marché
ARN	Acide RiboNucléique
BIOS	<i>Biological Innovation for Open Society</i>
CAMBIA	<i>Center for Application of Molecular Biology to International Agriculture</i>
CCP	Certificat Complémentaire de Protection
CHO	<i>Chinese Hamster Ovary</i>
CIB	Classification Internationale des Brevets
CIP	<i>Continuation-in-Part</i>
CLIR	<i>Cross-Lingual Information Retrieval</i>
CNKI	<i>China National Knowledge Information</i>
CNRS	Centre National de la Recherche Scientifique
ECLA	<i>European patent CLAssification</i>
EMA	<i>European MEDicines Agency</i>
EPO	<i>ErythroPOietin</i>
FIST	France Innovation Scientifique et Transfert
GDR	Groupeement De Recherche
GICC	Génétique-Immunothérapie-Chimie-Cancer
HUGO	<i>HUman Genome Organisation</i>
ICO	<i>In Computer Only</i>
IE	Intelligence Économique
IMT	Institut des Métiers et des Technologies

INPI	Institut National de la Propriété Industrielle
iPS	<i>induced Pluripotent Stem</i>
JPO	<i>Japan Patent Office</i>
KWIC	<i>Key Word In Context</i>
LCR	<i>Locus Control Region</i>
LEEM	Les Entreprises du Médicament
LTR	<i>Long Terminal Repeat</i>
MAR	<i>Matrix Attachment Region</i>
MB&L	Marketing Brevets & Licences
MeSH	<i>Medical Subject Headings</i>
PCRDT	Programme-Cadre de Recherche et Développement Technologique
PI	Propriété Intellectuelle
OPEN	<i>Oligomerized Pool ENgineering</i>
PCT	<i>Patent Community Treaty</i>
PIPRA	<i>Public Intellectual Property Resource for Agriculture</i>
PME	Petite et Moyenne Entreprise
PLuTO	<i>Patent Language Translations Online</i>
OMPI	Office Mondial de la Propriété Intellectuelle
OEB	Office Européen des Brevets
OCDE	Organisation pour la Coopération et le Développement Économiques
R&D	Recherche & Développement
SB	<i>Sleeping Beauty</i>
SAR	<i>Scaffold Attachment Region</i>
SIPO	<i>State Intellectual Property Office of the People's Republic of China</i>
STAR	<i>STabilizing and Anti-Repressor element</i>

STREP	<i>Specific Targeted Research Project</i>
UCOE	<i>Ubiquitously acting Chromatin Opening Element</i>
UMR	Unité Mixte de Recherche
USPTO	<i>United States Patent and Trademark Office</i>
WIPO	<i>World Intellectual Property Organization</i>
ZF	<i>Zinc Finger</i>

1. Introduction :

**Les brevets au cœur du *business* des
biotechnologies et des biomédicaments**

1.1. Origine des travaux de thèse

Mon projet de thèse trouve son origine dans le plan d'exploitation d'un projet européen du 6^{ème} Programme-Cadre de Recherche et Développement Technologique (PCRDT), SyntheGeneDelivery (2005-2009). Il s'agissait d'un projet STREP (*Specific Targeted Research Project*) qui répondait à la ligne d'action FP6-LIFESCIHEALTH LSH-2004-1.2.4-4 sur le développement de technologies innovantes de transfert de gène non virales pour la thérapie génique *ex vivo* [w1]. SyntheGeneDelivery a été conçu pour développer des stratégies de transfert de gènes pour les traitements de maladies héréditaires et acquises chez l'homme. La thérapie génique représente de nouvelles possibilités de traitements concernant des maladies pour lesquelles il n'existe pas actuellement de solutions thérapeutiques. Le projet concernait une stratégie de thérapie génique *ex vivo* qui regroupe plusieurs étapes successives :

- 1) le prélèvement et la culture de cellules d'un patient,
- 2) l'introduction d'un gène thérapeutique dans les cellules prélevées à l'aide d'un vecteur viral ou non-viral puis l'intégration dans le génome des cellules hôtes,
- 3) la réimplantation des cellules génétiquement modifiées chez le patient.

L'objectif de SyntheGeneDelivery était de développer un nouveau protocole pour fournir une expression stable et à long terme d'un transgène intégré dans le génome des cellules souches avec un vecteur non-viral. Un des aspects du projet était donc de développer un vecteur de transfert de gène. Par vecteur, on entend le résultat de l'assemblage de systèmes de transfection cellulaire et d'import nucléaire actif, associés à un système d'intégration efficace, indépendamment de la taille du transgène et du site d'insertion spécifique dans les chromosomes. Le programme de travail prévoyait notamment de résoudre trois verrous technologiques majeurs en matière de transfert de gène :

- 1) l'internalisation du transgène dans les cellules souches : le fragment d'ADN d'intérêt doit franchir la membrane plasmique de la cellule hôte,
- 2) le transport et la traversée de l'enveloppe nucléaire du transgène : l'ADN exogène doit parvenir jusqu'au au noyau,

- 3) l'intégration du gène thérapeutique à un locus génomique spécifique de la cellule hôte.

L'équipe 2 *Cellules usines, Ingénierie des protéines et Bio-informatique* du GICC au sein de laquelle j'ai réalisé mes travaux de thèse a focalisé ses travaux de recherche sur la mise au point d'un vecteur non-viral qui s'intègre à un site chromosomique spécifique. Dans cet objectif, les membres de l'équipe 2 ont cherché dans un premier temps à concevoir un vecteur qui permette d'insérer des transgènes d'intérêt dans les chromosomes. Dans un second temps, ils devaient développer des solutions pour attribuer au vecteur un caractère d'intégration site-spécifique, non seulement pour avoir une expression stable du transgène transporté par le vecteur, mais aussi pour limiter les effets délétères liés aux insertions multiples et aléatoires. L'importance de ce problème de génotoxicité a été soulignée lors des premiers essais de thérapie génique *ex vivo* réalisés sur des patients (« bébés bulles ») présentant un déficit immunitaire combiné sévère lié au chromosome X. Ces patients sont contraints de vivre en atmosphère stérile jusqu'à ce qu'ils bénéficient d'une greffe de moelle osseuse compatible. Lors des essais, un phénomène de mutagenèse insertionnelle délétère est apparu dans quelques cellules de toutes les populations cellulaires traitées (intégration du vecteur à proximité ou au sein même d'un oncogène LMO2 dont la mutagenèse peut conduire à une prolifération de type leucémique). Après réimplantation, les cellules souches génétiquement modifiées et mutées ont provoqué des leucémies chez les patients [Fischer *et al.*, 2004]. L'attribution d'un caractère site-spécifique aux vecteurs d'intégration non-viraux représente actuellement l'un des défis majeurs pour le transfert de gène, que ce soit en thérapie génique ou en bioproduction. En effet, une utilisation alternative des vecteurs issus de SyntheGeneDelivery concerne le transfert de gène *ex vivo* pour la production de cellules usines, des cellules génétiquement modifiées dans l'unique but d'exprimer une protéine recombinante. À la différence des cellules génétiquement modifiées pour la thérapie génique, celles-ci ne sont pas implantées chez un patient mais cultivées en bioréacteurs pour produire des protéines d'intérêt pharmaceutique, également appelées biomédicaments.

Pour remplir les objectifs de SyntheGeneDelivery, un consortium a été mis en place entre les partenaires publics et privés du projet. Il avait pour but d'organiser la gestion globale du projet, la coordination des activités de recherches, mais également la gestion des connaissances issues du projet, comme la gestion des droits de Propriété Intellectuelle (PI) [European Commission, 2003 ; European Commission, 2006]. Parmi les activités visant à promouvoir l'exploitation des résultats, des études de brevets permettent au projet d'atteindre pleinement ses objectifs en matière de Recherche & Développement (R&D). En effet, elles participent à la définition et à l'évolution des stratégies technologiques des projets et offrent la possibilité de localiser des niches favorables de développement technologique. Dans le cas de SyntheGeneDelivery, elles ont aussi apporté une meilleure compréhension du positionnement des acteurs dans le domaine du transfert de gène et, comme nous le verrons, elles ont permis de localiser des niches de développement technologique libres de PI. Elles permettent ainsi l'élaboration de solides stratégies de recherche pour se positionner dans l'environnement international de compétitions scientifique et économique, sans se retrouver confronté à l'existence de droits de PI appartenant à des tiers.

Dans cette optique, mes travaux de thèse ont consisté à analyser et interpréter les paysages brevets concernant les technologies en cours de développement dans le cadre du projet SyntheGeneDelivery. Cet objectif se déclinait à travers la mise en place d'une veille technologique sur les quatre thématiques scientifiques du laboratoire (Génétique, Immunothérapie, Chimie et Cancer), l'identification des acteurs publics et privés impliqués dans ces technologies et l'identification des financements publics locaux, nationaux et européens.

Mes travaux ont permis l'élaboration de cinq articles qui constituent, avec la méthodologie présentée dans la section 2. Matériel et méthodes, le cœur de mes résultats de thèse :

- *Sustained transgene expression using non-viral enzymatic systems for stable chromosomal integration.* Palazzoli F, Carnus E, Wells DJ, Bigot Y. *Curr Gene Ther.* 2008 Oct;8(5):367-90, [Palazzoli et al., 2008]

- *Transposon tools: worldwide landscape of intellectual property and technological developments.* Palazzoli E, Testu FX, Merly F, Bigot Y. *Genetica*. 2010 Mar;138(3):285-99, [[Palazzoli et al., 2010a](#)]
- *Zinc finger monopoly : quelles sont les règles du jeu ?* Palazzoli E, Carnus E, Merly F, Goiran S, Bigot Y. *Biofutur*. 2010;306:58-62, [[Palazzoli et al., 2010b](#)]
- *Characterization of monomeric protein domains binding specifically to a highly conserved 100-bp target within rDNA genes.* Carnus E, Demattei MV, Casteret S, Carpentier C, Palazzoli E, Bire F, Bressac C, Renault S, Bigot Y. 2011. *En revision*, [[Carnus et al., 2011](#)]
- *Patent landscape of chromatin control elements: position effects in the pharmaceutical bioproduction.* Palazzoli E, Bire S, Bigot Y, Bonnin-Rouleux F. *Nature Biotechnology*. 2011, accepté [[Palazzoli et al., 2011](#)]

Cette démarche originale pour un laboratoire de recherche public entre dans le cadre d'une veille stratégique et de l'Intelligence Économique (IE), qui est définie comme l'ensemble des actions coordonnées de recherche, de traitement, de distribution et de protection de l'information utile aux acteurs économiques obtenue légalement [[Martre et al., 1994](#)]. La veille stratégique quant à elle est un processus informationnel par lequel les entreprises ou plus généralement les acteurs de l'innovation, se mettent à l'écoute anticipative de leur environnement interne et externe, dans le but de saisir des opportunités, de maîtriser les facteurs critiques de succès et d'agir ou de réagir avant les autres [[Lesca, 1994](#) ; [Aron P & Petit, 1997](#)]. Elle représente donc un outil privilégié d'aide à la décision en matière de stratégie de R&D et se base en partie sur la PI. Une définition du brevet, indispensable pour comprendre son importance et son utilisation notamment en biotechnologies, est abordée dans la partie suivante.

1.2. Le brevet, outil de protection et outil d'information¹

1.2.1. Le brevet, outil de protection : généralités et cas des biomédicaments

1.2.1.1. Définition

Un brevet est un titre de PI qui offre à son titulaire la possibilité d'interdire à tout tiers l'exploitation de l'invention couverte par le brevet, ce qui revient à lui accorder un monopole d'exploitation. Ce droit est limité dans le temps, vingt ans à compter de la date de dépôt de la demande de brevet, et dans l'espace, car il est limité au(x) territoire(s) dans le(s)quel(s) la protection a été demandée et accordée. Un brevet se présente en trois parties : une première comprend toutes les informations bibliographiques (noms des déposants et inventeurs, date de dépôt...); une deuxième partie appelée « description » décrit les caractéristiques de l'invention, les moyens d'y parvenir, elle est accompagnée le cas échéant de figures (plasmides...) et de tableaux; une troisième partie appelée « revendications » définit la nature de la protection juridique souhaitée par le demandeur.

1.2.1.2. La procédure brevet

Des étapes clés jalonnent la vie du brevet : dépôt, extension, publication, examen et délivrance. La vie d'un brevet débute à la date de dépôt de la demande de brevet. La terminologie « demande de brevet » est à ne pas confondre avec « brevet » car, dans le premier cas, il s'agit de la protection qui a été demandée et dans le second de la protection qui a été accordée. Cette date de dépôt génère un droit dit « droit de priorité » qui permet pendant une durée maximum de douze mois de demander une protection pour la même invention dans d'autres pays que celui où la première demande a été déposée. Cette demande de protection supplémentaire est appelée « extension de la protection à l'étranger ». Toutes les demandes de brevets et les brevets qui concernent une seule et même invention font partie de la même « famille de brevets » (leur numéro de priorité est identique). Si le demandeur ne souhaite pas bénéficier de son droit de priorité, l'étape

¹ Cette section reprend une partie du chapitre *Le rôle des brevets dans la protection des biomédicaments et de leurs systèmes de production* [Palazzoli et al., 2010c], du livre Bio³ : Biotechnologies – Bioproduction – Biomédicaments.

d'extension de la protection à l'étranger n'a pas lieu. Dans les deux cas, l'étape suivante est la publication, qui a lieu dix-huit mois après le dépôt de la demande de brevet. La publication rend accessible aux tiers les informations contenues dans la demande de brevet telles que les informations techniques relatives à l'invention elle-même, les informations administratives (le demandeur, les inventeurs, la date de dépôt...). La demande de brevet entre ensuite dans un processus dit « d'examen » au cours duquel l'office de brevets du pays considéré détermine si elle répond aux exigences de la loi en matière de brevets. Notamment, l'office de brevet détermine si la demande de brevet répond à des critères dits « de brevetabilité », à savoir : la nouveauté, l'activité inventive et l'application industrielle. À l'issue de cette période d'examen, le brevet est accordé si tous les critères sont remplis ou la demande de brevet est rejetée si ce n'est pas le cas. Si le brevet est accordé, il expirera vingt ans après la date de dépôt (Figure 1).

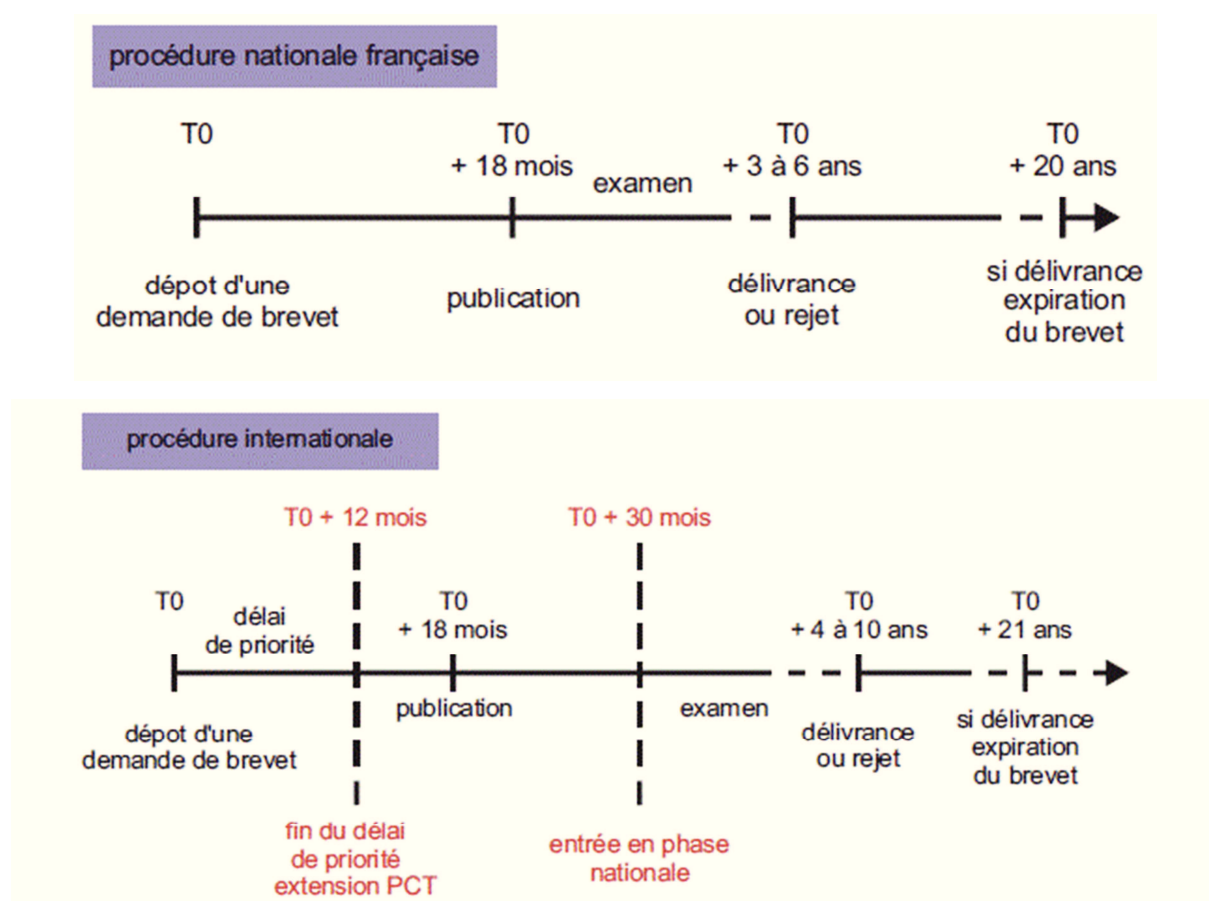


Figure 1 : Étapes majeures de la vie d'un brevet en procédures nationale française et internationale.

Si le demandeur a choisi de bénéficier de son droit de priorité, cela signifie qu'il a demandé dans un délai de douze mois suivant le dépôt d'une première demande, l'extension de la protection dans des pays supplémentaires. Classiquement, cette demande d'extension est réalisée par un système centralisé international appelé système du PCT (*Patent Cooperation Treaty*). Ce traité a pour but de simplifier les procédures de dépôts de brevet dans plus de 140 pays, en permettant à une unique demande dite « internationale » ou demande PCT, d'éclater en un faisceau de brevets dans les pays désignés. L'étape de publication de la première demande a lieu de la même manière dix-huit mois après le dépôt. Dans ce cas de figure, deux publications ont lieu dans le même temps, celle de la première demande et celle de la demande internationale. Lorsqu'une extension par voie internationale a été demandée, le demandeur doit désigner les pays où il souhaite obtenir une protection parmi les pays adhérents au système PCT, dans les trente mois à compter du dépôt de la première demande de brevet. Cette étape est appelée « entrée en phases nationales ». À l'issue de ce choix, la demande de brevet sera examinée dans chacun des pays désignés par l'office de brevets correspondant. Chacun de ces pays statuera sur la délivrance ou le rejet du brevet. Il est à noter qu'il n'existe pas de brevet international : une demande unique ne permet pas d'obtenir un titre unique. Seule la demande est internationale, mais elle aboutit en autant de brevets que de pays choisis. L'expiration des brevets interviendra vingt ans après la date de dépôt de la demande internationale soit vingt et un ans après le dépôt de la première demande.

1.2.1.3. Le Certificat Complémentaire de Protection (CCP)

Pour les médicaments et les produits phytopharmaceutiques, il existe un titre de PI supplémentaire : le Certificat Complémentaire de Protection (CCP). Il est organisé par le droit des brevets mais fait intervenir des aspects réglementaires liés aux médicaments car il ne peut être obtenu que si une Autorisation de Mise sur le Marché (AMM) a été accordée avant l'expiration du brevet. Le CCP permet au titulaire d'un brevet disposant d'une AMM pour le produit couvert par le brevet, de prolonger la durée de protection du produit en question. Cette disposition a été mise en place dans de nombreux pays et vise à compenser l'impossibilité d'exploiter une invention de médicaments durant les phases préalables et parfois longues d'essais cliniques. Il est important de souligner qu'un produit peut être

couvert par un brevet délivré et ne pas obtenir d'AMM et, à l'inverse, un produit ayant reçu une AMM peut ne pas être couvert par un brevet. En effet, les systèmes réglementaires et d'examen des brevets sont indépendants. Le CCP prolonge la durée de protection d'un brevet d'un maximum de cinq ans et son obtention ne peut pas générer une protection supérieure à quinze ans à compter de l'obtention de l'AMM en Europe (quatorze ans aux États-Unis).

1.2.1.4. Les conditions de brevetabilité

Un biomédicament est constitué de toute substance utilisée ou utilisable en thérapie issue de matière biologique (extraction) ou produite par un procédé de biotechnologies (qui incluent la recherche en ingénierie moléculaire des processus biologiques chez les procaryotes et les eucaryotes, mais aussi l'ensemble des approches regroupées sous le chapeau de recherche fondamentale et translationnelle concernant la santé humaine et animale). En matière de brevet, le biomédicament ne se distingue pas du médicament ou de toute autre invention. En effet, pour être protégeable par un brevet, le biomédicament doit être nouveau, impliquer une activité inventive et être susceptible d'application industrielle. Ces trois critères constituent le fondement de la brevetabilité d'une invention. En d'autres termes, pour pouvoir être protégeable par un brevet, un biomédicament :

- ne doit pas faire partie de l'état de la technique (concept de nouveauté), c'est-à-dire qu'il ne doit pas avoir fait l'objet d'une publication, d'une demande de brevet, d'une communication orale, d'un devis ou de tout autre type de divulgation, avant la date de dépôt de la demande de brevet ;
- ne doit pas découler de manière évidente de ce qui est déjà connu (concept d'activité inventive). C'est-à-dire qu'à la lecture de ce qui est déjà connu, un homme normalement qualifié dans le domaine technique (appelé « homme du métier » en droit des brevets) ne doit pas pouvoir parvenir à l'invention ;
- doit être susceptible d'application industrielle, c'est-à-dire être fabriqué ou utilisé dans l'industrie.

Il en va de même pour un procédé d'obtention ou un système de production d'un biomédicament que l'on souhaiterait protéger, ou encore d'une utilisation de ce biomédicament dans une indication donnée.

1.2.1.5. La Directive relative à la brevetabilité du vivant

Il est difficile de s'intéresser aux biomédicaments sans évoquer la brevetabilité du vivant. Au niveau européen, la Directive 98/44/CE du 6 juillet 1998 tente de poser les grands principes de la brevetabilité dans le domaine du vivant et d'harmoniser les législations nationales. Pour produire ses effets, dans les droits nationaux, la Directive 98/44/CE a dû être transposée dans chacun des états membres. Ce fut le cas en France en 2004, en conséquence de quoi, le code de la PI a été modifié. Dans cette Directive, le premier point à souligner est le fait que la matière biologique est brevetable à condition que l'invention respecte les trois principes fondateurs de la brevetabilité et que cette matière biologique ait été produite à l'aide d'un procédé technique, même si elle préexistait dans la nature. Les inventions qui découlent de matière biologique ne sont donc pas exclues de la brevetabilité. La directive définit ce qu'est la « matière biologique » comme une matière contenant des informations génétiques et qui est autoreproductible ou reproductible dans un système biologique. Cette même directive statue par ailleurs sur la non-brevetabilité du corps humain et de ses éléments. Elle précise que « le corps humain aux différents stades de sa constitution et de son développement, ainsi que la simple découverte d'un de ses éléments, y compris la séquence ou la séquence partielle d'un gène ne sont pas brevetables ». En effet, ceci constituerait une « simple découverte » et les découvertes font parties des éléments exclus de la brevetabilité. Elle définit aussi les éléments exclus du champ de la brevetabilité dans le domaine du vivant, à savoir :

- les procédés de clonage des êtres humains,
- les procédés de modification de l'identité génétique germinale de l'être humain,
- l'utilisation d'embryons humains à des fins industrielles ou commerciales,
- les procédés de modification de l'identité génétique des animaux de nature à provoquer chez eux des souffrances sans utilité médicale substantielle pour l'homme ou l'animal.

Cependant, la Directive 98/44/CE stipule aussi qu'un « élément isolé du corps humain ou autrement produit par un procédé technique, y compris la séquence ou la séquence partielle d'un gène, peut constituer un élément brevetable même si sa structure est identique à celle de l'élément naturel ». Cela signifie que le fait d'isoler une protéine ou un gène par la mise en œuvre d'un procédé technique peut constituer une invention brevetable si celle-ci répond aux critères de brevetabilité. Elle précise aussi que « l'application industrielle d'une séquence ou d'une séquence partielle d'un gène doit être concrètement exposée dans la demande de brevet ». Il faut donc que l'application d'une séquence soit définie pour qu'une séquence soit brevetable. La fonction d'une telle séquence n'est pas nécessairement décryptée mais son utilisation doit être décrite. La directive a introduit ici une limitation en ce qui concerne la brevetabilité des séquences. Du séquençage permettant de mettre en évidence des gènes ne pourrait pas donner lieu à un brevet si une application n'est pas associée à chaque séquence. Enfin, il est important de souligner qu'en droit français seule l'utilisation d'une séquence ou d'une séquence partielle d'un gène est brevetable. En conséquence, la portée d'une revendication couvrant une séquence ou une séquence partielle d'un gène est limitée à la partie de cette séquence directement liée à la fonction spécifique concrètement exposée dans la description.

À l'exception des exclusions ci-dessus, il ressort de cette directive que tout médicament dérivé ou constitué de matière biologique (biomédicament) peut être brevetable si sa mise en évidence fait intervenir un procédé technique et si, dans le cas d'une séquence, l'application de celle-ci est concrètement définie. Les biomédicaments doivent également répondre aux autres critères de brevetabilité au même titre que toute invention.

1.2.1.6. Les différents types de brevet

Plusieurs types de brevets sont envisageables : brevet de produit, de procédé ou d'utilisation. La protection la plus large est donnée par un brevet de produit qui interdit à tout tiers d'exploiter le produit en question, quel que soit le procédé utilisé pour sa production ou l'utilisation qui en est faite. Un brevet de procédé (de purification d'une

protéine par exemple) empêche un tiers de produire la protéine en question en reproduisant le procédé protégé, mais il n'empêche pas le tiers de produire ladite protéine par un autre procédé (de purification pour l'exemple donné). Le brevet d'utilisation a une portée moins large que les deux types précédents, en ce sens qu'il interdit aux tiers d'utiliser un biomédicament donné dans une indication thérapeutique donnée mais il n'empêche pas l'utilisation dudit biomédicament en tant que tel dans d'autres indications. Il est important de souligner qu'un même brevet peut contenir des revendications de produits, des revendications de procédé permettant l'obtention du produit et des revendications d'utilisation de ce produit. Mais ces différents types de protection peuvent aussi être présents dans des brevets différents et même détenus par des déposants différents.

1.2.2. Le brevet, outil d'information - Article 1 : *Les brevets : une source d'informations stratégiques pour les acteurs privés et académiques*

1.2.2.1. L'information brevets dans la veille technologique et stratégique

L'objectif du système des brevets est d'encourager l'innovation et la croissance économique [OMPI, 2003]. Ces droits de PI ont deux fonctions principales. La première a été présentée précédemment : permettre à un détenteur d'interdire tout tiers d'exploiter sur un plan commercial une invention protégée dans un pays donné et pour une durée déterminée. La seconde fonction repose sur la publication et par conséquent la divulgation de l'information technique nouvelle. En effet, comme je l'ai évoqué auparavant, une demande de brevet est publiée dix-huit mois après la date de dépôt et cette divulgation se fait à l'échelle mondiale. Autrement dit, c'est un moyen pour le public et donc pour la communauté scientifique, d'accéder à l'information technique nouvelle et d'exploiter ces connaissances. Les demandes de brevets et les brevets, rassemblés sous le terme unique « documents brevets », constituent alors une source unique d'information technologique. L'Office Européen des Brevets (OEB) estime même que pour certains domaines technologiques, jusqu'à 80% des connaissances techniques actuelles ne peuvent être trouvées que dans les documents brevets [European Commission & EPO, 2007]. Cette information est définie par l'OEB sous le terme d'« information brevets », et désigne les

informations techniques qui figurent dans les documents brevets ainsi que les informations juridiques qui s'y rapportent [w2]. L'Office Mondial de la Propriété Intellectuelle (OMPI) ajoute même qu'il s'agit du recueil de documents techniques sur les technologies nouvelles et innovantes le plus complet, le mieux organisé et le plus actualisé [w3].

L'information brevets comporte quatre dimensions (Figure 2) :

- des informations techniques : elles sont tirées de la description et des dessins de l'invention. La description comprend le domaine technique, l'état de l'art antérieur, un exposé de l'invention, des figures et les applications industrielles ;
- des informations juridiques : elles proviennent des revendications définissant l'étendue de la protection de la situation juridique du brevet (en vigueur ? Dans quel pays ?), et plus généralement de la vie de la demande de brevet c'est à dire de sa validité dans le temps et dans l'espace (dépôt, priorité, publication, retrait, abandon, opposition, délivrance, déchéance...) ;
- des informations stratégiques et commerciales (concurrentielles) : elles dérivent des informations bibliographiques et administratives concernant le dépôt (date de dépôt, noms et adresses des (co-)déposants, (co-)inventeurs...) ;
- des informations relatives à la politique des pouvoirs publics : leur origine vient de l'analyse de l'évolution des dépôts. Elles sont susceptibles d'être utilisées par les responsables de l'élaboration des politiques, par exemple dans le cadre de la stratégie industrielle d'un pays.

HYPERACTIVE, NON-PHOSPHORYLATED, MUTANT TRANSPOSASES OF MARINER MOBILE GENETIC ELEMENTS

Données bibliographiques	Description	Revendications	Mosaïque	Document original	Situation juridique INPADOC
Numéro de publication: WO2004078981 (A1) Date de publication: 2004-09-16 Inventeur(s): BIGOT YVES [FR]; AUGÉ-GOUILLOU CORINNE [FR]; HAMELIN MARIE-HELENE [FR]; BRILLET BENJAMIN [FR] + Demandeur(s): CENTRE NAT RECH SCIENT [FR]; UNIV FRANCOIS RABELAIS DE TOUR [FR]; BIGOT YVES [FR]; AUGÉ-GOUILLOU CORINNE [FR]; HAMELIN MARIE-HELENE [FR]; BRILLET BENJAMIN [FR] + Classification: - internationale C12N15/54; C12N15/55; C12N9/22; A61K38/00; C12N15/54; C12N15/55; C12N9/22; A61K38/00; (IPC1-7): A61K38/16; C12N15/55; C12N9/22 - européenne C12N9/22 Numéro de demande: WO2004FR00161 20040123 Numéro(s) de priorité: FR20030000905 20030128					Également publié en tant que: WO2004078981 (A8) FR2850395 (A1) US2007031967 (A1) US7732209 (B2) JP2006518220 (T) plus >>
Documents cités: US6368830 (B1) WO0073510 (A1) XP002256851 (A) XP002256852 (A) XP002256853 (A) Tout visualiser					

Visualiser la famille de brevets INPADOC
Visualiser la liste des citants
Visualiser le document dans le Registre européen ☒

Abrév. pour WO 2004078981 (A1)

Signaler une erreur concernant les données

INFORMATION STRATÉGIQUE ET COMMERCIALE

INFORMATION TECHNIQUE

INFORMATION JURIDIQUE

Figure 2 : Capture d'écran d'esp@cenet présentant la demande de brevet internationale WO2004078981. La navigation à l'aide des onglets permet de visualiser les différents types d'informations disponibles.

Dans ce contexte, l'information brevets représente une source de premier plan en matière d'informations techniques, juridiques et stratégiques, qui peut être utilisée à différentes fins [OMPI, 2003] :

- s'informer sur l'état de la technique ou l'état de l'art, et donc d'éviter de faire plusieurs fois le même travail sur le plan de la R&D ;
- évaluer la brevetabilité d'une invention et notamment sa nouveauté ;
- estimer la valeur des brevets ;
- collecter des données commerciales et s'informer sur les titulaires des brevets, les inventeurs et suivre les activités des concurrents (orientations de R&D) ;
- recenser les tendances clés et les concepts émergents dans le développement des technologies ;
- éviter d'être en position de contrefaçon en portant atteinte à des droits de PI appartenant à des tiers ;

- améliorer la planification des décisions commerciales telles que la concession de licences, la conclusion de partenariats technologiques, les acquisitions ;
- identifier des niches de développement technologiques, en exploitant des inventions divulguées provenant de demandes de brevets qui n'ont jamais été acceptées, de brevets qui ne sont pas valables dans certains pays ou de brevets qui ne sont plus en vigueur...

En dépit de la valeur du brevet comme indicateur concurrentiel et technologique, cette source reste peu exploitée à cette fin par les entreprises technologiques en Europe avec seulement 61% d'utilisateurs de l'information brevets contre 95% aux États-Unis (Figure 3).

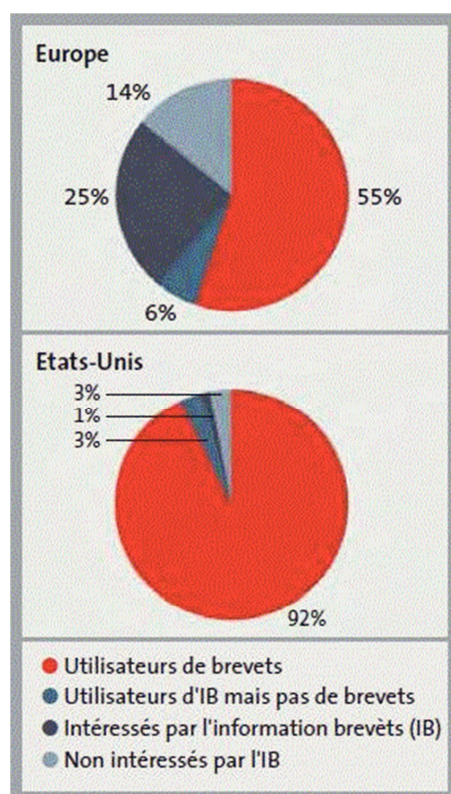


Figure 3 : Utilisation des brevets et de l'information brevets dans les entreprises technologiques en Europe et aux États-Unis [OEB, 2010a].

Le brevet fait partie intégrante de la stratégie des entreprises qui l'utilisent soit dans le cadre d'une politique active de PI soit en tant que source d'informations stratégiques et concurrentielles. Cependant, cette dernière fonction est largement méconnue dans les laboratoires de recherche publics, qui sont également des acteurs de l'innovation. La

recherche publique étant un réservoir non négligeable d'innovations, c'est dans ce contexte que l'article 1 s'intègre.

1.2.2.2. Article 1

EUROPE



LES BREVETS : UNE SOURCE D'INFORMATIONS STRATÉGIQUES POUR LES ACTEURS PRIVÉS ET ACADÉMIQUES

Le contexte actuel international de compétition scientifique et économique place les brevets comme un outil indispensable pour le *business* des biotechnologies. D'un côté, ils accordent aux titulaires un avantage concurrentiel en garantissant la protection d'une invention. De l'autre, ils fournissent à tous les acteurs impliqués dans ce secteur à haut degré d'innovation des informations essentielles pour la prise de décisions stratégiques.

Face à l'extrême complexité du vivant, les biotechnologies sont un secteur où la R&D est particulièrement intensive, ce qui implique des investissements financiers considérables. La protection des résultats revêt donc une importance capitale pour le développement industriel des innovations technologiques (1). Dans ce secteur de pointe, les brevets représentent la protection la plus efficace pour sécuriser les inventions. Par ailleurs, la valorisation de ces avantages concurrentiels va jouer un rôle clé dans l'économie et le développement des entreprises concernées. Celles-ci vont pouvoir, par exemple, amortir indirectement les coûts de R&D en exploitant elles-mêmes leur invention ou en attribuant des licences d'exploitation à des tiers. Ces droits exclusifs et la protection des brevets eux-mêmes vont être directement intégrés dans le *business model* et la stratégie de développement des entreprises de biotechnologies (2). En effet, un portefeuille de titres de propriété industrielle (PI) régulièrement renforcé (perfectionnements, applications

nouvelles) et qui s'inscrit dans une stratégie cohérente est un atout majeur pour convaincre les investisseurs potentiels lors des levées de fonds nécessaires au développement des sociétés de biotechnologies. De plus, il renforce les pouvoirs de négociation auprès des partenaires, clients et concurrents et lors de la phase finale de rachat de la société.

Toutefois, ils ne constituent pas uniquement un moyen de protection. En effet, les brevets représentent, au même titre que les articles scientifiques, une source d'informations incontournable pour l'analyse de l'état de l'art technologique mais également pour la prise de décisions stratégiques.

L'Organisation mondiale de la propriété intellectuelle (OMPI) rappelle que le système des brevets « vise à encourager l'innovation et la croissance économique ». C'est pourquoi, en contrepartie du monopole d'exploitation accordé au titulaire du brevet, la demande est rendue publique 18 mois après la date de dépôt, et ce quel que soit le pays. Les caractéristiques techniques nouvelles de

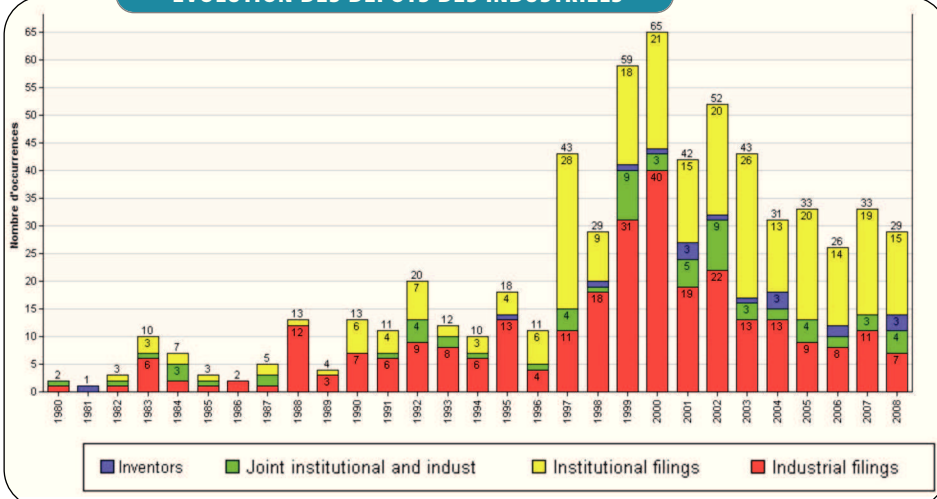
l'invention sont alors dévoilées à l'échelle mondiale. La divulgation de cette « information brevet », définie par l'OMPI comme « toute information qui a

Une PI RÉGULIÈREMENT RENFORCÉE est un ATOUT MAJEUR pour CONVAINCRE LES INVESTISSEURS

déjà été publiée dans un document de brevet ou qui peut être tirée de l'analyse de statistiques relatives aux brevets », permet ainsi à tout tiers d'en exploiter les connaissances, par exemple pour contourner le titre ou s'y opposer. Elle couvre plusieurs types de données. Celles contenues dans la description de l'invention et ses dessins sont tout d'abord des informations techniques qui ne sont souvent pas disponibles en dehors des documents brevets. Il est estimé que jusqu'à 80 % des connaissances techniques actuelles peuvent être trouvées dans les demandes de brevets et les brevets accordés. Au-delà, le titre recèle également des informations juridiques qui sont contenues dans les revendications délimitant l'étendue de la protection. Celles-ci peuvent être exploitées pour déterminer la liberté d'exploitation afin de ne pas être en situation de contrefaçon. Enfin, des données liées au *business* (pays désignés...) vont fournir des précisions essentielles pour la prise de décisions stratégiques. Ces dernières peuvent être exploitées sous la forme de paysages brevets, en tenant compte des facteurs technologiques et de l'analyse de l'environnement. Établies sous forme de représentations graphiques de l'ensemble

(suite p.7)

ÉVOLUTION DES DÉPÔTS DES INDUSTRIELS

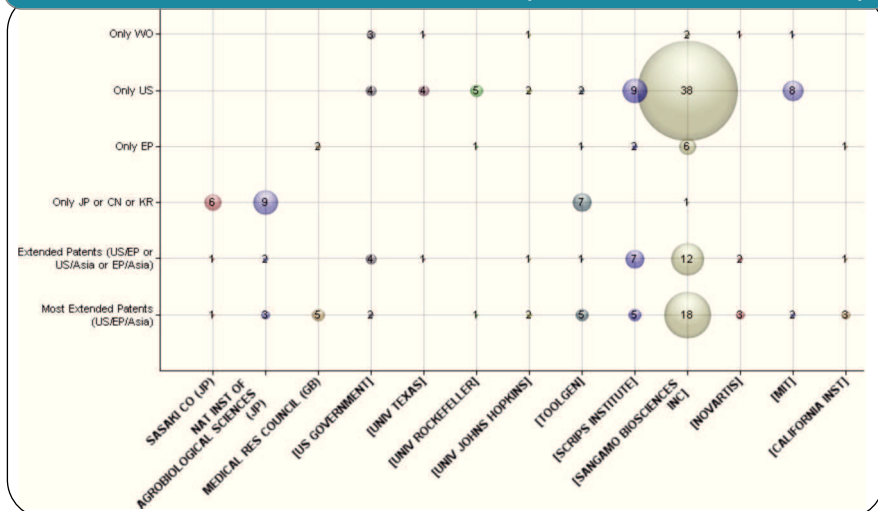


des brevets d'une technologie, elles vont leur permettre de manager leur portefeuille de titres de PI et d'identifier leur environnement en définissant les axes de recherche concurrents pour ainsi mieux définir les orientations stratégiques avant de lancer des projets de R&D (rachat de sociétés concurrentes, négociation de collaborations ou de licences d'exploitation...). L'analyse des paysages brevets permet aussi de déterminer les acteurs clés (déposants et inventeurs) et leurs réseaux de collaboration. Elle offre également la possibilité de détecter des équipes émergentes, définir des tendances technologiques, identifier des technologies complémentaires et estimer la valeur de certains actifs (brevets fondateurs). Ces outils d'intelligence économique apportent donc une meilleure compréhension du positionnement des compétiteurs d'un secteur et permettent de suivre l'évolution de leur environnement. Leur exploitation correspond purement à une démarche d'entreprise. Cependant, les acteurs du monde académique se retrouvent eux aussi confrontés à ce même contexte de mondialisation et de compétition et ont donc de plus en plus recours à l'information brevet.

Vers une exploitation de l'information brevet en laboratoires de recherche publics ?

Les laboratoires de recherche publics ont une vraie légitimité à exploiter eux aussi l'information brevets et les outils d'aide à la décision. Tout d'abord, les chercheurs académiques peuvent consulter les informations techniques des brevets au même titre que celles des articles scientifiques, dans le cadre de leur veille bibliographique. De plus, pour financer leurs recherches, ils sont de plus en plus amenés à répondre à des appels à projets (AAP), où la PI est au cœur même du processus de valorisation des résultats issus des laboratoires publics. Au même titre que les sociétés, la PI, sa solidité et la stratégie de valorisation des résultats font partie intégrante des critères d'éligibilité des projets publics financés. Par exemple, dans les AAP européens des Programmes cadres de recherche et de développement technologique, il est demandé de « *décrire l'état de l'art et les avancées proposées par le projet, mais*

STRATÉGIES DE PROTECTION DES DÉPOSANTS (nombre de brevets dans les bulles)



aussi de démontrer le caractère innovant du programme proposé, en se référant aux études de brevets préliminaires qui ont pu être réalisées ». Par ailleurs, les biotechnologies sont un secteur où l'émergence de *start-up* exploitant une innovation technologique issue de laboratoires académiques est fréquente. Le cadre de création de ces *spin-off* est

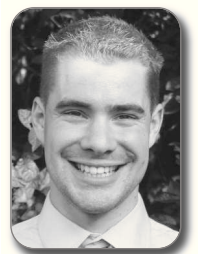
L'ANALYSE du paysage brevets est STRATÉGIQUE pour une ENTREPRISE

même rendu favorable grâce à diverses mesures comme l'AAP Émergence de l'Agence nationale de la recherche, le concours national d'aide à la création d'entreprises de technologies innovantes, ou les incubateurs. Enfin, dans le secteur des médicaments, en pleine mutation depuis une quinzaine d'années, les partenariats entre les sociétés pharmaceutiques et les laboratoires publics se multiplient pour faire face au défi des innovations thérapeutiques. L'exploitation de ces informations stratégiques par un acteur académique peut donc être un argument, en plus de son expertise scientifique, dans les négociations avec des partenaires industriels.

Avoir une telle démarche d'entreprise d'intelligence économique au sein même d'un laboratoire de recherche public est un pas supplémentaire vers l'optimisation des efforts d'innovation. Par exemple, à l'UMR 6239 - Génétique, Immunothérapie, Chimie et Cancer de Tours, l'exploitation de l'information brevets a contribué à l'obtention de

différents financements de projets du laboratoire et de plusieurs contrats. En outre, plusieurs travaux portant sur la PI de technologies issues de l'ingénierie génétique ont été présentés, sous la forme de communications orales et de posters, lors de divers congrès internationaux, européens et nationaux. Par ailleurs, le laboratoire a initié une collaboration avec la société FIST SA (France Innovation Scientifique et Transfert, filiale du CNRS), dont une des expertises porte précisément sur l'analyse de portefeuilles de brevets. L'intégration d'une telle structure exploitant l'information brevets et travaillant en étroite collaboration avec les équipes de recherche représente donc un réel enjeu. Cependant, cela reste peu développé en France, à l'inverse de nombreux laboratoires académiques chinois qui possèdent une cellule de veille stratégique et innovent fortement en biotechnologies. ■

Fabien Palazzoli
Doctorant en analyses de la propriété industrielle en sciences de la vie
UMR CNRS 6239 Génétique, Immunothérapie, Chimie et Cancer
fabien.palazzoli@etu.univ-tours.fr



1 - Lire « Savoir déposer des brevets solides dans l'urgence est une des contraintes des biotech » dans le n°404 de *Bf* du 02/02/2009

2 - Lire « Propriété intellectuelle : Collectis ne lâche rien » dans le n°459 de *Bf* du 26/04/2010

1.2.2.3. Utilisation d'outils d'analyse pour l'élaboration de paysages brevets

Dans un monde où le volume d'informations disponibles double presque tous les deux ans, des outils vont permettre de ressortir une information décisionnelle d'un volume important d'informations très variées, ce qui impliquera un tri à effectuer. L'information brevets est à la disposition du public à travers de nombreuses bases de données qui couvrent chacune des collections particulières de documents. Aucune base de données ne couvre la totalité des documents brevets publiés au niveau mondial et ne permet d'interroger toute l'information disponible. Par conséquent, la première difficulté pour exploiter l'information brevets prend son origine dans le choix des bases de données à même de receler l'information pertinente (quelle couverture ? Quelles limites ?). S'y ajoute la difficulté de la méthodologie d'interrogation des bases pour obtenir les stratégies de recherche les plus performantes (opérateurs booléens ou de proximité, combinaison des étapes de recherche entre elles...). Seules des requêtes d'un niveau de complexité élevé permettront d'atteindre les documents brevets les plus pertinents. Une fois les documents brevets identifiés, deux façons d'analyser l'information brevets se présentent à nous : une méthode qualitative et une méthode quantitative. La première visera l'analyse concise du contenu d'un petit nombre de documents brevets, dans le but par exemple d'évaluer le caractère de nouveauté d'une invention ou de rechercher des atteintes potentielles à des droits de PI. La seconde porte quant à elle sur le traitement statistique de grand volume de documents brevets. Dans ce cas, l'information brevets peut être définie comme « toute information qui peut être tirée de l'analyse de statistiques de documents brevets ». L'analyse des résultats se fait alors à l'aide de représentations visuelles telles que des diagrammes en barres, des graphiques linéaires ou à secteurs... Cette approche par représentation graphique est aisément compréhensible ce qui facilite la lecture et l'interprétation des résultats. En effet, une telle visualisation synthétise un maximum d'informations sous une forme structurée et met ainsi en évidence des relations de façon pertinente. On parle alors de « paysages brevets » ou de « cartographies brevets » [w4]. Ces représentations graphiques de l'ensemble des documents brevets d'une technologie vont permettre de comprendre les axes de recherche des concurrents et ainsi de mieux définir ses propres stratégies de R&D.

Pour faciliter et exploiter efficacement l'information brevets, il est possible d'exploiter les nombreux progrès des laboratoires des sciences de l'information et de la communication. Les outils qu'ils développent peuvent être programmés automatiquement pour effectuer une recherche et un traitement de l'information. Cependant, les outils logiciels ne sont pas doués de la faculté d'analyse stratégique nécessaire pour suivre les choix des acteurs en présence dans un secteur face à l'évolution technologique [Durand, 1993]. Le système d'information repose donc également sur la compétence des hommes et des experts [Dou *et al.*, 1993]. L'auteur Anthony Trippe définit la science de l'analyse de l'information brevets pour la découverte de relations et de tendances avec le terme « *patinformatics* » [Trippe, 2002 ; Trippe, 2003]. Plusieurs outils sont disponibles pour l'analyse de collections entières et volumineuses de documents brevets et ils présentent les résultats sous une forme rapide, analytique et détaillée. Bien entendu, il faut tenir compte des contraintes budgétaires des laboratoires, en recherche perpétuelle de financements pour leurs projets de recherche. Certains logiciels représentent un investissement important, allant de quelques milliers d'euros à plusieurs dizaines de milliers par an. Il existe cependant plusieurs outils libres ou à des prix moindres qui permettent de bien démarrer les activités d'une cellule de veille stratégique. C'est par exemple le cas des logiciels de la société Matheo Softwares qui proposent les outils Matheo Patent [w5] et Matheo Analyzer [w6] à des prix très abordables pour des laboratoires de recherche publics [Dou, 2004].

Bien qu'essentielle, l'information brevets ne correspond qu'à une partie des informations disponibles pour les acteurs de l'innovation. Les autres informations à intégrer dans le processus de veille technologique peuvent être de plusieurs natures : articles de journaux spécialisés, thèses, littérature grise, livres, résumés, comptes rendus de congrès, notices techniques et technologiques des fabricants, rapports d'activités des sociétés, bases de données spécialisées..., sans oublier l'information informelle souvent orale. Face aux difficultés à réaliser un état de l'art exhaustif, il est donc important de s'intéresser à toutes les sources d'informations disponibles, tout en tenant compte des différents supports disponibles : papier, électronique, oral, visuel...

1.3. Du médicament au biomédicament : évolution du paysage des produits thérapeutiques et importance des brevets

1.3.1. Impact des biotechnologies sur le marché des produits thérapeutiques

Le marché des produits thérapeutiques est en plein bouleversement et ce depuis une quinzaine d'années. Cette situation est en partie due à l'expiration de nombreux brevets protégeant principalement des médicaments issus de l'industrie chimique. À l'expiration d'un brevet, celui-ci tombe dans le domaine public, permettant ainsi les copies appelées génériques. Un générique est une copie conforme d'un médicament princeps qui comporte le même principe actif et a le même effet thérapeutique. La molécule princeps est protégée très en amont de son exploitation commerciale. En effet celle-ci intervient généralement dix à douze ans après le dépôt de la demande de brevet principale et les essais cliniques ont lieu entre cinq et dix ans après son dépôt (Figure 4).

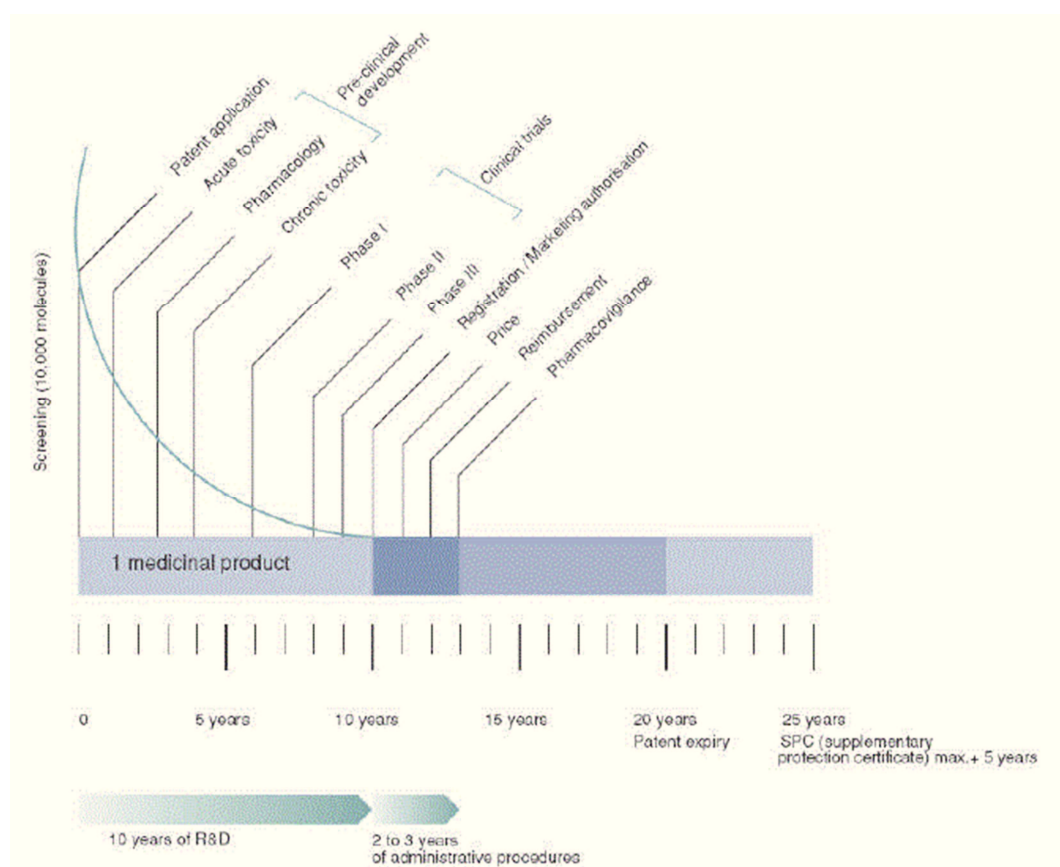


Figure 4 : Schéma de développement d'une molécule thérapeutique, du dépôt de brevet à son exploitation commerciale (Source : Fédération Européenne des Associations et Industries Pharmaceutiques 2008).

L'expiration des brevets et l'arrivée des génériques font partie intégrante de ce bouleversement du marché pharmaceutique. Cette évolution majeure concerne tous les acteurs liés de près ou de loin aux médicaments : les patients et associations de défense des malades, les médecins ainsi que tous les intervenants de la chaîne du médicament, de la R&D à leur mise sur le marché, en passant par les instances nationales et européennes qui gèrent les dépenses de santé. En effet, beaucoup de gouvernements comptent sur les génériques pour réduire les coûts de fonctionnement des systèmes de santé. Par conséquent, ils favorisent leur développement et leur mise en vente à des prix inférieurs à leurs équivalents princeps.

Les grandes entreprises pharmaceutiques, également appelées *Big Pharmas*, subissent quant à elles de plein fouet le développement des génériques, particulièrement lorsque le produit princeps correspondant est un *blockbuster*. Un *blockbuster* est une molécule dont l'exploitation génère un chiffre d'affaire de plus d'un milliard de dollars (plus de 660 millions d'euros) par an. L'expiration des brevets relatifs à ces *blockbusters* et la commercialisation des génériques correspondants impactent fortement le chiffre d'affaire global des sociétés pharmaceutiques. Pour exemple, le médicament Singulair® de Merck (traitement de l'asthme) va tomber dans le domaine public en 2012. Ce médicament génère actuellement plus de 4,5 milliards de dollars de vente. C'est également le cas de l'anticholestérol vedette Lipitor® de Pfizer, qui représente à lui seul douze milliards de dollars de vente et dont le brevet expire en 2011. L'arrivée de génériques engendre nécessairement une perte de chiffre d'affaire pour le produit princeps, ce qui oblige les sociétés de biotechnologies à réagir face à cette situation.

L'innovation est de plus en plus difficile dans le domaine des médicaments et ceci pousse les sociétés pharmaceutiques à s'adapter à cette évolution majeure de leur environnement, en redéfinissant leurs stratégies R&D et commerciale. À ce jour, plusieurs options peuvent être envisagées. Certaines sociétés ont opté pour produire et commercialiser des génériques de leurs molécules princeps. Ainsi, Sanofi-Aventis devrait commercialiser sa propre version générique de l'un des médicaments les plus vendus au monde : le Plavix® (antithrombotique avec plus de 2,5 milliards d'euros de ventes dans le monde en 2008). De son côté, Novartis réalise des acquisitions d'achats de sociétés

spécialisées dans les génériques (Azupharma, Lek). La firme Lilly a de son côté poursuivi en justice pour contrefaçon des fabricants de génériques pour son antidépresseur Prozac®, exemple d'une lutte à l'issue juridique incertaine. D'autres entreprises pharmaceutiques vont se rapprocher et fusionner afin de posséder un portefeuille de molécules en développement plus conséquent. C'est l'opération choisie par Pfizer et Wyeth (opération de 46,8 milliards de dollars) ou de Merck et Schering-Plough (pour un coût de 41 milliards de dollars). Enfin, d'autres encore vont se tourner vers les molécules innovantes issues du passage des technologies de l'industrie de la chimie à celles de l'industrie des biotechnologies.

L'Organisation pour la Coopération et le Développement Économiques (OCDE) propose une définition extrêmement large des biotechnologies : l'application de la science et de la technologie à des organismes vivants, de même qu'à ses composantes, produits et modélisations, pour modifier des matériaux vivants ou non-vivants aux fins de la production de connaissances, de biens et de services [w7]. Ceci peut se faire par l'acquisition de molécules innovantes par des accords de *licensing-in*, c'est-à-dire l'acquisition de produits à différents stades de développement (préclinique ou clinique, via des contrats de licence ou de cession des droits). Cela peut aussi se faire par le biais de rachats de sociétés de biotechnologies de taille plus ou moins importante. C'est l'option choisie par Roche qui a récemment racheté Genentech pour 46,8 milliards de dollars, l'une des plus grandes entreprises de biotechnologies au monde, mais aussi par Merck qui a annoncé son souhait d'acquérir GlycoFi pour 400 millions de dollars. Ces sociétés biopharmaceutiques vont y puiser de nouvelles sources d'innovation thérapeutique et se positionner sur des marchés émergents à forte croissance : ceux des biomédicaments. Leur développement va profiter pleinement des progrès scientifiques et des évolutions technologiques issus des biotechnologies, comme par exemple l'ingénierie génétique, les criblages à haut débit, les analyses bioinformatiques et les séquençages de génomes (dont le projet international *HUMAN Genome Organisation*, communément appelé HUGO). Cependant, les biotechnologies sont un des secteurs où la recherche est la plus intensive et la R&D exige des investissements financiers considérables. La protection des résultats qui en sont issus revêt donc une importance capitale pour leur développement industriel. Le brevet est donc un

élément encore plus essentiel pour la valorisation des biomédicaments que pour celle des molécules chimiques thérapeutiques.

1.3.2. Les biomédicaments : généralités

Comme défini précédemment, un biomédicament est constitué de toute substance utilisée ou utilisable en thérapie issue de matière biologique ou produite par un procédé de biotechnologies. Les biomédicaments sont soumis à une classification qui comprend quatre grands groupes :

- les peptides et protéines : ce groupe met en évidence la distinction à faire entre les protéines naturelles extraites de la matière biologique et les protéines dites recombinantes (éléments composés de matière biologique) produites par des moyens biotechnologiques. Les peptides et protéines sont subdivisés en fonction de leur nature :
 - les hormones (insuline...),
 - les facteurs de croissance (érythropoïétine...),
 - les facteurs plasmatiques : également appelés médicaments dérivés du plasma sanguin. Ce sont principalement des immunoglobulines polyvalentes (IgIV) de l'albumine et des facteurs plasmatiques de la coagulation.
 - les cytokines : ce sont des molécules qui régulent l'activité et la fonction de certaines cellules et interviennent notamment dans la réponse immunitaire et l'activation hépatocytaire lors de phénomènes inflammatoires (interférons et *Tumor Necrosis Factors...*),
 - les protéines de fusion : elles rassemblent les protéines chimériques obtenues par l'association (fusion) de domaines fonctionnels de plusieurs molécules (nucléases *zinc finger*, abordées dans l'article 4 de la section 3. Résultats...).
- les vaccins et anticorps : les anticorps peuvent être des biomédicaments extraits de matière biologique ou produits par des procédés de biotechnologies (par recombinaison ou par transgénèse).

- les acides nucléiques : l'acide désoxyribonucléique (ADN) et l'acide ribonucléique (ARN).

- la thérapie génique et cellulaire :
 - La Partie IV de la Directive 2003/63/CE définit les médicaments de thérapie génique comme « tout produit obtenu par un ensemble de procédés de fabrication visant au transfert, *in vivo* ou *ex vivo*, d'un gène prophylactique, diagnostique ou thérapeutique (à savoir un morceau d'acide nucléique), vers des cellules humaines/animales et son expression consécutive *in vivo* ».
 - La Partie IV de la Directive 2003/63/CE définit aussi les médicaments de thérapie cellulaire somatique : « les cellules vivantes somatiques autologues (émanant du patient lui-même), allogéniques (provenant d'un autre être humain) ou xénogéniques (provenant d'animaux) utilisées chez l'homme, dont les caractéristiques biologiques ont été sensiblement modifiées sous l'effet de leur manipulation pour obtenir un effet thérapeutique, diagnostique ou préventif s'exerçant par des moyens métaboliques, pharmacologiques et immunologiques ».

Ces quatre grands groupes de biomédicaments peuvent se scinder en deux vis-à-vis de leur production. En effet, les peptides et protéines, vaccins et anticorps, acides nucléiques vont être facilement industrialisables, à l'inverse des produits issus des thérapies génique et cellulaire, dont les technologies nécessitent d'être optimisées et validées. Par ailleurs, il est important de distinguer dans la production de médicaments issus des biotechnologies, le biomédicament lui-même et les technologies mises en œuvre pour produire ce biomédicament, c'est-à-dire le système d'expression ou de bioproduction (Figure 5).

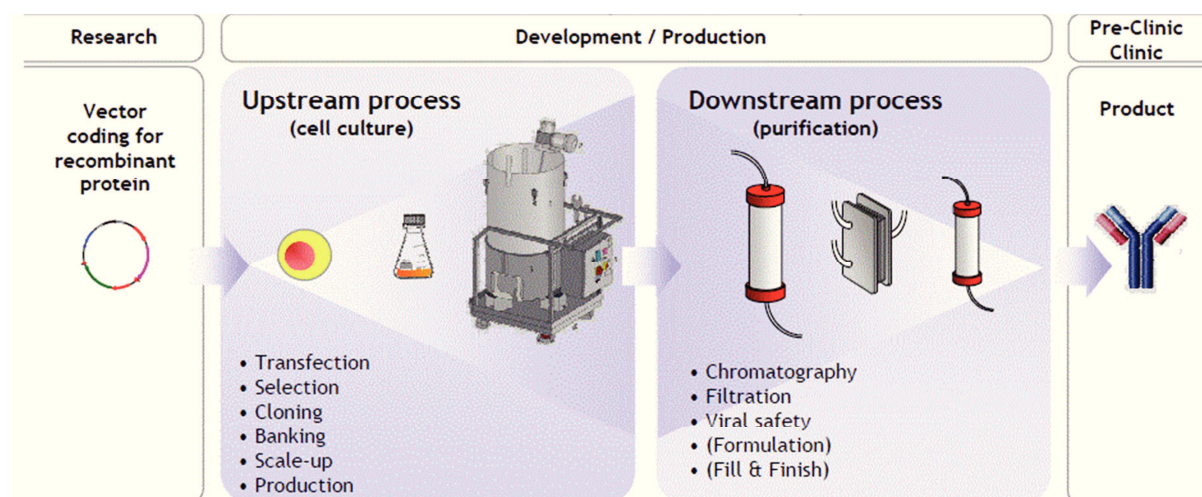


Figure 5 : Schéma d'un procédé standard de bioproduction [Cochet, 2009].

Dans ce manuscrit, nous nous intéresserons plus particulièrement à l'étape de mise au point du système d'expression, c'est-à-dire du vecteur et de la lignée cellulaire utilisés pour la synthèse du biomédicament. Les étapes d'*upstream process* ou culture cellulaire en bioréacteur [Parties 1 et 2 du livre Bio³, 2010], de *downstream process* ou purification de la molécule d'intérêt [Partie 2 du livre Bio³, 2010] et de mise sous forme pharmaceutique [Partie 3 du livre Bio³, 2010] ne seront donc pas abordées.

Le choix d'un biomédicament va conditionner le choix du système d'expression et de son procédé global de bioproduction. Chaque système de bioproduction est unique et optimisé spécifiquement pour l'expression d'un biomédicament. Ce système repose sur la production de protéines recombinantes par des organismes vivants modifiés génétiquement tels que des :

- Cellules bactériennes : les systèmes d'expression bactériens (*Escherichia coli*...) ont été largement employés dans les années 1980 pour la synthèse de petits polypeptides du type interleukine, interféron... Par exemple, pour synthétiser des protéines non-glycosylées telle que l'insuline, les cellules bactériennes conviennent parfaitement.
- Cellules de levures : comme pour les bactéries, les connaissances génétiques des levures en font un système d'expression bien connu pour la bioproduction (*Saccharomyces cerevisiae*...). Cependant, à la différence des cellules bactériennes, les levures peuvent réaliser des modifications post-

traductionnelles essentielles pour de nombreux biomédicaments. Elles peuvent donc synthétiser des protéines dont la maturation *in vivo* est plus complexe.

- Microalgues : bien que le développement de cette technologie soit actuellement moins avancé que les autres systèmes, les microalgues peuvent être utilisées pour la synthèse de protéines glycosylées.
- Cellules d'insectes : elles sont impliquées dans de nombreuses voies de recherche comme les anticorps murins, l'antigène VIH. Le vaccin Cervarix® contre certaines souches du Papillomavirus Humain est par exemple produit dans des cellules d'insectes.
- Cellules d'oiseaux : elles sont notamment utilisées pour produire des anticorps monoclonaux et des vaccins.
- Cellules de mammifères : les cellules de mammifères produisant des protéines recombinantes ont pris beaucoup d'importance au sein des systèmes de bioproduction. En effet, ces lignées reproduisent fidèlement les modifications post-traductionnelles nécessaires au bon fonctionnement des biomédicaments, comme c'est le cas pour les anticorps.
- Animaux et plantes transgéniques : d'autres voies de production de médicaments issus des biotechnologies se basent sur l'utilisation d'organismes entiers comme des animaux non-humains. Il existe par exemple des chèvres transgéniques qui produisent des protéines recombinantes dans leur lait (le produit antithrombotique Atryn®...) ou des maïs transgéniques qui expriment des molécules hétérologues dans leurs semences.

Le choix de la lignée cellulaire dépend directement des biomédicaments à synthétiser et notamment de leur profil de glycosylation. Pour les molécules les plus complexes, les cellules de mammifères seront privilégiées car elles sont à elles seules capables de réaliser toutes les modifications post-traductionnelles indispensables à l'expression du biomédicament. C'est le cas des anticorps qui peuvent être produits en cellules de hamster (lignée CHO dérivée de cellules d'ovaire de hamster chinois), de souris (lignée Sp2/0 dérivée de cellules de myélome murin) et en cellules humaines (lignée PER.C6® dérivée de cellules de rétine humaine). Les cellules d'oiseaux peuvent aussi être exploitées (lignée EBx® dérivée de cellules de souches embryonnaires de canard). Ces lignées agissent en tant que « cellules

usines » dont l'unique objectif n'est autre que l'expression d'une protéine recombinante, c'est-à-dire du biomédicament.

Pour cela, l'approche conventionnelle pour obtenir ces cellules usines repose sur l'utilisation de systèmes d'expression stables créés par intégration chromosomique du transgène dans les cellules hôtes. Le gène d'intérêt thérapeutique est isolé puis inséré dans un vecteur d'expression (virus, plasmides, chromosomes artificiels...). Ce vecteur est ensuite optimisé en fonction de la lignée cellulaire pour avoir les meilleurs taux d'expression possibles. Par exemple, les plasmides sont utilisés pour les cellules usines bactériennes tandis qu'un vecteur baculoviral sera préféré pour intégrer le transgène dans une lignée cellulaire d'insecte *Spodoptera frugiperda*. Sont également associés à chaque lignée de cellules usines, des méthodes de transfection, des méthodes de culture, des systèmes de sélection et d'amplification... Pour résumer, l'objectif est de développer un procédé permettant d'avoir une expression optimale du biomédicament à une échelle industrielle et économiquement viable.

Cette optimisation maximale du procédé se fait dans le but de raccourcir le temps nécessaire pour fabriquer les lignées cellulaires transgéniques et les sélectionner, afin de stimuler la productivité du système d'expression. L'amélioration des taux d'expression repose sur l'optimisation de l'intégration du transgène dans la cellule hôte, de sa transcription, de sa traduction et enfin de son expression par la cellule usine. Par exemple, des séquences peuvent être intégrées dans le vecteur d'expression pour réduire l'effet de position du transgène. La plupart des sites d'intégration sont silencés et obligent un *screening* important de clones pour isoler ceux ayant un taux d'expression acceptable. D'autres méthodes portent sur l'utilisation de marqueurs de sélection améliorés et optimisés, de promoteurs forts (promoteur de cytomégalo virus murin...), de la séquence de codons optimale et sur l'optimisation de la structure de l'ARN. Pour résumer, le procédé global de bioproduction est composé d'une multitude de technologies qui nécessitent d'être judicieusement sélectionnées et optimisées dans le but de répondre à des exigences strictes, tels que des contraintes de temps, des exigences réglementaires ou des droits de PI.

1.3.3. Perspectives du secteur des biomédicaments et nouvelle évolution de la situation de la PI

Depuis une quinzaine d'années, le secteur biopharmaceutique est devenu stratégique et incontournable pour la mise au point de produits thérapeutiques innovants. Associées à une meilleure connaissance des organismes vivants et une meilleure compréhension des pathologies, les biotechnologies ont été à l'origine d'une nouvelle génération de médicaments : les biomédicaments. Progressivement, un grand nombre de sociétés pharmaceutiques se sont orientées vers le développement de biomédicaments dont certains sont déjà qualifiés de *blockbusters* depuis quelques années (Figure 6). À l'exception du produit Enbrel® (traitement de l'arthrite rhumatoïde sévère), les chiffres d'affaires les plus importants sont observés pour la classe des anticorps, suivie par les facteurs de croissance (protéines).

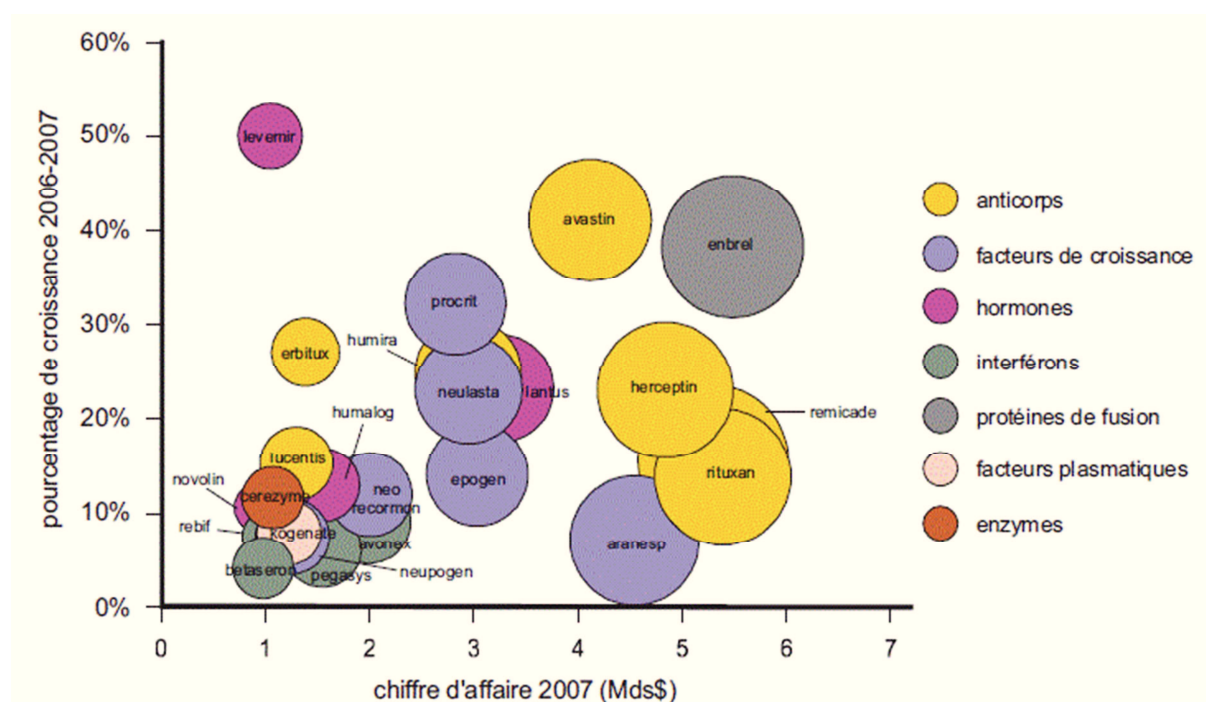


Figure 6 : Pourcentage de croissance et chiffre d'affaires de biomédicaments. La couleur indique les classes de biomédicaments. La taille du cercle représente leur part de marché (Source : Site internet des entreprises leaders, Développement & Conseil, 2008).

Vingt-cinq pour cent des *blockbusters* actuels sont des biomédicaments. Par exemple, la protéine de fusion Enbrel® d'Amgen, premier *blockbuster* des biomédicaments, a réalisé un chiffre d'affaires consolidé de plus de 5Mds\$ en 2007. Il en est de même pour les produits Herceptin® de Genentech, Rituxan® de Roche ou encore Remicade® de Shering-

Plough. Cependant, comme ce fut le cas pour les *blockbusters* issus de l'industrie chimique, les brevets protégeant certains biomédicaments arrivent bientôt à échéance : 2012 aux États-Unis et en Europe pour Enbrel®, 2013 aux États-Unis et en Europe pour Herceptin®. Au même titre que les médicaments issus de la chimie avec les génériques, les biomédicaments pourront être librement copiés à l'expiration des brevets correspondants et des médicaments biosimilaires arriveront sur le marché.

Le marché des médicaments biosimilaires est naissant (Figure 7), le premier biosimilaire (l'Omnitrope®, hormone de croissance) a été approuvé aux États-Unis en 2005 et en Europe en 2006. Les analystes estiment néanmoins qu'environ vingt-cinq biosimilaires seront disponibles sur le marché en 2012, associant un chiffre d'affaire d'environ 3,5 Mds\$.

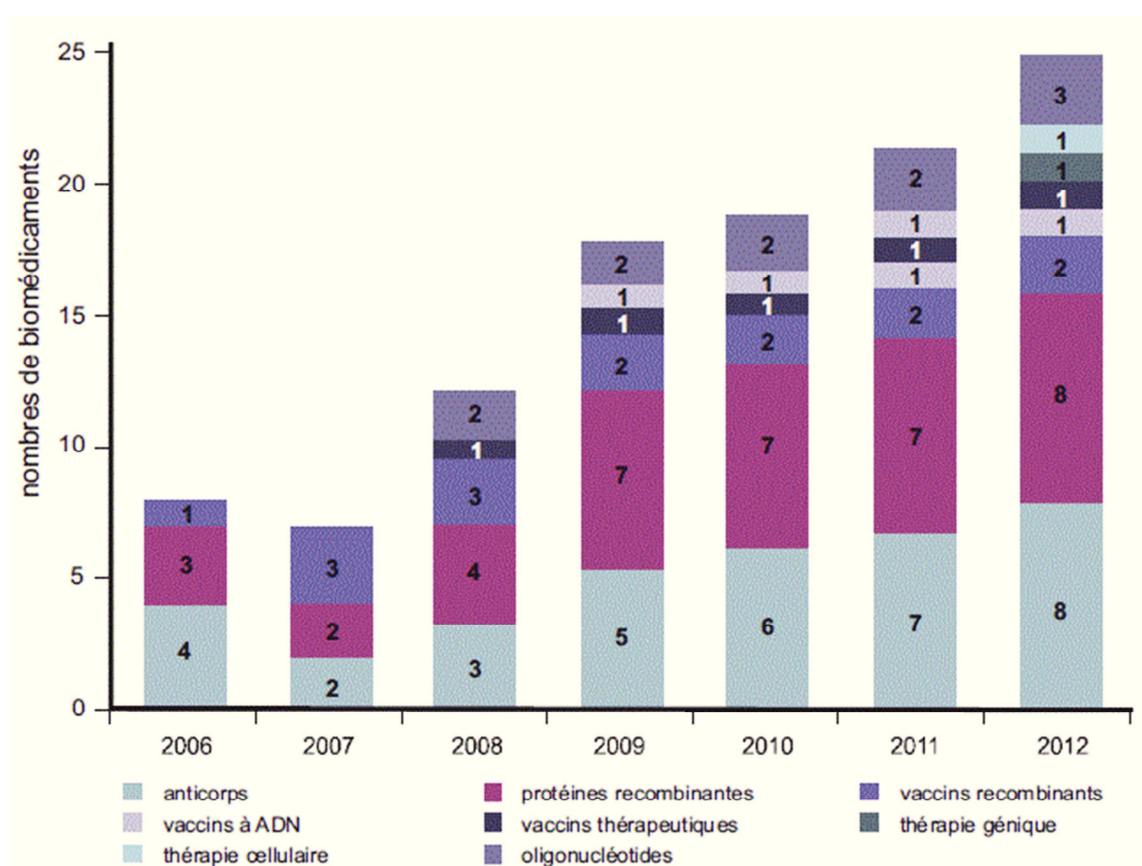


Figure 7 : Hypothèse haute du nombre de molécules biologiques qui sont ou seront commercialisées entre 2006 et 2012 (Source : Développement & Conseil, 2008).

Il est à noter que la terminologie de « biosimilaire » est préférée à celle de « biogénérique » car la reproduction exacte du médicament « vivant » est techniquement impossible. En conséquence, l'homologation d'un biosimilaire se fait sur la base d'une équivalence de résultats thérapeutiques basée sur des études cliniques englobant des

centaines de patients, et non pas uniquement sur la base d'une équivalence pharmaceutique (comme c'est le cas pour les médicaments génériques).

Dès 2004, les autorités européennes ont mis en place des règles propres aux biosimilaires (Directive 2004/27/CE du 31 mars 2004) alors qu'aux États-Unis, la réglementation des biosimilaires n'est pas clairement établie. L'article 10-4 de cette Directive définit la notion de biosimilaire :

Les médicaments biologiques similaires à des médicaments de référence ne remplissent habituellement pas toutes les conditions pour être considérés comme des médicaments génériques, en raison notamment des caractéristiques des procédés de fabrication, des matières premières utilisées, des caractéristiques moléculaires et des modes d'action thérapeutiques. Lorsqu'un médicament biologique ne remplit pas toutes les conditions pour être considéré comme un médicament générique, les résultats d'essais appropriés devraient être fournis afin de satisfaire aux conditions relatives à la sécurité (essais précliniques) ou à l'efficacité (essais cliniques), ou aux deux.

Le producteur d'un médicament biosimilaire doit prouver que son produit a le même effet thérapeutique que le produit biologique de référence. À ce titre, l'EMA (*European Medicines Agency*) a publié des lignes directrices pour la comparaison des biomédicaments en décembre 2003 et en septembre 2005 pour la bioéquivalence. Une étude récente du Comité Biotechnologies du LEEM et du Genopole® sur l'état des lieux de la bioproduction en France en 2008, estime que la somme liée à la perte des brevets de biomédicaments est estimée à plus de 10Mds\$ en 2010 et pourrait atteindre 20Mds\$ en 2015. Elle cite à titre d'exemple que les brevets sur les EPO expirent en moyenne aux USA vers 2013, ouvrant un marché potentiel de 8Mds\$ pour les biosimilaires.

Les évolutions du secteur pharmaceutique peuvent s'observer au travers des brevets. En effet, comme il a été précisé dans les parties précédentes, le biomédicament en tant que produit thérapeutique n'est pas le seul élément brevetable. Sont également protégeables par brevet les procédés de production, les lignées cellulaires utilisées dans le cadre de ces procédés ou encore les méthodes de culture en fermenteur, les vecteurs d'expression, les

méthodes d'extraction... L'arrivée des biomédicaments sur le marché et le fait qu'un biomédicament puisse être protégé par plusieurs brevets (brevets de produit, de procédé et d'utilisation) expliquent les observations faites en matière de dépôts de brevets liés aux biomédicaments. La classification des brevets couvrant les biomédicaments fait apparaître cinq catégories (Figure 8) : les peptides et protéines, les vaccins et tout élément pouvant servir à faire de l'immunothérapie comme les anticorps, les médicaments issus des thérapies génique et cellulaire, les acides nucléiques ou les substances provenant de la matière biologique (les procédés relatifs à chacun des différents éléments sont compris dans la catégorie correspondante).



Figure 8 : Pourcentage des brevets en fonction de la classe des biomédicaments (étude intitulée *Du médicament au biomédicament : tendances de l'innovation*, INPI, 2008).

Une étude de l'Institut National de la Propriété Industrielle (INPI) datée d'octobre 2008 montre l'importance des dépôts relatifs aux biomédicaments. Basée sur une analyse des demandes de brevet français déposées entre 1997 et 2004, la France se place en 5^{ème} position après les dépôts de demandes internationales, les demandes américaines, européennes et japonaises et 39% des demandes liées à un médicament couvrent un biomédicament. Parmi elles, environ 40% sont des demandes relatives à un peptide ou une

protéine, 20 à 25% couvrent un vaccin ou un anticorps et un peu moins de 20% sont relatives à la thérapie génique.

Bouleversant le schéma classique de R&D de la pharmacopée traditionnelle, les acteurs de ce secteur ont dû rapidement s'adapter et se positionner au sein d'un environnement en pleine mutation. Les brevets jouent donc un rôle essentiel dans l'économie et le développement des biotechnologies et des biomédicaments. Ils sont et resteront au cœur de la « bioéconomie », définie par l'OCDE comme un système dans lequel les biotechnologies assureront une part substantielle de la production économique.

C'est pour rester compétitif et pour valoriser ses programmes de R&D que le laboratoire GICC s'est intéressé à l'information brevets et a même établi une cellule de veille stratégique. Ce genre de démarches étant très rarement développé parmi les laboratoires de recherche publics, nous avons décidé d'établir une collaboration avec la société France Innovation Scientifique et Transfert, spécialisée dans le transfert de technologies innovantes et l'analyse de portefeuilles de brevets.

1.4. Collaboration avec la société France Innovation Scientifique et Transfert (FIST)

1.4.1. Présentation de l'UMR 6239 CNRS Génétique-Immunothérapie-Chimie & Cancer (GICC) et de la Cellule StratéGICC

Le projet initial du GICC pendant mes deux premières années de thèse (2007-2009) était de se positionner comme acteur majeur d'un secteur associant l'ingénierie génétique et l'immunothérapie par anticorps. Les compétences réunies au sein du laboratoire ont permis d'appréhender différents points clés du développement des biomédicaments. Chercheurs, cliniciens, ingénieurs et techniciens ont unis leurs efforts dans la réalisation de projets allant de la recherche fondamentale au développement préclinique.

L'interdisciplinarité du projet scientifique, établie comme ligne directrice du GICC, permettait également le développement d'autres voies de valorisation qui associaient tout

ou partie de l'UMR. L'enjeu était de réunir les conditions propices pour optimiser le fonctionnement de la pluridisciplinarité en favorisant l'innovation par le traitement de tout type d'information disponible puis en assurant son exploitation au sein même du GICC. Le laboratoire a adopté une politique de recherche collaborative avec un réseau étendu de partenaires. À l'échelle européenne, Yves Bigot (Directeur de l'UMR jusqu'à fin 2009) a coordonné le projet SyntheGeneDelivery regroupant huit partenaires dont deux industriels. En outre, l'équipe 2 *Cellules usines, ingénierie des protéines et bio-informatique* a participé aux activités du Réseau d'Excellence européen Clinigene (6^{ème} PCDRT), qui rassemble une communauté de vingt-huit responsables d'équipes de recherche de renommée internationale en sciences et en médecine, impliquées dans le développement du transfert de matériel génétique à des fins thérapeutiques. Pour s'inscrire dans une dynamique scientifique nationale, l'UMR participe à trois Groupements De Recherche (GDR), des réseaux développés et soutenus par le CNRS pour favoriser l'émergence de projets de recherche pluridisciplinaires. Depuis 2009, Hervé Watier, à l'époque en charge de l'équipe 6 *Immunogénomique et anticorps thérapeutiques*, coordonne le GDR ACCITH - *Anticorps et ciblage thérapeutique* qui rassemble les acteurs majeurs français de ce secteur multidisciplinaire (65 partenaires dont un tiers de privés). Par ailleurs, le GICC était aussi actif au sein de plusieurs structures : le Cancéropôle Grand Ouest (coordination d'un réseau structurant MabiImpact) et les pôles de compétitivité *Atlanpole Biotherapies* et *Cosmetic Valley*.

La volonté de faire entrer au laboratoire l'innovation comme moteur de la recherche a été portée par toute l'équipe dirigeante du GICC et principalement son ancien directeur Yves Bigot. Forte d'un portefeuille de onze brevets, la PI a été au centre de la stratégie de développement du laboratoire. Avec le soutien de la Direction de la Politique Industrielle du CNRS, la cellule de développement stratégique dénommée « StratéGICC », dans laquelle j'ai été intégré, a été mise en place au sein même de l'UMR en mars 2008. Delphine Vandame, ingénieur Valorisation, a alors été recrutée pour prendre en charge la responsabilité de cette cellule. Sa mission initiale était d'élaborer la stratégie de valorisation du laboratoire afin de créer des conditions favorables à la mise en place d'une structure mixte public-privé à partir des différentes démarches entreprises par le GICC [w8]. Sa mission finale était d'apporter un

soutien à la recherche en s'appuyant sur l'innovation et le transfert de technologies au travers de quatre activités : la veille technologique, la création d'entreprises innovantes, la valorisation de la recherche et le management de projet (Figure 9).

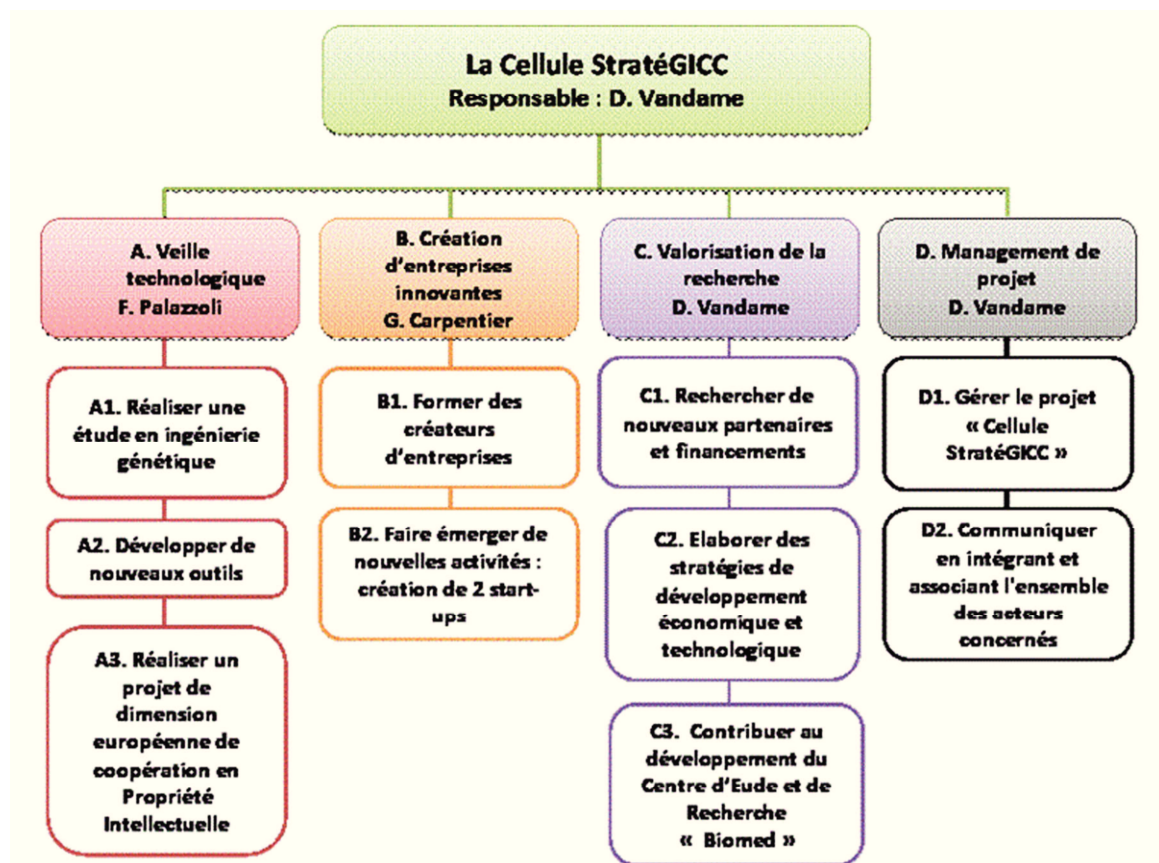


Figure 9 : Organigramme technique de la Cellule StratégICC au 1er janvier 2009. Il permet de visualiser l'organisation et les missions de la Cellule : veille technologique, création d'entreprises innovantes, valorisation de la recherche et management de projet.

À la suite de la démission d'Yves Bigot du poste de Directeur du GICC et de l'abandon des activités de StratégICC par la nouvelle équipe de direction du GICC, la Cellule a été dissoute à la fin de l'année 2009, soit au début de ma troisième année de thèse.

1.4.2. Présentation de la société partenaire FIST

Créée en 1992, FIST est une filiale de droit privé du CNRS (70%) et d'Oseo-Innovation (30%). C'est une société de transfert et de commercialisation de technologies innovantes comprenant une quarantaine de personnes [w9]. En étroite liaison avec la Direction de l'Innovation et des Relations avec les Entreprises du CNRS, la société effectue en son nom les

dépôts de brevets, le courtage technologique et le portage de son capital dans les jeunes pousses issues de ses laboratoires. Bénéficiant de dix-huit années d'expérience dans la gestion du portefeuille de brevets du CNRS, FIST est un acteur majeur de la valorisation de la recherche en Europe (valorisation de 1500 technologies, gestion d'un portefeuille de plus de 3200 brevets prioritaires et 800 accords d'exploitation...). La mission de FIST consiste à guider ses clients dans leur stratégie et leurs opérations de transfert de technologies et de commercialisation de droits d'exploitation de PI. Autrement dit, FIST apporte un conseil aux propriétaires d'innovations, qu'il s'agisse d'entreprises ou d'acteurs de la recherche publique, souhaitant capitaliser sur leur PI.

FIST est ainsi la plus importante structure française et européenne de ce type par le nombre de projets instruits et d'accords conclus. Structure incontournable en matière de valorisation de la recherche, FIST propose différents services à ses clients : définir la stratégie de PI et de valorisation des technologies, identifier les exploitants potentiels et négocier avec eux les contrats d'exploitation et optimiser la gestion des portefeuilles de brevets et de contrats d'exploitation.

Dans ce but, la société s'organise principalement autour de trois directions :

- La Direction Juridique et Propriété Industrielle : elle est en charge de la gestion administrative des dossiers de valorisation sur les plans juridiques et de la PI, c'est-à-dire du suivi de l'exécution des accords d'exploitation ;
- La Direction Transfert & Licensing : elle offre tous les services nécessaires à la réussite du transfert de technologie et à la commercialisation des innovations des clients de FIST. Elle intervient depuis l'évaluation des technologies jusqu'à la conclusion d'un accord de transfert ;
- La Direction Marketing Brevets & Licences : elle réalise des études sectorielles des portefeuilles de brevets et gère le suivi de l'exploitation des contrats. Elle fournit des audits de portefeuilles de brevets et d'innovations pour une meilleure maîtrise de leur environnement concurrentiel.

1.4.3. Présentation de la Direction Marketing Brevets & Licences (MB&L)

La Direction MB&L a été créée en 2004 pour accompagner les clients de FIST dans l'analyse de leur portefeuille et de leur environnement brevets. Elle propose des paysages brevets qui facilitent les prises de décisions relatives à la gestion d'un portefeuille de brevets. Ces analyses compétitives dressent un panorama global du paysage brevets dans un domaine d'activité [w10] et permettent de visualiser l'évolution des dépôts dans le domaine, détecter les experts, les équipes émergentes et les partenaires potentiels, comprendre les réseaux de collaboration existants et leur dynamique. Ces analyses sectorielles de portefeuille de brevets et d'innovations sont un moyen pour identifier les acteurs pertinents d'un secteur susceptibles d'être intéressés par l'une des innovations que FIST valorise.

La Direction MB&L comprend une équipe de chargés d'affaires scientifiques spécialisés dans l'analyse de brevets. Elle a développé une connaissance des stratégies de protection dans différents secteurs industriels grâce à la mise en place des outils dédiés à l'analyse de gros portefeuilles de brevets (~5000 documents brevets). Par exemple, FIST a réalisé des études sectorielles dans des domaines tels que la cancérologie, l'imagerie médicale, l'optique télécom, le photovoltaïque, le système nerveux central, la vectorisation, les nanomatériaux, les batteries, les solvants verts.... [Duparcq, 2008]. FIST propose également à ses clients la possibilité de demander l'analyse personnalisée d'un secteur particulier.

1.4.4. Contexte et historique de la collaboration

Yves Bigot (ancien Directeur du GICC, 2007-2009), les membres de la cellule StratéGICC et la Direction Transfert & *Licensing* de FIST entretiennent de bonnes relations depuis plusieurs années, à la suite de plusieurs projets de développement technologique portant sur de potentiels transferts de technologies développées au laboratoire. Par exemple, le projet ANR Magic Transpo s'appuyait sur la PI protégeant la transposase *MOS1* développée par le GICC (WO2004078981, WO2007063033, WO2007132096) pour faire du

transfert d'ADN *in vitro* dans un but d'exploitation industrielle (mutagenèse par insertion, clonage et séquençage). Yves Bigot (GICC) et Franck Merly (FIST) ont co-dirigé de janvier à juin 2007 mon stage de Master 2 Biotechnologies et Droit intitulé *Évaluation des conséquences de l'utilisation de domaines protéiques de type doigt de zinc, dans des protéines de fusion d'intérêt économique*. L'objectif était de déterminer l'impact de la PI existant sur les mécanismes d'intégration non-viraux dérivés de transposons et de leur modification pour devenir site-spécifique, sur la viabilité économique d'une stratégie de développement technologique en thérapie génique *ex vivo*. L'analyse du paysage brevets de la technologie basée sur les domaines de liaison à l'ADN de type doigt de Zinc a montré que l'exploitation de cette technologie, dont le marché est estimé à plusieurs milliards de dollars, était totalement bloquée par la société Sangamo Biosciences. Ce monopole a d'ailleurs fait l'objet deux ans après d'une publication qui a confirmés les résultats du stage et ceux publiés entre temps [Chandrasekharan *et al.*, 2009]. Les résultats obtenus ont démontré la nécessité pour le GICC d'étendre les travaux à l'ensemble de ses thématiques de recherche, et notamment sur des technologies utilisées en bioproduction et en thérapie génique. Cependant, la construction de paysages brevets a impliqué l'utilisation d'outils développés spécialement pour l'analyse d'importants portefeuilles de brevets, que nous n'avions pas au laboratoire. Par conséquent, le GICC a alors pris contact avec la Direction MB&L de FIST à la suite de l'article *Les brevets, une valeur en hausse* [Duparcq, 2008]. Cet article présente les nouvelles orientations de FIST comme les études sectorielles de PI réalisées par la Direction MB&L. Après plusieurs rencontres et discussions à la fin de l'année 2008 et au début de l'année 2009, un projet de collaboration entre FIST et le GICC a été rapidement évoqué à propos d'études sectorielles de PI en biotechnologies. La thèse *Exploitation de l'information brevets dans un laboratoire de recherche public : identification de niches de développement technologique en bioproduction et en thérapie génique* s'inscrit donc dans cette démarche. Coordonnée par Yves Bigot pour la partie scientifique et François-Xavier Testu pour la partie juridique, l'objectif principal a été de localiser les niches de développement technologiques les plus viables en termes de promesses technologiques et de financements.

1.4.5. Modalités de la collaboration

La collaboration comprenait deux types d'objectifs. Le premier portait sur la réalisation d'études sectorielles de PI sur les thématiques de recherche du GICC (Génétique, Immunothérapie, Chimie et Cancer) qui intéressaient également FIST. Plus précisément, les études devaient aborder des technologies spécifiques mises en œuvre dans les systèmes de bioproduction (vecteurs d'intégration, éléments de contrôle de la chromatine, lignées cellulaires...). Le second correspondait à un travail de réflexion sur l'exploitation, l'évolution et l'optimisation des études sectorielles de PI telles que les méthodologies à mettre en place pour la recherche de niches de développement technologique ou pour l'identification des brevets fondateurs et des brevets d'application d'une technologie.

L'un des avantages de cette collaboration est que les deux partenaires présentaient une bonne complémentarité (Figure 10).

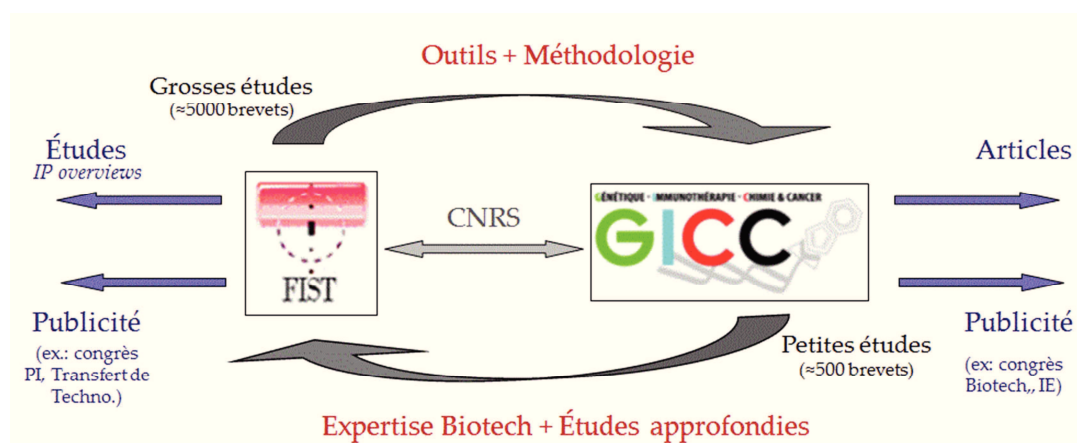


Figure 10 : Schéma présentant la collaboration et les caractéristiques des partenaires.

L'apport de FIST comprenait une méthodologie éprouvée : la définition des requêtes utilisées pour la recherche de documents brevets dans les bases de données, l'exploitation des résultats et une formation sur les outils d'extraction et d'analyse dont la société dispose (QPAT pour l'extraction de brevets et Intellixir pour l'analyse des résultats). Le GICC a donc profité d'un savoir-faire reconnu et d'outils commerciaux pour analyser l'information bibliographique concernant d'importants portefeuilles de brevets en bioproduction et en thérapie génique. L'apport du GICC portait sur le renforcement des compétences scientifiques de la Direction MB&L, à travers ses expertises en Génétique, Immunothérapie,

Chimie et Cancer, pour la réalisation et l'interprétation des paysages brevets. Par exemple, la collaboration a pris tout son sens lors de la réalisation d'un paysage brevets en immunothérapie (*Monoclonal antibodies against digestive system cancers* - 2010/09). Ce domaine multidisciplinaire et très vaste nécessite d'importants travaux de définition et de découpage en domaines, sous-domaines... Par ailleurs, plusieurs résultats issus de la collaboration ont été publiés sous la forme d'articles (deux articles présentés dans la section 3. Résultats), de posters (deux posters, Annexes 2 et 3) et sous la forme d'études *IP Overviews* de la même forme que celles déjà proposées par FIST. En outre, chaque partenaire a pu profiter de la communication réalisée autour de cette collaboration (remerciements dans les communications orales lors des congrès, Annexe 4 et autres articles, Annexe 5) pour valoriser ses travaux.

Pour résumer, la collaboration avec la Direction MB&L a été un tournant décisif pour les deux dernières années de thèse, grâce à l'utilisation d'outils très puissants et à un important savoir-faire. Cela m'a permis de passer de l'utilisation d'outils libres et limités à des outils commerciaux utilisés par les professionnels de l'information brevets pour proposer des études de brevets plus pertinentes et exhaustives. Ces avantages sont développés dans la section suivante 2. Matériel et méthodes.

2. Matériel et méthodes

2.1. Présentation de la méthode

Les recherches de brevets peuvent être exploitées à travers plusieurs applications : un état de l'art technologique, une étude de brevetabilité, une étude de liberté d'exploitation ou un paysage brevets. Leur objectif est de trouver les documents brevets pertinents sur un sujet qui a été préalablement défini. Dans le cas du GICC, nous avons principalement constitué des paysages brevets de technologies en rapport avec le transfert de gène. Par exemple, nous avons construit le paysage brevets sur les transposons, une méthode de vectorisation développée au laboratoire depuis bientôt dix ans.

L'élaboration d'un paysage brevets d'une technologie se compose de différentes étapes successives (Figure 11).

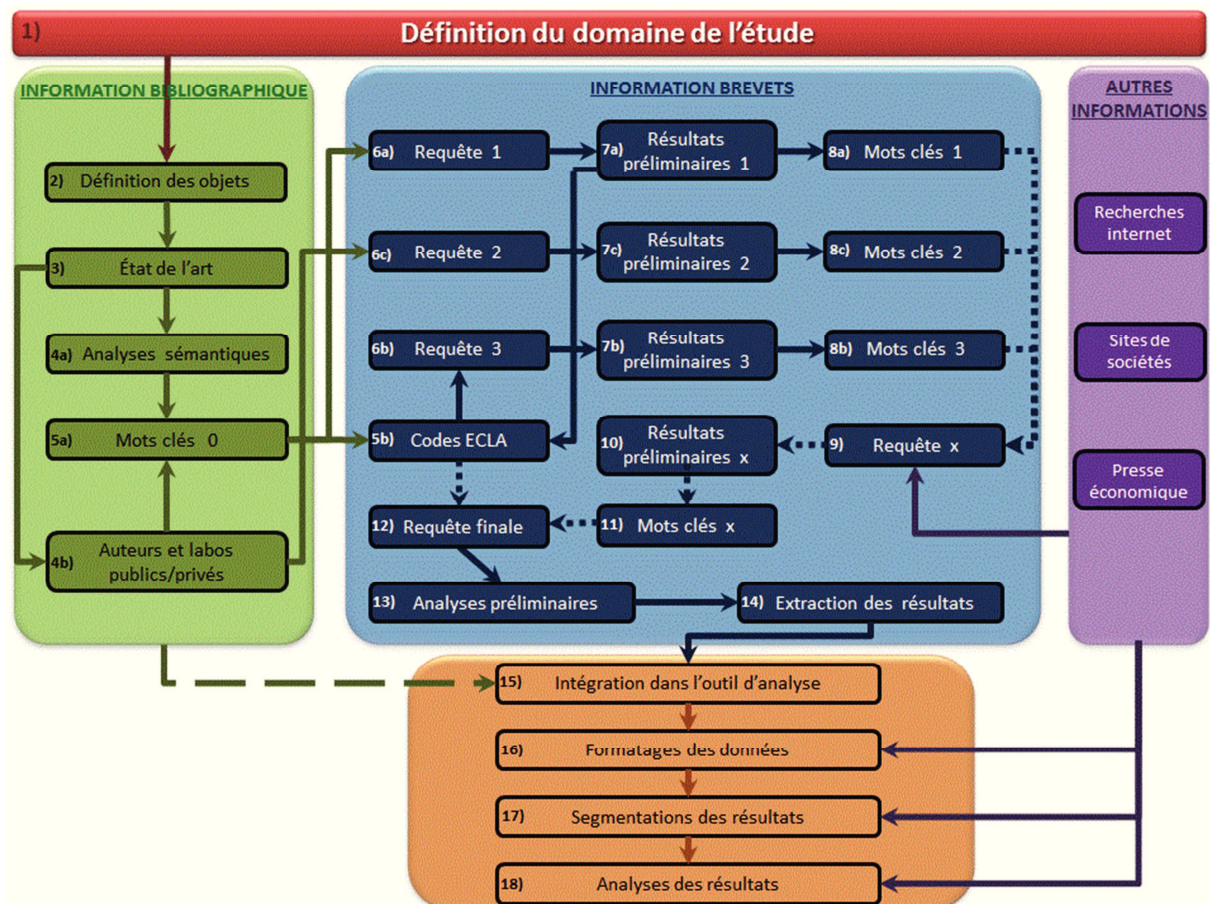


Figure 11 : Schéma des étapes successives de recherche d'informations et d'analyse pour la réalisation d'un paysage brevets. La requête est construite à partir de mots clés définis par les littératures bibliographiques et brevets et de symboles de classifications. Les résultats sont extraits de la base de données brevets puis intégrés dans l'outil d'analyse pour être formatés et segmentés.

Je tenterai ci-après de définir brièvement chaque étape intervenant dans l'élaboration d'un paysage brevets pour ensuite approfondir les phases les plus complexes.

Définition du domaine sur lequel l'étude doit être réalisée (étape 1 de la Figure 11) : le domaine est délimité par les chercheurs et les ingénieurs du laboratoire ou du service impliqués dans le développement technologique du projet. Par exemple, le domaine peut porter sur le transfert de gène ou la recombinaison homologue...

Les étapes 2, 3, 4a, 4b et 5a de la Figure 11 sont effectuées à partir de bases de données bibliographiques.

Définition des objets du domaine (étape 2 de la Figure 11) : le domaine est ensuite segmenté en « objets » (technologies, applications...) en concertation avec les chercheurs impliqués et à l'aide d'une base de données bibliographique comme PubMed [w11]. Il est important de répertorier chaque segmentation du domaine. En effet, un paysage brevets peut porter sur l'ensemble des objets mis en œuvre dans le domaine ou sur un seul d'entre eux. Cette segmentation peut être plus ou moins détaillée. Dans l'exemple précédent sur le transfert de gène, les vecteurs se scindent en deux catégories : viraux et non viraux. À leur tour, les vecteurs non viraux rassemblent entre autres les endonucléases, les recombinaisons et les transposases...

Définition de l'état de l'art bibliographique (étape 3 de la Figure 11) : les objets du domaine et le domaine lui-même sont définis par des mots et des notions recherchés dans la base de données bibliographique PubMed [w11]. Cette étape consiste à examiner toutes les informations existantes sur le domaine et sur chaque objet dans la littérature scientifique : les différents mots clés et notions en rapport avec le domaine, les applications technologiques, les dates de publication des articles correspondants, voire les informations relatives aux intérêts financiers des auteurs.

Analyses sémantiques (étape 4a de la Figure 11) : une analyse sémantique des publications est effectuée à la suite de l'état de l'art bibliographique. L'objectif est de déterminer les différents sens des mots clés utilisés par les auteurs des articles afin d'appréhender le mieux possible les connaissances dévoilées par les auteurs. En effet, le sens des mots utilisés peut

varier selon l’auteur, sa culture, la date à laquelle il a été employé, selon l’application dans laquelle il est cité...

Identification d’acteurs clés (étape 4b de la Figure 11) : il s’agit des acteurs qui sont les plus actifs en termes de publication d’articles sur le domaine. Ces auteurs majeurs (auteurs étant à l’origine d’une invention ou découverte et/ou auteurs ayant beaucoup publié) et les laboratoires publics et privés majeurs dont ils sont originaires sont recensés. Cela permet d’avoir une première idée des acteurs actifs dans le domaine et c’est également un moyen pour découvrir des mots clés potentiels.

Définition d’une liste préliminaire de mots clés (étape 5a de la Figure 11) : une première liste de mots clés est constituée à la suite des étapes 4a et 4b (« mots clés 0 »). Elle comporte les mots clés et leurs synonymes identifiés dans la littérature bibliographique. Cette liste est préliminaire car d’autres étapes vont révéler de nouveaux mots clés qui s’y ajouteront.

Les étapes 5b, 6a-c, 7a-c, 8a-c, 9, 10, 11, 12, 13 et 14 de la Figure 11 sont réalisées avec une base de données brevets.

Définition de symboles de classification (étape 5b de la Figure 11) : les mots clés de l’étape 5a sont recherchés dans la classification grâce à la base de données de brevets esp@cenet [w12]. Les symboles de classification sont une stratégie de recherche très efficace pour couvrir tous les documents brevets d’un domaine et d’un objet si le symbole correspondant existe. Les documents brevets appartenant au(x) symbole(s) de classification ainsi identifié(s) peuvent alors apporter de nouveaux mots clés ou être directement intégrés dans la stratégie de recherche.

Construction de requêtes préliminaires (étape 6a-c de la Figure 11) : des requêtes préliminaires (« requêtes 1-3 ») sont testées dans une base de données gratuite comme esp@cenet [w12] ou commerciale comme QPAT [w13], à l’aide de combinaisons de mots clés définis aux étapes 4a, 4b, 5a et de symboles de classification identifiés à l’étape 5b. Pour l’élaboration d’un paysage brevets, la sélection de la base de données de brevets a un impact direct sur la quantité et la qualité des documents brevets résultant de la recherche.

Elle doit être choisie en tenant compte de ses limitations : couverture géographique, recherche en texte intégral... Pour la recherche dans une base de données brevets, une requête est construite à l'aide d'un ou de plusieurs critères de recherche judicieusement sélectionnés (mots clés, symboles de classification, dates, noms d'inventeurs...) qui sont combinés à l'aide de différents opérateurs et jokers de troncature. L'objectif de cette étape est d'obtenir un maximum de documents brevets pertinents et un minimum de documents sans intérêt pour l'étude.

Obtention de résultats préliminaires (étape 7a-c de la Figure 11) : les requêtes des étapes 6a-c aboutissent à des « résultats préliminaires ». Il s'agit des documents brevets qui ressortent avec les requêtes recherchant les mots clés préliminaires identifiés avec la littérature bibliographique (étape 5a), les symboles de classification (étape 5b) et les mots clés utilisés par les acteurs majeurs du domaine (étape 4b). Les résultats comportent des documents brevets pertinents et des documents brevets considérés comme des faux positifs qui sont par conséquent à éliminer pour le paysage brevets. Généralement, il s'agit de documents brevets citant les mots recherchés mais sans aucun rapport avec les objets ou le domaine défini auparavant. Plus le nombre de documents pertinents est élevé et plus le nombre de documents appartenant au « bruit de fond » est faible, donc meilleure est la requête. Celle-ci se rapproche alors de la requête finale utilisée pour élaborer le paysage brevets.

Identification de nouveaux mots clés préliminaires (étape 8a-c de la Figure 11) : l'analyse des résultats de l'étape précédente révèle souvent de nouveaux mots clés qui s'ajoutent à ceux déjà identifiés (« mots clés 1-3 »). L'objectif de cette étape est d'avoir une liste la plus exhaustive possible des mots clés et notions utilisés par les auteurs et inventeurs du domaine.

Construction de requêtes intermédiaires, obtention de résultats intermédiaires et identification de nouveaux mots clés intermédiaires (étapes 9, 10 et 11 de la Figure 11) : de nouvelles requêtes (« requête x ») sont construites avec les nouveaux mots clés, qui apportent à leur tour des résultats inédits (« résultats préliminaires x ») dont sont extraits de nouveaux mots clés (« mots clés x »), théoriquement en quantité moindre. Ces trois étapes sont réitérées tant que de nouveaux documents brevets en rapport avec le domaine

résultent des requêtes intermédiaires. Ces trois étapes récurrentes sont arrêtées lorsqu'on estime que de nouveaux documents brevets ne ressortiront plus avec la requête composée de mots clés et de symboles de classification.

Construction de la requête finale (étape 12 de la Figure 11) : la requête finale comporte une combinaison de mots clés et/ou de symboles de classifications. La requête idéale est celle qui donne le plus de documents brevets pertinents possibles et le moins de documents brevets sans aucun rapport avec le sujet.

Analyses préliminaires des résultats issus de la requête finale (étape 13 de la Figure 11) : les résultats de la requête finale sont analysés un par un dans le but d'écarter les documents brevets qui citent mais ne revendiquent pas les mots clés comme des éléments de l'invention (« analyses préliminaires »). Par exemple, le mot clé anglais « *transposon* » est exclu dans la demande PCT WO9910513 (« *the universal integration and expression vector of claim 2 which does not include a transposon* ») et n'est pas impliqué dans l'invention de la demande PCT WO9618372 (« *the kenamycin resistance transposon Tn903* »).

Extraction des résultats issus de la requête finale (étape 14 de la Figure 11) : l'ensemble des documents brevets résultant de la requête finale est extrait sous une forme appropriée permettant son analyse. Cette étape est réalisée à l'aide de l'outil de recherche QPAT qui autorise l'extraction d'un grand volume de données, comme cela peut être le cas pour les paysages brevets.

Les étapes 15, 16, 17 et 18 de la Figure 11 sont accomplies à l'aide d'un outil d'analyse et de traitement de l'information structurée tel qu'Intellixir [w14].

Intégration des résultats dans l'outil d'analyse (étape 15 de la Figure 11) : les résultats extraits précédemment sont intégrés dans un outil d'analyse et de traitement (Intellixir dans le cas du GICC). Dans la Figure 11, les résultats qui sont intégrés sont des documents brevets. Cependant, l'outil Intellixir accepte également l'intégration d'articles issus de bases de données bibliographiques. Ces publications se présentent aussi, comme les documents brevets, sous une forme structurée. Il est donc envisageable de construire des « paysages articles » (non abordés au cours de la thèse).

Formatage des données (étape 16 de la Figure 11) : les documents brevets ont l'avantage d'être présentés sous une forme normalisée, ils peuvent donc être aisément formatés à l'aide des filtres automatiques de l'outil d'analyse (déposants, inventeurs, dates de priorité...). Un second formatage manuel est néanmoins nécessaire pour homogénéiser les différents noms d'un même déposant, afin de tenir compte des opérations d'acquisitions et de fusions entre les acteurs du domaine. C'est donc une étape clé dans la méthodologie d'élaboration des paysages brevets. Le formatage des données, réalisé avec l'outil Intellixir, doit refléter le plus exactement possible la réalité économique du secteur et les relations entre ses acteurs.

Segmentation des résultats (étape 17 de la Figure 11) : cette étape porte sur la segmentation de la base de données de brevets constituée dans Intellixir. Celle-ci peut être partagée selon plusieurs critères : les objets définis à l'étape 2 (technologies, applications des brevets...), la protection géographique, la qualité des déposants (industriels ou institutionnels)... et même selon un critère défini par l'utilisateur.

Analyses des résultats (étape 18 de la Figure 11) : dans cette dernière étape, les fonctionnalités de l'outil d'analyse Intellixir facilitent l'analyse des résultats grâce à un large éventail de représentations graphiques : tableaux, histogrammes, graphiques... qui permettent d'apprécier le paysage sous différents angles et ainsi d'avoir une analyse la plus fine possible.

Dans la section suivante sont reprises et détaillées les étapes les plus importantes pour l'élaboration d'un paysage brevets :

- la formulation de la stratégie de recherche :
 - l'identification de mots clés à partir des publications scientifiques et des documents brevets,
 - l'identification de symboles de classification ECLA à partir de mots clés,
- le choix de la base de données de brevets,
 - les bases gratuites esp@cenet et PATENTSCOPE,
 - la base commerciale FamPat exploitée par l'outil de recherche QPAT

- le formatage et l'analyse des résultats par l'outil d'analyse et de traitement Intellixir,
- la recherche d'informations complémentaires liées à l'industrie des biotechnologies.

2.2. Formulation de la stratégie de recherche

Plusieurs stratégies peuvent être employées pour rechercher des documents brevets dans les bases de données. En effet, il existe plusieurs critères de recherches : par mots clés, numéros de publication, numéros de demande, numéros de priorité, dates de publication, noms d'inventeur, noms de déposant et symboles de classification. Pour la réalisation des paysages brevets, nous avons principalement utilisé les recherches par mots clés et symboles de classification. Dans un premier temps, il est nécessaire de définir les limites du sujet de l'étude. Pour cela, un travail bibliographique est une étape préliminaire indispensable afin de formuler la stratégie de recherche. Après avoir défini le domaine sur lequel porte le paysage brevets, un état de l'art est réalisé grâce à la base de données bibliographiques PubMed [w11]. Cette étape doit permettre de déterminer le nombre d'objets du domaine.

2.2.1. Identification de mots clés à partir de la littérature bibliographique (étapes 4a, 4b et 5a)

L'analyse d'articles scientifiques sur le domaine permet de construire une première liste de mots clés qui seront utilisés pour la recherche de documents brevets. Le contenu technique des articles scientifiques et des documents brevets est défini par des combinaisons de concepts et de mots clés qui peuvent varier d'un document à un autre. Afin d'être le plus exhaustif possible dans la sélection des mots clés, il est important d'examiner des publications d'auteurs différents, publiées à des dates différentes, voire dans des journaux différents. Par exemple, en s'intéressant à des articles publiés sur une période de temps la plus large possible (de la découverte à l'époque actuelle), un maximum de mots clés devrait être recensé. En suivant l'évolution des citations de la littérature scientifique au

cours du temps, de nouveaux mots clés, qui ne sont plus utilisés actuellement, peuvent être mis en évidence. En effet, il est possible que les auteurs attribuent une première dénomination à leur découverte ou invention, avant qu'elle n'évolue au cours du temps. C'est le cas du transposon *piggyBac* [Fraser et al., 1996] initialement dénommé « IFP2 » [Cary et al., 1989] ou encore des *Locus Control Regions* qui étaient originellement qualifiées de *Dominant Control Regions* [Grosveld et al., 1987].

Par ailleurs, les initiales et acronymes, dont les utilisations sont courantes en biologie, peuvent également être employés en tant que mots clés. Dans ce cas, il convient de les associer à des mots clés complémentaires tels que « séquences, ADN, locus... ». C'est le cas de la demande PCT du transposon *Sleeping Beauty* (WO9840510) qui ne cite à aucun moment « *Sleeping Beauty* » dans les revendications, le titre et le résumé, mais seulement ses initiales « SB » et le mot clé « transposase ». Il peut être également intéressant de noter que certains brevets possèdent des erreurs d'orthographe dans leurs textes. Ils peuvent être identifiés grâce à d'autres mots clés ou d'autres critères de recherche (noms d'inventeurs), et le mot mal orthographié peut éventuellement rentrer dans la liste de mots clés. C'est par exemple le cas de deux demandes de brevets sur les transposons : « transposan » est cité à la place de « transposon » dans la demande PCT WO9909817 – *Use of Mariner transposan in the production of transgenic animals*, et « piggyback » est utilisé à la place de « piggyBac » dans la demande CN101343638 - *Toxoplasma transgenosis carrier based on piggyback transposon and construction method*. Il s'agit probablement d'erreurs de logiciel de traduction provenant de documents brevets asiatiques ou russes.

2.2.2. Identification de mots clés à partir du web

La première liste de mots clés peut être complétée par d'autres en variant les sources d'informations comme des sites internet de sociétés, de laboratoires de recherche ou plus généraux (Wikipedia...). Ces sources d'informations sont complémentaires des bases de données bibliographiques et de brevets car elles peuvent aussi apporter de nouvelles informations et des mots clés qui n'étaient pas encore apparus. Par exemple, le site Wikipédia peut permettre de mieux appréhender le domaine de l'étude [w15]. En illustration, un article français à nature encyclopédique est disponible pour les éléments

transposables [w16]. S'il existe un article Wikipédia en rapport avec les mots clés recherchés, il est généralement dans les premiers résultats donnés par les moteurs de recherche généralistes comme Google [w17] ou par les métamoteurs comme Copernic Agent [w18]. Les métamoteurs sont un bon moyen pour couvrir le web car ils interrogent en parallèle plusieurs moteurs de recherche, combinent les résultats et éliminent les doublons. Concernant les moteurs de recherche généralistes (Google, Yahoo, Bing, Exalead, Ask...), ils possèdent leurs propres caractéristiques et ne donnent pas de résultats identiques [w19 ; w20]. Cela est dû à l'algorithme de recherche spécifique à chaque moteur de recherche qui aboutit à ses propres résultats et à son propre classement. C'est pourquoi il est possible de trouver des informations différentes à partir de la même requête sur des moteurs de recherche différents. En outre, ils comportent des limitations comme c'est le cas pour les bases de données brevets. Par exemple, Google ne gère par le joker de troncature « * » [w21] et recherche donc exactement le(s) mot(s) demandé(s). Pour rechercher « transposon(s), transposase(s), transposable(s), transposition(s)... », le moteur de recherche Google nécessite de rentrer chaque mot. À l'inverse, l'outil QPAT autorise l'utilisation du joker « * » et récupère tous ces mots avec la requête « transpos* » dans les documents brevets. Dans cette étape, il est important de varier les sources d'informations provenant du web et de faire concorder les nombreuses informations qui en résultent.

2.2.3. Identification de mots clés à partir de la littérature brevets (étapes 8a et 8c)

La première liste de mots clés est ensuite utilisée pour réaliser des recherches tests de documents brevets. L'objectif est d'avoir le plus de documents pertinents par rapport à l'étude et le moins possible de « bruit de fond », c'est-à-dire de documents comportant des mots clés recherchés mais dénués d'intérêt pour le paysage brevets. Une analyse de ces nouveaux documents pertinents révèle à son tour de nouveaux mots clés qui s'ajoutent sur la liste. Reprenant l'exemple précédent sur les *Locus Control Regions* ou *Dominant Control Regions*, une troisième appellation, *Dominant Activator Sequences*, a été identifiée dans la demande PCT WO8901517 – *Vector for integration site independent gene expression in mammalian host cells*. En appliquant de nouvelles requêtes tests qui tiennent compte de ces

modifications, le nombre de documents brevets pertinents doit théoriquement augmenter et/ou le nombre de documents brevets inadéquats doit théoriquement diminuer. La requête est ainsi optimisée par récurrence, au fur et à mesure des requêtes tests, de l'identification de nouveaux mots clés et de nouveaux documents brevets rentrant dans le cadre de l'étude.

De plus, il est également possible d'exploiter l'information issue des descriptions des inventions des demandes de brevets. En effet, les demandes de brevets exposent dans leur description un état de l'art technologique. Certains citent même des brevets importants comme dans la demande PCT WO2006048459 - *Selection of host cells expressing protein at high levels*. Ainsi, les paragraphes [0059] à [0070] de la description de l'invention présentent un état de l'art des éléments de contrôle de la chromatine et donnent plusieurs numéros de brevets : MARs/SARs (WO02074969), insulateurs (WO9423046), UCOE (WO0005393) et séquences STAR (WO03004704). La qualité de la requête peut être testée en comparant ces numéros de documents brevets avec les résultats qui ressortent de la requête testée. Si la requête est adéquate, les numéros des documents brevets précédents doivent théoriquement faire partie des résultats. Dans le cas contraire, les numéros de brevets qui n'apparaissent pas dans les résultats avec la requête apportent à leur tour de nouveaux mots clés.

En outre, les citations de brevets peuvent être utilisées dans le même but. Les examinateurs des offices de brevets rédigent des rapports de recherche sur les critères de brevetabilité de l'invention pour laquelle une demande de brevet est déposée. Ils font souvent référence à des numéros de brevet appartenant à l'état de l'art de l'invention en question, par exemple des numéros de brevet antérieurs qui détruisent la condition de nouveauté. Ces informations donnent donc une idée sur la qualité de la requête qui comporte ou non dans ses résultats les numéros de brevets identifiés par les experts.

Enfin, il convient de rappeler que, pour chaque mot clé identifié par les méthodes précédentes, peuvent être associés des synonymes qui doivent entrer dans la formulation de la requête. En effet, des documents brevets appartenant à un domaine technologique peuvent avoir des concepts identiques mais des mots clés différents : les mots clés sont subjectifs et plusieurs d'entre eux peuvent décrire un seul et même concept. De plus, dans l'idéal, les mots clés doivent être recherchés dans toutes les sources d'informations

disponibles et dans toutes les langues. Cela représente donc un effort supplémentaire à fournir dans la recherche d'informations. Par exemple, des acteurs chinois en biotechnologies déposent des demandes de brevets uniquement en Chine, car avec plus d'un milliard d'habitants, ce pays représente un marché qui se suffit à lui seul. Par conséquent, le fait de ne pas traduire la requête dans la (les) langue(s) chinoise(s) représente un « trou noir » non négligeable dans l'information technologique disponible, dont il faudra tenir compte lors de l'interprétation des résultats.

Finalement, avoir une liste de mots clés la plus exhaustive possible pour une recherche de documents brevets représente un exercice fastidieux qui s'incrémente nécessairement en plusieurs étapes d'optimisation. L'absence de nouveaux résultats pertinents est un indicateur de la qualité de la requête et le seuil d'exhaustivité peut être considéré comme ayant été atteint. De ce fait, les utilisateurs de l'information brevets auront une requête finale lorsqu'ils estimeront que de nouvelles recherches de mots clés n'aboutiront pas à de nouveaux résultats pertinents. Bien que les méthodes exposées ci-dessus permettent de diminuer le risque, la recherche par mots clés peut ne pas identifier de documents brevets pertinents (mot clé absent de la requête, erreur de traduction...). Pour pallier ce défaut, elle peut être remplacée ou combinée à une recherche exploitant les symboles du système de classification des brevets.

2.2.4. Généralités sur le système de classification des brevets

Les bases de données de brevets contiennent des volumes de données très importants. Une classification commune pour les brevets a été mise en place pour faciliter les recherches de documents brevets. L'Arrangement de Strasbourg de 1971 a établi un système hiérarchique et normalisé des documents brevets à l'échelle internationale : la Classification Internationale des Brevets, communément appelée CIB. La CIB offre l'avantage d'être exploitable par les experts des offices de brevets ou par toute personne souhaitant évaluer le caractère de nouveauté d'une invention, apprécier l'activité inventive ou l'état de l'art des connaissances techniques. Le guide d'utilisation de la CIB [OMPI, 2009] précise qu'elle représente :

- un instrument d'organisation méthodique des documents brevets afin de rendre aisément accessible l'information technologique et juridique qu'ils contiennent ;
- un moyen permettant la diffusion sélective de l'information à tous les utilisateurs de l'information en matière de brevets ;
- un outil de recherche de l'état de la technique dans des domaines déterminés ;
- un moyen favorisant l'établissement de statistiques de PI qui rendent à leur tour possible l'analyse de l'évolution des techniques dans divers secteurs.

Dans ce système hiérarchique, les documents brevets sont classés par des symboles ou codes, en fonction des domaines technologiques auxquels ils se rapportent. Pour cela, l'Office de brevets qui effectue la publication d'une demande de brevet, lui attribue un symbole de classement CIB. Au niveau hiérarchique le plus élevé, la CIB comprend huit sections qui sont notées d'une lettre allant de A à H :

- A : NÉCESSITÉS COURANTES DE LA VIE
- B : TECHNIQUES INDUSTRIELLES; TRANSPORTS
- C : CHIMIE; MÉTALLURGIE
- D : TEXTILES; PAPIER
- E : CONSTRUCTIONS FIXES
- F : MÉCANIQUE; ÉCLAIRAGE; CHAUFFAGE; ARMEMENT; SAUTAGE
- G : PHYSIQUE
- H : ÉLECTRICITÉ

Ces huit sections se scindent ensuite en plus de 70 000 sous-groupes qui comprennent chacune un symbole. Les sections sont divisées en classes (représentées par deux chiffres), qui sont à leur tour segmentées en sous-classes (une lettre). Puis celles-ci sont partagées en groupes (un nombre de un à trois chiffres), qui sont eux-mêmes décomposés en sous-groupes, dernier niveau hiérarchique de la CIB (un nombre de un à trois chiffres). Pour résumer, un symbole de classification CIB (Figure 12) est composé successivement de :

- 1 lettre représentant la section CIB,
- 1 nombre de 2 chiffres indiquant la classe CIB,
- 1 lettre indiquant la sous-classe CIB,
- 1 nombre de 1 à 3 chiffres indiquant le groupe principal CIB,
- 1 barre de fraction oblique (caractère slash « / »),
- 1 nombre de 1 à 4 chiffres indiquant le sous-groupe CIB.

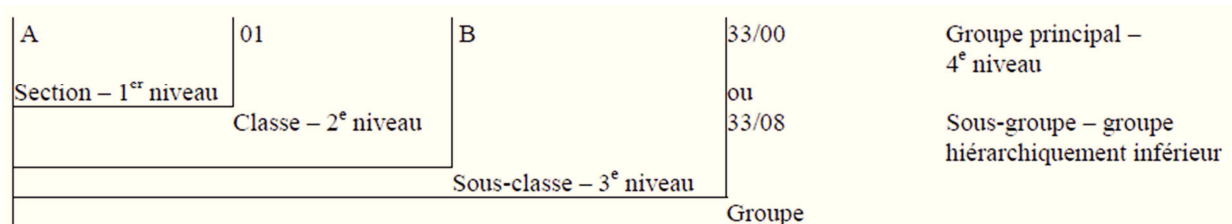


Figure 12 : Symbole complet de classement représentant un symbole CIB [OMPI, 2009].

Finalement, la CIB est un système hiérarchique qui peut être représenté sous la forme d'une arborescence (Figure 13).

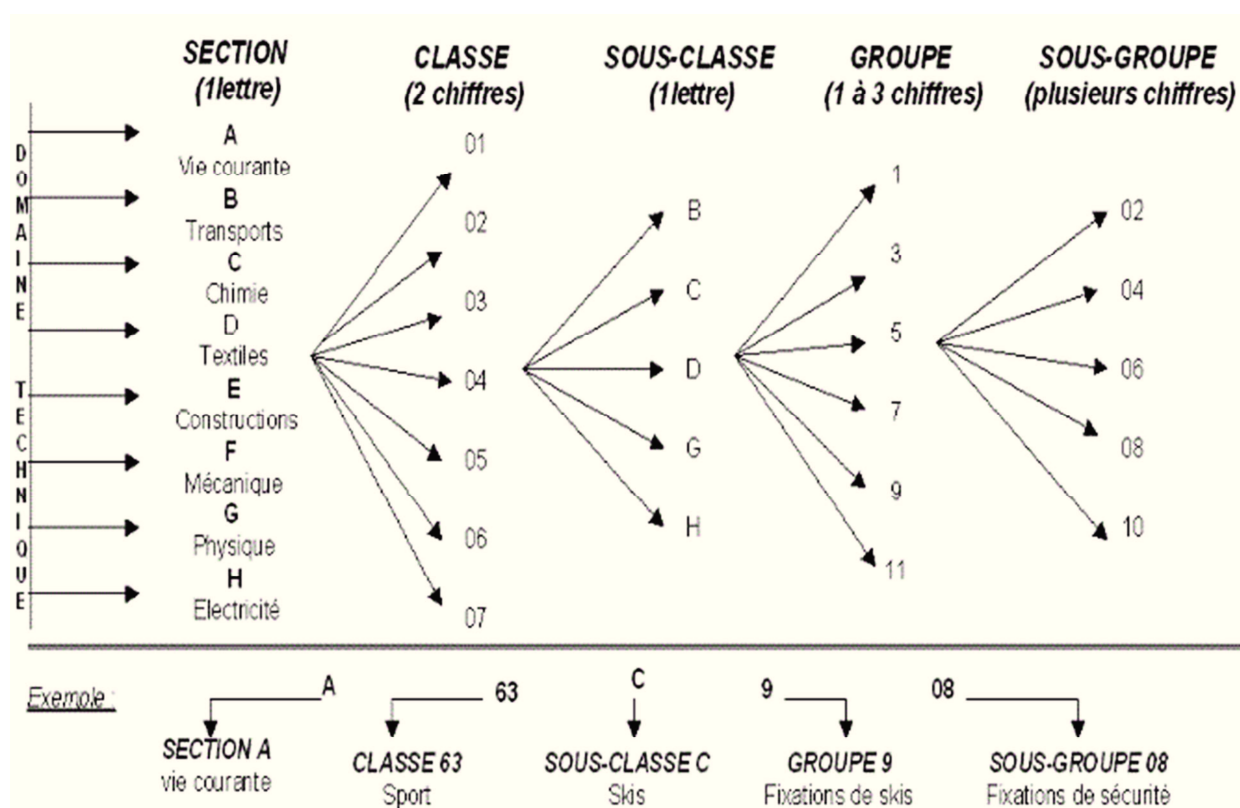


Figure 13 : Représentation graphique simplifiée du système de la CIB [w22].

L'OMPI classe les documents brevets des biotechnologies par plusieurs symboles de classifications (Annexe 6). Par exemple, le code C12N 15/00 porte sur les « *Techniques de mutation ou génie génétique; ADN ou ARN concernant le génie génétique, vecteurs, p.ex. plasmides, ou leur isolement, leur préparation ou leur purification; Utilisation d'hôtes pour ceux-ci (mutants ou micro-organismes modifiés par génie génétique C12N 1/00, C12N 5/00, C12N 7/00; nouveautés végétales A01H; reproduction de plantes par des techniques de culture de tissus A01H 4/00; nouvelles races d'animaux A01K 67/00; utilisation de*

*préparations médicinales contenant du matériel génétique qui est introduit dans des cellules du corps vivant pour traiter des maladies génétiques, thérapie génique A61K 48/00; peptides en général C07K) ». Cette définition illustre également le fait que les brevets peuvent appartenir à plusieurs symboles de classification. En effet, dans cet exemple, une invention peut être couverte par le groupe C12N15/00 ainsi que le groupe A61K 48/00 rassemblant les connaissances technologiques concernant les « *Préparations médicinales contenant du matériel génétique qui est introduit dans des cellules du corps vivant pour traiter des maladies génétiques; Thérapie génique* ».*

Depuis son entrée en vigueur en 1975, la CIB a fait l'objet de modifications à travers plusieurs révisions. Ces révisions sont nécessaires non seulement pour améliorer le classement mais aussi pour tenir compte de l'évolution des progrès technologiques. En dehors de la CIB (70 000 divisions), il existe d'autres classifications comme la classification américaine divisée en classes et sous-classes (*U.S. Patent Classification System*) ou le système japonais *F-term* qui complète la CIB. Cependant, ces classifications permettent de rechercher des documents brevets uniquement dans les bases de données américaine et japonaise. Une quatrième classification très utile pour la recherche de documents brevets est la classification européenne ou ECLA. Cette extension de la CIB comporte plus de 135 000 divisions [w23]. Elle est donc encore plus précise dans le classement des connaissances technologiques des documents brevets. La classification ECLA est divisée au-delà de la classification CIB. Ainsi, le symbole ECLA comporte le symbole CIB auquel s'ajoutent :

- une lettre indiquant le sous-groupe ECLA : C12N15/90B4...
- un chiffre suivi optionnellement d'une autre lettre : C12N15/85A1B...

En outre, la classification ECLA comprend une neuvième section, dont le symbole est Y, spécialement pour les développements innovants de technologies interdisciplinaires. Ainsi, le symbole Y01N est attribué à plus de 85 000 documents brevets sur les nanotechnologies. De plus, cette section supplémentaire s'ajoute à celles des classifications CIB et ECLA appartenant aux autres sections A-H déjà octroyées. Il est important de noter que les classifications ECLA sont généralement disponibles quelques mois après la date de publication. On estime qu'environ 90% des documents devant être classifiés sous ECLA se sont vus attribuer un symbole dans les huit mois suivant leur publication.

2.2.5. Identification de symboles de classification à l'aide de mots clés (étape 5b)

Deux situations peuvent être envisagées pour la recherche par symboles de classification. Premièrement, le symbole est déjà connu grâce à des documents brevets identifiés par une recherche de mots clés comme expliqué précédemment. La seconde possibilité est que le symbole n'est pas connu et doit donc être recherché, soit directement dans la classification, soit par le biais de mots clés. Dans ce cas, les mots clés permettent d'obtenir non pas des documents brevets mais des symboles de classification qui représentent les domaines technologiques auxquels se rapportent les mots clés utilisés. La base de données de brevets esp@cenet propose de trouver des classifications ECLA correspondant à des mots clés de la classification européenne. Par exemple, la recherche de symboles ECLA avec les mots clés « *matrix attachment regions* » aboutit au symbole C12N15/82B6 (Figure 14). Celui-ci est défini par : « *Reducing position variability, e.g. by the use of scaffold attachment region/matrix attachment region (SAR/MAR); Use of SAR/MAR to regulate gene expression* ».

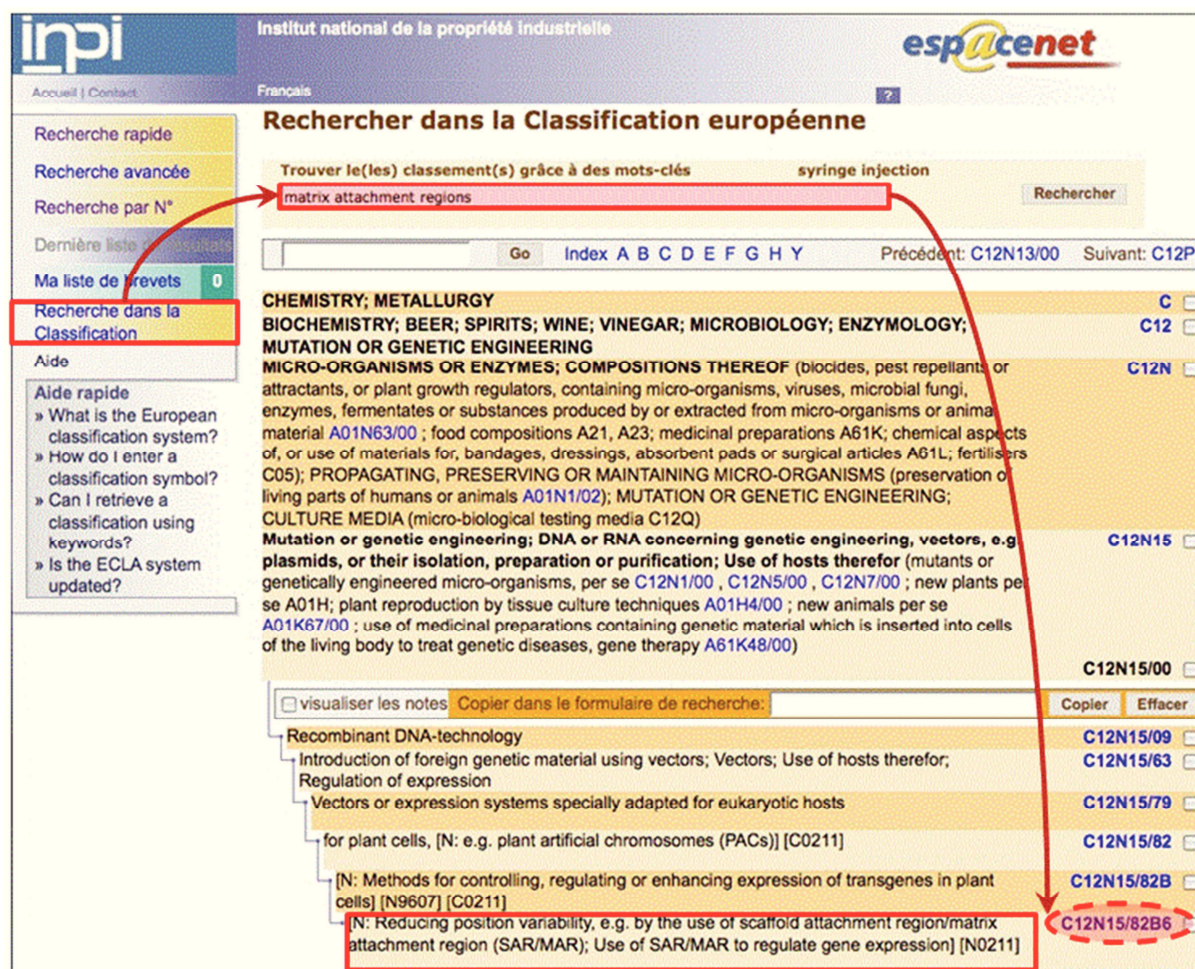


Figure 14 : Illustration d'une recherche de symbole ECLA à partir de mots clés. La requête « matrix attachment regions » aboutit au symbole de classification C12N15/82B6.

Dans le système hiérarchique de symboles de classification ECLA, le symbole C12N15/82B6 rassemble tous les documents brevets portant sur les MAR qui sont exploités chez les plantes (les symboles CIB C12N15/82 et ECLA C12N15/82B impliquent les plantes et les cellules de plantes). C'est le cas de la demande PCT WO0032800 - *Artificial matrix attachment region for increasing expression of genes introduced in plant cells*. Cependant, puisque ce symbole est spécifique aux plantes, une recherche complémentaire est nécessaire pour rechercher les documents brevets sur des MAR utilisés chez les animaux. Le symbole correspondant n'existant pas actuellement pour l'exploitation des MAR chez les animaux, une requête par mots clés est nécessaire pour trouver les documents brevets d'intérêt. La combinaison de symboles de classification et de mots clés est alors un moyen pour obtenir tous les documents brevets l'exploitation des MAR chez les plantes et les animaux.

Avec plus de 135 000 sous-divisions, la classification ECLA est un outil efficace et régulièrement exploité par les experts de l'information brevets car elle facilite les recherches sur l'état de l'art. D'un point de vue général, l'utilisation de la classification représente une stratégie de recherche particulièrement efficace pour couvrir tous les documents brevets d'un domaine technologique déterminé. Les symboles de classification sont les critères de recherche qui caractérisent le mieux les contenus techniques couverts par les documents brevets. Les classifications présentent l'avantage d'être indépendantes de la langue dans laquelle sont rédigées les demandes de brevet et de la terminologie utilisée au sein des documents brevets. Les symboles de classifications sont attribués par des examinateurs de brevets de l'OEB et non pas par les inventeurs. Par conséquent, ils sont donc moins subjectifs que les mots clés. Leur utilisation en tant que critère de recherche constitue une deuxième possibilité pour trouver des documents brevets pertinents.

2.2.6. Association de mots clés et de symboles de classification (étape 12)

Les recherches de documents brevets peuvent être optimisées en combinant plusieurs critères de recherche. En effet, les contenus techniques des documents brevets sont décrits par le résumé, les revendications et par les symboles de classification. Une stratégie de recherche efficace combine donc critères de recherche par mots clés et par symboles de classification (Figure 15).

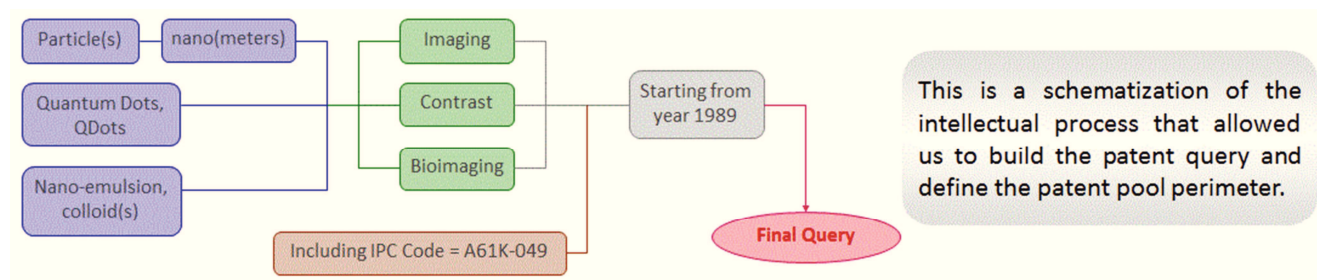


Figure 15 : Représentation schématique de la constitution d'une requête concernant une étude sur les nanoparticules utilisées en bio-imagerie (poster Patent landscape analysis "Nanoparticles for bio-imaging", Annexe 3). La requête combine des mots clés sur l'objet en lui-même (nanoparticules et synonymes), son utilisation (bio-imagerie et synonymes), un code CIB (A61K-049) et une date de début d'étude (1989).

Une telle stratégie est un moyen d'identifier des documents brevets qui n'auraient pas été trouvés uniquement avec des mots clés ou des symboles de classification. En effet, certains documents ne comportent pas de titres ou de résumés disponibles car ils n'ont pas été traduits en anglais. Ceux-ci peuvent donc être retrouvés avec une recherche de symboles de classification ECLA ou CIB. À l'inverse, d'autres documents brevets n'ont pas été hiérarchisés dans la classification ECLA et peuvent alors être récupérés dans les résultats avec une recherche par mots clés.

2.2.7. Utilisation d'autres critères de recherche

Les mots clés et les symboles de classification ne sont pas les seuls critères de recherche pour la formulation de requêtes. En effet, il existe aussi les recherches par numéros de publication, de demande ou de priorité, date de publication, noms d'inventeur et de déposant. Les auteurs de publications scientifiques peuvent également être recherchés en tant qu'inventeurs dans les demandes de brevets. La comparaison des résultats issus des recherches par nom d'inventeur et par mots clés peut donner une idée de la qualité de la requête. Soit la requête est confirmée, soit de nouveaux mots clés et/ou symboles de classification sont révélés. Dans l'exemple évoqué ci-dessus sur les *Locus Control Regions* (LCR), la recherche de « Grosveld » en tant que nom d'inventeur donne bien entendu tous les documents brevets dont l'inventeur s'appelle « Grosveld », y compris ceux couvrant les LCR. Il convient d'éliminer tous les résultats sans rapport avec les LCR, que ce soit pour l'inventeur « Grosveld » correspondant ou un inventeur homonyme. Dans ce cas, les documents brevets résultant de la recherche n'ont plus en commun des mots clés mais un nom d'inventeur. Cela permet alors d'obtenir tous les résultats dont Grosveld est l'inventeur, quels que soient les concepts technologiques, mots clés et synonymes décrits dans les documents brevets : *Locus Control Regions* (WO9508635), *Dominant Activator Sequences* (WO8901517), *Dominant Control Regions* [Grosveld et al., 1987]... La recherche par nom d'inventeur est aussi un moyen de vérifier que tous les brevets d'un auteur clé du domaine ressortent avec la requête et que tous les mots clés utilisés ont été répertoriés.

En conclusion, la définition d'une requête requiert d'intégrer tous les critères de recherche (mots clés, classification, noms d'inventeurs...) pour formuler la stratégie de recherche la plus complète et efficace. Une fois que les critères de recherche ont été définis, la requête doit être construite à l'aide d'opérateurs et de jokers. La formulation de la stratégie de recherche dépend de la base de données de brevets utilisée, c'est-à-dire des critères de recherche disponibles et surtout de ses limitations. Selon l'étude de brevets envisagée, le choix de la base de données est donc particulièrement important.

2.3. Choix de la base de données de brevets

L'une des premières étapes dans la recherche d'informations porte sur le choix de la source d'information : bases de données, sites internet, articles scientifiques... Dans le cas de la recherche d'information brevets, de nombreux outils profitent du développement d'internet. Certains outils gratuits sont accessibles à tous les utilisateurs qui peuvent ainsi exploiter l'information brevets. Par exemple, en France, l'INPI offre un libre accès au service de recherche FR-esp@cenet comprenant les demandes de brevets français, européens et internationaux des 30 dernières années [w24]. La base de données de l'*United States Patent and Trademark Office* (USPTO) recense quant à elle les demandes de brevets et les brevets accordés aux États-Unis [w25]. Cependant, la réalisation de paysages brevets et la recherche de l'état de l'art nécessitent d'utiliser des bases de données avec une couverture temporelle et géographique la plus large possible.

Le choix d'une base de données de brevets est déterminé par plusieurs critères : les couvertures temporelle et géographique, les limitations des champs de recherche (recherche en texte intégral, nombre d'opérateurs booléens utilisables, opérateurs de recherche complexes...), sans oublier la langue d'interrogation de la base. Bien entendu, le budget alloué à la recherche d'information brevets et à son analyse conditionnera également le choix d'une base gratuite ou commerciale. Au GICC, nous avons utilisé successivement les bases gratuites esp@cenet et PATENTSCOPE puis l'outil commercial QPAT.

2.3.1. La base de données esp@cenet (étapes 5b, 6a-c, 9 et 12)

La base esp@cenet est le réseau européen de bases de données de brevets [w26]. Elle recense plus de 60 millions de demandes de brevets et de brevets publiés dans plus de 80 pays [w27]. Accessible gratuitement sur internet, esp@cenet constitue l'une des principales et des plus vastes sources d'information brevets au monde. Les utilisateurs disposent de la même source d'information brevets que les experts de l'OEB. Ce réseau de bases de données brevets a été développé par l'OEB en collaboration avec les offices nationaux des brevets et la Commission Européenne. Lancé en 1998, il joue un rôle prépondérant dans la politique européenne d'information brevets de l'OEB et reflète la volonté de favoriser son utilisation et sa diffusion. Ses principaux objectifs sont les suivants :

- « offrir un accès à l'information brevets de base aux individus, PME, étudiants...,
- accroître la sensibilisation et l'utilisation de l'information brevets aux niveaux nationaux et européens,
- supporter l'innovation et réduire le gaspillage dans le cycle de l'innovation,
- compléter les ressources existantes pour la dissémination de l'information brevets ».

Pour résumer, esp@cenet a « révolutionné l'accès au public de l'information brevets internationale, en libérant les documents brevets de leur prison de papier et en changeant profondément leur dissémination, leur organisation, leur recherche et leur récupération » [w28].

La structure de ce réseau européen peut être schématisée comme un réseau de passerelles et de portails entre des bases de données de brevets connectées ensemble (Figure 16).

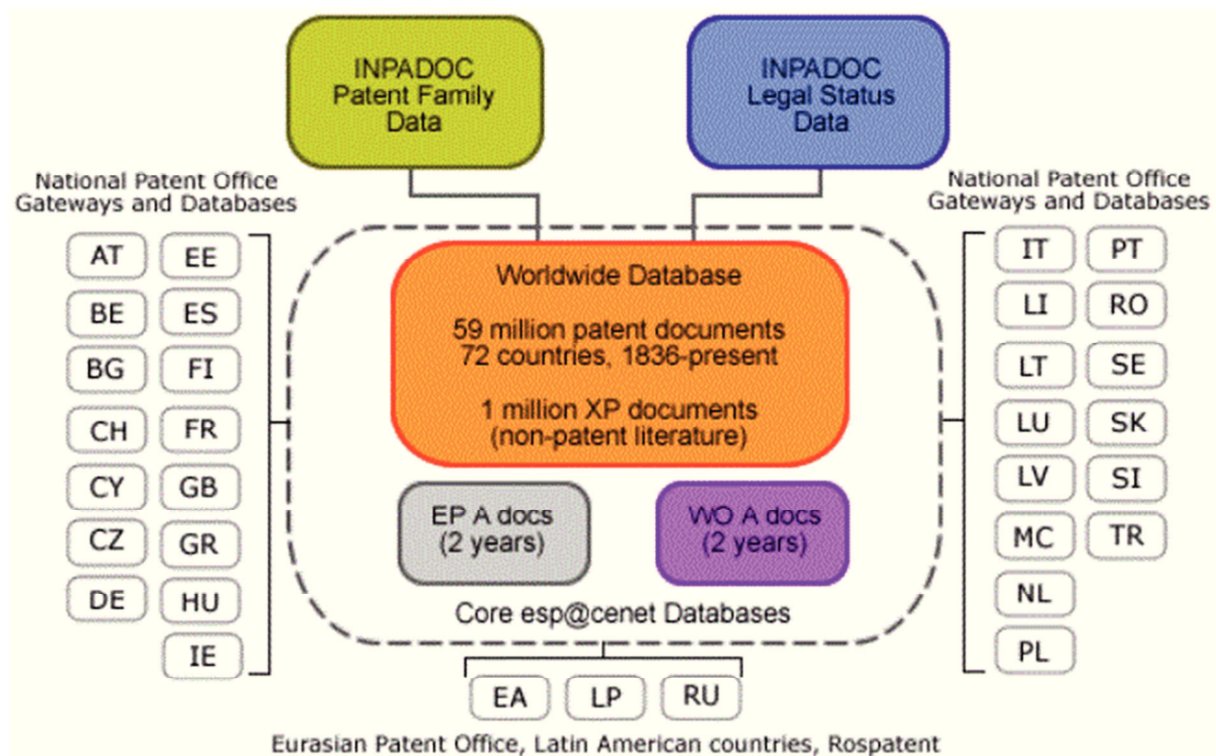


Figure 16 : Représentation schématique de la structure d'esp@cenet, le réseau européen de bases de données de brevets [w28]. Le cœur d'esp@cenet est constitué par les bases EP-esp@cenet, WIPO-esp@cenet (ou WO-esp@cenet) et Worldwide, la base qui comprend le plus de documents brevets provenant des bases d'offices nationaux (les pays sont codés par deux lettres : FR= France...).

Le cœur d'esp@cenet repose sur trois bases de données qui possèdent chacune une couverture différente de documents brevets :

- EP-esp@cenet qui permet de faire des recherches, y compris en texte intégral, dans toutes les demandes de brevets européens publiées par l'OEB (demandes EP) ;
- WIPO-esp@cenet qui contient toutes les demandes de brevets internationales (PCT) publiées par l'OMPI, avec la possibilité de faire des recherches dans le texte intégral de ces demandes WO ;
- *Worldwide* qui offre la collection de documents la plus complète avec des possibilités de recherches dans les demandes de brevets publiées dans plus de 90 pays et régions. Des informations supplémentaires peuvent aussi être présentées telles que la classe ECLA ou les références des documents brevets, lesquelles sont cités dans les rapports de recherche par les examinateurs, les déposants, les opposants... Il est important de noter que la couverture des

données disponibles varie selon les pays et selon le type d'information. Pour information, en mars 2007, la base *Worldwide* contenait des données sur plus de 60 millions de documents brevets provenant de 81 pays, dont 30,5 millions avec un titre, 19,5 millions avec un résumé en anglais et 29,5 millions avec un symbole de classification ECLA.

En outre, le réseau européen donne accès à plus de 30 bases nationales interrogeables dans la langue du pays désigné, mais aussi régionales² (OMPI, Organisation Africaine de la Propriété Intellectuelle, pays d'Amérique Latine...). Deux autres bases de données sont connectées à la base *Worldwide* d'esp@cenet. La base de données « Famille de brevets INPADOC » regroupe tous les documents brevets possédant la même priorité ou combinaison de priorités. La base de données « Statuts légaux INPADOC » fournit des informations sur la validité des brevets et les pays dans lesquels une protection a été demandée. Par conséquent, il est possible de connaître toutes les étapes qui jalonnent la vie des demandes de brevets (abandon, refus, rapport de recherche, délivrance, opposition, modification) et de vérifier si un titre de PI est en vigueur dans un pays déterminé.

De nombreux documents peuvent compléter les documents brevets présents dans les bases de données. La base mondiale *Worldwide* comprend ainsi environ un million de documents de « littérature non-brevet » qui correspondent à diverses sources d'informations : articles scientifiques, manuels... [w29]. Ces documents renforcent le contenu technique des documents brevets et détaillent ainsi encore plus l'état de l'art. Ils apparaissent avec le symbole « XP » avant leur référence bibliographique uniquement s'ils ont été classés par le système ECLA ou cités dans un rapport de recherche rédigé par un expert de l'OEB.

L'un des principaux avantages de ce réseau est qu'il repose sur une interface accessible et simple à utiliser qui devrait satisfaire tous types d'utilisateurs. D'un côté, les non-spécialistes retrouvent une interface proche de celle des bases de données bibliographiques comme Pubmed [w11], dont l'interrogation se fait par recherche de mots

² D'après le PCT, un brevet régional est délivré par une administration nationale ou intergouvernementale habilitée à délivrer des brevets ayant effet dans plus d'un état.

clés dans les titres et résumés des articles scientifiques. D'un autre côté, les utilisateurs expérimentés peuvent profiter de ses puissantes fonctionnalités telles que la classification ECLA ou la recherche en texte intégral dans les bases EP-esp@cenet et WIPO-esp@cenet. À ce sujet, la base esp@cenet propose quatre types de recherche différents : la recherche rapide, la recherche avancée, la recherche par numéros (numéros de publication, de demande ou de priorité) et la recherche dans la classification européenne (ECLA). Les recherches rapide, avancée et par numéro nécessitent au préalable de choisir l'une des trois bases de données brevets proposées : EP-esp@cenet, WIPO-esp@cenet et *Worldwide*, cette dernière étant majoritairement utilisée.

La recherche avancée propose les champs de recherches suivants : par mots clés dans le titre ou le résumé, numéros de publication, numéros de demande, numéros de priorité, dates de publication, noms d'inventeurs, noms de déposants (Figure 17).

Entrez vos mots-clés en anglais

Mots-clés dans le titre:	plastic and bicycle
Mots-clés dans le titre ou l'abrégé:	hair
Numéro de publication:	WO2008014520
Numéro de demande	DE19971031696
Numéro de priorité:	WO1995US15925
Date de publication:	yyyymmdd
Demandeur(s)	Institut Pasteur
Inventeur(s)	Smith
Classification européenne (ECLA):	F03G7/10
Classification Internationale des Brevets (CIB)	H03M1/12

Figure 17 : Copie d'écran de la recherche avancée d'esp@cenet [w30]. Après avoir choisi l'une des trois bases de données brevets parmi EP-esp@cenet, WIPO-esp@cenet et *Worldwide*, différents critères de recherche peuvent être utilisés pour trouver des documents brevets.

La recherche par symboles de classification ECLA ou CIB est également possible, ce qui n'est pas le cas pour les symboles de classification propres aux offices de brevets américain (USPTO) et japonais (*Japan Patent Office* ou JPO). Cependant, la recherche par symboles ECLA est disponible uniquement dans la base de données mondiale *Worldwide*.

Plus la requête de recherche est précise, meilleurs et pertinents seront les résultats. Par conséquent, les termes et mots clés peuvent être combinés entre eux avec des opérateurs booléens :

- l'opérateur « **AND** » retrouve les documents brevets comprenant tous les termes. C'est également l'opérateur par défaut des champs de recherche des titres, résumés, inventeurs, demandeurs (déposants) et des symboles de classification CIB et ECLA ;
- l'opérateur « **OR** » retrouve les documents brevets comprenant au moins un des termes. Cet opérateur est particulièrement utile pour les synonymes et les mots clés techniques proches : (*gene AND (transfer OR delivery)*)... Pour les champs de recherche concernant les numéros de publication, de demande, de priorité et dates de publication, c'est l'opérateur par défaut ;
- l'opérateur « **NOT** » retrouve les documents brevets comprenant le premier terme sauf le second. Cet opérateur est un moyen d'exclure des résultats sans intérêt pour l'étude.

Des jokers de troncature peuvent également être utilisés pour remplacer un ou plusieurs caractères :

- le joker « ***** » remplace une suite de caractère de longueur variable. Ce joker est utile pour étendre la recherche et tenir compte des formes plurielles et composées des mots clés. Par exemple, la requête « *transpos** » donne comme résultats les documents brevets citant « *transposon(s), transposase(s), transposable(s), transposition(s)*... », et tient même compte de certaines erreurs de frappe comme « *transposan(s)* » (WO9909817, Cf. précédemment).
- le joker « **#** » remplace exactement un caractère ;

- le joker « ? » remplace zéro ou un caractère. Il peut donc être utilisé pour obtenir les formes singulières et plurielles de mots. La recherche « transposon? » donnera les résultats avec « transposon » ou « transposons » ;

Une expression de mots clés peut être trouvée avec l'utilisation des guillemets mais ils doivent être manipulés avec précaution car ils permettent de retrouver l'expression exacte. En manipulant judicieusement les mots clés, les opérateurs et jokers de troncature, et en choisissant les champs de recherche adéquats (recherches dans les titres et résumés, classification...), un maximum de documents brevets pertinents pour l'étude de brevets devrait ressortir dans les résultats. Cependant, l'utilisation des opérateurs booléens et des jokers de troncature est limitée dans esp@cenet et notamment dans la base *Worldwide* qui nécessite une interrogation avec des mots clés en anglais [w31]. Par exemple, ces limitations sont : un maximum de quatre termes de recherche par champ de saisie sur la base mondiale *Worldwide*, les caractères spéciaux ne sont pas reconnus par le système, les jokers fonctionnent seulement dans les champs « Titre », « Titre ou résumé », « Inventeur » et « Demandeur », il n'est pas possible d'effectuer une recherche entre deux dates... Il est important de noter que dans le cas de l'utilisation de la base FR-esp@cenet, il existe quelques différences (interrogation avec des mots clés en français, pas de maximum de termes ni d'opérateurs par champ de recherche).

En tenant compte de ces limitations, il est possible de combiner des termes avec des opérateurs booléens et jokers de troncatures dans les divers champs de recherche. Définies sous l'appellation « recherches emboîtées », elles utilisent des parenthèses pour préciser l'ordre dans lequel les termes de recherche et les opérateurs doivent être interprétés. Ainsi, des termes sont recherchés dans d'autres termes de recherche. Dans ce cas, la règle est la priorité des parenthèses les plus centrales qui prévalent sur les parenthèses externes. La requête « *transpos* AND (gene (delivery OR transfer))* », exécutée dans la base *Worldwide* d'esp@cenet, donne tous les documents brevets pertinents qui associent des mots clés concernant le transfert de gènes aux transposons et transposases.

Par ailleurs, d'autres limitations sont également à prendre en compte : un maximum de 500 documents brevets trouvés est affiché, ce qui limite les résultats notamment pour l'élaboration de paysages brevets recensant un nombre important de documents brevets.

De plus, le nombre limité de combinaisons offert par esp@cenet et l'impossibilité d'effectuer une recherche en texte intégral dans la base *Worldwide* peuvent être problématiques pour la réalisation de paysages brevets, et plus généralement pour toute recherche de brevets complexe. Enfin, esp@cenet ne propose pas d'historique des recherches ni d'outil d'analyse statistique, à l'inverse de la base de données PATENTSCOPE.

2.3.2. La base de données PATENTSCOPE (étapes 6a-c, 9 et 12)

À l'instar du réseau européen esp@cenet, une autre base de données gratuite peut être utilisée : PATENTSCOPE. Elle offre un accès à plus de 1 700 000 demandes internationales de brevets publiées depuis 1978 [w32]. Les champs de recherche proposés sont un peu plus diversifiés que pour esp@cenet : numéro de publication ; numéro de demande ; date de publication ; date de la demande ; titres en anglais/français/japonais ; résumés en anglais/français/japonais ; nom, adresse, domicile et nationalité du déposant ; nom et adresse du mandataire ; symbole de classification CIB ; nom et adresse de l'inventeur ; pays, date d'entrée, numéro de la demande et type d'entrée en phase nationale ; numéro, date et pays de priorité ; états désignés ; langue dans laquelle la demande a été déposée ; langue de publication ; revendications et description.

La recherche peut aussi être structurée avec des jokers de troncature et les opérateurs booléens « AND », « OR » et « ANDNOT » (NOT dans esp@cenet). De plus, deux nouveaux opérateurs font leur apparition : « XOR » et « NEAR ». Séparant deux termes, l'opérateur « XOR » récupère les documents brevets qui ont soit un terme soit un autre, mais pas les deux. Quant à l'opérateur de proximité « NEAR », il permet de retrouver des documents qui contiennent des termes proches les uns des autres, avec une distance maximale de cinq mots pour PATENTSCOPE.

L'une des particularités pour cette base de données repose sur la visualisation des résultats de la recherche sous la forme de graphiques simples (Figure 18).

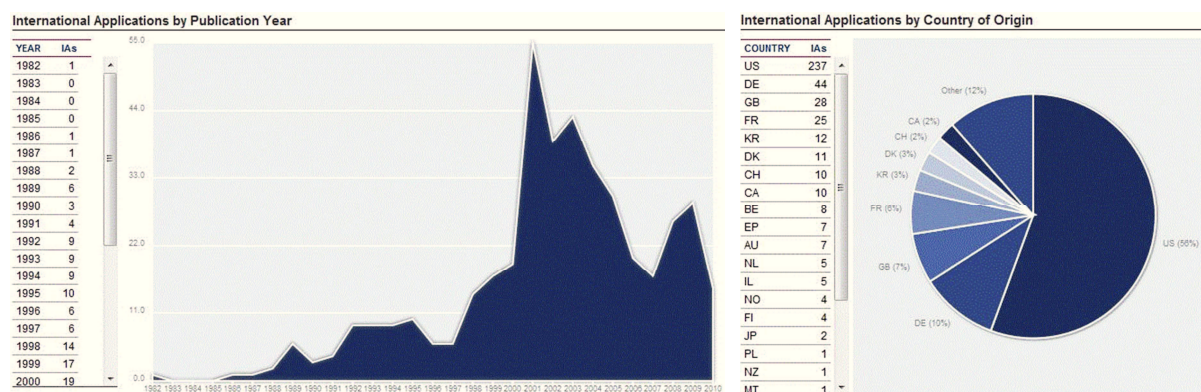


Figure 18 : Exemples de graphiques proposés par l'outil d'analyse statistique de PATENTSCOPE. Ceux-ci exposent l'évolution des dépôts par année de publication et les pays d'origine des demandes PCT pour les 426 demandes internationales de brevets obtenues grâce à la recherche dans les revendications du terme « transposon? ».

Ces représentations graphiques statistiques portent sur l'évolution des dépôts par année de publication, les pays d'origine des demandes PCT (soit le pays du premier déposant nommé), le nombre de demandes PCT par noms de déposants et le nombre de demandes internationales par sous-classes CIB (par exemple C12N). Une autre fonctionnalité de PATENTSCOPE appréciée par ses utilisateurs se base sur un historique interactif des recherches qui recense les vingt dernières recherches comprenant le détail des requêtes réalisées. Ainsi, les requêtes peuvent être sauvegardées puis comparées, retravaillées et ré-exécutées ultérieurement afin de voir si de nouveaux documents brevets pertinents ont été publiés.

Depuis la fin de l'année 2009, PATENTSCOPE a été élargie et renforcée de plus de trois millions de documents brevets provenant de collections d'informations nationales et régionales [w33]. Dorénavant, il est possible d'effectuer des recherches dans les collections nationales de brevets de neuf offices nationaux et régionaux (à la date du 21 juillet 2010) : Office de l'Organisation Régionale Africaine de la Propriété Intellectuelle et offices de l'Afrique du Sud, de Cuba, d'Israël, du Mexique, de la République de Corée, de Singapour, du Viêtnam et de l'Argentine. Des négociations sont actuellement en cours avec d'autres offices nationaux, et leur collection nationale de documents brevets devrait être à leur tour recueillie par l'OMPI, en échange d'une assistance technique. Comme pour esp@cenet, l'objectif d'un tel outil reste la promotion et la diffusion la plus vaste possible de l'information brevets.

Depuis le mois de mai 2010, PATENTSCOPE met à disposition un outil expérimental de recherches multilingues d'information brevets, appelé *Cross-Lingual Information Retrieval* ou CLIR [w34]. Il offre de nouvelles possibilités en matière de recherches multilingues en générant des mots clés dans d'autres langues, avec des niveaux de précision et d'exhaustivité qui sont à déterminer. En effet, à partir d'un ou plusieurs mots clés écrits en allemand, anglais, espagnol, français ou japonais, CLIR recherche leurs traductions les plus exactes possibles dans les autres de ces cinq langues et propose même des synonymes et des mots clés proches (Figure 19). Ainsi, non seulement il traduit le mot clé initial, mais en plus il apporte des synonymes et de nouveaux mots clés proches dans d'autres langues.

Results 1-10 of 72 for Criteria:((EN_Ti:("transposase") OR EN_AB:("transposase")) OR (DE_Ti:("Transposase") OR DE_AB:("Transposase")) OR (FR_Ti:("transposase") OR FR_AB:("transposase")) OR (JA_Ti:("のトランスボザーゼ") OR JA_AB:("のトランスボザーゼ")))) Office(s):all Language:EN Stemming: true

Figure 19 : Exemple de traduction du terme anglais « transposase » en allemand, français et japonais.

L'outil CLIR se base sur l'analyse statistique de documents brevets et ressources terminologiques sur le sujet et utilise des méthodes de corrélation entre les langues des collections de l'OMPI. De plus, les symboles de classification CIB jouent un rôle de filtre, permettant ainsi d'avoir des domaines technologiques en rapport avec les mots clés et d'avoir des résultats conformes aux attentes des utilisateurs. Cet outil est donc particulièrement efficace pour accroître la qualité et la quantité de résultats pertinents issus des collections de PATENTSCOPE, à partir du moment où il existe des entrées dans ces différentes langues. Par ailleurs, il est important de noter qu'il est toujours possible de récupérer la requête finale proposée par l'outil CLIR et de l'utiliser dans une autre base de données.

Depuis août 2010, le nouveau système PATENTSCOPE [w33] propose de parcourir les descriptions et revendications des documents brevets dans toutes les langues supportées par Google Translate. Il est ainsi possible d'obtenir une traduction automatique d'un document anglais en caractères chinois. Avec cette fonctionnalité combinée à l'outil CLIR, PATENTSCOPE fournit ainsi une solution complète pour la recherche et l'utilisation de l'information brevets dans différentes langues.

Les exemples d'esp@cenet, qui offre la collection de documents la plus complète avec la base *Worldwide*, et de PATENTSCOPE, qui propose des outils innovants, montrent que les bases de données de brevets gratuites sont en constante amélioration. Celles-ci cherchent à compléter leurs collections avec des documents brevets provenant d'offices nationaux et proposent des fonctionnalités toujours plus utiles et efficaces pour les utilisateurs de l'information brevets, qu'ils soient experts ou non.

Cependant, certaines de leurs limitations peuvent imposer d'utiliser des bases de données commerciales. En effet, les moteurs de recherche internes des bases de données de brevets gratuites sont souvent moins puissants que ceux de leurs homologues commerciaux. Ils ne proposent pas toujours la recherche en texte intégral (titre, résumé, revendications et description) et n'offrent qu'un nombre limité de combinaisons de mots clés, d'opérateurs et de jokers de troncature. De plus, l'extraction de documents brevets résultant de la recherche est souvent lourde et ne permet pas la récupération de gros volumes de données. Dans le cas des paysages brevets qui portent sur des centaines voire des milliers de documents, il est préférable de consulter des bases de données commerciales qui facilitent le téléchargement de grands volumes de données.

De même, les utilisateurs des bases de données gratuites peuvent se retrouver limités en raison des fonctionnalités réduites de ces bases par rapport aux bases commerciales. Par exemple, un outil comme Scifinder propose en plus de la recherche par mots clés, la recherche d'informations scientifiques à partir d'une structure chimique ou d'une séquence biologique [w35]. En outre, il est possible que les résumés ne soient pas toujours disponibles dans les bases de données gratuites comme esp@cenet, alors qu'ils figurent dans les bases de données commerciales. L'OEB conseille de tenir compte des limitations des systèmes gratuits et reconnaît même que l'utilisation des outils commerciaux prend moins de temps pour trouver des résultats pertinents [w36]. Il est ainsi précisé à ce sujet que « les coûts engendrés lorsqu'un professionnel en matière de brevets passe deux heures sur un système gratuit sont souvent plus élevés que les coûts totaux - la somme des coûts du temps de travail et des coûts de la base de données - occasionnés lorsque le même professionnel passe quinze minutes sur un système nécessitant le paiement d'une redevance ». Cependant, ces limitations sont un choix délibéré des offices nationaux qui ne

souhaitent pas être en concurrence déloyale avec les fournisseurs de bases de données commerciales [Lambert, 2004].

2.3.3. La base de données FamPat de Questel et l'outil de recherche QPAT (étapes 6a-c, 9 et 12)

Pour la réalisation d'études de brevets et de paysages brevets, il est préférable d'utiliser des bases de données de brevets commerciales. Bien qu'elles s'appuient souvent sur des bases de données provenant de l'OEB et d'autres offices des brevets, elles permettent non seulement de gagner du temps mais aussi de produire des résultats plus précis et pertinents. Leurs limitations sont moins importantes au niveau des critères de recherche et des manipulations de mots clés, ce qui rend possible la construction de requêtes complexes. De nombreuses sociétés fournissent des outils commerciaux pour la recherche de documents brevets : MicroPatent, Delphion, STN... [w37]. Au GICC, nous avons fait le choix d'utiliser l'outil de recherche QPAT, développé par la société Questel.Orbit [w38]. En effet, je l'avais déjà utilisée pendant mon stage de Master Professionnel de Biotechnologies & Droit de l'Université de Tours, grâce au soutien de la société FIST. Dans un premier temps, QPAT peut être utilisé pour rechercher des documents brevets, des modèles et des dessins, dont les résultats sont facilement identifiables et exploitables. Dans un second temps, un module d'analyse statistique facilite l'analyse des résultats à l'aide de graphiques dynamiques et interactifs. Il propose aussi la surveillance de marché (concurrents, statuts légaux...), la commande de documents originaux et même l'envoi de rapports pour partager les analyses.

Au laboratoire, j'ai surtout exploité QPAT pour ses puissantes fonctionnalités d'outil de recherche de documents brevets [w13] au sein de la base de données brevets FamPat. Cette base de données de brevets est présentée comme la plus exhaustive au monde [w39]. Elle couvre en effet les collections de 93 pays depuis le début du XIX^{ème} siècle (FR : 1842, US : 1836, GB : 1840, CH : 1888...) et recense plus de soixante millions de documents, quarante millions de résumés en anglais et dix-huit millions d'images. Elle est enrichie de textes intégraux de plus de vingt offices dont la France, l'Allemagne, le Canada, l'Europe, les États-Unis, le Japon, le Brésil, la Chine, l'Inde, la Russie, les demandes PCT... Dans FamPat, les

documents brevets sont regroupés en familles car « l'invention est au cœur de la famille ». La société Questel a développé sa propre définition d'une famille de brevets. Elle s'est basée sur la définition de la famille stricte de l'OEB qui comprend tous les documents ayant exactement la même priorité ou combinaison de priorités (pour esp@cenet) et a intégré des règles additionnelles :

- les dépôts qui ne sont pas compris dans la limite des douze mois,
- les liens entre les enregistrements EP et PCT,
- les demandes US provisoires qui partagent les mêmes priorités avec les dépôts US publiés.

Ce regroupement en familles FamPat présente notamment comme avantage une meilleure lisibilité de lien entre les documents brevets qui sont membres de la famille FamPat. Questel a également défini la notion d'« information clé » des documents brevets. Cette information clé est extraite de l'objet de l'invention, des avantages de l'invention comparés aux inconvénients de l'art antérieur et des revendications indépendantes. Elle propose donc uniquement les informations essentielles des documents brevets, d'où une recherche plus précise et une exploitation plus rapide.

QPAT permet d'effectuer facilement des recherches de brevets à l'aide des critères suivants (Figure 20) :

- mots clés dans les titres, résumés, revendications, information clé, description et texte intégral,
- symboles des classifications CIB, ECLA, US, et ICO (classification *In Computer Only* dérivée de la classification ECLA utilisée par les examinateurs de l'OEB pour décrire des caractéristiques mineures de l'invention ou des caractéristiques pour lesquelles il n'existe pas de code ECLA),
- noms des déposants,
- noms des inventeurs,
- numéros de publication, de dépôt et de priorité.

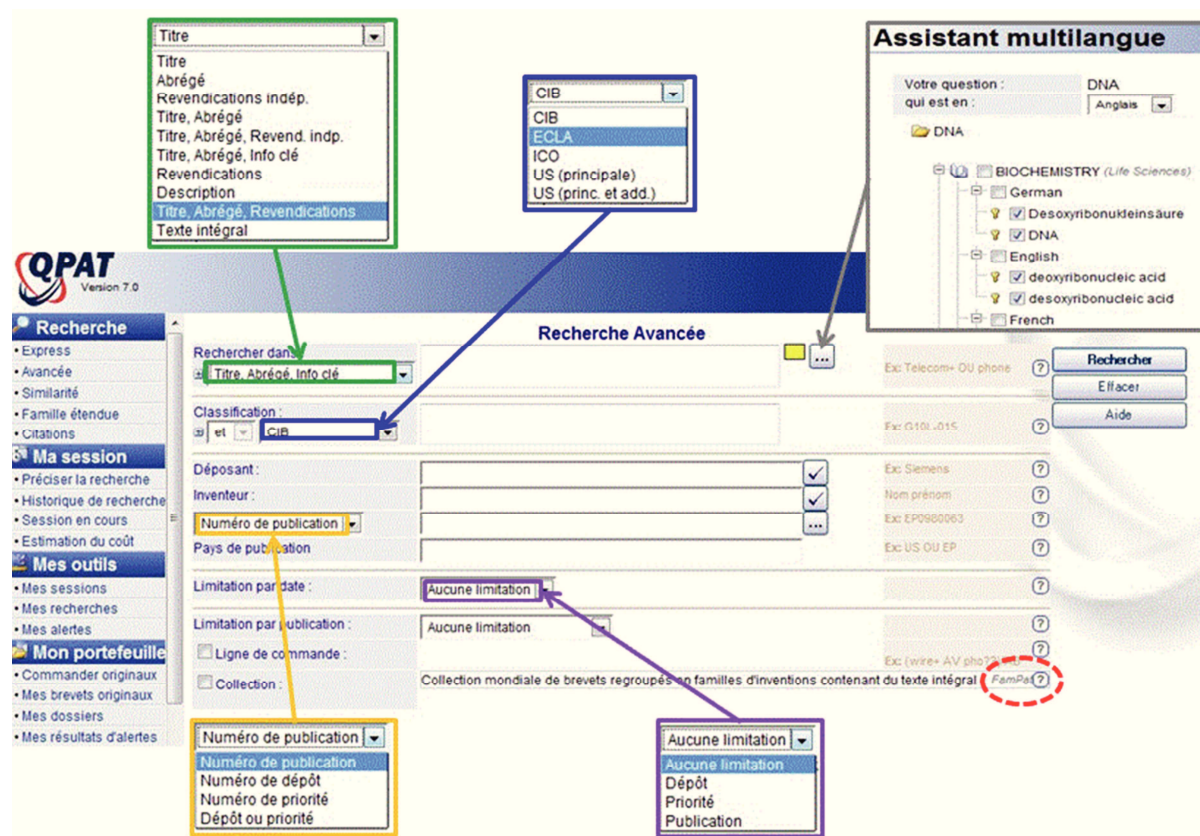


Figure 20 : Capture d'écran de l'interface de recherche avancée de QPAT, comprenant le détail des différents choix proposés pour certains critères de recherche. La collection mondiale de brevets sélectionnée est la base de données de brevets FamPat (cercle discontinu rouge).

Le choix des critères de recherche disponibles dans QPAT est donc plus large que pour les bases de données gratuites présentées précédemment. La recherche dans les revendications est particulièrement intéressante et importante pour les études de brevets comme les paysages brevets. En outre, QPAT accepte de limiter la recherche par des dates de priorité, de dépôt ou de publication. Aussi, les utilisateurs experts peuvent profiter du large choix de jokers de troncature et surtout d'opérateurs disponibles [Questel, 2010]. Les jokers de troncature sont les mêmes que pour esp@cenet, à l'exception que le joker « + » remplace le joker « * ». Concernant les opérateurs, ils sont plus nombreux que pour les bases de données gratuites esp@cenet et PATENTSCOPE et ils offrent la possibilité d'élaborer des requêtes complexes (Tableau 1).

OU	Retrouve les références contenant au moins un des termes	sulfur ou sulphur
ET	Tous les termes	plutonium et isotope
SAUF	Le premier terme sans le second terme	suv sauf vesicle
CHP	Les termes dans le même champ	sodium chp chlorine
PHR	Les termes dans la même phrase	sodium phr chlorine
PRG	Les termes dans le même paragraphe	sodium prg chlorine
M	Les termes juxtaposés quel que soit l'ordre	redundancy m check+
nM	Les termes adjacents, quel que soit l'ordre, et pouvant être séparés par un maximum de n mots (valeur de n comprise entre 1 et 9)	electric+ 2m conduct+ 2m adhesive
AV	Les termes juxtaposés dans l'ordre précisé ; traitement appliqué par défaut pour deux termes entrés sans opérateur	smart av card? smart card?
nAV	Les termes adjacents dans l'ordre précisé et pouvant être séparés par un maximum de n mots (valeur de n comprise entre 1 et 9)	friction 1av pad?
SDOC	Les termes dans le même membre de la famille FamPat Opérateur spécifique à la collection FamPat	us/pub sdoc lic/eg
Parenthèses	Parenthèses (mise en facteur) nécessaires lors de combinaison d'opérateurs différents	((wireless av application av protocol) ou wap) sauf (dna ou transgenic) (hair 1m (dye ou dyeing)) et oxidat+

Tableau 1 : Liste des opérateurs à disposition des utilisateurs de QPAT pour la construction de requêtes complexes [Questel, 2010].

L'un des avantages majeurs repose sur le fait que tous les documents sont en anglais d'origine ou traduits instantanément par machine. Par conséquent, il est possible d'effectuer des recherches en anglais dans les nombreux documents brevets des pays asiatiques (Chine, Inde, Japon..). Comme pour PATENTSCOPE, un outil multilingue propose des traductions et des synonymes pour chacun des termes recherchés, dans les différents domaines concernés par l'expression. De plus, les formes féminines et plurielles des mots sont automatiquement prises en compte. Cependant, les troncatures sont ignorées et les opérateurs ne sont pas autorisés. Par exemple, le mot anglais « DNA » peut être associé dans le domaine de la biochimie aux mots français « acide désoxyribonucléique » et « ADN », au mot allemand « desoxyribonukleinsäure » et aux mots anglais « deoxyribonucleic acid » et « desoxyribonucleic acid ». Un autre avantage de QPAT porte sur la gestion de la stratégie de recherche. En effet, l'historique des stratégies de recherches peut être sauvegardé et les requêtes ainsi conservées peuvent être aisément relancées ultérieurement. Une alerte peut également être programmée pour informer de toute nouvelle publication de brevet entrant

dans les résultats des recherches sauvegardées. De plus, il est possible de combiner les recherches effectuées pour créer une nouvelle recherche.

Par exemple, la recherche de documents brevets sur les *Locus Control Regions* porte sur une requête complexe :

Requête 4 : requête finale soumise dans QPAT = 395 résultats

(Requête 3 OU Requête 2) SAUF Requête 1

Requête 3 = 874 résultats

((((LOCUS 1V CONTROL+ 1AV (REGION+ OU SUBREGION+)) OU (DOMINANT 1AV ACTIVAT+ 1AV SEQUENCE+) OU (DOMINANT 1AV CONTROL+ 1AV (REGION+ OU SUBREGION+)) OU (LOCUS 1AV ACTIVAT+ 1AV (REGION+ OU SUBREGION+)) OU ((LCR OU LCRS OU DCR OU DCRS OU LAR OU LARS) CHP (GENE OU GENES OU GENETIC OU DNA OU GLOBIN OU CHROMATIN))))/BI/CLMS/OBJ)

Requête 2 = 35 résultats

((((DNASE+ OU DNAASE+) PHR ((SUPERSENSIT+ OU HYPERSENSIT+ OU SUPERHYPERSENSIT+) 1M (SITE OU SITES OU CORE OU LOCUS OU LOCI))))/BI/CLMS/OBJ)

Requête 1 = 1503 résultats

((((LIGA+ 1AV CHAIN 1AV REACTION) OU (LIQUIDE+ 1AV CEPHALO 1AV RACHIDIEN+) OU ((LCR OU LCRS OU DCR OU DCRS OU LAR OU LARS) PHR (PCR OU NASBA OU SDA OU AMPLIC+ OU PTP OU HPV)) OU ((LCR OU LCRS OU DCR OU DCRS OU LAR OU LARS) 2M (REDUCTASE+ OU RESPONSE+ OU SANDOSTATIN+ OU PROTEIN+ OU POLYPEPTIDE+ OU METHOD+ OU RELATED OU RECEPTOR+ OU ENZYME+ OU ANALYSIS OU PHOSPHATASE+))))/BI/CLMS/OBJ)

Dans cet exemple, les requêtes 2 et 3 permettent de récupérer les documents brevets sur les *Locus Control Regions* ou LCR. Cependant, en biotechnologies, le terme « LCR » peut également être utilisé pour définir « *Ligase Chain Reaction* » ou « Liquide Céphalo Rachidien ». L'élimination des résultats de la requête 1 est un moyen pour n'obtenir que des résultats pertinents.

Concernant les résultats, QPAT possède de nombreuses fonctionnalités très utiles. Par exemple, les mots clés recherchés sont surlignés, chaque mot clé pouvant être mis en évidence par une couleur distincte. La fonction KWIC (*Key Word In Context*) affiche uniquement les parties dans lesquels les termes sont présents. Cela améliore la visibilité et la rapidité de sélection de résultats pertinents et facilite ainsi l'élimination des documents brevets sans intérêt. QPAT intègre également l'information légale complète des demandes de brevets et des brevets accordés, y compris leur propriété, l'entrée en phases nationales, les actions en opposition, les accords de licences... Enfin, QPAT met à disposition de ses clients un module d'analyse statistique qui élabore des représentations graphiques dynamiques en se basant sur les résultats de la recherche de documents brevets. En effet, les sociétés Questel et Intellixir ont annoncé en 2007 la signature d'un accord sur l'intégration d'un module développé par Intellixir dans le système de gestion de portefeuille de brevets de Questel. Cependant, dans le cadre de l'élaboration de paysages brevets, ce module reste limité car il ne permet pas d'opérer de traitement des résultats ni d'approfondir leurs analyses en réalisant des segmentations. En revanche, l'association de QPAT et de la version intégrale d'Intellixir fournit une méthodologie complète pour la recherche et l'analyse de documents brevets. Les résultats provenant des recherches avec QPAT sont extraits puis intégrés dans Intellixir, offrant ainsi plus de possibilités d'analyses et de traitements. Pour paraphraser Questel, cette complémentarité offre un gain de productivité incontestable grâce à « l'exhaustivité et la structuration de la donnée, la puissance de la recherche, l'analyse graphique dynamique et interactive, le stockage de documents pertinents dans des portefeuilles ». C'est le choix que nous avons fait au GICC pour la réalisation des paysages brevets.

2.4. Formatage et analyses des documents brevets extraits

2.4.1. Présentation de l'outil d'analyse Intellixir

Le système Intellixir [w14] est une application web qui permet d'analyser de l'information structurée issue de bases de données brevets et bibliographique. Proposé par la société du même nom créée en 2002, il se base sur une application informatique dédiée au traitement de l'information structurée et issue des travaux du Commissariat à l'Énergie

Atomique. Destiné aux organismes qui souhaitent maîtriser leur environnement scientifique et technique, cet outil propose des solutions de traitement et d'analyse de l'information à des fins de veilles technologique et concurrentielle. Les applications visées sont :

- l'étude d'antériorité (étude de brevetabilité),
- l'étude d'innovation,
- la recherche de collaborations,
- la recherche de l'état de l'art,
- la création d'indicateurs et de tableaux de bords (nombre de publications sur une technologie dans le temps)...

Le système Intellixir applique des traitements statistiques et établit des corrélations entre les différentes données de l'information brevets. Les résultats correspondent à des représentations graphiques sur des pages web interactives et dynamiques qui présentent de nombreuses données sous des angles d'analyse différents. Cet outil d'analyse permet donc de localiser finement des informations stratégiques : évolution des dépôts dans un domaine, détection de tendances technologiques émergentes, localisation des centres de R&D et des marchés, identification de partenaires potentiels et d'experts, compréhension des réseaux de collaboration entre inventeurs et entre déposants, détection de sociétés qui émergent des laboratoires de recherche publics...

L'intégration de la collection de documents brevets résultant de la recherche avec QPAT se fait aisément en suivant les indications du système Intellixir. Celui-ci est parfaitement adapté au traitement de l'information brevets dont les informations sont bien structurées (numéros de publication, de priorité, noms d'inventeurs et de déposants, titres et résumés, description et revendications). Un premier formatage des données est ensuite effectué automatiquement par Intellixir. En effet, l'avantage offert par ce système est qu'il regroupe, pour le traitement de tous les documents brevets intégrés, plusieurs processus en une seule application : reformatage grâce à des filtres dédiés, dédoublonnage, analyses et représentations graphiques. Néanmoins, un second formatage, manuel, doit être réalisé à la suite de ce formatage automatique.

2.4.2. Formatage manuel des données (étape 16)

Malgré le formatage opéré par le système Intellixir, un formatage manuel doit être effectué sur les noms des inventeurs et des déposants (dénommés « affiliations » dans Intellixir). Il est nécessaire de regrouper et d'homogénéiser les données bibliographiques qui peuvent différer entre les documents brevets. En effet, un seul et même inventeur peut avoir plusieurs combinaisons pour son nom : prénom, diminutif du prénom, deuxième prénom... et nom. Bien qu'il s'agisse du même inventeur, l'outil Intellixir comprendra autant de noms d'inventeurs différents que de combinaisons dans la base. Par exemple, l'inventeur Franklin Grosveld est dénommé de plusieurs façons dans les documents brevets : Grosveld Franklin, Grosveld Frank, Grosveld Franklin Gerardus, Grosveld Franklin G, et même Grosveld Frankin G et Grosveld Franklin Geradus. Cela représente six dénominations différentes pour un seul et même inventeur (dont deux avec des erreurs). De même, un unique déposant peut avoir plusieurs appellations. Par exemple, le Centre National de la Recherche Scientifique peut aussi être écrit de plusieurs manières : CNRS, C.N.R.S., CENTRE NAT RECH SCIENT, CT NAT RECH SCIENT, CT NAT DE LA RECHERCHE SCIENTI... et même par des noms erronés tel que « SANTR NAS ONAL DE LJA RESHERSH » (RU2373415). Il est possible de rencontrer des erreurs dans le nom des déposants comme Sanagamo Biosciences (JP2009106276) à la place de Sangamo Biosciences (WO0042219).

Un tel regroupement offre ainsi la possibilité d'avoir une représentation plus réaliste du nombre de brevets de l'inventeur ou du déposant en question, de son réseau de collaborations, de la fréquence de ses dépôts dans le temps... Pour faciliter cette homogénéisation dans les noms de déposants, un dictionnaire a été mis en place par FIST dans Intellixir. Des entrées références ont été intégrées dans le dictionnaire qui regroupe automatiquement les différentes appellations d'un même déposant. Par exemple, la requête suivante regroupe les multiples noms de déposants correspondant au Centre National de la Recherche Scientifique ou CNRS : « (CNRS* OR C.N.R.S OR (CENT* AND NAT* AND RECH* AND SCI*) OR (CT* AND NAT* AND (SCIEN* OR SCENTIF* OR SCIENTIF* OR RECH*)) OR (C AND N AND R AND S) OR (SANTR AND NA* AND RESHERSH)) ».

2.4.3. Analyses proposées par le système Intellixir (étapes 17 et 18) et exemples de représentations graphiques

Illustrées par quelques représentations graphiques issues du paysage brevets sur les éléments de contrôle de la chromatine, les principales analyses proposées par Intellixir sont les suivantes :

A. Analyse globale de la base :

- 1) Couverture temporelle : évolution du nombre de publications par an mesurée sur la base.

La figure 21 montre l'évolution des dépôts de brevets pour la période 1987-2008.

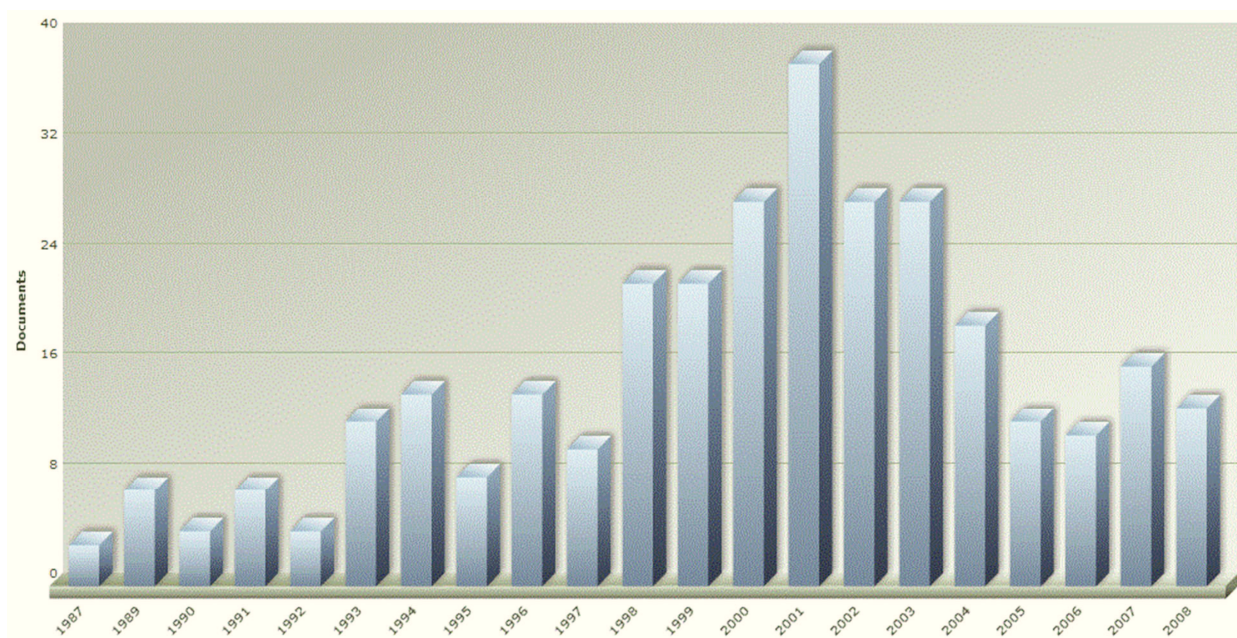


Figure 21 : Évolution des dépôts de demandes de brevets entre 1987 et 2008.

- 2) Couverture géographique : couverture de protection (tous les pays où l'invention est protégée), couverture d'émission (pays d'origine où le premier brevet a été déposé et qui est identifié à partir du numéro de priorité) et couverture d'extension (pays d'extension de la protection).

La figure 22 est une carte mondiale exposant les pays où des brevets ont été déposés. Généralement, les pays où une protection est demandée correspondent aux pays

représentant un marché d'intérêt pour les déposants ou aux pays dans lesquels des concurrents sont présents.

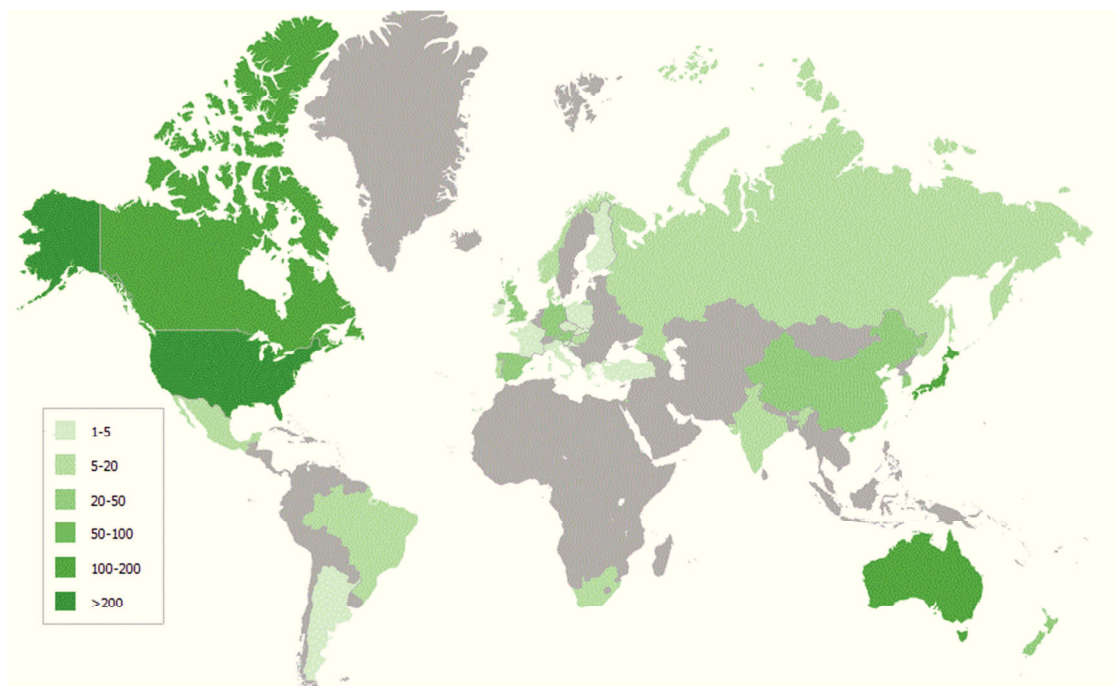


Figure 22 : Distribution géographique de la protection accordée par les brevets.

- 3) Innovation : évolution du nombre de publications par an comparée au nombre de nouveaux concepts par document par an et au nombre de nouveaux auteurs par document par an.
- 4) Recherche & Industrie : auteurs ou affiliations ayant rédigé des articles ET déposé des brevets.

B. Analyse des acteurs :

- 1) Auteurs (= inventeurs) :
 - a. Principaux, sélectionnés ou tous.

L'analyse des co-publications de brevets est un moyen pour identifier les équipes de recherche et les collaborations entre inventeurs. Par exemple, dans la figure 23, Otte Arie, Sewalt Richard, Van Blokland Henricus et Kruckeberg Arthur sont co-inventeurs pour de nombreux brevets : ils travaillent en effet pour la société Chromagenics.

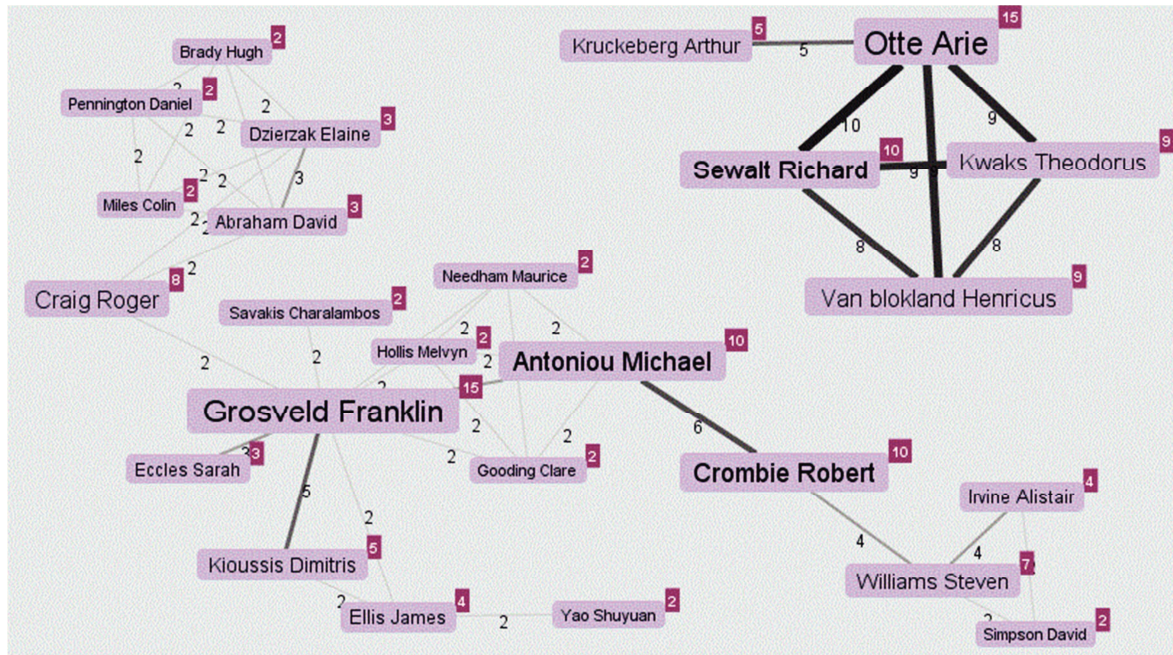


Figure 23 : Réseau de collaboration entre les inventeurs présents dans la base de données de brevets créée dans Intellixir. Les chiffres des cadres bordeaux représentent le nombre total de brevets de l'inventeur, qui est directement lié à la taille de police du nom. Les chiffres noirs entre deux inventeurs correspondent au nombre de brevets en commun.

- b. *Experts* : il s'agit des auteurs qui collaborent beaucoup, indépendamment des auteurs qui publient le plus. Le facteur d'expertise est calculé selon la formule suivante : Facteur d'expertise = Nombre de publications × Nombre de co-auteurs différents.

Les auteurs experts sont ceux dont le coefficient d'expertise est le plus important. Ce facteur d'expertise multiplie le nombre de documents de l'auteur par le nombre de co-auteurs différents de cet auteur, ce qui permet de faire « remonter » des auteurs qui collaborent beaucoup. Par ailleurs, la figure 24 ci-dessous corrobore le graphique ci-dessus sur les auteurs experts : l'inventeur Grosveld Franklin est au centre d'un important réseau de collaboration.

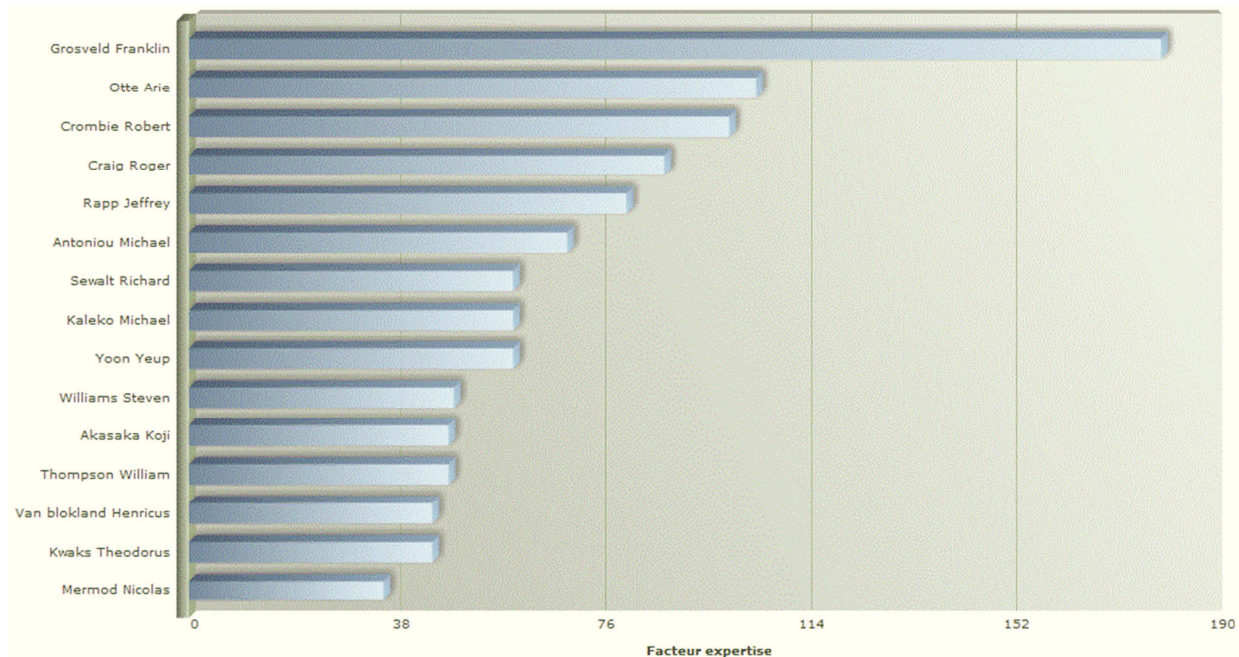


Figure 24 : Top 15 des inventeurs considérés comme experts par Intellixir. L'inventeur Grosveld Franklin est l'auteur expert incontournable dans cette étude.

- c. *Mobiles* : ce sont les inventeurs qui publient sous le nom de plusieurs déposants différents.
- d. *Apparition/disparition* : apparitions des auteurs pendant une période à déterminer.
- e. *Accélération* : auteurs qui ont la plus forte augmentation de fréquence d'utilisation dans une base (auteurs candidats à l'émergence).
- f. *Innovants* : auteurs premiers utilisateurs de concepts ou de codes CIB.

2) Affiliations (= déposants) :

- a. *Principales, sélectionnées ou toutes.*

La figure 25 présente les principaux acteurs de la technologie en question qui se sont positionnés sur le marché étudié. Le principal déposant est la société Vectura Group PLC avec un portefeuille de vingt et une familles de brevets sur le domaine.

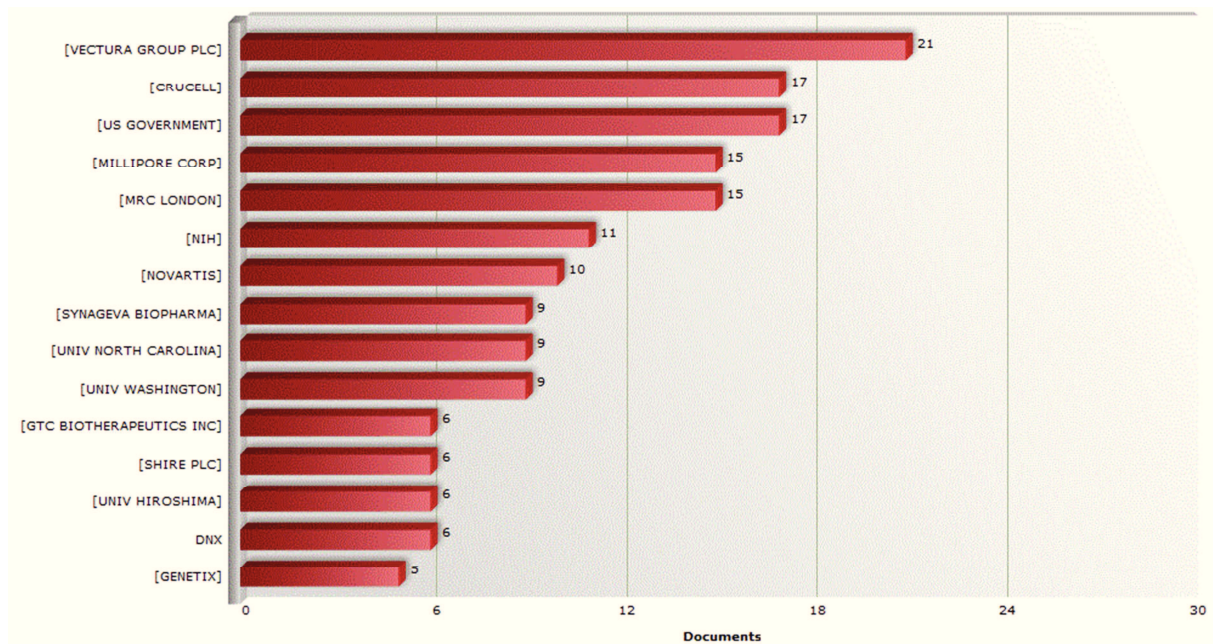


Figure 25 : Top 15 des déposants de brevets.

- b. *Apparition/disparition* : apparitions des affiliations pendant une période à déterminer.
- c. *Accélération* : affiliations qui ont la plus forte augmentation de fréquence d'utilisation dans une base (affiliations candidates à l'émergence).
- d. *Zoom top 20*.
- e. *Concepts communs* : concepts utilisés par les affiliations.
- f. *Innovantes* : premières utilisatrices, au travers de leur auteurs associés, de concepts ou de codes CIB.

C. Analyse du contenu :

- 1) Concepts : un concept est un mot ou un groupe de mots extrait des titres, des résumés et, dans le cas de brevets, des revendications. L'extraction se fait par la méthode des mots vides et triviaux qui supprime les mots vides (articles, adverbes...) et les mots triviaux isolés (mots vides de sens lorsqu'ils sont seuls comme « accumulation, additionnel ou formation ») et conserve les mots restants.
 - a. *Principaux ou tous*.

Intellixir définit un concept comme un « mot ou groupe de mots extrait des titres et/ou des résumés des documents brevets ». Plus la taille de la police du concept est grande, plus celui-ci possède d'occurrences dans la base Intellixir. Il s'agit de mots clés qui représentent les sujets principaux de l'étude de brevets. Ils peuvent être exploités pour segmenter les documents brevets intégrés dans la base Intellixir par des mots clés. Dans l'exemple ci-dessus, le concept *Mammalian cell* regroupe les documents brevets qui citent l'utilisation des cellules de Mammifères.

Adenovirus | Amino acid | Amino acid sequence | Animal cell | Bp | Cancel | Carrier | Cell culture | Cell line | Cell type | Cho | Cho cell | Chromatin | Chromosome | Clone | Cmv promoter | Code region | Code sequence | Complement | Copy | Cytokine | Delete | Deletion | Delivery | Derivative | Desirable | Dna construct | Dna molecule | Dna sequence | Downstream | Electroporation | Encode | Enhancer | Enzyme | Erythropoietin | Eukaryotic cell | Expression cassette | Expression level | Expression system | Expression vector | Flank | Functional fragment | Fusion protein | Gene encode | Gene expression | Gene product | Gene therapy | Genome | Globin | Globin gene | Goat | Green fluorescent protein | Harvest | High level | Hiv | Homologous recombination | Hormone | Host cell | Human beta | Human cell | Hybridize | Id | Immunoglobulin | Insertion | Instance | Insulator | Insulator element | Insulator sequence | Integration | Interferon | Intron | Ire | Isolate nucleic acid | Iv | Kb | Lcr | Lentiviral vector | Light chain | Locus control region | Ltr | Mammal | Mammalian cell | Mammalian host cell | Mar | Matrix attachment region | Milk | Naturally | Nucleic acid | Nucleic acid construct | Nucleic acid encode | Nucleic acid molecule | Nucleic acid sequence | Nucleic acid sequence encode | Nucleotide | Nucleotide sequence | Nucleotide sequence encode | Open read frame | Operably link | Operatively link | Organism | Peptide | Pharmaceutical composition | Pig | Plant cell | Plasmid | Polynucleotide | Polynucleotide sequence | Polypeptide | Position effect | Promoter | Promoter operably link | Promoter sequence | Protein production | Recombinant protein | Regulatory element | Regulatory sequence | Replication | Reporter gene | Retroviral vector | Rna | Sar | Scaffold attachment region | Secrete | Selectable marker | Seq | Seq id | Sequence encode | Sheep | Stably | Stem cell | Structural gene | Target cell | Target gene | Tissue specific promoter | Transcription | Transcription unit | Transfect | Transfect cell | Transfection | Transgene | Transgene expression | Transgenic animal | Transgenic plant | Translation | Ucoe | Upstream | Vector | Vector accord | Vertebrate | Viral vector |

Figure 26 : Nuage des mots clés ou concepts présents dans les documents brevets portant sur les éléments de contrôle de la chromatine.

- b. Concepts communs aux affiliations.
- c. Concepts communs aux auteurs.
- d. Apparition/disparition : apparitions des concepts pendant une période à déterminer.
- e. Accélération : concepts qui ont la plus forte augmentation de fréquence d'utilisation dans une base.
- f. Cartographie des concepts : information visuelle sur les concepts (nombre d'apparitions dans la base...).
- g. Signaux faibles : concepts apparaissant un grand nombre de fois dans peu de documents.

2) Brevets :

- a. *CIB, Sous-classes CIB ou Groupes CIB* : principaux ou tous, codes communs aux affiliations, apparition/disparition, évolution des codes.

Comme cela a été précisé précédemment, les symboles CIB classent et hiérarchisent les documents brevets selon les domaines technologiques auxquels ils se rapportent. Le Tableau 2 expose les différents domaines d'applications des inventions revendiquées dans les documents brevets.

CIB	NbDoc	Description
C12N15/85	157	Mutation or genetic engineering, Recombinant DNA-technology, Introduction of foreign genetic material using vectors, Vectors or expression systems specially adapted for eukaryotic hosts, for animal cells
C12N5/10	138	Undifferentiated human, animal or plant cells, e.g. cell lines, Cells modified by introduction of foreign genetic material, e.g. virus-transformed cells
C12N15/09	124	Mutation or genetic engineering, Recombinant DNA-technology
A61K48/00	119	Medicinal preparations containing genetic material which is inserted into cells of the living body to treat genetic diseases
C12N15/63	112	Mutation or genetic engineering, Recombinant DNA-technology, Introduction of foreign genetic material using vectors
C12N15/00	105	Mutation or genetic engineering
A01K67/027	94	Rearing or breeding animals, not otherwise provided for, New breeds of vertebrates
C07H21/04	80	Compounds containing two or more mononucleotide units having separate phosphate or polyphosphate groups linked by saccharide radicals of nucleoside groups, e.g. nucleic acids, with deoxyribosyl as saccharide radical
C07K14/435	72	Peptides having more than 20 amino acids, from animals
C12Q1/68	70	Measuring or testing processes involving enzymes or micro-organisms, involving nucleic acids
C12N5/06	67	
C12N15/11	53	Mutation or genetic engineering, Recombinant DNA-technology, DNA or RNA fragments
C07H21/00	51	Compounds containing two or more mononucleotide units having separate phosphate or polyphosphate groups linked by saccharide radicals of nucleoside groups, e.g. nucleic acids
C12N15/87	50	Mutation or genetic engineering, Recombinant DNA-technology, Introduction of foreign genetic material using processes not otherwise provided for, e.g. co-transformation
C12P21/02	49	Preparation of peptides or proteins, having a known sequence of two or more amino acids, e.g. glutathione
A61K38/00	48	Medicinal preparations containing peptides
C12N15/82	47	Mutation or genetic engineering, Recombinant DNA-technology, Introduction of foreign genetic material using vectors, Vectors or expression systems specially adapted for eukaryotic hosts, for plant cells
C12N15/86	46	Mutation or genetic engineering, Recombinant DNA-technology, Introduction of foreign genetic material using vectors, Vectors or expression systems specially adapted for eukaryotic hosts, for animal cells, Viral vectors
C12N15/12	45	Mutation or genetic engineering, Recombinant DNA-technology, DNA or RNA fragments, Genes encoding animal proteins
C12N15/67	44	Mutation or genetic engineering, Recombinant DNA-technology, Introduction of foreign genetic material using vectors, General methods for enhancing the expression

Tableau 2 : Description des principaux symboles CIB des brevets de la base créée dans Intellixir.

- b. *ECLA ou Groupes ECLA* : principaux ou tous, codes communs aux affiliations, apparition/disparition, évolution des codes.
- c. *Codes US* : principaux ou tous, codes communs aux affiliations, apparition/disparition, évolution des codes.
- d. *Citations* : brevets les plus cités présents et absents dans la base, liaisons entre les brevets cités et citant, liaisons entre les déposants cités et citant.
- e. *Familles* : familles de brevets les plus importants en nombre de brevets.

- f. *Statuts* : statuts des brevets définis par le *kind code* (code de type de brevet) et le code pays du numéro de brevet (par exemple le brevet JP3709485 B1 a un *kind code* B1 et un code pays JP).

3) Champs personnalisés :

- a. *Occurrences des champs*.
- b. *Cooccurrences des champs*.

L'utilisation des champs personnalisés est particulièrement performante car cela permet de faire des segmentations et ainsi d'affiner l'analyse du paysage brevets. Il est ainsi possible de découper la base de données constituée dans Intellixir selon des critères à définir : déposants industriels VS déposants institutionnels, segmentation des documents brevets selon leurs applications, focus sur un territoire déterminé comme l'Europe...

Par exemple, la figure 27 a été obtenue en segmentant l'ensemble de la base par des critères de mots clés.

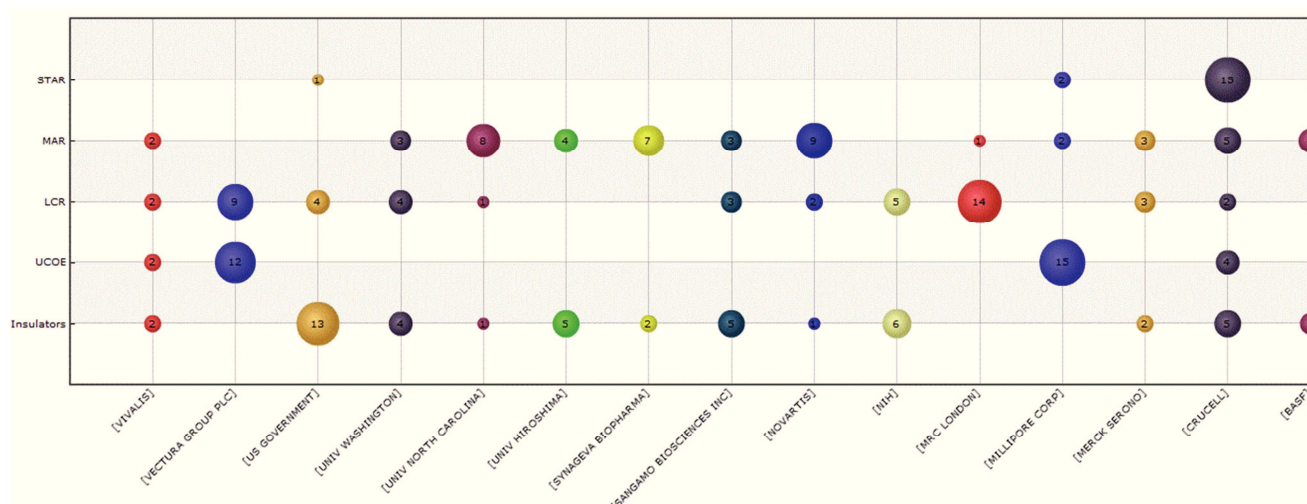


Figure 27 : Taille du portefeuille de brevets couvrant les systèmes d'éléments de contrôle de la chromatine pour chaque déposant. Le nombre au sein de la bulle indique le nombre de brevets pour l'élément en question, qui est directement proportionnel à la taille de la bulle.

Dans l'exemple, le graphique montre le positionnement technologique des déposants sur les différentes technologies favorisant l'expression de gène. L'analyse de ces résultats expose donc les déposants concurrents potentiels, comme les sociétés Synageva Biopharma et Novartis qui ont déposé respectivement sept et neuf brevets sur les éléments MAR.

Certains déposants se spécialisent avec des brevets portant principalement sur une technologie comme la société Millipore (quinze brevets sur dix-neuf concernent les éléments UCOE). D'autres déposants ne se spécialisent dans aucun système comme la société Vivalis qui a déposé des brevets couvrant les éléments MAR, LCR, UCOE et insulateurs. En recoupant avec d'autres informations comme les sites internet des déposants en question, il est possible de mieux comprendre et d'appréhender le positionnement technologique de chaque acteur. En effet, comme le signale la société Millipore sur son site, elle détient la PI et donc l'exploitation des éléments UCOE [w40].

Cette remarque montre qu'il est important de ne pas se limiter à l'information brevets pour les paysages brevets. En effet, les acquisitions, fusions et scissions sont des opérations courantes de l'économie de l'industrie des biotechnologies. Il est nécessaire de tenir compte de cette réalité économique dans l'analyse et l'interprétation des résultats fournis par l'outil Intellixir.

2.5. Recherche d'informations complémentaires

2.5.1. Informations économiques et industrielles

Pour comprendre le développement et les relations entre tous les acteurs des biotechnologies, que ce soit des personnes physiques (auteurs et inventeurs) ou des personnes morales (acteurs publics et privés), il est nécessaire de compléter les informations brevets et bibliographiques par des données d'autres natures. Il s'agit de toutes les informations stratégiques, économiques, financières, réglementaires voire politiques, qui portent sur le développement des biotechnologies. Les laboratoires (entreprises pharmaceutiques et de biotechnologies, laboratoires publics de recherche) interagissent avec les autres acteurs impliqués dans ce secteur comme par exemple les partenaires financiers (fonds d'investissements, *business angels*, capitaux risqués...), les agences réglementaires, les offices de brevet... Ces informations apportent des indices sur leur stratégie au sein de cet environnement. Elles regroupent toutes les opérations qui jalonnent les diverses activités des acteurs du secteur : opérations financières (levée de fonds, achat

d'une société, vente d'une branche de la société, introduction en bourse...), collaborations avec des partenaires complémentaires, stratégies de PI (*licensing-in* et *-out*, révocation d'un brevet concurrent...), intégrations d'experts scientifiques dans les directions des entreprises, réussites technologiques, autorisations réglementaires (accord de la *US Food and Drug Administration*...).

Ce type d'informations peut être trouvé en exploitant les sources citées à la partie 2.2.2 – Identification de mots clés à partir du web : les sites internet de sociétés et de laboratoires de recherche ou *via* les résultats donnés par les moteurs de recherche généralistes ou par les métamoteurs. Ici aussi, le choix des mots clés est important pour l'obtention de résultats pertinents. La pertinence de ces recherches d'informations complémentaires est illustrée avec la recherche des mots clés « *transposon company* » dans Google. Cette requête fait ressortir dans les dix premiers résultats, les sites internet des sociétés Minos Biosystems [w41] et Discovery Genomics [w42]. Ces dernières font partie des sociétés de biotechnologies créées à partir d'un système transposon [Palazzoli *et al.*, 2010a]. Le site de Discovery Genomics donne même une information sur la PI du transposon en question : *United States Patent 6,489,458; Hackett , et al. December 3, 2002; DNA-based transposon system for the introduction of nucleic acid into DNA of a cell*. Les sites internet des sociétés de biotechnologies peuvent donc comporter de nombreuses informations incontournables pour les paysages brevets. Ils comportent généralement une partie *Press Releases* ou *News* où des informations concernant des activités de la société sont publiées. Il peut s'agir par exemple de l'obtention d'un brevet [w43], d'une collaboration avec un partenaire [w44], de l'obtention ou de l'attribution d'une licence [w45], ou encore de l'acquisition d'une technologie [w46] ou d'une société [w47].

D'autres sites internet sont spécialisés dans le recensement et la publication de ces informations sur l'économie de l'industrie des biotechnologies et proposent régulièrement des résumés. C'est le cas en France et en Europe de la revue *Biotech Finances* [w48] qui rassemble chaque semaine l'essentiel des informations sur l'activité financière et les transactions dans le domaine des biotechnologies et informe des stratégies des acteurs des biotechnologies. C'est aussi le cas du site anglo-saxon *Bioworld* [w49] qui relate des actions en justice contre des brevets [w50] ou des opérations financières entre entreprises du secteur [w51]. Il peut également être intéressant de s'abonner à des newsletters dédiées

aux biotechnologies comme Buzz4Bio, anciennement Info Veille Biotech [w52], ou encore Flash Info Biotech [w53] dont l'un des objectifs est d'informer sur l'économie des biotechnologies.

Il est donc important de varier les sources d'informations provenant du web et de recouper les nombreuses informations qui en proviennent.

2.5.2. Informations liées à la littérature bibliographique

Comme cela a été expliqué précédemment, l'outil Intellixir permet d'analyser l'information brevets exportée depuis des bases de données brevets. Une autre exploitation possible d'Intellixir porte sur l'analyse d'articles scientifiques. Dans l'idéal, il serait intéressant de réaliser un paysage brevets dont les informations et l'analyse seraient complétées par un « paysage articles ». L'analyse de ce paysage bibliographique aboutit aux mêmes résultats que pour les brevets : mesure de l'activité de publication, orientations de R&D des centres de recherche, identification d'experts, de concurrents, de partenaires potentiels... Pour cela, les articles peuvent être extraits des bases de données Scopus, Web of science ou Medline. À la place de cette possibilité, j'ai utilisé le site GoPubMed [w54]. Cet outil ne remplace pas Intellixir mais propose quelques fonctionnalités qui s'en approchent comme des statistiques sur l'évolution des publications au cours du temps, les principaux pays dont sont originaires les publications, les principaux journaux dans lesquels les articles ont été publiés, les auteurs majeurs, les concepts exploités dans les articles ou encore les réseaux de collaboration entre co-auteurs (Figure 28).

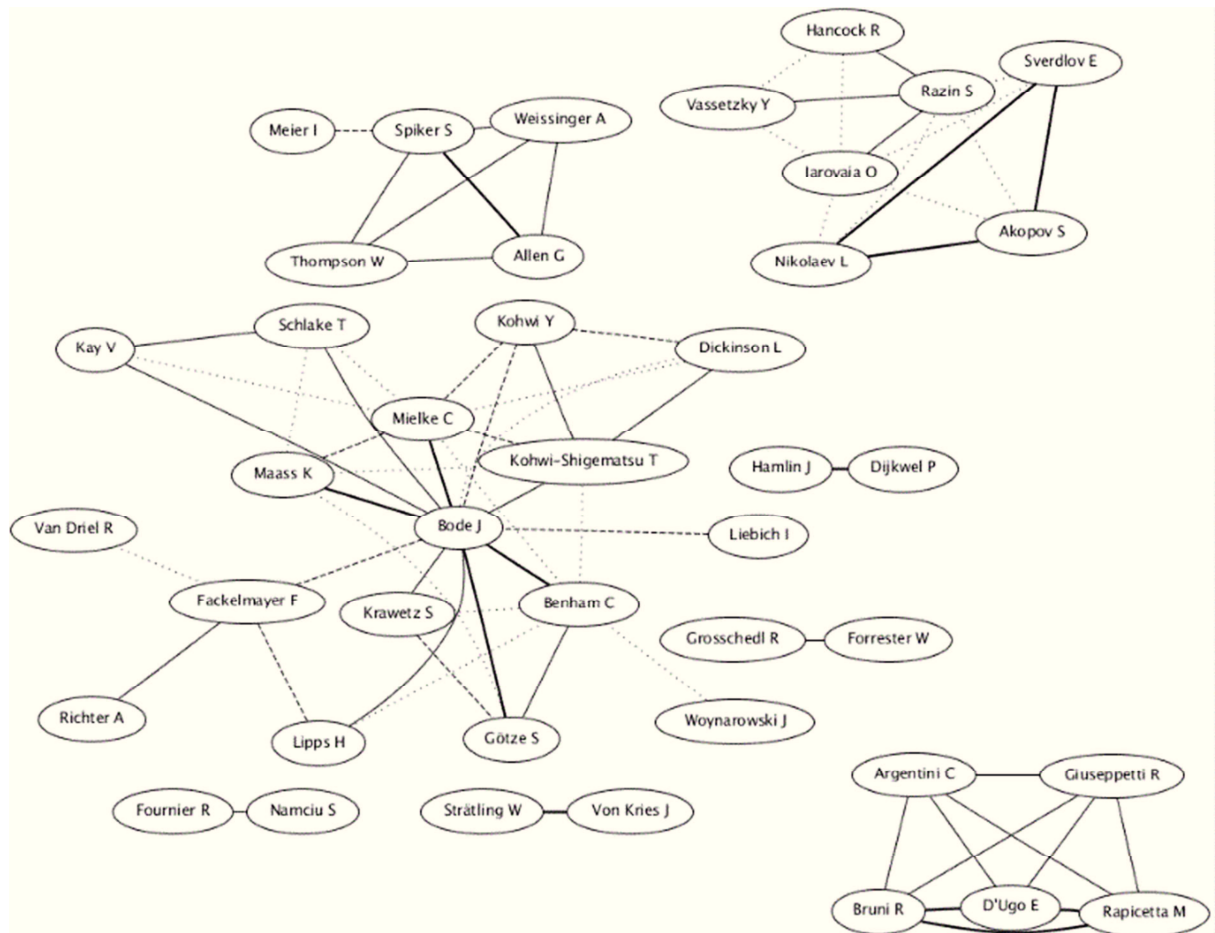


Figure 28 : Réseau de collaboration des auteurs d'articles pour la requête « Matrix Attachment Regions », créé par le site GoPubMed.

Les mots clés utilisés pour la requête dans GoPubMed sont soumis dans PubMed et les résumés qui en résultent sont ensuite classés selon le MeSH (*Medical Subject Headings*), le système hiérarchique de vocabulaire utilisé pour indexer les articles dans Medline/Pubmed. Concernant les statistiques, elles sont établies à partir des résumés des articles de PubMed.

3. Résultats

3.1. Article 2 : *Sustained transgene expression using non-viral enzymatic systems for stable chromosomal integration*

3.1.1. Contexte et objectif de l'article

L'utilisation d'un vecteur d'intégration représente une bonne stratégie pour avoir une expression à long terme et durable d'un gène d'intérêt thérapeutique. Les stratégies d'intégration de transgènes se scindent en deux approches selon l'origine du vecteur : virale ou non virale. Les virus sont des vecteurs efficaces mais leur utilisation chez l'homme n'est pas sans risque et soulève de nombreux problèmes d'innocuité et de sécurité pour les patients. Les vecteurs non-viraux représentent donc une alternative intéressante même s'ils sont moins efficaces pour réaliser une expression du transgène à long terme dans les cellules cibles.

L'objectif de cet article est de présenter les différentes stratégies enzymatiques non virales qui existent pour réaliser une intégration chromosomique stable. Nous avons comparé les propriétés de chaque système pour proposer une niche de développement technologique pertinente, tout en tenant compte de la PI qui les protège.

3.1.2. Résumé de l'article

L'article offre un panorama des stratégies non virales de transfert de gènes basées sur l'activité d'enzymes exogènes, pour réaliser l'intégration chromosomique d'un vecteur contenant un transgène. Dans une première partie sont synthétisées les propriétés naturelles et les caractéristiques des trois principales stratégies. La première option regroupe les nucléases à doigt de zinc et les méganucléases. Il s'agit d'endonucléases réalisant des coupures double brin site-spécifiques pouvant induire des modifications génétiques par recombinaison homologue. La deuxième alternative concerne les intégrases du type de celle du phage ϕ C31, qui sont des recombinases nécessitant la présence de sites spécifiques (*attB* et *attP*) pour intégrer un vecteur. La dernière solution est basée sur les transposases (*Sleeping Beauty*, *Tol2* et *piggyBac*, *Minos*, *FrogPrince*, *Himar1*, *Hsmar1* et *Mos1*, *ISY100* et *Tn7*) qui permettent à des fragments d'ADN mobiles, les transposons, de se

déplacer dans les génomes. Les vecteurs qui dérivent de ces transposons rendent possible l'insertion des transgènes qu'ils contiennent.

Dans la seconde partie de l'article, nous avons décrit les effets génotoxiques et délétères de l'utilisation des systèmes enzymatiques non viraux, qui proviennent principalement des activités de clivage non-spécifiques. Ils comprennent l'intégration non désirée à la suite d'une recombinaison non-spécifique, le silençage de l'expression des transgènes intégrés et l'impact de l'intégration sur l'environnement chromatinien du site d'intégration.

La troisième partie de l'article porte sur les définitions d'un bon site d'intégration et des niches de développement des systèmes d'intégration enzymatiques non viraux. L'article se termine par une section concernant la PI qui protège ces systèmes enzymatiques et qui complète la notion de niche de développement technologique. En effet, les systèmes enzymatiques d'intégration ont été protégés par des brevets qui couvrent de larges applications, allant de la thérapie génique jusqu'à leur utilisation en tant qu'outil de biologie moléculaire.

3.1.3. Conclusion de l'article

Les récents progrès en ingénierie sur les systèmes d'intégration non-viraux montrent que ceux-ci représentent des alternatives prometteuses aux stratégies virales de transfert de gènes. Ils offrent la possibilité de réparer des gènes défectueux et de réaliser des intégrations d'ADN exogène dans l'ADN chromosomique. Cependant, des investissements importants sont nécessaires pour permettre une exploitation efficace et adaptée de ces technologies non virales. Dans les prochaines années, comme c'est en partie le cas pour les vecteurs viraux, ces technologies devront être en mesure de répondre aux défis que représentent la gestion de la génotoxicité des vecteurs non-viraux *in vitro* et *in vivo*.

Un des facteurs limitant le développement de ces vecteurs porte sur la PI qui les couvre et qui risque de freiner les projets de R&D sur l'intégration enzymatique non virale. Souvent, ces systèmes sont protégés par des brevets appartenant à quelques acteurs du domaine. En effet, de nombreux systèmes font partie intégrante du plan d'affaires (*business*

plan) de sociétés de biotechnologies qui ont été créées dans le but d'exploiter les systèmes en question (Cellestis pour les méganucléases, Sangamo Biosciences pour les nucléases à doigts de zinc...). Par conséquent, il est indispensable d'en tenir compte avant d'élaborer et de financer un plan d'exploitation basé sur ces systèmes.

3.1.4. Conséquences pour le choix des études et la suite de mes travaux

Cet article [article 2 ; [Palazzoli et al., 2008](#)] et ceux de Harraghy et Kwaks & Otte sur les éléments de contrôle de la chromatine [[Kwaks & Otte, 2006](#) ; [Harraghy et al., 2008](#)] ont permis d'organiser mon travail de thèse en définissant les domaines et les objets des études de brevets (exemples de domaine : transposons ou éléments de contrôle de la chromatine, exemples d'objet : *Sleeping Beauty* ou insulateurs). Comme cela a été exposé précédemment, l'objectif de SyntheGeneDelivery était de développer un vecteur transposon. Par conséquent, un paysage brevets plus approfondi de ces systèmes a été réalisé [article 3 ; [Palazzoli et al., 2010a](#)]. L'article de *Current Gene Therapy* a aussi permis de clarifier les stratégies de gestion de la génotoxicité à travers deux stratégies, qui ont aussi fait l'objet de paysages brevets. La première porte sur la spécificité d'intégration du transgène. Les travaux se sont concentrés sur la spécificité d'insertion et sur les outils moléculaires qui permettraient une insertion dans les séquences codantes des gènes ribosomaux [articles 4 et 5 ; [Palazzoli et al., 2010b](#) ; [Carnus et al., 2011](#)]. La seconde stratégie repose sur l'utilisation d'éléments de contrôle de la chromatine qui sont intégrés dans le vecteur d'expression enzymatique non viral. Ces éléments sont un bon moyen pour assurer l'expression du transgène et pour limiter le nombre d'intégrations du vecteur à une seule insertion [article 6 ; [Palazzoli et al., 2011](#)]. L'objectif de ces panoramas mondiaux de la PI était d'identifier des niches de développement technologique pour le GICC, c'est-à-dire des technologies qui soient libres de droits de PI.

3.1.5. Remarque sur la méthodologie et les outils employés

Une remarque peut être faite quant à la méthodologie et aux outils exploités pour la recherche de brevets sur les systèmes d'intégration enzymatiques non viraux. En effet, les travaux ont été obtenus pendant ma première année de thèse, quand nous n'avions aucun outil commercial à disposition tel que l'outil de recherche QPAT ([w13], acquis au début de ma deuxième année de thèse). Nous avons utilisé la base esp@cenet [w30] qui, comme cela a été exposé auparavant, comporte certaines limitations. Par exemple, le fait de ne pas pouvoir rechercher dans les revendications peut créer un biais dans les résultats. En utilisant l'outil de recherche QPAT, le nombre de résultats aurait été plus important qu'avec esp@cenet. Cet inconvénient n'a cependant pas été un problème pour cet article dont le but était de proposer un panorama des systèmes existants et non pas de présenter un paysage brevets complet ou de réaliser une étude de liberté d'exploitation.

3.1.6. Article 2

Sustained Transgene Expression Using Non-Viral Enzymatic Systems for Stable Chromosomal Integration

Fabien Palazzoli^{1,2}, Elodie Carnus^{1,2}, Dominic J. Wells⁴ and Yves Bigot^{1,2,3,*}

¹Université François Rabelais de Tours, GICC, ²CNRS, UMR 6239, ³CHRU de Tours, UFR des Sciences & Techniques, Parc de Grandmont, 37200 Tours, France; ⁴Gene Targeting Group, Department of Cellular and Molecular Neuroscience, Hammersmith Hospital Campus, Imperial College London, London, United Kingdom

Abstract: Gene delivery technologies have been developed for various biotechnology applications. In gene therapy, they are promising for the treatment of several inherited and acquired human diseases. When therapies require the transfection of a transgene, the vector integration is one of the solutions that is used for maintaining and sustaining expression. On the basis of their origin, vectorisation technologies are currently divided in two fields, gathering on one hand viral vectors and, on the other hand, non-viral approaches. In the case of the non-viral therapies, three main sub-fields are in progress to integrate transgenes. The first uses oligonucleotides to stimulate targeted gene repair by homologous recombination. The second is based on site-specific endonucleases for which the cleavage activity is used to stimulate the host recombination mechanisms in the presence of a DNA vector. The third one is developed from phage and transposon enzymatic systems. The two last sub-fields use non-viral enzymes and are the scope of this review.

Here, our objective was to overview the main non-viral enzymatic systems able to integrate DNA cassettes. Their molecular and functional characteristics are summarized, and their properties and limits in the current state of the art highlighted. An overview of the safety and quality issues is also presented and discussed, taking into account the solutions that might circumvent problems, intellectual property and economic status for each system. As a conclusion, we propose projections of the future technological developments in the context of the different interests for public and private bodies.

Keywords: Nuclease, meganuclease, integrase, transposase, transposon, targeting, recombination.

INTRODUCTION

In 2007, over 1340 clinical trials have been completed, are on going or have been approved worldwide, using over 100 genes [1]. Since 2004, the percentage of trials using viral strategies remained stable (67%). For non-viral therapies using plasmid DNA, this rate has increased and has reached 18%. Although some studies in vector engineering aimed to develop systems made with viral and non-viral gene components [2, 3, 4, 5] and whereas most non-viral strategies mainly mimic what virus do, both approaches are commonly felt as being in rupture at levels of the concepts and technologies. Another way to distinguish strategies among viral and non-viral gene therapies is to consider the mode of maintaining gene expression in the genetically modified cells. Indeed, numerous viral and non-viral therapies used so far in clinical trials involved extra-chromosomal DNA vectors containing the transgenic constructs. To date, retrovectors are the only efficient vectors able to ensure maintenance of the transgene by integration within chromosomes. Wild adeno-associated virus (AAV) are able to integrate at a specific site within the human chromosome 19q under *in vitro* laboratory conditions [6]. However, this property would be deleterious if it could occur *in vivo* in patients [7], and is not conserved in recombinant AAV.

In the case of therapies concerning non-dividing cells with a low regeneration rate, the maintenance of an extra-chromosomal vector raises few problems. In tissues where cells are frequently regenerated, the maintenance of the vector is of crucial importance and is directly related to its ability to be accurately replicated and segregated during cell division. Viral mechanisms such as that involving the EBNA1 protein from EBV are able to ensure an extra-chromosomal genome maintained in dividing cells [8]. However, the deleterious effects of such a mechanism on the biology of the genetically modified cells do not allow their use in gene therapy yet [9]. Beside replication and maintenance in cells, the middle and long-term presences of extra-chromosomal DNA molecules within the cellular nucleus is the origin of other problems that can be highly deleterious, whatever the vector used. For example, integrations occur and create deleterious mutations in host chromosomes that can lead to the emergence of proliferative cancer cells if episomal papillomavirus, HBV genome [10, 11, 12], or rAAV vectors [13, 14] are middle or long-term maintained *ex vivo* and *in vivo*. Similarly, the introduction of extra-chromosomal molecules such as plasmid DNA in the nucleus of eukaryotic cells can lead, on one hand, to the creation and the maintenance of very large recombinant DNA molecules just after transfection [15] and, on the other hand, to random integration events [7, 16, 17, 18] that are stimulated by random single-stranded and double-stranded DNA cleavages in the chromosomal DNA [19]. These integration events, over middle and long-term time duration and at the population

*Address correspondence to this author at the GICC, UMR CNRS 6239, Université François Rabelais, UFR des Sciences & Techniques, Parc de Grandmont, 37200 Tours, France; Tel: +33 2 47 36 70 63; Fax: +33 2 47 36 70 42; E-mail: bigot@univ-tours.fr

scale of several million cells, can also lead to the emergence of proliferative cancer cells.

The integration of the vector is one of the solutions that can be used for maintaining, and sustaining transgene expression in the cells transfected for gene therapy purposes. In the case of the non-viral therapy, three main strategies are under investigation and none of them is for the moment involved in clinical trials. The first strategy uses oligonucleotides to stimulate targeted gene repair by homologous recombination (HR) in an attempt to correct point mutations. The two other strategies require the use of exogenous enzymes. The second is thus based on the use of site-specific endonucleases for which the cleavage activity is used to stimulate the host recombination mechanisms in the presence of a DNA vector. It allows two kinds of DNA rearrangements: to correct point mutations and to insert a DNA fragment at a precise locus. The third strategy was developed from two kinds of mobile genetic elements (MGE; so-called parasitic genetic elements (PGE)), phages and transposons, which use a DNA molecule as an intermediate. The processes involving MGE as vectors generally require a source of enzyme and a DNA vector donor that is a modified integron or a transposase donor. They are mostly used to insert DNA fragments.

The scope of the present review does not include approaches involving oligonucleotides that were recently reviewed [20, 21] but concentrates on approaches involving enzymes such as endonucleases, recombinases and transposases. Their natural properties of integration efficiency and specificity, their genotoxic effects and the problems met to sustain transgene expression are synthesized. Approaches that aim to circumvent the encountered problems are presented and the expected performances and perspectives of safety and quality for these non-viral integration systems are finally discussed.

1. DIVERSITY OF THE NON-VIRAL ENZYMATIC INTEGRATION SYSTEMS

To date, two kinds of non-viral enzymatic mechanisms have been investigated for gene therapy purposes. The first corresponds to artificial DNA endonucleases derived from type II restriction enzymes of bacterial origin and homing endonucleases, so-called meganucleases that are found in phages, bacteria, archaeobacteria and various eukaryotes [22]. These proteins are used to stimulate the host mechanisms of HR in order to correct point mutations or to insert DNA fragments. The second kind of proteins are DNA phosphotransferases that mostly originate from prokaryotic and eukaryotic MGE. Based on their chemistry, two sub-types of phosphotransferases are used, recombinases and transposases. In contrast to nucleases, the gene delivery processes using these enzymes do not require specific host factors to trigger the integration. They mobilize the host machinery to fill and repair the ends of the vector only after integration.

1.1. Endonucleases

1.1.1. Polydactyl Zinc Finger-Nucleases (pZF-Nucleases)

The pZF-nuclease system was mainly developed from a type II restriction enzyme, *FokI*, isolated from the bacteria

Flavobacterium okenokoites. Its homo-dimerization domain is required to recognize the asymmetric sequence 5'-GGATG(N)_{9/13}-3' and to cleave double-stranded DNA at staggered sites [23]. *FokI* comprises a N-terminal DNA recognition domain, containing three helix-turn-helix subdomains, and a C-terminal DNA cleavage domain, containing an Asp-450, Asp-467 and Lys-469 catalytic triad close to that of the restriction enzyme *BamHI* [24]. The modular structure of *FokI* allows construction of hybrid endonucleases with novel sequence specificities by substituting its DNA binding domain with other DNA binding domains. The potential of this modularity was firstly demonstrated by constructing a functional enzyme with a new cleavage specificity. This was obtained by linking the DNA-binding domain of the *Drosophila Ubx* homeodomain to the DNA cleavage domain of *FokI* [25]. More interestingly, hybrid enzymes were obtained with zinc finger DNA binding domains (ZFD) [26, 27].

ZFD are of special interest in engineering the cleavage specificity of hybrid endonucleases made with the *FokI* C-terminal DNA cleavage domain. Indeed, beside their diversity and prevalence, they are the most abundant DNA binding domains in eukaryotes [28]. A revolutionary technology of molecular engineering was developed 20 years ago to obtain tailor-made DNA binding domains able to specifically bind to any DNA sequence ranging in size from 3 to 18-bp. This technology was developed from a Cys₂His₂ ZFD backbone related to those found in the TFIIIA and Zif268 proteins [29, 30]. This backbone is a 30 amino acid unit consisting of a compact $\alpha\beta\beta$ fold internally stabilized by chelation of a single zinc ion (Fig. 1a). Structural analyses have revealed that one ZFD unit typically recognizes three contiguous base pairs of a DNA sequence, and 7 amino acid residues located from positions -1 to 6, with respect to the start of the α -helix, are responsible for contacting DNA. Carlos Barbas and colleagues have constructed a set of ZFD units that target all GNN, most ANN and CNN and a few TNN triplet subsites, thus opening the possibility to generate ZFD arrays, so-called pZFD, able to specifically bind to a DNA triplet oligomer [for review, see 31 and 32]. Different tools of molecular engineering and *in silico* design were developed to optimize the pZFD affinity for their DNA target [for review, see 33, 34, 35, and <http://www.zincfingers.org/>]. Because efficient double-stranded DNA cleavage requires dimerization through the *FokI* C-terminal DNA, the possibility to fuse it with a pZFD designed and selected to specifically bind to a sequence of 9 nucleotides (Fig. 1b), opens opportunities to target 18-bp long recognition sites, long enough to be unique in the human genome [http://www.sangamo.com/human/human_thera_overview.html]. A single pZF-nuclease can therefore target cleavage in a palindromic motif in which inverted 9-bp repeats are separated by an optimal spacer of 6 to 8 nucleotides (Fig. 1c) [36]. When no palindromic target is present, two different and complementary pZF-nucleases can also be used (Fig. 1d) [37, 38].

1.1.2. Meganucleases

Meganucleases are endonucleases that are often encoded by introns or inteins (so-called "protein-intron"), behaving as a MGE. They recognize large DNA target sites (>12-bp) where they produce a double-stranded cleavage [39,

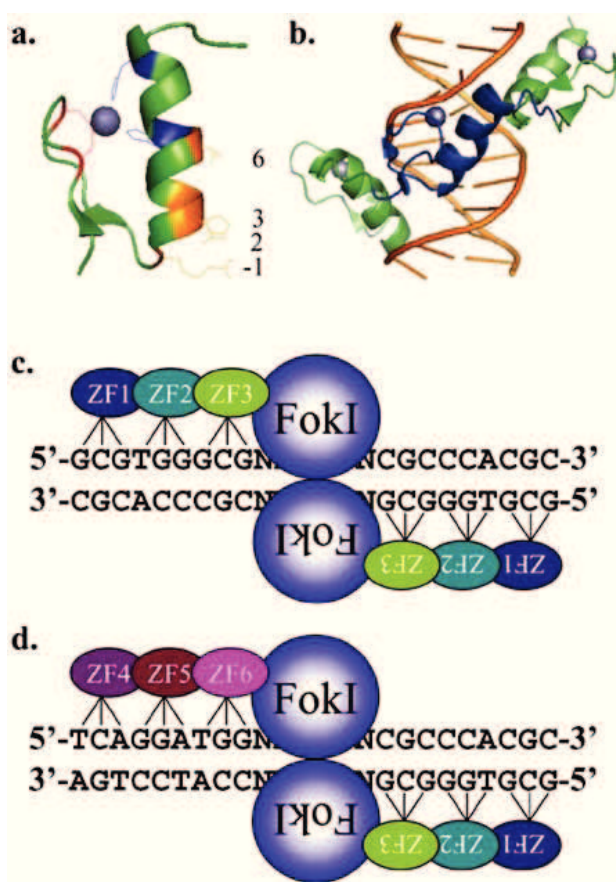


Fig. (1). Zinc finger and pZF-nuclease properties. **a)** Canonical structure of a zinc finger module consisted of two anti-parallel β -strands (green arrows) and an α -helix, with a zinc cation (grey sphere) coordinated by four conserved residues, two cysteines and two histines (respectively located in red and blue with their side chain). The numbered amino acid residues in orange with their side chain at position -1, +2, +3 and +6 relative to the amino-terminal end of the α -helix are important residues contacting in the major groove of the double-stranded target DNA. **b)** Structural organization of a trimeric zinc finger domain bound to the major groove of its double-stranded target DNA. In a and b, ribbon representations were drawn from the PDB file code 1AAY and MacPyMal facilities. **c)** Homodimeric zinc finger nuclease bound to a palindromic target site. **d)** Heterodimeric zinc finger nuclease bound to a non-palindromic target site. A pZF-nuclease target site consists of two similar (c) and different (d) binding sites on complementary strands, separated by a spacer of 5-6 nucleotides. In this configuration, recombinant FokI endonuclease monomers (dark blue spheres) fused to the C-terminal of each Zinc finger arrays can form an active dimeric nuclease able to cleave double-stranded DNA in the 5-6 nucleotides spacer between the two pZFD binding sites.

<http://www.collectis.com/homing.php>]. These cleavage sites usually correspond to intron-free or intein-free genes. Meganucleases are classified into four families named after conserved peptide motifs: the LAGLIDADG family, the His-Cys box family, the GIY-YIG family, and the HNH family. To date, only members belonging LAGLIDADG family, I-CreI and related proteins such as I-AniI, I-CeuI, I-MsoI and

I-SceI, have been engineered for gene transfer purposes, including gene therapy [for review, see 40]. Usually, these natural proteins cleave DNA as homodimers or single chains. In structure, the LAGLIDADG motifs are central with two characteristic $\alpha\beta\alpha\beta\alpha$ folds facing each other across a 2-fold pseudo-symmetry that separates monomers (Fig. 2a). In each monomer, a four-stranded β -sheet provides a DNA binding interface that drives the interaction of the protein with the target DNA sequence [41, 42]. A natural meganuclease therefore targets a double-stranded DNA cleavage within a palindromic motif (Fig. 2b).

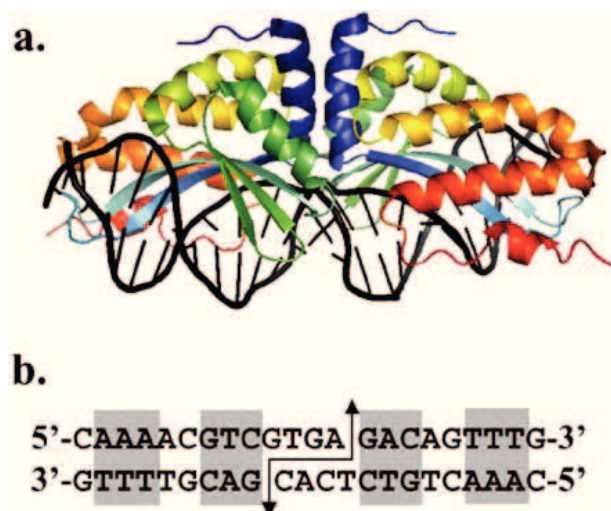


Fig. (2). Meganucleases and target site properties. **a)** Structural organization of I-CreI dimer bound to its DNA target. The α -helices and β -strands contained in each of both monomers are represented from N-terminal to C-terminal ends following the rainbow colours (blue to red). Ribbon representation was drawn from the PDB file code 1G9Y and MacPyMal facilities. **b)** Sequence of the natural I-CreI target site. Trinucleotides boxed in gray correspond to nucleotides that can be altered to engineer new I-CreI variants that can cleave DNA as a homodimer or a heterodimer, depending on the selection process used.

Although they result from approaches and research teams that are unrelated to those involved with the pZFD-nucleases, processes to engineer the binding specificity of the meganucleases have been developed. They now allow designing many new meganucleases with predictable substrate specificities by domain swapping and fusion of homing meganuclease domains. Mutagenesis also allowed creation of heterodimeric endonucleases that are able to cleave non-symmetric DNA targets [43, 44, 45]. Overall, these efforts have demonstrated that the plasticity of LAGLIDADG endonucleases allows extensive engineering, and provide a general method to create a second kind of nucleases with tailored specificities for sites that do not have to be palindromic.

1.1.3. Stimulating Point DNA Modifications and Targeting Insertion with Endonucleases

Genome DNA modifications with endonucleases use host mechanisms to repair DNA double-stranded breaks (DSB).

DNA DSB are naturally occurring events, potentially causing chromosomal rearrangement and cell death. There are two primary repair pathways in mammals: i) the non-homologous end-joining (NHEJ) that simply ligates back together broken ends and ii) homologous recombination (HR) that is a very conserved DNA maintenance pathway involved in the repair of DNA DSB or other lesions. To repair lesions, HR promotes exchange of genetic information between endogenous homologous sequences as a repair template, in a manner most consistent with the synthesis-dependent strand annealing (SSDA) model (Fig. 3a) [for review, see 34]. Usually, cells use the sister-chromatid as a template, but other homologous loci or exogenous DNA constructs can be involved in repair. In mammalian cells, the natural occurrence rate of HR between a locus and a homologous exogenous DNA construct ranges from 10^{-7} to 10^{-2} , depending on the cell type and the transfection method used [46].

One of the technologies developed to stimulate HR is to use sequence-specific endonucleases that create sites-specific DSB. Indeed, DSB and sequence similarities between the cleaved site and the DNA template elicit host repair machinery. Importantly, this technology is not limited to gene cor-

rection in order to restore a functional gene (Fig. 3a) or to mutate a gene (Fig. 3b), but it can also be used to target insertion or gene inactivation by insertional mutagenesis (Fig. 3c), thus opening a wide range of applications spanning beyond the field of the gene therapy. Using pZF-nucleases to stimulate HR, it was found that the rates of targeted human gene correction reach from 5 to 15% *ex vivo*, depending on the nature of the correction (point mutation correction of a functional gene or integration of a DNA fragment), within 72h of treatment and with no selection of the desired events [37, 38]. Under similar conditions, it was found that maximal rates of 1 to 10% are obtained *ex vivo* when gene corrections are simulated by the I-SceI meganucleases [40]. Overall, this indicates that both kinds of endonucleases have for the moment similar efficiencies to stimulate gene corrections *ex vivo* by HR. Due to the fact that HR is mostly active during cell division, and in the absence of published data on quiescent cells, pZF-nuclease and meganuclease systems can be considered efficient only in dividing cells [47]. Nevertheless, the actual challenge between both systems is not the rate of gene correction, but more likely the robustness of the design and selection processes to obtain efficient endonucleases and

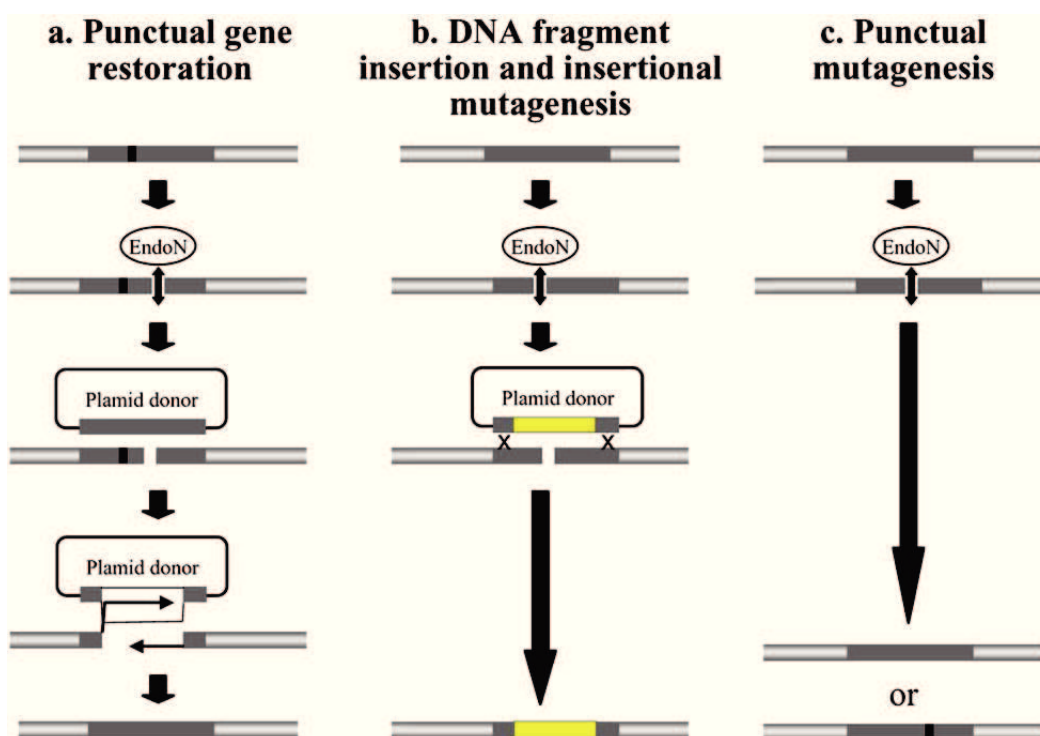


Fig. (3). Endonuclease-induced genetic modifications by HR. Endonuclease (EndoN)-induced homology-directed DSB repair in mammalian cells is based on the repair of the gene by HR after the induction of a specific DSB. After cleavage, the DSB is processed to form free 3' single-strand tails, a process that requires the Mre11/Rad50/Nbs1 complex. The HR machinery, through the actions of the strand invasion protein, Rad51, then uses the free 3' ends to invade a homologous repair template/donor. How the machinery identifies a homologous repair donor remains unclear but it is likely that simple physical proximity plays an important role. In the normal repair of a DSB, the repair donor is the sister-chromatid and thus the template is identical to the damaged allele. In gene targeting, the repair donor would be an extra-chromosomal piece of DNA, in general a plasmid containing the targeted DNA fragment with the modifications. After strand invasion, primed DNA synthesis occurs to generate new undamaged DNA using the undamaged DNA as a template. The process is completed by the annealing of the new strand of DNA with its original partner and subsequent use of that new DNA to template DNA synthesis. Endonuclease-induced gene targeting can be used for gene correction (a) or fragment DNA gene insertion (b) at the meganuclease cleavage site. The DSB can also be repaired by Non-Homologous End Joining (NHEJ), an error prone process that can either restore the initial sequence, or result in small insertions or deletions of various sizes (c).

the management of their genotoxicity.

1.1.4. Translocating Endonuclease Systems into the Nucleus for Gene Correction

To date, gene delivery assays with both endonuclease systems have been investigated *ex vivo*, using dividing mouse embryonic stem (ES) cells and common cell lineages. A two-plasmid system is usually used for transfection employing lipofectin or polyethylenimine (PEI) as transfecting agents, or by nucleofection ([48], Amaxa Biosystem, Germany). The first plasmid is an expression plasmid encoding a pZF-nuclease or meganuclease. The second plasmid contains the gene of interest that is used as donor template for HR. When using a heterodimeric system of endonucleases, two expression plasmids are co-transfected.

1.2. □□c□□bi□a□□a□d □□a□□□□a□□□□□□□□

Integration systems described in this section and investigated for gene therapy purposes were all isolated from MGE, such as phages and DNA transposons of bacterial and eukaryotic origins. Integrase and transposase systems however are distinguished by the origin of the enzyme and the chemistry of the reaction involved in the specific integration of the transgene. The first system uses site-specific recombinases [49], for which the best investigated system is the serine-integrase of the phage ϕ C31. The second involves transposons belonging to the subclass 1 of the Class II DNA transposons [50, 51], the best investigated elements in vectorization for gene therapy purposes being related to the *IS630-Tc1-mariner* (*ITm*), *hobo-Ac-Tam* (*hAT*) and *piggyBac* superfamilies. Although this feature has been rarely investigated to date, one characteristic of the recombinase and transposase systems is that they should be able to specifically integrate exogenous DNA into chromosomes in non-dividing cells. Indeed, they do not require DNA replication to occur and can mobilize DNA repair pathways other than HR and NHEJ [52-56].

1.2.1. Site-Specific Recombinases

All site-specific recombinases so far identified fall into one of just two families, the tyrosine recombinases and the serine recombinases [49, 57]. Both family names were defined by the amino acid residue that forms a covalent protein-linkage in the reaction intermediate. The main biochemical characteristic of these proteins that distinguish them from other recombinases is that the nucleophilic attack of the DNA backbone leads to the formation of an intermediate in which a tyrosine or a serine residue of the protein is covalently bound to the released phosphate on both DNA strands [58]. The main models of tyrosine recombinases that have contributed to understanding their architectural, biochemical and mechanistic properties in the recombination reaction are the lambda Int protein, the Cre protein of phage P1, the XerC/XerD proteins of *Escherichia coli* and the FLP protein of yeast. Whereas Cre/*lox* and FLP/*frt* have been widely used for genomic modification, so far tyrosine recombinases have rarely been investigated for development of transgene integration systems for gene delivery purposes. The requirements in host factors helping the phage integrases and the reversibility of the transgene integrations specifically

mediated by Cre and FLP are probably the main causes that have impeded their uses in this field of application [for review see 59].

Phylogenetic analyses revealed that there are three major groups of serine recombinases, represented by resolvase/invertase related to $\gamma\delta$, Hin and Gin proteins, relatives of IS607 transposases and large serine recombinases of which the best studied representative is the integrase of the *Streptomyces* phage ϕ C31 [49]. To date, the activity of three serine integrases have been studied in mammal cells, ϕ C31, R4 and TP901-1. That of the ϕ C31 phage has been the most investigated as an integration system, for gene therapy purposes.

1.2.1.1. ϕ C31 Integrase

At present, the structure of phage serine-integrase related to ϕ C31 has not been determined. However, sequence analyses reveal that the three groups of serine recombinases share a catalytic domain of about 120 amino acids for which a structure was determined from the $\gamma\delta$ serine resolvase. The catalytic domain contains a set of 4 motifs, named A, B, C and D, in which A and C were demonstrated to contain the critical residues of the serine recombinase active site. In the ϕ C31, the catalytic domain is located at the N-terminal end and is linked to a large C-terminal domain of about 400 amino acids whose sub-domain organization, structure and function remain to be clarified (Fig. □a) [60].

In its host *Streptomyces* genome, the ϕ C31 integrase needs to pair a phage *attP* site with a chromosomal or plasmid *attB* site to mediate integration. In nature, the DNA sequences of *attB* and *attP* sites are 30-40-bp long and are not perfectly similar, with only 50% of the nucleotide positions being similar (Fig. □b). The reaction between *attB* and *attP* sites is unidirectional, does not require any host co-factor, and leads to the formation of hybrid *att* sites at both ends of the integrated phage, named *attL* and *attR* sites which are unable to be remobilized by the ϕ C31 integrase (Fig. □c) [61, 62, <http://www.stanford.edu/~calos/research.shtml>].

1.2.1.2. Targeted Insertion with the ϕ C31 Integration in Mammal Cells

The ϕ C31 integration system has been shown to be active in human cells. Depending on the cell type, it is able to promote integrations in 0.1 to 15% of the transfected cells with no selection of the desired events [for review see 63]. When expressed from a mammalian promoter, the ϕ C31 integrase mediates precise integration of plasmids bearing an *attB* site into native chromosomal target sequences that share partial identities with *attP* (Fig. □d). Sequence and statistical analyses have revealed that between 202-764 pseudo-*attP* sites are present in the human genome [64]. Integration assays done on cell cultures indicate that about 20 sites are preferentially used by the ϕ C31 integration system, these preferences vary widely depending on cell type. Based on this observation, one of the arguments to promote the ϕ C31 integration system for gene therapy purposes is based on the fact that in the absence of coincidence between the pseudo-*attP* sites and the proximity of the genes known to lead to proliferative cancer (346 genes [65]), the integration properties of this system should be safer than systems mediating random integrations.

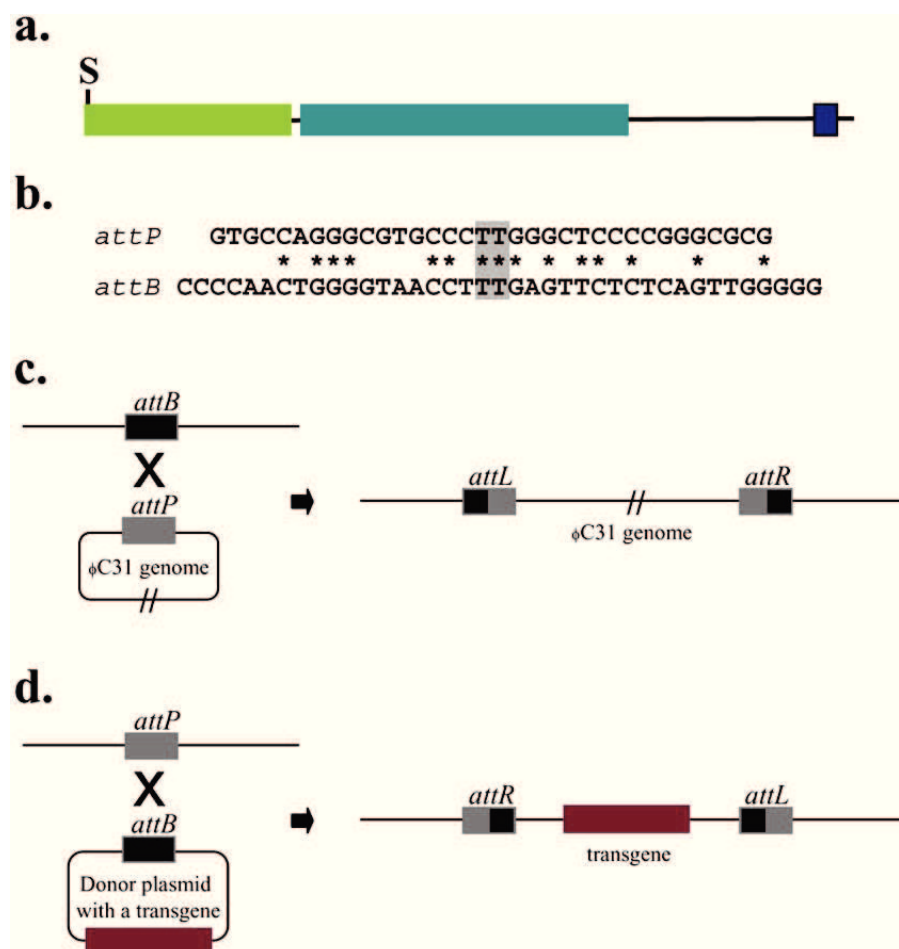


Fig. (1). Properties of the ϕ C31 vectors. **a)** Domain organization of the ϕ C31 recombinase. The green box represented the conserved catalytic domain of all serine-recombinases, with S showing the position of the serine nucleotide involved in cleavage activities. Turquoise and dark blue boxes corresponded to the cysteine and the Leu/Iso/Val/Met regions. The cysteine region is about 220 amino acid long and contains at least 10 highly conserved residues, including three cysteines, two of which being required for recombination. Data were obtained with a closely related protein, *spoIVCA*, recombined that the cysteine region would contain in its amino terminal part the DNA binding activity [49]. **b)** Sequence comparison of the *attP* and *attB* sites. (□) indicated identical positions. The central TT core shared by all *att* sites is boxed in gray. A 2-bp staggered double-stranded cleavage is made at each 3' end of the core during recombination. **c)** and **d)** Schematic diagrams of ϕ C31 integrase-mediated recombination [62]. **c)** In nature, ϕ C31 integrase performs precise recombination between an *attB* site located in the *Streptomyces* genome and an *attP* site located on the ϕ C31 phage genome. The outcome is integration of the phage into the host genome. The ϕ C31 genome is nearly 50 kb long, so integration of sizeable sequences by this enzyme is natural. **d)** In mammalian cells, the ϕ C31 integrase system works optimally if the plasmid to be integrated carries an *attB* site, rather than *attP*. ϕ C31 integrase performs recombination between *attB* and naturally-occurring genomic sequences related to *attP*. The typical outcome is integration of a single copy of the *attB*-containing plasmid into the mammalian genome.

To circumvent these problems, methods to engineer the binding specificity of the ϕ C31 recombinase have been developed [66, 67]. For example, Scilimenti and colleagues have selected prototypes of ϕ C31 that have a significantly increased integration frequency (2 to 3-fold) and specificity (5-6-fold) for one particular pseudo-*attP* site located on chromosome 8. Although the ϕ C31 integration system seems to be more specific, it remains difficult to evaluate whether this argument can be considered decisive because ϕ C31 vectors also mediate a significant number of aberrant events leading to local chromosome rearrangements [68] and the number of genetic elements responsible for cancer is potentially underestimated [69].

1.2.2. Transposases

In contrast to integrations obtained by HR-endonuclease systems and site-specific recombinases, those mediated by transposition do not require any sequence similarity between the transgene donor and the integration target. This ability to move a discrete DNA fragment from one locus to another using a “copy and paste” or a “cut and paste” mechanism means that DNA transposons are considered good candidates to develop vectors for mutagenesis and gene delivery purposes in animals. As early as the 1980's, the first animal transposon tools were developed from the *Drosophila P* elements [70, 71]. However, this is only since 1997 that, with the molecular reconstruction of the fish transposon *Sleeping*

Beauty (*SB*), and the demonstration of its activity in human cells [72] as well as other related vertebrate models [73], DNA transposons were subjected to an increasing interest. Since this breakthrough, extensive investigations were performed to develop transposon tools for mutagenesis and gene delivery in mammals. Although *SB* has been the most investigated laboratory model so far, alternative models have been developed with certain eukaryotic and rare prokaryotic elements, all of them belonging to the *IS630-Tc1-mariner* (*ITm*), *hobo-Ac-Tam* (*hAT*) and *piggyBac* families (Fig. 1a). Elements of these 3 families have a “cut and paste” transposition (Fig. 1b) that is mediated by a transposase oligomer that recognizes two inverted terminal repeats (ITR) located at both transposon ends, cuts both strands at the outer transposon ends and inserts the transposon into a new target site that is duplicated upon insertion. Depending on the performances of these various transposons and the investments in their development, they are classified here as reference and alternative systems, or prospects.

1.2.2.1. Transposon Vectors - the Reference System:

A database gathering all the references of the scientific and technical documents published on this system is available and regularly updated at <http://www.beckmancenter.umn.edu/media/SBTS-Bib7-07.pdf>. *SB* is a consensus vertebrate *Tc1*-like element (*TLE*) reconstructed from 12 closely related but inactive *TLE* sequences isolated from seven different fish species [72]. Like every *ITm* element, the transposase (340 amino acids) consists of an ITR binding domain at the N-terminal and a catalytic domain containing a DD34E triad at the C-terminal. The hinge region linking the N- and the C-terminal domains of *TLE* transposases contains a functional bipartite nuclear localization signal (NLS). The original version of *SB* transposase, so-called SB10, mediates the mobility of an element comprised of two ITR of about 230 bp that flank an inner DNA segment containing the open reading frame (ORF) encoding the transposase surrounded by short 5' and 3' untranslated regions (UTR). Each ITR

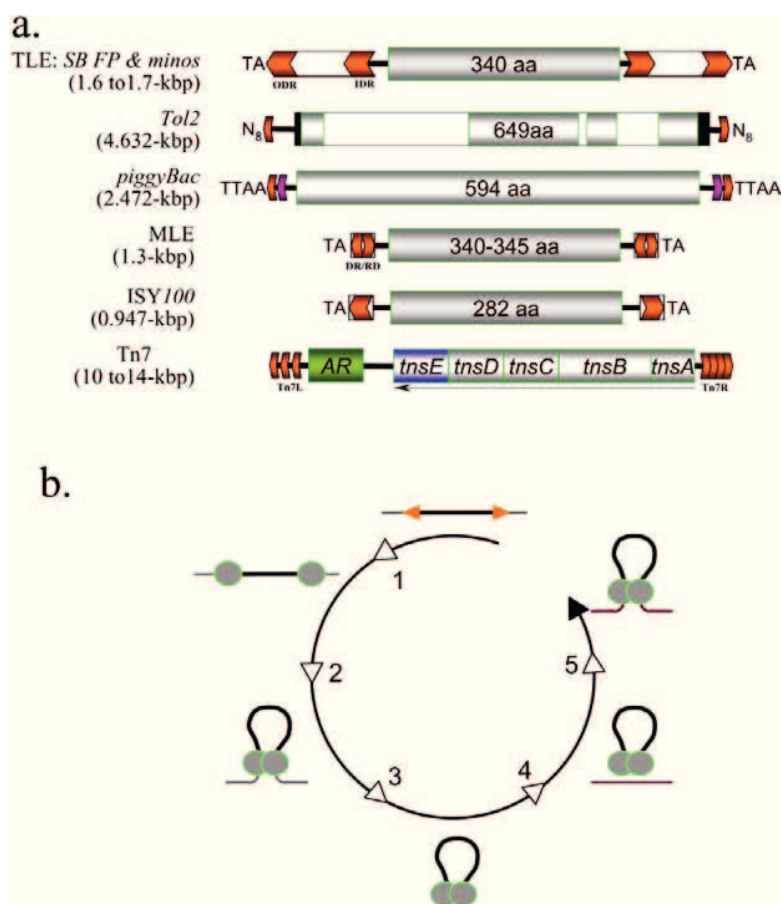


Fig. (1). **a)** Sequence organization of eukaryotic and prokaryotic transposons investigated so far to develop gene delivery vectors. Genes (gray boxes) with or without an intron (white boxes) are flanked by ITRs (white arrows) containing one or two transposase binding sites (DR: orange arrows). UTRs are indicated as black lines. Nucleotide motifs duplicated upon integration are indicated at both transposon ends. In TLE, the two transposase binding sites at both ITR ends are named outer and inner DR (ODR and IDR). In *piggyBac*, 13-bp inner inverted repeats are indicated with pink arrows. In Tn7, the green AR box locates the position where the antibiotic resistant genes carried by this element are located. **b)** Diagram representation of the five main steps required for the cut and paste mobility cycle of TLE, *Tol2*, *piggyBac*, MLE, ISY100 and Tn7. Step 1, binding of a transposase monomer or oligomer at both transposon ends; Step 2, assembly of the synaptic complex; Step 3, transposon excision; Step 4, integration target site binding; Step 5, integration. The transposon is represented with a thick black line and its ITRs with orange triangles. Transposase monomer or oligomers that bind to each ITR are represented with grey oval shapes. Grey and purple lines correspond to chromosomal DNA at the excision and integration sites, respectively.

contains two transposase binding sites repeated in a direct orientation (DR), and located at the outer (ODR) and the inner (IDR) ends (Fig. 1a). To specifically cleave the transposon at its ends, the transposase uses a cleavage signal sequence (5'-CAGT-3') located at the outer extremity of the ODR, where it spans 4 to 7-bp, depending on the element [for review see 74].

When used for gene delivery purposes, *SB* is a two-component gene vector system consisting of two plasmids, a transposon donor containing an expression cassette-of-interest cloned between the *SB* ITR and a transposase source corresponding to an expression vector. As the first versions of the *SB* vector system were only able to recover integration rates of about 5×10^{-4} in *ex vivo* cell assays, extensive efforts were made to optimize its efficiency. Several prototypes were generated, taking into account the ITR binding properties of the *SB* transposase [75], its similarities with other hyperactive *ITm* transposases [76], and the possibilities to humanize the *SB* ORF, in terms of codon usage [77]. The most active version, HSB17, has a transposition activity nearly 17-fold higher than that of *SB10*. The properties of the transposon DNA components were also investigated and solutions were found at two levels. First, *SB* vectors with two 5' ITRs were found to be 3-fold more efficient than the native version [78, 79]. Second, sandwich vectors consisting of two *SB* transposons flanking a large transgenic cassette (6 to 10-kpb) were developed to restore part of the transposition efficiency of vectors with a size over 2.5 kbp [76]. Indeed, it was demonstrated that, over 2.5 kbp, increasing the transgene size dramatically decreased the transposition efficiency of the *SB* vectors [73, 80]. Finally, combining these optimized solutions leads to a minimal increase of transposition efficiency of at least 33-fold in *ex vivo* integration assays. Depending on the cell type, this results in 2.5 to 17% of the transfected cells containing a *SB* vector integrated by transposition [81].

Like every *ITm* elements, *SB* has no real insertion site specificity, although it was demonstrated that its natural preferences are to insert at (TA)_n repeats located within a DNA segment with a elevated local bendability [82, 83, 84]. To circumvent this lack of integration specificity, and because random integration opens the risk of insertional mutagenesis of endogenous genes, two main strategies were developed to obtain a targeted *SB* integration system. The first has been investigated by several groups and uses *SB*-transposases fused at their N-terminal end with one DNA binding domain (DBD such as Gal4, or ZFD like E2C or ZNF202) to target integration (Fig. 6a). Although several DBD were fused using different peptide linkers to the *SB* transposase, very few fusion proteins kept activity for *ex vivo* transposition assays, and none of them was able to target integration into a specific integration site or fragment [85-87]. To obtain *SB* transposase targeting, Ivics and colleagues [87] have developed an alternative approach using a helper factor able to interact with both the *SB* transposase and a specific target chromosomal DNA site. This helper factor is a fusion protein made of the 57 amino acids of the N-terminal homo-oligomerization domain of the *SB* transposase fused to a DBD able to bind to a specific target chromosomal DNA site (Fig. 6b). In this way, they demonstrated that about 10% of the integrations occurred at the expected site.

A proof of principle was also provided for a second strategy investigated by the same research team [87]. Indeed, they showed that *SB* vectors could be targeted into a specific locus by targeting the vector donor rather than the transposase. Using as a co-factor, a fusion protein made with two different DBD (Fig. 6c), each of them being able to specifically bind to the transposon vector or a single chromosomal site, they demonstrated that at least part of the obtained vector integrations occur at the expected site. Beyond the engineering of the *SB* vector system, the principles developed by Ivics and colleagues is of a major interest because the targeting of the transgene donor at the proximity of the chromosomal site might also be a potential solution to improve the integration efficiency of the HR-endonuclease and recombinase systems and the integration specificity of any other transposon vector.

1.2.2.2. Transposon Vectors - Alternative Systems

Tol2 and *piggyBac* are the two main alternative systems to *SB*. Their theoretical interest is to bring solutions to circumvent three shortcomings of the *SB* system: i) a relatively modest cargo-capacity, ii) a decreased activity under high transposase concentrations (over-expression inhibition) and iii) the inability of *SB* transposases fused to DBD to target integrations. Although the data about the second shortcoming can be considered controversial or not fully elucidated for the moment [73, 88, 89], since they are still not confirmed by biochemical studies that unambiguously demonstrate that the decrease of activity is directly correlated with a dramatic transposase over concentration in transfected cells, the section below is mainly focused on the efficiency, the cargo-capacity and the integration specificity of *Tol2* and *piggyBac*.

Tol2

Tol2 is a naturally active and autonomous member of the *hAT* family that was characterized from the genome of a small teleost, the medaka fish *Oryzias latipes* [90]. It is 4.7-kbp in length, has ITRs of 17 and 19-bp, carries an internal gene, and is flanked by 8-bp target site duplication. *Tol2* does not seem to have any integration site preferences [84]. Its internal gene, consisting of four exons, encodes a transposase that catalyzes the transposition reaction of the *Tol2* elements (Fig. 1a). The structure of the *Tol2* transposase has not been determined, but that of a closely related element, *Hermes*, has been obtained [91]. *Hermes*-like transposases are three domains proteins. The N-terminal domain is likely the site-specific DNA binding domain for recognizing the ITR, the central region is a catalytic domain that contains a DDE triad, and the C-terminal is the dimerization domain [91-93]. Functional dissection of the *Tol2* transposable element has demonstrated that the design of *Tol2* vectors with efficient transposition also requires inclusion of regions inner to both ITRs, 200-bp from the 5' end and 150-bp from the 3' end [94].

By comparison with *SB*, the first quality of *Tol2* is to have excellent cargo-capacity, since it does not show significant differences in efficiency of gene delivery when the transgene carried by the vector varies from 2 to 11-kbp [94, 95]. Its second quality is that a *Tol2* kit is available to monitor assays for mutagenesis and gene delivery purposes [96].

However, the *Tol2* system, in its current version, only allows recovery of about 0.1 to 0.25% of the transfected cells with a vector integrated by transposition. Moreover, the few assays done for this purposes suggest that, similarly to the *SB* transposase, its fusion with a DBD leads to the construction of proteins inactive for the transposition [89].

piggyBac

The *piggyBac* transposable element, originally named IFP2, was isolated from the lepidopteran cell line TN-368 (TN = *Trichophusia ni*, the cabbage looper moth) as a gene-disrupting insertion within spontaneously arising baculovirus plaque morphology mutants [97]. Structurally, this is a 2472-bp element with 13-bp ITR and additional internal 19-bp inverted repeats flanking an ORF encoding a 594 amino acid transposase. For the moment, the structure of the *piggyBac* transposase is not determined, but sequence analyses revealed that it would also be a DD[E/D] transposase [56, 98, 99] whose modalities of the cut-and-paste transposition are not elucidated yet (Fig. □a). The analysis of a wide range of insertion sites has revealed that *piggyBac* vectors insert within a TTAA motif that is duplicated upon insertion. However, integrations do not fully integrate at random, as they have a significant preference for occurring in regions surrounding transcriptional start sites and within long terminal repeat elements [100].

Since the last decade, vectors derived from *piggyBac* have been successfully and extensively used to make transgenic insects. Recently, they were also found to be efficient for gene delivery purposes in mammal cells [101]. In a version close to that of the natural element, the efficiency of *piggyBac* vectors have been quite oversold [89]. Indeed, whether or not *piggyBac* vectors can be considered 2-fold more efficient than the SB11 version, data comparison indicates that it remains at least 15-fold less efficient than the HSB17 version [77]. Recent work on the optimization of *piggyBac* vectors have however increased the integration efficiency about 23-fold by humanizing the codon usage of the transposase [102]. Although no direct comparison has been published, it appears that the most optimized *SB* and *piggyBac* versions have very similar integration efficiencies.

In fact, the most interesting properties of the *piggyBac* system are its ability to carry a very large transgene (14-kbp) and the possibility to design transposases fused with one DBD that are active and able to target integration. Using *in vivo* transitory transposition assays in insect embryos, it was demonstrated that a DBD-*piggyBac* transposase had its integration preference dramatically modified. Moreover, such fusion proteins also have an integration efficiency that is about 10-fold increased compared to controls [103].

1.2.2.3. Transposon Vectors - Eukaryotic Systems Investigated as Prospects

Minos & Frog Prince

Minos and *Frog Prince* (FP) are transposable elements belonging to the same TLE subfamily as *SB*. *Minos* is a natural element that was characterized in the genome of *Drosophila hydei* (Fig. □a) [104]. Similarly to *SB*, FP was reconstructed from inactive TLE sequences isolated from the

Northern Leopard Frog, *Rana pipiens* [105]. Both systems have been shown to function in human cells, and even to be as active as *SB* [105, 106]. However, as their cargo properties and the solutions to modify their integration specificity are not significantly different of those of *SB*, their potential for gene therapy purposes have not been further investigated.

Mariner-Like Elements (MLE)

MLE, like TLE, consist of one of the *ITm* families. Depending on their sequence similarities, they are classified in five subfamilies named: *irritans*, *cecropia*, *elegans/briggsae*, *capitata/melifera* and *mauritiana* [107]. Their DNA sequence organization is close to those of other *ITm* elements, their main characteristics being to have short ITRs (20 to 40-bp) flanking an intronless ORF encoding a transposase (Fig. □a). The MLE signatures in the amino acid sequence of the transposase are the presence of two conserved motifs, WVPHEL and YSPDLAP, a DD34D catalytic triad, and a CRO/*singed*-like ITR binding domain that confers properties in the assembly of the transposition complex that are different to those of TLE [74]. MLE are potentially excellent candidates to develop gene delivery vectors. Indeed, they remain the transposons for which the amount of fundamental knowledge is probably the most important, and the only ones for which a complete set of biochemical, genetic, *ex vivo* and *in vivo* assays is available to develop any process of molecular engineering.

To date, the properties of three MLE have been investigated in mammalian cells. The two first, *Himar1* and *Hsmar1*, are elements belonging to the *irritans* and *cecropia* subfamilies. They were reconstructed and optimized from inactive MLE sequences isolated from the fly *Haematobia irritans* and the human genome, respectively [108, 109]. Both were assayed in HeLa cells and they were evaluated as being as efficient as the SB10 version [2, 109, 110]. The third element is a member of the *mauritiana* family, *Mos1*. In contrast to the extremely high activity of *Mos1* in the worm *Caenorhabditis elegans* (in which transposition rates are 1 to 10 events / cell division in somatic cells and 5×10^{-1} in the germ cells [111]), a very low integration activity is found in mammalian cells [89, 112]. This indicates that certain host factors negatively interfere with the MLE activity and prevent the release of their full potential for gene delivery purposes.

There is little chance for *Hsmar1*-derived vectors to be developed for gene therapy purposes, because the *Hsmar1* transposase causes chromosomal rearrangements by mobilizing the numerous *Hsmar1* ITR sequences in the human genome [113]. However, *Himar1* and *Mos1*-derived vectors have properties that remain widely unexplored. First, it should be possible to engineer the integration efficiency of these systems by circumventing the inhibiting effects of the host factors in an attempt to recover an activity close to that obtained in *C. elegans*. Second, they are the only eukaryotic elements, together with *piggyBac*, for which the transposase keeps its integration efficiency when it is fused at the N-terminal end, with most protein domains [114]. Like *piggyBac*, MLE transposases fused to some DBD have an increased integration efficiency and were demonstrated to be able to mediate integration within DNA fragment specifically recognized by the DBD [103].

1.2.2.4. Transposon Vectors - Prokaryotic Systems Investigated as Prospects

ISY100

ISY100 is a member of the *ITm* superfamily, belonging to the bacterial IS630 family, that was characterized in the genome of cyanobacterium *Synechocystis* sp. Overall, its sequence organization, the transposase's features and landmarks of mobility are very close to those of a *TLE* (Fig. 1a) [115]. Although its activity has never been assayed on mammal cells, the crucial interest is that a ISY100 transposase fused at its N-terminal end with a pZFD keeps its full activity, and has an almost specific integration activity modified by the presence of a functional pZFD [116].

Tn7

The bacterial transposon Tn7 is the most sophisticated element investigated so far for gene delivery purposes (Fig. 1a). Interest comes from its ability to specifically insert within a single 35-bp site in the *Escherichia coli* chromosome, called *attTn7*. This integration target is located 25-bp downstream of the coding sequence of an essential gene encoding the glucosamine synthetase (*glmS*). To keep integration specificity, Tn7 requires four proteins, encoded by the internal element. TnsA and TnsB together constitute the transposase and carry out the chemistry of transposition. TnD, the target selector, binds in a sequence-specific manner to *attTn7* and recruits TnsC, an ATP-dependent non-sequence-specific DNA-binding protein that stringently regulates the TnsD specific target binding.

The integration properties of Tn7 have already been used for some biotechnological applications [117, 118]. Although it requires four proteins for transposition, the Tn7 system remains interesting for gene therapy purposes because of the occurrence of *attTn7* sites downstream eukaryotic genes. Indeed, fructose-6-phosphate transaminases (*gfp*) are the homologs of *glmS* in eukaryotes, including human (Fig. 6d) [119]. These *attTn7* sites have a sequence conserved enough to be specifically bound by TnsD. *In vitro* human *attTn7* sites allow specific integration with a frequency of transposition comparable to the *E. coli glmS* target. However, transposition assays done in yeast or in the presence of human nucleosome extracts revealed that Tn7 transposition is reduced, likely impaired by factors in eukaryotic cells [120].

1.2.3. α - β - γ - δ - ϵ - ζ - η - θ - ι - κ - λ - μ - ν - ξ - $\omicron$ - π - ρ - σ - τ - υ - ϕ - χ - ψ - ω

To be assayed *ex vivo*, recombinase and transposase systems involve two components, a source of enzyme (that can be an expression plasmid or messenger RNA), and a donor plasmid carrying the transgene of interest flanked by nucleic acid motifs that are specifically bound by the enzyme in order to promote the specific recombination of the transgene within a chromosomal locus.

2. α - β - γ - δ - ϵ - ζ - η - θ - ι - κ - λ - μ - ν - ξ - $\omicron$ - π - ρ - σ - τ - υ - ϕ - χ - ψ - ω

In addition to highly specific genetic modifications or DNA integrations in chromosomes, several other side-effects

of the use of non-viral enzymatic systems need to be considered for safety issues. Here, only genetic consequences are taken in consideration. Due to the lack of data for some of the systems reviewed in the previous section, side-effects will be only reviewed for pZF-nucleases, meganucleases, ϕ C31 and *SB*, i.e. the most studied systems. Consequently, most of the issues raised in this section need to be verified in the other non-viral systems to allow accurate comparisons of their properties with the reference systems.

2.1. α - β - γ - δ - ϵ - ζ - η - θ - ι - κ - λ - μ - ν - ξ - $\omicron$ - π - ρ - σ - τ - υ - ϕ - χ - ψ - ω

Endonucleases, recombinases and transposases are enzymes that have DNA cleavage activities mostly controlled by their modes of specific binding to the DNA. In general, these enzymes are specific when they are bound to their accurate DNA binding sites as organized oligomers. However, when these enzymes are expressed in cells for middle or long-term durations, when they are not or are badly oligomerized, when their concentration over saturates the capacity of the cell to repair the DNA damage, their presence in the nucleus can have genotoxic effects related to non-specific single- and double-stranded DNA break activities.

Methods for evaluating the genotoxicity of DSB in chromosomes have been extensively developed for the analysis of various chemical and physical stresses such as ionizing radiation [121, 122]. The most commonly used method consists of following the presence of DSB *in situ*, through the presence of some histone H2AX forms. These assays take advantage of the fact that the H2AX molecules flanking DSB in the chromatin are rapidly phosphorylated at serine residue 139 (4 residues from the C-terminus), yielding a specifically modified form named γ -H2AX. Antibodies against γ -H2AX allow detection of the accumulation of γ -H2AX at the DSB sites, so-called γ -foci. Various techniques of immunochemistry or Fluorescence-Activated Cell Sorting (FACS) can then be used to quantify the number and/or the intensity of the γ -foci in cultured cells or tissues. Whatever the biological material, a basal level of γ -foci can be determined, since cells are permanently subject to DSB, then compared to the γ -foci level observed under various stresses. Moreover, a threshold from which the cell is not able to repair DNA, and in consequence dies, can be correlated with γ -foci amounts. Single-stranded DNA breaks (SSB) resulting from non-specific cleavage activities can also have deleterious effects on the cell viability. At present, there is no immunochemical assay available with properties similar to those of the γ -H2AX assay that allows quantifying SSB *in vivo* using the detection of protein involved in SSB signaling, such as the poly(ADP-ribose) polymerase-1 (PARP) [123]. Other methods for evaluating SSB amounts in cell populations have been developed [124, 125], but they have not been used so far to investigate the effects of the endonucleases, recombinases and transposases in the context of gene delivery. Finally, all the available methods allow overall quantification of SSB and/or DSB amounts randomly accumulated in the genome. Whether this DNA damage does not occur at random, but only in a limited population of cleavage sites, these methods likely comprise a bias that should lead to under-estimate the effects of SSB and DSB.

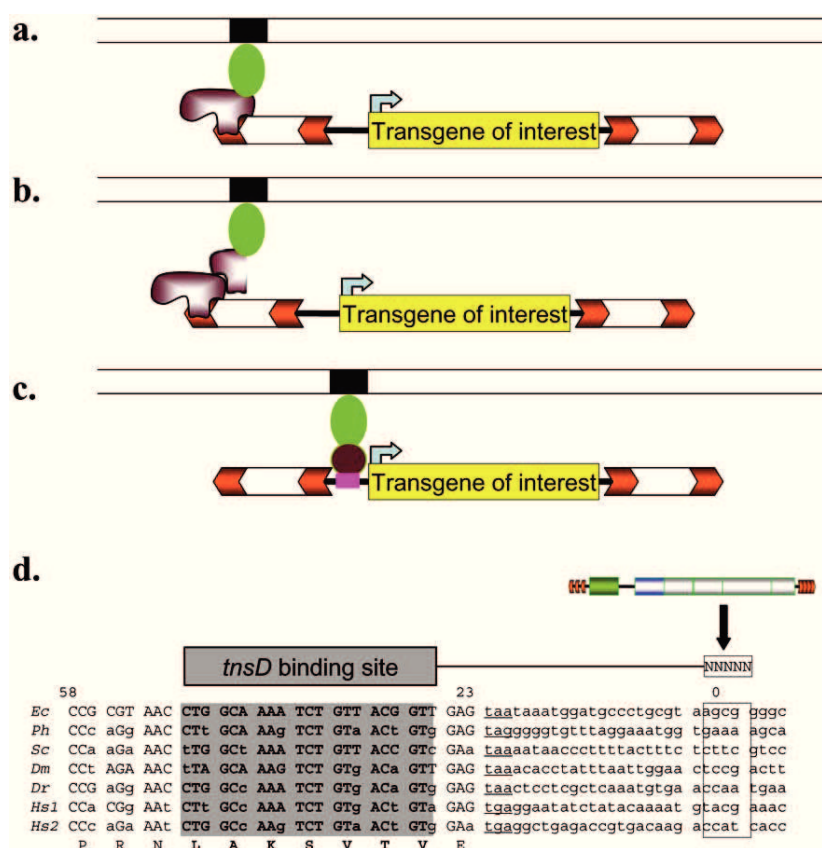


Fig. (6). Transposon vectors and insertion specificity. Two kinds of strategies were developed with *SB* to target integration into chromosomal DNA: to target the transposase (a and b) or transposon donor at the integration site (c). The components of the targeting systems comprise a transposable element consisted of two ITR flanking a gene-of-interest with a promoter for expression (yellow box), a transposase (purple shape) that binds to the ITRs and catalyzes transposition, and a DNA-binding protein domain (green oval) that recognizes a specific sequence (black box) in the target DNA (parallel lines). In b) and c), the targeting of the transposase is achieved by fusing a specific DNA-binding protein domain to the transposase or by using a protein co-factor mode with a DNA binding protein fused with a protein domain able to oligomerize with the transposase. For *SB*, the transposase domain that allows transposase-transposase interactions was used. In c), the transposase is not modified but a co-factor protein containing two DBD is added in the system. Targeting is achieved by fusing a specific DBD to another protein (dark purple oval) that binds to a specific DNA sequence within the transposable element (pink box). d) Sequence comparison of the *attTn7* sites located at the 3' end of the *glmS* ORF in the *Escherichia coli* (Ec) genome, and ortholog proteins in the genomes of the archaebacteria, *Pyrrococcus horikoshii* (Ph), the yeast *Saccharomyces cerevisiae* (Sc), the fly *Drosophila melanogaster* (Dm), the zebrafish, *Danio rerio* (Dr) and both genes in the human genome, *Hs1* and *Hs2*. These 63-bp regions both contain the DNA binding site of *tnsD* (nucleotides in bolded letters boxed in grey) and the Tn7 integration site in which the five nucleotides duplicated upon integration are boxed. The organisation of the site and the site in which Tn7 integrates are diagrammatized above sequence alignment. Nucleotides conserved in the *E. coli* ORF are in upper letters. The translation of the C-terminal region of the *glmS* ORF and its orthologs is indicated below the sequence alignments, stop codons are underlined.

2.1.1. Endonucleases Genotoxicity

The genotoxicity of the pZF-nuclease and meganuclease is a well-identified challenge [40]. Depending on the dose of endonuclease used and although the I-SceI meganuclease was found to be less toxic than pZF-nuclease, cell death and apoptosis likely resulting from elevated DSB and SSB rates are commonly observed [126-128]. To date, solutions to circumvent this point were only engineered for pZF-nuclease, and consist of improving the DNA binding specificity and/or the protein dimerization [129, 130].

2.1.2. Recombinase and Transposase Genotoxicity

Despite the fact that biochemists and molecular biologists working on these proteins know that their production is in

general difficult due to their toxicity for bacteria, the genotoxicity in mammal cells is a question that is rarely addressed in the literature. As a matter of fact, this question is also a neglected point for retro-vectors, although it has been demonstrated that the expression of the integrase of some endogenous retro-elements creates DSB in human cells [131]. Solutions have been developed to reduce the exposure time of the cells to recombinases and transposases after cell transfection. For the moment, they all propose to transfect the messenger RNA (mRNA) encoding the enzyme to the cells (instead of a DNA plasmid), *ex vivo* or *in vivo* [132-134]. Such solutions were found to limit the persistence of the enzyme source and in some case, to improve the efficiency rate of vector integration. Although a similar strategy should be applicable to endonucleases, this solution has not

been used to reduce their cytotoxic effect. Although some recombinases and transposases are difficult to produce fully active in mass, only one assay was so far reported in the literature about the possibility to directly transfect transposases, such as those of *Himar1* or *Mos1*, by the way of a penetrating peptide fused at their N-terminal domain to cross through the cellular membrane [135].

2.2.

Two kinds of undesired integrations can occur when a cell population is transfected with plasmids for gene correction or integration, mediated by any of the non-viral recombination systems evoked here. The first results from random integrations due to the overload of transfected DNA. The second is related to the ability of the endonucleases, recombinases and transposases to create non-specific SSB and DSB in the chromosomal DNA. Although these events occur two to three orders of magnitude less commonly than accurate cleavages, their occurrences are opportunities to stimulate incorrect recombination events. In both cases, this leads to non-specific single or tandem integrations of the transfected DNA molecules. At the scale of one cell, the occurrence of these events can be considered as negligible. However, when populations of 10^5 to 10^7 cells are transfected, these events become significant and the probability of promoting the emergence of proliferative clonal cancerous cells cannot be neglected. Since the detection of the undesired recombination events widely depends on the makers and the molecular tools, it can be considered that their presence remains under investigated in the genetically modified cells, whatever the non-viral gene delivery system used.

2.2.1. Endonucleases and Non-Specific Recombination

Two kinds of undesired side-effects resulting from inappropriate recombination can occur. In the literature, their occurrence was mainly illustrated for meganuclease systems, whereas their occurrence with pZF-nuclease remain largely ignored. The first relies on non-specific cleavage activities of endonucleases that increase SSB and DSB and in consequence increase the rate of non-specific integrations and local chromosomal rearrangements [136]. In addition, some recombination events result in tandem integrations of several copies of the plasmids. Although the integration of tandem repeats occurs at a locus without mutagenic consequence, the integration of concatemers often leads to their silencing by heterochromatin formation, and in consequence to the absence of expression of the integrated transgenes [137, 138]. Solutions to circumvent, or at least to decrease, the rate of non-specific integration events in part concern the engineering of the endonuclease specificity that was discussed previously. A complementary approach is to define optimal transfection conditions that allow reduction of the quantity of transfected DNA in cells, both by replacing the enzyme source by mRNA or purified proteins, and by decreasing the amount of plasmid donor of transgene.

The second kind of undesired side-effects occurs at the target site and corresponds to additional rearrangement on one side of the break. At these "one-side" events, the targeted insertions display one homologous junction and one non-homologous one. The occurrence of such events (a few

percent) appears to be frequent enough to seriously impair the use of endonuclease systems to integrate DNA fragments [40].

2.2.2. Recombinase and Non-Specific Recombination

In addition to simple integration events that occur at pseudo-*attP* sites, the ϕ C31 integrase appears to carry out aberrant events with a frequency of about 10% of genetically modified cells made *ex vivo*. Aberrant integration events are not a specificity of the ϕ C31 integrase since similar observations were made with another serine recombinase, Rec encoded by AAVs [139]. These events seem to take place at pseudo-*attP* sites and include intra-chromosomal deletions and apparent inter-chromosomal rearrangements [64, 68]. For the moment, the use of the ϕ C31 integrase in mouse models has not been associated with adverse events in the tissues in which it has been applied, up to 90-days post transfection [62, 140]. However, it must be noticed that the detection of cancer risks in animals require prolonged observation and appropriate genetic backgrounds for the mice to allow the detection of emerging tumours.

2.2.3. Transposase and Non-Specific Recombination

No case of chromosomal rearrangement has been described so far when transposon vectors were used for gene delivery purposes in mammalian cells. However, undesired side-effects representing from 10 to 85% of the integration events were observed, depending on the cell lineage, when using the SB10 system [73, 141]. These non-specific events were experimentally detected with genetic assays involving the HSV-TK suicide system [142] or LoxP/Cre recombinase system [143], which allow selection of the cells with integrations resulting from transposition, and eliminating those resulting from non-specific recombination events or maintenance of plasmid DNA in an extra-chromosomal form. Molecular analysis has confirmed that these events correspond to transposase-dependent non-specific recombination, thus integrating in tandem or not the plasmid donor containing the transposon vector. Similarly to endonucleases, these non-specific integrations are likely stimulated by the endonuclease activities of the transposase. Such observations were not restricted to *ex vivo* assays, and similar data were recovered when the properties of the SB10 system was investigated in mice [143].

Beside the fact that these non-specific integrations strongly impair the possibility of developing transposon vector systems with a controlled ability to integrate the transgene within a specific locus, they also raise serious problems with verifying the genomic quality of the genetically modified cells. In the absence of a counter-selection and controls done by Southern blot/hybridization with accurate genomic DNA digests, PCR controls done for verifying the transgene integration by transposition therefore obscure the non-specific integrations and mask the real content of the genetically modified cells. Among studies available in the literature, a nice example of the shifts encountered between PCR and Southern blot/hybridization data is presented in the Fig. 1 of Huang and colleagues [144].

In the last seven years, important steps have been developed to improve the integration efficiency of the transposon

after integration, it is not possible, to date, to define whether this efficiency results from an increased ability to transpose or from targeting of loci where the transgene is less efficiently silenced by the host machinery or both. In a similar fashion, this suggests that the differences of integration efficiency observed between transposon systems [89] may in fact result from their propensity to promote transposition in loci more or less under the control of the host silencing machinery rather than variation of “true” transposition activities.

One approach to circumvent the silencing of the integrated transgene is to include within the expression cassettes, matrix attachment region (MAR) elements to enhance the marker gene expression [155]. Because the study by Garrison and colleagues has only recently been published, the efficiency of this solution has not yet been investigated in the other non-viral systems.

2. ac g a i ic i

Although frequently silenced after integration, some of the integrated vectors can interfere with their close genomic environment. To date, this has only been investigated for *SB* vectors, and it was found that both *SB* ITRs retain a weak promoter activity, in both orientations [156]. In addition, the inner *SB* 5' and 3' regions that are juxtaposed to the ITR, the UTR sequences, are able to enhance gene expression in both orientations. The fact that *SB* ITR can promote transcriptional activity has two kinds of undesired side-effects. First, the expression from the *SB* ITR located at the 3' end of the vector is able to have a negative effect on the expression of the transgene, for example by RNA interference. Second, the ITR promoter activity may also interfere with the expression of the host sequences located in the close neighbourhood of the integration locus, and in consequence, to modify their expression. Finally, the impact of the expression of the marker cassette carried by the transposon vector was also investigated and it was demonstrated that it interferes with its close genomic environment [157].

To circumvent these problems, two solutions have been demonstrated to be effective with *SB* vectors in cell culture. First, the impact of the *SB* vector on the expression of its close genomic environment can be partly reduced by completely removing the UTR parts of the vector constructs [156]. Second, the expression cassette carried by transposon can be controlled by incorporating minute insulator elements, such as HS4, at both ends of the marker gene cassette [157].

3.1. $\square \square \square \square i \square \square \square i \square \square \square i$

The main challenge for the future of the non-viral integration systems in gene therapy begins by clarifying their real possibilities, taking into account their qualities and limits, but also the requirements related to each gene therapy strategy. As a consequence, the system chosen will depend on the nature of the genetic modification that it is required to perform (gene correction *versus* gene addition by integra-

tion), the status of the modified cells, their own specificity and innocuity.

When the aim is to correct (*in vivo* or *ex vivo*) a point mutation at the site of the error within the chromosomal DNA, systems involving stimulated HR appear to be the most appropriate candidates. Among non-viral strategies, they correspond to those using endonucleases, but also to non-viral alternatives ensuring DNA repairs by HR stimulated with either triplex-forming oligonucleotides, RNA-DNA chimeric oligodeoxynucleotides, single-stranded oligodeoxynucleotides or short fragment homologous replacement (Table 1) [158]. For the moment, only short fragment homologous replacements have been found to perform corrections in certain genes with efficiencies (up to 20%) similar to those obtained with HR systems stimulated by endonucleases (Table 2). This efficiency appears to be limited by the transfection systems, and efficiencies up to 50% have been reported with chimeric systems associating a HR system with an integrase-defective lentiviral vector delivery system [5]. When the objective is to supply a transgene by integration, endonuclease-stimulated HR also appears to be the solution with the most appropriate potential, even for transgenes with a size up to 8-kbp. Despite their lower efficiency, the integration specificity of these systems is a major advantage for performing genetic modifications both *in vivo* and *ex vivo* for gene therapy purposes. Moreover, it can reasonably be considered that the size of the integrated fragments will be able to be increased with the optimization of the performances of these systems. However there are limits. First, endonuclease-stimulated HR can only be used on cells and tissues that are cycling, thus excluding possibilities of *in vivo* treatments in rarely dividing and non-dividing cells, since they require an active DNA replication to be efficient. Alternatively, recombinase and transposase systems might supply solutions allowing integration in non-dividing cells [56].

The second limit that is currently under investigation for protein optimization concerns the non-specific nuclease activities of the endonucleases (Table 2). These points require more extensive investigation, both by precisely defining all the features of their genotoxicity and by determining optimized conditions to supply the endonuclease to the cell (mRNA transfection or protein translocation rather than transfection of a plasmid expressing the endonuclease). As a matter of fact, it must be pointed out that to date, in spite of the available background (for example, on the effects of ionizing radiations), there is no consensus agreement to set rules allowing the development of a pipeline for evaluating the full genotoxicity thresholds of the endonuclease systems, but also of those involving recombinases and transposases (Table 2). Similarly, our mining of the information available on the web has revealed that such a pipeline does not exist for the impact of the integrase used by the retrovectors, although it is demonstrated that the expression of certain endogenous retro-element integrases creates DSB in human cells [131], and putatively SSB and chromosome rearrangements.

3.2.

To date, the definition of the requirements to target the chromosomal integration of any vectors is mainly driven by

the need to avoid the deleterious consequences of random integrations in a genetically modified cell population of several million cells. In this perspective, although integration systems such as those of ϕ C31 and *SB* were shown to be better at avoiding integration into genes compared to retrovectors, they remain able to integrate into genes, thereby posing the potential danger of accidental activation or inactivation of an endogenous gene [62, 159]. In addition, the recent update of the definition of what are the physical boundaries of a gene, [160] questions the meaning of these integration patterns in supporting integration innocuity. This therefore raises a new question concerning the definition of what might be a good therapeutic integration site for gene delivery purposes. This issue is also raised for purposes other than circumvention of the putative deleterious effects of vector integration. Indeed, recent studies by Garrison and colleagues [153] have highlighted the need to have site specific vectors able to circumvent the extinction of the transgene expression upon integration by the host insulation machinery, and thus to reach unaltered transgene expression over time.

In this attempt, two strategies have so far been investigated. The first is to develop systems able to integrate DNA in repeated regions that are involved in the binding to the nuclear matrix, so-called MARs (Matrix Attachment Regions). Such a targeting was successfully achieved with *SB* vectors using a chimeric protein interacting with both the DNA vector and MARs (Fig. 6c) [87]. However, it is not clear whether integrations within or at the close proximity of such a sequence guarantee the transgene expression in all cases, and whether they alter or not the functionality of the MAR targets. The second strategy also concerns the use of repeated sequences as integration targets. These sequences are those encoding the ribosomal RNA (rRNA) 18S, 5.8S and 28S that are essential for the survival of the cells. In eukaryotic genomes, these genes are tandemly repeated and clustered in one or a few loci. The number of gene copies per haploid genome varies from 55 in the worm *C. elegans* to several thousands in the frog *Xenopus laevis*. The human genome contains approximate 400 copies of rRNA genes per haploid genome, clustered on the short arms of the chromosomes 13, 14, 15, 21 and 22. The proofs of principle demonstrating that these sequences are good integration targets (warranting innocuity and sustaining gene expression, whatever the promoter used for the transgene expression) were obtained by two kinds of experiment, in fields where the scientific backgrounds and objectives are quite different to those encountered in gene therapy. The first data set came from research groups working on the engineering of yeast as factory cells to produce recombinant proteins. They have demonstrated that the disruption of one or a limited number of the rRNA genes with vectors integrated by HR has no impact on the cell viability [161, 162]. They also showed that transgene integrations within rRNA genes allow permanent and elevated rates of expression, without any requirement of MAR sequences in the vector [163]. One similar evidence was also obtained in other unicellular eukaryotes such as protozoons [164]. The second data set were provided by research groups working on DNA transposons [165], and retro-transposons [166] that have strategies of specific integration into genomic niches containing repeated genes such

Table 1. Classification of non-viral enzymatic systems according to the type of integration and the type of vector used

Integration type	Enzymatic system	Vector type
Correction of point mutations at specific loci	pZFD-nuclease Meganuclease	Triplex-forming oligonucleotides RNA-DNA chimeric oligodeoxynucleotides Single-stranded oligodeoxynucleotides Short fragment homologous replacement
Sequence integration using a site-specific integrating vector	pZFD-nuclease Meganuclease	None
Sequence integration using an integrating vector	pZFD-nuclease Meganuclease ϕ C31 recombinase SB Tol2 PiggyBac MLE	Lentivirus vectors

Table 2. Comparison of the properties and undesired side-effects of the integrative viral vector of reference and the non-viral enzymatic integration system

Integration system	Properties			Undesired side-effects			
	Estimated integration efficiency	Maximal Cargo capacities	Integration specificities (Natural qualities / potential strategies to engineer specificity)	Non-specific DSB and SSB activities of the enzyme	Rate of non-specific DNA integration	Chromosomal rearrangements	Vector interferences with the genomic environment flanking IS
<i>Integrative viral vector of reference</i>							
Lentivirus [150]	99%	8-kbp*	Intragenic regions	NA	NA	NA	controllable
<i>Non-viral enzymatic integration system</i>							
pZFD-nuclease [37,38]	WS: 5 to 15%	≤ 8-kbp*	Site-specific	Elevated	NA	NA	NA
Meganuclease [40]	WS: 1 to 10%	DNA	Site-specific	weak	≤ 20%	NA	NA
ϕ C31 recombinase [62]	WS: 0.1 to 15%	≤ 50-kbp*	Degenerated site specificity	NA	NA	<10%	NA
SB [76,77,149]	SC: 2.5 to 17% WS: 55%	≤ 10-kbp	UD/TP	NA [#]	5 to 85 %	NA	controllable
Tol2 [94,95]	SC: 3% WS: ?	≤ 10-kbp*	UD/TP	NA [#]	NA [#]	NA	NA
piggyBac [89,102]	SC: 3% WS: ?	≤ 14-kbp *	UD/TP & Ftpase	NA [#]	NA [#]	NA	NA
MLE [110]	SC>1% WS: ?	≤ 5 to 13-kbp	UD/TP & Ftpase	NA [#]	NA [#]	NA	NA

SC, Under selective conditions; WS, without selection; (*) points out systems with a roughly constant integration efficiency whatever the transgene cassette size above the maximal limit; TP, targeted plasmid; Ftpase: Transposase fused with a DBD; UD, the natural specificity was notified as being undetermined since available data only from cell clones obtained from maker selections and do not account for integration biases revealed by Garrison and colleagues [149]; NA, not available; ([#]) locates the systems for which effects are suspected to be significant from biochemical knowledge of the enzyme; IS, Integration site.

as rRNA, tRNA, and snRNA, or repeats such as microsatellites, telomeric or sub-telomeric repeats [167]. Studies with these MGE have demonstrated that the disruption of few repeated genes does not alter the host viability, and that integration in genes such as rRNA guarantees innocuity and gene expression either by the RNA polymerases 1, 2 and 3.

Recent studies have started investigating rRNA genes as an integration target for gene therapy purposes. All of them were based on the use of vectors integrated by HR. Such an approach was for example used to specifically insert sequences in artificial chromosomes containing rRNA genes [168]. More interestingly, a research team in Medical Genetics of China (Changsha) directed by Professor Xia Jia-hui has developed a non-viral vector that allows the specific integration of exogenous DNA fragments into the human rRNA genes. This allowed them to demonstrate that a targeted, elevated and sustained expression of factor VIII and vascular endothelial growth factor (VEGF) can be obtained *ex vivo* in hepatocytes and endothelial cells [169-172]. Although these studies highlight the potential to use rRNA genes as integration targets, the fact that all to date have used vectors functioning by passive HR has decreased their interest because they only allow integration efficiencies ranging from 10^{-5} to 10^{-4} . Moreover, it is unlikely that HR systems stimulated with tailored endonucleases would be a valuable solution to increase the integration efficiency. Indeed, the coding sequence of the rRNA genes is perfectly conserved (100%) within the genome of each species. If a tailored endonuclease is designed to specifically cleave within rRNA genes, its activity will be very likely be fatal for the cell, since the enzyme would fragment most of the rRNA units by double-stranded DNA cleavage, thus destroying these loci which are essential for cell viability. From this standpoint, the use of transposon vectors targeted for integration into RNA genes, using for example, protein domains able to specifically bind to these genes [173], seems to be a safer and more valuable solution.

□□□□ Intellectual Property (IP) Status of Non-Viral Enzymatic Integration Systems

To check the future technological perspectives of non-viral enzymatic systems, an important issue, which is going to affect the private investments and those of the state scientific and medical bodies, is the possibility to develop exploitation plans for each technical and therapeutic application. In fact, these plans widely depend on the IP status of each technology. Here, a brief overview of a study concerning IP in this field is presented, the complete work will be presented in another manuscript [174].

3.3.1. Methods Used to State the Art for IP

Our study was developed using the esp@cenet® facilities at http://ep.espacenet.com/advancedSearch?locale=en_EP, and its database release of April 2008, except for the Zinc Finger technologies for which the February 2007 release of the QPAT database (Questel) was used at <http://www.qpat.com/index.htm>. Results are summarized in the Tables □a and □ and were obtained using the advanced search option of esp@cenet®. Keyword combinations about non-viral enzymatic integration systems were sought inside the patent titles and the abstracts. Results supplied by esp@cenet® are orga-

nized in patent families (patent applications and publications), all the patents of a family having the same priorities (i.e. a French and a US patent concerning the same invention are gathered in a single patent family). Results in Tables □ present patent families that only concern non-viral enzymatic systems for gene delivery purposes. Our investigations were restricted by two main factors. First, we used a limited selection of specific keywords, and two databases. Second, the mining process has several inherent limits. Indeed, patent laws and patent office jurisdictions are different, according to the countries. Patent mining therefore depends on the way of writing for the title and the abstract (describing the content of the invention), the last one putatively having a scope that is more or less general or ambiguous. This problem was circumvented for some related patent families for which no word in the patent title and abstract matched with the keywords used (e.g. in WO2008027384). They were located by using inventor references found in other patents or literature. The esp@cenet® website makes important provisions regarding to the terms and conditions of use for esp@cenet®, especially about availability and liability. Indeed, the European Patent Office (EPO) does “not accept any responsibility for the accuracy of data and information originating from other authorities than the EPO, including but not exclusively whether they are complete, up-to-date and fit for specific purposes”. In consequence, this led us to remove from our study several patent families for which no access to claims was available. This was met, for example, with several non translated Chinese and Japanese patents, for which no Patent Cooperation Treaty extensions were available.

3.3.2. IP Status

In the last 15 years, the non-viral enzymatic integration systems detailed in this review have been patented (Table □a): endonucleases (36 patent families), recombinases (7) and transposases (37). The main claims of these patents are focused on the processes to engineer the properties of each non-viral enzymatic integration system and their potential uses. It is striking that there are few patents specifically dedicated to the modification of mammalian cells, except for *I-SceI*, *Tol2* and *Sleeping Beauty*. This highlights the fact that non-viral enzymatic integration systems are patented for a wide range of applications since these technologies have an interest that spans beyond gene therapy, in various biotechnology fields that require genetic modifications.

Our results revealed that no patent was founded for the meganucleases *I-CeuI* and *I-MsoI*, and for the transposase *ISY100*. However, it must be kept in mind that “an absence of evidence is not an evidence of absence”, as reminded at http://ep.espacenet.com/?locale=en_EP&view=seven. Except for these systems, each non-viral enzymatic integration system reviewed in this manuscript comprises at least one patent family, with more than one patent application (patents filed in many countries, hence a resulting monopoly), except for the *Himar1* transposase (two patent families of one patent applications). Most of the applicants and inventors in these patents are specialists with very significant backgrounds in the enzymatic integration systems and these authors of scientific publications are considered as the milestones of this field. Interestingly, it must be noticed that the investment of the state and private bodies in patenting these technologies

□a□le □a□ □imited Intellectual Property Status of the Non-Viral Enzymatic Integration Systems

	Non-viral enzymatic integration systems	□eywords used for patent searches	Patent □amilies relative to the proof of concept in use and in engineering	Patent families for modifying mammalian cells	Applicants (universities□companies□)
NUCLEASES MEGANUCLEASES	FokI and variants	fokI, site specific nuclease*	JP2006254921 ⁺ , US5487994, WO9418313 ⁺ , EP1340812 ⁺ , US5356802	/	Univ Johns Hopkins (US), Univ Ohio (US), Sangamo Biosciences (US)
	pZFD-FokI		WO9640882 ⁺ , WO2007014181 ⁺ , US5436150, WO2007139898 ⁺	/	
	Meganucleases and variants	meganuclease*	US2007117128+, WO2007057781+, WO2007034262+, US2006206949, US2006153826, US2006078552, WO2005105989+, CA2514517+, US2004002092, WO03078619+	/	Collectis (FR), Precision Biosciences (US)
	I-SceI	SceI	US7309605 ⁺ , JP2007014347 ⁺ , US2005032223 ⁺ , US2003182670 ⁺ , US2003113887 ⁺ , WO0046386 ⁺ , WO0046385 ⁺ , US6238924, US5948678 ⁺ , US5962327 ⁺	US6822137, US6566579 ⁺	Pasteur Institut (FR), Univ Paris Curie (FR), Childrens Medical Center (US), Centre Nat Rech Scient (FR)
	I-CreI	CreI	WO2008010093 ⁺ , EP1863909 ⁺ , WO2007093918 ⁺ , WO2007060495 ⁺ , WO2006097784 ⁺	/	Collectis (FR)
	I-CeuI	CeuI	/	/	/
	I-MsoI	MsoI	/	/	/
INTEGRASES	ΦC31 recombinase	phi C31, phiC31, integrase* C31, recombinase* C31, phage* C31	EP1205490 ⁺ , WO0107572 ⁺ , US2003186291, US2004003420, WO0238613 ⁺ , US2005054106 ⁺ , WO03066867 ⁺	/	Artemis Pharmaceuticals (DE ; now TaconicArtemis, US), Univ California (US), US Agriculture (US)
TRANSPOSASES	<i>Tol2</i>	tol2	US2006212959 ⁺ , US2006211116 ⁺	US2005177890 ⁺	Japan Science & Tech Corp (JP)
	<i>piggyBac</i>	piggybac	EP1896578 ⁺ , US7129083, US7105343, US7005296 ⁺ , US2002173634 ⁺	/	Univ Fudan (CN), Univ Notre Dame du Lac (US), US Agriculture (US)
	<i>Himar1</i>	himar1 mariner transpos*	US6368830, CA2319168	/	Harvard College (US)
	<i>Hsmar1</i>	hsmar1, mariner transpos*	EP1869193 ⁺ , WO2006108525 ⁺	/	Max Delbrueck Centrum (DE)
	<i>Mos1</i>	mos1, mariner transpos*	WO2007063033 ⁺ , US2007031967 ⁺ , WO2004078962 ⁺ , WO9909817 ⁺	/	Centre Nat Rech Scient (FR), Univ de Tours (FR), Univ du Maine (FR), Biotech & Biolg Scien Res (GB)

(a) contd

	Non-viral enzymatic integration systems	Keywords used for patent searches	Patent families relative to the proof of concept in use and in engineering	Patent families for modifying mammalian cells	Applicants (universities/companies)
	TLE	mariner transpos*	US2003158129, US6051430, WO9907871 ⁺ , WO9729202 ⁺ , CA2244812	/	Kanker Inst NL, Introgen BV (NL), Het NL Kanker (NL)
	<i>Sleeping Beauty</i>	sleeping beauty, SB transpos*	US2005003542, US2005112764 ⁺ , US2004092471, US2002103152 ⁺ , US2003124668 ⁺ , WO9840510 ⁺	US2006252140	Leland Stanford Junior Univ (US), Max Delbrueck Centrum (DE), Univ Minnesota (US)
	<i>Frog Prince</i>	frog prince	US2005241007 ⁺	/	Max Delbrueck Centrum (DE), Netherland Cancer Inst (NL)
	<i>Minos</i>	minos	WO0171019 ⁺ , WO0129205 ⁺ , EP0955364 ⁺ , US6159717 ⁺ , US5840865, US2002026650 ⁺	/	Minos Biosystems Ltd (GB), Inst Molecular Biology and BioI (GR)
	<i>Tn7</i>	tn7	WO9905295 ⁺ , US5958775	/	Univ Jefferson (US)
	<i>ISY100</i>	isy100	/	/	/

(+) represents a patent family (documents sharing the same priority or combination of priorities, about the same invention), (*) is the usual wildcard character for truncation in Internet (e.g. "transpos*" can be transposase(s), transposon(s), transpos(ition), ...). The opinion given in this article engage neither examiners nor patent offices.

(a) Limited Intellectual Property Status of Molecular Processes Required for Modifying their integration Specificity of Non-Viral Enzymatic Integration Systems

Engineering process	Keywords used for patent searches		Patent families relative to the proof of concept in use and in engineering	Patent families for modifying mammalian cells	Applicants (universities/companies)
pZFD	zinc finger	design	US2006292621 ⁺ , US2003068675 ⁺ , WO0242459 ⁺ , WO0183751 ⁺ , WO0160970 ⁺ , US2003104526, WO0042219 ⁺ , GB2360285 ⁺ , WO9942474 ⁺ , WO9948909 ⁺ , WO9853057 ⁺ , WO9853059 ⁺ , WO9632475 ⁺ , WO9606166 ⁺	/	Sangamo Biosciences Inc (US), Massachusetts Inst Technology (US), Scripps Research Inst (US), Toolgen InC (KR), Syngenta Participations AG (CH), Gendaq Ltd (GB, acquired by Sangamo), Medical Res Council (GB), Genome Dynamics Inc (US), Univ Washington (US)
		linker	US2003068675 ⁺ , WO9945132 ⁺ ,	/	
		ZFP production	US2005084885, WO200316571 ⁺ , WO02057293 ⁺ , WO0208286 ⁺ , WO0027878 ⁺ , WO9947656 ⁺	/	
Fusion protein with targeted integration preferences	transpos* target*		US2008008687 ⁺ , US2006236413 ⁺ , US2006258603 ⁺ , US2006210977, CA2533708 ⁺	US2006252140	Max Delbrueck Centrum (DE), Univ Vanderbilt (US) Leland Stanford Junior Univ (US)

(+) represents a patent family (documents sharing the same priority or combination of priorities, about the same invention), (*) is the usual wildcard character for truncation in Internet (e.g. "transpos*" can be transposase(s), transposon(s), transpos(ition), ...). The opinion given in this article engage neither examiners nor patent offices

are very different, state bodies playing a strategic role for the emergence of the technology, whatever the country (c.f. 3.3.3 section). The IP of the pZF-nuclease and the ϕ C31 systems, and those considered so far as being the most valuable transposon systems (*Sleeping Beauty* and *PiggyBac*), belong of US bodies. In this IP overview, European bodies are only in a favourable position for the meganuclease, *Hsmar1* and *Mos1* systems, whereas Japanese bodies appear to set their sights upon the *Tol2* transposon systems. Patent families concerning the *FP* and *Minos* systems also belong to European bodies but the robustness of their IP seem to be more fragile, since their claims might be considered as being already covered by anterior patent families concerning a closely related element, *SB*. Similarly, certain patent families concerning TLE and MLE (e.g. WO9909817), seem to protect crumbly IP, as they could involve more than one invention in each patent family (the unity of invention being that TLE and MLE are now considered two different entities), and/or because the definition of the patented systems can be considered as inaccurate, in the light of the knowledge accumulated in the last 10 years.

Among the technology that can be used to improve the properties of the enzymatic integration systems, the IP concerning the engineering of the ZFD gather about twenty patent families (Table □□) that span from the ZF design, the linkers to associate ZFD, the processes to select pZFD with elevated DNA binding efficiency and specificity, to the production methods of ZF proteins. Overall, these patent families lock all the required elements to be in position to monitor the ZF technology. The majority applicants are state and private US bodies, and comprise the Massachusetts Institute of Technology (US), the Scripps Research Institute (US), Toolgen (KR), Syngenta Participations (CH), Gendaq (GB, acquired by Sangamo), the Medical Research Council (GB), Genome Dynamics (US), the University of Washington (US), and Sangamo Biosciences (US). 2005 saw the development of a world network, the Zinc Finger Consortium, in order to “ensure and to promote continued research and development of engineered zinc finger technology” (<http://www.zincfingers.org/default2.htm>). Sangamo Biosciences, the “worldwide leader in the design and development of engineered zinc finger DNA-binding proteins (ZFPs) for gene regulation and gene modification” (<http://www.sangamo.com/about/about.html>) and members of the Zinc Finger Consortium hold key patents for the whole technology of based-Zinc Finger Domains targeting. Therefore, all the uses of ZF technology depend on their patent families (gene regulation, gene transfer etc.), thus restricting, or at least slowing down, the possibilities to extensively commercially exploit this technology for various gene delivery purposes, including gene therapy [175]. To some extent, such an observation can also be considered valuable for the meganuclease systems.

Regarding fusion proteins with targeted integration preferences (Table □□), several patent families entitled “Transposon-Based Targeting System”, “Development of a transposon system for site-specific DNA integration in mammalian cells”, and “Transposon-based vectors and methods of nucleic acid integration” (various patent families have the same name and modifications in their claims) respectively belong to the Max Delbrueck Centrum (DE), the Leland

Stanford Junior University (US), and the Vanderbilt University (US). The first and the second patent families claim targeting systems with a “transposase” or a “transposon”, fused to a “DNA targeting domain” or a “site-specific DNA binding protein”. The third last patent family (“Transposon-based vectors and methods of nucleic acid integration”) claims a “composition” with “an integrating enzyme”, which can be “selected from the group consisting of transposase, integrase, retrotransposase, recombinase, bacteriophage integrase, integrin, terminase or retroviral integrase”. That is the reason why the international search authority has found five claimed inventions for the patent application WO2004009792 (family of the patent CA2533708), one for each transposase, recombinase, integrase, integrin, and terminase-based compositions. Overall, some ambiguities and the excessively wide coverage of the claims included in these patent families (more than one protected invention) should allow most bodies to circumvent the IP constraints associated with these documents, or at least will oblige the different bodies to negotiate the use rights for these technologies.

3.3.3. Impact of IP on the Development of Non-Viral Enzymatic Integration Systems

Esteban Burrone (Consultant, SMEs Division) explains on the World Intellectual Property Organization (WIPO) website (http://www.wipo.int/sme/en/documents/patents_biotech.htm) that “the strong growth of the biotechnology industry in recent years has been mirrored by a higher than average growth rate for patent applications and patent grants that relate to biotechnology inventions”. He also stated that “the growth in the number of patents in the field of biotechnology is largely due to the importance that life sciences and biotechnology companies attach to intellectual property, particularly patents”. Often, such companies arise from scientists that spin off biotech companies with the technology they developed in university laboratories. This is the case of Manoa Biosciences “a new life sciences company specializing in non-viral horizontal gene insertion as it pertains to animal transgenesis and gene therapy” (<http://www.manoabiosciences.com/>). This company has “the exclusive rights to six transgenesis technologies licensed from the University of Hawaii” and “has secured exclusive assignment rights for the chimeric transposon technology from Vanderbilt University”. This biotech company is one of the companies specialized in gene transfer using non-viral vectors, like Discovery Genomics Inc. and TranspoVec Corporation.

As also indicated by Esteban Burrone, “Patents is at the core of the Biotech Business”. The invention protection, by a patent, and the patent itself are key elements of the strategy for biotech companies. For example, Collectis (Table □a, meganucleases), “a worldwide leader in the research, development and commercialization of rational genome engineering technologies” (<http://www.collectis.com/>), have sued its US competitor Precision Biosciences (Table □a), for patent infringements about meganucleases technology (<http://www.collectis.com/news.php>). “The lawsuit seeks both monetary damages for infringement, as well as a permanent injunction preventing Precision BioSciences (Table □a, meganucleases) from any further making, using or selling of such meganucleases”. In 2007, Collectis won “a patent battle in the field

of targeted recombination and gene correction" with the invalidation of a European Patent owned by Johns Hopkins University (Table 1a, nucleases) and licensed to Sangamo Biosciences (in spite of an appeal of both them). These patent fights highlight the problem of defining the limits of biotech patents and rights to exclude invention exploitations to third parties.

CONCLUDING REMARKS

Recent advances in engineering non-viral integration systems indicate that they are not only promising alternatives to viral integrative vectors, but that they can offer unique solutions to repair genes, and to target "safe" integrations of exogenous DNA vectors into chromosomes. Two important issues should have to be investigated before entering into clinical phases. The first is to define niches of therapeutic applications and whether these technologies must be limited to strategies involving an *ex vivo* step then a grafting to the patient of the genetically modified cells, or whether they present real decisive advantages on viral vectors allowing use *in vivo*. In fact, the main part of the answer will depend of the technical means that are available to specifically address these integration systems to cells or tissues. The second important issue is related to the toxicity of these systems that remains widely under explored over the medium to long term duration of use in current *in vitro* laboratory models and in the pre-clinical models.

An efficient and adapted exploitation of these technologies will therefore require making important investments. To date, their development can be considered less extensive than that of the viral integrative vectors, partly because of the background and the size of the community working in this field which is much smaller than the number of virologists involved in gene therapy issues. However, there are also endogenous causes in this community that limit the investment in technological developments for gene therapy purposes. Indeed, there is an abundant literature demonstrating that non-viral integration systems have an interest that goes beyond gene therapy, in the various biotechnology fields that involve genetic modifications. This is due to the fact that these systems work in a wide host range, on potential applications that are easier to reach, and less exacting to develop in term of safety than those concerning gene therapy. In consequence, whereas the community working on the development of most viral vectors is "mainly" focused on human health issues, that working on non-viral integration systems can be considered less involved on gene therapy issues. Another inherent limiting factor finds its origin within the IP issues that can discourage and hinder new research and development bodies to invest and to reinforce the development of these technologies, since they will not be in position to recover any benefits of their investments. Indeed, considering what are so far the most promising solutions of non-viral integration technologies, they are heavily patented by a few bodies and, in consequence, locked. From a non-US point of view, in the position of a state or an organisation such as the European Commission, there will be a reluctance to stimulate investment in programs to develop these technologies when no valuable exploitation plan can be contemplated at the economic level.

In conclusion, there is much to be done to establish non-viral integration systems as an efficient and safe alternative to the use of integrating viral vectors. Such developments could be constrained by intellectual property issues and the ability of various state and private bodies to resolve these issues will be critical in the timely development of these technologies for treatment of human diseases.

ACKNOWLEDGEMENTS

This contribution was sustained by the University of François Rabelais of Tours and funded by grants from the European Commission (Project SyntheGeneDelivery, N°018716), the C.N.R.S., the Ministère de l'Education Nationale, de la Recherche et de la Technologie, the Association Française contre la Myopathie, and the Groupement de Recherche CNRS 2157. Fabien Palazzoli holds a doctoral fellowship from the Ministère de l'Education Nationale, de la Recherche et de la Technologie.

ABBREVIATIONS

AAV	=	Adeno-associated virus
DBD	=	DNA binding domain
DNA	=	Deoxyribonucleic acid
DR	=	Direct repeat
DSB	=	Double-strand break
EBV	=	Epstein-Barr virus
EPO	=	European Patent Office
ES cell	=	Embryonic stem cell
FACS	=	Fluorescence-Activated Cell Sorter
FP	=	<i>Frog Prince</i>
<i>hAT</i>	=	<i>Hobo-Ac-Tam</i>
glmS	=	Glucosamine synthetase
gfpt	=	Fructose-6-phosphate transaminases
HBV	=	Hepatis B virus
HR	=	Homologous recombination
IDR	=	Inner direct repeat
IP	=	Intellectual property
<i>ITm</i>	=	<i>IS630-Tc1-mariner</i>
ITR	=	Inverted terminal repeat
MAR	=	Matrix attachment region
MGE	=	Mobile genetic element
<i>MLE</i>	=	<i>Mariner</i> -like element
NHEJ	=	Non-homologous end-joining
NLS	=	Nuclear localization signal
ODR	=	Outer direct repeat
ORF	=	Open reading frame
PEI	=	Polyethylimide
PGE	=	Parasitic genetic element

pZFD = Polydactyl zinc finger domain
 SB = *Sleeping Beauty*
 RNA = Ribonucleic acid
 SSB = Single strand break
 SSDA = Synthesis-dependent strand annealing
 TLE = Tc1-like element
 UTR = Untranslated region
 WIPO = World Intellectual Property Organization.

REFERENCES

- [1] Edelstein ML, Abedi MR, Wixon J. Gene therapy clinical trials worldwide to 2007 - an update. *J Gene Med* 2007; 9: 8333-42.
- [2] Zhang L, Sankar U, Lampe DJ, Robertson HM, Graham FL. The Himar1 mariner transposase cloned in a recombinant adenovirus vector is functional in mammalian cells. *Nucleic Acids Res* 1998; 26: 3687-93.
- [3] Yant SR, Ehrhardt A, Mikkelsen JG, Meuse L, Pham T, Kay MA. Transposition from a gutless adeno-transposon vector stabilizes transgene expression *in vivo*. *Nat Biotechnol* 2002; 20: 999-1005.
- [4] Miller DG, Wang PR, Petek LM, Hirata RK, Sands MS, Russell DW. Gene targeting *in vivo* by adeno-associated virus vectors. *Nat Biotechnol* 2006; 24: 1022-26.
- [5] Lombardo A, Genovese P, Beausejour, *et al.* Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol* 2007; 25: 1298-306.
- [6] Schnepf BC, Jensen RL, Chen CL, Johnson PR, Clark KR. Characterization of adeno-associated virus genomes isolated from human tissues. *J Virol* 2005; 79: 14793-803.
- [7] Vasileva A, Jessberger R. Precise hit: adeno-associated virus in gene targeting. *Nat Rev Microbiol* 2005; 3: 837-347.
- [8] Kanda T, Kamiya M, Maruo S, Iwakiri D, Takada K. Symmetrical localization of extrachromosomally replicating viral genomes on sister chromatids. *J Cell Sci* 2007; 120: 1529-39.
- [9] Krauer KG, Burgess A, Buck M, Flanagan J, Sculley TB, Gabrielli B. The EBNA-3 gene family proteins disrupt the G2/M checkpoint. *Oncogene* 2004; 23: 1342-53.
- [10] Munger K. The role of human papillomaviruses in human cancers. *Front Biosci* 2002; 7: d641-49.
- [11] Nakai H, Montini E, Fuess S, Storm TA, Grompe M, Kay MA. AAV serotype 2 vectors preferentially integrate into active genes in mice. *Nat Genet* 2003; 34: 297-302.
- [12] Cougot D, Neuveut C, Buendia MA. HBV induced carcinogenesis. *J Clin Virol* 2005; 34: S75-78.
- [13] Donsante A, Miller DG, Li Y, *et al.* AAV vector integration sites in mouse hepatocellular carcinoma. *Science* 2007; 317: 477.
- [14] Kay MA. AAV vectors and tumorigenicity. *Nat Biotech* 2007; 10: 1111-3.
- [15] Perucho M, Hanahan D, Wigler M. Genetic and physical linkage of exogenous sequences in transformed cells. *Cell* 1980; 22: 309-17.
- [16] Drexler GA, Wilde S, Beisker W, Ellwart J, Eckardt-Schupp F, Fritz E. The rate of extrachromosomal homologous recombination within a novel reporter plasmid is elevated in cells lacking functional ATM protein. *DNA Repair* 2004; 3: 1345-53.
- [17] Little KC, Chartrand P. Genomic DNA is captured and amplified during double-strand break (DSB) repair in human cells. *Oncogene* 2004; 23: 4166-72.
- [18] Wang Z, Troilo PJ, Wang X, *et al.* Detection of integration of plasmid DNA into host genomic DNA following intramuscular injection and electroporation. *Gene Ther* 2004; 11: 711-24.
- [19] Kenyon J, Gerson SL. The role of DNA damage repair in aging of adult stem cells. *Nucleic Acids Res* 2007; 35: 7557-65.
- [20] Parekh-Olmedo H, Ferrara L, Brachman E, Kmiec EB. Gene therapy progress and prospects: targeted gene repair. *Gene Ther* 2005; 12: 639-46.
- [21] Parekh-Olmedo H, Kmiec EB. Progress and prospects: targeted gene alteration (TGA). *Gene Ther* 2007; 14: 1675-80.
- [22] Gogarten JP, Hilario E. Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol Biol* 2006; 6: 94.
- [23] Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I. FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci USA* 1998; 95: 10570-75.
- [24] Wah DA, Bitinaite J, Schildkraut I, Aggarwal AK. Structure of FokI has implications for DNA cleavage. *Proc Natl Acad Sci USA* 1998; 95: 10564-69.
- [25] Kim YG, Chandrasegaran S. Chimeric restriction endonuclease. *Proc Natl Acad Sci USA* 1994; 91: 883-87.
- [26] Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to FokI cleavage domain. *Proc Natl Acad Sci USA* 1996; 93: 1156-60.
- [27] Kim YG, Smith J, Dugesha M, Chandrasegaran S. Chimeric restriction enzyme: Gal4 fusion to FokI cleavage domain. *Biol Chem* 1998; 379: 489-95.
- [28] Gamsjaeger R, Liew CK, Loughlin FE, Crossley M, Mackay JP. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem Sci* 2007; 32: 63-70.
- [29] Lee MS, Gippert GP, Soman KV, Case DA, Wright PE. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* 1989; 245: 635-7.
- [30] Pavletich NP, Pabo CO. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 1991; 252: 809-17.
- [31] Wright DA, Thibodeau-Beganny S, Sander JD, *et al.* Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc* 2006; 1: 1637-52.
- [32] Papworth M, Kolasinska P, Minczuk M. Designer zinc-finger proteins and their applications. *Gene* 2006; 366: 27-38.
- [33] Beerli RR, Barbas CF3rd. Engineering polydactyl zinc-finger transcription factors. *Nat Biotech* 2002; 20: 135-141.
- [34] Durai S, Mani M, Kandavelou K, Wu J, Porteus MH, Chandrasegaran S. Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res* 2005; 33: 5978-990.
- [35] Sander JD, Zaback P, Joung JK, Voytas DF, Dobbs D. Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Res* 2007; 35: W599-605.
- [36] Bibikova M, Carroll D, Segal DJ, *et al.* Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* 2001; 21: 289-97.
- [37] Urnov FD, Miller JC, Lee YL, *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 2005; 435: 646-751.
- [38] Moehle EA, Rock JM, Lee YL, *et al.* Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci USA* 2007; 104: 3055-60.
- [39] Thierry A, Dujon B. Nested chromosomal fragmentation in yeast using the meganuclease I-SceI: a new method for physical mapping of eukaryotic genomes. *Nucleic Acids Res* 1992; 20: 5625-31.
- [40] Paques F, Duchateau P. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther* 2007; 7: 49-66.
- [41] Jurica MS, Monnat RJ Jr, Stoddard BL. DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. *Mol Cell* 1998; 2: 469-76.
- [42] Chevalier BS, Monnat RJ Jr, Stoddard BL. The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. *Nat Struct Biol* 2001; 8: 312-16.
- [43] Epinat JC, Arnould S, Chames P, *et al.* A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res* 2003; 31: 2952-62.
- [44] Smith J, Grizot S, Arnould S, *et al.* A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res* 2006; 34: e149.
- [45] Fajardo-Sanchez E, Stricher F, Paques F, Isalan M, Serrano L. Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences. *Nucleic Acids Res* 2008; 36(7): 2163-73.
- [46] Vasquez KM, Marburger K, Intody Z, Wilson JH. Manipulating the mammalian genome by homologous recombination. *Proc Natl Acad Sci USA* 2001; 98: 8403-10.
- [47] Delacote F, Lopez BS. Importance of the cell cycle phase for the choice of the appropriate DSB repair pathway, for genome stability maintenance: the trans-S double-strand break repair model. *Cell Cycle* 2008; 7: 33-38.

- [48] Lakshmiipathy U, Pelacho B, Sudo K, *et al.* Efficient transfection of embryonic and adult stem cells. *Stem Cells* 2004; 22: 531-43.
- [49] Smith MC, Thorpe HM. Diversity in the serine recombinases. *Mol Microbiol* 2002; 44: 299-307.
- [50] Chandler M, Mahillon J. Insertion sequences revisited. In: Eds Craig NL, Craigie R, Gellert M, Lambowitz AM, *Mobile DNA II*. Washington, DC, ASM Press. pp. 2002; 305-66.
- [51] Wicker T, Sabot F, Hua-Van A, *et al.* A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 2007; 8: 973-82.
- [52] Yant SR, Kay MA. Nonhomologous-end-joining factors regulate DNA repair fidelity during Sleeping Beauty element transposition in mammalian cells. *Mol Cell Biol* 2003; 23: 8505-18.
- [53] Izsvák Z, Stüwe EE, Fiedler D, Katzer A, Jeggo PA, Ivics Z. Healing the wounds inflicted by sleeping beauty transposition by double-strand break repair in mammalian somatic cells. *Mol Cell* 2004; 13: 279-90.
- [54] Walisko O, Izsvák Z, Szabó K, Kaufman CD, Herold S, Ivics Z. Sleeping Beauty transposase modulates cell-cycle progression through interaction with Miz-1. *Proc Natl Acad Sci USA* 2006 Mar 14; 103(11): 4062-7. Epub 2006 Mar 7.
- [55] Fortini P, Dogliotti E. Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair* 2007; 6: 398-409.
- [56] Mitra R, Fain-Thornton J, Craig NL. piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J* 2008; 27(7):1097-109.
- [57] Grainge I, Jayaram M. The integrase family of recombinase: organization and function of the active site. *Mol Microbiol* 1999; 33: 449-56.
- [58] Kolb AF. Genome engineering using site-specific recombinases. *Cloning Stem Cells* 2002; 4: 65-80.
- [59] Groth AC, Calos MP. Phage integrases: biology and applications. *J Mol Biol* 2004; 335: 667-78.
- [60] Grindley ND, Whiteson KL, Rice PA. Mechanisms of site-specific recombination. *Annu Rev Biochem* 2006; 75: 567-605.
- [61] Thorpe HM, Smith MC. *In vitro* site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci USA* 1998; 95: 5505-5010.
- [62] Calos MP. The phiC31 integrase system for gene therapy. *Curr Gene Ther* 2006; 6: 633-45.
- [63] Coates CJ, Kaminski JM, Summers JB, Segal DJ, Miller AD, Kolb AF. Site-directed genome modification: derivatives of DNA-modifying enzymes as targeting tools. *Trends Biotechnol* 2005; 23: 407-19.
- [64] Chalberg TW, Portlock JL, Olivares EC, *et al.* Integration specificity of phage phiC31 integrase in the human genome. *J Mol Biol* 2006; 357: 28-48.
- [65] Futreal PA, Coin L, Marshall M, *et al.* A census of human cancer genes. *Nat Rev Cancer* 2004; 4: 177-83.
- [66] Selimenti CR, Thyagarajan B, Calos MP. Directed evolution of a recombinase for improved genomic integration at a native human sequence. *Nucleic Acids Res* 2001; 29: 5044-51.
- [67] Olivares EC, Hollis RP, Chalberg TW, Meuse L, Kay MA, Calos MP. Site-specific genomic integration produces therapeutic Factor IX levels in mice. *Nat Biotechnol* 2002; 20: 1124-28.
- [68] Liu J, Jeppesen I, Nielsen K, Jensen TG. phiC31 integrase induces chromosomal aberrations in primary human fibroblasts. *Gene Ther* 2006; 13: 1188-90.
- [69] Starr TK, Largaespada DA. Cancer gene discovery using the Sleeping Beauty transposon. *Cell Cycle* 2005; 4: 1744-48.
- [70] Spradling AC, Rubin GM. Transposition of cloned P elements into *Drosophila* germ line chromosomes. *Science* 1982; 218: 341-7.
- [71] Rio DC, Barnes G, Laski FA, Rine J, Rubin GM. Evidence for *Drosophila* P element transposase activity in mammalian cells and yeast. *J Mol Biol* 1988; 200: 411-5.
- [72] Ivics Z, Hackett PB, Plasterk RH, Izsvák Z. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* 1997; 91: 501-10.
- [73] Izsvák Z, Ivics Z, Plasterk RH. Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. *J Mol Biol* 2000; 302: 93-102.
- [74] Brillet B, Bigot Y, Augé-Gouillou C. Assembly of the Tc1 and mariner transposition initiation complexes depends on the origins of their transposase DNA binding domains. *Genetica* 2007; 130: 105-20.
- [75] Yant SR, Park J, Huang Y, Mikkelsen JG, Kay MA. Mutational analysis of the N-terminal DNA-binding domain of sleeping beauty transposase: critical residues for DNA binding and hyperactivity in mammalian cells. *Mol Cell Biol* 2004; 24: 9239-47.
- [76] Zayed H, Izsvák Z, Walisko O, Ivics Z. Development of hyperactive sleeping beauty transposon vectors by mutational analysis. *Mol Ther* 2004; 9: 292-304.
- [77] Baus J, Liu L, Heggstad AD, Sanz S, Fletcher BS. Hyperactive transposase mutants of the Sleeping Beauty transposon. *Mol Ther* 2005; 12: 1148-56.
- [78] Izsvák Z, Khare D, Behlke J, Heinemann U, Plasterk RH, Ivics Z. Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in Sleeping Beauty transposition. *J Biol Chem* 2002; 277: 34581-88.
- [79] Cui Z, Geurts AM, Liu G, Kaufman CD, Hackett PB. Structure-function analysis of the inverted terminal repeats of the sleeping beauty transposon. *J Mol Biol* 2002; 318: 1221-35.
- [80] Karsi A, Moav B, Hackett P, Liu Z. Effects of insert size on transposition efficiency of the sleeping beauty transposon in mouse cells. *Mar Biotechnol* 2000; 3: 241-45.
- [81] Ortiz-Urda S, Lin Q, Yant SR, Keene D, Kay MA, Khavari PA. Sustainable correction of junctional epidermolysis bullosa via transposon-mediated nonviral gene transfer. *Gene Ther* 2003; 10: 1099-104.
- [82] Vidal TJ, Kaufman CD, Izsvák Z, Voytas DF, Ivics Z. Common physical properties of DNA affecting target site selection of Sleeping beauty and other Tc1/mariner transposable elements. *J Mol Biol* 2002; 323: 441-52.
- [83] Geurts AM, Hackett CS, Bell J, *et al.* Structure-based prediction of insertion-site preferences of transposons into chromosomes. *Nucleic Acids Res* 2006; 34: 2803-11.
- [84] Hackett CS, Geurts AM, Hackett PB. Predicting preferential DNA vector insertion sites: implications for functional genomics and gene therapy. *Genome Biol (Suppl. 8)* 2007; 1: S12.
- [85] Wilson MH, Kaminski JM, George AL Jr. Functional zinc finger/sleeping beauty transposase chimeras exhibit attenuated overproduction inhibition. *FEBS Lett* 2005; 579: 6205-09.
- [86] Yant SR, Huang Y, Akache B, Kay MA. Site-directed transposon integration in human cells. *Nucleic Acids Res* 2007; 35: e50.
- [87] Ivics Z, Katzer A, Stüwe EE, Fiedler D, Knespel S, Izsvák Z. Targeted Sleeping Beauty transposition in human cells. *Mol Ther* 2007; 15: 1137-44.
- [88] Geurts AM, Yang Y, Clark KJ, *et al.* Gene transfer into genomes of human cells by the sleeping beauty transposon system. *Mol Ther* 2003; 8: 108-17.
- [89] Wu SC, Meir YJ, Coates CJ, *et al.* piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. *Proc Natl Acad Sci USA* 2006; 103: 15008-13.
- [90] Koga A, Suzuki M, Inagaki H, Bessho Y, Hori H. Transposable element in fish. *Nature*. Sep 1996; 5: 383 (6595): 30.
- [91] Hickman AB, Perez ZN, Zhou L, *et al.* Molecular architecture of a eukaryotic DNA transposase. *Nat Struct Mol Biol* 2005; 12: 715-21.
- [92] Michel K, O'Brochta DA, Atkinson PW. The C-terminus of the Hermes transposase contains a protein multimerization domain. *Insect Biochem Mol Biol* 2003; 33: 959-70.
- [93] Zhou L, Mitra R, Atkinson PW, Hickman AB, Dyda F, Craig NL. Transposition of hAT elements links transposable elements and V(D)J recombination. *Nature* 2004; 432: 995-1001.
- [94] Urasaki A, Morvan G, Kawakami K. Functional dissection of the Tol2 transposable element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics* 2006; 174: 639-49.
- [95] Balciunas D, Wangenstein KJ, Wilber A, *et al.* Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. *PLoS Genet* 2006; 2: e169.
- [96] Kwan KM, Fujimoto E, Grabher C, *et al.* The Tol2 kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Dev Dyn* 2007; 236: 3088-99.
- [97] Fraser MJ, Cary L, Boonvisudhi K, Wang HG. Assay for movement of Lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA. *Virology* 1995; 211: 397-407.

- [98] Sarkar A, Sim C, Hong YS, Hogan JR, Fraser MJ, Robertson HM. Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences. *Mol Genet Genomics* 2003; 270: 173-80.
- [99] Hikosaka A, Kobayashi T, Kawahara A. Evolution of the *Xenopus* piggyBac transposon family TxpB: domesticated and untamed strategies of transposon subfamilies. *Mol Biol Evol* 2007; 24: 2648-56.
- [100] Wilson MH, Coates CJ, George AL Jr. PiggyBac Transposon-mediated Gene Transfer in Human Cells *Mol Ther* 2007; 15: 139-45.
- [101] Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T. Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* 2005; 122: 473-83.
- [102] Cadiñanos J, Bradley A. Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Res* 2007; 35: e87.
- [103] Maragathavally KJ, Kaminski JM, Coates CJ. Chimeric *Mos1* and piggyBac transposases result in site-directed integration. *FASEB J* 2006; 20: 1880-2.
- [104] Franz G, Savakis C. Minos, a new transposable element from *Drosophila hydei*, is a member of the Tc1-like family of transposons. *Nucleic Acids Res* 1991; 19: 6646.
- [105] Miskey C, Izsvák Z, Plasterk RH, Ivics Z. The Frog Prince: a reconstructed transposon from *Rana pipiens* with high transpositional activity in vertebrate cells. *Nucleic Acids Res* 2003; 31: 6873-81.
- [106] Zagoraiou L, Drabek D, Alexaki S, *et al.* *In vivo* transposition of Minos, a *Drosophila* mobile element, in mammalian tissues. *Proc Natl Acad Sci USA* 2001; 98: 11474-78.
- [107] Bigot Y, Brillet B, Augé-Gouillou C. Conservation of Palindromic and Mirror Motifs within Inverted Terminal Repeats of mariner-like Elements. *J Mol Biol* 2005; 351: 108-16.
- [108] Lampe DJ, Akerley BJ, Rubin EJ, Mekalanos JJ, Robertson HM. Hyperactive transposase mutants of the Himar1 mariner transposon. *Proc Natl Acad Sci USA* 1999; 96: 11428-33.
- [109] Miskey C, Papp B, Mátés L, *et al.* The ancient mariner sails again: transposition of the human Hsmar1 element by a reconstructed transposase and activities of the SETMAR protein on transposon ends. *Mol Cell Biol* 2007; 27: 4589-600.
- [110] Keravala A, Liu D, Lechman ER, *et al.* Hyperactive Himar1 transposase mediates transposition in cell culture and enhances gene expression *in vivo*. *Hum Gene Ther* 2006; 7: 1006-18.
- [111] Boulton T, Bessereau JL. *Mos1*-mediated insertional mutagenesis in *Caenorhabditis elegans*. *Nat Protoc* 2007; 2: 1276-87.
- [112] Fischer SE, Wienholds E, Plasterk RH. Regulated transposition of a fish transposon in the mouse germ line. *Proc Natl Acad Sci USA* 2001; 98: 6759-64.
- [113] Cordaux R, Udit S, Batzer MA, Feschotte C. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc Natl Acad Sci U S A* 2006; 103: 8101-6.
- [114] Augé-Gouillou C, Brillet B, Hamelin MH, Bigot Y. Assembly of the mariner *Mos1* synaptic complex. *Mol Cell Biol* 2005; 25: 2861-70.
- [115] Feng X, Colloms SD. *In vitro* transposition of ISY100, a bacterial insertion sequence belonging to the Tc1/mariner family. *Mol Microbiol* 2007; 65: 1432-43.
- [116] Feng X, Colloms SD. *In vitro* transposition of ISY100, a bacterial insertion sequence belonging to the Tc1/mariner family. Poster presentation at the Oxford workshop on Site-Specific Recombination, Transposition and DNA Dynamics 2006; 65(6):1432-43.
- [117] Leusch MS, Lee SC, Olins PO. A novel host-vector system for direct selection of recombinant baculoviruses (bacmids) in *Escherichia coli*. *Gene* 1995; 160: 191-94.
- [118] McKenzie GJ, Craig NL. Fast, easy and efficient: site-specific insertion of transgenes into enterobacterial chromosomes using Tn7 without need for selection of the insertion event. *BMC Microbiol* 2006; 6: 39.
- [119] Chakrabarti A, Desai P, Wickstrom E. Transposon Tn7 protein TnsD binding to *Escherichia coli* attTn7 DNA and its eukaryotic orthologs. *Biochemistry* 2004; 43: 2941-46.
- [120] Kuduvali PN, Mitra R, Craig NL. Site-specific Tn7 transposition into the human genome. *Nucleic Acids Res* 2005; 33: 857-63.
- [121] Qvarnström OF, Simonsson M, Johansson KA, Nyman J, Turesson I. DNA double strand break quantification in skin biopsies. *Radiother Oncol* 2004; 72: 311-17.
- [122] Nakamura A, Sedelnikova OA, Redon C, *et al.* Techniques for gamma-H2AX detection. *Methods Enzymol* 2006; 409: 236-50.
- [123] Malanga M, Althaus FR. The role of poly(ADP-ribose) in the DNA damage signaling network. *Biochem Cell Biol* 2005; 83: 354-64.
- [124] Holz O, Jörres R, Kästner A, Krause T, Magnusson H. Reproducibility of basal and induced DNA single-strand breaks detected by the single-cell gel electrophoresis assay in human peripheral mononuclear leukocytes. *Int Arch Occup Environ Health* 1995; 67: 305-10.
- [125] Hengstler JG, Fuchs J, Tanner B, Oesch-Bartlomowicz B, Hölz C, Oesch F. Analysis of DNA single-strand breaks in human venous blood: a technique which does not require isolation of white blood cells. *Environ Mol Mutagen* 1997; 29: 58-62.
- [126] Porteus MH, Baltimore D. Chimeric nucleases stimulate gene targeting in human cells. *Science* 2003; 300: 763.
- [127] Alwin S, Gere MB, Guhl E, *et al.* Custom zinc-finger nucleases for use in human cells. *Mol Ther* 2005; 12: 610-17.
- [128] Gouble A, Smith J, Bruneau S, *et al.* Efficient *in toto* targeted recombination in mouse liver by meganuclease-induced double-strand break. *J Gene Med* 2006; 8: 616-22.
- [129] Szczepek M, Brondani V, Büchel J, Serrano L, Segal DJ, Cathomen T. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol* 2007; 25: 786-93.
- [130] Cornu TI, Thibodeau-Beganny S, Guhl E, *et al.* DNA-binding Specificity Is a Major Determinant of the Activity and Toxicity of Zinc-finger Nucleases. *Mol Ther* 2008; 16: 352-8.
- [131] Gasior SL, Wakeman TP, Xu B, Deiningner PL. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol* 2006; 357: 1383-93.
- [132] Hollis RP, Stoll SM, Sclementi CR, Lin J, Chen-Tsay Y, Calos MP. Phage integrases for the construction and manipulation of transgenic mammals. *Reprod Biol Endocrinol* 2003; 1: 79.
- [133] Wilber A, Frandsen JL, Geurts JL, Largaespada DA, Hackett PB, McIvor RS. RNA as a source of transposase for Sleeping Beauty-mediated gene insertion and expression in somatic cells and tissues. *Mol Ther* 2005; 3: 625-30.
- [134] Wilber A, Wangenstein KJ, Chen Y, *et al.* Messenger RNA as a Source of Transposase for Sleeping Beauty Transposon-mediated Correction of Hereditary Tyrosinemia Type I. *Mol Ther* 2007; 15: 1280-87.
- [135] Foerg C, Merkle HP. On the biomedical promise of cell penetrating peptides: limits versus prospects. *J Pharm Sci* 2008; 97: 144-62.
- [136] Rideharson C, Jasin M. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* 2000; 405: 697-700.
- [137] Garrick D, Fiering S, Martin DI, Whitelaw I. Repeat-induced gene silencing in mammals. *Nat Genet* 1998; 18: 56-59.
- [138] Henikoff S. Conspiracy of silence among repeated transgene. *Bioassays* 1998; 20: 532-35.
- [139] Balagué C, Kalla M, Zhang WW. Adeno-associated virus Rep78 protein and terminal repeats enhance integration of DNA sequences into the cellular genome. *J Virol* 1997; 71: 3299-306.
- [140] Held PK, Olivares EC, Aguilar CP, Finegold M, Calos MP, Grompe M. *In vivo* correction of murine hereditary tyrosinemia type I by phiC31 integrase-mediated gene delivery. *Mol Ther* 2005; 11(3):399-408.
- [141] Converse AD, Belur LR, Gori JL, *et al.* Counterselection and co-delivery of transposon and transposase functions for Sleeping Beauty-mediated transposition in cultured mammalian cells. *Biosci Rep* 2004; 24: 577-94.
- [142] Bonini C, Ferrari G, Verzeletti S, *et al.* HSV-TK gene transfer into donor lymphocytes for control of allogeneic graft-versus-leukemia. *Science* 1997; 276: 1719-24.
- [143] Score PR, Belur LR, Frandsen JL, *et al.* Sleeping Beauty-mediated transposition and long-term expression *in vivo*: use of the LoxP/Cre recombinase system to distinguish transposition-specific expression. *Mol Ther* 2006; 13: 617-24.
- [144] Huang X, Wilber AC, Bao L, *et al.* Stable gene transfer and expression in human primary T cells by the Sleeping Beauty transposon system. *Blood* 2006; 107: 483-91.
- [145] Ivics Z, Izsvák Z. Transposons for gene therapy! *Curr Gene Ther* 2006; 6: 593-607.
- [146] Ehrhardt A, Xu H, Huang Z, Engler JA, Kay MA. A direct comparison of two nonviral gene therapy vectors for somatic integration: *in vivo* evaluation of the bacteriophage integrase phiC31 and the Sleeping Beauty transposase. *Mol Ther* 2005; 11: 695-706.

- [147] Tolar J, Osborn M, Bell S, *et al.* Real-time *in vivo* imaging of stem cells following transgenesis by transposition. *Mol Ther* 2005; 12: 42-48.
- [148] Wilber A, Linehan JL, Tian X, *et al.* Efficient and stable transgene expression in human embryonic stem cells using transposon-mediated gene transfer. *Stem Cells* 2007; 25: 2919-27.
- [149] Betzl G, Brem G, Weidle UH. Epigenetic modification of transgenes under the control of the mouse mammary tumor virus LTR: tissue-dependent influence on transcription of the transgenes. *Biol Chem* 1996; 377: 711-19.
- [150] Challita PM, Skelton D, el-Khoueiry A, Yu XJ, Weinberg K, Kohn DB. Multiple modifications in cis elements of the long terminal repeat of retroviral vectors lead to increased expression and decreased DNA methylation in embryonic carcinoma cells. *J Virol* 1995; 69: 748-55.
- [151] Schumacher A, Koetsier PA, Hertz J, Doerfler W. Epigenetic and genotype-specific effects on the stability of de novo imposed methylation patterns in transgenic mice. *J Biol Chem* 2000; 275: 37915-921.
- [152] Park CW, Park J, Kren BT, Steer CJ. *Sleeping Beauty* transposition in the mouse genome is associated with changes in DNA methylation at the site of insertion. *Genomics* 2006; 88: 204-13.
- [153] Garrison BS, Yant SR, Mikkelsen JG, Kay MA. Postintegrative gene silencing within the *Sleeping Beauty* transposition system. *Mol Cell Biol* 2007; 27: 8824-33.
- [154] Lu X, Humeau L, Slepishkin V, *et al.* Safe two-plasmid production for the first clinical lentivirus vector that achieves >99% transduction in primary cells using a one-step protocol. *J Gene Med* 2004; 6: 963-73.
- [155] Papapetrou EP, Ziros PG, Micheva ID, Zoumbos NC, Athanasiadou A. Gene transfer into human hematopoietic progenitor cells with an episomal vector carrying an S/MAR element. *Gene Ther* 2006; 13: 40-51.
- [156] Moldt B, Yant SR, Andersen PR, Kay MA, Mikkelsen JG. Cis-acting gene regulatory activities in the terminal regions of sleeping beauty DNA transposon-based vectors. *Hum Gene Ther* 2007; 18: 1193-204.
- [157] Walisko O, Schorn A, Rolfs F, *et al.* Transcriptional activities of the sleeping beauty transposon and shielding its genetic cargo with insulators. *Mol Ther* 2008; 16: 359-69.
- [158] Kolb AF, Coates CJ, Kaminski JM, Summers JB, Miller AD, Segal DJ. Site-directed genome modification: nucleic acid and protein modules for targeted integration and gene correction. *Trends Biotechnol* 2005; 23: 399-406.
- [159] Izsvák Z, Ivics Z. Sleeping beauty transposition: biology and applications for molecular therapy. *Mol Ther* 2004; 9: 147-56.
- [160] Gerstein MB, Bruce C, Rozowsky JS, *et al.* What is a gene, post-ENCODE? History and updated definition. *Genome Res* 2007; 17: 669-81.
- [161] Lopes TS, Klootwijk J, Veenstra AE, *et al.* High-copy-number integration into the ribosomal DNA of *Saccharomyces cerevisiae*: a new vector for high-level expression. *Gene* 1989; 79: 199-206.
- [162] Wery J, Gutker D, Rennie AC, Verdoes JC, van Ooyen AJ. High copy number integration into the ribosomal DNA of the yeast *Phaffia rhodozyma*. *Gene* 1997; 184: 89-97.
- [163] Klabunde J, Kunze G, Gellisen G, Hollenberg CP. Wide-range integrative expression vectors for fungi, based on ribosomal DNA elements. In: Eds Gellissen G, Production of recombinant proteins. Weinheim, Wiley-VCH. pp. 2004; 273-86.
- [164] dos Santos WG, Buck GA. Simultaneous stable expression of neomycin phosphotransferase and green fluorescence protein genes in *Trypanosoma cruzi*. *J Parasitol* 2000; 86: 1281-88.
- [165] Penton EH, Crease TJ. Evolution of the transposable element Pokey in the ribosomal DNA of species in the subgenus *Daphnia* (Crustacea: Cladocera). *Mol Biol Evol* 2004; 21: 1727-39.
- [166] Kojima KK, Fujiwara H. An extraordinary retrotransposon family encoding dual exonuclease. *Gen Res* 2005; 15: 1106-17.
- [167] Kojima KK, Fujiwara H. Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. *Mol Biol Evol* 2004; 21: 207-17.
- [168] Stewart S, MacDonald N, Perkins E, deJong G, Perez C, Lindebaum M. Retrofitting of a satellite repeat DNA-based murine artificial chromosome (ACes) to contain loxP recombination sites. *Gene Ther* 2002; 9: 719-23.
- [169] Liu X, Liu M, She H, *et al.* Expression of reconstructive hFVIII in the hrDNA by using hrDNA targeting vector. *Chin Sci Bull Engl* 2005; 50: 2187-92.
- [170] Liu X, Liu M, Xue Z, *et al.* Non-viral *ex vivo* transduction of human hepatocyte cells to express factor VIII using a human ribosomal DNA-targeting vector. *J Thromb Haemost* 2007; 5: 347-51.
- [171] Liang ZM, Luo H, Yang YF, *et al.* Targeted expression of vascular endothelial growth factor 165 in the hrDNA locus mediated by hrDNA targeting vector. *Chin Med J* 2007; 120: 427-30.
- [172] Wen L, Liang de S, Wu LQ, *et al.* Efficient protein expression from the endogenous RNA polymerase I promoter using a human ribosomal DNA targeting vector. *Biochem Biophys Res Commun* 2008; 367: 846-51.
- [173] Christensen SM, Bibillo A, Eickbush T. Role of the *Bombyx mori* R2 element N-terminal domain in the target-primed reverse transcription (TPRT) reaction. *Nucleic Acids Res* 2005; 33: 6461-68.
- [174] Palazzoli F, Testu FX, Merly F, Bigot Y. Transposon tools: world distribution of the intellectual property and technological development. *Genetica*, special issue "Mariner tools: from engineering to applications", *Gene Therapy*, in press (2009).
- [175] Kaiser J. Gene Therapy. Putting the fingers on gene repair. *Science* 2005; 310: 1894-96.

3.2. Article 3 : *Transposon tools: worldwide landscape of intellectual property and technological developments*

3.2.1. Contexte et objectif de l'article

Après avoir inventorié les différents systèmes d'intégration enzymatiques non viraux [Palazzoli *et al.*, 2008], nous nous sommes focalisés sur les vecteurs dérivés des transposons. Ces éléments sont très étudiés dans le cadre du transfert de fragments d'ADN exogène [Miskey *et al.*, 2005] et ils représentent de bons candidats pour élaborer des outils en biotechnologies ou pour développer des vecteurs de transgénèse [Palazzoli *et al.*, 2008]. Pendant une dizaine d'années, le modèle d'étude des équipes 1 et 2 du GICC a été le transposon *mariner* MOS1 [Medhora *et al.*, 1991]. L'un des objectifs du projet SyntheGeneDelivery prévoyait le développement technologique de vecteurs à partir de ce transposon pour le transfert de gène en cellules de mammifères. Cependant, nos travaux et ceux d'autres laboratoires ont montré que le taux d'intégration des vecteurs optimisés MOS1 dans les cellules de mammifères était extrêmement faible, sinon nul [Wu *et al.*, 2006]. À l'inverse des systèmes concurrents, MOS1 ne pouvait être employé efficacement pour réaliser du transfert de gène en thérapie génique ou en bioproduction. Par conséquent, nous avons recherché un système alternatif efficace en cellules de mammifères et bénéficiant d'une liberté d'exploitation (non protégé par des droits de PI). Dans ce but, nous avons construit le paysage brevets sur les transposons pour identifier une niche de développement technologique pour le laboratoire.

3.2.2. Résumé de l'article

Après une rapide présentation du système des brevets et des conditions de brevetabilité, nous avons présenté leur importance en tant qu'outil incontournable pour le développement industriel des biotechnologies. En effet, en contrepartie de la protection exclusive accordée, les brevets fournissent des informations essentielles pour la prise de décisions stratégiques comme la définition de niches de développement technologique. La partie suivante de l'article présente le paysage brevets concernant les demandes PCT qui couvrent les principaux systèmes transposons : *Tn5*, *Tn10*, *Sleeping Beauty*, *Tol2*, *Minos*,

Mos1, les éléments *Mariner*, l'élément *P*, *piggyBac*, *Himar*, *hobo* et les éléments *Tc1*. Les résultats portent sur l'évolution des demandes PCT au cours de la période 1986-2006, sur les déposants (nationalités, Top 10, évolution du nombre de déposants), sur les applications revendiquées dans les demandes de brevets et sur le rôle clé des auteurs d'articles scientifiques majeurs. Enfin, nous avons recensé les statuts légaux des demandes de brevets qui portent sur la preuve de concept et l'ingénierie de ces systèmes. Ces informations nous ont permis de remarquer que de nombreux transposons entraient dans le plan d'affaires de sociétés de biotechnologies spécialement dédiées (Discovery Genomics pour *Sleeping Beauty*, Minos Biosystems pour *Minos*...). La plupart des systèmes transposons ont été mis au point dans les laboratoires académiques des auteurs pionniers à l'origine de ces technologies. Ces derniers ont ensuite fait émerger des *spin-offs* pour exploiter commercialement les systèmes qu'ils avaient mis au point et protégés par des brevets. Cependant, un transposon se détache des autres systèmes. En effet, *piggyBac* est le seul élément efficace à être librement exploitable (deux brevets accordés aux États-Unis n'ont pas été étendus) à cause d'une divulgation antérieure au dépôt de brevet.

3.2.3. Conclusion de l'article

Les principaux systèmes transposons : *Tn5*, *Tn10*, *Sleeping Beauty*, *Tol2*, *Minos*, *Mos1*, les éléments *Mariner*, l'élément *P*, *piggyBac*, *Himar*, *Hobo* et les éléments *Tc1* ont fait l'objet de demandes internationales de brevets, à l'exception du transposon *Hobo*. Certains systèmes sont le sujet d'une exploitation économique par des sociétés de biotechnologies créées à cet effet (Tableau 3 issu du poster *Transposon tools: worldwide patent landscape and patent exploitation by key actors* – Annexe 2).

Transposon systems	Key applicants	Commercial exploitation	Creation	Nationality	Field	Purposes
<i>Mu</i>	Finzymes	Finzymes	1986		Molecular biology - kits	DNA sequencing Functional analysis of proteins
<i>Tn5</i>	Wisconsin Alumni Research Foundation	Epicentre	1987		Molecular biology - kits	DNA sequencing
<i>Minos</i>	Institute for Molecular Biology and Biotechnology & Minos Biosystems	Minos Biosystems	2000		Molecular biology - kits	in-mass production of proteins Functional gene discovery
<i>piggyBac</i>	US Government, University Notre Dame	Oxitec Minos Biosystems Others ?	2002 2000		Agriculture Molecular biology - kits	Control insect pests in-mass production of proteins Functional gene discovery
<i>Tn10</i>	Louisiana State University and Agricultural and Mechanical College & Transgenrx	Transgenrx	2002		Therapeutic	Production of protein-based drugs
<i>Sleeping Beauty</i>	University of Minnesota	Discovery Genomics	2002		Therapeutic	Gene delivery Gene discovery Gene therapy
<i>p element</i>	Leland Stanford Junior University Tosk	Tosk	1998		Therapeutic	Drug discovery
<i>piggyBac</i>	US Government, University Notre Dame	Manoa Biosciences Others ?	2006		Therapeutic	Gene therapy
<i>Tol2</i>	Japan Science and Technology Corp	?	?	?	?	?

Tableau 3 : Résultats issus du poster *Transposon tools: worldwide patent landscape and patent exploitation by key actors*, présenté lors du congrès international *Mobile DNA* à Montréal du 24 au 28 avril 2010 (Annexe 2). Les applications non-thérapeutiques sont en couleur verte et les applications thérapeutiques sont en bleu. Seule l'exploitation commerciale du transposon *Tol2* n'a pas été identifiée.

Des acteurs industriels se sont positionnés sur le marché du transfert de gène en utilisant des systèmes dérivés de transposons. Nous pouvons observer que les sociétés qui exploitent commercialement les transposons à but thérapeutique sont des sociétés américaines. À l'inverse, les transposons utilisés en tant qu'outils de biologie moléculaire (séquençage, mutagenèse insertionnelle...) sont exploités par des sociétés européennes. Nous pouvons nous attendre à une compétition intense pour produire le système transposon le plus efficace, principalement entre les sociétés américaines qui entreront en essais cliniques pour la thérapie génique ou en phase industrielle pour la bioproduction.

Finalement, deux systèmes représentent une niche de développement technologique en Europe : *Himar1* (protégé par un brevet américain et un brevet canadien) et *piggyBac* (protégé par deux brevets américains). Nous avons choisi d'exploiter le transposon *piggyBac* car il est aussi efficace en cellules de mammifères que les systèmes modèles *Sleeping Beauty* et *Tol2* [Ding et al., 2005]. De plus, à l'inverse du système le plus en vue, *Sleeping Beauty*, qui perd son efficacité lorsque des transgènes de grande taille sont transportés, le système *piggyBac* s'accommode de larges cassettes transgéniques (14 kbp) sans diminution d'efficacité [Ding et al., 2005]. Une autre propriété intéressante de *piggyBac* concerne la possibilité de concevoir une transposase chimérique fusionnée à un domaine de liaison à

l'ADN, pour intégrer le vecteur à des sites spécifiques [Maragathavally *et al.*, 2006]. Le transposon *piggyBac* représente donc une niche de développement technologique particulièrement intéressante pour le GICC ou tout acteur souhaitant développer un système transposon dans le cadre d'une exploitation en cellules de mammifères.

3.2.4. Remarque sur la méthodologie et les outils employés

Comme pour l'article 2 [Palazzoli *et al.*, 2008], la recherche de documents brevets a été réalisée avec un outil de recherche gratuit : PATENTSCOPE [w32]. L'un de ses avantages est qu'il offre la possibilité d'effectuer des recherches en texte complet, ce qui inclut les titres et résumés comme esp@cenet, mais aussi les revendications. Cependant, la recherche se fait dans des demandes internationales de brevets (numéros en WO), ce qui signifie que toute demande de brevet n'ayant pas fait l'objet d'une telle procédure n'apparaîtra pas dans les résultats. Par exemple, il peut s'agir d'une demande de brevet qui a été déposée uniquement aux États-Unis et qui n'a pas été étendue en PCT. Une telle situation représente donc un biais dont il faudra tenir compte dans l'analyse et l'interprétation du paysage brevets élaboré à partir des résultats fournis par PATENTSCOPE.

3.2.5. De l'importance de connaître les paysages brevets avant de financer des projets de R&D : exemple du projet INTHER

Le tableau de l'article 3 qui recense les demandes de brevets internationales sur les transposons comprend de nombreux brevets pour le modèle *Sleeping Beauty*, dont la famille de brevets fondateurs : la demande PCT WO9840510 – *DNA-based transposon system for the introduction of nucleic acid into DNA of a cell*. Il est important de noter que les brevets fondateurs de *Sleeping Beauty* ont été déposés par l'Université du Minnesota et qu'ils sont exploités par la société *spin-off* Discovery Genomics. Le transposon *Sleeping Beauty* provient d'une reconstruction moléculaire réalisée par Perry Hackett, Zsuzsanna Izsvak et Zoltan Ivics, lorsqu'ils travaillaient à l'Université du Minnesota [Ivics *et al.*, 1997]. Ceci a représenté une étape importante dans l'application de la transposition en thérapie génique [Ivics & Izsvak, 2006]. Cependant, bien que Zsuzsanna Izsvak et Zoltan Ivics fassent partie des inventeurs, les

titres de PI appartiennent à l'Université du Minnesota qui a ensuite accordé des licences d'exploitation à Discovery Genomics, dont l'un des cofondateurs est Perry Hackett. Dans le système de brevets américain, les inventeurs sont également déposants et donc propriétaires des titres de PI. Néanmoins, leurs contrats de travail incluent généralement des clauses de cession des brevets pour lesquels ils sont inventeurs (par exemple US2002016975 de la famille de brevets de WO9840510). Cette particularité existe uniquement dans le système américain et non pour les autres demandes de brevets nationales ou régionales. Par ailleurs, il est intéressant de noter que *Sleeping Beauty* a été intégré en tant que vecteur de transfert de gène dans des programmes d'essais cliniques de thérapie génique [Williams, 2008].

Un projet européen concurrent à SyntheGeneDelivery a également été financé par la Commission Européenne. Coordonné par Zsuzsanna Izsvak, le projet INTHER porte sur le développement et l'utilisation du transposon *Sleeping Beauty* et de technologies d'intégration site-spécifiques comme méthodes de transfert de gène non virales en thérapie génique *ex vivo* [w55]. Nous pouvons donc nous poser la question de la viabilité de l'exploitation économique du projet INTHER, étant donné que le vecteur transposon n'appartient à aucun des partenaires du projet. Une des possibilités envisagées pourrait être que les partenaires du projet INTHER aient négocié une licence d'exploitation de *Sleeping Beauty* pour la thérapie génique auprès de l'Université du Minnesota et de Discovery Genomics, qui ne sont pas partenaires du projet. En effet, le Guide de mise en œuvre des droits de PI pour les projets du 6^{ème} PCRDT prévoit dans l'article 7 – *Coûts éligibles pour les droits de PI* [European Commission, 2006], l'intégration des redevances comme coûts éligibles dans le financement projet, si les autres conditions applicables sont respectées. Cependant, cette éligibilité peut être limitée. Par exemple, l'exclusivité d'exploitation doit être prouvée car cela augmente les redevances. Le coût total du projet INTHER était de 3,54 millions d'euros, avec un financement de la Commission Européenne de 2,8 millions d'euros [w55]. Quant au projet SyntheGeneDelivery, la Commission Européenne a financé 2,4 millions d'euros sur un coût total de 3,01 millions d'euros [w1]. Pour ce projet, aucune licence d'exploitation n'était requise puisque les brevets couvrant chaque technologie mise en œuvre dans le projet SyntheGeneDelivery appartenaient en totalité aux partenaires. La comparaison des financements des deux projets nous incite à penser que les partenaires du

projet INTHER n'ont pas obtenu (négocié ?) de licence d'exploitation pour le système *Sleeping Beauty* en thérapie génique. Par conséquent, les partenaires devraient être bloqués pour le plan d'exploitation du système développé lors du projet INTHER. De plus, les essais cliniques ne peuvent pas être envisagés sans licence d'exploitation. L'Université du Minnesota et la *spin-off* Discovery Genomics seraient donc les bénéficiaires du projet INTHER, en cas de succès et de retombées économiques. Par ailleurs, un Symposium sur le transfert de gène non viral basé sur les recombinaisons a été organisé par les consortiums des projets INTHER et SyntheGeneDelivery, lors du XVIème Congrès Annuel de la Société Européenne de Thérapies Génique et Cellulaire (12-16 novembre 2008, Brugge). Scott McIvor, un autre chercheur de l'Université du Minnesota, l'un des cofondateurs de Discovery Genomics et qui est à la fois *Chief Executive Officer* et *Chief Science Officer* était présent. Cela nous permet de penser que l'Université du Minnesota et Discovery Genomics ont habilement fait financer la recherche et le développement technologique du système prometteur *Sleeping Beauty*. En effet, il était nécessaire d'optimiser ce système transposon afin de réaliser une intégration site-spécifique, ce qui était l'un des objectifs du projet INTHER. Cela renforce l'importance de l'information brevet et des paysages brevets avant de lancer tout projet de R&D, que ce soit en biotechnologies ou dans d'autres domaines.

3.2.6. Article 3

Transposon tools: worldwide landscape of intellectual property and technological developments

Fabien Palazzoli · François-Xavier Testu ·
Franck Merly · Yves Bigot

Received: 27 May 2009 / Accepted: 3 November 2009 / Published online: 3 December 2009
© Springer Science+Business Media B.V. 2009

Abstract DNA transposons are considered to be good candidates for developing tools for genome engineering, insertional mutagenesis and gene delivery for therapeutic purposes, as illustrated by the recent first clinical trial of a transposon. In this article we set out to highlight the interest of patent information, and to develop a strategy for the technological development of transposon tools, similar to what has been done in many other fields. We propose a patent landscape for transposon tools, including the changes in international patent applications, and review the leading inventors and applicants. We also provide an overview of the

potential patent portfolio for the prokaryotic and eukaryotic transposons that are exploited by spin-off companies. Finally, we discuss the difficulties involved in tracing relevant state-of-the-art of articles and patent documents, based on the example of one of the most promising transposon systems, including all the impacts on the technological development of transposon tools.

Keywords Gene delivery · Biotechnology · Recombination · Patent landscape

Electronic supplementary material The online version of this article (doi:10.1007/s10709-009-9426-3) contains supplementary material, which is available to authorized users.

F. Palazzoli
UFR des Sciences & Techniques, Université François Rabelais de Tours, GICC, Parc de Grandmont, 37200 Tours, France

F. Palazzoli · Y. Bigot (✉)
UFR des Sciences & Techniques, CNRS, UMR 6239,
Parc de Grandmont, 37200 Tours, France
e-mail: bigot@univ-tours.fr

F.-X. Testu
UFR de Droit, Economie et Sciences Sociales, Université François Rabelais de Tours, CRDP, 50, Avenue Jean Portalis, 37200 Tours, France

Y. Bigot
UFR des Sciences & Techniques, CHRU de Tours,
Parc de Grandmont, 37200 Tours, France

F. Merly
France Innovation Scientifique & Transfert SA,
83, Boulevard Exelmans, 75016 Paris, France

Abbreviations

AU	Australia
CA	Canada
DE	Germany
DK	Denmark
DBD	DNA-binding domain
DNA	Deoxyribo nucleic acid
EPO	European patent office
EU	European
FR	France
GB	United Kingdom
IP	Intellectual property
IPC	International patent classification
NL	Netherlands
NO	Norway
PCT	Patent cooperation treaty
R&D	Research and development
RNA	Ribo nucleic acid
SB	Sleeping Beauty
TE	Transposable element
US	United States of America
WIPO	World intellectual property organization
ZFD	Zinc finger domain

Introduction

Transposable elements (TEs) were discovered more than half a century ago by Barbara McClintock. They are genetic agents responsible for the chromosomal breaks and alterations of gene expression, and have considerably remodeled prokaryotic and eukaryotic genomes over the course of evolution. These DNA fragments have the particular characteristic of moving from one chromosomal locus to another within their host genome. TEs are divided into two main classes on the basis of their transposition mechanism. Class I consists of retrotransposons, whereas Class II includes the elements known as “DNA transposons”. Class-I TEs move *via* an RNA intermediate that undergoes reverse transcription in DNA before being reinserted into the host genome. We will not consider these elements any further here, since so far most transposon tools have been engineered from Class-II TEs. DNA transposons transpose by a cut-and-paste mechanism, using a double-stranded DNA molecule as intermediate. The only Class-II elements that transpose by a different mechanism are the *Helitrons* and related elements, that transpose by rolling-circle replication, *Marvericks* (also known as *Polinton*), whose the transposition mechanism is not yet known, and the bacterial *IS200/605* family of insertion sequences that transpose as a single-stranded transposon circle (Feschotte and Pritham 2007; Barabas et al. 2008). Because of their mobility, for over two decades DNA transposons have been considered as good candidates to be used to develop tools for insertional mutagenesis, genomic engineering and gene delivery purposes. Transposon-based Research and Development (R&D) programs have rapidly been in position to propose prototypes and suggest strategies to engineer them. The first transposon tools were developed from natural elements (Spradling and Rubin 1982). Once scientists had become aware of their potential, they tried to enhance their mobility by selecting hyperactive mutants (Lampe et al. 1999; Baus et al. 2005; Germon et al. 2009; Lampe 2009), or engineering molecular reconstructions (Lampe et al. 1996; Ivics et al. 1997; Miskey et al. 2003). An example of this is the *Sleeping Beauty* (SB; Ivics et al. 1997; Izsvák et al. 2000) reference transposon system, which is used to integrate exogenous DNA into cellular vertebrate genomes (Yant et al. 2000). Subsequently, attempts have been made to improve the properties of transposon tools by adding capabilities, such as site-specific integration by modifying transposases, or developing new strategies to target the transposon vector to its integration site using DNA-binding domains (DBDs; Wilson et al. 2005; Yant et al. 2007; Ivics et al. 2007). The most promising DBDs are the Zinc Finger Domains (ZFD; Beerli and Barbas 2002; Papworth et al. 2006; Gamsjaeger et al. 2007; Sander et al. 2007), because they can be

tailor-designed *in vitro* to be specific for any site in the host genome.

Transposon-based technologies have now been developed for a variety of biotechnology applications, including gene delivery for gene therapy purposes (Williams 2008). Recently the National Institutes of Health Recombinant DNA Advisory Committee announced the assessment of the first clinical trial involving a transposon vector, entitled “Adoptive immunotherapy for CD19+ B-lymphoid malignancies using SB transposition to express a CD19-specific chimeric antigen receptor in autologous ex vivo expanded T cells”. This first clinical trial with SB have been launched about a decade after the patenting of the SB vector invention (WO9840510). Clinical trials are therefore still in progress long after patent applications have been submitted, as these need to be submitted early in the development of research programs, before the researchers know whether their invention will in fact form the basis of an exploitation plan. This highlights the importance of patents in R&D, and in particular in Biotech, where the R&D costs are high. From this standpoint, there is an increasing need for the researchers in life and sciences to challenge the ins and outs of the patenting procedures, and their implications for the future of their research programs.

The “ins and outs” of patents

Patenting procedures & source developments

The World Intellectual Property Organization (WIPO) defines a patent as “an exclusive right granted for an invention, which is a product or a process that provides, in general, a new way of doing something, or offers a new technical solution to a problem” (ws1, Sup. Mat. 1). To be patentable, the invention must fulfill patentability conditions: it must be of practical use and must have an element of novelty with characteristics that are not already known in the body of existing knowledge, known as the “prior art”. The invention must correspond to something that cannot be deduced by a person of the art, who has an average knowledge of the technical field. Briefly, the procedure to obtain patent protection requires that the applicant, a public or industrial entity, submits a patent application, generally to a national office, such as the United States Patent and Trademark Office in the US, or the Japan Patent Office in Japan. This document contains a title, an abstract, a precise description of how the patented invention functions, illustrations, and above all, claims that define the scope of the invention protection (ws2, Sup. Mat. 1). After a formal scrutiny, if the patent application is not refused or withdrawn, it will be published after an immutable 18 month period (Fig. 1). Based on the prior art, the examiners in the

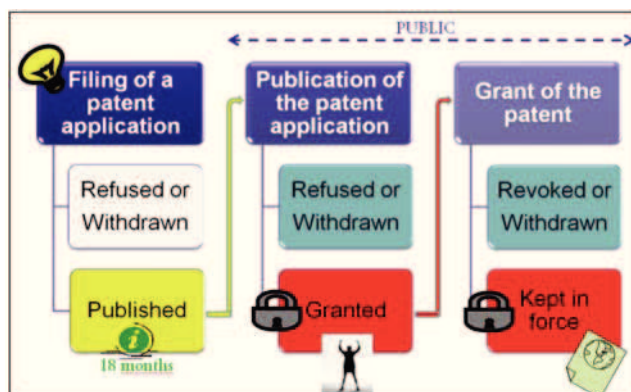


Fig. 1 A patent application becomes public after 18 months from the date of submission. If the patent is granted by the patent office, the applicant has to pay fees in order to keep the protection in force. At any moment, the patent can be withdrawn, and then the invention falls into the public domain, and becomes available for free commercial exploitation

patent offices carry out a substantive examination or scrutiny in order to decide whether they grant the patent (it can be withdrawn or refused depending on whether it fulfills the patentability conditions). If the applicant wants to get protection in other States, an appropriate international application must be submitted within 12 months from the date of the first application (priority date), designating the Patent Cooperation Treaty (PCT) contracting states. In other words, the applicant then holds a collection of national patents in the designated States (nowadays, an “international” patent no longer really exist). At this point, it is important to remember that patent laws are specific to each State, and there are divergences between national offices. For example, the US system holds that the priority for an invention is allocated to “the first person who invents”, whereas most other countries consider as decisive “the first who applies to patent”. In all cases, after paying the fees for submitting, examining and granting the patent, applicants also have to pay fees each year to keep their patents in force in each state where a patent has been granted. If applicants decide not to keep their patents in force, they will then fall into the public domain in some or all the states where they had been protected by patent.

PCT applications are useful and advisable, given the importance of patents in Biotechnology. As highlighted by the WIPO Consultant Esteban Burrone (ws3, Sup. Mat. 1), patents are at the core of the biotech business. He also pointed out that the growth in the number of patents in biotechnology is largely due to the added value that they confer on biotech companies. Indeed, patents are directly linked to the main challenge facing the biotech industry, that of obtaining the necessary finance. Biotechnology is characterized by a high level of innovation (involving technological progress in various domains, including

genetics) and the complexity of biological systems (e.g. interactions between genes) that generate very high costs for R&D. Moreover, the industrial means available are still under-developed, due to the fragility of the emerging markets for their products. This is why, although transposons are being investigated as candidate vectors for future gene delivery purposes, patents are key elements in a market forecast to reach \$5.73 billion by 2011 (ws4, Sup. Mat. 1). Also, no fully-designed product is yet available that has been validated by national or international regulatory agencies. Esteban Burrone also advances several other pertinent arguments (ws3, Sup. Mat. 1). Patents are the main assets of biotech companies, due to the facts that start-up companies are created on the basis of the patents developed, and because fundamental and applied research overlap. Protecting research findings is vital if the company wants to survive, and successful fundraising depends mainly on the ability to convince investors that there is a real IP strategy, and that the risks of infringement are negligible.

Why play the “patent game”?

As the WIPO summarizes, “Patent information represents a vast resource of technological and legal information presented in a standardized format, and frequently not available anywhere else” (WIPO 2008). Knowing what information is contained in patents can enable users to avoid patent infringement, to exploit a technology from patents that have not been kept in force or never granted in some countries, and also to locate the likely future strategic orientations of R&D competitors. Furthermore, the European Patent Office (EPO) observes that “up to 80% of current technical knowledge can be found in patent documents” (EU Commission and EPO 2007). To sum up, the study of the state of the art for an R&D project must not only focus on scientific articles, but also needs to take patent documents into account to be pertinent. The increasing importance of patent documents for the development of research programs is confirmed by the fact that specific patent-related journals appeared a few years ago in a format intended for scientists. These journals now include review articles authored by “patent experts” in various specialized fields of economic interest, such as Biotechnology or DNA & Gene Sequences, as exemplified in the collection of Recent patents on “x” (ws5, Sup. Mat. 1). In these journals, articles may also combine original scientific data with information drawn from recent important patents, including patent numbers and patent applicants, as exemplified in the article “Gene targeting for chromosome engineering applications in eukaryotic cells” (Lyznik and Dress 2008). Some journals, like *Nature in Biotechnology* and *Stem Cells* have adopted an editorial policy of keeping

some space for letters and articles that discuss patent landscapes, including critical and relevant analyses of the commercial R&D and academic research, open-science alternatives, and the strategy of the technology's owners (Vrtovec and Vrtovec 2007; Bergman and Graff 2007; Denker 2008; Vrtovec and Scott 2008; Chandrasekharan et al. 2009). In a few cases, most of the articles in purely scientific reviews have in fact included a section about IP (Palazzoli et al. 2008).

Patent contents and portfolios are of interest at two main levels (Fig. 2). On the one hand, patents give their owners (the applicants) a right to forbid and exclude third parties from economic exploitation of their inventions for a determined period and within a specific area. This allows the applicants to mobilize R&D investments that can then be amortized by exploiting the invention directly or indirectly, by awarding exclusive or non-exclusive licenses. In theory, this does not prohibit the use of the patented technology for non-commercial fundamental research purposes, as long as the regulations in each country make it possible to demonstrate unambiguously whether this research overlaps with economical and/or commercial activities. On the other hand, patents are also of public interest, since they involve the release of published documents that can be considered to provide easily accessible information belonging to the shared technological heritage (thus providing easily accessible technological information) (ws6, Sup. Mat. 1). Patents therefore provide an opportunity to construct an IP and a technological development strategy.

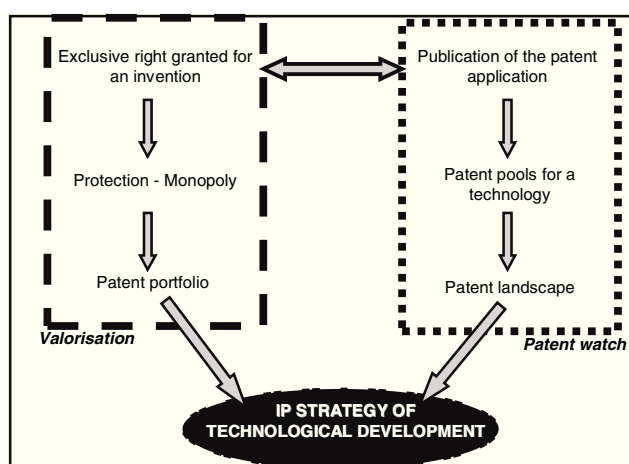


Fig. 2 In compensation for the patent that grants a monopoly of the invention (and hence of its valorization), the protected invention is published with a full description, which is fully available to the public. The publications of patent applications make it possible to compile a patent landscape of a given technology. Valorization and patent watch should be taken into account when thinking through the strategy of technological developments

The interest of using patent documents is, therefore, to exploit data extracted from patent databases. Consequently, a pool of patent documents resulting from a search for a specific research topic in patent databases produces not only the patents that protect the technology involved, including competitor models (e.g. *SB*, *piggyBac*, *Tol2*...), but also the technologies that can be used to exploit them. These searches, the results of which are known as patent landscapes, are performed using dedicated computing tools, and can be exploited at two levels (ws7, Sup. Mat. 1). The first level is only interested in the bibliographical data of patent documents, such as their publication number and date, priority data, applicants and inventors. A global patent landscape of this type is a competitive intelligence tool, used to optimize strategic choices and manage an IP portfolio, and it constitutes an asset for those who “play the international patent game” (Barrett 2003). Indeed, it can be used to identify the major players (applicants and inventors), to deduce the relationships between them, to follow the progress of patent applications, and to define trends in technological development. The second level is concerned with the claims and the description of invention itself. The meticulous dissection of claims can provide a way to find an original and pertinent way to get round the monopoly granted for an invention (by exploiting something that is not explicit or excluded in the claims). This makes it possible to develop potential new niches for a patented technology without infringing the patent. A development niche can be defined as an opportunity that is not patented/protected, because of at least one of the following reasons:

- (1) some aspect was never disclosed and is therefore potentially still patentable (depending on the patentability conditions),
- (2) the patent has been published, but never granted (because of the patentability conditions),
- (3) the applicant decided to withdraw the patent application before it was granted,
- (4) the invention was patented, but the patent has been abandoned (as the result of a strategic choice of the applicant), and the invention is therefore in the public domain, and can be exploited in all States where patents are not kept in force,
- (5) a third party has successfully opposed a granted patent, and had it revoked, because the patent does not fulfill to the patentability conditions, i.e. it contains an element that destroys the novelty condition. This last alternative is however the one least often used, because legal cases of this type are time-consuming and very expensive.

In this article, we provide an overview of PCT applications with at least one claim involving the use of transposon tools. We have tried to define the most important

applicants and inventors, and identify their various applications. Overall, our objective was to clarify what is currently not protected by patents, and to identify potential economic niches, that are available for the development of innovative and IP-free plans involving transposon tools. We took advantage of recent Internet facilities that now make it easier to access patent information, mainly because of the availability of free databases such as esp@cenet or PatentScope.

Data sources and methods

The searches for patents involving transposon tools were mainly carried out using the PatentScope facilities (searches for full text in more than 1.5 million international patent applications), and its database release of August 2008 (ws8, Sup. Mat. 1). It provides access to published PCT international applications, and to the latest bibliographic data and documents contained in the files of PCT international applications. All patent documents were downloaded and processed using Microsoft Office software. The terms “transposon(s)” or “transposase(s)” were used as the main keywords to mine claims. Limiting ourselves to data available on August 14, 2008, five hundred patent documents were obtained. After a preliminary study (Sup. Mat. 2), we retained for the patent landscape 176 PCT patent applications (between 1986 and 2006) used to determine the worldwide landscape of the IP and technological developments of transposons. All the patent numbers cited in the text are included in Supplemental materials 2.

Moreover, a bibliographic study was carried out to overview the scientific trends in the field of transposon tools, using GoPubMed[®] facilities (ws9, Sup. Mat. 1), which makes it possible to visualize the statistics for a search (e.g. authors, journals, years, countries). Because the first query gave too many results (transposon* OR transposase* with more than 11,660 results), more targeted searches were carried out on specific transposon families: (1) *Tc1*, *SB* and *Minos* (391 results), (2) *piggyBac* (91 results), (3) *Mos1*, *Himar* and *mariner* (345 results), (4) *hobo* and *tol2* (132 results), (5) *P* elements (304 results) and (6) *Tn5* and *Tn10* (1650 results).

Patent landscape for transposon tools

Evolution of patent applications

The patent study was limited to the period 1986–2007. However, as patent applications are published 18 months after the application date, we did not in fact consider the year 2007 (only one result). Figure 3 shows the progress of

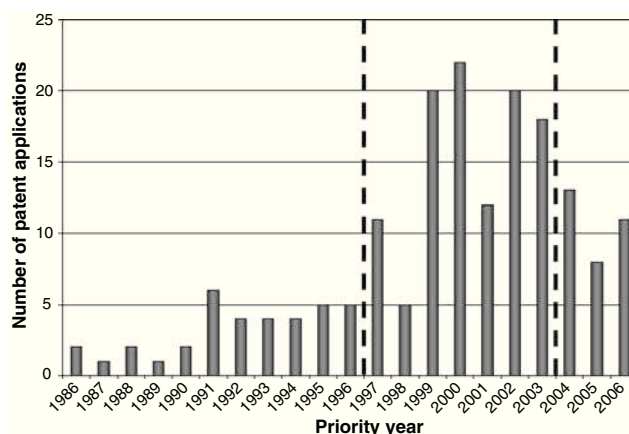


Fig. 3 Time coverage of the number of PCT applications per priority year. This diagram shows the number of PCT publications for each priority year, for the period 1986–2006, which claim at least one transposon and/or one transposase in the invention. The number of international applications increased sharply from 1999

the 176 patent applications per priority year (Sup. Mat. 2). The PCT applications submitted during the pioneering period (1986–1996) cover the use of transposons for insertional mutagenesis, immunization, gene identification, sequencing, usually for purposes that concern prokaryotic and plant cells. The first two international applications to claim a transposon as vector for integrating DNA into eukaryotic cells were WO/1995/001095–Transformed eukaryotic cells and transposon-based transformation vectors, and WO/1997/029202–Vectors and methods for providing cells with additional nucleic acid material integrated in the genome of said cells. A first peak of 11 patent publications occurred during the pivotal year 1997, which marks the beginning of the intermediate period (1997–2003), with an irregular increase in the number of submissions, with a peak of 22 PCT applications per year in 2000. It corresponds to important international applications that principally concern transposon-based gene delivery systems using integrative vectors made from derivatives of *Tc1/mariner* elements (WO/1999/007871–Vectors and methods for providing cells with additional nucleic acid material integrated in the genome of said cells; WO/1999/009817–Use of *mariner* transposon in the production of transgenic animals; WO/1998/040510–DNA-based transposon system for the introduction of nucleic acid into DNA of a cell and WO/1999/025817–*Tc1*-based transposon vectors), and *Tn7* (WO/1999/005295–Composition and method for targeted integration into cells). From 1999, the number of PCT applications increased markedly, then from 2003 fell back to the level reached in 1997. We can assume that this increase was due to the content of the applications published in 1997 that revealed that transposons can be used not only in insertional mutagenesis, sequencing, gene identification, but are also promising candidates as vectors for the delivery of

exogenous DNA into host chromosomes. Patents claim inventions that are based on transposons themselves, but also exploit the knowledge accumulated in recent years, especially in gene transfer (2004–2006). This explains why only 36 PCT applications were published during the period 1986–1996 (11 years), whereas 140 became publicly available between 1997 and 2006 (10 years).

Applicants

During the period 1986–2006, 176 PCT applications were published by 131 different applicants, 90% of whom (Fig. 4a) were based either in the US (57%), or Europe (33%). A few of these 131 applicants were particularly active, and are the major players in transposon-based technologies (Fig. 4b). Four applicants are involved in 6 PCT applications (Max-Delbrück-Centrum für Molekulare Medizin, Pioneer Hi-Bred International, Minos Biosystems, University of Minnesota), 2 applicants in 5 PCT

patent applications (Louisiana State University and Agricultural and Mechanical College, Wisconsin Alumni Research Foundation), 4 applicants in 4 PCT applications (Centre National de la Recherche Scientifique, and its co-applicant, University F. Rabelais de Tours, Johns Hopkins University, Transgenrx). Finally, 10, 12 and 99 applicants submitted 3, 2 and 1 PCT applications respectively. With regards to the co-authoring of PCT applications, 151 (85.8%) of the 176 PCT applications had only one applicant, and 25 (14.2%), two applicants. The co-applicants were based in different countries in only 16.7% of co-authored applications (DK-US, NO-GB, DE-NL, GB-NL, US-FR), the remaining 83.3% corresponding to patents with co-authors from the same country (US-US: 50%; FR-FR: 25%; NL-NL and AU-AU: 8.3%). Interestingly, whereas only 3 of the 36 (8.3%) PCT applications submitted between 1986 and 1996 were published by co-applicants, co-authoring expanded from 1997 to 2006, accounting for 21 out of 140 (15%) PCT applications.

Evolution in the number of applicants

During the studied period, comparison of the number of applicants quoted in the PCT applications (Fig. 5a, dark gray curve), and the number of new applicants (Fig. 5a, light gray curve) appearing each year in new published PCT applications revealed that the field of the transposon tools has become of increasing interest. Indeed, the average number of new applicants increased regularly until 2004. These new applicants give some idea of the number of players in the potential transposon tool market. We observe that between 1986 and 1996, most applicants were new applicants. After 1997, new applicants still appeared, and indeed doubled in number. However, it is noteworthy that the first applicants during the period 1986–1997 were still active from 1997 to 2006, both in protecting their transposon-based inventions and growing their international patent portfolio.

Comparing the shift in the absolute number of applicants (Fig. 5b, dark gray curve) and those from each nationality revealed that most of those concerned by the development of the transposon technologies for the period 1986–2001 were from the US (Fig. 5b, light gray curve). These data suggest that US applicants were less active from 2001, as they published the same number of PCT applications as other countries. A meticulous analysis of the documents published since 2001 revealed that the non-US applicants used transposon-based inventions patented before 2001 and developed by US bodies, to develop new products, medicines and tools. As a result, they actually contributed to reinforce and even to improving the transposon technology patented before 2001. If they went to use their inventions for business purposes, these non-US applicants would need

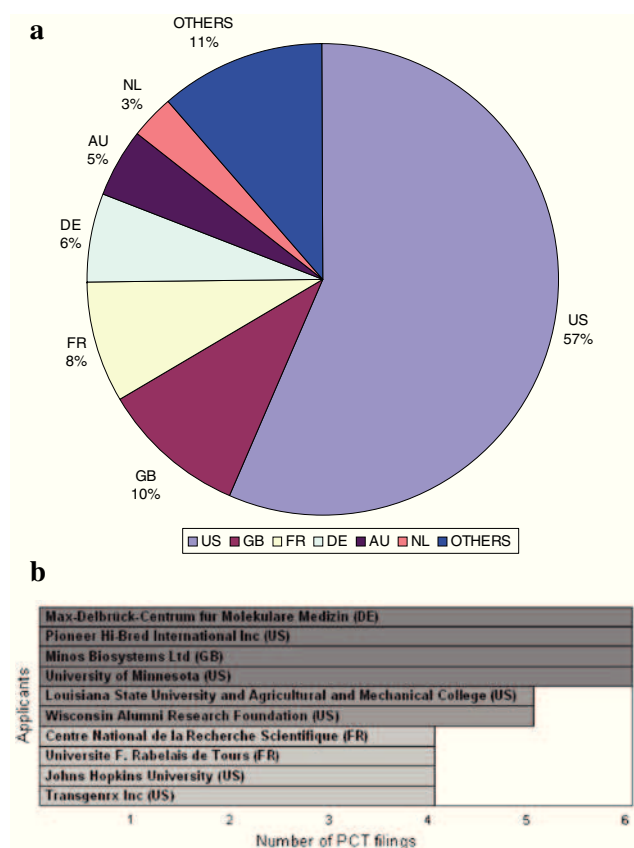


Fig. 4 **a** Nationality of the applicants. Applicants are mainly US or European. **b** The top 10 applicants. This diagram shows the number of PCT submissions submitted by the top 10 applicants, for the period 1986–2006, and claiming at least one transposon and/or one transposase in the invention

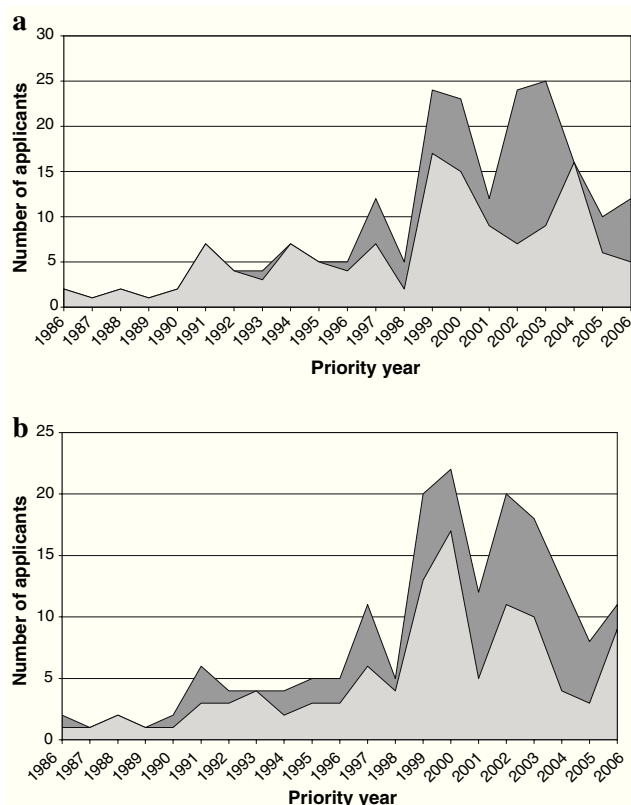


Fig. 5 **a** Change in the arrival of new PCT applicants during the period 1986–2006. The *light gray* area represents the applicants who had submitted a PCT application for the first time, defined as new applicants. **b** Time coverage of the change in the total number of applicants, and in US applicants. We can see that US applicants (*light gray*) remained the majority until 2001

to license the key patents, including the US key patents (covering the reference systems involving *Sleeping Beauty*, *P element*, *Tn10*...). From 2004, the drop in the number of applicants may be attributable to the fact that the entry barriers to the transposon tools sector are too high for new participants.

Potential fields of application for transposon tools

In our investigation of the technological fields of patents, we used the International Patent Classification (IPC) as recommended (World Intellectual Property Organization 2009). IPC is a tool primarily intended to provide “an effective search tool for the retrieval of patent documents”. It therefore offers a way to piece together the state-of-the-art in technological fields, as well as subsequently providing “a basis for the preparation of industrial property statistics which in turn permit the assessment of technological development in various areas”. The IPC is a hierarchical classification system, divided into Sections (the highest level), Classes, Sub-classes, Groups and Sub-groups (the lowest level). In this classification system, one

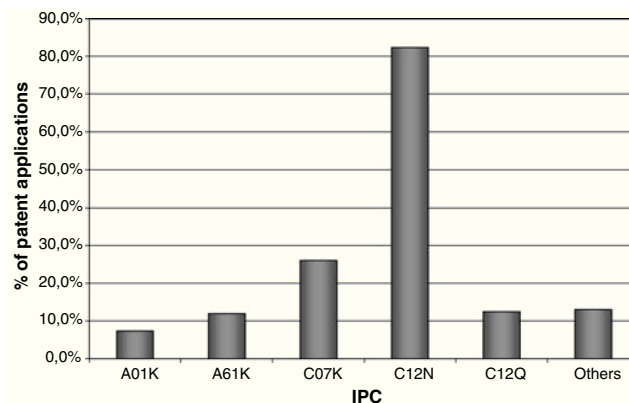


Fig. 6 Percentage share of the International Patent Classification. More than 80% of the results are international applications related to the biotechnological field (C07 K, C12 N, C12Q)

patent can have several classifications (the complete definition of IPC codes is given on the WIPO site; *ws10*, Sup. Mat. 1), and this is true for most of them. The main IPC sub-classes relevant to the potential uses of transposon tools are outlined in the Fig. 6. According to the IPC and technology concordance table, the potential uses for transposon tools are in the area of biotechnology, and mainly in three IPC sub-classes: C07 K, concerning peptides (26.1% of the patents), C12 N, concerning micro-organisms, enzyme composition, mutation or genetic engineering (82.4%), and C12Q, which involves measuring or testing processes involving enzymes or micro-organisms (12.5%). A minority of inventions concern the pharmaceutical field (A61 K, 11.9%), and animal husbandry (A01 K, 7.4%). The other IPC codes (A01H, A01 N, C02F, C07H, C12P, G01 N-033, G06F, accounting for 13.1% of the PCT applications; not detailed here) included fewer than 6 international patent applications.

The over time analysis of the number of PCT applications in each class (Fig. 7) revealed that A01 K, A61 K, C07 K, C12Q and other sub-classes displayed similar irregular changes between 1986–2006. In contrast, the progression of the C12 N sub-class can be correlated to that of PCT applications per priority year (Fig. 3). Since the C12 N sub-class comprised most of the 176 PCT applications (82.4%), with its main group, C12N15, representing 77.3% of them, we further focus our analysis on this sub-class. The IPC defines the C12N15 group as “Mutation or genetic engineering; DNA or RNA genetic engineering of vectors, or their isolation, preparation or purification; use of hosts therefore (mutants or genetically engineered micro-organisms; new plants; plant reproduction by tissue culture techniques; new animals; use of medicinal preparations containing genetic material which is inserted into cells of the living body to treat genetic diseases, gene therapy; peptides in general)”. Its importance in our results therefore confirms

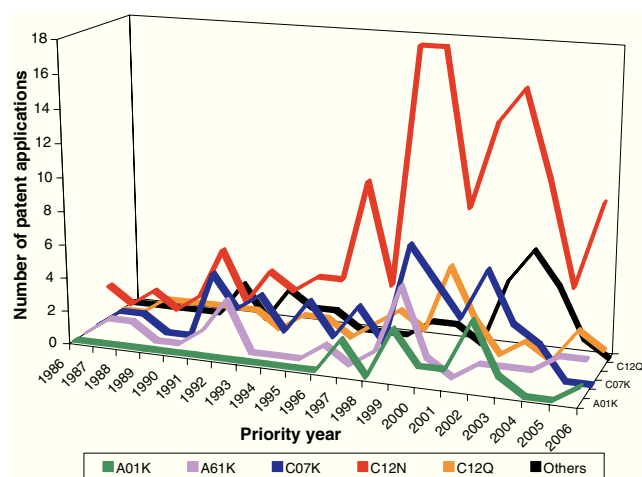


Fig. 7 Change in the number of submissions per priority year for the main sub-classes of the 176 PCT patents. We can see that there was an increase in the number of submissions from 1997, in the major sub-class C12 N, and in other minor and diversified sub-classes

that transposon tools are, from an IP standpoint, viewed as biotechnology tools mainly used for mutagenesis and genetic purposes.

Application targets of transposon-based inventions

In this section, we consider only PCT patent applications that refer to at least one transposon or transposase in the first claim (109 out of 176 PCT publications). A search for the applications of each invention allowed us to construct the tree of their uses (Fig. 8). Applications include molecular biology tools (e.g. gene tagging, insertional mutagenesis, gene identification, sequencing), and engineering (some patents involve targeting and insulation). The most frequent use is gene delivery for the production of transgenic invertebrates or vertebrates, and similar applications specifically for plants and bacteria, even though these latter are less often in the limelight.

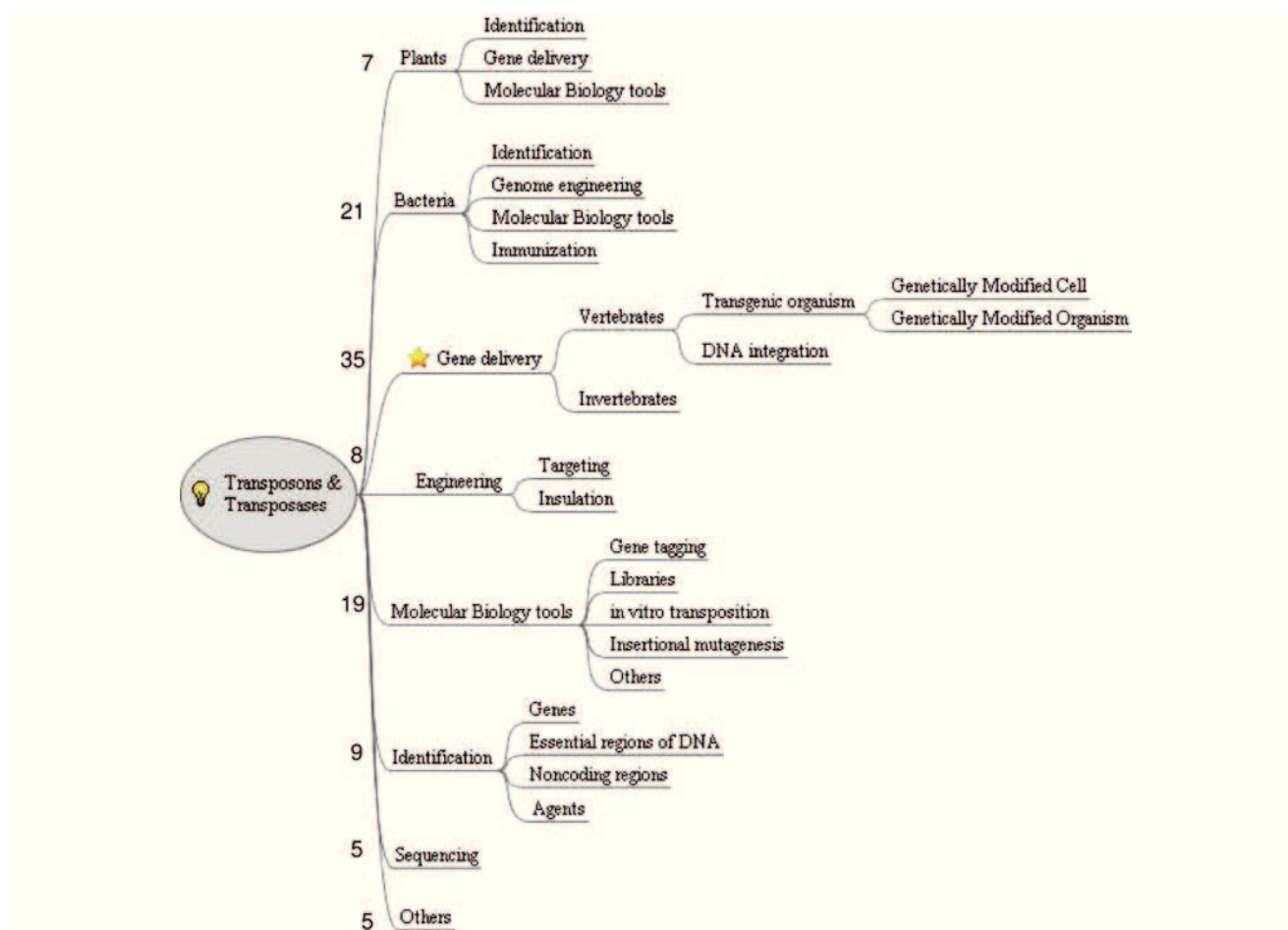


Fig. 8 The tree of application domains for transposon tools. Applications of transposon-based inventions include specific patents related to plants, bacteria, gene delivery, engineering, molecular biology tools, identification, sequencing and others

Table 1 Legal status of patent families, relative to the proof of concept in use and in engineering for specific transposon families (updated in October 09)

Transposon systems	PCT applications with issued patents in several states	PCT applications with EP applications under examination	Other publications
<i>Tn5</i>	WO/2000/017343 WO/1998/010077	/	WO/2004/093645 WO/2002/046444 WO/2001/009363
<i>Tn10</i>	WO/2004/067707 WO/2004/003157	/	/
<i>Sleeping Beauty</i>	WO/2001/081565 WO/1998/040510	/	WO/2003/089618 WO/2001/030965 WO/2000/068399 WO/1999/025817 WO/2007/082164
<i>Tol2</i>	WO/2003/068960* WO/2001/040477*	/	WO/2003/097826
<i>Minos</i>	WO/2003/056912 WO/2001/071019 WO/2001/044483	/	
<i>Mos1</i>	/	WO/2007/132096 WO/2007/063033 WO/2004/078981	/
<i>Mariner</i>	WO/2006/108525	/	WO/2004/078962 WO/1999/009817 WO/2000/065042 WO/2001/014537 WO/2008/098181 WO/2008/027384 WO/2007/100821 WO/2000/055346 WO/1999/050402
<i>p element</i>	WO/2003/027241	/	
<i>piggyBac</i>	/	WO/2006/122442	
<i>Himar</i>	/	/	WO/2001/029205 WO/1999/007871 WO/1997/029202
<i>Hobo</i>	/	/	
<i>Tc1</i>	/	/	

* Patents resulting from a previous study

The results show that there are granted patents, those currently undergoing examination, and other publications, including applications deemed to have been withdrawn and those that have been recently published

Patent landscape for specific transposon families

Among the existing transposons, we focused more particularly on certain members of the following families: *Tn5*, *Tn10*, *Tc1*, *Sleeping Beauty*, *Minos*, *piggyBac*, *mariner*, *Mos1*, *Himar*, *Hobo*, *P element*. Table 1 summarizes the legal status of the members of the patent families of the key PCT applications. They are relative to the proof of concept in use and in engineering, that belong to the C12N15 IPCs, that claim specifically or preferentially a particular transposon (Sup. Mat. 3). There are four different statutes: on the one hand (1) granted patents, and on the other hand the applications which are not issued: (2) publications that have been published (3), those currently undergoing examination, and (4) those deemed to have been withdrawn.

These findings confirm those of a previous study of the families of published patents concerning transposases and transposons (Palazzoli et al. 2008). The only PCT

applications that were not found by our patent search were about *Tol2* (WO/2003/068960 and WO/2001/040477); this was because the claims are in Japanese, and there is no English translation in the databases. Patents were granted in several states and involve *Tn5* (WO/2000/017343 and WO/1998/010077), *Tn10* (WO/2004/067707 and WO/2004/003157), *SB* (WO/2001/081565, WO/1998/040510), *Tol2* (WO/2003/068960 and WO/2001/040477), *Minos* (WO/2003/056912, WO/2001/071019 and WO/2001/044483), *Hsmar* (WO/2006/108525), and *P element* (WO/2003/027241). Other granted patents deal with only one or two states : for *Tn5* (WO/2004/093645, WO/2002/046444 and WO/2001/009363 for the US and AU), *SB* (WO/2001/030965 just for the US), *Minos* (WO/2001/044483 for states that are members of the African Intellectual Property Organization), *Mos1* (WO/2007/063033 and WO/2004/078981 for the FR), *Mariner* (WO/2004/078962), and *P element* (WO/2000/065042), *piggyBac* (WO/2001/014537),

Himar (WO/1999/050402), and *Tc1* (WO/1997/029202). However, some of the PCT applications notified in databases as being currently under examination, specially EP applications, and also need to be taken into account as important documents for *Mos1* (WO/2007/132096, WO/2007/063033 and WO/2004/078981), and *piggyBac* (WO/2006/122442). The data content also identified the main applicants with patents involving these prokaryotic or eukaryotic transposon systems, whatever their origins.

Prokaryotic transposon systems

The Wisconsin Alumni Research Foundation is the main applicant for the technologies developed from *Tn5*. Its portfolio is based on two main patent families including granted patents (WO/1998/010077 and WO/2000/017343), and three other PCT applications (WO/2001/009363, WO/2002/046444 and WO/2004/093645). *Tn5* technologies were mainly developed for in vitro transposon insertion purposes (ws11 and ws12, Sup. Mat. 1). The Epicentre website indicates that the company possesses licenses covering *Tn5* for the patents EP0927258 and US5948622 (the same family of WO/1998/010077), and “other patents issued or pending, exclusively licensed or assigned to Epicentre” (members of the family of WO/2000/017343, as US6159736 or EP1115856). This gives Epicentre exclusive rights to commercialize tool kits involving *Tn5*.

Two other co-authoring applicants in the top 10 (Fig. 4b) are the Transgenrx Company and the Louisiana State University and Agricultural and Mechanical College. Transgenrx can be expected to obtain licenses for the commercial exploitation of patents of the families of WO/2004/067707 and WO/2004/003157. They concern the preferential use of *Tn10* for gene delivery purposes in vertebrates (ws13, Sup. Mat. 1). These co-applicants have also tried to reinforce their position with two other patents concerning the therapeutic use of transposons (WO/2004/067743 and WO/2005/062881, not in the Table 1). The business model of Transgenrx is the production of protein-based drugs, with a market estimated to have reached \$70 billion by 2008 (ws14, Sup. Mat. 1).

An important applicant that is not currently included in the Top 10 is Finnzymes, which designs kits for the purposes of DNA sequencing and functional analysis of proteins, using the *Mu* transposon (ws15, Sup. Mat. 1).

Eukaryotic transposon systems

The main system for *SB* is WO/1998/040510, but the University of Minnesota has also tried to strengthen its portfolio with a series of international applications (WO/1999/025817, WO/2000/068399, and WO/2003/089618). It can be reasonably supposed that members of the PCT

application WO/1998/040510, which belong to the University of Minnesota, have actually been licensed by Discovery Genomics, not only because of the company founders are also the inventors, but also because patents US6489458 and WO/1998/040510 are in the same family (ws16, Sup. Mat. 1). In addition to the University of Minnesota's patents, Discovery Genomics has tried to reinforce its position with regard to *SB* by including non-transposon technologies, such as insulation in gene therapy contexts (the PCT application WO/2004/065581). Discovery Genomics is therefore an University of Minnesota spin-off company, that exploits the *SB*-based technology developed in the public research laboratory for purposes ranging from fundamental research to clinical trials (ws17, Sup. Mat. 1).

The status of the inventions developed from the *Tol2* transposon remains to be elucidated. Indeed, we found no license or company exploiting granted patents that are members of the two families of WO/2003/068960 and WO/2001/040477. This might be considered as surprising since *Tol2* is the most efficient transposon system in terms of cargo capacities (Balciunas et al. 2006). Interestingly, the capacities of *Tol2* vectors are being investigated by the research unit of S McIvor, P Hackett, N Somia and D Largaespada at the Center for Genome Engineering at the University of Minnesota (formerly known as the Beckman Center for Transposon Research; ws18, Sup. Mat. 1). They have developed “methods to more effectively insert therapeutic genes into chromosomes”, using the *SB* and *Tol2* transposons, including, for example, the development of the mini*Tol2* vector (Balciunas et al. 2006).

Concerning *Minos*, the applicant for the granted patents members of the two families WO/2003/056912 and WO/2001/071019 is Minos Biosystems. The company has also submitted several other PCT patent applications involving *Tc1-mariner* transposons: WO/2001/029205, including molecular biology tools: WO/2003/097826 and WO/2002/062991, and technologies using transposon: WO/2004/013171. Minos Biosystems, which is one of the top 10 applicants, is a company that exploits *Minos* transposon-based technology for the *in-mass* production of proteins and gene discovery (ws19, Sup. Mat. 1).

The PCT applications covering *Mos1* (WO/2007/132096, WO/2007/063033, and WO/2004/078981 (all corresponding EP applications are currently undergoing examination) belong to two co-applicants, the Centre National de la Recherche Scientifique and the University F. Rabelais of Tours. For the moment, these PCT applications have not been licensed by any company, and in consequence are not included in any economic project. The lack of activity of this system in mammal cells (Wu et al. 2006; Germon et al. 2009) means that new fields of application will have to be identified, for instance in nematodes and protozoons, in which other transposon systems have little or no activity.

The Tosk Company exploits the *P element* transposon under its granted patents of the two families WO/2003/027241 and WO/2000/065042, jointly with Leland Stanford Junior University, which has granted licenses to Tosk for several technological platforms including a DNA vector (ws20, Sup. Mat. 1). Here too, both applicants share the same inventor (P. Fogarty). Tosk was founded in 1998, and is developing its business in the field of gene delivery for drug discovery purposes (ws21, Sup. Mat. 1).

The case of the piggyBac system

To date, *piggyBac*, initially named in the scientific literature *IFP2* (Cary et al. 1989), appears to be the only element among the most efficient transposon systems for gene delivery not to be protected by issued patents except two US patents, in spite of five PCT applications (Table 1). However, this situation may be the result of an oral disclosure of the IFP2 transposon's usefulness for carrying active genes in infected insect cells (communication at the First International Workshop on Transgenesis of Invertebrates of medical, agricultural and aquacultural importance, in 1995 (Sup. Mat. 4)). Indeed, this disclosure, occurring several years before the filing of the first patent application known to us (WO 2001/014537, and the two other members of the family US7005296 and US6773914, with a priority in 1999), prevents patentability for general use in gene delivery. Nevertheless, several new PCT applications about *piggyBac* tools have been subsequently filed by US applicants such as the University of Hawai'i (WO/2008/027384), the University Notre Dame (WO/2007/100821), and the University of Utah research foundation (WO/2008/098181). Due to the weakening of any new patent that results from the disclosure in 1995, which means that the use of *piggyBac* as a transposon tool is not a patentable invention, it is reasonable to suppose that the existence of these patents might just be intended to raise funds to develop a company, such as Manoa Biosciences. Founded in 2006, Manoa is a spin-off from the University of Hawaii (ws22 and ws23, Sup. Mat. 1), and is interested in the non-viral horizontal gene insertion for animal transgenesis and gene therapy purposes. The freedom-to-operate with *piggyBac* has, for example, been exploited by the Oxitec Company, which develops its business models using genetics (WO/2005/003364) and molecular biology (RIDL technology; Alphey et al. 2008) to improve significantly the cost-effectiveness and safety of the Sterile Insect Technique, and to extend it to a broader range of insect pests.

One of our primary objectives was to find out whether there was freedom-to-operate for transposons, in particular for gene transfer purposes. Our analyses have revealed that the patents covering the *Tc1*-like elements (of the following families WO/2001/029205, WO/1999/007871, WO/1997/029202) and *mariner*-like elements (WO/1999/009817) are

all deemed to be withdrawn, which is what we expected in the light of the knowledge accumulated over a decade which has shown that these two kinds of transposons are now considered to be separate elements. No patent data concerning use as gene delivery vectors was found for the *Hobo* and *Himar* elements (PCT applications: WO/2000/055346 and WO/1999/050402, with just one US patent, confirmed by a previous study, Palazzoli et al. 2008). This implies that their exploitation may be free, either in all countries if the two patents are deemed to be withdrawn, or otherwise in countries other than US or CA. In conclusion, the only transposons that seems to be IP-free for gene delivery purposes are *Himar1*, *piggyBac* and, of course any new transposon system that has not been so far disclosed.

Landscape of the scientific literature for transposon tools

In an attempt to investigate the relationships between the patent applicants who are developing and exploiting transposon tool technologies and the scientific community, we carried out a survey of the scientific literature in this field. Searches were carried out for (1) *Tc1*, *SB* and *Minos* (391 results), (2) *piggyBac* (91 results), (3) *Mos1*, *Himar* and *mariner* (345 results), (4) *hobo* and *Tol2* (132 results), (5) *P element* (304 results) and (6) *Tn5* and *Tn10* (1650 results). Each bibliographic search yielded the major authors corresponding to the transposon system: R. Plasterk (*Tc1*), Z. Ivics and P. Hackett (*SB*), M. Fraser and M. Handler (*piggyBac*), D. Hartl and Y. Bigot, (*mariner*), K. Kawakami (*Tol2*), D. Rio (*p element*), and W. Reznikoff (*Tn5*). Interestingly, these authors, who are among the 20 most frequently published, are also the inventors behind the top 10 applicants: e.g. W. Reznikoff (Wisconsin Alumni Research Foundation), P. Hackett (University of Minnesota), Z. Ivics and Z. Izsvák (Max-Delbrück-Centrum für Molekulare Medizin; Sup. Mat. 5 versus Fig. 4b).

An analysis of the GoPubMed results revealed statistics about the author collaboration (i.e. co-authoring) networks involved in the initial searches (from 1 to 6, and even more collaborations, Fig. 9). For example, the collaboration networks for authors who have worked and published together on *Tc1*-like elements revealed that the people working on *Impala*, *Minos* and *SB* are not directly collaborating. Two small and independent scientific networks can be identified around *Minos* studies, whereas *SB* studies take advantage of a large and very active publishing network that has its epicentre among scientists at Minnesota University and Discovery Genomics (Fig. 9a, Authors highlighted by red ellipses). This network of scientists working on *SB* seems to be coordinated from its epicentre, and contributes to add to knowledge and know-how about this system (Table 1, WO/2001/030965, with M. Kay and S. Yant as inventors),

a

Minos

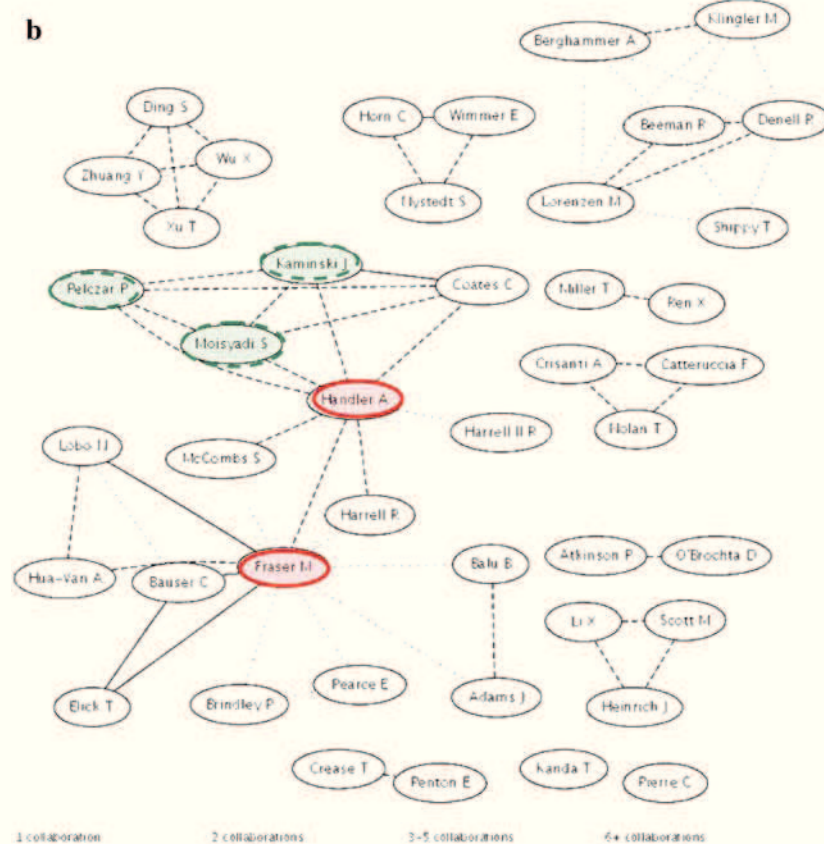
Tc1 SB

1 collaboration

2 collaborations

3-5 collaborations

6+ collaborations



thus providing the background that is vital for the development of an economic developmental plan involving this technology. Interestingly, Z. Ivics and Z. Izsvák, who are historically the two most innovative scientific pioneers in this technology, appear to have been marginalized in this network. This was an unexpected finding, as these authors are the inventors in the active patents, WO/1998/040510 from the University of Minnesota, and WO/2001/081565 that belongs to the Max-Delbrück-Centrum für Molekulare Medizin. Incidentally, rather surprisingly, WO/2001/081565 has been granted, even though the patentability conditions for the inventions it describes are not fulfilled since they had already been included in WO/1998/040510. The existence of WO/1998/040510 was known when the WO/2001/081565 PCT examination procedures were launched. This situation raises questions about the reliability of the PCT examination procedures, and is a source of potential conflict of interests if *SB* technology turns out to be economically viable.

GoPubMed results (Sup. Mat. 6) have revealed that networks similarly organized around an epicentre corresponding to the owner of the technology also exist for *Tol2* and *Tn5* technologies, even though the resulting links are less developed. They have also revealed that networks involving the *piggyBac* (Fig. 9b), *mariner* and *P* elements (Sup. Mat. 6) are organized differently, and have no single epicentre. The inventors of *piggyBac* system, M. Fraser and M. Handler collaborated with other authors, including the emergent inventors (WO/2008/051620 and WO/2008/027384 published before 2008.08.14 and WO/2008/100424 and WO/2008/137114 published after 2008.08.14, and so not included in the study) as J. Kaminski, P. Pelczar and S. Moisyadi who are related to Manoa Biosciences. It is noteworthy that M. Fraser is member of the Scientific Advisory Board of the company, founded in 2006 (ws24, Sup. Mat. 1), and that J. Kaminski, who was linked to the TranspoVec company (specialized in chimeric transposases, and insertion enzymes, and for which the domain name of the website has expired or been deleted) is a joint founder of this company with S Moisyadi.

Concluding remarks

The year 1997 marked a significant turning point in transposon-based technology. From this date, the number of PCT applications increased sharply during the 2000s, mainly for gene delivery purposes. The main authors of scientific publications about each of the transposon systems were also the pioneers of the technology based on the ability to introduce DNA into the host genome. Most of the international patents involving emergent technologies were submitted by public bodies. Indeed, the scientists who developed transposon-based technologies in University

laboratories, at the end of the 1990s, were also those who wrote the patents concerning the use of transposon as gene delivery vectors. These patents cover the proof of concept in use and in engineering for specific transposon families. The exploitation of these technologies was granted to spin-off biotech companies, through exclusive licenses, by the universities where they first emerged. Most of these companies were established in the 2000s, mostly in the US in the case of business plans concerning therapeutic purposes (Discovery Genomics, Manoa Biosciences, Transgenrx, Tosk), and mainly European for non-therapeutic, transposition-based applications in mutagenesis (Minos Biosystems, Finnzymes or even Oxitec). Currently, most of the transposon systems are protected by patents, including PCT applications for the most important ones. It is important to remember that licenses can be granted not only for granted patents, but also for pending patents (ws12, Sup. Mat. 1). These licenses are used to sustain a strategic position in a market where several transposon systems are already present, and for the business development of competitors. In such a competitive situation, the IP is crucial in order to develop a strong development plan, for example including clinical trials of medicines derived from biotechnology (defined as “innovative therapies”). From this standpoint, it can be expected that, when the race to produce the most efficient transposon system is over, a new competition will start between the different US companies involving clinical trials in an environment where research and development will be increasingly intertwined.

One of the main potential development niches that is currently being investigated is that of the targeted integration of transposon vector into the host genome. This is the approach currently followed by Z. Ivics and J. Kaminski, both of them are developing, with their respective colleagues, targeted systems and fusion proteins (PCT applications WO/2004/070042, WO/2004/069995 and WO/2004/069994, WO/2004/009792 and WO/2008/137114). However, they are not the first to work in this field, and this will very probably limit the interest of patenting in this field of investigation. PCT application WO/2002/008286 already proposed to fuse a transposase with a DBD, such as ZFD. The recent literature however has concluded that this approach constitutes a monopoly severely restricting access to the Zinc Finger technology, and for potential commercial R&D plans (Lyznik and Dress 2008; Chandrasekharan et al. 2009). For the moment, the best solution, an IP-free system, is probably provided by the PCT patent WO/1994/024300 (Transposition assembly for gene transfer in eukaryotes—applicant: Transgene SA, inventor: E. Jacobs, not included in the results), with two granted patents (FR2703996 and US6346414). This patent claimed a specific way to integrate vectors into rRNA genes, using tools derived from a non-LTR retrotransposon. Using an IP-free transposon, such as *piggyBac* for instance, such a targeting system might

provide a way of developing an exploitation plan in this field, without being blocked by IP rights.

The case of *piggyBac* is interesting for another reason. Indeed, it highlights the difficulties involved in carrying out an anteriority study of the state of the art (by patent examiners or others). The disclosure of a transposon-based invention is not included in the various publications (articles and patent applications), available in databases. However, if a patent has been granted, the person who would have this disclosure information, has to formulate an opposition in order to have the patent revoked.

As we have already mentioned, patent information is made available to the public in several patent databases. Many national and regional offices provide a way to carry out free patent searches. Each database covers specific patent information (e.g. the USPTO for the US patent applications), but with variable availability (see the various notices concerning patent searches). Database providers propose patent searching in databases with a real added value. For example, a full text search coverage, including description and claims, is available for the most important patent databases, such as the Questel QPAT database (ws25, Sup. Mat. 1). Complementary tools are available for performing quick and detailed analyses of patent results. These commercial tools provide maps of priority submissions, the segmentation profile of a portfolio, the major pioneer applicants, the collaborations between inventors and applicants. For legal matters, it is useful and advisable to consult one of these commercial patent databases. However, it is important to bear in mind that no data source covers all the available patent information, let alone all the available technological information. Consequently, a thorough knowledge of the technological field, including, for example, the requisite technical vocabulary, keywords, major authors, is advisable for any attempt to carry out judicious searches of patent documents, articles and congress abstracts.

Acknowledgments This contribution was supported by the University of François Rabelais of Tours and funded by grants from the European Commission (Project SyntheGeneDelivery, N°018716), the C.N.R.S., the Ministère de l'Éducation Nationale, de la Recherche et de la Technologie, the Association Française contre la Myopathie and the Groupement de Recherche CNRS 2157. Fabien Palazzoli holds a doctoral fellowship from the Ministère de l'Éducation Nationale, de la Recherche et de la Technologie.

References

- Alphey L, Nimmo D, O'Connell S, Alphey N (2008) Insect population suppression using engineered insects. *Adv Exp Med Biol* 627:93–103
- Balciunas D, Wangenstein KJ, Wilber A et al (2006) Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. *PLoS Genet* 2:e169
- Barabas O, Ronning DR, Guynet C, Hickman AB, Ton-Hoang B, Chandler M, Dyda F (2008) Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection. *Cell* 132:208–220
- Barrett WA (2003) The patent gamble: strategic insights for playing the worldwide patent game. *Nat Biotech* 21:1515–1517
- Baus J, Liu L, Heggstad AD et al (2005) Hyperactive transposase mutants of the *sleeping beauty* transposon. *Mol Ther* 12:1148–1156
- Beerli RR, Barbas CF III (2002) Engineering polydactyl zinc-finger transcription factors. *Nat Biotech* 20:135–141
- Bergman K, Graff GD (2007) The global stem cell patent landscape: implications for efficient technology transfer and commercial development. *Nat Biotechnol* 25:419–424
- Cary LC, Goebel M, Corsaro BG, Wang HG, Rosen E, Fraser MJ (1989) Transposon mutagenesis of baculoviruses: analysis of *Trichoplusia ni* transposon *IFP2* insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology* 172:156–169
- Chandrasekharan S, Kumar S, Valley CM et al (2009) Proprietary science, open science and the role of patent disclosure: the case of zinc-finger proteins. *Nat Biotech* 27:140–144
- Denker HW (2008) Totipotency/pluripotency and patentability. *Stem Cells* 26:1656–1657
- European Commission (DG Research) and European Patent Office (2007) Why researchers should care about patents? http://ec.europa.eu/invest-in-research/pdf/download_en/patents_for_researchers.pdf
- Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* 41:331–368
- Gamsjaeger R, Liew CK, Loughlin FE et al (2007) Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem Sci* 32:63–70
- Germon S, Casteret S, Carpentier G, Bouchet N, Adet J, Bigot Y, Augé-Gouillou C (2009) Transposase optimization of the *Mos1 mariner* transposon by rational mutagenesis. *Genetica*, Accepted
- Ivics Z, Hackett PB, Plasterk RH et al (1997) Molecular reconstruction of *Sleeping Beauty*, a *Tc1*-like transposon from fish, and its transposition in human cells. *Cell* 91:501–510
- Ivics Z, Katzer A, Stüwe EE et al (2007) Targeted *Sleeping Beauty* transposition in human cells. *Mol Ther* 15:1137–1144
- Izsvák Z, Ivics Z, Plasterk RH (2000) *Sleeping Beauty*, a wide host-range transposon vector for genetic transformation in vertebrates. *J Mol Biol* 302:93–102
- Lampe DJ (2009) Bacterial genetic methods to explore the biology of *mariner* transposons. *Genetica* (this issue)
- Lampe DJ, Churchill ME, Robertson HM (1996) A purified *mariner* transposase is sufficient to mediate transposition in vitro. *EMBO J* 15:5470–5479
- Lampe DJ, Akerley BJ, Rubin EJ, Mekalanos JJ, Robertson HM (1999) Hyperactive transposase mutants of the *Himar1 mariner* transposon. *Proc Natl Acad Sci U S A* 96:11428–11433
- Lyznik LA, Dress V (2008) Gene targeting for chromosome engineering applications in eukaryotic cells. *Recent Pat Biotechnol* 2:94–106
- Miskey C, Izsvák Z, Ronald H et al (2003) The *Frog Prince*: a reconstructed transposon from *Rana pipiens* with high transpositional activity in vertebrate cells. *Nucl Acids Res* 31:6873–6881
- Palazzoli F, Carnus E, Wells DJ et al (2008) Sustained transgene expression using non-viral enzymatic systems for stable chromosomal integration. *Cur Gene Ther* 8:367–390
- Papworth M, Kolasinska P, Minczuk M (2006) Designer zinc-finger proteins and their applications. *Gene* 366:27–38
- Sander JD, Zaback P, Joung JK et al (2007) Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucl Acids Res* 35:W599–W605

- Spradling AC, Rubin GM (1982) Transposition of cloned *P elements* into *Drosophila* germ line chromosomes. *Science* 218:341–347
- Vrtovec KT, Scott CT (2008) Patenting pluripotency: the next battle for stem cell intellectual property. *Nat Biotechnol* 26:393–395
- Vrtovec KT, Vrtovec B (2007) Commentary: is totipotency of a human cell a sufficient reason to exclude its patentability under the European law? *Stem Cells* 25:3026–3028
- Williams DA (2008) *Sleeping Beauty* vector system moves toward human trials in the United States. *Mol Ther* 16:1515–1516
- Wilson MH, Kaminski JM, George AL Jr (2005) Functional zinc finger/*Sleeping Beauty* transposase chimeras exhibit attenuated overproduction inhibition. *FEBS Lett* 579:6205–6209
- World Intellectual Property Organization (2008) Guide to Using Patent Information, pp 44
- World Intellectual Property Organization (2009) Guide to the international patent classification, version of 2009, pp 40
- Wu SC, Meir YJ, Coates CJ, Handler AM, Pelczar P, Moisyadi S, Kaminski JM (2006) *piggyBac* is a flexible and highly active transposon as compared to *Sleeping Beauty*, *Tol2*, and *Mos1* in mammalian cells. *Proc Natl Acad Sci U S A* 103:15008–15013
- Yant SR, Meuse L, Chiu W et al (2000) Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nat Genet* 25:35–41
- Yant SR, Huang Y, Akache B et al (2007) Site-directed transposon integration in human cells. *Nucl Acids Res* 35:e50

Transposon tools: worldwide landscape of intellectual property and technological developments

Fabien Palazzoli, François-Xavier Testu, Franck Merly and Yves Bigot

Electronic Supplementary Material 1: Websites cited in the core text

- [ws1] http://www.wipo.int/patentscope/en/patents_faq.html
- [ws2] http://www.wipo.int/sme/en/documents/drafting_patent_claims.htm
- [ws3] http://www.wipo.int/sme/en/documents/patents_biotech.htm
- [ws4] <http://www.frost.com/prod/servlet/report-brochure.pag?id=F300-01-00-00-00>
- [ws5] <http://www.bentham.org/Titlesbydiscipline.htm>
- [ws6] http://www.wipo.int/sme/en/ip_business/patents/advantages.htm
- [ws7] http://www.wipo.int/wipo_magazine/en/2008/04/article_0005.html
- [ws8] <http://www.wipo.int/pctdb/en/index.jsp>
- [ws9] <http://www.gpubmed.com/>
- [ws10] <http://www.wipo.int/classifications/ipc/ipc8/?lang=en>
- [ws11] <http://www.epibio.com/item.asp?ID=291>
- [ws12] <http://www.epibio.com/terms.asp>
- [ws13] <http://www.tgrxinc.com/>
- [ws14] <http://www.tgrxinc.com/business.htm>
- [ws15] http://www.finnzymes.com/sequencing_mutagenesis/sequencing_and_mutagenesis.html
- [ws16] <http://www.discoverygenomics.net/sbts.html>
- [ws17] http://www.discoverygenomics.net/clinical_fa.html
- [ws18] <http://www.cge.umn.edu/about.htm>
- [ws19] <http://www.minosbiosystems.com/business.htm>
- [ws20] <http://www.allbusiness.com/medicine-health/diseases-disorders-neurological/6696658-1.html>
- [ws21] <http://www.tosk.com/company.htm>
- [ws22] <http://manoabiosciences.com/Default.aspx?base>
- [ws23] <http://www.manoabiosciences.com/TechnologyScience/IntellectualProperty/tabid/69/Default.aspx>
- [ws24] <http://www.manoabiosciences.com/CorporateOverview/ScientificAdvisoryBoard/tabid/65/Default.aspx>
- [ws25] <http://www.questel.com/en/Prodsandservices/QPAT.htm>
- [ws26] http://ep.espacenet.com/advancedSearch?locale=en_EP

[ws27] <http://www.wipo.int/patentscope/en/data/terms.html>

Transposon tools: worldwide landscape of intellectual property and technological developments

Fabien Palazzoli, François-Xavier Testu, Franck Merly and Yves Bigot

Electronic Supplementary Material 

Patent documents used for the study

Details about the data sources and methods

Patent database

We opted for the PatentScope search service rather than that of esp@cenet, because it allows one to do searches within claims, which are the most informative sections in patents, as they specify the limits of patent protection. In contrast, esp@cenet only proposes searches for keywords in titles or abstracts ([ws26](#), [Sup. Mat. 1](#)). For example, the PatentScope result WO/2006/060314 does not contain the terms "transposase(s)" or "transposon(s)" in its title or abstract, even though a transposon is a key element in its first claim. PatentScope also offers complementary facilities, such as graphic tools representing the general term distribution of the keywords used, the citation number and the distribution of terms in the claims. However, searching in PCT applications only comes up with patent applications submitted in more than one state covered by the PCT. This implies that if there has been only a single national patent application, it will not appear. As this article focuses on the worldwide aspect of IP, we considered that a single national patent could be ignored, as such single national patents are a lot less restrictive than PCT patents.

The PatentScope website provides important information regarding the terms and conditions of use for PatentScope service ([ws27](#), [Sup. Mat. 1](#)), particularly about availability and liability. The description and claims that are available on PatentScope Search Service are supplied in text format, and were recovered by applying automatic Optical Character Recognition procedures to the scanned images of the documents. They therefore contain some discrepancies from the originals, and have no legal value. For example, the patent search came up with no data for the *Tn10* transposon. This could suggest that the automatic Optical Character Recognition process did not correctly identify "Tn10", but anomalies such as "TnIO" or "Tn/O", as a meticulous dissection of the claims bears out. This means that PatentScope documents, and indeed the present article, engage neither the examiners nor the patent offices.

Preliminary study

The most recent document was PCT application WO/2008/098181 (Publication Date: 14.AUG.2008), and the oldest one was PCT application WO/1982/003087 (Publication Date: 16.SEP.1982). A screening was carried out to eliminate any documents that did not contain or involve the keywords "transposase" or "transposon" in the claims, or which excluded their use

in the invention. In consequence, results were set aside, for one or more of the following reasons. First, in a few patent applications, the keywords "transposon" or "transposase" were not in fact cited in the claims, but only in the description and had been inappropriately incorporated into the claims (e.g. WO/2001/07077). Second, the use of a "transposon" or a "transposase", in the context of the invention was in fact excluded by the inventors (e.g. WO/1999/01051). Third, patent applications claiming a retrotransposon, Class-I TE (e.g. WO/2001/081598) were set aside, since in this article we are only analyzing Class-II TEs (except *Helitrons*). Fourth, if a transposon is the specific target of the invention (e.g. WO/2001/063540, WO/2001/089579), or one target among others (e.g. WO/2005/017205), as in the case of computer detection tools or chemical inhibitors. Fifth, if the invention did not directly concern a transposase or a transposon, but a gene derived from a transposon (e.g. WO/1989/005345), or a transposon site (e.g. WO/2007/022623). For many patent applications, a transposon or a transposase is one way among others to obtain mutagenesis (a gene is altered either by random mutagenesis, or site-directed mutagenesis; e.g. WO/1997/048719) or vectorizing exogenous DNA (the vector can be a plasmid, a virus, a pro-virus, a BAC, a YAC, a transposon, a bacteriophage, or a phagemid, e.g. WO/2001/073000). In general, this situation was encountered in sub-claims that depended on more general claims enumerating technical possibilities. Overall, all patents that did not claim at least one transposase or transposon as a key element of the invention were discarded.

Patent documents used for the study

NUMBER	PRIORITY	IPC	TITLE
WO/2008/098181	2007	C12N 15/74	IN VIVO GENOME-WIDE MUTAGENESIS
WO/2008/082429	2006	C12N 15/09	PLANT-SPECIFIC GENETIC ELEMENTS AND TRANSFER CASSETTES FOR PLANT TRANSFORMATION
WO/2008/079608	2006	C12N 15/63, C12N 15/11	COMPOSITIONS AND METHODS FOR THE EXPRESSION OF NUCLEIC ACIDS
WO/2008/054945	2006	C12N 9/10	ENDURACIDIN BIOSYNTHETIC GENE CLUSTER FROM STREPTOMYCES FUNGICIDICUS
WO/2008/049849	2006	C12N 15/82, C12N 15/01, C12N 15/29, C12N 15/87	MAIZE WITH GOOD DIGESTIBILITY AND DISEASE RESISTANT

WO/2008/027384	2006	C12N 15/06, C12N 15/09	METHODS AND COMPOSITIONS FOR TRANSPOSON-MEDIATED TRANSGENESIS
WO/2007/143681	2006	A61K 48/00, C12N 15/86	HELPER VIRUS-FREE HERPESVIRUS AMPLICON PARTICLES AND USES THEREOF
WO/2007/132096	2006	C12N 15/90, A61K 48/00, C12N 15/10, C12N 15/82, C12N 9/22	SYSTEM FOR TRANSPOSING HYPERACTIVE RECOMBINANT DERIVATIVES OF MOS-1 TRANSPOSON
WO/2007/109013	2006	C12Q 1/68	METHOD FOR GENETIC SELECTION OF HIGH-PLASMID PRODUCING E. COLI CLONES
WO/2007/103382	2006	C12N 15/82	PLANT-SPECIFIC GENETIC ELEMENTS AND TRANSFER CASSETTES FOR PLANT TRANSFORMATION
WO/2007/100821	2006	A01K 67/027, C12N 15/09	PIGGYBAC CONSTRUCTS IN VERTEBRATES
WO/2007/082164	2006	C12N 15/90, C12N 15/85	METHODS FOR IDENTIFYING FUNCTIONAL NONCODING SEQUENCES
WO/2007/063033	2005	C12N 15/55, A61K 38/46, C12N 9/22	PROCESS FOR THE PRODUCTION, IN PROKARYOTES, OF ACTIVE, STABLE TRANSPOSASES OF MARINER MOBILE GENETIC ELEMENTS
WO/2007/041251	2005	C12Q 1/68	SYSTEM FOR DETECTING PROTEIN-PROTEIN INTERACTIONS
WO/2006/135602	2005	A01N 63/00	HERPES VIRUS-BASED COMPOSITIONS AND METHODS OF USE IN THE PRENATAL AND PERINATAL PERIODS
WO/2006/124001	2005	C12N 15/29, C12N 15/85	TRANSPOSITION OF MAIZE AC/DS ELEMENTS IN VERTEBRATES
WO/2006/122442	2005	C12N 9/22, C12N 15/09, C12N 15/85, C12N 15/86, C12N 5/16, C12N 5/22	PIGGYBAC AS A TOOL FOR GENETIC MANIPULATION AND ANALYSIS IN VERTEBRATES
WO/2006/108525	2005	C12N 15/90, C12N 9/22	RECONSTRUCTED HUMAN MARINER TRANSPOSON CAPABLE OF STABLE GENE TRANSFER INTO CHROMOSOMES IN VERTEBRATES
WO/2006/089121	2005	A61K 39/09, A61K 39/02	STREPTOCOCCUS INIAE PHOSPHOGLUCOMUTASE IS A VIRULENCE FACTOR AND TARGET FOR VACCINE DEVELOPMENT
WO/2006/077411	2005	C12N 15/10, C12Q 1/68	POLYPEPTIDE MUTAGENESIS METHOD
WO/2006/060314	2004	C12N 15/861, C12N 15/86, C12N 7/00	GENERATION OF REPLICATION COMPETENT VIRUSES FOR THERAPEUTIC USE
WO/2006/055040	2004	G01N 33/68, C12N 15/09, C12N 15/90	IDENTIFICATION OF PROTEINS IN A GENOME
WO/2006/046152	2004	C12N 15/63, C12N 1/21	METHODS FOR HETEROLOGOUS EXPRESSION OF SECONDARY METABOLITES
WO/2006/035045	2004	A01H 5/10, C12N 15/82	MAIZE HAVING AN IMPROVED DIGESTIBILITY

WO/2006/029449	2004	C12N 1/21, C12N 15/12	CHROMOSOMAL INSERTION OF GFP INTO BACTERIA FOR QUALITY CONTROL
WO/2006/029076	2004	C12N 15/82, A01H 1/00	PLANT-SPECIFIC GENETIC ELEMENTS AND TRANSFER CASSETTES FOR PLANT TRANSFORMATION
WO/2006/017371	2004	C12N 15/10	METHODS OF PRODUCING MUTANT POLYNUCLEOTIDES
WO/2006/008468	2004	C12N 15/861, C12N 15/10, C12N 15/64	ADENOVIRUS VECTOR AND METHOD TO MANIPULATE THE ADENOVIRUS GENOME
WO/2005/111060	2004	C12N 5/10, C07K 14/47, C12N 15/12	MOLECULES INVOLVED IN PROTEIN FOLDING AND METHODS OF IDENTIFYING THEM
WO/2005/100585	2004	C12P 21/00, C12P 19/34, C12P 21/06	METHODS FOR OBTAINING DIRECTIONALLY TRUNCATED POLYPEPTIDES
WO/2005/087263	2004	A61K 39/245	VACCINE FOR PREVENTING AND/OR TREATING A HERPES VIRUS INFECTION
WO/2005/083092	2004	C12N 15/80, C12N 9/22	TRANSPOSON FROM ASPERGILLUS FUMIGATUS AND MUTAGENESIS THEREWITH
WO/2005/066344	2004	C07K 14/61, C12N 15/62, C12N 15/70, C12N 15/10	NEW PLASMIDS, THEIR DERIVATIVES AND FRAGMENTS, THEIR METHODS OF MANUFACTURE AND APPLICATION
WO/2005/062923	2003	C12P 23/00, C07H 21/04, C12N 1/21, C12N 9/10	GENE TARGETS FOR ENHANCED CAROTENOID PRODUCTION
WO/2005/062881	2003	A01K 67/00, A01K 67/027	GENE THERAPY USING TRANSPOSON-BASED VECTORS
WO/2005/049789	2003	C12N 15/63, C07H 21/02, C07H 21/04, C12N 15/00, C12N 15/09, C12N 15/85, C12N 5/00	SYNTHETIC MAMMALIAN RETROTRANSPOSON GENE
WO/2005/024062	2003	C12N 15/10, C12Q 1/68	METHOD FOR THE IDENTIFICATION OF ATTENUATING LESIONS IN PATHOGENS
WO/2005/017205	2003	C12Q 1/68	NUCLEIC ACID MAPPING USING LINEAR ANALYSIS
WO/2005/016504	2003	C12N 15/00, C07H 21/00, C12N 15/10, C12N 15/63	DISRUPTION OF ACC SYNTHASE GENES TO DELAY SENESCENCE IN PLANTS
WO/2005/003364	2003	C12N 15/68, C12N 15/79, C12N 15/90	STABLE INTEGRANDS
WO/2004/106360	2003	C12P 19/34, C07H 21/02, C07K 1/00, C12N 7/00	VIRAL VECTORS WITH IMPROVED PROPERTIES
WO/2004/093645	2003	C12N 15/87	TN5 TRANSPOSASE MUTANTS AND THE USE THEREOF

WO/2004/090146	2003	C12N 15/10, C12N 15/90, C12N 9/22	DELIVERY OF NUCLEIC ACIDS INTO EUKARYOTIC GENOMES USING IN VITRO ASSEMBLED MU TRANSPOSITION COMPLEXES
WO/2004/078981	2003	C12N 9/22, A61K 38/00	HYPERACTIVE, NON-PHOSPHORYLATED, MUTANT TRANSPOSASES OF MARINER MOBILE GENETIC ELEMENTS
WO/2004/078962	2003	A61K 31/7052, C12N 9/22	MOBILE GENETIC ELEMENTS BELONGING TO THE MARINER FAMILY IN HYDROTHERMAL EUKARYOTES
WO/2004/070042	2003	C12N 15/90	TRANSPOSON-BASED TARGETING SYSTEM
WO/2004/069995	2003	C12N 15/90	TRANSPOSON-BASED TARGETING SYSTEM
WO/2004/069994	2003	C12N 15/90	TRANSPOSON-BASED TARGETING SYSTEM
WO/2004/067743	2003	C07K 14/62, C07K 16/00, C07K 16/02, C07K 16/10, C07K 16/26, C12N 15/85, A61K 39/00	VACCINE PRODUCTION USING TRANSPOSON BASED VECTORS
WO/2004/067707	2003	C07K 14/62, C07K 16/00, C07K 16/02, C07K 16/10, C07K 16/26, C12N 15/85, A61K 39/00	ADMINISTRATION OF TRANSPOSON-BASED VECTORS TO REPRODUCTIVE ORGANS
WO/2004/065581	2003	C12N 15/85	TRANSPOSON-INSULATOR ELEMENT DELIVERY SYSTEMS
WO/2004/044150	2002	A01K 67/00, A01K 67/033, C12N 15/90, C12N 5/06, C12N 5/10	SYSTEMS FOR GENE TARGETING AND PRODUCING STABLE GENOMIC TRANSGEN E INSERTIONS
WO/2004/018624	2002	C07K 14/31, C12N 15/10, A61K 38/00	RANDOM TRANSPOSON INSERTION IN STAPHYLOCOCCUS AUREUS AND USE THEREOF TO IDENTIFY ESSENTIAL GENES
WO/2004/015088	2002	C12N 15/74, C12N 9/22, C12P 19/62	TRANSPOSON-BASED TRANSFORMATION SYSTEM
WO/2004/013171	2002	A01K 67/033, C07K 16/00, C07K 16/18, C07K 16/42	MULTI-SUBUNIT PROTEIN PRODUCTION SYSTEM IN TRANSGENIC INSECT
WO/2004/006855	2002	A61K 39/05, C12N 1/36	RHODOCOCCUS EQUI MUTANTS AND VACCINES COMPRISING SAME
WO/2004/003157	2002	C07K 14/62, C07K 16/00, C07K 16/02, C07K 16/10, C07K 16/26, C12N 15/85, A61K 39/00, A61K 48/00	GENE REGULATION IN TRANSGENIC ANIMALS USING A TRANSPOSON-BASED VECTOR

WO/2003/100070	2002	C12N 15/85, C12N 15/90	THE FROG PRINCE, A TRANSPOSON VECTOR FOR GENE TRANSFER IN VERTEBRATES
WO/2003/097826	2002	A01K 67/027, A01K 67/033, C12N 15/10, C12N 15/85, C12N 15/90, C12N 9/22	USE OF MINOS IN FUNCTIONAL GENOMICS
WO/2003/097813	2002	C07K 14/31	REGULATION OF STAPHYLOCOCCUS AUREUS BIOFILM FORMATION
WO/2003/093451	2002	C07H 21/04	TRANSPOSABLE ELEMENTS IN RICE AND METHODS OF USE
WO/2003/093428	2002	C12N 15/82, C12N 15/90, C12N 9/00, C12N 9/12	GENE TARGETING USING REPLICATING DNA MOLECULES
WO/2003/089639	2002	C12N 15/10, C12N 15/90	TRANSPOSON FOR BIDIRECTIONAL INTRAMOLECULAR GENOME DELETIONS, CONSTRUCTION OF NOVEL MICROORGANISM AND IDENTIFICATION OF NONESSENTIAL GENES USING THE SAME
WO/2003/089618	2002	C12N 15/90, C12N 9/22, A61K 48/00	TRANSPOSON SYSTEM AND METHODS OF USE
WO/2003/089572	2002	C07K 14/21, G06F 19/00, A61K 38/00	ESSENTIAL AND IMPORTANT GENES OF PSEUDOMONAS AERUGINOSA AND THE USE THEREOF TO DESIGN OR IDENTIFY ANTIBACTERIAL AGENTS
WO/2003/087370	2002	C12N 15/10, C12N 15/90	METHOD AND MATERIALS FOR PRODUCING DELETION DERIVATIVES OF POLYPEPTIDES
WO/2003/076464	2002	C07K 14/38, A61K 38/00	MUTAGENESIS OF ASPERGILLUS FUNGI AND IDENTIFICATION OF GENES ESSENTIAL FOR GROWTH
WO/2003/074700	2002	C12N 15/10, C12N 15/90	TRANSPOSON
WO/2003/070955	2002	C12N 15/63, C12N 15/90	CONSTRUCTION OF NOVEL STRAINS CONTAINING MINIMIZING GENOME BY Tn5-COUPLED Cre/loxP EXCISION SYSTEM
WO/2003/062390	2002	G01N 33/50	COMPOSITIONS AND METHODS FOR THE MODULATION OF SPHINGOLIPID METABOLISM AND/OR SIGNALING
WO/2003/060120	2001	C07K 14/22, C12N 15/10, C12Q 1/68	MEANS FOR IDENTIFYING NEISSEIRIA MENINGITIDIS SPECIFIC GENES
WO/2003/056912	2002	A01K 67/027, A01K 67/033, C12N 15/85, C12N 15/90, C12N 9/22	INDUCIBLE TRANSPOSITION IN TRANSGENIC ORGANISM USING TRANSPOSON VECTOR
WO/2003/031629	2001	C12N 15/63	METHODS OF PREPARING A TARGETING VECTOR AND USES THEREOF
WO/2003/027241	2001	C12N 15/85, C12N 15/90	ANIMAL INTEGRATION VECTOR AND METHODS FOR ITS USE

WO/2003/002738	2001	C12N 15/10, C12N 15/76	METHODS AND MATERIALS FOR GENERATING GENETIC DISRUPTIONS IN BACTERIAL CELLS
WO/2002/103513	2001	C07K 14/21, A61K 38/00	ESSENTIAL AND IMPORTANT GENES OF PSEUDOMONAS AERUGINOSA AND THE USE THEREOF TO DESIGN OR IDENTIFY ANTIBACTERIAL AGENTS
WO/2002/101061	2001	C12N 15/82	METHOD FOR OBTAINING A MONOCOTYLEDON PLANT CONTAINING A GENE OF INTEREST FREE OF FOREIGN ANCILLARY SEQUENCE
WO/2002/097100	2001	C12N 15/70	CLONING VECTORS FOR HOMOLOGOUS RECOMBINATION AND METHOD USING SAME
WO/2002/084280	2001	A01K 67/033	INSECT AMMUNITION VECTORS AND METHODS OF USE TO IDENTIFY PESTICIDE TARGETS
WO/2002/070666	2001	C07K 14/31, C12Q 1/14, C12Q 1/18	COMPOSITIONS AND METHODS FOR REGULATING AUTOLYTIC PROCESSES IN BACTERIA
WO/2002/068634	2001	C12N 15/10	METHOD FOR PRODUCING GENE LIBRARIES
WO/2002/068631	2001	C12N 15/74	PLASMID FOR INSERTIONAL MUTAGENESIS IN BACTERIA
WO/2002/066653	2000	C12N 15/10	PROCARYOTIC LIBRARIES AND USES
WO/2002/062991	2001	C12N 15/10, C12N 15/90	INSERTIONAL MUTAGENESIS TECHNIQUE
WO/2002/053732	2000	C12N 15/85, C12N 15/90	METHODS FOR MAKING POLYNUCLEOTIDE LIBRARIES, POLYNUCLEOTIDE ARRAYS, AND CELL LIBRARIES FOR HIGH-THROUGHPUT GENOMICS ANALYSIS
WO/2002/046444	2000	C12N 15/90, C12N 9/22	DOUBLE TRANSPOSITION METHODS FOR MANIPULATING NUCLEIC ACIDS
WO/2002/024890	2000	C07K 14/415, C12N 15/82	RICE MLH1 ORTHOLOG AND USES THEREOF
WO/2002/022876	2000	C07K 14/24, C07K 14/245, C07K 14/255, C12N 15/10, C12Q 1/68	METHOD OF IDENTIFYING PUTATIVE ANTIBIOTIC RESISTANCE GENES
WO/2002/022804	2000	C12N 15/82, C12N 15/90	COMPOSITIONS AND METHODS FOR STABLE TRANSFORMATION USING MU BACTERIOPHAGE CLEAVED DONOR COMPLEX
WO/2002/018601	2000	C07K 14/285	ESSENTIAL BACTERIA GENES AND GENOME SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF 'ESSENTIAL GENES'
WO/2002/008431	2000	C12N 15/10, C12N 15/74	COMPOSITIONS AND METHODS FOR USE THEREOF IN MODIFYING THE GENOMES OF MICROORGANISMS
WO/2002/008286	2000	C07K 14/47, C12N 15/82	ZINC FINGER DOMAIN RECOGNITION CODE AND USES THEREOF
WO/2002/004674	2000	C12N 15/09, C12N 15/10, C12Q 1/68	TRANSPOSON MEDIATED MULTIPLEX SEQUENCING

WO/2002/004629	2000	C12N 15/10, C12N 15/90, C12N 9/22	MOLECULAR BREEDING OF TRANSPOSABLE ELEMENTS
WO/2002/000916	2000	C12N 15/10	METHODS FOR IDENTIFYING AN ESSENTIAL GENE IN A PROKARYOTIC MICROORGANISM
WO/2001/089579	2000	A61K 47/48	COMPOSITION FOR DELIVERY OF COMPOUNDS TO CELLS
WO/2001/085965	2000	C12N 15/10, C12N 15/70	VECTORS FOR USE IN TRANSPOSON-BASED DNA SEQUENCING METHODS
WO/2001/083786	2000	C12N 15/74, C12N 15/90, C12N 9/10, C12N 9/12, C12N 9/92, C12P 7/06	STABLE ZYMOMONAS MOBILIS XYLOSE AND ARABINOSE FERMENTING STRAINS
WO/2001/081565	2000	C12N 15/90	SLEEPING BEAUTY, A TRANSPOSON VECTOR WITH A BROAD HOST RANGE FOR THE GENETIC TRANSFORMATION IN VERTEBRATES
WO/2001/079419	2000	C12Q 1/68	CELLULAR ARRAYS FOR THE IDENTIFICATION OF ALTERED GENE EXPRESSION
WO/2001/077319	2000	C07K 14/33, C12N 15/65, C12N 15/74	GENETIC MANIPULATION OF CLOSTRIDIUM DIFFICILE
WO/2001/071040	2000	C12N 15/10, C12Q 1/68	METHOD FOR DETERMINATION OF GENE FUNCTION
WO/2001/071019	2000	A01K 67/033, C12N 15/82, C12N 15/85, C12N 15/90	METHOD OF GENERATING TRANSGENIC ORGANISMS USING TRANSPOSONS
WO/2001/064926	2000	C12N 15/82	METHODS AND COMPOSITIONS TO REDUCE OR ELIMINATE TRANSMISSION OF A TRANSGENE
WO/2001/051649	2000	C12Q 1/18	IDENTIFICATION OF VIRULENCE DETERMINANTS
WO/2001/044483	1999	A01K 67/033, C12N 15/85	TRANSGENIC INSECT
WO/2001/034819	1999	C07K 14/415, C12N 15/82	GENES AND METHODS FOR MANIPULATION OF GROWTH
WO/2001/030965	1999	A61K 38/00, A61K 48/00, C12N 15/00, C12N 15/63, C12N 15/85, C12N 15/87	METHODS OF IN VIVO GENE TRANSFER USING A SLEEPING BEAUTY TRANSPOSON SYSTEM
WO/2001/029205	1999	A01K 67/033, C12N 15/90, C12N 9/22	METHOD FOR GENETIC MANIPULATION
WO/2001/021781	1999	C12N 15/82	METHODS FOR SITE-ASSOCIATED MODIFICATION OF GENE ACTIVITY AND NUCLEIC ACID STRUCTURE

WO/2001/019993	1999	C07K 14/35, C12N 9/00, A61K 38/00, A61K 39/00	VIRULENCE GENES OF M. MARINUM AND M. TUBERCULOSIS
WO/2001/014537	1999	A01K 67/033, C07K 14/435, C12N 15/85	PIGGYBAC TRANSFORMATION SYSTEM
WO/2001/011060	1999	C12N 15/10, C12N 15/62, C12N 15/74	METHOD OF ISOLATING SECRETION SIGNALS IN LACTIC ACID BACTERIA AND NOVEL SECRETION SIGNALS ISOLATED FROM LACTOCOCCUS LACTIS
WO/2001/011033	1999	C07K 14/285, C12N 15/10	IDENTIFICATION OF GENES ESSENTIAL FOR THE SURVIVAL OF HAEMOPHILUS INFLUENZAE THROUGH GENOME SCANNING BY TRANSPOSITION MUTAGENESIS
WO/2001/009363	1999	C12N 15/10, C12N 15/90, C12N 9/22	MUTANT TN5 TRANSPOSASE ENZYMES AND METHOD FOR THEIR USE
WO/2001/007651	1999	C12Q 1/68	TRANSPOSON MEDIATED DIFFERENTIAL HYBRIDISATION
WO/2001/002555	1999	C12N 15/10, C12N 15/74	METHOD OF MAKING AND IDENTIFYING ATTENUATED MICROORGANISMS, COMPOSITIONS UTILIZING THE SEQUENCES RESPONSIBLE FOR ATTENUATION, AND PREPARATIONS CONTAINING ATTENUATED MICROORGANISMS
WO/2000/075289	1999	C12N 15/82	COMPOSITIONS AND METHODS FOR TARGETED GENE INSERTION
WO/2000/073510	1999	C07K 14/435, C12N 9/22	METHOD OF TRANSPOSON-MEDIATED MUTAGENESIS IN THE NEMATODE CAENORHABDITIS ELEGANS
WO/2000/071158	1999	A61K 39/21, C07K 14/74, C12N 15/10, C12N 15/63, C12Q 1/68	TAGGED EPITOPE PROTEIN TRANSPOSABLE ELEMENT
WO/2000/068399	1999	A61K 48/00, C12N 15/63, C12N 15/861, C12N 15/90, C12N 9/22	VECTOR-MEDIATED DELIVERY OF INTEGRATING TRANSPOSON SEQUENCES
WO/2000/065042	1999	A61K 48/00, C07K 14/435, C12N 15/85, C12N 15/90	P ELEMENT DERIVED VECTOR AND METHODS FOR ITS USE
WO/2000/056902	1999	C12N 15/80	POLYNUCLEOTIDES FOR MUTAGENESIS IN FUNGUS COMPRISING A FUNCTIONAL GENE IN MAGNAPORTHE AND AN IMPALA TRANSPOSON
WO/2000/055346	1999	C12N 15/10, C12N 15/90, C12N 9/22	METHODS AND MATERIALS FOR THE RAPID AND HIGH VOLUME PRODUCTION OF A GENE KNOCK-OUT LIBRARY IN AN ORGANISM

WO/2000/051635	1999	A61K 35/74, A61K 39/00, A61K 39/39, C12P 19/04, C12P 7/64	EMULSAN AND EMULSAN ANALOGS IMMUNIZATION FORMULATIONS AND USE
WO/2000/017343	1998	C12N 15/10, C12N 15/90, C12N 9/22	METHOD FOR MAKING INSERTIONAL MUTATIONS
WO/2000/012734	1998	C12N 15/82	TRANSPOSON TAGGING AND GENE DELIVERY IN SMALL GRAIN CEREALS
WO/2000/010582	1998	A61K 35/74, A61K 45/06	PROBIOTIC, LACTIC ACID-PRODUCING BACTERIA AND USES THEREOF
WO/1999/061604	1998	C07K 14/435, C12N 15/10, C12N 15/85, C12Q 1/68	A VECTOR FOR GENE TRAP, AND A METHOD FOR GENE TRAPPING BY USING THE VECTOR
WO/1999/050402	1998	C12N 15/10, C12Q 1/68	SYSTEMATIC IDENTIFICATION OF ESSENTIAL GENES BY <i>in vitro</i> TRANSPOSON MUTAGENESIS
WO/1999/025817	1997	C12N 15/10, C12N 15/90	TC1-BASED TRANSPOSON VECTORS
WO/1999/016868	1997	C12N 15/74	TM4 CONDITIONAL SHUTTLE PHASMIDS AND USES THEREOF
WO/1999/009817	1997	A01K 67/027, C12N 15/87, C12N 15/90, A61K 48/00	USE OF <i>in</i> (MARINER) TRANSPOSAN IN THE PRODUCTION OF TRANSGENIC ANIMALS
WO/1999/007871	1997	C07K 14/435, C12N 15/90, C12N 9/22, A61K 48/00	VECTORS AND METHODS FOR PROVIDING CELLS WITH ADDITIONAL NUCLEIC ACID MATERIAL INTEGRATED IN THE GENOME OF SAID CELLS
WO/1999/005295	1997	C12N 15/90	COMPOSITION AND METHOD FOR TARGETED INTEGRATION INTO CELLS
WO/1999/005281	1997	C07K 14/415, C12N 15/82	NOVEL NUCLEIC ACID MOLECULES AND USES THEREFOR
WO/1999/004035	1997	C12Q 1/68	<i>in vitro</i> METHOD FOR PROVIDING TEMPLATES FOR DNA SEQUENCING
WO/1998/056903	1997	A01K 67/027, C12N 15/85, C12N 15/90	METHODS AND USES FOR TRANSPOSON-BASED GENE TARGETING
WO/1998/044141	1997	C07K 14/575, C07K 14/79, C12N 9/64, C12N 15/69, C12N 15/85	INSECT EXPRESSION VECTORS
WO/1998/040510	1997	C12N 15/10, C12N 15/90	DNA-BASED TRANSPOSON SYSTEM FOR THE INTRODUCTION OF NUCLEIC ACID INTO DNA OF A CELL
WO/1998/037205	1997	C12N 9/14, C12N 15/70, C12N 15/90	MUTATIONS IN ATP-DEPENDENT TRANSPOSITION PROTEINS THAT REDUCE TARGET-SITE SPECIFICITY
WO/1998/028431	1996	C12N 15/82	TRANSCRIPTIONAL REGULATION IN PLANTS

WO/1998/010077	1996	C12N 9/22, C12N 15/85, C12N 15/90	SYSTEM FOR ϕ (IN VITRO) TRANSPOSITION USING MODIFIED TN5 TRANSPOSASE
WO/1998/008960	1996	C12N 9/22, C12N 15/81	IDENTIFICATION OF AND CLONING A MOBILE TRANSPOSON FROM ϕ (ASPERGILLUS)
WO/1997/029768	1996	A61K 39/00, A61K 39/112, C07K 14/255	SALMONELLA TYPHIMURIUM VACCINE
WO/1997/029202	1996	C12N 15/90	VECTORS AND METHODS FOR PROVIDING CELLS WITH ADDITIONAL NUCLEIC ACID MATERIAL INTEGRATED IN THE GENOME OF SAID CELLS
WO/1997/015679	1995	C07K 14/16, C12N 15/00, C12N 15/861, C12N 15/867, C12N 15/90	RECOMBINANT VIRUSES CONTAINING MOBILE GENETIC ELEMENTS AND METHODS OF USE IN GENE THERAPY
WO/1996/041891	1995	C07K 14/205, C12Q 1/68	PROCESS FOR IDENTIFYING SECRETORY GENES FROM HELICOBACTER PYLORI
WO/1996/039803	1995	A01H 1/06, C07K 14/415, C12N 15/01	USE OF THE INDETERMINATE GAMETOPHYTE GENE FOR MAIZE IMPROVEMENT
WO/1996/029414	1995	C12N 9/22, C12N 15/81	IDENTIFICATION AND CLONING OF A MOBILE TRANSPOSON FROM ASPERGILLUS
WO/1996/023073	1995	C12N 15/75, C12N 15/90	DNA INTEGRATION BY TRANSPOSITION
WO/1995/032296	1994	C12N 9/02, C12N 9/88, C12N 15/52, C12N 15/74	SHUTTLE VECTORS FOR THE INTRODUCTION OF DNA INTO MYCOBACTERIA AND UTILIZATION OF SUCH BACTERIA AS VACCINES
WO/1995/030005	1994	C07K 14/415, C12N 15/82	GENES REGULATING THE RESPONSE OF ϕ (ZEA MAYS) TO WATER DEFICIT
WO/1995/023875	1994	C12N 15/90, C12Q 1/68	ϕ (IN VITRO) TRANSPOSITION OF ARTIFICIAL TRANSPOSONS
WO/1995/020040	1994	A01N 63/00	PSEUDOMONAS BIOCONTROL STRAINS
WO/1995/009923	1993	C12N 15/70, C12N 15/866	METHOD OF PRODUCING RECOMBINANT EUKARYOTIC VIRUSES IN BACTERIA
WO/1995/002058	1993	A01N 63/00, C07K 14/325, C12N 9/00, C12N 15/75	ϕ (BACILLUS THURINGIENSIS) TRANSPOSON T _n 5401 AND ITS USE IN A SITE-SPECIFIC RECOMBINATION SYSTEM FOR ϕ (BACILLUS THURINGIENSIS) STRAIN DEVELOPMENT
WO/1995/001095	1993	A61K 38/00, A61K 48/00, C07K 14/435, C07K 14/47, C12N 15/85, C12N 15/90	TRANSFORMED EUKARYOTIC CELLS, AND TRANSPOSON-BASED TRANSFORMATION VECTORS
WO/1994/023561	1993	A01H 5/02, C07K 14/415, C12N 15/82	PH GENES AND THEIR USES
WO/1994/011024	1992	A61K 39/102, C07K 14/285	COMPOSITION PROTECTIVE AGAINST ϕ (P. MULTOCIDA) PASTEURELLOSIS INFECTION

WO/1994/002599	1992	A61K 48/00, C12N 15/00, C12N 15/10, C12N 15/85, C12N 15/90, C12Q 1/68	ADVANCED GLYCOSYLATION ENDPRODUCTS TO CAPTURE TRANSPOSONS MODIFIED DNA
WO/1993/016728	1992	A61K 39/00, A61K 39/10	IMPROVED VACCIN AGAINST BRUCELLA ABORTUS
WO/1993/014211	1992	C07K 14/415, C12N 15/82	REGULATION OF PLANT GENES
WO/1993/007161	1991	C07K 14/435, C07K 14/705	MOLECULAR CLONING AND TRANSFORMATION OF CYCLODIENE RESISTANCE
WO/1993/006221	1991	C12N 15/10, C12N 15/63, C12N 15/82	METHOD FOR PROVIDING A DELETION OR INVERSION MUTATION IN A PLANT GENOME; RECOMBINANT DNA USABLE THEREFOR; MUTATED PLANT
WO/1992/022326	1991	A61K 39/00, C07K 14/35, C12N 15/74	INSERTIONAL MUTATIONS IN MYCOBACTERIA
WO/1992/019750	1991	A61K 38/00, A61K 39/10, C07K 14/235, C12N 15/63, C12N 15/74, C12N 15/90	HYBRID TRANSPOSON WITH A PT-OPERON, PLASMIDS AND BACTERIAL STRAINS THEREFOR
WO/1992/019738	1991	C02F 3/12, C02F 3/34, C12N 9/02, C12N 15/52	MICROBIAL DEGRADATION OF TRICHLOROETHYLENE, DICHLOROETHYLENES AND AROMATIC POLLUTANTS
WO/1992/015687	1991	C07K 14/195, C12N 9/02, C12N 15/62, C12Q 1/02, C12Q 1/68, G01N 33/00, G01N 33/58, G01N 33/84	FUSED GENES AND THEIR USE FOR DETERMINING THE PRESENCE OF METALS OR OF XENOBIOTIC COMPOUNDS
WO/1992/006205	1990	C12N 15/52, C12N 15/82	A PROCESS FOR THE GENE MANIPULATION OF PLANT CELLS, RECOMBINANT PLASMIDS, RECOMBINANT BACTERIA, PLANTS
WO/1992/001370	1990	C12N 15/64, C12N 15/82, C12Q 1/68	BIOLOGICALLY SAFE PLANT TRANSFORMATION SYSTEM
WO/1990/012891	1989	C12N 15/10, C12Q 1/68	METHOD OF PHYSICALLY MAPPING GENETIC MATERIAL
WO/1990/004041	1988	C07K 14/21, C12N 9/24, C12N 15/00, C12Q 1/02, C12Q 1/04, C12Q 1/70	BACTERIAL DETECTION BY PHAGE TRANSDUCTION OF DETECTABLE PHENOTYPE
WO/1989/009063	1988	A61K 39/112, C12N 15/00, C12N 15/74	NOVEL NON-REVERTING SHIGELLA LIVE VACCINES

WO/1989/003427	1987	A61K 39/04, C07K 14/315, C07K 14/35, C12N 9/02, C12N 15/00, C12N 15/68	A METHOD OF MAINTAINING A DESIRED RECOMBINANT GENE IN A GENETIC POPULATION OF CELLS
WO/1988/003169	1986	C12N 15/62, C12N 15/81, C12N 15/90	GENE AMPLIFICATION USING RETROTRANSPOSONS
WO/1988/001646	1986	C12N 15/10, C12N 15/76, C12N 15/81	UNIVERSAL SYSTEM FOR TRANSPOSON MUTAGENESIS

**Transposon tools: worldwide landscape of intellectual property
and technological developments**

Fabien Palazzoli, François-Xavier Testu, Franck Merly and Yves Bigot

Electronic Supplementary Material

**Legal status of patent families relative to the proof of concept in use and in
engineering for specific transposon families
(October 2009)**

□□a□□□□□ systems	□□□ patent application □	Issued patents	EP applications under examination	Other publication□
<i>Tn5</i>	WO/2000/017343	CA2343000 (C) US6159736 (A) EP1115856 (B1) US6294385 (B1) AU758960 (B2) RU2237715 (C2) CN100420747 (C)	/	JP2002531062 (T) CN1319135 (A) PL346772 A1 AU6057399 (A)
	WO/1998/010077	CA2265477 (C) EP0927258 (B1) US5925545 (A) US5948622 (A) US5965443 (A) AU732130 (B2) US6437109 (B1) RU2218406 (C2) CN1163605 (C)	/	PL332145 (A1) CN1251135 (A) JP2001507565 (T) AU4262097 (A)
	WO/2004/093645	US7083980 (B2)	/	US2004235103 (A1)
	WO/2002/046444	US7067644 (B2) US7338801 (B2)	/	CA2431246 (A1), US2002132349 (A1) EP1339863 (A2) US2006223186 (A1) AU2588802 (A)
	WO/2001/009363	US6406896 (B1) AU775043 (B2)	/	CA2380850 (A1) EP1198583 (A1) CN1367840 (A) JP2003505104 (T) PL353241 (A1) RU2002105508 (A)
<i>Tn10</i>	WO/2004/067707WO/2004/003157	EP1592789 (B1)EP1539785 (B1)AU2003261096 (B2)US7527966 (B2)	/	JP2006512922 (T)JP2006512921 (T)JP2005530516 (T)CA2490693 (A1)US2004172667 (A1)US2004197910 (A1)US2004235011 (A1)BR0305217 (A)EP1597378 (A2)US2008235813 (A1)AU6396700 (A)AU2003303831 (A1)AU2003303830 (A1)AU2003299959 (A1)
<i>Tol2</i>	WO/2003/068960*	EP1484395 (B1)	/	US2005177890 (A1) AU2003211996 (A1) JP2003235575 (A)
	WO/2001/040477*	US7468430 (B2) US7034115 (B1) EP1239042 (B1)	/	US2006212959 (A1) US2006212958 (A1) US2006211116 (A1) JP2001218588 (A)
	WO/2007/082164	/	/	/

ranpo on tem	P patent applications	Issued patents	EP applications under examination	Other publications
no	WO/2003/056912	AU2003201649 (B2) EP1465993 (B1)	/	US2005066376 (A1) US2008255002 (A1) JP2005512602 (T) CN1612931 (A) CA2470277 (A1) AU2003201649 (A1)
	WO/2001/071019	EP1268837 (B1) AU2001246507 (B2)	/	US2003150007 (A1) US2008193933 (A1) JP2003527864 (T) CA2402924 (A1) AU4650701 (A)
	WO/2003/097826	/	/	AU2003227945 (A1)
	WO/2001/044483	OA12120 (A)	/	GB9929681 (D0) JP2004500064 (T) MXPA02005960 (A) CN1409766 (A) US2003033622 (A1) BR0016398 (A) EP1242607 (A1) AU1871901 (A) AP200202514 (D0)
Intern n	WO/2001/081565	EP1276889 (B1)	/	US2005112764 (A1) CA2407651 (A1) AU6575601 (A)
	WO/1998/040510	AU745049 (B2) US6489458 (B2) US7148203 (B2) EP0973928 (A1)	/	CA2280997 (A1) JP2002511741 (T) US2002016975 (A1) US2003124668 (A1) AU6456998 (A) JP2008271981 (A)
	WO/2003/089618	/	/	US2004077572 (A1) AU2003231048 (A1)
	WO/2001/030965	US6613752 (B2)	/	US2002103152 (A1) AU1100201 (A)
	WO/2000/068399	/	/	CA2373121 (A1) JP2002543792 (T) AU4711000 (A)
	WO/1999/025817	/	/	EP1034258 (A2) CA2309000 (A1) JP2001523450 (T) AU1410399 (A)
TI	WO/2001/029205	/	/	AU1292301 (A)
	WO/1999/007871	/	/	CA2301198 (A1) EP1003899 (A1) JP2001512690 (T) AU8752898 (A)
	WO/1997/029202	AU731615 (B2)	/	CA2244812 (A1)EP0885307 (A1)JP2001503962 (T)AU1674697 (A)

Transposon systems	Patent applications	Issued patents	EP applications under examination	Other publications
n	WO/2006/108525	EP1869193 (B1)	/	EP1710316 (A1) US2008279838 (A1)
	WO/2004/078962	FR2850668 (B1)	/	/
	WO/1999/009817	/	/	CA2300972 (A1) JP2001513336 (T) EP1006790 (A1) AU8817798 (A)
oI	WO/2007/132096	/	EP2021485 (A2)#	FR2900935 (A1) CA2651413 (A1)
	WO/2007/063033	FR2893951 (B1)	EP1957649 (A1)#	JP2009517059 (T) CA2631400 (A1)
	WO/2004/078981	FR2850395 (B1)	EP1590460 (A1)#	US2007031967 (A1) JP2006518220 (T) CA2514730 (A1)
I n	WO/2003/027241	EP1436399 (B1)	/	US2003092179 (A1) US2005034180 (A1) US2005176145 (A1) AU2002331884 (A1)
	WO/2000/065042	US6291243 (B1) US6475798 (B2)	/	US2002028513 (A1) JP2002542781 (T) EP1171588 (A1) AU4806100 (A)
o	/	/	/	/
	WO/2000/055346	/	/	JP2003502015 (T) EP1161551 (A2) AU3900700 (A)
	WO/1999/050402	US6207384 (B1)	/	AU3196099 (A)
	WO/2006/122442	/	EP1896578 (A1)#	NO20076465 (A)MX2007014139 (A)JP2008545375 (T)CN101297031 (A)CA2608481 (A1)BRPI0520212 (A2)AU2005331864 (A1)KR20080041146 (A)IN5762/CHENP/2007 (A)
	WO/2008/098181	/	/	/
	WO/2008/027384	/	/	/
	WO/2007/100821	/	/	US2007204356 (A1)
	WO/2001/014537	US6773914 (B1) US7005296 (B1)	/	AU6909200 (A)

Transposon tools: worldwide landscape of intellectual property and technological developments

Fabien Palazzoli, François-Xavier Testu, Franck Merly and Yves Bigot

Electronic ¹supplementar² ³aterial 4⁵

Oral communication done in 1995 about piggyBac vectors



FIRST INTERNATIONAL
WORKSHOP ON TRANSGENESIS
OF INVERTEBRATES
OF MEDICAL , AGRICULTURAL ,
AND
AQUACULTURAL IMPORTANCE

ABSTRACTS

MONTPELLIER, FRANCE APRIL 21 - 26, 1995

HOTEL RÉSIDENCE LES PINS, CLAPIERS

EXPLORING THE UTILITY OF THE IFP2 TRANSPOSON OF LEPIDOPTERANS FOR GENETIC ENGINEERING OF INSECTS.

Malcolm J. FRASER, Teri A. ELICK, Christopher BAUSER, Nicole PRINCIPE,
Michael LIMON, Paul D. SHIRK*, and Joseph E. OTOUSA.

The Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556,
and *USDA/ARS, Insect Attractants, Behavior, and Basic Biology Research Laboratory,
Gainesville, FL 32604

We are studying the movement of TTAA-specific Lepidopteran Class II transposons originally identified as insertions within the genomes of Baculovirus mutants. Our current analyses have focused on the IFP2 element originally derived from the TN-368 cell line. The element is transcriptionally active in the TN-368 cell line of origin, and encodes a single large open reading frame. PCR analyses of genomic representatives of this element demonstrate that mobilization into the Baculovirus genome is mechanistically identical to movement within the TN-368 cell genome. Similar analyses have as yet failed to identify IFP2-specific sequences in either laboratory or natural populations of *Trichoplusia ni*. We have developed an assay for transpositional insertion of this element into the Baculovirus genome and determined that this element is capable of carrying active genes during movement from plasmids into the virus genome in infected insect cells. The assay results also indicate that the IFP2 element may encode some factor that participates in its mobilization. We have established that excision occurs in a precise fashion in cell lines that support transpositional movement of the element. Current analyses of IFP2 transposon constructs for excision in insect embryos and for establishment of transgenic insects will be discussed.

Transposon tools: worldwide landscape of intellectual property and technological developments

Fabien Palazzoli, François-Xavier Testu, Franck Merly and Yves Bigot

Electronic ¹supplementar² ³aterial 5⁴

**Authors top 20 for the scientific publication on each transposon system
(April 2009)**

Tc1-like elements: Tc1, Sleeping beauty, Minos

Author	Publications ▼
 Plasterk R	27
 Largaespada D	24
 Hackett P	23
 Ivics Z	23
 Izsvák Z	20
 McIvor R	17
 Savakis C	17
 Kay M	15
 Yant S	14
 Ekker S	10
 Takeda J	10
 Kren B	9
 Sasakura Y	9
 Horie K	9
 Loukeris T	9
 Van Luenen H	9
 Steer C	8
 Satoh N	8
 Dupuy A	8
 Emmons S	8

Mariner-like elements: Mos1, Himar, ...

Author	Publications ▼
 Hartl D	23
 Bigot Y	19
 Robertson H	18
 Lampe D	16
 Ivics Z	14
 Capy P	13
 Plasterk R	13
 Lohe A	12
 Izsvák Z	12
 Augé-Gouillou C	9
 Sasakura Y	8
 Langin T	7
 O'Brochta D	7
 Atkinson P	7
 Satoh N	7
 Beverley S	7
 Mekalanos J	7
 Finnegan D	7
 Daboussi M	6
 Awazu S	6

hAT: Tol2, hobo, ...

Author	Publications ▼
 Kawakami K	14
 Koga A	12
 Hori H	10
 O'Brochta D	9
 Atkinson P	9
 Zakharov I	6
 Shima A	6
 Zakharenko L	5
 Bazin C	5
 Higuet D	5
 Handler A	5
 Kovalenko L	4
 Lemeunier F	4
 Iida A	4
 Urasaki A	4
 Bonnivard E	4
 Warren W	4
 Ladevèze V	3
 Korzh V	3
 Chaminade N	3

piggyBac

Top 20 authors	
Author	Publications ▼
 Fraser M	13
 Handler A	9
 Lobo N	5
 Bauser C	4
 Wimmer E	4
 O'Brochta D	3
 Elick T	3
 Horn C	3
 Coates C	3
 Kanda T	3
 Kaminski J	3
 Denell R	2
 Miller T	2
 Li X	2
 Heinrich J	2
 Beeman R	2
 Scott M	2
 Lorenzen M	2
 Harrell R	2
 Hua-Van A	2

P elements

Author	Publications ▼
 Rio D	22
 Engels W	13
 Simmons M	12
 Hagemann S	10
 Anxolabéhère D	9
 Chovnick A	8
 Rubin G	8
 Pinsker W	7
 MacKay T	7
 Nouaud D	6
 Raymond J	6
 Kidwell M	6
 Bellen H	6
 Spradling A	6
 Handler A	5
 Margulies L	5
 Georgiev P	5
 Miller W	5
 Kassis J	5
 Anholt R	5

Tn5 and Tn10

Author	Publications ▼
 Reznikoff W	75
 Hillen W	29
 Kleckner N	27
 Berg D	25
 Haniford D	23
 Yamaguchi A	22
 Sawai T	20
 Goryshin I	15
 Ausubel F	14
 Timmis K	13
 Volk C	12
 Kleckner N	12
 Steinbüchel A	11
 De Lorenzo V	11
 Rolfe B	10
 De Bruijn F	10
 Pühler A	9
 Foster T	8
 Lupski J	8
 Wardle S	8

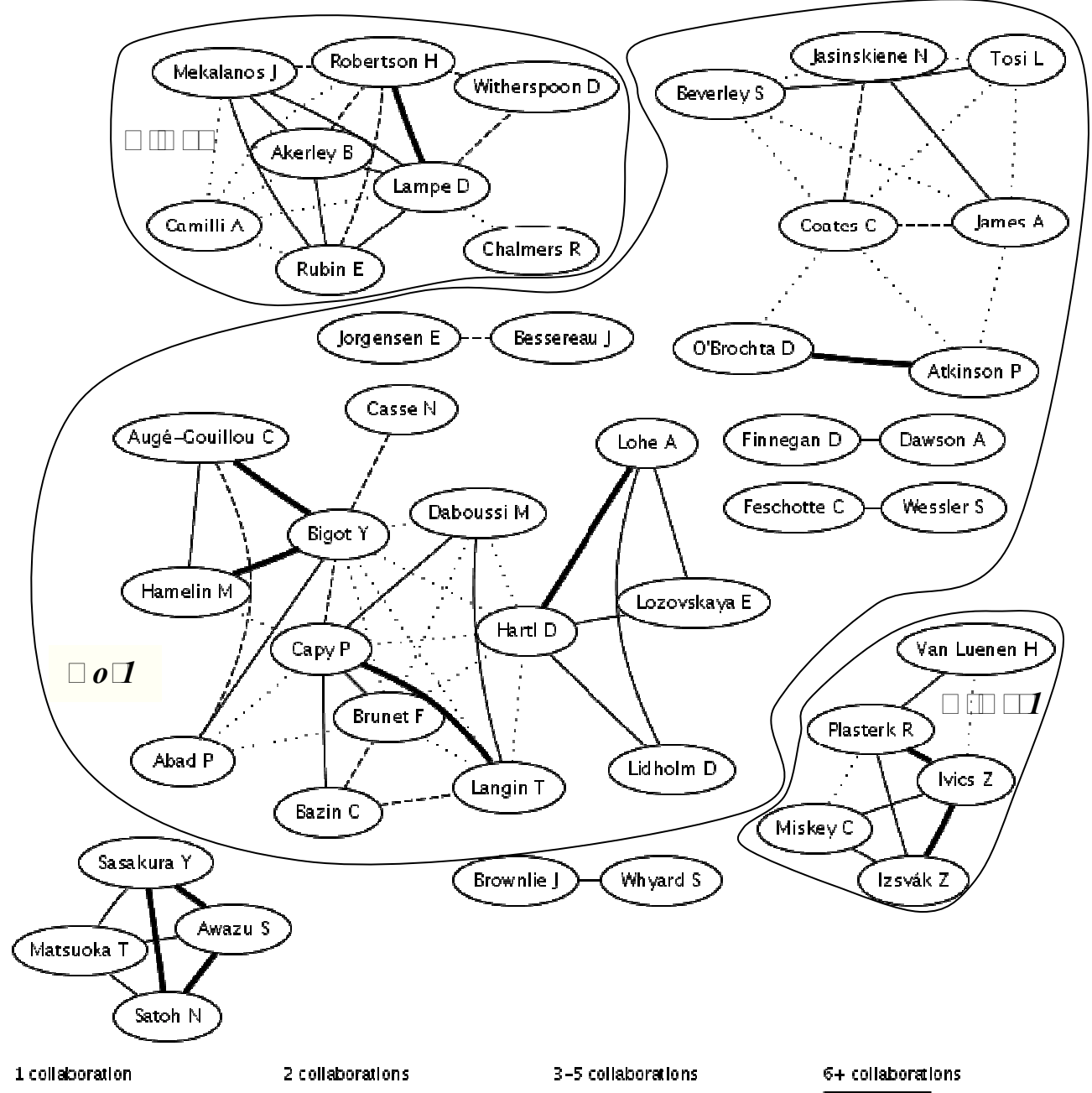
**Transposon tools: worldwide landscape of intellectual property
and technological developments**

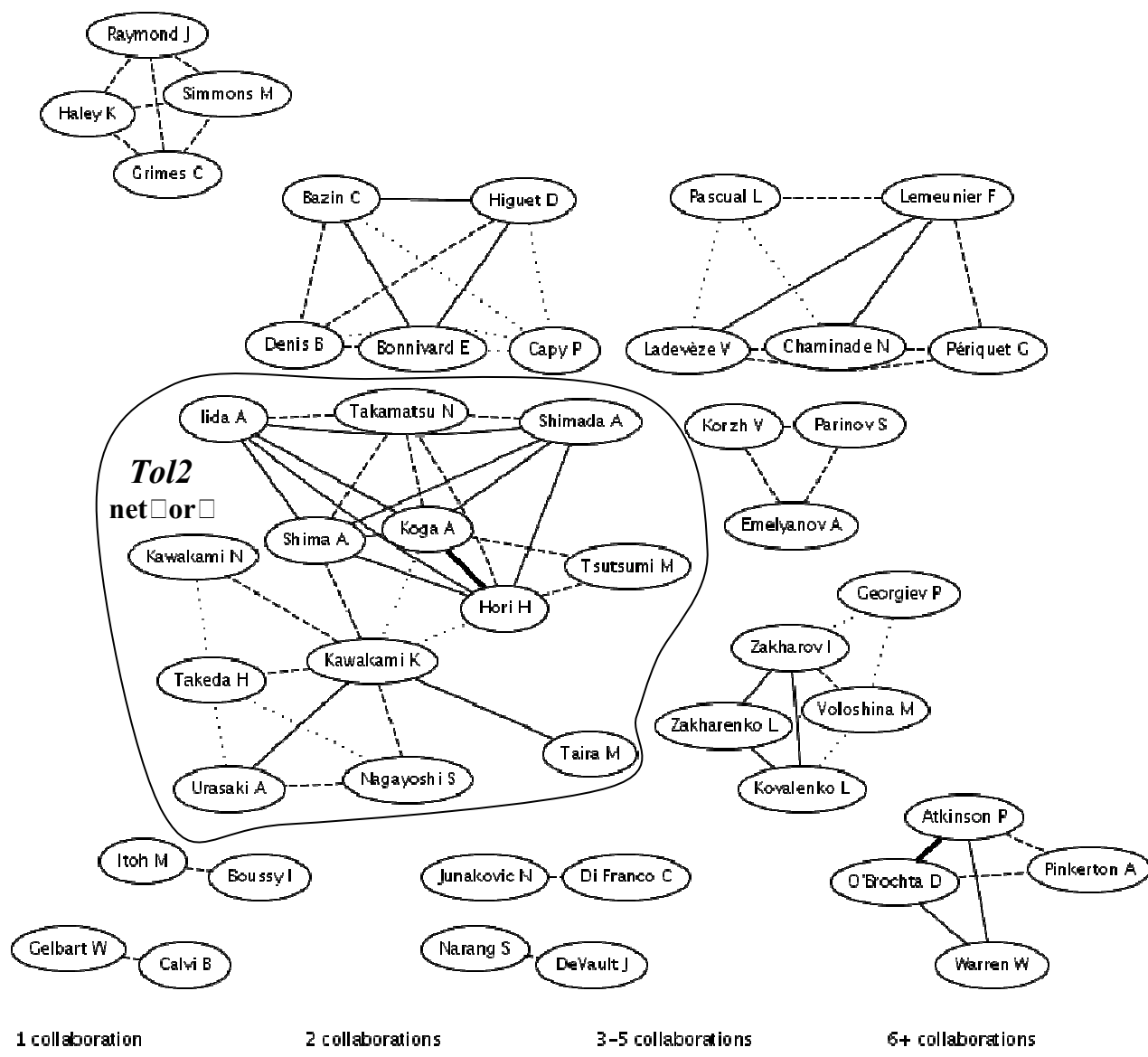
Fabien Palazzoli, François-Xavier Testu, Franck Merly and Yves Bigot

Electronic supplementary material 6

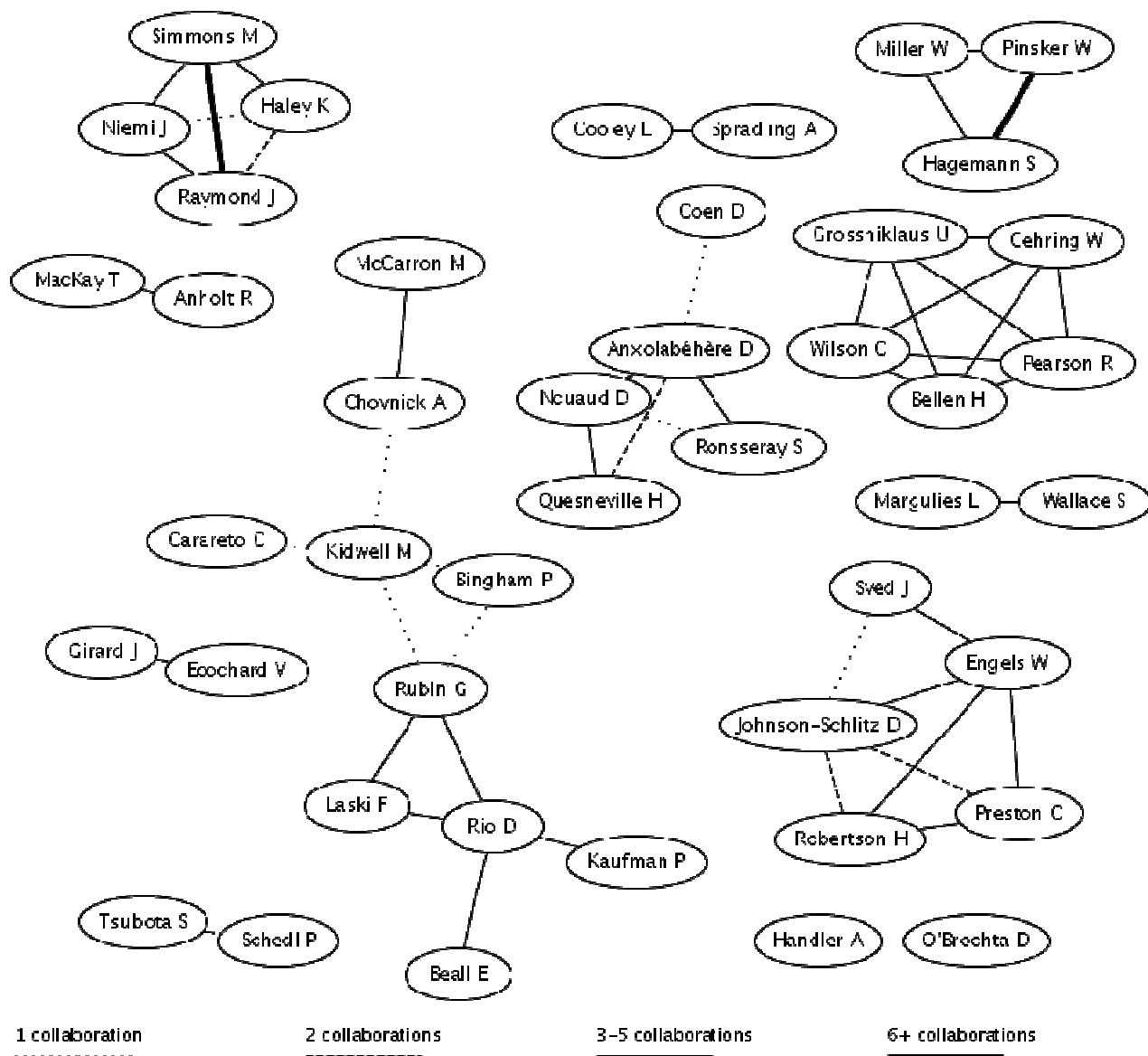
**Author networks for the scientific publications on each transposon system
(April 2009)**

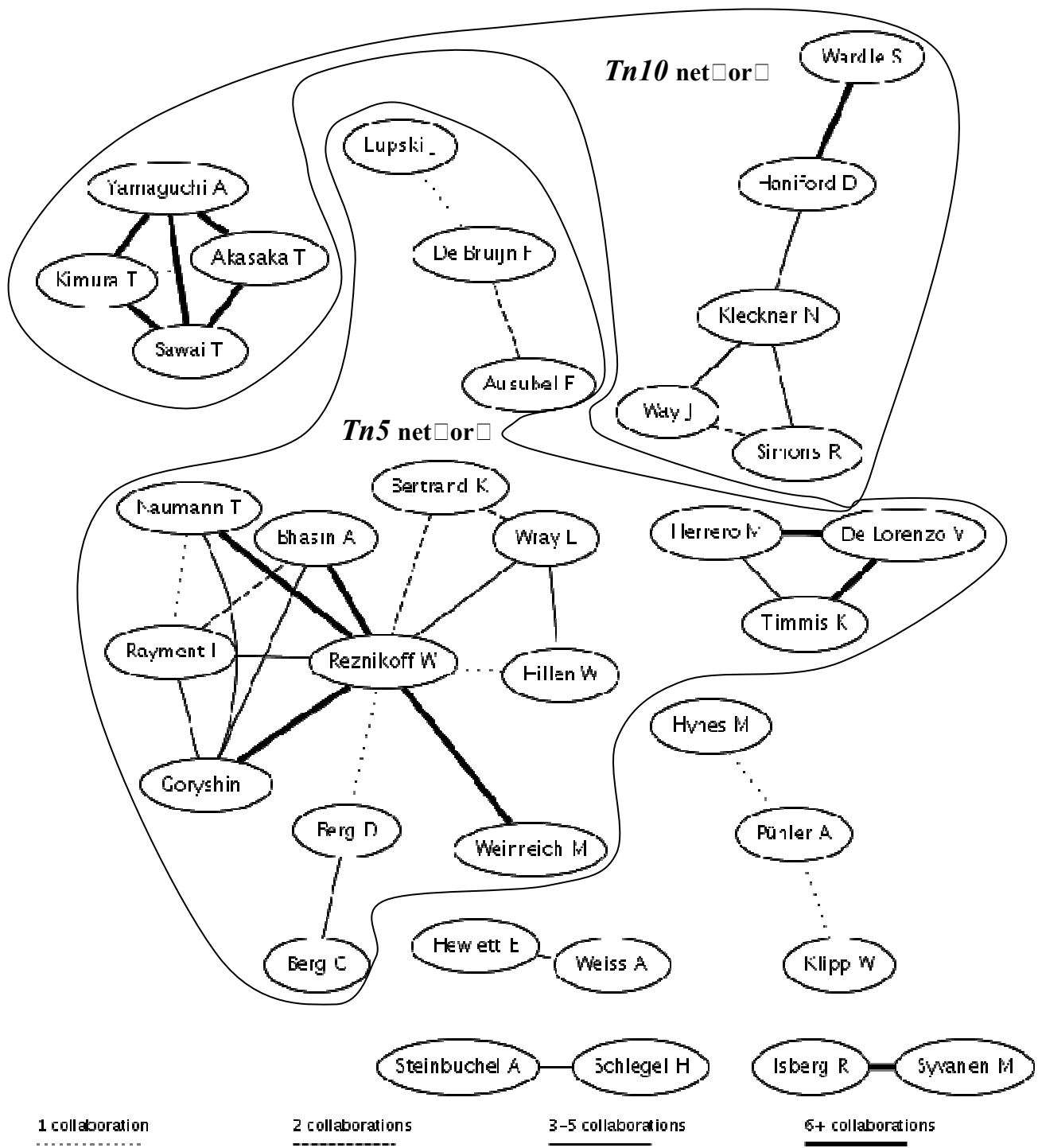
Mariner-like elements. Mos1, Himar, Hsmar





Pelements





3.3. Article 4 : *Zinc finger monopoly : quelles sont les règles du jeu ?*

3.3.1. Contexte et objectif de l'article

L'objectif de cibler l'intégration chromosomique de transgènes avec des vecteurs intégratifs site-spécifiques est principalement motivé par la nécessité d'éviter les conséquences délétères des intégrations aléatoires dans les cellules hôtes. En effet, les systèmes d'intégration rétroviraux et les systèmes non viraux du type ϕ C31 et *Sleeping Beauty* peuvent réaliser des intégrations au sein même des gènes pouvant amener à leur activation ou inactivation [Izsvák & Ivics, 2004 ; Calos, 2006]. Dans le programme de travail initial de SyntheGeneDelivery, le ciblage de l'intégration du transgène par le vecteur transposon devait être développé à partir de domaines de liaison à l'ADN de type doigt de zinc (*Zinc Finger* ou ZF). De janvier à juin 2007, j'ai effectué mon stage de seconde année de Master Professionnel Biotechnologies & Droit de l'Université de Tours dans l'équipe 2 de l'actuel GICC. Dans le cadre de ce stage, j'ai été amené à déterminer l'impact de la PI existant sur les mécanismes d'intégration non-viraux dérivés de transposons et de leur modification pour devenir site-spécifique. Mon objectif principal était de déterminer la PI protégeant la technologie basée sur les domaines ZF et d'identifier les propriétaires et les exploitants. Nos résultats ont permis de conclure que la société Sangamo Biosciences détient le contrôle de la technologie ZF avec 79 familles de brevets et plusieurs licences. Son portefeuille de titres de PI lui permet de se positionner comme l'acteur dominant incontournable pour le contrôle de la technologie ZF et ses multiples exploitations (nucléases, facteurs de transcription artificiels, criblage...). À la fin de l'année 2008, nous avons commencé à travailler sur la rédaction d'un article qui présenterait les résultats du stage. Celui-ci devait compléter l'article *The zinc finger nuclease monopoly* qui constatait la position dominante de la société Sangamo Biosciences sur la technologie des nucléases chimériques site-spécifiques [Scott, 2005]. Cependant, l'article *Proprietary science, open science and the role of patent disclosure: the case of zinc-finger proteins* a été publié au début de l'année 2009. Il présente un examen de la situation de la PI aux États-Unis couvrant l'ingénierie et l'utilisation des protéines ZF et l'impact que cela a sur la R&D des acteurs industriels et académiques

[Chandrasekharan *et al.*, 2009]. Dans cet article, les auteurs ont construit un paysage brevets de la technologie ZF à partir des demandes de brevets et des brevets déposés aux États-Unis.

Nous avons donc décidé d'approfondir notre étude et de proposer un article analysant le paysage brevets de la technologie ZF à l'échelle mondiale. L'objectif de l'article *Zinc finger monopoly : quelles sont les règles du jeu ?* porte donc sur l'impact de la situation de monopole mondial de la société Sangamo Biosciences et sur la liberté d'exploitation des acteurs du domaine.

3.3.2. Résumé de l'article

La technologie ZF se base sur l'utilisation de domaines de liaison à l'ADN de type doigt de zinc, capables de reconnaître et de se lier spécifiquement à un motif d'ADN de trois paires de bases. Ces domaines doigt de zinc présentent une bonne modularité, à la fois dans leur structure et dans leur fonction. L'assemblage de plusieurs domaines ZF en modules offre la possibilité de reconnaître une séquence nucléotidique plus longue. Par exemple, l'association de six domaines ZF permet de cibler une séquence de dix-huit paires de bases, c'est-à-dire un motif nucléique statistiquement unique dans le génome humain. De plus, les domaines ZF peuvent être fusionnés à d'autres domaines fonctionnels comme ceux des endonucléases [Palazzoli *et al.*, 2008], d'intégrases [Tan *et al.*, 2006] et même de facteurs de transcription artificiels [Rebar *et al.*, 2002]. La technologie ZF est donc prometteuse et permet de répondre à l'un des plus grands enjeux du transfert de gène : le ciblage d'une modification génétique. Elle a fait l'objet de deux articles qui illustrent l'impact des brevets couvrant l'ingénierie, les applications et les conséquences de la stratégie de la société Sangamo Biosciences sur la recherche académique et privée aux États-Unis [Scott, 2005 ; Chandrasekharan *et al.*, 2009]. Nos résultats confirment la position dominante de Sangamo Biosciences à l'échelle mondiale. Les stratégies de PI (brevets et licences) et de collaboration ont assuré à la société le contrôle exclusif à l'échelle mondiale de toute la R&D incluant l'ingénierie et toutes les applications de ces domaines de liaison à l'ADN. La situation de blocage est légèrement moins importante en dehors des États-Unis (Europe, Japon...). Néanmoins, les brevets fondateurs de la technologie ZF ont été étendus de manière importante et assurent une large couverture géographique. Par ailleurs, pour contrer la

stratégie de Sangamo Biosciences qui limite l'accès à l'information contenue dans les brevets, certains acteurs ont recherché des alternatives comme le contournement et la révocation de brevets, avec plus ou moins de succès. D'autres se sont rassemblés au sein du Zinc Finger Consortium pour proposer un projet *open source* collaboratif qui favorise la publication et le partage des connaissances, à l'inverse de la stratégie de la firme américaine. Finalement, un projet de R&D basé sur la technologie ZF peut être limité par la position de monopole que Sangamo Biosciences a su mettre en place dans le monde entier, et y échapper semble difficile. Par conséquent, devant le peu de liberté d'exploitation, d'autres stratégies doivent être envisagées telle que la mise au point d'une technologie de rupture comme a pu l'être la technologie ZF il y a une quinzaine d'années.

3.3.3. Conclusion de l'article

L'analyse du paysage brevets de la technologie ZF montre clairement que la société Sangamo Biosciences est en situation de monopole et qu'elle contrôle toutes les applications des domaines ZF. La société a su regrouper au sein de son portefeuille, tous les brevets fondateurs de la technologie de fabrication à façon de protéines ZF chimériques, ce qui bloque toute stratégie ou utilisation liée aux domaines ZF. Bien que la technologie ZF représente l'une des solutions les plus prometteuses pour réaliser l'insertion site-spécifique d'un fragment d'ADN, la liberté d'exploitation est très limitée. Malgré cet inconvénient, des acteurs majeurs des transposons ont développé et protégé des protéines de fusion comprenant un domaine de liaison à l'ADN ZF et un domaine fonctionnel de transposase, ce qui était prévu dans le projet SyntheGeneDelivery [Palazzoli *et al.*, 2008 ; Palazzoli *et al.*, 2010a ; WO2002008286, WO2004009792, WO2004070042...]. Cette situation a restreint l'intérêt de développer une transposase fusionnée aux domaines ZF. Les conséquences sur l'évolution du projet SyntheGeneDelivery a été la réorientation des recherches des équipes 1 et 2 du GICC, avec le développement d'une transposase chimérique site-spécifique qui ne soit pas basée sur la technologie ZF.

3.3.4. Conséquences pour la suite de mes travaux de thèse et le projet SyntheGeneDelivery

Du fait du monopole sur la technologie ZF, des solutions alternatives ont été recherchées à partir de juin 2007 pour envisager autrement le ciblage spécifique d'un transgène dans les chromosomes. L'une des niches de développement technologique est présentée par la demande PCT WO9424300 - *Transposition assembly for gene transfer in eukaryotes* [Palazzoli *et al.*, 2010a]. Cette invention tombée dans le domaine public porte sur l'intégration d'un vecteur dérivé d'un rétrotransposon non-LTR [Rangwala & Kazazian, 2009] dans les gènes des ARN ribosomiaux. Cette stratégie alternative qui consiste à cibler indirectement la transposase ou le transposon à un locus génomique spécifique est présentée dans l'article 5 - *Characterization of monomeric protein domains binding specifically to a highly conserved 100-bp target within rDNA genes* [Carnus *et al.*, 2011].

3.3.5. Remarque sur la méthodologie et les outils employés

Contrairement aux articles précédents 2 et 3 [Palazzoli *et al.*, 2008 ; Palazzoli *et al.*, 2010a], j'ai utilisé l'outil commercial QPAT pour les recherches de documents brevets. J'ai recherché les mots clés « *zinc ET finger* », dans les titres, les résumés et les revendications indépendantes. Cependant, l'outil QPAT a évolué récemment et ne se limite plus aux revendications indépendantes. Il est désormais possible de faire des recherches dans toutes les revendications, indépendantes et dépendantes. Cette nouvelle fonctionnalité améliore la quantité de brevets qui peuvent être exploités pour construire des paysages brevets.

Ainsi, en effectuant la même requête « *zinc ET finger* » dans toutes les revendications, il est possible de trouver de nouveaux documents brevets pertinents comme le brevet US20060252140 - *Development of a transposon system for site-specific DNA integration in mammalian cells*. Il revendique dans la revendication indépendante 3 : « *A source of transposase activity comprising a fusion protein comprising a Sleeping Beauty transposase and a site-specific DNA binding protein* » et dans la revendication dépendante 4 : « *The source of claim 3, wherein the site-specific DNA binding protein is a*

zinc-finger DNA binding protein ». Ainsi, le brevet US20060252140 apparaît dorénavant dans les résultats de QPAT, grâce à la recherche dans toutes les revendications.

3.3.6. Article 4

Biobusiness

Zinc finger monopoly : quelles sont les règles du jeu ?

Le ciblage de séquences d'ADN spécifiques par des domaines de liaison de type doigt de zinc est une technologie prometteuse et protégée par de nombreux brevets. Deux articles ont illustré l'impact de la stratégie de la société Sangamo Biosciences sur la recherche académique aux États-Unis. La situation est-elle comparable en Europe ? Pour les instituts travaillant dans ce domaine, quelle est la liberté d'exploitation de leurs découvertes ?

*1 Protéines naturelles très étudiées pour leurs doigts de zinc
*2 Représentation graphique de la liste de brevets d'une technologie, permettant de comprendre les axes de recherche concurrents et ainsi de mieux définir ses stratégies de recherche

(1) Miller J *et al.* (1985) *EMBO J* 4, 1609-14

(2) Rebar EJ *et al.* (2002) *Nat Med* 8, 1427-32

(3) Ivics Z *et al.* (2007) *Mol Ther* 15, 1137-44

(4) Tan W *et al.* (2006) *J Virol* 80, 1939-48

(5) Gordley RM *et al.* (2007) *J Mol Biol* 367, 802-13

(6) Porteus MH, Carroll D. (2005) *Nat Biotechnol* 23, 967-73

(7) Pavletich NP, Pabo CO (1991) *Science* 252, 809-17

(8) Diakun GP *et al.* (1986) *Nature* 324, 698-9

(9) Choo Y, Isalan M (2000) *Curr Opin Struct Biol* 10, 411-6

(10) Scott CT (2005) *Nat Biotechnol* 23, 915-18

Un des enjeux des technologies développées pour le transfert de gène est de pouvoir modifier spécifiquement l'expression des gènes. Dans le contexte de la thérapie génique, ce ciblage peut concerner des gènes impliqués dans le déterminisme de maladies génétiques acquises (cancers) ou innées. Plusieurs technologies proposant de cibler spécifiquement l'intégration ou la correction d'un gène d'intérêt reposent sur l'utilisation de domaines de liaison à l'ADN de type doigt de zinc (*zinc finger domain*, ZFD) (1). Elles sont fondées sur l'utilisation de protéines recombinantes fabriquées à façon, à partir de ZFD dont la liaison est spécifique d'un motif d'ADN et de domaines activateurs ou répresseurs de transcription (2), de transposases (3), d'intégrases (4) et de recombinases (5). Ces domaines de liaison à l'ADN ont également été beaucoup utilisés dans des fusions avec des nucléases (figure ci-contre) (6).

Caractéristiques de trois principales familles de ZFD naturels

- Les ZFD les plus répandus sont de type Cys₂-His₂ (noté aussi C2H2). Leur séquence consensus est de type (Tyr,Phe)-X-Cys-X₂-Cys-X₃-(Tyr,Phe)-X₅-Leu-X₂-His-X₃-His, dans laquelle X représente n'importe quel acide aminé et les acides aminés notés en gras chélatent l'ion Zn²⁺ afin de stabiliser la molécule.
- Certains ZFD, ceux de type Cys₂-Cys₂ (notés aussi Cys₄), ont la propriété de chélater l'ion Zn²⁺ en impliquant 4 résidus cystéine. Leur séquence consensus est : Cys-X₂-Cys-X₃-Cys-X₂-Cys. Ces domaines sont principalement présents dans les récepteurs stéroïdes nucléaires ou hormonaux comme le récepteur des glucocorticoïdes.
- Les ZFD de type Cys₆ ont une séquence consensus caractérisée par la présence d'un motif regroupant 6 cystéines Cys-X₂-Cys-X₆-Cys-X₆-Cys-X₂-Cys-X₆-Cys. Ils sont retrouvés dans des protéines correspondant à des régulateurs métaboliques comme le domaine Gal4 de la levure.

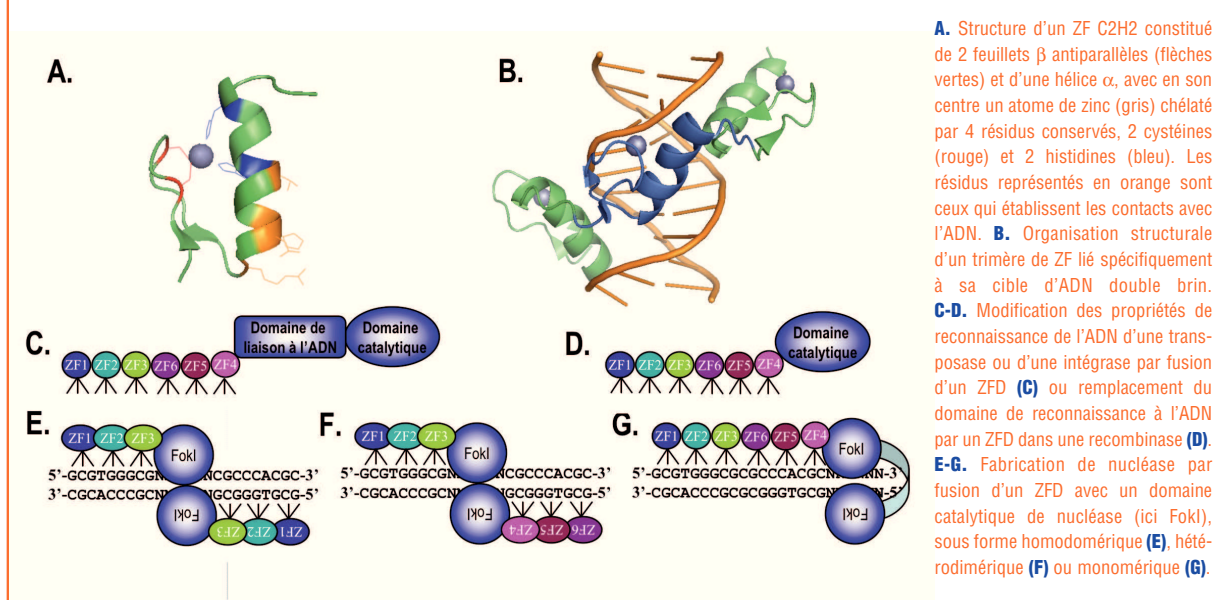
Chez les eucaryotes, les protéines possédant des ZFD sont fréquemment impliquées dans la régulation de l'expression des gènes. Les motifs ZFD qu'elles contiennent sont principalement classés en trois grandes familles sur la base de leur structure et des résidus impliqués dans la chélation d'un ion zinc (encadré). Les ZFD utilisés pour le ciblage sont de type Cys₂-His₂. Les modalités de leur reconnaissance spécifique de l'ADN ont été élucidées lors d'études cristallographiques des domaines murin et amphibien, Zif268 et TFIIIA*¹ (7,8). Un module en doigt de zinc (ZF) est ainsi capable de reconnaître spécifiquement un motif d'ADN de 3 pb. Dans le trimère Zif268, l'association de trois ZF permet d'étendre la spécificité de reconnaissance à un motif ADN de 9 pb. Par conséquent, l'assemblage de plusieurs ZF ayant chacun des spécificités différentes de reconnaissance de l'ADN permet de cibler des séquences nucléotidiques plus longues. Il est ainsi possible de fusionner jusqu'à 6 ZF pour reconnaître spécifiquement un motif de 18 pb, une taille suffisante pour être unique dans le génome humain. L'obtention par sélection de ZF capables de reconnaître chacun des 64 triplets de l'ADN a permis de développer une ingénierie de fabrication à façon de ZFD capables de reconnaître n'importe quelle séquence nucléotidique (9).

Suscitant de nombreux espoirs, cette technologie prometteuse a fait l'objet de deux articles sur la situation de la propriété intellectuelle (PI) aux États-Unis. Celui de Christopher Scott (10) introduit la situation de monopole que Sangamo Biosciences a su mettre en place à travers une stratégie basée sur la PI et le secret, deux éléments indispensables dans l'intense compétition internationale scientifique et économique. L'article conclut en proposant deux solutions complémentaires pour casser le monopole : réaliser un paysage brevet du domaine*²

Fabien Palazzoli^{*,**}, Élodie Carnus^{***}, Frank Merly^{***}, Sylvain Goiran^{***}, Yves Bigot^{*,****}

* Université François Rabelais de Tours, GICC, ** CNRS, UMR 6239, *** France Innovation Scientifique & Transfert SA, 83, Boulevard Exelmans, 75016 Paris
 **** CHRU de Tours, UFR des Sciences & Techniques, Parc de Grandmont, 37200 Tours bigot@univ-tours.fr

Caractéristiques structurales des ZF et ZFD et utilisations en ingénierie génétique



proposé par l'article de Subhashimi Chandrasekharan et de ses collaborateurs ⁽¹¹⁾ de l'université de Dubaï et en exploiter les faiblesses à travers un projet collaboratif *open source*^{**}. Ces deux publications posent la question de l'impact de la stratégie de la société sur la recherche académique et privée. Mais hors des États-Unis, quelle est la situation ?

Paysage brevet de la technologie zinc finger

La recherche de brevets a porté sur la période 1985-2007, les années 2008 et 2009 n'étant pas prises en compte car la publication des demandes de brevets s'effectue 18 mois après leur dépôt. La recherche des mots-clés « zinc finger » dans les résumés, titres et revendications indépendantes des brevets de la base de données QPAT a permis de localiser 557 familles de brevets^{**} impliquant l'utilisation de ZFD.

L'intervalle étudié, 1988-2007, peut être découpé en plusieurs périodes (figure p. 60). La période pionnière correspond aux années suivant la découverte des ZFD (1988-1993, 20 brevets). Jusqu'en 1993, aucune application n'est décrite. En 1994, la demande de brevet EP0682699 – *Functional domains in Flavobacterium Okeanokoites* (FokI) restriction endonuclease, déposée par l'université Johns Hopkins, fut la première à revendiquer la fusion du domaine catalytique de la nucléase FokI et de ZFD, marquant ainsi le début de la technologie d'ingénierie génétique utilisant des nucléases site-spécifiques. Au cours de la période intermédiaire 1994-2001, le nombre de demandes de brevet a augmenté exponentiellement (275 brevets). Parallèlement, le nombre de déposants s'est accru, les acteurs

principaux du développement de la technologie ZF se positionnant dès cette époque pour former les principaux déposants actuels (figure 3).

À partir de cette période, les brevets se distinguent en deux catégories : les brevets fondateurs de l'ingénierie des ZFD (conception, sélection, optimisation) et ceux concernant ses applications (régulation de l'expression et modification de la séquence codante de gènes). Enfin, au cours de la période 2002-2007, le nombre de brevets diminue puis stagne (262 brevets), indiquant soit une maturité de la technologie, soit un désintérêt des acteurs du fait du verrouillage du secteur. De fait, en analysant le nombre de dépôts réalisés par des industriels ou par des institutionnels, on observe un désintérêt marqué de la part des industriels à partir de 2002 qui est compensé en partie par de plus nombreux dépôts institutionnels. Le principal industriel continuant ainsi à déposer sur ce domaine étant Sangamo Biosciences.

Un acteur majeur incontournable : Sangamo Biosciences

Sangamo Biosciences se définit comme un leader mondial dans la conception, le développement et l'ingénierie des protéines de liaison à l'ADN de type doigt de zinc, pour la régulation et la modification de gènes ⁽¹²⁾. Créée en 1995, cette société a initié son activité en exploitant une licence exclusive sur l'utilisation des ZF nucléases FokI en recombinaison homologue. Sangamo Biosciences a également acheté la société anglaise Gendaq évaluée en 2001 à 30 M\$. Elle a profité de cette opération stratégique pour se renforcer avec une équipe de recherche leader dont les travaux sont à l'origine

⁽¹¹⁾ Chandrasekharan S et al. (2009) *Nat Biotechnol* 27, 140-4

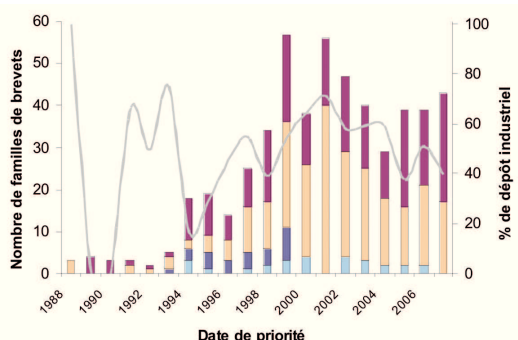
⁽¹²⁾ www.sangamo.com/about/about.html

^{**} Notion définie par l'Office mondial de la propriété intellectuelle comme un modèle de partage de technologies définissant des « champs d'innovation comprenant une plate-forme technologique que les tiers peuvent librement utiliser et adapter et sur la base de laquelle des innovations peuvent être à leur tour partagées »

^{**} 287 familles de brevets supplémentaires présentant la particularité de désigner comme inventeurs Y. Mao et Y. Xie ont été exclues car elles portent sur des protéines naturelles et humaines contenant des ZFD, dans un contexte d'exploitation totalement différent.

Biobusiness

Évolution des dépôts de brevets sur la technologie Zinc Finger au cours de la période 1988-2007



Dans chaque barre de l'histogramme est représenté le nombre de dépôts soumis par des organismes institutionnels (pourpre), des industriels (beige) et autres (bleu clair). Les brevets codéposés par des organismes institutionnels et des industriels sont indiqués en bleu foncé. La courbe en gris représente la cinétique d'évolution du pourcentage de dépôt industriel au cours de cette période.

© DR

de l'ingénierie des ZFD (1), et acquérir les titres de PI fondateurs de la technologie. Cette acquisition a été déterminante pour Sangamo Biosciences, éliminant son seul concurrent direct, renforçant son portefeuille de brevets et élargissant son rayon d'application à la régulation de gènes avec des facteurs de transcription artificiels.

Cette position dominante est également liée au fait que Sangamo Biosciences a su acquérir les brevets et des droits d'exploitation exclusifs et mondiaux qui étaient dispersés entre les autres acteurs majeurs (*Scripps Research Institute*, *Medical Research Council* de Londres, *Massachusetts Institute of Technology*, *Johns Hopkins University*...). Ils concernent la conception et l'optimisation des ZFD (bibliothèques de ZF, méthodes de construction, *linkers*^{**}), qui sont à la base de toute application en découlant, plaçant ainsi la société comme seul interlocuteur pour leur exploitation (10-11).

Ce portefeuille de brevets et licences met la société en position de force pour négocier des contrats à des conditions avantageuses. Diverses licences ont été accordées pour des applications thérapeutiques (à Merck pour la découverte et le développement de nouveaux médicaments en 2000, Pfizer pour la production de protéines recombinantes et d'anticorps monoclonaux en 2008) ou dans d'autres domaines (Dow AgroSciences pour des plantes en 2005, Sigma-Aldrich pour des réactifs de recherche en 2009). Cependant, Sangamo Biosciences reste décisionnaire dans le processus de sélection des sociétés avec lesquelles collaborer. L'exemple de la société Phytodyne illustre bien cet état de fait : Daniel Voytas, spécialiste de l'utilisation des ZF nucléases pour cibler les gènes chez les plantes, n'a pu négocier à un prix abordable l'obtention des licences indispensables à la survie de la société (10,11). Peu après, la société Dow AgroSciences obtenait un droit d'exploitation exclusif pour utiliser la technologie ZF en cellules végétales contre la somme de 50 M\$ (2005).

^{**} Chaîne de résidus d'acides aminés reliant deux ZFD

(13) Mandell JG, Barbas CF (2006) *Nucleic Acids Res* 34, W516-23

Pour Sangamo Biosciences, cette collaboration était l'opportunité d'entrer sur un nouveau champ d'application : l'ingénierie génétique des végétaux.

Comme exposé par les articles de Chandrasekharan et Scott, l'autre moyen utilisé par Sangamo Biosciences pour protéger son savoir-faire d'ingénierie des ZFD contre de potentiels contrefacteurs, serait de ne divulguer que le strict nécessaire au dépôt d'un brevet et de limiter l'accès aux informations essentielles pour la réalisation des inventions protégées. De plus, la commercialisation des ZF nucléases par Sigma-Aldrich ne permettrait pas non plus d'avoir un accès direct à la plate-forme.

Synthèse de ZFD à façon : outils et sites internet en accès libre

Pour fabriquer des domaines protéiques composés de trois ZF, plusieurs approches sont actuellement possibles. Trois stratégies de sélection *in vitro* sont décrites dans la littérature. La première, nommée sélection parallèle, consiste à sélectionner individuellement chacun des motifs ZF indépendamment les uns des autres, puis à les fusionner en utilisant un *linker* peptidique canonique. Si cette approche est rapide et accessible à la plupart des laboratoires, son efficacité reste limitée car elle ne prend pas en compte l'effet des interactions entre ZF adjacents sur la fonctionnalité du trimère.

La deuxième méthode, ou sélection séquentielle, contourne ce problème en sélectionnant chaque ZF dans le contexte des deux autres. Bien que les trois ZF obtenus de cette manière présentent une affinité très élevée pour le site choisi, cette stratégie est difficile d'accès car elle nécessite la mise en place de nombreuses banques.

La troisième approche combine les avantages des deux premières. Elle permet la sélection de chaque ZF en tenant compte des deux autres tout en limitant le nombre de banques requises. Cette stratégie bipartite propose d'utiliser deux banques ZF précréées avec, dans chacune, la moitié des trois motifs ZF partiellement « randomisée » au niveau des acides aminés clés en contact avec l'ADN.

Enfin, il existe une quatrième approche dont le détail des modalités techniques ne serait pas dévoilé par Sangamo Biosciences, notamment dans les brevets correspondants. On sait seulement de cette approche qu'elle permet de sélectionner des domaines à quatre doigts de Zinc reconnaissant ainsi un motif de 12 pb, la sélection s'effectuant par paire de ZF. Sur la base des résultats obtenus en sélection parallèle, des plates-formes *in silico* ont été développées et plusieurs sites internet permettent de concevoir des ZFD. Le laboratoire de Carlos Barbas (*Scripps Research Institute*) a ainsi été le premier à développer un site internet, zincfinger-tools.org, qui permet aux chercheurs d'identifier les motifs nucléiques susceptibles d'être ciblés par un ZFD dans une séquence d'ADN et donc de concevoir des ZFD spécifiques de cette région (13). L'accès à ce site est libre mais son utilisation nécessite l'ouverture d'un compte.

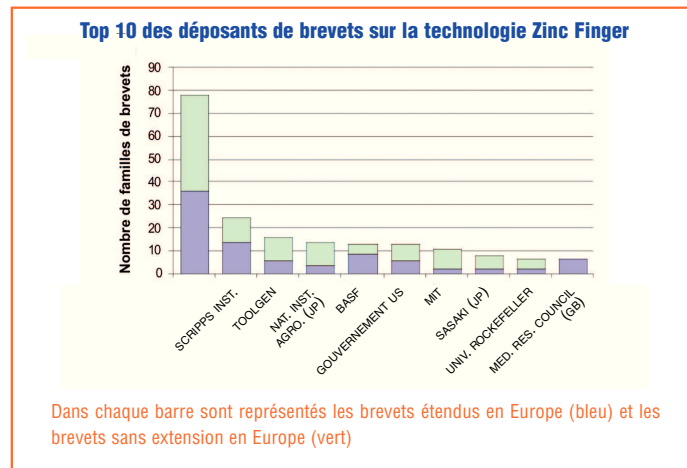
Dans la pratique, il s'avère que la position de chaque ZF au sein du ZFD influence fortement la spécificité de reconnaissance de la cible. Deux autres sites internet prennent en compte ces contraintes, ZiFiT (14) et ZiFDB (15). Leurs calculs s'appuient sur une base de données répertoriant les données publiées sur les ZFD utilisés et ceux présents dans les bases. Ces sites sont régulièrement mis à jour et leur accès est totalement gratuit. Cependant, l'affinité et la spécificité de liaison des ZFD conçus et optimisés in silico restent largement à démontrer et, bien que requérant plus de moyens, les approches in vitro sont actuellement considérées comme les plus efficaces.

Le Zinc Finger Consortium

Devant les limites des approches in silico et les difficultés pour accéder aux approches in vitro les plus efficaces développées par Sangamo Biosciences, les chercheurs Keith Joung (*Massachusetts General Hospital* et *Harvard Medical School* de Boston) et Daniel Voytas (*Iowa State University*, Ames) ont fondé le *Zinc Finger Consortium*^{**}, qui rassemble une quinzaine de laboratoires académiques. Le but de cette initiative est de promouvoir la R&D de la technologie ZF. En effet, le consortium souhaite rendre cette technologie accessible, en mettant régulièrement et librement à la disposition de la communauté scientifique des outils in vitro et in silico (ZiFiT et ZiFDB) et des protocoles de sélection de ZFD.

Dans un article récent, le consortium propose aussi une méthode d'ingénierie *open source* rapide et accessible au public dénommée OPEN (*Oligomerized Pool Engineering*) offrant aux chercheurs la possibilité de « faire progresser la technologie des ZFD » (16). De même, les plasmides utilisés par le consortium sont mis à disposition par Addgene, une organisation à but non lucratif qui facilite la diffusion de plasmides dans la communauté scientifique. Les méthodologies, informations et outils partagés par le consortium sont *open source* et donc non confidentiels. Mais, pour toute autre utilisation que la recherche, tout tiers souhaitant réaliser, utiliser, distribuer ou vendre la technologie ZF nécessite le consentement préalable et les licences appropriées auprès de Sangamo Biosciences : il n'y a pas d'*open access* aux droits de PI et à l'exploitation de la technologie. C'est pourquoi une page du site d'Addgene stipule que certaines utilisations de plasmides du consortium, portant sur des modules ZFD de Sangamo Biosciences nécessitent une licence d'exploitation (WO0153480, EP1364020, etc) (17).

L'intérêt ici n'est pas de contourner les brevets de Sangamo Biosciences, mais d'avoir un accès à la plate-forme et à son savoir-faire. Malgré le fait que certains modules semblent porter atteinte à divers brevets, Sangamo Biosciences n'a jamais empêché leur distribution ni fait valoir ses droits pour contrefaçon. En effet, cette plate-forme est également l'occasion de partager des innovations au travers de nouvelles applications (trois membres sont experts en manipulation génétique de cellules souches). Les diffusions de connaissances scientifiques (articles, communications



orales, sites internet) sont aussi un moyen efficace pour Sangamo Biosciences d'avoir de la publicité gratuite pour sa technologie de liaison à l'ADN.

Il est intéressant de constater que Sangamo Biosciences profite de l'exclusivité des exploitations de la technologie ZF sur la base de brevets, tout en limitant les possibilités à l'homme de l'art de reproduire les résultats divulgués. Paradoxalement, l'existence du ZF consortium et la distribution de ZFD à des prix abordables permettent ainsi aux chercheurs de valider les brevets. En effet, le consortium commercialise des réactifs basés sur des ZFD au prix de 5 000 \$. Ce tarif est plus accessible pour la plupart des laboratoires académiques en comparaison des 25 000 \$ réclamés par Sigma-Aldrich pour des réactifs utilisables à des fins de recherche uniquement. Cependant, en faisant cela, les membres du consortium seraient contrefacteurs de plusieurs brevets déposés par la firme qui ne s'est pas pour autant manifestée. Ce point pourrait évoluer prochainement si cette commercialisation devenait rentable et/ou si Sigma-Aldrich considérait que Sangamo Biosciences ne lui assure pas la libre jouissance de ses brevets. Il sera intéressant de suivre l'issue juridique d'une action en justice si la firme Sangamo Biosciences faisait valoir ses droits de PI.

Implication pour les acteurs américains et européens

On assiste ainsi à un système où les chercheurs du secteur public utilisent des technologies protégées par des brevets et sont donc potentiellement contrefacteurs, mais où les sociétés détentrices des brevets ou des licences n'imposent pas le respect de leurs droits. L'article de Chandrasekharan (11), qui repose sur une étude des brevets et licences détenus par Sangamo Biosciences outre-Atlantique, synthétise très bien le dilemme pour les acteurs américains, notamment publics. En étendant cette analyse aux extensions mondiales (figure p. 62), on constate un blocage légèrement moins étendu en Europe

(14) Sander JD *et al.* (2007) *Nucleic Acids Res* 35, W599-605

(15) Fu F *et al.* (2009) *Nucleic Acids Res* 37, D279-83

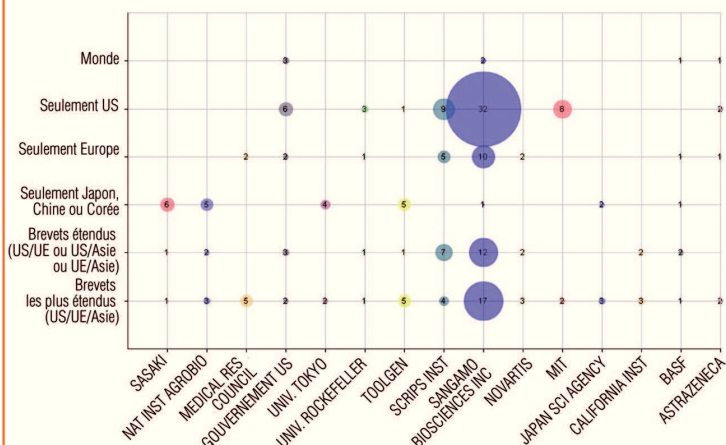
(16) Maeder ML *et al.* (2008) *Mol Cell* 31, 294-301

(17) www.addgene.org/pgvec1?f=a&cmd=showfile&file=sangamo

^{**} www.zincfingers.org

Biobusiness

Densité des extensions des familles de brevets soumis par les déposants (axe horizontal) en fonction des régions géographiques (axe vertical)



La taille du cercle est proportionnelle au nombre de familles étendues dans la région concernée.

et au Japon, avec notamment plus de la moitié des dépôts de Sangamo Biosciences non étendus en dehors des États-Unis.

Les brevets fondateurs, principalement de Sangamo Biosciences (incluant ceux de la société Gendaq), de l'université Johns Hopkins et du *Scripps Institute* ont toutefois été étendus de manière importante. Si les brevets déposés les plus importants assurent une couverture large de l'invention (c'est-à-dire Europe, Japon, Canada et États-Unis), plusieurs institutions, comme le MIT (US7153949, US7595376, US6410248, US5789538, brevets d'ingénierie), n'ont pas pris la peine de protéger leurs inventions en dehors du territoire américain, créant un « trou » dans la protection des inventions licenciées à Sangamo Biosciences.

Quelles stratégies de contournement ?

Est-il possible d'échapper au monopole de Sangamo Biosciences sur la technologie ZF ? Plusieurs solutions ont été envisagées pour réaliser des plans d'exploitation avec l'ingénierie modulaire des ZFD. Plutôt que d'attendre la chute dans le domaine public des brevets protégeant la fabrication à façon des ZFD (environ une dizaine d'années), certains ont essayé de contourner techniquement les brevets. Par exemple, l'équipe du Japonais Yokio Sugiura a modifié le squelette des ZFD de Zif268, échappant ainsi partiellement à certains brevets spécifiques aux ZFD de type Cys₂-His₂ (WO9853059, US6866997). En mutant une cystéine en histidine, ou inversement, il est possible d'obtenir les types naturels Cys₄ et Cys₃-His₁ et les types artificiels Cys₁-His₃ et His₄ (18).

Il est également possible d'échapper à certains brevets qui revendiquent des ZFP comprenant des ZFD associés par des *linkers* d'au plus 10 acides aminés (7 résidus glycine étant le plus souvent utilisés), mais l'efficacité

de cette solution n'a pas été confirmée. Une autre alternative est d'être dans une situation permettant de faire révoquer partiellement ou en totalité les brevets, grâce à des éléments destructeurs de la condition de nouveauté, ou pour défaut d'activité inventive, ce qui fut le cas pour Sangamo Biosciences (révocation du brevet EP0682699, réexamen des brevets US6265196 et US5792640). Une dernière solution consiste en l'analyse par pays dans lesquels a été demandée une protection par brevets. Pour les pays où la protection est absente (pas de dépôt ou retrait), une exploitation sera possible. Bien entendu, les pays où le marché est le plus important (principalement les États-Unis) ne rentrent pas dans ce cas de figure, ce qui rend aléatoire le succès de cette stratégie pour une application commerciale.

Quelle liberté d'exploitation ?

La volonté affichée du ZF Consortium est de favoriser le développement et l'optimisation de la technologie ZF en s'affranchissant de la position dominante de Sangamo Biosciences. Paradoxalement, l'initiative tend à renforcer cette dernière, d'une part en découvrant de nouvelles applications qui restent dépendantes des brevets fondateurs, et d'autre part en encourageant des institutions publiques à investir des ressources dans ces découvertes. Sangamo Biosciences reconnaît dans ses communications que cette initiative lui est favorable. Par ailleurs, l'un des plus gros avantages de la technologie ZF provient de l'exploitation de facteurs de transcription comprenant des ZFD agissant directement sur les gènes ciblés et ne portant pas atteinte aux multiples brevets d'autres sociétés protégeant les ADN complémentaires des gènes correspondants.

Au final, échapper au monopole sur la technologie ZF semble difficile pour les acteurs américains mais également européens et japonais. Il est important pour eux de négocier en amont avec Sangamo Biosciences la mise en place d'une liberté d'exploitation sur un investissement en R&D. Deux solutions peuvent donc être envisagées. La première concerne l'utilisation de niches technologiques, qui peuvent être découvertes en étudiant les bases de données brevets. Il s'agit de technologies qui existent déjà mais qui ne sont pas ou plus protégées. Ces inventions appartiennent dorénavant au domaine public pour plusieurs raisons : elles ont été divulguées avant la demande ou l'obtention d'une protection (dans un article, une présentation orale), ou le brevet a été abandonné par son déposant (choix stratégique).

Dans des domaines équivalents, la seconde option concerne la découverte d'une innovation de rupture, une innovation technologique qui va remplacer une technologie en position dominante sur un marché, apportant de nouveaux produits sur des marchés qui n'existaient pas initialement. Cette dernière solution tire donc partie du système des brevets qui stimule le développement économique et délimite un cadre économique favorable à la concurrence. En effet, quel serait l'autre but du système des brevets si ce n'était de dynamiser l'innovation ? ●

(18) Hori Y et al. (2000) *Nucleic Acids Symp Ser* 44, 295-6

3.4. Article 5 : *Characterization of monomeric protein domains binding specifically to a highly conserved 100-bp target within rDNA genes*

3.4.1. Contexte et objectif de l'article

L'analyse du paysage brevets sur les transposons et la technologie ZF associée à l'absence de liberté d'exploitation des domaines ZF nous a menés à aborder l'intégration site-spécifique sous un nouvel angle. L'exploitation de la technologie ZF, initialement envisagée dans le projet SyntheGeneDelivery, a été abandonnée et une nouvelle stratégie de ciblage de l'ADN a été mise en place. Celle-ci repose sur l'intégration du transgène d'intérêt dans une séquence située dans l'ADN nucléaire ribosomal de la cellule cible. Cette méthode connue de longue date par les levuristes est revendiquée dans la demande internationale de brevet WO9424300 - *Transposition assembly for gene transfer in eukaryotes*, dont l'invention appartient aujourd'hui au domaine public. Cela consiste à cibler le vecteur transposon en utilisant d'une part un peptide de ciblage du vecteur dans l'ADN ribosomal et, d'autre part, une transposase chimérique capable de se localiser dans le nucléole à proximité des mêmes gènes ribosomaux. Pour cela, l'équipe 2 du GICC tire profit des propriétés des domaines de liaison à l'ADN des éléments transposables R2, dont les sites naturels d'insertion sont les séquences répétées codant pour les ARN ribosomaux.

Contrairement aux articles 2, 3, 4 et 6 basés sur les littératures brevets et bibliographique, l'article 5 est un article de recherche dont l'objectif est de caractériser un domaine de liaison à l'ADN permettant un ciblage dans l'ADN ribosomal.

3.4.2. Résumé et conclusion de l'article

Lors des expériences d'intégration d'un transgène, il est important de cibler des régions qui soient toujours accessibles pour la transcription afin d'assurer en permanence l'expression du transgène. Pour résoudre le problème d'extinction de l'expression du transgène par la machinerie des cellules hôtes, l'une des stratégies porte sur l'utilisation de

séquences répétées pas ou peu silençables en tant que site cible d'intégration. Les séquences de l'ADN ribosomique 28S sont répétées en plusieurs centaines de copies dans le génome humain. L'insertion d'un transgène dans une des copies n'a donc pas de conséquence sur le niveau d'expression de ces gènes dans une cellule ce qui garantit l'innocuité de l'intégration. Les séquences de l'ADN ribosomique 28S ne sont pas sensibles au silençage par méthylation des cytosines. De plus, ces séquences répétées essentielles à la survie des cellules sont transcrites continuellement et de manière ubiquitaire, assurant ainsi l'expression d'un transgène. Ces répétitions de gènes constituent donc d'excellentes niches génomiques pour réaliser une intégration site-spécifique, d'autant plus que cela n'altère pas la viabilité des cellules hôtes. La preuve de concept, démontrant qu'une intégration d'ADN à ce locus est garante d'innocuité et d'expression durable, a été établie en levure, protozoaire, insecte et cellules de mammifères. Cependant, les travaux ont été menés avec des vecteurs intégratifs par recombinaison homologue ayant une efficacité d'intégration trop faible (10^{-5}) pour la thérapie génique. Néanmoins, cette solution a été acceptée comme une stratégie pertinente en bioproduction depuis une trentaine d'années [Dos Santos & Buck, 2004 ; Klabunde *et al.*, 2004]. Plus récemment, ce système a également été envisagé en thérapie génique [Stewart *et al.*, 2002 ; Liu *et al.*, 2007 ; Wen *et al.*, 2008]. Quels que soient le vecteur d'intégration et la stratégie de ciblage utilisés, une transposase chimérique fusionnée à un domaine de liaison à l'ADN doit être conçue pour localiser le vecteur au site de l'intégration [Ciuffi *et al.*, 2006 ; Ivics *et al.*, 2007]. Des domaines de liaison à l'ADN doivent donc être spécifiquement développés pour réaliser le ciblage dans les gènes codant l'ARN ribosomal.

Les travaux de cet article ont porté sur la conception et la synthèse de domaines de liaison à l'ADN issus de l'endonucléase du transposon R2, et capables de cibler et de lier spécifiquement les gènes codant l'ARN ribosomal. Les résultats montrent que les prototypes peuvent être considérés comme des candidats prometteurs du fait de la haute affinité de reconnaissance envers la séquence cible. Cependant, ces prototypes nécessitent d'être optimisés, notamment par l'ajout d'un motif d'import nucléolaire. Il sera alors nécessaire de vérifier si les propriétés de reconnaissance spécifique de ce domaine sont conservées dans un contexte cellulaire. De plus, l'efficacité d'une telle stratégie de ciblage requiert de nombreuses étapes de validation, afin de confirmer la spécificité et les conséquences sur les

différents vecteurs [Ciuffi *et al.*, 2006 ; Ivics *et al.*, 2007], avant d'être considérée sans effet délétère.

3.4.3. Article 5

Characterization of monomeric protein domains that bind specifically to a highly-conserved 100-bp DNA target within rRNA genes

Authors:

Elodie Carnus[‡], Marie-Véronique Demattei[‡], Sophie Casteret[‡], Guillaume Carpentier[‡], Fabien Palazzoli[‡], Solenne Bire[‡], Christophe Bressac[§], and Yves Bigot^{‡,*}

Affiliations

[‡] UMR CNRS 6239, GICC and [§] UMR CNRS 6035, IRBI, UFR des Sciences et Techniques, Parc de Grandmont, 37200 Tours - France

Authors **e-mails:** elodie.carnus@etu.univ-tours.fr, baud@univ-tours.fr,
sophie.casteret@univ-tours.fr, carpentier@med.univ-tours.fr, fabien.palazzoli@etu.univ-tours.fr,
solenne.bire@etu.univ-tours.fr, christophe.bressac@univ-tours.fr, yves.bigot@univ-tours.fr

*** Corresponding author:** UMR CNRS 6239, GICC, UFR des Sciences & Techniques, Parc de Grandmont, 37200 Tours - FRANCE
Tel: +33 2 47 36 70 63, Fax: +33 2 47 36 70 42, e-mail: yves.bigot@univ-tours.fr

Abbreviations

Bm, *Bombyx mori*

bp, base pair

BB, binding buffer

BS, binding site

DBD, DNA binding domain

Dm, *Drosophila melanogaster*

DMEM, Dulbecco's modified eagle medium

DNA, deoxyribonucleic acid

DTT, dithiothreitol

EDTA, ethylenediaminetetraacetic acid

EMSA, electrophoresis mobility shift assay

FBS, fetal bovine serum

His, histidine

HR, homologous recombination

IPTG, isopropyl β -D-1-thiogalactopyranoside

ITR, inverted terminal repeat

GFP, green fluorescent protein

K_d, “dissociation” constant

MBP, maltose binding protein

NeoR, neomycin resistance gene

OD, optical density

ORF, open reading frame

NLS, nuclear localization signal

³²P, phosphorus ³²

PAGE, polyacrylamide gel electrophoresis

PB, *piggyBac*

PBS, phosphate buffer saline

PCR, polymerase chain reaction

PEI, polyethylenimine

pZFD, polydactyl zinc finger domain

R2, R2 non-LTR retrotransposon

RNA, ribonucleic acid

rRNA, ribosomal ribonucleic acid

SB, Sleeping Beauty

SB10, Sleeping Beauty transposase version 10

SDS, sodium dodecyl sulfate

TACH, Temperature-assisted cyclic hybridization

TBE, tris borate EDTA

Tpase, transposase

ZBS, zinc finger domain binding site

ZF, zinc finger

ZFD, zinc finger domain

Abstract (305 words)

Background. Proofs of concept have shown that chromosomal gene clusters encoding ribosomal RNA (rRNA) constitute gene delivery integration loci that are optimal for transgene expression. However, because homologous recombination is efficient to integrate DNA segments into these genes in animals, new molecular tools are required to construct systems able to target molecules in the immediate vicinity of the rRNA genes.

Results. We investigated the properties of several DNA binding domains (DBDs) able to recognize specifically a motif within a 100-bp region of the rRNA genes that is 99-100% conserved among eukaryotes. Our findings demonstrate that two *Myb*-like DBDs originating from the endonucleases encoded by R2 non-LTR retrotransposons (R2DBD) are promising candidates since they i) specifically recognize, with high affinity, a 20-bp binding site located within the expected genomic rDNA target, ii) act as monomers, iii) contain a nuclear localization signal, iv) remain functional when fused to another domain and, v) do not alter the functionality of the protein to which they are fused. However, results obtained *in vivo* with several R2DBD fusions reveal that two properties remain to be engineered before these DBDs can be integrated into a molecular targeting system directed into rRNA genes. The first concerns the ability of R2DBD to locate within the nucleolus, the organelle in which the rRNA genes reside. The second is the tendency of R2DBD to accumulate in certain parts of the nuclei, which limits its diffusion within nuclei. Solutions are discussed to circumvent these current limitations.

Conclusions. Our results supply important information concerning the R2DBD properties and the targeting of plasmid DNA within nuclei. They will need to be further analyzed from three aspects; the unexploited advantages of the R2DBDs, the possibilities and limitations of fusion peptides for targeting integrations of non-viral vector, and the alternatives to fusion peptides for targeting vectors.

Keywords: polydactyl zinc finger domain / R2 non-LTR retrotransposon / recombination / transposition / targeting

Background

The ability to use of DNA binding domains (DBDs) able to interact with a specific DNA target in order to develop synthetic proteins capable of modifying gene expression or the functioning of DNA recombination machinery, such as those involving endonucleases, integrases or transposases (Tpase), constitutes a major challenge in the post genomic era. From this perspective, a DBD added to an enzymatic system has to combine several qualities, including i) being bound to a DNA target large enough to ensure specific and high-affinity recognition, ii) having a controllable ability to assemble (or not) as homo-oligomer or with other cellular partners, and iii) having a known and modifiable propensity to locate to certain cell compartments when it has to be used *in vivo*.

Here, our focus was on the characterization of DBD candidates able to bind in monomer form to a 100-bp DNA segment that is 99-100% conserved among eukaryotes (Fig. 1b), and located in the tandemly-repeated genes that encode the ribosomal 18S-5.8S-28S RNA (rRNA, Fig. 1a). Whether intended for use in bioproduction or gene corrections in unicellular eukaryotes [1,2] or in mammalian cells [4-5], these repeated genes have been shown to be good target candidates for the insertion of DNA fragments encoding recombinant proteins. The fact that only a few rRNA genes are disrupted accounts for the absence of deleterious effects in genetically-modified cells. Indeed, genetic investigations have demonstrated that the disruption of < 70% of the rRNA genes by integration of DNA segments is not lethal and does not impair the viability of invertebrate genomes [6,7]. The expression of the transgene also persists for long periods, because integrated DNA fragments are rarely subjected to expression silencing [8], and their expression can be carried out by the RNA Polymerase I promoter, but also by RNA Polymerase II and III promoters [3,5]. However, their interest has been considerably reduced in mammalian cells because the integrative vectors functioning with passive homologous recombination (HR) have integration efficiencies ranging from 10^{-5} to 10^{-4} [3,5,8]. A way to enhance HR at this target could be to increase the concentration of DNA fragments for insertion within the nucleoli where rRNA genes are located in nuclei. Targeting of this sort might be achieved by using, for example, two fused DBDs [9] or a *bis*-PNA oligonucleotide [10] cross-linked to a DBD, these molecules being able on the one hand to bind or anneal specifically to the DNA fragment to be integrated and, on the other hand, to bind to the integration target. Vectors derived from certain transposons could also be suitable. The *Pokey* DNA transposon [11], which is an element distantly related to *piggyBac* (*PB*) [12], has been shown to integrate frequently into rDNA in its host of origin. However, *Pokey* vectors are unable to transpose in vertebrate cells (personal data). At least three innovative

approaches have been developed to target transposon vectors, such as *or piggyBac* [13] or *Sleeping Beauty (SB)* [14-16]. The first consists of fusing the T_pase to a DBD that specifically binds to the DNA integration target. The second and third approaches consist of using two fused protein domains in which the first domain specifically binds to the integration target, and the second to the transposon vector via a binding site contained in its sequence, or via an interaction with the T_pase bound to the end of the transposon (known as an inverted terminal repeat (ITR)).

Overall, whatever the species and strategy used to integrate a DNA fragment into rRNA genes, it is noticeable that it always requires a protein domain that can target the enzyme and/or the DNA fragment in the proximity of the rRNA genes. To date, we have identified only two kinds of DBD able to bind it as a monomer to a motif present within the conserved 100-bp DNA target, this motif being so large (at least 18-bp) as to be nearly unique in a mammalian genome. The first are synthetic DBDs corresponding to polydactyl zinc finger domains (pZFDs; reviewed in [17-19]). pZFDs consist of the fusion of *n* peptides from thirty amino acids, each of which specifically recognizes a three base-pair DNA motif. Research carried out by Carlos Barbas 3rd and his colleagues has made it possible to develop a single ZF able to specifically recognize all GNNs, most CNNs and ANNs, and a few TNN DNA triplets. Under these conditions, the fusion of *n* ZFDs theoretically permits the specific recognition of a 3*n* base-pair nucleotide motif (Fig. 1c and Additional file1). In practice, the background knowledge that has been accumulating for more than 10 years now enables scientists to design pZFDs *in silico* using website facilities [20-22], or to manufacture them by *in-vitro* selection [23]. The second category consists of natural DBDs that originate from the endonuclease encoded by R2 mobile genetic elements, non-LTR retrotransposons [24]. The R2 endonuclease is related to type-IIIS restriction enzymes [25]. It specifically allows the integration of R2 non-LTR retrotransposons into a 20-bp target motif located within the conserved 100-bp DNA segment contained in rRNA genes (Fig. 1b; [26,27]). The minimum DBD of this enzyme (R2DBD), without its homo-dimerization activities, has recently been characterized [28]. It consists of one ZF associated with a *Myb*-like sub-domain [29] (Fig. 1d).

In this paper, we investigate the potential of three pZFDs and two R2DBDs. We first evaluated their specific and non-specific binding site numbers in the genome of models currently used for gene therapy and bioproduction purposes. We then compared their binding specificities to about 20-bp binding sites contained within the 100-bp DNA target. Under these conditions, the best candidates were both R2DBDs. In a second step, we analyzed the

activity of fusion peptides made from R2DBD *in vitro*, and *in vivo* in HeLa cells, using integration vector systems. Finally, we have found that R2DBDs combine several properties that ensure accurate nuclear traffic to rRNA genes within the nucleolus. Potential solutions to circumvent the current problems encountered are discussed.

Results

Genomic distributions of pZFD (ZBS) and R2DBD (R2BS) binding sites

Before designing pZFD proteins able to bind specifically within the 100-bp region that is 99 to 100% conserved in the gene encoding 28S rRNA from yeasts to human beings (Fig. 1), we used an *in silico* two-step procedure to locate an accurate binding site in this DNA segment. The first step consisted of finding out whether the 100-bp region contained a candidate nucleotide motif with sequence properties that matched with an accurate ZBS. The aim of the second step was to check whether the candidate ZBS occurred only within 100-bp region in the gene encoding 28S rRNA, or whether it was also present at other chromosomal locations in the genome of models frequently used for the purposes of gene therapy (mouse, rat, and human) or bioproduction (baker's yeast and chickens).

To target a unique sequence in the human genome, searches were performed to locate an appropriate 18-bp site within both strands of the 100-bp DNA segment. This was performed taking into account the fact that the pZFD had to contain two modules with three ZFs. In consequence, the aim of our search was to locate two 9-bp binding sites that were juxtaposed or separated by 1 to 12-bp, using the facilities of the Zinc Finger Tools website. Calculations revealed only one solution that included all the features required to design a bipartite pZFD within the 100-bp region. It consists of two 9-bp motifs separated by 4 bp (Fig. 1b and c). The distribution of this ZBS among the chromosomes of five models was determined on the one hand using WAPAM software with the CGGGTAAAC-x(4)-GGAGTAACT as a model to mine the database and, on the other hand, with BLASTn using the NCBI facilities, with a word size = 7, and an expected threshold of 100 as parameters. We obtained similar results with both types of software. The number of ZBSs located outside the chromosomal loci that contain the rRNA gene clusters in each model is summarized in Table 2a. The results revealed that there were a few ZBSs present outside the rDNA loci in the mammalian species, and none was detected in the chicken, drosophila or yeast genomes.

The number of R2BS located outside the rDNA loci was also evaluated using an *in silico* approach similar to that described above (Table 2b). R2 non-LTR retrotransposons are able to integrate specifically and only in the “TAAGGTAGCCAAATGCCTCG” motif that is usually present in the genes encoding 28S rRNA. However, when identical sites are present in a genome, as in *Bombyx mori*, in which R2BS occurs outside the rDNA loci, R2 non-LTR retrotransposons can also locate to and use these other sites for integration [30]. Here, although R2BSs were less abundant than ZBS, they were also found outside the rDNA loci.

Overall, these non-specific sites represented from 1% to 10% of the binding sites for both proteins. Such occurrence levels indicate that R2BS has, on average, only half as many binding targets as pZFD in the mammal genome. Since both domains have a population of specific and non-specific binding sites with equivalent quality in the different models, and previous studies [28,30] have shown that R2DBD was able to bind to R2BS, we limited our investments in designing and optimizing pZFD to *in silico* approaches to evaluate their potential so that functional pZFDS could be obtained quickly. With such an approach, the results obtained were less satisfactory as the binding efficiency of the three synthetic pZFDS was weak, whatever the *in silico* design strategy used (Additional file1). We therefore focused our work on the properties of the two R2DBDs.

Analysis of the R2DBD of the non-LTR retrotransposon endonucleases

The sequence encoding the DBD of the endonucleases encoded by the R2 non-LTR retrotransposons contained in the genome of the fruit fly *Drosophila melanogaster* [31] and the silkworm *Bombyx mori* [32] were located as described [25]. The codon usage was optimized *in silico* to avoid some restriction sites, and to allow their expression in mammalian cells. These DBDs were designated DmR2DBD and BmR2DBD, respectively. Genes were cloned to fuse them with an amino-terminal MBP or Histidine tag. The 4 fusion proteins, MBP-BmR2DBD, His-BmR2DBD, MBP-DmR2DBD and His-DmR2DBD, were produced in bacteria and purified by affinity chromatography. Preliminary EMSA showed that the four proteins were similarly able to bind specifically to the R2site probe (Fig. 2a,d and Fig. 3a).

Influence of tagging on the EMSA complexes assembled with MBP-R2DBD

Since the same results were obtained with BmR2DBD and DmR2DBD, only those obtained with BmR2DBD are illustrated in the figures. EMSA analyses revealed that MBP-BmR2DBD and MBP-DmR2DBD assembled several different complexes with the R2site probe under our EMSA conditions (Fig. 2a, lane 2). Protein-DNA complexes were therefore assembled under similar conditions, cross-linked with UV, denatured by boiling, and analyzed by SDS-PAGE to determine how many proteins were involved in each of these complexes (Fig. 2b). The results revealed that there were three cross-linked complexes, with molecular weights indicating that they should correspond to one molecule of the R2site probe (MW = 99 kDa) linked to one protein molecule (MW = 99 + (1 x 59) = 158 kDa), two protein molecules (MW = 99 + (2 x 59) = 217 kDa), and three protein molecules (MW = 99 + (3 x 59) = 276 kDa), respectively. We also observed that the number of complexes increased with the

protein concentration, indicating that they resulted from protein oligomers, each being bound to a single R2site probe (data not shown). The fact that only one R2site probe was present in each complex was confirmed by performing short-long probe experiments, using two probes of 150-bp and 240-bp in length, one of which was radio-labeled, as previously described [33] (data not shown).

This made it impossible to determine the apparent K_d of the MBP-R2DBDs. Since fusion with the MBP allowed oligomerization to occur, the MBP tag of the MBP-BmR2DBD and MBP-DmR2DBD proteins was removed by specific site enzymatic cleavage with factor Xa at the hinge peptide between MBP and R2DBD. We first used PAGE to check that this enzymatic treatment did not degrade the peptides by proteolysis (Fig. 2c). EMSA analysis revealed that the binding activity of the BmR2DBD released was reduced 1000-fold as a result of the cleavage treatment (Fig. 2d, lane 2 *versus* 3 to 6).

The apparent K_d of the cleaved protein was determined by EMSA. The saturation graph was plotted, and used to calculate a fixed concentration of 102 nM for the BmR2DBD protein, and of 116 nM for the DmR2DBD protein (data not shown). EMSA assays were then performed using the appropriate fixed concentration of the protein with variable concentrations of the R2 site probe. Under these assay conditions, the BmR2DBD and DmR2DBD were found to be low affinity proteins with an apparent K_d of more than 1 μ M. Taking into account the data presented below, our results indicated that the release of the MBP tag did not seem to change the specific activity of either R2DBD to any great extent, but did modify their affinity, probably by destabilizing them or modifying their folding.

Properties of R2DBD

Under our laboratory conditions, and at all protein concentrations, His-BmR2DBD and His-DmR2DBD each assembled only one complex with the R2site probe (Fig. 3a). To find out how many proteins were bound to each R2site molecule, UV cross-linking assays were carried out with both R2DBDs (Fig. 3b). The molecular weight of the only cross-linked complex formed indicated that only one protein was bound to each R2site probe.

Both purified proteins were also used to determine the apparent K_d . Saturation graphs were plotted and used to calculate a fixed concentration of 6.6 nM for the His-BmR2DBD protein and of 5 nM for His-DmR2DBD (Fig. 3c and d). EMSA assays were then performed using a fixed concentration of protein and variable concentrations of the R2site. The apparent K_d of His-BmR2DBD for the R2site probe was estimated to be 30 nM, and that of His-DmR2DBD to be 18 nM (Fig. 3e and f). These two apparent K_d values were similar, and

higher than those obtained using Factor Xa cleaved proteins. Overall, they showed that the tag used had a major impact on the binding of both R2DBDs to the R2site probe, and on their ability to oligomerize.

Target specificity of the His-BmR2DBD domain produced in eukaryotic cells

We found that under our operating conditions, the activity of the His-BmDBD produced in bacteria was more stable than that of the His-DmR2DBD. It was therefore retained as the most promising candidate for investigation in HeLa cells. The size of His-BmR2DBD and its ability to bind to the R2site probe were first investigated using HeLa cell extracts in which the protein was transiently expressed. Western blot hybridized with an anti-His₆ monoclonal antibody revealed a single band with a molecular weight (18kDa), confirming His-BmR2DBD was expressed in HeLa cells (data not shown). EMSA experiments were therefore performed under conditions similar to those used with the proteins produced in bacteria. Similarly to previous findings with this protein, only one main complex was assembled with the His-DmR2DBD and R2site probe, whatever the biological system in which it had been produced (Fig. 4a, lane 2 versus 3). HeLa cell extracts were then used to determine the apparent K_d of His-BmR2DBD. Saturation graphs were plotted to fix the concentration in HeLa cell extract (Fig. 4b), and the apparent K_d was estimated to be 31 nM (Fig. 4c). This value was identical to that obtained with the protein produced in bacteria, indicating that expression in eukaryotic cells did not alter its activity.

Since we had previously observed that fusion of both R2DBDs with the MBP tag modified their ability to oligomerize, we investigated the properties of a fusion protein model that had previously been investigated as a potential way to target a nucleic acid vector close to its chromosomal integration target [14]. This consisted of fusing the BmR2DBD at its C-terminal end with a LexA DBD, and including a SV40 Nuclear Localization Signal (NLS) between the DBDs (Fig. 4d). The LexA DBD was an interesting model, because its functionality had previously been shown to depend on the properties of the domain to which it is fused [34], making it possible to find out whether His-BmR2DBD can modify the properties of a fused DBD. The profile of the R2DBD-LexA expressed in HeLa cells was verified. Western blot analysis revealed that the fusion was expressed as a single protein of about 46 kDa (Fig. 4e; theoretical molecular weight = 45.6 kDa). EMSA analyses done with the R2site probe revealed that the BmR2DBD in the fusion protein was active, but that three complexes were observed (Fig. 4f, lane 3). Similarly, EMSA indicated that the LexA part was also active, and once again, three complexes were also observed (Fig. 4g, lane 3). The

presence of three complexes was very probably not due to BmR2DBD oligomerization, but to the fact that the LexA DBD can occur as a monomer, a dimer, or a tetramer in solution [34]. In contrast to MBP fusions, saturation graphs could be calculated for these proteins, although these fusions were able to assemble in several different complexes as a result of LexA DBD oligomerization. Similar UV cross-linking investigations and short-long probe EMSA indicated that they all contained a single DNA probe to which the fusion was oligomerized (data not shown). These controls made it possible to perform affinity analyses, which revealed that the values of the apparent K_d of both domains in the fusion proteins were similar to those of each isolated domain (about 20 nM; [34]).

Impact of R2DBD fusions on DNA integration in HeLa cells

The effects of several BmR2DBD fusions were assayed on the ability of *SB* vectors to integrate into chromosomes by transposition and that of a plasmid to integrate by random recombination. Integrations by transposition and random recombination occur at different rates during an assay monitored with a transposon vector system, depending on the cell lineage used and plasmid amounts transfected [35,36]. As a consequence, a NeoR cassette can integrate by transposing with the *SB* vector into chromosomes in presence of T_{pase}. A fraction of the plasmids carrying the *SB* vector will also integrate into chromosomes as a result of random recombination, but in general this happens less often than transposition. In the control assays carried out in absence of *SB* T_{pase}, only plasmid integrations by random recombination occur.

A first set of assays was monitored with an R2DBD fused to the N-terminal end of the *SB* T_{pase} (R2DBD-SB10; Additional file2). One objective of these assays was to find out whether R2DBD-SB10 was able to mediate integration into rRNA genes. Two *SB* vectors were used (Table 1). The first vector, pT2-pSV40-NeoR, carried a NeoR cassette with expression that depended on an RNA polymerase II promoter, pSV40, and the second vector, pT2-pPol1h-NeoR, had NeoR cassette expression that depends on the human RNA polymerase I promoter, pPol1h [37]. The interest of this second construct is that pPol1h contains cryptic RNA polymerase II promoters that allow expression of NeoR cassettes integrated into nucleolar rRNA genes or elsewhere in the genome [38]. Our results indicated that R2DBD-SB10 was unable to achieve transposition, whatever the *SB* vector used (Fig. 5a and b, white bars versus dark gray bars). These data agreed with those obtained with most of the *SB* T_{pases} so far fused to a DBD [14-16]. Interestingly, we found that the numbers of NeoR clones obtained in assays monitored with R2DBD-SB10 (Fig. 5a and b, dark gray bars)

were systematically lower than those in the controls done in the absence of T_{pase} (Fig. 5a and b, black bars). Statistical analyses indicated that this difference was only significant ($p > 95\%$) when pT2-pPol1hNeoR was used as the *SB* vector source.

In a second set of assays, we checked the impact of several peptides theoretically designed to target a DNA plasmid in the proximity of the rRNA genes [9,14]. First, we analyzed the effect of R2DBD-NLS-LexA peptide on the integration ability of *SB* vectors. Two plasmid sources of *SB* vectors were used: pT2-pSV40-NeoR and pT2-pSV40-NeoR-LexA/BS, in which a LexA binding site was cloned within the *SB* vector (Table 1). Our results indicated that the R2DBD-NLS-LexA expression in HeLa cells inhibited the transposition of both *SB* vectors (Fig. 5c and d, white bars versus light gray bars). Here too, we found that the numbers of NeoR clones obtained in the absence of SB10 T_{pase}, and in the presence of R2DBD-NLS-LexA (Fig. 5c and d, dark gray bars), were statistically lower than those of the control assays monitored in the absence of both proteins (Fig. 5c and d, black gray bars). We conclude that R2DBD-NLS-LexA was able to inhibit the *SB* vector transposition, and the plasmid integration by random recombination. Because a similar effect of R2DBD-NLS-LexA was observed with both *SB* vector sources, we concluded that LexA interfered non-specifically with both plasmids independently of the presence of LexA/BS (Fig. 5c and d, top bars). Since some studies in the literature [39] have confirmed that LexA DBD has lower DNA binding specificity, we assayed a second peptide R2DBD-NLS-pZFD^{bcr-abl} [40] in which pZFD^{bcr-abl} is composed of a concatemer of four ZFs that recognizes a 12-bp motif (GACGCAGAAGCC; ZBS). ZBS is not found in the human genome, except in the genomes of most patients with chronic myelogenous leukemia (CML), and in some patients with acute lymphoblastic leukemia (ALL) or acute myelogenous leukemia (AML), in which the BCR and ABL genes are fused. Two pBS-NeoRs were used for the integration assays: pBS-NeoR and one pBS-NeoR derivative in which was cloned a ZBS to produce pBS-NeoR-ZBS. Whichever plasmid was used, our results indicated that R2DBD-NLS-pZFD^{bcr-abl} also lacked binding specificity under our experimental conditions (Additional file3). These results were confirmed for both peptides in experiments in which the pSV40-NeoR cassette was replaced by a pPol1h-NeoR cassette (data not shown).

The properties of a third peptide that theoretically works in a different way were therefore investigated. This peptide was composed of an R2DBD fused to a domain corresponding to the first 57 N-terminal residues of SB10 T_{pase} (N57SB; Additional file4). N57SB has previously been proposed to be a T_{pase}-interacting domain that is able to interact with full-length *SB* T_{pase} complex bound to the *SB* vector ITRs [9,14]. The N57SB10-NLS-

TetR peptide has also been demonstrated to be able to modify the integration preferences of *SB* vectors in the proximity of a TetR binding site [14]. For the integration assays, two ratios of the plasmids pCS2-SB10 and pCS2-R2DBD-NLS-N57SB were used. The results revealed that R2DBD-NLS-N57SB inhibited *SB* vector transposition, the intensity of inhibition depended on the amount of R2DBD-NLS-N57SB expressed in the cells (Fig. 5e and f, white bars versus light gray bars). Unexpectedly, controls done in the absence of SB10 T_{pase} indicated that R2DBD-NLS-N57SB also inhibited plasmid integration by random recombination, at least under conditions in which R2DBD-NLS-N57SB was abundant (Fig. 5f, black bars versus dark gray bars). Complementary experiments indicated that this effect was statistically significant from a plasmid ratio of 1:4 (data not shown). In agreement with structural data [41,42], our results therefore support the suggestion that N57SB is above all a DBD that binds to the *SB* vector ends (ITRs) contained in pT2-pSV40-NeoR.

In conclusion, R2DBD-NLS-LexA, R2DBD-NLS-pZFD^{bcr-abl} and R2DBD-NLS-N57SB all have the ability to inhibit *SB* vector transposition and plasmid integration by random recombination. Preliminary results obtained with *piggyBac* vectors, R2DBD-NLS-LexA and R2DBD-NLS-pZFD^{bcr-abl} led to similar conclusions, indicating that the effect was independent of the transposons used (Data not shown; only one replicate of each of the transposition assays shown in Fig. 5 was done). In agreement with these conclusions, we also found that R2DBD-NLS-LexA, R2DBD-NLS-pZFD^{bcr-abl} and R2DBD-NLS-N57SB have the ability to decrease the expression of a gene beard by a co-transfected plasmid (Data not shown). These results were obtained using an internal control of our transposition assays (see Material and Methods). We checked 24 hours post transfection that plasmids were transfected by following the transient luciferase expression of the co-transfected pGL3 plasmid. We observed that luciferase activities were all 2 to 5-fold lower in HeLa cells transfected with plasmids expressing R2DBD-NLS-LexA or R2DBD-NLS-pZFD^{bcr-abl}, 5 to 20-fold lower in those expressing with R2DBD-NLS-N57SB, and similar to the controls in those expressing R2DBD-NLS, NLS-LexA or NLS-N57SB.

Overall, these inhibitory effects did not depend on the domain fused to R2DBD or on the promoter of the marker gene since we obtained similar results with all three peptides using a pBS-NeoR plasmid. However, their effects did depend on their cellular level of expression, which was to be expected since they do not have the same affinities for DNA binding sites. None of the domains fused here to R2DBD, i.e. LexA, pZFD^{bcr-abl} and N57SB, has previously been found to inhibit transposition and recombination [14-16] when associated with another DBD or enzyme. We therefore concluded that the inhibitory effect was allocated by R2DBD

to the fusion peptides. Two hypotheses can be proposed to explain our results. The first supposes that the expression of R2DBD peptides has a toxic effect on HeLa cells, and so killed the transfected cell population. The second proposes that the nuclear trafficking of the plasmid to the target was affected, and that this impaired plasmid access to the transposition and recombination machineries.

Impact of R2DBD peptides on HeLa cells

To find out whether R2DBD peptides were toxic for HeLa cells, the effect of R2DBD-NLS was assayed on the transposition and plasmid integration of *SB* vectors. The results showed that R2DBD-NLS did not have any effect on transposition (Fig. 5g, white bars *versus* light gray bars) or recombination (Fig. 5g, black bars *versus* dark gray bars). The effects of the NLS-LexA and NLS-N57SB were also assayed. Our findings also showed that they had no effects (Fig. 5h), even though we did detect a weak effect with NLS-LexA when a plasmid containing a LexA/BS was used (Fig. 5h, black bars *versus* white bars). The toxicities of R2DBD and the peptides in which it was fused to LexA and N57SB were then verified using a viability assay evaluating cell survival during the 48 hours following the plasmid transfection (Additional file5). We found that the viability of the HeLa cells was not altered, whatever R2DBD peptide was expressed. We therefore concluded that R2DBD peptides had no toxic effect on HeLa cells under our experimental conditions, and that their effects on transposition and recombination were attributable to some other cause.

The localization of R2DBD in nuclei was verified by transfecting HeLa cells with plasmids expressing an R2DBD-NLS-GFP fusion or a variant expressing an R2DBD-GFP fusion in which the SV40NLS had been removed by mutagenesis (PKKKRK to PKAIRK [43]). The expression profiles of both peptides were compared to those of HeLa cells transfected with control plasmids expressing GFP, a protein that passively diffuses into all cell compartments, or a MOS1-GFP, a protein fusion that is actively concentrated in the nucleus (personal data). Our findings indicated that fusions of both GFPs with R2DBD were actively accumulated in the nucleus (Fig. 6a). This indicated that the R2DBD sequence contained NLS information that was efficient enough to monitor a nuclear import as strong as that mediated by SV40 NLS. PSORT II facilities at <http://psort.hgc.jp/form2.html> allowed us to identify two NLS candidates (HKRR and PMMVKRR). The findings also revealed that the nuclear distribution of the green fluorescent signal was unexpected, since it was not homogenous with either of the GFP fusions, both of which assembled fluorescent “mottling” around and between the nucleoli. This feature was in striking contrast with the results

obtained in the controls in which the GFP or the MOS1-GFP fluorescences were found to diffuse homogenously within the nuclei (Fig. 6a). This was therefore further investigated by co-expressing both R2DBD-GFP fusions with a nucleolar marker, human nucleoline [44]. The composite image created by merging the GFP and DsRED signals confirmed that R2DBD-GFP was indeed assembled as mottling organized around and between nucleoli, and that R2DBD-GFP did not co-localize with nucleolin within the nucleoli (Fig. 6b). These observations led us to propose that R2DBD was located around and between the nucleoli, and probably around other small nucleolar accessory bodies [45]. Such a location could explain why the R2DBD-GFP signal formed mottled patterns, and suggests that R2DBD has an affinity for a nuclear protein that is abundant in nuclear organelles and molecular networks ensuring connections between them. Fibrillarin is a potential candidate [46].

Discussion

The positive and negative results reported here supply important information that needs to be analyzed from four aspects concerning the R2DBD properties; the unexploited advantages of the R2DBDs, the possibilities and limitations of fusion peptides for targeting integrations of non-viral vector, and the alternatives to fusion peptides for targeting vectors.

Properties and current limitations of R2DBDs

Our objective was to characterize at least one DBD that has the properties required to specifically target rRNA genes, whatever the eukaryotic genome in which it could be used. The interests of the strategy that consist in targeting insertions of transgenic cassettes into rRNA genes in eukaryotic genomes are summarized in Additional file6.

Here, our data show that R2DBDs have at least five of the properties that are required in such a peptide. Firstly, R2DBDs are able to bind specifically to a 20-bp motif located within the 100-bp region that is conserved among all eukaryotes. Moreover, the affinity of R2DBDs for the 20-bp target is rather high *in vitro*, their apparent Kds being around 20 nM. Secondly, R2DBDs bind to their DNA target in monomer form. This property is very important if R2DBDs are used to target other molecules, because it should prevent the assembly of homo-aggregates. Thirdly, the R2DBD sequences contain strong NLS information, which should avoid the addition of an NLS in the synthetic peptide. Fourthly, R2DBDs retain their DNA binding properties when there are fused to another protein or domain. Fifthly, the functionality of a protein fused to R2DBD is preserved, as illustrated here with LexA, pZFD^{bcr-abl}, N57SB and GFP. We found that the functionality of SB10 is in fact altered by its N-terminal fusion to R2DBD. However, this is very probably due to the fact that SB10 cannot tolerate N-terminal fusion, which is clearly illustrated in the literature with domains of various origins [14-16].

Our results also show that a few properties of the R2DBDs still need to be optimized before they become fully usable. Indeed, we observed that R2DBD-GFP fusions did not penetrate into all nuclear compartments, but were concentrated around nuclear organelles, nucleoli and other small nucleolar accessory bodies, that are connected to each other by proteins such as fibrillarin [45,46]. This property, that is probably due to the absence of a nucleolar localization signal (NoLS), means that in their present form R2DBDs are unable to reach the rRNA genes in the nucleoli *in vivo*. There are ways to circumvent both problems independently. Firstly, the localization of R2DBD in nucleoli can be determined by adding an NoLS at its N- or C-terminal ends, as described in [47]. Secondly, the potential affinity for the

surface of nuclear organelles could be eliminated by screening a library of punctually mutated R2DBDs, using a single or a two yeast-hybrid system to locate mutants that have lost the ability to interfere, for example, with fibrillarin. This second point still needs to be engineered after adding an NoLS at one end of R2DBD. Even if it is relatively simple to think up ways of doing this, the implementation of such mutant screening will require considerable investment, since the properties of several mutants that do not assemble around nuclear organelles, will all need to be verified in order to identify one or more optimal DBD candidates.

Unexploited properties of R2DBD

Previous studies [28,30] have together demonstrated that R2DBD is able to bind specifically to a 20-bp rRNA target. The findings reported here indicate that R2DBD has no dual DNA binding activities, i.e. one DNA binding activity with high affinity for its 20-bp target, and a second activity characterized by the absence of specificity and low binding affinity. Indeed, we observed that R2DBD expressed alone *in vivo* does not inhibit the DNA plasmid integrations by transposition and recombination (Fig. 5g), and does not hamper the luciferase expression from a plasmid. The absence of dual DNA binding activities is not common in either natural or synthetic DBDs. Hence, the three DBDs used here to be fused to R2DBD, LexA, pZFD^{bcr-abl} and N57SB, were also found to be able to interact non-specifically *in vivo* with plasmids that do not contain their specific binding target (Fig. 5h and Additional file3). For LexA and N57SB DBDs, our observations confirm observations already reported in the literature, even though there are few such examples [39,48]. In the case of pZFD^{bcr-abl}, this was a priori not expected for a tetra-ZF domain, although dual DNA binding activities had already been identified for pZFDs bigger than tri-ZF. The objective of most molecular methods so far developed to obtain pZFDs is to produce tailored tri-ZFs. Fused to a nuclease or a recombinase catalytic domain, such tri-ZFs are sufficient to bind to a 18-bp binding target by dimerizing two ZF-nuclease or ZF-recombinase monomers [49]. Few molecular methods have been developed to manufacture hexa- and nona-ZF monomers to bind to 18- and 30-bp DNA motifs, respectively [50-53]. If tailored hexa- and nona-ZF monomers have nM affinity for their binding target, their organization into several blocks of di- or tri-ZF means that they are also able to bind with a lower affinity to smaller motifs similar to submotifs contained in their specific 18- and 30-bp DNA target. In a eukaryotic genome, the ability of a hexa-ZF monomer to recognize a unique, 18-bp chromosomal motif is therefore going to be a balance between its specificity for its target and its ability to bind with lower affinity to hexa- and nona-bp motifs that are much more abundant in eukaryotic genomes. These features of the

pZFDs greater than tri-ZF explain why we found here that pZFD^{bcr-abl} is able to interact non-specifically *in vivo* with plasmids that do not contain its binding site (Additional file3).

Overall, these observations suggest that R2DBDs may have two very interesting properties for certain biotechnological purposes requiring highly specific DNA binding of a 20-bp target. First, R2DBDs seem to be unable to bind non-specifically to DNA. Second, R2DBDs seem unable to bind to DNA sub-motifs contained in the sequence of its 20-bp rDNA target. The 3D structure of R2DBDs is probably responsible for their binding as a single block to the 20-bp rDNA target.

To target one specific chromosomal locus requires that only one 18-bp DNA motif in eukaryote genomes is recognized. This means that pZFD monomers designed to target DNA motifs of at least 12-bp cannot ensure the uniqueness of their binding to their target as reliably as R2DBDs do to their 20-bp rDNA target. The core structure of R2DBDs consists of one ZF associated with an *Myb*-like sub-domain [28,29](Fig. 1d). Starting from the R2DBD backbone, it might therefore be of interest to develop the molecular engineering of a new type of monomeric DBDs that would be tailored to the 20-bp target sequence to which they have to bind. The design of molecular methods to select such tailored R2DBD derivatives might be facilitated by the fact that the ZF structure in R2DBD (modeled using I-TASSER [54]) is very similar to that used to design tailored ZF. Using data from the ZiFit database, it should therefore be possible to fit this ZF moiety to the nucleotide triplet to which it binds [21], and so “only” have to design a selection procedure on the *Myb*-like sub-domain. *Myb*-like sub-domains are composed of 3 main α -helices in which only the third is involved in DNA binding [29]. Obtaining a crystal structure of an R2DBD bound to its 20-bp DNA target should help to confirm the viability of this approach.

Alternatives to fusion peptides to target vectors

Data recovered here about the functioning of the LexA, pZFD^{bcr-abl} and N57SB confirm the need to have rigorous control procedures in place before attempting to use a protein domain in a synthetic peptide system. Nevertheless, our experience with such peptides suggests that these controls will very probably be far from sufficient to control the targeting. Indeed, whatever integration process is used (i.e. HR or transposition) and whatever locus is targeted (rRNA genes or other loci), the localization of the plasmid in the vicinity of the locus into which it must be targeted is going to depend to a large extent on the balance between two parameters: the peptide expression rate in the cell and the amount of plasmid able to reach the nuclei. A third parameter is that this balance is going to be complicated by the fact that the

plasmid transfection efficiency is going to vary over about 4 orders of magnitude in a single cell population (this is strongly suggested by data obtained from the FACS analysis of a cell population transfected with a plasmid expressing the GFP). To date, although the third parameter seems to be difficult to control, the transfection of a pre-assembled complex (plasmid-targeting peptide) is a technical solution that should be kept in mind, since it could make it possible to circumvent the problems related to the two first parameters. One way to assemble such a complex could be to use a *bis*-PNA oligonucleotide that is able, on the one hand, to be specifically annealed to the plasmid [10], and on the other hand, to be bound or cross-linked to a DBD. In such a procedure, the *bis*-PNA oligonucleotide might first be annealed to the plasmid by Temperature-assisted cyclic hybridization (TACH [10]). Depending on the presence of a biotin or SH group grafted at one end of the *bis*-PNA oligonucleotide, a streptavidin-DBD peptide or a polyNQ-DBD peptide might be then bound or chemically cross-linked to the *bis*-PNA/plasmid complex.

The efficiency of the TACH method to anneal *bis*-PNA oligonucleotide onto a plasmid, and of the binding of a monovalent streptavidin to a biotin [55], or of the cross-linking with the Sulfo-EMCS cross-linker between a SH group and a polyNQ tail mean that this might be a viable solution. As previously pointed out, even if it is relatively simple to think up such solutions, their implementation will require considerable investment.

Conclusion.

R2DBDs are promising candidates to target DNA molecules in the immediate vicinity of the rRNA genes since they i) specifically recognize, with high affinity, a 20-bp binding site located within the expected genomic rDNA target, ii) act as monomers, iii) contain a nuclear localization signal, iv) remain functional when fused to another domain and, v) do not alter the functionality of the protein to which they are fused. However, R2DBDs have two properties remain to be engineered before these DBDs can be integrated into a molecular targeting system directed into rRNA genes. The first concerns the ability of R2DBD to locate within the nucleolus, the organelle in which the rRNA genes reside. The second is the tendency of R2DBD to accumulate in certain parts of the nuclei, which limits its diffusion within nuclei. Technological solutions are available to engineer these DBDs. The results reported here also supply important information concerning the R2DBD properties and the plasmid DNA targeting within nuclei that will need to be further analyzed from three aspects; the unexploited advantages of the R2DBDs, the possibilities and limitations of fusion peptides

for targeting integrations of non-viral vector, and the alternatives to fusion peptides for targeting vectors.

Materials and methods

pZFD design

An optimal motif was first calculated using the Zinc Finger Tools website (<http://www.scripps.edu/mb/barbas/zfdesign/zfdesignhome.php>; [20]) to target a synthetic pZFD within the 100-bp region shown in Fig. 1. This 22-bp motif is located 11-bp upstream of the R2 endonuclease binding site. DNA fragments encoding pZFD able to bind to this 22-bp motif were designed using facilities available at websites proposing three calculation methods. The first pZFD, designated ZFDBarbas, was designed using the Zinc Finger Tools website. Two other pZFDs, designated ZFDSangamo and ZFDToolgen respectively, were designed at the ZiFiT V3.0 website (<http://bindr.gdcb.iastate.edu/ZiFiT/>; Sander et al. 2007), following the recommendations for optimization from the ZiFDB V1.0 website (<http://bindr.gdcb.iastate.edu:8080/ZiFDB/controller/searchObjects>; [22]).

DNA fragment synthesis

DNA fragments containing the 100-bp segment including the binding sites of the pZFD and the R2DBD proteins (Fig. 1b) were synthesized by ATGBiosynthetics (Germany) and cloned in pUC57 (Fermentas). The fragments encoding pZFD and the R2DBD proteins (Fig. 1d) were designed and synthesized with a codon usage avoiding codons that are rare in human cells. The R2DBD originated from sibling endonucleases encoded by the non-LTR R2 retrotransposon of *Bombyx mori* and *Drosophila melanogaster*. They were designated BmR2DBD and DmR2DBD respectively. The LexA binding site was constructed using two oligonucleotides (LexAfix: 5'- AATTCCATGGATCCGTCGACCTGTATATATATACACT CGAGATCTCCATGG -3' and LexAfix-Rev: 5'- AATTCCATGGAGATCTCGAGTGTATATATATACAGGTCGACGGATCCATGG -3') that were annealed, and then cloned into the *EcoRI* pBS-SK- plasmid. The DNA segment encoding the LexA-NLS was designed on the basis of data in the literature [14].

DNA probes

A 150-bp fragment containing the 100-bp segment with the protein binding sites was purified after plasmid digestion with *KpnI* and *SalI*, separation on an agarose gel and elution with the Wizard SV gel and PCR clean-up system kit (Promega). This probe was designated the R2site. A variant fragment of 240-bp was obtained by PCR using M13 universal and reverse primers. A 51-bp fragment containing the LexA binding site (LexA-BS) was similarly purified using *EcoRI* plasmid digests. A 70-bp fragment containing the 3'-ITR of *Mos1*

transposon was prepared as described [33]. DNA fragments were radiolabeled by filling using the DNA polymerase Klenow fragment (New England Biolabs), 30 μ Ci of [α - 32 P] dATP (3000 Ci/mmol) and unlabelled dC-G-TNPs. The probes were purified by electrophoresis on a 6% native polyacrylamide gel and then eluted in Tris-HCl 10 mM [pH8], 1 mM EDTA, and 50 mM NaCl buffer. After precipitation using tRNA as a carrier, the probes were resuspended in 50 μ l of H₂O, and the DNA concentration was estimated by Hoechst assay [56] using a Mithras LB 940 (Berthold Technologies).

Protein purification of bacterial recombinant protein

Each of the five fragments encoding a DBD was cloned into the pET14b vector (Novagen), in a frame with a Histidine tag (His₆) at the amino terminal end. The pET14b constructs were transformed in BL21 *Escherichia coli* bacteria containing a pRARE plasmid (Novagen), which encoded rare tRNA *E. coli* codons. They were also cloned into the pMalc2 system (New England Biolabs), and fused in frame with the Maltose Binding Protein (MBP) at the amino terminal end. The pMalc2 constructs were transformed in JM109 *E. coli* bacteria using the pRARE plasmid (Novagen).

Bacteria were grown at 37°C and 250 rpm either until the OD reached 0.4 to 0.6 (proteins tagged with His₆) or for 6 hours (proteins tagged with MBP). Expression was then induced by supplying 1 mM IPTG for 3 hours at 37°C and 250 rpm. The cells were harvested by centrifuging, resuspended in buffer A (50 mM Tris-HCl [pH 8], 500 mM NaCl, 1 mM dithiothreitol (DTT)) and ruptured by enzymatic digestion with 1 mg/ml lysozyme, for 30 minutes at 4°C. Insoluble material was removed by centrifuging (15,000 g for 10 min, at 4°C) and the supernatant was used as a crude extract containing the expected fusion protein. The crude extracts were loaded in buffer A onto a HisLink Protein Purification Resin (Promega) for His-tagged proteins and onto amylose resin (New England Biolabs) for MBP-fused protein. After washing the column with buffer A, the proteins were eluted with buffer A supplemented with 250 mM imidazole or 10 mM maltose, respectively. The protein concentration in each of the eluted fractions was determined by a Bradford assay using bovine serum albumin as the standard, and the purification quality was checked by electrophoresis on a sodium dodecyl sulfate - polyacrylamide gel (SDS-PAGE) with Coomassie blue staining. Proteins fused with MBP have a cleavage site for factor Xa located between the tag and the recombinant protein. Proteolyses with factor Xa of the MBP fusions (New England Biolabs) were done according to the instructions of the enzyme supplier.

DNA plasmid transfection in HeLa cells

HeLa cells were cultured in DMEM supplemented with 10% fetal bovine serum (FBS). About 5×10^4 cells were seeded onto each well of a 24-well plate, one day prior to transfection. Cells were transfected with jetPEI™ (Polyplus Transfection) at an N/P rate of 5, according to Manufacturer's instructions. Briefly, plasmid DNA (0.2 to 0.5 µg) and jetPEI™ were each diluted in 50 µl of 150 mM NaCl, then mixed gently together and incubated for 15 min. Cells were then incubated for 2 to 4 h with the complexes dropped into 500 µL of Opti-MEM medium (Invitrogen). The transfection solution was then discarded, replaced by fresh DMEM supplemented with 10% FBS, and incubated for 24 hours at 37°C.

Protein expression in HeLa cells

Each R2DBD fragment was cloned into the pCS2+ vector (Addgene). In each well of a 6-well-plate, 4×10^5 HeLa cells were plated one day prior to transfection, and were cultured in DMEM medium (GIBCO-BRL) supplemented with 10% fetal bovine serum at 37°C and 5% CO₂. Two µg of pCS2-R2DBD were transfected using JetPEI™ according to standard protocols (Polyplus transfection). One day post-transfection, the cells were scraped off and washed with Phosphate Buffer Saline (PBS) 1X solution. Cells were resuspended in buffer A, and disrupted with four 5-minute freeze (-80°C) and thaw (37°C) cycles, and sonicated (twice, for 10 sec at 50W). After centrifuging for 10 min at 15,000g and 4°C, the supernatant was used as a crude extract known to contain the protein of interest.

Integration assays in HeLa

The features of the plasmids used in integration assays, transfection, and microscopy analysis in HeLa cells are summarized in Table 1. About 5×10^5 HeLa cells were co-transfected with 500 ng of a 4 DNA plasmid mix. One fifth (100 ng) of the mix corresponded to the pGL3 plasmid (Promega) that was used to check that transfection had occurred. The remaining 400 ng consisted of a mix composed of one of the 3 different T_pase sources (pCS2, pCS2-SB10 or pCS2-SB10-R2DBD), one of the 5 sources of a cassette encoding a resistance to G418 (pBS-NeoR, pBS-NeoR-LexA/BS, pBS-NeoR-ZBS, pT2-pSV40-NeoR, pT2-pSV40-NeoR-LexA/BS, pT2-pPol1h-NeoR), and one of the 8 target peptide variants (pCS2, pCS2-R2DBD, pCS2-R2DBD-LexA, pCS2-R2DBD-NLS-LexA, pCS2-R2DBD-N57SB10, pCS2-R2DBD-pZFD^{bcr-abl}, pCS2-NLS-LexA, pCS2-NLS-N57SB10). Two days after transfection, 1/3 of cells were used to check the effectiveness of transfection by evaluating the luciferase activity with the Luciferase Assay System Kit (Promega). The remaining 2/3 of the cells were

transferred to 100-mm plates followed by G418 sulfate selection (800 µg/mL, PAA) for 14 days. Finally, the cells were fixed and stained with 70% EtOH - 0.5% methylene blue for 3h. Only colonies with a diameter > 0.5 mm were counted.

Microscopic analysis in HeLa cells

About 5×10^4 HeLa cells were co-transfected with 500 ng of a 2 DNA plasmid mix as described above. 50% of the mix corresponded to the plasmid expressing the R2DBD fused to the GFP (Table 1); the remaining 50% consisted of a plasmid expressing a nucleolus marker fused to the DsRed (pDsRed-C1Hsnucleoline; Becherel et al. 2006). Twenty-four hours post-transfection, cells were washed three times with PBS 1X, and nuclear DNA was stained with 2 µg/mL DAPI for 2h. Three more PBS 1X washes were performed before microscopic observation. Cells were observed under an epifluorescence microscope (Olympus BX51).

Electrophoretic mobility gel assay (EMSA)

Binding reactions were carried out in binding buffer (BB) containing 50 mM Tris-HCl [pH 8], 200 mM NaCl, 5 mM MgCl₂, 1 mM DTT, 11% glycerol, 0.1 mg/ml of bovine serum albumin (BSA), and 0.01% Triton X-100. In each 20 µl of reaction mixture, 0.03 pmol of the ³²P-labeled duplexes were mixed with appropriate amounts of purified proteins, and 1 µg of sheared DNA from herring sperm. Mixtures were incubated at 37°C for 1 hour. Reaction products were then separated using discontinuous 4 to 6 or 9% native polyacrylamide gels (depending on the protein size). Electrophoresis was carried out in 0.25X TBE at 200 V, dried and autoradiographed.

UV cross-linking of DNA-protein complexes

Complexes were assembled in 20 µl reaction mixtures containing 0.2 pmol of ³²P-labeled R2 site probe (150-bp), and 5 pmol of protein His-BmR2DBD or 0.6 pmol of MBP-BmR2DBD in BB. After incubating for 1 hour at 37°C, samples were exposed to UV irradiation (312 nm) for 30 min at 4°C. Reactions were stopped by adding 4 µl of 6X SDS-PAGE loading buffer. The samples were then boiled for 10 min.

Kd determination

Concentrated proteins or HeLa cell extracts were serially diluted and used in EMSA. Bound and unbound R2site probe was quantified by Instant Imager (Packard). The apparent Kd for each binding reaction was estimated from the plot of bound/unbound probe versus binding protein concentration using PRISM 4 software. Data from 4 or 5 replicates were pooled to calculate each apparent Kd.

Analysis of the expressed proteins by immunoblotting

Cells recovered from the cultures were washed three times with 1X PBS. Total protein extracts were separated by electrophoresis, adding 20 µg of each sample to a sodium dodecyl sulfate 8% miniPAGE, and then electro-blotted onto nitrocellulose filters (Bio-Rad Laboratories). After blocking with 5% skim milk in 1X PBS for 1 h, the filters were incubated overnight with rabbit polyclonal anti-6-His (1: 20,000; Bethyl Laboratories Inc.) or anti-LexA (1:1,000; Sigma). The filters were then incubated with horseradish peroxidase-conjugated anti-rabbit IgG (Santa Cruz Biotechnology, Santa Cruz, CA) before being developed using enhanced chemiluminescence (Amersham Pharmacia Biotech, Sunnyvale, CA).

Nucleotide sequence analyses

To make an inventory of the nucleotide motifs in the human genome, we used syntactic models to mine the database using WAPAM at the website http://genoweb.univ-rennes1.fr/Serveur-GPO/outils_acces.php3?id_syndic=185&lang=en. The syntactic models used were TAAGGTAGCCAAATGCCTCG to locate 100% conserved of R2 endonuclease binding sites, and CGGGTAAAC-x(0,12)-GGAGTAACT, for the different variants of the ZBS potentially bound by the pZFD used in this paper, and which has a variable spacer of 0 to 12 nucleotides between the 9-bp motifs bound by ZFD1 and ZFD2 respectively (Fig. 1c). The parameters were set to detect only exact matches (significance score of 100%). We confirmed results obtained with WAPAM by mining databases with BLASTn, using the NCBI facilities and non-default parameters that were the “Somewhat similar sequences (blastn)”, a word size = 7, and an expected threshold of 100.

Additional material

Additional file 1. Design and properties of pZFD proteins

Additional file 2. Amino acid and nucleotide sequence features of R2DBD-SB10

Additional file 3. Impact of R2DBD-NLS-ZFD on the integration by random recombination of pBS-NeoR or pBS-NeoR-ZBS (Ratio 1/1 (pCS2-SB10/pCS2-R2DBD-NLS-ZFD))

Additional file 4. Amino acid sequence of the R2DBD-NLS-N57SB fusion

Additional file 5. Impact of R2DBD peptides on cell viability

Additional file 6. Interest of targeting insertions of exogenous DNA fragments into rRNA genes

List of abbreviations

Bm, *Bombyx mori*

bp, base pair

BB, binding buffer

BS, binding site

DBD, DNA binding domain

Dm, *Drosophila melanogaster*

DMEM, Dulbecco's modified eagle medium

DNA, deoxyribonucleic acid

DTT, dithiothreitol

EDTA, ethylenediaminetetraacetic acid

EMSA, electrophoresis mobility shift assay

FBS, fetal bovine serum

His, histidine

HR, homologous recombination

IPTG, isopropyl β -D-1-thiogalactopyranoside

ITR, inverted terminal repeat

GFP, green fluorescent protein

K_d, “dissociation” constant

MBP, maltose binding protein

NeoR, neomycin resistance gene

OD, optical density

ORF, open reading frame

NLS, nuclear localization signal

³²P, phosphorus ³²

PAGE, polyacrylamide gel electrophoresis

PB, *piggyBac*

PBS, phosphate buffer saline

PCR, polymerase chain reaction

PEI, polyethylenimine

pZFD, polydactyl zinc finger domain

R2, R2 non-LTR retrotransposon

RNA, ribonucleic acid

rRNA, ribosomal ribonucleic acid

SB, *Sleeping Beauty*

SB10, Sleeping Beauty transposase version 10

SDS, sodium dodecyl sulfate

TACH, Temperature-assisted cyclic hybridization

TBE, tris borate EDTA

Tpase, transposase

ZBS, zinc finger domain binding site

ZF, zinc finger

ZFD, zinc finger domain

Authors'contributions

EC, FB and YB conceived the study and wrote the manuscript; EC, MVD performed all the molecular biology and biochemistry works; EC, SC and SC performed all the integration assays in HeLa cells works; CB and GC performed all the experiments in microscopy. All authors read and approved the final manuscript.

Acknowledgements and funding

We would like to thank Prof. Brian McStay for supplying human RNA polymerase I promoters and helpful scientific information and recommendations. The English text has been revised by Dr M. Ghosh. This work was supported by the C.N.R.S. and the François Rabelais University of Tours, and funded by grants from the European Commission (Project SyntheGeneDelivery, N°018716), the Ministère de l'Education Nationale, de la Recherche et de la Technologie, the Association Française contre la Myopathie and the Groupement de Recherche CNRS 2157.

References

- 1 Klabunde J, Kunze G, Gellissen G, Hollenberg CP: **Wide-range integrative expression vectors for fungi, based on ribosomal DNA elements.** In: Production of recombinant proteins. Edited by Gellissen G. Weinheim: Wiley-VCH; 2004: 273-286.
- 2 dos Santos WG, Buck GA: **Simultaneous stable expression of neomycin phosphotransferase and green fluorescence protein genes in *Trypanosoma cruzi*.** *J. Parasitol* 2000, **86**:1281-1288.
- 3 Liu X, Liu M, Xue Z, Pan Q, Wu L, Long Z, Xia K, Liang D, Xia J: **Non-viral *ex vivo* transduction of human hepatocyte cells to express factor VIII using a human ribosomal DNA-targeting vector.** *J. Thromb Haemost* 2007, **5**:347-351.
- 4 Stewart S, MacDonald N, Perkins E, deJong G, Perez C, Lindebaum M: **Retrofitting of a satellite repeat DNA-based murine artificial chromosome (ACes) to contain loxP recombination sites.** *Gene Ther* 2002, **9**:719-723.
- 5 Wen L, Liang de S, Wu LQ, Cai F, Pan Q, Long ZG, Dai HP, Xia K, Xia JH: **Efficient protein expression from the endogenous RNA polymerase I promoter using a human ribosomal DNA targeting vector.** *Biochem Biophys Res Commun* 2008 **367**:846-851.
- 6 Jakubczak JL, Burke WD, Eickbush TH: **Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects.** *Proc Natl Acad Sci USA* 1991, **88**:3295-3299.
- 7 Gentile KL, Burke WD, Eickbush TH: **Multiple lineages of R1 retrotransposable elements can coexist in the rDNA loci of *Drosophila*.** *Mol Biol Evol* 2001, **18**:235-245.
- 8 Robinett CC, O'Connor A, Dunaway M: **The repeat organizer, a specialized insulator element within the intergenic spacer of the *Xenopus* rRNA genes.** *Mol Cell Biol* 1997, **17**:2866-2875.
- 9 Voigt K, Izsvák Z, Ivics Z: **Targeted gene insertion for molecular medicine.** *J Mol Med* 2008, **86**:1205-1219.
- 10 Oprea II, Simonson OE, Moreno PM, Viola JR, Lundin KE, Smith CI: **Temperature-assisted cyclic hybridization (TACH), an improved method for supercoiled DNA hybridization.** *Mol Biotechnol* 2010, **45**:171-179. doi: 10.1007/s12033-010-9261-x.
- 11 Penton EH, Crease TJ: **Evolution of the transposable element *Pokey* in the ribosomal DNA of species in the subgenus *Daphnia* (Crustacea: Cladocera).** *Mol Biol Evol* 2004, **21**:1727-1739.

- 12 Wu SC, Meir YJ, Coates CJ, Handler AM, Pelczar P, Moisyadi S, Kaminski JM: ***piggyBac* is a flexible and highly active transposon as compared to *Sleeping beauty*, *Tol2*, and *Mos1* in mammalian cells.** *Proc Natl Acad Sci USA* 2006, **103**:15008-15013.
- 13 Maragathavally KJ, Kaminski JM, Coates CJ: **Chimeric *Mos1* and *piggyBac* transposases result in site-directed integration.** *FASEB J* 2006, **20**:1880-1882.
- 14 Ivics Z, Katzer A, Stüwe EE, Fiedler D, Knespel S, Izsvák Z: **Targeted *Sleeping Beauty* transposition in human cells.** *Mol Ther* 2007, **15**:1137-1144.
- 15 Wilson MH, Kaminski JM, George AL Jr: **Functional zinc finger/*sleeping beauty* transposase chimeras exhibit attenuated overproduction inhibition.** *FEBS Lett* 2005, **579**:6205-6209
- 16 Yant SR, Huang Y, Akache B, Kay MA: **Site-directed transposon integration in human cells.** *Nucleic Acids Res* 2007, **35**:e50.
- 17 Nomura W, Sugiura Y: **Design and synthesis of artificial zinc finger proteins.** *Methods Mol Biol* 2007, **352**:83-93.
- 18 Pabo CO, Peisach E, Grant RA: **Design and selection of novel Cys₂His₂ zinc finger proteins.** *Annu Rev Biochem* 2001, **70**:313-340.
- 19 Palazzoli F, Carnus E, Wells DJ, Bigot Y: **Sustained transgene expression using non-viral enzymatic systems for stable chromosomal integration.** *Curr Gene Ther* 2008, **8**:367-390.
- 20 Mandell JG, Barbas CF: **Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases.** *Nucleic Acids Res* 2006, **34**:W516–W523.
- 21 Sander JD, Zaback P, Joung JK, Voytas DF, Dobbs D: **Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool.** *Nucleic Acids Res* 2007, **35**:W599-605.
- 22 Fu F, Sander JD, Maeder M, Thibodeau-Beganny S, Joung JK, Dobbs D, Miller L, Voytas DF: **Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays.** *Nucleic Acids Res* 2009, **37**:D279-283.
- 23 Wright DA, Thibodeau-Beganny S, Sander JD, Winfrey RJ, Hirsh AS, Eichtinger M, Fu F, Porteus MH, Dobbs D, Voytas DF, Joung JK: **Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly.** *Nat Protoc* 2006, **1**:1637-1652.
- 24 Han JS: **Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions.** *Mob DNA* 2010, **1**:15.

- 25 Yang J, Eickbush TH: **RNA-induced changes in the activity of the endonuclease encoded by the R2 retrotransposable element.** *Mol Cell Biol* 1998, **18**:3455-3465.
- 26 Burke WD, Malik HS, Jones JP, Eickbush TH: **The domain structure and retrotransposition mechanism of R2 elements are conserved throughout arthropods.** *Mol Biol Evol* 1999, **16**:502-511.
- 27 Kojima KK, Fujiwara H: **Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets.** *Mol Biol Evol* 2004, **21**:207-217.
- 28 Christensen SM, Bibillo A, Eickbush TH: **Role of the Bombyx mori R2 element N-terminal domain in the target-primed reverse transcription (TPRT) reaction.** *Nucleic Acids Res* 2005, **33**:6461-6468.
- 29 Ogata K, Morikawa S, Nakamura H, Sekikawa A, Inoue T, Kanai H, Sarai A, Ishii S, Nishimura Y: **Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices.** *Cell* 1994, **79**:639-648.
- 30 Xiong Y, Burke WD, Jakubczak JL, Eickbush TH: **Ribosomal DNA insertion elements R1Bm and R2Bm can transpose in a sequence specific manner to locations outside the 28S genes.** *Nucleic Acids Res* 1988, **16**:10561-10573.
- 31 Roiha H, Glover DM: **Duplicated rDNA sequences of variable lengths flanking the short type I insertions in the rDNA of Drosophila melanogaster.** *Nucleic Acids Res* 1981, **9**:5521-5532.
- 32 Fujiwara H, Ogura T, Takada N, Miyajima N, Ishikawa H, Maekawa H: **Introns and their flanking sequences of Bombyx mori rDNA.** *Nucleic Acids Res* 1984, **12**:6861-6869.
- 33 Augé-Gouillou C, Brillet B, Hamelin MH, Bigot Y: **Assembly of the mariner Mos1 synaptic complex.** *Mol Cell Biol* 2005, **25**:2861-2870.
- 34 Schnarr M, Granger-Schnarr M, Hurstel S, Pouyet J: **The carboxy-terminal domain of the LexA repressor oligomerises essentially as the entire protein.** *FEBS Let* 1988, **234**:56-60.
- 35 Izsvák Z, Ivics Z, Plasterk RH: **Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates.** *J Mol Biol* 2000, **302**:93-102.
- 36 Grabundzija I, Irgang M, Mátés L, Belay E, Matrai J, Gogol-Döring A, Kawakami K, Chen W, Ruiz P, Chuah MK, VandenDriessche T, Izsvák Z, Ivics Z: **Comparative analysis of transposable element vector systems in human cells.** *Mol Ther* 2010, **18**:1200-1209.

- 37 Palmer TD, Miller AD, Reeder RH, McStay: **Efficient expression of a protein coding gene under the control of an RNA polymerase I promoter.** *Nucleic Acids Res* 1993, **21**:3451-3457.
- 38 Smale ST, Tjian R: **Transcription of Herpes Simplex Virus tk Sequences Under the Control of Wild-Type and Mutant Human RNA Polymerase I Promoters.** *Mol Cell Biol* 1985, **5**:352-362.
- 39 Wade JT, Reppas NB, Church GM, Struhl K: **Genomic analysis of LexA binding reveals the permissive nature of the *Escherichia coli* genome and identifies unconventional target sites.** *Genes Dev* 2005, **19**:2619-2630.
- 40 McNamara AR, Ford KG: **A novel four zinc-finger protein targeted against p190(BcrAbl) fusion oncogene cDNA: utilisation of zinc-finger recognition codes.** *Nucleic Acids Res* 2000, **28**:4865-4872.
- 41 Yant SR, Park J, Huang Y, Mikkelsen JG, Kay MA: **Mutational analysis of the N-terminal DNA-binding domain of sleeping beauty transposase: critical residues for DNA binding and hyperactivity in mammalian cells.** *Mol Cell Biol* 2004, **24**:9239-9247.
- 42 Watkins S, van Pouderooyen G, Sixma TK: **Structural analysis of the bipartite DNA-binding domain of Tc3 transposase bound to transposon DNA.** *Nucleic Acids Res* 2004, **32**:4306-4312.
- 43 Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, Corbett AH: **Classical nuclear localization signals: definition, function, and interaction with importin alpha.** *J Biol Chem* 2006, **282**:5101-5105.
- 44 Becherel OJ, Gueven N, Birrell GW, Schreiber V, Suraweera A, Jakob B, Taucher-Scholz G, Lavin MF: **Nucleolar localization of aprataxin is dependent on interaction with nucleolin and on active ribosomal DNA transcription.** *Hum Mol Genet* 2006, **15**:2239-2249.
- 45 Nizami Z, Deryusheva S, Gall JG: **The Cajal body and histone locus body.** *Cold Spring Harb Perspect Biol* 2010, **2**:a000653.
- 46 Liu JL, Murphy C, Buszczak M, Clatterbuck S, Goodman R, Gall JG: **The *Drosophila melanogaster* Cajal body.** *J Cell Biol* 2006, **172**:875-884.
- 47 Emmott E, Hiscox JA: **Nucleolar targeting: the hub of the matter.** *EMBO Rep* 2009, **10**:231-238.

- 48 Daniel DC, Thompson M, Woodbury NW: **DNA-binding interactions and conformational fluctuations of Tc3 transposase DNA binding domain examined with single molecule fluorescence spectroscopy.** *Biophys J* 2002, **82**:1654-1666.
- 49 Papworth M, Kolasinska P, Minczuk M: **Designer zinc-finger proteins and their applications.** *Gene* 2006, **366**:27-38.
- 50 Liu Q, Segal DJ, Ghiara JB, Barbas CF 3rd: **Design of polydactyl zinc-finger proteins for unique addressing within complex genomes.** *Proc Natl Acad Sci USA* 1997, **94**:5525-5530.
- 51 Kamiuchi T, Abe E, Imanishi M, Kaji T, Nagaoka M, Sugiura Y: **Artificial nine zinc-finger peptide with 30 base pair binding sites.** *Biochemistry.* 1998, **37**:13827-13834.
- 52 Imanishi M, Yan W, Morisaki T, Sugiura Y: **An artificial six-zinc finger peptide with polyarginine linker: selective binding to the discontinuous DNA sequences.** *Biochem Biophys Res Commun* 2005, **333**:167-173.
- 53 Yan W, Imanishi M, Futaki S, Sugiura Y: **α -helical linker of an artificial 6-zinc finger peptide contributes to selective DNA binding to a discontinuous recognition sequence.** *Biochemistry* 2007, **46**:8517-8524.
- 54 Zhang Y: **I-TASSER server for protein 3D structure prediction.** *BMC Bioinformatics* 2008, **9**:40.
- 55 Howarth M, Chinnapen DJ, Gerrow K, Dorrestein PC, Grandy MR, Kelleher NL, El-Husseini A, Ting AY: **A monovalent streptavidin with a single femtomolar biotin binding site.** *Nat Methods* 2006, **3**:267-273.
- 56 Rengarajan K, Cristol SM, Mehta M, Nickerson JM: **Quantifying DNA concentrations using fluorometry: a comparison of fluorophores.** *Mol Vis* 2002, **8**:416-421.
- 57 Gonzalez IL, Sylvester JE: **Complete sequence of the 43-kb human ribosomal DNA repeat: analysis of the intergenic spacer.** *Genomics* 1995, **27**:320-328.
- 58 Ivics Z, Hackett PB, Plasterk RH, Izsvák Z: **Molecular reconstruction of *Sleeping Beauty*, a *TcI*-like transposon from fish, and its transposition in human cells.** *Cell* 1997, **91**:501-510.
- 59 Bloom SE, Bacon LD: **Linkage of the major histocompatibility (B) complex and the nucleolar organizer in the chicken: Assignment to a microchromosomes.** *J Heredity* 1985, **76**:146-154.
- 60 Ide S, Miyazaki T, Maki H, Kobayashi T: **Abundance of ribosomal RNA gene copies maintains genome integrity.** *Science* 2010, **327**:693-696.

- 61 Ito T, Osawa S, Shibata H, Kanda N: **Identification of rRNA Gene Loci in the Wild Mouse (*Mus musculus molossinus*) Captured at Hachioji, Tokyo.** *J Vet Med Sci* 2007, **69**:1277-1279.
- 62 Su MH, Delany ME: **Ribosomal RNA Gene copy number and nucleolar-size polymorphisms within and among chicken lines selected for enhanced growth.** *Poult Sci* 1998, **77**:1748-1754.
- 63 Szabo P, Lee MR, Elder FB, Prensky W: **Constancy of ribosomal RNA genes during aging of mouse heart cells and during serial passage of WI-38 cells.** *Arch Gerontol Geriatr* 1978, **65**:161-172.

Figure Legends

Fig. 1 Features of rRNA genes, integration and cleavage sites, and the DNA binding domains targeted within the region encoding the 28S rRNA. (a) General map of the human ribosomal DNA repeat unit including the gene region transcribed by RNA polymerase I and the intergenic non-transcribed spacer [57]. In the transcribed region, white boxes correspond to transcribed spacers. Black boxes locate the region encoding the 18S, 5.8S and 28S rRNA. Dotted lines from a to b delineate the 100-bp conserved region. (b) Sequence comparison of the 100-bp region in the gene encoding the 28S rRNA that is 99 to 100% conserved from yeast to human beings. The TTAA tetranucleotide in which the *Pokey* DNA transposon integrates [11] is typed in italics and bold, and the binding site of the endonuclease encoded by the R2 non-LTR retrotransposon [28] is highlighted in gray, with the AG dinucleotide in which the endonuclease cleavage occurs highlighted in black and typed in white. The bipartite DNA binding site of the three pZFDs described in b and c are boxed and in bold, the linker between these sites being shown in lower-case letters. Since this 100-pb DNA segment contained the binding site of our pZFD (ZBS) and that of the R2 endonuclease, it was used as probe in our EMSA analysis, and designated the R2 site (c) Organization of the bipartite pZFD in which both ZFD trimers, ZFD1 and ZFD2, are separated by seven glycine residues. Their binding to the bipartite DNA 9-bp binding site is shown below the pZFD representation. (d) Amino acid sequences of the two DNA binding domains contained in the R2 non-LTR retrotransposons isolated from (1) *Drosophila melanogaster* and (2) *Bombyx mori* (Accession numbers: M16558 and X51967). The zinc fingers are boxed, and the residues involved in the C₂H₂ bonds highlighted in gray. The conserved motifs within the Myb sub-domain are also boxed, and the most highly conserved residues are highlighted in black and typed in white [29].

Fig. 2 EMSA analysis of MBP-BmR2DBD. (a) Shifted complexes assembled with the R2site probe in the absence (lane 1) or presence of 10 nM MBP-BmR2DBD (lane 2). FP: Free R2site Probe. (b) SDS-PAGE (6%) analysis of UV cross-linked complexes resulting from the assembly of the MBP-BmR2DBD and the radiolabeled R2site probe. Lane 1: R2site probe after UV treatment; lane 2: complexes with no UV treatment; lane 3: UV-treated complexes. U: unbound free R2site probe. MW (in kDa) were scaled with the Precision Plus Protein Pre-stained Standards (BioRad). (c) PAGE stained with colloidal Coomassie blue. Lane 1: MBP-BmR2DBD; Lane 2: MBP-BmR2DBD cleaved by Factor Xa. (d) Shifted complexes assembled with the R2site probe in the absence (lane 1) or presence (lane 2) of

5 nM MBP-BmR2DBD, 5, 50, 500 and 5000 nM of MBP-BmR2DBD cleaved with Factor Xa (for 2 hours at 25°C before complex assembly; lanes 3 to 6). Lanes 7 and 8 are negative controls obtained after incubating the R2site probe with MBP and Factor Xa proteins respectively. FP: Free Probe.

Fig. 3 EMSA analysis of His-BmR2DBD. (a) Pattern of the complexes assembled with 50 nM (lane 1), 100 nM (lane 2), 200 nM (lane 3) and 250 nM (lane 4) of His-BmR2DBD and the R2site probe. Lane 5: R2site probe without protein. FP: R2site Free Probe. (b) SDS-PAGE (8%) analysis of UV cross-linked complexes resulting from the assembly of the His-BmR2DBD and the radiolabeled R2site probe. Lane 1: R2site probe after UV treatment; lane 2: complexes with no UV treatment; lane 3: UV-treated complexes. U: unbound free R2site probe. MW in kDa were scaled with the Prestained Protein Molecular Weight Marker (Fermentas). (c) and (d) Saturation graphs at the binding equilibrium with a fixed probe concentration and variable concentrations of His-BmR2DBD and His-DmR2DBD, respectively. (e) and (f) Scatchard plots of the single complexes obtained with His-BmR2DBD and His-DmR2DBD, respectively. The bound/free (B/F) R2site probe was plotted *versus* concentration of bound R2site probe [nM], leading to the Scatchard plot. $-1/\text{Slope}$ of the line gives the apparent K_d . The R^2 correlation coefficients calculated from the data used for the Scatchard linear regression were 0.92 for His-BmR2DBD and 0.86 for His-DmR2DBD.

Fig. 4 Features of His-BmR2DBD expressed in HeLa cells. (a) Shifted complexes assembled with the R2site probe in the absence of protein (lane 1), in the presence of purified protein produced in bacteria (lane 2) or HeLa cell extract expressing His-BmR2DBD (lane 3). FP: Free Probe. (b) Saturation graphs set at the binding equilibrium with a fixed concentration probe and variable amounts of HeLa cell extracts containing His-BmR2DBD. (c) Scatchard plot of the single complex obtained with His-BmR2DBD. The bound/Free (B/F) R2site probe was plotted *versus* concentration of bound R2site probe [nM], leading to the Scatchard plot. $-1/\text{Slope}$ of the line gives the apparent K_d . The R^2 correlation coefficient calculated from data used for the Scatchard linear regression was 0.96. (d) Sequence of the His-BmR2DBD-LexA fusion. The peptide containing the Histidine tag is typed in dark gray and bold. The linker is typed in italics and bold, the flexible poly-Glycine tracts are typed in light gray, and the NLS in black. BmR2DBD and the LexA part are typed in black, the latter being underlined. (e) Western Blot analysis of the His-BmR2DBD-LexA protein expressed in HeLa cells. Lanes 1

and 3: HeLa cell extract transfected with pCS2. Lanes 2 and 4: HeLa cell extract transfected with pCS2-BmR2DBD-LexA vector. Revelation was done with an anti-Histidine tag antibody (lanes 1 and 2) and anti-LexA antibody (lanes 3 and 4). **(f and g)** His-BmR2DBD-LexA complexes with R2site **(f)** and LexA-BS **(g)** probes. Lane 1: probe alone; lane 2: protein extract from cells transfected with pCS2; lane 3: protein extract from cells transfected with pCS2-BmR2DBD-LexA. Black arrows indicate the shifted complexes. Those in gray indicate dissociated complexes that varied in intensity for the various experiments and binding conditions. FP: Free Probe.

Fig. 5 Impact of R2DBD peptides on *in vivo* NeoR cassette integration by random recombination of the plasmid carrier or transposition of the SB vector carrier. Transposition assays monitored in the absence (black bar) or presence of T_pase (SB10, white bar; R2DBD-SB10, dark gray bar) with a transposon source plasmid carrying a pSV40-NeoR **(a)** or a pPol1hNeoR cassette **(b)**. Impact of the R2DBD-NLS-LexA peptide in transposition assays monitored in the absence (black and dark gray bars) or presence of SB10 T_pase (white and light gray bars) with an *SB* transposon plasmid carrying a pSV40-NeoR cassette **(c and d)** and a LexA binding site (LexA/BS; **d**). **(e)** and **(f)**, impact of the R2DBD-NLS-N57SB10 peptide in transposition assays monitored in the absence (black and dark gray bars) or presence of SB10 T_pase (white and light gray bars) with an *SB* transposon plasmid carrying a pSV40-NeoR cassette. 100 ng of pT2-pSV40-NeoR was transfected into 5x10⁴ cells, but a 1/1 and 1/400 plasmid ratio between pCS2-SB10 and pCS2-R2DBD-NLS-N57SB10 were respectively used in **(e)** and **(f)** (100ng/100ng *versus* 1ng/400ng). **(g)**, impact of the R2DBD-NLS peptide in transposition assays monitored in the absence (black and dark gray bars) or presence of SB10 T_pase (white and light gray bars) with a transposon source plasmid carrying a pSV40-NeoR. **(h)**, impact of the NLS-LexA (black, dark gray and white bars) and N57SB10 peptides (light gray bar) on the integration by random recombination of pBS plasmids carrying a pSV40-NeoR cassette containing (white bars) or not a LexA/BS (white bars *versus* black, dark gray and light gray bars). In a, b, c, d, g and h, 100 ng of each plasmid was transfected into 5x10⁴ cells. Data were obtained from at least 6 replicates resulting from at least 2 different experiments. Percentages (%) of G418-resistant clones are calculated for each replicate using the sample represented by a black bar as a reference. They are represented with median and errors bars corresponding to quartiles 1 and 3. * and brackets indicate significant differences using a non-parametric, two-tailed Wilcoxon/Kruskal-Wallis test with a significance threshold of 0.05 (p>0.95).

Fig. 6 Location of R2DBD-NLS-GFP and R2DBD-GFP fusions expressed in HeLa cells.

(a) Analysis of DAPI and GFP-fluorescence distribution in HeLa cells transfected with plasmids expressing GFP, MOS1-GFP or R2DBD-NLS-GFP proteins. GFP and MOS1-GFP⁵² were used as negative and positive controls of active nuclear import, respectively. The nuclear DNA was stained with DAPI. Merge pictures were used to confirm the co-localization between the nucleus and GFP proteins. Since the results were similar, only those obtained with R2DBD-NLS-GFP are shown. **(b).** Analysis of GFP and DsRed-fluorescence distribution in HeLa cells co-transfected with plasmids each of which expressed the R2DBD-NLS-GFP and pDsRed-C1Hsnucleoline fusions. pDsRed-C1Hsnucleoline was used to label nucleoli. Merge pictures were used to verify the co-localization between nucleoli and GFP signal in nuclei. Since results were similar with both fusions, only those obtained with R2DBD-NLS-GFP are shown in a and b.

Table 1: Plasmids used for transfections into HeLa cells

Plasmid name	Plasmid features
<i>Transposase source for integration assays</i>	
pCS2-SB10	Plasmid encoding the version 10 of the <i>sleeping beauty</i> transposase (SB10 [58])
pCS2-SB10-R2DBD	Plasmid encoding SB10 fused with R2DBD at its terminal end
<i>Sources of NeoR cassette for integration assays</i>	
pBS-NeoR	pBSKS- in which a gene encoding a resistance to G418 flanked at its 5' and 3' ends respectively by a SV40 promoter (pSV40) and a SV40 polyadenylation signal is cloned between the <i>EcoRI</i> and <i>BamHI</i> sites of the multicloning site (MCS)
pBS-NeoR-LexA/BS	pBS-NeoR in which a binding site to LexA is cloned at the <i>EcoRI</i> site in MCS
pBS-NeoR-ZBS	pBS-NeoR in which is cloned a binding site (12-mer) to the DNA binding domain ZFD ^{bcr-abl} [40] at the <i>EcoRI</i> site in the MCS
pT2-pSV40-NeoR	Plasmid coding an SB vector in which a NeoR cassette similar to that of the pBS-NeoR is cloned [58]
pT2-pSV40-NeoR-LexA/BS	pT2-pSV40-NeoR in which a binding site to LexA at the <i>EcoRI</i> site located upstream to the NeoR cassette is cloned
pT2-pPol1h-NeoR	pT2-pSV40-NeoR in which pSV40 was replaced by an RNA polymerase I promoter extracted from the pHENA plasmid [37]
<i>Targeting peptide sources for integration assays</i>	
pCS2-R2DBD-NLS	Plasmid encoding R2DBD fused to an SV40 NLS at its C-terminal end
pCS2-R2DBD-NLS-LexA/BS	Plasmid encoding R2DBD-NLS fused to the LexA DNA binding domain at its C-terminal end
pCS2-R2DBD-LexA	pCS2-R2DBD-NLS-LexA/BS in which the region encoding the NLS was mutagenized to remove it (from a PKKKRK motif to PKAIRK)
pCS2-R2DBD-NLS-N57SB10	Plasmid encoding R2DBD-NLS fused to the 57 first residues of SB10 (N57SB10) at its C-terminal end [28,41]
pCS2-R2DBD-NLS-N57SB10	pCS2-R2DBD-NLS-N57SB10 in which the region encoding the NLS was mutagenized to remove it (from a PKKKRK motif to PKAIRK)
pCS2-R2DBD-NLS- ZFD ^{bcr-abl}	Plasmid encoding R2DBD-NLS fused to the DNA binding domain ZFD ^{bcr-abl}
pCS2-NLS-LexA/BS	Plasmid encoding SV40 NLS fused to the LexA DNA binding domain at its C-terminal end
pCS2-NLS-N57SB10	Plasmid encoding SV40 NLS fused to N57SB10 domain at its C-terminal end
<i>GFP or DsRed fusion proteins for epifluorescent microscopy analyses</i>	
pCS2-R2DBD-NLS-GFP	Plasmid encoding R2DBD-NLS fused to the GFP protein at its C-terminal end
pCS2-R2DBD-GFP	pCS2-R2DBD-NLS-GFP in which the region encoding the NLS was mutagenized to remove it (from a PKKKRK motif to PKAIRK)
pCS2-MOS1-GFP	Plasmid encoding <i>Mos1</i> transposase fused to the GFP protein at its C-terminal end
pDsRed-C1Hsnucleoline	Plasmid encoding the Red fluorescent protein (DsRed) fused with the human nucleoline at its C-terminal end [44]

Table 2a - Number of ZBSs (CGGGTAAACggcgGGAGTAACT) outside the chromosomal loci that contain the rRNA genes.

Model and features of the rDNA loci [‡]	N° of ZBS	N° exact hits	Locations in chromosomes
<i>Homo sapiens</i> (GRCh37.p2 primary reference assembly) rDNA loci on chromosomes 13, 14, 15, 21 and 22. 350 rDNA copies/haploid genome	6	5	UGC-117088 UGC-161060 chr1-237766392 chr7- 68527469 chr11-77597559 chrX-108297433*
<i>Mus musculus</i> (strain C57BL/6) rDNA loci on chromosomes 12, 15, 16, 18 and 19. 100 rDNA copies/haploid genome	10	7	chr1-46132978 chr1-19294485 chr3-5860423 chr4-45359522 chr4-71689154 chr7-127897713* chr17-7146256 chr17-8145263 chr17-24802931* chr18-54858951*
<i>Rattus norvegicus</i> (strain BN/SsNHsdMCW) rDNA loci on chromosomes 3, 11, 12. 100-360 rDNA copies/haploid genome	3	3	chr5-24497292 chr12-34422429 chr16-68375450
<i>Gallus gallus</i> rDNA loci on chromosome 16. 261-331 rDNA copies/haploid genome	0	0	-
<i>Drosophila melanogaster</i> rDNA loci on chromosome X and Y. 240 rDNA copies/haploid genome	0	0	-
<i>Saccharomyces cerevisiae</i> rDNA loci on chromosome XII. 150 rDNA copies/haploid genome	0	0	-

UGC, unplaced genomic contig; *, one mismatch into the ggcg linker; [‡], [59-63].

Table 2b - Number of R2DBD binding sites (TAAGGTAGCCAAATGCCTCG**) outside the chromosomal loci that contain the rRNA genes

Model***	N° of BS	Locations in chromosomes
<i>Homo sapiens</i>	3	chr2-133037437 chr8-70602505 chrX-108297465
<i>Mus musculus</i>	9	chr1-46132946 chr1-192944886 chr2-5300263 chr3-5860455 chr3-120952860 chr4-71689122 chr14-57312991 chr17-7146288 chr18-54858919
<i>Rattus norvegicus</i>	3	chr5-24497324 chr12-34422461 chr16-68375482
<i>Gallus gallus</i>	0	-
<i>Drosophila melanogaster</i>	0	-
<i>Saccharomyces cerevisiae</i>	0	-

, 100% similar over 100% of the sequence motif; *, model features are similar to those in Table 1a

Fig. 1

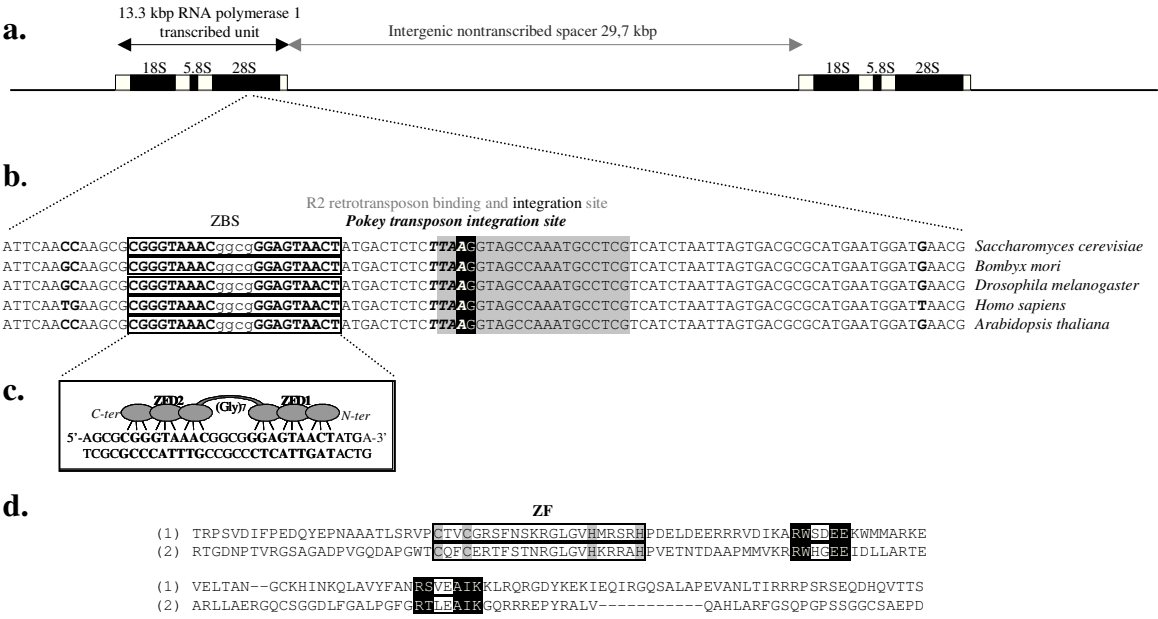


Figure 1

Fig. 2

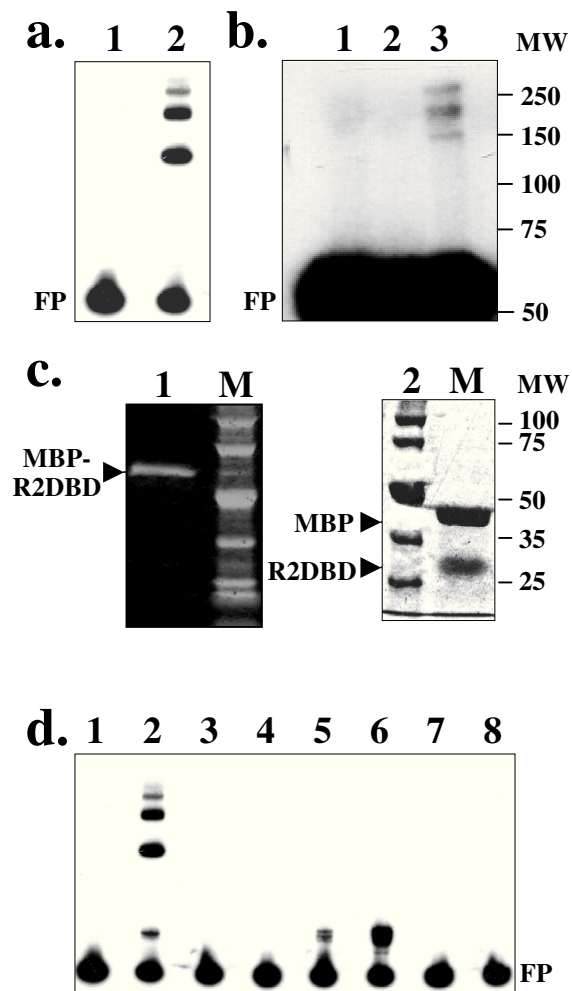


Fig. 3

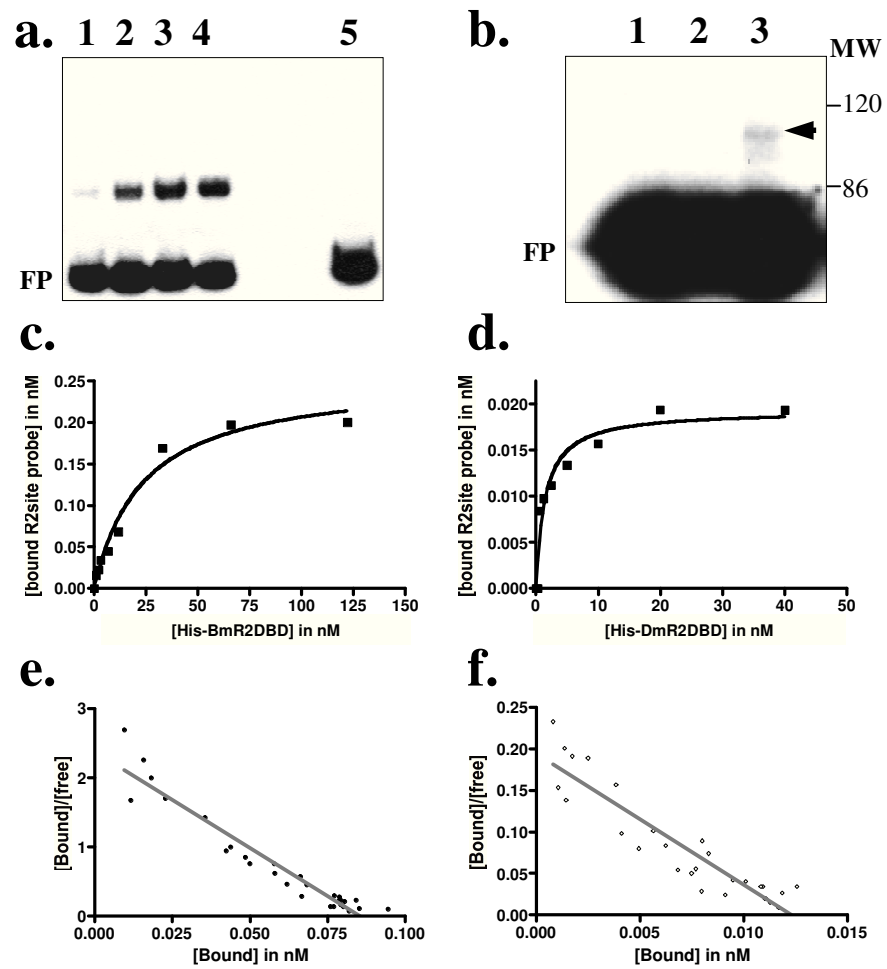


Fig. 4

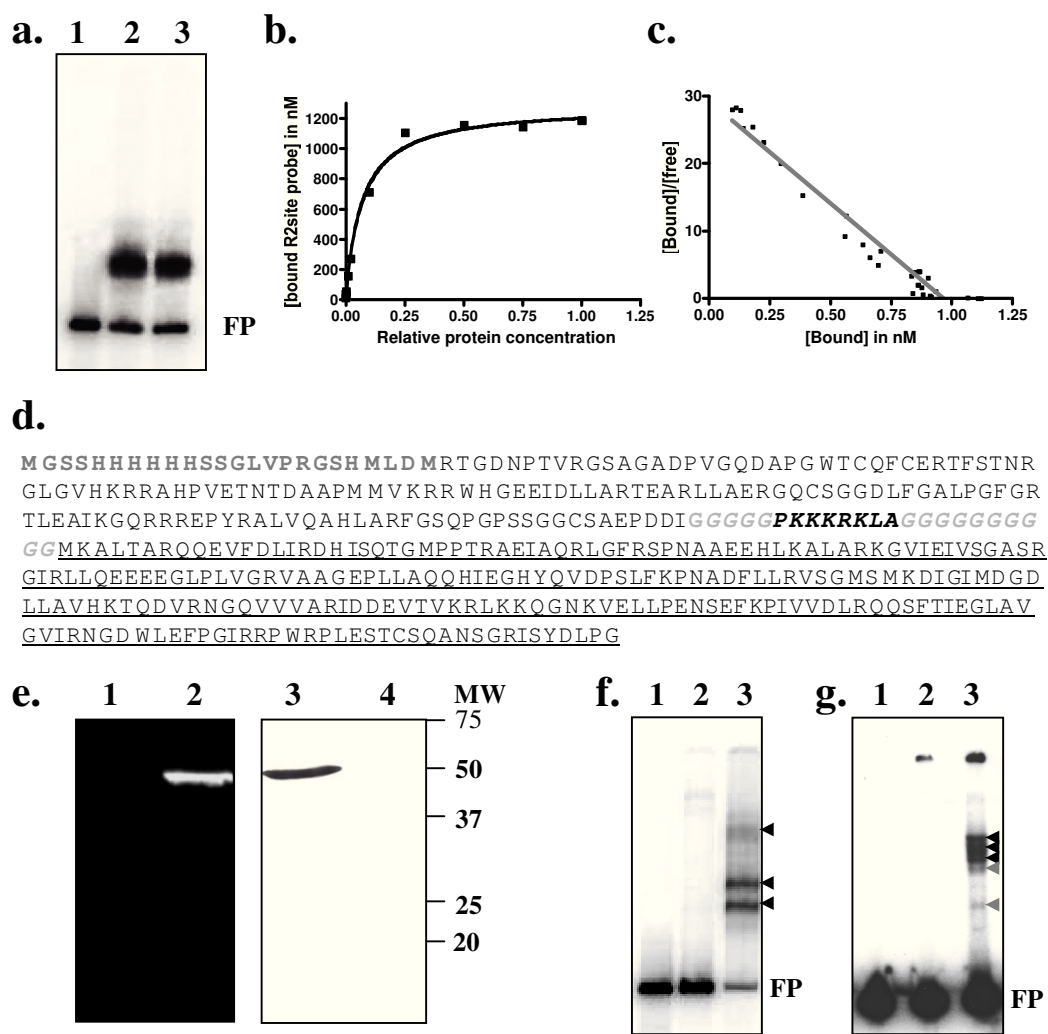


Figure 4

Fig. 5

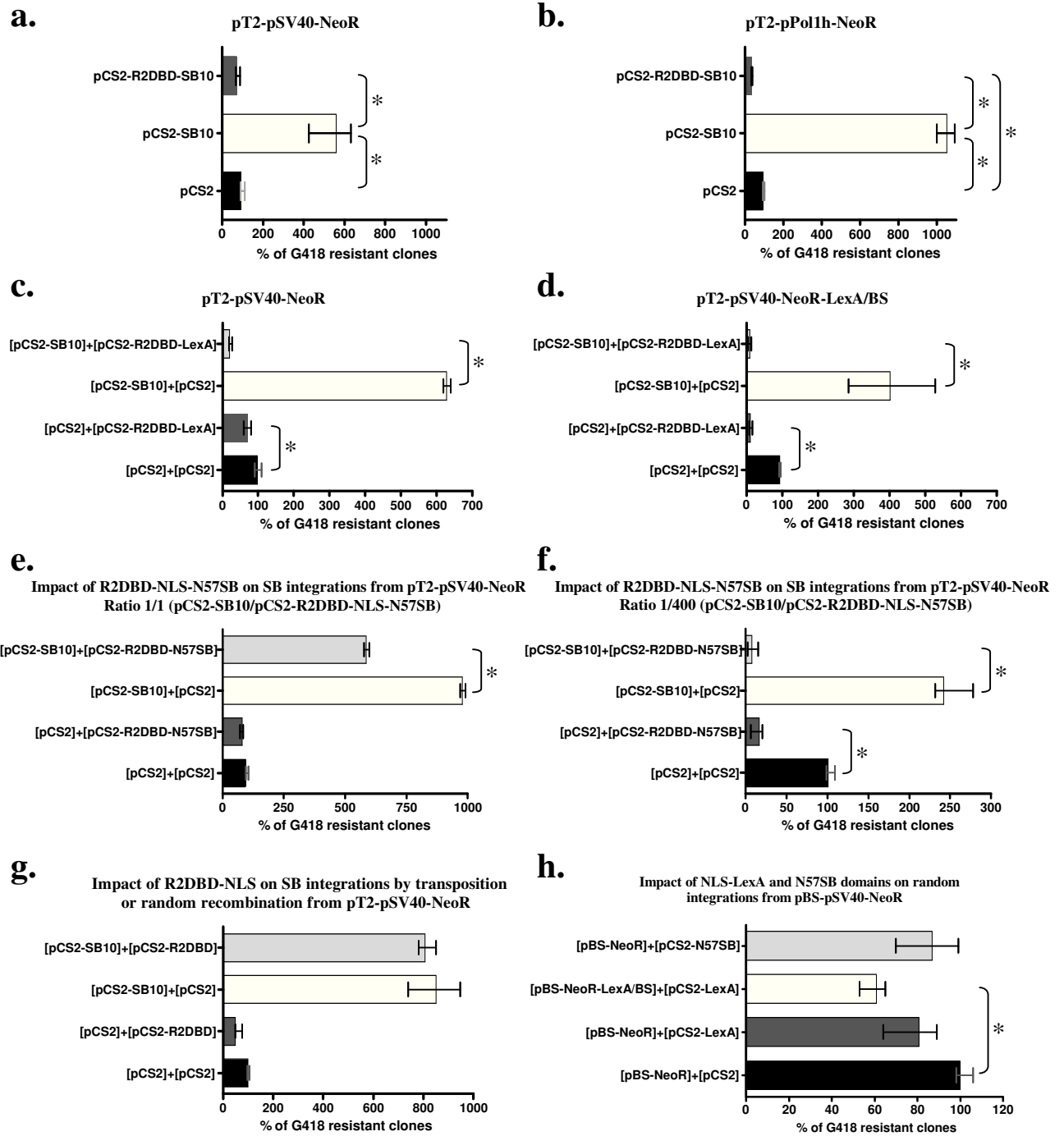
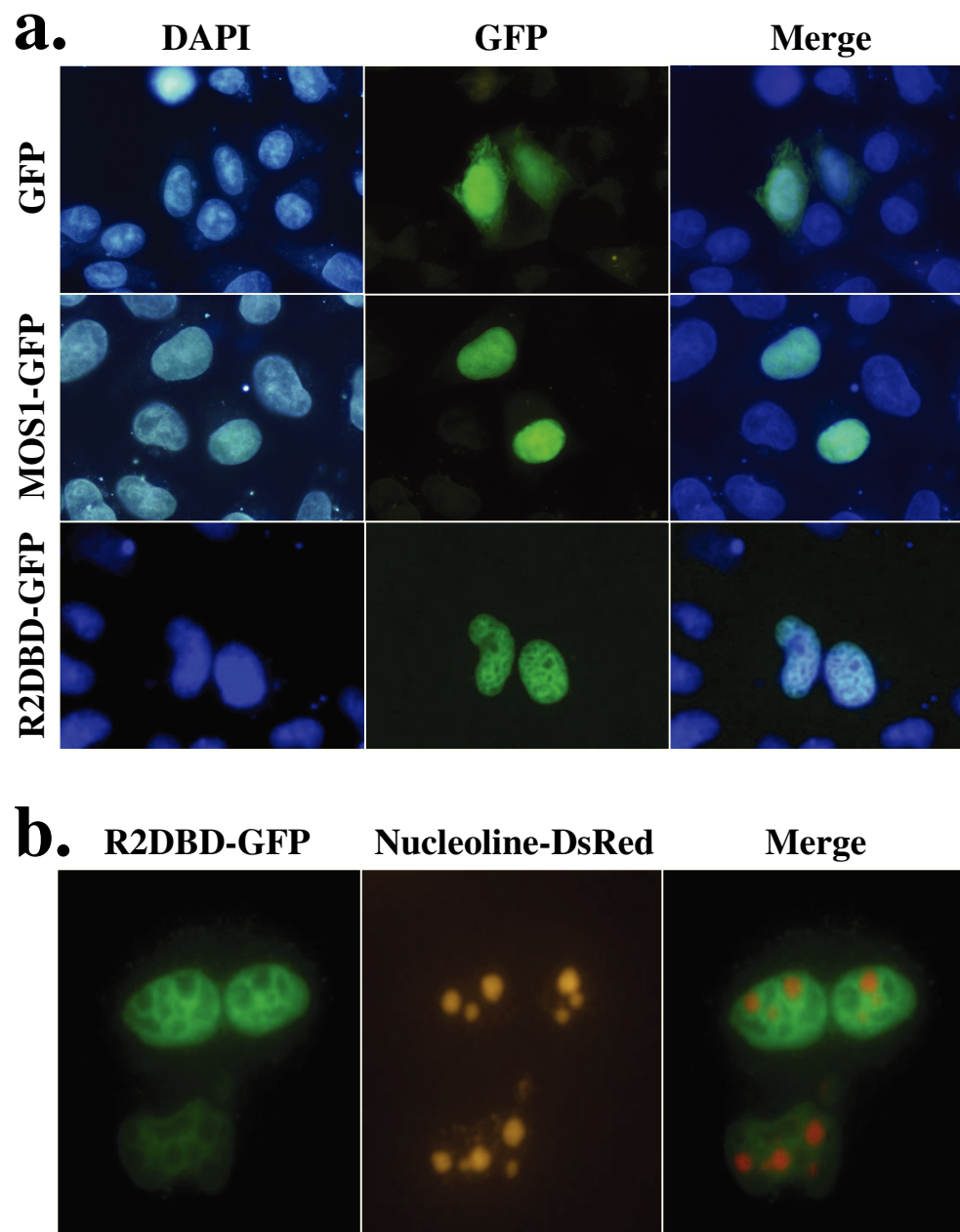


Fig. 6



Characterization of monomeric protein domains that bind specifically to a highly-conserved 100-bp target within rRNA genes

Elodie Carnus, Marie-Véronique Demattei, Sophie Casteret, Guillaume Carpentier, Fabien Palazzoli, Solenne Bire, Christophe Bressac, and Yves Bigot

Additional material

Additional file. Design and properties of pZFD proteins

In silico design of pZFD proteins.

Three different pZFDs were designed *in silico* following the method described by Mandell and Barbas¹, and using the backbone sequence YKCPECGKSFS-X₇-HQRT¹, in which X₇ corresponds to the variable amino acid motifs that allocate their trinucleotide binding specificity to each ZF. In each of the ZF trimers (**Figure 1c**, ZFD1 and ZFD2), the ZFs were linked to the consensus linker, TGEKP. A flexible linker consisting of seven glycines (Gly)₇ was used to link the ZF trimers². Finally, the specific peptides LEPGEKP and TGKKTS were added at the N- and C-terminal ends, respectively, of the three pZFD proteins (**Figure S1a**). In the first pZFD, known as ZFDBarbas, the ZF with the greatest affinity for each of the 6 trinucleotides (CGG, GTA, AAC, GGA, GTA and ACT) was selected to make the fusion protein (**Figure S1a**, lane 1). Two other pZFDs were optimized using ZiFiT facilities²⁸. The second protein, called ZFDSangamo, was designed using the Sangamo method that optimized each ZF trimer, taking into account the position occupied by each ZF in the fusion (**Figure S1b**, lane 2). The Toolgen method, that optimized pZFD from data recovered from natural pZFD encoded by the human genome, was used for the third pZFD, and designated ZFDToolgen (**Figure S1a**, lane 3). When the trimer of ZF could not be optimized, the ZF corresponding to ZFDBarbas was used. In view of its position in each ZF trimer, calculations done using ZiFDB facilities³ could not be used to further optimize the choice of each ZF in our pZFD. It nevertheless revealed that all the ZFs selected in ZFDSangamo corresponded to the theoretically optimal solutions.

a.

```
(1) MLEPGEKPYKPECGKSFSSTHLDLIRHQRTHTGEKPYKPECGKSFSQSSSLVRHQRTHTGEKPYKPECGKSFSQRAHLRHQRTGGG
(2) MLEPGEKPYKPECGKSFSSTHLDLIRHQRTHTGEKPYKPECGKSFSQSGALARRHQRTHTGEKPYKPECGKSFSQSGHLQRHQRTGGG
(3) MLEPGEKPYKPECGKSFSSTHLDLIRHQRTHTGEKPYKPECGKSFSQSSSTLTRHQRTHTGEKPYKPECGKSFSQSTHLRHQRTGGG

(1) GGGGYKPECGKSFSDSGNLRVHQRTHTGEKPYKPECGKSFSQSSSLVRHQRTHTGEKPYKPECGKSFSRSDKLTETHTTGKKTS
(2) GGGGYKPECGKSFSDSGNLRVHQRTHTGEKPYKPECGKSFSQSGALARRHQRTHTGEKPYKPECGKSFSRSDNLTETHTTGKKTS
(3) GGGGYKPECGKSFSDSGNLRVHQRTHTGEKPYKPECGKSFSQSSSTLTRHQRTHTGEKPYKPECGKSFSRSDNLTETHTTGKKTS
```

Figure S1a. Sequence features of pZFD that were designed to bind to the conserved 100-bp region located within the rRNA genes encoding the 28S rRNA. Amino acid sequences of the three bipartite pZFD used in our work: (1) ZFDBarbas, (2) ZFDSangamo and (3) ZFDToolgen. Variable regions in each ZF are highlighted in gray. C₂H₂ motifs are highlighted in yellow or green in each ZF. The linkers between ZFD1 and ZFD2 are highlighted in black and typed in white.

DNA fragment synthesis

DNA fragments containing the 100-bp segment including the binding sites of pZFD and R2DBD proteins (**Figure 1a and S1a**) were synthesized by ATGBiosynthetics (Germany) and cloned in pUC57 (Fermentas) on the basis of a codon usage avoiding codons that are rare in human cells.

Binding specificity of pZFD protein to ZBS

The three pZFD proteins were cloned in the expression plasmids: in pMalc2 to produce protein fused with the Maltose binding protein (MBP), and in pET14b to generate fusions with a His₆ tag at their N-terminal ends. These fusion proteins were produced in bacteria and then purified. Their abilities to bind to the R2site probe containing ZBS (**Figure 1b**) were evaluated by EMSA (electrophoretic mobility shift assay). Results revealed the formation of similar shifted complexes with MBP-ZFDBarbas, MBP-ZFDSangamo and MBP-ZFDToolgen (**Figure S2, lanes 1 to 3**), the most efficient binding being obtained with the first of these proteins. In contrast, no shifted complex was obtained with the three His₆-tagged proteins, indicating that the tag prevented either the activity or the folding of these fusions, as previously reported for other basic proteins such as transposases⁷. Although the specificity of the shifted complex obtained with the three MBP-tagged fusions had already been confirmed

by the presence of a saturating amount of non-specific DNA competitor in each sample assayed, it was nevertheless confirmed that no shifted complex was obtained with a non-specific probe, such as the 3' inverted terminal repeat of the *MosI* transposon (data not shown).

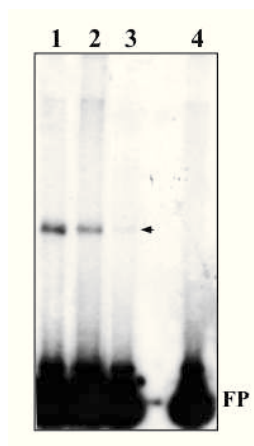


Figure S1b. EMSA analysis of the three pZFDs fused to the MBP. Lane 1, MBP-ZFDBarbas; lane 2, MBP-ZFDSangamo; lane 3, MBP-ZFDToolgen and lane 4, R2site probe alone. FP, indicates the free R2site probe in gel, and the arrow, the MBP-pZFD/R2site complexes. Similar results were obtained whatever the binding conditions (temperature of 20, 30 or 37°C, and 0 or 5 mM ZnCl₂).

Because of its greater binding activity, the MBP-ZFDBarbas protein was identified as being the best candidate, and its apparent K_d was determined. EMSA experiments were first carried out using a fixed concentration of the R2site probe, and serially diluted protein. A saturation graph was obtained that made it possible to specify a fixed concentration of 39 nM of the MBP-ZFDBarbas protein (data not shown). Under these conditions, the apparent K_d was estimated to be over 1 μM, a value characteristic of a fusion protein with low affinity for its binding site.

Overall, our data indicate that it was not possible to design an efficient pZFD *in silico* that would be able to bind to the conserved 100-bp segment of the gene encoding 28S rRNA. An alternative approach might therefore be to use a selection procedure to obtain an efficient pZFD. However, this raises questions about the pertinence of using a molecular selection procedure with, for example, the open source of standardized reagents and protocols for engineering ZF by modular assembly⁹. Indeed, since an open source of standardized reagents and protocols for engineering ZF by modular assembly⁸ is now available, one solution might have been to pursue our investigations by selecting a more efficient pZFD *in vitro*. It must be pointed out that this approach would be still more time-consuming and expensive than the *in silico* design approach. Beside the fact we had the R2DBD alternative solution available in the lab, one of the reasons that we did not further develop pZFD are the conditions to which they are subject when they have been developed or exploited for commercial purposes. Any pZFD engineered by public or private research laboratories is in fact controlled by private bodies, due to the intellectual property (IP) issues that surround and constrain this technology¹⁰⁻¹³. Since our aim is to develop DBD tools for gene targeting that can subsequently be made freely available to the scientific and medical communities, we focused our investigation on the BmR2DBD and DmR2DBD. From this perspective, it must be pointed out that the strategy consisting of integrating vectors specifically into rRNA genes, and using molecular tools derived from a non-LTR retrotransposon is IP-free, because a 1994 patent belonging to Transgene S.A. has been abandoned (WO/1994/024300 - Transposition assembly for gene transfer in eukaryotes. Applicant: Transgene S.A.; inventor Jacobs, E), thus liberating this strategy for free exploitation.

References cited in S1

1. Mandell, JG and Barbas, CF (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* **34** (Web Server issue): W516–W523.
2. Imanishi, M, Hori, Y, Nagaoka, M and Sugiura, Y (2000) DNA-bending finger: artificial design of 6-zinc finger peptides with polyglycine linker and induction of DNA bending. *Biochemistry* **39**: 4383-4390.
3. Fu, F, Sander, JD, Maeder, M, Thibodeau-Beganny, S, Joung, JK, Dobbs, D, Miller, L and Voytas, DF (2009) Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. *Nucleic Acids Res* **37** (Database issue): D279-283.
4. Penton, EH and Crease, TJ (2004) Evolution of the transposable element *Pokey* in the ribosomal DNA of species in the subgenus *Daphnia* (Crustacea: Cladocera). *Mol Biol Evol* **21**: 1727-1739.
5. Christensen, SM, Bibillo, A and Eickbush, TH (2005) Role of the Bombyx mori R2 element N-terminal domain in the target-primed reverse transcription (TPRT) reaction. *Nucleic Acids Res* **33**:6461–6468.
6. Ogata K, Morikawa S, Nakamura H, Sekikawa A, Inoue T, Kanai H, Sarai A, Ishii, S and Nishimura, Y (1994) Solution structure of a specific DNA complex of the *Myb* DNA-binding domain with cooperative recognition helices. *Cell* **79**: 639-648.
7. Augé-Gouillou, C, Hamelin, MH, Demattei, MV, Periquet, G and Bigot, Y (2001) The ITR binding domain of the *Mariner Mos-1* transposase. *Mol Genet Genomics* **265**: 58-65.
8. Wright, DA, Thibodeau-Beganny, S, Sander, JD, Winfrey, RJ, Hirsh, AS, Eichinger, M, Fu, F, Porteus, MH, Dobbs, D, Voytas, DF and Joung, JK (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc* **1**: 1637-1652.
9. Kojima, KK and Fujiwara, H (2004) Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. *Mol Biol Evol* **21**: 207-217.
10. Palazzoli, F, Carnus, E, Wells, DJ and Bigot, Y (2008) Sustained transgene expression using non-viral enzymatic systems for stable chromosomal integration. *Curr Gene Ther* **8**: 367-390.
11. Lyznik, LA and Dress, V (2008) Gene targeting for chromosome engineering applications in eukaryotic cells. *Recent Pat Biotechnol* **2**: 94-106.
12. Palazzoli F (2007) Analyses of the consequences of using Zinc Finger Domains for fusion proteins, in a R&D exploitation plan - Impact of the intellectual property on the economic viability for a technological development strategy in *ex vivo* gene therapy, of : i) transposon-based integration mechanisms, ii) their modifications to become site-specific. *Master thesis in Biotechnology and laws*. University François Rabelais of Tours, 60 pages.
13. Palazzoli, F, Carnus, E, Merly, F, Goiran, S and Bigot, Y. (2010) Zinc finger monopoly : quelles sont les règles du jeu ? *Biofutur* **306**: 58-62

Additional file 2. Amino acid and nucleotide sequence features of R2DBD-SB10

S2a. Amino acid sequence of the His-R2DBD-S3N10-SB protein resulting from the BmR2DBD (highlighted in yellow) fused at its C-terminal end with the SB10 transposase (blue letters). Between these two moieties there is a linker similar to that in the pMalc2vectors system (New England Biolabs) that produces protein fusion. This linker consists of a flexible S₃N₁₀ peptide (letters in red type), plus a Factor Xa cleavage site (green letters). We selected this linker because it has been empirically shown to retain the activity of both moieties in MBP fusion. The peptide containing the Histidine tag is typed in light gray, the His₆ tag is in bold.

MGSS**HHHHHH**SSGLVPRGSHMLD**MRTGDNPTVRGSAGADPVGQDAPGWTCQFCERTFSTNRGLGVHKR**
RAHPVETNTDAAPMMVKRRWHGEEIDLLARTEARLLAERGQCSGGDLFGALPGFGRTLEAIKGQRRRE
PYRALVQAHLARFGSQPGPSSGGCSAEPDDGN**SSSNNNNNNNNNN**LG**IEGRW**MGKSKEISQDLRKKIV
DLHKSGLSLGAISKRLKVPRSSVQ**TIVRKYKHHGTTQPSYRSGRRRVLSPRDERTLVRKVQINPRTTA**
KDLVKMLEETG**TKVSISTVKRVLYRHN**LKGRSARKKPLLQNRHKKARLRFATAHGDKDRFTWRNVLWS
DETKIELFGHNDHRYVWRKKGEACKPKNTIPTVKHGGGSIMLWGCFAAGGTGALHKIDGIMRKENYVD
ILKQHLKTSVRKLKLGRKWV**FQMDNDPKHTSKVVAKWLKDNKVKVLEWPSQSPDLNPIENLWAE**LKKR
VRARRPTNLTQLHQLCQEEWAKIHPTYCGKLVEGYPKRLTQVKQFKGNATKY

S2b. Nucleotide sequence of the gene encoding the His-R2DBD-S3N10-SB protein that was cloned in the mammal expression plasmid pCS2+. The color codes are the same than those used in S2a. *Eco*RI sites used to subclone the DNA segment encoding the His-R2DBD moiety are underlined.

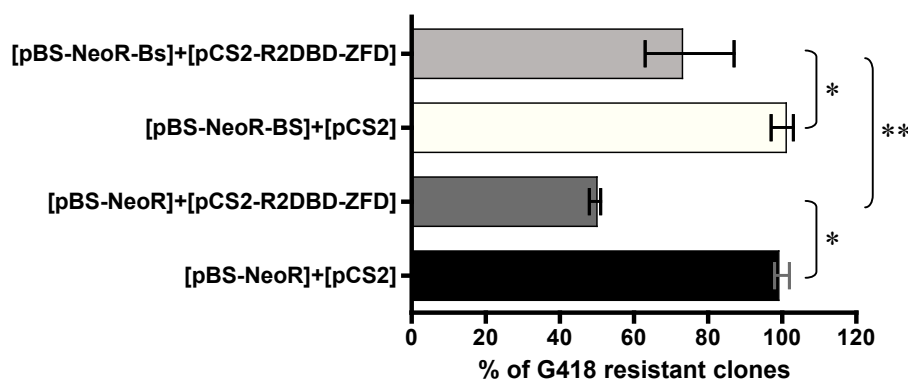
GACCATAAGCTTCTTTTTGAGGATCCCATCGATT**CGAATTCGCCGTCATGGGCAGCAGCCATCATCATCATCATC**
ACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCTCGAC**ATGCGCACCGGCGACAACCCACCGTGCGCGGCA**
GCGCCGGCGCCGACCCCGTGGGCCAGGACGCCCCGGCTGGACCTGCCAGTTCTGCGAGCGCACCTTCAGCACCA
ACCGCGGCCTGGGCGTGACACAAGCGCCGCGCCACCCCGTGGAGACCAACACCGACGCCGCCCCCATGATGGTGA
AGCGCCGCTGGCACGGCGAGGAGATCGACCTGCTGGCCCGCACCGAGGCCCCGCTGCTGGCCGAGCGCGGCCAGT
GCAGCGGCGGCGACCTGTTTCGGCGCCCTGCCCGGCTTCGGCCGCACCTTGGAGGCCATCAAGGGCCAGCGCCGCC
GCGAGCCCTACCGCGCCCTGGTGACGGCCACCTGGCCCGCTTCGGCAGCCAGCCCGGCCCCAGCAGCGGCGGCT
GCAGCGCCGAGCCCGAC**GATGGGAATTCGAGCTCGAACAACAACAACAATAACAATAACAACAAC****CTCGGGATCG**
AGGGAAGGTGGATGGGAAAATCAAAAGAAATCAGCCAAGACCTCAGAAAAAAATTTGTAGACCTCCACAAGTCTG
GTTTCATCCTTGGGAGCAATTTCCAAACGCCTGAAAGTACCACGTTTCATCTGTACAAACAATAGTACGCAAGTATA
AACACCATGGGACCACGCAGCCGTCATACCGCTCAGGAAGGAGACGCGTTCTGTCTCCTAGAGATGAACGTACTT
TGGTGCGAAAAGTGCAAATCAATCCAGAACAACAGCAAAGGACCTTGTGAAGATGCTGGAGGAAAACAGGTACAA
AAGTATCTATATCCACAGTAAAACGAGTCCTATATCGACATAACCTGAAAGGCCGCTCAGCAAGGAAGAAGCCAC
TGCTCCAAAACCGACATAAGAAAGCCAGACTACGGTTTGCAACTGCACATGGGGACAAAGATCGTACTTTTGGGA
GAAATGTCCTCTGGTCTGATGAAACAAAAATAGAAGTGTGGCCATAATGACCATCGTTATGTTTGGAGGAAGA
AGGGGGAGGCTTGCAAGCCGAAGAACACCATCCCAACCGTGAAGCACGGGGGTGGCAGCATCATGTTGTGGGGGT
GCTTTGCTGCAGGAGGGACTGGTGCACTTCAAAAATAGATGGCATCATGAGGAAGGAAAATATGTGGATATAT
TGAAGCAACATCTCAAGACATCAGTCAGGAAGTTAAAGCTTGGTCGCAAATGGGTCTTCCAAATGGACAATGACC
CCAAGCATACTTCCAAAGTTGTGGCAAAATGGCTTAAGGACAACAAAGTCAAGGTATTGGAGTGGCCATCACAAA
GCCCTGACCTCAATCCTATAGAAAATTTGTGGGCAGAACTGAAAAAGCGTGTGCGAGCAAGGAGGCCTACAAACC
TGACTCAGTTACACCAGCTCTGTGAGGAGGAATGGGCCAAAATTCACCCAACCTATTGTGGGAAGCTTGTGGAAG
GCTACCCGAAACGTTTGACCCAAGTTAAACAATTTAAAGGCAATGCTACCAAATACTAG

Additional file 3. Impact of R2DBD-NLS-ZFD on the integration by random recombination of pBS-NeoR or pBS-NeoR-ZBS (Ratio 1/1 (pCS2-SB10/pCS2-R2DBD-NLS-ZFD))

S3a. Sequence of the R2DBD-NLS-ZFD fusion. The peptide containing the Histidine tag is typed in dark gray and bold. The linker between the two moieties is typed in italics and bold, the flexible poly-Glycine tracts being typed in light gray, and the NLS in blue. The R2DBD and ZFD^{bcr-abl} parts are typed in black, the latter being underlined.

MGSSHHHHHSSGLVPRGSHMLDMRTGDNPTVRGSAGADPVGQDAPGWTCQFCERTFSTNRGLGVHKRRRAHPVET
NTDAAPMMVKRRWHGEEIDLLARTEARLLAERGQCSGGDLFGALPGFGRTLEAIKGQRRREPYPALVQAHLARFG
SQPGPSSGGCSAEPDDI **GGGGGPKKKRKL** **GGGGGGGGGG**EFMAEEKPFQCRICMRNFSDRSSLTRHTRHTTGEK
PFQCRICMRNFSQGGNLVRHLRTHHTGEKPFQCRICMRNFSQAATLQRHLKTHHTGEKPFQCRICMRNFSDRSNLER
HTRHTTGEKLEDRSGGGGGGGGGSGGGGGSSGEFI

S3b. Impact of the targeting peptide, R2DBD-NLS-ZFD^{bcr-abl}, on the production of NeoR clones after the transfection of 200 ng of a plasmid DNA mix into 5 x 10⁴ HeLa cells, followed by selection in G418 for two-weeks. The composition of each DNA plasmid mix is indicated in the left margin. It consisted of 100 ng of a plasmid source of NeoR cassette containing or not a specific binding site for ZFD^{bcr-abl} (pBS-NeoR or pBS-NeoR-ZBS) and 100 ng of a plasmid expressing or not the targeting peptide (pCS2-R2DBD-NLS-ZFD^{bcr-abl} or pCS2). Data were obtained from 9 replicates resulting from 3 different experiments. They are represented with median and the errors bars correspond to quartiles 1 and 3. For each plasmid source of NeoR cassette, (*) indicates significant differences (p>0.95). (**) indicates significant differences (p>0.95) between both plasmid sources of the NeoR cassette in the absence or presence of R2DBD-NLS-ZFD^{bcr-abl}.



The results show that the presence of the R2DBD-NLS-ZFD^{bcr-abl} peptide significantly reduced the rate of integration of pBS-NeoR and pBS-NeoR-ZBS (known as pBS-NeoR-Bs) by random recombination. Surprisingly, they also indicate that this decrease was greater (x2) with a plasmid that did not contain ZBS. Overall, the extents of the differences obtained at a 1/1 ratio between both plasmids are similar to those obtained with an R2DBD-NLS-N57SB10 peptide in similar plasmid conditions (**Figure 5h**).

Additional file 4. Amino acid sequence of the R2DBD-NLS-N57SB fusion

Sequence of the R2DBD-NLS-N57SB fusion. The peptide containing the Histidine tag is typed in dark gray and bold. The linker between the two moieties is typed in italics and bold, the flexible poly-Glycine tracts being typed in light gray, and the NLS in blue. The R2DBD and N57SB parts are typed in black, the latter being underlined.

MGSSHHHHHHSSGLVPRGSHMLDMRTGDNPTVRGSAGADPVGQDAPGWTCQFCERTFSTNRG
LGVHKRRRAHPVETNTDAAPMMVKRRWHGEEIDLLARTEARLLAERGQCSGGDLFGALPGFGR
TLEAIKGQRRREPYPYRALVQAHLARFGSQPGPSSGGCSAEPDDI **GGGGGPKKKRKL****AGGGGG**
GGGGEFEFMGKSKEISQDLRKKIVDLHKSGSSLGAISKRLKVPRSSVQTIVRKYKHHGTTQPSYR

Additional file 5. Impact of R2DBD peptides on cell viability

The viability of the HeLa cells transfected with plasmids expressing different R2DBD peptides was evaluated by monitoring their ability to proliferate during the 48 hours following their plasmid transfection.

Briefly, about 2.5×10^5 cells were seeded onto each 24-well plate, one day prior to transfection. Cells were transfected with jetPEI™ (Polyplus Transfection) at an N/P rate of 5, according to the Manufacturer's instructions (Polyplus Transfection). Plasmid DNA (0.5 µg) and jetPEI™ (1 µl) were each diluted in 50 µl of 150 mM NaCl, and then mixed together. After incubating for 15 min, the mixture was dropped into 500 µL of Opti-MEM medium (Invitrogen). Cells were then incubated with the complexes for 4 h. The transfection solution was then discarded, replaced by 1 mL fresh DMEM supplemented with 10% FBS, and incubated for 24 hours at 37°C. At 0, 24 and 48 hours post-transfection, the size of each treated cell population was evaluated with the "Cell Titer 96 Non-Radioactive Cell Proliferation Assay" kit, under the conditions recommended by the producer (Promega). Briefly, 45 µl of dye solution was added in each well, then incubated for 1h at 37°C, in an atmosphere containing 5% CO₂. 300 µl of the stopping solution were added and mixed with each sample. The OD was finally read at $\lambda = 620$ nm using a Mithras LB 940 (Berthold Technologies).

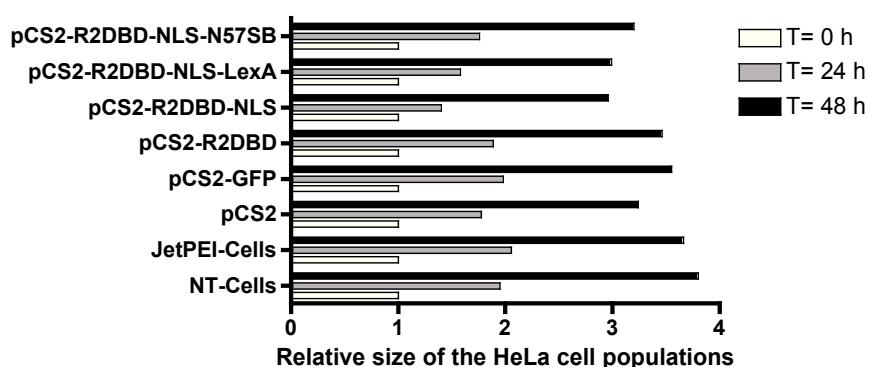


Figure S4a. HeLa cell proliferations under various transfection conditions. No treatment (NT-Cells), treatment with JetPEI (JetPEI-Cells), and transfection with pCS2 or pCS2-GFP were used as controls to verify the effect of the transfection on cell proliferation. Plasmid transfections with pCS2-R2DBD, pCS2-R2DBD-NLS, pCS2-R2DBD-NLS-LexA and pCS2-R2DBD-NLS-N57SB were done to evaluate the effect of R2DBD peptides on cell proliferation. For each treatment, data were obtained from 4 replicates resulting from 2 different experiments. They are represented using averaged relative size for each HeLa cells population. Under our experimental conditions and in agreement with jetPEI producer recommendations, FACS analyses allowed us to verify that more than 50% of the cells transfected with a pCS2-GFP expressed GFP, 24 hours and 48-hours post transfection (data not shown).

Results indicated that there was no difference in cell proliferation between control samples and samples in which cells expressed a R2DBD peptide. Similar results were obtained when plasmids expressing R2DBD fusion peptides were co-transfected with the SB vector system (pCS2-SB10 + pT2-pSV40-NeoR).

Additional file 6. Interest of targeting insertions of exogenous DNA fragment into rRNA genes

The need to perform targeted vector integrations into specific chromosomal sites is a major challenge in vectorology, for gene therapy purposes, but most of all in fields such as the bioproduction of therapeutic proteins in mammal cells. Indeed, the requirements related to transgene expression and the genotoxicity of the vectors used to make genetic modifications are similar in both fields, although the safety requirements are dramatically more stringent for gene therapy.

Apart from the important issues raised by producing a targeted vector for each purpose, two issues that do not depend on whether the integrative vector is of viral or non-viral origin, need to be addressed to comply with the safety and effectiveness requirements. Firstly, transgene expression must be sustained; most chromosomal integrations are in fact followed by fairly rapid silencing of the expression of the transgene ([Chen et al. 1997](#)), which has so far led to considerable underestimation of the integration of certain vectors, as recently illustrated for *Sleeping Beauty* in mammalian cells (SB; [Garrison et al. 2007](#); [Dalsgaard et al. 2009](#)). Methods for circumventing the silencing of transgene expression have been developed, and usually consist of including an insulator, such as the MAR, UCOE, STAR and LCR elements, in the integrative vector (reviewed in [Harraghy et al. 2008](#)). However, the effectiveness of these elements depends considerably on the chromosomal location where the vector is integrated, its integration mode, and the features of the vector sequence. Secondly, the genotoxicity resulting from random integrations of the vector into chromosomes must be controlled, because this can cause deleterious genetic side effects, such as the emergence of proliferative cancer cells in the gene therapy context. Developing vectors that integrate specifically into a non-genic euchromatic target is currently thought to offer a promising way to circumvent this problem. However, in the light of the difficulties encountered in defining the genomic boundaries of a gene (reviewed in [Gerstein et al. 2007](#)) and, incidentally, the fact that this definition depends on the cell type, developing viral and non-viral vectors with secure insertion specificity appears to be a challenging goal.

The proposition that an optimum way to integrate a vector is to target it into rRNA genes has been accepted as a valuable strategy in bioproduction for more than thirty years ([Klabunde et al. 2004](#); [dos Santos and Buck 2000](#)). Incidentally, it must be remembered that the rRNA genes located in the nuclei are the sequences most transcribed in the genomes, several orders of magnitude above than any other genes. The proofs of principle that this solution is also useful for gene therapy purposes have been provided more recently ([Liu et al. 2007](#); [Stewart et al. 2002](#); [Wen et al. 2008](#)). Important information supporting the concept that specific integrations of DNA fragments into a few rRNA genes are non-deleterious, have also been provided by several long-term studies carried out to elucidate the transposition machinery of various molecular parasites. Indeed, some of them have a strategy for maintaining themselves in eukaryotic genomes that consists of having few or no deleterious post-integrative effects, by inserting within conserved repeated genes ([Kojima and Fujiwara 2004](#)). Data recovered from the group-I self-splicing introns that encode I-*PpoI*, the R1 to R7 non-LTR retrotransposons ([Kojima and Fujiwara 2005](#)), and the *Pokey* DNA transposon ([Penton et al. 2004](#)), indicate that rRNA genes have been selected as a non-deleterious solution several times during eukaryote evolution, since these elements have no evolutionary relationships, and occur in entirely distinct lineages of eukaryotes. These convergences are striking, since these three kinds of molecular parasites all have their insertion site in a 24-bp segment located within a 100-bp region that is 99 to 100% conserved from yeast to humans. Another interesting point is that transgenes integrated into these genes can be expressed, whatever kind of promoter is involved. Their expression level is optimal, because they are not

silenced when only a few rRNA genes copies are interrupted, due to the natural presence of insulators in these loci (Robinett et al. 1997).

From an environmental standpoint, the production of genetically modified organisms (GMO) by specifically integrating transgene cassettes into rRNA genes might help to control the diffusion of the transgene among non-genetically modified conspecifics or members of sibling species able to cross-hybridize. Indeed, some of the properties of rRNA genes might be useful to prevent their diffusion. Although their number can vary by a factor 2 in the genome of one species, rRNA genes are highly repetitive because they are essential for protein synthesis and thus for the viability of the organism. Their need is so essential that the evolution has retained, since the eukaryote origin, two properties for these genes. First, there are far more of them than the organism needs. Second, their functionality is preserved by a mechanism for monitoring the homogeneity of their sequence, which is known as magnification (Terracol 1987). This mechanism is specific to the germ line, and eliminates defective rRNA gene copies disrupted by DNA fragment integrations or deletions within about 2-5 generations. In the absence of selection, the vectors integrated into the repeated rRNA genes should therefore be labile and so not maintained in natural environments. This strategy might therefore be a technical solution to avoid the spread of transgenes into non-GM crops and natural populations of plants and animals.

References in S6 that are not included in the main text

- Chen WY, Bailey EC, McCune SL, Dong JY, Townes TM (1997) Reactivation of silenced, virally transduced genes by inhibitors of histone deacetylase. *Proc Natl Acad Sci USA* 94:5798-5803.
- Garrison BS, Yant SR, Mikkelsen JG, Kay MA (2007) Postintegrative gene silencing within the *Sleeping Beauty* transposition system. *Mol Cell Biol* 27:8824-8833.
- Dalsgaard T, Moldt B, Sharma N, Wolf G, Schmitz A, Pedersen FS, Mikkelsen JG (2009) Shielding of *Sleeping beauty* DNA transposon-delivered transgene cassettes by heterologous insulators in early embryonal cells. *Mol Ther* 17:121-130.
- Harraghy N, Gaussin A, Mermod N (2008) Sustained transgene expression using MAR elements. *Curr Gene Ther* 8:353-366.
- Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M (2007) What is a gene, post-ENCODE? History and updated definition. *Genome Res* 17:669-681.
- Kojima KK, Fujiwara H (2005) An extraordinary retrotransposon family encoding dual endonucleases. *Gen Res* 15:1106-1117.
- Terracol R (1987) Differential magnification of rDNA gene types in bobbed mutants of *Drosophila melanogaster*. *Mol Gen Genet* 208:168-176.

Characterization of monomeric protein domains that bind specifically to a highly-conserved 100-bp target within rRNA genes

Elodie Carnus, Marie-Véronique Demattei, Sophie Casteret, Guillaume Carpentier, Fabien Palazzoli, Solenne Bire, Christophe Bressac, and Yves Bigot

Additional material

Additional file. Design and properties of pZFD proteins

In silico design of pZFD proteins.

Three different pZFDs were designed *in silico* following the method described by Mandell and Barbas¹, and using the backbone sequence YKCPECGKSFS-X₇-HQRT¹, in which X₇ corresponds to the variable amino acid motifs that allocate their trinucleotide binding specificity to each ZF. In each of the ZF trimers (**Figure 1c**, ZFD1 and ZFD2), the ZFs were linked to the consensus linker, TGEKP. A flexible linker consisting of seven glycines (Gly)₇ was used to link the ZF trimers². Finally, the specific peptides LEPGEKP and TGKKTS were added at the N- and C-terminal ends, respectively, of the three pZFD proteins (**Figure S1a**). In the first pZFD, known as ZFDBarbas, the ZF with the greatest affinity for each of the 6 trinucleotides (CGG, GTA, AAC, GGA, GTA and ACT) was selected to make the fusion protein (**Figure S1a**, lane 1). Two other pZFDs were optimized using ZiFiT facilities²⁸. The second protein, called ZFDSangamo, was designed using the Sangamo method that optimized each ZF trimer, taking into account the position occupied by each ZF in the fusion (**Figure S1b**, lane 2). The Toolgen method, that optimized pZFD from data recovered from natural pZFD encoded by the human genome, was used for the third pZFD, and designated ZFDToolgen (**Figure S1a**, lane 3). When the trimer of ZF could not be optimized, the ZF corresponding to ZFDBarbas was used. In view of its position in each ZF trimer, calculations done using ZiFDB facilities³ could not be used to further optimize the choice of each ZF in our pZFD. It nevertheless revealed that all the ZFs selected in ZFDSangamo corresponded to the theoretically optimal solutions.

a.

```
(1) MLEPGEKPYKPECGKSFSSTHLDLIRHQRTHTGEKPYKPECGKSFSQSSSLVRHQRTHTGEKPYKPECGKSFSQRAHLRHQRTGGG
(2) MLEPGEKPYKPECGKSFSSTHLDLIRHQRTHTGEKPYKPECGKSFSQSGALARRHQRTHTGEKPYKPECGKSFSQSGHLQRHQRTGGG
(3) MLEPGEKPYKPECGKSFSSTHLDLIRHQRTHTGEKPYKPECGKSFSQSSSTLTRHQRTHTGEKPYKPECGKSFSQSTHLRHQRTGGG

(1) GGGGYKPECGKSFSDSGNLRVHQRTHTGEKPYKPECGKSFSQSSSLVRHQRTHTGEKPYKPECGKSFSRSDKLTETHTTGKKTS
(2) GGGGYKPECGKSFSDSGNLRVHQRTHTGEKPYKPECGKSFSQSGALARRHQRTHTGEKPYKPECGKSFSRSDNLTETHTTGKKTS
(3) GGGGYKPECGKSFSDSGNLRVHQRTHTGEKPYKPECGKSFSQSSSTLTRHQRTHTGEKPYKPECGKSFSRSDNLTETHTTGKKTS
```

Figure S1a. Sequence features of pZFD that were designed to bind to the conserved 100-bp region located within the rRNA genes encoding the 28S rRNA. Amino acid sequences of the three bipartite pZFD used in our work: (1) ZFDBarbas, (2) ZFDSangamo and (3) ZFDToolgen. Variable regions in each ZF are highlighted in gray. C₂H₂ motifs are highlighted in yellow or green in each ZF. The linkers between ZFD1 and ZFD2 are highlighted in black and typed in white.

DNA fragment synthesis

DNA fragments containing the 100-bp segment including the binding sites of pZFD and R2DBD proteins (**Figure 1a and S1a**) were synthesized by ATGBiosynthetics (Germany) and cloned in pUC57 (Fermentas) on the basis of a codon usage avoiding codons that are rare in human cells.

Binding specificity of pZFD protein to ZBS

The three pZFD proteins were cloned in the expression plasmids: in pMalc2 to produce protein fused with the Maltose binding protein (MBP), and in pET14b to generate fusions with a His₆ tag at their N-terminal ends. These fusion proteins were produced in bacteria and then purified. Their abilities to bind to the R2site probe containing ZBS (**Figure 1b**) were evaluated by EMSA (electrophoretic mobility shift assay). Results revealed the formation of similar shifted complexes with MBP-ZFDBarbas, MBP-ZFDSangamo and MBP-ZFDToolgen (**Figure S2, lanes 1 to 3**), the most efficient binding being obtained with the first of these proteins. In contrast, no shifted complex was obtained with the three His₆-tagged proteins, indicating that the tag prevented either the activity or the folding of these fusions, as previously reported for other basic proteins such as transposases⁷. Although the specificity of the shifted complex obtained with the three MBP-tagged fusions had already been confirmed

by the presence of a saturating amount of non-specific DNA competitor in each sample assayed, it was nevertheless confirmed that no shifted complex was obtained with a non-specific probe, such as the 3' inverted terminal repeat of the *MosI* transposon (data not shown).

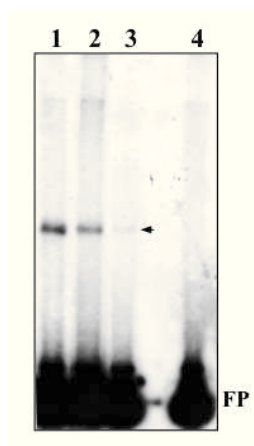


Figure S1b. EMSA analysis of the three pZFDs fused to the MBP. Lane 1, MBP-ZFDBarbas; lane 2, MBP-ZFDSangamo; lane 3, MBP-ZFDToolgen and lane 4, R2site probe alone. FP, indicates the free R2site probe in gel, and the arrow, the MBP-pZFD/R2site complexes. Similar results were obtained whatever the binding conditions (temperature of 20, 30 or 37°C, and 0 or 5 mM ZnCl₂).

Because of its greater binding activity, the MBP-ZFDBarbas protein was identified as being the best candidate, and its apparent K_d was determined. EMSA experiments were first carried out using a fixed concentration of the R2site probe, and serially diluted protein. A saturation graph was obtained that made it possible to specify a fixed concentration of 39 nM of the MBP-ZFDBarbas protein (data not shown). Under these conditions, the apparent K_d was estimated to be over 1 μ M, a value characteristic of a fusion protein with low affinity for its binding site.

Overall, our data indicate that it was not possible to design an efficient pZFD *in silico* that would be able to bind to the conserved 100-bp segment of the gene encoding 28S rRNA. An alternative approach might therefore be to use a selection procedure to obtain an efficient pZFD. However, this raises questions about the pertinence of using a molecular selection procedure with, for example, the open source of standardized reagents and protocols for engineering ZF by modular assembly⁹. Indeed, since an open source of standardized reagents and protocols for engineering ZF by modular assembly⁸ is now available, one solution might have been to pursue our investigations by selecting a more efficient pZFD *in vitro*. It must be pointed out that this approach would be still more time-consuming and expensive than the *in silico* design approach. Beside the fact we had the R2DBD alternative solution available in the lab, one of the reasons that we did not further develop pZFD are the conditions to which they are subject when they have been developed or exploited for commercial purposes. Any pZFD engineered by public or private research laboratories is in fact controlled by private bodies, due to the intellectual property (IP) issues that surround and constrain this technology¹⁰⁻¹³. Since our aim is to develop DBD tools for gene targeting that can subsequently be made freely available to the scientific and medical communities, we focused our investigation on the BmR2DBD and DmR2DBD. From this perspective, it must be pointed out that the strategy consisting of integrating vectors specifically into rRNA genes, and using molecular tools derived from a non-LTR retrotransposon is IP-free, because a 1994 patent belonging to Transgene S.A. has been abandoned (WO/1994/024300 - Transposition assembly for gene transfer in eukaryotes. Applicant: Transgene S.A.; inventor Jacobs, E), thus liberating this strategy for free exploitation.

References cited in S1

1. Mandell, JG and Barbas, CF (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* **34** (Web Server issue): W516–W523.
2. Imanishi, M, Hori, Y, Nagaoka, M and Sugiura, Y (2000) DNA-bending finger: artificial design of 6-zinc finger peptides with polyglycine linker and induction of DNA bending. *Biochemistry* **39**: 4383-4390.
3. Fu, F, Sander, JD, Maeder, M, Thibodeau-Beganny, S, Joung, JK, Dobbs, D, Miller, L and Voytas, DF (2009) Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. *Nucleic Acids Res* **37** (Database issue): D279-283.
4. Penton, EH and Crease, TJ (2004) Evolution of the transposable element *Pokey* in the ribosomal DNA of species in the subgenus *Daphnia* (Crustacea: Cladocera). *Mol Biol Evol* **21**: 1727-1739.
5. Christensen, SM, Bibillo, A and Eickbush, TH (2005) Role of the Bombyx mori R2 element N-terminal domain in the target-primed reverse transcription (TPRT) reaction. *Nucleic Acids Res* **33**:6461–6468.
6. Ogata K, Morikawa S, Nakamura H, Sekikawa A, Inoue T, Kanai H, Sarai A, Ishii, S and Nishimura, Y (1994) Solution structure of a specific DNA complex of the *Myb* DNA-binding domain with cooperative recognition helices. *Cell* **79**: 639-648.
7. Augé-Gouillou, C, Hamelin, MH, Demattei, MV, Periquet, G and Bigot, Y (2001) The ITR binding domain of the *Mariner Mos-1* transposase. *Mol Genet Genomics* **265**: 58-65.
8. Wright, DA, Thibodeau-Beganny, S, Sander, JD, Winfrey, RJ, Hirsh, AS, Eichinger, M, Fu, F, Porteus, MH, Dobbs, D, Voytas, DF and Joung, JK (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc* **1**: 1637-1652.
9. Kojima, KK and Fujiwara, H (2004) Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. *Mol Biol Evol* **21**: 207-217.
10. Palazzoli, F, Carnus, E, Wells, DJ and Bigot, Y (2008) Sustained transgene expression using non-viral enzymatic systems for stable chromosomal integration. *Curr Gene Ther* **8**: 367-390.
11. Lyznik, LA and Dress, V (2008) Gene targeting for chromosome engineering applications in eukaryotic cells. *Recent Pat Biotechnol* **2**: 94-106.
12. Palazzoli F (2007) Analyses of the consequences of using Zinc Finger Domains for fusion proteins, in a R&D exploitation plan - Impact of the intellectual property on the economic viability for a technological development strategy in *ex vivo* gene therapy, of : i) transposon-based integration mechanisms, ii) their modifications to become site-specific. *Master thesis in Biotechnology and laws*. University François Rabelais of Tours, 60 pages.
13. Palazzoli, F, Carnus, E, Merly, F, Goiran, S and Bigot, Y. (2010) Zinc finger monopoly : quelles sont les règles du jeu ? *Biofutur* **306**: 58-62

Additional file 2. Amino acid and nucleotide sequence features of R2DBD-SB10

S2a. Amino acid sequence of the His-R2DBD-S3N10-SB protein resulting from the BmR2DBD (highlighted in yellow) fused at its C-terminal end with the SB10 transposase (blue letters). Between these two moieties there is a linker similar to that in the pMalc2vectors system (New England Biolabs) that produces protein fusion. This linker consists of a flexible S₃N₁₀ peptide (letters in red type), plus a Factor Xa cleavage site (green letters). We selected this linker because it has been empirically shown to retain the activity of both moieties in MBP fusion. The peptide containing the Histidine tag is typed in light gray, the His₆ tag is in bold.

MGSS**HHHHHH**SSGLVPRGSHMLD**MRTGDNPTVRGSAGADPVGQDAPGWTCQFCERTFSTNRGLGVHKR**
RAHPVETNTDAAPMMVKRRWHGEEIDLLARTEARLLAERGQCSGGDLFGALPGFGRTLEAIKGQRRRE
PYRALVQAHLARFGSQPGPSSGGCSAEPDDGN**SSSNNNNNNNNNN**LG**IEGRW**MGKSKEISQDLRKKIV
DLHKSGLSLGAISKRLKVPRSSVQ**TIVRKYKHHGTTQPSYRSGRRRVLSPRDERTLVRKVQINPRTTA**
KDLVKMLEETG**TKVSISTVKRVLYRHN**LKGRSARKKPLLQNRHKKARLRFATAHGDKDRFTWRNVLWS
DETKIELFGHNDHRYVWRKKGEACKPKNTIPTVKHGGGSIMLWGCF**AAGGTGALHKIDGIMRKENYVD**
ILKQHLKTSVRKLKLGRKWV**FQMDNDPKHTSKVVAKWLKDNKVKVLEWPSQSPDLNPIENLWAE**LKKR
VRARRPTNLTQLHQLCQEEWAKIHPTYCGKLVEGYPKRLTQVKQFKGNATKY

S2b. Nucleotide sequence of the gene encoding the His-R2DBD-S3N10-SB protein that was cloned in the mammal expression plasmid pCS2+. The color codes are the same than those used in S2a. *Eco*RI sites used to subclone the DNA segment encoding the His-R2DBD moiety are underlined.

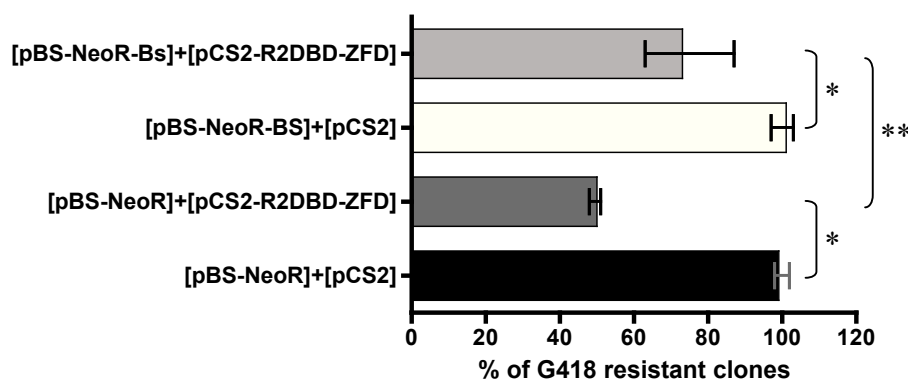
GACCATAAGCTTCTTTTTGAGGATCCCATCGATT**CGAATTCGCCGTCATGGGCAGCAGCCATCATCATCATCATC**
ACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCTCGAC**ATGCGCACCGGCGACAACCCACCGTGCGCGGCA**
GCGCCGGCGCCGACCCCGTGGGCCAGGACGCCCCGGCTGGACCTGCCAGTTCTGCGAGCGCACCTTCAGCACCA
ACCGCGGCCTGGGCGTGACACAAGCGCCGCGCCACCCCGTGGAGACCAACACCGACGCCGCCCCCATGATGGTGA
AGCGCCGCTGGCACGGCGAGGAGATCGACCTGCTGGCCCGCACCGAGGCCCCGCTGCTGGCCGAGCGCGGCCAGT
GCAGCGGCGGCGACCTGTTTCGGCGCCCTGCCCGGCTTCGGCCGCACCTGGAGGCCATCAAGGGCCAGCGCCGCC
GCGAGCCCTACCGCGCCCTGGTGACGGCCACCTGGCCCGCTTCGGCAGCCAGCCCGGCCCCAGCAGCGGCGGCT
GCAGCGCCGAGCCCGACGATGGGAATTCGAGCTCGAACAACAACAATAACAATAACAACAACCTCGGGATCG
AGGGAAGGTGGATGGGAAAATCAAAAGAAATCAGCCAAGACCTCAGAAAAAAATTTGTAGACCTCCACAAGTCTG
GTTTCATCCTTGGGAGCAATTTCCAAACGCCTGAAAGTACCACGTTTCATCTGTACAAACAATAGTACGCAAGTATA
AACACCATGGGACCACGCAGCCGTCATACCGCTCAGGAAGGAGACGCGTTCTGTCTCCTAGAGATGAACGTACTT
TGGTGCGAAAAGTGCAAATCAATCCAGAACAACAGCAAAGGACCTTGTGAAGATGCTGGAGGAAAACAGGTACAA
AAGTATCTATATCCACAGTAAAACGAGTCTATATCGACATAACCTGAAAGGCCGCTCAGCAAGGAAGAAGCCAC
TGCTCCAAAACCGACATAAGAAAGCCAGACTACGGTTTGCAACTGCACATGGGGACAAAGATCGTACTTTTGGGA
GAAATGTCCTCTGGTCTGATGAAACAAAAATAGAAGTGTGGCCATAATGACCATCGTTATGTTTGGAGGAAGA
AGGGGGAGGCTTGCAAGCCGAAGAACACCATCCCAACCGTGAAGCACGGGGGTGGCAGCATCATGTTGTGGGGGT
GCTTTGCTGCAGGAGGGACTGGTGCACTTCAAAAATAGATGGCATCATGAGGAAGGAAAAATTATGTGGATATAT
TGAAGCAACATCTCAAGACATCAGTCAGGAAGTTAAAGCTTGGTCGCAAATGGGTCTTCCAAATGGACAATGACC
CCAAGCATACTTCCAAAGTTGTGGCAAAATGGCTTAAGGACAACAAAGTCAAGGTATTGGAGTGGCCATCACAAA
GCCCTGACCTCAATCCTATAGAAAATTTGTGGGCAGAACTGAAAAAGCGTGTGCGAGCAAGGAGGCCTACAAACC
TGACTCAGTTACACCAGCTCTGTGAGGAGGAATGGGCCAAAATTCACCCAACCTATTGTGGGAAGCTTGTGGAAG
GCTACCCGAAACGTTTGACCCAAGTTAAACAATTTAAAGGCAATGCTACCAAATACTAG

Additional file 3. Impact of R2DBD-NLS-ZFD on the integration by random recombination of pBS-NeoR or pBS-NeoR-ZBS (Ratio 1/1 (pCS2-SB10/pCS2-R2DBD-NLS-ZFD))

S3a. Sequence of the R2DBD-NLS-ZFD fusion. The peptide containing the Histidine tag is typed in dark gray and bold. The linker between the two moieties is typed in italics and bold, the flexible poly-Glycine tracts being typed in light gray, and the NLS in blue. The R2DBD and ZFD^{bcr-abl} parts are typed in black, the latter being underlined.

MGSSHHHHHSSGLVPRGSHMLDMRTGDNPTVRGSAGADPVGQDAPGWTCQFCERTFSTNRGLGVHKRRRAHPVET
NTDAAPMMVKRRWHGEEIDLLARTEARLLAERGQCSGGDLFGALPGFGRTLEAIKGQRRREPYPALVQAHLARFG
SQPGPSSGGCSAEPDDI **GGGGGPKKKRKL** **GGGGGGGGGG**EFMAEEKPFQCRICMRNFSDRSSLTRHTRHTTGEK
PFQCRICMRNFSQGGNLVRHLRTHHTGEKPFQCRICMRNFSQAATLQRHLKTHHTGEKPFQCRICMRNFSDRSNLER
HTRHTTGEKLEDRSGGGGGGGGGSGGGSSGEFI

S3b. Impact of the targeting peptide, R2DBD-NLS-ZFD^{bcr-abl}, on the production of NeoR clones after the transfection of 200 ng of a plasmid DNA mix into 5 x 10⁴ HeLa cells, followed by selection in G418 for two-weeks. The composition of each DNA plasmid mix is indicated in the left margin. It consisted of 100 ng of a plasmid source of NeoR cassette containing or not a specific binding site for ZFD^{bcr-abl} (pBS-NeoR or pBS-NeoR-ZBS) and 100 ng of a plasmid expressing or not the targeting peptide (pCS2-R2DBD-NLS-ZFD^{bcr-abl} or pCS2). Data were obtained from 9 replicates resulting from 3 different experiments. They are represented with median and the errors bars correspond to quartiles 1 and 3. For each plasmid source of NeoR cassette, (*) indicates significant differences (p>0.95). (**) indicates significant differences (p>0.95) between both plasmid sources of the NeoR cassette in the absence or presence of R2DBD-NLS-ZFD^{bcr-abl}.



The results show that the presence of the R2DBD-NLS-ZFD^{bcr-abl} peptide significantly reduced the rate of integration of pBS-NeoR and pBS-NeoR-ZBS (known as pBS-NeoR-Bs) by random recombination. Surprisingly, they also indicate that this decrease was greater (x2) with a plasmid that did not contain ZBS. Overall, the extents of the differences obtained at a 1/1 ratio between both plasmids are similar to those obtained with an R2DBD-NLS-N57SB10 peptide in similar plasmid conditions (**Figure 5h**).

Additional file 4. Amino acid sequence of the R2DBD-NLS-N57SB fusion

Sequence of the R2DBD-NLS-N57SB fusion. The peptide containing the Histidine tag is typed in dark gray and bold. The linker between the two moieties is typed in italics and bold, the flexible poly-Glycine tracts being typed in light gray, and the NLS in blue. The R2DBD and N57SB parts are typed in black, the latter being underlined.

MGSSHHHHHHSSGLVPRGSHMLDMRTGDNPTVRGSAGADPVGQDAPGWTCQFCERTFSTNRG
LGVHKRRRAHPVETNTDAAPMMVKRRWHGEEIDLLARTEARLLAERGQCSGGDLFGALPGFGR
TLEAIKGQRRREPYRALVQAHLARFGSQPGPSSGGCSAEPDDI ***GGGGGPKKKRKL*** ***GGGGG***
GGGG EFEFMGKSKEISQDLRKKIVDLHKSGSSLGAISKRLKVPRSSVQTIVRKYKHHGTTQPSYR

Additional file 5. Impact of R2DBD peptides on cell viability

The viability of the HeLa cells transfected with plasmids expressing different R2DBD peptides was evaluated by monitoring their ability to proliferate during the 48 hours following their plasmid transfection.

Briefly, about 2.5×10^5 cells were seeded onto each 24-well plate, one day prior to transfection. Cells were transfected with jetPEI™ (Polyplus Transfection) at an N/P rate of 5, according to the Manufacturer's instructions (Polyplus Transfection). Plasmid DNA (0.5 µg) and jetPEI™ (1 µl) were each diluted in 50 µl of 150 mM NaCl, and then mixed together. After incubating for 15 min, the mixture was dropped into 500 µL of Opti-MEM medium (Invitrogen). Cells were then incubated with the complexes for 4 h. The transfection solution was then discarded, replaced by 1 mL fresh DMEM supplemented with 10% FBS, and incubated for 24 hours at 37°C. At 0, 24 and 48 hours post-transfection, the size of each treated cell population was evaluated with the "Cell Titer 96 Non-Radioactive Cell Proliferation Assay" kit, under the conditions recommended by the producer (Promega). Briefly, 45 µl of dye solution was added in each well, then incubated for 1h at 37°C, in an atmosphere containing 5% CO₂. 300 µl of the stopping solution were added and mixed with each sample. The OD was finally read at $\lambda = 620$ nm using a Mithras LB 940 (Berthold Technologies).

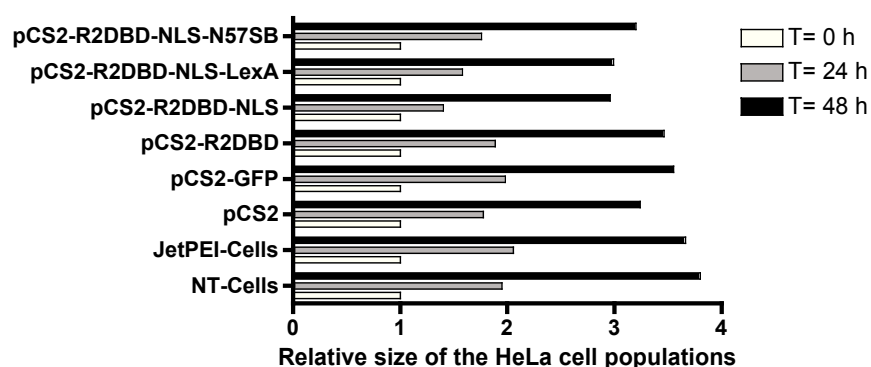


Figure S4a. HeLa cell proliferations under various transfection conditions. No treatment (NT-Cells), treatment with JetPEI (JetPEI-Cells), and transfection with pCS2 or pCS2-GFP were used as controls to verify the effect of the transfection on cell proliferation. Plasmid transfections with pCS2-R2DBD, pCS2-R2DBD-NLS, pCS2-R2DBD-NLS-LexA and pCS2-R2DBD-NLS-N57SB were done to evaluate the effect of R2DBD peptides on cell proliferation. For each treatment, data were obtained from 4 replicates resulting from 2 different experiments. They are represented using averaged relative size for each HeLa cells population. Under our experimental conditions and in agreement with jetPEI producer recommendations, FACS analyses allowed us to verify that more than 50% of the cells transfected with a pCS2-GFP expressed GFP, 24 hours and 48-hours post transfection (data not shown).

Results indicated that there was no difference in cell proliferation between control samples and samples in which cells expressed a R2DBD peptide. Similar results were obtained when plasmids expressing R2DBD fusion peptides were co-transfected with the SB vector system (pCS2-SB10 + pT2-pSV40-NeoR).

Additional file 6. Interest of targeting insertions of exogenous DNA fragment into rRNA genes

The need to perform targeted vector integrations into specific chromosomal sites is a major challenge in vectorology, for gene therapy purposes, but most of all in fields such as the bioproduction of therapeutic proteins in mammal cells. Indeed, the requirements related to transgene expression and the genotoxicity of the vectors used to make genetic modifications are similar in both fields, although the safety requirements are dramatically more stringent for gene therapy.

Apart from the important issues raised by producing a targeted vector for each purpose, two issues that do not depend on whether the integrative vector is of viral or non-viral origin, need to be addressed to comply with the safety and effectiveness requirements. Firstly, transgene expression must be sustained; most chromosomal integrations are in fact followed by fairly rapid silencing of the expression of the transgene (Chen et al. 1997), which has so far led to considerable underestimation of the integration of certain vectors, as recently illustrated for *Sleeping Beauty* in mammalian cells (SB; Garrison et al. 2007; Dalsgaard et al. 2009). Methods for circumventing the silencing of transgene expression have been developed, and usually consist of including an insulator, such as the MAR, UCOE, STAR and LCR elements, in the integrative vector (reviewed in Harraghy et al. 2008). However, the effectiveness of these elements depends considerably on the chromosomal location where the vector is integrated, its integration mode, and the features of the vector sequence. Secondly, the genotoxicity resulting from random integrations of the vector into chromosomes must be controlled, because this can cause deleterious genetic side effects, such as the emergence of proliferative cancer cells in the gene therapy context. Developing vectors that integrate specifically into a non-genic euchromatic target is currently thought to offer a promising way to circumvent this problem. However, in the light of the difficulties encountered in defining the genomic boundaries of a gene (reviewed in Gerstein et al. 2007) and, incidentally, the fact that this definition depends on the cell type, developing viral and non-viral vectors with secure insertion specificity appears to be a challenging goal.

The proposition that an optimum way to integrate a vector is to target it into rRNA genes has been accepted as a valuable strategy in bioproduction for more than thirty years (Klabunde et al. 2004; dos Santos and Buck 2000). Incidentally, it must be remembered that the rRNA genes located in the nuclei are the sequences most transcribed in the genomes, several orders of magnitude above than any other genes. The proofs of principle that this solution is also useful for gene therapy purposes have been provided more recently (Liu et al. 2007; Stewart et al. 2002; Wen et al. 2008). Important information supporting the concept that specific integrations of DNA fragments into a few rRNA genes are non-deleterious, have also been provided by several long-term studies carried out to elucidate the transposition machinery of various molecular parasites. Indeed, some of them have a strategy for maintaining themselves in eukaryotic genomes that consists of having few or no deleterious post-integrative effects, by inserting within conserved repeated genes (Kojima and Fujiwara 2004). Data recovered from the group-I self-splicing introns that encode I-*PpoI*, the R1 to R7 non-LTR retrotransposons (Kojima and Fujiwara 2005), and the *Pokey* DNA transposon (Penton et al. 2004), indicate that rRNA genes have been selected as a non-deleterious solution several times during eukaryote evolution, since these elements have no evolutionary relationships, and occur in entirely distinct lineages of eukaryotes. These convergences are striking, since these three kinds of molecular parasites all have their insertion site in a 24-bp segment located within a 100-bp region that is 99 to 100% conserved from yeast to humans. Another interesting point is that transgenes integrated into these genes can be expressed, whatever kind of promoter is involved. Their expression level is optimal, because they are not

silenced when only a few rRNA genes copies are interrupted, due to the natural presence of insulators in these loci (Robinett et al. 1997).

From an environmental standpoint, the production of genetically modified organisms (GMO) by specifically integrating transgene cassettes into rRNA genes might help to control the diffusion of the transgene among non-genetically modified conspecifics or members of sibling species able to cross-hybridize. Indeed, some of the properties of rRNA genes might be useful to prevent their diffusion. Although their number can vary by a factor 2 in the genome of one species, rRNA genes are highly repetitive because they are essential for protein synthesis and thus for the viability of the organism. Their need is so essential that the evolution has retained, since the eukaryote origin, two properties for these genes. First, there are far more of them than the organism needs. Second, their functionality is preserved by a mechanism for monitoring the homogeneity of their sequence, which is known as magnification (Terracol 1987). This mechanism is specific to the germ line, and eliminates defective rRNA gene copies disrupted by DNA fragment integrations or deletions within about 2-5 generations. In the absence of selection, the vectors integrated into the repeated rRNA genes should therefore be labile and so not maintained in natural environments. This strategy might therefore be a technical solution to avoid the spread of transgenes into non-GM crops and natural populations of plants and animals.

References in S6 that are not included in the main text

- Chen WY, Bailey EC, McCune SL, Dong JY, Townes TM (1997) Reactivation of silenced, virally transduced genes by inhibitors of histone deacetylase. *Proc Natl Acad Sci USA* 94:5798-5803.
- Garrison BS, Yant SR, Mikkelsen JG, Kay MA (2007) Postintegrative gene silencing within the *Sleeping Beauty* transposition system. *Mol Cell Biol* 27:8824-8833.
- Dalsgaard T, Moldt B, Sharma N, Wolf G, Schmitz A, Pedersen FS, Mikkelsen JG (2009) Shielding of *Sleeping beauty* DNA transposon-delivered transgene cassettes by heterologous insulators in early embryonal cells. *Mol Ther* 17:121-130.
- Harraghy N, Gaussin A, Mermoud N (2008) Sustained transgene expression using MAR elements. *Curr Gene Ther* 8:353-366.
- Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M (2007) What is a gene, post-ENCODE? History and updated definition. *Genome Res* 17:669-681.
- Kojima KK, Fujiwara H (2005) An extraordinary retrotransposon family encoding dual endonucleases. *Gen Res* 15:1106-1117.
- Terracol R (1987) Differential magnification of rDNA gene types in bobbed mutants of *Drosophila melanogaster*. *Mol Gen Genet* 208:168-176.

3.5. Article 6 : *Patent landscape of chromatin control elements: position effects in the pharmaceutical bioproduction*

3.5.1. Contexte et objectif de l'article

Le transfert de gènes dans les cellules eucaryotes est limité par des effets épigénétiques entraînant un silençage de l'expression du transgène et de fortes variations d'expression entre les cellules. La structure de la chromatine au site d'insertion semble être un paramètre critique à l'égard de l'expression du transgène, dont le niveau de transcription dépend du site d'intégration. L'intégration dans l'euchromatine (chromatine « ouverte » donc activement transcrite) favorise une expression élevée et stable du transgène, à l'inverse d'une intégration dans l'hétérochromatine (chromatine « condensée » non transcrite). Les séquences d'ADN intégrées peuvent être reconnues comme étant des séquences exogènes par la machinerie de défense nucléaire, provoquant ainsi une hétérochromatinisation et donc l'extinction de son expression. Des études ont montré l'importance de générer des vecteurs d'intégration capables de surmonter l'extinction de l'expression du transgène après son intégration chromosomique [Pfeifer *et al.*, 2002]. L'article de Chen montre que 80% des transgènes sont silencés dans les cellules hôtes HeLa [Chen *et al.*, 1997]. Comme tous les vecteurs intégratifs, les vecteurs non viraux tels que les transposons sont soumis aux effets de silençage du transgène [Garrison *et al.*, 2007].

Pour résoudre ce problème, deux solutions peuvent être envisagées. La première porte sur l'augmentation du nombre d'intégrations afin d'obtenir une expression stable du transgène. Cependant, multiplier les insertions revient à amplifier une mutagenèse insertionnelle, ce qui peut se révéler génotoxique pour la cellule [Liu *et al.*, 2009 ; Mátés *et al.*, 2009]. Une autre alternative est d'utiliser des technologies qui maintiennent le transgène intégré dans un état de chromatine active et exprimée et/ou qui le conduisent dans le noyau au niveau des sites d'expression [Kwaks & Otte, 2006 ; Harraghy *et al.*, 2008]. Les éléments de contrôle de la chromatine ont un impact sur l'organisation de la chromatine avoisinant le transgène, et donc sur son taux d'expression [Recillas-Targa 2004]. Ils empêchent le silençage du transgène intégré dans le génome nucléaire en maintenant l'ouverture de la chromatine, permettant ainsi une expression stable du transgène au cours du temps. Cette stratégie est un bon moyen pour réduire le nombre d'insertions nécessaires

pour obtenir une expression suffisante du transgène. En effet, en théorie, une seule intégration suffit à une expression stable lorsque le transgène n'est pas silencé. Par exemple, l'utilisation de régulateurs épigénétiques tels que *les MAR* a montré que ces éléments représentaient une solution intéressante pour diminuer ces effets indésirables [Zahn-Zabal *et al.*, 2001 ; Girod *et al.*, 2007].

Un des objectifs de l'équipe 2 du GICC étant la mise au point de cellules usines pour la bioproduction, une expression importante et stable dans le temps du transgène est essentielle. Dans ce cadre, j'ai réalisé, comme pour les transposons, le paysage brevets sur des éléments de contrôle de la chromatine, avant de rechercher des éléments qui offrent une liberté d'exploitation.

3.5.2. Résumé de l'article

Les éléments de contrôle de la chromatine sont particulièrement utiles en bioproduction, dans le cadre de la mise au point des cellules usines. En effet, l'optimisation du processus de culture cellulaire pour la production de protéines thérapeutiques recombinantes représente l'un des facteurs les plus déterminants de la faisabilité commerciale. Le succès de cette étape réside dans l'identification des cellules transfectées qui produisent les protéines recombinantes à des taux d'expression industriellement intéressants, et dans l'élaboration de lignées cellulaires stables au cours temps. C'est pourquoi des éléments de contrôle de la chromatine ont été caractérisés et utilisés pour améliorer l'efficacité et la stabilité de la production de protéines recombinantes à partir de cellules de mammifères. L'intégration de tels éléments dans les vecteurs de transfert de gènes est une solution pour éliminer le silençage de son expression et pour rendre l'expression du transgène indépendante du site d'intégration.

D'après la littérature, il existe cinq types d'éléments de contrôle de la chromatine : les *Locus Control Regions* ou LCR, les *Matrix Attachment Regions* ou MAR, les *insulators*, les *Ubiquitously acting Chromatin Opening Elements* ou UCOE et les *STabilizing and Anti-Repressor elements* ou STAR [Kwaks & Otte, 2006 ; Harraghy *et al.*, 2008].

Cet article porte sur l'analyse du paysage brevets des éléments de contrôle de la chromatine. L'étude de l'évolution du nombre de brevets déposés pendant la période 1987-2008, associée à celle du nombre de déposants sur la même période, montre que les technologies de contrôle de la chromatine sont mûres. Nous nous attendions donc à ce que des industriels aient mis sur le marché des éléments qu'ils auraient auparavant protégés par des brevets. Cette hypothèse a été vérifiée en étudiant les revendications des 320 familles de brevets du paysage brevets. Les portefeuilles de brevets de tous les déposants ont été regroupés en fonction du type d'élément et de la nature de l'invention protégée. Un premier groupe rassemble les brevets de preuve de concept et d'ingénierie des éléments de contrôle de la chromatine. Le second regroupe les brevets d'application qui vont exploiter les brevets du premier groupe. L'analyse du positionnement technologique des déposants nous a permis d'observer la présence de trois leaders technologiques majeurs sur le marché : Selexis (MAR), Millipore (UCOE) et Crucell (STAR). Cela nous a aussi permis d'identifier les concurrents pour chaque élément. Nous nous sommes également intéressés à la stratégie de protection géographique. Les résultats montrent que les brevets les plus importants ont été étendus dans de nombreux pays, en Europe, en Amérique du Nord et du Sud, en Chine, au Japon... Dans une dernière partie, nous avons complété le paysage brevets par l'étude de la stratégie des sociétés impliquées sur le marché des éléments de contrôle de la chromatine, et plus particulièrement des trois leaders technologiques.

3.5.3. Conclusion sur l'article

Pour chaque acteur souhaitant entrer sur le marché de la bioproduction, il sera nécessaire de naviguer dans le paysage brevets de chaque technologie exploitée dans ce domaine (lignées cellulaires, vecteurs...). Concernant les éléments de contrôle de la chromatine, les sociétés Selexis, Millipore et Crucell représentent les trois leaders technologiques mondiaux du domaine. Grâce à une stratégie en PI très poussée mise en place dès leur émergence, ces acteurs ont su asseoir une position incontournable sur le marché. De plus, ils continuent à déposer des brevets, soit sur de nouveaux éléments insulateurs, soit sur des techniques d'ingénierie de bioproduction qui ne se limitent pas à leur élément respectif mais incluent les autres éléments protégés par leurs concurrents.

Ainsi, les autres acteurs ont dû se positionner afin de pouvoir eux aussi entrer sur le marché, sans être en position de contrefaçon vis-à-vis de brevets déjà existants. Pour cela, plusieurs options peuvent être envisagées. Premièrement, les sociétés négocient des contrats de licences auprès des leaders technologiques pour la production de protéines recombinantes comme les anticorps thérapeutiques. Dans ce cas, elles ont à leur disposition un élément de contrôle de la chromatine qui est déjà optimisé et n'ont pas à développer leur propre technologie. Deuxièmement, d'autres sociétés ont développé leur propre élément, en profitant par exemple du savoir-faire de chercheurs expérimentés dans le domaine et ayant déjà été inventeurs pour d'autres déposants. Enfin, la troisième et dernière solution nous intéresse plus particulièrement : elle concerne l'utilisation d'éléments qui sont librement exploitables car pas ou peu protégés par des droits de PI. En nous intéressant aux statuts des brevets et des demandes de brevets, j'ai déterminé que l'élément le plus utilisé et le plus testé, l'insulateur cHS4 [[Chung 1993](#) ; [Chung 1997](#) ; [Pikaart 1998](#) ; [Potts 2000](#) ; [Recillas-Targa 2002](#) ; [Mutskov 2002](#) ; [Guglielmi 2003](#)], est protégé uniquement aux États-Unis (US5610053 - *DNA sequence which acts as a chromatin insulator element to protect expressed genes from cis-acting regulatory sequences in mammalian cells*). En effet, la demande internationale de brevet n'a pas abouti (WO9423046) ce qui offre une niche de développement technologique pour tout acteur souhaitant bénéficier d'un élément de contrôle de la chromatine bien étudié. C'est le cas de l'équipe 2 du GICC, qui étudie les effets de l'insulateur cHS4 intégré au vecteur *piggyBac* depuis quelques mois.

3.5.4. Article 6

Landscape of chromatin control element patents: positioning effects in pharmaceutical bioproduction

Fabien Palazzoli, Solenne Bire, Yves Bigot & Florence Bonnin-Rouleux

Characteristics of the current patent landscape on chromatin control elements indicate the intellectual property strategies of technological leaders.

Sales of therapeutic recombinant proteins produced by biopharmaceutical companies have increased from \$49 billion in 2004 to an estimated \$121 billion in 2010. Two of the main priorities for manufacturers attempting to meet the challenge of biopharmaceutical production are shortening the time taken to manufacture new cell lines and selecting them for elevated productivity. The conventional approach for obtaining these cell lines is to use stable expression systems based on chromosomal integration into host cells by transfection. In general, no attempt is made to target the transgenic DNA cassette that expresses the protein of interest at a specific chromosomal site in the transfected cells. Indeed, so far there is no evidence that targeted integration is synonymous with higher protein productivity and long-term stability of expression in mammal cells. Nevertheless, the chromatin structure at the site of insertion does appear to be a critical parameter with regard to the expression of the integrated transgene. Integration into open chromatin (euchromatin), which is actively transcribed, favors high and stable expression, whereas integration into condensed chromatin (heterochromatin), which has low transcription activity, results in little or no gene expression.

Another source of variation is the silencing of the expression of the newly integrated transgene in the host cell chromosomes. Indeed, some or all of the integrated vector copies are recognized as exogenous inserts by the nuclear defense machinery, which leads them to be heterochromatinized, resulting in silencing of the

expression of the transgene¹. One way to render the transgene expression independent of the vector integration site, thus circumventing expression silencing, is to include *cis*-regulatory elements, such as chromatin control elements (CCEs), in the components of a vector designed to genetically modify the factory cells. The scientific literature^{2,3} indicates that CCEs include DNA sequences that have an impact on the chromatin organization, and thus, on the level of expression of the transgenes in their vicinity. Indeed, these DNA *cis*-regulatory elements can act as positional enhancer or silencer blockers, by impairing the enhancer- or silencer-mediated activation of a promoter. Alternatively, they can protect a transgene against chromatin position effects at its genomic integration site, thus maintaining transgene expression over time. CCEs have been identified in various eukaryotic genomes and divided into five categories^{2,3}: locus control regions (LCRs), matrix attachment regions (MARs), insulators (such as the chicken hypersensitive site 4 element, CHS4), ubiquitously acting chromatin opening elements (UCOE) and stabilizing and anti-repressor elements (STARs), according to their properties and specificities. Their properties make them of major interest in bioproduction, and their characterization and engineering have been subject to extensive patenting since their discovery.

Here we analyze the patent landscape (see the methodology in **Supplementary Data 1**) for the main classes of CCEs with the aim of tracking their emergence and current development. We have determined the degree of maturity of these technologies and the geographic coverage of patent protection. We have also identified the key applicants and their technological positioning within the environment of biotech companies and public-sector laboratories working on CCEs.

Changes in the number of applicants

The shifts over the last 30 years in patent ownership between industrial and institutional applicants combined with the arrival of new applicants illustrate the attractiveness of the CCE sector since 1987 (**Fig. 1**). When CCE development began, patent applications were filed by academic bodies acting alone or with an industrial partner. Overall, patenting by academic bodies has remained stable over the last 30 years. Between 1987 and 2001, technological developments expanded, showing that the field of CCEs was attracting increasing interest. The first maturity phase for CCE technologies during the period 1998–2003 (**Supplementary Data 2**) is confirmed by the number of patent filings submitted by industrial applicants. Private industry's interest in CCEs has resulted in its becoming massively dominant in filing patent applications. During these years of more intense patent filings, industry held 70% of patents with CCE claims. In other words, the development of the five chromatin control technologies, and particularly those of insulators and MARs, was essentially undertaken by industry. Finally, the year 2005 marked the return of the public sector, with nearly 50% of applicants being institutional or academic bodies, most of them working on insulation system development.

The changes in the numbers of new and total applicants (**Fig. 1**) also highlight the attractiveness of this field of innovation. The number of new applicants has paralleled the total number of applicants since 1987, the former corresponding to almost 50% of the latter throughout the period. It should also be pointed out that 2001 was a pivotal year, with the highest numbers of both patent families (patents and patent applications sharing the same priority or combination of priorities, about the same invention) (38) and

Fabien Palazzoli, Solenne Bire, Yves Bigot and Florence Bonnin-Rouleux are at Genetic Immunotherapy Chemistry and Cancer (GICC), François Rabelais University, Tours, France. e-mail: yves.bigot@univ-tours.fr

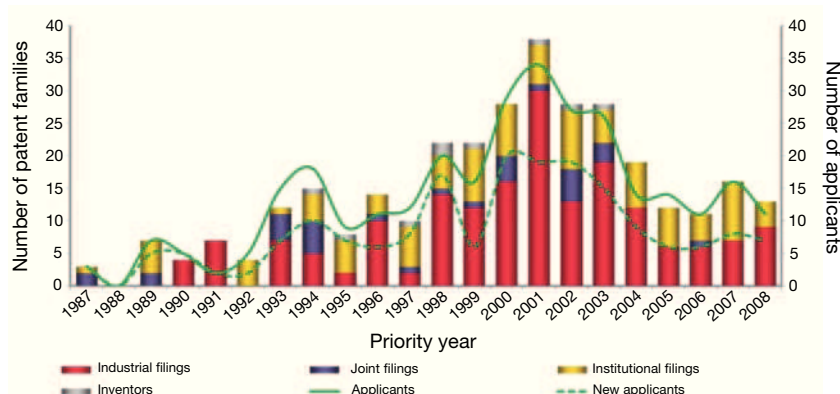


Figure 1 Change in the breakdown of industrial patents versus institutional patents and in the number of applicants. The different colored sections in each bar represent the proportion of filings stemming from academic institutions (yellow), industrial research (red), joint (academic and industrial) filings (blue) and inventors (gray). The filings in the inventors section (gray) correspond to the numerous US patents that are filed by inventors as applicants, under the US patent system. The dashed green line shows the number of new applicants per year defined as applicants who have filed a patent application for the first time, and the continuous green line the total number of applicants.

applicants (34). From 2002 to 2008, the number of new applicants followed the general downward trend until 2005, and has subsequently remained relatively stable. It is noteworthy that the pioneering applicants are still actively patenting, reinforcing their patent portfolio by improving their technologies. It appears necessary today to develop novel elements or alternative solutions—a challenge that is usually undertaken by public-sector fundamental research, sometimes to the detriment of its own best interests. Indeed, government laboratories with limited financial means are often trapped by the need to avoid disclosing to later control the intellectual property of their innovations and recover the benefits of their exploitation, and to validate their works through international journals or conferences to be able to raise funds to continue their projects.

Positioning of the main pioneering applicants

All five families of CCEs have been the subject of R&D projects and patenting activities (Supplementary Data 2). It is interesting to rank each major applicant on the basis of the size of their portfolio for each technology type (Supplementary Data 3). We therefore compared the different technologies covered by the portfolios of the top pioneering applicants to identify those that are in competition within each technology and to identify their likely fields of expertise (Fig. 2).

The analysis of the UCOE-based patent portfolios indicates that Millipore (Billerica, MA, USA) is the main technological player. Millipore's current dominant position results from several successive acquisitions of companies and technology transfers. Therexsys was a

gene therapy company set up in 1992 to exploit the research projects of the Medical Research Council (MRC) London, including LCR technology. By the end of the 1990s, Therexsys had grown and become Cobra Therapeutics, following its fusion with Cobra Biosciences. Cobra Therapeutics went on to develop UCOE technology in collaboration with the King's College London. Cobra was bought a few years later by ML Lab, which wanted to reinforce its gene therapy and cell-targeting projects. To focus and reinforce its activities on inhaled drugs for respiratory diseases, ML Lab became Innovata in 2005, following its acquisition of Quadrant Technologies. Innovata was then acquired in 2007 by the Vectura Group (Chippenhams, UK), which shared similar technological objectives. The orientation taken by Innovata in 2005 led the company to sell its non-core activities, including those concerning its DNA production business. Consequently, the UCOE gene expression technology owned by Innovata was sold to Celliance, a wholly owned affiliate of Serologicals, which is itself owned by Millipore. Finally, compared to the two other major applicants in this technology, Crucell (Leiden, the Netherlands) and GlaxoSmithKline (GSK; London), Millipore can be considered to be the undeniable leader in UCOE technology, with 15 patent families (e.g., WO0005393, WO02081677 and WO2006095156).

Our analysis of STAR technology shows that it has been strongly protected by Crucell (through the acquisition of Chromagenics in 2004), and that the company has concentrated its research and patenting activities. To date, Crucell holds a portfolio of 15 patent families (e.g., WO03004704, WO2004056986 and

WO2006005718). There are only two other relatively minor applicants—Millipore and the Massachusetts Institute of Technology (MIT; Cambridge, MA, USA); thus, Crucell is incontrovertibly the sole leader for STAR elements.

The situations for MARs and insulators are quite similar. The major applicants for insulators and MAR are US government agencies with nine patent families (e.g., WO9423046, WO0102553 and WO2008101216), and the University of North Carolina (Chapel Hill, NC, USA) with eight patent families (e.g., WO9407902, WO9907866 and WO9805757), respectively. This situation contrasts sharply with those of UCOEs and STARS, as no private sector entity has a major patent portfolio based on these technologies. However, minor applicants have emerged from final uses of MARs, as illustrated by Neurotech Pharma (Lincoln, RI, USA) and its MAR derived from the human β -interferon gene (WO200248379), and Selexis (Plan-les-Ouates, Switzerland), with a MAR derived from the chicken lysozyme gene (WO02074969). Similar situations are also encountered with Dow Agrosciences (Indianapolis, IN, USA), and the use of artificial MARs in plants (WO0032800), Synageva Biopharma (Lexington, MA, USA) and the development of a MAR from the chicken ovalbumin gene (WO2004094640) and Novartis (Basel, Switzerland) with its MAR originating from the gene encoding human interferon- β , for which they have developed uses in retroviral vectors for gene therapy purposes (WO9746687). Yet, to date, no major industrial leader has emerged in MAR technology, at least based on the somewhat arbitrary criterion of the number of patent applications filed. It should be noted that within the industry Selexis is considered to occupy a leadership position in MARs even though it has only a few patents, because these have high-value content for bio-production.

Some companies have focused their technological interests on insulators, such as Boehringer Ingelheim (Ingelheim, Germany), which exploits expression-enhancing elements originating from a DNA sequence segment isolated from Chinese hamster ovary cells and located upstream of the coding region of the gene encoding ubiquitin/ribosomal protein S27a (WO2008012142). GTC Biotherapeutics (Framingham, MA, USA) does not use its own genetic element, as this company has claimed the β -globin insulator sequence (so-called cHS4, covered by WO9423046) as a 'preferred insulator' for a method of making a transgenic fusion protein (WO0119846, WO0119842). Similarly, Aulix Biopharma (Berkeley, CA, USA) uses insulators, including cHS4, to screen for compounds having effects on various

tissues and diseases (WO03046133) or to identify cellular targets (WO2005078069).

In the case of LCRs, our results show that the MRC London is the major player, with ten patent families (e.g., WO8901517, WO9010077 and WO9533841; **Fig. 2**). To a lesser extent, the National Institutes of Health (Bethesda, MD, USA) and the Vectura Group are also important applicants with five patent families each. Other companies have filed some patent applications on LCRs, including major pharma such as AstraZeneca (London), GSK and Novartis and biotechs such as Genetix (San Jose, CA, USA), Sangamo Biosciences (Richmond, CA, USA), the Mogam Research Institute (Yongin, Korea) and DNX (Princeton, NJ, USA).

In conclusion, the key players fall into two categories: The first consists of specialists in a single technology, whereas those in the second category tend to diversify their activities by patenting several CCE uses. For example, Millipore and Crucell are two leaders that control the ownership and exploitation of UCOEs and STARs, respectively. However, these two companies have also filed several patent applications covering the engineering of other elements, as illustrated by the four patent families that cover UCOE uses that belong to Crucell, and the two patent families on MAR uses held by Millipore. These two applicants have already developed their CCEs, which they probably now consider to be technically mature. As a result, they have also developed technologies for improving gene expression that can be integrated into a global manufacturing process for bioproduction. For example, in the international patent application WO2006048459, Crucell has disclosed methods for obtaining host cells that express a polypeptide of interest, including an expression cassette comprising at least one chromatin-influencing element selected from the group consisting of a MAR, an insulator sequence, and a UCOE or STAR sequence.

Corporate strategies of CCE-based companies

Reconstructing the history of the main CCE-based companies and the evolution of their corporate strategies is a way to complete the picture provided by the patent landscape. Here, results from our patent studies and analysis of the players involved in the development of CCEs have allowed us to show that several different types of competing companies are involved in this field.

The first type corresponds to spin-offs that have exploited CCEs originally discovered in the laboratories from which they emerged; for example, Chromagenics (now Crucell), which emerged from the University of Amsterdam; Selexis from the University of Lausanne;

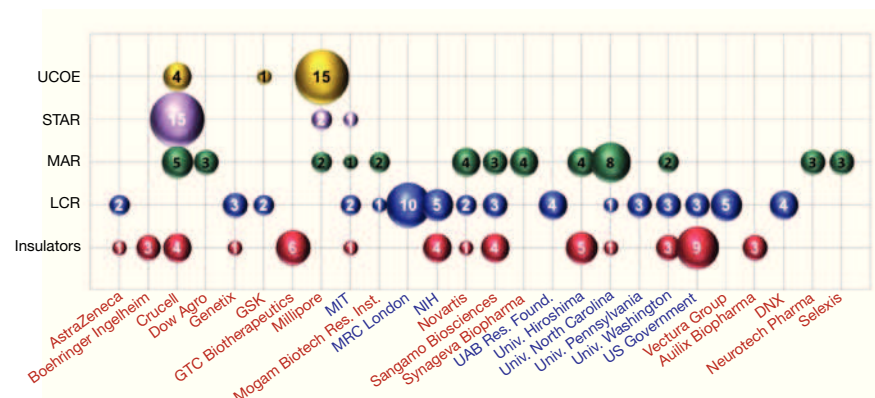


Figure 2 Technological positioning of applicants who have filed at least three patent families covering CCEs, their engineering and their final uses. The ball size represents the technological importance of each applicant, based on the size of its portfolio for each CCE type. The number inside the colored balls corresponds to the number of patent families. Affiliates have been included with their parent companies.

Therexsys from the MRC London and Cobra Therapeutics (ML Lab/Innovata/Millipore) through a fruitful collaboration with King's College London. These technology-based companies have focused on engineering their respective CCEs, which they have protected with robust patent portfolios. To develop their technologies up to a commercial level, companies have had to maintain collaboration with the laboratory from which the CCE emerged, and initiate new interactions with players in the bioproduction field that are developing complementary technologies (e.g., vectors, cell lines). This is why spin-offs that have been successful in the field of CCE technology have often used research license agreements that allow a potential new customer to test the technology for an evaluation period (e.g., ML Laboratories with Maxygen in 2003, Selexis with Pierre Fabre in 2006 and Crucell with Celltrion in 2008). Generally, spin-offs have adopted an out-licensing strategy allowing them to avoid being deprived of market share. In the cases cited, these agreements have proved very useful for these emerging companies because they allowed them to obtain validation of their technology for the manufacture of biopharmaceuticals and, to some extent, to subsequently subcontract the validation to third parties. Once validation was obtained, the research license agreements were replaced by commercial license agreements intended to generate income in the form of commercial turnover and royalties (e.g., ML Laboratories with Maxygen in 2005, Selexis with Ganymed in 2007 and Crucell with Centocor in 2009). These licenses have therefore been used to reinforce a strategic position in a market where several competing CCE systems are present. These validations have also been used by companies to organize part of the communication

system to promote their technology, first by publishing credible communications of their results in scientific journals, and second, by divulging their improvements to the scientific community⁴⁻⁹.

Evolution of CCE-based companies

After a few years of existence, the two emergent companies developing UCOE and STAR technologies were both taken over by a bigger company involved in manufacturing processes for the biopharmaceutical sector, that is, Millipore and Crucell. By these acquisitions, the second type of company expected to reinforce and expand its strategic and economic position in the bioproduction value chain either by absorbing a possible future competitor, or by incorporating a technology that it lacked. Crucell not only proposed the exploitation of the STAR technology, but also that of its PER.C6 cell line within a platform for recombinant protein manufacturing. Moreover, the company estimated that "license fees on the STAR technology will be far more valuable than the licenses Crucell issues on its PER.C6 technology." Such companies propose technology packages once they have been successfully optimized, and concentrate on marketing products or platforms that are protected by a strong patent portfolio. One advantage of their strategies is to reach agreements in biopharmaceutical production that should generate short- and medium-term revenue streams.

Besides their positioning in CCE technologies, companies such as Crucell and Millipore have reinforced their leadership positions by completing their patent portfolio with complementary patent families exploiting other technologies that make it possible to increase protein production in different but complementary ways (WO2007096399 and WO2008085956,

respectively). In this second group, some companies may have used back roads to establish their position. For example, Selexis has so far limited its proposition of services to the use of its MARs under conditions predefined for optimal expression of proteins in mammal cells. In spite of the fact that the company was founded in the pivotal year of 2001 and so far remains a fairly small company (in terms of number of employees), it has not been taken over by a larger company as were the former owners of the STAR and UCOE technologies. This situation makes it difficult to decipher Selexis' corporate strategy. If Selexis thinks that its technology could give an unprecedented advantage to a company involved in bioproduction, they could be waiting for a large deal from big pharma, or with a company that develops entire manufacturing processes for potential customers. Alternatively, Selexis may see its future as a developer or assembler of bioproduction processes and thus be developing its own tailored cell lines, vectors and screening processes. It will be interesting to follow the evolution of the patent portfolio of this company and its competitors in the years to come (with a potential license for WO2010046493 family in the name of the University of Lausanne (which is not included in this patent landscape) that deals with new insulators and not MARs).

Intellectual property strategy of the three technological leaders

In spite of strategy differences between the current leaders in this field, an analysis of their CCE patent portfolios reveals that they have all exploited all the advantages of patent systems since their origin. In particular, they have cleverly made use of the advantages of divisional applications to reinforce their patent portfolio (EP1404872 for Crucell, EP1373535 and EP1587934 for Millipore and EP1675952 for Selexis). A divisional application is a patent application derived from a parent application. Although this cannot be claimed in the form, some parent applications have been written in such a way that unavoidably leads a patent office examiner to conclude, for example, that the unity of the invention is not fulfilled. As a consequence, the applicants have to subdivide their parent application into as many divisional applications as there are unique inventions that they still wish to patent after expert assessment. This also leads to numerous patents, which strategically reinforces the patent portfolio of the applicant. Above all, this allows applicants to extend the period between the date of the patent submission and its publication, thus gaining time to work on the scientific and technological content of their invention and therefore to pinpoint the most pertinent parts in the parent

application that need to be protected by patenting. For example, parent application EP1404872 claims a method for selecting a DNA sequence with a gene transcription-modulating quality in various species, including DNA sequences obtained by the applicant using this method. Following the patent office examination, this parent application was divided into several divisional applications, each covering one DNA sequence obtained by the applicant using their method of selection (e.g., EP1842919 and EP1806407). These intellectual property (IP) strategies are based on the facilities available for the national innovation policies of companies in the Netherlands (Chromagenics is Dutch) or in Switzerland¹⁰ (Selexis is Swiss).

Positioning of the third group of applicants

Currently, there is fierce competition between Crucell, Millipore and Selexis to win over potential bioproduction customers¹¹. However, they may also be faced by a third type of player, which has developed a global process for gene expression purposes, following another strategy that consists of supplying its own CCE in a specific context. For example, Novo Nordisk (Bagsværd, Denmark) has developed an approach to producing Factor VII polypeptides or Factor VII-related polypeptides using its own MAR isolated from hamster cells (WO2004027072). In a similar way, Synageva Biopharma has developed technological packages that use transgenic birds as bioreactors, the production of recombinant protein being linked to a chicken MAR isolated from the ovalbumin gene (WO2004094640). For its part, Avesthagen (Bangalore, India) has developed its own MAR in association with certain expression-enhancing sequences the company had already identified, as illustrated by patent family WO2009150517. In general, these companies correspond to contract manufacturing organizations, offering services involving the manufacture of recombinant proteins on fully optimized cell systems. They also offer upstream and downstream services and make the transfer process much easier, as potential consumers do not need to deal with multiple service providers. At present, they represent emerging-stage growth companies that have accumulated promising technologies to supply a growing worldwide market. They can therefore be viewed as good investment opportunities for venture capital funds, as in the case of the Pender NDI Life Sciences Fund, which invested in Chromos Molecular Systems (Burnaby, BC, Canada; WO9423046) and Fuqua Ventures, which invested in Avigenics, a company acquired by Synageva Biopharma (WO02079447).

Which strategies suit new entrants?

For existing firms that wish to enter the field of biopharmaceutical manufacturing, setting up an R&D and/or exploitation plan first requires the ability to navigate accurately through the patent landscape where competitors are already positioned. An additional difficulty arises from the fact that the technological leaders in the field are continuously developing their technology, so that the patent landscape is forever shifting. Thus, new entrants will have to position themselves to avoid infringing patents held by their technological competitors, who have already developed a strong patent portfolio worldwide (**Supplementary Data 4**). An attempt to penetrate this business by a company owning technologies based on new CCEs also seems to be a high-risk strategy in the light of the data presented above, and the barrier to entry into this market can be considered high.

At least three solutions are exemplified (**Supplementary Data 5**). The first involves the granting of licenses by the technology leaders. The second is to obtain know-how from pioneering inventors. The third solution is to use CCEs that are not patented. So, for new companies such as spin-offs, entering the field of the biopharma manufacturing appears to be a stiff challenge. Indeed, the IP for CCEs as a whole is owned by the pioneering companies, allowing them to control future developments.

The integration of CCEs into current production processes is also well developed. This situation is also indirectly and unintentionally supported by the regulatory agencies that have so far validated only a few production systems for pharmaceutical purposes. From this point of view, only major legislative changes would be able to modify the rules and regulations that would allow the reactivation of innovative solutions in this field. Modifications of the reproduction processes used by the biopharma manufacturers may be initiated from one or several technology breakthroughs: (i) the availability of new and cheaper production systems (animals, plants or fungi) that have been validated by national and international safety agencies; (ii) technological innovations such as the availability of controllable and stable vectors with large cargo capacities (100–200 kbp) that would be able to carry large, tissue-specific elements and inducible regulatory elements; or by major global problems and/or constraints that result from (iii) the discovery of safety issues in the existing production technologies; or (iv) a change in the rules following the economic and political management of world conflicts of interest (e.g., changes in trust, infringement of IP rights).

Conclusion

The market for therapeutic substances that are derived from biological materials or produced by a biotechnological process continues to grow rapidly. Nearly 1,000 products are currently being evaluated in clinical studies, and the estimated sales for 2010 are about \$50 billion. In the absence of major upheaval in the bioproduction landscape, the use of CCEs to improve and optimize factory cell lines appears to have a bright future. Moreover, succeeding in the cell line-engineering business may be an essential first step to achieving some other long-term objective of the companies that have developed CCEs. Indeed, the ability to sustain heterologous gene expression in gene therapy is a new market for players with chromatin control technologies to master, and can be expected to generate additional revenues from this promising new market. New players will

get involved in clinical trials to develop and optimize vectors for gene transfer for therapeutic purposes. This is illustrated by disputes over the ownership and exploitation in gene therapy of the technology based on the LCR from the β -globin gene in the 1990s, opposing the MRC London-Therexsys (EP0332667) and the MIT-Genetix (EP0377676) patents. Competition between CCE technologies has turned out to be economically viable and effective for the production of biopharmaceuticals, and a second leg of the race can be expected to involve producing the most efficient CCEs for gene therapy purposes.

ACKNOWLEDGMENTS

We would like to thank A. Dacheux, D. Cardi, S. Moreau and U. Streichenberger for critical discussion of this paper. This work was supported by the François Rabelais University, Tours, and funded by grants from the European Commission

(SyntheGeneDelivery Project, no. 018716), the C.N.R.S., the French Ministère de l'Éducation Nationale, de la Recherche et de la Technologie, the Association Française contre la Myopathie and the Groupement de Recherche CNRS 2157.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Chen, W.Y. *et al. Proc. Natl. Acad. Sci. USA* **94**, 5798–5803 (1997).
2. Harraghy, N. *et al. Curr. Gene Ther.* **8**, 353–366 (2008).
3. Kwaks, T. & Otte, A. *Trends Biotechnol.* **24**, 137–142 (2006).
4. Benton, T. *et al. Cytotechnology* **38**, 43–46 (2002).
5. Williams, S. *et al. BMC Biotechnol.* **5**, 17 (2005).
6. Kwaks, T.H. *Nat. Biotechnol.* **21**, 553–558 (2003).
7. van Blokland, H.J. *et al. J. Biotechnol.* **128**, 237–245 (2007).
8. Girod, P.A. *et al. Nat. Methods* **4**, 747–753 (2007).
9. Buceta, M. *et al. Gene Ther.* **18**, 7–13 (2010).
10. Ledergerber, W. & Kurt, A. *World Pat. Inf.* **25**, 57–62 (2003).
11. Williams, S. & Cranenburgh, R. *Innov. Pharm. Technol.* **26**, 52–56 (2008).

Supplementary Data 1: Methodology and key words on chromatin control technologies

We have compiled a comprehensive dataset of patent documents by collecting and analyzing CCE patent filings worldwide. To do this, we extracted patent documents from the FamPat patent database (Questel). This lists the patent applications within given patent families published by 93 patent offices. The mining step yielded 320 patent families published between 1980 and March, 2010.

The search method associates standard truncation symbols, Boolean operators, complex search operators and combinations of key words that are used to select the patent documents by probing titles, abstracts and claims. Key words for each of the five element families were defined using the forms of words used in the patents and the scientific literature:

- LCRs: locus control*, dominant activat*, dominant control*, locus activat*, region?, subregion?, sequence?, LCR, LCRs, DCR, DCRs, LAR, LARs, HS, HSS, HS1, HS2, HS3, HS4, HS5, dnase?, dnaase?, supersensit*, hypersensit*, superhypersensit, gene?, genetic, DNA, globin, chromatin, locus, loci, core, site?.
- MARs: scaffold, matri*, scaffold/matri*, attach*, region?, element?, site?, sequence?, DNA, domain?, SAR, SARs, MAR, MARs, S/MAR, S/MARs, A-element?.
- Insulators: insulat*, insulator, insulating, barrier, boundary, regulatory, chromatin, chromosomal, position, effect?, variegation, enhancer-blocking, cis-acting, silencing, expression, element?, gene?, transgene?, vector?, regions?, site?, sequence?, element?, DNA, domain?, SEQ, nucleic acid?, locus, nucleotide?, polynucleotide?, flank*, globin, globulin.
- UCOEs: chromatin*, open*, element?, region?, sequence?, structure?, site?, DNA, domain?, UCOE, not methylated, methylat* free, unmethyl*, CPG island?.
- STARs: stabiliz*, anti, repress*, anti-repress*, region?, element?, site?, sequence?, DNA, domain?, activity, chromatin, STAR.

Supplementary Data 2: Changes in the number of patent filings for the chromatin control technologies

The number of patent filings reflects contemporary technological developments and levels of innovation. **Figure S2** illustrates the number of patent applications involving the five major CCE technologies: LCRs, insulators, MARs, UCOEs and STARs. For the period 1987-2008, our analysis of the claims in the 320 patent families reveal that 89 of them cite the use of LCRs, 127 that of MARs, 28 of UCOEs, 20 of STARs and 133 of insulators. The number of patent applications submitted (black line) has followed a pattern involving 3 distinct phases that began in 1987 with the patenting of dominant activator sequences, also known as LCRs (WO8901517). The 1987-1997 period was a pioneering phase, during which the founding patent families mostly covered CCEs of the LCR type, but also some MARs (US5731178 in 1990) and insulators, such as *chs4* (WO9423046 in 1993). The 1998-2003 period corresponded to a phase of development in which some of these technologies began to mature, with a steady growth in patent filings between 1998 and the pivotal year of 2001, during which 38 patent applications were filed. This period of growth corresponds to a shift in the patent applications filed on CCEs. Relative to the total number of patent families, filings concerning LCR technology tended to decrease or at least remain stable, whereas filings about development activities involving MARs and insulators increased in importance. The years 1998 and 2001 saw the emergence of two new expression-improving elements, the natural (WO0005393 in 1998) and artificial UCOEs (WO0224930 in 2000) respectively, and the STARs (WO03106684 in 2001). After a plateau in 2002 and 2003, 2004 saw a downturn in the filing rate which marked the beginning of a phase (2004-2008) during which the total number of patent applications fell. The patent filings for MARs and insulator technologies were still dominant, compared to UCOEs, STARs and LCRs. In the case of LCRs, this may be explained by the fact that these had been the first elements to be exploited 20 years earlier. As a consequence, their potential had already been explored, and the prospects for innovation were probably limited with regard to the current state of the art biopharmaceutical production. However, in spite of the decreased number of filings, it is noteworthy that some new elements have been protected recently, such as the 5'UTR region of the *ZAM* retrotransposon of *D. melanogaster* (EP1887083 in 2006) or the D4Z4 element (WO2009016206 in 2007). From the beginning of the 2000's, and particularly during the third phase (2004-2008), patent applications have covered the use of one to five elements. For example, they deal with expression cassettes comprising either an LCR, a MAR, an insulator or a UCOE for producing recombinant proteins in animal cells, such as

avian stem cell lineages (e.g. WO2008142124). On the basis of these findings, CCEs seem to have reached a first technically mature stage, with most of them being protected and already widely used in various eukaryotic systems.

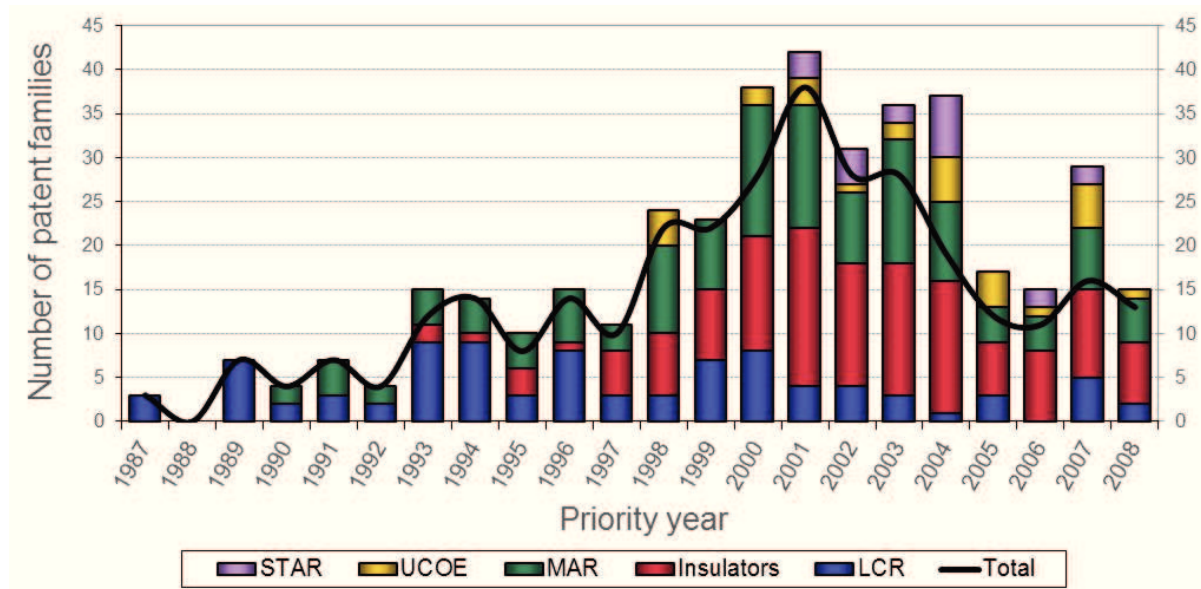


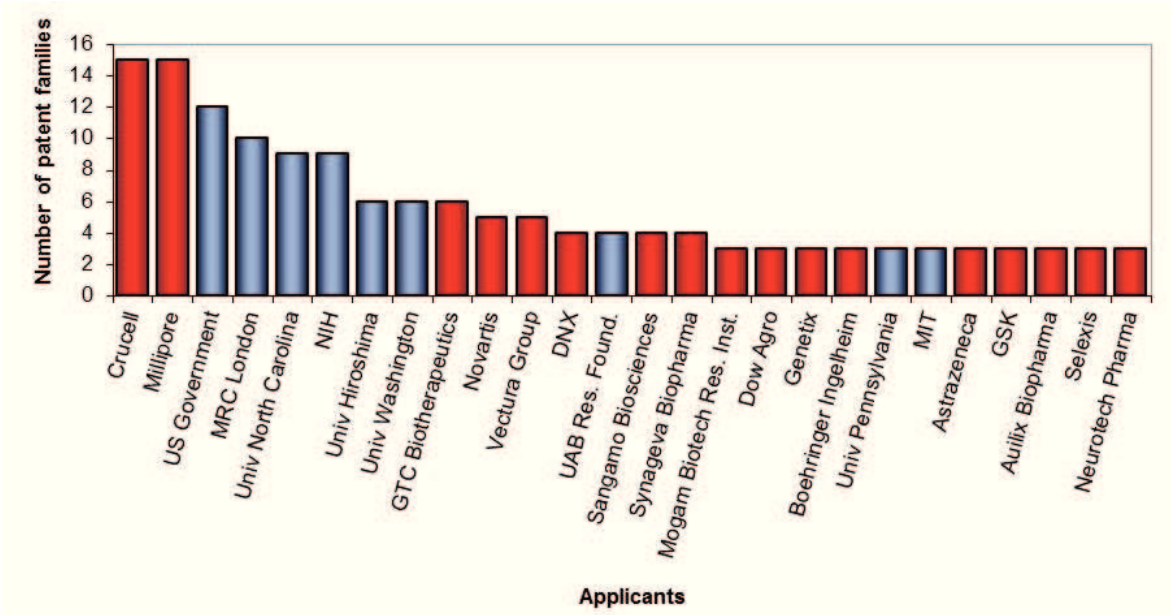
Figure S2 Breakdown of patent families claiming at least one of the following technologies: LCRs, insulators, MARs, UCOEs and STARS. Patent filings covering the five technologies for improving gene expression are shown for the period 1987-2008. The black line represents the 320 patent families. 2001 emerges as a pivotal year for these technologies. LCRs (blue, 89 patent families) were mainly developed from the end of the 1980s to the early 1990s, MARs (green, 127) and insulators (red, 133) from the end of the 1990s to the early 2000s, and UCOEs (yellow, 28) and STARS (purple, 20) during the 2000s. It should be noted that the same patent application can cover several CCEs, and so may appear several times.

Supplementary Data 3: Main applicants submitting pioneering patents and patents of final use

During the period 1987-2008, a total of 180 distinct applicants filed 320 patent families concerning CCEs. An analysis of the claims of each patent family reveals that they can be divided into two groups. The first included a collection of 185 patent families covering the nucleic sequences of the CCEs, their mode of use and/or their engineering. This group obviously comprises all the pioneering patent documents. For example, international patent application WO9727207 describes the features of a plant nuclear MAR that increases the expression of foreign genes in cells, but also includes a way to use it efficiently in DNA constructs and vectors. Similarly, international patent application WO2004054512 concerns engineering an LCR so as to adapt it for use in an optimized retroviral vector for erythroid-specific expression and the treatment of hemoglobinopathy. The second group consists of applicants who filed 135 patent families and used CCEs previously developed by the pioneering applicants as it can be seen in claims, and so the corresponding pioneer key patents have to be licensed before they can be exploited.

The main applicants in this technological field can be identified by comparing the size of their patent portfolios. It should be noted that a company's portfolio includes portfolios belonging to companies it has subsequently acquired (for example the acquisition of Chromagenics by Crucell). Some of the pioneering applicants have developed a strong patent portfolio centered on their CCEs. Twenty-six of the 105 pioneer applicants have filed at least 3 patent families protecting their elements (**Fig. S3a**). These pioneers consist of 17 companies and 9 public-sector laboratories. The top two industrial applicants are Crucell and Millipore, which have a clear leadership over other companies, and each has a portfolio of 15 patent families. US government agencies and the MRC London are joint first place for the number of patents filed by an academic body, with 12 and 10 patent families respectively. Of the applicants in the second group, 99 have owned 135 patent families, and 34 have filed at least 2 patent families (**Fig. S3b**). This group consist of 22 industrial companies, with Shire having the largest portfolio (6 patent families), and 12 institutional bodies, US government agencies being the top public applicant (5 patent families). In both groups, it should be pointed out that some private applicants, such as Novartis or Synageva Biopharma, have developed a strong patent portfolio including both pioneering patents and patents of final use (**Fig. S3a and b**). In conclusion, these data confirm that Crucell and Millipore are the dominant players in the CCE market.

S3a



S3b.

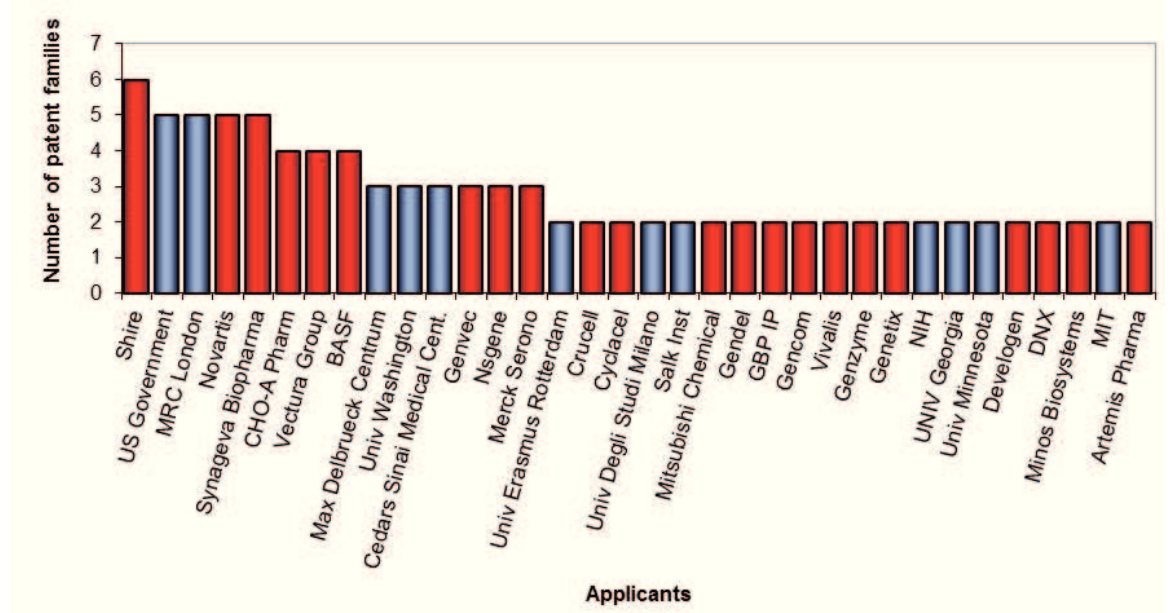
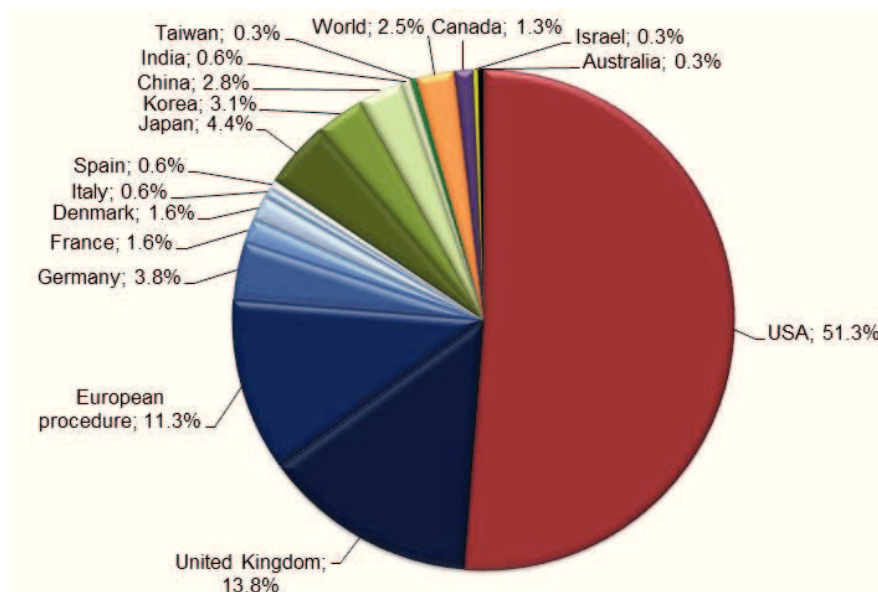


Figure S3 **a** Ranking of applicants who have filed at least 3 patent families covering engineering and CCEs for the period 1987-2008 (industrial applicants are shown in red and institutional applicants in blue). **b** Ranking of applicants that have filed at least 2 patent families covering uses of CCEs for the period 1987-2008. Affiliates have been included with their parent company in both graphs.

Supplementary Data 4: Geographic protection of research and patenting activities and protection strategies for each type of chromatin control element

Applicants file patent applications in the countries they consider to be strategic for protecting their novel CCEs. Priority patents usually depend on the geographic location of the applicant (e.g. NIH in USA, University of Hiroshima in Japan). The breakdown of country origins indicates that the USA is indisputably the country where the most priority patents have been filed in this domain (**Fig. S4a**, 51.3%). This is mostly due to US leadership in bioproduction and gene therapy, supported by the critical role played by US universities and US-based multinational biotech companies. The UK and European procedures are respectively in second and third positions for the number of priority patent filings (13.8% and 11.3%). The countries designated for granting protection provides some insight into the expanding geographic protection of research and patenting activities for these genetic elements (**Fig. S4b**). The extension of priority filings indicates the existence of a potential market, of the applicants' production sites and/or of competitors. Most extensions are filed in Asia (172), Europe (130), Canada (109) and USA (83).

S4a.



S4b.

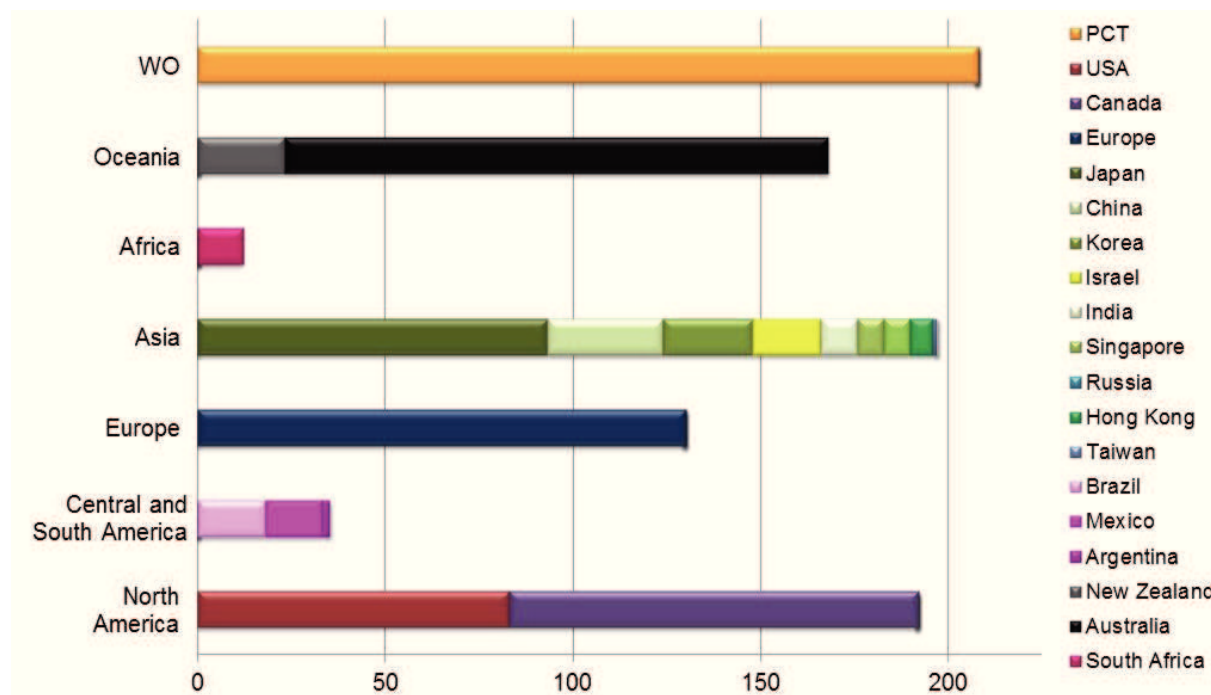


Figure S4 a Breakdown of the priority filings claiming at least one of the following technologies: LCRs, insulators, MARs, UCOEs and STARS. Priority patents correspond to the first patent applications in the patent family, before extension of the protection. Priority countries are identified from the priority number of the initial filing in each patent family. **b** Extensions of priority patents over the entire period. The graph shows the countries where the priority patents have been extended. WO corresponds to PCT (Patent Cooperation Treaty) applications and EP corresponds to European applications. The number of extensions in Australia must be understood in the light of the policy of the Australian Patent Office, which automatically issues an Australian patent number if the PCT mentions Australia, even if no extension has been filed in Australia during the national phase.

To compare the protection strategies used by the main pioneering applicants in the different parts of the world, we looked at the countries to which protection has been extended for each patent. This allows us to identify the countries where the main pioneering applicants have protected their insulation systems. In the current climate of international competition, the number of times a country occurs within patent protections reveals its interest to applicants as a potential market, or the presence of competitors and potential infringers. As a result, this scrutiny also identifies where systems can be freely exploited, without having to negotiate commercial licenses with the owners.

Figure S4c summarizes the protection strategies for each type of CCE, which have been designated by the applicants in the USA, Europe, and Asia and in other areas. Protection strategies also depend on the status of the applicant. Industrial companies (e.g. Crucell, Millipore and Novartis) generally extend their patents more widely than public-sector institutions. US government agencies and universities have filed most of their patent applications in the USA (data not shown). Most pioneering applicants, especially companies, prefer to establish extensive protection of their

elements in numerous countries, including the USA and countries in Europe and Asia. For example, the pioneering patent families of WO03106684 for STARs, WO02074969 for MARs, and WO0005393 for UCOEs have been expanded in numerous countries by Crucell, Selexis and Millipore. Extending protection is a way to establish a monopoly over exploiting technologies. To be able to exploit CCEs in an R&D plan, it is necessary to negotiate an agreement for non-exclusive or exclusive licenses.

S4c.

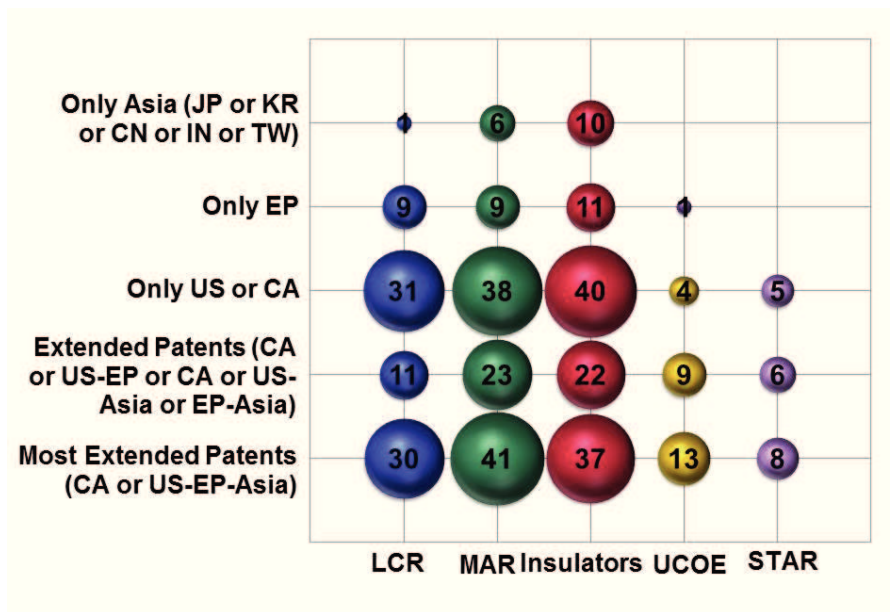


Figure S4 c Protection strategies for each CCE type: LCRs, insulators, MARs, UCOEs and STARs. The number inside the colored balls corresponds to the number of patent families. CA, Canada; CN, China; EP, Europe; IN, India; JP, Japan; KR, Republic of Korea; TW, Taiwan and US, USA.

Supplementary Data 5: Which strategies suit new entrants?

At least three solutions can be envisaged. The first involves the granting of licenses by the technology leaders. This option has been chosen by some companies, such as Vivalis, which have chosen to exploit several CCEs selected among existing LCRs, MARs, insulators or UCOEs (WO2008142124). Other similar examples of licenses granted by Crucell, Millipore and Selexis are those allocated to Bayer Healthcare, Novartis and Neogenix respectively for the manufacture of biopharmaceuticals, and more specifically monoclonal antibodies.

The second solution has been adopted by other applicants that own patents of final use (Roche - WO2007021353; Boehringer Ingelheim - WO2008012142; Abbott - WO2008121324). These companies have an interest in obtaining know-how from pioneering inventors. This can be detected in the patent information in the case of Gerald Hall who discovered MARs, and first filed patents on MARs in the name of the North Carolina State University (WO9407902 and WO9727207), and then in the name of Mycogen plant sciences (owned by Dow Agrosiences; WO9844139 and WO0006757). Another example is that of Philippe Chatellard and Markus Imhof, who are cited as inventors in the WO02074969 international patent application in the name of the University of Lausanne, before appearing in the WO2004081167 and WO2005040384 patent families that were filed by Applied Research Systems (owned by Merck KGa).

The third solution is to use CCEs that are IP-free. This can be envisaged in cases where i) certain CCEs have been the subject of communications, and are therefore not patentable, ii) a patent was never granted or iii) the application was abandoned, iv) the patent is more than 20-years old and has fallen into the public domain. For example, the first patent families on β -globin LCRs had priority years in the late 1980s (e.g. WO8901517, WO9010077 and WO9211380). At present they have fallen into the public domain or will soon do so, thus allowing third parties IP-free exploitation for business purposes. Such a situation is encountered with one of the most widely studied and widely used insulators in the scientific literature, the cHS4 insulator that is disclosed in the patent family of US5610053 (WO9423046 family). Surprisingly, our investigations indicate that patent US5610053 has been granted for the USA, but that no patent extension has been granted for Europe, Canada or Asia. Unlike other promising insulator systems that have led to the emergence of insulation-technology based companies from public research labs (Chromagenics for STARS, Cobra Therapeutics for UCOEs, Selexis for MARs Therexsys for LCRs), no spin-off directly exploiting the cHS4 insulator has emerged. Nevertheless, the public Canadian-based company Chromos Molecular Systems has developed and commercialized an Artificial Chromosome Expression (ACE)

system containing cHS4 insulators to enhance gene expression^{1,2}. Similarly, the use of a cHS4 insulator is cited in a few patents that propose to use it for bioproduction purposes (e.g. WO0119846 in the name of GTC Biotherapeutics).

1. Lindenbaum, M. *et al. Nucleic Acids Res.* **32**(21), e172 (2004).
2. Katona, R.L. *et al. Cell Mol. Life Sci.* **65**(23), 3830-8 (2008).

4. Discussion

Mes travaux de thèse m'ont amené à analyser l'information brevets pour la recherche de niches de développement technologique en bioproduction et en thérapie génique.

Tout d'abord, plusieurs paysages brevets ont été élaborés. Leur analyse critique à un instant t a permis de mettre en évidence des choix technologiques à valider ou non, des partenariats potentiels... pour les projets de recherche du laboratoire. Dans toutes ces études de PI, l'aspect temporel est essentiel : ce qui est vrai à l'instant t ne l'est pas obligatoirement à l'instant $t+1$. Ainsi, après trois années de thèse, les résultats des paysages brevets et leur interprétation ont certainement évolué.

Par ailleurs, en analysant plus en détails les panoramas mondiaux de la PI et en m'intéressant aux stratégies des leaders technologiques positionnés sur les différents marchés étudiés, j'ai identifié des stratégies de PI propres à certains déposants. Des déposants industriels profitent du système pour renforcer leurs avantages concurrentiels sur les autres acteurs du domaine. Des acteurs institutionnels tentent d'échapper aux monopoles imposés par les industriels en contournant les brevets ou en favorisant librement le partage des technologies, ce qui peut profiter finalement aux industriels.

Enfin, dans un dernier point, je souligne l'importance croissante de l'information brevets asiatique. En plus de la barrière de la langue et du volume d'informations disponibles difficilement analysable, il faut également tenir compte des spécificités des systèmes de brevets en question comme le système chinois dans le cadre d'une veille stratégique.

4.1. Importance de la veille pour les paysages brevets

4.1.1. Exemple des transposons

Les résultats pour la période 1986-2007 de l'article 3 - *Transposon tools: worldwide landscape of intellectual property and technological developments* [Palazzoli et al., 2010a] confirme l'évolution permanente des paysages brevets. Une recherche récente des mots clés « *transposase(s) OR transposon(s)* » dans esp@cenet et PATENTSCOPE révèle que certains

déposants déjà identifiés ont renforcé leur portefeuille de brevets avec de nouvelles demandes. C'est le cas des sociétés Transgenrx (WO2010118360, WO2010036979, WO2010036978 et WO2010036976) et Epicentre (WO2010048605). Des déposants institutionnels comme l'Université d'Hawaï (WO2008100424 et WO2008137114) ou le Max-Delbrück-Centrum für Molekulare Medizin (WO2009003671 et WO2009071334) sont également retrouvés. Ces demandes de brevets concernent de nouvelles applications comme la génération de mutations chez des cellules souches spermatogoniales (WO2009071334), mais aussi l'amélioration de systèmes existants telles que des transposases *Sleeping Beauty* hyperactives (WO2009003671) ou de nouveaux systèmes comme le transposon *Passport* (WO2010008564). De nouveaux acteurs revendiquent l'utilisation de transposons et/ou de transposases dans leurs demandes de brevets. C'est par exemple le cas de l'Université d'Aarhus (WO2008106985), de la société Recombinetics (WO2010008562 et WO2010008564), mais aussi de la société Transposagen Biopharmaceuticals (WO2010099296 et WO2010124200) qui a attiré mon attention. Celle-ci prétend être le leader mondial dans la création de modèles uniques génétiquement modifiés (*Knockout*) chez le rat [w56]. Transposagen Biopharmaceuticals précise sur son site internet qu'elle a étendu sa licence d'exploitation auprès de l'Université Notre Dame en Indiana, de l'Université de Floride et du Département américain de l'Agriculture pour couvrir toutes les applications commerciales de la technologie basée sur *piggyBac*. Cette société propose aussi de sous-licencier la technologie de ce transposon pour diverses applications prometteuses telles que la production de cellules iPS (*induced pluripotent stem cells*), la thérapie génique, la thérapie cellulaire et la modification génétique de plantes [w57]. L'analyse de la situation de la PI sur le transposon *piggyBac* illustre bien qu'un paysage brevets doit être analysé au moment où il a été construit, puis exploité rapidement et mis à jour régulièrement à l'aide d'une veille. Dans l'article 3 [Palazzoli *et al.*, 2010a], nous avons proposé *piggyBac* comme niche de développement technologique pour tout acteur souhaitant développer un plan d'exploitation à partir d'un vecteur transposon. Notre conclusion qui date de 2010 est donc aujourd'hui en contradiction avec les informations données par Transposagen Biopharmaceuticals.

Les familles de brevets de *piggyBac* au nom de l'Université Notre Dame, de l'Université de Floride et du Département américain de l'Agriculture, dont les inventeurs sont M. Fraser ou A. Handler, sont les suivantes : US6551825, US7005296 (famille de WO0114537), US7129083, US7105343, US6962810, US2007204356 (famille de WO2007100821) et US2010221824. Cinq familles de brevets ne comportent que des brevets américains et ne protègent l'utilisation de *piggyBac* qu'aux États-Unis. Deux autres familles incluent des demandes internationales de brevets, l'une étant abandonnée (WO0114537, avec deux brevets US accordés), l'autre poursuivant la procédure PCT (WO2007100821). Notre interprétation vis-à-vis du positionnement de Transposagen Biopharmaceuticals est que cette société communique des informations ambiguës dans lesquelles elle n'indique pas si elle détient une licence mondiale exclusive pour *piggyBac*, bien que sa fiche d'information décrive des droits mondiaux exclusifs pour créer des rats modèles *Knockout*. En conséquence, nous pouvons faire l'hypothèse que la société détient une licence exclusive pour les brevets couvrant *piggyBac* uniquement pour les États-Unis, voire pour ceux qui seront issus de la demande internationale WO2007100821. Il sera également intéressant de suivre l'évolution de la procédure internationale pour cette demande. En effet, l'examineur de brevet a cité dans le rapport de recherche deux documents détruisant la condition de nouveauté ou celle d'activité inventive : le brevet US6962810 - *Methods and compositions for transposition using minimal segments of the eukaryotic transformation vector piggybac* et l'article *Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice* [Ding et al., 2005].

Ne pas avoir de protection étendue pour le système *piggyBac*, alors que c'est le cas pour les brevets couvrant les méthodes de création de rats modèles *Knockout*, peut être une faiblesse dans le plan de développement de la société Transposagen Biopharmaceuticals. C'est peut-être pour cette raison que la société a déposé une demande internationale de brevet concernant des transposases *piggyBac* hyperactives (WO2010099296 – *Hyperactive piggyBac transposases*). De plus, Nancy Craig de l'Université Johns Hopkins est une collaboratrice clé de Transposagen Biopharmaceuticals [Transposagen Biopharmaceuticals, 2009]. Cette dernière est inventrice dans deux récentes demandes internationales de brevets concernant des variants du transposon *piggyBac* (WO2010099301) ou des transposons *piggyBac* de Mammifères (WO2010085699), déposées par l'Université Johns

Hopkins. Il sera intéressant de vérifier dans les mois à venir si les demandes de brevets et les brevets accordés qui en découlent seront licenciés à Transposagen Biopharmaceuticals.

Transposagen Biopharmaceuticals affirme détenir les droits d'exploitation sur toutes les applications commerciales de la technologie basée sur *piggyBac*. En suivant cette hypothèse, la société Manoa Biosciences a dû, ou devra, négocier des licences d'exploitation pour le transposon afin de ne pas être en position de contrefaçon. Fondée en 2006, Manoa Biosciences est une *spin-off* de l'Université d'Hawaï créée par Stefan Moisyadi. Celui-ci est co-inventeur dans plusieurs demandes PCT au nom de l'Université d'Hawaï (WO2008137114, WO2008100424, WO2008051620, WO2008027384). Ces dépôts revendiquent l'utilisation d'une transposase ou d'un transposon, préférentiellement *piggyBac* ou de transposons qui lui sont apparentés, pour la création d'animaux transgéniques. En 2007, la société avait annoncé que Malcolm Fraser de l'Université Notre Dame était membre de son conseil consultatif scientifique [Palazzoli *et al.*, 2010a]. Manoa Biosciences n'a pas développé son propre système transposon et revendique donc dans les brevets que la transposase est sélectionnée parmi *piggyBac*, *Sleeping Beauty*, *Mosl*, *Tcl/mariner*, *Tol2*, *Tc3*, *MuA*, and *HimarI*. Dans un journal économique, la société a été citée comme exemple de start-up ayant des difficultés à se développer à cause de l'absence de financements privés [w58]. Son *Chief Executive Officer* (CEO ou Président-Directeur Général) Williams Ettouati reconnaît que la société n'est pas supportée par des investissements hawaïens et a été développée à partir d'une technologie provenant de l'extérieur d'Hawaï. Aucun investisseur local ne souhaitant les financer, Williams Ettouati s'est installé à San Diego afin d'être physiquement proche d'investisseurs potentiels. Le site internet de la société Manoa Biosciences [w59], qui était accessible lors de la conception de l'article 3, a expiré en 2009. Cette situation pourrait suggérer que la société n'existe plus aujourd'hui. Pourtant, le dernier article sur *piggyBac* publié en mai 2010 cite Stefan Moisyadi comme appartenant à Manoa Biosciences [Urschitz *et al.*, 2010]. Cependant, il n'est pas précisé dans les remerciements de l'article si la société a contribué au financement du projet et les auteurs ont déclaré n'avoir aucun conflit d'intérêt.

Cet exemple illustre la fragilité de la durée de vie d'une niche de développement technologique et l'ambiguïté de sa localisation au cours du temps. Il est donc primordial de réaliser une veille pour mettre à jour continuellement les paysages brevets et suivre l'évolution de l'environnement PI.

4.1.2. Exemple des éléments de contrôle de la chromatine

La veille brevets nous a aussi permis d'identifier récemment une nouvelle demande de brevet au nom de l'Université de Lausanne couvrant des éléments de contrôle de la chromatine : WO2010046493 – *Gene transfer vectors comprising at least one isolated DNA molecule having insulator and or boundary properties and methods to identify the same*. Cette nouvelle information devra être incluse dans l'analyse du paysage brevets correspondant [Palazzoli *et al.*, 2011]. Il sera intéressant de voir si les brevets de cette famille seront licenciés ou vendus à Selexis par l'Université de Lausanne, comme cela a été le cas auparavant pour la famille WO02074969 – *Matrix attachment regions and methods for use thereof*. Jusqu'à présent, Selexis était considérée comme une société spécialisée dans les éléments MAR [Williams & Cranenburgh, 2008 ; Jones, 2009]. Or, cette demande de brevet ne couvre pas d'éléments MAR mais des insulateurs. Ce brevet est donc un premier jalon pour se positionner comme leader technologique des éléments insulateurs, dans une niche où aucune société ne s'est véritablement positionnée [Palazzoli *et al.*, 2011]. Si la famille de brevets est licenciée à Selexis, sa position de leader technologique sur le marché des éléments de contrôle de la chromatine sera renforcée vis-à-vis des deux autres concurrents Crucell et Millipore qui détiennent un portefeuille de brevets plus important.

4.2. Stratégies de PI propres à certains déposants

4.2.1. Contournement des brevets considérés comme bloquants

L'analyse de l'information brevets est un moyen pour déterminer la liberté d'exploitation d'une invention. Celle-ci est établie en étudiant les revendications et les statuts légaux des brevets pertinents et en s'intéressant aux pays désignés pour l'obtention d'une protection. Dans l'article 4 *Zinc finger monopoly : quelles sont les règles du jeu ?*, nous avons confirmé le contrôle mondial de la société Sangamo Biosciences sur le développement de la technologie ZF [Palazzoli *et al.*, 2010b]. Devant ce constat, des chercheurs ont tenté de contourner le monopole en recherchant des failles technologiques dans le verrouillage de la technologie. Ils ont essayé de contourner techniquement des brevets sur l'ingénierie de ces

domaines de liaison à l'ADN. Par exemple, l'équipe du Japonais Yokio Sugiura a tenté de modifier le squelette des domaines ZF de Zif268 [Hori *et al.*, 2000 ; Dhanasekaran *et al.*, 2006] pour échapper partiellement à certains brevets spécifiques aux domaines ZF de type Cys₂-His₂ (WO9853059, WO9853060).

Diverses substitutions offrent la possibilité de construire de nouveaux motifs ZF, et d'échapper à certains brevets comme le brevet US6866997 – *Nucleic acid binding proteins* (famille de la demande PCT WO9853059) qui revendique « *A method for preparing a nucleic acid binding protein that binds to a target nucleotide sequence, wherein the binding protein comprises a plurality of zinc fingers of the Cys2-His2 class [...]* ». Les protéines ZF d'un motif différent de Cys₂-His₂ ne sont pas couvertes par cette famille de brevets. Une autre solution pour échapper à certains brevets est de faire varier la longueur et la séquence des acides aminés faisant les liaisons entre les domaines ZF (WO9945132 – *Poly Zinc Finger proteins with improved linkers*). Cependant, ces stratégies lorsqu'elles se sont avérées être réellement fonctionnelles, ne contournent que quelques brevets et ne permettent globalement pas d'échapper aux brevets fondateurs de la technologie ZF.

4.2.2. Stratégies de PI de certains acteurs privés profitant du système des brevets

4.2.2.1. Hyperactivité des déposants

Esteban Burrone, consultant à la Division des PME de l'OMPI, nous rappelle que les brevets sont au cœur du *business* des biotechnologies [w60]. Les brevets, qui sont les seuls actifs des entreprises de ce secteur à haut degré d'innovation, vont être exploités par les déposants, notamment industriels, pour occuper un secteur technologique. Les entreprises profitent pleinement du système et déposent de très nombreuses demandes afin d'occuper une place stratégique sur le marché visé et pour obtenir la délivrance d'au moins un brevet solide. De plus, la contestation juridique d'un brevet occasionne des dépenses très élevées dues aux procédures judiciaires, notamment aux États-Unis [w61]. Pour résumer, l'activité intense de dépôts de brevets par les industriels va leur apporter un réel avantage concurrentiel qu'il sera difficile d'attaquer. Les brevets font donc partie intégrante du plan

d'affaires des entreprises et ils constituent la clé du succès de leurs détenteurs dans un environnement extrêmement compétitif.

4.2.2.2. Demandes divisionnaires

L'analyse du paysage brevets sur les éléments de contrôle de la chromatine et des portefeuilles de brevets des acteurs du domaine m'a permis d'identifier une stratégie de PI particulière : l'utilisation « détournée » des demandes divisionnaires. Celle-ci n'est, à ma connaissance, que rarement mise en exergue par les organismes de recherche publics français, à l'exception de quelques-uns comme Collectis. Une demande divisionnaire est une demande plus réduite que la demande initiale que le déposant veut désormais voir traitée de manière autonome. Les sociétés Crucell, Millipore et Selexis, leaders technologiques des éléments de contrôle de la chromatine [Williams & Cranenburgh, 2008 ; Jones, 2009], ont une stratégie de dépôt et de rédaction des demandes de brevets remarquablement bien pensée, basée sur les demandes divisionnaires³ (EP1404872 pour Crucell, EP1587934 et EP1373535 pour Millipore et EP1675952 pour Selexis). Ces demandes qui ne sont pas affichées comme telles lors du dépôt initial vont exploiter légalement certaines caractéristiques des procédures de dépôt pour optimiser au cours du temps le développement et le contrôle d'une technologie.

Lors de la procédure de dépôt en Europe, l'examineur établit un rapport de recherche dans lequel il présente l'état de la technique dans le domaine de la demande. Il va apprécier l'unité de l'invention. L'examineur peut émettre une objection de non unité d'invention lorsqu'il considère que les revendications définissent plus d'une invention. Pour éviter toute perte des droits, le déposant est alors invité à poursuivre la procédure pour chaque invention identifiée dans la demande initiale. La demande initiale est alors scindée en autant de demandes divisionnaires que d'inventions identifiées, à condition que les demandes divisionnaires déposées ne dépassent pas le contenu de la demande initiale. L'un des avantages repose sur le fait que les demandes divisionnaires sont considérées comme étant déposées à la date de dépôt de la demande initiale, elles bénéficient donc du droit de priorité. Elles suivent leur propre procédure d'examen, même si la demande parente est

³ Je me suis intéressé ici plus particulièrement aux demandes divisionnaires européennes.

abandonnée ou révoquée. La demande divisionnaire ne permet en aucun cas de prolonger la durée de vie d'un brevet, mais protège d'autres aspects de l'invention. Elle peut être utilisée, lorsqu'on dépose une demande de brevet couvrant une invention, pour se réserver la possibilité de revendiquer d'autres aspects de cette invention tant que la demande est en instance. Dans la pratique et jusqu'au 1^{er} avril 2010, une demande divisionnaire pouvait être déposée à tout moment tant que la demande initiale était en examen.

Bien que les demandes divisionnaires n'aient pas été prévues pour ça, il est possible de rédiger intentionnellement une demande de brevet en incluant plusieurs inventions, en sachant pertinemment que celle-ci sera refusée pour défaut de non unité d'invention. L'examineur sera donc amené à faire une objection à ce sujet et le déposant éclatera sa demande initiale en autant de demandes divisionnaires que d'inventions. C'est la stratégie utilisée par Crucell avec la demande EP1404872 qui a été scindée en une multitude de demandes divisionnaires : EP1842919, EP1806407... La demande initiale EP1404872 – *DNA sequences comprising gene transcription regulatory qualities and methods for detecting and using such DNA sequences* décrit une méthode pour sélectionner une séquence d'ADN ayant la capacité de moduler la transcription chez plusieurs espèces, et protégeant lesdites séquences ainsi obtenues. À la suite de l'examen statuant le défaut d'unité d'invention, des demandes divisionnaires ont été déposées pour chaque séquence identifiée par la méthode de sélection et avec un titre identique : *DNA sequences having anti-repressor activity*. Grâce à cette procédure, la société peut protéger uniquement les séquences qui lui auront donné satisfaction après les avoir testées pendant les périodes d'examen et d'échanges avec l'office de brevets, plutôt que de les avoir toutes protégées dès le départ (et d'avoir déposé une demande pour chacune d'entre elles).

Déposer des demandes divisionnaires est également un moyen de gagner du temps pour répondre aux objections des examinateurs des offices de brevets lorsque la demande initiale a été refusée ou abandonnée. Les déposants profitent du délai accordé pour l'examen des demandes divisionnaires pour répondre aux objections contre la brevetabilité de la demande initiale. Cela leur permet de préparer les réponses de manière adéquate. C'est pourquoi les concurrents de déposants de demandes divisionnaires font remarquer que la « période d'incertitude juridique » est prolongée (moment entre le dépôt et la délivrance). Il leur est donc pratiquement impossible de prévoir le moment où la demande

divisionnaire sera éventuellement accordée et ils risquent donc d'être en position de contrefaçon et de payer d'importants dommages-intérêts. De plus, en cas de demandes divisionnaires successives, le risque est d'autant plus important. Finalement, l'utilisation des demandes divisionnaires permet à un déposant d'avoir un positionnement optimal sur un marché en bloquant stratégiquement ses concurrents.

Depuis le 1^{er} avril 2010, l'OEB a publié des directives qui visent à limiter le temps dans lequel on peut déposer une demande divisionnaire. Auparavant, il n'existait pas de restriction au dépôt de demandes divisionnaires. Dans sa décision G1/05, la Grande chambre de recours⁴ de l'OEB a précisé qu'il n'existait pas de base légale pour imposer des restrictions au dépôt de demandes divisionnaires, même à celles de n^{ième} génération. Selon elle toutefois, « il n'est pas satisfaisant que des séries de demandes divisionnaires où chacune contient le même exposé de large portée que la demande d'origine du fait qu'elle conserve au moins une description non modifiée, puissent rester en instance pendant une période allant jusqu'à vingt ans [...] ». Dorénavant, à la suite de ces nouvelles directives, une demande divisionnaire peut maintenant être déposée deux ans au plus tard à compter de l'émission de la première notification de la demande initiale (et de la plus ancienne dans le cas où il y a plus d'une demande divisionnaire) et deux ans au plus tard à compter de la notification mentionnant le défaut d'unité d'invention. Passé ce délai, aucune demande divisionnaire ne pourrait être déposée. L'introduction de cette limitation reflète la volonté de l'OEB de s'attaquer aux demandes divisionnaires dites « abusives ». Les déposants qui exploitaient cette stratégie vont donc devoir tenir compte de ce délai dans leur stratégie de PI [Jongste, 2010].

Dans le système américain, il existe une autre procédure dénommée « *Continuation-in-Part* » ou plus communément CIP, qui est également exploitée par les déposants de brevets dans un but stratégique semblable. Pour résumer, si un complément doit absolument être déposé pour soutenir une demande défailante, il faut avoir recours au

⁴ La Grande Chambre de recours peut être saisie afin d'assurer une application uniforme du droit ou si une question de droit d'importance fondamentale se pose.

dépôt d'une seconde demande dite demande CIP. Celle-ci reprend l'essentiel ou la totalité d'une demande antérieure et y ajoute de nouveaux éléments qui n'y figuraient pas. Les revendications qui s'appuient sur la matière nouvelle ajoutée par la CIP ne peuvent bénéficier que de la date de dépôt de la CIP (si, à nouveau, la condition de suffisance de description est remplie) et ne peuvent pas bénéficier de la date de dépôt de la première demande. Cependant, les revendications qui s'appuient sur une divulgation suffisante figurant déjà dans la première demande bénéficient de la date de dépôt de celle-ci. Lorsqu'on est en présence d'un brevet qui a été délivré sur la base d'une CIP, il faut donc faire le tri dans les revendications pour déterminer les dates respectives dont elles bénéficient (date de dépôt de la première demande ou de la CIP).

Face à ce constat pour le moins nouveau en matière de PI lorsqu'on évolue dans un organisme de recherche public français, je me suis renseigné auprès de deux entreprises de biotechnologies renommées, d'une société de transfert de technologies innovantes et d'un cabinet de conseil en PI sur l'ubiquité d'utilisation de ces stratégies. Une nuance peut être apportée sur les stratégies basées sur les demandes divisionnaires. En effet, il est important de distinguer les effets d'une procédure de protection attentive, c'est-à-dire l'objectif initial d'une demande divisionnaire et une réelle stratégie de PI. Toutes les personnes interrogées, sous réserve de discrétion, ont reconnu le caractère stratégique des demandes divisionnaires, malgré le fait que cette notion ne soit jamais enseignée. Il semblerait donc que l'utilisation de telles stratégies repose directement sur la profondeur des définitions des politiques d'innovation et des stratégies de PI nationales. Une telle profondeur est par exemple très bien affichée par des états comme la Suisse ou le Japon [[Ledergerber & Kurt, 2003](#) ; [Arai, 2006](#)]. Les entreprises de biotechnologies de ces pays sont sensibilisées à ce type de stratégie, ce qui leur permet lorsqu'elles les mettent en œuvre, d'avoir un avantage concurrentiel qui va bien au-delà de celui procuré par les dépôts de brevet. Ce constat ayant été fait lors des derniers mois de ma thèse, il serait intéressant d'approfondir l'utilisation des demandes divisionnaires, notamment en France. En effet, alors que le giron des juristes spécialistes des questions de PI est à notre sens parfaitement au courant de l'intérêt stratégique de ces procédures, la question de savoir pourquoi les demandes divisionnaires ne sont pas ou peu utilisées comme outil de valorisation par les laboratoires de recherche

publics français demeure sans réponse. Dans le cas des organismes de recherche publics, l'implication des chercheurs face à leur invention les amène à être le plus précis possible avec les ingénieurs brevets des cabinets de conseil en PI dans la description et la rédaction des revendications pour déposer des « brevets simples ». Aucune approche stratégique en amont ou en aval n'est réellement évoquée par les tutelles, si ce n'est de se retourner vers le chercheur après le dépôt pour lui demander quel type d'exploitation il envisage. Une hypothèse explicative de ce comportement tient au fait que mettre en place une stratégie de demandes divisionnaires requiert des moyens financiers beaucoup plus importants qui sont peut-être hors de portée des organismes publics de recherche.

4.2.3. Stratégie originale exploitée par des acteurs académiques : *l'open innovation/source*

Tandis que la recherche académique valorise essentiellement ses travaux au travers de publications scientifiques, la protection des résultats de la recherche par des droits de PI incite les industriels à investir dans la R&D. Pour la recherche académique, la question peut cependant être reformulée en se demandant si les droits de PI constituent un obstacle ou au contraire un encouragement pour l'industrie des biotechnologies. Le nombre croissant de brevets et de licences et la complexité de l'exploitation des paysages brevets peuvent en effet être un obstacle à l'innovation. Les problèmes rencontrés portent sur les restrictions imposées par les brevets sur l'exploitation d'une technologie comme un contrôle excessif (exemple de Sangamo Biosciences), la durée trop longue des brevets, l'importance des frais de procédure de dépôt, le prix des licences d'exploitation et le coût financier des batailles juridiques (opposition, défense de ses propres brevets)... Un équilibre est donc nécessaire entre le domaine public et les monopoles accordés par les brevets contrôlés par les industriels, qui ne doivent pas limiter ou empêcher les acteurs académiques d'utiliser les inventions à des fins de recherche fondamentale. À ce niveau, il faut souligner que la différence entre recherche fondamentale et R&D est ténue. En effet, quel que soit le type de recherche, le chercheur se situe toujours à la frontière entre le savoir et l'inconnu. Pour progresser dans l'inconnu, il va donc devoir exploiter des outils dont certains sont protégés par des brevets à leurs limites, donc dans des conditions innovantes. En fonction du regard

posé sur le travail du chercheur, on va donc pouvoir conclure qu'il fait de la recherche fondamentale ou de la R&D dans des conditions potentiellement illicites (contrefaçon). Face à ce constat, des personnes se sont organisées pour proposer, à leur sens, plusieurs stratégies pour gérer et réglementer les droits de PI en vue d'obtenir des résultats optimaux tant pour les innovateurs que pour la société dans son ensemble. L'objectif est de partager les technologies pour relever un défi commun [w62]. L'une de leurs finalités est de supplanter les intérêts particuliers ou nationaux pour répondre aux besoins des pays en développement (maladies, famine, pauvreté, changements climatiques...).

Il n'existe pas une façon unique de mettre en œuvre cette notion mais les propositions sont souvent proches. L'un des modèles de partage des technologies repose sur la notion d'*open source*. La notion est définie par l'OMPI comme un modèle de partage de technologies définissant des « champs d'innovation comprenant une plate-forme technologique que les tiers peuvent librement utiliser et adapter, et sur la base de laquelle des innovations peuvent être à leur tour partagées » [w63]. Autrement dit, il s'agit de plateformes de connaissances, alimentées par des acteurs, qui font avancer toute la communauté scientifique. La notion d'*open biotechnology* est également apparue : des chercheurs proposent des licences ouvertes permettant l'utilisation immédiate de techniques brevetées. Le but est de « créer un réservoir commun de techniques de pointe fondamentales que les preneurs de licences pourront utiliser gratuitement à condition que les améliorations qu'ils apporteront auxdites techniques soient aussi partagées » [w63]. Sont ainsi visées des techniques indispensables à la transformation génétique et des instruments de recherche tels que les marqueurs génétiques. Le généticien Richard Jefferson est l'un des leaders précurseurs de ce mouvement [Herrera, 2005]. Pour cela, il a créé une association à but non lucratif promouvant la R&D dans le domaine des biotechnologies agricoles dans l'intérêt des pays en développement : CAMBIA (*Center for Application of Molecular Biology to International Agriculture*) [w64]. CAMBIA a lancé une initiative d'*open source* pour l'innovation en biotechnologies dénommée BIOS (*Biological Innovation for Open Society*) [w65] qui s'inspire des logiciels *open source* [Isaac & Park, 2009]. Les participants à ce projet ont convenu de partager l'accès à la technique fondamentale tout en utilisant celle-ci pour apporter des améliorations et créer des applications dérivées. Cela ne les empêche pas de détenir des brevets mais ceux-ci ne sont pas directement utilisés pour réaliser des bénéfices.

Une contrepartie financière peut être requise pour récupérer les sommes investies et subventionner un accès équitable (recherche non lucrative et recherche visant à répondre aux besoins des pays en développement).

Dans le cadre de l'initiative BIOS, CAMBIA a créé une base dénommée *Patent Lens* [Herrera, 2005 ; Editorial, 2006 ; w66] qui propose des ressources globales *open source* pour améliorer la transparence du système des brevets. Il s'agit d'une base de données de brevets indépendante et en accès libre qui rassemble des données de brevets issus de plusieurs offices nationaux. Le site *Patent Lens* a ajouté dans sa base de données le texte intégral des brevets australiens accordés, à l'aide de la reconnaissance de caractère optique. Cent quinze mille brevets peuvent ainsi être sondés dans cette base de données en libre accès [List, 2008]. Des paysages brevets sur des domaines précis sont mis gratuitement à la disposition de la communauté scientifique : adjuvants, gènes de résistance aux antibiotiques et utilisations pour la transformation génétiques de plantes, promoteurs utilisés pour la régulation de l'expression de gènes... [w67]. Par exemple, l'analyse du paysage brevets sur la technologie développée à partir d'*Agrobacterium tumefaciens* a poussé les auteurs à rechercher des alternatives aux méthodes usuelles brevetées massivement par les acteurs majeurs du domaine comme Monsanto [Herrera, 2005]. Richard Jefferson précise que la complexité de la PI représente un obstacle pour les acteurs publics et privés qui souhaitent utiliser cette technologie pour améliorer l'industrie agricole. Par conséquent, il présente plusieurs espèces de bactéries en dehors du genre *Agrobacterium* qui rentrent dans le projet *open source* de CAMBIA et qui peuvent être modifiées pour réaliser des transferts de gènes chez les plantes [Broothaerts et al., 2005]. Les souches *Sinorhizobium meliloti*, *Rhizobium sp.* et *Mesorhizobium loti* représentent ainsi de bonnes alternatives au transfert de gène par *Agrobacterium tumefaciens*.

Une autre initiative peut être mentionnée : PIPRA (*Public Intellectual Property Resource for Agriculture*). Il s'agit d'une coalition regroupant plus de cinquante universités et instituts de recherche qui militent pour de meilleurs échanges sur les droits de PI en biotechnologies végétales afin de faciliter la R&D d'applications délaissées, c'est-à-dire dans les cas où les brevets affectent l'innovation en agriculture et le bien-être social.

4.2.4. Déviance et manipulation de la notion d'*open innovation/source*

Dans l'article 4 *Zinc finger monopoly : quelles sont les règles du jeu ?* [Palazzoli *et al.*, 2010b], j'ai évoqué le Zinc Finger Consortium, dont le but est de promouvoir l'ingénierie et la R&D de la technologie ZF [w68]. Pour cela, contrairement à la stratégie de monopole de la société Sangamo Biosciences, les membres se sont engagés à développer une plate-forme d'ingénierie de la technologie ZF qui soit à la libre disposition de la communauté scientifique académique. Ils mettent ainsi à disposition le logiciel ZiFit de design de motifs ZF [w69], la base de données ZiFDB de motifs ZF [w70], ainsi que de nombreuses publications. Les membres du consortium proposent gratuitement une méthode d'ingénierie qu'ils présentent comme une *open source* dénommée OPEN (*Oligomerized Pool Engineering*) offrant aux chercheurs la possibilité de faire progresser la technologie des ZFD [Maeder *et al.*, 2008]. De même, les plasmides utilisés par le consortium sont mis à disposition par Addgene, une organisation à but non lucratif qui facilite la diffusion de plasmides dans la communauté scientifique. Les méthodologies, informations et outils partagés par le consortium sont *open source*. Cependant, pour toute autre utilisation que la recherche, tout tiers souhaitant réaliser, utiliser, distribuer ou vendre la technologie ZF a l'obligation d'obtenir le consentement préalable et les licences appropriées auprès de Sangamo Biosciences. Il n'y a donc pas d'*open access* aux droits de PI et à l'exploitation de la technologie. C'est pourquoi une page du site d'Addgene stipule que certaines utilisations de plasmides du consortium portant sur des modules ZF de Sangamo Biosciences, nécessitent une licence d'exploitation (WO0153480, EP1364020...) [w71]. L'intérêt ici n'est pas de contourner les brevets de Sangamo Biosciences mais d'avoir un accès à la plate-forme et à son savoir-faire [Chandrasekharan *et al.*, 2009 ; OEB, 2010b]. Finalement, l'impact de ce consortium est limité et Sangamo Biosciences reconnaît même que cette initiative lui est favorable. L'existence du Zinc Finger Consortium et la distribution de modules ZF à des prix abordables permettent ainsi aux chercheurs de valider les brevets. Les nombreuses diffusions de connaissances scientifiques (articles, communications orales, sites internet) sont aussi un moyen efficace pour Sangamo Biosciences de faire développer la R&D de sa technologie à moindre frais et d'avoir de la publicité gratuite.

Dans un autre exemple de modèle de partage de technologie, une équipe de chercheurs allemands a développé en collaboration avec des chercheurs indiens une

méthode innovante pour la synthèse d'insuline à faible coût pour le traitement du diabète [w72]. Elle repose sur l'utilisation de la levure *Pichia pastoris* qui synthétise dans son milieu de culture le précurseur de l'insuline. Avec cette nouvelle méthode, le rendement de production d'insuline est deux fois supérieur à celui de la méthode usuelle brevetée basée sur cette levure. Fin 2009, il y avait plus de 220 millions de diabétiques dans le monde [w73], dont environ 50 millions en Inde. Le choix de publier plutôt que de breveter est donc un premier effort pour faciliter l'accès à l'insuline aux populations les plus défavorisées. Ils ont fait le choix de ne pas déposer de demande de brevet pour protéger la méthode mais au contraire de la publier pour qu'elle soit librement accessible et exploitable par tous [Gurramkonda et al., 2010]. Ces équipes de chercheurs avaient déjà eu une telle démarche en publiant en 2009 une méthode de production du vaccin contre l'hépatite B, afin que les populations les plus défavorisées puissent elles aussi avoir un accès à ces médicaments [Gurramkonda et al., 2009]. Cependant, ce type d'initiative demeure très idéaliste et peut s'avérer être très limité en terme de liberté d'exploitation face à la complexité des paysages brevets et de l'enchevêtrement des droits de PI. Par exemple, il existe des brevets dans la procédure globale de production de l'insuline, tel que le brevet US4916212 - *DNA-sequence encoding biosynthetic insulin precursors and process for preparing the insulin precursors and human insulin*. La nouvelle méthode exposée dans l'article concerne le précurseur de l'insuline. Il s'agit donc d'un produit intermédiaire qui nécessite d'être converti par des réactions enzymatiques pour obtenir le produit final, l'insuline humaine, *via* une méthode qui peut être protégée par des droits de PI (US4489159 - *Process for preparing esters of human insulin*).

L'open source représente une réelle volonté des acteurs académiques de supplanter le monopole des acteurs industriels en mettant savoir-faire et technologies à la disposition de la communauté scientifique. Cependant, la complexité du système des brevets associée à celle des paysages brevets entraîne une certaine déviance au profit de ces mêmes acteurs industriels.

4.3. Problème d'accès à l'information asiatique

4.3.1. Accès à l'information brevets chinoise

L'élaboration des paysages brevets et la veille technologique qui lui est étroitement liée sont d'une importance cruciale pour se positionner au sein d'une niche de développement technologique. Cependant, la recherche exhaustive d'informations, qu'elles proviennent des documents brevets ou des publications scientifiques, peut être entravée par la barrière de la langue notamment asiatique. Par exemple, le terme transposase se traduit en chinois par « 转座酶 » [w74], se lit « Zhuǎn zuò méi » d'après Google Translate]. Nous pouvons ensuite vérifier par traduction inverse que « 转座酶 » correspond effectivement à « enzyme transposase » dans Google Translate [w75]. À partir de ce terme, nous avons comparé les résultats entre des bases de données en anglais et en chinois. Ainsi, la recherche de « 转座酶 » aboutit à ≈184000 résultats en anglais [w76], ≈103000 résultats avec le site chinois de Google (redirigé depuis mars 2010 vers le site Google de Honk Kong [w77]) et ≈320000 (Annexe 7, Figure 29) avec le moteur de recherche chinois Baidu [w78]. Ce dernier est le site internet le plus visité en Chine, loin devant Google. Concernant les articles scientifiques, PubMed propose 3822 publications tandis que le portail d'informations CNKI - *China National Knowledge Information* [w79 ; Guénec & Dou, 2008] propose 131 articles en anglais et 2578 en chinois (Annexe 7, Figure 30).

Dans l'article 4 *Zinc finger monopoly : quelles sont les règles du jeu ?* [Palazzoli et al., 2010b], nous avons exclu du paysage brevets 287 familles de brevets qui présentent la particularité de désigner comme inventeurs Y. Mao et Y. Xie. Elles portent sur des protéines naturelles contenant des domaines ZF et nous en avons déduit qu'elles s'inscrivent dans un contexte d'exploitation totalement différent de celui du paysage brevets de l'article. Par exemple, les mots clés « *zinc finger* » sont mentionnés dans le titre de la demande CN1406958 - *Polypeptide-human zinc finger protein-61.27 and polynucleotide for encoding it*. En étudiant le brevet avec esp@cenet, nous pouvons remarquer que les onglets « Description », « Revendications » et « Mosaïque » sont grisés et ne sont pas accessibles. En revanche, l'onglet « Document original » offre la possibilité de naviguer dans la demande de brevet rédigée en chinois. De même, dans les autres paysages brevets, j'ai été régulièrement confronté à des documents brevets en langue asiatique, comme le coréen (KR20020047006,

paysage brevets sur les éléments de contrôle de la chromatine) ou le japonais (JP2006238804, paysage brevets sur les transposons).

Pour résumer, nous sommes retrouvés face à des demandes de brevet qui n'ont été déposées que dans un pays asiatique et donc rédigées dans la langue correspondante. Ces documents ne sont disponibles que dans la langue de la demande brevet initiale soit parce qu'ils n'ont pas été étendus dans d'autres pays, soit parce que la demande PCT pour étendre la protection dans d'autres pays n'est pas arrivée à son terme. Par exemple, la demande internationale WO2005012519 – *Transposon transferase and use thereof* n'a pas été traduite et n'est donc pas disponible en anglais dans esp@cenet (comme l'autre membre de la famille JP2005046093). Dans le cas de l'élaboration des paysages brevets, je me suis basé sur les titres et les résumés des documents brevets qui n'étaient pas rédigés en anglais pour déterminer si je devais les exclure ou les inclure dans l'étude. Cependant, pour apprécier le caractère de nouveauté d'une invention, il serait indispensable de traduire les documents brevets qui semblent casser cette condition de brevetabilité après l'analyse des titres et résumés.

Ces exemples nous montrent que le potentiel de la R&D chinoise est plus vaste que ne le laisse supposer l'information scientifique disponible en anglais. L'information asiatique représente aujourd'hui une source de données scientifiques et technologiques qu'il est de plus en plus difficile de ne pas prendre en compte. L'OMPI estime que les dépôts de brevets ont doublé entre 1995 et 2008, augmentant de près de 1,05 millions de dépôts par an à 1,9 millions [Nombre total de demandes de brevet ; w80]. Mais les évolutions les plus marquantes sont observées auprès des offices de PI des pays émergents en forte croissance économique comme la Chine ou l'Inde [OEB, 2008a ; Demandes de brevet par office des brevets ; w80]. Depuis des dizaines d'années, les premiers offices de brevets sont les offices américains (USPTO, plus de 450 000 dépôts en 2008) et Japonais (JPO, plus de 391 000 dépôts en 2008). Cependant, la Chine est en passe de devenir une superpuissance en termes de nombre de dépôts de brevets [w81]. En effet, la Chine sera le premier pays déposant en nombre de brevets au plus tard à la fin de l'année 2011.

La particularité de ce dynamisme peut être aussi illustrée par la nationalité des déposants [Demandes de brevet par office des brevets ; w80]. En 2003, il y avait à peu près autant de dépôts faits en Chine par des résidents chinois (~57000) que par des non-résidents (~49000). Depuis 2003, le nombre de déposants chinois augmente tous les ans et atteint en 2008 ~195000 dépôts contre ~95000 par des non-résidents, soit deux fois plus. En comparaison, il y a en 2008 aux États-Unis ~231000 dépôts fait par des résidents américains, soit quasiment autant que les ~225000 dépôts des non-résidents. Ainsi, la société chinoise qui s'est développée en copiant les technologies et les produits des entreprises occidentales, est dorénavant capable d'innover fortement. L'augmentation du nombre de dépôts par les résidents chinois peut être expliquée par la politique industrielle menée par le gouvernement pour promouvoir l'innovation nationale. Cela est d'autant plus remarquable que ce pays ne possédait pas de loi relative à la PI il y a une vingtaine d'années. Depuis, la Chine a modifié à plusieurs reprises ses textes de loi sur les brevets afin de les mettre en conformité avec les ADPIC⁵ et les normes internationales. Par exemple, le 1^{er} février 2010, le nouveau règlement d'application de la loi chinoise sur les brevets est entré en vigueur [OEB, 2008b ; INPI, 2010]. Dans ce texte, deux types de mesures se distinguent. Les premières ont pour objectif de se conformer aux pratiques en vigueur dans les autres offices. Par exemple, l'une des évolutions les plus significatives concerne le passage au critère de la nouveauté absolue, qui inclut désormais dans l'état de la technique toute forme de divulgation en Chine mais également à l'étranger. Auparavant, seule l'utilisation ou la divulgation publique en Chine cassait le caractère de nouveauté d'une invention. D'autres mesures de ce texte répondent aussi à des préoccupations purement nationales, comme un examen de confidentialité qui doit être demandé à l'Office des Brevets Chinois (SIPO – *State Intellectual Property Office of the People's Republic of China*) avant toute demande de brevet à l'étranger portant sur une invention faite en Chine. Le non-respect de cette mesure peut aboutir par la perte du droit au brevet sur le territoire chinois. Par conséquent, les acteurs qui souhaiteront se positionner sur le marché chinois devront en tenir compte dans leurs stratégies commerciale et économique. Une autre mesure concerne la protection des ressources génétiques : sont exclues de la brevetabilité les inventions dont la mise en œuvre repose sur des ressources génétiques dont l'acquisition et l'utilisation sont interdites en

⁵ Aspects des Droits de Propriété Intellectuelle qui touchent au Commerce, qui introduisent des règles relatives à la Propriété Intellectuelle dans le système commercial multilatéral.

Chine. La Chine a également adopté une disposition qui autorise la fabrication, l'utilisation, ou l'importation de médicaments ou d'appareils médicaux brevetés afin d'obtenir les informations nécessaires à l'obtention des autorisations administratives. Cette exception, déjà prévue dans certains pays, facilite la commercialisation des médicaments génériques en permettant de débiter la procédure d'autorisation de mise sur le marché et les essais cliniques sans attendre l'expiration du brevet.

Le système des brevets chinois comporte quelques caractéristiques particulières. La moitié des brevets sont en réalité des certificats d'utilité, des titres de PI qui protègent une invention pour une durée moins importante que les brevets et à moindres frais. De plus, ils ne nécessitent pas d'examen de validité pour être délivrés [w81]. Par ailleurs, selon le directeur du Centre de recherche de la PI de l'Université de Jiaotong à Shanghai, la plupart des brevets déposés en Chine ne font que changer l'apparence ou proposer de nouveaux modèles, ce qui ne requiert pas une grande technique innovante [w82]. Des mesures ont été mises en place pour inciter les inventeurs chinois à déposer des brevets : un professeur d'université obtient plus rapidement sa titularisation, l'impôt sur les sociétés peut être réduit de 25% à 15% pour les entreprises déposantes, les examinateurs du SIPO sont mieux payés s'ils approuvent davantage de brevets [w81]. C'est pourquoi, selon un conseil en brevets à Shanghai, trouver un brevet solide chinois est plus difficile que d'en obtenir un. Cependant, d'après une étude publiée en 2009 [w83], même si la qualité des brevets délivrés est plus faible que celles des offices européens, américains ou coréens, elle est en constante amélioration (18% des offices interrogés et avocats estiment que la qualité des brevets est très bonne ou excellente contre 70% pour l'OEB).

4.3.2. Solution envisagée et actions proposées

Compte tenu de l'explosion du nombre de documents brevets et d'articles publiés en Chine, l'information scientifique contenue dans ces documents est devenue une source d'information technologique incontournable dans le monde entier [Guénec, 2010]. La détermination de l'état de l'art, l'appréciation de la nouveauté et de l'activité inventive

doivent tenir compte de ces informations chinoises et, par extrapolation, de toutes les informations asiatiques (coréens [OEB, 2008c ; OEB, 2009a ; OEB, 2009b], indiens [OEB, 2010c]...). Les traductions automatiques sont donc une réalité et une nécessité pour la recherche de brevets, y compris pour les examinateurs des offices de brevets [OEB, 2010d]. Par exemple, l'OEB travaille sur un service de traduction automatisée rendant le contenu technique d'un brevet suffisamment compréhensible par une personne techniquement qualifiée (appelée « homme de métier »). Dans ce but, l'OEB a créé des dictionnaires bilingues associant l'anglais à d'autres langues européennes (allemand, français, italien, espagnol, suédois, portugais, grec, roumain et néerlandais) et qui se basent sur les symboles de la CIB [OEB, 2008d]. Cette méthode fonctionne également pour des langues qui ne sont pas européennes mais certaines posent problème, comme le chinois dont une ligne de texte peut comporter jusqu'à trois polices différentes alternant d'un caractère à l'autre [Kister & Dou, 2010]. Les traductions chinois-anglais comprennent de nombreuses erreurs relatives au style et à l'ordre des mots. L'OEB possède sur son site une page internet spécialement dédiée à l'information brevets des pays d'Asie [w84]. On y trouve des prestations de traduction du japonais, chinois ou coréen vers l'anglais et un outil payant de traduction automatique coréen-anglais : K2E-PAT (*Korean to English Automatic Machine Translation*) [OEB, 2010e ; w85]. Cet outil fournit des traductions immédiates en anglais de documents de brevets et de publications PCT coréennes. De son côté, le SIPO a mis des produits pilotes à la disposition des utilisateurs souhaitant les tester [w86 ; w87 ; w88]. D'autres outils sont disponibles comme l'outil expérimental de recherches multilingues CLIR de PATENTSCOPE offrant des traductions et synonymes de QPAT ou Google Translate. Enfin, un dernier outil peut être aussi cité : PLuTO - *Patent Language Translations Online* [OEB, 2010f ; w89]. Il concerne une approche intégrée de traduction en ligne qui s'adapte aux domaines techniques et aux langues. À la fin du développement du projet prévu en 2013, cet outil ne se limitera plus uniquement à la traduction mais il pourra aussi être utilisé comme un moteur de recherche pour l'état de l'art.

Les traductions automatiques de textes techniques en langues asiatiques n'en sont qu'à leurs débuts mais les efforts se poursuivent pour rendre les traductions plus compréhensibles. Néanmoins, les traductions automatiques ne parviennent que rarement à égaler la qualité du travail réalisé par un traducteur professionnel « humain ». Dans certains

cas comme l'analyse de revendications, une traduction manuelle exacte du contenu technique est alors indispensable [OEB, 2010e].

Concernant les paysages brevets, il est important de savoir que les chercheurs chinois qui émigrent en occident peuvent continuer à produire des publications sous un nom plus occidentalisé (notamment au niveau des prénoms). Ils peuvent donc avoir produit et être enregistrés sous plusieurs identités. De même, nous pouvons citer l'exemple du nom « Zhao Ying », porté par 233 chercheurs chinois (dont les caractères chinois des noms sont différents). L'un d'entre eux est un chercheur expert dans le domaine photovoltaïque. La recherche de ce nom donne 408 articles en 2008, 356 en 2009... mais seulement 51 articles sur la période 2008-2010 font ressortir ce spécialiste en photovoltaïque dans les auteurs [Guénec, 2010]. Une analyse réalisée directement à partir de la langue chinoise permet donc d'identifier sans ambiguïté les bons auteurs et inventeurs. Certaines bases de données comme CNKI comprennent des champs de métadonnées suffisamment organisés et renseignés pour pouvoir faire l'objet d'un traitement automatisé. Par exemple, nous pouvons constater avec l'annexe 7 que l'architecture de la page de résultats est proche de celle de PubMed et comprend la liste des articles. Cela va pouvoir donner lieu à un traitement grâce à des outils de traitement automatique de l'information qui prendront en compte les caractères chinois en se basant sur leur codage informatique.

Outre les barrières de langue qui rendent l'information brevets asiatique difficile d'accès et à interpréter, il faut également tenir compte des spécificités des systèmes de brevets en question comme les systèmes chinois [OEB, 2008b ; INPI, 2010] ou coréen [OEB, 2009a ; OEB, 2009b]. C'est pourquoi l'OEB a installé un bureau avec des experts des systèmes de brevets chinois, japonais et coréens [w84]. Leurs missions reposent sur des conseils pour une utilisation efficace des bases de données gratuites et la recherche de documents brevets dans la langue concernée. De ce fait, il serait intéressant de développer un centre de recherche en information spécialisé sur la Chine [Guénec & Dou, 2008]. Finalement, l'accès à l'information scientifique chinoise est un premier pas pour anticiper et contrer la stratégie chinoise de saturation de dépôts qui lui permet d'acquérir une puissance

mondiale avec une position de leader en nouvelles technologies [w90]. L'importance d'une analyse de l'information en langue chinoise avant traduction et la valeur inestimable de la connaissance des réseaux chinois prennent alors tout leur sens pour aboutir à une information pertinente [Guénec, 2010].

5. Conclusion :

Une cellule de veille stratégique a-t-elle sa place au sein d'un laboratoire public de recherche en biotechnologies ?

Au cours de ma thèse je me suis intéressé à répondre à cette question : une cellule de veille stratégique a-t-elle sa place au sein d'un laboratoire public de recherche en biotechnologies ? J'ai tenté d'y apporter une réponse en montrant l'intérêt de cette démarche au niveau local au GICC mais aussi en démontrant ses limites. Dans un dernier point, je positionnerai l'intérêt d'une telle stratégie au niveau national en soulignant les efforts qu'il reste à accomplir.

Grâce à l'exploitation de l'information brevets que j'ai mise en place durant ma thèse, j'ai contribué à l'obtention de plusieurs contrats et de différents financements de projets du laboratoire. En effet, en démontrant par des études de brevets le caractère innovant des projets proposés par le GICC et en associant cette démarche à une vraie stratégie de valorisation des résultats, les projets soumis présentent alors un raisonnement qui apparaît comme un gage de confiance et de sérieux pour les évaluateurs des projets. En effet, les chercheurs des laboratoires publics sont de plus en plus amenés à répondre à des Appels À Projets (AAP) pour financer leurs recherches avec des projets où la PI est au cœur même du processus de valorisation des résultats. La sélection des projets retenus est effectuée sur des critères de qualité pour l'aspect scientifique auxquels s'ajoute la pertinence économique pour les entreprises. Les laboratoires de recherche publics sont, au même titre que les entreprises, de plus en plus confrontés à l'instabilité de leur environnement : concurrence, financements, politique et réglementation... Un autre exemple est le développement des pôles de compétitivité qui modifient profondément les relations entre les organismes de recherche publics et les entreprises. L'enjeu est de s'appuyer sur des projets collaboratifs innovants pour permettre aux acteurs impliqués, notamment les entreprises, de prendre une position de premier plan dans leurs domaines et de renforcer leur compétitivité. L'objectif est, comme pour les AAP, d'augmenter l'accès à la connaissance en partageant les résultats issus de la recherche et en améliorant le transfert d'innovations entre la recherche publique et les entreprises. Dans un contexte international où la compétition scientifique et économique s'intensifie, les acteurs de l'innovation qui réussiront seront ceux qui utiliseront l'information avec le plus de pertinence. Dans ce cadre, l'IE consiste à donner la bonne information à la bonne personne, au bon moment, pour prendre la bonne décision. Puisque tout organisme doit apprendre à maîtriser les flots informationnels avant de prendre une quelconque décision stratégique, les laboratoires de

recherche publics ont donc tout autant intérêt à miser sur la veille stratégique pour être compétitif et innover. Par conséquent, une cellule dédiée à la veille stratégique et à l'exploitation de l'information brevets devient un élément important pour un laboratoire dans les échanges avec les industriels et les financeurs, parce que celui-ci sera à même de comprendre et de tenir compte de cette source d'information pour ses projets de R&D.

Dans cette dernière partie de ma thèse⁶, mon objectif est d'examiner l'importance capitale que pourrait avoir une cellule de veille stratégique au sein même des laboratoires publics de recherche, notamment dans le domaine des biotechnologies, à travers l'expérience acquise par la direction de l'UMR CNRS 6239 – GICC au cours de la période 2006-2009.

L'exploitation de l'information brevets au cours de ma thèse s'est inscrite dans le cadre de la mise en place d'une cellule de Développement Stratégique au sein même du laboratoire. L'idéal pour mettre en place une telle cellule dans un laboratoire de recherche public serait de se baser sur ce qui existe déjà, de le compléter et de l'optimiser. En effet, généralement, des bases de données spécifiques, des outils et des méthodes sont déjà présents au laboratoire. Les chercheurs et les ingénieurs connaissent bien les sources fiables ainsi que les moyens d'y accéder car ils sont traditionnellement habitués à faire de la recherche d'informations. Bien que cette approche de la veille soit informelle et souvent peu organisée, les chercheurs et les ingénieurs sont les acteurs directs de la veille stratégique puisque ce sont eux qui décident de l'évolution des projets de R&D. L'un des éléments à ne pas négliger dans la mise en place d'une cellule de veille stratégique porte sur le réseau et l'environnement. En effet, les personnels des laboratoires de recherche publics, notamment les chercheurs et ingénieurs, sont en interaction permanente avec une multitude d'acteurs et sont donc impliqués dans de nombreux réseaux formels (organismes de tutelle, organismes de financements...) et informels (partenaires, contacts lors de congrès...). Un des facteurs de succès de la veille stratégique repose précisément sur la capacité à bien gérer les différents réseaux du laboratoire, toujours dans l'objectif d'anticiper les actions et les réactions des acteurs présents dans l'environnement du laboratoire.

⁶ Il s'agit uniquement d'un point de vue non exhaustif de personnes travaillant au sein d'un laboratoire de recherche public en biotechnologies qui ne prétendent pas être des experts de la veille stratégique et de l'IE.

Selon l'expérience que nous avons eue au GICC, la cellule de veille stratégique qui coordonnera les actions de veille devra être pluri/multidisciplinaire. Elle comportera des experts qui connaissent les subtilités des biotechnologies, des outils des sciences de l'information et de la communication, des systèmes de brevets, des données juridiques, économiques, industrielles et commerciales... Seule une telle cellule saura mettre en place les outils adéquats pour collecter, analyser et diffuser les informations stratégiques utiles, offrant à la direction du laboratoire les moyens de prendre une décision rationnelle. Trois principaux groupes d'acteurs apparaissent donc dans ce schéma : les décideurs (besoin d'information pour agir), les veilleurs (spécialistes, ils font remonter l'information et la synthétisent) et les animateurs (intermédiaires entre les spécialistes et les décideurs). Comme ses activités doivent permettre de localiser des niches de développement innovantes, cet outil, lorsqu'il est à la disposition du ou des responsables de l'unité, peut être un excellent support à la définition des priorités et au développement d'une politique scientifique. La mise en place d'une cellule de veille est totalement dépendante de la direction du laboratoire, laquelle définit la stratégie du laboratoire. Autrement dit, la mise en place d'une telle cellule, son efficacité et son succès reposent directement sur l'appui de l'équipe dirigeante, qui exploitera les informations décisionnelles résultant du processus de veille. Par ailleurs, les succès obtenus sont étroitement liés à la façon dont seront intégrées les informations dans le développement des projets de R&D des équipes de recherche du laboratoire.

La présence d'une cellule de veille stratégique au sein d'une unité de recherche est un outil dédié à l'aide à la décision qui participe au processus d'innovation et à la progression de la connaissance. Le meilleur exemple est probablement l'unité *Strategic Intelligence Survey* de l'UMR CNRS 6171 - Systèmes Chimiques Complexes, grâce à son organisation autour de la veille stratégique et de l'IE. Considérée comme une expérience de référence, cette unité est entièrement autofinancée. L'une de ses retombées est le prix du concours national du ministère de l'Enseignement et de la Recherche avec OSEO et l'ANR, lauréat 2009 en catégorie Émergence. Le travail qui a été primé a même conduit à la création d'une Jeune Entreprise Innovante au sein d'un incubateur.

Malheureusement, l'unité *Strategic Intelligence Survey* est l'un des trop rares laboratoires de recherche publics français à suivre une démarche semblable. Cette absence de stratégie et d'organisation de la part des laboratoires de recherche publics en biotechnologies n'est pas étonnante. D'après l'expérience vécue au GICC, nous pouvons affirmer que les moyens et le positionnement d'un tel outil dans l'organisation de la recherche publique posent un problème qui prend son origine au niveau national. En effet, dans un rapport de 2005 sur la place des biotechnologies en France et en Europe, le député Jean-Yves Le Déaut recommande l'urgence de reconnaître en France la dimension stratégique des sciences de la vie et des biotechnologies et la nécessité d'identifier des leviers essentiels pour leur développement. Il constate que la France est en train de décrocher en termes de R&D, en particulier dans l'industrie pharmaceutique. Pour paraphraser l'auteur du rapport, le cheminement vers une « économie de la connaissance » ne s'est pas accompagné d'un progrès dans les connaissances de « l'économie du vivant ». Heureusement, ce retard peut être comblé en élaborant une stratégie cohérente en faveur des sciences de la vie et des biotechnologies. L'un des moyens pour que la France rattrape son retard, par rapport aux acteurs historiques en biotechnologies comme les États-Unis ou les acteurs émergents comme la Chine, est d'avoir une politique ambitieuse d'innovation. Dans ce but, Jean-Yves Le Déaut propose quinze recommandations telles que la gestion des droits de PI issus des travaux de recherche financés sur fonds publics et un meilleur soutien de la création et du développement des sociétés françaises de biotechnologies. Nul doute que rattraper le terrain perdu passera également par la mise en place de cellules de veille stratégique utilisées comme outil de compétitivité favorisant l'innovation.

En 1994, le rapport d'Henri Martre sur l'IE et la stratégie des entreprises jugeait le système français en retrait par rapport à ses concurrents : conception partielle de l'IE, veille technologique encore balbutiante dans l'immense majorité des PME... De nos jours, les entreprises françaises semblent avoir pris conscience de l'importance d'intégrer cette activité de veille dans leur stratégie de développement. Malheureusement, nous ne pouvons pas faire le même constat pour les laboratoires de recherche publics. Il est donc nécessaire aujourd'hui pour la compétitivité des laboratoires de recherche publics français, que ce soit en biotechnologies ou dans d'autres domaines technologiques, de mettre en place des cellules de veille stratégique. Cette démarche permettrait de réaliser des projets de R&D

pertinents en phase avec les besoins stratégiques du pays.⁷ Il reste à faire comprendre que ces ingrédients conduisent à terme à une augmentation du potentiel scientifique, pas simplement dans le domaine initialement visé mais, par le jeu de fertilisations croisées, dans d'autres domaines de collaboration où d'autres objectifs peuvent être atteints. Le résultat n'est pas la simple réponse à une question industrielle ou sociétale spécifique mais une avancée scientifique, base de nouveaux potentiels. Cela devrait permettre d'optimiser les savoir-faire et les compétences créées dans les laboratoires à la fois pour les rendre utiles au secteur productif mais aussi pour les orienter et les faire évoluer en posant de nouveaux problèmes à résoudre sur le plan fondamental. C'est, dans cette spirale vertueuse que devraient s'engager les laboratoires mais aussi l'industrie et les pouvoirs publics pour créer une dynamique nécessaire à la recherche française.

Pour paraphraser le généticien Axel Kahn, je dirais que « les biotechnologies en France ont un passé, elles ont un présent, il reste à leur inventer un avenir ». La mise en place de cellules de veille stratégique est un premier pas pour y parvenir.

⁷ Le paragraphe suivant est tiré du chapitre 10. Intégration de l'IE et de la veille dans un laboratoire de recherche académique scientifique, de J. Kister et H. Dou provenant du livre IE et problèmes décisionnels (2010). [Kister & Dou, 2010]

Bibliographie

Références bibliographiques

- Arai H.** Japan's intellectual property strategy. *World Patent Information*. 2006, 28;323–326
- Aron P & Petit C.** L'info, nerf de la guerre. Dossier Intelligence Économique, *Le Monde Informatique*. 1997, 731
- Bio³ : Biotechnologies – Bioproduction – Biomédicaments.** IMT Éditions, Tours, 2010. 464p
- Broothaerts W, Mitchell HJ, Weir B, Kaines S, Smith LM, Yang W, Mayer JE, Roa-Rodríguez C, Jefferson RA.** Gene transfer to plants by diverse species of bacteria. *Nature*. 2005 Feb 10;433(7026):583-4
- Calos MP.** The *phiC31* integrase system for gene therapy. *Curr Gene Ther*. 2006 Dec;6(6):633-45
- Carnus E, Demattei MV, Casteret S, Carpentier C, Palazzoli F, Bire F, Bressac C, Renault S, Bigot Y.** Characterization of monomeric protein domains binding specifically to a highly conserved 100-bp target within rDNA genes. *En révision*
- Cary LC, Goebel M, Corsaro BG, Wang HG, Rosen E, Fraser MJ.** Transposon mutagenesis of baculoviruses: analysis of *Trichoplusia ni* transposon *IFP2* insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology*. 1989 Sep;172(1):156-69
- Chandrasekharan S, Kumar S, Valley CM, Rai A.** Proprietary science, open science and the role of patent disclosure: the case of zinc-finger proteins. *Nat Biotechnol*. 2009 Feb;27(2):140-4
- Chen WY, Bailey EC, McCune SL, Dong JY, Townes TM.** Reactivation of silenced, virally transduced genes by inhibitors of histone deacetylase. *Proc Natl Acad Sci USA*. 1997 May 27;94(11):5798-803
- Chung JH, Bell AC, Felsenfeld G.** Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci USA*. 1997 Jan 21;94(2):575-80

- Chung JH, Whiteley M, Felsenfeld G.** A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell*. 1993 Aug 13;74(3):505-14
- Ciuffi A, Diamond TL, Hwang Y, Marshallb HM, Bushman FD.** Modulating target site selection during human immunodeficiency virus DNA integration in vitro with an engineered tethering factor. *Hum Gene Ther*. 2006 Sep;17(9):960-7
- Cochet O,** Pierre Fabre. Development and production of recombinant proteins for therapeutics applications. *Université de Tours*. 16 octobre 2009
- Dhanasekaran M, Negi S, Sugiura Y.** Designer zinc finger proteins: tools for creating artificial DNA-binding functional proteins. *Acc Chem Res*. 2006 Jan;39(1):45-52
- Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T.** Efficient transposition of the *piggyBac* (PB) transposon in mammalian cells and mice. *Cell*. 2005 Aug 12;122(3):473-83
- Dos Santos WG & Buck GA.** Simultaneous stable expression of neomycin phosphotransferase and green fluorescence protein genes in *Trypanosoma cruzi*. *J Parasitol*. 2000 Dec;86(6):1281-8
- Dou H JM.** Benchmarking R&D and companies through patent analysis using free databases and special software: a tool to improve innovative thinking. *World Patent Information*. 2004, 26;297–309
- Dou H, Hassanaly P, Quoniam L, La Tela A.** *La veille technologique*. Les cahiers du Lera - Sciences de la Société. 1993, 29;213-216
- Duparcq S.** Les brevets, une valeur en hausse. *Le journal du CNRS n°224*. Septembre 2008
- Durand T, de Brabant C, Gomez E.** *Pour une approche intelligente de l'intelligence stratégique*. IDT93
- Editorial.** Patently transparent. *Nat Biotechnol*. 2006 May;24(5):474

- European Commission**, FP6 Instruments Task Force. *Provisions for implementing specific targeted research projects*. Edition: 12 May 2003
- European Commission**, Research DG. *Guide to Intellectual Property Rights provisions for FP6 projects*. Version 2: 28 June 2006
- European Commission & European Patent Office**. *Why researchers should care about patents*. 2007
- Fischer A, Hacein-Bey-Abina S, Cavazzana-Calvo M**. Gene therapy of children with X-linked severe combined immune deficiency: efficiency and complications. *Med Sci (Paris)*. 2004 Jan;20(1):115-7
- Fraser MJ, Ciszczon T, Elick T, Bauser C**. Precise excision of TTAA-specific lepidopteran transposons *piggyBac* (IFP2) and *tagalong* (TFP3) from the baculovirus genome in cell lines from two species of *Lepidoptera*. *Insect Mol Biol*. 1996 May;5(2):141-51
- Garrison BS, Yant SR, Mikkelsen JG, Kay MA**. Postintegrative gene silencing within the *Sleeping Beauty* transposition system. *Mol Cell Biol*. 2007 Dec;27(24):8824-33
- Girod PA, Nguyen D, Calabrese D, Puttini S, Grandjean M, Regamey A, Martinet D, Regamey A, Saugy D, Beckmann JS, Bucher P, Mermoud N**. Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells. *Nat Methods*. 2007 Sep;4(9):747-53
- Grosveld F, van Assendelft GB, Greaves DR, Kollias G**. Position-independent, high-level expression of the human β -globin gene in transgenic mice. *Cell*. 1987 Dec 24;51(6):975-85
- Guénec N**. *Comment accéder au cœur des laboratoires chinois*. 2010. www.hec.fr/eurasia
- Guénec N & Dou H**. Intérêt et méthode d'extraction de l'information scientifique chinoise. *Cahiers de la documentation*. 2008;4

- Guglielmi L, Truffinet V, Cogné M, Denizot Y.** The beta-globin HS4 insulator confers copy-number dependent expression of IgH regulatory elements in stable B cell transfectants. *Immunol Lett.* 2003 Oct 31;89(2-3):119-23
- Gurramkonda C, Adnan A, Gäbel T, Lünsdorf H, Ross A, Nemani SK, Swaminathan S, Khanna N, Rinas U.** Simple high-cell density fed-batch technique for high-level recombinant protein production with *Pichia pastoris*: Application to intracellular production of Hepatitis B surface antigen. *Microb Cell Fact.* 2009 Feb 10;8:13
- Gurramkonda C, Polez S, Skoko N, Adnan A, Gäbel T, Chugh D, Swaminathan S, Khanna N, Tisminetzky S, Rinas U.** Application of simple fed-batch technique to high-level secretory production of insulin precursor using *Pichia pastoris* with subsequent purification and conversion to human insulin. *Microb Cell Fact.* 2010 May 12;9:31
- Harraghy N, Gaussin A, Mermod N.** Sustained transgene expression using MAR elements. *Curr Gene Ther.* 2008 Oct;8(5):353-66
- Herrera S.** Profile: Richard Jefferson. *Nat Biotechnol.* 2005 June;23(6):643
- Hori Y, Suzuki K, Okuno Y, Nagaoka M, Futaki S, Sugiura Y.** Artificial Zinc Finger Peptide Containing a Novel His₄ Domain. *J. Am. Chem. Soc.*, 2000, 122(32);7648–7653
- Institut National de la Propriété Industrielle.** Le coordonnateur régional propriété intellectuelle pour l'Asie du Nord, Service Économique de l'Ambassade de France en Chine. *La nouvelle loi chinoise sur les brevets.* 2010
- Isaac AG & Park WG.** *Open Development: Is the "Open Source" Analogy Relevant to Biotechnology?* The Role of Intellectual Property Rights in Biotechnology Innovation. December 2009
- Ivics Z, Hackett PB, Plasterk RH, Izsvák Z.** Molecular reconstruction of *Sleeping Beauty*, a Tc1-like transposon from fish, and its transposition in human cells. *Cell.* 1997 Nov 14;91(4):501-10
- Ivics Z & Izsvák Z.** Transposons for Gene Therapy! *Curr Gene Ther.* 2006 Oct;6(5):593-607

- Ivics Z, Katzer A, Stüwe EE, Fiedler D, Knespel S, Izsvák Z.** Targeted Sleeping Beauty transposition in human cells. *Mol Ther.* 2007 Jun;15(6):1137-44
- Izsvák Z & Ivics Z.** *Sleeping beauty* transposition: biology and applications for molecular therapy. *Mol Ther.* 2004 Feb;9(2):147-56
- Jones SD.** Technologies to improve cell line development and engineering. *Cambridge Healthtech Institute PEptalk Conference.* January 14, 2009
- Jongste H.** Europe - New rules for divisional patent applications: patent strategy will need to change. *Building and enforcing intellectual property value.* 2010
- Kister J & Dou H.** Intégration de l'IE et de la veille dans un laboratoire de recherche académique scientifique. *Intelligence économique et problèmes décisionnels.* 2010;239-261
- Klabunde J, Kunze G, Gellisen G, Hollenberg CP.** Wide-Range Integrative Expression Vectors for Fungi, Based on Ribosomal DNA Elements, in *Production of Recombinant Proteins: Novel Microbial and Eukaryotic Expression Systems.* ed G. Gellissen, 2004, Wiley-VCH, Weinheim, p273-286
- Kwaks TH & Otte AP.** Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. *Trends Biotechnol.* 2006 Mar;24(3):137-42
- Lambert N.** Internet patent information in the 21st century: a comparison of Delphion, micropatent and QPAT. *International Chemical Information Conference & Exhibition.* Annecy, France, 17-20 October, 2004
- Ledergerber W & Kurt A.** The Swiss Federal Institute of Intellectual Property's new search services to assist corporate strategic decision-making. *World Patent Information.* 2003, 25;57-62
- Lesca H.** *Veille stratégique.* Ed. Aster, Villeurbanne, 1994
- List J.** Free patent databases come of age. *World Patent Information.* 2008, 30;185-186

- Liu X, Liu M, Xue Z, Pan Q, Wu L, Long Z, Xia K, Liang D, Xia J.** Non-viral *ex vivo* transduction of human hepatocyte cells to express factor VIII using a human ribosomal DNA-targeting vector. *J Thromb Haemost.* 2007 Feb;5(2):347-51
- Liu J, Skjørringe T, Gjetting T, Jensen TG.** *PhiC31* integrase induces a DNA damage response and chromosomal rearrangements in human adult fibroblasts. *BMC Biotechnol.* 2009 Apr 2;9:31
- Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, Unger-Wallace E, Sander JD, Müller-Lerch F, Fu F, Pearlberg J., Göbel C., Dassie JP, Pruett-Miller SM, Porteus MH, Sgroi DC, Iafrate AJ, Dobbs D, McCray PB, Cathomen T, Voytas DF and Joung JK.** Rapid "Open-Source" Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification. *Mol Cell.* 2008 Jul 25;31(2):294-301
- Maragathavally KJ, Kaminski JM, Coates CJ.** Chimeric *Mos1* and *piggyBac* transposases result in site-directed integration. *FASEB J.* 2006 Sep;20(11):1880-2
- Martre H et al.** *Intelligence économique et stratégie des entreprises.* Commissariat Général du Plan, travaux du groupe présidé par Henri Martre, La documentation française, 1994
- Mátés L, Chuah MK, Belay E, Jerchow B, Manoj N, Acosta-Sanchez A, Grzela DP, Schmitt A, Becker K, Matrai J, Ma L, Samara-Kuko E, Gysemans C, Pryputniewicz D, Miskey C, Fletcher B, VandenDriessche T, Ivics Z, Izsvák Z.** Molecular evolution of a novel hyperactive *Sleeping Beauty* transposase enables robust stable gene transfer in vertebrates. *Nat Genet.* 2009 Jun;41(6):753-61
- Miskey C, Izsvák Z, Kawakami K, Ivics Z.** DNA transposons in vertebrate functional genomics. *Cell Mol Life Sci.* 2005 Mar;62(6):629-41
- Medhora M, Maruyama K, Hartl DL.** Molecular and functional analysis of the *mariner* mutator element *Mos1* in *Drosophila*. *Genetics.* 1991 Jun;128(2):311-8

- Mutskov VJ, Farrell CM, Wade PA, Wolffe AP, Felsenfeld G.** The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. *Genes Dev.* 2002 Jun 15;16(12):1540-54
- Office Européen des Brevets.** La Chine bientôt devant les États-Unis et le Japon. *Patent Information News.* 2008a;3
- Office Européen des Brevets.** Troisième révision en cours de la loi chinoise sur les brevets. *Patent Information News.* 2008b;1
- Office Européen des Brevets.** Recherche de documents de brevets coréens avec des mots-clés anglais. *Patent Information News.* 2008c;1
- Office Européen des Brevets.** Des traductions automatiques pour la recherche de brevets ? *Patent Information News.* 2008d;4
- Office Européen des Brevets.** Information brevets coréenne Partie 1 : Décrypter les documents brevets coréens. *Patent Information News.* 2009a;3
- Office Européen des Brevets.** Information brevets coréenne Partie 2 : Données sur le statut juridique en Corée. *Patent Information News.* 2009b;4
- Office Européen des Brevets.** Oui, oui, non - Premiers résultats d'une vaste enquête auprès de l'industrie européenne. *Patent Information News.* 2010a;2
- Office Européen des Brevets.** La divulgation des brevets sous la loupe des universitaires. *Patent Information News.* 2010b;3
- Office Européen des Brevets.** iPairs – portail de l'information brevets indienne. *Patent Information News.* 2010c;2
- Office Européen des Brevets.** Comment les examinateurs de l'OEB utilisent-ils les documents chinois ? *Patent Information News.* 2010d;1
- Office Européen des Brevets.** Lost in Translation ? Traduction automatique et traduction manuelle de documents de brevets coréens. *Patent Information News.* 2010e;1

- Office Européen des Brevets.** PLUTO : traductions multilingues de brevets en ligne. *Patent Information News*. 2010f;2
- Organisation Mondiale de la Propriété Intellectuelle.** Guide de l'OMPI sur l'utilisation de l'information en matière de brevets. 2003. *WIPO Publication* No. L434/3(F)
- Organisation Mondiale de la Propriété Intellectuelle.** *Guide d'utilisation de la classification internationale des brevets*. Version 2009
- Palazzoli F, Bigot Y, Dacheux A.** Le rôle des brevets dans la protection des biomédicaments et de leurs systèmes de production. *Bio³: Biotechnologies – Bioproduction – Biomédicaments*. Tours : IMT Éditions, 2010c, p421-449
- Palazzoli F, Bire S, Bigot Y, Bonnin-Rouleux F.** Patent landscape of chromatin control elements: position effects in the pharmaceutical bioproduction. *Nat Biotechnol*. 2011 ; accepté
- Palazzoli F, Carnus E, Merly F, Goiran S, Bigot Y.** Zinc finger monopoly : quelles sont les règles du jeu ? *Biofutur*. 2010b ;306:58-62
- Palazzoli F, Carnus E, Wells DJ, Bigot Y.** Sustained transgene expression using non-viral enzymatic systems for stable chromosomal integration. *Curr Gene Ther*. 2008 Oct;8(5):367-90
- Palazzoli F, Testu FX, Merly F, Bigot Y.** Transposon tools: worldwide landscape of intellectual property and technological developments. *Genetica*. 2010a Mar;138(3):285-99
- Pfeifer A, Ikawa M, Dayn Y, Verma IM.** Transgenesis by lentiviral vectors: lack of gene silencing in mammalian embryonic stem cells and preimplantation embryos. *Proc Natl Acad Sci USA*. 2002 Feb 19;99(4):2140-5
- Pikaart MJ, Recillas-Targa F, Felsenfeld G.** Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev*. 1998 Sep 15;12(18):2852-62

- Potts W, Tucker D, Wood H, Martin C.** Chicken beta-globin 5'HS4 insulators function to reduce variability in transgenic founder mice. *Biochem Biophys Res Commun.* 2000 Jul 14;273(3):1015-8
- Questel.** *Manuel d'utilisation de QPAT.* Version 7. Avril 2010
- Rangwala SH, Kazazian HH Jr.** The *L1* retrotransposition assay: a retrospective and toolkit. *Methods.* 2009 Nov;49(3):219-26
- Rebar EJ, Huang Y, Hickey R, Nath AK, Meoli D, Nath S, Chen B, Xu L, Liang Y, Jamieson AC, Zhang L, Spratt SK, Case CC, Wolffe A, Giordano FJ.** Induction of angiogenesis in a mouse model using engineered transcription factors. *Nat Med.* 2002 Dec;8(12):1427-32
- Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bell AC, Litt MD, West AG, Gaszner M, Felsenfeld G.** Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci USA.* 2002 May 14;99(10):6883-8
- Recillas-Targa F, Valadez-Graham V, Farrell CM.** Prospects and implications of using chromatin insulators in gene therapy and transgenesis. *Bioessays.* 2004 Jul;26(7):796-807
- Scott CT.** The zinc finger nuclease monopoly. *Nat Biotechnol.* 2005 Aug;23(8):915-8
- Stewart S, MacDonald N, Perkins E, deJong G, Perez C, Lindebaum M.** Retrofitting of a satellite repeat DNA-based murine artificial chromosome (ACes) to contain *loxP* recombination sites. *Gene Ther.* 2002 Jun;9(11):719-23
- Tan W, Dong Z, Wilkinson TA, Barbas CF 3rd, Chow SA.** Human immunodeficiency virus type 1 incorporated with fusion proteins consisting of integrase and the designed polydactyl zinc finger protein E2C can bias integration of viral DNA into a predetermined chromosomal region in human cells. *J Virol.* 2006 Feb;80(4):1939-48
- Transposagen Biopharmaceuticals.** *Better animal models – Better results.* Factsheet. Aug2009

- Trippe AJ.** Patinformatics: Identifying haystacks from space. *Searcher*. 2002;10(9):28
- Trippe AJ.** Patinformatics: Tasks to tools. *World Patent Information*. 2003, 25;211–221
- Urschitz J, Kawasumi M, Owens J, Morozumi K, Yamashiro H, Stoytchev I, Marh J, Dee JA, Kawamoto K, Coates CJ, Kaminski JM, Pelczar P, Yanagimachi R, Moisyadi S.** Helper-independent *piggyBac* plasmids for gene delivery approaches: strategies for avoiding potential genotoxic effects. *Proc Natl Acad Sci USA*. 2010 May 4;107(18):8117-22
- Wen L, Liang de S, Wu L, Cai F, Pan Q, Long ZG, Dai HP, Xia K and Xia JH.** Efficient protein expression from the endogenous RNA polymerase I promoter using a human ribosomal DNA targeting vector. *Biochem Biophys Res Commun*. 2008 Mar 21;367(4):846-51
- Williams DA.** *Sleeping Beauty* vector system moves toward human trials in the United States. *Mol Ther*. 2008 Sep;16(9):1515-6
- Williams S & Cranenburgh R.** Improving Clone Production for Increased Protein Yield from Mammalian Cell Lines. *Innovations in Pharmaceutical Technology*. 2008, 26:52-56
- Wu SC, Meir YJ, Coates CJ, Handler AM, Pelczar P, Moisyadi S, Kaminski JM.** *piggyBac* is a flexible and highly active transposon as compared to *Sleeping Beauty*, *Tol2*, and *Mos1* in mammalian cells. *Proc Natl Acad Sci USA*. 2006 Oct 10;103(41):15008-13
- Zahn-Zabal M, Kobr M, Girod PA, Imhof M, Chatellard P, de Jesus M, Wurm F, Mermod N.** Development of stable cell lines for production or regulated expression using matrix attachment regions. *J Biotechnol*. 2001 Apr 27;87(1):29-42

Sites web consultés

- w1** http://cordis.europa.eu/fetch?CALLER=FP6_PROJ&ACTION=D&DOC=1&CAT=PROJ&QUERY=012a9f333a37:b8c0:5a74d19e&RCN=79830 [consulté le 08 novembre 2010]
- w2** http://www.epo.org/patents/patent-information/about_fr.html [consulté le 24 janvier 2011]
- w3** http://www.wipo.int/sme/fr/documents/patent_information.htm [consulté le 24 janvier 2011]
- w4** http://www.epo.org/patents/patent-information/business/stats/faq_fr.html [consulté le 24 janvier 2011]
- w5** <http://www.matheo-patent.com/> [consulté le 24 janvier 2011]
- w6** <http://www.matheo-analyzer.com/> [consulté le 24 janvier 2011]
- w7** http://www.oecd.org/document/41/0,3343,en_2649_34537_35534441_1_1_1_1,00.html [consulté le 24 janvier 2011]
- w8** <http://gicc.univ-tours.fr/laboratoire/presentation/equipe8.php?connect=&lang=fr> [consulté le 17 décembre 2010]
- w9** <http://www.fist.fr/> [consulté le 24 janvier 2011]
- w10** <http://www.fist.fr/fr/ip-overview/index-2.html> [consulté le 24 janvier 2011]
- w11** <http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed> [consulté le 17 décembre 2010]
- w12** <http://www.espacenet.com/info/index.fr.htm> [consulté le 17 décembre 2010]
- w13** <http://www.qpat.com/index.htm> [consulté le 17 décembre 2010]
- w14** <http://www.intellixir.com/default.asp> [consulté le 17 décembre 2010]
- w15** http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Accueil_principal [consulté le 17 décembre 2010]
- w16** http://fr.wikipedia.org/wiki/%C3%89l%C3%A9ment_transposable [consulté le 17 décembre 2010]
- w17** <http://www.google.fr/> [consulté le 17 décembre 2010]
- w18** <http://www.copernic.com/fr/products/agent/index.html> [consulté le 17 décembre 2010]
- w19** <http://lesmoteursderecherche.com/tabella.htm> [consulté le 17 décembre 2010]
- w20** <http://www.journaldunet.com/0411/041116googleyahoo.shtml> [consulté le 17 décembre 2010]
- w21** <http://outils.abondance.com/google.html> [consulté le 17 décembre 2010]
- w22** <http://taistoiquandtuparles.over-blog.com/article-28338370.html> [consulté le 17 décembre 2010]
- w23** http://fr.espacenet.com/help?locale=fr_FR&method=handleHelpTopic&topic=ecla [consulté le 17 décembre 2010]
- w24** <http://www.inpi.fr/fr/services-et-prestations/bases-de-donnees-gratuites/base-brevets.html> [consulté le 17 décembre 2010]

- w25** <http://patft.uspto.gov/> [consulté le 17 décembre 2010]
- w26** <http://www.espacenet.com/info/index.fr.htm> [consulté le 17 décembre 2010]
- w27** http://fr.espacenet.com/help?locale=fr_FR&method=handleHelpTopic&topic=bibliocoverage [consulté le 17 décembre 2010]
- w28** <http://www.istl.org/06-summer/electronic3.html> [consulté le 17 décembre 2010]
- w29** http://fr.espacenet.com/help?locale=fr_FR&method=handleHelpTopic&topic=npl [consulté le 17 décembre 2010]
- w30** http://ep.espacenet.com/advancedSearch?locale=fr_EP [consulté le 17 décembre 2010]
- w31** http://fr.espacenet.com/help?locale=fr_FR&method=handleHelpTopic&topic=limitations [consulté le 17 décembre 2010]
- w32** <http://www.wipo.int/pctdb/fr/index.jsp> [consulté le 17 décembre 2010]
- w33** <http://www.wipo.int/patentscope/search/en/search.jsf> [consulté le 17 décembre 2010]
- w34** <http://www.wipo.int/patentscope/search/en/clir/clir.jsp> [consulté le 17 décembre 2010]
- w35** <http://www.cas.org/products/sfacad/index.html> [consulté le 17 décembre 2010]
- w36** http://ep.espacenet.com/?locale=fr_EP&view=seven [consulté le 17 décembre 2010]
- w37** <http://www.piug.org/vendors.php#DBProducers> [consulté le 17 décembre 2010]
- w38** <http://www.questel.fr/FR/index.htm> [consulté le 17 décembre 2010]
- w39** <http://www.questel.fr/produitsetservices/fampat.htm> [consulté le 17 décembre 2010]
- w40** <http://www.millipore.com/processdev/pd3/ucoc&open&cid=S0348> [consulté le 17 décembre 2010]
- w41** http://www.minosbiosystems.com/sci_tec.htm [consulté le 17 décembre 2010]
- w42** <http://www.discoverygenomics.net/sbts.html> [consulté le 17 décembre 2010]
- w43** <http://investor.sangamo.com/releasedetail.cfm?ReleaseID=247051> [consulté le 17 décembre 2010]
- w44** http://www.selexis.com/media/news_060523.html [consulté le 17 décembre 2010]
- w45** http://www.millipore.com/press/pr3/press_release12042008 [consulté le 17 décembre 2010]
- w46** http://www.millipore.com/pressroom/pr3/innovata_ucoc [consulté le 17 décembre 2010]
- w47** http://www.cruceell.com/Investors-Press_Releases# [consulté le 17 décembre 2010]
- w48** <http://www.biotech-finances.com/fr/> [consulté le 17 décembre 2010]
- w49** <http://www.bioworld.com/> [consulté le 17 décembre 2010]
- w50** <http://www2.prnewswire.com/cgi-bin/stories.pl?ACCT=BIOWORLD.story&STORY=/www/story/10-05-2010/0005324992&EDATE=TUE+Oct+05+2010,+08:30+AM> [consulté le 17 décembre 2010]

- w51** <http://www2.prnewswire.com/cgi-bin/stories.pl?ACCT=BIOWORLD.story&STORY=/www/story/10-04-2010/0005323733&EDATE=MON+Oct+04+2010,+02:41+AM> [consulté le 17 décembre 2010]
- w52** <http://www.buzz4bio.com/> [consulté le 17 décembre 2010]
- w53** <http://www.actinbiotech.com/FIB/index.php> [consulté le 17 décembre 2010]
- w54** <http://www.gopubmed.com/web/gopubmed/> [consulté le 17 décembre 2010]
- w55** http://cordis.europa.eu/fetch?CALLER=FP6_PROJ&ACTION=D&DOC=1&CAT=PROJ&QUERY=012a9f30725f:3951:57d4af6c&RCN=78563 [consulté le 11 novembre 2010]
- w56** <http://www.transposagenbio.com/> [consulté le 11 novembre 2010]
- w57** <http://www.transposagenbio.com/page.php?page=piggybac-licenses> [consulté le 11 novembre 2010]
- w58** <http://www.bizjournals.com/pacific/stories/2009/08/10/story2.html?page=2> [consulté le 11 novembre 2010]
- w59** <http://www.manoabiosciences.com/> [consulté le 11 novembre 2010]
- w60** http://www.wipo.int/sme/fr/documents/patents_biotech.htm [consulté le 11 novembre 2010]
- w61** <http://www.cncpi.fr/iaa145-45-propriete-intellectuelle-monde.htm?level=> [consulté le 11 novembre 2010]
- w62** http://www.wipo.int/wipo_magazine/fr/2009/02/article_0002.html [consulté le 11 novembre 2010]
- w63** http://www.wipo.int/patent-law/fr/developments/open_source.html [consulté le 11 novembre 2010]
- w64** <http://www.cambia.org/daisy/cambia/home.html> [consulté le 11 novembre 2010]
- w65** <http://www.openinnovation.org/daisy/ioi/home.html> [consulté le 11 novembre 2010]
- w66** <http://www.patentlens.net/> [consulté le 11 novembre 2010]
- w67** <http://www.patentlens.net/daisy/patentlens/landscapes-tools.html> [consulté le 11 novembre 2010]
- w68** <http://www.zincfingers.org/> [consulté le 11 novembre 2010]
- w69** <http://bindr.gdcb.iastate.edu/ZiFiT/> [consulté le 11 novembre 2010]
- w70** <http://bindr.gdcb.iastate.edu:8080/ZiFDB/> [consulté le 11 novembre 2010]
- w71** www.addgene.org/pgvec1?f=a&cmd=showfile&file=sangamo [consulté le 11 novembre 2010]
- w72** http://www.helmholtz-hzi.de/de/presse_und_oeffentlichkeit/pressemitteilungen/ansicht/article/complete/neuer_weg_zur_insulinherstellung/ [consulté le 11 novembre 2010]
- w73** <http://www.who.int/mediacentre/factsheets/fs312/fr/index.html> [consulté le 11 novembre 2010]
- w74** <http://www.systranet.fr/> [consulté le 21 janvier 2011]
- w75** <http://translate.google.fr/> [consulté le 21 janvier 2011]
- w76** <http://www.google.com/> [consulté le 21 janvier 2011]

-
- w77** <http://www.google.com.hk/> [consulté le 21 janvier 2011]
- w78** <http://www.baidu.com/> [consulté le 21 janvier 2011]
- w79** <http://search.cnki.net/> [consulté le 21 janvier 2011]
- w80** <http://www.wipo.int/ipstats/fr/statistics/patents/> [consulté le 21 janvier 2011]
- w81** <http://www.ip-sharing.com/Papers/news0000000765.php> [consulté le 21 janvier 2011]
- w82** <http://www.latribune.fr/actualites/economie/international/20101012trib000559284/comment-la-chine-va-devenir-la-championne-du-monde-des-brevets-.html> [consulté le 21 janvier 2011]
- w83** <http://www.iam-magazine.com/blog/detail.aspx?g=11f9cec8-4af8-42eb-9338-2f149504d60d&q> [consulté le 21 janvier 2011]
- w84** <http://www.epo.org/patents/patent-information/east-asian.html> [consulté le 21 janvier 2011]
- w85** http://www.epo.org/patents/patent-information/east-asian/translation/K2E-PAT_fr.html [consulté le 21 janvier 2011]
- w86** www.cnpat.com.cn [consulté le 21 janvier 2011]
- w87** www.sipo.gov.cn/sipo_English/ [consulté le 21 janvier 2011]
- w88** <http://english.cnipr.com> [consulté le 21 janvier 2011]
- w89** <http://www.pluto-patenttranslation.eu/> [consulté le 21 janvier 2011]
- w90** <http://breese.blogs.com/pi/2010/11/chine.html> [consulté le 21 janvier 2011]

Annexe 1

**Le rôle des brevets dans la protection des
biomédicaments et de leurs systèmes de production.**

**Chapitre 6 de l'ouvrage Bio³ : Biotechnologies –
Bioproduction – Biomédicaments**

PARTIE 6

Protection de la propriété industrielle



BIOPRODUCTION ET PROPRIÉTÉ INDUSTRIELLE

LE RÔLE DES BREVETS DANS LA PROTECTION DES BIOMÉDICAMENTS ET DE LEURS SYSTÈMES DE PRODUCTION

Fabien PALAZZOLI
UMR 6239 CNRS et
Université FR37

Yves BIGOT
UMR 6239 CNRS et
Université FR37

Angélique DACHEUX
Ingénieur Brevet – LFB

1. DU MÉDICAMENT AU BIOMÉDICAMENT : ÉVOLUTION DU PAYSAGE DES PRODUITS THÉRAPEUTIQUES ET IMPORTANCE DES BREVETS

1.1. Impact des Biotechnologies sur le marché des produits thérapeutiques

Le marché des produits thérapeutiques est en plein bouleversement et ce depuis une quinzaine d'années. Cette situation est en partie due à l'expiration de nombreux brevets protégeant principalement des médicaments issus de l'industrie chimique. À l'expiration d'un brevet, celui-ci « tombe » dans le domaine public, permettant ainsi les copies aussi appelées génériques. Un générique est une copie conforme d'un médicament princeps, comportant le même principe actif et ayant le même effet thérapeutique.

La molécule princeps est protégée très en amont de son exploitation commerciale. En effet celle-ci intervient généralement 10 à 12 ans après le dépôt de la demande de brevet principale. Les essais cliniques ont lieu entre 5 et 10 ans après le dépôt de cette même demande de brevet (Figure 1). L'expiration des brevets et l'arrivée des génériques font partie intégrante de ce bouleversement du marché pharmaceutique.

Cette évolution majeure concerne tous les acteurs liés de près ou de loin aux médicaments : les patients et associations de défense de malades, les médecins ainsi que tous les intervenants de la chaîne du médicament, de la R&D à la mise sur le marché de ces médicaments, en passant par les instances nationales et européenne qui gèrent les dépenses de santé. En effet, beaucoup de gouvernements comptent sur les génériques pour réduire les coûts de fonctionnement des systèmes de santé. Par conséquent, ils favorisent leur développement et leur mise en vente à des prix inférieurs à leurs équivalents princeps.

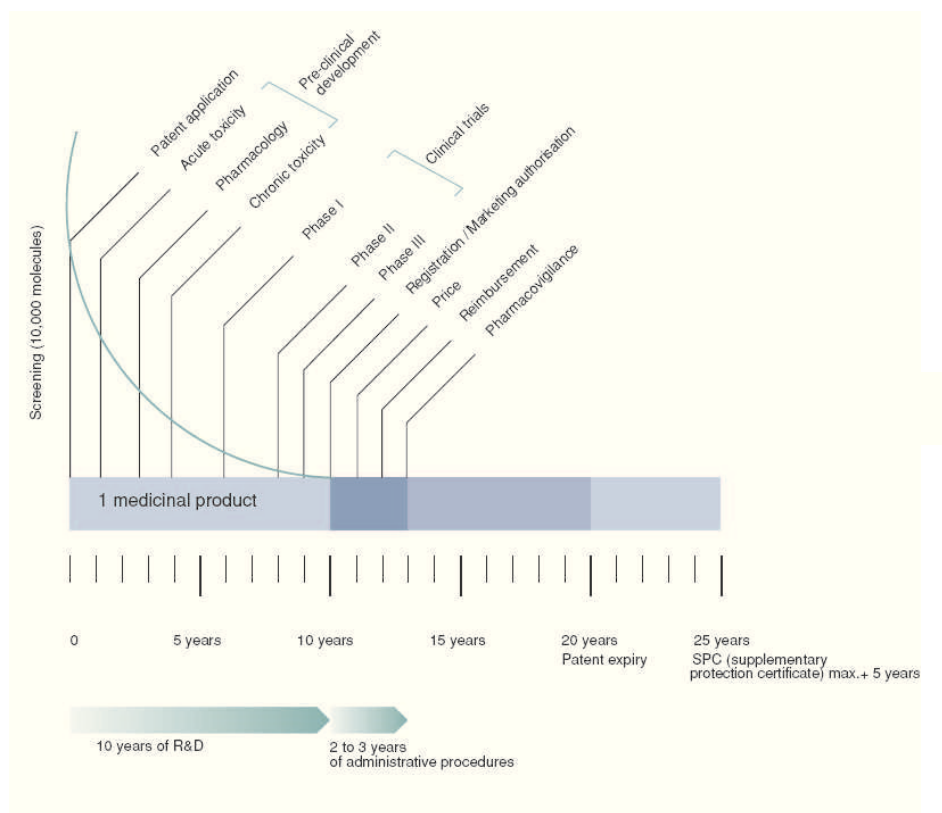


Figure 1 : Schéma de développement d'une molécule thérapeutique, du dépôt de brevet à son exploitation commerciale (Source: *Fédération Européenne des Associations et Industries Pharmaceutiques 2008*).

Les grandes entreprises pharmaceutiques, également appelées « Big Pharmas », subissent quant à elles de plein fouet le développement des génériques, particulièrement lorsque le produit princeps correspondant est un blockbuster. Un blockbuster est une molécule dont l'exploitation génère un chiffre d'affaire de plus de 1 milliard de \$US (plus de 660 millions €) par an. L'expiration des brevets relatifs à ces blockbusters et la commercialisation des génériques correspondants impactent fortement le chiffre d'affaire global des sociétés pharmaceutiques. Pour exemple, le médicament Singulair® de Merck (traitement de l'asthme) va tomber dans le domaine public en 2012. Ce médicament génère actuellement plus de 4,5 milliards de vente. C'est également le cas de l'anticholestérol vedette Lipitor® de Pfizer, qui représente à lui seul 12 milliards de vente, et dont le brevet expire en 2011. L'arrivée de génériques engendrera nécessairement une perte de chiffre d'affaire pour le produit princeps.

L'innovation est de plus en plus difficile dans le domaine des médicaments, et ceci pousse les sociétés pharmaceutiques à s'adapter à cette évolution majeure de leur environnement, en redéfinissant leurs stratégies R&D et commerciale. A ce jour, plusieurs options peuvent être envisagées. Certaines sociétés ont opté pour produire et commercialiser des génériques de leurs molécules princeps. Ainsi, Sanofi-Aventis devrait commercialiser sa propre version générique de l'un des médicaments les plus vendus au monde : le Plavix® (antithrombotique avec plus de 2,5 milliards d'euros de ventes dans le monde en 2008). De son côté, Novartis réalise des acquisitions d'achats de sociétés spécialisées dans les génériques (Azupharma, Lek). La firme Lilly a de son côté poursuivi en justice pour contrefaçon des fabricants de génériques pour son antidépresseur Prozac®, exemple d'une lutte à l'issue juridique incertaine. D'autres entreprises pharmaceutiques vont se rapprocher et fusionner afin de posséder un portefeuille de molécules en développement plus conséquent. C'est l'opération choisie par Pfizer et Wyeth (opération de 46,8 milliards US\$), ou de Merck et Schering-Plough (pour un coût de 41 milliards US\$).

Enfin, d'autres encore vont se tourner vers les molécules innovantes issues du passage des technologies de l'industrie de la chimie à celles de l'industrie des Biotechnologies (appartenant alors aux « Biotechnologies rouges »). Cela peut se faire par l'acquisition de molécules innovantes par des accords de licensing-in, c'est-à-dire l'acquisition de produits à différents stades de développement (préclinique ou clinique, via des contrats de licence ou de cession des droits). Ceci peut aussi se faire par le biais de rachats de sociétés de biotechnologies de taille plus ou moins importante. C'est l'option choisie par Roche qui a récemment racheté pour 46,8 milliards US\$, Genentech, l'une des plus grandes entreprises de Biotechnologies au monde, et de Merck qui a annoncé son souhait d'acquérir GlycoFi pour 400 millions de US\$.

Ces sociétés biopharmaceutiques vont donc y puiser de nouvelles sources d'innovation thérapeutique, et se positionner sur des marchés émergents à forte croissance : celui des biomédicaments. Le développement de ces derniers va profiter pleinement des progrès scientifiques et des évolutions technologiques issus des biotechnologies, comme par exemple l'ingénierie génétique, les criblages à haut débit, les analyses bioinformatiques et les séquençages de génomes (dont le projet international Human Genome Organisation, communément appelé HUGO). Cependant, les Biotechnologies sont un des secteurs où la recherche est la plus intensive, et la R&D exige des investissements financiers considérables. La protection des résultats qui en sont issus revêt donc une importance capitale pour leurs développements industriels. Le brevet est donc, là encore, un élément essentiel, encore plus que pour les molécules chimiques thérapeutiques, pour la valorisation des biomédicaments.

1.2. Protection par brevet : généralités et cas des biomédicaments

1.2.1. Définition

Un brevet d'invention est un titre de Propriété Industrielle qui offre à son titulaire la possibilité d'interdire à tout tiers l'exploitation d'une invention couverte par un brevet. Ce droit d'interdire revient à accorder au titulaire un monopole d'exploitation de son invention. Ce droit est limité dans le temps - 20 ans à compter de la date de dépôt de

la demande de brevet – et dans l’espace – car il est limité au(x) territoire(s) dans le(s)quel(s) la protection a été demandée, puis accordée.

Un brevet se présente en trois parties : une première partie qui comprend toutes les informations bibliographiques (noms des déposants et inventeurs, date de dépôt, etc.). Une deuxième partie appelée « description » qui comme son nom l’indique, décrit les caractéristiques de l’invention, les moyens d’y parvenir, et est accompagnée le cas échéant de figures (plasmides, etc.), de tableaux. Enfin, la troisième partie, appelée « revendications », définit la nature de la protection juridique souhaitée par le demandeur.

1.2.2. La procédure brevet

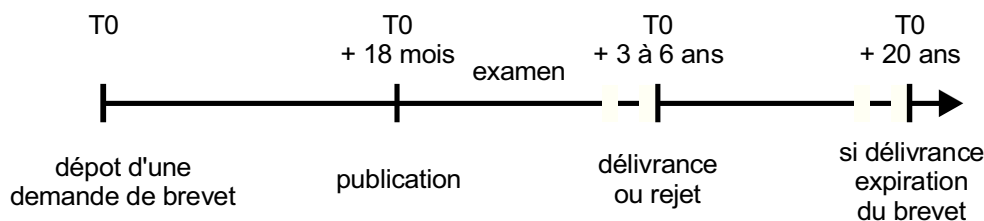
Des étapes clés jalonnent la vie du brevet : dépôt, extension, publication, examen, délivrance. La vie d’un brevet débute à la date de dépôt de la demande de brevet (la terminologie « demande de brevet » est à ne pas confondre avec « brevet » car dans le premier cas, il s’agit de la protection qui a été demandée et dans le second de la protection qui a été accordée).

Cette date de dépôt génère un droit dit « droit de priorité » qui permet pendant une durée maximum de 12 mois de demander une protection pour la même invention, dans d’autres pays que celui où la première demande a été déposée. Cette demande de protection supplémentaire est appelée « extension de la protection à l’étranger ». Si le demandeur ne souhaite pas bénéficier de son droit de priorité, l’étape d’extension de la protection à l’étranger n’a pas lieu (Figure 2).

Dans les deux cas, l’étape suivante est la publication, qui a lieu 18 mois après le dépôt de la demande de brevet. La publication rend accessible aux tiers les informations contenues dans la demande de brevet, informations techniques relatives à l’invention elle-même, détaillées dans la description, informations administratives (le demandeur, les inventeurs, la date de dépôt par exemple) et de ce fait parfois des informations sur la stratégie de développement d’une entreprise. Les demandes de brevet et brevets sont donc une source essentielle d’informations.

La demande de brevet entre ensuite dans un processus dit « d’examen » au cours duquel l’office de brevets du pays considéré détermine si la demande de brevet répond aux exigences de la loi en matière de brevets. Notamment, l’office de brevet détermine si la demande de brevet répond à des critères dits « de brevetabilité », à savoir : la nouveauté, l’activité inventive et l’application industrielle. A l’issue de cette période d’examen, le brevet est accordé (si tous les critères sont remplis) ou la demande de brevet est rejetée (si ce n’est pas le cas). Si le brevet est accordé, il expirera 20 ans après la date de dépôt.

procédure nationale française



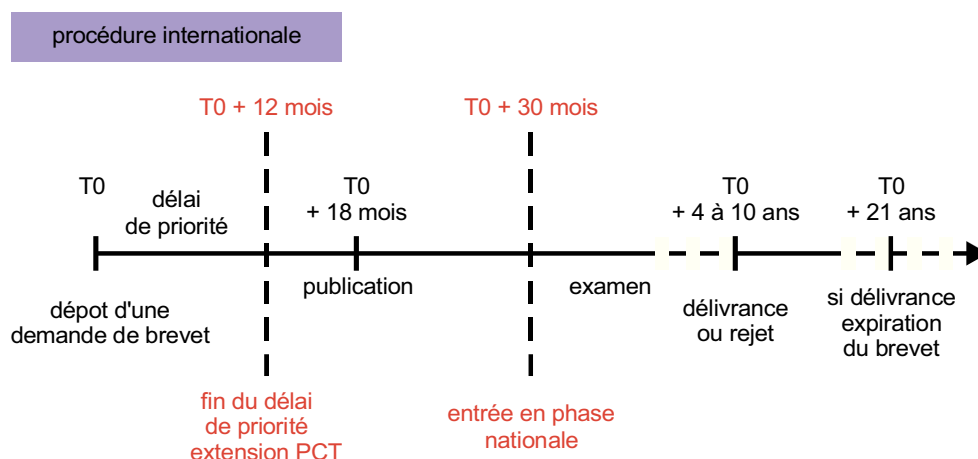


Figure 2 : Étapes majeures de la vie d'un brevet.

Si le demandeur a choisi de bénéficier de son droit de priorité, cela signifie qu'il a demandé (dans un délai de 12 mois suivant le dépôt d'une première demande), l'extension de la protection dans des pays supplémentaires. Classiquement, cette demande d'extension est réalisée par un système centralisé international appelé système du PCT (Patent Community Treaty). Ce système permet de demander la protection dans plus de 140 pays en n'effectuant qu'un seul dépôt d'une demande dite «internationale». L'étape de publication de la première demande a lieu de la même manière 18 mois après le dépôt. Dans ce cas de figure, deux publications ont lieu dans le même temps, celle de la première demande et celle de la demande internationale.

Lorsqu'une extension par voie internationale a été demandée, le demandeur doit désigner les pays où il souhaite obtenir une protection parmi les pays adhérents au système PCT dans les 30 mois à compter du dépôt de la première demande de brevet. Cette étape est appelée «entrée en phases nationales». A l'issue de ce choix, la demande de brevet sera examinée dans chacun des pays désignés, par l'office de brevet correspondant. Chacun de ces pays statuera sur la délivrance ou non du brevet. Il est à noter qu'il n'existe pas de brevet international (une demande unique ne permet pas d'obtenir un titre unique), seule la demande est internationale, mais elle aboutit en autant de brevets que de pays choisis. L'expiration des brevets interviendra 20 ans après la date de dépôt de la demande internationale soit 21 ans après le dépôt de la première demande.

1.2.3. Le Certificat Complémentaire de Protection (CCP)

Pour les médicaments et les produits phytopharmaceutiques, il existe un titre de Propriété Industrielle supplémentaire : le Certificat Complémentaire de Protection (CCP). Il est organisé par le droit des brevets mais fait intervenir des aspects réglementaires liés aux médicaments, car il ne peut être obtenu que si une Autorisation de Mise sur le Marché (AMM) a été accordée avant l'expiration du brevet. Le CCP permet au titulaire d'un brevet disposant d'une AMM pour le produit couvert par le brevet, de prolonger la durée de protection du produit en question. Cette disposition a été mise en place dans

de nombreux pays et vise à compenser l'impossibilité d'exploiter une invention de médicaments durant les phases préalables parfois longues d'essais cliniques. Il est important de souligner qu'un produit peut être couvert par un brevet délivré et ne pas obtenir d'AMM et à l'inverse, un produit ayant reçu une AMM peut ne pas être couvert par un brevet (les systèmes réglementaires et d'examen des brevets sont indépendants). Le CCP prolonge la durée de protection d'un brevet d'un maximum de 5 ans et son obtention ne peut pas générer une protection supérieure à 15 ans à compter de l'obtention de l'AMM en Europe (14 ans aux États-Unis).

1.2.4. Les conditions de brevetabilité

En matière de brevet, le biomédicament ne se distingue pas du médicament ou de toute autre invention. En effet, pour être protégeable par un brevet, le biomédicament doit être nouveau, impliquer une activité inventive et être susceptible d'application industrielle. Ces trois critères constituent le fondement de la brevetabilité d'une invention. En d'autres termes, pour pouvoir être protégeable par un brevet, un biomédicament :

- ne doit pas faire partie de l'état de la technique (concept de nouveauté), c'est-à-dire qu'il ne doit pas avoir fait l'objet d'une publication, d'une demande de brevet, d'une communication orale, d'un devis, d'une soumission de dossier d'AMM ou de tout autre type de divulgation, avant la date de dépôt de la demande de brevet.
- ne doit pas découler de manière évidente de ce qui est déjà connu (concept d'activité inventive). C'est-à-dire qu'à la lecture de ce qui est déjà connu, un homme normalement qualifié dans le domaine technique (appelé « homme du métier » en droit des brevets) ne doit pas pouvoir parvenir à l'invention.
- doit être susceptible d'application industrielle, c'est-à-dire être fabriqué ou utilisé dans l'industrie.

Il en va de même pour un procédé d'obtention ou un système de production d'un biomédicament que l'on souhaiterait protéger, ou encore d'une utilisation de ce biomédicament dans une indication donnée.

1.2.5. La Directive 98/44/CE relative à la brevetabilité du vivant

Il est difficile de s'intéresser aux biomédicaments sans évoquer la brevetabilité du vivant. Au niveau européen, la Directive 98/44/CE du 6 juillet 1998 tente de poser les grands principes de la brevetabilité dans le domaine du vivant et d'harmoniser les législations nationales. Pour produire ses effets, dans les droits nationaux, la Directive 98/44/CE a dû être transposée dans chacun des états membres. Elle a été transposée en France en 2004, en conséquence de quoi, le code de la propriété intellectuelle a été modifié.

Dans cette directive, le premier point à souligner (article 3) est le fait que la matière biologique est brevetable à condition que :

- l'invention respecte les trois principes fondateurs de la brevetabilité et,
- que cette matière biologique ait été produite à l'aide d'un procédé technique même si elle préexistait dans la nature.

→ ***Les inventions qui découlent de matière biologique ne sont donc pas exclues de la brevetabilité.***

La directive définit ce qu'est la « matière biologique » comme une matière contenant des informations génétiques et qui est autoreproductible ou reproductible dans un système biologique.

Cette même directive statue par ailleurs (article 5) sur la non-brevetabilité du corps humain et de ses éléments. Elle précise que « *le corps humain aux différents stades de sa constitution et de son développement, ainsi que la simple découverte d'un de ses éléments, y compris la séquence ou la séquence partielle d'un gène ne sont pas brevetables* »

→ ***En effet, ceci constituerait une « simple découverte », et les découvertes font parties des éléments exclus de la brevetabilité. (article L. 611-10 du Code de la Propriété Intellectuelle).***

Elle définit aussi les éléments exclus du champ de la brevetabilité dans le domaine du vivant, à savoir :

- les procédés de clonage des êtres humains,
- les procédés de modification de l'identité génétique germinale de l'être humain
- l'utilisation d'embryons humains à des fins industrielles ou commerciales,
- les procédés de modification de l'identité génétique des animaux de nature à provoquer chez eux des souffrances sans utilité médicale substantielle pour l'homme ou l'animal.

Cependant, la Directive 98/44/CE stipule aussi qu'« *un élément isolé du corps humain ou autrement produit par un procédé technique, y compris la séquence ou la séquence partielle d'un gène, peut constituer un élément brevetable même si sa structure est identique à celle de l'élément naturel* ».

→ ***Cela signifie que le fait d'isoler une protéine ou un gène par la mise en œuvre d'un procédé technique peut constituer une invention brevetable si cette invention répond aux critères de brevetabilité.***

Elle précise aussi que « *l'application industrielle d'une séquence ou d'une séquence partielle d'un gène doit être concrètement exposée dans la demande de brevet* ».

Il faut donc que l'application d'une séquence soit définie pour qu'une séquence soit brevetable. La fonction d'une telle séquence n'est pas nécessairement décryptée, mais son utilisation doit être décrite. La directive a introduit ici une limitation en ce qui concerne la brevetabilité des séquences. Du séquençage permettant de mettre en évidence des gènes ne pourraient pas donner lieu à un brevet si une application n'est pas associée à chaque séquence.

Enfin, il est important de souligner qu'en droit français seule l'utilisation d'une séquence ou d'une séquence partielle d'un gène est brevetable. En conséquence la portée d'une revendication couvrant une séquence ou une séquence partielle d'un gène est limitée à la partie de cette séquence directement liée à la fonction spécifique concrètement exposée dans la description.

À l'exception des exclusions ci-dessus, il ressort de cette directive, que tout médicament dérivé ou constitué de matière biologique (biomédicament) peut être brevetable si sa mise en évidence fait intervenir un procédé technique et si dans le cas d'une séquence, l'application de celle-ci est concrètement définie. Les biomédicaments doivent également répondre aux autres critères de brevetabilité, au même titre que toute invention.

1.2.6. Les différents types de brevet

Plusieurs types de brevets sont envisageables (brevet de produit, de procédé ou d'utilisation). La protection la plus large est donnée par un brevet de produit qui interdit à tout tiers d'exploiter le produit en question, quelque soit le procédé utilisé pour sa production ou l'utilisation qui en est faite. Un brevet de procédé (de purification d'une protéine par exemple) empêche un tiers de produire la protéine en question en reproduisant le procédé protégé, mais il n'empêche pas le tiers de produire ladite protéine par un autre procédé (de purification pour l'exemple précédent). Le brevet d'utilisation a une portée moins large que les deux types précédents, en ce sens qu'il interdit aux tiers d'utiliser un biomédicament donné dans une indication thérapeutique donnée mais n'empêche pas l'utilisation dudit biomédicament en tant que tel dans d'autres indications. Il est important de souligner qu'un même brevet peut contenir des revendications de produits, des revendications de procédé permettant l'obtention du produit et des revendications d'utilisation de ce produit. Mais ces différents types de protection peuvent aussi être présents dans des brevets différents et même détenus par des entités différentes.

2. LES BIOMÉDICAMENTS BREVETABLES : EXEMPLIFICATION

Pour rappel, un biomédicament est constitué de toute substance utilisée ou utilisable en thérapie issue de matière biologique (extraction) ou produite par un procédé de biotechnologies. Les biomédicaments sont soumis à une classification, qui comprend quatre grands groupes :

- les peptides et protéines
- les vaccins et anticorps
- les acides nucléiques
- la thérapie génique et cellulaire.

Les quatre grands groupes de biomédicaments peuvent se scinder en deux vis-à-vis de leur production. Les peptides et protéines, vaccins et anticorps, acides nucléiques vont être facilement industrialisables, à l'inverse des produits issus des thérapies génique et cellulaire, dont les technologies nécessitent d'être optimisées et validées.

Par ailleurs, il est important de distinguer dans la production de médicaments issus des biotechnologies :

- le biomédicament lui-même,
- les technologies mises en œuvre pour produire ce biomédicament.

Les deux prochaines parties du chapitre ont pour objectif de donner quelques exemples de revendications de brevets portant sur les biomédicaments eux-mêmes et sur leurs systèmes de production. Les brevets et demandes de brevets peuvent être consultés sur la base de données de l'Office Européen des Brevets.

(http://ep.espacenet.com/advancedSearch?locale=fr_EP).

2.1. Les peptides et protéines

Les peptides et les protéines sont subdivisés en fonction de leurs natures : les hormones, les facteurs de croissance, les facteurs plasmatiques, les cytokines ou encore les protéines de fusion.

2.1.1. Les hormones

Ce groupe met en évidence la distinction à faire entre les protéines naturelles extraites de la matière biologique et les protéines dites recombinantes (éléments composés de matière biologique) produites par des moyens biotechnologiques.

Parmi les hormones, l'exemple de l'insuline est particulièrement intéressant. Historiquement, l'insuline était extraite de matière biologique, à savoir de pancréas de bœuf ou de porc. Des procédés de purification de l'insuline ont donc fait l'objet de demandes de brevet. Pour exemple, un brevet de 1971 déposé par la société Eli Lilly (FR 2 045 435) revendique :

Procédé pour purifier l'insuline d'un extrait pancréatique aqueux contenant de l'insuline ou d'une autre solution aqueuse contenant de l'insuline, qui est caractérisé par le fait qu'on ajuste l'alcalinité de la solution contenant l'insuline à un pH d'environ 7,2 à environ 10, 0 et la concentration en cation métal alcalin ou ammonium S à une valeur d'environ 0,2 à environ 1,0 M du même cation, provoquant ainsi la cristallisation de l'insuline du métal alcalin ou d'ammonium.

Les technologies évoluant, la source d'insuline, passe d'un produit naturel extrait à un produit synthétique. Pour preuve, le brevet FR 2 047 763 au nom de Hoechst AG, vise à protéger un procédé de préparation de peptides arginiques (comme l'insuline) au moyen de synthèse chimique.

Dès la fin des années 80, apparaissent les premiers brevets visant à protéger la production d'insuline recombinante, et notamment dans des bactéries. Le brevet FR 2 422 717 au nom de Genentech revendique par exemple :

Plasmide recombinant amélioré permettant la transformation d'un hôte bactérien et son utilisation comme vecteur de clonage, caractérisé en ce que le plasmide comprend: (a) un régulon homologue à l'hôte bactérien dans son état non transformé; et (b) en phase de lecture avec le régulon, un ADN inséré codant la séquence d'acides-amino d'un polypeptide hétérologue, de sorte que les bactéries transformées par le plasmide peuvent exprimer ladite séquence d'acides-amino sous forme récupérable.

Le polypeptide hétérologue pouvant être en outre de la pré-pro insuline humaine, de la pro insuline humaine, la chaîne A ou B de l'insuline humaine.

L'insuline passe donc d'un statut de biomédicament extrait d'une matière biologique à un statut de biomédicament produit à partir de procédés biotechnologiques. Mais les deux formes n'en sont pas moins brevetables, la première (même si l'insuline préexiste à l'état naturel) car l'extraction d'insuline à partir de matière biologique met en œuvre un procédé technique et la seconde parce qu'elle fait appel à un procédé biotechnologique (la production de protéines recombinantes).

2.1.2. Les facteurs de croissance

Dans le groupe des facteurs de croissance, le paysage brevets autour de l'érythropoïétine reproduit sensiblement celui de l'insuline. En effet, en premier lieu les brevets revendiquent de l'érythropoïétine naturelle extraite de matière biologique, à l'image du brevet EP 0 209 539 au nom de Genetics Institute Inc, déposé en 1985 qui revendique un procédé pour purifier l'érythropoïétine d'origine naturelle, caractérisée par un poids moléculaire défini et une activité spécifique.

Dans un second temps, des brevets revendiquant de l'érythropoïétine recombinante sont déposés à l'image du brevet EP 0 902 085 (au nom de Aventis Pharma GMBH) :

Polypeptide ayant la conformation de structure primaire de l'érythropoïétine, ledit polypeptide étant le produit de l'expression eucaryote d'une séquence d'ADN exogène utilisant un système d'expression BPV-1.

2.1.3. Les facteurs plasmatiques

Le biomédicaments constitués de facteurs plasmatiques sont appelés les médicaments dérivés du plasma sanguin (MDPS). Ce sont principalement des immunoglobulines polyvalentes (IgIV) de l'albumine et des facteurs plasmatiques de la coagulation.

Un facteur plasmatique extrait à partir de plasma sanguin peut être brevetable à condition que son isolement résulte de la mise en œuvre d'un procédé technique et ce même si cette protéine préexistait dans la nature (plasma humain). Pour exemple, le brevet EP 1 718 673 (déposé par le LFB) revendique un procédé de purification de l'albumine, des compositions d'albumine et une utilisation pour la stabilisation de produits thérapeutiques :

1. Procédé de purification d'albumine caractérisé en ce qu'il comprend une étape consistant à soumettre une solution aqueuse d'albumine, de concentration 15 g/l à 80 g/l et de pH non inférieur à 7, à une nanofiltration dans une plage de température allant de 15°C à 55°C.

18. Composition d'albumine à usage thérapeutique obtenue par un traitement, selon la revendication 14, d'adaptation d'une solution aqueuse d'albumine, selon l'une quelconque des revendications 15 à 17, à un usage clinique.

19. Utilisation d'une composition d'albumine à usage thérapeutique selon la revendication 18, pour la stabilisation d'au moins un membre choisi dans le groupe constitué par les protéines de faible concentration et d'activité spécifique élevée, les immunoglobulines spécifiques, les anticorps monoclonaux, les vaccins, les allergènes, les cytokines et les hormones peptidiques.

Le procédé mis en œuvre pour isoler une telle protéine est comme indiqué ci-dessus également brevetable.

Dans la catégorie des biomédicaments résultant d'une extraction de matière biologique, on peut citer par exemple le brevet FR 0 506 440 au nom du LFB qui revendique un procédé de séparation des protéines fibrinogènes, facteur XIII et colle biologique mais également les concentrés de ces protéines obtenues par le procédé initialement revendiqué.

Comme pour d'autres types de protéines déjà évoquées, les facteurs plasmatiques produits par la mise en œuvre de procédé de biotechnologies sont également protégés par des brevets. Ils peuvent être de nature recombinante mais aussi transgénique.

Pour exemple, le facteur VII est une protéine constitutive du sang humain, protéine qui peut être un produit naturel issu du plasma ou un produit obtenu par recombinaison ou par transgénèse. Le brevet FR 2 904 558 au nom du LFB protège en effet des compositions de FVII recombinants ou transgéniques possédant un profil de glycosylation particulier :

Composition de facteur VII recombinant ou transgénique, chaque molécule de facteur VII de la composition comportant des formes glycaniques liées aux sites de N-glycosylation, caractérisé en ce que, parmi toutes les molécules de facteur VII de ladite composition, les formes glycaniques biantennées, bisialylées et non fucosylées sont majoritaires par rapport à toutes les formes glycaniques liées aux sites de N-glycosylation du facteur VII de la composition.

2.1.4. Les cytokines

Parmi le groupe des cytokines, les interférons et les TNF (Tumor Necrosis Factors) ont fait l'objet de nombreux brevets. On peut citer comme exemple le brevet US 7 615 615 au nom de Merck qui revendique :

A modified human interferon beta (INF β) which is less immunogenic than human INF β (SEQ ID NO: 1) when administered in vivo to a human; wherein the modified human INF β comprises an amino acid residue sequence that differs from SEQ ID NO: 1 by an amino acid residue substitution selected from the group consisting of L57A, L57C, L57D, L57E, L57G, L57H, L57K, L57N, L57P, L57Q, L57R, L57S, and L57T and an additional substitution selected from the group consisting of the H140A, H140C, H140G, and H140P.

Ou encore, le brevet EP 0 536 520 au nom de Yeda qui revendique :

Utilisation d'interféron- β 2A humain pour préparer un médicament pour le traitement de maladies dans les domaines d'inflammation et/ou de réponse phase.

Le premier exemple illustre la protection d'un INF β (matière biologique produite par un procédé biotechnologique) dont la séquence en acides aminés a été modifiée pour diminuer son immunogénicité. Le second exemple illustre la protection d'une indication, à savoir, l'utilisation de l'INF β humain dans le traitement des maladies inflammatoires.

2.1.5. Les protéines de fusion

Comme toute protéine, une protéine de fusion est également brevetable. Pour exemple, le brevet EP 0 227 938 au nom de Hoechst revendique :

Protéine de fusion, caractérisée par un segment C-ou N-terminal qui correspond essentiellement aux 100 premiers aminoacides de l'interleukine-2, mais ne présente pas d'activité d'interleukine-2.

Par ailleurs, toute protéine (ou tout autre élément brevetable) peut être protégée en tant que telle ou par son utilisation dans le même brevet ou dans un brevet distinct.

Dans l'exemple suivant, le brevet EP 0 489 116 au nom d'Immunex Corp illustre une revendication dite de produit (la protéine de fusion) :

Protéine de fusion ayant une formule choisie dans le groupe formé par R1-R2, R2-R1, R1-L-R2 et R2-L-R1 dans lequel R1 représente GM-CSF; R2 représente IL-3; et L représente une séquence de peptide de liaison.

et une revendication dite d'utilisation :

Utilisation d'une protéine de fusion selon l'une quelconque des revendications 1 à 6 en préparant un médicament pour réguler les réponses immunitaires chez un être humain.

2.2. Les vaccins et anticorps

2.2.1. Les anticorps

A l'image des biomédicaments détaillés dans le groupe peptides/protéines, les anticorps peuvent être des biomédicaments extraits de matière biologique ou produits par des procédés de biotechnologies.

Le brevet FR 2 899 111 au nom du LFB illustre le cas des anticorps extraits de matière biologique en ce qu'il revendique :

Concentré d'immunoglobulines spécifiques du virus chikungunya en tant que médicament.

Ce brevet revendique également le procédé de préparation d'un tel concentré de la manière suivante :

Procédé de préparation d'un concentré selon la revendication 1 ou 2, caractérisé par les étapes suivantes : - constitution d'un lot d'au moins 1000 dons de plasma, chaque don présentant un titre suffisant en Ig anti chikungunya - précipitation des contaminants lipidiques et protéiques en une seule étape - récupération du concentré d'Ig dans le surnageant.

Les anticorps peuvent aussi être produits par recombinaison ou par transgénèse. Quelque soit leur mode de production, ils peuvent être protégés par différents types de revendications.

Revendication de type 1 : *Anticorps dirigé contre l'actine du myocarde, caractérisé en ce que l'anticorps ne réagit pas avec l'actine du muscle squelettique (EP 0 820 471, au nom de PROGEN BIOTECHNIK GMBH).*

Ce type de revendication permet de protéger n'importe quel anticorps dirigé contre l'actine du myocarde quelque soit sa séquence, sa nature (naturel, recombinant ou transgénique) et quelque soit le procédé utilisé pour le produire, sous réserve qu'il ne reconnaisse pas l'actine du muscle squelettique.

Revendication de type 2 : *Anticorps monoclonal dirigé contre l'antigène CD20, caractérisé en ce que la région variable de chacune de ses chaînes légères est codée par une séquence possédant au moins 70% d'identité avec la séquence d'acide nucléique murine SEQ ID NO: 5, la région variable de chacune de ses chaînes lourdes est codée par une séquence possédant au moins 70% d'identité avec la séquence d'acide nucléique murine SEQ ID NO: 7, et les régions constantes de ses chaînes légères et de ses chaînes lourdes sont des régions constantes provenant d'une espèce non-murine (WO 2006 064 121- au nom du LFB).*

Ce type de revendication permet de protéger plusieurs anticorps anti-CD20 présentant tous au moins 70% d'homologie avec une séquence définie codant les chaînes légères (séquence n°5) et au moins 70% d'homologie avec une séquence définie codant les chaînes lourdes (séquence n°7). La revendication de type 2 a une portée moins large que la revendication de type 1 en ce sens qu'elle ne couvre pas n'importe quel anticorps anti-CD20.

Revendication de type 3 : *Anticorps monoclonal qui est sécrété par la lignée cellulaire ayant un numéro de dépôt ATCC PTA-6652 ou PTA-6653 (brevet EP 1 745 129 au nom de IDEXX LAB Inc).*

Cette revendication définit un anticorps par un numéro de dépôt à l'ATCC (American Type Culture Collection). En conséquence, la protection est octroyée uniquement pour l'anticorps produit par cette lignée cellulaire.

Cependant, si un brevet A revendique l'utilisation d'un anticorps de séquence S1, dirigé contre une cible A1 et utilisé dans la préparation d'un médicament destiné au traitement d'une pathologie P3, il sera difficile pour une société B de défendre l'activité inventive d'un brevet B revendiquant l'utilisation d'un anticorps de séquence S2, dirigé contre la même cible que l'anticorps de séquence S1 et indiqué dans le traitement de la même pathologie P3.

Les anticorps peuvent aussi être définis par leur structure et notamment par leur profil de glycosylation. Pour exemple le brevet EP 1 272 527 au nom du LFB revendique :

Anticorps monoclonal caractérisé en ce qu'il possède sur son site de glycosylation (Asn 297) du Fcγ, des structures glycaniques de type biantennées, avec des chaînes courtes, une faible sialylation, des mannoses terminaux et/ou des GlcNAc terminaux non intercalaires.

Ce type de revendication couvre l'ensemble des anticorps quelle que soit leur cible et quel que soit leur mode de production, possédant le profil glycanique spécifié.

Revendication de type 4 : *A therapeutic antibody comprising a VH chain having the sequence set forth in SEQ ID No:26 and a VL domain having the sequence set forth in SEQ ID No:32. (brevet EP 1 996 621 au nom de Glaxo Group LTD).*

Ce type de revendication limite la protection à un anticorps particulier défini par les séquences des chaînes lourdes et légères. La portée de ce type de revendication est la même que celle de type 3 où l'anticorps est défini par l'hybridome qui le produit.

2.2.2. Les vaccins

Dans le domaine des vaccins, les revendications peuvent être rédigées en faisant apparaître clairement le mot « vaccin ». Pour exemple, le brevet EP 1 439 856 au nom de l'Université de Jefferson revendique :

Un vaccin antirabique vivant comprenant un génome de virus rabique recombinant, dans lequel le génome rabique recombinant comprend au moins deux gènes de protéine G.

Une revendication de vaccin peut aussi être formulée à l'aide des termes « composé immunogène » ou « composition immunogène ». Pour exemple, le brevet EP 1 629 004 au nom de Wyeth Corp revendique :

Composition immunogène comprenant une population de particules du réplicon de virus (VRP) de l'encéphalite équine vénézuélienne comprenant des gènes de réplicase du virus de l'encéphalite équine vénézuélienne, des protéines de réplicase du virus de l'encéphalite équine vénézuélienne, une glycoprotéine E1 du virus de l'encéphalite équine vénézuélienne, une glycoprotéine E2 du virus de l'encéphalite équine vénézuélienne, un gène de glycoprotéine F du virus parainfluenza de type 3 et un gène de glycoprotéine HN du virus parainfluenza de type 3.

La terminologie « vaccin » n'apparaît pas dans le texte de cette revendication. La terminologie « composé immunogène » la remplace tout en ayant une portée plus large. Cet exemple permet de souligner l'importance du choix et du croisement des mots clés pour toute recherche de documents dans un domaine donné.

La notion de vaccin est très couramment associée à la prévention d'une pathologie liée à un agent pathogène. Or, se développent depuis quelques années, des vaccins anticancéreux. Pour exemple, le brevet EP 1 465 658 au nom de Bio Life Science Forschungs revendique :

Médicament contenant une substance agissant comme antigène tumoral en association avec une substance à activité d'antiacide gastrique, lesdites substances étant destinées, séparément ou ensemble, simultanément ou successivement, à une application orale pour effectuer une vaccination contre des cancers et/ou des tumeurs, l'antigène en tant que substance active étant un mimotope d'antigène avec la séquence d'acides aminés SEQ.IDNO :11: Gln-Met-Trp-Ala-Pro-Gln-Trp-Gly-Pro-Asp.

Cet exemple permet également de montrer qu'une revendication de vaccin peut être formulée sous la forme « médicament.... destiné à une vaccination contre... ».

Enfin, le brevet EP 0 862 634 au nom de Transgène revendique :

[Composition pharmaceutique destinée au traitement ou à la prévention d'une infection ou tumeur à papillomavirus qui comprend ...]

La rédaction sous la forme « composition pharmaceutique destinée à la prévention » doit être interprétée à l'image du mot « vaccin ».

2.3. Les acides nucléiques

Pour rappel, la directive 98/44/CE précise qu'une séquence ou une séquence partielle d'un gène est brevetable à condition que l'application de cette séquence soit concrètement exposée dans la demande de brevet (Cf. 1.2.5. La Directive 98/44/CE relative à la brevetabilité du vivant).

Pour exemple, le brevet EP 1 543 127 au nom du Alfred Wegener Institute for Polar and Marine Research présente une revendication 1 comme suit :

*Séquence nucléotidique codant pour des enzymes protéolytiques sous forme de protéases particulières, caractérisée en ce que ladite séquence nucléotidique est issue de la diatomée marine psychrotolérante *Fragilariopsis cylindrus* et code pour une protéase calpaïne-7 selon la SEQ ID n° 1.*

Dans le même brevet, une autre revendication est la suivante :

Utilisation de la séquence nucléotidique selon la revendication 3 pour exprimer ou surexprimer l'enzyme protéase calpaïne-7 dans des organismes hôtes.

Ce brevet répond clairement à la Directive Européenne 98/44/CE en ce sens qu'il revendique une séquence d'acide nucléique et son application, à savoir l'expression ou la surexpression de l'enzyme protéase calpaïne 7.

Dans les brevets français, les revendications basées sur des acides nucléiques peuvent se présenter sous les formes suivantes :

Le brevet FR 2 930 152 au nom du CNRS revendique : *Composition pharmaceutique ou cosmétique comprenant au moins un acide nucléique comprenant une séquence capable de s'hybrider spécifiquement avec un gène ou un ARNm codant pour une kinésine interagissant avec le complexe adaptateur AP-1 et de diminuer ou supprimer l'expression de cette protéine.*

Le brevet FR 2 861 596 au nom du CNRS et de l'ENS revendique : *Utilisation d'un acide nucléique, simple brin ou double brin, comprenant au moins un résidu éthénocytosine, dans la préparation d'un médicament destiné au traitement des affections liées à l'expression de l'alkyl-N-purine-ADN glycosylase (ANPG) dans les cellules.*

Le droit français diverge du droit européen en ce sens qu'il ne permet pas la brevetabilité d'une séquence en tant que telle. Il faut noter qu'il existe des brevets français revendiquant des séquences nucléotidiques, mais la portée de ce type de revendication sera limitée à l'utilisation de la séquence telle que décrite dans la description.

Enfin, des brevets revendiquent aussi des oligonucléotides de type morpholino, comme dans le brevet EP 1 155 140 co-déposé par le CNRS et le CEA :

*Procédé de fabrication d'un fragment d'acide nucléique (ADN ou ARN) marqué en 3', qui comprend l'incorporation enzymatique d'un dérivé de nucléotide ayant pour précurseur un composé de formule : dans laquelle R**1 représente une base nucléique et R**2 représente un groupe répondant à l'une des formules suivantes : $-(CH_2)_n-NH_2$ $-(CH_2)_n-SH$ $-(CH_2)_n-COOH$ $-(CH_2)_n-OH$ $-(CH_2)_n-NH-R^{**3}$ $-(CH_2)_n-SR^{**3}$ $-(CH_2)_n-CO-R^{**3}$ $-(CH_2)_n-OR^{**3}$ dans lesquelles n est un nombre entier allant de 1 à 12 et R**3 est un groupe dérivé d'un marqueur, d'une protéine, d'une enzyme, d'un acide gras ou d'un peptide, à l'extrémité 3' OH du fragment d'acide nucléique.*

*Morpholino-nucléotide répondant à la formule : dans laquelle R**1 est l'adénine et R**2 représente $-CH_2-COOH$, $-(CH_2)_4-NH_2$ ou $-(CH_2)_4-NH-R^{**3}$ avec R**3 représentant un groupe dérivé de la fluorescéine.*

2.4. Les thérapies cellulaire et génique

Dans le domaine de la thérapie génique, des revendications de vecteurs ou d'utilisation de ces vecteurs sont couramment retrouvées. Pour exemple, le brevet EP 1 348 030 revendique :

Utilisation d'un vecteur comprenant les séquences de nucléotides définies par les positions 30811-31788 et 18254-21100 dans SEQ ID N° 1 pour la production d'un médicament à utiliser dans la thérapie génique.

Des procédés ou utilisations visant à introduire une séquence nucléotidique dans une cellule sont également brevetables, à l'image du brevet EP 1 136 083 au nom de Anges MG Inc qui revendique :

Utilisation d'un gène du facteur de croissance des hépatocytes (HGF) pour la production d'un agent thérapeutique.... utilisé pour administrer directement le gène du HGF à une partie affectée du muscle cardiaque en utilisant une assistance échocardiographique sans incision de la partie affectée ou thoracotomie.

Dans le domaine de la thérapie cellulaire, les différents types de brevets, produits, procédés, utilisations peuvent être retrouvés. Pour exemple, le brevet WO 2008 145 866 au nom de LFB Biotechnologies présente la revendication suivante :

Ensemble de moyens pour le traitement d'une pathologie maligne, d'une maladie auto-immune ou d'une maladie infectieuse, comprenant une cellule effectrice qui exprime le récepteur Fcγ RIII (CD16) à sa surface, et un anticorps monoclonal, dans lequel l'affinité de la région Fc dudit anticorps monoclonal pour le CD16 est supérieure à l'affinité de la région Fc des immunoglobulines polyclonales pour le CD16.

Ce type de revendication permet d'illustrer un autre concept important dans le domaine des biomédicaments, à savoir que les méthodes de traitement chirurgical ou thérapeutique et les méthodes de diagnostic ne sont pas brevetables en Europe. Cette disposition a pour but de ne pas entraver la liberté d'action des médecins. En effet, une revendication de méthode de traitement d'une maladie A empêcherait toute personne n'ayant pas de droit sur cette technologie de soigner un patient atteint de la maladie A avec la méthode protégée. En revanche, le produit ou la composition pharmaceutique relevant de ce traitement peut bien entendu être brevetable si elle répond aux critères de

brevetabilité. C'est pourquoi, en Europe, les revendications sont écrites sous la forme « utilisation.... pour le préparation d'un médicament destiné à » ou « composition pour le traitement ».

Enfin, la thérapie cellulaire introduit la question de la brevetabilité des cellules souches. L'exemple ci-après, issu du brevet EP 1 694 354 au nom de DEVELOGEN AG introduit de lui-même la limitation concernant la brevetabilité des cellules souches.

Utilisation d'un produit de type neurture pour stimuler et/ou induire la différenciation de cellules produisant de l'insuline à partir de cellules progénitrices, en particulier de cellules souches à l'exception des cellules souches embryonnaires humaines, par exemple de cellules souches somatiques, in vitro,
...

En effet, la question de la brevetabilité des cellules souches fait débat. A ce jour, le droit européen précise qu'un brevet ne peut être délivré pour une invention biotechnologique qui a pour objet l'utilisation d'embryons humains à des fins industrielles ou commerciales. Par ailleurs, une décision de jurisprudence rejette la brevetabilité de produits ou de méthodes si cela nécessite la destruction d'un embryon. Il en ressort que les cellules souches embryonnaires humaines ne sont pas brevetables, encore faudrait-il qu'une définition claire du mot « embryon » soit énoncée. Cependant, les cellules souches embryonnaires non humaines et les cellules souches humaines non embryonnaires de type somatique restent des éléments brevetables au sens du droit européen et du droit français.

3. EXEMPLES DE BREVETS PROTEGEANT LES DIFFERENTS SYSTEMES D'EXPRESSION DES BIOMEDICAMENTS

L'Organisation de Coopération et de Développement Economiques (OCDE) définit les biotechnologies comme « l'application des sciences et des techniques à des organismes vivants, qu'il s'agisse d'éléments ou de produits pour transformer les matériaux vivants ou non, dans le but de produire des connaissances, des biens et des services ». Autrement dit, les médicaments issus des biotechnologies vont être principalement issus de la production de protéines recombinantes par des organismes vivants modifiés génétiquement tels que des bactéries, des levures, et des cellules végétales ou animales. Pour produire un biomédicament, plusieurs étapes sont nécessaires. Dans la partie suivante, nous nous intéresserons uniquement à l'étape d'Upstream (chapitre 2a), les étapes de Downstream (chapitre 2b) et de Mise Sous Forme Pharmaceutique (chapitre 3) ne seront donc pas abordées.

Un gène d'intérêt thérapeutique va être isolé puis être inséré dans un vecteur d'expression (virus, plasmide...), dans le but d'être produit par une cellule ou un organisme génétique modifié. Dans le cas d'une cellule, on parlera de cellule usine (Chapitre 1.3.), celle-ci exprimant la protéine recombinante. Parmi les technologies de bioproduction développées (chapitre 1.4), il existe les systèmes d'expression en bactéries, en levures, en cellules d'insectes, d'oiseaux, de mammifères et même dans des animaux et plantes transgéniques. C'est le biomédicament qui conditionnera le choix du système d'expression. Par exemple, pour synthétiser des protéines non-

glycosylées telle que l'insuline, des bactéries comme *Escherichia coli* conviendront parfaitement.

Comme pour la partie précédente, au travers d'exemple de revendications, les paragraphes ci-après visent à illustrer la brevetabilité des systèmes de production des biomédicaments et les différents types de protection qu'il est possible d'obtenir.

3.1. La bioproduction en cellules bactériennes

Les systèmes d'expression bactériens ont été largement employés dans les années 1980, pour la synthèse de petits polypeptides du type interleukine, interféron... Par exemple, le brevet FR 2 635 113 au nom de Sanofi SA revendique la culture de cellules d'*Escherichia coli* modifiées pour produire des protéines recombinantes d'interleukine-1 :

Procédé pour l'obtention d'IL-1 p mature recombinante, caractérisé en ce qu'il consiste à cultiver des cellules d'E. coli transformées par un gène codant pour ladite protéine, à soumettre les cellules d'E. coli à un choc osmotique et à séparer la protéine recombinante du surnageant du choc osmotique.

Les plasmides d'expression des systèmes bactériens font également l'objet d'une protection, à l'image du brevet EP 1 022 339 au nom de Bayer AG qui revendique :

Vecteur pour la production de IL-4 et de mutéines de IL-4 dans une souche d'Escherichia coli, comprenant, dans l'ordre 5' à 3', les éléments liés de manière fonctionnelle suivants : un promoteur régulable consistant en le promoteur de phage T5 de E. coli et deux séquences d'opérateur lac, un site de liaison ribosomale provenant de g10 du phage T7 de E. coli, un codon d'initiation de traduction, un gène structural pour la IL-4 ou une mutéine de IL-4 et, en aval de ce gène structural, un terminateur de transcription.

Dans certains brevets, comme le brevet EP 0 043 980 au nom de Genentech Inc, tout le système de bioproduction est protégé par un seul et même brevet : sont revendiqués les polypeptides (interférons pour l'exemple), les compositions pharmaceutiques, le procédé d'obtention, les cellules bactériennes et les véhicules d'expression microbiens.

3.2. La bioproduction en levures

Comme pour la bactérie *Escherichia coli*, les connaissances génétiques des levures en font un système d'expression bien connu pour la bioproduction. Cependant, à la différence des cellules bactériennes, les levures peuvent réaliser des modifications post-traductionnelles, essentielles pour de nombreux biomédicaments. Elles peuvent donc synthétiser des protéines dont la maturation *in vivo* est plus complexe. Par exemple, le brevet EP 0 655 503 au nom de Green Cross Corp revendique un procédé d'obtention d'une protéine recombinante en cellules de levure :

Procédé pour la production de sérum-albumine humaine recombinante comprenant la culture d'une souche de levure productrice de sérum-albumine humaine à une température de 21 à 25 DEG.C.

Dans le même brevet, une autre revendication précise que la souche de levure est dérivé de *Saccharomyces cerevisiae* :

*Procédé selon la revendication 1, dans lequel l'hôte produisant la sérum-albumine humaine est dérivé de *Saccharomyces cerevisiae* AH 22.*

Le brevet EP 1 109 922 au nom de Novo Nordisk revendique quant à lui la transformation de la levure *Saccharomyces cerevisiae* pour réaliser de la bioproduction à une échelle industrielle :

*Un procédé industriel de fermentation pour produire un produit hétérologue, comprenant (a) la mise en culture dans des conditions industrielles d'une souche de levure *Saccharomyces* Crabtree-négative existant à l'état naturel qui comprend un plasmide ou un ADN codant pour le produit hétérologue, où la souche utilise le glucose plus efficacement et présente un métabolisme de fermentation moindre par rapport à celui observé pour *Saccharomyces cerevisiae* dans les mêmes conditions de croissance, et (b) récupérer le produit hétérologue à partir du milieu de culture.*

Les levures du genre *Pichia* peuvent également être utilisées dans le même but, le brevet EP 0 226 752 au nom de Phillips Petroleum Co revendique un procédé de production de polypeptides basé sur un tel système.

Les vecteurs disponibles pour les systèmes d'expression en levures comprennent des plasmides, comme illustré par le brevet EP 0 312 159 déposé par Merck & Co Inc :

*A plasmid expression vector containing yeast-derived sequences for the selection and amplification of the plasmid in a species of yeast derived from the families *Saccharomycetaceae* or *Cryptococcaceae*, [...].*

Aussi, comme pour les bactéries, des brevets revendiquent des méthodes de culture des levures. C'est le cas du brevet EP 0 706 562 au nom de Merck & Co Inc :

*A medium for the growth of *Saccharomyces cerevisiae* which comprises per liter: (NH₄)₂SO₄ 10 g, KH₂PO₄ 10 g, CaCl₂p2H₂O 0.5 g, NaCl 0.5 g [...].*

3.3. La bioproduction dans les microalgues

Un autre système d'expression porte sur l'utilisation de microalgues pour la synthèse de protéines glycosylées. Bien que le développement de cette technologie soit actuellement moins avancé, elle a fait l'objet de plusieurs dépôts de brevets, dont le brevet EP 2 090 648, co-déposé par l'IFREMER, le CNRS et l'université de Rouen. Celui-ci vise à la fois la protection des microalgues transformées, et des méthodes de production :

Transformed microalgae comprising a nucleotide sequence operably linked to a promoter that drives expression in said microalgae, wherein said nucleotide sequence encodes a glycosylated polypeptide that is expressed in the transformed microalgae.

A method for producing at least one glycosylated polypeptide, comprising transforming microalgae or transformed microalgae according to anyone of claim 15 to 21 with a nucleotide sequence operably linked to a promoter that drives expression in said microalgae, wherein said nucleotide sequence encodes a glycosylated polypeptide that is expressed in the transformed microalgae.

3.4. La bioproduction en cellules d'insectes

Les cellules d'insectes sont impliquées dans de nombreuses voies de recherche comme par exemple les anticorps murins, l'antigène VIH, l'EPO. Le vaccin Cervarix® contre certaines souches du Papillomavirus Humain est par exemple produit dans des cellules d'insectes. Comme pour tous les systèmes d'expression, les lignées cellulaires vont être protégées par des brevets, par exemple une lignée de *Spodoptera frugiperda* dans le brevet EP 1 119 612 au nom de Protein Sciences Corp :

Insect cell line Sf900+ (ATCC CRL-12579).

À ceux-ci sont associés d'autres protections portant sur les procédés de bioproduction, comme proposé par le brevet WO9426087 au nom de US of America :

A process for production of a recombinant polypeptide comprising the steps of (a) selecting insect cells for culture; (b) transforming the insect cells to include a selected DNA sequence encoding a selected polypeptide; and (c) culturing the insect cells in a horizontally rotating culture vessel modulated to create low shear conditions during which the insect cells recombinantly produce the selected polypeptide.

Bien entendu, pour les cellules d'insectes un vecteur d'expression est également nécessaire. A titre d'exemple, le brevet EP 0 127 839 au nom de The Texas A&M University a pour objectif de protéger la mise au point d'un vecteur baculoviral pour la production de protéines recombinantes :

A method for producing a recombinant baculovirus expression vector, capable of expressing a selected gene or portion thereof in a host insect cell [...].

Aussi, le brevet EP 0 833 933 déposé par MG PMC Llc revendique précisément le baculovirus pour la production en cellules d'insectes d'hémagglutinines utilisées comme vaccin :

Vecteur pour la préparation d'une protéine hémagglutinine HAO grippale recombinante comprenant les séquences 5'→3' suivantes : un promoteur polyhédrine provenant d'un baculovirus, un codon d'initiation de traduction ATG, un peptide signal, des séquences codant pour l'hémagglutinine mature provenant d'une souche de virus grippal, un codon de terminaison de traduction, et un signal de polyadénylation d'ARN de polyhédrine, caractérisé en ce que le peptide signal comprend un peptide signal de baculovirus comprenant les aminoacides 1 à 18 de la SEQ ID No7.

3.5. La bioproduction en cellules d'oiseaux

Les cellules d'oiseaux font également l'objet d'une protection par un brevet. Par exemple, le brevet EP 1 446 004 et dont le déposant est Vivalis, revendique le procédé d'obtention de cellules aviaires transgéniques :

Procédé d'obtention d'une cellule souche embryonnaire (ES) aviaire modifiée par recombinaison homologue, ledit procédé comprenant les étapes suivantes : A) l'introduction d'un vecteur de recombinaison homologue dans ladite cellule ES aviaire par une méthode de transfection ; B) la sélection des cellules par l'addition d'un agent de sélection dans le milieu de culture ; et C) le criblage des clones résistants et amplification [...].

Comme illustré par le brevet FR 2 884 255 au nom de Vivalis, les lignées ainsi produites sont à leur tour protégées :

Lignées de cellules aviaires EBx diploïdes, non-tumorigènes, immortelles et non transformées, caractérisées en ce que lesdites cellules présentent: une morphologie ronde et compacte de cellules souches embryonnaires aviaires et un rapport nucléo-cytoplasmique élevé ; un temps de doublement de sa population comprise entre 15 et 30 heures [...].

Les utilisations potentielles des lignées cellulaires vont également être protégées, comme dans le brevet EP 1 685 243 déposé par Probiogen AG :

Utilisation de la lignée cellulaire selon l'une quelconque des revendications 1 à 7, pour la production de produits biologiques ou de virus.

Ou encore dans le brevet EP 1 446 004 au nom de Vivalis :

Oeuf susceptible d'être obtenu à partir d'un animal obtenu par un procédé selon l'une des revendications 18 à 20, caractérisé en ce qu'une partie de l'ovalbumine, ou du lysozyme est partiellement ou totalement remplacée par la protéine d'intérêt exogène.

3.6. La bioproduction en cellules de mammifères

Les cellules de mammifères produisant des protéines recombinantes ont pris beaucoup d'importance au sein des systèmes de bioproduction. En effet, ces lignées reproduisent fidèlement les modifications post-traductionnelles nécessaires au bon fonctionnement des biomédicaments. Le biomédicament produit peut être par exemple un anticorps, comme dans le brevet EP 1 167 537 au nom de Japan Tobacco Inc et Abgenix qui revendique des méthodes de production :

Méthode pour produire un anticorps monoclonal, dans laquelle la méthode comprend les étapes suivantes : (a) introduire dans une cellule in vitro un ADN exogène, ladite cellule (i) comprenant un gène de la chaîne lourde d'immunoglobuline endogène réarrangé et un gène de la chaîne légère d'immunoglobuline endogène réarrangé, (ii) sécrétant un anticorps monoclonal comprenant un polypeptide de la chaîne lourde d'immunoglobuline dérivé dudit gène de la chaîne lourde d'immunoglobuline endogène réarrangé et un polypeptide de la chaîne légère d'immunoglobuline dérivé dudit gène de la chaîne légère d'immunoglobuline endogène réarrangé,[...].

Comme pour les autres systèmes cellulaires, des brevets protègent spécifiquement les lignées cellulaires exploitées à des fins de bioproduction. Par, exemple, le brevet EP 0 216 846 au nom de Celltech Ltd revendique :

A myeloma cell-line transformed with a vector including a gene coding for a eukaryotic polypeptide and a viral promoter, such that expression occurs of the gene coding for the eukaryotic polypeptide, directed by the viral promoter, such that the eukaryotic polypeptide is produced at a level greater than 1 mg/L.

Sont associés à ces cellules usines, des systèmes de sélection et d'amplification. Par exemple, le système GS est l'un des plus utilisés, il porte sur le marqueur de sélection de

la glutamine synthétase (GS), et est revendiqué par le brevet EP 0 256 055 co-déposé par Celltech Ltd et l'Université de Glasgow :

A method for co-amplifying a recombinant DNA sequence which encodes the complete amino acid sequence of a desired protein other than a glutamine synthetase (GS), which method comprises: (a) providing a vector capable, in a transformant host cell, of expressing both a recombinant DNA sequence which encodes an active GS enzyme and the recombinant DNA sequence which encodes the complete amino acid sequence of the desired protein other than GS; (b) providing a eukaryotic host cell which is a glutamine prototroph; (c) transforming said host cell with said vector; and (d) culturing said host cell under conditions which allow transformants containing an amplified number of copies of the vector-derived GS-encoding recombinant DNA sequence to be selected, which transformants also contain an amplified number of copies of the desired protein-encoding DNA sequence.

Le brevet EP 0 731 845 au nom de Merck & Co Inc revendique un locus spécifique dans le génome d'une cellule murine qui favorise l'expression d'un gène recombinant après une intégration stable par recombinaison homologue :

Vecteur d'expression de recombinaison homologue pour l'expression de gènes recombinants dans des cellules de mammifères, caractérisé en ce que ledit vecteur comprend un promoteur pour l'expression d'un gène recombinant, une unité de transcription codant pour un marqueur de sélection et des séquences d'ADN spécifiques du locus de l'immunoglobuline murine gamma 2A pour un ciblage de la recombinaison homologue.

Certains brevets vont protéger uniquement des vecteurs d'expression spécifiques, comme dans le brevet EP 0 302 429 déposé par Ono Pharmaceutical Co (vecteur d'expression pour l'interleukine humaine 4).

Pour les vecteurs d'expression non-viraux, il est possible de faire pénétrer ceux-ci dans les cellules et leurs noyaux, à l'aide de plusieurs technologies. Par exemple, d'un côté le brevet EP 0 789 564 au nom de Supratek Pharma Inc est un exemple de protection d'une composition polynucléotidique pour la transfection chimique :

Composition de polynucléotide comprenant: (a) un polynucléotide ou un dérivé de polynucléotide; et (b) en dessous de 5% d'un copolymère séquencé de polyéther (p/v) comprenant un segment polymère de type A [...] et un segment polymère de type B [...]; où (i) la composition ne forme pas un gel; et (ii) la composition forme des micelles de 10 nm à 100 nm de diamètre.

D'un autre côté, le brevet EP 1 297 119 dont le déposant est Amaxa GmbH revendique un procédé de transfection physique à l'aide d'un champ électrique :

Procédé d'incorporation de molécules biologiquement actives dans le noyau cellulaire de cellules eucaryotiques supérieures à l'aide de courant électrique, l'incorporation dans le noyau de cellules primaires étant obtenue indépendamment de la division cellulaire au moyen d'une impulsion d'une force de champ de 2 à 10 kV/cm et d'une durée d'au moins 10 µs et d'une intensité de courant d'au moins 1 A.

3.7. La bioproduction par des animaux et plantes transgéniques

D'autres voies de production de médicaments issus des biotechnologies se basent sur l'utilisation d'organismes entiers, comme des animaux non-humains. Le brevet EP 0 741 515 au nom de GTC Biotherapeutics revendique en effet un procédé de production d'une protéine recombinante d'intérêt dans un mammifère non-humain, comme une chèvre :

Procédé pour l'obtention d'une immunoglobuline hétérologue et assemblée dans le lait d'un mammifère transgénique non humain caractérisé en ce qu'il comprend : (a) l'introduction dans une lignée germinale dudit mammifère d'ADN comprenant séparément les séquences codantes de protéines des chaînes lourdes et légères de ladite immunoglobuline [...] ; (b) l'obtention de lait comprenant ladite immunoglobuline hétérologue et assemblée à partir dudit mammifère non humain ; ladite immunoglobuline hétérologue et assemblée étant dans une configuration fonctionnelle et étant produite à une concentration d'au moins 1 mg/ml dans le lait dudit mammifère.

De même, des plantes transgéniques sont utilisées comme bioproduleurs, comme divulgué dans le brevet EP 0 577 598 déposé par l'Université de Californie :

Procédé pour la production d'une plante transgénique comprenant un gène d'intérêt qui est exempt de séquences de gène marqueur, ledit procédé comprenant: a) la fourniture d'une construction d'ADN comprenant le gène d'intérêt, des acides nucléiques auxiliaires étrangers comprenant des séquences de gène marqueur et un transposon fonctionnant dans les plantes [...].

Certains brevets sont spécifiques à une plante : le brevet EP 1 808 491 déposé par Meristem Therapeutics protège la production de molécules hétérologues dans des cellules de maïs :

A method for producing an heterologous molecule in maize seeds, comprising cultivation of a maize plant wherein cells contain stably introduced into their genome a first nucleic acid which specifically decreases the production of [alpha]-zein and/or a second nucleic acid which specifically decreases the production of [gamma]-zein, and a third nucleic acid which expresses the heterologous molecule, and obtaining a seed containing the heterologous molecule.

4. CONCLUSION

Depuis une quinzaine d'années, le secteur biopharmaceutique est devenu un secteur stratégique et incontournable pour la mise au point de produits thérapeutiques innovants. Associées à une meilleure connaissance des organismes vivants et une meilleure compréhension des pathologies, les biotechnologies ont été à l'origine d'une nouvelle génération de médicaments : les biomédicaments.

Progressivement un grand nombre de sociétés pharmaceutiques se sont orientées vers le développement de biomédicaments dont certains sont déjà qualifiés de blockbusters depuis quelques années (Figure 3). A l'exception du produit Enbrel®, les chiffres d'affaires les plus importants sont observés pour la classe des anticorps, suivie par les facteurs de croissance (protéines).

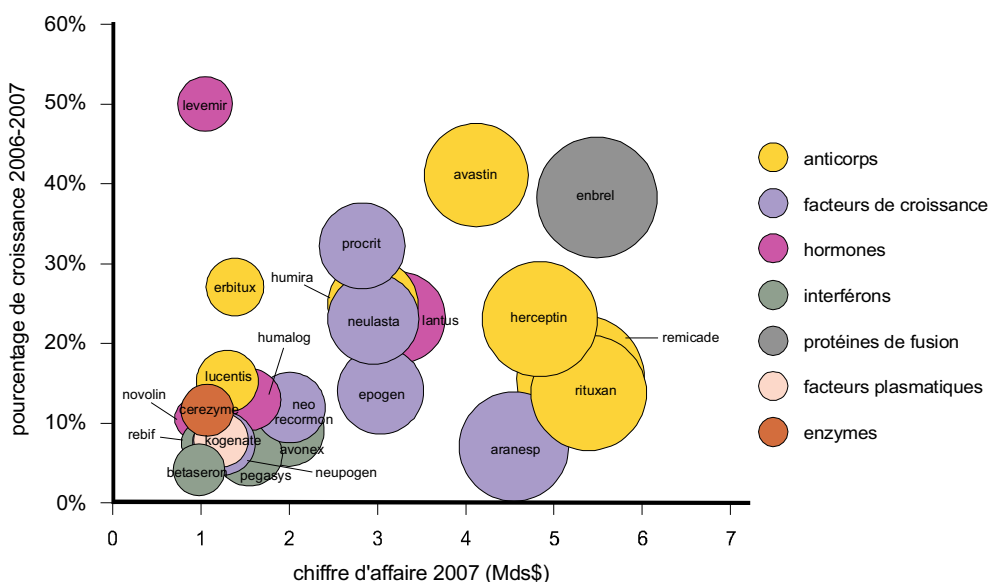


Figure 3 : Pourcentage de croissance et chiffre d'affaires de biomédicaments. La couleur indique les classes de biomédicaments. La taille du cercle représente leur part de marché (*Site internet des entreprises leaders, Développement & Conseil, 2008*).

25% des blockbusters actuels sont des biomédicaments. Par exemple, la protéine de fusion Enbrel® d'Amgen (traitement de l'arthrite rhumatoïde sévère), premier blockbuster des biomédicaments, a réalisé un chiffre d'affaires consolidé de plus de 5Mds\$ en 2007. Il en est de même pour les produits Herceptin® de Genentech, Rituxan® de Roche ou encore Remicade® de Shering-Plough.

Cependant, comme ce fut le cas pour les blockbusters issus de l'industrie chimique, les brevets protégeant certains biomédicaments arrivent bientôt à échéance : 2012 aux États-Unis et en Europe pour Enbrel®, 2013 aux États-Unis et en Europe pour Herceptin®. Au même titre que les médicaments issus de la chimie avec les génériques, les biomédicaments pourront être librement copiés à l'expiration des brevets correspondants et des médicaments biosimilaires arriveront sur le marché.

Le marché des médicaments biosimilaires est naissant (figure 4), le premier biosimilaire (l'Omnitrope®, hormone de croissance) a été approuvé aux États-Unis en 2005, et en Europe en 2006 mais les analystes estiment qu'environ 25 biosimilaires seront disponibles sur le marché en 2012 associant un chiffre d'affaire d'environ 3.5 Mds\$.

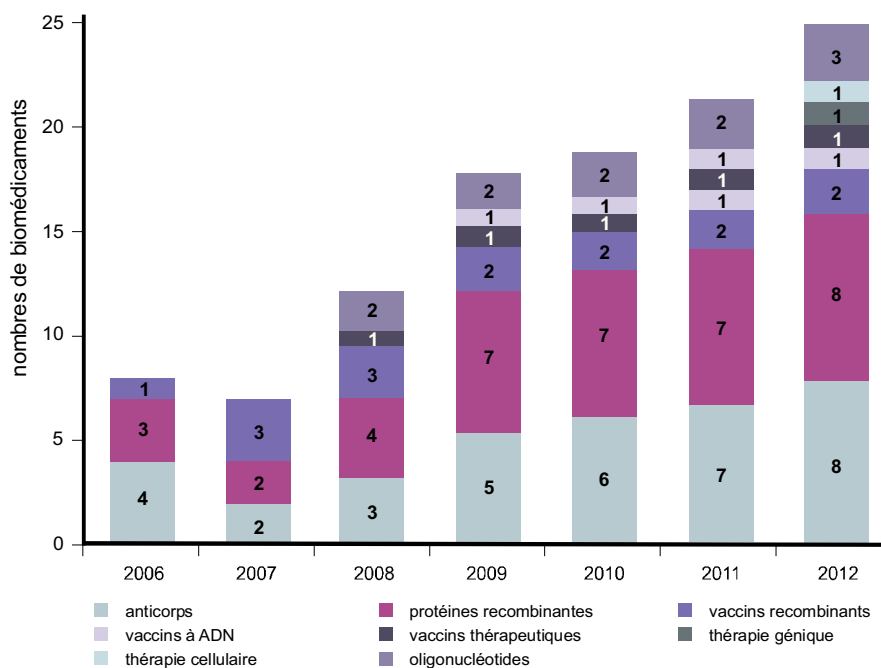


Figure 4 : Hypothèse haute du nombre de molécules biologiques qui sont et qui seront commercialisées entre 2006 et 2012 (Source : Développement & Conseil, 2008).

Il est à noter que la terminologie de « biosimilaire » est préférée à celle de biogénérique car la reproduction exacte du médicament « vivant » est techniquement impossible. En conséquence, l'homologation d'un biosimilaire se fait sur la base d'une équivalence de résultats thérapeutiques, basés sur des études cliniques englobant des centaines de patients, et non pas uniquement sur la base d'une équivalence pharmaceutique (comme c'est le cas pour les médicaments génériques).

Dès 2004, les autorités européennes mettent en place des règles propres aux biosimilaires, (Directive 2004/27/CE du 31 mars 2004) alors qu'aux États-Unis, la réglementation des biosimilaires n'est pas clairement établie. L'article 10-4 de cette Directive définit la notion de biosimilaire :

Les médicaments biologiques similaires à des médicaments de référence ne remplissent habituellement pas toutes les conditions pour être considérés comme des médicaments génériques, en raison notamment des caractéristiques des procédés de fabrication, des matières premières utilisées, des caractéristiques moléculaires et des modes d'action thérapeutiques. Lorsqu'un médicament biologique ne remplit pas toutes les conditions pour être considéré comme un

médicament générique, les résultats d'essais appropriés devraient être fournis afin de satisfaire aux conditions relatives à la sécurité (essais précliniques) ou à l'efficacité (essais cliniques), ou aux deux.

Le producteur d'un médicament biosimilaire doit prouver que son produit a le même effet thérapeutique que le produit biologique de référence. À ce titre, l'EMA (European Medicines Agency) a publié des lignes directrices pour la comparaison des biomédicaments en décembre 2003 et en septembre 2005 pour la bioéquivalence.

Une étude récente du Comité Biotechnologies du Leem et du Genopole® sur l'état des lieux de la bioproduction en France en 2008, estime que « la somme liée à la perte des brevets de biomédicaments est estimée à plus de 10Mds\$ en 2010 et pourrait atteindre 20Mds\$ en 2015 ». Elle cite à titre d'exemple que « les brevets sur les EPO expirent en moyenne aux USA vers 2013, ouvrant un marché potentiel de 8Mds\$ aux biosimilaires ».

Les évolutions du secteur pharmaceutique peuvent s'observer au travers des brevets. En effet, comme il a été précisé dans les parties précédentes, le biomédicament en tant que produit thérapeutique n'est pas le seul élément brevetable ; en effet, sont également protégeables par brevet, les procédés de production, les lignées cellulaires utilisées dans le cadre de ces procédés ou encore les méthodes de culture en fermenteur, les vecteurs d'expression, ou les méthodes d'extraction. L'arrivée des biomédicaments sur le marché et le fait qu'un biomédicament puisse être protégé par plusieurs brevets (brevets de produit, de procédé et d'utilisation) expliquent les observations faites en matière de dépôts de brevets liés aux biomédicaments.

La classification des brevets couvrant les biomédicaments fait apparaître 5 catégories (figure 5) : les peptides et protéine, les vaccins et tout élément pouvant servir à faire de l'immunothérapie comme les anticorps, les médicaments issus des thérapies génique et cellulaire, les acides nucléiques ou une substance provenant de la matière biologique (les procédés relatifs à chacun des différents éléments sont compris dans la catégorie correspondante).

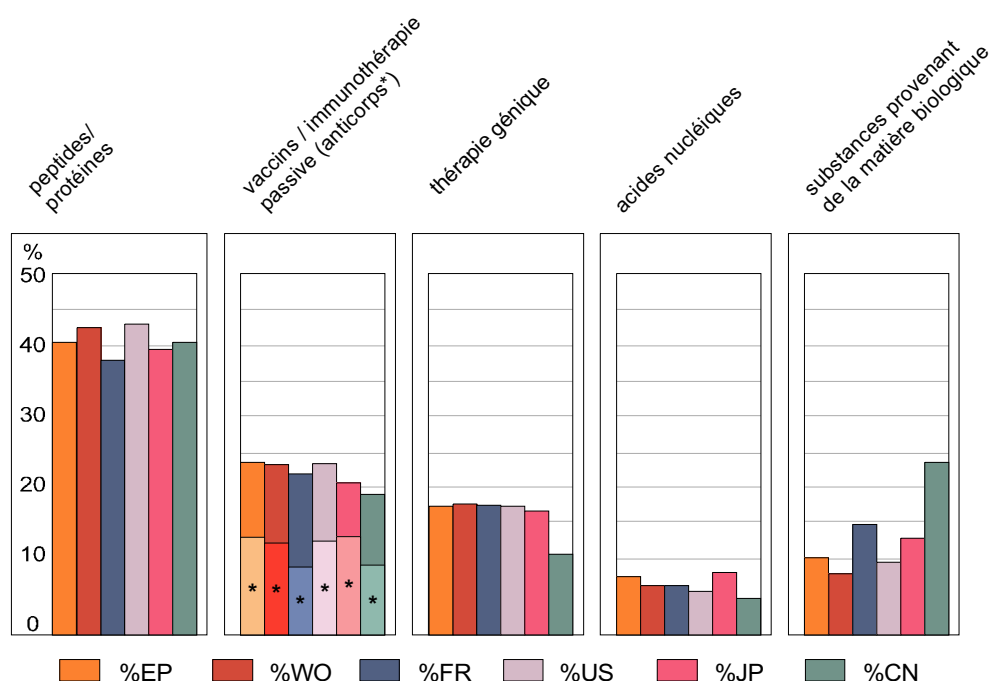


Figure 5 : Pourcentage des brevets en fonction de la classe des biomédicaments (étude de 2008 de l'Institut National de la Propriété Industrielle, intitulée « Du médicament au biomédicament : tendances de l'innovation »).

Une étude de l'Institut National de la Propriété Industrielle (INPI) datée d'octobre 2008 montre l'importance des dépôts relatifs aux biomédicaments. Basée sur une analyse des demandes de brevet français déposées entre 1997 et 2004, la France se place en 5^{ème} position après les dépôts de demandes internationales, les demandes américaines, européennes et japonaises et 39% de demandes liées à un médicament couvrent en fait un biomédicament. Parmi elles, environ 40% sont des demandes relatives à un peptide ou une protéine, 20 à 25 % couvrent un vaccin ou un anticorps et un peu moins de 20% sont relatives à la thérapie génique.

Cette étude montre également que le nombre de demandes de Certificat Complémentaire de Protection (voir paragraphe 1.2.3) liées à un biomédicament augmente pour atteindre en 2007, près de 60% des demandes de CCP.

Bouleversant le schéma classique de R&D de la pharmacopée traditionnelle, les acteurs de ce secteur ont dû rapidement s'adapter et se positionner au sein d'un environnement en pleine mutation.

Les brevets jouent donc un rôle essentiel dans l'économie et le développement des biotechnologies et des biomédicaments. Ils sont et resteront au cœur de la bioéconomie, définie par l'OCDE comme « un système dans lequel les biotechnologies assureront une part substantielle de la production économique ».

Annexe 2

***Poster Transposon tools: worldwide patent
landscape and patent exploitation by key actors***

Transposon tools: worldwide patent landscape and patent exploitation by key actors

F. PALAZZOLI¹, D. CARDI², U. STREICHENBERGER², Y. BIGOT¹

¹GICC - CNRS/Univ. F. Rabelais, Tours, FRANCE, ²France Innovation Scientifique & Transfert SA, Paris, FRANCE
fabien.palazzoli@etu.univ-tours.fr or delphine.cardi@fist.fr

Introduction : why did we do such a study ?

Nowadays, in the international scientific and economic competition, the protection of Biotech inventions by Industrial Property (IP) rights should be taken into account: **patents are at the core of the Biotech business**. A patent can be defined as an exclusive right granted for an invention to an applicant, for a limited period of time (generally 20 years to the maximum) and on a territorial basis (a specific country or a group of countries).

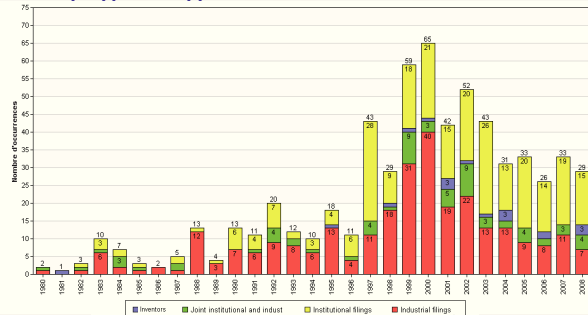
In the general context of the gene delivery, vectors are composed by several elements as insulators, targeting systems... that are potentially locked by IP rights, what is of a major importance to plan a strategy of technological development. The proposition of the first clinical trial of gene transfer system based on a transposon asks the question of the property of the technology and its economic exploitation.

Are transposon-based systems protected by patents ? Which patents are kept in force for an exploitation strategy ?

Methodology

The patent extraction and overall statistics were respectively performed using FamPat patent database (Questel) and Intellixir software. The patent query was built with the following keywords: transposase(s), mobile or transposable genetic element(s) or transposition; and the search was carried out in the titles, the abstracts and the independent claims of patent documents.

1) Cross-analysis of the evolution in patent filings over time by type of applicants

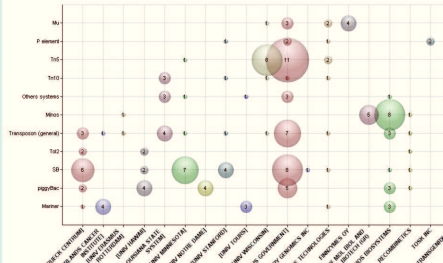


629 patents were filed during the period 1980-2008. The year 1997 appears as a pivotal year with 43 patents and marks the emergence of the main transposon-based system as vectors for providing cells with additional nucleic acid material (*SB*, *Tol2*, *Mariner* elements...).

Since 2001, the number of institutional applicants is stable whereas the number of industrial applicants drops. It may be due to the fact that the entry barriers to the transposon tool sector are too high.

2) Protection strategies of key applicants

a) Transposon systems claimed in patents by key applicants



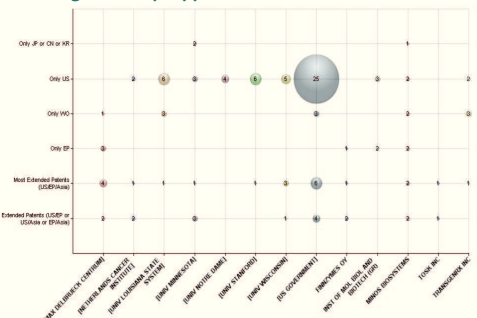
Pioneer applicants filed several patents to protect the transposon system they developed (Univ. Minnesota on *SB*, Finnzymes on *Mu*...), covering the proof of concept in usage and in engineering.

Other applicants (Epicentre, Recombinetics...) claimed in their patents an invention using a "general" transposon system, and would need to obtain licenses of key patents of transposons.

b) Geographic strategies of filings of key applicants

The United States of America are indisputably the country where the most patents have been filed because of the nationality of applicants and because it also represents a major market.

However, the most important patents were extended in other key countries (EU, Japan...).



3) Patent exploitation for therapeutic or other purposes

Transposon systems	Key applicants	Commercial exploitation	Creation	Nationality	Field	Purposes
<i>Mu</i>	Finnzymes	Finnzymes	1986	Finland	Molecular biology - kits	DNA sequencing Functional analysis of proteins
<i>Tol2</i>	Wisconsin Alumni Research Foundation	Epicentre	1987	USA	Molecular biology - kits	DNA sequencing
<i>Minos</i>	Institute for Molecular Biology and Biotechnology & Minos Biosystems	Minos Biosystems	2000	USA	Molecular biology - kits	in-situ production of proteins Functional gene discovery
<i>piggyBac</i>	US Government, University Notre Dame	Oxitec Minos Biosystems Others ?	2002 2000	USA	Agriculture Molecular biology - kits	Control insect pests in-situ production of proteins Functional gene discovery
<i>Tel2</i>	Louisiana State University and Agricultural and Mechanical College & Transgene	Transgene	2002	France	Therapeutic	Production of protein-based drugs
<i>Sleeping Beauty</i>	University of Minnesota	Discovery Genomics	2002	USA	Therapeutic	Gene delivery Gene discovery Gene therapy
<i>p elements</i>	Harvard Medical School Telix	Telix	1998	USA	Therapeutic	Drug discovery
<i>piggyBac</i>	US Government, University Notre Dame	Minos Biosciences Others ?	2006	USA	Therapeutic	Gene therapy
<i>Tol2</i>	Japan Science and Technology Corp.	?	?	?	?	?

The main transposon systems are included in economic projects mostly by US companies that concern therapeutic purposes or European companies for non-therapeutic developments (mutagenesis...). Indeed, the universities generally licensed their transposon-based technologies to spin-off companies (Discovery Genomics from the University of Minnesota...).

Among DNA transposons, the main promising systems, such as *Sleeping beauty* and *Tol2* are strongly protected by patents except *piggyBac* that has interesting capacities in terms of cargo. The first patents covering this transposon vector were a priori only filed in USA. So, *piggyBac* seems to be the only system for gene delivery that is freedom to operate for R&D plans, that may explain why *piggyBac* is integrated in the business models of companies as Oxitec or Manoa Biosciences.

Conclusion

As patent applications are published 18 months after their filings, they represent a vast and non-negligible source of:

- **scientific data** → The claims and the description disclose all technical details of the invention (which transposon systems are protected by patents ?).
- **legal information** → Are these patents kept in force ? In which countries ?
- **business-relevant information** → Global patent landscape knowledge of the environment of transposon tools facilitates decision making and R&D orientations.

The analysis of this three types of information allow to build a strong strategy of technological development.

Annexe 3

Poster Patent landscape analysis

“Nanoparticles for bio-imaging”

This patent landscape analysis on the field of “nanoparticles used in bio-imaging” shows a boom of patent filings in this sector. However, this recent technology is confronted to a precarious market in the field of medical imaging. This can be explained by the difficulties encountered by companies to lead clinical trials with such particles and therefore the cost of such development and commercialization. As a result, only a limited number of major companies (GE, Philips, Bayer, Fujifilm, Konica Minolta) and companies specifically focused on this topic (AMAG Pharmaceuticals, Guerbet, Visen Medical) developed this type of nanoparticles. Due to the importance of basic research needed to improve nanoparticles (residence time, biodegradability, functionalization...) and the relative independence of the institutional applicants regarding market consideration, an important number of institutional player positioned themselves on this sector. Although patents claiming magnetic nanoparticles are in majority, there is a clear tendency to describe nanoparticles combining several functionalities (i.e. magnetic AND luminescent, etc...) and applications (data not shown): strong specific patents are in small number.

The analysis of the patent landscape of this sector shows the necessity of maturation of this technology and confirms that places for new entrants are still available provided that their patent portfolio is consistent enough.

Annexe 4

Communications orales réalisées

Communications orales réalisées au cours de la thèse

- **Les brevets : une source d'informations stratégiques pour les acteurs des Biotechs** – *Rencontres technologiques IMT 2010, Technologies de bioproduction & lancement de l'ouvrage Bio³ : Biotechnologies – Bioproduction – Biomédicaments*, Tours, France, 13-14 octobre 2010
- **Insulator patents: position effects of biopharmaceuticals in their environment** – *Institute of Biotechnology, University of Lausanne*, Lausanne, Suisse, 16 septembre 2010
- **Table ronde « Témoignages de docteurs sur leur insertion professionnelle »** – *Doctoriales des Pays de la Loire 2010*, Sablé sur Sarthe, France, 02 juillet 2010
- **Intérêt des paysages brevets dans le développement d'une politique scientifique : la recherche de niches de développement technologique au GICC** – *Symposium d'analyse et de réflexion sur les : modèles, méthodes, ingénierie de l'Intelligence Compétitive*, Beaulieu sur Mer, France, 25-26 novembre 2009
- **Transposon tools: worldwide landscape of intellectual property and technological developments** – *XVI^{ème} colloque national sur les éléments transposables*, Le Mans, France, 1-3 juillet 2009
- **Industrial Property backgrounds of the non-viral integration systems: Preliminary strategic study for the establishment of R&D plans in non-viral gene therapy** – *Symposium on Recombinase-based non-viral gene transfer, XVIth Annual Congress of the European Society of Gene and Cell Therapy*, Bruges, Belgique, 13-16 novembre 2008

Annexe 5

Article BIOTECH.INFO n°437

Fist analyse le « paysage brevets » sectoriel

Grâce à la nouvelle activité de Fist (France Innovation Scientifique et Transfert) les brevets fournissent une information essentielle à la prise de décision stratégique.

« Il est nécessaire de mieux comprendre son environnement brevets pour mieux innover. Aujourd'hui, 80 % de l'information technologique est disponible dans les textes de brevets existants mais toutes ces données sont disséminées parmi plus de 40 millions de textes. Il est donc nécessaire de disposer d'une méthodologie et d'instruments adaptés pour faire d'une analyse de brevets un outil d'intelligence économique », explique Sylvain Goiran, responsable du département Études chez Fist (France Innovation Scientifique et Transfert).

Créé en 1992, Fist SA, filiale du CNRS et d'Oséo Innovation, est une société de transfert et de commercialisation de technologies innovantes qui gère un portefeuille de 3000 brevets et négocie près de 100 licences d'exploitation par an. Fist dispose depuis 2004 d'une équipe spécialisée dans l'analyse globale de brevets et a développé un savoir-faire qui lui permet d'analyser et de visualiser l'environnement brevets d'un secteur technologique afin de comprendre la dynamique des réseaux existants.

Une meilleure compréhension de l'environnement brevets permet à tous les acteurs de mieux appréhender leur positionnement, d'identifier de nouveaux partenaires ou concurrents et de détecter les équipes et les technologies émergentes. Aujourd'hui, Fist commercialise deux types d'études : « IP On Demand » réalisée sur mesure et « IP Overview » qui dresse un panorama global sur des thématiques d'actualité ou émergentes.

UN OUTIL STRATÉGIQUE

La première étude « IP Overview » dans le domaine des sciences du vivant est aujourd'hui commercialisée via le site Internet de Fist. Elle concerne les « Thérapies contre le VIH » et regroupe, segmente et analyse, les 6800 familles de brevets liées à ce domaine. Cette étude met en évidence les stratégies mises en place par les acteurs et les tendances actuelles du domaine. Il apparaît par exemple que les efforts en

R&D faiblissent depuis 2005 malgré la forte croissance du marché du VIH et le besoin constant de nouvelles thérapies, laissant ainsi ce secteur en manque de nouvelles perspectives (40 % de diminution du nombre de dépôts de brevets). Cet essoufflement affecte même certains leaders historiques, mais donne l'opportunité à de nouveaux acteurs de s'imposer sur des segments spécifiques encore peu explorés (Cyclophilin, Trim5alpha, TSG 101, etc.).

Les prochaines « IP Overview » porteront sur les thématiques « Paludisme » et « Inhibiteurs de Cyclin-Dependant-Kinases ». Fist a par ailleurs initié cette année une collaboration inédite avec la cellule de développement stratégique du laboratoire GICC (Génétique Immunothérapie Chimie & Cancer) de Tours, pour analyser l'environnement brevets de leurs thématiques de recherche notamment en thérapie génique (études sur les domaines de liaison à l'ADN de type doigt de zinc, systèmes de transfection et cellules souches).

« Nos études constituent de véritables outils opérationnels d'aide à la prise de décision et s'adressent aux responsables brevets des sociétés, mais également aux investisseurs, cabinets de conseils, structures de recherche et organismes de valorisation, conclut Sylvain Goiran. En plus d'apporter aux décideurs un nouvel élément essentiel d'informations au même titre que les informations économiques, elles permettent au sein d'une même entreprise d'avoir un outil et un langage communs entre les équipes de recherche, les équipes dirigeantes et les financiers. » ■ CÉLINE HÉCHARD

FICHE D'IDENTITÉ

LIEU : Paris
CRÉATION : 1992
FINANCEMENT : CNRS et Oséo Innovation
EFFECTIF : 45
SPÉCIALITÉ : Analyse de portefeuilles de brevets et transfert de technologie
CONTACT : ipoverview@frinnov.fr
SITE INTERNET : <http://www.Fist.fr>

EN BREF

L'Allemand Medac a acheté une licence mondiale auprès de l'École Médicale de Hanovre (MHH), pour un diagnostic qui aide à orienter le traitement de la leucémie lymphoblastique aiguë. L'accord a été négocié par Ascenion GmbH, le partenaire exclusif de MHH pour la propriété intellectuelle.

Une contamination accidentelle au bacille de charbon s'est produite dans un laboratoire de l'Agence française de sécurité sanitaire des aliments (Afssa). Des laborantins ont été pris en charge par les secours et la mise en place du protocole de décontamination a été immédiate.

Le Parlement européen a adopté un règlement, qui entrera en application entre 2010 et 2013, sur les produits cosmétiques. Il impose une évaluation renforcée des risques, y compris pour les nanomatériaux, de plus en plus utilisés par l'industrie. Le règlement inclut le principe de l'élimination progressive des tests sur les animaux.

GlaxoSmithKline (Royaume-Uni) va partager 800 brevets avec d'autres instances qui cherchent de nouveaux traitements dans 16 maladies tropicales négligées, hors Sida. Le groupe baisse de 45 % le prix de 110 thérapies brevetées allant du paludisme à l'asthme. Cette politique globale vise à aider les pays pauvres.

Le Californien Arena Pharmaceuticals a levé 50 millions de dollars auprès de Azimuth Opportunity Ltd. Au bout de 18 mois, la société pourra éventuellement vendre des actions à Azimuth. Arena se situe dans l'obésité, les troubles du sommeil, les maladies neurodégénératives, la schizophrénie, la dépression et l'anxiété.

Annexe 6

New concept of technology classification - WIPO

Table 2: New concept of technology classification, update: May 2008

Area, field	IPC code
I Electrical engineering	
1 Electrical machinery, apparatus, energy	F21#, H01B, H01C, H01F, H01G, H01H, H01J, H01K, H01M, H01R, H01T, H02#, H05B, H05C, H05F, H99Z
2 Audio-visual technology	G09F, G09G, G11B, H04N-003, H04N-005, H04N-009, H04N-013, H04N-015, H04N-017, H04R, H04S, H05K
3 Telecommunications	G08C, H01P, H01Q, H04B, H04H, H04J, H04K, H04M, H04N-001, H04N-007, H04N-011, H04Q
4 Digital communication	H04L
5 Basic communication processes	H03#
6 Computer technology	(G06# not G06Q), G11C, G10L
7 IT methods for management	G06Q
8 Semiconductors	H01L
II Instruments	
9 Optics	G02#, G03B, G03C, G03D, G03F, G03G, G03H, H01S
10 Measurement	G01B, G01C, G01D, G01F, G01G, G01H, G01J, G01K, G01L, G01M, (G01N not G01N-033), G01P, G01R, G01S; G01V, G01W, G04#, G12B, G99Z
11 Analysis of biological materials	G01N-033
12 Control	G05B, G05D, G05F, G07#, G08B, G08G, G09B, G09C, G09D
13 Medical technology	A61B, A61C, A61D, A61F, A61G, A61H, A61J, A61L, A61M, A61N, H05G
III Chemistry	
14 Organic fine chemistry	(C07B, C07C, C07D, C07F, C07H, C07J, C40B) not A61K, A61K-008, A61Q
15 Biotechnology	(C07G, C07K, C12M, C12N, C12P, C12Q, C12R, C12S) not A61K
16 Pharmaceuticals	A61K not A61K-008
17 Macromolecular chemistry, polymers	C08B, C08C, C08F, C08G, C08H, C08K, C08L
18 Food chemistry	A01H, A21D, A23B, A23C, A23D, A23F, A23G, A23J, A23K, A23L, C12C, C12F, C12G, C12H, C12J, C13D, C13F, C13J, C13K
19 Basic materials chemistry	A01N, A01P, C05#, C06#, C09B, C09C, C09F, C09G, C09H, C09K, C09D, C09J, C10B, C10C, C10F, C10G, C10H, C10J, C10K, C10L, C10M, C10N, C11B, C11C, C11D, C99Z
20 Materials, metallurgy	C01#, C03C, C04#, C21#, C22#, B22#

21	Surface technology, coating	B05C, B05D, B32#, C23#, C25#, C30#
22	Micro-structure and nano-technology	B81#, B82#
23	Chemical engineering	B01B, B01D-000#, B01D-01##, B01D-02##, B01D-03##, B01D-041, B01D-043, B01D-057, B01D-059, B01D-06##, B01D-07##, B01F, B01J, B01L, B02C, B03#, B04#, B05B, B06B, B07#, B08#, D06B, D06C, D06L, F25J, F26#, C14C, H05H
24	Environmental technology	A62D, B01D-045, B01D-046, B01D-047, B01D-049, B01D-050, B01D-051, B01D-052, B01D-053, B09#, B65F, C02#, F01N, F23G, F23J, G01T, E01F-008, A62C
IV Mechanical engineering		
25	Handling	B25J, B65B, B65C, B65D, B65G, B65H, B66#, B67#
26	Machine tools	B21#, B23#, B24#, B26D, B26F, B27#, B30#, B25B, B25C, B25D, B25F, B25G, B25H, B26B
27	Engines, pumps, turbines	F01B, F01C, F01D, F01K, F01L, F01M, F01P, F02#, F03#, F04#, F23R, G21#, F99Z
28	Textile and paper machines	A41H, A43D, A46D, C14B, D01#, D02#, D03#, D04B, D04C, D04G, D04H, D05#, D06G, D06H, D06J, D06M, D06P, D06Q, D99Z, B31#, D21#, B41#
29	Other special machines	A01B, A01C, A01D, A01F, A01G, A01J, A01K, A01L, A01M, A21B, A21C, A22#, A23N, A23P, B02B, C12L, C13C, C13G, C13H, B28#, B29#, C03B, C08J, B99Z, F41#, F42#
30	Thermal processes and apparatus	F22#, F23B, F23C, F23D, F23H, F23K, F23L, F23M, F23N, F23Q, F24#, F25B, F25C, F27#, F28#
31	Mechanical elements	F15#, F16#, F17#, G05G
32	Transport	B60#, B61#, B62#, B63B, B63C, B63G, B63H, B63J, B64#
V Other fields		
33	Furniture, games	A47#, A63#
34	Other consumer goods	A24#, A41B, A41C, A41D, A41F, A41G, A42#, A43B, A43C, A44#, A45#, A46B, A62B, B42#, B43#, D04D, D07#, G10B, G10C, G10D, G10F, G10G, G10H, G10K, B44#, B68#, D06F, D06N, F25D, A99Z
35	Civil engineering	E02#, E01B, E01C, E01D, E01F-001, E01F-003, E01F-005, E01F-007, E01F-009, E01F-01#, E01H, E03#, E04#, E05#, E06#, E21#, E99Z

Note: This table is available in Excel format on: www.wipo.int/ipstats/en/statistics/patents
Users are requested cite WIPO as the source in the following manner: "Source: WIPO IPC-Technology Concordance Table".

Annexe 7

**Recherche d'informations en chinois sur Baidu et
sur le portail d'informations CNKI**

Recherche d'informations en chinois sur Baidu et sur le portail d'informations CNKI

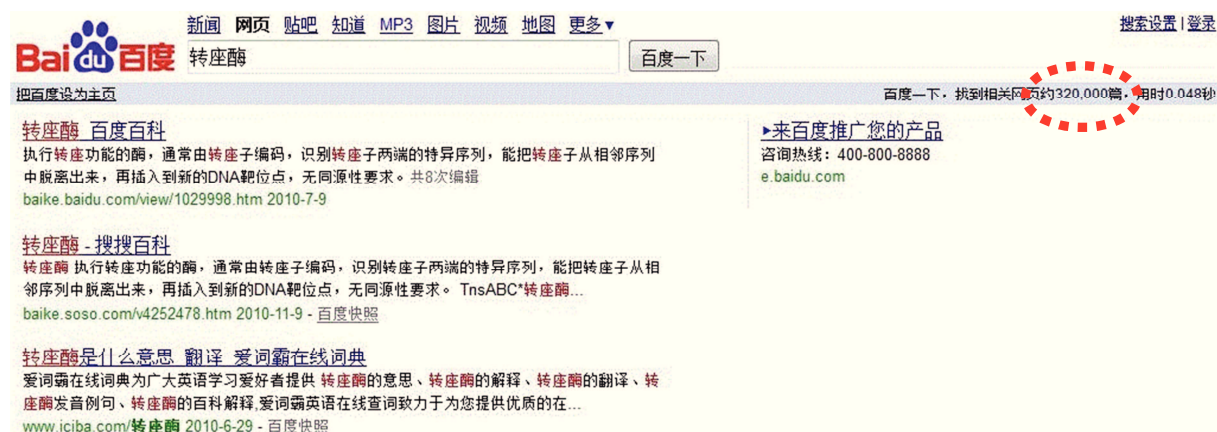


Figure 29 : Page de résultats de la requête « 转座酶 », c'est-à-dire « transposase » en chinois, avec le moteur de recherche chinois Baidu. La recherche donne 320 000 résultats.



Figure 30 : Page de résultats de la requête « 转座酶 », c'est-à-dire « transposase » en chinois, sur le portail d'information CNKI. Les résultats sont de 131 articles en anglais (en vert) et 2578 en chinois (en rouge).

Exploitation de l'information brevets dans un laboratoire de recherche public : identification de niche de développement technologique en bioproduction et en thérapie génique

Résumé

Dans un monde où la course à l'innovation est de plus en plus rapide, il est important pour une entreprise innovante ou un laboratoire de recherche public de mettre en place une stratégie de protection et de valorisation de ses inventions qui soit performante. La protection des résultats par des brevets revêt une importance capitale pour le développement industriel des biotechnologies qui forment un secteur innovant et prometteur, et où la R&D exige des investissements financiers considérables. Au-delà de cet intérêt fondamental, les brevets sont aussi une source de premier plan en matière d'informations technologiques, juridiques et stratégiques, pouvant être exploitées à travers des paysages brevets. Ces études constituent un outil privilégié d'aide à la décision en matière de stratégie de R&D puisqu'elles permettent de définir les axes de recherche des concurrents et les niches de développement technologique libres de droits de Propriété Intellectuelle.

Dans cette optique, mes travaux de thèse ont consisté en l'élaboration et l'analyse de plusieurs paysages brevets sur des technologies utilisées en bioproduction et en thérapie génique : les transposons, les domaines de liaison à l'ADN de type doigt de zinc et les éléments de contrôle de la chromatine. Ces études ont permis l'identification de niches de développement technologique qui ont été intégrées dans les projets de recherche du laboratoire. Cependant, l'information brevets comme indicateur concurrentiel et technologique est largement méconnue dans les laboratoires de recherche publics pourtant eux-aussi des acteurs de l'innovation. J'ai donc montré par mes travaux que les laboratoires de recherche publics ont tout autant intérêt à miser sur l'information brevets et la veille stratégique pour rester compétitifs et innover. En effet, dans un contexte international où la compétition scientifique et économique s'intensifie, il est fort probable que les acteurs de l'innovation qui réussiront seront ceux qui utiliseront l'information disponible avec le plus de pertinence.

Mots clés : information brevets, paysages brevets, veille stratégique, biotechnologies, bioproduction, thérapie génique.

Résumé en anglais

In a world where the innovation race is increasing fast, it is of economic importance for an innovative company or a public research laboratory to develop a strategy for the protection and enhancement of its inventions is efficient. Protection of results through patents is critical for the industrial development of biotechnology which are an innovative and promising sector where R&D requires considerable financial investments. Beyond this fundamental interest, patents are also a source of information on technological, legal and strategic, which can be exploited through patent landscapes. These studies are a key tool for decision support in R&D since they allow to identify research strategies of competitors and technological niches free from of Intellectual Property rights.

In this context, my thesis work has consisted in the development and analysis of several landscapes patents on different technologies used in bioproduction and gene therapy: transposons, zinc finger DNA-binding domains and chromatin control elements. These studies allowed the identification of niches of technological development that have been then incorporated within the laboratory's research projects. However, patent information as a competitive and technology indicator is largely unknown in the public research laboratories, yet they are also actors of innovation. I have therefore showed in my PhD thesis that public research laboratories have as much interest to capitalize on the patent information and intelligence to stay competitive and innovate. Indeed, in an international context where the scientific and economic competition intensifies, it is likely that the innovation actors that succeed will be those who will use the information available with the most relevance.

Keywords: patent information, patent landscape, strategic survey, biotechnology, bioproduction, gene therapy.