





UNIVERSITÉ FRANÇOIS - RABELAIS DE TOURS

ÉCOLE DOCTORALE SSBCV

Equipe « Biologie et Bioinformatique des Systèmes de Signalisation » INRA 0085-CNRS UMR 7247-Université de Tours-Haras Nationaux

THÈSE présentée par :

Pauline GLOAGUEN

soutenue le : 14 décembre 2012

pour obtenir le grade de : Docteur de l'Université François - Rabelais de Tours

Discipline/ Spécialité : Sciences de la vie et santé/ Bioinformatique

Inférence automatique de modèles de voies de signalisation à partir des données expérimentales

THÈSE dirigée par :

Mme POUPON Anne Directeur de recherche CNRS, Nouzilly

RAPPORTEURS:

Mme DEVIGNES Marie-Dominique Chargée de recherche CNRS, Docteur d'état, Nancy

Mr PAIN Bertrand Directeur de recherche INSERM, Bron

JURY:

Mme DEVIGNES Marie-Dominique Mme DUITTOZ Anne Mme FROIDEVAUX Christine Mr PAIN Bertrand Mme POUPON Anne Mme TESSERAUD Sophie Mr TRUBUIL Alain Chargée de recherche CNRS, Docteur d'état, Nancy Professeur à l'Université François Rabelais, Tours Professeur à l'Université Paris Sud 11, Orsay Directeur de recherche INSERM, Bron Directeur de recherche CNRS, Nouzilly Directeur de recherche INRA, Nouzilly Ingénieur de recherche INRA, Jouy-en-Josas

A ma famille

Remerciements

Pour commencer, je souhaite remercier les membres du jury d'avoir accepté d'évaluer mes travaux de thèse. Marie-Dominique Devignes et Bertrand Pain pour avoir accepté d'en être les rapporteurs et Christine Froidevaux, Anne Duittoz, Sophie Tesseraud et Alain Trubuil pour avoir accepté d'en être examinateurs.

Ensuite, je tiens à remercier Anne, ma directrice de thèse. Les mots me manquent pour t'exprimer ma reconnaissance car tu as été la seule à me laisser une chance au moment où même moi j'avais perdu toute confiance. J'ai eu un grand plaisir à travailler avec toi pendant ces trois années, autant pour le côté scientifique que pour le côté humain car ton sourire et ta bonne humeur sont remarquables.

Un grand merci à Pascale et Eric pour vos conseils et votre soutien. Sans vous, la revue FSH n'aurait jamais vu le jour, merci encore. Joël, j'espère que tu n'as pas trop souffert en partageant ton bureau avec moi. Sache que ce fût, pour ma part, un grand plaisir de travailler en ta compagnie. Olivier, grâce à toi, j'ai découvert la conception des modèles de données et en parallèle je t'ai « sâoulé » avec ma formalisation des données. Merci d'avoir partagé ton expérience et d'avoir surmonté la complexité de la formalisation et surtout de son stockage. A Domitille, pour ton accueil il y a maintenant trois ans et pour tout le reste. A Christophe, pour tes passages sympas jusqu'à ce que tu nous dises qu'en bioinformatique, on ne glande rien. Merci à tous les membres de l'équipe BIOS, ça a été tellement agréable de travailler dans la bonne humeur. Au départ, je croyais qu'une thèse représentait un travail personnel permettant d'obtenir un doctorat, mais grâce à vous tous, je me suis rendue compte qu'une thèse ne peut être réussie sans un travail personnel certes mais surtout sans un travail d'équipe.

Christine, je suis heureuse d'avoir eu l'occasion de collaborer avec toi. Merci pour tous tes conseils et remarques tout au long de ma thèse ainsi que d'avoir pris le temps de m'expliquer la logique. Merci Adrien d'avoir participer à ces travaux, grâce à toi, nous avons pu avancer et surtout je suis heureuse de savoir que ces travaux vont continuer au travers de ta propre thèse. Merci Sarah d'avoir contribué à ce travail et pour tous tes conseils. Merci à la promotion de Master 2 Bioinformatique et Biostatistiques de l'Université de Paris Sud pour leur participation à l'extraction des données expérimentales.

Barbara, Liliane, Marie-Françoise, Stéphanie et toutes les autres personnes de l'INRA, merci pour tous ces petits moments passés avec vous. A Camille, non seulement nous avons partagé nos expériences de thèse mais nous sommes aussi devenues amies. Discuter avec toi a toujours été réconfortant dans les moments de doute et souvent très drôles dans les autres moments. Merci Peggy, tu as été la première personne de l'INRA que j'ai rencontré par hasard lors d'une des promenades de nos toutous. J'ai eu beaucoup de plaisir à partager des discussions avec toi pendant les balades qui ont suivi. A tous les autres thésards, Anthony, Daphné, Sarah (courage pour la suite), Kelly, Leslie... Vincent, n'oublie pas que lorsque tu veux automatiser quelque chose, il faut structurer tes fichiers et éviter les tremblements de terre. Merci à toutes les personnes de l'INRA qui par leur présence m'ont aidée à avancer. Un grand merci au département PHASE et à la région Centre d'avoir financé ces travaux de thèse.

Je veux aussi remercier toute ma promotion de Master, sans qui je ne serai jamais arrivée là : Aurélie, Florent, Léa, Maxime et Tristan. La découverte de la bioinformatique a été tellement plus simple à vos côtés. A tous mes amis : Karine, Suzanne, Julien, Fabienne ... d'avoir toujours cru en moi et d'avoir toujours été présents.

A ma famille. Je vous dédie cette thèse car sans vous tous, rien n'aurait été possible. Maman et Papa, vous avez toujours été présents pour moi, aussi bien dans les moments de déception comme mes trois premières années de fac, que dans les moments de joie. Je maintiens tout de même que la première année de médecine, ça ne compte pas car rares sont ceux qui réussissent du premier coup. J'espère, aujourd'hui, ne vous apporter que le bonheur que vous méritez. Les frangins, Arnaud et Erwan, votre contribution est très indirecte bien sûr, mais sans vous, je ne serai jamais devenue suffisamment forte pour affronter les obstacles de la vie. A mes grand-parents, ceux qui n'ont pas été là lorsque cette aventure a commencé, pépé petit Louis qui n'aura pas connu la fin et mémé Monique qui heureusement est toujours à mes côtés. Merci à vous tous d'avoir toujours été là, Marie-Josée (ma vieille tante), Manon, Paul, Patricia, Patrick, les nains de jardin et tous les autres membres de ma famille.

Willy, tu n'as pas eu de chance, nous nous sommes rencontrés juste au début de ma troisième année de thèse, ce qui comme tu l'as appris, est la période la plus difficile. Je suis si fière de t'avoir à mes côtés car je ne pensais pas rencontrer, un jour, une personne aussi bien que toi. Les mots ne seront jamais suffisants pour te remercier. Mais surtout j'espère rester à tes côtés pour très longtemps.

Résumé

Les réseaux biologiques, notamment les réseaux de signalisation déclenchés par les hormones, sont extrêmement complexes et leur construction représente un challenge important. Les méthodes expérimentales à haut débit permettent d'aborder cette complexité, mais la prise en compte de l'ensemble des données générées requiert la mise au point de méthodes automatiques pour la construction des réseaux.

Nous avons développé une nouvelle méthode d'inférence automatique reposant sur la formalisation, sous forme de règles logiques, du raisonnement de l'expert sur les données expérimentales. Cela nécessite la constitution d'une base de connaissances, ensuite exploitée par un moteur d'inférence afin de déduire les conclusions permettant de construire les réseaux.

Les récepteurs couplés aux protéines G (RCPG) étant au cœur des recherches de l'équipe BIOS, et en particulier le récepteur de l'hormone folliculo-stimulante (FSH-R), nous avons donc choisi le réseau de signalisation intracellulaire déclenché par cette hormone comme modèle d'élaboration de notre méthode automatique d'inférence de réseaux. En parallèle, ce réseau a également été construit manuellement pour évaluer notre méthode.

La création des règles utilisées dans notre méthode d'inférence reposait sur les expériences rencontrées dans le cadre du réseau de signalisation induit par la FSH, c'est pourquoi nous avons choisi un modèle de validation : le réseau induit par le facteur de croissance épidermique (EGF), se liant à un récepteur tyrosine kinase. Ce réseau de signalisation est très étudié et par conséquent connu, c'est pourquoi il nous semblait être un bon modèle. Ce test a pour but de démontrer que notre méthode ne se limite pas à la construction d'un seul type de réseau de signalisation.

Ces modèles ont permis de mettre en évidence les différents points qui devront être améliorés mais surtout de démontrer la pertinence de notre nouvelle méthode d'inférence automatique de réseaux de signalisation.

<u>Mots-clés</u>: Inférence automatique, Réseaux de signalisation, Règles logiques, Logique du premier ordre

Abstract

Biological networks, including signalling networks induced by hormones, are very complex

and their construction represents a big challenge. High-throughput experimental methods

permit to approach this complexity, but to be able to use all generated data, it is necessary to

create automated inference methods to build networks.

We have developped a new automated inference method based on the formalization of the

expert's reasoning on experimental data. This reasoning is converted into logical rules. This

work requires the creation of a knowledge base, which is used by an inference engine to

deduce conclusions to build networks.

The BIOS team works on G protein-coupled receptors (GPCR), particularly on the follicle

stimulating hormone receptor (FSH-R), that is why we have chosen the intracellular

signalling network triggered by this hormone as an elaboration model of our automated

method for the networks inference. In parallel, this network has also been built manually to

evaluate our method.

The creation of rules in our inference method was based on experiments encountered in the

context of the signalling network triggered by FSH, for this reason we have decided to use a

validation model: the network induced by the epidermal growth factor (EGF) which binds to a

tyrosine kinase receptor. This signalling network is already well known, and its automated

conctruction aims to demonstrate that our method is generic.

These models, although they have highlighted different points that need to be improved,

demonstrate the pertinence of our new automated method to infer molecular networks.

Keywords: Automated inference, Signalling networks, Logical rules, First order logic

7

Table des matières

Remerciements	4
Résumé	6
Abstract	7
Table des matières	8
Liste des tableaux	12
Liste des figures	13
Liste des annexes	17
Liste des abréviations	18
Introduction	20
I- Les récepteurs couplés aux protéines G	22
A- Généralités	22
B- Structure	23
C- Classification	24
D- Mécanismes de signalisation cellulaire	27
1- Signalisation dépendante des protéines G	28
2- Désensibilisation des récepteurs	33
3- Internalisation et devenir des RCPG	34
4- Signalisation indépendante des protéines G	35
E- Récepteur de l'hormone folliculo-stimulante (FSHR)	39
II- Construction de réseaux biologiques : Etat de l'art	41
A- Approche manuelle	42
1- Bibliographie	42
2- Banques de données	42
3- Conclusion sur l'approche manuelle	44
B- Approche semi-automatique	44
C- Approche automatique	45
1- Modèle reposant sur des statistiques	45
2- Modèles logiques	51
3- Systèmes d'équations différentielles	55
4- Méthodes à base de règles	57
5- Autres méthodes	60
6- Conclusion sur les méthodes actuelles de construction des réseaux moléculaires	61

III- Visualisation graphique des réseaux biologiques	62
A- SBML	63
B- SBGN	64
C- BioPax	66
D- Logiciels de visualisation graphique des réseaux biologiques	68
Problématique scientifique et objectif de la thèse	70
Matériel et méthodes	72
I- Formalisation et inférence	74
A- Inférence	74
1- Calcul de prédicats ou Logique du premier ordre	74
2- Moteur d'inférence SOLAR	77
B- Méthodologie de la formalisation	80
1- Classification hiérarchique des molécules	80
2- Hiérarchie des relations	86
3- Création de règles	88
II- Stockage des données	94
A- Règles et prédicats	94
B- Molécules	96
C- Données expérimentales	98
III- Visualisation des réseaux	103
1- CellDesigner	103
2- Transformation des faits en éléments du réseau	103
Signalisation induite par l'hormone folliculo-stimulante (FSH)	108
I- Les voies classiques	110
II- D'autres voies de signalisation	124
Une nouvelle méthode d'inférence de réseaux de signalisation	134
I- Présentation de la méthode	137
II- Base de connaissances	146
A- Modèle de données	146
B- Données expérimentales	149
C- Interface utilisateur	149
D- Lien avec le moteur d'inférence	150
E- Choix de visualisation graphique	151
F- Liens avec d'autres bases de données	154

III- Application aux réseaux de signalisation de la FSH et de l'EGF	157
A- Réseau FSH	158
1- Voie dépendante de la protéine G stimulatrice	159
2- Voie dépendante des β-arrestines :	161
3- Voie dépendante de la PI3K :	163
4- Réseau obtenu	164
B- Réseau EGF	165
1- Présentation du récepteur EGF	165
2- Réseau EGF obtenu versus réseau EGF de référence	168
IV- Les leçons de ces premiers tests	173
A- Pourquoi manque-t'il des relations dans les réseaux obtenus ?	173
B- Le problème des ARN interférents	175
C- Utilisation du savoir implicite	176
D- Les faits contradictoires	177
1- Hypothèses contradictoires	177
2- Contradictions issues de l'interprétation des expériences	179
3- Problème des conditions expérimentales	181
V- Les limites de la formalisation	183
A- Les formes modifiées	183
B- Le problème des hypothèses	184
C- Des règles de plus en plus complexes	185
D- Difficultés pour la formalisation	187
1- Agonistes et analogues structuraux	187
2- Mutants et variants naturels	188
VI- Inférence	190
A- Utilisation du 'OU' et problème d'égalité	190
1- Le 'OU'	190
2- L'égalité	191
B- Temps de calcul	191
C- Indice de confiance des données	192
D- Règles d'élagage	193
Conclusion et perspectives	196
Bibliographie	200
Annexes	218

Résumé	283
Abstract	283

Liste des tableaux

Tableau 1 : Signalisation dépendante des β -arrestines	37
Tableau 2 : Méthodes d'inférence reposant sur la théorie de l'information	46
Tableau 3 : Méthodes d'inférence reposant sur le modèle bayésien	50
Tableau 4 : Méthodes d'inférence reposant sur le modèle booléen	53
Tableau 5 : Modifications post-traductionnelles et leur formalisation	84
Tableau 6 : Temps d'exécution de SOLAR avec un enchaînement de quatre r	ègles maximum
	191
Tableau 7 : Transcription induite par la FSH	258
Tableau 8 : Transcription inhibée par la FSH	260
Tableau 9 : Traduction induite par la FSH	260
Tableau 10 : Nombre de relations obtenues dans le réseau FSH	266
Tableau 11 · Nombre de relations obtenues dans le réseau EGF	275

Liste des figures

Figure 1 : Représentation 3D en 'ruban' de la structure de la Rhodopsine [11]	23
Figure 2 : Structure générale en 2D d'un RCPG	24
Figure 3 : Classification des RCPG selon [17, 18]	25
Figure 4 : Inhibition du récepteur Smoothened par le récepteur Patched	27
Figure 5 : Schématisation des différents mécanismes induits par l'activation d'un	RCPG [26]
	28
Figure 6 : Rencontre d'un RCPG activé avec une protéine G	28
Figure 7 : Mécanismes d'activation et de désactivation d'une protéine G	29
Figure 8 : Voies de signalisation induites par les différentes voies G	31
Figure 9 : Représentation schématique d'une transactivation d'un récepteur tyrosir	ne kinase. 32
Figure 10 : Représentation schématique des désensibilisations hétérologue et hom	ologue 33
Figure 11 : Internalisation et devenir intracellulaire des RCPG [95]	35
Figure 12 : Modèle de spécialisation fonctionnelle des différentes GRK [100]	36
Figure 13 : Inférence du réseau de kinases dépendantes des β -arrestines en aval	du récepteur
AT1aR	38
Figure 14 : Exemples de banques de données répertoriant des interactions molécul	laires43
Figure 15 : Passage de la matrice de précision au graphe correspondant	47
Figure 16 : Représentation d'un réseau bayésien de 4 gènes.	49
Figure 17 : Schéma représentant la régulation du cycle cellulaire	52
Figure 18 : Diagramme logique décrivant l'activité de la protéine Rb à partir de 4	entrées 52
Figure 19 : Principe d'un système expert	58
Figure 20 : Représentation graphique de modèles entre deux variables [178]	60
Figure 21 : Incohérences et ambiguïtés causées par le manque de standardisation	de notations
de représentation des réseaux biologiques	62
Figure 22 : Exemple la cascade MAP Kinase ERK en représentation graphique SE	3ML 64
Figure 23 : Symbolique utilisée dans le cadre du langage dédié aux diagrammes d	de processus
	65
Figure 24 : Exemple de la voie de MAP Kinase ERK sous forme de diagramme d	le processus.
	66
Figure 25 : Exemple de molécules et réactions décrites en BioPax issue de [185]	67
Figure 26 : Exemple de calcul de tableau avec un champ de production [191]	78
Figure 27 : Hiérarchie au sein de la classe ADN	81

Figure 28 : Hiérarchie au sein de la classe ARN	82
Figure 29 : Hiérarchie au sein de la classe Protéine	83
Figure 30 : Hiérarchie au sein de la classe Petites Molécules	85
Figure 31 : Complexation	86
Figure 32 : Phosphorylation directe	87
Figure 33 : Déroulement de la traduction	88
Figure 34 : Enzymatic Assay	89
Figure 35 : Phosphorylation Assay	91
Figure 36 : Schéma de données pour le stockage des règles et des prédicats	95
Figure 37 : Schéma de données pour le stockage des molécules	97
Figure 38 : Schéma de données pour le stockage des expériences	. 100
Figure 39 : Schéma de données pour le stockage des faits	. 101
Figure 40 : Convention de représentation utilisée pour la représentation des molécules	dans
CellDesigner	. 104
Figure 41 : Gène représenté avec une région régulatrice	. 104
Figure 42 : Exemples de prédicats convertis en représentation graphique CellDesigner	. 105
Figure 43 : Extraction des relations dans la base de données	. 106
Figure 44 : La FSH induit des voies anti-apoptotiques conjointement avec IGF1	. 125
Figure 45 : Stéroïdogenèse induite par la FSH	. 127
Figure 46 : Production de glutathion induite par la FSH	. 127
Figure 47: Voie PI3K/Akt	. 128
Figure 48 : Signalisation induite par la FSH, conduisant à la traduction des protéines dans	s les
cellules de Sertoli.	. 129
Figure 49 : Exemples d'actions de l'AMPc	. 131
Figure 50 : Différences de localisation cellulaire des molécules	. 132
Figure 51 : Structure générale de la base de données	. 146
Figure 52 : Phosphorylation Assay d'Akt et de STAT1 provenant d'Andl et al [212]	. 148
Figure 53 : Lien entre la base de données et le moteur d'inférence	. 150
Figure 54 : Différentes représentations graphiques d'une partie du réseau EGF	. 153
Figure 55 : Résultats obtenus dans différentes banques de données en effectuant	une
recherche sur le FSHR humain	. 155
Figure 56 : Statistiques du réseau FSH disponibles dans NetPath	. 156
Figure 57 : Voie G manuelle versus voie G automatique	. 160
Figure 58 : Voie β-arrestines manuelle versus voie β-arrestines automatique	. 162

Figure 59 : Voie PI3K manuelle versus voie PI3K automatique	164
Figure 60 : Dimérisation du récepteur de l'EGF [216]	165
Figure 61 : Voies de signalisation induites par l'EGF [216]	166
Figure 62 : Schématisation du réseau EGF issu de la banque de données Pathway	Interaction
Database, mis au format CellDesigner	167
Figure 63 : Parallèle entre le réseau issu de la PID et le réseau obtenu	170
Figure 64 : Complexation entre le récepteur EGFR et Grb2:SOS1	171
Figure 65 : Exemple de relations manquantes	174
Figure 66 : Déduction d'une nouvelle hypothèse	175
Figure 67 : Incompatibilité des hypothèses, l'exemple de la PKA	177
Figure 68 : Phosphorylation Assay de p38MAPK de Yu et al [204]	180
Figure 69 : Phosphorylation Assay de p38MAPK de Wayne et al [77]	180
Figure 70 : Hypothèses obtenues à partir d'un ELISA sur la production d'AMPc	en présence
de FSH	184
Figure 71 : Différents réseaux pouvant conduire au même résultat	184
Figure 72 : Fusion des trois réseaux pouvant être obtenus à partir d'un Phosphory	lation Assay
avec inhibiteur	185
Figure 73 : Possibilités de réseaux provenant d'un Phosphorylation Assay avec	une double
perturbation	186
Figure 74 : Fusion des réseaux pouvant être obtenus à partir d'un Phosphorylation	Assay avec
une double perturbation	187
Figure 75 : Schématisation des conclusions obtenues à partir d'un PANWB	190
Figure 76 : Exemple d'élagage possible	194
Figure 77 : Exemple d'élagage problématique	195
Figure 78 : Effet sur la complexation	220
Figure 79 : Dissociation	220
Figure 80 : Modification	221
Figure 81 : Réaction enzymatique	221
Figure 82 : Influence d'une molécule sur la transcription	223
Figure 83 : Exemple de l'AMPc et ses différentes formes modifiées	225
Figure 84 : Conclusions obtenues à partir d'un test ELISA avec inhibiteur	226
Figure 85 : Principe du FRET	227
Figure 86 : Western Blot	229
Figure 87 · Exemple de déduction d'une distance indirecte	231

Figure 88: Déduction d'une relation implicite à partir de faits déduits	validés
expérimentalement	232
Figure 89 : Activation du FSHR telle qu'inférée par notre méthode, recrutement de la μ	orotéine
Gs et production de l'AMPc	266
Figure 90 : Actions de l'AMPc dans la signalisation induite par la FSH	267
Figure 91 : Actions d'EPAC dans la signalisation induite par la FSH	267
Figure 92 : Actions de la PKA dans la signalisation induite par la FSH	268
Figure 93 : Actions de la p38 MAP Kinase dans la signalisation induite par la FSH	269
Figure 94 : Activation du FSHR, recrutement de la protéine Gi	269
Figure 95 : Voie β-arrestines	270
Figure 96 : Voie PI3K	270
Figure 97 : Activation de Src par la FSH	271
Figure 98 : Cascade ERK MAPK	272
Figure 99 : Régulation de la synthèse des hormones stéroïdiennes par la FSH	272
Figure 100 : Relations impliquée dans la régulation du cycle cellulaire	273
Figure 101 : Expression des protéines régulée par la signalisation induite par la FSH	274
Figure 102 : Actions impliquant le récepteur EGFR dans le réseau EGF	275
Figure 103 : Actions impliquant les protéines Grb2 et SOS dans le réseau EGF	276
Figure 104 : Voie PI3K induite par l'EGF	277
Figure 105 : Voie ERK MAPK induite par l'EGF (partie 1)	277
Figure 106 : Voie ERK MAPK induite par l'EGF (partie 2)	278
Figure 107 : Actions impliquant les protéines Gab dans le réseau EGF	278
Figure 108 : Actions impliquant les protéines Shc dans le réseau EGF	279
Figure 109 : Actions impliquant les sous-unités Gαi dans le réseau EGF	279
Figure 110 : Voie JAK/STAT induite par l'EGF	280
Figure 111 : Actions impliquant la protéine Src dans le réseau EGF	281
Figure 112 : Actions impliquant les protéines SHP dans le réseau EGF	281
Figure 113 : Signalisation au site d'adhésion focale	282

Liste des annexes

Annexe 1 : Relations formalisées non présentées de la partie « Matériel et Méthodes »	220
Annexe 2 : Création des règles non présentée dans la partie « Matériel et Méthodes »	224
Annexe 3 : Dictionnaire des données	234
Annexe 4 : Liste des règles	246
Annexe 5 : Régulation de la transcription et de la traduction par la FSH	258
Annexe 6: Modules	265

Liste des abréviations

binding phosphoprotein-50

EGFR:

EGF: Epidermal Growth Factor

Epidermal

Growth

3D: 3 dimensions Receptor 5HT: 5-hydroxytryptamine ELISA: Enzyme-Linked ImmunoSorbent 8-CPT-2ME-cAMP: 8-(4-Assay chlorophenylthio)-20-O-methyl adenosine-Exchange EPAC: Protein directly 30,50-cyclic monophosphate Activated by cAMP ERK: Extracellular signal-regulated kinase ADN: Acide Désoxyribonucléique AMP: Adénosine monophosphate FRET: Fluorescence Resonance Energy Adénosine AMPc: monophosphate Transfer cyclique *FSH* : Follicle Stimulating Hormone AMPK: AMP-activated protein Kinase FSHR: Follicle Stimulating Hormone AP-2: β 2-adaptin Receptor API: Application Programming Interface APPL1: Adaptor Protein, Phosphotyrosine GDP: Guanosine diphosphate GHS-R1a: growth hormone secretagogue interaction, PH domain and Leucine zipper receptor 1a containing 1 **GIP**: GPCR Interating Protein APPL2: Adaptor Protein, Phosphotyrosine **GnRH**: Gonadotropin Releasing Hormone interaction, PH domain and Leucine zipper GPR15: G protein-coupled receptor 15 containing 2 Grb2: Growth factor receptor-bound AT1R: Angiotensin 1 Receptor ATP: Adénosine triphosphate protein 2 AR: Adrenergic Receptor **GRK**: G protein Receptor Kinase GSK3: Glycogen synthase kinase 3 ARN: Acide Ribonucléique GTP: Guanosine triphosphate ARNm: ARN messager <u>hCG</u>: Human chorionic gonadotrophin βARK : β Adrenergic Receptor Kinase HEK293: Human Embryonic Kidney 293 BiNoM: Biological Network Manager HSP90: Heat Shock Protein 90 Biocham: Biochemical Abstract Machine **BioPax**: Biological Pathway Exchange *IGF*: Insulin-like Growth Factor IKK: IκB Kinase ChiBE: Chisio BioPax Editor IP: Immunoprécipitation **CLR**: Context Likelihood Relatedness IP3: Inositol 1,4,5-triphosphate cortBP1: cortactin-binding protein 1 CREB: cAMP Responsive element binding protein JAK: Janus kinase JNK: c-Jun-N-terminal kinase CXCR4: C-X-C chemokine receptor type KEGG: Kyoto Encyclopedia of Genes and Genomes D2R: Dopamine 2 Receptor KO: Knock Out *DAG* : Diacylglycerol LH: Luteinizing Hormone EA: Enzymatic Assay *EBP50*: erzin-radixin-moesin(ERM)-MAPK: Mitogen-activated protein kinase

Factor

MathML:

Language

Mathematical

Markup

Rap1: Ras-related protein 1 MEK: Mitogen-activated protein kinase RasGAP: Ras GTPase activating protein kinase mGLUR1: Metabotropic RCPG: Récepteur couplé aux protéines G glutamate Rheb: Ras homolog enriched in brain receptor 1 mTOR: mammalian Target Of Rapamycin RIA: RadioImmunoAssay mTORC1: mTOR complex 1 **RNAi**: ARN interférent *MOR* : μ-opioid receptor ROCK: Rho-associated protein kinase rpS6: ribosomal protein S6 RTK : Récepteur Tyrosine Kinase NFκB: nuclear factor-kappa B Reverse Transcription Nrf2: Nuclear factor (erythroid-derived RTPCR polymerase Chain Reaction 2)-like 2 SBGN: **Systems** Biology Graphical <u>ODE</u>: Ordinary Differential Equation Notation OWL: Web Ontology Language SBML: **Systems Biology** Markup *p70S6K* : p70S6 kinase Language SBO: Systems Biology Ontology p38MAPK: p38 Mitogen-activated protein SBW: Systems Biology Workbench kinase **PA**: Phosphorylation Assay SHP1: Src homology region 2 domaincontaining phosphatase-1 PAR2: Protease activated receptor 2 SHP2: Src homology region 2 domain-PDE: Partial Differential Equation *PDE*: Phosphodiesterase containing phosphatase-2 siRNA: petit ARN interférent <u>PDGFR</u>: Platelet-derived growth factor SNP : Single Nucleotide Polymorphism receptors SOLAR : Skip Ordered Linear for PDK: Phosphoinositide-Dependent Advanced Reasoning Kinase SOS: Son of Sevenless PDZ: PSD-95, dics-large et ZO-1 Src: du mot sarcome *PF* : Production Field STAT: Signal Transducers and Activators *PI3K*: Phosphatidyl inositol 3-kinase of Transcription PID: Pathway Interaction Database Sybil: Systems Biology Linker <u>PIP2 :</u> Phosphatidyl-inositol-4-5diphosphate *PIP3* : Phosphatidylinositol <u>TAP-tag</u>: Tandem Affinity Purification (3,4,5)triphosphate tag *PIP5K1C* : phosphatidylinositol-4-TPR1: Tetratricopeptide repeat 1 TSC2: Tuberous Sclerosis Protein 2 phosphate 5-kinase, type I, gamma *TSH*: Thyroid-stimulating hormone PKA: Protein Kinase A PKAc : Sous-unité catalytique de la PKA PKB: Protein Kinase B UML: Unified Modeling Language PKC: Protein Kinase C PLC: Phospholipase C *V2R*: Vasopressin type 2 Receptor PMA: Procréation médicalement assistée WB: Western Blot PP2A: Protein Phosphatase 2A PRA: Phosphorylation Radioactive Assay Wnt: fusion des mots wingless et Int

XML: eXtensible Markup Language

PTHR: Parathyroid Hormone Receptor

<u>PTP</u>: Protein Tyrosine Phosphatase <u>PYK2</u>: Protein tyrosine kinase 2

Introduction

I- Les récepteurs couplés aux protéines G

A- Généralités

Les récepteurs couplés aux protéines G (RCPG ou Récepteurs à sept domaines transmembranaires) constituent une vaste famille de récepteurs membranaires qui représente plus d'3% du génome chez les vertébrés. Par exemple, le génome humain contient plus de 800 gènes codant pour ce type de récepteurs [1].

Les RCPG jouent un rôle important dans la réception, le décodage et la transduction des signaux extracellulaires vers le milieu intracellulaire. De façon plus précise, un ligand extracellulaire se complexe avec le récepteur membranaire qui lui est spécifique et ce dernier va transmettre le message à l'intérieur de la cellule. En réponse à ce signal, la cellule s'adapte à son environnement.

Les RCPG sont des récepteurs présentant une grande diversité de ligands, ils sont capables de se lier à des molécules de natures très variées telles que des catécholamines [2], des eicosanoïdes [3], des peptides [4], des protéines [5] ou encore des ions [6]. Ces récepteurs sont présents dans un grand nombre d'espèces et sont exprimés de façon ubiquiste. La présence de cette famille de récepteurs dans un grand nombre de tissus leur confère un rôle important dans virtuellement tous les processus physiologiques, tels que la vision, l'immunité, la contraction musculaire et la reproduction.

Les RCPG présentent donc une grande diversité de fonctions biologiques et de ligands, ce qui leur confère un grand intérêt pharmacologique. Aujourd'hui, environ 40% des médicaments ciblent cette famille protéique [7]. Les RCPG sont capables d'induire des réseaux de signalisation cellulaire d'une telle complexité qu'il n'est actuellement pas possible de les décrire de façon exhaustive. La compréhension détaillée des mécanismes moléculaires et cellulaires engendrés par ces récepteurs pourrait permettre de développer des traitements capables de les moduler, de les activer ou de les inhiber totalement ou partiellement. Cette complexité pour être maîtrisée, nécessite de développer des méthodes de construction des réseaux moléculaires.

B-Structure

En 1986, Dixon et al [8] ont étudié la structure du récepteur β-adrénergique par analyse d'hydropathie, ce qui a révélé l'existence de sept hélices transmembranaires. Ce travail a ensuite été effectué, en 1992, pour 39 RCPG, aboutissant au même résultat [9]. Cependant, c'est à partir d'expériences de cryo-microscopie électronique et de structure cristalline haute résolution obtenue par diffraction aux rayons X, que la structure de la rhodopsine fut validée [10, 11] (structure cristalline en 3D disponible dans la banque de données *Protein DataBank*: http://www.pdb.org/pdb/explore/explore.do?pdbId=1F88) (Figure 1). Ensuite, différentes structures de RCPG ont été découvertes telles que la structure cristalline de l'hormone folliculo-stimulante complexée à son récepteur, du récepteur β2 adrénergique complexé avec un agoniste de l'adrénaline ou encore complexé avec une protéine G stimulatrice, de la rhodopsine non stimulée [12-15].

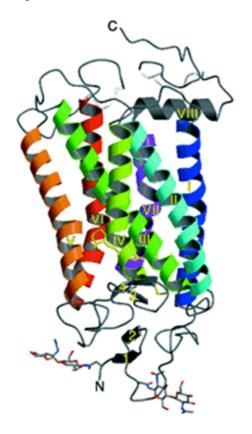


Figure 1 : Représentation 3D en 'ruban' de la structure de la Rhodopsine [11]

Vue en parallèle de la membrane plasmique

Les RCPG présentent donc une structure commune avec sept hélices α transmembranaires reliées par trois boucles intracellulaires et trois boucles extracellulaires (Figure 2). La queue amino-terminale (ou N-terminale) du récepteur se situe au niveau extracellulaire et est

impliquée dans la fixation du ligand. Les sept hélices α transmembranaires participent à la transduction du signal depuis le domaine extracellulaire vers le domaine intracellulaire. La queue carboxy-terminale (ou C-terminale) est intracellulaire et contient une huitième hélice α [11, 16]. Cette partie du récepteur présente également de multiples sites de phosphorylation. Cette queue C-terminale ainsi que les boucles intracellulaires sont impliquées dans l'activation de voies de signalisation et dans la désensibilisation du récepteur.

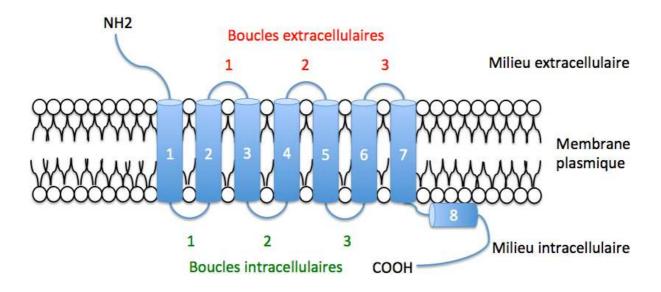


Figure 2 : Structure générale en 2D d'un RCPG

La structure générale est commune à tous les RCPG mais il existe tout de même des différences, ce qui a permis de mettre au point une classification.

C- Classification

Les RCPG ont été classés avant la découverte de leur structure tridimensionnelle. Cette classification repose sur des alignements de séquences mais aussi en prenant en compte leurs différents ligands. Bockaert et Pin [17] proposent une classification en six grandes familles (Figure 3).

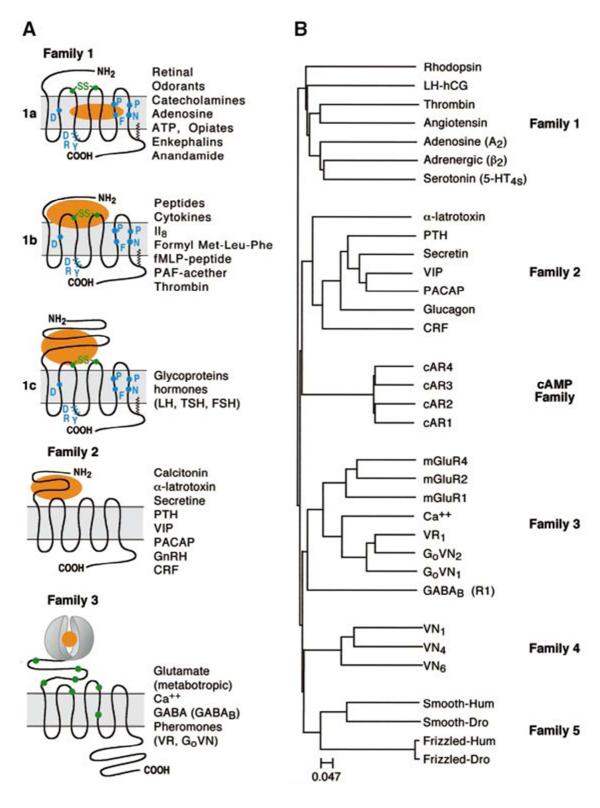


Figure 3: Classification des RCPG selon [17, 18]

(A) Représentation des 3 principales familles dont le site de fixation du ligand a été caractérisé. (B) Arbre phylogénétique des 6 familles de RCPG

<u>Famille 1 :</u> Cette famille est aussi connue sous le nom de *rhodopsin-like*. Les RCPG de cette famille présentent quelques résidus conservés au niveau des hélices α, par exemple, ils contiennent un motif E/D-R-Y/W au niveau de la deuxième boucle intracellulaire. De plus, il existe un pont disulfure entre les boucles extracellulaires 1 et 2.

Cette famille est subdivisée en trois groupes en fonction du site de fixation et de la taille des ligands. Le groupe 1a présente un site de fixation au niveau des hélices α transmembranaires, les récepteurs de ce groupe sont capables de fixer des ligands de petite taille. Le groupe 1b présente un site de fixation du ligand à cheval entre la queue N-terminale, les boucles extracellulaires et la partie supérieure des hélices transmembranaires. Ces récepteurs ont pour ligands des peptides. Le groupe 1c présente une queue N-terminale bien plus longue que les autres et celle-ci prend en charge un ligand de grande taille tel que des hormones glycoprotéiques. Cette extrémité N-terminale est dans ce cas un domaine globulaire.

Pour cette famille de RCPG, deux « poches » de fixation des ligands ont été déterminées de façon plus précise. La première, qui concerne le groupe 1a, est située entre les hélices 3, 4, 5, 6 et 7 et la deuxième, qui concerne le groupe 1c, est localisée entre les hélices 1, 2, 3 et 7 [19].

Famille 2 : Cette famille est aussi connue sous le nom de *secretin-like*. Elle ne présente pas les acides aminés conservés retrouvés dans la famille 1. Ces récepteurs présentent une queue N-terminale structurée de façon à se lier à des peptides de grande taille tels que le glucagon. Le site de fixation du ligand se situe en partie sur la première boucle extracellulaire.

<u>Famille 3 :</u> La troisième famille de RCPG présente une queue C-terminale plus longue que les autres familles. La particularité de cette famille est la présence d'une structure dite « *Venus Flytrap* » (référence à la plante carnivore du même nom) qui prend en charge la fixation du ligand, neuropeptide ou phéromone.

<u>Famille 4 :</u> Cette famille de RCPG est très peu connue. Ils n'ont été localisés qu'au niveau de neurones sensoriels de l'organe voméronasal et pourraient avoir pour ligand des phéromones [18]. Une étude sur le récepteur VN1R montre qu'il se lie à un ligand volatile et présente une fonction proche de celle des récepteurs olfactifs [20].

Famille 5 : Cette famille est composée de récepteurs homologues de *frizzled* et de *smoothened*, qui sont des protéines impliquées dans le développement de la Drosophile [18]. Les récepteurs *frizzled-like* ont pour ligand des glycolipoprotéines Wnt qui contrôlent le

développement embryonnaire et l'homéostasie chez l'adulte [21]. Quant au récepteur *smoothened*, il a la particularité d'être soumis à une inhibition par un autre récepteur *patched*. Cette inhibition est levée par la fixation de la protéine *Hedgehog* sur *patched* (Figure 4) [18]. Cette protéine *Hedgehog* est impliquée dans le développement embryonnaire.

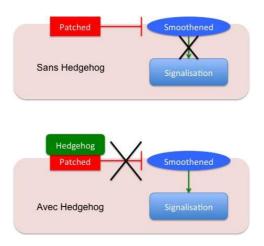


Figure 4 : Inhibition du récepteur Smoothened par le récepteur Patched

Famille AMPc: Cette famille de RCPG a été caractérisée chez *Dictyostelium discoideum* [22], *Dictyostelium minutum* [23], *Polysphondylium pallidum* et *Dictyostelium aureostipes* [24]. Ce type de récepteur est impliqué dans le développement de l'organisme et la chimiotaxie [24, 25]. Il n'existe pas d'homologue chez les vertébrés.

Les RCPG sont donc une superfamille de récepteurs présentant des caractéristiques communes mais aussi des différences structurales. Néanmoins, les mécanismes de signalisation cellulaire induits par ces récepteurs restent similaires.

D- Mécanismes de signalisation cellulaire

Tous les RCPG induisent des voies de signalisation cellulaire lorsqu'ils sont activés par la fixation de leur ligand. Les phénomènes observés sont l'activation d'une voie dépendante des protéines G, suivie d'une désensibilisation des récepteurs à cette voie. Cette désensibilisation est suivie de l'internalisation du RCPG et de l'activation de voies de signalisation indépendantes des protéines G (Figure 5).

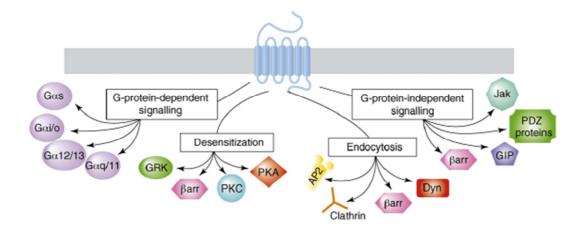


Figure 5 : Schématisation des différents mécanismes induits par l'activation d'un RCPG [26]

L'activation du récepteur induit la voie G dépendante, sa désensibilisation, l'endocytose du récepteur mais également des voies de signalisation indépendantes des protéines G.

1- Signalisation dépendante des protéines G

Les RCPG, comme leur nom l'indique, sont capables de se coupler aux protéines G qui ont une activité GTPase. Il existe deux théories quant à ce couplage, la première est le précouplage RCPG-Protéine G et la deuxième est la « rencontre » entre le récepteur activé et la protéine G par diffusion grâce aux mouvements membranaires. Les expériences de Mueller et al [27] ont montré une amplification de l'activation des protéines G par un seul récepteur. Cette amplification montre qu'il n'y a pas une seule protéine G activée par le récepteur mais plusieurs, ce qui signifie que le complexe récepteur-protéine G n'est pas figé, donc cela favorise la théorie de la rencontre par diffusion (Figure 6).

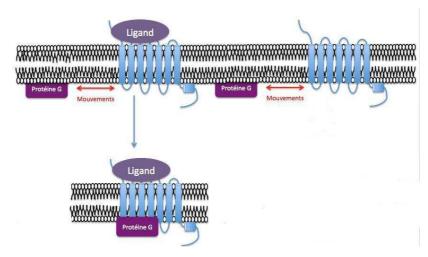


Figure 6 : Rencontre d'un RCPG activé avec une protéine G

Le récepteur activé par la fixation de son ligand et la protéine G se rencontrent grâce aux mouvements des lipides constituant la membrane plasmique. Alors qu'un RCPG inactif ne se lie pas à la protéine G.

La fixation du ligand à son récepteur provoque un changement conformationnel de ce dernier, ce qui explique qu'un récepteur actif est capable de se complexer à une protéine G alors qu'un récepteur inactif n'en est pas capable.

Les molécules communément nommées protéines G sont en réalité des complexes hétérotrimériques composés d'une sous-unité α , une sous-unité β et une sous-unité γ (Figure 7) [28, 29].

La protéine G dans son état inactif est liée à une molécule de GDP (Figure 7). Lorsqu'elle est activée par fixation à un RCPG actif, la molécule de GDP est échangée contre une molécule de GTP, ce qui provoque la dissociation de la protéine G en deux parties, un complexe composé de la sous-unité α liée au GTP et un autre complexe composé des sous-unités β et γ . Ainsi libérée, la sous-unité α agit sur ses cibles jusqu'à ce que le GTP qu'elle porte soit hydrolysé en GDP, ce qui provoque la reformation de l'hétérotrimère inactif de proteine G.

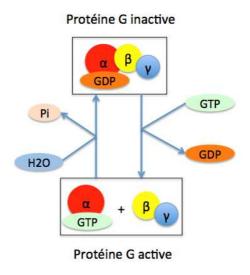


Figure 7 : Mécanismes d'activation et de désactivation d'une protéine G (RCPG non représenté)

Il existe différents types de sous-unités α qui ont différents effecteurs. Des alignements de séquences de ces sous-unités α ont été réalisés, ce qui a permis de construire un arbre phylogénétique qui permet d'obtenir les regroupements suivants [30] :

- Gαs (stimulation) / Gαolf (olfaction),
- Gαi (inhibition) / Gα0 / Gαz / Gαt (transducine),
- $-G\alpha q / G\alpha 11 / G\alpha 14 / G\alpha 15 / G\alpha 16$,
- Gα12 / Gα13.

Les différentes protéines G agissent sur un grand nombre d'effecteurs (Figure 8). La sous-famille Gαs/Gαolf induit la production d'AMPc grâce à l'activation de l'adénylate cyclase [31-36]. L'AMPc est un second messager qui active la PKA et EPAC par complexation [37-40]. Ces derniers participent notamment à l'activation de la cascade des MAP Kinase ERK [40-42]. La PKA est également un régulateur de la transcription et de la traduction grâce à son action sur CREB et sur p70S6Kinase respectivement. De plus, cette enzyme est capable d'auto-régulation par activation des phosphodiestérases qui dégradent l'AMPc en AMP [32, 43].

La sous-famille Gαi/0/t/z induit une inhibition de la production d'AMPc mais aussi une activation de la cascade MAP Kinase ERK, la cascade Wnt/JNK, les signalisations dépendantes de Src et PI3K [44-51].

La sous-famille $G\alpha 12/13$ induit une signalisation dépendante de Src, de Rho/ROCK [52, 53], mais aussi la voie JNK qui conduit à l'activation du facteur de transcription Nrf2 [54] et la voie JAK/STAT [55].

La sous-famille $G\alpha q/11/14/15/16$ induit la production des seconds messagers IP3 et DAG par l'intermédiaire de la phospholipase C [56-62]. IP3 induit l'ouverture de canaux ioniques Ca^{2+} et ce qui provoque l'augmentation de la quantité des ions Ca^{2+} dans le cytosol [63]. Les ions Ca^{2+} et DAG sont nécessaires pour l'activation de l'enzyme PKC [64]. La sous-famille $G\alpha q$ est capable d'activer différentes voies de signalisation impliquant notamment Rho, p38MAPK, MAP Kinase ERK et IKK [65, 66], d'inhiber la voie PI3K [67, 68] et de réguler la transcription en agissant sur les facteurs NF κ B et STAT [66].

Pendant très longtemps, les sous-unités α ont été considérées comme étant les seules sousunités des protéines G à être capables de déclencher des voies de signalisation, or il s'est avéré que le complexe $\beta\gamma$ est impliqué dans l'ouverture des canaux potassiques [69] et dans le recrutement des GRK pour phosphoryler les RCPG [70, 71]. De plus, complexe $\beta\gamma$ induit l'activation du complexe PI3K et des phosphodiestérases [72, 73].

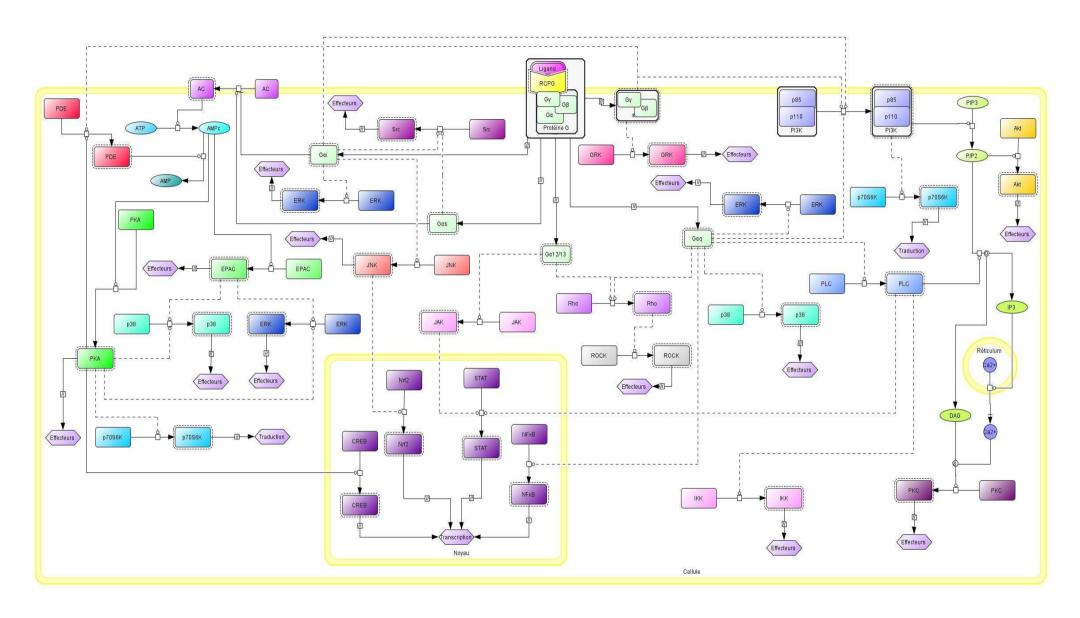


Figure 8 : Voies de signalisation induites par les différentes voies G

Les sous-unités Gαs et Gαi sont capables d'activer la protéine Src directement (Figure 8) [74]. Cette protéine est très importante car elle est impliquée dans différentes voies de signalisation, notamment la transactivation des récepteurs des facteurs de croissance par les RCPG [75-77].

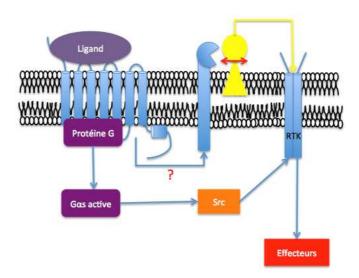


Figure 9 : Représentation schématique d'une transactivation d'un récepteur tyrosine kinase

Un RCPG actif induit l'activation de Src via la sous-unité $G\alpha$ s de la protéine G. Src phosphoryle ensuite le récepteur tyrosine kinase (RTK). Cette étape n'est pas la seule induite par un RCPG car il est aussi capable d'activer des métalloprotéases, qui sont des protéines transmembranaires (le mécanisme exact est inconnu). Ces métalloprotéases (représentées par une partie transmembranaire et une tête de Pacman) sont capables de cliver un pré-facteur lié à la membrane en facteur libre (site de clivage représenté par la double flèche rouge). Ce facteur vient ensuite se fixer à son récepteur, ce qui avec la phosphorylation induite par Src induit son activation.

Un récepteur est transactivé si le signal initial est reçu par un autre récepteur et conduit à son activation. Les récepteurs des facteurs de croissance (récepteurs tyrosines kinases) doivent être phosphorylés pour être totalement activés. Src participe donc à cette transactivation car elle phosphoryle les récepteurs tyrosines kinases sur des résidus tyrosines.

Un RCPG peut activer préférentiellement un type de protéine G mais peut également activer un ou plusieurs autres types tels que :

- Récepteur dopaminergique 1 (Gαs, Gαolf) [78]
- Récepteur à l'adénosine A2aR (Gαolf) [36, 78],
- Récepteur de la FSH (Gas, Gai) [45, 79],
- Récepteurs β adrénergiques (Gαs) [80],
- Récepteurs de type Frizzled-like (Gi/o) [49],
- Récepteur adrénergique α2 (Gz) [47],
- Récepteur dopaminergique 4 (Gαi, Gαo et Gαt) [48, 78].
- Récepteur dopaminergique 5 (Gas,Gaq) [78],
- Récepteur de la sérotonine 5HTR2a (Gα11) [81],

- Récepteur de l'angiotensine AT1a (Gαq) [82],
- Récepteur de la thyrotropine ($G\alpha 15$ et $G\alpha 16$) [83, 84],
- Récepteur purinergique P2Y2 (Gα15) [63].
- Récepteur du thromboxane TXA₂ (Gα12/13) [52],
- Récepteur de l'hormone thyroïdo-stimulante (Gα12/13) [54].

La voie G est ensuite désensibilisée, c'est-à-dire que des mécanismes se mettent en place pour arrêter cette voie de signalisation.

2- Désensibilisation des récepteurs

Un court instant après l'activation d'un RCPG, des mécanismes de régulation sont enclenchés afin d'empêcher une production excessive de seconds messagers, c'est la désensibilisation. Cette désensibilisation peut être réalisée suivant deux procédés :

- Elle peut être ligand-indépendante, c'est la désensibilisation hétérologue,
- Elle peut être ligand-dépendante, c'est la désensibilisation homologue.

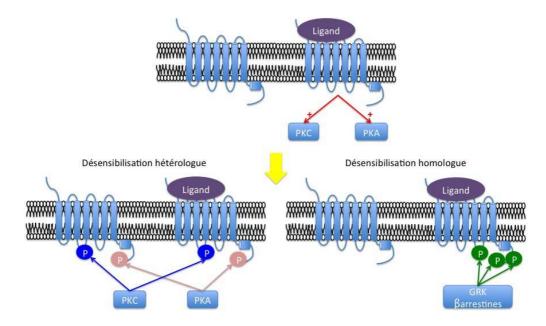


Figure 10 : Représentation schématique des désensibilisations hétérologue et homologue.

Les kinases PKA et PKC activées par le RCPG, viennent le phosphoryler. Elles phosphorylent le récepteur à l'état actif mais aussi à l'état inactif, c'est la désensibilisation hétérologue. Les GRK phosphorylent uniquement le récepteur à l'état actif, c'est la désensibilisation homologue. Reproduction à partir de [85].

<u>Désensibilisation hétérologue</u>: Ce type de désensibilisation des RCPG fait intervenir des kinases dont l'activation est dépendante des seconds messagers, la PKA et la PKC (Figure 10). Ces kinases phosphorylent les RCPG sur les résidus sérines et thréonines ce qui entraîne une impossibilité de couplage des protéines G. Ce processus est défini comme une désensibilisation hétérologue car elle n'est pas spécifique aux RCPG activés, les RCPG

inactifs sont également soumis à cette désensibilisation (Figure 10). Certains RCPG sont phosphorylés uniquement par une des deux kinases. C'est le cas pour le récepteur muscarinique de l'acétylcholine, le récepteur dopaminergique 3 ou encore le récepteur P2Y1 [86-88].

Les travaux de Bouvier et al [89] ont montré que la suppression des sites de phosphorylation du récepteur $\beta 2$ adrénergique induit un retard mais pas une abolition de la désensibilisation. Ceci signifie qu'il existe un autre mécanisme impliqué dans la désensibilisation des RCPG.

Désensibilisation homologue : La désensibilisation des RCPG n'est pas contrôlée seulement par la PKA et/ou la PKC. *In vitro*, le récepteur β adrénergique est phosphorylé et montre une activation de la voie dépendante des protéines G réduite en présence de la kinase βARK (GRK) purifiée [90]. Cette famille protéique des GRK est composée de sept membres, de GRK1 à GRK7. GRK1 et GRK7 sont localisées uniquement dans la rétine et GRK4 dans les testicules [91, 92]. Les autres GRK sont exprimées de façon ubiquiste.

L'activation d'un RCPG induit le recrutement à la membrane des protéines GRK2 et GRK3. Elles se lient au complexe formé par les sous-unités βγ de la protéine G activée et sont ainsi recrutée à proximité des RCPG activés. Les GRK5 et GRK6 sont constitutivement présentes à la membrane et viennent interagir avec le RCPG actif. Ce mode de désensibilisation dépend de l'activation du RCPG par son ligand, elle est donc appelée désensibilisation homologue ou encore désensibilisation ligand-dépendante (Figure 10).

Les GRK phosphorylent des sérines et des thréonines au niveau de la queue C-terminale et de la troisième boucle intracellulaire du récepteur. Ces phosphorylations induisent le recrutement des β-arrestines qui sont des protéines adaptatrices. Ces protéines, comme les GRK, entrent en jeu uniquement dans le cas de la désensibilisation homologue. Elles viennent se fixer aux récepteurs phosphorylés par les GRK et empêchent le couplage aux protéines G [93]. De plus, elles sont impliquées dans le processus d'internalisation des RCPG [94].

3- Internalisation et devenir des RCPG

L'internalisation des RCPG est une conséquence de l'action des GRK et des β -arrestines. Elle se déroule de la façon présentée sur la Figure 11. Les β -arrestines permettent la fixation de la clathrine et d'AP-2. Ceci entraı̂ne la formation d'une vésicule avec un « manteau » de clathrine. Lors de ce processus, la dynamine est recrutée afin de « fermer » la vésicule.

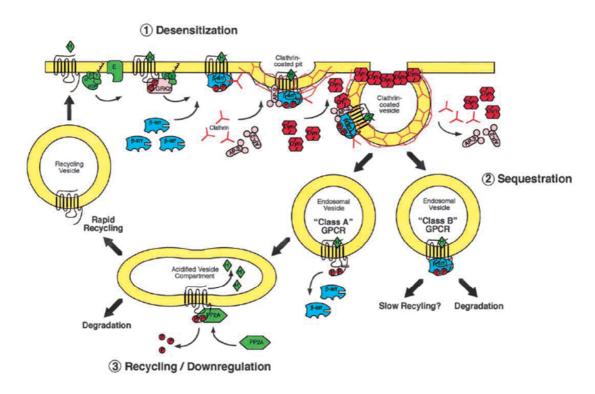


Figure 11 : Internalisation et devenir intracellulaire des RCPG [95]

Une fois la vésicule d'endocytose formée (endosome), les β-arrestines se dissocient des RCPG, ce qui va permettre à des phosphatases de déphosphoryler les récepteurs. En parallèle, les endosomes s'acidifient, ce qui permet la dissociation entre le ligand et son récepteur. Lorsque cette étape est réalisée, le récepteur peut être soit recyclé à la membrane plasmique pour recommencer un cycle d'activation, soit dégradé. La poly-ubiquitination des protéines conduit à leur adressage au protéasome, système de dégradation. Si le récepteur est poly-ubiquitiné, il sera dégradé, sinon il sera recyclé.

Les GRK et les β -arrestines sont ainsi des acteurs directs de l'endocytose des RCPG mais aussi d'une signalisation indépendante des protéines G.

4- Signalisation indépendante des protéines G

La voie dépendante des protéines G a longtemps été considérée comme l'unique voie de transduction du signal de cette famille de récepteurs. Mais il existe aussi différents modes de signalisation indépendants des protéines G notamment via les β -arrestines.

En fonction des motifs de phosphorylation des RCPG par les GRK, les β -arrestines adoptent des conformations différentes. Il y a un « code barre » sur le récepteur, imposé par les GRK

qui a pour effet de diriger les β -arrestines dans la conformation adaptée afin de se diriger vers un évènement cellulaire ou un autre [82, 96].

Différents travaux ont montré que la déplétion en GRK2/3 par siRNA renforce fortement le signal MAP Kinase ERK et la déplétion en GRK5/6 affaiblit ce signal. Ceci a été réalisé pour les récepteurs AT1AR, V₂ et le récepteur FSHR [97-99]. Heitzler et al [82] ont également montré ce découplage entre les GRK2/3 et GRK5/6 pour les récepteurs de l'angiotensine, le récepteur β2 adrénergique, le récepteur de la vasopressine V₂, le récepteur de la FSH et le récepteur de la neurokinine 1. Ces travaux montrent que les GRK2/3 inhibent la transduction du signal par les β-arrestines alors que les GRK5/6, au contraire, la stimulent.

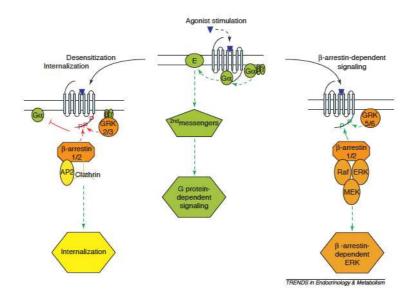


Figure 12 : Modèle de spécialisation fonctionnelle des différentes GRK [100]

Les RCPG activent une voie de signalisation dépendante des protéines G avec production de seconds messagers, une désensibilisation et une internalisation via les GK2/3 et les β -arrestines ainsi qu'une voie de signalisation dépendante des β -arrestines.

La Figure 12 montre le modèle que Reiter et Lefkowitz ont proposé quant à la fonction des différentes GRK [100]. Les GRK2/3 sont impliquées dans la désensibilisation et l'internalisation des RCPG, ce qui explique qu'elles inhibent le signal MAP Kinase ERK dans les expériences précédemment citées, alors qu'au contraire, les GRK5/6 sont impliquées dans l'activation de la voie de signalisation MAP Kinase ERK. Les RCPG sont donc capables d'induire une signalisation dépendante des protéines G mais également une signalisation indépendante des protéines G.

Tableau 1 : Signalisation dépendante des β -arrestines induite par les RCPG (reproduction à partir de [101])

 \uparrow = activation, \downarrow inhibition

RCPG	Activité β-arrestines dépendante
β2AR	ERK1/2 ↑
	NF-κB↓
A_2AR	Src ↑
	ERK1/2 ↑
AT1R	ERK1/2 (β-arrestine 2) ↑
	RhoA (β-arrestine 1) ↑
	Jnk ↑
PAR-2	ERK1/2 ↑
	Cofilin ↑
	LIMK ↓
	PI3K↓
PTHR	ERK1/2 ↑
V2R	ERK1/2 ↑
CXCR4	p38MAPK ↑
D2R	Akt↓
MOR	ERK1/2 (β-arrestine 2) ↑
mGLUR1	ERK1/2 ↑
GHS-R1a	ERK1/2↑

<u>β-arrestines</u>: De nombreux travaux ont été réalisés afin de connaître l'implication des β -arrestines dans les voies de signalisation induites par les RCPG. Dans une revue, Defea [101] a répertorié quelques régulations des β -arrestines sur différentes voies (Tableau 1).

D'autres travaux ont également montré que les β -arrestines sont impliquées dans l'activation des MAP Kinase ERK et de la rpS6 dans le réseau induit par la FSH [99, 102]. Le tableau montre que les β -arrestines activent la kinase Src dans le cas des récepteurs $\alpha 2$ adrénergiques, ce qui est démontré également par les travaux de Chun et al [103] dans le cas du récepteur Prostaglandine E2.

Une étude sur l'influence des β -arrestines dans la signalisation induite par les RCPG a été menée sur le phosphoprotéome induit par ces β -arrestines en aval du récepteur AT1aR et cela montre une grande diversité d'actions de ces protéines (Figure 13) [104]. Les β -arrestines agissent dans différents mécanismes cellulaires tels que la voie MAP Kinase ERK et la voie PI3K.

Les études sur ces différents RCPG montrent l'étendue de la signalisation dépendante des β -arrestines. En fonction des voies qu'elles activent, elles sont impliquées dans divers processus cellulaires, tels que le cycle cellulaire ou la réorganisation du cytosquelette.

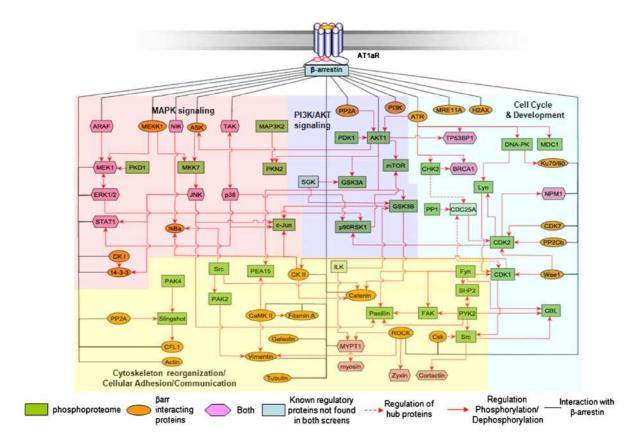


Figure 13 : Inférence du réseau de kinases dépendantes des β -arrestines en aval du récepteur AT1aR

Construction à partir du phosphoprotéome dépendant des β -arrestines et de l'interactome (issue de [104]) Rectangle vert : protéine identifiée dans le phosphoprotéome induit par les β -arrestines, ellipse orange : protéine interagissant avec les β -arrestines, hexagone rose : protéine identifiée dans l'interactome et le phosphoprotéome des β -arrestines, rectangle gris : protéines régulatrices connues mais non identifiées dans l'interactome et le phosphoprotéome des β -arrestines, flèche rouge : régulation des phosphorylation et déphosphorylation, ligne noire : interaction avec les β -arrestines, en pointillés : régulations des protéines régulatrices.

Autres voies de signalisation : Différentes études ont permis de mettre en évidence des interactions d'autres molécules avec les RCPG et qui pourraient induire une signalisation indépendante des protéines G et des β -arrestines. Par exemple :

<u>Src</u>: L'activation de Src est complexe. Il a été expliqué plus haut que cette protéine peut être activée via les protéines G, via les β -arrestines mais d'autres études montrent aussi que son activation pourrait se faire directement via le RCPG lui-même, par exemple les récepteurs β 2 adrénergiques et le récepteur à la sérotonine 5HT₄ [105, 106].

<u>APPL1 et APPL2</u>: Les recherches sur le récepteur FSHR montrent une complexation avec APPL1 [107] et avec APPL2 [108]. Il est donc possible que ces interactions induisent des

voies de signalisation, d'autant que le complexe APPL1:FSHR induit la production d'IP3 [109]. Cette interaction n'est cependant pas dépendante de la fixation de la FSH à son récepteur.

<u>14-3-3</u>: Des études montrent que les RCPG peuvent se complexer avec la protéine 14-3-3 (PAR2, FSHR, GPR15) [110-112]. Le rôle exact de cette complexation n'est pas connu mais 14-3-3 est importante car elle est impliquée dans un grand nombre d'évènements cellulaires.

<u>Protéines à domaine PDZ</u>: Des travaux ont montré que le récepteur de la somatostatine de type 2 est capable de se lier au domaine PDZ de la protéine cortBP1 [113]. Ce récepteur interagit avec la cortactine et les filaments d'actine, ce qui révèle un lien entre les RCPG et le cytosquelette. La protéine à domaine PDZ, EBP50 se lie au récepteur β2 adrénergique et participe à son recyclage à la membrane plasmique [114].

E- Récepteur de l'hormone folliculo-stimulante (FSHR)

La signalisation induite par l'hormone FSH est au cœur des recherches au sein de l'équipe BIOS. C'est une hormone glycoprotéique composée de deux sous-unités, α qui est commune avec la LH (Hormone lutéinisante), la TSH (Thyréostimuline) et la hCG (Gonadotropine chorionique), et une sous-unité β qui lui est propre [115, 116]. La FSH est sécrétée par l'hypophyse sous l'action de la GnRH et circule dans le sang jusqu'aux cellules présentant le récepteur adapté, FSHR, qui est un RCPG [79]. Ces cellules cibles sont les cellules de Sertoli dans les testicules et les cellules de la Granulosa dans les ovaires [117]. La FSH participe au contrôle de la reproduction. Elle permet le bon déroulement de la spermatogenèse chez les mâles ainsi que la croissance et la maturation folliculaires chez les femelles [118].

Une mutation de la FSH peut avoir pour conséquence une forte diminution de la fertilité de l'individu ou une stérilité. Lors d'un Knock Out (KO) du gène codant la sous-unité β de la FSH chez des souris, les femelles sont stériles [119]. De plus, lorsque le gène de la sous-unité β de la FSH est muté, les femmes peuvent présenter une aménorrhée primaire alors que les hommes peuvent montrer une azoospermie. Si c'est le gène codant pour le récepteur qui est muté, les femmes présentent une insuffisance ovarienne primaire et les hommes une altération de la spermatogenèse [120]. De même, des expériences de KO du gène FSHR montrent un défaut de développement testiculaire chez les souris [121]. Ceci montre bien que si la FSH ou

son récepteur est déficient, les individus présentent une fertilité altérée, voire même une infertilité totale.

Du fait de son importance dans le contrôle de la reproduction, la FSH est utilisée dans le cadre de la procréation médicalement assistée (PMA). Les femmes sont traitées avec des hormones gonadotropes : FSH et LH ou hCG, ce qui peut avoir pour conséquence l'apparition d'un syndrome d'hyperstimulation ovarienne qui, dans sa forme la plus sévère peut être mortel [122].

Dans le domaine de la pharmacologie, que ce soit en médecine de la reproduction ou en zootechnie, il serait intéressant de créer des molécules capables de mimer les effets de la FSH sans conduire aux mêmes complications.

Pour répondre à une telle ambition, il faut d'abord être en mesure de connaître les mécanismes cellulaires induits par la FSH et notamment les mécanismes de signalisation. Une revue sera présentée dans la partie « Signalisation induite par l'hormone folliculo-stimulante (FSH) » représentant une cartographie non exhaustive de la signalisation induite par la FSH.

II- Construction de réseaux biologiques : Etat de l'art

Un réseau est un ensemble d'éléments en interaction les uns avec les autres, les éléments sont appelés nœuds, reliés par des arcs (les interactions). Dans le domaine de la biologie, on distingue différents types de réseaux :

- Moléculaires (régulations géniques, métaboliques, voies de signalisation) qui représentent les interactions entre les molécules,
- Neuronaux qui représentent les interactions entre les neurones,
- Phylogénétiques qui représentent des interactions évolutives, pour indiquer quelle espèce dérive de quel ancêtre,
- Ecosystémiques qui représentent les interactions dans un écosystème ou entre les écosystèmes.

Ces travaux de thèse reposent sur l'inférence de modèles de voies de signalisation, c'est pourquoi seules les méthodes de construction des réseaux moléculaires seront présentées.

Les réseaux moléculaires sont utilisés pour répondre à différentes problématiques. Ils permettent de décrire, d'identifier et de comprendre des systèmes biologiques tels que les réseaux de gènes ou les voies de signalisation. Ils peuvent aider dans la prédiction de la fonction des gènes ou pour élucider les propriétés des protéines. Ils permettent de prédire le comportement d'un système en cas de perturbations et d'identifier les molécules touchées par ces perturbations. Ils peuvent servir de base pour l'identification des interactions physiques entre les composés. Et ils participent à la découverte de nouvelles cibles thérapeutiques.

Il est possible de classer les différentes méthodes utilisées dans l'acquisition de réseaux moléculaires en trois groupes distincts. D'une part, il y a la construction manuelle, c'est-à-dire la description manuelle de ce qui est connu. D'autre part, il y a la construction semi-automatique reposant sur des outils de fouille de texte. Pour finir, il y a la construction automatique. Cette approche repose sur l'élaboration de programmes informatiques pour construire les réseaux. De plus, certaines méthodes conduisent à la prédiction d'évènements inconnus et à la génération d'hypothèses à partir de ce qui est connu.

A- Approche manuelle

Les biologistes peuvent construire des réseaux moléculaires à partir de la littérature scientifique et/ou des banques de données.

1- Bibliographie

Un chercheur peut créer manuellement son propre réseau à partir de la littérature, de ses propres expériences et de son expertise dans le domaine.

Il existe des réseaux construits manuellement, comme par exemple, Caron et al [123] ont construit un réseau concernant mTOR et ses régulations. D'autres chercheurs voulaient connaître le réseau moléculaire des cellules souches manuellement de façon à valider leur méthode automatique [124]. Nous avons nous-même réalisé une cartographie de la signalisation induite par la FSH à partir de la littérature, afin d'avoir un modèle permettant d'évaluer l'inférence automatique de ce même réseau [125]. Nous n'avons d'ailleurs pas été les seuls à construire ce réseau manuellement [126]. Une étude a également été réalisée sur le contrôle de la traduction par les RCPG [127].

Malheureusement, cette méthode est très coûteuse en temps et soulève le problème des biais dus à l'interprétation et aux choix du chercheur. Le chercheur se réfère à ce qu'il connaît et peut donc passer à côté d'éléments importants du réseau qui l'intéresse. L'exemple concret de la limitation de l'exploitation des données réside dans les expériences à haut débit, l'expérimentateur regarde les quelques gènes ou protéines qu'il connaît et ne traite pas le reste qui peut cacher des données importantes. Dans le cas d'une puce à ADN pangénomique, il est difficile d'analyser l'expression de milliers de gènes sans oublis, ni erreurs. De plus, il y a des possibilités d'erreurs dues à l'interprétation des chercheurs. Par exemple, Gonzalez-Robayna et al [128] présentent un *Western Blot* de la protéine Raf1 qui montre que sous l'action de la FSH, son expression augmente. Les auteurs affirment que la FSH active la protéine Raf1 or cette protéine pour être activée doit être phosphorylée sur la sérine 338 et l'expérience montre uniquement une augmentation de la quantité de Raf1.

2- Banques de données

Il existe des banques de données généralistes qui regroupent un grand nombre d'informations sur les interactions moléculaires. La Figure 14 présente quelques exemples de banques de données où des informations sur les réseaux peuvent être extraites. Par exemple, il est possible d'utiliser les voies métaboliques présentées dans la banque de données KEGG ou

encore les voies de signalisation décrites dans Reactome. Certains chercheurs combinent même les informations issues de différentes banques de données pour créer les réseaux qui les intéressent [129].

Category	Content	Examples
Protein-protein interactions	Mainly pairwise interactions between proteins	DIP, BIND, STRING, HPRD
Metabolic pathways	Biochemical reactions in metabolic pathways	KEGG, Reactome, ENZYME
Signaling pathways	Molecular interactions and chemical modifications in regulatory pathways	STKE, Reactome, TRANSPATH
Transcription factors/Gene regulatory networks	Transcription factors and the genes they regulate	JASPAR, TRANSFAC, RegulonDB
Pathway diagrams	Hyperlinked pathway images	KEGG, HPRD, SPAD
Protein-compound interactions	Interactions between proteins and compounds	ResNet, CLiBE
Protein sequence focused	Diverse pathway information in relation with sequence data	REBASE
Genetic interaction networks	Genetic interactions, such as epistasis	BIND, BioGRID

Figure 14 : Exemples de banques de données répertoriant des interactions moléculaires (Tableau issu de [130])

Il existe également des banques de données spécialisées dans un groupe taxonomique particulier, par exemple, RegulonDB regroupe toutes les informations concernant *Escherichia coli*.

Certaines équipes de recherche ont tout de même remis en question la qualité de ces banques de données. C'est pourquoi des tests de cohérence ont été réalisés, en confrontant ces données à des profils d'expression afin d'évaluer la complétude et la qualité des informations. Dans le cas d'*Escherichia coli*, un test de cohérence a été fait entre des profils d'expression de gènes acquis suivant différentes conditions et la base RegulonDB construite via les connaissances issues de la littérature, mais les résultats se sont montrés peu satisfaisants [131]. Ce genre d'observations ouvre de nouvelles discussions concernant la concordance entre les connaissances stockées et les expériences. Dans ce but, Shin et al [132] ont développé une méthode afin de tester automatiquement la cohérence entre des données de puces à ADN et les connaissances.

Un autre problème qui peut être rencontré dans les banques de données est la traçabilité de l'information, en particulier l'absence d'un lien entre l'interaction présentée et une référence bibliographique. Par exemple, dans la banque de données Transpath version Professional 2008.4, 88892 interactions étaient liées à une ou des références Pubmed sur un total de

161813 interactions. Il y avait donc un problème de traçabilité pour 46% du contenu de cette ancienne version de Transpath [133].

De temps à autre, les banques de données généralistes peuvent également contenir des erreurs : par exemple, une interaction peut être décrite avec un lien vers sa source bibliographique et, lorsque la source est étudiée, l'interaction et les molécules ne sont pas présentes dans la source donnée. Notamment, une erreur de saisie de la référence peut se produire en cas de curation manuelle des données.

3- Conclusion sur l'approche manuelle

La construction manuelle de réseaux moléculaires soulève plusieurs problèmes. D'une part, le traitement des données à haut débit est humainement impossible ou extrêmement long. De plus, lorsqu'un réseau est construit manuellement, il peut être biaisé par l'interprétation de la personne qui le construit. Il peut également contenir des erreurs. Et surtout, comme toutes les données ne peuvent pas être intégrées, il est très incomplet.

De plus, lorsqu'un travail est effectué à partir des banques de données, il est nécessaire d'évaluer la qualité et la traçabilité de la source d'information.

En conclusion, ce type de construction de réseaux est réellement long, difficile et incomplet, c'est pourquoi une approche automatique est nécessaire aujourd'hui.

B- Approche semi-automatique

Pour construire des réseaux moléculaires, différents outils ont été développés. Par exemple, il est possible d'utiliser des données de puces à ADN dans l'outil proposé par l'entreprise *Ingenuity* qui grâce aux données disponibles dans la banque de données *Ingenuity Knowledge Database* (construite grâce à de la fouille de texte : *text mining*), reconstruit le réseau moléculaire correspondant. Cela a été réalisé pour la construction de réseaux induits par des allergènes dans les lymphocytes T [134]. Il est également possible de créer un réseau à partir du nom des protéines avec des logiciels tels que *Pathway Studio* qui contient une base de donnée créée à partir d'un outil de *text mining* du logiciel *MedScan* [135].

Ces systèmes permettent la création semi-automatique des réseaux à partir de la saisie manuelle de l'utilisateur.

Ces méthodes sont rapides et peuvent constituer une bonne aide. Mais comme elles reposent souvent sur le *text mining* qui est de la fouille de texte, ces méthodes ne sont pas totalement

fiables et tiennent compte uniquement du texte et malheureusement pas des expériences décrites sous forme de figures.

C- Approche automatique

La construction automatique des réseaux moléculaires peut être abordée, aujourd'hui, grâce à un grand nombre de modèles reposant sur des formalismes particuliers. Les modèles les plus « populaires » vont être expliqués avec leur formalisme de base ainsi que quelques méthodes d'application. Ces méthodes de construction de réseaux sont qualifiées d'inférence de réseaux.

<u>Définition</u>: L'inférence est une opération par laquelle on passe d'une assertion considérée comme vraie à une autre assertion au moyen d'un système de règles qui rend cette deuxième assertion également vraie. (Dictionnaire Larousse)

Un modèle d'inférence de réseaux permet donc d'obtenir des éléments (nœuds et arcs) d'un réseau à partir d'autres éléments en répondant aux règles imposées par le formalisme utilisé. Les différents formalismes peuvent avoir pour source les statistiques, les réactions ou la logique.

Dans cette partie, les arcs des réseaux sont décrits grâce à deux termes qu'il est important de connaître :

- Influence : c'est le sens de l'interaction, c'est-à-dire quelle molécule agit sur quelle autre molécule,
- Type d'influence : c'est le type d'interaction (inhibition ou activation).

1- Modèle reposant sur des statistiques

<u>Théorie de l'information et information mutuelle</u>: Cette approche, comme son nom l'indique, repose sur des informations théoriques, telles que l'information mutuelle. Ce modèle est probabiliste et a pour but de quantifier la dépendance statistique entre des variables.

C'est donc une mesure des probabilités de réalisation d'un évènement Y en fonction de la quantité d'information apportée par la réalisation de l'évènement X. Pour l'information mutuelle, on calcule l'information inconnue d'un phénomène aléatoire à partir de l'entropie ou par la divergence de Kullback-Leibler en se basant sur l'information connue.

Ce modèle de la théorie de l'information a été utilisé dans la création de trois algorithmes ARACNE [136], MINDy [137] et MI3 [138], capables de traiter des données de puces à ADN et permettant l'inférence de réseaux de gènes. De plus, Ciaccio et al [139] ont utilisé ARACNE à partir de données issues de *Micro Western Array* de façon à comparer les résultats avec leur propre méthode de construction des réseaux. Le Tableau 2 présente un bref récapitulatif citant la référence de chaque méthode, les données expérimentales utilisées, l'algorithme et le type de validation.

Tableau 2 : Méthodes d'inférence reposant sur la théorie de l'information

^{*} Cette méthode n'a pas été validée expérimentalement ou à partir des connaissances car elle représente le moyen de validation des auteurs pour évaluer leur propre méthode.

Référence	Données expérimentales	Algorithme	Validation
[136]	Puces à ADN	ARACNE	Connaissances
[137]	Puces à ADN	MINDy	Expérimentale
[138]	Puces à ADN	MI3	Littérature
[139]	Micro Western Array	ARACNE	*

Le modèle de la théorie de l'information peut donc être appliqué à différents types d'expériences. Le résultat est un graphe non orienté, dans lequel des coefficients de corrélation sont affectés aux arcs. Ceci signifie que ce modèle permet d'obtenir le squelette du réseau sans connaître l'influence ni le type d'influence des relations qui le composent.

Modèle graphique gaussien: Le modèle graphique gaussien se base sur le formalisme de la loi normale multidimensionnelle (ou loi de Gauss à plusieurs variables). Cette loi de probabilité est une généralisation multidimensionnelle de la loi normale. Ce modèle repose donc sur le calcul d'une matrice de corrélation, les coefficients de corrélation sont calculés sur les variables prises deux à deux. Les données utilisées pour les calculs sont en général, des données d'expression [140-142].

Si les coefficients de corrélation utilisés sont les coefficients de Pearson alors le résultat est une matrice des variances-covariances de variables réduites.

Pour un réseau basé sur le traitement de données d'expression de gènes, si le coefficient de corrélation entre deux gènes est élevé, cela peut indiquer :

- Une interaction directe entre ces deux gènes,
- Une interaction indirecte entre ces deux gènes,
- Une régulation de ces deux gènes par un autre gène qui serait un régulateur commun.

Un coefficient de corrélation (Pearson) se situe entre les valeurs -1 et 1. Si le coefficient de corrélation de deux variables se situe entre -1 et -0,5 ou entre 0,5 et 1 alors les variables sont fortement corrélées. Si ce coefficient se situe entre -0,5 et 0 ou entre 0 et 0,5, alors les variables sont très faiblement corrélées. Si le coefficient est nul alors les variables ne sont pas corrélées. Ensuite, une matrice de précision est calculée, elle correspond à une matrice de corrélation inverse caractérisée par l'emplacement des zéros. Il est possible de voir, à partir de la matrice de précision représentée sur la Figure 15, que si le coefficient de précision est égal à 0 alors les gènes n'interagissent pas et si le coefficient est différent de 0, les gènes interagissent. Par exemple, le coefficient de précision entre le gène 1 et le gène 2 est différent de 0, ils interagissent donc sur le graphe alors que le coefficient entre le gène 1 et le gène 3 est égal à 0, il n'existe donc pas de lien entre ces deux gènes sur le graphe.

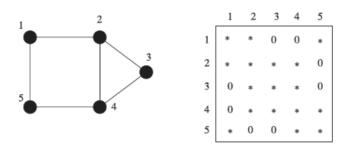


Figure 15 : Passage de la matrice de précision au graphe correspondant Les étoiles sont les coefficients différents de zéro. (Figure issue de [143])

Chiquet et al [142] ont développé un ensemble de fonctions intégré au logiciel R : package SIMoNe. Ce package permet une inférence basée sur des coefficients de corrélation partielle. Une étape de biclustering¹ est réalisée afin de classer les gènes en fonction de leur niveau d'expression. Chu et al [144] ont, quant à eux, élaboré une méthodologie reposant sur ce modèle graphique gaussien associé à des données de corrélation gènes-SNP.

Andrei et Kendziorski [140], quant à eux, estiment que le modèle graphique gaussien entraîne une inférence erronée, c'est-à-dire qu'une interaction indirecte pourrait être représentée comme une interaction directe. Ces auteurs proposent donc une alternative quant à l'utilisation du modèle graphique gaussien, qui consiste à construire deux matrices de coefficients de corrélation partielle, celles-ci étant ensuite associées pour chaque interaction possible. Puis, une ré-estimation des corrélations partielles est effectuée pour enfin identifier

.

¹ Méthode d'exploration des données permettant leur classification.

les vraies interactions. Cette méthode est guidée par des connaissances et permet d'identifier les relations directes et indirectes entre les nœuds.

Le modèle graphique gaussien a été plus récemment utilisé pour la construction de réactions métaboliques à partir de données métabolomiques [145]. Les auteurs affirment que ce travail pourrait être prolongé par la comparaison de coefficients de corrélation partielle de l'activité enzymatique ou par la validation expérimentale des interactions.

Ce modèle permet également d'aborder l'aspect dynamique grâce à la prise en compte de données de puces en séries temporelles.

Le graphe résultat ne contient ni les influences, ni les types d'influences, le modèle graphique gaussien permet donc de construire uniquement le squelette du réseau moléculaire. Donc, comme le modèle reposant sur la théorie de l'information, le réseau obtenu manque d'informations importantes.

<u>Modèle bayésien</u>: Le modèle bayésien repose sur le calcul de probabilités utilisant le théorème de Bayes dont la formule est :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Les termes P(A) et P(B) sont les probabilités de l'évènement A et de l'évènement B respectivement. Le terme P(A|B) est la probabilité de l'évènement A sachant que l'évènement B est vrai. Si cette probabilité est différente de P(A), alors l'évènement A n'est pas indépendant de l'évènement B. De même, le terme P(B|A) est la probabilité de l'évènement B sachant que A est vrai. Si B est connu, ce terme est dit 'fonction de vraisemblance de A'. En biologie, cela permet de calculer la probabilité des données non observées en fonction des données observées.

Le théorème de Bayes est donc un outil important en inférence statistique, et consiste en la mise à jour de probabilités à partir d'observations et de leurs propres probabilités. L'utilisation de ce formalisme dans le domaine de l'inférence de réseaux est nommée modèle ou réseau bayésien. Un réseau bayésien représente les connaissances et permet de calculer des probabilités conditionnelles.

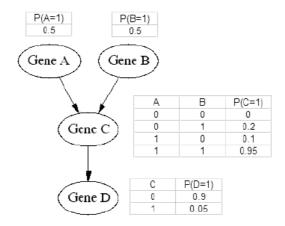


Figure 16 : Représentation d'un réseau bayésien de 4 gènes.

Les gènes A et B agissent de façon combinée pour activer l'expression du gène C, qui lui, réprime le gène D. Lorsque la probabilité est égale ou proche de 0 alors le gène ne s'exprime pas et lorsqu'elle est égale ou proche de 1 le gène s'exprime. Figure issue de [146].

Dans l'exemple présenté à la Figure 16, les auteurs s'intéressent à la probabilité d'expression du gène C connaissant l'expression des gènes A et B et à la probabilité d'expression du gène D connaissant l'expression du gène C. Les probabilités obtenues permettent de dire que l'expression de C est induite par l'expression de A et de B conjointement et que l'expression de D est réprimée par l'expression de C.

Cette technique permet de modéliser des réseaux en se basant sur des statistiques, ce qui permet de faire face à l'aspect aléatoire des données ou à des données bruitées [147]. Le modèle bayésien est également capable de fonctionner même s'il y a des données manquantes ou des variables cachées. Cependant, la construction d'un modèle bayésien est un problème NP-difficile, c'est-à-dire que la construction de ce modèle présente une complexité algorithmique importante.

Le modèle bayésien ne prend pas en compte l'aspect dynamique des réseaux, c'est pourquoi le modèle bayésien dynamique a été développé. Ce modèle est une répétition du modèle bayésien mais avec l'addition de lien causal à chaque pas de temps. Le modèle dynamique bayésien est une cumulation du modèle bayésien donc sa complexité algorithmique en est d'autant plus grande.

Le Tableau 3 répertorie quelques méthodologies basées sur le modèle bayésien. Ce modèle est très utilisé en inférence de réseaux moléculaires. Dans un grand nombre de cas, les données

expérimentales prises en compte sont des résultats de puces à ADN. Certaines méthodes peuvent être guidées par des connaissances ou de la classification par clustering. L'exploitation des données expérimentales ne se limite tout de même pas uniquement aux puces à ADN, des méthodologies basées sur des données de cytométrie de flux avec des conditions de perturbations ou des données de *Micro Western Array* ont été développées par Sachs et al [148] et Ciaccio et al [139] respectivement. La validation des méthodes se fait à l'aide des connaissances ou expérimentalement. Ce qui est original ici, c'est le choix de Ciaccio et al, qui contrairement aux autres auteurs, ont utilisé des algorithmes existants et validés pour effectuer une comparaison avec leur propre méthode.

Tableau 3 : Méthodes d'inférence reposant sur le modèle bayésien

MOC = model-based overlapping clustering algorithm, VBEM = Variational Bayesian Expectation Maximization, RNAi = RNA interférent, CLR = Context Likelihood of Relatedness, ? = inconnu

Référence	Données expérimentales	Algorithme	Validation
[149]	Puces à ADN	MOC	Littérature
[150]	Puces à ADN	VBEM	Littérature, YeasTract
[146]	Puces à ADN	Metropolis Hasting	NON
[151]	Données RNAi	Metropolis Hasting	Littérature
[148]	Cytométrie en flux et perturbations	Bayesian network learning algorithm [152]	Expérimentale
[139]	Micro Western array	Dynamic programming Algorithm	ARACNE, CLR
		[153]	[154]
[155]	Expression de gènes	Banjo	Connaissances
[149]	Données d'expression de	?	KEGG
	gènes		

Comme il a été dit précédemment, il existe une variante dynamique du modèle bayésien. Ce modèle dynamique peut faire appel à l'algorithme Banjo qui utilise des données d'expression de gènes [155] ou encore à l'algorithme Metropolis-Hasting sur des données d'ARN interférents [151].

En résumé, le modèle bayésien peut donc être utilisé sur différents types de données expérimentales en suivant un grand nombre de méthodologies distinctes. Ce modèle se présente toujours sous forme de réseau orienté acyclique, ce qui en biologie présente un

inconvénient car il existe des rétro-régulations. Néanmoins, le modèle bayésien permet de connaître l'influence et le type d'influence peut aussi être déduit car, en fonction des probabilités, il est possible de savoir si l'influence est une activation ou une inhibition.

La variante dynamique du modèle bayésien, quant à elle, permet d'obtenir le même type de réseaux et elle permet, en plus, de modéliser les rétro-régulations, ce qui signifie que les réseaux obtenus peuvent être cycliques. Le problème posé par ce modèle est sa grande complexité algorithmique.

L'inférence de réseaux biologiques peut donc se faire grâce à des modèles statistiques comme décrits précédemment mais les modélisateurs peuvent également utiliser une autre facette des mathématiques : la logique.

2- Modèles logiques

<u>Modèle booléen</u>: Ce modèle est basé sur la logique binaire. L'algèbre de Boole ne contient que deux éléments aussi appelés variables booléennes. Ces variables ont deux états possibles 0 ou 1 (faux ou vrai respectivement). Donc, en fonction des conditions, une variable peut être vraie ou fausse.

Exemple:

Doctorat = Manuscrit + Soutenance

Un doctorat est vrai si le manuscrit est vrai et si la soutenance est vraie.

Dans l'exemple, la fonction logique Doctorat dépend de deux variables Manuscrit et Soutenance.

L'algèbre booléenne, lorsqu'elle est utilisée dans le cadre de l'inférence de réseaux moléculaires, permet de modéliser les molécules dans deux états : soit la molécule est active, état *ON*, soit elle est inactive, état *OFF*. Ce formalisme permet également l'utilisation d'opérateurs logiques tels que le ET (conjonction), le OU (disjonction) et le NON (négation). *Exemple de conjonction :*

x ET y est VRAI si et seulement si x est VRAI ET si y est VRAI.

Exemple de disjonction :

x OU y est VRAI si et seulement si x est VRAI OU y est VRAI.

Exemple de négation :

Le contraire de *x* est VRAI si et seulement si *x* est FAUX.

Voici un exemple d'application du modèle booléen sur le cycle cellulaire selon Shmulevich et al [156]. La Figure 17 montre les interactions impliquées dans la régulation du cycle cellulaire. L'association cdk7/cycline H forme le complexe CAK, celui-ci active la protéine cdk2. Cdk2 en se complexant avec la cycline E inhibe la protéine Rb, dans ce cas la synthèse de l'ADN peut se faire et le cycle cellulaire suit donc son cours. Or le complexe p21/WAF1 inhibe le complexe cdk2/cycline E donc, dans ce cas, la protéine Rb est active et inhibe la synthèse de l'ADN.

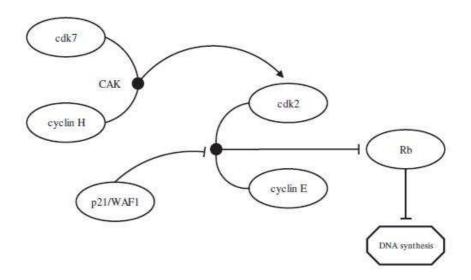


Figure 17 : Schéma représentant la régulation du cycle cellulaire.

Les flèches représentent les activations et les barres représentent les inhibitions. Figure extraite de [156]

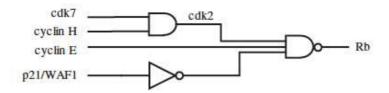


Figure 18 : Diagramme logique décrivant l'activité de la protéine Rb à partir de 4 entrées cdk7, cycline H, cycline E et p21/WAF1 [156]

Avec représentant un AND, représentant un NOT et représentant NEGATED AND.

Cette régulation de l'activité de la protéine Rb peut être retranscrite en diagramme logique. Le diagramme (Figure 18) représente le réseau suivant : si cdk7 ET cycline H sont *ON* alors cdk2 est *ON*. Si cdk2 est *ON* ET cycline E est *ON* alors Rb est *OFF*. Si p21/WAF1 est *ON* alors le complexe formé par la cdk2 et la cycline E est *OFF*, ce qui confère un état *ON* à Rb. La symbolique *NEGATED AND* est utilisée ici afin de définir que p21/WAF1, cdk2 et cycline E déterminent tous ensemble l'état de Rb. Ce qui se traduit par : si cdk2 est *ON* et cycline E est *ON* et si p21/WAF1 est *OFF* alors Rb est *OFF* alors que si cdk2 est *OFF* et cycline E est *OFF* et si p21/WAF1 est *ON* alors Rb est *ON*.

Certains travaux d'inférence de réseaux reposent sur le formalisme booléen, le Tableau 4 regroupe quelques méthodes rencontrées dans la littérature avec le type de réseau moléculaire obtenu, l'outil utilisé et le mode de validation.

Tableau 4 : Méthodes d'inférence reposant sur le modèle booléen

Référence	Type de réseau	Algorithme/ Logiciel	Validation
[157]	Réseau de régulation de gènes	Propre aux auteurs	Connaissances
[158]	Réseau de régulation de gènes	GinSim	NON
[159]	Réseau de signalisation	CellNetOptimizer	Expérimentale
[160]	Réseau de signalisation	Integer Linear Program	Suppression/ Prédiction

Les méthodes utilisant le formalisme booléen permettent l'intégration de connaissances de façon à diriger le réseau sortant. Le Tableau 4 montre que les auteurs utilisant ce formalisme obtiennent des réseaux de régulation des gènes mais aussi des réseaux de signalisation cellulaire. La validation des différentes méthodes peut se faire expérimentalement ou à partir de connaissances. En ce qui concerne les travaux de Mitsos et al [160], la méthode de validation repose sur la prédiction du réseau lorsque 20% des données expérimentales sont supprimées dans différents scénarios.

Le modèle booléen permet uniquement la modélisation topologique des réseaux. Une adaptation dynamique a donc été créée : le modèle booléen dynamique.

De plus, il existe aussi une variante du modèle booléen reposant sur l'insertion de probabilités : le modèle booléen probabiliste. Ce modèle permet de prendre en compte un aspect dynamique mais aussi des incertitudes [156].

En conclusion, cette méthode permet d'obtenir un graphe orienté, c'est-à-dire qu'il est possible de connaître l'influence. De plus, ce modèle permet de connaître le type d'influence, ce qui signifie que l'on sait si A active ou inhibe B.

Modèle Fuzzy Logic (ou logique floue): Le modèle Fuzzy Logic permet, par rapport à la logique booléenne, qu'une condition puisse être présentée dans un état autre que vrai ou faux. Par exemple, la moyenne d'une classe est de 14/20 et on considère qu'une note au-dessus de 15 est une très bonne note et une note en dessous de 13 est une note très moyenne. En logique booléenne, ceci se traduirait par tout ce qui est égal à ou au-dessus de 15 est à 100% une très bonne note et à 0% si la note est en dessous de 15. Alors qu'en logique floue, il sera possible de dire qu'une note est moyenne à 100% si elle est égale à 14, une note en dessous de 13 ou au-dessus de 15 est à 0% moyenne et une note égale à 13 ou à 15 est à 50% moyenne.

Ce modèle, lorsqu'il est appliqué à des réseaux moléculaires ne limite donc pas l'état des molécules à *ON* et *OFF*. Ce formalisme ressemble au formalisme booléen en étant tout simplement moins restrictif.

Le modèle *Fuzzy Logic* est surtout utilisé sur des réseaux de signalisation. Les auteurs utilisent ce modèle pour l'analyse de croisements de voies de signalisation induites par différents facteurs [161, 162]. Le modèle *Fuzzy Logic* peut également servir à affiner un réseau moléculaire existant [163].

Comme le modèle booléen, le modèle *Fuzzy Logic* permet d'obtenir un réseau comportant l'influence et le type d'influence des relations.

<u>Conclusion sur les modèles logiques</u>: Il existe d'autres méthodes basées sur la logique telles que la création de systèmes ODE à partir de la logique [164] ou encore un modèle standardized qualitative dynamical systems [165].

La logique ne se limite pas à l'algèbre booléenne ou à la logique floue. Il existe aussi la logique d'ordre 0 (ou calcul de propositions), la logique de premier ordre (ou calcul de prédicats) et la logique d'ordre supérieur. Le calcul de prédicats représente la formalisation du langage en introduisant des notions de variables, fonctions, relations, prédicats, connecteurs logiques et quantificateurs alors que le calcul de propositions est en fait un calcul des prédicats mais sans les quantificateurs. La logique d'ordre supérieur est une extension de la logique de premier ordre.

Une présentation de la logique du premier ordre sera réalisée dans la partie Matériel et Méthodes.

Les modèles logiques permettent d'obtenir des réseaux complets, c'est-à-dire que l'influence et le type d'influence des relations sont obtenus. De plus, c'est un formalisme adapté à tout type de réseaux moléculaires.

L'inférence de réseaux biologiques peut donc se faire grâce à des modèles mathématiques comme décrits précédemment mais les modélisateurs peuvent également faire appel à un autre formalisme : les systèmes d'équations différentielles.

3- Systèmes d'équations différentielles

<u>Définition</u>: Une équation différentielle est une égalité qui permet de faire le lien entre une fonction et sa dérivée.

Les équations différentielles peuvent être de deux types :

- Equations différentielles ordinaires (ODE),
- Equations différentielles partielles (PDE).

Dans le cas de la modélisation dynamique de réseaux moléculaires, ce sont les systèmes ODE qui sont les plus utilisés. Un système ODE permet de décrire quantitativement les processus biochimiques.

De façon à illustrer l'utilisation des systèmes ODE, nous allons prendre l'exemple de la complexation entre un ligand L et son récepteur R que l'on peut décrire de la façon suivante :

$$L + R \stackrel{k_1}{\rightleftharpoons} L:R$$

L:R est le complexe formé par L et R. Les k1 et k2 sont respectivement la constante de vitesse d'association et la constante de vitesse de dissociation.

A partir de ces données, il est possible de calculer les vitesses initiales des réactions. Pour cette étape, il est nécessaire de choisir une loi de vitesse qui est le lien entre la vitesse et les

paramètres qui ont une influence sur elle. Dans notre exemple, nous avons choisi la loi d'action de masse.

Ces données permettent le calcul des vitesses initiales de réactions :

$$v_1 = k_1 * [L] * [R] et v_2 = k_2 * [L:R]$$

Avec v₁, [L], [R], v₂, [L:R] représentant respectivement la vitesse initiale d'association, la concentration en ligand, la concentration en récepteur, la vitesse initiale de dissociation et la concentration en complexe L:R.

Grâce aux calculs de vitesses précédemment décrits, les ODE suivantes sont créées :

$$\frac{d[R]}{d_t} = v_2 - v_1 \text{ et } \frac{d[L:R]}{d_t} = v_1 - v_2$$

Ces équations permettent ensuite de déterminer l'évolution de la concentration de complexe L:R et de récepteur libre R en fonction du temps. Bien sûr, ici le modèle est très simple puisqu'il n'y a que trois espèces moléculaires. Dans la réalité, les équations se complexifient en fonction du nombre de variables à prendre en compte.

Il faut noter que les paramètres de départ peuvent être mesurés ou estimés. Si les données sont mesurées alors elles sont dites observables.

Le formalisme ODE est le plus utilisé dans la modélisation dynamique de réseaux moléculaires et peut être appliqué à tout type d'expériences. Les systèmes ODE permettent de modéliser des réseaux de gènes mais aussi des réseaux de signalisation.

Les méthodes d'inférence reposant sur les ODE peuvent utiliser différents algorithmes ou logiciels, il est possible de choisir NIR (*Network Identification by multiple Regression* [166]), MNI (*Microarray Network Identification* [167]) ou TSNI (*Time Series Network Identification* [168]). L'algorithme NIR a pour point de départ des mesures d'expression de gènes à l'état stable ou des mesures en séries temporelles avec des perturbations transcriptionnelles. Dans le cas de données avec perturbations, NIR nécessite en plus des connaissances sur les gènes directement perturbés. MNI utilise le même type de données que NIR mais ne nécessite pas de connaissances sur les gènes touchés par les perturbations. MNI permet l'identification des voies de signalisation mais aussi des gènes touchés par une perturbation. TSNI, quant à lui ne prend en entrée que des données en séries temporelles. Il permet de prédire les cibles directes

des perturbations, les cibles des facteurs de transcription lors d'une surexpression de ces facteurs [169].

Les systèmes PDE ne sont pas présentés ici mais le principe est similaire aux systèmes ODE or dans ce cas la dérivée de la fonction est partielle. Ces systèmes sont utilisés par exemple, dans la modélisation de phénomènes d'endocytose et de transports cellulaires [170].

En résumé, les modèles reposant sur les réactions ne sont pas limités à un ou quelques types d'expériences. Comme les données de départ sont des mesures, il est possible d'exploiter un grand nombre de résultats expérimentaux. Les réseaux obtenus en utilisant ce formalisme comportent les influences et leur type. De plus, ce modèle autorise les boucles et modélise donc les rétro-régulations. L'atout majeur des systèmes d'équations est l'étude de l'évolution des quantités moléculaires dans le temps, grâce à laquelle il est possible de mettre en avant des amplifications du signal par exemple [82]. Les modèles topologiques ne permettent pas de faire la lumière sur de tels phénomènes. Les réseaux moléculaires sont très vastes, ce qui entraîne une grande complexité (en temps et en espace) pour la modélisation dynamique, c'est pourquoi les modèles nécessitent d'être réduits. Ce qui signifie que la totalité d'un réseau ne peut pas être prise en compte.

Tous les formalismes précédemment décrits sont mathématiques. Il existe d'autres méthodes d'inférence de réseaux reposant sur un formalisme à base de règles.

4- Méthodes à base de règles

Les méthodes à base de règles sont surtout connues sous le nom de systèmes experts.

<u>Définition</u>: Un système expert est un programme informatique capable d'imiter le raisonnement d'un humain expert dans le domaine d'intérêt [171].

Ce formalisme est composé de trois parties :

- Une base de règles,
- Une base de faits.
- Un moteur d'inférence.

La base de règles et la base de faits représentent une base de connaissances [172]. Les experts créent la base de règles et les utilisateurs présentent des faits (Figure 19). La base de connaissances est prise en charge par le moteur d'inférence pour déduire des conclusions qui sont de nouveaux faits. Ces conclusions sont ensuite intégrées à la base de faits.

Une règle prend la forme suivante :

SI condition(s) ALORS conclusion(s)

Les règles sont composées de ce que l'on appelle des prédicats qui sont en quelque sorte des « faits génériques ». La partie condition comporte un ou plusieurs prédicats et il en est de même pour la partie conclusion. Les faits sont quant à eux, des prédicats instanciés (voir exemple ci-dessous).

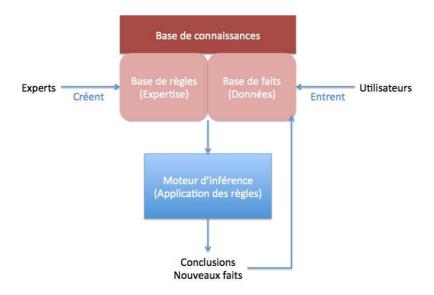


Figure 19 : Principe d'un système expert

Exemple:

<u>Règle</u>: SI x est un récepteur membranaire ET SI x se lie à la protéine G ALORS x est un RCPG.

Faits : FSHR est un récepteur membranaire. FSHR se lie à la protéine G.

<u>Instanciation des prédicats dans la règle et conclusion obtenue</u>: SI FSHR est un récepteur membranaire ET SI FSHR se lie à la protéine G ALORS FSHR est un RCPG.

Le moteur d'inférence effectue l'opération décrite dans l'exemple, c'est-à-dire l'instanciation des prédicats dans la règle à partir des faits qui lui sont présentés, afin d'obtenir la conclusion correspondante.

Les moteurs d'inférence peuvent utiliser les modes de chaînage avant ou de chaînage arrière [147]. Le chaînage avant est une méthode de déduction. A partir des règles et des faits, il est possible d'obtenir de nouvelles conclusions. Alors que le chaînage arrière se fait dans l'autre sens, c'est-à-dire qu'il part des conclusions pour tenter de remonter vers les faits.

Blinov et al [173] ont développé le logiciel BioNetGen. Ce logiciel utilise des règles formalisées dans un langage créé par ses auteurs (langage BioNetGen). Les règles utilisées par BioNetGen concernent les réactions enzymatiques, les modifications post-traductionnelles potentielles et les interactions. En entrée, l'utilisateur donne des quantités moléculaires. Colvin et al [174] ont développé RuleMonkey, c'est un logiciel similaire à la méthode de Gillespie² dont les modèles peuvent être encodés dans le langage BioNetGen. Comme pour le logiciel BioNetGen, la formalisation sous forme de règles ne prend en compte que les relations entre les molécules et permet de déduire le réseau à partir de quantités de molécules. De même, Sneddon et al [175] ont développé le logiciel NFsim (Network-Free Stochastic Simulator) dont le principe repose sur une modélisation grâce à l'algorithme de Gillespie sous une forme à base de règles et une extension du langage BioNetGen. La validation a été réalisée grâce à la confrontation des résultats avec ceux d'autres logiciels pour RuleMonkey et expérimentalement pour NFsim. Une autre méthode à base de règle a été élaborée par Ollivier et al [176], le principe étant de formaliser sous forme de règles l'allostérie et les réactions biochimiques. Cette méthode est validée expérimentalement mais avec une formalisation de l'allostérie, il est possible de modéliser uniquement des réactions directes, or la plupart des expériences ne permettent pas de savoir si une réaction est directe ou indirecte. Les différentes méthodes présentées montrent qu'il est possible de formaliser les relations et même l'allostérie sous forme de règles. Les méthodes présentées sont capables de déduire un réseau à partir de données numériques. L'inconvénient est que ce type de données n'est pas toujours disponible.

Aucune des méthodes rencontrées n'utilise de système à base de règles pour formaliser les expériences directement. Cependant, ceci permettrait d'obtenir les conclusions que peut tirer un biologiste de ses observations.

² La méthode de Gillespie permet la simulation du comportement des systèmes qui évoluent en suivant une équation maîtresse.

5- Autres méthodes

Les chercheurs ont le choix en ce qui concerne les méthodes d'inférence de réseaux moléculaires. Il existe d'autres approches que les formalismes précédemment décrits.

Boosting approach : BoostiGraph est une « *boosting approach* » pour l'apprentissage automatique de graphes [177]. Cette technique repose sur l'autorégression et utilise les données de puces à ADN. C'est un principe regroupant de nombreux algorithmes qui s'appuient sur des ensembles d'hypothèses binaires : le *boosting* optimise leurs performances. Ce principe est issu de la combinaison d'hypothèses. C'est une méthode réalisée par itérations successives. Cette méthode permet l'introduction de connaissances supplémentaires (en plus des données de puces). Le résultat de la méthode représente uniquement le squelette du réseau.

Relevance networks: Les 'relevance networks' sont des groupes de variables, celles-ci étant connectées les unes aux autres. Cette technique se fait par le calcul de coefficients de corrélation r entre deux variables. La Figure 20 représente le principe des relevance networks selon Butte et Kohane [178]. L'idée est que, plus il y a de mesures (n) et plus les coefficients de corrélation (r²) sont élevés, alors la relation est validée (arc très épais sur la Figure 20). Lorsque le r² est faible (arc fin sur la Figure 20) alors la relation est hypothétique. Cette méthode permet d'obtenir uniquement le squelette du réseau.

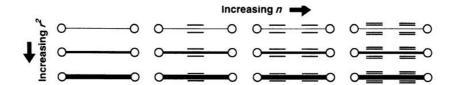


Figure 20 : Représentation graphique de modèles entre deux variables [178].

Le nombre de mesures utilisées pour créer le modèle, ou n, augmente dans la représentation de la gauche vers la droite, et est représenté par les petits traits. Le r² du modèle augmente dans la représentation du haut vers le bas et est représenté par l'épaisseur de la ligne.

<u>Context Likehood Relatedness (CLR) [154]</u>: Certains auteurs combinent plusieurs méthodes afin de créer une méta-méthode capable de mieux répondre aux attentes sur l'acquisition des réseaux. L'algorithme CLR est né de cette initiative et regroupe :

- relevance network,
- ARACNE,
- Modèle bayésien.

Cet algorithme a été appliqué à des données d'expression chez *E. coli* (4345 gènes, 445 observations) pour construire un réseau de gènes. Aucune donnée complémentaire n'est ajoutée. La validation de la méthode a été effectuée de façon expérimentale en association avec la banque de données RegulonDB. Cet algorithme a été utilisé pour évaluer une autre méthode de construction de réseaux mais à partir de données de *Micro Western Array* [139]. Le graphe obtenu est orienté mais il n'y a pas de précision sur les types d'interactions.

6- Conclusion sur les méthodes actuelles de construction des réseaux moléculaires

Un des buts de la modélisation des réseaux moléculaires est de construire des réseaux les plus complets possibles (avec un maximum de données, de molécules, de relations). C'est pourquoi des méthodes (Information mutuelle, modèle gaussien, Relevance networks...) aboutissant uniquement au squelette des réseaux ou des méthodes ne prenant en compte qu'un type de données expérimentales, ne sont pas les plus adaptées. Le modèle booléen, quant à lui, permet d'obtenir des réseaux complets mais le problème réside dans son aspect binaire. Une molécule peut présenter différents états, donc il est difficile de se limiter aux deux états possibles comme dans le modèle booléen. La modélisation bayésienne ne permet pas d'obtenir des cycles dans le réseau, ce qui, comme il l'a déjà été cité, pose un problème quant à la modélisation des rétro-régulations. La version dynamique des réseaux bayésiens, quant à elle, permet de modéliser des cycles mais au niveau algorithmique, son utilisation présente une trop grande complexité. Le formalisme qui présente un grand nombre d'avantages est le modèle basé sur des ODE. Ce formalisme présente tout de même un souci majeur, la complexité. Or les réseaux biologiques sont très vastes, ce qui rend impraticable ce type de formalisation [179]. Ceci entraîne la réduction des modèles et donc des réseaux incomplets. Les modèles qui permettent la meilleure modélisation topologique des réseaux moléculaires sont les modèles logiques non binaires et les méthodes à base de règles. Il faut noter que les systèmes à base de règles permettent une plus grande liberté de formalisation car il est possible de créer des règles directement à partir des expériences de façon à imiter le raisonnement du chercheur. De cette façon, la construction d'un réseau directement liée aux expériences peut être élaborée. Ce réseau pourra ensuite être visualisé grâce à des logiciels développés pour la visualisation.

III- Visualisation graphique des réseaux biologiques

La construction des réseaux et leur exploitation par les biologistes nécessitent une étape de transformation des conclusions obtenues en une visualisation graphique compréhensible et exploitable.

Le choix d'une méthode de visualisation des réseaux moléculaires est crucial car comme le montre la Figure 21, il existe des incohérences et/ou des ambiguïtés avec des notations non standardisées.

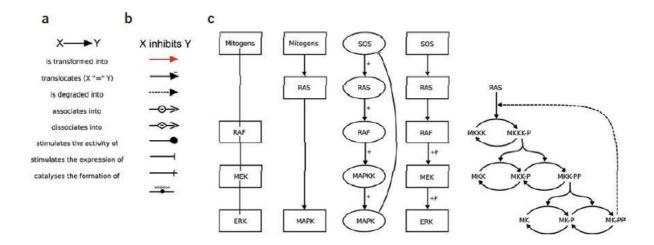


Figure 21 : Incohérences et ambiguïtés causées par le manque de standardisation de notations de représentation des réseaux biologiques

(a) Huit significations différentes associées au même symbole dans un graphique décrivant le rôle de la cycline dans les régulations cellulaires. (b) Neuf symboles différents, trouvés dans la littérature, pour représenter la même signification. (c) Cinq graphiques différents de la cascade MAP kinase trouvés dans la littérature scientifique, représentant des niveaux progressifs des connaissances biologiques et biochimiques. Figure extraite de [180].

La Figure 21 montre bien que les représentations graphiques rencontrées sont parfois incohérentes, comme par exemple, un même symbole avec plusieurs significations ou différents symboles avec la même signification. Il est donc nécessaire d'homogénéiser les représentations graphiques des réseaux moléculaires, de façon à faciliter la compréhension mais aussi l'échange des informations.

Dans ce but, les scientifiques créent des langages de représentation graphique des réseaux qui sont standardisés. Ceci représente, dans le cadre de la bioinformatique, un challenge et un

effort d'homogénéisation des données. De cette façon, il est plus facile d'échanger des données si toute la communauté s'accorde sur le vocabulaire/la syntaxe à utiliser et les outils à développer. Il existe donc des langages standardisés, dont les plus utilisés aujourd'hui sont probablement les langages SBML (*Systems Biology Markup Language*), SBGN (*Systems Biology Graphical Notation*) et BioPax (*Biological Pathway Exchange*).

A- SBML

Le SBML (http://sbml.org/) est le langage le plus ancien. Il est défini à partir du langage de modélisation UML (UML (Unified Modeling Language). Ce langage est ensuite décrit sous forme d'XML (extens.com/ (extens.co

Le SBML niveau 1 (version 2) permet ainsi de mettre en place une représentation graphique grâce à un système de balises pour définir les espèces moléculaires, les compartiments cellulaires et les réactions. Il fournit une possibilité d'ajouter également les quantités initiales des espèces moléculaires, la définition de paramètres des réactions comme la vitesse initiale, la vitesse maximale et la définition d'unités (moles, litres, secondes...). Ceci signifie que le SBML peut être utilisé non seulement pour visualiser la topologie des réseaux moléculaires mais également pour aller vers une modélisation dynamique de ces réseaux. Ainsi, il est possible d'étudier le comportement du modèle en fonction du temps. Le SBML niveau 2 propose des améliorations comme par exemple, la définition des formules mathématiques qui est dorénavant décrite avec le langage MathML (*Mathematical Markup Language*), alors que ces formules étaient décrites sous forme de texte dans la version précédente [182].

La Figure 22 montre un exemple de représentation en SBML avec deux types de définitions. La définition des espèces moléculaires présente un identifiant (sI et s7 dans notre exemple), un nom (Ras et pERK), l'appartenance à un compartiment cellulaire (ici 'default' car aucun compartiment n'a été défini en dehors de la fenêtre par défaut) et la quantité initiale (non définie dans l'exemple donc '0'). La définition des réactions comporte aussi un identifiant et un nom (reI dans notre exemple), la réversibilité de la réaction (ici 'false' car elle n'est pas réversible). Il peut y avoir des annotations mais l'exemple n'en présente pas. Ensuite, on définit la réaction en elle-même avec la liste des substrats (ici Raf1 identifiée s2), la liste des produits (pRaf1 identifiée s3) et la liste des molécules ayant une action sur cette réaction (ici Ras identifiée s1).

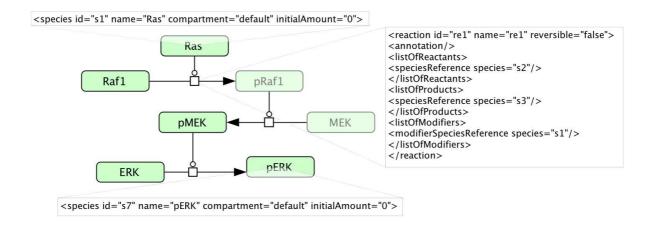


Figure 22 : Exemple la cascade MAP Kinase ERK en représentation graphique SBML

Les rectangles arrondis représentent les espèces moléculaires (ici, définitions SBML de Ras et pERK uniquement) et les réactions (ici, uniquement la réaction de transformation de Raf1 en pRaf1 catalysée par Ras).

Le niveau 3 du SBML est actuellement en cours de développement. Les spécifications du noyau du SBML niveau 3 sont disponibles [183]. Ce nouveau niveau de SBML devrait être capable de gérer par exemple, les complexes, ce qui n'est pas le cas avec la version actuelle. Le format supporté par les logiciels actuels reste le SBML niveau 2 version 4.

Le SBML ne fait pas de différences entre les espèces moléculaires, les ions sont représentés de la même façon que les protéines par exemple. De plus, les régulations se limitent à l'activation, il n'existe pas d'inhibition. C'est pourquoi pour combler les lacunes de ce langage, il peut être associé au langage SBGN.

B- SBGN

Le langage SBGN est devenu un standard pour la visualisation des réseaux biologiques (http://www.sbgn.org/). En réalité, le SBGN est constitué de trois sous-langages [180]. Le premier concerne les diagrammes de processus (*process diagram*) dont le but est de représenter les processus de transformation de molécules en d'autres molécules, de changement d'état de ces molécules ou encore de changement de leur localisation. Le deuxième sous-langage concerne les diagrammes d'interactions des entités (*Entity relationship diagram*), c'est donc une représentation des relations entre les molécules et les règles que suivent ces relations. Le dernier sous-langage concerne les diagrammes d'activité (*Activity flow diagram*) et il permet la représentation de l'influence des activités biologiques sur les autres.

Chaque sous-langage respecte une symbolique précise pour la représentation graphique des réseaux. Nous allons prendre l'exemple des diagrammes de processus. La Figure 23 représente la légende d'un diagramme de processus. Une représentation graphique propre à chaque espèce moléculaire est définie, cela va de la simple molécule chimique au complexe moléculaire. Il en est de même pour les processus tels que l'association ou la dissociation. Chaque régulation d'une molécule sur un processus est représentée de façon distincte, par exemple, il n'est pas possible de confondre une catalyse avec une inhibition.

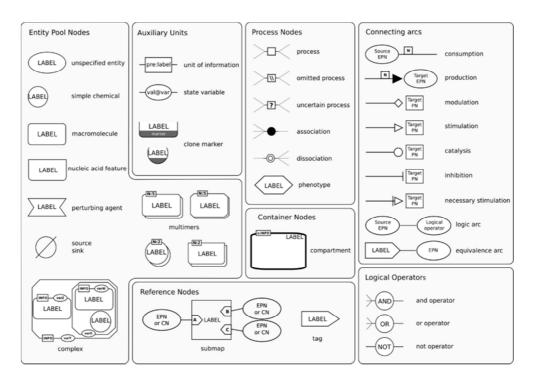


Figure 23 : Symbolique utilisée dans le cadre du langage dédié aux diagrammes de processus Figure extraite du site http://www.sbgn.org/.

Pour faciliter l'explication, la Figure 24 montre la cascade des MAP Kinase ERK sous forme de diagramme de processus SBGN, les rectangles arrondis sont les macromolécules, ici ce sont des protéines. Les ronds contenant des 'P' représentent des variations d'état de ces protéines, si le 'P' est présent, la protéine est phosphorylée sinon elle ne l'est pas. Les ronds sont quant à eux de simples molécules chimiques, ici le bas des ronds est noir, ceci signifie tout simplement que la molécule est présente en plusieurs exemplaires dans le schéma. Cet exemple ne comporte qu'un seul type de processus, la transition avec un seul type d'arc de connexion, la catalyse. Nous avons donc, par exemple, un processus de transition qui est la phosphorylation de MEK pour devenir phosphoMEK, catalysée par phosphoRaf1.

L'exemple des diagrammes de processus montre la capacité du langage SBGN de représenter les réseaux biologiques de façon standardisée. De plus, un autre atout du SBGN est la possibilité d'intégrer un vocabulaire contrôlé et adapté à chaque nœud et à chaque arc du réseau grâce à l'ontologie SBO (*Systems Biology Ontology*) [184].

Le SBGN se présente comme une version modifiée du langage XML. Le SBGN peut aussi être associé au SBML.

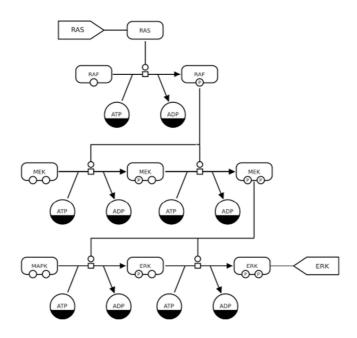


Figure 24 : Exemple de la voie de MAP Kinase ERK sous forme de diagramme de processus. Figure extraite du site http://www.sbgn.org/.

C- BioPax

BioPax est le langage le plus récent. Il prend la forme d'une ontologie, c'est-à-dire que c'est un système formel de connaissances [185] (http://www.biopax.org/), ou plus exactement BioPax peut être définie comme une classification hiérarchique de connaissances sur les réseaux biologiques. BioPax se présente sous la forme OWL (Web Ontology Language). Les créateurs de BioPax expliquent leur langage grâce à l'exemple décrit sur la Figure 25. L'exemple montre la protéine AKT1 qui en bref, présente deux sites de phosphorylation T308 et S473, phosphorylés respectivement par PDK1 et PDK2. Ces sites sont déphosphorylés par la phosphatase PP2A. La forme d'AKT1 comportant deux phosphorylations se complexe à la protéine HSP90. La description BioPax d'AKT1 dans sa forme non modifiée est :

- AKT1 sans modification porte l'identifiant AKT1.1,

- est une protéine (is a),
- a pour référence la protéine rAKT1 (has proteinReference),
- n'a pas la caractéristique d'être phosphorylée à la position 308 (has notFeature p@308),
- n'a pas la caractéristique d'être phosphorylée à la position 473.

La forme phosphorylée sur la position 308 présente deux différences par rapport à AKT1.1 :

- elle a pour identifiant AKT1.2,
- elle a la caractéristique d'être phosphorylée sur la position 308 (has Feature).

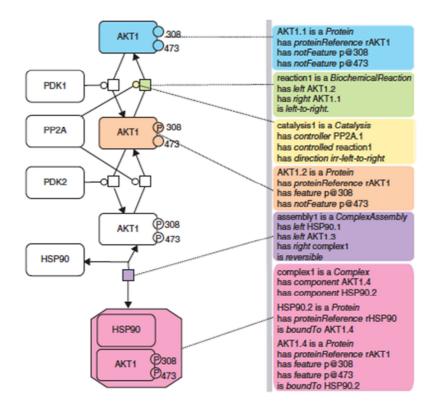


Figure 25 : Exemple de molécules et réactions décrites en BioPax issue de [185]

Chaque molécule est ainsi définie. Ensuite, il y a les réactions. La Figure 25 montre la déphosphorylation d'AKT1 sur la position 308. En BioPax, ceci est décrit de la façon suivante :

- la réaction porte l'identifiant reaction 1,
- c'est une réaction biochimique (is a BiochemicalReaction),
- le substrat (*has left*) est AKT1.2, c'est-à-dire la forme phosphorylée d'AKT sur la position 308,
- le produit (has right) est AKT1.1, c'est-à-dire la forme non modifiée d'AKT1,
- la réaction a une direction, c'est-à-dire qu'elle va du substrat vers le produit et non l'inverse (*is left-to-right*).

Cette réaction est catalysée par la phosphatase PP2A et cette catalyse est définie :

- avec un identifiant *catalysis* 1,
- c'est donc une catalyse (is a),
- elle est contrôlée par PP2A.1 (has controller),
- elle contrôle la réaction 1 (has controlled),
- elle est irréversible et la réaction se fait du substrat vers le produit (*has direction irr-left-to-right*).

La complexation de deux (ou plus) molécules se définit de la même façon.

BioPax représente surtout un vocabulaire précis de représentation graphique des processus biologiques. Il ne gère pas les aspects dynamiques et quantitatifs des processus biologiques comme le fait le SBML. Dans le but d'y remédier, une coordination BioPax/SBML est envisagée. De plus, les différents concepts du BioPax peuvent être visualisés en utilisant le SBGN grâce à une collaboration entre les différentes équipes de développement [185].

D- Logiciels de visualisation graphique des réseaux biologiques

Les langages présentés précédemment nécessitent des logiciels capables de les interpréter car la compréhension d'un réseau se fait par la sortie graphique et non par la lecture de fichiers exprimés dans un format quelque peu difficile à déchiffrer.

BioPax peut être utilisé grâce à certains logiciels tels que Paxtools, une API Java qui gère les fichiers BioPax. Sybil (*Systems Biology Linker*) permet la visualisation des réseaux en BioPax mais aussi la transformation des fichiers en SBML (Sybil fait partie du logiciel Virtual Cell [186, 187]). BiNoM (*Biological Network Manager*) qui est un plugin de Cytoscape permet, quant à lui, l'analyse des réseaux en BioPax, en SGBN et en SBML grâce à sa capacité d'importation et exportation de ces langages [188, 189]. Un outil très utilisé dans la communauté des biologistes est CellDesigner. Ce logiciel est capable d'interpréter le SBML mais surtout, il possède un langage qui lui est propre, c'est une combinaison de SBML et de SBGN. Ce langage permet à CellDesigner d'obtenir des réseaux biologiques « plus » complets qu'avec seulement un des deux langages. De plus, ce langage propre à CellDesigner est également interprétable dans Cytoscape grâce à BiNoM.

Bien sûr, la liste des outils permettant l'interprétation des langages n'est pas complète, il en existe bien d'autres comme par exemple, Biocham ou BioUML pour le SBGN et le SBML, ChiBE pour BioPax ou encore COPASI pour le SBML. En ce qui concerne le SBML, les

logiciels et bases de données l'utilisant sont répertoriés sur le site officiel (http://sbml.org/SBML_Software_Guide/SBML_Software_Matrix).

Ces langages permettent un échange standardisé des connaissances sur les réseaux biologiques. Dans ce but, il existe des banques de données généralistes donnant accès aux représentations graphiques des réseaux mais aussi parfois aux fichiers BioPax, SBML qui leur ont servi.

Problématique scientifique et objectif de la thèse

Les RCPG sont au cœur des recherches de l'équipe BIOS, et en particulier le récepteur de l'hormone FSH. Par une démarche expérimentale, les biologistes de l'équipe étudient les différentes voies de signalisation induites par cette hormone dans le but de pouvoir comprendre son mode de fonctionnement, construire un modèle de ce réseau et pouvoir ainsi aller vers le développement d'une nouvelle pharmacologie grâce à l'élaboration de molécules capables de mimer les effets de la FSH sans reproduire ses effets indésirables, ce type de molécules est appelé « agoniste biaisé ».

Mais comme cela a été décrit précédemment, les récepteurs tels que les RCPG sont capables d'activer un grand nombre de voies de signalisation cellulaires. La complexité de ces voies rend difficile la création de modèles de réseaux moléculaires.

Les méthodes d'inférence automatique de réseaux citées plus haut sont dans un grand nombre de cas validées expérimentalement ou à partir de la littérature, mais elles sont surtout appliquées aux réseaux de régulation des gènes. De plus, les réseaux ne sont pas toujours complets (rétro-régulations absentes, orientations inconnues, simplifications du modèle, complexité algorithmique).

L'acquisition de réseau à partir des banques de données n'est pas envisageable non plus, car des incohérences existent avec les données expérimentales. L'acquisition manuelle est quant à elle possible mais cela représente une source d'erreurs et d'oublis ainsi qu'une charge de travail considérable.

Aujourd'hui, un des plus grands challenges en bioinformatique et en biologie systémique est le développement d'une méthode capable de reconstruire automatiquement les réseaux d'une grande envergure directement à partir des expériences réalisées par les biologistes.

L'objectif de cette thèse est précisément de mettre au point une méthode permettant de créer des réseaux de signalisation, de façon automatique, directement à partir des données expérimentales, la réalisation de ce travail a nécessité différentes étapes :

- Une grande partie du réseau FSH a été construite manuellement de façon à disposer d'un modèle publié.
- Le développement d'une méthode d'inférence de réseaux à base de règles a été réalisé grâce à la formalisation du raisonnement. Les résultats obtenus ont pu être comparés à des réseaux existants. Ce travail est effectué en étroite collaboration avec l'équipe « Algorithms and Models for Integrative Biology » (LRI/INRIA, Orsay) dans le cadre du projet ASAM (INRA/INRIA). Christine Froidevaux a adapté le moteur d'inférence utilisé avec l'aide de Sarah Cohen Boulakia et Adrien Rougny.
- Une base de données avec une interface web est en cours de développement pour donner accès aux différentes informations de ce projet.

Matériel et méthodes

I- Formalisation et inférence

Dans un premier temps, nous définirons les notions de logique utilisées dans notre méthode. Dans un second temps, nous allons décrire la méthodologie qui a permis la formalisation du raisonnement et des données biologiques. Et pour terminer, nous aborderons la transformation des faits logiques en représentation graphique.

A- Inférence

Comme cela a été décrit dans l'introduction, les systèmes à base de règles permettent, à partir d'un ensemble de règles et d'un ensemble de faits initiaux, de déduire des faits nouveaux au moyen d'un moteur d'inférence. Cette partie est développée par nos collaborateurs du Laboratoire de Recherche en Informatique à Orsay.

1- Calcul de prédicats ou Logique du premier ordre

Nous avons choisi d'utiliser la logique du premier ordre qui permet une généricité des règles et des constantes ainsi que des symboles de fonction nous permettant de construire des objets complexes tels que les complexes moléculaires. Ce type de formalisation repose sur une syntaxe précise basée sur un alphabet $\mathcal A$ représentant un ensemble de symboles. Ils symbolisent des opérateurs, des quantificateurs, des termes et des prédicats.

Opérateurs : Il en existe quatre :

- ¬ signifiant 'NON' aussi noté ~,
- A signifiant 'ET' aussi noté 'AND',
- V signifiant 'OU' aussi noté 'OR',
- ullet signifiant 'IMPLIQUE' aussi noté IF ... THEN.

Quantificateurs: Ils sont au nombre de deux:

- ▼ pour 'quel que soit',
- \exists pour 'il existe'.

<u>Termes</u>: Un terme peut être une constante, une variable ou une fonction qui prend pour arguments d'autres termes. Ceci est défini de la façon suivante :

• Une constante est un terme,

- Une variable est un terme,
- $f(t_1,...,t_n)$ est un terme si f est un symbole de fonction d'arité³ n et si $t_1,...,t_n$ sont des termes.

<u>Prédicats</u>: Ce sont des formules simples. Les prédicats que nous utilisons contiennent des arguments. Ces derniers sont obligatoirement des termes.

<u>Formules atomiques</u>: Si \mathcal{P} est un prédicat d'arité n et si $t_1,...,t_n$ sont des termes alors l'expression $\mathcal{P}(t_1,...,t_n)$ est une formule simple (ou formule atomique).

Formules : En logique de premier ordre, sont des formules :

- 1. Une formule atomique,
- 2. $(f_1 \land f_2)$ si f_1 et f_2 sont des formules,
- 3. $(f_1 \lor f_2)$ si f_1 et f_2 sont des formules,
- 4. $\neg f_1 \operatorname{si} f_1 \operatorname{est} \operatorname{une} \operatorname{formule}$,
- 5. $(f_1 \rightarrow f_2)$ si f_1 et f_2 sont des formules,
- 6. $\forall x f \text{ si } f \text{ est une formule,}$
- 7. $\exists x \ f \ \text{si} \ f \ \text{est} \ \text{une formule}.$

Remarque : Les quantificateurs ne sont pas utilisés dans notre méthode.

Ce vocabulaire permet maintenant de construire les règles. Une règle est de la forme:

Si condition(s) Alors conclusion(s)

La partie condition de la règle r est une formule et est décrite par :

Condition(r):
$$cond_1 \land cond_2 \land ... \land cond_p$$

Où chaque condition est une formule. L'ensemble des prédicats présents dans la condition est noté $\mathcal{P}(\text{Condition}(r))$.

De la même façon, la partie conclusion d'une règle est une formule et est décrite par :

Conclusion(r):
$$conc_1 \land conc_2 \land ... \land conc_q$$

Où chaque conclusion est une formule. L'ensemble des prédicats présents dans la conclusion est noté $\mathcal{P}(\text{Conclusion}(r))$.

En logique, une règle est donc représentée par la formule suivante :

³ L'arité est le nombre d'arguments.

$$(\operatorname{cond}_1 \wedge \operatorname{cond}_2 \wedge ... \wedge \operatorname{cond}_p) \longrightarrow (\operatorname{conc}_1 \wedge \operatorname{conc}_2 \wedge ... \wedge \operatorname{conc}_q)$$

<u>Illustration avec une expérience de test enzymatique ou Enzymatic Assay :</u>

La formalisation du raisonnement permettant d'interpréter le résultat d'une expérience de « *Enzymatic Assay* » (EA, voir la description dans la section Création des règles) conduit à la règle :

• **IF** $EA(x,[(n_y,y),(n_z,z)],[(n_a,a),(n_b,b)],e)$ **THEN** $CATALYZE(x,[(n_y,y),(n_z,z)],[(n_a,a),(n_b,b)],direct,confirmed,e)$

Signifiant: "Si en faisant un EA des molécules y et z en présence de la molécule x, on constate un effet e sur la production des molécules a et b alors x a un effet e direct sur la transformation de y et z en a et b."

Cette écriture ne correspond pas au formalisme du premier ordre. Par exemple, nous représentons les listes par des crochets, ce qui n'est pas reconnu dans ce formalisme. On peut cependant transformer la règle de manière automatique ce qui conduit à :

Condition(r): $EA(x,liste((n_y,y),(n_z,z)),liste((n_a,a),(n_b,b)),effect)$

 $Conclusion(r): CATALYZE(x, liste((n_y, y), (n_z, z)), liste((n_a, a), (n_b, b)), direct, confirmed, effect)\\$

Dans la suite du manuscrit, nous utiliserons le premier formalisme qui permet une meilleure lisibilité des règles.

Les règles sont transformées en règles de Horn, c'est-à-dire en règles comportant une seule conclusion positive. Par exemple, si une règle comporte plusieurs conclusions, on obtient une règle par conclusion.

Application à la règle portant sur une expérience de Western Blot :

• IF PANWB(x;a;e) AND Transcription(z;w) AND Translation(w;y) AND PanAntibody(a;y) THEN TCE(x;z;w;unknown;hypothesis;e) AND TLE(x;w;y;unknown;hypothesis;e) AND DEGRADATION(x;w;unknown;hypothesis;e) AND DEGRADATION(x;y;unknown;hypothesis;e)

<u>Signifiant</u>: "Si un WB montre une variation e de la quantité de la protéine y, reconnue par l'anticorps a pan y, entre la condition sans la molécule x dans l'environnement cellulaire et avec la molécule x alors la molécule x a un effet e hypothétique sur la transcription du gène z en ARNm w avec une distance inconnue et x a un effet e hypothétique sur la traduction de l'ARNm w en protéine y avec une distance inconnue et x a un effet e hypothétique sur la

dégradation de l'ARNm w avec une distance inconnue et x a un effet e hypothétique sur la dégradation de la protéine y avec une distance inconnue. »

Sous la forme de règles de Horn, on obtient :

- (PANWB(x;a;e) \land Transcription(z;w) \land Translation(w;y) \land PanAntibody(a;y)) \rightarrow (TCE(x;z;w;unknown;hypothesis;e))
- (PANWB(x;a;e) \land Transcription(z;w) \land Translation(w;y) \land PanAntibody(a;y)) \rightarrow (TLE(x;w;y;unknown;hypothesis;e))
- (PANWB(x;a;e) ∧ Transcription(z;w) ∧ Translation(w;y) ∧ PanAntibody(a;y)) → (DEGRADATION(x;w;unknown;hypothesis;e))
- (PANWB(x;a;e) \land Transcription(z;w) \land Translation(w;y) \land PanAntibody(a;y)) \longrightarrow (DEGRADATION(x;y;unknown;hypothesis;e))

La création de ces règles de Horn a pour conséquence une augmentation du nombre de règles dans la base.

2- Moteur d'inférence SOLAR

Un moteur d'inférence est un système automatique qui permet, à partir d'un ensemble de règles et d'un ensemble de faits initiaux de déduire de nouvelles conséquences.

Le moteur d'inférence choisi pour ce travail est SOLAR (*Skip Ordered Linear for Advanced Reasoning*) [190]. SOLAR a été choisi car ce moteur d'inférence est capable de raisonner sur n'importe quel type de clauses. En effet, outre les règles de Horn utilisées pour construire les réseaux, des règles permettant de définir les types moléculaires des arguments sont également utilisées. Dans le cas d'un *Enzymatic Assay*, la catalyse de la réaction peut être réalisée uniquement si la molécule x est de type Complex ou Protein ou Ribozyme donc il est possible d'écrire la règle « **IF** $EA(x,[(n_y,y),(n_z,z)],[(n_a,a),(n_b,b)],e)$ **THEN** COMPLEX(x) **OR** PROTEIN(x) **OR** RIBOZYME(x) ». Cette dernière n'est pas une règle de Horn. De plus, des règles de Horn non définies (conclusion unique mais négative) sont également utilisées comme contrainte d'intégrité telle que « Si x est un gène alors x n'est pas une protéine », ce qui s'écrit « **IF** GENE(x) **THEN** ¬PROTEIN(x) ».

SOLAR est une implémentation de la procédure *Skip Ordered Linear* qui est une méthode permettant de trouver les clauses caractéristiques d'une théorie. SOLAR permet l'élagage des conséquences redondantes, ce qui réduit l'espace de recherche grâce à l'implémentation de différentes méthodes (*Skip-regularity, Identical reduction, Local failure caching for length condition...*).

SOLAR est un démonstrateur de théorème basé sur la résolution logique du premier ordre. Il est aussi capable de créer de nouvelles règles grâce au raisonnement déductif. Par exemple, si on fournit les règles « SI A et B ALORS C » et « SI C et D ALORS E », SOLAR sera capable de créer la règle « SI A et B et D ALORS E ».

<u>Stratégie utilisée [191]</u>: Il existe de nombreuses stratégies pour l'utilisation de SOLAR. Nous expliciterons ici uniquement la stratégie utilisée par nos collaborateurs du LRI dans le cadre du stage de Master d'Adrien Rougny.

SOLAR utilise le calcul des tableaux afin de déduire des faits nouveaux à partir d'un ensemble de règles et de faits initiaux (top_clauses). SOLAR permet l'obtention de clauses finales à partir de l'extension de clauses initiales (top_clauses). Une top_clause est étendue de façon itérative par l'élimination des littéraux de cette clause et par l'addition de nouveaux littéraux, ce qui aboutit à chaque itération à de nouvelles clauses (Figure 26). Toutes les clauses, top_clause, clauses obtenues et clauses initiales, constituent un tableau. SOLAR étend la top_clause et les clauses qu'il déduit jusqu'à ce que le tableau soit résolu ou jusqu'à ce qu'il n'y ait plus aucune extension possible.

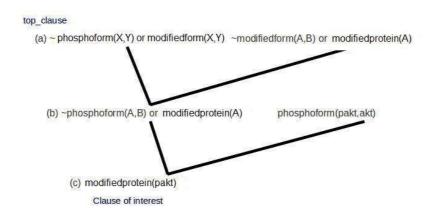


Figure 26: Exemple de calcul de tableau avec un champ de production [191]

SOLAR permet de ne rechercher qu'un sous-ensemble des conclusions possibles par la définition d'un champ de production (PF). Un tableau résolu est donc un tableau qui a pour clause finale la forme déclarée dans le PF. Par exemple, on peut choisir de rechercher uniquement les nouveaux faits du type « MODIFIEDPROTEIN ».

La Figure 26 permet d'illustrer la procédure suivie par SOLAR. Sont supposés vrais les trois prédicats :

- 1. PHOSPHOFORM(X,Y) : X est une forme phosphorylée de Y,
- 2. MODIFIEDFORM(X,Y): X est une forme modifiée de Y,
- 3. MODIFIEDPROTEIN(X): X est une molécule modifiée.

Deux règles sont fournies à SOLAR :

- règle 1 : **IF** PHOSPHOFORM(X,Y) **THEN** MODIFIEDFORM(X,Y) formalisée par ~*phosphoform*(X,Y) *or modifiedprotein*(X,Y),
- règle 2 : **IF** MODIFIEDFORM(X,Y) **THEN** MODIFIEDPROTEIN(X) formalisée par *~modifiedform*(X,Y) *or modifiedprotein*(X).

Le fait initial est *PHOSPHOFORM*(*pAkt*,*Akt*), et on réduit le champ de production aux faits de la forme *MODIFIEDPROTEIN*.

SOLAR va appliquer les règles au fait initial et en déduire deux faits nouveaux :

- MODIFIEDFORM(pAkt,Akt) en utilisant la règle 1,
- *MODIFIEDPROTEIN(pAkt)* en utilisant la règle 2.

Cet exemple aboutit sur un fait d'intérêt (correspondant à ce qui est défini dans le PF) : MODIFIEDPROTEIN(pAkt), le tableau est donc résolu puisque d'une part il n'y a plus d'extension possible et d'autre part, la forme d'intérêt a été obtenue.

Un grand nombre d'options sont disponibles dans SOLAR, elles permettent de suivre différentes stratégies de raisonnement. Celles qui ont été utilisées sont :

- *Option used* : Cette option permet de savoir quelles clauses ont été utilisées pour aboutir à un fait donné, c'est donc l'historique des faits déduits.
- Choix de la top_clause: Cette option contraint SOLAR à utiliser cette clause comme commencement, ceci signifie que dans tous les faits déduits, il y en aura obligatoirement un qui aura la top_clause dans son historique⁴.
- Choix du PF: Cette option permet de déterminer la forme d'intérêt des faits déduits. Par exemple, si l'utilisateur veut uniquement des relations de phosphorylation, tous les faits déduits seront de la forme PHOSPHORYLATE.
- Option –df n: Cette option permet de définir une limite quant à la profondeur du raisonnement pour être sûr que SOLAR s'arrête. La profondeur du raisonnement est l'enchaînement des règles, si la profondeur est définie à deux alors il y aura un enchaînement de deux règles au plus. Cette option évite que SOLAR fasse des boucles infinies à cause du fait que certaines clauses peuvent s'étendre par elles-mêmes⁵.

⁴ Il peut y avoir plusieurs top_clauses mais chaque fait déduit dépend d'au moins une.

⁵ La présence de règles récursives peut empêcher de finaliser le raisonnement.

B- Méthodologie de la formalisation

La formalisation du raisonnement et des données expérimentales se déroule en trois étapes. La première consiste en une classification hiérarchique des molécules, la seconde est la formalisation des relations qu'il est possible de rencontrer dans un réseau moléculaire, et pour terminer, l'étape la plus critique, l'écriture sous forme de règles du raisonnement qui permet à l'expert de déduire des relations entre les molécules à partir de faits expérimentaux.

1- Classification hiérarchique des molécules

Les molécules sont les nœuds des réseaux que nous étudions. Il est important de définir précisément des classes (aussi appelées types moléculaires dans ce manuscrit) car certaines expériences ne peuvent être appliquées qu'à certains types moléculaires et pas à d'autres : par exemple, dans une expérience de *Western Blot*, la molécule observée est une protéine.

Il est important de préciser qu'à chaque niveau de la hiérarchie, les classes sont mutuellement exclusives. Ainsi, l'appartenance d'une molécule à une classe donnée permet de conclure que cette molécule n'appartient à aucune autre classe de même niveau dans la hiérarchie.

Cette classification est réalisée en fonction de la nature biochimique de ces molécules et en fonction des représentations possibles dans le langage propre au logiciel CellDesigner. Nous avons défini cinq classes :

- ADN.
- ARN.
- Protéine,
- Petite molécule,
- Complexe.

La seule différence qui existe entre cette classification et celle du langage de CellDesigner réside dans le fait que la représentation de l'ADN se limite aux gènes avec la possibilité d'ajouter des régions régulatrices.

<u>ADN</u>: L'ADN est composé de nucléotides, c'est-à-dire de bases azotées liées à une molécule de désoxyribose portant un groupement phosphate. Les bases azotées contenues dans l'ADN sont l'adénine, la cytosine, la guanine et la thymine.

L'ADN contient l'information génétique et présente une localisation précise en fonction du type d'organisme :

- Procaryotes: ils contiennent un seul ADN circulaire dans le cytosol⁶,
- Eucaryotes : l'ADN est divisé en plusieurs chromosomes⁷. La plus grande partie de l'ADN est localisée dans le noyau, une petite quantité est présente dans les mitochondries. Dans le cas des végétaux, il y a une petite quantité située dans les chloroplastes.

L'appartenance d'une molécule x à la classe ADN est représentée par le prédicat DNA(x). Pour la suite de la description des prédicats, ils seront notés entre parenthèses.

La classe ADN est divisée en sous-classes car chaque 'partie' de l'ADN ne présente pas les mêmes fonctions biologiques. La Figure 27 présente la hiérarchisation des sous-classes d'ADN.

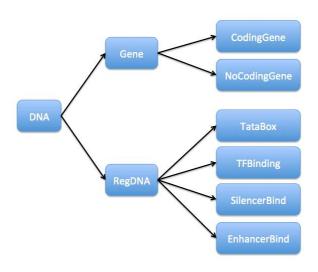


Figure 27 : Hiérarchie au sein de la classe ADN

L'ADN contient des gènes (GENE(x)), ce sont les régions transcrites en ARN. Ces gènes peuvent être codants (CODINGGENE(x)), c'est-à-dire qu'ils codent pour une protéine. Mais ces gènes peuvent aussi être non codants (NOCODINGGENE(x)), ces derniers sont transcrits en ARN mais ceux-ci n'aboutissent pas à une traduction en protéines.

L'ADN contient également des régions régulatrices (REGDNA(x)). Elles représentent les régions non transcrites en ARN. Une région régulatrice peut être une boîte TATA (TATABOX(x)), c'est une portion de l'ADN présente dans la région promotrice d'un gène et elle sert en partie à la reconnaissance par l'ARN polymérase, complexe enzymatique responsable de la transcription en ARN. Une région régulatrice peut être un domaine de

_

⁶ Les procaryotes contiennent aussi des plasmides mais ils ne sont pas nécessaires pour leur survie.

⁷ Le nombre de chromosomes dépend de l'état de ploïdie de l'organisme (humain : 46, souris : 40 et chien : 78).

fixation d'un facteur de transcription (TFBIND(x)). Un facteur de transcription est une protéine indispensable pour l'initiation ou la régulation de la transcription. Il existe deux autres types de régions régulatrices dans notre classification, les régions de fixation d'amplificateurs de la transcription (ENHANCERBIND(x)) et les régions de fixation de répresseurs de la transcription (SILENCERBIND(x)).

Cette hiérarchie permet de créer un ensemble de règles permettant de faire le lien entre la sous-classe d'origine et la classe générale, comme par exemple :

- IF GENE(x) THEN DNA(x)
- IF CODINGGENE(x) THEN GENE(x)

Bien sûr, cette classification n'est pas figée, il sera possible d'ajouter, par exemple, le site d'initiation de la transcription.

ARN:

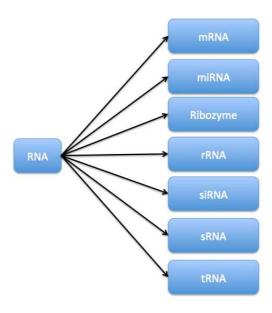


Figure 28 : Hiérarchie au sein de la classe ARN

Les ARN proviennent de la transcription des gènes. Ils sont aussi composés de nucléotides, mais dans leur cas, les bases azotées sont liées à du ribose portant un groupement phosphate. Les bases azotées constituant les ARN sont l'adénine, la cytosine, la guanine et l'uracile. Les ARN sont divisés en sept sous-classes en fonction de leur rôle biologique (Figure 28).

Les ARN provenant de la transcription de gènes codants sont les ARN messagers (MRNA(x)). Ces ARN messagers sont ensuite traduits en protéines. Aucun des autres types

d'ARN n'est traduit. Les microARN (MIRNA(x)) sont impliqués dans la régulation des gènes. Les ribozymes (RIBOZYME(x)) sont des ARN dotés d'une activité catalytique. Les ARN ribosomaux (RRNA(x)) entrent dans la composition des ribosomes qui sont de gros complexes permettant la synthèse des protéines à partir des ARN messagers. Les ARN de transfert (TRNA(x)) sont les ARN qui transportent les acides aminés jusqu'aux ribosomes lors de la traduction des ARN messagers. Les petits ARN (SRNA(x)) sont impliqués dans diverses régulations. Les petits ARN interférents (SIRNA(s)) sont des ARN courts qui s'hybrident à des ARN messagers de manière spécifique et inhibent ainsi leur traduction en protéine. Ces siRNA sont très utilisés dans les expériences biologiques, le prédicat SIRNAAGAINST(s,x) signifie que la siRNA s permet d'inhiber la synthèse de la protéine x. La hiérarchie des ARN permet de créer des règles permettant le lien entre la sous-classe et la classe, par exemple :

- IF TRNA(x) THEN RNA(x)
- IF RIBOZYME(x) THEN RNA(x)

Protéine : Les protéines (PROTEIN(x)) sont des macromolécules composées d'acides aminés. Elles sont issues de la traduction des ARN messagers. Cette classe regroupe toutes les protéines natives et trois sous-classes, les protéines modifiées, les mutants et les anticorps (Figure 29).

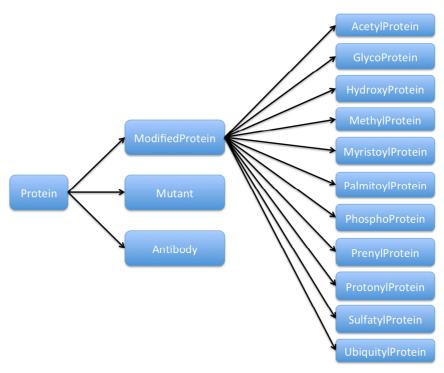


Figure 29 : Hiérarchie au sein de la classe Protéine

Nous défini modifiées avons une sous-classe regroupant les protéines (MODIFIEDPROTEIN(x)). Les protéines peuvent exister sous différentes formes, par exemple elles peuvent contenir une ou plusieurs modifications post-traductionnelles, ou présenter des changements conformationnels, ou encore faire partie d'un complexe. La sousclasse des protéines modifiées permet donc de marquer le fait qu'une protéine peut avoir plusieurs états qui ne sont pas obligatoirement définis précisément. Parmi les protéines modifiées, il y a des phosphoprotéines (PHOSPHOPROTEIN(x)), ce sont des protéines qui ont subi l'ajout d'un groupement phosphate au niveau d'un résidu sérine, thréonine ou tyrosine. Il existe des prédicats permettant de faire le lien entre les protéines modifiées et leur forme non modifiée:

- MODIFIEDFORM(b,x) qui signifie que *b* est une forme modifiée de la protéine *x*, le type de modification et sa position sont ici inconnus,
- PHOSPHOFORM(b,x,pos) qui signifie que *b* est une forme phosphorylée de la protéine *x* à la position *pos*.

Tableau 5 : Modifications post-traductionnelles et leur formalisation

Modification	Prédicat de la protéine	Prédicat de lien avec la
	modifiée	protéine native
Acétylation	ACETYLPROTEIN(b)	ACETYLFORM(b,x,pos)
Glycosylation	GLYCOSYLPROTEIN(b)	GLYCOSYLFORM(b,x,pos)
Hydroxylation	HYDROXYPROTEIN(b)	HYDROXYFORM(b,x,pos)
Méthylation	METHYLPROTEIN(b)	METHYLFORM(b,x,pos)
Myristoylation	MYRISTOYLPROTEIN(b)	MYRISTOYLFORM(b,x,pos)
Palmitoylation	PALMITOYLPROTEIN(b)	PALMITOYLFORM(b,x,pos)
Phosphorylation	PHOSPHOPROTEIN(b)	PHOSPHOFORM(b,x,pos)
Prénylation	PRENYLPROTEIN(b)	PRENYLFORM(b,x,pos)
Protonation	PROTOPROTEIN(b)	PROTOFORM(b,x,pos)
Sulfatation	SULFATYLPROTEIN(b)	SULFATYLFORM(b,x,pos)
Ubiquitination	UBIQUITYLPROTEIN(b)	UBIQUITYLFORM(b,x,pos)
Inconnue	MODIFIEDPROTEIN(b)	MODIFIEDFORM(b,x)

La position d'une modification est le code à une lettre de l'acide aminé portant la modification et son numéro dans la séquence. Le Tableau 5 regroupe les différentes modifications rencontrées ainsi que les prédicats qui leur sont associés. Les modifications

post-traductionnelles sont très importantes car, en fonction de l'état de modification d'une protéine, cette dernière peut être active ou inactive.

Dans la hiérarchie des protéines, nous avons défini deux autres sous-classes : les anticorps et les mutants. Ces sous-classes ont été différenciées car leur description nécessite des prédicats supplémentaires. Dans le cas des mutants, un prédicat définit le lien entre le mutant et la protéine sauvage : MUTANT(m,z) qui signifie que la protéine m est un mutant de la protéine m. Dans le cas d'un anticorps, un prédicat permet de le relier à sa molécule cible. Pour l'écriture de ce prédicat, il existe une distinction entre un anticorps m0 capable de reconnaître toutes les formes (la forme native et les forme modifiées) d'une protéine m0. PANANTIBODY(a,x), et un anticorps m0 capable de reconnaître spécifiquement m1, une forme modifiée de m2. ANTIBODYAGAINST(a,y).

<u>Petite molécule</u>: Cette classe regroupe toutes les petites molécules (SMALLMOLECULE(x)) qui sont réparties en deux sous-classes (Figure 30), les ions (ION(x)) et les molécules simples (SIMPLEMOLECULE(x)).

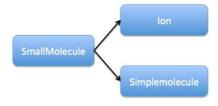


Figure 30 : Hiérarchie au sein de la classe Petites Molécules

<u>Complexe</u>: La dernière classe contient les assemblages de plusieurs molécules (COMPLEX(x)), les composants du complexe pouvant appartenir à n'importe quelle classe moléculaire, y compris Complexe.

Il est important de bien classer les différentes molécules qui entrent dans la composition d'un réseau moléculaire car un grand nombre de relations n'existe qu'entre des types moléculaires bien définis. Cela permet de s'assurer d'une certaine forme de cohérence, celle liée aux types moléculaires utilisés en entrée, ainsi que de la signification biologique des faits déduits.

2- Hiérarchie des relations

Les relations sont les liens biochimiques entre les molécules d'un réseau. Dans cette description, nous nous limiterons à deux exemples fréquemment rencontrés dans les réseaux de signalisation, les relations de complexation et de phosphorylation, ainsi qu'un cas formalisé de façon particulière la traduction. Les autres relations sont décrites dans l'Annexe 1.

Les relations que nous allons présenter contiennent un ou trois paramètres particuliers notés :

- *Status*, permet d'indiquer si la relation est validée (*confirmed*), invalidée (*invalidated*) ou hypothétique (*hypothesis*),
- *Distance*, indique si la relation est directe (*direct*), indirecte (*indirect*) ou si cette information est inconnue (*unknown*),
- Effect, est l'effet observé, qui peut être une augmentation (increase), une diminution (decrease) ou nul (noeffect).

<u>Complexation</u>: La complexation est une relation entre deux molécules ou plus qui s'assemblent pour former un complexe (Figure 31).

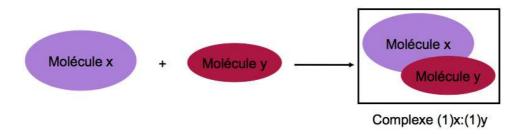


Figure 31 : Complexation

Une molécule x et une molécule y se lient pour former le complexe (1)x:(1)y.

La formalisation d'une complexation est représentée par le prédicat :

$$CO([(n_x,x),(n_y,y)],(n_x)x:(n_y)y,status)$$

Où x et y sont des molécules qui composent le complexe $(n_x)x$: $(n_y)y$ avec n_x et n_y des entiers représentant la stœchiométrie des molécules x et y respectivement. Les molécules qui entrent dans la composition d'un complexe peuvent appartenir à n'importe quelle classe (ADN, ARN, Protéine, Petite molécule et Complexe). Bien sûr, le produit de la complexation appartient à la classe Complexe. Les paramètres *distance* et *effect* sont absents puisque dans ce cas la relation est nécessairement directe, et la notion d'effet n'a pas de sens.

Phosphorylation: C'est l'addition d'un groupement phosphate (PO₄) sur une protéine⁸ (Figure 32).

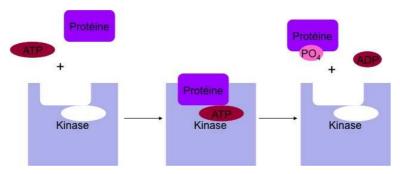


Figure 32: Phosphorylation directe

La kinase est une protéine ayant une activité enzymatique, elle fixe sa protéine substrat et une molécule d'ATP. La kinase hydrolyse l'ATP en ADP, ce qui permet la libération d'un groupement phosphate qui est ajouté de manière covalente à la protéine substrat.

La phosphorylation d'une protéine est très importante car dans un grand nombre de cas cette modification est intimement liée à l'état d'activité de la protéine. Par exemple, une protéine peut être active si elle est phosphorylée sur une position particulière (type de résidu et son numéro dans la séquence) et être inactive si cette phosphorylation est absente. Ou au contraire, une phosphorylation peut être responsable de l'inactivation d'une protéine et sa déphosphorylation peut entraîner son activation. Ceci est particulièrement fréquent dans les réseaux de signalisation cellulaire, qui constituent notre principal domaine d'application.

La formalisation se fait par le prédicat suivant :

PHOSPHORYLATE(x,y,b,distance,status,effect)

Où *x* est la molécule qui a un effet sur la phosphorylation de *y* en *b*. La molécule *x* peut appartenir à n'importe quelle classe moléculaire. La molécule *y* fait obligatoirement partie de la classe Protéines et *b* est obligatoirement une phosphoprotéine.

La distance peut être directe, indirecte ou inconnue. L'effet peut être une augmentation, une diminution ou sans effet.

<u>Traduction</u>: C'est la lecture d'un ARN messager par un ribosome pour synthétiser la protéine correspondante (Figure 33).

-

⁸ Actuellement, la phosphorylation est considérée uniquement comme une modification posttraductionnelle, cela poura être modifié si nécessaire.

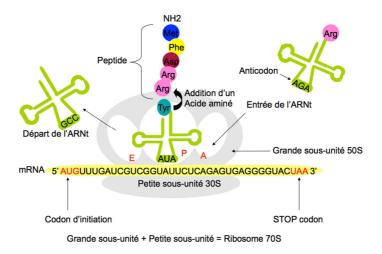


Figure 33 : Déroulement de la traduction

A pour AminoAcyl, P pour Peptidyl et E pour Exit. L'ARN de transfert entre dans le ribosome par la cavité A, puis s'il est complémentaire du codon en cours de lecture, il se déplace dans la cavité P où l'acide aminé associé est lié au peptide en cours de synthèse, puis il sort du ribosome par la cavité E.

L'ARN est composé de nucléotides qui sont lus trois par trois (par codon) à l'intérieur du ribosome. Un codon correspond à un seul acide aminé.

Le ribosome reconnaît le codon d'initiation AUG. Ce codon est en principe toujours le point de départ de la traduction. Quand un ribosome lit un codon, un ARN de transfert se joint au complexe par liaison de son anticodon au codon en cours de lecture (Figure 33).

S'il y a complémentarité anticodon/codon, alors l'acide aminé apporté par l'ARN de transfert se lie à la chaîne peptidique en cours de synthèse, sinon il repart avec son acide aminé pour laisser la place à un autre. L'élongation du peptide continue jusqu'à ce que le ribosome lise un codon STOP.

La traduction est formalisée par le prédicat :

TRANSLATION(w,y)

Où w est l'ARN messager traduit en la protéine y.

La traduction est un mécanisme cellulaire très complexe qui nécessite un grand nombre de régulations. Le prédicat formalise uniquement le fait qu'un ARNm w est traduit en protéine y.

3- Création de règles

Les règles définissent comment des faits nouveaux peuvent être déduits à partir de faits préexistants. Dans notre méthode, les faits initiaux sont les résultats d'expériences biologiques qui permettent à l'expert de déduire des relations entre molécules. Ces règles doivent donc permettre de reproduire le raisonnement de l'expert.

Nous avons distingué deux types de règles : les règles simples qui sont l'interprétation directe des résultats d'expériences biologiques, et les règles complexes dont les conditions sont des faits nouveaux.

L'utilisation d'un moteur d'inférence automatique apporte une contrainte très forte : toute molécule présente dans la conclusion doit être également présente dans la condition. Par exemple, si on observe une augmentation de la phosphorylation d'une protéine y en présence d'une stimulation alors la stimulation induit la phosphorylation de y pour donner y phosphorylée. Or y phosphorylée est une nouvelle molécule qui si elle n'apparaît pas dans la condition comme étant une forme phosphorylée de y, il n'est pas possible de l'obtenir dans la conclusion.

Nous avons choisi de présenter la méthodologie qui a permis de créer les règles les plus fréquemment utilisées. Cependant, la création des règles non décrite ici, est disponible en Annexe 2.

3.1- Création de règles simples

Les règles ont été créées à partir des expériences rencontrées dans la littérature qui a permis de créer manuellement le réseau induit par la FSH. A partir de ces expériences nous avons créé dans un premier temps des prédicats qui sont utilisables dans différents types d'expériences comme par exemple ANTIBODYAGAINST(a,x) et INHIBITOR(i,x). Ces prédicats ont été présentés dans la Classification hiérarchique des molécules et sont également disponibles dans l'Annexe 3. Dans un second temps, les prédicats de relations présentés dans la Hiérarchie des relations ont été élaborés. Ensuite, nous avons formalisé le raisonnement relatif aux différents types d'expériences rencontrés.

Test enzymatique ou Enzymatic Assay:

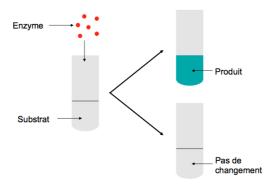


Figure 34 : Enzymatic Assay

Le principe d'un *Enzymatic Assay* (EA) est de mesurer l'activité d'une enzyme. Cette dernière est placée en présence de son (ses) substrat(s) et on mesure la production du (des) produit(s) (Figure 34).

Le prédicat de l'EA contient donc l'enzyme, la liste de substrats, la liste de produits et l'effet. Dans ce cas, l'effet *e* peut être une augmentation ou nul (Figure 34). Il n'est pas possible d'avoir une diminution car soit l'enzyme catalyse une réaction soit elle ne le fait pas. En conclusion, on obtient une réaction enzymatique. La règle créée est :

• IF $EA(x,[(n_y,y)],[(n_z,z)],e)$ THEN $CATALYZE(x,[(n_y,y)],[(n_z,z)],direct,confirmed,e)$

<u>Immunoprécipitation (IP)</u>: L'immunoprécipitation consiste à isoler une protéine d'intérêt en utilisant un anticorps dirigé contre elle. Dans le cas où la protéine d'intérêt forme un complexe avec d'autres molécules, ces dernières seront co-purifiées avec cette protéine (co-immunoprécipitation). La règle correspondante est :

• IF IP([x,y,...],increase) THEN CO([(?,x),(?,y)],(?)x:(?)y,confirmed)

L'immunoprécipitation est également utilisée pour comparer les quantités de protéines ou de complexes en présence et en absence d'un signal z. Dans le cas d'une quantification de complexe, ceci se traduit par la règle suivante :

• **IF** IPD(z,[x,y],increase) **THEN**ECO(z,[(?,x),(?,y)],(?)x:(?)y,unknown,confirmed,increase) **AND**CO([(?,x),(?,y)],(?)x:(?)y,confirmed)

Avec ECO signifiant effet sur la complexation.

Mesure de la phosphorylation ou Phosphorylation Assay (PA): Un PA est un Western Blot permettant de visualiser l'état de phosphorylation d'une protéine. Dans ce cas, on utilisera pour la détection un anticorps spécifique de la forme phosphorylée de y que l'on veut visualiser.

Cette expérience permet d'évaluer l'effet de la molécule x sur la phosphorylation de la protéine y. L'exemple de la Figure 35 montre une augmentation de la quantité de b (forme phosphorylée de y), donc on peut conclure que x induit la phosphorylation de y en b avec une distance inconnue.

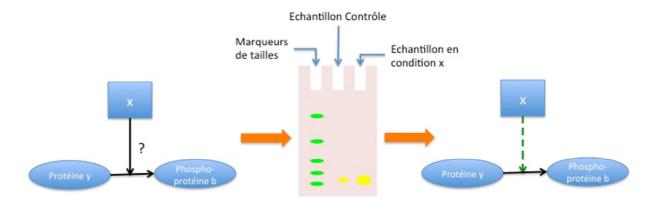


Figure 35: Phosphorylation Assay

On cherche à mesurer l'action de la molécule x sur la phosphorylation de la protéine y en b. Par *Phosphorylation Assay*, on constate une augmentation de la quantité de b en présence de x. On peut en conclure que x induit la phosphorylation de y en b avec une distance inconnue. Flèche pointillée verte = relation confirmée avec une distance inconnue.

Le *Phosphorylation Assay* est formalisé par le prédicat PA(x,y,a,e). On précise également que b est une forme phosphorylée de y (PHOSPHOFORM(b,y,pos)) et que a reconnaît b (ANTIBODYAGAINST(a,b)).

Comme il n'y a qu'une seule hypothèse possible pour expliquer le résultat de l'expérience, la conclusion aura le statut « confirmed », par contre l'expérience ne permet pas de conclure à une action directe ou indirecte : PHOSPHORYLATE(x,y,b,unknown,confirmed,e).

L'assemblage de ces prédicats permet de construire la règle suivante :

• **IF** PA(x,y,a,e) **AND** ANTIBODYAGAINST(a,b) **AND** PHOSPHOFORM(b,y,pos) **THEN** PHOSPHORYLATE(x,y,b,unknown,confirmed,e)

3.2- Création de règles complexes

Les règles complexes permettent de formaliser le raisonnement par la combinaison de faits déduits. Elles permettent de déduire de nouvelles relations, d'étendre les propriétés de molécules ou de préciser certaines relations notamment en ce qui concerne leur distance. Elles permettent également de « transférer » les propriétés démontrées pour un agoniste au ligand naturel.

<u>Création de nouvelles relations avec une précision sur la distance :</u> Souvent les conclusions obtenues comportent une distance inconnue, or il est possible de conclure qu'une relation est indirecte si les molécules concernées ne se situent pas dans le même compartiment cellulaire. Ainsi, un prédicat a été créé pour exprimer la distance entre compartiments dans lesquels se trouvent les molécules x et y: COMPARTDIST(x,y,distance). Cette distance prend pour

valeur *same*, *neighbour* ou *distant*, permettant d'indiquer si les compartiments sont les mêmes, connexes, ou disjoints. Ce fait peut être déduit de la règle suivante :

IF LOCALIZATION(y,cytosol) AND NOLOCALIZATION(y,extracellular) AND NOLOCALIZATION(y,plasmamembrane) AND LOCALIZATION(x,extracellular) AND NOLOCALIZATION(x,plasmamembrane) AND NOLOCALIZATION(x,cytosol) THEN COMPARTDIST(x,y,distant)

<u>Signification</u>: Si la molécule *y* est localisée dans le cytosol et n'est pas localisée dans le milieu extracellulaire ni dans la membrane plasmique et si la molécule *x* est localisée dans le milieu extracellulaire et n'est pas localisée dans le cytosol ni dans la membrane plasmique alors les deux molécules *x* et *y* sont distantes l'une de l'autre.

Ces nouveaux faits COMPARTDIST peuvent ainsi être utilisés afin de préciser des relations :

• **IF** PHOSPHORYLATE(x,y,b,unknown,confirmed,e) **AND** COMPARTDIST(x,y,distant) **THEN** PHOSPHORYLATE(x,y,b,indirect,confirmed,e) **AND** PHOSPHORYLATE(x,y,b,direct,invalidated,e)

Signification: Si la molécule x a un effet e confirmé sur la phosphorylation de la protéine y pour donner b avec une distance inconnue et si les molécules x et y sont distantes l'une de l'autre alors la molécule x a un effet e indirect confirmé sur la phosphorylation de la protéine y pour donner b.

<u>Utilisation des données concernant les agonistes</u>: Un agoniste est une molécule capable de mimer les effets d'une molécule naturelle grâce à la fixation sur un même récepteur⁹. Il est donc possible de « transférer » à une molécule z les faits déduits pour son agoniste x (prédicat AGONIST(x,z)):

• **IF** PHOSPHORYLATE(x,y,b,d,s,e) **AND** AGONIST(x,z) **THEN** PHOSPHORYLATE(z,y,b,d,s,e)

Signification: Si x a un effet e sur la phosphorylation de la protéine y pour donner b avec une distance d et un statut s et si x est un agoniste de z alors z a un effet e sur la phosphorylation de la protéine y pour donner b avec une distance d et un statut s.

Il faut noter que cette règle est valable pour les agonistes complets, elle ne peut en aucun cas être utilisée dans le cas d'agonistes biaisés.

_

⁹ Il est également possible dire agoniste du récepteur.

Tous les prédicats utilisés dans notre méthode sont disponibles dans le dictionnaire des données en Annexe 3. La description de la création de toutes les règles est disponible dans l'Annexe 4.

II- Stockage des données

La méthode d'inférence automatique présentée dans cette thèse implique l'utilisation d'un nombre important de règles et de faits initiaux, et permet l'obtention d'un nombre également important de faits déduits. De plus, comme nous l'avons vu, la syntaxe des règles et des faits est contraignante et peu intuitive. Afin de faciliter l'ensemble du processus, nous avons créé une base de données qui contient à la fois les règles, les faits initiaux (ainsi que des détails sur leur provenance) et les faits déduits. Nous avons également généré un ensemble de scripts qui permettent d'exporter vers SOLAR les règles et les faits initiaux, d'importer les faits déduits et de les traduire en un fichier lisible par CellDesigner.

A-Règles et prédicats

La Figure 36 montre le schéma de base de données pour le stockage des règles et des prédicats. Les prédicats sont stockés dans la table predicate. Les règles sont stockées dans la table rule. Comme cela a été décrit plus haut, une règle est composée de un ou plusieurs prédicats en condition et un ou plusieurs prédicats en conclusion. La table predicate_compose_rule permet de faire le lien entre les règles et les prédicats qui les composent, le champ condition de cette table permet de préciser si le prédicat apparaît en condition ou en conclusion de la règle. Ainsi une règle est liée à au moins deux prédicats, un prédicat est lié à zéro, une ou plusieurs règles. Un prédicat peut provenir d'un protocole expérimental (table experimenttype) mais ce n'est pas obligatoirement le cas (cardinalité 0...*)¹⁰. Un protocole est lié à au moins un prédicat. La cardinalité de la relation est définie comme 0...*, alors qu'elle devrait être 1...*, pour permettre de créer des types d'expériences pour lesquelles les prédicats et les règles n'ont pas encore été définis. La table predicate_experimenttype permet de faire la liaison. Dans le cas où un perturbateur est utilisé, son type est défini dans la table perturbatortype (actuellement les types de perturbateurs définis sont : Inhibiteur, siRNA, Antagoniste). L'expérience perturbée (par exemple un Western Blot en présence d'un siRNA) est définie comme un nouveau type d'expérience dans la table *experimenttype*, et est liée à l'expérience non perturbée (par l'intermédiaire de la table perturbated_experimenttype) et au type de perturbateur (par l'intermédiaire de la table

¹⁰ L'étoile * signifie « plusieurs ».

experimenttype_perturbatortype). Une expérience peut être non perturbée ou comprendre plusieurs perturbations et, une perturbation peut être présente dans 0 à plusieurs types d'expériences. Un résultat d'expérience (table *experiment*) est lié à un et un seul type d'expérience. Au contraire, un type d'expérience peut être utilisé dans de nombreuses expériences.

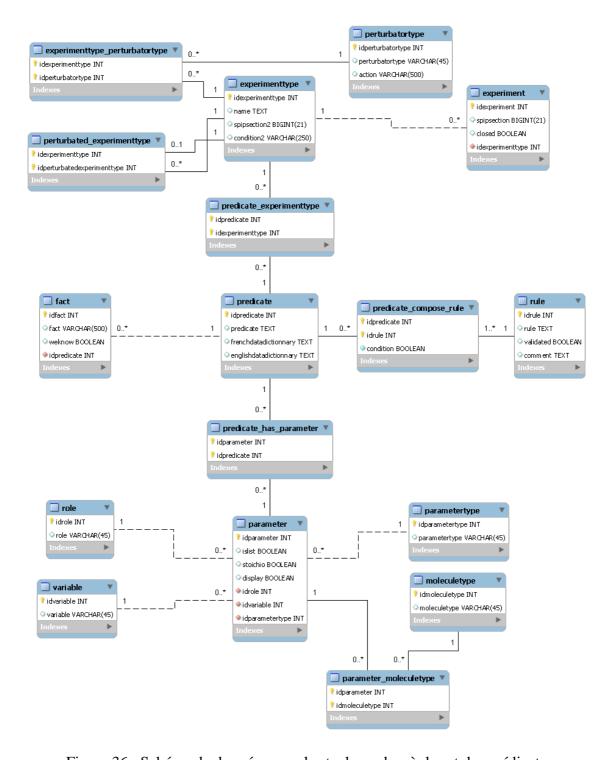


Figure 36 : Schéma de données pour le stockage des règles et des prédicats

Les faits sont des prédicats instanciés, il existe donc un lien entre les prédicats (table *predicate*) et les faits (table *fact*). La structure concernant les faits sera expliquée plus loin.

Les prédicats peuvent contenir de 0 à plusieurs paramètres (0...*). Un paramètre peut ne pas être utilisé dans un prédicat ou peut être utilisé dans un ou plusieurs prédicat (0...*). Les paramètres ont obligatoirement un rôle (molécule ajoutée, molécule d'intérêt...), la cardinalité 0...* est utilisée pour faciliter l'insertion des données. Les paramètres ont obligatoirement un type (distance, effet, molécule, compartiment...). Les paramètres sont obligatoirement représentés par une variable (x,y,z,...). Si le paramètre est une molécule, cette molécule peut provenir d'une ou plusieurs classes. Par exemple, un paramètre qui a pour rôle *catalyseur* dans une réaction enzymatique peut provenir des classes Protéine, Complexe ou Ribozyme. De cette façon, nous mettons en place un système de restriction des types moléculaires possibles pour chaque argument (si c'est une molécule) d'un prédicat. Le lien entre les paramètres et les types moléculaires (table *moleculetype*) est fait par l'intermédiaire de la table *parameter_moleculetype*.

B- Molécules

Le schéma des données pour le stockage des molécules est présenté sur la Figure 37. Les molécules sont répertoriées dans une table *molecule* et chacune d'entre elles est reliée à une des classes présentées précédemment (table *moleculetype*). La hiérarchisation des molécules est prise en charge par la table *moleculetypehierarchy*, c'est-à-dire que dans cette table, les sous-classes sont liées à la classe supérieure.

Au sein des cellules, il existe une compartimentation (cytosol, noyau, membrane plasmique...), celle-ci est répertoriée dans la table *cellcompartment*. Cette compartimentation permet d'associer aux molécules une localisation cellulaire quand l'information est disponible (table *moleculelocation*). La table *molecule* contient les champs qui sont communs à tous les types de molécules, les tables *antibody*, *smallmolecule*, *regDNA*, *RNA*, *gene* et *protein* ont été créées de façon à prendre en compte les données qui sont spécifiques au type correspondant. Une molécule ne peut faire partie que d'une et une seule classe, de ce fait les liaisons sont notées 0...1 sur les tables représentant les différentes classes, cela signifie par exemple, qu'une molécule est un gène ou ne l'est pas.

Il existe des liens biologiques entre certaines molécules, en particulier la table *gene_to_RNA* faisant le lien entre un gène et l'ARN correspondant, la table *RNA_to_protein* fait le lien entre un ARN messager et la (les) protéine(s) traduite(s). Les complexes sont des molécules et sont liés à leurs composants (table *complexmolecule*).

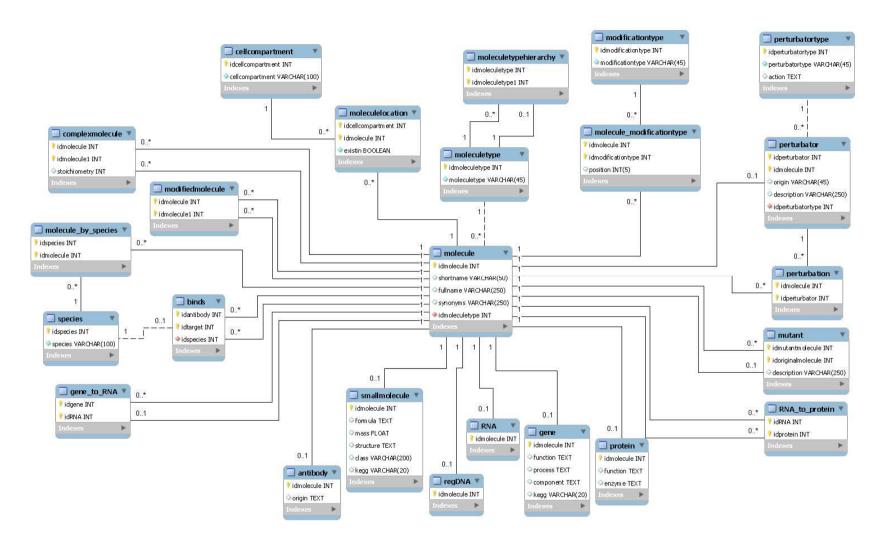


Figure 37 : Schéma de données pour le stockage des molécules

Une molécule modifiée est répertoriée dans la table *molecule* et est liée à la molécule non modifiée (table *modifiedmolecule*), ainsi qu'au type de modifidation (table *modificationtype*), il y a donc une table intermédiaire faisant le lien entre la molécule modifiée et le type de modification qu'elle porte, ainsi que sa position (table *molecule_modificationtype*).

Les molécules, telles que les gènes, les ARN, les protéines et les complexes, ne sont pas identiques d'une espèce à l'autre, c'est pourquoi la base de données permet une liaison entre les molécules et les espèces (table *species*) grâce à la table *molecule_by_species*.

En ce qui concerne les anticorps, il est nécessaire de connaître la (les) molécule(s) qu'ils ciblent, la table *binds* permet de relier un anticorps à sa ou ses cibles.

La table *perturbator* regroupe la description des molécules perturbatrices, elle est liée à la table *perturbatortype*. Si une molécule est présente dans cette table, elle l'est également dans la table *molecule*. Lorsqu'une molécule perturbatrice a une ou plusieurs cibles, la table *perturbation* permet de faire le lien.

C- Données expérimentales

La Figure 38 montre le schéma de données permettant le stockage des expériences. Les expériences, qui permettent de générer les faits, sont stockées dans la table *experiment*. Chaque expérience est liée à un type d'expérience (table *experimenttype* dans la partie Règles et prédicats).

Les expériences doivent avoir une source, qui peut soit être une publication (table *publication*), soit provenir d'une autre source, par exemple de notre équipe (table *otherreference*).

Une expérience suit une méthode (table *method*). Elle peut être réalisée sur un type cellulaire particulier (table *cellularmaterial*). Il existe des kits pour certaines expériences, ils sont stockés dans la table *kit* et dans le cas où une expérience nécessite un kit, le couple kit/expérience est donné dans la table *experiment_kit*. Il peut arriver que les chercheurs utilisent un autre matériel (exemple : un logiciel) pour leurs expériences. Il sera stocké dans la table *othermaterial* et le lien entre lui et l'expérience est effectué grâce à la table *experiment_othermaterial*.

Si les auteurs des expériences veulent ajouter un fichier, par exemple une image, la table *file* stocke le chemin pour accéder à l'image et le lien avec l'expérience se fait via la table *experiment_file*.

Il est possible que les expériences soient réalisées à partir de matériel provenant d'un compartiment cellulaire particulier (tables *cellcompartment* et *experiment_cellcompartment*). La table *experiment_molecule* permet de relier une expérience avec les molécules concernées. Ces molécules ont un rôle au sein de l'expérience tel que 'molécule observée' par exemple. Les rôles sont répertoriés dans la table *role* et l'identifiant d'un rôle est repris dans la table *experiment_molecule* pour chaque couple molécule/expérience.

Souvent, les chercheurs veulent connaître les variations cellulaires que peut entraîner un changement des conditions environnementales (exemples: ajout d'une hormone, lumière/obscurité...), nous avons choisi de nommer *signal* ce type de changement. Ils sont stockés dans une table *signal* et les couples expérience/signal sont présents dans la table *experiment_signal*.

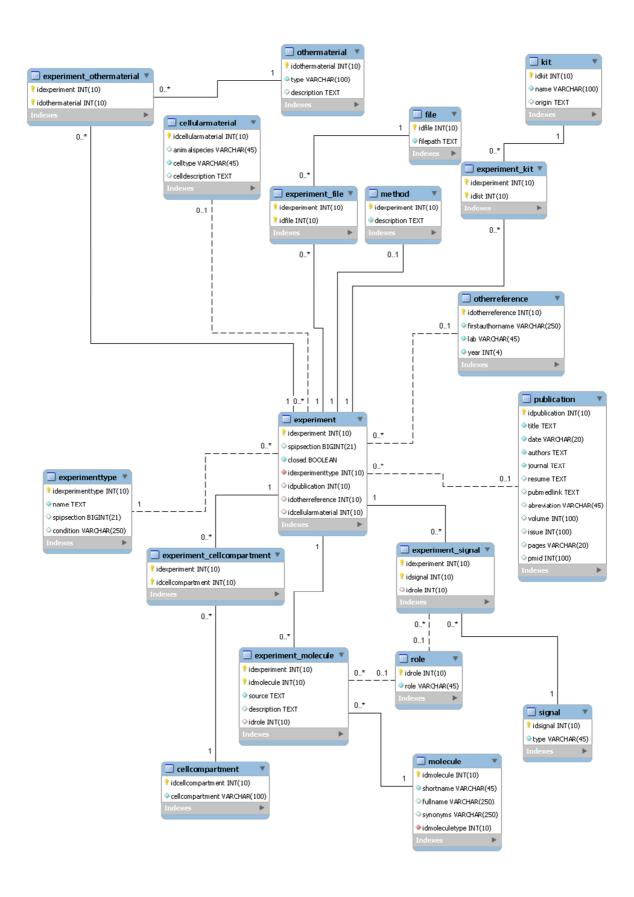


Figure 38 : Schéma de données pour le stockage des expériences

Les expériences conduisent à des faits expérimentaux, la Figure 39 correspond au schéma de données pour le stockage des faits.

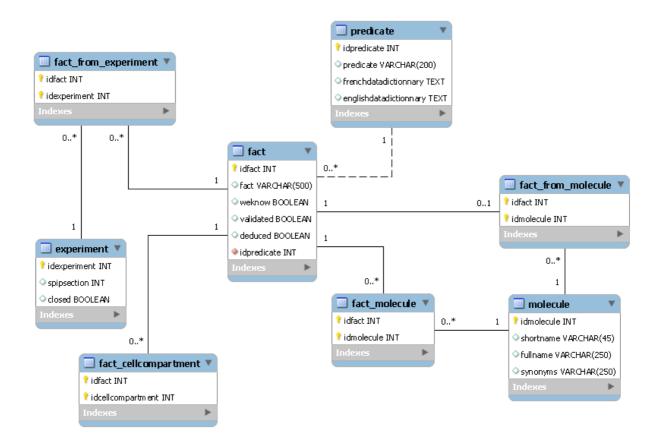


Figure 39 : Schéma de données pour le stockage des faits

Les faits sont stockés dans la table fact et sont liés aux expériences grâce à la table fact_from_experiment. Un fait peut provenir de 0 ou plusieurs expériences et une expérience peut aboutir à 0 ou plusieurs faits. Comme cela a été dit plus tôt, un fait est un prédicat instancié donc un fait est lié à un et un seul prédicat (table predicate). Différents faits peuvent représenter l'instanciation d'un même prédicat. Un fait peut être valable uniquement dans un compartiment cellulaire, par exemple, il existe des expériences de localisation qui aboutissent à ce type de faits, c'est pourquoi une table fact_cellcompartment a été créée de façon à stocker les couples fait/compartiment. Dans le cas où les expériences sont réalisées sur des molécules, les faits contiennent obligatoirement ces molécules en tant que variables, les couples fait/molécule sont présentés dans la table fact_molecule. Un fait ne provient pas obligatoirement d'une expérience, par exemple l'instanciation des prédicats PHOSPHOFORM(b,y,pos) ou ANTIBODYAGAINST(a,y), n'est pas liée à des expériences.

La table *fact_from_molecule* présente les couples fait/molécule dans le cas où les faits sont relatifs uniquement à des molécules comme dans les deux exemples cités.

Une interface est également en cours de développement, elle permettra la saisie des données et notamment la création automatique de faits. Par exemple, lorsqu'un utilisateur ajoutera une protéine comportant une phosphorylation, le fait PHOSPHOFORM sera créé automatiquement à partir des données fournies. De plus, l'insertion d'une protéine à partir de son identifiant Uniprot permet la récupération de l'identifiant de son gène et de son ARN messager d'origine. Ces données alimentent les tables <code>gene_to_RNA</code> et <code>RNA_to_protein</code> et la création automatique des faits portant sur les prédicats TRANSCRIPTION et TRANSLATION. L'insertion de perturbateurs sera également accompagnée de la génération automatique des faits tels que INHIBITOR, SIRNAAGAINST ou ANTAGONIST.

La méthode que nous avons suivie permet, si cela est nécessaire, d'étendre la base de données. Par exemple, si les tables actuelles ne sont pas suffisantes en ce qui concerne les molécules, il sera tout à fait possible d'en créer de nouvelles. Il faut noter aussi que le fait de limiter les types moléculaires (natures biochimiques) possibles pour les arguments des prédicats au moment de la saisie des données permet d'éviter l'entrée d'erreurs lors de l'instanciation de ces prédicats.

III- Visualisation des réseaux

Pour présenter les réseaux de signalisation obtenus, nous avons choisi le logiciel CellDesigner (choix expliqué dans la partie « Une nouvelle méthode d'inférence de réseaux de signalisation »). Dans cette partie, nous allons présenter le logiciel et expliquer la transformation des faits en éléments du réseau.

1- CellDesigner

CellDesigner est un éditeur gratuit et multi-plateforme de représentation graphique des réseaux. Les créateurs de CellDesigner ont développé un langage qui est propre à ce logiciel, c'est une extension du SBML par le langage SBGN décrivant les diagrammes de processus (http://www.celldesigner.org/) [192, 193]. Il est également lié à des packages de simulation et d'analyse de réseaux par l'intermédiaire de *Systems Biology Workbench* (SBW). CellDesigner peut également faire de l'analyse de paramètres grâce au *SBML ODE Solver* ou COPASI. CellDesigner permet aussi de relier les éléments d'un réseau à diverses bases de données, il est donc possible d'associer une protéine à la page de description de son gène d'origine dans

est donc possible d'associer une protéine à la page de description de son gène d'origine dans la banque de données EntrezGene grâce à son identifiant ou encore de lier les relations à des articles répertoriés dans la banque de données PubMed. Chaque élément du réseau peut également porter un identifiant de l'ontologie SBO.

2- Transformation des faits en éléments du réseau

La transformation des faits en réseau est réalisée en deux étapes. Tout d'abord, il faut représenter les nœuds du réseau (molécules), puis les liaisons entre les molécules grâce aux faits initiaux ou déduits.

Molécules: Elles sont, comme dit précédemment, les nœuds du réseau. La Figure 40 présente la convention de représentation utilisée pour les molécules dans CellDesigner. La classification présentée plus tôt est retrouvée ici. Seules les molécules d'ADN correspondant à des régions régulatrices sont absentes mais dans CellDesigner, il est possible de les ajouter directement sur les gènes (Figure 41).

Pour les protéines, il est possible de visualiser des protéines dites « génériques », et il existe une symbolique particulière pour les récepteurs et les canaux ioniques.

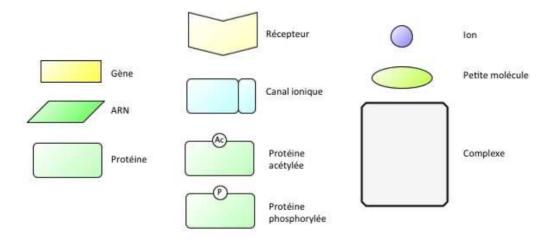


Figure 40 : Convention de représentation utilisée pour la représentation des molécules dans CellDesigner

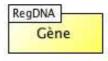


Figure 41 : Gène représenté avec une région régulatrice

Les molécules sont extraites de la table *molecule* de la base de données. La représentation graphique est déterminée par la classe de la molécule. Les molécules extraites correspondent aux classes Gène, ARN, Protéine, Petite molécule et Complexe. En revanche, sont exclues les protéines appartenant aux sous-classes Anticorps et Mutant. Sont aussi exclues toutes les perturbations (Inhibiteur, siRNA, Antagoniste) et les agonistes/analogues structuraux. Ces différents types moléculaires sont des 'outils' d'aide à la construction du réseau mais n'en font, en réalité, pas partie.

Lorsque les régions régulatrices seront utilisées, elles seront extraites de la même façon.

Les protéines sont représentées par la notation correspondant aux récepteurs si le mot 'receptor' apparaît dans leur nom complet. Les canaux ioniques ne sont pas apparus dans nos premiers réseaux, mais il sera possible de les visualiser avec la représentation graphique appropriée en cherchant 'channel' dans leur nom complet. Pour les protéines modifiées une étape supplémentaire est nécessaire : on les fait apparaître comme des protéines (génériques, récepteurs, canaux ioniques), ensuite la (les) modification(s) ainsi que sa (leur) position sont récupérées dans la table molecule_modificationtype et ajoutées à la visualisation.

Relations: Les faits nouveaux obtenus par inférence sont des relations entre les molécules du réseau. Cependant, certaines de ces relations ne sont pas incluses dans la représentation graphique car elles ne sont pas pertinentes pour l'étude biologique. En particulier, les relations dont l'effet est nul (noeffect) ne sont pas représentées, de même que les relations impliquant les agonistes, antagonistes et mutants car ces molécules ne font pas réellement partie des réseaux. En effet, les faits qui démontrent la présence de ces relations sont nécessaires au raisonnement, mais ne sont pas utiles à la compréhension du système. En ce qui concerne les prédicats convertibles en schémas CellDesigner, la Figure 42 illustre quelques exemples.

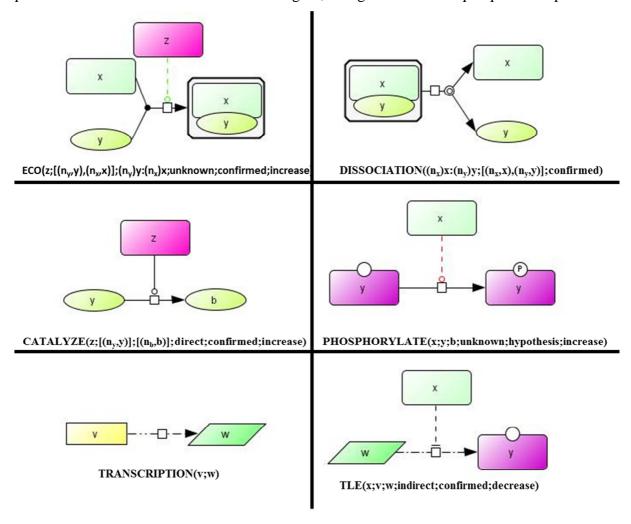


Figure 42 : Exemples de prédicats convertis en représentation graphique CellDesigner ECO : Effet sur la complexation, TLE : Effet sur la traduction

Légende des représentations graphiques :

Un code couleur a été mis au point pour mettre en évidence les différents statuts et distances. Relations confirmées :

• Si la relation est directe, elle apparaît en trait plein noir,

- Si la relation est indirecte, elle apparaît en pointillés noirs,
- Si la relation a une distance inconnue, elle apparaît en pointillés verts,
- Si la relation est admise (issue du savoir de l'expert mais pas inférée à partir de faits expérimentaux), elle apparaît en parme.

Relations hypothétiques:

- Si la relation est directe, elle apparaît en trait continu bleu,
- Si la relation est indirecte, elle apparaît en pointillés bleus,
- Si la relation a une distance inconnue, elle apparaît en pointillés rouge.

Extraction des relations:

Toutes les relations sont extraites de la base de données à partir de la table *fact*. Dans les cas de modification, transcription et traduction, la relation est extraite à partir des tables *modifiedmolecule*, *gene_to_RNA* et *RNA_to_protein* respectivement (partie A Figure 43). Ensuite, l'action d'une molécule sur ces relations est ajoutée si un fait le précise (partie B Figure 43).

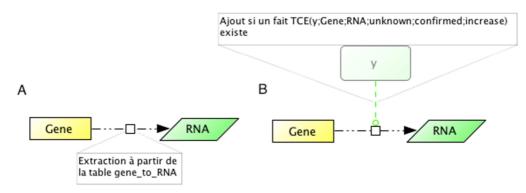


Figure 43 : Extraction des relations dans la base de données

Les relations de complexation sont extraites à partir de la table *complexmolecule*, cette table fait le lien entre les complexes et leurs composants. Ensuite, ces relations sont directement converties en représentation CellDesigner. Si un effet sur une complexation existe, cette régulation est ajoutée sur la complexation concernée. En ce qui concerne les relations de dissociation, la situation est différente car il est possible d'avoir un complexe formé à partir de molécules « isolées » et ce même complexe pourrait se dissocier en, par exemple, deux complexes plus petits. C'est pourquoi, un fait de dissociation est extrait à partir de la table *fact* et est directement converti en graphique.

Signalisation induite par l'hormone folliculostimulante (FSH) L'hormone FSH est au cœur du projet de recherche de l'équipe BIOS, c'est pourquoi nous avons choisi le réseau de signalisation intracellulaire déclenché par cette hormone comme modèle d'élaboration de notre méthode automatique d'inférence de réseaux. Afin de valider les capacités de notre méthode à exploiter au maximum les données expérimentales disponibles, mais également dans le but d'estimer ses limitations en termes de taille des réseaux inférés, nous avons cherché à reconstituer le réseau le plus large possible. Les données expérimentales qui ont permis la construction manuelle de ce réseau sont issues de la littérature et résultent de l'analyse de plus de 100 articles. Une partie du réseau ainsi construit a été publiée dans une revue incluse ci-après.

Dans les figures de l'article, les schémas des voies de signalisation (réalisés avec CellDesigner) montrent non seulement les relations du réseau, mais également pour chacune la ou les publications dans lesquelles apparaissent des expériences qui justifient son existence. L'objectif est non seulement de permettre à l'expert qui étudie le réseau de vérifier lui-même l'information, mais également d'évaluer le degré de confiance qui peut être accordé à l'existence de cette relation. Ceci sera particulièrement important dans les perspectives de la méthode automatique.

I- Les voies classiques

Titre: Mapping the follicle-stimulating hormone-induced signaling networks.

Pauline Gloaguen, Pascale Crépieux, Domitille Heitzler, Anne Poupon et Eric Reiter

Mapping the follicle-stimulating hormone-induced signaling networks

Pauline Gloaguen^{1,2,3,4}, Pascale Crépieux^{1,2,3,4}, Domitille Heitzler^{1,2,3,4}, Anne Poupon^{1,2,3,4} and Eric Reiter^{1,2,3,4} **

- [†] BIOS Group, INRA, UMR85, Unité Physiologie de la Reproduction et des Comportements, Nouzilly, France
- ² UMR6175, CNRS, Nouzilly, France
- 3 Université François Rabelais, Tours, France
- 4 L'Institut Français du Cheval et de l'Équitation, Nouzilly, France

Edited by:

Sandhya Srikant Visweswariah, Indian Institute of Science, India

Reviewed by:

Suraj Unniappan, York University, Canada Rajan R. Dighe, Indian Institute of Science, India

*Correspondence:

Eric Reiter, INRA UMR85, CNRS-Université François Rabelais UMR6175, 37380, Nouzilly, France. e-mail: Eric.Reiter@tours.inra.fr Follicle-stimulating hormone (FSH) is a central regulator of male and female reproductive function. Over the last decade, there has been a growing perception of the complexity associated with FSH-induced cellular signaling. It is now clear that the canonical Gs/cAMP/PKA pathway is not the sole mechanism that must be considered in FSH biological actions. In parallel, consistent with the emerging concept of biased agonism, several examples of ligand-mediated selective signaling pathway activation by gonadotropin receptors have been reported. In this context, it is important to gain an integrative view of the signaling pathways induced by FSH and how they interconnect to form a network. In this review, we propose a first attempt at building topological maps of various pathways known to be involved in the FSH-induced signaling network. We discuss the multiple facets of FSH-induced signaling and how they converge to the hormone integrated biological response. Despite of their incompleteness, these maps of the FSH-induced signaling network represent a first step toward gaining a system-level comprehension of this hormone's actions, which may ultimately facilitate the discovery of novel regulatory processes and therapeutic strategies for infertility and non-steroidal contraception.

Keywords: follicle-stimulating hormone, receptor, signaling network, topological map, systems biology

INTRODUCTION

Follicle-stimulating hormone (FSH) plays a central role in the control of reproduction. FSH is a heterodimeric pituitary glycoprotein consisting of an α-subunit, which is common to other glycoprotein hormones, and a specific β-subunit (Papkoff and Ekblad, 1970). FSH binds to and activates the FSH receptor (FSHR), which belongs to the 7 transmembrane domains receptor (7TMR) family, also known as G protein-coupled receptors. The FSHR is expressed in Sertoli cells in testis and granulosa cells in ovaries (Simoni et al., 1997). FSH is required for normal growth and maturation of ovarian follicles in women and for normal spermatogenesis in men (Themmen and Huhtaniemi, 2000). Knock-out of the FSH β-subunit or the FSHR genes in mice result in significant reproductive defects in both sexes (Kumar et al., 1997; Dierich et al., 1998). Consistently, inactivating mutations in either FSH β subunit or FSHR led to similar reproductive defects (Matthews et al., 1993; Aittomaki et al., 1995; Layman et al., 1997; Huhtaniemi et al., 2006). Depending on the physiological situation, FSH has to control distinct, sometime opposite, integrated biological responses in its target cells, ranging from differentiation, cellular metabolism, steroidogenesis, proliferation, and apoptosis (Dias et al., 2010).

Due to its ability to control reproduction, either native or recombinant FSH preparations have been extensively used in reproductive medicine and animal breeding (Lunenfeld, 2004; Macklon et al., 2006). However, the use of FSH remains associated with significant drawbacks such as the risk of triggering ovarian hyperstimulation syndrome (OHSS; Vloeberghs et al., 2009)

or heterogeneous responsiveness (Loutradis et al., 2003, 2004). Conversely, it has been proposed that FSHR blockers could potentially represent a novel non-steroidal approach for contraception (Naz et al., 2005). A causal role of FSH in the etiology of ovarian epithelial cancer has also been proposed many years ago (Choi et al., 2007). In addition, it has been recently reported that FSHR is ectopically expressed by endothelial cells associated with the angiogenesis of a wide panel of tumors, pointing out to the FSHR as a common marker for early diagnosis of most cancers (Radu et al., 2010). In addition, this finding also raises the intriguing possibility of a direct role played by FSH in promoting early tumoral angiogenesis. In this general context, selective pharmacological modulators of FSHR would be of great interest.

Noteworthy, it is increasingly accepted that certain ligands can lead to selective activation of signaling pathways by binding at their cognate 7TMRs in a process referred to as biased agonism (Kenakin, 2005; Rajagopal et al., 2010; Whalen et al., 2011). In line with this emerging concept, it has been recently reported that certain hormone glycosylation variants, potentiating antibodies and small molecule ligands can trigger biased responses at the FSHR (Arey et al., 2008; Wehbi et al., 2010a,b; Dias et al., 2011; Ulloa-Aguirre et al., 2011). It has also been shown that FSHR can be directly stimulated by antibodies, expending even more the potential to develop biased agonists at this receptor (Agrawal and Dighe, 2009). Paralleling the notion of biased agonism, it is now generally accepted that the canonical G protein-dependent coupling is not the unique mechanism leading to 7TMRs signaling

(Lefkowitz and Shenoy, 2005; Reiter and Lefkowitz, 2006). The FSHR coupling to the classical $G\alpha_s$ /cAMP/protein kinase A (PKA) signaling pathway, which had been acknowledged as the sole effector mechanism of FSH for more than 20 years, is now viewed as one of several mechanisms contributing to the activation of a complex integrated signaling network. Deciphering the organization and dynamic functioning of this network is an important challenge that the field has now to address as it may help identify novel regulatory processes and therapeutic strategies for infertility and non-steroidal contraception.

In the present review paper, we attempted to integrate the data available on FSH-induced signaling by building a series of topological maps in the CellDesigner modeling environment (Kitano et al., 2005). CellDesigner is a structured diagram editor for drawing biochemical and gene-regulatory networks. Networks are drawn based on the graphical notation system proposed by Kitano et al. (2005). Models are stored using the Systems Biology Markup Language (SBML), a standard for representing models of biochemical and gene-regulatory networks. Because they are coded in SBML, CellDesigner models are easily shared or linked with simulation and other analysis packages through Systems Biology Workbench (SBW). For each reaction, the relevant references are indicated within the network representation, providing a unique and user-friendly integration of the available knowledge. For simplification purpose, we chose to only represent monomers in the maps throughout the paper although it is now well established that FSHR is expressed as dimers at the plasma membrane (Thomas et al., 2007; Guan et al., 2010). In the future, as the data become more abundant, it will be possible and more rigorous to draw separate models for each physiological situation (i.e., Sertoli vs. granulosa cells, young vs. mature, as a function of the species, endogenous vs. heterogeneously expressed receptors, etc.). However, in the present paper, in order to deal with the relative scarcity of available data for certain mechanisms, we decided to aggregate data even if they were generated in different cellular models.

MULTIPLE TRANSDUCTION MECHANISMS ARE ACTIVATED BY FSH

The $G\alpha_s$ /cAMP/PKA signaling pathway has been considered for more than 20 years as the key mechanism relaying FSH biological action inside target cells (Means et al., 1974; Dattatreyamurty et al., 1987). However, it is now amply documented that FSHR also engages other transduction mechanisms upon FSH binding. **Figure 1** shows, in CellDesigner format, the known G protein-dependent and -independent transduction mechanisms engaged at the FSHR.

Early assumption of Gαs/cAMP/PKA-independent signaling at the FSHR comes from the observation that adenovirus-mediated transduction of a constitutively active Gαs (Gαs Q227L) in undifferentiated rat granulosa cells was not capable of inducing the expression of differentiation markers such as aromatase or luteinizing hormone receptor (Zeleznik et al., 2003). Overexpression of a constitutively active form of protein kinase B (PKB/Akt) was required to fully restore the expression of these two markers (Zeleznik et al., 2003). Besides, PKB activation had been reported to depend in some instance on cAMP (Meroni et al., 2002; Alam

et al., 2004), independently of PKA (Gonzalez-Robayna et al., 2000), suggesting that exchange protein directly activated by cAMP (EPAC) was also involved in the process. A role for EPAC in FSH biological action has been later demonstrated (Wayne et al., 2007). To conclude, PKB seems to be activated both by G α s-dependent mechanisms, likely via EPAC, and by G α s-independent mechanisms, in FSH-stimulated cells. Interestingly, quantitatively weaker gene expression was observed with a lentiviral vector overexpressing a constitutively active PKA (PKA-CQR), when compared to FSH (Escamilla-Hernandez et al., 2008a), suggesting again that both PKA-dependent and PKB-dependent signaling pathways are required for full FSH biological activity.

Consistent with the idea of Gas-independent transduction mechanisms at the FSHR, it has been reported that this receptor couples to other G protein subtypes. Indeed, FSHR has been demonstrated to trigger pertussis toxin-sensitive pathways upon activation by certain hormonal variants (Arey et al., 1997) or in particular developmental stage of the target cells (Crépieux et al., 2001). In addition, the FSHR has also been shown to activate the inositol trisphosphate (IP3) signaling pathway, particularly at high doses of agonist (Quintana et al., 1994). Consistent with these data, FSHR directly interacts with $G\alpha_q$ subunit overexpressed in granulosa cells (Escamilla-Hernandez et al., 2008b). An alternative transduction mechanism has been reported in Sertoli cells and could also explain the observed FSH-induced IP3 response. According to this mechanism, FSHR functionally couples to Gαh, also known as tissue transglutaminase, which leads to PLC8 activation and IP3 accumulation (Lin et al., 2006).

Beside heterotrimeric G proteins, two protein families have been reported to specifically interact with the FSHR upon FSH stimulation: G protein-coupled receptor kinases (GRKs) and βarrestins. Originally, GRKs and β-arrestins have been shown to control the desensitization, internalization, and recycling of FSHR (Nakamura et al., 1998; Lazari et al., 1999; Troispoux et al., 1999; Reiter et al., 2001; Kishi et al., 2002; Marion et al., 2002, 2006; Krishnamurthy et al., 2003a,b; Piketty et al., 2006). Over the last 10 years, the perception of β-arrestins' functions has expended as they have been shown to act as G protein-independent signal transducers at many 7TMRs (Lefkowitz and Shenoy, 2005; Reiter and Lefkowitz, 2006) including the FSHR (Kara et al., 2006; Wehbi et al., 2010a,b; Tranchant et al., 2011). In the case of the FSHR, βarrestin-dependent and G protein-independent activation of ERK and rpS6 have been reported so far. However, it is likely that β arrestins are involved in the G protein-independent activation of a wide array of signaling pathways at the FSHR since they act as multifunctional scaffolds interacting with many protein partners (Xiao et al., 2007) and facilitating the phosphorylation of numerous intracellular targets (Xiao et al., 2010) at other 7TMRs (see Whalen et al., 2011 for a recent review).

The adaptor protein containing a PH domain, PTB domain, and leucine zipper motif 1 (APPL1) has also been reported to bind directly to the FSHR and trigger downstream signaling mechanisms. APPL1 has been suggested to interact with the first intracellular loop of the FSHR and to mediate FSH-dependent PI3K signaling (Nechamen et al., 2004). Recently, Thomas et al. (2011) have shown that the interaction of APPL1 with the FSHR is required for the activation of the inositol–phosphate calcium

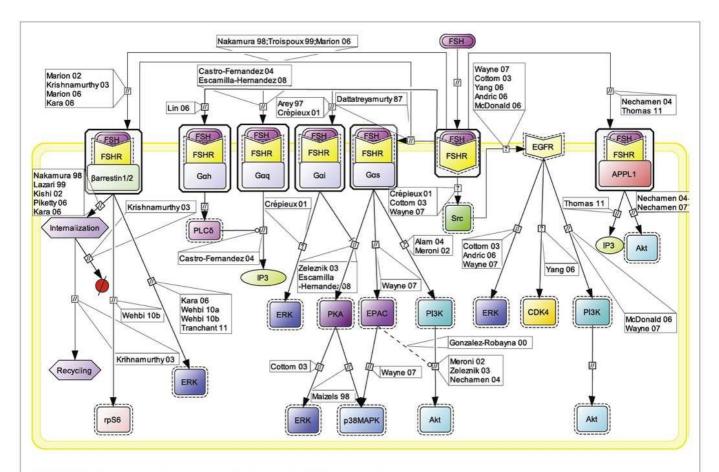


FIGURE 1 | Multiple transduction mechanisms induced at the FSHR. The $G\alpha$ s/cAMP/protein kinase A (PKA) signaling pathway has classically been considered as the key signaling mechanism triggered at the FSHR. Over the last decade, mounting evidence have shown that FSHR also engages other transduction mechanisms mainly coupling to other $G\alpha$ subunits, β -arrestin-dependent signaling, EGFR transactivation, and APPL1-mediated signals. Bibliographic references are indicated on the map

by the first author's name with two digits for the year of publication. The

CellDesigner program has been used to represent signaling pathways as shown in **Figures 1–7**. Complexes are surrounded by a box. Dashed lines indicate indirect reactions. The following semantic has been used: proteins (); active protein (); receptor (); transcript (); gene (); catalysis (); association (); dissociation (); inhibition (); phosphorylation (); degradation (); known transition omitted (); unknown transition ()

pathways upon FSH exposure. In addition, Src family members have also been reported to induce the PI3K pathway in granulosa cells (Wayne et al., 2007) and the ERK pathway in both Sertoli and granulosa cells (Crépieux et al., 2001; Cottom et al., 2003) upon FSH exposure.

Transactivation of the epithelial growth factor receptor (EGFR) also seems to play a role in the transmission of FSH signal within the target cells. Indeed, FSH has been shown to trigger EGFR autophosphorylation in granulosa cells through the activation of Src (Cottom et al., 2003; Wayne et al., 2007). Moreover, when EGFR is inhibited, a decrease in the ability of FSH to induce ERK or Akt phosphorylation and CDK4 activation has been observed in various models (Cottom et al., 2003; Andric and Ascoli, 2006; Shimada et al., 2006; Yang and Roy, 2006; Wayne et al., 2007).

In the remaining of this review, detailed topological maps for these different transduction mechanisms and their downstream signaling cascades will be presented and discussed.

THE CANONICAL GαS-DEPENDENT PATHWAY

As already stated, the G α s pathway has been the most studied and is associated with various intracellular events (**Figure 2**). Upon FSH activation, FSHR functionally couples to G α s subunit which in turn induces adenylate cyclase activity (Northup et al., 1980). The adenylate cyclase-mediated cAMP production is counterbalanced by phosphodiesterase (PDE) activity (Fakunding et al., 1976). Accumulated cAMP binds to and activates two distinct downstream effectors: PKA and EPAC.

Upon cAMP binding, PKA's catalytic subunits are released and activated (Rangel-Aldao and Rosen, 1976; Landmark et al., 1991). Active catalytic subunits of PKA subsequently phosphory-late a number of targets, either in the cytosol, or in the nucleus. Nuclear actions of PKA are treated in a separate section of this review. In the cytosol, FSH-induced ERK MAPK activation has been shown in several models to be sensitive to PKA inhibition (Crépieux et al., 2001; Kara et al., 2006). Two somehow contradictory mechanisms have been proposed to explain PKA-mediated

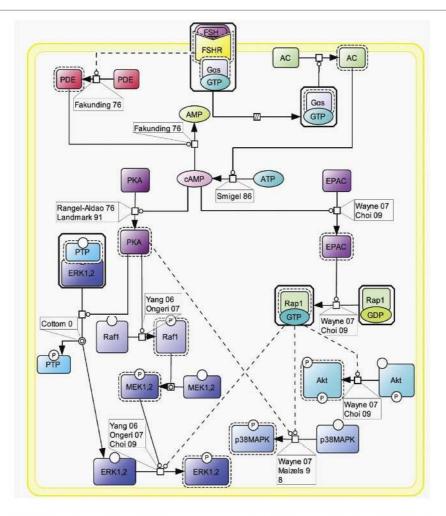


FIGURE 2 | Canonical Gas/cAMP/PKA pathway at the FSHR. The Gas/cAMP/PKA pathway has been the most studied and is associated with various intracellular events. It is now acknowledged that PKA is not the sole

target of cAMP accumulation. Indeed, EPAC is also activated upon FSH stimulation. Both PKA and EPAC contribute to the activation of MAPK ERK and p38, whereas EPAC also leads to Akt activation.

activation of ERK upon FSH stimulation. First, PKA-mediated phosphorylation of a protein phosphatase (PTP) has been shown to disrupt the PTP/ERK complex, thereby leading to an increase in ERK phosphorylation with MEK being constitutively phosphorylated (Cottom et al., 2003). Second, it has been reported that PKA promotes Raf1 activation, which in turn activates MEK and finally ERK (Yang and Roy, 2006; Ongeri et al., 2007). Further studies will be necessary to clarify this situation. In addition, PKA inhibition has also been shown to impair FSH-induced MAPK p38 phosphorylation and the associated cell rounding and steroidogenesis in granulosa cells (Maizels et al., 1998; Yu et al., 2005).

Evidence have been provided that EPAC acts as an effector downstream of FSH-induced cAMP accumulation. Upon cAMP binding, EPAC promotes Rap1-GDP to GTP exchange and its subsequent activation in granulosa (Wayne et al., 2007) and in surface epithelium ovary (Choi et al., 2009) cells. Once activated, Rap1 relays FSH-dependent activation of Akt, p38, and ERK (Wayne et al., 2007; Choi et al., 2009). Again, further investigation will be

necessary in order to delineate the exact contributions of PKA and EPAC in FSH signaling.

THE β-ARRESTIN-DEPENDENT PATHWAY

Initially, the β -arrestin pathway was viewed as controlling FSHR's desensitization and recycling. Similar to many other 7TMRs, this perception has progressively evolved toward a more general role of β -arrestins as adapters and transducers (Reiter and Lefkowitz, 2006; **Figure 3**).

The FSHR has been reported to be phosphorylated by GRKs 2, 3, 5, and 6 upon FSH binding to the receptor in various models (Nakamura et al., 1998; Lazari et al., 1999; Troispoux et al., 1999; Marion et al., 2002; Krishnamurthy et al., 2003a; Kara et al., 2006). A cluster of five serines and threonines located in the C terminus of the FSHR has been shown to account for the bulk of FSH-induced phosphorylation as a result of GRK2 action (Kara et al., 2006). It is also well documented that β -arrestins are recruited to the GRK-phosphorylated and agonist-occupied FSHR (Nakamura et al., 1998; Lazari et al., 1999; Troispoux et al., 1999; Marion et al.,

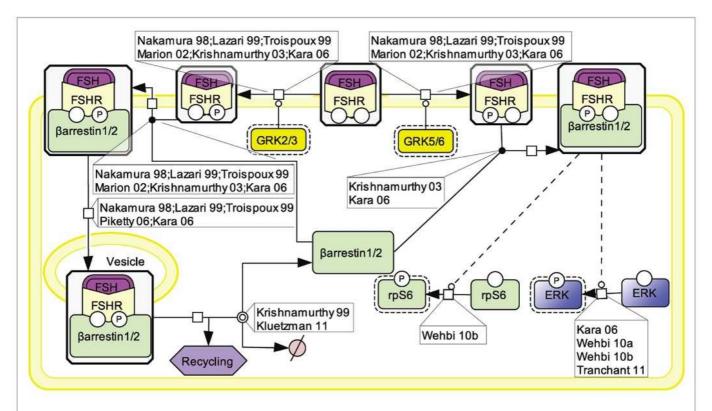


FIGURE 3 | FSH-induced β -arrestin-dependent signaling. Initially, the perception of β -arrestins' role was restricted to the control of FSHR's desensitization and recycling. This view has progressively evolved toward a more general role of β -arrestins as adapters and transducers leading to the

activation of MAPK ERK and rpS6 independently of G proteins upon FSH stimulation. GRK2/3 and GRK5/6 control the fate of the activated FSHR (i.e., desensitization vs. signaling) presumably through phosphorylation of distinct serines and threonines within the receptor's C-tail.

2002; Krishnamurthy et al., 2003a; Kara et al., 2006). In addition to GRK2-dependent phosphorylation, which plays a major role in FSHR phosphorylation and β-arrestin recruitment, GRK5 and 6 have also been found to contribute to the same processes in HEK293 cells, though to a lesser extent (Kara et al., 2006). Interestingly, β-arrestins recruited to GRK2- or GRK5/6-phosphorylated FSHR have been suggested to exert distinct intracellular functions (Kara et al., 2006; Reiter and Lefkowitz, 2006). It is well established that β-arrestin 1 and 2 binding to GRK-phosphorylated FSHR leads to the internalization and recycling of the receptor (Nakamura et al., 1998; Lazari et al., 1999; Kishi et al., 2002; Kara et al., 2006; Piketty et al., 2006). In HEK293 cells, GRK2phosphorylated FSHR has been reported to predominate in the β-arrestin-mediated internalization process (Kara et al., 2006). While most of the internalized FSHR is recycled back to the plasma membrane, a modest proportion of the receptor is routed to the lysosomal degradation pathway (Krishnamurthy et al., 2003b; Kluetzman et al., 2011).

A growing number of 7TMRs, including the FSHR, have been demonstrated to elicit signals independently of heterotrimeric G protein coupling, through direct interaction with β -arrestins (Kara et al., 2006; Reiter and Lefkowitz, 2006; Wehbi et al., 2010a,b; Tranchant et al., 2011). G protein-mediated ERK activation is rapid (reaching a maximum within \sim 2–5 min) and transient. In contrast, ERK activation via β -arrestins is slower in onset (reaching a maximum within \sim 5–10 min) but sustained (t1/2 > 1 h; Kara

et al., 2006). In addition, β -arrestins have been shown to contribute to rpS6 phosphorylation in HEK293 cells upon FSHR activation (Wehbi et al., 2010b). As also reported for other 7TMRs, GRK5 and 6-induced phosphorylation of the activated FSHR is required for β -arrestin-dependent signaling pathway in HEK293 cells (Kara et al., 2006; Reiter and Lefkowitz, 2006). Interestingly, partially deglycosylated equine LH has recently been shown to preferentially activate β -arrestin-dependent signaling at the FSHR, presumably acting as a biased ligand at this receptor (Wehbi et al., 2010b; Ulloa-Aguirre et al., 2011).

THE PI3K/mTOR PATHWAY

An array of evidence, gathered in various cell models, supports the notion that the PI3K/mTOR pathway plays an important role in FSH-induced actions including proliferation, regulation of gene transcription as well as of protein translation (**Figure 4**). FSH has been shown to lead to PI3K activation, PIP3 accumulation, and FSH-dependent proliferation in both granulosa cells and primary Sertoli cells from newborn rat (Park et al., 2005; Musnier et al., 2009; Dupont et al., 2010). Interestingly, in differentiating rat Sertoli cells, PIP3 accumulation is negatively regulated by PTEN whose expression is strongly and rapidly induced upon FSH stimulation, which results in a blockade of FSH-induced proliferation (Dupont et al., 2010).

This FSH-induced PI3K/PIP3 pathway has been reported by numerous authors to trigger Akt phosphorylation and activation

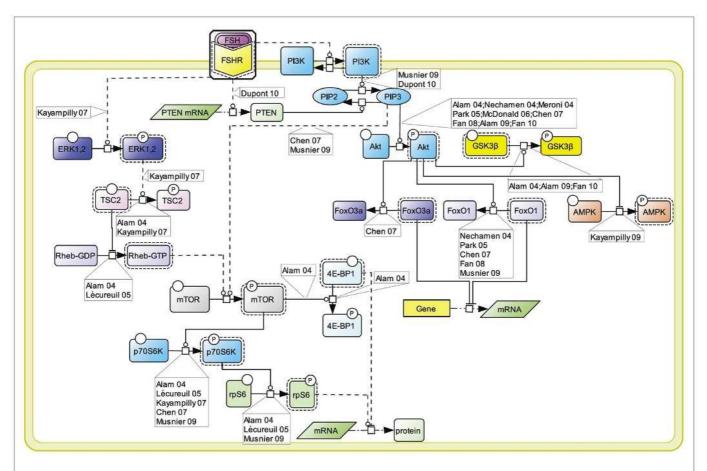


FIGURE 4 | PI3K/mTOR signaling at the FSHR. The PI3K/mTOR pathway plays an important role in FSH-induced actions, including proliferation, regulation of gene transcription as well as of protein translation.

(Gonzalez-Robayna et al., 2000; Alam et al., 2004, 2009; Meroni et al., 2004; Nechamen et al., 2004; Park et al., 2005; McDonald et al., 2006; Chen et al., 2007; Fan et al., 2008, 2010). In turn Akt leads to GSK3β phosphorylation and deactivation (Alam et al., 2009; Fan et al., 2010). It also triggers AMPK phosphorylation and its deactivation (Kayampilly and Menon, 2009). In addition, Akt has been reported to inactivate FoxO3a (Chen et al., 2007) and FoxO1 (Cunningham et al., 2003; Nechamen et al., 2004; Park et al., 2005; Chen et al., 2007; Fan et al., 2008; Musnier et al., 2009), two transcription factors which differentially impact on gene regulation.

The PI3K/PIP3 pathway has also been shown to induce mTOR phosphorylation and activation (Chen et al., 2007; Musnier et al., 2009). Another mechanism contributing to the full activation of mTOR upon FSH stimulation involves ERK activation which leads to TSC2 phosphorylation, thereby releasing its negative regulation on Rheb. Rheb-GTP in turn induces mTOR activity (Alam et al., 2004; Lécureuil et al., 2005; Kayampilly and Menon, 2007).

Once activated, mTOR controls p70S6K activity (Alam et al., 2004; Lécureuil et al., 2005; Chen et al., 2007; Kayampilly and Menon, 2007; Musnier et al., 2009). Active mTOR also leads to the phosphorylation and inactivation of 4E-BP1, a factor negatively controlling protein translation (Alam et al., 2004). In parallel,

FSH-induced p70S6K activity triggers a robust phosphorylation of rpS6 (Alam et al., 2004; Lécureuil et al., 2005; Musnier et al., 2009). The combination of rpS6 activation and 4E-BP1 inhibition support the notion that FSH controls and activates protein translation in addition to its well-known effects on gene expression.

FSHR INTERACTING PROTEINS

A number of proteins has been reported to interact with the FSHR and to impact on FSH-induced signaling pathways (Dias et al., 2005). One of the alternative transduction mechanism used by the FSHR involves the APPL1 (Figure 5). APPL1 has been shown to interact with the first and second intracellular loop of the FSHR, and has been proposed to mediate FSH-dependent PI3K and Akt signaling (Nechamen et al., 2004). Furthermore, recent studies in HEK293 cells and KGN granulosa cells have shown that residues K376, L377, and F382 in the first intracellular loop of the FSHR are implicated in the interaction of the receptor with APPL1 and that K376 particularly, links the activated FSHR to the inositol-phosphate pathway and FSH-stimulated intracellular calcium mobilization (Thomas et al., 2011). Other studies have reported calcium accumulation upon FSH stimulation in different cell models (Flores et al., 1990; Jayes et al., 2000; Lin et al., 2006; Lai et al., 2008). Further studies will be necessary to determine to

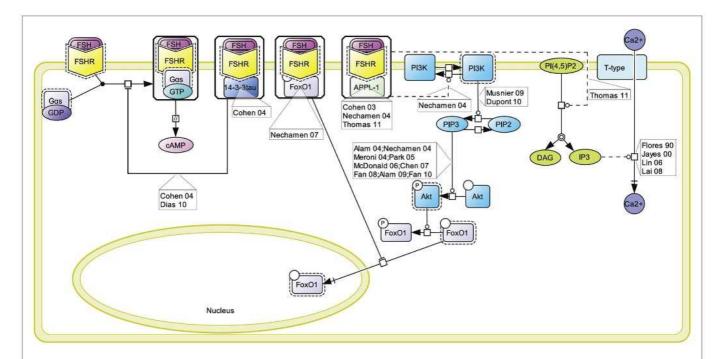


FIGURE 5 | Proteins interacting at the FSHR and their role in signaling. A number of proteins have been reported to interact with the FSHR and to impact on FSH-induced signaling pathways. These include 14-3-3τ, FoxO1, and APPL1, the latter two being involved in the control of the PI3K/Akt pathway downstream of the FSHR.

what extent APPL1/FSHR interaction is a general requirement for FSH-induced calcium signaling.

A direct interaction between the transcription factor FoxO1a and the FSHR has been evidenced (Nechamen et al., 2007). It has been suggested that, while interacting with the FSHR, FoxO1a cannot translocate to the nucleus and affect gene transcription (Dias et al., 2010). Interestingly, APPL1 also potentially contributes to the nuclear exclusion of FoxO1a though a distinct mechanism. Indeed, APPL1-mediated PI3K activation leads to Akt-dependent phosphorylation of FoxO1a, impairing its nuclear localization. It has been proposed that FoxO1a phosphorylation could occur within FSHR-containing signalosome since APPL1 is well-known to interact with Akt (Dias et al., 2010).

The FSHR interacts with the scaffolding protein $14-3-3\tau$ (Cohen et al., 2004). This interaction has been mapped in the second intracellular loop, in a region encompassing the ERW motif, where it is predicted to compete with G α s coupling. Consistently, when overexpressed, $14-3-3\tau$ has been reported to dampen FSH-induced cAMP accumulation (Cohen et al., 2004; Dias et al., 2010).

EGFR TRANSACTIVATION

Similar to other GPCRs, many lines of evidence point to EGFR transactivation as an important transduction mechanism used by FSHR (Cottom et al., 2003; McDonald et al., 2006; **Figure 6**). Two mechanisms have been shown to contribute to EGFR activation upon FSH stimulation. First, the metalloprotease ADAM17 has been shown to be phosphorylated in response to FSH stimulation (Yamashita et al., 2007, 2009, 2010). Once activated, ADAM17 splits a pro-EGF-like protein, releasing an EGF-like ligand which

subsequently binds to and activates EGFR (Yamashita et al., 2007). The mechanism by which FSHR induces ADAM17 phosphorylation is not yet identified. However, data from other receptors suggest that either p38 (Xu and Derynck, 2010) or Src (Zhang et al., 2006) could be involved in this process. A second mechanism of EGFR transactivation by the FSHR involves Src which has been proposed to be activated upon FSH exposure and to directly phosphorylate EGFR thereby eliciting EGF-independent activation of this receptor (Wayne et al., 2007).

When EGFR activation is pharmacologically inhibited, a significant decrease in phosphorylated ERK is observed (Cottom et al., 2003; Andric and Ascoli, 2006; McDonald et al., 2006; Wayne et al., 2007; Shupe et al., 2011). EGFR transactivation activates Ras which in turn induces the Raf1/MEK/ERK MAP kinase module, ultimately resulting in ERK phosphorylation (Cottom et al., 2003; Andric and Ascoli, 2006; McDonald et al., 2006; Ongeri et al., 2007; Wayne et al., 2007; Shupe et al., 2011). In granulosa cells, it has been shown that, upon sustained FSH exposure, a self-activation loop involving the MAP kinase module as well as phospholipase A2 and PKC, leads to CDK4 stimulation, retinoblastoma (RB1) inactivation, and ultimately DNA synthesis (Yang and Roy, 2006). EGFR transactivation has also been reported to trigger PI3K (Wayne et al., 2007).

NUCLEAR EVENTS CONTROLLED BY FSH

For simplification purpose, we decided to aggregate all the knowledge accumulated in the literature about FSH-induced nuclear events in a separate map (**Figure 7**). Gene transcription has long been known to be affected by FSH, as the hormone is known to control for instance genes implicated in steroidogenesis

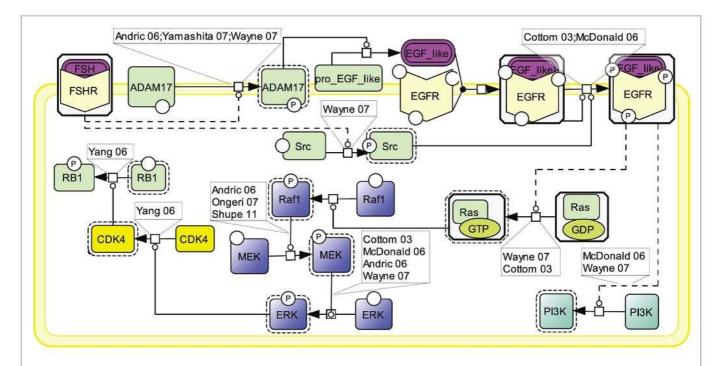


FIGURE 6 | Transactivation of the EGFR by FSH. Many lines of evidence point to EGFR transactivation as an important transduction mechanism used by FSHR. Two mechanisms have been reported: first, FSH stimulates the ADAM17 metalloprotease that leads to the release of an EGF-like ligand,

which subsequently binds to and activates EGFR; second, FSH activates Src which directly phosphorylates EGFR, thereby eliciting EGF-independent activation of this receptor. MAPK ERK and PI3K pathways have been shown to be activated as a result of EGFR transactivation by FSHR.

(Escamilla-Hernandez et al., 2008a) or in cell cycle via Smad proteins (Wang et al., 2011).

Upon FSH activation, PKA catalytic subunit translocates to the nucleus, where it activates CREB through phosphorylation on S133, thereby controlling cAMP response element (CRE)containing genes (Cottom et al., 2003; Fan et al., 2010). It also promotes the recruitment of the activator protein 1 (AP1) transcription factor to their cognate promoter regions (Yang et al., 2008). Nuclear PKA also leads to histone H3 phosphorylation and acetylation (Salvador et al., 2001). Histone H3 phosphorylation is known to favor cell division (Hans and Dimitrov, 2001) and ChIP experiments have revealed FSH-induced interaction between phosphorylated, acetylated histone H3 and the promoter region of c-Fos, a well-known signal transducer of cell proliferation and differentiation (Salvador et al., 2001). Interestingly, it has recently been reported that cytosolic PKA is able to bind the retinoic acid receptor alpha (RARA) and by doing so, to inhibit RARA translocation to the nucleus, hence the effects this nuclear receptor exerts on gene transcription (Santos and Kim, 2010).

Akt controls the activity of several transcription factors through PKA-independent mechanisms. Akt has been reported to suppress the inhibition of some transcription factors and to inhibit the activity of some others. For example, upon FSH exposure, Akt phosphorylates FoxO1a which is then excluded from the nucleus and is not able to inhibit MDM2-dependent HIF-1α activation anymore (Nechamen et al., 2004; Park et al., 2005; Chen et al., 2007; Fan et al., 2008; Alam et al., 2009). Akt also phosphorylates FoxO3a provoking its exclusion from the nucleus, blocking the pro-apoptotic effects of this transcription factor (Chen et al.,

2007). In addition, Akt has been reported to promote GSK3 β phosphorylation and deactivation which indirectly favors LEF-dependent transcription (Gonzalez-Robayna et al., 2000; Fan et al., 2010). The nuclear translocation of NF κ B is also promoted by Akt (Wang et al., 2002).

Once phosphorylated, MAP kinases ERK and p38 affect various transcriptional regulations in the nucleus (Cameron et al., 1996), participating to the regulation of AP1 and CREB activities.

FUTURE DIRECTIONS

Reconstructing the very complex intracellular networks triggered by FSH from the literature represents a huge challenge, partly due to the tremendous heterogeneity of experimental approaches and models, and partly because of the subjectivity and "bias" which cannot be avoided when manually assembling such network topologies. In the future, automated methods will probably allow building complex networks from the available data in a completely unbiased manner. In parallel, mass-spectrometry-based highthroughput methods now allow to grasp the phosphoproteome with an unprecedented power and in an unbiased fashion. Data generated using such experimental approaches will soon become available for FSH-responsive cells and will probably deliver an exhaustive coverage of the FSH-mediated signaling mechanisms induced in different cellular context. It will also be of importance to continue to bridge the gap between intracellular signaling pathways and integrated cellular responses such as gene regulation, metabolism, differentiation, proliferation, or apoptosis. Once the architecture of the FSH-induced signaling network will be deciphered, dynamical modeling will allow numerical simulations to

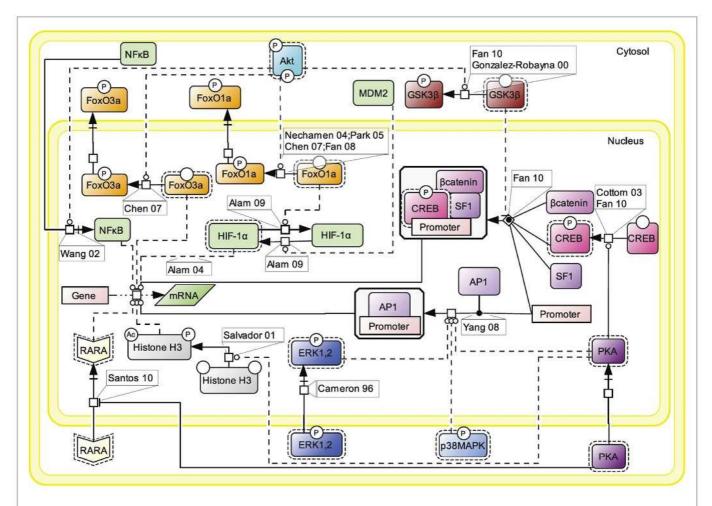


FIGURE 7 | Nuclear events controlled by FSH. Gene transcription has long been known to be affected by FSH. Multiple signaling pathways that are activated upon stimulation (i.e., PKA, p38, ERK, and Akt) subsequently trigger the activation or suppression of various transcription factors' activities at the nuclear level.

be carried out and functional prediction to be made. In that context, the generation of high quality dynamical data of the key FSH-induced signaling events will likely become a prerequisite in order to calibrate and validate the dynamical models.

Together, these technological breakthroughs will likely help reconstructing the global FSH-induced signaling mechanisms, predicting the way it processes information and delivers adapted physiological outcomes. It is expected that all these computational resources will help rationalize the development of pathway-selective pharmacological approaches at the FSHR. In addition, several genetic studies have reported a number of mutations and polymorphisms in the FSHR. Until now, the functional consequences of these genetic alterations were evaluated solely by measuring their effects on the classical Gs/cAMP/PKA pathway.

REFERENCES

Agrawal, G., and Dighe, R. R. (2009). Critical involvement of the hinge region of the follicle-stimulating hormone receptor in the activation of the receptor. J. Biol. Chem. 284, 2636–2647. Aittomaki, K., Lucena, J. L., Pakarinen, P., Sistonen, P., Tapanainen, J., Gromoll, J., Kaskikari, R., Sankila, E. M., Lehvaslaiho, H., Engel, A. R., Nieschlag, E., Huhtaniemi, I., and De La Chapelle, A. (1995). Mutation in the follicle-stimulating hormone However, as we have recently shown for the A189V mutation of the FSHR, multiplexed assessment of the mutants' and/or variants' functionality can allow uncovering subtle perturbations/imbalances within the signaling networks (Tranchant et al., 2011). Therefore, with the availability of computational and experimental tools, it will become possible to achieve better assessment of the real functional impact of genetic alterations encountered, not only in the FSH or FSHR but also anywhere within the activated networks.

ACKNOWLEDGMENTS

This work was supported by the Action d'Envergure (AE) INRIA/INRA Regate, and the ASAM project (INRA/INRIA). Pauline Gloaguen is recipient of a thesis fellowship from Région Centre.

receptor gene causes hereditary hypergonadotropic ovarian failure. *Cell* 82, 959–968.

Alam, H., Maizels, E. T., Park, Y., Ghaey, S., Feiger, Z. J., Chandel, N. S., and Hunzicker-Dunn, M. (2004). Follicle-stimulating hormone activation of hypoxia-inducible factor-1 by the phosphatidylinositol 3-kinase/AKT/Ras homolog enriched in brain (Rheb)/ mammalian target of rapamycin (mTOR) pathway is necessary for induction of select protein

- markers of follicular differentiation. J. Biol. Chem. 279, 19431–19440.
- Alam, H., Weck, J., Maizels, E., Park, Y., Lee, E. J., Ashcroft, M., and Hunzicker-Dunn, M. (2009). Role of the phosphatidylinositol-3kinase and extracellular regulated kinase pathways in the induction of hypoxia-inducible factor (HIF)-1 activity and the HIF-1 target vascular endothelial growth factor in ovarian granulosa cells in response to follicle-stimulating hormone. Endocrinology 150, 915–928.
- Andric, N., and Ascoli, M. (2006).
 A delayed gonadotropin-dependent and growth factor-mediated activation of the extracellular signal-regulated kinase 1/2 cascade negatively regulates aromatase expression in granulosa cells. Mol. Endocrinol. 20, 3308–3320.
- Arey, B. J., Stevis, P. E., Deecher, D. C., Shen, E. S., Frail, D. E., Negro-Vilar, A., and Lopez, F. J. (1997). Induction of promiscuous G protein coupling of the follicle-stimulating hormone (FSH) receptor: a novel mechanism for transducing pleiotropic actions of FSH isoforms. Mol. Endocrinol. 11, 517–526.
- Arey, B. J., Yanofsky, S. D., Claudia Perez, M., Holmes, C. P., Wrobel, J., Gopalsamy, A., Stevis, P. E., Lopez, F. J., and Winneker, R. C. (2008). Differing pharmacological activities of thiazolidinone analogs at the FSH receptor. Biochem. Biophys. Res. Commun. 368, 723–728.
- Cameron, M. R., Foster, J. S., Bukovsky, A., and Wimalasena, J. (1996). Activation of mitogen-activated protein kinases by gonadotropins and cyclic adenosine 5'-monophosphates in porcine granulosa cells. *Biol. Reprod.* 55, 111–119.
- Chen, Y.-J., Hsiao, P.-W., Lee, M.-T., Mason, J. I., Ke, F.-C., and Hwang, J.-J. (2007). Interplay of PI3K and cAMP/PKA signaling, and rapamycin-hypersensitivity in TGFbetal enhancement of FSH-stimulated steroidogenesis in rat ovarian granulosa cells. J. Endocrinol. 192, 405–419.
- Choi, J.-H., Chen, C.-L., Poon, S. L., Wang, H.-S., and Leung, P. C. K. (2009). Gonadotropin-stimulated epidermal growth factor receptor expression in human ovarian surface epithelial cells: involvement of cyclic AMP-dependent exchange protein activated by cAMP pathway. Endocr. Relat. Cancer 16, 179–188.
- Choi, J. H., Wong, A. S., Huang, H. F., and Leung, P. C. (2007). Gonadotropins and ovarian cancer. *Endocr. Rev.* 28, 440–461.

- Cohen, B. D., Nechamen, C. A., and Dias, J. A. (2004). Human follitropin receptor (FSHR) interacts with the adapter protein 14-3-3tau. Mol. Cell. Endocrinol. 220, 1–7.
- Cottom, J., Salvador, L. M., Maizels, E. T., Reierstad, S., Park, Y., Carr, D. W., Davare, M. A., Hell, J. W., Palmer, S. S., Dent, P., Kawakatsu, H., Ogata, M., and Hunzicker-Dunn, M. (2003). Follicle-stimulating hormone activates extracellular signalregulated kinase but not extracellular signal-regulated kinase kinase through a 100-kDa phosphotyrosine phosphatase. J. Biol. Chem. 278, 7167–7179.
- Crépieux, P., Marion, S., Martinat, N., Fafeur, V., Vern, Y. L., Kerboeuf, D., Guillou, F., and Reiter, E. (2001). The ERK-dependent signalling is stagespecifically modulated by FSH, during primary Sertoli cell maturation. Oncogene 20, 4696–4709.
- Cunningham, M. A., Zhu, Q., Unterman, T. G., and Hammond, J. M. (2003). Follicle-stimulating hormone promotes nuclear exclusion of the forkhead transcription factor FoxO1a via phosphatidylinositol 3-kinase in porcine granulosa cells. *Endocrinology* 144, 5585–5594.
- Dattatreyamurty, B., Figgs, L. W., and Reichert, L. E. (1987). Physical and functional association of follitropin receptors with cholera toxin-sensitive guanine nucleotidebinding protein. J. Biol. Chem. 262, 11737–11745.
- Dias, J. A., Bonnet, B., Weaver, B. A., Watts, J., Kluetzman, K., Thomas, R. M., Poli, S., Mutel, V., and Campo, B. (2011). A negative allosteric modulator demonstrates biased antagonism of the follicle stimulating hormone receptor. *Mol. Cell. Endocrinol*. 333, 143–150.
- Dias, J. A., Mahale, S. D., Nechamen, C. A., Davydenko, O., Thomas, R. M., and Ulloa-Aguirre, A. (2010). Emerging roles for the FSH receptor adapter protein APPL1 and overlap of a putative 14-3-3tau interaction domain with a canonical Gprotein interaction site. Mol. Cell. Endocrinol. 329, 17–25.
- Dias, J. A., Nechamen, C. A., and Atari, R. (2005). Identifying protein interactors in gonadotropin action. *Endocrine* 26, 241–247.
- Dierich, A., Sairam, M. R., Monaco, L., Fimia, G. M., Gansmuller, A., Lemeur, M., and Sassone-Corsi, P. (1998). Impairing folliclestimulating hormone (FSH) signaling in vivo: targeted disruption of the FSH receptor leads to aberrant

- gametogenesis and hormonal imbalance. *Proc. Natl. Acad. Sci. U.S.A.* 95, 13612–13617.
- Dupont, J., Musnier, A., Decourtye, J., Boulo, T., Lécureuil, C., Guillou, H., Valet, S., Fouchécourt, S., Pitetti, J.-L., Nef, S., Reiter, E., and Crépieux, P. (2010). FSH-stimulated PTEN activity accounts for the lack of FSH mitogenic effect in prepubertal rat Sertoli cells. Mol. Cell. Endocrinol. 315, 271–276.
- Escamilla-Hernandez, R., Little-Ihrig, L., Orwig, K. E., Yue, J., Chandran, U., and Zeleznik, A. J. (2008a). Constitutively active protein kinase A qualitatively mimics the effects of follicle-stimulating hormone on granulosa cell differentiation. Mol. Endocrinol. 22, 1842–1852.
- Escamilla-Hernandez, R., Little-Ihrig, L., and Zeleznik, A. J. (2008b). Inhibition of rat granulosa cell differentiation by overexpression of Galphaq. *Endocrine* 33, 21–31.
- Fakunding, J. L., Tindall, D. J., Dedman, J. R., Mena, C. R., and Means, A. R. (1976). Biochemical actions of follicle-stimulating hormone in the Sertoli cell of the rat testis. *Endocrinology* 98, 392–402.
- Fan, H.-Y., Liu, Z., Cahill, N., and Richards, J. S. (2008). Targeted disruption of Pten in ovarian granulosa cells enhances ovulation and extends the life span of luteal cells. *Mol. Endocrinol.* 22, 2128–2140.
- Fan, H.-Y., O'Connor, A., Shitanaka, M., Shimada, M., Liu, Z., and Richards, J. S. (2010). Beta-catenin (CTNNB1) promotes preovulatory follicular development but represses LH-mediated ovulation and luteinization. Mol. Endocrinol. 24, 1529–1542.
- Flores, J. A., Veldhuis, J. D., and Leong, D. A. (1990). Follicle-stimulating hormone evokes an increase in intracellular free calcium ion concentrations in single ovarian (granulosa) cells. *Endocrinology* 127, 3172–3179.
- Gonzalez-Robayna, I. J., Falender, A. E., Ochsner, S., Firestone, G. L., and Richards, J. S. (2000). Follicle-Stimulating hormone (FSH) stimulates phosphorylation and activation of protein kinase B (PKB/Akt) and serum and glucocorticoid-Induced kinase (Sgk): evidence for A kinase-independent signaling by FSH in granulosa cells. Mol. Endocrinol. 14, 1283–1300.
- Guan, R., Wu, X., Feng, X., Zhang, M., Hebert, T. E., and Segaloff, D. L. (2010). Structural determinants underlying constitutive dimerization of unoccupied human

- follitropin receptors. *Cell. Signal.* 22, 247–256.
- Hans, F., and Dimitrov, S. (2001). Histone H3 phosphorylation and cell division. Oncogene 20, 3021–3027.
- Huhtaniemi, I., Ahtiainen, P., Pakarainen, T., Rulli, S. B., Zhang, F. P., and Poutanen, M. (2006). Genetically modified mouse models in studies of luteinising hormone action. *Mol. Cell. Endocrinol.* 252, 126–135.
- Jayes, F. C., Day, R. N., Garmey, J. C., Urban, R. J., Zhang, G., and Veldhuis, J. D. (2000). Calcium ions positively modulate follicle-stimulating hormone- and exogenous cyclic 3',5'-adenosine monophosphate-driven transcription of the P450(scc) gene in porcine granulosa cells. *Endocrinology* 141, 2377–2384.
- Kara, E., Crépieux, P., Gauthier, C., Martinat, N., Piketty, V., Guillou, F., and Reiter, E. (2006). A phosphorylation cluster of five serine and threonine residues in the C-terminus of the follicle-stimulating hormone receptor is important for desensitization but not for beta-arrestin-mediated ERK activation. Mol. Endocrinol. 20, 3014–3026.
- Kayampilly, P. P., and Menon, K. M. J. (2007). Follicle-stimulating hormone increases tuberin phosphorylation and mammalian target of rapamycin signaling through an extracellular signal-regulated kinase-dependent pathway in rat granulosa cells. *Endocrinology* 148, 3950–3957.
- Kayampilly, P. P., and Menon, K. M. J. (2009). Follicle-stimulating hormone inhibits adenosine 5'-monophosphate-activated protein kinase activation and promotes cell proliferation of primary granulosa cells in culture through an Akt-dependent pathway. Endocrinology 150, 929–935.
- Kenakin, T. (2005). New concepts in drug discovery: collateral efficacy and permissive antagonism. Nat. Rev. Drug Discov. 4, 919–927.
- Kishi, H., Krishnamurthy, H., Galet, C., Bhaskaran, R. S., and Ascoli, M. (2002). Identification of a short linear sequence present in the Cterminal tail of the rat follitropin receptor that modulates arrestin-3 binding in a phosphorylationindependent fashion. J. Biol. Chem. 277, 21939–21946.
- Kitano, H., Funahashi, A., Matsuoka, Y., and Oda, K. (2005). Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.* 23, 961–966.

- Kluetzman, K. S., Thomas, R. M., Nechamen, C. A., and Dias, J. A. (2011). Decreased degradation of internalized follicle-stimulating hormone caused by mutation of aspartic acid 6.30550 in a protein kinase-CK2 consensus sequence in the third intracellular loop of human follicle-stimulating hormone receptor. *Biol. Reprod.* 84, 1154–1163.
- Krishnamurthy, H., Galet, C., and Ascoli, M. (2003a). The association of arrestin-3 with the follitropin receptor depends on receptor activation and phosphorylation. Mol. Cell. Endocrinol. 204, 127–140.
- Krishnamurthy, H., Kishi, H., Shi, M., Galet, C., Bhaskaran, R. S., Hirakawa, T., and Ascoli, M. (2003b). Postendocytotic trafficking of the folliclestimulating hormone (FSH)-FSH receptor complex. Mol. Endocrinol. 17, 2162–2176.
- Kumar, T. R., Wang, Y., Lu, N., and Matzuk, M. M. (1997). Follicle stimulating hormone is required for ovarian follicle maturation but not male fertility. *Nat. Genet.* 15, 201–204.
- Lai, T.-H., Lin, Y.-F., Wu, F.-C., and Tsai, Y.-H. (2008). Folliclestimulating hormone-induced Galphah/phospholipase C-delta1 signaling mediating a noncapacitative Ca2+ influx through T-type Ca2+ channels in rat sertoli cells. Endocrinology 149, 1031–1037.
- Landmark, B. F., Fauske, B., Eskild, W., Skalhegg, B., Lohmann, S. M., Hansson, V., Jahnsen, T., and Beebe, S. J. (1991). Identification, characterization, and hormonal regulation of 3', 5'-cyclic adenosine monophosphate-dependent protein kinases in rat Sertoli cells. Endocrinology 129, 2345–2354.
- Layman, L. C., Lee, E. J., Peak, D. B., Namnoum, A. B., Vu, K. V., Van Lingen, B. L., Gray, M. R., Mcdonough, P. G., Reindollar, R. H., and Jameson, J. L. (1997). Delayed puberty and hypogonadism caused by mutations in the follicle-stimulating hormone beta-subunit gene. N. Engl. J. Med. 337, 607–611.
- Lazari, M. F., Liu, X., Nakamura, K., Benovic, J. L., and Ascoli, M. (1999). Role of G protein-coupled receptor kinases on the agonist-induced phosphorylation and internalization of the follitropin receptor. *Mol. Endocrinol.* 13, 866–878.
- Lécureuil, C., Tesseraud, S., Kara, E., Martinat, N., Sow, A., Fontaine, I., Gauthier, C., Reiter, E., Guillou, F., and Crépieux, P. (2005). Folliclestimulating hormone activates p70

- ribosomal protein S6 kinase by protein kinase A-mediated dephosphorylation of Thr 421/Ser 424 in primary Sertoli cells. *Mol. Endocrinol.* 19, 1812–1820.
- Lefkowitz, R. J., and Shenoy, S. K. (2005). Transduction of receptor signals by beta-arrestins. *Science* 308, 512–517.
- Lin, Y. F., Tseng, M. J., Hsu, H. L., Wu, Y. W., Lee, Y. H., and Tsai, Y. H. (2006). A novel follicle-stimulating hormone-induced G alpha h/phospholipase C-delta1 signaling pathway mediating rat sertoli cell Ca2+-influx. Mol. Endocrinol. 20, 2514–2527.
- Loutradis, D., Drakakis, P., Milingos, S., Stefanidis, K., and Michalas, S. (2003). Alternative approaches in the management of poor response in controlled ovarian hyperstimulation (COH). *Ann. N. Y. Acad. Sci.* 997, 112–119.
- Loutradis, D., Elsheikh, A., Kallianidis, K., Drakakis, P., Stefanidis, K., Milingos, S., and Michalas, S. (2004). Results of controlled ovarian stimulation for ART in poor responders according to the short protocol using different gonadotrophins combinations. Arch. Gynecol. Obstet. 270, 223–226.
- Lunenfeld, B. (2004). Historical perspectives in gonadotrophin therapy. Hum. Reprod. Update 10, 453–467.
- Macklon, N. S., Stouffer, R. L., Giudice, L. C., and Fauser, B. C. (2006). The science behind 25 years of ovarian stimulation for in vitro fertilization. *Endocr. Rev.* 27, 170–207.
- Maizels, E. T., Cottom, J., Jones, J. C., and Hunzicker-Dunn, M. (1998). Follicle stimulating hormone (FSH) activates the p38 mitogen-activated protein kinase pathway, inducing small heat shock protein phosphorylation and cell rounding in immature rat ovarian granulosa cells. Endocrinology 139, 3353–3356.
- Marion, S., Kara, E., Crepieux, P., Piketty, V., Martinat, N., Guillou, F., and Reiter, E. (2006). G protein-coupled receptor kinase 2 and beta-arrestins are recruited to FSH receptor in stimulated rat primary Sertoli cells. J. Endocrinol. 190, 341–350.
- Marion, S., Robert, F., Crepieux, P., Martinat, N., Troispoux, C., Guillou, F., and Reiter, E. (2002). G protein-coupled receptor kinases and beta arrestins are relocalized and attenuate cyclic 3',5'-adenosine monophosphate response to folliclestimulating hormone in rat primary Sertoli cells. Biol. Reprod. 66, 70–76.

- Matthews, C. H., Borgato, S., Beck-Peccoz, P., Adams, M., Tone, Y., Gambino, G., Casagrande, S., Tedeschini, G., Benedetti, A., and Chatterjee, V. K. (1993). Primary amenorrhoea and infertility due to a mutation in the beta-subunit of folliclestimulating hormone. *Nat. Genet.* 5, 83–86.
- McDonald, C. A., Millena, A. C., Reddy, S., Finlay, S., Vizcarra, J., Khan, S. A., and Davis, J. S. (2006). Folliclestimulating hormone-induced aromatase in immature rat Sertoli cells requires an active phosphatidylinositol 3-kinase pathway and is inhibited via the mitogen-activated protein kinase signaling pathway. Mol. Endocrinol. 20, 608–618.
- Means, A. R., Macdougall, E., Soderling, T. R., and Corbin, J. D. (1974). Testicular adenosine 3':5'-monophosphate-dependent protein kinase. Regulation by folliclestimulating hormone. J. Biol. Chem. 249, 1231–1238.
- Meroni, S. B., Riera, M. F., Pellizzari, E. H., and Cigorraga, S. B. (2002). Regulation of rat Sertoli cell function by FSH: possible role of phosphatidylinositol 3-kinase/protein kinase B pathway. J. Endocrinol. 174, 195–204.
- Meroni, S. B., Riera, M. F., Pellizzari, E. H., Galardo, M. N., and Cigorraga, S. B. (2004). FSH activates phosphatidylinositol 3kinase/protein kinase B signaling pathway in 20-day-old Sertoli cells independently of IGF-I. J. Endocrinol. 180, 257–265.
- Musnier, A., Heitzler, D., Boulo, T., Tesseraud, S., Durand, G., Lécureuil, C., Guillou, H., Poupon, A., Reiter, E., and Crépieux, P. (2009). Developmental regulation of p70 S6 kinase by a G protein-coupled receptor dynamically modelized in primary cells. Cell. Mol. Life Sci. 66, 3487–3503.
- Nakamura, K., Krupnick, J. G., Benovic, J. L., and Ascoli, M. (1998). Signaling and phosphorylation-impaired mutants of the rat follitropin receptor reveal an activation- and phosphorylation-independent but arrestin-dependent pathway for internalization. J. Biol. Chem. 273, 24346–24354.
- Naz, R. K., Gupta, S. K., Gupta, J. C., Vyas, H. K., and Talwar, G. P. (2005). Recent advances in contraceptive vaccine development: a mini-review. *Hum. Reprod.* 20, 3271–3283.
- Nechamen, C. A., Thomas, R. M., Cohen, B. D., Acevedo, G., Poulikakos, P. I., Testa, J. R., and Dias, J. A. (2004). Human

- follicle-stimulating hormone (FSH) receptor interacts with the adaptor protein APPL1 in HEK 293 cells: potential involvement of the PI3K pathway in FSH signaling. *Biol. Reprod.* 71, 629–636.
- Nechamen, C. A., Thomas, R. M., and Dias, J. A. (2007). APPL1, APPL2, Akt2 and FOXO1a interact with FSHR in a potential signaling complex. Mol. Cell. Endocrinol. 260–262, 93–99.
- Northup, J. K., Sternweis, P. C., Smigel, M. D., Schleifer, L. S., Ross, E. M., and Gilman, A. G. (1980). Purification of the regulatory component of adenylate cyclase. *Proc. Natl. Acad.* Sci. U.S.A. 77, 6516–6520.
- Ongeri, E. M., Verderame, M. F., and Hammond, J. M. (2007). The TATA binding protein associated factor 4b (TAF4b) mediates FSH stimulation of the IGFBP-3 promoter in cultured porcine ovarian granulosa cells. *Mol. Cell. Endocrinol.* 278, 29–35.
- Papkoff, H., and Ekblad, M. (1970). Ovine follicle stimulating hormone: preparation and characterization of its subunits. *Biochem. Biophys. Res. Commun.* 40, 614–621.
- Park, Y., Maizels, E. T., Feiger, Z. J.,
 Alam, H., Peters, C. A., Woodruff,
 T. K., Unterman, T. G., Lee, E. J.,
 Jameson, J. L., and Hunzicker-Dunn,
 M. (2005). Induction of cyclin
 D2 in rat granulosa cells requires
 FSH-dependent relief from FOXO1
 repression coupled with positive signals from Smad. J. Biol. Chem. 280,
 9135–9148.
- Piketty, V., Kara, E., Guillou, F., Reiter, E., and Crepieux, P. (2006). Follicle-stimulating hormone (FSH) activates extracellular signal-regulated kinase phosphorylation independently of beta-arrestinand dynamin-mediated FSH receptor internalization. Reprod. Biol. Endocrinol. 4, 33–44.
- Quintana, J., Hipkin, R. W., Sanchez-Yague, J., and Ascoli, M. (1994). Follitropin (FSH) and a phorbol ester stimulate the phosphorylation of the FSH receptor in intact cells. *J. Biol. Chem.* 269, 8772–8779.
- Radu, A., Pichon, C., Camparo, P., Antoine, M., Allory, Y., Couvelard, A., Fromont, G., Hai, M. T., and Ghinea, N. (2010). Expression of follicle-stimulating hormone receptor in tumor blood vessels. N. Engl. J. Med. 363, 1621–1630.
- Rajagopal, S., Rajagopal, K., and Lefkowitz, R. J. (2010). Teaching old receptors new tricks: biasing seventransmembrane receptors. *Nat. Rev. Drug Discov.* 9, 373–386.

- Rangel-Aldao, R., and Rosen, O. M. (1976). Dissociation and reassociation of the phosphorylated and nonphosphorylated forms of adenosine 3':5'-monophosphate-dependent protein kinase from bovine cardiac muscle. *J. Biol. Chem.* 251, 3375–3380.
- Reiter, E., and Lefkowitz, R. J. (2006). GRKs and beta-arrestins: roles in receptor silencing, trafficking and signaling. Trends Endocrinol. Metab. 17, 159–165.
- Reiter, E., Marion, S., Robert, F., Troispoux, C., Boulay, F., Guillou, F., and Crepieux, P. (2001). Kinase-inactive G-protein-coupled receptor kinases are able to attenuate follicle-stimulating hormone-induced signaling. Biochem. Biophys. Res. Commun. 282, 71–78.
- Salvador, L. M., Park, Y., Cottom, J., Maizels, E. T., Jones, J. C., Schillace, R. V., Carr, D. W., Cheung, P., Allis, C. D., Jameson, J. L., and Hunzicker-Dunn, M. (2001). Follicle-stimulating hormone stimulates protein kinase A-mediated histone H3 phosphory-lation and acetylation leading to select gene activation in ovarian granulosa cells. J. Biol. Chem. 276, 40146–40155.
- Santos, N. C., and Kim, K. H. (2010). Activity of retinoic acid receptor-alpha is directly regulated at its protein kinase A sites in response to follicle-stimulating hormone signaling. *Endocrinology* 151, 2361–2372.
- Shimada, M., Hernandez-Gonzalez, I., Gonzalez-Robayna, I., and Richards, J. S. (2006). Paracrine and autocrine regulation of epidermal growth factor-like factors in cumulus oocyte complexes and granulosa cells: key roles for prostaglandin synthase 2 and progesterone receptor. Mol. Endocrinol. 20, 1352–1365.
- Shupe, J., Cheng, J., Puri, P., Kostereva, N., and Walker, W. H. (2011). Regulation of sertoli-germ cell adhesion and sperm release by FSH and nonclassical testosterone signaling. Mol. Endocrinol. 25, 238–252.
- Simoni, M., Gromoll, J., and Nieschlag, E. (1997). The follicle-stimulating hormone receptor: biochemistry, molecular biology, physiology, and pathophysiology. *Endocr. Rev.* 18, 739–773.
- Themmen, A. P. N., and Huhtaniemi, I. T. (2000). Mutations of gonadotropins and gonadotropin

- receptors: elucidating the physiology and pathophysiology of pituitary-gonadal function. *Endocr. Rev.* 21, 551–583.
- Thomas, R. M., Nechamen, C. A., Mazurkiewicz, J. E., Muda, M., Palmer, S., and Dias, J. A. (2007). Follice-stimulating hormone receptor forms oligomers and shows evidence of carboxyl-terminal proteolytic processing. *Endocrinology* 148, 1987–1995.
- Thomas, R. M., Nechamen, C. A., Mazurkiewicz, J. E., Ulloa-Aguirre, A., and Dias, J. A. (2011). The Adapter Protein APPL1 Links FSH Receptor to inositol 1,4,5trisphosphate production and is implicated in intracellular Ca2+ mobilization. Endocrinology 152, 1691–1701.
- Tranchant, T., Durand, G., Gauthier, C., Crépieux, P., Ulloa-Aguirre, A., Royère, D., and Reiter, E. (2011). Preferential β-arrestin signalling at low receptor density revealed by functional characterization of the human FSH receptor A189 V mutation. *Mol. Cell. Endocrinol.* 331, 109–118.
- Troispoux, C., Guillou, F., Elalouf, J. M., Firsov, D., Iacovelli, L., Blasi, A. D., Combarnous, Y., and Reiter, E. (1999). Involvement of G protein-coupled receptor kinases and arrestins in desensitization to follicle-stimulating hormone action. *Mol. Endocrinol.* 13, 1599–1614.
- Ulloa-Aguirre, A., Crépieux, P., Poupon, A., Maurel, M.-C., and Reiter, E. (2011). Novel pathways in gonadotropin receptor signaling and biased agonism. *Rev. Endocr. Metab. Disord.* doi: 10.1007/s11154-011-9176-2. [Epub ahead of print].
- Vloeberghs, V., Peeraer, K., Pexsters, A., and D'Hooghe, T. (2009). Ovarian hyperstimulation syndrome and complications of ART. Best Pract. Res. Clin. Obstet. Gynaecol. 23, 691–709.
- Wang, W., Chen, X., Li, X., Wang, L., Zhang, H., He, Y., Wang, J., Zhao, Y., Zhang, B., and Xu, Y. (2011). Interference RNA (RNAi)-based silencing of endogenous Smad4 in porcine granulosa cells resulted in decreased FSH-mediated GCs proliferation and steroidogenesis. Reproduction 141, 643–651.
- Wang, Y., Chan, S., and Tsang, B. K. (2002). Involvement of inhibitory nuclear factor-kappaB (NFkappaB)-independent NFkappaB activation in the gonadotropic regulation of X-linked inhibitor of apoptosis expression during ovarian follicular

- development in vitro. Endocrinology 143, 2732–2740.
- Wayne, C. M., Fan, H.-Y., Cheng, X., and Richards, J. S. (2007). Folliclestimulating hormone induces multiple signaling cascades: evidence that activation of Rous sarcoma oncogene, RAS, and the epidermal growth factor receptor are critical for granulosa cell differentiation. Mol. Endocrinol. 21, 1940–1957.
- Wehbi, V., Decourtye, J., Piketty, V., Durand, G., Reiter, E., and Maurel, M.-C. (2010a). Selective modulation of follicle-stimulating hormone signaling pathways with enhancing equine chorionic gonadotropin/antibody immune complexes. Endocrinology 151, 2788–2799.
- Wehbi, V., Tranchant, T., Durand, G., Musnier, A., Decourtye, J., Piketty, V., Butnev, V. Y., Bousfield, G. R., Crépieux, P., Maurel, M.-C., and Reiter, E. (2010b). Partially deglycosylated equine LH preferentially activates betaarrestin-dependent signaling at the follicle-stimulating hormone receptor. Mol. Endocrinol. 24, 561–573.
- Whalen, E. J., Rajagopal, S., and Lefkowitz, R. J. (2011). Therapeutic potential of beta-arrestin- and G protein-biased agonists. *Trends. Mol. Med.* 17, 126–139.
- Xiao, K., Mcclatchy, D. B., Shukla, A. K., Zhao, Y., Chen, M., Shenoy, S. K., Yates, J. R. III, and Lefkowitz, R. J. (2007). Functional specialization of beta-arrestin interactions revealed by proteomic analysis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12011–12016.
- Xiao, K., Sun, J., Kim, J., Rajagopal, S., Zhai, B., Villen, J., Haas, W., Kovacs, J. J., Shukla, A. K., Hara, M. R., Hernandez, M., Lachmann, A., Zhao, S., Lin, Y., Cheng, Y., Mizuno, K., Ma'ayan, A., Gygi, S. P., and Lefkowitz, R. J. (2010). Global phosphorylation analysis of beta-arrestin-mediated signaling downstream of a seven transmembrane receptor (7TMR). Proc. Natl. Acad. Sci. U.S.A. 107, 15299–15304.
- Xu, P., and Derynck, R. (2010). Direct activation of TACE-mediated ectodomain shedding by p38 MAP kinase regulates EGF receptordependent cell proliferation. Mol. Cell 37, 551–566.
- Yamashita, Y., Hishinuma, M., and Shimada, M. (2009). Activation of PKA, p38 MAPK and ERK1/2 by

- gonadotropins in cumulus cells is critical for induction of EGF-like factor and TACE/ADAM17 gene expression during in vitro maturation of porcine COCs. J. Ovarian Res. 2, 20.
- Yamashita, Y., Kawashima, I., Gunji, Y., Hishinuma, M., and Shimada, M. (2010). Progesterone is essential for maintenance of Tace/Adam17 mRNA expression, but not EGFlike factor, in cumulus cells, which enhances the EGF receptor signaling pathway during in vitro maturation of porcine COCs. J. Reprod. Dev. 56, 315–323.
- Yamashita, Y., Kawashima, I., Yanai, Y., Nishibori, M., Richards, J. S., and Shimada, M. (2007). Hormone-induced expression of tumor necrosis factor alphaconverting enzyme/A disintegrin and metalloprotease-17 impacts porcine cumulus cell oocyte complex expansion and meiotic maturation via ligand activation of the epidermal growth factor receptor. Endocrinology 148, 6164-6175.
- Yang, P., and Roy, S. K. (2006). A novel mechanism of FSH regulation of DNA synthesis in the granulosa cells of hamster preantral follicles: involvement of a protein kinase C-mediated MAP kinase 3/1 selfactivation loop. *Biol. Reprod.* 75, 149–157.
- Yang, W., Lu, C.-L., Yu, F.-Q., Liu, T., Hu, Z.-Y., and Liu, Y.-X. (2008). Mitogen-activated protein kinase regulates FSH-induced expression of tissue-type plasminogen activator through an activator protein 1 response element. *Endocrine* 34, 101–107.
- Yu, E.-Q., Han, C.-S., Yang, W., Jin, X., Hu, Z.-Y., and Liu, Y.-X. (2005). Activation of the p38 MAPK pathway by follicle-stimulating hormone regulates steroidogenesis in granulosa cells differentially. J. Endocrinol. 186, 85–96.
- Zeleznik, A. J., Saxena, D., and Little-Ihrig, L. (2003). Protein kinase B is obligatory for follicle-stimulating hormone-induced granulosa cell differentiation. *Endocrinology* 144, 3985–3994.
- Zhang, Q., Thomas, S. M., Lui, V. W. Y., Xi, S., Siegfried, J. M., Fan, H., Smithgall, T. E., Mills, G. B., and Grandis, J. R. (2006). Phosphorylation of TNFalpha converting enzyme by gastrinreleasing peptide induces amphiregulin release and EGF receptor activation. Proc. Natl. Acad. Sci. U.S.A. 103, 6901–6906.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 July 2011; accepted: 14 September 2011; published online: 05 October 2011.

Citation: Gloaguen P, Crépieux P, Heitzler D, Poupon A and Reiter E (2011) Mapping the follidestimulating hormone-induced signaling networks. Front. Endocrin. 2:45. doi: 10.3389/fendo.2011.00045
This article was submitted to Frontiers in Cellular Endocrinology, a specialty of Frontiers in Endocrinology.

Copyright © 2011 Gloaguen, Crépieux, Heitzler, Poupon and Reiter. This is an open-access article subject to a nonexclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with. Ce travail bibliographique nous a permis de répertorier les voies de signalisation induites par la FSH les plus étudiées telles que les différentes voies G, la voie PI3K, la voie β-arrestine ou encore la transactivation du récepteur EGFR. Les régulations de la FSH sur la prolifération cellulaire, la traduction des protéines et la transcription des gènes sont possibles grâce l'induction de ces différentes voies. De plus, ce travail montre qu'il existe des molécules carrefour dans le réseau (exemple : ERK MAPK), qui appartiennent à plusieurs voies différentes, et de ce fait jouent un rôle très important.

Un certain nombre de points avaient été soulevés au cours de cette étude, mais n'ont pas été abordés dans l'article. En particulier, certaines voies ont été écartées pour des raisons qui sont explicitées dans la suite. D'autre part, si dans la revue nous avons exploité des données venant aussi bien des cellules de Sertoli que des cellules de la Granulosa ou même de lignées cellulaires telles que les cellules HEK293, certaines parties du réseau peuvent être spécifiques de l'un ou l'autre. De même, des différences ont pu être observées dans les mêmes cellules mais à différents stades : prolifération et différenciation.

II- D'autres voies de signalisation

La FSH induit différentes voies de signalisation interconnectées, dont les principales sont présentées dans la revue. En plus de ces mécanismes bien étudiés et relativement consensuels, d'autres travaux ont montré que la FSH pouvait aussi induire d'autres voies que celles décrites dans la publication, en particulier une voie anti-apoptotique, la stéroïdogenèse ou encore une voie anti-oxydante.

Voie anti-apoptotique: Tunjung et al [194] ont émis l'hypothèse que la FSH induit une voie anti-apoptotique via la production d'acide hyaluronique dans les cellules de la Granulosa. Ces travaux montrent que lorsque le récepteur de l'acide hyaluronique CD44 est bloqué par un anticorps, la quantité de procaspases diminue. L'hypothèse déduite est que la complexation de l'acide hyaluronique avec son récepteur empêche le clivage des procaspases 3, 8 et 9 en caspases (Figure 44). Les caspases jouent un rôle important dans l'apoptose. L'hypothèse de cette voie est appuyée par des travaux plus anciens. Tout d'abord, les travaux de Němcová et al [195] montrent que la FSH induit la production d'acide hyaluronique en synergie avec IGF1 à travers les voies PI3K/Akt et MAP Kinase ERK dans les cellules de cumulus (Figure

44). De plus, une augmentation des quantités de caspases 8 et 9 a été observée dans les cellules testiculaires, lors de la suppression de la FSH [196]. Ces auteurs ont aussi vérifié l'action de la FSH sur la transcription des gènes Casp9, Bcl2l2, Bax, Casp8 et Fas, qui sont fortement impliqués dans le processus d'apoptose. Aucune différence significative n'a été observée au niveau de la transcription, ceci implique que l'action de la FSH sur ces molécules se situe bien au niveau protéique.

Dans les cellules de la Granulosa, la FSH, en synergie avec IGF1, active aussi une voie antiapoptotique par le biais de la PKA [197].

Les travaux de Gonzalez et al [198] montrent, quant à eux, que la FSH inhibe la transcription du gène PAR4, qui code pour une protéine pro-apoptotique et que la suppression de ce gène induit l'activation d'une voie anti-apoptotique via PKCζ (Figure 44). Ces travaux ont également montré qu'il y a moins de cellules en apoptose en présence de FSH.

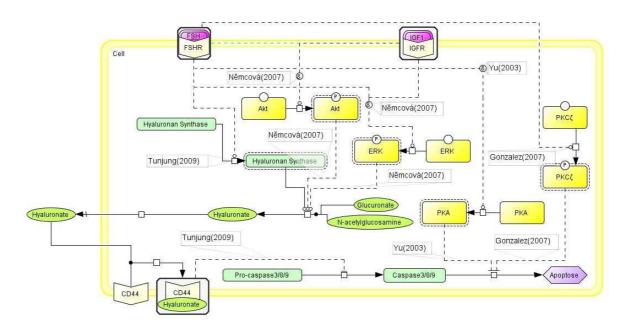


Figure 44: La FSH induit des voies anti-apoptotiques conjointement avec IGF1

Cette voie n'a pas été reportée dans la revue car les preuves expérimentales sur lesquelles elle repose sont encore fragiles. Par exemple, l'inhibition du clivage des procaspases en caspases par l'acide hyaluronique (sous l'effet de la FSH) n'a réellement été montrée qu'une seule fois [194]. D'autres études vont dans le sens d'une voie anti-apoptotique induite par la FSH mais le mécanisme n'est pas encore totalement élucidé. De plus, l'inhibition de l'apoptose a le plus souvent été décrite comme un effet de la synergie entre la FSH et l'IGF, il est donc suspecté

que l'effet anti-apoptotique soit le résultat d'une stimulation combinée de la cellule par la FSH et l'IGF plutôt que de la stimulation par la FSH seule.

L'apoptose étant la mort cellulaire programmée, elle permet de supprimer par exemple, les cellules germinales porteuses d'anomalies génétiques qui sont incapables d'assurer une reproduction normale. Or l'apoptose peut également survenir au niveau de cellules « normales ». Par exemple, le nombre de follicules est défini mais un grand nombre d'entre eux sont dégradés par apoptose au cours de la vie reproductive d'une femelle. La FSH fait partie des facteurs permettant de limiter la perte folliculaire grâce à l'activation d'une voie anti-apoptotique, provoquant ainsi la survie de ces cellules cibles et des cellules environnantes. Il a été démontré que la FSH provoque la survie des follicules primordiaux (*In Vitro*) grâce à la réduction de l'apoptose [199]. Ces auteurs mettent en avant que ce phénomène est indirect puisque seules les cellules de la Granulosa expriment le récepteur de la FSH. Grâce à cette régulation, la FSH participe au maintien du *pool* de gamètes.

<u>Stéroïdogenèse</u>: Le rôle de la FSH dans la stéroïdogenèse est bien connu. Les hormones stéroïdes sont dérivées du cholestérol et sont divisées en trois grands groupes :

- Glucocorticoïdes,
- Minéralocorticoïdes.
- Stéroïdes sexuels.

La FSH induit la synthèse des stéroïdes sexuels tels que les androgènes (testostérone), les œstrogènes (estradiol) et les progestatifs (progestérone) (Figure 45) [200, 201]. Ces stéroïdes sexuels jouent un rôle important dans le développement des caractères sexuels primaires et secondaires. De plus, elles exercent un rétrocontrôle sur la sécrétion des gonadotrophines par l'hypophyse.

De nombreux gènes dont la transcription est induite par la FSH, codent pour les enzymes de synthèse de ces hormones (cf Annexe 5).

La voie PI3K décrit une action de la FSH sur l'AMPK (Figure 4 de la revue). En effet, la FSH inhibe la phosphorylation de l'AMPK sur la position T172 et induit une phosphorylation sur le site S485/491, ce qui la rend inactive [202]. Tosca et al [203] montrent que si on active l'AMPK en présence de FSH, l'expression des enzymes impliquées dans la stéroïdogenèse diminue, ce qui suggère une levée d'inhibition de l'expression de ces protéines par la FSH.

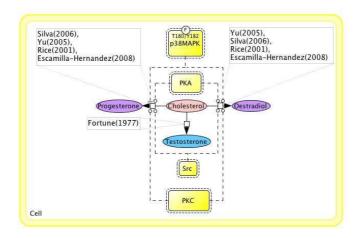


Figure 45 : Stéroïdogenèse induite par la FSH

Cette production est activée par différentes voies, par exemple la progestérone est produite via la p38MAPK [204], la PKA [200], Src [205] et la PKC [206]. La production d'œstradiol est induite par les mêmes voies, excepté la p38MAPK qui dans ce cas exerce une inhibition. Il est possible de déduire que ces différentes activations/inhibitions de la synthèse des stéroïdes sont induites par la régulation de l'expression des gènes codant pour les enzymes de synthèse, par exemple, la PKA est un intermédiaire pour l'activation de la transcription des enzymes 3β-HSD, p450scc et p450arom (les autres régulations sont disponibles en Annexe 5).

Voie anti-oxydante:

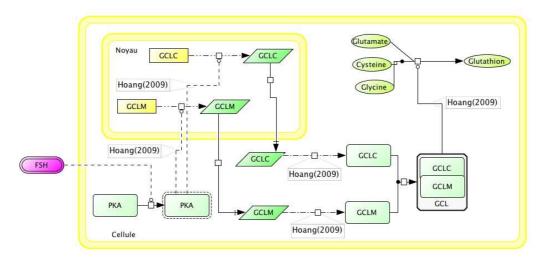


Figure 46: Production de glutathion induite par la FSH

Une autre voie a été mise en évidence, elle est impliquée dans la production de glutathion [207], molécule connue pour son effet antioxydant. Cette voie n'est pas représentée dans

l'article car elle est induite à travers la transcription des gènes codant les sous-unités du complexe enzymatique permettant cette production (Figure 46) et dans la revue, nous avons choisi de présenter uniquement l'aspect signalisation en amont des régulations transcriptionnelles provoquées par la FSH or l'unique intermédiaire connu est la PKA.

Une voie anti-oxydante est un système de défense contre l'accumulation de radicaux libres issus du métabolisme de l'oxygène ou de l'azote, qui seraient responsables du vieillissement précoce des cellules. Le vieillissement conduit à une diminution de la qualité des gamètes. La FSH induit donc un système de protection indirect des gamètes.

<u>Voie PI3K</u>: La revue présente une phosphorylation de mTOR induite par PIP3, ce qui conduit à son activation (Figure 4 de la revue). En plus des relations déjà indiquées, PIP3 participe à l'activation d'Akt qui phosphoryle ensuite TSC2 pour l'inhiber [208]. Ceci conduit à une levée d'inhibition sur la voie Rheb/mTOR (Figure 47).

La voie PI3K, par son action sur mTOR, participe à l'activation de la traduction de certaines protéines.

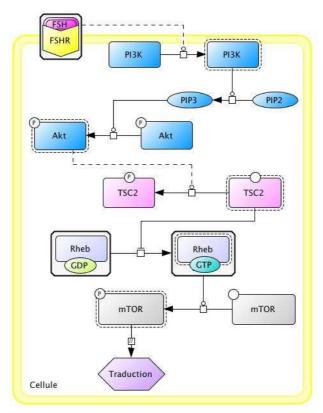


Figure 47: Voie PI3K/Akt

<u>Autres voies</u>: La FSH est impliquée dans d'autres processus cellulaires, mais son action se fait à travers une régulation de la transcription et/ou de la traduction comme pour la synthèse des hormones stéroïdes et du glutathion. Ceci dépend de la signalisation mais ne relève pas de des mécanismes de signalisation *stricto sensus*. Les régulations sur quelques gènes ont tout de même été répertoriées, les tableaux sont disponibles dans l'Annexe 5.

Interconnections de différentes voies de signalisation: Comme cela a été montré dans la revue, la FSH est capable de réguler des protéines clés grâce à différentes voies de signalisation. Par exemple, la cascade des ERK MAPK peut être induite via les voies Gs, Gi, β-arrestines ou encore la transactivation du récepteur EGFR. De même, les régulations sur TSC2 (inhibiteur de la voie mTOR) impliquent les voies ERK MAPK et PI3K.

Ainsi, les différentes voies de signalisation ne sont pas indépendantes les unes des autres puisqu'elles contiennent toutes des « molécules carrefour » qui les relient entre elles.

III- Types cellulaires

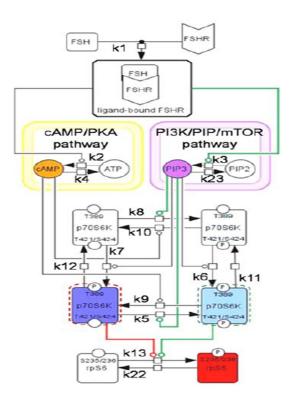


Figure 48 : Signalisation induite par la FSH, conduisant à la traduction des protéines dans les cellules de Sertoli.

En vert, les relations propres aux cellules en prolifération et en rouge les relations propres aux cellules en différenciation. Figure issue de [209].

Nous avons choisi de représenter un réseau de signalisation induit par la FSH en ne tenant pas compte des différents types cellulaires et des différents stades de développement mais comme il l'a été dit dans l'article, il serait plus rigoureux à l'avenir de les prendre en compte car il peut exister des différences de signalisation. Par exemple, Musnier et al [209] ont démontré des différences de mécanismes de signalisation conduisant à la traduction des protéines sous l'action de la FSH dans les cellules Sertoli en prolifération et en différenciation (Figure 48). Cette représentation montre que mTOR induit la phosphorylation du site T421/S424 de la p70S6Kinase uniquement dans les cellules de Sertoli en prolifération alors que cette phosphorylation n'existe pas dans les cellules en différenciation. De plus, la p70S6Kinase active est doublement phosphorylée dans les cellules en prolifération alors qu'il n'y a qu'une phosphorylation sur le site T389 pour les cellules en différenciation.

Il existe également des différences entre les types cellulaires. Dans les cellules de la Granulosa, la revue de Musnier et al [127] ne décrit pas d'implication de la PKA dans la déphosphorylation du site T421/S424 de la p70S6Kinase contrairement à ce qui est décrit pour les cellules de Sertoli. En réalité, cette déphosphorylation n'a jamais été étudiée/publiée dans les cellules de la Granulosa dans le cadre d'une stimulation FSH.

Les différents types cellulaires ne présenteraient pas tous les mêmes réseaux de signalisation en réponse à la stimulation par la FSH. Etant donné que les effets biologiques observés entre les différents modèles cellulaires ne sont pas identiques, ceci soulève plusieurs questions :

- La topologie du réseau peut-elle changer d'une cellule à l'autre, et d'un état physiologique à l'autre ?
- Les différents modèles cellulaires présentent-ils la même dynamique ?

Toutes les molécules ont chacune, une fonction bien définie. Par exemple, quel que soit le type cellulaire, la protéine Gs induit toujours l'activation de l'adénylate cyclase qui produit toujours l'AMPc qui a toujours les mêmes effecteurs. Ces différences de signalisation entre les types cellulaires ne devraient donc pas être dues à des différences topologiques du réseau.

Ceci nous conduit à la deuxième question, les différentes cellules présentent-elles une dynamique différente ? C'est-à-dire, est-ce que les quantités moléculaires sont différentes d'un modèle cellulaire à un autre et si oui, comment pourraient-elles varier et comment cela peut-il entraîner des différences de signalisation ?

Pour confirmer une différence de quantités moléculaires d'un type cellulaire à un autre, des expériences de quantification précise sont nécessaires. Cette possibilité paraît plus

vraisemblable qu'une différence de topologie car si une molécule est moins présente dans un cas par rapport à un autre, la signalisation qui en dépend varie. Pour illustrer cette idée, prenons l'exemple hypothétique de l'AMPc. La Figure 49 illustre le mode d'action de l'AMPc sur ses effecteurs, PKA (à gauche) et EPAC (à droite). La PKA est en réalité un complexe composé de deux sous-unités régulatrices et de deux sous-unités catalytiques. L'activation de la PKA dépend de la disponibilité en AMPc car, il faut quatre molécules d'AMPc pour permettre la libération des sous-unités catalytiques, ainsi capables d'agir sur leurs effecteurs. En ce qui concerne EPAC, son activation dépend d'une seule molécule d'AMPc. Comme la protéine EPAC nécessite moins d'AMPc que la PKA pour être activée, il semble possible de déduire qu'une cellule contenant une faible quantité d'AMPc, montre une signalisation dépendante d'EPAC plutôt qu'une signalisation dépendante de PKA.

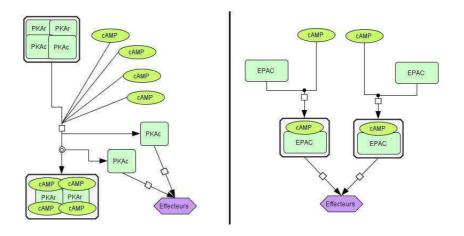


Figure 49: Exemples d'actions de l'AMPc

Cette hypothèse pourrait expliquer certaines différences de signalisation d'un type cellulaire à un autre. Par exemple, sur la Figure 48, il est possible de voir que la p70S6K phosphorylée sur le site T421/S424 et sur la position T389 est la forme capable d'induire l'activation de rpS6 dans les cellules de Sertoli en prolifération alors que dans les cellules de Sertoli en différenciation, c'est la forme phosphorylée sur la position T389 qui active rpS6. Cette différence pourrait s'expliquer par des variations de quantités : il serait possible d'émettre l'hypothèse que la quantité de PKA active (qui déphosphoryle le site T421/S424) soit plus importante dans les cellules de Sertoli en prolifération que dans les cellules de Sertoli en différenciation, ce qui impliquerait que la quantité de p70S6Kinase phosphorylée sur le site T421/S424 soit beaucoup plus faible voire indétectable dans les cellules en différenciation.

Cette explication est hypothétique et constitue uniquement une aide à la description du raisonnement.

Un autre facteur pourrait également expliquer ces différences : la répartition des molécules dans la cellule. La Figure 50A illustre le cas où un signal induit l'activation d'une enzyme. Si son substrat se situe dans un autre compartiment de la cellule, l'enzyme ne pourra pas catalyser sa transformation. La Figure 50B montre l'activation de la même enzyme sous l'action du même signal. La réaction enzymatique est possible dans ce cas car le substrat se situe à proximité de l'enzyme.

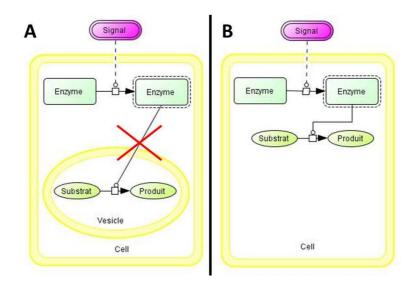


Figure 50 : Différences de localisation cellulaire des molécules

Il existe une dernière hypothèse pour expliquer les différences de réseaux de signalisation entre les types cellulaires. Comme nous l'avons évoqué dans la description de la voie apoptotique, la FSH est capable d'agir sur la cellule en synergie avec le facteur IGF1. Ce dernier peut interférer avec la signalisation FSH grâce à une stimulation autocrine. De plus, la FSH induit une levée d'inhibition de la fixation d'IGF1 avec son récepteur dans les cellules de Sertoli [210, 211]. Il est donc possible que différents facteurs soient présents dans l'environnement cellulaire et interfèrent donc avec le réseau de signalisation induit par la FSH. Ces interférences pourraient conduire à des différences observées entre les différents réseaux car les récepteurs des facteurs ne sont pas obligatoirement exprimés dans tous les types cellulaires.

Une nouvelle méthode d'inférence de réseaux de signalisation

En nous basant sur la logique du premier ordre, nous avons créé une méthode originale d'inférence de réseaux de signalisation directement à partir du raisonnement de l'expert biologiste et des données expérimentales. Cette méthode permet la traduction des conclusions en un réseau décrit dans le langage utilisé par le logiciel CellDesigner.

Comme cela a été indiqué précédemment, l'hormone FSH est au cœur du projet de recherche de l'équipe BIOS, il était donc naturel de prendre le réseau de signalisation induit par cette hormone comme modèle de mise au point de notre nouvelle méthode d'inférence. Les règles ont été créées à partir des expériences rencontrées dans la bibliographie concernant la FSH. Ensuite, nous avons choisi de prendre un autre modèle de réseau de signalisation, différent de celui induit par la FSH, afin d'évaluer la généricité des règles créées. Nous voulions vérifier que notre méthode ne se limiterait pas à un seul type de réseau de signalisation et que les règles existantes suffisaient pour construire un autre réseau. Pour cette validation, le réseau induit par le facteur EGF semblait être un bon modèle car c'est l'un des plus étudiés, et l'ensemble des données permettant de le construire est facilement accessible. Nous avons utilisé le réseau induit par l'EGF présenté dans la banque de données *Pathway Interaction Database* (http://pid.nci.nih.gov/) car les références bibliographiques utilisées par les curateurs de cette banque sont fournies, ce qui nous a permis d'utiliser la même littérature pour la création des faits initiaux permettant l'inférence du réseau EGF.

I- Présentation de la méthode

Un article décrivant notre méthode a été accepté pour le Workshop de ECML (*European Conference on Machine Learning*) 'Learning and Discovery in Symbolic Systems Biology 2012' (https://www.cs.bris.ac.uk/~oray/LDSSB12/).

<u>Titre</u>: Towards a logic-based method to infer provenance-aware molecular networks.

Zahira Aslaoui-Errafi, Sarah Cohen-Boulakia, Christine Froidevaux, **Pauline Gloaguen**,

Anne Poupon, Adrien Rougny, Meriem Yahiaoui

Towards a logic-based method to infer provenance-aware molecular networks

Zahira Aslaoui-Errafi^{1,2}, Sarah Cohen-Boulakia^{1,2}, Christine Froidevaux^{1,2*}, Pauline Gloaguen³, Anne Poupon³, Adrien Rougny^{1,2}, Meriem Yahiaoui^{1,2}

¹ Laboratoire de Recherche en Informatique (LRI), CNRS UMR 8623 Université Paris Sud, F-91405 Orsay, Cedex, France
² AMIB group, INRIA Saclay

³ BIOS group, INRA, UMR85, Unité Physiologie de la Reproduction et des Comportements, F-37380 Nouzilly, France; CNRS, UMR7247, F-37380 Nouzilly, France; Université François Rabelais, F-37041 Tours, France

Abstract. Providing techniques to automatically infer molecular networks is particularly important to understand complex relationships between biological objects. We present a logic-based method to infer such networks and show how it allows inferring signalling networks from the design of a knowledge base. Provenance of inferred data has been carefully collected, allowing quality evaluation. More precisely, our method (i) takes into account various kinds of biological experiments and their origin; (ii) mimics the scientist's reasoning within a first-order logic setting; (iii) specifies precisely the kind of interaction between the molecules; (iv) provides the user with the provenance of each interaction; (v) automatically builds and draws the inferred network.

1 Context

Biological objects (proteins, genes, small molecules, etc.) interact with each other (physical interactions, activations, inhibitions, catalyses, etc.) and form biological networks. Studying such networks allows the discovery of emerging properties and the understanding of complex biological systems. With the rise of highthroughput experimental methods and the rapid development of bioinformatics analysis methods, there has been a dramatic increase in the quantity and heterogeneity of available data. Providing methods able to automatically construct molecular networks from experimental data is thus of paramount importance. While the series of DREAM workshops competitions have been attempting to address these issues since 2007 (http://www.the-dream-project.org/), they are still challenging. In the ASAM project, we aim to meet the challenging task of designing performant algorithms to infer the topology (or influence graph) of signalling networks (subtask of current DREAM7), a first step in the process of elaborating predictive dynamical models. Given the components of the network,

Corresponding author: chris@lri.fr. (Authors in alphabetical order). This work has been supported by the INRA-INRIA ASAM project.

building the influence graph consists in determining which relationships they share (e.g., enzymatic catalysis, (de-)phosphorylation, etc.).

Among the approaches available, some provide focused solutions: on connectivity reconstruction (finding which molecules interact with each others from a dataset, without specifying the type of relationship), or on causal reconstruction (finding effects that molecules may have on each others)[20]. Some approaches proceed by refining existing molecular networks [19], which requires the knowledge of an initial network to be analyzed. Others proceed by inferring the network, considering only a single type of experimental method at a time, mostly DNA micro-arrays [1] [4] [3], but also phosphoproteomics [21] [22] [18], or metabolomics [2] etc. Therefore they cannot infer complete pathways involving interactions of different types, failing to provide an holistic view necessary to the understanding of the complexity of living cells [17].

Approaches integrating a variety of different experimental data types have also been proposed [11][24], but they reconstruct only partial network topologies, and define relations without causality (the interactions are not marked as being activations or inhibitions).

Automatic inference of topological networks is performed using various computational approaches that range from statistics [4] to machine learning techniques (e.g. bayesian networks [1]) through logic-based techniques such as boolean networks [25] [12] or fuzzy logic networks [23]. Most of these techniques allow performing model-based prediction as well. In addition, logic-based approaches are particularly well-suited to analyze the consistency of a network towards experimental data by model-checking [9], [7]. Our approach to automatically build the inference graph underlying signalling networks leads to the design of a logic-based model (see Figure 1), in the same spirit as [5] but with a more expressive logical formalism which allows making explicit the expert reasoning. The rest of this paper is organized as follows. Section 2 introduces the Knowledge Base where expert's reasoning is made explicit and a wide variety of experimental data is stored, and shows how the network can be deduced. Section 3 is dedicated to the process of building and drawing the network automatically. Section 4 presents our validation approach while Section 5 draws conclusions and perspectives.

2 Building the Knowledge Base (KB)

Our network inference method is based on the application of reasoning rules to experimental facts. These general rules formalize how experts deduce from experimental facts the belonging of a given molecule to the network and the relationships it shares with other molecules of the network. To achieve the proofof-concept for this method, we have formalized the knowledge necessary for inferring the signalling network triggered by the FSH receptor [8]. This involved the creation of first-order rules for seven types of experiments.

The design of the KB classically consists of a set of predicates, rules, and facts (ground predicates) described here after.

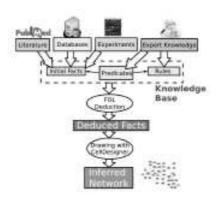


Fig. 1. Workflow of our logic-based method to infer molecular networks

2.1 Predicates

We distinguish three classes of predicates, depending on their biological meaning.

Basic predicates provide ontological types or indicate simple relationships between data (e.g., AntibodyAgainst(A,X) means that A is an antibody recognizing molecule X).

Experiment predicates describe a wide panorama of experiments on molecular data from simple to more complex experiments involving perturbators. An example of a simple experiment is PA(X,Y,A,E), meaning that "in a phosphorylation assay (PA) we observe more (resp. less, the same quantity) of the molecule revealed by antibody A, which is a phosphorylated form of molecule Y, in presence of X than in absence of X, where E can be either 'increase', 'decrease' or 'noeffect'". An example with a perturbator is ICPPA(X,Y,I,A,E) meaning that "in a phosphorylation assay (PA) in presence of X we observe more (resp. less, the same quantity) of the molecule revealed by antibody A, which is a phosphorylated form of molecule Y, in presence of inhibitor I than in absence of I, where E can be either 'increase', 'decrease' or 'noeffect's. Table I shows all the experiment predicates relative to phosphorylation assays (in the variables, the antibody variable A is replaced by the molecule variable Z it triggers). All the experiments have been formalized by predicates divided into three types depending on the kind of conclusion they allow one to obtain: modulation of a reaction, structure, or localization, and each predicate has been carefully designed and documented as depicted in Table 1.

Network Predicates formalize the different types of biological relationships that can exist between two or more biochemical species (modification, phosphorylation, translation, transcription, complexation, dissociation) taking into account the modulation of these relationships by another species (activation, inhibition or no effect). The modulation of the relationships may be associated with a level of confidence, expressed by the status variable (taking values among: 'invalidated', 'confirmed' or 'hypothesis'). Additionally, the precision of the mod-

Table 1. Phosphorylation assays predicates.

Pred name	Deduction	Method	Variables	Res type	Type (y)	Type (z)	Perturbator (i)
PA, PRA	Modulation of a	Measure	x,y,z,e x,y,i,z,e	Phosphoryl. of y into z	Protein	100100000	None Antagonist

ulation may vary, which is expressed through the distance variable (taking the values 'direct', 'indirect', or 'unknown'), meaning that a molecule has a direct effect on a relationship (for example, it catalyzes a reaction) or it has an effect via intermediaries, or that this information is currently not available.

2.2 Rules

We have designed two sets of rules. Simple rules use experimental data to draw conclusions on new relationships between biochemical species. Complex rules use the conclusions obtained by other rules and either provide new data relationships or refine existing relationships. They include transitive rules and rules that deduce an indirect distance (see above) from an unknown distance for a network fact. A simple rule is: IF PA(X,Y,A,E) AND AntibodyAgainst(A,B) AND PhosphoForm(B,Y,POS) THEN PHOSPHORYLATE(X,Y,B,unknown,confirmed,E) meaning that "If in a phosphorylation assay where A is an antibody that recognizes B, B is a phosphorylated form of Y at position POS, we observe more (resp. less, the same quantity) of B in presence of X than in absence of X, then X activates (resp. inhibits, does not have any effect on) the phosphorylation of Y in B, with an unknown distance, where E can be 'increase', 'decrease' or 'noeffect'".

2.3 Initial and Deduced Facts

Based on the study of the literature and data from public databases (e.g., SBEAMS, http://www.sbeams.org/), predicates have then been instantiated into facts. Facts have been carefully attached to their provenance, specifying the article(s) and/or the data base(s) from which they have been extracted. Various data quality features have been made as explicit as possible in particular the level of confidence to be associated to a set of experimental results.

Initial facts have then been used to produce deduced facts by triggering rules. As expert rules are mostly Horn rules, we have mainly considered using forward chaining procedures for rules-base systems [6], leading to a saturation of the KB. In order to tackle more expressive rules (e.g. typing rules) and facts involving functional terms that describe molecular complexes, we have also been considering first-order consequence-finding techniques. In these settings, provenance information of deduced facts is obtained from their proof(s) making it possible to assess their quality from the quality of initial facts used in the deduction process.

3 Automatic generation of the signalling network

The network topology is automatically built from our initial and deduced facts using the software CellDesigner [15], that is widely used to model molecular networks. CellDesigner uses (and extends) two standards, SBML (Systems Biology Markup Language) [10] and SBGN (Systems Biology Graphical Notation) [16], for the description and representation of molecular networks. CellDesigner can represent: cell compartments, biochemical species (nodes of the network) and reactions (arcs). Each biochemical species is localized in one or more compartment, and assigned to a biological type (e.g., complex, protein). Each reaction has a type (e.g., translation, transcription) and at least one reactant species and one product species, and can be modulated (e.g., activated/catalyzed) by one or more modifier species. The various species, reactions and modulation types are represented in the network by different forms of nodes and arcs (e.g. ovals for simple molecules; plain lines for state transitions).

Our facts (initial and deduced) are automatically translated into the extended version of SBML that CellDesigner uses. Initial facts are used to describe the chemical species of the networks (name, type and residue modifications, e.g. a phosphorylation on the residue S132). The deduced facts are used to describe the reactions (type, reactants, products and modulations) and species localizations. The status and the distance of the reaction modulations are modelled by different line-styles of arcs linking a modifier species to a reaction (e.g. black line for a confirmed status or red line for a hypothesis status) in the network. The layout of the network is then automatically calculated by CellDesigner. This automatic process is illustrated by a short example.

Example. Let phosphorylate(fsh,mek1,pmek1s17,indirect,confirmed,increase) be a deduced fact to be represented with CellDesigner. First, the fact has to be translated into (extended) SBML. It describes a relation between the molecules (first three arguments) fsh, mek1 and pmek1s17 (the latter being phosphorylated-mek1 at position S17). To declare precisely these molecules into CellDesigner, the KB is queried to get their type (protein(fsh), protein(mek1) and phosphoprotein (pmek1s17)), their localization in the cell (localization(fsh,extracellular), localization(mek1,cytosol) and localization(pmek1s17, cytosol)) and their possible residue modifications (phosphoform(pmek1s17, mek1, s17)). This is given by initial facts of the KB that are retrieved by exploiting the predicates typology. Second, the deduced fact is translated into an SBML reaction. Here, the type of reaction is a state transition (phosphorylation), the reactant is the second argument (mek1). the product the third (pmek1s17) and the reaction is activated (increase) by the first argument (fsh). Because the modulation of the reaction has a status confirmed and is indirect, the line-style of the arc representing the modulation is declared as being black and dashed.

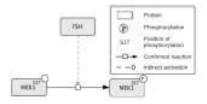


Fig. 2. Building elements of network with CellDesigner

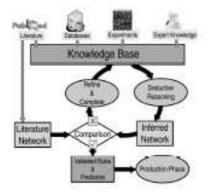


Fig. 3. Workflow for validation of our method to infer molecular networks

4 Validation

A phase of trial and error may be necessary in order to validate the method, before using it in production phase. This validation phase will require an important accumulation of data. Indeed, we need to infer automatically the networks from experimental data and compare them to networks of the literature to seek for missing links (links absent while expected), indicating missing reasoning expert rules which will need to be written. This upstream work has already been done for the follicle-stimulating hormone receptor (FSH-R) [8]. Even with this quite simple network, over 400 initial facts and 100 expert rules were necessary. Once processed, the KB contained over 1,000 (initial and deduced) facts. Two different serotonin receptors (5HT2 and 5HT4) and Epidermal Growth Factor receptor (EGF-R) are currently considered. We envisage to speed up the detection of biological objects in the literature that will be used within facts by using text mining tools such as PathText [14]. We expect that after studying a few networks, expert rules will be complete, validating the KB. During this validation phase and the following production phase, the inferred networks will be analyzed and all the unexpected conclusions will be validated experimentally.

5 Conclusion and Discussion

We have introduced a bottom-up method to infer molecular networks whose originality lies in its knowledge-driven basis. More precisely: (i) domain knowledge is collected from experts, literature and public databases, and data from classical and high-throughput experiments; (ii) the scientist's reasoning is mimicked within an expressive logical setting (first-order logic); (iii) the kind of interaction between the molecules is precisely specified; (iv) provenance of deduced facts (translated into interactions) is provided, allowing users to evaluate the quality of the inferred pieces of knowledge; (v) the inferred network is automatically built and drawn, offering the possibility of interacting with it.

Such an automated method, allowing automated exploitation of experimental data and resulting in molecular networks, will become increasingly necessary with the spread of huge amounts of data produced by large-scale experimental methods.

Ongoing work include considering techniques to formalize efficiently expressive rules and facts (including considering encoding into a propositional setting). Importantly, while all our current efforts have been made on deductive reasoning, we need also to consider exploring abductive reasoning to help scientists understanding how the biological system can be perturbed to reach a desired behaviour. Users may wish to know which facts or rules are to be added to reinforce the reliability of an existing arc or to add a new arc in the network. Determining the rules and pieces of data to be added to the initial KB to get the desired behaviour is obtained by the abduction reasoning that will provide all the possible hypotheses expressed in terms of initial facts and logical rules. Advices will thus be given to the scientists about the new experiments to be conducted to validate some hypothesis. As a consequence, we are currently looking at systems able to provide both deductive and abductive reasoning and are considering the SOLAR system as a first excellent candidate [13].

References

- M. Bansal, V. Belcastro, A. Ambesi-Impiombato and D. di Bernardo. How to infergene networks from expression profiles, Mol Syst Biol., 3:78, 2007.
- T. Cakir, M. Hendriks, J.A. Westerhuis and A.K. Smilde, Metabolic network discovery through reverse engineering of metabolome data., Metabolomics, 5:318-29, 2009.
- J. Chiquet, A. Smith, G. Grasseau, C. Matias and C. Ambroise, SIMoNe: Statistical Inference for MOdular NEtworks., Bioinformatics, 25:417-8, 2009.
- I. Drozdov, B. Svejda, B.I. Gustafsson, S. Mane, R. Pfragner, M. Kidd and I.M. Modlini, Gene Network Inference and Biochemical Assessment Delineates GPCR Pathways and CREB Targets in Small Intestinal Neuroendocrine Neoplasia., PLoS ONE, 6(8), 2011.
- F. Eduati, A. Corradin, B. Di Camillo and G. Toffolo, A Boolean approach to linear prediction for signaling network modeling, PLoS ONE, 5(9), 2010.
- C. Forgy, Rete: A Fast Algorithm for the Many Pattern-Many Object Pattern Match Problem, Artificial Intelligence, 19:17-37, 1982.

- M. Gebser, T. Schaub, S. Thiele and Ph. Veber, Detecting inconsistencies in large biological networks with answer set programming. Theory and Practice of Logic Programming, 11(2-3):323-360, 2011.
- P. Gloaguen, P. Crépieux, D. Heitzler, A. Poupon and E. Reiter, Mapping the folliclestimulating hormone-induced signaling networks., Front Endocrinol., 2:45, 2011.
- C. Guziolowski, J. Gruel, O. Radulescu and A. Siegel, Curating a Large-Scale Regulatory Network by Evaluating Its Consistency with Expression Datasets, CIBB, 144-155, 2008.
- M. Hucka et al., The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models., Bioinformatics, 19(4):524-31, 2003.
- D. Hwang, J.J. Smith, et al., A data integration methodology for systems biology: Experimental verification, PNAS, 102(48):17302-7, 2005.
- 12. K. Inoue, Logic Programming for Boolean Networks., Proc. of IJCAI, 924-930, 2011.
- K. Iwanuma, K. Inoue, O. Ray, SOLAR: An automated deduction system for consequence finding. AI Commun. 23(2-3): 183-203, 2010.
- B. Kemper, T. Matsuzaki, et al. PathText: a text mining integrator for biological pathway visualizations. Bioinformatics [ISMB]. 26(12):374-381, 2010.
- H. Kitano, A. Funahashi, Y. Matsuoka and K. Oda, Using process diagrams for the graphical representation of biological networks., Nat Biotechnol., 23(8):961-6, 2005.
- N. Le Novère et al., The Systems Biology Graphical Notation., Nat Biotechnol., 27(8):735-41, 2009.
- F. Markowetz and R. Spang, Inferring cellular networks a review, BMC Bioinformatics, 8 Suppl 6:S5, 2007.
- A. Mitsos, I.N. Melas, P. Siminelakis, A.D. Chairalaki, J. Saez-Rodriguez and L.G. Alexopoulos, Identifying drug effects via pathway alterations using an integer linear programming optimization formulation on phosphoproteomic data., PLoS Comput Biol., 5(12), 2009.
- M.K. Morris, J. Saez-Rodriguez, D.C. Clarke, P.K. Sorger and D.A. Lauffenburger, Training Signaling Pathway Maps to Biochemical Data with Constrained Fuzzy Logic: Quantitative Analysis of Liver Cell Responses to Inflammatory Stimuli, PLoS Comput Biol., 7(3), 2011.
- N. Papin, H. Tony, O.P. Bernhard and S. Shankar, Reconstrution of cellular signalling networks and analysis of their properties, Nat Rev Mol Cell Biol., 6(2):99-111, 2005.
- R.J. Prill, J. Saez-Rodriguez, L.G. Alexopoulos, P.K. Sorger and G. Stolovitzky, Crowdsourcing Network Inference: The DREAM Predictive Signaling Network Challenge., Sci Signal, 4:mr7, 2011.
- K. Sachs, O. Perez, D. Pe'er, D.A. Lauffenburger and G.P. Nolan, Causal protein-signaling networks derived from multiparameter single-cell data., Science, 308(5721):523-9, 2005.
- J. Saez-Rodriguez, L.G. Alexopoulos, J. Epperlein, R. Samaga, D.A. Lauffenburger, S. Klamt and P.K. Sorger. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction., Mol Syst Biol. 5:331, 2009.
- C.H. Yeang, T. Ideker and T. Jaakkola, Physical Network Models, J Comput Biol., 11(2-3):243-62, 2004.
- D. Wittmann, J. Krumsiek, J. Saez-Rodriguez, D.A. Lauffenburger, S. Klamt and F. Theis, From Qualitative to Quantitative Modeling, BMC Syst. Biol. 3:98, 2009.

II- Base de connaissances

Dans cette première phase de développement, nous avons appliqué notre méthode à des jeux de données de petite taille. Il est cependant apparu clairement que la manipulation manuelle des données deviendra très rapidement impossible. Nous avons donc choisi de développer dès le départ une base de données pour gérer à la fois les règles, les données expérimentales et les faits, que ce soient les faits initiaux, directement déduits des expériences, ou les faits déduits par inférence automatique. De plus, cette base permet l'exportation du réseau déduit vers un logiciel donnant une représentation graphique, permettant ainsi à l'utilisateur d'appréhender et d'exploiter ce réseau. La base de connaissance permet une traçabilité rigoureuse des données, ainsi que le stockage d'expériences négatives, souvent considérées comme inintéressantes alors qu'elles peuvent permettre d'éliminer des hypothèses.

A ce jour, le modèle de données est défini, et la base de données est fonctionnelle. En revanche, l'interface utilisateur est encore en cours de développement. Le détail du modèle de données est décrit dans la partie Matériel et Méthodes. Nous ne présenterons ici que les caractéristiques globales du modèle.

A- Modèle de données

La structure générale de la base de données est présentée sur la Figure 51.

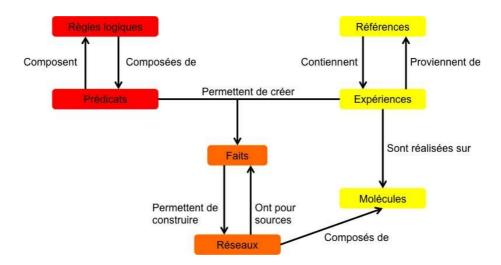


Figure 51 : Structure générale de la base de données

Le modèle de données est conçu de manière à ce qu'une expérience soit obligatoirement reliée à une référence (une source). Cette référence peut être une publication ou un auteur ayant réalisé l'expérience en question.

Les expériences sont liées à des prédicats. Ces prédicats sont liés aux règles logiques qui les utilisent. Les prédicats liés à une expérience sont utilisés pour créer les faits relatifs à l'expérience. Lorsque ces faits sont des relations entre des molécules, ils sont étiquetés comme tels, et permettront ensuite de construire les réseaux moléculaires.

La traçabilité est un élément essentiel pour obtenir des réseaux moléculaires fiables. Nous voulons permettre aux utilisateurs de remonter jusqu'à la source de l'information. Un fait initial est nécessairement relié à une source. Pour les faits déduits les choses sont bien évidemment un peu plus complexes puisqu'ils peuvent résulter du déclenchement successif de plusieurs règles. De plus, ces déductions sont faites par le raisonneur (moteur d'inférence), et donc en dehors de la base. Comme cela a été cité dans les options utilisées dans SOLAR, il est possible d'attacher à un fait déduit tout l'enchaînement des règles déclenchées. Il sera ainsi possible dans la base de relier un fait déduit à tous les faits initiaux qui ont été utilisées pour aboutir à cette conclusion, et également à toutes les règles qui ont été utilisées.

Pour illustrer cette traçabilité des données, nous allons prendre l'exemple de deux contradictions obtenues lors de l'inférence du réseau EGF :

- pJAK2 active et n'a pas d'effet sur la phosphorylation de Akt à la position S473 : PHOSPHORYLATE(pJAK2,Akt,pAktS473,indirect,confirmed,increase) et PHOSPHORYLATE(pJAK2,Akt,pAktS473,unknown,confirmed,noeffect),
- même chose pour pJAK1 : PHOSPHORYLATE(pJAK1,Akt,pAktS473,indirect,confirmed,increase) et PHOSPHORYLATE(pJAK1,Akt,pAktS473,unknown,confirmed,noeffect).

Afin de comprendre la raison de ces contradictions, nous avons étudié l'enchaînement des règles et des faits qui ont permis d'aboutir à ces différentes conclusions. Les relations ayant un effet nul ont été obtenues directement grâce à des expériences de *Phosphorylation Assay* avec un inhibiteur des protéines JAK1/2 phosphorylées [212] alors que les conclusions sur l'augmentation de la phosphorylation d'Akt par les phosphoJAK1/2 résultent du déclenchement de la règle suivante :

• IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase) AND PHOSPHORYLATE(b,z,d,unknown,confirmed,increase) THEN PHOSPHORYLATE(x,z,d,indirect,confirmed,increase) : si x active la

phosphorylation de y en b et si b active la phosphorylation de z en d, alors x active de manière indirecte la phosphorylation de z en d.

Les faits ayant permis de déclencher cette règle sont :

- PHOSPHORYLATE(pSTAT1Y701,Akt,pAktS473,unknown,confirmed,increase) qui lui-même est déduit d'une expérience de *Phosphorylation Assay* en présence de Fludara, un inhibiteur de STAT1 phosphorylée sur la position Y701.
- PHOSPHORYLATE(pJAK1/2,STAT1,pSTAT1Y701,unknown,confirmed,increase) qui lui-même est déduit d'une expérience de *Phosphorylation Assay* en présence de AG490, un inhibiteur des phosphoJAK1/2.

Ainsi, le déclenchement de la règle conduit aux faits PHOSPHORYLATE(pJAK1,Akt,pAktS473,indirect,confirmed,increase) et PHOSPHORYLATE(pJAK2,Akt,pAktS473,indirect,confirmed,increase).

Il y a donc un conflit dans le réseau. Grâce aux liens présents dans la base de données, il est possible de revenir aux publications ou aux expériences qui aboutissent à ces deux conclusions contradictoires. Si les faits expérimentaux qui ont été entrés au départ sont effectivement justes, alors on est en présence d'une vraie contradiction, et dans ce cas de nouvelles procédures devront être développées (voir plus loin). Par contre, s'il s'agit d'une erreur à l'entrée des données, ou bien d'une erreur d'interprétation de l'une des expériences, cette traçabilité permettra une correction rapide.

Ceci a été réalisé pour la contradiction présentée dans l'exemple ci-dessus. Tous les faits utilisés pour aboutir sur cette incohérence proviennent de la même expérience (Figure 52).

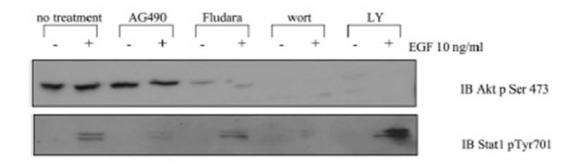


Figure 52: Phosphorylation Assay d'Akt et de STAT1 provenant d'Andl et al [212]

La phosphorylation d'Akt sur la position S473 augmente en présence d'EGF et il est indiscutable qu'en présence d'EGF et de Fludara, cette phosphorylation diminue fortement. Ceci permet d'affirmer que le fait déduit PHOSPHORYLATE(pSTAT1Y701, Akt,

pAktS473, unknown, confirmed, increase) est vrai. De même, les faits déduits PHOSPHORYLATE(pJAK1, Akt, pAktS473, unknown, confirmed, noeffect) et PHOSPHORYLATE(pJAK2, Akt, pAktS473, unknown, confirmed, noeffect) sont indiscutables car en présence d'AG490, inhibiteur des JAK, la phosphorylation d'Akt ne varie pas de façon significative (Figure 52). Nous sommes donc en présence d'une vraie contradiction et il revient à l'utilisateur de prendre une décision.

Cet exemple illustre parfaitement la capacité de notre système à remonter des faits déduits à leur source.

B- Données expérimentales

Le modèle de données est conçu de manière à pouvoir stocker les données venant de tous les types d'expériences. Nous avons conçu un scénario, qui sera intégré à l'interface utilisateur de la base, et qui permettra de générer les faits correspondants. Dans ce scénario, l'interface va chercher dans la base les différents prédicats qui se rapportent au type d'expérience considéré, puis demande à l'utilisateur de choisir différentes valeurs à chacune des variables présentes dans le prédicat afin de l'instancier. Les variables seront renseignées de manière automatique chaque fois que cela sera possible. Nous créerons également des procédures permettant de lire les fichiers issus d'expériences à haut-débit, d'insérer les données correspondantes dans la base et de générer les faits.

C- Interface utilisateur

L'interface de saisie des données est en cours de développement. Elle permettra la création de nouvelles règles et l'insertion des données expérimentales.

L'interface d'insertion des données expérimentales représentera un premier système de contrôle des données. Comme nous avons défini des types moléculaires bien précis pour chaque protocole expérimental, les utilisateurs ne pourront pas insérer d'incohérences dans la base. Par exemple, si un utilisateur veut insérer des données sur un *Phosphorylation Assay*, il devra instancier le prédicat PA(x,y,a,e). L'interface lui permettra de saisir uniquement les éléments x, a et e correspondants respectivement au signal, à l'anticorps et à l'effet. L'utilisateur aura un large choix concernant le signal, en revanche, la saisie de l'anticorps sera limitée uniquement à ceux qui ciblent des protéines phosphorylées. L'argument y sera récupéré automatiquement grâce au choix de l'anticorps car la base de données contient les

liens entre les protéines et leurs formes phosphorylées. En ce qui concerne l'effet, l'utilisateur devra choisir entre '*increase*', '*decrease*' et '*noeffect*'.

Nous sommes conscients du fait qu'il faut entrer un grand nombre de données manuellement dans la base, ce qui fait de l'acquisition des données une étape limitante. Cependant, nous pensons que la plus-value liée à l'inférence automatique du réseau sera une motivation suffisante pour les biologistes. Nous envisageons également de créer des interfaces avec les cahiers de laboratoire électroniques, en particulier celui qui est en cours de développement dans notre équipe.

Pendant cette phase de développement, la base est ouverte uniquement à notre équipe, mais notre objectif à terme est de l'ouvrir largement à la communauté scientifique. Nous espérons que des chercheurs intéressés par d'autres réseaux moléculaires pourront également contribuer à l'élargissement de la base de données.

Ces limitations ne concernent cependant que les données à bas débit puisque les données à haut débit seront traitées de façon automatique.

D- Lien avec le moteur d'inférence

La base est conçue pour permettre l'utilisation transparente de la méthode d'inférence. Les faits initiaux ainsi que les règles peuvent être exportés depuis la base dans le format de SOLAR. Les faits déduits par la méthode d'inférence peuvent ensuite être réinsérés dans la base (Figure 53), et un graphique CellDesigner peut être généré à partir de l'ensemble des faits. Comme nous l'avons expliqué précédemment, la création d'un historique des règles et faits utilisés par SOLAR pour parvenir à un fait déduit nous permettra, lors de l'insertion de ce fait déduit dans la base, de le lier à tous les faits initiaux utilisés lors du raisonnement. Cela permettra de remonter aux expériences en cas d'apparition d'incohérences dans la base.

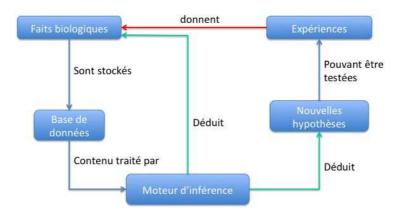


Figure 53 : Lien entre la base de données et le moteur d'inférence

E- Choix de visualisation graphique

Pour faciliter l'exploitation des conclusions que SOLAR fournit, il a fallu choisir un langage de représentation graphique et un logiciel de visualisation des réseaux. La Figure 54 montre différentes représentations à partir de trois langages et trois logiciels.

Le langage SBML est, depuis quelques années, un standard dans le domaine de la représentation graphique des réseaux biologiques car il facilite l'échange des données. La version actuelle est le niveau 2 version 4 mais cette représentation est limitée. Le SBML ne prend pas en compte plusieurs choses importantes dans un réseau moléculaire. Il n'y a pas de différence entre la catalyse et l'inhibition au niveau de la représentation graphique, cela se limite à des 'modifiers', c'est-à-dire des molécules qui ont une action non déterminée sur la réaction. Il est possible de voir sur la Figure 54B que le SBML ne fait également pas de différence entre les types moléculaires (protéine, complexe...), ils sont tous représentés de la même façon. Lorsque le SBML est combiné avec le langage de notation SBGN (langage CellDesigner), il est possible de constater que la représentation du réseau est beaucoup plus précise et compréhensible (Figure 54A). De plus, le langage CellDesigner est importable dans le logiciel Cytoscape grâce au plugin BiNoM, ce qui permet de visualiser un réseau créé à partir de CellDesigner dans Cytoscape (Figure 54C). Ce dernier permet également l'exportation en langage BioPax. La représentation de ce langage est illustrée par la Figure 54 parties D et E obtenues avec les logiciels Cytoscape et Sybil respectivement.

Le langage BioPax n'est pas réellement adapté à la représentation graphique, les schémas sont alourdis par les phrases qui décrivent les relations. Le SBML quant à lui manque de précision. Nous avons donc choisi d'utiliser le logiciel CellDesigner et son langage car comme illustré sur la Figure 54, il permet une représentation plus compréhensible et surtout plus précise que les autres langages. De plus, les biologistes de l'équipe utilisent ce logiciel depuis quelques années, il est donc plus aisé de présenter des réseaux dans une version qu'ils connaissent parfaitement. En ce qui concerne les éventuels utilisateurs qui préfèrent l'utilisation de Cytoscape, ce choix ne pose pas de problème puisqu'il est possible d'importer le langage CellDesigner dans ce logiciel. Notre choix permet aussi l'obtention des réseaux en BioPax, toujours à partir de Cytoscape. De plus, notre choix n'est pas figé puisqu'actuellement le niveau 3 du langage SBML est en cours de développement et compensera sûrement les lacunes que présente le niveau 2 version 4.

CellDesigner ne présente pas uniquement l'avantage de fournir une meilleure représentation graphique des réseaux, il permet également de faire le lien entre les molécules et des bases de

données telles que 'Entrez Gene'. De plus, il est possible d'ajouter des identifiants PubMed sur les relations, ce qui permet de remonter directement à la source bibliographique des informations. Ce logiciel permet également d'effectuer des simulations dynamiques.

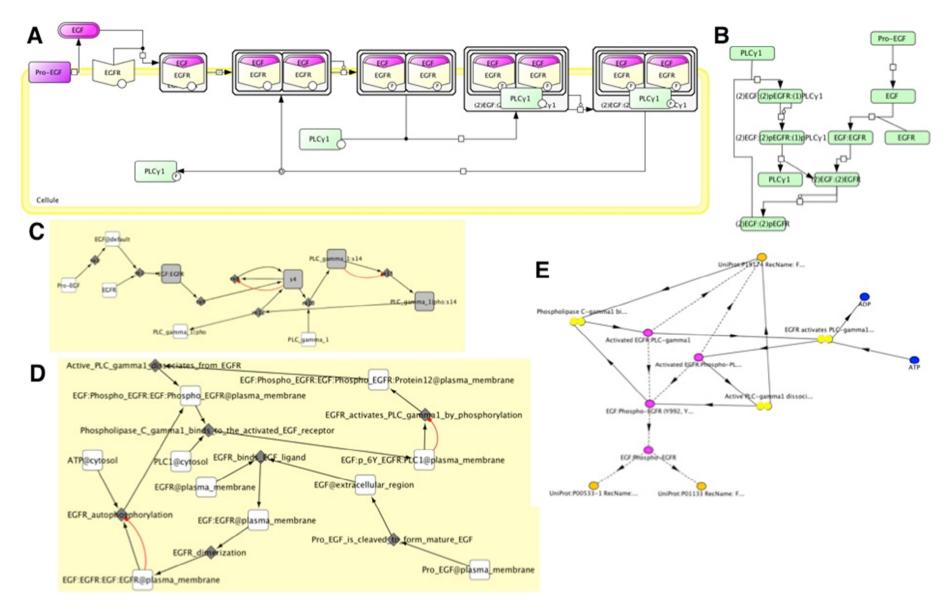


Figure 54 : Différentes représentations graphiques d'une partie du réseau EGF

(A) Représentation en langage CellDesigner dans CellDesigner. (B) Représentation en SBML niveau 2 version 4 dans CellDesigner. (C) Représentation en langage CellDesigner dans Cytoscape. (D) Représentation en langage BioPax dans Cytoscape. (E) Représentation en langage BioPax dans Sybil.

F- Liens avec d'autres bases de données

En introduction de ce manuscrit, différentes bases de données décrivant les réseaux moléculaires ont été évoquées, il est donc normal de se demander ce que notre nouvelle base de données pourrait apporter de plus.

Dans un premier temps, il existe des banques de données spécialisées dans les RCPG :

- GPCRDB (http://www.gpcr.org/7tm/, mise à jour le 26 mars 2012): référence les RCPG par espèce, les mutations existantes et les structures cristallines tridimensionnelles,
- gpDB (http://biophysics.biol.uoa.gr/gpDB/, mise à jour le 20 mars 2008) : référence les RCPG par espèce et les types de protéines G avec lesquelles ils se couplent,
- GLIDA (http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/, mise à jour le 10 octobre 2010) : référence les ligands,
- GPCR NaVa (http://nava.liacs.nl/, mise à jour le 16 novembre 2009) : référence les variants (mutants) naturels des RCPG,
- Sevens database (http://sevens.cbrc.jp/, mise à jour le 27 avril 2010) : référence la position des gènes codant pour les RCPG sur les chromosomes par espèce,
- IUPHAR (http://www.iuphar-db.org/DATABASE/ReceptorFamiliesForward?type=GPCR, mise à jour en juillet 2012): référence un grand nombre d'informations y compris la ou les protéines G pouvant se coupler aux RCPG.

Parmi ces six bases de données, seules les banques de données GPCRDB et IUPHAR sont encore mises à jour et aucune d'entre elles ne contient d'informations sur les réseaux de signalisation.

Nous établirons des liens avec ces différentes bases de données, permettant notamment d'aller chercher les caractéristiques du récepteur étudié.

Il existe également un grand nombre de bases de données généralistes, non commerciales, spécialisées dans le stockage des réseaux moléculaires : KEGG, REACTOME, STRING, IntAct. La Figure 55 représente les résultats que l'on obtient dans ces banques de données en effectuant une recherche sur le FSHR humain.

La banque de données KEGG¹¹ fournit uniquement la complexation de la FSH avec son récepteur. En matière de réseaux de signalisation, KEGG ne contient à ce jour que le réseau lié à la protéine p53. Cependant, il est probable que ce répertoire s'agrandisse dans les années à venir.

_

¹¹ http://www.genome.jp/kegg/

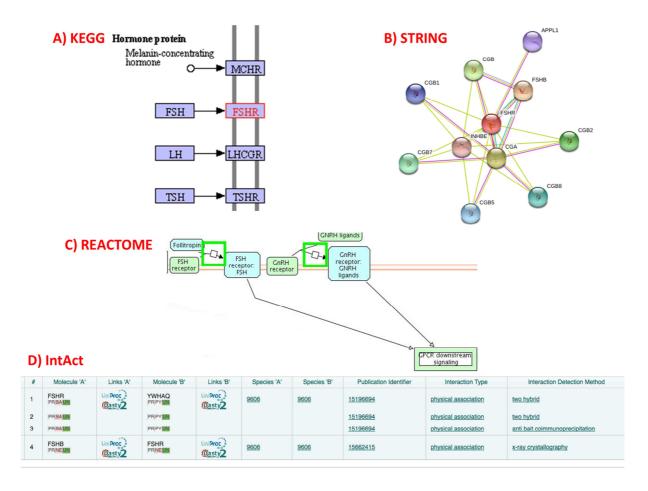


Figure 55 : Résultats obtenus dans différentes banques de données en effectuant une recherche sur le FSHR humain

REACTOME¹² contient également cette complexation puis un lien est proposé à l'utilisateur pour visualiser les voies G liées à ce type de récepteur. On signalera cependant l'absence de toutes les autres voies déclenchées par ce récepteur, ainsi que par un grand nombre d'autres RCPG, notamment les voies dépendantes des β-arrestines. Ici encore, on peut supposer que ce répertoire ira en s'élargissant, et il sera important pour nous de créer un lien avec cette base. IntAct¹³ et STRING¹⁴ permettent d'obtenir une liste de molécules interagissant directement avec le récepteur FSHR. NetPath dispose d'un réseau FSH plus complet (statistiques du réseau disponibles sur la Figure 56, la représentation graphique est disponible dans la publication de Telikicherla et al [126]).

12 http://www.reactome.org/

¹³ http://www.ebi.ac.uk/intact/

¹⁴ http://string-db.org/

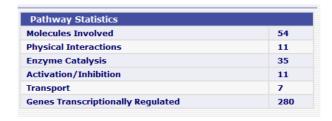


Figure 56 : Statistiques du réseau FSH disponibles dans NetPath¹⁵

Ces banques généralistes contiennent encore peu de données sur la signalisation, mais cela ira en augmentant. Ainsi il serait intéressant que nous soyons à l'avenir capables de récupérer ces données pour les présenter à l'utilisateur. Nous prendrons bien soin cependant de ne pas intégrer ces données avec le même statut que celles entrées directement dans la base. En effet, ces données ne sont pas accompagnées des données expérimentales qui ont permis de les établir. De plus, elles sont souvent issues de prédictions, ou d'un transfert de connaissances d'un réseau à un autre, et ne sont par conséquent pas toujours fiables. Cela peut cependant apporter une aide précieuse à l'utilisateur, en lui donnant des pistes. Il est possible par exemple, lorsqu'une relation intéressante existe dans l'une de ces bases de données, de chercher dans la littérature des justifications expérimentales, puis de les ajouter à notre base. Les informations stockées dans notre système seront ainsi moins nombreuses que dans une banque de données généraliste, mais plus complètes, et surtout reposant sur des données expérimentales fiables et identifiables. Etant donné que nous disposerons du traitement automatique des données à haut débit, notre banque de données pourra être remplie plus facilement et de façon plus complète qu'une banque de données telle que NetPath où les informations sont toutes ajoutées manuellement. Enfin, comme cela a été mentionné plus haut, le choix de visualisation en langage CellDesigner permettra la mise à disposition des fichiers de représentation des réseaux.

15 http://www.netpath.org/

_

III- Application aux réseaux de signalisation de la FSH et de l'EGF

Notre méthode nécessite d'être validée et comme nous l'avons dit précédemment, nous avons choisi de le faire sur deux réseaux de signalisation déclenchés au travers de deux types de récepteurs différents : le FSHR (récepteur type RCPG) et l'EGFR (récepteur type RTK). Lors de la recherche bibliographique, les expériences rencontrées dans la littérature sur le réseau FSH ont servi de guide pour créer les premières règles. De cette même littérature nous avons extrait des faits expérimentaux. Avec cet ensemble de règles et de faits nous avons pu reconstruire le modèle du réseau de signalisation de la FSH. Ensuite, nous avons réutilisé les mêmes 111 règles (disponibles en Annexe 4) pour inférer le réseau induit par l'EGF pour lequel les faits initiaux ont été extraits de la liste de publications fournie par la banque de données *Pathway Interaction Database*.

Le procédé qui a été utilisé a ainsi consisté à écrire des règles permettant d'interpréter de manière automatique les différentes expériences rencontrées dans les articles portant sur le réseau de signalisation de la FSH, puis des règles permettant de combiner les conclusions de ces premières règles élémentaires, et cela jusqu'à disposer de toutes les règles nécessaires pour reconstruire un réseau comparable à celui qu'on peut construire manuellement à partir des mêmes données. Il est bien évidemment très difficile de ne rien oublier dans ce processus, et les premiers essais d'inférence automatique nous ont amenés à ajouter des règles et des faits initiaux. Ce processus itératif a conduit à un premier ensemble de règles qui était suffisant pour reconstruire entièrement le réseau de signalisation lié à l'hormone FSH.

Nous avons dans un deuxième temps pu constater que cet ensemble de règles est également suffisant pour reconstruire le réseau de signalisation de l'EGF.

Par souci de lisibilité et de compréhension, les réseaux obtenus sont découpés en modules. Nous n'allons décrire ici que quelques-uns de ces modules, les autres modules sont disponibles dans l'Annexe 6. Dans une première partie, nous décrirons les résultats obtenus par l'inférence automatique, par rapport au réseau de référence (celui construit par les experts

du domaine). Dans une deuxième partie, nous discuterons certains points, notamment les cas où les faits déduits peuvent sembler *a priori* contradictoires.

Les molécules correspondent à la représentation SBGN décrite dans la partie Matériel et Méthodes. La FSH et l'EGF sont les seules molécules représentées par un symbole non décrit précédemment, la symbolique 'DRUG' a été utilisée dans ce cas afin de les différencier du reste du réseau. Les relations sont représentées telles qu'elles ont été introduites dans la partie Matériel et Méthodes.

A- Réseau FSH

Pour construire le réseau FSH, plus de 100 articles ont été analysés, parmi lesquels 62 ont été retenus pour l'écriture de la revue, 8 autres articles ont permis la création des voies supplémentaires décrites précédemment (stéroïdogenèse...) et 18 articles supplémentaires ont permis de répertorier les régulations de transcription et de traduction présentées en Annexe 5. Une partie des publications n'a pas été prise en compte car il n'y avait pas d'expériences permettant la déduction de relations du réseau. Dans les 88 publications utilisées, une partie des expériences sont redondantes ou nous ne disposons pas des règles pour les interpréter, c'est pourquoi l'inférence du réseau FSH a été finalement réalisée à partir de 38 publications, qui ont permis de créer :

- 542 faits concernant les molécules regroupant :
 - o le type moléculaire de chaque molécule,
 - o la localisation cellulaire de chaque molécule si cela était possible.
 - o les couples Anticorps/cible(s), Inhibiteur/cible(s), Molécule modifiée/Molécule sauvage...
- 240 faits issus d'expériences (PA, ELISA, RIA, EA, ICPPA...),
- 25 relations gène/ARN et ARNm/protéine sont issues de banques de données,
- 9 faits issus du savoir implicite¹⁶.

L'utilisation de ces faits par le moteur d'inférence SOLAR nous a permis d'obtenir 572 conclusions, parmi lesquelles, 381 peuvent être représentées graphiquement. Il faut ajouter à cela 137 relations ayant un effet nul 'noeffect' et 54 relations qui concernent des agonistes et des analogues structuraux.

¹⁶ Le savoir implicite signifie que la relation est admise. Une telle relation provient d'une ou plusieurs expériences difficilement formalisables, cependant elle doit obligatoirement été démontrée plusieurs fois.

Une comparaison des trois principales voies décrites dans la revue (Voies G stimulatrice, β-arrestine et PI3K) avec celles que nous avons obtenues par inférence automatique permet d'évaluer la pertinence de notre méthode. Dans la revue, nous avons considéré différentes formes de certaines protéines comme une seule entité : c'est le cas des ERK1 et ERK2 qui deviennent ERK1/2. Nous avons donc fusionné ces protéines dans le réseau obtenu pour faciliter la comparaison.

1- Voie dépendante de la protéine G stimulatrice

Les deux voies G stimulatrices obtenues manuellement et automatiquement sont présentées sur la Figure 57. Il manque certaines relations dans la voie générée automatiquement mais cela s'explique par l'absence des règles ou des faits permettant leur création.

Par exemple, la phosphorylation de la protéine Raf1 par la PKAc est absente car il existe des désaccords concernant cette relation. La PKAc phosphoryle Raf1 sur la position S259 et certains auteurs pensent que cette phosphorylation est activatrice alors que d'autres auteurs pensent que c'est la phosphorylation sur la position S338 qui permet l'activation de Raf1. Ongeri et al [213] montrent qu'il y a augmentation de l'activité de Raf1 et en parallèle une augmentation de sa phosphorylation sur la position S338 dans les 30 premières minutes de stimulation et, qu'après 30 minutes, cette phosphorylation ainsi que l'activité de l'enzyme diminuent. En parallèle, ils quantifient la phosphorylation sur la position S259, et l'expérience montre une diminution de cette phosphorylation dans les 30 premières minutes puis une augmentation après 30 minutes de stimulation. Ceci confirmerait que la phosphorylation sur S259 serait bien inhibitrice de l'activité de Raf1. Comme il existe une ambiguïté sur la phosphorylation de cette molécule par PKAc et comme l'action de cette dernière a été déduite à partir de publications ne portant pas sur la FSH, nous n'avons pas considéré le fait permettant sa création.

Il manque également la phosphorylation de la PTP par la PKAc, ce qui provoque une levée d'inhibition sur les ERK MAPK. Cette relation est en réalité une hypothèse émise par les auteurs à partir des résultats d'une expérience de « *In-gel PTP assay* » [42]. Bien que cette relation soit absente, le réseau inféré automatiquement montre tout de même une augmentation de la phosphorylation des ERK MAPK sous l'action indirecte de la PKAc.

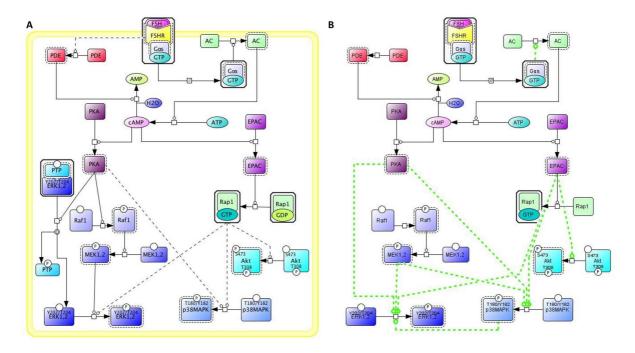


Figure 57: Voie G manuelle versus voie G automatique

(A) Voie G construite manuellement, (B) Voie G construite automatiquement avec élagage manuel.

Une autre différence présente sur la Figure 57 concerne EPAC. La protéine EPAC induit directement l'activation de Rap1, qui va ensuite, de manière indirecte, induire la phosphorylation d'Akt, de p38MAPK et des ERK1/2. Or les expériences permettent de mettre en évidence l'action d'EPAC (mutant constitutivement inactif), et non l'action de Rap1. Ainsi, même s'il est couramment admis que l'action d'EPAC passe nécessairement par Rap1, nous n'avons pas trouvé dans les publications d'expérience qui puisse le prouver. De ce fait, on ne voit pas l'action de Rap1 sur les molécules qui sont en aval dans notre réseau. Dans le réseau inféré automatiquement, les MEK1/2 phosphorylées induisent la phosphorylation de p38MAPK et p38MAPK phosphorylée induit la phosphorylation des ERK1/2, ces relations sont confirmées, alors qu'elles n'apparaissent pas dans le réseau manuel. Nous avons obtenu ces conclusions grâce à trois expériences de *Phosphorylation Assay* de :

- p38MAPK en présence de FSH et d'inhibiteurs des MEK1/2, PD98059 ou U0126 [77],
- ERK1/2 en présence de FSH et d'un inhibiteur de p38MAPK, SB203580 [214].

Dans les différents cas, les inhibiteurs ont provoqué une diminution de la phosphorylation de la protéine considérée, c'est pourquoi nous avons obtenu ces relations. Ces conclusions n'apparaissent pas dans l'article publié car il faut tenir compte de la spécificité des inhibiteurs et nous n'avons pas trouvé suffisamment d'expériences justifiant ces relations.

2- Voie dépendante des β-arrestines :

Comme il est possible de le voir sur la Figure 58, les voies β -arrestines manuellement et automatiquement obtenues présentent des différences importantes.

La différence la plus notable est le fait que l'élément central dans le réseau obtenu est la FSH, alors que ce devrait être le récepteur couplé à la FSH. Ce phénomène est retrouvé dans chacun des modules, et ce sera également le cas pour le récepteur EGF. Les molécules telles que la FSH ou l'EGF ne sont pas capables d'agir sur une cellule si cette dernière n'exprime pas leur récepteur. La formalisation actuelle prend en compte la stimulation (FSH ou EGF dans nos modèles) sans tenir compte de l'obligation d'une liaison avec un récepteur. Il sera nécessaire d'exprimer de façon formelle que les molécules telles que la FSH et l'EGF agissent uniquement sous forme de complexe avec leur récepteur.

L'action des GRK et des β-arrestines sur la phosphorylation des ERK MAPK est hypothétique dans le réseau inféré automatiquement, ceci est dû au fait que les données expérimentales utilisées pour obtenir ces conclusions reposent sur des expériences dans lesquelles des siRNA ont été utilisés. Or, l'utilisation des siRNA conduit à l'absence d'une protéine et de toutes ses formes modifiées, on ne peut donc pas savoir à partir de ces expériences quelle forme est active. En effet, l'application des règles mène à faire l'hypothèse que n'importe laquelle des formes de la protéine ciblée par le siRNA peut être la forme active. Pour déterminer la forme active, des expériences supplémentaires sont nécessaires.

En ce qui concerne l'action des GRK sur le récepteur, l'hypothèse la plus étayée est que ces protéines agissent directement sur le récepteur, en le phosphorylant à des positions déterminées, et que ce « code barre » de phosphorylations dirige le récepteur soit dans la voie de désensibilisation, soit dans la voie de signalisation. Cependant, les expériences qui nous auraient permis d'obtenir ces hypothèses reposent sur l'utilisation de mutants. Or, ces dernières sont particulièrement difficiles à formaliser, et nous n'avons pas encore écrit les règles correspondantes.

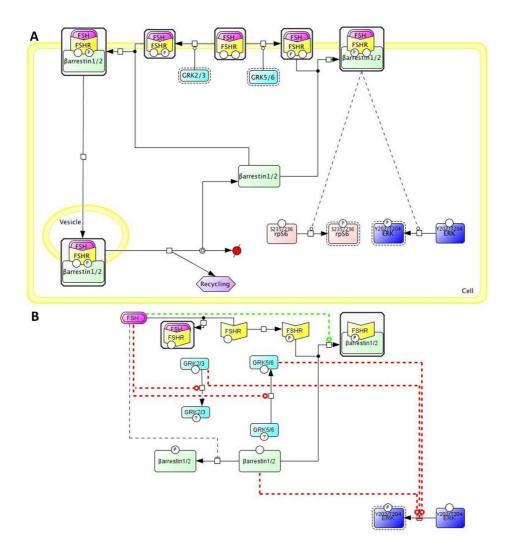


Figure 58 : Voie β -arrestines manuelle versus voie β -arrestines automatique

(A) Voie β -arrestines construite manuellement, (B) Voie β -arrestines construite automatiquement avec élagage manuel.

Sur la Figure 58A, les β-arrestines agissent sur la phosphorylation des ERK MAPK via leur complexation avec le récepteur FSHR phosphorylé lié à la FSH. Sur la Figure 58B, on voit qu'on a bien obtenu la complexation avec le récepteur FSHR phosphorylé sous l'action de la FSH mais pas l'action du complexe sur la phosphorylation des ERK MAPK. Dans le réseau manuellement construit, ceci est défini par la combinaison des expériences permettant de déduire la complexation des β-arrestines avec le récepteur et l'action des β-arrestines sur la phosphorylation des ERK MAPK. Il serait possible de tenter une formalisation de ce raisonnement mais il est nécessaire d'être prudent car il faut que la règle soit toujours vraie. Or la règle « si x se complexe avec z et si x induit la phosphorylation de y en b alors c'est le complexe x:z qui induit la phosphorylation de y en b » ne peut pas être toujours vraie. Par

exemple, il existe deux formes différentes du récepteur qui se lient avec les β -arrestines, mais induisent des événements différents. Donc, pour obtenir cette relation, il serait nécessaire d'avoir une expérience permettant de déterminer quel complexe est actif sur quelle voie.

3- Voie dépendante de la PI3K :

De la même façon, on peut noter des différences entre la voie PI3K de référence et celle obtenue automatiquement (Figure 59). Une différence importante est la perte de l'ordre d'action des molécules dans la cascade ERK/TSC2/Rheb/mTOR.

Cette cascade a été étudiée dans le cas de la FSH, mais les expériences ne permettent pas de déduire le lien entre TSC2 et Rheb et entre Rheb et mTOR. Ces relations sont décrites dans le réseau manuellement construit grâce au savoir obtenu dans d'autres réseaux.

On peut également constater que le modèle automatique conclut sur l'existence de trois formes de mTOR: mTOR, mTOR phosphorylée et mTORC1, alors que dans le modèle de référence apparaissent seulement deux formes: mTOR et mTOR activée. En fait, mTOR activée correspond à mTORC1, qui est un complexe dont mTOR phosphorylée est partenaire. Néanmoins, mTOR est également phosphorylée pour entrer dans la composition d'un autre complexe mTORC2. Cependant, il semblerait qu'il y ait une dominance de la phosphorylation de mTOR sur la position S2448 dans le complexe mTORC1 et une dominance de la phosphorylation de mTOR sur la position S2481 dans le complexe mTORC2 [215], mais la différence des quantités de phosphorylation ne nous permet pas d'utiliser ces données.

De même, la méthode automatique conclut sur l'existence de quatre formes d'Akt, une forme non phosphorylée et trois formes phosphorylées (Annexe 6 : Réseau FSH, voie PI3K). Parmi les trois formes phosphorylées, il y en a une pour laquelle la position du résidu touché par la phosphorylation n'est pas définie car elle n'est pas indiquée dans la publication utilisée. L'existence de ces différentes formes met en évidence l'utilité de créer une règle permettant l'agrégation des différentes modifications post-traductionelles sur une seule représentation de la protéine impliquée.

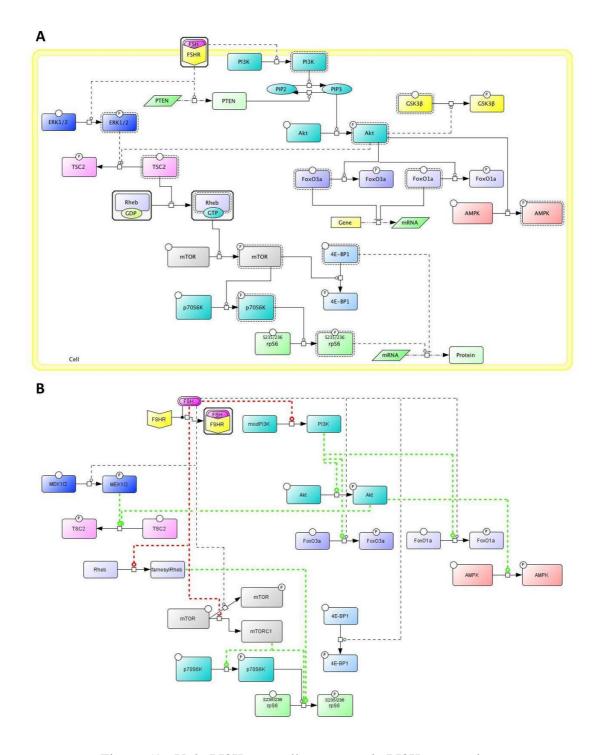


Figure 59: Voie PI3K manuelle versus voie PI3K automatique

(A) Voie PI3K construite manuellement, (B) Voie PI3K construite automatiquement avec élagage manuel.

4- Réseau obtenu

La totalité du réseau obtenu est présentée sous forme de modules dans l'Annexe 6. L'analyse de ces différents modules permet de constater l'existence de contradictions entre certaines

relations hypothétiques obtenues. Par exemple, le module PKA contient les relations suivantes :

- La FSH induit hypothétiquement l'activation de la PKA pour donner PKAc;
- La FSH inhibe hypothétiquement cette même relation.

L'inférence a aussi permis de générer des conclusions de type « *noeffect* » qui ne sont pas représentées sur les graphes. La grande majorité de ces relations est cohérente avec le réseau obtenu car elles concernent surtout les formes inactives des molécules telles que PKA, EPAC ou encore MEK1.

Il existe toutefois deux exceptions:

- Dans le module « Actions impliquant la PKA » (Annexe 6), on obtient d'une part le fait que la PKAc active la phosphorylation de p38MAPK, et d'autre part le fait qu'elle ne l'active pas, les deux faits ayant le statut « confirmed ».
- Dans le module « Actions impliquant la PKA » (Annexe 6), on obtient la conclusion que la PKAc active la phosphorylation de la tubérine, et d'autre part le fait qu'elle ne l'active pas, les faits ayant le statut « confirmed ».

La génération de toutes ces contradictions sera discutée dans la section « Les leçons de ces premiers tests ».

B- Réseau EGF

1- Présentation du récepteur EGF

L'Epidermal Growth Factor (EGF) est un facteur de croissance. Il se lie à un récepteur transmembranaire de type tyrosine kinase, l'EGFR. Cette famille de récepteurs a la particularité de faire partie des récepteurs-enzymes. Lorsqu'ils fixent leur ligand, ils se dimérisent (Figure 60), ce qui provoque une autophosphorylation des résidus tyrosines des domaines intracellulaires du dimère.

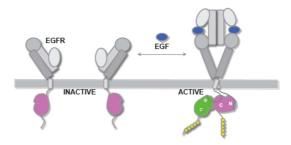


Figure 60 : Dimérisation du récepteur de l'EGF [216]

Ces phosphorylations entraînent un changement de conformation, ce qui permet à des protéines de venir se lier aux récepteurs. Cette complexation aboutit à la transduction du signal et l'activation de voies de signalisation (Figure 61).

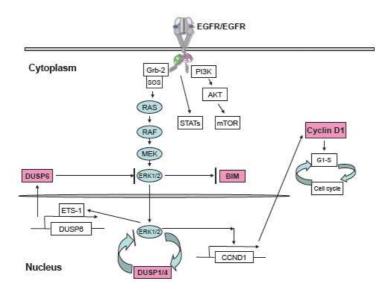


Figure 61 : Voies de signalisation induites par l'EGF [216]

Comme la création des règles avait été guidée par les expériences rencontrées dans l'étude de la signalisation dépendante de la FSH, il était important de vérifier qu'elles pouvaient être appliquées à l'inférence d'un autre réseau moléculaire. D'autre part, il est également nécessaire de déterminer si la création de nouvelles règles est ou non nécessaire lorsqu'on aborde un nouveau réseau, ou si au contraire les règles sont suffisamment génériques pour permettre l'inférence de n'importe quel réseau moléculaire. Le réseau induit par l'EGF est très étudié et fait l'objet d'un large consensus, c'est pourquoi il semblait approprié de le prendre comme modèle de validation. La banque de données *Pathway Interaction Database* (version du 15/02/2012) contient un réseau de signalisation induit par l'EGF (http://pid.nci.nih.gov/search/pathway_landing.shtml?what=graphic&jpg=on&pathway_id=2 00101) avec les références bibliographiques qui ont permis de le construire. Cela nous a permis d'extraire les faits afin de construire le réseau à partir des mêmes sources. Ce réseau a également été mis au format CellDesigner pour faciliter la comparaison (Figure 62).

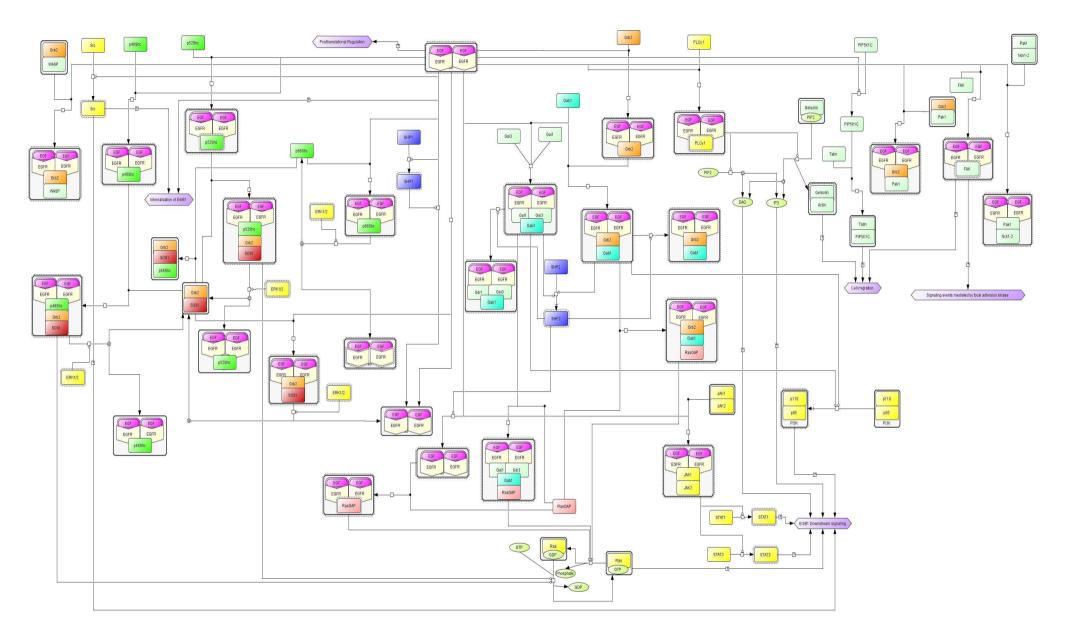


Figure 62 : Schématisation du réseau EGF issu de la banque de données *Pathway Interaction Database*, mis au format CellDesigner

2- Réseau EGF obtenu versus réseau EGF de référence

Pour créer les faits initiaux du réseau EGF, 55 publications ont été utilisées sur les 85 proposées. Une partie des publications présentait des expériences que nous n'avons pas encore formalisées (par exemple, immunoprécipitation en présence de mutants), ou alors, plus surprenant, ne citait à aucun moment l'EGF ou un de ces agonistes. Le réseau EGF a donc été construit à partir de :

- 572 faits concernant les molécules regroupant :
 - o le type moléculaire de chaque molécule,
 - o la localisation cellulaire de chaque molécule si cela était possible,
 - o les couples Anticorps/cible(s), Inhibiteur/cible(s), Molécule modifiée/Molécule sauvage...
- 323 faits issus d'expériences (PA, ELISA, RIA, EA,...),
- 5 faits issus du savoir implicite.

L'application de la méthode d'inférence nous a permis d'obtenir 561 conclusions en utilisant les mêmes règles que pour le réseau FSH. Parmi ces conclusions, 113 ont un effet nul et 40 prennent des agonistes en arguments, et ne sont pas prises en compte dans la représentation graphique du réseau.

Le réseau que nous avons obtenu grâce à notre méthode est plus large que le réseau de référence (totalité du réseau disponible en Annexe 6). Ceci est du au fait que dans la PID, le réseau a été divisé en deux parties dont certaines publications sont communes. Nous obtenons donc très logiquement des relations qui sont représentées dans la deuxième partie du réseau dans la *Pathway Interaction Database* (PID), et n'apparaissent donc pas sur la Figure 62.

La Figure 63 représente le réseau issu de la PID auquel un code couleur a été ajouté de façon à mettre en évidence les points communs et les différences avec le réseau inféré automatiquement.

La comparaison du réseau obtenu avec le réseau de référence (réseau de la PID) fait ressortir plusieurs différences. La première est que nous n'avons pas obtenu la liaison de l'EGF sur le récepteur. Ceci est dû au fait que dans les publications citées, aucune ne démontre expérimentalement la liaison de l'EGF au récepteur. Nous considèrerons que cela fait partie du savoir pré-existant.

De même, l'activation de PIP5K1C n'est pas obtenue car aucune preuve expérimentale n'apparaît dans les publications citées, cette relation est seulement décrite dans le résumé de Sun et al [217]. Ces mêmes auteurs ont démontré expérimentalement la complexation entre

PIP5K1C et la Taline, mais cette relation n'a pas pu être obtenue car ce résultat provient d'une expérience d'immunoprécipitation mettant en jeu des mutants pour lesquels le raisonnement n'a pas encore été formalisé. Il en est de même pour les relations suivantes :

- la déphosphorylation du récepteur par SHP1 a été démontrée par Haj et al [218] par *Knock-Out* ;
- la dissociation du complexe Gelsoline-PIP2 entraînant la complexation de la Gelsoline avec l'actine a été montrée par Chou et al [219] par des expériences de colocalisation par imagerie cellulaire (microscopie confocale, fluorescence...);
- l'activation de la PI3K par le complexe EGF-EGFR-Gαi1-Gαi3-Gab1 a été montrée par *Knock-Out* de Gab1 [220] ;
- la complexation du récepteur EGFR actif avec la protéine RasGAP a été obtenue par *Far Western Blot* [221].

Le réseau de référence fait apparaître un complexe PAK1-Nck1-2 qui s'associe au récepteur EGFR actif. L'existence du complexe PAK1-Nck1-2 n'est pas démontrée en présence d'EGF, et son association a été montrée pour le récepteur PDGFR et non EGFR [222].

De nombreuses actions des complexes présents dans le réseau issu de la PID sont absentes ou sont obtenues de façon différente (ce qui sera expliqué plus loin). Ceci s'explique par l'absence, dans les publications utilisées, d'expériences permettant de déduire ces relations :

- L'activation de SHP2 par le complexe EGF-EGFR-Gai1-Gai3-Gab1,
- L'activation des STAT1 et 3 par le complexe EGF-EGFR-JAK1/2,
- L'hydrolyse de PIP2 en IP3 et DAG sous l'action du complexe EGF-EGFR-PLCy,
- Les différentes régulations sur la protéine Ras.

Pour finir, le réseau issu de la PID contient un grand nombre de dissociations de complexes régulées par les ERK MAPK, mais aucune expérience présente dans la bibliographie proposée ne les démontre expérimentalement, elles n'ont par conséquent pas été obtenues.

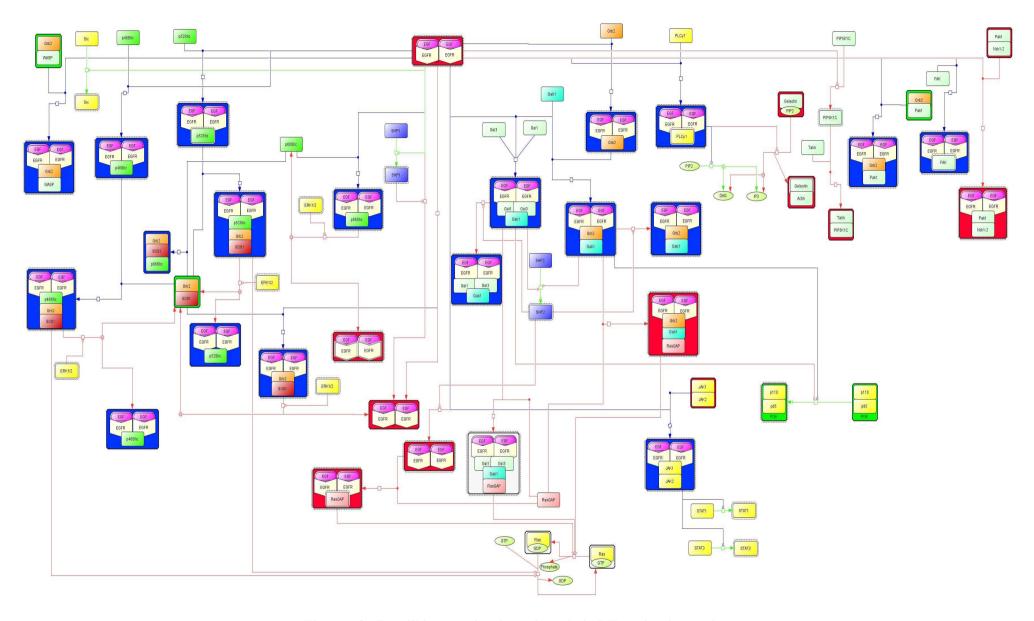


Figure 63 : Parallèle entre le réseau issu de la PID et le réseau obtenu

Rouge: molécules et relations non obtenues, Bleu: molécules et relations obtenues avec une différence, Vert: molécules et relations obtenues 170

Les complexes représentés avec un fond bleu sur la Figure 63 ont été obtenus mais ne sont pas le résultat des mêmes relations. Si on prend l'exemple du complexe (EGF-EGFR)₂-Grb2-SOS1 (Figure 64, partie de gauche), on constate que dans le réseau de référence il est issu de la complexation d'un dimère de EGF-EGFR et du complexe Grb2-SOS1. Dans le réseau obtenu, on a seulement un complexe pEGFR-Grb2-SOS1, et il est issu de la complexation directe d'un récepteur phosphorylé (phosphorylation induite par l'EGF avant la complexation avec Grb2 et SOS1) avec une protéine Grb2 et une protéine SOS1. L'activation (représentée par des pointillés autour des molécules) de ces récepteurs dépend aussi de leur état de phosphorylation. Lorsqu'ils sont phosphorylés alors la complexation avec Grb2:SOS1 est possible. Cette relation est obtenue automatiquement et est représentée par la Figure 64 (partie de droite).

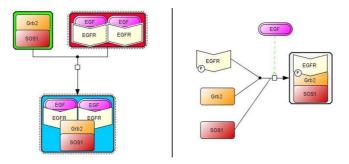


Figure 64 : Complexation entre le récepteur EGFR et Grb2:SOS1

L'EGF induit la complexation du récepteur EGFR phosphorylé et des protéines Grb2 et SOS1 avec une distance inconnue. Les expériences qui ont permis cette déduction sont des co-immunoprécipitations, dans lesquelles les auteurs ne révèlent jamais la présence d'EGF dans le complexe d'intérêt et comme cela a été précisé plus tôt, il n'existe pas encore de règle permettant d'affirmer que l'EGF agit uniquement grâce à la complexation avec son récepteur. Le résultat obtenu se limite donc à une régulation de la complexation par l'EGF avec une distance inconnue au lieu de sa présence dans le complexe. En ce qui concerne le remplacement du complexe Grb2:SOS1 par les molécules Grb2 et SOS1 isolées comme point de départ de la complexation, les expériences utilisées (co-immunoprécipitations dans tous les cas rencontrés) ne permettent pas de savoir si le complexe Grb2:SOS1 existe avant la formation du complexe EGF:EGFR:Grb2:SOS1. C'est pourquoi notre méthode permet d'inférer la complexation des trois molécules isolées en un complexe pEGFR:Grb2:SOS1, mais pas la complexation entre un récepteur EGFR phosphorylé et un complexe Grb2:SOS1. La totalité du réseau obtenu est présenté dans l'Annexe 6 sous forme de modules. Comme dans le cas du réseau FSH, les relations « noeffect » n'ont pas été représentées. Il est

cependant important de les analyser pour voir si aucune n'est en contradiction avec les relations représentées.

L'analyse montre qu'il y a effectivement des contradictions de ce type :

- Dans le module « Voie de signalisation JAK/STAT » (Annexe 6), phosphoJAK1 et 2 activent la phosphorylation d'Akt, on obtient également qu'elles sont sans effet sur ces mêmes réactions (cette contradiction a été discutée dans la démonstration de la traçabilité des données) ;
- Dans le module « Voie de signalisation JAK/STAT » (Annexe 6), PYK2 active les deux phosphorylations de STAT3, on obtient également qu'elle est sans effet sur ces mêmes réactions.

Comme dans le modèle FSH, différentes hypothèses contradictoires sont générées telles que l'EGF induirait l'activation de la PI3K et l'EGF inhiberait cette même relation.

On trouve également des contradictions entre des relations confirmées représentées dans le réseau et des hypothèses « *noeffect* » :

- Dans le module « Les actions impliquant les sous-unités Gαi » (Annexe 6), les complexes Gαi1-GTP et Gαi3-GTP activent les phosphorylations de GSK3β et Akt, on obtient également l'hypothèse qu'elle est sans effet sur ces mêmes réactions ;
- Dans le module « Les actions impliquant les sous-unités Gαi » (Annexe 6), Gαi1 et Gαi3 activent les phosphorylations de GSK3β et Akt, on obtient également l'hypothèse qu'elle est sans effet sur ces mêmes réactions.

La section « Les leçons de ces premiers tests » permet de discuter des différents types de contradictions que nous avons obtenus.

IV- Les leçons de ces premiers tests

Les résultats obtenus avec les réseaux FSH et EGF nous permettent d'évaluer la qualité de notre nouvelle méthode d'inférence. Les résultats ont permis de mettre en évidence l'existence de différences entre les réseaux de référence et les réseaux obtenus grâce à notre méthode. D'une part, certaines relations sont absentes dans les réseaux inférés automatiquement. D'autre part, nous avons à plusieurs reprises constaté que des relations, confirmées ou hypothétiques, sont contradictoires et que leur génération altère l'intégrité des réseaux obtenus. Nous allons expliquer les raisons de ces différents phénomènes qui sont communs aux deux réseaux inférés automatiquement.

A-Pourquoi manque-t'il des relations dans les réseaux obtenus ?

L'absence de certaines relations dans les réseaux FSH et EGF obtenus est essentiellement due au fait que les règles existantes ne permettent pas encore de raisonner sur tous les types d'expériences qui sont rencontrés dans l'étude de ces réseaux.

Pour illustrer cela nous allons prendre l'exemple de la voie Gs du modèle FSH car le résultat présente deux exemples distincts. Il est possible de voir sur la Figure 65 que la plupart des relations attendues sont obtenues. Néanmoins, il manque deux relations :

- L'activation indirecte de la PDE par la FSH,
- La formation du complexe Adénylate Cyclase:Gαs:GTP à partir de l'Adénylate cyclase et du complexe Gαs:GTP.

Pour déterminer l'activité des enzymes PDE, un *Enzymatic Assay* est utilisé : les cellules sont stimulées avec de la FSH puis les PDE sont extraites afin de quantifier leur activité. Or, nous avons formalisé ce type d'expérience sans stimulation, c'est pourquoi pour obtenir la relation entre la stimulation et l'activité de l'enzyme d'intérêt en réponse au signal, il faut créer un prédicat SEA(w,x,[(n_y,y)],[(n_z,z)],e) où w est le signal, x est le catalyseur, y est le substrat, z est le produit et e est l'effet observé. Ce prédicat signifie qu'en présence de w, on observe, par *Enzymatic Assay*, une augmentation de la production de z. Si de plus on sait que x catalyse de manière directe la conversion de y en z, cette augmentation ne peut provenir que de l'augmentation de la quantité de x, qui elle-même provient de l'augmentation de la conversion

de c (forme inactive de x) en x, où c est une forme modifiée de x qui ne possède pas cette activité enzymatique. Ainsi, on peut écrire la règle :

• IF $CATALYZE(x,[(n_y,y)],[(n_z,z)],direct,confirmed,increase)$ AND $SEA(w,x,[(n_y,y)],[(n_z,z)],increase)$ AND MODIFIEDFORM(x,c) THEN MODIFY(w,c,x,unknown,confirmed,increase) AND $CATALYZE(c,[(n_y,y)],[(n_z,z)],direct,confirmed,noeffect)$

Cette nouvelle règle permettra d'obtenir les liens manquants entre l'activation d'une enzyme et le signal qui a permis cette activation.

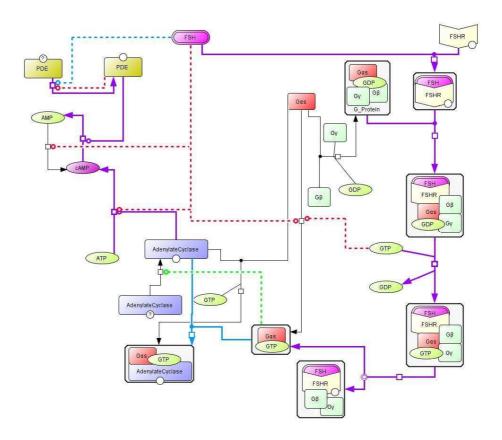


Figure 65 : Exemple de relations manquantes

Les relations attendues et présentes sont représentées en mauve et les relations attendues mais absentes sont représentées en bleu.

En ce qui concerne la formation du complexe Adénylate Cyclase:Gαs:GTP, cette complexation est démontrée par une structure cristalline tridimensionnelle, cela ne donne donc pas d'indication sur la forme sous laquelle les partenaires du complexe se présentent avant sa formation.

Nous avons également évoqué plus tôt l'action d'EPAC sur Rap1 et l'absence d'expériences permettant de définir le rôle de Rap1 dans le réseau. La protéine Rap1 est l'effecteur direct

d'EPAC [40], il serait donc intéressant de créer une règle permettant de déduire des hypothèses quant à l'action de Rap1. Par exemple :

• **IF** MODIFY(x,y,b,direct,confirmed,increase) **AND** MODIFY(x,z,c,unknown,confirmed,increase) **THEN** MODIFY(b,z,c,unknown,hypothesis,increase) signifiant: si *x* modifie *y* en *b* de manière directe, et si *x* modifie *z* en *c*, alors l'hypothèse que *b* modifie *z* en *c* peut être émise (Figure 66).

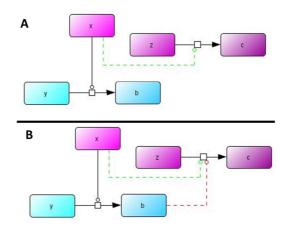


Figure 66 : Déduction d'une nouvelle hypothèse

(A) La molécule x modifie directement y en b et x modifie z en c avec une distance inconnue (B) A partir des relations précédentes, on déduit l'hypothèse que b induit la modification de z en c avec une distance inconnue

Grâce à cette règle, il serait possible d'obtenir des hypothèses supplémentaires et donc de compléter le réseau.

B- Le problème des ARN interférents

Etant donné qu'un petit ARN interférent (siRNA) diminue la présence de toutes les formes d'une même protéine dans la cellule touchée, il faudrait écrire la règle « IF SIRNAAGAINST(s,x) AND MODIFIEDFORM(y,x) THEN SIRNAAGAINST(s,y) ». Cependant, cette règle va nous poser un certain nombre de problèmes. Par exemple, nous obtiendrons des relations supplémentaires dans la voie β-arrestines du réseau FSH et dans la voie Grb2 du réseau EGF. Dans le modèle FSH, toutes les régulations hypothétiques des différentes GRK et des β-arrestines sur la phosphorylation des ERK MAPK seront transférées à toutes leurs formes. Par exemple, le résultat de l'inférence contiendra en plus les hypothèses selon lesquelles GRK2 modifiée inhibe la phosphorylation PHOSPHORYLATE(modGRK2,ERK1,pERK1,unknown,hypothesis,decrease) et l'hypothèse GRK2 n'a pas d'effet selon la sur cette même phosphorylation: PHOSPHORYLATE(GRK2,ERK1,pERK1,unknown,hypothesis,noeffect). Sachant que nous avons déjà dans les faits déduits l'hypothèse selon laquelle GRK2 inhibe cette phosphorylation: PHOSPHORYLATE(GRK2,ERK1,pERK1,unknown,hypothesis, decrease) et l'hypothèse selon laquelle la GRK2 modifiée n'a pas d'effet dessus: PHOSPHORYLATE(modGRK2,ERK1,pERK1,unknown,hypothesis,noeffect). Ainsi, les hypothèses générées seront contradictoires. Cependant, cette situation n'est pas problématique dans le sens où ce ne sont que des hypothèses. L'apport d'une expérience de validation supplémentaire pourra ensuite permettre d'éliminer des hypothèses. Par contre, cela soulève le problème de la multiplication des hypothèses qui pourrait rapidement mener à une explosion du nombre de relations inférées.

C- Utilisation du savoir implicite

Dans le réseau de signalisation induit par la FSH obtenu, nous avons utilisé le savoir implicite pour obtenir toutes les relations mettant en jeu la protéine G. Il est important de préciser que l'utilisation des « on sait que » est faite uniquement dans les cas où les relations ont été démontrées un grand nombre de fois. Dans le cas de la protéine Gs, différents travaux ont montré sa complexation avec le récepteur [79, 223], de plus l'action de la protéine Gs activée sur l'adénylate cyclase a été démontrée en présence de FSH [224]. Ces travaux prouvent le rôle de la protéine Gs dans le réseau induit par la FSH mais ils sont issus d'expériences de Gel Diffusion, or cette expérience est très complexe à formaliser, c'est pourquoi nous avons choisi d'admettre cette complexation.

Il est très important de limiter le nombre de faits admis. En effet, l'une de nos motivations dans le développement de la méthode était de limiter les biais dus à l'utilisation du savoir du domaine, notamment parce qu'une partie des processus admis par la communauté de la signalisation cellulaire ne repose pas sur des preuves indiscutables mais sur des faisceaux d'éléments convergeants. Or ce genre de dogme est régulièrement invalidé dans le domaine de la biologie.

L'autre phénomène auquel nous avons été confrontés est que certains processus ont été démontrés il y a plusieurs dizaines d'années. Or, les protocoles expérimentaux utilisés, en particulier dans les premiers temps de la biologie moléculaire, étaient souvent très indirects par rapport au phénomène à mettre en évidence, et il est souvent difficile de formaliser le raisonnement. Enfin, ces protocoles ont le plus souvent été abandonnés depuis au profit de protocoles plus directs. Pour toutes ces raisons, nous n'envisageons pas de créer les règles qui permettraient d'intégrer ces protocoles à la méthode.

D- Les faits contradictoires

Dans l'analyse des réseaux inférés nous avons signalé plusieurs contradictions, soit entre des faits sans effet et des faits présents dans le graphe, soit entre des hypothèses.

1- Hypothèses contradictoires

Dans l'analyse des deux réseaux inférés nous avons signalé des cas où des hypothèses contradictoires sont générées.

Le réseau inféré pour la FSH contient les hypothèses que la FSH active et inhibe la réaction PKA-> PKAc (Figure 67).

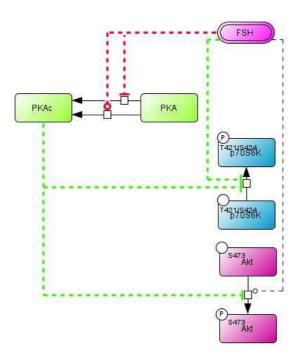


Figure 67 : Incompatibilité des hypothèses, l'exemple de la PKA

Trait pointillé rouge = relation hypothétique avec une distance inconnue, trait pointillé vert = relation confirmée avec une distance inconnue, trait pointillé noir = relation confirmée indirecte.

Ce petit réseau est obtenu grâce à différentes expériences de *Phosphorylation Assay* de p70S6Kinase et d'Akt en présence de FSH et d'inhibiteurs. Dans un premier temps, nous avons obtenu les conclusions :

- La FSH induit indirectement la phosphorylation d'Akt sur la position S473 [225],
- La FSH inhibe indirectement la phosphorylation de p70S6Kinase sur le site T421/S424 [209].

Ensuite, les mêmes expériences ont été reproduites en présence d'un inhibiteur de la PKA active. Cet inhibiteur a entraîné une augmentation de la quantité d'Akt phosphorylée (fait ICPPA(FSH,Akt,a_pAktS473,H89,increase) [225] et une augmentation de p70S6Kinase phosphorylé (fait ICPPA(FSH,p70S6K,MyrPKI,a_pp70S6KT421/S424,increase) [209].

Le fait impliquant Akt a entraîné le déclenchement de la règle :

IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase)
 ICPPA(x,y,i,a,increase)
 AND INHIBITOR(i,z)
 AND ANTIBODYAGAINST(a,b)
 AND MODIFIEDFORM(z,d)
 AND NOTMODIFIED(x)
 THEN PHOSPHORYLATE(z,y,b,unknown,confirmed,decrease)
 AND MODIFY(x,d,z,unknown,hypothesis,decrease)
 PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)

Ce qui a permis d'obtenir les conclusions :

- PHOSPHORYLATE(PKAc,Akt,pAktS473,unknown,confirmed,decrease)
- MODIFY(FSH,PKA,PKAc,unknown,hypothesis,decrease)
- PHOSPHORYLATE(PKA,Akt,pAktS473,unknown,confirmed,noeffect)

Et le fait impliquant p70S6Kinase a entraîné le déclenchement de la règle :

• IF PHOSPHORYLATE(x,y,b,unknown,confirmed,decrease) AND ICPPA(x,y,i,a,increase) AND INHIBITOR(i,z) AND ANTIBODYAGAINST(a,b) AND NOTMODIFIED(x) AND MODIFIEDFORM(z,d) THEN PHOSPHORYLATE(z,y,b,unknown,confirmed,decrease) AND MODIFY(x,d,z,unknown,hypothesis,increase) AND PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)

Ce qui a permis d'obtenir les conclusions :

- PHOSPHORYLATE(PKAc,pp70S6K,pp70S6KT421/S424,unknown,confirmed,decre ase)
- MODIFY(FSH,PKA,PKAc,unknown,hypothesis,increase)
- PHOSPHORYLATE(PKA,pp70S6K,pp70S6KT421/S424,unknown,confirmed, noeffect)

Ceci conduit donc à une contradiction hypothétique sur l'effet de la FSH sur PKA.

Toutes les hypothèses contradictoires qui ont été obtenues dans nos deux modèles sont générées de la même façon que les hypothèses générées concernant la PKA. Ces contradictions entre hypothèses ne posent cependant pas de problème d'intégrité des données puisque ce sont des hypothèses. Pour résoudre ces contradictions il est nécessaire d'apporter des expériences supplémentaires. Ceci peut être très important pour les utilisateurs puisque

cela leur permettra de mettre en évidence les parties du réseau dans lesquelles les preuves expérimentales sont encore insuffisantes.

2- Contradictions issues de l'interprétation des expériences

Dans le cadre de ce travail, la lecture des résultats expérimentaux est d'une importance capitale car c'est de cette étape que dépend le réseau obtenu, cependant elle n'est pas toujours aisée. Afin de valider cette étape dans le cas du réseau induit par la FSH, nous avons fait participer une promotion de Master 2 (Bioinformatique et Biostatistiques) de l'Université de Paris Sud. Dans le cadre de leur unité d'enseignement « Intégration et Analyse de Données Biologiques issues du Web », nous avons expliqué aux étudiants la procédure d'extraction des données biologiques à partir des publications, puis nous leur avons donné différentes publications ainsi qu'un dictionnaire des données afin qu'ils puissent extraire les informations et les mettre dans le format que nous avons édité. Au total, il y avait huit binômes ayant chacun trois ou quatre publications (en fonction du nombre total d'expériences) représentant un total de 11 publications : une des publications était commune à tous les binômes, les autres ont été distribuées de sorte que chaque binôme ait une combinaison d'articles différente des autres, et que chaque article soit vu par au moins trois binômes. L'objectif était de voir si la lecture des résultats expérimentaux variait en fonction du lecteur.

Cette expérience a permis de constater que dans les cas où les expériences montrent une variation importante entre la situation contrôle (sans FSH) et en présence de FSH, les lectures convergent. Au contraire, lorsque les résultats montrent une faible variation, alors les lectures varient en fonction de la personne. Nous avons donc obtenu un grand nombre de réponses identiques aux nôtres. Il y avait cependant une expérience qui a donné lieu à des lectures différentes. C'est le cas de l'expérience permettant d'établir une relation entre la PKA et la p38MAPK.

Lorsque nous avons analysé le réseau FSH inféré, nous avons mis en évidence la contradiction suivante : PKAc active et n'a pas d'effet sur la phosphorylation de p38MAPK, les deux faits étant confirmés.

La première conclusion est obtenue à partir d'une expérience de *Phosphorylation Assay* réalisée par Yu et al [204] pour détecter p38MAPK phosphorylée en présence de l'inhibiteur de la PKA, H89 (Figure 68, les bandes utilisées sont encadrées en rouge).

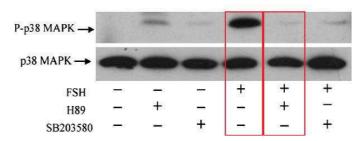


Figure 68 : Phosphorylation Assay de p38MAPK de Yu et al [204]

La lecture de ce résultat ne présente aucune difficulté, l'ajout de l'inhibiteur de la PKA activée, H89, induit une diminution très prononcée de la phosphorylation du site T180/Y182 de la p38MAPK.

La deuxième conclusion provient des travaux de Wayne et al [77]. L'expérience de *Phosphorylation Assay* a été cette fois effectuée en présence de l'inhibiteur PKI (Figure 69, les spots encadrés en rouge sont ceux qui sont utilisés).

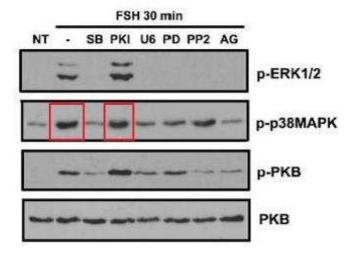


Figure 69: Phosphorylation Assay de p38MAPK de Wayne et al [77]

Cette expérience est plus difficile à lire que la précédente. Trois des binômes de Master avaient cet article. Deux binômes ont considéré que les deux bandes étaient de même intensité, ce qui les a conduits à conclure que la PKA n'a pas d'action sur la phosphorylation. Le troisième binôme a lu une diminution de la phosphorylation de la p38MAPK ce qui les a conduits à conclure que PKA active la phosphorylation.

Ceci montre bien que la lecture des résultats expérimentaux peut varier d'une personne à l'autre surtout lorsque qu'il s'agit d'évaluer une faible différence d'intensité entre deux spots. Ceci montre que si la lecture est difficile, il peut être utile de demander plusieurs avis et

surtout prendre en compte les statistiques présentées si elles sont disponibles. Dans cet exemple, il y a seulement 10% de différence mais pour en être sûr, il faudrait plusieurs expériences de quantification. Ceci montre également que si la lecture est difficile et qu'on dispose d'autres expériences avec des résultats plus tranchés pour la même relation, alors il vaut mieux éliminer la première.

3- Problème des conditions expérimentales

En biologie, ce que l'on appelle les « conditions expérimentales » recouvre une réalité très complexe. Il peut s'agir de points majeurs tels que le type cellulaire ou le stade de développement. Cependant, d'autres points plus subtils peuvent également s'avérer déterminants, comme par exemple la durée de sevrage avant stimulation. Dans le cas du réseau de signalisation de la FSH, il existe par exemple des différences importantes entre les cellules de Sertoli et les cellules de la Granulosa. Dans notre revue, nous avons mélangé les données provenant de ces deux types cellulaires, ainsi que des résultats obtenus dans des lignées. Cependant, dans la base de données, le type cellulaire est attaché à l'expérience, et il est donc tout à fait envisageable de réaliser l'inférence à partir de données provenant d'un unique type cellulaire. Nous nous retrouvons néanmoins rapidement confrontés au problème que chaque relation n'a pas été démontrée dans chaque type cellulaire.

La prise en considération de la molécule utilisée pour stimuler les cellules est également très importante. Par exemple, Gonzalez-Robayna et al [128] obtiennent une forte augmentation de la phosphorylation de la p38MAPK lors de l'ajout de H89, inhibiteur de la PKA active, alors que cette inhibition provoque une diminution dans d'autres expériences. Ceci est du au fait que les auteurs utilisent de la FSH associée à de la testostérone, or il est établi que la testostérone seule est capable d'induire la phosphorylation de différentes MAP kinases, dont p38MAPK [226].

La qualité de la stimulation et le temps de mesure sont également des facteurs importants. Une expérience dans laquelle la stimulation est une protéine recombinante peut donner des résultats différents d'une expérience ayant pour stimulation une molécule purifiée car la purification n'est jamais parfaite. En ce qui concerne les temps de mesure, nous l'avons déjà évoqué, le fait qu'une relation puisse exister à un temps i et ne plus exister à un temps i+1. Il serait intéressant de prendre note des temps auxquels les mesures servant à créer les faits sont

réalisées. Néanmoins ceci introduirait un niveau de complexité que nous ne pouvons pas encore gérer.

Même si nous ne sommes pas toujours en mesure de prendre en compte l'ensemble des facteurs, il nous a semblé primordial de les stocker avec l'expérience elle-même dans la base de données. D'une part, cela permet de s'y référer lorsque des contradictions apparaissent dans le réseau pour tenter de les résoudre. D'autre part, la méthode d'inférence est appelée à évoluer, et prendra en compte un nombre croissant de facteurs.

V- Les limites de la formalisation

La formalisation du raisonnement et des données expérimentales est une phase complexe et certaines étapes posent des problèmes difficiles.

A- Les formes modifiées

En biologie, on parle généralement de « forme active » et « forme inactive » d'une protéine. Ce changement provient nécessairement d'une modification, qu'elle soit chimique (modification post-traductionnelle), ou conformationnelle, lorsque la protéine se lie à un complexe, ou encore lors d'un changement de compartiment cellulaire.

Nous avons choisi de représenter cela par le prédicat MODIFIEDFORM(x,y), qui signifie que x est une forme modifiée de y. Cette notation permet de formaliser tous les types de modifications sans qu'il soit nécessaire de les définir précisément.

Les formes modifiées ne présentent pas de problème en elles-mêmes dans le raisonnement. L'inconvénient provient du fait que dans une règle, chaque élément présent dans la conclusion doit être présent dans la condition.

Par exemple, il a été montré que la FSH induit indirectement la phosphorylation des protéines ERK MAPK, et que cet effet est absent lorsqu'on ajoute un inhibiteur des MEK. Ainsi on peut faire l'hypothèse que la FSH active les MEK, qui à leur tour phosphorylent les ERK MAPK. Par cette phrase, on dit implicitement qu'il existe une forme inactive des MEK et une forme non phosphorylée des ERK. Pour que les règles qui aboutissent à la bonne conclusion puissent être déclenchées, il est nécessaire de définir cette forme inactive.

La situation devient plus problématique lorsque plusieurs formes modifiées différentes et inconnues sont nécessaires pour déclencher toutes les règles nécessaires à l'obtention du résultat correct, car il faut définir dès le début un « réservoir » suffisant de formes modifiées. De plus, un certain nombre de règles va être déclenché par chacune de ces formes hypothétiques. Ce raisonnement nous a fait craindre une explosion du nombre d'hypothèses générées, mais cela n'a pas été le cas jusqu'à présent. C'est cependant une réflexion sur laquelle il faudra revenir.

Il existe également des molécules qui ne peuvent pas être modifiées dans le cadre d'un réseau de signalisation. C'est par exemple le cas de la FSH dans le cadre du réseau de signalisation : on sait qu'elle ne peut pas être modifiée par la signalisation qu'elle induit. Pour ce type de cas particulier, nous avons créé le prédicat NOTMODIFIED(x). Cela permet d'éliminer d'emblée un certain nombre d'hypothèses qui n'auraient pas de sens biologique.

B- Le problème des hypothèses

L'inférence automatique génère un grand nombre d'hypothèses, beaucoup plus que l'expert à partir des mêmes données. Comme indiqué plus haut, la présence de nombreuses formes modifiées par exemple entraîne la création de nombreuses hypothèses. Par exemple, si une expérience d'ELISA est réalisée pour quantifier l'AMPc en présence de FSH et qu'il existe deux formes modifiées de l'AMPc : l'AMP et l'ATP, le réseau obtenu correspond à la Figure 70. Or il s'avère que l'AMPc est produit à partir de l'ATP et non de l'AMP.

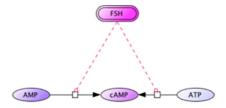


Figure 70 : Hypothèses obtenues à partir d'un ELISA sur la production d'AMPc en présence de FSH

Prenons aussi l'exemple d'une expérience de *Phosphorylation Assay* en présence d'un inhibiteur. On sait que *x* active la modification de *y* en *b*. On ajoute un inhibiteur d'une molécule *z* et on constate une diminution de *b*. Si on n'a défini qu'une seule forme modifiée pour chacune des molécules impliquées dans l'expérience, trois réseaux sont compatibles avec ce résultat (Figure 71), et les hypothèses correspondant à ces trois réseaux seront générées automatiquement.

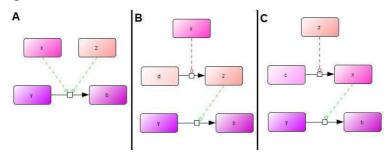


Figure 71 : Différents réseaux pouvant conduire au même résultat

Dans le cas où on a défini un plus grand nombre de formes modifiées pour les différentes molécules impliquées, le nombre de solutions augmente.

Nous avons choisi de faire les règles les plus complètes possibles, c'est-à-dire que le déclenchement d'une règle va générer toutes les hypothèses qui peuvent mener au résultat observé, cela sans prendre en considération la probabilité de chaque hypothèse au regard de la biologie. Dans l'exemple précédent, cela se traduit par le réseau de la Figure 72.

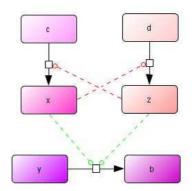


Figure 72 : Fusion des trois réseaux pouvant être obtenus à partir d'un *Phosphorylation Assay* avec inhibiteur

Ceci est très important et est certainement une des forces de la méthode. En effet, trop souvent en biologie, lorsque deux hypothèses sont possibles, et qu'au regard de ce que l'on sait déjà sur le système l'une est nettement plus plausible que l'autre, on va simplement éliminer l'hypothèse la moins probable. La raison essentielle est que les systèmes deviennent rapidement très complexes, et donc il est impossible d'envisager toutes les hypothèses de manière simultanée. De ce fait, dans un système complexe on finit par avoir un échafaudage d'hypothèses qui sont chacune localement la plus probable, mais l'ensemble a quant à lui une probabilité assez faible d'être juste. Notre système permet justement de n'écarter aucune hypothèse tant qu'il n'a pas été démontré qu'elle était fausse. De plus, la représentation de toutes ces hypothèses dans un même graphe permet de se concentrer sur les points où des données supplémentaires sont nécessaires pour éliminer des hypothèses concurrentes et incompatibles entre elles, en faisant ainsi un outil pour rationaliser le travail expérimental.

C- Des règles de plus en plus complexes

Les règles actuelles ne permettent d'interpréter qu'un nombre limité de protocoles expérimentaux. En particulier, nous nous sommes limités au cas où au plus un perturbateur est

utilisé. Or il est très fréquent d'utiliser plusieurs perturbateurs. Si on reprend l'exemple précédent, mais en introduisant cette fois des inhibiteurs de deux molécules z et w, le nombre de réseaux qui mènent à l'observation d'une baisse de la quantité de y en présence de ces inhibiteurs est beaucoup plus élevé (Figure 73).

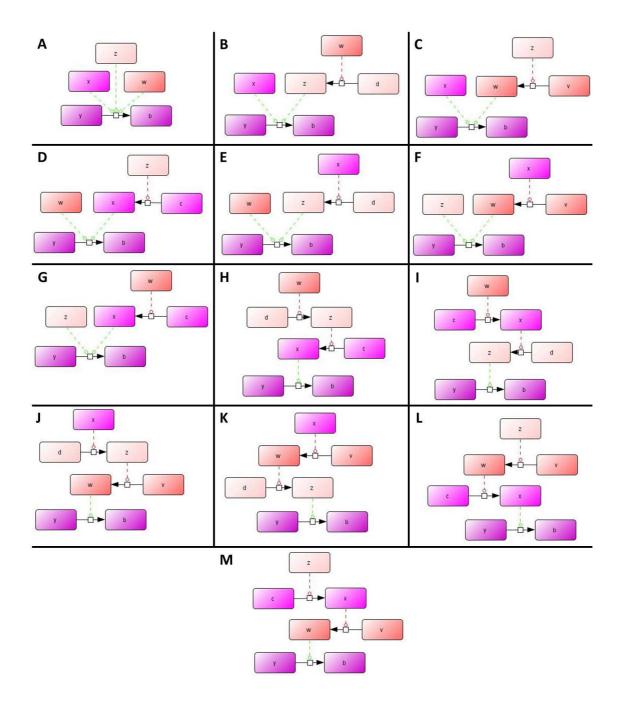


Figure 73 : Possibilités de réseaux provenant d'un *Phosphorylation Assay* avec une double perturbation

(Relation rouge = hypothèse avec une distance inconnue, relation verte = confirmée avec une distance inconnue)

Les différentes hypothèses peuvent également être représentées par le graphe de la Figure 74.

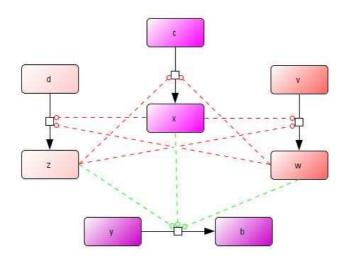


Figure 74 : Fusion des réseaux pouvant être obtenus à partir d'un *Phosphorylation Assay* avec une double perturbation

(Relation rouge = hypothèse avec une distance inconnue, relation verte = confirmée avec une distance inconnue)

Ceci va incontestablement entraîner la création de règles de plus en plus complexes et ainsi augmenter la complexité des réseaux obtenus.

D- Difficultés pour la formalisation

Certains points sont apparus comme particulièrement difficiles à formaliser. En particulier dans les protocoles utilisant des molécules très particulières comme les agonistes, les analogues structuraux, les mutants ou les variants naturels.

1- Agonistes et analogues structuraux

On distingue deux types d'agonistes : les agonistes complets, qui sont capables de reproduire exactement tous les effets de la molécule dont ils sont agonistes, et les agonistes biaisés, qui reproduisent une partie de ces effets. Les agonistes complets (AGONIST(x,y) : x est un agoniste de y) ne posent aucune difficulté particulière. Des règles ont été créées pour « transférer » toutes les propriétés de l'agoniste à la molécule naturelle. Il en est de même pour les analogues structuraux des molécules.

Pour les agonistes biaisés et les analogues structuraux mimant une partie des effets d'une molécule, les choses sont beaucoup plus complexes puisque l'on ne peut pas transférer l'ensemble des propriétés de la même manière.

Par exemple, lors de la construction du réseau FSH, un analogue structural important de l'AMPc a été rencontré, le 8-CPT-2ME-cAMP. Il est capable d'activer EPAC mais pas la PKA. Cette molécule permet l'identification de la voie de signalisation dépendante d'EPAC.

Pour le moment, nous avons assimilé 8-CPT-2ME-cAMP à un activateur d'EPAC de façon à pouvoir utiliser les informations, mais il est nécessaire de trouver une alternative pour formaliser l'utilisation de cette molécule comme étant un analogue structural de l'AMPc mais avec un effet partiel.

De la même façon, il est difficile de formaliser les agonistes biaisés car ces derniers entraînent un changement de conformation du récepteur différent de celui induit par le ligand naturel. Par exemple, le récepteur FSHR peut être activé par la LH équine partiellement déglycosylée et dans ce cas, le récepteur induit préférentiellement la voie dépendante des β-arrestines par rapport à l'action de la FSH [102]. Ce qui signifie que cette LH est un agoniste biaisé de la FSH et sa fixation au récepteur déclenche un changement conformationnel du récepteur pour l'activer et cette conformation active est différente de celle induite par la FSH.

2- Mutants et variants naturels

En biologie, l'utilisation de protéines mutantes est très répandue. Les mutants permettent d'étudier une caractéristique précise d'une molécule d'intérêt pour déterminer la fonction qui est due à cette caractéristique. Par exemple, si une hypothèse est que la phosphorylation d'une protéine sur un site donné permet son activation, on peut créer un mutant à cette position, soit mimant une phosphorylation constitutive, soit mimant une déphosphorylation constitutive, et utiliser ce mutant pour démontrer l'hypothèse.

Les difficultés à formaliser les mutants viennent justement du fait qu'il y a une « intention » derrière la création d'un mutant, et que c'est cette intention qui doit être formalisée.

D'autre part, ces mutants peuvent avoir des conséquences similaires à celles des analogues structuraux partiels, puisqu'ils sont capables de remplir une partie seulement des fonctions de la protéine sauvage. Prenons l'exemple du récepteur FSHR, lorsqu'il est activé par la fixation de la FSH, il est phosphorylé par les GRK2/3 ou GRK5/6, puis il se complexe avec les β -arrestines. En fonction de ces phosphorylations, il sera dirigé vers la désensibilisation ou vers la signalisation dépendante des β -arrestines. Dans le cas où cinq sérines sont remplacées par des alanines, ce récepteur présente une phosphorylation très diminuée et des expériences de co-immunoprécipitation montrent que cette diminution de phosphorylation du récepteur entraîne une forte diminution de la complexation avec les β -arrestines [99]. Cependant, on constate toujours à la fois un recyclage du récepteur et une signalisation dépendante des β -arrestines. Ainsi, la mutation du FSHR n'entraîne ni l'activation constitutive, ni l'inactivation constitutive de ce récepteur qui peut donc bien être défini comme un analogue structural

partiel du récepteur sauvage. On revient donc aux problèmes de formalisation évoqués dans le chapitre précédent.

Au sein des cellules, il peut aussi exister des mutants naturels des protéines, c'est ce que l'on appelle les variants naturels. Ces derniers peuvent induire des différences dans la signalisation cellulaire. Une étude menée sur un mutant du récepteur FSH (Alanine 189 remplacée par une Valine) montre que la cascade ERK MAPK est induite préférentiellement par la voie dépendante des β -arrestines plutôt que par la voie dépendante des protéines G [227]. Les résultats ont montré que cette mutation entraîne un défaut d'adressage du récepteur à la membrane plasmique et que lorsque le récepteur FSHR naturel est exprimé aussi faiblement que le mutant A189V alors il se comporte de la même façon. Ce type d'exemple montre la difficulté de formalisation liée à ces variants naturels car les différences sont essentiellement quantitatives, or actuellement, notre méthode ne prend pas en compte les données quantitatives.

VI- Inférence

A- Utilisation du 'OU' et problème d'égalité

1- Le 'OU'

Comme nous l'avons déjà signalé, nous avons choisi de ne pas utiliser le OU dans les conclusions des règles. Un certain nombre de raisonneurs dédiés à la programmation logique, ne sont pas en mesure d'utiliser des règles contenant en conclusion un OU, mais SOLAR le peut. Or, si dans un certain nombre de cas déjà cités, les hypothèses pourraient être remplacées par l'ajout d'un OU, la règle pourrait devenir fausse car il est possible de rencontrer les différents cas de figures :

- Une des hypothèses est validée et l'autre invalidée, dans ce cas, l'utilisation du 'OU' n'est pas problématique,
- Les deux hypothèses sont validées ce qui implique que la présence d'un 'OU' rendrait la règle fausse,
- Les deux hypothèses sont invalidées, ce qui conduit également à une règle erronée.

Par exemple, lorsqu'un *Western Blot* est effectué avec un anticorps Pan, le but est de quantifier une protéine et donc son expression dans la cellule. Comme nous l'avons expliqué dans le Matériel et Méthodes, si une augmentation de la quantité de la protéine d'intérêt y est observée en présence de la molécule x alors il est possible d'en déduire les quatre hypothèses : x active la transcription du gène v en ARNm w OU x inhibe la traduction de l'ARNm w en protéine y OU x inhibe la dégradation de y (Figure 75).

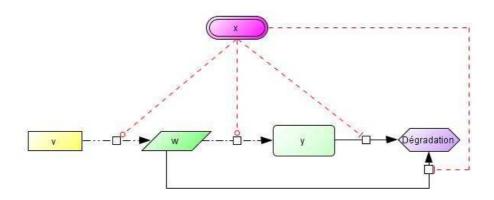


Figure 75 : Schématisation des conclusions obtenues à partir d'un PANWB

Cependant, il est tout à fait possible que *x* agisse simultanément sur deux de ces phénomènes, voire les quatre. Ainsi dans ce cas, deux, trois ou quatre hypothèses sont vraies simultanément, et on ne peut pas remplacer les règles concluant sur ces différentes hypothèses par une règle avec un OU.

2- L'égalité

En logique du premier ordre, il est possible d'introduire le symbole d'égalité. Cela est formalisé grâce à un prédicat d'égalité permettant de lier des termes. Ce type de prédicat serait nécessaire notamment pour définir l'égalité entre deux complexes telle que : $(n_x)x:(n_y)y$ = $(n_y)y:(n_x)x$.

Cependant, cette égalité n'est pas implémentée dans SOLAR mais les développeurs de ce moteur d'inférence envisagent d'intégrer cette possibilité.

B- Temps de calcul

Le Tableau 6 contient les temps d'exécution de SOLAR observés pour obtenir la totalité des faits déduits pour les modèles FSH et EGF avec une profondeur de raisonnement limité à l'enchaînement de quatre règles (profondeur optimale pour la déduction de tous les faits).

Tableau 6 : Temps d'exécution de SOLAR avec un enchaînement de quatre règles maximum

Modèle PF	FSH	EGF
MODIFY	1991s	3316s
PHOSPHORYLATE	2876s	3326s
TCE	17,25s	6,78s
TLE	13,81s	6,65s
Complexation	4,25s	51,61s
Réaction enzymatique	3,81s	3,36s
Temps total	1h18	1h48

Les différents temps obtenus permettent d'affirmer que SOLAR ne permet pas une déduction rapide de tous les faits.

Cependant, Adrien Rougny a testé SOLAR avec une méthode différente. Pour ce nouveau test, la profondeur de raisonnement choisie correspond à une seule règle (et non un enchaînement de quatre règles comme précédemment) et le PF contient tous les prédicats. Les faits initiaux ont servi de *top_clause* et les règles sont considérées comme des axiomes afin d'obtenir les faits déduits. Ensuite, à chaque itération, les faits déduits remplacent les faits initiaux en *top_clause*. Les faits initiaux utilisés lors d'une itération sont considérés comme des axiomes à l'itération suivante. De cette façon SOLAR est exécuté jusqu'à ce qu'il n'y ait plus aucun fait nouveau déduit.

Cette deuxième technique est beaucoup plus rapide que la première car elle permet d'obtenir tous les faits déduits des modèles FSH et EGF en environ une minute. Cette nouvelle technique de raisonnement est valable car toutes les clauses manipulées par SOLAR pour obtenir les relations d'un réseau sont issues de règles de Horn.

C- Indice de confiance des données

Comme nous l'avons discuté à plusieurs reprises, des contradictions vont apparaître dans le réseau chaque fois que des expériences dont les résultats sont contradictoires vont être utilisées. Ces contradictions représentent potentiellement un outil intéressant puisque qu'il est possible de faire ressortir des contradictions entre des expériences utilisant des protocoles différents et effectuées à différents niveaux du réseau, qui pourraient sans cela, passer inaperçu. Résoudre ces contradictions demande cependant de revenir aux expériences initiales et de « décider » d'écarter certaines d'entre elles.

Etant donné que cette situation est relativement fréquente en biologie, nous avons commencé à réfléchir à des moyens de guider l'utilisateur dans ce choix en donnant un « indice de confiance » aux relations du réseau.

Plusieurs facteurs nous semblent particulièrement pertinents :

- L'estimation de la fiabilité des protocoles expérimentaux,
- Le nombre de faits initiaux menant de manière indépendante à cette conclusion,
- Le nombre d'équipes de recherche dont proviennent ces faits.

Par exemple, lorsqu'il s'agit de montrer une interaction entre deux molécules, la cristallographie est plus fiable qu'une expérience d'immunoprécipitation. L'indice de confiance d'une relation de complexation déduite d'une expérience de cristallographie sera donc plus élevé que celui d'une relation de complexation déduite d'une expérience d'immunoprécipitation.

De plus, l'indice de confiance pourrait augmenter avec le nombre d'expériences indépendantes démontrant la même relation. Ainsi, si une relation de complexation est démontrée par immunoprécipitation, l'indice de confiance sera plus faible que celui d'une relation de complexation démontrée par immunoprécipitation et FRET. La cristallographie quant à elle conduira toujours à un indice de confiance très élevé.

En ce qui concerne la prise en compte du nombre de faits initiaux menant à la conclusion à évaluer, la procédure à suivre est plus complexe. Doit-on compter +1 à chaque fois que la conclusion est rencontrée ? D'autre part, il faut être très vigilant quant à la source des expériences. En effet, il est fréquent de voir apparaître une même expérience dans plusieurs articles publiés par la même équipe. De plus, les habitudes et les pratiques de chaque équipe introduisent parfois des biais similaires d'une expérience à l'autre, et donc d'une publication à l'autre.

D- Règles d'élagage

SOLAR permet d'obtenir tous les faits déduits qui ont pu être générés à partir des faits initiaux. Or certaines conclusions ne sont pas utiles pour construire le réseau. Par exemple, les conclusions contenant un analogue structural (ou un agoniste) en argument ne sont pas représentées dans la visualisation du réseau car l'analogue structural (ou l'agoniste) ne fait pas partie du réseau de signalisation cellulaire, il permet uniquement d'obtenir des informations sur la molécule dont il est l'analogue structural (ou l'agoniste).

D'autres cas sont plus difficiles à traiter. Par exemple, dans l'inférence du réseau de la signalisation de la FSH, nous avons obtenu les conclusions suivantes :

- Raf1 phosphorylée sur la position S338 catalyse directement la phosphorylation des MEK,
- MEK phosphorylées catalysent directement la phosphorylation des ERK,
- Raf1 phosphorylée sur la position S338 induit indirectement la phosphorylation des ERK.

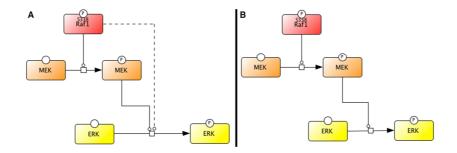


Figure 76: Exemple d'élagage possible

Dans la partie A, il existe une relation indirecte qui a été supprimée dans la partie B

On obtient le réseau représenté sur la Figure 76A. Dans ce réseau, on a deux chemins allant de Raf1 aux ERK, l'un constitué de réactions directes et passant par les MEK, l'autre indiquant une action indirecte de Raf1 sur les ERK. Dans ce cas, cette dernière relation n'apporte aucune information, et pour rendre le graphe plus clair on peut la supprimer pour obtenir le réseau de la Figure 76B.

Dans cet exemple, aucune information n'est perdue, le superflu a tout simplement été élagué dans le graphique.

Il est important de faire très attention car ce type d'élagage ne peut pas être réalisé systématiquement. C'est pourquoi, si cette information n'est plus représentée sur le graphe, il est important de ne pas la supprimer dans la base de données. Par exemple, à partir des conclusions suivantes :

- FSH agit sur la PKA,
- FSH agit sur les β-arrestines,
- La PKA induit indirectement la phosphorylation des ERK,
- Les β -arrestines induisent indirectement la phosphorylation des ERK,
- La FSH induit indirectement la phosphorylation des ERK.

On obtient le réseau de la Figure 77A. En absence d'information supplémentaire, le lien indirect entre FSH et la phosphorylation des ERK est inutile. Cependant, il existe effectivement une autre voie indirecte entre la FSH et la phosphorylation d'ERK, présentée sur la Figure 77B dans laquelle le complexe AMPc:EPAC est un intermédiaire.

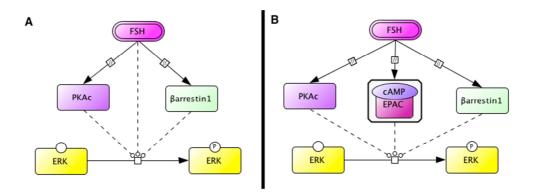


Figure 77 : Exemple d'élagage problématique

Le symbole carré contenant deux traits signifie que les étapes intermédiaires ont été omises.

Ces exemples font apparaître le fait que nous aurons besoin de créer des règles d'élagage, mais qui interviendront uniquement pour la visualisation, et pas dans le raisonnement qui permet d'inférer le réseau.

Conclusion et perspectives

Les travaux menés aux cours de ma thèse ont permis l'élaboration d'une nouvelle méthode d'inférence automatique des réseaux de signalisation directement à partir de données expérimentales massives et hétérogènes, ce qui représente une innovation importante dans le domaine de la biologie systémique. Les premiers modèles de réseaux induits par la FSH et l'EGF ont permis de démontrer la pertinence de la démarche et feront prochainement l'objet d'une publication.

Actuellement, seul le raisonnement permettant d'interpréter les expériences les plus fréquemment rencontrées dans l'étude de la signalisation induite par la FSH a été formalisé. La formalisation du raisonnement permettant l'interprétation d'autres types d'expériences est en cours, notamment pour l'exploitation des données à haut débit telles que les puces à ADN ou à protéines. Il sera alors possible d'exploiter la totalité des résultats de façon automatique, ce qui est actuellement impossible.

En ce qui concerne les représentations graphiques dans le langage CellDesigner, les conclusions obtenues dans le cadre des réseaux induits par la FSH et l'EGF ont été transformées manuellement en graphique de façon à vérifier que nous obtenions bien toutes les conclusions que nous attendions. L'automatisation de cette étape est en cours, conjointement avec la finalisation de la base de données dont elle dépend. Lorsque cette base sera prête, nous pourrons compléter le programme d'automatisation qui pourra ainsi prendre en compte toutes les relations entre les molécules d'un réseau donné. Il sera également possible d'intégrer la localisation cellulaire des molécules et donc de représenter graphiquement les différents compartiments cellulaires.

La base de données que nous développons sera riche en informations sur les réseaux qui y seront intégrés et surtout ces informations seront toutes reliées à une source minimum. Cette possibilité de remonter aux sources bibliographiques ou aux auteurs des expériences sera une caractéristique majeure de notre approche.

Nous espérons étendre la base de données, et donc l'accès à la méthode d'inférence, à d'autres réseaux de signalisation, comme par exemple les réseaux induits par la sérotonine à travers les RCPG 5HT₂ et 5HT₄ dans le cadre du projet ANR *GPCRNet* dans lequel notre équipe est impliquée. Nous voulons également appliquer ce procédé d'inférence à d'autres types de réseaux de moléculaires.

Sur le long terme, notre méthode permettra de construire les réseaux statiques qui serviront de points de départ pour la modélisation dynamique. Cet outil pourra également être très utile dans le *design* de protocoles expérimentaux. En effet, SOLAR permet de faire de l'abduction, c'est-à-dire trouver les règles ou les faits manquants pour qu'une conclusion prédéterminée puisse être déduite. Cela permettrait par exemple de déterminer l'ensemble des expérimentations qui doivent être menées pour démontrer qu'une molécule appartient à un réseau. Les capacités de déduction de l'outil peuvent également permettre de déterminer l'apport en information d'une expérience, en fonction des résultats possibles, avant même de réaliser l'expérience, et de déterminer ainsi si cet apport justifie de la réaliser. La méthode d'inférence peut ainsi devenir un outil d'aide à la décision pour le biologiste.

Aujourd'hui, d'autres développements méthodologiques sont en cours dans notre équipe, en particulier sur la modélisation dynamique et l'amarrage protéine-protéine. Ainsi, la méthode d'inférence des réseaux que j'ai développée prend toute sa place dans une suite de méthodes qui permettront de créer des réseaux dynamiques des processus cellulaires de manière quasi-automatique à partir des données expérimentales. Cela ouvrira la voie à une meilleure compréhension de ces processus complexes, et à la possibilité de maximiser l'information que les données expérimentales sont susceptibles d'apporter.

Bibliographie

- 1. Lagerstrom MC, Schioth HB. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nature reviews. Drug discovery* 2008,**7**:339-357.
- 2. Lefkowitz RJ, Stadel JM, Caron MG. Adenylate cyclase-coupled beta-adrenergic receptors: structure and mechanisms of activation and desensitization. *Annual review of biochemistry* 1983,**52**:159-186.
- 3. Felder CC, Briley EM, Axelrod J, Simpson JT, Mackie K, Devane WA. Anandamide, an endogenous cannabimimetic eicosanoid, binds to the cloned human cannabinoid receptor and stimulates receptor-mediated signal transduction. *Proceedings of the National Academy of Sciences of the United States of America* 1993, **90**:7656-7660.
- 4. Masu Y, Nakayama K, Tamaki H, Harada Y, Kuno M, Nakanishi S. cDNA cloning of bovine substance-K receptor through oocyte expression system. *Nature* 1987,**329**:836-838.
- 5. Parmentier M, Libert F, Maenhaut C, Lefort A, Gerard C, Perret J, Van Sande J, Dumont JE, Vassart G. Molecular cloning of the thyrotropin receptor. *Science* 1989,**246**:1620-1622.
- 6. Brown EM, Gamba G, Riccardi D, Lombardi M, Butters R, Kifor O, Sun A, Hediger MA, Lytton J, Hebert SC. Cloning and characterization of an extracellular Ca(2+)-sensing receptor from bovine parathyroid. *Nature* 1993,**366**:575-580.
- 7. Zhao Q, Wu BL. Ice breaking in GPCR structural biology. *Acta pharmacologica Sinica* 2012,**33**:324-334.
- 8. Dixon RA, Kobilka BK, Strader DJ, Benovic JL, Dohlman HG, Frielle T, Bolanowski MA, Bennett CD, Rands E, Diehl RE, Mumford RA, Slater EE, Sigal IS, Caron MG, Lefkowitz RJ, Strader CD. Cloning of the gene and cDNA for mammalian beta-adrenergic receptor and homology with rhodopsin. *Nature* 1986,321:75-79.
- 9. Trumpp-Kallmeyer S, Hoflack J, Bruinvels A, Hibert M. Modeling of G-protein-coupled receptors: application to dopamine, adrenaline, serotonin, acetylcholine, and mammalian opsin receptors. *Journal of medicinal chemistry* 1992, **35**:3448-3462.
- 10. Unger VM, Hargrave PA, Baldwin JM, Schertler GF. Arrangement of rhodopsin transmembrane alpha-helices. *Nature* 1997, **389**:203-206.
- 11. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 2000, **289**:739-745.
- 12. Fan QR, Hendrickson WA. Structure of human follicle-stimulating hormone in complex with its receptor. *Nature* 2005, **433**:269-277.
- 13. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 2007,**318**:1258-1265.
- 14. Park JH, Scheerer P, Hofmann KP, Choe HW, Ernst OP. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* 2008, **454**:183-187.
- 15. Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah ST, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK. Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* 2011,477:549-555.
- 16. Choi G, Guo J, Makriyannis A. The conformation of the cytoplasmic helix 8 of the CB1 cannabinoid receptor using NMR and circular dichroism. *Biochimica et biophysica acta* 2005,**1668**:1-9.
- 17. Bockaert J, Pin JP. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *The EMBO journal* 1999, **18**:1723-1729.

- 18. Bockaert J, Pin JP. [Use of a G-protein-coupled receptor to communicate. An evolutionary success]. *Comptes rendus de l'Academie des sciences. Serie III, Sciences de la vie* 1998,**321**:529-551.
- 19. Ulloa-Aguirre A, Crepieux P, Poupon A, Maurel MC, Reiter E. Novel pathways in gonadotropin receptor signaling and biased agonism. *Reviews in endocrine & metabolic disorders* 2011,**12**:259-274.
- 20. Corin K, Baaske P, Geissler S, Wienken CJ, Duhr S, Braun D, Zhang S. Structure and function analyses of the purified GPCR human vomeronasal type 1 receptor 1. *Scientific reports* 2011,**1**:172.
- 21. MacDonald BT, Tamai K, He X. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Developmental cell* 2009,**17**:9-26.
- 22. Bakthavatsalam D, Meijer HJ, Noegel AA, Govers F. Novel phosphatidylinositol phosphate kinases with a G-protein coupled receptor signature are shared by Dictyostelium and Phytophthora. *Trends in microbiology* 2006,**14**:378-382.
- 23. Alvarez-Curto E, Rozen DE, Ritchie AV, Fouquet C, Baldauf SL, Schaap P. Evolutionary origin of cAMP-based chemoattraction in the social amoebae. *Proceedings of the National Academy of Sciences of the United States of America* 2005,**102**:6385-6390.
- 24. Kawabe Y, Morio T, James JL, Prescott AR, Tanaka Y, Schaap P. Activated cAMP receptors switch encystation into sporulation. *Proceedings of the National Academy of Sciences of the United States of America* 2009,**106**:7089-7094.
- 25. Saxe CL, 3rd, Ginsburg GT, Louis JM, Johnson R, Devreotes PN, Kimmel AR. CAR2, a prestalk cAMP receptor required for normal tip formation and late development of Dictyostelium discoideum. *Genes & development* 1993,**7**:262-272.
- 26. Galandrin S, Oligny-Longpre G, Bouvier M. The evasive nature of drug efficacy: implications for drug discovery. *Trends in pharmacological sciences* 2007, **28**:423-430.
- 27. Mueller H, Weingarten R, Ransnas LA, Bokoch GM, Sklar LA. Differential amplification of antagonistic receptor pathways in neutrophils. *The Journal of biological chemistry* 1991, **266**:12939-12943.
- 28. Wall MA, Coleman DE, Lee E, Iniguez-Lluhi JA, Posner BA, Gilman AG, Sprang SR. The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2. *Cell* 1995,83:1047-1058.
- 29. Baehr W, Morita EA, Swanson RJ, Applebury ML. Characterization of bovine rod outer segment G-protein. *The Journal of biological chemistry* 1982, **257**:6452-6460.
- 30. Wilkie TM, Gilbert DJ, Olsen AS, Chen XN, Amatruda TT, Korenberg JR, Trask BJ, de Jong P, Reed RR, Simon MI. Evolution of the mammalian G protein alpha subunit multigene family. *Nature genetics* 1992,**1**:85-91.
- 31. Weiss B, Costa E. Adenyl cyclase activity in rat pineal gland: effects of chronic denervation and norepinephrine. *Science* 1967,**156**:1750-1752.
- 32. Fakunding JL, Tindall DJ, Dedman JR, Mena CR, Means AR. Biochemical actions of follice-stimulating hormone in the sertoli cell of the rat testis. *Endocrinology* 1976,**98**:392-402.
- 33. Pace U, Hanski E, Salomon Y, Lancet D. Odorant-sensitive adenylate cyclase may mediate olfactory reception. *Nature* 1985,**316**:255-258.
- 34. Tesmer JJ, Sunahara RK, Gilman AG, Sprang SR. Crystal structure of the catalytic domains of adenylyl cyclase in a complex with Gsalpha.GTPgammaS. *Science* 1997, **278**:1907-1916.

- 35. Insel PA, Ostrom RS. Forskolin as a tool for examining adenylyl cyclase expression, regulation, and G protein signaling. *Cellular and molecular neurobiology* 2003,**23**:305-314.
- 36. Corvol JC, Studler JM, Schonn JS, Girault JA, Herve D. Galpha(olf) is necessary for coupling D1 and A2a receptors to adenylyl cyclase in the striatum. *Journal of neurochemistry* 2001,**76**:1585-1588.
- 37. Kim C, Cheng CY, Saldanha SA, Taylor SS. PKA-I holoenzyme structure reveals a mechanism for cAMP-dependent activation. *Cell* 2007,**130**:1032-1043.
- 38. Kerlavage AR, Taylor SS. Site-specific cyclic nucleotide binding and dissociation of the holoenzyme of cAMP-dependent protein kinase. *The Journal of biological chemistry* 1982, **257**:1749-1754.
- 39. Paveto C, Passeron S, Corbin JD, Moreno S. Two different intrachain cAMP sites in the cAMP-dependent protein kinase of the dimorphic fungus Mucor rouxii. *European journal of biochemistry / FEBS* 1989,**179**:429-434.
- 40. de Rooij J, Zwartkruis FJ, Verheijen MH, Cool RH, Nijman SM, Wittinghofer A, Bos JL. Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. *Nature* 1998, **396**:474-477.
- 41. Quinn JC, Johnson-Farley NN, Yoon J, Cowen DS. Activation of extracellular-regulated kinase by 5-hydroxytryptamine(2A) receptors in PC12 cells is protein kinase C-independent and requires calmodulin and tyrosine kinases. *The Journal of pharmacology and experimental therapeutics* 2002,**303**:746-752.
- 42. Cottom J, Salvador LM, Maizels ET, Reierstad S, Park Y, Carr DW, Davare MA, Hell JW, Palmer SS, Dent P, Kawakatsu H, Ogata M, Hunzicker-Dunn M. Folliclestimulating hormone activates extracellular signal-regulated kinase but not extracellular signal-regulated kinase kinase through a 100-kDa phosphotyrosine phosphatase. *The Journal of biological chemistry* 2003, **278**:7167-7179.
- 43. Liu S, Li Y, Kim S, Fu Q, Parikh D, Sridhar B, Shi Q, Zhang X, Guan Y, Chen X, Xiang YK. Phosphodiesterases coordinate cAMP propagation induced by two stimulatory G protein-coupled receptors in hearts. *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**:6578-6583.
- 44. Noda M, Yasuda S, Okada M, Higashida H, Shimada A, Iwata N, Ozaki N, Nishikawa K, Shirasawa S, Uchida M, Aoki S, Wada K. Recombinant human serotonin 5A receptors stably expressed in C6 glioma cells couple to multiple signal transduction pathways. *Journal of neurochemistry* 2003,84:222-232.
- 45. Crepieux P, Marion S, Martinat N, Fafeur V, Vern YL, Kerboeuf D, Guillou F, Reiter E. The ERK-dependent signalling is stage-specifically modulated by FSH, during primary Sertoli cell maturation. *Oncogene* 2001, **20**:4696-4709.
- 46. Yang J, Wu J, Kowalska MA, Dalvi A, Prevost N, O'Brien PJ, Manning D, Poncz M, Lucki I, Blendy JA, Brass LF. Loss of signaling through the G protein, Gz, results in abnormal platelet activation and altered responses to psychoactive drugs. *Proceedings of the National Academy of Sciences of the United States of America* 2000,**97**:9984-9989.
- 47. Kimple ME, Nixon AB, Kelly P, Bailey CL, Young KH, Fields TA, Casey PJ. A role for G(z) in pancreatic islet beta-cell biology. *The Journal of biological chemistry* 2005,**280**:31708-31713.
- 48. Yamaguchi I, Harmon SK, Todd RD, O'Malley KL. The rat D4 dopamine receptor couples to cone transducin (Galphat2) to inhibit forskolin-stimulated cAMP accumulation. *The Journal of biological chemistry* 1997, **272**:16599-16602.

- 49. Bikkavilli RK, Feigin ME, Malbon CC. G alpha o mediates WNT-JNK signaling through dishevelled 1 and 3, RhoA family members, and MEKK 1 and 4 in mammalian cells. *Journal of cell science* 2008,**121**:234-245.
- 50. Dauphinee SM, Voelcker V, Tebaykina Z, Wong F, Karsan A. Heterotrimeric Gi/Go proteins modulate endothelial TLR signaling independent of the MyD88-dependent pathway. *American journal of physiology. Heart and circulatory physiology* 2011,**301**:H2246-2253.
- 51. Banquet S, Delannoy E, Agouni A, Dessy C, Lacomme S, Hubert F, Richard V, Muller B, Leblais V. Role of G(i/o)-Src kinase-PI3K/Akt pathway and caveolin-1 in beta-adrenoceptor coupling to endothelial NO synthase in mouse pulmonary artery. *Cellular signalling* 2011,**23**:1136-1143.
- 52. Klages B, Brandt U, Simon MI, Schultz G, Offermanns S. Activation of G12/G13 results in shape change and Rho/Rho-kinase-mediated myosin light chain phosphorylation in mouse platelets. *The Journal of cell biology* 1999,**144**:745-754.
- 53. Wirth A, Benyo Z, Lukasova M, Leutgeb B, Wettschureck N, Gorbey S, Orsy P, Horvath B, Maser-Gluth C, Greiner E, Lemmer B, Schutz G, Gutkind JS, Offermanns S. G12-G13-LARG-mediated signaling in vascular smooth muscle is required for salt-induced hypertension. *Nature medicine* 2008,**14**:64-68.
- 54. Cho MK, Kim WD, Ki SH, Hwang JI, Choi S, Lee CH, Kim SG. Role of Galpha12 and Galpha13 as novel switches for the activity of Nrf2, a key antioxidative transcription factor. *Molecular and cellular biology* 2007, **27**:6195-6208.
- 55. Versteeg HH, Spek CA, Slofstra SH, Diks SH, Richel DJ, Peppelenbosch MP. FVIIa:TF induces cell survival via G12/G13-dependent Jak/STAT activation and BclXL production. *Circulation research* 2004,**94**:1032-1040.
- 56. Taylor SJ, Smith JA, Exton JH. Purification from bovine liver membranes of a guanine nucleotide-dependent activator of phosphoinositide-specific phospholipase C. Immunologic identification as a novel G-protein alpha subunit. *The Journal of biological chemistry* 1990, **265**:17150-17156.
- 57. Taylor SJ, Exton JH. Two alpha subunits of the Gq class of G proteins stimulate phosphoinositide phospholipase C-beta 1 activity. *FEBS letters* 1991,**286**:214-216.
- 58. Blank JL, Ross AH, Exton JH. Purification and characterization of two G-proteins that activate the beta 1 isozyme of phosphoinositide-specific phospholipase C. Identification as members of the Gq class. *The Journal of biological chemistry* 1991,**266**:18206-18216.
- 59. Yung LY, Joshi SA, Chan RY, Chan JS, Pei G, Wong YH. GalphaL1 (Galpha14) couples the opioid receptor-like1 receptor to stimulation of phospholipase C. *The Journal of pharmacology and experimental therapeutics* 1999, **288**:232-238.
- 60. Ho MK, Yung LY, Chan JS, Chan JH, Wong CS, Wong YH. Galpha(14) links a variety of G(i)- and G(s)-coupled receptors to the stimulation of phospholipase C. *British journal of pharmacology* 2001,**132**:1431-1440.
- 61. Su Y, Ho MK, Wong YH. A hematopoietic perspective on the promiscuity and specificity of Galpha16 signaling. *Neuro-Signals* 2009,**17**:71-81.
- 62. Liu AM, Lo RK, Lee MM, Wang Y, Yeung WW, Ho MK, Su Y, Ye RD, Wong YH. Galpha16 activates Ras by forming a complex with tetratricopeptide repeat 1 (TPR1) and Son of Sevenless (SOS). *Cellular signalling* 2010,**22**:1448-1458.
- 63. Giannone F, Malpeli G, Lisi V, Grasso S, Shukla P, Ramarli D, Sartoris S, Monsurro V, Krampera M, Amato E, Tridente G, Colombatti M, Parenti M, Innamorati G. The puzzling uniqueness of the heterotrimeric G15 protein and its potential beyond hematopoiesis. *Journal of molecular endocrinology* 2010,44:259-269.

- 64. Shinohara H, Kurosaki T. Comprehending the complex connection between PKCbeta, TAK1, and IKK in BCR signaling. *Immunological reviews* 2009,**232**:300-318.
- 65. Gutkind JS, Offermanns S. A new G(q)-initiated MAPK signaling pathway in the heart. *Developmental cell* 2009,**16**:163-164.
- 66. Lee MM, Wong YH. CCR1-mediated activation of Nuclear Factor-kappaB in THP-1 monocytic cells involves Pertussis Toxin-insensitive Galpha(14) and Galpha(16) signaling cascades. *Journal of leukocyte biology* 2009,**86**:1319-1329.
- 67. Ballou LM, Lin HY, Fan G, Jiang YP, Lin RZ. Activated G alpha q inhibits p110 alpha phosphatidylinositol 3-kinase and Akt. *The Journal of biological chemistry* 2003,**278**:23472-23479.
- 68. Yeung WW, Wong YH. Galpha16 interacts with Class IA phosphatidylinositol 3-kinases and inhibits Akt signaling. *Cellular signalling* 2010,**22**:1379-1387.
- 69. Logothetis DE, Kurachi Y, Galper J, Neer EJ, Clapham DE. The beta gamma subunits of GTP-binding proteins activate the muscarinic K+ channel in heart. *Nature* 1987,**325**:321-326.
- 70. Haga K, Haga T. Activation by G protein beta gamma subunits of agonist- or light-dependent phosphorylation of muscarinic acetylcholine receptors and rhodopsin. *The Journal of biological chemistry* 1992, **267**:2222-2227.
- 71. Pitcher JA, Inglese J, Higgins JB, Arriza JL, Casey PJ, Kim C, Benovic JL, Kwatra MM, Caron MG, Lefkowitz RJ. Role of beta gamma subunits of G proteins in targeting the beta-adrenergic receptor kinase to membrane-bound receptors. *Science* 1992, **257**:1264-1267.
- 72. Brock C, Schaefer M, Reusch HP, Czupalla C, Michalke M, Spicher K, Schultz G, Nurnberg B. Roles of G beta gamma in membrane recruitment and activation of p110 gamma/p101 phosphoinositide 3-kinase gamma. *The Journal of cell biology* 2003,**160**:89-99.
- 73. Marcantoni A, Levi RC, Gallo MP, Hirsch E, Alloatti G. Phosphoinositide 3-kinasegamma (PI3Kgamma) controls L-type calcium current (ICa,L) through its positive modulation of type-3 phosphodiesterase (PDE3). *Journal of cellular physiology* 2006,**206**:329-336.
- 74. Ma YC, Huang J, Ali S, Lowry W, Huang XY. Src tyrosine kinase is a novel direct effector of G proteins. *Cell* 2000,**102**:635-646.
- 75. Tsuchioka M, Takebayashi M, Hisaoka K, Maeda N, Nakata Y. Serotonin (5-HT) induces glial cell line-derived neurotrophic factor (GDNF) mRNA expression via the transactivation of fibroblast growth factor receptor 2 (FGFR2) in rat C6 glioma cells. *Journal of neurochemistry* 2008,**106**:244-257.
- 76. Kasina S, Scherle PA, Hall CL, Macoska JA. ADAM-mediated amphiregulin shedding and EGFR transactivation. *Cell proliferation* 2009, **42**:799-812.
- 77. Wayne CM, Fan HY, Cheng X, Richards JS. Follicle-stimulating hormone induces multiple signaling cascades: evidence that activation of Rous sarcoma oncogene, RAS, and the epidermal growth factor receptor are critical for granulosa cell differentiation. *Molecular endocrinology* 2007, **21**:1940-1957.
- 78. Beaulieu JM, Gainetdinov RR. The physiology, signaling, and pharmacology of dopamine receptors. *Pharmacological reviews* 2011,**63**:182-217.
- 79. Dattatreyamurty B, Figgs LW, Reichert LE, Jr. Physical and functional association of follitropin receptors with cholera toxin-sensitive guanine nucleotide-binding protein. *The Journal of biological chemistry* 1987,**262**:11737-11745.
- 80. Kume H, Hall IP, Washabau RJ, Takagi K, Kotlikoff MI. Beta-adrenergic agonists regulate KCa channels in airway smooth muscle by cAMP-dependent and independent mechanisms. *The Journal of clinical investigation* 1994,**93**:371-379.

- 81. Oufkir T, Arseneault M, Sanderson JT, Vaillancourt C. The 5-HT 2A serotonin receptor enhances cell viability, affects cell cycle progression and activates MEK-ERK1/2 and JAK2-STAT3 signalling pathways in human choriocarcinoma cell lines. *Placenta* 2010,**31**:439-447.
- 82. Heitzler D, Durand G, Gallay N, Rizk A, Ahn S, Kim J, Violin JD, Dupuy L, Gauthier C, Piketty V, Crepieux P, Poupon A, Clement F, Fages F, Lefkowitz RJ, Reiter E. Competing G protein-coupled receptor kinases balance G protein and beta-arrestin signaling. *Molecular systems biology* 2012,**8**:590.
- 83. Offermanns S, Negulescu P, Hu YH, Simon MI. Conditionally expressed G alpha 15 couples to endogenous receptors in GH3 cells. *Naunyn-Schmiedeberg's archives of pharmacology* 2001,**364**:140-148.
- 84. Quick MW, Lester HA, Davidson N, Simon MI, Aragay AM. Desensitization of inositol 1,4,5-trisphosphate/Ca2+-induced Cl- currents by prolonged activation of G proteins in Xenopus oocytes. *The Journal of biological chemistry* 1996,**271**:32021-32027.
- 85. Durand G. Importance relative des voies de signalisation dépendantes de l'AMPc/PKA ou des β-arrestines dans la réponse cellulaire à l'hormone folliculo-stimulante (FSH). *Manuscrit de thèse* 2011.
- 86. Montiel M, Quesada J, Jimenez E. Activation of second messenger-dependent protein kinases induces muscarinic acetylcholine receptor desensitization in rat thyroid epithelial cells. *Molecular and cellular endocrinology* 2004,**223**:35-41.
- 87. Cho EY, Cho DI, Park JH, Kurose H, Caron MG, Kim KM. Roles of protein kinase C and actin-binding protein 280 in the regulation of intracellular trafficking of dopamine D3 receptor. *Molecular endocrinology* 2007,**21**:2242-2254.
- 88. Hardy AR, Conley PB, Luo J, Benovic JL, Poole AW, Mundell SJ. P2Y1 and P2Y12 receptors for ADP desensitize by distinct kinase-dependent mechanisms. *Blood* 2005,**105**:3552-3560.
- 89. Bouvier M, Hausdorff WP, De Blasi A, O'Dowd BF, Kobilka BK, Caron MG, Lefkowitz RJ. Removal of phosphorylation sites from the beta 2-adrenergic receptor delays onset of agonist-promoted desensitization. *Nature* 1988, 33:370-373.
- 90. Benovic JL, Kuhn H, Weyand I, Codina J, Caron MG, Lefkowitz RJ. Functional desensitization of the isolated beta-adrenergic receptor by the beta-adrenergic receptor kinase: potential role of an analog of the retinal protein arrestin (48-kDa protein). *Proceedings of the National Academy of Sciences of the United States of America* 1987.84:8879-8882.
- 91. Zhao X, Huang J, Khani SC, Palczewski K. Molecular forms of human rhodopsin kinase (GRK1). *The Journal of biological chemistry* 1998,**273**:5124-5131.
- 92. Virlon B, Firsov D, Cheval L, Reiter E, Troispoux C, Guillou F, Elalouf JM. Rat G protein-coupled receptor kinase GRK4: identification, functional expression, and differential tissue distribution of two splice variants. *Endocrinology* 1998,**139**:2784-2795.
- 93. Lohse MJ, Benovic JL, Codina J, Caron MG, Lefkowitz RJ. beta-Arrestin: a protein that regulates beta-adrenergic receptor function. *Science* 1990,**248**:1547-1550.
- 94. Ferguson SS. Evolving concepts in G protein-coupled receptor endocytosis: the role in receptor desensitization and signaling. *Pharmacological reviews* 2001,**53**:1-24.
- 95. Luttrell LM, Lefkowitz RJ. The role of beta-arrestins in the termination and transduction of G-protein-coupled receptor signals. *Journal of cell science* 2002,**115**:455-465.
- 96. Nobles KN, Xiao K, Ahn S, Shukla AK, Lam CM, Rajagopal S, Strachan RT, Huang TY, Bressler EA, Hara MR, Shenoy SK, Gygi SP, Lefkowitz RJ. Distinct

- phosphorylation sites on the beta(2)-adrenergic receptor establish a barcode that encodes differential functions of beta-arrestin. *Science signaling* 2011,**4**:ra51.
- 97. Kim J, Ahn S, Ren XR, Whalen EJ, Reiter E, Wei H, Lefkowitz RJ. Functional antagonism of different G protein-coupled receptor kinases for beta-arrestin-mediated angiotensin II receptor signaling. *Proceedings of the National Academy of Sciences of the United States of America* 2005,**102**:1442-1447.
- 98. Ren XR, Reiter E, Ahn S, Kim J, Chen W, Lefkowitz RJ. Different G protein-coupled receptor kinases govern G protein and beta-arrestin-mediated signaling of V2 vasopressin receptor. *Proceedings of the National Academy of Sciences of the United States of America* 2005,**102**:1448-1453.
- 99. Kara E, Crepieux P, Gauthier C, Martinat N, Piketty V, Guillou F, Reiter E. A phosphorylation cluster of five serine and threonine residues in the C-terminus of the follicle-stimulating hormone receptor is important for desensitization but not for beta-arrestin-mediated ERK activation. *Molecular endocrinology* 2006, **20**:3014-3026.
- 100. Reiter E, Lefkowitz RJ. GRKs and beta-arrestins: roles in receptor silencing, trafficking and signaling. *Trends in endocrinology and metabolism: TEM* 2006, **17**:159-165.
- 101. Defea K. Beta-arrestins and heterotrimeric G-proteins: collaborators and competitors in signal transduction. *British journal of pharmacology* 2008,**153 Suppl 1**:S298-309.
- 102. Wehbi V, Tranchant T, Durand G, Musnier A, Decourtye J, Piketty V, Butnev VY, Bousfield GR, Crepieux P, Maurel MC, Reiter E. Partially deglycosylated equine LH preferentially activates beta-arrestin-dependent signaling at the follicle-stimulating hormone receptor. *Molecular endocrinology* 2010,**24**:561-573.
- 103. Chun KS, Lao HC, Langenbach R. The prostaglandin E2 receptor, EP2, stimulates keratinocyte proliferation in mouse skin by G protein-dependent and {beta}-arrestin1-dependent signaling pathways. *The Journal of biological chemistry* 2010,**285**:39672-39681.
- 104. Xiao K, Sun J, Kim J, Rajagopal S, Zhai B, Villen J, Haas W, Kovacs JJ, Shukla AK, Hara MR, Hernandez M, Lachmann A, Zhao S, Lin Y, Cheng Y, Mizuno K, Ma'ayan A, Gygi SP, Lefkowitz RJ. Global phosphorylation analysis of beta-arrestin-mediated signaling downstream of a seven transmembrane receptor (7TMR). *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**:15299-15304.
- 105. Sun Y, Huang J, Xiang Y, Bastepe M, Juppner H, Kobilka BK, Zhang JJ, Huang XY. Dosage-dependent switch from G protein-coupled to G protein-independent signaling by a GPCR. *The EMBO journal* 2007, **26**:53-64.
- 106. Barthet G, Carrat G, Cassier E, Barker B, Gaven F, Pillot M, Framery B, Pellissier LP, Augier J, Kang DS, Claeysen S, Reiter E, Baneres JL, Benovic JL, Marin P, Bockaert J, Dumuis A. Beta-arrestin1 phosphorylation by GRK5 regulates G protein-independent 5-HT4 receptor signalling. *The EMBO journal* 2009, **28**:2706-2718.
- 107. Nechamen CA, Thomas RM, Cohen BD, Acevedo G, Poulikakos PI, Testa JR, Dias JA. Human follicle-stimulating hormone (FSH) receptor interacts with the adaptor protein APPL1 in HEK 293 cells: potential involvement of the PI3K pathway in FSH signaling. *Biology of reproduction* 2004,**71**:629-636.
- 108. Nechamen CA, Thomas RM, Dias JA. APPL1, APPL2, Akt2 and FOXO1a interact with FSHR in a potential signaling complex. *Molecular and cellular endocrinology* 2007, **260-262**:93-99.
- 109. Thomas RM, Nechamen CA, Mazurkiewicz JE, Ulloa-Aguirre A, Dias JA. The adapter protein APPL1 links FSH receptor to inositol 1,4,5-trisphosphate production and is implicated in intracellular Ca(2+) mobilization. *Endocrinology* 2011,**152**:1691-1701.

- 110. Cunningham MR, McIntosh KA, Pediani JD, Robben J, Cooke AE, Nilsson M, Gould GW, Mundell S, Milligan G, Plevin R. Novel role for proteinase-activated receptor 2 (PAR2) in membrane trafficking of proteinase-activated receptor 4 (PAR4). *The Journal of biological chemistry* 2012.
- 111. Cohen BD, Nechamen CA, Dias JA. Human follitropin receptor (FSHR) interacts with the adapter protein 14-3-3tau. *Molecular and cellular endocrinology* 2004,**220**:1-7.
- 112. Okamoto Y, Shikano S. Phosphorylation-dependent C-terminal binding of 14-3-3 proteins promotes cell surface expression of HIV co-receptor GPR15. *The Journal of biological chemistry* 2011,**286**:7171-7181.
- 113. Zitzer H, Honck HH, Bachner D, Richter D, Kreienkamp HJ. Somatostatin receptor interacting protein defines a novel family of multidomain proteins present in human and rodent brain. *The Journal of biological chemistry* 1999, **274**:32997-33001.
- 114. Cao TT, Deacon HW, Reczek D, Bretscher A, von Zastrow M. A kinase-regulated PDZ-domain interaction controls endocytic sorting of the beta2-adrenergic receptor. *Nature* 1999,**401**:286-290.
- 115. Papkoff H, Ekblad M. Ovine follicle stimulating hormone: preparation and characterization of its subunits. *Biochemical and biophysical research communications* 1970.**40**:614-621.
- 116. Pierce JG, Parsons TF. Glycoprotein hormones: structure and function. *Annual review of biochemistry* 1981, **50**:465-495.
- 117. Simoni M, Gromoll J, Nieschlag E. The follicle-stimulating hormone receptor: biochemistry, molecular biology, physiology, and pathophysiology. *Endocrine reviews* 1997,**18**:739-773.
- 118. Themmen APN, Huhtaniemi IT. Mutations of gonadotropins and gonadotropin receptors: elucidating the physiology and pathophysiology of pituitary-gonadal function. *Endocrine reviews* 2000,**21**:551-583.
- 119. Kumar TR, Wang Y, Lu N, Matzuk MM. Follicle stimulating hormone is required for ovarian follicle maturation but not male fertility. *Nature genetics* 1997,**15**:201-204.
- 120. Huhtaniemi IT, Aittomaki K. Mutations of follicle-stimulating hormone and its receptor: effects on gonadal function. *European journal of endocrinology / European Federation of Endocrine Societies* 1998, **138**:473-481.
- 121. O'Shaughnessy PJ, Monteiro A, Abel M. Testicular development in mice lacking receptors for follicle stimulating hormone and androgen. *PloS one* 2012,**7**:e35136.
- 122. Vloeberghs V, Peeraer K, Pexsters A, D'Hooghe T. Ovarian hyperstimulation syndrome and complications of ART. *Best practice & research. Clinical obstetrics & gynaecology* 2009, **23**:691-709.
- 123. Caron E, Ghosh S, Matsuoka Y, Ashton-Beaucage D, Therrien M, Lemieux S, Perreault C, Roux PP, Kitano H. A comprehensive map of the mTOR signaling network. *Molecular systems biology* 2010,**6**:453.
- 124. Xu H, Schaniel C, Lemischka IR, Ma'ayan A. Toward a complete in silico, multi-layered embryonic stem cell regulatory network. *Wiley interdisciplinary reviews*. *Systems biology and medicine* 2010,**2**:708-733.
- 125. Gloaguen P, Crépieux P, Heitzler D, Poupon A, Reiter E. Mapping the follicle-stimulating hormone-induced signaling networks. *Frontiers in cellular endocrinology* 2011,**2**.
- 126. Telikicherla D, Ambekar A, Palapetta SM, Dwivedi SB, Raju R, Sharma J, Prasad Ts K, Ramachandra Y, Mohan SS, Maharudraiah J, Mukherjee S, Pandey A. A comprehensive curated resource for follicle stimulating hormone signaling. *BMC research notes* 2011,**4**:408.

- 127. Musnier A, Blanchot B, Reiter E, Crepieux P. GPCR signalling to the translation machinery. *Cellular signalling* 2010,**22**:707-716.
- 128. Gonzalez-Robayna IJ, Falender AE, Ochsner S, Firestone GL, Richards JS. Follicle-Stimulating hormone (FSH) stimulates phosphorylation and activation of protein kinase B (PKB/Akt) and serum and glucocorticoid-Induced kinase (Sgk): evidence for A kinase-independent signaling by FSH in granulosa cells. *Molecular endocrinology* 2000,14:1283-1300.
- 129. Bonnet A, Lagarrigue S, Liaubet L, Robert-Granie C, Sancristobal M, Tosser-Klopp G. Pathway results from the chicken data set using GOTM, Pathway Studio and Ingenuity softwares. *BMC proceedings* 2009,**3 Suppl 4**:S11.
- 130. Hecker M, Lambeck S, Toepfer S, van Someren E, Guthke R. Gene regulatory network inference: data integration in dynamic models-a review. *Bio Systems* 2009.**96**:86-103.
- 131. Gutierrez-Rios RM, Rosenblueth DA, Loza JA, Huerta AM, Glasner JD, Blattner FR, Collado-Vides J. Regulatory network of Escherichia coli: consistency between literature knowledge and microarray profiles. *Genome research* 2003, **13**:2435-2443.
- 132. Shin DG, Kazmi SA, Pei B, Kim YA, Maddox J, Nori R, Wong A, Krueger W, Rowe D. Computing consistency between microarray data and known gene regulation relationships. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society* 2009,**13**:1075-1082.
- 133. Gloaguen P. Détermination de la meilleure technique d'acquisition d'un réseau de régulation à partir des connaissances avec pour application le métabolisme des lipides. *Rapport de Stage Master 2* 2009.
- 134. Benson M, Carlsson L, Guillot G, Jernas M, Langston MA, Rudemo M, Andersson B. A network-based analysis of allergen-challenged CD4+ T cells from patients with allergic rhinitis. *Genes and immunity* 2006,**7**:514-521.
- 135. Nikitin A, Egorov S, Daraselia N, Mazo I. Pathway studio--the analysis and navigation of molecular networks. *Bioinformatics* 2003,**19**:2155-2157.
- 136. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC bioinformatics* 2006,**7 Suppl 1**:S7.
- 137. Wang K, Saito M, Bisikirska BC, Alvarez MJ, Lim WK, Rajbhandari P, Shen Q, Nemenman I, Basso K, Margolin AA, Klein U, Dalla-Favera R, Califano A. Genomewide identification of post-translational modulators of transcription factor activity in human B cells. *Nature biotechnology* 2009, **27**:829-839.
- 138. Luo W, Hankenson KD, Woolf PJ. Learning transcriptional regulatory networks from high throughput gene expression data using continuous three-way mutual information. *BMC bioinformatics* 2008,**9**:467.
- 139. Ciaccio MF, Wagner JP, Chuu CP, Lauffenburger DA, Jones RB. Systems analysis of EGF receptor signaling dynamics with microwestern arrays. *Nature methods* 2010,**7**:148-155.
- 140. Andrei A, Kendziorski C. An efficient method for identifying statistical interactors in gene association networks. *Biostatistics* 2009,**10**:706-718.
- 141. Jaffrezic F, Tosser-Klopp G. Gene network reconstruction from microarray data. *BMC proceedings* 2009,**3 Suppl 4**:S12.
- 142. Chiquet J, Smith A, Grasseau G, Matias C, Ambroise C. SIMoNe: Statistical Inference for MOdular NEtworks. *Bioinformatics* 2009,**25**:417-418.
- 143. Verzelen N. Modèles graphiques gaussiens et sélection de modèles. *Manuscrit de thèse* 2008.

- 144. Chu JH, Weiss ST, Carey VJ, Raby BA. A graphical model approach for inferring large-scale networks integrating gene expression and genetic polymorphism. *BMC* systems biology 2009,**3**:55.
- 145. Krumsiek J, Suhre K, Illig T, Adamski J, Theis FJ. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC* systems biology 2011,5:21.
- 146. Le Phillip P, Bahl A, Ungar LH. Using prior knowledge to improve genetic network reconstruction from microarray data. *In silico biology* 2004,**4**:335-353.
- 147. de Jong H. Modeling and simulation of genetic regulatory systems: a literature review. Journal of computational biology: a journal of computational molecular cell biology 2002.9:67-103.
- 148. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 2005, **308**:523-529.
- 149. Jing L, Ng MK, Liu Y. Construction of gene networks with hybrid approach from expression profile and gene ontology. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society* 2010,**14**:107-118.
- 150. Lin K, Husmeier D. Modelling transcriptional regulation with a mixture of factor analyzers and variational bayesian expectation maximization. *EURASIP journal on bioinformatics & systems biology* 2009, **2009**:601068.
- 151. Kaderali L, Dazert E, Zeuge U, Frese M, Bartenschlager R. Reconstructing signaling pathways from RNAi data using probabilistic Boolean threshold networks. *Bioinformatics* 2009, **25**:2229-2235.
- 152. Pe'er D. Bayesian network analysis of signaling networks: a primer. *Science's STKE*: signal transduction knowledge environment 2005, **2005**:pl4.
- 153. Koivisto M, Sood K. Exact Bayesian Structure Discovery in Bayesian Networks. *Journal of Machine Learning Research* 2004,**5**:549-573.
- 154. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS biology* 2007,5:e8.
- 155. Yu J, Smith VA, Wang PP, Hartemink AJ, Jarvis ED. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* 2004, **20**:3594-3603.
- 156. Shmulevich I, Dougherty ER, Kim S, Zhang W. Probabilistic Boolean Networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* 2002, **18**:261-274.
- 157. Martin S, Zhang Z, Martino A, Faulon JL. Boolean dynamics of genetic regulatory networks inferred from microarray time series data. *Bioinformatics* 2007, 23:866-874.
- 158. Hegde SR, Rajasingh H, Das C, Mande SS, Mande SC. Understanding communication signals during mycobacterial latency through predicted genome-wide protein interactions and boolean modeling. *PloS one* 2012,7:e33893.
- 159. Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, Klamt S, Sorger PK. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Molecular systems biology* 2009,**5**:331.
- 160. Mitsos A, Melas IN, Siminelakis P, Chairakaki AD, Saez-Rodriguez J, Alexopoulos LG. Identifying drug effects via pathway alterations using an integer linear programming optimization formulation on phosphoproteomic data. *PLoS computational biology* 2009,5:e1000591.

- 161. Zielinski R, Przytycki PF, Zheng J, Zhang D, Przytycka TM, Capala J. The crosstalk between EGF, IGF, and Insulin cell signaling pathways--computational and experimental analysis. *BMC systems biology* 2009,**3**:88.
- 162. Aldridge BB, Saez-Rodriguez J, Muhlich JL, Sorger PK, Lauffenburger DA. Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling. *PLoS computational biology* 2009,**5**:e1000340.
- 163. Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, Lauffenburger DA. Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. *PLoS computational biology* 2011,7:e1001099.
- 164. Wittmann DM, Krumsiek J, Saez-Rodriguez J, Lauffenburger DA, Klamt S, Theis FJ. Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling. *BMC systems biology* 2009,**3**:98.
- 165. Philippi N, Walter D, Schlatter R, Ferreira K, Ederer M, Sawodny O, Timmer J, Borner C, Dandekar T. Modeling system states in liver cells: survival, apoptosis and their modifications in response to viral infection. *BMC systems biology* 2009,**3**:97.
- 166. Gardner TS, di Bernardo D, Lorenz D, Collins JJ. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 2003,**301**:102-105.
- 167. di Bernardo D, Thompson MJ, Gardner TS, Chobot SE, Eastwood EL, Wojtovich AP, Elliott SJ, Schaus SE, Collins JJ. Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nature biotechnology* 2005, **23**:377-383.
- 168. Bansal M, Della Gatta G, di Bernardo D. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics* 2006, **22**:815-822.
- 169. Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D. How to infer gene networks from expression profiles. *Molecular systems biology* 2007,**3**:78.
- 170. Birtwistle MR, Kholodenko BN. Endocytosis and signalling: a meeting with mathematics. *Molecular oncology* 2009,**3**:308-320.
- 171. Jackson P. Introduction To Expert Systems (3 ed.); 1998.
- 172. Hayes-Roth F, Waterman DA, Lenat DB. An Overview of Expert Systems. *Building Expert Systems* 1983:3-29.
- 173. Blinov ML, Faeder JR, Goldstein B, Hlavacek WS. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics* 2004, **20**:3289-3291.
- 174. Colvin J, Monine MI, Gutenkunst RN, Hlavacek WS, Von Hoff DD, Posner RG. RuleMonkey: software for stochastic simulation of rule-based models. *BMC bioinformatics* 2010,**11**:404.
- 175. Sneddon MW, Faeder JR, Emonet T. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nature methods* 2011,**8**:177-183.
- 176. Ollivier JF, Shahrezaei V, Swain PS. Scalable rule-based modelling of allosteric proteins and biochemical networks. *PLoS computational biology* 2010,**6**:e1000975.
- 177. Anjum S, Doucet A, Holmes CC. A boosting approach to structure learning of graphs with and without prior knowledge. *Bioinformatics* 2009, **25**:2929-2936.
- 178. Butte AJ, Kohane IS. Unsupervised knowledge discovery in medical databases using relevance networks. *Proceedings / AMIA ... Annual Symposium*. *AMIA Symposium* 1999:711-715.
- 179. Danos V, Feret J, Fontana W, Harmer R, Krivine J. Rule-Based Modelling of Cellular Signalling. *Proceedings of the 18 th International Conference on Concurrency Theory (CONCUR'07), Lecture Notes in Computer Science* 2007.

- 180. Le Novere N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, Bergman FT, Gauges R, Ghazal P, Kawaji H, Li L, Matsuoka Y, Villeger A, Boyd SE, Calzone L, Courtot M, Dogrusoz U, Freeman TC, Funahashi A, Ghosh S, Jouraku A, Kim S, Kolpakov F, Luna A, Sahle S, Schmidt E, Watterson S, Wu G, Goryanin I, Kell DB, Sander C, Sauro H, Snoep JL, Kohn K, Kitano H. The Systems Biology Graphical Notation. *Nature biotechnology* 2009, 27:735-741.
- 181. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin, II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novere N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 2003, 19:524-531.
- 182. Finney A, Hucka M. Systems biology markup language: Level 2 and beyond. *Biochemical Society transactions* 2003,**31**:1472-1473.
- 183. Hucka M, Frank Bergmann, Stefan Hoops, Sarah Keating, Sven Sahle, James Schaff, Lucian Smith, Wilkinson D. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core. 2010.
- 184. Le Novere N. Model storage, exchange and integration. *BMC neuroscience* 2006,**7 Suppl 1**:S11.
- Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, Wu G, D'Eustachio P, 185. Schaefer C, Luciano J, Schacherer F, Martinez-Flores I, Hu Z, Jimenez-Jacinto V, Joshi-Tope G, Kandasamy K, Lopez-Fuentes AC, Mi H, Pichler E, Rodchenkov I, Splendiani A, Tkachev S, Zucker J, Gopinath G, Rajasimha H, Ramakrishnan R, Shah I, Syed M, Anwar N, Babur O, Blinov M, Brauner E, Corwin D, Donaldson S, Gibbons F, Goldberg R, Hornbeck P, Luna A, Murray-Rust P, Neumann E, Ruebenacker O, Samwald M, van Iersel M, Wimalaratne S, Allen K, Braun B, Whirl-Carrillo M, Cheung KH, Dahlquist K, Finney A, Gillespie M, Glass E, Gong L, Haw R, Honig M, Hubaut O, Kane D, Krupa S, Kutmon M, Leonard J, Marks D, Merberg D, Petri V, Pico A, Ravenscroft D, Ren L, Shah N, Sunshine M, Tang R, Whaley R, Letovksy S, Buetow KH, Rzhetsky A, Schachter V, Sobral BS, Dogrusoz U, McWeeney S, Aladjem M, Birney E, Collado-Vides J, Goto S, Hucka M, Le Novere N, Maltsev N, Pandey A, Thomas P, Wingender E, Karp PD, Sander C, Bader GD. The BioPAX community standard for pathway data sharing. Nature biotechnology 2010**,28**:935-942.
- 186. Schaff JC, Slepchenko BM, Loew LM. Physiological modeling with virtual cell framework. *Methods in enzymology* 2000,**321**:1-23.
- 187. Ruebenacker O, Moraru, II, Schaff JC, Blinov ML. Integrating BioPAX pathway knowledge with SBML models. *IET systems biology* 2009,**3**:317-328.
- 188. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 2003, **13**:2498-2504.
- 189. Zinovyev A, Viara E, Calzone L, Barillot E. BiNoM: a Cytoscape plugin for manipulating and analyzing biological networks. *Bioinformatics* 2008, **24**:876-877.
- 190. Nabeshima H, Iwanuma K, Inoue K, Ray O. SOLAR: An automated deduction system for consequence finding. *AI Communications 23* 2010:183-203.
- 191. Rougny A. Automatic construction of GPCR signalling networks. *Internship report* 2012.

- 192. Funahashi A, Matsuoka Y, Jouraku A, Morohashi M, Kikuchi N, Kitano H. CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks. *Proceedings of the IEEE* 2008, **96**:1254-1265.
- 193. Funahashi A, Morohashi M, Kitano H, Tanimura N. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO* 2003,**1**:159-162.
- 194. Tunjung WA, Yokoo M, Hoshino Y, Miyake Y, Kadowaki A, Sato E. Effect of hyaluronan to inhibit caspase activation in porcine granulosa cells. *Biochemical and biophysical research communications* 2009, **382**:160-164.
- 195. Nemcova L, Nagyova E, Petlach M, Tomanek M, Prochazka R. Molecular mechanisms of insulin-like growth factor 1 promoted synthesis and retention of hyaluronic acid in porcine oocyte-cumulus complexes. *Biology of reproduction* 2007, **76**:1016-1024.
- 196. Ruwanpura SM, McLachlan RI, Stanton PG, Loveland KL, Meachem SJ. Pathways involved in testicular germ cell apoptosis in immature rats after FSH suppression. *The Journal of endocrinology* 2008,**197**:35-43.
- 197. Yu Y, Li W, Han Z, Luo M, Chang Z, Tan J. The effect of follicle-stimulating hormone on follicular development, granulosa cell apoptosis and steroidogenesis and its mediation by insulin-like growth factor-I in the goat ovary. *Theriogenology* 2003,**60**:1691-1704.
- 198. Gonzalez IH, Santana P, Gonzalez-Robayna I, Ferrer M, Morales V, Blanco FL, Fanjul LF. Regulation of the expression of prostate apoptosis response protein 4 (Par-4) in rat granulosa cells. *Apoptosis : an international journal on programmed cell death* 2007, 12:769-779.
- 199. Cossigny DA, Findlay JK, Drummond AE. The effects of FSH and activin A on follicle development in vitro. *Reproduction* 2012,**143**:221-229.
- 200. Escamilla-Hernandez R, Little-Ihrig L, Orwig KE, Yue J, Chandran U, Zeleznik AJ. Constitutively active protein kinase A qualitatively mimics the effects of follicle-stimulating hormone on granulosa cell differentiation. *Molecular endocrinology* 2008,22:1842-1852.
- 201. Fortune JE, Armstrong DT. Androgen production by theca and granulosa isolated from proestrous rat follicles. *Endocrinology* 1977,**100**:1341-1347.
- 202. Kayampilly PP, Menon KM. Follicle-stimulating hormone inhibits adenosine 5'-monophosphate-activated protein kinase activation and promotes cell proliferation of primary granulosa cells in culture through an Akt-dependent pathway. *Endocrinology* 2009.**150**:929-935.
- 203. Tosca L, Chabrolle C, Uzbekova S, Dupont J. Effects of metformin on bovine granulosa cells steroidogenesis: possible involvement of adenosine 5' monophosphate-activated protein kinase (AMPK). *Biology of reproduction* 2007,**76**:368-378.
- 204. Yu FQ, Han CS, Yang W, Jin X, Hu ZY, Liu YX. Activation of the p38 MAPK pathway by follicle-stimulating hormone regulates steroidogenesis in granulosa cells differentially. *The Journal of endocrinology* 2005, **186**:85-96.
- 205. Rice VM, Chaudhery AR, Oluola O, Limback SD, Roby KF, Terranova PF. Herbimycin, a tyrosine kinase inhibitor with Src selectivity, reduces progesterone and estradiol secretion by human granulosa cells. *Endocrine* 2001,**15**:271-276.
- 206. Silva JM, Hamel M, Sahmi M, Price CA. Control of oestradiol secretion and of cytochrome P450 aromatase messenger ribonucleic acid accumulation by FSH involves different intracellular pathways in oestrogenic bovine granulosa cells in vitro. *Reproduction* 2006,**132**:909-917.

- 207. Hoang YD, Nakamura BN, Luderer U. Follicle-stimulating hormone and estradiol interact to stimulate glutathione synthesis in rat ovarian follicles and granulosa cells. *Biology of reproduction* 2009,**81**:636-646.
- 208. Kayampilly PP, Menon KM. Follicle-stimulating hormone increases tuberin phosphorylation and mammalian target of rapamycin signaling through an extracellular signal-regulated kinase-dependent pathway in rat granulosa cells. *Endocrinology* 2007,**148**:3950-3957.
- 209. Musnier A, Heitzler D, Boulo T, Tesseraud S, Durand G, Lecureuil C, Guillou H, Poupon A, Reiter E, Crepieux P. Developmental regulation of p70 S6 kinase by a G protein-coupled receptor dynamically modelized in primary cells. *Cellular and molecular life sciences: CMLS* 2009,**66**:3487-3503.
- 210. Smith EP, Dickson BA, Chernausek SD. Insulin-like growth factor binding protein-3 secretion from cultured rat sertoli cells: dual regulation by follicle stimulating hormone and insulin-like growth factor-I. *Endocrinology* 1990,**127**:2744-2751.
- 211. Rappaport MS, Smith EP. Insulin-like growth factor (IGF) binding protein 3 in the rat testis: follicle-stimulating hormone dependence of mRNA expression and inhibition of IGF-I action on cultured Sertoli cells. *Biology of reproduction* 1995,**52**:419-425.
- 212. Andl CD, Mizushima T, Oyama K, Bowser M, Nakagawa H, Rustgi AK. EGFR-induced cell migration is mediated predominantly by the JAK-STAT pathway in primary esophageal keratinocytes. *American journal of physiology. Gastrointestinal and liver physiology* 2004,**287**:G1227-1237.
- 213. Ongeri EM, Verderame MF, Hammond JM. The TATA binding protein associated factor 4b (TAF4b) mediates FSH stimulation of the IGFBP-3 promoter in cultured porcine ovarian granulosa cells. *Molecular and cellular endocrinology* 2007,**278**:29-35.
- 214. Shimada M, Hernandez-Gonzalez I, Gonzalez-Robayna I, Richards JS. Paracrine and autocrine regulation of epidermal growth factor-like factors in cumulus oocyte complexes and granulosa cells: key roles for prostaglandin synthase 2 and progesterone receptor. *Molecular endocrinology* 2006, **20**:1352-1365.
- 215. Copp J, Manning G, Hunter T. TORC-specific phosphorylation of mammalian target of rapamycin (mTOR): phospho-Ser2481 is a marker for intact mTOR signaling complex 2. *Cancer research* 2009,**69**:1821-1827.
- 216. Zhang Z, Stiegler AL, Boggon TJ, Kobayashi S, Halmos B. EGFR-mutated lung cancer: a paradigm of molecular oncology. *Oncotarget* 2010,**1**:497-514.
- 217. Sun Y, Ling K, Wagoner MP, Anderson RA. Type I gamma phosphatidylinositol phosphate kinase is required for EGF-stimulated directional cell migration. *The Journal of cell biology* 2007,**178**:297-308.
- 218. Haj FG, Markova B, Klaman LD, Bohmer FD, Neel BG. Regulation of receptor tyrosine kinase signaling by protein tyrosine phosphatase-1B. *The Journal of biological chemistry* 2003, **278**:739-744.
- 219. Chou J, Stolz DB, Burke NA, Watkins SC, Wells A. Distribution of gelsolin and phosphoinositol 4,5-bisphosphate in lamellipodia during EGF-induced motility. *The international journal of biochemistry & cell biology* 2002,**34**:776-790.
- 220. Cao C, Huang X, Han Y, Wan Y, Birnbaumer L, Feng GS, Marshall J, Jiang M, Chu WM. Galpha(i1) and Galpha(i3) are required for epidermal growth factor-mediated activation of the Akt-mTORC1 pathway. *Science signaling* 2009,**2**:ra17.
- 221. Agazie YM, Hayman MJ. Molecular mechanism for a role of SHP2 in epidermal growth factor receptor signaling. *Molecular and cellular biology* 2003, **23**:7875-7886.

- 222. Galisteo ML, Chernoff J, Su YC, Skolnik EY, Schlessinger J. The adaptor protein Nck links receptor tyrosine kinases with the serine-threonine kinase Pak1. *The Journal of biological chemistry* 1996, **271**:20997-21000.
- 223. Grasso P, Reichert LE, Jr. Follicle stimulating hormone (FSH) induces G protein dissociation from receptor-G protein complexes **FSH** in reconstituted proteoliposomes. Biochemical and biophysical research communications 1989,**162**:1214-1221.
- 224. Arey BJ, Stevis PE, Deecher DC, Shen ES, Frail DE, Negro-Vilar A, Lopez FJ. Induction of promiscuous G protein coupling of the follicle-stimulating hormone (FSH) receptor: a novel mechanism for transducing pleiotropic actions of FSH isoforms. *Molecular endocrinology* 1997, 11:517-526.
- 225. Alam H, Maizels ET, Park Y, Ghaey S, Feiger ZJ, Chandel NS, Hunzicker-Dunn M. Follicle-stimulating hormone activation of hypoxia-inducible factor-1 by the phosphatidylinositol 3-kinase/AKT/Ras homolog enriched in brain (Rheb)/mammalian target of rapamycin (mTOR) pathway is necessary for induction of select protein markers of follicular differentiation. *The Journal of biological chemistry* 2004, **279**:19431-19440.
- 226. Jin H, Wang DY, Mei YF, Qiu WB, Zhou Y, Wang DM, Tan XR, Li YG. Mitogen-activated protein kinases pathway is involved in physiological testosterone-induced tissue factor pathway inhibitor expression in endothelial cells. *Blood coagulation & fibrinolysis : an international journal in haemostasis and thrombosis* 2010,21:420-424.
- 227. Tranchant T, Durand G, Gauthier C, Crepieux P, Ulloa-Aguirre A, Royere D, Reiter E. Preferential beta-arrestin signalling at low receptor density revealed by functional characterization of the human FSH receptor A189 V mutation. *Molecular and cellular endocrinology* 2011,**331**:109-118.

Annexes

Annexe 1 : Relations formalisées non présentées de la partie « Matériel et Méthodes »

Effet sur la Complexation : L'effet sur la complexation est une relation de régulation sur la formation d'un complexe (Figure 78).

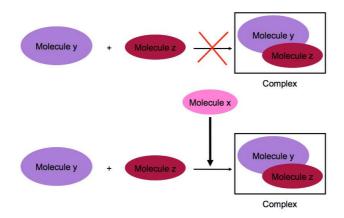


Figure 78: Effet sur la complexation

Une molécule x induit la complexation d'une molécule y avec une molécule z pour former le complexe (1)y:(1)z.

La formalisation d'un effet sur la complexation est représentée par le prédicat :

$ECO(x,[(n_v,y),(n_z,z)],(n_v)y:(n_z)z,distance,status,effect)$

Où x agit sur la formation du complexe $(n_y)y:(n_z)z$, où n_y et n_z sont des entiers représentant le nombre de molécules y et z respectivement. La molécule x peut appartenir à n'importe quelle classe moléculaire. Les molécules qui entrent dans la composition d'un complexe peuvent appartenir à n'importe quelle classe (ADN, ARN, Protéine, Petite molécule et Complexe). Bien sûr, le produit de la complexation appartient à la classe Complexe. La distance peut être directe, indirecte ou inconnue. Le statut peut être confirmé, infirmé ou hypothétique. L'effet, quant à lui, peut être une augmentation, une diminution ou nul.

<u>Dissociation</u>: C'est la séparation d'un complexe en partenaires isolés (Figure 79).

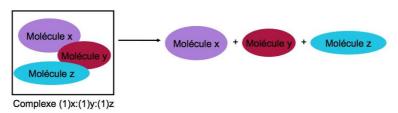


Figure 79: Dissociation

Le complexe (1)x:(1)y:(1)z se dissocie pour donner les molécules x, y et z.

La formalisation d'une dissociation est représentée par le prédicat :

DISSOCIATION($(n_x)x:(n_y)y:(n_z)z,[x,y,z]$,status)

Où $(n_x)x:(n_y)y:(n_z)z$ est un complexe et [x,y,z] représente la liste des molécules obtenues après dissociation.

Les molécules issues de la dissociation peuvent appartenir à n'importe quelle classe de molécules. La molécule dissociée appartient à la classe Complexe.

Modification: C'est la modification d'une molécule par une autre molécule (Figure 80).

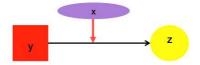


Figure 80: Modification

La molécule x induit la modification de la molécule y pour donner la molécule z.

La formalisation d'une modification est représentée par le prédicat :

MODIFY(x,y,b,distance,status,effect)

Où *x* est la molécule qui a un effet sur la modification de *y* en *z*. Les molécules *x*, *y* et *z* peuvent appartenir aux classes Protéine, Complexe ou Petite molécule. La distance peut être directe, indirecte ou inconnue. Le statut peut être confirmé, infirmé ou hypothétique. L'effet, quant à lui, peut être une augmentation, une diminution ou nul.

Les réactions enzymatiques ou encore les modifications post-traductionnelles des protéines sont des cas particuliers de modifications.

Réaction enzymatique : C'est la transformation d'un substrat en un produit par une enzyme (Figure 81).

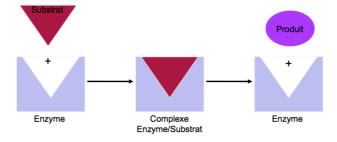


Figure 81 : Réaction enzymatique

Le substrat vient se fixer sur l'enzyme qui catalyse sa transformation en produit.

La formalisation d'une réaction enzymatique est représentée par le prédicat :

CATALYZE(x,[(n_v ,y)],[(n_z ,z)],direct,status,effect)

Où x est l'enzyme, y est le substrat et z est le produit. La distance d'une réaction enzymatique est toujours directe et l'effet ne peut prendre que deux des trois valeurs indiquées auparavant, *increase* ou *noeffect*. L'enzyme appartient à la classe Protéine, Complexe ou Ribozyme, le ou les substrats peuvent appartenir à n'importe quelle classe moléculaire tout comme le ou les produits.

Acétylation: C'est l'ajout d'un groupement acétyle (COCH₃) sur une protéine. La formalisation de cette relation suit la même logique que la phosphorylation mais se traduit par le prédicat :

ACETYLATE(x,y,b,distance,status,effect)

Où la molécule x peut faire partie de n'importe quelle classe moléculaire, la molécule y est une protéine et b est une protéine acétylée.

Autres modifications post-traductionnelles: Les autres modifications ne sont pas encore formalisées mais elles suivront le même raisonnement que la phosphorylation et l'acétylation.

<u>Transcription</u>: C'est la synthèse d'un ARN à partir d'une matrice d'ADN par une polymérase. Les séquences d'ADN concernées sont uniquement les gènes.

La formalisation de la transcription est décrite par le prédicat :

TRANSCRIPTION(v,w)

Où *v* représente le gène transcrit et *w* l'ARN obtenu.

La transcription est un phénomène cellulaire très complexe qui nécessite un grand nombre de régulations. Le prédicat formalise uniquement le fait qu'un gène *v* est transcrit en ARN *w*.

Effet sur la transcription : Une molécule peut avoir une action sur la transcription d'un gène. Par exemple, il est possible qu'un gène v ne puisse pas être transcrit en ARN w si une molécule z est absente (Figure 82).

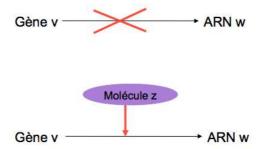


Figure 82 : Influence d'une molécule sur la transcription

La formalisation d'une action sur la transcription est définie par le prédicat :

TCE(z,v,w,distance,status,effect)

Où z est la molécule agissant sur la transcription du gène v en ARN w. z peut appartenir à n'importe quelle classe moléculaire. La molécule v est forcément un gène et w un ARN.

Effet sur la traduction : Une molécule peut avoir une action sur la traduction d'un ARN messager. Par exemple, il est possible qu'un ARN messager w ne puisse pas être traduit en protéine y si une molécule z est absente.

La formalisation d'une action sur la traduction est définie par le prédicat :

TLE(z,w,y,distance,status,effect)

Où z est la molécule agissant sur la traduction de l'ARN messager w en protéine y. z peut provenir de n'importe quelle classe moléculaire. La molécule w est forcément un ARN messager et y est une protéine.

<u>Dégradation d'une molécule</u>: Une molécule peut avoir une action sur la dégradation d'une autre molécule.

La formalisation d'une action sur la dégradation est définie par le prédicat :

DEGRADATION(x,y,distance,status,effect)

Où x est la molécule agissant sur la dégradation de la molécule y. x peut provenir de n'importe quelle classe moléculaire, de même pour y.

Annexe 2 : Création des règles non présentée dans la partie « Matériel et Méthodes »

Règles simples

Immunolocalisation: L'immunolocalisation est une technique de détection de la localisation cellulaire d'une molécule grâce à un anticorps dirigé contre cette molécule. Cette localisation peut être réalisée de deux façons différentes, soit par immunofluorescence (prédicat FLUO) soit par immunohistochimie (prédicat IHC). Les règles étant de la même forme pour les deux techniques, nous allons prendre l'exemple de l'immunofluorescence.

La détection étant réalisée grâce à un anticorps, il est nécessaire d'indiquer la cible de cet anticorps par le prédicat ANTIBODYAGAINST(a,x). Le prédicat FLUO a pour arguments l'anticorps a, un compartiment cellulaire *compartment* et la présence ou non de la molécule dans le compartiment : *positive* ou *negative*. Lorsque la présence de la molécule est confirmée dans un compartiment, il est possible de conclure sur la localisation (prédicat LOCALIZATION(x,compartment)), ce qui se traduit par la règle :

• **IF** FLUO(a,compartment,positive) **AND** ANTIBODYAGAINST(a,x) **THEN** LOCALIZATION(x,compartment)

Lorsque c'est l'absence de la molécule dans le compartiment cité qui est observée, il est possible de conclure sur la non localisation de la molécule dans le compartiment cité (prédicat NOLOCALIZATION(x,compartment)), ce qui permet d'obtenir la règle :

• **IF** FLUO(a,compartment,negative) **AND** ANTIBODYAGAINST(a,x) **THEN** NOLOCALIZATION(x,compartment)

Enzyme-Linked ImmunoSorbent Assay (ELISA): Un ELISA permet le dosage d'une molécule d'intérêt grâce à un anticorps (primaire) dirigé contre cette molécule. L'ELISA est utilisé pour réaliser un dosage absolu de la molécule d'intérêt en présence et en absence d'une autre molécule y. Un anticorps secondaire couplé à une enzyme (ou un fluorochrome) permet la détection de l'anticorps primaire. Lorsque le substrat de l'enzyme est ajouté, il est transformé en un produit coloré. Cette coloration permet la quantification de la molécule d'intérêt car plus elle est présente, plus la quantité de l'anticorps primaire fixé sera élevée donc plus la coloration due à l'enzyme couplée au deuxième anticorps sera élevée. Ainsi, la quantification

est possible grâce à la spectrophotométrie dans le cas du couplage à une enzyme ou par fluorescence dans le cas du couplage avec un fluorochrome.

Le prédicat ELISA a pour arguments la molécule signal y, l'anticorps primaire a et l'effet e observé. Il est nécessaire d'ajouter le prédicat ANTIBODYAGAINST(a,x) pour indiquer que l'anticorps primaire a reconnaît la molécule x.

Etant donné qu'il est possible d'observer des quantités différentes de x en fonction de la présence ou de l'absence de y, cela signifie que x résulte de la modification d'une autre molécule l (MODIFIEDFORM(x,l)), qui n'est pas détectée par l'anticorps. Cette expérience ne permet pas de savoir quelle forme modifiée de x est concernée, la conclusion est donc hypothétique. Cette expérience ne permet pas non plus de savoir si la relation est directe ou indirecte. La Figure 83 illustre l'exemple de l'AMPc et ses deux formes modifiées.

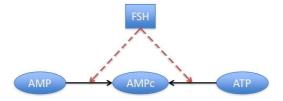


Figure 83 : Exemple de l'AMPc et ses différentes formes modifiées

L'AMPc est une molécule qui présente deux formes modifiées : l'ATP et l'AMP. Donc les prédicats MODIFIEDFORM(cAMP,ATP) et MODIFIEDFORM(cAMP,AMP) existent, ce qui entraîne, lors d'un test ELISA en présence de FSH, la création de deux hypothèses : la FSH induirait la modification de l'AMP en AMPc avec une distance inconnue et la FSH induirait la modification de l'ATP en AMPc avec une distance inconnue.

La conclusion est représentée par le prédicat MODIFY(y,l,x,unknown,hypothesis,e). Et on peut écrire la règle suivante :

• **IF** ELISA(y,a,e) **AND** ANTIBODYAGAINST(a,x) **AND** MODIFIEDFORM(x,l) **THEN** MODIFY(y,l,x,unknown,hypothesis,e)

L'effet *e* peut prendre pour valeur *increase*, *decrease* ou *noeffect* et il est le même dans la condition et dans la conclusion.

ELISA avec ajout d'un inhibiteur (ICELISA): Il est également possible d'introduire un inhibiteur d'une molécule z dont on veut connaître l'effet sur la production de la molécule quantifiée par ELISA en présence du signal y. Cette expérience n'est utile que dans la mesure où l'on sait déjà que y induit la modification de la molécule l en x: MODIFY(y,l,x,unknown,confirmed,increase). La détection de la molécule x est toujours réalisée par l'anticorps x (prédicat ANTIBODYAGAINST(x,x). Le prédicat ICELISA prend

pour arguments le signal y, l'anticorps utilisé a, l'inhibiteur utilisé i et l'effet observé e: ICELISA(y,a,i,e). Si l'inhibition de z (prédicat INHIBITOR(i,z)) conduit à une diminution de la quantité de x alors on peut conclure que z induit la modification de l en x avec une distance inconnue.

Cette observation permet également de faire deux hypothèses sur les relations entre z et x: i) z induit la modification de l en x indépendemment de y, ii) z induit la modification de c en y (MODIFIEDFORM(y,c)) avec une distance inconnue, avec c qui n'a pas d'effet sur la modification de l en x, ainsi z augmente la quantité de y, ce qui augmente la quantité de x, iii) y induit la modification de d en z (MODIFIEDFORM(z,d)) avec une distance inconnue, avec d qui n'a pas d'effet sur la modification de l en x, dans ce cas z est un intermédiaire entre y et x (Figure 84).

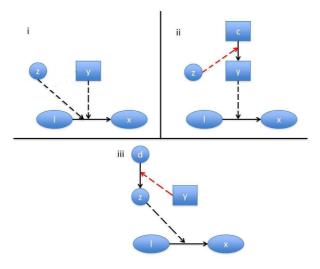


Figure 84: Conclusions obtenues à partir d'un test ELISA avec inhibiteur

On sait que la molécule y induit la modification de l en x avec une distance inconnue. On veut connaître l'action de la molécule z sur cette relation. On ajoute un inhibiteur de z et on constate que la quantité de x diminue. On peut en conclure que z induit la modification de l en x avec une distance inconnue. Trois hypothèses différentes permettent d'expliquer ce résultat.

L'assemblage des différents prédicats permet d'obtenir la règle suivante :

•	IF MODIFY(y,l,x,unknown,confirmed	l,increase)(i) AND ICELISA(y,a,i,decrease)
	AND ANTIBODYAGAINST (a,x)	\mathbf{AND} INHIBITOR(i,z) \mathbf{AND}
	MODIFIEDFORM(y,c) AND	MODIFIEDFORM(z,d) THEN
	MODIFY(z,l,x,unknown,confirmed,incre	ease)(i) AND
	MODIFY(y,d,z,unknown,hypothesis,inca	rease)(iii) AND
	MODIFY(z,c,y,unknown,hypothesis,incr	rease)(ii) AND
	MODIFY(d,l,x,unknown,hypothesis,noe	effect)(iii) AND
	MODIFY(c,l,x,unknown,hypothesis,noe	ffect)(ii)

Remarque : Cette règle est valable si et seulement si la molécule y augmente la modification de *l* en *x* et si l'inhibition par *i* la diminue. D'autres règles sont disponibles dans l'Annexe 4. D'autres types d'expériences sont formalisés en suivant le même raisonnement :

- Phosphorylation Assay avec inhibiteur/antagoniste/siRNA,
- RadioImmunoAssay avec inhibiteur,
- Western Blot avec inhibiteur.

Fluorescence Resonance Energy Tranfer (FRET)¹⁷: Le FRET permet de mettre en évidence une complexation entre deux molécules. Le principe est de coupler un fluorophore donneur (e.g. : CFP) à une molécule y et un fluorophore accepteur (e.g. : YFP) à une molécule z.

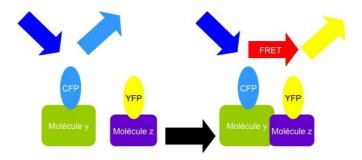


Figure 85: Principe du FRET

Lorsque les deux fluorophores sont éloignés l'un de l'autre, l'excitation du donneur provoque une réémission de fluorescence à la fréquence spécifique du donneur. Lorsque les deux molécules forment un complexe, l'énergie lumineuse émise par le donneur est transférée à l'accepteur, qui réémet à sa longueur d'onde spécifique, différente de celle du donneur (Figure 85). On peut ainsi évaluer la proximité de deux molécules fluorescentes en mesurant la réémission de fluorescence par l'accepteur.

La règle correspondante est :

• **IF** FRET(y,z) **THEN** CO([(?,y),(?,z)],(?)y:(?)z,confirmed)

Remarque: Si la stechiométrie est inconnue, les n_x et n_y sont remplacés par des points d'interrogation.

¹⁷ Au sein de l'équipe BIOS, le FRET est utilisé avec des fluorophores présents sur une même

molécule, c'est le changement conformationnel dû à la complexation qui provoque la fluorescence du receveur. Cette technique n'est pas encore formalisée.

<u>Phosphorylation Radioactive Assay (PRA)</u>: Un PRA est formalisé de la même façon qu'un Phosphorylation Assay, seul l'anticorps a dirigé contre la phosphoprotéine b est remplacé par une forme radioactive r de la phosphoprotéine b (prédicat RADIOLABELLEDFORM(r,b)). Les groupements phosphates ajoutés à la protéine sont radioactifs, c'est donc cette radioactivité qui permet de quantifier l'état de phosphorylation de la protéine.

<u>RadioImmunoAssay (RIA)</u>: Le RIA est formalisé de la même façon que le test ELISA, les expériences changent uniquement au niveau de la quantification. Dans le cas du RIA, cette dernière ne se fait pas grâce à la fluorescence ou à l'enzyme liée à l'anticorps mais grâce à la radioactivité de la molécule d'intérêt. De même, la formalisation d'un RIA en présence d'un inhibiteur suit la même logique que le test ELISA avec ajout d'un inhibiteur.

Reverse Transcription Polymerase Chain Reaction (RT-PCR): La RT-PCR permet de quantifier l'expression de l'ARN y, en présence ou en absence d'un signal x. Si l'expérience de RT-PCR montre un effet e sur la quantité d'ARN y en présence de x (par comparaison à la condition en absence de x): RTPCR(x,y,e), on peut en conclure que x a l'effet e sur la transcription du gène z en ARN y: TCE(x,y,y,distance,status,effect)). On doit également préciser que y est le résultat de la transcription de z: TRANSCRIPTION(z,y). Ce qui est décrit par la règle suivante :

• **IF** RTPCR(x,y,e) **AND** TRANSCRIPTION(z,y) **THEN** TCE(x,z,y,unknown,confirmed,e)

L'effet peut prendre pour valeur increase, decrease ou noeffect.

<u>Structure cristalline tridimensionnelle (3D)</u>: La structure cristalline d'un complexe permet de connaître tous les composants d'un complexe ainsi que sa stœchiométrie.

• IF 3D([(nx,x),(ny,y)]) THEN CO([(nx,x),(ny,y),...],(nx)x:(ny)y:...,confirmed)

Western Blot avec utilisation d'un anticorps pan (PANWB): Le PANWB est une expérience permettant de quantifier toutes les formes d'une protéine d'intérêt y en utilisant un anticorps capable de reconnaître la forme sauvage de la protéine ainsi que toutes ses formes modifiées. Si on observe une augmentation (resp. une diminution, une stabilisation) de la quantité de y lorsque x est présente (PANWB(x,a,effect)) (Figure 86), on peut en déduire que : i) soit x active (resp. inhibe, n'a pas d'effet sur) la transcription du gène codant pour y (TCE(x,z,w,unknown,hypothesis,effect)), soit ii) x active (resp. inhibe, n'a pas d'effet sur) la

traduction de l'ARNm correspondant (TLE(x,w,y,unknown,hypothesis,effect)), soit iii) x inhibe (resp. active, n'a pas d'effet sur) la dégradation de l'ARNm correspondant, soit iv) x inhibe (resp. active, n'a pas d'effet sur) la dégradation de la protéine observée. On précise également que z est le gène transcrit en ARNm w (TRANSCRIPTION(z,w)) qui est traduit en la protéine y (TRANSLATION(w,y)).

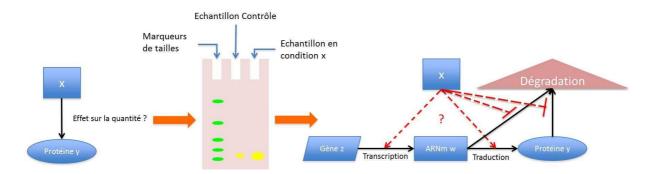


Figure 86: Western Blot

On cherche à mesurer l'effet d'une molécule x sur l'expression de la protéine y. Pour cela, on effectue un PANWB qui, ici, montre une augmentation de la quantité de la protéine dans la condition avec la molécule x par rapport au contrôle sans x. On peut donc conclure que x induit l'augmentation de la quantité de la protéine y, ce qui peut résulter de deux phénomènes : x active la transcription du gène codant pour y ou x active la traduction de l'ARN messager codant pour y ou x inhibe la dégradation de l'ARN messager w avec une distance inconnue ou x inhibe la dégradation de la protéine y avec une distance inconnue. Flèche noire continue = relation directe confirmée, flèche pointillée rouge = relation hypothétique avec une distance inconnue.

On peut ainsi écrire les quatre règles suivantes :

- **IF** PANWB(x,a,effect) **AND** TRANSCRIPTION(z,w) **AND** TRANSLATION(w,y) **AND** PANANTIBODY(a,y) **THEN** TCE(x,z,w,unknown,hypothesis,effect),
- **IF** PANWB(x,a,effect) **AND** TRANSLATION(w,y) **AND** PANANTIBODY(a,y) **THEN** TLE(x,w,y,unknown,hypothesis,effect),
- **IF** PANWB(x,a,effect) **AND** TRANSLATION(w,y) **AND** PANANTIBODY(a,y) **THEN** DEGRADATION(x,w,unknown,hypothesis,effect).
- **IF** PANWB(x,a,effect) **AND** TRANSLATION(w,y) **AND** PANANTIBODY(a,y) **THEN** DEGRADATION(x,y,unknown,hypothesis,effect).

<u>Western Blot avec utilisation d'un anticorps pan et ajout d'un inhibiteur de relation</u>
(<u>RIPANWB</u>): En biologie, il existe des inhibiteurs particuliers, que nous appellerons inhibiteurs de relations, qui permettent d'inhiber un processus donné dans la cellule, comme par exemple la transcription et la traduction. Dans l'exemple précédent, le WB simple ne permet pas de savoir si l'effet se situe au niveau de la traduction ou au niveau de la traduction. L'ajout d'un inhibiteur de la traduction (RELATIONINHIBITOR(i,TRANSLATION)) permet de le déterminer.

Ainsi, si on fait un PANWB en présence de cet inhibiteur, et que l'on observe une diminution de la quantité de *y* en présence de *x* : RIPANWB(x,a,i,decrease), on peut déduire que le signal *x* augmente la traduction de l'ARNm *w* en protéine *y* avec une distance inconnue. Ceci permet d'obtenir la règle :

• **IF** TLE(x,w,y,unknown,hypothesis,increase) **AND** RIPANWB(x,a,i,decrease) **AND** PANANTIBODY(a,y) **AND** RELATIONINHIBITOR(i,TRANSLATION) **THEN** TLE(x,w,y,unknown,confirmed,increase)

<u>Remarque</u>: Attention, si une des hypothèses TCE et TLE générées par un PANWB est validée avec ce type d'inhibiteur, cela ne veut pas dire que l'autre hypothèse est invalidée. Un signal peut très bien agir sur la transcription ET la traduction.

Règles complexes

Création de nouveaux prédicats concernant les molécules :

Il peut arriver que deux prédicats différents aient une même signification mais elle est implicite. Par exemple, la forme phosphorylée/acétylée *b* d'une protéine *y* est également une forme modifiée de cette protéine. Il est donc nécessaire de rendre explicite cette idée, ce qui se traduit par les règles :

- **IF** PHOSPHOFORM(b,y,pos) **THEN** MODIFIEDFORM(b,y)
- **IF** ACETYLFORM(b,y,pos) **THEN** MODIFIEDFORM(b,y)

La création de ces deux nouveaux faits va permettre l'utilisation des formes phosphorylées/acétylées dans les règles où ce sont les formes modifiées qui sont nécessaires. Si la molécule *x* est une forme modifiée de la molécule *y* alors l'inverse est également vrai :

• **IF** MODIFIEDFORM(x,y) **THEN** MODIFIEDFORM(y,x)

De même, si une relation de modification de l en x est confirmée alors il est possible de déduire que l est une forme modifiée de x:

• **IF** MODIFY(y,l,x,d,confirmed,e) **THEN** MODIFIEDFORM(l,x)

S'il y a complexation entre les molécules x et y pour donner le complexe $(n_x)x:(n_y)y$ alors le complexe $(n_x)x:(n_y)y$ existe :

• IF $CO([(n_x,x),(n_y,y),...],(n_x)x:(n_y)y:...,confirmed)$ THEN $COMPLEX((n_x)x:(n_y)y:...)$

Cette création de nouveaux faits moléculaires est également très utile en ce qui concerne les anticorps. Par exemple, lorsque les biologistes utilisent un anticorps pan protéine x, ils savent que l'anticorps reconnaît toutes les formes de x, mais cela est implicite. Donc il est nécessaire de préciser que cet anticorps reconnaît x mais aussi ses autres formes :

• **IF** PANANTIBODY(a,x) **AND** MODIFIEDFORM(c,x) **THEN** PANANTIBODY(a,c)

Cette explicitation permet également de définir que le blocage qu'induit un siRNA sur l'expression d'une protéine, a aussi des conséquences sur les formes modifiées de cette protéine :

• **IF** SIRNAAGAINST(s,x) **AND** MODIFIEDFORM(y,x) **THEN** SIRNAAGAINST(s,y)

Tous ces nouveaux prédicats peuvent être utilisés par les règles présentées précédemment, ce qui permet de déduire de nouvelles relations.

<u>Création de nouvelles relations avec une précision sur la distance :</u> Un grand nombre de faits déduits concernant les relations ont une distance inconnue. En les combinant avec d'autres faits, il est possible d'obtenir une précision sur la distance, ce qui permet d'avoir des relations complètes. Par exemple, comme l'illustre la Figure 87, il est possible de déduire une action indirecte de x sur la phosphorylation de la protéine z en d, ce qui sous forme de règle se décrit de la façon suivante :

• IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase) PHOSPHORYLATE(b,z,d,unknown,confirmed,increase) THEN PHOSPHORYLATE(x,z,d,indirect,confirmed,increase)

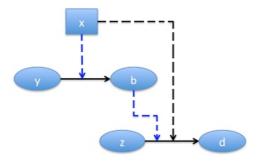


Figure 87 : Exemple de déduction d'une distance indirecte

La molécule x induit la phosphorylation de la protéine y pour donner b avec une distance inconnue et la phosphoprotéine b induit la phosphorylation de la protéine z pour donner d avec une distance inconnue. La molécule x induit indirectement la phosphorylation de la protéine z pour donner d. Flèche pleine noire = relation directe confirmée, flèche pointillée bleue = relation confirmée avec une distance inconnue, flèche pointillée noire = relation indirecte confirmée.

<u>Création de relations implicites à partir de faits déduits validés:</u> Il peut arriver qu'en combinant les faits déduits, de nouvelles hypothèses émergent. Par exemple, deux conclusions obtenues : x induit la transformation de z en y avec une distance inconnue et e catalyse directement la transformation de z en y (Figure 88). Ces deux conclusions permettent de supposer que l'action de x se situe au niveau de l'activation de x, donc on émet l'hypothèse que x induit la modification de x (x) en x0 avec une distance inconnue (Figure 88).

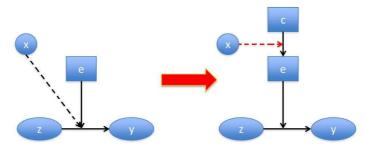


Figure 88 : Déduction d'une relation implicite à partir de faits déduits validés expérimentalement

Ce raisonnement conduit à la création de la règle suivante :

IF MODIFY(x,z,y,unknown,confirmed,increase)
 CATALYZE(e,[(nz,z)],[(ny,y)],direct,confirmed,increase)
 MODIFIEDFORM(e,c) THEN MODIFY(x,c,e,unknown,hypothesis,increase)
 CATALYZE(c,[(nz,z)],[(ny,y)],direct,hypothesis,noeffect)
 MODIFY(x,z,y,indirect,hypothesis,increase)

Signification: Si la molécule x induit la modification de z en y avec une distance inconnue et si la réaction enzymatique ayant pour catalyseur la molécule e, pour substrat la molécule z et pour produit la molécule y, est confirmée et si e (forme active) est une forme modifiée de la molécule e (forme inactive) alors e induirait la modification de e en e (activation) avec une distance inconnue, e n'a pas d'effet sur la réaction enzymatique ayant pour substrat e et pour produit e et la molécule e induirait indirectement la modification de e en e.

<u>Réversibilité des relations</u>: La majorité des relations sont considérées réversibles, et l'activation d'une relation est considérée équivalente à l'inhibition de la relation inverse. Ceci se traduit par les règles suivantes :

- **IF** MODIFY(x,y,z,d,s,increase) **THEN** MODIFY(x,z,y,d,s,decrease)
- **IF** MODIFY(x,y,z,d,s,noeffect) **THEN** MODIFY(x,z,y,d,s,noeffect)
- **IF** MODIFY(x,y,z,d,s,decrease) **THEN** MODIFY(x,z,y,d,s,increase)

En biologie, une seule des deux relations est possible. Cependant, au niveau du fonctionnement du réseau, l'existence des deux hypothèses ne pose pas de problème d'intégrité. De plus, les expériences considérées ne permettent pas de discriminer les deux hypothèses. Ceci pourra être reconsidéré si nécessaire à l'avenir.

Annexe 3 : Dictionnaire des données

Le dictionnaire des données permet de définir la signification de chaque prédicat.

0...9

$3D([(n_x,x),(n_y,y),...])$

(Structure cristalline tridimensionnelle)

La structure tridimensionnelle d'un complexe contenant n_x molécules x et n_y molécules y a été résolue. « ... » signifie que la liste peut s'étendre.

```
x = All; y = All
```

A

AA(x,y,a,effect)

(Acetylation Assay)

L'acetylation assay montre l'effet de la molécule x sur l'acétylation de la protéine y détectée par l'anticorps a reconnaissant la forme acétylée de y.

```
effect = increase, decrease, noeffect x = \text{All}; y = \text{Protein}; a = \text{Antibody}
```

ACETYLATE(x,y,b,distance,status,effect)

(Acétylation)

La molécule x a un effet sur l'acétylation de la protéine y pour donner la protéine acétylée b.

effect = increase, decrease, noeffect

status = confirmed, hypothesis, invalidated

distance = direct, unknown, indirect

x = All; y = Protein; b = AcetylProtein

ACETYLFORM(b,y,pos)

(Forme acétylée)

b est la forme acétylée de la protéine y à la position pos.

y = Protein; b = AcetylProtein

ACETYLPROTEIN(x)

(Protéine acétylée)

La molécule *b* est une protéine acétylée.

$ACPEA(x,[(n_v,y),...],[(n_z,z),...],i,effect)$

(Enzymatic Assay en présence d'un antagoniste)

L'enzymatic assay montre l'effet que peut avoir i, un antagoniste d'une molécule d'intérêt, sur la réaction enzymatique catalysée par x, ayant pour substrat une liste de molécules (ici n_y molécules y) et pour produit une liste de molécules (ici n_z molécules y).

```
effect = increase, decrease, noeffect
```

x = Protein, Complex, Ribozyme ; y = All ; z = All ; i = Protein, Complex, SmallMolecule

ACPPA(x,y,i,a,effect)

(Phosphorylation Assay en présence d'un antagoniste)

Le *phosphorylation assay* montre l'effet que peut avoir *i*, un antagoniste d'une molécule d'intérêt, sur la phosphorylation de la protéine *y* sur laquelle agit la molécule *x* détectée par l'anticorps *a* de la forme phosphorylée de *y*.

effect = increase, decrease, noeffect

x = All; y = Protein; a = Antibody; i = Protein; Complex, SmallMolecule

AGONIST(x,y)

(Agoniste)

La molécule x est un agoniste de la molécule y.

x =Protein, Complex, SmallMolecule ; y =Protein, Complex, SmallMolecule

ANALOGUE(x,y)

(Analogue structural)

La molécule x est un analogue structural de la molécule y.

x =Protein, Complex, SmallMolecule ; y =Protein, Complex, SmallMolecule

ANTAGONIST(i,z)

(Antagoniste)

La molécule *i* est un antagoniste de la molécule *z*.

i = Protein, Complex, SmallMolecule; z = Protein, Complex, SmallMolecule

ANTIBODY(a)

(Anticorps)

a appartient à la classe Anticorps.

ANTIBODYAGAINST(a,b)

(Anticorps)

La molécule *a* est un anticorps dirigé contre *b*.

a = Antibody; b = Protein, Complex, SmallMolecule

В

BDPULLDOWN (x,[y,...])

(Binding Domain Pulldown)

Par PullDown utilisant un domaine de liaison immobilisé sur une colonne, on récupère la molécule x qui se fixe au domaine de liaison et on vérifie les molécules [y,...] liées à x.

x =Protein, SmallMolecule ; y =Protein, SmallMolecule

C

CATALYZE($x,[(n_y,y),...],[(n_z,z),...]$,distance,status,effect)

(Réaction enzymatique)

La molécule x catalyse ou non la réaction enzymatique ayant pour substrat n_y molécules y et pour produit n_z molécules z. Attention les substrats et produits sont des listes de molécules.

effect = increase, noeffect

status = confirmed, hypothesis, invalidated

```
distance = direct
 x = Protein, Complex, Ribozyme ; y = All ; z = All
```

$CO([(n_x,x),(n_y,y)],(n_x)x:(n_y)y,status)$

(Complexation)

Les molécules x et y forment un complexe de n_x molécules x et de n_y molécules y.

status = confirmed, invalidated, hypothesis

x = All; y = All; $(n_x)x:(n_y)y = \text{Complex}$

CODINGGENE(x)

(Gène codant)

La molécule x est un gène codant.

$COMPLEX((n_x)x:(n_y)y)$

(Complexe)

 $(n_x)x:(n_y)y$ est un Complexe.

x = All; y = All

COMPARTDIST(x,y,distance)

Distance entre 2 molécules

Les molécules x et y sont situées dans des compartiments étant à une distance l'un de l'autre.

distance = distant, neighbour, same

D

DEGRADATION(x,y,distance,status,effect)

(Dégradation)

L'ajout de la molecule x a un effet sur la degradation de y.

x = All; y = All

DISSOCIATION($(n_x)x:(n_y)y,[x,y]$,status)

(Dissociation)

Le complexe $(n_x)x:(n_y)y$ se dissocie en ses composants x et y.

status = confirmed, invalidated, hypothesis

 $(n_x)x:(n_y)y = \text{Complex}$; x = All; y = All

DM2(x,y,effect)

(Puce à ADN entre 2 conditions)

L'ajout de la molécule x induit un effet sur la transcription de y.

effect: increase, decrease, noeffect

x = All; y = RNA

DNA(x)

(ADN)

La molécule x est un ADN.

 \mathbf{E}

 $EA(x,[(n_y,y),...],[(n_z,z),...],effect)$

(Enzymatic Assay)

On fait un *Enzymatic Assay* pour savoir si la molécule x catalyse ou pas, la réaction enzymatique ayant pour substrat n_y molécules y et pour produit n_z molécules z (attention les substrats et produits sont des listes).

```
effect = increase, noeffect
```

x =Protein, Complex, Ribozyme ; y =All ; z =All

$ECO(x,[(n_y,y),(n_z,z),...],(n_y)y:(n_z)z:...,distance,status,effect)$

(Effet sur la Complexation)

La molécule x a un effet sur la Complexation de n_y molécules y avec n_z molécules z pour former le complexe $(n_y)y:(n_z)z$.

effect = increase, decrease, noeffect

status = confirmed, hypothesis, invalidated

distance = direct, unknown, indirect

x = All; y = All; z = All; $(n_y)y:(n_z)z = \text{Complex}$

ELISA(y,a,effect)

(Enzyme-Linked ImmunoSorbent Assay)

Un test *ELISA* permet de savoir si l'ajout de la molécule y a un effet sur la quantité de la molécule ciblée par l'anticorps a.

effect = increase, decrease, noeffect

y = All; a = Antibody

ENHANCERBIND(x)

(Enhancer)

La molécule x est un enhancer.

$EXCHANGE(x,(n_x)x:(n_y)y:(n_z)z,((n_z)z,(n_u)u))$

(Echange)

La molécule x induit un échange de n_z molécules z contre n_u molécules u dans le complexe $(n_x)x:(n_y)y:(n_z)z$.

```
x = \text{All}; (n_x)x:(n_y)y:(n_z)z = \text{Complex}; y = \text{All}; z = \text{All}; u = \text{All}
```

 \mathbf{F}

FLUO(a,compartment,effect)

(ImmunoFluorescence)

Une molécule est visualisée dans un compartiment *compartment* grâce à un anticorps spécifique *a*.

compartment = PlasmaMembrane, CytoplasmicVesicles, Cytoplasm, ...

effect = positive, negative

a = Antibody

FRET(y,z)

(Fluorescence Resonance Energy Transfer)

S'il y a *FRET*, on met en évidence une Complexation entre la molécule y et la molécule z sinon il n'y a pas de Complexation.

y =Protein, SmallMolecule ; z =Protein, SmallMolecule

GENE(x)

(Gène)

La molécule x est un gène.

GLYCOPROTEIN(x)

(Protéine glycosylée)

La molécule x est une protéine glycosylée.

GR(x,y,effect)

(Gène rapporteur)

Par gène rapporteur, on vérifie si l'ajout de la molécule x a un effet sur l'expression de la protéine y.

effect = increase, decrease, noeffect x = All; y = Protein

GSTPULLDOWN(x,[y,...])

(Gluthatione-S-transferase pulldown)

On immobilise la molécule x qui porte une étiquette gluthathione transferase sur une colonne de gluthation et on vérifie si d'autres molécules [y,...] sont liées à x.

x = Protein; y = Protein, SmallMolecule

Η

HYDROXYPROTEIN(x)

(Protéine hydroxylée)

La molécule *x* est une protéine hydroxylée.

Ι

ICELISA(y,a,i,effect)

(Enzyme-Linked ImmunoSorbent Assay en présence d'un inhibiteur)

L'*ELISA* montre l'effet que peut avoir un inhibiteur *i* sur l'action que la molécule y a sur la production d'une molécule ciblée par l'anticorps *a*.

effect = increase, decrease, noeffect

y = All; a = Antibody; i = Protein, Complex, SmallMolecule

ICPEA($x,[(n_v,y),...],[(n_z,z),...],i,effect$)

(Enzymatic Assay en présence d'un inhibiteur)

L'*enzymatic assay* montre l'effet que peut avoir i, un inhibiteur d'une molécule d'intérêt, sur la réaction enzymatique catalysée par x, ayant pour substrat une liste de molécules (ici n_y molécules y) et pour produit une liste de molécules (ici n_z molécules y).

effect = increase, decrease, noeffect

x = Protein, Complex, Ribozyme ; y = All ; z = All ; i = Protein, Complex, SmallMolecule

ICPPA(x,y,i,a,effect)

(Phosphorylation Assay en présence d'un inhibiteur)

Le phosphorylation assay montre l'effet que peut avoir i, un inhibiteur d'une molécule

d'intérêt, sur la phosphorylation de la protéine y sur laquelle agit la molécule x, la détection se fait par l'anticorps a ciblant la forme phosphorylée de y.

```
effect = increase, decrease, noeffect
```

```
x = \text{All}; y = \text{Protein}; a = \text{Antibody}; i = \text{Protein}, Complex, SmallMolecule
```

ICPPRA(x,y,r,i,effect)

(Phosphorylation Radioactive Assay en présence d'un inhibiteur)

Le *phosphorylation radioactive assay* montre l'effet que peut avoir un inhibiteur *i* sur la phosphorylation de la protéine *y* sur laquelle agit la molécule *x* en mesurant la forme radioactive *r* de la forme phosphorylée de *y*.

```
effect = increase, decrease, noeffect
```

```
x = \text{All }; y = \text{Protein }; r = \text{PhosphoProtein }; i = \text{Protein, Complex, SmallMolecule}
```

ICPANWB(x,a,i,effect)

(Western Blot en présence d'un inhibiteur)

Le *western blot* montre l'effet que peut avoir *i*, un inhibiteur d'une molécule d'intérêt, sur l'expression d'une protéine sur laquelle agit la molécule *x*, la détection se fait par l'anticorps pan *a* dirigé contre cette protéine.

```
effect = increase, decrease, noeffect
```

```
x = All; a = Antibody; i = Protein, Complex, SmallMolecule
```

ICRIA(y,a,i,effect)

(RadioImmunoAssay en présence d'un inhibiteur)

Le *RIA* montre l'effet que peut avoir i, un inhibiteur d'une molécule d'intérêt, sur l'action que la molécule y a sur la production d'une molécule ciblée par l'anticorps a.

```
effect = increase, decrease, noeffect
```

```
x = All; a = Antibody; i = Protein, Complex, SmallMolecule
```

IHC(a,compartment,effect)

(*ImmunoHistoChemistry*)

Une molécule est localisée dans un compartiment compartment grâce à un anticorps spécifique a.

```
compartment = PlasmaMembrane, CytoplasmicVesicles, Cytoplasm, ...
```

```
effect = positive, negative
```

```
a = Antibody
```

INHIBITOR(i,z)

```
(Inhibiteur)
```

La molécule *i* est un inhibiteur de la molécule *z*.

```
z = All; i = Protein, Complex, SmallMolecule
```

ION(x)

(Ion)

La molécule *x* est un ion.

IP([x,y,...],increase)

(Immunoprécipitation)

L'immunoprécipitation d'une protéine *x* entraı̂ne la coimmunoprécipitation d'une liste de protéines *y*,...

```
x = Protein ; y = All
```

IPD(z,[x,y,...],effect)

(Immunoprécipitation avec ajout d'une molécule)

La molécule z a un effet sur l'immunoprécipitation de x entraînant la coimmunoprécipitation de y.

effect = increase, decrease, noeffect

z = All; x = Protein; y = All

IRTPCR(x,y,i,effect)

(Reverse Transcription Polymerase Chain Reaction en présence d'un inhibiteur)

La *RT-PCR* montre l'effet que peut avoir *i*, un inhibiteur d'une molécule d'intérêt, sur l'action que la molécule *x* a sur l'expression de l'ARN *y*.

effect = increase, decrease, noeffect

x = All; y = RNA; i = Protein, Complex, SmallMolecule

 \mathbf{L}

LOCALIZATION(x,compartment)

(Localisation)

La molécule *x* est localisée dans le compartiment *compartment*.

compartment = PlasmaMembrane, CytoplasmicVesicles, Cytoplasm, ...

x = A11

M

METHYLPROTEIN(x)

(Protéine méthylée)

La molécule *x* est une protéine méthylée.

MODIFIEDFORM(d,z)

(Forme modifiée)

La molécule d est la forme modifiée de la molécule z.

d = All; z = All

MODIFIEDPROTEIN(x)

(Protéine modifiée)

La molécule x est une protéine modifiée.

MODIFY(x,z,d,distance,status,effect)

(Modification)

La molécule x a un effet sur la modification de la molécule z en d.

effect = increase, decrease, noeffect

status = confirmed, invalidated, hypothesis

distance = direct, indirect, unknown

x = All; d = All; z = All

MRNA(x)

(ARN messager)

La molécule *x* est un ARN messager.

MIRNA(x)

(micro ARN)

La molécule *x* est un micro ARN.

MUTANT(x)

(Mutant)

La molécule x est un mutant.

MYRISTOYLPROTEIN(x)

(Protéine myristoylée)

La molécule *x* est une protéine myristoylée.

N

NB(x,y,effect)

(Northern Blot)

Par NB, on vérifie si l'ajout de la molécule x induit un effet sur l'expression de l'ARN y

effect = increase, decrease, noeffect

x = All; y = RNA

NOCODINGGENE(x)

(Gène non codant)

La molécule x est un gène non codant.

NOLOCALIZATION(x,compartment)

(Non Localisation)

La molécule x n'est pas localisée dans le compartiment *compartment*. compartment = PlasmaMembrane, CytoplasmicVesicles, Cytoplasm, ... x = All

NOTMODIFIED(x)

(Non modifiable)

La molécule x n'est pas modifiable.

P

PA(x,y,a,effect)

(Phosphorylation Assay)

Par PA, on vérifie si l'ajout de la molécule x induit un effet sur la phosphorylation de la protéine y qui est suivie grâce à l'anticorps a dirigé contre la forme phosphorylée de y.

effect = increase, decrease, noeffect

x = All; y = Protein; a = Antibody

PALMITOYLPROTEIN(x)

(Protéine palmitoylée)

La molécule *x* est une protéine palmitoylée.

PANANTIBODY(a,x)

(Anticorps Pan)

a est un anticorps pan protéine x, c'est-à-dire qu'il reconnaît la forme sauvage de x mais aussi ses formes modifiées.

```
a = Antibody ; x = Protein
```

PANWB(x,a,effect)

(Western Blot avec l'utilisation un anticorps pan)

Par WB, on vérifie si l'ajout de la molécule x induit un effet sur l'expression d'une protéine suivie grâce à l'anticorps pan a la ciblant.

```
effect = increase, decrease, noeffect x = \text{All}; a = \text{Antibody}
```

PHOSPHOFORM(b,y,pos)

```
(Forme phosphorylée)
b est la forme phosphorylée de la protéine y à la position pos.
y = Protein; b = PhosphoProtein
```

PHOSPHOPROTEIN(x)

(Protéine phosphorylée)

La molécule x est une protéine phosphorylée.

PHOSPHORYLATE(x,y,b,distance,status,effect)

(Phosphorylation)

La molécule x a un effet sur la phosphorylation de la protéine y pour donner la protéine phosphorylée b.

```
effect = increase, decrease, noeffect

status = confirmed, invalidated, hypothesis

distance = direct, indirect, unknown

x = All; y = Protein; b = PhosphoProtein
```

PPM(x,y,r,effect)

(PhosphoPeptide Mapping)

Par PPM, on vérifie si l'ajout de la molécule x induit un effet sur la phosphorylation de la protéine y qui est suivie grâce à la forme radioactive r de la forme phosphorylée de y.

```
effect = increase, decrease, noeffect x = \text{All}; y = \text{Protein}; r = \text{PhosphoProtein}
```

PRA(x,y,r,effect)

(Phosphorylation Radioactive Assay)

Par PRA, on vérifie si l'ajout de la molécule x induit un effet sur la phosphorylation de la protéine y qui est suivie grâce à la forme radioactive r de la forme phosphorylée de y.

```
effect = increase, decrease, noeffect x = \text{All }; y = \text{Protein }; r = \text{PhosphoProtein}
```

PRENYLPROTEIN(x)

(Protéine prénylée)

La molécule *x* est une protéine prénylée.

PROTONYLPROTEIN(x)

(*Protéine protonée*)

La molécule *x* est une protéine protonée.

R

RADIOLABELLEDFORM(r,b)

(Forme radioactive)

La molécule r est la forme radioactive de la molécule b.

r = All; x = All

REGDNA(x)

(Région d'ADN régulatrice)

La molécule *x* est une région d'ADN régulatrice.

RELATIONINHIBITOR(i,Relation)

(Inhibiteur de relation)

La molécule i est un inhibiteur de relation par exemple transcription ou traduction.

Relation = TRANSCRIPTION, TRANSLATION, PHOSPHORYLATE ...

i = Protein, Complex, SmallMolecule

RIA(y,a,effect)

(RadioImmunoAssay)

Par RIA, on vérifie si l'ajout de la molécule y a un effet sur la quantité de la molécule ciblée par l'anticorps a.

effect = increase, decrease, noeffect

y = All; a = Antibody

RIBOZYME(x)

(Ribozyme)

La molécule x est un ribozyme.

RIPANWB(x,a,i,effect)

(Western Blot en présence d'un inhibiteur de relation)

Le *western blot* montre l'effet que peut avoir un inhibiteur de relation (transcription ou traduction) *i* sur l'expression d'une protéine sur laquelle agit la molécule *x* en suivant l'anticorps pan *a* dirigé contre cette protéine.

effect = increase, decrease, noeffect

x = All; a = Antibody; i = Protein, Complex, SmallMolecule

RL(r,compartment,effect)

(RadioLocalisation)

Par RL, une molécule est localisée ou pas dans un compartiment compartment grâce à sa forme radioactive r.

compartment = PlasmaMembrane, CytoplasmicVesicles, Cytoplasm, ...

effect = positive, negative

r = Protein, Complex, SmallMolecule

RNA(x)

(ARN)

La molécule x est un ARN.

RRNA(x)

(ARN ribosomal)

La molécule *x* est un ARN ribosomal.

RTPCR(x,y,effect)

(Reverse Transcription Polymerase Chain Reaction)

Par *RT-PCR*, on vérifie si l'ajout de la molécule *x* induit un effet sur l'expression de l'ARN *y*.

effect = increase, decrease, noeffect

x = All; y = RNA

S

SCPPA(x,y,s,a,effect)

(Phosphorylation Assay en présence d'un siRNA)

Le phosphorylation assay montre l'effet que peut avoir un siRNA s sur la phosphorylation de la protéine y sur laquelle agit la molécule x, suivi à l'aide d'un anticorps a de la forme phosphorylée de y.

effect = increase, decrease, noeffect

x = All; y = Protein; a = Antibody; s = siRNA

SILENCERBIND(x)

(Site de fixation de silencers)

La molécule *x* est un site de fixation de *silencers*.

SIMPLEMOLECULE(**x**)

(Molécule simple)

La molécule b est une molécule simple.

SIRNA(x)

(siARN)

La molécule x est un siARN.

SIRNAAGAINST(s,z)

(siRNA)

s est un siRNA bloquant l'expression de la protéine z.

z = Protein; s = siRNA

SMALLMOLECULE(x)

(Petite molécule)

La molécule *x* est une petite molécule.

SRNA(x)

(small ARN)

La molécule x est un petit ARN.

SULFATYLPROTEIN(x)

(Protéine sulfatée)

La molécule x est une protéine sulfatée.

TATABOX(x)

(Boîte TATA)

La molécule *x* est une boîte TATA.

TCE(x,z,y,distance,status,effect)

(Effet sur la transcription)

La molécule x a un effet sur la transcription du gène z en ARN y.

effect = increase, decrease, noeffect

status = confirmed, invalidated, hypothesis

distance = direct, indirect, unknown

x = All; z = Gene; y = RNA

TFBINDING(x)

(Site de fixation de facteurs de transcription)

La molécule *x* est un site de fixation de facteurs de transcription.

TLE(x,w,y,distance,status,effect)

(Effet sur la traduction)

La molécule x a un effet sur la traduction de l'ARNm w en protéine y.

effect = increase, decrease, noeffect

status = confirmed, invalidated, hypothesis

distance = direct, indirect, unknown

x = All; w = mRNA; y = Protein

TRANSCRIPTION(z,y)

(Transcription)

Le gène z est transcrit en ARN y.

z = CodingGene; y = mRNA

TRANSLATION(w,y)

(Traduction)

L'ARNm w est traduit en protéine y.

w = mRNA; y = Protein

TRNA(x)

(ARN de transfert)

La molécule *x* est un ARN de transfert.

U

UBIQUITYLPROTEIN(x)

(*Protéine ubiquitinée*)

La molécule x est une protéine ubiquitinée.

Annexe 4 : Liste des règles

Les noms attribués aux différentes règles représentent uniquement le prédicat principal utilisé.

Acetylation Assay

• IF AA(x,y,a,e) AND ANTIBODYAGAINST(a,b) AND ACETYLFORM(b,y,pos) THEN ACETYLATE(x,y,b,unknown,confirmed,e)

Binding Domain PullDown

• IF BDPULLDOWN(x,[y,...]) THEN CO([(?,x),(?,y),...],(?)x:(?)y:...,confirmed)

Enzymatic assay

• IF $EA(x,[(n_y,y),...1],[(n_z,z),...2],effect)$ THEN $CATALYZE(x,[(n_y,y),...1],[(n_z,z),...2],direct,confirmed,effect)$

Enzymatic assay en présence d'un antagoniste

• IF CATALYZE(x,[(n_y,y),...1],[(n_z,z),...2],direct,confirmed,increase) AND ACPEA(x,[(n_y,y),...1],[(n_z,z),...2],i,decrease) AND ANTAGONIST(i,w) AND MODIFIEDFORM(x,c) THEN MODIFY(w,c,x,unknown,confirmed,increase) AND CATALYZE(c,[(n_y,y),...1],[(n_z,z),...2],direct,confirmed,noeffect)

Enzymatic assay en présence d'un inhibiteur

• IF $CATALYZE(x,[(n_y,y),...1],[(n_z,z),...2],direct,confirmed,increase)$ AND $ICPEA(x,[(n_y,y),...1],[(n_z,z),...2],i,decrease)$ AND INHIBITOR(i,w) AND MODIFIEDFORM(x,c) THEN MODIFY(w,c,x,unknown,confirmed,increase) AND $CATALYZE(c,[(n_y,y),...1],[(n_z,z),...2],direct,confirmed,noeffect)$

ELISA

• IF ELISA(y,a,e) AND ANTIBODYAGAINST(a,x) AND MODIFIEDFORM(x,l) THEN MODIFY(y,l,x,unknown,hypothesis,e)

ELISA en présence d'un inhibiteur

- IF MODIFY(y,l,x,unknown,hypothesis,increase) AND ICELISA(y,a,i,decrease) AND ANTIBODYAGAINST(a,x) AND INHIBITOR(i,z) AND MODIFIEDFORM(y,c) AND MODIFIEDFORM(z,d) THEN MODIFY(z,l,x,unknown,hypothesis,increase) AND MODIFY(y,d,z,unknown,hypothesis,increase) AND MODIFY(z,c,y,unknown,hypothesis,increase) AND MODIFY(d,l,x,unknown,hypothesis,noeffect) AND MODIFY(c,l,x,unknown,hypothesis,noeffect)
- IF MODIFY(y,l,x,unknown,confirmed,increase) AND ICELISA(y,a,i,decrease) AND ANTIBODYAGAINST(a,x) AND INHIBITOR(i,z) AND MODIFIEDFORM(y,c) AND MODIFIEDFORM(z,d) THEN MODIFY(z,l,x,unknown,confirmed,increase)

	AND MODIFY(y,d,z,unknown,hypothesis,increase)	AND
	MODIFY(z,c,y,unknown,hypothesis,increase)	AND
	MODIFY(d,l,x,unknown,hypothesis,noeffect)	AND
	MODIFY(c,l,x,unknown,hypothesis,noeffect)	
•	IF MODIFY(y,l,x,unknown,hypothesis,increase) AND ICELISA(y,a,i,decrease)	AND
	ANTIBODYAGAINST(a,x) AND INHIBITOR(i,z) AND NOTMODIFIED(y)	
	MODIFIEDFORM(z,d) THEN MODIFY(z,l,x,unknown,hypothesis,increase)	AND
	MODIFY(y,d,z,unknown,hypothesis,increase)	AND
	MODIFY(d,1,x,unknown,hypothesis,noeffect)	711112
•	IF MODIFY(y,l,x,unknown,confirmed,increase) AND ICELISA(y,a,i,decrease)	AND
•	ANTIBODYAGAINST(a,x) AND INHIBITOR(i,z) AND NOTMODIFIED(y)	
	MODIFIEDFORM(z,d) THEN MODIFY(z,l,x,unknown,confirmed,increase)	AND
		AND
	MODIFY(y,d,z,unknown,hypothesis,increase)	AND
	MODIFY(d,1,x,unknown,hypothesis,noeffect)	
•	IF MODIFY(y,l,x,unknown,hypothesis,increase) AND ICELISA(y,a,i,noeffect)	
		THEN
	MODIFY(z,l,x,unknown,confirmed,noeffect)	
•	IF MODIFY(y,l,x,unknown,hypothesis,increase) AND ICELISA(y,a,i,increase)	
	ANTIBODYAGAINST(a,x) AND INHIBITOR(i,z) AND MODIFIEDFORM	M(n,y)
	AND MODIFIEDFORM(m,z) THEN MODIFY(z,l,x,unknown,hypothesis,dec	rease)
	AND MODIFY(z,y,n,unknown,hypothesis,increase)	AND
	MODIFY(y,z,m,unknown,hypothesis,increase)	AND
	MODIFY(n,l,x,unknown,confirmed,noeffect)	AND
	MODIFY(m,l,x,unknown,confirmed,noeffect)	
•	IF MODIFY(y,1,x,unknown,confirmed,increase) AND ICELISA(y,a,i,increase)	AND
	ANTIBODYAGAINST(a,x) AND INHIBITOR(i,z) AND MODIFIEDFORM	
	AND MODIFIEDFORM(m,z) THEN MODIFY(z,l,x,unknown,confirmed,dec	
	AND MODIFY(z,y,n,unknown,hypothesis,increase)	AND
	MODIFY(y,z,m,unknown,hypothesis,increase)	AND
	MODIFY(n,l,x,unknown,confirmed,noeffect)	AND
	MODIFY(m,l,x,unknown,confirmed,noeffect)	
	IF MODIFY(y,l,x,unknown,hypothesis,increase) AND ICELISA(y,a,i,increase)	AND
•	ANTIBODYAGAINST(a,x) AND INHIBITOR(i,z) AND MODIFIEDFORM	
		` ' '
	AND NOTMODIFIED(y) THEN MODIFY(z,l,x,unknown,hypothesis,decrease)	
	MODIFY(y,z,m,unknown,hypothesis,increase)	AND
	MODIFY(m,l,x,unknown,confirmed,noeffect)	
•	IF MODIFY(y,1,x,unknown,confirmed,increase) AND ICELISA(y,a,i,increase)	
	ANTIBODYAGAINST(a,x) AND INHIBITOR(i,z) AND MODIFIEDFORM	
	AND NOTMODIFIED(y) THEN MODIFY(z,l,x,unknown,confirmed,decrease)	
	MODIFY(y,z,m,unknown,hypothesis,increase)	AND
	MODIFY(m,l,x,unknown,confirmed,noeffect)	

FRET

• IF FRET(y,z) THEN CO([(?,y),(?,z)],(?)y:(?)z,confirmed)

GST-PullDown

• IF GSTPULLDOWN(x,[y,...]) THEN CO([(?,x),(?,y),...],(?)x:(?)y:...,confirmed)

ImmunoFluorescence

- IF FLUO(a,compartment,positive) AND ANTIBODYAGAINST(a,x) THEN LOCALIZATION(x,compartment)
- IF FLUO(a,compartment,negative) AND ANTIBODYAGAINST(a,x) THEN NOLOCALIZATION(x,compartment)
- IF FLUO(a,compartment,positive) AND PANANTIBODY(a,x) THEN LOCALIZATION(x,compartment)
- IF FLUO(a,compartment,negative) AND PANANTIBODY(a,x) THEN NOLOCALIZATION(x,compartment)

ImmunoHistoChimie

- IF IHC(a,compartment,positive) AND ANTIBODYAGAINST(a,x) THEN LOCALIZATION(x,compartment)
- IF IHC(a,compartment,negative) AND ANTIBODYAGAINST(a,x) THEN NOLOCALIZATION(x,compartment)

Immunoprécipitation

- IF IP([x,y,...],increase) AND THEN CO([(?,x),(?,y),...],(?)x:(?)y:...,confirmed)
- IF IPD(z,[x,y,...],increase) THEN ECO(z,[(?,x),(?,y),...],(?)x:(?)y:...,unknown,confirmed,increase) AND CO([(?,x),(?,y),...],(?)x:(?)y:...,confirmed)
- IF IPD(z,[x,y,...],decrease) THEN ECO(z,[(?,x),(?,y),...],(?)x:(?)y:...,unknown,confirmed,decrease)
- IF IPD(z,[x,y,...],noeffect) THEN ECO(z,[(?,x),(?,y),...],(?)x:(?)y:...,unknown,confirmed,noeffect)

Phosphorylation Assay

• IF PA(x,y,a,e) AND ANTIBODYAGAINST(a,b) AND PHOSPHOFORM(b,y,pos) THEN PHOSPHORYLATE(x,y,b,unknown,confirmed,e)

Phosphorylation Assay en présence d'un inhibiteur

- PHOSPHORYLATE(x,y,b,unknown,confirmed,increase) AND ICPPA(x,y,i,a,decrease) AND INHIBITOR(i,z) AND ANTIBODYAGAINST(a,b) MODIFIEDFORM(z,d) AND MODIFIEDFORM(x,c)THEN AND PHOSPHORYLATE(z,v,b,unknown,confirmed,increase) **AND** MODIFY(x,d,z,unknown,hypothesis,increase) **AND** MODIFY(z,c,x,unknown,hypothesis,increase) AND PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect) **AND** PHOSPHORYLATE(c,y,b,unknown,confirmed,noeffect)
- IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase) AND ICPPA(x,y,i,a,decrease) AND INHIBITOR(i,z) AND ANTIBODYAGAINST(a,b) AND MODIFIEDFORM(z,d) AND NOTMODIFIED(x) THEN PHOSPHORYLATE(z,y,b,unknown,confirmed,increase) AND MODIFY(x,d,z,unknown,hypothesis,increase) AND PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)

•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase)	AND
	ICPPA(x,y,i,a,noeffect) AND INHIBITOR(i,z) AND ANTIBODYAGAI	NST(a,b)
	THEN PHOSPHORYLATE(z,y,b,unknown,confirmed,noeffect)	
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase)	AND
	ICPPA(x,y,i,a,increase) AND INHIBITOR(i,z) AND ANTIBODYAGAI	NST(a,b)
	AND MODIFIEDFORM(z,d) AND MODIFIEDFORM(x,c)	THEN
	PHOSPHORYLATE(z,y,b,unknown,confirmed,decrease)	AND
	MODIFY(x,d,z,unknown,hypothesis,decrease)	AND
	MODIFY(z,c,x,unknown,hypothesis,decrease)	AND
	PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)	AND
	PHOSPHORYLATE(c,y,b,unknown,confirmed,noeffect)	
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase)	AND
	ICPPA(x,y,i,a,increase) AND INHIBITOR(i,z) AND ANTIBODYAGAI	NST(a,b)
	AND MODIFIEDFORM(z,d) AND NOTMODIFIED(x)	THEN
	PHOSPHORYLATE(z,y,b,unknown,confirmed,decrease)	AND
	MODIFY(x,d,z,unknown,hypothesis,decrease)	AND
	PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)	
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,decrease)	AND
	ICPPA(x,y,i,a,increase) AND INHIBITOR(i,z) AND ANTIBODYAGAI	
	AND MODIFIEDFORM(x,c) AND MODIFIEDFORM(z,d)	THEN
	PHOSPHORYLATE(z,y,b,unknown,confirmed,decrease)	AND
	MODIFY(x,d,z,unknown,hypothesis,increase)	AND
	MODIFY(z,c,x,unknown,hypothesis,increase)	AND
	PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)	AND
	PHOSPHORYLATE(c,y,b,unknown,confirmed,noeffect)	
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,decrease)	AND
	ICPPA(x,y,i,a,increase) AND INHIBITOR(i,z) AND ANTIBODYAGAI	NST(a,b)
	AND NOTMODIFIED(x) AND MODIFIEDFORM(z,d)	THEN
	PHOSPHORYLATE(z,y,b,unknown,confirmed,decrease)	AND
	MODIFY(x,d,z,unknown,hypothesis,increase)	AND
	PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)	
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,decrease)	AND
	ICPPA(x,y,i,a,noeffect) AND INHIBITOR(i,z) AND ANTIBODYAGAI	NST(a,b)
	THEN PHOSPHORYLATE(z,y,b,unknown,confirmed,noeffect)	
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,decrease)	AND
	ICPPA(x,y,i,a,decrease) AND INHIBITOR(i,z) AND ANTIBODYAGAI	NST(a,b)
	AND $MODIFIEDFORM(x,c)$ AND $MODIFIEDFORM(z,d)$	THEN
	PHOSPHORYLATE(z,y,b,unknown,confirmed,increase)	AND
	MODIFY(z,c,x,unknown,hypothesis,decrease)	AND
	MODIFY(x,d,z,unknown,hypothesis,decrease)	AND
	PHOSPHORYLATE(c,y,b,unknown,confirmed,noeffect)	AND
	PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)	
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,decrease)	AND
	ICPPA(x,y,i,a,decrease) AND INHIBITOR(i,z) AND ANTIBODYAGAI	NST(a,b)
	AND NOTMODIFIED(x) AND MODIFIEDFORM(z,d)	THEN
	PHOSPHORYLATE(z,y,b,unknown,confirmed,increase)	AND
	MODIFY(x,d,z,unknown,hypothesis,decrease)	AND
	PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)	

Phosphorylation Assay en présence d'un antagoniste

1 hospitol viation Assay en presence a un antagoniste		
• IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase)	AND	
ACPPA(x,y,i,a,decrease) AND ANTAGONIST(i,z)	AND	
ANTIBODYAGAINST(a,b) AND MODIFIEDFORM(x,c)	AND	
MODIFIEDFORM(z,d)	THEN	
PHOSPHORYLATE(z,y,b,unknown,confirmed,increase)	AND	
MODIFY(x,d,z,unknown,hypothesis,increase)	AND	
MODIFY(z,c,x,unknown,hypothesis,increase)	AND	
PHOSPHORYLATE(d,y,b,unknown,hypothesis,noeffect)	AND	
PHOSPHORYLATE(c,y,b,unknown,hypothesis,noeffect)		
• IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase)	AND	
ACPPA(x,y,i,a,decrease) AND ANTAGONIST(i,z)	AND	
ANTIBODYAGAINST(a,b) AND NOTMODIFIED(x) AND MODIFIEDFO	RM(z,d)	
THEN PHOSPHORYLATE(z,y,b,unknown,confirmed,increase)	AND	
MODIFY(x,d,z,unknown,hypothesis,increase)	AND	
PHOSPHORYLATE(d,y,b,unknown,hypothesis,noeffect)		
Dhogahogulation Aggay on magagago dina siDNA		
Phosphorylation Assay en présence d'un siRNA		
• IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase)	AND	
SCPPA(x,y,s,a,decrease) AND $SIRNAAGAINST(s,z)$	AND	
ANTIBODYAGAINST(a,b) AND MODIFIEDFORM(z,d)	AND	
MODIFIEDFORM(x,c)	THEN	
PHOSPHORYLATE(z,y,b,unknown,hypothesis,increase)	AND	
MODIFY(x,d,z,unknown,hypothesis,increase)	AND	

PHOSPHORYLATE(c,y,b,unknown,confirmed,noeffect)
 IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase)
 SCPPA(x,y,s,a,decrease) AND SIRNAAGAINST(s,z)

SCPPA(x,y,s,a,decrease) AND SIRNAAGAINST(s,z) AND ANTIBODYAGAINST(a,b) AND MODIFIEDFORM(z,d) AND NOTMODIFIED(x) THEN PHOSPHORYLATE(z,y,b,unknown,hypothesis,increase) AND

AND

AND

AND

MODIFY(x,d,z,unknown,hypothesis,increase)

AND

PHOSPHORYLATE(d,y,b,unknown,hypothesis,noeffect)

PHOSPHORYLATE(d,y,b,unknown,hypothesis,noeffect)

MODIFY(z,c,x,unknown,hypothesis,increase)

MODIFY(x,d,z,unknown,hypothesis,decrease)

AND

MODIFY(z,c,x,unknown,hypothesis,decrease)
PHOSPHORYLATE(d,y,b,unknown,hypothesis,noeffect)
AND

PHOSPHORYLATE(c,y,b,unknown,confirmed,noeffect)

• IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase) AND SCPPA(x,y,s,a,increase) AND SIRNAAGAINST(s,z) AND

ANTIBODYAGAINST(a,b) AND MODIFIEDFORM(z,d) AND NOTMODIFIED(x)

THEN PHOSPHORYLATE(z,y,b,unknown,hypothesis,decrease) AND

MODIFY(x,d,z,unknown,hypothesis,decrease) AND

PHOSPHORYLATE(d,y,b,unknown,hypothesis,noeffect)

IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase) AND SCPPA(x,y,s,a,noeffect) AND SIRNAAGAINST(s,z) AND ANTIBODYAGAINST(a,b) THEN PHOSPHORYLATE(z,y,b,unknown,confirmed,noeffect)

Phosphorylation Radioactive Assay

• IF PRA(x,y,r,e) AND RADIOLABELLEDFORM(r,b) AND PHOSPHOFORM(b,y,pos) THEN PHOSPHORYLATE(x,y,b,unknown,confirmed,e)

Phosphorylation Radioactive Assay en présence d'un inhibiteur

•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase)	AND	
	ICPPRA(x,y,i,r,decrease) AND INHIBITOR(i,z)	AND	
	RADIOLABELLEDFORM(r,b) AND MODIFIEDFORM(z,d)	AND	
	MODIFIEDFORM(x,c)	THEN	
	PHOSPHORYLATE(z,y,b,unknown,confirmed,increase)	AND	
	MODIFY(x,d,z,unknown,hypothesis,increase)	AND	
	MODIFY(z,c,x,unknown,hypothesis,increase)	AND	
	PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)	AND	
	PHOSPHORYLATE(c,y,b,unknown,confirmed,noeffect)		
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase)	AND	
	ICPPRA(x,y,i,r,decrease) AND INHIBITOR(i,z)	AND	
	RADIOLABELLEDFORM(r,b) AND MODIFIEDFORM(z,d)	AND	
	NOTMODIFIED(x) THEN PHOSPHORYLATE(z,y,b,unknown,confirmed	d,increase)	
	AND MODIFY(x,d,z,unknown,hypothesis,increase)	AND	
	PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)		
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase)	AND	
	ICPPRA(x,y,i,r,noeffect) AND $INHIBITOR(i,z)$	AND	
	RADIOLABELLEDFORM(r,b)	THEN	
	PHOSPHORYLATE(z,y,b,unknown,confirmed,noeffect)		
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase)	AND	
	ICPPRA(x,y,i,r,increase) AND INHIBITOR(i,z)	AND	
	RADIOLABELLEDFORM(r,b) AND MODIFIEDFORM(z,d)	AND	
	MODIFIEDFORM(x,c)	THEN AND	
	PHOSPHORYLATE(z,y,b,unknown,confirmed,decrease)		
	MODIFY(x,d,z,unknown,hypothesis,decrease)	AND	
	MODIFY(z,c,x,unknown,hypothesis,decrease)	AND	
	PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)	AND	
	PHOSPHORYLATE(c,y,b,unknown,confirmed,noeffect)		
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase)	AND	
	ICPPRA(x,y,i,r,increase) AND INHIBITOR(i,z)	AND	
	RADIOLABELLEDFORM(r,b) AND MODIFIEDFORM(z,d)	AND	
	NOTMODIFIED(x) THEN PHOSPHORYLATE(z,y,b,unknown,confirmed		
	AND MODIFY(x,d,z,unknown,hypothesis,decrease)	AND	
	PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)		
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,decrease)	AND	
	ICPPRA(x,y,i,r,increase) AND INHIBITOR(i,z)	AND	
	RADIOLABELLEDFORM(r,b) AND MODIFIEDFORM(x,c)	AND	
	MODIFIEDFORM(z,d)	THEN	

	PHOSPHORYLATE(z,y,b,unknown,confirmed,decrease)	AND
	MODIFY(x,d,z,unknown,hypothesis,increase)	AND
	MODIFY(z,c,x,unknown,hypothesis,increase)	AND
	PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)	AND
	PHOSPHORYLATE(c,y,b,unknown,confirmed,noeffect)	
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,decrease)	AND
	ICPPRA(x,y,i,r,increase) AND INHIBITOR(i,z)	AND
	RADIOLABELLEDFORM(r,b) AND NOTMODIFIED(x)	AND
	MODIFIEDFORM(z,d)	THEN
	PHOSPHORYLATE(z,y,b,unknown,confirmed,decrease)	AND
	MODIFY(x,d,z,unknown,hypothesis,increase)	AND
	PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)	
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,decrease)	AND
	ICPPRA(x,y,i,r,noeffect) AND INHIBITOR(i,z)	AND
	RADIOLABELLEDFORM(r,b)	THEN
	PHOSPHORYLATE(z,y,b,unknown,confirmed,noeffect)	
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,decrease)	AND
	ICPPRA(x,y,i,r,decrease) AND INHIBITOR(i,z)	AND
	RADIOLABELLEDFORM (r,b) AND MODIFIEDFORM (x,c)	AND
	MODIFIEDFORM(z,d)	THEN
	PHOSPHORYLATE(z,y,b,unknown,confirmed,increase)	AND
	MODIFY(z,c,x,unknown,hypothesis,decrease)	AND
	MODIFY(x,d,z,unknown,hypothesis,decrease)	AND
	PHOSPHORYLATE(c,y,b,unknown,confirmed,noeffect)	AND
	PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)	
•	IF PHOSPHORYLATE(x,y,b,unknown,confirmed,decrease)	AND
	ICPPRA(x,y,i,r,decrease) AND $INHIBITOR(i,z)$	AND
	RADIOLABELLEDFORM (r,b) AND NOTMODIFIED (x)	AND
	MODIFIEDFORM(z,d)	THEN
	PHOSPHORYLATE(z,y,b,unknown,confirmed,increase)	AND
	MODIFY(x,d,z,unknown,hypothesis,decrease)	AND
	PHOSPHORYLATE(d,y,b,unknown,confirmed,noeffect)	

RadioImmunoAssay

• IF RIA(y,a,e) AND ANTIBODYAGAINST(a,x) AND MODIFIEDFORM(x,l) THEN MODIFY(y,l,x,unknown,hypothesis,e)

RadioImmunoAssay en présence d'un inhibiteur

- IF MODIFY(y,l,x,unknown,hypothesis,increase) AND ICRIA(y,a,i,decrease) AND ANTIBODYAGAINST(a,x) AND INHIBITOR(i,z) AND MODIFIEDFORM(y,c) AND MODIFIEDFORM(z,d) THEN MODIFY(z,l,x,unknown,hypothesis,increase) AND MODIFY(y,d,z,unknown,hypothesis,increase) AND MODIFY(z,c,y,unknown,hypothesis,increase) AND MODIFY(d,l,x,unknown,confirmed,noeffect) AND MODIFY(c,l,x,unknown,confirmed,noeffect)
- IF MODIFY(y,l,x,unknown,confirmed,increase) AND ICRIA(y,a,i,decrease) AND ANTIBODYAGAINST(a,x) AND INHIBITOR(i,z) AND MODIFIEDFORM(y,c) AND MODIFIEDFORM(z,d) THEN MODIFY(z,l,x,unknown,confirmed,increase)

	AND MODIFY(y,d,z,unknown,hypothesis,increase)	AND
	MODIFY(z,c,y,unknown,hypothesis,increase)	AND
	MODIFY(d,l,x,unknown,confirmed,noeffect)	AND
	MODIFY(c,l,x,unknown,confirmed,noeffect)	
•	IF MODIFY(y,l,x,unknown,hypothesis,increase) AND ICRIA(y,a,i,decrease)	AND
		AND
	MODIFIEDFORM(z,d) THEN MODIFY(z,l,x,unknown,hypothesis,increase)	AND
	MODIFY(y,d,z,unknown,hypothesis,increase)	AND
	MODIFY(d,l,x,unknown,confirmed,noeffect)	111,12
•	IF MODIFY(y,l,x,unknown,confirmed,increase) AND ICRIA(y,a,i,decrease)	AND
	ANTIBODYAGAINST(a,x) AND INHIBITOR(i,z) AND NOTMODIFIED(y)	AND
	MODIFIEDFORM(z,d) THEN MODIFY(z,l,x,unknown,confirmed,increase)	AND
	MODIFY(y,d,z,unknown,hypothesis,increase)	AND
	MODIFY(d,l,x,unknown,confirmed,noeffect)	
•	IF MODIFY(y,l,x,unknown,hypothesis,increase) AND ICRIA(y,a,i,noeffect)	AND
		ΓHEN
	MODIFY(z,l,x,unknown,confirmed,noeffect)	III I
•	IF MODIFY(y,l,x,unknown,hypothesis,increase) AND ICRIA(y,a,i,increase)	AND
	ANTIBODYAGAINST(a,x) AND INHIBITOR(i,z) AND MODIFIEDFORM	
	AND MODIFIEDFORM(z,d) THEN MODIFY(z,l,x,unknown,hypothesis,dec	
	AND MODIFY(z,c,y,unknown,hypothesis,decrease)	AND
	MODIFY(y,d,z,unknown,hypothesis,decrease)	AND
	MODIFY(c,l,x,unknown,confirmed,noeffect)	AND
	MODIFY(d,1,x,unknown,confirmed,noeffect)	
•	IF MODIFY(y,l,x,unknown,confirmed,increase) AND ICRIA(y,a,i,increase)	AND
	ANTIBODYAGAINST(a,x) AND INHIBITOR(i,z) AND MODIFIEDFORM	
	AND MODIFIEDFORM(z,d) THEN MODIFY(z,l,x,unknown,confirmed,dec	
	AND MODIFY(z,c,y,unknown,hypothesis,decrease)	AND
	MODIFY(y,d,z,unknown,hypothesis,decrease)	AND
	MODIFY(c,l,x,unknown,confirmed,noeffect)	AND
	MODIFY(d,l,x,unknown,confirmed,noeffect)	
•	· · · · · · · · · · · · · · · · · · ·	AND
	ANTIBODYAGAINST(a,x) AND INHIBITOR(i,z) AND NOTMODIFIED(y)	
	MODIFIEDFORM(z,d) THEN MODIFY(z,l,x,unknown,hypothesis,decrease)	AND
	MODIFY(y,d,z,unknown,hypothesis,decrease)	AND
	MODIFY(d,1,x,unknown,confirmed,noeffect)	71112
•	IF MODIFY(y,1,x,unknown,confirmed,increase) AND ICRIA(y,a,i,increase)	AND
•		AND
	MODIFIEDFORM(z,d) THEN MODIFY(z,l,x,unknown,confirmed,decrease)	AND
	MODIFY(y,d,z,unknown,hypothesis,decrease)	AND
	MODIFY(d,l,x,unknown,confirmed,noeffect)	AND
	MODIFT (u,1,x,ulikilowii,coliilillicu,ilocifect)	

REASONING

- IF PHOSPHOFORM(b,y,pos) THEN MODIFIEDFORM(b,y)
- IF MODIFIEDFORM(x,y) THEN MODIFIEDFORM(y,x)
- IF ACETYLFORM(b,y,pos) THEN MODIFIEDFORM(b,y)
- IF SIRNAAGAINST(s,x) AND MODIFIEDFORM(y,x) THEN SIRNAAGAINST(s,y)

- IF PANANTIBODY(a,x) THEN ANTIBODYAGAINST(a,x)
- IF PANANTIBODY(a,x) AND MODIFIEDFORM(c,x) THEN PANANTIBODY(a,c)
- IF CO([(?,x),(?,y),...],(?)x:(?)y:...,confirmed) THEN COMPLEX((?)x:(?)y:...)
- IF MODIFY(y,l,x,d,confirmed,e) THEN MODIFIEDFORM(l,x)
- IF PHOSPHORYLATE(x,y,b,d,confirmed,e) AND AGONIST(x,z) THEN PHOSPHORYLATE(z,y,b,d,confirmed,e)
- IF MODIFY(x,y,b,d,s,e) AND AGONIST(x,z) THEN MODIFY(z,y,b,d,s,e)
- IF PHOSPHORYLATE(x,y,b,d,confirmed,e) AND ANALOGUE(x,z) THEN PHOSPHORYLATE(z,y,b,d,confirmed,e)
- IF MODIFY(x,y,b,d,s,e) AND ANALOGUE(x,z) THEN MODIFY(z,y,b,d,s,e)
- IF LOCALIZATION(y,cytosol) AND NOLOCALIZATION(y,extracellular) AND NOLOCALIZATION(y,plasmamembrane) AND LOCALIZATION(x,extracellular) AND NOLOCALIZATION(x,plasmamembrane) AND NOLOCALIZATION(x,cytosol) THEN COMPARTDIST(x,y,distant)
- IF LOCALIZATION(y,nucleus) AND NOLOCALIZATION(y,cytosol) AND NOLOCALIZATION(y,extracellular) AND NOLOCALIZATION(y,plasmamembrane) AND LOCALIZATION(x,extracellular) AND NOLOCALIZATION(x,plasmamembrane) AND NOLOCALIZATION(x,cytosol) AND NOLOCALIZATION(x,nucleus) THEN COMPARTDIST(x,y,distant)
- IF PHOSPHORYLATE(x,y,b,unknown,confirmed,increase) AND PHOSPHORYLATE(b,z,d,unknown,confirmed,increase) THEN PHOSPHORYLATE(x,z,d,indirect,confirmed,increase)
- IF PHOSPHORYLATE(x,y,b,unknown,confirmed,e) AND COMPARTDIST(x,y,distant) THEN PHOSPHORYLATE(x,y,b,indirect,confirmed,e) AND PHOSPHORYLATE(x,z,d,direct,invalidated,increase)
- IF MODIFY(x,y,b,unknown,hypothesis,e) AND COMPARTDIST(x,y,distant) THEN MODIFY(x,y,b,indirect,hypothesis,e)
- IF COMPLEX((nx)x:(ny)y:(nz)z:...) AND EXCHANGE(x,(nx)x:(ny)y:(nz)z:...,(nz,z),(na,a)) THEN COMPLEX((nx)x:(ny)y:(na)a:...)
- IF MODIFY(x,z,y,unknown,confirmed,increase) AND CATALYZE(e,[(nz,z),...1],[(ny,y),...2],direct,confirmed,increase) AND MODIFIEDFORM(e,c) THEN MODIFY(x,c,e,unknown,hypothesis,increase) AND CATALYZE(c,[(nz,z),...1],[(ny,y),...2],direct,hypothesis,noeffect) AND MODIFY(x,z,y,indirect,hypothesis,increase)
- IF CATALYZE(x,[(ny,y)],[(nz,z)],direct,confirmed,increase) AND PHOSPHOFORM(z,y,pos) THEN PHOSPHORYLATE(x,y,z,direct,confirmed,increase)
- IF MODIFY(x,y,z,d,s,increase) THEN MODIFY(x,z,y,d,s,decrease)
- IF MODIFY(x,y,z,d,s,noeffect) THEN MODIFY(x,z,y,d,s,noeffect)
- IF MODIFY(x,y,z,d,s,decrease) THEN MODIFY(x,z,y,d,s,increase)
- IF PHOSPHORYLATE(x,y,z,d,s,e) THEN MODIFY(x,y,z,d,s,e)

Reverse-Transcriptase Polymerase Chain Reaction

• IF RTPCR(x,y,e) AND TRANSCRIPTION(z,y) THEN TCE(x,z,y,unknown,confirmed,e)

Reverse-Transcriptase Polymerase Chain Reaction en présence d'un inhibiteur

•	ir TCE(x,z,y,ulkilowii,collillilled,illclease) AND IRTFCK(x,y,i,declease)	AND
	INHIBITOR(i,p) AND MODIFIEDFORM(x,c) AND MODIFIEDFORM(p,d)	THEN
	TCE(p,z,y,unknown,hypothesis,increase)	AND
	MODIFY(x,d,p,unknown,hypothesis,increase)	AND
	MODIFY(p,c,x,unknown,hypothesis,increase)	AND
	TCE(d,z,y,unknown,confirmed,noeffect)	AND
	TCE(c,z,y,unknown,confirmed,noeffect)	
•	IF TCE(x,z,y,unknown,confirmed,increase) AND IRTPCR(x,y,i,decrease)	AND
	INHIBITOR(i,p) AND NOTMODIFIED(x) AND MODIFIEDFORM(p,d)	THEN
	TCE(p,z,y,unknown,confirmed,increase)	AND
	MODIFY(x,d,p,unknown,hypothesis,increase)	AND
	TCE(d,z,y,unknown,confirmed,noeffect)	
•	IF TCE(x,z,y,unknown,confirmed,increase) AND IRTPCR(x,y,i,increase)	AND
	INHIBITOR(i,p) AND MODIFIEDFORM(x,c) AND MODIFIEDFORM(p,d)	THEN
	TCE(p,z,y,unknown,confirmed,decrease)	AND
	MODIFY(p,x,c,unknown,hypothesis,increase)	AND
	MODIFY(x,p,d,unknown,hypothesis,increase)	AND
	TCE(c,z,y,unknown,confirmed,noeffect)	AND

IF TCF(x z y unknown confirmed increase) AND IRTPCR(x y i decrease) AND

- IF TCE(x,z,y,unknown,confirmed,increase) AND IRTPCR(x,y,i,increase) AND INHIBITOR(i,p) AND NOTMODIFIED(x) AND MODIFIEDFORM(p,d) THEN TCE(p,z,y,unknown,confirmed,decrease) AND MODIFY(x,p,d,unknown,hypothesis,increase) AND TCE(d,z,y,unknown,confirmed,noeffect)
- IF TCE(x,z,y,unknown,confirmed,increase) AND IRTPCR(x,y,i,noeffect) AND INHIBITOR(i,p) THEN TCE(p,z,y,unknown,confirmed,noeffect)

Structure cristalline tridimensionnelle

• IF 3D([(nx,x),(ny,y)]) THEN CO([(nx,x),(ny,y),...],(nx)x:(ny)y:...,confirmed)

Western Blot avec utilisation d'un anticorps Pan

TCE(d,z,y,unknown,confirmed,noeffect)

- IF PANWB(x,a,e) AND TRANSCRIPTION(z,w) AND TRANSLATION(w,y) AND PANANTIBODY(a,y) THEN TCE(x,z,w,unknown,hypothesis,e)
- IF PANWB(x,a,e) AND TRANSLATION(w,y) AND PANANTIBODY(a,y) THEN TLE(x,w,y,unknown,hypothesis,e)
- IF PANWB(x,a,e) AND TRANSLATION(w,y) AND PANANTIBODY(a,y) THEN DEGRADATION(x,w,unknown,hypothesis,e)
- IF PANWB(x,a,e) AND TRANSLATION(w,y) AND PANANTIBODY(a,y) THEN DEGRADATION(x,y,unknown,hypothesis,e)

Western Blot avec utilisation d'un anticorps Pan en présence d'un inhibiteur

IF TCE(x,v,w,unknown,hypothesis,increase) AND ICPANWB(x,a,i,decrease) AND INHIBITOR(i,z) AND PANANTIBODY(a,y) AND MODIFIEDFORM(x,c) AND MODIFIEDFORM(z,d) AND TRANSLATION(w,y) THEN TCE(z,v,w,unknown,hypothesis,increase) AND

	MODIFY(x,d,z,unknown,hypothesis,increase)	AND
	MODIFY(z,c,x,unknown,hypothesis,increase)	AND
	TCE(d,v,w,unknown,confirmed,noeffect)	AND
	TCE(c,v,w,unknown,confirmed,noeffect)	
•	IF TLE(x,w,y,unknown,hypothesis,increase) AND ICPANWB(x,a,i,decrease)	AND
	INHIBITOR(i,z) AND PANANTIBODY(a,y) AND MODIFIEDFORM(x,c)	
	MODIFIEDFORM(z,d) THEN TLE(z,w,y,unknown,hypothesis,increase)	AND
	MODIFY(x,d,z,unknown,hypothesis,increase)	AND
	· · · · · · · · · · · · · · · · · · ·	
	MODIFY(z,c,x,unknown,hypothesis,increase)	AND
	TLE(d,w,y,unknown,confirmed,noeffect)	AND
	TLE(c,w,y,unknown,confirmed,noeffect)	
•	IF $TCE(x,v,w,unknown,hypothesis,increase)$ AND $ICPANWB(x,a,i,decrease)$	
	INHIBITOR(i,z) AND PANANTIBODY(a,y) AND NOTMODIFIED(x)	AND
	MODIFIEDFORM(z,d) AND TRANSLATION(w,y)	THEN
	TCE(z,v,w,unknown,hypothesis,increase)	AND
	MODIFY(x,d,z,unknown,hypothesis,increase)	AND
	TCE(d,v,w,unknown,confirmed,noeffect)	
•	IF TLE(x,w,y,unknown,hypothesis,increase) AND ICPANWB(x,a,i,decrease)	AND
	INHIBITOR(i,z) AND PANANTIBODY(a,y) AND NOTMODIFIED(x)	AND
	MODIFIEDFORM(z,d) THEN TLE(z,w,y,unknown,hypothesis,increase)	AND
	MODIFY(x,d,z,unknown,hypothesis,increase)	AND
	TLE(d,w,y,unknown,confirmed,noeffect)	THILD
	IF TCE(x,v,w,unknown,hypothesis,increase) AND ICPANWB(x,a,i,noeffect)	AND
•	INHIBITOR(i,z) AND PANANTIBODY(a,y) AND TRANSLATION(w,y)	
	TCE(z,v,w,unknown,confirmed,noeffect)	TILLIN
_		ANID
•	IF TLE(x,w,y,unknown,hypothesis,increase) AND ICPANWB(x,a,i,noeffect)	
	() J	THEN
	TLE(z,w,y,unknown,confirmed,noeffect)	
•	IF TCE(x,v,w,unknown,hypothesis,increase) AND ICPANWB(x,a,i,increase)	
	INHIBITOR(i,z) AND PANANTIBODY(a,y) AND MODIFIEDFORM(x,c)	
	() , , , , , , , , , , , , , , , , , ,	THEN
	TCE(z,v,w,unknown,hypothesis,decrease)	AND
	MODIFY(z,c,x,unknown,hypothesis,decrease)	AND
	MODIFY(x,d,z,unknown,hypothesis,decrease)	AND
	TCE(c,v,w,unknown,confirmed,noeffect)	AND
	TCE(d,v,w,unknown,confirmed,noeffect)	
•	IF TLE(x,w,y,unknown,hypothesis,increase) AND ICPANWB(x,a,i,increase)	AND
	INHIBITOR(i,z) AND PANANTIBODY(a,y) AND MODIFIEDFORM(x,c)	AND
	MODIFIEDFORM(z,d) THEN TLE(z,w,y,unknown,hypothesis,decrease)	AND
	MODIFY(z,c,x,unknown,hypothesis,decrease)	AND
	MODIFY(x,d,z,unknown,hypothesis,decrease)	AND
	TLE(c,w,y,unknown,confirmed,noeffect)	AND
	TLE(d,w,y,unknown,confirmed,noeffect)	
•	IF TCE(x,v,w,unknown,hypothesis,increase) AND ICPANWB(x,a,i,increase)	AND
-	INHIBITOR(i,z) AND PANANTIBODY(a,y) AND NOTMODIFIED(x)	AND
	* · · • · · · · · · · · · · · · · · · ·	THEN
	TCE(z,v,w,unknown,hypothesis,decrease)	AND
	MODIFY(x,d,z,unknown,hypothesis,decrease)	AND
	TCE(d,v,w,unknown,confirmed,noeffect)	AND
	1 CL(u, v, w, ulikilowii, Collii ilicu, ilociicu)	

IF TLE(x,w,y,unknown,hypothesis,increase) AND ICPANWB(x,a,i,increase) AND INHIBITOR(i,z) AND PANANTIBODY(a,y) AND NOTMODIFIED(x) AND MODIFIEDFORM(z,d) THEN TLE(z,w,y,unknown,hypothesis,decrease) AND MODIFY(x,d,z,unknown,hypothesis,decrease)
 TLE(d,w,y,unknown,confirmed,noeffect)

Western Blot avec utilisation d'un anticorps Pan en présence d'un inhibiteur de relation

- IF TCE(x,v,w,unknown,hypothesis,increase) AND RIPANWB(x,a,i,decrease) AND PANANTIBODY(a,y) AND RELATIONINHIBITOR(i,TRANSCRIPTION) AND TRANSLATION(w,y) THEN TCE(x,v,w,unknown,confirmed,increase)
- IF TLE(x,w,y,unknown,hypothesis,increase) AND RIPANWB(x,a,i,decrease) AND PANANTIBODY(a,y) AND RELATIONINHIBITOR(i,TRANSLATION) THEN TLE(x,w,y,unknown,confirmed,increase)
- IF TCE(x,v,w,unknown,hypothesis,increase) AND RIPANWB(x,a,i,noeffect) AND PANANTIBODY(a,y) AND RELATIONINHIBITOR(i,TRANSCRIPTION) AND TRANSLATION(w,y) THEN TCE(x,v,w,unknown,confirmed,noeffect)
- IF TLE(x,w,y,unknown,hypothesis,increase) AND RIPANWB(x,a,i,noeffect) AND PANANTIBODY(a,y) AND RELATIONINHIBITOR(i,TRANSLATION) THEN TLE(x,w,y,unknown,confirmed,noeffect)

Annexe 5 : Régulation de la transcription et de la traduction par la FSH

L'augmentation de traduction induite par la FSH peut être la conséquence de l'augmentation de la transcription. La FSH peut tout de même induire la traduction sans avoir d'effet sur la transcription, une thèse est en cours sur ce sujet au sein de l'équipe.

Tableau 7: Transcription induite par la FSH

Impliqué dans	Gène codant pour	Intermédiaire	Références
Stéroïdogenèse	3β-HSD		(Tosca,2005) (Escamilla-Hernandez,2008)
		PKA	(Saxena,2004)
	11β-HSD		(Tetsuka,2010)
	p450scc		(Doyle,2010)(Escamilla-
			Hernandez,2008)(Park,2005)
		PKA	(Wayne,2007)(Saxena,2004)
		ERK, Src, EGF	(Wayne,2007)
		Smad4	(Wang,2011)
	p450arom		(Kwintkiewicz,2010),(Escamilla-
	P		Hernandez,2008)(Park,2005)(Tetsuka,2010)
		Ras, ERK	(Wayne,2007)
		PKA	(Wayne,2007)(Silva,2006)(Saxena,2004)
		CTNNB1	(Fan,2010)
		Smad3	(Gong,2009)
		Smad4	(Wang,2011)
	StAR		(Tosca,2005)(Yu,2005)(Escamilla-
			Hernandez,2008)
		Ras,ERK,EGFR,PI3K	(Wayne,2007)
Régulation de l'hypophyse	α-Inhibin		(Escamilla-Hernandez,2008)
<i>і пурорпуѕе</i>		PKA	(Saxena,2004)(Park,2005)
		HIF-1α	(Alam,2004)
		CREB	(Pei,1990)

	Inhibin-βA		(Escamilla-Hernandez,2008)
Facteurs de	c-JUN		(Yang,2008)
transcription	TAF4B		(Ongeri,2007)
	SF-1		(Park,2005)
Cycle cellulaire	Cend2		(Kayampilly,2009)
		Smad3	(Gong,2009)
		Smad4	(Wang,2011)
		mTOR	(Kayampilly,2007)
	Cdk2	Smad4	(Wang,2011)
	Cdk4	Smad4	(Wang,2011)
Récepteurs	FSHR	Smad3	(Gong,2009)
		Smad4	(Wang,2011)
		CTNNB1	(Fan,2010)
	LHR		(Escamilla-Hernandez,2008)
		PKA	(Saxena,2004)
		HIF-1α	(Alam,2004)
Hormones et Facteurs	АМН	PKA	(Taieb,2011)
racieurs	IGF-1		(Kwintkiewicz,2010)
		PKA	(Yu,2003)
	IGF-2	PKA	(Yu,2003)
	VEGF		(Alam,2009)
		HIF-1α	(Alam,2004)
	AREG		(Escamilla-Hernandez,2008)(Freimann,2004)
		EGFR, ERK, PI3K	(Wayne,2007)
	EREG		(Freimann, 2004) (Park, 2005)
Régulation de facteurs	IGFBP-3		(Rappaport,1995),(Meachem,2005),(Smith,1990),(Abel,2009)
Voie anti-	GCLC	PKA	(Hoang,2009)
oxydante	GCLM	PKA	(Hoang,2009)
Autres	Carbonic anhydrase-2		(Escamilla-Hernandez,2008)

	(Escamilla-Hernandez,2008)
	(Escamilla-Hernandez,2008)
	(Escamilla-Hernandez,2008)
cAMP	(Kazutaka,2011)
	(Levallet,2007)
	(Levallet,2008)
	(Lekha-Dahia,2005)
	(Lekha-Dahia,2005)
cAMP	(Hall,1990)
NFκB	(Wang,2002)
	cAMP

Tableau 8 : Transcription inhibée par la FSH

Impliqué dans	Protéine	Intermédiaire	Références
Stéroïdogenèse	p450arom	PKC, ERK	(Silva,2006)
Transcription	FOXO1		(Fan,2010)
Recepteurs	FSHR	cAMP	(Kazutaka,2011)
Cycle cellulaire	p27 kip		(Kayampilly,2009)
Apoptose	Par-4		(Hernandez Gonzalez,2007)
Hormones et Facteurs	IL-6		(Lekha-Dahia,2005)
Autres	Caveolin-1		(Escamilla-Hernandez,2008)

Tableau 9 : Traduction induite par la FSH

Impliquée dans	Protéine	Intermédiaire	Références
Stéroïdogenèse	3β-HSD		(Tosca,2005)
		PKA	(Escamilla-Hernandez,2008)
	p450arom	PKA	(Escamilla-Hernandez,2008)
	p450scc		(Tosca,2005)
		PKA	(Escamilla-Hernandez,2008)
	StAR		(Cottom,2003)(Tosca,2005)(Yu,2005)
Régulation de l'hypophyse	α-Inhibin		(Escamilla-Hernandez,2008)
Prolifération	Raf-A		(Gonzales-Robayna,2000)

	Raf1		(Gonzales-Robayna,2000)
Facteurs de transcription	β-catenin		(Fan,2010)
	HIF-1α		(Alam,2004)
	LRH-1	p38MAPK	(Yu,2005)
	SF-1		(Park,2005)
	c-Fos		(Musnier,2012)
Récepteurs	CD44	Hyaluronan Synthase	(Tunjung,2009)
	LHR	PKA	(Escamilla-Hernandez,2008)
Hormones et facteurs	VEGF		(Musnier,2012)
Cycle cellulaire	PCNA		(Tosca,2005)(Park,2005)
	PP2A catalytic subunit		(Levallet,2008)
Autres	AKAP80		(Carr,1993)
		ERK	(Cottom,2003)
	ΑΜΡΚα1		(Tosca,2005)
	MAP2D	PI3K,mTOR	(Alam,2004)(Salvador,2004)
	Rap1		(Gonzales-Robayna,2000)
	RIIβ	PI3K,mTOR	(Alam,2004)(Carr,1993)
	SCF		(Kazutaka,2011)
	PDE4D1		(Levallet,2008)
	ABP	cAMP	(Hall,1990)

Bibliographie additionnelle:

- Abel, M. H., et al. (2009), 'Effects of FSH on testicular mRNA transcript levels in the hypogonadal mouse', *Journal of molecular endocrinology*, 42 (4), 291-303.
- Alam, H., et al. (2004), 'Follicle-stimulating hormone activation of hypoxia-inducible factor-1 by the phosphatidylinositol 3-kinase/AKT/Ras homolog enriched in brain (Rheb)/mammalian target of rapamycin (mTOR) pathway is necessary for induction of select protein markers of follicular differentiation', *The Journal of biological chemistry*, 279 (19), 19431-40.
- Alam, H., et al. (2009), 'Role of the phosphatidylinositol-3-kinase and extracellular regulated kinase pathways in the induction of hypoxia-inducible factor (HIF)-1 activity and the HIF-1 target vascular endothelial growth factor in ovarian granulosa cells in response to follicle-stimulating hormone', *Endocrinology*, 150 (2), 915-28.
- Carr, D. W., et al. (1993), 'Follicle-stimulating hormone regulation of A-kinase anchoring proteins in granulosa cells', *The Journal of biological chemistry*, 268 (28), 20729-32.

- Cottom, J., et al. (2003), 'Follicle-stimulating hormone activates extracellular signal-regulated kinase but not extracellular signal-regulated kinase kinase through a 100-kDa phosphotyrosine phosphatase', *The Journal of biological chemistry*, 278 (9), 7167-79.
- Dahia, C. L. and Rao, A. J. (2006), 'Regulation of FSH receptor, PKIbeta, IL-6 and calcium mobilization: Possible mediators of differential action of FSH', *Molecular and cellular endocrinology*, 247 (1-2), 73-81.
- Doyle, L. K., Walker, C. A., and Donadeu, F. X. (2010), 'VEGF modulates the effects of gonadotropins in granulosa cells', *Domestic animal endocrinology*, 38 (3), 127-37.
- Escamilla-Hernandez, R., et al. (2008), 'Constitutively active protein kinase A qualitatively mimics the effects of follicle-stimulating hormone on granulosa cell differentiation', *Molecular endocrinology*, 22 (8), 1842-52.
- Fan, H. Y., et al. (2010), 'Beta-catenin (CTNNB1) promotes preovulatory follicular development but represses LH-mediated ovulation and luteinization', *Molecular endocrinology*, 24 (8), 1529-42.
- Freimann, S., et al. (2004), 'EGF-like factor epiregulin and amphiregulin expression is regulated by gonadotropins/cAMP in human ovarian follicular cells', *Biochemical and biophysical research communications*, 324 (2), 829-34.
- Gong, X. and McGee, E. A. (2009), 'Smad3 is required for normal follicular follicle-stimulating hormone responsiveness in the mouse', *Biology of reproduction*, 81 (4), 730-8.
- Gonzalez, I. H., et al. (2007), 'Regulation of the expression of prostate apoptosis response protein 4 (Par-4) in rat granulosa cells', *Apoptosis : an international journal on programmed cell death*, 12 (4), 769-79.
- Gonzalez-Robayna, I. J., et al. (2000), 'Follicle-Stimulating hormone (FSH) stimulates phosphorylation and activation of protein kinase B (PKB/Akt) and serum and glucocorticoid-Induced kinase (Sgk): evidence for A kinase-independent signaling by FSH in granulosa cells', *Molecular endocrinology*, 14 (8), 1283-300.
- Hoang, Y. D., Nakamura, B. N., and Luderer, U. (2009), 'Follicle-stimulating hormone and estradiol interact to stimulate glutathione synthesis in rat ovarian follicles and granulosa cells', *Biology of reproduction*, 81 (4), 636-46.
- Kayampilly, P. P. and Menon, K. M. (2007), 'Follicle-stimulating hormone increases tuberin phosphorylation and mammalian target of rapamycin signaling through an extracellular signal-regulated kinase-dependent pathway in rat granulosa cells', *Endocrinology*, 148 (8), 3950-7.
- --- (2009), 'Follicle-stimulating hormone inhibits adenosine 5'-monophosphate-activated protein kinase activation and promotes cell proliferation of primary granulosa cells in culture through an Akt-dependent pathway', *Endocrinology*, 150 (2), 929-35.
- Kazutaka, S., et al. (2011), 'Regulation of Sertoli cell activin A and inhibin B by tumour necrosis factor alpha and interleukin 1alpha: interaction with follicle-stimulating hormone/adenosine 3',5'-cyclic phosphate signalling', *Molecular and cellular endocrinology*, 335 (2), 195-203.
- Kwintkiewicz, J., et al. (2010), 'Peroxisome proliferator-activated receptor-gamma mediates bisphenol A inhibition of FSH-stimulated IGF-1, aromatase, and estradiol in human granulosa cells', *Environmental health perspectives*, 118 (3), 400-6.
- Levallet, G., Levallet, J., and Bonnamy, P. J. (2007), 'Alterations in proteoglycan synthesis selectively impair FSH-induced particulate cAMP-phosphodiesterase 4 (PDE4) activation in immature rat Sertoli cells', *Biochimica et biophysica acta*, 1770 (4), 638-48.

- --- (2008), 'FSH-induced phosphoprotein phosphatase 2A-mediated deactivation of particulate phosphodiesterase-4 activities is abolished after alteration in proteoglycan synthesis in immature rat Sertoli cells', *The Journal of endocrinology*, 197 (1), 45-54.
- Meachem, S. J., et al. (2005), 'Developmentally distinct in vivo effects of FSH on proliferation and apoptosis during testis maturation', *The Journal of endocrinology*, 186 (3), 429-46.
- Musnier, A., et al. (2012), 'mRNA-selective translation induced by FSH in primary Sertoli cells', *Molecular endocrinology*, 26 (4), 669-80.
- Ongeri, E. M., Verderame, M. F., and Hammond, J. M. (2007), 'The TATA binding protein associated factor 4b (TAF4b) mediates FSH stimulation of the IGFBP-3 promoter in cultured porcine ovarian granulosa cells', *Molecular and cellular endocrinology*, 278 (1-2), 29-35.
- Park, Y., et al. (2005), 'Induction of cyclin D2 in rat granulosa cells requires FSH-dependent relief from FOXO1 repression coupled with positive signals from Smad', *The Journal of biological chemistry*, 280 (10), 9135-48.
- Rappaport, M. S. and Smith, E. P. (1995), 'Insulin-like growth factor (IGF) binding protein 3 in the rat testis: follicle-stimulating hormone dependence of mRNA expression and inhibition of IGF-I action on cultured Sertoli cells', *Biology of reproduction*, 52 (2), 419-25.
- Salvador, L. M., et al. (2004), 'Neuronal microtubule-associated protein 2D is a dual a-kinase anchoring protein expressed in rat ovarian granulosa cells', *The Journal of biological chemistry*, 279 (26), 27621-32.
- Saxena, D., et al. (2004), 'Liver receptor homolog-1 stimulates the progesterone biosynthetic pathway during follicle-stimulating hormone-induced granulosa cell differentiation', *Endocrinology*, 145 (8), 3821-9.
- Silva, J. M., et al. (2006), 'Control of oestradiol secretion and of cytochrome P450 aromatase messenger ribonucleic acid accumulation by FSH involves different intracellular pathways in oestrogenic bovine granulosa cells in vitro', *Reproduction*, 132 (6), 909-17.
- Smith, E. P., Dickson, B. A., and Chernausek, S. D. (1990), 'Insulin-like growth factor binding protein-3 secretion from cultured rat sertoli cells: dual regulation by follicle stimulating hormone and insulin-like growth factor-I', *Endocrinology*, 127 (6), 2744-51.
- Taieb, J., et al. (2011), 'FSH and its second messenger cAMP stimulate the transcription of human anti-Mullerian hormone in cultured granulosa cells', *Molecular endocrinology*, 25 (4), 645-55.
- Tetsuka, M., et al. (2010), 'Gene expression of 11beta-HSD and glucocorticoid receptor in the bovine (Bos taurus) follicle during follicular maturation and atresia: the role of follicular stimulating hormone', *The Journal of reproduction and development*, 56 (6), 616-22.
- Tosca, L., et al. (2005), 'Adenosine 5'-monophosphate-activated protein kinase regulates progesterone secretion in rat granulosa cells', *Endocrinology*, 146 (10), 4500-13.
- Tunjung, W. A., et al. (2009), 'Effect of hyaluronan to inhibit caspase activation in porcine granulosa cells', *Biochemical and biophysical research communications*, 382 (1), 160-4.
- Wang, W., et al. (2011), 'Interference RNA-based silencing of endogenous SMAD4 in porcine granulosa cells resulted in decreased FSH-mediated granulosa cells proliferation and steroidogenesis', *Reproduction*, 141 (5), 643-51.
- Wang, Y., Chan, S., and Tsang, B. K. (2002), 'Involvement of inhibitory nuclear factor-kappaB (NFkappaB)-independent NFkappaB activation in the gonadotropic regulation

- of X-linked inhibitor of apoptosis expression during ovarian follicular development in vitro', *Endocrinology*, 143 (7), 2732-40.
- Wayne, C. M., et al. (2007), 'Follicle-stimulating hormone induces multiple signaling cascades: evidence that activation of Rous sarcoma oncogene, RAS, and the epidermal growth factor receptor are critical for granulosa cell differentiation', *Molecular endocrinology*, 21 (8), 1940-57.
- Yang, W., et al. (2008), 'Mitogen-activated protein kinase regulates FSH-induced expression of tissue-type plasminogen activator through an activator protein 1 response element', *Endocrine*, 34 (1-3), 101-7.
- Yu, F. Q., et al. (2005), 'Activation of the p38 MAPK pathway by follicle-stimulating hormone regulates steroidogenesis in granulosa cells differentially', *The Journal of endocrinology*, 186 (1), 85-96.
- Yu, Y., et al. (2003), 'The effect of follicle-stimulating hormone on follicular development, granulosa cell apoptosis and steroidogenesis and its mediation by insulin-like growth factor-I in the goat ovary', *Theriogenology*, 60 (9), 1691-704.

Annexe 6: Modules

Cette annexe est dédiée aux résultats obtenus grâce à l'inférence automatique mais qui ne sont pas présentés dans la partie Résultats. Cette annexe permet de visualiser la totalité des réseaux induits par la FSH et l'EGF que nous avons obtenus automatiquement.

Dans les deux modèles inférés, les conclusions obtenues grâce au moteur d'inférence peuvent parfois être redondantes. Par exemple, dans le réseau FSH, les conclusions suivantes sont obtenues :

- PHOSPHORYLATE(FSH,βArrestin1,pβArrestin1,unknown,confirmed,decrease)
- PHOSPHORYLATE (FSH,\beta\Arrestin1,\p\beta\Arrestin1,indirect,confirmed,decrease)
- MODIFY(FSH,\beta\text{Arrestin1},\text{p\beta\text{Arrestin1},unknown,hypothesis,decrease)

Lorsque ce cas de figure est rencontré, la relation la plus précise est choisie pour être représentée dans le réseau. Dans l'exemple précédent, la conclusion représentée dans le réseau est PHOSPHORYLATE (FSH, βArrestin1, pβArrestin1, indirect, confirmed, decrease).

De plus, dans la partie « Matériel et Méthodes », nous avons évoqué la réversibilité des relations de modification. Comme nous l'avons précisé, seules les activations sont représentées.

Légende des figures présentées dans cette preuve de concept :

Relations confirmées:

- Si la relation est directe, elle apparaît en trait plein noir,
- Si la relation est indirecte, elle apparaît en pointillés noirs,
- Si la relation a une distance inconnue, elle apparaît en pointillés verts,
- Si la relation est admise, elle apparaît en parme.

Relations hypothétiques :

- Si la relation est directe, elle apparaît en trait continu bleu,
- Si la relation est indirecte, elle apparaît en pointillés bleus,
- Si la relation a une distance inconnue, elle apparaît en pointillés rouges.

A- Réseau FSH

Le réseau FSH obtenu contient un grand nombre de relations dont le détail est présenté dans le Tableau 10. Parmi ces relations, 191 ne sont pas représentées graphiquement car elles

présentent un effet nul (137) ou présentent en argument un agoniste (54) dont la présence n'est pas « naturelle » dans le réseau.

Tableau 10 : Nombre de relations obtenues dans le réseau FSH

Type de relation	Nombre
Complexation	16
Dissociation	2
Echange	2
Effet sur la traduction	38
Effet sur la transcription	60
Modification	183
Phosphorylation	271
Réaction enzymatique	8

1- Voie Gs

1.1- Voie Gs et production de l'AMPc

Ce module ne présente aucun conflit entre les relations déduites.

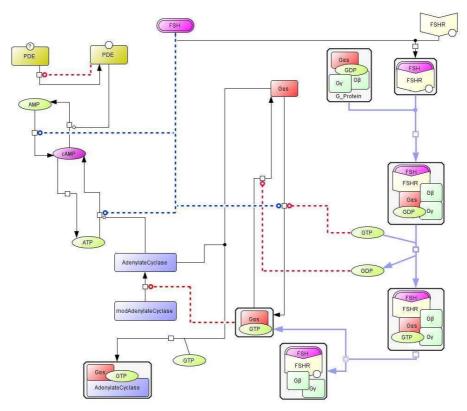


Figure 89 : Activation du FSHR telle qu'inférée par notre méthode, recrutement de la protéine Gs et production de l'AMPc

1.2- Actions de l'AMPc

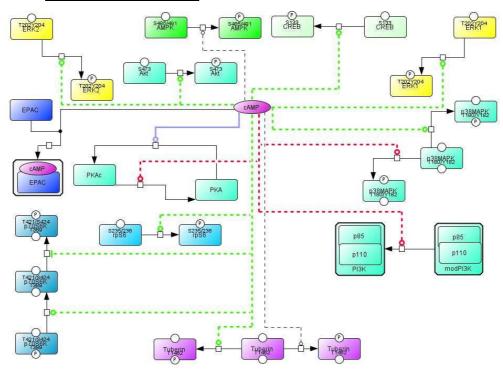


Figure 90: Actions de l'AMPc dans la signalisation induite par la FSH

Il existe des faits déduits qui peuvent sembler contradictoires : l'AMPc induit directement l'activation de la PKA en libérant la PKAc (*status=confirmed*), or l'un des faits déduits (*status=hypothesis*) est que l'AMPc inhibe cette réaction.

1.3- Actions d'EPAC | Rap1 |

Figure 91: Actions d'EPAC dans la signalisation induite par la FSH

Il existe des faits déduits qui peuvent sembler contradictoires : le complexe EPAC:AMPc induit l'activation de la PKA en libérant la PKAc avec une distance inconnue (status=hypothesis), or l'un des faits déduits (status=hypothesis) est que le complexe EPAC:AMPc inhibe cette réaction.

1.4- Actions impliquant la PKA

Il existe des faits déduits qui peuvent sembler contradictoires : la FSH induit l'activation de la PKA en libérant la PKAc avec une distance inconnue (*status=hypothesis*), or l'un des faits déduits (*status=hypothesis*) est que la FSH inhibe cette réaction.

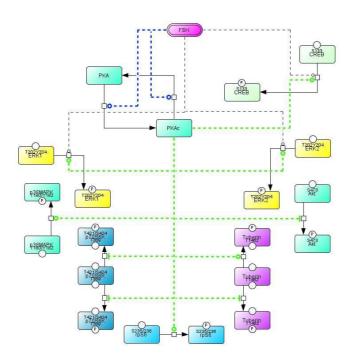


Figure 92: Actions de la PKA dans la signalisation induite par la FSH

1.5- Actions de la p38 MAP Kinase

Il existe des faits déduits qui peuvent sembler contradictoires : la FSH induit indirectement la phosphorylation de la p38MAPK (*status=confirmed*), or l'un des faits déduits (*status=hypothesis*) est que la FSH inhibe cette réaction.

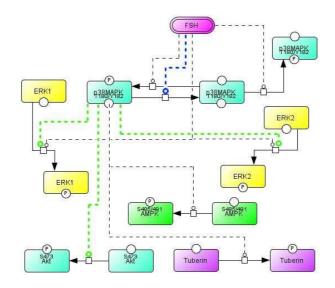


Figure 93 : Actions de la p38 MAP Kinase dans la signalisation induite par la FSH

2- Voie Gi

Ce module ne présente aucun conflit entre les relations déduites.

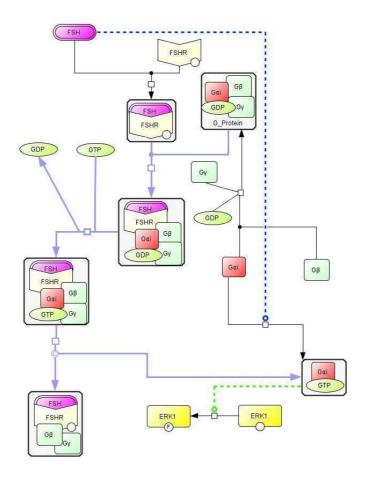


Figure 94 : Activation du FSHR, recrutement de la protéine Gi

3- Voie β -arrestines

Ce module ne présente aucun conflit entre les relations déduites.

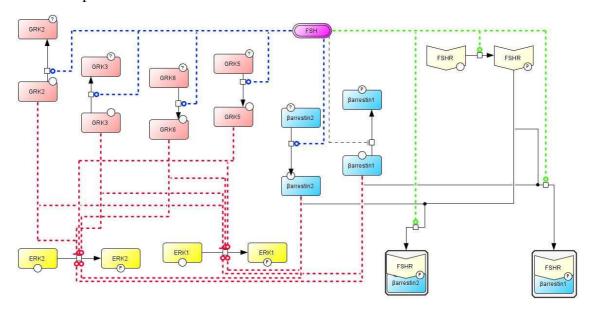


Figure 95 : Voie β -arrestines

3- Voie PI3K

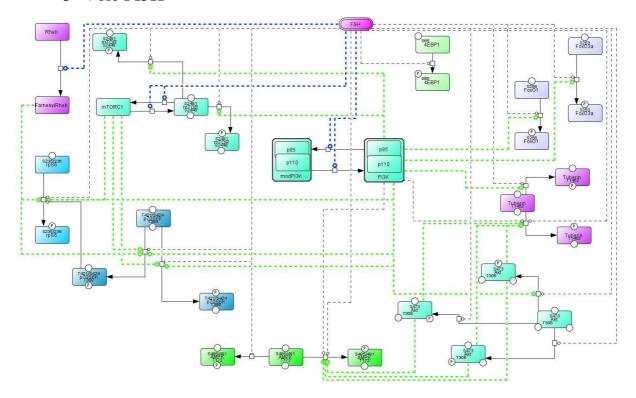


Figure 96 : Voie PI3K

Il existe des faits déduits qui peuvent sembler contradictoires : la FSH induit l'activation de mTOR (*status=hypothesis*), or l'un des faits déduits (*status=hypothesis*) est que la FSH inhibe cette réaction. Le même type de contradiction est observé dans le cas de l'activation de la PI3K.

4- Activation et actions de Src

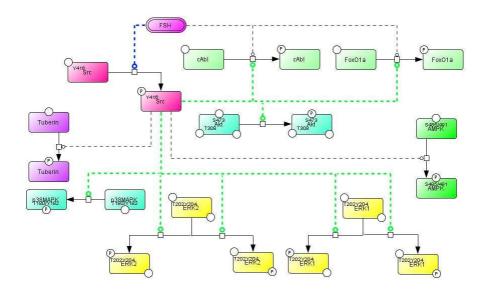


Figure 97 : Activation de Src par la FSH

Ce module ne contient aucun conflit.

5- Cascade ERK MAPK

Contradictions entre des hypothèses :

- La FSH induit la modification de Ras en Ras:GTP mais induit également la modification inverse,
- La FSH induit la phosphorylation de MEK1 mais également la déphosphorylation.

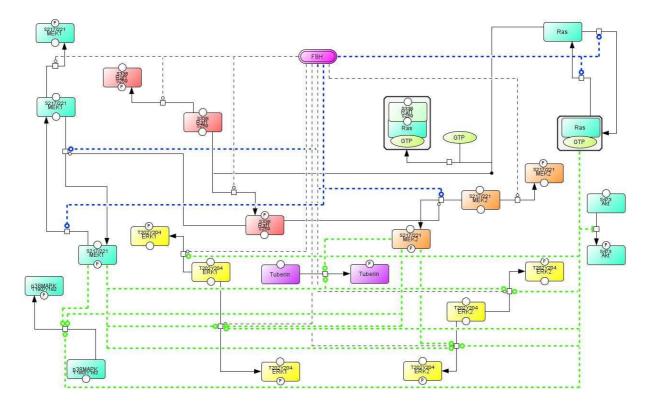


Figure 98 : Cascade ERK MAPK

6- Synthèse des hormones stéroïdes

Il n'existe pas de conflit dans ce module.

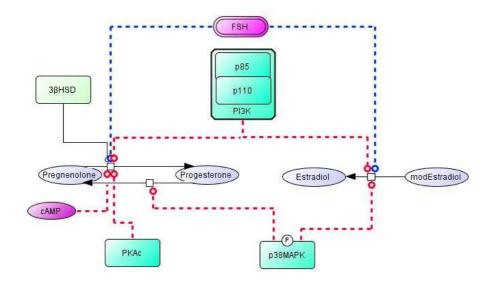


Figure 99 : Régulation de la synthèse des hormones stéroïdiennes par la FSH

7- Molécules impliquées dans le cycle cellulaire

Il n'y a pas de contradictions entre les relations.

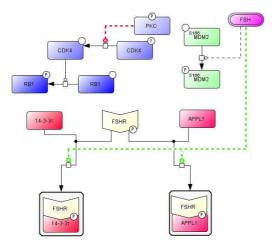


Figure 100 : Relations impliquée dans la régulation du cycle cellulaire

8- Expression des protéines régulées par la signalisation FSH Ce module ne présente aucune contradiction.

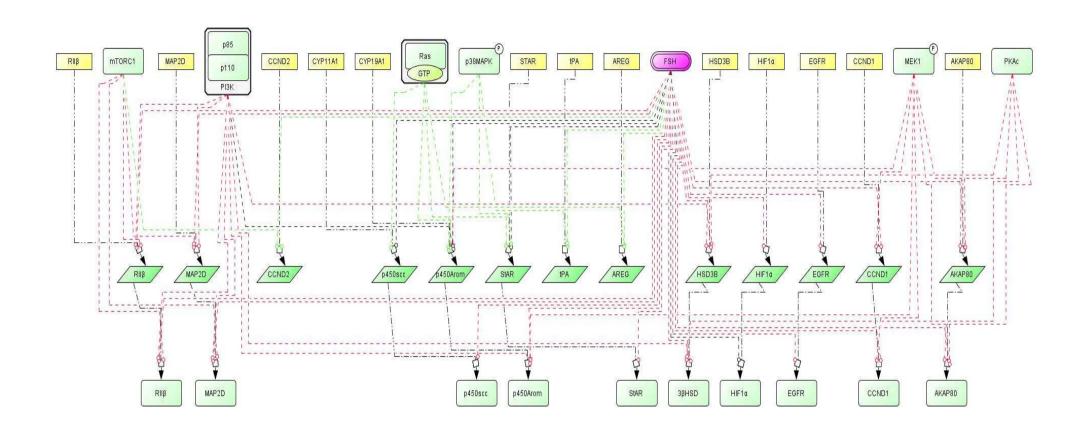


Figure 101 : Expression des protéines régulée par la signalisation induite par la FSH

B- Réseau EGF

Le réseau EGF obtenu contient un grand nombre de relations dont le détail est présenté dans le Tableau 11. Parmi ces relations, 153 ne sont pas représentées graphiquement car elles présentent un effet nul (113) ou présentent en argument un agoniste (40) dont la présence n'est pas « naturelle » dans le réseau.

Tableau 11 : Nombre de relations obtenues dans le réseau EGF

Type de relation	Nombre
Complexation	80
Dissociation	0
Echange	0
Effet sur la traduction	8
Effet sur la transcription	8
Modification	118
Phosphorylation	332
Réaction enzymatique	16

1- Les actions impliquant le récepteur EGFR

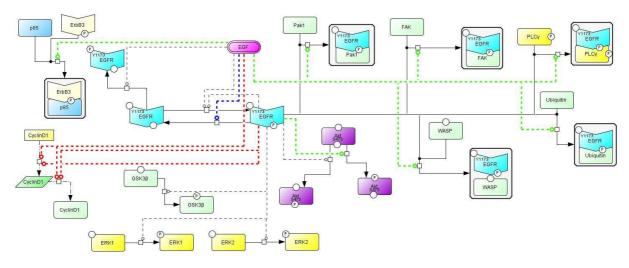


Figure 102 : Actions impliquant le récepteur EGFR dans le réseau EGF

Il existe des faits déduits qui peuvent sembler contradictoires : L'EGF induit la phosphorylation de son récepteur (*status=confirmed*), or l'un des faits déduits (*status=hypothesis*) est que l'EGF inhibe cette réaction.

2- Les actions impliquant les protéines Grb2 et SOS

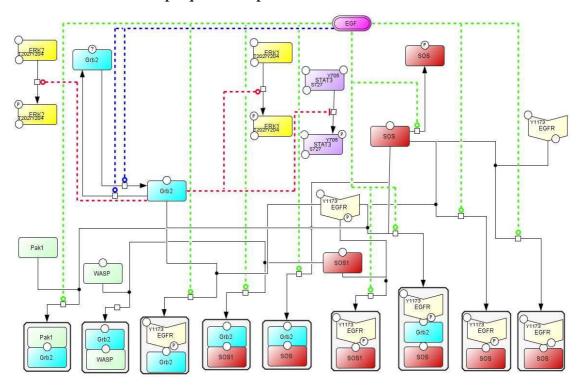


Figure 103 : Actions impliquant les protéines Grb2 et SOS dans le réseau EGF Il existe des faits déduits qui peuvent sembler contradictoires : L'EGF induit une modification de Grb2 (*status=hypothesis*), or l'un des faits déduits (*status=hypothesis*) est que l'EGF inhibe cette réaction.

3- Voie de signalisation PI3K

Il existe des faits déduits qui peuvent sembler contradictoires : L'EGF induit l'activation de la PI3K (*status=hypothesis*), or l'un des faits déduits (*status=hypothesis*) est que l'EGF inhibe cette réaction.

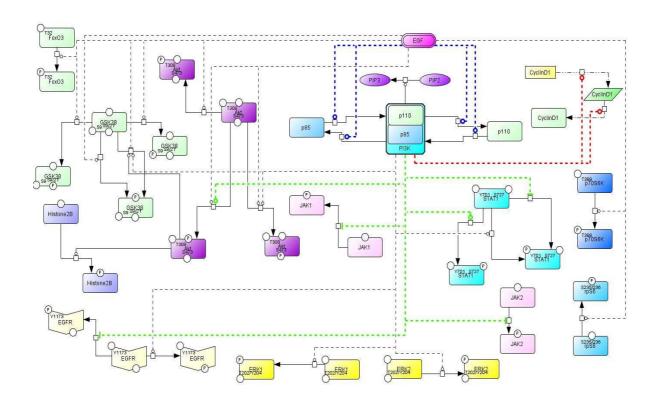


Figure 104 : Voie PI3K induite par l'EGF

4- Cascade ERK MAPK

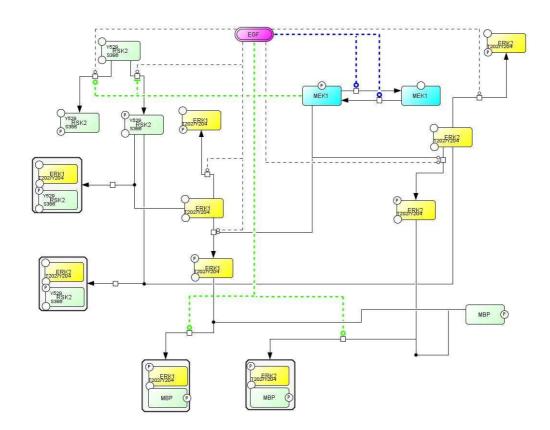


Figure 105 : Voie ERK MAPK induite par l'EGF (partie 1)

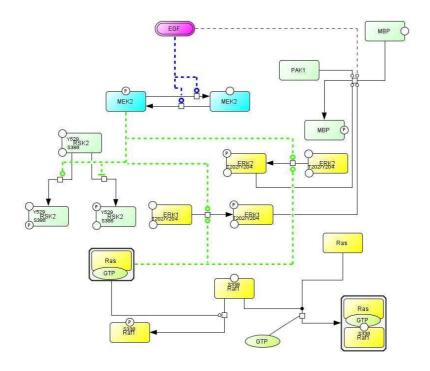


Figure 106 : Voie ERK MAPK induite par l'EGF (partie 2)

Il existe des faits déduits qui peuvent sembler contradictoires : L'EGF induit la phosphorylation des MEK 1 et 2 (*status=hypothesis*), or l'un des faits déduits (*status=hypothesis*) est que l'EGF inhibe ces réactions.

5- Les actions impliquant les protéines Gab Aucun conflit n'est détecté dans ce module.

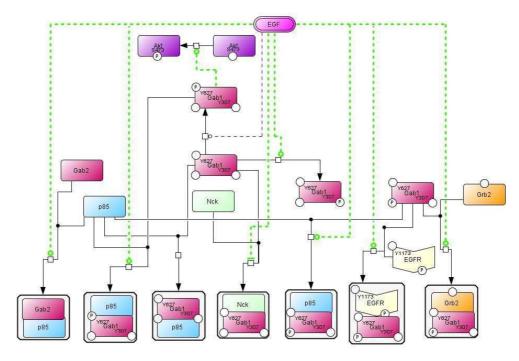


Figure 107 : Actions impliquant les protéines Gab dans le réseau EGF

6- Les actions impliquant les protéines Shc

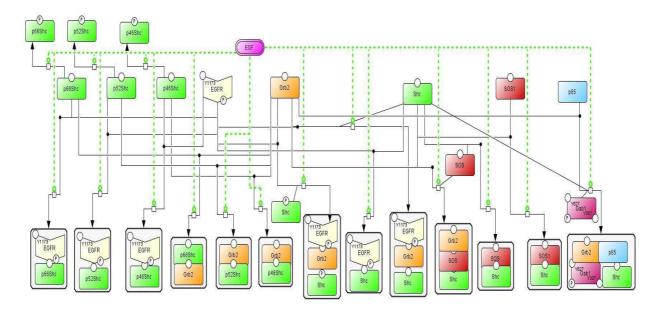


Figure 108: Actions impliquant les protéines Shc dans le réseau EGF

Aucun conflit n'est détecté dans ce module.

7- Les actions impliquant les sous-unités Gai

Il existe des faits déduits qui peuvent sembler contradictoires : L'EGF induit l'activation des Gai1 et Gai3 (*status=hypothesis*), or l'un des faits déduits (*status=hypothesis*) est que l'EGF inhibe ces réactions.

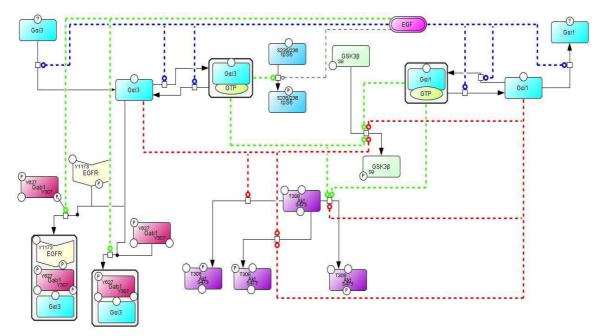


Figure 109 : Actions impliquant les sous-unités Gai dans le réseau EGF

8- Voie de signalisation JAK/STAT

Il existe des faits déduits qui peuvent sembler contradictoires : L'EGF induit la phosphorylation de STAT1 sur la position Y701 (*status=confirmed*), or l'un des faits déduits (*status=hypothesis*) est que l'EGF inhibe cette réaction.

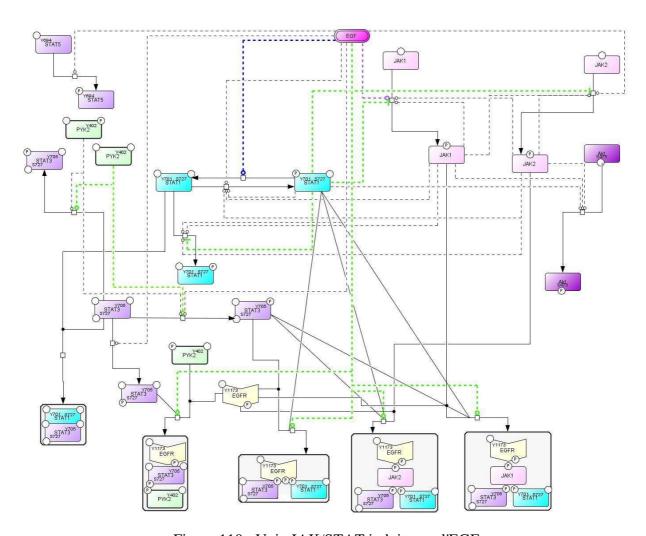


Figure 110 : Voie JAK/STAT induite par l'EGF

9- Les actions impliquant la protéine Src

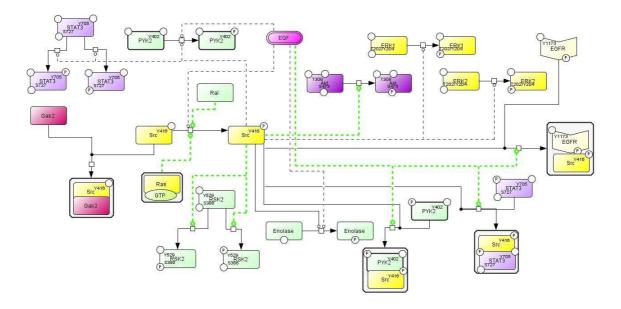


Figure 111 : Actions impliquant la protéine Src dans le réseau EGF

Aucun conflit n'est détecté dans ce module.

10- Les actions impliquant les protéines SHP

Aucun conflit n'est détecté dans ce module.

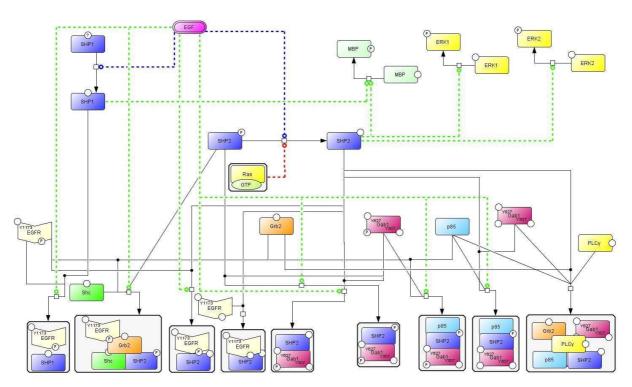


Figure 112 : Actions impliquant les protéines SHP dans le réseau EGF

11- Molécules impliquées dans la signalisation au site d'adhésion focale

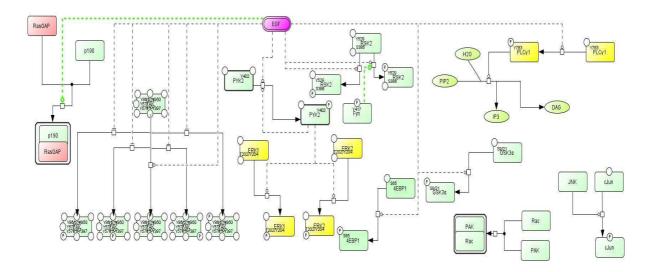


Figure 113 : Signalisation au site d'adhésion focale

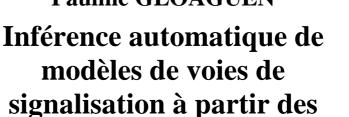
La protéine présentant un grand nombre de sites de phosphorylation est FAK.

Aucun conflit n'est détecté dans ce module.



Pauline GLOAGUEN

données expérimentales





Résumé

Les réseaux biologiques, notamment les réseaux de signalisation déclenchés par les hormones, sont extrêmement complexes. Les méthodes expérimentales à haut débit permettent d'aborder cette complexité, mais la prise en compte de l'ensemble des données générées requiert la mise au point de méthodes automatiques pour la construction des réseaux. Nous avons développé une nouvelle méthode d'inférence reposant sur la formalisation, sous forme de règles logiques, du raisonnement de l'expert sur les données expérimentales. Cela nécessite la constitution d'une base de connaissances, ensuite exploitée par un moteur d'inférence afin de déduire les conclusions permettant de construire les réseaux. Notre méthode a été élaborée grâce au réseau de signalisation induit par l'hormone folliculo-stimulante dont le récepteur fait partie de la grande famille des récepteurs couplés aux protéines G. Ce réseau a également été construit manuellement pour évaluer notre méthode. Un contrôle a ensuite été réalisé, en utilisant les mêmes règles, sur le réseau induit par le facteur de croissance épidermique, se liant à un récepteur tyrosine kinase, de façon à montrer la généricité de notre méthode.

Mots-clés: Inférence automatique, Réseaux de signalisation, Règles logiques, Logique du premier ordre

Abstract

Biological networks, including signalling networks induced by hormones, are very complex. High-throughput experimental methods permit to approach this complexity, but to be able to use all generated data, it is necessary to create automated inference methods to build networks. We have developped a new inference method based on the formalization of the expert's reasoning on experimental data. This reasoning is converted into logical rules. This work requires the creation of a knowledge base which is used by an inference engine to deduce conclusions to build networks. Our method has been elaborated by the construction of the signalling network induced by the follicle stimulating hormone whose receptor belongs to the G protein-coupled receptors family. This network has also been built manually to assess our method. Then, a test has been conducted, using same set of rules, on the network induced by the epidermal growth factor, which binds to a tyrosine kinase receptor, to demonstrate the genericity of our method.

Keywords: Automated inference, Signalling networks, Logical rules, First order logic