

UNIVERSITÉ DE TOURS

ÉCOLE DOCTORALE « Santé, Science Biologiques et Chimie du Vivant »

U1253 iBrain

EA6295 « Nano Médicaments et Nano Sondes »

THÈSE présentée par **Clément BRUNO**

soutenue le: **14 Septembre 2021**

Pour obtenir le grade de : **Docteur de l'université de Tours**

Discipline/ Spécialité : **Science de la vie et de la santé**

DEVELOPPEMENT ET MISE AU POINT DE NOUVEAUX OUTILS ANALYTIQUES DANS UNE APPROCHE METABOLOMIQUE

THÈSE dirigée par :

**Mme BLASCO Hélène
M. BONNIER Franck**

PU-PH, université de Tours
MCU-HDR, université de Tours

PRESIDENT DU JURY :

M. TFAYLI Ali

PU, Université Paris Saclay

RAPPORTEURS :

**M. SOCKALINGUM Ganesh
M. SIRE Olivier**

PU, Université de Reims Champagne-Ardenne
PU, Université de Bretagne-Sud

JURY :

**Mme BLASCO Hélène
M. BONNIER Franck
M. SOCKALINGUM Ganesh
M. SIRE Olivier
M. REYNIER Pascal
M. ANDRES Christian
M. TFAYLI Ali**

PU-PH, université de Tours
MCU-HDR, université de Tours
PU, Université de Reims Champagne-Ardenne
PU, Université de Bretagne-Sud
PU-PH, Université d'Angers
PU-PH, Université de Tours
PU, Université Paris Saclay

« Non omnibus aegris eadem auxilia conveniunt »

(On ne soigne pas tous les malades avec les mêmes remèdes)

Celse, I^{er} siècle

REMERCIEMENTS

Tout d'abord, je tiens à remercier mes directeurs de thèse, le Pr Hélène Blasco et le Dr Franck Bonnier, pour leur implication dans ce projet, leur soutien sans faille, leur accompagnement au quotidien ainsi que leurs conseils avisés tout au long de ce travail de recherche. Ce travail n'aurait pu être mené à bien sans eux.

Un grand merci au Pr Sire et au Pr Sockalingum d'avoir accepté d'être rapporteurs de ce travail ainsi qu'aux autres membres du jury pour juger ce travail. Un petit mot particulier à l'intention de Pascal Reynier avec qui j'ai éprouvé beaucoup de plaisir à travailler et pour toutes nos discussions très intéressantes.

Je tiens également à remercier le Pr. Igor Chourpa, directeur de l'équipe EA6295 NMNS de l'université de Tours, ainsi que le Pr Catherine Belzung, directrice de l'Unité Inserm 1253 Imagerie et Cerveau, pour m'avoir accueilli au sein de leurs unités de recherche.

Je souhaite également remercier l'ensemble des biologistes des laboratoires de Pharmacologie-Toxicologie du Pr Paintaud et de Biochimie et Biologie Moléculaire du Pr Andres pour l'ensemble des connaissances et compétences qui m'ont été transmises et que j'ai pu mettre en application dans ce travail.

Je remercie bien évidemment tous ceux qui m'ont apporté la bonne humeur au quotidien, qui m'ont permis d'évacuer les moments de doutes et de raz-le-bol (je pense en particulier à toi Céline et nos fameux groupes de paroles).

Je souhaite adresser mes derniers remerciements aux trois personnes qui font de ma vie ce qu'elle est aujourd'hui. Ceux qui me font me rappeler au quotidien qu'il y a toujours des bonnes choses, même dans les moments difficiles. Merci mon cœur pour ton soutien au quotidien, tu as toujours été attentive pour me ménager durant ces années compliquées. Merci également à Théo et Oscar qui m'ont toujours apporté leur amour et qui en un clignement d'œil ou un sourire sont capable d'embellir ma vie au quotidien et de me donner la force de continuer.

LISTE DES FIGURES

Figure 1 : Système nerveux central et Sclérose Latérale Amyotrophique

Figure 2 : Chronologie de la génétique dans la SLA. Le pourcentage représente la proportion des cas de SLA expliqués par la génétique

Figure 3 : Schéma représentant l'approche thérapeutique « classique » en A et l'approche de la médecine de précision en B

Figure 4 : Nombre de publications annuelles traitant de génomique, transcriptomique, protéomique et métabolomique ces 20 dernières années

Figure 5 : Cascade « omique » (reproduit d'après Dettmer et al, 2007) : de ce qu'il peut se passer jusqu'à ce qu'il se passe et s'est passé

Figure 6 : Schéma représentant la polarisation des spins et leur relaxation permettant l'analyse en RMN

Figure 7 : Schéma des paramètres d'identification en RMN

Figure 8 : Représentation des multiplicités de signaux visibles en RMN 1H

Figure 9 : Schéma d'un sélecteur de masse de type quadripôle

Figure 10 : Schéma des différents phénomènes physiques impactant le trajet lumineux

Figure 11 : Modes de vibrations des liaisons covalentes

Figure 12 : Diagramme de Jablonski représentant les phénomènes d'absorption infrarouge, de diffusion Raman Stokes, Raman anti-Stokes et Rayleigh

Figure 13 : Spectre d'absorption ATR-IR obtenu à partir d'un plasma humain

Figure 14 : Spectre d'absorption Raman obtenu à partir d'un plasma humain

Figure 15 : Phénomène d'absorption et mise en évidence par le calcul de la transmittance

Figure 16 : Schéma représentant les différents modes d'analyse en FTIR

Figure 17 : Schéma représentant les phénomènes de diffusion inélastique Stokes, inélastique anti-stokes et de diffusion élastique Rayleigh

Figure 18 : Représentation schématique du trajet optique dans un spectromètre Raman confocal

Figure 19 : Impact de la séparation chromatographique et de l'ionisation sur le nombre de pics d'intérêts obtenu à partir d'un même échantillon analysé selon 4 modes différents (C18(+), C18(-), HILIC(+)) and HILIC(-))

Figure 20 : Évaluation de la linéarité du signal du plasma humain sous forme liquide en fonction des volumes d'échantillon déposés

Figure 21 : Évaluation de la linéarité du signal du plasma humain sous forme liquide en fonction de la concentration en protéines.

Figure 22 : Spectres ATR-IR obtenu à partir de solutions à 1g/L d'albumine, d'immunoglobulines, de glucose et d'urée

Figure 23 : Spectres ATR-IR obtenu à partir d'un plasma patient analysé sous forme liquide et après séchage

Figure 24 : Exemple de corrélations obtenues entre les concentrations de référence de différents paramètres étudiés

Figure 25 : Régression PLSR et comparaison des coefficients de régression et des spectres de référence des protéines totales et du glucose après analyse sous forme liquide

Figure 26 : Régression PLSR et comparaison des coefficients de régression et des spectres de référence des protéines totales et du glucose après analyse sous forme séchée

Figure 27: Impact du mouillage de la zone de mesure sur la forme de la goutte et sur la surface de dépôt

Figure 28 : Schéma représentant les différents systèmes d'ultrafiltration

Figure 29: Histogramme représentant les pourcentages de recouvrement des composants de la fraction protéique pour chaque protocole d'élimination des protéines

Figure 30: Histogramme représentant les pourcentages de recouvrement des composants de la fraction de bas poids moléculaire pour chaque protocole d'élimination des protéines

Figure 31 : Spectres d'absorbance mesurés par ATR-IR d'un sérum, d'une solution d'albumine et d'un filtrat préparé à l'aide du filtre Centrisart®

Figure 32 : Analyse multivariée par composante principale (PCA) mettant en évidence une modification de la composition de l'échantillon avant ultrafiltration et après ultrafiltration

Figure 33 : Amélioration de la discrimination par ATR-IR après filtration

Figure 34 : Spectres de diffusion Raman des fractions sEV et IEV des échantillons patients SLA et témoins

LISTE DES TABLEAUX

Tableau I : Essais cliniques étudiant de nouvelles thérapies ciblées sur des anomalies identifiées par analyse métabolomique

Tableau II : Caractéristiques et spécificités des techniques d'analyses métabolomiques

Tableau III : Biomarqueurs identifiées dans la Sclérose Latérale Amyotrophique à partir de prélèvements sanguins ou de LCR

Tableau IV : Concentrations de références des échantillons patients

Tableau V : Attribution des signaux spectraux aux catégories moléculaires

Tableau VI : Erreurs moyennes des concentrations prédites par les modèles de régression partielle des moindres carrés en analyse liquide pour la fraction protéique

Tableau VII : Erreurs moyennes des concentrations prédites par les modèles de régression partielle des moindres carrés en analyse liquide la fraction de bas poids moléculaire

Tableau VIII : Erreurs moyennes des concentrations prédites par les modèles de régression partielle des moindres carrés en analyse sous forme séchée pour la fraction protéique

Tableau IX : Erreurs moyennes des concentrations prédites par les modèles de régression partielle des moindres carrés en analyse sous forme séchée pour la fraction de bas poids moléculaire

Tableau X : Caractéristiques des filtres utilisés pour l'ultrafiltration lors de l'étude

Tableau XI : Masse molaire et concentrations des biomarqueurs mesurées par méthodes de référence

Tableau XII : Biomarqueurs protéiques identifiées dans la Sclérose Latérale Amyotrophique à partir de prélèvements sanguins ou de LCR

SOMMAIRE

Contexte de la recherche.....	1
Introduction.....	1
I) La Sclérose Latérale Amyotrophique	1
I.1) Présentation et caractéristiques	1
I.2) Diverses étiologies	3
I.3) Une prise en charge complexe.....	5
I.4) La nécessité de biomarqueurs de diagnostic et pronostic.....	6
I.4.1) Intérêt des biomarqueurs.....	6
I.4.2) Vers la médecine personnalisée.....	8
I.4.2.a Description des sous-populations d'une pathologie hétérogène	9
I.4.2.b Suivi de l'évolution de la maladie	11
I.4.3) Modalités de recherche de biomarqueurs.....	11
I.4.3.a) Les approches « classiques »	11
I.4.3.b) Les approches « omiques »	12
I.5) Les méthodes conventionnelles	15
RATIONNEL ET OBJECTIFS.....	18
I) Applications des méthodes conventionnelles à la recherche de biomarqueurs dans la SLA.....	18
II) La spectroscopie vibrationnelle en biologie	20
III) OBJECTIFS.....	23
MATERIEL ET METHODES.....	25
I) Les méthodes conventionnelles.....	25
I.1) La RMN	25
I.2) La chromatographie	28
I.2.1) Détection.....	28
I.2.2) Séparation chromatographique.....	30

I.2.2.a) La phase stationnaire	31
I.2.2.b) La phase mobile	32
II) La spectroscopie vibrationnelle.....	33
II.1) Principes	33
II.2) Les modes de vibrations	35
II.3) Complémentarité de la spectroscopie IR et Raman	41
II.4) Spectroscopie infrarouge (IR)	42
II.4.a) Principe.....	42
II.4.b) Spectroscopie infrarouge et identification	43
II.4.c) Spectroscopie infrarouge et quantification	43
II.4.d) Modes de mesure	44
II.5) Spectroscopie Raman	46
II.5.a) Principe.....	46
II.5.b) Application de l'effet Raman	49
II.4.c) Spectromètre Raman	49
II.5.1) Source d'excitation	49
II.5.2) Détection	50
RESULTATS	51
Partie I : Evaluation de la combinaison des méthodes conventionnelles	51
I) Recherche de biomarqueurs musculaires dans la SLA.....	51
II) Intérêt de la combinaison de différentes méthodes analytiques lors d'une analyse métabolomique.....	53
III) Article 1 : Exploration du métabolome du muscle squelettique à travers la combinaison de 4 méthodes analytiques : meilleure couverture métabolique ou argument commercial ?	55
Partie II : Optimisation de l'analyse de plasmas humains par ATR-IR et évaluation de la capacité à quantifier spécifiquement un/des biomarqueur(s)	67

I) Utilisation de la spectroscopie infrarouge comme outil diagnostic: état de l'art	67
I.1) La complexité de l'analyse.....	67
I.2) La place de la spectroscopie infrarouge dans le diagnostic biologique ...	67
II) Article n°2 : Quantification des protéines dans le sérum humain par ATR-IR.	69
II.2) Etude préliminaire : protocoles analytiques en mode ATR	70
II.1.a) Influence du volume de dépôt sur le signal collecté	70
II.1.b) Relation concentration/signal en fonction de la concentration en protéines	71
II.2) Mesures des échantillons « patients »	73
II.2.a) Préparation des échantillons par la méthode des ajouts dosés	73
II.2.b) Obtention des valeurs de référence	74
II.2.c) Analyse par ATR-IR	74
III) Etude n°3 : Quantification simultanée des protéines et de petites molécules dans le plasma humain	87
IV) Etude n°4 : Impact de l'ultrafiltration centrifuge sur la composition qualitative et quantitative du plasma humain.....	112
V) Etude n°5 : Discrimination d'échantillons plasmatiques de patients SLA et de témoins.....	128
DISCUSSION GENERALE.....	134
La combinaison des méthodes conventionnelles permet une amélioration de la couverture métabolique.....	134
Quantification de biomarqueurs plasmatiques par spectroscopie vibrationnelle .	135
Analyse sous forme liquide et sous forme séchée	135
Utilisation de la spectroscopie vibrationnelle comme outil diagnostique en biologie.....	136
La spectroscopie vibrationnelle dans une approche métabolomique.....	137
AUTRES PROJETS	139

Optimisation de l'analyse de plasmas humains par Spectroscopie Raman et évaluation de la capacité à quantifier spécifiquement un/des biomarqueur(s)....	139
Etude n°1 : Analyse des fluides biologiques à l'aide de la spectroscopie vibrationnelle : une comparaison directe des spectroscopies de diffusion Raman et d'absorption infrarouge pour le cas du glucose dans le plasma	139
Etude n°2 : Analyse par spectroscopie Raman des fractions de bas poids et de haut poids moléculaires du plasma humain.....	156
CONCLUSION.....	181
REFERENCES	184

Contexte de la recherche

Introduction

I) La Sclérose Latérale Amyotrophique

I.1) Présentation et caractéristiques

La Sclérose Latérale Amyotrophique (ou SLA) est aussi connue sous le nom de maladie de Charcot (d'après le premier médecin à avoir décrit les symptômes en 1865). C'est une maladie neurodégénérative caractérisée par la mort des motoneurones supérieurs (cortex cérébral projetant dans le bulbe rachidien et la moelle épinière) et inférieurs (bulbe rachidien et/ou corne antérieure de la moelle projetant vers les muscles) conduisant à la paralysie musculaire progressive du patient par un phénomène d'atrophie puis à la mort. L'incidence est très faible (entre 1 et 2,6 nouveaux cas / 100 000 habitants (1) avec un pic entre 51 et 66 ans (2). Les études décrivent un sex-ratio d'environ 1,5 en faveur de la population masculine (3).

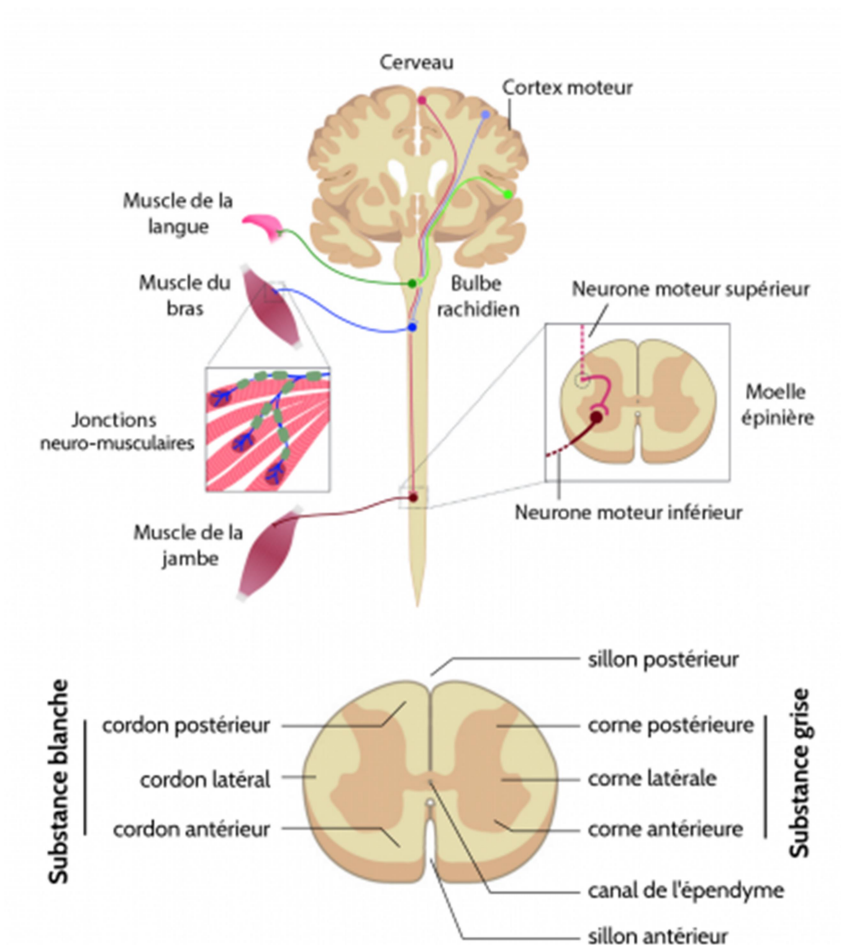


Figure 1 : Système nerveux central et Sclérose Latérale Amyotrophique (4)

Les symptômes sont variables d'un patient à un autre selon le premier site d'atteinte des motoneurones. La forme bulbaire (atteinte des motoneurones supérieurs) caractérisée par une atrophie primaire des muscles de la face. Cette faiblesse musculaire conduit à des troubles de la parole (dysarthrie spastique) ainsi que des troubles de la déglutition (dysphagies, fausses routes). L'atteinte primaire des motoneurones inférieurs, qui conduit à la forme spinale de la SLA, est caractérisée par une atrophie musculaire des membres (inférieurs ou supérieurs). Cette atrophie conduit à des fasciculations accompagnées de crampes amenant le patient à avoir des troubles de la coordination et des troubles de l'équilibre. Enfin, la forme pulmonaire avec une atteinte primaire des motoneurones innervant les muscles de la

respiration. C'est la forme la plus grave avec nécessité d'une assistance respiratoire très rapide (Figure 1).

Pour chaque forme de la SLA, l'évolution est progressive avec une généralisation des symptômes plus ou moins rapide d'un patient à l'autre. Quelle que soit l'atteinte primaire, l'évolution de la pathologie tend vers une atteinte globale et une forme complète (spinale, bulbaire et pulmonaire)

1.2) Diverses étiologies

Environ 10% des cas de SLA sont d'origine familiale avec un mode de transmission autosomique dominant. Parmi les cas de SLA sporadiques, 11% sont expliqués par une cause génétique contre 68% des cas pour les SLA familiales (Figure 2). La première mutation décrite a été la mutation de *SOD1* qui code la Superoxyde Dismutase 1 (5). D'autres facteurs génétiques ont également été découverts comme les mutations de *FUS* (6), *TARDBP* (7) et le nombre de répétitions d'un hexanucléotide dans l'intron du gène *C9orf72* (8). Ces mutations ne sont cependant retrouvées que dans 48% des formes familiales et environ 8% des formes sporadiques (9). D'autres gènes ont été identifiés comme potentiellement impliqués dans la SLA comme *CCNF*, *CHCHD10*, *MATR3*, *OPTN*, *PFN1*, *SQSTM1*, *TBK1*, *TUBA4A*, *UBQLN2* et *VCP* (10–13).

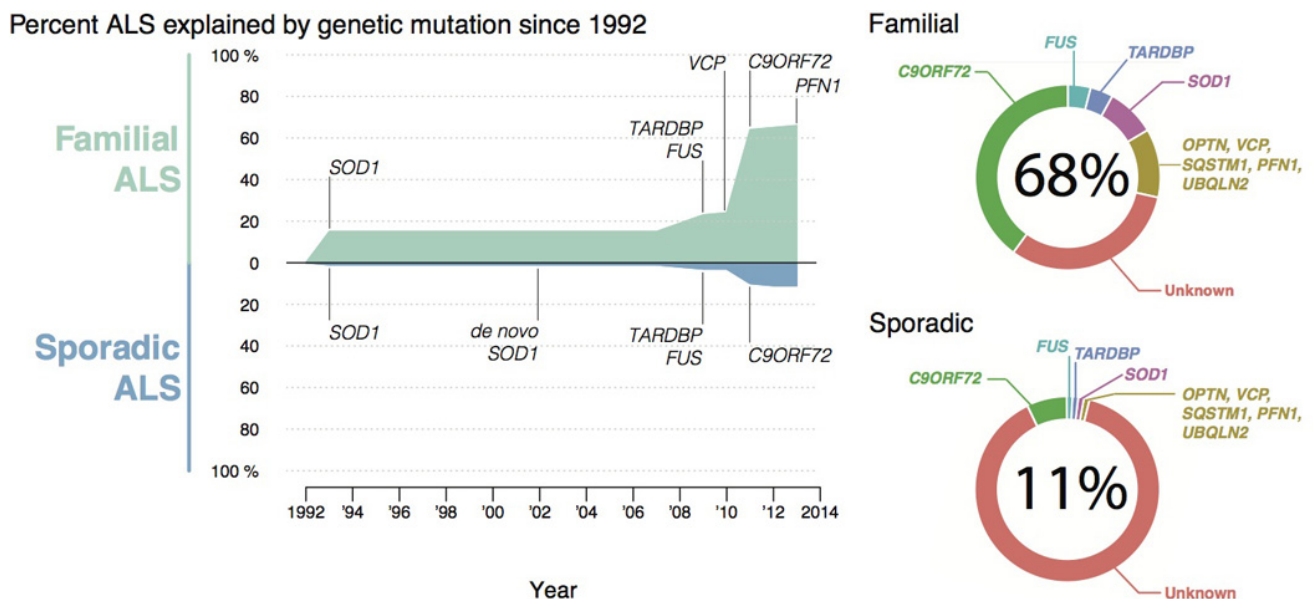


Figure 2 : Chronologie de la génétique dans la SLA. Le pourcentage représente la proportion des cas de SLA expliqués par la génétique (d'après Renton et al, 2014(10))

Les connaissances sur les voies physiopathologiques potentiellement impliquées dans la SLA ne cessent de progresser. Le stress oxydant (14) ainsi que l'excitotoxicité médiée par le glutamate (15,16) semblent admis comme ayant un rôle prépondérant dans la pathogénicité (17). D'autres mécanismes semblent également impliqués comme par exemple le dysfonctionnement mitochondrial (12), la neuro-inflammation (18,19) ainsi qu'une agrégation protéique anormale (20–22). Aujourd'hui, il est clair que tous ces mécanismes peuvent être impliqués dans la maladie mais leur rôle dans la pathogénèse et leur interaction restent non élucidés (23,24). L'impact de l'environnement sur l'apparition et l'évolution de la SLA a également été étudié. L'exposition notamment aux métaux lourds ou d'autres produits toxiques (pesticides, solvants) a été associée à un risque accru de développer la maladie (25).

1.3) Une prise en charge complexe

Le diagnostic de cette pathologie est très difficile (délai diagnostique d'environ 9 à 12 mois) et l'absence de traitement efficace conduit à une survie moyenne courte (médiane de survie : 36 mois). L'évolution de la pathologie est évaluée par le suivi de la capacité vitale forcée (CVF) et le score ALSFRS-R (Amyotrophic Lateral Sclerosis Functionnated Rating Scale-Revised) (26). Aujourd'hui, les patients sont traités par du Riluzole (RILUTEK®) qui permet d'augmenter leur espérance de vie de quelques mois (27–29). La recherche de nouvelles thérapies est très active mais en pratique l'amélioration de la prise en charge des patients est d'avantage basée sur un traitement symptomatique multidisciplinaire (gestion de la douleur, orthophonie, suppléance des fonctions respiratoires et digestives...). Récemment, quelques études ont montré l'intérêt d'une nouvelle molécule (l'Edaravone) pour ralentir la progression de la pathologie avec une évolution moyenne du score ALS-FRS-R de - 5.01 vs 7.50 dans le bras placebo (30). Cependant, d'autres essais cliniques n'ont pas démontré le moindre effet concernant l'évolution de la maladie ou la survie (31). Cependant, l'EMA (European Medicines Agency) a autorisé l'utilisation de cette molécule en thérapeutique sous forme d'une ATU (Autorisation Temporaire d'Utilisation) en France depuis juin 2018 pour la bonne tolérance des patients à la molécule. De nombreuses autres molécules ont également été évaluées pour leurs propriétés anti-glutamatergiques (32), anti-inflammatoires (33), anti-oxydantes (34), neuroprotectrices (35), ou encore un inhibiteur du récepteur CSF1 (36). Cependant, avec un rationnel parfois faible et des méthodologies pré-cliniques et cliniques critiquables, la majorité des essais cliniques ont été décevants (Tableau I) (16).

Tableau I : Essais cliniques étudiant de nouvelles thérapies ciblées sur des anomalies identifiées par analyse métabolomique (publié par Lanznaster et coll (16))

Target	Intervention	Clinical trials for ALS
Glutamatergic overactivation	Perampanel	Phase II, NCT03377309 (Lebanon); NCT03019419 (Japan); NCT03020797 (Unites States).
	Memantine	Ongoing (phase II, NCT02118727, Unites States). [No effect observed in phase II-III; (53)].
Oxidative stress	Inosine	Phase I, NCT02288091 (United States).
	CC100	Phase I, NCT03049046 (United States).
Hypermetabolism	Triheptanoïn	Phase I-II, NCT03506425 (United States).
	High caloric fatty diet	NCT02306590 (Germany).
	Oral nutritional supplementation (high fat and protein)	NCT02152449 (France).

Information available in clinicaltrials.gov.

1.4) La nécessité de biomarqueurs de diagnostic et pronostic

1.4.1) Intérêt des biomarqueurs

Selon le National Institute of Health, un biomarqueur est « une caractéristique biologique mesurée de façon objective et évaluée comme un indicateur soit de processus biologiques normaux ou pathologiques, soit de réponses pharmacologiques résultant d'une intervention thérapeutique » (37). Les

biomarqueurs les plus utilisés sont les paramètres biochimiques ou moléculaires qui peuvent être détectés dans les tissus ou fluides biologiques. Aujourd'hui, l'étude des biomarqueurs est un domaine très vaste (38,39) et devient incontournable dans le domaine de la santé.

Ces biomarqueurs sont classés en catégories selon leur utilisation.

- Les biomarqueurs physiopathologiques permettent l'amélioration des connaissances sur les voies impliquées dans la survenue et dans l'évolution de la maladie
- Les biomarqueurs diagnostiques permettent d'aider le clinicien dans sa démarche diagnostique soit directement soit par un diagnostic d'exclusion pour une pathologie étudiée (40)
- Les biomarqueurs pronostiques ont pour objectif de prédire l'évolution probable de la pathologie ainsi que le niveau de risque associé pour un patient (41)
- Les biomarqueurs prédictifs de la pharmacodynamie et de l'effet thérapeutique permettent de prévoir *a priori* l'efficacité d'un traitement (41,42). Les biomarqueurs prédictifs de la pharmacodynamie sont ceux qui permettent l'évaluation de la dose nécessaire pour obtenir l'effet thérapeutique tout en évitant si possible la survenue d'effets indésirables (médecine personnalisée) alors que les biomarqueurs prédictifs de l'effet thérapeutique permettent de prédire la présence ou l'absence de réponse à un traitement selon les cibles présentes ou par la mise en évidence de voies de shunt métabolique par exemple (43).

Les biomarqueurs peuvent être recherchés dans différents types d'échantillon, que ce soit des cellules isolées (44), des tissus tel que le muscle (45), la peau (46), le foie (47) ou les fluides biologiques tel que le sang (48), le liquide céphalo-rachidien (LCR) (48), les urines (49) ou les larmes (50). Le choix de la matrice est basé sur sa proximité avec l'organe principalement touché par la pathologie. Par exemple, pour les pathologies affectant le système nerveux central, les échantillons provenant de LCR semblent être les plus adaptés. La ponction lombaire pour le recueil de LCR étant un acte très invasif, une alternative est l'utilisation de prélèvements de sang. L'analyse de cet échantillon permet également d'avoir une vision générale du métabolisme et de ne pas se focaliser uniquement sur un organe. Cependant, le manque de spécificité de l'échantillon sanguin vis-à-vis de l'organe impacté par la pathologie peut avoir pour conséquence la diminution de la sensibilité pour la détection et/ou quantification d'un biomarqueur spécifique par un phénomène de dilution ainsi qu'une baisse de la spécificité pour la pathologie étudiée. Pour cela, il est souhaitable de combiner une analyse d'échantillon sanguin à un échantillon plus spécifique comme le LCR lorsque cela est possible.

I.4.2) Vers la médecine personnalisée

La médecine personnalisée (ou médecine de précision) est apparue avec l'essor de la génétique et de la pharmacogénétique. En effet, cette médecine se base sur la personnalisation du traitement et/ou la prévention de la maladie selon le génôme du patient. L'augmentation du nombre de médicaments notamment d'anticancéreux et de biomédicaments utilisables pour le traitement de pathologies similaires sur le plan clinique mais différentes sur le plan biologique ou moléculaire a encouragé le

développement de la médecine personnalisée. En effet il a été démontré en 2001 qu'un traitement par anticancéreux était inefficace dans 75% des cas, probablement à cause d'une stratégie thérapeutique inadaptée (51). Le but de la médecine de précision est d'utiliser différents biomarqueurs pour personnaliser qualitativement et/ou quantitativement le traitement médical d'un individu en fonction de son profil. Aujourd'hui, cette médecine a été associée à la médecine des preuves pour donner la médecine des 5P. Cela signifie qu'en absence de preuves scientifiques de l'efficacité d'un traitement curatif personnalisé et/ou d'un traitement préventif, le rôle de la médecine personnalisée qui permet à un individu de connaître très tôt ses risques génétiques est discutable.

I.4.2.a Description des sous-populations d'une pathologie hétérogène

L'inefficacité d'une thérapie est parfois due à un traitement inadapté à la pathologie ou au patient lui-même. La médecine personnalisée s'intéresse aux caractéristiques du traitement, de la cible et du patient. Avant la mise en place de certains traitements, il est parfois nécessaire de réaliser des tests préalables pour adapter la posologie administrée. L'une des utilisations les plus courantes de ce principe est basée sur l'apport de la pharmacogénétique. Par exemple, avant une mise en place d'un traitement par clozapine, il est recommandé de réaliser l'étude de l'activité du cytochrome 2C9 qui est impliqué dans son métabolisme (52). La posologie pourra être alors adaptée d'emblée dès l'instauration du traitement. Dans le même principe, l'utilisation d'un ou de plusieurs biomarqueurs endogènes pourrait permettre d'évaluer l'efficacité du traitement voire de l'adapter qualitativement (changement de molécule) ou quantitativement (modification de la posologie) (48,53).

Cette approche s'applique également dans des contextes de pathologies complexes et hétérogènes comme la SLA pour lesquelles plusieurs étiologies ont pour conséquence une symptomatologie similaire. Le défi de la médecine personnalisée est de décrire plus finement ces pathologies afin de définir des sous-groupes homogènes et permettre de proposer une thérapie plus adaptée et plus efficace (Figure 3). Cependant, face à la difficulté de discriminer ces sous-groupes par l'examen clinique, il devient nécessaire de trouver un ou plusieurs biomarqueurs permettant de les différencier et de les identifier. Dans ce contexte, les approches « omiques » sont des outils puissants pour déterminer des combinaisons de biomarqueurs spécifiques des mécanismes physiopathologiques selon les stades de la maladie et selon les individus.

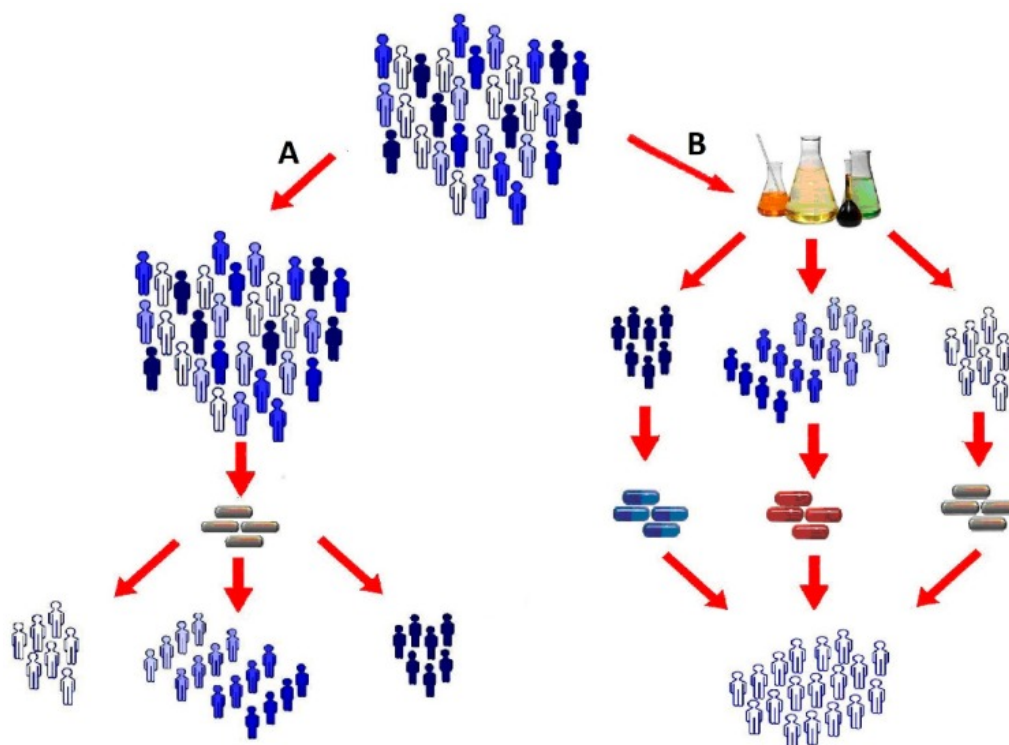


Figure 3 : Schéma représentant l'approche thérapeutique « classique » en A et l'approche de la médecine de précision en B (d'après Balashova et coll, 2018 (54))

I.4.2.b Suivi de l'évolution de la maladie

De manière similaire aux applications mentionnées ci-dessus, l'efficacité de l'approche thérapeutique peut être évaluée par la comparaison de la signature métabolique avant et après l'exposition au médicament. Cette stratégie peut être ciblée sur une voie d'intérêt ou non ciblée. En effet, la correction des anomalies du profil métabolique pourrait être un marqueur indirect de l'amélioration de la pathologie et donc de l'efficacité du traitement. Ce type de suivi pourrait permettre d'évaluer très tôt si le traitement est adapté qualitativement et/ou quantitativement. L'obtention de ces profils de signature métabolique est basée sur l'utilisation d'un ou plusieurs biomarqueurs généralement issues d'études basées sur des approches « omiques ».

I.4.3) Modalités de recherche de biomarqueurs

I.4.3.a) Les approches « classiques »

La plupart des méthodologies évaluant les biomarqueurs sont des approches ciblées, axées sur des perturbations biochimiques connues et identifiées par des expériences *in vivo*, *ex vivo* et *in vitro*. Une analyse ciblée est une analyse développée spécifiquement pour un biomarqueur ou une liste déterminée de biomarqueurs (protéine, métabolite, gène...). Elle a pour avantage de permettre l'obtention d'un résultat rapide, tout en restant spécifique et sensible. L'inconvénient de ce type d'approche est une vision étroite et avec *a priori* de l'état métabolique du patient. L'autre type de méthode est l'approche globale : l'approche « omique ». Cette approche repose sur l'analyse de l'ensemble des constituants de l'échantillon

(que ce soit un tissu, un fluide biologique ou des cellules en culture) à un instant donné. L'approche globale se divise en deux stratégies différentes :

- l'analyse globale ciblée, sans *a priori* permettant l'analyse d'un grand nombre de composés pour lesquels les paramètres d'identification sont déjà connus et recensés dans une base de donnée. Ce type d'approche permet également de comparer les profils obtenus mais comportant spécifiquement un panel de molécules identifiées. Les autres molécules non présentes dans la base de données ne seront pas analysées, ce qui implique une information incomplète.
- l'analyse non ciblée qui repose sur l'analyse de l'ensemble de tous les constituants sans identification préalable. L'objectif de ce type d'analyse n'est pas d'identifier chacune des molécules observées mais de comparer les profils obtenus. Les paramètres pré-analytiques et analytiques doivent être suffisamment peu sélectifs pour permettre une vision complète sans perdre d'informations. Les composés d'intérêts nécessitent ensuite d'être identifiés et pourront être confirmés voire quantifier par des analyses ciblées pour comprendre les voies physiologiques perturbées.

I.4.3.b) Les approches « omiques »

Parmi les approches « omiques », on recense quatre approches reflétant le patrimoine génétique et son expression : la génomique (ensemble des gènes), le séquençage de l'exome complet (ensemble des zones codantes du gène (ou exons)), la transcriptomique (ensemble des séquences génétiques codantes transcrites sous forme d'ARN messager (ou transcrits)) et la protéomique (ensemble des protéines) (Figure 5). Deux autres approches globales de plus en plus utilisées

sont la métabolomique (ensemble des métabolites) qui permet d'obtenir une signature métabolique composée d'une combinaison de petites molécules (<1000 Da) et la lipidomique, ciblée sur l'étude des lipides (Figure 4).

De plus, l'utilisation de ces méthodes complémentaires (55) (Figure 5) a permis

a) d'améliorer la compréhension de nombreuses pathologies dans des domaines variés (56–58)

b) de découvrir de nouveaux biomarqueurs mais également étudier l'impact des modifications génétiques (59,60), de l'environnement (60), ou de toxiques exogènes (55,61)).

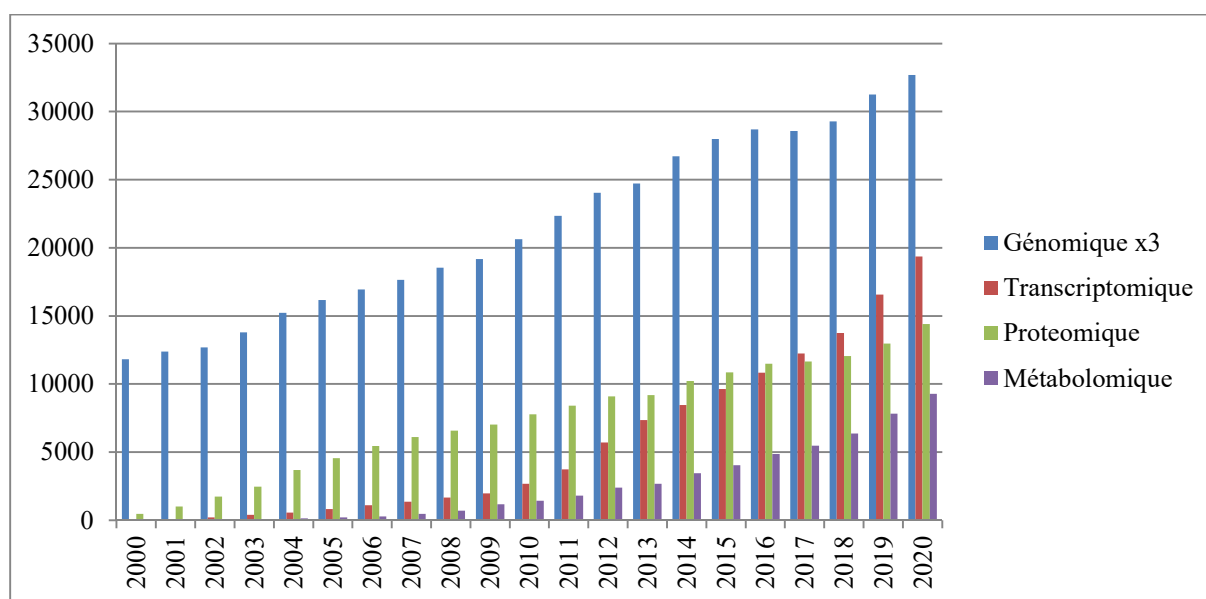


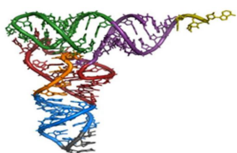
Figure 4: Nombre de publication annuel traitant de génomique, transcriptomique, protéomique et métabolomique ces 20 dernières années (source PubMed ; 07/2021)

La cascade « omics »



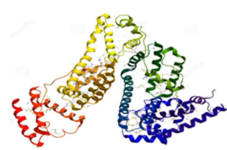
Génome

Ce qu'il peut se passer



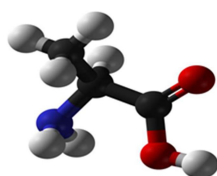
Transcriptome

Ce qu'il semble se passer



Protéome

Ce qu'il fait que cela se passe



Métabolome

Ce qu'il se passe et qu'il s'est passé

Figure 5 : Cascade « omique » (reproduit d'après Dettmer et al, 2007) : de ce qu'il peut se passer jusqu'à ce qu'il se passe et s'est passé

L'analyse de l'échantillon passe par une première étape de préparation (phase pré-analytique), puis par l'étape d'analyse (phase analytique).

Phase pré-analytique

L'étape pré-analytique est l'étape de préparation d'échantillon. Elle consiste à purifier l'échantillon en éliminant les composants de la matrice qui ne sont pas compatibles avec l'analyse comme les globules rouges ou les protéines dans le cas d'échantillons sanguins ainsi que les composants pouvant interférer avec les analytes. Cette étape de préparation doit également permettre d'extraire un maximum de métabolites et de les concentrer selon la sensibilité de l'équipement utilisé pour réaliser l'analyse. Pour

cela, un des grands défis de cette phase est d'extraire un maximum de métabolites en les concentrant tout en éliminant un maximum d'interférents. En pratique, il n'existe pas de protocole de préparation permettant de respecter toutes ces contraintes. Pour cela, le défi est de trouver un équilibre entre la quantité de composés extraits, du rendement d'extraction de ces différents composés et l'élimination des autres éléments de la matrice.

Phase analytique

L'analyse métabolomique s'applique à l'ensemble des métabolites parmi un large spectre de concentrations, polarités et masses (62), grâce à des techniques puissantes telles que la Résonance Magnétique Nucléaire du proton (RMN- ^1H) sans séparation préalable (63) ou la spectrométrie de masse couplée à différents types de technique séparative (électrophorèse capillaire (64), chromatographie gazeuse (65), chromatographie liquide phase inverse ou par interaction hydrophile (HILIC) (49)), différents modes d'ionisation (positive et négative (49)). Chacune de ces techniques est caractérisée par des avantages et des inconvénients. Aujourd'hui, aucune ne permet une vision complète du métabolome, et de nombreuses études les utilisent en combinaison. Malgré tout, cela conduit à une augmentation du coût et de la difficulté à traiter les données recueillies.

1.5) Les méthodes conventionnelles

Actuellement, plus de 114 000 métabolites appartenant à plus de 52 classes différentes et plus de 20 000 protéines ont été décrits en biologie humaine (66,67). L'analyse métabolomique est complexe due à la variabilité des métabolites en terme de classes chimiques (lipides, sucres, acides aminés, acides organiques,

nucléotides...) mais aussi en terme de concentration (du picomolaire au millimolaire) (68,69). Cette complexité d'échantillon qualitative et quantitative nécessite l'utilisation d'outils différents adaptés à chaque famille de molécules. L'utilisation des approches « omiques » a nécessité le développement d'outils analytiques capables d'analyser un grand nombre de molécules en un temps relativement court. Ces technologies sont caractérisées par des critères de performance, d'exigence, et permettent de détecter, d'identifier et de quantifier des molécules à l'aide de propriétés physico-chimiques distinctes. Le tableau II recense les différentes technologies utilisées ainsi que leurs caractéristiques.

Tableau II : Caractéristiques et spécificités des techniques d'analyses métabolomiques (d'après Verpoorte et al) (70)

	HPLC or TLC-UV	GC-MS	LC-MS	MS ⁿ	NMR
Sample preparation	++	—	—	+	+++
Reproducibility	—	+	—	+	+++
Absolute quantitation	—	—	—	—	+++
Relative quantitation	+	++	+	++	+++
Identity	+	++	++	++	++
Compound number	ca. 30	ca. 1000	ca. 200	ca. 1000	ca. 200
Sensitivity	+	++	++	+++	—

La RMN :

La RMN est une technologie très utilisée dans l'analyses métabolomique pour sa capacité d'analyse des échantillons sans préparation préalable, et sa forte reproductibilité (70,71) à moindre coût. De plus, cette analyse est non destructrice, ce qui permet également une analyse dans des tissus intacts (biopsies (72)). De nombreuses études réalisées sur différents fluides biologiques (sérum (73), urine (74), LCR (75), salive (76)) ont démontré la place de la RMN dans l'analyse métabolomique.

L'utilisation de la RMN permet une quantification ou semi-quantification des composés identifiés avec une préparation d'échantillon simple. La RMN ne permet pas de détecter simultanément plus de quelques centaines de molécules (70). De plus, sa sensibilité de l'ordre du micromolaire rend cette technique insuffisante pour une vision complète du métabolome (77). Ce type d'analyse étant réalisé sans séparation des différents composés de l'échantillon, il peut donner lieu à des superpositions de signaux rendant l'identification et/ou la quantification plus difficile. La RMN à deux dimensions (RMN 2D) peut être utilisée pour surmonter le problème du chevauchement des résonances en séparant les signaux dans une deuxième dimension basée sur une autre propriété physique «orthogonale» de l'atome ou des atomes d'intérêt (par exemple, un voisin attaché de manière covalente, un temps de relaxation, une constante de couplage, etc.). La résolution supplémentaire disponible par RMN 2D permet potentiellement de détecter et d'identifier plus de métabolites que ne le permet la RMN à une dimension (RMN 1D).

La spectrométrie de masse :

La spectrométrie de masse est une technologie reconnue pour sa grande spécificité, sa haute sensibilité et sa capacité à quantifier les différents composés. La séparation préalable des composés permet d'accroître sa sensibilité, et ses capacités d'identification et de quantification. De nombreuses études ont démontré l'intérêt de la spectrométrie de masse dans l'analyse métabolomique de divers types d'échantillons biologiques tels que le plasma (78), l'urine (79), le LCR (80), la salive (81) ou encore le tissu cérébral (82) ou musculaire (83).

A l'inverse de la RMN, la spectrométrie de masse connue pour sa grande sensibilité nécessite une préparation d'échantillon au préalable (précipitation des protéines,

dérivation pour une analyse par GC-MS), ainsi qu'une séparation chromatographique des différents composés. La sélectivité de ces étapes mène à une perte d'information. Pour cela, l'analyse par spectrométrie de masse nécessite de combiner plusieurs analyses avec différents protocoles de préparation d'échantillon. De plus, les signaux par spectrométrie de masse ne sont pas toujours corrélés aux concentrations des métabolites. En effet, la spectrométrie de masse est particulièrement sensible aux effets matrice et à une instabilité inter-jour nécessitant l'ajout d'étalons internes pour corriger ces variations. Malheureusement, dans une analyse globale de type métabolomique, il n'est pas possible de proposer un étalon interne pour chaque métabolite étudié.

RATIONNEL ET OBJECTIFS

La recherche de biomarqueurs représente actuellement un enjeu majeur dans le contexte de pathologies en impasse diagnostique et pronostique, telle que la SLA. Les progrès dans ce domaine sont accompagnés de développements technologiques au niveau pré-analytique et analytique mais les limites nécessitent d'adapter les stratégies d'utilisation.

I) Applications des méthodes conventionnelles à la recherche de biomarqueurs dans la SLA

Devant la multitude de paramètres potentiellement impliqués dans l'étiologie de la SLA, il n'existe pas de marqueurs biologiques validés ni même de panels de biomarqueurs pour aider au diagnostic. Certains biomarqueurs diagnostiques et/ou pronostiques potentiels ont été identifiés comme la créatinine plasmatique (84) et le

neurofilament dans le sang et le liquide céphalo-rachidien (LCR) (85). En effet, une diminution de la créatinine a été rapportée dans plusieurs études (86–88). Cependant, cette baisse semble plus être une conséquence de l'atrophie musculaire qu'une cause pathologique. Ainsi, la créatinine plasmatique semble plus être un marqueur de suivi de la dégénérescence musculaire. Les neurofilaments sont reconnus comme les biomarqueurs indispensables à explorer chez tous patients suspects de SLA. Il faut noter la présence de chaînes légères de neurofilament au niveau plasmatique et dans le LCR de patients dans différentes pathologies caractérisées par une atteinte neuronale comme la maladie de Parkinson (89), la sclérose en plaque (90), ou encore lors de phénomènes de neurodégénération (91). Le manque de standardisation des méthodes de dosage et l'absence de validation rigoureuse sur une cohorte limitent son utilisation, mais il est aujourd'hui évalué dans tous les essais cliniques pour suivre la réponse aux traitements (92). Ainsi les biomarqueurs les plus utilisés aujourd'hui dans la SLA souffrent d'un manque de spécificité.

Par ailleurs, les références pour juger de l'évolution clinique des malades peuvent être discutées car leur sensibilité ne permet pas de déceler rapidement une amélioration clinique subtile. Les échelles fonctionnelles utilisées pour évaluer la motricité chez les patients SLA (ALSFRS-R, capacité vitale forcée et tests musculaires) peuvent ne pas être suffisamment adaptées pour le suivi de la réponse médicamenteuse. Ainsi, la recherche et l'identification de biomarqueurs fiables pour le diagnostic et le pronostic de la SLA sont de la plus haute importance, car leur suivi aiderait à l'identification des phénotypes de réponse aux médicaments, améliorant ainsi l'évaluation de l'efficacité du traitement.

Dans ce contexte, les études utilisant les approches omiques se sont succédées depuis plus de 10 ans. Les perturbations du métabolisme du glutamate ainsi que l'impact du stress oxydant restent les mécanismes physiopathologiques qui sont le plus souvent retrouvés (93–96). En effet, des augmentations du glutathion (97) et des radicaux ascorbiques (98) ont été retrouvées, ces deux molécules étant fortement impliquées dans la défense au stress oxydatif. Tohgi et coll ont également observé une augmentation des nitrates (97), produit de la réaction du NO et des espèces réactives de l'oxygène, et de la 3-nitro-tyrosine (99), produit final de l'oxydation des peroxynitrites. Des altérations du métabolisme lipidique ont également pu être mises en évidence (100). Bien que ces études aient démontré des perturbations dans les mêmes voies métaboliques, elles n'ont pas permis d'identifier un ou plusieurs métabolites communs entre ces travaux (101,102).

Ainsi, malgré les efforts déployés, aucun profil métabolique n'a encore été validé en tant que biomarqueur pronostique ou diagnostique et aucune thérapeutique ciblant un biomarqueur ou une voie métabolique impactée n'a montré son efficacité (Tableau I, p10). Cet échec souligne le besoin urgent de progrès dans les stratégies « omiques » en combinant les informations issues de différentes analyses (103) provenant de différentes matrices chez les patients SLA, comme le LCR, le sang et le muscles (104) en combinant différentes approches (protéomique, transcriptomique, lipidomique, etc.) (105).

II) La spectroscopie vibrationnelle en biologie

Les techniques de spectroscopie vibrationnelle, que ce soit l'absorption infrarouge (IR) ou la diffusion Raman, ont été largement explorées au cours des deux dernières

décennies pour obtenir la composition biochimique de fluides corporels dans le domaine de l'analyse biomédicale (106–115). Leur application aux biomolécules et les tissus a été réalisée pour la première fois dès les années 1960 et très rapidement dans les années 70 des applications dans le domaine biomédical ont été documentées (116–118).

Les deux techniques sont rapides, peu coûteuses, faciles à utiliser, non destructives et permettent d'obtenir une signature moléculaire unique à partir d'un faible volume d'échantillon sans étape de préparation et sans utilisation de solvants ou autres réactifs. Le développement d'approches analytiques basées sur ces techniques sont particulièrement attractives pour l'analyse de routine de divers fluides corporels (106–110). La spectroscopie d'absorption infrarouge et la spectroscopie de diffusion Raman sont des approches complémentaires qui permettent d'enregistrer des signatures spectrales spécifiques de la composition moléculaire de l'échantillon (aussi appelée empreinte moléculaire). Les avancées technologiques dans le domaine de l'instrumentation et le perfectionnement du traitement chimiométrique de données (analyses statistiques multivariées) ont renforcé leur positionnement sur des problématiques de détection et de quantification de biomarqueurs dans des matrices biologiques complexes. Ainsi, une analyse rapide et peu invasive pourrait permettre le dépistage précoce de pathologies (119–125). Les techniques sont communément décrites comme complémentaires, ce qui se justifie par les phénomènes physiques fondamentaux très différents qui les régissent avec des défis techniques spécifiques pour leur mise en œuvre clinique (111).

La sensibilité pour détecter les changements subtils dans la composition biochimique peut être utilisée pour identifier des biomarqueurs spécifiques pour le diagnostic

précoce de diverses pathologies comme par exemple pour la malaria ou la maladie d'Alzheimer (126,127). De plus certaines études ont démontré que cet outil pouvait également être utilisé dans le cadre de la surveillance des thérapeutiques (128–133).

III) OBJECTIFS

Ainsi, l'objectif de ce travail était d'optimiser l'utilisation de la métabolomique pour la recherche de biomarqueurs, en prenant l'exemple de la recherche de biomarqueurs dans la SLA par 1) l'évaluation de l'intérêt de la combinaison des méthodes conventionnelles, 2) l'exploration de nouvelles méthodes analytiques et 3) l'application de ces méthodes dans la SLA.

Objectif 1 (Partie I)

La première partie de ce travail consiste à évaluer si une combinaison de différentes techniques analytiques permet d'avoir un profil métabolique plus riche et plus robuste. Nous avons étudié 4 analyses différentes : une méthode d'analyse ciblée par FIA-MSMS (Flow Injection Analysis–MSMS), une méthode d'analyse non ciblée par GC-MS, une méthode d'analyse ciblée par LC-HRMS et une méthode d'analyse par RMN. Les résultats obtenus nous ont confirmé l'intérêt de combiner différents protocoles analytiques pour augmenter la couverture métabolique.

Objectif 2 (Partie II)

Pour pallier ce déficit inhérent à la préparation de l'échantillon et aux méthodes d'analyse utilisées, nous nous sommes orientés vers de nouvelles technologies qui permettent d'obtenir une vision complète de la composition de l'échantillon et qui ont démontré leur intérêt dans cette approche complexe. Nous avons aussi l'ambition d'évaluer l'intérêt d'une nouvelle technique analytique pouvant être complémentaire des outils standards, à utiliser en première approche ou en combinaison. Dans ce but, nous avons travaillé dans un premier temps sur l'optimisation de l'analyse du

plasma humain par spectroscopie infrarouge par réflectance totale atténuée (Article n°2 et études n°3 et 4). Nous avons ensuite appliqué le protocole déterminé à des échantillons patients dans la troisième partie de ce travail pour évaluer l'apport des données spectroscopiques vis-à-vis de celles obtenues par les méthodes conventionnelles seule ou en combinaison (Etude n°5). Pour finir, une collaboration avec une équipe de recherche de Dublin nous a permis de commencer cette même étude en utilisant la spectroscopie Raman (Projets en cours).

MATERIEL ET METHODES

I) Les méthodes conventionnelles

I.1) La RMN

La RMN (ou résonnance magnétique nucléaire) est basée sur le comportement des atomes sous l'influence d'un champ magnétique. Le noyau d'un atome est constitué de protons et de neutrons. Chacune de ces particules est considérée comme un micro-aimant caractérisée par un spin (grandeur quantique). Le spin correspond au mouvement intrinsèque d'une particule élémentaire de l'atome (neutron, proton et électron) et correspond au moment magnétique de la particule. On observe généralement le spectre RMN du proton (^1H) mais peuvent également être réalisés des spectres RMN sur le ^{13}C , ^{15}N , ^{31}P .

En absence de champs magnétique, la répartition des moments magnétiques des spins est aléatoire et le moment magnétique total résultant est nul. Lorsque l'échantillon est placé dans un champ magnétique B_0 , l'ensemble des spins s'alignent parallèlement à l'axe B_0 soit dans la même direction que B_0 (spin positif) ou dans la direction opposé (spin négatif). La phase d'excitation correspond à l'application d'un second champ B_1 perpendiculaire à B_0 (onde radio fréquentée).

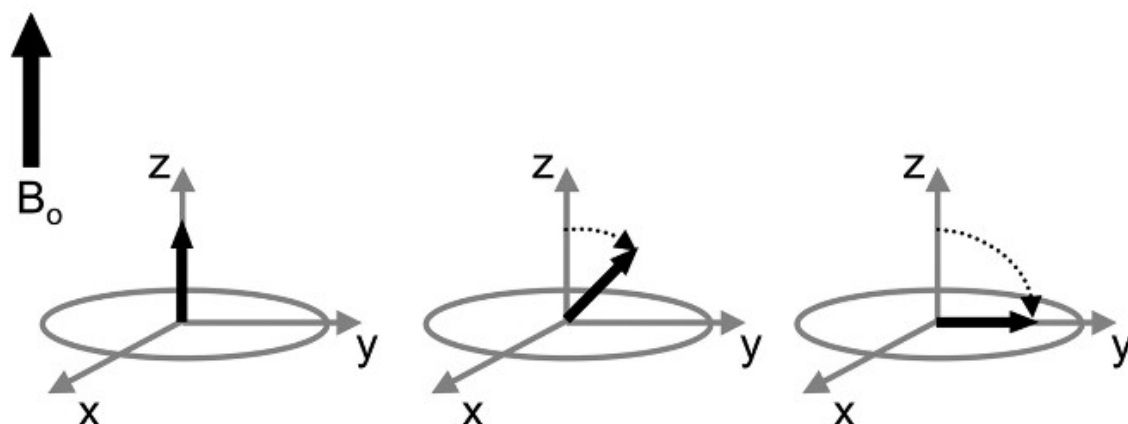


Figure 6 : Schéma représentant la polarisation des spins et leur relaxation permettant l'analyse en RMN. A gauche, avant l'impulsion RF, la petite flèche noire représente l'aimantation microscopique, parallèle au champ B_0 . Au centre et à droite, le pulse RF produit à la fréquence de Larmor provoque un déplacement de l'axe de rotation du proton de z vers y (134)

Dans un troisième temps, à l'arrêt du champ magnétique, les spins vont relaxer sous la forme d'un signal spécifique par son temps de décroissance et sa fréquence de rotation (Figure 6). Ce signal émis est capté par une bobine de direction perpendiculaire à B_0 . Le signal recueilli sous forme d'une onde magnétique donne ensuite un spectre RMN par un processus mathématique (la transformée de Fourier) qui déconvolue l'onde mesurée en plusieurs ondes de fréquences de résonances différentes.

En RMN, chaque signal est caractérisé par un déplacement chimique, une constante de couplage, une multiplicité et une intensité (Figure 7).

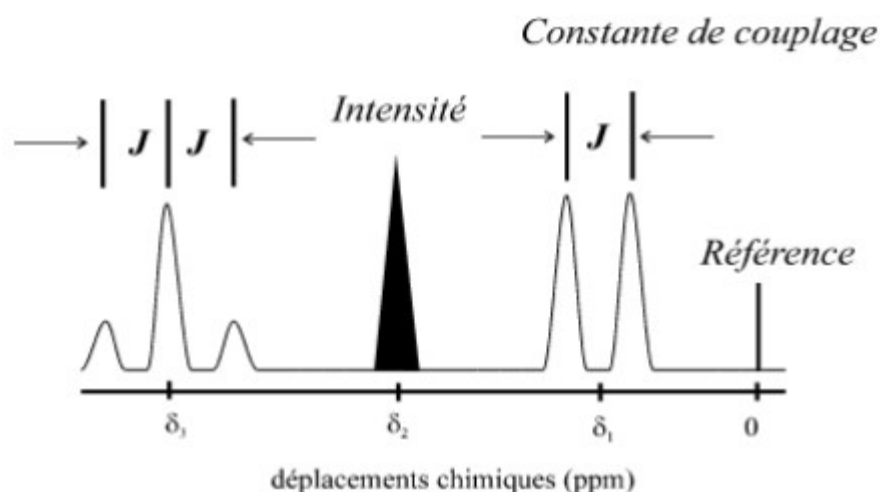


Figure 7 : Schéma des paramètres d'identification en RMN : le déplacement chimique δ en abscisse (en ppm), la multiplicité (ou constante de couplage) et l'intensité (proportionnelle à la concentration)

Le déplacement chimique (δ) définit la position du signal selon l'axe des abscisses (fréquence mesurée). Il s'exprime en ppm (partie par million) et est normalisé par le déplacement chimique du TSP (acide trimethylsilylpropanoïque) qui est un analyte de référence universel. La multiplicité d'un signal correspond au nombre de pic appartenant à un seul atome. Il dépend de l'environnement chimique de l'atome (Figure 8). La constante de couplage correspond à l'écart entre les différents déplacements chimiques d'un massif de signaux correspondant à un seul atome (doublet, triplet...). Enfin, l'intensité dépend du nombre de proton ou d'atome qui possède le même signal mais également à la concentration de la molécule dans l'échantillon analysé.

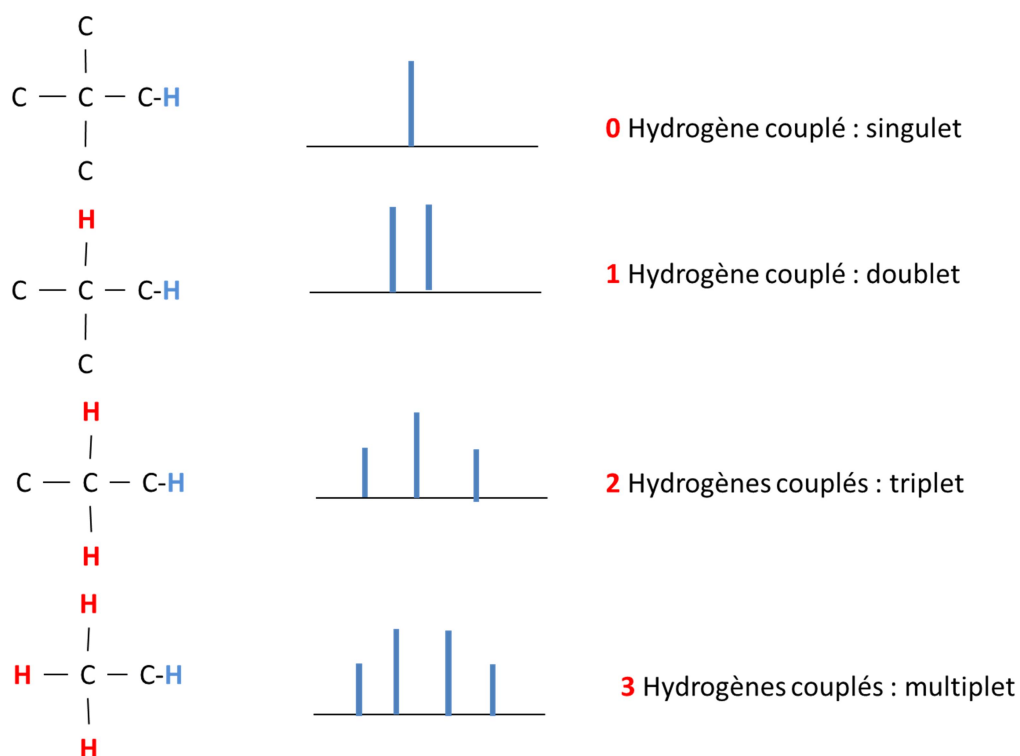


Figure 8 : Représentation des multiplicités de signaux visibles en RMN 1H

Un des inconvénients de la RMN est la superposition de signaux provenant de différents métabolites rendant l'identification et la quantification plus difficile. De plus, sa sensibilité de l'ordre du micromolaire ne permet pas une étude complète du métabolome.

I.2) La chromatographie

I.2.1) Détection

En chimie analytique, la partie chromatographique de l'analyse permet uniquement la séparation des composés. Il est alors nécessaire de coupler un détecteur après la colonne chromatographique pour pouvoir identifier et/ou quantifier le ou les

molécule(s) d'intérêt. En fonction des caractéristiques du ou des analytes concernés, il est possible de coupler des détecteurs spectroscopiques (détecteur d'absorption UV-visible, détecteur à fluorescence, détecteur d'absorption infrarouge), des détecteurs électrochimiques (ampérométriques, coulométriques, potentiométrique), un détecteur évaporatif à diffusion de lumière (DEDL) ou des détecteurs spectraux (spectrométrie de masse ou spectres atomiques). Dans ce travail, seul le détecteur par spectrométrie de masse a été utilisé et sera détaillé.

La spectrométrie de masse est une méthode d'analyse basée sur le comportement d'une particule chargée dans un champ magnétique en fonction de sa charge électrique et de sa masse. Un spectromètre de masse est capable de détecter, identifier et/ou quantifier les composants d'un échantillon en les ionisant puis en séparant les ions selon leur rapport masse sur charge (m/z). Quel que soit l'analyseur, la spectrométrie de masse nécessite une source d'ionisation pour produire les ions moléculaires, un sélecteur de masse (selon le rapport m/z) puis un détecteur.

Le rôle de la source d'ionisation est de transformer les molécules présentes en phase liquide, solide ou gazeuse en un mélange d'ions en phase gazeuse. Différentes sources d'ionisations sont utilisables en spectrométrie de masse, chacune ayant ses avantages et ses inconvénients. Certaines sources permettent une ionisation douce (ex : ionisation par électrospray, MALDI) tandis que d'autres sont plus énergétiques (ionisation par impact électronique).

Les ions générés dans la source sont ensuite conduits au sélecteur puis au détecteur. Dans ce travail, seul le sélecteur de type quadripôle a été utilisé. Le

quadripôle est un sélecteur de masse constitué de quatre électrodes en forme de barreaux. Un courant alternatif parcourt chaque barreau de manière à créer un champ magnétique tournant au centre du quadripôle. L'intensité du courant ainsi que la fréquence d'alternance permet la sélection des ions d'un rapport m/z choisi. En effet, les ions du rapport masse/charge souhaité ont une trajectoire en spirale entre les quatre barreaux métalliques du quadripôle jusqu'au détecteur tandis que les autres rentrent en contact avec un barreau et sont annihilés (Figure 9).

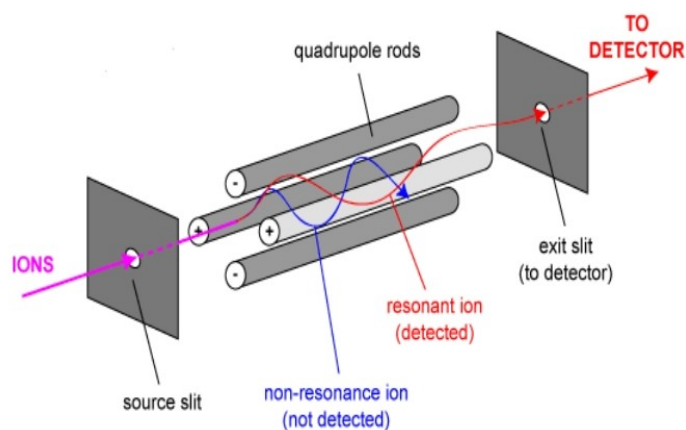


Figure 9 : Schéma d'un sélecteur de masse de type quadripôle

I.2.2) Séparation chromatographique

La chromatographie est une technique d'analyse chimique et/ou de purification qui vise à séparer les différents composants d'un mélange en exploitant leurs propriétés physico-chimiques. La séparation est réalisée grâce à l'utilisation d'une phase stationnaire (colonne chromatographique) et d'une phase mobile (liquide ou gazeuse). Il existe différentes techniques chromatographiques qui s'appuient sur la polarité (chromatographie d'adsorption), l'encombrement stérique (chromatographie d'exclusion stérique), les interactions hydrophiles (chromatographie par interactions hydrophiles), la volatilité (chromatographie en phase gazeuse) ou encore des

affinités antigène-anticorps (chromatographie d'affinité) pour séparer les constituants de l'échantillon. Dans le contexte des analyses métabolomiques, les techniques couramment utilisées sont la chromatographie en phase gazeuse (CPG), la chromatographie d'adsorption et la chromatographie par interactions hydrophiles (HILIC). Seule la chromatographie d'adsorption sera évoquée par la suite.

I.2.2.a) La phase stationnaire

La chromatographie d'adsorption utilise un phénomène de partage entre deux phases : la phase mobile et la phase stationnaire. La phase stationnaire est constituée de microparticules de silice greffées. La phase stationnaire est apparentée à un film liquide non miscible à la phase mobile. Elle permet la séparation de composés qui possèdent une affinité pour la phase stationnaire permettant la rétention et pour la phase mobile permettant leur élution. Les composés se répartissent entre les deux phases selon leur polarité. Ce partage dépend du facteur K qui correspond au rapport des activités chimiques d'un soluté entre deux phases non miscibles. Dans ce cas, l'activité chimique correspond à la concentration du soluté dans chacune des deux phases. Ce coefficient de partage est dépendant des conditions chromatographiques telles que la température, et la composition des phases mobile et stationnaire.

$$K = \frac{C_s}{C_M}$$

Equation 1 : Relation entre le coefficient de partage et les concentrations dans la phase mobile (C_M) et dans la phase stationnaire (C_S)

Il existe deux modes de séparation qui reposent principalement sur les propriétés de la phase stationnaire (colonne) :

- La chromatographie en phase normale
- La chromatographie en phase inverse

Dans le cas de la chromatographie en phase normale, la phase stationnaire est constituée de silice greffée avec des groupements polaires de type amine (NH_2), nitrile (CN) ou dialcool ($\text{CHOH-CH}_2\text{OH}$). Cette phase a la particularité de retenir les molécules ayant un caractère plutôt polaire.

Dans le cas de la chromatographie en phase inverse, la phase stationnaire est constituée de silice greffée avec des chaînes carbonées (C_8 , C_{18}). Cette phase possède plus d'affinité pour les composés apolaires.

En biologie, les composés d'intérêt sont majoritairement polaires. Ils sont fortement retenus par les colonnes utilisées en chromatographie en phase normale, ce qui rend leur séparation difficile et longue. Au contraire, la chromatographie en phase inverse les retient faiblement et permet une séparation plus rapide et plus efficace. Cependant la chromatographie en phase inverse n'est pas adaptée pour l'analyse des lipides. Pour cela, lors des analyses métabolomiques, ce mode d'analyse est parfois complété par une séparation en mode HILIC (49).

I.2.2.b) La phase mobile

En chromatographie en phase liquide, la phase mobile est généralement composée d'un mélange d'un solvant fort (minoritaire) favorisant l'élution et d'un solvant faible (majoritaire) ayant peu d'affinité pour les composés à séparer alors qu'en

chromatographie en phase gazeuse, la phase mobile est un gaz vecteur neutre (généralement de l'hélium). En phase normale, la phase mobile est composée majoritairement d'un solvant organique (par exemple l'hexane) dit solvant faible et minoritairement d'un solvant polaire (eau). En phase inverse, la phase mobile est composée majoritairement d'eau et minoritairement de solvant organique (méthanol, acétonitrile).

Deux modes de séparations existent : le mode isocratique pour lequel la composition de la phase mobile reste identique pendant toute la durée de l'analyse et le mode gradient pour lequel la proportion de solvant fort (ou la température du gaz vecteur) augmente progressivement au cours de l'analyse. Ce dernier permet de modifier la composition de la phase mobile au cours de l'analyse et donc de modifier le coefficient de partage. Ainsi l'analyse de composés très différents devient possible sur une durée d'analyse relativement courte. Pour les analyses métabolomiques, la complexité de la matrice et le nombre de métabolite à analyser nécessite l'utilisation du mode de séparation en gradient.

II) La spectroscopie vibrationnelle

II.1) Principes

La spectroscopie vibrationnelle fait partie des spectroscopies optiques mettant en jeu des interactions entre un rayonnement électromagnétique et un échantillon. Lorsqu'un faisceau lumineux est envoyé sur un échantillon, celui-ci subit plusieurs phénomènes physiques : des phénomènes de transmission, d'absorption, de diffusion, et de réfraction (Figure 10). L'absorption peut ensuite induire des

phénomènes de désexcitation non-radiatifs mais également une photoluminescence (fluorescence ou phosphorescence).

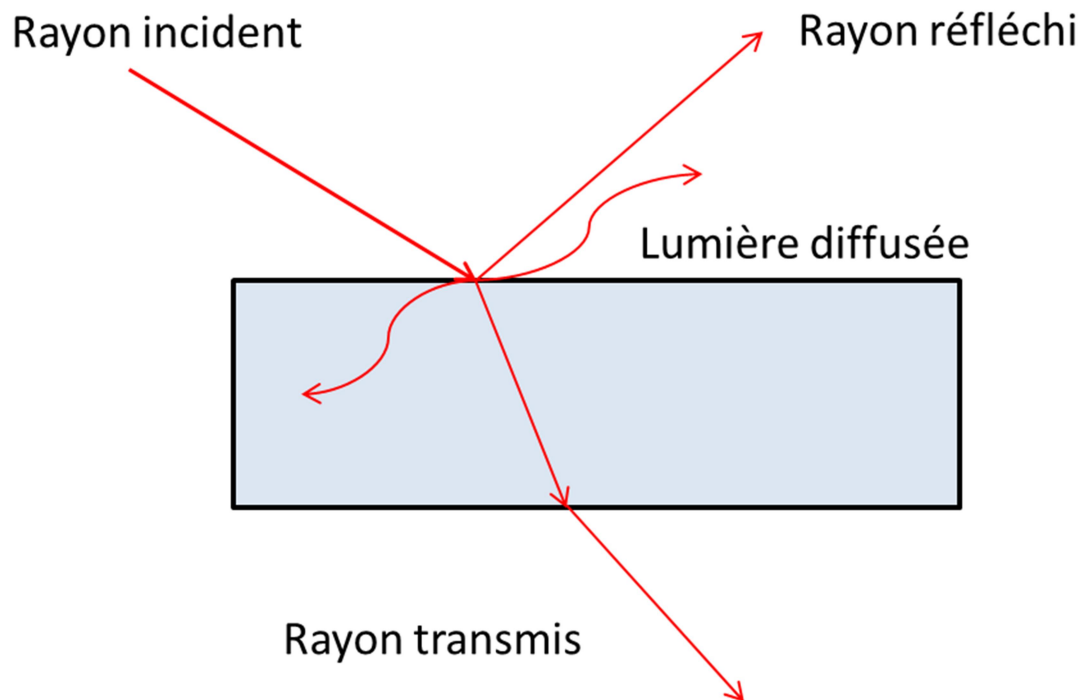


Figure 10 : Schéma des différents phénomènes physiques impactant le trajet lumineux

Le terme de spectroscopie vibrationnelle fait référence principalement à trois techniques de spectroscopie (spectroscopie d'absorption infrarouge, spectroscopie de diffusion Raman et spectroscopie Rayleigh).

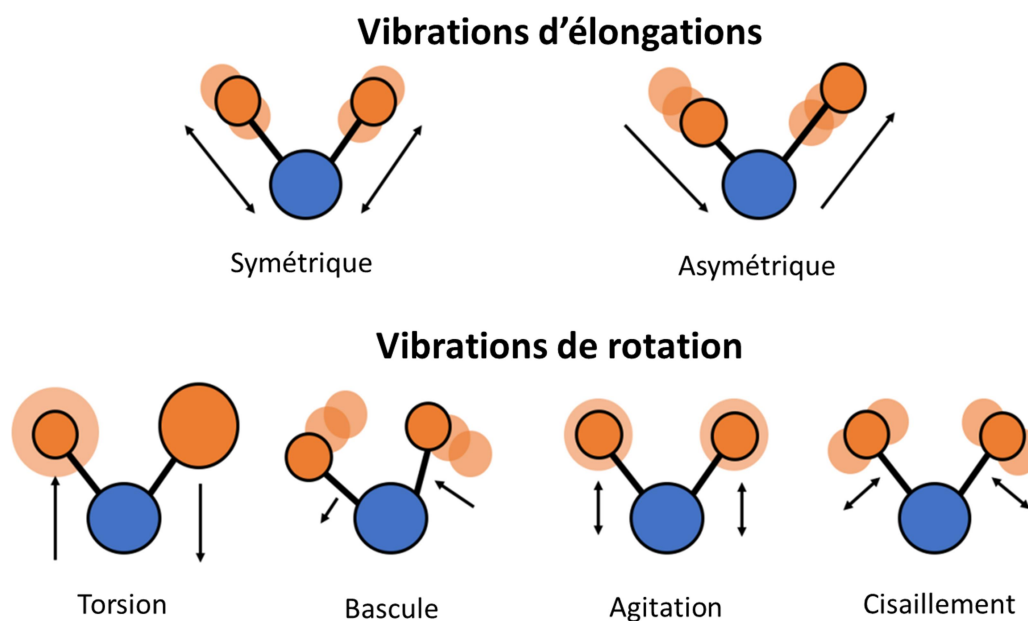
Ces technologies sont basées sur une absorption ou une émission d'un signal lumineux impliquant un phénomène de vibrations de liaisons covalentes. Pour qu'une vibration soit active en infrarouge, un changement du moment dipolaire doit se produire alors qu'il faut un changement de polarisabilité en spectroscopie Raman. En effet les liaisons polaires ont un signal de forte intensité en infrarouge alors que les liaisons apolaires « symétriques » ont un meilleur signal en Raman. Par exemple,

l'absorbance de la molécule d'eau est très forte en infrarouge alors que le signal Raman est très faible en comparaison.

Bien que la spectroscopie infrarouge mesure directement l'absorption des rayons infrarouge et nécessite donc une source d'émission dans l'infrarouge, la spectroscopie Raman doit utiliser une source laser monochromatique dans l'UV, dans le spectre visible ou proche infrarouge. Cependant, il est parfois préférable d'utiliser une source d'émission dans le proche infrarouge pour limiter les interférences dues à la fluorescence ou la photo dégradation (135).

II.2) Les modes de vibrations

Les modes vibrationnels sont classés en deux catégories : vibrations d'élongation et vibration de déformations angulaires. Les vibrations d'élongation concernent les vibrations pour lesquelles la longueur de la liaison covalente varie. Elles regroupent les vibrations d'élongation symétriques et asymétriques. Les vibrations de rotations concernent les vibrations pour lesquelles l'angle entre les liaisons covalentes varie. Elles regroupent les vibrations de torsion, de bascule, d'agitation et de cisaillement (Figure 11). Chaque liaison possède des énergies spécifiques de vibration pour lesquelles elles vibrent ou tournent. Ces énergies sont dépendantes des masses atomiques des atomes, de la force des liaisons (simple, double, triple) ainsi que l'environnement chimique.



*Figure 11 : Modes de vibrations des liaisons covalentes (Cameron et al, 2020
(136))*

Lorsqu'une liaison reçoit un photon dont l'énergie est équivalente à la différence énergétique entre deux niveaux d'excitation, il existe une probabilité que ce dernier soit absorbé et que l'amplitude de la vibration de la liaison augmente (Figure 12). Concernant les rayons réfléchis, on peut voir que l'effet Raman Stokes est d'abord constitué d'une phase d'excitation du système de l'état fondamental à un niveau d'énergie virtuel. Ensuite, le système se désexcite vers un niveau vibrationnel différent de celui de départ. Le système retourne ensuite à son état initial par une dissipation de l'énergie restante par des phénomènes vibrationnels. Concernant l'effet Raman anti-Stokes, le système doit être dans un état vibrationnel excité. Puis, le système se désexcite jusqu'au niveau énergétique fondamental. La probabilité que le système soit dans un état fondamental est toujours plus forte que la probabilité qu'il soit dans un état vibrationnel excité. En conséquence, en spectroscopie Raman,

l'intensité de la diffusion Raman Stokes est toujours supérieure à celle de la diffusion Raman anti-Stokes.

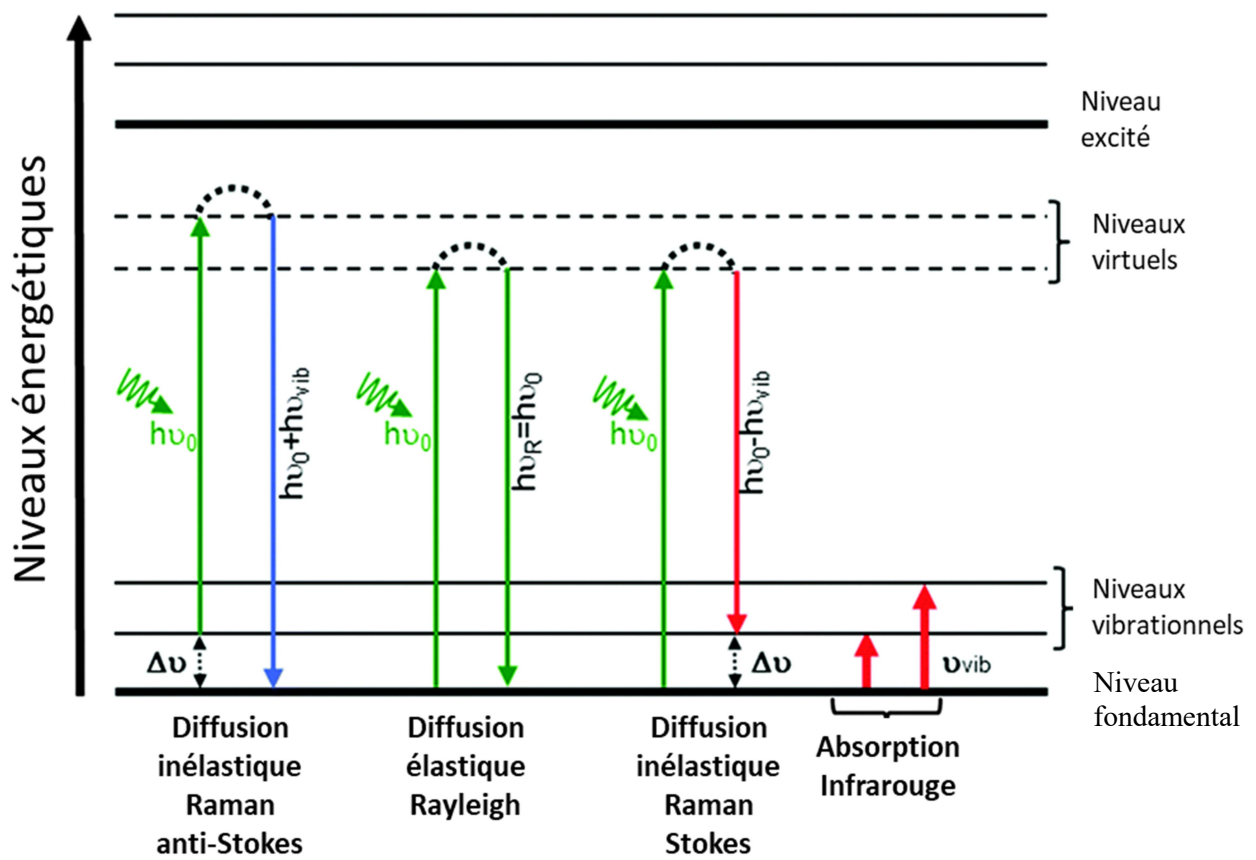


Figure 12 : Diagramme de Jablonski représentant les phénomènes d'absorption infrarouge, de diffusion Raman Stokes, Raman anti-Stokes et Rayleigh (adapté de Baker et al, 2016 (108))

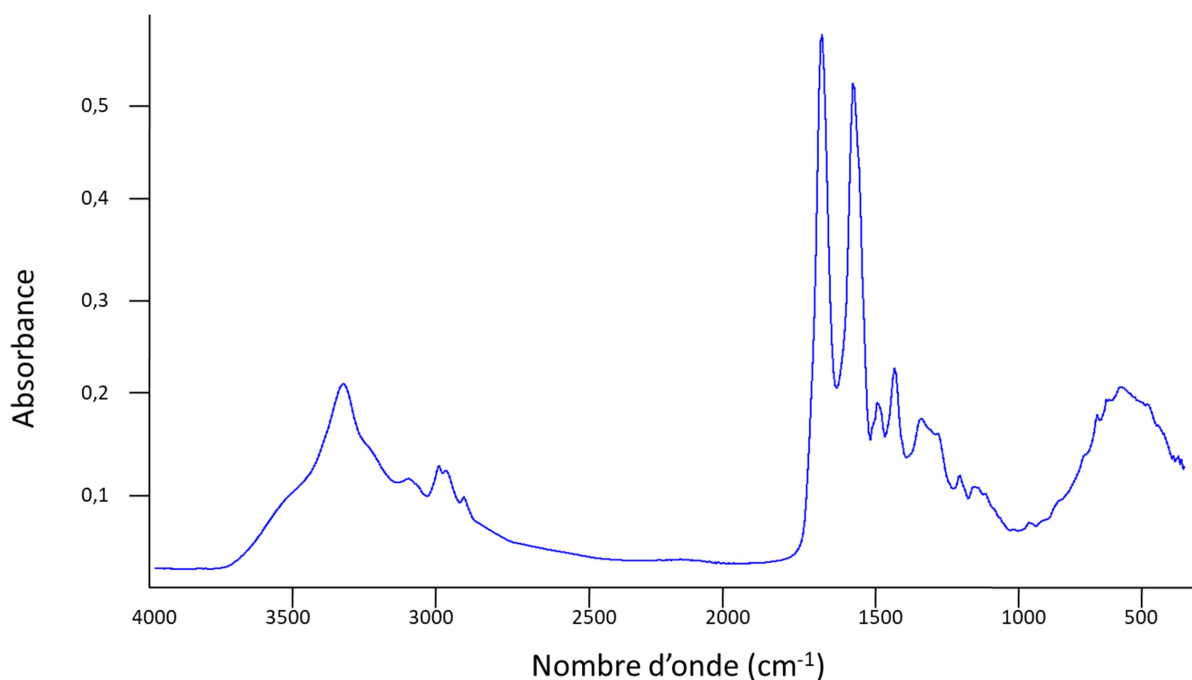


Figure 13 : Spectre d'absorption ATR-IR obtenu à partir d'un plasma humain

Le spectre ATR-IR d'un plasma humain peut être divisé comme suit (Figure 13):

- L'absorption entre 2000 cm^{-1} et 4000 cm^{-1} est due aux vibrations d'élongation. La présence d'une forte et large bande avec un maximum à 3280 cm^{-1} qui peut être attribuée aux vibrations d'élongation des liaisons N-H des groupements amides. Les groupements CH_2 des chaînes carbonées à 2855 cm^{-1} et 2926 cm^{-1} représentent respectivement les vibrations d'élongation symétrique et asymétrique. Dans cette zone du spectre, on observe également une large bande entre 3100 et 3600 cm^{-1} correspondant aux vibrations d'élongation des fonctions OH.
- Les bandes amides I et II se caractérisent par deux pics de forte intensité à 1640 cm^{-1} et 1580 cm^{-1} .
- La région de l'empreinte digitale rassemble également d'autres types de vibrations d'élongation et de déformation pouvant se trouver entre 600 cm^{-1} et

1500 cm^{-1} . Le pic à 1447 cm^{-1} peut être attribué aux vibrations de déformation dans le plan (cisaillement) des CH_2 et des CH_3 . Les groupements carboxylates ainsi que les déformations des liaisons CH_3 peuvent être responsables de la bande d'absorption à 1395 cm^{-1} . Les liaisons amides provoquent également une bande d'absorption de faible intensité vers 1250 cm^{-1} . Les vibrations d'élongations des liaisons C-C se caractérisent par des pics à 1108 cm^{-1} et 938 cm^{-1} .

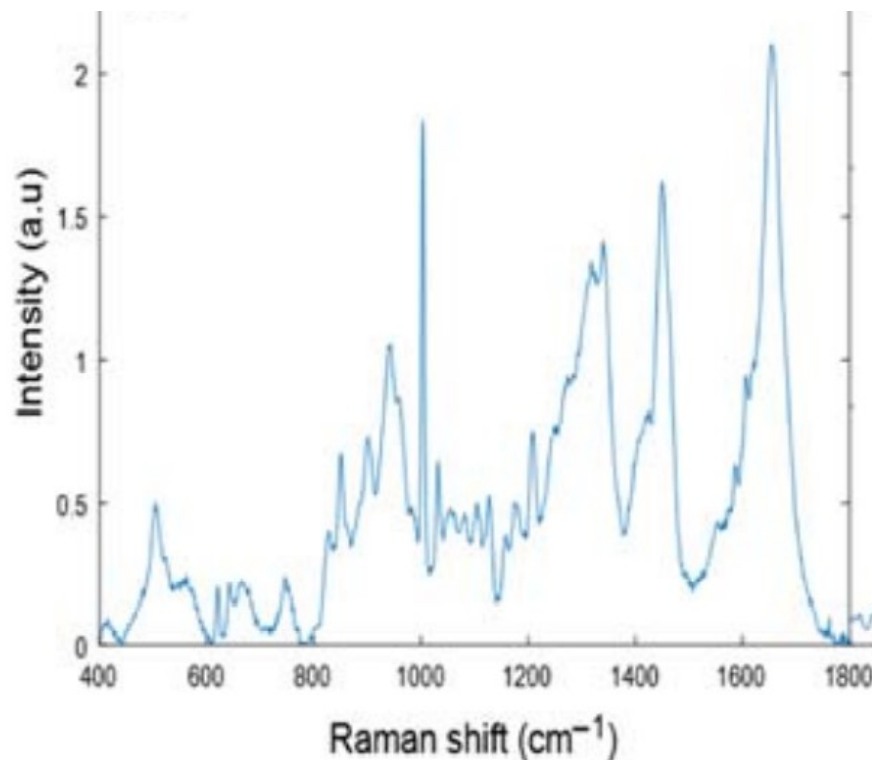


Figure 14 : Spectre Raman obtenu à partir d'un plasma humain

Le spectre Raman d'un plasma humain peut être divisé comme suit (Figure 14):

- Le signal observé à 780 cm^{-1} est associé aux vibrations d'élongations de liaison PO^{2-} .

- La présence d'un signal de forte intensité à 1004 cm^{-1} est due aux vibrations d'élongation de cycles aromatiques tels que présents dans la phénylalanine est peut donc être associé aux protéines.
- On observe également un signal à 1340 cm^{-1}
- Les bandes amides I et II se caractérisent par deux pics de forte intensité à 1640 cm^{-1} et 1580 cm^{-1} .
- La région de l'empreinte digitale rassemble également d'autres types de vibrations d'élongation et de déformation pouvant se trouver entre 600 cm^{-1} et 1500 cm^{-1} . Le pic à 1447 cm^{-1} peut être attribué aux vibrations de déformation dans le plan (cisaillement) des CH_2 et des CH_3 . Les groupements carboxylates ainsi que les déformations des liaisons CH_3 peuvent être responsables de la bande d'absorption à 1395 cm^{-1} . Les liaisons amides provoquent également une bande d'absorption de faible intensité vers 1250 cm^{-1} . Les vibrations d'élongations des liaisons C-C se caractérisent par des pics à 1108 cm^{-1} et 938 cm^{-1} .

II.3) Complémentarité de la spectroscopie IR et Raman

Toutes les molécules possèdent des signatures spectrales spécifiques en spectroscopie Raman et infrarouge selon les groupements chimiques qui les composent. Les spectroscopies infrarouge et Raman s'utilisent de manière complémentaire mais les principes physiques impliqués restent très différents. En pratique, il est observé que les molécules possèdent généralement des signatures spectrales très différentes en infrarouge et en Raman, ce qui est un plus pour affiner une caractérisation sélective de celle-ci. Il existe des règles de sélection qui caractérisent des modes de vibration moléculaire qui sont plus actifs en IR ou en Raman. Certains modes peuvent même être inactifs, donc non détectable, pour l'une ou l'autre des approches. Brièvement, les liaisons pourvues d'un fort moment dipolaire, généreront des absorptions infrarouges importantes, alors que les liaisons covalentes, plus polarisables, participeront à des modes de vibration avec une activité Raman significative. Ces deux types de spectroscopies permettent donc de révéler les niveaux énergétiques de vibration par des mécanismes différents et apportent souvent des informations complémentaires. De plus, lors de l'étude d'une molécule solubilisée dans un solvant, le choix de ce solvant se fait par rapport à la technique vibrationnelle utilisée. On essaiera a priori d'éviter l'eau pour une analyse infrarouge (trop forte absorption), et si l'eau est inévitable pour cette étude, on préférera la spectroscopie Raman, le signal Raman de l'eau étant faiblement intense.

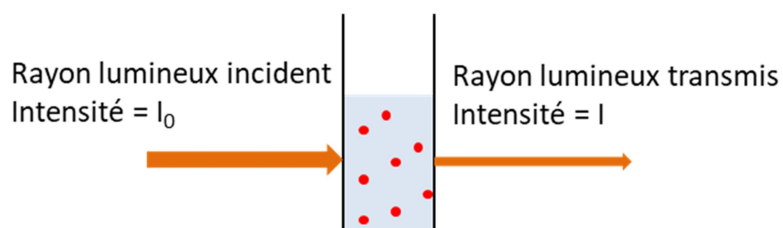
II.4) Spectroscopie infrarouge (IR)

II.4.a) Principe

La spectroscopie infrarouge (ou spectroscopie IR) est une méthode d'analyse appartenant à la catégorie des spectroscopies d'absorption. En effet, elle étudie la réaction vibrationnelle de l'échantillon après absorption de photon de longueur d'onde spécifique. La spectroscopie infrarouge utilise les rayonnements du moyen infrarouge (2,5-25 μ m) qui sont suffisamment énergétiques pour amener les liaisons à un niveau énergétique correspondant aux niveaux vibrationnels.

Ce phénomène d'absorption est mis en évidence par la mesure de l'intensité du faisceau transmis à travers l'échantillon. La transmittance est calculée par le rapport entre l'intensité du faisceau lumineux transmis et celle du faisceau lumineux incident (Figure 15). Il est ensuite possible de convertir la transmittance en absorbance selon la relation suivante :

$$A = \log(T)$$



$$\text{Transmittance (T)} = \frac{I}{I_0} \times 100$$

Figure 15 : Phénomène d'absorption et mise en évidence par le calcul de la transmittance

II.4.b) Spectroscopie infrarouge et identification

Historiquement, la spectroscopie infrarouge n'était utilisée qu'en astronomie et astrophysique (137). Il a fallu attendre les années 1900 pour qu'Abney et Festing enregistrent les spectres d'absorption infrarouge de 52 composés et établissent une corrélation entre des bandes d'absorbances avec des structures moléculaires (138). Les spectromètres infrarouges dispersifs ont été peu utilisés en biologie jusqu'au développement dans les années 1970 des spectromètres infrarouge à transformée de Fourier (FTIR) ainsi que celui des microscopes FTIR dans les années 1980 (139,140) qui ont désormais remplacés les équipements traditionnels à dispersion (141).

Le moyen infrarouge est préféré dans le domaine médical du à la capacité de la majorité des molécules biologiques à absorber entre 4000 et 400 cm^{-1} (142). La spectroscopie d'absorption infrarouge mesure l'ensemble des vibrations moléculaires (élongation et rotation) actives dans le domaine de l'infrarouge. La spécificité des bandes d'absorption permet une analyse qualitative des spectres IR par identifications des groupements fonctionnels ou par comparaison à une librairie de spectre existante.

II.4.c) Spectroscopie infrarouge et quantification

La spectroscopie d'absorption infrarouge est soumise à la loi de Beer Lambert. Intrinsèquement il existe une relation entre l'absorbance (A), la longueur du trajet

optique dans l'échantillon (l), le coefficient d'absorption molaire (ϵ) et la concentration de l'échantillon sous la forme :

$$A = \epsilon \times l \times c$$

La loi de Beer Lambert est le principe de base qui démontre une relation linéaire entre l'absorbance (A) et la concentration (C) si les autres paramètres restent constants. Notamment, la longueur du trajet optique (l) dans l'échantillon qui est le paramètre critique qui reste souvent difficile à maîtriser en spectroscopie IR. Il est encore très courant de retrouver dans la littérature scientifique que la technique est semi-quantitative (proportion des constituants dans un mélange) puisqu'une quantification au sens strict du terme demandera des conditions expérimentales parfaitement maîtrisées.

II.4.d) Modes de mesure

Les spectromètres FTIR peuvent utiliser trois modes de mesures : le mode transmission, le mode réflexion externe et le mode réflectance totale atténuée (Figure 16).

Historiquement, l'analyse infrarouge était principalement réalisée en mode transmission. Dans ce type d'analyse, le faisceau incident doit traverser l'échantillon. Cette technique permet la réalisation d'analyses sur des pastilles de type CaF_2 , ou KBr . Cette analyse est adaptée pour les composés sous forme de poudre qui peuvent être comprimés pour la fabrication de la pastille échantillon. La pastille réalisée doit être suffisamment fine pour que le faisceau traverse la totalité de l'échantillon sans phénomène de saturation du signal. De plus, cette étape de préparation est sujette à une variabilité dans la fabrication de la poudre ainsi que dans la mesure si le mélange n'est pas suffisamment homogène. En effet, plus la

pastille est épaisse, plus le faisceau va être absorbé, ce qui rend cette analyse peut utilisable pour des analyses quantitatives. A l'inverse, ce mode d'analyse est limité pour les solutions aqueuses étant donné le signal intense de l'eau (143), ce qui nécessite une étape préalable de séchage pour former un film fin. Ces problématiques rencontrées font que le mode de transmission n'est pas utilisé aujourd'hui pour des analyses de fluides biologiques.

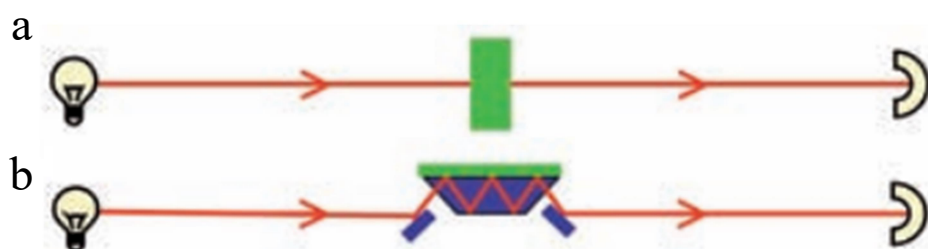


Figure 16 : Schéma représentant les différents modes d'analyse en FTIR : le mode transmission (a) et le mode réflectance totale atténuée (b)

Le mode réflectance totale atténuée (ATR) est apparu comme une alternative au mode de mesure par transmission. Dans ce mode de mesure, l'échantillon est déposé directement sur la zone de mesure appelée cristal ATR sous forme liquide ou solide. Ce mode d'analyse repose sur les mises en œuvre des conditions de réflexion totale. Le rayonnement incident arrive par le dessous de l'échantillon. Lorsqu'un rayon passe d'un milieu d'incidence de réfraction n_1 supérieur à l'indice n_2 de l'échantillon, il y a réflexion totale pour un angle incident supérieur à un angle θ tel que $\sin \theta = n_2 / n_1$. Cette condition peut être vérifiée en plaçant l'échantillon au contact d'un cristal de fort indice de réfraction (germanium, diamant ou silicone) appelé élément de réflexion interne (IRE). Malgré le nom de réflectance totale, une partie du rayonnement pénètre dans l'échantillon sur une profondeur de l'ordre du micromètre sous la forme d'une onde évanescente. L'onde évanescente subit les phénomènes d'absorption énergétique et permet d'obtenir le spectre des composés

qu'elle a traversé. Le reste du rayon est réfléchi vers le détecteur qui mesure l'intensité résiduelle.

Par opposition au mode transmission pour lequel le rayon doit traverser l'échantillon, il n'existe pas de phénomène de saturation en ATR.

Une des limites de cette approche est le nettoyage du cristal ATR pour éviter les contaminations inter-échantillons. L'utilisation du diamant est préférée pour éviter l'usure de la surface. En effet, des éraflures sur la surface de réflexion sont connues comme impactant le contact entre l'échantillon et l'IRE diminuant la pénétration de l'onde évanescente et donc la sensibilité du système (138). Par opposition à l'analyse par transmission, le mode ATR ne nécessite pas de préparation d'échantillon puisque celui-ci peut être analysé directement sous forme liquide ou solide. Ce mode permet l'analyse des échantillons liquides, pâtes et films souples. De plus, la longueur du chemin parcouru par le rayon lumineux étant relativement petite, l'absorbance due à la présence d'eau est plus faible qu'en mode transmission (144).

II.5) Spectroscopie Raman

II.5.a) Principe

Cet effet a été observé pour la première fois en 1928 par Chandrashekhara Venkata Raman et lui a valu l'obtention du prix Nobel en 1930 (146). Le phénomène physique qu'utilise la spectroscopie Raman est la diffusion de la lumière. Il peut être séparé en trois phénomènes distincts : la diffusion Rayleigh, la diffusion Raman Stokes et la diffusion Raman anti-Stokes (Figure 17).

La diffusion Rayleigh est une diffusion de type élastique. C'est-à-dire que l'énergie du photon diffusé est la même que celle du photon incident. Ce phénomène concerne la majorité des photons diffusés. En pratique, le photon entre en contact avec la molécule analysée qui va absorber l'énergie $h\nu_{\text{exc}}$ et passer d'un état stable à un état virtuel excité. Puis, elle se relaxe pour retourner à l'état initial en réémettant un photon de même énergie $h\nu_{\text{em}} = h\nu_{\text{exc}}$ (Figure 17).

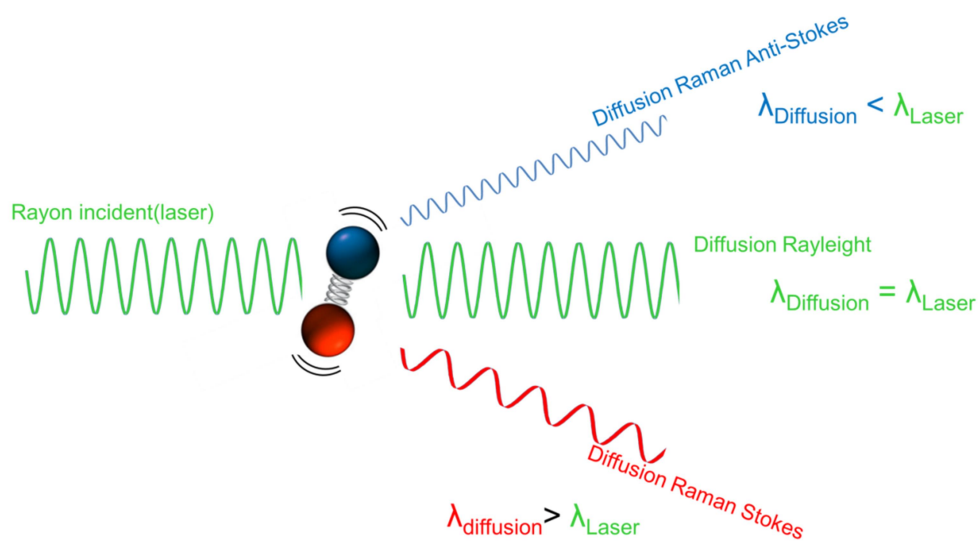


Figure 17 : Schéma représentant les phénomènes de diffusion inélastique Stokes, inélastique anti-stokes et de diffusion élastique Rayleigh (adapté de Cameron et coll (136))

La diffusion Raman est une diffusion de type inélastique basée sur l'effet Raman, c'est-à-dire le couplage des polarisabilités électroniques oscillantes de la liaison moléculaire avec le champ électromagnétique de la source (145). Brièvement, l'énergie du photon diffusé est différente de celle du photon incident. Ce phénomène concerne une minorité des photons diffusés (environ 1/1 000 000). Dans les deux cas, il existe un décalage d'énergie entre le photon émis et le photon incident. Ce

décalage peut être un gain d'énergie (décalage anti-Stokes) ou une perte d'énergie (décalage Stokes). On parlera respectivement de diffusion Raman anti-Stokes et de diffusion Raman Stokes (Figure 17).

Dans le cas de la diffusion Raman Stokes, le photon incident rentre en contact avec une molécule dans un état stable. Cette dernière absorbe l'énergie $h\nu_{\text{exc}}$ et passe dans un état excité virtuel. Ensuite, elle relaxe pour redescendre à un niveau énergétique équivalent à un niveau vibrationnel en émettant un photon d'énergie $h\nu_{\text{em}} < h\nu_{\text{exc}}$. L'énergie résiduelle $h\nu_{\text{final}}$ est ensuite éliminée par la vibration de la molécule qui revient à l'état initial stable (Figure 17). L'énergie mesurée en spectroscopie de diffusion Raman Stokes est l'énergie $h\nu_{\text{em}}$. Le signal est ensuite converti en $h\nu_{\text{final}}$, également appelée Raman shift selon l'équation suivante :

$$\Delta h\nu_{\text{final}} = h\nu_{\text{exc}} - h\nu_{\text{diff}}$$

Dans le cas de la diffusion Raman anti-Stokes, le photon incident entre en contact avec une molécule dans un état de transition vibrationnel. Cette dernière absorbe l'énergie $h\nu_{\text{exc}}$ et passe dans un état excité virtuel. Ensuite, elle relaxe pour redescendre à son niveau énergétique stable en émettant un photon d'énergie $h\nu_{\text{em}} > h\nu_{\text{exc}}$ (Figure 15). L'énergie mesurée en spectroscopie de diffusion Raman Stokes est l'énergie $h\nu_{\text{em}}$. Le signal est ensuite converti en $h\nu_{\text{final}}$, également appelée Raman shift selon l'équation suivante :

$$\Delta h\nu_{\text{final}} = h\nu_{\text{diff}} - h\nu_{\text{exc}}$$

II.5.b) Application de l'effet Raman

Cette technologie est restée relativement peu utilisée jusqu'au développement des lasers dans les années 1960 ainsi que des détecteurs à transfert de charge (« charged coupled detector » ; CCD) dans les années 1980. De plus, l'apparition des lasers monochromatiques à bande étroite a permis l'éviction des monochromateurs pour un temps d'acquisition plus court mais également un meilleur rapport signal/bruit et donc une meilleure sensibilité (147). De nos jours, la spectroscopie Raman est principalement utilisée dans des domaines tels que le contrôle de production, de matières premières, ou la caractérisation de cristaux (analyse de la structure précise, présence éventuelle de défauts). Les récentes avancées technologiques ont néanmoins permis l'utilisation des spectromètres Raman dans les domaines de la biologie ou encore de la médecine légale (148).

II.4.c) Spectromètre Raman

Un spectromètre Raman classique est composé d'une source d'excitation, d'un réseau dispersif et d'un détecteur (Figure 18). Pour les équipements dédiés à la microspectrométrie Raman, la source est montée sur un microscope.

II.5.1) Source d'excitation

L'effet Raman nécessite l'utilisation d'une lumière d'excitation polarisée monochromatique. De ce fait, l'excitation est réalisée à l'aide de lasers. Une large gamme de laser peut être utilisée comme source d'excitation en fonction de la longueur d'onde souhaitée. Classiquement, les lasers utilisés sont des lasers bleus (488 nm), vert (532 nm), rouge (633 nm) ou dans le proche-infrarouge (785 nm). En

biologie, les lasers de longueur d'onde 633 nm et 785 nm sont les plus utilisés pour réduire les phénomènes de fluorescence.

II.5.2) Détection

Le rayon diffusé est dans un premier temps focalisé par des lentilles optiques sur un filtre bande-coupante (« *notch filter* » en anglais) qui permet d'éliminer la longueur d'onde d'excitation du laser tout en laissant passer le signal de diffusion Raman. Le signal filtré est ensuite redirigé à l'aide de miroirs vers un réseau dispersif (grating en anglais). Ce réseau permet de séparer les différentes longueurs d'ondes diffusées. Les réseaux de diffraction principalement utilisés contiennent 1800 traits/mm ou 600 traits/mm. L'utilisation d'un réseau dispersif qui possède un nombre de traits/mm plus important permet d'obtenir une meilleure résolution spectrale. L'intensité des rayons diffractés est ensuite mesurée par un détecteur CCD pour afin de reconstruire le spectre (Figure 20).

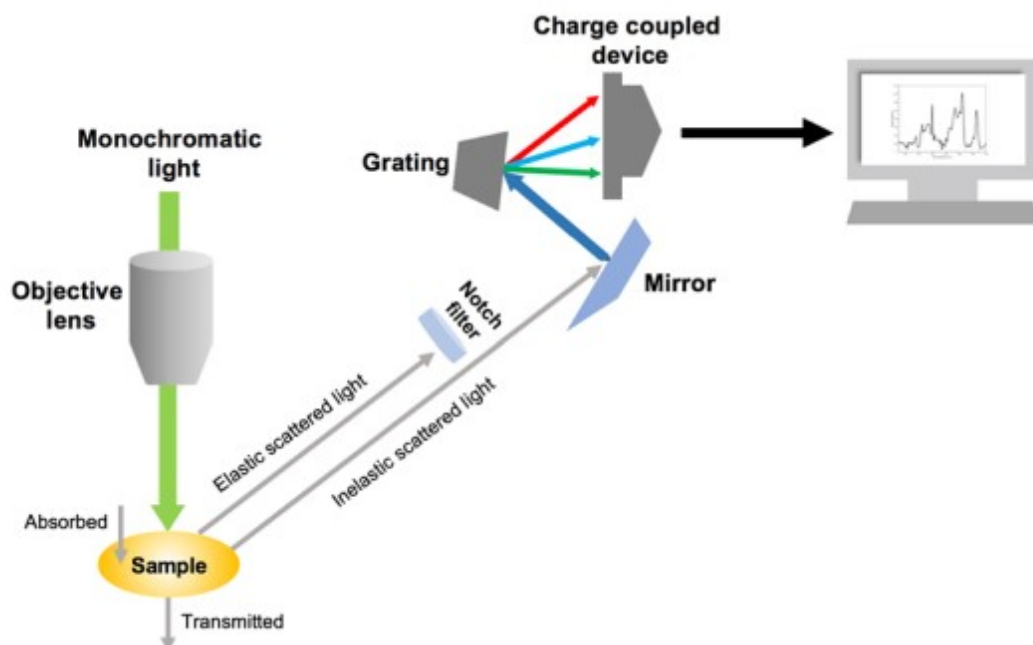


Figure 18 : Représentation schématique du trajet optique dans un spectromètre Raman (Cameron et al, 2020

(136))

RESULTATS

Partie I : Evaluation de la combinaison des méthodes conventionnelles

I) Recherche de biomarqueurs musculaires dans la SLA

La SLA est une pathologie complexe très hétérogène avec dérèglement métabolique important (80,149). Les études publiées ont permis de mettre en évidence des perturbations dans le métabolisme des acides aminés (150), des glucides (151), ainsi que des lipides (80). Dans la physiopathologie de la SLA, il est reconnu que la mort des motoneurones a pour conséquence l'atrophie des muscles qu'ils innervent. Cependant, il semble que le muscle ait également un rôle actif dans ces mécanismes de mort cellulaire et qu'il soit à l'origine de la mort des motoneurone par le phénomène de dying-back (152).

Les voies métaboliques principales sont différentes entre les différents tissus/organes, de même que les compositions biochimiques varient entre les différents fluides biologiques (Tableau III). Par exemple, le muscle squelettique a une plasticité métabolique lui permettant d'adapter son métabolisme énergétique entre le métabolisme du glucose ou des acides gras selon la disponibilité des différents substrats (153–155). Les principales voies métaboliques utilisées pour la production de l'énergie sont celle de la créatine kinase, la glycolyse et la lipolyse qui conduisent à la synthèse d'ATP grâce à la chaîne respiratoire mitochondriale. Ce phénomène implique également le métabolisme des purines. De plus, la synthèse protéique très

active au niveau du tissu musculaire implique également un métabolisme et un catabolisme des acides aminés importants. Finalement, bien que le muscle puisse être une source de biomarqueurs potentiels, celui-ci a été très peu étudié.

Quelques acteurs pathologiques identifiés comme biomarqueurs de diagnostic ou de pronostic putatifs, ou impliqués dans les mécanismes de neurodégénérescence ont cependant été identifiés dans des échantillons musculaires de patients SLA, tels que les protéines Nogo-A (150), SCD1 (140), Akt et le facteur XIIIB (151), la protéine UCP1 (152), SOD1 (153), neurotrophine 4 (154), protéines BMP (155) ou les Smads (156). Aucun de ces biomarqueurs n'a été validé à ce jour pour le diagnostic ou le pronostic.

Tableau III : Biomarqueurs identifiés dans la Sclérose Latérale Amyotrophique à partir de prélèvements sanguins ou de LCR

Biomarqueur	LCR	Sérum	Plasma	Références
Pyruvate	↑		↑	(23,156)
Ascorbate	↑			(98,156)
Glutamine	↓	↓	↓	(24,157,158)
Aspartate	↑		↑	(159,160)
Sérine	↑		↑	(24,159)
Glucose	↑			(161,162)
Créatine	↑		↑	(23,24,161)
Créatinine	↓		↓	(23,24,161)
Alpha-hydroxybutyrate	↑		↑	(23,24,161)
Cholestérol	↑		↑	(24,163,164)
Homocystéine	↑	↑	↑	(165–167)
Glutamate		↑	↑	(158,160,168)

Il nous paraît donc primordial d'avoir une vision métabolique la plus large possible avec une couverture de l'ensemble des métabolismes cités précédemment pour observer les perturbations métaboliques dues à la maladie au niveau musculaire.

II) Intérêt de la combinaison de différentes méthodes analytiques lors d'une analyse métabolomique

Une des solutions analytique permettant l'augmentation de la couverture métabolique est la combinaison des différents outils décrits précédemment. L'équipe d'Anton Ribbenstedt et coll. a montré l'intérêt de combiner la chromatographie en phase inverse (C18) et la chromatographie HILIC pour augmenter la couverture métabolique de l'analyse (Figure 19) sur deux tissus différents (169). Cette combinaison a permis d'augmenter le nombre de métabolites analysés. En effet, dans cette étude, le nombre de métabolites observés a été doublé lors de l'étude du métabolisme cérébral et multiplié par 1,5 environ lors de l'étude du métabolisme hépatique. La combinaison des différents modes d'ionisation positive et négative en chromatographie liquide couplée à la spectrométrie de masse a par exemple permis de doubler le nombre d'analytes sur les deux tissus.

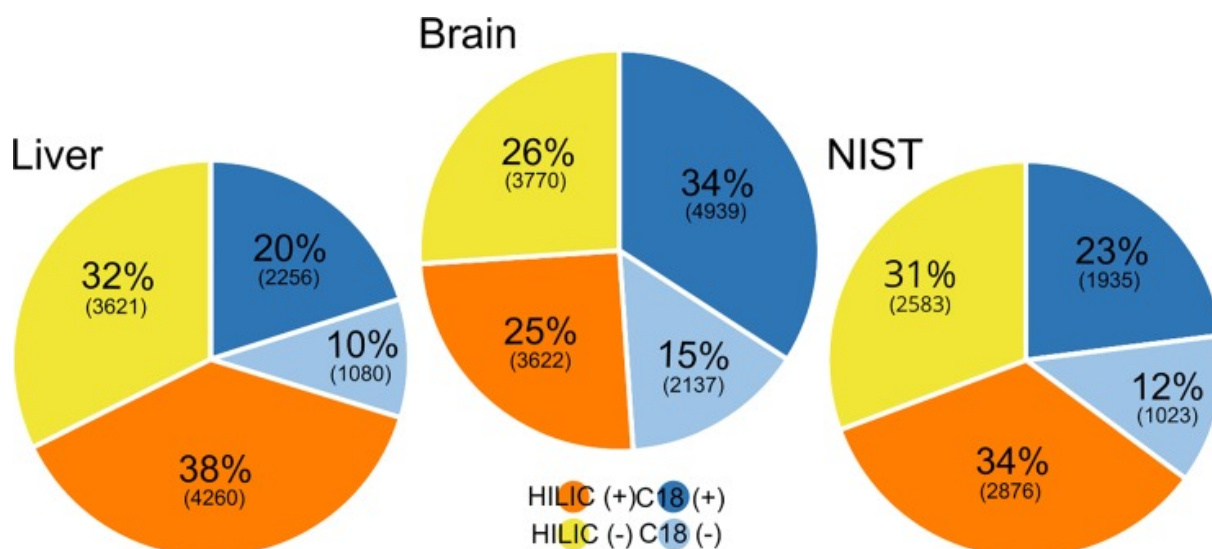


Figure 19: Impact de la séparation chromatographique et de l'ionisation sur le nombre de pics d'intérêts obtenu à partir d'un même échantillon analysé selon 4 modes différents (C18(+), C18(-), HILIC(+)) and HILIC(-)) (d'après Ribbenstedt et al (169)).

Ce résultat s'explique par les techniques analytiques mises en place. En effet, les techniques chromatographiques actuelles (phase inverse ou HILIC) ont des sélectivités différentes en termes de polarité des molécules analysées. Cette particularité technique implique de s'affranchir d'une partie des composés (les composés non retenus sur la colonne chromatographique). Pour cela, la combinaison de plusieurs méthodes séparatives couplées à la spectrométrie de masse semble conseillée. En effet, de par leur structure, la majorité des molécules biologiques ainsi que des médicaments sont analysées par chromatographie en phase inverse mais la combinaison des deux modes de séparations est susceptible de permettre à la fois une analyse des molécules hydrophiles et à la fois des molécules lipophiles. Ajoutée à cela, l'association de ces techniques avec la RMN est prometteuse car elle permet d'étendre la gamme de molécules détectées (170).

Ces résultats nous amènent à penser qu'il est primordial de combiner un maximum de différentes méthodes à disposition pour ce type d'analyse. D'autres études ont également montré l'intérêt d'utiliser la chromatographie en phase gazeuse couplée à la spectrométrie de masse en combinaison avec la RMN (171) ou à la chromatographie liquide couplée à la spectrométrie de masse (172). Les auteurs ont cependant décrit un manque dans la détection de certaines voies métaboliques et la nécessité de compléter l'analyse.

Finalement, les différentes méthodes évaluées dans ces différentes études ne permettent pas de détecter les mêmes familles de produits chimiques ni les mêmes concentrations de composés selon les spécificités de techniques analytiques utilisées. L'utilisation de l'ensemble des techniques en combinaison semble intéressante mais mène à plusieurs écueils tels que l'augmentation du coût et du temps des analyses et de la quantité d'échantillon nécessaire ainsi que les difficultés d'analyses et de traitement de l'information obtenue.

III) Article 1 : Exploration du métabolome du muscle squelettique à travers la combinaison de 4 méthodes analytiques : meilleure couverture métabolique ou argument commercial ?

Dans cette étude, nous avons analysé des échantillons musculaires de souris sauvages à l'aide de différentes méthodes analytiques : la chromatographie ultra-haute performance couplée à la spectrométrie de masse haute résolution (UPLC-HRMS), la chromatographie gazeuse couplée à la spectrométrie de masse (GC-MS), la RMN ainsi qu'une méthode en injection directe couplée à la spectrométrie de

masse en tandem (Flow injection Analysis-MSMS ; FIA-MSMS). La qualité et la fiabilité de l'analyse étant directement dépendantes de l'étape de préparation d'échantillon, nous avons étudié différents protocoles d'extraction pour chaque technologie étudiée. Pour chaque méthode, nous avons déterminé le nombre de métabolites observés ainsi que le nombre de métabolites quantifiés avec une variabilité de mesure acceptable pour ce type d'analyse (CV <30%). Le protocole permettant la meilleure couverture métabolique a ensuite été conservé pour l'analyse en combinaison.

Pour évaluer l'intérêt de combiner l'ensemble des techniques analytiques utilisées dans cette étude, nous avons évalué la spécificité et la redondance de chaque outil vis-à-vis de l'ensemble des métabolites observés.



Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

Short communication

The combination of four analytical methods to explore skeletal muscle metabolomics: Better coverage of metabolic pathways or a marketing argument?



C. Bruno^{a,b}, F. Patin^{a,b}, C. Bocca^c, L. Nadal-Desbarats^b, F. Bonnier^d, P. Reynier^c,
P. Emond^b, P. Vourc'h^{a,b}, K. Joseph-DelaFont^a, P. Corcia^{b,e,f}, C.R. Andres^{a,b}, H. Blasco^{a,b,*}

^a CHRU de Tours, Laboratoire de Biochimie et Biologie Moléculaire, Tours, France^b UMR INSERM U930, Université François Rabelais de Tours, France^c Institut MITOVASC, CNRS 6015, INSERM U1083, Université d'Angers, Angers, France^d Université François-Rabelais de Tours, Faculté de Pharmacie, EA 6295 Nanomédicaments et Nanosondes, Tours, France^e Centre de Ressources et de Compétences SLA, CHU Tours, France^f Fédération des Centres de Ressources et de Compétences de Tours et Limoges, Litorals, France

ARTICLE INFO

Article history:

Received 1 July 2017

Received in revised form 12 October 2017

Accepted 13 October 2017

Available online 18 October 2017

Keywords:

Metabolomics

Muscle

NMR

Mass spectrometry

ABSTRACT

Objectives: Metabolomics is an emerging science based on diverse high throughput methods that are rapidly evolving to improve metabolic coverage of biological fluids and tissues. Technical progress has led researchers to combine several analytical methods without reporting the impact on metabolic coverage of such a strategy. The objective of our study was to develop and validate several analytical techniques (mass spectrometry coupled to gas or liquid chromatography and nuclear magnetic resonance) for the metabolomic analysis of small muscle samples and evaluate the impact of combining methods for more exhaustive metabolite covering.

Design and methods: We evaluated the muscle metabolome from the same pool of mouse muscle samples after 2 metabolite extraction protocols. Four analytical methods were used: targeted flow injection analysis coupled with mass spectrometry (FIA-MS/MS), gas chromatography coupled with mass spectrometry (GC-MS), liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS), and nuclear magnetic resonance (NMR) analysis. We evaluated the global variability of each compound i.e., analytical (from quality controls) and extraction variability (from muscle extracts). We determined the best extraction method and we reported the common and distinct metabolites identified based on the number and identity of the compounds detected with low analytical variability (variation coefficient < 30%) for each method. Finally, we assessed the coverage of muscle metabolic pathways obtained.

Results: Methanol/chloroform/water and water/methanol were the best extraction solvent for muscle metabolome analysis by NMR and MS, respectively. We identified 38 metabolites by nuclear magnetic resonance, 37 by FIA-MS/MS, 18 by GC-MS, and 80 by LC-HRMS. The combination led us to identify a total of 132 metabolites with low variability partitioned into 58 metabolic pathways, such as amino acid, nitrogen, purine, and pyrimidine metabolism, and the citric acid cycle. This combination also showed that the contribution of GC-MS was low when used in combination with other mass spectrometry methods and nuclear magnetic resonance to explore muscle samples.

Conclusion: This study reports the validation of several analytical methods, based on nuclear magnetic resonance and several mass spectrometry methods, to explore the muscle metabolome from a small amount of tissue, comparable to that obtained during a clinical trial. The combination of several techniques may be relevant for the exploration of muscle metabolism, with acceptable analytical variability and overlap between methods. However, the difficult and time-consuming data pre-processing, processing, and statistical analysis steps do not justify systematically combining analytical methods.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Metabolomics is an emerging science that has largely evolved since its development, due to impressive progress in analytical

* Corresponding author at: Laboratoire de Biochimie et Biologie Moléculaire, Hôpital Bretonneau, Bâtiment B2A, 2 Boulevard Tonnellé, 37044 Tours, France.
E-mail address: helene.blasco@univ-tours.fr (H. Blasco).

<https://doi.org/10.1016/j.jpba.2017.10.013>

0731-7085/© 2017 Elsevier B.V. All rights reserved.

techniques and the promise of its use in the healthcare field [1–6]. This approach has been developed using a variety of analytical techniques, including mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy. To date, 40 153 metabolites have been described, identified, and referenced in the Human Metabolome DataBase (HMDB, <http://www.hmdb.ca/>), based on the use of these techniques, alone, or in combination [4,5,7]. There is a large overlap in the metabolites identified by these techniques and such experiments are costly and time-consuming. It is thus not clear whether combining several analytical techniques are generally useful or more attractive for the investigation of specific tissues, characterized by several over- and under-represented biochemical pathways. Exploration of the muscle metabolome is of utmost importance for a better understanding of metabolic and/or neuromuscular diseases, but no standard method has been described [8–12]. The objectives of our study were to develop and validate several analytical techniques (GC–MS, LC–MS, and NMR) for performing metabolomic studies on small muscle samples and evaluate their combination. We also aim to purpose an “optimized” combination of methods, especially for muscular tissue in this study.

2. Materials and methods

2.1. Sample collection

Experiments were performed on medial hamstring muscle samples collected from 10 wild type mice previously used in a closed preclinical study [13]. A total of 500 mg of muscle specimens was collected from ten mice just after sacrifice. The samples were pooled and stored in 10- and 30-mg aliquots at -80°C (Fig. 1(1)).

2.2. Sample preparation

Metabolites were extracted following two different protocols (Fig. 1), the NMR protocol and the MS protocol, for analysis by several mass spectrometry techniques, including flow injection analysis coupled with mass spectrometry (FIA–MS/MS), gas chromatography coupled with mass spectrometry (GC–MS), and ultrahigh performance liquid chromatography coupled with high-resolution mass spectrometry (LC–HRMS). Details of the extraction protocols are provided in Supplemental data. For each protocol, we systematically performed 15 different extractions protocols of muscle samples to evaluate the global variability by each technique.

2.3. NMR

A semi targeted metabolomics approach was performed by NMR. Each extract was analysed as previously described [9] on a Bruker AVANCE III 600MHz spectrometer[®] (see Supplemental data for more details). We only compared the qualitative and quantitative content of muscle extracts between the two extraction protocols, as intra-assay precision has been previously evaluated [9].

2.4. FIA–MS/MS

A semi-quantitative targeted metabolic fingerprint strategy by FIA–MS/MS was performed for the determination of amino acids and acylcarnitines using the Perkin Elmer Neobase Kit (Perkin Elmer Life and Analytical Sciences, Turku, Finland). Additional data concerning sample preparation and the MS methods are detailed in Supplemental data. Analytical variability has been previously validated for routine use following the *Comité Français d'Accréditation* (COFRAC) recommendations [14] using quality controls (QC). Thus,

we only evaluated the global variability of the experiment from the analysis of 15 muscle sample extracts.

2.5. GC–MS

We performed a semi-quantitative targeted approach by GC–MS (see details in Supplemental data). We only evaluated the global variability between 15 different muscle sample extracts, as this protocol has already been validated and published for cell extracts [15], and is used in routine practice for organic acid chromatography.

2.6. LC–HRMS

We used a recently published non-targeted approach [16] performed on a Dionex UltiMate 3000 UHPLC system (Dionex, Sunnyvale, USA) coupled to a Thermo Scientific Q Exactive HRMS (Thermo Fisher Scientific, Bremen, Germany). Additional data for this experiment are provided in Supplemental data. We assessed the analytical variability from 10 QC samples (obtained after pooling 10 μL of each extract) and global variability from the analysis of extracts from 15 muscle samples.

2.7. Building of the final dataset

We collected the area or concentration of all peaks detected for each experiment. We normalised metabolite levels using internal standards (IS) for each technique. In addition, we tested two additional normalisation methods, considering the dried muscle mass or the concentration of hydrophilic proteins, measured by Nanodrop[™] (ThermoFisher[®], Waltham, USA). Only metabolites with a coefficient of variation <30% were used to build the final dataset for further analysis [13,15,17]. We compared the methods based on the number, intensity, and variability of the area of the peaks.

2.8. Evaluation of combining analytical methods

We built a Venn diagram (Venny 2.1) to represent common and distinct metabolites identified using the various analytical techniques. The identification of corresponding metabolic pathways was performed using MetaboAnalyst (<http://www.metaboanalyst.ca/>).

3. Results

3.1. NMR

We identified 38 metabolites by NMR (Table A, Annexe A). Extracts from protocol NMR.b contained several contaminants, and several metabolites were extracted only using protocol NMR.a (not shown). There was no difference in peak intensity between both methods (not shown). Thus, we considered protocol NMR.a to be superior.

3.2. FIA–MS/MS

We semi-quantified 42 metabolites, regardless of the method of extraction used, but the relative abundance of several types varied between the two methods (Fig. 2). More metabolites had a CV < 30% when protocol MS.a was used than protocol MS.b (Supplemental Table A). Normalisation against the amount of water-soluble proteins enabled the analysis of a total of 37 metabolites with acceptable CV ($15.5 \pm 6.1\%$) with MS.a (Table A, Supplemental table A).

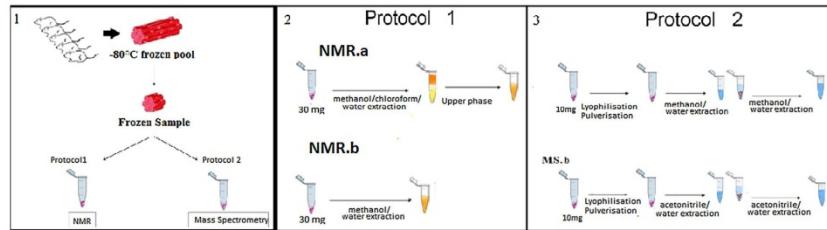


Fig. 1. Schematic overview of protocols used to extract metabolites for NMR (2) and Mass Spectrometry (3) analysis. We splitted our muscle sample into two parts. One part was used for NMR analysis (Protocol 1) and the other part was used for the different methods of mass analysis (Protocol 2). Then we tested two solvents for each method to compare and evaluate these methods.

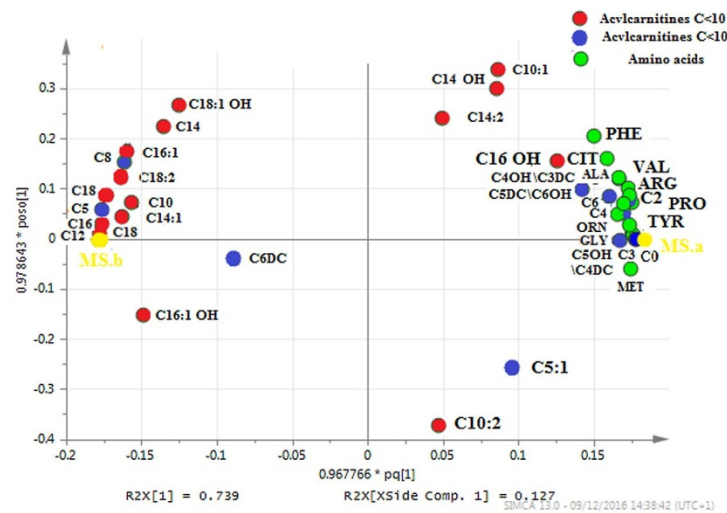


Fig. 2. Plot representing the contribution of each metabolite detected by FIA-MS after the two extraction methods. Amino acids are represented in green, acylcarnitines with a carbon chain <10 in blue, and acylcarnitines with a carbon chain >10 in red. Negative X Value represents contribution to acetonitrile/water profile whereas positive X value represent contribution to methanol/water profile. Acetonitrile/water is more appropriate to analyse long-chain acylcarnitines whereas methanol/water is more suitable to analyse aminoacids and short-chain acylcarnitines. (C0: Carnitine; C2: Acetylcarnitine; C3: Propionylcarnitine; C4: Butyrylcarnitine; C4OH/C3DC: Malonylcarnitine; C5: Isovalerylcarnitine; C5OH/C4DC: Methylmalonylcarnitine; C5DC/C6OH: Glutaryl carnitine; C6: Hexanoylcarnitine; C6DC: C8: Octanoylcarnitine; C10: Decanoylcarnitine; C10:1: Decenoylcarnitine; C12: Dodecanoylcarnitine; C14: Tetradecanoylcarnitine; C14OH: Hydroxytetradecanoylcarnitine; C14:1: Tetradecenoylcarnitine; C14:2: Tetradecadienylcarnitine; C16: hexadecanoylcarnitine; C16:1: Hexadecenoylcarnitine; C16OH: Hydroxyhexadecanoylcarnitine; C16:1 OH: Hydroxyhexadecanoylcarnitine; C18: Octadecanoylcarnitine; C18:2: Octadecadienylcarnitine; C18:1 OH: Hydroxyoctadecanoylcarnitine; ALA: Alanine; ARG: Arginine; GLY: Glycine; MET: Methionine; ORN: Ornithine; PHE: Phenylalanine; PRO: Proline; TYR: Tyrosine; VAL: Valine). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

Table representing the final characteristics of the various analytical methods used. The parameters tested and results are summarised for each technique (sample weight, extraction solvent, extract volume, normalization).

	NMR	FIA-MS/MS	GC-MS	LC-HRMS
Sample Weight	30 mg	10 mg		
Extraction Solvent	water/methanol/chloroform	water/methanol		
Extract Volume		150 µL	700 µL	500 µL
Normalization	TSP	water-soluble proteins		
Metabolites detected (CV < 30%)	38	42 (37)	76 (18)	120 (80)

3.3. GC-MS

We identified approximately 70 metabolites using protocol MS.a or MS.b (Supplemental data, Annexe A). MS.a provided better performance (19 metabolites vs 7 with CV < 30%). There were no

major differences between the two extraction methods (Supplemental Table B). The final conditions are presented in Table 1.

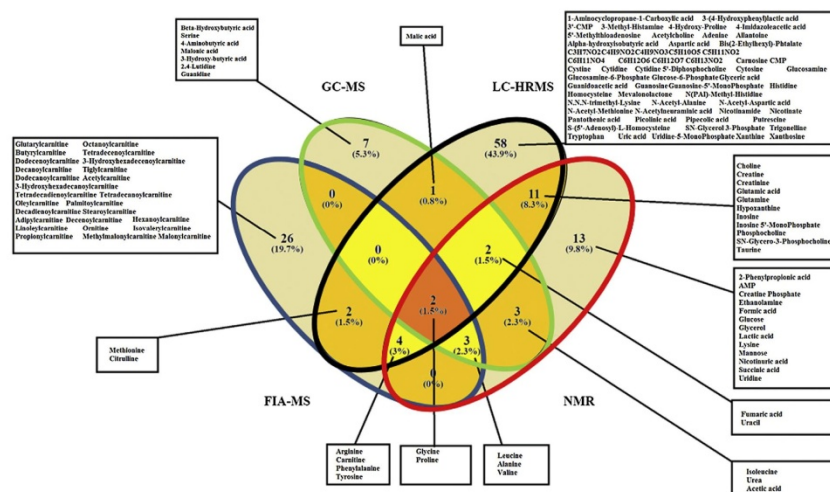


Fig. 3. Venn diagram representing specificity and overlap of metabolites reliably detected by each analytical method. Red, blue, green and black circles represent metabolites analysed respectively by NMR, FIA-MS, GC-MS and LC-HRMS. Crossed zone represents overlaps between methods. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

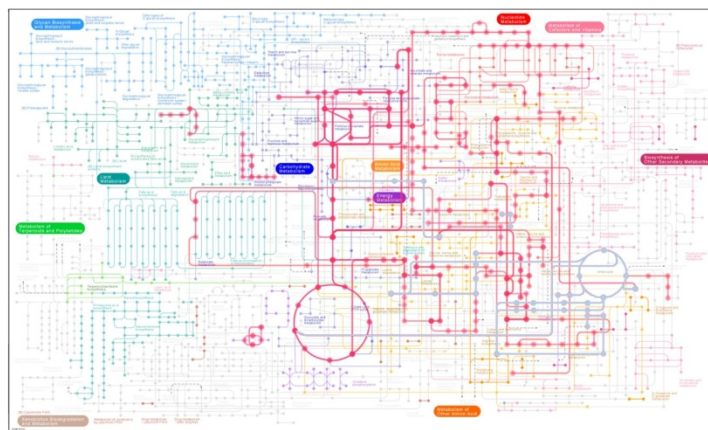


Fig. 4. Schematic mapping of Human metabolome. Each metabolite is represented by a circle. Pathways composed by at least one metabolite detected by our protocol are underlined in pink. Arginine and proline; alanine, aspartate, and glutamate; and glycine, serine, and threonine metabolism are underlined in grey (pathways with high impact and high coverage). There was good coverage of energy, amino acid, and nucleotide metabolism. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4. LC-HRMS

We identified 61 metabolites in positive and 19 in negative ionization (Supplemental data A). We used the MS.a extraction protocol for UHPLC-HRMS as it performed better than the MS.b protocol in the FIA-MS/MS and GC-MS experiments. There was low variability (3.5%/[− 7.1%]) and no difference between normalisation methods (Table A, Supplemental Table C).

3.5. Data integration and metabolic pathway analysis

Combining all the data (Table 1), we analysed 38 metabolites by NMR, 37 by FIA-MS/MS, 18 by GC-MS, and 80 by LC-HRMS (Fig. 3). Almost 43.9% of the metabolites were identified only by LC-HRMS and 5% were specifically identified by GC-MS. Altogether, these metabolites cover 58 metabolic pathways (Fig. 4, Supplementary data B). Three high impact amino acids metabolic pathways were highly covered: arginine and proline; alanine, aspartate, and glutamate; and glycine, serine, and threonine metabolism.

Table 2

Review of the previously published analytical methods used to study muscle and the metabolic pathways observed and a comparison with our analytical methods [9,19–23].

	Palma et al. [19]	Sato et al. [20]	Petucci et al. [21]	Karlsson et al. [22]	Beaucercq et al. [9]	Le et al. [23]	our study
Instrument	NMR	CE-MS	IC-MS	not known	NMR	MS	LC-HRMS/NMR/FIA-MS
Alanine, aspartate and glutamate metabolism		x	x		x		x
Aminoacyl-tRNA biosynthesis	x	x					x
Ammonia recycling					x		
Arginine and Proline metabolism					x	x	x
beta-alanine metabolism					x		x
Butanoate metabolism		x	x			x	x
Creatine metabolism					x	x	x
Citrate cycle (TCA cycle)	x	x	x	x	x	x	x
D-Glutamine and D-glutamate metabolism		x	x			x	x
Fructose and mannose degradation					x		
Galactose metabolism					x		
Glutathione metabolism	x	x			x	x	x
Glycine, serine and threonine metabolism					x		x
Glycolysis or Gluconeogenesis	x	x		x	x	x	x
Glyoxylate and dicarboxylate metabolism		x					x
Histidine metabolism					x	x	x
Lysine biosynthesis			x				x
Methane metabolism	x	x					x
Methionine metabolism					x		x
Nicotinate metabolism						x	x
Nitrogen metabolism	x	x					x
Nucleotide sugars metabolism					x		
Pentose phosphate pathway					x		x
Phenylalanine metabolism	x	x	x				x
Phenylalanine, tyrosine and tryptophan biosynthesis	x				x		x
Propanoate metabolism	x						x
Protein biosynthesis					x		
Purine metabolism	x	x			x	x	x
Pyrimidine metabolism							x
Pyruvate metabolism	x	x					x
Thiamine metabolism	x						x
Urea cycle					x	x	
Valine, leucine and isoleucine biosynthesis/degradation	x	x				x	x

4. Discussion

4.1. Each analytical method performs well on a small muscle sample

Each analytical method provided high quality data from small muscle samples (10 or 30 mg), smaller than those used in other published studies [8,9,12]. Furthermore, we only focused on metabolites which had small variability, after redundancy deletion, and with reliable identification, so we detected a limited number of metabolites but with high confidence. In this study, there was overlap for some identified metabolites between methods, but others were more method-specific. As previously described [9], NMR provided good coverage of amino acid metabolism and the TCA cycle, and detected 13 metabolites not detected by any of the MS-

based methods. It is difficult to compare our results for the MS methods with those reported in the literature due to the different extraction and separation methods used and the lack of published variability data [8,9,11,12,15] (Table 2). The analytical conditions used for LC-HRMS were not suitable for the separation of hexoses, but this method enabled the detection of metabolites of interest in muscle, such as several polar and semi-polar compounds (such as cystine, hydroxy-proline, and acetylcholine, not detected using other methods). FIA-MS/MS provided data concerning acylcarnitines, which are very important in muscle energetic metabolism [18]. GC-MS, following the common metabolite extraction procedure, contributed little to the overall metabolic profile when combined with the data generated using the other techniques. However metabolites identification depends on the libraries used as references. For GC-MS, LC-HRMS and NMR, we used respectively

NIST libraries, MSMLS and Chemomx Reference Libraries. FIA/MS was targeted on 42 known metabolites (aminoacids and acylcarnitines).

4.2. The multimodal approach is feasible for the exploration of small muscle samples

Few studies have described the combined use of NMR spectroscopy and mass spectrometry methods to analyse the muscle metabolome [24–27]. Targeted strategies, focused on a few dozen metabolites, were performed in most of these studies [24,26,27], and all but one used tissue samples of 100 mg or more [26]. Samples of limited size, such as human biopsies performed for diagnosis or clinical trials purposes, have not been previously analysed using a multimodal metabolomic approach. Several issues, such as the overlap of identified metabolites across techniques and the management of big data, remain significant challenges facing such research, even if data fusion strategies are being developed to improve the combined multivariate analysis of these data [28,29]. Our study shows the feasibility of combining several different analytical methods, but GC–MS may not be an ideal tool for targeted metabolomics [15]. However, this technic can be useful for untargeted metabolomics [30]. Furthermore, it would be interesting to pursue our evaluation of the interest of combined analytical methods for metabolomics toward other “omics”. Transcriptomics and proteomics could be used to confirm the involvement of pathways identified by metabolomics or to discover other pathways due to the complementary of all these methods [31].

4.3. The multimodal approach enables the exploration of the main metabolic pathways in muscle

We were able to combine several analytical methods to analyse numerous metabolic pathways including amino acids, the TCA cycle, nitrogen metabolism, energy metabolism, purine metabolism, and acylcarnitines (Supplementary data A). These experiments provided a global overview of skeletal muscle metabolism. Indeed, these highly metabolic organs require energy derived from carbohydrates by aerobic (mitochondrial fatty acids oxidation and TCA cycle) or anaerobic degradation (cytoplasmic glycolysis), amino acid degradation, and lipolysis. The pivotal role of ATP and phosphocreatine as energy substrates in anaerobic fibres also highlight the importance of amino acid and purine metabolism [32]. Skeletal muscle is highly plastic and can adapt in response to functional demands [33–36]. This process involves macroscopic organisation, as well as metabolic modifications [37]. Our strategy is adapted to detect metabolic adaptations to physiological or pathological conditions [18,38–41]. Importantly, metabolism pathways known to be affected in metabolic diseases such as Pump disease, Lactate dehydrogenase deficiency or also Carnitine deficiency, may be indirectly detected by this method (respectively anaerobic glycolysis, aerobic glycolysis and acylcarnitines). This approach, based on the use of small muscle samples, is also compatible with clinical experiments.

5. Conclusion

We present, for the first time, the development of a multimodal approach to analyse the muscle metabolome from a small muscle sample. The combination of several different techniques may be scientifically justified, but must be systematically evaluated to avoid its being driven strictly by commercial interests. In our study it allowed us to explore a total of 132 metabolites with very few metabolites detected by all techniques. The analytical approach that we have developed is compatible with human experiments, especially clinical trials. Other tools, based on various

physical properties, may be used to perform such exploratory studies, such as Fourier-transformation infra-red (FT-IR) and Raman spectroscopy [42–46]. These methods are sufficiently powerful to discriminate different metabolic profiles and could be used in addition to the standard tools of metabolomic analysis.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jpba.2017.10.013>.

References

- [1] L. Pauling, A.B. Robinson, R. Teranishi, P. Cary, Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography, *Proc. Natl. Acad. Sci.* 68 (1971) 2374–2376.
- [2] B. Dieme, S. Mavel, H. Blasco, G. Tripi, F. Bonnet-Brihault, J. Malvy, C. Bocca, C.R. Andres, L. Nadal-Desbarats, P. Emond, Metabolomics study of urine in Autism spectrum disorders using a multiplatform analytical methodology, *J. Proteome Res.* 14 (2015) 5273–5282.
- [3] S. Jäger, D.Y. Kim, J.F. Hultquist, K. Shindo, R.S. LaRue, E. Kwon, M. Li, B.D. Anderson, L. Yen, D. Stanley, Vif hijacks CBF- β to degrade APOBEC3G and promote HIV-1 infection, *Nature* 481 (2012) 371–375.
- [4] J. Sun, J. Jin, R.D. Beger, C.E. Cerniglia, M. Yang, H. Chen, Metabolomics evaluation of the impact of smokeless tobacco exposure on the oral bacterium *Capnocytophaga sputigena*, *Toxicol. In Vitro* 36 (2016) 133–141.
- [5] A.I. Pardo-García, K.L. Wilkinson, J.A. Culbert, N.D. Lloyd, G.L. Alonso, M.R. Salinas, Accumulation of guaiacol glycoconjugates in fruit, leaves and shoots of *Vitis vinifera* cv. Monastrell following foliar applications of guaiacol or oak extract to grapevines, *Food Chem.* 217 (2017) 782–789.
- [6] L. Pretali, L. Bernardo, T.S. Butterfield, M. Trevisan, L. Lucini, Botanical and biological pesticides elicit a similar Induced Systemic Response in tomato (*Solanum lycopersicum*) secondary metabolism, *Phytochemistry* 130 (2016) 56–63.
- [7] M.O. Sarfaraz, R.P. Myers, C.S. Coffin, Z.-H. Gao, A.A.M. Shaheen, P.M. Crotty, P. Zhang, H.J. Vogel, A.M. Weljie, A quantitative metabolomics profiling approach for the noninvasive assessment of liver histology in patients with chronic hepatitis C, *Clin. Transl. Med.* 5 (2016) 33.
- [8] M.S. Dotzert, M.R. Murray, M.W. McDonald, T.D. Oliver, T.J. Velenosi, A. Hennop, E.G. Noble, B.L. Urquhart, C.J. Melling, Metabolomic response of skeletal muscle to aerobic exercise training in insulin resistant type 1 diabetic rats, *Sci. Rep.* 6 (2016).
- [9] S. Beaulercq, L. Nadal-Desbarats, C. Hennequet-Antier, A. Collin, S. Tesserand, M. Bourin, E. Le Bihan-Duval, C. Berri, Serum and muscle metabolomics for the prediction of ultimate pH, a key factor for chicken-meat quality, *J. Proteome Res.* 15 (2016) 1168–1178.
- [10] Y. Hatazawa, N. Senoo, M. Tadaishi, Y. Ogawa, O. Ezaki, Y. Kamei, S. Miura, Metabolomic analysis of the skeletal muscle of mice overexpressing PGC-1 α , *PLoS One* 10 (2015) e0129084.
- [11] J. Welzenbach, C. Neuhoff, C. Looft, K. Schellander, E. Tholen, C. Große-Brinkhaus, Different statistical approaches to investigate porcine muscle metabolome profiles to highlight new biomarkers for pork quality assessment, *PLoS One* 11 (2016) e0149758.
- [12] J.A. Carrillo, Y. He, Y. Li, J. Liu, R.A. Erdman, T.S. Sonstegard, J. Song, Integrated metabolomic and transcriptome analyses reveal finishing forage affects metabolic pathways related to beef quality and animal welfare, *Sci. Rep.* 6 (2016).
- [13] F. Patin, T. Baranek, P. Vourc'h, L. Nadal-Desbarats, J.-F. Goossens, S. Marouillat, A.-F. Dessein, A. Descat, B.M. Hounoum, C. Bruno, Combined metabolomics and transcriptomics approaches to assess the IL-6 blockade as a therapeutic of ALS: deleterious alteration of lipid metabolism, *Neurotherapeutics* 13 (2016) 905–917.
- [14] I. Ich, Q2 (R1): Validation of Analytical Procedures: Text and Methodology, 2005.
- [15] C. Veyrat-Durebex, P. Corcia, E. Piver, D. Devos, A. Dangoumau, F. Gouel, P. Vourc'h, P. Emond, F. Laumonnier, L. Nadal-Desbarats, Disruption of TCA cycle and glutamate metabolism identified by metabolomics in an in vitro model of amyotrophic lateral sclerosis, *Mol. Neurobiol.* 53 (2016) 6910–6924.
- [16] J. Kouassi Nzoughet, C. Bocca, G. Simard, D. Prunier-Mirebeau, J.M. Chao de la Barca, D. Bonneau, V. Procaccio, F. Prunier, G. Lenaers, P. Reynier, A nontargeted UHPLC-HRMS metabolomics pipeline for metabolite identification: application to cardiac remote ischemic preconditioning, *Anal. Chem.* 89 (2017) 2138–2146.
- [17] H. Blasco, C. Veyrat-Durebex, M. Bertrand, F. Patin, F. Labarthe, H. Henique, P. Emond, C. Andres, C. Antar, C. Landon, A Multiplatform Metabolomics Approach to Characterize Plasma Levels of Phenylalanine and Tyrosine in Phenylketonuria, 2016.
- [18] P.P. Devarshi, S.M. McNabney, T.M. Henagan, Skeletal muscle nucleo-Mitochondrial crosstalk in obesity and type 2 diabetes, *Int. J. Mol. Sci.* 18 (2017) 831.
- [19] M. Palma, T. Scanlon, T. Kilminster, J. Milton, C. Oldham, J. Greeff, M. Matzapetakis, A.M. Almeida, The hepatic and skeletal muscle ovine

- metabolomes as affected by weight loss: a study in three sheep breeds using NMR-metabolomics, *Sci. Rep.* 6 (2016).
- [20] E. Sato, T. Mori, E. Mishima, A. Suzuki, S. Sugawara, N. Kurasawa, D. Saigusa, D. Miura, T. Morikawa-Ichinose, R. Saito, Metabolic alterations by indoxyl sulfate in skeletal muscle induce uremic sarcopenia in chronic kidney disease, *Sci. Rep.* 6 (2016).
 - [21] C. Petucci, A. Zelenin, J.A. Culver, M. Gabriel, K. Kirkbride, T.T. Christison, S.J. Gardell, Use of ion chromatography/mass spectrometry for targeted metabolite profiling of polar organic acids, *Anal. Chem.* 88 (2016) 11799–11803.
 - [22] M. Karlsson, J.K. Ehinger, S. Piel, F. Sjövall, J. Henriksnäs, U. Höglund, M.J. Hansson, E. Elmer, Changes in energy metabolism due to acute rotenone-induced mitochondrial complex I dysfunction—An in vivo large animal model, *Mitochondrion* 31 (2016) 56–62.
 - [23] L. Le, B. Jiang, W. Wan, W. Zhai, L. Xu, K. Hu, P. Xiao, Metabolomics reveals the protective of Dihydromyricetin on glucose homeostasis by enhancing insulin sensitivity, *Sci. Rep.* 6 (2016).
 - [24] P.P. Dzeja, K. Hoyer, R. Tian, S. Zhang, E. Nemutlu, M. Spindler, J.S. Ingwall, Rearrangement of energetic and substrate utilization networks compensate for chronic myocardial creatine kinase deficiency, *J. Physiol.* 589 (2011) 5193–5211.
 - [25] R.D. Warner, R.H. Jacob, K. Rosenfold, S. Rochfort, C. Trener, T. Plozza, M.B. McDonagh, Altered post-mortem metabolism identified in very fast chilled lamb M. longissimus thoracis et lumborum using metabolomic analysis, *Meat Sci.* 108 (2015) 155–164.
 - [26] H.J. Atherton, N.J. Bailey, W. Zhang, J. Taylor, H. Major, J. Shockor, K. Clarke, J.L. Griffin, A combined 1H-NMR spectroscopy-and mass spectrometry-based metabolomic study of the PPAR- α null mutant mouse defines profound systemic changes in metabolism linked to the metabolic syndrome, *Physiol. Genomics* 27 (2006) 178–186.
 - [27] H.J. Atherton, M.K. Gulston, N.J. Bailey, K. Cheng, W. Zhang, K. Clarke, J.L. Griffin, Metabolomics of the interaction between PPAR- α and age in the PPAR- α -null mouse, *Mol. Syst. Biol.* 5 (2009) 259.
 - [28] E. Acar, A.J. Lawaetz, M.A. Rasmussen, R. Bro, Structure-revealing data fusion model with applications in metabolomics, in: *IEEE*, 2013, pp. 6023–6026.
 - [29] E. Acar, E.E. Papalexakis, G. Gürdeniz, M.A. Rasmussen, A.J. Lawaetz, M. Nilsson, R. Bro, Structure-revealing data fusion, *BMC Bioinf.* 15 (2014) 239.
 - [30] B. Diémé, A. Lefèvre, L. Nadal-Desbarats, L. Galineau, B.M. Hounoum, F. Montigny, H. Blasco, C.R. Andres, P. Emond, S. Mavel, Workflow methodology for rat brain metabolome exploration using NMR, LC-MS and GC-MS analytical platforms, *J. Pharm. Biomed. Anal.* 142 (2017) 270–278.
 - [31] K. Yugi, H. Kubota, A. Hatano, S. Kuroda, Trans-omics: how to reconstruct biochemical networks across multiple omiclayers, *Trends Biotechnol.* 34 (2016) 276–290.
 - [32] H. Westerblad, J.D. Bruton, A. Katz, Skeletal muscle: energy metabolism, fiber types, fatigue and adaptability, *Exp. Cell Res.* 316 (2010) 3093–3099.
 - [33] B.M. Doucet, A. Lam, L. Griffin, Neuromuscular electrical stimulation for skeletal muscle function, *Yale J. Biol. Med.* 85 (2012) 201–215.
 - [34] M. Maffei, E. Longa, R. Qaisar, V. Agoni, J. Desaphy, D.C. Camerino, R. Bottinelli, M. Canepari, Actin sliding velocity on pure myosin isoforms from hindlimb unloaded mice, *Acta Physiol.* 212 (2014) 316–329.
 - [35] J. Zhang, A.R. Light, C.L. Hoppel, C. Campbell, C.J. Chandler, D.J. Burnett, E.C. Souza, G.A. Casazza, R.W. Hughen, N.L. Keim, Acylcarnitines as markers of exercise-associated fuel partitioning, xenometabolism, and potential signals to muscle afferent neurons, *Exp. Physiol.* 102 (1) (2016) 48–69.
 - [36] A. Cristea, R. Qaisar, P.K. Edlund, J. Lindblad, E. Bengtsson, L. Larsson, Effects of aging and gender on the spatial organization of nuclei in single human skeletal muscle cells, *Aging Cell* 9 (2010) 685–697.
 - [37] D.P. Enot, M. Niso-Santano, S. Durand, A. Chery, F. Pietrocola, E. Vacchelli, F. Madeo, L. Galluzzi, G. Kroemer, Metabolomic analyses reveal that anti-aging metabolites are depleted by palmitate but increased by oleate in vivo, *ABV Cell Cycle* 14 (2015) 2399–2407.
 - [38] W. Dott, J. Wright, K. Cain, P. Mistry, K.E. Herbert, Integrated metabolic models for xenobiotic induced mitochondrial toxicity in skeletal muscle, *Redox Biol.* 14 (2018) 198–210.
 - [39] K. Yoshida, C.K. Imamura, K. Hara, M. Mochizuki, Y. Tanigawara, Effect of everolimus on the glucose metabolic pathway in mouse skeletal muscle cells (C2C12), *Metabolomics* 13 (2017) 98.
 - [40] L.S. Quinn, B.G. Anderson, Interleukin-15, IL-15 receptor- α , and obesity: concordance of laboratory animal and human genetic studies, *J. Obes.* (2011) 2011.
 - [41] T. Gamberi, T. Fiaschi, E. Valocchia, A. Modesti, P. Mantuano, J.-F. Rolland, F. Sanarica, A. De Luca, F. Magherini, Proteome analysis in dystrophic mdx mouse muscle reveals a drastic alteration of key metabolic and contractile proteins after chronic exercise and the potential modulation by anti-oxidant compounds, *J. Proteomics* (2017) (In press, Corrected proofs).
 - [42] A.-L. Chateigner-Boutin, J.J. Ordaz-Ortiz, C. Alvarado, B. Bouchet, S. Durand, Y. Verherbruggen, Y. Barrière, L. Saulnier, Developing pericarp of maize: a model to study arabinoxylan synthesis and feruloylation, *Front. Plant Sci.* 7 (2016).
 - [43] Q. Zhao, T. Yin, J. Peng, Y. Zou, J. Yang, A. Shen, J. Hu, Noninvasive metabolomic profiling of human embryo culture media using a simple spectroscopy adjunct to morphology for embryo assessment in in vitro fertilization (IVF), *Int. J. Mol. Sci.* 14 (2013) 6556–6570.
 - [44] Z. Farhane, F. Bonnier, A. Casey, A. Maguire, L. O'Neill, H.J. Byrne, Cellular discrimination using in vitro Raman micro spectroscopy: the role of the nucleolus, *Analyst* 140 (2015) 5908–5919.
 - [45] R. Buckmaster, F. Asphahani, M. Thein, J. Xu, M. Zhang, Detection of drug-induced cellular changes using confocal Raman spectroscopy on patterned single-cell biosensors, *Analyst* 134 (2009) 1440–1446.
 - [46] P.T. Winnard Jr., C. Zhang, F. Vesuna, J.W. Kang, J. Garry, R.R. Dasari, I. Barman, V. Raman, Organ-specific isogenic metastatic breast cancer cell lines exhibit distinct Raman spectral signatures and metabolomes, *Oncotarget* 8 (12) (2017) 20266.

Discussion et perspectives

Une bonne couverture métabolique

Notre étude nous a permis de proposer une préparation d'échantillon qui permettait une bonne extraction des acides aminés, des purines, des aminoacyls-t-RNA, des intermédiaires du cycle de Krebs et des métabolites impliqués dans la réponse au stress oxydatif. Les résultats obtenus en FIA-MSMS nous ont par ailleurs confirmé la bonne reproductibilité de l'analyse. En effet, les concentrations mesurées sont très satisfaisantes pour 37 métabolites sur les 42 présents dans la méthode (CV = 15.5% +/- 6.1%). Cette étude a permis d'observer un total de 132 métabolites répartis entre 58 voies métaboliques. Sur les 132 métabolites, 80 sont détectés par l'HPLC-HRMS, 34 par la FIA-MSMS, 38 par la RMN et 18 par la GC-MS. Au vu de la redondance des métabolites détectés par chaque technique lors de cette étude et du faible

nombre de métabolites spécifiquement détectés par la GC-MS (7 métabolites), son intérêt a été remis en question pour l'analyse du métabolisme musculaire. Malgré cela, nous avons pu mettre en évidence l'importance de combiner différents outils. Cela confirme les résultats de Ribbenstedt et coll bien que, dans cette étude, l'absence d'identification des métabolites ne permettait pas de juger de la redondance entre les différentes analyses (169).

Le second point majeur relevé est la bonne couverture de l'ensemble du métabolisme énergétique via notamment le métabolisme des acides aminés, le métabolisme glucidique, le métabolisme des purines, synthèse des aminoacyls-t-RNA, le cycle de Krebs mais aussi le métabolisme oxydatif. Les voies métaboliques observées correspondent aux voies métaboliques retrouvées perturbées dans le plasma, ce qui nous conforte dans la pertinence du protocole analytique proposé (149) dans le contexte de la SLA.

Un manque de couverture du métabolisme lipidique

L'ensemble des méthodes d'analyse utilisées ne nous ont pas permis d'accéder aux informations concernant le métabolisme lipidique et phospholipidique. Or des dérèglements du métabolisme lipidique ont été retrouvés dans le liquide céphalo-rachidien (80) ainsi qu'au niveau plasmatique (173). En effet, la stéatose hépatique rapportée chez 16% des patients SLA suggère une perturbation globale du métabolisme lipidique (164). De plus, Certains auteurs ont rapporté une hyperlipidémie (cholestérol total, LDL) chez des patients SLA (164,174), l'association entre l'hyperlipidémie et la survie reste controversée (175–179). Une étude a rapporté qu'un rapport LDL / HDL élevé était associé à une meilleure survie (164),

mais une autre étude a trouvé une relation entre la diminution des lipides sanguins ou le rapport LDL / HDL et une insuffisance respiratoire (175). En revanche, certaines études n'ont pas confirmé cette association entre l'hyperlipidémie et la survie ou la progression de la maladie (174,180–182). Une étude récente a également montré une signature lipidique différente entre les patients SLA et un groupe contrôle au niveau du LCR (80). De plus d'autres lipides comme le TG (16:0/16:0/18:1) et le TG (18:0/16:0/18:1) étaient associés à une meilleure survie pour des concentrations plus basses.

Vers une amélioration de la préparation d'échantillon

La qualité et la fiabilité des données analytiques recueillies lors d'une analyse dépendent fortement de la préparation d'échantillon. Des études ont montré que la préparation d'échantillon pouvait modifier la composition de l'échantillon par des réactions chimiques (hydrolyse, oxydation...) (183) ou n'extraire qu'une partie des métabolites d'intérêt. Par exemple, une étude publiée récemment combinant une analyse métabolomique et une analyse lipidomique a dû utiliser deux protocoles d'extraction différents (184). En effet, l'analyse lipidomique nécessitait une extraction par le méthanol suivi d'une étape d'estérification alors que l'extrait pour l'analyse métabolomique était préparé à partir d'un mélange composé à 50% de méthanol et 50% d'acétonitrile sans étape de dérivation. De plus, dans cette étude, l'analyse métabolomique n'a pas permis d'observer de lipides, de même que l'analyse lipidomique n'a pas permis d'observer de métabolites polaires. Une seconde étude a également montré l'impact de la préparation d'échantillon sur les métabolites détectés ainsi que sur la variabilité de l'analyse (185). Les auteurs ont notamment

observé que les oxystérols, glycerophosphatidylcholines et les sphingomyélines sont significativement mieux extraits par un solvant éthanol/dichlorométhane alors que d'autres classes plus polaires comme les acides aminés, les amines biogènes et les acylcarnitines sont significativement mieux extraits par un tampon PBS ou un mélange PBS/méthanol. Cette étude confirme l'impact de la préparation d'échantillon sur l'analyse. Malgré tout, elle n'a pas permis de déterminer un protocole d'extraction compatible avec l'ensemble des métabolites. Il apparaît que la préparation d'échantillon, nécessaire avant de réaliser une analyse par spectrométrie de masse, est l'une des étapes limitant la couverture métabolique. Dans notre étude, la préparation d'échantillon a été réalisée par du méthanol uniquement, ce qui peut être l'une des raisons à l'absence de couverture du métabolisme des lipides. Il est donc envisageable de réaliser une analyse supplémentaire avec une extraction de type éthanol/dichlorométhane pour compléter la couverture métabolique.

Partie II : Optimisation de l'analyse de plasmas humains par ATR-IR et évaluation de la capacité à quantifier spécifiquement un/des biomarqueur(s)

I) Utilisation de la spectroscopie infrarouge comme outil diagnostic: état de l'art

I.1) La complexité de l'analyse

La spectroscopie infrarouge et la spectroscopie Raman sont deux techniques qui ont l'avantage de nécessiter peu ou pas de traitement des échantillons avant analyse et qui permettent d'avoir une vision globale de la composition biochimique. De plus, la rapidité d'acquisition des spectres permet d'envisager cette technologie pour des analyses nécessitant un haut débit analytique telle que le contrôle de production dans l'industrie ou l'analyse de plasmas en routine en milieu hospitalier pour des analyses de dépistage « au lit du malade » de la même manière qu'un glucomètre avec un rendu de résultat de quelques secondes.

I.2) La place de la spectroscopie infrarouge dans le diagnostic biologique

La majorité des publications traitant de la spectroscopie vibrationnelle dans le domaine biomédical est basée sur l'analyse de tissus humains, avec des études pilotes montrant qu'il est possible de différencier les tissus sains et cancéreux, ainsi que les tumeurs bénignes et malignes (186). Plus récemment, l'intérêt pour la spectroscopie dans l'analyse des fluides biologiques s'est développé en raison de la facilité de manipulation, et de préparation des échantillons. Backhaus et al. ont utilisé

la spectroscopie infrarouge sur des sérums de patientes pour différencier celles en bonne santé de celles atteintes d'un cancer du sein. En appliquant des méthodes de traitement de données non supervisées et supervisées, ils ont rapporté des sensibilités et des spécificités >92% (187). Dans le domaine des maladies neurodégénératives, Paraskevaidi et al. ont démontré que l'ATR-IR était capable de différencier les patients atteints de certaines pathologies. La maladie d'Alzheimer (MA) a été identifiée avec une sensibilité et une spécificité de 70 %, et les patients atteints ont été séparés de ceux atteints de démence avec une précision de 90 % (188). D'autres études ont montré l'intérêt de la spectroscopie infrarouge dans la détection sérologique de l'encéphalopathie spongiforme bovine (189), dans le diagnostic de la polyarthrite rhumatoïde (190) et comme outil pronostique pour déterminer la sévérité des pancréatites aiguës (191). Malgré tout, bien que ces études aient démontré leur intérêt, la majorité a été réalisée sur de faibles effectifs sans validation externe sur une seconde population indépendante. Ces résultats doivent donc être interprétés avec précaution. De plus, le manque d'harmonisation des protocoles utilisés complique la standardisation pour une utilisation multicentrique permettant la constitution de cohortes plus importantes (192).

Ces outils peuvent être utilisés seules (119–125) mais pourraient également être utilisées comme méthode de screening de classes de molécules (lipides, glucides, acides aminés, nucléotides...) potentiellement impactées dans la physiopathologie. Une dernière approche pourrait également être d'utiliser ces technologies en combinaison avec les méthodes conventionnelles (193).

II) Article n°2 : Quantification des protéines dans le sérum humain par ATR-IR.

Dans cette étude, nous avons analysé différents protocoles de mesures et évalué leur capacité à quantifier les protéines dans le plasma humain. L'analyse sous forme liquide permet de réaliser une acquisition du spectre rapide en quelques secondes et ainsi de réaliser un grand nombre d'analyse en peu de temps. De plus, le dépôt d'un volume d'échantillon sous forme liquide peut être facilement standardisé à l'aide d'équipements contrôlés ou par l'utilisation de capillaires standardisés. La spectroscopie IR étant très sensible à l'interférence de l'eau, certaines études préconisent de réaliser l'analyse après une étape de séchage permettant l'élimination de l'eau présente dans l'échantillon. Ce séchage permet d'augmenter la sensibilité de l'analyse. Cependant, la standardisation de l'étape de séchage est complexe à cause de la nécessité de maîtriser différents paramètres influençant tel que la température et l'hygrométrie. De plus, une inhomogénéité chimique et physique peut en résulter, due à l'effet «coffee ring» ou encore l'effet de craking. Un manque de maîtrise de cette étape peut conduire à une hétérogénéité du film pour un même échantillon ou entre plusieurs échantillons, ce qui peut engendrer des incohérences dans les résultats obtenus (194–196). De plus, l'absorption en spectroscopie infrarouge ATR sous forme séchée est régie par la loi de Beer-Lambert, or une variation de l'épaisseur du film pourrait avoir pour conséquence une augmentation de la variabilité de mesure.

Dans la littérature, peu d'études comparent les résultats obtenus en fonction des différents modes d'acquisition. La majorité des études sont réalisées sur des

prélèvements après une étape de séchage sur sang total ou sur plasma (98,143). Cependant, Perez-Guaita et al ont montré que l'analyse sous forme liquide permettait également la quantification d'albumine, d'hémoglobine et d'immunoglobuline (133).

Dans ce contexte, nous avons souhaité évaluer la capacité de la spectroscopie infrarouge par ATR-IR à quantifier des biomarqueurs selon différents protocoles étudiés. Etant donné l'impact potentiel de la préparation de l'échantillon sur une analyse globale, notre volonté était de proposer une étape pré-analytique la plus simple possible.

Les objectifs de cette étude sont

- 1) L'optimisation des conditions d'acquisition des spectres ATR-IR sous forme liquide et après séchage à l'air libre
- 2) L'évaluation de la quantification des protéines par spectroscopie infrarouge en mode ATR selon ces deux modes de mesures
- 3) L'application de la régression partielle des moindres carrés pour comparer les résultats quantitatifs obtenus dans les deux conditions.

II.2) Etude préliminaire : protocoles analytiques en mode ATR

II.1.a) Influence du volume de dépôt sur le signal collecté

Le volume déposé sur le cristal ATR influence l'intensité du spectre infrarouge enregistré. La figure 20 présente l'aire sous le spectre ($1800-1492\text{ cm}^{-1}$) pour des volumes de 0,25, 0,5, 1, 2, 5, 7,5 et 10 μL d'un sérum commercial poolé analysés sous forme liquide et après séchage. Sous forme liquide, on observe une

augmentation linéaire du signal jusqu'à un volume de 2 μ L à partir duquel l'intensité est maximale. L'épaisseur de la goutte étant largement supérieure à la pénétration de l'onde évanescente, l'augmentation de l'intensité est probablement liée à l'augmentation de la surface de contact entre l'échantillon et la zone de mesure. Lors de l'analyse après séchage de différents volumes de plasma déposés, on observe également une intensité croissante jusqu'à un volume de 2 μ L. Le temps de séchage augmentant considérablement avec l'augmentation du volume de dépôt, la condition conservée pour la suite de l'étude est le volume le plus petit avec lequel nous avons obtenus la plus grande intensité, soit 2 μ L.

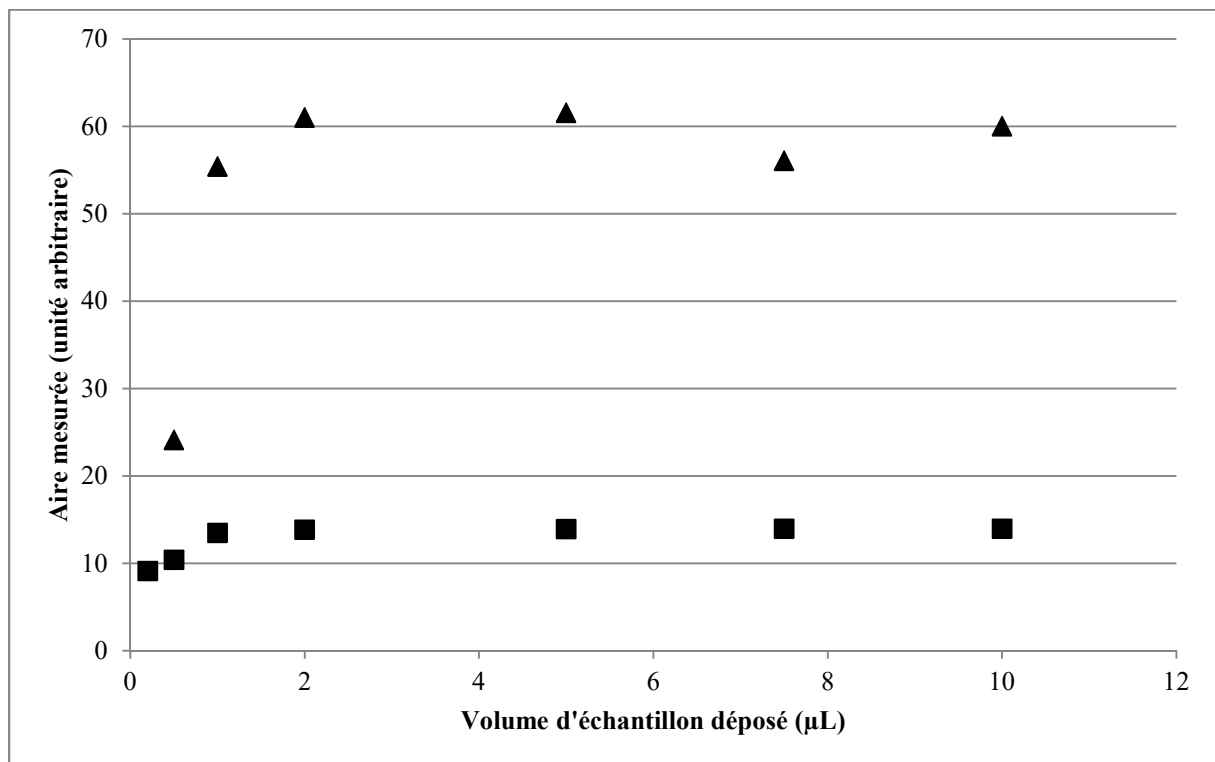


Figure 20: Évaluation de la linéarité du signal des spectres infrarouge de plasmas humains sous forme liquide (carrés) et séchée (triangles) en fonction des volumes d'échantillon déposés

II.1.b) Relation concentration/signal en fonction de la concentration en protéines

L'absorption en spectroscopie infrarouge ATR étant régie par la loi de Beer-Lambert, il était également nécessaire d'évaluer une éventuelle saturation du signal et donc d'évaluer le domaine de linéarité pour l'analyse du plasma humain. Pour mettre en évidence une éventuelle saturation du signal, nous avons étudié l'évolution du spectre ATR-IR en fonction de la concentration en protéines d'un mélange commercial de sérum humains (Figure 21). Plusieurs échantillons de concentrations différentes (de 1.875 g/L à 120 g/L) ont été réalisés en diluant le mélange commercial dans de l'eau MilliQ.

Sous forme liquide, on observe une augmentation linéaire du signal quelle que soit la concentration en protéine dans le domaine étudié. Lors de l'analyse des sérums après séchage, on observe également une intensité croissant jusqu'à une concentration de 15g/L puis un phénomène de plateau. L'épaisseur du film formé est probablement supérieure à la profondeur de pénétration de l'onde évanescente pour les échantillons dont la concentration en protéines est supérieure à 15 g/L.

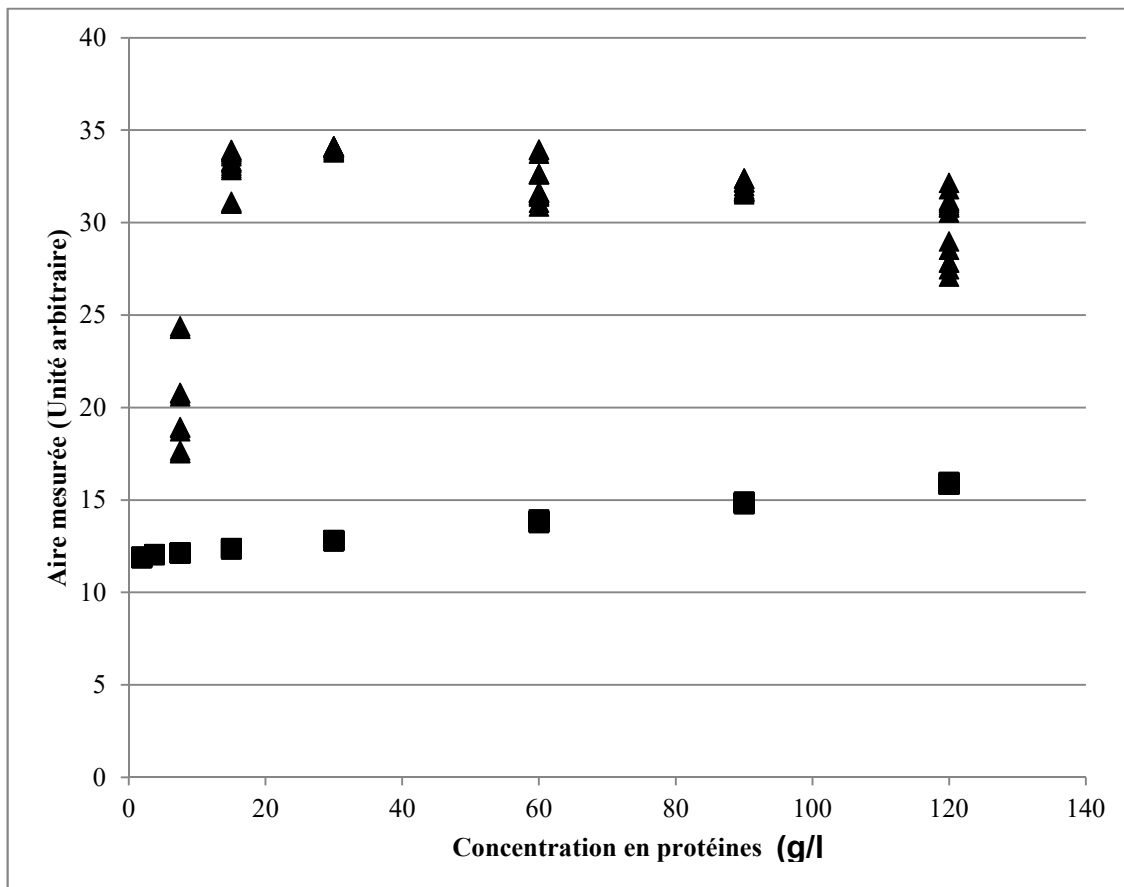


Figure 21: Évaluation de l'intensité des spectres infrarouge de plasma humain sous forme liquide (carrés) et séchée (triangles) en fonction de la concentration en protéines

II.2) Mesures des échantillons « patients »

II.2.a) Préparation des échantillons par la méthode des ajouts dosés

Dans un premier temps, nous avons analysé un mélange commercial de plasma auquel nous avons ajouté des quantités connues d'albumine et d'immunoglobuline G selon la méthode des ajouts dosés. L'albumine humaine et les immunoglobulines G ont été achetées auprès de Sigma-Aldrich (Saint-Louis, Missouri, États-Unis). Des solutions ont été préparées à des concentrations respectives de 70 g/L et 30 g/L. Ensuite, cette solution a été utilisée pour obtenir des concentrations de 46, 66, 76,

86, 96, 106 et 116 g/L en albumine et 13, 18, 24, 29, 33, 39 et 44 g/L en immunoglobuline G.

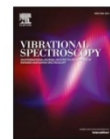
II.2.b) Obtention des valeurs de référence

Dans cette étude, nous avons évalué la fiabilité du résultat en termes de linéarité entre le signal mesuré et la concentration de l'échantillon ainsi que la variabilité de mesure. Pour cela, la collaboration entre notre unité et le laboratoire de Biochimie et Biologie Moléculaire du CHU de Tours nous a permis d'obtenir 20 échantillons de sérum recueillis dans le cadre de leur prise en charge. La quantification des protéines totales, de l'albumine et des immunoglobulines G a été réalisée sur l'ensemble de ces échantillons par des méthodes de mesures validées et accréditées sur des chaînes automatisées cobas® c501 de chez Roche® (Bâle, Suisse). L'ensemble des méthodes respecte les critères de validation de méthodes publiées par la Société Française de Biologie Clinique (SFBC) (197). Les concentrations mesurées sont considérées comme concentrations de référence pour l'étude. Puis, nous avons appliqué le même protocole aux échantillons patients. Les concentrations prédites sont ensuite comparées aux concentrations de références. Les résultats obtenus ont ensuite été mis en regard de l'interprétation biologique qui en découle.

II.2.c) Analyse par ATR-IR

Pour chaque échantillon, le spectre d'absorption ATR a été enregistré en triplicat sous forme liquide, séchée et séchée après avoir réalisé une dilution au dixième. Ensuite, la moitié des spectres mesurés a été utilisée pour construire le modèle PLSR (Partial Least Square Regression). Les concentrations en protéines, albumine

et immunoglobuline G sont ensuite prédites en appliquant le modèle de PLSR sur la moitié des spectres restante. Les concentrations prédites sont ensuite comparées aux concentrations théoriques.



Enabling quantification of protein concentration in human serum biopsies using attenuated total reflectance – Fourier transform infrared (ATR-FTIR) spectroscopy

Katie Spalding^a, Franck Bonnier^b, Clément Bruno^b, Hélène Blasco^{c,d}, Ruth Board^e,
Isabelle Benz-de Bretagne^{c,d}, Hugh J. Byrne^f, Holly J. Butler^a, Igor Chourpa^b,
Pretheepan Radhakrishnan^{a,g}, Matthew J. Baker^{a,*}

^a WestCHEM, Department of Pure & Applied Chemistry, Technology and Innovation Centre, University of Strathclyde, Glasgow, G1 1RD, UK

^b Université François-Rabelais de Tours, Faculté de Pharmacie, EA 6295 Nanomédicaments et Nanosondes, 31 Avenue Monge, 37200 Tours, France

^c Laboratoire de biochimie et biologie moléculaire, CHRU Tours, France

^d Inserm U930, Université François Rabelais, Tours, France

^e Rosemere Cancer Centre, Lancashire Teaching Hospitals NHS Trust, Royal Preston Hospital, Sharoe Green Lane, Preston, PR2 9HT, UK

^f FOCAS Research Institute, Dublin Institute of Technology (DIT), Kevin Street, Dublin 8, Ireland

^g CDT in Medical Devices & Health Technologies, Department of Biomedical Engineering, Technology and Innovation Centre, University of Strathclyde, Glasgow, G1 1RD, UK

ARTICLE INFO

Keywords:
Protein
Infrared
Attenuated total reflection
Serum
Clinical
Quantification

ABSTRACT

Changes in protein concentrations within human blood are used as an indicator for nutritional state, hydration and underlying illnesses. They are often measured at regular clinical appointments and the current analytical process can result in long waiting times for results and the need for return patient visits. Attenuated total reflectance – Fourier transform infrared (ATR-FTIR) spectroscopy has the ability to detect minor molecular differences, qualitatively and quantitatively, in biofluid samples, without extensive sample preparation. ATR-FTIR can return an analytical measurement almost instantaneously and therefore could be deemed as an ideal technique for monitoring molecular alterations in blood within the clinic.

To determine the suitability of using ATR-FTIR spectroscopy to enable protein quantification in a clinical setting, pooled human serum samples spiked with varying concentrations of human serum albumin (HSA) and immunoglobulin G (IgG) were analysed, before analysing patient clinical samples. Using a validated partial least squares method, the spiked samples (IgG) produced a linearity as high as 0.998 and a RMSEV of $0.49 \pm 0.05 \text{ mg mL}^{-1}$, with the patient samples producing R^2 values of 0.992 and a corresponding RMSEV of $0.66 \pm 0.05 \text{ mg mL}^{-1}$. This claim was validated using two blind testing models, leave one patient out cross validation and k-fold cross validation, achieving optimum linearity and RMSEV values of 0.934 and $1.99 \pm 0.79 \text{ mg mL}^{-1}$, respectively.

This demonstrates that ATR-FTIR is able to quantify protein within clinically relevant complex matrices and concentrations, such as serum samples, rapidly and with simple sample preparation. The ability to provide a quantification step, along with rapid disease classification, from a spectroscopic signature will aid clinical translation of vibrational spectroscopy to assist with problems currently faced with patient diagnostic pathways.

1. Introduction

The analysis of biofluids such as serum using vibrational spectroscopy is considered a potential solution to current problems with early and accurate diagnosis of many diseases [1] and promises improved patient mortality, morbidity and quality of life [2]. Biofluids are

routinely obtained following a minimally invasive procedure, providing a large sample volume that contains biomolecular components such as proteins, amino-acids, lipids and carbohydrates in relative concentrations which are highly dependent on demographical characteristics and physiological or pathological status [3]. Clinicians establish a diagnosis from several criteria, including; medical history, clinical symptoms,

* Corresponding author.

E-mail address: matthew.baker@strath.ac.uk (M.J. Baker).

<https://doi.org/10.1016/j.vibspec.2018.08.019>

Received 18 June 2018; Received in revised form 21 August 2018; Accepted 25 August 2018

Available online 01 September 2018

0924-2031/ © 2018 Elsevier B.V. All rights reserved.

imaging data and biological exploration. Numerous diseases are characterised by a qualitative or quantitative modification of a specific biological parameter, while others are associated with a biological signature, changes in multiple biological parameters [4,5].

Proteomics, peptidomics and metabolomics are often studied through nuclear magnetic resonance [6], mass spectrometry [7] or capillary electrophoresis [8]. A large number of proof-of-principle studies have identified diagnostic markers for cancers [9–12]. However, there is extensive sample preparation associated with these techniques. ATR-FTIR can provide a spectral profile of all the macromolecular classes contained within serum and a signature, as opposed to single markers, could be advantageous when analysing a heterogeneous disease such as cancer. Vibrational spectroscopic investigations have resulted in a large number of proof of principle studies that show promising results [13].

The diagnosis of gliomas (high-grade and low-grade) from non-cancer through a combination of ATR-FTIR and multivariate support vector machine analysis (SVM), was achieved with average sensitivities and specificities of 93.75 and 96.53% respectively for human serum samples [14]. In 2016, a large serum study using FTIR spectroscopy was completed, reporting the discrimination of cancer vs non-cancer patients with a sensitivity of 91.5% and specificity of 83.0%, as well as deciphering cancer severity and the primary site of metastasis [15]. These classification values were then improved to 92.8 and 91.5%, sensitivity and specificity, by executing Random forest and 2D correlation analysis in combination [16]. The application of vibrational spectroscopy to analyse tissue sections, as well as single cells [17,18], has also been hugely successful. The advantages of vibrational spectroscopy, such as ATR-FTIR, and high classification values demonstrates a potential use as the gold standard for patient disease screening using serum [19–21].

To facilitate the translation of an infrared spectroscopy based diagnostic test, the incorporation of a quantification step could be regarded as beneficial and complementary to current clinical practice as the majority of clinical tests are currently based upon quantitative values as opposed to signatures or fingerprints. Protein vibrations are often the most prominent in a biological infrared spectrum [22]. Furthermore, protein concentrations are systematically measured in routine practice; they are useful to interpret biological parameters, discuss nutritional status, extracellular hydration status or to help in the diagnosis of some diseases. Specific proteins such as human serum albumin (HSA) and immunoglobulinG (IgG), (as well as the ratio of the two), may be altered in the case of inflammation, infection, unexplained weight loss, fatigue or act as symptoms of kidney or liver disease [23,24]. HSA constitutes between 57–71 % of the serum composition, and globulins 8–26 % [25]. HSA and IgG could be regarded as ideal to produce models in order to demonstrate an ATR-FTIR spectroscopic test capable of quantifying proteins.

Infrared spectroscopy enables the production of a unique spectrum representative of the fundamental molecular vibrations that occur within the sample, that provides a ‘fingerprint’ of the sample [26,27]. The combination of the rapid collection method obtained through the FTIR systems and spectroscopic method development has accelerated biomedical research using infrared spectroscopy. In particular, ATR-FTIR spectroscopy has been shown to be suitable for biological materials, due to the minimal sample preparation and the ability to analyse a variety of samples types, including serum [1,28–30]. An advantageous property of IR based techniques, is that they obey the principles of the Beer Lambert law [31], allowing quantification of a given molecule relative to the absorbance of light in the sample it is travelling through. This enables ATR-FTIR spectroscopy to quantify specific biomolecule concentrations, as the proportion of light absorbed by the sample will correlate with the concentration of molecules within a sample.

This is evident from the wide variety of research carried out, quantifying particular biomarkers from biofluid samples [32–34]. For example, the analysis of dried serum deposits using transmission

spectroscopy highlighted the ability to quantify eight serum analytes [35] and the simultaneous quantification of glucose and urea analytes in addition to malaria parasitaemia from a single drop of blood dried on a glass slide [36]. The latter highlights the capability of using ATR-FTIR spectroscopy to determine disease and metabolic state, through the identification and quantification of chemical parameters associated with the disease diagnosis. Furthermore, the concentration of *in situ* DNA within cells [37], as well as the metabolite concentrations in urine [38] and saliva [39], could be determined using ATR-FTIR and bovine IgG was quantified using transmission and ATR-FTIR spectroscopy [40]. The quantification of glycine, a low molecular weight fraction (LMWF), provided evidence that ATR-FTIR spectroscopy can monitor systemic spectral modifications created by spiking human serum with lyophilised glycine [41]. Additionally, the removal of high molecular weight fractions (HMWF), through centrifugal filtration, led to an increased precision and accuracy of the quantitative models based on the partial least squares algorithm [42]. Research carried out by Perez-Guaita in 2012 [43], showed the possibility of determining total albumin, total globulin and immunoglobulin concentrations through the analysis of 50 μ l liquid serum samples deposited on an ATR crystal cell. This work highlighted the potential for ATR-FTIR to act as a green alternative to current methods used within hospitals, through the removal of reagents and implementation of relatively cheap and simple instrumentation. However, no sample preparation study was performed in order to establish the optimum sample preparation with minute volumes of serum.

Infrared spectral datasets are information rich, highlighting underlying biological and structural differences. Coupled with powerful multivariate analysis approaches, they have the ability to differentiate between disease classes by extracting relevant information. Multiple data mining approaches have been used in spectral data analysis, such as Principal Component Analysis (PCA), Random Forest (RF) and Support Vector Machine (SVM), all demonstrating the ability to discriminate diseased from non-diseased biofluid samples [44]. Currently, Partial Least Squares Regression analysis (PLSR) is one of the most frequently used techniques for the production of quantitative models, due to its ability to identify systematic variations of contributing factors and generate quantitative predictive models. This allows the prediction of unknowns, using the latent variables extracted from the regression model [40,32,45,46].

ATR-FTIR spectroscopy has the ability to detect minor differences in biofluid samples, with minimal sample preparation, and multiple proof-of-principle studies have highlighted the potential clinical use for such a technique. However, translation of ATR-FTIR spectroscopy has not occurred due to multiple factors, including the lack of acceptance from clinical environments.

We show, for the first time, an optimised methodology to enable protein quantification in single and complex mixtures using a PLSR approach, detailing the in-depth progression of determining protein concentration from spiked samples, to patient samples, before blind testing methods. The incorporation of this new quantification step within biofluid diagnostic methodologies would enable a direct comparison to gold standard diagnostic methods and highlight the clinical excellence of vibrational spectroscopic analysis of biofluids and facilitate translation.

2. Materials and methods

2.1. Sample preparation methodology

For the first time, an in depth methodological investigation was performed in order to establish the optimum sample preparation protocol for quantification from serum based ATR-FTIR spectroscopy. This study was performed using 2 models samples sets 1) Whole Serum Dilution Study and 2) Spiked Human Serum Models, before moving onto patient samples. Table 1 and subsections, 2.1.1 - 2.1.3, below provide further information on experimental details.

Table 1
Experimental details.

	Whole Serum Dilution Study	Spiked Human Serum Models	
		HSA	IgG
Sample Preparation	2-fold dilutions from pure human pooled serum	0.14 g HSA/2000 µl human pooled serum	0.06 g IgG/2000 µl human pooled serum
Sample Concentrations	100, 50, 25, 12.5, 6.25, 3.125, 0 %	116.3, 106.29, 96.28, 86.27, 76.33, 66.32, 46.3 mg mL ⁻¹	43.53, 38.58, 33.48, 28.53, 23.58, 18.48, 13.53 mg mL ⁻¹
Sample States Analysed	10 µl liquid, 1 µl pure air dried	10 µl liquid, 1 µl pure air dried, 2 µl 10% diluted air dried	10 µl liquid, 1 µl pure air dried, 2 µl 10% diluted air dried

2.1.1. Whole serum dilution study

To determine the ability of ATR-FTIR spectroscopy to detect variable protein concentrations, 1000 µl of commercially available, whole, sterile, filtered, mixed pool human serum (TCS Biosciences, UK) were used to create a set of seven 2-fold dilutions using deionised water (Milli-Q water (Millipore Elix S)).

2.1.2. Spiked human serum models

HSA and IgG (Sigma-Aldrich, UK) were used to create two spiked pooled serum models. Due to their abundant nature within human blood, current clinical use and detection over a range of concentrations from patient samples, these two proteins were deemed ideal for comparative measurements. Preparation of the two separate spiked models were created by the addition of each protein independently, while maintaining the concentration of the other. The initial concentrations of HSA and IgG of 46.3 and 13.53 mg/ml, respectively, were also taken into consideration. Further details of this can be found in Table 1 and Supporting Information (SI Materials and Methods, *Spiked Human Serum Models and Table S-1*).

2.2. Patient sample protein levels

Serum samples collected at the Biochemical laboratory at the University Hospital CHU Bretonneau de Tours, for the measurement of total protein, HSA and IgG were used for research, obeying the ethical procedures implemented by the hospital. The concentrations of total protein, HSA and IgG were measured using a Cobas 6000 analyser series (Roche Diagnostics) with a measurement precision of 1 g/l – (Table S-2). The samples were obtained following an in-house standard operating procedure, developed by the hospital for the routine analysis of serum samples. Whole blood was collected using a dry tube with separate gel and coagulation activator. After at least one hour of clotting, blood was centrifuged for 10 min at 3000 g to isolate the serum from the other blood components. Sera were then analysed by immunoturbidimetry (IgG and HSA) and a colorimetric assay based on copper reaction for total proteins. The remainder of the blood serum was stored at –20 °C until ATR-FTIR experiments were carried out.

2.3. Data collection using ATR – FTIR spectrometer

ATR-FTIR spectra were recorded using a diamond crystal and a single reflection golden gate accessory (Specac, UK) attached to a Bruker Vector 22 (Bruker, Germany). 32 co-added scans, covering a wavenumber range of 4000 – 400 cm⁻¹, were combined to produce the spectrum, using a spectral resolution of 4 cm⁻¹. A background spectrum (32 co-added scans), using the same spectral range, of the ambient conditions was automatically subtracted by the OPUS package version 4.2 (Bruker, Germany) to create the sample spectrum.

The sample preparation approaches used are liquid, air dried and liquid samples which have been diluted by 10% using deionised water, and then air dried (10% air dried). Spectra from liquid serum samples were obtained in triplicate, immediately after the drop was deposited on to the crystal. Dry serum samples (air dried and 10% air dried) were also obtained in triplicate, following a drying time of between 5–8 minutes. The triplicate analysis of each serum drop accounted for any instrumental variance. This process was repeated five times to encompass any biological variance between the samples. The serum was dropped at a perpendicular angle using a micropipette to ensure a high level of reproducibility. This led to the production of a spiked data set containing 105 spectra and a patient data set containing 300 spectra, for analysis. Information on how the samples were confirmed as dry and the triplicate analysis is included in Supporting Information (Materials and Methods, *Data Collection Using ATR – FTIR Spectrometer*).

2.4. Data pre-processing and analysis

Matlab (Mathworks, USA) was used to carry out all pre-processing and data analysis. A rubber-band baseline correction and vector normalisation using University of Strathclyde, in house written software was applied to the fingerprint region (1800 – 900 cm⁻¹) (Fig. 1). Pre-processing allowed the systematic increase of the two protein concentrations to be observed and assessed, by removing any non-biochemical components of the spectra and enabling clearer analysis of spectral variations in the amide region.

PLSR was used to quantify the prepared protein concentrations from

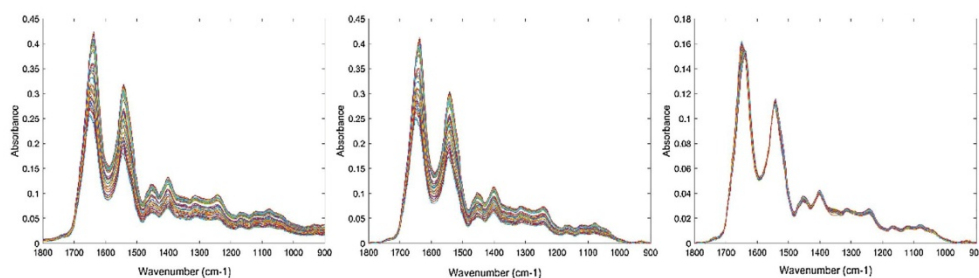


Fig. 1. Highlighting spectral pre-processing steps. From left to right, raw data, baseline corrected data and finally baseline corrected and vector normalised.

the spiked samples as well as estimate the serum protein levels in patient samples. The algorithm is a supervised method, whereby the concentrations are provided to the model prior to running the analysis. The PLSR models discussed have been built from the pre-processed data sets. The analysis gives an estimated value for the model accuracy [47] and is termed the Root Mean Square Error (RMSE), as well as an R^2 value indicating the linearity between the experimental and the predicted concentrations.

2.5. Whole serum dilution study PLSR optimisation

During the PLSR optimisation step, mean spectra were identified following pre-processing to calculate the area under the curve (AUC) and determine whether ATR-FTIR can detect and quantify protein concentrations. This was carried out on pure air dried and pure liquid, 2-fold dilution, samples to determine if a dilution factor was required.

2.6. Spiked and patient model validation

To validate the robustness of the PLSR predictive models, the optimum number of cross validation loops was determined, by re-sampling 512 cross validation iterations, 1000 times. This produced three convergence plots, for the outputs of the RMSE calibration (RMSEC), RMSE validation (RMSEV) and R^2 values (Fig. 2). From this plot, the optimum number of cross validation loops could be determined to produce minimal variation in the output from the PLSR analysis. This was done by curve fitting a one term power series model to the data and calculating when the gradient was < 0.0001 . The highest number of iterations was selected from the three convergence plots, with the result that the number of cross validation loops was not the same for each protein, but was consistently selected as the largest. For each iteration, the calibration set was compiled from 50% of the data, selected randomly, leaving the remaining 50% to be used for the validation set for the quantitative predictions. The mean and standard deviation of the RMSE and R^2 were calculated from each iteration. This methodology was carried out prior to all PLSR analysis in order to validate any results obtained, as PLSR is a supervised method and may be prone to over-fitting the data. Once the ability of ATR-FTIR spectroscopy to determine the protein concentration of spiked as well as patient samples was identified, the patient set was used to blindly test the models.

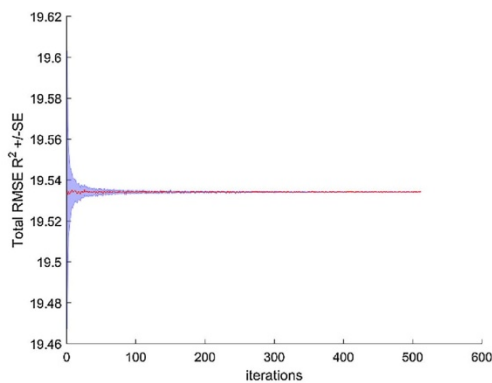


Fig. 2. A representative convergence plot of the R^2 value $\pm$ SE vs. the no. of iterations from the 10% diluted air dried globulin analysis. This particular plot led to the selection of 26 iterations which was compared to the number selected from the RMSEC and the RMSEV plots, before the highest value was selected and taken forward to the PLSR analysis.

2.7. Blind testing model validation

The patient sample dataset was then used to create both the calibration and validation sets. To begin with, a leave one patient out cross validation (LOPOCV) method was employed, whereby 19 patient samples were used as the training set and the remaining 1 was used to test the model. A similar methodology was then repeated, whereby 15 patients were selected to act as the training set, leaving the remaining 5 to act as the test set and be blindly predicted, in a process termed K-fold cross validation. Both approaches were optimised ensuring the maximum number of combinations were carried out as cross validation iterations. As such, this led to the former approach being repeated to cover all 20 possible combinations of selecting one patient out of 20, and the latter approach carried out over all the 15,504 possible combinations of selecting five from 20 patients. As the IgG concentration from patient 18 was not available (Table S-2), the model validation for the IgG concentrations was based on 19 patients as opposed to 20. Thus, the training sets contained 18 and 14 patients, for the LOPOCV and the K-fold methodologies respectively, reducing the number of possible combinations to 19 and 11,628, respectively.

3. Results and discussion

3.1. Quantification of protein concentrations in spiked human serum

3.1.1. Determining dilution factor

The analysis of biofluids, such as serum, using ATR-FTIR, produces high quality spectra with clearly defined spectral features [48]. In an ATR-FTIR spectrum, spectral peaks can be assigned to particular biomolecules, in order to allow the function, structure and biochemical signature of the sample to be identified [28]. Due to the strong water absorbance of IR light, air dried samples are generally preferred over liquid samples, although the biomolecular composition of the serum is unchanged [49]. The spectra of air dried pooled human serum exhibit the expected spectral features and assignments associated with human serum (Fig. 3). These can be briefly described as; 3280 cm^{-1} (H–O–H stretching), 2957 cm^{-1} (asymmetric CH_3 stretching), 2920 cm^{-1} (asymmetric CH_2 stretching), 2872 cm^{-1} (symmetric CH_3 stretching), 1650 cm^{-1} (amide I of proteins), 1536 cm^{-1} (amide II of proteins), 1453 cm^{-1} (CH_2 scissoring), 1394 cm^{-1} (C=O stretch of COO^-), 1242 cm^{-1} (asymmetric PO_2 stretch), 1171 cm^{-1} (ester C–O symmetric stretch) and 1080 cm^{-1} (CO stretch) [50]. The spectra are strongly dominated by the abundant proteins contained in the serum, which are present in high concentration compared to the other low molecular weight (LMW) components. In fact, the amide I peak at 1650 cm^{-1} has the highest intensity within the entire spectrum.

Fig. 3 also shows the spectra of serum solutions which have been serially diluted before drying. The spectral peak centroids identified and assigned above remain unchanged by the dilution process, and the impact of the dilution procedure can be monitored by plotting the integrated area under the curve of the fingerprint region, as shown in Fig. 3 (right). The curve shows an approximate linear dependence of integrated absorbance as a function of concentration in the low concentration region, but the behaviour rapidly deviates from linearity above 30% dilution. Notably, after 50% dilution of the stock solution, the integrated absorbance decreases by only less than 5%. The non-linearity and saturation of the absorbance of dried deposits measured by ATR-FTIR, as a function of solution concentration has previously been discussed by Bonnier et al. [41]. Importantly, for the methodology employed in the current study, in order to produce the models spiked with protein, further protein will need to be added to the pooled serum to incorporate a wide concentration range. The minimal change in absorption above a 50% dilution factor, shown in Fig. 3 suggests the identification of an upper detection limit for the volume deposited. Therefore, for the analysis of the air dried serum, a dilution factor of 10%, is required to ensure different protein concentrations are

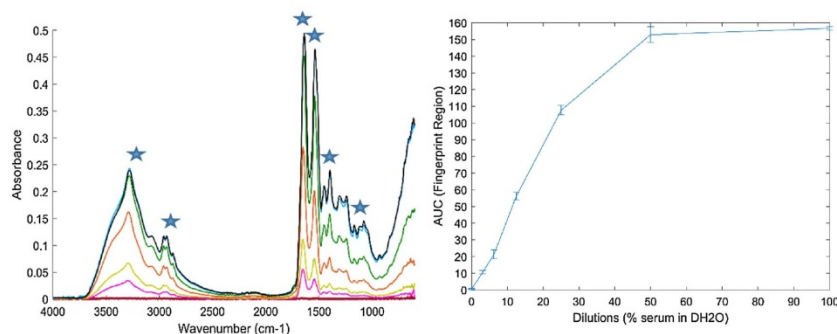


Fig. 3. Left: Mean ATR-FTIR spectra collected from the analysis of the air dried 2-fold dilution set of pooled serum. Red: 0% serum, Pink: 3.125%, Yellow: 6.25%, Orange: 12.5%, Green: 25%, Blue: 50% and Black: 100% Stars highlight peaks of interest. Right: AUC plot from the fingerprint region (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

observable. This could also have been combatted experimentally by depositing smaller volumes. However, to satisfy the requirement to cover the entire internal reflection element (IRE) and reproducible pipetting, diluting the larger volume was determined to be the optimum experimental approach.

In contrast, following the analysis of the liquid samples, biomolecular spectral assignments are difficult, due to the dominant water contribution from the broad O–H stretching band around 3300 cm^{-1} and bending vibration around 1680 cm^{-1} , (Fig. 4). The AUC plot, shown in Fig. 4 (right), indicates that serum dilution has a minimal effect on the integrated absorbance over the concentration range, although a linear decrease is observed below the 6.25% dilution, with an R^2 value of 0.9979, which may be associated with a disaggregation phenomenon [51], therefore no dilution is necessary.

3.1.2. Construction of the quantitative model: partial least square regression (PLSR)

Prior to the analysis of patient samples, a quantitative model using the PLSR algorithm was produced to evaluate the ability of ATR-FTIR to quantify protein concentration within a complex medium, such as human serum. This was applied to the protein spiked human serum models that reflect the clinically relevant protein concentrations, which tend to lie outside the normal ranges of 34–54 mg/ml for HSA and 8.1–23 mg/ml for IgG, within human blood in order to optimise the protocol.

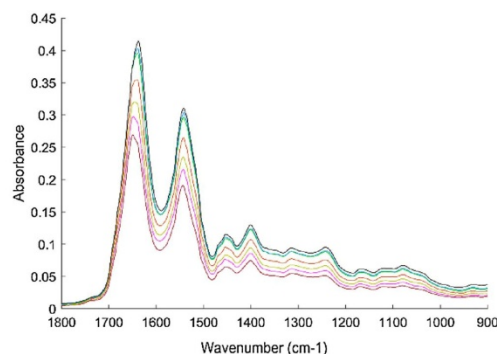


Fig. 5. Mean ATR-FTIR fingerprint spectra following the analysis of the 10% diluted air dried, IgG spiked samples. Red: 13.53 mg/ml, Pink: 18.48 mg/ml, Yellow: 23.58 mg/ml, Orange: 28.53 mg/ml, Green: 33.48 mg/ml, Blue: 38.58 mg/ml and Black: 43.53 mg/ml (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

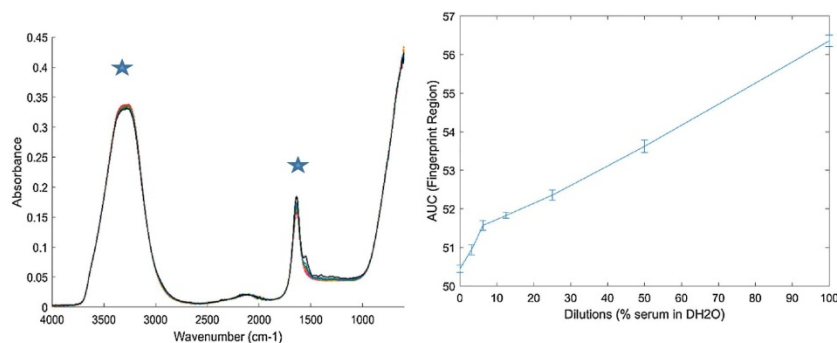


Fig. 4. Left: Mean ATR-FTIR spectra collected from the analysis of the liquid 2-fold dilution set of pooled serum Red: 0% serum, Pink: 3.125%, Yellow: 6.25%, Orange: 12.5%, Green: 25%, Blue: 50% and Black: 100% Stars highlight peaks of interest. Right: AUC plot from the fingerprint region (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

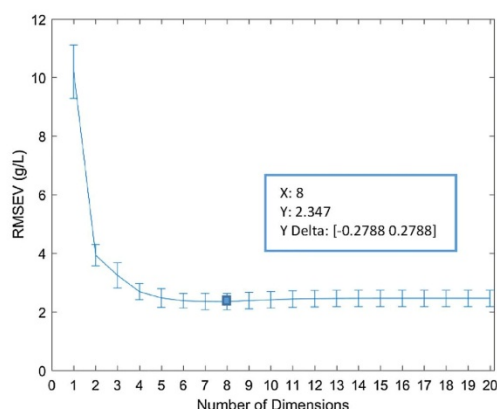


Fig. 6. Evolution of the root mean square error on the validation set (RMSEV). In this case, values are averaged from the 234 cross validations.

Fig. 5 shows the mean ($n = 9$) ATRFTIR spectral fingerprint region for spectra obtained from IgG spiked, 10% diluted air dried serum samples. The data shows an increasing absorbance trend moving from the stock solution (red) to the highest concentration of protein (black), highlighting the systematic increase in the protein amide bands at 1640 and 1560 cm^{-1} associated with the increased concentration of IgG.

To determine any relationship between variations in the spectra and the protein concentrations (HSA and IgG), PLSR analysis was conducted. The optimum number of dimensions was selected by plotting the RMSE from the validation set vs. the number of dimensions, and an example of such a plot, for the 10% diluted air dried albumin analysis, is shown in Fig. 6. For consistency, the minimum point on the curve was selected, in this case the 8th dimension, determined at $2.347 \pm 0.2788 \text{ mg mL}^{-1}$. This information is fed into a predictive model to compare the estimated concentrations from the spectral data set to the known concentrations from the produced solutions.

Table 2 summarises the predictive values from the protein spiked models, also provided in Supporting Information (SI Results and Discussion, Figure S-2). By comparing the R^2 values, as well as the RMSE of the validation set (RMSEV), the best overall result came from the 10% diluted air dried samples. These results show that concentrations can be estimated unambiguously and that dilution ensures that the protein absorbances are within the range of validity of the Beer-Lambert Law. Results highlight that, after air drying, consistent and reproducible spectra are obtained. The best individual result came from the pure air dried IgG spiked samples, the linearity being, $R^2 = 0.998$ and the RMSEV being $0.49 \pm 0.05 \text{ mg mL}^{-1}$.

Summarising, the standard deviations across all the predictive values represent good repeatability between the cross-validation iterations. As the 10% diluted air dried samples produce the best overall predictive values, this sample preparation protocol was adopted for the analysis of the patient samples. Interestingly, the PLSR models from the liquid samples produce comparable predictive results to the air dried samples. Due to the speed and ease of acquiring liquid ATR-FTIR

spectra, given the removal of the rate determining step (5–8 minute drying time), the liquid sample state was also considered in the patient sample analysis steps.

It was necessary to carry out these methodological analysis steps before progressing to patient samples and model blind testing, to establish the optimum sample preparation protocol, leading to the best possible predictive values. While the work carried out by Perez et al. showed excellent results, this particular type of methodology development procedure was not considered. This led to protein quantification of 50 μL liquid serum samples using an ATR crystal cell, potentially missing the demonstrated potential of diluted serum sample analysis [43].

3.2. Protein level quantification in patient samples

Due to the natural biological variation between individuals, the analysis of patient samples can be considered more challenging than spiking commercially available pooled human serum. Spiking a sample with a known amount of a specific biological component can model one physiological change, whilst everything else remains consistent. Between patients, the composition of blood can vary for multiple reasons, including diet, time of sample collection, as well as their disease state. During routine blood analysis, multiple biomolecular concentrations are measured, including the total protein concentration made up mainly of HSA and immunoglobulins. It is therefore important to analyse patient samples in order to understand the potential variance in the spectral response in order for these spectra to be used for clinical purposes. As the 10% diluted air dried samples produced the best predictive models, the patient samples were analysed in this sample state. In addition, the patient samples were also analysed in the liquid form, due to the promising performance of these samples during calibration and the shorter collection time which is an important parameter in clinical situations.

3.2.1. 10% diluted air dried patient samples

A new PLSR model was calculated based on the 20 patient samples using the 10% diluted air dried sample preparation, as listed in Table 3 and further detailed in Supporting Information (SI Results and Discussion, Figure S-3). For the quantification of total protein concentration, a RMSEV of $0.662 \pm 0.046 \text{ mg mL}^{-1}$ and an R^2 value of 0.992 was achieved. This result suggests that, despite moving to a more complex serum sample, a high level of predictive power is maintained and the relationship between the spectral variations and the total protein concentration is linear, within standard deviation. Quantification of the individual protein concentrations resulted in the HSA performing better than the spiked model and the IgG performing poorer than the spiked model, when examining the RMSEV and R^2 values detailed in Table 2 (spiked) and 3 (patient).

3.2.2. Liquid patient samples

Similarly, the results from the PLSR analysis of the 20 patient liquid samples are detailed in Table 3 and provided in Supporting Information (SI Results and Discussion, Figure S-3). It is evident that the RMSEV values are relatively consistent with those achieved for the spiked samples. The HSA patient model produced a result of $2.56 \pm 0.35 \text{ mg mL}^{-1}$, compared to the HSA spiked models result of $3.065 \pm 0.290 \text{ mg mL}^{-1}$.

Table 2
Summary of the RMSEV $\pm$ STD and R^2 values from the predictive models, for the two protein spikes.

Protein	Air Dried		Liquid		10 % Diluted Air Dried	
	RMSECV $\pm$ STD (mg mL^{-1})	R^2	RMSECV $\pm$ STD (mg mL^{-1})	R^2	RMSECV $\pm$ STD (mg mL^{-1})	R^2
HSA	4.585 ± 0.568	0.959	3.065 ± 0.290	0.982	2.347 ± 0.287	0.989
IgG	0.487 ± 0.053	0.998	2.365 ± 0.194	0.947	0.861 ± 0.104	0.993

Table 3

Summary of the RMSEV $\pm$ STD and R^2 values from the three predictive models, for the two patient sample states.

Protein	10 % Diluted Air Dried		Liquid	
	RMSECV $\pm$ STD (mg mL ⁻¹)	R^2	RMSECV $\pm$ STD (mg mL ⁻¹)	R^2
Total	0.662 $\pm$ 0.046	0.992	3.080 $\pm$ 0.483	0.831
HSA	0.848 $\pm$ 0.064	0.976	2.556 $\pm$ 0.351	0.780
IgG	1.945 $\pm$ 0.134	0.812	2.982 $\pm$ 0.346	0.566

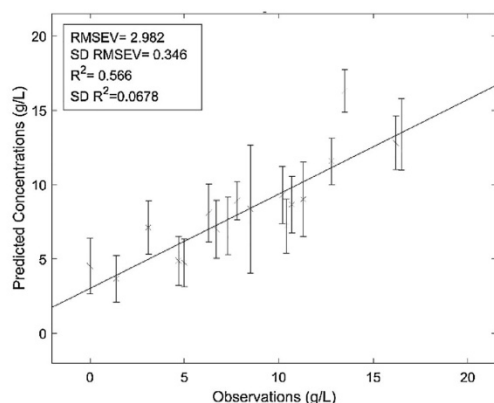


Fig. 7. Predictive model built from the PLS analysis of the liquid IgG patient samples. For each concentration the values displayed are an average of the concentration predicted from the iterations of the cross validation. Shown on the plot is the RMSEV, R^2 and the standard deviation corresponding to each of the values.

mL⁻¹. However, when comparing the RMSEV values of the patient liquid to the patient 10% diluted air dried samples, the results are dramatically higher. This suggests that the analysis of the liquid patient samples produce models with a reduced predictive power.

When considering the R^2 values, results show that the values for the liquid patient samples are considerably lower than both the spiked models and the 10% diluted air dried patient samples. The drop-in linearity implies the spectral variations and the protein concentrations show less correlation; the best result achieved was 0.831 for the total protein concentration, dramatically less than the 0.962 achieved for the analysis of the sample patient set in the 10% diluted air dried state. The error bars displayed on the plot (Fig. 7) are increased in size and in some areas, show overlap. From this analysis, it is evident that the patient sample concentrations cannot be quantified unambiguously.

3.3. Model validation

Following the determination of the optimal sample states from predicting the spiked samples and then clarifying using the patient sample concentrations, the next stage was to determine the ability of ATR-FTIR spectroscopy to predict unknown protein concentrations from serum. This was done by blind testing, removing information of the concentrations from the model.

3.3.1. Leave one patient out cross validation (LOPOCV) of patient based model

The first method of validating the use of ATR-FTIR spectroscopy to predict serum protein concentrations involved the use of a LOPOCV process. From the results in Table 4, it is evident that the prediction of

Table 4

Summary of the RMSEV $\pm$ STD and R^2 values from the two blind predictive models, for the three different protein concentrations of the 10% diluted air dried samples.

Protein	LOPOCV		K-Fold CV	
	RMSEV $\pm$ STD (mg mL ⁻¹)	R^2	RMSEV $\pm$ STD (mg mL ⁻¹)	R^2
Total	1.534 $\pm$ 1.140	0.926	1.986 $\pm$ 0.778	0.934
HSA	2.029 $\pm$ 1.260	0.890	2.491 $\pm$ 0.849	0.805
IgG	3.582 $\pm$ 2.140	0.827	4.464 $\pm$ 1.460	0.454

the total protein concentration produced the best results, with an RMSEV of 1.534 $\pm$ 1.14 mg mL⁻¹ and an R^2 value of 0.926. The prediction of individual HSA and IgG concentrations were not as effective as the total protein concentration, represented by higher RMSECV values for both proteins. The high standard deviation of the IgG data ($\pm$ 2.14), results in the model not being able to identify individual patient concentrations with precision. The R^2 values for both proteins are also lower than those of the total protein concentration, showing correlation between spectral variations and concentrations decreases.

3.3.2. K-fold cross validation of patient based model

A similar trend to the LOPOCV is apparent in Table 4. The total protein content allows for the best predictive values, followed by the HSA and then the IgG. The RMSECV for the total protein concentration prediction is higher (1.99 $\pm$ 0.78 mg mL⁻¹) than the LOPOCV model, but the lower standard deviation suggests that this is a more precise method. For this reason, the R^2 value of 0.934 is also higher than that of the previous method, showing more linearity between the predicted and observed concentrations. The prediction of the individual proteins shows the same trend. For both HSA and IgG, the k-fold blind testing produces less accurate results, in comparison to the true result but with more precision and reduced statistical variability. The linearity of the models decreases and again highlights that a linear relationship between the spectral variations and the concentrations.

Both the LOPOCV and k-fold blind testing methods produced promising results with similar trends, specifically for the prediction of total protein concentration. The reason for the poor IgG results could be due to the inability to differentiate between the variable contributions of the five major types of immunoglobulin present within human blood (IgA, IgG, IgM, IgE and IgD) [52].

4. Conclusions

Vibrational spectroscopy is widely used for the analysis of biofluids and many proof-of-principle studies have shown the capability of techniques like ATR-FTIR spectroscopy to enable disease detection, as well as quantification of biomolecules. However, the translation of vibrational spectroscopy into a clinical environment is dramatically impacted by the inability to perform a direct comparison to the current quantitative diagnostic measurements. Current clinical practice uses blood protein concentration as a non-specific disease indicator, possibly leading to further investigation and potentially a diagnosis, highlighting the advantageous nature of protein quantification.

The work presented showcases for the first time the development of the optimal methodology for the quantification of protein biomarkers in a complex background (namely 10% diluted air dried serum analysis), the inclusion of this methodology into vibrational spectroscopic diagnostics of biofluids could bridge the gap between vibrational spectroscopy and clinical practice. In addition, the drying process could be accelerated through the implementation of heating mantle, the use of a smaller sample volume or batch drying, before analysis.

This study shows how ATR-FTIR spectroscopy can be used to

quantify proteins in spiked and patient samples, rapidly, economically and with simple sample preparation. Linearity values as high as 0.992, in addition to high accuracy and precision demonstrated by RMSEV values such as $0.662 \pm 0.046 \text{ mg mL}^{-1}$, indicate that quantification of clinically relevant molecules can be conducted using this approach.

The blind testing of patient clinical samples, while maintaining desirable linearity ($R^2 = 0.934$), precision and accuracy (RSMEV = $1.986 \pm 0.778 \text{ mg mL}^{-1}$), illustrates the potential use of this technique within a clinical setting and its incorporation could bridge the gap between vibrational spectroscopy and current clinical analyses. The development of a quantification step in addition to disease differentiation shows great promise to enable a dynamic clinical diagnostic platform that can improve the current patient diagnostic pathway.

Competing interests statement

The authors declare that no competing interests exist.

Acknowledgements

The authors would like to acknowledge funding from the Postgraduate and Early Career Researcher Exchanges (PECRE) and the Pool Engagement in European Research (PEER) schemes, as well as support from Rosemere Cancer Foundation and the Spectral Analytics Laboratory, University of Strathclyde.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.vibspec.2018.08.019>.

References

- [1] M.J. Baker, S.R. Hussain, L. Lovergne, V. Untereiner, C. Hughes, R.A. Lukaszewski, G. Thieffin, G.D. Sockalingum, Developing and understanding biofluid vibrational spectroscopy: a critical review, *Chem. Soc. Rev.* 45 (2015) 1803–1818, <https://doi.org/10.1039/C5CS00585J>.
- [2] J.R. Hands, P. Abel, K. Ashton, T. Dawson, C. Davis, R.W. Lea, A.J.S. McIntosh, M.J. Baker, Investigating the rapid diagnosis of gliomas from serum samples using infrared spectroscopy and cytokine and angiogenesis factors, *Anal. Bioanal. Chem.* 405 (2013) 7347–7355, <https://doi.org/10.1007/s00216-013-7163-z>.
- [3] D.W. Greening, R.J. Simpson, Serum/Plasma Proteomics 728 (2011), pp. 259–265, <https://doi.org/10.1007/978-1-61779-068-3>.
- [4] J. Liu, Y. Duan, Saliva: a potential media for disease diagnostics and monitoring, *Oral Oncol.* 48 (2012) 569–577, <https://doi.org/10.1016/j.oraloncology.2012.01.021>.
- [5] R.J. Perry, V.T. Samuel, K.F. Petersen, G.I. Shulman, N. Haven, N. Haven, HHS Public Access 510 (2015), pp. 84–91, <https://doi.org/10.1038/nature13478>.
- [6] S. Beaulacq, L. Nadal-Desbarats, C. Hennequet-Antier, A. Collin, S. Tesserand, M. Bourin, E. Le Bihan-Duval, C. Berri, Serum and muscle metabolomics for the prediction of ultimate pH, a key factor for chicken-meat quality, *J. Proteome Res.* 15 (2016) 1168–1178, <https://doi.org/10.1021/acs.jproteome.5b01050>.
- [7] C. Bruno, D. Dufour-Rainfray, F. Patin, P. Vourch, D. Guilloteau, F. Maillot, F. Labarthe, M. Tardieu, C.R. Andres, P. Emond, H. Blasco, Validation of amino-acids measurement in dried blood spot by FIA-MS/MS for PKU management, *Clin. Biochem.* 49 (2016) 1047–1050, <https://doi.org/10.1016/j.clinbiochem.2016.07.008>.
- [8] E. Sato, T. Mori, E. Mishima, A. Suzuki, S. Sugawara, N. Kurasawa, D. Saigusa, D. Miura, T. Morikawa-Ichinoe, R. Saito, I. Oba-Yabana, Y. Oe, K. Kisu, E. Naganuma, K. Koizumi, T. Mokudai, Y. Niwano, T. Kudo, C. Suzuki, N. Takahashi, H. Sato, T. Abe, T. Niwa, S. Ito, Metabolic alterations by indoxyl sulfate in skeletal muscle induce uremic sarcopenia in chronic kidney disease, *Sci. Rep.* 6 (2016) 36618, <https://doi.org/10.1038/srep36618>.
- [9] G.L. Wright Jr, L. Cazares, S.-M. Leung, S. Nasim, B.-L. Adam, T. Yip, P. Schellhammer, L. Gong, A. Vlahou, ProteinChip® Surface Enhanced Laser desorption/ionization (SELDI) Mass Spectrometry: A Novel Protein Biochip Technology for Detection of Prostate Cancer Biomarkers in Complex Protein Mixtures, (2000), <https://doi.org/10.1038/sj.pcan.4500384>.
- [10] A. Vlahou, P.F. Schellhammer, S. Mendrinis, K. Patel, F.I. Kondylis, L. Gong, S. Nasim, G.L. Wright, Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine, *Am. J. Pathol.* 158 (2001) 1491–1502, [https://doi.org/10.1016/S0002-9440\(10\)64100-4](https://doi.org/10.1016/S0002-9440(10)64100-4).
- [11] E.F. Petricoin, A.M. Ardekani, B.A. Hitt, P.J. Levine, V.A. Fusaro, S.M. Steinberg, G.B. Mills, C. Simone, D.A. Fishman, E.C. Kohn, L.A. Liotta, Use of proteomic patterns in serum to identify ovarian cancer, *Lancet.* 359 (2002) 572–577, [https://doi.org/10.1016/S0140-6736\(02\)07746-2](https://doi.org/10.1016/S0140-6736(02)07746-2).
- [12] H. Zhang, G. Wu, H. Tu, F. Huang, Discovery of serum biomarkers in astrocytoma by SELDI-TOF MS and proteinchip technology, *J. Neurooncol.* 84 (2007) 315–323, <https://doi.org/10.1007/s11060-007-9376-5>.
- [13] K. Spalding, R. Board, T. Dawson, M.D. Jenkinson, M.J. Baker, A review of novel analytical diagnostics for liquid biopsies: spectroscopic and spectrometric serum profiling of primary and secondary brain tumors, *Brain Behav.* 6 (2016) 1–8, <https://doi.org/10.1002/brb3.502>.
- [14] J.R. Hands, K.M. Dorling, P. Abel, K.M. Ashton, A. Brodbelt, C. Davis, T. Dawson, M.D. Jenkinson, R.W. Lea, C. Walker, M.J. Baker, Attenuated total reflection fourier transform infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples, *J. Biophotonics* 7 (2014) 189–199, <https://doi.org/10.1002/jbio.201300149>.
- [15] J.R. Hands, G. Clemens, R. Stables, K. Ashton, A. Brodbelt, C. Davis, T.P. Dawson, M.D. Jenkinson, R.W. Lea, C. Walker, M.J. Baker, Brain tumour differentiation: rapid stratified serum diagnostics via attenuated total reflection Fourier-transform infrared spectroscopy, *J. Neurooncol.* 127 (2016) 463–472, <https://doi.org/10.1007/s11060-016-2060-x>.
- [16] B.R. Smith, K.M. Ashton, A. Brodbelt, T. Dawson, M.D. Jenkinson, N.T. Hunt, D.S. Palmer, M.J. Baker, Combining random forest and 2D correlation analysis to identify serum spectral signatures for neuro-oncology, *Analyst* 141 (2016) 3668–3678, <https://doi.org/10.1039/C5AN02452H>.
- [17] G. Clemens, J.R. Hands, K.M. Dorling, M.J. Baker, Vibrational spectroscopic methods for cytology and cellular research, *Analyst* 139 (2014) 4411–4444, <https://doi.org/10.1039/c4an00636d>.
- [18] Z. Farhane, F. Bonnier, A. Casey, A. Maguire, L. O'Neill, H.J. Byrne, Cellular discrimination using in vitro Raman micro spectroscopy: the role of the nucleolus, *Analyst* 140 (2015) 5908–5919, <https://doi.org/10.1039/C5AN01157D>.
- [19] J. Backhaus, R. Mueller, N. Formanski, N. Szlama, H.G. Meerpohl, M. Eit, P. Bugert, Diagnosis of breast cancer with infrared spectroscopy from serum samples, *Vib. Spectrosc.* 52 (2010) 173–177, <https://doi.org/10.1016/j.vibspec.2010.01.013>.
- [20] G.L. Owens, K. Gajjar, J. Trevisan, S.W. Fogarty, S.E. Taylor, B. Da Gama-Rose, P.L. Martin-Hirsch, F.L. Martin, Vibrational biospectroscopy coupled with multivariate analysis extracts potentially diagnostic features in blood plasma/serum of ovarian cancer patients, *J. Biophotonics* 7 (2014) 200–209, <https://doi.org/10.1002/jbio.201300157>.
- [21] M.J. Baker, C.S. Hughes, K.A. Hollywood, Biophotonics: Vibrational Spectroscopic Diagnostics, (2016), <https://doi.org/10.1088/978-1-6817-4071-3>.
- [22] A. Barth, Infrared spectroscopy of proteins, *Biochim. Biophys. Acta - Bioenerg.* 1767 (2007) 1073–1101, <https://doi.org/10.1016/j.bbabi.2007.06.004>.
- [23] R.A. M, M.R. Pincus, Henry's Clinical Diagnosis and Management by Laboratory Methods, 23rd edition, (2011).
- [24] J.T. Busher, Serum albumin and globulin, *Clin. Methods Hist. Phys. Lab. Exam.* (1990) 497–499.
- [25] J. Barth, J.K. Rae, Harmonisation of Reference Intervals President, Association for Clinical Biochemistry, 2011.
- [26] P. Larkin, Introduction, *Infrared Raman Spectrosc.* (2011), pp. 1–5, <https://doi.org/10.1016/B978-0-12-386984-5.10001-1>.
- [27] R.D. Schoenwald, Basic principles, *Ther. Drug Monit.* 1 (2002) 4–33, <https://doi.org/10.1016/B978-0-12-386984-5.10002-3>.
- [28] M.J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H.J. Butler, K.M. Dorling, P.R. Fielden, S.W. Fogarty, N.J. Fullwood, K.A. Heys, C. Hughes, P. Lasch, P.L. Martin-Hirsch, B. Obinaju, G.D. Sockalingum, J. Sulé-Suso, R.J. Strong, M.J. Walsh, B.R. Wood, P. Gardner, F.L. Martin, Using Fourier transform IR spectroscopy to analyze biological materials, *Nat. Protoc.* 9 (2014) 1771–1791, <https://doi.org/10.1038/nprot.2014.110>.
- [29] A.L. Mitchell, K.B. Gajjar, G. Theophilou, F.L. Martin, P.L. Martin-Hirsch, Vibrational spectroscopy of biofluids for disease screening or diagnosis: translation from the laboratory to a clinical setting, *J. Biophotonics* 7 (2014) 153–165, <https://doi.org/10.1002/jbio.201400018>.
- [30] C. Hughes, M. Brown, G. Clemens, A. Henderson, G. Monjardez, N.W. Clarke, P. Gardner, Assessing the challenges of Fourier transform infrared spectroscopic analysis of blood serum, *J. Biophotonics* 7 (2014) 180–188, <https://doi.org/10.1002/jbio.201300167>.
- [31] P.R. Griffiths, J.A. de Haseth, Chapter-2 Theoretical Background, (2007), <https://doi.org/10.1002/047010631X>.
- [32] R.A. Shaw, H.H. Mantsch, Infrared spectroscopy in clinical and diagnostic analysis, *Encycl. Anal. Chem.* (2006) 1–20, <https://doi.org/10.1002/9780470027318.a0106>.
- [33] A. Oleszko, J. Hartwich, A. Wójciewicz, M. Gąsior-Głogowska, H. Huras, M. Komorowska, Comparison of FTIR-ATR and Raman spectroscopy in determination of VLDL triglycerides in blood serum with PLS regression, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 183 (2017) 239–246, <https://doi.org/10.1016/j.saa.2017.04.020>.
- [34] G. Sankari, E. Krishnamoorthy, S. Jayakumaran, S. Gunasekaran, V. Vishnu Priya, S. Subramaniam, S. Subramaniam, S.K. Mohan, Analysis of serum immunoglobulins using Fourier transform infrared spectral measurements, *Biol. Med.* 2 (2010) 42–48.
- [35] R.A. Shaw, S. Kotowich, M. Leroux, H.H. Mantsch, Multianalyte serum analysis using mid-infrared spectroscopy, *Ann. Clin. Biochem. An Int. J. Biochem. Lab. Med.* 35 (1998) 624–632, <https://doi.org/10.1177/000456329803500505>.
- [36] S. Roy, D. Perez-Guaita, D.W. Andrew, J.S. Richards, D. McNaughton, P. Heraud, B.R. Wood, Simultaneous ATR-FTIR based determination of malaria parasitemia, glucose and urea in whole blood dried onto a glass slide, *Anal. Chem.* 89 (2017) 5238–5245, <https://doi.org/10.1021/acs.analchem.6b04578>.

- [37] D.R. Whelan, K.R. Bamberg, L. Puskar, D. McNaughton, B.R. Wood, Quantification of DNA in simple eukaryotic cells using Fourier transform infrared spectroscopy, *J. Biophotonics* 6 (2013) 775–784, <https://doi.org/10.1002/jbio.201200112>.
- [38] H.M. Heise, G. Voigt, P. Lampen, L. Küpper, S. Rudloff, G. Werner, Multivariate calibration for the determination of analytes in urine using mid-infrared attenuated total reflection spectroscopy, *Appl. Spectrosc.* 55 (2001) 434–443, <https://doi.org/10.1366/0003702011951948>.
- [39] S. Khaustova, M. Shkurnikov, E. Tonevitsky, V. Artyushenko, A. Tonevitsky, Noninvasive biochemical monitoring of physiological stress by Fourier transform infrared saliva spectroscopy, *Analyst* 135 (2010) 3183–3192, <https://doi.org/10.1039/C0AN00529K>.
- [40] I. Elsohaby, J.T. McClure, C.B. Riley, R.A. Shaw, G.P. Keefe, Quantification of bovine immunoglobulin G using transmission and attenuated total reflectance infrared spectroscopy, *J. Vet. Diagn. Invest.* 28 (2016) 30–37, <https://doi.org/10.1177/1040638715613101>.
- [41] F. Bonnier, G. Brachet, R. Duong, T. Sojinrin, R. Respaud, N. Aubrey, M.J. Baker, H.J. Byrne, I. Chourpa, Screening the low molecular weight fraction of human serum using ATR-IR spectroscopy, *J. Biophotonics* 9 (2016) 1085–1097, <https://doi.org/10.1002/jbio.201600015>.
- [42] F. Bonnier, H. Blasco, C. Wasselet, G. Brachet, R. Respaud, L.F.C.S. Carvalho, D. Bertrand, M.J. Baker, H.J. Byrne, I. Chourpa, Ultra-filtration of human serum for improved quantitative analysis of low molecular weight biomarkers using ATR-IR spectroscopy, *Analyst* (2017), <https://doi.org/10.1039/C6AN01888B>.
- [43] D. Perez-Guaita, J. Ventura-Gayete, C. Pérez-Rambla, M. Sancho-Andreu, S. Garrigues, M. De La Guardia, Protein determination in serum and whole blood by attenuated total reflectance infrared spectroscopy, *Anal. Bioanal. Chem.* 404 (2012) 649–656, <https://doi.org/10.1007/s00216-012-6030-7>.
- [44] H.J. Byrne, M. Baranska, G.J. Puppels, N. Stone, B. Wood, K.M. Gough, P. Lasch, P. Heraud, J. Sulé-Suso, G.D. Sockalingum, Spectropathology for the next generation: quo vadis? *Analyst* 140 (2015) 2066–2073, <https://doi.org/10.1039/c4an02036g>.
- [45] Y. Xu, H. Muhamadali, A. Sayqal, N. Dixon, R. Goodacre, Partial least squares with structured output for modelling the metabolomics data obtained from complex experimental designs: a study into the Y-block coding, *Metabolites* 6 (2016), <https://doi.org/10.3390/metabo6040038>.
- [46] A. Khoshmanesh, M.W.A. Dixon, S. Kenny, L. Tilley, D. McNaughton, B.R. Wood, Detection and quantification of early-stage malaria parasites in laboratory infected erythrocytes by attenuated total reflectance infrared spectroscopy and multivariate analysis, *Anal. Chem.* 86 (2014) 4379–4386, <https://doi.org/10.1021/ac500199x>.
- [47] B. a Walther, J.L. Moore, The concept of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimators, *Ecography (Cop.)* 28 (2005) 815–829, <https://doi.org/10.1111/j.2005.0906-7590.04112.x>.
- [48] J. Coates, Interpretation of infrared spectra, a practical approach, *Encycl. Anal. Chem.* (2000) 10815–10837, <https://doi.org/10.1002/9780470027318>.
- [49] F. Bonnier, F. Petitjean, M.J. Baker, H.J. Byrne, Improved protocols for vibrational spectroscopic analysis of body fluids, *J. Biophotonics* 7 (2014) 167–179, <https://doi.org/10.1002/jbio.201300130>.
- [50] Z. Movasaghi, S. Rehman, D.I. ur Rehman, Fourier transform infrared (FTIR) spectroscopy of biological tissues, *Appl. Spectrosc. Rev.* 43 (2008) 134–179, <https://doi.org/10.1080/05704920701829043>.
- [51] B.R. Priya, H.J. Byrne, Investigation of sodium dodecyl benzene sulfonate assisted dispersion and debundling of single-wall carbon nanotubes, *J. Phys. Chem. C* 112 (2008) 332–337, <https://doi.org/10.1021/jp0743830>.
- [52] A.A. Marghoob, K. Koenig, F.V. Bittencourt, A.W. Kopf, R.S. Bart, Breslow thickness and Clark level in melanoma: support for including level in pathology reports and in American Joint Committee on cancer staging, *Cancer* 88 (2000) 589–595, [https://doi.org/10.1002/\(SICI\)1097-0142\(20000201\)88:3<589::AID-CNCR15>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-0142(20000201)88:3<589::AID-CNCR15>3.0.CO;2-I).

Discussion et perspectives

Analyse des échantillons par la méthode des ajouts dosés

Analyses sur échantillons liquides : Les essais réalisés au cours de ce travail sur l'analyse liquide montrent une corrélation entre l'augmentation de l'intensité mesurée et la concentration en protéines dans l'échantillon analysé. Les analyses réalisées sur les échantillons enrichies en IgG et en albumine ont mis en évidence un RMSE moyen de 3,06 g/L pour l'albumine et 2,365 g/L pour les immunoglobulines G sous forme liquide. Ces résultats montrent une erreur de mesure supérieure à celles recommandées par les sociétés savantes pour la quantification de ces deux paramètres (respectivement 1,83 g/L et 1,04 g/L). Ce résultat est probablement dû à la forte contribution de l'eau dans le spectre infrarouge venant interférer avec le signal des protéines.

Analyses sur échantillons séchés : Le séchage de l'échantillon avant analyse a permis d'améliorer la précision de mesure des immunoglobulines G (RMSECV 4,59 g/L en liquide vs 0,49 g/L après séchage). En effet, les erreurs moyennes calculées par le modèle PLSR pour les analyses réalisées après séchage et après dilution et séchage sont inférieures à 1 g/L. Or, la précision des concentrations en protéines totales et en albumine rendues par un laboratoire de biologie médicale n'excède pas 1 g/L. Ces résultats sont donc tout à fait comparables à ceux obtenus par les méthodes de références.

Analyses des échantillons patients

Les analyses réalisées sur les échantillons patients montrent une bonne corrélation entre les concentrations de référence et celles estimées par l'analyse infrarouge. Concernant la quantification de l'albumine, c'est l'analyse sur échantillon dilué après séchage qui a permis d'obtenir l'erreur moyenne de mesure la plus faible de 0,86 g/L. Ces résultats sont très satisfaisants en regard des recommandations SFBC. Par ailleurs, ces résultats ont été obtenus sur un petit échantillonnage (n=20) et pourraient être améliorés en augmentant le nombre d'échantillons utilisés (133).

L'erreur moyenne de quantification des immunoglobulines G est de l'ordre de 3 g/L sous forme liquide et 1,95 g/L après dilution et séchage. Ce résultat montre les limites de la spectroscopie infrarouge pour quantifier les immunoglobulines G dans le plasma. Comme pour les échantillons dopés, on observe une amélioration de la précision de mesure grâce à l'étape de séchage. Cependant, l'erreur de mesure obtenue sur les échantillons après dilution et séchage est supérieure à celle retrouvée sur les échantillons dopés. Cette erreur est probablement due à la complexité de l'échantillon. En effet, les immunoglobulines G ne sont pas les seules

présentes dans le sérum. Or le spectre d'absorption des différentes immunoglobulines est similaire. Il est également possible que d'autres composants du sérum interfèrent. Lors de l'analyse des échantillons dopés, seules les concentrations en albumine et immunoglobulines changent, alors que sur les échantillons patients, la totalité des composants peuvent être présents dans des concentrations différentes. Ces variations ne sont probablement pas prises en compte dans le modèle de prédiction, probablement dû à un effectif trop petit pour avoir « rencontré » les différentes combinaisons de concentrations en immunoglobulines. En effet, la PLSR est un modèle basé sur un apprentissage puis d'application sur une seconde population. Plus le nombre d'échantillons est important, plus la composante principale extraite pour quantifier un biomarqueur sera spécifique. Ainsi la quantification des échantillons sera également plus précise car sujette à moins d'interférence.

Finalement, la dilution préalable de l'échantillon ne permet pas d'améliorer la prédiction des concentrations vis-à-vis de l'analyse sur sérum séché. De plus, cette étape préliminaire peut également avoir pour conséquence une diminution de la sensibilité et pourrait avoir un effet néfaste sur l'analyse de composés moins concentrés.

III) Etude n°3 : Quantification simultanée des protéines et de petites molécules dans le plasma humain

INTRODUCTION

Ces dernières années, les études cherchant à quantifier précisément une ou plus molécule dans un fluide biologique se sont succédées. En 2014, Bonnier et al. ont montré que les intensités des bandes d'absorbances sur une spectre ATR-IR sont reliées aux concentrations en protéines de l'échantillon (198). La même équipe a ensuite démontré dans une seconde étude la relation entre les intensités d'absorbance et les concentrations de glucose dans un plasma dopé à des concentrations connues (199). Au vu des résultats satisfaisants obtenus sur la fraction protéique, nous avons souhaité évaluer la capacité de la spectroscopie infrarouge par ATR-IR à quantifier des biomarqueurs de bas poids moléculaire en suivant le protocole utilisé lors de la première étude. En effet ces molécules sont présentes en quantité plus faible dans l'échantillon et sont moins accessibles que les protéines.

L'objectif de cette étude est de confirmer les résultats encourageants de l'étude précédente et ceux retrouvées dans la littérature en évaluant la quantification de plusieurs biomarqueurs dans le plasma.

MATERIEL ET METHODES

1) Produit de référence

Les solutions de référence sont des solutions aqueuses à une concentration de 1g/L, réalisées à partir de poudres d'albumine humaine, d'immunoglobulines G, de D-glucose et d'urée provenant de chez Sigma-Aldrich (Saint-Louis, Missouri, Etats-Unis).

2) Echantillons patients et valeur de références

Pour cette étude, nous avons obtenu 20 échantillons de plasma de patients du CHU de Tours, recueillis dans le cadre de leur prise en charge. La quantification des protéines totales, de l'albumine, des immunoglobulines G, du glucose, de l'urée et de l'acide urique a été réalisé sur l'ensemble de ces échantillons par des méthodes de mesures validées et accréditées sur des chaînes automatisées cobas® c501 de chez Roche® (Bâle, Suisse). L'ensemble des méthodes respecte les critères de validation de méthodes publiées par la Société Française de Biologie Clinique (SFBC) (197). Les concentrations mesurées font office de concentration de référence pour l'étude (Tableau IV).

Tableau IV : Concentrations de références des échantillons patients

	Protéines totales (g/L)	Albumine (g/L)	Immunoglobuline G (g/L)	Glucose (mmol/L)	Urée (mmol/L)	Acide urique (μ mol/L)
Concentrations normales (197)	60-76	35-55	8-18	4.0-5.8	2.5-7.5	150-500
1	47	22	5.73	5.9	8.64	157
2	62	34	10.18	3.2	38.35	644
3	56	30	6.89	6.8	8.99	238
4	58	26	8.53	4.0	2.93	136
5	60	27	9.71	8	6.4	12
6	59	34	6.31	16.6	2.68	65
7	53	30	5.62	13.9	6.3	450
8	58	37	7.07	7.9	5.89	207
9	57	36	6.75	14.1	2.89	103
10	54	32	7.24	4.7	6.29	141
11	69	38	9.65	6.2	9.74	526
12	58	34	7.25	4.4	6.84	351
13	36	18	5.35	8	7.9	99
14	97	31	40.54	4.8	5.64	177
15	81	45	19.6	3.5	4.18	254
16	59	38	6.38	12.7	3.34	280
17	93	40	42.96	12.9	12.13	277
18	69	33	5.29	2.8	43.42	644
19	88	32	33.94	4.8	1.09	139
20	58	32	4.88	4.8	1.41	144

3) Paramètres d'acquisition :

Les spectres IR ont été acquis en mode de réflectance totale atténuée (ATR) à l'aide d'un spectromètre infrarouge Frontier (Perkin Elmer, France) entre 4000-400 cm^{-1} avec une résolution spectrale à 4 cm^{-1} . Un spectre de référence de l'air a été enregistré (4 scans) et automatiquement soustrait du spectre de l'échantillon (4 scans moyennés) par le logiciel. Chaque échantillon a été déposé 5 fois et 3 spectres ont été acquis à chaque fois. Enfin, 15 spectres ont été enregistrés à partir de chaque échantillon, permettant d'évaluer la variabilité inter-dépôt et intra-dépôt. Les spectres de référence obtenus à partir de solutions de composés purs ont également été collectés en utilisant des paramètres similaires.

4) Prétraitement et analyses des données

Le logiciel Matlab® a été utilisé pour construire le modèle PLSR (Partial Least Square Regression) à partir des concentrations de référence connues. Le modèle prédictif est construit en 2 temps avec la mise en place d'une validation croisée. La première étape d'apprentissage ou de calibration est réalisée sur la moitié des données sélectionnées de manière aléatoire. La deuxième étape consiste à déterminer les concentrations à partir de la 2^{ème} moitié des données qui sont utilisées projetées en aveugle. La qualité de la régression linéaire est évaluée par les valeurs R^2 et RMSECV (Root Mean Square Error of Cross Validation) qui reflète l'erreur moyenne dans les concentrations estimées. Les résultats obtenus ont ensuite été mis en regard de l'interprétation biologique qui en découle.

RESULTATS

1) Caractérisation des signatures spectrales de référence

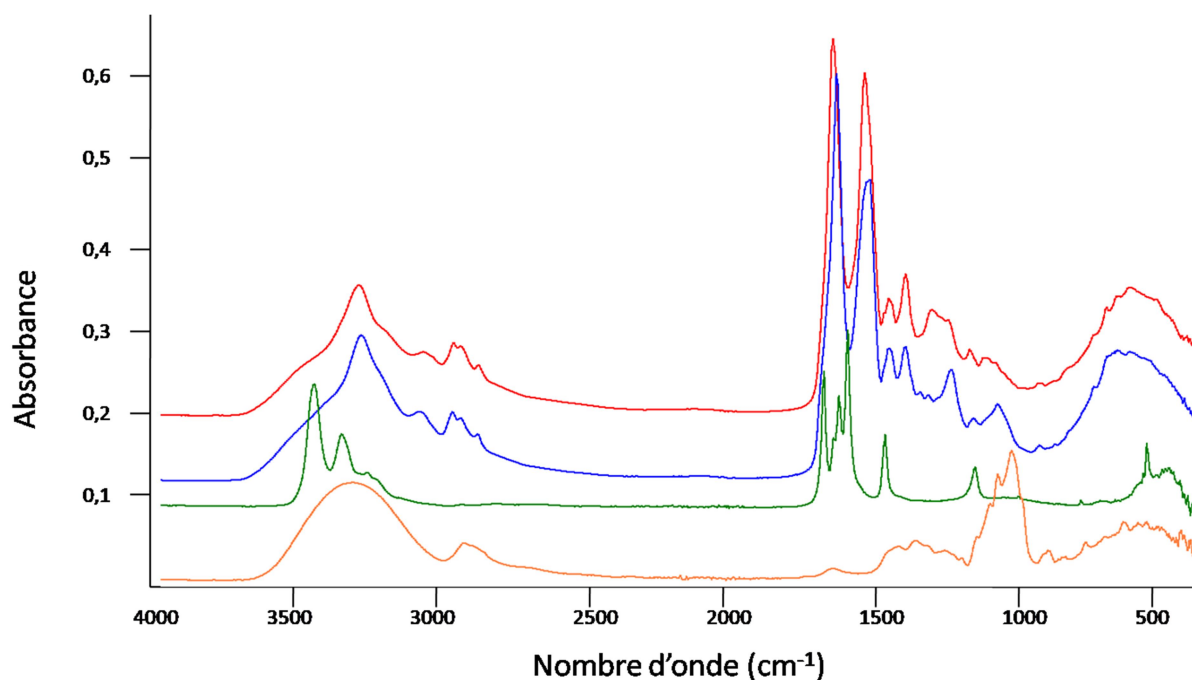


Figure 22 : Spectres ATR-IR obtenu après séchage des solutions à 1g/L d'albumine (rouge), d'immunoglobulines (bleu), de glucose (orange) et d'urée (vert). Les spectres sont volontairement décalés pour clarté.

Le spectre ATR-IR de l'immunoglobuline peut être divisé comme suit (Figure 22 bleue):

- La présence d'une forte et large bande avec un maximum à 3280cm^{-1} peut être attribuée aux vibrations N-H des liaisons amides. Les groupements CH_2 des chaînes carbonées se caractérisent par les pics à 2855 cm^{-1} et 2926 cm^{-1} qui représentent respectivement les vibrations d'élongation symétrique et asymétrique. Dans cette zone du spectre, on observe également une large bande entre 3100 et 3600 cm^{-1} correspondant aux vibrations d'élongation des fonctions OH.

- Les bandes amides I et II se caractérisent par deux pics de forte intensité à 1630 cm^{-1} et 1520 cm^{-1} .
- Le pic à 1447 cm^{-1} peut être attribué aux vibrations de déformation dans le plan (cisaillement) des CH_2 et des CH_3 . Les groupements carboxylates ainsi que les déformations des liaisons CH_3 peuvent être responsables de la bande d'absorption à 1395 cm^{-1} . Les liaisons amides provoquent également une bande d'absorption de faible intensité vers 1250 cm^{-1} . Les vibrations d'élongations des liaisons C-C se caractérisent par des pics à 1108 cm^{-1} et 938 cm^{-1} .

Le spectre ATR-IR de l'albumine est relativement proche de celui des immunoglobulines (Figure 22, rouge) :

- La présence d'une forte et large bande avec un maximum à 3280 cm^{-1} qui peut être attribuée aux vibrations des liaisons amides. Les groupements CH_2 des chaînes carbonées à 2855 cm^{-1} et 2926 cm^{-1} représentant respectivement les vibrations d'élongation symétrique et asymétrique sont également visibles. Dans cette zone du spectre, on observe également une large bande entre 3100 et 3600 cm^{-1} correspondant aux vibrations d'élongation des fonctions OH.
- Les bandes amides I et II se caractérisent par deux pics de forte intensité à 1640 cm^{-1} et 1580 cm^{-1} . On observe cependant une déformation de ces deux bandes. Le maximum de la bande amide I se trouve à 1645 cm^{-1} contre 1630 cm^{-1} pour les immunoglobulines et celui de la bande amide II à 1536 cm^{-1} contre 1520 cm^{-1} .

- Le pic à 1447 cm^{-1} peut être attribué aux vibrations de déformation dans le plan (cisaillement) des CH_2 et des CH_3 . Les groupements carboxylates ainsi que les déformations des liaisons CH_3 peuvent être responsables de la bande d'absorption à 1395 cm^{-1} . Les liaisons amides provoquent également une bande d'absorption de faible intensité vers 1250 cm^{-1} . Les vibrations d'élongations des liaisons C-C se caractérisent par des pics à 1108 cm^{-1} et 938 cm^{-1} . Dans cette région, on observe cependant certaines différences dans les intensités des bandes d'absorbances. On voit que l'absorbance à 1305 cm^{-1} est plus importante pour l'albumine que pour les immunoglobulines. A l'inverse, les bandes d'absorption à 1235 cm^{-1} et 1075 cm^{-1} sont plus intenses pour les immunoglobulines que pour l'albumine.

Le spectre ATR-IR du glucose peut être divisé comme suit (Figure 22, orange) :

- Dans la première partie du spectre, on observe une large bande entre 3000 et 3600 cm^{-1} correspondant aux vibrations d'élongation des fonctions OH. Comme pour les spectres précédemment décrits, les vibrations d'élongation symétrique et asymétrique des groupements CH_2 des chaînes carbonées sont visibles à 2855 cm^{-1} et 2926 cm^{-1} .
- Dans la région de l'empreinte digitale, on observe une bande d'absorption à 1645 cm^{-1} qui peut être due à de l'eau résiduelle après séchage ainsi qu'aux vibrations d'agitation des liaisons O-H. On peut également voir 2 massifs d'absorption entre 1500 et 1180 cm^{-1} et entre 1180 et 940 cm^{-1} . Le premier de ces massifs peut être attribué aux vibrations de cisaillement des groupements CH_2 , d'élongation des liaisons C=O et C-C ainsi qu'aux vibrations d'agitation

des liaisons C-O. Le deuxième massif entre 1180 et 940 cm^{-1} peut être attribué aux vibrations d'élongation des groupements C-O et C-OH.

Le spectre ATR-IR de l'urée peut être divisé comme suit (Figure 22, vert):

- Dans la première partie du spectre, on observe une première bande avec un maximum d'absorption à 3440 cm^{-1} , une seconde à 3344 cm^{-1} et une à 3260 cm^{-1} . Ces trois bandes peuvent être attribuées aux vibrations d'élongations des liaisons N-H.
- Dans la région de l'empreinte digitale, on observe une bande d'absorption à 1678 cm^{-1} pouvant représenter les groupements carbonyle (C=O) de l'urée. Les liaisons C-N et C-C sont probablement responsables de la présence de deux bandes à 1630 cm^{-1} et 1600 cm^{-1} . Finalement, une dernière bande d'absorption est visible à 1466 cm^{-1} . Cette bande peut être attribuée aux vibrations d'élongation asymétrique des groupements CH_3 .

2) Caractérisation des signatures spectrales du plasma

L'analyse du plasma avec ATR-IR permet l'obtention de caractéristiques spectrales définies (tableau II). En effet, chaque pic spectral peut être affecté à une ou plusieurs catégories de biomolécules. En raison de la forte absorbance de l'eau, la plupart d'entre eux ne sont pas visibles sur les spectres liquides (Figure 23). Cependant, on peut observer les bandes amide I et amide II de protéines (1650 et 1550 cm^{-1}) et une faible déformation spectrale entre 1480 et 1000 cm^{-1} . Les spectres de plasma

humain séché présentent les caractéristiques spectrales attendues et les affectations associées au plasma humain (Tableau V).

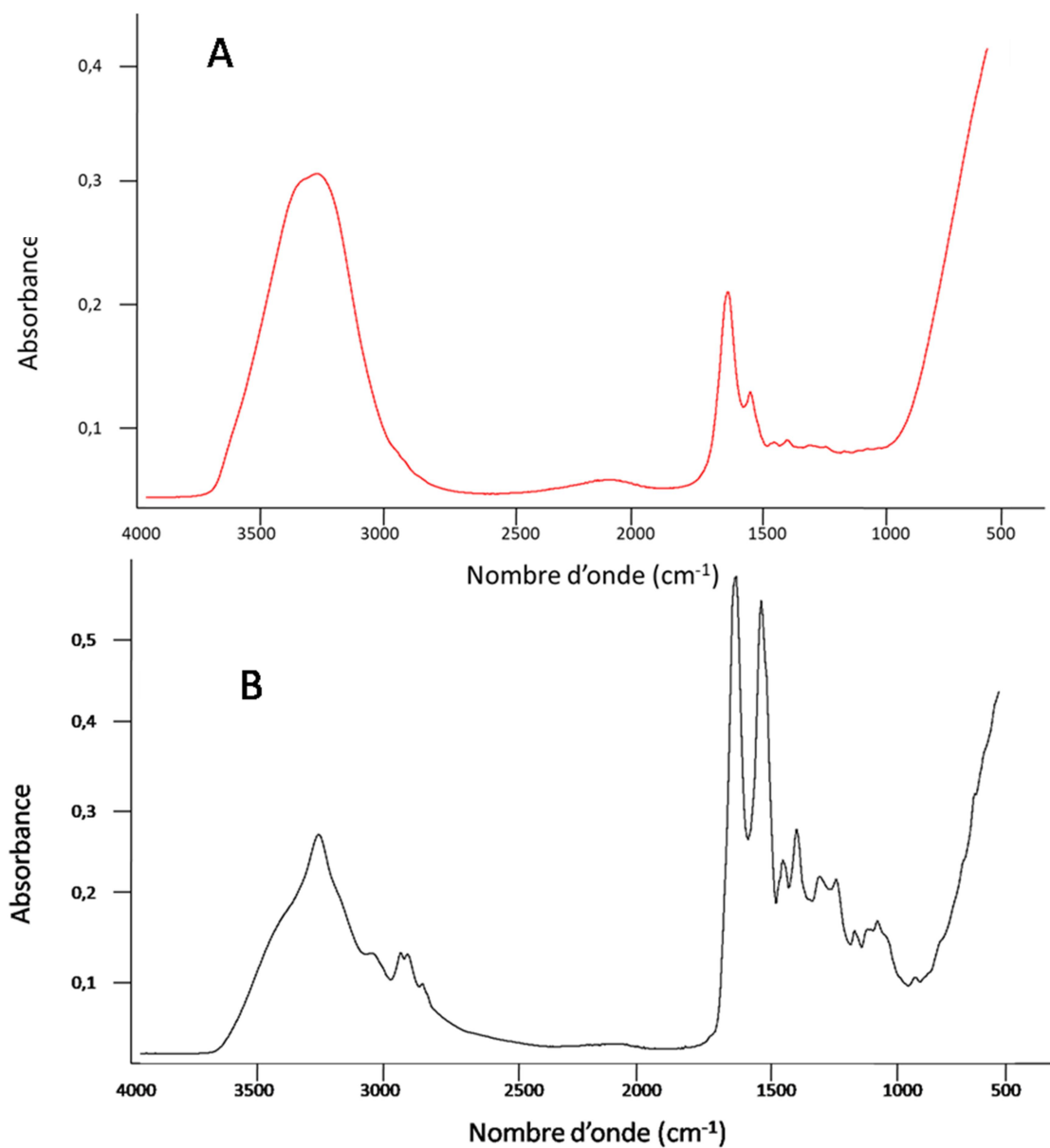


Figure 23 : Spectres ATR-IR obtenu à partir d'un plasma patient analysé sous forme liquide (A) et après séchage (B)

Tableau V : Attribution des signaux spectraux aux catégories moléculaires (186,200–203). ν : élongation, δ : cisaillement , γ : agitation, torsion et bascule, as: asymétrique, s: symétrique ; def : déformations

Nombre d'onde	Vibration	Attribution biochimique
3300	$\nu(\text{N-H})$	Amide A des protéines/peptides
3100	$\nu(\text{N-H})$	Amide B des protéines/peptides
2957	$\nu_{\text{as}}(\text{CH}_3)$	Lipides
2920	$\nu_{\text{as}}(\text{CH}_2)$	
2872	$\nu_{\text{s}}(\text{CH}_2)$	
2850	$\nu_{\text{s}}(\text{CH}_2)$	
1740	$\nu(\text{C=O})$	Esters de phospholipides
1715-1680	$\nu(\text{C=O})$	Acides nucléiques
1650	$\nu(\text{C=O})$	Amide I des protéines
1645	$\nu(\text{C-N}), \delta(\text{N-H})$	
1550	$\gamma(\text{HOH})$	Eau
1453	$\nu(\text{C-N}), \delta(\text{N-H})$	Amide II des protéines
1450	$\nu(\text{C-C}), \delta(\text{C-O})$	
1395	$\gamma(\text{CH}_2)$	Cisaillement CH ₂
1395	$\delta_{\text{as}}(\text{CH}_3)$	Lipides/protéines
1380	$\delta_{\text{s}}(\text{CH}_3)$	Lipides/protéines
1350-1250	$\nu(\text{C=O})$	Carboxylates COO ⁻
	$\gamma_{\text{s}}(\text{CH}_3)$	Phospholipides/triglycérides
	$\nu(\text{C-N}), \delta(\text{N-H})$	Amide III-
	$\nu(\text{C-C}), \gamma(\text{C-O})$	peptides/protéines/collagène
1242	$\nu_{\text{as}}(\text{PO}_2)$	ADN/ARN/phospholipides
1170	$\nu_{\text{as}}(\text{C-O})$	Esters
1150	$\nu(\text{C-O}), \nu(\text{C-OH})$	Sucres
1090	$\nu_{\text{s}}(\text{PO}_2)$	ADN/ARN/phospholipides
1086	$\nu(\text{C-O}), \nu(\text{C-C}),$	Sucres
1079	$\nu(\text{C-C}),$	Glycogène
1065	$\nu(\text{C-O}),$	ADN/ARN/ribose
1050	$\nu(\text{C-O}),$	Estersde phosphate
1028	def (CHO)	Glycogène
965	$\nu(\text{PO}_3)$	ADN/ARN/ribose
710-620	def (O=C-N)	Amide IV

La figure 23 présente les spectres obtenus par l'analyse d'un sérum sous forme liquide (A) ou après l'étape de séchage (B). Le spectre liquide est composé d'une large bande d'absorption entre 3800 et 2800 cm^{-1} pouvant être principalement associée aux liaisons O-H de l'eau, 2 bandes d'absorbances à 1650 et 1550 cm^{-1} reliées aux vibrations d'élongations des liaisons N-H des fonctions amides des protéines (Tableau IV).

En comparant le spectre du sérum séché (Figure 23) et celui de l'albumine et des immunoglobulines, on observe une grande ressemblance, signe que les protéines sont les composants de l'échantillon qui contribuent le plus à sa signature spectrale avec le pic amide I à 1650 cm^{-1} qui domine.

Ceci s'explique en partie par leur présence en concentration élevée par rapport aux autres composants de faible poids moléculaire (voir tableau IV).

2) Analyses quantitatives

a) Validation de la population utilisée :

Les concentrations de référence en protéines, albumine, immunoglobuline G, glucose, urée et acide urique ont été mesurées respectivement entre 36 et 93 g/L, 18 et 45 g/L, 4,88 et 42,96 g/L, 2,8 et 16,6 mmol/L, 1,09 et 43,42 mmol/L et 12 et 644 $\mu\text{mol/L}$ (Tableau V). Excepté pour l'albumine, ces différentes plages de concentrations recouvrent les valeurs usuelles retrouvées au niveau plasmatique de qui regroupent des populations saines et pathologiques couramment rencontrées en milieu hospitalier en routine. (197).

Une attention particulière a été donnée à l'absence de corrélation entre les différents marqueurs étudiés afin d'éviter les biais dans les résultats de la PLSR. Si les concentrations de tous les marqueurs évoluent avec les mêmes tendances, les

bandes spectrales identifiées ne seront pas spécifiques à un seul des composés et les résultats PLSR « optimistes ». La figure 24 présente des projections qui illustrent l'absence de corrélation entre les différents biomarqueurs ($p > 0,05$).

L'ensemble des résultats ci-dessus valident le lot d'échantillons pour l'étude.

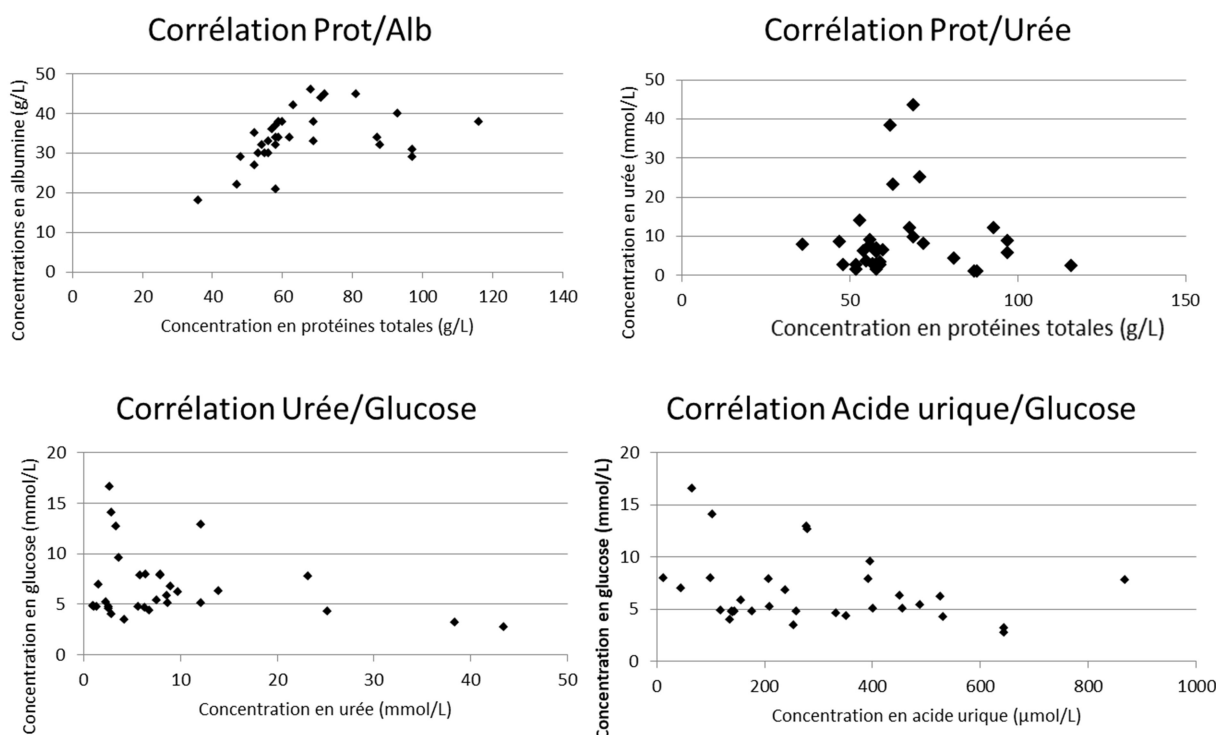


Figure 24 : Exemple de corrélations obtenues entre les concentrations de référence des différents métabolites étudiés

b) Quantification des protéines et des métabolites dans les échantillons de patients par ATR-IR sous forme liquide

La figure 25 présente d'une part les régressions linéaires entre les concentrations prédites (en ordonnées) par l'analyse par ATR-IR entre 1800 et 900 cm^{-1} sur les échantillons liquides (protéines totales (A) et glucose (C)) et les concentrations mesurées selon les méthodes de référence (en abscisse) entre 1800 and 900 cm^{-1} .

Une comparaison entre le coefficient de régression de la PLSR et du spectre de référence pour la prédiction des protéines totales et du glucose est également proposée. Les similarités entre les spectres sont surlignées en gris.

Analyse de la fraction protéique

L'analyse des protéines totales, de l'albumine et des immunoglobulines G a permis d'obtenir des coefficients de corrélation respectivement de 0,941, 0,923 et 0,873. Dans le tableau VI les valeurs de RMSECV en valeur absolue et en valeur relative sont données pour les différents analytes composants de la fraction protéique (protéines totales, albumine et Ig). Le modèle construit dans cette étude a permis de prédire les concentrations en protéines, en albumine et en immunoglobulines G avec des erreurs moyennes respectives de 4,55%, 3,87% et 33,7% (3,72 g/L et 1,68 g/L et 4,22 g/L).

Tableau VI : Erreurs moyennes des concentrations prédites par les modèles de régression partielle des moindres carrés en analyse liquide pour la fraction protéique

	Protéines totales	Albumine	Immunoglobuline G
RMSECV (g/L)	3.72	1.68	4.22
RMSECV (%)	4.55	3.87	33.7

De plus, la concentration la plus basse a été mal prédite par le modèle avec une erreur de 9 g/L ainsi que 2 échantillons pour lesquels la concentration sont de 53 et 54 g/L et pour lesquels les concentrations prédites sont surestimées de 5 g/L. La

comparaison entre les coefficients de régression et les spectres de référence a montré que les modèles prédictifs sont basés sur les bandes d'absorption spécifiques et non sur d'autres signaux interférents d'autres composants de la matrice (Figure 24). Les bandes d'absorption sur le coefficient de régression aux nombres d'onde de 1580 et 1640 cm^{-1} qui sont assimilables aux bandes amide I et amide II identiques à celles du spectre de référence de l'albumine. De plus, on observe d'autres similarités entre ces deux spectres entre 1500 et 1000 cm^{-1} .

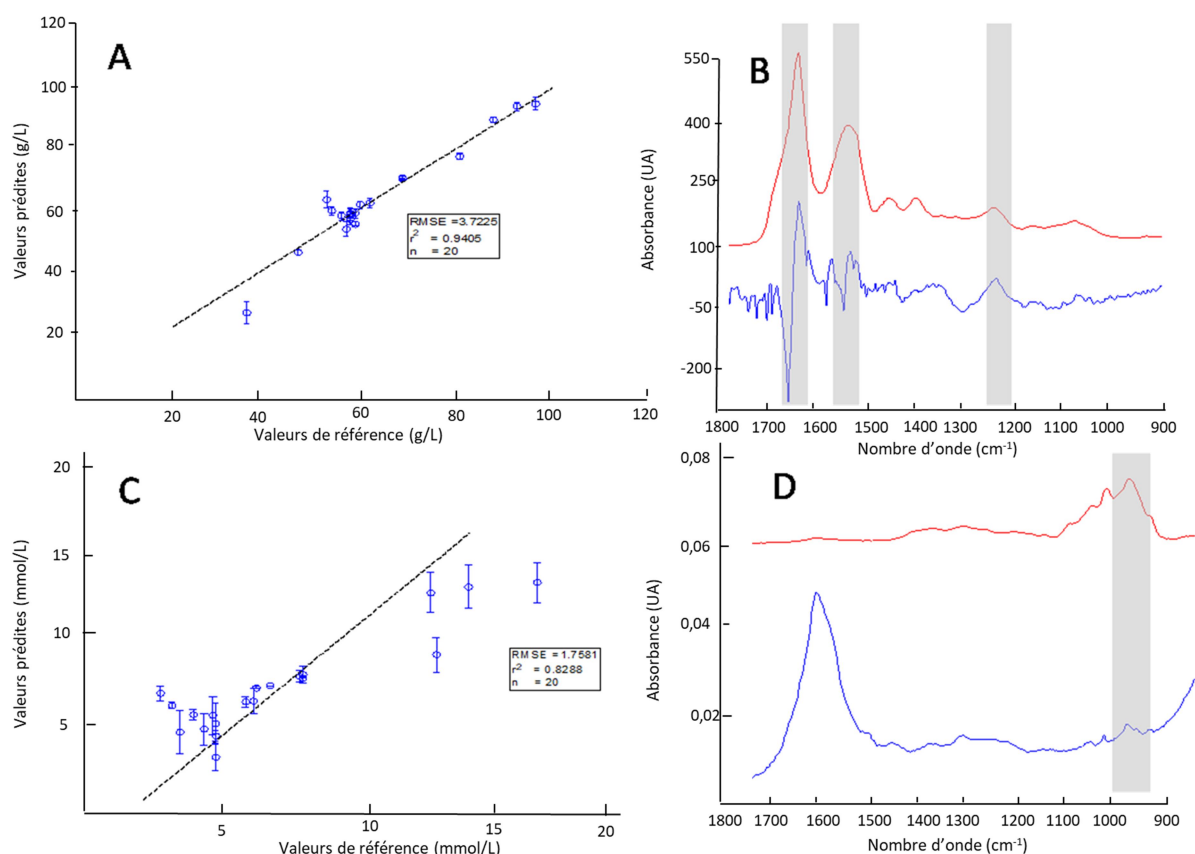


Figure 25 : A) Régression PLSR pour les protéines totales ; B) Coefficient de régression (bleu) comparé au spectre de référence de d'albumine (rouge) ; C) Régression PLSR pour le glucose ; B) Coefficient de régression (bleu) comparé au spectre de référence du glucose (rouge). .Pour chaque échantillon, le cercle représente la moyenne des concentrations prédites et la barre représente l'erreur moyenne absolue.

Analyse de la fraction de bas poids moléculaire

Dans cette étude, l'analyse de la fraction de bas poids moléculaire est réalisée via l'analyse du glucose, de l'urée et de l'acide urique. La prédiction des concentrations a été moins bonne avec des coefficients de corrélation respectifs de 0.8288 (Figure 25 C), 0.962 et 0.854. Concernant le glucose, les concentrations basses sont généralement surestimées alors que les valeurs fortes sont plutôt sous-estimées. Pour l'acide urique, les concentrations hautes sont généralement surestimées alors que les valeurs basses sont plutôt sous-estimées.

Contrairement à l'étude de la fraction protéique, les similitudes entre le spectre de référence du glucose et le coefficient de régression de la PLSR sont moins évidentes. La présence d'une bande d'absorption entre 1800 cm^{-1} et 1500 cm^{-1} qui ne correspond pas à l'absorbance du glucose. Ce signal indique la présence d'une interférence avec un autre composant de l'échantillon. Cette bande d'absorption ayant un maximum aux alentours de 1640 cm^{-1} , celle-ci peut être assimilée à la bande amide I des protéines ou à l'eau et indique probablement une influence d'au moins un de ces paramètres sur la quantification du glucose sous forme liquide. Néanmoins....on discerne le massif d'absorption entre 1180 et 940 cm^{-1} du glucose tel que décrit précédemment (Figure 25).

Dans le tableau VII les valeurs de RMSECV en valeur absolue et en pourcentage sont données pour les différents analytes composants de la fraction de bas poids moléculaire (urée, glucose, acide urique). A l'inverse des protéines, les petites molécules comme le glucose, l'urée et l'acide urique sont mal prédites avec des erreurs moyennes respectives de 23,7%, 47,7% et 50,4% (Tableau VII). Selon les recommandations SFBC, les erreurs totales admissibles en biologie pour ces

paramètres sont respectivement de 5,5%, 15,55% et 11.97% pour une analyse ciblée quantitative (204). Pour l'ensemble de ces paramètres, les erreurs moyennes calculées sont largement supérieures aux recommandations. Cependant, concernant l'urée, nous observons des erreurs respectives de 549,9%, 44,3%, 28,99% et 128,4% pour les échantillons les moins concentrés (respectivement 1,04 mmol/ L, 1,41 mmol/L, 2,68 mmol/L et 3,34 mmol/L). Ces résultats démontrent une incapacité de la méthode à prédire correctement ces faibles concentrations. Lorsque nous retirons ces échantillons du modèle, l'erreur moyenne calculée diminue fortement à 9,8%. Cela suggère que la limite basse de quantification de l'urée dans cette condition pourrait être proche de 4 mmol/L. De plus dans ces conditions, les erreurs moyennes mesurées sont inférieures à l'erreur totale admissible. Toutefois, les valeurs physiologiques de l'urée étant comprises entre 2,5 et 7,5 mmol/L (197), cette méthode ne serait en mesure que de mettre en évidence une hyperurémie et ne permettrait pas un suivi fin dans les valeurs physiologiques ou inférieures.

Tableau VII : Erreurs moyennes des concentrations prédites par les modèles de régression partielle des moindres carrés en analyse liquide la fraction de bas poids moléculaire

	Glucose	Urée	Acide urique
RMSE	1.76 mmol/L	2.17 mmol/L	72 µmol/L
RMSECV (%)	23.7	47.7	50.4

c) Quantification des protéines et des métabolites dans les échantillons de patients par ATR-IR après séchage

La figure 26 présente d'une part les régressions linéaires entre les concentrations prédites (en ordonnées) par l'analyse par ATR-IR entre 1800 and 900 cm^{-1} sur les échantillons patients après séchage (protéines totales (A) et glucose (C)) et les concentrations mesurées selon les méthodes de référence (en abscisse). On peut également observer grâce, à la comparaison entre les coefficients de régression et les spectres de référence, que les modèles prédictifs sont basés sur les bandes d'absorption spécifiques et non sur d'autres signaux interférents d'autres composants de la matrice. Les similarités entre les spectres sont surlignées en gris.

Analyse de la fraction protéique

L'analyse des protéines totales a permis d'obtenir un coefficient de corrélation de 0,9751 avec une erreur moyenne de 2,46 g/L. En outre, la comparaison entre la composante principale et le spectre de référence a montré que les modèles prédictifs sont basés sur les bandes d'absorption spécifiques et non sur d'autres signaux interférents d'autres composants de la matrice (Figure 26). Comme pour l'analyse liquide, les modèles PLS sont basés sur des bandes d'absorption spécifiques (1645 cm^{-1} , 1535 cm^{-1} et 1245 cm^{-1}) (Figure 26). Les similarités observées sont beaucoup plus marquées que celles observées lors de l'analyse sous forme liquide.

Dans le tableau VIII, les valeurs de RMSECV en valeur absolue et en pourcentage sont données pour les différents analytes composants de la fraction protéique (protéines totales, albumine et Ig). Le modèle construit dans cette étude a permis de prédire les concentrations en protéines et en albumine avec des erreurs moyennes

respectives de 4,55% et 3,87% (3,72 g/L et 1,68 g/L). Les erreurs moyennes calculées de prédiction des concentrations en protéines totales, albumine et immunoglobulines G sont respectivement de 2,46%, 2,95% et 15,7%. Ces résultats confirment le pouvoir prédictif élevé de l'ATR pour les protéines. On observe une meilleure justesse des concentrations prédites concernant les protéines totales (2,46% vs 4,55%), l'albumine (2,95% vs 3,87%) et les immunoglobulines G (15,7% vs 33,7%) (Tableau VIII). Ces valeurs sont inférieures à celles recommandées pour la quantification de ces analytes excepté les immunoglobulines G, pour lesquelles, malgré l'amélioration de la prédiction, l'erreur de mesure reste largement supérieure aux recommandations (8,0%).

Tableau VIII : Erreurs moyennes des concentrations prédites par les modèles de régression partielle des moindres carrés en analyse sous forme séchée pour la fraction protéique

	Protéines	Albumine	Immunoglobuline G
RMSE (g/L)	2.46	1.69	3.12
RMSECV (%)	2.46	2.95	15.7

Fraction de bas poids moléculaire

Comme pour la fraction protéique, nous observons une meilleure corrélation avec un coefficient de corrélation de 0,9609 pour le glucose, 0,967 pour l'urée et 0,900 pour l'acide urique. De la même manière que pour l'étude de la fraction protéique, on observe des similitudes entre le spectre de référence du glucose et le spectre vecteur du modèle PLSR. En effet, on discerne le massif d'absorption entre 1180 et 940 cm^{-1} du glucose tel que décrit précédemment (Figure 26). Par rapport aux

modèles liquides, nous avons observé moins de variations spectrales non spécifiques sur le spectre de la composante principale.

Dans le tableau IX, les valeurs de RMSECV en valeur absolue et en pourcentage sont données pour les différents analytes composants de la fraction de bas poids moléculaire (urée, glucose, acide urique). Les acquisitions réalisées après séchage permettent d'améliorer les concentrations prédites en glucose, urée et acide urique. Les erreurs moyennes calculées sont respectivement de 6,89%, 26,55% et 56,0% (Tableau IX). Pour l'ensemble de ces paramètres, les erreurs moyennes calculées restent supérieurs aux recommandations (204). Comme pour l'analyse sous forme liquide, nous observons des erreurs de prédiction particulièrement importantes pour les échantillons les moins concentrés en urée (152% pour l'échantillon contenant 1,09 mmol/L d'urée et 251% pour celui contenant 1,41 mmol/L d'urée).

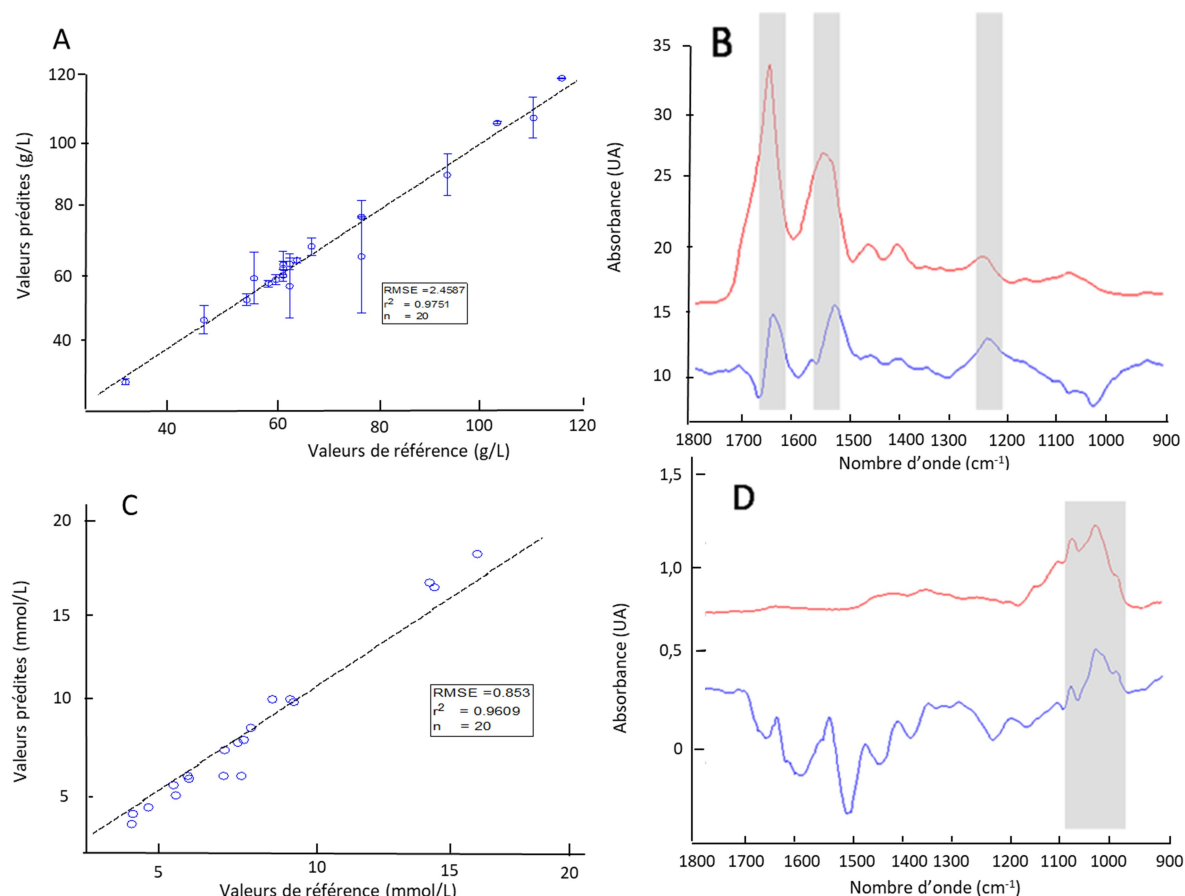


Figure 26 : A) Régression PLSR pour les protéines ; B) Coefficient de régression (bleu) comparé au spectre de référence de d'albumine (rouge) ; C) Régression PLSR pour le glucose ; B) Coefficient de régression (bleu) comparé au spectre de référence du glucose (rouge). Pour chaque échantillon, le cercle représente la moyenne des concentrations prédites et la barre représente l'erreur moyenne absolue.

Ces résultats suggèrent une incapacité de la méthode à prédire correctement ces faibles concentrations. Lorsque nous retirons ces échantillons du modèle, l'erreur moyenne calculée diminue fortement à 7,1%. Cela suggère que la limite basse de quantification de l'urée dans cette condition pourrait être entre 1,5 et 2,5 mmol/L. De plus dans ces conditions, les erreurs moyennes mesurées sont inférieures à l'erreur totale admissible.

Tableau IX : Erreurs moyennes des concentrations prédites par les modèles de régression partielle des moindres carrés en analyse sous forme séchée pour la fraction de bas poids moléculaire

	Glucose	Urée	Acide urique
RMSE	0.85mmol/L	2.27mmol/L	58 µmol/L
RMSECV (%)	6.89	26.55	56.0

d) Comparaison entre l'analyse sous forme liquide et l'analyse après séchage

L'analyse du plasma après séchage montre des résultats intéressants concernant la quantification des protéines totales et de l'albumine. Cependant, nous avons également observé les limites de cette analyse pour détecter des variations faibles dans le spectre infrarouge tel que le décalage de la bande amide I pour l'analyse des immunoglobulines G (Figure 26). En effet les déformations spectrales au niveau de la bande amide I des protéines dues aux immunoglobulines G sont probablement masquées par l'intensité de la bande d'absorption de l'eau entre 1800 cm^{-1} et 1500 cm^{-1} . L'analyse après séchage a permis d'améliorer la prédiction des protéines totales, de l'albumine et des immunoglobulines G. Par exemple, celle des protéines totales diminue de 2,45 à 3,72 g/L. L'élimination de l'eau et l'intensification des signaux propres aux protéines qui en a découlé est probablement à l'origine de ces résultats. Les erreurs de prédictions observées avec les protéines totales et l'albumine sont compatibles avec des analyses quantitatives. En revanche, celle calculée à partir des concentrations prédites en immunoglobulines G reste trop importante pour ce type d'examen. Nous supposons que cette difficulté à prédire correctement les concentrations en immunoglobulines G peut venir d'une interférence avec les autres protéines telles que l'albumine. En effet, les différences

entre les deux spectres infrarouges sont minimales (Figure 22). Barth et al ont démontré que cette différence est due à une composition différente des protéines en feuillets bêta et en hélice alpha (205). Le plasma étant plus concentré en albumine qu'en immunoglobuline, il est possible que le décalage de la bande amide ne soit pas suffisant pour être détecté.

Concernant l'analyse de la fraction de bas poids moléculaire, les modèles de prédictions réalisés à partir des spectres obtenus sur l'échantillon sous forme après séchage ont permis de mettre en évidence une erreur moyenne de mesure plus faible suggérant une meilleure prédiction avec moins de variations spectrales non spécifiques.

Par exemple, celle du glucose a diminué de 1,76 à 0,85 mmol/L. De plus les résultats obtenus sur l'urée suggèrent une limite basse de quantification entre 1,5 et 2,5 mmol/L après séchage au lieu d'environ 4 mmol/L sous forme liquide. Ceci est en faveur d'une meilleure sensibilité de la spectroscopie ATR-IR sur dépôt séché.

DISCUSSION

1) Performances actuelles de l'analyse par ATR-IR

Bien que l'ensemble de ces résultats soient en défaveur de la capacité de l'ATR-IR à quantifier l'ensemble des composants dans des échantillons complexes tels que le plasma dans ces conditions, les imprécisions de mesures sont relativement faibles (de l'ordre de 15%). De plus le protocole analytique utilisé est composé de plusieurs étapes pour lesquels certains paramètres ne sont pas maîtrisés et qui augmentent la variabilité de la mesure. D'une part, 2µL d'échantillon est déposé manuellement sur

le cristal de mesure à l'aide d'une micropipette. Nous nous sommes rendu compte qu'entre les dépôts, la forme de la goutte n'est pas toujours identique malgré un volume supposé constant. Lorsqu'une goutte est déposée sur un solide, celle-ci va soit s'étaler, soit former une lentille plus ou moins bombée selon les tensions superficielles qui s'y appliquent. Ces tensions superficielles vont dépendre des trois phases en contact : le liquide (l'échantillon), le solide (le cristal) et le gaz (l'air de la pièce). Or cette tension superficielle varie en fonction de la température, de la densité du liquide et des constituants d l'échantillon. D'autre part, le taux d'humidité de l'air influence également la mouillabilité du cristal ATR et donc la forme de la goutte (Figure 27). Dans notre cas, la densité de l'échantillon liquide déposé dépend directement de sa composition notamment en protéines. Les paramètres de température et d'humidité de la pièce n'étant pas régulés et donc pas maîtrisés, ceux-ci variaient fortement au cours de la journée ou entre les différentes journées d'acquisitions. D'après la loi de Beer-Lambert qui s'applique à l'absorption infrarouge, la mesure de l'absorbance dépend directement de l'épaisseur de l'échantillon. Or cette variabilité de la forme de la goutte a pour conséquence une variation de l'épaisseur du film ainsi formé après séchage, ce qui impacte directement les absorbances mesurées lors de l'acquisition du spectre ATR-IR. Une automatisation de ces étapes de dépôts et séchage en maîtrisant ces différents paramètres ou l'utilisation d'un étalon interne qui permettrait de corriger cette variabilité semblent être nécessaires pour la quantification de biomarqueurs par ATR-IR.

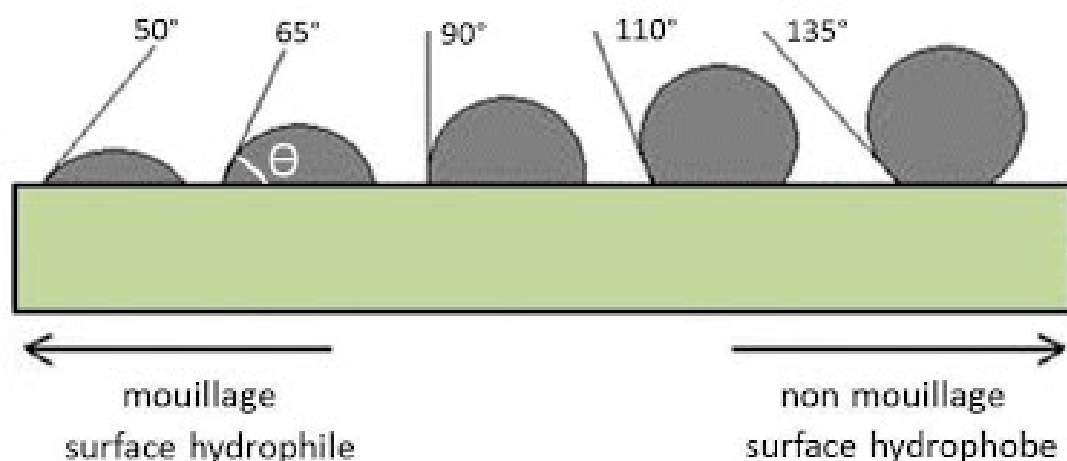


Figure 27: Impact du mouillage de la zone de mesure sur la forme de la goutte et sur la surface de dépôt

2) L'infrarouge comme outil diagnostique en biologie

La Société Française de Biologie Clinique (SFBC) publie régulièrement des recommandations concernant les performances des méthodes analytiques pour chaque biomarqueur comme les protéines totales, l'albumine, l'immunoglobuline G, le glucose, l'urée et le glucose (197). Malheureusement, nous n'avons pas pu tester notre protocole sur des échantillons très faibles ou très concentrés pour chaque biomarqueur comme l'exige la société de biologie clinique pour une analyse spécifique. Néanmoins, les bons résultats trouvés sur ces pages de mesure conduisent à une interprétation biologique sans ambiguïté. Concernant l'analyse des différentes protéines étudiées et du glucose, malgré l'imprécision de la mesure, chaque échantillon analysé prédit conduit à une interprétation biologique correcte. De plus, vis-à-vis de l'urée, seuls quatre échantillons ne respectent pas les recommandations biologiques appliquées à des analyses spécifiques pour une utilisation quotidienne dans un laboratoire accrédité. Malgré tout, les performances

analytiques dans ces conditions ne permettent pas d'utiliser cet outil pour la quantification de ces biomarqueurs dans un laboratoire d'analyse médicale (197). Cependant, ces résultats démontrent le potentiel de l'ATR-IR pour une analyse semi-quantitative qui peut être utilisée comme outil de dépistage. En effet, nous avons démontré que la capacité à classer les concentrations de protéines, d'immunoglobuline G, de glucose, et d'urée en concentrations physiologiques, anormalement augmentées ou anormalement abaissées. De plus, l'analyse ATR-IR est capable d'obtenir ces résultats semi-quantitatifs pour plusieurs biomarqueurs en une seule analyse par opposition au processus actuel de laboratoire. Cela positionne l'ATR-IR comme outil potentiel dans la réalisation d'un dépistage biochimique. En effet, cette technologie pourrait être utilisée pour réaliser un bilan d'entrée en hospitalisation et obtenir un premier résultat en quelques minutes et à moindre coût. De plus, mis à part le manque de sensibilité mis en évidence sur l'urée dont la limite basse de quantification est estimée entre 1,5 et 2,5mmol/L après séchage, les résultats obtenus en termes de variabilité est en faveur de l'utilisation de ce type d'outil pour une approche métabolomique pour lesquelles les limites classiquement utilisées sont de 30%.

Ajouté à cela, ce résultat permet d'envisager la quantification de ces biomarqueurs en corrigeant la variabilité en maîtrisant les paramètres environnementaux durant l'étape de séchage.

IV) Etude n°4 : Impact de l'ultrafiltration centrifuge sur la composition qualitative et quantitative du plasma humain

INTRODUCTION

En raison de la composition complexe du plasma (qualitative et quantitative), la quantification d'un biomarqueur de bas poids moléculaire reste difficile en ATR-IR sans une étape pré-analytique de préparation des échantillons. Dans ce travail, nous avons montré que même si l'ATR-IR est capable d'estimer les concentrations de protéines totales et d'albumine dans le plasma sous forme liquide, la contribution importante de l'eau interfère avec la détermination des concentrations de glucose, d'urée et d'acide urique. A titre d'exemple, les concentrations de glucose sont difficiles à estimer sous forme liquide avec une erreur moyenne de 23,7%. Cependant, les représentations spectrales des coefficients de régression ont montré que les modèles PLSR sont capables d'extraire des variations spectrales spécifiques à partir d'échantillons sous forme liquide. De plus, après séchage et élimination de ce signal dû à l'eau, la justesse de mesure est fortement améliorée et permet d'obtenir une erreur moyenne de 6,89%. En revanche, les prédictions des concentrations d'acide urique par les modèles PLSR conservent une erreur moyenne importante de l'ordre de 58 $\mu\text{mol/L}$ pour la plage habituelle entre 150 et 500 $\mu\text{mol/L}$ même après séchage à l'air. De plus, les échantillons les moins concentrés en urée sont très mal prédits et ont été exclus du modèle pour en améliorer les performances. Ces résultats suggèrent que les biomarqueurs faiblement concentrés ont une faible contribution dans le spectre plasmatique global ou que les signaux spécifiques de

ces biomarqueurs sont complètement masqués par celui des composants majoritaires de la matrice tel que les protéines. La superposition des spectres d'absorbance de l'urée et de celui des protéines montre que les bandes d'absorption de l'urée entre 1700 cm^{-1} et 1600 cm^{-1} sont masquées par la bande amide I des protéines, de la même manière que la forte contribution de l'eau dans le spectre infrarouge masquait les signaux des immunoglobulines G et du glucose lors de l'analyse du plasma sous forme liquide. De ce fait, éliminer les protéines de l'échantillon est une étape qui pourrait améliorer la détection et la quantification des composants de la fraction plasmatique de bas poids moléculaire. Nous suggérons donc une étape de déplétion des protéines avant de réaliser l'analyse pour une meilleure quantification des petites molécules.

Notre hypothèse est que l'élimination des protéines de l'échantillon pourrait améliorer la sensibilité et la spécificité de l'infrarouge pour ces composés minoritaires. A cet effet, différents modes de préparations d'échantillons physiques et chimiques ont été proposées dans la littérature pour éliminer les protéines et améliorer la précision et l'exactitude. Bonnier et al (199) ont utilisé l'ultrafiltration à l'aide de filtres de 10KDa pour isoler la fraction de faible poids moléculaire. Ils ont montré que la précision et l'exactitude des modèles quantitatifs construits à partir de l'analyse PLS étaient fortement améliorées après ce prétraitement. Cependant, cette démonstration n'a été faite que sur le glucose et devrait être étendue à différents métabolites. D'autre part, Kuligowski et al (206) ont utilisé des réactifs chimiques pour précipiter les protéines. Malheureusement, cette étude a montré que l'acétonitrile, le méthanol ou l'acétone interfèrent avec l'interprétation des variations spectrales attribuées aux biomolécules du sérum. De plus, ces modifications ont été confirmées sur LC-MSMS suggérant que ce prétraitement élimine les protéines de l'échantillon mais modifie également la

concentration d'autres métabolites. Par exemple, la plupart des concentrations d'acides aminés, de monosaccharides, d'imidazopyrimidines et d'indoles identifiés étaient plus faibles sur les échantillons prétraités à l'acétone, tandis que les intensités d'acides gras et de glycérophospholipides étaient plus élevées. Des travaux précédemment publiés sont en faveur d'une séparation physique qui pourrait induire moins de modification de la composition de la fraction de bas poids moléculaire du plasma que la précipitation chimique des protéines par l'utilisation de solvants organiques. L'objectif de cette étude est donc d'évaluer les modifications de la composition du plasma suite à une étape de préparation par ultrafiltration centrifuge. Nous avons également comparé ce protocole de déprotéinisation physique à un protocole chimique à l'acide sulfosalicylique pas encore publié à notre connaissance.

MATERIEL ET METHODES

1) Echantillons

Un pool de plasma a été constitué à partir d'échantillons sanguins de patients recueillis sur des tubes héparinés au sein du laboratoire de biochimie et biologie moléculaire du CHU de Tours dans le cadre de leur prise en charge. Pour la première étape d'optimisation de l'étape de filtration, nous avons préparé 35 aliquots identiques pour comparer les différentes conditions testées.

Dans un second temps, nous avons également recueillis 8 échantillons sanguins de patients qui ont été centrifugés 3 minutes à 10 000g. Les plasmas ont été récupérés, aliquotés en deux exemplaires puis congelés à -20°C jusqu'à analyse.

I.2) Elimination de la fraction protéique

I.2.a) Ultrafiltration centrifuge

Dans cette étude, nous avons testé différents dispositifs de filtration centrifuges munies de différents types de membrane (Tableau X) pour fractionner le plasma (Figure 28) :

- filtres Amicon Ultra-0,5 ml (Millipore Merck, Allemagne), avec un seuil à 30 kDa
- filtres HY-2mL (Sartorius, France) avec un seuil à 20 kDa
- filtres PES-2mL (Sartorius, France) avec un seuil à 20 kDa
- filtres CTA-2mL (Sartorius, France) avec un seuil à 20 kDa
- filtres Centrisart-2,5mL (Sartorius, France) avec un seuil à 20 kDa

Les filtres Amicon, HY, PES et CTA permettent l'ultrafiltration des plasmas selon le sens de la force centrifuge. Dans un premier temps, l'échantillon est déposé dans la partie supérieure du filtre. Ensuite, l'ultracentrifugation permet le passage de l'échantillon à travers le filtre. Le filtrat est recueilli dans le tube collecteur (Figure 28 A).

Le filtre Centrisart fonctionne différemment. En effet, l'échantillon est déposé dans le tube puis l'insert est rajouté (Figure 28 B). La centrifugation permet le passage de l'échantillon au travers du filtre dans le sens opposé à la force centrifuge. Le filtrat est recueilli dans l'insert.

Tableau X : Caractéristiques des filtres utilisés pour l'ultrafiltration lors de l'étude

	<u>Type de membrane</u>	<u>Direction de l'ultrafiltration</u> (Figure 28)	<u>Volume filtré</u>	<u>Temps et vitesse de centrifugation</u>
<u>Amicon</u>	Cellulose régénérée	Sens de la force centrifuge	0,5mL	15min, 12000g
<u>HY</u>	Cellulose régénérée	Sens de la force centrifuge	2mL	15min, 12000g
<u>PES</u>	Polyethersulfone	Sens de la force centrifuge	2mL	15min, 12000g
<u>CTA</u>	Cellulose triacétate	Sens de la force centrifuge	2mL	15min, 12000g
<u>Centrisart</u>	Cellulose triacétate	Sens inverse de la force centrifuge	2,5mL	15min, 12000g

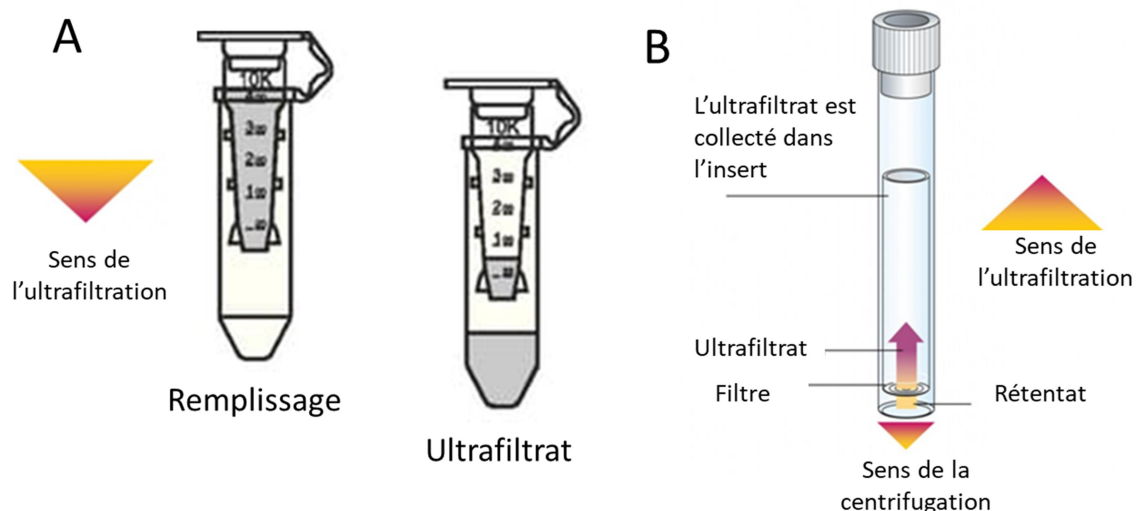


Figure 28 : Schéma représentant les différents systèmes d'ultrafiltration

L'étape de filtration a été réalisée selon les paramètres recommandés par les fournisseurs qui sont recensées dans le tableau X.

I.2.b) Déprotéinisation chimique

Une solution d'acide sulfosalicylique à 10% dans de l'eau ultrapure à l'aide d'acide sulfosalicylique acheté chez Sigma-Aldrich (Saint-Louis, Missouri, États-Unis). 50µL de cette solution sont ajoutés à 500µL de plasma. L'échantillon a ensuite été vortexé pendant 1 minute puis centrifugé 5 minutes à 10000g. Le surnageant a ensuite été transvasé dans un tube secondaire et congelé à -20°C jusqu'à analyse.

I.3) Analyse par des méthodes de référence

Pour évaluer le comportement du filtre vis-à-vis de l'ensemble des composants du plasma, nous avons sélectionné des molécules avec des masses allant de 90 g/mol à 155 kDa (155 000 g/mol). Pour cela nous avons mesuré les concentrations des

molécules suivantes : l'acide urique, l'acide lactique, la bilirubine, la créatinine, le glucose, l'urée, l'albumine, la ferritine, l'haptoglobine, l'insuline, les immunoglobulines G, le NT-proBNP, le peptide C, la préalbumine, la protéine C réactive, les protéines totales, la transferrine, la troponine, la bêta2-microglobuline ainsi que les activités enzymatiques de l'amylase, de la créatine kinase, de la lactate déshydrogénase, et de la lipase ont été quantifiées sur l'ensemble des échantillons préparés tel que décrit précédemment ainsi que sur 5 échantillons de sérum par des méthodes validées et accréditées sur Cobas c501® (Roche, Bâle, Suisse).

Tableau XI : Masse molaire et concentrations des biomarqueurs mesurées par méthodes de référence

Biomarqueur	Masse molaire	Concentration/Activité mesurée avant déprotéinisation
Urée	60,06 g/mol	4,0 mmol/L
Acide lactique	90,08 g/mol	8,3 mmol/L
Créatinine	113,12 g/mol	84 µmol/L
Acide urique	168,11 g/mol	352 µmol/L
Glucose	180,16 g/mol	4,0 mmol/L
Bilirubine	584,66 g/mol	5,8 µmol/L
Peptide C	3 kDa	0,30 nM
Insuline	5,8 kDa	320 mUI/L
NT-proBNP	8,5 kDa	19,6 ng/L
Béta2-microglobuline	11,8 kDa	1,7 mg/L
Ferritine	27,5 kDa	129 µg/L
Troponine T	30,5 kDa	3,2 ng/L
Préalbumine	35 kDa	0,27 g/L
Haptoglobine	45 kDa	0,96 g/L
Lipase	50 kDa	29,6 UI/L
Amylase	58 kDa	106 UI/L
Albumine	65 kDa	48 g/L
Transferrine	77 kDa	2,9 g/L
Créatine kinase	85 kDa	141 UI/L
Protéine C réactive (CRP)	115 kDa	1,5 mg/L
Lactate déshydrogénase	140 kDa	133 UI/L
Immunoglobulines G	155 kDa	13,7 g/L

I.4) Analyse par LCMSMS

Seize échantillons de plasma de patients ont été séparés en deux aliquots. Selon les résultats obtenus lors de la première étape de cette étude, nous avons éliminé la fraction protéique d'un aliquot en le centrifugeant pendant 10 minutes à 12000g à travers le filtre Centrisart®. Le surnageant a ensuite été transvasé puis conservé à -20°C jusqu'à la réalisation de l'analyse. Les échantillons (16 filtrats et 16 plasmas)

ont ensuite été analysés en LC-MSMS à l'aide du kit AbsoluteIDQ® p180 (Biocrates Life Science AG, Innsbruck, Autriche). Brièvement, les échantillons ainsi que des échantillons de calibration et de contrôle sont déposés sur une plaque munie de filtres dans chaque puit. Après séchage sous azote, les échantillons sont dérivés au phénylthiocyanate (PTC). Après une incubation de 30 minutes, les échantillons sont séchés sous azote puis élués dans une plaque de collection. Les échantillons sont ensuite injectés sur un système LC-MSMS de type TQS (Waters, Milford, Massachussets, Etats-Unis). L'analyse par LC-MSMS permet l'obtention des concentrations de 180 métabolites dont des acides aminés, des amines biogènes, des lipides et dérivées nucléotidiques.

I.5) Analyse par ATR-IR

L'analyse par ATR-IR de l'ensemble des échantillons étudiés a été réalisée tel que décrit précédemment. Brièvement, 2µL de plasma ou de filtrat a été déposé sur le cristal ATR en diamant. Après séchage à l'air ambiant, nous avons réalisé les acquisitions spectrales de 3 dépôts entre 4000cm⁻¹ et 400 cm⁻¹.

I.6) Analyses statistiques

Pour chaque protocole d'élimination des protéines, 5 échantillons ont été préparés et analysés. Les valeurs mesurées sur le sérum sont considérées comme la valeur de référence. Nous avons calculé les pourcentages de recouvrement moyen de cette valeur de référence ainsi que les coefficients de variations pour chacun des autres modes de préparation d'échantillons. Le protocole choisi est celui qui élimine la totalité des protéines (recouvrement = 0%), avec la meilleure robustesse. Le dernier

critère de choix est le meilleur recouvrement sur les autres paramètres (le glucose, l'acide urique, l'urée, le lactate, et la créatinine).

Pour les analyses par LC-MSMS, les concentrations ont été calculées à l'aide du logiciel MetIDQ®. Pour les analyses par spectroscopie infrarouge, une matrice contenant l'ensemble des intensités de chaque échantillon pour chaque nombre d'onde a été réalisée à l'aide du logiciel Matlab®.

Les données recueillies en spectroscopie IR et par LC-MSMS ont été analysées par une approche multivariée en utilisant Simca P + version 13.0 (Umetrics, Umeå, Suède). Des analyses multivariées par composante principale ou discriminantes par régression des moindres carrés partiels (OPLS-DA ou PLS-DA) ont ensuite été réalisées. L'OPLS-DA identifie les variations des intensités ou des concentrations en métabolites entre les groupes : la variation des données mesurées a été partitionnée en deux blocs par le programme, l'un contenant des variations qui correspondent à l'identifiant de classe et l'autre contenant des variations orthogonales au premier bloc et donc non contributives à la discrimination de groupe.

RESULTATS

a) Elimination des protéines

L'ensemble des pourcentages de recouvrement calculés dans les différents échantillons sont représentées dans la figure 29. Quel que soit le protocole utilisé, la totalité des protéines ayant une masse molaire supérieure ou égale à 40kDa a été complètement éliminée de l'échantillon. L'ensemble des filtres ayant un seuil à 20kDa (HY, CTA, PES, Centrisart) se comportent de la même façon vis-à-vis de l'ensemble des paramètres. Ils permettent d'éliminer la majorité des protéines testées de l'échantillon dont l'insuline, la β 2-microglobuline et la préalbumine de

masses molaires respectives 5,8kDa, 11,8kDa et 16kDa. A l'inverse, entre 10 et 30% de certaines protéines de masse molaire plus faible comme le peptide C (3kDa) ou le NT-proBNP (3,5kDa) sont conservés dans le filtrat.

Le filtre Amicon a un seuil de 30kDa, ce qui est supérieur à celui des autres filtres étudiés. En utilisant ce filtre, les protéines de masse molaire supérieure ou égale à 16kDa ont été retenues en totalité dans le concentrat. Celles ayant une masse molaire intermédiaire tel que le peptide C, le NT-proBNP, l'insuline et la préalbumine sont restées quantifiables dans le filtrat avec également des concentrations supérieures à celles mesurées dans les autres filtrats pour le peptide C et le NT-proBNP.

La précipitation des protéines en condition acide a permis d'éliminer l'ensemble des protéines excepté la ferritine, le peptide C et le NT-proBNP qui sont restées quantifiables.

Au final, l'élimination des protéines de notre filtrat semble moins bonne en utilisant le filtre Amicon. Cependant, les seuils de filtration n'étant pas les mêmes entre les différents produits, il est difficile de conclure quant à l'impact de la chimie de membrane.

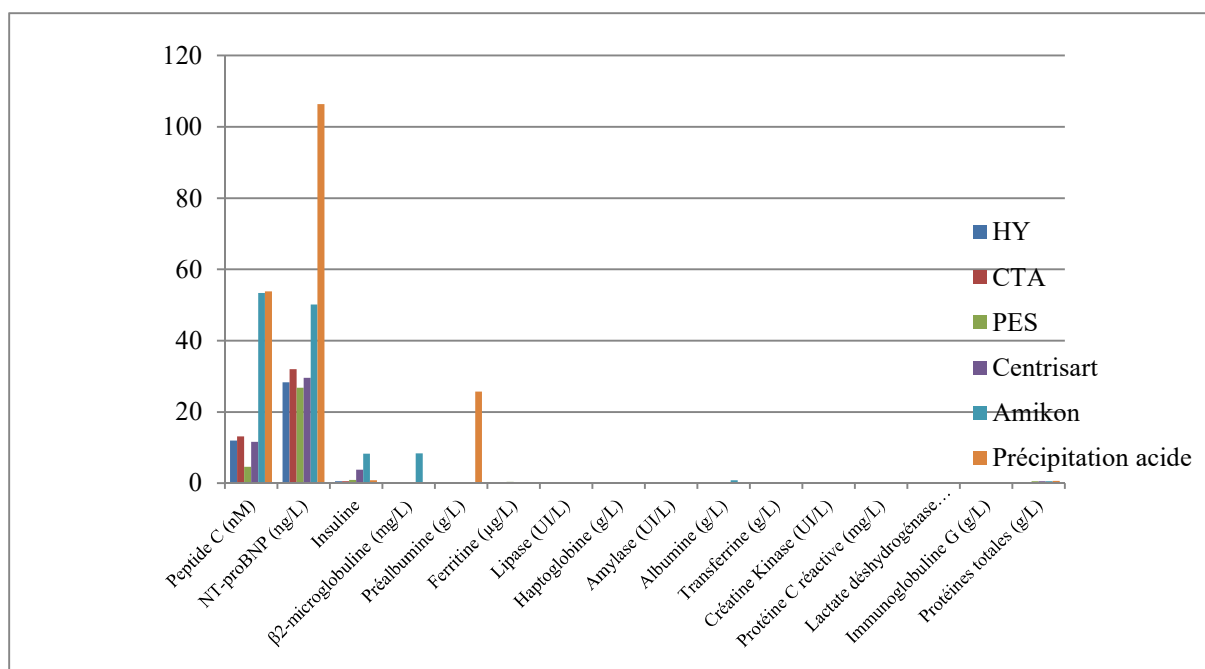


Figure 29 : Histogramme représentant les pourcentages de recouvrement des composants de la fraction protéique pour chaque protocole d'élimination des protéines

b) Absence de modifications de la fraction de bas poids moléculaire

L'ensemble des pourcentages de recouvrement calculés dans les différents échantillons sont représentées dans la figure 30. Pour l'ensemble des protocoles de préparation utilisés, les molécules de faible poids moléculaire sont conservées dans le filtrat avec un recouvrement proche de 100% à l'exception de l'acide urique pour lequel le recouvrement est de 80%. Les recouvrements observés des molécules de bas poids moléculaire avec le filtre Amikon est supérieur à celui observé avec les autres filtres d'environ 20% avec des recouvrements dépassant les 100% pour l'urée, le lactate, la créatinine et le glucose. Les coefficients de variations calculés pour les paramètres de la fraction de bas poids moléculaire sont tous inférieurs à 7%, ce qui démontre une bonne reproductibilité du processus de filtration.

La déprotéinisation acide a permis de conserver l'ensemble des molécules de bas poids moléculaire excepté l'acide urique.

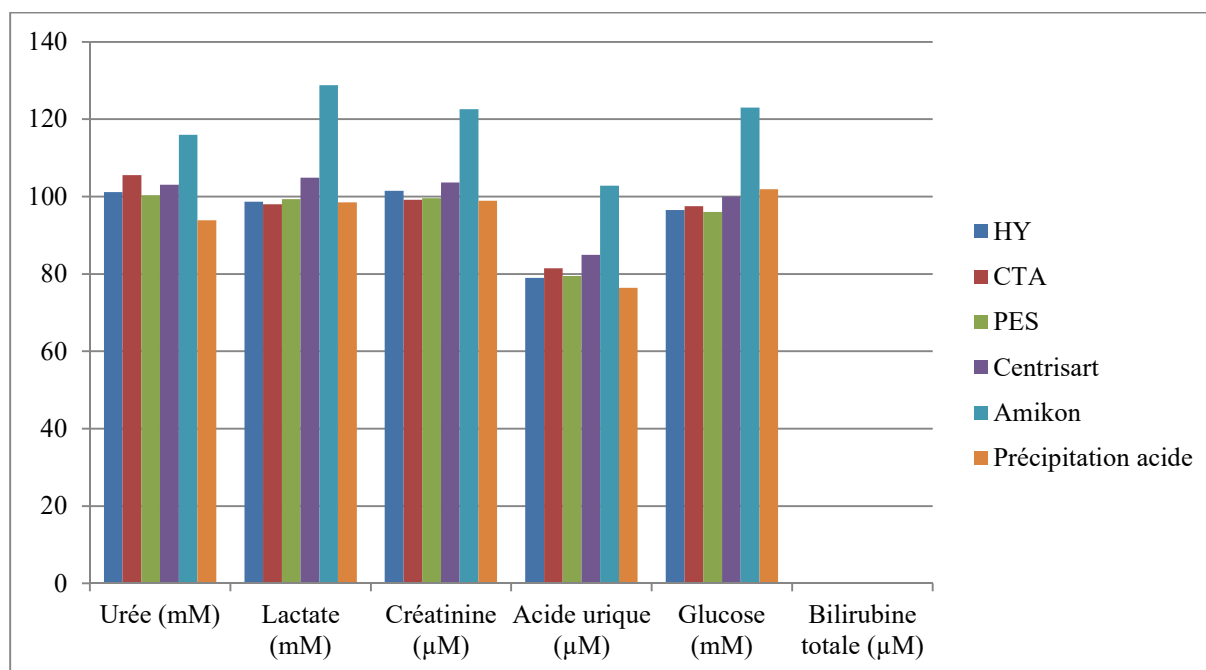


Figure 30 : Histogramme représentant les pourcentages de recouvrement des composants de la fraction de bas poids moléculaire pour chaque protocole d'élimination des protéines

c) Discrimination entre les échantillons patients avant et après filtration

Les bandes principales d'absorption des protéines comme les bandes amide I et amide II à 1640cm^{-1} et 1550cm^{-1} sont absentes sur le spectre d'absorbance mesuré par analyse ATR-IR d'un filtrat obtenu à l'aide d'un dispositif de type Centrisart® (Figure 31).

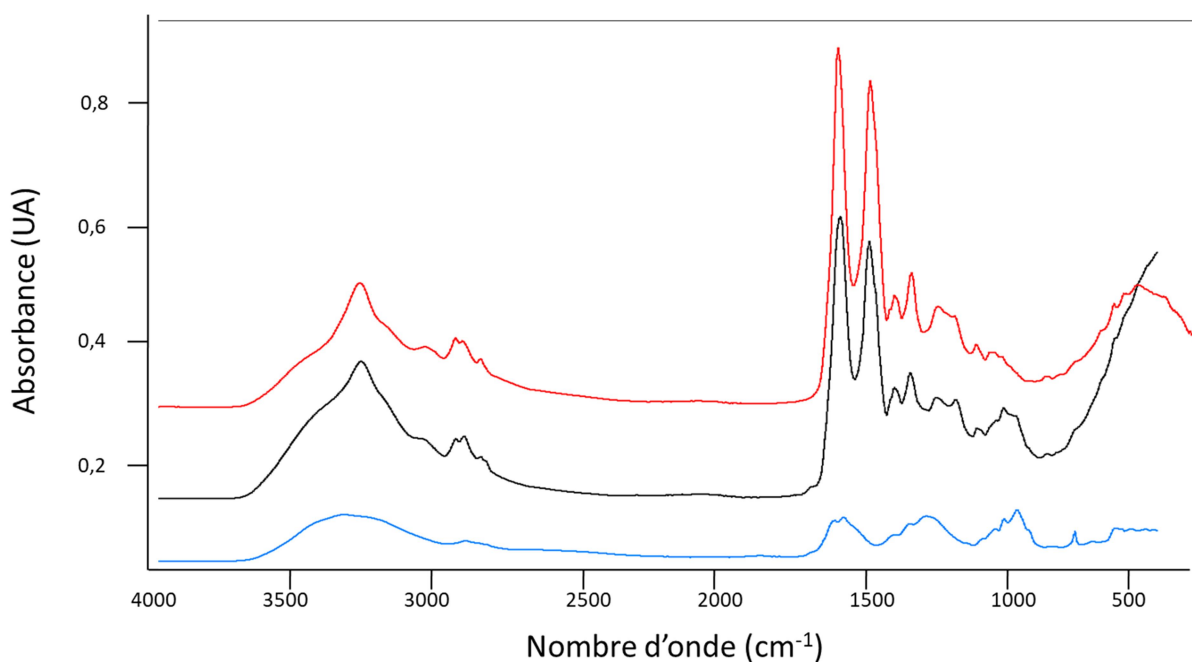


Figure 31 : Spectres ATR-IR mesuré à partir d'un sérum humain en noir, d'une solution d'albumine en rouge et d'un filtrat préparé à l'aide du filtre Centrisart® en bleu

Les résultats obtenus par l'analyse en LC-MSMS montrent une modification de la composition de l'échantillon due à l'étape de filtration. En effet, l'analyse par composante principale montre (Figure 32) que les métabolites dont la concentration est modifiée sont des acides aminés (alanine, citruline, glutamine, histidine, isoleucine, leucine, lysine, methionine, ornitine, sérine, tryptophane, tyrosine et valine) ainsi que l'acétylcarnitine avec une diminution des concentrations pour la condition filtrée. Les concentrations sont globalement divisées de moitié pour l'ensemble des prélèvements. Cependant les proportions entre les échantillons restent conservées, ce qui confirme la compatibilité de la préparation d'échantillon et l'analyse.

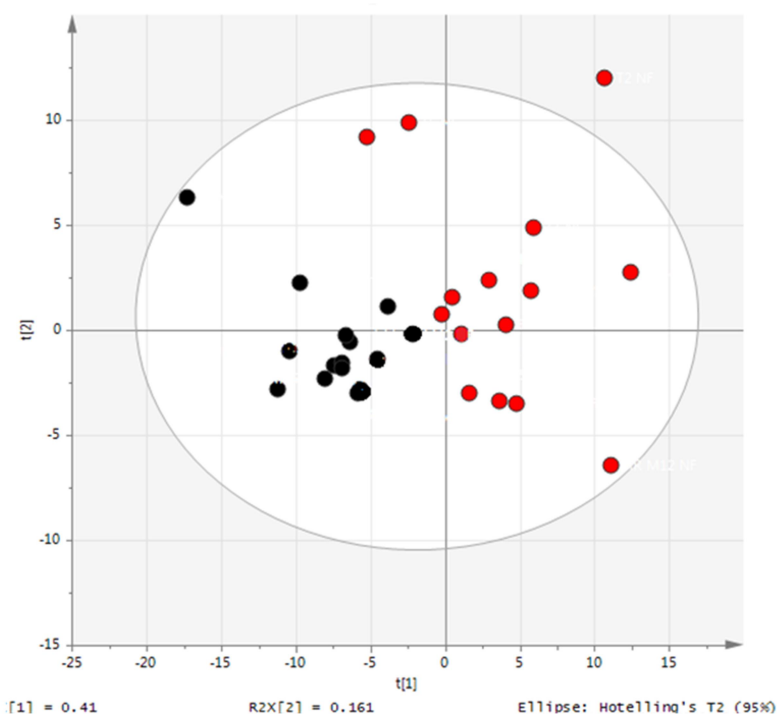


Figure 32 : Analyse en composantes principales (ACP) mettant en évidence une modification de la composition de l'échantillon avant ultrafiltration (rouge) et après ultrafiltration (noir)

DISCUSSION

a) Précipitation acide de la fraction protéique

Les résultats obtenus dans cette étude ont confirmé la bonne élimination globale de la fraction de haut poids moléculaire alliée à une bonne conservation de la fraction de bas poids moléculaire. De façon surprenante, la précipitation des protéines à l'aide d'acide sulfosalicylique n'élimine pas la totalité des protéines testées. En effet, les pourcentages de recouvrement du peptide C, de la préalbumine et de l'insuline sont respectivement de l'ordre de 50%, 105% et 25%. Les conditions de pH ne semblent pas être suffisamment dénaturantes pour ces protéines et il est envisageable que ce soit également le cas pour d'autres protéines. La limite de

quantification des protéines totales étant de 2g/L, nous n'avons pas pu déterminer la fraction protéique restante dans notre échantillon. Il serait nécessaire de compléter ces résultats par une autre méthode de quantification des protéines plus sensible. Concernant la fraction non protéique de l'échantillon, l'ensemble des molécules d'intérêt a été conservé excepté la bilirubine. La bilirubine est un composé peu soluble dans l'eau qui est présent soit sous forme libre (liée à l'albumine) soit sous forme conjuguée en formant un éther. Or l'ajout de l'acide dénature l'albumine et rompt les liaisons éther, libérant la bilirubine et la rendant insoluble dans notre échantillon.

b) Validation de l'ultrafiltration

Les concentrations mesurées pour chaque paramètre montrent une bonne élimination globale de la fraction protéique exceptée pour le peptide C et le NT-proBNP qui restent quantifiables dans le filtrat. Les masses molaires respectives de ces deux protéines sont de 3 et 3.45kDa, ce qui est inférieur à la valeur seuil indiquée par le fournisseur. La membrane du filtre ne devrait pas du tout retenir ces protéines et les recouvrements attendus étaient proches de 100%. Cependant, ces filtres sont commercialisés comme des concentrateurs de fraction protéique et la valeur seuil certifie la masse molaire à partir de laquelle la totalité des protéines est retenue dans le filtre. Il existe également des filtres avec un seuil de 3 kDa qui pourrait être intéressant pour éliminer la totalité de la fraction protéique.

Les recouvrements mesurés sur les différents composants de la fraction de bas poids moléculaire testés sont cependant satisfaisants (entre 80 et 100%). Le recouvrement plus important du filtre Amicon est probablement dû à une rétention d'eau plus importante dans le filtre ou dans le concentrat. La faible variabilité de

l'étape de préparation est compatible avec la réalisation d'une analyse quantitative et semi-quantitative. L'analyse par LC-MSMS a cependant mis en évidence une différence entre les profils métaboliques obtenus sur les échantillons filtrés et non filtrés. Ces différences ne concernent que 4 métabolites sur les 180 analysés. L'impact sur le profil métabolique complet est donc faible. Ce résultat est un argument supplémentaire en faveur de l'utilisation de cette préparation d'échantillon pour l'analyse du métabolome. De plus, la compatibilité avec la méthode d'analyse en LC-MSMS nous permet également d'envisager l'utilisation d'un protocole de préparation unique pour les deux analyses réalisées séquentiellement ou combinées.

V) Etude n°5 : Discrimination d'échantillons plasmatiques de patients SLA et de témoins

Les résultats obtenus lors de l'étude précédente nous ont permis de démontrer un meilleur accès à la fraction de bas poids moléculaire sans altération de sa composition. Dans ce contexte, nous avons étudié la discrimination de plasmas de patients atteints de la SLA et de plasmas témoins grâce à la spectroscopie IR en mode ATR. Pour cela, nous avons analysé les plasmas sans préparation d'échantillon et les plasmas après élimination de la fraction protéique.

MATERIEL ET METHODES

1) Echantillon

Dans cette étude, nous avons utilisé 5 échantillons de patients et 3 échantillons de témoins provenant d'un essai clinique. Les échantillons sanguins ont été collectés au diagnostic, centrifugés 10 minutes à 3000 g puis le plasma a été transvasé puis congelé à -80°C jusqu'à analyse.

2) Elimination de la fraction protéique

Le fractionnement du plasma a été réalisé à l'aide des filtres Centrisart-2,5mL (Sartorius, France) avec un seuil à 20 kDa selon le protocole décrit précédemment.

3) Analyse par ATR-IR

L'analyse par ATR-IR de l'ensemble des échantillons étudiés a été réalisée tel que décrit précédemment. Brièvement, 2µL de plasma ou de filtrat a été déposé sur le cristal ATR en diamant. Après séchage à l'air ambiant, nous avons réalisé les acquisitions spectrales de 3 dépôts entre 4000cm⁻¹ et 400 cm⁻¹.

4) Analyses statistiques

Les données recueillies en spectroscopie IR et par LC-MSMS ont été analysées par une approche multivariée en utilisant Simca P + version 13.0 (Umetrics, Umeå, Suède). Des analyses multivariées par composante principale ou discriminantes par régression des moindres carrés partiels (OPLS-DA ou PLS-DA) ont ensuite été réalisées. L'OPLS-DA identifie les variations des intensités ou des concentrations en

métabolites entre les groupes : la variation des données mesurées a été partitionnée en deux blocs par le programme, l'un contenant des variations qui correspondent à l'identifiant de classe et l'autre contenant des variations orthogonales au premier bloc et donc non contributives à la discrimination de groupe.

RESULTATS

L'analyse du plasma complet par spectroscopie infrarouge ATR-IR et le traitement statistique des données par analyse en composantes principales a permis d'obtenir une discrimination imparfaite entre les échantillons témoins et les échantillons patients (Figure 33 A). En effet, l'un des trois témoins analysés a été prédit comme un échantillon malade.

L'analyse réalisée sur les échantillons après l'étape d'ultrafiltration et le traitement statistique des données par analyse en composantes principales permet d'améliorer la discrimination des deux groupes d'échantillon. En effet, l'ensemble des échantillons analysés sont correctement prédit par le modèle (Figure 33B).

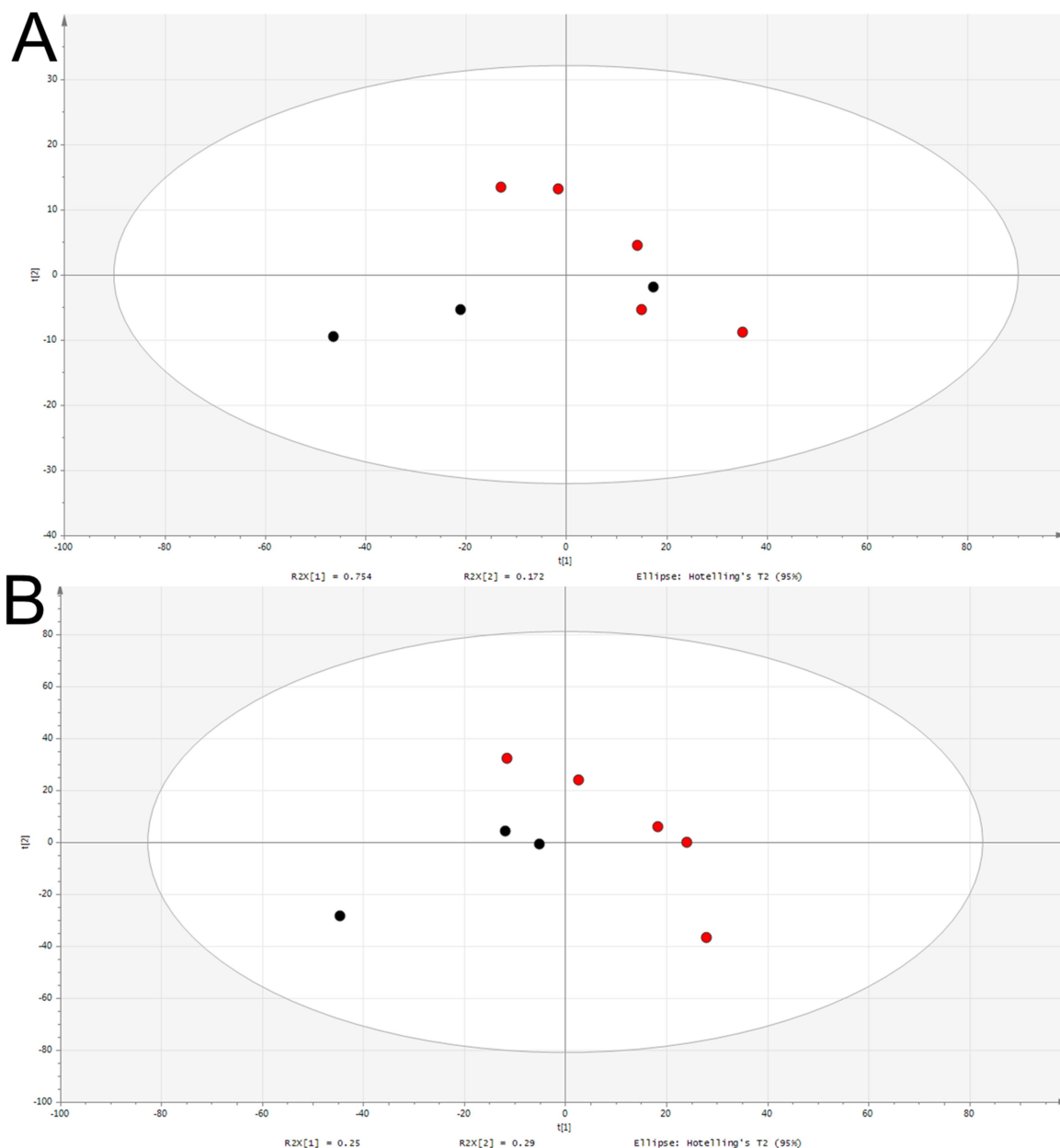


Figure 33 : Amélioration de la discrimination par ATR-IR après filtration. Les cercles noirs représentent les échantillons témoins et les cercles rouges les échantillons patients. A : Score scatter plot : PCA des échantillons avant filtration. B : Score scatter plot : PCA des échantillons après filtration

L'ATR-IR et l'analyse globale

L'analyse par spectrométrie infrarouge associée à l'analyse en composantes principales semble être un outil pertinent pour la construction de modèle permettant de discriminer des échantillons provenant de populations différentes. Les essais réalisés ont cependant montré la limite de l'analyse sur plasma brut. En effet, l'un des trois échantillons était mal prédit par le modèle. Cependant, le protocole d'ultrafiltration que nous avons utilisé a amélioré le modèle et permis de prédire correctement l'ensemble des échantillons analysés. Cependant, cette étape d'ultrafiltration conduit à perdre l'information contenue dans la fraction de haut poids moléculaire. De ce fait, la combinaison des deux types d'analyse augmenterait l'information extraite de l'échantillon et pourrait améliorer le modèle permettant la discrimination entre les deux groupes. En effet, la fusion de données mène à une amélioration de la séparation des groupes mais pose également la question de la prise en compte des redondances entre les différentes analyses. Le spectre d'absorbance du plasma complet dépend aussi de la composition de la fraction de bas poids moléculaire, il serait pertinent d'analyser le filtrat et le concentrat obtenus par l'ultrafiltration. Cela permettrait de fusionner les informations contenues dans chaque fraction du plasma tout en limitant la redondance des données. Or dans la littérature, plusieurs biomarqueurs protéiques de la SLA ont été décrits dans le plasma (Tableau XII). La préalbumine (ou transthyréthine) a notamment été retrouvée diminuée dans le LCR des patients SLA ainsi que la CRP augmentée dans le sérum et le LCR.

*Tableau XII : Biomarqueurs protéiques identifiées dans la Sclérose Latérale
Amyotrophique à partir de prélèvements sanguins ou de LCR*

Biomarqueur	LCR	Plasma	Sérum	Leucocytes	Références
Transthyréthine	↓↓				(207,208)
Cystatine C	↓↓↓↓↓				(207–211)
CRP	↑	↑			(208,212)
Chaîne lourde phosphorylée du neurofilament	↑↑↑↑	↑	↑		(210,213–216)
Chaîne légère du neurofilament	↑↑				(213,217)
Chitotriosidase	↑↑				(210,218)
MCP1- α		↑	↑		(219,220)
TDP43				↑↑	(221,222)

Lors de nos essais de filtration, nous avons établi que ces différents biomarqueurs sont retenus par le filtre et ne passent pas dans le filtrat. L'équipe de Dai a également démontré que la fusion de données obtenues par ATR-IR et par LC-HRMS permettait d'améliorer les modèles OPLS-DA de prédiction (193). Il se pose cependant la question de la bonne stratégie pour fusionner des données de concentrations de métabolites et des intensités d'absorbances. L'ensemble de ces résultats est très encourageant pour l'utilisation de l'ATR-IR pour une analyse de type métabolomique seule ou combinée à des outils conventionnels tel que la spectrométrie de masse. Ces résultats restent néanmoins préliminaires et nécessitent une confirmation sur un nombre d'échantillon plus grand.

DISCUSSION GENERALE

La combinaison des méthodes conventionnelles permet une amélioration de la couverture métabolique

Les différentes analyses réalisées ont mis en évidence l'importance de la LC-HRMS qui a permis l'identification de 80 métabolites dont 58 qui n'ont été observés par aucune autre technique. L'ensemble de ces métabolites permet de couvrir principalement le métabolisme des acides aminés et le métabolisme des purines. L'étude menée dans ce travail a également pu démontrer la spécificité de chaque méthode utilisée malgré le faible apport de la GC-MS dans ce contexte. Cependant, certaines molécules sont également observées et/ou quantifiées par différentes méthodes avec parfois des concentrations mesurées différentes. Dans ces conditions, la combinaison de ces différents outils pose la question de la manière de traiter l'ensemble des données obtenues.

Le second résultat majeur de cette étude est la couverture incomplète du métabolome. En effet, nous avons pu observer un défaut de couverture au niveau du métabolisme lipidique. Pour pallier à ce déficit inhérent à la préparation de l'échantillon et à la séparation chromatographique, il est possible de rajouter une analyse avec un protocole d'extraction dédié (185). Une seconde alternative est de proposer de nouvelles méthodes d'analyse à celles utilisées. Dans la littérature, l'analyse des lipides peut être envisagée à l'aide de la spectroscopie Raman, la spectroscopie infrarouge ou encore par spectrométrie de masse avec une séparation préalable par électrophorèse capillaire (223). Finalement la combinaison des outils utilisée doit également être adaptée au type d'échantillon utilisé. Dans notre étude,

nous avons observé un faible apport de l'analyse par GC-MS et ce, malgré l'utilisation d'un protocole d'extraction adapté au acides aminés, acides organiques, acides gras et nucléotides (224). Cependant d'autres études ont également montré un grand nombre de métabolites observés avec cette même méthode d'analyse. Finalement la combinaison à utiliser est différente selon l'échantillon étudié et nécessite une étude préalable pour simplifier l'analyse des données et diminuer la quantité d'échantillon nécessaire.

Quantification de biomarqueurs plasmatiques par spectroscopie vibrationnelle

Analyse sous forme liquide et sous forme séchée

L'ensemble des études menées dans ce travail a permis de démontrer le potentiel de la spectroscopie infrarouge par réflectance totale atténuée (ATR-IR) en tant qu'outil bioanalytique pour détecter de faibles variations spectrales dans des échantillons biologiques complexes comme le sérum.

Dans ce travail nous avons démontré l'intérêt d'analyser le sérum en ATR-IR sous ses différentes formes (liquides et après une étape de séchage). L'analyse des échantillons après séchage permet d'augmenter la sensibilité de l'analyse en éliminant l'eau présente dans l'échantillon. En revanche, bien que la plupart des études utilisant la spectroscopie ATR-IR soient réalisées sur les échantillons séchés, nous avons démontré que l'analyse sous forme liquide permet d'avoir accès à la fraction protéique. Ces deux acquisitions sont complémentaires et nécessitent d'être utilisées en combinaison. En effet, nous avons pu observer une meilleure fiabilité de l'analyse de la fraction protéique sous forme liquide. A l'inverse, les concentrations

des molécules de faible poids moléculaire sont mieux estimées lors de l'analyse de l'échantillon sous forme séchée.

Utilisation de la spectroscopie vibrationnelle comme outil diagnostique en biologie

Malgré les résultats satisfaisants observés, la spectroscopie ATR-IR et la spectroscopie Raman ont toutes les deux montré des limites dans la quantification spécifique de biomarqueurs. En effet, les erreurs de prédictions des concentrations sont supérieures aux valeurs recommandées par les sociétés savantes pour une analyse quantitative ciblée. Ces résultats démontrent l'incapacité de ces deux techniques à remplacer les outils analytiques utilisés actuellement pour la réalisation des analyses quantitatives de biomarqueurs dans le domaine médical dans les conditions actuelles. En revanche, ces imprécisions n'ont pas eu d'impact sur l'interprétation biologique des résultats. Cependant, ces outils permettent une analyse rapide avec l'obtention d'un spectre représentatif de la composition complète de l'échantillon. Cet avantage associé aux résultats obtenus dans ce travail permet de positionner ces technologies comme outils de dépistage permettant une orientation rapide du patient lors de son entrée à l'hôpital.

De plus ces résultats ont été obtenus après une étape de dépôt non automatisée et une étape de séchage qui est sensible aux conditions environnementales. La standardisation de cette dernière étape est complexe à cause de la nécessité de maîtriser différents paramètres tel que la température et l'hygrométrie ainsi que l'inhomogénéité chimique et physique de l'échantillon qui en découle. Cette inhomogénéité de la préparation d'échantillon pourrait être levée par une automatisation du processus et une maîtrise de ces facteurs environnementaux. Pour

finir, le fractionnement du plasma et l'analyse de la fraction de bas poids moléculaire pourrait permettre d'améliorer les performances analytiques en éliminant l'interférence des protéines.

Finalement, l'automatisation des étapes de dépôts et de séchage des échantillons voire de fractionnement de l'échantillon permettraient d'améliorer la précision de la mesure, ce qui pourrait permettre d'envisager la quantification simultanée de plusieurs biomarqueurs en quelques secondes. En comparaison, une analyse quantitative d'un seul biomarqueur par une technique automatisée peut durer entre une trentaine de secondes et une trentaine de minutes.

La spectroscopie vibrationnelle dans une approche métabolomique

Dans ce travail, l'analyse de plasma par ATR-IR a permis de discriminer les échantillons de patients témoins et ceux de patients atteints de la SLA. Cependant, contrairement aux autres outils disponibles, la détermination de bandes d'absorbances discriminantes en ATR-IR ne permet pas d'identifier les molécules. Malgré cela, la spécificité de certaines bandes pour des liaisons particulières permet d'identifier une classe de molécule impactée. Dans cet objectif, la spectroscopie infrarouge pourrait être utilisée comme test préalable. Cette analyse permettrait d'identifier des classes de molécules perturbées et de réaliser par la suite une analyse métabolomique ciblée sur ces différentes familles. Ce processus conduirait ainsi à une analyse plus simple et moins coûteuse combinant moins de méthodes analytiques sans perdre d'information pertinente.

Elle permet également d'obtenir des données qualitatives et semi-quantitatives à partir d'un très faible volume de liquide biologique (225). Cette caractéristique rend

cet outil particulièrement attractif pour l'analyse de fluides biologiques précieux comme le LCR, les larmes, l'humeur aqueuse, la sueur ou encore la salive. En effet, les études métabolomiques sur des liquides biologiques présents en faible quantité sont de plus en plus fréquentes. Malgré tout, la problématique de ce genre d'analyse est la faible quantité d'échantillon disponible. Yazdani et al ont récemment publié que le volume de larme prélevé est de l'ordre de $100\mu\text{L}$, ce qui limite grandement le nombre d'analyse pouvant être réalisée (226). Selon Travo et al, une analyse des larmes par ATR-IR serait possible sur seulement $2\mu\text{L}$ d'échantillon, ce qui rend cet outil particulièrement intéressant dans ce contexte.

D'autres études ont également développé des senseurs pour une analyse ciblée par ATR-IR. En 2019, Budde et al ont développé un outil de diagnostic de la maladie d'Alzheimer réutilisable à partir d'un diamant de germanium sur lequel étaient fixées des immunoglobulines ciblant la protéine Tau (227). Dans cette étude, les auteurs ont pu détecter la présence de protéine Tau anormale à partir de $100\mu\text{L}$ de LCR contre plusieurs centaines de microlitres par une méthode ELISA.

AUTRES PROJETS

Optimisation de l'analyse de plasmas humains par Spectroscopie Raman et évaluation de la capacité à quantifier spécifiquement un/des biomarqueur(s)

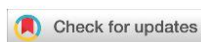
Etude n°1 : Analyse des fluides biologiques à l'aide de la spectroscopie vibrationnelle : une comparaison directe des spectroscopies de diffusion Raman et d'absorption infrarouge pour le cas du glucose dans le plasma

Similairement à la spectroscopie infrarouge, la spectroscopie Raman permet une caractérisation moléculaire des fluides biologiques pour accéder à une information biochimique. Intrinsèquement, la spectroscopie Raman est plus adaptée à l'analyse d'échantillons sous forme liquide ce qui lui donne un petit avantage expérimental et une meilleure transférabilité en milieu clinique. Cependant, les spectres obtenus peuvent présenter un bruit de fond important avec des déformations de la ligne de base. Des études complémentaires doivent donc être menées sur les fluides biologiques et dans ce cas sur le plasma pour évaluer la capacité de la spectroscopie Raman à extraire les informations cliniquement pertinentes.

Le but de cette étude est de comparer les résultats obtenus précédemment par spectroscopie ATR-IR et ceux par spectroscopie Raman. Nous avons donc souhaité

mettre en place en spectroscopie Raman la même étude que celle réalisée précédemment par ATR-IR.

Comme pour la spectroscopie infrarouge, le spectre obtenu par l'analyse d'un échantillon de plasma est la somme des contributions spectrales de l'ensemble des composants de l'échantillon. Bien que la réponse des molécules diffère entre IR et Raman, la problématique de la fraction de haut poids moléculaire et des protéines abondantes reste applicable. Les spectres Raman seront aussi dominés par les protéines. Les échantillons utilisés dans cette étude ont donc également été filtrés pour ne récupérer et analyser que la fraction de bas poids moléculaire du sérum. Dans cette étude, nous avons dans un premier temps suivi l'évolution du signal mesuré par spectrométrie Raman selon les concentrations de glucose dans des solutions aqueuses pures. Le modèle obtenu a ensuite été confirmé à l'aide de sérum chargé par des concentrations connues de glucose. Pour finir, nous avons appliqué le même protocole sur 25 échantillons patients dont les concentrations étaient comprises entre 2,9 et 24,1 mmol/L. Nous avons finalement comparé les concentrations prédites par l'analyse Raman à celles obtenues à l'aide d'une méthode enzymatique à l'hexokinase sur un automate Cobas c501 (Roche), validée et accréditée qui fait office de méthode de référence.


 Cite this: *Analyst*, 2019, **144**, 3334

Analysis of bodily fluids using vibrational spectroscopy: a direct comparison of Raman scattering and infrared absorption techniques for the case of glucose in blood serum†

 Drishya Rajan Parachalil,^{a,b} Clément Bruno,^{c,d,e} Franck Bonnier,^c Hélène Blasco,^{d,f} Igor Chourpa,^g Matthew J. Baker,^h Jennifer McIntyre^a and Hugh J. Byrne^a

Analysis of biomarkers present in the blood stream can potentially deliver crucial information on patient health and indicate the presence of numerous pathologies. The potential of vibrational spectroscopic analysis of human serum for diagnostic purposes has been widely investigated and, in recent times, infrared absorption spectroscopy, coupled with ultra-filtration and multivariate analysis techniques, has attracted increasing attention, both clinical and commercial. However, such methods commonly employ a drying step, which may hinder the clinical work flow and thus hamper their clinical deployment. As an alternative, this study explores the use of Raman spectroscopy, similarly coupled with ultra-filtration and multivariate analysis techniques, to quantitatively monitor diagnostically relevant changes of glucose in liquid serum samples, and compares the results with similar analysis protocols using infrared spectroscopy of dried samples. The analysis protocols to detect the imbalances in glucose using Raman spectroscopy are first demonstrated for aqueous solutions and spiked serum samples. As in the case of infrared absorption studies, centrifugal filtration is utilised to deplete abundant analytes and to reveal the spectral features of Low Molecular Weight Fraction analytes in order to improve spectral sensitivity and detection limits. Improved Root Mean Square Error of Cross Validation (RMSECV) was observed for Raman prediction models, whereas slightly higher R^2 values were reported for infrared absorption prediction models. Summarising, it is demonstrated that the Raman analysis protocol can yield accuracies which are comparable with those reported using infrared absorption based measurements of dried serum, without the need for additional drying steps.

 Received 19th January 2019,
Accepted 2nd April 2019

DOI: 10.1039/c9an00125e

rsc.li/analyst

1. Introduction

Human bodily fluids (e.g. blood serum/plasma, urine, saliva, tears and cerebrospinal fluid) are considered to be a rich reservoir of clinical biomarkers and are an interesting alternative to

cells and tissues in terms of disease diagnosis and prognosis, owing to advantages such as minimal invasiveness, low cost, and rapid sample collection and processing.^{1–4} The biochemical composition of human serum can provide crucial information on patient health and indicate the presence of numerous pathologies, as it encompasses a vast range of proteins and biochemical products accumulated while perfusing various organs.⁵ Moreover, alterations in the biochemical composition of the serum/plasma could reflect changes of physiological states due to disease, enabling early disease diagnosis and treatment.^{6–8} For example, it is reported that carcinoembryonic antigens, CA 15-3 and CA 27.29, can be considered as serum biomarkers for breast cancer,⁹ KL-40 and MMP-9 have been found to be potential serum biomarkers for high grade gliomas,¹⁰ prostate specific antigen was found in higher levels in the serum of patients with prostate cancer,^{11,12} and carcinoembryonic antigen has been reported to be serum biomarker for colorectal cancer.¹³ In recent decades, extensive studies have explored the detection of serum biomarkers for

^aFOCAS Research Institute, Technological University Dublin, Kevin Street, Dublin 8, Ireland

^bSchool of Physics and Optometric & Clinical Sciences, Technological University Dublin, Kevin Street, Dublin 8, Ireland. E-mail: drishyarajan.parachalil@mydit.ie

^cUniversité de Tours, UFR sciences pharmaceutiques, EA 6295 Nanomédicaments et Nanosondes, 31 avenue Monge, 37200 Tours, France

^dCHRU de Tours, Laboratoire de Biochimie et Biologie Moléculaire, Tours, France

^eUniversité de Tours, iBrain, UMR INSERM U1253, 37032, France

^fWestCHEM, Department of Pure & Applied Chemistry, Technology and Innovation Centre, University of Strathclyde, Glasgow, G1, 1RD, UK

^gClinSpec Diagnostics Ltd (ClinSpec Dx), Level 7, Technology and Innovation Centre, 99 George Street, Glasgow, G1, 1RD, UK

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9an00125e

various diseases, and quite naturally the reference tools used are the conventional, complex analytical techniques such as chromatography, electrophoresis or mass spectroscopy.^{14–18} More recently, the field of serum proteomics has exploded in the literature, as this emerging field has gained world-wide attention due to its potential to reveal important information regarding the pathogenesis or progression of diseases.^{19–22}

Analysis of serum biomarkers based on protein content is inherently challenging, because of the low concentrations and the vast variety and dynamic range of protein abundance. The complex nature of the serum poses as a huge problem for the detection of small molecule biomarkers.^{2,23} Human serum contains more than 10 000 different proteins, with an overall concentration ranging from 60–80 mg mL⁻¹. Furthermore, circulating species present in the serum, such as metabolites, peptides, sugars, and lipids, add to its complexity. Conventional proteomic methods struggle to handle large dynamic range of abundances of its constituent components.¹⁷ The characteristics of serum are usually dominated by the high molecular weight fraction (HMWF) of proteins, which includes albumin (57–71%) and globulin (8–26%) and masks the features of low molecular weight analytes which are present in trace amounts (<5%).²⁴ Chromatographic fractionation can significantly enhance the sensitivity and specificity of the data recorded, and coupling with spectroscopic analysis techniques such as IR absorption in hybrid techniques such as Liquid Chromatography (LC-IR) or Gas Chromatography (GC-IR)²⁵ can improve the performance further. Such chromatographic techniques, although extensively used in pharmaceutical, chemical or food science applications are time consuming and costly, however,^{26–28} and commercially available centrifugal filters have been employed to deplete the HMWF before analysis to facilitate analysis of the low molecular weight fraction (LMWF), including proteins and molecular biomarkers.²⁹

Vibrational spectroscopic techniques, both Raman and infrared absorption, have emerged over the past 20 years as increasingly routine analytical techniques for a wide range of applications, as they reveal specific biochemical information without the use of extrinsic labels. Although they are often considered complementary techniques³⁰ there are also specific considerations for each, for specific applications such as measurement of bodily fluids, as described by Bonnier *et al.*³¹ They provide intrinsic vibrational signatures of the material of interest in a non-destructive fashion, and the potential for diagnostic applications has been well demonstrated, notably in human serum and plasma.^{7,32–34} However, although both Raman, Fourier-Transform Infrared (FTIR) and Attenuated Total Reflectance-FTIR (ATR-FTIR) spectroscopy have been widely explored to study bodily fluids over the last two decades, most of these studies have been carried out on air dried samples, in order to avoid the water contribution in the case of FTIR, and to increase the concentration of the analytes in the case of Raman.^{35–40} Two major limiting factors in the use of dried samples are the drying time⁴¹ and also the so-called “coffee-ring” effect, or, specifically in terms of blood

serum, the Vroman effect,^{42–44} whereby different analytes precipitate from solution at different rates, giving rise to variations in the spectral features due to chemical and physical inhomogeneity. Previous studies have clearly shown that the IR spectrum of dried (aggregates of) molecular species is not the same as that in solution.⁴⁵ It has further been shown that, for dried deposits from solutions of varying concentrations of analyte, the linearity of the Beer–Lambert law, and therefore the quantitative nature of the measurement is compromised as the concentration of the analyte is increased.⁴⁵ In the case of the “coffee ring effect”, the thickness, and therefore the quantitative accuracy of the measurement is spatially inhomogeneous, and the technique is not ideally suited for quantitative measurements, such as those considered here. Notably, Spalding *et al.* have used a serum dilution technique to improve the reliability of the technique for quantitative measurement.⁴⁶ The issues with inhomogeneity can potentially be overcome by micropipetting and sampling the whole drop, and there have been studies that show excellent specificity/sensitivity of classification of diseased state using dried samples.^{40,47,48} Nevertheless, the requirement of a drying step adds considerably to the sampling time and complexity of the workflow.^{41,49}

Protocols for monitoring changes in HMWF serum proteins in their native liquid form using Raman spectroscopy have recently been demonstrated, however, without the need for drying.⁵⁰ Analysing in the native liquid state, the chemical composition is averaged out by molecular motion over the measurement time, greatly reducing the variability of the measurement. Bonnier *et al.* compared FTIR and Raman for measuring gelatin in solution and described protocols for isolation of LMWF from serum.³¹ This study indicated that Raman in the liquid form could deliver similar sensitivities to ATR-FTIR for measurement of LMWF species.

The aim of this study is to investigate the sensitivity and accuracy of Raman spectroscopy in the liquid state, coupled with centrifugal filtration, as a biochemical tool to detect clinically relevant changes in the biochemical composition of serum and compare to the ATR-FTIR technique, using the specific example of glucose. The study is specifically designed according to the ATR-FTIR study of glucose in serum by Bonnier *et al.*, using identical parameters and protocols of ultra-filtration and multivariate regression analysis, such that a direct comparison of the two techniques can be made.²⁹ The protocol is first demonstrated using aqueous solutions and human serum spiked with systematically varied concentrations of glucose, before exploring the sensitivity and accuracy of the technique in patient samples.

2. Materials and methods

2.1 Preparation of varying concentration of glucose in distilled water model

D-Glucose (G8769) was purchased from Sigma Aldrich, Ireland and 6 glucose solutions were prepared over the concentration

range 100 mg dL⁻¹ to 1000 mg dL⁻¹. Amicon Ultra 0.5 mL centrifugal filter devices (Millipore-Merck, Germany), with 10 kDa cut off point, were employed to concentrate and fractionate the serum samples. The centrifugation procedure previously reported by Bonnier *et al.* was followed.³¹ A further study published by Bonnier *et al.* reported 100% recovery of LMWF using 10 kDa, hence only 10 kDa cut off filtration has been used in the present study.⁴⁵ Pre-rinsing of the filter devices with 0.1 M NaOH prior to plasma analysis is essential to avoid glycerine interference in the analysis.³² The optimised washing and rinsing procedure includes spinning 0.5 mL 0.1 M NaOH at 14 000g for 30 minutes followed by three rinses with distilled water by spinning 0.5 mL distilled water for 30 minutes at 14 000g. Every 30 minutes wash and rinse must be followed by spinning the device in the inverted position at 1000g for 2 minutes, to remove the residual solution contained in the filter. After washing, 0.5 mL of glucose solution is transferred to the 10 kDa filter and centrifuged at 14 000g for 30 minutes. The solution that passes through the 10 kDa filter is the filtrate, which contains mostly water and molecules smaller than 10 kDa. The remainder of the serum, known as the concentrate, is collected by placing the filter device upside down and spinning for 1000g for 2 minutes. The resultant concentrate, ~50 µL, contains molecules with molecular weight larger than 10 kDa, concentrated by a factor of ~10, and can be employed for study of the HMWF.⁵⁰ All the filtrate solutions were analysed using Raman spectroscopy and five replicate measurements from different positions have been performed for each sample of ~50 µL. In subsequent analysis, each patient is represented by all the spectra recorded from that patient, rather than the mean.

2.2 Preparation of glucose spiked in serum model

Sterile, human serum (H6194) and D-glucose (G8769) were purchased from Sigma Aldrich, Ireland, for the preparation of *in vitro* spiked models. The commercial human serum was spiked with glucose in the concentration range that matches of the study of Bonnier *et al.*²⁹ Since glucose is already present in normal human serum at concentrations of 70–110 mg dL⁻¹,⁵¹ the final concentration of the spiked samples covers the physiologically relevant ranges of normal (80–120 mg dL⁻¹),⁵² and hyperglycaemia (>120 mg dL⁻¹). The centrifugal processing step described in section 2.1 was performed on the spiked serum samples to obtain the filtrate. Note, in the analysis of Fig. 3, the regression is performed over spiked, rather than total glucose concentration, consistent with the approach of Bonnier *et al.*²⁹

2.3 Glucose levels in patient serum samples

Patient serum samples were donated by the University Hospital (CHU) Bretonneau de Tours (France) and the ethical procedures were followed. The blood samples were collected from the individuals as routine blood check-ups and 1 mL per patient was provided for spectroscopic analysis. A total of 25 patient samples were included in the present study. Samples

were collected by personnel of CHU, under standard clinical protocols and approved ethical procedures (Comité de Protection des Personnes, Tours Region Central Oest 1-PP/ANSM-PHA015-HB-METABOMU, registered internationally: ClinicalTrials.gov ID: NCT02670226). The samples were serologically profiled, for other purposes, and the anonymised, residual discard samples, along with their serological profiles were donated to the Université François-Rabelais de Tours, for further study. No further specific ethical approval or patient consent is required. Glucose concentrations were obtained by routine biochemical analysis using a COBAS analyser, following the in house guidelines for routine biochemical analysis. The principle of the test is based on the enzymatic reference method with hexokinase, which catalyses the phosphorylation of glucose to glucose-6-phosphate by ATP methods.^{53,54} Subsequently, glucose-6-phosphate is oxidised by glucose-6-phosphate dehydrogenase, in the presence of NADP, to gluconate-6-phosphate. This reaction is specific, with no other carbohydrate being oxidised. The rate of NADPH formation during the reaction is directly proportional to the glucose concentration and is measured photometrically in the UV. Measured glucose levels in the patient samples are listed in Table 1, and can be seen to cover the range 55–435 mg dL⁻¹. Note, that the distribution of glucose levels covers a broader range than that of the study of similar (*n* = 15) patients by Bonnier *et al.* (61–208 mg dL⁻¹).²⁹

Table 1 List of measured glucose levels in patient samples. Glucose blood levels are quoted in terms of the SI unit of mmol L⁻¹ as well as mg dL⁻¹, commonly used in serology literature and in the study of Bonnier *et al.*²⁹

Sample number	Glucose blood levels	
	mmol L ⁻¹	mg dL ⁻¹
1	2.9	52.25
2	3.1	55.85
3	3.9	70.27
4	3.9	70.27
5	4.2	75.67
6	4.2	75.67
7	4.3	77.47
8	4.5	81.08
9	5.1	91.89
10	5.1	91.89
11	5.2	93.69
12	5.7	102.70
13	6	109.10
14	6.1	109.90
15	6.4	115.31
16	6.4	115.31
17	7.2	129.72
18	8.8	158.55
19	9.9	178.37
20	11.5	207.20
21	11.7	210.81
22	13.4	241.44
23	15.7	282.88
24	15.8	284.68
25	24.1	434.23

2.4 Data collection using Raman spectrophotometer

The measurement conditions used for screening HMWF proteins in solution have recently been detailed.⁵⁰ Raman spectra of all the liquid samples were recorded at stabilised room temperature (18 °C) using a Horiba Jobin-Yvon LabRam HR800 spectrometer with a 16-bit dynamic range Peltier cooled CCD detector. The spectrometer was coupled to an Olympus 1X71 inverted microscope and a ×60 water immersion objective (LUMPlanF1, Olympus) was employed. The substrate used was a Lab-Tek plate (154534) with a 0.16–0.19 mm thick glass bottom, 1.0 borosilicate cover glass, and was purchased from Thermo Fischer Scientific, Ireland. Since the scattering efficiency is inversely proportional to the fourth power of the wavelength, it is always desirable to use a short wavelength Raman source. In the filtered serum, there are no molecular species which are resonant with 532 nm, and therefore neither fluorescence nor photodamage are limiting factors.

2.5 Data pre-processing and analysis

The raw spectra were subjected to pre-processing techniques in Matlab before further analysis, to remove the background signal and reduce the noise. Smoothing of the raw data was done using the Savitzky-Golay method, with a polynomial order of 5 and window 13 and the Extended Multiplicative Signal Correction (EMSC) algorithm was applied to the smoothed spectra to remove the underlying water spectrum, whose OH bending feature at 1640 cm⁻¹ can interfere with the protein spectra, particularly at low concentrations.⁵⁵

The principle of EMSC for subtraction of a specific measurable background spectrum and the associated Matlab codes have previously been published by Kerr and Hennelly, 2016,⁵⁵ and their description has been adapted in the following. In the case of measurement in aqueous solution, the raw sample spectrum, S , consists of Raman spectrum of the analyte of interest, R , a baseline signal, B , and the water signal, W .⁵⁵

$$S = R + B + W \quad (1)$$

The Raman spectrum of interest can be represented by a reference spectrum of the analyte of interest, r , and it can be assumed that R is the product of this reference spectrum and a certain scalar weight, C_r , which describes the concentration dependence.^{55,56,57}

$$R \sim C_r \times r \quad (2)$$

Similarly, a spectrum, w , is recorded from water directly in order to represent the spectral contribution of water in W , as the product of pure water spectrum and a certain scalar weight.⁵⁵

$$W = C_w \times w \quad (3)$$

The baseline, B , is now represented by a polynomial:⁵⁵

$$B_N = C_0 + C_1X + C_2X^2 + \dots + C_NX^N \quad (4)$$

where N is the order of polynomial and C_m for $m = 0 \rightarrow N$ represents various coefficients of polynomial.⁵⁸ The EMSC algorithm is used to obtain estimates of the scalar values C_r , C_m and C_w . These estimates are obtained from an optimal fit of the various vectors in eqn (5).⁵⁵

$$S \sim [C_r \times r] + [C_w \times w] + \left[\sum_{m=0}^N C_m X^m \right] \quad (5)$$

The background corrected, concentration dependent analyte spectra, T , can be represented as:⁵⁵

$$T = \frac{S - [C_w \times w] - \sum_{m=0}^N C_m X^m}{C_w} \quad (6)$$

Note, that division by C_w has the effect of scaling the analyte spectra, assuming a constant water contribution to all sample spectra. The Raman spectrum of the stock glucose solution (450 g L⁻¹) is used as the reference for EMSC, and polynomial of order 5 was used in all cases. The glucose reference spectrum and water spectrum used in EMSC were recorded using the 532 nm laser line and Lab-Tek plate as substrate. The ×60 objective brings the laser to a focus within the liquid, beyond the thin (0.16 mm–0.19 mm) glass substrate. No significant contribution from the glass to the recorded spectrum was apparent, and therefore no correction was required. The Rubberband method⁵⁹ was used to baseline correct the glucose reference spectrum after smoothing it using the Savitzky-Golay method.

2.5.1 Partial least squares regression. Partial Least Squares Regression (PLSR) was employed to establish a model that relates the variations of the spectral data to a series of concentrations. This regression model can be used to improve the limit of detection of Raman bio-sensing.⁶⁰ Constructed based on the spectra of samples of known glucose content, either solutions of varying concentrations of glucose (in water or in commercial serum), or those of the patient serum, the model is then validated using a rigorous cross validation procedure which evaluates its performance in accurately predicting glucose concentrations. A 20 fold cross validation approach has been employed to validate the robustness of the method. This approach involves randomly dividing the set of observations into approximately equal size, 50% of the spectral data were randomly selected as test set, while the remaining 50% is used as the training set.⁶¹ In the current case, (5 × 25 spectra) were divided into two groups of 65 (test) and 60 (training) spectra. The cross-validation process is then repeated 20 times (the folds), with all observations are used for both training and testing, and each observation is used for testing exactly once. The results from the folds can then be averaged to produce a single estimation. The Root Mean Square Error of Cross Validation (RMSECV) is calculated from the 20 iterations to measure the performance of the model for the unknown cases within the calibration set. The correlation between the concentration and spectral intensity is given by the R^2 value. The standard deviation was calculated to find the variation between

each spectrum calculated from the same sample. The number of latent variables used for building the PLSR model is optimised by finding the value that is equivalent to the minimum of the RMSECV.

3. Results

In this study, Raman spectra of the samples were recorded in the inverted geometry using a $\times 60$ water immersion objective with a 532 nm laser and the Lab-Tek plate was used as the substrate. The 532 nm laser was chosen as it is compatible with the thin glass bottomed Lab-Tek plate and provides a strong Raman signal of the sample with minimal background interference. This set-up was previously reported by Bonnier *et al.*³¹ to yield better analysis of serum using Raman spectroscopy when the sample was analysed in the inverted geometry using a water immersion objective with a 785 nm laser and CaF_2 substrate. The added advantage of this setup is that it provides high quality, consistent Raman spectra from sample volumes as low as 1 μL . The protocol using 532 nm was more recently further explored for analysis of HMWF serum proteins.⁵⁰

Fig. 1 presents the spectra of the fingerprint region of the pure glucose solution recorded in the inverted geometry. The raw spectra of the glucose were baseline corrected using the rubberband method and smoothed using the Savitzky-Golay algorithm (polynomial 5, window 13). Example signature peaks of glucose (indicated by asterisks) appear at $\sim 450\text{ cm}^{-1}$, associated with an endocyclic $\delta(\text{C}-\text{C}-\text{O})$ ring mode, $\delta(\text{C}1-\text{H}1)$ vibration at 911 cm^{-1} , a peak at 1060 cm^{-1} due to $\nu(\text{C}1-\text{OH})$ stretching, a relatively sharp peak at 1125 cm^{-1} which can be assigned to the $\delta(\text{C}-\text{O}-\text{C})$ angle-bending mode and the sharp

peaks at 1340 cm^{-1} and 1460 cm^{-1} , related to the $\delta(\text{C}-\text{C}-\text{H})$ vibration and pure CH_2 group vibration, respectively.⁶²

3.1 Monitoring the concentration dependence of glucose in distilled water

In order to establish the analysis protocol for the patient samples, a PLSR prediction model was first built and applied to the set of varying concentrations of glucose in distilled water, as well as spiked into serum. The first step in this study was to optimise the measurement protocol and also to evaluate the efficacy of the centrifugal filtration technique in separating and concentrating glucose from the HMWF proteins. For this, different amounts of pure glucose were spiked into the distilled water. The normal glucose concentration and the concentration and hyper-glycaemia were deliberately included to simulate physiologically relevant concentrations. Fig. 2A displays examples of the EMSC (polynomial order 5) corrected low ($5 \times 100\text{ mg dL}^{-1}$), medium ($5 \times 450\text{ mg dL}^{-1}$) and high ($5 \times 1000\text{ mg dL}^{-1}$) datasets, showing glucose features (indicated by asterisks) of greater intensity at higher concentration, whereas the lower concentrations shows weaker features of glucose. The groups of spectra are off set for clarity. The concentration 1000 mg dL^{-1} was deliberately included in the dataset to evaluate the consistency of the glucose spectral features as the concentration increases from 100 mg dL^{-1} to 1000 mg dL^{-1} . In order to analyse the spectral variations and the glucose concentrations, the PLSR algorithm was applied.

Based on the percent variance explained by the latent variables and the minimum value of RMSECV (Fig. 2B), the optimum number of latent variables to reach the best performance is determined to be 4. The PLSR coefficient plot displayed in Fig. 2C confirms the correlation of the data in Fig. 2D is based on glucose features, such as the peaks at $\sim 1060\text{ cm}^{-1}$, $\sim 1125\text{ cm}^{-1}$, 1450 cm^{-1} and $\sim 1340\text{ cm}^{-1}$. Finally, after selecting the optimum number of components for the data set analysed, a predictive model is built from the PLSR analysis (Fig. 2D), to compare the observations to the known concentrations of glucose in the samples with the estimated concentrations from the spectral data sets. Fig. 2D indicates that a satisfactory linear model could be obtained with the raw data set and that the concentration dependence of the sample set is conserved by centrifugal filtration. From the results shown in Fig. 2, it is evident that the predicted values are in good agreement with the reference concentrations and the corresponding correlation coefficient (R^2) is calculated as 0.9705. Note, each concentration point has five independent measurements, and the mean standard deviation of each measurement is 4.8 mg dL^{-1} . The RMSECV calculated from the 20 iteration of cross validation is 10.93 mg dL^{-1} , thereby indicating that PLSR provides accurate predictions for the glucose Raman data over the entire concentration range of 100 mg dL^{-1} to 1000 mg dL^{-1} .

3.2 Monitoring the glucose concentration in spiked serum

In an attempt to extend the optimised protocol to a more complex environment, PLSR analysis was performed on the EMSC corrected data set recorded from the filtrate of the

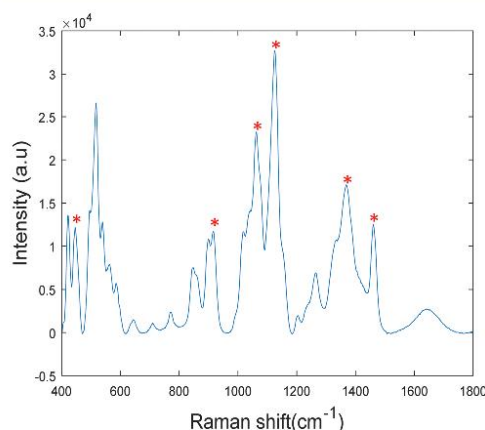


Fig. 1 Raman spectrum of an aqueous glucose solution, concentration 450 g L^{-1} . Example signature peaks at 450 cm^{-1} , 911 cm^{-1} , 1125 cm^{-1} , 1340 cm^{-1} and 1460 cm^{-1} are labelled in the figure.

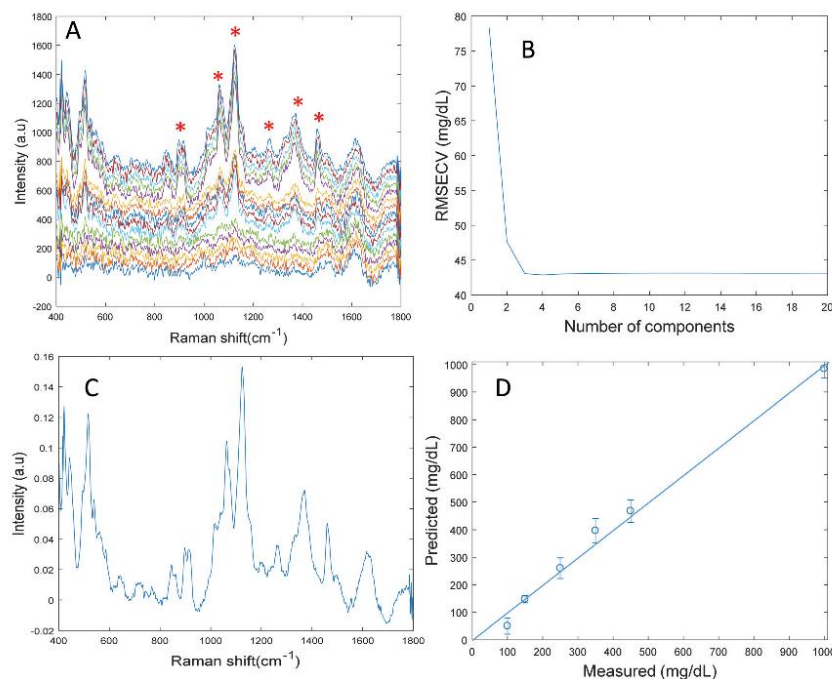


Fig. 2 (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10 kDa filters of varying concentrations of glucose ($5 \times 100 \text{ mg dL}^{-1}$, $5 \times 450 \text{ mg dL}^{-1}$ and $5 \times 1000 \text{ mg dL}^{-1}$, offset for clarity), in distilled water and signature peaks of glucose are highlighted with asterisks, (B): evolution of the RMSECV on the validation model, (C): plot of PLSR coefficient with glucose features, (D): predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation. The RMSECV and R^2 values were calculated as 10.93 mg dL^{-1} and 0.9705 respectively.

serum samples with spiked glucose concentrations varying from 0 mg dL^{-1} to 220 mg dL^{-1} . When the entire fingerprint region was selected for PLSR analysis, the resultant PLSR coefficient displayed a negative peak at $\sim 1000 \text{ cm}^{-1}$ which could potentially derive from other LMWF species such as urea⁶³ (Fig. S1C in ESI†). Therefore, the spectral region from 1030 cm^{-1} to 1400 cm^{-1} , which contains strong glucose features at $\sim 1050 \text{ cm}^{-1}$ and at $\sim 1340 \text{ cm}^{-1}$ but minimal interference from urea, has been chosen for PLSR analysis.⁶²

Fig. 3A shows the glucose data set after background correction using the EMSC algorithm (offset for clarity by 5.0 units). The optimum number of latent variables were chosen by calculating the lowest value of RMSECV (Fig. 3B). Six latent variables were chosen for this model and the resultant PLSR coefficient exhibits strong glucose features, as shown in Fig. 3C. A linear predictive model can be defined from the EMSC corrected data set of varying concentration of glucose in serum Fig. 3D. The RMSECV was found to be 1.66 mg dL^{-1} and R^2 value was calculated as 0.9914 . The mean standard deviation of each measurement from 20 iterations of cross validation was calculated to be 3.2 mg dL^{-1} . The results suggest that this optimised proto-

col can be applied to the patient samples to build a quantitative model.

Bonnier *et al.* have previously illustrated the strategy of centrifugal filtration to fractionate the samples to eliminate the influence of HMW in order to screen potential LMWF biomarkers using glucose as a model analyte, measured by ATR-FTIR.²⁹ The results of the analysis of similarly centrifugally filtered, glucose spiked samples of commercial serum are reproduced in Table 2, and directly compared to the results of the current study using Raman spectroscopic analysis in the native liquid state. Although each method yields similar R^2 values, the RMSECV of Raman spectroscopic analysis is significantly lower, which indicates an increased accuracy of the measurement protocol in the liquid state.

3.3 Monitoring the glucose concentration in patient samples

Fig. 4A displays the Raman spectra from patient samples after performing background correction using the EMSC algorithm, in the reduced spectral range of $1030\text{--}1400 \text{ cm}^{-1}$. The spectra are offset for clarity by 5.0 units. The glucose bands at 1340 cm^{-1} and 1460 cm^{-1} , related to the $\delta(\text{C-C-H})$ vibration

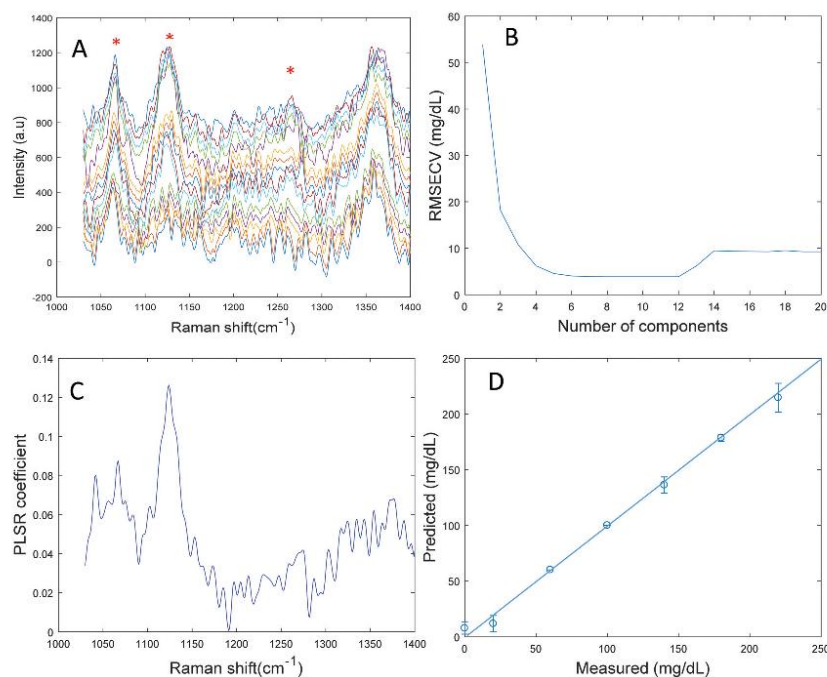


Fig. 3 (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10 kDa filters of glucose spiked in serum (spiked concentrations $5 \times 0 \text{ mg dL}^{-1}$, $5 \times 120 \text{ mg dL}^{-1}$ and $5 \times 220 \text{ mg dL}^{-1}$, offset for clarity) and the signature peaks of glucose are highlighted by asterisks, (B): evolution of RMSECV of the data set (C): plot of PLSR coefficient with glucose features, (D): predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation. The RMSECV and R^2 values were calculated as 1.66 mg dL^{-1} and 0.9914.

Table 2 Comparison of the results of ATR-FTIR and Raman spectroscopic analysis of human serum spiked with varying concentrations of glucose

Measurement type	Concentration range (mg dL^{-1})	RMSECV (mg dL^{-1})	Standard deviation	R^2
FTIR (min-max normalised)	0–220	2.199	0.250	0.995
Raman spectroscopy	0–220	1.665	3.2	0.991

and pure CH_2 group vibration, and the peak at 1060 cm^{-1} due to $\nu(\text{C1-OH})$ stretching can be clearly seen. Based on the minimum RMSECV value (Fig. 4B), 12 latent variables were found to be optimal for constructing a PLSR based model. The PLSR coefficient clearly shows glucose features (Fig. 4C), indicating that the prediction is based on the variation in the glucose peak intensities. In the prediction model of Fig. 4D, the minimum value of RMSECV is 1.84 mg dL^{-1} and the R^2 value is calculated as 0.82, indicating a high prediction capacity. The mean standard deviation was determined to be

3.48 mg dL^{-1} , indicating acceptable repeatability between the cross validation iterations.

Table 3 directly compares the results of analysis of a similar patient sample-set, with glucose levels which varied over the range $0\text{--}210 \text{ mg dL}^{-1}$, similarly centrifugally filtered and analysed using ATR-FTIR²⁹ and the results of the current study of patient samples over the same (patients 1–21, Table 1) and extended (patients 1–25, Table 1) range using Raman spectroscopic analysis in the native liquid state. It is noteworthy that Raman spectroscopy yields significantly lower values of RMSECV for all the Raman prediction models, suggesting higher sensitivity and accuracy. However, the standard deviation was found to be higher and the R^2 value was found to be considerably lower for the shorter as well as larger concentration range. The reduced R^2 value and higher standard of deviation could be attributed to the variability in the spectral response of the patient samples. Nevertheless, the precision of the model expressed by the RMSECV values indicates the suitability of this technique to discriminate patients with very close concentrations of blood glucose. The results demonstrate that Raman spectroscopy is able to detect subtle variations in

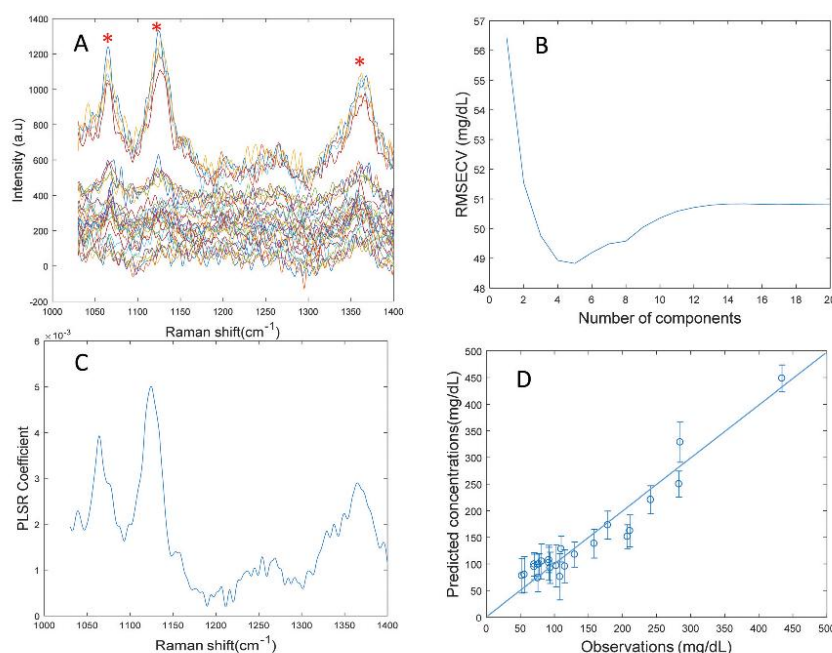


Fig. 4 (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10 kDa filters of patient samples ($5 \times 52.25 \text{ mg dL}^{-1}$, $5 \times 75.67 \text{ mg dL}^{-1}$, $5 \times 93.69 \text{ mg dL}^{-1}$, $5 \times 210.81 \text{ mg dL}^{-1}$ and $5 \times 434.35 \text{ mg}$, offset for clarity) and the signature peaks are marked by asterisks, (B): evolution of RMSECV of the data set, (C): plot of PLSR coefficient with glucose features, (D): predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation. The RMSECV and R^2 values were calculated as 1.84 mg dL^{-1} and 0.84 respectively.

Table 3 Comparison of the results of ATR-FTIR²⁹ and Raman spectroscopic analysis of patient sample set for monitoring the glucose levels. FTIR results are normalised

Measurement type	Concentration range (mg dL^{-1})	RMSECV (mg dL^{-1})	Standard deviation (mg dL^{-1})	R^2
ATR-FTIR (min-max normalised)	61.25–210	3.1	1.90	0.9957
Raman	52.25–210	1.6	2.31	0.91
Spectroscopy	52.25–440	1.84	3.48	0.84

the glucose concentrations with similar accuracy in the native liquid state, to the ATR-FTIR method in dried samples.

In order for a glucose detection method to be viable, it should be able to detect glucose in the clinically relevant range ($10\text{--}450 \text{ mg dL}^{-1}$). Post PLSR analysis, the dataset is presented in the Clarke's error grid (Fig. 5), the most common standard for evaluating the performance of a glucose detection method used since 1987.⁶⁴ Data points that fall in zone A and B are acceptable values. Values that fall outside A and B result in erroneous diagnosis. On the Clarke's error grid of patient

samples (Fig. 5), 98% of the PLSR validation dataset falls within zone A and B, which is the zone of clinical accurate measurement with no effect on clinical actions. Error can be attributed to the intrinsic variability of patient samples which reflects their physiological state on the day. However, the results from this study are promising, indicating that Raman spectroscopy coupled with multivariate analysis and centrifugal filtration techniques can be used as a biochemical tool for detecting potential small biomarkers from human serum/plasma.

4. Discussion

The potential advantages of using vibrational spectroscopy for biomarker assessment in bodily fluids have been extensively explored in the last two decades. However, little consideration has been given so far to protocols involving Raman analysis in the native liquid state of proteins. Liakat *et al.* reported the *in vitro* prediction of physiologically relevant concentration of glucose using mid-IR transmission light with respect to a Clarke error grid.⁵² Near IR and Raman spectroscopy was also

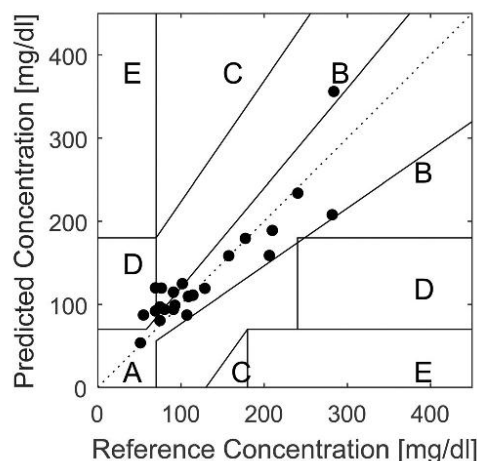


Fig. 5 PLSR validation of patient samples on Clarke's error grid. The RMSECV was found to be 1.84 mg dL^{-1} and R^2 value was calculated as 0.84.

used for measurement of glucose from artificial plasma with high precision and accuracy.⁶⁵ Although these are initial steps towards developing Raman spectroscopy into a biochemical tool for serum/plasma analysis, measurements should be performed on patient samples in order to ensure the relevancy of these methods. Using Raman spectroscopy as a biochemical tool, it has been possible to detect differences in peak intensities of altered serum compared to normal ones for glucose and lipid compounds,⁶⁶ multicomponent blood analysis⁶⁷ and also to determine blood glucose concentration of blood samples with above-physiological levels of glucose within 5 min.⁶⁸

Using glucose as a model analyte, this study successfully demonstrated the feasibility of employing Raman spectroscopy for detecting small biomarkers in serum after depletion of HMWF proteins. It has been shown that optimal experimental set up for Raman analysis for this experiment is Lab-Tek plates as substrate and measurement in the inverted geometry using water immersion objective and the sample volume can be as small as $1 \mu\text{L}$. This experimental set up is advantageous for clinical purposes where the volumes of patient samples are minimal. After the depletion of the abundant proteins, the dominant water peak from the filtrate collected after centrifugal filtration using 10 kDa can be removed by using the EMSC algorithm, and PLSR analysis applied to obtain a prediction model relating the glucose concentrations and the intensity of glucose features. Even though the EMSC algorithm removed the underlying water spectra effectively, there could be interference from other LMWF analytes, namely, urea ($7\text{--}20 \text{ mg dL}^{-1}$). Thus, as presented in the present study, the spectral range from 1030 cm^{-1} to 1400 cm^{-1} was chosen for data

analysis, as this region does not contain signature peaks of urea.

The depletion of HMWF proteins using centrifugal filtration to detect glucose in serum using ATR-FTIR was previously reported by Bonnier *et al.*²⁹ While the work carried out by Bonnier *et al.* showed excellent results, the requirement of a drying step is potentially a major drawback. Indeed, the drying process required for ATR-FTIR has been identified by Cameron *et al.* as a potentially significant impediment to translation of the technique to clinical applications.⁴¹ Since Raman is compatible with aqueous samples, sample drying can be avoided and data can be recorded from the native environment, which makes the proposed method an ideal alternative to IR. The results summarised in Table 2 and Fig. 3 suggest that Raman spectroscopy maintained high level of accuracy and predictive power and the relationship between spectral variation and protein concentrations is linear, with minimal standard deviation. The PLSR model built on varying concentration of glucose spiked in serum provided an accurate prediction model ($R^2 = 0.9914$, RMSECV = 1.66 mg dL^{-1}) after applying pre-processing steps using the EMSC based algorithm. Having established the optimal sample preparation and analysis protocol using the spiked serum model, the same protocol was applied to patient samples. In the case of patient samples, RMSECV and R^2 values were calculated to be 1.84 mg dL^{-1} and 0.84 respectively. Although ATR-FTIR provides better standard deviation and higher R^2 value, the Raman prediction model gives lower a RMSECV value, indicating higher accuracy. The PLSR coefficient plot shows clear features of glucose, indicating the prediction model is built on variations in glucose concentrations. The added advantage of Raman spectroscopy is that the analysis can be performed on liquid samples and no additional time delays associated with sample drying are introduced. Although the quantitative capability of Raman can be easily demonstrated using glucose spiked with a pooled serum model, the analysis of patient samples can be more complicated, the reason being the intrinsic variability of individual samples depending upon the physiological state of the individual on that day.

For the purposes of a direct comparison of the two techniques, glucose concentration in human serum was chosen, and the results detailed in Table 3 indicate that Raman in the liquid state provides higher accuracy than FTIR-ATR. Although vibrational spectroscopic techniques are unlikely to replace current techniques for routine glucose monitoring, in a more general sense, an argument for serological applications of vibrational spectroscopic techniques has been made.^{28,37} The advantages of using Raman spectroscopy over common biochemical assays to quantify urea and creatinine has previously been reported.⁶³ Moreover, Raman spectroscopy as a biochemical tool for serum analysis is cost effective, rapid and a non-destructive method as compared to currently employed gold standard clinical methods such as spectrophotometric analysis (e.g. COBAS analyser).⁶⁹ The COBAS analyser has a standard deviation of 0.04 mmol L^{-1} (which is equivalent to 0.721 mg dL^{-1}), as mentioned in the Material and Methods

section, and the lower detection limit is 2 mg dL⁻¹ (ref. 70) with a correlation coefficient of 0.975 reported for immunoassays⁷¹ and R^2 of 0.990 for urea, creatinine, sugar, total protein and calcium.⁷² Similar R^2 values were calculated for pure glucose solutions and glucose spiked serum solutions using Raman spectroscopy, suggesting that Raman spectroscopy is well-suited for routine use as a biochemical tool for glucose analysis. However, the analysis using COBAS analyser is complex and uses various enzymatic reagents, increasing the cost and chance of inaccuracy in the results obtained. Employing well trained personnel to operate the equipment enhances the reliability but increases the cost. Hence, Raman spectroscopy offers several advantages as it is a onetime investment, easy to operate and provides rapid results with wider information without destroying the medium. This could be translated as an alternative method for glucose monitoring, especially in the case of hyperglycaemia. Further studies need to be conducted to investigate other LMWF metabolites from human serum using Raman spectroscopy. However, to further ensure relevancy of the results, the study should ultimately be conducted on large number of patient samples.

5 Conclusion

Summarising, the work presented showcases the development of an optimal methodology for the detection of LMWF analytes from human serum using Raman spectroscopy with minimal sample preparation steps and without the use of any extrinsic labels. The proposed approach can be expeditiously employed for early detection of pathological disorders associated with high or low serum/plasma proteins/biomarkers. Disease diagnosis from bodily fluids can be developed into a dynamic diagnostic environment that will enable early disease diagnosis even before the disease becomes symptomatic. Thus, analysis of bodily fluids has emerged as one of the promising approaches to deliver crucial information about patient health and monitor disease progression and/or therapy. Given the remarkable advances in the field over the last two decades, including sample preparation, protein fractionation, quantitation and chemometrics, it is conceivable that vibrational spectroscopic techniques can be developed as a point-of-care disease monitoring system. Ultimately, the proof of concept presented in this study can be easily transferable to any other low molecular weight biomarkers or therapeutic drugs.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Drishya Rajan Parachalil was funded by DIT Fiosraigh scholarship. J. McIntyre was funded by Science Foundation Ireland, PI/11/08. The Irish-French PHC Ulysses 2018

Collaborative project is jointly funded by the Irish Research Council and Campus France with participation of the French embassy in Ireland.

References

- 1 T. D. Veenstra, T. P. Conrads, B. L. Hood, A. M. Awellino, R. G. Ellenbogen and R. S. Morrison, Biomarkers: Mining the Biofluid Proteome, *Mol. Cell. Proteomics*, 2005, **4**(4), 409–418.
- 2 K. Kong, C. Kendall, N. Stone and I. Nottingher, Raman spectroscopy for medical diagnostics - From in-vitro biofluid assays to in-vivo cancer detection, *Adv. Drug Delivery Rev.*, 2015, **89**, 121–134.
- 3 S.-B. Su, T. Chuen, W. Poon and V. Thongboonkerd, Human Body Fluid, *BioMed Res. Int.*, 2013, **2013**, 2–4.
- 4 J. T. Busher, Serum Albumin and Globulin, *Clin. Methods Hist. Phys. Lab. Exam.*, 1990, 497–499.
- 5 D. W. Greening and R. J. Simpson, Low-Molecular Weight Plasma Proteome Analysis Using Centrifugal Ultrafiltration, *Methods Mol. Biol.*, 2011, **728**, 109–124.
- 6 C. E. Parker and C. H. Borchers, Mass spectrometry based biomarker discovery, verification, and validation—Quality assurance and control of protein biomarker assays, *Mol. Oncol.*, 2014, **8**(4), 840–858.
- 7 M. J. Baker, C. S. Hughes and K. A. Hollywood, *Biophotonics: Vibrational Spectroscopic Diagnostics*, Morgan & Claypool Publishers, 2016.
- 8 J. Li, Z. Zhang, J. Rosenzweig, Y. Y. Wang and D. W. Chan, Proteomics and Bioinformatics Approaches for Identification of Serum Biomarkers to Detect Breast Cancer, *Clin. Chem.*, 2002, **48**(8), 1296–1304.
- 9 R. C. Bast, P. Ravdin, D. F. Hayes, S. Bates, H. Fritsche, J. M. Jessup, *et al.*, 2000 Update of Recommendations for the Use of Tumor Markers in Breast and Colorectal Cancer: Clinical Practice Guidelines of the American Society of Clinical Oncology, *J. Clin. Oncol.*, 2001, **19**(6), 1865–1878.
- 10 A. Hormigo, B. Gu, S. Karimi, E. Riedel, K. S. Panageas, M. A. Edgar, *et al.*, YKL-40 and matrix metalloproteinase-9 as potential serum biomarkers for patients with high-grade gliomas, *Clin. Cancer Res.*, 2006, **12**(19), 5698–5704.
- 11 F. Labrie, A. Dupont, R. Suburu, L. Cusan, M. Tremblay, J. L. Gomez, *et al.*, Serum prostate specific antigen as pre-screening test for prostate cancer, *J. Urol.*, 1992, **147**(3 Pt 2), 842–846.
- 12 W. J. Catalona, D. S. Smith, T. L. Ratliff, K. M. Dodds, D. E. Coplen, J. J. Yuan, *et al.*, Measurement of prostate-specific antigen in serum as a screening test for prostate cancer, *N. Engl. J. Med.*, 1991, **324**(17), 1156–1161.
- 13 G. Y. Locker, S. Hamilton, J. Harris, J. M. Jessup, N. Kemeny, J. S. Macdonald, *et al.*, ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer, *J. Clin. Oncol.*, 2006, **24**(33), 5313–5327.
- 14 D. B. Liesenfeld, N. Habermann, R. W. Owen, A. Scalbert and C. M. Ulrich, Review of mass spectrometry-based meta-

- bolomics in cancer research, *Cancer Epidemiol., Biomarkers Prev.*, 2013, **22**(12), 2182–2201.
- 15 T. Kimhofer, H. Fye, S. Taylor-Robinson, M. Thursz and E. Holmes, Proteomic and metabonomic biomarkers for hepatocellular carcinoma: A comprehensive review, *Br. J. Cancer*, 2015, **112**(7), 1141–1156.
 - 16 M. Nowak, L. Janas, G. Stachowiak, T. Stetkiewicz and J. R. Wilczyński, Current clinical application of serum biomarkers to detect ovarian cancer, *Przegl. Menopauzalny*, 2015, **14**(4), 254–259.
 - 17 J. L. Luque-Garcia and T. A. Neubert, Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry, *J. Chromatogr. A*, 2007, **1153**(1–2), 259–276.
 - 18 A. Fernandez-Olavarria, R. Mosquera-Perez, R. Diaz-Sanchez, M. Serrera-Figallo, J. Gutierrez-Perez and D. Torres-Lagares, The role of serum biomarkers in the diagnosis and prognosis of oral cancer: A systematic review, *J. Clin. Exp. Dent.*, 2016, **8**(2), 184–193.
 - 19 Z. J. Sahab, S. M. Semaan and Q.-X. A. Sang, Methodology and applications of disease biomarker identification in human serum, *Biomarker Insights*, 2007, **2**, 21–43.
 - 20 J. N. Adkins, S. M. Varnum, K. J. Auberry, R. J. Moore, N. H. Angell, R. D. Smith, *et al.*, Toward a Human Blood Serum Proteome, *Mol. Cell. Proteomics*, 2002, **1**(12), 947–955.
 - 21 R. L. Lundblad, Considerations for the use of blood plasma and serum for proteomic analysis, *Internet J. Genomics Proteomics*, 2005, **1**(2), 1–8.
 - 22 R. Pieper, C. L. Gatlin, A. J. Makusky, P. S. Russo, C. R. Schatz, S. S. Miller, *et al.*, The human serum proteome: Display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins, *Proteomics*, 2003, **3**(7), 1345–1364.
 - 23 C. Lacombe, V. Untereiner, C. Gobinet, M. Zater, G. D. Sockalingum and R. Garnotel, Rapid screening of classic galactosemia patients: A proof-of-concept study using high-throughput FTIR analysis of plasma, *Analyst*, 2015, **140**(7), 2280–2286.
 - 24 F. Di Girolamo, J. Alessandroni, P. Somma and F. Guadagni, Pre-analytical operating procedures for serum Low Molecular Weight protein profiling, *J. Proteomics*, 2010, **73**(3), 667–677.
 - 25 K. N. Patel, J. K. Patel, M. P. Patel, G. C. Rajput and H. A. Patel, *Introduction to hyphenated techniques and their applications in pharmacy*. Pharm Methods, Medknow Publications & Media Pvt Ltd, India, 2010, vol. 1(1), pp. 2–13.
 - 26 J. Kuligowski, M. Cascant, S. Garrigues and M. De La Guardia, An infrared spectroscopic tool for process monitoring: Sugar contents during the production of a depilatory formulation, *Talanta*, 2012, **99**, 660–667.
 - 27 A. Edelmann, J. Diewok, J. R. Baena and B. Lendl, High-performance liquid chromatography with diamond ATR-FTIR detection for the determination of carbohydrates, alcohols and organic acids in red wine, *Anal. Bioanal. Chem.*, 2003, **376**(1), 92–97.
 - 28 A. Ioannou, Real Time Monitoring the Maillard Reaction Intermediates by HPLC- FTIR, *J. Phys. Chem. Biophys.*, 2016, **6**(2), 6–10.
 - 29 F. Bonnier, H. Blasco, C. Wasselet, G. Brachet, R. Respaud, L. F. C. S. Carvalho, *et al.*, Ultra-filtration of human serum for improved quantitative analysis of low molecular weight biomarkers using ATR-IR spectroscopy, *Analyst*, 2017, **142**(8), 1285–1298.
 - 30 H. Byrne, G. Sockalingum and N. Stone, Raman Microscopy: Complement or Competitor, *Biomed. Appl. Synchrotron Infrared Microspectrosc.*, 2011, (11), 105–142.
 - 31 F. Bonnier, F. Petitjean, M. J. Baker and H. J. Byrne, Improved protocols for vibrational spectroscopic analysis of body fluids, *J. Biophotonics*, 2014, **7**(3–4), 167–179.
 - 32 F. Bonnier, M. J. Baker and H. J. Byrne, Vibrational spectroscopic analysis of body fluids: avoiding molecular contamination using centrifugal filtration, *Anal. Methods*, 2014, **6**(14), 5155.
 - 33 A. A. Bunaciu, Ş. Fleschin, V. D. Hoang and H. Y. Aboul-Enein, Vibrational Spectroscopy in Body Fluids Analysis, *Crit. Rev. Anal. Chem.*, 2017, **47**(1), 67–75.
 - 34 A. L. Mitchell, K. B. Gajjar, G. Theophilou, F. L. Martin and P. L. Martin-Hirsch, Vibrational spectroscopy of biofluids for disease screening or diagnosis: Translation from the laboratory to a clinical setting, *J. Biophotonics*, 2014, **7**(3–4), 153–165.
 - 35 K. T. Nyuwi, C. H. Gyan Singh, S. Khumukcham, R. Rangaswamy, Y. S. Ezung, S. R. Chittivolu, *et al.*, The role of serum fibrinogen level in the diagnosis of acute appendicitis, *J. Clin. Diagn. Res.*, 2017, **11**(1), PC13–PC15.
 - 36 I. O. Tekin, B. Pocan, A. Borazan, E. Ucar, G. Kuvandik, S. Ilikhan, *et al.*, Positive correlation of CRP and fibrinogen levels as cardiovascular risk factors in early stage of continuous ambulatory peritoneal dialysis patients, *Renal Failure*, 2008, **30**(2), 219–225.
 - 37 J. J. Stec, H. Silbershatz, G. H. Tofler, T. H. Matheney, P. Sutherland, I. Lipinska, *et al.*, Association of fibrinogen with cardiovascular risk factors and cardiovascular disease in the Framingham Offspring Population, *Circulation*, 2000, **102**(14), 1634–1638.
 - 38 R. A. S. Ariens, Elevated fibrinogen causes thrombosis, *Blood*, 2011, **117**(18), 4687–4688.
 - 39 L. F. Hong, X. L. Li, S. H. Luo, Y. L. Guo, C. G. Zhu, P. Qing, *et al.*, Association of fibrinogen with severity of stable coronary artery disease in patients with type 2 diabetic mellitus, *Dis. Markers*, 2014, **2014**, 485687.
 - 40 M. Paraskevaidi, C. L. M. Morais, K. M. G. Lima, J. S. Snowden and J. A. Saxon, Differential diagnosis of Alzheimer ' s disease using spectrochemical analysis of blood, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**(38), 929–938.
 - 41 J. M. Cameron, H. J. Butler, D. S. Palmer and M. J. Baker, Biofluid spectroscopic disease diagnostics: A review on the processes and spectral impact of drying, *J. Biophotonics*, 2018, **11**(4), 1–12.

- 42 R. A. Shaw, S. Kotowich, M. Leroux and H. H. Mantsch, Multianalyte Serum Analysis Using Mid-Infrared Spectroscopy, *Ann. Clin. Biochem.*, 1998 Sep 1, 35(5), 624–632.
- 43 K. Gajjar, J. Trevisan, G. Owens, P. J. Keating, N. J. Wood, H. F. Stringfellow, *et al.*, Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer, *Analyst*, 2013, 138(14), 3917–3926.
- 44 R. A. Shaw, S. Low-Ying, A. Man, K.-Z. Liu, C. Mansfield and C. B. Rileg, *et al.*, *Infrared Spectroscopy of Biofluids in Clinical Chemistry and Medical Diagnostics*, in *Biomedical Vibrational Spectroscopy*, John Wiley & Sons, Inc., 2007. pp. 79–103.
- 45 F. Bonnier, G. Brachet, R. Duong, T. Sojinrin, R. Respaud, N. Aubrey, *et al.*, Screening the low molecular weight fraction of human serum using ATR-IR spectroscopy, *J. Biophotonics*, 2016, 9(10), 1085–1097.
- 46 K. Spalding, F. Bonnier, C. Bruno, H. Blasco, R. Board, I. Benz-de Bretagne, *et al.*, Enabling quantification of protein concentration in human serum biopsies using attenuated total reflectance – Fourier transform infrared (ATR-FTIR) spectroscopy, *Vib. Spectrosc.*, 2018, 99, 50–58.
- 47 A. Nabers, L. Perna, J. Lange, U. Mons, J. Schartner, J. Güldenhaupt, *et al.*, Amyloid blood biomarker detects Alzheimer's disease, *EMBO Mol. Med.*, 2018, 1–11.
- 48 S. Roy, D. Perez-Guaita, D. W. Andrew, J. S. Richards, D. McNaughton, P. Heraud, *et al.*, Simultaneous ATR-FTIR Based Determination of Malaria Parasitemia, Glucose and Urea in Whole Blood Dried onto a Glass Slide, *Anal. Chem.*, 2017 May 16, 89(10), 5238–5245.
- 49 J. R. Hands, P. Abel, K. Ashton, T. Dawson, C. Davis, R. W. Lea, *et al.*, Investigating the rapid diagnosis of gliomas from serum samples using infrared spectroscopy and cytokine and angiogenesis factors, *Anal. Bioanal. Chem.*, 2013, 405(23), 7347–7355.
- 50 D. R. Parachalil, B. Brankin, J. McIntyre and H. J. Byrne, Raman spectroscopic analysis of high molecular weight proteins in solution – considerations for sample analysis and data pre-processing, *Analyst*, 2018, 143(24), 5987–5998.
- 51 T. L. Daines and K. W. Morse, Determination of Glucose in Blood Serum, *J. Chem. Educ.*, 1976, 53(2), 126–127.
- 52 S. Liakat, K. A. Bors, T. Huang, A. P. M. Michel, E. Zanghi, C. F. Gmachl, *et al.*, In vitro measurements of physiological glucose concentrations in biological fluids using mid-infrared light, *Biomed. Opt. Express*, 2013, 4(7), 233–239.
- 53 J. Illingworth, *Methods of enzymatic analysis*, Biochemical Education, 3rd edn, 1985, vol. 13(1), p. 38.
- 54 C. Bell, Clinical Guide to Laboratory Tests, in *Transfusion*, ed. N. W. Tietz, Wiley/Blackwell, 3rd edn, 2018, vol. 35(11), p. 972.
- 55 L. T. Kerr and B. M. Hennelly, A multivariate statistical investigation of background subtraction algorithms for Raman spectra of cytology samples recorded on glass slides, *Chemom. Intell. Lab. Syst.*, 2016, 158, 61–68.
- 56 A. Kohler, C. Kirschner, A. Oust and H. Martens, Extended multiplicative signal correction as a tool for separation and characterization of physical and chemical information in Fourier transform infrared microscopy images of cryo-sections of beef loin, *Appl. Spectrosc.*, 2005 Jun, 59(6), 707–716.
- 57 K. H. Liland, A. Kohler and N. K. Afseth, Model-based pre-processing in Raman spectroscopy of biological samples, *J. Raman Spectrosc.*, 2016, 47(6), 643–650.
- 58 L. Joss and E. A. Müller, Machine Learning for Fluid Property Correlations: Classroom Examples with MATLAB, *J. Chem. Educ.*, 2019, 96(4), 697–703.
- 59 S. Xianchun, X. Liang, Y. Shubin, H. Rong, J. Ling, X. Hanyang and L. Wenqing, Automatic baseline correction method for the open-path Fourier transform infrared spectra by using simple iterative averaging, *Opt. Express*, 2018, 26(10), 609–614.
- 60 S. Wold, M. Sjöström and L. Eriksson, PLS-regression: A basic tool of chemometrics, *Chemom. Intell. Lab. Syst.*, 2001, 58(2), 109–130.
- 61 S. Russell and P. Norvig, in *Artificial Intelligence: A Modern Approach*, Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edn, 2009.
- 62 S. Söderholm, Y. H. Roos, N. Meinander and M. Hotokka, Raman spectra of fructose and glucose in the amorphous and crystalline states, *J. Raman Spectrosc.*, 1999, 30(11), 1009–1018.
- 63 C. J. Saatkamp, M. L. de Almeida, J. A. M. Bispo, A. L. B. Pinheiro, A. B. Fernandes and L. Silveira, Quantifying creatinine and urea in human urine through Raman spectroscopy aiming at diagnosis of kidney disease, *J. Biomed. Opt.*, 2016, 21(3), 037001.
- 64 W. L. Clarke, D. Cox, L. A. Gonder-Frederick, W. Carter and S. L. Pohl, Evaluating Clinical Accuracy of Systems for Self-Monitoring of Blood Glucose, *Diabetes Care*, 1987 Sep 1, 10(5), 622 LP–622628.
- 65 J. Xue, H. Chen, D. Xiong, G. Huang, H. Ai, Y. Liang, *et al.*, Noninvasive Measurement of Glucose in Artificial Plasma with Near-Infrared and Raman Spectroscopy, *Appl. Spectrosc.*, 2014, 68(4), 428–433.
- 66 R. De Cássia, F. Borges, R. S. Navarro, H. E. Giana, F. G. Tavares, A. B. Fernandes, *et al.*, Detecting alterations of glucose and lipid components in human serum by near-infrared Raman spectroscopy, *Res. Biomed. Eng.*, 2015, 31(2), 160–168.
- 67 A. J. Berger, T. Koo, I. Itzkan, G. Horowitz and M. S. Feld, Multicomponent blood analysis by near-infrared Raman spectroscopy, *Appl. Opt.*, 1999, 38(13), 2916–2926.
- 68 A. J. Berger, I. Itzkan and M. S. Feld, Feasibility of measuring blood glucose concentration by near-infrared Raman spectroscopy, *Spectrochim. Acta, Part A*, 1997, 53, 287–292.
- 69 K. Imai, Clinical Chemistry and Immunoassay Testing Supporting the Individual Healthy Life, *Hitachi Rev.*, 2008, 57, 1–7.
- 70 S. Pas, R. Molenkamp, J. Schinkel, S. Rebers, C. Copra, S. Seven-Deniz, *et al.*, Performance evaluation of the new

- Roche cobas AmpliPrep/cobas TaqMan HCV test, version 2.0, for detection and quantification of hepatitis C virus RNA, *J. Clin. Microbiol.*, 2013 Jan, 51(1), 238–242.
- 71 A. J. Van Gammeren, N. Van Gool, M. J. M. De Groot and M. Christa, Analytical performance evaluation of the Cobas 6000 analyzer – special emphasis on trueness verification, *Clin. Chem. Lab. Med.*, 2008, 46(6), 863–871.
- 72 A. Ahmed, J. M. Alam, H. Ali, I. Sultana and S. Fazal, Comparative precision analysis of Precision-controls on automated chemistry analyzers, *International Journal of Innovative Science Engineering and Technology*, 2015, 2(8), 1–13.

Discussion et perspectives

Analyse en solutions aqueuses

Le spectre Raman d'une solution de glucose a montré plusieurs modes vibrationnels à 450cm^{-1} , 911cm^{-1} , 1125cm^{-1} , 1340cm^{-1} et 1460cm^{-1} . Nous avons observé que l'intensité des modes vibrationnels augmente avec l'augmentation des concentrations en glucose des solutions. Le modèle permettant de prédire les concentrations de solutions contenant entre 1 g/L et 10 g/L de glucose a montré une erreur de 109,3 mg/L. De plus, la composante principale du modèle confirme que la prédiction est basée sur bandes spectrales spécifiques du glucose. On retrouve en effet les 5 signaux décrits précédemment.

Méthode des ajouts dosés

Dans un deuxième temps, nous avons surchargé un sérum commercial avec du glucose. Les concentrations des échantillons obtenus variaient de 0 g/L à 2,2 g/L. Lors de l'analyse sur les échantillons plasmatiques surchargés, le modèle de prédiction a permis d'obtenir une meilleure prédiction avec une erreur moyenne de 16,6 mg/L. Le modèle de prédiction n'utilisait cette fois-ci que le mode vibrationnel à 1125cm^{-1} . Ce résultat met en évidence un manque de sensibilité de la technique pour les autres modes de vibration et/ou une interférence de la matrice. En effet, les concentrations inférieures testées en solution aqueuses étaient nettement supérieures à celles des échantillons surchargés.

Analyse des échantillons patients

Le résultat obtenu sur les échantillons surchargés a été confirmé par ceux obtenus par l'analyse des échantillons patients. Nous avons observé 3 modes de vibrations à 1060cm^{-1} , 1125 cm^{-1} et 1360cm^{-1} . Malgré une difficulté supérieure pour extraire les informations spécifiques du spectre d'émission Raman, l'erreur moyenne de prédiction mesurée de 18,4 mg/L, soit environ 1 mmol/L. Ce résultat montre une imprécision de mesure similaire entre la spectroscopie d'émission Raman sur échantillon liquide après ultrafiltration centrifuge et la spectroscopie d'absorption ATR-IR sous forme séchée sans préparation préalable (0,85 mmol/L). Comme pour la spectroscopie infrarouge, l'analyse par spectroscopie Raman semble intéressante pour la réalisation d'une analyse de type métabolomique mais montre ses limites dans ces conditions pour une analyse quantitative. En effet, les imprécisions de mesures sont supérieures à celles recommandées pour le suivi du glucose en routine (197) mais l'absence d'impact sur l'interprétation biologique permet de positionner cette technologie comme un outil de dépistage.

Etude n°2 : Analyse par spectroscopie Raman des fractions de bas poids et de haut poids moléculaires du plasma humain

Lors de nos essais de filtration en ATR-IR, nous avons établi que parmi les différents biomarqueurs analysés, les protéines ayant un poids moléculaire supérieur au seuil des filtres sont complètement retenus par le filtre et ne passent pas dans le filtrat. D'autre part, nous avons également observé que les molécules de bas poids moléculaires sont peu retenues. L'étude précédente a permis de mettre en évidence des résultats satisfaisants pour une analyse semi-quantitative du glucose par spectroscopie Raman. Cependant, la qualité de l'analyse de la fraction de bas poids moléculaire ne peut pas être estimée uniquement par une étude réalisée sur le glucose. Cette étude nécessite donc la confirmation sur un panel de biomarqueurs plus large. D'autre part, l'analyse de la fraction de haut poids moléculaire est également une source d'informations importante en biologie ainsi que dans le contexte de la recherche de biomarqueurs dans la SLA. Dans ce contexte, il est important d'évaluer la relation entre les signaux mesurés et les concentrations protéiques de l'échantillon, d'autant plus que la perte d'eau dans le concentrat est susceptible de faire varier les concentrations sans conservation des proportions.

Nous avons donc réalisé une étude similaire à la précédente en étudiant cette fois-ci d'autres composants de la fraction de bas poids moléculaire ainsi que des biomarqueurs de haut poids moléculaire. Pour ce faire, nous avons mesuré les concentrations en protéines totales, albumine, immunoglobulines G, glucose et urée de 25 échantillons de patients par des méthodes de mesures validées et accréditées sur des chaînes automatisées cobas® c501 de chez Roche® (Bâle, Suisse).

L'ensemble des méthodes respecte les critères de validation de méthodes publiées par la Société Française de Biologie Clinique (SFBC) (197). Les concentrations mesurées font office de concentration de référence pour l'étude.

Ces mêmes échantillons ont été également filtrés puis analysés par spectroscopie Raman à celles mesurées selon le protocole utilisé dans l'étude précédente. Ensuite, la moitié des spectres mesurés a été utilisée pour construire le modèle PLSR (Partial Least Square Regression). Les concentrations correspondantes à l'autre moitié des spectres ont ensuite été prédites et comparées aux concentrations de références.

Cite this: *Analyst*, 2019, **144**, 4295

Raman spectroscopic screening of high and low molecular weight fractions of human serum†

Drishya Rajan Parachalil,^{a,b} Clément Bruno,^{c,d,e} Franck Bonnier,^c H          ,^{d,e} Igor Chourpa,^c Jennifer McIntyre^a and Hugh J. Byrne^a

This study explores the suitability of Raman spectroscopy as a bioanalytical tool, when coupled with ultra-filtration and multivariate analysis, to detect imbalances in both high molecular weight (total protein content, γ globulins and albumin) and low molecular weight (urea and glucose) fractions of the same samples of human patient serum, in the native liquid form. Ultracentrifugation was employed to separate and concentrate the high and low molecular weight fractions of the serum. Initially, aqueous solutions of the respective molecular species, covering physiologically relevant concentration ranges, were analysed to optimise the measurement protocols. An adapted Extended Multiplicative Signal Correction (EMSC) algorithm was applied to raw spectra to remove water background signal and spectral interferents (β -carotene). Using a validated partial least squares regression modelling method, R^2 values, Root Mean Square Error of Cross Validation (RMSECV) and standard deviations were established for the quantification of γ globulin, total protein, albumin, urea and glucose content of the patient serum samples. The study demonstrates that Raman spectroscopy in the liquid form is a viable alternative and/or adjunct to current clinical practice for the parallel analysis of high and low molecular weight fractions, and simultaneous analysis of multiple analytes in the low molecular weight fraction, of human serum for diagnostic applications.

Received 3rd April 2019,
Accepted 7th June 2019
DOI: 10.1039/c9an00599d

rsc.li/analyst

1. Introduction

Increasingly, medical diagnostic tests are performed on bodily fluids, as sample collection is minimally invasive, compared to histology or cytology based techniques. However, conventional methods for biofluid analysis can be inconsistent and are associated with high cost and long-time delays.^{1,2} Therefore, there is a need for more sensitive and cost effective methods with higher accuracy for early disease diagnosis from bodily fluids in a point-of-care clinical setting. Vibrational spectroscopic techniques, both Infrared absorption and Raman scattering, are among the most important analytical techniques available to scientists, as they provide detailed molecularly specific fingerprints of both organic and inorganic compounds, without the use of extrinsic labels and without being

extremely invasive or destructive to the system studied. Since both techniques are truly label-free, their potential for medical diagnostic applications has been well investigated and demonstrated.^{3–7} Such techniques are particularly attractive for routine analysis of biofluids, as they are easy to apply, require minimal sample preparation and are readily adaptable to analysis of various bodily fluids.^{1–5} Notably, it has been demonstrated that, whereas IR analysis of biofluids usually requires a drying step,^{8–10} Raman spectroscopic analysis can be applied to bodily fluids in their native liquid state.^{11,12} As water is a relatively weak scatterer, Raman spectroscopy is much more amenable than IR for analysis of biofluids in their native aqueous state,^{10,11} and its potential has been well demonstrated, notably in human serum and plasma.^{13–16}

Estimating serum total protein, γ globulin and albumin content is important to assess the nutritional status of patients.^{17,18} Although total serum protein estimation has limited diagnostic potential when compared to albumin or globulin, its relevance in the evaluation of patients with clinical conditions such as malnutrition, renal malfunction, liver diseases and immune disorders cannot be ignored.^{18,19} A normal level of total serum protein (6000–8300 mg dL^{−1}) indicates healthy nutritional status and normal liver function. Reduced serum total protein is predominantly found in patients with kidney disorders, HIV and aging in the elderly.^{20,21} The biuret

^aFOCAS Research Institute, Technological University Dublin, Kevin Street, Dublin 8, Ireland. E-mail: drishyarajan.parachalil@mydit.ie^bSchool of Physics and Optometric & Clinical Sciences, Technological University Dublin, Kevin Street, Dublin 8, Ireland^cUniversit   de Tours, UFR sciences pharmaceutiques, EA 6295 Nanom  dicaments et Nanosondes, 31 avenue Monge, 37200 Tours, France^dCHRU de Tours, Laboratoire de Biochimie et Biologie Mol  culaire, Tours, France^eUniversit   de Tours, iBrain, UMR INSERM U1253, 37032, France

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9an00599d

assay is the most common method used to quantify total protein levels in blood serum.¹

γ globulins (<150 kDa,²² ~38% of serum proteins), produced by lymphocytes and plasma cells in lymphoid tissue, are large protein molecules that include the immunoglobulins: IgM, IgA, and IgG.²³ The most characteristic abnormality in serum proteins in liver diseases and carcinoma of the gastrointestinal tract or breast is an elevation in γ -globulins.^{24,25} Testing globulin levels in serum routinely provides key information that helps diagnose various conditions and diseases that affect the immune status. Liver diseases, chronic inflammatory diseases, haematological disorders, infections and malignancies cause excess globulin levels (>1600 mg dL⁻¹),²⁶ whereas humoral immunodeficiencies cause low globulin levels (<700 mg dL⁻¹).²⁷ Radial immunodiffusion (RID)²⁸ is the gold standard method for measuring globulins. All the conventional methods used for testing total protein content and globulin measurement make use of expensive disposables and are labour intensive. With escalating medical costs and budget constraints, a cost effective alternative technology is desirable.

Albumin is the most abundant high molecular weight fraction (HMWF) serum protein, normally constituting about 50% of the total serum content, and has a molecular weight of 66 kDa.²⁹ The normal concentration of albumin in the human body is 3000 mg dL⁻¹, although it dramatically decreases in critically ill patients and does not increase again until the recovery phase of the illness.¹⁷ Several studies have demonstrated that the functions of albumin, such as ligand binding and transport of various molecules, can be applied to the treatment of cirrhotic patients and patients suffering from other end stage liver diseases.^{30–32} Strong correlation between cardiovascular disease and the level of albumin concentration have also been reported,^{33–38} which means a normal concentration of albumin in bodily fluids is considered a sign of good health. It is clear therefore that closely monitoring the variation in albumin concentration could act as an indicator of liver diseases and other related pathologies.^{39–44} Conventional methods used to determine the level of albumin include absorption spectroscopy or electrochemical based assays, immunoassays and high performance liquid chromatography (HPLC) and can be time consuming and very expensive.^{20,26–32} Therefore, a sensitive, rapid, cost-effective method such as Raman spectroscopy is highly desirable to quantitatively analyse albumin.

Urea is a colourless crystalline compound of molecular weight 60.056 g mol⁻¹ (60.056 Da) and is the main nitrogenous by-product produced by the liver when the body metabolises proteins.⁴⁵ The normal concentration of urea in human blood serum is 5–20 mg dL⁻¹. The kidneys normally filter out this waste from the body, and therefore it is important to monitor blood urea as higher (<20 mg dL⁻¹) or lower levels (>5 mg dL⁻¹) could indicate various pathologies such as kidney or liver malfunction.^{46–49} The standard technique for assessment of urea from blood or urine is based on colorimetric measurement, by which a specific reagent reacts with the sample and

absorption at a given wavelength is used to identify urea concentration.^{48,50}

Glucose is a low molecular weight carbohydrate that must be monitored on a regular basis in the case of diabetes patients.⁵¹ The normal concentration of glucose in blood is 70 to 130 mg dL⁻¹ and changes in this level would lead to hyper (>130 mg dL⁻¹) or hypoglycaemia (<70 mg dL⁻¹).⁵¹ While currently available methods of routinely self-monitoring blood glucose have become low cost, such glucose meters often suffer from significant errors that can lead to unreliable results.^{52,53} Hence, there is an unmet need for an accurate glucose monitoring tool, notably in a clinical setting.

Over the past decades, there have been numerous studies of analytes in biofluids using vibrational spectroscopy, and, in recent years, attenuated total reflection (Fourier Transform) IR (ATR-FTIR) has become popular for rapid screening of biofluids, particularly blood serum.⁵⁴ Multianalyte serum analysis has previously been reported using mid-infrared spectroscopy, for the simultaneous quantitation of eight serum analytes: total protein content, albumin, triglycerides, cholesterol, glucose, urea, creatinine and uric acid⁵⁵ and for simultaneous quantification of glucose and urea analytes along with malaria parasitemia quantification using ATR-FTIR.⁵⁶ Notably, however, both the analyses were conducted on dried serum samples. Berger *et al.*, as long ago as 1999, used Raman microscopic analysis of liquid whole human blood and serum samples to quantify the content of six analytes, namely glucose, cholesterol, triglyceride, urea, total protein and albumin.⁵⁷ Rohleder *et al.*, performed a direct comparison of Raman and FTIR spectroscopies of multiple analytes in human serum, concluding that the techniques produced similar accuracies.⁵⁸ More recently, Parachalil *et al.*, demonstrated that, using the same serum fractionation and regression analysis protocols, Raman microspectroscopic analysis in the liquid state performed at least as well as ATR-FTIR of dried samples, in the quantification of glucose levels in human serum.¹¹

This study aims to further evaluate the potential of Raman spectroscopy for the analysis of blood serum in the liquid state to simultaneously detect and quantify subtle variations in the whole serum (total protein content and γ globulins), HMWF (albumin) and LMWF (urea and glucose) as specific biomarkers linked to numerous pathologies, adding improved sampling techniques, sample fractionation and selected spectral ranges to improve over the previous work using ATR-FTIR and Raman spectroscopy.

Human serum is highly complex and diverse, however, due to its heterogeneous nature, consisting of dynamic range of biomolecules such as proteins, lipids, carbohydrates *etc.*⁵⁹ Raman spectra of serum are dominated by features of the abundant proteins (notably globulins and albumin) and water, making it difficult to visualise the spectral features of less abundant analytes and small molecules which can act as biomarkers. Both urea and glucose are classified as Low Molecular Weight Fraction (LMWF) analytes, and their low abundance makes it harder to detect these analytes from the

full serum samples. Therefore, measures are taken to make data interpretation easier, such as fractionation of serum using centrifugal filters prior to Raman analysis, followed spectral pre-processing.

2. Materials and methods

2.1 Sample preparation

γ globulins (G4386), albumin (A9511), urea (F3879) and β -carotene (C9750-5G) were purchased from Sigma Aldrich, Ireland. Individual solutions of varying concentrations of urea (1–1000 mg dL⁻¹) were prepared in distilled water by varying the concentrations over a physiologically relevant range.

Patient serum samples were donated by the University Hospital (CHRU) Bretonneau de Tours (France) and the ethical procedures were followed. The blood samples were collected from the individuals as routine blood check-ups and 1 mL per patient was provided for spectroscopic analysis. A total of 25 patient samples were included in the present study. Samples were collected by personnel of CHRU, under standard clinical protocols and approved ethical procedures (Comité de Protection des Personnes, Tours Region Central Oest 1-PP/ANSM-PHAO15-HB-METABOMU, registered internationally: ClinicalTrials.gov ID: NCT02670226). The samples were serologically profiled, for other purposes, and the anonymised, residual discard samples, along with their serological profiles were donated to the Université de Tours, for further study. No further specific ethical approval or patient consent is required. Albumin, γ globulins (IgG, IgM and IgA), total protein, urea and glucose concentrations were obtained by routine biochemical analysis using a COBAS analyser, following the CHRU guidelines for routine biochemical analysis. The principle for the test of albumin utilises the formation a blue-green complex to test for albumin, and a purple-coloured biuret complex to test for total protein. While albumin displays a sufficiently cationic character to be able to bind with an anionic dye, Bromocresol green (BCG), at a pH of 4.1⁶⁰ and divalent copper reacts in alkaline solution with protein peptide bonds in the case of total protein.⁶¹ The urea test is based on a coupled enzyme reaction (urease, followed by glutamate dehydrogenase), whereby measurement of NADH (converting to NAD⁺) is made at 340 nm.⁶² γ globulin concentrations were obtained using immunoturbidometric assays, and were summed to provide a value for total γ globulin content.⁶³ Measured analyte levels in the patient samples are listed in Table 1. Concentrations are expressed in mg dL⁻¹ for consistency with other studies.⁹ Notably, no correlation was found between the concentrations of HMWF and LMWF analytes in the patient serum samples (Fig. S1 and S2 in ESI†).

Amicon Ultra 0.5 mL centrifugal filter devices (Merck, Germany), of cut-off point 100 kDa, 50 kDa and 10 kDa, were employed to concentrate and separate the analytes of interest from the individual solutions and patient samples. The centrifugation procedure that has been previously reported by Bonnier *et al.* was followed.¹³ The optimised washing and

Table 1 Measured analyte levels in patient samples

Sample number	Total protein (mg dL ⁻¹)	γ globulin (mg dL ⁻¹)	Albumin (mg dL ⁻¹)	Urea (mg dL ⁻¹)	Glucose (mg dL ⁻¹)
1	6300	1327	3130	78.99	108.0
2	6900	1010	4510	15.97	81.0
3	6400	2254	2710	21.85	91.8
4	6700	1259	3580	7.00	115.2
5	5800	404	3050	8.40	241.2
6	5600	670	3120	5.88	207.0
7	7600	1395	4220	50.14	115.2
8	6700	712	4450	75.63	210.6
9	6000	875	3690	59.38	75.6
10	6200	1164	2970	5.88	282.6
11	6900	1409	3690	2.52	75.6
12	7500	1843	3910	29.41	102.6
13	5800	726	3880	21.57	129.6
14	6100	840	3340	10.64	93.6
15	4200	541	2600	11.20	178.2
16	7900	1478	3890	17.65	77.4
17	6600	1187	3950	10.92	91.8
18	7600	1201	4780	17.28	70.2
19	7700	1484	4420	26.33	158.4
20	6500	1396	3380	52.94	52.2
21	7400	1663	4360	8.12	109.8
22	7600	924	4250	39.50	70.2
23	6600	1556	3150	19.05	284.4
24	6500	1171	3750	8.12	55.8
25	7100	1236	4140	25.49	433.8

rinsing procedure includes spinning 0.5 mL 0.1 M NaOH at 14 000g for 30 minutes followed by three rinses with distilled water by spinning 0.5 mL distilled water for 30 minutes at 14 000g. Every 30 minutes wash and rinse must be followed by spinning the device in the inverted position at 1000g for 2 minutes, to remove the residual solution contained in the filter.

Fig. 1 is a schematic representation of the patient serum fractionation steps employed in this study. Non-fractionated serum was directly analysed to collect data for total protein content and γ globulin. In order to fractionate the patient serum samples, 0.5 mL sample was first spun for 30 minutes using 100 kDa filters to remove the analytes larger than 100 kDa (γ globulin) in the concentrate and the transmitted filtrate, which contains molecules smaller than 100 kDa, was used for further fractionation. For albumin isolation, the filtrate obtained was spun using 50 kDa filters for 30 minutes and the resultant concentrate was collected for Raman analysis. The filtrate transmitted by the 50 kDa filter was collected and was spun using 10 kDa filter for 30 minutes and the resultant filtrate was collected for Raman analysis to detect the presence of urea/glucose. Using a model sample based on glycine spiked serum, Bonnier *et al.* demonstrated the reproducibility of the HMW proteins depletion to be quite efficient and acceptable.⁸

2.2 Data collection using Raman spectrometer

The measurement conditions used were the same as those recently reported for screening HMWF proteins¹¹ and glucose⁶⁴ in solution. Raman spectra of the liquid samples

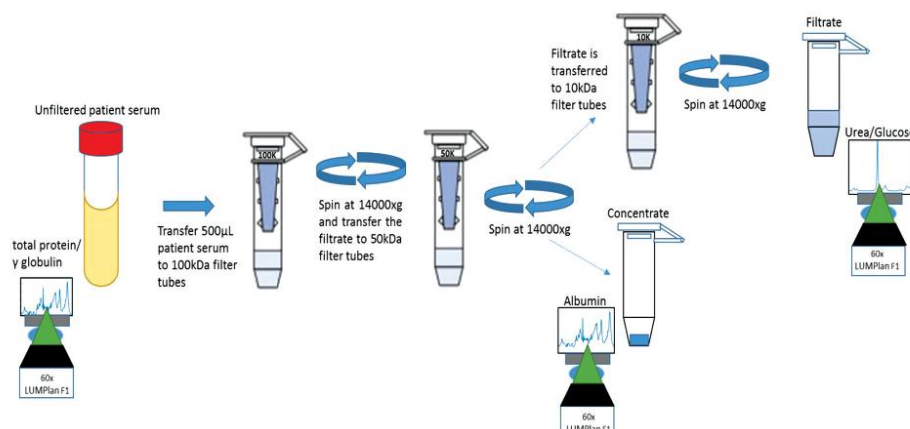


Fig. 1 Schematic overview of steps in fractionation of patient serum samples to separate γ globulin, albumin, and urea/glucose.

were recorded at stabilised room temperature (18 °C) using a Horiba Jobin-Yvon LabRam HR800 spectrometer with a 16-bit dynamic range Peltier cooled Charged Coupled Device. The spectrometer was coupled to an Olympus 1X71 inverted microscope and a $\times 60$ water immersion objective (LUMPlanF1, Olympus) was employed. A 532 nm laser was used, which had a power of ~ 30 mW at the sample, with a 600 lines per mm grating and the backscattered Raman signal was integrated for 3×80 seconds over the spectral range from 400–1800 cm^{-1} . The substrate used was a Lab-Tek plate (catalog number 154534) with a 0.16–0.19 mm thick glass bottom, 1.0 borosilicate cover glass, and was purchased from Thermo Fischer Scientific, Ireland.

2.3 Data pre-processing and analysis

The raw spectra were subjected to pre-processing techniques in Matlab before further analysis, to remove the background signal and reduce the noise. Smoothing of the raw data was done using the Savitzky–Golay method (polynomial order of 5 and window 13) and the rubberband method¹¹ was found to be appropriate to baseline correct the smoothed reference spectra of all the analytes and the smoothed spectra of varying concentrations of albumin spiked in distilled water.¹¹ The 'rubberband' correction was carried out by wrapping a 'rubberband' of defined length around the ends of the spectrum to be corrected and fitting against the curved profile of the spectrum. An adapted Extended Multiplicative Signal Correction (EMSC) algorithm⁶⁵ with a 5th order polynomial was applied to the raw dataset to remove the spectral interferences from the data. EMSC is applied to remove the underlying water spectrum from all the dataset, whose OH bending feature at 1640 cm^{-1} can interfere with the protein spectra and also scales the analyte spectra, assuming a constant water contri-

bution to all sample spectra.⁶⁵ The principle of EMSC for subtraction of a specific measurable background spectrum and the associated Matlab codes have previously been published by Kerr and Hennelly, 2016,⁶⁵ and their description is adapted in the following. The raw spectrum, S , consists of Raman spectrum of interest, R , a baseline signal, B , and the water signal, W .⁶⁵

$$S = R + B + W \quad (1)$$

The Raman spectrum of interest can be represented by a reference spectrum of the material of interest, r , and it can be assumed that R is the product of this reference spectrum and a certain scalar weight, C_r , which describes the concentration dependence^{65–67}

$$R \sim C_r \times r \quad (2)$$

Similarly, a spectrum, w , is recorded from water directly in order to represent the spectral contribution of water in W , as the product of pure water spectrum and a certain scalar weight.⁶⁵

$$W = C_w \times w \quad (3)$$

The baseline, B , is now represented by an appropriate order of polynomial (N) as:⁶⁵

$$B_N = C_0 + C_1X + C_2X^2 + \dots + C_NX^N \quad (4)$$

where N is the order of polynomial and C_m for $m = 0 \rightarrow N$ represents various coefficients of polynomial. The EMSC algorithm is used to obtain estimates of the scalar values C_r , C_m and C_w . These estimates are obtained from an optimal fit of the various vectors in eqn (5).⁶⁵

$$S \sim [C_r \times r] + [C_w \times w] + \left[\sum_{m=0}^N C_m X^m \right] \quad (5)$$

The background corrected, concentration dependent analyte spectra, T , can be represented as:⁶⁵

$$T = \frac{S - [C_w \times w] - \left[\sum_{m=0}^N C_m X^m \right]}{C_w} \quad (6)$$

Note, that division by C_w has the effect of scaling the analyte spectra, assuming a constant water contribution to all sample spectra.

In the case of total protein, γ globulins and albumin, EMSC is also applied to remove the β -carotene. Since the Lab-Tek plate is made up of thin glass bottom, no glass correction was required. Raman spectra of the pure serum, γ globulin ($\sim 100 \text{ mg mL}^{-1}$), albumin ($\sim 100 \text{ mg mL}^{-1}$), glucose (45 mg mL^{-1}) and urea ($\sim 100 \text{ mg mL}^{-1}$) prepared with minimal amount of water are used as the reference for EMSC.

2.3.1 Partial least squares regression. Partial Least Squares Regression (PLSR) is a multivariate statistical method which aims to establish a model that relates the variations of the spectral data to a series of relevant targets.

The PLSR model attempts to elucidate factors which account for the systematic majority of variation in predictors 'X' (spectral data) versus associated responses 'Y' (target values of protein concentration).⁶⁸ The spectral data (X matrix) is thus related to the targets (Y matrix) according to the linear equation $Y = XB + E$, where B is a matrix of regression coefficients and E is a matrix of residuals. The PLSR algorithm allows for the construction of a regression model which can be used to predict the outcome in varying concentration of analytes, and the performance of the PLSR model in predicting varying analytes concentration was evaluated in this study. In this case, the examples used are concentration and Raman signal, and therefore the algorithm can be used to predict the detection of Raman signal for a particular analytes concentration. The performance of the regression model in predicting varying concentrations of the analyte over a particular range was evaluated and this method can be employed to improve the limit of detection of Raman bio-sensing.⁶⁸ Constructed based on the spectra of samples of known analyte content, either solutions of varying concentrations in distilled water or those of the patient serum, the model is then validated using a rigorous cross validation procedure which evaluates its performance in accurately predicting analyte concentrations. For consistency with previous studies,⁶⁴ a 20 fold cross validation approach has been employed to validate the robustness of the method. This approach involves randomly dividing the set of observations into approximately equal size, $\sim 50\%$ of the spectral data were randomly selected as test set, while the remaining $\sim 50\%$ is used as the training set.⁶⁹ In the current case, (5×25 spectra) were divided into two groups of 65 (test) and 60 (training) spectra. The cross-validation process is then repeated 20 times (the folds), whereby all observations are used for both training and testing, and each observation is used for testing exactly once. The results from the folds can then be averaged to produce a single estimation. The Root

Mean Square Error of Cross Validation (RMSECV) is calculated from the 20 iterations to measure the performance of the model for the unknown cases within the calibration set. The correlation between the concentration and spectral intensity is given by the R^2 value. The standard deviation was calculated to find the variation between each spectrum calculated from the same sample. The number of latent variables used for building the PLSR model is optimised by finding the value that is equivalent to the minimum of the RMSECV.

3. Results and discussion

Fig. 2A shows the mean spectra of the whole serum samples derived from patients. For comparison, Fig. 3 shows the spectra of globulin and albumin. Unexpectedly, additional strong Raman bands at 1167 cm^{-1} and 1535 cm^{-1} (ref. 12 and 70) are observed, which dominate the protein features of serum, which can be ascribed to the presence of β -carotene (Fig. S3†). β -Carotene is a red-orange-coloured carotenoid widely distributed in fruits and vegetables and plays a major role in the maintenance of normal healthy skin, vision, immune system and mucous membranes.^{71,72} Regular intake of β -carotene rich foods results in higher levels in blood serum/plasma^{73,74} and is also associated with decreased risk of cardiovascular disease, heart disease, cancer, and other causes of mortality.^{75,76} As shown by a community-based study of French adults, serum β -carotene and vitamin C concentrations are useful biomarkers of vegetable and fruit consumption in the French diet.⁷⁷ β -Carotene features became less apparent in the spectra recorded from the concentrate of 50 kDa filtration, as shown in Fig. 2B and no β -carotene features were found in the spectra collected from the filtrate of 10 kDa filtration, as seen in Fig. 2C. The molecular size of

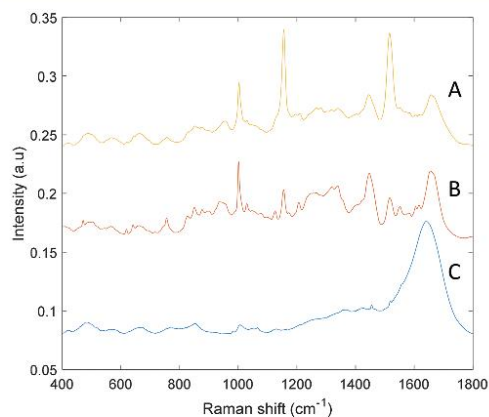


Fig. 2 Raman spectra of patient serum collected using Raman spectroscopy. (A) Whole serum, (B) concentrate from 50 kDa filtration and (C) filtrate from 10 kDa filtration. Spectra have been off set for clarity.

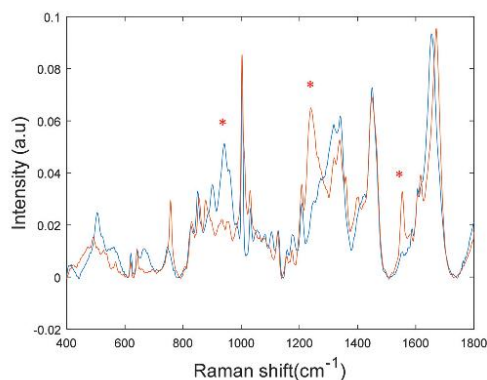


Fig. 3 Spectra of γ globulins (red – ~38% of serum) and albumin (blue – ~50% of serum) showing similar spectral features. Identifying signature peaks of γ globulins at 1240 cm^{-1} and 1553 cm^{-1} and of albumin at 940 cm^{-1} are highlighted with asterisks.

β -carotene is less than 10 kDa (~537 Da),⁷⁸ and therefore the presence of β -carotene in the concentrate of 50 kDa rather than filtrate of 10 kDa must be attributed to the binding of β -carotene to albumin.⁷⁹ Some studies have shown high binding affinity of β -carotene to albumin due to its identical left-right symmetric structure that has great potential for association of albumin molecules and protects β -carotene against oxidative degradation.^{79,80} In terms of quantitative analysis of the high molecular weight constituents, the strong contribution of the β -carotene is undesirable. Therefore, EMSC was applied to subtract the water signal, β -carotene signal as well as background noise from the initial serum data, and from that of the 50 kDa filtered data.

3.1 Quantification of total protein concentration and γ globulins in whole serum

A total serum protein test analysis measures the total amount of protein in the serum, as well as the amount of two major proteins in the serum; albumin and globulin. The normal concentration of total protein content in human serum is between 6000–8300 mg dL^{-1} and a reduced serum protein level is an indication of kidney disorders, HIV and aging in the elderly.^{20,21} A pure serum spectrum was used as the reference for the EMSC correction (Fig. S4†), along with spectra of pure β -carotene and water, measured under identical conditions. Fig. 4A presents the EMSC corrected Raman spectra of all the 25 patients from unfiltered serum samples. As expected, the characteristic bands of both globulin and albumin, highlighted in Fig. 3, can be observed, as indicated by the asterisks.^{81,82} Notably, the strong bands of β -carotene are no longer as prominent. The background-subtracted, smoothed spectra of the whole patient serum from the range 400 cm^{-1} to 1800 cm^{-1} were provided to the PLSR algorithm, regressed against total serum protein concentrations of Table 1, and a

prediction model for total protein concentrations was built. The RMSECV plot of the PLSR model shows a steady decrease within the first 10 components and stabilises after 12th component, indicating the data is well modelled (Fig. 4B). Note that the data range is not well balanced, and particularly that there is only one patient with total serum content below 5500 mg dL^{-1} . Nevertheless, the PLSR co-efficient of Fig. 4C shows spectral features at 940 cm^{-1} , 1553 cm^{-1} and 1176 cm^{-1} , indicating the prediction model (Fig. 4D) was built on the spectral features of albumin (50% of total protein) and γ globulin (38% of total protein). The RMSECV, R^2 and standard deviation values were calculated as 114.7 mg dL^{-1} , 0.82 and 5.69 mg dL^{-1} . For comparison, the prediction accuracy achieved by Berger *et al.*, for measuring total protein using near infrared Raman spectroscopy with 830 nm was reported to be 190 mg dL^{-1} ,⁵⁷ whereas Shaw *et al.*,⁵⁵ and Rohleder *et al.*,⁸³ reported prediction accuracies of 310 mg dL^{-1} , 176 mg dL^{-1} using Mid-IR spectroscopy of dried films indicating that the present method is more feasible for total protein analysis from serum.

In the case of determination of γ globulin content in patient serum, the background-subtracted and smoothed fingerprint region of the unfiltered serum spectra was also analysed, but regressed this time against the γ globulin concentration. As expected, the RMSECV prediction plot (Fig. 5B) shows a steady decrease until the 12th component, indicating that 12 components should be utilised in this model. The lowest value of RMSECV was recorded as 126 mg dL^{-1} and the linearity of the model was calculated as $R^2 = 0.88$. However, it should be noted that the spectral profile of the regression coefficient of Fig. 5C is not strikingly similar to the spectrum of γ globulin in Fig. 3, and the characteristic features of at 1240 cm^{-1} and 1553 cm^{-1} are not prominent. However, it should be noted that the spectral profile of the regression coefficient of Fig. 5C is not strikingly similar to the spectrum of γ globulin in Fig. 3, and the characteristic features of at 1240 cm^{-1} and 1553 cm^{-1} are not prominent. γ globulin is composed of IgG, IgM and IgA, the relative contributions of which are seen to vary considerably from patient to patient. Immunoglobulin G (IgG), itself, for example, is well known for its heterogeneity, which in itself makes it an excellent biomarker of a person's general state of health.^{84,85} Regression against each of the individual components did not produce a good correlation. The spectrum of each of them, and the composite g-globulin can also be influenced by protein/protein and any other interactions in the complex serum mixture, which can give rise to conformational changes in the immunoglobulin structure. Thus, it may not be surprising that the individual patient globulin spectral contributions are variable from patient to patient, and that the regression co-efficient of Fig. 5C, effectively an average of them all, does not exactly match the spectral of the commercial sample in Fig. 3. Nevertheless, the R^2 of this model is comparable to that of the method reported by Perez-Guaita *et al.*⁸⁶ to detect total globulin using ATR-FTIR from blood serum, and it should be noted that no correlation between globulin content and that of the

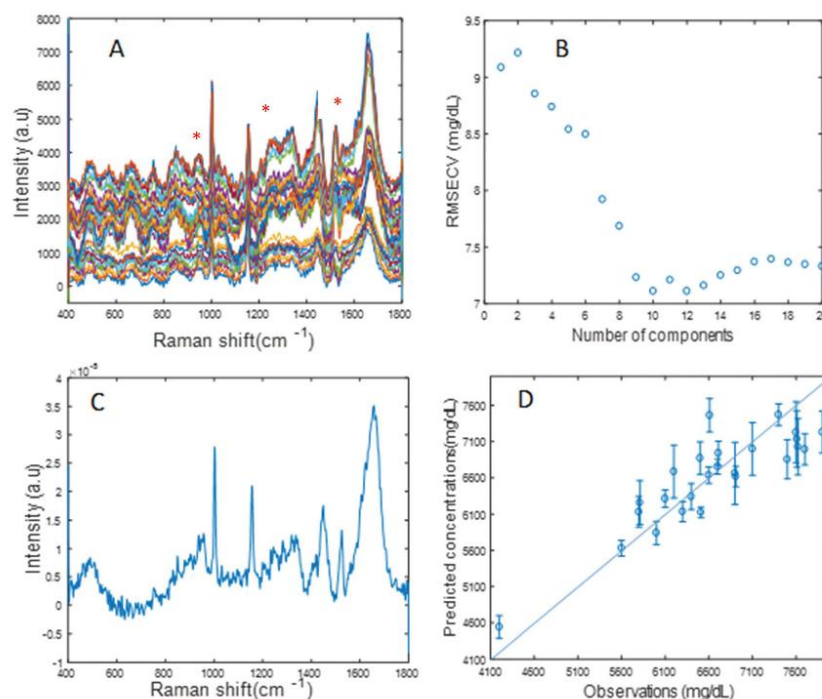


Fig. 4 (A) EMSC corrected Raman spectra of total protein content from patient serum samples (4200 mg dL⁻¹, 5800 mg dL⁻¹, 6400 mg dL⁻¹ and 7900 mg dL⁻¹). The spectra have been offset for clarity (B): evolution of RMSECV of the data set (C): plot of PLSR coefficient for regression against total serum protein, (D): linear predictive model for total serum protein built from the PLSR analysis. The RMSECV, R^2 and standard deviation values were calculated as 114.7 mg dL⁻¹, 0.82 and 5.69 mg dL⁻¹.

other major protein, albumin was observed (Fig. S1†). The background-subtracted and smoothed fingerprint region of the unfiltered serum spectra was also analysed over a reduced spectral region (1150 cm⁻¹ to 1600 cm⁻¹) to avoid spectral interference from albumin, but regressed this time against the γ globulin concentration. Analysis over this reduced range failed to produce an improved performance, however (Fig. S5†). Excess globulin levels (>1600 mg dL⁻¹) are usually an indicator of liver diseases, chronic inflammatory diseases, haematological disorders, infections and malignancies,²⁶ whereas humoral immunodeficiencies cause low globulin levels (<700 mg dL⁻¹).²⁷

As shown in Fig. 3, the Raman spectra of albumin and γ globulin are extremely similar and, since the concentrations of albumin and γ globulin are not correlated (Fig. S1 in ESI†), PLSR cannot predict both from the same dataset; invalid correlation for the entire population from which the training set is drawn would lead to spurious prediction models.⁵⁷ Hence, patient serum was spun using 100 kDa centrifugal filter tubes to separate and remove the γ globulins in the concentrate and the remainder filtrate was transferred to 50 kDa filter tubes for

further concentration of albumin. The spectra recorded from the concentrate of 50 kDa filtrate was used for building a PLSR prediction model for albumin from patient samples.

3.2 Quantification of albumin from the HMWF concentrate

For albumin determination, the spectra recorded from albumin paste (Fig. 3 blue) and β -carotene (Fig. S3†) were used as the reference spectrum for the EMSC algorithm to perform background correction and the background-corrected and smoothed spectra were fed into the PLSR algorithm to build the prediction model by regressing against albumin concentration (Table 1). The signature albumin bands that can be seen are the amide I band around ~1659 cm⁻¹, a relatively sharp band at 1003 cm⁻¹ associated with phenylalanine, intense bands at ~1336 cm⁻¹ and ~1450 cm⁻¹ due to C-H deformation, bands at 899 cm⁻¹ and 1102 cm⁻¹, which can be related to ν (CC) and ν (CN) and a vibration band at ~940 cm⁻¹, related to C-C stretching mode backbone of α -helix structure.⁸⁷ The absence of the Raman peaks at 1240 cm⁻¹ and 1553 cm⁻¹ confirms that there is no interference from γ globulin. PLSR analysis of aqueous solutions of albumin over the range

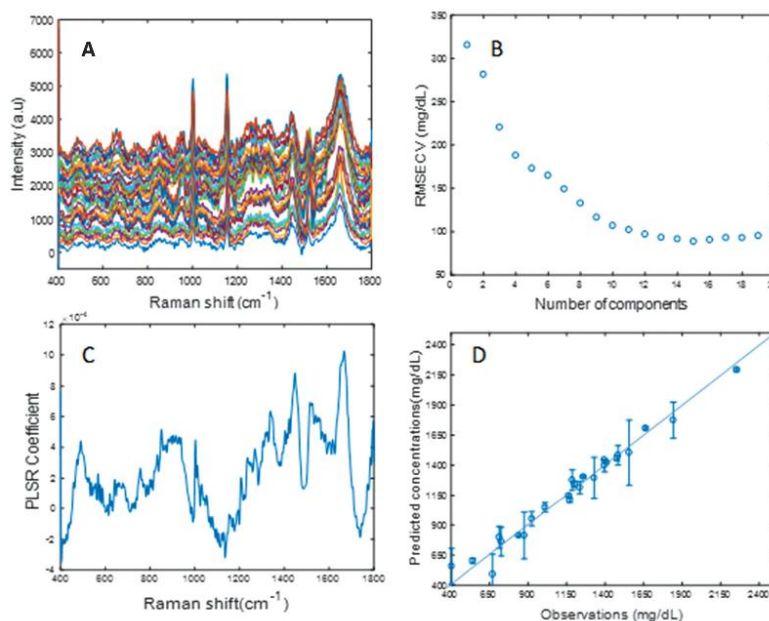


Fig. 5 (A) EMSC corrected Raman spectra of γ globulin of patient serum samples (329 mg dL⁻¹, 690 mg dL⁻¹, 836 mg dL⁻¹ and 1404 mg dL⁻¹), the spectra has been offset for clarity. (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient for regression against γ globulin concentration shows the signature peaks of γ globulin (highlighted by asterisk), (D): linear predictive model for γ globulin concentration built from the PLSR analysis. The RMSECV, R^2 and standard deviation values were calculated as 126 mg dL⁻¹, 0.88 and 4.62 mg dL⁻¹.

(500–5000 mg dL⁻¹) has previously been reported, and the results of the prediction model are summarised by Fig. S6 in the ESI†. Leave-one-out cross validation was used to test the robustness of this model and the RMSECV value calculated was 158 mg dL⁻¹.¹¹

EMSC successfully subtracted the background without altering the albumin features (Fig. 6A), such as bands at ~ 1336 cm⁻¹ due to C–H deformation and a vibration band at ~ 940 cm⁻¹, related to C–C stretching mode backbone of α -helix structure. The strength of the albumin features is, however, seen to vary considerably from patient to patient, as expected, due to the variation in albumin content from 2710 mg dL⁻¹ to 4780 mg dL⁻¹ (Table 1). The normal level of albumin in human blood is 3000 mg dL⁻¹,¹⁷ any decrease in this level could be an indicator of liver diseases.^{30–32} Applying PLSR to the dataset, a strong decrease in the RMSECV is observed within the first 8 latent variables, as shown in Fig. 6B, which is followed by a stabilisation of the values after 12 latent variables, indicating that 12 latent variables should be used to build the prediction model. The PLSR coefficient plot displayed in Fig. 6C shows visible albumin features, indicating that the prediction model was built on their variation. Finally, a linear predictive model is built from the PLSR analysis (Fig. 6D) to compare the observations to the known con-

centrations of albumin in the samples with the estimated concentrations from the spectral data sets, yielding an R^2 value of 0.9072 and RMSECV value of 90.097 mg dL⁻¹. The prediction accuracy of the present method for measuring albumin from serum is higher than the prior studies on Raman (120 mg dL⁻¹)⁵⁷ and Mid IR spectroscopy (220 mg dL⁻¹).⁵⁵ The overall standard deviation is calculated to be 1.1692 mg dL⁻¹. These values are comparable to the previously reported values obtained from varying concentration of pure albumin in water.¹¹

3.3 Quantification of glucose and urea from the LMWF filtrate

Glucose and urea are both present in the filtrate obtained after depletion of the HMWF of human serum using filtration. Notably, there is no obvious correlation between the levels of the two analytes per patient, as shown in Fig. S2 in ESI†.

Fig. 7 shows the reference spectra of urea and glucose used for EMSC correction of the raw spectral data. The signature peak of urea is a relatively sharp band seen at 1006 cm⁻¹, as seen in Fig. 7A, which can be attributed to the symmetric stretching of C–N.⁸⁸ Fig. 7B shows Raman peaks of glucose at 1060 cm⁻¹, due to ν (C1–OH) stretching, and a sharp peak at

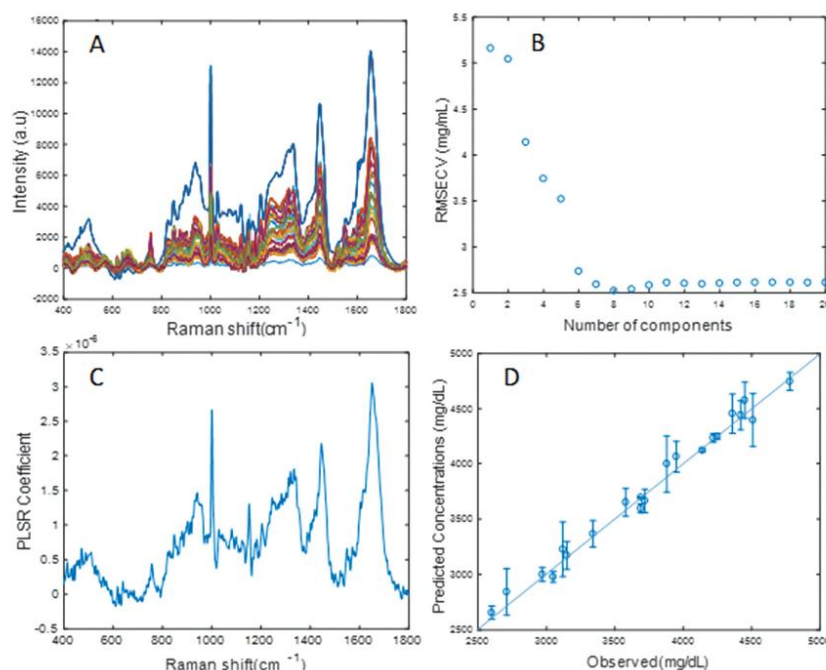


Fig. 6 (A): EMSC corrected Raman spectra of concentrate obtained after filtration with 50 kDa filters of patient sample (2710 mg dL⁻¹, 3580 mg dL⁻¹, 4140 mg dL⁻¹ and 4780 mg dL⁻¹), the spectra have been offset for clarity (B): evolution of RMSECV of the data set (C): plot of PLSR coefficient from regressing against albumin concentrations, (D): linear predictive model for albumin built from the PLSR analysis. The RMSECV, R^2 and standard deviation values were calculated as 90.097 mg dL⁻¹, 0.9072 and 1.1692 mg dL⁻¹.

1125 cm⁻¹, which can be assigned to the $\delta(\text{C-O-C})$ angle-bending mode.⁸⁹

In the case of glucose, Fig. S7, S8 and S9 in the ESI† display the previously reported prediction models built from varying concentrations of glucose in distilled water, spiked in commercial serum, and patient serum, respectively.⁶⁴ The RMSECV value calculated for the model built from varying concentrations in distilled water (100–1000 mg dL⁻¹) was 10.93 mg dL⁻¹ and R^2 was calculated as 0.9705. Notably, in the patient samples, the overlap of the strong spectral features of urea and glucose can interfere with the PLSR prediction of glucose using the full spectral range, as shown in Fig. S9†. In order to facilitate efficient prediction of glucose from the filtrate, the spectral region from 1030 cm⁻¹ to 1400 cm⁻¹ was chosen to build a PLSR prediction model of glucose from patient samples (Fig. S8†), (52.25–440 mg dL⁻¹) resulting in an RMSECV value of 1.84 mg dL⁻¹ and R^2 value of 0.84, comparable with the values obtained for the pure aqueous solutions.⁶⁴ The results suggest that this method could detect the concentration of glucose in the ranges of hyper (>130 mg dL⁻¹) or hypoglycaemia (<70 mg dL⁻¹).⁵¹

Prior to the analysis of urea in patient samples, aqueous solutions of varying concentrations of urea were made up and Raman analysis with the optimised protocol was performed. The recorded spectra were pre-processed and PLSR analysis was performed. EMSC efficiently subtracts the water spectrum, which has an OH bending vibration at ~1640 cm⁻¹ (ref. 64) and obscures the protein signals at low concentrations. Fig. S11† shows the spectra corrected with the EMSC algorithm, showing strong features of urea increasing as the concentration is increased from 1 mg dL⁻¹ to 1000 mg dL⁻¹. The PLSR algorithm, using 6 latent variables, was applied to the smoothed and background-subtracted spectra and a linear predictive model was built, yielding a lowest value of RMSECV of 70.4 mg dL⁻¹ and an R^2 value of 0.9048. The overall standard deviation was calculated to be 1.0975 mg dL⁻¹.

Having verified the protocol for analysis of urea in water, it was applied to the filtrate obtained after filtration from patient serum. Higher (<20 mg dL⁻¹) or lower levels (>5 mg dL⁻¹) of urea in the patient serum could indicate various pathologies such as kidney or liver malfunction.^{46–49} Fig. 8A shows the EMSC-corrected data of the fingerprint region and Fig. 8B indi-

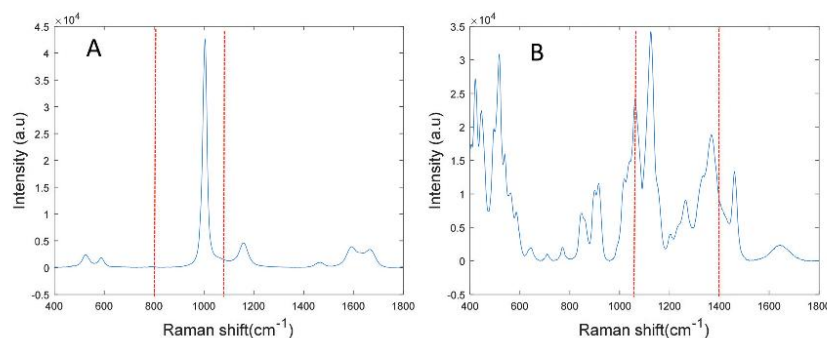


Fig. 7 (A) Reference spectrum of urea (1000 mg dL⁻¹) and (B) reference spectrum of glucose (45 000 mg dL⁻¹). The reduced spectral regions selected for PLSR analysis are indicated by dotted lines in each spectrum.

icates 12 latent variables should be used to build the model. The PLSR coefficient shows the features of urea, as displayed in Fig. 8C and D shows the linear model built from this dataset. The linearity is indicated by $R^2 = 0.9232$ and the RMSECV was calculated as 1.736 mg dL⁻¹. The overall standard deviation was calculated to be 2.89 mg dL⁻¹.

Notably, the typical serum concentrations of urea are considerably higher than those of glucose, as shown in Table 1, and therefore the characteristic features of urea dominate over those of glucose. Nevertheless, having observed the improvement of the sensitivity of the model for the case of glucose, a reduced spectral region, from 800 cm⁻¹ to 1030 cm⁻¹, was also tested for the urea regression. Fig. 9A shows the EMSC corrected spectra of urea over the reduced spectral range, and Fig. 9C displays the PLSR coefficient, which is dominated by the 1006 cm⁻¹ peak of urea. 12 latent variables were used to build the model after calculating the lowest value of RMSECV, as seen in Fig. 9B. Fig. 9D confirms that the linear predictive model is based on systematic variations of the features of urea. The value of RMSECV, R^2 and overall standard deviation calculated are 1.1418 mg dL⁻¹, 0.9722 and 2.52 mg dL⁻¹. Better R^2 and RMSECV values were obtained when the PLSR analysis was performed over the reduced spectral range of urea as compared to the full fingerprint region, suggesting that the sensitivity and the accuracy of the model could be considerably improved. The proposed method also has a better prediction accuracy than the previously reported methods on Raman by Berger *et al.* (3.8 mg dL⁻¹)⁵⁷ and mid IR spectroscopy by Shaw *et al.* (3.08 mg dL⁻¹)⁹⁰ Petrich *et al.* (16 mg dL⁻¹)⁵⁸ and Rohleder *et al.* (2.1 mg dL⁻¹).⁸³

4. Discussion

The study demonstrates that Raman spectroscopy, coupled with ultra-filtration and PLSR analysis, can be employed to detect variations in multiple analytes from the same serum samples with a high degree of higher accuracy (Table 2). The

routine serum analysis techniques used for the measurement of total protein (biuret method) and γ globulin (RID and Turbidometric Immuno assay (TIA)) are expensive, require technically competent laboratory personnel and cannot measure multiple analytes from the same serum samples.⁹¹ The strategy demonstrated in this study enables the simultaneous estimation of total protein level and detection of imbalance in γ globulin concentration accurately from whole serum, without the use of any reagents and without destroying the sample being studied. The proposed method has many advantageous over the biuret method, as the required sample volume can be as low as 10 μ L, it is rapid and non-destructive to the medium being studied, whereas the biuret method is considered impractical due to the requirement of large sample volume and laborious sample processing steps.^{1,92} Current methods used for determination of γ globulins are time and temperature dependent and are time consuming, which is a major impediment for early diagnosis.^{93–96} Moreover, the linearity reported for a comparative study of RID and TIA is 0.59,¹ lower than the R^2 calculated in the present study. The gold standard methods used for albumin measurements, such as BCG and Bromocresol purple assays have been reported to be susceptible to overestimation of albumin content, especially at lower albumin concentrations.^{97,98} In this study, it has been demonstrated that PLSR can be used to extract concentration predictions in order to build a prediction model from blood serum. With the use of filtration, the albumin was isolated from patient serum and the prediction model had a prediction accuracy significantly superior to that of the BCG method (2.2 g dL⁻¹) used to determine albumin concentrations from cirrhotic patients.⁹⁹

Serum fractionation considerably enhanced the capability to detect and quantify LMWF analytes⁹ and it has been shown that reducing the spectral region helped to avoid the spectral interference from other analytes and yielded improved accuracy.⁶⁴ It has previously been reported that selecting the spectral region from 1030 cm⁻¹ to 1400 cm⁻¹ improved the sensitivity and specificity for the prediction model of glucose from

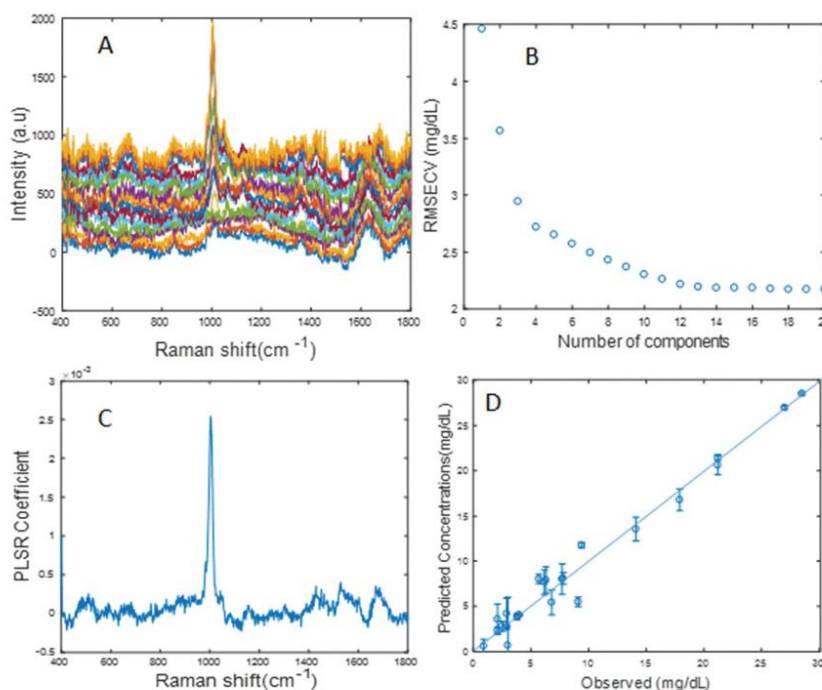


Fig. 8 (A): EMSC corrected Raman spectra of filtrate obtained after filtration with 10 kDa filters of patient samples (2.5 mg dL⁻¹, 10.64 mg dL⁻¹, 19.04 mg dL⁻¹ and 78.99 mg dL⁻¹). The spectra have been offset for clarity (B): evolution of RMSECV of the data set regressed against urea concentrations (C): plot of PLSR coefficient with strong features of urea, (D): linear predictive model built from the PLSR analysis. The RMSECV, R^2 , and standard deviation values were calculated as 1.736 mg dL⁻¹, 0.9232 and 2.89 mg dL⁻¹.

patient samples over the concentration range 52.5–434.2 mg dL⁻¹, and the technique was demonstrated to be at least as accurate as ATR-FTIR of similar patient samples,⁹ measured in the dried state and closer to the accuracy of colorimetric methods, 1.4 mg dL⁻¹ urea¹⁰⁰ and 2 mg dL⁻¹ glucose.¹⁰¹ As anticipated, similarly higher prediction accuracy (RMSECV = 1.14 mg dL⁻¹) was attained when PLSR analysis was performed on a reduced range for urea from patient samples over the concentration range 2.52–78.99 mg dL⁻¹, compared to the full range (RMSECV = 1.73 mg dL⁻¹).

The strategy illustrated in this and previous studies enables simultaneous detection of various analytes from human serum using Raman spectroscopy with minimal sample preparation, no labelling and no additional sample drying steps. Depletion of the HMWF is a non-optional step in the investigation of protein imbalances or disease related biomarkers in the LMWF of serum. Mass spectroscopy, the most commonly used technique for LMWF analysis also makes use of the HMWF depletion in order to potentially target small circulating biomarkers.¹⁰² While most of these methods found in the literature are based on chemical extraction/precipitation, the

impacts of these solvents are unknown and cross contamination from these solvents is highly probable.^{102–104} Therefore, ultra-filtration coupled with chemometrics is a viable option to investigate the less studied LMWF and to interpret the results by overcoming the intricacies of the multidimensional dataset.

Several methods have been previously described in literature for the quantification of multiple analytes in serum samples. Perez-Guaita *et al.*, established models for the determination, in serum samples, of albumin, γ globulin, total globulin, and albumin/globulin coefficients, using ATR-FTIR. Values of RMSECV determined for albumin (126 mg dL⁻¹) and γ globulin (138 mg dL⁻¹) are comparable, or larger than the RMSECV determined in the current study (albumin 90 mg dL⁻¹, γ globulin 126 mg dL⁻¹). Although the spectral co-efficient of prediction of Fig. 5C is somewhat unsatisfactory, in that the specific distinctive features of γ globulin are not prominent, the γ globulin levels are not correlated with albumin levels, and the predictive model shows a relatively high degree of linearity and low standard deviation, giving some assurance that it is indeed predicting γ globulin levels. In the study reported by Roy *et al.*, ATR-FTIR spectroscopy enabled the sim-

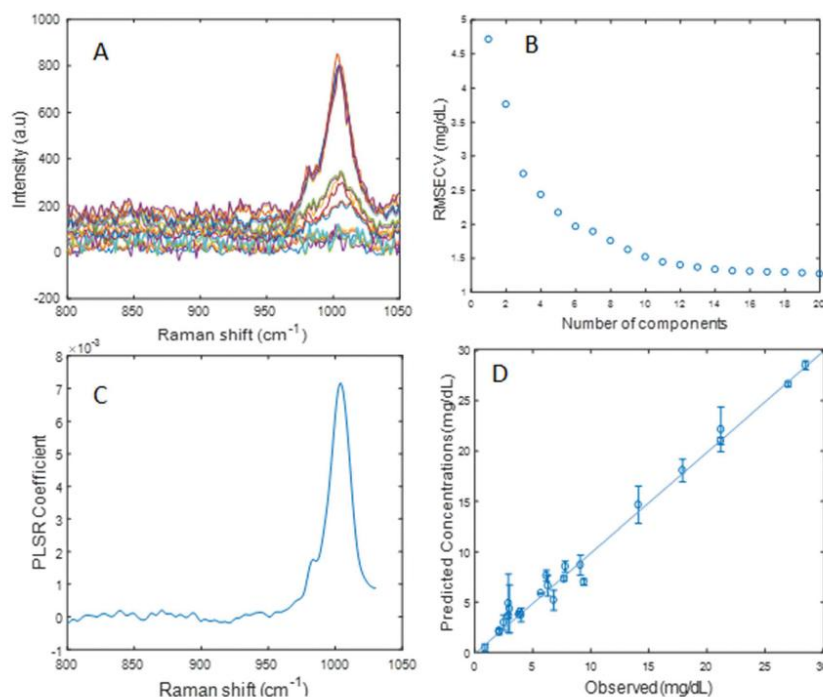


Fig. 9 (A): EMSC corrected Raman spectra of filtrate obtained after filtration with 10 kDa filters of patient samples (2.5 mg dL⁻¹, 10.64 mg dL⁻¹, 19.04 mg dL⁻¹ and 78.99 mg dL⁻¹). Spectra have been offset for clarity. (B): evolution of RMSECV of the data set regressed against urea concentration (C): plot of PLSR coefficient with strong features of urea, (D): linear predictive model for urea concentration built from the PLSR analysis. The RMSECV, R^2 and standard deviation values were calculated as 2.52 mg dL⁻¹, 0.9722 and 1.1418 mg dL⁻¹.

Table 2 Summary of the results obtained from the patient samples

Analyte	Analysis range	Concentration range	R^2	RMSECV	Standard deviation
Total protein (whole serum)	400–1800 cm ⁻¹	4200–7900 mg dL ⁻¹	0.82	115 mg dL ⁻¹	5.7 mg dL ⁻¹
γ globulin (whole serum)	1150–1600 cm ⁻¹	329–1404 mg dL ⁻¹	0.88	126 mg dL ⁻¹	4.6 mg dL ⁻¹
Albumin (HMWF)	400–1800 cm ⁻¹	2600–4780 mg dL ⁻¹	0.90	90 mg dL ⁻¹	1.2 mg dL ⁻¹
Glucose (LMWF)	1030–1400 cm ⁻¹	52.5–434.2 mg dL ⁻¹	0.84	1.8 mg dL ⁻¹	3.5 mg dL ⁻¹
Urea (LMWF)	400–1800 cm ⁻¹	2.52–78.99 mg dL ⁻¹	0.92	1.7 mg dL ⁻¹	2.9 mg dL ⁻¹
Urea (LMWF)	800–1030 cm ⁻¹	2.52–78.99 mg dL ⁻¹	0.97	1.1 mg dL ⁻¹	2.5 mg dL ⁻¹

ultaneous quantification of glucose and urea analytes along with malaria parasitemia quantification from a spectrum obtained from a dried drop of blood on a glass slide. The specificity for the PLS-DA was found to be 98% for parasitemia levels but low sensitivity of 70% was achieved because of the negative samples in the model. The RMSECV for parasite concentration (0–5%), glucose (0–400 mg dL⁻¹) and urea (0–250 mg dL⁻¹) spiked samples were 0.58%, 16% and 17%, respectively,⁵⁶ whereas the present study showed a considerably lower RMSECV value of 1.84 mg dL⁻¹ (52–440 mg dL⁻¹) and 1.69 mg dL⁻¹ (2–79 mg dL⁻¹) for glucose and urea from

patient samples. Another study conducted by Shaw *et al.*, based upon the infrared spectra for 8 serum analytes, reported standard errors of 2.8 g L⁻¹ (total protein), 22 mg dL⁻¹ (albumin), 0.23 mmol L⁻¹ (triglycerides), 0.28 mmol L⁻¹ (cholesterol), 7.4 mg dL⁻¹ (glucose) and 6.6 mg dL⁻¹ for urea, with correlation coefficients of 0.95.⁵⁵ In an investigation conducted on 247 serum donors by Rohleder *et al.* using Raman spectroscopy of $\lambda_{\text{ex}} = 785$ nm, the standard errors for 7 analytes were reported as 176 mg dL⁻¹ (total protein), 20.7 mg dL⁻¹ (triglycerides), 11.0 mg dL⁻¹ (high density lipoprotein), 15.7 mg dL⁻¹ (low density lipoprotein), 0.81 mg dL⁻¹ (uric acid),

2.1 mg dL⁻¹ (urea) and 6.8 mg dL⁻¹ (glucose) and the efficiency of ultrafiltration technique in improving the prediction accuracy of glucose and urea was demonstrated.⁸³ Although the samples were dried in all these, superior prediction accuracy was afforded by Raman analysis of liquid samples. Berger *et al.* demonstrated the use of Raman spectroscopy to measure concentrations of serum and whole blood components to simultaneously predict the content of 6 analytes in serum from a 66 patient data set. The quoted prediction errors for albumin, urea and glucose were 120 mg dL⁻¹, 3.8 mg dL⁻¹ and 26 mg dL⁻¹, respectively,⁵⁷ and therefore the combination of improved measurement protocols, serum fractionation, and sectioning the spectral region for regression results in considerable improvement in the prediction accuracy.

It should be noted that direct comparisons of the two techniques of ATR-FTIR and Raman, and even different studies using the same technique, can only be tenuous at best, given the number of potential variables in samples, measurement protocols and data analysis techniques. Notably, a systematic study of the PLSR model construction, validation and testing protocols has not as yet been carried out. The previous study of Parachalil *et al.*,¹¹ tried to minimise these variabilities by utilising the same sample preparation/processing as well as data analysis techniques in the direct comparison of ATR-FTIR of dried serum with Raman of the liquid state, for the quantification of glucose in serum. The present study has further demonstrated the capacity for prediction of serum content for total protein, γ globulin, albumin, urea and glucose from the same patient samples, with a high degree of accuracy, consolidating the prospect of establishing Raman spectroscopy as a biomedical tool to rival and/or augment conventional approaches such as mass spectroscopy or chromatography, currently used to deliver crucial information relevant for diagnosis. Further improvements in the sensitivities and variabilities of the techniques ultimately rely on the reproducibility of the measurement, and the signal to noise ratio. The former could potentially be improved by an automated focussing and sampling methodology, while the latter is largely an instrumental consideration. As longer accumulation times to reduce the noise are not recommended, because of sample evaporation and also speed of measurement throughput, optimisation of instrumentation for higher signal throughput could be explored, for example by sacrificing spectral resolution.

5. Conclusion

In summary, the potential of Raman spectroscopy combined with filtration and chemometrics to detect variations in total serum protein, γ globulin, HMWF and LMWF analytes from the same patient serum is successfully demonstrated. Firstly, the prediction model was built in the spiked samples of the analytes in water and was then translated to the patient serum. Although Raman spectroscopy can build quantitative models with higher accuracy in spiked samples, the analysis of patient samples can be affected by numerous factors such as multi-

parametrical variability, normally observed in clinical application. Nevertheless, the proposed approach successfully built the prediction model from the whole serum (total protein, γ globulin), concentrate (albumin) by removing β -carotene and filtrate (urea and glucose) of patient samples at higher accuracy and sensitivity. Given its low cost, easy implementation, rapid results and higher sensitivity and specificity, the technique may become an alternative for the existing laboratory screening methods. Furthermore, the methodology presented in this work can be applied to a wider range of bodily fluids and can be implemented as a next generation point of care biochemical diagnostic tool in a clinical setting.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Drishya Rajan Parachalil was funded by DIT Fiosraigh scholarship. J. McIntyre was funded by Science Foundation Ireland, PI/11/08. The Irish-French PHC Ulysses 2018 Collaborative project is jointly funded by the Irish Research Council and Campus France with participation of the French Embassy in Ireland.

References

- 1 B. Okutucu, Ö. Habib and Z. Figen, Comparison of five methods for determination of total plasma protein concentration, *J. Biochem. Biophys. Methods*, 2007, **70**, 709–711.
- 2 C. Parikh, R. Yalavarthy, A. Gurevich, A. Robinson and I. Teitelbaum, Discrepancies in serum albumin measurements vary by dialysis modality, *Renal Failure*, 2003, **25**(5), 787–796.
- 3 P. Crow, B. Barrass, C. Kendall, M. Hart-Prieto, M. Wright, R. Persad, *et al.*, The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines, *Br. J. Cancer*, 2005, **92**, 2166–2170.
- 4 P. J. Caspers, G. W. Lucassen, E. A. Carter, H. A. Bruining and G. J. Puppels, In vivo confocal raman microspectroscopy of the skin: Noninvasive determination of molecular concentration profiles, *J. Invest. Dermatol.*, 2001, **116**(3), 434–442.
- 5 C. G. Atkins, K. Buckley, M. W. Blades and R. F. B. Turner, Raman Spectroscopy of Blood and Blood Components, *Appl. Spectrosc.*, 2017, **71**(5), 767–793.
- 6 J. Sulé-Suso, N. R. Forsyth, V. Untereiner and G. D. Sockalingum, Vibrational spectroscopy in stem cell characterisation: Is there a niche?, *Trends Biotechnol.*, 2014, **32**(5), 254–262.
- 7 R. Gautam, S. Vanga, F. Ariese and S. Umapathy, Review of multidimensional data processing approaches for

- Raman and infrared spectroscopy, *EPJ Tech. Instrum.*, 2015, **2**(1), 8.
- 8 F. Bonnier, G. Brachet, R. Duong, T. Sojinrin, R. Respaud, N. Aubrey, *et al.*, Screening the low molecular weight fraction of human serum using ATR-IR spectroscopy, *J. Biophotonics*, 2016, **9**(10), 1085–1097.
 - 9 F. Bonnier, H. Blasco, C. Wasselet, G. Brachet, R. Respaud, L. F. C. S. Carvalho, *et al.*, Ultra-filtration of human serum for improved quantitative analysis of low molecular weight biomarkers using ATR-IR spectroscopy, *Analyst*, 2017, **142**(8), 1285–1298.
 - 10 F. Bonnier, F. Petitjean, M. J. Baker and H. J. Byrne, Improved protocols for vibrational spectroscopic analysis of body fluids, *J. Biophotonics*, 2014, **7**(3–4), 167–179.
 - 11 D. R. Parachalil, B. Brankin, J. McIntyre and H. J. Byrne, Raman spectroscopic analysis of high molecular weight proteins in solution – considerations for sample analysis and data pre-processing, *Analyst*, 2018, **143**(24), 5987–5998.
 - 12 C. A. Jenkins, R. A. Jenkins, M. M. Pryse, K. A. Welsby, M. Jitsumura, C. A. Thornton, *et al.*, A high-throughput serum Raman spectroscopy platform and methodology for colorectal cancer diagnostics, *Analyst*, 2018, **143**(24), 6014–6024.
 - 13 F. Bonnier, M. J. Baker and H. J. Byrne, Vibrational spectroscopic analysis of body fluids: avoiding molecular contamination using centrifugal filtration, *Anal. Methods*, 2014, **6**(14), 5155.
 - 14 A. A. Bunaciu, Ş. Fleschin, V. D. Hoang and H. Y. Aboul-Enein, Vibrational Spectroscopy in Body Fluids Analysis, *Crit. Rev. Anal. Chem.*, 2017, **47**(1), 67–75.
 - 15 M. J. Baker, C. S. Hughes and K. A. Hollywood, *Biophotonics: Vibrational Spectroscopic Diagnostics [Internet]*, Morgan & Claypool Publishers, 2016.
 - 16 A. L. Mitchell, K. B. Gajjar, G. Theophilou, F. L. Martin and P. L. Martin-Hirsch, Vibrational spectroscopy of biofluids for disease screening or diagnosis: Translation from the laboratory to a clinical setting, *J. Biophotonics*, 2014, **7**(3–4), 153–165.
 - 17 J. T. Busher, Serum Albumin and Globulin, *Clin. Methods Hist. Phys. Lab. Exam.*, 1990, 497–499.
 - 18 M. Z. Rahman and B. A. Begum, Serum Total Protein and Albumin Levels in Different Grades of Protein Energy Malnutrition, *Mymensingh Med. J.*, 2005, **14**(1), 38–40.
 - 19 K. Hayden and C. van Heyningen, Measurement of Total Protein Is a Useful Inclusion in Liver Function Test Profiles, *Clin. Chem.*, 2001, **47**(4), 793–794.
 - 20 B. G. Gazzard, HIV disease and the gastroenterologist, *Gut*, 1988, **29**(11), 1497–1505.
 - 21 C. Tian, L. Qian, X. Shen, J. Li and J. Wen, Distribution of Serum Total Protein in Elderly Chinese, *PLoS One*, 2014, **9**(6), e101242.
 - 22 P. J. Roberts-Thomson and K. Shepherd, Molecular size heterogeneity of immunoglobulins in health and disease, *Clin. Exp. Immunol.*, 1990, **79**(3), 328–334.
 - 23 E. Merler and F. S. Rosen, The Gamma Globulins, *N. Engl. J. Med.*, 1966, **275**(10), 536–542.
 - 24 T. B. Tomasi and W. A. Tisdale, Serum Gamma-globulins in Acute and Chronic Liver Diseases, *Nature*, 1964, **201**(4921), 834–835.
 - 25 W. Gross and R. S. Snell, The Serum Gamma-Globulin-Level in Malignant Disease, *Nature*, 1956, **178**(4538), 855, DOI: 10.1038/178855a0. Available from.
 - 26 A. Dispenzieri, M. A. Gertz, T. M. Therneau and R. A. Kyle, Retrospective cohort study of 148 patients with polyclonal gammopathy, *Mayo Clin. Proc.*, 2001, **76**(5), 476–487.
 - 27 R. H. Buckley, Humoral immunodeficiency, *Clin. Immunol. Immunopathol.*, 1986, **40**(1), 13–24.
 - 28 J. T. Whicher, C. Warren and R. E. Chambers, Immunochemical assays for immunoglobulins, *Ann. Clin. Biochem.*, 1984, **21**(Pt 2), 78–91.
 - 29 J. P. Nicholson, M. R. Wolmarans and G. R. Park, The role of albumin in critical illness, *Br. J. Anaesth.*, 2000, **85**(4), 599–610.
 - 30 V. Arroyo, R. Garcia-Martinez and X. Salvatella, Human serum albumin, systemic inflammation, and cirrhosis, *J. Hepatol.*, 2014, **61**(2), 396–407.
 - 31 T. B. Vree, M. Shimoda, J. J. Driessen, P. J. Guelen, T. J. Janssen, E. F. Termond, *et al.*, Decreased plasma albumin concentration results in increased volume of distribution and decreased elimination of midazolam in intensive care patients, *Clin. Pharmacol. Ther.*, 1989, **46**(5), 537–544.
 - 32 V. Arroyo, Review article: albumin in the treatment of liver diseases—new features of a classical treatment, *Aliment. Pharmacol. Ther.*, 2002, **16**(Suppl 5), 1–5.
 - 33 G. Fanali, A. Di Masi, V. Trezza, M. Marino, M. Fasano and P. Ascenzi, Human serum albumin: From bench to bedside, *Mol. Aspects Med.*, 2012, **33**(3), 209–290.
 - 34 A. T. Høstmark, Serum albumin and prevalence of coronary heart disease : A population-based, cross sectional study, *Norsk Epidemiologi*, 2003, **13**(1), 107–113.
 - 35 O. Karahan, H. Acet, F. Ertaş, O. Tezcan, A. Çalişkan, M. Demir, *et al.*, The relationship between fibrinogen to albumin ratio and severity of coronary artery disease in patients with ST-elevation myocardial infarction, *Am. J. Emerg. Med.*, 2016, **34**(6), 1037–1042.
 - 36 H. C. Beck, M. Overgaard and L. Melholt Rasmussen, Plasma proteomics to identify biomarkers - Application to cardiovascular diseases, *Transl. Proteomics*, 2015, **7**, 40–48.
 - 37 R. F. Gillum, Assessment of Serum Albumin Concentration As a Risk Factor for Stroke and Coronary Disease in African Americans and Whites, *J. Natl. Med. Assoc.*, 2000, **92**, 3–9.
 - 38 J. J. Nelson, D. Liao, A. R. Sharrett, A. R. Folsom, L. E. Chambless, E. Shahar, *et al.*, Serum albumin level as a predictor of incident coronary heart disease: the Atherosclerosis Risk in Communities (ARIC) study, *Am. J. Epidemiol.*, 2000, **151**(5), 468–477.
 - 39 Z. Liu, S. Fan, H. Liu, J. Yu, R. Qiao, M. Zhou, *et al.*, Enhanced detection of low-abundance human plasma

- proteins by integrating polyethylene glycol fractionation and immunoaffinity depletion, *PLoS One*, 2016, **11**(11), 1–17.
- 40 J. S. Lee, Albumin for end-stage liver disease, *Korean J. Intern. Med.*, 2012, **27**(1), 13–19.
 - 41 P. K. Drain, J. M. Baeten, J. Overbaugh, M. H. Wener, D. D. Bankson, L. Lavreys, *et al.*, Low serum albumin and the acute phase response predict low serum selenium in HIV-1 infected women, *BMC Infect. Dis.*, 2006, **6**(85), 1–6.
 - 42 S. Choi, E. Y. Choi, D. J. Kim, J. H. Kim, T. S. Kim and S. W. Oh, A rapid, simple measurement of human albumin in whole blood using a fluorescence immunoassay (I), *Clin. Chim. Acta*, 2004, **339**(1–2), 147–156.
 - 43 A. Artigas, J. Wernerman, V. Arroyo, J. L. Vincent and M. Levy, Role of albumin in diseases associated with severe systemic inflammation: Pathophysiologic and clinical evidence in sepsis and in decompensated cirrhosis, *J. Crit. Care*, 2016, **33**, 62–70.
 - 44 A. Carfray, K. Patel, P. Whitaker, P. Garrick, G. J. Griffiths and G. L. Warwick, Albumin as an outcome measure in haemodialysis in patients: the effect of variation in assay method, *Nephrol., Dial., Transplant.*, 2000, **15**(11), 1819–1822.
 - 45 F. Kurzer and P. M. Sanderson, Urea in the history of organic chemistry: Isolation from natural sources, *J. Chem. Educ.*, 1956, **33**(9), 452.
 - 46 D. Aronson, M. A. Mittleman and A. J. Burger, Elevated blood urea nitrogen level as a predictor of mortality in patients admitted for decompensated heart failure, *Am. J. Med.*, 2004, **116**(7), 466–473.
 - 47 J. L. Orsonneau, C. Massoubre, M. Cabanes and P. Lustenberger, Simple and sensitive determination of urea in serum and urine, *Clin. Chem.*, 1992, **38**(5), 619–623.
 - 48 A. Kazory, Emergence of blood urea nitrogen as a biomarker of neurohormonal activation in heart failure, *Am. J. Cardiol.*, 2010, **106**(5), 694–700.
 - 49 C. Higgins, *Urea and the clinical value of measuring blood urea concentration*, 2016, available from: <https://acute-care-testing.org/~media/acute-care-testing/files/pdf/urea-and-the-clinical-value-of-measuring-blood-ans-approved.pdf>.
 - 50 J. Lindenfeld and R. W. Schrier, Blood Urea Nitrogen, *J. Am. Coll. Cardiol.*, 2011, **58**(4), 383–385.
 - 51 J. M. Mcmillin, Blood Glucose, *Clin. Methods Hist. Phys. Lab. Exam.*, 1990, 662–665.
 - 52 B. H. Ginsberg, Factors affecting blood glucose monitoring: Sources of errors in measurement, *J. Diabetes Sci. Technol.*, 2009, **3**(4), 903–913.
 - 53 E. Boland, T. Monsod, M. Delucia, C. A. Brandt, S. Fernando and W. V. Tamborlane, Limitations of conventional methods of self-monitoring of blood glucose lessons learned from 3 days of continuous glucose sensing in pediatric patients with type 1 diabetes, *Diabetes Care*, 2001, **24**(11), 1858–1862.
 - 54 K. Spalding, F. Bonnier, C. Bruno, H. Blasco, R. Board, I. Benz-de Bretagne, *et al.*, Enabling quantification of protein concentration in human serum biopsies using attenuated total reflectance – Fourier transform infrared (ATR-FTIR) spectroscopy, *Vib. Spectrosc.*, 2018, **99**, 50–58.
 - 55 R. A. Shaw, S. Kotowich, M. Leroux and H. H. Mantsch, Multianalyte Serum Analysis Using Mid-Infrared Spectroscopy, *Ann. Clin. Biochem.*, 1998, **35**(5), 624–632.
 - 56 S. Roy, D. Perez-Guaita, D. W. Andrew, J. S. Richards, D. McNaughton, P. Heraud, *et al.*, Simultaneous ATR-FTIR Based Determination of Malaria Parasitemia, Glucose and Urea in Whole Blood Dried onto a Glass Slide, *Anal. Chem.*, 2017, **89**(10), 5238–5245.
 - 57 A. J. Berger, T. Koo, I. Itzkan, G. Horowitz and M. S. Feld, Multicomponent blood analysis by near-infrared Raman spectroscopy, *Appl. Opt.*, 1999, **38**(13), 2916–2926.
 - 58 D. Rohleder, G. Kocherscheidt, K. Gerber, W. Kiefer, W. Kohler, J. Mocks, *et al.*, Comparison of mid-infrared and Raman spectroscopy in the quantitative analysis of serum, *J. Biomed. Opt.*, 2005, **10**(3), 31108.
 - 59 H. A. Krebs, Chemical Composition of Blood Plasma and Serum, *Annu. Rev. Biochem.*, 1950, **19**(1), 409–430.
 - 60 J. Fine, The biuret method of estimating albumin and globulin in serum and urine, *Biochem. J.*, 1935, **29**(3), 799–803.
 - 61 M. M. Lubran, The Measurement of Total Serum Proteins by the Biuret Method, *Ann. Clin. Lab. Sci.*, 1978, **8**(2), 106–110.
 - 62 E. J. Sampson, M. A. Baird, C. A. Burtis, E. M. Smith, D. L. Wltte and D. D. Bayse, A Coupled-Enzyme Equilibrium Method for Measuring Urea in Serum : Optimization and Evaluation of the AACC Study Group on Urea Candidate Reference Method, *Clin. Chem.*, 1980, **26**(7), 816–826.
 - 63 A. Tvarijonaviciute, S. Mart, M. Caldin, F. Tecles and J. J. Ceron, Evaluation of automated assays for immunoglobulin G, M, and A measurements in dog and cat serum, *Vet. Clin. Pathol.*, 2013, **42**(3), 270–280.
 - 64 D. R. Parachalil, C. Bruno, F. Bonnier, H. Blasco, I. Chourpa, M. J. Baker, *et al.*, Analysis of bodily fluids using vibrational spectroscopy: a direct comparison of Raman, *Analyst*, 2018, **143**(24), 5987–5998.
 - 65 L. T. Kerr and B. M. Hennelly, A multivariate statistical investigation of background subtraction algorithms for Raman spectra of cytology samples recorded on glass slides, *Chemom. Intell. Lab. Syst.*, 2016, **158**, 61–68.
 - 66 A. Kohler, C. Kirschner, A. Oust and H. Martens, Extended multiplicative signal correction as a tool for separation and characterization of physical and chemical information in Fourier transform infrared microscopy images of cryo-sections of beef loin, *Appl. Spectrosc.*, 2005, **59**(6), 707–716.
 - 67 K. H. Liland, A. Kohler and N. K. Afseth, Model-based pre-processing in Raman spectroscopy of biological samples, *J. Raman Spectrosc.*, 2016, **47**(6), 643–650.
 - 68 S. Wold, M. Sjöström and L. Eriksson, PLS-regression: A basic tool of chemometrics, *Chemom. Intell. Lab. Syst.*, 2001, **58**(2), 109–130.

- 69 S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edn, 2009.
- 70 N. Tschirner, M. Schenderlein, K. Brose, E. Schlodder, A. Mroginiski and P. Hildebrandt, Resonance Raman spectra of β -carotene in solution and in photosystems revisited: an experimental and theoretical study, *Phys. Chem. Chem. Phys.*, 2009, **11**(48), 11471–11478.
- 71 B. Tan and D. N. Soderstrom, Qualitative aspects of UV-vis spectrophotometry of β -carotene and lycopene, *J. Chem. Educ.*, 1989, **66**(3), 258.
- 72 K. A. Steinmetz and J. D. Potter, Vegetables, fruit, and cancer. II. Mechanisms, *Cancer, Causes Control*, 1991, **2**(6), 427–442.
- 73 M. C. Martini, D. R. Campbell, M. D. Gross, G. A. Grandits, J. D. Potter and J. L. Slavin, Plasma carotenoids as biomarkers of vegetable intake: the University of Minnesota Cancer Prevention Research Unit Feeding Studies, *Cancer Epidemiol., Biomarkers Prev.*, 1995, **4**(5), 491–496.
- 74 L. Le Marchand, J. H. Hankin, F. S. Carter, C. Essling, D. Luffey, A. A. Franke, *et al.*, A pilot study on the use of plasma carotenoids and ascorbic acid as markers of compliance to a high fruit and vegetable dietary intervention, *Cancer Epidemiol., Biomarkers Prev.*, 1994, **3**(3), 245–251.
- 75 P. Hu, D. B. Reuben, E. M. Crimmins, T. B. Harris, M. Huang and T. E. Seeman, The Effects of Serum β -Carotene Concentration and Burden of Inflammation on All-Cause Mortality Risk in High-Functioning Older Persons: MacArthur Studies of Successful Aging, *J. Gerontol., Ser. A*, 2004, **59**(8), 849–854.
- 76 J. Huang, S. J. Weinstein, K. Yu, S. Männistö and D. Albanes, A prospective study of serum metabolites and glioma risk, *Oncotarget*, 2017, **8**(41), 70366–70377.
- 77 L. Rock and A. Henderson, Serum α -carotene and vitamin C as biomarkers vegetable and fruit intakes in a community-based of French adults13 of sample, *Am. J. Clin. Nutr.*, 1997, **65**(6), 1796–1802.
- 78 P. Hanson, S. Lu, J. Wang, W. Chen, L. Kenyon, C. Tan, *et al.*, Scientia Horticulturae Conventional and molecular marker-assisted selection and pyramiding of genes for multiple disease resistance in tomato, *Sci. Hortic.*, 2016, **201**, 346–354.
- 79 P. Thi, T. Phuong, S. Lee, C. Lee, B. Seo and S. Park, Colloids and Surfaces B : Biointerfaces β -carotene-bound albumin nanoparticles modified with chlorin e6 for breast tumor ablation based on photodynamic therapy, *Colloids Surf., B*, 2018, **171**(June), 123–133.
- 80 H.-T. Chang, H. Cheng, R.-M. Han, J.-P. Zhang and L. H. Skibsted, Binding to Bovine Serum Albumin Protects β -Carotene against Oxidative Degradation, *J. Agric. Food Chem.*, 2016, **64**(29), 5951–5957.
- 81 P. C. Painter and J. L. Koenig, Raman Spectroscopic Study of the Structure of Antibodies, *Biopolymers*, 1975, **14**(3), 457–468.
- 82 A. M. K. Enejder, T.-W. Koo, J. Oh, M. Hunter, S. Sasic, M. S. Feld, *et al.*, Blood analysis by Raman spectroscopy, *Opt. Lett.*, 2002, **27**(22), 2004–2006.
- 83 D. Rohleder, W. Kiefer and W. Petrich, Quantitative analysis of serum and serum ultrafiltrate by means of Raman spectroscopy, *Analyst*, 2004, **129**(10), 906–911.
- 84 D. Zhang, B. Chen, Y. Wang, P. Xia, C. He and Y. Liu, Disease-specific IgG Fc N-glycosylation as personalized biomarkers to differentiate gastric cancer from benign gastric diseases, *Sci. Rep.*, 2016, **6**, 25957.
- 85 I. Gudelj, G. Laue and M. Pezer, Immunoglobulin G glycosylation in aging and diseases, *Cell. Immunol.*, 2018, **333**(January), 65–79.
- 86 D. Perez-Guaita, J. Ventura-Gayete, C. Pérez-Rambla, M. Sancho-Andreu, S. Garrigues and M. De La Guardia, Protein determination in serum and whole blood by attenuated total reflectance infrared spectroscopy, *Anal. Bioanal. Chem.*, 2012, **404**(3), 649–656.
- 87 A. Lykina, D. Artemyev and I. Bratchenko, Analysis of albumin Raman scattering registration efficiency from different volume and shape cuvette, *J. Biomed. Photonics Eng.*, 2017, **3**(2), 020309.
- 88 R. L. Frost, J. Kristof, L. Rintoul and J. T. Klopogge, Raman spectroscopy of urea and urea-intercalated kaolinites at 77 K, *Spectrochim. Acta, Part A*, 2000, **56**(9), 1681–1691.
- 89 S. Söderholm, Y. H. Roos, N. Meinander and M. Hotokka, Raman spectra of fructose and glucose in the amorphous and crystalline states, *J. Raman Spectrosc.*, 1999, **30**(11), 1009–1018.
- 90 R. A. Shaw, S. Kotowich, M. Leroux and H. H. Mantsch, Multianalyte Serum Analysis Using Mid-Infrared Spectroscopy, *Ann. Clin. Biochem.*, 1998, **35**(5), 624–632.
- 91 T. R. Koch, G. F. Johnson and M. E. Chilcote, Kinetic Determination of Total Serum Protein with a Centrifugal Analyzer, *Clin. Chem.*, 1974, **20**(3), 392–394.
- 92 R. A. Shaw, S. Low-Ying, A. Man, K.-Z. Liu, C. Mansfield and C. B. Rileg, *et al.*, *Infrared Spectroscopy of Biofluids in Clinical Chemistry and Medical Diagnostics*, in *Biomedical Vibrational Spectroscopy*, John Wiley & Sons, Inc., 2007, pp. 79–103.
- 93 D. G. Davis, D. M. W. Schaefer, K. W. Hinchcliff, M. L. Wellman, V. E. Willet and J. M. Fletcher, Measurement of Serum IgG in Foals by Radial Immunodiffusion and Automated Turbidimetric Immunoassay, *J. Vet. Intern. Med.*, 2005, 93–96.
- 94 R. A. Ferris and P. M. McCue, How to Use a Quantitative Turbidimetric Immunoassay Assay to Determine Immunoglobulin G Concentrations in Neonatal Foals, *AAEP Proc.*, 2009, **55**, 45–47.
- 95 S. L. Branch and P. N. Levett, Evaluation of Four Methods for Detection of Immunoglobulin M Antibodies to Dengue Virus, *Clin. Diagn. Lab. Immunol.*, 1999, **6**(4), 555–557.
- 96 R. Alende, Serum levels of immunoglobulins (IgG, IgA, IgM) in a general adult population and their relationship

- with alcohol consumption, smoking and common metabolic abnormalities, *Clin. Exp. Immunol.*, 2007, **42**, 42–50.
- 97 V. Garcia Moreira, N. Beridze Vaktangova, M. D. Martinez Gago, B. Laborda Gonzalez, S. Garcia Alonso and E. Fernandez Rodriguez, Overestimation of Albumin Measured by Bromocresol Green vs Bromocresol Purple Method: Influence of Acute-Phase Globulins, *Lab. Med.*, 2018, **49**(4), 355–361.
 - 98 Y. Uchida, Y. Okuzumi, M. Fujishiro, K. Kawamura, M. Shibasaki, N. Shimetani, *et al.*, Controversies in the determination of serum albumin concentration in chronic liver diseases, *Rinsho Byori*, 2006, **54**(10), 1008–1012.
 - 99 A. Watanabe, S. Matsuzaki and H. Moriwaki, Problems in Serum Albumin Measurement and Clinical Significance of Albumin Microheterogeneity in Cirrhotics, *Nutrition*, 2004, **20**(4), 351–357.
 - 100 S. Smolcic, Validation of methods performance for routine biochemistry analytes at Cobas 6000 analyzer series module c501, *Biochem. Med.*, 2011, **21**(2), 182–190.
 - 101 J. L. Luque-Garcia and T. A. Neubert, Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry, *J. Chromatogr. A*, 2007, **1153**(1–2), 259–276.
 - 102 O. Chertov, A. Biragyn, L. W. Kwak, J. T. Simpson, T. Boronina, V. M. Hoang, *et al.*, Organic solvent extraction of proteins and peptides from serum as an effective sample preparation for detection and identification of biomarkers by mass spectrometry, *Proteomics*, 2004, **4**(4), 1195–1203.
 - 103 R. L. Sparrow, R. J. Simpson and D. W. Greening, A Protocol for the Preparation of Cryoprecipitate and Cryo-depleted Plasma for Proteomic Studies, *Methods Mol. Biol.*, 2017, **1619**, 23–30.
 - 104 I. Finoulst, M. Pinkse, W. Van Dongen and P. Verhaert, Sample preparation techniques for the untargeted LC-MS-based discovery of peptides in complex biological matrices, *J. Biomed. Biotechnol.*, 2011, 2011.

Discussion et perspectives

Quantification des protéines

Caractéristiques spectrales des composés protéiques

Les spectres Raman obtenus par l'analyse de solutions d'albumine et de gammaglobulines sont similaires. On retrouve notamment des modes vibrationnels communs à 760 cm^{-1} , 1450 cm^{-1} et 1650 cm^{-1} . A l'inverse, on retrouve des signaux spécifiques de chaque protéine (940 cm^{-1} pour l'albumine et 1240 cm^{-1} et 1550 cm^{-1} pour les gammaglobulines). L'ensemble des signaux observés permet d'envisager la quantification de l'albumine et des gammaglobulines grâce aux bandes spectrales spécifiques mais également la quantification des protéines totales grâce aux signaux communs entre les différents spectres.

Quantification de l'albumine dans le concentrat

Les bandes du spectre Raman de l'albumine situées à 900 cm^{-1} , 940 cm^{-1} , 1000 cm^{-1} , 1100 cm^{-1} , 1340 cm^{-1} , 1450 cm^{-1} et 1650 cm^{-1} sont visibles sur le coefficient de régression du modèle PLSR. De plus l'absence de signal à 1240 cm^{-1} et 1550 cm^{-1} est en faveur de l'absence d'interférence avec les immunoglobulines de l'échantillon. Dans cette étude, le modèle PLSR construit à partir des spectres Raman a permis de prédire les concentrations en albumine des échantillons avec une erreur moyenne d'environ $0,9\text{ g/L}$. Cette erreur de moyenne est faible au regard des recommandations pour une analyse quantitative, d'autant plus que la précision des valeurs de références sont de l'ordre de 1 g/L . Ce résultat permet d'envisager l'utilisation de ce protocole et de cet outil pour une analyse quantitative d'albumine.

Ce résultat permet également de confirmer la conservation de la proportionnalité entre la concentration d'albumine du sérum avant filtration et celle du concentrat après filtration.

Quantification du glucose et de l'urée dans le filtrat

Caractéristiques spectrales des composés

Le spectre de référence de l'urée par spectroscopie Raman est composé de modes vibrationnels à 540 cm^{-1} , 600 cm^{-1} , 1000 cm^{-1} , 1180 cm^{-1} , 1470 cm^{-1} . Le pic de plus grande intensité étant celui situé à 1000 cm^{-1} , imputable aux vibrations d'élongations des liaisons C-N (228)). Pour l'étude de l'urée, le modèle PLSR a été construit à partir de la région spectrale allant de 800 à 1030 cm^{-1} .

Le spectre de référence du glucose par spectroscopie Raman est composé d'un grand nombre de modes de vibrations. La majorité se situe entre 1000 cm^{-1} et 1500 cm^{-1} . Les intensités maximum sont observées à 1030 cm^{-1} , 1100 cm^{-1} , 1380 cm^{-1} et 1450 cm^{-1} . Pour s'affranchir de l'interférence potentielle avec l'urée à 1000 cm^{-1} , le modèle PLSR pour l'étude du glucose a été construit à partir de la région spectrale allant de 1030 cm^{-1} à 1400 cm^{-1} .

Quantification dans le filtrat

Les modèle PLSR construit à partir des spectres Raman ont permis de prédire les concentrations en glucose et en urée des échantillons avec des erreurs moyennes respectives d'environ 18 mg/L et 11 mg/L . Ces erreurs de prédiction sont faible au regard des recommandations pour une analyse quantitative. Ce résultat permet

d'envisager l'utilisation de ce protocole et de cet outil pour une analyse quantitative de la fraction de bas poids moléculaire. Ce résultat permet également de confirmer la conservation de la proportionnalité entre la concentration de glucose et d'urée du sérum avant filtration et celle du filtrat après filtration. Bien que la spectroscopie ATR-IR et la spectroscopie Raman soient toutes les deux des spectroscopies vibrationnelles, les intensités des signaux mesurés ne sont pas identiques entre les deux techniques. En effet les liaisons polaires ont un signal de forte intensité en infrarouge alors que les liaisons apolaires « symétriques » ont un meilleur signal en Raman. Les études menées dans ce travail ont confirmé ce phénomène. En effet, les concentrations en urée ont été mieux prédites par spectroscopie Raman que par spectroscopie ATR-IR (respectivement 0,18 mmol/L et 2,7 mmol/L). A l'inverse, la spectroscopie ATR-IR a permis de prédire avec une grande précision la concentration en protéines totales dans le sérum, contrairement à la spectroscopie Raman.

Conclusion

L'analyse des différentes sous-fractions du plasma a montré que la spectroscopie de diffusion Raman est capable d'extraire des informations spécifiques d'une matrice complexe à l'aide des modèles de traitement statistiques par PLS. En effet, les modèles de prédiction des concentrations de glucose et d'urée sont construits à partir de signaux spécifiques. Cette étude démontre que la spectroscopie Raman, associée à l'ultrafiltration et à l'analyse PLSR, peut être utilisée pour détecter les variations de plusieurs analytes à partir d'un unique échantillon de plasma avec une bonne précision. Comme précédemment démontré pour l'analyse par spectroscopie ATR-IR (Partie II, Etude n°2, Bonnier, 2017 (199)), le fractionnement du sérum a

considérablement amélioré la capacité de la spectroscopie Raman à détecter et quantifier les analytes de la fraction de bas poids moléculaire. L'approche proposée a permis de construire des modèles de prédiction à partir du plasma entier (protéine totale, gammaglobuline), du concentrat (albumine) et du filtrat (urée et glucose) des échantillons de patients tout en améliorant les performances analytiques. Ces résultats confirment également l'intérêt d'analyser les différentes fractions du plasma pour obtenir l'information la plus complète sur sa composition. Compte tenu de son faible coût, de sa mise en œuvre facile, de ses résultats rapides et de sa sensibilité et spécificité plus élevées, cette technique pourrait devenir une alternative aux méthodes de quantification en laboratoire existantes et proposer un bilan de dépistage global.

Hackshaw et coll ont récemment réussi à mettre en évidence des différences de composition du plasma chez des patients atteints de fibromyalgie, de lupus érythémateux disséminé et de polyarthrite rhumatoïde à l'aide de la spectroscopie de diffusion Raman (229). Selon les auteurs, les signaux discriminants sont principalement dus aux vibrations du squelette carboné des protéines. Or d'après nos résultats concernant la spectroscopie infrarouge mais également la spectroscopie de diffusion Raman, l'analyse du plasma sans préparation est limitée pour l'analyse des composés minoritaires dont les signaux sont masqués par celui des protéines. L'acquisition des spectres provenant des différentes sous fractions permettrait d'obtenir une vision plus complète de la composition de l'échantillon et d'augmenter le pouvoir discriminant de ces outils. Une étude a confirmé cette hypothèse en étudiant le plasma, ainsi que les petites et grosses vésicules extracellulaires circulantes (sEV et IEV) dans le plasma chez le patient SLA et en comparant les spectres de diffusion Raman à ceux obtenus sur les mêmes sous

fractions de plasmas témoins (230). Les résultats obtenus ont mis en évidence une complémentarité des analyses réalisées sur les sEV et sur les IEV. En effet, malgré les similarités observées sur les spectres Raman mesurés sur chaque fraction, on peut observer également des différences représentatives de variations dans leur composition (Figure 34).

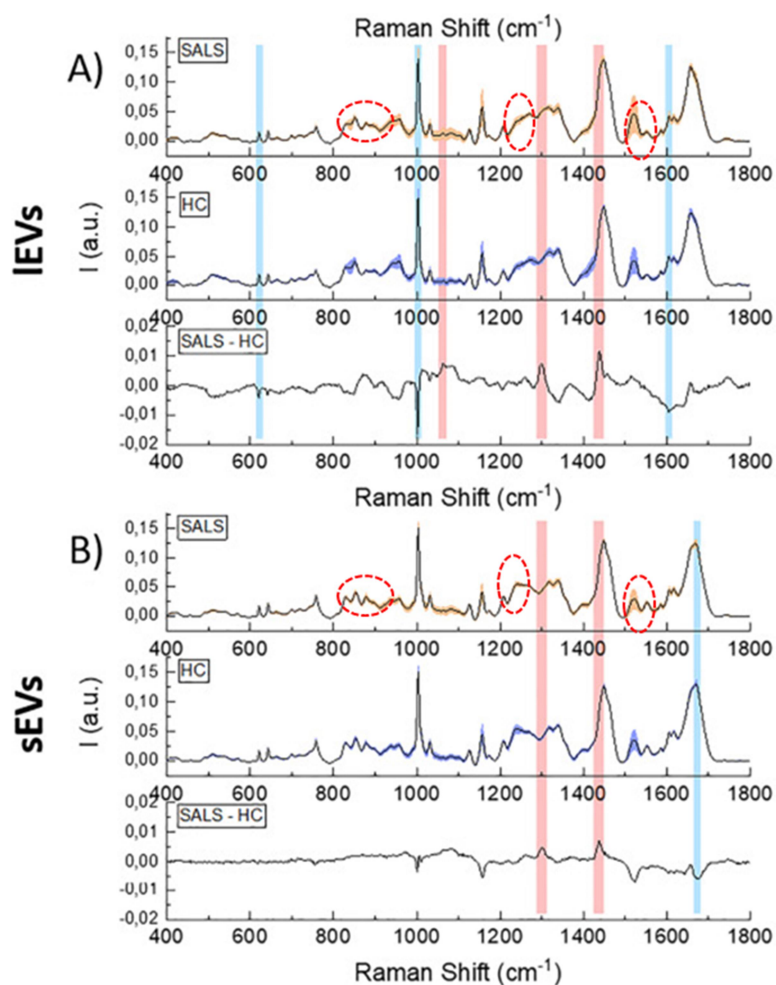


Figure 34 : Spectres de diffusion Raman des fractions sEV et IEV des échantillons patients SLA (SALS) et témoins (HC). Les spectres SALS-HC représentent les différences retrouvées entre les deux groupes (adapté de Morasso et coll (230))

De plus, des différences concernant la composition protéique dans la bande amide I (1670cm^{-1}) n'a été retrouvé que dans la fraction sEV et des différences concernant la composition des protéines en phénylalanine (621cm^{-1} , 1002cm^{-1} , 1604cm^{-1}) n'a été

retrouvé que dans la fraction IEV. Les courbes ROC obtenues montrent une AUC comprise entre 64% pour les fractions sEV et le plasma et 85% pour la fraction IEV. Malheureusement, les auteurs n'ont pas essayé de combiner les données provenant des différentes fractions pour améliorer le modèle.

CONCLUSION

Au cours de ce travail, trois axes majeurs ont été abordés à savoir : (1) l'étude de la combinaison des outils analytiques conventionnels utilisés dans une approche métabolomique, (2) le développement d'approches quantitatives basées sur les spectroscopies vibrationnelles (Raman et ATR-IR), couplées aux méthodes prédictives quantitatives, et (3) l'application de l'ATR-IR dans la SLA.

Dans le cadre du premier axe, nous avons déterminé qu'il n'y a pas actuellement de méthode universelle basée sur les techniques conventionnelles. De plus l'utilisation d'un grand nombre de méthodes d'analyse permet d'augmenter la couverture métabolique mais avec l'inconvénient d'augmenter le coût et la difficulté pour interpréter les données. Nous avons également pu montrer que la combinaison considérée comme le meilleur compromis est différente d'un tissu ou d'un fluide biologique à l'autre. La spectroscopie vibrationnelle telle que la spectroscopie ATR-IR et la spectroscopie Raman nécessitant peu de préparation d'échantillon, elle nous a paru comme étant intéressante comme analyse préalable ainsi qu'en complément de la spectrométrie de masse.

Dans le cadre du deuxième axe, différents aspects ont été abordés du point de vue des conditions expérimentales et de l'analyse de données spectrales. Les potentialités des spectroscopies Raman et moyen IR couplées à des méthodes statistiques multivariées ont été mises en évidence, pour l'analyse quantitative directe des molécules présentes dans le plasma. L'intérêt du prétraitement des spectres a également été soulevé dans le cadre de ce travail. Nous avons également

pu observer l'intérêt des différentes analyses sous forme liquide, séchée ou après fractionnement du plasma. Il a été mis en évidence que la précision des modèles prédictifs quantitatifs peut être influencée par la quantité de données spectrales présentes dans la base de référence ainsi que leur qualité.

Pour finir, dans le troisième axe de ce travail nous avons confirmé le potentiel de la spectroscopie ATR-IR à discriminer deux populations différentes selon leur composition chimique, et ce malgré l'absence d'identification des molécules d'intérêt. Les résultats obtenus dans cette dernière partie nécessitent une confirmation sur un nombre d'échantillons plasmatiques plus important. De plus, l'utilisation de l'ATR-IR pour l'analyse de fluides précieux tels que le LCR semble prometteuse, notamment dans le cas d'une pathologie neurodégénérative comme la SLA. Combiner les données obtenues sur les échantillons plasmatiques et ceux de LCR pourrait permettre la mise en évidence de corrélation entre les différentes matrices. Dans la SLA, il est également connu que la protéine TDP-43 s'agrège dans les cellules motoneuronales. Dans ce contexte, l'ATR-IR et la spectroscopie Raman seraient également capables d'observer ces agrégats à l'aide de microscopes confocaux.

En dernier lieu, un des principaux atouts de ces techniques est la faible quantité d'échantillon nécessaire pour l'analyse. Elles sont des candidates très intéressantes pour l'analyse de fluides rares tels que les larmes ou la périlymphe. En effet, le liquide de périlymphe est présent en volume particulièrement faible puisque seuls quelques microlitres d'échantillon peuvent être récupérés lors de l'ouverture de la cochlée et que ce liquide n'est pas régénéré.

Une étude préliminaire a été réalisée à Tours dans le cadre d'un Master 2 en partenariat avec le service d'Ophtalmologie-ORL et a montré une très bonne reproductibilité analytique ainsi qu'une excellente capacité à observer des différences entre plusieurs échantillons. De plus, le service d'Ophtalmologie-ORL s'intéresse également à l'approche métabolomique pour l'analyse du liquide de périlymphe. Les résultats encourageants se sont confrontés à une limite de sensibilité de la spectrométrie de masse qui pourrait être palliée par l'utilisation de l'ATR-IR ou de la spectroscopie Raman.

REFERENCES

1. Talbott EO, Malek AM, Lacomis D. The epidemiology of amyotrophic lateral sclerosis. *Handb Clin Neurol*. 2016;138:225-38.
2. Longinetti E, Fang F. Epidemiology of amyotrophic lateral sclerosis: an update of recent literature. *Curr Opin Neurol*. oct 2019;32(5):771-6.
3. Couratier P, Marin B, Lautrette G, Nicol M, Preux P-M. Épidémiologie, spectre clinique de la SLA et diagnostics différentiels. *La Presse Médicale*. 1 mai 2014;43(5):538-48.
4. Pourquoi la sclérose latérale amyotrophique reste-t-elle encore incurable ? [Internet]. Planet-Vie. [cité 4 oct 2020]. Disponible sur: <https://planet-vie.ens.fr/thematiques/sante/pathologies/pourquoi-la-sclerose-laterale-amyotrophique-reste-t-elle-encore>
5. Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature*. 4 mars 1993;362(6415):59-62.
6. Vance C, Rogelj B, Hortobágyi T, De Vos KJ, Nishimura AL, Sreedharan J, et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science*. 27 févr 2009;323(5918):1208-11.
7. Sreedharan J, Blair IP, Tripathi VB, Hu X, Vance C, Rogelj B, et al. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science*. 21 mars 2008;319(5870):1668-72.
8. Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*. 20 oct 2011;72(2):257-68.
9. Therrien M, Dion PA, Rouleau GA. ALS: Recent Developments from Genetics Studies. *Curr Neurol Neurosci Rep*. juin 2016;16(6):59.

10. Renton AE, Chiò A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci.* janv 2014;17(1):17-23.
11. Chia R, Chiò A, Traynor BJ. Novel genes associated with amyotrophic lateral sclerosis: diagnostic and clinical implications. *Lancet Neurol.* 2018;17(1):94-102.
12. Bannwarth S, Ait-El-Mkadem S, Chaussenot A, Genin EC, Lacas-Gervais S, Fragaki K, et al. A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. *Brain.* août 2014;137(Pt 8):2329-45.
13. Ajroud-Driss S, Siddique T. Sporadic and hereditary amyotrophic lateral sclerosis (ALS). *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease.* 1 avr 2015;1852(4):679-84.
14. Cluskey S, Ramsden DB. Mechanisms of neurodegeneration in amyotrophic lateral sclerosis. *MP, Mol Pathol.* déc 2001;54(6):386-92.
15. Doble A. The role of excitotoxicity in neurodegenerative disease: implications for therapy. *Pharmacol Ther.* mars 1999;81(3):163-221.
16. Lanznaster D, de Assis DR, Corcia P, Pradat P-F, Blasco H. Metabolomics biomarkers: A strategy toward therapeutics improvement in ALS. *Frontiers in neurology.* 2018;9.
17. Musarò A. State of the art and the dark side of amyotrophic lateral sclerosis. *World J Biol Chem.* 26 mai 2010;1(5):62-8.
18. McGeer PL, McGeer EG. Inflammatory processes in amyotrophic lateral sclerosis. *Muscle Nerve.* oct 2002;26(4):459-70.
19. Thonhoff JR, Simpson EP, Appel SH. Neuroinflammatory mechanisms in amyotrophic lateral sclerosis pathogenesis. *Curr Opin Neurol.* oct 2018;31(5):635-9.
20. Watanabe M, Dykes-Hoberg M, Culotta VC, Price DL, Wong PC, Rothstein JD. Histological evidence of protein aggregation in mutant SOD1 transgenic mice

and in amyotrophic lateral sclerosis neural tissues. *Neurobiol Dis.* déc 2001;8(6):933-41.

21. Maurel C, Dangoumau A, Marouillat S, Brulard C, Chami A, Hergesheimer R, et al. Causative Genes in Amyotrophic Lateral Sclerosis and Protein Degradation Pathways: a Link to Neurodegeneration. *Mol Neurobiol.* août 2018;55(8):6480-99.
22. Kurtishi A, Rosen B, Patil KS, Alves GW, Møller SG. Cellular Proteostasis in Neurodegeneration. *Mol Neurobiol.* 4 sept 2018;
23. Lawton KA, Cudkowicz ME, Brown MV, Alexander D, Caffrey R, Wulff JE, et al. Biochemical alterations associated with ALS. *Amyotroph Lateral Scler.* janv 2012;13(1):110-8.
24. Lawton KA, Brown MV, Alexander D, Li Z, Wulff JE, Lawson R, et al. Plasma metabolomic biomarker panel to distinguish patients with amyotrophic lateral sclerosis from disease mimics. *Amyotroph Lateral Scler Frontotemporal Degener.* sept 2014;15(5-6):362-70.
25. Wang M-D, Little J, Gomes J, Cashman NR, Krewski D. Identification of risk factors associated with onset and progression of amyotrophic lateral sclerosis using systematic review and meta-analysis. *Neurotoxicology.* 2017;61:101-30.
26. Gordon PH, Miller RG, Moore DH. ALSFRS-R. *Amyotroph Lateral Scler Other Motor Neuron Disord.* sept 2004;5 Suppl 1:90-3.
27. Miller RG, Mitchell JD, Lyon M, Moore DH. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). *Amyotroph Lateral Scler Other Motor Neuron Disord.* sept 2003;4(3):191-206.
28. Miller RG, Mitchell JD, Moore DH. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). *Cochrane Database Syst Rev.* 14 mars 2012;(3):CD001447.
29. Oskarsson B, Gendron TF, Staff NP. Amyotrophic Lateral Sclerosis: An Update for 2018. *Mayo Clin Proc.* nov 2018;93(11):1617-28.

30. Bhandari R, Kuhad A, Kuhad A. Edaravone: a new hope for deadly amyotrophic lateral sclerosis. *Drugs Today*. juin 2018;54(6):349-60.
31. Abe K, Aoki M, Tsuji S, Itoyama Y, Sobue G, Togo M, et al. Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial. *The Lancet Neurology*. 2017;16(7):505-12.
32. de Carvalho M, Pinto S, Costa J, Evangelista T, Ohana B, Pinto A. A randomized, placebo-controlled trial of memantine for functional disability in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler*. oct 2010;11(5):456-60.
33. Miller RG, Block G, Katz JS, Barohn RJ, Gopalakrishnan V, Cudkowicz M, et al. Randomized phase 2 trial of NP001-a novel immune regulator: Safety and early efficacy in ALS. *Neurol Neuroimmunol Neuroinflamm*. juin 2015;2(3):e100.
34. Kaufmann P, Thompson JLP, Levy G, Buchsbaum R, Shefner J, Krivickas LS, et al. Phase II trial of CoQ10 for ALS finds insufficient evidence to justify phase III. *Ann Neurol*. août 2009;66(2):235-44.
35. Lenglet T, Lacomblez L, Abitbol JL, Ludolph A, Mora JS, Robberecht W, et al. A phase II-III trial of olesoxime in subjects with amyotrophic lateral sclerosis. *Eur J Neurol*. mars 2014;21(3):529-36.
36. Martínez-Muriana A, Mancuso R, Francos-Quijorna I, Olmos-Alonso A, Osta R, Perry VH, et al. CSF1R blockade slows the progression of amyotrophic lateral sclerosis by reducing microgliosis and invasion of macrophages into peripheral nerves. *Sci Rep*. 13 2016;6:25663.
37. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*. mars 2001;69(3):89-95.
38. Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron*. 1976;16(1):31-41.

39. Selvin E, Marinopoulos S, Berkenblit G, Rami T, Brancati FL, Powe NR, et al. Meta-analysis: glycosylated hemoglobin and cardiovascular disease in diabetes mellitus. *Ann Intern Med.* 21 sept 2004;141(6):421-31.
40. Reflexion prospective autour des biomarqueurs. 2009;4.
41. Fraser GAM, Meyer RM. Biomarkers and the design of clinical trials in cancer. *Biomark Med.* oct 2007;1(3):387-97.
42. McShane LM, Hunsberger S, Adjei AA. Effective incorporation of biomarkers into phase II trials. *Clin Cancer Res.* 15 mars 2009;15(6):1898-905.
43. Harbeck N, Gnant M. Breast cancer. *Lancet.* 18 2017;389(10074):1134-50.
44. Birger A, Ben-Dor I, Ottolenghi M, Turetsky T, Gil Y, Sweetat S, et al. Human iPSC-derived astrocytes from ALS patients with mutated C9ORF72 show increased oxidative stress and neurotoxicity. *EBioMedicine.* déc 2019;50:274-89.
45. Bruno C, Patin F, Bocca C, Nadal-Desbarats L, Bonnier F, Reynier P, et al. The combination of four analytical methods to explore skeletal muscle metabolomics: Better coverage of metabolic pathways or a marketing argument? *J Pharm Biomed Anal.* 30 janv 2018;148:273-9.
46. Ren Y, Liu W, Li Y, Sun B, Li Y, Yang F, et al. Cutaneous somatic and autonomic nerve TDP-43 deposition in amyotrophic lateral sclerosis. *J Neurol.* août 2018;265(8):1753-63.
47. McGill MR, Jaeschke H. Biomarkers of Drug-induced Liver Injury. *Adv Pharmacol.* 2019;85:221-39.
48. Lanznaster D, de Assis DR, Corcia P, Pradat P-F, Blasco H. Metabolomics Biomarkers: A Strategy Toward Therapeutics Improvement in ALS. *Front Neurol* [Internet]. 18 déc 2018 [cité 14 sept 2020];9. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6305341/>
49. Contrepois K, Jiang L, Snyder M. Optimized Analytical Procedures for the Untargeted Metabolomic Profiling of Human Urine and Plasma by Combining

Hydrophilic Interaction (HILIC) and Reverse-Phase Liquid Chromatography (RPLC)-Mass Spectrometry. *Mol Cell Proteomics*. juin 2015;14(6):1684-95.

50. Cicalini I, Rossi C, Pieragostino D, Agnifili L, Mastropasqua L, di Iorio M, et al. Integrated Lipidomics and Metabolomics Analysis of Tears in Multiple Sclerosis: An Insight into Diagnostic Potential of Lacrimal Fluid. *Int J Mol Sci* [Internet]. 13 mars 2019 [cité 21 nov 2020];20(6). Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6471885/>
51. Spear BB, Heath-Chiozzi M, Huff J. Clinical application of pharmacogenetics. *Trends Mol Med*. mai 2001;7(5):201-4.
52. Papetti F, Morel-Pingault V, Buisse V, Maziere L, Banayan M, Thauby S, et al. [Clozapine-resistant schizophrenia related to an increased metabolism and benefit of fluvoxamine: four case reports]. *Encephale*. oct 2007;33(5):811-8.
53. Blasco H, Patin F, Descat A, Garçon G, Corcia P, Gelé P, et al. A pharmacometabolomics approach in a clinical trial of ALS: Identification of predictive markers of progression. *PLoS One* [Internet]. 5 juin 2018 [cité 14 sept 2020];13(6). Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5988280/>
54. Balashova EE, Maslov DL, Lokhov PG. A Metabolomics Approach to Pharmacotherapy Personalization. *J Pers Med*. 5 sept 2018;8(3).
55. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass spectrometry reviews*. 2007;26(1):51-78.
56. Günther UL. Metabolomics Biomarkers for Breast Cancer. *Pathobiology*. 2015;82(3-4):153-65.
57. Hoffman JM, Lyu Y, Pletcher SD, Promislow DEL. Proteomics and metabolomics in ageing research: from biomarkers to systems biology. *Essays in biochemistry*. 11 juill 2017;61(3):379-88.

58. Lauwen S, de Jong EK, Lefeber DJ, den Hollander AI. Omics Biomarkers in Ophthalmology. *Investigative Ophthalmology & Visual Science*. 21 juill 2017;58(6):BIO88-98.
59. Fuhrer T, Zampieri M, Sévin DC, Sauer U, Zamboni N. Genomewide landscape of gene-metabolome associations in *Escherichia coli*. *Mol Syst Biol*. 16 2017;13(1):907.
60. Tang W, Hazebroek J, Zhong C, Harp T, Vlahakis C, Baumhover B, et al. Effect of genetics, environment, and phenotype on the metabolome of maize hybrids using GC/MS and LC/MS. *Journal of agricultural and food chemistry*. 2017;65(25):5215-25.
61. Zaitsev K, Hayashi Y, Kusano M, Tsuchihashi H, Ishii A. Application of metabolomics to toxicology of drugs of abuse: A mini review of metabolomics approach to acute and chronic toxicity studies. *Drug Metab Pharmacokinet*. févr 2016;31(1):21-6.
62. Wang JH, Byun J, Pennathur S. Analytical approaches to metabolomics and applications to systems biology. *Semin Nephrol*. sept 2010;30(5):500-11.
63. Beauchercq S, Nadal-Desbarats L, Hennequet-Antier C, Collin A, Tesseraud S, Bourin M, et al. Serum and Muscle Metabolomics for the Prediction of Ultimate pH, a Key Factor for Chicken-Meat Quality. *Journal of proteome research*. 2016;15(4):1168-78.
64. Ramautar R. Capillary Electrophoresis-Mass Spectrometry for Clinical Metabolomics. *Adv Clin Chem*. 2016;74:1-34.
65. Koek MM, Jellema RH, van der Greef J, Tas AC, Hankemeier T. Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. *Metabolomics*. sept 2011;7(3):307-28.
66. Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, et al. HMDB: a knowledgebase for the human metabolome. *Nucleic acids research*. janv 2009;37(Database issue):D603-10.

67. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic acids research*. 4 janv 2018;46(D1):D608-17.
68. Bigler MB, Egli SB, Hysek CM, Hoenger G, Schmied L, Baldin FS, et al. Stress-Induced In Vivo Recruitment of Human Cytotoxic Natural Killer Cells Favors Subsets with Distinct Receptor Profiles and Associates with Increased Epinephrine Levels. Vivier E, éditeur. *PLOS ONE*. 23 déc 2015;10(12):e0145635.
69. Park SK, Ryoo J-H, Oh C-M, Choi J-M, Choi Y-J, Lee KO, et al. The risk of type 2 diabetes mellitus according to 2-h plasma glucose level: The Korean Genome and Epidemiology Study (KoGES). *Diabetes Research and Clinical Practice*. déc 2018;146:130-7.
70. Verpoorte R, Choi YH, Mustafa NR, Kim HK. Metabolomics: back to basics. *Phytochemistry Reviews*. 1 oct 2008;7(3):525-37.
71. Nicholson JK, O'Flynn MP, Sadler PJ, Macleod AF, Juul SM, Sönksen PH. Proton-nuclear-magnetic-resonance studies of serum, plasma and urine from fasting normal and diabetic subjects. *Biochem J*. 15 janv 1984;217(2):365.
72. Nicholson JK, Lindon JC. Metabonomics. *Nature*. 22 oct 2008;455:1054.
73. Soininen Pasi, Kangas Antti J., Würtz Peter, Suna Teemu, Ala-Korpela Mika. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics. *Circulation: Cardiovascular Genetics*. 1 févr 2015;8(1):192-206.
74. Vignoli A, Rodio DM, Bellizzi A, Sobolev AP, Anzivino E, Mischitelli M, et al. NMR-based metabolomic approach to study urine samples of chronic inflammatory rheumatic disease patients. *Analytical and Bioanalytical Chemistry*. 1 févr 2017;409(5):1405-13.
75. Blasco H, Nadal-Desbarats L, Pradat P-F, Gordon PH, Antar C, Veyrat-Durebex C, et al. Untargeted ¹H-NMR metabolomics in CSF: toward a diagnostic biomarker for motor neuron disease. *Neurology*. 1 avr 2014;82(13):1167-74.

76. Schirra HJ, Ford PJ. NMR-Based Metabolomics of Oral Biofluids. In: Seymour GJ, Cullinan MP, Heng NCK, éditeurs. Oral Biology: Molecular Techniques and Applications [Internet]. New York, NY: Springer New York; 2017. p. 79-105. Disponible sur: https://doi.org/10.1007/978-1-4939-6685-1_5
77. Wishart DS. Metabolomics for Investigating Physiological and Pathophysiological Processes. *Physiol Rev*. 01 2019;99(4):1819-75.
78. Bocca C, Kouassi Nzoughet J, Leruez S, Amati-Bonneau P, Ferré M, Kane M-S, et al. A Plasma Metabolomic Signature Involving Purine Metabolism in Human Optic Atrophy 1 (OPA1)-Related Disorders. *Invest Ophthalmol Vis Sci*. 01 2018;59(1):185-95.
79. Khamis MM, Adamko DJ, El-Aneed A. Mass spectrometric based approaches in urine metabolomics and biomarker discovery. *Mass Spectrom Rev*. 2017;36(2):115-34.
80. Blasco H, Veyrat-Durebex C, Bocca C, Patin F, Vourc'h P, Kouassi Nzoughet J, et al. Lipidomics Reveals Cerebrospinal-Fluid Signatures of ALS. *Sci Rep*. 15 2017;7(1):17652.
81. Zheng J, Dixon RA, Li L. Development of isotope labeling LC-MS for human salivary metabolomics and application to profiling metabolome changes associated with mild cognitive impairment. *Anal Chem*. 18 déc 2012;84(24):10802-11.
82. Gonzalez-Riano C, Garcia A, Barbas C. Metabolomics studies in brain tissue: A review. *Journal of Pharmaceutical and Biomedical Analysis*. 25 oct 2016;130:141-68.
83. Xiang L, Zhang H, Wei J, Tian XY, Luan H, Li S, et al. Metabolomics studies on db/db diabetic mice in skeletal muscle reveal effective clearance of overloaded intermediates by exercise. *Analytica Chimica Acta*. 11 déc 2018;1037:130-9.
84. Rafiq MK, Lee E, Bradburn M, McDermott CJ, Shaw PJ. Creatine kinase enzyme level correlates positively with serum creatinine and lean body mass,

- and is a prognostic factor for survival in amyotrophic lateral sclerosis. *Eur J Neurol*. 2016;23(6):1071-8.
85. Lu C-H, Macdonald-Wallis C, Gray E, Pearce N, Petzold A, Norgren N, et al. Neurofilament light chain: a prognostic biomarker in amyotrophic lateral sclerosis. *Neurology*. 2015;84(22):2247-57.
 86. Lawton KA, Brown MV, Alexander D, Li Z, Wulff JE, Lawson R, et al. Plasma metabolomic biomarker panel to distinguish patients with amyotrophic lateral sclerosis from disease mimics. *Amyotroph Lateral Scler Frontotemporal Degener*. sept 2014;15(5-6):362-70.
 87. Kori M, Aydın B, Unal S, Arga KY, Kazan D. Metabolic Biomarkers and Neurodegeneration: A Pathway Enrichment Analysis of Alzheimer's Disease, Parkinson's Disease, and Amyotrophic Lateral Sclerosis. *OMICS*. nov 2016;20(11):645-61.
 88. Wuolikainen A, Jonsson P, Ahnlund M, Antti H, Marklund SL, Moritz T, et al. Multi-platform mass spectrometry analysis of the CSF and plasma metabolomes of rigorously matched amyotrophic lateral sclerosis, Parkinson's disease and control subjects. *Mol Biosyst*. avr 2016;12(4):1287-98.
 89. Wang H, Wang W, Shi H, Han L, Pan P. Cerebrospinal fluid and blood levels of neurofilament light chain in Parkinson disease: A protocol for systematic review and meta-analysis. *Medicine (Baltimore)*. 31 juill 2020;99(31):e21458.
 90. Varhaug KN, Torkildsen Ø, Myhr K-M, Vedeler CA. Neurofilament Light Chain as a Biomarker in Multiple Sclerosis. *Front Neurol*. 2019;10:338.
 91. Brureau A, Blanchard-Bregeon V, Pech C, Hamon S, Chaillou P, Guillemot J-C, et al. NF-L in cerebrospinal fluid and serum is a biomarker of neuronal damage in an inducible mouse model of neurodegeneration. *Neurobiol Dis*. août 2017;104:73-84.
 92. Kuhle J, Barro C, Andreasson U, Derfuss T, Lindberg R, Sandelius Å, et al. Comparison of three analytical platforms for quantification of the neurofilament light chain in blood samples: ELISA, electrochemiluminescence immunoassay

- and Simoa. *Clinical Chemistry and Laboratory Medicine (CCLM)*. 1 oct 2016;54(10):1655-61.
93. Blasco H, Mavel S, Corcia P, Gordon PH. The glutamate hypothesis in ALS: pathophysiology and drug development. *Curr Med Chem*. 2014;21(31):3551-75.
 94. Spreux-Varoquaux O, Bensimon G, Lacomblez L, Salachas F, Pradat PF, Le Forestier N, et al. Glutamate levels in cerebrospinal fluid in amyotrophic lateral sclerosis: a reappraisal using a new HPLC method with coulometric detection in a large cohort of patients. *J Neurol Sci*. 15 janv 2002;193(2):73-8.
 95. Zhao W, Xie W, Le W, Beers DR, He Y, Henkel JS, et al. Activated microglia initiate motor neuron injury by a nitric oxide and glutamate-mediated mechanism. *J Neuropathol Exp Neurol*. sept 2004;63(9):964-77.
 96. Blasco H, Garcon G, Patin F, Veyrat-Durebex C, Boyer J, Devos D, et al. Panel of Oxidative Stress and Inflammatory Biomarkers in ALS: A Pilot Study. *Can J Neurol Sci*. janv 2017;44(1):90-5.
 97. Tohgi H, Abe T, Yamazaki K, Murata T, Ishizaki E, Isobe C. Increase in oxidized NO products and reduction in oxidized glutathione in cerebrospinal fluid from patients with sporadic form of amyotrophic lateral sclerosis. *Neurosci Lett*. 5 févr 1999;260(3):204-6.
 98. Ihara Y, Nobukuni K, Takata H, Hayabara T. Oxidative stress and metal content in blood and cerebrospinal fluid of amyotrophic lateral sclerosis patients with and without a Cu, Zn-superoxide dismutase mutation. *Neurol Res*. janv 2005;27(1):105-8.
 99. Tohgi H, Abe T, Yamazaki K, Murata T, Ishizaki E, Isobe C. Remarkable increase in cerebrospinal fluid 3-nitrotyrosine in patients with sporadic amyotrophic lateral sclerosis. *Ann Neurol*. juill 1999;46(1):129-31.
 100. Cutler RG, Pedersen WA, Camandola S, Rothstein JD, Mattson MP. Evidence that accumulation of ceramides and cholesterol esters mediates oxidative stress-induced death of motor neurons in amyotrophic lateral sclerosis. *Ann Neurol*. oct 2002;52(4):448-57.

101. Lanznaster D, de Assis DR, Corcia P, Pradat P-F, Blasco H. Metabolomics Biomarkers: A Strategy Toward Therapeutics Improvement in ALS. *Front Neurol* [Internet]. 18 déc 2018 [cité 14 sept 2020];9. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6305341/>
102. Blasco H, Vourc'h P, Pradat PF, Gordon PH, Andres CR, Corcia P. Further development of biomarkers in amyotrophic lateral sclerosis. *Expert Rev Mol Diagn.* août 2016;16(8):853-68.
103. Atassi N. When a negative trial in ALS has a positive effect on research. *Lancet Neurol.* 2018;17(8):653-5.
104. Ryberg H, Bowser R. Protein biomarkers for amyotrophic lateral sclerosis. *Expert Rev Proteomics.* avr 2008;5(2):249-62.
105. Mitropoulos K, Katsila T, Patrinos GP, Pampalakis G. Multi-Omics for Biomarker Discovery and Target Validation in Biofluids for Amyotrophic Lateral Sclerosis Diagnosis. *OMICS.* 2018;22(1):52-64.
106. Hughes C, Brown M, Clemens G, Henderson A, Monjardez G, Clarke NW, et al. Assessing the challenges of FTIR spectroscopic analysis of blood serum. *Journal of Biophotonics.* avr 2014;7(3-4):180-8.
107. Baker M. Biophotonics: Vibrational Spectroscopic Diagnostics [Internet]. IOP Publishing; 2016 [cité 19 juill 2017]. Disponible sur: <http://iopscience.iop.org/book/978-1-6817-4071-3>
108. Baker MJ, Hussain SR, Lovergne L, Untereiner V, Hughes C, Lukaszewski RA, et al. Developing and understanding biofluid vibrational spectroscopy: a critical review. *Chem Soc Rev.* 2016;45(7):1803-18.
109. Hughes C, Baker MJ. Can mid-infrared biomedical spectroscopy of cells, fluids and tissue aid improvements in cancer survival? A patient paradigm. *Analyst.* 21 janv 2016;141(2):467-75.

110. Kong K, Kendall C, Stone N, Notingher I. Raman spectroscopy for medical diagnostics — From in-vitro biofluid assays to in-vivo cancer detection. *Advanced Drug Delivery Reviews*. juill 2015;89:121-34.
111. Bunaciu AA, Fleschin Ş, Hoang VD, Aboul-Enein HY. Vibrational Spectroscopy in Body Fluids Analysis. *Critical Reviews in Analytical Chemistry*. 2 janv 2017;47(1):67-75.
112. Shaw RA, Low-Ying S, Man A, Liu K-Z, Mansfield C, Rileg CB, et al. 'Infrared Spectroscopy of Biofluids in Clinical Chemistry and Medical Diagnostics. *Biomedical Vibrational Spectroscopy* Hoboken, NJ: John Wiley and Sons, Inc. 2008;79-103.
113. Byrne HJ, Baranska M, Puppels GJ, Stone N, Wood B, Gough KM, et al. Spectropathology for the next generation: Quo vadis? *The Analyst*. 2015;140(7):2066-73.
114. Mitchell AL, Gajjar KB, Theophilou G, Martin FL, Martin-Hirsch PL. Vibrational spectroscopy of biofluids for disease screening or diagnosis: translation from the laboratory to a clinical setting: Vibrational spectroscopy of biofluids: laboratory to clinical setting. *Journal of Biophotonics*. avr 2014;7(3-4):153-65.
115. Eberhardt K, Stiebing C, Matthäus C, Schmitt M, Popp J. Advantages and limitations of Raman spectroscopy for molecular diagnostics: an update. *Expert Rev Mol Diagn*. juin 2015;15(6):773-87.
116. Walton AG, Deveney MJ, Koenig JL. Raman spectroscopy of calcified tissue. *Calcified Tissue Research*. 1 déc 1970;6(1):162-7.
117. Lord RC, Yu NT. Laser-excited Raman spectroscopy of biomolecules. I. Native lysozyme and its constituent amino acids. *J Mol Biol*. 14 juin 1970;50(2):509-24.
118. Tobin MC. Raman Spectra of Crystalline Lysozyme, Pepsin, and Alpha Chymotrypsin. *Science*. 5 juill 1968;161(3836):68-9.
119. Tomasi TB, Tisdale WA. SERUM GAMMA-GLOBULINS IN ACUTE AND CHRONIC LIVER DISEASES. *Nature*. 22 févr 1964;201:834-5.

120. Gazzard BG. HIV disease and the gastroenterologist. *Gut*. nov 1988;29(11):1497-505.
121. Tian C-R, Qian L, Shen X-Z, Li J-J, Wen J-T. Distribution of serum total protein in elderly Chinese. *PLoS ONE*. 2014;9(6):e101242.
122. Mistek E, Lednev IK. Identification of species' blood by attenuated total reflection (ATR) Fourier transform infrared (FT-IR) spectroscopy. *Anal Bioanal Chem*. sept 2015;407(24):7435-42.
123. Barlev E, Zelig U, Bar O, Segev C, Mordechai S, Kapelushnik J, et al. A novel method for screening colorectal cancer by infrared spectroscopy of peripheral blood mononuclear cells and plasma. *J Gastroenterol*. mars 2016;51(3):214-21.
124. Lopes J, Correia M, Martins I, Henriques AG, Delgadillo I, da Cruz E Silva O, et al. FTIR and Raman Spectroscopy Applied to Dementia Diagnosis Through Analysis of Biological Fluids. *J Alzheimers Dis*. 08 2016;52(3):801-12.
125. Martin M, Perez-Guaita D, Andrew DW, Richards JS, Wood BR, Heraud P. Detection and Quantification of Plasmodium falciparum in Aqueous Red Blood Cells by Attenuated Total Reflection Infrared Spectroscopy and Multivariate Data Analysis. *Journal of Visualized Experiments* [Internet]. 2 nov 2018 [cité 30 janv 2019];(141). Disponible sur: <https://www.jove.com/video/56797/detection-quantification-plasmodium-falciparum-aqueous-red-blood>
126. Roy S, Perez-Guaita D, Andrew DW, Richards JS, McNaughton D, Heraud P, et al. Simultaneous ATR-FTIR Based Determination of Malaria Parasitemia, Glucose and Urea in Whole Blood Dried onto a Glass Slide. *Anal Chem*. 16 2017;89(10):5238-45.
127. Paraskevaïdi M, Morais CLM, Lima KMG, Snowden JS, Saxon JA, Richardson AMT, et al. Differential diagnosis of Alzheimer's disease using spectrochemical analysis of blood. *Proceedings of the National Academy of Sciences*. 19 sept 2017;114(38):E7929-38.

128. Chowdhury MSI, Akhter N, Haque M, Aziz R, Nahar N. Serum Total Protein and Albumin Levels in Different Grades of Protein Energy Malnutrition. *Journal of Bangladesh Society of Physiologist*. 1 janv 1970;3:58-60.
129. Yang J, Tan X, Shih W-C, Cheng MM-C. A sandwich substrate for ultrasensitive and label-free SERS spectroscopic detection of folic acid / methotrexate. *Biomed Microdevices*. oct 2014;16(5):673-9.
130. Fornasaro S, Marta SD, Rabusin M, Bonifacio A, Sergo V. Toward SERS-based point-of-care approaches for therapeutic drug monitoring: the case of methotrexate. *Faraday Discuss*. 23 2016;187:485-99.
131. Panikar SS, Ramírez-García G, Sidhik S, Lopez-Luke T, Rodriguez-Gonzalez C, Ciapara IH, et al. Ultrasensitive SERS Substrate for Label-Free Therapeutic-Drug Monitoring of Paclitaxel and Cyclophosphamide in Blood Serum. *Anal Chem*. 05 2019;91(3):2100-11.
132. Spalding K, Bonnier F, Bruno C, Blasco H, Board R, Benz-de Bretagne I, et al. Enabling quantification of protein concentration in human serum biopsies using attenuated total reflectance – Fourier transform infrared (ATR-FTIR) spectroscopy. *Vibrational Spectroscopy*. nov 2018;99:50-8.
133. Perez-Guaita D, Ventura-Gayete J, Pérez-Rambla C, Sancho-Andreu M, Garrigues S, de la Guardia M. Protein determination in serum and whole blood by attenuated total reflectance infrared spectroscopy. *Anal Bioanal Chem*. août 2012;404(3):649-56.
134. Le Bas JF, Estève F, Grand S, Rubin C, Rémy C, Benabid AL, et al. [NMR spectroscopy and brain diseases. Clinical applications]. *J Neuroradiol*. mars 1998;25(1):55-69.
135. Gross W, Snell RS. The serum gamma-globulin-level in malignant disease. *Nature*. 20 oct 1956;178(4538):855.
136. Cameron JM, Bruno C, Parachalil DR, Baker MJ, Bonnier F, Butler HJ, et al. Vibrational spectroscopic analysis and quantification of proteins in human blood

- plasma and serum. *Vibrational Spectroscopy in Protein Research: From Purified Proteins to Aggregates and Assemblies*. 2020;269-314.
137. Rieke GH. History of infrared telescopes and astronomy. *Experimental Astronomy*. 2009;25(1-3):125-41.
 138. Smith BC. *Fundamentals of Fourier transform infrared spectroscopy* [Internet]. Boca Raton, Fla.: CRC Press; 2011 [cité 31 janv 2018]. Disponible sur: <http://www.crcnetbase.com/isbn/9781420069303>
 139. *Applied Infrared Spectroscopy: Fundamentals Techniques and Analytical Problem-Solving* [Internet]. Wiley.com. [cité 20 mars 2019]. Disponible sur: <https://www.wiley.com/en-fr/Applied+Infrared+Spectroscopy%3A+Fundamentals+Techniques+and+Analytical+Problem+Solving-p-9780471043782>
 140. Messerschmidt R, Harthcock MA. *Infrared Microspectroscopy: Theory and Applications*. 1988.
 141. Vo-Dinh T, Gauglitz G, éditeurs. *Handbook of spectroscopy*. Weinheim; [Cambridge]: Wiley-VCH; 2003. 2 p.
 142. NIR Versus Mid-IR: How to Choose | Spectroscopy [Internet]. [cité 20 mars 2019]. Disponible sur: <http://www.spectroscopyonline.com/nir-versus-mid-ir-how-choose?id=&sk=&date=&%0A%09%09%09&pageID=2>
 143. Glassford SE, Byrne B, Kazarian SG. Recent applications of ATR FTIR spectroscopy and imaging to proteins. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. déc 2013;1834(12):2849-58.
 144. Lovergne L, Clemens G, Untereiner V, Lukaszewski RA, Sockalingum GD, Baker MJ. Investigating optimum sample preparation for infrared spectroscopic serum diagnostics. *Anal Methods*. 2015;7(17):7140-9.
 145. Gerrard DL, Birnie J. Raman spectroscopy. *Anal Chem*. juin 1990;62(12):140R-50R.

146. Raman CV, Krishnan KS. A New Type of Secondary Radiation. *Nature*. mars 1928;121(3048):501-2.
147. Adjouri C, Elliasmine A, Le Duff Y. *Spectroscopy*. 1996;44(16).
148. Das RS, Agrawal YK. Raman spectroscopy: Recent advancements, techniques and applications. *Vibrational Spectroscopy*. 1 nov 2011;57(2):163-76.
149. Blasco H, Patin F, Madji Hounoum B, Gordon PH, Vourc'h P, Andres CR, et al. Metabolomics in amyotrophic lateral sclerosis: how far can it take us? *Eur J Neurol*. mars 2016;23(3):447-54.
150. Bame M, Grier RE, Needleman R, Brusilow WSA. Amino acids as biomarkers in the SOD1(G93A) mouse model of ALS. *Biochim Biophys Acta*. janv 2014;1842(1):79-87.
151. Manzo E, Lorenzini I, Barrameda D, O'Conner AG, Barrows JM, Starr A, et al. Glycolysis upregulation is neuroprotective as a compensatory mechanism in ALS. *Elife*. 10 2019;8.
152. Loeffler J-P, Picchiarelli G, Dupuis L, Gonzalez De Aguilar J-L. The Role of Skeletal Muscle in Amyotrophic Lateral Sclerosis. *Brain Pathol*. mars 2016;26(2):227-36.
153. Sylow L, Kleinert M, Richter EA, Jensen TE. Exercise-stimulated glucose uptake - regulation and implications for glycaemic control. *Nature Reviews Endocrinology*. 2017;13(3):133-48.
154. Eaton S. Control of mitochondrial beta-oxidation flux. *Progress in Lipid Research*. mai 2002;41(3):197-239.
155. Béland-Millar A, Takimoto M, Hamada T, Messier C. Brain and muscle adaptation to high-fat diets and exercise: Metabolic transporters, enzymes and substrates in the rat cortex and muscle. *Brain Res*. 15 sept 2020;1749:147126.

156. Blasco H, Corcia P, Moreau C, Veau S, Fournier C, Vourc'h P, et al. 1H-NMR-based metabolomic profiling of CSF in early amyotrophic lateral sclerosis. *PLoS One*. 8 oct 2010;5(10):e13223.
157. Wuolikainen A, Andersen PM, Moritz T, Marklund SL, Antti H. ALS patients with mutations in the SOD1 gene have an unique metabolomic profile in the cerebrospinal fluid compared with ALS patients without mutations. *Mol Genet Metab*. mars 2012;105(3):472-8.
158. Kumar A, Bala L, Kalita J, Misra UK, Singh RL, Khetrpal CL, et al. Metabolomic analysis of serum by (1) H NMR spectroscopy in amyotrophic lateral sclerosis. *Clin Chim Acta*. 2 avr 2010;411(7-8):563-7.
159. Rothstein JD, Tsai G, Kuncel RW, Clawson L, Cornblath DR, Drachman DB, et al. Abnormal excitatory amino acid metabolism in amyotrophic lateral sclerosis. *Ann Neurol*. juill 1990;28(1):18-25.
160. Iwasaki Y, Ikeda K, Kinoshita M. Plasma amino acid levels in patients with amyotrophic lateral sclerosis. *J Neurol Sci*. févr 1992;107(2):219-22.
161. Wuolikainen A, Jonsson P, Ahnlund M, Antti H, Marklund SL, Moritz T, et al. Multi-platform mass spectrometry analysis of the CSF and plasma metabolomes of rigorously matched amyotrophic lateral sclerosis, Parkinson's disease and control subjects. *Mol Biosyst*. avr 2016;12(4):1287-98.
162. Gray E, Larkin JR, Claridge TDW, Talbot K, Sibson NR, Turner MR. The longitudinal cerebrospinal fluid metabolomic profile of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener*. 2015;16(7-8):456-63.
163. Abdel-Khalik J, Yutuc E, Crick PJ, Gustafsson J-Å, Warner M, Roman G, et al. Defective cholesterol metabolism in amyotrophic lateral sclerosis. *J Lipid Res*. 2017;58(1):267-78.
164. Dupuis L, Corcia P, Fergani A, Gonzalez De Aguilar J-L, Bonnefont-Rousselot D, Bittar R, et al. Dyslipidemia is a protective factor in amyotrophic lateral sclerosis. *Neurology*. 25 mars 2008;70(13):1004-9.

165. Valentino F, Bivona G, Butera D, Paladino P, Fazzari M, Piccoli T, et al. Elevated cerebrospinal fluid and plasma homocysteine levels in ALS. *Eur J Neurol.* janv 2010;17(1):84-9.
166. Levin J, Bötzel K, Giese A, Vogeser M, Lorenzl S. Elevated levels of methylmalonate and homocysteine in Parkinson's disease, progressive supranuclear palsy and amyotrophic lateral sclerosis. *Dement Geriatr Cogn Disord.* 2010;29(6):553-9.
167. Zoccolella S, Simone IL, Lamberti P, Samarelli V, Tortelli R, Serlenga L, et al. Elevated plasma homocysteine levels in patients with amyotrophic lateral sclerosis. *Neurology.* 15 janv 2008;70(3):222-5.
168. Cieslarova Z, Lopes FS, do Lago CL, França MC, Colnaghi Simionato AV. Capillary electrophoresis tandem mass spectrometry determination of glutamic acid and homocysteine's metabolites: Potential biomarkers of amyotrophic lateral sclerosis. *Talanta.* 1 août 2017;170:63-8.
169. Ribbenstedt A, Ziarrusta H, Benskin JP. Development, characterization and comparisons of targeted and non-targeted metabolomics methods. *PLoS ONE.* 2018;13(11):e0207082.
170. Mandal R, Guo AC, Chaudhary KK, Liu P, Yallou FS, Dong E, et al. Multi-platform characterization of the human cerebrospinal fluid metabolome: a comprehensive and quantitative update. *Genome Med.* 30 avr 2012;4(4):38.
171. Hamerly T, Tripet BP, Tigges M, Giannone RJ, Wurch L, Hettich RL, et al. Untargeted metabolomics studies employing NMR and LC-MS reveal metabolic coupling between *Nanoarchaeum equitans* and its archaeal host *Ignicoccus hospitalis*. *Metabolomics.* 1 août 2015;11(4):895-907.
172. Wilkins J, Sakrikar D, Petterson X-M, Lanza IR, Trushina E. A comprehensive protocol for multiplatform metabolomics analysis in patient-derived skin fibroblasts. *Metabolomics.* 23 2019;15(6):83.
173. Trostchansky A. Overview of Lipid Biomarkers in Amyotrophic Lateral Sclerosis (ALS). *Adv Exp Med Biol.* 2019;1161:233-41.

174. Ikeda K, Hirayama T, Takazawa T, Kawabe K, Iwasaki Y. Relationships between disease progression and serum levels of lipid, urate, creatinine and ferritin in Japanese patients with amyotrophic lateral sclerosis: a cross-sectional study. *Intern Med.* 2012;51(12):1501-8.
175. Chiò A, Calvo A, Ilardi A, Cavallo E, Moglia C, Mutani R, et al. Lower serum lipid levels are related to respiratory impairment in patients with ALS. *Neurology.* 17 nov 2009;73(20):1681-5.
176. Yang JW, Kim S-M, Kim H-J, Kim J-E, Park KS, Kim S-H, et al. Hypolipidemia in patients with amyotrophic lateral sclerosis: a possible gender difference? *J Clin Neurol.* avr 2013;9(2):125-9.
177. Kim S-M, Kim H, Kim J-E, Park KS, Sung J-J, Kim SH, et al. Amyotrophic lateral sclerosis is associated with hypolipidemia at the presymptomatic stage in mice. *PLoS One.* 25 mars 2011;6(3):e17985.
178. Yoshii Y, Hadano S, Otomo A, Kawabe K, Ikeda K, Iwasaki Y. Lower serum lipid levels are related to respiratory impairment in patients with ALS. *Neurology.* 15 juin 2010;74(24):2027; author reply 2027-2028.
179. Murai A, Miyahara T, Tanaka T, Kaneko T, Sako Y, Kameyama M. Abnormalities of lipoprotein and carbohydrate metabolism in degenerative diseases of the nervous system--motor neuron disease and spinocerebellar degeneration. *Tohoku J Exp Med.* avr 1983;139(4):365-76.
180. Dedic SIK, Stevic Z, Dedic V, Stojanovic VR, Milicev M, Lavrnic D. Is hyperlipidemia correlated with longer survival in patients with amyotrophic lateral sclerosis? *Neurol Res.* juill 2012;34(6):576-80.
181. Paganoni S, Deng J, Jaffa M, Cudkowicz ME, Wills A-M. Body mass index, not dyslipidemia, is an independent predictor of survival in amyotrophic lateral sclerosis. *Muscle Nerve.* juill 2011;44(1):20-4.
182. Huang R, Guo X, Chen X, Zheng Z, Wei Q, Cao B, et al. The serum lipid profiles of amyotrophic lateral sclerosis patients: A study from south-west China

- and a meta-analysis. *Amyotroph Lateral Scler Frontotemporal Degener.* 2015;16(5-6):359-65.
183. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, et al. The human serum metabolome. *PLoS ONE.* 16 févr 2011;6(2):e16957.
 184. Xue Y, Wang X, Zhao Y-Y, Ma X-T, Ji X-K, Sang S-W, et al. Metabolomics and Lipidomics Profiling in Asymptomatic Severe Intracranial Arterial Stenosis: Results from a Population-Based Study. *Journal of Proteome Research.* 05 2020;19(6):2206-16.
 185. Urban M, Enot DP, Dallmann G, Körner L, Forcher V, Enoh P, et al. Complexity and pitfalls of mass spectrometry-based targeted metabolomics in brain research. *Analytical Biochemistry.* 15 nov 2010;406(2):124-31.
 186. Movasaghi Z, Rehman S, ur Rehman Drl. Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. *Applied Spectroscopy Reviews.* févr 2008;43(2):134-79.
 187. Backhaus J, Mueller R, Formanski N, Szlama N, Meerpohl H-G, Eidt M, et al. Diagnosis of breast cancer with infrared spectroscopy from serum samples. *Vibrational Spectroscopy.* mars 2010;52(2):173-7.
 188. Paraskevaïdi M, Ashton KM, Stringfellow HF, Wood NJ, Keating PJ, Rowbottom AW, et al. Raman spectroscopic techniques to detect ovarian cancer biomarkers in blood plasma. *Talanta.* nov 2018;189:281-8.
 189. Schmitt J, Beekes M, Brauer A, Udelhoven T, Lasch P, Naumann D. Identification of scrapie infection from blood serum by Fourier transform infrared spectroscopy. *Anal Chem.* 1 août 2002;74(15):3865-8.
 190. Staib A, Dolenko B, Fink DJ, Früh J, Nikulin AE, Otto M, et al. Disease pattern recognition testing for rheumatoid arthritis using infrared spectra of human serum. *Clin Chim Acta.* juin 2001;308(1-2):79-89.

191. Petrov MS, Gordetzov AS, Kukosh MV. Early prediction of severity in acute pancreatitis using infrared spectroscopy of serum. *Pancreatology*. 2007;7(5-6):451-8.
192. Byrne HJ, Baranska M, Puppels GJ, Stone N, Wood B, Gough KM, et al. Spectropathology for the next generation: Quo vadis? *The Analyst*. 2015;140(7):2066-73.
193. Dai S, Lin Z, Xu B, Wang Y, Shi X, Qiao Y, et al. Metabolomics data fusion between near infrared spectroscopy and high-resolution mass spectrometry: A synergetic approach to boost performance or induce confusion. *Talanta*. 1 nov 2018;189:641-8.
194. Hands JR. Investigating the use of Attenuated Total Reflection Fourier-Transform Infrared (ATR-FTIR) Spectroscopy for the Rapid Diagnosis of Brain Tumours using Human Blood Serum [Doctor of Philosophy]. [Glasgow]: University of Strathclyde; 2015.
195. Cameron JM, Butler HJ, Palmer DS, Baker MJ. Biofluid spectroscopic disease diagnostics: A review on the processes and spectral impact of drying. *Journal of Biophotonics*. 5 mars 2018;e201700299.
196. Bonnier F, Brachet G, Duong R, Sojinrin T, Respaud R, Aubrey N, et al. Screening the low molecular weight fraction of human serum using ATR-IR spectroscopy. *J Biophotonics*. 2016;9(10):1085-97.
197. Vassault A, Grafmeyer D, De Graeve J, Cohen R, Beaudonnet A, Bienvenu J. John Libbey Eurotext - Annales de Biologie Clinique - Analyses de biologie médicale : spécifications et normes d'acceptabilité à l'usage de la validation de techniques [Internet]. [cité 9 juill 2020]. Disponible sur: https://www.jle.com/fr/revues/abc/e-docs/analyses_de_biologie_medicale_specifications_et_normes_dacceptabilite_a_lusage_de_la_validation_de_techniques_50188/article.phtml
198. Bonnier F, Petitjean F, Baker MJ, Byrne HJ. Improved protocols for vibrational spectroscopic analysis of body fluids: Improved protocols for vibrational

- spectroscopic analysis of body fluids. *Journal of Biophotonics*. avr 2014;7(3-4):167-79.
199. Bonnier F, Blasco H, Wasselet C, Brachet G, Respaud R, Carvalho LFCS, et al. Ultra-filtration of human serum for improved quantitative analysis of low molecular weight biomarkers using ATR-IR spectroscopy. *Analyst*. 10 avr 2017;142(8):1285-98.
 200. Garidel P, Schott H. Fourier-Transform Midinfrared Spectroscopy for Analysis and Screening of Liquid Protein Formulations. 2006;6.
 201. Baker MJ, Gazi E, Brown MD, Shanks JH, Gardner P, Clarke NW. FTIR-based spectroscopic analysis in the identification of clinically aggressive prostate cancer. *British Journal of Cancer*. 2 déc 2008;99(11):1859-66.
 202. Bellisola G, Sorio C. Infrared spectroscopy and microscopy in cancer research and diagnosis. :21.
 203. Naumann D. FT-INFRARED AND FT-RAMAN SPECTROSCOPY IN BIOMEDICAL RESEARCH. *Applied Spectroscopy Reviews*. 30 juin 2001;36(2-3):239-98.
 204. Current databases on biological variation: pros, cons and progress: *Scandinavian Journal of Clinical and Laboratory Investigation*: Vol 59, No 7 [Internet]. 2020 [cité 12 févr 2020]. Disponible sur: <https://www.tandfonline.com/doi/abs/10.1080/00365519950185229>
 205. Barth A. Infrared spectroscopy of proteins. *Biochim Biophys Acta*. sept 2007;1767(9):1073-101.
 206. Kuligowski J, Pérez-Guaita D, Escobar J, Lliso I, de la Guardia M, Lendl B, et al. Infrared biospectroscopy for a fast qualitative evaluation of sample preparation in metabolomics. *Talanta*. 1 sept 2014;127:181-90.
 207. Ranganathan S, Williams E, Ganchev P, Gopalakrishnan V, Lacomis D, Urbinelli L, et al. Proteomic profiling of cerebrospinal fluid identifies biomarkers for amyotrophic lateral sclerosis. *J Neurochem*. déc 2005;95(5):1461-71.

208. Ryberg H, An J, Darko S, Lustgarten JL, Jaffa M, Gopalakrishnan V, et al. Discovery and verification of amyotrophic lateral sclerosis biomarkers by proteomics. *Muscle Nerve*. juill 2010;42(1):104-11.
209. Wilson ME, Boumaza I, Lacomis D, Bowser R. Cystatin C: a candidate biomarker for amyotrophic lateral sclerosis. *PLoS One*. 9 déc 2010;5(12):e15133.
210. Chen X, Chen Y, Wei Q, Ou R, Cao B, Zhao B, et al. Assessment of a multiple biomarker panel for diagnosis of amyotrophic lateral sclerosis. *BMC Neurol*. 15 sept 2016;16:173.
211. Pasinetti GM, Ungar LH, Lange DJ, Yemul S, Deng H, Yuan X, et al. Identification of potential CSF biomarkers in ALS. *Neurology*. 25 avr 2006;66(8):1218-22.
212. Lunetta C, Lizio A, Maestri E, Sansone VA, Mora G, Miller RG, et al. Serum C-Reactive Protein as a Prognostic Biomarker in Amyotrophic Lateral Sclerosis. *JAMA Neurol*. 01 2017;74(6):660-7.
213. Rossi D, Volanti P, Brambilla L, Colletti T, Spataro R, La Bella V. CSF neurofilament proteins as diagnostic and prognostic biomarkers for amyotrophic lateral sclerosis. *J Neurol*. mars 2018;265(3):510-21.
214. Gonçalves M, De Carvalho M, Peixoto C, Alves P, Barreto C, Oliva A, et al. Phosphoneurofilament heavy chain and vascular endothelial growth factor as cerebrospinal fluid biomarkers for ALS. *Amyotroph Lateral Scler Frontotemporal Degener*. 2017;18(1-2):134-6.
215. Boylan KB, Glass JD, Crook JE, Yang C, Thomas CS, Desaro P, et al. Phosphorylated neurofilament heavy subunit (pNF-H) in peripheral blood and CSF as a potential prognostic biomarker in amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry*. avr 2013;84(4):467-72.
216. Gendron TF, C9ORF72 Neurofilament Study Group, Daugherty LM, Heckman MG, Diehl NN, Wu J, et al. Phosphorylated neurofilament heavy chain: A

- biomarker of survival for C9ORF72-associated amyotrophic lateral sclerosis. *Ann Neurol.* juill 2017;82(1):139-46.
217. Gaiani A, Martinelli I, Bello L, Querin G, Puthenparampil M, Ruggero S, et al. Diagnostic and Prognostic Biomarkers in Amyotrophic Lateral Sclerosis: Neurofilament Light Chain Levels in Definite Subtypes of Disease. *JAMA Neurol.* 01 2017;74(5):525-32.
218. Thompson AG, Gray E, Thézénas M-L, Charles PD, Evetts S, Hu MT, et al. Cerebrospinal fluid macrophage biomarkers in amyotrophic lateral sclerosis. *Ann Neurol.* 2018;83(2):258-68.
219. Mitchell RM, Simmons Z, Beard JL, Stephens HE, Connor JR. Plasma biomarkers associated with ALS and their relationship to iron homeostasis. *Muscle Nerve.* juill 2010;42(1):95-103.
220. Simpson EP, Henry YK, Henkel JS, Smith RG, Appel SH. Increased lipid peroxidation in sera of ALS patients: a potential biomarker of disease burden. *Neurology.* 25 mai 2004;62(10):1758-65.
221. Nardo G, Pozzi S, Pignataro M, Lauranzano E, Spano G, Garbelli S, et al. Amyotrophic lateral sclerosis multiprotein biomarkers in peripheral blood mononuclear cells. *PLoS One.* 2011;6(10):e25545.
222. Filareti M, Luotti S, Pasetto L, Pignataro M, Paoletta K, Messina P, et al. Decreased Levels of Foldase and Chaperone Proteins Are Associated with an Early-Onset Amyotrophic Lateral Sclerosis. *Front Mol Neurosci.* 2017;10:99.
223. Analytical Techniques in Lipidomics: State of the Art: Critical Reviews in Analytical Chemistry: Vol 47, No 5 [Internet]. [cité 23 juill 2021]. Disponible sur: <https://www.tandfonline.com/doi/abs/10.1080/10408347.2017.1310613>
224. Beale DJ, Pinu FR, Kouremenos KA, Poojary MM, Narayana VK, Boughton BA, et al. Review of recent developments in GC-MS approaches to metabolomics-based research. *Metabolomics.* 17 nov 2018;14(11):152.

225. Travo A, Paya C, Délérís G, Colin J, Mortemousque B, Forfar I. Potential of FTIR spectroscopy for analysis of tears for diagnosis purposes. *Anal Bioanal Chem.* avr 2014;406(9-10):2367-76.
226. Yazdani M, Elgstøen KBP, Rootwelt H, Shahdadfar A, Utheim ØA, Utheim TP. Tear Metabolomics in Dry Eye Disease: A Review. *Int J Mol Sci.* 1 août 2019;20(15):3755.
227. Budde B, Schartner J, Tönges L, Kötting C, Nabers A, Gerwert K. Reversible Immuno-Infrared Sensor for the Detection of Alzheimer's Disease Related Biomarkers. *ACS Sens.* 26 juill 2019;4(7):1851-6.
228. Frost RL, Kristof J, Rintoul L, Klopogge JT. Raman spectroscopy of urea and urea-intercalated kaolinites at 77 K. *Spectrochim Acta A Mol Biomol Spectrosc.* août 2000;56A(9):1681-91.
229. Hackshaw KV, Aykas DP, Sigurdson GT, Plans M, Madiar F, Yu L, et al. Metabolic fingerprinting for diagnosis of fibromyalgia and other rheumatologic disorders. *J Biol Chem.* 15 févr 2019;294(7):2555-68.
230. Morasso CF, Sproviero D, Mimmi MC, Giannini M, Gagliardi S, Vanna R, et al. Raman spectroscopy reveals biochemical differences in plasma derived extracellular vesicles from sporadic Amyotrophic Lateral Sclerosis patients. *Nanomedicine.* oct 2020;29:102249.

Résumé en français

La Sclérose Latérale Amyotrophique (SLA) est une maladie neurodégénérative caractérisée par la mort des motoneurons. L'absence de biomarqueurs et de thérapies a pour conséquence un retard au diagnostic. Actuellement, la recherche de biomarqueur s'appuie principalement sur des approches globales comme l'approche métabolomique, cet outil permettant une vision complète du métabolisme. Cependant les techniques utilisées actuellement nécessitent de multiplier et combiner des analyses pour augmenter la couverture métabolique de l'analyse, sans pour autant parvenir à une couverture totale. Malgré un grand nombre d'étude réalisées, les analyses métabolomiques n'ont pas permis de valider un ou une combinaison de biomarqueurs. Des études récentes ont montré l'émergence de nouveaux outils tels que la spectroscopie infrarouge et la spectroscopie Raman avec des applications directes sur le plasma ou d'autres matrices biologiques. Ils permettent d'obtenir un spectre représentatif de la totalité des composants de l'échantillon sans nécessiter d'étape de préparation au préalable. Le développement méthodologique réalisé dans le cadre de cette thèse a mis en évidence le potentiel de la spectroscopie infrarouge ATR-IR et de la spectroscopie Raman dans la quantification de différents biomarqueurs plasmatiques. D'autre part, nous avons montré l'amélioration des performances analytiques dans l'analyse des métabolites par l'ultrafiltration. Pour finir, nous avons appliqué la méthodologie mise en place à des échantillons de patients témoins et de patients atteints de SLA et démontré la capacité de la spectroscopie infrarouge à discriminer les profils métaboliques des patients de ceux des témoins. En conclusion, les travaux réalisés ont permis de montrer l'utilisation potentielle de ces outils dans une approche globale pour améliorer la prise en charge du patient, cela dans un but diagnostique et/ou pronostique.

Mots-clés : spectroscopie infrarouge, spectroscopie Raman, métabolomique, SLA

Résumé en anglais

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease characterized by motoneuronal death. The lack of biomarkers and therapeutic solutions lead to diagnosis and prognosis delays. Actually, biomarker research is principally based on global approach as for example metabolomics approach. The objective of these tools is the obtention of a complete metabolism vision. However, conventional approach needs to multiply and combine different analytical methods. Recent studies showed infrared spectroscopy and Raman spectroscopy as emergent tools in biology. Spectra obtained by infrared and raman spectroscopy represents the whole sample composition without a sample pre-treatment before analysis. Firstable, methodology developed in this work highlighted Fourier transformed infrared spectroscopy and Raman spectroscopy capacity for some plasmatic biomarkers quantification. Furthermore, we showed that splitting sample in high molecular weight and low molecular weight fractions should improve analytical performances for metabolites. Finally, developed methodology was applied to control and ALS patients' sample in a metabolomics approach. To conclude, this work demonstrated the potential interest of these analyses as diagnosis and/or prognosis tools in global approach in health field.

Keywords: infrared spectroscopy, Raman spectroscopy, metabolomics, ALS