



UNIVERSITE DE REIMS CHAMPAGNE-ARDENNE  
ECOLE DOCTORALE SCIENCES TECHNOLOGIE SANTE

## **THESE**

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITE DE REIMS CHAMPAGNE-ARDENNE**

**Discipline : Biophysique**

Présentée et soutenue publiquement

Par

**Imane TALEB**

Le 18 Décembre 2013

---

**Apport de la spectroscopie vibrationnelle, infrarouge et Raman,  
appliquée au sérum pour le diagnostic du carcinome hépatocellulaire  
chez les patients atteints de cirrhose.**

---

### JURY

Président:	Pr. Michel MANFAIT (Reims)
Rapporteurs:	Pr. Erik GOORMAGHTIGH (Bruxelles, Belgique) Dr. François Le NAOUR (Villejuif)
Examineurs:	Dr. Patrick DUCOROY (Dijon)
Directeurs de thèse:	Pr. Gérard THIEFIN (Reims) Pr. Ganesh D SOCKALINGUM (Reims)

*« La vérité de demain se nourrit de l'erreur d'hier »*

*Antoine Saint-Exupéry*

## REMERCIEMENTS

Tout d'abord je tiens à remercier la Ville de Reims pour le financement de mon doctorat pendant ces trois années.

Un grand merci au Professeur Michel MANFAIT et au Professeur Olivier PIOT, directeurs successifs du laboratoire MÉDIAN, de m'avoir accueillie dans votre équipe pendant ces trois années de thèse et de m'avoir donné les moyens d'effectuer ma thèse dans les meilleures conditions. Je voudrais vous remercier également pour m'avoir donné l'occasion de présenter mes travaux dans des congrès internationaux.

Je tiens à remercier le Professeur Erik GOORMAGHTIGH et le Docteur François LENOIR d'avoir accepté de prendre connaissance de mon travail de thèse et d'en être les rapporteurs. Veuillez accepter mes plus vifs remerciements pour votre présence dans ce jury et soyez assurés de tout mon respect et de ma profonde gratitude.

Je tiens à remercier le Docteur Patrick DUCOROY d'avoir accepté de juger ce travail et d'en être l'examineur. Veuillez accepter l'expression de ma profonde reconnaissance.

J'aimerais exprimer toute ma reconnaissance à mes directeurs de thèse le Professeur Gérard THIEFIN et le Professeur Ganesh SOCKALINGUM, pour votre aide précieuse, vos conseils et votre soutien. Vous avez toujours veillé à suivre le bon déroulement de mon travail malgré vos nombreuses charges. Merci pour la confiance que vous m'avez accordé pendant ces trois années.

Un énorme merci au Docteur Cyril GOBINET pour son aide précieuse sur les traitements de données. Merci pour tout le temps que tu m'as consacré. PROMIS des petits Gâteaux bien sucrés et bien gras tu en auras ;) .

Je tiens à remercier le Docteur Roselyne Garnotel pour tous les conseils, la gentillesse, et la disponibilité.

Je tiens à remercier Monsieur Thomas FERON pour son aide lors de la collecte des échantillons des différents centres de recrutement. Merci pour ta disponibilité et ta bonne humeur.

Merci à Vincent Gaydou pour son aide sur le traitement PLS-DA et à Teddy Happillon et Michael Fere pour leurs interventions à l'apparition des fameuses « lignes rouges » sur Matlab.

Une mention spéciale au Professeur Hamid MORJANI. Il est difficile de trouver des qualificatifs assez forts pour souligner tes conseils, ta gentillesse, et ta bonne humeur quotidienne.

Merci à l'Unité MEDyC pour la bonne ambiance qui régnait, c'était un élément crucial pour le bon déroulement de ce travail.

Durant ces trois années de thèse, j'ai eu la chance d'être entourer de nombreuses personnes attachantes : que toutes soient remerciées pour les bons moments partagés. Je pense notamment à Mahdi, Sana, Que, Hanane, Saad, Ali, Mohammed, Jennifer et Mme Cousinat. Merci pour votre aide, votre soutien, votre écoute et vos encouragements.

Enfin, les mots les plus simples étant les plus forts, j'adresse toute mon affection à ma famille (mima, papa, mama, Hind, Meryem, Bissouu et Mohssine). Malgré mon éloignement depuis de nombreuses années, votre confiance, votre tendresse et votre amour me portent et me guident tous les jours. Merci pour avoir fait de moi ce que je suis aujourd'hui. Est-ce un bon endroit pour dire ce genre de choses? Je n'en connais en tout cas pas de mauvais. Je vous aime.

Une pensée pour terminer ces remerciements pour toi papi et toi petite Asmaa. Vous n'avez pas vu l'aboutissement de mon travail, mais je sais que vous en auriez été très fiers!!! Paix a vos âmes.

À tous ceux que je ne nomme pas, mais qui se reconnaîtront

Merci...

# RÉSUMÉ

## **Apport de la spectroscopie vibrationnelle, infrarouge et Raman, appliquée au sérum pour le diagnostic de carcinome hépatocellulaire chez les patients atteints de cirrhose.**

Le carcinome hépatocellulaire (CHC) est la 3ème cause de mortalité par cancer dans le monde et son pronostic est sombre en raison d'un diagnostic souvent tardif. Cela explique la nécessité d'identifier de nouveaux marqueurs sériques pour obtenir un diagnostic précoce et améliorer le pronostic.

Dans ce travail de thèse, nous avons évalué l'intérêt de la spectroscopie vibrationnelle appliquée au sérum pour le diagnostic du CHC. La spectroscopie vibrationnelle, Raman ou infrarouge, permet de détecter les caractéristiques spectrales liées aux différents constituants moléculaires présents dans le matériel étudié, de façon qualitative et quantitative. Notre hypothèse de travail était que le sérum de patients cirrhotiques atteints de CHC présente une signature spectrale, Raman ou infrarouge, permettant de le différencier du sérum de patients cirrhotiques n'ayant pas de CHC. Ce travail a utilisé la sérothèque constituée dans le cadre de l'étude cas-témoins multicentrique CiRCE (étude des facteurs de risque environnementaux, nutritionnels et métaboliques de carcinome hépatocellulaire chez les malades atteints de cirrhose) réunissant six centres hospitaliers universitaires de la région grand-est de la France (Reims, Dijon, Besançon, Nancy, Strasbourg et Metz).

Dans un premier temps, nous avons réalisé deux études pilotes pour évaluer la faisabilité et le potentiel de ces deux approches biophotoniques pour le diagnostic sérique du CHC chez des patients cirrhotiques. L'intérêt de la microspectroscopie Raman appliquée au sérum a été évalué chez 37 patients cirrhotiques atteints de CHC et 34 patients cirrhotiques sans CHC. L'analyse des données spectrales a été effectuée par une méthode de classification supervisée, les séparateurs à vaste marge (SVM), associée à une technique de validation croisée par "leave one out" en raison du nombre limité de données. Cette analyse a permis de classer les deux groupes de patients avec un taux global d'exactitude diagnostique de 85 à 91%. Nous avons également démontré dans une deuxième étude pilote le potentiel diagnostique de la spectroscopie IRTF dans une population de 40 patients avec CHC et 39 patients sans CHC. Dans cette étude, le taux global d'exactitude diagnostique obtenu avec la même méthode de classification était de 82 à 86 %.

Dans un deuxième temps, nous avons mené une étude sur une large cohorte de patients afin de valider les résultats prometteurs obtenus dans l'étude pilote IRTF. Les caractéristiques spectrales IRTF du sérum de 308 patients cirrhotiques avec CHC ont été comparées à celles du sérum de 509 patients cirrhotiques sans CHC. Après prétraitement par EMSC, l'analyse des données spectrales a été réalisée en utilisant deux méthodes de classification supervisée: la méthode SVM et l'analyse discriminante PLS (partial least squares). Le modèle de classification a été établi sur un groupe d'entraînement et de validation interne. La performance diagnostique a ensuite été testée sur une groupe de validation externe. L'analyse par ces deux méthodes de l'ensemble du spectre 900-4000  $\text{cm}^{-1}$ , de l'empreinte digitale 900-1800  $\text{cm}^{-1}$ , avec et sans dérivation des données spectrales, n'a pas permis de confirmer sur une large population de patients les résultats obtenus dans l'étude pilote. Avec un taux global de bonne classification entre 46 et 61 %, l'analyse spectrale du sérum entier n'apparaît pas discriminante pour distinguer les patients cirrhotiques avec et sans CHC. Ce résultat souligne la nécessité de confirmer sur un grand nombre de patients les résultats obtenus dans des études pilote limitées à de petits effectifs. L'analyse spectrale ciblée sur des fractions du sérum plutôt que sur le sérum entier pourrait permettre d'identifier plus efficacement des marqueurs diagnostiques en évitant la superposition des informations spectrales liées aux multiples molécules présentes dans le sérum.

**Mots clés:** Spectroscopie vibrationnelle, Carcinome hépatocellulaire, Sérum, Raman et IRTF, Chimiométrie.

# ABSTRACT

## **Application of infrared and Raman vibrational spectroscopy to serum analysis for the diagnosis of hepatocellular carcinoma in cirrhotic patients**

Hepatocellular carcinoma (HCC) is the third leading cause of cancer death in the world. Its prognosis is poor due to late diagnosis. This explains the need to identify new serum markers for early diagnosis and improve prognosis.

In this study, we have evaluated the potential of vibrational spectroscopy applied to serum for the diagnosis of HCC. Vibrational spectroscopy, Raman and infrared, allows detecting spectral characteristics associated with different molecular components present in the biological sample, both qualitatively and quantitatively. Our working hypothesis was that the serum of cirrhotic patients with HCC has a different spectral signature from that obtained from the serum of cirrhotic patients without HCC, which can then allow its differentiation. This work has used the serum bank established as part of the case-control CiRCE study (the study of environmental, nutritional and metabolic risk factors of hepatocellular carcinoma in patients with cirrhosis). Sera were recruited from six university hospitals in the East region of France (Reims, Dijon, Besançon, Nancy, Strasbourg and Metz).

In the first step, two pilot studies were conducted to evaluate the feasibility and the potential of these two biophotonic approaches for the serologic diagnosis of HCC in cirrhotic patients. Raman microspectroscopy applied to serum was tested in 37 cirrhotic patients with HCC and 34 cirrhotic patients without HCC. Analysis of spectral data was performed by support vector machine (SVM) a supervised method of classification. This method was combined with a technique of cross validation "leave-one-out" to overcome the risk of overfitting due to the small number of patients. This analysis allowed discriminating between the two groups of patients with a diagnostic accuracy rate of 85% to 91%. We also demonstrated in a second pilot study the diagnostic potential of FTIR spectroscopy in a population of 40 patients with HCC and 39 patients without HCC. In this study, the diagnostic performance accuracy obtained with the same method of classification was 82% to 86%.

In a second step, an FTIR study on a larger number of patients was performed to validate the promising results obtained in the pilot study. FTIR spectral characteristics of 308 sera from cirrhotic patients with HCC were compared with those of 509 cirrhotic patients without HCC.

After pre-processing with EMSC, the spectral data analysis was performed using two supervised classification methods: SVM and PLS-DA (Partial Least Squares Discriminant Analysis). The classification model was established on training and internal validation sets. The model was optimized and classification was then tested on an external validation set. The analysis using these two methods on the spectral range 900-4000  $\text{cm}^{-1}$  and 900-1800  $\text{cm}^{-1}$ , with and without derivation of the spectral data did not confirm the result obtained in the FTIR pilot study. The diagnostic accuracy was between 46% and 61%. FTIR spectral analysis of the whole serum does not appear discriminant enough to differentiate cirrhotic patients with and without HCC. This result highlights the need to confirm on a large number of patients the results obtained in pilot studies. Spectral analysis of serum fractions rather than whole serum could be an alternative to more effectively identify diagnostic markers, avoiding overlapping spectral information related to the complex composition of whole serum.

**Keywords:** Vibrational Spectroscopy, Hepatocellular carcinoma, Serum, Raman and FTIR, Chemometrics.

# TABLE DES MATIÈRES

Remerciements .....	1
Résumé .....	3
Abstract .....	5
Liste des abréviations .....	10
Liste des figures .....	12
Liste des tableaux .....	15
Avant-propos .....	17
Chapitre 1 : Introduction générale.....	19
I. Le carcinome hépatocellulaire (CHC) .....	20
1. Généralités.....	20
2. L'épidémiologie du CHC .....	21
3. Les facteurs étiologiques du CHC.....	22
4. La survenue et la progression du CHC.....	23
5. Les marqueurs biologiques du CHC .....	24
5.1. Alpha-foetoprotéine (AFP) .....	24
5.2. Des-gamma-carboxyprothrombine (DCP).....	24
5.3. Glypicane-3 .....	25
5.4. Alpha-1-fucosidase .....	25
6. Les symptômes liés au CHC .....	25
7. Le diagnostic du CHC .....	26
8. Le bilan d'extension extrahépatique du CHC .....	26
9. Le dépistage du CHC .....	27
10. La prévention du CHC .....	27
11. Le traitement du CHC .....	27
II. La spectroscopie vibrationnelle .....	30
1. Introduction .....	30

2.	Les modes de vibrations moléculaires .....	30
2.1.	Élongation .....	31
2.2.	Déformations dans et hors du plan.....	31
3.	La spectroscopie infrarouge à transformée de Fourier (IRTF) .....	33
3.1.	Principe de la spectroscopie IR.....	33
3.2.	Spectromètre IRTF.....	34
3.3.	Génération d'un spectre IRTF .....	37
3.4.	Résolution spectrale .....	39
3.5.	Loi de Beer-Lambert.....	40
3.6.	Avantages de la spectroscopie IRTF.....	40
3.7.	Domaines d'application de la spectroscopie IRTF .....	41
3.8.	Principales applications de la spectroscopie IRTF en cancérologie .....	41
4.	La microspectroscopie Raman .....	42
4.1.	Introduction.....	42
4.2.	Principe de la diffusion Raman.....	43
4.3.	Le microspectromètre Raman .....	45
4.4.	Avantages de la spectroscopie Raman .....	47
4.5.	Domaines d'applications de la spectroscopie Raman.....	47
4.6.	Principales applications de la spectroscopie Raman en cancérologie .....	47
5.	Analyses des biofluides par spectroscopie vibrationnelle.....	48
5.1.	Spectroscopie infrarouge des biofluides .....	49
5.2.	Spectroscopie Raman des biofluides .....	50
III.	Traitement des données spectrales .....	50
1.	Les prétraitements spectraux .....	50
2.	La chimiométrie .....	53
2.1.	Analyse hiérarchique en cluster (HCA).....	55
2.2.	Analyse en composante principale (ACP) .....	55
2.3.	Séparateurs à vaste marges (SVM).....	58
2.4.	Validation croisée (Leave-one-out) .....	62
2.5.	Régression aux moindres carrés partiels (PLS) .....	63
	Chapitre 2 : Matériels et méthodes.....	66
I.	Recrutement des patients .....	67

II. Microspectroscopie Raman appliquée au sérum.....	68
1. Préparation des sérums.....	68
1.1. Sérum en goutte séchée.....	69
1.2. Sérum lyophilisé .....	69
2. Paramètres d’acquisitions.....	70
3. Prétraitement des spectres Raman.....	70
3.1. Le test de qualité .....	70
3.2. Les prétraitements des spectres Raman .....	71
4. Analyses statistiques .....	73
III. Spectroscopie IRTF à haut débit appliquée au sérum .....	73
1. Préparation des sérums.....	73
2. Paramètres d’acquisitions.....	75
3. Prétraitement des spectres infrarouge .....	76
3.1. Le test de qualité .....	76
3.2. Le prétraitement des données spectrales.....	78
4. Analyses statistiques .....	78
Chapitre 3 : Résultats et discussion.....	80
Partie 3.1 : Etude pilote Raman.....	81
Partie 3.2 : Etude pilote IRTF .....	100
Partie 3.3 : Validation de l’approche IRTF dans une cohorte de 817 patients.....	109
Conclusions et perspectives .....	141
Références bibliographiques .....	145
Annexes.....	155
Publications et communications.....	175

## LISTE DES ABRÉVIATIONS

<b>A</b>	Absorbance
<b>ACP</b>	Analyse en composante principale
<b>AFP</b>	Alpha-foetoprotéine
<b>ANN</b>	Réseau de neurones artificiels
<b>ARN</b>	Acide Ribonucléique
<b>ATR</b>	Réflexion totale atténuée
<b>BCLC</b>	Classification de Barcelona Clinic Liver Cancer
<b>CaF<sub>2</sub></b>	Fluorure de calcium
<b>CCD</b>	Charged Coupled Device
<b>CHC</b>	Carcinome hépatocellulaire
<b>CLT</b>	Cadaveric liver transplantation
<b>CP</b>	Composante principale
<b>DCP</b>	Des-gamma-carboxyprothrombine
<b>EMSC</b>	Extended Multiplicative Signal Correction
<b>GGT</b>	Gamma-glutamyltransferase
<b>GPC3</b>	Glypicane-3
<b>HCA</b>	Hierarchical Cluster Analysis (Analyse hiérarchique ascendante)
<b>I</b>	Intensité
<b>IR</b>	Infrarouge
<b>IRM</b>	Imagerie par résonance magnétique
<b>IRTF</b>	Infrarouge à Transformée de Fourier
<b>LDA</b>	Analyse Discriminante Lineaire
<b>LDLT</b>	Living Donor Liver Transplantation
<b>LOOCV</b>	Leave-one-out cross-validation
<b>NASH</b>	Non alcoholic steato-hepatitis
<b>NIR</b>	Near-Infrared
<b>PDGF</b>	Platelet-derived growth factor 2
<b>PEI</b>	Percutaneous Ethanol Injection

<b>Phe</b>	Phénylalanine
<b>PLS-DA</b>	Partial Least Squares Discriminant Analysis
<b>PST</b>	Performance status test
<b>RF</b>	Radiofréquence
<b>RMN</b>	Résonance magnétique nucléaire
<b>S/B</b>	Signal/Bruit
<b>SERS</b>	Surface-Enhanced Raman spectroscopy
<b>Si</b>	Silicium
<b>SVM</b>	Séparateurs à vastes marges
<b>TACE</b>	Transarterial embolization
<b>Trp</b>	Tryptophane
<b>Tyr</b>	Tyrosine
<b>VEGF</b>	Vascular endothelial growth factor

## LISTE DES FIGURES

Figure 1 : Anatomie du foie et des organes adjacents.....	20
Figure 2 : Incidence et mortalité du cancer du foie en France. ....	21
Figure 3 : Classification de Barcelona Clinic Liver Cancer (BCLC) avec algorithme thérapeutique. ....	28
Figure 4 : Les modes de vibrations moléculaires des molécules H <sub>2</sub> O et CH <sub>2</sub> . ....	32
Figure 5 : Le spectre électromagnétique. ....	34
Figure 6 : Schéma représentant les principaux éléments d'un spectromètre IRTF.....	35
Figure 7 : Schéma d'un interféromètre de Michelson.....	36
Figure 8 : Schéma simplifié des modes d'acquisition spectrales en spectroscopie IR. ....	37
Figure 9 : Schéma montrant les étapes de l'obtention d'un spectre IRTF. ....	38
Figure 10 : Distance « d » entre deux signaux spectraux et l'interférogramme correspondant. ....	40
Figure 11 : Diagramme de Jablonski illustrant les transitions impliquées dans les phénomènes d'absorption, de diffusion et d'émission. ....	43
Figure 12 : Spectre de transition du niveau fondamental aux états virtuels.....	44
Figure 13 : Schéma d'un microspectromètre Raman.....	46
Figure 14: Dendrogramme représentant les distances euclidiennes entre les objets.....	55
Figure 15 : Analyse en composante principale (ACP).....	57
Figure 16 : Exemple de représentation graphique des échantillons A, B et C en prenant 2 composantes principales. V1 et V2 représentent les % de variance de chaque CP. ....	58
Figure 17: Schéma illustrant le principe de la SVM pour des données linéairement séparables. ....	59
Figure 18 : Schéma illustrant le principe de la SVM pour des exemples qui ne sont pas linéairement séparables. ....	61
Figure 19 : Schéma représentant les étapes de la construction du modèle PLS-DA. ....	64

Figure 20 : Classification hiérarchique du sérum d'un patient montrant l'effet de la congélation-décongélation sur les spectres IRTF.....	68
Figure 21 : Image visible d'une goutte séchée de sérum d'un patient montrant les zones d'acquisition : centre et périphérie .....	69
Figure 22 : Micro-spectromètre Raman LabRam 785 nm. ....	70
Figure 23 : Exemple de l'application du test qualité à spectre Raman de sérum.....	71
Figure 24 : Exemple des acquisitions enregistrées de la calibration Raman. ....	72
Figure 25 : Spectres infrarouge d'un sérum pur (bleu) et dilué au tiers (rouge).....	74
Figure 26 : Sérums déposés sur une plaque silicium de 384 puits pour l'analyse IRTF à haut débit.....	75
Figure 27 : Le spectromètre IRTF Tensor 27 couplé au module HTS-XT à haut débit. ....	75
Figure 28 : Description des zones spectrales utilisées pour évaluer les critères du test de qualité appliqué aux spectres IRTF.....	77
Figure 29 : Exemple de spectres IRTF conforme et non conforme du sérum. ....	78
Figure 30: Schéma décrivant de la procedure expérimentale utilisée pour la preparation des échantillons et l'acquisition spectrale. ....	84
Figure 31 : Spectre Raman du sérum humain. ....	84
Figure 32 : Exemple de prétraitements appliqués sur un spectre Raman d'une goutte de sérum séché.....	86
Figure 33: Image optique du centre (A) et de la périphérie (B) de la goutte séchée du sérum.....	89
Figure 34: Comparaison des moyennes des spectres Raman des deux sites d'acquisitions : le centre (C) et la périphérie (P) de la goutte séchée de sérum. ....	90
Figure 35: Comparaison des moyennes des spectres Raman des patients avec CHC (CHC <sup>+</sup> ) et sans CHC (CHC <sup>-</sup> ). ....	91
Figure 36: Représentation 3D de l'ACP de l'ensemble des spectres Raman du sérum de patients cirrhotiques avec CHC (rouge) et sans CHC (bleu). ....	93
Figure 37: Analyse Raman du sérum lyophilisé. ....	97

Figure 38 : Schéma décrivant de la procédure expérimentale utilisée pour la préparation des échantillons et l'acquisition spectrale. ....	103
Figure 39 : Attribution de pics du spectre infrarouge du sérum humain.....	104
Figure 40 : Spectres médians IR prétraités par EMSC des groupes de patients cirrhotiques avec CHC (CHC <sup>+</sup> ) et sans CHC (CHC <sup>-</sup> ). ....	105
Figure 41 : Représentation graphique des 3 CPs les plus explicatives sur l'ensemble des spectres infrarouge du sérum de patients cirrhotiques avec (rouge) et sans (bleu) CHC.....	106
Figure 42 : Spectres moyens IRTF des sérums de patients avec et sans CHC. ....	116
Figure 43 : Analyse en composante principale des données IRTF des patients avec (rouge) et sans (bleu) CHC. ....	117
Figure 44 : Analyse en composante principale des données IRTF des patients avec (rouge) et sans (bleu) CHC. ....	123
Figure 45 : Spectres moyens IRTF des sérums de patients avec et sans CHC. ....	128
Figure 46 : Analyse en composante principale des données IRTF des patients avec (rouge) et sans (bleu) CHC. ....	130
Figure 47 : Analyse en composante principale des données IRTF des patients avec (rouge) et sans (bleu) CHC. ....	135

## LISTE DES TABLEAUX

Tableau 1 : Répartition géographique des principaux facteurs de risque de CHC (adapté d'après (3)) .....	22
Tableau 2: Attribution des pics Raman du sérum humain. ....	85
Tableau 3: Les caractéristiques démographiques et clinico-biologiques des patients cirrhotiques avec et sans CHC. ....	88
Tableau 4: Classification SVM-LOOCV des sérums en gouttes séchées des patients cirrhotiques avec et sans CHC. ....	95
Tableau 5: Classification SVM-LOOCV des sérums lyophilisés de patients cirrhotiques avec et sans CHC.....	98
Tableau 6: Caractéristiques démographiques et clinico-biologiques des patients cirrhotiques avec et sans CHC.....	101
Tableau 7 : Classification SVM-LOOCV des 79 patients cirrhotiques avec et sans CHC. Le modèle a été appliqué à l'ensemble des 790 spectres (données 1), aux 739 spectres après l'exclusion des spectres aberrants (données 2) et aux 79 spectres médians de tous les patients. ....	107
Tableau 8 : Nombre de patients par centre de recrutement de la région Grand Est de France. ....	111
Tableau 9 : Caractéristiques clinico-biologiques des patients cirrhotiques avec et sans CHC. ....	112
Tableau 10 : Répartition aléatoire des patients en 3 jeux de données utilisés pour l'analyse SVM et l'analyse PLS-DA.....	115
Tableau 11 : Classification SVM sur la médiane, la moyenne et tous les spectres des patients cirrhotiques avec et sans CHC. Analyse sur la région spectrale 900-4000 $\text{cm}^{-1}$ sans dérivation seconde. ....	119
Tableau 12 : Classification PLS-DA sur la médiane, la moyenne et l'ensemble des spectres des patients cirrhotiques avec et sans CHC. Analyse sur la région spectrale 900-4000 $\text{cm}^{-1}$ sans dérivation seconde. ....	121

Tableau 13 : Classification SVM sur la médiane, la moyenne et l'ensemble des spectres des patients cirrhotiques avec et sans CHC. Analyse avec une dérivation seconde sur la région spectrale 900-4000 $\text{cm}^{-1}$ .	125
Tableau 14 : Classification PLS-DA sur la médiane, la moyenne et l'ensemble des spectres des patients cirrhotiques avec et sans CHC. Analyse avec dérivation seconde sur la région spectrale 900-4000 $\text{cm}^{-1}$ .	127
Tableau 15 : Classification SVM sur la médiane, la moyenne et l'ensemble des spectres des patients cirrhotiques avec et sans CHC. Analyse sans dérivation seconde sur la région spectrale 900-1800 $\text{cm}^{-1}$ .	132
Tableau 16 : Classification PLS-DA sur la médiane, la moyenne et l'ensemble des spectres des patients cirrhotiques avec et sans CHC. Analyse sans dérivation seconde sur la région spectrale 900-1800 $\text{cm}^{-1}$ .	133
Tableau 17: Classification SVM sur la médiane, la moyenne et l'ensemble des spectres des patients cirrhotiques avec et sans CHC. Analyse avec dérivation seconde sur la région spectrale 900-1800 $\text{cm}^{-1}$ .	137
Tableau 18 : Classification PLS-DA sur la médiane, la moyenne et l'ensemble des spectres des patients cirrhotiques avec et sans CHC. Analyse avec dérivation seconde sur la région spectrale 900-1800 $\text{cm}^{-1}$ .	138
Tableau 19 : Résultats des analyses SVM et PLS-DA sur la cohorte de 817 patients.	139

## AVANT-PROPOS

Le carcinome hépatocellulaire (CHC) est la troisième cause de mortalité par cancer dans le monde (1). En France, il constitue le deuxième cancer digestif chez l'homme après les cancers colorectaux. Son incidence, croissante depuis 1980, est estimée à 10.4/100.000 chez l'homme et de 2/100.000 chez la femme. Il représente actuellement plus de 6000 nouveaux cas par an (2).

Le CHC se développe dans 90 % des cas chez des malades atteints de cirrhose dont la cause est en France essentiellement alcoolique, virale (virus C et virus B) ou métabolique. Le pronostic est très sombre en raison d'un diagnostic habituellement tardif. Lorsque la tumeur est symptomatique, l'espérance de survie est inférieure à 12 mois. Selon les recommandations actuelles, les malades atteints de cirrhose sont surveillés par échographie hépatique semestrielle (3). Toutefois, l'efficacité de l'échographie de dépistage est insuffisante pour la détection du carcinome hépatocellulaire à un stade précoce et le dosage associé du marqueur tumoral alpha-foetoprotéine n'ajoute pas de bénéfice significatif. Cela souligne la nécessité de développer de nouvelles méthodes de diagnostic précoce pour améliorer le pronostic (3).

Au cours de la dernière décennie, l'approche protéomique a été largement utilisée pour identifier de nouveaux marqueurs du CHC (4). Jusqu'à présent, aucune des molécules candidates n'a été validée pour la pratique clinique. D'autres techniques telles que la spectroscopie vibrationnelle pourraient permettre d'identifier des marqueurs sériques permettant de détecter précocement ce cancer. En effet la spectroscopie vibrationnelle du sérum permet d'obtenir une empreinte spectrale des composés présents dans le sérum à la fois qualitativement et quantitativement. Cette technique permet ainsi d'identifier, en utilisant des méthodes d'analyse chimométrique, des marqueurs spectroscopiques caractérisant divers états pathologiques.

Dans ce travail de thèse, nous avons évalué le potentiel des spectroscopies Raman et infrarouge appliquées au sérum pour le diagnostic du CHC chez les patients atteints de cirrhose.

Dans le premier chapitre de ce travail, nous décrivons les principaux aspects épidémiologiques, étiologiques et cliniques du CHC ainsi qu'un rappel du principe des différentes techniques de spectroscopie et des analyses chimométriques utilisées dans le cadre de cette étude.

Dans le deuxième volet de cette thèse, nous présenterons la méthodologie utilisée pour les analyses Raman et infrarouge de nos échantillons biologiques, à savoir le sérum des patients cirrhotiques avec ou sans carcinome hépatocellulaire. Nous développerons aussi les différents prétraitements et analyses chimiométriques appliqués dans ce travail.

Dans le dernier volet, nous décrirons les résultats obtenus pour les études pilotes Raman et infrarouge dont le but était d'évaluer le potentiel de ces deux approches spectroscopiques pour le diagnostic du CHC. Ensuite nous présenterons les résultats de l'étude IR menée sur un nombre plus important de patients et dont le but était de valider les résultats obtenus dans l'étude de faisabilité. Enfin, nous terminerons avec une conclusion générale ainsi qu'une présentation des perspectives futures qu'ouvre ce travail.

# CHAPITRE 1 : INTRODUCTION GÉNÉRALE

### I. Le carcinome hépatocellulaire (CHC)

#### 1. Généralités

Le foie est un organe abdominal logé chez l'homme sous le diaphragme, dans la partie supérieure droite de l'abdomen. Le foie est le plus volumineux des viscères humains. Sa masse moyenne est de 1500 grammes environ, et il représente 2% de la masse corporelle totale chez l'adulte. Il est formé de deux lobes droit et gauche (Figure 1). Le foie remplit un nombre important de fonctions essentielles à la vie. Les hépatocytes, cellules principales du foie (80 à 90%), sont impliquées dans de nombreuses fonctions biochimiques et métaboliques :

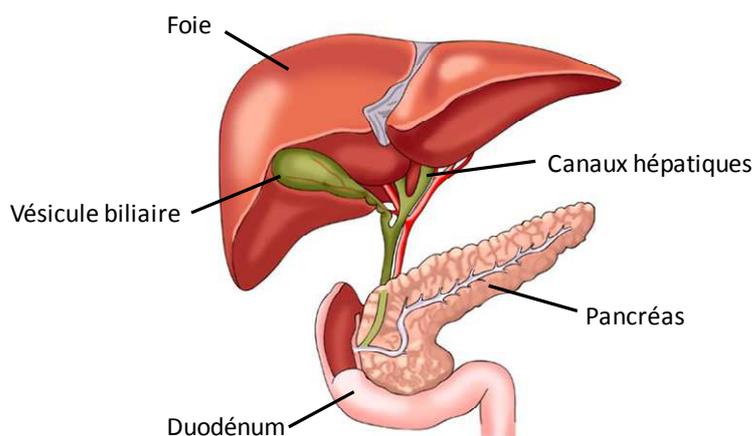
1. Une fonction d'épuration : la détoxification de l'organisme des substances toxiques : alcool, médicaments, polluants divers auxquels il est exposé.

2. Une fonction de synthèse : la production de nombreuses molécules comme par exemple l'albumine, des facteurs de coagulation, les acides biliaires...

3. Une fonction de transformation et de stockage des divers nutriments : glucides, lipides, protéines et le stockage des vitamines et du glycogène.

4. une fonction exocrine : la sécrétion de la bile.

Le parenchyme hépatique, est également composé de cellules biliaires, de cellules endothéliales permettant les échanges entre les hépatocytes et le sang, de cellules de Küpffer, qui sont des macrophages et de cellules étoilées.

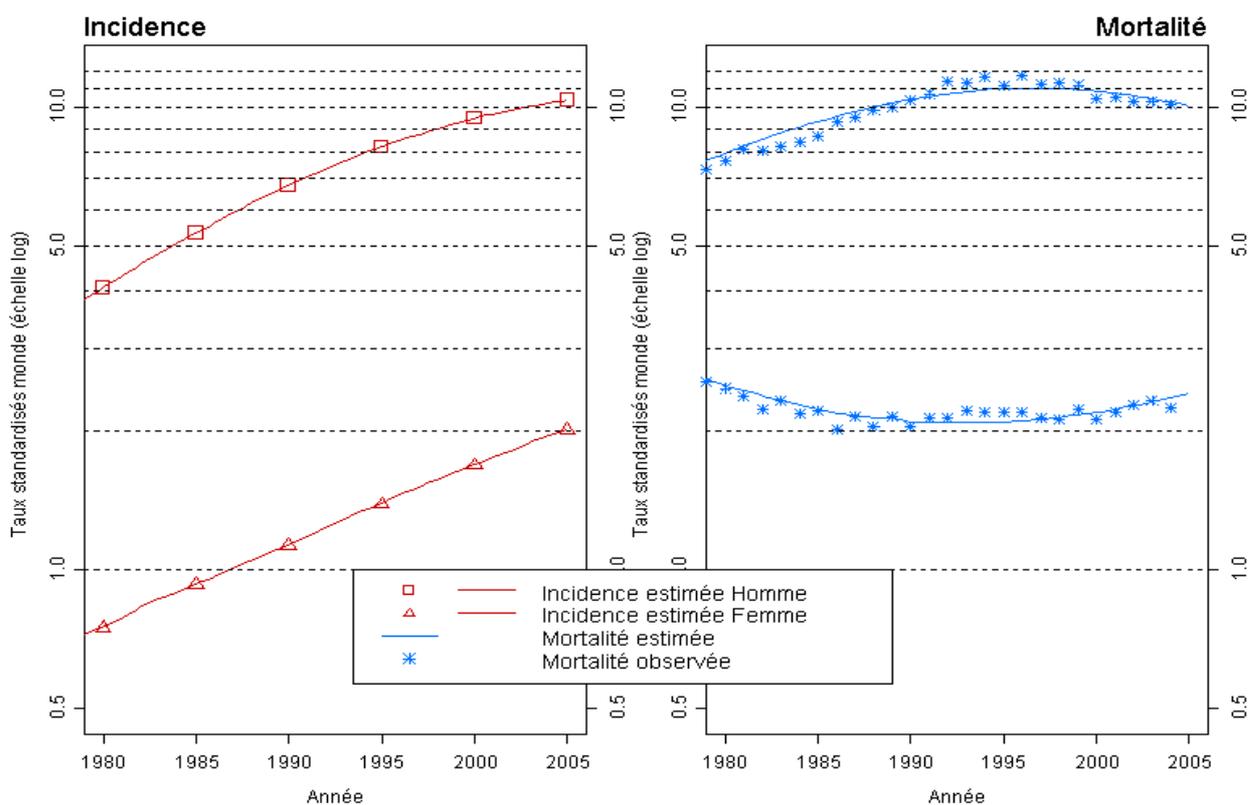


**Figure 1 : Anatomie du foie et des organes adjacents**

### 2. L'épidémiologie du CHC

A l'échelle mondiale, le cancer primitif du foie est le cinquième cancer le plus fréquent chez l'homme et le 7<sup>ème</sup> chez la femme, représentant au total environ 750 000 nouveaux cas par an (1). Du fait de son mauvais pronostic (le ratio de la mortalité à l'incidence est de 0.93), il représente la troisième cause de mortalité par cancer dans le monde soit environ 700 000 décès par an. Le sex ratio homme/femme est de 2,4 (1). La forme histologique la plus fréquente des cancers primitifs du foie est le carcinome hépatocellulaire représentant plus de 90 % de l'ensemble des cancers primitifs du foie.

En France, l'incidence du cancer primitif du foie a augmenté au cours des dernières décennies. L'augmentation annuelle sur la période 1980-2005 a été de 3.8 % chez l'homme et 4% chez la femme. Le nombre de nouveaux cas diagnostiqués en 2005 était de 5104 chez l'homme et 1329 chez la femme (2) (figure2). L'incidence standardisée sur l'âge du cancer primitif du foie estimée en 2012 pour la population française était de 16,6 pour 100 000 chez l'homme et de 3.6 pour 100 000 chez la femme (5).



**Figure 2 : Incidence et mortalité du cancer du foie en France.**

Figure tirée de : A. Belot, M. Velten, P.Grosclaude, N. Bossard, G. Launoy, L. Remontet, E. Benhamou, L. Chérié-Challine et al. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2005. INVS. Accès internet: [http://www.invs.sante.fr/publications/2009/estimation\\_cancer\\_1980\\_2005/estimation\\_cancer\\_1980\\_2005.pdf](http://www.invs.sante.fr/publications/2009/estimation_cancer_1980_2005/estimation_cancer_1980_2005.pdf)

### 3. Les facteurs étiologiques du CHC

Dans la majorité des cas, une maladie préexistante du foie est à l'origine du développement du CHC. Les principaux facteurs de risque à l'échelle mondiale sont représentés par les infections virales chroniques B et C, la consommation excessive d'alcool, l'aflatoxine et le syndrome métabolique. Ces facteurs de risque interviennent dans la carcinogénèse hépatique le plus souvent via la formation d'une cirrhose. Des études épidémiologiques ont établi que le risque de CHC chez les patients cirrhotiques était de 1 à 8% par an selon la cause de la cirrhose (3). La répartition géographique des facteurs de risque est très hétérogène (3) (Tableau 1).

En Asie de l'Est et en Afrique, l'infection chronique par le virus B est impliquée dans plus de la moitié des cas de CHC. L'infection par le virus B y est le plus souvent contractée à la naissance et le CHC survient chez l'adulte jeune. La contamination fréquente dans ces régions des aliments par la mycotoxine aflatoxine jouerait un rôle inducteur supplémentaire. Dans les pays occidentaux, et notamment en France, l'infection par le virus C est la première cause de CHC avec la consommation excessive d'alcool. Dans ces populations, le CHC complique habituellement une cirrhose après l'âge de 50 ans. Le syndrome métabolique est un facteur de risque qui émerge dans les pays développés. Le CHC sur foie sain est exceptionnel. Il représente moins de 10 % de l'ensemble des CHC. C'est pourquoi les patients atteints de cirrhose représentent la population à haut risque de CHC, ce qui en fait la population cible des programmes de dépistage du CHC.

**Tableau 1 :** Répartition géographique des principaux facteurs de risque de CHC (adapté d'après (3))

Zone géographique	Facteurs de risque		Alcool (%)	Autres (%)
	Virus C (%)	Virus B (%)		
Europe	60-70	10-15	20	10
Amérique du nord	50-60	20	20	10 (NASH)
Asie et Afrique	20	70	10	10 (aflatoxine)
<i>Japon</i>	70	10-20	10	10
Monde entier	31	54	15	

## *Chapitre 1 : Introduction générale*

---

Chez un patient atteint d'une maladie chronique du foie, un tissu cicatriciel remplace les cellules hépatiques endommagées : c'est la fibrose hépatique. L'agression chronique des hépatocytes (par l'alcool, un virus, une réaction immunitaire....) provoque l'apparition d'une fibrose qui s'accroît au fil du temps en détruisant l'architecture du foie et en créant des nodules structurellement anormaux.

Selon l'ampleur des dommages subis par le foie, la fibrose peut-être plus ou moins importante, et l'on distingue plusieurs stades. Selon la classification METAVIR (6), le stade F1 désigne une fibrose légère, le stade 2 une fibrose moyenne, le stade F3, une fibrose extensive, précirrhotique. Le stade F4, correspond au stade de cirrhose définie par l'existence d'une fibrose diffuse du parenchyme hépatique qui désorganise l'architecture lobulaire normale en formant des nodules d'hépatocytes de taille variable structurellement anormaux.

La cirrhose est réversible quand la cause de la maladie est traitée et les lésions à un stade précoce. Cependant, en l'absence de traitement, la cirrhose peut se compliquer et prédispose à la survenue du CHC.

### **4. La survenue et la progression du CHC**

Le développement du CHC est un processus composé de trois phases primordiales : une phase d'initiation, de promotion, et de progression tumorale. En effet, l'exposition répétée et prolongée à un ou plusieurs facteurs de risque, induit une mort cellulaire par nécrose, ce qui provoque une inflammation dans le foie conduisant progressivement à la formation d'une cirrhose via l'activation des cellules étoilées du foie et des fibroblastes des espaces portes. Ces cellules se différencient en myofibroblastes responsables de la fibrogénèse dans le parenchyme hépatique. Dans un foie cirrhotique, les hépatocytes subissent des cycles de nécrose et de régénération, accumulant progressivement des anomalies moléculaires. Ces anomalies causent plusieurs altérations fonctionnelles qui donnent aux clones cellulaires un avantage de croissance.

C'est dans ce contexte que peuvent apparaître des foyers d'hépatocytes dysplasiques. Lorsque ces foyers deviennent macroscopiques et détectables à l'imagerie, on parle alors de nodules dysplasiques. Ces nodules peuvent dégénérer en CHC si les altérations génétiques sont suffisamment importantes pour entraîner la transformation et l'immortalisation de ces hépatocytes (7).

### **5. Les marqueurs biologiques du CHC**

Les biomarqueurs, détectables dans le sang, l'urine et les tissus, jouent un rôle important dans le diagnostic de diverses pathologies en particulier le cancer. Les fluctuations de leur concentration permettent de suivre la progression de la maladie et ainsi de diriger le traitement. Le carcinome hépatocellulaire est un des cancers les plus étudiés pour la recherche de biomarqueurs diagnostiques, thérapeutiques et pronostiques. Parmi ces biomarqueurs, nous pouvons citer :

#### **5.1. Alpha-foetoprotéine (AFP)**

L'alpha-foetoprotéine (AFP) est le marqueur sérique le plus largement utilisé dans le diagnostic et le suivi clinique des patients atteints de CHC. La performance diagnostique de ce marqueur chez des patients cirrhotiques ayant une tumeur hépatique est dépendante du seuil de concentration sérique retenu. Avec un seuil de 20 ng/mL, la sensibilité a été évaluée à 60% et la spécificité à 90%. Lorsque le seuil retenu est de 200 ng/mL la spécificité est excellente (99%) mais la sensibilité chute à 22% (8). D'autre part la performance de ce marqueur est également dépendante de la taille de la tumeur. Seulement 10 à 20 % des tumeurs à un stade précoce ont une augmentation de la concentration sérique, alors que ces tumeurs représentent la cible du dépistage chez les patients cirrhotiques (3). D'autre part, une augmentation du taux sérique d'AFP peut être observée au cours des maladies chroniques du foie en l'absence de CHC. Enfin, il a été démontré que la combinaison de l'AFP au dépistage par échographie hépatique n'apporte pas un bénéfice significatif en regard du surcoût (3). Pour toutes ces raisons, l'usage de ce marqueur dans le dépistage du CHC chez les patients cirrhotiques n'est plus recommandé par les sociétés savantes européennes (3) et américaines (9). Il reste toutefois utilisé comme test diagnostique des tumeurs, comme marqueur de suivi et comme facteur pronostique du CHC.

L'intérêt du dosage d'une isoforme fucosylée de l'AFP appelé AFP-L3 a été testé chez des patients ayant un taux élevé d'AFP pour améliorer la valeur diagnostique de ce marqueur. L'utilisation du ratio de l'AFP-L3 sur AFP totale n'améliore pas la performance du dosage de l'AFP seule dans le dépistage du CHC chez les patients cirrhotiques. En revanche, il existe des données en faveur de sa valeur pronostique péjorative (10).

#### **5.2. Des-gamma-carboxyprothrombine (DCP)**

DCP est une forme moléculaire anormale de prothrombine résultant de l'absence de carboxylation post translationnelle du précurseur de la prothrombine. La DCP est produite

## ***Chapitre 1 : Introduction générale***

---

par les hépatocytes malins par suite d'un déficit de l'expression du gène gamma-carboxylase mais également par des hépatocytes normaux chez les malades ayant un déficit en vitamine K (cholestase) ou traités par médicaments antivitamine K. La DCP a une valeur proche de la l'AFP en tant que marqueur diagnostique de CHC chez les malades cirrhotiques ayant une tumeur hépatique (11, 12). D'autre part, comme l'AFP, la DCP est un marqueur pronostique du CHC, significativement associée aux tumeurs agressives envahissant le système vasculaire (13). En revanche, sa valeur en tant que test de dépistage est médiocre avec une sensibilité inférieure à 50 % qui le rend inadéquat pour un programme de dépistage (14).

### **5.3. Glypicane-3**

La Glypicane-3 (GPC3) est une protéoglycane membranaire qui joue un rôle important dans la régulation de l'activité des facteurs de croissance. Il a été démontré que l'ARNm de cette molécule était significativement plus élevé dans le CHC que dans le foie normal et les tumeurs bénignes (15). Il est maintenant bien établi que la protéine est exprimée dans la majorité des CHC et la recherche de GPC-3 par immunohistochimie est désormais recommandée pour le diagnostic histopathologique de CHC en cas de difficultés pour le diagnostic différentiel entre CHC et nodule dysplasique. Le dosage sérique du GPC-3 a également été proposé comme test diagnostique et de dépistage. Dans une série limitée de 38 patients cirrhotiques atteints de CHC, la présence de GPC-3 dans le sérum a été observée dans 53 % des cas et seulement dans un cas chez 20 patients ayant une hépatite virale (16). Toutefois aucune étude sur un grand nombre de patients et en situation de dépistage n'a été menée pour valider ce marqueur en pratique clinique courante. Son dosage n'est pas recommandé en dehors d'études de recherche clinique (3, 9)

### **5.4. Alpha-1-fucosidase**

L'Alpha-1-fucosidase est une enzyme lysosomale dont la fonction est d'hydrolyser les liaisons glucosidiques du fucose et des glycolipides. Il a été rapporté que sa concentration sérique était plus élevée chez les patients cirrhotiques ayant un CHC que chez les patients cirrhotiques sans CHC avec une sensibilité et une spécificité de 75 à 90% (17). Son dosage combiné à l'AFP a été proposé pour le diagnostic des CHC de petite taille. Aucune étude de dépistage sur un grand nombre de patients n'a validé son intérêt pour la pratique médicale courante.

## **6. Les symptômes liés au CHC**

Le carcinome hépatocellulaire peut être longtemps asymptomatique, compliquer une cirrhose connue, ou être la cause d'une décompensation révélant la cirrhose jusqu'alors compensée et

## *Chapitre 1 : Introduction générale*

---

non connue. La décompensation peut se manifester par une aggravation de l'hypertension portale avec hémorragie digestive révélatrice, l'apparition ou la majoration d'une ascite (qui devient irréductible ou hémorragique) ou d'une insuffisance hépatocellulaire qui se traduit par un ictère et/ou une encéphalopathie. Il peut exister des douleurs liées à l'extension ou à la nécrose de la masse tumorale.

### **7. Le diagnostic du CHC**

- Les examens biologiques sont aspécifiques, sauf l'augmentation du taux de l'alpha-fœtoprotéine (AFP) mais, comme indiqué précédemment, ce marqueur est peu performant dans les stades précoces de la maladie.

- Le diagnostic radiologique : L'échographie est utilisée pour dépister l'apparition de nodules dans le parenchyme hépatique chez les patients cirrhotiques. Selon les dernières recommandations des sociétés savantes (3, 9) le diagnostic de CHC peut être posé sur la base des caractéristiques vasculaires des nodules sans avoir recours à la biopsie dirigée. Ces caractéristiques sont l'hypervascularisation au temps artériel et le wash-out veineux. Elles peuvent être mises en évidence par le scanner et/ou l'imagerie par résonance nucléaire avec injection de produit de contraste. Ces examens permettent de préciser également le nombre et la taille des lésions ainsi qu'un éventuel envahissement porte ou sus-hépatique. Lorsque la tumeur est de petite taille < 2 cm, le diagnostic non invasif par imagerie peut être pris en défaut (3). Dans cette situation, la réalisation d'une biopsie dirigée doit être envisagée.

- La biopsie dirigée par échographie ou scanner: elle n'est indiquée que lorsque le diagnostic de CHC ne peut être posé par les examens d'imagerie comme indiqué précédemment. Le diagnostic différentiel entre CHC et dysplasie de haut grade peut être difficile. Dans ce cas, une étude immunohistochimique de la glypican-3, la protéine du choc thermique 70 et la glutamine synthétase est recommandée : la positivité de deux parmi ces marqueurs est très évocatrice du diagnostic de CHC (9).

### **8. Le bilan d'extension extrahépatique du CHC**

Le bilan d'extension inclut un scanner ou une IRM abdominale et un scanner thoracique à la recherche de métastases viscérales (surrénales et pulmonaires principalement) ou ganglionnaires. D'autres examens d'imagerie peuvent être réalisés en fonction de points d'appel clinique (scintigraphie si douleurs osseuses par exemple).

### **9. Le dépistage du CHC**

Le traitement curatif du CHC n'est possible qu'au stade précoce du CHC ce qui justifie la mise en œuvre d'un programme de dépistage. La population cible est représentée principalement par les patients cirrhotiques puisque 90% des CHC se développent sur foie de cirrhose. Les recommandations récentes associent deux autres populations, les patients porteurs chroniques du virus B avec hépatite chronique active ou une histoire familiale de CHC et les patients infectés par le virus C ayant une fibrose extensive (3).

Le dépistage repose sur une échographie hépatique tous les six mois. Comme on l'a vu précédemment, l'association du dosage systématique de l'AFP ou d'un autre marqueur sérique n'est pas recommandée actuellement (3, 9).

### **10. La prévention du CHC**

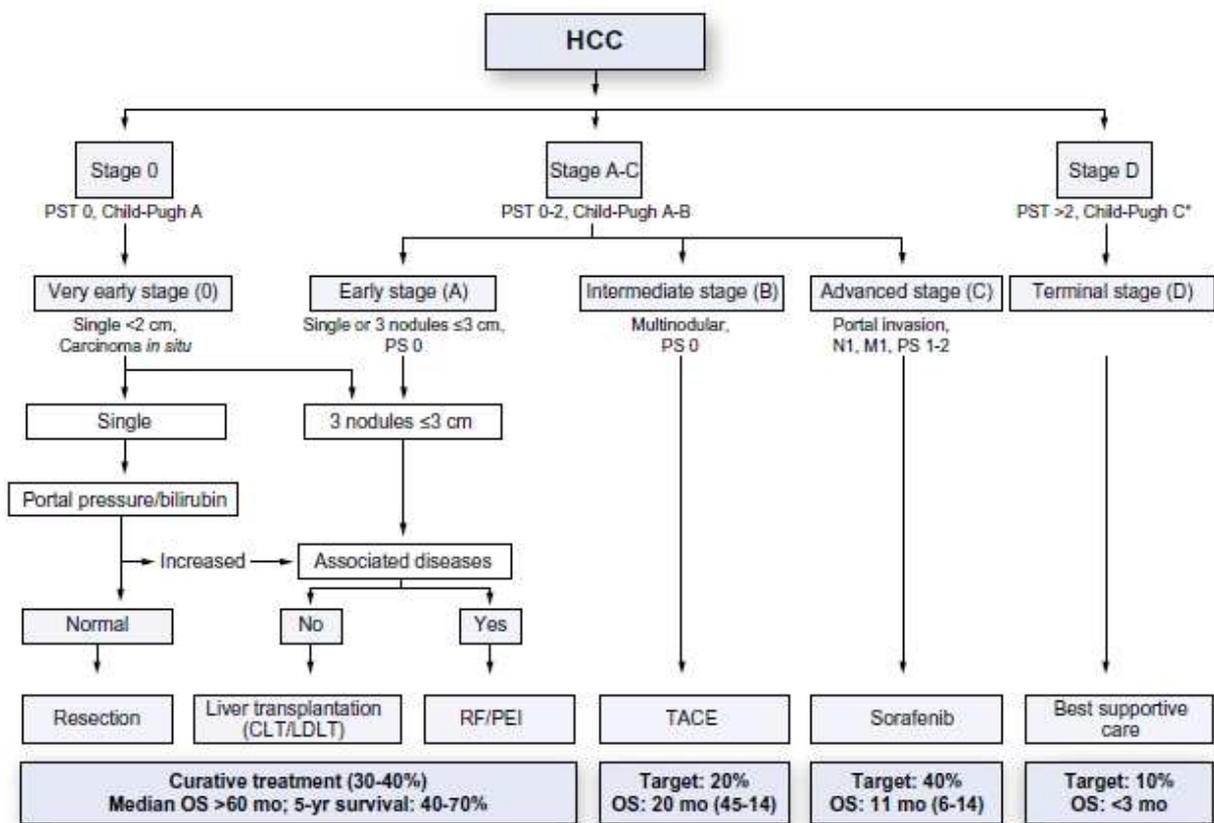
La prévention du développement d'un CHC constitue la première ligne efficace pour combattre cette maladie. La vaccination contre l'hépatite B prévient le CHC lié à ce facteur de risque. La vaccination est recommandée par l'OMS chez tous les nouveaux nés et chez les adultes à risque (3). Il n'y a pas actuellement de vaccin contre l'hépatite C mais les mesures de prévention de l'infection (notamment l'usage de matériel à usage unique par les toxicomanes, le dépistage des donneurs de sang, les règles de bonnes pratiques dans les soins) sont efficaces. Chez les patients infectés par le virus B et le virus C, il a été démontré que le traitement antiviral réduisait significativement le risque d'évolution vers le CHC (3).

La prévention du CHC repose également sur le contrôle des autres facteurs de risque d'hépatopathie chronique, notamment la consommation alcoolique chronique excessive, le syndrome métabolique et la correction d'une surcharge en fer hépatique.

### **11. Le traitement du CHC**

La prise en charge du CHC requiert des compétences étendues et une collaboration multidisciplinaire (chirurgien, gastro-entérologue, radiologue, anesthésistes...).

La stratégie thérapeutique repose sur la classification de Barcelona Clinic Liver Cancer (BCLC) qui propose un algorithme de traitement en fonction de l'extension de la maladie, de la sévérité de la maladie hépatique sous-jacente et de l'état général du patient (Figure 3).



**Figure 3 : Classification de Barcelona Clinic Liver Cancer (BCLC) avec algorithme thérapeutique.**

HCC: Hepatocellular carcinoma, PST: Performance status test, RF: Radiofréquence, PEI: Percutaneous ethanol injection, TACE: Transarterial embolization, CLT: Cadaveric liver transplantation, LDLT: Living donor liver transplantation (3)

Les traitements curatifs du CHC sont représentés par la transplantation hépatique, la résection chirurgicale de la tumeur et par les techniques de destruction focalisée (3, 9).

La transplantation hépatique a l'avantage de traiter le CHC et en même temps la maladie hépatique sous-jacente. Compte tenu de la pénurie de greffons, la sélection des patients est très stricte. Elle est réservée essentiellement aux patients répondant aux critères de Milan c'est-à-dire une tumeur unique de moins de 5 cm ou 3 tumeurs de moins de 3 cm. Chez les patients remplissant ces critères, la survie après transplantation est de 65 à 78% à 5 ans.

La possibilité de résection de la tumeur est dépendante de la maladie hépatique sous-jacente et en particulier du degré d'hypertension portale et du degré d'insuffisance hépatocellulaire. Elle est réservée aux tumeurs de moins de 5 cm développées dans un foie de cirrhose non décompensée.

## *Chapitre 1 : Introduction générale*

---

La destruction in situ de la tumeur est obtenue principalement par une technique de radiofréquence réalisée par voie percutanée ou préopératoire. Elle évite la perte de tissu fonctionnel. Elle est réservée idéalement aux petites tumeurs de moins de 3 cm.

Les traitements palliatifs du CHC sont représentés principalement par la chimioembolisation et le sorafénib (3, 9).

La chimioembolisation est l'association d'une chimiothérapie intra-artérielle sélective et d'une embolisation par gélatine résorbable ou microparticules. Elle est réservée aux patients ayant un CHC multifocal sans envahissement vasculaire.

Le sorafénib est une molécule anti-angiogénique inhibitrice des récepteurs du vascular endothelial growth factor (VEGF) et du platelet-derived growth factor (PDGF) et inhibitrice des MAP-kinases. Il a été démontré que ce traitement augmentait significativement la survie chez les patients qui ne peuvent bénéficier d'aucun des traitements précédents.

Chez les malades ayant un CHC évolué développé sur une maladie du foie sévère, le traitement symptomatique est le seul possible.

Cette courte revue des options thérapeutiques souligne l'importance d'un diagnostic précoce qui seul permet d'envisager un traitement à visée curative. D'où l'importance d'identifier de nouveaux marqueurs de la maladie pour un dépistage plus efficace.

## **II. La spectroscopie vibrationnelle**

### **1. Introduction**

Les méthodes de spectroscopie vibrationnelle d'absorption infrarouge (IR) et de diffusion Raman sont devenues aujourd'hui des outils indispensables dans le milieu de la recherche. Ces deux techniques sont non seulement complémentaires d'autres méthodes de caractérisation de la matière (par exemple la spectroscopie de masse, la Résonance magnétique nucléaire (RMN)), mais fournissent aussi des informations, inaccessibles par d'autres techniques, sur la structure, la composition moléculaire, et les interactions au sein d'un échantillon. Ces techniques sont de plus en plus utilisées dans le domaine de l'industrie, notamment pour l'identification et la caractérisation des molécules dans le cadre du contrôle qualité.

La spectroscopie vibrationnelle est un outil d'analyse qui permet de déterminer la composition chimique d'une substance (solide, liquide ou gaz) par l'interaction entre une onde électromagnétique excitatrice et les liaisons moléculaires qui la composent. Elle fournit des informations qualitatives et quantitatives de l'échantillon étudié.

### **2. Les modes de vibrations moléculaires**

Dans une molécule, les atomes sont liés entre eux par l'intermédiaire des liaisons (covalente, ionique, Van der Waals...). Si la molécule est dans son état fondamental (état de stabilité maximale), les atomes sont quasiment figés dans leur position d'équilibre. En revanche, si un facteur extérieur vient perturber la molécule, un photon par exemple, les atomes ont la possibilité de vibrer autour de leur position d'équilibre si ce photon peut être absorbé. Il est également possible d'observer des mouvements de rotation des atomes autour du centre de masse de la molécule.

Les vibrations moléculaires dépendent de la géométrie de la molécule, des atomes qui la composent et de la nature des liaisons entre les atomes. En étudiant ces vibrations, on est par conséquent capable d'extraire des informations sur la structure de la molécule. Ceci forme la base de la spectroscopie vibrationnelle.

Une vibration moléculaire se produit lorsque les atomes d'une molécule sont dans un mouvement périodique pendant que la molécule dans son ensemble subit un mouvement de translation et de rotation. La fréquence du mouvement périodique est appelée fréquence de vibration. Une molécule non-linéaire constituée de  $n$  atomes possède  $3n-6$  modes normaux de

## *Chapitre 1 : Introduction générale*

vibration, alors qu'une molécule linéaire n'en possède que  $3n-5$ , puisque la rotation autour de son axe moléculaire ne peut être observée. Une molécule diatomique ne possède ainsi qu'un mode normal de vibration. Les modes normaux des molécules poly-atomiques sont indépendants les uns des autres, chacun d'entre eux impliquant des vibrations simultanées des différentes parties de la molécule.

Les vibrations simples peuvent être classées en deux grands groupes : les vibrations d'élongation (stretching) et les vibrations de déformation (bending) et qui se déclinent en fonction de leur symétrie.

### **2.1. Élongation**

Appelée aussi vibration de valence ou "stretching", ce mode concerne la vibration de la molécule le long de l'axe des liaisons. Ce mouvement implique une variation de la distance interatomique.

Les vibrations d'élongation sont représentées par « $\nu$ ». Ces vibrations se situent dans la région du spectre allant de 4000 à 1000  $\text{cm}^{-1}$ . Ce mode de vibration peut être symétrique ou asymétrique (Figure 4):

1. Vibration symétrique ( $\nu_s$ ) : c'est une vibration avec conservation de la symétrie moléculaire.
2. Vibration asymétrique ( $\nu_{as}$ ): c'est une vibration avec une perte d'un ou plusieurs éléments de symétrie de la molécule ce qui exige plus d'énergie.

$$\nu_{as} > \nu_s$$

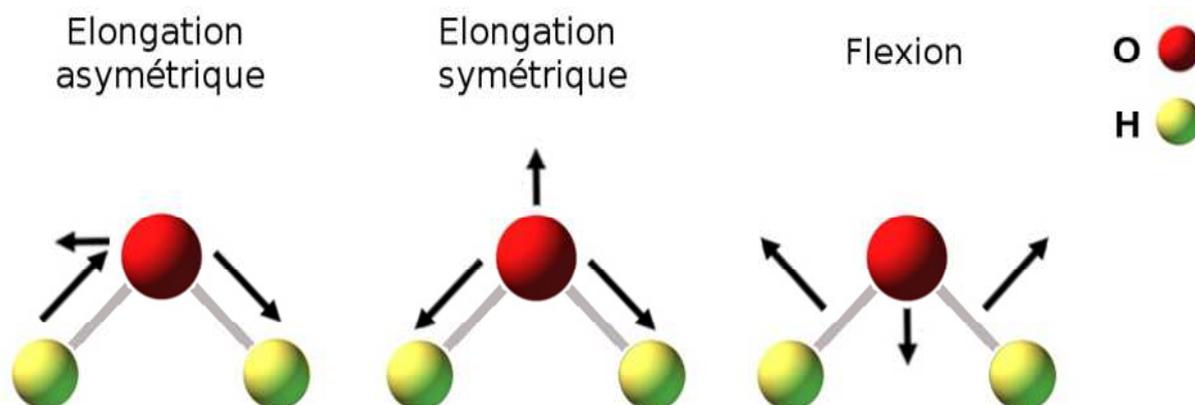
### **2.2. Déformations dans et hors du plan**

Considérons la molécule  $\text{H}_2\text{O}$  (Figure 4.a). En plus de la vibration d'élongation, l'angle des liaisons peut varier : il y a flexion ou déformation. Lors de ce mouvement, la distance interatomique reste constante. Ces déformations peuvent avoir lieu dans le plan des deux liaisons concernées ou hors du plan.

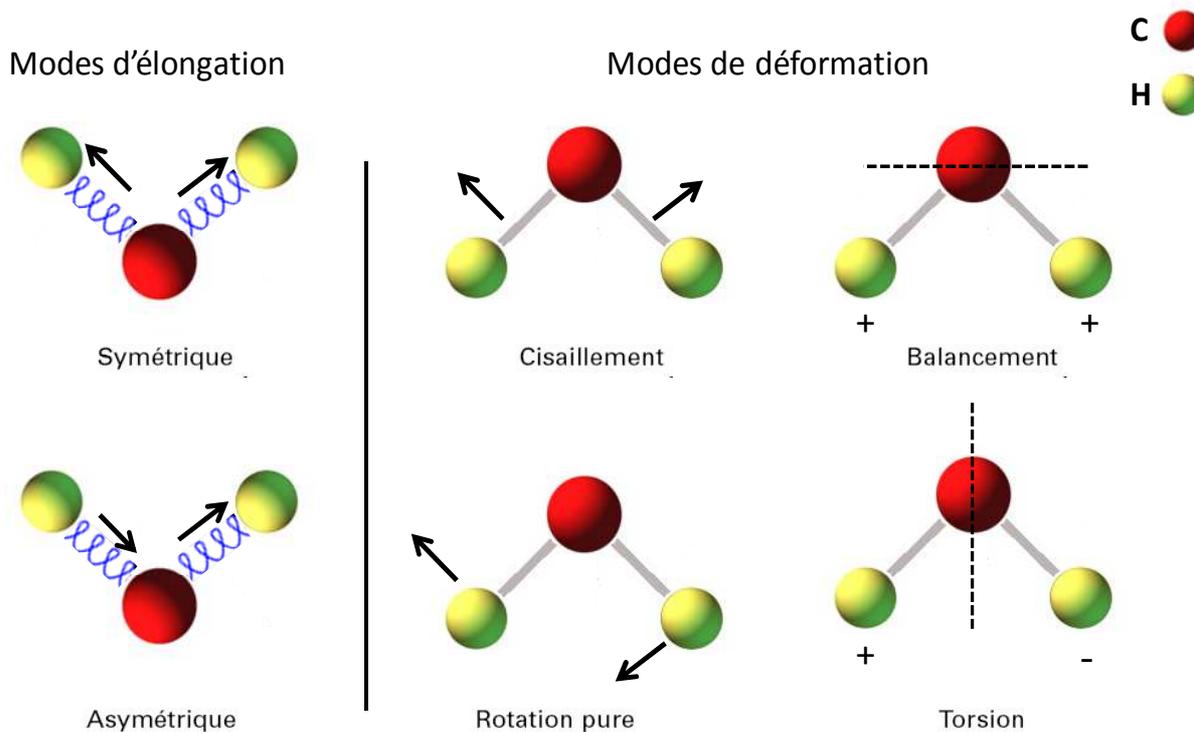
La Figure 4 présente les modes de vibrations fréquemment rencontrés en spectroscopies vibrationnelles pour deux molécules  $\text{H}_2\text{O}$  et  $\text{CH}_2$ :

## Chapitre 1 : Introduction générale

a. Exemple des différents modes de vibration de  $\text{H}_2\text{O}$  :



b. Exemple des différents modes de vibration de  $\text{CH}_2$  :



**Figure 4 : Les modes de vibrations moléculaires des molécules  $\text{H}_2\text{O}$  et  $\text{CH}_2$ .**

Les vibrations de déformations sont d'intensité plus faible que celle d'élongation. Elles sont nombreuses et beaucoup plus sensibles à l'environnement car elles ont besoin pour se produire d'un volume plus important et risquent donc d'être entravées par la présence d'atomes voisins.

### **3. La spectroscopie infrarouge à transformée de Fourier (IRTF)**

#### **3.1. Principe de la spectroscopie IR**

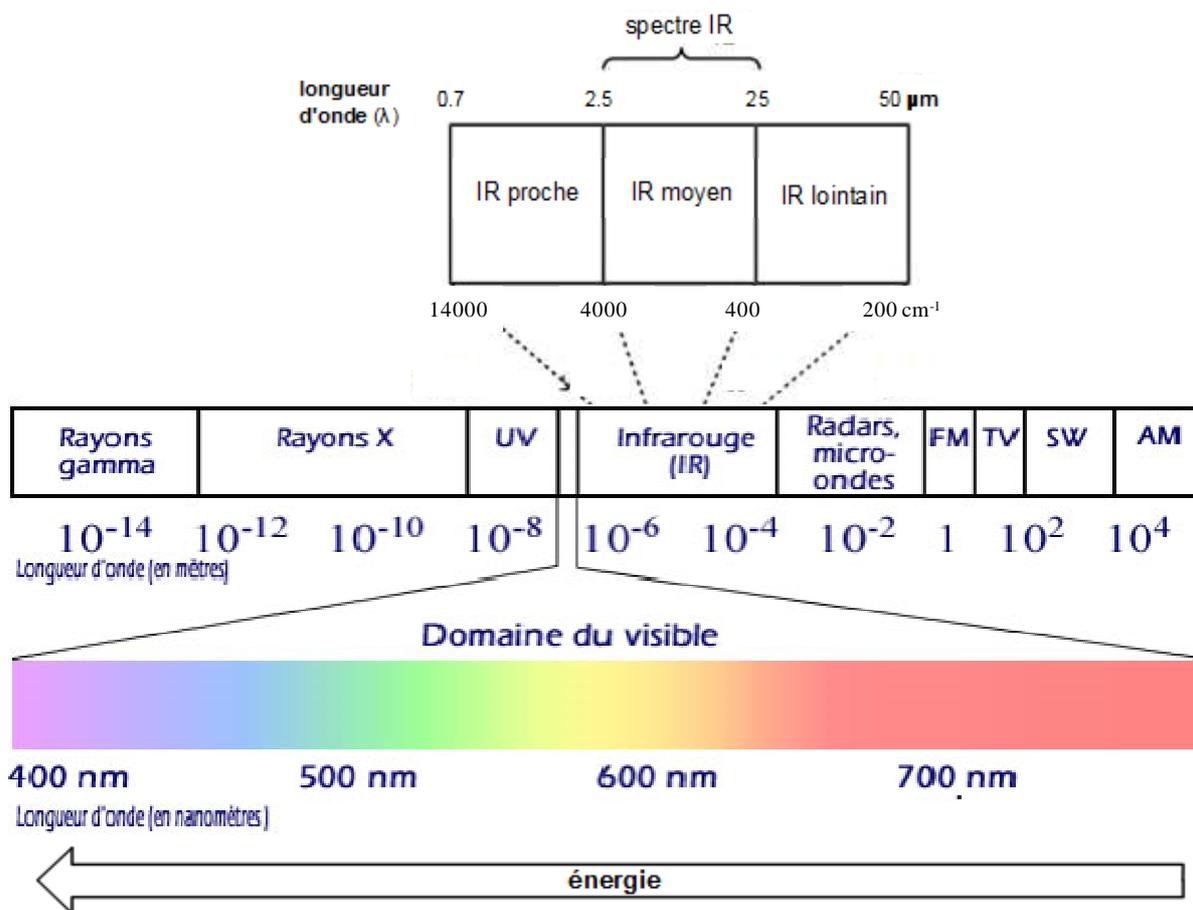
La spectroscopie IR est une technique d'analyse permettant de déterminer la nature des liaisons chimiques présentes dans un échantillon et d'en caractériser les groupements moléculaires afin d'obtenir de nombreuses informations sur la conformation et les éventuelles interactions. Dans le cas de molécules biologiques, elle permet de détecter des variations spectrales liées à différents constituants macromoléculaires comme les acides nucléiques, les lipides, les glucides ou encore les protéines présents dans le spécimen étudié, de façon qualitative et quantitative. Notons que seule une molécule ayant une variation de son moment dipolaire, peut être active en IR.

La région infrarouge est localisée entre la région du spectre visible et des ondes hertziennes (Figure 5). Le domaine infrarouge s'étend de 0,8  $\mu\text{m}$  à 1000  $\mu\text{m}$ . Il comprend 3 zones : le proche infrarouge (0,8 à 2,5  $\mu\text{m}$  soit 12500-4000  $\text{cm}^{-1}$ ), le moyen infrarouge (2,5 à 25  $\mu\text{m}$  soit 4000-400  $\text{cm}^{-1}$ ) et le lointain infrarouge (25 à 1000  $\mu\text{m}$  soit 400-10  $\text{cm}^{-1}$ ).

Le spectre infrarouge représente l'absorbance « A » en fonction de la longueur d'onde «  $\lambda$  » qui est traditionnellement exprimée sous la forme du nombre d'onde «  $\nu$  ».

$$\nu = 10^4 / \lambda \text{ en cm}^{-1} \quad \text{Avec } \lambda \text{ en } \mu\text{m}.$$

Les régions du proche IR et du lointain IR ont suscité un certain intérêt et ont été utilisées dans plusieurs études. Cependant, le moyen IR reste le plus adapté et le plus utilisé pour l'étude de la composition moléculaire d'un échantillon.

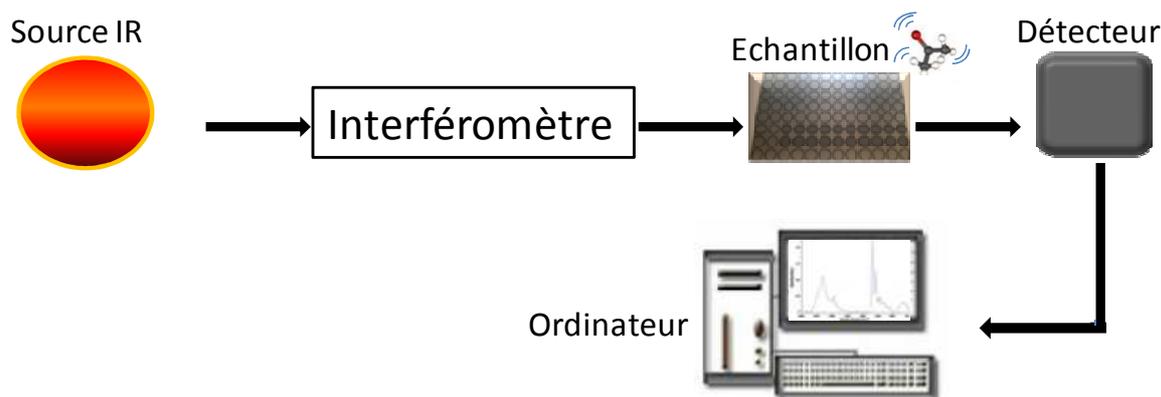


**Figure 5 : Le spectre électromagnétique.**

### 3.2. Spectromètre IRTF

Un spectromètre IRTF comporte essentiellement 5 parties comme montré dans la Figure 6:

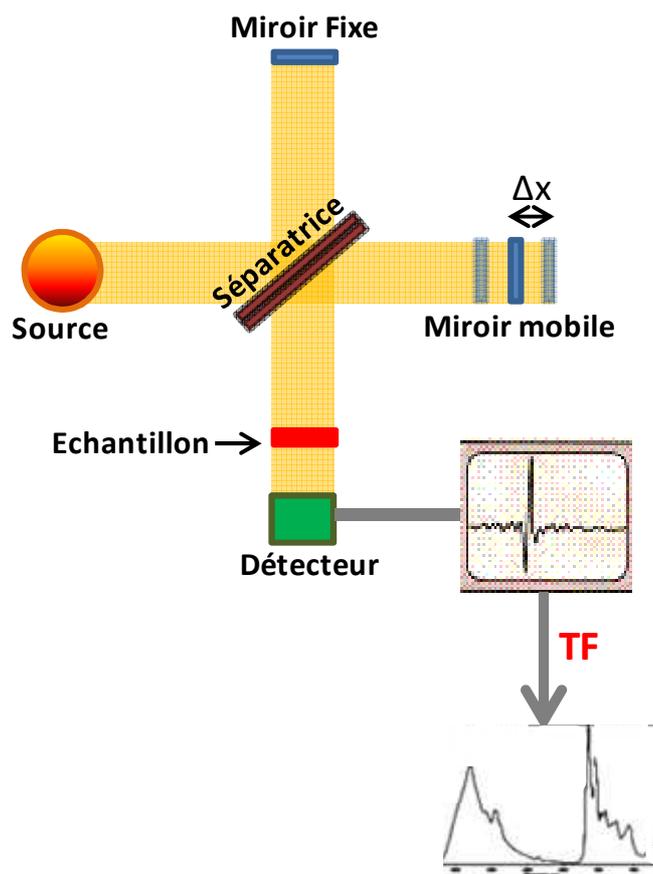
- Une source lumineuse polychromatique, type Globar<sup>®</sup>
- Un interféromètre, un dispositif permettant de moduler le signal et générer les interférences
- Un compartiment échantillon
- Un détecteur ou capteur photosensible
- Un convertisseur analogique numérique qui interroge le détecteur à des intervalles réguliers et transforme le signal analogique en un signal numérique manipulable par le système informatique.



**Figure 6 : Schéma représentant les principaux éléments d'un spectromètre IRTF.**

L'interféromètre est l'élément essentiel d'un spectromètre infrarouge à transformée de Fourier. Il permet de mesurer toutes les fréquences simultanément. L'interféromètre le plus connu est l'interféromètre de Michelson qui se compose de deux plans perpendiculaires à l'axe optique (Figure 7). Ce dispositif est composé d'un miroir fixe et d'un miroir mobile qui peut se déplacer le long de l'axe optique.

Une source émet un faisceau de lumière infrarouge qui est divisée en deux parties grâce à une lame séparatrice comportant une face semi-réfléchissante. La moitié du faisceau est réfléchi et l'autre moitié est transmise. Chacun de ces deux faisceaux est ensuite réfléchi par les miroirs puis retombe sur la séparatrice où les deux faisceaux se recombinent, mais les différences de chemin optique créent des interférences constructives et destructives formant l'interférogramme. Le faisceau recombinaison traverse l'échantillon qui absorbe à différentes longueurs d'onde, caractéristiques de son spectre et qui sont soustraites de l'interférogramme. Le faisceau est recueilli par un détecteur qui voit une variation de l'énergie en fonction du temps pour toutes les longueurs d'onde simultanément. Un faisceau laser (He:Ne) y est superposé et sert de référence pour le fonctionnement de l'appareil, c'est-à-dire pour la conversion analogique-digitale. L'interférogramme, qui représente l'intensité en fonction du temps (ou distance), est ensuite converti en un spectre infrarouge, représenté par l'intensité en fonction de la fréquence, par une transformation mathématique appelée transformée de Fourier (TF), d'où la dénomination « Infrarouge à transformée de Fourier ».



**Figure 7 : Schéma d'un interféromètre de Michelson.**

Deux modes de mesure sont généralement possibles en spectroscopie IR : la transmission et la réflexion (Figure 8).

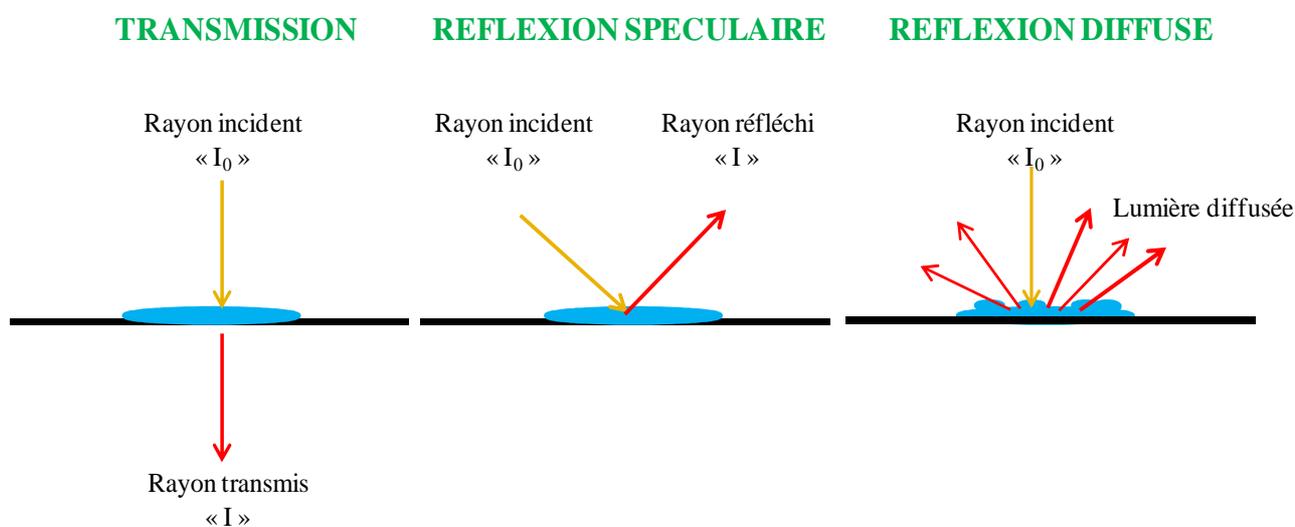
- En mode « transmission », le faisceau infrarouge passe à travers l'échantillon et l'énergie provenant de ce dernier est mesurée. Le signal est recueilli et envoyé au détecteur. En effet, avant l'échantillon, l'intensité de la lumière est  $I_0$ . Elle est égale à  $I$  après avoir traversé l'échantillon. On obtient alors une transmission  $T = I/I_0$ . L'absorbance est donc :

$$A = -\log T = \log (I_0/I)$$

Ce mode de mesure nécessite l'utilisation de supports transparents dans le moyen infrarouge et peu absorbant. Il en existe une multitude parmi ceux-ci, le fluorure de calcium ( $\text{CaF}_2$ ), le bromure de potassium (KBr), le séléniure de zinc ( $\text{ZnSe}$ ), le fluorure de baryum ( $\text{BaF}_2$ ) et le germanium (Ge). Avant toute mesure, une référence doit être enregistrée sur le support vide.

## Chapitre 1 : Introduction générale

- En mode réflexion, la lumière infrarouge est réfléchiée par l'échantillon. Le rayonnement incident est collecté par la même optique et ensuite acheminé vers le détecteur. Dans cette configuration, le signal est alors exprimé en % de réflexion par rapport au "100% de réflexion" obtenu en présentant un échantillon de "référence" considéré comme réfléchissant à 100% (par exemple : miroir en or ou en argent). Sur des échantillons de surface non homogène ou rugueuse, ou de poudres, il s'agit alors de réflexion "diffuse". Sur des échantillons polis ou de surface très plane, la réflexion "spéculaire" est largement prépondérante (comportement miroir), le faisceau réfléchi est monodirectionnel (Figure 8).



**Figure 8 : Schéma simplifié des modes d'acquisition spectrales en spectroscopie IR.**

### 3.3. Génération d'un spectre IRTF

Un spectre infrarouge est obtenu selon les étapes illustrées dans la Figure 9 :

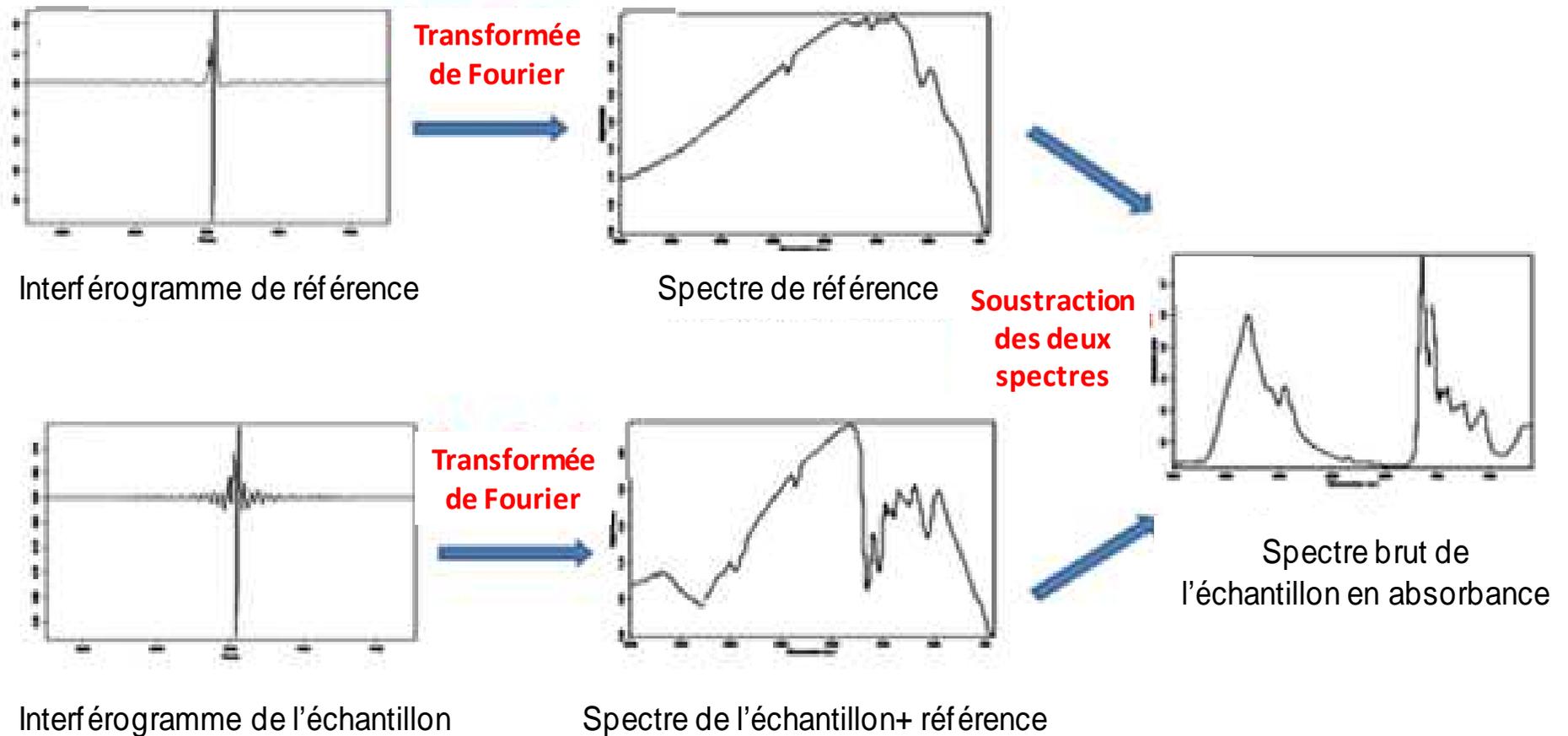


Figure 9 : Schéma montrant les étapes de l'obtention d'un spectre IRTF.

## *Chapitre 1 : Introduction générale*

La première étape est l'enregistrement d'un interférogramme de référence sur le support porte-échantillon (spectre de référence), ensuite le spectromètre procède à l'enregistrement d'un interférogramme simple-faisceau de l'échantillon à analyser. Ces deux signaux sont ensuite transformés par l'équation mathématique de transformée de Fourier en spectres. Le spectre infrarouge de l'échantillon est obtenu après la soustraction des deux spectres [(spectre (échantillon+ référence) –spectre de référence)].

Le spectre IRTF a la caractéristique de fournir des informations qualitatives et quantitatives de l'échantillon analysé :

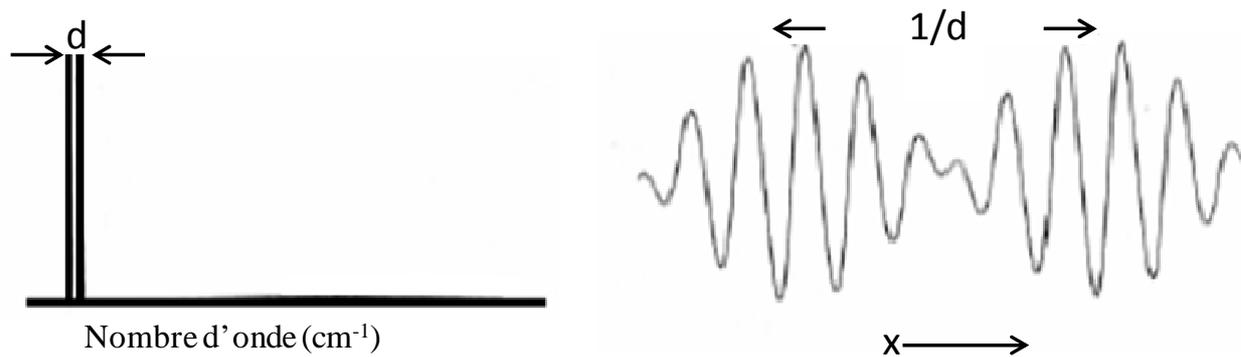
- Informations qualitatives : L'échantillon absorbe à des longueurs d'onde précises et qui sont caractéristiques des groupements chimiques qui le composent.
- Informations quantitatives : L'intensité d'absorption à une longueur d'onde précise est proportionnelle à la concentration du groupement chimique qui est à l'origine de cette absorption (mesure de l'aire du pic caractéristique).

### **3.4. Résolution spectrale**

Ce paramètre est illustré sur la Figure 10. Un écart « d » entre deux bandes sur l'échelle des nombres d'onde est observable lorsqu'une modulation périodique de « 1/d » est constatée sur l'interférogramme : pour obtenir une séparation correcte de « d » entre deux bandes dans le domaine des nombres d'onde, il faut au minimum mesurer l'interférogramme sur une distance  $D=1/d$ .

La résolution spectrale est un paramètre très important qui définira la précision du signal enregistré. C'est-à-dire, plus la résolution est grande, plus le spectre est bien défini. On observera alors le maximum de détail de l'enregistrement. La résolution (en  $\text{cm}^{-1}$ ) est liée au parcours du miroir mobile et est donc un paramètre que l'on règlera selon le type de précision d'analyse que l'on désire. L'appareil effectue une mesure tous les  $X \text{ cm}^{-1}$  du spectre infrarouge. X est définie en début de mesure.

Exemple : pour une résolution spectrale de  $4 \text{ cm}^{-1}$  il faudra au minimum un déplacement du miroir mobile de 0,25 cm.



**Figure 10 : Distance « d » entre deux signaux spectraux et l'interférogramme correspondant.**

### 3.5. Loi de Beer-Lambert

Le spectre d'absorption de la matière est caractérisé par le coefficient d'extinction moléculaire  $\epsilon$  à chaque longueur d'onde. Cette loi stipule que l'absorbance est directement proportionnelle à la longueur  $\ell$  (épaisseur de l'échantillon traversée par le faisceau) et la concentration  $C$ , et est décrite par l'équation suivante :

$$A_{\lambda} = -\log(I/I_0) = \epsilon_{\lambda} \cdot \ell \cdot C$$

- $I/I_0$  : la transmittance de la solution
- $A$  : l'absorbance à une longueur d'onde  $\lambda$  en unité arbitraire.
- $\epsilon_{\lambda}$  : le coefficient d'extinction molaire. exprimé en  $L \cdot mol^{-1} \cdot cm^{-1}$ . Il dépend de la longueur d'onde, la nature chimique de l'entité et la température.
- $\ell$  : la longueur du trajet optique en cm.

### 3.6. Avantages de la spectroscopie IRTF

La spectroscopie infrarouge offre plusieurs avantages d'analyse en comparaison avec les méthodes dispersives :

- Avantage de Jacquinot : Cela résulte du fait que, dans un instrument dispersif, le monochromateur dispose des fentes d'entrée et de sortie qui limitent la quantité de lumière qui passe à travers elles. La quantité de lumière entrant dans ou sortant de l'interféromètre est déterminée par le diamètre du faisceau collimaté issu de la source. Il en résulte très bon rapport signal/bruit.

## ***Chapitre 1 : Introduction générale***

---

- Avantage de Fellgett (ou multiplexe): Ceci provient du fait que l'information de toutes les longueurs d'onde sont collectées simultanément. Il en résulte un rapport signal-bruit élevé pour un temps d'acquisition équivalente.
- Avantage de Connes : donné par la grande précision de l'échelle des fréquences ; ce qui facilite les opérations mathématiques simples comme la correction de ligne de base, la somme et la différence des spectres, ou encore le calcul des moyennes.
- Rapide et à haut débit : l'utilisation d'une plaque multipuits permet d'analyser plusieurs échantillons en même temps.
- Auto-calibration interne automatique.
- Bonne résolution spectrale.

### **3.7. Domaines d'application de la spectroscopie IRTF**

La spectroscopie infrarouge est très répandue dans la recherche académique et dans l'industrie en tant que technique simple et sûre, de contrôle de qualité et de mesure dynamique. Elle est, par exemple, utilisée en :

- Médecine légale : identification des composés.
- Industrie de la micro-électronique : pour la caractérisation des matériaux et/ou quantification du stress subi lors de leur fonctionnement.
- Environnement : pour la caractérisation de polluants ou de contaminants.
- Pharmacie : pour la caractérisation des composés et notamment la répartition du principe actif.
- Géologie : pour l'analyse de minéraux par micro-spectrométrie infrarouge.
- Agroalimentaire : pour l'identification de contaminants (champignons, bactéries...).

### **3.8. Principales applications de la spectroscopie IRTF en cancérologie**

Les limites du dépistage précoce des tests biologiques ont conduit à l'application de la spectroscopie vibrationnelle telle que l'infrarouge pour discriminer entre les cellules normales et cancéreuses ou encore entre différents stades d'évolution du cancer au niveau tissulaire. Le nombre important de publications relatif aux applications de la spectroscopie IRTF montre l'intérêt de cette approche au niveau cellulaire et tissulaire. Elle a été largement utilisée dans une perspective diagnostique et pronostique, en particulier pour détecter les modifications structurales précoces de la tumorigénèse.

## *Chapitre 1 : Introduction générale*

---

Au niveau cellulaire, la spectroscopie IRTF a été appliquée pour l'étude des cellules isolées (18, 19), ainsi que pour l'interaction cellule-médicament (18, 19). Dans leur étude des cellules leucémiques, Ramesh et al (20) ont rapporté l'intérêt de cette approche à différencier les cellules cancéreuses et les cellules saines. Dans le même contexte, G. Bellisola et al, en 2010, ont montré le potentiel de la spectroscopie IR dans l'évaluation des effets des médicaments anticancéreux chez les patients atteints de leucémie (21).

Au niveau tissulaire, le potentiel de la technique IRTF, comme un outil efficace pour étudier les modifications moléculaires, a été rapporté dans de nombreuses études diagnostiques du cancer. Dans leur étude du cancer gastrique, en utilisant des coupes tissulaires de patients atteints de cancer et des coupes de patients sains, Qingbo Li et al ont pu obtenir un taux de bonne classification de 92,2% en utilisant la spectroscopie IRTF combinée à des méthodes de classifications avancées (22).

La différenciation entre un tissu ovarien sain et un tissu pathologique par spectroscopie IRTF a été également rapportée par Krishna et al. (23).

En ce qui concerne le cancer de la prostate, il a été démontré que les changements biochimiques associés à ce cancer pourraient être discriminés par FTIR ce qui permettrait d'améliorer le diagnostic et le pronostic de cette pathologie (24, 25).

L'intérêt de la spectroscopie IRTF à diagnostiquer les tumeurs a été aussi étudié pour la peau (26), l'œsophage (27), l'estomac (28), le poumon (29), l'intestin (30) ou encore pour le carcinome de la vésicule biliaire (31).

Couplée à des algorithmes chimiométriques automatiques, la spectroscopie vibrationnelle IRTF peut fournir une solution haut débit, objective et à faible coût pour le diagnostic précoce du cancer. Cependant, il est nécessaire de disposer d'un échantillon tissulaire suffisamment grand pour être représentatif de la population étudiée. Pour mieux mener ces études, une collaboration entre histopathologistes et biophysiciens serait donc très indispensable.

### **4. La microspectroscopie Raman**

#### **4.1. Introduction**

Parmi les spectroscopies vibrationnelles on distingue la diffusion Raman (Figure 11). En effet, cette technique, découverte en 1928 par le physicien indien Chandrasekhara Venkata Raman, repose sur un effet de diffusion de la lumière. Les émissions observées reflètent les

## Chapitre 1 : Introduction générale

polarisations induites par un faisceau exciteur. L'effet Raman est à la base du développement de méthodes de spectrométrie Raman. Les sources de lumière laser, particulièrement bien adaptées à l'excitation de l'effet Raman, ont été à l'origine des progrès importants de la technique, en permettant par exemple l'analyse d'échantillons microscopiques en un temps très court.

Le diagramme de Jablonski représente les différents phénomènes induits suite à l'interaction entre les moments dipolaires des liaisons moléculaires et une onde électromagnétique, tout en considérant les transitions électroniques (Figure 11).

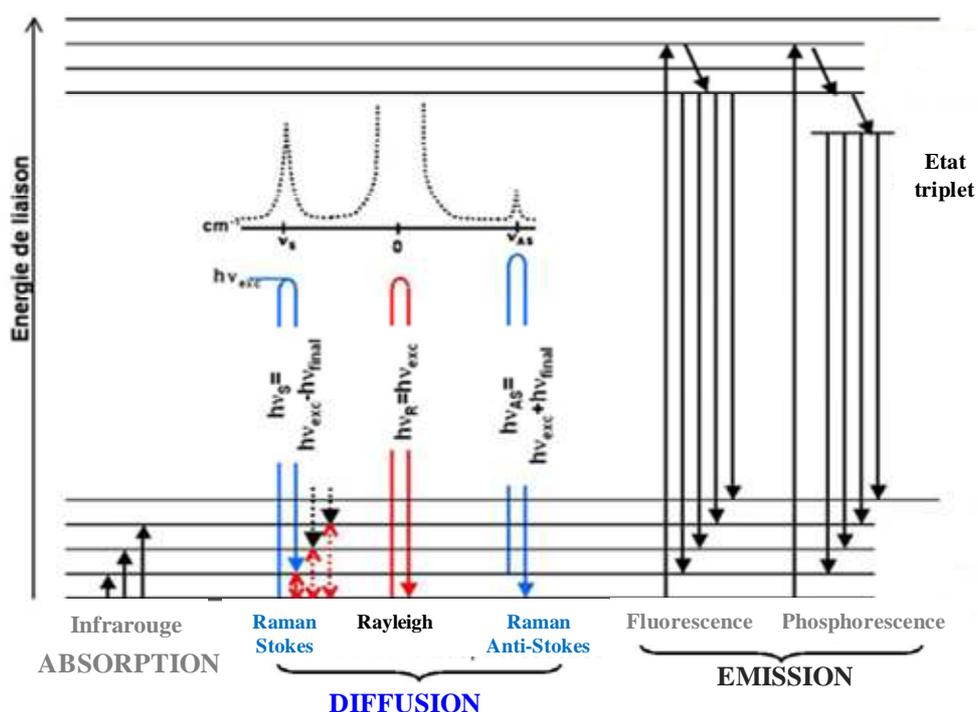


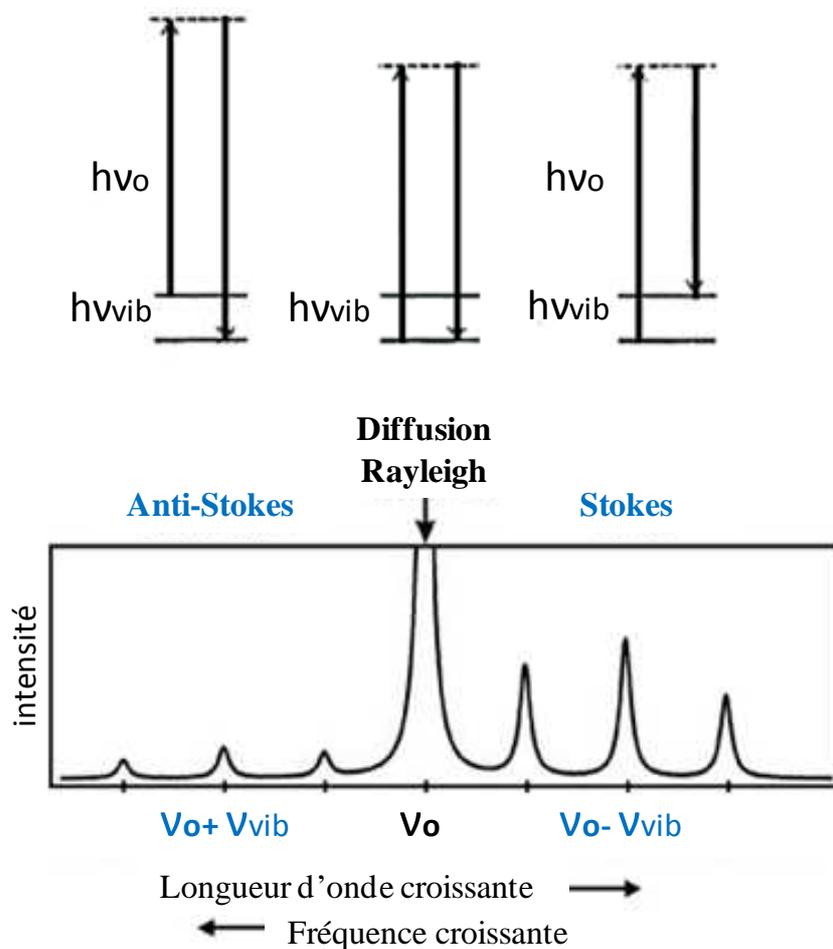
Figure 11 : Diagramme de Jablonski illustrant les transitions impliquées dans les phénomènes d'absorption, de diffusion et d'émission.

### 4.2. Principe de la diffusion Raman

Lorsqu'on soumet un milieu (solide, liquide, gaz) à une onde électromagnétique monochromatique, ce dernier est porté à un niveau énergétique virtuel. Ensuite, la majeure partie du faisceau incident est transmise, mais une partie de la lumière est diffusée dans toutes les directions de l'espace, avec un maximum d'intensité dans les directions perpendiculaires au faisceau incident. Ces radiations sont collectées par un détecteur. Elles comportent deux types de signaux (Figure 12) :

## Chapitre 1 : Introduction générale

- Diffusion Rayleigh : majoritaire, la radiation incidente est diffusée élastiquement sans changement d'énergie donc de longueur d'onde.
- Diffusions Stokes et anti-Stokes : ces diffusions sont inélastiques et correspondent à un échange d'énergie entre le rayonnement et la matière. Celle-ci absorbe (ou cède) de l'énergie aux photons incidents. La variation d'énergie observée sur le photon renseigne alors sur les niveaux énergétiques de vibration de la molécule concernée. L'ensemble des radiations diffusées, analysées par un spectromètre, forme le spectre Raman de l'échantillon.



**Figure 12 : Spectre de transition du niveau fondamental aux états virtuels.**

Notons qu'en spectroscopie Raman, seule une molécule avec une variation de la polarisation autour de sa position d'équilibre, peut être active.

Au fil des années, différentes formes de spectroscopie Raman ont été développées et utilisées pour des applications biologiques. La spectroscopie Raman conventionnel avec une excitation

## *Chapitre 1 : Introduction générale*

laser dans l'ultraviolet ou le visible a souvent posé des problèmes soit de fluorescence (32) de photodégradation ou les deux quand elle est appliquée aux systèmes biologiques complexes. Dans les années 1980, la spectroscopie Raman à transformée de Fourier a été développée pour pallier aux problèmes de fluorescence et de photodégradation, mais sa sensibilité était diminuée à cause de l'utilisation d'un laser proche-infrarouge à 1064 nm (33). La spectroscopie Raman de résonance avec excitations ultraviolet ou visible peut être utilisée pour des molécules spécifiques ciblées par la sélection de la longueur d'onde d'excitation permettant le phénomène de résonance, donnant ainsi lieu à de forts signaux Raman. Ces signaux de résonance étaient spécifiques qu'à quelques pics du spectre. La spectroscopie Raman exaltée de surface (DRES) ou SERS (Surface-Enhanced Raman Spectroscopy), découverte il y a 40 ans, est une technique qui permet de d'augmenter la sensibilité de la technique Raman des fois jusqu'à un facteur  $10^6$  (34, 35). L'effet SERS peut être utilisé pour développer des sondes permettant de détecter des signatures moléculaires sous forme de traces et a été utilisée pour des applications médicales. Grâce à des fibres optiques, la spectroscopie Raman peut être déportée, et donc utilisable dans des applications médicales in vivo. Un laser dans le proche-infrarouge est ici recommandé pour des mesures in vivo.

### **4.3. Le microspectromètre Raman**

La source laser délivre un faisceau monochromatique vers le microscope où l'on retrouve les différents objectifs chargés de l'illumination optimale de l'échantillon et de la collection la plus efficace possible de la lumière diffusée par effet Raman, ainsi que de son transfert vers l'entrée du spectromètre. La Figure 13 décrit le système utilisé dans ce travail, nous y trouvons:

Une source laser excitatrice à 785 nm. Il faut noter que le choix de la longueur d'onde d'excitation en spectroscopie Raman est très crucial et résulte d'un compromis : plus l'onde est énergétique (longueur d'onde basse), plus le signal de diffusion est important, mais on peut introduire alors plus de fluorescence parasite ou de photodégradation.

La focalisation du laser sur l'échantillon se fait avec l'objectif Olympus Bx40 d'un microscope optique (Olympus). La mesure est ensuite collectée à l'aide d'un objectif 100X long focal. Le couplage spectromètre / microscope constitue la caractéristique la plus importante du microspectromètre Raman permettant d'atteindre une résolution spatiale de l'ordre du micromètre (dépend du laser utilisé et l'ouverture numérique de l'objectif).

## Chapitre 1 : Introduction générale

Le système est alimenté de filtres atténuateurs qui permettent de réduire, en cas de besoins, l'intensité du signal incident évitant ainsi une photodégradation de l'échantillon. La raie excitatrice laser (réflexion), la diffusion Raman anti-Stokes et la diffusion Rayleigh (beaucoup plus intense que la diffusion Raman) de même longueur d'onde que l'excitatrice, sont éliminées à l'aide d'un filtre Notch.

Le signal Raman-Stokes est dirigé vers le spectrographe en passant par un trou confocal qui permet de sélectionner le signal émis par un point précis de l'échantillon. Ce signal est ensuite focalisé sur la fente d'entrée du spectrographe. La largeur de la fente d'entrée définit la résolution spectrale de l'appareil. La réduction de la largeur de la fente diminue l'intensité du signal Raman. Le signal est ensuite capté par un détecteur photoélectrique constitué d'une mosaïque du type CCD (charge coupled device) de 1064x256 pixels. Ce système assure la mesure simultanée de l'intensité des différentes longueurs d'onde du spectre.

La dernière partie de l'équipement permet de mesurer et de traiter les signaux électriques issus du détecteur. Le développement des techniques numériques de traitement et la puissance des systèmes micro-informatiques disponibles ont permis d'accroître la qualité des mesures spectroscopiques et d'assurer, en même temps, la gestion des différentes fonctions du spectromètre.

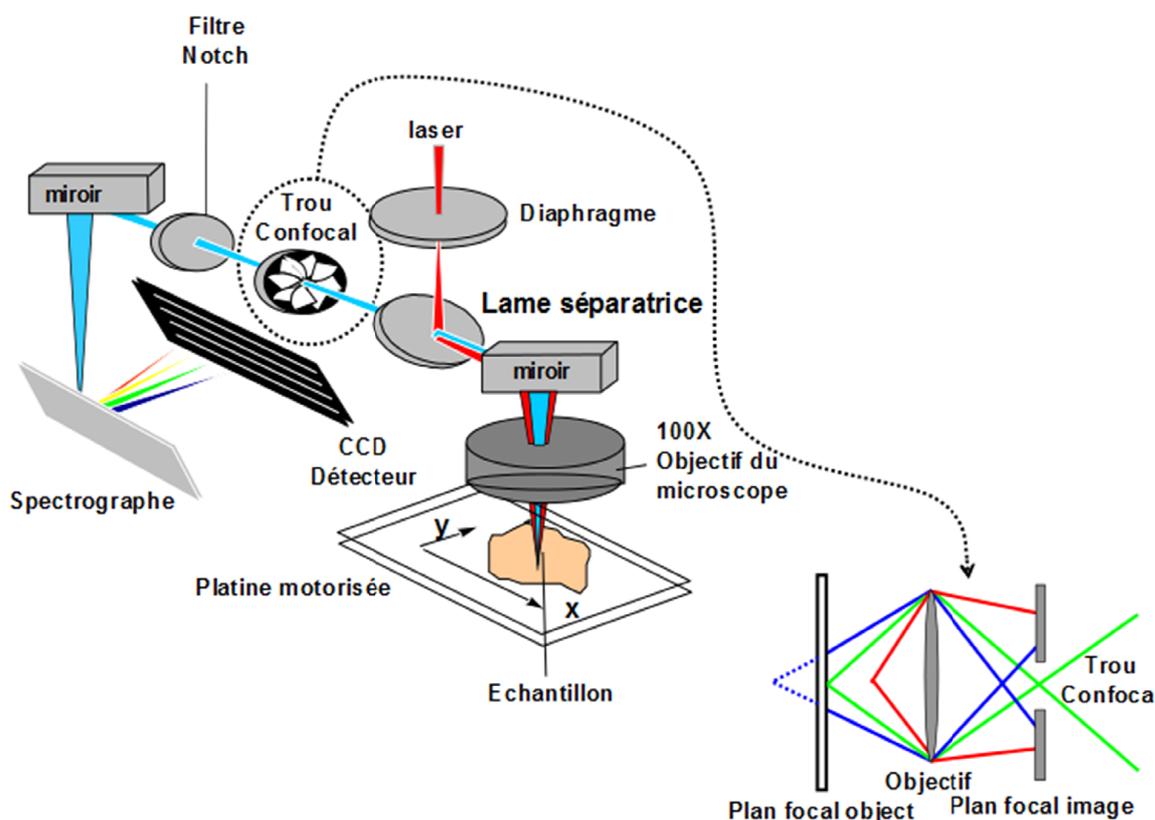


Figure 13 : Schéma d'un microspectromètre Raman.

### **4.4. Avantages de la spectroscopie Raman**

Par rapport aux autres techniques d'analyse, la spectroscopie Raman apporte certains avantages parmi lesquels:

- Pas/peu de préparation de l'échantillon avant analyse.
- Analyse non destructive (choix de la longueur d'onde excitatrice): l'échantillon peut être utilisé pour d'autres analyses.
- Etude de solutions aqueuses possible (l'eau diffuse peu en Raman).
- Bonne résolution spatiale (1 à 2  $\mu\text{m}$ ).
- La confocalité permet de faire des mesures en profondeur de l'échantillon.

Cependant, cette technique présente quelques inconvénients tels:

- La fluorescence qui masque le signal.
- Bibliothèque de spectres encore très incomplète.
- Analyse quantitative : avec précaution (plutôt semi-quantitative).
- Sensibilité inférieure qu'en infrarouge.

### **4.5. Domaines d'applications de la spectroscopie Raman**

Depuis de nombreuses années, la spectroscopie optique est appliquée à un certain nombre de domaines, y compris :

- La chimie analytique : La détermination de la structure chimique (C=O, C=C, O-H, C-O,... etc), la configuration des polymères, la détermination de la cristallinité et de l'épaisseur lamellaire pour les polymères, l'étude des forces intra- et inter-moléculaires (liaison hydrogène) et l'étude de l'orientation des molécules (polarisation)...
- La géologie, la minéralogie et la gemmologie.
- L'histoire de l'art : l'analyse des peintures dans les œuvres d'art.
- La santé : diagnostic et pronostic.

### **4.6. Principales applications de la spectroscopie Raman en cancérologie**

Au cours des dernières années, l'utilisation de la spectroscopie optique pour des applications biomédicales a considérablement augmentée. Son attrait vient de sa capacité à fournir des

## *Chapitre 1 : Introduction générale*

---

informations qualitatives et quantitatives sur les états biochimiques et morphologiques des tissus. La spectroscopie Raman a suscité beaucoup d'intérêt dans le diagnostic des pathologies, notamment le cancer. En effet, cette technique est non invasive et sensible aux petites modifications moléculaires qui sont associées à cette pathologie, comme l'augmentation de la proportion du noyau dans le cytoplasme, l'élévation de l'activité métabolique, et les changements dans les taux des lipides et des protéines (36).

Parmi les applications de la spectroscopie Raman, on peut citer l'étude de N. Stone et al pour la détection de tumeurs malignes dans les ganglions lymphatiques de la tête et du cou. Au cours de ce travail une sensibilité et une spécificité de 81% et 89% ont été obtenues (37).

De plus, de nombreuses études ont porté sur l'application de la spectroscopie Raman pour la différenciation des tissus mammaires cancéreux et sains (38-40). Dans leur étude du diagnostic des lésions bénignes et malignes dans le tissu mammaire humain, Haka et al ont rapporté le potentiel de la spectroscopie Raman à mettre en évidence les différences moléculaires existantes entre un tissu sain et un tissu cancéreux avec une sensibilité de 94% et une spécificité de 96% (39).

E. Widjaja et al se sont intéressés à l'analyse du cancer du colon et ont ainsi discriminé entre un tissu cancéreux et un tissu normal avec une sensibilité et une spécificité supérieure à 98% (41).

D'autres cancers ont fait l'objet d'une analyse par spectroscopie Raman dans un but diagnostique dans plusieurs types de tissus distincts tels que la peau (42), l'ovaire (43), l'œsophage (44) et la région cervicale (45).

Combinée à des analyses statistiques multivariées et des algorithmes de discrimination, la spectroscopie Raman permet une classification automatisée des spectres dans les cas pathologiques pertinents. Bien que des limites existent, la technique est une perspective intéressante dans le diagnostic du cancer.

### **5. Analyses des biofluides par spectroscopie vibrationnelle**

Le sérum a été choisi comme l'échantillon d'intérêt dans ce projet car il offre plusieurs avantages par rapport aux échantillons tissulaires pour la recherche des biomarqueurs du CHC. Il est facilement disponible avec une quantité suffisante pour l'analyse et permet de donner une idée complète sur l'état physiologique du patient vu qu'il contient des milliers de

molécules. Le choix de travailler sur du sérum et non pas sur le sang entier ou le plasma est basé sur le fait que cette solution ne contient pas de cellules ou de facteurs de coagulation qui pourraient donner un signal spectral important et ainsi cacher l'information des petites protéines qui pourraient être des biomarqueurs potentiels du CHC.

### **5.1. Spectroscopie infrarouge des biofluides**

La spectroscopie infrarouge a été appliquée à plusieurs biofluides tels que la bile, le liquide céphalo-rachidien, la salive, les larmes, l'urine, le plasma (46). L'analyse du sérum en spectroscopie infrarouge n'a fait l'objet que d'un nombre limité de travaux. Certaines équipes se sont intéressées au dosage par spectroscopie IRTF du glucose dans le sérum (47, 48). D'autres ont démontré l'intérêt de la spectroscopie IR pour le dosage des lactates, de l'urée, des triglycérides et des protéines dans le sérum de sportifs de haut niveau afin d'évaluer leur performance à l'effort (49, 50). D'autres études ont montré son intérêt dans la détection sérologique de l'encéphalopathie spongiforme bovine (51), dans le diagnostic de la polyarthrite rhumatoïde (52) et comme outil pronostique pour déterminer la sévérité des pancréatites aiguës (53).

La spectroscopie infrarouge a été aussi appliquée sur le sérum pour le diagnostic du cancer. Dans leur analyse du sérum, J. Backhaus, et al (54) ont démontré que la spectroscopie infrarouge combinée à une analyse chimiométrique avancée permet de discriminer entre des échantillons provenant de patients atteints de cancer du sein et des patients contrôles avec une sensibilité de 96% et une spécificité de 93%. De même, en 2013, l'équipe de Gajjar (23) ont mis en évidence le potentiel de cette technique dans le dépistage du cancer ovarien avec des taux de sensibilité et de spécificité de 96%.

Plus récemment, l'intérêt de la spectroscopie infrarouge a été rapporté pour le diagnostic du cancer de la région cervicale (55). En effet, une analyse ATR-IRTF du sérum a été combinée à un traitement SVM afin de séparer les patients avec et sans cancer et a donné une sensibilité et une spécificité de 87.5% et 100%, respectivement.

Une autre étude réalisée sur 75 sérums (n=50 cas et n=25 témoins) par James R. Hands en 2013 a présenté, pour la première fois, l'utilisation de l'ATR-IRTF combinée avec une SVM non linéaire pour le diagnostic des gliomes. La sensibilité et la spécificité obtenues étaient de 87.5% et 100% respectivement (56).

Dans une étude pilote réalisée dans notre équipe (57) la spectroscopie infrarouge IRTF a été utilisée afin d'évaluer la fibrose hépatique chez les patients atteints d'hépatite C chronique.

L'étude a été réalisée sur le sérum de 23 patients (Cas =11, Témoin=12). La sensibilité et la spécificité obtenues étaient respectivement de 90.1% et de 100%.

### **5.2. Spectroscopie Raman des biofluides**

L'intensité Raman est  $10^{-6}$  à  $10^{-9}$  fois inférieure à celle de la diffusion Rayleigh, ce qui désavantage la technique en termes de sensibilité. Néanmoins, la spectroscopie Raman présente certains avantages par rapport à la spectroscopie IR, en particulier dans le domaine médical, en raison de l'abondance de l'eau dans la plupart des cellules, des tissus et des liquides biologiques.

En effet, des travaux récents ont démontré l'application de la spectroscopie Raman aux biofluides. Elle a été utilisée pour analyser les métabolites dans les larmes (58), la bile (59) et le sérum (60). Toutefois, l'intérêt majeur dans le domaine de la recherche biomédicale est l'identification de biomarqueurs diagnostiques et pronostiques du cancer ou d'autres pathologies. Dans la littérature, peu d'études ont évalué le potentiel de la micro-spectroscopie Raman appliquée au sérum pour le diagnostic du cancer. Dans leur analyse, Pichardo-Molina et al, (61) ont démontré que la spectroscopie Raman combinée à une analyse chimiométrique avancée permet de discriminer entre des sérums provenant de patients atteints de cancer du sein et les sérums de patients contrôles avec une sensibilité de 97% et une spécificité de 78%. De même, l'équipe de Harris en 2009 (62) ont montré le potentiel de cette technique dans le dépistage du cancer du cou et de la tête avec des taux de sensibilité et de spécificité de 75%.

## **III. Traitement des données spectrales**

### **1. Les prétraitements spectraux**

Après l'acquisition, les données spectrales doivent être traitées. Une question principale se pose donc: comment l'information pertinente peut-elle être extraite de la quantité de données collectées?

L'inspection visuelle des spectres sur les différentes longueurs d'onde est longue et impossible lorsque l'on tente de comparer des spectres très similaires ou encore plusieurs spectres en même temps. Une des premières manipulations sur les données est l'étape du prétraitement.

Ces prétraitements permettent la suppression des artefacts spectraux et spatiaux tels que les différences d'épaisseur de l'échantillon, les effets optiques, l'influence de l'appareil et le bruit de fond du détecteur. Ces techniques sont désormais considérées comme essentielles pour

## *Chapitre 1 : Introduction générale*

---

développer des modèles de classification robustes. Ainsi, les spectres doivent être prétraités avant de les soumettre à des méthodes statistiques multivariées ou chimiométriques parce que les résultats obtenus par les techniques de classification (par exemple : ACP, LDA, SVM...) en dépendent. Ces prétraitements spectraux incluent :

### Le lissage :

Il permet de réduire le bruit dans le signal. Le lissage utilisé dans ce travail est basé sur l'algorithme Savitzky-Golay (63). Cet algorithme a l'avantage de réaliser un lissage de la courbe tout en préservant l'existence et la largeur des pics d'intensité d'intérêt. Il utilise un ensemble différent de coefficients pré-calculés connus dans le domaine de la chimie. Il s'agit d'un type de lissage polynomial commandé par deux paramètres: l'ordre du polynôme et le nombre de points utilisés. Les valeurs possibles pour les points de lissage sont entre 5 et 25. Dans notre étude, nos données ont été lissées avec un polynôme 3 et 11 points. A noter qu'un lissage trop important peut induire une perte d'information spectrale.

### La correction de ligne de base :

En plus de l'absorption (IR) et la diffusion (Raman), d'autres phénomènes physiques peuvent être enregistrés tels que les aberrations chromatiques, la diffraction, la fluorescence...

Ces signaux sont étroitement liés à la nature de l'échantillon et de la source lumineuse. Ainsi la ligne de base d'un spectre brut se présente sous une forme inclinée (ou autres formes) d'où la nécessité de la corriger afin de corriger le spectre. Cette correction se fait selon des paramètres bien définis : degré du polynôme et nombre de points et se fait en estimant une ligne de base passant par les minima du spectre et qui sera ensuite soustraite à ce dernier.

### La normalisation :

Ce traitement est appliqué pour éliminer les facteurs liés aux variations de la source ou du détecteur et pour corriger les différences d'épaisseur de l'échantillon ce qui permettra de simplifier la comparaison entre les données spectrales qualitativement et quantitativement. Les spectres sont donc ramenés à la même échelle d'intensité. La priorité est ainsi donnée à la forme des spectres et non à leurs intensités relatives. Il existe plusieurs types de normalisation :

## *Chapitre 1 : Introduction générale*

1. Normalisation Min / Max : l'intensité d'un pic du spectre est ramenée à une échelle où l'unité d'absorbance minimale sera de 0 et la maximale de 1.
2. Normalisation vectorielle : Cette méthode calcule la moyenne « y » des absorbances du spectre qui sera ensuite soustraite du spectre. La somme des carrés de toutes les valeurs « y » est calculée et le spectre est divisé par la racine carrée de cette somme. L'écart-type est égale à 1.
3. Correction de l'offset : Cette méthode déplace la ligne de base du spectre pour que l'unité d'absorbance minimale soit à 0.

### La dérivation :

La dérivation joue un rôle important dans la séparation des informations spectrales. En effet, les petites différences entre deux spectres très corrélés (des pics qui se recouvrent) ne sont pas visibles à l'œil nu. La dérivation permet donc de les accentuer et les révéler au manipulateur. Cependant, l'inconvénient de ce prétraitement est qu'il est sensible au bruit de fond. Les dérivés les plus communément utilisés sont celles d'ordre 1 ou 2.

### EMSC (Extended Multiplicative Signal Correction)

Quand la lumière infrarouge interagit avec l'échantillon biologique, elle peut être réfléchi, transmise, absorbée ou diffusée par divers facteurs physiques. La méthode de traitement EMSC permet d'estimer la contribution de certains de ces phénomènes physiques et les corrige (64).

Pour un échantillon de « j » constituants. L'absorbance peut être formulée comme :

$$A(\nu) = \left[ \sum_{j=1}^J K_j(\nu) \cdot c_j \right] \cdot b$$

Où : -  $K_j(\nu)$  : l'absorbance caractéristique de l'échantillon à un nombre d'onde  $\nu$ .

-  $b$  : la longueur du chemin optique.

-  $c$  : la concentration du constituant  $j$ .

En présence d'un facteur physique externe, la lumière transmise collectée varie. Ceci peut être dû à des variations de diffusion au sein de l'échantillon analysé.

Cette variation peut être représentée par une constante "multiplicative"  $\alpha$ . L'absorbance est donc exprimée sous la forme :

## Chapitre 1 : Introduction générale

$$A(\nu) = -\log \left[ \frac{\alpha I(\nu)}{I_0(\nu)} \right] = -\log[\alpha] - \log \left[ \frac{I(\nu)}{I_0(\nu)} \right]$$

$$A(\nu) = b \cdot \left[ \sum_{j=1}^J K_j(\nu) c_j \right] + a$$

(MSC : multiplicative signal correction)

La variation de la diffusion peut avoir d'autres effets dépendant du nombre d'onde et qui peuvent être modélisés par des fonctions polynomiales et non linéaires (fonction du nombre d'onde). L'absorbance sera représentée sous la forme suivante :

$$A(\nu) = b \cdot \left[ \sum_{j=1}^J K_j(\nu) c_j \right] + a + d \cdot \nu + e \cdot \nu^2$$

La connaissance des paramètres a, b, d et e permet donc une correction optimale et l'accès à une évaluation des spectres des constituants moléculaires. En pratique, il est nécessaire de déterminer un spectre de référence (en règle générale on prend la moyenne des spectres), ainsi que des spectres d'interférences.

### 2. La chimiométrie

Le terme « chimiométrie » peut être défini comme « la discipline utilisant les mathématiques, les statistiques, ainsi que d'autres méthodes de logique formelle pour définir ou sélectionner des expériences et procédés de mesures optimaux, et pour fournir un maximum d'informations pertinentes par l'analyse de données chimiques » (65).

C'est un outil utilisé afin d'extraire de l'information pertinente et utile à partir de données physico-chimiques mesurées ou connues brutes. Il est basé sur la construction, puis l'exploitation d'un modèle de comportement à l'aide d'outils statistiques. Il peut traiter des systèmes complexes et donc généralement multivariés.

La chimiométrie recouvre aujourd'hui l'ensemble des applications de la chimie, de la physique, de la biologie, de l'économie, de la sociologie, des méthodes statistiques et de l'informatique.

En spectroscopie, la chimiométrie est très utilisée pour obtenir des résultats quantitatifs et qualitatifs sur des systèmes complexes. En effet, cette discipline permet d'extraire, à partir des variables réelles, l'information en condensant les données sous forme de variables latentes. Le

## *Chapitre 1 : Introduction générale*

---

modèle différenciera donc les données utiles et les variables aléatoires comme le bruit de fond.

Cependant, les méthodes chimiométriques doivent être maniées avec beaucoup d'esprit critique. Ces méthodes peuvent présenter des limitations telles que le risque du sur-apprentissage (overfitting) ou encore l'obtention de modèles très spécifiques aux données de calibration et donc devient inefficace pour l'analyse de nouveaux échantillons ce qui est un facteur limitant.

Si les méthodes de traitement de données sont nombreuses, on peut cependant les regrouper en trois catégories principales : Les méthodes descriptives, prédictives et de classification.

Dans ce chapitre nous allons décrire le principe des différentes méthodes de classification utilisées au cours des présents travaux de thèse.

Les méthodes de classification :

Pour une exploitation des résultats, et parce que l'interprétation visuelle des données spectrales est difficile (chevauchement de pics), une multitude d'algorithmes d'analyse multivariée sont actuellement disponibles pour être appliquées aux mesures spectroscopiques obtenus.

Ces méthodes de classification peuvent être classées en deux groupes:

1. Classification non supervisée : appelée aussi partitionnement de données ou clustering, consiste à regrouper les données au sein d'ensembles homogènes nommés classes. Ces classes sont créées automatiquement par le logiciel. Pas de connaissance *a priori*. Parmi ces méthodes on trouve : l'analyse en composante principale (ACP) et la classification hiérarchique ascendante (CHA).
2. Classification supervisée : appelée aussi méthode d'affectation ou classifieur, est caractérisée par une phase d'apprentissage qui consiste à établir des règles de classification à partir des connaissances *a priori* disponibles sur un jeu d'échantillon. Parmi ces méthodes, on retrouve les séparateurs à vaste marges, l'analyse discriminante linéaire et le réseau de neurones artificiels...

A noter que le choix des paramètres de prétraitement affecte fortement la nature de l'information acquise.

### 2.1. Analyse hiérarchique en cluster (HCA)

L'analyse hiérarchique en cluster (HCA) appelée aussi « clustering » est un ensemble de classification non supervisée qui regroupe les données en clusters sans prédéfinir de classe. Chaque objet est comparé avec les objets restants et les deux clusters qui vont fusionner sont ceux dont la distance est la plus faible. La distance spectrale indique le degré de similitude spectrale. Cette distance est définie comme étant leur distance Euclidienne. Plus cette dernière est grande, plus les groupes sont hétérogènes. Graphiquement les distances entre les groupes sont décrites comme un arbre hiérarchique, le dendrogramme (Figure 14).

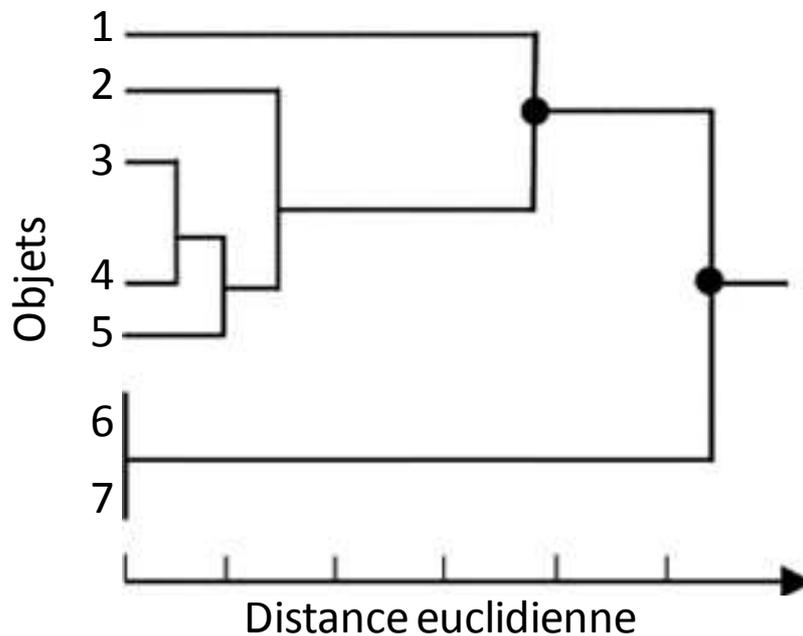


Figure 14: Dendrogramme représentant les distances euclidiennes entre les objets.

### 2.2. Analyse en composante principale (ACP)

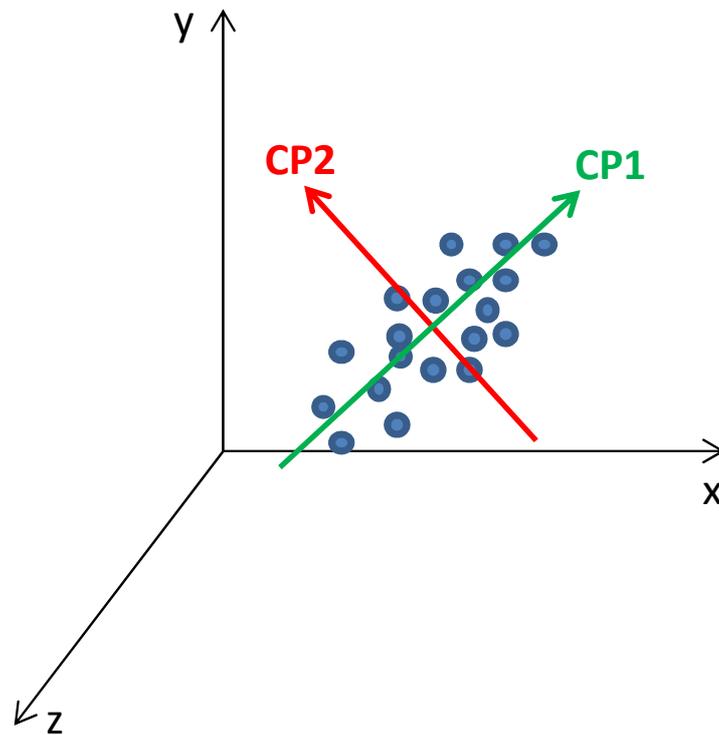
Publiée par Pearson en 1901 (66) et développée ensuite par Hotelling en 1933 (67), l'analyse en composante principale (ACP) consiste essentiellement à trouver les relations existant entre les individus, entre les variables, et entre les individus et les variables. Il s'agit alors de réduire le nombre des variables afin de mieux visualiser les données. Pour ceci, l'ACP est très employée comme une méthode statistique exploratoire permettant de décrire un grand tableau de données de type individus / variables.

## Chapitre 1 : Introduction générale

Individus	V <sub>1</sub>	V <sub>2</sub>	V <sub>3</sub>	,	,	,	V <sub>p</sub>
I <sub>1</sub>							
I <sub>2</sub>							
I <sub>3</sub>							
,							
,							
I <sub>n</sub>							

En présence d'un nombre important de variables dans des données spectrales, la difficulté réside dans l'impossibilité de représenter les individus dans un plan à 2 ou 3 dimensions car ils sont représentés dans un espace de plusieurs dimensions. D'où la nécessité d'utiliser un modèle mathématique tel que l'ACP permettant de réduire un système complexe de corrélations en un plus petit nombre de dimensions en éliminant les variables redondantes tout en conservant la quasi-totalité de l'information spectrale liée à la variabilité des données. Par exemple, pour les variables, on cherchera donc quelles sont celles qui sont très corrélées entre elles et celles qui, au contraire ne sont pas corrélées aux autres.

Lorsqu'elle est appliquée aux données, l'ACP classe les échantillons en groupes en produisant des nuages de points. Un ensemble de données sera constituée de composantes principales (CP). Ces composantes principales contiennent des informations moléculaires sur les composés chimiques. En effet, la première CP est construite de telle sorte à expliquer le maximum de variance, la seconde CP est contrainte à être orthogonale à la première CP et d'expliquer la variance résiduelle non prise en compte par la CP précédente et ainsi de suite (Figure 15).

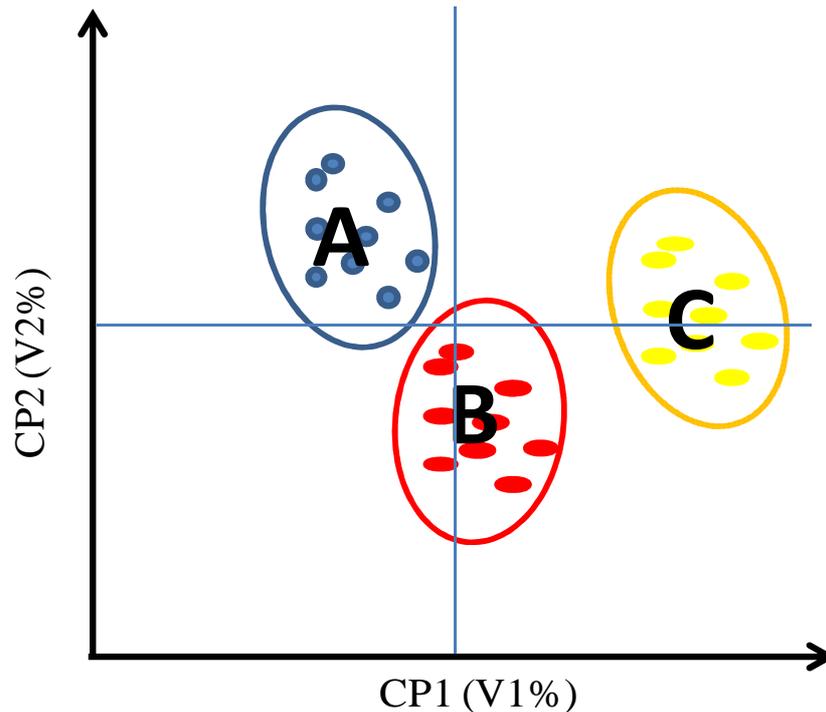


**Figure 15 : Analyse en composante principale (ACP).**

Une fois les CPs calculées, il est nécessaire de connaître les coordonnées des observations dans le nouvel espace défini. Pour cela, les « loadings », correspondant à la contribution ou au poids de chaque variable, et les scores, correspondant aux coordonnées des observations sur les composantes, peuvent être déterminés.

L'ACP cherche donc, à représenter graphiquement les relations entre individus par l'évaluation de leurs ressemblances, ainsi que les relations entre variables par l'évaluation de leurs liaisons. Il s'agit donc d'obtenir le résumé le plus pertinent possible des données initiales. L'interprétation des graphiques ACP permet de bien comprendre la relation entre les données analysées. Cette analyse se fait en tenant compte des qualités de représentation, en termes de regroupement ou dispersions par rapport aux axes factoriels et projections des variables initiales.

La Figure 16 montre un exemple de représentation graphique de 3 échantillons (A, B, C). Ces derniers forment des nuages de points plus ou moins rassemblés et localisés en fonction de la contribution de chaque composante principale représentée dans le schéma.



**Figure 16 : Exemple de représentation graphique des échantillons A, B et C en prenant 2 composantes principales. V1 et V2 représentent les % de variance de chaque CP.**

Cependant, le principal problème lors de l'application de l'ACP à un ensemble de données est l'évaluation du nombre suffisant de CPs. L'utilisation de plusieurs CPs peut conduire à un sur-apprentissage des données, ce qui entraîne une très bonne séparation qui ne reflète toutefois pas la justesse des résultats. D'autre part, l'application de suffisamment de CPs est nécessaire pour représenter le maximum de variance des données analysées.

Le pourcentage de la variance expliquée peut être un bon moyen pour décider du nombre de CPs suffisant pour l'analyse. Il faut noter aussi que l'ACP sert de fondement aux autres méthodes statistiques plus développées. En effet, cette méthode a été largement utilisée dans le domaine de la détection du cancer afin d'expliquer la structure de variance-covariance d'un ensemble de variables à travers quelques combinaisons linéaires de ces derniers.

### **2.3. Séparateurs à vaste marges (SVM)**

Introduite par Vladimir Vapnik en 1995 (68), la SVM est une méthode de classification supervisée. Elle a été adoptée pour sa capacité à travailler avec des données de grandes dimensions.

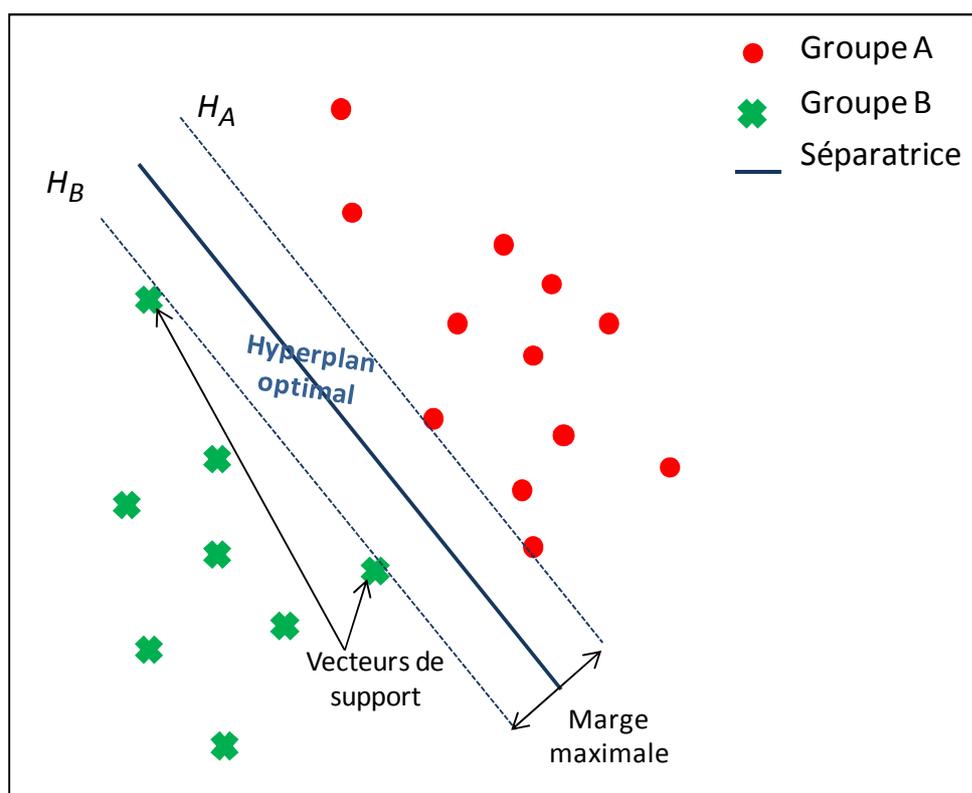
## Chapitre 1 : Introduction générale

Les SVMs sont dans leur origine utilisées pour la classification binaire et la régression. Aujourd'hui, elles sont utilisées dans différents domaines de recherche et d'ingénierie tel que le marketing, la biologie, la reconnaissance de caractères manuscrits, de visages humains et pour le diagnostic médical.

En supposant qu'il existe un hyperplan permettant de séparer les données du groupe A des données du groupe B, nous n'allons pas nous contenter d'en trouver un, mais nous allons en plus chercher parmi ceux-ci celui qui passe « au milieu » des points des deux classes de données. Intuitivement, cela revient à chercher l'hyperplan le « plus sûr ». En effet, supposons qu'une donnée n'ait pas été décrite parfaitement, une petite variation ne modifiera pas sa classification si sa distance à l'hyperplan est grande.

Cela revient donc à chercher un hyperplan dont la distance minimale aux échantillons d'apprentissage est maximale (Figure 17). On appelle cette distance « marge ». Comme on cherche à maximiser cette dernière, on parlera de méthode des séparateurs à vaste marge.

Exemple de données linéairement séparables:



**Figure 17: Schéma illustrant le principe de la SVM pour des données linéairement séparables.**

## *Chapitre 1 : Introduction générale*

---

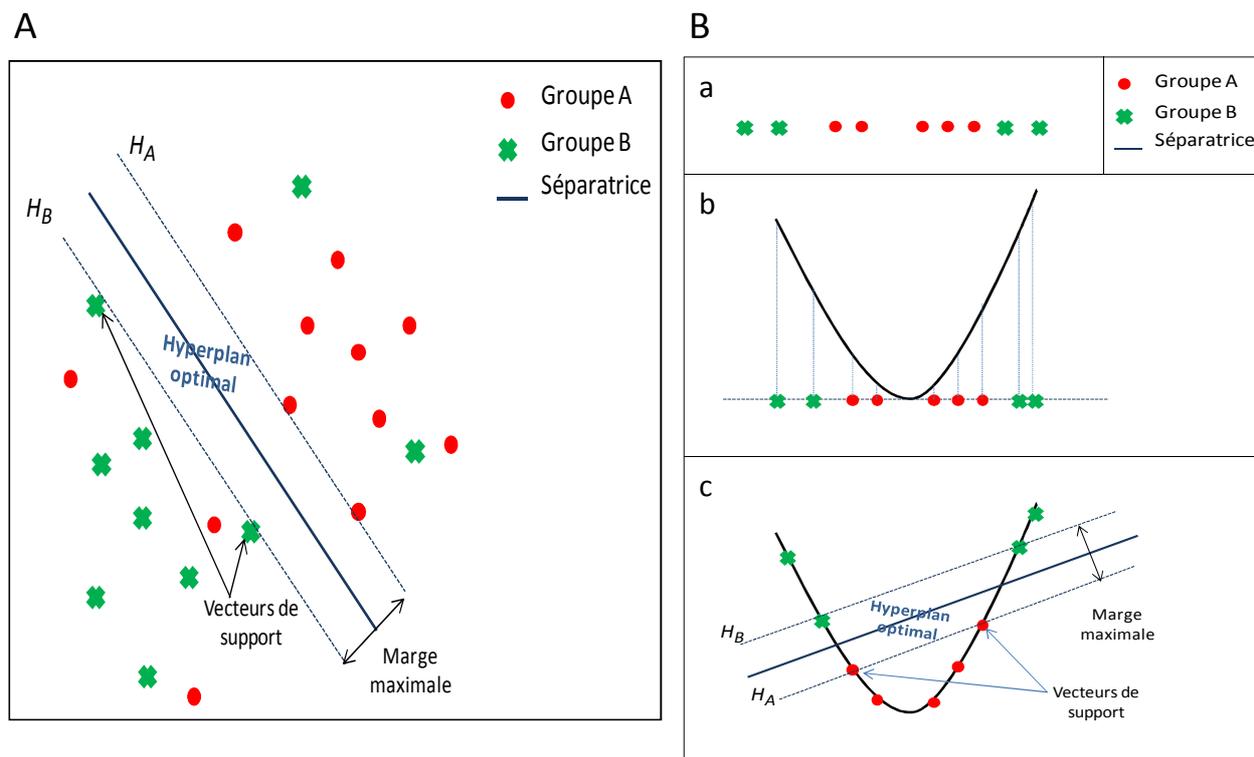
Il existe parmi les modèles des SVM, le cas des données linéairement séparables et le cas des données non linéairement séparables. Le premier cas est le plus simple et permet de trouver facilement un classifieur linéaire. Cependant, dans la plupart des données il n'y a pas de séparation linéaire possible et donc l'utilisation des SVMs non linéaires est très recommandée.

Dans le cas des données non séparables linéairement, nous obtenons trois sortes de données :

- les données mal classées
- les données correspondant aux vecteurs supports (bien classées)
- les données bien classées, qui ne sont pas des vecteurs supports.

Dans ce cas de figure, la SVM permet de trouver une transformation de l'espace des données dans un autre espace dans lequel les données sont à nouveau linéairement séparables. Généralement, ce nouvel espace, appelé espace de re-description, a une dimension plus grande que l'espace des données initial; il peut même être de dimension infinie. En effet, plus la dimension de l'espace de re-description est importante, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les données est élevée. On a donc une transformation d'un problème de séparation non-linéaire dans l'espace de représentation initial en un problème de séparation linéaire dans un espace de re-description de plus grande dimension (Figure 18). En pratique, cette transformation est réalisée à l'aide de fonctions noyaux, ces fonctions peuvent avoir plusieurs formes : polynômiale, sigmoïdale ou gaussienne.

### Exemple de données non linéairement séparables:



Dans notre étude, nous avons utilisé la méthode SVM linéaire. Ce modèle a été construit sur 3 jeu de données : un jeu d'entraînement, un jeu de validation et un jeu test. C'est par cette répartition que l'on peut s'assurer de la capacité à généraliser le modèle.

#### Jeu d'entraînement :

Supposons que nous ayons une séquence d'observations décrites par des mesures sur un ensemble prédéfini d'attributs (par exemple : le poids, l'âge, stade de la maladie...), chacune de ces observations étant affectée à une classe  $C$  prise dans «  $C1, C2$  » (par exemple dans ce travail: patients cirrhotiques sans CHC, patients cirrhotiques avec CHC). À partir de l'échantillon d'apprentissage, nous cherchons à trouver les paramètres du modèle qui

## *Chapitre 1 : Introduction générale*

---

permettra de prédire la classe de nouvelles observations à l'avenir. Il s'agit d'une tâche d'apprentissage supervisé.

En effet, le but de ce modèle est de parvenir à classer correctement les futurs objets grâce à la connaissance apprise à partir d'un échantillon limité de données.

### Jeu de validation (interne) :

Pour ce set de données, l'utilisation du jeu de validation intègre de l'information dans l'algorithme d'apprentissage. En effet ce jeu de données représente une « validation interne » permettant de construire et d'optimiser le modèle statistique.

Dans le cas de notre étude, pour chaque composante principale, la méthode d'analyse SVM est répétée 10 fois pour le jeu d'entraînement et de validation. Dix résultats sont donc obtenus et nous en gardons le meilleur résultat en jeu de validation et son jeu d'entraînement correspondant.

- ⇒ Le jeu de validation influence d'une manière très importante le modèle d'où la qualification de « validation interne » et l'intérêt d'utiliser un set de données externe « jeu de test ou de validation externe » pour l'évaluation de la performance du modèle statistique adopté.

### Jeu de test (validation externe) :

Ce jeu de données, jamais utilisé dans l'apprentissage, permet l'évaluation finale des performances du modèle SVM.

En résumé : pour une base de données  $F$ , on entraîne sur le sous-ensemble d'entraînement, on optimise sur le sous-ensemble de validation, et on considère sa performance sur le sous-ensemble du test.

## **2.4. Validation croisée (Leave-one-out)**

Pour minimiser l'influence du choix du partitionnement de l'ensemble des exemples, la validation croisée subdivise l'ensemble d'entraînement initial en  $k$  sous ensemble disjoints  $E_1, E_2, \dots, E_k$  de même taille. L'entraînement et la validation sont effectués  $k$  fois. A l'itération «  $i$  » le sous-ensemble  $E_i$  est réservé pour le test et le reste des exemples sont utilisés pour entraîner le modèle. La précision finale du modèle est égale à la moyenne des  $k$  précisions de validation.

## *Chapitre 1 : Introduction générale*

La méthode « leave-one-out » (69) est un cas particulier de la validation croisée où  $k=N$ . A chaque itération, le modèle est entraîné sur  $N-1$  objets et testé sur l'objet exclu de l'entraînement. On obtient à la fin  $N$  précisions ; la précision du modèle est égale à leur moyenne.

### **2.5. Régression aux moindres carrés partiels (PLS)**

Une méthode statistique supervisée introduite par le statisticien suédois Herman Wold (70), et est une variante de l'ACP. Cette méthode permet de construire des modèles prédictifs lorsque les variables explicatives sont nombreuses et très corrélées. La PLS (Partial Least Squares Regression) a connu de grands succès dans le domaine de la chimie et la spectrométrie, où, souvent, les données répondent à ce type de propriétés.

Cette une méthode de modélisation qui permet d'identifier les relations entre plusieurs variables. Il y a toujours, d'une part, les variables explicatives (notées généralement  $A_1, A_2, \dots, A_p$ ), et les variables expliquées (notées généralement  $C_1, \dots, C_q$ ). Ces variables sont, dans une régression PLS, toutes étudiées sur les mêmes « individus ». On distingue la régression PLS univariée et la régression PLS multivariée. Dans le premier cas, la régression ne porte que sur une seule variable expliquée. Dans le second, il peut y avoir plusieurs variables (71).

Le principe de la PLS consiste à remplacer une matrice des données prédictives « X » contenant  $n$  lignes (individus) et  $m$  colonnes (variables), par une nouvelle matrice désignée par « W », avec le même nombre de lignes que « X », mais un nombre de colonnes  $h$  inférieur à  $m$  représentant des vecteurs de poids. Une matrice supplémentaire est nécessaire pour une description complète de la procédure de régression PLS ; la matrice « E » qui contient la partie non expliquée du score de « X ».

Une première étape d'étalonnage (calibration) est nécessaire pour la construction du modèle mathématique reliant les matrices X et W. Le modèle est construit à l'aide d'un ensemble d'échantillons à partir duquel les matrices X et W sont clairement connues. En effet, pour la calibration, un algorithme mathématique est paramétré afin qu'il puisse associer les variables à expliquer aux variables explicatives. Pour obtenir le modèle le plus représentatif des échantillons, Il est important que les valeurs à expliquer des données de calibration soient réparties de manière homogène entre les deux extrema des valeurs à expliquer de l'ensemble des échantillons.

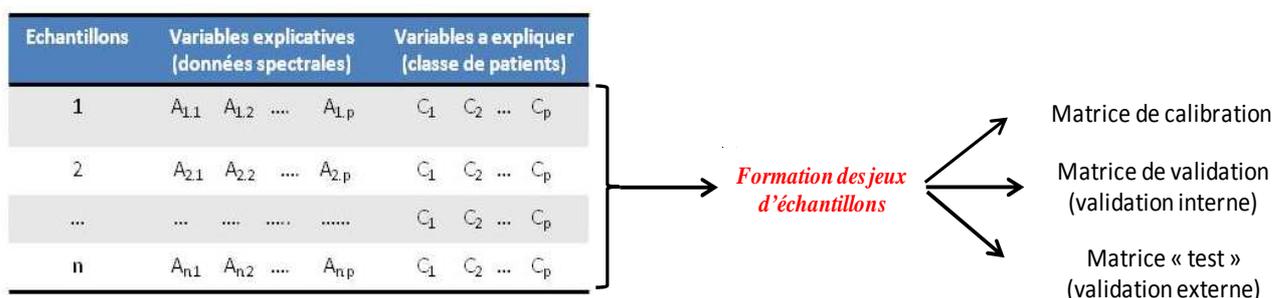
## *Chapitre 1 : Introduction générale*

Une fois développé, la précision du modèle doit être vérifiée à l'aide d'un nouvel ensemble de valeurs connues avant d'être en mesure de prédire les échantillons inconnus. C'est l'étape de validation.

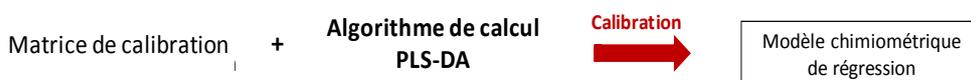
### PLS-analyse discriminante (PLS-DA)

La PLS-DA est une méthode de la toolbox PLS qui a été développée pour établir un modèle permettant de classer au mieux les individus dans les groupes prédéfinis en se basant sur leurs variables latentes. Le nombre optimal de variables latentes à retenir dans le modèle PLS-DA a également été déterminé par validation croisée de type "leave-one-out". Dans le cas de la PLS-DA, les variables ne sont plus quantitatives, mais qualitatives (variables binaires 0/1). Les coordonnées des individus sur les variables latentes du modèle PLS-DA ont ensuite été utilisées pour calculer la matrice de confusion. Une matrice de confusion est un outil de visualisation typiquement utilisé dans les analyses supervisées, servant à mesurer la qualité d'une discrimination. Chaque colonne de la matrice représente les classes prédites, alors que chaque rangée représente les classes observées.

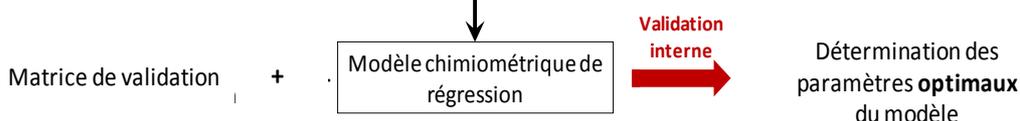
#### - Matrice des données spectrales



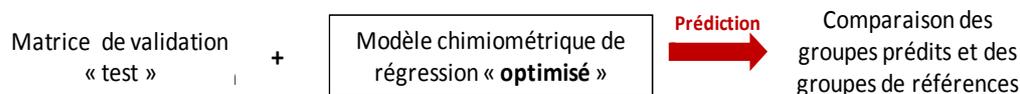
#### - Etape 1: calibration (jeu d'entraînement)



#### - Etape 2: validation (interne)



#### - Etape 3: test (validation externe)



**Figure 19 : Schéma représentant les étapes de la construction du modèle PLS-DA.**

## *Chapitre 1 : Introduction générale*

---

La Figure 19 reprend les différentes étapes suivies pour la détermination des paramètres du modèle PLS-DA :

- La calibration : permet de construire le modèle de régression, c'est à dire qu'un algorithme mathématique est paramétré afin qu'il puisse associer les variables à expliquer aux variables explicatives.
- La validation (interne) : permet de tester le pouvoir prédictif du modèle et de déterminer les paramètres optimaux pour ce dernier. Parmi ces paramètres, la sélection du nombre de dimensions (itérations) utilisé pour la construction du modèle. Ce critère est très important. En effet, un nombre de dimensions trop élevé peut induire un surentraînement « overfitting » suite à une modélisation du bruit. Cependant, un nombre trop faible de dimensions peut induire un sous-apprentissage « underfitting » et on obtient alors un modèle incomplet.
- La validation test : permet de tester le modèle avec un jeu de données externe jamais utilisé pour la construction du modèle un groupe. C'est une validation dite externe.

Dans les études menées dans le présent travail, et afin de comparer les différentes méthodes de classification (SVM, PLS-DA), nous avons utilisé les mêmes jeux d'échantillons de calibration, de validation et test.

## CHAPITRE 2 : MATÉRIELS ET MÉTHODES

## ***Chapitre 2 : Matériels et méthodes***

Ce projet fait partie d'une étude multicentrique « ProSpec » visant à identifier les biomarqueurs sériques du CHC chez les patients atteints de cirrhose selon deux approches: la protéomique et la spectroscopie vibrationnelle. ProSpec est une étude ancillaire de la grande étude CiRCE qui est une étude cas-témoins du Canceropole Grand Est comprenant six centres hospitaliers (Nancy, Reims, Besançon, Strasbourg, Metz, Dijon), dont le but est d'analyser les facteurs environnementaux, nutritionnels et métaboliques impliqués dans le carcinome hépatique chez les patients atteints de cirrhose. L'étude a été approuvée par le Comité de Protection des Personnes Est (agreement no. 2008/09) et par l'Agence Française de Sécurité Sanitaire des Produits de Santé (AFSSAPS; agreement no. 2008-A00023-52). Un consentement écrit a été approuvé par tous les patients de cette étude.

### **I. Recrutement des patients**

Les patients ont été recrutés dans les six centres cités ci-dessous. Pour les patients cirrhotiques de cette étude, la cirrhose a été diagnostiquée en s'appuyant sur les examens histologiques obtenus par biopsie du foie ou sur les différents signes cliniques combinés aux caractéristiques endoscopiques. L'absence de CHC était constatée par des techniques d'imagerie (échographie et/ou tomodensitométrie et/ou imagerie par résonance magnétique), et par un dosage de l'alpha-fœtoprotéine inférieure à 100 ng.ml<sup>-1</sup>. Un dépistage de CHC par échographie est prescrit aux patients cirrhotiques tous les six mois. Les témoins de cette étude n'ont pas développé de CHC au cours d'un an de suivi. En ce qui concerne le diagnostic du CHC, il a été fait selon les critères de l'Association européenne pour l'étude du foie.

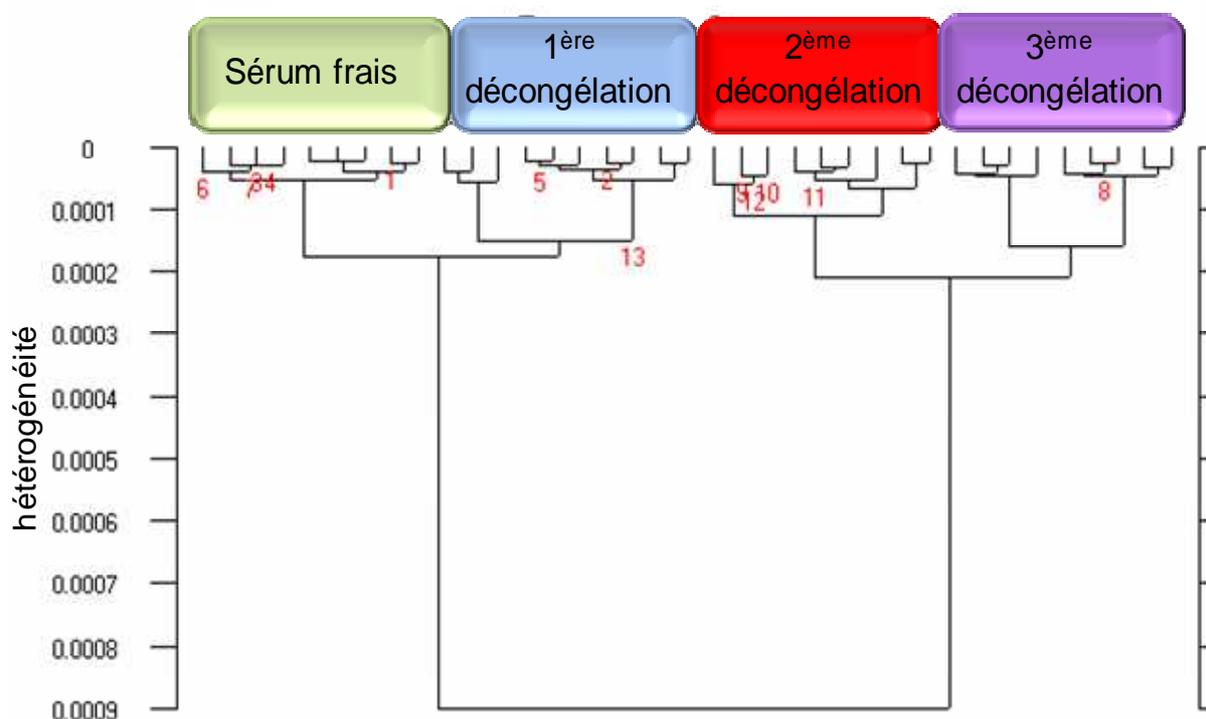
Pour l'étude pilote Raman, nous avons analysé 71 sérums de patients cirrhotiques avec et sans cancer. En ce qui concerne l'étude pilote IRTF, les sérums de 79 patients avec et sans CHC ont été recrutés pour l'analyse. Pour ces études, les échantillons utilisés ont été recrutés au centre hospitalier de Reims.

Pour l'étude multicentrique de validation IRTF, elle a été menée dans une cohorte de 817 sérums de patients cirrhotiques avec et sans CHC recrutés dans les cinq centres hospitalo-universitaires de la région Grand Est de France (Nancy, Reims, Strasbourg, Metz, Dijon).

### II. Microspectroscopie Raman appliquée au sérum

#### 1. Préparation des sérums

Les sérums utilisés dans cette étude ont été congelés à  $-80^{\circ}\text{C}$  après la collecte dans les différents centres hospitaliers. En effet, dans une étude antérieure réalisée au sein de notre équipe (Scaglia, 2009), il a été montré que le sérum peut être modifié pendant le stockage, suite à plusieurs congélations et décongélations pour d'autres expériences. Dans cette étude nous avons comparé des sérums frais (jamais congelés) à d'autres sérums ayant déjà subi une à plusieurs décongélations. La Figure 20 montre l'analyse en classification hiérarchique des spectres sériques d'un patient en passant par 3 cycles de congélation-décongélation. Les spectres obtenus se séparent en deux clusters, l'un regroupant les spectres du sérum à l'état frais et les spectres de la première décongélation, et le second regroupant ceux de la deuxième et de la troisième décongélation. Les résultats obtenus HCA, montrent qu'un sérum congelé et décongelé une seule fois est très proche d'un sérum frais (distance spectrale très faible). Pour les spectres obtenus après le deuxième et le troisième cycle de congélation-décongélation, les distances spectrales sont plus importantes, reflétant ainsi des altérations et des modifications significatives des composés du sérum.



**Figure 20 : Classification hiérarchique du sérum d'un patient montrant l'effet de la congélation-décongélation sur les spectres IRTF. (Scaglia 2009)**

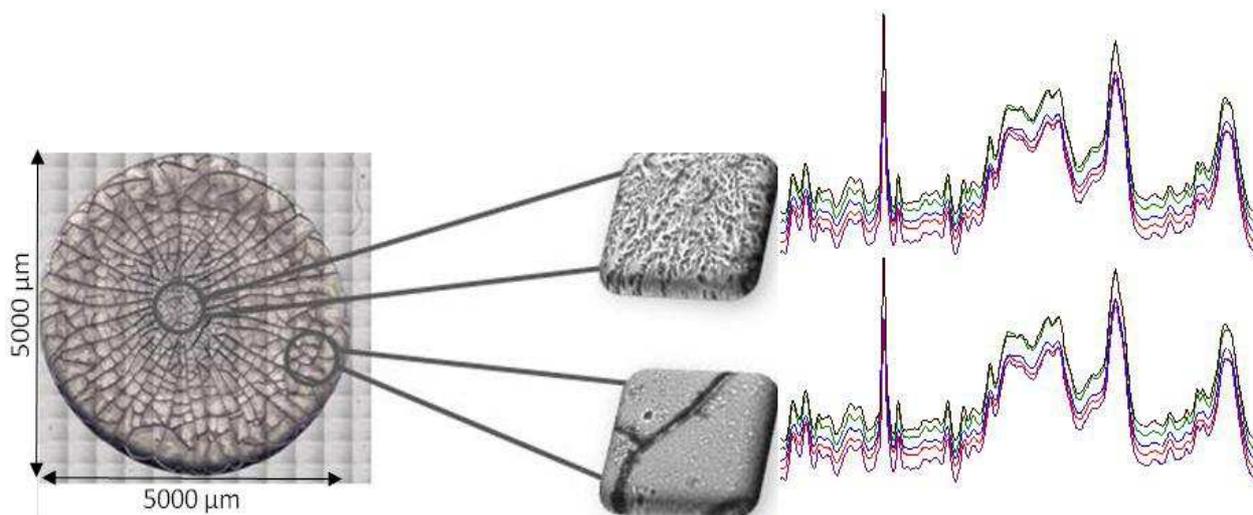
## Chapitre 2 : Matériels et méthodes

Dans toutes les études réalisées dans ce travail, nous avons utilisé un sérum ayant subi qu'une « 1<sup>ère</sup> décongélation » afin de s'approcher au maximum de l'état frais et donc des conditions physiologiques.

Pour l'étude en microspectroscopie Raman, deux préparations de sérum, en goutte séchée et lyophilisée, ont été analysées et les résultats comparés.

### 1.1. Sérum en goutte séchée

Pour la première série d'expériences, les sérums stockés à  $-80^{\circ}\text{C}$  ont été décongelés à température ambiante pendant 30 min. Dix microlitres de chaque échantillon ont été déposés sur une fenêtre  $\text{CaF}_2$  (compatible Raman) puis laissés sécher à température ambiante pendant 30 min. Dix spectres ont été enregistrés : 5 spectres au centre de la goutte séchée et 5 spectres à sa périphérie (Figure 21).



**Figure 21 : Image visible d'une goutte séchée de sérum d'un patient montrant les zones d'acquisition : centre et périphérie**

### 1.2. Sérum lyophilisé

Pour la deuxième série d'expériences, un aliquot de  $200\ \mu\text{l}$  de chaque échantillon de sérum a été lyophilisé pendant 24 h, sous vide, en utilisant un lyophilisateur équipé d'une pompe à palettes rotatives RV8 (Edwards, SuperModulyo 1,5 litre, Thermo Fisher Scientific). Les sérums lyophilisés ont été stockés à  $-80^{\circ}\text{C}$  jusqu'à l'utilisation. Pour l'analyse Raman, les sérums lyophilisés ont été décongelés à température ambiante pendant 10 minutes, puis aplatis sur une fenêtre  $\text{CaF}_2$  (compatible Raman) pour obtenir une surface plate avec une

## *Chapitre 2 : Matériels et méthodes*

épaisseur uniforme. Les spectres Raman ont été enregistrés en choisissant, aléatoirement, cinq points d'acquisition sur la surface du lyophilisat.

### **2. Paramètres d'acquisitions**

Les spectres Raman ont été enregistrés avec un spectromètre LabRam (Horiba Jobin Yvon SAS, Villeneuve d'Ascq, France) équipé d'un microscope Olympus BX40 et un objectif long focal 100x (Olympus, NA 0.9, Rungis, France) (Figure 22). Pour l'excitation des échantillons, une diode laser proche-infrarouge a été utilisée (Toptica photonique, Munich, Allemagne) générant une puissance de 300 mW à une longueur d'onde de 785 nm. La taille du spot laser était de 2  $\mu\text{m}$  et sa puissance à l'échantillon était de 60 mW (évite la photodégradation). Une caméra CCD (1024 x 256 pixels, Andor Tech., Belfast, Royaume-Uni) a été utilisée pour détecter le signal Raman. Tous les spectres ont été enregistrés en mode point avec un temps d'acquisition de 2 x 20s dans la région spectrale 600 - 1800  $\text{cm}^{-1}$ . La résolution spectrale était de 6  $\text{cm}^{-1}$ . Tous les spectres ont été acquis à l'aide du logiciel Labspec (Horiba Scientific Jobin Yvon; Lille; France).



**Figure 22 : Micro-spectromètre Raman LabRam 785 nm.**

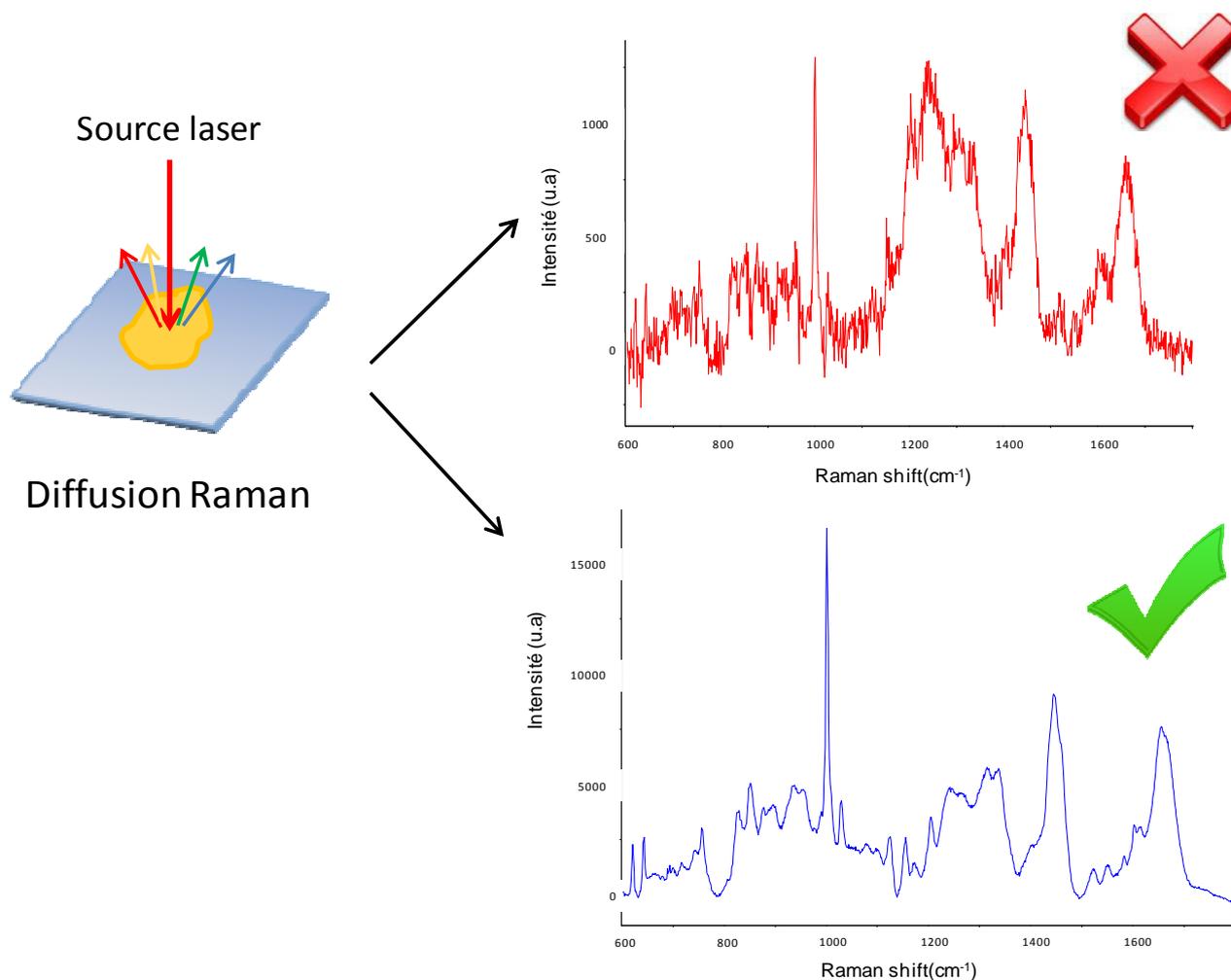
### **3. Prétraitement des spectres Raman**

#### **3.1. Le test de qualité**

L'un des avantages de la spectroscopie Raman est qu'elle peut être faite sur des échantillons liquides et est donc pas gênée par la présence d'eau dans le milieu. Le critère le plus important donc, dans la présélection des spectres Raman est le rapport S/B. en effet nous

## Chapitre 2 : Matériels et méthodes

avons mis en place un algorithme « test qualité » dans le logiciel Matlab qui permet de sélectionner les spectres bruités selon un seuil de 50 (seuil utilisé en biologie). Ce rapport est calculé sur la région 990-1010  $\text{cm}^{-1}$  comportant la bande de la phénylalanine et la région 1720-1800  $\text{cm}^{-1}$  ne comportant aucune signature moléculaire et représente le signal du bruit de fond uniquement. La Figure 23 montre un exemple de spectre Raman bruité rejeté par le test qualité (spectre rouge) et un exemple de spectre remplissant les critères du test (spectre bleu).



**Figure 23 : Exemple de l'application du test qualité à spectre Raman de sérum.**

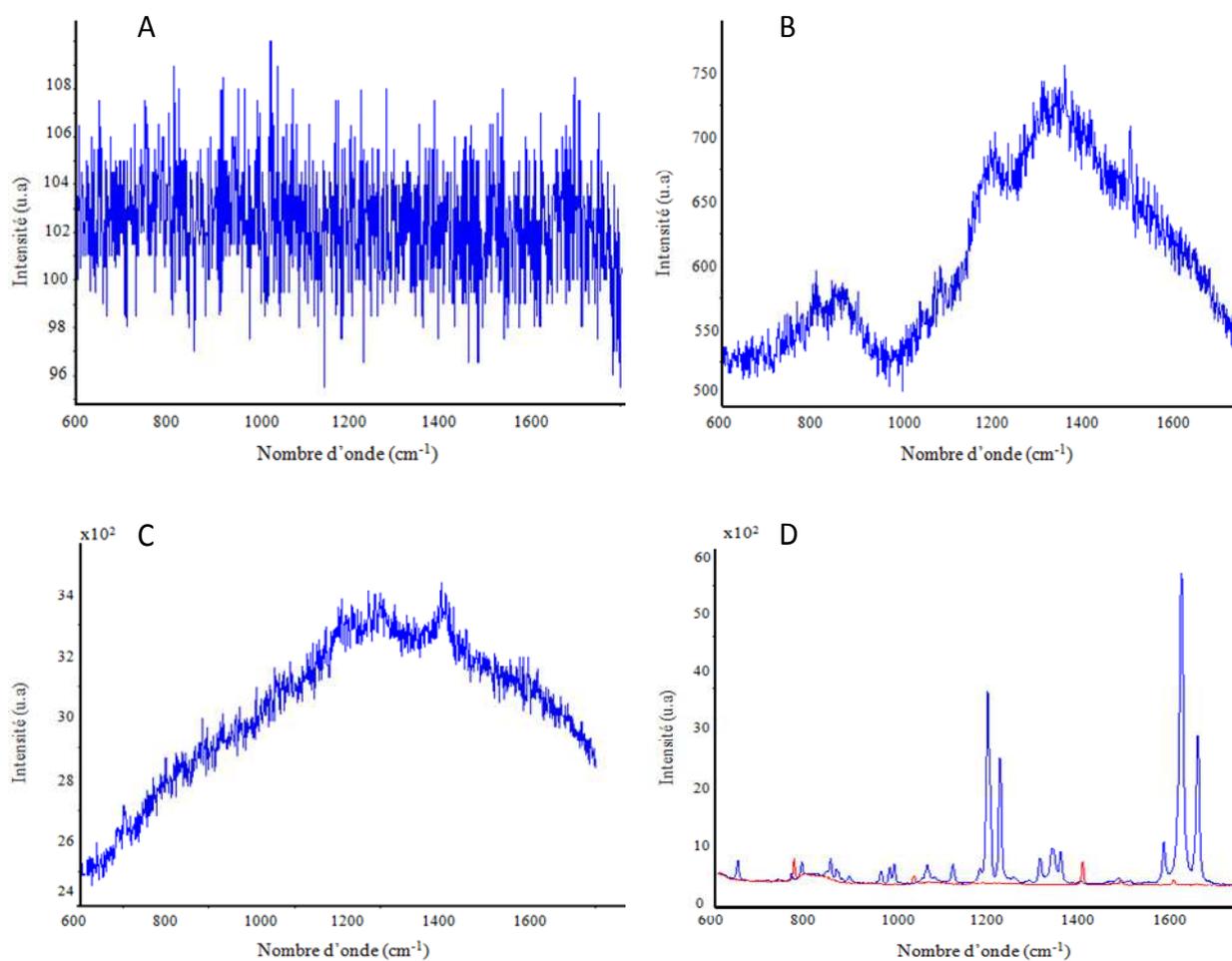
### 3.2. Les prétraitements des spectres Raman

Après le test qualité les spectres sont prétraités afin d'améliorer la qualité du signal et d'en homogénéiser les intensités sans perdre l'information spectrale de l'échantillon étudié.

Dans un premier temps, les données Raman sont corrigées de la réponse instrumentale (courant noir et réponse du détecteur). Pour cela, des mesures de la réponse du détecteur et du

## Chapitre 2 : Matériels et méthodes

courant noir ont été enregistrées au début des acquisitions dans les mêmes conditions expérimentales que l'échantillon biologique. Ensuite, l'intensité relative des spectres a été calibrée avec le NIST qui est un matériel de référence standard spécifique à la longueur d'onde utilisée (785 nm dans notre étude). Le décalage ou «shift» du spectre Raman a été ensuite corrigé en utilisant des spectres référence de composés chimiques bien définis : le 1,4bis (2-méthylstyryl) benzène et le naphthalène. Les pics de ces derniers sont comparés à leurs spectres de référence, ce qui permettra de recalibrer les spectres des échantillons. La Figure 24 montre un exemple des différents signaux de calibration obtenus.



**Figure 24 : Exemple des acquisitions enregistrées de la calibration Raman.**

A) Courant noir. B) Réponse du détecteur. C) NIST. D) Produit chimique de calibration.

Dans un deuxième temps, les spectres Raman ont été prétraités et pour ceci, un lissage avec un polynôme d'ordre 3 et une largeur de fenêtre de 11 points en utilisant l'algorithme Savitsky-Golay a été appliqué. Ensuite, tous les spectres ont été corrigés de la ligne de base avec un polynôme d'ordre 4. Enfin, une normalisation vectorielle suivi d'une correction

« offset » ont été appliquées afin d'uniformiser les intensités et ainsi pouvoir comparer les spectres entre eux.

### **4. Analyses statistiques**

Toutes les données démographiques, cliniques et biologiques ont été évaluées pour les patients de cette étude. Les données quantitatives ont été exprimées en moyenne  $\pm$  écartype et ont été comparées en utilisant le test de Student. Les données qualitatives ont été exprimées en pourcentage et ont été comparées en utilisant le test de Khi 2. Une valeur de p inférieure à 0,05 était considérée comme significative.

Les données Raman ont été analysées par l'analyse en composante principale. Cette méthode multivariée permet de réduire l'information spectrale et donc de sélectionner les CPs les plus discriminantes. Un test statistique a été appliqué sur les scores de chaque CP. La distribution des scores étant inconnue, un test statistique non paramétrique, appelé test Kruskal-Wallis a été utilisé et a permis de réorganiser les CPs dans l'ordre croissant de leur valeur de p. La première CP étant la plus discriminante entre les deux groupes de patients. Après la détermination des CPs les plus discriminants, la méthode de séparation à vaste marges (SVM), a été réalisée en utilisant la bibliothèque LIBSVM (72). Seuls les CPs expliquant 99,99% de la variance totale ont été utilisées pour optimiser le modèle. Les autres CPs (classées dernières) représentant le bruit de fond ont été supprimées. La sensibilité et la spécificité du classifieur ont été calculées pour tous les spectres. Ce calcul a été effectué sur la moyenne, la médiane, et sur tous les spectres de chaque patient.

A noter que pour notre étude pilote, vu le nombre restreint de patients analysés, le modèle SVM a été combiné à une validation croisée.

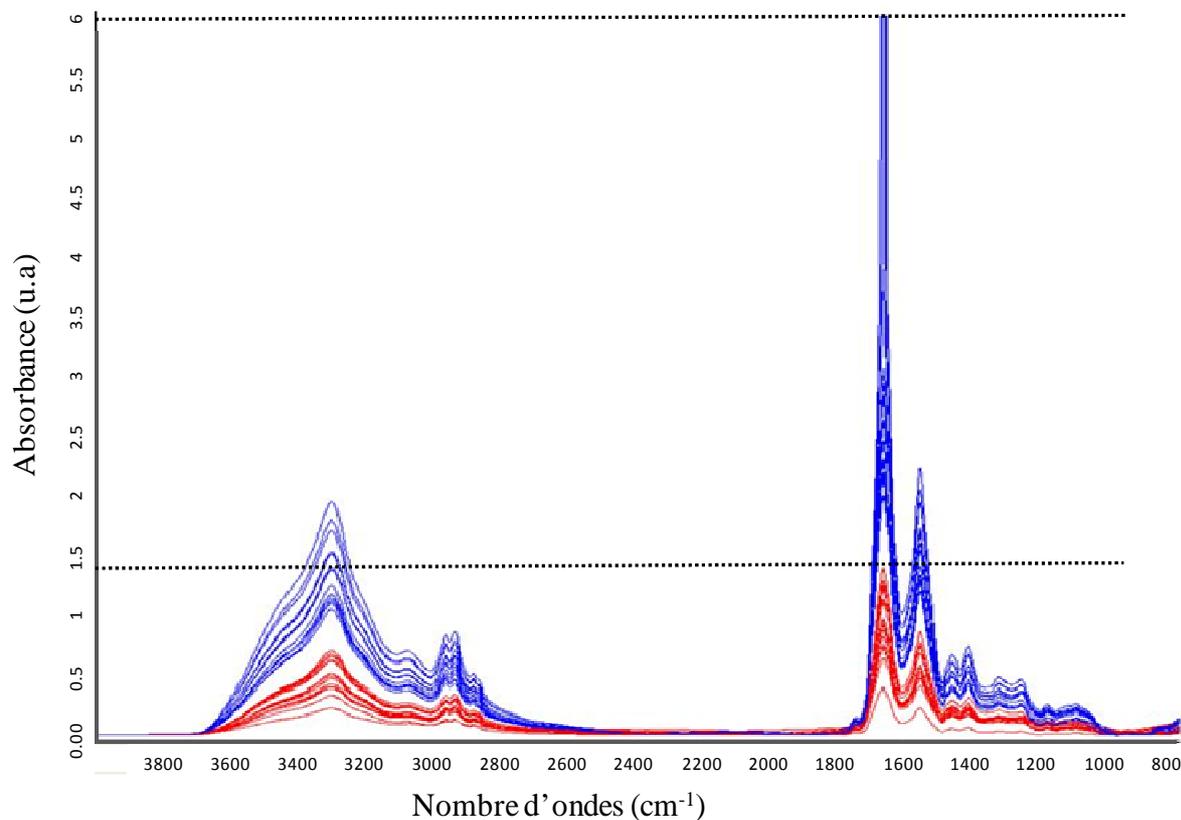
## **III. Spectroscopie IRTF à haut débit appliquée au sérum**

### **1. Préparation des sérums**

Les échantillons de sérum utilisés dans cette étude ont été décongelés pendant 30 min à température ambiante, puis dilués au 1/3 dans du sérum physiologique afin d'être conforme au critère d'absorbance du test qualité.

La Figure 25 présente des spectres infrarouge de sérum pur et d'un sérum dilué au 1/3. Les spectres représentés en bleu sont saturés au niveau du pic d'absorbance maximale dans la région spectrale  $1600-1700\text{ cm}^{-1}$  correspondant aux protéines et donc ne répondent pas au test

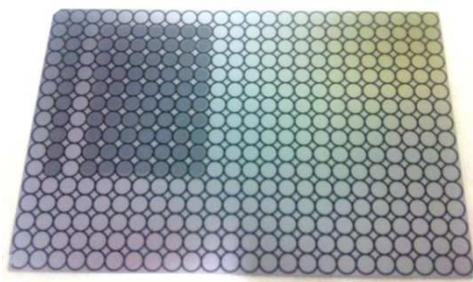
qualité contrairement aux spectres rouge (dilué au tiers) dont l'intensité d'absorbance est comprise entre 0,35 et 1,8.



**Figure 25 : Spectres infrarouge d'un sérum pur (bleu) et dilué au tiers (rouge).**

Dix dépôts de 5  $\mu$ l sont ensuite déposés sur une plaque en silicium 384 puits adaptée à l'analyse infrarouge à haut débit (Figure 26). L'analyse de dix dépôts par échantillon permet la détection des valeurs aberrantes qui pourraient être dues à la plaque de silicium et/ou à des variations dues aux dépôts (reproductibilité intra-plaque et des répliqués instrumentaux).

Les sérums sont ensuite déshydratés pendant 1h à température ambiante. Le séchage des puits est une étape indispensable en spectroscopie IRTF car elle permet de réduire la contribution de l'eau dans les spectres.



**Figure 26 : Sérums déposés sur une plaque silicium de 384 puits pour l'analyse IRTF à haut débit.**

### 2. Paramètres d'acquisitions

Après la phase de séchage, la plaque de silicium est placée dans le module à haut débit (HTS-XT, Bruker Optics GmbH, Ettlingen, Allemagne) couplé au spectromètre IRTF (Tenseur 27, Bruker Optics GmbH, Ettlingen, Allemagne) pour l'analyse des échantillons sériques (Figure 27).



**Figure 27 : Le spectromètre IRTF Tensor 27 couplé au module HTS-XT à haut débit.**

Les spectres ont été enregistrés en mode transmission. Les paramètres d'acquisitions étaient :

- Région spectrale : 400 - 4000  $\text{cm}^{-1}$
- Résolution spectrale : 4  $\text{cm}^{-1}$
- Facteur « zéro-filling » : 2
- Nombre de scans : 32 (environ 30s)

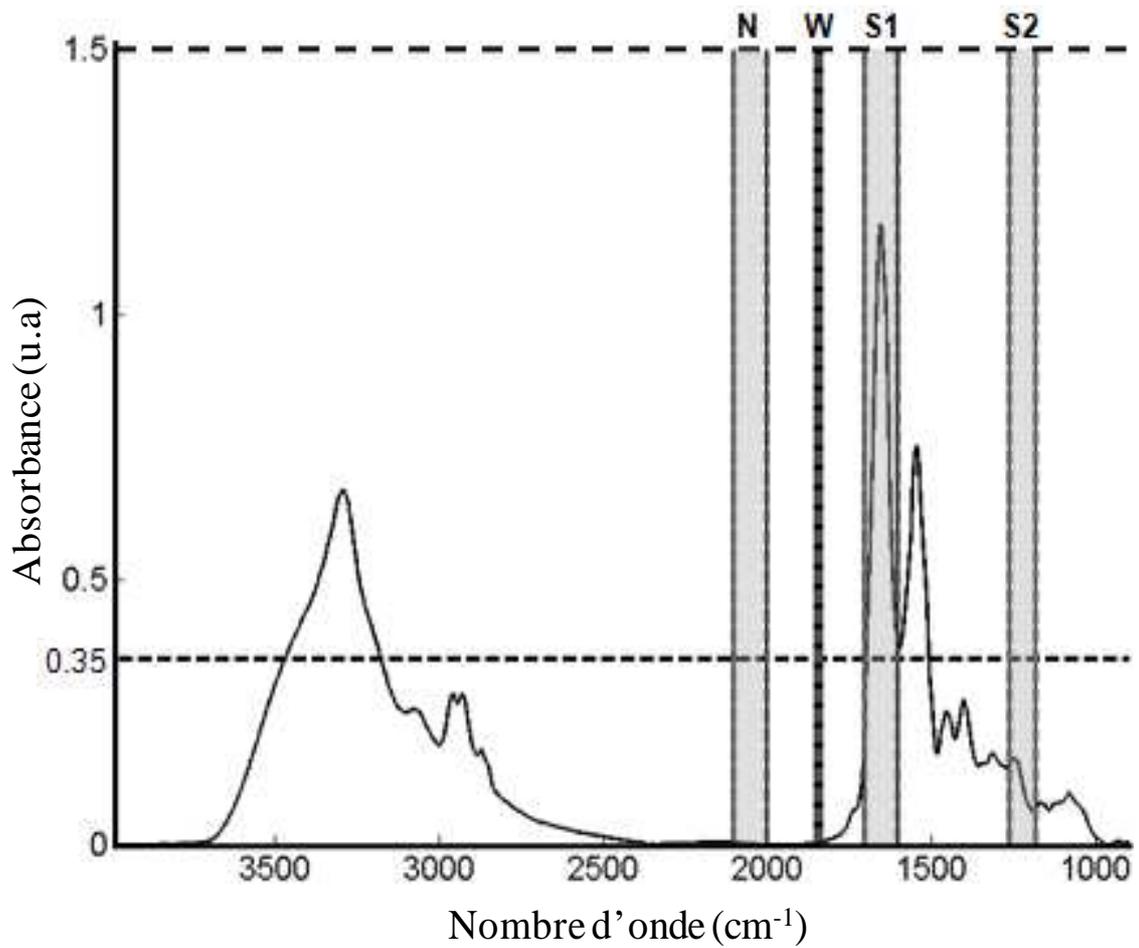
### 3. Prétraitement des spectres infrarouge

#### 3.1. Le test de qualité

Le test qualité permet d'évaluer la qualité des spectres en mesurant leur teneur en vapeur d'eau, leur rapport signal/bruit ainsi que leur intensité. En effet, La vapeur d'eau et le dioxyde de carbone (CO<sub>2</sub>) sont des composés atmosphériques qui absorbent fortement en IR et entraînent des pics indésirables dans la région spectrale 3400 cm<sup>-1</sup> et 1640 cm<sup>-1</sup> pour l'eau et 2349 cm<sup>-1</sup> et 667 cm<sup>-1</sup> pour le CO<sub>2</sub>.

Dans le cadre de notre étude IRTF, le test qualité est effectué à l'aide du logiciel OPUS (BRUKER Bruker Optics GmbH, Ettlingen, Allemagne). Il est adapté d'après les travaux de Helm et coll (73). Ce test repose sur la validation des critères suivants (Figure 28):

- L'absorbance maximale du spectre : elle doit être comprise entre 0,35 et 1,8. Cette bande correspond à la bande Amide I située ~1650 cm<sup>-1</sup> dans le cas des systèmes biologiques.
- Le rapport signal sur bruit (S/B) : il est calculé sur deux régions de signal important entre 1600 et 1700 cm<sup>-1</sup> (valeur S1), et entre 1178 et 1260 cm<sup>-1</sup> (valeur S2) et sur le bruit de fond du spectre (valeur N) en utilisant sa dérivée première. Plus le S/B est élevé, plus le spectre est de bonne qualité et sera donc conservé. Les spectres ont été retenus lorsque les rapports S1 / N et S2/ N sont au-dessus de 50 et 10 respectivement.
- La contenance en vapeur d'eau dans l'échantillon (valeur W) : ce paramètre a été évalué dans la région 1837 à 1847 cm<sup>-1</sup>. Les spectres ont été validé lorsque les rapports S1/W et S2/W sont supérieurs à 20 et à 4 respectivement.



**Figure 28 : Description des zones spectrales utilisées pour évaluer les critères du test de qualité appliqué aux spectres IRTF.**

La Figure 29 présente un exemple de spectre de sérum conforme (spectre noir) et non conforme (spectre rouge) aux critères du test qualité :

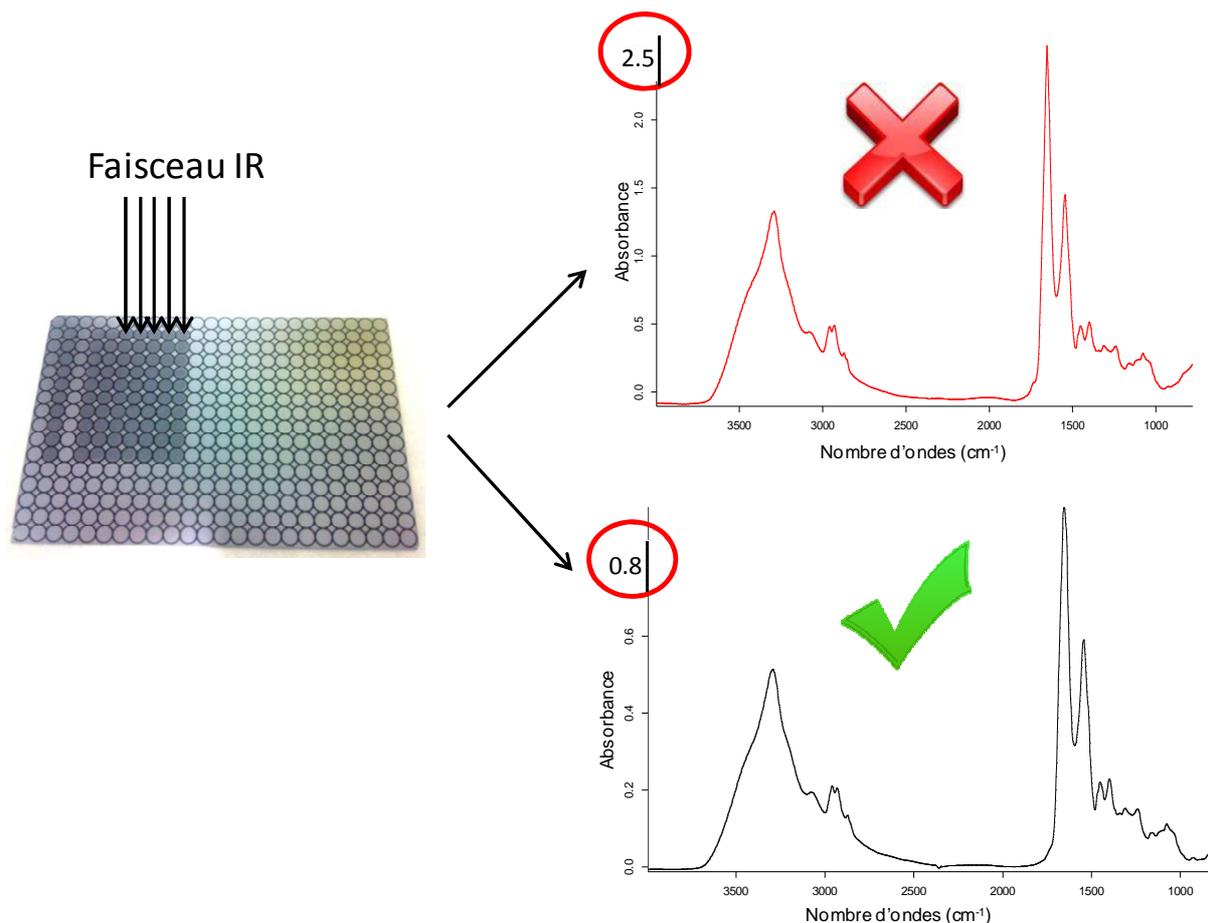


Figure 29 : Exemple de spectres IRTF conforme et non conforme du sérum.

### 3.2. Le prétraitement des données spectrales

Les spectres ayant été validés par le test de qualité, ont ensuite été prétraités en utilisant la méthode EMSC dans le logiciel Matlab (Matlab, The MathWorks, Inc, Natick, Massachusetts, Etats-Unis). Les prétraitements sont effectués sur la région 900 à 4000 cm<sup>-1</sup>, car la zone 400 - 900 cm<sup>-1</sup> n'apporte pas d'information et est très bruitée.

Le spectre moyen a été ensuite soustrait de chacun des spectres.

### 4. Analyses statistiques

Toutes les données démographiques, cliniques et biologiques ont été évaluées pour les patients de cette étude. Les données quantitatives ont été exprimées en moyenne  $\pm$  écart-type et ont été comparées en utilisant le test de Student ou test de  $t$ . Les données qualitatives ont été exprimées en pourcentage et ont été comparées en utilisant le test de Khi 2. Une valeur de  $p$  inférieure à 0,05 était considérée comme significative.

## *Chapitre 2 : Matériels et méthodes*

---

Les spectres ont été analysés par ACP afin de réduire les données spectrales et sélectionner les CPs les plus discriminantes, un test statistique a été appliqué sur les scores de chaque CP. Le test Kruskal-Wallis a été utilisé afin de réorganiser les CPs dans l'ordre croissant de leur valeur de p, c'est-à-dire de la CP la plus discriminante à la moins discriminante entre les deux groupes de patients.

Après la détermination des CPs les plus discriminants, la SVM a été appliquée sur les données. Seuls les CPs expliquant 99,99% de la variance totale ont été utilisées dans ce classifieur. La sensibilité et la spécificité ont été calculées pour la moyenne, la médiane, et tous les spectres de chaque patient.

## CHAPITRE 3 : RÉSULTATS ET DISCUSSION

### **Partie 3.1 : Etude pilote Raman**

#### **L'apport de la microspectroscopie Raman pour le diagnostic du carcinome hépatocellulaire chez les patients cirrhotiques.**

##### Contexte de l'étude

Le CHC est le sixième cancer le plus fréquent et la troisième cause de mortalité par cancer dans le monde (5). Son pronostic est très péjoratif en raison du diagnostic souvent tardif de la tumeur, à un stade qui ne permet pas d'envisager un traitement à visée curative. Les programmes de dépistage chez les patients cirrhotiques qui représentent la population à haut risque, reposent actuellement sur l'échographie du foie répétée tous les 6 mois (3). Les résultats obtenus par cette méthode de dépistage sont insuffisants. L'identification de biomarqueurs sériques permettant d'obtenir un diagnostic de CHC à un stade précoce est reconnue comme un objectif majeur pour améliorer le pronostic de cette tumeur.

Le potentiel de la spectroscopie Raman appliquée au sérum pour le diagnostic de CHC n'a jamais été évalué auparavant. Dans ce chapitre, l'hypothèse de travail était que les spectres Raman des sérums de patients cirrhotiques atteints de CHC présentent des caractéristiques spectrales (qualitatives et quantitatives) qui peuvent être utilisées pour prédire la présence d'un cancer.

⇒ **Le développement de cette méthode, non invasive, d'aide au diagnostic du carcinome hépatocellulaire serait donc une avancée significative.**

Nous avons réalisé une étude pilote dont le but était donc de déterminer si l'analyse du profil spectral Raman du sérum reflétant sa composition moléculaire permettait de différencier les sérums de patients avec ou sans CHC.

##### Patients et méthodes

###### *Préparations des échantillons*

Cette étude fait partie de l'étude multicentrique « ProSpec » visant à identifier des biomarqueurs sériques de CHC chez les patients atteints de cirrhose en utilisant deux approches: la protéomique et la spectroscopie vibrationnelle. Cette étude prospective est une étude ancillaire du projet CirCE, qui est une étude cas-témoins du canceropôle de la région

### *Chapitre 3 : Résultats et discussion*

Grand Est de la France visant à analyser le rôle des facteurs environnementaux, nutritionnels et métaboliques impliqués dans la carcinogenèse hépatique chez les patients cirrhotiques. Ce travail a été approuvé par le Comité de Protection des Personnes Est (convention n° 2008/09) et par l'Agence Française de Sécurité Sanitaire des Produits de Santé (AFSSAPS; convention n° 2008-A00023-52). Un consentement signé a été obtenu de tous les patients.

Des prélèvements sanguins ont été obtenus dans deux groupes de patients: des patients cirrhotiques atteints de CHC (n = 37) et des patients cirrhotiques sans CHC (n = 34). Après centrifugation, le sérum a été isolé et stocké à -80°C. Le diagnostic de cirrhose reposait sur l'examen histopathologique d'une biopsie hépatique ou sur l'association de signes clinico-biologiques d'insuffisance hépatocellulaire ou d'hypertension portale et/ou de signes endoscopiques d'hypertension portale et/ou et de caractéristiques d'imagerie de cirrhose. L'absence de CHC à l'inclusion était affirmée par des techniques d'imagerie (échographie et/ou tomodensitométrie et/ou imagerie par résonance magnétique), et une concentration sérique d'alpha-fœtoprotéine inférieure à 100 ng/mL. Une surveillance des patients cirrhotiques sans CHC a été faite tous les six mois. Aucun de ces patients n'a développé un cancer pendant la durée du suivi (un an). Le diagnostic de CHC a été fait selon les critères définis par l'Association Européenne pour l'Etude du Foie (3). Les critères d'exclusion étaient l'âge inférieur à 35 ans, l'infection par le VIH, les cancers extra-hépatiques et les maladies psychiatriques qui n'ont pas permis l'obtention d'un consentement éclairé.

Deux préparations de sérum ont été utilisées dans cette étude :

- Dans la première série d'expériences, les sérums ont été décongelés et 10 µl de chaque échantillon ont été déposés sur une fenêtre CaF<sub>2</sub> et séchés à l'air libre (température ambiante) pendant 30 min. Dix spectres ont été enregistrés, 5 au centre de la goutte séchée et 5 à sa périphérie.
- Dans la 2<sup>ème</sup> série d'expériences, un aliquot de 200 µl de sérum de chaque patient a été lyophilisé pendant 24 h sous vide dans un lyophilisateur équipé d'une pompe RV8 (Edwards, Supermodulyo 1,5 L, Thermo Fisher Scientific). Les produits lyophilisés ont été stockés à -80°C. Pour l'analyse Raman, les sérums lyophilisés ont été décongelés à température ambiante pendant 10 minutes, et ensuite pressée sur une fenêtre CaF<sub>2</sub> pour obtenir une surface plate d'une épaisseur uniforme. Cinq spectres

### *Chapitre 3 : Résultats et discussion*

---

ont été enregistrés sur des sites choisis de façon aléatoire sur la surface de l'échantillon.

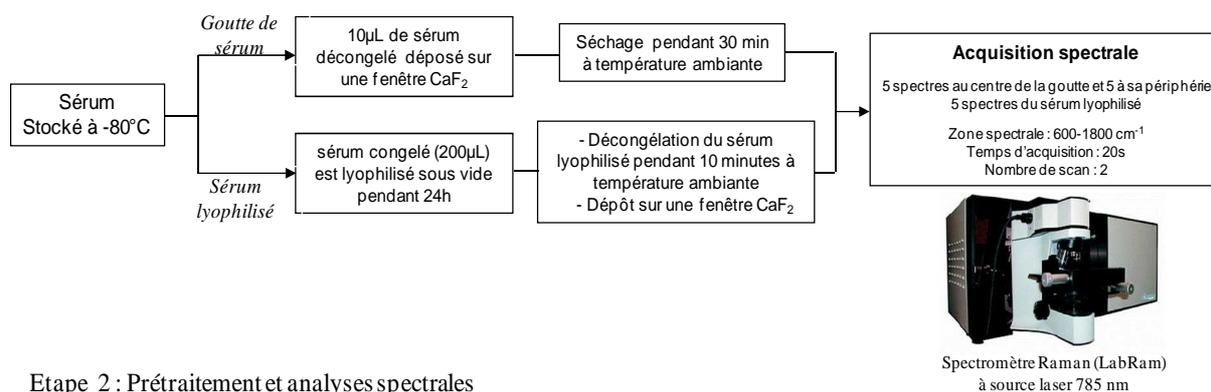
#### *Acquisition des spectres Raman*

Pour les deux préparations de sérum, Les spectres Raman ont été enregistrés avec un microspectromètre LabRam (Horiba Jobin Yvon SAS, Villeneuve d'Ascq, France) équipé d'un microscope Olympus BX40 et un objectif 100x longue distance de travail (Olympus, NA 0.9, Rungis, France). La source excitatrice utilisée était une diode laser proche infrarouge à 785 nm (Toptica Photonics, Munich, Allemagne) générant une puissance de 300 mW. La puissance du laser sur l'échantillon biologique était de 60 mW et la taille du spot laser était de 2  $\mu\text{m}$ . Pour détecter le signal Raman, une caméra CCD (1024x256 pixels, Andor Tech., Belfast, Royaume-Uni) était utilisée. Tous les spectres ont été enregistrés dans le mode point. Le temps d'acquisition était 2x20s sur la zone spectrale s'étalant de 600 à 1800  $\text{cm}^{-1}$  avec une résolution spectrale de 6  $\text{cm}^{-1}$ . La Figure 30 décrit les différentes préparations d'échantillons et les paramètres d'acquisition Raman correspondants.

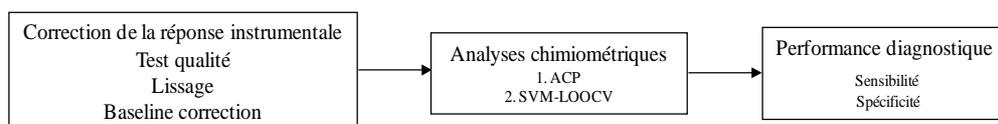
Avant chaque utilisation, le spectromètre Raman était calibré à l'aide d'un échantillon de silicium émettant un signal Raman à 520  $\text{cm}^{-1}$  et de deux composés de référence (1,4 bis-(2-méthylstyril) benzène et de naphtalène). L'intensité relative des spectres était corrigée par rapport à un étalon NIST pour 785 nm (matériau standard de référence # 2241; NIST, Gaithersburg, MD, USA). Toutes les acquisitions spectrales étaient enregistrées avec le logiciel LabSpec 4 (Horiba Jobin Yvon scientifique, Lille, France).

## Chapitre 3 : Résultats et discussion

Etape 1 : Préparation des échantillons et acquisition spectrale

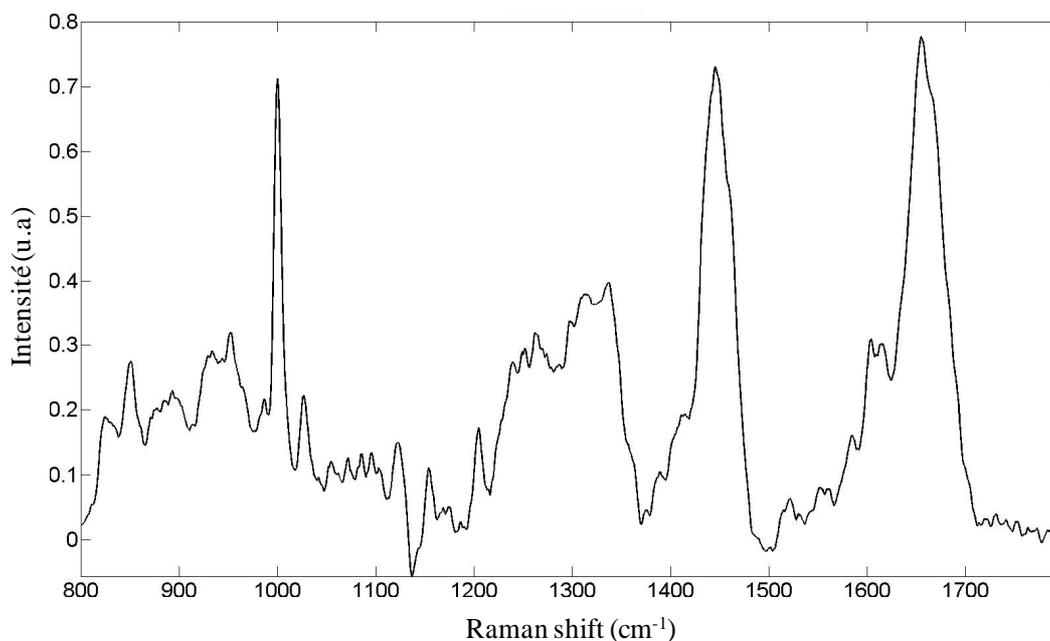


Etape 2 : Prétraitement et analyses spectrales



**Figure 30: Schéma décrivant de la procédure expérimentale utilisée pour la préparation des échantillons et l'acquisition spectrale.**

La figure 31 montre un exemple de spectre Raman de sérum humain. Les attributions des principaux pics sont citées dans le Tableau 2. Ces attributions sont tirées des travaux de Pichardo-Molina (61).



**Figure 31 : Spectre Raman du sérum humain.**

## *Chapitre 3 : Résultats et discussion*

**Tableau 2:** Attribution des pics Raman du sérum humain.

<i>Nombre d'onde (cm<sup>-1</sup>)</i>	<i>Attribution de bandes</i>
1658	Protéines (amide I)
1609	Tyr, Phe
1582	Protéines, Tyr, Arg, Adénine
1557	Protéine (Amide II)
1523	Betacarotène
1446	Protéines, Phospholipides
1340	Trp, Adénine, phospholipides
1311	Adénine
1271	Phospholipides, Amide III
1236	Amide III
1176	Trp, Phe
1155	Betacarotène
1126	Protéines, phospholipides
1103	Phe
1082	Phospholipides
1061	Phe
1032	Phe
1003	Phe
938	Protéines
876	Trp
852	Tyr
757	Protéines
741	Phospholipides
714	Polysaccharides

Arg: arginine, Phe: phenylalanine, Trp: Tryptophane.

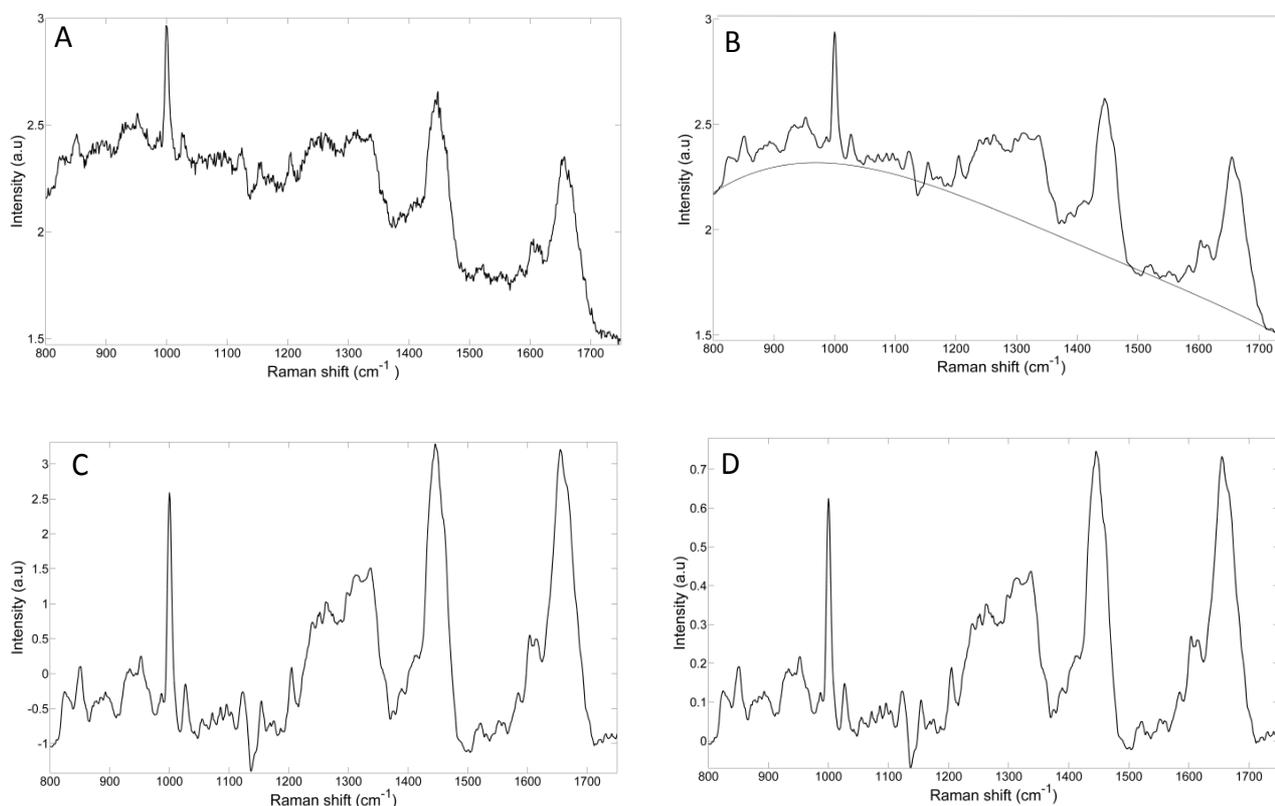
### ***Prétraitement des données Raman***

Tous les prétraitements ont été réalisés à l'aide du logiciel MatLab (version 7.6.0, MathWorks, Natick, Mass, USA). Les spectres Raman ont été corrigés de la réponse instrumentale : courant noir, optique et intensité relative en utilisant le NIST pour 785nm. Ensuite, tous les spectres ont été soumis à un test de qualité basé sur le rapport S/B. L'intensité du signal était celle du pic phénylalanine à 1300cm<sup>-1</sup> et celle du bruit était mesurée dans la zone 1720-1800 cm<sup>-1</sup> qui ne contient aucun signal. Les spectres ayant un rapport signal/bruit inférieur à 50 étaient écartés de l'analyse.

Les spectres validés par le test qualité ont été lissés en utilisant l'algorithme de Savitsky-Golay avec une largeur de fenêtre de 11 points et un polynôme d'ordre 3. La ligne de base a été ensuite corrigée avec un polynôme d'ordre 4 et enfin, les spectres ont été normalisés sur

## Chapitre 3 : Résultats et discussion

toute la gamme spectrale afin de pouvoir les comparer entre eux. La Figure 32 montre un exemple de prétraitements utilisés sur un spectre Raman d'une goutte séchée de sérum.



**Figure 32 : Exemple de prétraitements appliqués sur un spectre Raman d'une goutte de sérum séché.**

A) correction de la réponse instrumentale. B) lissage et estimation de la ligne de base. C) correction de la ligne de base (polynôme 4). D) normalisation et correction du décalage du spectre normalisé.

### Analyse statistique

Les données démographiques et clinico-biologiques ont été recueillies et notées dans les dossiers médicaux des patients. Les données quantitatives ont été exprimées sous forme de moyenne  $\pm$  écart-type et ont été comparées en utilisant le test t de Student. Les données qualitatives ont été exprimées en pourcentage et ont été comparées en utilisant le test du chi-carré. Une valeur de p inférieure à 0,05 était considérée significative.

Afin de réduire la dimensionnalité des données, une ACP a été appliquée aux deux groupes de données spectrales obtenues à partir des sérums en gouttes séchées et des sérums lyophilisés.

## *Chapitre 3 : Résultats et discussion*

Cette méthode est la plus appropriée pour diminuer le nombre de variables prédictives compte tenu du grand nombre de variables dans chaque spectre Raman et du nombre relativement faible de patients ( $n = 71$ ). Ensuite, les CPs expliquant 99,99% de la variance ont été sélectionnées et les plus discriminantes pour différencier les spectres des patients cirrhotiques avec et sans CHC ont été identifiées en utilisant test de Mann-Whitney

L'analyse par cluster a été effectuée avec les CPs les plus explicatives et les CPs discriminantes. Pour ceci, les scores des CPs ont été portés sur des graphiques à trois dimensions pour révéler les relations potentielles entre les spectres des deux groupes de patients.

La méthode de classification SVM a ensuite été appliquée sur les données spectrales en utilisant les CPs les plus discriminantes. Dans le but d'éviter un phénomène de surentrainement en raison du petit nombre de patients ( $n = 71$ ), une validation croisée « leave-one-out » a été utilisée. Cette méthode consiste à placer à chaque fois, les spectres d'un patient donné dans le jeu de validation, alors que les spectres de tous les patients restants sont utilisés en jeu d'entraînement. Le processus est répété de telle sorte que les spectres de chaque patient sont utilisés une fois en jeu de validation. La sensibilité, la spécificité et le taux global de bonne classification du modèle ont été calculés en utilisant l'ensemble des spectres, la moyenne et la médiane de chaque patient.

### Résultats et discussion

#### *1. Population des patients*

Pour 90% des patients inclus dans cette étude, la cirrhose était liée à une consommation excessive d'alcool, le reste des patients avaient une cirrhose dont l'origine était métabolique ou virale (virus hépatique B ou C). Le Tableau 3 montre les caractéristiques démographiques et clinico-biologiques des patients cirrhotiques avec et sans CHC.

Comme nous pouvons le noter, aucune différence significative n'a été relevée dans la distribution des facteurs étiologiques de cirrhose entre les patients cirrhotiques avec et sans cancer. L'âge moyen dans le groupe de patients atteints de CHC était plus élevé ( $66 \pm 9$ ) que dans le groupe sans CHC ( $61 \pm 9$ ,  $p = 0,015$ ) avec une prédominance des hommes dans les deux groupes.

Quant à la sévérité de la cirrhose, elle a été évaluée par le score de Child-Pugh. Ce score était plus élevé, mais pas de manière significative, chez les patients sans CHC. En ce qui concerne

## *Chapitre 3 : Résultats et discussion*

le taux sérique d'alpha-foetoprotéine, il était plus élevé chez les patients atteints de CHC que chez les patients témoins (Tableau 3).

**Tableau 3:** Les caractéristiques démographiques et clinico-biologiques des patients cirrhotiques avec et sans CHC.

	Patients cirrhotiques avec CHC (n=37)	Patients cirrhotiques sans CHC (n=34)	p*
Age	66 ± 9	61±9	0.015
Sex ratio (M/F)	32/5	28/6	0.75
Score de Child-Pugh	6.3 ±1.7	7.3±2.6	0.07
Cause de la cirrhose**			0.23
Alcoolique	32/37	32/34	
virus C	3/37	3/34	
virus B	4/37	0/34	
Métabolique	3/37	3/34	
Alpha-fetoprotéine >500 ng/ml	6/37(16)	0/34	0.03

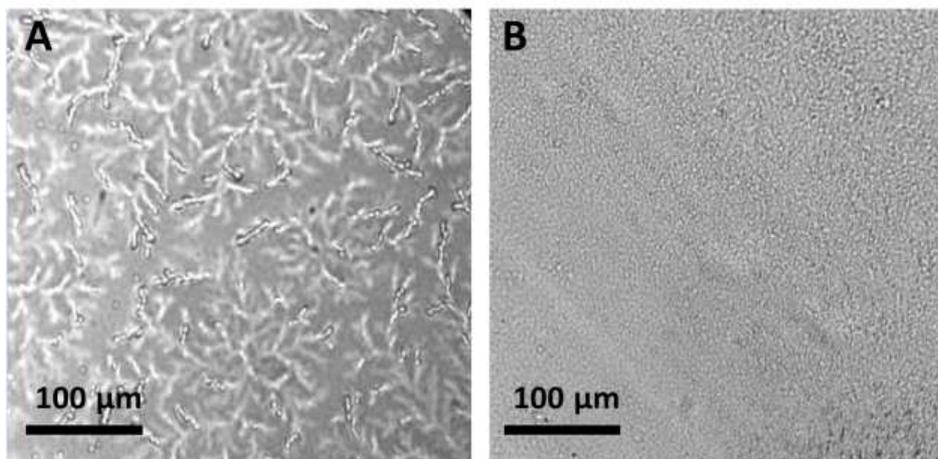
\*Test de Student pour les valeurs quantitatives (moyenne ± SEM) et test  $\chi^2$  pour les valeurs qualitatives.

\*\* Plusieurs causes peuvent coexister chez le même patient.

### *2. Analyse du sérum en goutte séchée*

Dans cette étude, deux préparations d'échantillons ont été réalisées dans le but de tester le potentiel de la spectroscopie Raman. Tout d'abord, une analyse a été faite sur le sérum en goutte séchée. Dans la mesure où il a été démontré que le processus de séchage peut induire une migration des constituants moléculaires du sérum du centre vers la périphérie de la goutte (74), des spectres Raman ont été enregistrés au niveau des deux sites (centre et périphérie de la goutte) et ils ont été analysés séparément.

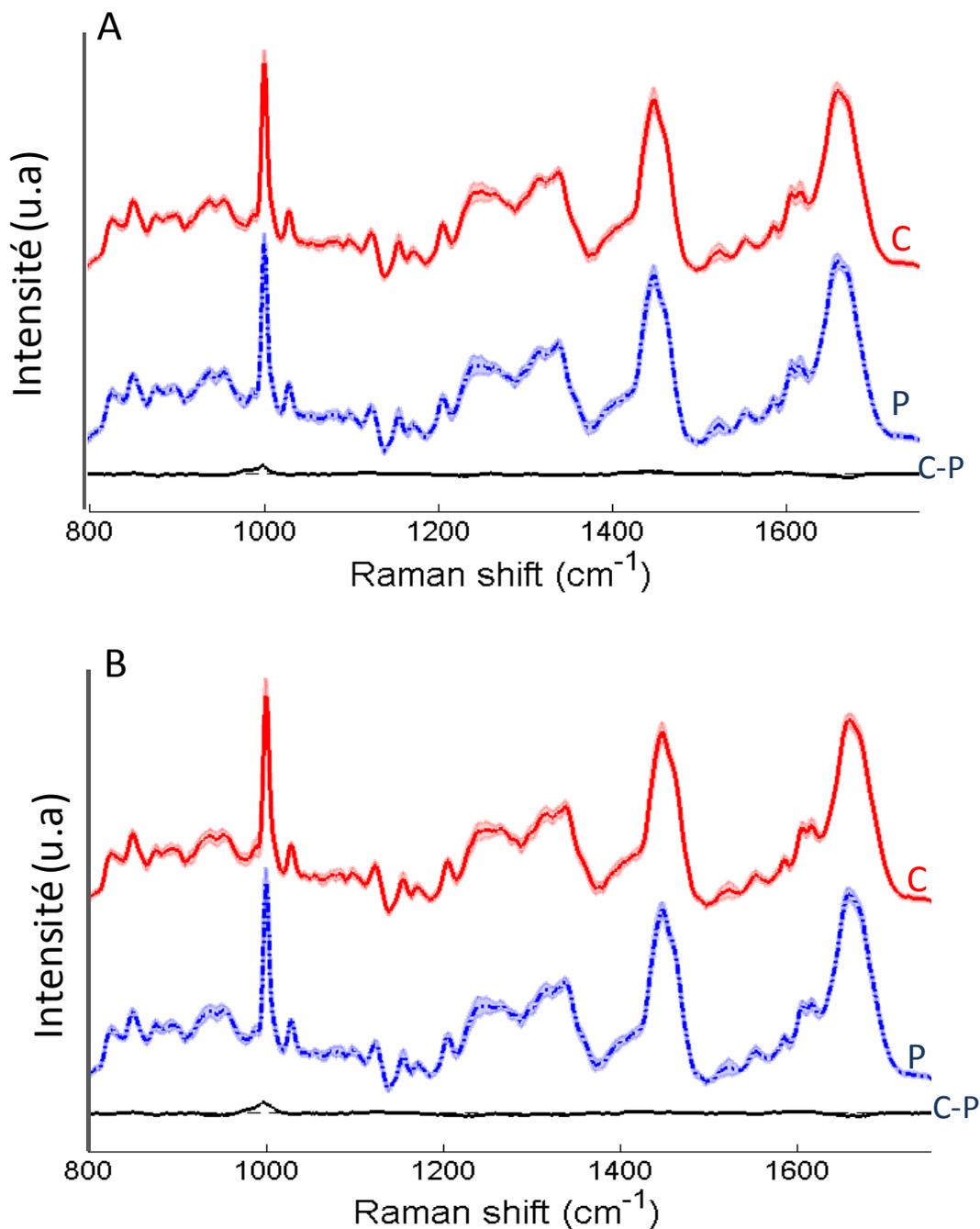
Après acquisition des spectres Raman, ces derniers ont été soumis à un test qualité pour éliminer les spectres bruités. Pour chacun des sites d'acquisition (centre et périphérie), 353 parmi 355 spectres remplissaient le test qualité. La Figure 33A et la Figure 33B montrent, respectivement, l'image en microscopie optique du centre et de la périphérie de la goutte séchée de sérum.



**Figure 33: Image optique du centre (A) et de la périphérie (B) de la goutte séchée du sérum.**

Comme illustré dans la Figure 33, la goutte séchée du sérum présentait deux aspects différents, un aspect cristallisé « en fougère » au centre et un aspect amorphe homogène en périphérie. Ces observations sont en accord avec des études antérieures effectuées sur diverses solutions protéiques (75, 76) ainsi que sur le liquide lacrymal (58, 75). Dans une étude Raman sur les larmes, Filik et al ont montré que les protéines étaient principalement déposées de manière homogène en périphérie (58) alors que le centre de la goutte contenait une petite quantité de protéines, d'électrolytes et d'urée (58, 77).

Dans notre travail, les spectres Raman acquis au centre et à la périphérie de la goutte séchée de sérum étaient très semblables visuellement (Figure 34). Aucune différence ne pouvait être mise en évidence en examinant la moyenne des spectres du centre et la moyenne des spectres de la périphérie de la goutte. Les mêmes observations ont été notées dans les deux groupes de patients cirrhotiques avec CHC (Figure 34A) et sans CHC (Figure 34B).



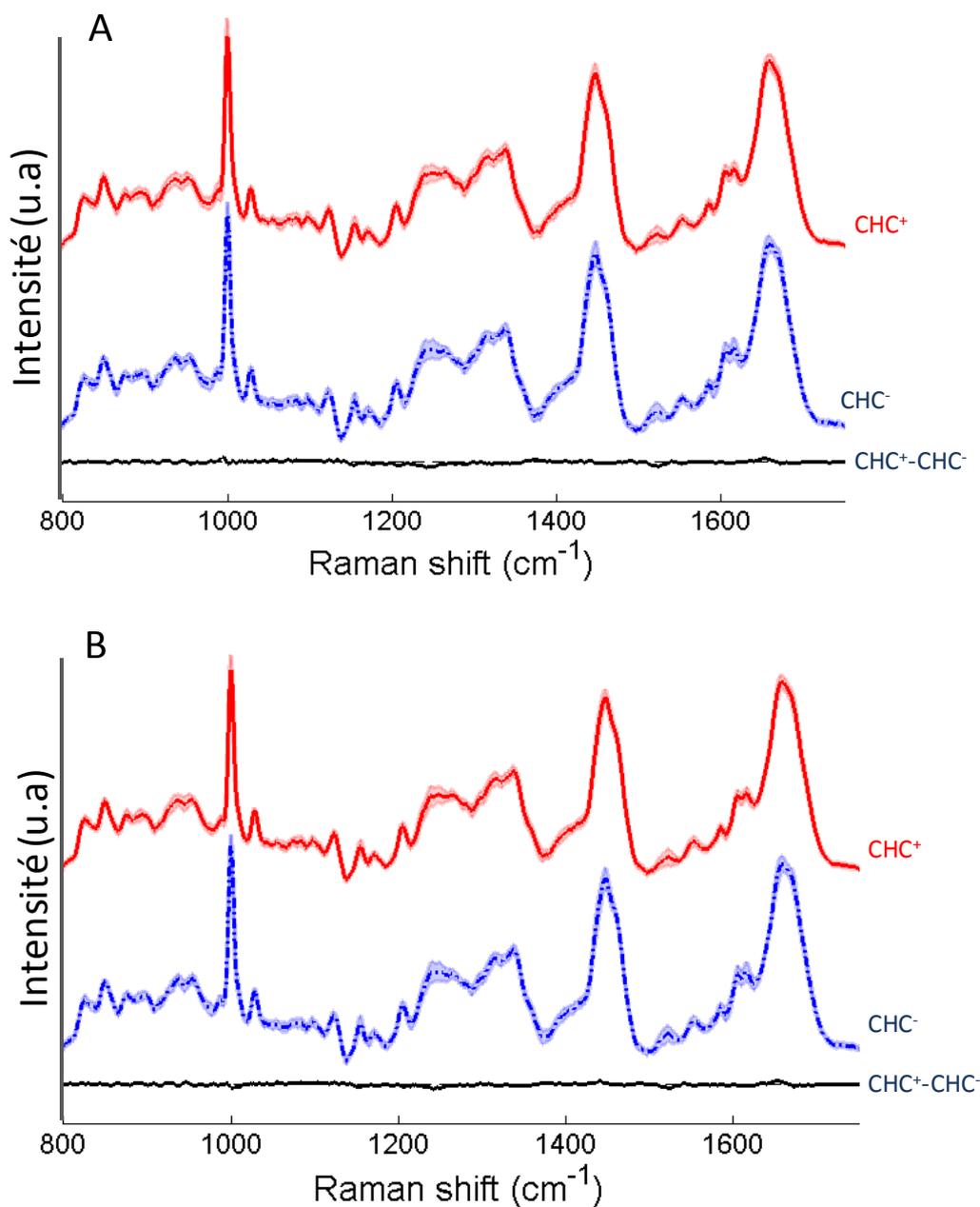
**Figure 34: Comparaison des moyennes des spectres Raman des deux sites d'acquisitions : le centre (C) et la périphérie (P) de la goutte séchée de sérum.**

(A): les patients sans CHC et (B) : les patients avec CHC. Les écart-types (zone grisée) sont superposés aux spectres moyens.

Nous avons comparé, ensuite, les moyennes des spectres Raman du sérum des patients avec CHC et sans CHC. Visuellement, nous n'avons noté aucune différence entre les spectres, tant pour les données spectrales du centre de la goutte séchée (Figure 35A) que pour celles de la

### Chapitre 3 : Résultats et discussion

périphérie (Figure 35B). L'examen des spectres de différence, notés «  $\text{CHC}^+ - \text{CHC}^-$  », confirmaient ces observations.



**Figure 35: Comparaison des moyennes des spectres Raman des patients avec CHC ( $\text{CHC}^+$ ) et sans CHC ( $\text{CHC}^-$ ).**

Le graphique (A) montre les spectres obtenus au centre de la goutte séchée de sérum et le graphique (B) montre les spectres obtenus à la périphérie. Les écart-types (zone grisée) sont superposés aux spectres Raman.

### *Chapitre 3 : Résultats et discussion*

---

Dans le but de mettre en évidence les différences spectrales permettant de distinguer les deux groupes de patients avec et sans CHC, une méthode de classification non supervisée, la classification en clustering basée sur l'ACP a été mise en œuvre dans un premier temps.

#### *Résultats de l'ACP*

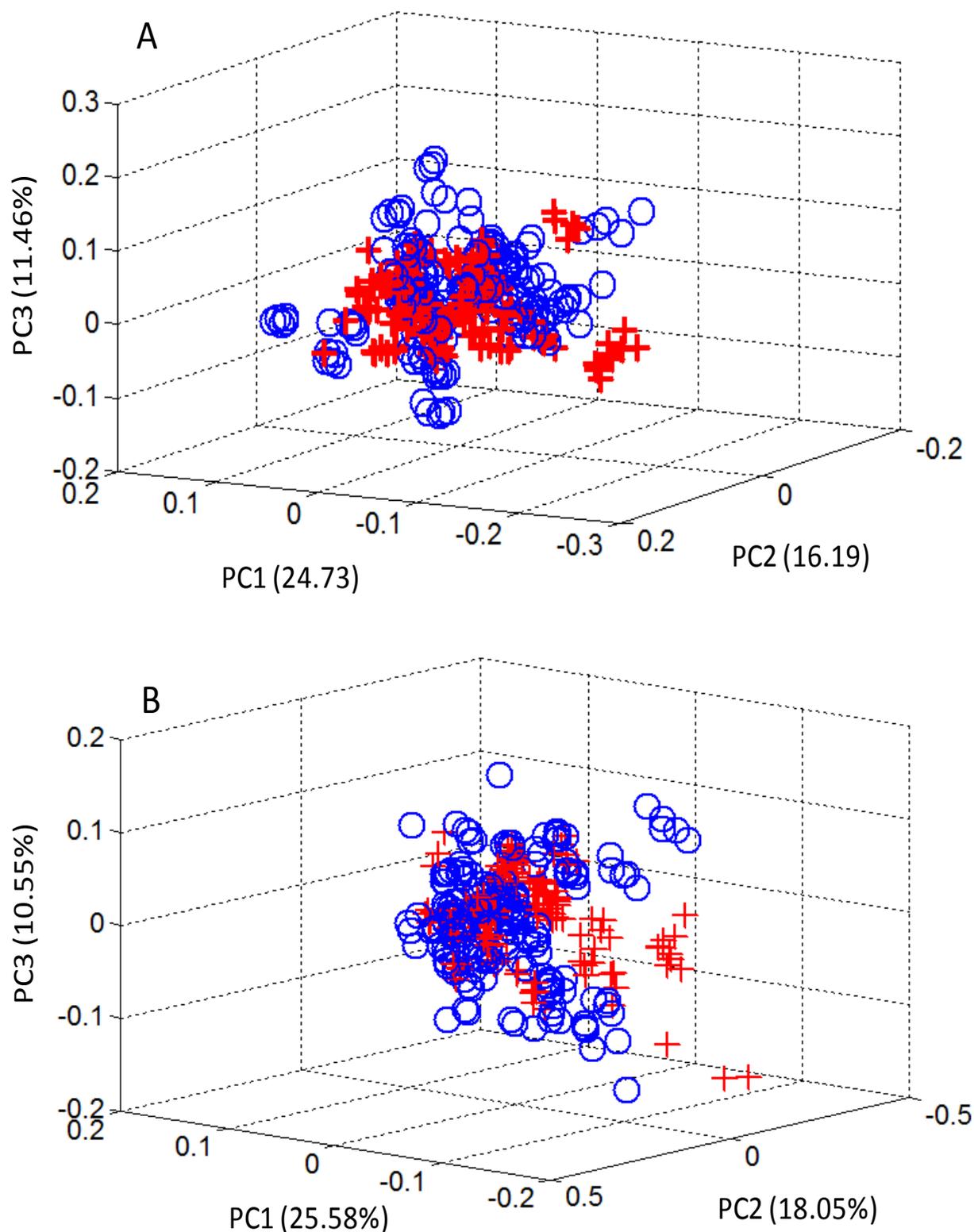
L'ACP a d'abord été appliquée aux spectres Raman enregistrés au centre de la goutte séchée de sérum. Au total, 287 CPs représentaient 99,99% de la variance totale dans le groupe de l'ensemble des 353 spectres individuels, 69 CPs dans le groupe de 71 spectres moyens et 68 CPs dans le groupe de 71 spectres médians.

En ce qui concerne les spectres de la périphérie de la goutte séchée de sérum, 286 CPs, représentaient 99,99% de la variance totale dans le groupe de l'ensemble des 353 spectres individuels. Pour l'analyse des spectres moyens et médians, respectivement, 66 CPs et 68 CPs expliquaient 99.99% de la variance.

La Figure 36 montre la représentation tridimensionnelle obtenue avec les trois CPs les plus explicatives de l'ensemble des spectres provenant du centre (Figure 36A) et de la périphérie (Figure 36B) de la goutte séchée de sérum. Les trois premières CPs représentaient, respectivement, 24.73%, 16.19% et 11.46% de la variance totale pour l'analyse du centre et de 25.58%, 18.05% et 10.55% pour l'analyse de la périphérie.

Comme nous pouvons le voir sur la Figure 36, l'ACP ne permettait pas de séparer les spectres des patients avec et sans CHC. Des résultats similaires ont été obtenus en utilisant les 3 CPs les plus discriminantes.

Ces résultats suggèrent une grande similitude entre les spectres des deux groupes de patients. Pour cette raison, une méthode de classification plus sophistiquée a été mise en œuvre pour discriminer les patients avec et sans cancer.



**Figure 36: Représentation 3D de l'ACP de l'ensemble des spectres Raman du sérum de patients cirrhotiques avec CHC (rouge) et sans CHC (bleu).**

ACP sur l'ensemble des spectres pris au centre (A) et à la périphérie (B) de la goutte séchée.

## *Chapitre 3 : Résultats et discussion*

### *Résultats SVM- LOOCV*

Une analyse SVM-LOOCV a été appliquée aux données Raman en utilisant les CPs les plus discriminantes préalablement définies par l'ACP. Le Tableau 4 présente les résultats de classification obtenus avec la SVM-LOOCV. Cette méthode a été appliquée aux spectres Raman enregistrés au centre et à la périphérie des gouttes séchées de sérum provenant des patients cirrhotiques avec et sans CHC. Pour chaque analyse, la médiane, la moyenne et tous les spectres de chaque patient ont été testés.

Comme nous pouvons le voir sur le Tableau 4, cette procédure de classification a permis de séparer les patients avec et sans CHC avec une sensibilité comprise entre 81 et 92% et une spécificité entre 82 et 92%. L'utilisation de l'ensemble des spectres plutôt que les moyennes ou les médianes de chaque patient n'améliore pas les performances du classifieur. Ce résultat peut s'expliquer par la bonne reproductibilité des spectres du même échantillon.

D'autre part, des données similaires ont été obtenues avec les spectres du centre et de la périphérie de la goutte (Tableau 4). Ceci suggère que les biomarqueurs du CHC ne se séparent pas de manière significative au cours du séchage de la goutte. Dans leur étude, Filik et Stone (58) ont analysé la distribution spatiale des molécules constituant les larmes au cours du séchage. Bien que la composition du fluide lacrymal soit différente du sérum, il représente également une solution biologique complexe contenant plusieurs centaines de protéines distinctes, de lipides, de métabolites et d'électrolytes. Cependant, La concentration des protéines dans le liquide lacrymal est environ sept fois plus faible que dans le sérum. Pour ces raisons, les données Raman obtenues à partir du modèle lacrymal ne peuvent être extrapolées à l'analyse du sérum sans précaution. Néanmoins, Filik et Stone (58) ont montré en utilisant l'ACP que le signal des protéines était prédominant à la périphérie de la goutte, mais était également présent à son centre.

Nos résultats nous amènent à supposer que les marqueurs d'intérêt sont présents au niveau des deux sites d'acquisition, au centre comme à la périphérie de la goutte séchée à des concentrations suffisantes pour donner d'importantes informations spectrales.

### *Chapitre 3 : Résultats et discussion*

**Tableau 4:** Classification SVM-LOOCV des sérums en gouttes séchées des patients cirrhotiques avec et sans CHC.

	Goutte séchée de sérum-centre			Goutte séchée de sérum -périphérie		
	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres
Nombre de CPs*	16	11	32	16	13	41
Patients avec CHC	<b>37</b>	<b>37</b>	<b>37</b>	<b>37</b>	<b>37</b>	<b>37</b>
<i>Bien classés</i>	33	33	33	34	33	30
<i>Mal classés</i>	4	4	4	3	4	7
Patients sans CHC	<b>34</b>	<b>34</b>	<b>34</b>	<b>34</b>	<b>34</b>	<b>34</b>
<i>Bien classés</i>	28	30	30	30	31	30
<i>Mal classés</i>	6	4	4	4	3	4
Sensibilité (%)	<b>89.19</b>	<b>89.19</b>	<b>89.19</b>	<b>91.89</b>	<b>89.19</b>	<b>81.08</b>
Spécificité (%)	<b>82.35</b>	<b>88.24</b>	<b>88.24</b>	<b>88.24</b>	<b>91.18</b>	<b>88.24</b>
Taux de patients bien classés (%)	<b>85.91</b>	<b>88.74</b>	<b>88.74</b>	<b>90.14</b>	<b>90.14</b>	<b>84.51</b>

#### *3. Analyse du sérum lyophilisé*

La lyophilisation est une technique de déshydratation utilisée pour préserver et/ou améliorer la stabilité des protéines (78). Toutefois, il est bien établi que ce processus peut conduire à des changements conformationnels des structures secondaires des protéines (79, 80). Chang et al.(81) ont montré par l'analyse IRTF, une diminution de la teneur en hélice  $\alpha$  après la lyophilisation de l'albumine sérique humaine.

Dans ce travail, et dans le but de s'affranchir de l'hétérogénéité de la goutte du sérum, nous avons analysé les échantillons sous leur forme lyophilisée.

Les 355 spectres Raman des 71 patients ont tous rempli les critères du test qualité. Les cinq spectres enregistrés à différents points d'acquisition de l'échantillon étaient très similaires avec un écart-type très faible reflétant l'homogénéité du lyophilisat pour les deux groupes de patients.

Les spectres moyens du groupe de patients cirrhotiques avec et sans CHC sont présentés dans la Figure 37A. Comme nous pouvons le noter, le spectre de différence ne montrait pas de variations significatives entre les deux groupes de patients avec et sans cancer.

Nous avons ensuite procédé de la même manière que précédemment pour analyser nos données en appliquant une ACP suivie d'une classification SVM-LOOCV sur les médianes, les moyennes et l'ensemble des spectres individuels de chaque patient.

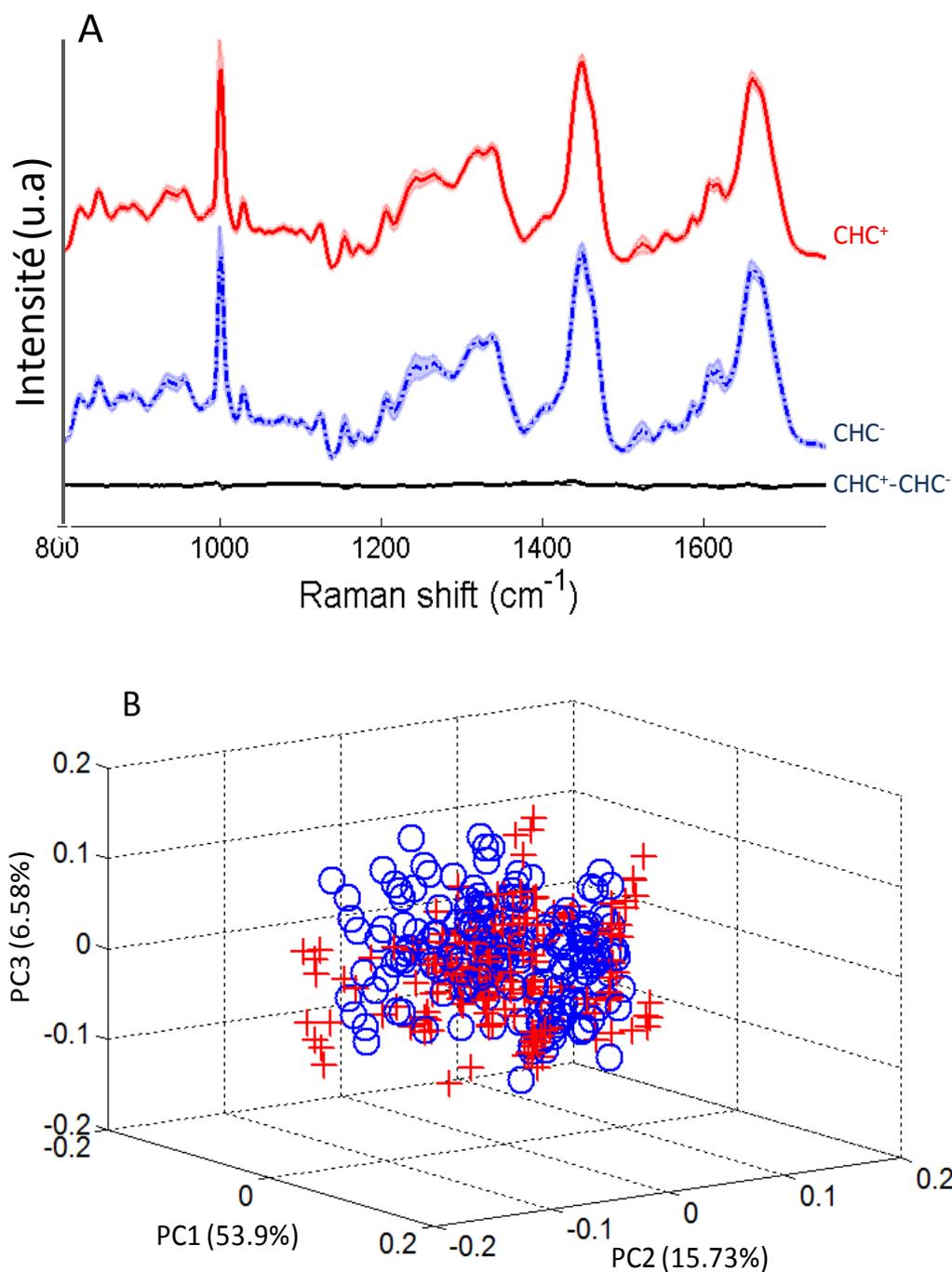
## *Chapitre 3 : Résultats et discussion*

---

### *Résultats de l'ACP*

En appliquant l'ACP aux spectres Raman des sérums lyophilisés, 267 CPs parmi 355 CPs représentaient 99,99% de la variance totale pour l'ensemble des spectres, 65 CPs parmi 71 pour l'analyse des spectres moyens des patients et 68 CPs parmi 71 pour les spectres médians. La Figure 37B montre la représentation graphique de l'ACP du sérum lyophilisé en utilisant les trois composantes les plus explicatives de la variance. Les contributions des CPs (en pourcentages) étaient respectivement de 53,9%, 15,73% et 6,58% de la variance totale.

Comme nous pouvons le voir, aucune discrimination des deux groupes de patients, avec et sans CHC, n'a été observée (les spectres sont mélangés). Le même résultat a été noté en prenant les 3 CPs les plus discriminantes. Ceci reflète la grande similitude entre les spectres des deux groupes de patients.



**Figure 37: Analyse Raman du sérum lyophilisé.**

A) moyenne des spectres Raman du sérum lyophilisé de patients cirrhotiques avec CHC (HCC<sup>+</sup>) et sans CHC (HCC<sup>-</sup>) et leur spectre de différence correspondant (HCC<sup>+</sup>-HCC<sup>-</sup>). La zone grisée représente les écarts-types des spectres de chaque groupe. Tous les spectres sont tracés en utilisant la même échelle verticale et un offset a été appliqué pour plus de clarté. B) ACP sur l'ensemble des spectres Raman du sérum lyophilisé de patients cirrhotiques avec (rouge) et sans (bleu) CHC.

### Chapitre 3 : Résultats et discussion

L'analyse en composantes principales du sérum lyophilisé montre clairement, comme c'était le cas pour le sérum en goutte séchée, ses limites pour séparer les patients cirrhotiques avec et sans CHC. En effet, les données semblent présenter des signatures spectrales très similaires qui ne peuvent être séparées par une telle méthode. L'utilisation d'une méthode de classification plus sophistiquée était donc nécessaire pour obtenir une distinction des deux groupes de patients.

Les résultats de l'ACP ont été utilisés pour la classification supervisée afin de réduire le nombre de variables des données analysées.

#### Résultats SVM- LOOCV

La classification de nos données Raman avec la méthode SVM-LOOCV montre une bonne discrimination entre les spectres des patients avec et sans CHC (Tableau 5). Pour chaque jeu de données utilisé (tous les spectres, moyennes ou médianes), la sensibilité était de 86% à 95% et la spécificité de 82% à 89%. Les taux de bonne classification obtenus avec l'ensemble des spectres étaient légèrement mais pas significativement plus élevés que ceux obtenus pour l'analyse des moyennes et des médianes.

Comme obtenu dans l'analyse du sérum en goutte séchée, l'utilisation de l'ensemble des spectres n'a pas amélioré la performance du classifieur par rapport à l'analyse sur la base de spectres moyens ou médians. Encore une fois, cela pourrait être dû à la bonne reproductibilité des spectres du même échantillon. Le taux global d'exactitude diagnostique se situait entre 85% et 92%.

**Tableau 5:** Classification SVM-LOOCV des sérums lyophilisés de patients cirrhotiques avec et sans CHC.

Sérum lyophilisé			
	Médiane	Moyenne	Tous les spectres
Nombre de CPs	13	9	31
Patients avec CHC	<b>37</b>	<b>37</b>	<b>37</b>
<i>Bien classés</i>	33	32	35
<i>Mal classés</i>	4	5	2
Patients sans CHC	<b>34</b>	<b>34</b>	<b>34</b>
<i>Bien classés</i>	28	30	30
<i>Mal classés</i>	6	4	4
Sensibilité (%)	<b>89.19</b>	<b>86.49</b>	<b>94.59</b>
Spécificité (%)	<b>82.35</b>	<b>88.24</b>	<b>88.24</b>
Taux de patients bien classés (%)	<b>85.91</b>	<b>87.33</b>	<b>91.55</b>

### Conclusion

Notre étude soutient l'hypothèse de travail que la spectroscopie Raman appliquée au sérum et combinée à la méthode SVM-LOOCV, peut être une approche appropriée pour un classement rapide des patients cirrhotiques avec et sans CHC.

Deux préparations d'échantillons, le sérum en goutte séchée et le sérum lyophilisé, ont été analysées et ont donné des performances diagnostiques similaires, à savoir une précision diagnostique globale de 85 à 90%. En dépit de l'aspect visuel hétérogène de la goutte séchée de sérum, le site de mesure n'a pas d'impact significatif sur le taux de patients bien classés. En ce qui concerne les échantillons lyophilisés, le procédé présente l'avantage de donner une poudre de sérum homogène. Cependant, ce procédé nécessite un volume plus important d'échantillon, nécessite plus de 24 heures de préparation des échantillons, et implique un traitement supplémentaire de la matière biologique. Enfin, notre étude indique que les méthodes chimiométriques de classification sont nécessaires pour une analyse spectrale optimale. Dans notre travail, l'ACP seule n'a pas permis la distinction entre les deux groupes de patients. En revanche, la méthode de classification SVM-LOOCV combinée à l'ACP pour réduire la dimensionnalité des données s'est révélée efficace pour détecter des différences spectrales extrêmement subtiles permettant de distinguer les deux groupes de patients. Les résultats obtenus dans cette étude pilote justifient d'étendre la base de données pour construire un modèle SVM avec un set d'entraînement (calibration), un ensemble de validation et un ensemble test afin d'améliorer la robustesse du classifieur.

### **Partie 3.2 : Etude pilote IRTF**

#### **Diagnostic sérologique du carcinome hépatocellulaire chez les patients cirrhotiques par spectroscopie infrarouge à transformée de Fourier.**

##### Contexte de l'étude

Le rationnel de cette étude a été exposé précédemment dans cette thèse. Rappelons brièvement que le pronostic péjoratif du CHC est lié à un diagnostic souvent tardif. L'identification de nouveaux biomarqueurs est donc un enjeu majeur pour obtenir un diagnostic précoce et améliorer la survie. L'utilisation des techniques biophotoniques est une approche innovante pour atteindre cet objectif. L'étude pilote précédente (82) a exploré le potentiel de la spectroscopie Raman appliquée au sérum de patients cirrhotiques avec et sans CHC. Dans ce deuxième travail, l'intérêt de la spectroscopie IRTF a été évalué dans la même population de patients.

⇒ **L'objectif de cette étude pilote était de montrer le potentiel de la spectroscopie IRTF combinée à des méthodes chimiométriques pour la discrimination des sérums de patients cirrhotiques avec CHC et sans CHC.**

##### Patients et méthodes

Cette étude pilote s'intègre dans le cadre de l'étude ProSpec décrite en détail précédemment (82). Les sérums ont été collectés chez 79 patients cirrhotiques : 39 patients avec CHC et 40 patients sans CHC. Les caractéristiques démographiques et clinico-biologiques de ces patients sont détaillées dans le Tableau 6.

### Chapitre 3 : Résultats et discussion

**Tableau 6:** Caractéristiques démographiques et clinico-biologiques des patients cirrhotiques avec et sans CHC.

	Patients cirrhotiques avec CHC (n=39)	Patients cirrhotiques sans CHC (n=40)	p*
Age	65.8 ± 10.5	61.9±9	NS
Sex ratio (M/F)	34/5	31/9	NS
Score de Child-Pugh	6.3 ±1.7	7.4±2.6	0.025
Cause de la cirrhose**			NS
Alcoolique	34/39	37/40	
virus C	3/39	3/40	
virus B	4/39	0/40	
Métabolique	3/39	3/40	
Alpha-foetoprotéine >500 ng/ml	6/39	0/40	0.03

\*Test de student pour les valeurs quantitatives (moyenne ± SEM) et test  $\chi_2$  pour les valeurs qualitatives.

\*\* Plusieurs causes peuvent coexister chez le même patient.

Comme le montre le Tableau 6, aucune différence significative n'a été observée pour l'âge, le sex ratio, et la distribution des facteurs étiologiques entre les patients cirrhotiques avec et sans CHC. La sévérité de la cirrhose évaluée par le score de Child-Pugh était significativement plus importante chez les patients cirrhotiques sans CHC que chez ceux ayant un CHC (7,4 ± 0,4 contre 6,3 ± 0,3, p = 0,025). Concernant l'alpha-foetoprotéine, un taux sérique supérieur à 500 ng/ml, n'était observée que chez les patients cirrhotiques atteints de CHC confirmant l'excellente spécificité de ce seuil sérique, au prix toutefois d'une faible sensibilité (15%). Le CHC était unifocal chez 35,9% des patients, bifocal chez 15,4% et multifocal chez 48,7% d'entre eux. Une extension extrahépatique de la tumeur était détectée chez 10,3% des patients.

#### Préparation des échantillons

Tous les échantillons sériques ont été analysés par spectroscopie IRTF. Chaque échantillon de sérum a été décongelé puis dilué 2 fois dans l'eau distillée. Dix aliquots de 5 microlitres ont été déposés sur les puits de la plaque de silicium 384 puits puis séchés pendant 30 minutes à température ambiante. Après séchage, la plaque a été insérée dans un module à haut débit

### *Chapitre 3 : Résultats et discussion*

couplé à un spectromètre FTIR (HTS-XT-Tensor 27) (Bruker Optics GmbH, Ettlingen, Allemagne).

Les spectres IRTF ont été acquis en mode transmission (OPUS v6.5, Bruker Optics GmbH, Ettlingen, Allemagne) sur la région spectrale 400-4000  $\text{cm}^{-1}$ , avec une résolution spectrale de 4  $\text{cm}^{-1}$  et un nombre de scans de 32.

Les spectres ont été ensuite soumis à des tests de qualité afin de vérifier leur intensité d'absorption, leurs rapports S/B, ainsi que leur teneur en vapeur d'eau.

Les spectres validés par le test qualité ont été prétraités sur la région spectrale 890 à 4000 $\text{cm}^{-1}$ . Après la correction de la ligne de base, les spectres ont été lissés avec une largeur de fenêtre de 9 points à l'aide de l'algorithme de Savitsky-Golay puis convertis en dérivée seconde. Les spectres dérivés étaient ensuite normalisés vectoriellement. Ces prétraitements spectraux ont été réalisés en utilisant le logiciel OPUS (v6.5, Bruker Optics GmbH, Ettlingen, Allemagne).

Les spectres corrigés ont été analysés avec la méthode de classification hiérarchique (HCA) en utilisant le calcul de la distance euclidienne et l'algorithme de Ward afin d'exclure les spectres aberrants (« outliers »). Le partitionnement des données (« data clustering ») a été réalisé en utilisant l'information spectrale contenue dans les régions de 2800-3100  $\text{cm}^{-1}$  et 900-1800  $\text{cm}^{-1}$ .

#### *Analyse statistique*

Les données démographiques, cliniques et biologiques ont été collectées chez tous les patients. Les données quantitatives ont été exprimées en moyenne  $\pm$  écart-type et ont été comparées en utilisant le test t de Student. En ce qui concerne les données qualitatives, elles ont été exprimées en pourcentage et ont été comparées en utilisant le test Khi 2. Une valeur de p inférieure à 0,05 était considérée comme significative.

L'analyse spectrale a été réalisée en utilisant le logiciel Matlab (Matlab, The MathWorks, Inc., Natick, Massachusetts, USA). Après exclusion des spectres aberrants (« outliers »), tous les spectres bruts ayant passé le test qualité ont été traités par la méthode EMSC avec un polynôme d'ordre 4. Les différences spectrales entre les deux groupes de patients étant très faibles, le spectre moyen, non informatif, a été soustrait (information commune à tous les individus).

Une ACP des spectres corrigés a été appliquée et les CPs les plus discriminantes ont été

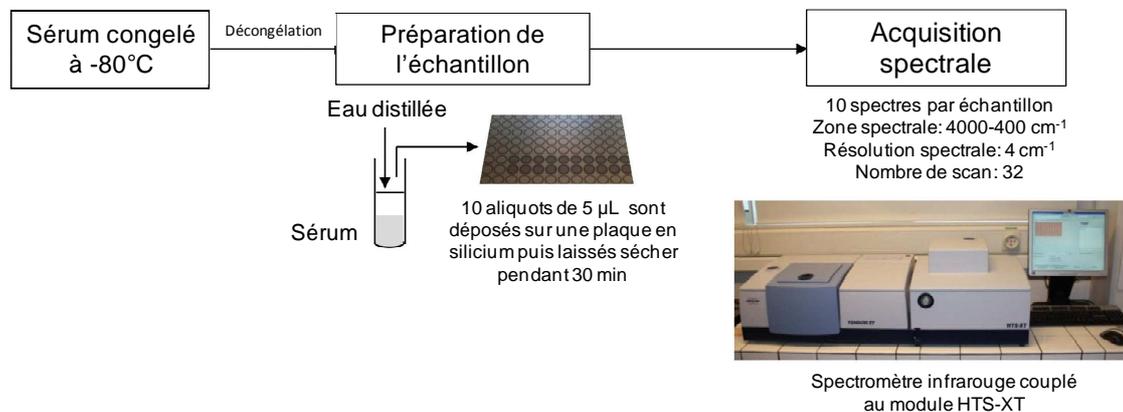
### Chapitre 3 : Résultats et discussion

déterminées à l'aide d'un test statistique appliqué sur les scores de chacune des CPs. La distribution de probabilité des scores étant inconnue, le test statistique non paramétrique de Mann-Whitney a été utilisé. Ce test permet la réorganisation des CPs dans l'ordre croissant de leur valeur de « p », la première étant la plus discriminante entre les deux groupes de patients.

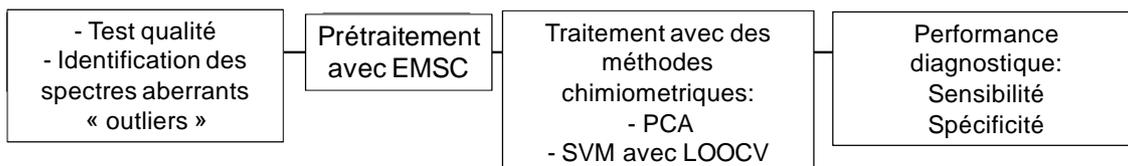
Dans un premier temps, la méthode de classification non supervisée basée sur l'ACP a été utilisée pour analyser les spectres de patients atteints ou non de CHC et déterminer si les deux ensembles de spectres pouvaient être séparés. Cette méthode a été appliquée en utilisant les CPs les plus explicatives de la variance puis les CPS les plus discriminantes.

Dans un deuxième temps, la méthode SVM combinée à une validation croisée a été appliquée compte tenu du petit nombre d'échantillons étudiés. L'analyse a été faite sur les médianes et tous les spectres des patients. Seules les CPs expliquant 99,99% de la variance totale ont été utilisés pour optimiser le modèle SVM. Les autres CPs représentatives du bruit ont été écartées de l'analyse. La sensibilité et la spécificité du classifieur ont été calculées pour deux sets de données, avec et sans spectres aberrants « outliers ». La Figure 38 récapitule les étapes de travail de la préparation des échantillons à l'analyse spectrale par SVM-LOOCV.

#### Etape 1: Préparation des échantillons et acquisition des spectres



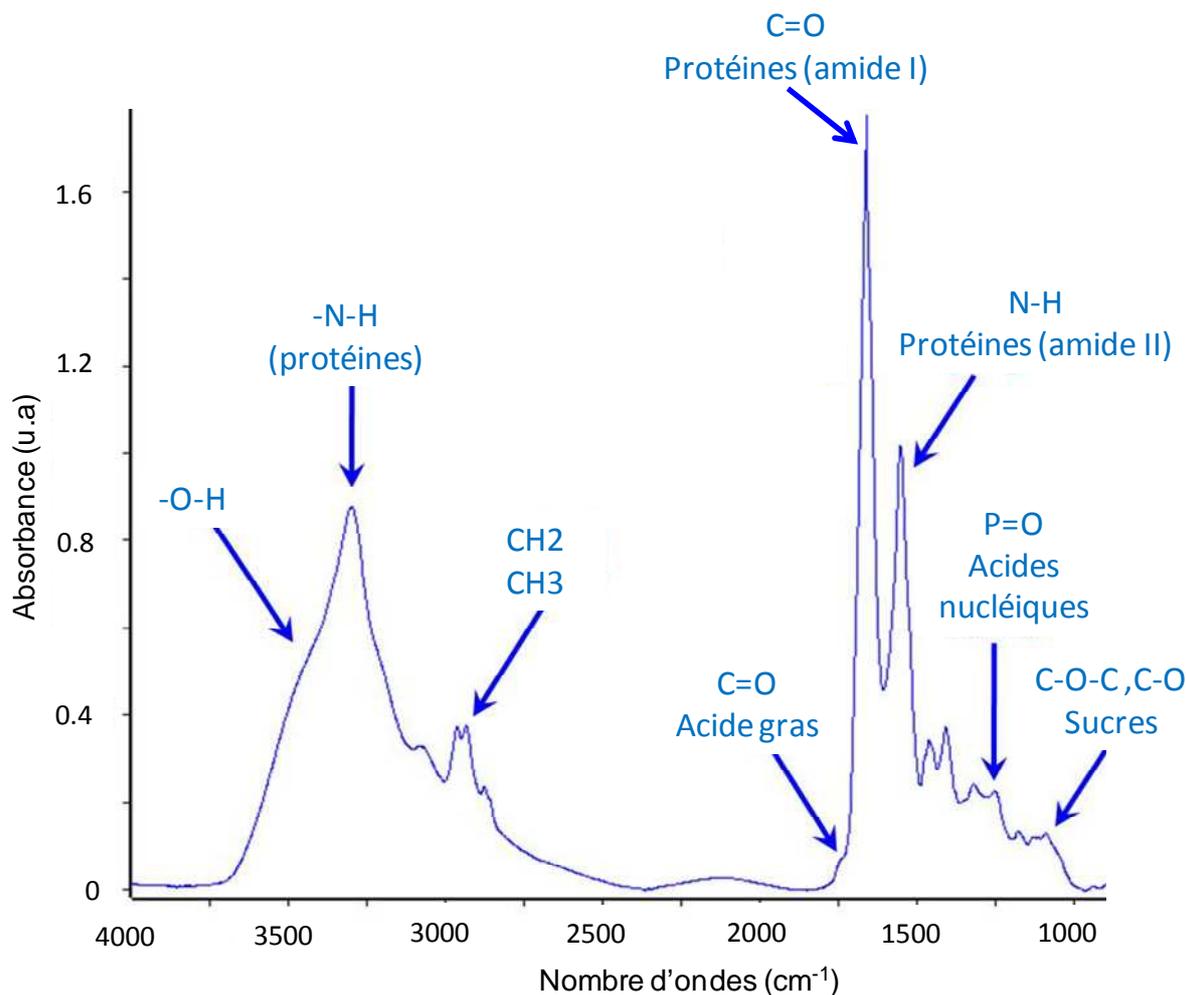
#### Etape 2: Prétraitements et analyses spectrales



**Figure 38 : Schéma décrivant de la procédure expérimentale utilisée pour la préparation des échantillons et l'acquisition spectrale.**

## Chapitre 3 : Résultats et discussion

La Figure 39 ci-dessous présente un exemple de spectre IR du sérum humain et les attributions de pics correspondantes:



**Figure 39 : Attribution de pics du spectre infrarouge du sérum humain.**

### Résultats et discussion

Après l'acquisition des spectres infrarouges, ces derniers ont été tous validés par le test qualité, soit 790 spectres (79 patients). Cependant, l'analyse HCA a permis de sélectionner 51 spectres aberrants: 24 spectres du groupe de patients avec CHC et 27 spectres appartenant au groupe de patients sans CHC.

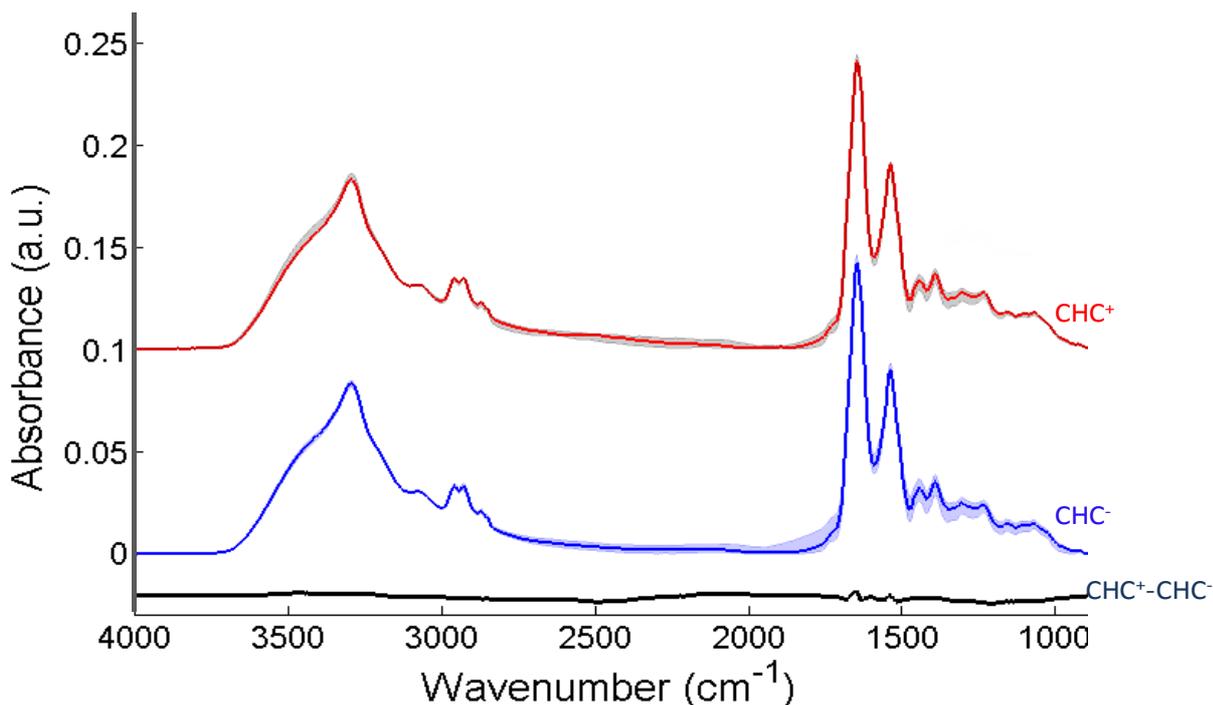
Sur la base de ces résultats, trois ensembles de données ont été créés:

1. Les données contenant l'ensemble des 790 spectres (données 1).
2. Les données excluant les « outliers », soit 739 spectres (données 2).
3. Les données contenant les spectres médians des 79 patients, soit 79 spectres (données 3).

### *Chapitre 3 : Résultats et discussion*

Les spectres bruts ont été prétraités par EMSC dans le but de corriger la ligne de base et de les normaliser.

La Figure 40 montre une comparaison des spectres médians des deux groupes de patients, avec CHC ( $\text{CHC}^+$ ) et sans CHC ( $\text{CHC}^-$ ) après prétraitements. Comme nous pouvons le voir sur la figure, les spectres sont très similaires et les différences spectrales entre les deux groupes de patients sont très faibles.



**Figure 40 : Spectres médians IR prétraités par EMSC des groupes de patients cirrhotiques avec CHC ( $\text{CHC}^+$ ) et sans CHC ( $\text{CHC}^-$ ).**

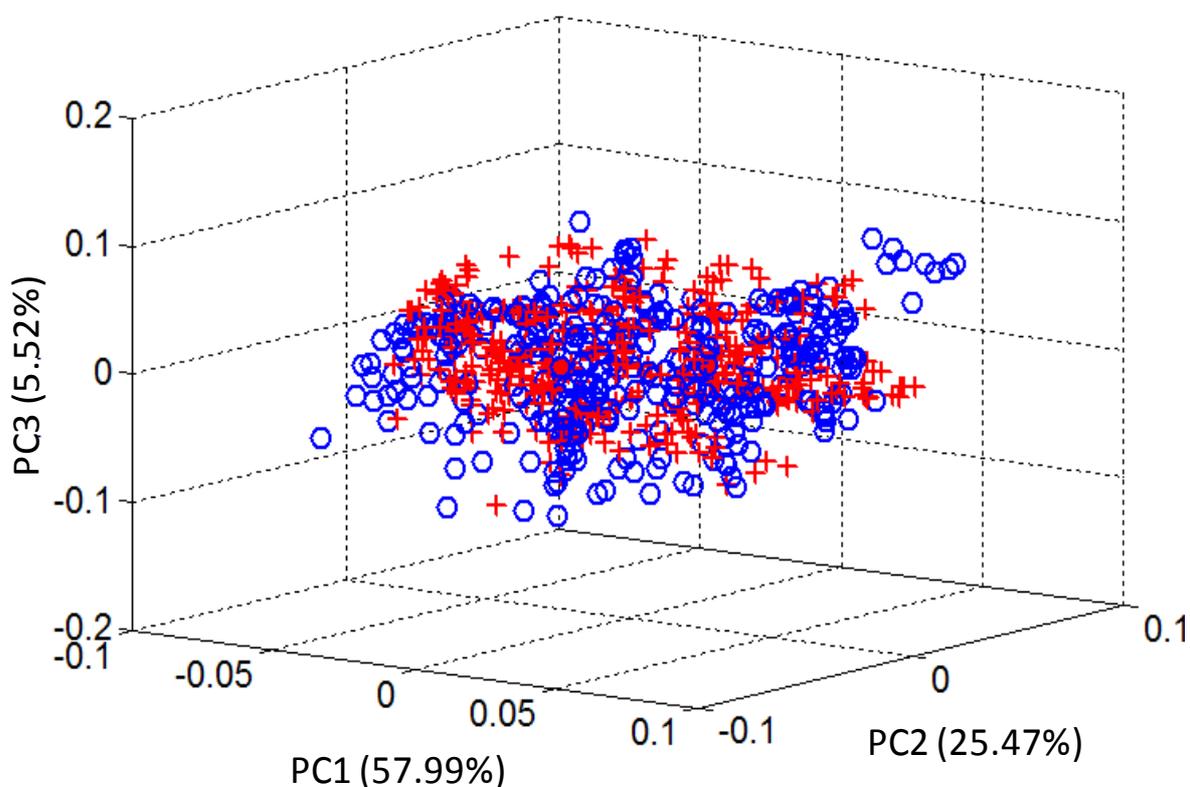
Le spectre de différence ( $\text{CHC}^+ - \text{CHC}^-$ ) montre de très légères variations. Les spectres sont décalés pour plus de clarté, mais sont à la même échelle.

#### Résultats de l'ACP

L'ACP a été appliquée aux trois groupes de spectres précédemment décrits: 131 CPs expliquaient 99.99% de la variance totale pour l'ensemble des 790 spectres (données 1), 130 CPs pour les données 2 (sans outliers) et 51 CPs pour l'ensemble des spectres médians (données 3). L'analyse par clusters a été effectuée en utilisant soit les CPs les plus explicatives, soit les CPs les plus discriminantes dans chacune des 3 bases de données spectrales.

### *Chapitre 3 : Résultats et discussion*

La Figure 41 montre un exemple de représentation graphique tridimensionnelle de l'ACP appliquée à l'ensemble des 790 spectres (données 1).



**Figure 41 : Représentation graphique des 3 CPs les plus explicatives sur l'ensemble des spectres infrarouge du sérum de patients cirrhotiques avec (rouge) et sans (bleu) CHC.**

Comme nous pouvons le voir, la classification par cette méthode n'a pas permis de séparer les sérums de patients avec et sans CHC, en utilisant les CPs les plus explicatives et les CPs les plus discriminantes. Le même résultat était noté pour les autres séries de spectres.

Ce résultat suggère que les différences spectrales sont très subtiles et nécessitent l'utilisation de méthodes de classification plus sophistiquées.

#### Résultats SVM- LOOCV

La méthode de classification supervisée SVM-LOOCV a été appliquée en utilisant les CPs les plus discriminantes préalablement déterminées par ACP. Pour les séries de données 1 et 2, les patients ont été classés conformément à l'affectation de la majorité de leurs spectres (supérieur

### *Chapitre 3 : Résultats et discussion*

ou égal à 6 spectres). En ce qui concerne la série de données 3, les patients ont été classés en fonction de l'affectation de leur spectre médian.

Lorsque la méthode SVM-LOOCV a été appliquée à l'ensemble des spectres (données 1), le modèle optimal de classification a été construit avec les 74 CPs les plus discriminantes expliquant 42% de la variance. La sensibilité du classifieur était de 82,05% et la spécificité de 82,5%. Le taux global de prédiction était de 82,3%. Dans le cas d'une analyse sur l'ensemble des spectres sans « outliers » (données 2), le modèle SVM optimal a été construit avec 54 CPs expliquant 42% de la variance. Ce modèle a permis d'obtenir une sensibilité et une spécificité de 87,18% et 85% respectivement. La précision était de 86,1%. Enfin, lorsque nous avons appliqué la même procédure aux spectres médians de tous les patients (données 3), le modèle a été construit avec 21 CPs soit 10,94% de la variance. La sensibilité et la spécificité étaient respectivement de 84,62% et 82,50% et le taux global de patients bien classés de 83,54%.

Le Tableau 7 reprend l'ensemble des résultats obtenus pour les 3 séries de données :

**Tableau 7 :** Classification SVM-LOOCV des 79 patients cirrhotiques avec et sans CHC. Le modèle a été appliqué à l'ensemble des 790 spectres (données 1), aux 739 spectres après l'exclusion des spectres aberrants (données 2) et aux 79 spectres médians de tous les patients.

<b>Analyse IRTF</b>			
	<b>Données 1 (790 spectres)</b>	<b>Données 2 (739 spectres)</b>	<b>Médianes (79 spectres)</b>
Nombre de CPs	74	54	21
Patients avec CHC	<b>39</b>	<b>39</b>	<b>39</b>
<i>Bien classés</i>	32	34	33
<i>Mal classés</i>	7	5	6
Patients sans CHC	<b>40</b>	<b>40</b>	<b>40</b>
<i>Bien classés</i>	33	34	33
<i>Mal classés</i>	7	6	7
Sensibilité (%)	<b>82.05</b>	<b>87.18</b>	<b>84.62</b>
Spécificité (%)	<b>82.50</b>	<b>85</b>	<b>82.50</b>
Taux de patients bien classés (%)	<b>82.3</b>	<b>86.1</b>	<b>83.54</b>

Ce travail a mis en évidence le potentiel de la spectroscopie IR appliquée au sérum pour le diagnostic de CHC chez des patients cirrhotiques.

Le potentiel diagnostique de la technique appliquée à l'analyse des tissus a été rapporté précédemment pour le diagnostic de tumeurs de la peau (83), la région cervicale (84), de la

## *Chapitre 3 : Résultats et discussion*

---

vésicule biliaire (31), et des lymphomes (85). Elle a été appliquée pour analyser des cellules Elle a été appliquée également pour l'analyse cytologique (18, 86).

L'application de la spectroscopie IRTF sur le sérum est moins répandue, mais des résultats prometteurs ont été rapportés pour le diagnostic de plusieurs pathologies.

La spectroscopie IR a été appliquée dans le cadre du diagnostic du cancer du sein (54). Quatre vingt seize sérums de patients sains et 97 sérums de patients cancéreux ont été analysés. Une analyse par réseau de neurones artificiels (ANN) a été appliquée sur les données spectrales et a permis d'obtenir 90% de sensibilité et de spécificité.

Enfin, Dans une étude pilote, E. Scaglia et al 2011 ont utilisé la spectroscopie IRTF afin d'évaluer la fibrose hépatique chez les patients atteints d'hépatite C chronique. Vingt-trois échantillons de sérums ont été analysés : 11 sérums de patients avec fibrose et 12 sérums de patients sans fibrose. La classification supervisée SVM appliquée aux données spectrales a permis une classification correcte des spectres dans 95.65% des cas (57).

### **Conclusion**

Notre étude confirme l'hypothèse de travail que le sérum des patients cirrhotiques atteints de CHC présente une signature spectrale infrarouge qui permet de les différencier des patients cirrhotiques sans CHC. Une petite quantité de sérum, sans réactif, est suffisante à cet effet. Pour développer ce nouvel outil diagnostique dans la pratique clinique, la prochaine étape est de valider les résultats sur un plus grand nombre d'échantillons, de construire un classifieur robuste avec une méthode telle que la SVM et le tester de façon prospective sur un ensemble indépendant d'échantillons de sérum.

### **Partie 3.3 : Validation de l'approche IRTF dans une cohorte de 817 patients.**

#### **Evaluation de la spectroscopie IRTF appliquée au sérum pour le diagnostic de CHC : étude multicentrique dans une cohorte de 817 patients cirrhotiques**

##### Contexte de l'étude

Le cancer primitif du foie est le sixième cancer le plus répandu dans le monde avec une large prédominance dans les pays en voie de développement (1). Le carcinome hépatocellulaire est le type histologique le plus fréquent, représentant environ 80% des cas. Alors que son incidence diminue dans la plupart des pays asiatiques en raison de la réduction de la transmission du virus de l'hépatite B, elle augmente dans d'autres parties du monde comme les États-Unis et l'Europe, en raison de l'extension épidémique de l'hépatite C et de l'obésité au cours des dernières décennies (1, 87). En 2008, la mortalité liée au CHC a été estimée à 700 000 décès dans le monde (1, 87). Ce pronostic est dû au diagnostic tardif de la tumeur excluant tout traitement potentiellement curatif.

La grande majorité des CHC compliquent l'évolution d'une maladie chronique du foie sous-jacente (88). La cirrhose est le principal facteur de risque avec une incidence annuelle du CHC estimée entre 1 et 8% en fonction de la cause de cirrhose (3). Pour cette raison, un programme de surveillance est recommandé pour cette population à haut risque (3). Le but est d'obtenir une diminution de la mortalité liée à la maladie grâce à la détection de petites tumeurs pouvant bénéficier de traitements curatifs.

Selon les recommandations actuelles, la surveillance des patients cirrhotiques est basée sur l'échographie hépatique effectuée tous les six mois (3). Cependant, la sensibilité de cette technique pour détecter les tumeurs à un stade précoce est limitée. Elle est estimée à seulement 63% dans une méta-analyse récente (89). Dans le but d'améliorer l'efficacité du dépistage, la performance diagnostique de nombreux marqueurs sérologiques dont l'alpha-fœtoprotéine (AFP), la lectine liée à l'AFP et des gamma-carboxy prothrombine a été évaluée (11, 12). Bien que certains experts recommandent l'utilisation de l'AFP sérique en combinaison avec l'analyse échographique (90), les sociétés médicales scientifiques et professionnelles n'ont pas retenu cette stratégie en raison de la faible sensibilité de l'AFP sérique (augmentation du nombre des faux positifs) et de son coût (3). L'identification de nouveaux biomarqueurs permettant un diagnostic précoce du CHC apparaît aujourd'hui

### *Chapitre 3 : Résultats et discussion*

---

comme un objectif majeur de santé publique afin d'améliorer la survie des patients atteints de ce cancer (3).

Plusieurs approches telles que la génomique, la protéomique et la métabolomique ont été utilisées pour atteindre ce but. Bien que certaines études aient permis d'améliorer la connaissance des mécanismes de la cancérogenèse du CHC, aucun des biomarqueurs candidats n'a franchi le l'étape du laboratoire à la pratique clinique. Compte tenu de ces résultats décevants, nous avons évalué l'intérêt d'une nouvelle approche basée sur la biophotonique, c'est-à-dire les interactions entre la lumière et les nuages d'électrons des liaisons interatomiques au sein des molécules présentes dans l'échantillon étudié. La spectroscopie vibrationnelle, Raman ou infrarouge, permet d'analyser les caractéristiques biochimiques du sérum sous la forme de spectres qui reflètent sa composition moléculaire quantitativement et qualitativement. Les deux techniques sont non-destructives, sans contact et n'utilisent pas de réactifs. Alors que la spectroscopie Raman est basée sur le principe de la diffusion de la lumière par l'échantillon, la spectroscopie infrarouge à transformée de Fourier repose sur le principe d'absorption du rayonnement infrarouge permettant d'obtenir une signature moléculaire de l'échantillon étudié. Ces technologies ont été utilisées dans le domaine de la recherche en cancérologie, principalement pour l'analyse des tissus et des cellules, avec l'objectif de développer de nouveaux outils diagnostiques ou pronostiques. Des résultats prometteurs ont été rapportés. En revanche, l'application de la spectroscopie vibrationnelle aux biofluides, particulièrement le plasma / sérum, est beaucoup plus limitée. Nous avons rapporté précédemment dans deux études pilotes, le potentiel de la spectroscopie Raman et de la spectroscopie IRTF à différencier les échantillons de sérum de patients cirrhotiques avec et sans CHC (82, 91). Dans ces études, nous avons analysé un nombre limité de patients en utilisant la méthode de classification SVM combinée à la validation croisée par « leave-one-out » (LOOCV) afin de discriminer les patients avec et sans CHC. Un taux d'exactitude diagnostique globale de 80 à 90 % a été obtenu pour les deux techniques d'analyse justifiant une étude de validation sur un plus grand nombre de patients.

Cette étude supplémentaire est rapportée dans cette 3ème partie du travail de thèse. Plus de 800 échantillons de sérum de patients cirrhotiques avec et sans CHC ont été analysés par spectroscopie FTIR. Cette technique permet une acquisition spectrale rapide avec une collecte de données à haut débit. Des méthodes avancées de classification telles que la SVM et la PLS-DA ont été appliquées aux données spectrales dans le but de valider les résultats obtenus dans nos études pilotes antérieures (82, 91) et de construire un modèle de classification robuste qui pourrait être utilisé dans la pratique clinique.

## Chapitre 3 : Résultats et discussion

### Matériels et méthodes

Ce travail s'inscrit dans le cadre de l'étude prospective ProSpec dont l'objectif était d'identifier des biomarqueurs sériques ou plasmatiques de CHC chez des patients cirrhotiques en utilisant deux approches: la protéomique et la spectroscopie vibrationnelle. C'est une étude ancillaire du projet inter-régional CiRCE (Cirrhose et Risque de Carcinome hépatocellulaire dans le grand Est) visant à évaluer le rôle des facteurs environnementaux, nutritionnels et métaboliques dans la survenue de CHC chez les patients cirrhotiques. Ces patients ont été recrutés dans les hôpitaux universitaires du Cancéropôle de la région Grand Est de France (Reims, Strasbourg, Metz, Besançon, Dijon et Nancy).

L'étude a été approuvée par le Comité de Protection des Personnes Est (convention n° 2008/09.) et par l'Agence Française de Sécurité Sanitaire des Produits de Santé (AFSSAPS. Accord n° 2008-A00023-52). Tous les patients ont signé un consentement pour participer à l'étude.

Pour chaque patient, les données suivantes ont été recueillies : âge, sexe, cause de la cirrhose, stade de la cirrhose, présence ou non d'un diabète, données anthropométriques, et taux sérique d'alpha-foetoprotéine.

### *Population étudiée*

Huit cent dix sept patients cirrhotiques ont été inclus dans cette étude, 308 patients atteints de CHC et 509 sans CHC. Le tableau 8 indique la répartition des patients selon les centres.

**Tableau 8 :** Nombre de patients par centre de recrutement de la région Grand Est de France.

	<i>Centre 1</i>	<i>Centre 2</i>	<i>Centre 3</i>	<i>Centre 4</i>	<i>Centre 5</i>
	<i>(Dijon)</i>	<i>(Metz)</i>	<i>(Nancy)</i>	<i>(Reims)</i>	<i>(Strasbourg)</i>
<i>Patients sans CHC</i>	110	79	92	135	102
<i>Patients avec CHC</i>	101	18	59	65	56

*\*Noter que les patients du centre 6 (Besançon) n'ont pas été inclus dans l'étude en raison du non respect des modalités de stockage des prélèvements sanguins telles que prévues par le protocole CiRCE.*

Selon les critères définis dans l'étude de CiRCE, le diagnostic de cirrhose reposait soit sur l'examen histopathologique d'une biopsie du foie ou sur la combinaison de signes clinico-biologiques d'insuffisance hépato-cellulaire ou d'hypertension portale et/ou de signes endoscopiques d'hypertension portale et/ou de caractéristiques d'imagerie évocatrices de

### *Chapitre 3 : Résultats et discussion*

cirrhose. L'absence de CHC était confirmée par des techniques d'imagerie (échographie et / ou tomodensitométrie et/ou tomodensitométrie et/ou imagerie par résonance magnétique), et par une concentration d'alpha-foetoprotéine sérique inférieur à 100 ng.ml<sup>-1</sup>. Le diagnostic de CHC était fondé sur les critères définis par l'Association Européenne pour l'Etude du Foie (3). Les critères d'exclusion étaient l'âge inférieur à 35, l'infection par le VIH, les cancers extra-hépatiques et les maladies psychiatriques ne permettant pas l'obtention d'un consentement éclairé.

Les caractéristiques clinico-biologiques des patients sont présentées dans le Tableau 9.

**Tableau 9 :** Caractéristiques clinico-biologiques des patients cirrhotiques avec et sans CHC.

	<b>Tous les patients</b>	<b>Patients cirrhotiques avec CHC (n=37)</b>	<b>Patients cirrhotiques sans CHC (n=34)</b>	<b>p*</b>
Age (n=791)	60.7±10.3	64.3 ± 9.4	58.7 ± 10.1	<0.0001
Prédominance masculine	76.2%	87%	70.2%	<0.0001
Score de Child-Pugh (n=769)	7.3 ± 2.3	6.5 ± 1.8	7.3 ± 2.3	
Grade A	53.8%	64.5%	47.7%	
Grade B	31.3%	26.2%	34.3%	<0.0001
Grade C	14.8%	9.3%	18%	
Cause de la cirrhose** (n=791)				
Alcoolique	67.8%	69.5%	66.8%	0.067
Virus C	24.9%	22.3%	26.5%	
Virus B	5.8%	8.1%	4.5%	
Diabète (n=791)	38.3%	45.6%	34.2%	0.0015
BMI (n=789)	27.5± 5.4	27.7 ± 5.1	27.3 ± 5	0.28
Obésité (n=789)	28%	31%	26.3%	0.16
AFP>500 ng/ml (n=784)	6.6%	18.7%	0%	<0.0001

\*Test de Student pour les valeurs quantitatives (moyenne ± SEM) et test  $\chi^2$  pour les valeurs qualitatives. \*\* Plusieurs causes peuvent coexister chez le même patient

Ces patients avaient des caractéristiques épidémiologiques concordantes avec celles rapportées dans la population des patients cirrhotiques en France (92). Il existait une prédominance masculine chez les patients cirrhotiques en raison de la prévalence plus élevée des facteurs de risques de la cirrhose chez les hommes, notamment la consommation excessive d'alcool. La répartition des causes de cirrhose était dominée par la consommation excessive d'alcool qui concernait deux patients sur trois et par l'infection virale C qui concernait environ 25 % des patients (92). La forte proportion des patients cirrhotiques

### *Chapitre 3 : Résultats et discussion*

infectés par le virus C s'explique par l'épidémie d'hépatite C qui s'est développée en France à partir des années 60-70 avec la diffusion des pratiques de transfusion sanguine et de la toxicomanie intraveineuse. La prévalence du diabète était plus élevée chez les patients cirrhotiques avec CHC, de l'ordre de 30-40%, confirmant le rôle favorisant du diabète pour le développement du CHC (93).

L'augmentation du taux sérique d'alpha-foetoprotéine au dessus de 500 ng /ml, considéré comme pathognomonique du CHC chez les patients cirrhotiques ayant un nodule hépatique, était observée chez moins d'un patient sur 5 dans le groupe CHC et chez aucun patient cirrhotique sans CHC.

#### *Préparation des échantillons*

Après le prélèvement sanguin, les échantillons ont été centrifugés immédiatement. Les sérums ont été ensuite congelés à  $-80^{\circ}\text{C}$  jusqu'à l'utilisation. Après décongélation, chaque sérum a été dilué au tiers dans du sérum physiologique. Dix aliquots de  $5\ \mu\text{L}$  d'échantillon ont été déposés sur une plaque de silicium de 384 puits (Bruker Optics, Ettlingen, Allemagne) puis séchés à température ambiante pendant environ 30 min. La plaque a été ensuite transférée dans le spectromètre à transformée de Fourier (tenseur 27, Bruker Optics GmbH, Ettlingen, Allemagne) couplé à un système à haut débit (HTS-XT, Bruker Optics GmbH, Ettlingen, Allemagne).

Dix dépôts par échantillon ont été analysés afin de détecter des spectres aberrants liés à la plaque de silicium elle-même et/ou à la procédure d'échantillonnage.

#### *Acquisition des spectres*

Les spectres IRTF des échantillons ont été acquis en mode transmission. Chaque spectre a été enregistré dans la région spectrale comprise entre 400 et  $4000\ \text{cm}^{-1}$ . La résolution spectrale était de  $4\ \text{cm}^{-1}$  avec un nombre de scans de 32. Après l'acquisition spectrale, un test qualité a été appliqué à tous les spectres (rapport S/B, rapport Signal/Eau et intensité d'absorbance). Dans le groupe de patients cirrhotiques sans CHC, quarante six spectres parmi 5090 (soit  $\sim 0,9\%$ ) ne remplissaient pas les critères du test qualité. Dans ce groupe, deux patients pour lesquels plus de 6 spectres du même sérum ne remplissaient pas les critères de qualité ont été exclus. Concernant les patients cirrhotiques atteints de CHC, onze spectres parmi 3080 n'ont pas été validés par le test qualité (soit  $\sim 0,35\%$ ).

### *Prétraitements des spectres*

Les spectres IRTF validés par le test de qualité ont été analysés en conservant d'abord l'ensemble de la gamme spectrale  $900\text{-}4000\text{ cm}^{-1}$  puis secondairement en sélectionnant la région  $900\text{-}1800\text{ cm}^{-1}$  (empreinte spectrale). Une analyse spectrale avec et sans dérivée seconde a été appliquée sur les deux zones sélectionnées.

Dans un premier temps, les spectres ont été lissés en utilisant l'algorithme Savitzky-Golay (94) avec un polynôme du troisième ordre et une largeur de fenêtre de 9 points. Ensuite, les données ont été prétraités avec la méthode EMSC (The MathWorks, Natick, MA). Nous avons utilisés un polynôme du huitième ordre pour les données sans dérivation et un polynôme du second ordre pour les spectres dérivés. Enfin, le spectre moyen a été soustrait de chaque spectre pour enlever l'information redondante et commune à tous les spectres. Tous les prétraitements et analyses ont été réalisés avec le logiciel Matlab (version 7.6.0, MathWorks, Natick, Mass, USA).

### *Analyse statistique*

Concernant les caractéristiques démographiques et clinico-biologiques, les données quantitatives ont été exprimées en moyenne  $\pm$  écart-type et ont été comparées en utilisant le test t de Student. Les données qualitatives ont été exprimées en pourcentage et ont été comparées en utilisant le test du chi-carré. Une valeur de p inférieure à 0,05 était considérée comme significative.

Plusieurs méthodes de classification ont été utilisées pour classer les spectres IRTF des deux groupes de patients, l'analyse en composantes principales ACP et les méthodes SVM et PLS-DA. Dans un premier temps, une réduction de la dimension spectrale (775 longueurs d'onde) a été obtenue par une ACP afin de minimiser le risque de surentrainement lié au phénomène de Hughes (95). L'ACP est une méthode non supervisée qui réduit le nombre de dimensions par une transformation linéaire conduisant à une base de dimension réduite dans laquelle la variance des données est maximale.

Les CPs les plus discriminantes entre les spectres de patients cirrhotiques avec et sans CHC ont été identifiées en utilisant le test de Mann-Whitney. Une analyse en clustering a ensuite été réalisée avec les trois CPs les plus significatives puis les trois CPs les plus discriminantes de la variance.

Dans un deuxième temps, les méthodes de classification SVM linéaire et non linéaire puis par PLS-DA ont été appliquées en utilisant les CPs les plus discriminantes. Ces méthodes ont

## Chapitre 3 : Résultats et discussion

été retenues pour leur capacité à traiter un volume important de données. Pour analyser les 817 patients de cette étude IR, les modèles SVM et PLS-DA ont été construits et validés avec 3 ensembles de données réparties aléatoirement en utilisant le logiciel Matlab: un jeu d'entraînement (calibration), un jeu de validation (validation interne) et un jeu test (validation externe). Les effectifs de ces trois jeux de données sont indiqués dans le Tableau 10.

**Tableau 10** : Répartition aléatoire des patients en 3 jeux de données utilisés pour l'analyse SVM et l'analyse PLS-DA.

	Jeu d'entraînement (calibration)	Jeu de validation (validation interne)	Jeu test (validation externe)
Patients sans CHC	160	80	269
Patients avec CHC	160	80	68

Le modèle a été construit avec le même nombre de patients avec et sans CHC. Après calibration du modèle par le set d'entraînement, celui-ci a été optimisé en utilisant le set de validation (validation interne). Enfin, le set 'jeu test' a permis de tester le modèle en aveugle et de calculer la spécificité et la sensibilité du classifieur obtenu.

La sensibilité et la spécificité des classifieurs ont été calculées successivement pour la médiane, la moyenne et pour l'ensemble des spectres de chaque patient.

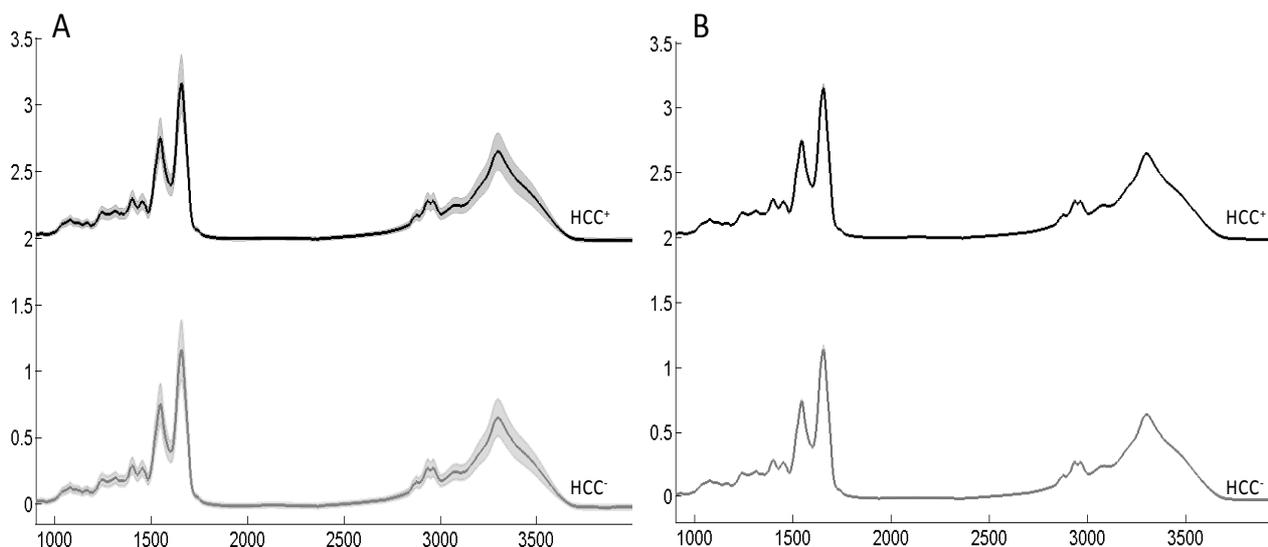
### Résultats

#### 1. Analyse des données sur l'ensemble du spectre IR ( 900-4000 $\text{cm}^{-1}$ )

##### **a. Analyse des spectres IRTF – sans dérivation**

Les données spectrales ont été prétraitées selon les paramètres suivants :

- Lissage des spectres : polynôme d'ordre 3 et 9 points
- Dérivée : 0
- Correction EMSC : polynôme d'ordre 8
- Soustraction du spectre moyen



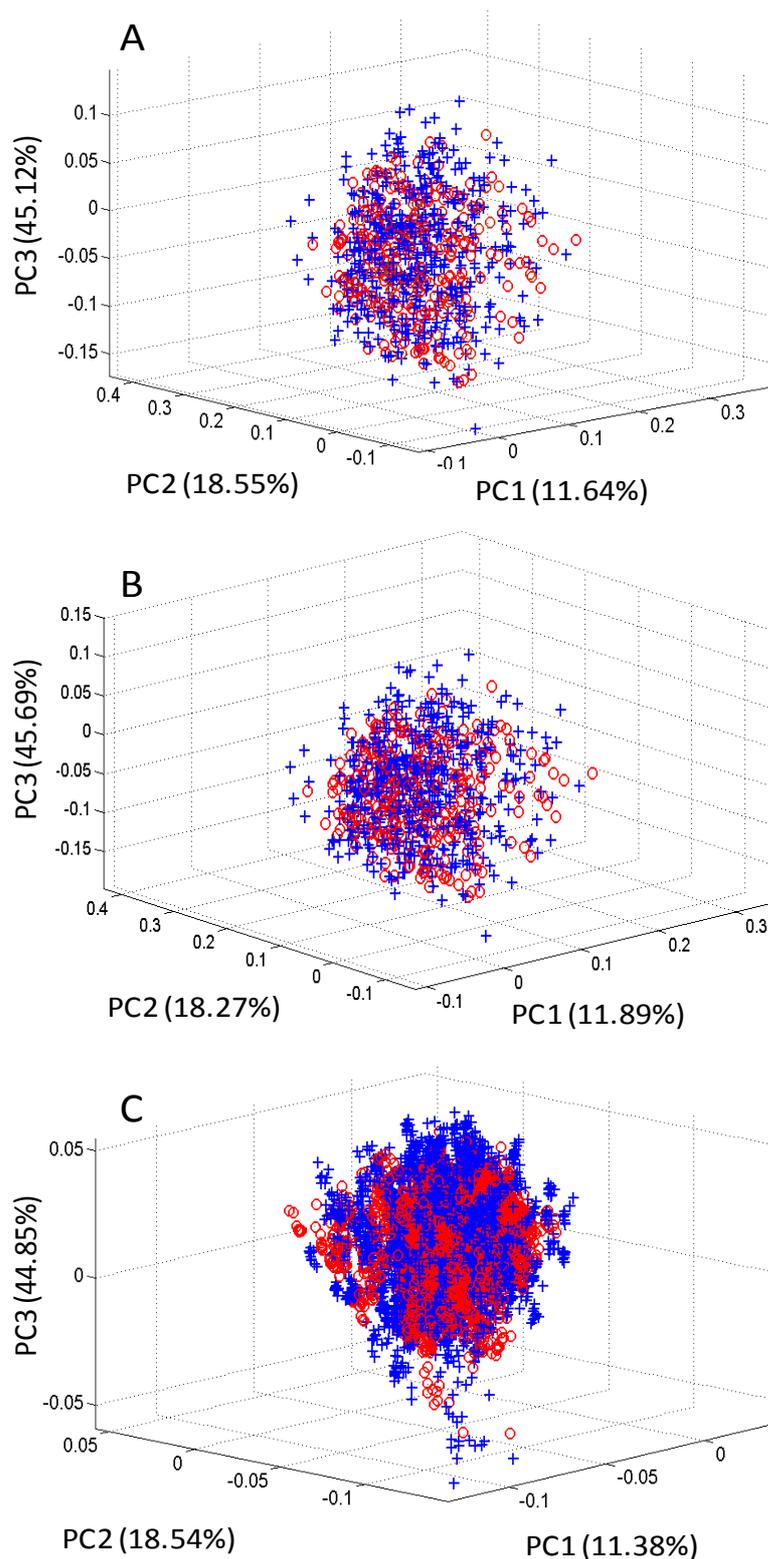
**Figure 42 : Spectres moyens IRTF des sérums de patients avec et sans CHC.**

(A) Spectres bruts. (B) Spectres prétraités par EMSC avec un polynôme d'ordre 8 sur la région spectrale 900-4000  $\text{cm}^{-1}$ .

La Figure 42 montre les spectres moyens des groupes de patients avec et sans CHC. La Figure 42A présente la moyenne des spectres bruts de chaque groupe de patients avec l'écart-type correspondant. La Figure 42B montre les moyennes de ces deux groupes de patients après prétraitement par EMSC. L'analyse visuelle des spectres ne montre aucune différence significative entre les moyennes des deux groupes (spectres très similaires). L'application de méthodes de classification plus puissantes est donc nécessaire pour détecter d'éventuelles différences spectrales permettant de discriminer les deux groupes de sujets analysés.

### Analyse par composantes principales

L'ACP a été menée successivement sur les spectres médians de chaque patient (320 spectres), les spectres moyens (320 spectres) puis l'ensemble des spectres de chaque patient (3180 spectres) du groupe d'entraînement. La Figure 43 montre la représentation tridimensionnelle de l'ACP des médianes, des moyennes et de l'ensemble des spectres de chaque patient en utilisant les 3 composantes les plus explicatives de la variance.



**Figure 43 : Analyse en composante principale des données IRTF des patients avec (rouge) et sans (bleu) CHC.**

A) Analyse des médianes. B) Analyse des moyennes. C) Analyse de l'ensemble des spectres de chaque patient sur la région spectrale  $900-4000\text{ cm}^{-1}$ , sans dérivée seconde.

### *Chapitre 3 : Résultats et discussion*

---

Comme illustré sur la Figure 43, aucune discrimination entre les deux groupes de patients, avec et sans CHC, n'a été observée (les spectres sont mélangés et se superposent). Le même résultat a été obtenu en prenant les 3 CPs les plus discriminantes pour chaque analyse. Ce résultat reflète la grande similitude entre les spectres des deux groupes de patients cirrhotiques avec et sans cancer et justifie la mise en œuvre de méthodes de classification plus sophistiquées pour distinguer les patients avec et sans CHC.

#### *Analyse SVM*

Pour construire le modèle SVM, les sets de données (entraînement, validation interne et validation externe) ont été sélectionnés au hasard, en utilisant le logiciel Matlab. La composition des différents groupes est mentionnée dans le Tableau 10

Dans cette partie du travail, nous rapportons les résultats de la méthode de classification SVM obtenus sur la région spectrale  $900-4000\text{ cm}^{-1}$  sans dérivation seconde (Tableau 11). Le modèle de classification a été construit en utilisant les CPs les plus discriminantes représentant 99.99% de la variance totale de chaque base de données.

### *Chapitre 3 : Résultats et discussion*

**Tableau 11** : Classification SVM sur la médiane, la moyenne et tous les spectres des patients cirrhotiques avec et sans CHC. Analyse sur la région spectrale 900-4000 cm<sup>-1</sup> sans dérivation seconde.

	Jeu d'entraînement			Jeu de validation			Jeu test		
	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres
Nombre de CPs	56	33	47	56	33	47	56	33	47
Patients avec CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>68</b>	<b>68</b>	<b>68</b>
<i>Bien classés</i>	<i>114</i>	<i>116</i>	<i>114</i>	<i>51</i>	<i>51</i>	<i>51</i>	<i>44</i>	<i>37</i>	<i>42</i>
<i>Mal classés</i>	<i>46</i>	<i>44</i>	<i>46</i>	<i>29</i>	<i>29</i>	<i>29</i>	<i>24</i>	<i>31</i>	<i>26</i>
Patients sans CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>269</b>	<b>269</b>	<b>269</b>
<i>Bien classés</i>	<i>108</i>	<i>118</i>	<i>119</i>	<i>50</i>	<i>53</i>	<i>49</i>	<i>112</i>	<i>151</i>	<i>150</i>
<i>Mal classés</i>	<i>52</i>	<i>42</i>	<i>41</i>	<i>30</i>	<i>27</i>	<i>31</i>	<i>157</i>	<i>118</i>	<i>119</i>
Sensibilité (%)	<b>71,25</b>	<b>72,5</b>	<b>71,25</b>	<b>63,75</b>	<b>63,75</b>	<b>63,75</b>	<b>64,71</b>	<b>54,41</b>	<b>61,76</b>
Spécificité (%)	<b>67,5</b>	<b>73,75</b>	<b>74,38</b>	<b>62,5</b>	<b>66,25</b>	<b>61,25</b>	<b>41,64</b>	<b>56,13</b>	<b>55,76</b>
Taux de patients bien classés (%)	<b>69,38</b>	<b>73,13</b>	<b>72,82</b>	<b>63,13</b>	<b>65</b>	<b>62,50</b>	<b>46,30</b>	<b>55,78</b>	<b>56,97</b>

### *Chapitre 3 : Résultats et discussion*

---

Le modèle de classification optimal de la méthode SVM pour l'analyse des médianes a été construit avec les 93 CPs les plus discriminantes. Après optimisation avec le jeu de validation interne, le modèle retenu, basé sur 56 CPs, a été testé sur le set de validation externe. La sensibilité du classifieur était de 64,7% et la spécificité de 41.6%. Le taux global de patients bien classés était de 46.3%.

Dans le cas de l'analyse des moyennes des spectres des patients, le modèle a été construit sur 54 CPs les plus discriminantes. Le modèle optimal obtenu après validation interne était basé sur 33 CPs. La sensibilité du classifieur testé sur le groupe de validation externe était de 54.4% et la spécificité de 56.1% pour un taux global de patients bien classés de 55.8%. Enfin, lorsque nous avons appliqué la même procédure à l'ensemble des spectres des 817 patients, nous avons construit le modèle avec les 154 CPs les plus discriminantes sur le groupe d'entraînement. Le meilleur modèle après validation interne était basé sur les 47 CPs les plus discriminantes. La sensibilité de ce classifieur en validation externe était de 61.76%, la spécificité de 55.76% et le taux de patients bien classés de 56.97%.

#### *Analyse PLS-DA*

Les données IR ont été traitées par PLS-DA dans le but de discriminer les patients avec et sans carcinome hépatocellulaire. Le Tableau 12 présente les résultats obtenus avec le classifieur optimisé par l'utilisation du jeu de validation interne. Ce classifieur a ensuite été testé par le jeu de données externe (jeu test). L'analyse a été faite sur la médiane, la moyenne et l'ensemble des spectres de chaque patient sur la région spectrale  $900-4000\text{ cm}^{-1}$  sans dérivation seconde.

### *Chapitre 3 : Résultats et discussion*

**Tableau 12 :** Classification PLS-DA sur la médiane, la moyenne et l'ensemble des spectres des patients cirrhotiques avec et sans CHC. Analyse sur la région spectrale 900-4000 cm<sup>-1</sup> sans dérivation seconde.

	Jeu d'entraînement			Jeu de validation			Jeu test		
	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres
<i>Nombre de dimensions</i>	23	24	23	23	24	23	23	24	23
Patients avec CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>68</b>	<b>68</b>	<b>68</b>
<i>Bien classés</i>	123	118	109	56	59	57	44	45	47
<i>Mal classés</i>	37	42	51	24	21	23	24	23	21
Patients sans CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>269</b>	<b>269</b>	<b>269</b>
<i>Bien classés</i>	127	125	121	49	47	49	139	134	134
<i>Mal classés</i>	33	35	39	31	33	31	130	135	135
Sensibilité (%)	<b>76.9</b>	<b>73.8</b>	<b>68.3</b>	<b>70</b>	<b>73.8</b>	<b>70.8</b>	<b>64.7</b>	<b>66.2</b>	<b>69.6</b>
Spécificité (%)	<b>79.4</b>	<b>78.1</b>	<b>75.7</b>	<b>61.3</b>	<b>58.8</b>	<b>60.8</b>	<b>51.7</b>	<b>49.8</b>	<b>49.9</b>
Taux de patients bien classés (%)	<b>78.15</b>	<b>75.95</b>	<b>72.00</b>	<b>65.65</b>	<b>66.3</b>	<b>65.8</b>	<b>54.32</b>	<b>53.11</b>	<b>53.88</b>

## *Chapitre 3 : Résultats et discussion*

Le Tableau 12 montre les résultats PLS-DA obtenus sur les 817 patients avec et sans carcinome hépatocellulaire. Le modèle optimal pour l'analyse des médianes était un modèle à 23 dimensions. La sensibilité de ce classifieur sur le groupe de validation externe était de 64.7% et la spécificité de 51.7% correspondant à un taux de patients bien classés de 54.32%. En ce qui concerne l'analyse des moyennes, le modèle optimal a été établi avec 24 dimensions. Sa sensibilité et sa spécificité étaient de 66.2% et 49.8% respectivement. Le taux global de bonne classification était de 53.11%. Pour l'analyse de tous les spectres des patients pris individuellement, le modèle optimal construit avec 23 dimensions permettait d'obtenir une sensibilité de 69.6%, une spécificité de 49.9% et un taux global de bonne classification de 53.88%.

Globalement, les résultats obtenus par la méthode PLS-DA étaient similaires à ceux obtenus avec l'analyse SVM.

### **b. Analyse des spectres IRTF – avec la dérivation seconde**

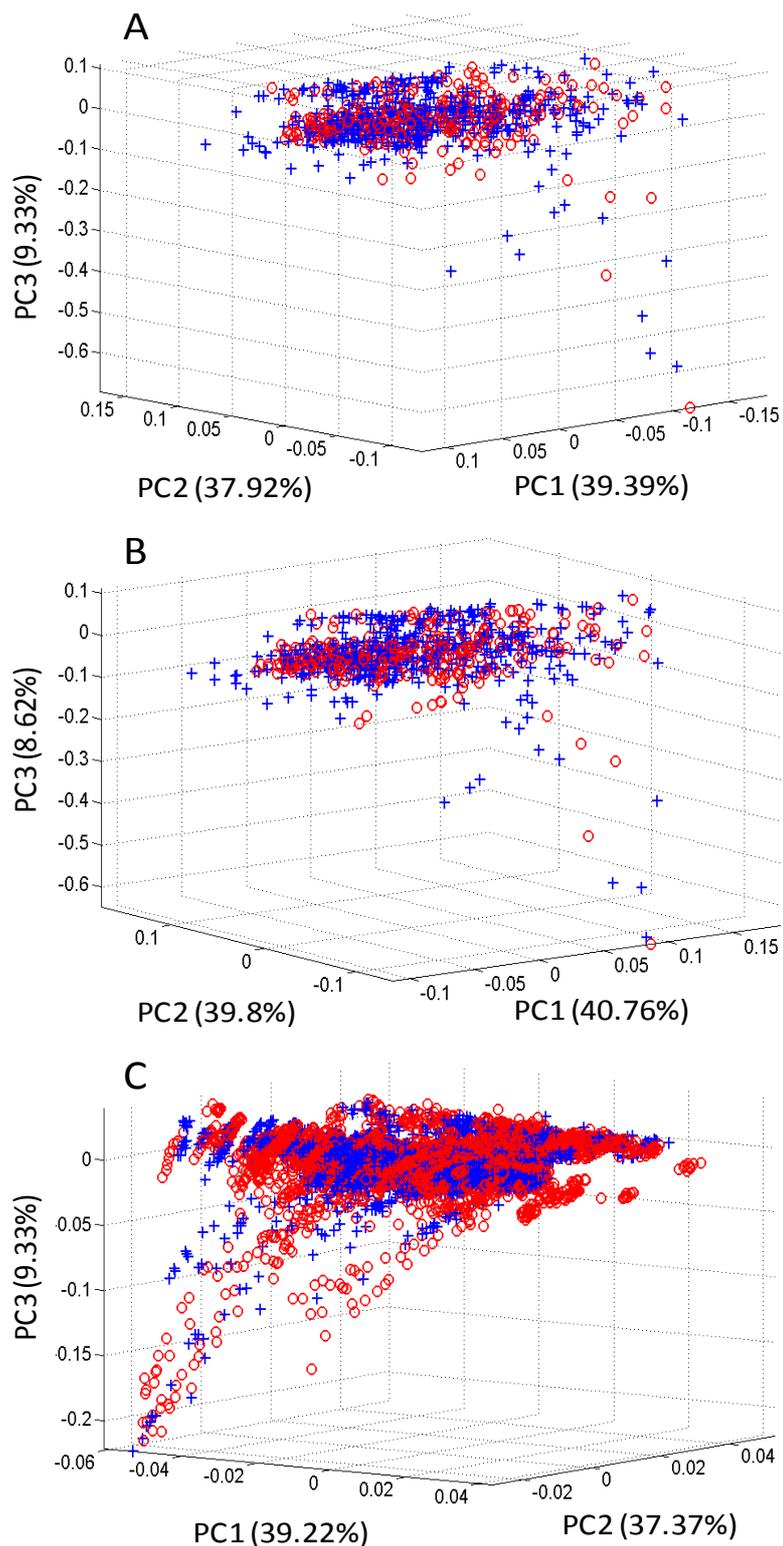
Les données spectrales ont été prétraitées selon les paramètres suivants :

- Lissage des spectres : polynôme d'ordre 3 et 9 points
- Dérivée : 2
- Correction EMSC : polynôme d'ordre 2
- Soustraction du spectre moyen

Les mêmes traitements et analyses (ACP, SVM et PLS-DA) utilisés pour l'étude précédente ont été appliqués sur les spectres dérivés afin de discriminer les patients avec et sans CHC. En effet, en accentuant le contraste spectral, la dérivation seconde pourrait permettre d'identifier les différences spectrales subtiles permettant de distinguer les deux groupes de patients.

### **Analyse en composante principales**

L'ACP a été réalisée sur le groupe des spectres médians (320 spectres) et moyens (320 spectres) et sur l'ensemble des spectres (3180 spectres) de chaque patient. La Figure 44 montre la représentation tridimensionnelle de l'ACP en utilisant les 3 composantes les plus explicatives de la variance.



**Figure 44 : Analyse en composante principale des données IRTF des patients avec (rouge) et sans (bleu) CHC.**

A) Analyse des médianes. B) Analyse des moyennes. B) Analyse de l'ensemble des spectres de chaque patient sur la région spectrale  $900-4000\text{ cm}^{-1}$  avec une dérivation seconde.

### *Chapitre 3 : Résultats et discussion*

---

L'analyse ACP avec les 3 composantes principales les plus explicatives n'a pas permis de séparer les patients cirrhotiques avec et sans CHC. Le même résultat a été obtenu en prenant les 3 CPs les plus discriminantes. Ce résultat peut être expliqué par les grandes similitudes spectrales entre les deux groupes de patients.

#### *Analyse SVM*

Le modèle SVM utilisant les dérivées secondes a été construit sur les mêmes groupes de patients que ceux du modèle SVM appliqué sur les données spectrales sans dérivation (Tableau 10). Les résultats sont présentés en détail le Tableau 13.

Les taux globaux de bonne classification étaient de 50,15 % pour l'analyse des médianes, de 52,53% pour l'analyse des moyennes et de 55,49% pour l'analyse des spectres. Ces résultats étaient similaires à ceux obtenus avec les données spectrales non dérivées.

### *Chapitre 3 : Résultats et discussion*

**Tableau 13 :** Classification SVM sur la médiane, la moyenne et l'ensemble des spectres des patients cirrhotiques avec et sans CHC. Analyse avec une dérivation seconde sur la région spectrale 900-4000 cm<sup>-1</sup>.

	Jeu d'entraînement			Jeu de validation			Jeu test		
	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres
Nombre de CPs	45	20	33	45	20	33	45	20	33
Patients avec CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>68</b>	<b>68</b>	<b>68</b>
<i>Bien classés</i>	<i>127</i>	<i>122</i>	<i>117</i>	<i>50</i>	<i>56</i>	<i>52</i>	<i>46</i>	<i>47</i>	<i>45</i>
<i>Mal classés</i>	<i>33</i>	<i>38</i>	<i>43</i>	<i>30</i>	<i>24</i>	<i>28</i>	<i>22</i>	<i>21</i>	<i>23</i>
Patients sans CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>269</b>	<b>269</b>	<b>269</b>
<i>Bien classés</i>	<i>123</i>	<i>113</i>	<i>103</i>	<i>48</i>	<i>57</i>	<i>51</i>	<i>123</i>	<i>130</i>	<i>142</i>
<i>Mal classés</i>	<i>37</i>	<i>47</i>	<i>57</i>	<i>32</i>	<i>23</i>	<i>29</i>	<i>146</i>	<i>139</i>	<i>127</i>
Sensibilité (%)	<b>79,38</b>	<b>76,25</b>	<b>73,13</b>	<b>62,5</b>	<b>70</b>	<b>65</b>	<b>67,65</b>	<b>69,12</b>	<b>66,18</b>
Spécificité (%)	<b>76,88</b>	<b>70,63</b>	<b>64,38</b>	<b>60</b>	<b>71,25</b>	<b>63,75</b>	<b>45,72</b>	<b>48,33</b>	<b>52,79</b>
Taux de patients bien classés (%)	<b>78,13</b>	<b>73,44</b>	<b>68,76</b>	<b>61,25</b>	<b>70,63</b>	<b>64,38</b>	<b>50,15</b>	<b>52,53</b>	<b>55,49</b>

### *Chapitre 3 : Résultats et discussion*

---

Comme pour l'analyse précédente, l'analyse SVM a été complétée par une analyse PLS-DA.

#### *Analyse PLS-DA*

L'analyse PLS-DA a été appliquée sur les médianes, les moyennes et l'ensemble des spectres des 817 patients. Le Tableau 14 présente le détail des résultats obtenus sur la région spectrale 900-4000  $\text{cm}^{-1}$  avec dérivation seconde des données spectrales.

Les taux globaux de bonne classification obtenus étaient de 52.19%, 57.01% et 57.26% pour l'analyse des médianes, des moyennes et de tous les spectres des patients respectivement.

⇒ **En résumé, dans cette première partie de l'étude menée sur l'ensemble du spectre IR (900-4000  $\text{cm}^{-1}$ ), les deux méthodes de classification utilisées (SVM et PLS-DA) ne permettaient pas de distinguer efficacement les deux groupes de patients cirrhotiques avec et sans CHC. Le taux global d'exactitude diagnostique variait entre 46 et 57% sans différence significative selon l'application ou non d'une dérivation seconde des données spectrales et quelle que soit la base de données utilisée (médianes, moyennes et l'ensemble des spectres).**

### *Chapitre 3 : Résultats et discussion*

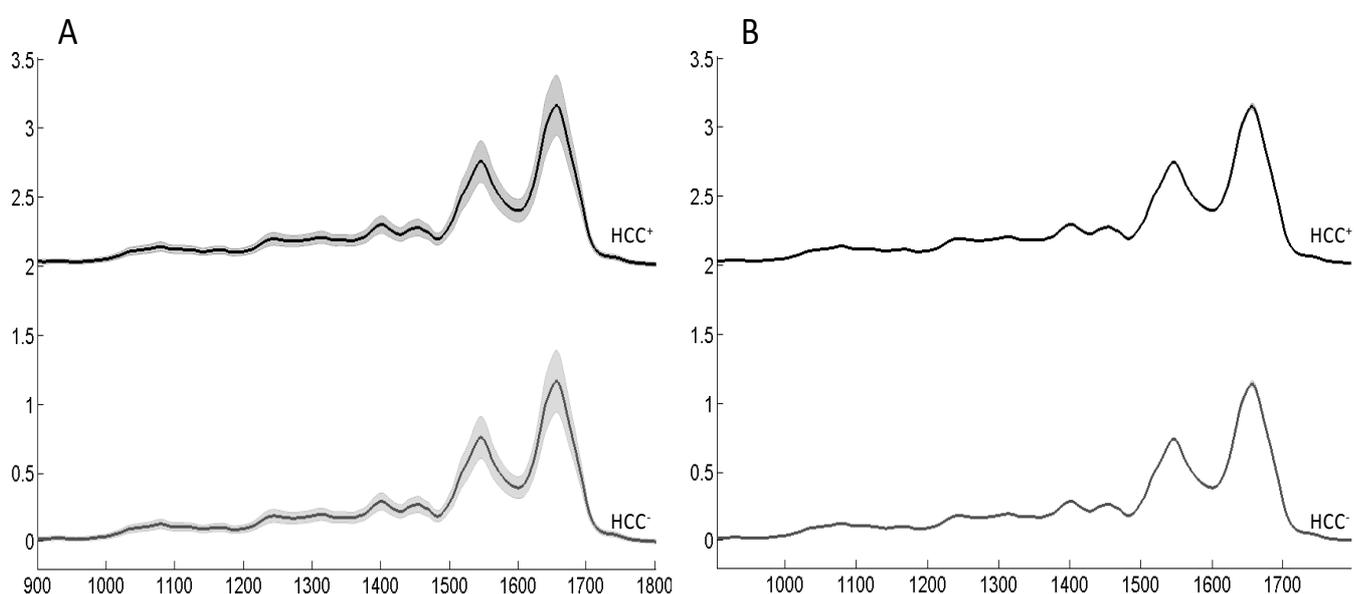
**Tableau 14 :** Classification PLS-DA sur la médiane, la moyenne et l'ensemble des spectres des patients cirrhotiques avec et sans CHC. Analyse avec dérivation seconde sur la région spectrale 900-4000 cm<sup>-1</sup>.

	Jeu d'entraînement			Jeu de validation			Jeu test		
	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres
<i>Nombre de dimensions</i>	16	20	21	16	20	21	16	20	21
Patients avec CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>68</b>	<b>68</b>	<b>68</b>
<i>Bien classés</i>	<i>120</i>	<i>127</i>	<i>120</i>	<i>55</i>	<i>58</i>	<i>57</i>	<i>46</i>	<i>49</i>	<i>50</i>
<i>Mal classés</i>	<i>40</i>	<i>33</i>	<i>40</i>	<i>25</i>	<i>22</i>	<i>23</i>	<i>22</i>	<i>19</i>	<i>18</i>
Patients sans CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>269</b>	<b>269</b>	<b>269</b>
<i>Bien classés</i>	<i>121</i>	<i>128</i>	<i>124</i>	<i>53</i>	<i>52</i>	<i>50</i>	<i>129</i>	<i>143</i>	<i>143</i>
<i>Mal classés</i>	<i>39</i>	<i>32</i>	<i>36</i>	<i>27</i>	<i>28</i>	<i>30</i>	<i>140</i>	<i>126</i>	<i>126</i>
Sensibilité (%)	<b>75</b>	<b>79.4</b>	<b>75.1</b>	<b>68.8</b>	<b>72.5</b>	<b>70.8</b>	<b>67.6</b>	<b>72.1</b>	<b>73.7</b>
Spécificité (%)	<b>75.6</b>	<b>80</b>	<b>77.3</b>	<b>66.3</b>	<b>65</b>	<b>62.8</b>	<b>48.3</b>	<b>53.2</b>	<b>53.1</b>
Taux de patients bien classés (%)	<b>75.3</b>	<b>79.7</b>	<b>76.2</b>	<b>67.55</b>	<b>68.75</b>	<b>66.8</b>	<b>52.19</b>	<b>57.01</b>	<b>57.26</b>

### 2. Analyse des données IR sur la région 900-1800 cm<sup>-1</sup>

Les mêmes traitements effectués précédemment ont été appliqués sur la région 900-1800 cm<sup>-1</sup> afin d'en comparer les résultats avec ceux obtenus en analysant la totalité du spectre infrarouge.

Après le test qualité, la zone spectrale 900-1800 cm<sup>-1</sup> qui inclue la majorité des pics caractéristiques des protéines a été sélectionnée puis prétraitée par la méthode EMSC.



**Figure 45 : Spectres moyens IRTF des sérums de patients avec et sans CHC.**

(A) spectres bruts, (B) spectres prétraités par EMSC avec un polynôme d'ordre 8 sur la région spectrale 900-1800 cm<sup>-1</sup>.

La Figure 45 présente les spectres moyens des deux groupes de patients avec et sans CHC. Les spectres apparaissent similaires et ne peuvent être différenciés par un examen visuel simple justifiant le recours à des analyses chimiométriques sophistiquées.

#### a. Analyse des spectres IRTF – sans dérivation

Les données spectrales ont été prétraitées selon les paramètres suivants :

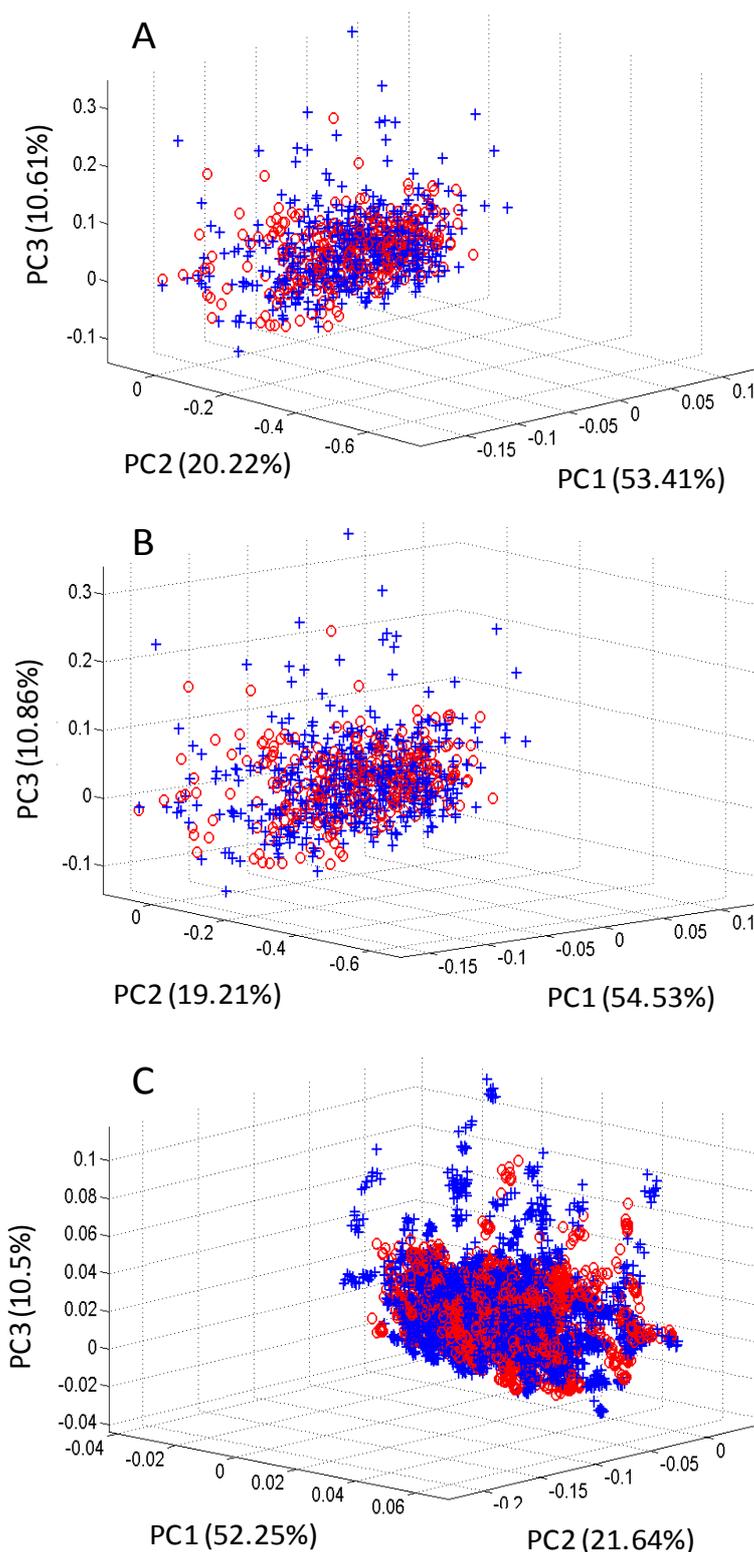
## *Chapitre 3 : Résultats et discussion*

---

- Lissage des spectres : polynôme d'ordre 3 et 9 points
- Dérivée : 0
- Correction EMSC : polynôme d'ordre 8
- Soustraction du spectre moyen

### *Analyse en composantes principales*

L'ACP a été appliquée sur les spectres médians (320 spectres), moyens (320 spectres) et l'ensemble des spectres (3180 spectres) du groupe d'entraînement. La Figure 46 montre la représentation tridimensionnelle de l'ACP en utilisant les 3 composantes les plus explicatives de la variance.



**Figure 46 : Analyse en composante principale des données IRTF des patients avec (rouge) et sans (bleu) CHC.**

A) Analyse des médianes. B) Analyse des moyennes. B) Analyse de l'ensemble des spectres de chaque patient sur la région spectrale  $900-1800\text{ cm}^{-1}$  sans dérivation seconde.

### *Chapitre 3 : Résultats et discussion*

---

Comme le montre la Figure 46, la méthode ACP ne permet pas de séparer les sérums des patients avec et sans CHC.

#### *Analyse SVM*

L'analyse SVM a été faite sur les CPs les plus discriminantes représentant 99,99% de la variance totale. Les jeux d'entraînement, de validation et de test étaient les mêmes que ceux utilisés dans l'analyse de la totalité du spectre IR (Tableau 10). Les résultats des différentes analyses sont détaillés dans le Tableau 15.

Les taux globaux de bonne classification étaient de 50,44 % pour l'analyse des médianes, de 56,97% pour l'analyse des moyennes et de 56,38% pour l'analyse de l'ensemble des spectres. Ces résultats sont similaires à ceux obtenus avec l'analyse SVM de la totalité du spectre 900-4000  $\text{cm}^{-1}$  (Tableau 15).

#### *Analyse PLS-DA*

L'analyse PLS-DA a été appliquée sur les médianes, les moyennes et l'ensemble des spectres des 817 patients. Le Tableau 16 présente le détail des résultats obtenus sur la région spectrale 900-1800  $\text{cm}^{-1}$  sans dérivation seconde des données spectrales.

Les taux globaux de bonne classification étaient de 50.18%, 53.14% et 54.19% pour l'analyse des médianes, des moyennes et de tous les spectres des patients respectivement.

### *Chapitre 3 : Résultats et discussion*

**Tableau 15 :** Classification SVM sur la médiane, la moyenne et l'ensemble des spectres des patients cirrhotiques avec et sans CHC. Analyse sans dérivation seconde sur la région spectrale 900-1800 cm<sup>-1</sup>.

	Jeu d'entraînement			Jeu de validation			Jeu test		
	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres
Nombre de CPs	2	9	3	2	9	3	2	9	3
Patients avec CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>68</b>	<b>68</b>	<b>68</b>
<i>Bien classés</i>	79	112	109	53	48	52	30	42	46
<i>Mal classés</i>	81	48	51	27	32	28	38	26	22
Patients sans CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>269</b>	<b>269</b>	<b>269</b>
<i>Bien classés</i>	105	90	94	49	49	50	140	150	144
<i>Mal classés</i>	55	70	66	31	31	30	129	119	125
Sensibilité (%)	<b>49,38</b>	<b>70</b>	<b>68,13</b>	<b>66,25</b>	<b>60</b>	<b>65</b>	<b>44,12</b>	<b>61,76</b>	<b>67,65</b>
Spécificité (%)	<b>65,63</b>	<b>56,25</b>	<b>58,75</b>	<b>61,25</b>	<b>61,25</b>	<b>62,5</b>	<b>52,04</b>	<b>55,75</b>	<b>53,53</b>
Taux de patients bien classés (%)	<b>57,51</b>	<b>63,13</b>	<b>63,44</b>	<b>63,75</b>	<b>60,63</b>	<b>63,75</b>	<b>50,44</b>	<b>56,97</b>	<b>56,38</b>

### Chapitre 3 : Résultats et discussion

**Tableau 16 :** Classification PLS-DA sur la médiane, la moyenne et l'ensemble des spectres des patients cirrhotiques avec et sans CHC. Analyse sans dérivation seconde sur la région spectrale 900-1800  $\text{cm}^{-1}$ .

	Jeu d'entraînement			Jeu de validation			Jeu test		
	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres
Nombre dimensions	19	24	22	19	24	22	19	24	22
Patients avec CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>68</b>	<b>68</b>	<b>68</b>
<i>Bien classés</i>	<i>109</i>	<i>110</i>	<i>112</i>	<i>51</i>	<i>56</i>	<i>54</i>	<i>40</i>	<i>47</i>	<i>45</i>
<i>Mal classés</i>	<i>51</i>	<i>50</i>	<i>48</i>	<i>29</i>	<i>24</i>	<i>26</i>	<i>28</i>	<i>21</i>	<i>23</i>
Patients sans CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>269</b>	<b>269</b>	<b>269</b>
<i>Bien classés</i>	<i>113</i>	<i>110</i>	<i>118</i>	<i>45</i>	<i>48</i>	<i>51</i>	<i>129</i>	<i>132</i>	<i>138</i>
<i>Mal classés</i>	<i>47</i>	<i>50</i>	<i>42</i>	<i>35</i>	<i>32</i>	<i>29</i>	<i>140</i>	<i>137</i>	<i>131</i>
Sensibilité (%)	<b>68.1</b>	<b>68.8</b>	<b>69.7</b>	<b>63.8</b>	<b>70</b>	<b>68</b>	<b>58.8</b>	<b>69.1</b>	<b>66</b>
Spécificité (%)	<b>70.6</b>	<b>68.8</b>	<b>74</b>	<b>56.3</b>	<b>60</b>	<b>64.2</b>	<b>48</b>	<b>49.1</b>	<b>51.2</b>
Taux de patients bien classés (%)	<b>69.35</b>	<b>68.8</b>	<b>71.85</b>	<b>60.05</b>	<b>65</b>	<b>66.1</b>	<b>50.18</b>	<b>53.14</b>	<b>54.19</b>

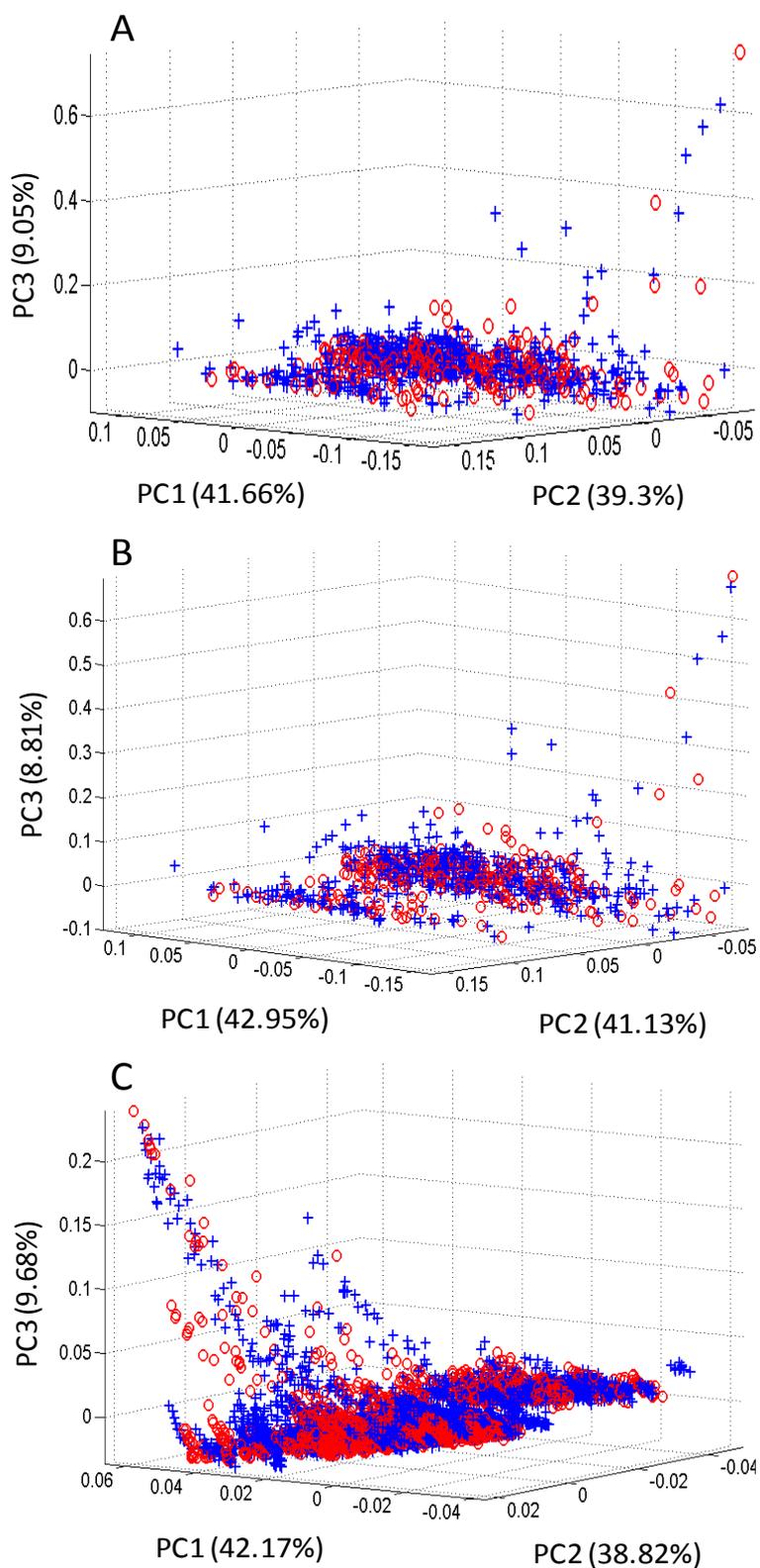
### **b. Analyse des spectres IRTF – avec la dérivation seconde**

Les données spectrales ont été prétraitées selon les paramètres suivants :

- Lissage des spectres : polynôme d'ordre 3 et 9 points
- Dérivée : 2
- Correction EMSC : polynôme d'ordre 2
- Soustraction du spectre moyen

### **Analyse en composantes principales**

L'ACP a été réalisée sur le groupe des spectres médians (320 spectres) et moyens (320 spectres) et sur l'ensemble des spectres (3180 spectres) de chaque patient. La Figure 47 montre la représentation tridimensionnelle de la méthode ACP en utilisant les 3 composantes les plus explicatives de la variance.



**Figure 47 : Analyse en composante principale des données IRTF des patients avec (rouge) et sans (bleu) CHC.**

A) Analyse des médianes. B) Analyse des moyennes. B) Analyse de l'ensemble des spectres de chaque patient sur la région spectrale  $900-1800\text{ cm}^{-1}$  avec dérivée seconde.

### *Chapitre 3 : Résultats et discussion*

La Figure 47 montre que l'ACP ne permet pas de séparer les patients avec et sans carcinome hépatocellulaire.

#### *Analyse SVM*

L'analyse SVM a été appliquée sur les données en prenant les CPs les plus discriminantes expliquant 99.99% de la variance totale. Les jeux d'entraînement, de validation et de test étaient les mêmes que ceux utilisés pour les analyses précédentes (Tableau 10).

Les taux globaux de bonne classification étaient de 51,04 % pour l'analyse des médianes, de 61,42% pour l'analyse des moyennes et de 51,93% pour l'analyse de l'ensemble des spectres. Ces résultats sont similaires à ceux obtenus avec l'analyse SVM de la totalité du spectre 900-4000  $\text{cm}^{-1}$  (Tableau 17).

#### *Analyse PLS-DA*

Une analyse PLS-DA a été appliquée sur les données IR pour discriminer les patients avec et sans CHC. Le Tableau 18 présente les résultats obtenus pour l'analyse médiane, moyenne et l'ensemble des spectres.

Les taux globaux de bonne classification étaient de 52,49 % pour l'analyse des médianes, de 55,77% pour l'analyse des moyennes et de 55,01% pour l'analyse de l'ensemble des spectres. Ces résultats sont similaires à ceux obtenus avec l'analyse PLS-DA de la totalité du spectre 900-4000  $\text{cm}^{-1}$ .

**=>En résumé, dans cette deuxième partie de l'étude, L'analyse des données IR (ACP, SVM et PLS-DA) faite sur la région 900-1800  $\text{cm}^{-1}$  avec et sans dérivation seconde n'améliore pas significativement les résultats obtenus sur les spectres entiers. Le taux global de bonne classification variait de 50 à 61% versus 46 à 57 % pour l'analyse des spectres entiers.**

### *Chapitre 3 : Résultats et discussion*

**Tableau 17:** Classification SVM sur la médiane, la moyenne et l'ensemble des spectres des patients cirrhotiques avec et sans CHC. Analyse avec dérivation seconde sur la région spectrale 900-1800 cm<sup>-1</sup>.

	Jeu d'entraînement			Jeu de validation			Jeu test		
	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres
Nombre de CPs	41	8	84	41	8	84	41	8	84
Patients avec CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>68</b>	<b>68</b>	<b>68</b>
<i>Bien classés</i>	<i>121</i>	<i>104</i>	<i>121</i>	<i>49</i>	<i>50</i>	<i>52</i>	<i>36</i>	<i>44</i>	<i>59</i>
<i>Mal classés</i>	<i>39</i>	<i>56</i>	<i>39</i>	<i>31</i>	<i>30</i>	<i>28</i>	<i>32</i>	<i>24</i>	<i>9</i>
Patients sans CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>269</b>	<b>269</b>	<b>269</b>
<i>Bien classés</i>	<i>100</i>	<i>118</i>	<i>105</i>	<i>46</i>	<i>54</i>	<i>52</i>	<i>136</i>	<i>163</i>	<i>116</i>
<i>Mal classés</i>	<i>60</i>	<i>42</i>	<i>55</i>	<i>34</i>	<i>26</i>	<i>28</i>	<i>133</i>	<i>106</i>	<i>153</i>
Sensibilité (%)	<b>75,63</b>	<b>65</b>	<b>75,63</b>	<b>61,25</b>	<b>62,5</b>	<b>65</b>	<b>52,94</b>	<b>64,71</b>	<b>86,76</b>
Spécificité (%)	<b>62,5</b>	<b>73,75</b>	<b>65,63</b>	<b>57,5</b>	<b>62,5</b>	<b>65</b>	<b>50,56</b>	<b>60,59</b>	<b>43,12</b>
Taux de patients bien classés (%)	<b>69,07</b>	<b>69,38</b>	<b>70,63</b>	<b>59,38</b>	<b>65,00</b>	<b>65,00</b>	<b>51,04</b>	<b>61,42</b>	<b>51,93</b>

### *Chapitre 3 : Résultats et discussion*

**Tableau 18 :** Classification PLS-DA sur la médiane, la moyenne et l'ensemble des spectres des patients cirrhotiques avec et sans CHC. Analyse avec dérivation seconde sur la région spectrale 900-1800  $\text{cm}^{-1}$ .

	Jeu d'entraînement			Jeu de validation			Jeu test		
	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres	Médiane	Moyenne	Tous les spectres
Nombre de dimensions	25	24	23	25	24	23	25	24	23
Patients avec CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>68</b>	<b>68</b>	<b>68</b>
<i>Bien classés</i>	<i>126</i>	<i>122</i>	<i>112</i>	<i>53</i>	<i>58</i>	<i>54</i>	<i>40</i>	<i>47</i>	<i>47</i>
<i>Mal classés</i>	<i>34</i>	<i>38</i>	<i>48</i>	<i>27</i>	<i>22</i>	<i>26</i>	<i>28</i>	<i>21</i>	<i>21</i>
Patients sans CHC	<b>160</b>	<b>160</b>	<b>160</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>269</b>	<b>269</b>	<b>269</b>
<i>Bien classés</i>	<i>123</i>	<i>119</i>	<i>120</i>	<i>56</i>	<i>51</i>	<i>51</i>	<i>137</i>	<i>141</i>	<i>138</i>
<i>Mal classés</i>	<i>37</i>	<i>41</i>	<i>40</i>	<i>24</i>	<i>29</i>	<i>29</i>	<i>132</i>	<i>128</i>	<i>131</i>
Sensibilité (%)	<b>78.8</b>	<b>76.3</b>	<b>70.3</b>	<b>66.3</b>	<b>72.5</b>	<b>67.6</b>	<b>58.8</b>	<b>69.1</b>	<b>69.7</b>
Spécificité (%)	<b>76.9</b>	<b>74.4</b>	<b>74.8</b>	<b>70</b>	<b>63.8</b>	<b>63.7</b>	<b>50.9</b>	<b>52.4</b>	<b>51.3</b>
Taux de patients bien classés (%)	<b>77.85</b>	<b>75.35</b>	<b>72.55</b>	<b>68.15</b>	<b>68.15</b>	<b>65.65</b>	<b>52.49</b>	<b>55.77</b>	<b>55.01</b>

## Chapitre 3 : Résultats et discussion

### Discussion des résultats

Cette étude menée dans une large cohorte de 817 patients n'a pas confirmé l'hypothèse de travail. L'analyse spectroscopique IR du sérum complet combinée à des méthodes de classification sophistiquées n'a pas permis dans la large population étudiée de différencier efficacement les patients cirrhotiques avec et sans CHC. Ces résultats sont en désaccord avec ceux obtenus précédemment dans une étude pilote sur un nombre limité de sujets qui rapportait un taux d'exactitude de 82% à 86% (91), ce qui soulève un certain nombre de questions méthodologiques. Les résultats obtenus dans le présent travail apparaissent robustes en raison du nombre élevé de patients inclus dans l'étude et de l'utilisation de deux méthodes de classification différentes, la méthode SVM et la PLS-DA, appliquées sur un jeu d'entraînement, un jeu de validation interne et un jeu de validation externe. Le taux global de bonne classification variait de 46 à 61% selon l'algorithme utilisé (Tableau 19).

**Tableau 19** : Résultats des analyses SVM et PLS-DA sur la cohorte de 817 patients.

	Analyse SVM			Analyse PLS-DA		
	Médiane	Moyenne	Tous les Spectres	Médiane	moyenne	Tous les spectres
<b>900-4000 cm<sup>-1</sup></b>						
<i>Sans dérivé</i> (%)	46.3	55.8	57	54.3	53.1	53.9
<i>Avec dérivé</i> seconde (%)	50.4	57	56.4	50.2	53.1	54.2
<b>900-1800 cm<sup>-1</sup></b>						
<i>Sans dérivé</i> (%)	50.1	52.5	55.5	52.2	57	57.3
<i>Avec dérivé</i> seconde(%)	51	61.4	51.9	52.5	55.8	55

Il n'existait pas de différence significative entre les deux méthodes de classification, quelle que soit la zone spectrale analysée. La dérivé seconde, souvent utilisée pour accentuer le

### *Chapitre 3 : Résultats et discussion*

---

contraste spectral, ne permettait pas non plus d'améliorer significativement la performance diagnostique des deux classifieurs (Tableau 19).

Dans les études spectroscopiques, le ratio entre le nombre habituellement faible d'échantillons et le nombre très élevé de variables analysées (nombre de longueurs d'onde dans les spectres étudiés) expose au risque de surentrainement lié au phénomène de Hughes (95) encore appelé « malédiction de la dimensionnalité » (« curse of dimensionality »). Dans notre travail, l'ACP des données spectrales a permis de réduire le nombre de dimensions par une transformation linéaire préservant la variance maximale des données. D'autre part, le grand nombre de patients dans l'étude de validation (n=817) nous a permis d'appliquer les techniques de classification SVM et PLS-DA en formant 3 groupes de patients pour l'entraînement du modèle, sa validation interne et sa validation externe. L'application de cette méthodologie permet de réduire le risque de surentrainement mentionné précédemment. Dans l'étude pilote antérieure (91), l'ACP a également été utilisée pour réduire le nombre de dimensions mais le petit nombre de patients (n=79) n'a pas permis la constitution de ces trois groupes (entraînement, validation interne et validation externe) et l'analyse a été réalisée en utilisant la méthode de validation croisée avec LOOCV. Les résultats positifs de cette étude pilote n'ont pas été confirmés par l'étude de validation sur un grand nombre de patients suggérant que la méthode de classification avec validation croisée LOOCV expose à un risque de surentrainement (« overfitting »).

En conclusion, la spectroscopie IR appliquée au sérum dans les conditions décrites dans ce travail ne permet pas d'envisager une application clinique du classifieur.

## CONCLUSIONS ET PERSPECTIVES

## *Conclusions et perspectives*

Le carcinome hépatocellulaire est le 6ème cancer le plus fréquent dans le monde et la 3ème cause de mortalité par cancer. C'est le principal cancer primitif du foie et il est associé à un très mauvais pronostic. L'identification de nouveaux marqueurs sériques est cruciale pour obtenir un diagnostic précoce et améliorer la survie.

Cette thèse est la première à évaluer le potentiel de la spectroscopie vibrationnelle IRTF et Raman couplée à des analyses statistiques multivariées (SVM, PLS-DA) pour discriminer des échantillons sériques de patients cirrhotiques avec et sans CHC.

Dans un premier temps, nous avons caractérisé la signature spectrale du sérum entier des patients cirrhotiques avec et sans CHC en utilisant la microspectroscopie Raman. Pour cela, une étude pilote a été menée sur 71 patients cirrhotiques avec et sans cancer. Deux approches ont été évaluées, le sérum en goutte séchée et le sérum lyophilisé. En analysant les données par une méthode chimiométrique telle que la SVM-LOOCV, nous avons obtenu des performances diagnostiques très prometteuses, à savoir un taux global de bonne classification variant entre 85 et 90%. Nous observons aussi qu'il n'y a pas de différences significatives entre les résultats obtenus sur goutte séchée et lyophilisat à partir des trois séries de données, la médiane, la moyenne, ou tous les spectres individuels de chaque patient. Ce résultat reflète la bonne reproductibilité des spectres Raman enregistrés avec les deux préparations. Cette étude pilote suggère aussi que la microspectroscopie Raman appliquée à la goutte séchée du sérum est plus appropriée car elle implique un protocole de préparation simplifié.

Dans le deuxième volet de ce travail, nous avons étudié la signature spectrale du sérum entier de 79 patients cirrhotiques avec et sans cancer par spectroscopie IRTF à haut débit. Dans cette étude pilote, nous avons obtenu, en utilisant également une analyse SVM-LOOCV, un taux global de bonne classification de 82 à 86%. Comme pour l'étude en microspectroscopie Raman, nous avons aussi démontré à travers ce travail, qu'il n'y a pas de différences significatives entre les résultats obtenus en analysant la médiane, la moyenne, ou tous les spectres individuels de chaque patient. Ceci reflète la bonne reproductibilité des spectres IRTF enregistrés.

Dans la perspective du développement d'un test de diagnostic clinique, la spectroscopie IR paraît plus appropriée que la spectroscopie Raman. En effet, elle opère déjà en mode "haut débit" et offre la possibilité de mesurer plusieurs échantillons (jusqu'à 384) avec un temps de mesure d'environ 30s par spectre. L'analyse IR est simple, directe et ne nécessite aucune préparation préalable de l'échantillon sérique. Ces deux études pilotes montrent aussi que les

## *Conclusions et perspectives*

---

méthodes chimiométriques avancées sont nécessaires pour une exploitation optimale des données spectrales. L'analyse en composante principale seule n'a pas permis de discriminer les deux groupes de patients. Cependant, l'utilisation des scores de composantes principales comme données d'entrée pour la classification SVM a permis de différencier les deux groupes de patients avec une bonne performance diagnostique.

Dans le troisième volet de cette thèse, compte tenu des résultats très prometteurs obtenus dans l'étude pilote IR et afin de les confirmer, une étude de validation IR a été réalisée sur une large cohorte de 817 patients cirrhotiques avec et sans CHC. Deux modèles de classification SVM et PLS-DA ont été construits avec un groupe d'entraînement, un groupe de validation interne et un groupe de validation externe. Le taux global de bonne classification obtenu était entre 46 et 61 %. Ce résultat ne confirme donc pas celui obtenu dans l'étude pilote IR sur 79 patients. Cette discordance pourrait s'expliquer soit par un effet de surentrainement (« overfitting ») de la méthode LOOCV sur les petits nombres de patients, soit par l'existence d'une hétérogénéité au sein de la grande cohorte de patients. Pour circonvenir le problème de surentrainement, une procédure de simulations de Monte Carlo pourrait être systématiquement mise en œuvre pour vérifier la performance du classifieur, à savoir que ce dernier sépare bien les spectres sur la base des caractéristiques biochimiques du sérum et non sur le bruit spectral. En effet, le test Monte Carlo est basé sur la formation de deux populations de patients distribuées aléatoirement sans tenir compte de la pathologie (avec ou sans CHC). La performance diagnostique du classifieur serait évaluée sur les deux populations ainsi formées et la procédure répétée N fois de manière à obtenir une distribution des performances diagnostiques des classifieurs. Si les valeurs diagnostiques sont centrées à 50%, le modèle de LOOCV n'est pas surentrainé. En revanche, si les valeurs sont centrées à 80% ou plus, on peut conclure que le modèle de LOOCV est surestimé. Ce test permettrait donc d'éviter les résultats faussement positifs des études pilotes.

Une perspective pour s'affranchir de l'hétérogénéité de la cohorte serait d'améliorer la classification des patients en analysant des populations plus homogènes. Les nouveaux groupes de patients seraient donc formés en tenant compte de la sévérité de la maladie, de la cause de cirrhose et des différents facteurs métaboliques. D'autre part, il serait intéressant de coupler différentes approches d'analyse pour la classification de nos données spectrales comme les données de spectroscopie IR et les données de spectroscopie Raman ou encore les données de spectroscopie vibrationnelle avec celles obtenues en protéomique par exemple.

## *Conclusions et perspectives*

---

Les données clinico-biologiques peuvent aussi être intégrées au classifieur pour prendre en compte différentes informations comme le sexe, l'âge...

Toutes les études IRTF ont été réalisées sur le sérum complet qui contient des milliers de molécules différentes. Dans ces conditions expérimentales, le spectre IR reflète la complexité moléculaire du sérum avec de nombreuses superpositions de pics d'absorbance qui peuvent masquer les informations d'intérêt. Une ultrafiltration du sérum éliminant de façon plus ou moins sélective les plus grosses molécules telles que l'albumine et les globulines pourrait permettre de révéler des caractéristiques spectrales sous-jacentes potentiellement discriminantes. Néanmoins, cette procédure a l'inconvénient d'introduire des étapes supplémentaires dans le protocole de préparation des échantillons sériques.

Une deuxième voie qui pourrait être envisagée pour l'analyse du sérum est l'approche par l'utilisation des substrats SERS fonctionnalisés. En effet, le signal Raman des molécules serait très fortement exalté lorsque celles-ci sont au contact des structures nanométriques d'or ou d'argent. Cette technique ouvre des perspectives en termes de sélectivité moléculaire et de gain de sensibilité. La principale limite de cette approche est la reproductibilité de la technique. L'approche SERS doit être bien cadrée car il est connu que la préparation des substrats SERS influe sur la reproductibilité des spectres.

Au total, notre étude n'a évalué qu'un aspect de l'approche biophotonique du sérum comme outil diagnostique et des approches complémentaires plus ciblées restent à explorer tant sur le versant technologique que sur le versant chimiométrique.

## RÉFÉRENCES BIBLIOGRAPHIQUES

## *Références bibliographiques*

---

1. Ferlay J, Shin H, Bray F, Forman D, Mathers C, Parkin D. *Int J Cancer*. 2010;127(12):2893-917.
2. Belot A, Grosclaude P, Bossard N, Jouglu E, Benhamou E, Delafosse P, et al. Cancer incidence and mortality in France over the period 1980-2005. *Rev Epidemiol Sante Publique*. 2008;56(3):159-75.
3. European Association for the study of the Liver and European Organisation for Research and Treatment of Cancer. *JHepatol*. 2012;56:908-43.
4. Chignard N, Beretta L. Proteomics for hepatocellular carcinoma marker discovery. *Gastroenterology*. 2004;127(5 Suppl 1): S120-5.
5. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh J, Comber H, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer Prev*. 2013;49(6):1374-403.
6. The French METAVIR Cooperative Study Group. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. *Hepatology*. 1994;20:15-20.
7. Farazi P, DePinho R. Hepatocellular carcinoma pathogenesis: from genes to environment. *Nat Rev Cancer*. 2006; 6:674-87.
8. Trevisani F, D'Intino PE, Morselli-Labate AM, Mazzella G, Accogli E, Caraceni P, et al. Serum alpha-fetoprotein for diagnosis of hepatocellular carcinoma in patients with chronic liver disease: influence of HBsAg and anti-HCV status. *J Hepatol*. 2001;34(4):570-5.
9. Bruix J, Sherman M, Diseases. AAftSoL. Management of hepatocellular carcinoma: an update. *Hepatology*. 2011;53(3):1020-2.
10. Yamashita F, Tanaka M, Satomura S, Tanikawa K. Prognostic significance of Lens culinaris agglutinin A-reactive alpha-fetoprotein in small hepatocellular carcinomas. *Gastroenterology*. 1996;111(4):996-1001.
11. Lok A, Sterling R, Everhart J, Wright E, Hoefs J, Di Bisceglie A, et al. Des-gamma-carboxy Prothrombin and Alpha fetoprotein as Biomarkers for the Early Detection of Hepatocellular Carcinoma. *Gastroenterology*. 2010;138(2):493-502.

## *Références bibliographiques*

12. Marrero J, Su G, Wei W, Emick D, Conjeevaram H, Fontana R, et al. Des-gamma carboxyprothrombin can differentiate hepatocellular carcinoma from nonmalignant chronic liver disease in american patients. *Hepatology*. 2003;37(5):1114-21.
13. Miyaaki H, Nakashima O, Kurogi M, Eguchi K, Kojiro M. Lens culinaris agglutinin-reactive alpha-fetoprotein and protein induced by vitamin K absence II are potential indicators of a poor prognosis: a histopathological study of surgically resected hepatocellular carcinoma. *J Gastroenterol*. 2007;42(12):962-8.
14. Sherman M. Serological surveillance for hepatocellular carcinoma: time to quit. *J Hepatol*. 2010;52(4):614-5.
15. Hsu H, Cheng W, Lai P. Cloning and expression of a developmentally regulated transcript MXR7 in hepatocellular carcinoma: biological significance and temporospatial distribution. *Cancer Res* 57. 1997:5179-84.
16. Capurro M, Wanless I, Sherman M, Deboer G, Shi W, Miyoshi E, et al. Glypican-3: a novel serum and histochemical marker for hepatocellular carcinoma. *Gastroenterology*. 2003;125(1):89-97.
17. Deugnier Y, David V, Brissot P, Mabo P, Delamaire D, Messner M, et al. Serum alpha-L-fucosidase: a new marker for the diagnosis of primary hepatic carcinoma. *Hepatology*. 1984;4(5):889-92.
18. Draux F, Jeannesson P, Gobinet C, Sule-Suso J, Pijanka J, Sandt C, et al. IR spectroscopy reveals effect of non-cytotoxic doses of anti-tumour drug on cancer cells. *Anal Bioanal Chem*. 2009;395(7):2293-301.
19. Pijanka J, Kohler A, Yang Y, Dumas P, Chio-Srichan S, Manfait M, et al. Spectroscopic signatures of single, isolated cancer cell nuclei using synchrotron infrared microscopy. *Analyst*. 2009;134(6):1176-81.
20. Ramesh J, Kapelushnik J, Mordehai J, Moser A, Huleihel M, Erukhimovitch V, et al. Novel methodology for the follow-up of acute lymphoblastic leukemia using FTIR microspectroscopy. *J Biochem Biophys Methods*. 2002;51(3):251-61.
21. Bellisola G, Della Peruta M, Vezzalini M, Moratti E, Vaccari L, Birarda G, et al. tracking infrared signatures of drugs in cancer cells by Fourier transform microspectroscopy. *Analyst*. 2010;135(12):3077-86.

## *Références bibliographiques*

---

22. Li Q, Wang W, Ling X, Wu J. Detection of Gastric Cancer with Fourier Transform Infrared Spectroscopy and Support Vector Machine Classification. *BioMed Research International*. 2013.
23. Gajjar K, Trevisan J, Owens G, Keating P, Wood N, Stringfellow H, 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):3917-26.
24. Mackanos M, Contag C. FTIR microspectroscopy for improved prostate cancer diagnosis. *Trends Biotechnol* 2009;27(12):661-3.
25. Gazi E, Dwyer J, Gardner P, Ghanbari-Siahkali A, Wade A, J. M, et al. Applications of Fourier transform infrared microspectroscopy in studies of benign prostate and prostate cancer. A pilot study. *J Pathol*. 2003;201(1):99-108.
26. Tosi G, Conti C, Giorgini E, P. F, Garavaglia M, Sabbatini S, et al. FTIR microspectroscopy of melanocytic skin lesions: a preliminary study. *Analyst*. 2010;135(12):3213-9.
27. Wang T, Triadafilopoulos G, Crawford J, Dixon L, Bhandari T, Sahbaie P, et al. Detection of endogenous biomolecules in Barrett's esophagus by Fourier transform infrared spectroscopy. *Proc Natl Acad Sci U S A*. 2007;104(40):15864-9.
28. Li Q, Yang L, Ling X, Wang J, Zhou X, Shi J, et al. Application of the SIMCA method to cancer diagnosis with Fourier-transform infrared spectroscopy. *Guang Pu Xue Yu Guang Pu Fen Xi*. 2004;24(4):414-7.
29. Yano K, Ohoshima S, Gotou Y, Kumaido K, Moriguchi T, Katayama H. Direct measurement of human lung cancerous and noncancerous tissues by fourier transform infrared microscopy: can an infrared microscope be used as a clinical tool? *Anal Biochem*. 2000;287(2):218-25.
30. Argov S, Sahu R, Bernshtain E, Salman A, Shohat G, Zelig U, et al. Inflammatory bowel diseases as an intermediate stage between normal and cancer: a FTIR-microspectroscopy approach. *Biopolymers*. 2004;75(5):384-92.
31. Du J, Shi J, Sun X, Wang J, Xu Y, Wu J, et al. Fourier transform infrared spectroscopy of gallbladder carcinoma cell line. *Hepatobiliary Pancreat Dis Int*. 2009;8(1):75-8.

## *Références bibliographiques*

32. Kagan M, McCreery R. reduction of fluorescence interference in Raman spectroscopy via analyte adsorption on graphitic carbon. *AnalChem*. 1994;66:4159-65.
33. Yang H, Irudayaraj J. Rapid detection of foodborne microorganismes on food surface using fourier transform Raman spectroscopy. *Journal of Molecular Structure*. 2003; 646(1-3):35-43.
34. Fleischmann M, Hendra P, McQuillan A. Raman Spectra of Pyridine Adsorbed at a Silver Electrode. *Chemical Physics Letters* 1974; 26 (2):163-6.
35. Le Ru E, Blackie E, Meyer M, Etchegoin P. Surface Enhanced Raman Scattering Enhancement Factors: A Comprehensive Study. *J Phys Chem C*. 2007;111 (37)::13794-803.
36. Mahadevan-Jansen A, Richards-Kortum R. Raman spectroscopy for the detection of cancers and precancers. *J Biomed Opt* 1996;1(1):31-70.
37. Lloyd G, Orr L, Christie-Brown J, McCarthy K, Rose S, Thomas M, et al. Discrimination between benign, primary and secondary malignancies in lymph nodes from the head and neck utilising Raman spectroscopy and multivariate analysis. *Analyst*. 2013;138:3900-8.
38. Brozek-Pluska B, Musial J, Kordek R, Bailo E, Dieing T, Abramczyk H. Raman spectroscopy and imaging: applications in human breast cancer diagnosis. *Analyst*. 2012;137(16):3773-80.
39. Haka A, Shafer-Peltier K, Fitzmaurice M, Crowe J, Dasari R, Feld M. Diagnosing breast cancer by using Raman spectroscopy. *Proc Natl Acad Sci U S A*. 2005;102:12371-6.
40. Keller M, Vargis E, de Matos Granja N, Wilson R, Mycek M, Kelley M, et al. Development of a spatially offset Raman spectroscopy probe for breast tumor surgical margin evaluation. *J Biomed Opt* 2011;16(7):077006:16.
41. Widjaja E, Zheng W, Huang Z. Classification of colonic tissues using near-infrared Raman spectroscopy and support vector machines. *Int J Oncol*. 2008;32:653-62.
42. Calin M, Parasca S, Savastru R, Calin M, Dontu S. Optical techniques for the noninvasive diagnosis of skin cancer. *J Cancer Res Clin Oncol*. 2013;139(7):1083-104.
43. Krishna CM, Sockalingum GD, Bhat RA, Venteo L, Kushtagi P, Pluot M, et al. FTIR and Raman microspectroscopy of normal, benign, and malignant formalin-fixed ovarian tissues. *Anal Bioanal Chem*. 2007;387(5):1649-56.

## *Références bibliographiques*

---

44. Kendall C, Day J, Hutchings J, Smith B, Shepherd N, Barr H, et al. Evaluation of Raman probe for oesophageal cancer diagnostics. *Analyst*. 2010;135(12):3038-41.
45. Tan K, Herrington C, Brown C. Discrimination of normal from pre-malignant cervical tissue by Raman mapping of de-paraffinized histological tissue sections. *J Biophotonics*. 2011;4(1-2):40-8.
46. Elkins K. Rapid presumptive "fingerprinting" of body fluids and materials by ATR FT-IR spectroscopy. *J Forensic Sci*. 2011 56(6):1580-7.
47. Petibois C, Gionnet K, Gonçalves M, Perromat A, Moenner M, Déléris G. Analytical performances of FT-IR spectrometry and imaging for concentration measurements within biological fluids, cells, and tissues. *Analyst*. 2006;131(5):640-7.
48. Khaustova S, Davydov I, Trushkin E, Shkurnikov M, Mueller R, Backhaus J, et al. Application of mid-infrared molecular spectroscopy for assessment of biochemical parameters of blood serum. *Bull Exp Biol Med*. 2009;148(6):943-7.
49. Petibois C, Déléris G. Fourier-transform infrared spectrometry determination of the metabolic changes during a maximal 400-meter swimming test. *Int J Sports Med*. 2003;24(5):313-9.
50. Petibois C, Déléris G. Alterations of lipid profile in endurance over-trained subjects. *Arch Med Res*. 2004;35(6):532-9.
51. 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*. 2002;74(15):3865-8.
52. Staib A, Dolenko B, Fink D, Früh J, Nikulin A, Otto M, et al. Disease pattern recognition testing for rheumatoid arthritis using infrared spectra of human serum. *Clin Chim Acta* 2001;308(1-2):79-89.
53. Petrov M, Gordetzov A, Kukosh M. Early prediction of severity in acute pancreatitis using infrared spectroscopy of serum. *Pancreatology*. 2007;7(5-6):451-8.
54. Backhaus J, Mueller R, Formanski N, Szlama N, Meerpohl H, Eindt M, et al. diagnosis of breast cancer with infrared spectroscopy from serum samples. *Vibrational Spectroscopy* 2010;52:173-7.

## *Références bibliographiques*

---

55. Hands J, Abel P, Ashton K, Dawson T, Davis C, Lea R, 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):7347-55.
56. Hands J, Abel P, Ashton K, Dawson T, Davis C, Lea RW, et al. Investigating the rapid diagnosis of gliomas from serum samples using infrared spectroscopy and cytokine and angiogenesis factors. *Anal Bioanal Chem*. 2013.
57. Scaglia E, Sockalingum G, Schmitt J, Gobinet C, Schneider N, Manfait M, et al. Noninvasive assessment of hepatic fibrosis in patients with chronic hepatitis C using serum Fourier transform infrared spectroscopy. *Anal Bioanal Chem*. 2011;401(9):2919-25.
58. Filik J, Stone N. Analysis of human tear fluid by Raman spectroscopy. *Anal Chim Acta*. 2008;616(2):177-84.
59. Lamcharfi E, Cohen-Solal C, Parquet M, Lutton C, Dupre J, Meyer C. Determination of molecular associations of some hydrophobic and hydrophilic bile acids by infrared and Raman spectroscopy. *Eur Biophys J*. 1997;25(4):285-91.
60. Rohleder D, Kiefer W, Petrich W. Quantitative analysis of serum and serum ultrafiltrate by means of Raman spectroscopy. *Analyst*. 2004;129(10):906-11.
61. Pichardo-Molina JL, Frausto-Reyes C, Barbosa-Garcia O, Huerta-Franco R, Gonzalez-Trujillo JL, Ramirez-Alvarado CA, et al. Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients. *Lasers Med Sci*. 2007;22(4):229-36.
62. Harris AT, Lungari A, Needham CJ, Smith SL, Lones MA, Fisher SE, et al. Potential for Raman spectroscopy to provide cancer screening using a peripheral blood sample. *Head Neck Oncol*. 2009;1:34.
63. Savitzky A, Golay M. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*. 1964;36:1627-39.
64. Kohler A, Bertrand D, Martens H, Hannesson K, Kirschner C, Ofstad R. Multivariate image analysis of a set of FTIR microspectroscopy images of aged bovine muscle tissue combining image and design information. *Anal Bioanal Chem*. 2007;389(4):1143-53.
65. Massart D, Buydens L. Chemometrics in pharmaceutical analysis. *J Pharm Biomed Anal*. 1988;6(6-8):535-45.

## *Références bibliographiques*

---

66. Pearson K. On lines and planes of closestfit to systems of points in space. *Philosophy Magazine*. 1901;2:559-72.
67. Hotelling H. Analysis of complex of statistical variables into principal components. *J Educ Psychol*. 1933;24(6) 417-41.
68. Vapnik V. *The nature of statistical learning theory*. N-Y, Springer-Verlag. 1995.
69. Hamel L. *knowledge discovery with support vector machines*. Wiley edition. 2009.
70. Wold H. systems under indirect observation,. 1982;2:589-91.
71. Pérez-Enciso M, Tenenhaus M. Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Hum Genet*. 2003;112(5-6):581-92.
72. Chang C, Lin C. *ACM transactions on Intelligent Systems and Technology*. 2011;2:1-27.
73. Helm D, Labischinski H, Schallehn G, Naumann D. Classification and identification of bacteria by Fourier-transform infrared spectroscopy. *J Gen Microbiol*. 1991 137(1):69-79.
74. Deegan RD, Bakajin O, Dupont T, Huber G, Nagel S, Witten T. Capillary flow as the cause of ring stains from dried liquid drops. *Nature*. 1997.
75. Filik J, Stone N. Drop coating deposition Raman spectroscopy of protein mixtures. *Analyst*. 2007;132(6):544-50.
76. Ortiz C, Zhang D, Xie Y, Ribbe AE, Ben-Amotz D. Validation of the drop coating deposition Raman method for protein analysis. *Anal Biochem*. 2006;353(2):157-66.
77. Pearce EI, A.Tomlinson. *Ophthalmic Physiol Opt*. 2000;20(4):306-13.
78. Tang XC, Pikal MJ. The effect of stabilizers and denaturants on the cold denaturation temperatures of proteins and implications for freeze-drying. *Pharm Res*. 2005;22(7):1167-75.
79. Hedoux A, Paccou L, Achir S, Guinet Y. In situ monitoring of proteins during lyophilization using micro-Raman spectroscopy: a description of structural changes induced by dehydration. *J Pharm Sci*. 2012;101(7):2316-26.
80. Tang XC, Pikal MJ. Measurement of the kinetics of protein unfolding in viscous systems and implications for protein stability in freeze-drying. *Pharm Res*. 2005;22(7):1176-85.

## *Références bibliographiques*

---

81. Chang L, Shepherd D, Sun J, Ouellette D, Grant KL, Tang XC, et al. Mechanism of protein stabilization by sugars during freeze-drying and storage: native structure preservation, specific interaction, and/or immobilization in a glassy matrix? *J Pharm Sci.* 2005;94(7):1427-44.
82. Taleb I, Thiéfin G, Gobinet C, Untereiner V, Bernard-Chabert B, Heurgué A, et al. Diagnosis of hepatocellular carcinoma in cirrhotic patients: a proof-of-concept study using serum micro-Raman spectroscopy. *Analyst.* 2013 138(14):4006-14.
83. Ly E, Piot O, Wolthuis R, Durlach A, Bernard P, Manfait M. Combination of FTIR spectral imaging and chemometrics for tumour detection from paraffin-embedded biopsies. *Analyst.* 2008;133(2):197-205.
84. Murali Krishna C, Sockalingum G, Vidyasagar M, Manfait M, Fernanades D, Vadhiraja B, et al. An overview on applications of optical spectroscopy in cervical cancers. *J Cancer Res Ther.* 2008;4(1):26-36.
85. Babrah J, McCarthy K, Lush R, Rye A, Bessant C, Stone N. Fourier transform infrared spectroscopic studies of T-cell lymphoma, B-cell lymphoid and myeloid leukaemia cell lines. *Analyst.* 2009;134(4):763-8.
86. Pijanka J, Sockalingum G, Kohler A, Yang Y, Draux F, Parkes G, et al. Synchrotron-based FTIR spectra of stained single cells. Towards a clinical application in pathology. *Lab Invest.* 2010;90(5):797-807.
87. Forner A, Bruix J. Biomarkers for early diagnosis of hepatocellular carcinoma. *Lancet Oncol.* 2012;13(8):750-1.
88. Forner A, Llovet J, Bruix J. Hepatocellular carcinoma. *Lancet.* 2012 379(9822):1245-55.
89. Singal A, Volk M, Waljee A, Salgia R, Higgins P, Rogers M, et al. Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis. *Aliment Pharmacol Ther.* 2009 30(1):37-47.
90. El-Serag H. Hepatocellular carcinoma. *NEJM.* 2011;365:1118-27.
91. Zhang X, Thiéfin G, Gobine C, Untereiner V, Taleb I, Bernard-Chabert B, et al. Profiling serologic biomarkers in cirrhotic patients via high-throughput Fourier transform infrared spectroscopy: toward a new diagnostic tool of hepatocellular carcinoma. *Transl Res.* 2013;5244(13):00215-6.

## *Références bibliographiques*

---

92. Naveau S, Perlemuter G, Balian A. Epidemiology and natural history of cirrhosis : Cirrhosis. La revue du praticien. 2005;55:1527-32.
93. Forner A, Llovet J, Bruix J. Hepatocellular carcinoma. Lancet. 2012;379:1245-55.
94. Savitzky A, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical chemistry. 1964;36(8):1627-39.
95. Hughes G. On the mean accuracy of statistical pattern recognizers. IEEE Transactions. 1968;14:55-63.

# ANNEXES

# ARTICLE 1

ETUDE PILOTE RAMAN

## Diagnosis of hepatocellular carcinoma in cirrhotic patients: a proof-of-concept study using serum micro-Raman spectroscopy

Cite this: DOI: 10.1039/c3an00245d

I. Taleb,<sup>a</sup> G. Thiéfin,<sup>ab</sup> C. Gobinet,<sup>a</sup> V. Untereiner,<sup>a</sup> B. Bernard-Chabert,<sup>b</sup> A. Heurgué,<sup>b</sup> C. Truntzer,<sup>c</sup> P. Hillon,<sup>c</sup> M. Manfait,<sup>a</sup> P. Ducoroy<sup>c</sup> and G. D. Sockalingum<sup>\*a</sup>

Hepatocellular carcinoma (HCC) is the third most common cause of cancer death worldwide. The development of novel diagnostic methods is needed to detect tumours at an early stage when patients are eligible for curative treatments. The purpose of this proof-of-concept study was to determine if micro-Raman spectroscopy applied to the sera of cirrhotic patients may be an alternative method for rapidly discriminating patients with and without HCC. Serum samples were collected from 2 groups of patients: cirrhotic patients with HCC ( $n = 37$ ) and without HCC ( $n = 34$ ). Two different approaches were used, dried serum drops and freeze-dried serum, and micro-Raman spectra were acquired in the point-mode with a 785 nm laser excitation in the spectral range of 600–1800  $\text{cm}^{-1}$ . Spectra were quality-tested and pre-processed (smoothing, baseline subtraction, vector normalization). Using principal component analysis, the 2 classes, corresponding to cirrhotic patients with and without HCC, could not be differentiated. In contrast, the support vector machine method using the leave-one-out cross validation procedure was able to correctly classify the two groups of patients with an overall rate of accuracy of 84.5% to 90.2% for dried serum drops and 86% to 91.5% for freeze-dried serum. These results are promising and support the concept that serum micro-Raman spectroscopy may become a useful diagnostic tool to detect biomarkers in the field of cancer, as described here for distinguishing between cirrhotic patients with and without HCC.

Received 31st January 2013

Accepted 21st March 2013

DOI: 10.1039/c3an00245d

[www.rsc.org/analyst](http://www.rsc.org/analyst)

### Introduction

Primary liver cancer is one of the most common cancers worldwide, accounting for about 7% of all cancers, and the third leading cause of cancer death with an estimated 694 000 deaths annually.<sup>1</sup> Hepatocellular carcinoma (HCC) represents more than 90% of primary liver cancers. In spite of therapeutic advancements, the prognosis of patients with HCC remains very poor with a median survival of less than one year, mainly because the majority of patients are diagnosed at an advanced stage of the tumor.<sup>2</sup> The main risk factors for HCC include chronic viral hepatitis B and C, alcoholic liver disease and nonalcoholic fatty liver disease. Most cases of HCC develop in patients with an underlying liver disease at the stage of cirrhosis

irrespective of the cause. Therefore, clinical practice guidelines recommend surveillance programs in cirrhotic patients to detect the development of HCC at an early stage, when patients are eligible for potentially curative treatments.<sup>3,4</sup>

According to these guidelines, surveillance of cirrhotic patients is based on hepatic ultrasonography every 6 months.<sup>3,4</sup> However, ultrasound examination is operator-dependent and the diagnostic performance for detecting HCC at an early stage is limited by the presence of fibrous septa and regenerative nodules in cirrhotic liver. In addition, the patient's morphology, particularly overweight, may not allow an adequate liver ultrasonographic assessment. For these reasons, there is a need for biomarkers of HCC. A number of serological biomarkers have been tested. Alpha-fetoprotein (AFP) is the most widely used in clinical practice. Although serum AFP measurement is of interest for the diagnosis of HCC in cirrhotic patients with a mass in the liver, its value as a screening test is much more limited. Only a minority of HCC at an early stage displays increased serum concentrations of AFP. On the other hand, increased concentrations of AFP may be observed in cirrhotic patients without HCC, particularly those infected with hepatitis virus B or C.<sup>3,4</sup> A number of other serum markers have been tested such as the fucosylated isoform of AFP, des-gamma-carboxyprothrombin,

<sup>a</sup>Equipe MEdIAN, Biophotonique et Technologies pour la Santé, Université de Reims Champagne-Ardenne, CNRS FRE 3481-MEDyC, UFR de Pharmacie, 51 rue Cognacq-Jay, 51096 Reims Cedex, France. E-mail: ganesh.sockalingum@univ-reims.fr; Fax: +33 (0)3 26 91 35 50; Tel: +33 (0)3 26 91 35 53

<sup>b</sup>Service d'Hépatogastroentérologie, Centre Hospitalo-Universitaire de Reims, Hôpital R Debré, 51092 Reims Cedex, France

<sup>c</sup>Service d'Hépatogastroentérologie and Clinical Innovation Proteomic Platform, Centre Hospitalo-Universitaire de Dijon, Hôpital du Bocage, 21079 Dijon Cedex, France

glypican-3, and alpha-1-fucosidase. None of these molecules is currently considered suitable for HCC screening and their use is not recommended in recently published guidelines from scientific and professional medical organizations.<sup>3,4</sup>

During the last decade, the development of vibrational spectroscopic techniques such as infrared or Raman spectroscopy combined with advanced chemometrics has opened up new perspectives for identification of novel cancer biomarkers. Raman spectroscopy is a molecular analytical tool which provides biochemical and structural information on the analyzed sample. It is non-destructive, non-contact, and reagent-free.<sup>5</sup> In the field of cancer research, the potential of Raman micro-spectroscopy has been explored mainly for tissue and cell molecular analyses. It has been shown to provide a biochemical tissue fingerprint that allows distinguishing normal and tumoral tissue. A number of studies have demonstrated the interest of micro-Raman spectroscopy as a diagnostic tool in a variety of cancers including skin,<sup>6,7</sup> ovarian,<sup>8</sup> breast,<sup>9,10</sup> esophagus,<sup>11</sup> cervix,<sup>12,13</sup> and colorectal<sup>14,15</sup> cancers.

In addition to studies on tissues and cells, Raman spectroscopy has been applied to biofluids. It has been used to analyze and measure metabolites in several biological fluids such as tears,<sup>16</sup> bile<sup>17,18</sup> and serum.<sup>17</sup> However, the leading interest in the field of biomedical research is the identification of diagnostic and prognostic cancer biomarkers. Serum appears to be of particular interest for this purpose as it is easily accessible and is expected to contain markers shed from tumor cells into the circulation.<sup>19</sup> In the literature, few studies have evaluated the potential of conventional micro-Raman spectroscopy applied to the serum for cancer diagnosis. Pichardo-Molina *et al.*<sup>20</sup> have demonstrated that serum Raman spectroscopy combined with principal component analysis (PCA) and linear discriminant analysis allowed the discrimination between samples from breast cancer patients and healthy volunteer controls with a sensitivity of 97% and a specificity of 78%. Similarly, Harris *et al.*<sup>21</sup> have shown the potential of this technique for neck and head cancer screening with sensitivity and specificity rates of 75%. More recently, in order to overcome the very low intensity of Raman scattering and the strong autofluorescence interference, blood serum surface enhanced Raman spectroscopy (SERS) has been developed, taking interest of advancements in nanotechnology.<sup>22</sup> Potential applications of this technique to the serum and plasma have been reported for detection of gastric,<sup>23</sup> colorectal,<sup>24</sup> and naso-pharyngeal<sup>25</sup> cancers.

To our knowledge, serum micro-Raman spectroscopy has never been evaluated for the diagnosis of HCC. Our working hypothesis was that serum micro-Raman spectra from cirrhotic patients with HCC show spectral characteristics (qualitatively and quantitatively) which can be used to predict the presence of cancer.

Given all these data, the aim of this pilot study was to investigate the potential of micro-Raman spectroscopy coupled with advanced chemometric methods to highlight the differences between sera from cirrhotic patients with and without HCC. The idea was to use the molecular signatures from different macromolecules (whole spectral fingerprint) rather than specific markers like those used for bio-assays.

## Materials and methods

### Study design

This proof-of-concept Raman study is part of the multicentric ProSpec study aimed at identifying serum biomarkers of HCC in patients with cirrhosis using two approaches: proteomics and vibrational spectroscopy. The ProSpec study is ancillary to the CirCE study, an on-going case-control study in the Great Eastern French Canceropole, the purpose of which is to analyze environmental, nutritional, and metabolic factors involved in liver carcinogenesis among cirrhotic patients. The study was approved by the Comité de Protection des Personnes Est (agreement no. 2008/09) and by the Agence Française de Sécurité Sanitaire des Produits de Santé (AFSSAPS; agreement no. 2008-A00023-52). A written informed consent was obtained from all patients.

Blood samples were collected from two groups of patients: cirrhotic patients with HCC ( $n = 37$ ) and cirrhotic patients without HCC ( $n = 34$ ). After centrifugation, the serum was isolated, frozen, and stored at  $-80\text{ }^{\circ}\text{C}$ . The diagnosis of cirrhosis relied either on histology obtained by liver biopsy or on combination of typical clinico-biological signs, endoscopic features, and imaging characteristics. The absence of HCC was ascertained by imaging techniques (ultrasonography and/or computed tomography and/or magnetic resonance imaging), and a serum alpha-fetoprotein concentration below  $100\text{ ng ml}^{-1}$ . Cirrhotic patients without HCC were screened for HCC every six months. None of them developed HCC during a one-year follow-up. The diagnosis of HCC was made according to the criteria defined by the European Association for Study of the Liver.<sup>26</sup> Exclusion criteria were age below 35, HIV infection, extrahepatic cancers and psychiatric diseases that did not allow obtaining an informed consent.

### Sample preparation (Fig. 1)

Two serum sample preparations were used in this study. In a first set of experiments, frozen serum samples were thawed and  $10\text{ }\mu\text{l}$  of each sample were deposited on a  $\text{CaF}_2$  window and allowed to dry in air at room temperature for 30 min. Ten spectra were recorded, 5 at the centre and 5 at the periphery of the dried drop. In a second set of experiments, an aliquot of  $200\text{ }\mu\text{l}$  of each serum sample was freeze-dried for 24 h under vacuum in a freeze dryer equipped with a RV8 Rotary Vane Pump (Edwards, SuperModulyo 1.5 liter, Thermo Fisher Scientific). The lyophilized products were stored at  $-80\text{ }^{\circ}\text{C}$ . For Raman analysis, the lyophilized sera were thawed at room temperature for 10 minutes, and then pressed on a  $\text{CaF}_2$  window to obtain a flat surface of uniform thickness. Five point spectra were randomly recorded on the sample surface.

### Micro-Raman spectral acquisition (Fig. 1 – step 1)

Raman spectra were recorded with a LabRam micro-spectrometer (Horiba Jobin Yvon SAS, Villeneuve d'Ascq, France) equipped with an Olympus BX40 microscope and a  $100\times$  long working distance objective (Olympus, NA 0.9, Rungis, France). The excitation was a near-infrared diode laser

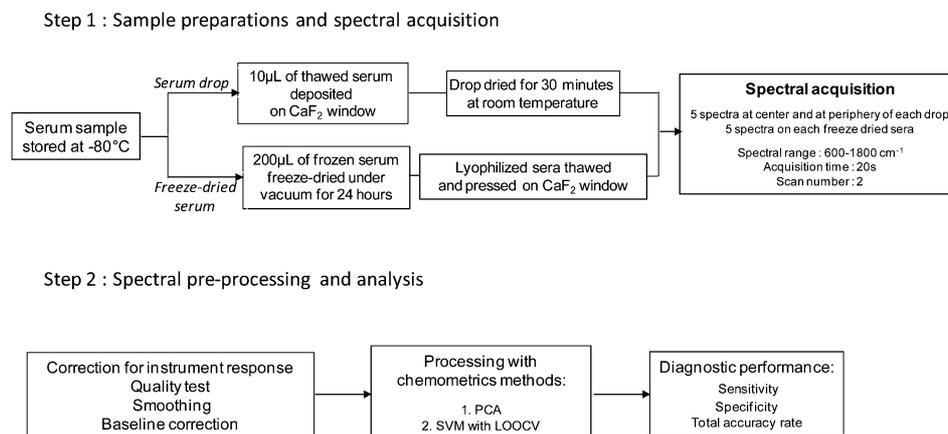


Fig. 1 Stepwise schematic representation of the procedure used from sample preparation to spectral diagnosis.

(Toptica Photonics, Munich, Germany) generating a 300 mW single mode line at 785 nm. The laser power on the sample was 60 mW and the spot size was 2  $\mu\text{m}$ . A CCD camera (1024  $\times$  256 pixels, Andor Tech., Belfast, UK) was used to detect the Raman signal. All spectra were recorded in the point-mode. The acquisition time was 2  $\times$  20 s and the wavenumber range was from 600 to 1800  $\text{cm}^{-1}$  with a 6  $\text{cm}^{-1}$  spectral resolution for both sample preparations. The instrument was calibrated using a silicon sample (Raman peak at 520  $\text{cm}^{-1}$ ) and two reference compounds (1,4-bis(2-methylstyryl) benzene and naphthalene). Spectra were calibrated in relative intensity with a NIST calibration standard for 785 nm (standard reference material #2241; NIST, Gaithersburg, MD, USA). Spectral acquisitions were recorded using the Labspec 4 software (Horiba Scientific Jobin Yvon; Lille; France).

### Raman data preprocessing (Fig. 1 – step 2)

All spectral preprocessing were performed using MatLab software (version 7.6.0, MathWorks, Natick, Mass., USA). Fig. 2 depicts an example of the pre-processing steps used for a micro-Raman spectrum taken from the centre of a dried serum drop. In the first step, the Raman spectra were corrected for the dark current, optics and a NIST standard for 785 nm in order to correct for instrument response (Fig. 2A). Then, all sample spectra were subjected to a quality test in order to verify the signal-to-noise ratio. To calculate the  $S/N$  ratio, the signal intensity was taken at the phenylalanine peak near 1003  $\text{cm}^{-1}$  and the noise intensity was measured in the 1720–1800  $\text{cm}^{-1}$  region, which does not contain any signal. In this study, all spectra with a signal-to-noise ratio lower than 50 were eliminated.<sup>27</sup>

Spectra that passed the quality test were smoothed using the Savitzky–Golay algorithm<sup>28</sup> with a window width of 11 datapoints and a third-order polynomial to increase the spectral contrast (Fig. 2B). Then, the background of each spectrum was corrected by using a fourth-order polynomial (Fig. 2C).<sup>29</sup> Finally, standard normal variate normalisation was performed in the full spectral range (Fig. 2D). The first step is to calculate the mean spectrum of the dataset, which is subtracted from each individual spectrum. Then normalization is achieved by

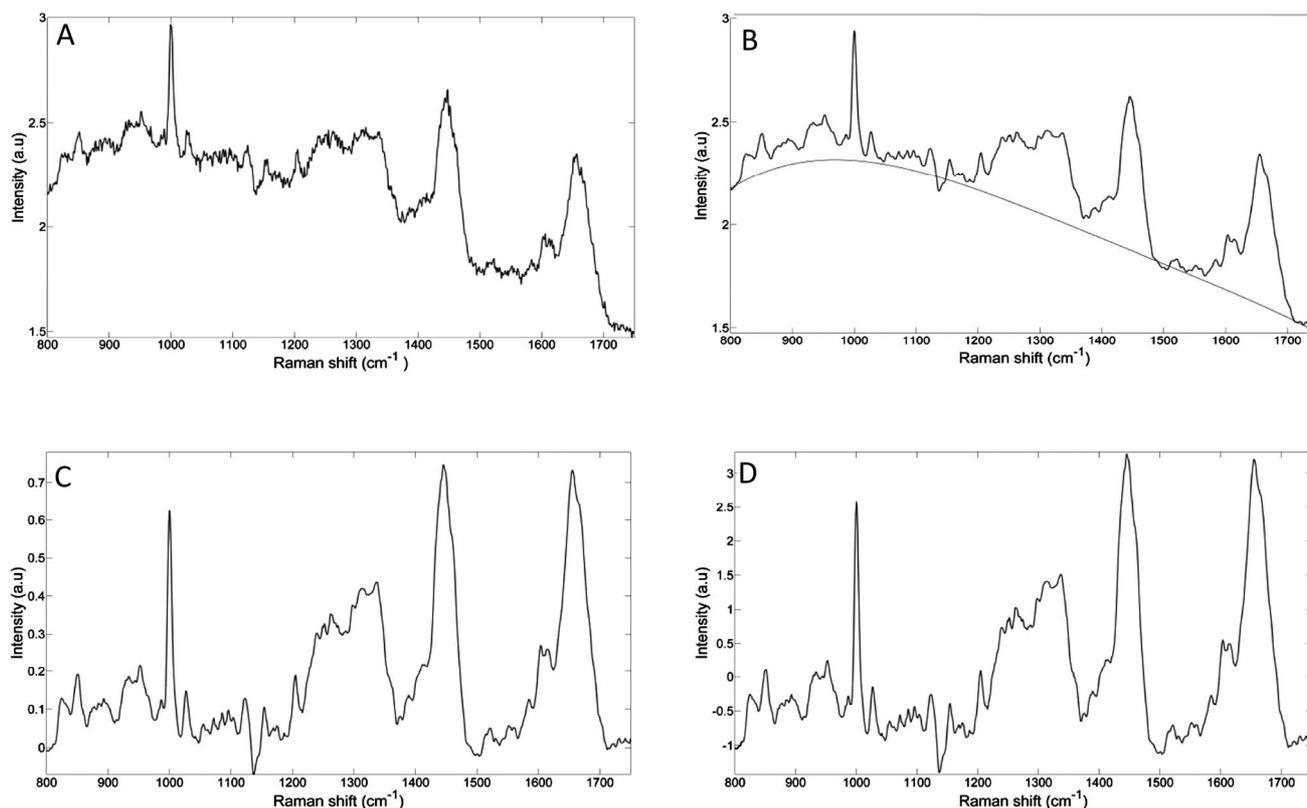
dividing the mean-centered spectra by the standard deviation over the spectral intensities giving the resulting spectra a unit standard deviation of one. The pre-processing was completed with a zero offset, the value of which is chosen such that at least one point of the corrected spectrum equals zero.

### Statistical analysis

Demographic and clinico-biological data were collected from patients' medical files. Quantitative data were expressed as mean  $\pm$  standard deviation and were compared using the Student's *t*-test. Qualitative data were expressed as a percentage and were compared using the chi-square test. A *p* value less than 0.05 was considered significant.

Spectral analysis was performed using the Matlab software. Firstly, principal component analysis (PCA), an unsupervised classification method, was applied to both datasets of dried drops and freeze-dried sera. PCA is a multivariate data analytical method used to reduce the dimensionality of the data and create an alternative set of coordinates, the principal components (PCs). Given the large number of variables in our Raman spectra (1013 wavelengths per spectrum) and the relatively low number of patients ( $n = 71$ ), PCA was appropriate to reduce the number of predictor variables. PCs are linear combinations of initial variables with the first PC explaining the highest variance of the data. The following PCs were determined in such a way that they were uncorrelated and each one successively had the highest variance possible. Then, the most discriminant PCs for differentiating spectra from both groups of cirrhotic patients with and without HCC were identified using the Mann–Whitney *U*-test. Finally, PCA clustering was performed using either the most significant PCs (those explaining the highest variances) or the most discriminant ones. For this purpose, PC scores were plotted to reveal potential relations between groups of spectra.

Secondly, the support vector machine (SVM), a supervised classification method, was applied to the most discriminant vectors selected by the PCA using the LIBSVM library.<sup>30</sup> To overcome the risk of overfitting due to a small number of patients ( $n = 71$ ), we used a leave-one-out cross-validation procedure (LOOCV). In this procedure, all replicate spectra from each patient were successively used as the validation data



**Fig. 2** Example of the pre-processing steps used for a micro-Raman spectrum taken from the centre of a dried serum drop. (A) Raman spectrum corrected for instrument response, (B) smoothed spectrum and representation of the 4<sup>th</sup> order polynomial used for baseline correction, (C) spectrum after baseline subtraction and normalised using the standard normal variate method, and (D) offset correction of the normalised spectrum.

whereas spectra from all remaining patients were used as the training set. The process was repeated so that spectra from each patient were used once as the validation set. The sensitivity, specificity and overall accuracy rate of the classifier were calculated using the whole set of validated spectra, the mean, and the median spectra.

## Results and discussion

### Patient characteristics

Ninety percent of the patients included in the study had cirrhosis related to excessive alcohol consumption. Other causes were hepatitis virus B or C infection and a metabolic syndrome. There was no significant difference in the distribution of these etiological factors between cirrhotic patients with and without HCC. The mean age was higher in the group of patients with HCC ( $66 \pm 9$ ) than in the group without HCC ( $61 \pm 9$ ;  $p = 0.015$ ). There was a predominance of males in both groups. Severity of cirrhosis as assessed by the Child Pugh score was more pronounced in patients without HCC although not significant. Finally, as expected, the serum level of alpha-fetoprotein was higher in patients with HCC than in those without HCC (Table 1).

### Analysis of micro-Raman spectra from dried serum drops

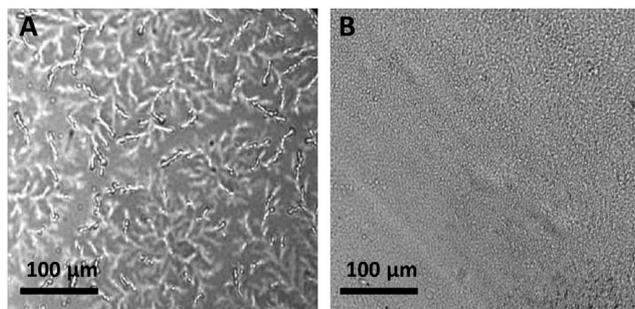
In this study, two kinds of sample preparations were used to test the Raman technique for HCC diagnosis. Firstly, micro-Raman

spectroscopy was performed on dried serum drops and secondly on freeze-dried sera. Since the drying process of a drop may induce an outward flow of analytes from the centre to the periphery of the drop,<sup>31</sup> we recorded micro-Raman spectra from both sites and analyzed them separately. In the sets of micro-Raman spectra taken at the centre and the periphery of dried serum drops, 353 out of 355 and 353 out of 355 respectively met the quality test. Fig. 3 shows a white light image of the centre of

**Table 1** Demographic and clinico-biological characteristics of cirrhotic patients with and without HCC

	Cirrhotic patients with HCC ( $n = 37$ )	Cirrhotic patients without HCC ( $n = 34$ )	$p$ Value <sup>a</sup>
Age	$66 \pm 9$	$61 \pm 9$	0.015
Sex ratio (males/females)	32/5	28/6	0.75
Child-Pugh score	$6.3 \pm 1.7$	$7.3 \pm 2.6$	0.07
Cause of cirrhosis <sup>b</sup>			0.23
Alcoholic	32/37	32/34	
C virus	3/37	3/34	
B virus	4/37	0/34	
Metabolic	3/37	3/34	
Alpha-fetoprotein >500 ng ml <sup>-1</sup>	6/37(16)	0/34	0.03

<sup>a</sup> Student's  $t$ -test for quantitative data (mean  $\pm$  SE) and chi-square test for qualitative data. <sup>b</sup> Several causes may be combined in some patients.



**Fig. 3** White light images of the dried serum drop with the centre displaying a fern-like pattern (A) and its periphery appearing as an amorphous ring (B).

the dried drop displayed a fern-like pattern (Fig. 3A), whereas its periphery appeared as an amorphous ring (Fig. 3B). This observation is in accordance with previous studies performed on various protein solutions<sup>32,33</sup> and tears.<sup>16,32</sup> Drop coating deposition Raman spectroscopy of tear fluid, an aqueous solution containing proteins, lipids, electrolytes and various metabolites, showed that proteins were deposited mainly and rather homogeneously in the ring.<sup>16</sup> The ferns in the centre of the drop contained a smaller amount of proteins, electrolytes and urea.<sup>16,34</sup> In our study, spectra from the centre and the periphery of the serum dried drop looked very similar upon visual inspection (Fig. 4). When the periphery mean spectrum was subtracted from the centre mean spectrum, the resulting difference spectrum did not exhibit any clear differences. This was observed for both groups of cirrhotic patients, with HCC (Fig. 4A) and without HCC (Fig. 4B).

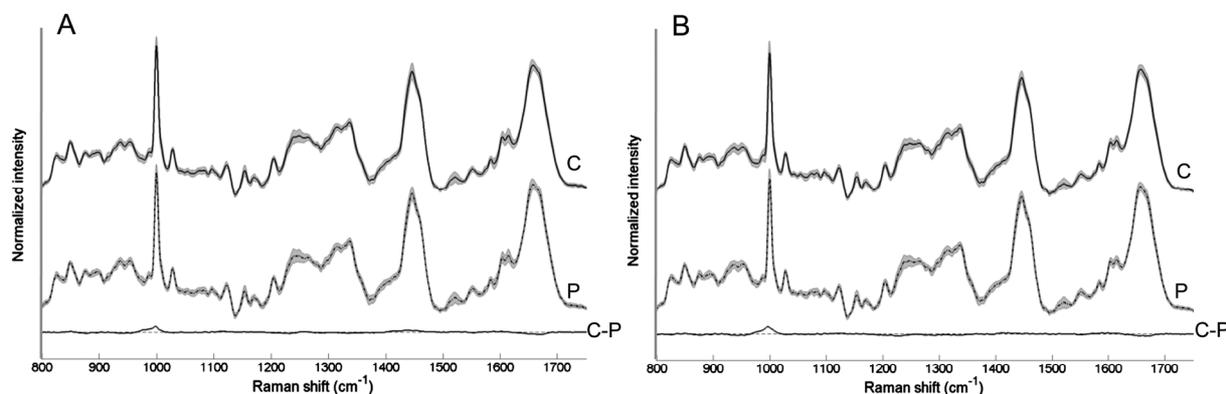
In Fig. 5, the mean micro-Raman serum spectra from patients with HCC and without HCC were compared. Visually, there was no difference between them either in the set taken at the centre of the dried drop (Fig. 5A) or at the periphery (Fig. 5B). The difference spectra support these observations showing no spectral variations between patients with and without HCC. Alternatively, to highlight potential discriminant spectral differences between the two groups, the commonly used chemometrics approach based on PCA was applied as a first clustering method.

**PCA results.** When PCA was applied to micro-Raman spectra taken at the centre of the dried serum drop, 287 PCs accounted for 99.99% of the total variance in the whole set of 353 individual spectra, 66 PCs in the set of 71 mean spectra and 68 PCs in the set of 71 median spectra. Regarding spectra taken at the periphery of the dried serum drop, 286 PCs accounted for 99.99% of the total variance in the whole set of 353 spectra, 66 PCs in the set of 71 mean spectra and 68 PCs in the set of 71 median spectra.

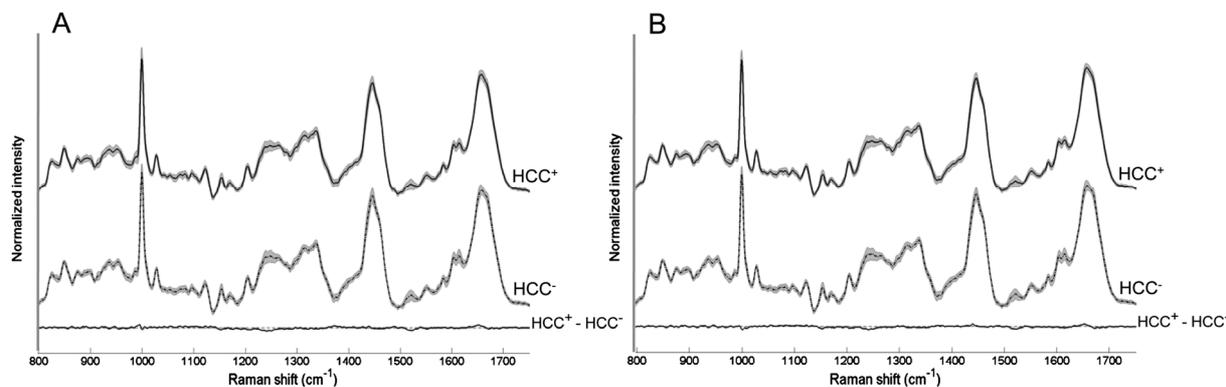
Fig. 6 shows the 3D scatter plots obtained using the three PCs with the highest explained variance in the sets of micro-Raman spectra from the centre (Fig. 6A) and the periphery of the dried serum drop (Fig. 6B). The first three PCs respectively accounted for 24.73%, 16.19% and 11.46% of total variance in the first set of spectra and for 25.58%, 18.05%, and 10.55% in the second set.

The PC score plots did not show any clustering tendency and did not allow separation of serum spectra from patients with and without HCC. The same analysis using the most discriminant PCs gave similar results (data not shown). This suggests a high similarity between the spectra from both groups of patients. Due to this limitation of PCA, the use of a more robust classifier was necessary to discriminate between patients with and without HCC.

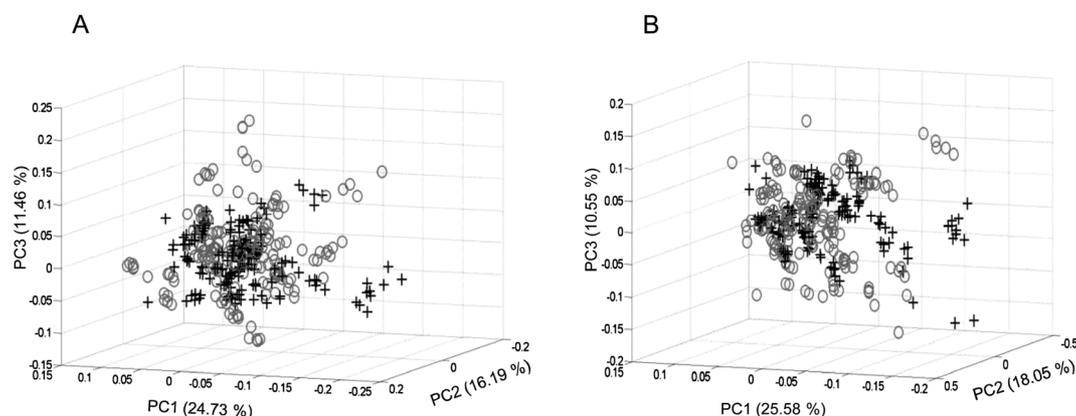
**LOOCV-SVM results.** LOOCV-SVM was applied to the Raman data using the most discriminant PCs previously determined by PCA. Table 2 summarizes the results of LOOCV-SVM method applied to the micro-Raman spectra taken at the centre and the periphery of dried serum drops from cirrhotic patients with and without HCC. For each analysis, median, mean, and all individual spectra were tested. As shown in Table 2, this classification procedure allowed discriminating patients with and without HCC with a sensitivity between 81 and 92% and a specificity between 82 and 92%. The use of the whole set of spectra rather than the mean or the median spectra did not improve the performance of the classifier. This may be explained by the good reproducibility of spectra from the same sample. On the other hand, similar results were achieved with spectra from the centre or the periphery of the drop (Table 2). This suggests that biomarkers of HCC did not segregate



**Fig. 4** Comparison of the mean micro-Raman spectra of the dried serum drop centre (c) and the periphery (p). The graphic on the left (A) refers to the group of patients without HCC and the graphic on the right (B) to the group of patients with HCC. Standard deviations (shaded area) are shown overlaying the mean micro-Raman spectra. All spectra are plotted using the same vertical scale and offset for clarity.



**Fig. 5** Comparison of the mean micro-Raman spectra from patients with HCC (HCC<sup>+</sup>) and without HCC (HCC<sup>-</sup>). The graphic on the left (A) refers to micro-Raman spectra obtained in the centre of the dried serum drop and the graphic on the right (B) to micro-Raman spectra obtained at the periphery of the dried serum drop. Standard deviations (shaded area) are shown overlaying the mean micro-Raman spectra. All spectra are plotted using the same vertical scale and offset for clarity.



**Fig. 6** PCA score plots of the whole set of micro-Raman spectra of serum dried drop from cirrhotic patients with (black cross) and without (gray circle) HCC. PCA was performed on the set of spectra taken in the centre (A) and in the periphery (B) of the dried drop.

significantly during drop deposition. Filik and Stone<sup>16</sup> have analyzed the spatial distribution of components in the tear drying pattern. Although the tear fluid composition is different from serum, it represents also a complex aqueous solution containing several hundreds of distinct proteins, lipids, metabolites and electrolytes. The concentration of proteins in tear fluid is about seven times lower than that in serum. For these reasons, Raman point mapping data obtained from tear drying patterns cannot be extrapolated to dried serum analysis without caution. However, Filik and Stone<sup>16</sup> have shown using PCA that protein signals were predominant in the ring but were also significantly present in the fern pattern at the centre of the drop. Our results lead us to assume that the markers of interest were present at the centre as well as at the periphery of the dried drop in sufficient concentrations to give spectral signals.

#### Analysis of micro-Raman spectra from freeze-dried serum

Freeze-drying (lyophilization) is a dehydration technique used to preserve and improve the stability of proteins.<sup>35</sup> However, it is well established that this process may lead to conformational changes of the secondary structures of proteins.<sup>36,37</sup> Chang

*et al.*<sup>38</sup> had shown by FTIR analysis a decrease in the  $\alpha$ -helix content after lyophilization of human serum albumin. In our study, we also tested the interest of Raman spectroscopy applied to the freeze-dried serum from patients with and without HCC.

All the 355 spectra from 71 serum samples met the quality test. The five replicates recorded from different areas arbitrarily chosen in each sample looked similar and the low standard deviation reflects the homogeneity of freeze-dried serum spectra in both groups of patients. The mean Raman spectra of freeze-dried serum from cirrhotic patients with and without HCC are shown in Fig. 7A. As is observed, the difference spectrum shows no important variations between the two groups of patients with and without HCC. As previously described, both PCA and LOOCV-SVM were applied to the dataset obtained with freeze-dried serum.

**PCA results.** When PCA was applied to micro-Raman spectra of freeze-dried serum drops, 267 PCs accounted for 99.99% of the total variance in the whole set of 355 spectra, 65 PCs in the set of 71 mean spectra and 68 PCs in the set of 71 median spectra. Fig. 7B shows the PCA score plots of the freeze-dried serum using the three components with the highest explained variance accounting for 53.9%, 15.73% and 6.58% of the total

**Table 2** LOOCV-SVM classification of cirrhotic patients with and without HCC. The SVM classification model was applied separately to the group of micro-Raman spectra taken in the periphery and the centre of the dried drop

	Dried serum drop-centre			Dried serum drop-periphery		
	Median	Mean	All spectra	Median	Mean	All spectra
Number of PCs <sup>a</sup>	16	11	32	16	13	41
Patients with HCC	37	37	37	37	37	37
Well classified	33	33	33	34	33	30
Wrongly classified	4	4	4	3	4	7
Patients without HCC	34	34	34	34	34	34
Well classified	28	30	30	30	31	30
Wrongly classified	6	4	4	4	3	4
Sensitivity (%)	89.19	89.19	89.19	91.89	89.19	81.08
Specificity (%)	82.35	88.24	88.24	88.24	91.18	88.24
Total rate of well-classified patients (%)	85.91	88.74	88.74	90.14	90.14	84.51

<sup>a</sup> Number of principal components used to build the classification model (common to both groups).

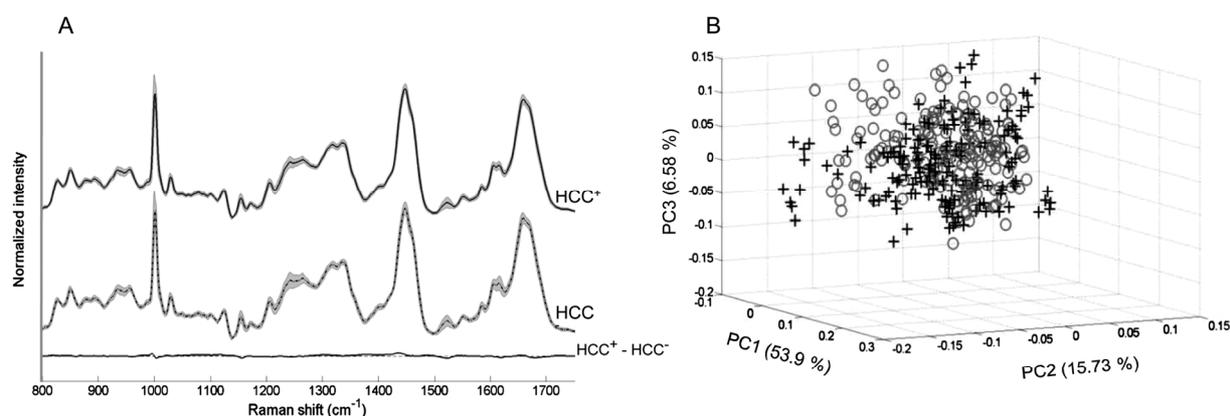
variance respectively. As previously observed, no clear discrimination was achieved and instead all the spectra overlapped. The same procedure was repeated using the most discriminant PCs. It did not allow discriminating between patients with and without HCC (data not shown). This suggests a high similarity between the spectra from both groups of patients. PCA seems to show here the same limitations as for the dried serum drop because the micro-Raman spectra of freeze-dried sera from patients with and without HCC have very similar spectral signatures and cannot be delineated by any such method. The use of a more robust classifier was therefore necessary to discriminate between the two groups of patients. The PCA outcome was then used in the next step to reduce the number of variables for the SVM-LOOCV analysis.

**LOOCV-SVM results.** As shown in Table 3, a good discrimination between micro-Raman spectra from patients with and without HCC was obtained using the LOOCV-SVM classification method. Depending on the dataset used (all spectra, mean or median spectra), the sensitivity was between 86% and 95% and

**Table 3** LOOCV-SVM classification of cirrhotic patients with and without HCC. The SVM classification model was applied to the set of micro-Raman spectra of freeze-dried serum

	Freeze-dried serum		
	Median	Mean	All spectra
Number of PCs <sup>a</sup>	13	9	31
Patients with HCC	37	37	37
Well classified	33	32	35
Wrongly classified	4	5	2
Patients without HCC	34	34	34
Well classified	28	30	30
Wrongly classified	6	4	4
Sensitivity (%)	89.19	86.49	94.59
Specificity (%)	82.35	88.24	88.24
Total rate of well-classified patients (%)	85.91	87.33	91.55

<sup>a</sup> Number of principal components used to build the classification model (common to both groups).



**Fig. 7** Mean micro-Raman analysis of freeze-dried serum. (A) Mean micro-Raman spectra of freeze-dried serum from patients with HCC (HCC<sup>+</sup>) and without HCC (HCC<sup>-</sup>) and the corresponding difference spectrum (HCC<sup>+</sup> minus HCC<sup>-</sup>). Standard deviations are shown overlaying the mean micro-Raman spectra (shaded area). All spectra are plotted using the same vertical scale and offset for clarity. (B) PCA score plot of the whole set of micro-Raman spectra of freeze-dried serum from cirrhotic patients with (black cross) and without (gray circle) HCC.

the specificity in the range 82% to 89%. The rates obtained with the whole set of spectra were slightly but not significantly higher than those obtained with mean or median spectra. As observed in the case of the dried serum drop, the use of the whole set of spectra did not improve the diagnostic performance compared with the analysis based on mean or median spectra. Again, this may be due to the good reproducibility of spectra from the same sample. The overall diagnostic accuracy rate was between 85% and 92%.

## Conclusions

Our study supports the working hypothesis that serum micro-Raman spectroscopy combined with SVM-LOOCV may be a suitable approach for rapid classification of cirrhotic patients with and without HCC. Two sample preparations, dried serum drops and the freeze-dried serum, were analyzed and had similar diagnostic performances, *i.e.*, an overall diagnostic accuracy of 85 to 90%. Raman analysis of dried serum drops is simple, rapid, and with minimal sample preparation. In spite of the heterogeneous visual aspect of the dried serum drops, the measurement site had no significant impact on the rate of well-classified patients. Regarding the freeze-dried samples, the process has the advantage of giving a homogeneous serum powder. However, this process requires more sample volume, takes more than 24 hours, and involves additional processing of the biological material. Finally, our study indicates that advanced chemometric methods are needed for an optimal spectral analysis. In our work, PCA alone failed to discriminate between the two groups of patients. The combination of PCA, to reduce data dimensionality, with SVM as the classification method, revealed to be more promising in the situation where spectra showed very subtle differences. Given the results obtained in this pilot study, it appears of interest to extend the database to build a training set, a test set, and a validation set to improve the robustness of the classifier.

## Acknowledgements

We are grateful to Mr Thomas Feron for his excellent technical support. Funding from the Institut National du Cancer, Fondation de France, Ligue contre le Cancer and Cancéropole Grand-Est is acknowledged. IT thanks the city of Reims for funding her PhD fellowship. The IBiSA technological platform "Imagerie Cellulaire et Tissulaire" is gratefully acknowledged. The authors acknowledge the involvement of the following members of the CiRCE ProSpec Study Group: J. P. Bronowicki (CHU Nancy), V. Di Martino (CHU Besançon), M. Doffoel (CHU Strasbourg), A. Jeannin (CHU Dijon), J. Liu (CHU Dijon), D. Pecqueur (CHU Dijon) and J. J. Raabe (CH Metz).

## References

- 1 J. Ferlay, H. Shin, F. Bray, D. Forman, C. Mathers and D. Parkin, *Int. J. Cancer*, 2010, **127**(12), 2893–2917.
- 2 H. Nordenstedt, D. L. White and H. B. El-Serag, *Dig. Liver Dis.*, 2010, **42**(suppl 3), S206–S214.
- 3 J. M. Llovet, M. Ducreux, R. Lencioni, A. M. Di Bisceglie, P. R. Galle, J. F. Dufour, T. F. Greten, E. Raymond, T. Roskams, T. De Baere, V. Mazzaferro, M. Bernadi, J. Bruix, M. Colombo and A. Zhu, *J. Hepatol.*, 2012, **56**(4), 908–943.
- 4 M. Sherman, K. Burak, J. Maroun, P. Metrakos, J. J. Knox, R. P. Myers, M. Guindi, G. Porter, J. R. Kachura, P. Rasuli, S. Gill, P. Ghali, P. Chaudhury, J. Siddiqui, D. Valenti, A. Weiss and R. Wong, *Curr. Oncol.*, 2011, **18**, 228–240.
- 5 C. Kendall, M. Isabelle, F. Bazant-Hegemark, J. Hutchings, L. Orr, J. Babrah, R. Baker and N. Stone, *Analyst*, 2009, **134**(6), 1029–1045.
- 6 E. Ly, A. Durlach, F. Antonicelli, P. Bernard, M. Manfait and O. Piot, *Exp. Dermatol.*, 2010, **19**, 68–73.
- 7 N. S. Eikje, K. Aizawa and Y. Ozaki, *Biotechnol. Annu. Rev.*, 2005, **11**, 191–225.
- 8 C. M. Krishna, G. D. Sockalingum, R. A. Bhat, L. Venteo, P. Kushtagi, M. Pluot and M. Manfait, *Anal. Bioanal. Chem.*, 2007, **387**, 1649–1656.
- 9 A. Saha, I. Barman, N. C. Dingari, S. McGee, Z. Volynskaya, L. H. Galindo, W. Liu, D. Plecha, N. Klein, R. R. Dasari and M. Fitzmaurice, *Biomed. Opt. Express*, 2011, **2**(10), 2792–2803.
- 10 J. S. Soares, I. Barman, N. C. Dingari, Z. Volynskaya, W. Liu, N. Klein, D. Plecha, R. R. Dasari and M. Fitzmaurice, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **110**(2), 471–476.
- 11 L. M. Almond, J. Hutchings, N. Shepherd, H. Barr, N. Stone and C. Kendall, *J. Biophotonics*, 2011, **4**, 685–695.
- 12 N. Amharref, A. Beljebbar, S. Dukic, L. Venteo, L. Schneider, M. Pluot and M. Manfait, *Biochim. Biophys. Acta*, 2007, **1768**, 2605–2615.
- 13 E. M. Kanter, E. Vargis, S. Majumder, M. D. Keller, E. Woeste, G. G. Rao and A. Mahadevan-Jansen, *J. Biophotonics*, 2009, **2**, 81–90.
- 14 A. Beljebbar, O. Bouche, M. D. Diebold, P. J. Guillou, J. P. Palot, D. Eudes and M. Manfait, *Crit. Rev. Oncol. Hematol.*, 2009, **72**, 255–264.
- 15 X. Li, T. Yang and S. Li, *Appl. Opt.*, 2012, **51**, 5038–5043.
- 16 J. Filik and N. Stone, *Anal. Chim. Acta*, 2008, **616**, 177–184.
- 17 E. Lamcharfi, C. Cohen-Solal, M. Parquet, C. Lutton, J. Dupre and C. Meyer, *Eur. Biophys. J.*, 1997, **25**, 285–291.
- 18 D. Rohleder, W. Kiefer and W. Petrich, *Analyst*, 2004, **129**, 906–911.
- 19 E. F. Petricoin, C. Belluco, P. R. Araujo and L. A. Liotta, *Nat. Rev. Cancer*, 2006, **6**(12), 961–967.
- 20 J. L. Pichardo-Molina, C. Frausto-Reyes, O. Barbosa-Garcia, R. Huerta-Franco, J. L. Gonzalez-Trujillo, C. A. Ramirez-Alvarado, G. Gutierrez-Juarez and C. Medina-Gutierrez, *Lasers. Med. Sci.*, 2007, **22**, 229–236.
- 21 A. T. Harris, A. Lungari, C. J. Needham, S. L. Smith, M. A. Lones, S. E. Fisher, X. B. Yang, N. Cooper, J. Kirkham, D. A. Smith, D. P. Martin-Hirsch and A. S. High, *Head Neck Oncol.*, 2009, **1**, 34.
- 22 F. Domenici, A. R. Bizzarri and S. Cannistraro, *Anal. Biochem.*, 2012, **421**(1), 9–15.
- 23 J. Lin, R. Chen, S. Feng, J. Pan, Y. Li, G. Chen, M. Cheng, Z. Huang, Y. Yu and H. Zeng, *Nanomedicine*, 2011, **7**(5), 655–663.

- 24 D. Lin, S. Feng, J. Pan, Y. Chen, J. Lin, G. Chen, S. Xie, H. Zeng and R. Chen, *Opt. Express*, 2011, **9**(14), 13565–13577.
- 25 S. Feng, R. Chen, J. Lin, J. Pan, G. Chen, Y. Li, M. Cheng, Z. Huang, J. Chen and H. Zeng, *Biosens. Bioelectron.*, 2010, **25**(11), 2414–2419.
- 26 J. Bruix and M. Sherman, *Hepatology*, 2011, **53**, 1020–1022.
- 27 J. R. Beattie, J. J. McGarvey and A. W. Stitt, *Methods Mol. Biol.*, 2013, **965**, 297–312.
- 28 A. Savitzky and M. J. E. Golay, *Anal. Chem.*, 1964, **36**(8), 1627–1639.
- 29 C. A. Lieber and A. M. Jansen, *Appl. Spectrosc.*, 2003, **57**(11), 1363–1367.
- 30 C. C. Chang and C. J. Lin, *ACM Transactions on Intelligent Systems and Technology*, 2011, vol. 2, pp. 1–27.
- 31 D. R. Deegan, O. Bakajin, T. F. Dupont, G. Huber, S. R. Nagel and T. A. Witten, *Nature*, 1997, **389**, 827–829.
- 32 J. Filik and N. Stone, *Analyst*, 2007, **132**, 544–550.
- 33 C. Ortiz, D. Zhang, Y. Xie, A. E. Ribbe and D. Ben-Amotz, *Anal. Biochem.*, 2006, **353**, 157–166.
- 34 E. I. Pearce and A. Tomlinson, *Ophthalmic Physiol. Opt.*, 2000, **20**(4), 306–313.
- 35 X. C. Tang and M. J. Pikal, *Pharm. Res.*, 2005, **22**, 1167–1175.
- 36 A. Hedoux, L. Paccou, S. Achir and Y. Guinet, *J. Pharm. Sci.*, 2012, **101**, 2316–2326.
- 37 X. C. Tang and M. J. Pikal, *Pharm. Res.*, 2005, **22**, 1176–1185.
- 38 L. Chang, D. Shepherd, J. Sun, D. Ouellette, K. L. Grant, X. C. Tang and M. J. Pikal, *J. Pharm. Sci.*, 2005, **94**, 1427–1444.

# ARTICLE 2

ETUDE PILOTE IR

# ORIGINAL ARTICLES

## Profiling serologic biomarkers in cirrhotic patients via high-throughput Fourier transform infrared spectroscopy: toward a new diagnostic tool of hepatocellular carcinoma

XIAOQING ZHANG, GÉRARD THIÉFIN, CYRIL GOBINET, VALÉRIE UNTEREINER, IMANE TALEB, BRIGITTE BERNARD-CHABERT, ALEXANDRA HEURGUÉ, CAROLINE TRUNTZER, PATRICK DUCOROY, PATRICK HILLON, and GANESH D. SOCKALINGUM

REIMS AND DIJON, FRANCE; AND BEIJING, PEOPLE'S REPUBLIC OF CHINA

**Identification of novel serum biomarkers of hepatocellular carcinoma (HCC) is needed for early-stage disease detection and to improve patients' survival. The aim of this study was to evaluate the potential of serum Fourier transform infrared (FTIR) spectroscopy for differentiating sera from cirrhotic patients with and without HCC. Serum samples were collected from 2 sets of patients: cirrhotic patients with HCC (n = 39) and without HCC (n = 40). The FTIR spectra (10 per sample) were acquired in the transmission mode, and data homogeneity was tested by cluster analysis to exclude outliers. After data preprocessing by extended multiplicative signal correction and principal component analysis, the Support Vector Machine (SVM) method was applied using a leave-one-out cross-validation algorithm to classify the spectra into 2 classes of cirrhotic patients with and without HCC. When SVM was applied to all spectra (n = 790), the sensitivity and the specificity for the diagnosis of HCC were, respectively, 82.02% and 82.5%. When applied to the subset of spectra excluding the outliers (n = 739), SVM classification led to a sensitivity and specificity of 87.18% and 85%, respectively. Using median spectra for each patient instead of all replicates, the sensitivity and specificity were 84.62% and 82.50%, respectively. The overall accuracy rate was 82%–86%. In conclusion, this study suggests that FTIR spectroscopy combined with advanced methods of pattern analysis**

From MéDIAN-Biophotonique et Technologies pour la Santé, Université de Reims Champagne-Ardenne, Reims, France; Service d'Hépatogastroentérologie, Centre Hospitalo-Universitaire de Reims, Hôpital R Debré, Reims, France; Clinical Innovation Proteomic Platform, Centre Hospitalo-Universitaire de Dijon, Dijon, France; Service d'Hépatogastroentérologie, Centre Hospitalo-Universitaire de Dijon, Hôpital du Bocage, Dijon, France; Department of Anesthesiology, Peking University Third Hospital, Beijing, People's Republic of China.

Conflicts of Interest: All authors have read the journal's policy on disclosure of potential conflicts of interest and have nothing to declare.

This study was supported by the Institut National du Cancer, Fondation de France, Ligue contre le Cancer, and Cancéropôle Grand-Est (CirCE ProSpec study).

X. Zhang and C. Gobinet contributed equally to this manuscript.

The ProSpec Study group includes, in addition to the authors, J. P. Bronowicki (Centre Hospitalo-Universitaire de Nancy), V. Di Martino

(Centre Hospitalo-Universitaire de Besançon), M. Doffoel (Centre Hospitalo-Universitaire de Strasbourg), J. J. Raabe (Centre Hospitalo-Universitaire de Metz), D. Pecqueur (Centre Hospitalo-Universitaire de Dijon), A. Jeannin (Centre Hospitalo-Universitaire de Dijon), and J. Liu (Centre Hospitalo-Universitaire de Dijon).

Submitted for publication March 13, 2013; revision submitted July 9, 2013; accepted for publication July 11, 2013.

Reprint requests: Prof. Ganesh D. Sockalingum, Equipe MéDIAN, Université de Reims Champagne-Ardenne, CNRS FRE 3481-MEDyC, UFR de Pharmacie, 51 rue Cognacq-Jay, 51096 Reims Cedex, France; e-mail: ganesh.sockalingum@univ-reims.fr.

1931-5244/\$ - see front matter

© 2013 Mosby, Inc. All rights reserved.

<http://dx.doi.org/10.1016/j.trsl.2013.07.007>

**shows potential for differentiating sera from cirrhotic patients with and without HCC. (Translational Research 2013;162:279–286)**

**Abbreviations:** AFP = alpha-fetoprotein; AFSSAPS = Agence Française de Sécurité Sanitaire des Produits de Santé; EMSC = Extended Multiplicative Signal Correction; FTIR = Fourier transform infrared; HCA = Hierarchical Cluster Analysis; HCC = hepatocellular carcinoma; LOOCV = leave-one-out cross-validation; PC = principal component; PCA = principal component analysis; SVM = support vector machine

**AT A GLANCE COMMENTARY**

Zhang X, et al.

**Background**

This study describes the use of a novel approach based on biophotonics to investigate the biochemistry of serum composition and to use this information in the form of spectra to search for diagnostic markers of hepatocellular carcinoma (HCC). The method is rapid, low cost, non-contact, non-destructive, and requires a few microliters and no consumables and reagents.

**Translational Significance**

This proof-of-concept study of 79 patients demonstrates the diagnostic capabilities of high-throughput Fourier transform infrared spectroscopy. The work is interdisciplinary between physical methods and clinical. It is also part of a multicenter study in which 7 hospital centers of regions in greater eastern France are taking part to collect serum samples. About 200 control subjects and 200 patients will be studied to increase the strength of the method. Its implementation in clinical settings will then be envisaged.

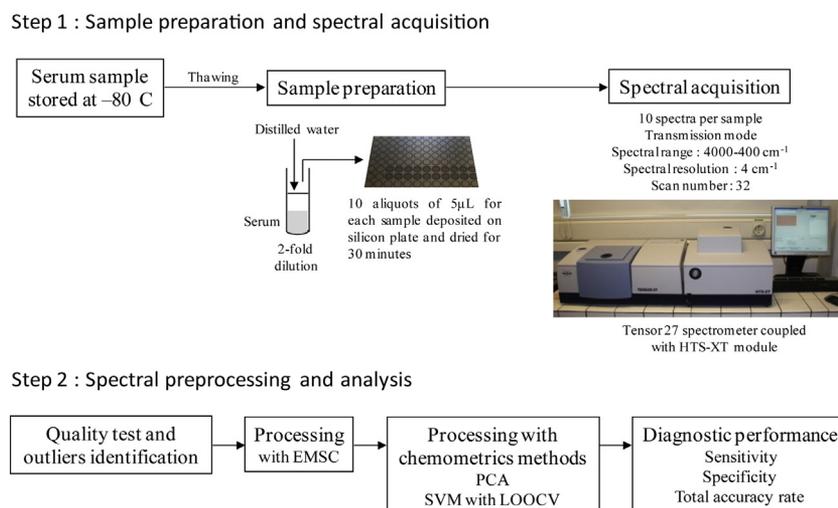
Hepatocellular carcinoma (HCC) is the sixth most commonly diagnosed cancer worldwide<sup>1</sup> and its incidence is increasing in western industrialized countries.<sup>2</sup> It is the third leading cause of cancer death.<sup>1</sup> The prognosis is very poor, with a 5-year survival of less than 10%, mainly because of the late detection of the tumor. Only a minority of patients is diagnosed at an early stage, allowing the application of potentially curative treatments such as surgical resection, liver transplantation, or local ablation with radio frequency. Therefore, identification of novel serum biomarkers of HCC is crucial to obtain an early diagnosis and to improve survival. Because most cases of HCC occur in cirrhotic patients (80%–90% of all patients), screening programs are dedicated mainly to this at-risk population.<sup>2,3</sup>

A number of serum biomarkers with potential diagnostic significance have been identified.<sup>4</sup> In the clinical

setting, alpha-fetoprotein (AFP) is the most widely used. However, a majority of HCC shows no increased AFP concentration at an early stage, and some patients with chronic liver disease have elevated AFP concentration in the absence of HCC.<sup>2,4</sup> Other biomarkers such as des-gamma-carboxy prothrombin and lectin-bound AFP have been tested.<sup>3,5</sup> They add little benefit and are not recommended in clinical practice.<sup>2</sup> There is, therefore, a consensus about the need for more reliable biomarkers.

During the past decade, proteomic technologies have been developed and used widely to detect new HCC biomarkers. Analysis of protein profiles in hepatic tissues obtained at different stages of the disease has brought a better understanding of the cellular mechanisms of HCC carcinogenesis.<sup>6,7</sup> Other studies conducted on serum and/or plasma have demonstrated the potential of the protein profiling approach to discriminate cirrhotic patients with and without HCC.<sup>8,9</sup> However, to date, none of the candidate molecules has been validated for use in clinical practice.

Besides proteomics, other techniques, including biophotonic approaches such as vibrational spectroscopy, could be useful to identify early serum spectral markers of HCC. Fourier transform infrared (FTIR) spectroscopy is a technique based on the interaction between infrared light and the electron cloud of chemical bonds of the studied sample. It is sensitive to changes in molecular compositions and structures, and has proved to be a powerful tool in detecting changes at the molecular level. When applied to tissues, cells, or biofluids, it allows the detection of spectral variations associated with different macromolecular constituents such as proteins, glucids, lipids, and nucleic acids, both qualitatively and quantitatively. Analysis of serum spectral characteristics using advanced methods of pattern recognition has been shown to be effective in identifying disease-specific signatures. The interest in the technique has been reported for serologic diagnosis of bovine spongiform encephalopathy,<sup>10</sup> rheumatoid arthritis,<sup>11</sup> myocardial infarction,<sup>12</sup> and breast cancer.<sup>13</sup> It has also been shown to be successful when used as a prognostic tool for assessing the severity of acute pancreatitis.<sup>14</sup> More recently, we have demonstrated the potential of serum FTIR spectroscopy for assessing the degree of hepatic fibrosis in patients with chronic liver disease.<sup>15</sup>



**Fig 1.** A stepwise schematic representation of the procedures used from sample preparation to spectral diagnosis. EMSC, extended multiplicative scattering correction; PCA, principal component analysis; SVM, Support Vector Machine; LOOCV, leave-one-out cross-validation.

The objective of this pilot study was to determine whether the whole-composition molecular information obtained by FTIR spectroscopy of serum can give specific spectral profiles that allow differentiating samples from cirrhotic patients with and without HCC.

## PATIENTS AND METHODS

**Study population.** This work was part of the ProSpec study, which is aimed at identifying markers of HCC in blood samples from patients with cirrhosis. The ProSpec study is an ancillary project to the CiRCE (Cirrhose et Risque de Cancer dans le Grand-Est) study, an ongoing case-control study in the northeast of France with the aim of analyzing environmental, nutritional, and metabolic factors involved in liver carcinogenesis among cirrhotic patients. The study was approved by the Comité de Protection des Personnes Est (agreement no. 2008/09) and by the Agence Française de Sécurité Sanitaire des Produits de Santé (agreement no. 2008-A00023-52). Patients gave written informed consent to participate in this study.

Serum samples for the pilot study were collected from 2 sets of patients: cirrhotic patients without HCC ( $n = 40$ ) and cirrhotic patients with HCC ( $n = 39$ ), regardless of the cause of liver disease. Immediately after blood samples were taken, they were frozen and stored at  $-80^{\circ}\text{C}$ . Diagnosis of cirrhosis was made either by histologic analysis of a liver biopsy or by a combination of clinical-biologic signs of hepatocellular failure or portal hypertension and/or endoscopic features of portal hypertension and/or imaging characteristics of cirrhosis. The absence of HCC in cirrhotic patients at inclusion was assessed through good quality imaging (ultrasonography and/or computed tomography and/or magnetic

resonance imaging) and a serum AFP level less than 100 ng/mL. None of the 40 cirrhotic patients without HCC at inclusion was diagnosed with HCC during a follow-up of at least 1 year after inclusion. The follow-up included hepatic ultrasonography and dosage of AFP every 6 months. The diagnosis of HCC relied on histologic analysis or on imaging technique characteristics as established by the European Association for Study of the Liver.<sup>16</sup> Patients younger than 35 years were not included, nor were patients with human immunodeficiency virus infection or extrahepatic cancer, and patients unable to give written informed consent to participate in the study. For each patient, the following data were collected: age, sex, cause of cirrhosis, staging of cirrhosis, serum AFP measurement, and staging of the tumor in the case of HCC.

**Infrared spectral acquisition and preprocessing.** A schematic of the sample preparation and spectral acquisition steps are shown in Fig 1. After thawing, each serum sample was diluted 2-fold with distilled water, and 10 aliquots of 5  $\mu\text{L}$  were deposited onto the wells of a 96-well silicon plate. The plate was then dried for 30 minutes at room temperature. After drying, the plate was inserted into a high-throughput module (HTS-XT) coupled to an FT-IR spectrometer (Tensor 27) both from Bruker Optics GmbH (Ettlingen, Germany). The FTIR spectra were acquired in the transmission mode (OPUS v6.5 software; Bruker Optics GmbH) in the wavenumber range of 400–4000/cm, at a nominal physical resolution of 4/cm, and averaged over 32 scans. One minute was enough to record a reference spectrum and the sample spectrum. The former was subtracted automatically from the latter to give the final absorbance spectrum. A zero-filling factor of 2

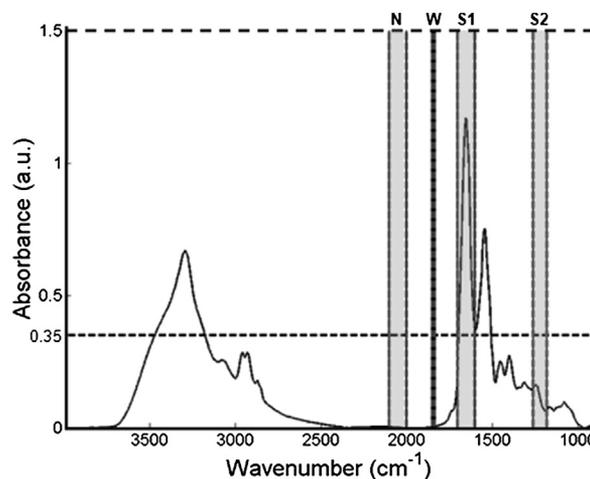
was applied, and a Blackman-Harris 3-term function was used for Fourier transformation.

All spectra were subjected to quality test to check absorbance intensity threshold, signal-to-noise ratio, and presence of water vapor content (Fig 2). Spectra with a maximum absorbance less than 0.35 and more than 2 in arbitrary units (or a.u.) were discarded. To calculate the signal-to-noise ratio, the signal was taken as the maximum absorbance of the amide I band between 1600/cm and 1700/cm (S1 value), and between 1170/cm and 1260/cm (S2 value). Noise intensity (N value) was calculated in the 2000–2100/cm region, which is devoid of spectral signature. Spectra were validated when the ratios S1/N and S2/N were above 50 and 10, respectively. Water vapor content (W value) was evaluated in the 1837–1847/cm range. Spectra were validated when S1/W and S2/W were more than 20 and 4, respectively.

Spectra that passed the quality tests were preprocessed on the wavenumber range from 890 to 4000/cm. After baseline correction, spectra were converted to second derivatives using the Savitsky-Golay algorithm, with 9 smoothing points to increase the spectral feature contrast. Then, their second derivatives were vector normalized. This spectral preprocessing was performed using the OPUS software (v6.5, Bruker Optics GmbH). Last, the processed spectra were saved and examined with hierarchical cluster analysis using Euclidean distance calculation and Ward's algorithm to exclude outliers. Hierarchical cluster analysis was performed using the spectral information contained in the regions of 2800–3100/cm and 900–1800/cm.

**Statistical analysis.** Demographic, clinical, and biologic data were evaluated for all patients. Quantitative data are expressed as mean  $\pm$  standard error and were compared using Student's *t* test. Qualitative data are expressed as a percentage and were compared using the  $\chi^2$  test. A *P* value less than 0.05 was considered significant.

Spectral analysis was performed using the Matlab software (Matlab, MathWorks, Inc., Natick, Mass., USA). After exclusion of outliers, all raw spectra that passed the quality test were processed by extended multiplicative signal correction to perform a baseline correction and to normalize them.<sup>17</sup> This method modeled each acquired spectrum as a weighted sum of a target spectrum that was the mean spectrum of all spectra, of a polynomial function modeling light-scattering effects that we chose as the fourth order, and of the modeling error.<sup>18</sup> The corrected spectra resulted from the subtraction of the estimated polynomial function from the original spectra. Normalization was achieved by dividing the corrected spectra by the estimated weights of the target spectrum. Fig 3 displays



**Fig 2.** The spectral quality test is performed for each spectrum by calculating the ratios of band intensities S1/N, S2/N, S1/W, and S2/W, where S1 is the highest absorbance between 1600/cm and 1700/cm, S2 is the highest absorbance between 1170/cm and 1260/cm, N is the noise in the 2000–2100/cm range, and W is the residual water in the 1837–1847/cm region. A spectrum passed the quality test if the following conditions were satisfied: S1/N greater than 50, S2/N greater than 10, S1/W greater than 20, and S2/W greater than 4.

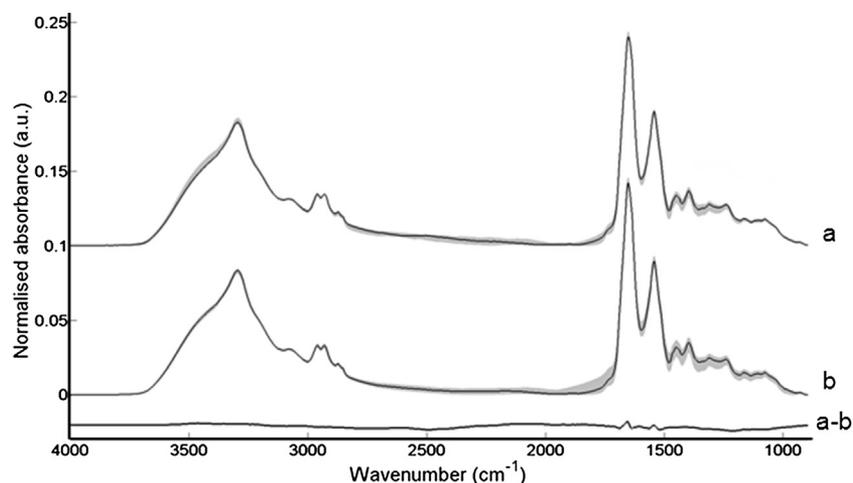
the median spectra with their interquartile range from each group of patients after preprocessing.

With the corrected spectra normalized around the same target spectrum and with weak the spectral differences between the 2 groups of patients (see difference spectrum in Fig 3), we decided to subtract the noninformative target spectrum from all the individual spectra.

To recover the remaining spectral information, a principal component analysis (PCA) was performed. This multivariate method decomposes the data into a weighted sum of uncorrelated principal components (PCs).<sup>19</sup> In the data set coordinate system, the first PC becomes the first axis of a new coordinate system and explains the highest variance of the data set. The second PC becomes the second axis, orthogonal to the first one, and explains the next highest variance of the data set, and so on. The PCs are also called loading vectors whereas their weights are known as the PC scores.<sup>19</sup>

To select the most discriminating PCs, a statistical test was conducted on the scores of each PC. The probability distribution of the scores being unknown, the Mann-Whitney nonparametric statistical test was used, which allows reorganizing PCs in ascending order of their *P* value, with the first one being the most discriminating between the 2 patient groups.

In a first step, an unsupervised classification method based on PCA was used to analyze the spectra from patients with and without HCC, and to determine whether the 2 sets of spectra could be distinguished. The method



**Fig 3.** Preprocessed median serum infrared spectra with their interquartile range (shaded regions) from the group of cirrhotic patients with hepatocellular carcinoma (HCC) (a) and from the group of cirrhotic patients without HCC (b). The difference spectrum (a-b) shows very minor variations. Spectra are offset for clarity but are at the same scale.

was applied using the most significant PCs and then the most discriminating ones.

In a second step, Support Vector Machine (SVM), a supervised classification method,<sup>20</sup> was performed using the library for support vector machines (LIBSVM).<sup>21</sup> Given the small number of samples, the SVM model was applied with a cross-validation using a leave-one-out cross-validation (LOOCV) procedure, either by removing all replicate spectra or the median spectrum for a given patient. Only PCs explaining 99.99% of the total variance were used to optimize the SVM model. PCs that accounted for background noise were discarded. The sensitivity and specificity of the classifier were calculated first by using all spectra and then by using a subset of spectra that excluded outliers. The spectral processing and SVM classification was also applied to the median spectra calculated for each patient.

## RESULTS

**Patient characteristics.** There was no significant difference in age, sex ratio, and distribution of etiologic factors between cirrhotic patients with and without HCC (Table I). According to the Child-Pugh score, cirrhosis was significantly more severe in cirrhotic patients without HCC ( $7.4 \pm 0.4$  pt vs  $6.3 \pm 0.3$  pt;  $P = 0.025$ ). As expected, a serum level of AFP more than 500 ng/mL, a level considered pathognomonic of HCC in cirrhotic patients, was more frequent in cirrhotic patients with HCC than in those without HCC (15% vs 0%;  $P = 0.03$ ; Table I). HCC was unifocal in 35.9% of the patients, bifocal in 15.4%, and multifocal in 48.7%. Extrahepatic spreading of the tumor was diagnosed in 10.3% of the patients.

**FTIR spectral analysis.** All spectra from 79 patients ( $n = 790$  spectra) met the quality test. Fifty-one spectra were assigned as outliers by cluster analysis: 24 in the group of cirrhotic patients with HCC and 27 in the group of cirrhotic patients without HCC. Based on these results, 3 data sets were created: data set 1 containing all 790 spectra; data set 2, a subset excluding the aforementioned 51 outliers; and data set 3, median spectra from all patients. After extended multiplicative signal correction preprocessing and PCA, the SVM method was applied using an LOOCV procedure. When the SVM classification was applied to the whole set of spectra, patients were classified directly according to the assignment of the majority of their spectra. If half or less than 50% of the spectra from 1 patient were classified incorrectly, the patient classification was considered to be wrong. For obvious reasons, this procedure did not apply to SVM classification based on the median spectra from all patients. In this case, patients were classified according to the assignment of their median spectrum.

When PCA was applied to the 3 data sets of spectra, 131 PCs accounted for 99.99% of the total variance in the whole set of 790 spectra (data set 1), 130 PCs in the subset of 739 spectra excluding the outliers (data set 2), and 51 PCs in the set of 79 median spectra from all patients (data set 3). The unsupervised classification method PCA did not allow us to separate serum from patients with and without HCC, using either the most significant PCs (not shown) or the most discriminating PCs in the 3 sets of spectra (data not shown).

When the LOOCV-SVM method was applied to the whole set of spectra (data set 1), the optimal classification model was built with the 74 most discriminating

**Table I.** Demographic, clinical, and biological characteristics of cirrhotic patients with and without HCC

	Cirrhotic patients with HCC (n = 39)	Cirrhotic patients without HCC (n = 40)	P value*
Age, y	65.8 ± 10.5	61.9 ± 9	NS
Sex ratio (males/females), n	34/5	31/9	NS
Child-Pugh score, pt	6.3 ± 1.7	7.4 ± 2.6	0.025
Cause of cirrhosis, n <sup>†</sup>			NS
Alcoholic	34/39	37/40	
C virus	3/39	3/40	
B virus	4/39	0/40	
Metabolic	3/39	3/40	
Alpha-fetoprotein level > 500 ng/mL	6/39	0/40	0.03

Abbreviations: HCC, hepatocellular cancer; NS, not significant.

\*Student's *t* test for quantitative data (mean ± standard deviation) and the  $\chi^2$  test for qualitative data were used.

<sup>†</sup>Several causes may be combined in some patients.

**Table II.** Sensitivity and specificity of the SVM classification model to differentiate serum spectra of patients with and without HCC

	Data set 1 (790 spectra)	Data set 2 (739 spectra)	Data set 3 (79 median spectra)
Patients with HCC, n	39	39	39
Correctly classified	32	34	33
Wrongly classified	7	5	6
Patients without HCC, n	40	40	40
Correctly classified	33	34	33
Wrongly classified	7	6	7
Sensitivity, %	82.05	87.18	84.62
Specificity, %	82.50	85	82.50
Total accuracy, %	82.3	86.1	83.54

Abbreviations: SVM, Support Vector Machine; HCC, hepatocellular cancer.

The SVM classification model was applied to the entire set of 790 spectra (data set 1), to the subset of 739 spectra excluding outliers (data set 2), and to the set of 79 median spectra (data set 3) from all patients.

PCs, explaining 42% of the variance. The classification procedure led to an overall sensitivity of 82.05% and an overall specificity of 82.5%. The total accuracy rate was 82.3% (Table II). When analyzing the set of spectra excluding the outliers (data set 2), the optimal SVM model was built with 54 PCs, explaining 42% of the variance. Using these PCs, the LOOCV-SVM procedure led to sensitivity and specificity rates of 87.18% and 85%, respectively. The total accuracy rate was 86.1% (Table II). Last, when using the same procedure with median spectra from all patients (data set 3) instead of all replicates, the LOOCV-SVM model was built with 21 PCs

(10.94% of the variance). The overall sensitivity and specificity were 84.62% and 82.50%, respectively. The total accuracy rate was 83.54% (Table II).

## DISCUSSION

Despite monitoring programs, most cirrhotic patients with HCC are diagnosed at an advanced stage of the tumor. This situation justifies continuous research to identify new biomarkers or to develop new and rapid methods for biomarker profiling to obtain an early diagnosis allowing curative treatments. Our study shows that serum FTIR spectroscopy can potentially differentiate cirrhotic patients with and without HCC, although specific biomarkers cannot be identified because the methodology is based on the global molecular information of the whole serum composition.

During the past decade, proteomic technologies have been used to identify new biomarkers specific to HCC. A number of studies have analyzed protein profiles in hepatic tissues obtained from patients with or without HCC.<sup>6,7,22-24</sup> Of interest, some studies have shown that proteins expressed differentially between normal and tumor tissues, such as human liver carboxylesterase<sup>123</sup> or vimentin,<sup>24</sup> were also expressed differentially in the serum of the same patients. Other studies using surface-enhanced laser desorption ionization/time-of-flight mass spectrometry have been conducted exclusively on serum and/or plasma from patients with or without HCC. Potential biomarkers such as the C-ter fragment of vitronectin,<sup>9</sup> cystatin C,<sup>25</sup> and clusterin<sup>26</sup> have been identified. However, these candidate molecules have not been validated as serologic biomarkers for early HCC diagnosis in clinical practice. Recently, several studies have shown the significance of posttranslational modifications (mainly glycosylation and phosphorylation) in the carcinogenesis process and as potential prognostic markers.<sup>27,28</sup>

In our study, the working hypothesis was that the serum from cirrhotic patients with HCC exhibits infrared spectral characteristics reflecting the presence of tumor spectral biomarkers. In the field of oncology, FTIR spectroscopy has been used mainly for studying structure and molecular composition of macromolecules present in tumor cells and tissues. The potential of the technique has been reported for skin,<sup>7</sup> cervical,<sup>29</sup> gallbladder carcinomas,<sup>30</sup> and lymphomas.<sup>31</sup> It has also been applied to analyze single cancer cells.<sup>32,33</sup> In addition, it has been reported that infrared imaging can identify molecular changes in the tissue preceding the morphologic features of cancer, thus allowing an early diagnosis.<sup>34,35</sup> Application of FTIR spectroscopy to the serum has been much more limited, but promising results have been reported for the diagnosis of noncancerous diseases.<sup>10,11,15</sup> Very few data are available in the field of oncology. Backhaus et al<sup>13</sup>

demonstrated that this technique combined with artificial neural networks classification or with cluster analysis was able to differentiate patients with and without breast cancer with sensitivity and specificity rates greater than 90%.

Application of vibrational spectroscopy to plasma/serum has a number of advantages. Blood sampling is a very noninvasive method, is easily available at a low cost, and can be repeated conveniently for diagnostic or prognostic purposes during the evolution of the disease. Because blood perfuses all organs and tissues, plasma/serum composition is assumed to reflect the status of the body. In case of a tumor, proteins and peptides are shed from the tumor microenvironment into the newly formed microcirculation. As a result, tumor markers are expected to be found in the circulatory system, either as whole functional proteins or as cleaved proteins. FTIR spectroscopy provides a spectral fingerprint of all molecules contained in the plasma/serum sample. Given the high number of spectral variables (about 1000 absorption wavelengths), the use of sophisticated mathematical and chemometric methods such as artificial neuronal networks or SVM<sup>10</sup> is necessary to identify discriminating spectral zones reflecting different molecular composition, thus enabling the construction of robust classifiers. It must be emphasized that spectral changes reflect differences in the absorbance of particular infrared wavelengths; they do not allow identification of specific macromolecules, such as specific proteins, underlying these changes.

In our study, the unsupervised classification method PCA did not allow us to distinguish serum from cirrhotic patients with HCC from serum from patients without HCC, indicating that spectral differences are subtle and necessitate the use of more sophisticated classification methods. Our relatively small number of patients led us to use an LOOCV procedure for the SVM classification to avoid overfitting the model. A PCA was performed to reduce redundant spectral information, and the classification was performed on the whole set of spectra, on the set of spectra excluding the outliers, and on median spectra from each patient. The results appear robust, with a rate of total diagnostic accuracy ranging from 82% to 86% using the 3 sets of data. Median spectra have the advantage of being less influenced by outliers and they avoid the redundant use of replicates from the same sample during the classification process.

Our study adds support to the working hypothesis that serum from cirrhotic patients with HCC exhibits an infrared spectral signature that allows its differentiation from cirrhotic patients without HCC. A small amount of serum, without reagent, is sufficient for this purpose. To develop this new diagnostic tool further in clinical

practice, the next step is to validate the results in a larger number of samples, to build a classifier with an advanced method of pattern recognition such as SVM, and to test this classifier prospectively with an independent set of serum samples.

The authors are grateful to Thomas Feron for his excellent technical support. The Infrastructures Biologie, Santé, Agronomie technological platform "Imagerie Cellulaire et Tissulaire" is acknowledged.

## REFERENCES

1. Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomarkers Prev* 2010;19:1893–907.
2. El-Serag HB. Hepatocellular carcinoma. *N Engl J Med* 2011;365:1118–27.
3. Lok AS, Sterling RK, Everhart JE, et al. Des-gamma-carboxy prothrombin and alpha-fetoprotein as biomarkers for the early detection of hepatocellular carcinoma. *Gastroenterology* 2010;138:493–502.
4. Malaguarnera G, Giordano M, Paladina I, Berretta M, Cappellani A, Malaguarnera M. Serum markers of hepatocellular carcinoma. *Dig Dis Sci* 2010;55:2744–55.
5. Marrero JA, Feng Z, Wang Y, et al. Alpha-fetoprotein, des-gamma-carboxy prothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. *Gastroenterology* 2009;137:110–8.
6. Corona G, De Lorenzo E, Elia C, et al. Differential proteomic analysis of hepatocellular carcinoma. *Int J Oncol* 2010;36:93–9.
7. Li N, Long Y, Fan X, et al. Proteomic analysis of differentially expressed proteins in hepatitis B virus-related hepatocellular carcinoma tissues. *J Exp Clin Cancer Res* 2009;28:122–31.
8. Schwegler EE, Cazares L, Steel LF, et al. SELDI-TOF MS profiling of serum for detection of the progression of chronic hepatitis C to hepatocellular carcinoma. *Hepatology* 2005;41:634–42.
9. Paradis V, Degos F, Dargère D, et al. Identification of a new marker of hepatocellular carcinoma by serum protein profiling of patients with chronic liver diseases. *Hepatology* 2005;41:40–7.
10. Martin TC, Moecks J, Belousov A. Classification of signatures of bovine spongiform encephalopathy in serum using infrared spectroscopy. *Analyst* 2004;129:897–901.
11. Staib A, Dolenko B, Fink DJ, et al. Disease pattern recognition testing for rheumatoid arthritis using infrared spectra of human serum. *Clin Chim Acta* 2001;308:79–89.
12. Haas SL, Müller R, Fernandes A, et al. Spectroscopic diagnosis of myocardial infarction and heart failure by Fourier transform infrared spectroscopy in serum samples. *Appl Spectrosc* 2010;64:262–7.
13. Backhaus J, Mueller R, Formanski N, et al. Diagnosis of breast cancer with infrared spectroscopy from serum samples. *Vibration Spectrosc* 2010;52:173–7.
14. Petrov MS, Gordetzov AS, Kukosh MV. Early prediction of severity in acute pancreatitis using infrared spectroscopy of serum. *Pancreatol* 2007;7:451–8.
15. Scaglia E, Sockalingum GD, Schmitt J, et al. Noninvasive assessment of hepatic fibrosis in patients with chronic hepatitis C using serum Fourier transform infrared spectroscopy. *Anal Bioanal Chem* 2011;401:2919–25.
16. Bruix J, Sherman M, Llovet JM, et al. Clinical management of hepatocellular carcinoma: conclusions of the Barcelona-2000 EASL conference: European Association for the Study of the Liver. *J Hepatol* 2001;35:421–30.
17. Kohler A, Kirschner C, Oust A, Martens H. 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;59:707–16.
18. Ly E, Piot O, Durlach A, Bernard P, Manfait M. Differential diagnosis of cutaneous carcinomas by infrared spectral micro-imaging combined with pattern recognition. *Analyst* 2009;134:1208–14.
  19. Abdi H, Williams LJ. Principal component analysis. Wiley online library, *WIREs Comp Stat* 2010;2:433–59. <http://dx.doi.org/10.1002/wics.101>.
  20. Cristianini N, Shawe-Taylor J. Support vector machines and other kernel-based learning methods. Cambridge, UK: Cambridge University Press, 2000.
  21. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. New York, NY: ACM Transactions on Intelligent Systems and Technology 2011;2:1–27. Software available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
  22. Yeo M, Mi Na Y, Kim DK, et al. The loss of phenol sulfotransferase 1 in hepatocellular carcinogenesis. *Proteomics* 2010;10:266–76.
  23. Na K, Lee EY, Lee HJ, et al. Human plasma carboxylesterase 1, a novel serologic biomarker candidate for hepatocellular carcinoma. *Proteomics* 2009;9:3989–99.
  24. Sun S, Poon RT, Lee NP, et al. Proteomics of hepatocellular carcinoma: serum vimentin as a surrogate marker for small tumors (< or = 2 cm). *J Proteome Res* 2010;9:1923–30.
  25. Zinkin NT, Grall F, Bhaskar K, et al. Serum proteomics and biomarkers in hepatocellular carcinoma and chronic liver disease. *Clin Cancer Res* 2008;14:470–7.
  26. Kimura A, Sogawa K, Satoh M, et al. The application of a three-step serum proteome analysis for the discovery and identification of novel biomarkers of hepatocellular carcinoma. *Int J Proteomics* 2012;2012:623190.
  27. Dai Z, Zhou J, Qiu SJ, Liu YK, Fan J. Lectin-based glycoproteomics to explore and analyze hepatocellular carcinoma-related glycoprotein markers. *Electrophoresis* 2009;30:2957–66.
  28. Lee HJ, Na K, Kwon MS, Kim H, Kim KS, Paik YK. Quantitative analysis of phosphopeptides in search of the disease biomarker from the hepatocellular carcinoma specimen. *Proteomics* 2009;9:3395–408.
  29. Murali Krishna C, Sockalingum GD, Vidyasagar MS, et al. An overview on applications of optical spectroscopy in cervical cancers. *J Cancer Res Ther* 2008;4:26–36.
  30. Du JK, Shi JS, Sun XJ, et al. Fourier transform infrared spectroscopy of gallbladder carcinoma cell line. *Hepatobiliary Pancreat Dis Int* 2009;8:75–8.
  31. Babrah J, McCarthy K, Lush RJ, Rye AD, Bessant C, Stone N. Fourier transform infrared spectroscopic studies of T-cell lymphoma, B-cell lymphoid and myeloid leukaemia cell lines. *Analyst* 2009;134:763–8.
  32. Draux F, Jeannesson P, Gobinet C, et al. IR spectroscopy reveals effect of non-cytotoxic doses of anti-tumour drug on cancer cells. *Anal Bioanal Chem* 2009;395:2293–301.
  33. Pijanka JK, Sockalingum GD, Kohler A, et al. Synchrotron based FTIR spectra of stained single cells: towards a clinical application in pathology. *Lab Invest* 2010;90:797–807.
  34. Nallala J, Gobinet C, Diebold MD, et al. Infrared spectral imaging as a novel approach for histopathological recognition in colon cancer diagnosis. *J Biomed Opt* 2012;17:116013.
  35. Nallala J, Piot O, Diebold MD, et al. Infrared imaging as a cancer diagnostic tool: introducing a new concept of spectral barcodes for identifying molecular changes in colon tumors. *Cytometry A* 2013;83:294–300.

## PUBLICATIONS ET COMMUNICATIONS

Ces travaux ont fait l'objet des publications et communications suivantes :

### **Articles de revues**

---

- I.Taleb, G.Thiéfin, C. Gobinet, V.Gaydou, B. Bernard-Chabert, Bronowicki, Raabe, Doffoel, Di martino, P Hillon P.Ducoroy, G.D. Sockalingum, for the ProSpec Study Group., Hepatocellular carcinoma diagnosis from blood serum using FTIR spectroscopy and chemometric methods. **Article en préparation.**
- Taleb I, Thiéfin G, Gobinet C, Untereiner V, Bernard-Chabert B, Heurgué A, Truntzer C, Hillon P, Manfait M, Ducoroy P, Sockalingum GD., Diagnosis of hepatocellular carcinoma in cirrhotic patients: a proof-of-concept study using serum micro-Raman spectroscopy, **Analyst**. 2013.
- Zhang X, Thiéfin G, Gobinet C, Untereiner V, Taleb I, Bernard-Chabert B, Heurgué A, Truntzer C, Ducoroy P, Hillon P, Sockalingum GD.; Profiling serologic biomarkers in cirrhotic patients via high-throughput Fourier transform infrared spectroscopy: toward a new diagnostic tool of hepatocellular carcinoma, **Trans Res**. 2013.

### **Communications orales**

---

- Taleb I, Sockalingum GD, Gobinet C, Untereiner V, Heurgué A, Ducoroy P, Hillon P, Manfait M, Thiéfin G, ProSpec Study Group. Diagnosis of hepatocellular carcinoma in cirrhotic patients: a proof-of-concept study using serum Raman microspectroscopy. **SPEC2012**, Chiang Mai, Thailand, 11 - 16 November 2012.
- Zhang X, Gobinet C, Scaglia E, Untereiner V, Taleb I, P Ducoroy, P Hillon, Manfait M, Sockalingum GD, Thiéfin G, on behalf of the PROSPEC Study Group. Identification of serum biomarkers predictive of hepatocellular carcinoma in cirrhotic patients. **Oncotrans 2011**, 2ème Colloque Inter-Régional Grand-Est de Recherche Translationnelle en Oncologie, Nancy, 8-9 septembre 2011.

**Communications affichées**

---

- Taleb I, Sockalingum GD, Gobinet C, Untereiner V, Heurgué A, Ducoroy P, Hillon P, Manfait M, Thiéfin G, ProSpec Study Group. Diagnosis of hepatocellular carcinoma in cirrhotic patients: a proof-of-concept study using serum Raman microspectroscopy. **Oncotrans 2013**, 3ème colloque inter-Régional Grand-Est de Recherche Translationnelle en Oncologie. Du laboratoire au patient et du patient au laboratoire. Reims, 27-28 juin 2013.
- Taleb I, Sockalingum GD, Gobinet C, Untereiner V, Heurgué A, Ducoroy P, Hillon P, Manfait M, Thiéfin G, ProSpec Study Group. Diagnosis of hepatocellular carcinoma in cirrhotic patients using serum Raman microspectroscopy. **47 th annual meeting of the EASL**, Barcelone, 18-22 avril 2012.
- Zhang X, Sockalingum GD, Gobinet C, Untereiner V, Scaglia E, Taleb I, Bernard-Chabert B, Ducoroy P, Hillon P, Manfait M, Thiéfin G. Intérêt de l'analyse du sérum par spectroscopie infrarouge à transformée de Fourier pour le diagnostic de carcinome hépatocellulaire chez les patients atteints de cirrhose. **Journées Francophones d'Hépatogastroentérologie et d'Oncologie Digestive**, paris, 15-18 mars 2012.
- Untereiner V, Sockalingum GD, Garnotel R, Gobinet C, Taleb I, Ramaholimihaso F, Ehrhart F, Heurgué A, Bernard-Chabert B, M. Manfait M, Thiéfin G. Bile analysis using Fourier-transform infrared spectroscopy combined with chemometrics for the diagnosis of malignant biliary strictures: a pilot study in 57 patients. 3ème colloque inter-Régional Grand-Est de Recherche Translationnelle en Oncologie. Du laboratoire au patient et du patient au laboratoire. **Oncotrans 2013**. Reims, 27-28 juin 2013.

---

**Apport de la spectroscopie vibrationnelle, infrarouge et Raman, appliquée au sérum pour le diagnostic de carcinome hépatocellulaire chez les patients atteints de cirrhose.**

---

Le carcinome hépatocellulaire (CHC) est la 3<sup>ème</sup> cause de mortalité par cancer dans le monde. L'identification de nouveaux marqueurs sériques est cruciale pour améliorer le pronostic. Dans ce travail, nous avons évalué l'intérêt de la spectroscopie vibrationnelle, infrarouge et Raman, appliquée au sérum pour le diagnostic du CHC. Dans un premier temps, nous avons réalisé 2 études pilotes pour évaluer le potentiel de ces deux approches. L'intérêt de la spectroscopie Raman appliquée au sérum a été évalué chez 37 patients cirrhotiques avec CHC et 34 patients cirrhotiques sans CHC. L'analyse des données spectrales a permis de classer les patients avec un taux d'exactitude diagnostique de 85 à 91%. Nous avons également démontré l'intérêt diagnostique de la spectroscopie IRTF dans une population de 40 patients avec CHC et 39 patients sans CHC. Dans cette étude, le taux d'exactitude diagnostique était de 82 à 86%. Dans un deuxième temps, nous avons mené une étude sur un plus grand nombre de patients afin de valider les résultats obtenus dans l'étude pilote IRTF. Les caractéristiques spectrales IRTF du sérum de 308 patients cirrhotiques avec CHC ont été comparées à celles du sérum de 509 patients cirrhotiques sans CHC. L'analyse des données spectrales par deux méthodes de classification supervisée, SVM et PLS-DA, n'a pas permis de confirmer les résultats obtenus dans cette étude pilote. Avec un taux d'exactitude diagnostique entre 46 et 61 %, l'analyse spectrale IRTF du sérum entier n'apparaît pas discriminante pour distinguer les patients cirrhotiques avec et sans CHC. Ce résultat souligne la nécessité de confirmer sur un grand nombre de patients les résultats obtenus dans des études pilotes. L'analyse spectrale ciblée sur des fractions du sérum pourrait permettre d'identifier plus efficacement des marqueurs diagnostiques en évitant la superposition des informations spectrales liées aux multiples molécules présentes dans le sérum entier.

---

**Application of infrared and Raman vibrational spectroscopy to serum analysis for the diagnosis of hepatocellular carcinoma in cirrhotic patients.**

---

Hepatocellular carcinoma (HCC) is the third cause of cancer death in the world. The identification of novel serum markers is crucial to improve the prognosis. In this work, we evaluated the potential of vibrational spectroscopy, infrared and Raman spectroscopy, applied to serum to diagnose HCC. On a first step, two pilot studies were conducted to evaluate these two approaches. Raman spectroscopy applied to the serum was tested in 37 cirrhotic patients with HCC and 34 cirrhotic patients without HCC. Analysis of spectral data showed that it was possible to classify patients with a diagnostic accuracy rate of 85 to 91%. We also demonstrated the diagnostic performance of FTIR spectroscopy in a population of 40 patients with HCC and 39 patients without HCC. In this study, the diagnostic accuracy rate was 82 to 86%. In a second step, an FTIR study on a larger number of patients was performed to validate the results obtained in the pilot study. FTIR spectral characteristics of 308 sera from cirrhotic patients with HCC were compared with those of 509 cirrhotic patients without HCC. The supervised classification methods, SVM and PLS -DA were applied but did not confirm the results obtained in the pilot study. The diagnostic accuracy was between 46 and 61%, FTIR spectral analysis of whole serum does not appear discriminant enough to differentiate cirrhotic patients with and without HCC. This result highlights the need to confirm on a large number of patients results in pilot studies. Spectral analysis of serum fractions could be an alternative to more effectively identify diagnostic markers avoiding overlapping spectral information related to the complex composition of whole serum.

---

**Discipline:** Biophysique

---

**Mots clés:** spectroscopie vibrationnelle, sérum, carcinome hépatocellulaire, Raman et IRTF, chimie

---

**Unité de recherche:** Equipe MéDIAN, Biophotoniques et Technologies pour la Santé, Unité MEDyC, CNRS FRE3481-Université de Reims Champagne-Ardenne, UFR Pharmacie, SFR CAP-Santé FED4231--51 rue Cognac-Jay, Reims, France.

---