Université de Reims Champagne-Ardenne UFR Pharmacie

Année 2010

N°

THESE

Présentée pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE DE REIMS CHAMPAGNE-ARDENNE

Mention : Ingénierie de la santé

Soutenue publiquement le 29/01/2010

Par

Adrian TRAVO

né le 13 décembre 1980 à Strasbourg

Caractérisation de marqueurs du développement tumoral par spectroscopie Infra-Rouge

Unité MéDYC CNRS UMR 6237 – IFR 53 UFR Pharmacie de Reims

JURY

Rapporteurs :Pr. Gérard DELERIS (Bordeaux)
Pr. Marie-Christine RIO (Strasbourg)Examinateurs :Dr. Dominique GUENOT (Strasbourg)
Pr. Marie-Danièle DIEBOLD (Reims)
Pr. Dieter NAUMANN (Berlin)Directeurs de thèse :Pr. Pierre JEANNESSON (Reims)
Pr. Michel MANFAIT (Reims)

Université de Reims Champagne-Ardenne UFR Pharmacie

Année 2010

N°

THESE

Présentée pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE DE REIMS CHAMPAGNE-ARDENNE

Mention : Ingénierie de la santé

Soutenue publiquement le 29/01/2010

Par

Adrian TRAVO

né le 13 décembre 1980 à Strasbourg

Caractérisation de marqueurs du développement tumoral par spectroscopie Infra-Rouge

Unité MéDYC CNRS UMR 6237 – IFR 53 UFR Pharmacie de Reims

JURY

Rapporteurs :Pr. Gérard DELERIS (Bordeaux)
Pr. Marie-Christine RIO (Strasbourg)Examinateurs :Dr. Dominique GUENOT (Strasbourg)
Pr. Marie-Danièle DIEBOLD (Reims)
Pr. Dieter NAUMANN (Berlin)Directeurs de thèse :Pr. Pierre JEANNESSON (Reims)
Pr. Michel MANFAIT (Reims)

Remerciements

A Monsieur le Professeur Michel Manfait,

Merci de m'avoir fait confiance en m'intégrant dans votre équipe, et en me proposant ce qui était pour moi un réel défi. Je suis arrivé à l'unité MéDIAN en ne connaissant presque rien à la spectroscopie vibrationnelle...j'en ressors avec une nouvelle vision de la biologie. Les connaissances que j'ai pu acquérir au cours de ces trois années ont remis en cause certains de mes acquis pour mieux consolider mon savoir, et me permettront, j'en suis certain, d'aborder mes futurs projets de recherche avec un regard original et un esprit ouvert. Encore une fois, merci.

A Madame le Professeur Marie-Christine Rio,

Je vous remercie sincèrement pour l'honneur que vous me faites en acceptant d'être rapporteur de ce travail. Veuillez accepter mes remerciements, et soyez assurée de mon respect et de ma gratitude.

A Monsieur le Professeur Gérard Déléris,

Vos publications ont souvent côtoyé mon bureau au cours de ces dernières années, c'est donc un honneur de vous voir rapporteur de mon travail. Je vous remercie sincèrement d'avoir accepté cette invitation, soyez également assuré de mon respect et de ma gratitude.

A Madame le Docteur Dominique Guénot,

Je vous adresse mes remerciements les plus sincères pour avoir accepté d'être membre du Jury. Merci également de m'avoir accueilli dans votre labo à de nombreuses reprises, et toujours avec le sourire.

Für Professor Dieter Naumann,

Ich möchte Sie, für ihre nette Prüfung meiner Arbeit, recht herzlich bedanken. Ihre Anwesenheit in dieser Prünfungskomission ist eine große Ehre für mich.

A Madame le Professeur Marie-Danièle Diebold,

Je vous adresse mes remerciements les plus respectueux pour avoir accepté d'être membre de mon jury de thèse.

A Monsieur le Professeur Pierre Jeannesson,

Au cours de ces trois années passées ensemble, et je ne pourrais jamais te remercier assez pour cela, tu m'auras appris à prendre du recul, à être critique, à me remettre en cause, à être curieux, à organiser mes idées...bref, à réfléchir. Nos discussions, souvent détendues, parfois moins, furent toujours, pour moi,

source de progrès dans mon travail. Je souhaite à tout thésard de tomber sur un encadrant comme toi, même si ce n'est pas facile tout les jours! ;) Merci infiniment pour tout ce que tu as fais pour moi, pour ta patience, ton soutien, ta répartie et ton humour. Tu fais partie de ces personnes qui, j'en suis déjà certain, me marqueront et m'inspirerons pour la vie. Dorénavant, chaque fois que je devrais mettre en ordre mes idées, rédiger des articles ou préparer des communications, je penserais à toi et à ce que tu m'as appris...Tu resteras pour moi une référence.

A Monsieur le Docteur Olivier Piot,

Tu m'auras apporté un soutien morale (et matériel, merci pour le matelas gonflable!) que tu ne soupçonnes peut-être pas. De tes paroles réconfortantes lors de mes débuts pour le moins difficile, à tes nombreux encouragements quand les idées et les résultats sont arrivés, tu m'as toujours aidé à surmonter les petits (et grands) moments de découragements. Alors pour cela, pour ta bonne humeur perpétuelle, pour tes conseils et ta disponibilité, pour toutes les corrections d'articles et autres manuscrits, pour notre folle et mémorable course à travers Paris suite à notre excursion brésilienne...pour tout cela, et pour le reste, je te dis : MERCI!

A Monsieur le Professeur Jacques Bara et Mme Elisabeth Forgue-Lafitte,

Je tiens à vous remercier encore une fois, mais cela ne sera jamais assez, de votre collaboration enthousiaste. Ce travail je vous le dois, vous n'avez cessé de me soutenir et de m'encourager dans mon approche...je vous suis infiniment reconnaissant.

Un grand merci à mes amis et collègues : Valérie (merci pour les coups de fils, l'agrafeuse, les conseils, le sourire et tout le reste!), Mme Pizani, Ali, Flo, Julien, Mathilde, Marie, Ganesh, aux fous furieux du traitement des données Cyril, David et Rolf, à Elodie (merci pour tout !)...et à tous les étudiants, stagiaires et chercheurs de l'Unité grâce à qui j'ai pu, pendant trois ans, travailler dans un environnement sympathique et efficace.

Je tiens également à remercier tout mes amis d'ici ou d'ailleurs : Hadrien, Teddy, Gaspard, John et Steph, Beber, Ben, Bilou (tu es roux mais je t'aime quand même), Valentine, Marlène et Moïse, tout le « Cheinsow Crew » (Keep on rockin'!) et les métalleux croisés sur les routes du Sud, sans oublier mes amis d'enfance DJ Bastos, Alex, Chach, Gitan, Seb...

Enfin, je remercie du plus profond de mon cœur que mes très chers parents, mon frère et ma bien-aimée Amandine, sans qui je ne serais pas ici à écrire ces quelques lignes...

Je dédie cette thèse à ma sœur...

Publications internationales

IR spectral imaging for histopathological characterization of xenografted human colon carcinomas.

Wolthuis R, **Travo A**, Nicolet C, Neuville A, Gaub MP, Guenot D, Ly E, Manfait M, Jeannesson P, Piot O.

Anal Chem, 80, 8461-8469, 2008.

FTIR spectral imaging of secreted mucus: a promising new tool for histopathological recognition of human colonic adenocarcinomas.

Travo A, Piot O, Wolthuis R, Gobinet C, Manfait M, Bara J, Forgue-Lafitte ME, Jeannesson P. (Accepté pour publication dans *Histopathology*, 2009).

(publications antérieures)

Lipoplex and peptide-based strategies for the delivery of steric-block oligonucleotides. Resina S, Abes S, Turner JJ, Prevot P, **Travo A**, Clair P, Gait MJ, Thierry AR, Lebleu B. *Int J Pharm*, 344, 96-102, 2007.

Switching on transgene expression by correcting aberrant splicing using multi-targeting steric-blocking oligonucleotides.

Resina S, Kole R, Travo A, Lebleu B, Thierry AR.

J Gene Med, 9, 498-510, 2007.

Comparison of basic peptides- and lipid-based strategies for the delivery of splice correcting oligonucleotides.

Thierry AR, Abes S, Resina S, **Travo A**, Richard JP, Prevot P, Lebleu B. *Biochim Biophys Acta*, 1758, 364-74, 2006.

Communications orales

Characterization of secreted mucus by FTIR Spectral Imaging.
<u>Travo A</u>, Piot O, Wolthuis R, Ly E, Manfait M, Bara J, Forgue-Lafitte ME, Jeannesson P. *SPEC 2008*, São José Dos Campos, Brésil, Octobre 2008.

Characterization of secreted mucus by FTIR Spectral Imaging.

<u>Travo A</u>, Piot O, Wolthuis R, Ly E, Manfait M, Bara J, Forgue-Lafitte ME, Jeannesson P. *Réunion IFR 53 Axe 1, contrat CPER : cellules cancéreuses – microenvironnement et progression tumorale*, Reims, France, Novembre 2008.

Infra-Red spectral imaging of secreted mucus: a promising tool for the histopathological recognition of human colon adenocarcinomas. (*)

<u>Travo A</u>, Jeannesson P, Wolthuis R, Gobinet C, Manfait M, Bara J, Forgue-Lafitte M, Piot O. *CECED 2009*, Strasbourg, France, Février 2009.

Infra-Red spectral imaging of secreted mucus: a promising tool for the histopathological recognition of human colon adenocarcinomas.

<u>Travo A</u>, Jeannesson P, Wolthuis R, Gobinet C, Manfait M, Bara J, Forgue-Lafitte M, Piot O. *ONCOTRANS 2009*, Nancy, France, Février 2009.

IR spectral imaging of mucus : a potential tool for histopathological recognition of colon adenocarcinomas. (*)

<u>Travo A</u>, Jeannesson P, Wolthuis R, Gobinet C, Manfait M, Bara J, Forgue-Lafitte M, Piot O. Journée des Jeunes Chercheurs de l'IFR 53, Reims, France, Avril 2009.

Implémentation de l'imagerie spectrale infrarouge en oncologie.

Ly E, <u>Travo A</u>, Jeannesson P, Manfait M, Piot O. *Colloque IMVIE 2009*, Mulhouse, France, Juin 2009.

Communications par voie d'affiche

Infrared micro spectral imaging of xenografted colon tumors.
Wolthuis R, <u>Travo A</u>, Jeannesson P, Guénot D, Gaub MP, Abecassis J, Manfait M, Piot O.
1^{er} forum du Canceropôle Grand-Est, Vittel, Octobre 2007.

Infrared imaging based-histology for molecular characterization of tumoral tissue.
Piot O, Jeannesson P, Ly E, <u>Travo A</u>, Untereiner V, Wolthuis R, Manfait M.
99th AACR Annual Meeting, San Diego, USA, Avril 2008.

IR spectral imaging for histopathological characterization of tumoral tissue.

Wolthuis R, <u>**Travo** A</u>, Nicolet C, Neuville A, Gaub MP, Guenot D, Manfait M, Jeannesson P, Piot O. 2^{ème} forum du Canceropôle Grand-Est, Vittel, Octobre 2008.

In-situ analysis of mucus by FTIR spectral imaging in colon adenocarcinomas: towards a potential tumor marker.

<u>**Travo** A</u>, Piot O, Wolthuis R, Ly E, Gobinet C, Manfait M, Bara J, Forgue-Lafitte ME, Jeannesson P. *100th AACR Annual Meeting*, Denver, USA, Avril 2009.

In-situ analysis of mucus by FTIR spectral imaging in colon adenocarcinomas: towards a potential tumor marker.

<u>**Travo** A</u>, Piot O, Wolthuis R, Ly E, Gobinet C, Manfait M, Bara J, Forgue-Lafitte ME, Jeannesson P. 3^{ème} forum du Canceropôle Grand-Est, Strasbourg, Novembre 2009.

(*) <u>Récompenses</u>

- Lauréat d'un des deux prix de présentation orale **ONCOTRANS 2009**
- ✤ Lauréat du prix de la meilleure présentation orale JJC IFR-53 2009

LISTE DES FIGURES

Figure 1. Représentation schématique de la progression tumorale...6

- Figure 2. Exemple de tissus colorés à l'HE et à l'HES...8
- Figure 3. Le diagramme de Jablonski...19
- Figure 4. Exemple de modes de vibration moléculaire d'élongation et de déformation...20

Figure 5. Exemple d'un spectre IR...21

Figure 6. Schéma de principe d'un microspectromètre IRTF...24

Figure 7. Influence de l'utilisation d'un module de purge sur le signal de l'eau et du CO₂ atmosphérique...**25**

Figure 8. Spectre IRTF de référence...26

Figure 9. Obtention du spectre d'un échantillon en spectroscopie IRTF...26

Figure 10. Exemple de spectres IR de mono- et di-saccharides sur la gamme spectrale 1000-1200 cm⁻¹...**32**

Figure 11. Spectres IR de tissus de foie humain et exemple d'attribution des régions spectrales...35

Figure 12. Exemple d'image spectrale IR de coupe tissulaire...36

Figure 13. Exemple de supports pour l'histologie IR...42

Figure 14. Influence de la résolution spectrale sur les données IR...45

Figure 15. Influence de la résolution spatiale sur les populations de spectres obtenus par imagerie IR...47

Figure 16. Illustration du principe de classification KM de données spectrales...51

Figure 17. Illustration du principe de classification HA de données spectrales...52

Figure 18. Comparaison de l'histologie IR d'un adénocarcinome colorectal selon la méthode de

classification des données spectrales employée...53

Figure 19. Processus de mise au point d'un modèle de prédiction de l'histologie IR...57

Figure 20. Diagnostic différentiel de lésions tumorales de la peau par histologie IR supervisée...58

Figure 21. Anatomie du colon...61

Figure 22. Histologie du colon...62

Figure 23. Modèle de Volgelstein illustrant les étapes de la progression de la tumorogénèse

colorectale...64

Figure 24. Classification TNM des cancers coliques...66

Figure 25. Aspect structural des mucines sécrétées...70

Figure 26. Spectro-imageur IRTF Spectrum Spotlight 300...79

Figure 27. Modélisation des interférences pour le pré-traitement des spectres IR de coupes

paraffinées...82

Figure 28. Coefficient de régression (a) et (e) du modèle EMSC pour tous les spectres d'une image IR de tissu...84

Figure 29. Effet du pré-traitement EMSC sur les données spectrales de coupes tissulaires paraffinées...84

Figure 30. Effet du pré-traitement EMSC sur une classification des données par classification KM...85

Figure 31. Exemple de marquage anti-Ki67 d'un échantillon de xénogreffe de carcinome de colon... 94

Figure 32. Image IRPC d'une coupe de xénogreffe de carcinome du colon...96

Figure 33. Signatures spectrales des clusters déterminés par classification KM...97

Figure 34. Comparaison des images IRPC d'une coupe de xénogreffe de carcinome de colon obtenues par classification KM ou modèle de prédiction LDA...**99**

Figure 35. Images IRPC de deux échantillons de l'ensemble de validation obtenues avec le modèle de prédiction LDA...**100**

Figure 36. Probabilités d'attributions des spectres aux clusters 5 et 9 dans les échantillons #241, #314 et #250...**102**

Figure 37. Organisation spatiale des clusters spectraux...104

Figure 38. Représentation schématique des deux principales séquences d'organisation tissulaire incluant les clusters de tumeur 5, 9 et 10...**105**

Figure 39. Différences spectrales dans la principale séquence de clusters...106

Figure 40. Différences spectrales dans la principale séquence de clusters incluant les tissus

sécréteurs...107

Figure 41. Le spectromètre IRTF Tensor 37 HTS-XT et micro-plaques 96 puits...116

Figure 42. Spectres IR moyens de mucines gastro-intestinales mammifères purifiées et lyophilisées...118

Figure 43. Spectres IR moyens d'échantillons lyophilisés de mucus ovariens de kystes bénins...119

Figure 44. Spectres IR moyens de mucus des voies respiratoires issus de cultures cellulaires...120

Figure 45. Spectres IR moyens de mucus colorectaux humains prélevés sur des pièces opératoires...121

Figure 46. Exemple de coupe de colon sain et tumoral en « Swiss roll » colorées à l'HE...129

Figure 47. Images IRPC et signatures IR des clusters spectraux de coupes tissulaires de colons sains et d'adénocarcinomes...**133**

Figure 48. Organisation spatiale et attribution des clusters spectraux 1/6, 2/6 et 3/6 dans les coupes de colons sains...**134**

Figure 49. Images IRPC spécifiques des mucus de coupes tissulaires de colons sains et

d'adénocarcinomes...136

Figure 50. Pourcentages relatifs des sous-clusters, en fonction du nombre de sous-clusters fixé lors la classification KM...137

Figure 51. Dérivées secondes (x-1) des spectres moyens des sous-clusters 1/3 et 2/3...139

Figure 52. Exemples d'images IRPC de mucus d'un adénocarcinome colique, après classification KM en 3 sous-clusters...**140**

Figure 53. Pourcentages relatifs des sous-clusters en fonction du nombre de sous-clusters et des gammes spectrales utilisés pour la classification des spectres de mucus...**141**

Figure 57. Discrimination des échantillons sains et tumoraux basée sur la distribution des sous-clusters après classification sur gammes spectrales spécifiques...143

Figure 58. Dérivées secondes des spectres moyens des sous-clusters obtenus après classification sur gammes spectrales spécifiques...144

Figure 59. Illustration représentative de la comparaison des images IRPC des mucus avec les marquages IHC de MUC2 et MUC5AC...**147**

Figure 1A. Exemples de modes de vibration de molécules complexes...167

Figure 2A. Principe du mode de mesure en transmission...168

Figure 3A. Principes de la réflexion diffuse et spéculaire...169

Figure 4A. Principe du mode de mesure en réflexion-absorption...170

Figure 5A. Principe du mode de mesure en réflexion totale atténuée (ATR)...171

Figure 6A. Ensemble d'individus N décrit par 3 variables x, y et z...172

Figure 7A. Ensemble d'individus N décrit par ses 2 premières composantes principales, x et y...173

Figure 8A. Ensemble d'individus N décrit par sa première composante principale...173

LISTE DES TABLEAUX

Tableau 1. Principales bandes d'absorption IR de la liaison peptidique...29

Tableau 2. Exemple de composantes des bandes d'absorption IR Amides I et II...30

Tableau 3. Exemples de bandes d'absorption IR des acides nucléiques...33

Tableau 4. Exemple de bandes d'absorption IR des phospholipides...33

Tableau 5. Types de supports utilisables en histologie IR...42

Tableau 6. Comparaison du temps de calcul total pour la classification KM, FCM, et HA d'une image spectrale...**54**

Tableau 7. Modifications de l'expression des mucines dans les adénocarcinomes coliques...73

Tableau 8. Caractéristiques histologiques des xénogreffes...92

 Tableau 9. Caractéristiques cliniques et histologiques des adénocarcinomes coliques étudiés...128

Tableau 10. Comparaison entre l'évaluation immunohistochimique de l'expression de MUC2 et

MUC5AC et les pourcentages de sous-clusters dans les images IRPC des échantillons tumoraux...146

ABREVIATIONS

ACP: Analyse en Composantes Principales

ADN: Acide Désoxyribonucléïque

APC: Adenopolyposis Coli

ARN: Acide Ribonucléïque

ARNm: ARN messager

CaF2: Fluorure de Calcium

HA: Classification Hiérarchique Ascendante

EMSC: Extended Multiplicative Signal Correction

HE: Hématoxyline-Eosine

IHC: *Immunohistochimie*

IR: Infra-Rouge

IRPC: *IR pseudo-couleur*

IRTF: Infra-Rouge à Transformé de Fourier

KM: K-Means (K-Moyennes)

LDA: Linear Discriminant Analysis

S/B: Signal sur Bruit (rapport)

TNM: Tumor Node Metastasis

ZnSe: Séléniure de Zinc

Sommaire

SOMMAIRE

CONTEXTE SCIENTIFIQUE ET TECHNIQUE	AVANT-PROPOS	1
ANATOMOPATHOLOGIE ET CARACTERISATION DES LESIONS TUMORALES	CONTEXTE SCIENTIFIQUE ET TECHNIQUE	4
TUMORALES 5 I. Préparation des échantillons tissulaires 7 II. Evaluation diagnostique des lésions tumorales 9 III. Evaluation pronostique et prédictive. 9 IV. Les nouvelles techniques d'analyses des prélèvements tissulaires 11 V. Les limites de la caractérisation des lésions tumorales en anatomo-pathologie 14 A. La subjectivité de l'examen anatomopathologique 14 B. Un diagnostic basé sur la morphologie 15 C. Des techniques chronophages et non automatisables 16 LA MICROSPECTROSCOPIE VIBRATIONNELLE IR. 18 I. Principe de la Microspectroscopie IR 18 I. Microspectroscopie IRTF et macromolécules biologiques 29 A. Protéines. 29 B. Carbohydrates. 31 C. Les acides nucléques 32 D. Les lipides 33 III. Microspectroscopie IRTF et échantillons biologiques 34 IV. Imagerie par microspectroscopie IRTF 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 L'HISTOLOGIE SPECTRALE IR 41 I. Les échantillons 47 L'A gaunespectrale <t< th=""><th>ANATOMOPATHOLOGIE ET CARACTERISATION DES LESIONS</th><th></th></t<>	ANATOMOPATHOLOGIE ET CARACTERISATION DES LESIONS	
I. Préparation des échantillons tissulaires 7 II. Evaluation diagnostique des lésions tumorales 9 III. Evaluation pronostique et prédictive. 9 IV. Les nouvelles techniques d'analyses des prélèvements tissulaires 11 V. Les limites de la caractérisation des lésions tumorales en anatomo-pathologie 14 A. La subjectivité de l'examen anatomopathologique 14 B. Un diagnostic basé sur la morphologie. 15 C. Des techniques chronophages et non automatisables 16 LA MICROSPECTROSCOPIE VIBRATIONNELLE IR. 18 I. Principe de la Microspectroscopie IR. 18 I. Microspectroscopie IRTF et macromolécules biologiques 29 B. Carbohydrates. 29 B. Carbohydrates. 31 C. Les acides nucléiques 32 D. Les lipides 33 II. Microspectroscopie IRTF et échantillons biologiques 34 V. Imagerie par microspectroscopie IRTF. 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 I. HISTOLOGIE SPECTRALE IR 41 I. Acquisition de l'image spectrale d'un tissu 43 A. Gamme spectrale 44 <tr< th=""><th>TUMORALES</th><th> 5</th></tr<>	TUMORALES	5
II. Evaluation diagnostique des lésions tumorales 9 III. Evaluation pronostique et prédictive. 9 IV. Les nouvelles techniques d'analyses des prélèvements tissulaires 11 V. Les limites de la caractérisation des lésions tumorales en anatomo-pathologie. 14 A La subjectivité de l'exame natomopathologique. 14 B. Un diagnostic basé sur la morphologie. 15 C. Des techniques chronophages et non automatisables 16 LA MICROSPECTROSCOPIE VIBRATIONNELLE IR. 18 I. Principe de la Microspectroscopie IR. 18 I. Microspectroscopie IRTF et macromolécules biologiques 29 A. Protéines 29 B. Carbohydrates. 31 C. Les acides nucléiques 32 D. Les lipides 33 III. Microspectroscopie IRTF et échantillons biologiques 34 IV. Imagerie par microspectroscopie IRTF 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 L'HISTOLOGIE SPECTRALE IR 41 I. Les échantillons 43 A. Gamme spectrale 43 A. Gamme spectrale 43 B. Résolution spatiale 45	I. Préparation des échantillons tissulaires	7
III. Evaluation pronostique et prédictive 9 IV. Les nouvelles techniques d'analyses des prélèvements tissulaires 11 V. Les limites de la caractérisation des lésions tumorales en anatomo-pathologie 14 A. La subjectivité de l'examen anatomopathologique 14 B. Un diagnostic basé sur la morphologie 15 C. Des techniques chronophages et non automatisables 16 LA MICROSPECTROSCOPIE VIBRATIONNELLE IR 18 I. Principe de la Microspectroscopie IR 18 I. Microspectroscopie IRTF et macromolécules biologiques 29 A. Protéines 29 B. Carbohydrates 31 C. Les acides nucléiques 32 D. Les lipides 33 III. Microspectroscopie IRTF et échantillons biologiques 34 IV. Imagerie par microspectroscopie IRTF 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 L'HISTOLOGIE SPECTRALE IR 41 I. Acquisition de l'image spectrale d'un tissu 43 A. Gamme spectrale 43 B. Résolution spatiale 44 C. Résolution spectrale 47 D. Le nombre d'accumulations <td< td=""><td>II. Evaluation diagnostique des lésions tumorales</td><td>9</td></td<>	II. Evaluation diagnostique des lésions tumorales	9
IV. Les nouvelles techniques d'analyses des prélèvements tissulaires 11 V. Les limites de la caractérisation des lésions tumorales en anatomo-pathologie 14 A. La subjectivité de l'examen anatomopathologique 14 B. Un diagnostic basé sur la morphologie 15 C. Des techniques chronophages et non automatisables 16 LA MICROSPECTROSCOPIE VIBRATIONNELLE IR 18 I. Principe de la Microspectroscopie IR 18 I. Microspectroscopie IRTF et macromolécules biologiques 29 B. Carbohydrates 31 C. Les acides nucléques 32 D. Les lipides 33 III. Microspectroscopie IRTF et échantillons biologiques 34 IV. Imagerie par microspectroscopie IRTF 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 L'HISTOLOGIE SPECTRALE IR 41 I. Les échantillons 43 A. Gamme spectrale 43 B. Résolution spectrale 43 B. Résolution spectrale 44 C. Résolution spectrale 43 B. Résolution spectrale 43 B. Résolution spectrale 43	III. Evaluation pronostique et prédictive	9
V. Les limites de la caractérisation des lésions tumorales en anatomo-pathologie 14 A. La subjectivité de l'examen anatomopathologique 14 B. Un diagnostic basé sur la morphologie 15 C. Des techniques chronophages et non automatisables 16 LA MICROSPECTROSCOPIE VIBRATIONNELLE IR 18 I. Principe de la Microspectroscopie IR 18 I. Microspectroscopie IRTF et macromolécules biologiques 29 A. Protéines 29 B. Carbohydrates 31 C. Les acides nucléiques 32 D. Les lipides 33 II. Microspectroscopie IRTF et échantillons biologiques 34 IV. Imagerie par microspectroscopie IRTF 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 L'HISTOLOGIE SPECTRALE IR 41 I. Les échantillons 43 A. Gamme spectrale 43 B. Résolution spectrale 43 B. Résolution spectrale 44 C. Résolution spectrale 43 B. Résolution spectrale 43 B. Résolution spectrale 43 B. Résolution spectrale 43	IV. Les nouvelles techniques d'analyses des prélèvements tissulaires	11
A. La subjectivité de l'examen anatomopathologique 14 B. Un diagnostic basé sur la morphologie. 15 C. Des techniques chronophages et non automatisables 16 LA MICROSPECTROSCOPIE VIBRATIONNELLE IR. 18 I. Principe de la Microspectroscopie IR. 18 I. Microspectroscopie IRTF et macromolécules biologiques. 29 A. Protéines. 29 B. Carbohydrates. 31 C. Les acides nucléiques 32 D. Les lipides 33 III. Microspectroscopie IRTF et échantillons biologiques 34 IV. Imagerie par microspectroscopie IRTF 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 L'HISTOLOGIE SPECTRALE IR 41 I. Acquisition de l'image spectrale d'un tissu 43 A. Gamme spectrale 44 C. Résolution spectrale 44 C. Résolution spectrale 44 M. Casifications non-supervisées : vers l'histologie IR 47 III. Le pré-traitement des données 48 IV. Traitements chimiométriques pour la classification des données 50 A. Les classifications non-supervisées : vers l'histologie IR	V. Les limites de la caractérisation des lésions tumorales en anatomo-pathologie	14
B. Un diagnostic basé sur la morphologie 15 C. Des techniques chronophages et non automatisables 16 LA MICROSPECTROSCOPIE VIBRATIONNELLE IR 18 I. Principe de la Microspectroscopie IR 18 I. Microspectroscopie IRTF et macromolécules biologiques 29 A. Protéines 29 B. Carbohydrates 31 C. Les acides nucléiques 32 D. Les lipides 33 III. Microspectroscopie IRTF et échantillons biologiques 34 IV. Imagerie par microspectroscopie IRTF 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 L'HISTOLOGIE SPECTRALE IR 41 I. Acquisition de l'image spectrale d'un tissu 43 A. Gamme spectrale 43 B. Résolution spatiale 47 III. Le pré-traitement des données 47 III. Les classifications non-supervisées : vers l'histologie IR 50 A. Les classifications non-supervisées : vers l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 55	A. La subjectivité de l'examen anatomopathologique	14
LA MICROSPECTROSCOPIE VIBRATIONNELLE IR	B. Un diagnostic basé sur la morphologie	15
LA MICROSPECTROSCOPIE VIBRATIONNELLE IR	C. Des techniques chronophages et non automatisables	16
I. Principe de la Microspectroscopie IR	I A MICROSPECTROSCOPIE VIRRATIONNELLE IR	18
I. Initipe de la introspectroscopie IRTF et macromolécules biologiques 29 A. Protéines 29 B. Carbohydrates 31 C. Les acides nucléiques 32 D. Les lipides 33 III. Microspectroscopie IRTF et échantillons biologiques 34 IV. Imagerie par microspectroscopie IRTF 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 L'HISTOLOGIE SPECTRALE IR 41 I. Les échantillons 43 A. Gamme spectrale 43 B. Résolution spectrale 43 B. Résolution spectrale 44 C. Résolution spectrale 45 D. Le nombre d'accumulations 47 III. Le pré-traitement des données 48 IV. Traitements chimiométriques pour la classification des données 50 A. Les classifications non-supervisées : vers l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 50	I Princine de la Microspectrosconie IR	10 18
A. Protéines. 29 A. Protéines. 29 B. Carbohydrates. 31 C. Les acides nucléiques 32 D. Les lipides 33 III. Microspectroscopie IRTF et échantillons biologiques 34 IV. Imagerie par microspectroscopie IRTF 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 L'HISTOLOGIE SPECTRALE IR 41 I. Les échantillons 41 I. Acquisition de l'image spectrale d'un tissu 43 A. Gamme spectrale 43 B. Résolution spectrale 44 C. Résolution spectrale 45 D. Le nombre d'accumulations 47 III. Le pré-traitement des données 48 IV. Traitements chimiométriques pour la classification des données 50 A. Les classifications non-supervisées : vers l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 50	II Microspectroscopie IRTE et macromolécules biologiques	10 20
B. Carbohydrates 31 C. Les acides nucléiques 32 D. Les lipides 33 III. Microspectroscopie IRTF et échantillons biologiques 34 IV. Imagerie par microspectroscopie IRTF 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 L'HISTOLOGIE SPECTRALE IR 41 I. Les échantillons 41 I. Acquisition de l'image spectrale d'un tissu 43 A. Gamme spectrale 43 B. Résolution spectrale 44 C. Résolution spectrale 44 C. Résolution spectrale 44 M. Cassifications non-supervisées : vers l'histologie IR 48 IV. Traitements chimiométriques pour la classification des données 50 A. Les classifications non-supervisées : vers l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 50	A. Protéines	
C. Les acides nucléiques 32 D. Les lipides 33 III. Microspectroscopie IRTF et échantillons biologiques 34 IV. Imagerie par microspectroscopie IRTF 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 L'HISTOLOGIE SPECTRALE IR 41 I. Les échantillons 41 I. Les échantillons 41 I. Acquisition de l'image spectrale d'un tissu 43 A. Gamme spectrale 43 B. Résolution spectrale 44 C. Résolution spectrale 47 III. Le pré-traitement des données 48 IV. Traitements chimiométriques pour la classification des données 50 A. Les classifications non-supervisées : vers l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 50 B. Les modèles de prédiction de la caractérisation des lésions tumorales 57	B. Carbohydrates	31
D. Les lipides 33 III. Microspectroscopie IRTF et échantillons biologiques 34 IV. Imagerie par microspectroscopie IRTF 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 L'HISTOLOGIE SPECTRALE IR 41 I. Les échantillons 41 I. Les échantillons 41 I. Les échantillons 43 A. Gamme spectrale 43 B. Résolution spectrale 44 C. Résolution spectrale 45 D. Le nombre d'accumulations 47 III. Le pré-traitement des données 48 IV. Traitements chimiométriques pour la classification des données 50 A. Les classifications non-supervisées : vers l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 50 S. Les modèles de prédiction de l'histologie IR 50 Y. Applications dans le domaine de la caractérisation des lésions tumorales 57	C. Les acides nucléiques	
III. Microspectroscopie IRTF et ecnantilions biologiques 34 IV. Imagerie par microspectroscopie IRTF 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 L'HISTOLOGIE SPECTRALE IR 41 I. Les échantillons 41 I. Acquisition de l'image spectrale d'un tissu 43 A. Gamme spectrale 43 B. Résolution spectrale 44 C. Résolution spectrale 45 D. Le nombre d'accumulations 47 III. Le pré-traitement des données 48 IV. Traitements chimiométriques pour la classification des données 50 A. Les classifications non-supervisées : vers l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 55 V. Applications dans le domaine de la caractérisation des lésions tumorales 57	D. Les lipides	
IV. Imagerie par microspectroscopie IRTF 35 V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical 37 L'HISTOLOGIE SPECTRALE IR 41 I. Les échantillons 41 I. Acquisition de l'image spectrale d'un tissu 43 A. Gamme spectrale 43 B. Résolution spectrale 44 C. Résolution spectrale 45 D. Le nombre d'accumulations 47 III. Le pré-traitement des données 48 IV. Traitements chimiométriques pour la classification des données 50 A. Les classifications non-supervisées : vers l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 50 S. Les modèles de prédiction de l'histologie IR 50 S. Les modèles de prédiction de l'histologie IR 50 S. Les modèles de prédiction de l'histologie IR 50 S. Les modèles de prédiction de l'histologie IR 50 S. Les modèles de prédiction de l'histologie IR 50 S. Les modèles de prédiction de l'histologie IR 50 S. Les modèles de prédiction de l'histologie IR 50	III. Microspectroscopie IR IF et echantilions biologiques	34
v. Les applications de la inicrospectroscopie IKTF dans les domaines de la biologie et du biomédical	IV. Imagerie par microspectroscopie IKIF	
Jointeurcat 57 L'HISTOLOGIE SPECTRALE IR 41 I. Les échantillons. 41 II. Acquisition de l'image spectrale d'un tissu 43 A. Gamme spectrale 43 B. Résolution spectrale 43 D. Le nombre d'accumulations 47 III. Le pré-traitement des données 48 IV. Traitements chimiométriques pour la classification des données 50 A. Les classifications non-supervisées : vers l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 55 V. Applications dans le domaine de la caractérisation des lésions tumorales 57	v. Les applications de la microspectroscopie IRTF dans les domaines de la biologie	et du 27
L'HISTOLOGIE SPECTRALE IR 41 I. Les échantillons 41 II. Acquisition de l'image spectrale d'un tissu 43 A. Gamme spectrale 43 B. Résolution spectrale 43 C. Résolution spatiale 45 D. Le nombre d'accumulations 47 III. Le pré-traitement des données 48 IV. Traitements chimiométriques pour la classification des données 50 A. Les classifications non-supervisées : vers l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 55 V. Applications dans le domaine de la caractérisation des lésions tumorales 57	Diometrical	
I. Les échantillons	L'HISTOLOGIE SPECTRALE IR	41
II. Acquisition de l'image spectrale d'un tissu 43 A. Gamme spectrale 43 B. Résolution spectrale 44 C. Résolution spatiale 45 D. Le nombre d'accumulations 47 III. Le pré-traitement des données 48 IV. Traitements chimiométriques pour la classification des données 50 A. Les classifications non-supervisées : vers l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 55 V. Applications dans le domaine de la caractérisation des lésions tumorales 57	L Ing i OLOGIE DI ECTRALE IR	/11
A. Gamme spectrale 43 B. Résolution spectrale 43 C. Résolution spectrale 44 C. Résolution spatiale 45 D. Le nombre d'accumulations 47 III. Le pré-traitement des données 48 IV. Traitements chimiométriques pour la classification des données 50 A. Les classifications non-supervisées : vers l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 55 V. Applications dans le domaine de la caractérisation des lésions tumorales 57	I. Les condition de l'image spectrale d'un tissu	/13
B. Résolution spectrale	A. Gamme spectrale	
C. Résolution spatiale	B. Résolution spectrale	
D. Le nombre d'accumulations	C. Résolution spatiale	45
III. Le pré-traitement des données 48 IV. Traitements chimiométriques pour la classification des données 50 A. Les classifications non-supervisées : vers l'histologie IR 50 B. Les modèles de prédiction de l'histologie IR 55 V. Applications dans le domaine de la caractérisation des lésions tumorales 57	D. Le nombre d'accumulations	
IV. Traitements chimiométriques pour la classification des données	III. Le pre-traitement des données	48
 A. Les classifications non-supervisees : vers i instologie IR	IV. I raitements chimiometriques pour la classification des données	50
V. Applications dans le domaine de la caractérisation des lésions tumorales	A. Les classifications non-supervisées : vers i histologie in	
	V. Applications dans le domaine de la caractérisation des lésions tumorales	57

LE CANCER DU COLON	60
I. Anatomie du colon	
II. Histologie du colon	61
III. Formes histologiques du cancer du colon	63
IV. Développement et hétérogénéité du cancer du colon	63
V. Facteurs pronostiques et prédictifs	65
A. La classification TNM et autres facteurs morphologiques	
B. Les marqueurs moleculaires	6/
VI. Les mucines coliques secrétées	
A. Aspect structural des mucines secrétées	
 B. Fonctions spécifiques jouées par les mucines coliques secrétées dans la barrière épithéliale C. Les altérations des mucines pendant la tumorogénèse colique 	71 73
CHAPITRE 1: Pré-traitement de données spectrales issues à coupes tissulaires paraffinées pour l'histologie IR	le 76
INTRODUCTION	77
MATERIELS ET METHODES	
I. Echantillons tissulaires	
II. Acquisition des données spectrales IR	
III. Traitement des données	79
RESULTATS	81
I Mise au noint d'un modèle de correction des données basée sur l'EMSC	81
II. Effet du modèle de correction sur les données spectrales	
DISCUSSION	87
CHAPITRE 2: Caractérisation histopathologique de xénogr	effes
de carcinome du colon humain par histologie IR	89
INTRODUCTION	90
MATERIELS ET METHODES	92
I. Echantillons tissulaires	
II. Acquisition des données spectrales IR	
III. Traitement des données	
IV. Immunohistochimie	
RESULTATS	96
I. Construction des images IRPC de l'ensemble d'entrainement	
II. Construction et validation d'un modèle de prédiction de l'histologie IR	

III. Analyse de la probabilité d'attribution des spectres dans le modèle	
IV. Analyse spectrale de l'organisation des tissus	
V. Interprétation moléculaire des différences spectrales	
VI. Corrélation avec l'IHC et les données cliniques	
DISCUSSION	109
CHAPITRE 3. Correctórication spectrale IR d'échaptille	ms do
CHAI II KE 5. Caracterisation spectrale IK u echantuno	
mucus et de mucines	
INTRODUCTION	114
MATERIELS ET METHODES	115
I. Origines des échantillons de mucus et de mucines	
II. Acquisition et pré-traitement des données spectrales	
	110
RESULTATS	
I. Analyse IR de mucines gastro-intestinales mammifères purifiées	
II. Analyse IR d'échantillons de mucus	119
DISCUSSION	122
CHAPITRE 4: Histologie IR des mucus secrétés pour la caractérisation d'adénocarcinomes du colon	
INTRODUCTION	127
MATERIELS ET METHODES	
I Fchantillons tissulaires	128
II. Acquisition des données spectrales IR	
III. Traitement des données	
IV. Détection de l'expression IHC de MUC2 et MUC5AC	
V. Analyse statistique	
	101
KESULIAIS I. Histologie IR non-supervisée de tissus coliques sains et d'adénocarcinome	131 S
II. Analyse de l'hétérogénéité spectrale des mucus et discrimination des éch	antillons 135
III. Mise en évidence d'un nouveau marqueur de discrimination entre tissu	s coliques
sains et adénocarcinomes du colon	
A. Investigation des différences spectrales	143
B. Comparaison entre profil d'expression des mucines et distribution des sous-clusters	145
DISCUSSION	148

CONCLUSION GENERALE	
RÉFÉRENCES	157
ANNEXES	166
Publications associées	

Avant-propos

Le cancer est aujourd'hui la première cause de mortalité en France en étant responsable de plus de 145 000 décès au cours de la seule année 2008 (Grünfeld 2009). Malgré une mobilisation très importante de la communauté scientifique mondiale sur cette pathologie, le processus de développement de la plupart des cancers est encore mal compris, bien qu'il soit admis que la plupart d'entre eux se développent à partir d'un état pré-cancéreux via de multiples altérations génétiques successives. De par ces altérations, les cancers apparaissent comme des pathologies très complexes, pouvant suivre différentes voies biomoléculaires de développements, ceci donnant naissance à une hétérogénéité intra- et inter-tumorale pouvant expliquer l'échec des démarches de diagnostic et d'évaluation pronostique basées sur l'étude de l'expression de marqueurs protéiques ou moléculaires ciblés. En effet, actuellement, le diagnostic des lésions tumorales repose essentiellement sur l'analyse morphologique histopathologique de prélèvements tissulaires par un anatomopathologiste. Le rôle de cette analyse morphologique est également étendu à l'évaluation de l'évolution des lésions tumorales (aspect pronostique) et de la prise en charge thérapeutique (aspect prédictif) la plus adéquate pour chaque patient, grâce aux systèmes de classification des lésions cancéreuses. Cependant, de par l'aspect morphologique de ces évaluations, l'hétérogénéité tumorale ainsi que les changements biomoléculaires précoces pouvant précéder les altérations morphologiques des tissus ne sont alors pas considérées.

Il existe donc aujourd'hui un réel besoin pour des outils d'investigation tissulaire pouvant prendre en compte la complexité et l'hétérogénéité biomoléculaire des pathologies tumorales. Dans ce contexte, les techniques d'analyses haut-débit du génome et du protéome, permettant d'établir de véritables cartes d'identités des tumeurs, suscitent un fort engouement tant dans le milieu de la recherche que dans le milieu clinique, même si de nombreuses contraintes techniques n'autorisent pas encore leur utilisation en routine. Par ailleurs, un autre type de technique, encore peu connu des biologistes et anatomopathologistes, connaît un très fort développement dans le cadre de la caractérisation des lésions tumorales : l'imagerie par microspectroscopie vibrationnelle Infrarouge (IR), permettant de sonder la constitution biochimique d'un tissu. En effet, depuis plus de 10 ans maintenant, les études démontrant le potentiel de cette technique pour l'analyse de tissus tumoraux se sont multipliées, et établissent aujourd'hui la microspectroscopie IR comme un outil permettant de reproduire l'histologie tissulaire sur la base de la composition moléculaire intrinsèque des échantillons. (Diem 2004 ; Lasch 2004 ; Kendall 2009 ; Wang 2007).

Les travaux présentés dans ce manuscrit visent à démontrer que la microspectroscopie IR, par delà cette capacité de reproduction automatique et objective de l'histologie conventionnelle, peut révéler de subtiles variations biomoléculaires indépendantes de la morphologie des structures tissulaires. La détection *in situ* de telles variations, pouvant potentiellement précéder certaines altérations morphologiques, peut alors permettre l'obtention d'un nouveau type d'histologie moléculaire, plus informatif que l'histologie classique basée sur des colorations de type Hématoxyline-Eosine (HE).

Ainsi la partie **contexte scientifique et technique** de ce manuscrit rappellera tout d'abord les méthodes actuelles de diagnostics et d'évaluations des lésions tumorales et exposera certaines de leurs limites. Par la suite, les notions essentielles liées à la spectroscopie IR et au concept d'histologie spectrale seront également développées. Enfin, puisque les modèles d'études principalement utilisés ici proviennent d'échantillons de cancers coliques, une dernière partie abordera certains aspects histologiques et moléculaires de la barrière colique et de sa tumorogénèse.

Le **Chapitre I** abordera la mise au point d'un modèle mathématique de modélisation des interférences spectrales permettant d'étudier spectralement les coupes tissulaires incluses dans la paraffine, principale source d'échantillons dans le cadre d'approches rétrospectives.

Le **Chapitre II** présentera un modèle original d'histologie IR de xénogreffes de cancers coliques, apportant des données sur l'hétérogénéité du tissu tumoral non visible par l'histologie classique.

Dans le **Chapitre III**, les travaux présentés porteront sur la caractérisation de la signature spectrale IR d'échantillons de mucus et de mucines, étape préliminaire à leur analyse IR *in situ*.

Enfin, le **Chapitre IV** présentera une étude du potentiel de l'analyse IR *in situ* de mucus colique pour la discrimination d'échantillons de colon sains et tumoraux.

Contexte scientifique et technique

<u>Anatomopathologie et caractérisation des lésions</u> <u>tumorales</u>

L'anatomopathologie est une discipline médicale qui étudie les lésions d'organes, tissus ou cellules, engendrées par des pathologies ou associées à celles-ci en utilisant des méthodes principalement basées sur l'analyse morphologique macroscopique et microscopique de prélèvements tissulaires (Stevens 1997 ; Cabanne 1986). Une lésion est définie comme étant une altération morphologique d'un organe, décelable par tout moyen d'observation. Le terme de « lésion tumorale », ou tumeur, désigne une prolifération cellulaire excessive aboutissant à une masse tissulaire ressemblant plus ou moins au tissu normal homologue adulte ou embryonnaire, ayant tendance à persister et à croître, témoignant de son autonomie biologique.

Les tumeurs dites bénignes se développent localement et restent cantonnées au tissu dans lequel elles ont pris naissance. Leur croissance est lente mais elles peuvent toutefois atteindre un volume et un poids important. Elles ne récidivent pas après ablation chirurgicale, à condition que l'exérèse soit complète. Ces tumeurs ne métastasent jamais et leur évolution est généralement favorable. Les caractères des tumeurs dites malignes, ou cancers, s'opposent point par point à ceux des tumeurs bénignes. Les tumeurs malignes ont habituellement une croissance rapide et donnent naissance à une dissémination tumorale à distance, principalement par voie lymphatique et sanguine, avec éclosion et développement de tumeurs secondaires dans d'autres organes : les métastases. Les tumeurs malignes ont tendance à récidiver après éradication locale. Leur évolution, en dehors de tout traitement, se fait spontanément vers la mort.

L'histoire naturelle d'un cancer peut être divisée schématiquement en plusieurs étapes (fig. 1) :

- la transformation tumorale d'une cellule et son expansion clonale (dysplasie).
- la croissance de la masse tumorale qui devient cliniquement détectable.
- l'invasion locale avec envahissement locorégional par le tissu cancéreux.
- la dissémination des cellules cancéreuses à distance du foyer tumoral initial et la formation de foyers tumoraux secondaires, les métastases.





Cette progression tumorale serait liée à l'instabilité génétique des cellules cancéreuses. Des modifications génétiques spontanées vont survenir progressivement entraînant l'apparition de variants du clone initial ainsi qu'un changement progressif de la morphologie cellulaire et tissulaire de la zone atteinte. Ces clones variants auront des comportements prolifératifs, invasifs, antigéniques, et métastatiques différents, ou encore une sensibilité inégale à la chimiothérapie, rendant le tissu tumoral particulièrement hétérogène (Konishi 2005 ; Owoniko 2002 ; Tremmel 2005). Les variations de la morphologie cellulaire et tissulaire survenant pendant le développement d'une lésion tumorale, qu'elle soit bénigne ou maligne, sont à la base de l'établissement du diagnostic de ces lésions, du pronostique de leur évolution ainsi que du mode de prise en charge thérapeutique des patients. La démarche de l'anatomie pathologique est donc basée sur une analyse sémiologique comparative des tissus normaux et des tissus pathologiques. Les lésions sont confrontées aux données cliniques, biologiques et d'imagerie : c'est la corrélation anatomo-clinique qui est indispensable pour permettre une interprétation synthétique qui aboutit à un diagnostic certain, probable ou incertain. Ainsi, les pathologistes sont des acteurs indispensables et indiscutables du diagnostic et de la qualité des soins en cancérologie (Stevens 1997 ; Cabanne 1986 ; Grünfeld 2009).

I. Préparation des échantillons tissulaires

L'analyse anatomopathologique de lésions tumorales requièrt le prélèvement d'échantillons tissulaires pouvant être effectué selon trois modalités : la biopsie, l'exérèse d'organes ou l'autopsie.

La technique standard de préparation des prélèvements tissulaires comporte plusieurs étapes : la fixation, l'inclusion en paraffine, la confection de coupes et leur coloration. Avant la fixation, il est possible d'effectuer sur le tissu frais des appositions sur lames pour une étude cytopathologique, et des prélèvements pour des techniques particulières : congélation, fixation adaptée à la microscopie électronique, mise en culture pour étude cytogénétique ou en suspension cellulaire pour étude par cytométrie en flux... En ce qui concerne les pièces opératoires, une étape d'analyse macroscopique est indispensable, avant (idéalement) ou après la fixation de la pièce.

Fixation

Indispensable pour conserver la morphologie cellulaire des tissus, elle doit être immédiate ou au moins très rapidement initiée après l'obtention du prélèvement. Toute fixation défectueuse rend l'étude anatomo-pathologique difficile voire impossible. Si le fixateur le plus habituellement utilisé est le formaldéhyde, des fixateurs à base d'alcool peuvent être utilisés pour les biopsies de petite taille. La fixation est alors plus rapide mais présente un effet délétère sur certains antigènes pouvant nuire à des techniques particulières d'immunohistochimie (IHC).

Imprégnation et inclusion

Après fixation, les prélèvements sont déposés dans des cassettes en plastique, directement s'il s'agit de biopsies ou, s'il s'agit de pièces opératoires, après un examen macroscopique au cours duquel sont prélevés des fragments de petite taille (en moyenne 2x2x0.3 cm). Puis les tissus contenus dans les cassettes sont déshydratés par passage dans des solutions alcooliques de degré croissant. L'alcool est ensuite éliminé par des solvants (par exemple, le xylène) puis la paraffine liquide à 56° imprègne les tissus et est refroidie. Ces étapes sont automatisées dans des appareils

à inclusion. L'étape finale de l'inclusion est manuelle et consiste à réorienter convenablement le fragment tissulaire dans le sens de la coupe dans un moule de paraffine.

Coupes et colorations

Le bloc solide de paraffine contenant le tissu est coupé grâce à un microtome, les coupes de 3 à 5 microns d'épaisseur sont étalées sur des lames.

Après dissolution de la paraffine, puis réhydratation, le tissu est coloré (fig. 2). La coloration usuelle associe un colorant basique nucléaire (hématéine, hématoxyline) et un colorant acide cytoplasmique (éosine, érythrosine, ou phloxine). Cette coloration de type Hématoxyline-Eosine (coloration HE) peut être complétée par l'ajout de safran se fixant sur le collagène (coloration HES). La coupe colorée est protégée par une lamelle de verre ou par un film plastique transparent. Elle est analysée au microscope par un médecin anatomo-pathologiste qui établit un compte-rendu. Les blocs et les lames sont ensuite archivés.



Figure 2. Exemple de tissus colorés à l'HE (gauche, xénogreffes de carcinome du colon) et à l'HES (droite, fémur de souris).

Des colorations spéciales ont pour but de mettre en évidence des constituants particuliers des cellules (glycogène, mucus, pigments...) ou de la matrice extracellulaire (collagène, fibres élastiques, amylose...) ainsi que des agents infectieux (bactéries, parasites, champignons). Ces colorations sont très variées et leur mise en œuvre, le plus souvent manuelle, rallonge le processus technique.

II. Evaluation diagnostique des lésions tumorales

Le diagnostic des lésions tumorales s'appuie principalement sur l'étude morphologique des tissus prélevés. Cette étude se fait tout d'abord par un examen macroscopique, suivi d'un examen microscopique.

L'examen macroscopique détaillé est une partie essentielle de l'étude d'une pièce opératoire : la pièce est examinée, mesurée, pesée, palpée puis disséquée. L'examen macroscopique donne des indications pour le pronostique de la maladie (notamment la taille et la localisation de la tumeur) mais il permet surtout de sélectionner les territoires à prélever pour l'étude microscopique : zones lésées, zones d'aspect macroscopique sain et limites d'exérèse. Après le choix des prélèvements destinés à l'analyse microscopique, les restes de la pièce opératoire sont conservés pendant quelques jours ou semaines afin de pouvoir, en cas de nécessité, effectuer des prélèvements complémentaires.

L'examen microscopique des coupes histologiques colorées à l'HE constitue l'élément essentiel du diagnostic anatomopathologique. En effet, les anomalies fonctionnelles des cellules tumorales (résistance à l'apoptose, prolifération illimitée, capacité d'invasion tissulaire...) s'accompagnent de modifications morphologiques permettant le plus souvent de caractériser ces cellules anormales par observation microscopique. Parmi ces modifications, on peut citer celles du noyau (multinucléation, mitose anormale...) ou du cytoplasme (anomalie du système sécrétoires ou du cytosquelette). Les variations du stroma ou l'angiogénèse sont également des facteurs pris en compte dans le diagnostic des lésions tumorales.

Enfin, l'immunohistochimie et les colorations spéciales, permettant la mise en évidence de particularités des cellules tumorales ou du stroma, sont souvent utiles à l'affinement du diagnostic.

III. Evaluation pronostique et prédictive

Si l'examen anatomopathologique a pour objectif premier d'établir le diagnostic d'une lésion tumorale, il contribue également à l'évaluation du pronostique d'évolution de la lésion tumorale ainsi qu'à l'évaluation prédictive de sa réponse à un type donné de traitement. En effet, de cet examen microscopique de la morphologie tissulaire sera déterminé non seulement le type histologique de la tumeur mais aussi son « grade » et son « stade ». Cette classification des tumeurs en fonction de leur degré de différenciation (grade) et de leur degré d'extension (stade) peut fournir des informations importantes pour leur évaluation pronostique et prédictive.

La stadification TNM

La stadification TNM (Tumor Node Metastasis) a été initialement développée pour prédire le pronostique d'évolution des lésions tumorales, mais son rôle a été étendu à l'évaluation prédictive de la réponse thérapeutique. Le stade (ou degré d'extension) d'un cancer se fonde sur la taille de la tumeur primitive et/ou :

- Son extension aux tissus et organes de voisinage (T).
- L'importance de la dissémination aux ganglions lymphatiques régionaux (N).
- La présence ou l'absence de métastases (M).

Le système TNM est actuellement le plus communément utilisé dans le monde (Sobin 2002). Chacune de ces trois lettres est suivie d'un chiffre variant de 0 (absent) à 4, ou d'un X en cas d'impossibilité d'évaluation. Le score est précédé de la lettre c si l'évaluation du stade est clinique ou de la lettre p si elle est faite par un pathologiste (ex : adénocarcinome colique stade pT2N2MX, le pathologiste a détecté une infiltration tumorale de la musculeuse et de plus de 3 ganglions, mais ignore s'il existe des métastases à distance). Le score TNM permet ensuite de déterminer le stade (ex : adénocarcinome colique T3N0M0 = stade IIA, ou T3N1M0 = stade IIIB).

Après une résection chirurgicale, le stade d'une tumeur doit être complété par une évaluation de la qualité de la résection, qui est exprimée par la lettre R, suivie des chiffres 0 (exérèse complète, 1 (envahissement microscopique des limites) ou 2 (envahissement macroscopique des limites). Si l'évaluation du stade est faite après un traitement (ex radio- ou chimiothérapie, le score TNM est précédé de la lettre y (ex : adénocarcinome rectal réséqué après radiothérapie ypT0N1MX).

Le grade

Le grade histopronostique d'un cancer se fonde sur des critères tels le degré de différenciation tumorale, le degré d'atypies cyto-nucléaires ou l'extension de la nécrose. Ces critères varient pour chaque type de tumeur, et peuvent présenter un intérêt pronostique et prédictif.

Les marqueurs protéiques

La mise en évidence de marqueurs protéiques au sein de tumeurs peut permettre d'affiner l'évaluation pronostique et prédictive basée sur leur morphologie. Dans ce cadre, l'IHC est à l'heure actuelle la seule technique utilisée en routine. L'utilisation de combinaisons d'anticorps dont le choix est orienté par l'étude histologique permet de préciser dans la plupart des cas la nature des tumeurs peu différenciées et l'origine primitive des métastases.

L'IHC présente donc plusieurs intérêts :

- Intérêt pronostique : par exemple, la classification précise de nombreuses tumeurs par la mise en évidence d'antigènes de différenciation cellulaire (ex : cytokératine témoignant d'une différenciation épithéliale dans un carcinome) ou la mise en évidence de protéines impliqués dans la prolifération cellulaire (ex : Ki67).
- Intérêt prédictif : par exemple, la mise en évidence par immunohistochimie des récepteurs nucléaires aux œstrogènes et à la progestérone afin de déterminer le statut potentiellement hormono-sensible des cancers du sein.

IV. Les nouvelles techniques d'analyses des prélèvements tissulaires

De nouvelles techniques d'investigation histologique se sont considérablement développées ces dernières années, notamment grâce au fort engouement relatif aux domaines de la biologie moléculaire et de l'imagerie biologique. Si certaines de ces techniques peuvent être d'ores et déjà utilisées en milieu clinique (morphométrie, microscopie confocale par exemple), la plupart restent actuellement du domaine de la recherche (microdissection, analyse chromosomique *insitu*, tissue array, PCR *in-situ*, techniques d'analyse génomique et protéomique haut-débit). Pour toutes ces techniques développées et utilisées en recherche l'application en routine n'est pas encore évidente en raison de problèmes de coût, de temps ou de fiabilité.

Parmi ces techniques, celles d'analyses haut-débit dites «-omiques» (génomique, transcriptomique, protéomique) suscitent un engouement certain pour l'amélioration du diagnostic ainsi que de l'évaluation pronostique et prédictive des lésions tumorales. Ces techniques permettent d'obtenir une vue d'ensemble du matériel génétique (génomique), des transcrits (transcriptomique) ou de la composition protéique (protéomique) d'un échantillon, y

compris dans le cas de tissus inclus dans la paraffine (Gillepsie 2004 ; Kondo 2008 ; Rezaul 2009).

Ces approches ne se focalisent donc non plus sur l'identification d'un marqueur unique, mais sur une multitude de marqueurs co-exprimés, débouchant ainsi sur l'établissement d'un profil moléculaire, faisant office de véritable « carte d'identité » pour chaque tumeur. Il est admis qu'une telle signature moléculaire, associée aux données cliniques et anatomopathologiques, pourrait grandement améliorer le dépistage précoce de cancers (Adekani 2008), ainsi que leur prise en charge thérapeutique (Anderson 2006). Le profilage haut-débit de gènes (par des techniques tel que les puces ADNc ou SAGE par exemple) a permis d'acquérir de nombreuses informations sur l'expression de gènes pouvant faire office de biomarqueurs utiles dans la compréhension de processus biologiques liés aux pathologies tumorales. Mais de la complexité des processus de transcription et de traduction des gènes en protéines résultent une absence de relation simple entre la séquence génétique et le protéome. Ainsi un même gène peut produire différentes isoformes d'une même protéine via, par exemple, des mécanismes d'épissages alternatifs des transcrits ou de modifications post-traductionnelles. Par conséquent, si l'approche génomique présente un intérêt certain, c'est la protéomique qui, aujourd'hui, offre le plus fort potentiel pour une application dans le diagnostic clinique de lésions tumorales. De plus, certains protocoles d'études protéomiques, non seulement utilisées pour déterminer le niveau d'expression des protéines, mais également leurs structures, modifications et interactions, sont depuis peu applicables sur les échantillons tissulaires paraffinés et peuvent être combinées avec les techniques de microdissection. Ceci offrant l'opportunité d'associer informations moléculaires et spatiales (Rezaul 2009).

Pour illustrer le potentiel de la protéomique haut-débit dans le cadre de l'amélioration du diagnostic et de la prise en charge des pathologies tumorales, nous pouvons citer les études réalisées par spectroscopie de masse SELDI-TOF (Surfaced-Enhanced Laser Desorption/Ionization-Time Of Flight). Cette technique s'avère à l'heure actuelle la mieux placée pour une éventuelle application clinique (Ciorda 2006; Solassol 2005), notamment en raison de sa relative simplicité d'utilisation, de sa sensibilité et de sa capacité d'analyse à haut débit des échantillons (près de 800/jour). Cette approche permet la séparation, la détection et l'analyse de protéines, directement à partir de l'échantillon biologique avec une sensibilité de l'ordre de la femtomole. On obtient alors une vue d'ensemble des peptides et des protéines

Contexte scientifique et technique

présents dans un échantillon donné. L'analyse différentielle et statistique de l'ensemble des données fournies par les profils protéiques des groupes témoins (sujets sains sans cancer) et pathologiques est une des étapes les plus importantes de ces études protéomiques. Pour cela, il existe une multitude d'algorithmes informatiques fondés sur des analyses statistiques multivariées (analyse discriminante, classification hiérarchique...). Ces algorithmes permettent d'analyser des données complexes et d'extraire des spectres de masse la meilleure combinaison de marqueurs capables de discriminer chacun des deux groupes.

Les différentes études du génome, du transcriptome et du protéome dans le cadre de l'identification de nouveaux marqueurs tumoraux ont suscité un intérêt certain pour l'application des ces techniques haut-débit en clinique, avec notamment des perspectives à terme pour l'amélioration de la caractérisation des lésions tumorales et le diagnostic des formes précoces des cancers. Néanmoins, si ces études ont mis en avant l'intérêt de l'établissement d'une signature tumorale comme outil diagnostique, elles ont également soulevé de nombreuses critiques sur les limitations réelles ou supposées de ce genre d'approche. Pour reprendre l'exemple de la technique SELDI-TOF, une méta-analyse réalisée à partir des données issues de cinq publications indépendantes suggère que cette approche est peu reproductible d'un laboratoire à l'autre (Diamandis 2002). Dans cette analyse, les profils protéiques permettant la discrimination de sérums témoins et de sérums de patients atteints de cancer de la prostate s'avèrent différents d'une publication à l'autre. En fait, il est apparu que l'établissement des profils protéiques.

V. Les limites de la caractérisation des lésions tumorales en anatomo-pathologie

En pathologie tumorale, le compte-rendu anatomo-pathologique constitue l'élément de décision le plus important pour la prise en charge thérapeutique du patient. Cependant, si l'examen anatomopathologique de routine (actuellement défini par la séquence colorations standards, éventuellement colorations spéciales et IHC) est un outil indispensable à la prise en charge des patients en cancérologie, cette méthode présente de nombreuses limites.

A. La subjectivité de l'examen anatomopathologique

La reproductibilité du diagnostic anatomopathologique est imparfaite avec des taux de discordances variables suivant le type de tumeur en cause et le niveau de la ou des questions posées. Cette discordance entre pathologistes est due à la difficulté du diagnostic de certaines tumeurs, liée principalement à trois facteurs : le type de lésions tumorales, la qualité du prélèvement et l'expérience du pathologiste (Coindre 2001).

La lésion examinée

Elle peut appartenir à un domaine spécialisé de la pathologie tumorale (hématopathologie, neuropathologie, tumeurs osseuses, tumeurs des tissus mous, etc.) ou correspondre à un type tumoral dont le diagnostic repose sur des critères qui ont évolué ou dont la validité n'est pas assurée. Elle peut aussi, par exemple, correspondre à une lésion frontière ou à un cancer débutant, dont le diagnostic entraîne des attitudes thérapeutiques différentes, mais pour lesquelles il n'existe pas de consensus.

La qualité du prélèvement

Elle peut dépendre par exemple de la taille du prélèvement (la petite taille du prélèvement est une difficulté fréquente lors du diagnostic) ou de son altération morphologique (par exemple, fragment tumoral écrasé ou cautérisé lors du prélèvement).

Le pathologiste

Les difficultés rencontrées lors de l'interprétation de l'examen d'une tumeur sont en partie liées aux connaissances du pathologiste et en particulier à son expérience dans le domaine en cause, éventuellement à son degré de confiance en soi (Wright 1975). Dans certains cas, la variabilité inter- et intra-observateur du diagnostic de lésions tumorales peut être insatisfaisante (Bird 2008). Sachant que du rapport du pathologiste dépendra la prise en charge médicale du patient, l'aspect subjectif de la caractérisation des lésions tumorales représente une problématique clef pour l'amélioration de la qualité des soins en cancérologie.

B. Un diagnostic basé sur la morphologie

Comme expliqué précédemment, l'établissement d'un diagnostic en anatomopathologie revient pour la plus grande part à l'étude de la morphologie des prélèvements tissulaires. Par essence, cette approche s'avère limitative pour une prise en charge précoce des patients présentant des lésions tumorales. Il est admis que le facteur le plus important dans la prévention et le traitement des cancers est leur détection précoce, facilitant grandement l'éradication des premières cellules anormales, avant l'invasion systémique. Mais à partir du moment où une tumeur est visible et que des changements morphologiques significatifs de la malignité de la tumeur sont détectables, celle-ci est souvent à un stade avancé et de nature invasive (Kendall 2009). De plus, cette approche morphologique ne prend pas en compte la nature hétérogène des voies biomoléculaires de développement des lésions tumorales. En effet, en raison de cette hétérogénéité, certaines tumeurs histologiquement semblables peuvent présenter un pronostique d'évolution et une réponse à la prise en charge thérapeutique très variables. Or, il est nécessaire pour une pathologie d'être correctement caractérisée et classifiée, non seulement sur la base de données morphologiques mais aussi moléculaires, dans l'optique de l'amélioration de la prise en charge des patients tout comme pour l'investigation des mécanismes sous-jacents à son développement (Jass 2007).

L'immunohistochimie a été introduite dans les examens anatomopathologiques de routine dans le but de détecter des changements fonctionnels, pouvant éventuellement précéder certains changements morphologiques, faisant office de marqueurs tumoraux diagnostiques, pronostiques ou prédictifs. Cependant, parce que cette technique ne peut cibler qu'un nombre limité de protéines, l'immunohistochimie n'a jamais abouti à l'obtention de marqueurs tumoraux définitifs, et n'est utilisée que pour affiner ou orienter les conclusions tirées de l'examen morphologique.

Le développement des techniques haut-débit de génomiques et protéomiques, applicables sur coupes tissulaires (Rezaul 2008), laisse entrevoir la possibilité d'exploiter de nouveaux types de marqueurs, qualifiés de « profils » tumoraux, par la détection d'un très grand nombre de protéines, ARNm ou gènes. Ce type de marqueurs pourrait donc prendre en compte l'hétérogénéité des voies biomoléculaires du développement tumoral permettant ainsi une amélioration de la prise en charge des patients. Cependant, à l'heure actuelle, ces techniques s'avèrent fort complexes et encore non applicables en routine. Elles présentent également un niveau de reproductibilité inter-laboratoire non satisfaisant (Diamandis 2002).

C. Des techniques chronophages et non automatisables

Les différents processus appliqués à un prélèvement tissulaire, tel que la fixation, l'inclusion, la découpe, les colorations et le(s) marquage(s), ainsi que l'analyse morphologique elle-même sont associés à l'utilisation de moyens techniques et humains pouvant engendrer un délai de plusieurs jours entre la date du prélèvement et le rapport du pathologiste. En effet, le délai de réponse nécessaire, en raison de ces diverses contraintes techniques, est généralement de l'ordre de 48h au minimum. En cas de délai prolongé (examen en attente de techniques complémentaires ou demande d'avis auprès d'un expert), un compte-rendu provisoire peut être adressé, mais une décision thérapeutique ne peut s'appuyer que sur le compte-rendu définitif. Ces délais sont, à l'heure actuelle, difficilement compressibles en raison notamment de la grande difficulté d'automatisation de la plupart des taches, dont l'analyse morphologique sur coupes colorées, visant à la caractérisation d'un prélèvement tissulaire.

Les limites de l'histopathologie ont amené à investiguer de nouveaux moyens de diagnostic et de caractérisation tissulaire applicable en milieu clinique. Si dans ce cadre, les techniques biomoléculaires haut-débit s'avèrent prometteuses, leur complexité et leur coût ainsi que la reproductibilité des résultats obtenus ne sont pas encore compatibles avec une application clinique. Parmi les autres techniques étudiées dans le cadre de la caractérisation pathologique de

prélèvements biologiques, la microspectroscopie vibrationnelle d'absorption Infrarouge (IR), reposant sur l'interaction non destructive entre un rayonnement électromagnétique et la matière a connu un fort développement au cours des 15 derniers années (Diem 2004 ; Ellis 2006 ; Krafft 2009). Cette technique, rapide, non subjective, relativement simple d'utilisation et ne nécessitant pas de marquages, permet d'obtenir des signaux hautement spécifiques de la composition biomoléculaire intrinsèque de l'échantillon étudié. Elle peut en conséquence révéler de subtiles variations moléculaires au sein de cellules et de tissus indépendamment de leur morphologie, particulièrement durant le processus de carcinogénèse.
La microspectroscopie vibrationnelle IR

I. Principe de la Microspectroscopie IR

La région spectrale IR est adjacente à celle du visible et s'étend de 0,78 μ m à environ 1000 μ m. Cette région peut être subdivisée en 3 domaines distincts : le « proche-IR » (de 780 nm à 2,5 μ m), le « moyen-IR » (de 2,5 μ m à 25 μ m), le « lointain-IR » (de 25 μ m à 1000 μ m).

Le domaine moyen-IR, situé entre 2,5 et 25 μ m (4000 cm⁻¹ et 400 cm⁻¹), correspond au domaine d'énergie de vibration des liaisons de la majorité des molécules aussi bien organiques qu'inorganiques. Ainsi, si les régions du proche-IR et du lointain-IR peuvent susciter un intérêt dans certain domaine (imagerie cérébrale, analyses pharmacologiques, astronomie...), l'utilisation de la spectroscopie moyen-IR reste la plus adaptée pour l'analyse de la composition moléculaire d'un échantillon. Dans la suite de se manuscrit, il sera uniquement fait référence à la spectroscopie moyen-IR, sauf indication contraire.

Dans les phénomènes d'émission de fluorescence et phosphorescence, l'énergie du rayonnement, monochromatique, provoque une transition électronique. Le retour à l'état fondamental engendre alors une émission de lumière. Dans le cas de la spectroscopie IR, basée sur l'interaction de la lumière IR avec le nuage électronique des liaisons chimiques, le rayonnement polychromatique émis par la source n'est pas assez énergétique pour provoquer des transitions électroniques mais il induit des transitions entre les niveaux d'énergie vibrationnelle. Le concept quantique des niveaux d'énergies vibrationnelles et électroniques est schématisé par le diagramme de Jablonski (fig. 3).



Pour qu'il y ait absorption dans l'IR, il faut non seulement que l'énergie du rayon corresponde à l'énergie de vibration de la liaison, mais également que la vibration engendrée induise une variation du moment dipolaire de la liaison moléculaire.

Une molécule diatomique ne présentera qu'un seul mode de vibration qui est l'élongation de la liaison. Le modèle le plus simple de cette élongation est un ressort qui s'apparente à une oscillation harmonique. Dans le cas des molécules polyatomiques, le nombre de liaisons augmente et leur géométrie se complexifie. Si on considère n atomes isolés, chacun possède 3 degrés de liberté de translation et 3 degré de liberté de rotation dans un espace à 3 dimensions. Lorsque ces atomes sont reliés entre eux par des liaisons d'angle et de longueur variables, la molécule possède 3n degrés de liberté dont 6 peuvent être attribués à la translation et la rotation de la molécule entière. Il reste donc 3n-6 degrés de liberté pour les modes de vibrations des liaisons. Ce nombre est ramené à 3n-5 pour les molécules linéaires du fait de la rotation selon l'axe de la molécule.

Les fréquences de vibration dépendent des masses des atomes et des forces de liaisons. Deux groupes de vibrations se distinguent (fig. 4) : i) les vibrations de valence ou élongation (y), symétriques ou asymétriques, qui présentent des modifications des longueurs de liaisons sans

changement des angles ; ii) les vibrations de déformation (δ), qui inversement, présentent des liaisons de même longueur mais une variation des angles intramoléculaires.



Figure 4. Exemple de modes de vibration moléculaire d'élongation (gauche) et déformation (droite).

De nombreux autres modes de vibration existent pour les molécules complexes (cf. Annexe 1, p167). Chacun des ces modes de vibration se comporte comme un oscillateur harmonique. A chaque mode de vibration normal est associé un nombre d'onde ou fréquence qui est exprimée en cm⁻¹. Puisque les fréquences vibrationnelles et la probabilité d'absorption de l'IR dépend de la force et de la polarité des liaisons chimiques, elles sont directement influencées par les interactions inter- et intra-moléculaires. La position, en nombre d'onde, d'une bande d'absorption IR est déterminée par les atomes constituant la liaison, le type de liaison (simple, double, triple), les forces inter- et/ou intra-moléculaires s'exerçant sur la liaison, et le couplage éventuelle à d'autres vibrations.

La spectroscopie IR permet donc d'obtenir des spectres qui correspondent à un ensemble de bandes d'absorption mettant en évidence les vibrations caractéristiques des groupements chimiques d'un échantillon (fig. 5). Un spectre peut donc être considéré comme l'empreinte chimique spécifique d'un échantillon. Ces spectres IR présentent l'absorbance A (ou la transmittance T) en fonction du nombre d'onde v ou σ avec :

$$\sigma = 1 / \lambda$$
 (cm) en cm⁻¹



L'ensemble des liaisons polaires contribuent à l'absorption IR, ceci constituant à la fois la faiblesse et la force de la spectroscopie IR : i) une faiblesse, parce que le spectre de molécules complexes est alors composé de multiples bandes se superposant, ceci ayant pour conséquence de masquer des informations potentiellement pertinentes au profit de bandes d'absorption peu caractéristiques ; ii) une force, parce que quasiment toutes les biomolécules présentent alors une absorption dans l'IR. Par conséquent, en spectroscopie IR, toutes les biomolécules sont potentiellement détectables sans nécessiter de marquage extrinsèque.

Les informations obtenues à partir des spectres IR sont de deux natures :

• Informations qualitatives :

Un spectre d'absorption IR, y compris celui provenant d'une molécule simple, comporte un certain nombre de pics ; chacun de ces pics correspondant à la vibration caractéristique des groupements chimiques de l'échantillon analysé.

• Informations quantitatives :

L'intensité de l'absorption IR à un nombre d'onde donnée est reliée à la concentration du groupement chimique à l'origine de cette absorption. Cependant, la linéarité de cette

relation absorbance-concentration n'est vérifiée que dans un domaine limité d'absorbance, situé en général entre 0,3 et 2.

La spectroscopie était à l'origine fortement limitée par la lenteur d'acquisition des spectres, les mesures étant réalisées successivement pour chaque longueur d'onde par un monochromateur dispersif (prisme ou réseau). Les spectromètres modernes sont en règle générale des spectromètres IR à Transformée de Fourier (IRTF). Ce nom à pour origine la fonction mathématique appliquée au signal du détecteur et permettant l'obtention d'un spectre IR. Le cœur d'un spectromètre IRTF est l'interféromètre, autorisant le multiplexage des acquisitions et permettant ainsi de mesurer toutes les longueurs d'onde du spectre de l'IR simultanément. Les principaux avantages de la spectroscopie IRTF sont sa vitesse de collection des données ainsi qu'un haut rapport signal sur bruit (S/B).

Le couplage d'un spectromètre IRTF avec un microscope utilisant des optiques purement réfléchissantes de type Cassegrain permet l'obtention d'un microspectromètre autorisant l'analyse IR d'échantillons avec une résolution spatiale de l'ordre de 10 μ m. Un microspectromètre IRTF comporte essentiellement 6 parties (fig. 6) :

Une source lumineuse

Dans le cas du moyen-IR, la source utilisée est le Globar, constituée d'un bâtonnet en carbure de silicium (SiC) fonctionnant à des températures de l'ordre de 1500°C. Le Globar rayonne un spectre continu entre 250 et 10000 cm⁻¹.

L'interféromètre

C'est un dispositif permettant de générer des interférences et autorisant le multiplexage des acquisitions, c'est-à-dire l'analyse de plusieurs nombres d'onde simultanément. Le plus utilisé est l'interféromètre dit de « Michelson » : il est composé de deux miroirs, un fixe et l'autre mobile, et d'une lame séparatrice. Le faisceau IR entrant est alors divisé en deux et le décalage de marche optique créé par le mouvement du miroir engendre des variations d'intensité arrivant sur le détecteur qui enregistre alors l'interférogramme. Associée à ce dispositif, une formule mathématique permet de représenter le spectre sous la forme de l'absorbance en fonction de la longueur d'onde : la Transformée de Fourier.

Un compartiment « échantillon »

Il permet d'accueillir plusieurs types d'accessoires (porte-échantillon) dépendant de la technique spectroscopique employée et du mode de mesure utilisé (cf. Annexe 2, p168).

Un détecteur ou un capteur photosensible

Un microspectromètre IRTF peut comporter un ou plusieurs détecteurs. Ils peuvent être de type pyroélectrique (générant un courant proportionnel au différentiel de température entre les 2 surfaces du détecteur) tel que les détecteurs DTGS (Deuterated Tri-glycine Sulfate), ou photoélectrique (générant une différence de potentiel par l'absorption de photons) tel que les détecteurs MCT (Mercure Cadmium Tellure), constitués d'un monocristal en alliage de mercure cadmium tellure déposé sur un support inerte. Pour une plus grande sensibilité, les détecteurs de type MCT sont refroidis à l'azote liquide.

Un convertisseur analogique-numérique

Il permet de transformer le signal analogique en signal numérique manipulable par un système informatique.

Un microscope

Le rôle du microscope est de conduire le rayon infrarouge de l'interféromètre à l'échantillon de manière colinéaire au chemin optique visible. L'échantillon mesuré correspond alors à l'échantillon observé. Les lentilles en verre étant opaque dans l'IR, un microscope IR est donc doté d'optiques réfléchissantes (miroirs dorés ou argentés, objectifs de type Cassegrain). La plupart des objectifs utilisés sur ce type d'instrumentation possèdent des ouvertures numériques comprises ente 0.5 et 0.7 donnant des résolutions spatiales allant de quelques microns à plusieurs dizaines de microns suivant l'ouverture utilisée. La résolution spatiale peut être définie comme la capacité à mesurer le spectre d'un objet sans contamination par des informations provenant d'objets placés hors du champ délimité par l'ouverture (Messerschmidt, 1988).

Dans tous les microspectromètres IRTF, le faisceau provenant de la source est dirigé vers l'interféromètre qui va moduler chaque longueur d'onde du faisceau à une fréquence différente. Le faisceau ainsi modulé est réfléchi au travers du microscope vers l'échantillon, où

interviennent des absorptions fonctions de sa composition chimique. Le faisceau arrive ensuite sur le détecteur pour être transformé en signal électrique qui sera par la suite converti en spectre IR.



Parce que la vapeur d'eau et le CO_2 atmosphérique entrainent une absorption significative non désirable en IR, un microspectromètre IRTF peut se voir équipé d'un module de purge. Cet équipement consiste en une boite scellée, le plus souvent composée de plexiglass, entourant le compartiment de l'échantillon ainsi que le système optique, et purgé par de l'air sec ou du N₂.





Les mesures spectrales se dérouleront alors dans un environnement limitant les effets indésirables de la vapeur d'eau et du CO2 atmosphérique (fig. 7), tout en permettant de maintenir une atmosphère constante autorisant une comparaison plus pertinente entre différents échantillons dont les données spectrales ont été acquises sur un large laps de temps (de plusieurs jours à plusieurs semaines).

La première étape nécessaire à l'obtention d'un spectre IR caractéristique d'un échantillon est la mesure du spectre dit de référence (« ou bruit de fond », fig. 8), avant de passer à la mesure de l'échantillon proprement dite. Ce spectre de référence va contenir toutes les interventions extérieures à l'échantillon mesuré et sera soustrait à tous les spectres. Une grande partie du signal de référence provient de la vapeur d'eau et du CO₂ atmosphérique. Ces contributions peuvent être minimisées un système de purge, tel que décrit précédemment.



Dans un premier temps un interférogramme de référence (fig. 9*A*) est calculé et le spectre correspondant est obtenu par Transformée de Fourier (fig. 9*C*). Dans un deuxième temps un interférogramme est également mesuré sur l'échantillon (fig. 9*B*) et converti en signal (fig. 9*D*). Dans l'étape finale, une opération mathématique simple permet de faire la différence entre les deux signaux et d'obtenir des spectres correspondant aux caractéristiques chimiques propres de l'échantillon (fig. 9*D*).



Figure 9. Obtention du spectre d'un échantillon en spectroscopie IRTF. *A*, interférogramme de référence. *B*, interférogramme échantillon. *C*, spectre simple faisceau de référence. *D*, spectre simple faisceau échantillon. *E*, spectre résultant exprimé en absorbance.

Rq : *Si les spectres sont générés initialement en unité de transmittance, l'affichage du spectre final se fait généralement en absorbance pour en faciliter la visualisation.*

Les spectres IR sont très généralement affectés par les propriétés physiques des échantillons tels que la taille et la distribution des particules qui le compose (Bertrand 2000). De plus, il faut noter

que la soustraction du spectre de référence à ceux provenant de l'échantillon ne suffit pas dans tous les cas à pleinement corriger le bruit de fond issu de l'environnement de l'échantillon. Des pré-traitements sont donc nécessaires pour l'optimisation du signal et se révèlent indispensables pour une interprétation pertinente des données. Les pré-traitements standards utilisés en spectroscopie IR sont la correction de la ligne de base, le lissage, la normalisation et la dérivation.

Correction de la ligne de base

Lors de l'interaction entre la lumière et le l'échantillon étudié, des phénomènes s'associent à l'absorption des radiations IR : diffusion, aberrations chromatiques et diffraction sont observées. Une fraction significative de la lumière est alors déviée de son trajet et n'atteint pas le détecteur. Ces phénomènes sont variables en fonction de la longueur d'onde et de l'épaisseur de l'échantillon. Ces interactions se manifestent par une déformation du spectre qui s'incline de façon prononcée pour les grands nombres d'onde.

L'opération de « correction de la ligne de base » permet d'éliminer l'influence négative de ces phénomènes parasites sur le potentiel informatif des spectres. Cette correction repose sur la détermination d'un certains nombre de points, définissant le nombre de minima du spectre, et du degré du polynôme utilisé. La procédure implique la définition d'une ligne de base passant par n minimima, n étant fixé par l'opérateur. L'allure de la ligne de base passant par ces points dépendra, quant à elle, du degré du polynôme choisi pour son calcul. La ligne de base ainsi définie est alors soustraite point par point au spectre acquis. Cette procédure est toutefois potentiellement génératrice d'artefacts et doit donc être strictement contrôlée par l'utilisateur.

Le lissage

Cette fonction vise à diminuer le bruit surajouté aux données spectrales spécifiques de la composition biochimique l'échantillon. Elle présente cependant deux aspects contradictoires : d'un coté le lissage tend à éliminer les signaux parasites non informatifs (ou le « bruit de fond ») enregistrés par le détecteur ; d'autre part ce pré-traitement affecte toujours les informations du spectre IRTF, essentiellement quand leur amplitude est faible. Le lissage a un effet équivalent à une diminution de la résolution spectrale (voir p. 44).

La normalisation

L'intensité de l'absorption IR peut varier selon l'épaisseur et la topologie des différentes régions composant l'échantillon, modifiant la quantité de matière interagissant avec le rayonnement et donc le signal résultant. Afin de permettre la comparaison de tous les spectres provenant d'un échantillon, il s'avère donc nécessaire de compenser ces variations d'absorption.

La procédure dites de « normalisation » consiste à ramener les spectres à la même intensité (i.e. la même échelle), afin de pouvoir les comparer aussi bien sur le plan qualitatif que quantitatif. Si cette opération mathématique permet de rendre comparable des spectres entre eux, elle les ramène tous à une échelle arbitraire éliminant la composante quantitative absolue contenue dans les informations spectrales. En revanche, les intensités relatives inter- et intra-spectrales (rapport de bandes ou de pics) peuvent toujours être comparées après normalisation. Cette opération peut être réalisée sur l'intensité d'un pic (normalisation min-max) ou sur l'aire d'une portion ou de la totalité du spectre (normalisation vectorielle). Dans ce cas, l'aire de chaque spectre est calculée dans un intervalle défini par l'utilisateur et tous les spectres sont ramenés à la même aire.

Les dérivées de premier et second ordre

La principale utilité de la dérivation, qu'elle soit de premier ou de second ordre, est de rendre certaines informations spectrales plus distinctes en augmentant virtuellement la résolution spectrale.

La dérivé première détermine la pente en chaque point du spectre et renseigne donc sur la façon dont varie l'absorbance en fonction du nombre d'onde, en s'affranchissant des problèmes liées à la variation de la ligne de base. La dérivée seconde, quant à elle, renseigne sur la variabilité de la pente du spectre et correspond en fait à son rayon de courbure. Elle permet également de s'affranchir des problèmes de lignes de bases. En revanche, dans les deux cas, l'information en termes d'absorbance pure est alors perdue, au profit de la mise en lumière des pics composants les bandes peu résolues des spectres.

Cette méthode est particulièrement utilisée dans l'évaluation de la structure secondaire des protéines d'échantillons biologiques, en révélant de subtiles variations des bandes Amides I et II relatives aux différentes conformations de type hélice α , feuillet- et tournant- β , ou « random coil » (aussi appelé structure désordonnée ou pelote statique non périodique).

II. Microspectroscopie IRTF et macromolécules biologiques

La microspectroscopie IRTF est un outil puissant pour l'analyse des biomolécules (Jackson 1996), permettant de sonder la composition et/ou la structure des protéines, carbohydrates, acides nucléiques et lipides.

A. Protéines

La microspectroscopie IRTF est une des méthodes classique pour l'analyse structurale des protéines (Barth 2007). Elle permet notamment d'étudier leur structure primaire et secondaire grâce aux nombreux modes de vibration de la liaison peptidique. En effet, les liaisons peptidiques possèdent des modes de vibration engendrant 9 larges bandes d'absorption IR (Amide I à Amide VII ainsi que Amide A et B).

Nomber deader (amf) Attribution			
Nombre d'Ordes (cm·)			
3250 - 3300	N-H) avec amide II) بر Arnide A résonance (N-H) avec amide II		
3090 - 3100	N-H) avec amide II) بر Amide B résonance) (N-H)		
1600 - 1700	C=O) couplé avec ע (C-N) et δ (N-H) (C=O) و C=N) et δ		
1500 - 1600	Amide II δ (N-H) couplé avec μ (C-N)		
1230 - 1330	C-N)ر Amide III δ (N-H) couplé avec)		
625 - 767	Amide IV 5 (O=C-N) et autres modes		
640 - 800	Amide V δ (O=C-N) et autres modes		
537 — 606	Amide VI ō (N-H) hors du plan		
200	Amide VII Torsion du squelette		

Tableau 1. Principales bandes d'absorption IR de la liaison peptidique.y Vibration d'élongation. δ Vibration de déformation.

Les bandes Amide I et II apparaissent comme les plus informatives pour l'étude des protéines en spectroscopie IR : l'analyse de leur position et de leurs composantes sous-jacentes (grâce à des techniques telles que la décomposition en vibrations élémentaire par dérivation) permet d'obtenir des informations relatives à la structure secondaire des protéines de l'échantillon étudié, notamment sur le pourcentage des acides aminés impliqués dans les différents types de structures secondaires.

Amide I (1600-1700 cm⁻¹)

La vibration de la bande Amide I provient essentiellement de la vibration d'élongation du groupement C=O. La bande Amide I est très peu affectée par la nature des chaines latérales des acides aminés, et dépend par conséquent de la structure secondaire des protéines de l'échantillon étudié. En effet, les fréquences des composantes de cette bande sont étroitement corrélées aux différents éléments structuraux secondaires que sont les hélices α , feuillet- et coude- β ou « random coil ». Ainsi, la bande Amide I est la plus couramment utilisée pour l'étude de la structure secondaire des protéines.

Amide II (1500-1600 cm⁻¹)

La bande Amide II provient essentiellement de la vibration de déformation du groupement NH et de la vibration d'élongation du groupement CN. Tout comme la bande Amide I, la bande Amide II n'est que peu affectée par la nature des chaines latérales des acides aminés mais la corrélation entre la structure secondaire des protéines et la fréquence de ses composantes est moins directe. Néanmoins, la bande Amide II fournit des informations structurales valables permettant l'évaluation de la structure secondaire d'échantillons protéiques.

Structure	Fréquence (cm-1)		
secondaire	Amide i	Amide II	
Hélice a	1648-1667	1516-1540	
Feuillet β	1623-1641	1530-1550	
	1674-16 9 5		
Tournant β	1662-1686		
Désordonnée	1642-1667	1535	

Tableau 2. Exemples de composantes des bandes d'absorption IR Amide I et II.

Autres bandes Amides

Les autres bandes Amides apparaissent comme très complexes, dépendantes de nombreux facteurs tels que la nature des chaines latérales des acides aminés ou des liaisons hydrogènes et n'ont, par conséquent, que peu d'intérêts pratiques dans l'étude conformationelle des protéines.

Outre cet aspect d'analyse de la structure secondaire des protéines, la spectrométrie IR peut aussi renseigner sur la présence de certains acides aminés impliqués dans leur structure primaire. Cependant, il s'avère difficile d'accéder à ce type d'information lors de l'analyse d'échantillons complexes, comme les cellules ou les tissus, où le recouvrement des bandes spectrales ne peut permettre la discrimination des différents acides aminés.

B. Carbohydrates

La spectroscopie IR est une technique permettant l'étude des carbohydrates en raison de leur absorption spécifique dans le domaine 1000-1580 cm⁻¹. Les marqueurs IR des carbohydrates les plus utilisés se situent néanmoins dans la gamme 1000-1200 cm⁻¹ (fig. 10) (Khajehpour 2006). La glycosylation des protéines est une des modifications post-transcriptionnelles les plus répandue : la plupart des protéines extracellulaires ainsi que certaines protéines intracellulaires

répandue : la plupart des protéines extracellulaires ainsi que certaines protéines intracellulaires possèdent des résidus glycosylés. De par l'absorbance caractéristique des carbohydrates dans la gamme 1000-1200 cm⁻¹, la spectroscopie IR est capable de mettre en évidence la présence de sucres sur des protéines, mais se révèle limitée quant à leur identification. La comparaison de l'empreinte spectrale spécifique des sucres (1000-1200 cm⁻¹) d'une protéine glycosylée avec le spectre individuel des différents sucres constituant ses glycannes permet d'obtenir des informations intéressantes notamment sur le pourcentage de glucose et de galactose (pour le collagène), la présence de liaisons faisant intervenir des mannoses (dans le cas de l'avidine), ou la présence de groupements carboxyliques d'acides sialiques (dans le cas des mucines). En revanche, il est important de noter que le spectre IR d'un sucre oligomèrique n'est pas la simple somme des spectres des monomères le constituant (fig. 10). Ceci complique sensiblement l'étude des résidus carbohydrates des glycoprotéines car, dans la majorité des cas, ces résidus sont constitués de plusieurs monosaccharides différents liés par des liaisons glycosydiques.



La spectroscopie IR est donc un outil adéquat pour estimer grossièrement, de manière quantitative et qualitative, le contenu en carbohydrates d'une protéine. Ceci peut servir de données préliminaires pour des études plus complexes et chronophages visant à déterminer la glycosylation de manière précise par des techniques telles que la spectroscopie de masse. Ces données préliminaires peuvent également aider à choisir les conditions analytiques optimales pour ces techniques complexes.

C. Les acides nucléiques

La spectroscopie IRTF est une méthode pouvant être appliquée à l'étude de nombreuses problématiques concernant les acides nucléiques (Banyay 2003). Des études en IR ont été menées notamment pour l'évaluation des changements de conformation d'acides nucléiques ou pour déterminer la nature et la force des liaisons impliquées dans leur interaction avec un ligand. Des bandes caractéristiques des conformations A, B et Z de l'ADN peuvent aussi être observées.

Attribution	N	bre d'onde (cm	r')
	ADN A	ADN B	ADN Z
ע(C=O) carbonyl	1710	1717	1692
ע(P=O)	1240	1220	1265
ע(P=O)	1090	1085	1090
O=P=O diester	810- 820	820- 825	

Tableau 3. Exemples de bandes d'absorption IR des acides nucléiques.

D. Les lipides

Les applications de la spectroscopie IR pour l'étude des lipides portent principalement sur l'analyse de l'organisation des membranes biologiques. Plus particulièrement, ce type d'études se focalise sur l'interaction entre les phospholipides et les protéines transmembranaires, la fluidité des membranes et l'orientation des molécules les constituant en réponse à diverses conditions comme des changements de polarité ou de degré d'hydratation par exemple (Silvestro 1998). L'activité vibrationnelle des lipides en IR est effective principalement dans la région 2700-3000 cm⁻¹, avec également des vibrations attribuable aux triglycérides entre 1700 et 1800 cm⁻¹.

Attribution	Fréquence (cm ⁻¹)	Attribution	Fréquence (cm ⁻¹)
ע(C-H) de CH ₃	1950-1960	ō (C-H) de CH₂	1470-1480
y(C-H) de CH ₂	2920-2930	δ (C-H) de CH ₃	1760
ע(C-H) de CH ₃	2865-2880	ע(P=O) de PO ₂	1230
ע(C-H) de CH ₂	2940-2860	ע (C-O-C) des esters	1170-1200
ע (C=O) des esters	1730-1760	ע(P=O) de PO ₂	1080
Tableau 4. Exemp vibration و	le de bandes d'ab n d'élongation ; δ V	sorption IR des pho	ospholipides.

Cependant, en raison de leurs larges pics d'absorptions entrainant des chevauchements importants, l'attribution de ces bandes dans un système complexe tel qu'un tissu reste particulièrement délicate.

III. Microspectroscopie IRTF et échantillons biologiques

Les échantillons biologiques complexes (fluides, cellules, tissus) contiennent par définition une multitude de molécules (protéines, lipides, carbohydrates et acides nucléiques). Ces échantillons complexes présentent donc de nombreuses bandes d'absorption dans l'IR qui peuvent se chevaucher et/ou se superposer. L'analyse des spectres IR, dans ce cas, se révèle complexe, d'autant plus que l'environnement de chacune des molécules constituantes d'un échantillon peut provoquer des décalages de leurs fréquences de vibrations. Cependant, la spectroscopie IR permet dans ce cadre d'obtenir une information biochimique globale et complète, quoique grossière, d'un échantillon biologique (Jackson 1996 ; Diem 2004).

Le spectre IRTF d'un échantillon biologique (fig. 11) est généralement dominé par les bandes Amides I et II des protéines (1500 – 1700 cm⁻¹), par une bande des phosphates à 1240 cm⁻¹ et par la gamme 1000-1200 cm⁻¹, présentant une forte absorption dans le cas de la présence de carbohydrates. Malgré cette relative segmentation, cette même région peut aussi contenir des pics associés au collagène ou à l'élastine, ou bien encore des bandes phosphates (~1080 et 1121 cm⁻¹). Les bandes phosphates sont souvent associées, particulièrement dans le cas de l'étude de tissus, à la présence d'acides nucléiques. Cependant, il est à noter que la capacité de la spectroscopie IRTF à détecter l'ADN sous forme de chromatine est encore sujette à débat (Diem 2002). Enfin, la gamme 1250-1500 cm⁻¹, présentant la bande Amide III (1250-1350 cm⁻¹) est une région complexe difficilement interprétable. De par le nombre important d'informations vibrationnelles contenue dans la zone 900-1500 cm⁻¹, cette gamme spectrale est aussi nommée « empreinte digitale ».

Cette segmentation grossière est valable pour la plupart des tissus mais doit être modulée en fonction de leur type. Par exemple, un tissus riche en mucines mais pauvre en lipides pourra voir une absorption dans le segment attribué généralement aux esters de lipides mais qui correspondra alors au groupement carboxyle des acides sialiques ; un tissu très riche en élastine ou collagène tel que la peau pourra présenter des absorptions issu de ces deux molécules dans la gamme 1000-1200 cm⁻¹).



Rq : Les échantillons biologiques contiennent un fort pourcentage d'eau, ceci apparaissant très limitant dans le cadre d'une étude en IR puisque les bandes d'absorptions de l'eau couvrent de larges domaines de la gamme spectrale de l'IR. Pour résoudre ce problème, les spectres IRTF sont généralement enregistrés directement sur des échantillons déshydratés selon un protocole rigoureux et spécifique à chaque échantillon. Les mesures en ATR (Réflexion Totale Atténuée, cf. Annexe 2, p170) permettent également de minimiser efficacement les problèmes liés à la présence d'eau, permettant ainsi l'analyse des liquides et des semi-solides par exemple.

IV. Imagerie par microspectroscopie IRTF

Initialement, les mesures en microspectroscopie IRTF étaient réalisées en mode ponctuel, spectres par spectres. Ce mode d'acquisition était particulièrement limitant dans le cadre de l'analyse d'échantillons fortement hétérogènes comme les tissus biologiques, notamment en raison du temps nécessaire pour obtenir une population de spectres représentatifs de ces

échantillons. L'apparition des détecteurs de type FPA (Focal Plane Array) a permis de palier à cette limitation en autorisant l'enregistrement de cartographies spectrales par imagerie IR. Les détecteurs FPA sont formés d'un nombre important de pixels de détection. Par exemple, un détecteur de 64x64 pixels permet l'acquisition simultanée de 4096 spectres, aboutissant ainsi à un gain de temps très important par rapport au mode de mesure ponctuel. Les détecteurs FPA permettent la mesure dans une gamme allant de 10000 à 500 cm⁻¹ et certains d'entre eux peuvent compter plus d'un million de pixels (1024x1024).

Un spectro-imageur IRTF moderne est donc constitué d'un microspectromètre IRTF équipé d'un détecteur FPA et d'une platine porte-échantillon motorisée de précision sub-micrométrique. L'acquisition d'une image spectrale IR peut être décrite simplement par un screening « bloc par bloc » d'un échantillon, dans notre cas une coupe tissulaire. Chaque « bloc » correspondant à un nombre de spectres déterminé par le nombre de pixels du détecteur FPA.



Une image spectrale (ou carte spectrale) exprime donc une information multiparamétrique de la composition chimique de l'échantillon associée à une information spatiale (fig. 12). L'ensemble de données correspond alors à une information sur l'absorbance IR de l'échantillon, pour chacune des longueurs d'ondes étudiées, à laquelle sont associées les coordonnées spatiale 2D

(x,y) de chacun des spectres. Il est convenu de qualifier cet ensemble multidimensionnel de données d'hypercube spectral.

Ces données peuvent être exploitées selon deux approches :

- Une approche univariée : cette méthode d'analyse ne tient compte que d'une seule fréquence des spectres ou, au mieux, d'un rapport d'absorbance entre deux bandes particulières. Cette méthode fait néanmoins preuve d'un pauvre usage de l'information contenue dans les spectres IRTF.
- Une approche multivariée : le développement de nouvelles méthodes de traitements des données dites « multivariées », prenant en compte l'ensemble des fréquences des spectres, permettent aujourd'hui la pleine exploitation de l'information contenue dans un ensemble de données spectrales (Wang 2008). Cette approche a aboutit au concept d'histologie IR, qui sera développé dans une section à part entière (« L'histologie spectrale IR », p41).

V. Les applications de la microspectroscopie IRTF dans les domaines de la biologie et du biomédical

La microspectroscopie IRTF permet d'apporter des informations sur la composition biochimique globale (acides nucléiques, protéines, carbohydrates et lipides) d'un échantillon. En conséquence cette méthode possède un potentiel intéressant pour la mise en évidence de variations qualitatives et semi-quantitatives des différents constituants moléculaires d'échantillons biologiques pouvant caractériser certaines pathologies. Les applications de la microspectroscopie IRTF en biologie et dans le domaine biomédical peuvent être catégorisées en 4 groupes : biochimie clinique (analyse des fluides biologiques), microbiologie, cytologie et caractérisation tissulaire.

Biochimie clinique

La biochimie clinique basée sur la microspectroscopie-IR permet essentiellement l'analyse des constituants des fluides biologiques tels que le sang, le sérum, la salive, les urines ou le liquide amniotique. Ces analyses sont généralement conduites dans le but de l'évaluation de la concentration de certaines molécules dans ces fluides (Petibois 2001; Low Ying 2002).

L'objectif est alors de diagnostiquer certaines pathologies associées à une concentration anormale de certains constituants par la voie d'un « dosage » spectrométrique. Dans la majorité de ces études, ces analyses sont réalisées à partir de film déshydratés, afin de s'affranchir des interférences spectrales dues à l'absorption de l'eau dans le domaine de l'IR.

Microbiologie

En microbiologie, la spectroscopie IR peut être utilisée comme outil performant pour :

- la classification des microorganismes (Karadenizli 2007).
- la différenciation de souches isogéniques (Bouhedja 1997).
- la différenciation de souches génétiquement modifiées (Sockalingum 1997).
- la détection de souches particulières sur des produits alimentaires (Irudayaraj 2002).
- une identification rapide et efficace de sous-espèces bactériennes (Naumann 2000).

Cytologie

Au niveau cellulaire, la microspectroscopie IR a été appliquée avec succès dans divers types d'études :

- Suivi de la production d'exoproduits et de changements fonctionnels en réponse à des variations environnementales (Naumann 1996 ; Meade 2007).
- Investigation des interactions cellule/médicament (Draux 2009).
- Etude de la différenciation de cellules embryonnaire (Ami 2008).
- Etude de la prolifération cellulaire (Boydston-White 2006).
- Caractérisation de cellules malignes (Ramesh 2001 ; Lasch 2002 ; Romeo 2006).
- Détection du phénotype MDR (Multi Drug Resistance) (Le Gal 1993 ; Gaigneaux 2002 ; Krishna 2006).

Ces différentes études servent généralement à mettre au point des modèles en vue de faciliter le diagnostic et/ou le suivi de pathologies, mais également pour prévoir l'efficacité de certains traitements.

Caractérisation tissulaire

Dans le domaine de la caractérisation tissulaire, ces 10 dernières années ont été marquées par la multiplication des études portant sur le potentiel diagnostic de la microspectroscopie-IR. Parmi les applications possibles de cette approche on peut noter l'étude de l'encéphalopathie spongiforme bovine dans le tissu nerveux (Kneipp 2000), l'étude de la minéralisation du tissu osseux (Bohic 1998) ainsi que des études sur l'anévrisme aortique (Bonnier 2008). Cependant, l'engouement pour les applications de la microspectroscopie IRTF au niveau tissulaire concerne particulièrement la caractérisation des néoplasies (Argov 2002 ; Diem 2004 ; Sahu 2005 ; Ellis 2006 ; Kendall 2009). En effet, le processus de carcinogénèse est accompagné de modifications biochimiques au sein des tissus que l'on peut détecter de façon non destructive par spectroscopie vibrationnelle. Parmi les cancers étudiés, citons à titre d'exemple le sein (Fabian 2006 ; Bird 2008), l'utérus (Chang 2003 ; Romeo 2002), la peau (Ly 2009), l'estomac (Li 2005), l'œsophage (Wang 2007), la prostate (Baker 2009), la thyroïde (Liu 2003) ou encore le colon (Salman 2001 ; Lasch 2002 ; Lasch 2004 ; Sahu 2004 ; Conti 2008). De manière générale, on distingue 3 approches :

• Etude par collection de spectres « points »

Dans ce cas, un nombre limité de spectres est collecté à des endroits précis du tissu. En raison du faible nombre de spectres obtenus et de la taille des pixels utilisés (entre 50 et 100 μ m), les populations spectrales ainsi acquises sont peu représentatives de la forte hétérogénéité morphologique et biochimique des tissus, en particulier des tissus cancéreux.

Ce type d'étude a néanmoins permis la discrimination entre tissus sains et tissus cancéreux à partir de différences spectrales telles que des variations de ratios de certains composants tissulaires. Par exemple, des études ont montré que le ratio ARN/ADN (1121/1080 cm⁻¹) est plus élevé pour certains tissus tumoraux que pour les tissus sains correspondants (Andrus 2006). Des changements aux niveaux des bandes Amide I et II, sensibles à la structure primaire et secondaire des protéines, ont également été décrits (Petibois 2006 ; Wehbe 2008).

• Etude par imagerie IR et méthodes de traitements univariées

Une avancée technologique importante dans le cadre de l'étude de tissus par microspectroscopie-IR est le développement des systèmes d'imagerie IR permettant une cartographie spectrale des tissus. Plusieurs milliers de spectres sont alors collectés avec une résolution spatiale de l'ordre de $10 \,\mu m$.

L'imagerie IR permet donc de construire des cartographies chimiques de certains marqueurs de la progression tumorale à partir des absorptions caractéristiques de molécule comme le glucose, le glycogène, l'acide lactique ou les acides aminés (Yano 1996 ; Andrus 2006 ; Petibois 2007). Cependant, ce type d'étude basé sur des analyses « univariées » (ne prenant en compte qu'une seule bande du spectre ou, au mieux, un rapport entre deux bandes) ne permet pas d'exploiter la totalité des informations des spectres IRTF. Or, les changements spectraux associés à des pathologies sont généralement multiples et non pas limités à une seule vibration.

• Etude par imagerie IR couplée à des méthodes de traitements multivariées

Ces méthodes de traitements « multivariées » des données spectrales permettent aujourd'hui d'exploiter pleinement toutes les informations contenues dans un ensemble de données spectrales. De plus, le développement de la puissance de calcul des systèmes informatiques autorise aujourd'hui l'utilisation de ces méthodes pour le traitement de populations spectrales pouvant contenir plusieurs millions de spectres. L'imagerie IR, couplée à ces méthodes de traitements multivariées, à donner naissance au concept d'histologie IR, ou histologie moléculaire, par analogie avec l'histologie conventionnelle de type HE (Lasch 2004; Diem 2004).

Le travail présenté dans cette thèse se focalisant sur l'étude de tissus par histologie spectrale IR, ce concept est développé dans la partie suivante de ce manuscrit.

L'histologie spectrale IR

Les récents développements des méthodes de traitements multivariés des données spectrales autorisent aujourd'hui la pleine exploitation de l'information contenue dans une image IR. Grâce aux nouveaux outils de classification multivariées des données, la microspectroscopie IR peut être développée comme un nouvel outil pour l'histologie, offrant un compromis entre microscopie de transmission (informations spatiales) et techniques de biologie moléculaire (informations biochimiques de nature génomique, transcriptomique et protéomique), sans nécessité l'utilisation de quelconques marquages extrinsèques. L'histologie spectrale IR, ou histologie classique, permettant de visualiser un tissu selon sa composition biochimique (Diem 2004). Cette méthode peut être décomposée en trois étapes : acquisition d'une image spectrale à partir d'un échantillon tissulaire à l'aide d'un microspectro-imageur, le pré-traitement des données spectrales et la classification statistique des ces données par traitements « multivariés ». Cette approche permet de construire des images IR en pseudo-couleur (IRPC) comparables à l'histologie classique auxquelles sont associées des informations relatives à la composition biochimique des échantillons étudiés.

I. Les échantillons

Les échantillons utilisés pour l'histologie IR sont des coupes tissulaires montées sur des supports spéciaux, n'absorbant pas dans l'IR.

Types de tissus

Les coupes tissulaires utilisées en histologie IR doivent être préparées, tout comme pour l'histologie classique, par section au microtome de blocs de tissus. Usuellement, les études en histologie IR porte sur l'analyse de coupes cryogénisés en raison des problèmes engendrés par l'absorption IR de la paraffine ou de son déparaffinage chimique. Ces coupes, d'une épaisseur avoisinant généralement 8 µm, doivent être montées sur des supports n'absorbant pas dans l'IR, sans utilisation d'agent adhérent (tel que l'albumine).

Il est essentiel de noter que la fixation des tissus, qu'elle soit réalisée au formaldéhyde ou à l'alcool peut entrainer des modifications spectrales, particulièrement au niveau des bandes Amides I et II, sensible à la structure des protéines. Ainsi, les données IR d'échantillons fixés selon des protocoles différents ne sont pas directement comparables.

Types de supports

Il existe 4 grands types de matériaux particulièrement adéquats pour la fabrication de supports autorisant l'étude IR de tissus : séléniure de zinc (ZnSe), fluorure de calcium (CaF₂) et de baryum (BaF₂), lames de verres recouvert d'atomes d'argent et d'oxyde de d'étain (Ag/SnO₂) (fig.13).



Figure 13. Exemple de supports pour l'histologie IR. CaF₂ (gauche), ZnSe (centre) et lame Ag/ SnO₂ (droite)

Les différentes caractéristiques et conditions d'utilisations de ces supports en spectroscopie IR sont résumées dans le tableau 5.

Type de support	Transparence dans l'IR	Mode de mesure	Coloration (HE, IHC)	Prix (Euros)
BaF ₂	8 50- 40 00 cm ⁻¹	Transmission	Impos sible	~ 30
CaF ₂	95 0- 40 00 cm ⁻¹	Transmission	Possible	~ 30
ZnSe	400- 4000 cm ⁻¹	Transmi ssion	Possible Visibilité médiacre	~ 80
Ag/SnO2	400- 400 0 cm ⁻¹	Réflexion-absorption	Possible Visibilité médiocre	~ 1

Tableau 5. Caractéristiques des supports pour l'histologie IR.

L'ensemble des acquisitions effectuées dans le cadre de cette thèse ont été réalisées sur des fenêtres en ZnSe et CaF₂.

II. Acquisition de l'image spectrale d'un tissu

Lors de l'acquisition de données spectrales par imagerie IR différents paramètres modulables doivent être pris en compte selon le contexte de l'expérience.

A. Gamme spectrale

La gamme spectrale est définie comme la plage de nombres d'onde au sein de laquelle sera analysée l'absorption IR d'un échantillon. Dans le cas d'échantillons biologiques, la gamme spectrale la plus informative se situe entre 720 et 1800 cm⁻¹ (Parker 1983). Les gammes spectrales s'étendant de 400 à 750 cm⁻¹ et de 1800 à environ 2500 cm⁻¹ ne présentent généralement pas d'informations pertinentes sur la composition d'un tissu. La plage 2800-3100 cm⁻¹ quant à elle peut fournir des informations sur la composition lipidique d'un tissu.

Le choix d'une gamme spectrale en histologie IR pour une expérimentation donnée dépend de plusieurs facteurs :

Le type de support

Comme indiqué dans le tableau 5, certains supports IR peuvent présenter une forte absorption sur une partie de la gamme spectrale 400-4000 cm⁻¹. C'est notamment le cas du CaF₂, présentant une absorption IR significative entre 400 et 900 cm⁻¹.

Le mode de conservation du tissu

Il existe deux grands modes de conservation des tissus : l'inclusion dans la paraffine (coupes paraffinées) et la cryogénisation (coupes congelées).

De par l'altération des lipides engendrée par l'inclusion dans la paraffine, l'étude de l'absorption IR de coupes paraffinées sur la gamme spectrale 1800-4000 cm⁻¹ ne fournit pas d'informations pertinentes. Ces coupes ne sont donc étudiables que sur la gamme 400-1800 cm⁻¹. En revanche,

les coupes congelées peuvent être étudiés sur la gamme spectrale complète du moyen-IR (400-4000 cm⁻¹).

L'objectif de l'expérience

La gamme spectrale pourra être ajustée en fonction des informations recherchées. Par exemple, l'obtention d'une image spectrale d'un tissu en fonction de sa composition en carbohydrate pourra être obtenue en analysant uniquement la gamme spectrale 1000-1200 cm⁻¹. Un tel ajustement de la gamme spectrale présente l'avantage majeur de réduire la taille des données obtenues.

B. Résolution spectrale

On appelle résolution spectrale le pouvoir de séparation d'un spectromètre. Plus la résolution spectrale est grande, plus le spectre obtenu sera détaillé, et plus on pourra y distinguer des bandes de vibrations très proches en fréquence ou profondeur. Un spectre de haute résolution comportera donc une information plus détaillée, mais présentera l'inconvénient d'être plus sensible au bruit de fond. La résolution spectrale sélectionnée doit-être en adéquation avec le rapport S/B nécessaire à une interprétation non biaisée des résultats d'une expérimentation. La figure 14 montre clairement qu'une forte résolution spectrale (1 cm⁻¹) engendre l'obtention de spectres bruités par des signaux parasites non informatifs (a). A contrario, une faible résolution entraine l'obtention d'un spectre (b) fort peu détaillé où l'information spectrale semble « gommée », ceci en raison de l'effet de lissage qui est fonction de la résolution spectrale utilisée. En revanche, le spectre obtenu avec une résolution spectrale de 6 cm⁻¹ (c) présente un bon compromis entre ces deux extrêmes : il affiche de nombreuses bandes clairement résolues, sans pour autant présenter un bruit de fond significatif (Lasch 2007).

Par ailleurs, plus un spectre présentera une résolution importante, plus le temps d'acquisition des données pour une même surface tissulaire sera long. De manière générale, on considère qu'une résolution spectrale comprise entre 4 et 8 cm⁻¹ est adéquate pour la plupart des études sur tissus et offre un bon compromis entre qualité du signal et temps d'acquisition.



C. Résolution spatiale

En imagerie IR, comme pour n'importe quel type de microscopie optique, le paramètre de résolution spatiale est un des facteurs critiques définissant la fidélité d'une image (Lasch 2006). En effet, de la résolution spatiale d'un système optique dépend directement le type d'objet étudiable. Les échantillons biologiques tel que les tissus sont composés par diverses unités structurales exhibant une hétérogénéité sur plusieurs niveaux microscopiques : i) entre types de tissus distincts ; ii) entre différents types de cellules ; iii) entre différentes structures cellulaires. Ainsi, la résolution spatiale d'un système optique détermine quels types d'objets peuvent être examinés. Ce fait évident est bien connu en histopathologie clinique où les échantillons tissulaires sont analysés à différents grossissements, permettant tout d'abord d'évaluer les changements morphologiques les plus grossiers pour finir par l'étude de la morphologie de structures cellulaires telles que le noyau dans les cellules tumorales.

L'une des limitations de l'imagerie IR vient de la faible résolution spatiale inhérente à l'utilisation du rayonnement IR. En effet, en raison de la longueur d'onde importante du moyen IR ($\lambda = 2,5$ à 25 µm), la résolution spatiale des microscopes IR sera toujours inférieure à celles des microscopes utilisant la lumière visible ($\lambda = 400$ nm à 700 nm). Cet inconvénient est néanmoins compensé par la grande quantité d'informations contenues dans le domaine spectral de l'IR. Alors que les microscopes utilisant le domaine du visible fournissent usuellement 3 (Rouge-Vert-Bleu) ou 4 (Cyan-Magenta-Jaune-Noir) valeurs de couleurs par pixel, la microspectroscopie IR produit, pour chacun des pixels d'une image, un spectre complet pouvant contenir plusieurs milliers de valeurs d'absorbance IR. Cette quantité énorme d'informations qualitatives et semi-quantitatives, est directement liée à la composition biochimique l'échantillon ainsi qu'à ses caractéristiques structurales (Lasch 2006).

Les microspectro-imageurs actuels permettent généralement des acquisitions d'images présentant des pixels d'une taille minimum de $6,25x6,25 \mu m$. Le paramètre de résolution spatiale en histologie IR influe sur 3 points majeurs : la finesse de l'image, le type de spectres obtenus, et le rapport S/B.

Finesse de l'image

Logiquement, plus la résolution spatiale sera importante, plus l'image IRPC finale sera fine et détaillée, permettant ainsi de mieux distinguer les différentes structures tissulaires et ainsi d'obtenir une corrélation précise avec l'histologie classique.

Type de spectres obtenus

Une résolution spatiale élevée permet également d'obtenir des spectres qui seront très majoritairement associés à une seule et unique structure tissulaire, représentant donc la signature biochimique de celle-ci. Quand la résolution spatiale s'abaisse, un nombre croissant de spectres pourra correspondre au signal mixte de deux structures accolées. Les populations spectrales obtenues varieront donc selon la résolution spatiale employée (fig. 15).



Figure 15. Influence de la résolution spatiale sur les populations de spectres obtenus par imagerie IR. *RS*, résolution spatiale; *Cadre en pointillé*, zone du tissu à partir de laquelle est calculée l'absorbance d'un

Il est donc évident que le choix de la résolution spatiale doit tenir compte de l'architecture du tissu étudié. Plus le tissu sera hétérogène, plus la résolution devra être importante pour qu'une grande majorité des spectres obtenus correspondent au signal spectral « pur » des différentes structures le composant (RS = Y dans la figure 15). A contrario, l'utilisation d'une faible résolution spectrale dans un tissu très hétérogène aboutira à l'obtention de nombreux spectres « contaminés », c'est-à-dire correspondant au mélange des caractéristiques spectrales de plusieurs structures (RS = 2Y et 4Y dans la figure 15).

Rapport S/B

La résolution spatiale est un des paramètres qui a le plus d'influence sur le rapport S/B (Lasch 2006). En effet, plus la résolution spatiale sera importante, plus le signal en chaque point mesuré de l'échantillon sera faible. Ainsi, la part du bruit de fond sera d'autant plus importante que la résolution spatiale sera élevée.

D. Le nombre d'accumulations

Le nombre d'accumulations est défini comme le nombre de mesures réalisées pour l'obtention d'un spectre. Par exemple, pour une accumulation égale à 4, l'absorbance d'un point donné du tissu sera mesurée 4 fois. Ces 4 signaux sont ensuite moyennés en vue de l'obtention du spectre représentatif de cette zone tissulaire. Plus le nombre d'accumulations sera important, plus le rapport S/B sera élevé. En revanche, le temps d'acquisition d'une image spectrale est lié de façon linéaire avec le nombre d'accumulations. En imagerie spectrale, on choisit généralement un nombre d'accumulation de 2 ou 4, offrant un bon compromis entre rapport S/B et temps d'acquisition.

III. Le pré-traitement des données

La transformation d'une image spectrale IR brute en une image IR pseudo-couleur (IRPC) apportant des informations histologiques et biochimiques facilement interprétables nécessite une étape cruciale de pré-traitement des données spectrales. Ces pré-traitements, visant à l'optimisation des données spectrales, se basent essentiellement sur une approche multivariée, appelée également chimiométrie. La chimiométrie est un outil utilisé afin d'extraire des informations pertinentes et utiles à partir de données physicochimiques issus de systèmes complexes et donc généralement multivariables. Cette discipline recouvre des domaines d'applications très variées en chimie, physique, sciences de la vie, économie, sociologie ou informatique.

En spectroscopie, les pré-traitements chimiométriques sont utilisés pour corrigés les signaux caractéristiques de la composition chimique de l'échantillon étudié des signaux interférents originaires de ses particularités physiques ainsi que de son environnement. En effet, les spectres IR d'un tissu ne sont pas composés uniquement par les informations biochimiques relatives à sa composition intrinsèque. De nombreux autres facteurs contribuent à ce signal, et certains d'entre eux peuvent masquer des informations d'intérêt. Dans le cas de l'étude de tissus paraffinés en spectroscopie, les facteurs interférant les plus importants sont :

- la vapeur d'eau atmosphérique, présentant des bandes d'absorption dans la gamme spectrale 1300-1900 cm⁻¹ dont l'intensité dépend du taux l'humidité.
- la paraffine utilisée dans la préparation des coupes tissulaires et présentant des bandes d'absorption à 888, 1377 et 1464 cm⁻¹. Chacune de ces bandes est en réalité composée de multiples bandes se superposant et dont l'intensité relative peut être influencée par le protocole d'inclusion dans la paraffine ou par la présence d'autres composés tel que le tissu biologique. De par sa variabilité, le signal IR de la paraffine est difficilement

modélisable. Ainsi, les études en histologie IR portent usuellement sur des tissus congelés ou déparaffinés chimiquement.

 les effets de dispersion de la lumière, tel que l'effet Mie qui peut être défini par la dispersion de la lumière par des particules sphériques de même taille que la longueur d'onde du rayonnement (Mohlenhoff 2005).

Avant toute étude basée sur l'analyse spectrale de coupes tissulaires, il sera donc nécessaire de corriger ces interférences des données spectrales par pré-traitement chimiométrique pour ne retenir que la variabilité spectrale propre à la constitution biochimique de l'échantillon. Les méthodes d'analyses en composante principale et d'EMSC sont des outils de choix pour la modélisation et la correction des interférences spectrales IR.

L'analyse en composantes principales

L'analyse en composantes principales (ACP) est la plus simple et la plus connue des techniques d'Analyse de Données multivariées. L'ACP est une méthode mathématique visant à réduire le nombre de variables d'un ensemble complexe de données en une somme de données simplifiés (les composantes principales) mais exprimant la plus grande partie de la variance de l'échantillon initial. L'ACP permet donc de simplifier la représentation, la modélisation et l'interprétation de données spectrales en éliminant les informations redondantes et non caractéristiques, au profit d'une synthèse des informations caractérisant le mieux les données initiales. Plus de détails sur cette méthode sont disponibles en Annexe 3, p172.

L'EMSC

L'algorithme EMSC (Extended Multiplicative Signal Correction) est une extension de la méthode de « correction de dispersion multiplicative », ou MSC (Multiplicative Signal Correction) et qui vise à minimiser l'influence des facteurs d'interférences cités préalablement (Martens 1991 ; Köhler 2005). L'EMSC est une approche intégrative de minimisation de ces facteurs basée sur la modélisation des données spectrales et des effets de dispersion de la lumière. Ainsi, après correction par l'EMSC la variance inter-spectrale due aux effets d'interférences et de dispersions est supprimée et les spectres sont normalisés, autorisant ainsi leur étude comparative. Plus de détails sur cette méthode sont disponibles en Annexe 4, p174.

IV. Traitements chimiométriques pour la classification des données

La classification des spectres est une étape clef du traitement d'une image spectrale. Cette opération a pour but de regrouper les spectres en clusters (ou classes) afin de mettre en évidence des zones tissulaires présentant des caractéristiques spectrales, et donc biochimiques, communes. On peut regrouper les méthodes de classification en deux catégories :

- les classifications non-supervisées, qui visent à séparer automatiquement l'image en clusters naturels, c'est-à-dire sans aucune connaissance préalable de ceux-ci.
- les classifications supervisées, qui s'opèrent à partir de clusters définis au préalable par l'utilisateur, et qui visent à établir des « modèles de prédictions » de l'histologie IR..

C'est à partir de ces analyses par « clustering » que seront construites les images IR pseudocouleurs (IRPC) représentant l'histologie spectrale IR des tissus. En effet, suite à la classification des spectres, une couleur distincte est arbitrairement assignée à tous les spectres appartenant à un même cluster. Ces couleurs seront affichées sur une image en tant que pixels aux coordonnées (x,y) desquelles le spectre a été acquis, permettant donc de construire une image en pseudocouleur sur la base du clustering des données (Lasch 2004).

A. Les classifications non-supervisées : vers l'histologie IR

Plusieurs algorithmes de classifications non-supervisées des données spectrales sont disponibles mais tous sont basés sur le même concept général. Ce type de classification vise à grouper les données spectrales en classes par maximisation de la variance inter-cluster et minimisation de la variance intra-cluster, regroupant ainsi les spectres en fonction de leur « ressemblance ». Cette analyse est purement descriptive dans le sens où elle permet de mettre en évidence différentes structures au sein d'un tissu, en y associant une signature spectrale reflétant leur composition moléculaire. Ci-dessous sont présentés les aspects théoriques des 3 procédures de classifications non-supervisées les plus utilisées en spectroscopie.

Classification par K-moyennes (KM)

La classification par la méthode des K-moyennes (ou K-means, KM) est une méthode de classification non hiérarchique conduisant à une classification « dure » des données spectrales

(Steinley 2006), i.e. la probabilité d'appartenance d'un spectre à une classe donnée ne peut être que de 0 ou 1. Cette méthode, s'appuyant sur un algorithme itératif, est considérée comme nonsupervisée excepter au niveau du choix du nombre de classes

Les étapes de cet algorithme peuvent être donc être décrites comme suit (fig. 16) :

- Choix du nombre de clusters k par l'utilisateur.
- Choix aléatoire de k spectres qui représenteront alors les centres initiaux des clusters.
- Attribution de chaque point au centre du cluster le plus proche.
- Calcul des nouveaux centres, correspondant au spectre moyen de chaque cluster.
- Répétition des deux étapes précédentes jusqu'à qu'un critère de convergence soit atteint (habituellement, absence de changement d'assignation des spectres après plusieurs itérations).

Distance euclidienne



Figure 16. Illustration du principe de classification KM de données spectrales.

Classification hiérarchique ascendante (HA)

Tout comme la classification par KM, la classification hiérarchique ascendante (HA) est dite « dure » (la probabilité d'appartenance d'un spectre à une classe donnée ne peut être que de 0 ou 1). Au cours de la classification HA (fig. 17), deux objets (spectres dans notre cas) calculés comme étant les plus similaires au sein de l'ensemble de données spectrales sont fusionnés au sein d'un nouvel objet (cluster). Par la suite, les distances entre cet objet nouvellement formé et les objets restant sont réévaluées de façon à combiner de nouveau les objets les plus proches. Ce processus de fusion des objets peut être visualisé dans un diagramme arborescent, nommé dendrogramme. La partition finale des objets en clusters est définie en « coupant » ce

dendrogramme. Bien que le processus de classification HA soit complètement non-supervisé, cette étape, qui définira le nombre de classes composants l'image, est subjective.



Classification floue (FCM)

La classification « floue », ou Fuzzy C-means (FCM), est comparable à la classification par la méthode des KM. Cependant, à la différence de celle-ci et plus généralement des classifications dites "dures", associant chaque élément d'un jeu de données à un cluster unique, les FCM permettent d'affecter de manière non-supervisée chaque élément à plusieurs clusters différents selon des degrés d'appartenance compris entre 0 et 1 (dont la somme doit être égale à 1) introduisant ainsi une notion de nuance dans l'appartenance d'un élément aux clusters. Cependant, comme pour les classifications KM, les FCM nécessitent le choix par l'utilisateur du nombre de classes C. On notera toutefois que notre équipe travaille actuellement sur le développement d'algorithmes permettant la définition automatique et non-supervisée d'un nombre optimal de clusters pour la classification de données.

Les FCM sont contrôlés par un paramètre de fuzzification m. Avec m = 1 les FCM deviennent identiques aux KM. Quand m augmente, les classifications deviennent floues, et lorsqu'il tend vers l'infini toutes les classes sont identiques. En pratique, la valeur usuelle de m est égale à 2. Les images pseudo-couleurs alors obtenus apparaissent nuancées, chaque pixels d'un même cluster apparaitra plus ou moins estomper, renseignant alors sur la probabilité d'appartenance du spectre correspondant au cluster auquel il a été attribué.

Cette méthode de classification est actuellement en cours d'évaluation dans notre laboratoire en tant qu'outil pour la mise en évidence d'un front de progression tumorale, zone intermédiaire entre tissu sain et néoplasique, dans les cancers de la peau.

Comparaison des méthodes de classifications non-supervisées

La figure 18, tiré de Lasch *et al.* (2004), présente une comparaison entre la classification KM, FCM et HA d'une même image spectrale. Dans les trois cas, les images IRPC sont relativement similaires et peuvent être corrélées avec l'histologie classique.



Figure 18. Comparaison de l'histologie IR d'un adénocarcinome colorectal selon la méthode de classification des données spectrales employée (Lasch 2004). *A*, coupe colorée à l'HE. *B*, histologie IR obtenue par classification KM (6 clusters). *C*, histologie IR obtenue par classification FCM (6 clusters). *D*, histologie IR obtenue par classification HA (6 clusters).

De manière générale, la meilleure corrélation avec l'histologie classique est retrouvée après traitement par classification HA des données spectrales. Cette méthode présente également l'avantage d'être indépendante de l'initialisation randomisée des k ou C centres des clusters initiaux dans le cas des classifications par KM ou FCM, respectivement. Cependant, la méthode
HA présente l'inconvénient majeur de nécessiter d'importantes ressources informatiques, et s'avère donc particulièrement chronophage en comparaison des méthodes KM et FCM (tableau 6).

-	i ype cranalyse	Temps de calcul	Modélisation du temps de calcul
	KM FCM HCA	7 min	t (KM)~const ₁ x k x n
		30 min 4.5 h	t (FCM)~const ₂ x k x n
			t (HCA)~const ₃ x n²

classification KM, FCM, et HA d'une image spectrale.

En effet, dans le cas de l'algorithme de classification HA, le temps de calcul nécessaire à la partition des données spectrales augmente en fonction du carré du nombre de spectres contrairement au cas des KM et FCM où le temps de calcul augmente linéairement en fonction de la taille des données. De plus, il faut savoir que l'espace alloué à une application dans un système informatique est limité, ainsi la classification d'un grand ensemble de données spectrales par classification HA peut être tout simplement impossible en l'état actuel des performances des ressources informatiques. En pratique, les images spectrales traitées étant de plus en plus larges grâce au développement des systèmes d'acquisitions, de telles limites sont peu acceptables.

Ainsi, il est généralement considéré que les classifications par KM et dans une moindre mesure par FCM, sont actuellement les méthodes les plus adéquates pour le traitement de larges ensembles de données spectrales dans le cadre de l'histologie IR (Lasch 2004).

L'identification des structures tissulaires dans une image IRPC obtenue via une classification non-supervisée des données spectrales passe nécessairement, quelque soit la méthode utilisée, par une étape cruciale de comparaison avec les coupes histologiques correspondantes colorées à l'HE, requérant alors l'expertise d'un partenaire anatomopathologiste. De plus, ces méthodes de classification non-supervisées peuvent nécessiter actuellement un temps de calcul de plusieurs heures, ceci étant peu compatible avec une application en milieu clinique. En revanche, cette analyse permet de construire une banque de marqueurs spectroscopiques spécifiques des différentes structures tissulaires normales, bénignes et/ou malignes pour le type d'échantillon étudié.

B. Les modèles de prédiction de l'histologie IR

Sur la base des marqueurs spectroscopiques établis par classification non-supervisée, l'utilisation de méthodes de classification supervisées permet de mettre au point des modèles de « prédiction » de l'histologie IR pour l'examen « en aveugle » d'échantillons dont l'examen histologique conventionnel et le diagnostic sont inconnus, autorisant potentiellement l'application de cette méthode en milieu clinique (Bhargava, 2007). Actuellement, 3 types de classifications supervisées sont principalement utilisés dans le cadre de la spectroscopie IR : l'Analyse Discriminante Linéaire (LDA), les Réseaux de Neurones Artificiels (ANN) et les Séparateurs à Vaste Marge (SVM). Toutes ces méthodes sont basées sur le même concept général : ces classifications permettent, à partir d'un jeu de données X classées dans k clusters de déterminer l'appartenance d'un élément Y inconnu à l'un de ces k clusters, d'où l'utilisation du terme « superviser ». Ces modèles visent à l'automatisation fiable et rapide (en quelques minutes), compatible avec une application clinique de routine, du processus de reconnaissance histopathologique basé sur la classification non-supervisée des données spectrales issues de coupes tissulaires.

Quelle que soit la méthode employée, la mise au point d'un modèle prédictif de l'histologie suit un même schéma générale (fig. 19) (Bird 2008). Un tel modèle est en fait un ensemble de règles d'affectation de spectres inconnus à des clusters définis préalablement. Sa construction et sa validation doivent se faire sur des données sélectionnées. Une partie des données, nommée ensemble d'entrainement, provient d'échantillons représentatifs de la pathologie tumorale étudiée. Cet ensemble d'entrainement X va servir à construire l'algorithme de classification supervisée, c'est-à-dire les règles d'affectations des spectres à k clusters préalablement déterminés par classification non-supervisée. L'autre partie des données, appelée ensemble de validation et provenant d'échantillons non caractérisés histologiquement, va servir à évaluer et à valider le potentiel prédictif de cet algorithme, c'est-à-dire sa capacité à prédire l'histologie IR d'échantillons non inclus dans l'ensemble d'entrainement.



- * Les différents clusters sont corrélés avec l'histologie HE pour la construction du modèle
- ** Les différents clusters sont corrélés avec l'histologie HE pour la validation de la valeur prédictive du modèle

Figure 19. Processus de mise au point d'un modèle de prédiction de l'histologie IR.

Une fois ce modèle validé, il permettra l'analyse automatique en « aveugle » de l'histologie d'échantillons inconnus, en quelques minutes et sans nécessité l'utilisation de réactifs quelconque ni de contrôle avec l'histologie classique. Ces classifications supervisées, aussi appelées Analyses Discriminantes, permettent non seulement de générer des règles d'affectation, mais aussi de calculer les probabilités, pour chaque spectre, d'appartenance à chacun des clusters (probabilités dites "à posteriori").

V. Applications dans le domaine de la caractérisation des lésions tumorales

L'histologie IR peut donc permettre, en se basant sur des méthodes de classification nonsupervisées et supervisées des données IR, le diagnostic automatique et objectif de lésions tumorales. Ceci est illustré dans la figure 20 présentant une méthode de diagnostic différentiel de lésions de type BCC (Basal Cell Carcinoma), SCC (Squamous Cell Carcinoma) et maladie de Bowen, sur coupes tissulaires de peau humaine non déparaffinées. Cette méthode a été développée dans notre laboratoire (Ly 2009) et repose sur la combinaison d'une classification KM, pour l'identification des clusters spectraux, et d'une classification supervisée de type LDA, pour la génération d'un modèle de prédiction automatique de l'histologie IR.

Après acquisition d'images spectrales brutes provenant d'un ensemble d'entrainement et correction chimiométrique des spectres les composants (fig. 20A), les données spectrales sont classifiées par KM sur la base de 11 clusters (fig. 20B). La comparaison des images obtenues avec l'histologie classique (fig. 20C) permet alors d'attribuer à chaque cluster une seule et unique structure histologique (fig. 20D). Les clusters ainsi déterminés de manière non-supervisée pourront alors être compilés, pour l'établissement d'un modèle de prédiction, en fonction des informations recherchées. Dans cet exemple, le modèle de prédiction est construit selon 5 nouveaux clusters supervisés (fig. 20E): 1 cluster par type de lésions étudiées (BCC, SCC, Bowen), 1 cluster pour les structures tissulaire non-lésionnelles, et 1 cluster largement minoritaire de spectres ayant une probabilité d'attribution non satisfaisante aux clusters précédents (P < 0.75). A chacun de ces clusters sera attribuée une couleur distincte. Lors de l'analyse de nouveaux échantillons (fig. 20F), les nouvelles données spectrales obtenues seront attribuées à ces clusters selon des règles d'affectations établies par LDA (fig. 20G). Une fois ce modèle prédictif validé, l'histologie IR permet alors, dans ce cas, de diagnostiquer automatiquement les différents types de lésions cancéreuses étudiées (BCC, SCC et maladie de Bowen) dans des échantillons inconnus, sans besoin de recourir à l'œil expert d'un pathologiste (fig. 20*G*).



Figure 20. Diagnostic différentiel de lésions tumorales de la peau par histologie IR supervisée (Ly 2009). *P*, paraffine. *Ti*, tissu.

Ce type de démarche a été appliqué avec succès pour la caractérisation automatique de différents tissus. On citera en exemple la caractérisation histopathologique de coupes tissulaires de prostate (Fernandez 2005 ; Bhargava 2007) ou de sein (Fabian 2006), de micro-métastases du cancer du sein dans les ganglions lymphatiques (Bird 2009), ou de carcinome utérin (Steller 2006).

Le cancer du colon

Avec plus de 33 000 nouveaux cas recensés en France durant l'année 2005, le cancer du colon se situe au 3ème rang des cancers dans la population, et au 2ème rang de la mortalité par cancer avec près de 17 000 décès. Il s'agit essentiellement d'un cancer associé à l'âge mûr, 85 % des cas survenant après 65 ans. On citera parmi les facteurs de risque, l'âge, les antécédents familiaux, les maladies inflammatoires chroniques comme la maladie de Crohn ou la colite ulcéreuse, ou bien encore des causes environnementales comme une alimentation riche en graisse et/ou sensiblement contaminée par certaines substances toxiques (pesticides, organochlorés ...).

I. Anatomie du colon

Les fonctions principales du colon sont la déshydratation du bol alimentaire et l'évacuation des déchets de la digestion. Il s'étend du cæcum jusqu'au rectum et constitue la partie terminale de l'intestin, appartenant à l'appareil digestif. Il se dispose en cadre dans la cavité abdominale et mesure de 1 m à 1,5 m avec un diamètre de 8 cm (cæcum) à 4 cm (rectum).

En considérant les différents changements de direction qu'il effectue, il peut être divisé en 5 segments (fig. 21) :

- Le colon ascendant (ou droit) qui est fixe, montant verticalement du caecum jusqu'à la face inférieure du foie.
- Le colon transverse, qui est mobile, dirigé transversalement de droite à gauche et légèrement de bas en haut, jusqu'au dessous de la rate.
- Le colon descendant (ou gauche) qui descend verticalement jusqu'à un plan horizontal qui correspond à la partie la plus élevée de la crête iliaque.
- Le colon sigmoïde qui descend verticalement jusqu'à croiser le muscle psoas gauche pour se continuer dans le rectum au niveau de la troisième vertèbre sacrée.



Contexte scientifique et technique

II. Histologie du colon

Les différents segments du colon sont histologiquement semblables et se caractérisent par l'absence d'éléments d'amplification de surface : pas d'anses, pas de valvules conniventes ni de villosités. La paroi du colon est constituée de 5 couches concentriques (fig. 22*A*) :

- La séreuse, ou adventice, est une couche de tissu conjonctif dense vascularisée et comportant de nombreux adipocytes. Cette couche se termine par un mésothélium, qui est le feuillet viscéral du péritoine.
- La musculeuse est formée de cellules musculaires lisses disposées selon deux axes formant ainsi une couche circulaire interne et une couche longitudinale externe.
- La sous-muqueuse (fig. 22*B*), est composée d'un tissu conjonctif plus dense contenant des vaisseaux sanguins et un réseau de nerfs sympathiques, le plexus de Meissner, qui commande la motilité du tube digestif.
- La musculaire-muqueuse, aussi appelée muscularis-mucosae, est formée de cellules musculaires lisses (fig. 22*B*).
- La muqueuse (fig. 22*B*) est délimitée par un épithélium de revêtement s'invaginant dans la muqueuse en glandes de Lieberkühn et située au dessus d'un chorion composé d'un tissu conjonctif lâche, très vascularisé, riche en cellules immunitaires organisées en formations lymphoïdes.



Contexte scientifique et technique

Les glandes de Lieberkühn (fig. 22*C*) sont essentiellement composées de cellules caliciformes, sécrétrices de mucus, et de quelques entérocytes qui jouent un rôle dans l'absorption de l'eau et des sels afin de concentrer les matières fécales. Le mucus est destiné à faciliter la progression du contenu intestinal et à protéger l'épithélium.

III. Formes histologiques du cancer du colon

La forme commune du cancer du colon est l'adénocarcinome, représentant près de 95% des cas. Les adénocarcinomes se développent à partir des glandes de Lieberkühn et sont définis par leur grade, correspondant à des degrés de différenciations variables de la tumeur et fonction de sa ressemblance avec l'épithélium glandulaire normal. Ainsi, 3 grades principaux sont définis :

- Les tumeurs bien différenciées sont uniquement constituées de structures glandulaires, proches des glandes de Lieberkühn.
- Les tumeurs moyennement différenciées comportent de nombreuses structures glandulaires (avec lumière) et des zones plus compactes.
- Les tumeurs peu différenciées présentent peu de lumière glandulaire et un stroma qui est souvent réduit, entrainant un aspect très compact.

Selon leur aspect morphologique, on distingue différents types de cancer du colon. On peut citer l'adénocarcinome muqueux (ou colloïde) et le carcinome à cellules isolées qui sont également observés dans d'autres organes, et deux formes spécifiques au colon, le carcinome médullaire et l'adénocarcinome festonné (Selves 2009).

IV. Développement et hétérogénéité du cancer du colon

Dans le colon, il existe une véritable filiation entre tumeurs bénignes (adénomes) et tumeurs malignes (adénocarcinomes/carcinomes). Dix à quinze pour cent des adénomes sont susceptibles de cancérisation. Ainsi, la grande majorité des adénocarcinomes coliques dérivent d'un polype adénomateux. L'évolution vers la malignité d'une tumeur bénigne est donc fréquente dans la muqueuse colique, ce qui ne permet pas de conserver l'opposition absolue, entre tumeurs bénignes et tumeurs malignes : il s'agit seulement de tumeurs présentant des grades différents d'évolutivité. Ce continuum appelé « progression tumorale » correspond à l'acquisition progressive par la tumeur d'un phénotype de malignité, d'anomalies chromosomiques et géniques en nombre croissant (Michor 2005).



Figure 23. Modèle de Volgelstein illustrant les étapes de la progression de la tumorogénèse colorectale (Rajagopalan 2003).

Le cancer colorectal a été considéré pendant plusieurs années comme étant la résultante d'une séquence d'altérations génétiques relativement uniforme et linéaire. Ce modèle de progression, établit par Fearon et Volgelstein (Fearon 1990) (fig. 23), décrit l'inactivation du gène APC (Adenopolyposis Coli) comme l'élément initiateur de la progression tumorale. Par la suite, une accumulation de mutations spécifiques d'oncogènes tels que KRAS (Kirsten Rat Sarcoma viral oncogene homolog) ou de gènes suppresseurs de tumeurs tels que TP53 entraine l'apparition de clones présentant une agressivité croissante. L'instabilité génétique est alors supposée intervenir durant une de ces étapes de la tumorogénèse, accélérant ainsi le taux de mutations dans les cellules cancéreuses (Rajagopalan 2003). Ce facteur additionnel implique la perte d'un ou plusieurs mécanismes critiques, non seulement pour le maintien de la fidélité génomique durant la division cellulaire, mais également pour le déclenchement du phénomène d'apoptose lors d'une accumulation d'altérations génétiques.

La séquence linéaire d'altérations génétiques illustrée dans le modèle de Volgelstein procure une illustration facilement compréhensible des principes fondamentaux sous-jacents à la tumorogénèse colorectale. Cependant, ce concept s'appuyant sur l'inactivation du gène suppresseur de tumeur APC comme potentiel élément initial de tous les cancers colorectaux apparaît aujourd'hui comme une simplification excessive de la tumorogénèse colorectale (Smith 2002 ; Samowitz 2007). En effet, si ce modèle s'avérait correct, la plupart des cancers colorectaux devrait présenter une trace de la mutation de ces gènes clefs (Jass 2007). Dans les faits, de nombreuses études ont montré que seul 10% des cancers du colon présentent des mutations de ces 3 gènes et que cette maladie est caractérisée par une hétérogénéité génétique

considérable. Ainsi, les cancers colorectaux comprennent donc plusieurs sous-groupes avec des caractéristiques cliniques, pathologiques et moléculaires différentes. De nouveaux types de classification multiparamétrique de ces cancers apparaissent donc, s'appuyant par exemple sur l'instabilité génétique, la méthylation des ilots CpG, et les mutations de BRAF (V-raf murine sarcoma viral oncogene homolog B1) et KRAS, et ayant pour finalité une meilleure compréhension des mécanismes de tumorogénèse ainsi qu'une amélioration de la prise en charge des patients (Jass 2007). Dans le même esprit, 2 sous-groupes d'adénocarcinomes muqueux, histologiquement semblables, ont pu être mis en évidence en corrélant données moléculaires (instabilité micro-satellitaire, présence de p27, Fhit, Cox-2) et données cliniques (Sarli 2008).

En parallèle à cette hétérogénéité inter-tumorale, une hétérogénéité intratumorale à également été décrite (Losi 2005). Comme expliqué précédemment, au cours de la progression d'un cancer colorectal chaque nouvelle étape est initiée par l'acquisition d'altérations contribuant à la progression de la lésion tumorale. Ceci engendre une expansion clonale de la cellule anormale par l'acquisition d'un avantage prolifératif. Par conséquent, ceci implique une certaine hétérogénéité génétique entre les cellules constituant la tumeur. Cette hétérogénéité génétique a été démontrée, dans les cancers colorectaux, pour les mutations de gène KRAS et p53. Au sein de tumeurs primaires colorectales, l'hétérogénéité génétique est plus souvent observée dans les stades précoces que dans les stades avancés. En effet, en raison de la sélection clonale, l'hétérogénéité génétique intratumorale, au moins pour les mutations points, décroit significativement pendant la progression tumorale.

L'hétérogénéité génétique intratumorale pourrait expliquer les données controversées concernant la significativité pronostique de marqueurs moléculaires potentiels dans les cancers colorectaux. En effet, l'analyse d'une seule petite zone tissulaire issue d'une tumeur hétérogène ne peut être représentative de la tumeur dans son entier.

V. Facteurs pronostiques et prédictifs

A l'heure actuelle, l'évaluation pronostique et prédictive des cancers coliques repose principalement sur le système de stadification TNM, tel qu'il est défini pour ce cancer dans la classification UICC (Union Internationale Contre le Cancer) de 2002 (Zlobec 2009). Cependant, les cancers coliques doivent être considérés comme une maladie hétérogène pouvant emprunter de multiples voies biomoléculaires de progression. Cette observation est corroborée par le fait que des tumeurs histologiquement semblables peuvent présenter des pronostiques d'évolution et des réponses aux traitements différents. Ainsi la détection de marqueurs moléculaires génétiques ou protéiques est couramment étudiée comme un moyen d'améliorer l'évaluation pronostique et prédictive des lésions tumorales coliques.

A. La classification TNM et autres facteurs morphologiques

Les critères de classification TNM des cancers coliques sont schématisés ci-dessous.



Figure 24. Classification TNM des cancers coliques.
T _{is} : carcinome <i>in situ</i> .
T1 : envahissement de la sous-muqueuse sans la dépasser.
T2 : envahissement de la musculeuse sans la dépasser.
T3 : envahissement de la sous-séreuse et du tissu péricolique.
T4 : la tumeur envahit les organes voisins.
N1 : métastases ganglionnaires $< ou = 3$.
N2 : métastases ganglionnaires > 3.
M0 : absence de métastases.
M1 : présence de métastases.

A partir de la classification obtenue, le stade de la maladie est défini comme suit :

```
Stade 0 : T<sub>is</sub>Stade 1 : T1N0M0<br/>T2N0M0Stade IIA : T3N0M0Stade IIB : T4N0M0Stade III : T(1-4) N1M0<br/>Stade IV : T(1-4) N(1-4) M1Stade IV : T(1-4) N(1-4) M1
```

La dernière révision de la stadification TNM des cancers coliques inclut une subdivision du stade II en IIA (T3N0M0) et IIB (T4N0M0). Les patients au stade IIB présentent un pronostique très mauvais comparé aux patients au stade IIA. Et, alors que les thérapies adjuvantes (chimio- ou radiothérapies suivant une résection chirurgicale) sont largement utilisées dans le cas de cancer au stade III, il n'est pas recommandé de les appliquer aux cancers de stade II.

La gradification des tumeurs, qui se base principalement sur le pourcentage de formations glandulaires peut également être utilisée comme approche additionnelle à l'évaluation pronostique des tumeurs coliques. Mais l'existence de différents systèmes de gradification complique la réalisation de méta-analyses, nécessaires pour une détermination pertinente de leurs significativités pronostique et prédictive.

On peut également citer d'autres facteurs morphologiques tels que l'invasion lymphatique et veineuse ou le bourgeonnement tumoral (Zlobec 2008 ; Sablin 2009).

B. Les marqueurs moléculaires

L'évaluation du potentiel pronostique et prédictif d'altérations génétiques détectées durant la progression du cancer colique est une problématique à l'origine de très nombreuses études. L'instabilité micro-satellitaire (MSI), la perte allélique du bras long du chromosome 18q (LOH18q) et les mutations de p53 ont été particulièrement étudiées.

• MSI : une récente méta-analyse (Popat 2005) a permis de mettre en évidence un taux de survie significativement amélioré pour les malades présentant cette altération. Cependant,

ces résultats sont issus d'études rétrospectives et doivent être confirmés par une approche prospective.

- LOH18q : les patients atteints des cancers de stade II ou III présentent un pronostique significativement plus mauvais lorsque cette altération est détectée (Zlobec 2008).
- TP53 : malgré plus de 20 ans d'investigation, l'impact de la mutation de TP53 sur le pronostique clinique est loin d'être concluant (Zlobec 2008).

C. Les marqueurs protéiques

Pour les cancers coliques, l'IHC n'a pas encore trouvé d'application de routine dans le cadre de leur évaluation pronostique et prédictive. Les marqueurs protéiques actuellement étudiées pour leur potentiel pronostique sont généralement impliqués dans la large gamme des voies de signalisation en jeu dans la progression et la dissémination des tumeurs coliques. La détection immunohistochimique de VEGF (Vascular Endothelial Growth Factor) et d'EGFR (Epidermal Growth Factor Receptor) sont les marqueurs protéiques pronostiques les plus prometteurs dans le cas du cancer du colon (Zlobec 2008). Mais les études initiales présentant de nouveaux marqueurs prometteurs sont souvent suivies de publications contredisant ces résultats préliminaires. Un certain nombre de raisons expliquant ces résultats contradictoires sont reconnues : notamment, les différences méthodologiques, le piètre design de certaines études, la petite taille des échantillons, ainsi que les différences dans l'évaluation des marquages immunohistochimiques. En effet, la significativité pronostique d'un marqueur protéique est grandement affecté par le choix de la méthode de comptage et des valeurs seuils déterminant la « positivité » de l'expression (McShane 2006).

Il existe donc de nombreux paramètres d'évaluations pronostiques et prédictives complémentaires à l'analyse morphologique standard, mais aucun d'entre eux n'est actuellement recommandé pour une utilisation en milieu clinique (Duffy 2007). En effet, si les outils biomoléculaires et immunohistochimiques ont permis d'améliorer notre compréhension de la progression des tumeurs coliques, les bénéfices pour les patients restent limités. Ceci est du en grande partie au caractère réductionniste et ponctuel de ces analyses qui, pour des raisons techniques, se focalisent traditionnellement sur un nombre limité de gènes ou de protéines, ignorant la nature complexe et hétérogène de la maladie au niveau moléculaire (Chabannon

2004). Le développement des techniques d'analyse « haut-débit », tant génomique que protéomique, présente donc dans ce cadre une piste intéressante pour la découverte de marqueurs tumoraux prenant mieux en compte l'hétérogénéité et la complexité du cancer colique.

VI. Les mucines coliques secrétées

Le mucus colique est un élément clé de la barrière colique puisqu'il se situe à l'interface entre la flore luminale et la muqueuse colique en elle-même. La première grande fonction du gel de mucus et des mucines qui le composent est de protéger physiquement la muqueuse colique. En effet, le gel de mucus tapisse et lubrifie la paroi intestinale, ce qui facilite le passage des selles et protège l'épithélium contre l'érosion mécanique. Le mucus est typiquement composé de mucines, qui sont des glycoprotéines de haut poids moléculaire. Elles sont constituées d'un squelette protéique appelé apomucine sur lequel sont greffées de nombreuses chaînes glycanniques (Byrd 2004).

On distingue deux grandes familles de mucines dont le profil d'expression est tissu-spécifique : les mucines sécrétées et les mucines transmembranaires. Dans le colon, MUC2 est la principale mucine sécrétée, et MUC1, MUC3 et MUC4 les principales mucines transmembranaires. Si les mucines membranaires ne sont pas secrétées, elles jouent néanmoins un rôle dans la cohésion et le maintien de la couche de mucus. Les caractéristiques physico-chimiques et structurales des mucines secrétées, en particulier leur glycosylation permettent au mucus de jouer un rôle de « tampon » entre la lumière et la muqueuse colique, assurant un rôle de protection physique, chimique et antimicrobienne de cette dernière. Certaines mucines peuvent également jouer un rôle dans les phénomènes de différenciation et croissance cellulaire, et subir des altérations qualitatives et quantitatives pendant la tumorogénèse colique (Gaudier 2006 ; Cone 2009).

A. Aspect structural des mucines secrétées

D'un point de vue structural, les mucines secrétées consistent en un polypeptide densément substitué avec des chaines latérales d'oligosaccharides O-liés et, dans une moindre mesure, Nliés. Chaque monomère de mucine, d'une longueur de 0.2 à 0.6 μ m, est lié au suivant par sa partie terminale via des ponts disulfures. Les multimères ainsi obtenus peuvent présenter une taille de plusieurs microns, et sont généralement plus longs que le diamètre de la plupart des cellules. De manière schématique, on peut rapprocher une molécule de mucine d'un écouvillon, sur lequel on trouverait des centaines de carbohydrates liés à l'apomucine (fig. 25) (Dekker 2002).



Figure 25. Aspect structural des mucines secrétées. *A*, représentation schématique de MUC2. *B*, représentation schématique de chaines latérales de carbohydrates des mucines. *GalNAc*, N-acetyl-galactosamine ; *GlcNAc*, N-acetyl-glucosamine ; *F*, fucose ; *M*, mannose ; *G*, galactose ; *SA*, acide sialique ; *S*, résidu sulfate. *Ser*, sérine ; *Thr*, thréonine ; *Asp*, asparagine.

Comme mentionné ci-dessus et illustré dans le figure 25, la majorité des chaines glycanniques sont jointes à l'apomucine par une liaison O-Glycosydique entre un résidu sérine ou thréonine et une molécule de N-acetyl-galactosamine (GalNAc). Ces séquences d'acides aminés O-Glycosylés, hautement répétitives et connues sous le nom de régions PTS, sont riches en proline, thréonine et sérine, et peuvent compter plus de 6000 acides aminés. Dans les parties N-ter et Cter des polypeptides de mucines, au-delà des régions PTS, se trouvent les régions non-PTS. Elles contiennent généralement tous les résidus cystéines des mucines. Ces séquences, sur lesquelles on retrouve les chaines de carbohydrates N-liés via des molécules de N-acetyl-glucosamine (GlcNAc), sont peu répétées et peuvent être caractéristiques du type de mucines. A coté de GalNAc et GlcNAc, les mucines contiennent du fucose, du mannose, du galactose, ou des acides sialiques tels que l'acide N-acetyl-neuraminique. La présence de mannose est caractéristique des glycanes N-liés. L'assemblage de ces différents sucres, via des glycosyl transférases, engendre des chaines latérales de carbohydrates présentant une hétérogénéité considérable au niveau de leur composition et de leur longueur, pouvant varier de 1 à 20 résidus. Enfin, en raison de la sialylation et de la présence de résidus sulfates sur les glycannes, les mucines sont chargées négativement.

La composition en acides aminés et en carbohydrates des mucines provenant d'organes différents est proche, mais pas identique. La significativité biochimique de ces différences entre mucines n'est pas encore clairement été définie. Une explication possible serait que différents gènes MUC pourraient être exprimés dans différent organes dans l'optique de remplir des fonctions spécifiques.

B. Fonctions spécifiques jouées par les mucines coliques secrétées dans la barrière épithéliale

Il est maintenant admis que limiter les fonctionnalités des mucines uniquement à leur rôle protecteur serait bien trop restrictif. En effet, des travaux récents ont permis de montrer que certaines mucines remplissent des fonctions spécifiques d'échange d'informations pour les cellules épithéliales, et interviennent dans les mécanismes de renouvellement des épithélia (Gaudier 2006).

Pour le colon, il est maintenant établi que MUC2 est impliquée dans les phénomènes de différenciation et croissance cellulaire. L'invalidation de l'expression du gène MUC2 chez la

souris apporte une preuve du rôle de cette mucine dans la structuration des cryptes et la maturation cellulaire : les souris knock-out pour MUC2 présentent une morphologie aberrante des cryptes intestinales, une absence de cellules à mucus reconnaissables tout au long de l'intestin, une maturation et une migration cellulaires altérées (Velcich 2002). Le rôle de MUC2 sur la régulation de la croissance et de la différenciation cellulaire est crucial, puisque ces mêmes souris invalidées pour le gène MUC2 développent des adénomes dans l'intestin grêle dès six mois, et des adénocarcinomes coliques et rectaux invasifs à l'âge de un an. Ainsi, MUC2 aurait des propriétés anti-tumorales. Un faisceau de preuves indirectes conforte ce rôle de MUC2 dans la différenciation de l'épithélium colique. Par exemple, l'étude de l'expression des différentes apomucines au cours du développement fœtal apporte des renseignements sur la mise en place de l'épithélium colique (Buisine 1998; Buisine 2000). MUC2 apparaît en effet être exprimé très précocement dans l'intestin en développement. Cette expression de MUC2 correspond à la différenciation antéropostérieure et cryptovillositaire de l'intestin. De plus, par la présence de domaines riches en cystéine de type CK (Cystein Knot, nœud cystéine), l'analogie structurale de MUC2 avec de nombreux facteurs de croissance (activines, inhibines, NGFs (Nerve Growth Factors), GDNF (Glial Cell Derived Neutrophic Factor) conforte son rôle potentiel dans la différenciation et croissance cellulaire. Enfin, il est également suggéré que MUC2 pourrait avoir un rôle dans les processus de réparation de l'épithélium, mais les preuves directes de cette fonction de MUC2 sont encore inexistantes (Gaudier 2006). Même si les mécanismes d'action de MUC2 ne sont pas connus, son rôle central dans la croissance et la différenciation cellulaire de l'épithélium colique est donc aujourd'hui admis, tout comme ses propriétés anti-tumorales, et son rôle potentiel dans la réparation de l'épithélium.

Les autres mucines sécrétées du colon, MUC5B et MUC6, sont faiblement exprimées dans le colon. Leur principale fonction repose vraisemblablement dans leur participation à la formation du gel de mucus protégeant la muqueuse. Pour MUC5B, aucune autre fonction spécifique n'a été jusqu'ici clairement démontrée. En revanche, l'implication de MUC6 dans la réparation épithéliale est suggérée par l'observation, dans la maladie de Crohn, d'une surexpression de MUC6 dans les cellules de type UACL (Ulceration-Associated Cell Lineage), qui se développent lors de la restitution de l'épithélium (Wright 1990).

C. Les altérations des mucines pendant la tumorogénèse colique

Il a été démontré à plusieurs reprises que les mucines peuvent subir des altérations qualitatives et quantitatives, notamment lors le la tumorogénèse colique (Bara 1986 ; Jass 2001 ; Aksoy 2004 ; Byrd 2004 ; Perçinel 2007 ; Forgue-Lafitte 2007). Celles-ci sont principalement de 3 natures : altérations des acides sialiques, altérations de l'assemblage des chaines latérales de carbohydrates, et altérations de l'expression des apomucines.

Le tableau 7 présente les altérations liées à la tumorogénèse colique les plus communément admises à l'heure actuelle pour les adénocarcinomes Lieberkühniens et les adénocarcinomes Lieberkühniens muqueux.

APOMUCINES	+ MUC1 - MUC2 -/ n MUC3 +/- MUC4	- MUC1 +/n MUC2	
	- MUC2 -/ n MUC3 +/- MUC4	+/n MUC2	
	-/ n MUC3 + / - MUC4		
	+ / - MUC4		
	.,		
	A MUC5AC	A MUC5AC	
GLYCOSYLATION	Motifs modifiés	Motifs modifiés	•
SIALYLATION	- O-acétylation des acides sialiques - Slalylation	- O-acétylation des acides sialiques	

Altérations des acides sialiques

Comme noté précédemment, des acides sialiques sont présents sur les chaines latérales de carbohydrates des mucines coliques, notamment au niveau latérale et périphérique. Des substituants O-acétylés sur C4 et C7-9 caractérisent les acides sialiques coliques. La perte de ces groupements peut se produire spontanément lors de pathologies tumorales et peut être évaluée par des procédures histochimiques. Cette perte de substituants O-acétylés entraine une altération de l'antigénicité des mucines. En effet, la plupart des anticorps réagissant avec les structures sialylées des carbohydrates tel que STn, SLex ou SLea, reconnaissent ces structures seulement si elles sont dépourvues d'O-acétylation. Par le passé, des articles se sont focalisés sur l'étude de

les

l'expression de STn, SLex et SLea en tant que marqueurs associés aux tumeurs. En fait, ces structures sont exprimées dans les mucines du colon normal, mais sont antigéniquement masquées par l'influence des substituants O-acétylés (Jass 2001).

Altérations de l'assemblage des chaines latérales de carbohydrates

L'extension, le raccourcissement ou le branchement anormal des chaines latérales de carbohydrates des mucines coliques ont été décrits pour différentes pathologies, incluant la tumorogénèse. Ce phénomène peut s'expliquer par l'échec de l'activation et de l'orchestration transcriptionnelle ordonnée des glycosyl transférases et/ou par l'altération de l'expression des apomucines (Jass 2001; Aksoy 2004; Sternberg 2004). Ces modifications structurelles pourraient altérer les fonctions ainsi que les propriétés adhésives et antigéniques des mucines (Mechenin 2000; Brockhausen 2006). Des études comparatives ont également pu montrer, pour les cancer coliques, une réduction du contenu total en carbohydrate pouvant atteindre 50% ainsi qu'une modification de la longueur moyenne des chaines de carbohydrates, passant 10 à 6 sucres dans les cancers coliques (Byrd 2004).

Altérations de l'expression des apomucines

L'altération du profil d'expression des apomucines a largement été décrite dans les cancers coliques (Byrd 2004 ; Aksoy 2004). Ces altérations peuvent correspondre soit à une sur- ou sousexpression d'une mucine normalement présente dans le colon (MUC2 par exemple), soit à l'apparition de mucines non détectables dans le tissu sain, mais anormalement exprimées dans les tissus tumoraux, comme MUC5AC (tableau 7). En effet, MUC5AC, un produit de la muqueuse gastrique normale, n'est pas exprimé dans le colon sain mais peut être détecté dans certains adénomes et adénocarcinomes coliques, tel que démontré par Forgue-Lafitte *et al.* (2007). En outre, le profil d'expression des mucines pourrait être associé à des changements subtils du phénotype des tumeurs.

Ces changements de profil d'expression, tout comme ceux concernant la glycosylation des mucines, pourraient avoir un impact sur de multiples mécanismes cellulaires tel que la croissance, la différenciation, l'adhésion ou l'invasion (Jass 2001 ; Gaudier 2006). Par exemple, une étude suggère le potentiel pro-métastatique de MUC2 lors de la tumorogénèse colique (Sternberg 1999). Cet effet pro-métastatique de MUC2 pourrait en partie expliquer le fait que les

adénocarcinomes muqueux, qui surexpriment MUC2, présentent un fort pouvoir métastatique et un pronostic plus grave que les adénocarcinomes non muqueux, qui présentent une expression de MUC2 diminuée (Corfield 2000). L'effet pro-métastatique de MUC2, lorsqu'elle est exprimée par des cellules cancéreuses pourrait être lié à l'altération de sa glycosylation. En particulier, le motif SLex est détectable sur MUC2 dans le cas de cancers coliques, et l'expression de ce motif est corrélée avec une augmentation des métastases et une diminution de la survie des malades (Byrd 2004). Les variations de profil des mucines ont ainsi fait l'objet de nombreuses analyses IHC dans le cadre de l'évaluation de leur potentiel en tant que marqueurs tumoraux, engendrant des résultats contradictoires. Ceci pourrait s'expliquer en partie par l'épissage alternatif des ARNm codés par les gènes MUC. Ainsi, pour MUC2, la présence de 2 isoformes MUC2 et MUC2.1 a été mise en évidence et pourrait expliquer les contradictions des études portant sur la corrélation entre la relative abondance de MUC2 et/ou de ses ARNm avec l'évolution des cellules cancéreuses du colon (Sternberg 2004 ; Schwartz 1992 ; Bresalier 1991 ; Cho 1997). Cependant, à l'heure actuelle, l'existence de variations d'épissages de MUC2 entre colon normal et adénocarcinomes, et/ou entre adénocarcinomes muqueux et non-muqueux, n'a pas été prouvée.

Si aujourd'hui, il est largement reconnu que l'expression des apomucines, la composition de leurs chaines latérales de carbohydrates ainsi que leur sialylation sont altérées durant la tumorogénèse colique, le fait que ces altérations soient des causes ou des conséquences de ce processus pathologique n'a pas encore été clairement établi (Gaudier 2006).

Chapitre 1

Pré-traitement de données spectrales issues de coupes tissulaires paraffinées pour l'histologie IR

INTRODUCTION

Dans le cadre du développement de l'histologie IR pour la mise en évidence de marqueurs tumoraux pronostiques et prédictifs, une étape importante de validation de l'approche doit être effectuée par la réalisation d'études rétrospectives. De telles études s'appuient sur des échantillons disponibles dans les banques de tissus, et majoritairement conservés par inclusion dans la paraffine. En effet, ces tissus représentent une source d'informations histopathologiques, moléculaires et cliniques considérable. Cependant, la paraffine présente de fortes bandes d'absorptions IR ainsi qu'un signal très variable dépendant du protocole d'inclusion ou de l'interaction avec d'autres constituants, en particulier le tissu biologique. Si il est possible de déparaffiner chimiquement les échantillons tissulaires, cette méthode est à proscrire dans le cas d'études IR. En effet, cette procédure fastidieuse nécessitant l'utilisation de solvants hautement toxiques de type xylène peut s'avérer incomplète (Faoláin 2005). Elle peut également altérer la composition biochimique des tissus et par conséquent l'information spectrale collectée lors de l'analyse IR.

Cette partie présente un modèle mathématique original de correction des données spectrales basée sur l'EMSC. Ce modèle vise à corriger non seulement les interférences dues à la vapeur d'eau atmosphérique et aux effets de dispersion de la lumière mais également celles engendrées par la paraffine. La modélisation du signal de la paraffine et sa soustraction à l'ensemble des données spectrales permet un « déparaffinage numérique » des échantillons, de façon à obtenir une information spectrale uniquement associée à la composition biochimique intrinsèque du tissu.

MATERIELS ET METHODES

I. Echantillons tissulaires

Après résection chirurgicale de tumeurs coliques, les tissus prélevés sont transférés dans du DMEM (Dulbecco's Modified Eagle's Medium) supplémenté avec 200 UI/ml de pénicilline, 200 μ g/ml de streptomycine et 5 μ /ml de fungizone (Gibco, France). Suite à l'élimination des portions fibreuses et nécrotiques, le tissu cancéreux restant est émincé avec des ciseaux et centrifugé pendant 10 min à 800g. La préparation obtenue est ensuite injectée par voie sous-cutanée dans des souris « nude » (Charles River Laboratories, EU) à hauteur de 0,4 ml par spécimen murin. Les xénogreffes sont ensuite collectées une fois leur volume approchant 1 cm³, puis fixées au formaldéhyde et inclus en paraffine selon les protocoles standard de routine.

Huit échantillons de xénogreffes de carcinomes du colon ainsi préparés ont été utilisés dans cette étude. Des sections de 7 μ m d'épaisseur ont été coupées et montées sur des lames de ZnSe de 2 mm d'épaisseur pour chacun de ces échantillons (Crystal GmbH, Allemagne) en vue des acquisitions spectrales IR. Une section adjacente à celle utilisée pour les mesures IR a été colorée à l'HE pour examen histologique et attribution des clusters spectraux.

II. Acquisition des données spectrales IR

Les images spectrales IR ont été acquises à l'aide un système Spotlight 300 (fig. 26, Perkin-Elmer Life Sciences, Courtaboeuf, France) avec une taille de pixel de 6,25 μ m, sur la gamme spectrale 648-2000 cm⁻¹, une résolution spectrale de 8 cm⁻¹ et 2 accumulations par spectre. Pour chaque coupe, une zone ne contenant que de la paraffine (environ 30 000 spectres) a également été acquise pour être incluse dans la matrice d'interférences du modèle de correction EMSC. Avant chaque acquisition, un spectre de référence des fenêtres ZnSe est enregistré avec une résolution spectrale de 8 cm⁻¹ et 240 scans co-additionnés. Ce spectre est automatiquement soustrait de l'ensemble des données.



Les spectres de vapeur d'eau on été mesurés avec le détecteur Perkin-Elmer Spectrum One qui est une partie du système Spotlight 300. Pour cela, le compartiment échantillon a été tout d'abord vidangé avec de l'azote pour éliminer toutes traces d'eau et de dioxyde de carbone. Un spectre de bruit de fond a été mesuré dans cette condition. La vidange est ensuite stoppée de façon à laisser l'air ambiant diffuser à l'intérieur du compartiment. Durant cette augmentation progressive d'humidité, des mesures ont été enregistrées en continue jusqu'à atteindre un point de stabilisation. Cette expérimentation, menée plusieurs fois, a abouti à la collecte d'un total de 58 spectres.

III. Traitement des données

Ajustement de la gamme spectrale

Avant traitement, tous les spectres sont coupés sur une gamme spectrale s'étendant de 724 à 1720 cm⁻¹, considérée comme la plus informative pour les échantillons biologiques (Parker 1983).

EMSC

Le modèle de correction des données spectrales développé ici s'appuie sur le principe de l'EMSC, extension de la MSC introduit en 1983 pour le traitement des spectres procheinfrarouge (Martens 1983 ; Geladi 1985). Le principe général de l'EMSC est explicité en Annexe 4, p174.

Classification des données spectrales pour l'histologie IR

La classification des données spectrales issues de chaque coupe tissulaire a été réalisée par un algorithme KM. Le nombre de clusters a été déterminé de façon à retrouver l'histologie HE avec un minimum de clusters, chacun d'eux ne pouvant être attribué qu'à une seule structure histologique.

Plate-forme logiciel

Le traitement des données spectrales a été réalisé à l'aide du logiciel Matlab 2006a (Mathworks, EU). Les routines ont été en partie écrites dans notre laboratoire ainsi qu'au Center for Optical Diagnostics and Therapy de l'Erasmus Medical Center (Rotterdam, Hollande). La routine PCA est issue de la PLS Toolbox v1.5.3 (Eigenvector, EU).

RESULTATS

I. Mise au point d'un modèle de correction des données basé sur l'EMSC

Le modèle développé ici permet de minimiser la contribution spectrale de la paraffine, ainsi que celle de la vapeur d'eau atmosphérique et des effets de dispersion de la lumière IR. Pour une telle correction, il est nécessaire d'obtenir un signal représentatif des données spectrales à corriger ainsi que de modéliser les sources d'interférences des données spectrales issus de l'environnement (vapeur d'eau atmosphérique...) et de certains paramètres physiques des coupes tissulaires (épaisseur et topographie des échantillons, présence de paraffine...). Ainsi dans l'équation de l'EMSC :

 $s_i = a_i \,\mathbf{\hat{s}} + b_i \mathbf{I} + c_i \mathbf{P} + e_i \qquad (1)$

ŝ est l'estimation des données spectrales à corriger et correspond en pratique au spectre moyen de ces données.

I est la matrice d'interférences. En pratique, I correspond à la modélisation du signal IR de l'eau et de la paraffine.

• Modélisation du signal de la paraffine :

Les spectres IR de paraffine présentent une variabilité importante. Ainsi, pour une meilleure description de ce signal et donc une modélisation optimale de celui-ci dans l'EMSC, nous avons choisi d'utiliser les 8 premières composantes principales d'un jeu de données généré à partie d'images spectrales acquises sur des zones de paraffine pure, plutôt que d'utiliser un spectre moyen unique (fig. 27*A*). Un total de 255 000 spectres de paraffine issus de 8 coupes différentes a donc été inclus dans ce modèle. Les spectres de paraffine présentant un bruit de fond élevé (au-delà de 0.35 UA sur la gamme 1800-1900 cm⁻¹) ainsi que les spectres présentant un absorption maximum inférieure à 0.3 (signal faible) ou supérieure à 1 (saturation du détecteur) de la principale bande de paraffine (1460-1470 cm⁻¹) sont exclus du modèle EMSC.

Lorsque de la paraffine et du tissu sont en conjonction, l'intensité relative des bandes spectrales de la paraffine peut changer selon des modalités non couvertes par cette analyse en composantes principales. Ces caractéristiques spectrales pourraient être dues à des interactions moléculaires entre le tissu et la paraffine (Vrabie 2007). Afin de compenser ces variations, un ensemble de fonctions mathématiques Gaussiennes, calculées à partir de la décomposition des bandes spectrales de la paraffine, est ajouté à la matrice d'interférences (fig. 27*B*, les 6 premières courbes en partant du haut).



Figure 27. Modélisation des interférences pour le pré-traitement des spectres IR de coupes paraffinées. *A*, de haut en bas les 8 composantes principales du signal de la paraffine utilisé comme spectres d'interférences dans le modèle EMSC. *B*, de haut en bas, 6 courbes Gaussiennes additionnelles modélisant les bandes de la paraffine et les 2 composantes principales de spectres de vapeurs d'eau.

• Description du signal de l'eau :

Les deux premières composantes principales issues d'un ensemble de 58 spectres de vapeur d'eau atmosphérique sont ajoutées à l'ensemble de spectres interférents pour prendre en compte les variations possible d'humidité (fig. 27*B*, les deux dernières courbes en partant du haut), et ce malgré l'utilisation d'un système de purge.

Enfin, dans l'équation (1), **P** est un polynôme modélisant le bruit de fond et les effets de dispersion de la lumière IR. En pratique, un polynôme de 4ème ordre est inclus dans le modèle d'EMSC. Cette approche a été utilisée au lieu de celles décrites par Mohlenhoff *et al.*

(2005) ou par Kohler *et al.* (2008) car elle permet de diminuer les temps de calcul et s'est révélée efficace sur la gamme spectrale 724-1720 cm⁻¹ étudiée.

L'EMSC, en corrigeant la contribution IR de l'eau, de la paraffine et des effets de dispersion de la lumière, permet l'obtention d'un signal biochimique pertinent non-biaisé par les caractéristiques physiques des tissus analysés et par les interférences de l'environnement. L'EMSC va également permettre, sur la base des coefficients a_i et e_i, d'exclure les spectres indésirables pour l'analyse des données. Ces spectres présentent : i) un signal de paraffine uniquement et/ou un faible signal tissulaire, ii) une forme particulièrement déviante par rapport au spectre moyen des données spectrales corrigées, ou bien encore iii) un bruit de fond important.

En pratique, la détection de ces spectres indésirables se fait donc sur les coefficients du modèle EMSC. La répartition des coefficients a_i et des résidus e_i pour l'ensemble des spectres d'un image IR est représentée graphiquement dans la figure 28.

Les règles d'exclusion sont les suivantes :

- Le premier coefficient de l'EMSC a_i doit être suffisamment élevé pour permettre d'éliminer les spectres contenant majoritairement de la paraffine et/ou un faible signal tissulaire. Sur la figure 28A, les spectres présentant un coefficient a_i inférieur à 7 correspondent à ces spectres indésirables.
- Le coefficient e_i, une estimation de l'erreur du modèle EMSC, doit être plus petit qu'un seuil déterminé empiriquement. En se basant sur la figure 28*B*, ce seuil est dans notre cas de 0.2. Ceci permet l'exclusion des spectres bruités et/ou présentant une forme particulièrement déviantes.
- Les autres coefficients (b_i et c_i) doivent être égaux, pour chaque spectre, à + /- 3 fois la déviation standard par rapport à leur moyenne calculée sur l'ensemble des données. Ceci permet également d'exclure les spectres déviants.



II. Effet du modèle de correction sur les données spectrales

La figure 29 illustre l'effet du pré-traitement sur les spectres composant une image IR : ces spectres sont normalisés, le bruit de fond est corrigé, les interférences sont minimisées et les spectres aberrants (c'est-à-dire les spectres ne respectant pas les 3 règles) sont exclus. Si les pics à 1377 et 1464 cm⁻¹ sont encore présents, la variabilité due au signal de la paraffine est corrigée, autorisant ainsi la comparaison de données spectrales intra- et inter-échantillons.



Figure 29. Effet du pré-traitement EMSC sur les données spectrales de coupes tissulaires paraffinées. *A*, spectres bruts non traités (le spectre moyen apparaît en gras). *B*, spectres après correction par le modèle EMSC : le bruit de fond et les spectres aberrants ou ne contenant pas de signal tissulaire sont éliminés ; la variabilité due à la paraffine et à la vapeur d'eau est réduite.

Les variations visibles dans la figure 29*B* proviennent d'un bruit de fond résiduel et sont visuellement magnifiés par le processus d'EMSC. Ces données bruitées ne représentent qu'une proportion négligeable de l'ensemble de données et peuvent être éliminées par des réglages plus rigoureux des paramètres d'exclusion des spectres. Cependant, un compromis doit être réalisé entre les critères de qualité des données spectrales et la complétude des images IRPC.

La figure 30 illustre, quant à elle, l'effet du pré-traitement des données spectrales sur le résultat de leur classification par k-moyennes.



Figure 30. Effet du pré-traitement EMSC sur une classification des données par classification KM. A, histologie HE d'une coupe de xénogreffe de carcinome du colon humain. B, image IR pseudo-couleur (5 clusters) correspondante (coupe adjacente à A), sans pré-traitement EMSC des données. C, image IRPC (5 clusters) correspondante (coupe adjacente à A), après pré-traitement EMSC des données.

Sans pré-traitement (fig. 30*B*), la classification des donnés spectrales n'aboutit pas à une image comparable à l'histologie. Les spectres présentant un signal élevé de paraffine interfèrent avec le reste des données spectrales et chacun des clusters obtenus peut être attribué à plusieurs structures histologiques. Par exemple, le cluster jaune (fig. 30*B*) peut être

associé à 3 structures histologiques différentes : nécrose, stroma, et tissu tumoral. En revanche, après pré-traitement, les spectres contenant un fort signal de paraffine sont éliminés, le reste des données spectrales a été corrigé permettant ainsi d'obtenir une image correspondant à l'histologie. Dans ce cas, chacun des clusters est attribué à une seule structure histologique.

DISCUSSION

Le modèle de correction des données spectrales présenté dans cette partie permet pour la première fois, à notre connaissance, de réaliser un « déparaffinage numérique » d'échantillons tissulaires dans le cadre de leur étude par histologie IR. Le signal IR de la paraffine, particulièrement variable à l'échelle d'une coupe tissulaire, a pu être modélisé par une approche basée sur l'EMSC, en incluant dans la matrice des interférences les 8 premières composantes principales d'un large ensemble de spectres de paraffine pure. Cette modélisation ne prend pas en considération les interactions moléculaires entre la paraffine et les éléments tissulaires. Ces interactions induisent un léger décalage en nombre d'ondes des bandes d'absorption de la paraffine et/ou de légères variations dans leur intensité relative. L'ajout de fonctions gaussiennes dans la matrice d'interférences de l'EMSC a permis de modéliser efficacement ces variations. Ainsi, ce modèle de correction permet d'obtenir, directement sur des coupes paraffinées, des données spectrales spécifiques de la composition moléculaire intrinsèque du tissu. Ces données, non biaisées par le signal de la paraffine ou par les altérations moléculaires consécutives au déparaffinage chimique, permettent la construction d'images IRPC comparables à l'histologie HE, sans nécessité de marquages ni de solvants.

Cette méthode originale de pré-traitement de données pourrait permettre d'exploiter pleinement le potentiel de l'histologie IR en tant qu'outil d'investigation biomoléculaire des tissus, et non plus seulement comme un outil de reproduction automatique de l'histologie HE. Usuellement, l'histologie IR est appliquée sur des coupes cryogénisées en raison des problèmes inhérents à l'absorption de la paraffine dans l'IR ou engendrés par la procédure de déparaffinage chimique. Ces études, visant à une reproduction stricte de l'analyse histopathologique conventionnelle (Diem 2004 ; Lasch 2004 ; Bird 2008 ; Bird 2009), sont de plus réalisées après dessiccation des coupes tissulaires, pouvant nuire au maintien de la structure moléculaire des tissus, notamment au niveau de la conservation des structures secondaires des protéines (Van de Weert 2001), et interdisant tout étude immunochimique postérieure sur la coupe acquise en IR. Ainsi, l'étude IR de tissus paraffinés présente des avantages certains par rapport à l'utilisation de coupes congelées. Tout d'abord, l'inclusion dans la paraffine préserve la structure moléculaire des tissus et s'avère compatible avec l'immunohistochimie ainsi qu'avec les techniques moléculaires haut-débit (Rezaul 2008).

Surtout, les tissus inclus en paraffine représentent la principale source d'échantillons disponible dans les tumorothèques. Ceci offre l'opportunité de mener des études rétrospectives à grande échelle de façon à identifier et à valider de nouveaux marqueurs spectroscopiques à viser diagnostique, pronostique et prédictive.

Ainsi, le modèle de déparaffinage numérique présenté ici ouvre la voie pour des études utilisant l'histologie IR en tant que véritable outil d'investigation moléculaire, et non plus comme un outil visant uniquement la reproduction de l'histologie conventionnelle.

Chapitre 2

Caractérisation histopathologique de xénogreffes de carcinome du colon humain par histologie IR
INTRODUCTION

La caractérisation histopathologique de tissus est une étape clef du diagnostique des lésions tumorales et de leur évaluation pronostique et prédictive. Contrairement à l'histologie classique principalement basée sur l'analyse morphologique ou sur l'immunochimie ciblant spécifiquement un nombre limité de protéines, la spectroscopie vibrationnelle IR permet de sonder la composition biomoléculaire intrinsèque des tissus et d'en détecter de subtils changements. En se basant sur ces informations moléculaires, elle permet également d'éviter la variabilité, inhérente à l'opérateur, observée dans le cadre des processus classique d'évaluation morphologique et immunochimique. Pour différent types de tumeurs, l'analyse spectroscopique de tissus à permis la discrimination entre tissus normaux et tissus bénins ou malins (Krishna 2007; De Jong 2006; Steller 2006; Krishna 2006). Ce type d'approche démontre une des grandes forces de l'histologie IR, à savoir sa capacité à détecter, de manière automatique, la présence de lésions tumorales malignes. En 2004, Lasch et al. ont pour la première fois démontré que l'imagerie en microspectroscopie IR pouvait être utilisée en tant qu'outil de caractérisation histologique de coupe tissulaire, en exploitant les données spectrales par des méthodes d'analyse statistique multivariée, donnant alors naissance au concept d'histologie spectrale IR. La possibilité de caractériser ainsi différents types de lésions tumorales a par la suite été démontrée, notamment pour le diagnostic différentiel de différentes lésions tumorales de la peau (Ly 2007) ou la détection de micro-métastases ganglionnaires chez des patientes présentant un cancer du sein (Bird 2009). Mais en choisissant de reproduire l'histologie classique, les informations additionnelles, relatives à la composition biochimique des tissus, qui pourraient être présentes dans les données spectrales ne sont pas pleinement exploitées.

Par ailleurs, il est admis que de nouveaux outils sont nécessaires pour l'étude de l'hétérogénéité intra-tumorale et l'évaluation de son intérêt pronostique et prédictif. Ainsi, à partir de l'analyse d'un ensemble de xénogreffes de carcinomes de colon humain, l'étude présentée dans ce chapitre vise à une exploration approfondie des possibilités de l'histologie IR pour la caractérisation histopathologique de tissus, en mettant l'accent sur l'hétérogénéité tumorale. Ces xénogreffes ont été partiellement décrites auparavant par Guénot *et al.* (2006). De tels échantillons constituent un compromis entre les modèles d'études cellulaires et tissulaires, présentant la complexité de la tumeur originelle tout en offrant une liberté

expérimentale proche des lignées cellulaires. En effet, les xénogreffes de carcinomes du colon maintiennent leurs caractéristiques morphologiques et génétiques même après de multiples passages.

MATERIELS ET METHODES

I. Echantillons tissulaires

Vingt-deux xénogreffes de carcinomes du colon et de métastases hépatiques, préparées selon le protocole décrit dans le Chapitre I (p76) ont été incluses dans notre étude. Ces échantillons, stadifiés de T2N0M0 à T4N1M1, proviennent de 11 hommes et 7 femmes. Pour 4 de ces patients, 2 passages différents des xénogreffes sont inclus dans cet ensemble d'échantillons. Les caractéristiques cliniques et histologiques de ces xénogreffes sont présentées dans le tableau 7. Ces échantillons ont été fournis gracieusement par l'unité INSERM U-682 (Strasbourg, France).

Numéro échantillon	Numéro patient	Sexe	Localisation	Classification TNM	Nombre de passages	Ens. d'entrainement (E) ou de validation (V)
125	B680	м	CD	T2N0	13	Е
212	B756	М	CG	T4N0M1	13	Е
119	B849	М	Tr	T3N1	9	Е
126	B866	F	Tr	T4N2	5	Е
176	B868	F	CG	T2N1	5	v
107	B 870	М	CD	T2N2	9	v
166	B87 0				13	Y
164	B1123	F	CS	T4N0	1	Е
285	B1123				6	Е
241	B 1174	F	CG	T3N1	4	Е
230	B 1194	F	CD	T3N1	1	v
263	B1200	М	Caecum	T3N2	б	Е
194	B1206	М	CG	T3N2M1	2	v
276	B1206				10	V
251	B1209	М	CD	T4N2	2	Е
293	B1209				3	Е
274	B1213	F	MH	Métastase	6	Е
215	B 1 21 7	F	CD	T4N1M1	4	V
250	B1249	М	CG	T3N1	4	v
279	B1248	М	MH	Métastase	3	v
314	B1394	М	MH	Métastase	3	V
309	B1408	М	RS	T3N2M1	1	V

Tableau 8. Caractéristiques cliniques et histologiques des xénogreffes.M = Masculin, F = Féminin, CD = Colon droit, CG = Colon gauche, Tr = Colontransverse, MH = Métastase Hépatique, RS = Recto-sigmoïde, TNM = Histopathologicalclassification (American Joint Committee on Cancer, AJCC).

Une partie de ces xénogreffes, issues de 9 patients (11 échantillons), a été utilisée comme ensemble d'entrainement pour générer un modèle LDA prédictif de l'histologie. Les autres échantillons (9 patients, 11 échantillons) ont été inclus dans l'ensemble de validation pour évaluer la pertinence de ce modèle prédictif. Ces ensembles ont été construits de tel sorte que

les 2 groupes soient aussi homogènes que possible tant au niveau du genre des patients que de la localisation des tumeurs et de leurs stades TNM.

II. Acquisition des données spectrales IR

Les xénogreffes de tumeurs fixées au formaldéhyde et incluses dans la paraffine ont été coupés en sections de 7 μ m d'épaisseur et montées sur des lames de ZnSe (Crystal GmbH, Berlin, Allemagne). Aucun déparaffinage chimique n'a été effectué. Les coupes consécutives à celles utilisées pour les acquisitions IR ont été montées sur lames de verres standards pour être colorées à l'HE ou être utilisées en IHC. Les images spectrales IR ont été acquises à l'aide un système Spotlight 300 (fig. 26, Perkin-Elmer Life Sciences, Courtaboeuf, France) avec une taille de pixel de 6,25 μ m sur la gamme spectrale 648-2000 cm⁻¹, une résolution spectrale de 8 cm⁻¹ et 2 accumulations par spectre. Par ailleurs, pour chaque coupe, une zone contenant environ 30 000 spectres de paraffine a également été acquise dans les mêmes conditions que pour les tissus. Avant chaque acquisition, un spectre de référence des fenêtres ZnSe est enregistré avec une résolution spectrale de 8 cm⁻¹ et 240 scans co-additionnés. Ce spectre est automatiquement soustrait de l'ensemble des données.

Un total de 68 images spectrales a été collecté, contenant en moyenne 135 000 spectres par image, et couvrant au total une surface d'environ 9 mm² par coupe. Cet ensemble de données comprend plus de 9 millions de spectres.

III. Traitement des données

Correction des données spectrales

Le modèle de correction basée sur l'EMSC, présenté dans le Chapitre I (cf. p76), a été appliqué sur l'ensemble des images spectrales acquises. Les interférences de la paraffine sont modélisées à partir des spectres acquis sur des zones périphériques au tissu ne contenant que de la paraffine.

Classification des données spectrales pour l'histologie IR

Un algorithme KM (cf. p51) a été appliqué sur les images spectrales issues des échantillons de l'ensemble d'entrainement dans le but de déterminer une base de clusters spectraux représentatifs des différentes structures histologiques rencontrées dans ces échantillons. L'attribution histologique de chaque cluster est réalisée par comparaison avec les coupes

tissulaires colorées à l'HE. Le nombre de clusters a été choisi de telle sorte que chacun d'eux corresponde à une seule structure histologique, tout en autorisant qu'une structure histologique soit représentée par plusieurs clusters. La classification KM obtenue a été ensuite utilisée pour générer un modèle prédictif LDA. Préalablement à la classification LDA, la dimensionnalité du jeu de données est réduite par ACP pour faciliter les processus de calcul informatique.

Plate-forme logiciel

Voir Matériels et Méthodes partie I, p80.

IV. Immunohistochimie

Détection de l'expression de Ki67

L'expression de Ki67 a été analysée pour chacune des xénogreffes. L'immunomarquage a été réalisé par un protocole de coloration DAB de routine, à l'aide d'un anticorps anti-Ki67 MIB-1 (Novocastra, Newcastle, Angleterre).

Evaluation

Après révélation du marquage Ki67, 4 à 6 sites pour chaque xénogreffe ont été examinés par microscopie à transmission classique.



Figure 31. Exemple de marquage anti-Ki67 d'un échantillon de xénogreffe de carcinome de colon.

Le score immunohistochimique a été déterminé pour chaque site par le comptage d'au moins 700 cellules tumorales, et établi en tant que pourcentage de noyaux marqués. Les marquages nucléaires plus intenses que les réactions non spécifiques (bruit de fond) ont été considérés comme positifs.

RESULTATS

I. Construction des images IRPC de l'ensemble d'entrainement

Une classification KM sur la base de 12 clusters été appliquée aux images IR acquises à partir des échantillons de l'ensemble d'entrainement (tableau 8). Ce nombre de clusters, dans le cas des xénogreffes étudiées, permet une reproduction non équivoque de l'histologie HE (fig. 32), tout en mettant en évidence l'existence de 2 clusters associés au tissu tumoral.



Figure 32. Image IRPC d'une coupe de xénogreffe de carcinome du colon. *A*, image IRPC d'un échantillon (#212, tableau 7) construite par classification KM des données après pré-traitement EMSC ; les pixels colorés en blanc correspondent aux spectres exclus sur la base des règles listées dans le Chapitre I. *B*, histologie HE du même tissu.

Dans la figure 33, l'attribution histologique des clusters tel que déterminée par comparaison avec l'histologie HE est présentée, ainsi qu'un dendrogramme visualisant la distance spectrale entre les spectres moyens de chaque cluster. Les spectres sont présentés sous forme de dérivées seconde afin de permettre une meilleure mise en évidence des différences spectrales. Ce dendrogramme est issu du calcul des distances euclidiennes entre chaque spectre moyen des clusters sur la base de l'algorithme de Ward, communément utilisé dans les analyses par classification hiérarchique.



Figure 33. Signatures spectrales des clusters déterminés par classification KM. *A*, dendrogramme obtenu par classification hiérarchique des clusters spectraux. Ce dendrogramme permet l'évaluation des distances inter-spectrales (calculées à partir de l'algorithme de Ward). *B*, spectre moyen des clusters. Les dérivées secondes sont présentées pour améliorer la visualisation des différences spectrales ; les couleurs correspondent à celles utilisées dans les images IRPC. *C*, attribution histologique des 11 clusters spectraux. Les pourcentages indiquent la proportion de pixels appartenant à chacun des clusters. Ces pourcentages ont été calculés à partir des images IRPC de l'ensemble d'entrainement.

En se basant sur les images IRPC générées, l'attribution histologique des clusters spectraux peut être fait comme suit : 3 clusters pour le tissu tumoral, 4 pour le stroma, 4 pour le tissu nécrotique, et 1 pour le mucus secrété. Il est utile de rappeler que, dans le cas des xénogreffes étudiées, le stroma est originaire de l'hôte. Il est également essentiel de noté que les clusters 5, 9 et 10, attribués au tissu tumoral, peuvent être présents simultanément dans un même échantillon, révélant potentiellement une hétérogénéité biomoléculaire intratumorale.

Comme le montre la branche 1A du dendrogramme de la figure 33, les deux clusters les plus proches sont les clusters de tumeurs 5 et 9. Ils sont directement liés au cluster 4, attribué au stroma. Ces trois clusters représentent 62% des spectres de l'ensemble d'entrainement. De plus, on peut noter que le cluster de tumeur 10 est séparé des clusters tumoraux 5 et 9. Cette

distance est due à la signature spectrale du cluster 10 dans la gamme 1000-1200 cm⁻¹ révélant la présence de mucines (Khajehpour 2006) et confirmant ainsi l'attribution de ce cluster aux zones tumorales contenant des cellules sécrétrices de mucus. La branche 1B est composée par le cluster 3 (mucus sécrété) et des clusters 2 et 12, tout deux attribués au tissu nécrotique. Les spectres correspondant au tissu nécrotique appartiennent à différentes branches du dendrogramme (cluster 1 et 6 dans la branche 1A, clusters 2 et 12 dans la branche 1B), reflétant différent degrés de nécrose. Le dendrogramme est également composé par la branche 2, contenant uniquement les clusters 8 et 11, attribués au stroma.

Ces données mettent en évidence la présence de plusieurs clusters pour une même structure histologique. Ceci révèle potentiellement une hétérogénéité marquée de la composition biochimique de chacun des sous-types tissulaires, supportant l'hypothèse de l'existence d'une organisation tissulaire non détectable par l'histologie HE.

II. Construction et validation d'un modèle de prédiction de l'histologie IR

Les clusters spectraux définis par la classification KM des données issues de l'ensemble d'entrainement (tableau 8) ont ensuite été utilisés comme base de données dans le cadre du développement d'un modèle LDA (Ripley 1996) de prédiction de l'histologie IR. Ce modèle vise à classifier automatiquement les données spectrales issues d'échantillons inconnus selon les clusters précédemment établis par classification KM (fig. 33), permettant d'obtenir leur histologie IR en quelques minutes et sans nécessité de validation externe par comparaison avec l'histologie HE.

Pour des raisons de gestion de la capacité de calcul des systèmes informatiques utilisés, le modèle LDA est construit en utilisant 10% des spectres disponibles dans l'ensemble d'entrainement. Le modèle de correction EMSC a été appliqué à ces spectres et la dimension du jeu de données a été réduite en appliquant une ACP. Les 25 premières composantes principales on été prises en compte, représentant plus de 99% de la variance de ces données spectrales.

Dans un premier temps, la validation du modèle LDA consiste en une comparaison des résultats de la classification LDA avec ceux obtenues par classification KM. Tel que montré dans la figure 34, le modèle prédictif LDA permet l'obtention d'images IRPC similaires à celles obtenues par classification KM. En effet, seules des différences mineures apparaissent à

la marge des clusters. Les pixels correspondant présentent une probabilité inférieure à 0,75 d'appartenir au cluster auxquels ils ont été attribués. Ceci reflète le fait que ces pixels possèdent une signature spectrale intermédiaire. Les autres pixels présentent usuellement une probabilité d'attribution supérieure à 0,9.



Figure 34. Comparaison des images IRPC d'une coupe de xénogreffe de carcinome de colon obtenues par classification KM ou modèle de prédiction LDA. *A*, image IRPC d'un échantillon (#212, tableau 7) construite par classification KM des données après pré-traitement EMSC. *B*, image IRPC du même tissu construite à l'aide du modèle de prédiction LDA.

Dans un second temps, ce modèle LDA doit être également validé sur les échantillons de l'ensemble de validation (échantillons inconnus). Cette validation sur des données indépendantes est nécessaire pour prouver que le modèle LDA généré permet d'établir l'histologie IR d'échantillons autres que ceux de l'ensemble d'entrainement. La valeur prédictive du modèle LDA a donc été évaluée en comparant le résultat de la classification LDA de l'ensemble de validation (tableau 8), i.e. un ensemble d'échantillons « inconnus », avec l'histologie conventionnelle HE. Les données de la figure 35 illustrent que l'histologie IR obtenue par notre modèle de prédiction LDA est totalement corrélée avec l'histologie conventionnelle. Par ailleurs, l'ensemble de validation inclut différents passages d'une même xénogreffe (tableau 8). L'analyse IR de ces échantillons indique une stabilité de l'histologie IR entre les différents passages. En effet, les images IRPC montrent que la contribution relative des clusters de tumeur 5, 9 et 10 est similaire entre ces échantillons. Cette stabilité au

niveau de l'histologie spectrale confirme les résultats démontrés par Guénot *et al.* (2006), démontrant la stabilité des xénogreffes, non seulement au niveau de leur morphologie mais aussi au niveau de leurs altérations chromosomiques.



Figure 35. Images IRPC de deux échantillons de l'ensemble de validation obtenues avec le modèle de prédiction LDA. A, image IR pseudo-couleur de l'échantillon #250 (tableau 8). B, coupe tissulaire colorée par HE et correspondant à l'image présentée en A. C, image IRPC de l'échantillon #314 (tableau 7). D, coupe tissulaire colorée par HE et correspondant à l'image présentée en C. Echelle, 400 µm.

III. Analyse de la probabilité d'attribution des spectres dans le modèle

Parce que le modèle LDA développé se base sur une classification non-supervisée (KM) des données, les différences spectrales entre les clusters attribués à une même structure histologique pourraient ne pas être liées à un processus ou événement biologique. En effet,

l'algorithme de classification KM distribue les données spectrales en un nombre de clusters déterminé par l'utilisateur. La segmentation des données peut alors se basée sur des différences spectrales si fines qu'elles pourraient ne pas présenter de caractère significatif, la classification obtenue pouvant alors être basée sur du bruit de fond résiduel voir même être fortuite. Il apparaît donc nécessaire de démontrer que les populations spectrales de chaque cluster sont significativement différentes. Ceci peut être réalisé en étudiant les probabilités d'attribution des spectres à un cluster donné dans le modèle LDA. En effet, dans un tel modèle, la classification des données spectrales est réalisée à partir des probabilités d'attribution de chaque spectre aux différents clusters. Un spectre est ainsi associé au cluster pour lequel il présente la plus forte probabilité d'attribution. Ces probabilités peuvent être visualisées en construisant des images en fausse-couleurs pour chacun des clusters et en estimant la distribution de probabilité d'attribution des pixels (fig. 36).

Parce que les clusters 5 et 9 apparaissent comme étant les 2 clusters les plus proches (voir le dendrogramme, fig. 33), nous avons calculé les probabilités d'attribution de ces clusters dans 3 échantillons représentatifs : les échantillons #241 et #314, présentant principalement le cluster 5 ou 9 respectivement, ainsi que l'échantillon #250 contenant simultanément un pourcentage élevé de ces deux clusters.



Figure 36. Probabilités d'attributions des spectres aux clusters 5 et 9 dans les échantillons #241 (A, B, G), #314 (C, D, H) et #250 (E, F, I). A - F, images en fausse-couleurs présentant la probabilité d'attribution des données spectrales aux cluster 5 (A, C, E) ou 9 (B, D, F). Echelle, 400 µm.

Concernant les échantillons #241 et #314, les probabilités de distribution et les images en fausse-couleurs montrent que les probabilités d'attribution sont élevées (>0,9), indiquant que l'attribution des spectres est faite statistiquement de manière non-ambigüe entre les clusters 5 et 9. Pour l'échantillon #250, la probabilité que les spectres appartiennent aux clusters auxquels ils sont attribués sont plus faibles (0,5). Les images en fausse-couleurs montrent, pour cet échantillon, qu'il existe une décroissance progressive de la probabilité d'attribution au cluster 5 (ou 9) à partir du centre d'une zone associée au cluster 5 (ou 9) vers le centre d'une autre zone associée au cluster 9 (ou 5). Ceci met en évidence l'absence de transition nette entre les clusters 5 et 9 dans cet échantillon.

Les probabilités d'attribution élevées dans les échantillons #241 et #314 ainsi que la distribution de ces probabilités, particulièrement dans l'échantillon #250 prouvent que la classification des données spectrales obtenus dans cette étude n'est ni fortuite, ni basée sur un éventuel bruit de fond résiduel.

IV. Analyse spectrale de l'organisation des tissus

L'attribution de plusieurs clusters spectraux à une même structure histologique (par exemple clusters 5 et 9 pour le tissu tumoral) suggère qu'une organisation tissulaire non révélée par l'histologie HE peut être détectée dans les images IRPC de nos échantillons de xénogreffes. Cette organisation a été étudiée de manière quantitative en appliquant une méthode d'analyse originale des images IRPC obtenues. Cette approche se base sur la construction d'histogrammes illustrant la probabilité pour un pixel attribué à un cluster donné de trouver un pixel appartenant à un autre cluster dans son voisinage. Ces histogrammes sont calculés comme suit : pour chaque pixel, un recensement de l'attribution des pixels directement voisin est effectué. Si un pixel est attribué au même cluster que le pixel initial, l'événement est alors omis. Le résultat de cette analyse est normalisé et représenté dans une série d'histogrammes (fig. 37). Chaque histogramme donne la probabilité pour qu'un pixel attribué à un cluster 5 bordent le cluster. Par exemple, 32% (en moyenne) des pixels attribués au cluster 5 bordent le cluster 4, et 63% bordent le cluster 9. En revanche le cluster 5 est très rarement en contact avec des clusters autre que le 4 ou le 9. A partir de ces histogrammes, des « séquences morphologiques » peuvent être déterminées en fonction des règles suivantes :

• Le point de départ de la séquence correspond à une colonne d'un cluster donné (par exemple le cluster 4).

- On détermine le cluster qui a le plus de probabilité de border le cluster de départ (i.e, le cluster 5, en rouge).
- On se réfère alors à la colonne de ce dernier cluster, et on détermine quel nouveau cluster a le plus de probabilité de le border (i.e cluster 9, en vert foncé), comme précédemment. Si ce nouveau cluster correspond au cluster précédent (i.e cluster 5), on prend en compte le cluster ayant la seconde plus forte probabilité.
- On continue ainsi tant que l'on ne retombe pas sur un cluster déjà présent dans la séquence.



Figure 37. Organisation spatiale des clusters spectraux. Les pourcentages de pixels voisins (*neighboring clusters*) ont été calculés à partir des images construites par le modèle LDA et en considérant la totalité des échantillons analysés (ensemble d'entrainement et de validation). Les couleurs des histogrammes correspondent aux couleurs des clusters. Les pourcentages indiquant la proportion de pixels appartenant à chacun des clusters sont aussi rappelés (*clusters number and contribution*).

Par exemple, en partant du cluster 4, cluster attribué au stroma, cette analyse permet d'obtenir la séquence suivante : $4 \rightarrow 5 \rightarrow 9 \rightarrow 1 \rightarrow 12$.

En comparant les séquences obtenues, on observe qu'elles ont toutes en commun l'enchainement $4 \rightarrow 5 \rightarrow 9$. Considérant que le cluster 8, attribué au stroma, borde quasiexclusivement le cluster 4, la séquence complète peut être démarrée au cluster $8: 8 \rightarrow 4 \rightarrow 5$ $\rightarrow 9 \rightarrow 1 \rightarrow 12$. Cette séquence suggère que le stroma représenté par les clusters 8 et 4, ainsi que le cluster tumoral 5 ne sont jamais en contact direct avec les clusters attribués au tissu nécrotique 1 et 12. Les clusters 8, 4 et 5 sont alors toujours séparés des clusters 1 et 12 par le cluster 9, également attribué au tissu tumoral. Une autre séquence d'intérêt, bien que moins probable, est celle impliquant le cluster 10, attribué au tissu tumoral contenant du mucus, et le cluster 3, attribué au mucus. Cette séquence met en évidence le tissu tumoral sécréteur de mucines : $8 \rightarrow 4 \rightarrow 5 \rightarrow 9 \rightarrow 10 \rightarrow 3$.



V. Interprétation moléculaire des différences spectrales

Les caractéristiques spectrales des séquences comprenant les clusters de tumeurs prédominants (5 et 9) ainsi que les clusters 1 et 12 (attribués au tissu nécrotique) ont également été analysées. Pour cela, les différences spectrales entre les spectres moyens des clusters successifs $5 \rightarrow 9$, $9 \rightarrow 1$, $1 \rightarrow 12$, ainsi que leurs dérivées secondes ont été calculées (fig. 39).



Figure 39. Différences spectrales dans la principale séquence de clusters (fig. 37). *A*, spectres de différence entre les clusters ; cluster 1 minus 12 (*rouge*) ; cluster 9 minus 1 (*vert*) ; cluster 5 minus 9 (*bleu*). *B*, Dérivée seconde des spectres de différences. Les bandes de vibrations apparaissent négatives sur ces spectres dérivés. *C*, insert montrant les spectres dérivés magnifiés dans la gamme 750-1350 cm-1. Les lignes verticales renseignent sur l'attribution des vibrations à l'ARN (*rouge*), l'ADN, (*bleu*) ou aux protéines (*vert*).

Les différences entre les clusters 5 et 9 (spectre bleu) mettent en évidence des bandes à 996, 1116 et 1240 cm⁻¹ (lignes rouges verticales), associées à l'ARN (Ly 2008). Cette observation indique que le cluster 5 est plus riche en ARN, ceci traduisant une activité transcriptionnelle plus importante dans le cluster 5 que dans le cluster 9. Concernant l'ADN, les vibrations spécifiques qui lui sont attribuées en IR n'apparaissent pas sur ce spectre de différence, indiquant un contenu constant en ADN entre les deux clusters de tumeurs 5 et 9. Les différences entre le cluster 9 et 1(spectre vert) ainsi qu'entre les clusters 1 et 12 (spectre rouge) se situent quant à elles au niveau des bandes spécifiques à l'ADN (1055, 1088 et 1718 cm⁻¹) (Arakawa 2001). La décroissance des bandes attribuées à l'ADN entre ces clusters traduit une perte progressive d'ADN entre le cluster de tumeur 9 et le tissu nécrotique (cluster 1 et 12).

Enfin, si l'on compare les cluster de tumeurs 5, 9 et 10 (contenant du mucus), on s'aperçoit que ce dernier possède un signal très particulier, caractériser notamment par la présence

simultanée de 3 bandes intense à 1048, 1078 et 1124 cm⁻¹ (fig. 40) Le cluster 3, attribué au mucus secrété, présente entre 1000 et 1200 cm⁻¹, 3 bandes proches de celles du cluster 10 mais localisées exactement à 1044, 1076, et 1125 cm⁻¹. La soustraction du cluster 9 au cluster 10 (fig. 40,a) permet de retrouvé ce signal caractéristique du mucus tandis que la soustraction du cluster 3 au cluster 10 (fig. 40, b) permet de retrouver un spectre très similaire à celui du cluster 9. Ainsi la présence de mucus semble être caractérisable par la présence d'un marqueur spectral spécifique facilement visuable sans recours à la dérivation ou à des méthodes statistiques. La spécificité de ce marqueur spectrale sera exploitée dans le cadre des travaux présentés dans les parties 3 et 4.





VI. Corrélation avec l'IHC et les données cliniques

Afin de donner un sens à l'hétérogénéité spectrale détectée par histologie IR, plusieurs tentatives de corrélation entre les images IRPC des xénogreffes et de données obtenus par IHC ont été tentées. Les marquages réalisés ont principalement visé à la détection de l'expression de Ki-67 (marqueur de prolifération). Les résultats obtenus n'ont montré aucune corrélation significative entre ces marqueurs et l'hétérogénéité spectrale des sous-types tissulaires des xénogreffes. L'explication de l'échec de cette approche comparative entre IHC et histologie IR sera discutée dans la partie Discussion à suivre (cf. p109).

De la même manière, la comparaison entre les niveaux d'hétérogénéité tumorale tels que révélés par l'histologie IR et les stades TNM déterminés pour chacun des échantillons

(tableau 8) n'a aboutit à aucune corrélation significative sur cet ensemble d'échantillons. On notera toutefois que les 3 échantillons métastasiques (#314, #279, #274, tableau 8) ne présentent apparemment pas d'hétérogénéité intra-tumorale, les images IRPC de ces tissus n'affichant qu'un seul cluster de tumeur (5 ou 9) (fig. 35C).

DISCUSSION

Dans cette étude, l'histologie IR de coupes paraffinées de xénogreffes de carcinomes du colon a été établie par une classification KM des données spectrales après pré-traitements de cellesci par le modèle EMSC présenté dans le Chapitre I. Sur la base de cette classification, un modèle LDA a été construit, permettant la prédiction de l'histologie IR d'échantillons inconnus.

Le modèle d'histologie IR développé ici permet la mise la mise en évidence d'une hétérogénéité spectrale jamais rapportée auparavant. En effet, l'histologie IR est usuellement limitée à la reproduction de l'histologie HE, permettant alors l'identification de marqueurs spectraux associés uniquement aux sous-types de tissus (tumeur, stroma, nécrose...). Dans notre cas, 3 clusters spectraux ont pu être attribués au tissu tumoral : les clusters 5, 9 et 10, ce dernier correspondant aux cellules tumorales sécrétrices de mucus. Ces clusters peuvent être présents simultanément dans un même échantillon mais à des degrés variables.

La nature non-supervisée de la classification KM amène à s'interroger sur la significativité des différences spectrales entre les clusters assignés à une même structure histologique. Cette méthode de classification « dure» dépend en effet du nombre de clusters déterminé par l'utilisateur. Ainsi, la classification des données de notre modèle peut ne pas être basée sur des différences spectrales biologiquement significatives, mais pourrait être fortuite ou basé sur le bruit de fond. Afin d'évaluer la pertinence de cette classification, les probabilités d'attribution des spectres aux clusters 5 et 9 dans le modèle LDA ont été calculées pour 3 échantillons représentatifs. Les probabilités d'attributions élevées (> 0,9) des clusters 5 et 9 dans les échantillons #241 et #314 prouvent que la classification des données spectrales obtenues dans cette étude est bien basée sur des différences spectrales significatives. Si dans les échantillons présentant simultanément les clusters 5 et 9 (#250) ces probabilités d'attributions sont plus faibles (0,5<p<0,9), leurs distributions particulières tend également à appuyer la significativité des résultats obtenus. Cette distribution met également en lumière l'existence d'une population spectrale intermédiaire à celles des clusters 5 et 9 et illustre un changement progressif de la composition biochimique du tissu tumoral. Enfin, ces probabilités, associées au fait que l'acquisition répétée de données spectrales sur une même zone tissulaire permet d'obtenir des résultats reproductibles au niveau de l'histologie IR, prouvent que la classification obtenue n'est ni fortuite ni basée sur le bruit de fond. Ainsi, notre modèle d'histologie IR permet de mettre en évidence 3 grands groupes d'échantillons tumoraux selon la distribution des clusters 5 et 9 : un groupe d'échantillons présentant majoritairement le cluster 5 (rouge), un groupe présentant majoritairement le cluster 9 (vert), et un groupe mixte. La présence possible de ces deux clusters au sein d'un même échantillon, et les variations de leur distribution inter-échantillons sont en adéquation avec le concept d'hétérogénéité intra- et inter- tumorale des cancers coliques.

Le développement d'un traitement original des images IRPC a permis l'obtention d'informations supplémentaires sur l'hétérogénéité tissulaire, tout en appuyant la pertinence biologique de notre modèle. Ce traitement est basé sur la connectivité des pixels afin de quantifié la probabilité que deux clusters soit en contact direct. Ces probabilités de connexion reflètent l'organisation tissulaire de nos échantillons tel que révélée par l'histologie IR. Cette analyse d'image à permis de mettre en évidence deux grandes séquences d'organisation tissulaire :

- La séquence 8→ 4→ 5→ 9→ 1→ 12 révèle un gradient du stroma au tissu tumoral suivi de deux niveaux progressif de nécrose.
- La séquence 8→ 4→ 5→ 9→ 10→ 3 révèle un gradient du stroma au tissu tumoral suivi du tissu tumoral sécréteur de mucines puis du mucus secrété.

Il apparait peu probable que de telles séquences puissent être retrouvées sans que celles-ci ne trouvent leurs origines dans une hétérogénéité biomoléculaire significative des sous-structures tissulaires, indépendante de leur morphologie puisque non-visible par histologie classique. Cette organisation spectrale renforce donc la pertinence biologique de notre modèle d'histologie IR. Adjointe à l'analyse des différences spectrales entre les clusters 5 et 9, elle suggère que le cluster 5 pourrait être associé au tissu tumoral présentant la plus forte activité proliférative. En effet, les différences spectrales révèlent que le cluster 5 serait plus riche en ARN, suggérant une activité transcriptionnelle plus élevée dans le tissu tumoral qui lui est associé. La décroissance progressive des bandes d'ADN indique une perte progressive de matériel génétique entre les clusters de la séquence $9 \rightarrow 1 \rightarrow 12$, suggérant la présence de différents stades progressifs de nécrose. Par ailleurs, l'analyse IR de différents passages d'une même tumeur mène à des images IRPC présentant une distribution similaire de leurs clusters, indiquant que l'hétérogénéité intra-tumorale est maintenue entre des passages successifs. Un tel résultat est en concordance avec ceux publiées par Guénot *et al.* (2006) rapportant que les xénogreffes de carcinomes du colon maintiennent leurs caractéristiques génétiques et leur

morphologie tissulaire au cours du temps. Ces résultats nous indique que notre modèle d'histologie IR peut non seulement révéler une hétérogénéité intra- et inter tumorale, mais aussi détecter une véritable organisation tissulaire non détectable par histologie HE. Ceci démontre le potentiel de la spectroscopie IR pour fournir un nouveau type d'histologie potentiellement plus informative que les procédures d'analyses anatomopathologiques de routines, et présentant l'avantage d'être non subjective et automatisable tout en ne nécessitant ni solvants, ni marquages.

Le sens biologique de l'hétérogénéité spectrales intra- et inter- tumorale révélée par notre modèle d'histologie IR n'a pu être élucidée dans cette étude. Les différences spectrales entre les clusters 5 et 9, suggérant une différence de vitalité de ces 2 types de tissus tumoraux, nous ont poussé à analyser l'expression du marqueur de prolifération Ki67, exprimé dans les cellules en cours de division. Aucune corrélation n'a pu être mise en évidence entre l'expression de ce marqueur et la distribution des clusters 5 et 9. Ce résultat peut facilement être expliqué par les limites de cette approche comparative. Tout d'abord, un marquage IHC prononcé pour Ki67 reflète seulement le nombre de cellules en cours de division et non pas le temps requit pour l'accomplissement du cycle cellulaire (Hilska 2005). Par exemple, une tumeur avec des cycles cellulaires courts et une prolifération importante peut ne pas présenter de marquage IHC pour Ki67 en raison du faible nombre de cellule en cours de division à un instant t (Jansson 1997). Ceci expliquerait d'ailleurs les résultats contradictoires entre l'expression de Ki-67 et les variables cliniques, pathologiques et pronostiques dans les cancers coliques (Brown 2002; Hilska 2005). De plus, les approches immunohistochimiques et spectroscopiques s'avèrent diamétralement opposées. En effet, l'immunohistochimie vise à la détection d'un nombre très limité d'antigènes mettant ainsi en évidence des événements biomoléculaires très précis. A contrario, la spectroscopie IR permet la détection de changements biochimiques multiples, prenant en compte notamment l'expression des protéines, leur modifications post-traductionnelles et leur structure, ainsi que des changements dans le contenu en acides nucléiques. Ainsi, l'étude par immunohistochimie d'un changement moléculaire précis pouvant illustrer à lui seul les variations moléculaires multiples mises en évidence par l'histologie IR apparaît être une démarche peu pertinente.

De plus, aucune corrélation significative entre l'hétérogénéité intra- et inter-tumorale révélée par l'histologie IR et les données cliniques (stade TNM) n'a pu être établie. Il est toutefois intéressant de noter que les échantillons métastasiques ne présentent qu'un seul cluster tumoral (5 ou 9), et par conséquent une faible hétérogénéité intra-tumorale. Ceci est en accord

avec le fait que l'hétérogénéité intra-tumorale aurait tendance à décroitre dans les stades avancés de la progression d'une lésion, en raison de la sélection clonale (Losi 2005). Le fait de détecter une hétérogénéité tumorale non visible par histologie classique et non corrélable à la classification TNM ou à une approche réductionniste par IHC est en accord avec de récentes publications portant sur la carcinogenèse colorectale (Jass 2007 ; Zlobec 2009). En effet, aujourd'hui, la carcinogenèse colique est reconnue comme un processus pouvant emprunter différentes voies biomoléculaires. L'hétérogénéité tumorale qui en découle est un concept aujourd'hui bien établi en oncologie médicale et s'avère corroboré par le fait que des tumeurs histologiquement semblables peuvent présenter des pronostiques et des réponses aux traitements différents. La reconnaissance de cette hétérogénéité implique une caractérisation multiparamétrique aussi précise que possible de la tumeur, associant des données cliniques, protéiques et génétiques. Ainsi, l'acceptation du concept d'hétérogénéité tumorale a engendré la publication de travaux visant à établir de nouvelles classifications multifactorielles pour une meilleure prise en charge des cancers colorectaux (Jass 2007). En effet, une classification performante des tumeurs prenant en compte leur hétérogénéité moléculaire tendrait non seulement à une amélioration de leur évaluation pronostique, mais également à une meilleure personnalisation des traitements. L'identification de profils moléculaires définissant des sousgroupes que les facteurs classiques ignorent est d'ailleurs un des objectifs principaux de l'application des outils d'analyse moléculaire à haut-débit en cancérologie (Chabannon 2004). Notre modèle d'histologie IR permettant l'identification de profils tumoraux définis à partir de changements biochimiques multiples, englobant notamment informations génomiques et protéomiques, se rapprocherait ainsi, de manière conceptuelle, des nouvelles approches hautdébit dédiées à la détection de nouveaux marqueurs tumoraux, tout en conservant l'information spatiale propre aux techniques basées sur la microscopie. De plus, il est admis que la sensibilité de la spectroscopie IR est telle que de subtiles variations spectrales pourraient mettre en évidence des changements biomoléculaires précédant les changements morphologiques (Wang 2007 ; Kendall 2009). La combinaison de la localisation par histologie IR de sous-types de tissus tumoraux non détectable par histologie conventionnelle, avec des techniques de microdissection (Ericksson 2008) et de profilage haut-débit, en particulier protéomique (Rezaul 2008), semble donc être une approche prometteuse dans le cadre de l'identification de nouveaux marqueurs tumoraux tissulaires, et dans l'établissement de systèmes de classification prenant en compte l'hétérogénéité biomoléculaire intra- et intertumorale.

Chapitre 3

Caractérisation spectrale IR d'échantillons de mucus et de mucines

Introduction

Les résultats du chapitre précédent ont mis en évidence, outre l'hétérogénéité des tissus tumoraux analysés, le caractère très spécifique de la signature IR des structures tissulaires riches en mucines. Les mucines, protéines hautement glycosylés, constituent le composant majoritaire des mucus. Un grand nombre d'études, principalement basées sur l'immunochimie, rapportent l'intérêt potentiel de ces mucines en tant que marqueurs tumoraux. En effet, durant les processus de tumorogénèse et carcinogénèse, l'expression de ces glycoprotéines peut être altérée de manière qualitative et quantitative sans que l'on sache encore clairement si ces changements précèdent le développement tumoral ou en sont une conséquence (Jass 2001 ; Byrd 2004 ; Gaudier 2006). De tels changements pourraient non seulement altérer certaines fonctions cellulaires ou leur antigénicité mais pourraient également jouer un rôle dans les processus d'invasion et de dissémination des cellules tumorales. Malgré l'engouement certain vis à vis de l'étude de ces glycoprotéines, aucune analyse spectroscopique IR des mucines n'a été publiée. Pourtant, il a été démontré que cette technique possède un potentiel intéressant pour la caractérisation des glycoprotéines par sa capacité à apporter des informations qualitative et semi-quantitative sur leur glycosylation, leur sialylation, et leur structure (Khajehpour 2006). Ainsi, nous avons cherché à évaluer l'intérêt de la spectroscopie IR pour la caractérisation spectrale des mucines ainsi que pour mettre en évidence leurs altérations liées au développement tumoral. Dans ce chapitre, le travail présenté porte sur l'analyse IR in vitro d'extraits de mucines purifiées et de mucus, étape préliminaire nécessaire à la caractérisation spectrale *in situ* des mucus de tissus coliques sains et tumoraux, présentée dans le Chapitre 4 (cf. p126).

Matériels et méthodes

I. Origines des échantillons de mucus et de mucines

Mucines gastro-intestinales

Des échantillons de mucines gastro-intestinales mammifères purifiées et lyophilisées ont été utilisés en tant que produits de référence : 2 extraits de mucines gastriques de porc (mucine porcine type I et II, Sigma) et 1 extrait de mucines sous-maxillaires bovines (Sigma).

Mucus de kystes ovariens bénins lyophilisés

Cinq échantillons de mucus ovariens lyophilisés, extraits de kystes bénins de 5 patients différents, ont également été inclus dans cette étude. Ces mucus ont été précédemment caractérisés comme exprimant majoritairement MUC2 ou MUC5AC par dosage radioimmunologique (Bara 1979). Ces échantillons ont été gracieusement fournis par l'Unité INSERM U-482 (Paris).

Mucus de culture de cellules des voies respiratoires

Au cours de cette étude, nous avons également analysé des échantillons de mucus provenant de cultures cellulaires en interface air-liquide de cellules isolées d'une bronchiole morphologiquement saines prélevée dans une zone péri-tumorale et de cellules isolées d'un polype nasal bénin. Les mucus ont été prélevés à l'aide de capillaires et directement congelés à -20°C. Ces échantillons ont été fournis par l'Unité INSERM UMRS 903 (Reims).

Mucus coliques

Six échantillons de mucus coliques prélevés sur des pièces opératoires ont également été analysés. Trois de ces échantillons proviennent de patients présentant un adénocarcinome colorectal, les 3 autres provenant de patients présentant une diverticulite, pathologie inflammatoire non maligne. Directement après la résection chirurgicale, les mucus sont raclés à partir de la muqueuse (en amont de la tumeur pour les spécimens cancéreux) avec un racloir à cellules et congelés à -20°C. Ces échantillons ont été gracieusement fournis par l'Unité INSERM U-482 (Paris).

II. Acquisition et pré-traitement des données spectrales

Les spectres IR ont été collectés à l'aide du spectromètre IRTF Tensor 37 (Brüker Optics GmbH, Ettlingen, Allemagne) équipé de l'extension haut-débit HTS-XT. Cet accessoire est un lecteur de micro-plaques développé pour l'analyse spectroscopique IR d'ensemble d'échantillons de taille importante.



Figure 41. Le spectromètre IRTF Tensor 37 HTS-XT (gauche) et micro-plaques 96 puits (droite).

Echantillons lyophilisés

Une solution aqueuse de chacun des échantillons (30% m/vol) a été préparée. Pour chacune de ces solutions, 15 μ l ont été déposés sur une micro-plaque HTS-XT de 96 puits, avec 6 réplicas par solution. Les échantillons ainsi déposés sont séchés pendant une nuit à température ambiante.

Echantillons congelés

Ces échantillons ont été décongelés à température ambiante, homogénéisés de façon douce à l'aide d'un capillaire et directement déposés sur la micro-plaque HTS-XT. Les échantillons ainsi déposés sont séchés pendant une nuit à température ambiante.

Les spectres d'absorptions de chaque dépôt de mucus et de mucines ont été mesurés avec 64 scans co-additionnés et pré-traités par correction de leur ligne de base et normalisation vectorielle sur les bandes Amide I et II (1500-1700 cm⁻¹). Le spectre moyen a été alors calculé en vue de la comparaison des signatures spectrales de chacun des échantillons.

Plate-forme logiciel

La collection des données ainsi que le pré-traitement et l'évaluation des spectres ont été réalisés avec la version 6.5 du logiciel Brüker Optics User Software (Brüker Optics GmbH, Ettlingen, Allemagne).

Résultats

I. Analyse IR de mucines gastro-intestinales mammifères purifiées

La figure 42 présente la signature spectrale IR d'échantillons de mucines gastro-intestinales purifiées provenant de mammifères : 2 types d'extraits de mucines gastriques de porc et 1 extrait de mucines sous-maxillaires bovines (Roussel 1988).



Figure 42. Spectres IR moyens de mucines gastro-intestinales mammifères purifiées et lyophilisées. Gamme spectrale : 950-1800 cm⁻¹.

Les principales bandes spectrales observées correspondent :

- aux bandes Amides I et II à 1550 et 1656 cm⁻¹, attribuées aux vibrations des liaisons peptidiques (Barth 2007).
- aux bandes de glycosylation dans la gamme spectrale 1000-1200 cm⁻¹ (Khajehpour 2006).

Concernant les bandes Amides, on visualise des variations spectrales marquées entre les 3 échantillons, illustrant potentiellement des différences de structure secondaire.

Dans la région 1000-1200 cm⁻¹, les trois principaux pics d'absorption retrouvés (1044, 1076 et 1124 cm⁻¹) sont communs aux trois échantillons bien que l'on notera des variations au niveau de l'intensité de ces pics d'absorbances entre les échantillons. La mucine porcine de type I, par exemple, présente un pic à 1124 cm⁻¹ nettement plus intense que pour les 2 autres échantillons. Ces variations d'intensité relative des pics de la zone 1000-1200 cm⁻¹ illustrent un changement qualitatif de la glycosylation des échantillons. Enfin, le pic à 1728 cm⁻¹

détecté dans l'échantillon de mucine porcine de type I peut être attribué au groupement carboxyle de l'acide sialique (Khajehpour 2006).

II. Analyse IR d'échantillons de mucus

Les échantillons de mucines purifiées, malgré leurs origines différentes, présentent une signature spectrale typique, par la présence notamment d'un signal caractéristique de leur glycosylation dans la gamme 1000-1200 cm⁻¹. Les mucines étant décrites comme le constituant largement majoritaire du mucus (Aksoy 2004), nous avons par la suite évalué la contribution spectrale des mucines au sein d'échantillons de mucus. Différents types d'échantillon ont été analysés : mucus de kystes ovariens bénins, de voies respiratoires et mucus colorectaux provenant de pièces opératoires.

Mucus ovarien

Les échantillons de mucus ovarien extraits de kystes bénins et originaires de 5 patients différents, ont été caractérisés comme exprimant majoritairement soit MUC2, la mucine sécrétée majeure du colon, soit MUC5AC, mucine gastrique anormalement exprimée dans les lésions précancéreuses du colon et dans certains de ses adénocarcinomes (Forgue-Lafitte 2007).



Figure 43. Spectres IR moyens d'échantillons lyophilisés de mucus ovariens de kystes bénins. *A*, spectres présentés sur la gamme 950-1800 cm⁻¹. *B*, dérivées secondes (x-1) sur la gamme 1600-1700 cm⁻¹. Les échantillons ont été caractérisés comme contenant majoritairement soit MUC2 (MUC2+), soit MUC5AC (MUC5AC+).

Les signatures spectrales obtenues à partir de ces 5 échantillons présentent les mêmes pics principaux que les spectres de mucines purifiées (fig. 43*A*) :

- 1044, 1076 et 1125 cm⁻¹, attribués à la glycosylation des apomucines.
- 1550 et 1656 cm⁻¹, attribués aux vibrations des liaisons peptidiques.

On retrouve également des variations d'intensité de ces pics principaux selon les échantillons. Les différences spectrales obtenues notamment dans la région 1000-1200 cm⁻¹, région des sucres, et 1600-1700 cm⁻¹, sensible à la structure secondaire des protéines de nos échantillons (fig. 43*B*), ne permettent pas de discriminer les échantillons contenant majoritairement MUC2 de ceux contenant majoritairement MUC5AC. Enfin, aucun pic à 1728 cm⁻¹ n'est visualisable sur ce spectre.

Mucus de voies respiratoires

Ces mucus proviennent de cultures de cellules issues i) d'une bronchiole morphologiquement saine prélevée dans une zone péritumorale et ii) d'un polype nasal bénin (fig. 44).



Figure 44. Spectres IR moyens de mucus des voies respiratoires issus de cultures cellulaires. Gamme spectrale : 950-1800 cm⁻¹.

Une fois encore, la signature spectrale de ces échantillons présentent les mêmes pics principaux que les spectres de mucines purifiées (1044, 1076 et 1124 cm⁻¹ pour les bandes de glycosylation ; 1550 et 1656 cm⁻¹ pour les bandes Amides I et II). Ici, on notera surtout des

différences très importantes au niveau des bandes Amides I et II, reflétant des variations structurales marquées des protéines entre ces 2 échantillons.

Mucus colorectaux

Enfin, nous avons analysé des échantillons de mucus colorectaux directement prélevés sur des pièces opératoires d'adénocarcinomes et de diverticulites (fig. 45). En raison de la nature des opérations et des prélèvements (à l'aide d'un racloir à cellule, directement sur les pièces opératoires), ces mucus coliques sont contaminés par des traces de sang, de fèces ainsi que par des cellules exfoliées.



Figure 45. Spectres IR moyens de mucus colorectaux humains prélevés sur des pièces opératoires. *A*, Spectres présentés sur la gamme 950-1800 cm⁻¹. *B*, dérivées secondes (x-1) des spectres sur la gamme 1000-1200 cm⁻¹.

Un seul pic défini est retrouvé dans la zone 1000-1200 cm⁻¹, et se situe en moyenne à 1080 cm⁻¹. Les dérivées secondes des spectres sur cette zone mettent en évidence l'absence apparente des 3 bandes caractéristiques des mucines dans chacun de ces échantillons. Les différences spectrales entre ces échantillons ne permettent pas non plus de discriminer les mucus provenant d'adénocarcinomes et ceux provenant de diverticulite. Les absorptions principales des bandes Amides sont quant à elles retrouvées à 1548 et 1660 cm⁻¹.

Discussion

Le potentiel diagnostique et pronostique des altérations des mucines au cours de la séquence adénome-carcinome du colon a fait l'objet de nombreuses études (Jass 2001 ; Byrd 2004 ; Aksoy 2004 ; Mall 2008). En effet, il a été démontré à de nombreuses reprises que les mucines subissent des altérations qualitative et quantitative lors de la carcinogénèse du colon. Si ces études n'ont pas encore abouti à la découverte de marqueurs clinique potentiels, les résultats obtenus s'avèrent toutefois intéressants. Notamment, la détection de l'expression aberrante de MUC5AC dans des états pré-cancéreux et certains adénocarcinomes suggère le potentiel de l'analyse des mucines pour le diagnostic des lésions tumorales précoces (Forgue-Lafitte 2007). D'autres études suggèrent également le potentiel pronostique de ces altérations (Perçinel 2007 ; Wakatsuki 2008). Cependant, ces études ne portent majoritairement que sur la détection d'altérations ponctuelles en IHC, ceci pouvant par ailleurs expliquer certains résultats contradictoires quant à leurs valeurs diagnostiques ou pronostiques (Sarli 2008). De manière générale, les mucines restent difficilement étudiables, notamment par des techniques d'analyses structurales comme la cristallographie, en raison de leur taille très importante et de leur glycosylation complexe dissimulant une grande part de leur structure (Dekker 2002).

La spectroscopie IRTF est un outil intéressant pour l'étude des glycoprotéines, dont les mucines, fournissant des données sur leur glycosylation, sialylation ainsi que sur leurs structures primaires et secondaires (Khajehpour 2006). Il est donc étonnant que les mucines et les mucus n'aient pas encore fait l'objet d'étude IRTF comparative visant à la caractérisation de leurs signatures spectrales et aux variations de celles-ci selon la nature pathologique des échantillons tissulaires dont ils proviennent.

Dans cette partie nous avons donc cherché à caractériser le signal IR de mucines gastrointestinales purifiées de mammifères, de mucus ovariens humains présentant des profils d'expression des apomucines différents, de mucus respiratoires issus de cultures cellulaires et de mucus coliques directement prélevés sur pièces opératoire.

La comparaison de la signature spectrale des 3 échantillons de mucines gastro-intestinales purifiées nous à tout d'abord permis d'identifier des vibrations caractéristiques de ces glycoprotéines au niveau de la gamme spectrale 1000-1200 cm⁻¹. En effet, malgré leurs origines différentes, ces 3 échantillons présentent des bandes localisés précisément à 1044, 1076 et 1124 cm⁻¹. Ces bandes sont associées à la composition caractéristique des chaines

glycanniques des mucines (Khajehpour 2006). Cependant, si ces 3 bandes sont présentent dans les 3 échantillons, leurs intensités relatives s'avèrent variables. Ces données spectrales sont en accord avec les différentes études portant sur la glycosylation des mucines. En effet, la nature des oligosaccharides constituant les chaines latérales de carbohydrates est comparable entre les différentes mucines secrétées (Dekker 2002), expliquant ainsi la présence de 3 pics caractéristiques à 1044, 1076 et 1124 cm⁻¹. Les variations d'intensité relative observées entre ces 3 pics traduisent par ailleurs un changement quantitatif de leur composition en oligosaccharides. Ces variations peuvent être mises en parallèle avec la très grande hétérogénéité des chaines latérales au niveau de l'assemblage des oligosaccharides constitutifs des glycannes (Jass 2001; Dekker 2002; Brockhausen 2006), induisant des variations de proportions relatives de chacun d'eux entre les molécules de mucines. Au niveau des bandes Amides I et II, un maximum d'absorption est retrouvé aux mêmes nombres d'ondes pour les 3 échantillons : 1656 et 1550 cm⁻¹, respectivement. Les variations des bandes Amide I (1600-1700 cm⁻¹), Amide II (1500-1600 cm⁻¹) et Amide III (1230-1330 cm⁻¹) reflètent des variations de structure secondaire entre ces échantillons (Barth 2007). Ce résultat, suggérant la présence de polypeptides différents, est corrélé au fait que le profil d'expression des mucines est dépendant des organes et des différents types de cellules les constituant (Mall 2008). Ainsi cette analyse IRTF met en évidence une signature spectrale caractéristique des mucines, malgré l'hétérogénéité structurale qui leur est propre. On notera également une absorbance marquée à 1728 cm⁻¹ pour un des deux échantillons de mucines porcines. Cette absorbance, reflétant la présence des groupements carboxyliques des acides sialiques, met en évidence la forte sialylation de cet extrait de mucines purifiées en comparaison des autres échantillons étudiés (Khajehpour 2006).

Si les mucines sont largement décrites comme étant le composant principal des mucus, d'autres molécules, bien que très minoritaires, peuvent y être détectées. En effet, les mucus peuvent contenir des IgA, des lysozymes ou lactoferrines (Aksoy 2004) pouvant potentiellement présenter une absorption significative dans l'IR. L'analyse IRTF de mucus kystiques ovariens a permis l'obtention d'une signature spectrale similaire à celles des mucines purifiées. On retrouve les pics caractéristiques de leur glycosylation ainsi qu'un maximum d'absorbance des bandes Amides I et II situé à 1656 et 1550 cm⁻¹. Ces résultats confirment donc que le signal IR des mucus correspond très majoritairement au signal IR des mucines, et que les constituants mineurs du mucus présentent une contribution spectrale négligeable. Ainsi la spectroscopie IRTF permet l'étude de certaines caractéristiques

biochimiques des mucines (glycosylation, structure secondaire, sialylation) par analyse directe d'échantillons de mucus.

Les mucus ovariens analysés ont été précédemment caractérisés comme exprimant majoritairement soit MUC2, soit MUC5AC (Bara 1979). Cependant, en se basant sur leur signature spectrale, ces deux types d'échantillons n'ont pu être discriminés. En effet, les variations observées sur les bandes de glycosylation ne semblent pas être caractéristiques du profil d'expression de ces échantillons. Ceci n'est guère étonnant, la glycosylation des mucines dépendant surtout de l'activité des glycosyl transférases qui est spécifique des organes et des cellules, et non pas des apomucines (Mall 2008). Les bandes Amide I et II présentent des variations significatives dans l'absorbance de leurs différentes composantes (accentuées par dérivation secondaire des spectres, fig. 43B), reflétant des variations de structure secondaire des mucines entre ces échantillons, mais sans corrélation avec leur profil d'expression. Ainsi, des mucus présentant un profil d'expression de mucines équivalent semble contenir des proportions différentes d'acide aminés impliqués dans des structures de type hélice- α , feuillet- et coude- β ou « random coil ». Cependant, ces résultats se doivent d'être pondérés de par la nature des échantillons analysés, i.e. des mucus lyophilisés selon des protocoles non implémentés pour la spectroscopie IR. En effet, dans le cas d'échantillons lyophilisés, il a été démontré que certaines variations spectrales des bandes Amides peuvent être attribuées non pas à des variations de structure secondaire mais à des artefacts dus à l'élimination de l'eau (van de Weert 2001).

L'analyse de mucus issu de cultures cellulaires nous a permis de comparer des mucus provenant de cellules prélevées chez des patients atteints de pathologies tumorales bénignes et malignes. En effet, un premier échantillon est issu de cellules de bronchiole morphologiquement saine mais prélever dans une zone péri-tumorale, tandis que le second échantillon provient d'une culture de cellules de polype nasal bénin. Ces deux échantillons présentent des différences spectrales très importantes sur les bandes Amides I et II, sensibles à leur structure protéique. Le très faible nombre d'échantillons ne permet pas de tirer de réelles conclusions sur ces résultats. Cependant si ceux-ci pouvaient être confirmés par l'analyse d'un nombre satisfaisant d'échantillons, ils pourraient témoigner d'une altération de structure secondaire des mucines dans une pathologie bénigne, ou au contraire suggérer la mise en évidence d'altérations structurales des mucines en relation avec la présence périphérique de tissu tumoral invasif.

Enfin, nous avons cherché à évaluer le potentiel de l'analyse de mucus pour la discrimination entre tissus coliques présentant une pathologie de type tumorale (adénocarcinomes colorectaux) ou inflammatoire (diverticulites). Cependant, ces mucus, prélevés directement sur des pièces opératoires, se sont avérés physiquement altérés par les méthodes de prélèvements appliquées. Ces échantillons présentent notamment des traces de fèces, de sang ainsi qu'une contamination par des cellules exfoliées lors du raclage de la muqueuse des pièces opératoires. Ces contaminations apparaissent comme masquant le signal des mucines. En effet, aucune de leur bandes caractéristiques n'est retrouvée dans ces échantillons, y compris suite à l'analyse des composantes spectrales de la bandes 1000-1200 cm⁻¹ par dérivation seconde des spectres.
Chapitre 4

Histologie IR des mucus secrétés pour la caractérisation d'adénocarcinomes du colon

INTRODUCTION

La caractérisation spectrale IR des mucines et des mucus présentée dans le chapitre précèdent a mis en évidence la possibilité d'évaluer directement certaines caractéristiques biochimiques des mucines (structure, sialylation et glycosylation des apomucines) par analyse IR des mucus. Ces caractéristiques ont été décrites comme pouvant être altérées lors de la tumorogénèse colique. En effet, une sur- ou sous-expression de MUC2, la mucine prépondérante dans cet organe, a été décrite selon le caractère muqueux ou non des adénocarcinomes (Gaudier 2006). L'apparition de MUC5AC, une mucine gastrique non exprimée dans le colon normal, a également été rapportée non seulement dans certains adénocarcinomes mais également dans de fréquents cas de lésions précancéreuses (Forgue-Lafitte 2007). De plus, indépendamment du profil d'expression des apomucines, l'altération des acides sialiques et la modification des chaines latérales d'oligosaccharides ont également été démontrées (Jass 2001; Byrd 2004). Par ailleurs, on notera qu'il n'existe pas actuellement de données sur d'éventuelles altérations de la structure secondaire des mucines, information accessible en spectroscopie IR. .

Prenant en compte le rôle potentiel des mucines en tant que marqueurs de la carcinogenèse colique, l'étude présentée consiste en une analyse *in-situ* des mucus sécrétés par histologie IR. L'objectif est la caractérisation spectrale IR de tissus coliques sains et tumoraux en vu de l'évaluation du potentiel des marqueurs spectraux associés aux mucines pour la discrimination de ces deux types d'échantillons.

MATERIELS ET METHODES

I. Echantillons tissulaires

Seize échantillons de tissus coliques ont été inclus dans cette étude. Neuf de ces échantillons sont originaires de patients présentant un adénocarcinome muqueux (tableau 8), les 7 autres provenant d'individus sains. Les échantillons d'adénocarcinomes ont été préalablement classifiés de T2N0 à T3N1.

Sex	Age	Description	TNM
М	72	Lieberkühnian colon adenocarcinoma with mucinous colloide	T3, N0
F	64	Lieberkühnian colon adenocarcinoma with mucinous colloide	T2, N0
F	71	Lieberkühnian colon adenocarcinoma with mucinous colloide	T2, N1, N2
М	74	Lieberkühnian colon adenocarcinoma with mucinous colloide	T3, N0
М	68	Colon adenocarcinoma with mucosal colloide	T3, N1
М	74	Lieberkühnian colon adenocarcinoma with mucinous colloide	T3,N0
F	69	Lieberkühnian colon adenocarcinoma with mucinous colloide	T3, N1
М	45	Lieberkühnian colon adenocarcinoma with mucinous colloide	T3, N0
М	82	Lieberkühnian colon adenocarcinoma with mucinous colloide	T2, N1

Tableau 9. Caractéristiques histologiques des adénocarcinomes coliques étudiés. *M*, masculin. *F*, féminin. *TNM*, classification histopathologique.

Les échantillons de parois de colons sains et tumoraux enroulés sur eux-mêmes (« Swiss roll », Bara 1996) ont été fixés à l'alcool et inclus dans la paraffine. Ces blocs ont par la suite été coupés en section de 8 μ m d'épaisseur et montées sur des lames de CaF₂ de 2 mm d'épaisseur (Crystal GmbH, Berlin, Allemagne). Les 2 coupes précédentes et consécutives à celle utilisée pour les acquisitions IR ont été montées sur lames de verres standards pour être colorées à l'HE (fig. 44) ou être utilisées en IHC. L'avantage de la disposition des tissus en « Swiss roll » est de permettre l'examen de la longueur entière de l'échantillon pour chacune des coupes.



tumoral (droite) en « Swiss roll » colorées à l'HE.

II. Acquisition des données spectrales IR

Comme précédemment, aucun déparaffinage chimique n'a été effectué préalablement à l'acquisition spectrale des coupes tissulaires. Les images spectrales IR ont été acquises à l'aide du système Spotlight 300 (fig. 26, Perkin-Elmer Life Sciences, Courtaboeuf, France) avec une résolution spatiale de 6,25 μ m sur la gamme spectrale 950-1750 cm⁻¹, une résolution spatiale de 8 cm⁻¹ et 2 accumulations par spectre. Par ailleurs, pour chaque coupe, une zone contenant environ 30 000 spectres de paraffine a également été acquise dans les mêmes conditions que pour les tissus. Avant chaque acquisition, un spectre de référence des fenêtres CaF₂ fut enregistré avec une résolution spectrale de 8 cm⁻¹ et 240 scans co-additionnés, et automatiquement soustrait de l'ensemble des données.

Environ 4 images IR ont été acquises par échantillon, couvrant une surface tissulaire moyenne de 2mm² par image. Cet ensemble de données comprend plus de 6 millions de spectres.

III. Traitement des données

Correction des données spectrales

Le modèle de correction basée sur l'EMSC, présenté dans le Chapitre I (cf. p76), a été appliqué sur l'ensemble des images spectrales inclus dans cette étude. Les interférences de la paraffine sont modélisées à partir des spectres collectés dans des zones périphériques aux tissus et contenant uniquement de la paraffine.

Classification des données spectrales pour l'histologie IR

Un algorithme KM a été appliqué simultanément à toutes les images spectrales issues des échantillons de tissus coliques sains et tumoraux. Cette classification a été comparée avec les

coupes colorées à l'HE afin de réaliser l'attribution histologique de chaque cluster. Le nombre de cluster a été défini de sorte que la classification des données spectrales permette la représentation des tissus sécréteurs en fonction de leur concentration en mucines. Les spectres ainsi attribués exclusivement aux mucus ont ensuite été isolés et à nouveau classifiés par KM. Ce procédé, que l'on nommera sous-clustering, permet d'étudier l'hétérogénéité d'une population spectrale attribuée à une structure donnée, en évitant tout risque d'interférences avec les spectres associés à d'autres structures tissulaires pendant le processus de classification.

Plate-forme logiciel

Voir Chapitre I, Matériels et Méthodes, p80.

IV. Détection de l'expression IHC de MUC2 et MUC5AC

L'immunohistomarquage de ces mucines a été réalisé par un protocole de coloration DAB (Di-Aminobenzidine) de routine, à l'aide d'un anticorps monoclonal anti-M1/MUC5AC et d'un anticorps polyclonal anti-MUC2, précédemment décrits par Forgue-Lafitte *et al.* (2007). Un score IHC a été attribué à chaque zone tissulaire correspondant aux acquisitions IR. Ce score prend en compte l'intensité et l'étendue des marquages MUC2 et MUC5AC. Ces scores, allant de 0 à 3 sont définis comme suit : 0 = pas de marquage, 1= marquage localisé de faible intensité ; 2= marquage étendu d'intensité modérée ; 3= marquage étendu de forte intensité.

V. Analyse statistique

Les analyses statistiques de la distribution des sous-clusters a été menée à l'aide du test de Mann-Whitney. Les résultats sont considérés comme statistiquement significatifs pour P < 0,1.

RESULTATS

I. Histologie IR non-supervisée de tissus coliques sains et d'adénocarcinomes muqueux

Nous avons tout d'abord effectué une analyse comparative des populations spectrales issues des échantillons coliques sains et tumoraux. Pour cela, l'ensemble des données acquises à partir des échantillons paraffinés de tissus sains et d'adénocarcinomes muqueux a été classifié par KM sur la gamme spectrale 950-1750 cm⁻¹, reconnue comme la plus informative pour l'analyse d'échantillons biologiques montés sur lames de CaF₂. Tel qu'illustré dans la figure 47*A*, les images pseudo-couleurs sont construites en utilisant un nombre variable de clusters (2, 4 (non montré) ou 6). Ceci afin de mettre en évidence les principales populations spectrales et de localiser les clusters attribuables aux différentes structures histologiques, avec une attention particulière pour le mucus sécrété. Sont également présentés une photographie des coupes tissulaires colorées par HE (fig. 47*A*) correspondantes aux images IRPC, ainsi que les dérivées secondes des spectres moyens de chaque cluster (fig. 47*B*-*C*).

Chapitre 4



Figure 47. Images IRPC et signatures IR des clusters spectraux de coupes tissulaires de colons sains et d'adénocarcinomes. *A*, images IRPC construites par classification KM et histologie HE des coupes adjacentes à celles utilisées pour les acquisitions spectrales ; échelle, 200 μ m. *B*, dérivées secondes (x-1) des spectres moyens des clusters correspondant aux images construites sur la base de 2 clusters. *C*, dérivées secondes (x-1) des spectres moyens des clusters correspondant aux images construites correspondant aux images construites sur la base de 2 clusters. *L*es couleurs des spectres correspondent aux couleurs utilisées dans les images IRPC.

Concernant la classification sur la base de 2 clusters (fig. 47*A-B*), le cluster 1/2 (vert foncé) présente les 3 bandes caractéristiques des chaînes glycanniques des mucines (1044, 1076 et 1125 cm⁻¹). Par contre, le cluster 2/2 (vert clair) ne présente pas simultanément ces 3 bandes caractéristiques mais affiche des bandes à 1028 et 1080 cm⁻¹ attribuées au glycogène (Wong 1991), à 1055 cm-1 attribuée à l'ADN (Wolthuis 2008) et à 1200, 1280 et 1335 cm⁻¹ associées au collagène (Lasch 2004). Ces caractéristiques spectrales suggèrent donc que contrairement au cluster 2/2, le cluster 1/2 peut être associé aux zones riches en mucines incluant le mucus secrété et les structures cellulaires sécrétrices. Ces résultats ont été confirmés en comparant les images IRPC à l'histologie conventionnelle HE.

Sur la base d'une classification des données en 6 clusters, le signal spécifique des mucines permet de discriminer efficacement le mucus secrété du mucus intracellulaire. Comme le montre la figure 47*C*, le spectre moyen du cluster 1/6 (bleu ciel) présente la plus forte absorption au niveau des bandes caractéristiques des chaines glycanniques des mucines (1044, 1076 et 1125 cm⁻¹), sans présenter de bandes attribuées aux acides nucléiques ou au collagène. Ceci suggère que le cluster 1/6 est représentatif du mucus secrété que l'on retrouve dans la lumière intestinale et dans la partie centrale des cryptes du colon normal, ou dans les structures glandulaires des tissus tumoraux. Les spectres moyens des clusters 2/6 et 3/6 (rose et noir, respectivement) présentent également les 3 bandes caractéristiques des mucines, mais de manière moins intense que le cluster 1/6. De plus, on notera que le cluster 3/6 présente une composante spectrale plus importante à 1055 cm⁻¹ (ADN) que les clusters 1/6 et 2/6. Le spectre moyen des 3 autres clusters (4/6, 5/6 et 6/6) ne présentent pas ces 3 bandes caractéristiques des mucines d'onde attribués au glycogène (1028 et 1080 cm⁻¹), à l'ADN (1055 cm⁻¹) et au collagène (1200, 1280, 1335 cm⁻¹). Ces caractéristiques spectrales impliquent que les clusters 4/6, 5/6 et 6/6 correspondent

aux structures histologiques dépourvues de mucus, comme confirmé par comparaison des images IRPC avec l'histologie classique.

L'organisation spatiale des clusters 1/6, 2/6 et 3/6 a également été analysée. Ces 3 clusters présentent un signal de mucines décroissant de manière progressive et sont organisés selon une séquence unique. En effet, ces clusters sont distribués de la manière suivante : le cluster 1/6 borde exclusivement le cluster 2/6, qui lui-même borde le cluster 3/6. Les clusters 4/6, 5/6 quant à eux, sont uniquement en contact avec le cluster 3/6. Enfin, le cluster 6/6 borde uniquement les clusters 4/6 et 5/6. Si l'on prend en compte la structure histologique des cryptes dans le tissu colique normal (fig. 47A) ainsi que les signatures spectrales de chaque cluster (fig. 47*C*), cette organisation spatiale suggère que le cluster 1/6 correspond effectivement au mucus sécrété alors que les clusters 2/6 et 3/6 peuvent être attribués respectivement aux parties apicale et basale des cellules composant les structures sécrétrices (fig. 48).



Cette organisation et attribution des clusters peuvent être retrouvées dans les adénocarcinomes malgré la forte altération de l'architecture morphologique de ces tissus.

II. Analyse de l'hétérogénéité spectrale des mucus et discrimination des échantillons

Nous avons démontré, dans les chapitres II et III (cf. p89 et p113), la possibilité de détecter l'hétérogénéité biochimique de structures histologiques par histologie IR, tout comme celle d'étudier certaines caractéristiques biochimiques des mucines par analyse IR des mucus. Le mucus est une mixture hétérogène principalement constitué de diverses mucines, connues pour être altérées durant la tumorogénèse colique. En conséquence, nous avons par la suite évalué le potentiel de l'histologie IR pour la discrimination des échantillons coliques sains et tumoraux, sur la base de l'hétérogénéité spectrale des mucus. Pour cela, une classification KM a été appliquée exclusivement aux spectres précédemment attribués au mucus sécrété (cluster 1/6, fig.47). Cette démarche originale que nous avons nommé « sous-clustering » permet d'étudier aisément l'hétérogénéité d'une population spectrale attribuée à une structure histologique donnée en évitant d'éventuelles interférences avec les spectres associés aux autres structures histologique pendant le processus de classification. Ainsi, cette démarche permet d'obtenir un « marquage spectral » du mucus représentatif de l'hétérogénéité biochimique de son contenu en mucines.

La figure 49 propose une illustration représentative des images IRPC obtenues par sousclustering des données spectrales attribuées au mucus secrété, pour différents nombres de sous-clusters. Les pourcentages de pixels traduisent les surfaces relatives occupées par chacun des sous-clusters dans chacune des images IRPC présentées.

Chapitre 4



Quelque soit le nombre de sous-clusters testé (de 2 à 5), chacun d'eux peut être retrouvé dans les deux types de tissus, sans séquence spatiale particulière et indépendamment de la localisation des mucus, i.e. dans la lumière de l'intestin ou au sein des cryptes du colon normal, ou encore dans les structures glandulaires des adénocarcinomes. Cependant on peut constater, par simple inspection visuelle des images IRPC et comparaison des pourcentages de pixels attribués à chaque sous-cluster que la distribution de ces derniers varie entre échantillons sains et tumoraux. Ces différences de distribution pourraient refléter des variations biochimiques du mucus entre ces deux types d'échantillons. Par conséquent, nous avons analysé la distribution des sous-clusters en calculant le pourcentage de pixels de chacun d'eux dans chacun des échantillons, c'est-à-dire la surface relative occupée par chaque sous-cluster du mucus entre l'hétérogénéité spectrale du mucus et le caractère sain ou tumoral des échantillons de tissus coliques. Pour l'ensemble des échantillons, les pourcentages de pixels des sous-clusters sont comparés de façon à évaluer le potentiel discriminant de cette approche (fig. 50).



Figure 50. Pourcentages relatifs des sous-clusters, en fonction du nombre de sousclusters fixé lors de la classification KM. Gamme spectrale : 950-1750 cm⁻¹. *Triangles bleus*, colon normal ; *cercles rouges*, adénocarcinomes. *Barres bleus*, pourcentages moyens des sous-clusters dans les échantillons de colon normaux. *Barres rouges*, pourcentages moyens des sous-clusters dans les échantillons d'adénocarcinomes. *, 0.05 < P < 0.1; **, 0.01 < P < 0.05; ***, P < 0.01; NS, non significatif. La classification des spectres attribués au mucus à l'aide de 2 sous-clusters met en évidence une différence significative (P < 0,01) de la distribution moyenne des sous-clusters selon les échantillons. Le sous-cluster 1/2 est présent en moyenne à une hauteur de 60,5% dans les tissus tumoraux contre 22,6% dans les tissus sains. Cette différence permet de discriminer les 7 échantillons sains de 7 des 9 échantillons tumoraux, deux d'entre eux présentant une distribution similaire à celle retrouvée dans les échantillons sains.

Suite à une classification à l'aide de 3 sous-clusters, 2 d'entre eux affichent une différence significative (P < 0.05) de leur distribution moyenne selon le type d'échantillons. En particulier, le sous-cluster 1/3 apparaît en moyenne significativement plus présent dans les tissus normaux que dans les tissus tumoraux (39,8% et 19,4% respectivement ; P < 0.01), les 2 populations d'échantillons ne présentant, dans ce cas, qu'un faible recouvrement. Le sous-cluster 2/3 apparaît quant à lui moins présent dans les tissus normaux en comparaison des tissus tumoraux (34,5% et 54,9% respectivement ; P < 0.1). Enfin, le sous-cluster 3/3 est retrouvé dans des proportions similaires dans les deux types d'échantillons (25,6% et 29,7%), mais avec une déviation plus importante dans les tissus tumoraux (25%) que dans les tissus sains (8,9%).

Concernant les classifications par 4 et 5 sous-clusters, même si la plupart des sous-clusters obtenus présentent une différence de leur distribution moyenne selon le type d'échantillon, aucune de ces variations n'affichent une significativité statistique (0,01 < P < 0,05) et un pouvoir discriminant égal à celui observé pour les sous-clusters 1/2 ou 1/3.

L'analyse de la distribution des populations spectrales après classification des données à l'aide de 2 ou 3 sous-clusters permet donc une discrimination partielle mais statistiquement significative des échantillons sains et tumoraux étudiés. La classification en 3 sous-clusters permet un recouvrement minimal des deux types d'échantillons. Afin d'identifier l'origine de cette distribution caractéristique, nous avons comparé les différences spectrales existantes entre le sous-cluster 1/3 (significativement plus présent dans les tissus normaux que dans les tissus tumoraux) et le sous-cluster 2/3 (significativement plus présent dans les tissus tumoraux que dans les tissus normaux) (fig. 51).



L'inspection des dérivées secondes des sous-clusters 1/3 et 2/3 permet de relevé des différences marquées au niveau des gammes spectrales spécifiques 1000-1200 cm⁻¹, 1500-1700 cm⁻¹ et 1700-1750 cm⁻¹ respectivement attribuées à la glycosylation, la structure secondaire et lasialylation des mucines. Afin d'étudier de manière approfondie ces différences spectrales et d'en évaluer le potentiel discriminant, nous avons analysé les spectres IR attribués au mucus de nos échantillons par un sous-clustering se basant sur ces gammes spectrales spécifiques uniquement. Cette approche permet de classer nos données IR en fonction de la composition en carbohydrates, de la structure secondaire ou de la sialylation des mucines, permettant alors d'obtenir des cartographies chimiques multivariées des mucus sur la base de ces caractéristiques biochimiques (fig. 52).



Figure 52. Exemples d'images IRPC de mucus d'un adénocarcinome colique, après classification KM en 3 sous-clusters. *A*, gamme spectrale 950-1750 cm⁻¹. *B*, gamme spectrale 1000-1200 cm⁻¹ (glycosylation). *C*, gamme spectrale 1700-17500 cm⁻¹ (sialylation). *D*, gamme spectrale 1600-1700 cm⁻¹ (Amide I).

L'utilisation de gammes spectrales spécifiques permet d'obtenir des résultats sensiblement différents au niveau de la distribution des sous-clusters. Ainsi la figure 53 présente les pourcentages de pixels attribués à chacun des sous-clusters obtenus pour chaque échantillon, en fonction des 3 gammes spectrales étudiées correspondant aux bandes de glycosylation (fig. 53A), de sialylation (fig. 53B) et Amide I (fig. 53C) ainsi que du nombre de sous-clusters utilisé (2 ou 3).



Figure 53. Pourcentages relatifs des sous-clusters en fonction du nombre de sousclusters et des gammes spectrales utilisés lors de la classification KM. Classification sur les gammes spectrales associées à la glycosylation (A, 1000-1200 cm⁻¹), sialylation (B, 1700-1750 cm⁻¹) ou à la bande Amide I (C, 1600-1700 cm⁻¹) des mucines. *Triangles bleues*, colon normal ; *cercles rouges*, adénocarcinomes. *Barres bleus*, pourcentages moyens des sous-clusters dans les échantillons de colon normaux. *Barres rouges*, pourcentages moyens des sous-clusters dans les échantillons d'adénocarcinomes. *, 0.05 <P < 0.1; **, 0.01 <P < 0.05; ***, P < 0.01; *NS*, non significatif.

sous-clusters

sous-clusters

La sous-classification se basant sur 2 sous-clusters dans les gammes 1000-1200 cm⁻¹ (fig. 53*A*) ou 1700-1750 cm⁻¹ (fig. 53*B*) ne permet pas d'observer des différences significatives de distributions entre échantillons sains et tumoraux. En revanche, des différences significatives peuvent être observées pour une sous-classification sur la gamme spectrale correspondant à la bande Amide I, reconnue comme la plus informative pour l'étude de la structure secondaire

des protéines (fig. 53*C*) (Barth 2007). Le cluster 1 est alors présent à hauteur de 59,5% dans les tissus tumoraux contre 24,6% dans les tissus sains (P < 0.01). Cependant, la discrimination entre les deux types de tissus dans ce cas reste tout de même partielle puisque deux échantillons tumoraux présentent une distribution des sous-clusters similaires à celle des échantillons sains.

Le sous-clustering se basant sur 3 sous-clusters, permet d'obtenir, pour les gammes spectrales 1000-1200 cm⁻¹ (fig. 56*A*) et 1700-1750 cm⁻¹ (fig. 56*B*), une différence significative de distribution (P < 0,05) pour un seul des sous-clusters obtenus, sans pour autant autoriser une discrimination net de nos deux types d'échantillons. Pour la gamme spectrale 1600-1700 cm⁻¹ (fig. 53*C*), les clusters 1 et 3 sont moins représentés dans le tissus tumoraux que dans les tissus sains (32% et 14,2% dans les tissus tumoraux ; 51.4% et 31,9% dans les tissus sains ; P < 0,05). A contrario, le cluster 2 est plus représenté dans les tissus tumoraux en comparaison des tissus sains (53,9% et 16,6%, respectivement ; P < 0,05). Si dans ce cas la distribution des sous-clusters 1 et 3 ne permet qu'une discrimination partielle de nos 2 types d'échantillons, la distribution du sous-cluster 2 permet de les représenter en tant que 2 populations distinctes ne présentant pas de recouvrement.

Afin de prendre en compte la distribution des 3 clusters de manière simultanée, pour chacune des gammes spectrales étudiée, la figure 54 présente les mêmes résultats sous la forme de graphique 3D. Sur de tels graphiques, les 2 populations d'échantillons ne peuvent être clairement discriminées lorsque le sous-clustering se base sur les gammes spectrales attribuées à la glycosylation ou à la sialylation. En revanche, dans le cas de la gamme spectrale correspondant à la bande Amide I, il apparaît clairement que les deux populations d'échantillons ne présentent pas de recouvrement.



Figure 54. Discrimination des échantillons sains et tumoraux basée sur la distribution des sous-clusters après classification sur des gammes spectrales spécifiques. Les graphiques 3D de dispersion sont construits dans les mêmes conditions que la figure 51, mais après classification sur les gammes spectrales spécifiques associées à la glycosylation (A, 1000-1200 cm⁻¹), sialylation (B, 1700-1750 cm⁻¹), ou la bande Amide I (C, 1600-1700 cm⁻¹) des mucines. *Triangles bleus*, colon normal. *Cercles rouges*, adénocarcinomes. Les zones colorées recouvrant les données sont présentées en tant que guide visuel.

III. Mise en évidence d'un nouveau marqueur de discrimination entre tissus coliques sains et adénocarcinomes du colon

A. Investigation des différences spectrales

Il est généralement difficile de relier des variations spectrales IR avec des événements ou processus biologiques lors de l'analyse d'échantillons complexes tels que des tissus. Cependant, l'investigation visuelle de la dérivée seconde des spectres moyens des sous-

clusters peut fournir des informations intéressantes en vue de l'interprétation de nos résultats (fig. 55).



Figure 55. Dérivées secondes des spectres moyens des sous-clusters obtenus après classification sur gammes spectrales spécifiques. *A*, classification sur les bandes de glycosylation (1000-1200 cm⁻¹). *B*, classification sur les bandes de sialylation (1700-1750 cm⁻¹). *C*, classification sur la bande Amide I (1600-1700 cm⁻¹). *Violet, Bleu, Vert*, sous-clusters 1, 2 et 3 présentés dans la figure 53.

Concernant les signatures spectrales des sous-clusters obtenus sur la gamme 1000-1200 cm⁻¹, attribuée à la glycosylation des mucines (fig. 58A), on observe des variations de rapport d'intensité entre les différents sous-clusters. En effet, le rapport de bandes 1044/1076 cm⁻¹ est proche de 1 pour les sous-clusters bleu (2) et vert (3) ; et inférieur à 1 pour le sous-cluster violet (1). En revanche, le rapport 1125/1163 cm⁻¹ est proche de 1 pour les sous-clusters violet

(1) et vert (3) mais nettement supérieur à 1 pour le sous-cluster bleu (2). Ces variations relatives d'intensité des différents pics caractéristiques de la glycosylation des mucines nous informent des changements qualitatifs, intra- et inter-échantillons, de composition des chaînes latérales de carbohydrates des mucines.

Concernant les signatures spectrales des sous-clusters obtenus après classification sur la gamme 1700-1750 cm⁻¹ (fig. 55*B*), différents pics d'absorption sont observables selon le souscluster. Cette gamme spectrale reflète la sialylation des mucines via la mise en évidence du groupement carboxyle des acides sialiques (Khajehpour 2006). Un décalage en nombre d'onde d'une vibration associée à groupement chimique d'une même molécule peut traduire un changement dans l'environnement chimique de ce groupement. En effet, la fréquence vibration d'un groupement chimique dépend de ses interactions moléculaires avec son environnement proche.

Enfin, pour le sous-clustering sur la gamme 1600-1700 cm⁻¹ (fig. 55*C*), attribuée à la bande Amide I, différents pics sont également observables selon le sous-cluster, après dérivation secondaire de leurs spectres moyens. L'analyse de ces composantes spectrales permet d'évaluer la part relative des structures protéiques de types hélice- α , feuillet- et coude- β ou encore « random coil » dans un échantillon (Barth 2007 ; Kong 2007 ; Petibois 2006). Dans ce cas, le sous-cluster 2 (bleu), significativement plus représenté dans les tissus tumoraux comparé aux tissus sains (53,9% et 16,6%, respectivement, P = 0.001), présente une absorption marquée pour les vibrations attribuées aux feuillet- (1628 et 1684 cm⁻¹) et coude- β (1660 cm⁻¹). Par contre, les sous-clusters 1 et 3 (violet et vert, respectivement), présentent une absorption marquée pour les vibrations attribuées aux « random coil » (1651 cm⁻¹) et aux hélices- α (1656 cm⁻¹). Ces deux clusters sont significativement plus représentés dans les tissus normaux en comparaison des tissus tumoraux (51.4% et 31.9% dans les tissus normaux ; 32.0% et 14.2% dans les tissus tumoraux ; P = 0.011).

B. Comparaison entre profil d'expression des mucines et distribution des sous-clusters

Le profil d'expression des mucines dans nos échantillons a été analysé par IHC, puis comparé à la distribution des sous-clusters. Cette comparaison a été réalisée pour déterminer si les différences spectrales au niveau de la bande Amide I, reflétant la structure secondaire des mucines, pouvaient être attribuées aux variations de leur profil d'expression. En effet, de nombreuses études ont décrit, au sein d'adénocarcinomes mucineux, une sur-expression de MUC2 et une expression aberrante de MUC5AC, mucine gastrique non exprimée dans le colon sain mais anormalement présente dans les adénomes et certains adénocarcinomes.

Le tableau 10 présente la comparaison entre les pourcentages de pixels de chaque sous-cluster obtenu après classification des données sur la gamme 1600-1700 cm⁻¹, avec l'expression IHC de MUC2 et MUC5AC.

Samples	Immunohistochemical scores			Spectral subclusters (% pixels)		
	MUC5AC		MUC2	1	2	3
	(a)	з	2	34.1	47.8	18.1
		1	2	24.7	61.2	14.1
	(b)	0	2	43.0	42.2	14.8
Adenocarcinomas		2	2	32.0	37.5	30.5
(n=9)		0	2	23.5	70.4	6.1
		1	2	28.6	58.4	13.0
		3	2	36.1	60.6	3.3
		2	2	47.8	32.6	19.6
		0	2	17.2	74.2	8.6

Tableau 10. Comparaison entre l'évaluation immunohistochimique de l'expression de MUC2 et MUC5AC et les pourcentages de sous-clusters dans les images IRPC des échantillons tumoraux. La classification est basée sur la bande Amide I avec 3 sous-clusters. Les scores d'évaluation IHC vont de 0 à 3 (0 = pas de marquage ; 1 = marquage localise de faible intensité ; 2 = marquage répandu d'intensité modérée ; 3 = marquage répandu de forte intensité). Les pourcentages des sous-clusters 1 (violet), 2 (bleu) et 3 (vert) correspondent à ceux presentés dans les figures 54*C*, 55*C* et 56*C*. Les cas des adénocarcinomes (a) et (b) sont illustrés dans la figure 57.

Dans notre cas, les échantillons tumoraux présentent un niveau équivalent d'expression de MUC2 et une très grande disparité d'expression pour MUC5AC. Rappelons que les échantillons de colon sain, par définition, exprime uniquement MUC2 (résultats non montrés). Aucune relation entre les scores évaluant l'expression de MUC2 et MUC5AC et les pourcentages de pixels de chaque sous-cluster n'a pu être définie. La figure 56 illustre le cas des adénocarcinomes notés (a) et (b) dans le tableau 10. Ces deux adénocarcinomes

présentent des pourcentages de sous-clusters proches, permettant de les discriminer des échantillons sains. En revanche, leur profil d'expression des mucines est opposé, l'échantillon (a) exprimant fortement MUC5AC tandis que cette mucine est indétectable dans l'échantillon (b).



Figure 56. Illustration représentative de la comparaison des images IRPC des mucus avec les marquages IHC de MUC2 et MUC5AC. Les coupes tissulaires présentés sont adjacentes les unes aux autres. La classification des spectres de mucus est basé sur la bande Amide I (1600-1700 cm⁻¹) et 3 sous-clusters. Echelle, 200 μ m. L'expression de MUC2 est observée dans les 2 échantillons ; l'expression de MUC5AC n'est détectée que dans l'échantillon (a). Comme illustré par les images IRPC des mucus et les pourcentages des sous-clusters 1 (violet), 2 (bleu) ou 3 (vert), les deux échantillons présentent des profils spectraux proches.

Les résultats de ce chapitre démontrent la possibilité de discriminer des échantillons tissulaires coliques sains et tumoraux sur la base de l'hétérogénéité spectrale des mucus au niveau de la gamme spectrale 1600-1700 cm⁻¹, correspondant à la bande Amide I. Nous avons également démontré que les variations spectrales de cette bande, sensible à la structure secondaire des protéines, ne sont pas associées à une variation de profil d'expression des mucines dans ces échantillons.

DISCUSSION

Afin d'évaluer le potentiel de l'histologie IR des mucus pour la discrimination entre colon normal et adénocarcinomes coliques, nous avons analysé 9 coupes tissulaires d'adénocarcinomes et 7 de tissus sains. Cette approche a été conduite directement sur coupes paraffinées à l'aide du modèle de correction des données spectrales basé sur l'EMSC présenté dans le chapitre I. Une classification KM, basée sur la gamme spectrale 950-1750 cm⁻¹, a été effectuée simultanément sur l'ensemble des données issues des tissus sains et tumoraux. Quelque soit le nombre de clusters utilisé lors de la classification KM (de 2 à 6), aucun des clusters obtenus ne s'avère spécifique du tissu tumoral. Les différentes structures tissulaires sont toujours représentées par les mêmes clusters spectraux dans les tissus sains comme tumoraux. Particulièrement, les structures glandulaires aberrantes des adénocarcinomes sont représentées par les mêmes clusters que les glandes de Lieberkühn des tissus sains. Ceci met en évidence la faible intensité des variations spectrales existant entre ces deux types d'échantillons. Ce résultat est en adéquation avec la seule étude portant sur la caractérisation histologique d'un adénocarcinome colique par histologie IR (Lasch 2004), et dans laquelle aucune distinction entre structures glandulaires saines et aberrantes au sein d'une même coupe tissulaire n'a pu être effectuée.

Une classification basée sur 6 clusters a permis d'obtenir des images IRPC offrant une distinction claire entre mucus sécrété, partie basale et partie apicale des structures glandulaires des tissus sains et tumoraux. Cette distinction repose sur la variation de la contribution du signal spécifique des mucines entre les différents clusters. Ces résultats démontrent le potentiel de la microspectroscopie IR pour une caractérisation histologique automatique, sans marquage, de tissus sécréteurs selon leur concentration en mucines ; et ceci directement depuis des coupes tissulaires paraffinées. Dans le Chapitre 3, nous avons démontré la possibilité d'étudier certaines caractéristiques biochimiques des mucines par analyse IR du mucus sécrété. Or, ces glycoprotéines, représentant le principal constituant du mucus, sont présentes de manière hétérogène dans les mucus coliques et sont connues comme pouvant subir des altérations quantitatives et qualitatives dès les stades précoces de la carcinogénèse colique (Forgue-Lafitte 2007 ; Gaudier 2006 ; Aksoy 2004). Ainsi, nous avons cherché à étudier l'hétérogénéité biochimique des mucus par histologie IR pour évaluer son potentiel en tant que marqueur pour la discrimination entre échantillons coliques sains et tumoraux.

L'analyse spécifique in-situ de l'hétérogénéité du mucus colique par histologie IR a été réalisée en appliquant la méthode de sous-clustering. Comme mentionné précédemment, le sous-clustering consiste à isoler une population spectrale, associée à une structure tissulaire unique, afin d'étudier son hétérogénéité en s'affranchissant des interférences avec les spectres associées aux autres structures tissulaires durant le processus de classification. Cette approche originale possède aussi des avantages pratiques comme une réduction importante des temps de calcul et de la taille des fichiers générés. Quelque soit les paramètres (nombre de sous-clusters, gamme spectrale...) utilisés pendant cette classification, tous les sous-clusters obtenus ont pu être retrouvés dans les deux types de tissus, mais dans des proportions très variables. Le fait de retrouver des mucus spectralement identiques à la fois dans les échantillons sains et tumoraux n'est pas surprenant. En effet, on peut mettre en parallèle ce résultat avec ceux obtenus lors de l'identification, par des techniques conventionnelles comme l'IHC, de marqueurs tumoraux protéiques, détectés dans les tissus sains comme tumoraux mais dans des proportions différentes (Zlobec 2009).

Le sous-clustering des spectres de mucus sur la gamme 950-1750 cm⁻¹, reflétant la composition biochimique globale des mucines, met en évidence la pertinence de notre approche. En effet, sur la base de 2 ou 3 sous-clusters, les images IRPC permettent d'obtenir des différences significatives de distribution des sous-clusters entre tissus sains et tumoraux. Ceci est particulièrement marqué pour les images construites sur la base de 3 sous-clusters. Les différences spectrales entre les sous-clusters majoritairement présents dans les tissus sains et ceux majoritairement présents dans les tissus tumoraux se situent au niveau des gammes spectrales attribuées aux caractéristiques des mucines connues pour être altérées pendant la carcinogénèse, à savoir : 1000-1200 cm⁻¹ (glycosylation), 1500-1700 cm⁻¹ (structure des apomucines) et 1700-1750 cm⁻¹ (sialylation). Cependant, dans ce cas, la discrimination entre échantillons sains et tumoraux reste tout de même partielle, les deux populations d'échantillons présentant alors un recouvrement au niveau de la distribution moyenne des sous-clusters.

Nous avons par la suite analysé les données spectrales associées au mucus par sous-clustering sur ces gammes spectrales spécifiques, afin d'en évaluer le potentiel discriminant. Sur la base d'une classification en 3 sous-clusters, des différences significatives de distribution des sousclusters peuvent être observés pour les gammes spectrales associées à la glycosylation et à la sialylation des mucines. Ce résultat s'avère en adéquation avec la littérature. En effet, la variation de composition en oligosaccharides des glycannes des apomucines ainsi que l'altération de leurs acides sialiques, notamment au niveau de leur O-acétylation, ont été démontrés (Jass 2001 ; Byrd 2004 ; Aksoy 2004) lors de la carcinogénèse colique. Cependant, ces différences de distribution, bien que statistiquement significatives, n'ont pas permis une discrimination non ambiguë entre échantillons et tumoraux. Ceci est à mettre en parallèle avec le fait que la valeur diagnostique et pronostique de la détection de ces altérations par des méthodes d'études conventionnelles (principalement en IHC) n'a jamais été clairement établi, et n'est donc pas recommandée en clinique (Duffy 2007 ; Zlobec 2009).

La classification des données spectrales associées au mucus sur la bande Amide I (1600-1700 cm⁻¹) à quant à elle permit l'obtention de différences de distribution significatives permettant la séparation de nos échantillons sains et tumoraux en 2 populations distinctes. La dérivation du second ordre des spectres moyens de ces sous-clusters permet l'obtention des informations sur la part des structures secondaires (hélice- α , feuillet- ou coude- β , « random coil ») constituant les apomucines. L'analyse des composantes spectrales de ces clusters, couplée à leur variation de distribution entre tissus sains et tumoraux, met en évidence que les mucines sécrétées dans les échantillons d'adénocarcinomes muqueux présentent une part plus importante de structures de type feuillet- et coude- β par rapport aux mucines sécrétées dans les échantillons de tissus coliques sains. De telles variations structurales n'ont, à notre connaissance, encore jamais été décrites.

Nous avons évalué par la suite si ces changements structuraux des mucines secrétées pouvaient être en relation avec des variations de leur profil d'expression entre les échantillons sains et tumoraux. En effet, une surexpression de MUC2 et l'expression aberrante de MUC5AC ont été démontrées dans le cas des adénocarcinomes muqueux. La comparaison entre les images IHC et IRPC correspondantes n'a montré aucune relation entre les profils d'expression des mucines et les profils spectraux pour chacun des échantillons. Ainsi, l'origine des changements structuraux décris ici ne semble pas reposer sur un changement de profil d'expression des mucines.

Deux hypothèses peuvent être émises pour expliquer ce résultat original. Tout d'abord, il se pourrait que ces variations structurales soit en relation directe avec l'épissage alternatif des apomucines (Sternberg 2004). En effet, ce mécanisme d'expression a été mis en évidence pour plusieurs mucines transmembranaires telles que MUC1, MUC3 et MUC4 mais également pour la mucine sécrétée MUC2. Or, un épissage alternatif d'un ARNm se traduit par définition par un changement de structure primaire, et par conséquent secondaire des protéines. Cependant, une éventuelle variation d'épissage entre mucines issus de tissus sains et mucines issues de lésions tumorales n'a pas encore été démontrée.

Ces résultats pourraient également être expliqués par un phénomène d'altération du repliement de ces protéines (Toupet 2009). L'analyse des phénomènes prions a permis de confirmer l'existence d'une transmission d'une information biologique basée sur la conformation des protéines. Cette découverte représente une facette originale des mécanismes épigénétique qui laisse envisager pour ce mécanisme « prion » des fonctions physiologiques importantes. Ainsi, une altération du repliement des mucines, en particulier MUC2, pourrait avoir une conséquence directe sur les fonctions spécifiques de ces glycoprotéines. Cependant, si l'étude des mécanismes d'agrégations des protéines trouve un écho dans une large gamme de pathologies allant de la maladie de Parkinson au diabète, ceux-ci n'ont encore jamais été associés au développement tumoral. Mais on notera cependant un fait intéressant vis à vis des résultats obtenus dans la présente étude : les protéines dont le repliement est altéré dans le cadre des protéinopathies ont une structure commune appelée « structure β super-plissée », correspondant à l'association en parallèle de plusieurs feuillet- β . Ceci concorde avec nos résultats, démontrant un accroissement de la part des structures de type feuillet- β dans les mucines sécrétées provenant d'adénocarcinomes.

Quelle qu'en soit l'origine, les altérations de structures des mucines secrétées pourraient avoir un impact direct sur leurs fonctions spécifiques, en particulier pour MUC2, connue pour avoir une influence sur les mécanismes cellulaires de croissance, de différenciation ou d'invasion. Plus précisément, ces hypothèses pourraient potentiellement expliquer le potentiel prométastatique de MUC2 dans les cancers coliques muqueux, alors que cette même apomucine est décrite comme présentant des propriétés anti-tumorales dans la muqueuse colique saine. Il sera nécessaire de procéder à de plus amples investigations pour appuyer la pertinence de ces hypothèses. Mais si ces altérations de structures pouvaient être démontrées comme influençant les fonctions spécifiques des mucines, il est possible qu'elles puissent être détectées à des stades précoces de la séquence adénome-carcinome ou être mis en relation avec certain sous-types biomoléculaires de cancers. Ainsi, l'analyse IR des mucus pourrait représenter dans le futur un outil attractif d'aide au diagnostic et à la classification des lésions cancéreuses épithéliales.

L'histologie IR, par delà sa capacité à reproduire la plus grande part des informations fournit par l'analyse morphologique de coupes tissulaires colorées à l'HE, est donc capable de révéler des changements biochimiques discriminants entre tissus sains et tumoraux, non accessibles par coloration HE ou marquages IHC. L'étude présentée ici confirme donc les résultats obtenus dans le Chapitre 2, à savoir que l'histologie IR permet l'obtention d'une « histologie moléculaire» potentiellement plus informative que les procédures conventionnelles d'analyses histologiques.

Conclusion Générale

Vers une nouvelle génération d'histologie

L'amélioration des connaissances sur les mécanismes du développement des lésions tumorales et de l'hétérogénéité biomoléculaire qui en découle a mis en exergue les limites des approches classiques de caractérisation de ces pathologies. En effet, le diagnostic ainsi que l'évaluation pronostique et prédictive des lésions tumorales reposent principalement sur une étude subjective de la morphologie des tissus et sur la mise en évidence de marqueurs moléculaires ciblés qui, s'ils peuvent présenter un intérêt dans l'affinement de la caractérisation des lésions, ne prennent pas en compte la nature complexe et hétérogène des cancers. Dans ce cadre, le développement de nouveaux types de classifications multiparamétriques et/ou des techniques d'analyse haut-débit permettant d'établir de véritables « cartes d'identités » des tumeurs sont des approches présentant un fort potentiel pour une meilleure caractérisation des lésions tumorales et, par conséquent, une amélioration de la prise en charge clinique des patients. La spectroscopie IR, par sa capacité à sonder globalement la composition biochimique d'un tissu, représente également un outil innovant et original pour l'analyse des tissus. En effet, le développement conjoint de l'instrumentation et des méthodes de traitement des données ont permis de démontrer, à de nombreuses reprises, la capacité de cette technique pour la caractérisation de lésions tumorales. Ces évolutions techniques ont permis de développer la spectroscopie IR comme un moyen de reproduction, automatique et non subjectif, de l'histologie conventionnelle. Mais cette approche d'histologie IR, parce que cherchant à reproduire strictement l'histologie HE, en présente les même principaux inconvénients et ne tire pas profit des informations additionnelles relatives à la composition biochimique des tissus. En effet, il a été suggéré que la sensibilité de la spectroscopie est telle qu'elle pourrait révéler des variations biomoléculaires globale au sein d'une même structure tissulaire malgré une apparente homogénéité morphologique.

Les travaux présentés dans ce manuscrit démontrent le potentiel de la spectroscopie IR pour la définition d'une nouvelle génération d'histologie, plus informative que l'histologie conventionnelle parce qu'apportant des informations, résolues spatialement, relatives à l'hétérogénéité biomoléculaire intra- et inter-tumorale. Ainsi, nous avons pu mettre en évidence, au sein de xénogreffes de carcinomes de colon humain, une organisation tissulaire et une hétérogénéité du tissu tumoral non détectables par coloration HE. Nous avons également démontré, par l'analyse d'adénocarcinomes coliques muqueux, que l'étude de l'hétérogénéité spectrale de structures tissulaires pouvait aboutir à la découverte de marqueurs tumoraux originaux non accessibles par les techniques conventionnelles de type IHC ou hybridation *in situ*. Nous avons pu discriminer, par analyse directe de coupes tissulaires paraffinées, des échantillons de colons sains et tumoraux sur la base d'une information

spectrale relative à la structure secondaire des mucines. Ces différents résultats n'ont, à notre connaissance, jamais été rapporté auparavant.

Par conséquent, l'histologie IR peut être aujourd'hui considérée comme un nouvel outil d'investigation moléculaire des tissus, présentant un compromis entre techniques d'analyses haut-débit (par sa capacité à sonder globalement la constitution biochimique des tissus) et techniques microscopique (par sa capacité à reproduire les informations obtenues avec l'histologie classique), et offrant ainsi la possibilité de visualiser et d'analyser les tissus de manière originale.

Deux niveaux principaux d'applications sont alors envisageables pour ce type d'approche. Tout d'abord, l'histologie IR pourrait servir de base à l'amélioration de la compréhension des pathologies tumorales. En effet, l'identification de sous-types de structures tissulaires par histologie IR, couplée à leur microdissection et à leur analyse par techniques haut-débit, pourrait révéler certains mécanismes et/ou interactions moléculaires liés au développement tumoral. Ceci pourrait potentiellement mener à l'identification de nouveaux marqueurs moléculaires ou/et cibles thérapeutiques. De plus, la corrélation entre l'hétérogénéité des tissus tumoraux révéler par histologie IR, et les données cliniques (pronostic d'évolution et réponse aux thérapeutiques) pourrait permettre l'établissement de nouveaux systèmes de classification des lésions tumorales pour une meilleure prise en charge des patients. Ce type d'approche nécessite des étapes de validation à grande échelle, aujourd'hui réalisable grâce au modèle de déparaffinage numérique des coupes tissulaires. En effet, ce modèle autorise l'utilisation des très nombreux échantillons paraffinés disponibles dans les tumorothèques, permettant ainsi de mener des études rétrospectives.

L'histologie IR, de par son coût, sa simplicité et sa capacité à produire automatiquement un nouveau type d'histologie potentiellement plus informative que les colorations classiques, est d'ores et déjà bien placée pour une éventuelle application en milieu clinique. Si certains aspects techniques (temps d'acquisitions des images IR, puissance des ressources informatiques) sont aujourd'hui encore limitant, l'amélioration constante des systèmes d'acquisitions et de traitements des données devrait pallier à ces problèmes. Mais d'autres facteurs sont également à prendre en compte pour une application clinique de l'histologie IR. Dans ce cadre, la génération de modèle de prédiction de l'histologie IR nécessite la construction et la validation d'importantes bases de données de tissus, induisant alors une collaboration entre établissement hospitalier. Par conséquent, les protocoles d'acquisitions des

spectres inclus dans ces bases de données doivent être très rigoureusement standardisés. De plus, parce que l'histologie IR est sensible aux méthodes de fixations et d'inclusions, ces dernières doivent elles aussi être rigoureusement standardisées au niveau, par exemple, du type de paraffine utilisée ou des temps d'exposition aux fixateurs. Cette application en routine de l'histologie IR nécessiterait donc par définition une collaboration étroite entre biologistes, anatomopathologistes et biophysiciens. Or l'histologie IR reste à l'heure actuelle, malgré son potentiel énorme dans le domaine biomédical, un outil peu adapté aux biologistes. Ceci est particulièrement flagrant lorsque l'on sait qu'il n'existe pas, à notre connaissance, d'interface informatique d'utilisation dédié à ce genre d'approche. Ainsi, pour l'application clinique de l'histologie IR, un effort important doit être réalisé quant à la démocratisation de cette approche, comme cela à pu être fait pour les techniques de microscopie de fluorescence dont l'utilisation se base aujourd'hui sur des logiciels simples et intuitifs malgré la complexité des concepts physiques sous-jacents à ces techniques.

Références

- Adekani AM, Akhondi MM, and Sadeghi MR. Application of genomic and proteomic technologies to early detection of cancer. *Arch Iranian Med* **2008**; (11): 427-34.
- Aksoy N, and Akinci FO. Mucin macromolecules in normal, adenomatous, and carcinomatous colon: evidence for the neotransformation. *Macromol Biosci* **2004**; (4): 483-96.
- Ami D, Neri T, Natalello A, *et al.* Embryonic stem cell differenciation studied by FT-IR spectroscopy. *Biochim Biophys Acta* 2008; (1783): 98-106.
- Anderson JE, Hansen LL, Mooren FC, *et al.* Methods and biomarkers for the diagnosis and prognosis of cancer and other diseases: Towards personalized medicine. *Drug Rsistance Updates* **2006**; (9): 198-210.
- Andrus PG. Cancer monitoring by FTIR spectroscopy. Technol Cancer Res Treat 2006; 157-67.
- Arakawa H, Neault JF, and Tajmir-Riahi HA. Silver(I) complexes with DNA and RNA studied by Fourier transform infrared spectroscopy and capillary electrophoresis. *Biophys J* **2001**; (81): 1580-87.
- Argov S, Ramesh J, Salman A, *et al.* Diagnostic potential of Fourier-transforminfrared microspectroscopy and advanced computational methods in colon cancer patients. *J Biochem Opt* **2002**; (7): 1-7.
- Baker MJ, Gazi E, Brown MD, *et al.* Investigating FTIR based histopathology for the diagnosis of prostate cancer. *J Biophoton* **2009**; (2): 104-13.
- Banyay M, Sarkar M, and Gräslund A. A library of IR bands of nucleic acids in solution. *Biophys Chem* **2003**; (104): 477-88.
- Bara J, Gautier R, Daher N, Zaghouani H, and C D. Monocolonal antobodies against oncofeta mucins M1 antigens associated with pre-cancerous colonic mucosae. *Can Res* **1986**; (48): 3983-89.
- Bara J, Loisillier F, and Burtin P. Correlation between the presence of gastrointestinal antigens and the histological type of human ovarian mucinous cysts. *Protides of biological fluids, 27th colloquium*, In: Peeters H,editor **1979**. 339-42.
- Barth A. Infrared spectroscopy of proteins. *Biochim Biophys Acta* 2007; (1767): 1073-101.
- Bertrand D, and Dufour E. La spectroscopie infrarouge et ses applications analytiques. *Tec & Doc* 2000;
- Bhargava R. Toward a practical Fourier transform infrared chemical imaging protocol for cancer histopathology. *Anal Bioanal Chem* **2007**; (389): 1155-69.
- Bird B, Miljkovic M, Romeo MJ, *et al.* Infrared micro-spectral imaging: distinction of tissue types in axillary lymph node histology. *BMC Clin Pathol* **2008**; (8): 8-22.

- Bird B, Romeo M, Laver N, and Diem M. Spectral detection of micro-metastases in lymph node histopathology. *J Biophoton* **2009**; (2): 37-46.
- Bohic S, Heymann D, Pouezat JA, Gauthier O, and Daculsi G. Transmission FT-IR microspectroscopy of mineral phases in calcified tissues. *CR Acad Sci III* **1998**; (321): 865-76.
- Bonnier F, Rubin S, Debelle L, *et al.* FTIR protein secondary structure analysis of human ascending aortic tissues. *J Biophotonics* **2008** (1): 204-14.
- Bouhedja W, Sockalingum GD, Pina P, et al. ATR-FTIR spectroscopic investigation of E. Coli transconjugants beta-lactams-resustance phenotype. FEBS Letters 1997; (412): 39-42.
- Boydston-White S, Romeo M, Chernenko T, *et al.* Cell-cycle-dependent variations in FTIR microspectra of single proliferating HeLa cells: Principal componant and artificial neural network analysis. *Biochim Biophys Acta* **2006**; (1758): 908-14.
- Bresalier RS, Niv Y, Byrd JC, *et al.* Mucin production by human colonic carcinoma cells correlates with their metastatic potential in animal models of colon cancer metastasis. *J Clin Invest* **1991**; (87): 1037-45.
- Brockhausen I. Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO reports* **2006**; (7): 599-604.
- Brown DC, and Gatter KC. Ki67 protein: the immaculate deception? *Histopathology* 2002; (40): 2-11.
- Buisine MP, Colombel JF, Lecomte-Houcke M, *et al.* Abnormal mucus in cap polyposis. *Gut* **1998**; (42): 135-38.
- Buisine MP, Devisme L, Degand P, *et al.* Developmental mucin gene expression in the gastroduodenal tract and accessory digestive glands. II. Duodenum and liver, gallbladder, and pancreas. *J Histochem Cytochem* **2000**; (48): 1667-76.
- Byrd JC, and Bresalier RS. Mucins and mucins binding proteins in colorectal cancer. *Cancer Metastasis Rev* 2004; (23): 77-99.
- Cabanne F, and Bonenfant JL. Anatomie pathologique: principes de pathologie générale, de pathologie spéciale et d'aetopathologie. In : Maloine SA, Editeur, Paris, **1986**.
- Chabannon C, Thouvenin D, Jacquemier J, et al. Les banques de tissus tumoraux: pourquoi? Comment? Oncologie 2004; (6): 93-98.
- Chang JI, Huang YB, Wu PC, *et al.* Characterization of human cervical precancerous tissue through the fourier transform infrared microscopy with mapping method. *Gynecol Oncol.* **2003** (91): 577-83.
- Cho M, Dahiya R, Choi SR, *et al.* Mucins secreted by cell lines derived from colorectal mucinous carcinoma and adenocarcinoma. *Eur J Cancer* **1997** (33): 931-41.
- Ciorda S, de los Rios V, and Albar JP. Contributions of advanced proteomics technologies to cancer diagnosis. *Clin Transl Oncol* **2006**; (8): 566-80.
- Coindre M, Blanc-Vincent MP, Collin F, *et al.* Standards, options and recommendations: Practice guidelines for difficult diagnosis in surgical pathology or cytopathology in cancer patients. *Bull Cancer* **2001**; (88): 765-73.

Cone RA. Barrier properties of mucus. Adv Drug Deliv Rev 2009; (61): 75-85.

- Conti C, Ferraris P, Giorgini E, *et al.* FT-IR microimaging spectroscopy: a comparison between healthy and neoplastic human colon tissues. *J Mol Struct* **2008**; (881): 46-51.
- Corfield AP, Myerscough N, Longman R, *et al.* Mucins and mucosal protection in the gastrointestinal tract: new prospects for mucins in the pathology of gastorintestinal disease. *Gut* **2000**; (47): 589-94.
- De Jong WD, Bakker Schut TC, Maquelin K, *et al.* Discrimination between nontumor bladder tissue and tumor by Raman spectroscopy. *Anal Chem* **2006**; (78): 7761–69.
- Dekker J, Rossen JWA, Büller HA, and C EAW. The MUC family: an obituary. *Trends Biochem Sci* **2002**; (27): 126-31.
- Diamandis EP. Proteomic patterns in serum and identification of ovarian cancer. *Lancet* **2002**; (360): 170.
- Diem M. Infrared spectroscopy of fully and partially condensed chromatine: evidence for optically dense (dark) DNA. *FTIR spectroscopy in microbiological and medical diagnostic* **2002**. FTIR Workshop, 5, Berlin.
- Diem M, Romeo M, Boydston-White S, Miljkovic M, and Matthäus C. A decade of vibrationnal micro-spectroscopy of human cells and tissue (1994-2004). *Analyst* **2004**; (129): 880-85.
- Draux F, Jeannesson P, Gobinet C, *et al.* IR spectroscopy reveals effect of non-cytotoxic doses of antitumour drug on cancer cells. *Anal Bioanal Chem* **2009**; (PMID: 19798486).
- Duffy MJ, Van-Dalen A, Haglund C, *et al.* Tumour markers in colorectal cancer: European group on tumour markers (EGTM) guidelines for clinical use. *Eur J Cancer* **2007**; (43): 1348-60.
- Ellis DI, and Goodacre R. Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. *Analyst* **2006**; (131): 875-85.
- Erickson HS, Gillepsie JW, and Emmert-Buck MR. Tissue Microdissection. *Methods Mol Biol* **2008**; (424): 433-48.
- Fabian H, Anh_Ngo_Thi N, Eiden M, *et al.* Diagnosing begnin and malignat lesions in breast tissue sections by using IR-microspectroscopy. *Biochim Biophys Acta* **2006**; (1758): 874-82.
- Faoláin EO, Hunter MB, Byrne JM, *et al.* Raman spectroscopic evaluation of efficacy of current paraffin wax section dewaxing agents. *J Histochem Cytochem* **2005**; (53): 121-29.
- Fearon ER, and B V. A genetic model for colorectal tumorigenesis. Cell 1990; (61): 759-67.
- Fernandez DC, Bhargave R, Hewitt SM, and Levin IW. Infrared spectroscopic imaging for histopathologic recognition. *Nat Biotechnol* **2005**; (23): 469-74.
- Forgue-Lafitte ME, Fabiani B, Levy PP, et al. Abnormal expression of M1/MUC5AC mucin in distal colon of patients with diverticulitis, ulcerative colitis and cancer. *Int J Cancer* 2007; (121): 1543-49.
- Gaigneaux A, Ruysschaert JM, and Goormaghtigh E. Infrared spectroscopy as a tool for discrimination between sensitive and multiresistant K562 cells. *Eur J Chem* **2002**; (269): 1968-73.

- Gaudier E, and Hoebler C. Rôles physiologiques des mucines dans la barrière colique. *Gastroenterol Clin Biol* **2006**; (30): 965-74.
- Geladi P, McDougall D, and Martens H. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Appl Spectrosc* **1985**; (39): 491-500.
- Gillepsie JW, Gannot G, Tangrea MA, *et al.* Molecular profiling of cancer. *Toxicol Pathol* **2004**; (32): 67-71.
- Götte K, Tremmel SC, Popp S, *et al.* Intratumoral genomic heterogeneity in advanced head and neck cancer detected by comparative genomic hybridization. *Adv Otorhinolaryngol* **2005**; (62):
- Grünfeld JP. Recommandations pour le Plan Cancer 2009-2013. *Rapport au Président de la République* **2009**, 6.
- Guenot D, Guerin E, Aguillon-Romain S, *et al.* Primary tumour genetic alterations and intra-tumoral heterogeneity are maintained in xenografts of human colon cancers showing chromosome instability. *J Pathol* **2006**; (208): 643-52.
- Hilska M, Yrjö, Laine V, *et al.* The significance of tumor markers for proliferation and apoptosis in predicting survival in colorectal cancer. *Dis Colon Rectum* **2005**; (48): 2197-208.
- Irudayaraj JJ, Hong Y, and S S. Differenciation and detection of microorganisms using Fouriertransform infrared photoacoustic spectroscopy. *J Mol Struct* **2002**; (606): 181-88.
- Jackson M, and Mantsch HH. Infrared spectroscopy of biomolecules. In: Chapman, editor., New-York, Wiley-Liss **1996**; 311-40.
- Jansson A, and Sun XF. Ki-67 expression in relation to clinicopathological variables and prognosis in colorectal adenocarcinomas. *APMIS* **1997**; (105):
- Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology* **2007**; (50): 113-30.
- Jass JR. Molecular heterogeneity of colorectal cancer: implications for cancer control. *Surg Oncol* **2007**; (16): S7-S9.
- Jass JR, and Walsch MD. Altered mucin expression in the gastrointestinal tract: a review. *J Cell Mol Med* **2001**; (5): 327-51.
- Karadenizili A, Kolayli F, and Ergen K. A novel application of Fourier-transformed infrared spectrsoscopy: classification of slime from *staphylocicci. Biofouling* **2007**; (23): 63-71.
- Kendall C, Isabelle M, Bazant-Hegemark F, *et al.* Vibrationnal spectroscopy: a clinical tool for cancer diagnosis. *Analyst* **2009**; (134): 1029-45.
- Khajehpour M, Dashnau JL, and Vanderkooi JM. Infrared spectroscopy used to evaluate glycosylation of proteins. *Anal Biochem* **2006**; (348): 40-48.
- Kneipp J, Lasch P, Baldauf B, Beekes M, and Naumann D. Detection of pathological molecular alterations in scrapie-infected hamster brain by Fourier transform infrared (FT-IR) soectroscopy. *Biochim Biophys Acta* **2000**; (1501): 189-99.

- Kohler A, Kirschner C, Oust A, and 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.
- Kohler A, Sulé-Suso J, Sockalingum GD, *et al.* Estimating and correcting mie scattering in synchrotron-based microscopic fourier transform infrared spectra by extended multiplicative signal correction. *Appl Spectrosc* **2008**; (62): 259-66.
- Kondo T. Tissue proteomics for cancer biomarker development Laser microdissection and 2D-DIGE. BMB reports 2008; (41): 626-34.
- Kong J, and YU S. Fourier transform infrared spectroscopic analysis of protein secondary structures. *Acta Biochim Biophys Sin* **2007**; (39): 549-59.
- Konishi N, Hiasa Y, Matsuda H, *et al.* Intratumor cellular heterogeneity and alterations in ras oncogene and p53 tumor suppressor gene in human prostate carcinoma. *Am J Pathol* **1995**; (147):
- Krafft C, Steiner G, Beleites C, and Salzer R. Disease recognition by infrared and Ramant spectroscopy. *J Biophoton* **2009**; (2): 13-28.
- Krishna MC, Kegelaer G, ADT I, *et al.* Combined Fourier transform infrared and Raman spectroscopic approach for identification of Multidrug Resistance phenotype in cancer cells lines. *Biopolymers* **2006**; (82): 462-70.
- Krishna CM, Sockalingum GD, Bhat RA, *et al.* FTIR and Raman microspectroscopy of normal, benign, and malignant formalin-fixed ovarian tissues. *Anal Bioanal Chem* **2007**; (387): 1649–56.
- Lasch P. Diagnostic Potentials of FT-IR Microspectrometry in the Examination of Colorectal Adenocarcinomas. *1st International Workshop On Spectral Diagnostics (SD-1), Boston, June 21st 23rd,* **2007**.
- Lasch P, Chiriboga L, Yee H, and Diem M. Infrared spectroscopy of human cells and tissue: Detection of disease. *Technol Cancer Res Treat* **2002**; (1): 1-7.
- Lasch P, Haensch W, Lewis NE, Kidder LH, and Naumann D. Characterization of colorectal adenocarcinoma sections by spatially resolved FT-IR microspectroscopy. *Appl Spectrosc* **2002**; (56): 1-10.
- Lasch P, Haensch W, Naumann D, and M D. Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis. *Biochim Biophys Acta* **2004**; (1688): 176-86.
- Lasch P, and Naumann D. Spatial resolution in infrared microspectroscopic imaging of tissues. *Biochim Biophys Acta* **2006**; (1758): 814-29.
- Le Gal JM, Morjani H, and Manfait M. Ultrastructural appraisal of the Multidrug Resistance in K562 and LR73 cell lines from Fourier transform spectroscopy. *Can Res* **1993**; (53): 3681-86.
- Li QB, Sun XJ, Xu YZ, *et al.* Use of Fourier-transform infrared spectroscopy to rapidly diagnose gastric endoscopic biopsies. *World J Gastroenterol* **2005**; (25): 3842-45.
- Liu KZ, Schultz CP, Salamon EA, Man A, and Mantsch HH. Infrared spectroscopic diagnosis of thyroid tumors. *J Mol Struct* **2003**; (661): 397-404.
- Losi L, Baisse B, Bouzourene H, and Benhattar J. Evolution of intratumoral genetic heterogeneity during colorectal cancer progression. *Carcinogenesis* **2005**; (26): 916-22.
- Low Ying S, Shaw RA, M L, and Mantsch HH. Quantification of glucose and urea in whole blood by mid-infrared spectroscopy of dry films. *Vib Spectrosc* **2002**; (28): 111-16.
- Ly E, Piot O, Durlach A, Bernard P, and Manfait M. Differential diagnosis of cutaneous carcinomas by infrared spectral micro-imaging combined with pattern recognition. *Analyst* **2009** (134): 1208-14.
- Ly E, Piot O, Wolthuis R, *et al.* Combination of FTIR spectral imaging and chemometrics for tumour detection from paraffin-embedded biopsies. *Analyst* **2008**; (133): 197-205.
- Mall AS. Analysis of mucins: role in laboratory diagnosis. J Clin Pathol 2008; (61): 1018-24.
- Martens H, Jensen SA, and Geladi P. Multivariate linearity transformation for near-infrared reflectance spectrometry. *In: Proc. Nordic. Symp. Applied Statistics, Stokkand Forlag Publisher: Stavanger* **1983**; 205-34.
- Martens H, and Stark E. Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. *J Pharm Biomed Anal* **1991**; (9): 625-35.
- McShane LM, Altman DG, Sauerbrei W, et al. Reporting recommendations for tumor MARKer prognostic studies (REMARK). Exp Oncol 2006; (28): 99-105.
- Meade AD, Lyng FM, Knief P, and Byrne HJ. Growth substrate induced functional changes elucidated by FTIR and Raman spectroscopy in in-vitro cultured human keratinocytes. *Anal Bioanal Chem* **2007**; (387): 1717-28.
- Meichenin M, Rocher J, Galanina O, *et al.* Tk, a new colon tumor-associated antigen resulting from altered *O*-Glycosylation. *Can Res* **2000**; (60): 5499-507.
- Messerschmidt RG, and Harthcock MA. Infrared microspectroscopy: Theory and applications. In: New-York Marcel Dekker Inc, 1988. (Practical spectroscopy series ; 6) **1988**.
- Michor F, Iwasa Y, Lengauer C, and Nowak MA. Dynamics of colorectal cancer. *Semin Cancer Biol* **2005**; (15): 484-93.
- Mohlenhoff B, Romeo M, M D, and Wood BR. Mie-type scattering and non-Beer-Lambert absorption behavior of human cells in infrared microspectroscopy. *Biophys J* 2005; (88): 3635-40.
- Naumann D. Infrared spectroscopy in microbiology (Encyclopedia of analytical chemistry). In: R.A Meyers, Editors, Chichester, Wiley & Sons Ltd 2000; 102-31.
- Naumann D, Schultz CP, and Helm D. What can infrared spectroscopy tell us about the structure and composition of pathogenic bacteria using Fourier transform infrared spectroscopic and multivariate analysis. *Infrared spectroscopy of biomolecule*. *In: HH Martens and D Chapman eds.*, *Wiley-Liss Inc, New-York* **1996**; 279-310.
- Owonikoko T, Rees M, Gabbert HE, and M S. Intratumoral genetic heterogeneity in Barrett adenocarcinoma. *Am J Clin Pathol* **2002**; (117):
- Parker FS. Applications of infrared, Raman and resonance Raman spectroscopy in biochemistry. *Plenum Press, New-York* **1983**.

- Pelton JT, and McLean LR. Spectroscopic methods for analysis of protein structure. *Anal Biochem* **2000**; (277): 167-76.
- Perçinel S, Savas B, Ensari A, *et al.* Mucins in the colorectal neoplastic spectrum with reference to conventionnal and serrated adenomas. *Turk J Gastroenterol* **2007**; (18): 230-38.
- Petibois C, Carzola G, Cassaigne A, and Déléris G. Plasma protein content determined by Fourier-transform infrared spectrometry. *Clin Chem* **2001**; (47): 730-38.
- Petibois C, and Déléris G. Chemical mapping of tumor progression by FT-IR spectroscopy: towards molecular histopathology. *Trends Biotechnol* **2006**; (24): 455-62.
- Petibois C, Drogat B, Bikfalvi A, Déléris G, and Moenner M. Histological mapping of biochemical changes in solid tumors by FT-IR spectral imaging. *FEBS Letters* **2007**; (581): 5469-74.
- Popat S, Hubner R, and Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol* **2005**; (23): 609-18.
- Rajagopalan H, Nowak MA, Vogelstein B, and Lengauer C. The significance of unstable chromosomes in colorectal cancer. *Nat Rev Cancer* **2003**; (3): 695-701.
- Ramesh J, Salman A, Hammody Z, *et al.* Application of FTIR microspectroscopy for the characterization of malignancy: H-ras transfected murine fibroblastes as an exemple. *J Biochem Biophys Methods* **2001**; (50): 33-42.
- Rezaul K, Wilson LL, and Han DK. Direct tissue proteomics in human diseases: potential applications to melanoma research. *Expert Rev Proteomics* **2008**; (5): 405-12.
- Ripley BD. Linear discriminant analysis. *Pattern recognition and neural networks*. *In: Cambridge University Press: Cambridge* **1996**; 91-120.
- Romeo M, Mohlenhoff B, Jennings M, and Diem M. Infrared micro-spectroscopic study of epithelial cells. *Biochim Biophys Acta* **2006**; (1758): 915-22.
- Romeo MJ, Quinn MA, Burden FR, and McNaughton D. Influence of begnin cellular changes in diagnosis of cervical cancer using microspectroscopy. *Biopolymers* **2002**; (67): 362-66.
- Roussel P, Lamblin G, Lhermitte M, et al. The complexity of mucins. Biochimie 1988; (70): 1471-82.
- Sablin MP, Italiano A, and Spano JP. Les cancers coliques: facteurs pronostiques et prédictifs de la réponse au traitement. *Bull Cancer* **2009**; (96): 417-23.
- Sahu RK, Argov S, Bernshtain E, *et al.* Detection of abnormal proliferation in histologically 'normal' colonic biopsies using FTIR-Microspectroscopy. *Scand J Gastroenterol* **2004**; (39): 557-66.
- Sahu RK, and Moderchai S. Fourier transform infrared spectroscopy in cancer detection. *Future Oncol* **2005**; (1): 635-47.
- Salman A, Argov S, Ramesh J, *et al.* FT-IR microscopic characterization of normal and malignant human colonic tissues. *Cell Mol Biol* **2001**; (47): 159-66.
- Samowitz WS, Slattery ML, Sweeney C, *et al. APC* mutations and other genetic and epigenetic changes in colon cancer. *Mol Cancer Res* **2007**; (5): 165-70.

- Sarli L, Bottarelli L, Azzoni C, *et al.* Two subtypes of mucinous adenocarcinoma of the colorectum: Clinicopathological and genetic features. *Ann Surg Oncol* **2008**; (15): 1429-39.
- Schmitt A, Le Blanc B, Corsini MM, Lafond C, and Bruzek J. Les réseaux de neurones artificiels. *Bull et Mém de la Société d'Anthropologie de Paris* **2001**; (13): 143-50.
- Schwartz B, Bresalier R, and Kim YS. The role of mucin in colon-cancer metastasis. *Int J Cancer* **1992**; (52): 60-65.
- Selves J. Aspects histologiques et critères d'évaluation pronostique. *MS Médecine sciences* **2009**; (25): 9-11.
- Silvestro L, and Axelsen PH. Infrared spectroscopy of supported lipid monolayer, bilayer, and multibilayer membranes. *Chem Phys Lip* **1998**; (96): 69-80.
- Singh AP, Senapati S, Ponnusamy MP, *et al.* Clinical potential of mucins in dignosis, prognosis, and therapy of ovarian cancer. *Lancet Oncol* **2008**; (9): 1018-24.
- Smith G, Carey FA, Beattie J, *et al.* Mutations in APC, Kirsten-ras an p53 alternative genetic pathways to colorectal cancer. *Proc Natl Acad Sci USA* **2002**; (99): 9433-38.
- Sobin L, Wittekind C, and eds. TNM classification of malignant tumors. *In: New-York: Willey-Li.* **2002**.
- Sockalingum GD, Bouhedja W, Pina P, *et al.* ATR-FTIR spectroscopic investiguation of imepenemsusceptible and -resistant Pseudomonas aeruginosa isogenic strains. *Biochem Biophys Res commun* **1997**; (232): 240-46.
- Solassol J, Boulle N, Maudelonde T, and Mangé A. Protéomique clinique : vers la détection précoce des cancers ? *MS Médecine sciences* **2005**; (21): 722-28.
- Steinley D. K-means clustering: a half-century synthesis. Br J Math Stat Psychol 2006; (59): 1-34.
- Steller W, Einenkel J, Horn LC, *et al.* Delimitation of squamous cell cervical carcinoma using infrared microspectroscopic imaging. *Anal Bioanal Chem* **2006**; (384): 145-54.
- Sternberg LR, Byrd JC, Hansson GC, Liu KF, and Bresalier RS. Alternative splicing of the human MUC2 gene. *Arch Biochem Biophys* **2004**; (421): 21-33.
- Sternberg LR, Byrd JC, Yunker CK, *et al.* Liver colonization by human colon cancer cells is reduced by antisense inhibition of MUC2 mucin synthesis. *Gastroenterology* **1999**(116): 363-71.
- Stevens A, and Lowe J. Anatomie pathologique générale et spéciale. In : De boeck et Larcier s.a éditeurs, Bruxelles 1997.
- Toupet K, and Lehmann S. L'information cachée dans le repliement des protéines. *MS Médecine sciences* **2009**; (25): 307-10.
- Van-de-Weert M, Haris PI, Hennink WE, and Crommenlin DJA. Fourier transform infrared spectrometric analysis of protein conformation: effect of sampling method and stress factors. *Anal Biochem* **2001**; (297): 160-69.
- Velcich A, Yang W, Heyer J, *et al.* Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science* **2002**; (295): 1726-29.

- Vrabie V, Gobinet C, Piot O, *et al.* Independent component analysis of Raman spectra: application on paraffin-embedded skin biopsies. *Biomed Signal Process Control* **2007**; (2): 40-50.
- Wakatsuki K, Yamada Y, Narikiyo M, *et al.* Clinicopathological and prognostic significance of mucin phenotype in gastric cancer. *J Surg Oncol* **2008**; (98): 124-29.
- Wang L, and Mizaikoff B. Application of multivariate data-analysis techniques to biomedical diagnostics based on mid-infrared spectroscopy. *Anal Bioanal Chem* **2008**; (391): 1641-54.
- Wang TD, Triadafilopoulos G, Crawford JM, *et al.* Detection of endogenous biomolecules in Barrett's esophagus by Fourier transform infrared spectroscopy. *Proc Natl Acad Sci USA* **2007**; (104): 15864-69.
- Wehbe K, Pinneau R, Moenner M, Déléris G, and Petibois C. FT-IR spectral imaging of blood vessels reveals protein secondary structure deviations induced by tumor growth. *Anal Bioanal Chem* **2008**; (392): 129-35.
- Wolthuis R, Travo A, Nicolet C, *et al.* IR spectral imaging for histopathological characterization of xenografted human colon carcinomas. *Anal Chem* **2008**; (80): 8461-69.
- Wong PT, Wong RK, Caputo TA, Godwin TA, and Rigas B. Infred spectroscopy of exfoliated cervical cells: Evidence of extensive structural changes during carcinogenesis. *Proc Natl Acad Sci USA* 1991; (88): 10988-92.
- Wright EA. Quality control in histopathology. Proc Roy Soc Med 1975; (68): 619-20.
- Wright NA, Pike C, and Elia G. Induction of a novel epidermal growth factor-secreting cell lineage by mucosal ulceration in human gastrointestinal stem cells. *Nature* **1990**; (343): 82-85.
- Yano K, Ohoshima S, Shimizu Y, Moriguchi T, and Katayama H. Evaluation of glycogen level in human lung carcinoma tissues by an infrared spectroscopic method. *Cancer Letters* **1996**; (20): 29-34.
- Zlobec I, and Lugli A. Prognostic and predictive factors in colorectal cancer. *J Clin Pathol* **2008**; (61): 561-69.

ANNEXES

Annexe 1

Modes de vibrations moléculaires



Annexe 2

Les modes de mesure en microspectroscopie IRTF

Absorption/Transmission

La mesure en absorption consiste à quantifier l'énergie transmise au travers de l'échantillon. Des matériaux transparents dans l'IR doivent alors être utilisés comme supports. Parmi ces matériaux on trouve de nombreux sels tel que le KBr (qui pressé à 1000N peut emprisonner une poudre dans un cristal transparent) ; le séléniure de Zinc (ZnSe), le fluorure de calcium (CaF₂) ou le fluorure de baryum (BaF₂) sont quant à eux utilisés pour réaliser des supports transparents. L'échantillon étudié, quant à lui, doit présenter une épaisseur compatible avec une absorbance raisonnable, de l'ordre d'une unité d'absorbance.

La mesure en absorption est la méthode IR la plus simple pour l'analyse d'échantillons de toute nature et la plus utilisée dans l'étude de tissus biologiques.



Figure 2A. Principe du mode de mesure en transmission. $I_{source}(X)$, intensité globale émise par la source ; I(X), intensité transmise après passage à travers l'échantillon.

Réflexion externe

En réflexion externe, le rayon IR est directement réfléchi de la surface de l'échantillon. Deux cas se présentent alors :

- Si l'énergie qui pénètre une ou plusieurs particules est réfléchie dans toutes les directions, c'est de la réflexion diffuse. La réflexion diffuse peut être considérée comme la résultante de multiples réflexions, réfractions et diffractions sur des particules sans orientation préférentielle. Cette méthode est particulièrement adéquate pour l'analyse de poudres (solides, polymères), de surfaces mates, ou de solides intraitables (durs). Cependant, les spectres de réflexion diffuse s'avèrent complexes car contenant à la fois des éléments engendrés par la réflectance, la diffusion et l'absorbance de l'échantillon.
- Si l'angle de la lumière réfléchie est égal à l'angle d'incidence du rayon, c'est de la réflexion spéculaire. Cette méthode est utilisée le plus généralement pour l'étude des revêtements de surface (tels que les métaux traités en surface, les peintures et les polymères).



La réflexion spéculaire peut s'avérer gênante en venant perturber la mesure de la composante diffuse. Dans ce cas il est possible de la minimiser grâce à une configuration optique désaxée. La collection de la lumière diffusée s'effectue grâce à des miroirs elliptiques ou des sphères d'intégrations autorisant le recueil d'un maximum de lumière et permettant ainsi une amélioration de la sensibilité, des étalonnages plus robustes et une meilleure reproductibilité des mesures d'échantillon à échantillon. La réflexion diffuse permet d'obtenir des mesures précises d'échantillons inhomogènes ou constituées de particules de grandes tailles.

Réflexion-absorption

Ce mode correspond à une mesure d'échantillons minces montés sur une surface réfléchissante. Le faisceau traverse alors deux fois l'échantillon. Le support le plus utilisé dans ce type de mesure sont des lames de verres recouvertes d'atomes d'argent et d'oxyde d'étain transparente à la lumière visible mais réfléchissante dans le domaine de l'IR.

Ce mode de mesure permet l'étude de cellules issues de cultures et peut être également utilisé pour l'étude de tissus.



Figure 4A. Principe du mode de mesure en réflexion-absorption. $I_{source}(X)$, intensité globale émise par la source ; I(X), intensité transmise après passage à travers l'échantillon.

Réflexion totale atténuée (ATR)

Dans cette méthode de mesure, l'échantillon est placé sur un cristal transparent dans l'IR mais possédant un indice de réfraction élevé (ZnSe, diamant, germanium, KRS5) appelé aussi élément à réflexion interne. Lorsqu'il est éclairé au travers du cristal par un faisceau sous une incidence supérieure à un angle limite mais proche de celui-ci, une absorption se produit à l'interface entre le cristal et l'échantillon. L'efficacité de l'absorption dépend de la fréquence du rayonnement incident et de son angle ainsi que de l'intimité du contact entre le cristal et l'échantillon. La pénétration du faisceau est faible et dépend de la longueur d'onde. Pour des échantillons ayant le même indice de réfraction, la profondeur de pénétration est la même, permettant de comparer aisément des matériaux de même nature.



Cette méthode de mesure est utilisée pour l'étude des liquides, des semi-solides ainsi que des films et connaît à l'heure actuelle un développement orienté vers l'étude de tissus *in vivo*. En effet, l'ATR permet de grandement minimiser l'influence de l'absorption IR de l'eau présente dans l'échantillon.

Annexe 3

L'analyse en composante principale

Etant donné un ensemble d'observations décrites par des variables exclusivement numériques $\{x1, x2, ..., xp\}$, l'ACP a donc pour objectif de décrire ce même ensemble de données par de nouvelles variables en nombre réduit. Ces nouvelles variables seront des combinaisons linéaires des variables originales, et porteront le nom de Composantes Principales. Les composantes principales sont donc de nouvelles variables indépendantes, qui sont les combinaisons linéaires des variables initiales, rendant compte de la variation maximale.

L'Analyse en Composantes principales peut donc être vue comme une technique de réduction de dimensionnalité.

Pour illustré simplement le principe de l'ACP, prenons un ensemble d' « individus » N décrit par 3 variables $\{x, y, z\}$. On peut donc les représentés par un « nuage de points » dans un espace à 3 dimensions.



Cependant, à l'examen, il apparaît que ces données sont reparties sur un plan. Or, il suffit de 2 nombre pour repérer un point sur un plan, celui-ci n'ayant par définition que deux dimensions. En fait, dans ce cas, toute paire d'axe (non parallèle) dans ce plan formerait un repère acceptable, et permettrait donc de décrire ces individus à l'aide de 2 variables seulement.

Devant une pareille situation, l'ACP identifierait le plan et un repère dans ce plan, et permettrait donc une représentation bi-dimensionnelle des données. Les variables x et y sont alors considérées comme les deux composantes principales des « individus ».



Supposons alors que, dans ce plan, les données aient la forme d'un nuage allongé et étroit. On ne peut pas espérer trouver une représentation parfaitement fidele de ces données avec une seule variable. Mais si un des deux axes de ce repère est orienté dans la direction de la plus grande élongation du nuage de point, et si nous projetons tous les points du nuage sur cet axe, le « nuage » à une dimension alors obtenu sera une représentation approximative mais relativement fidele des données initiales à 3 dimensions. La variable X est alors considérée comme la composante principale des « individus ».



Figure 8A. Ensemble d'individus N décrit par sa première composante principale variables, x.

Annexe 4

L'EMSC

L'algorithme EMSC (Extended Multiplicative Signal Correction) est une extension de la méthode de « correction de dispersion multiplicative », ou MSC (Multiplicative Signal Correction) et qui vise à minimiser l'influence des facteurs d'interférences. L'EMSC fut introduit par Martens et Stark afin de supprimer les signaux interférents et les effets de dispersions de la lumière IR pour les données spectrales collectées dans le proche IR. Elle se base sur la modélisation des données spectrales et des données interférentielles environnementale ainsi que sur celle des effets de dispersion de la lumière. En résumé, chaque spectre s_i d'un ensemble de données peut être décrit de la manière suivante :

 $s_i = a_i \ \boldsymbol{\hat{s}} + b_i \boldsymbol{I} + c_i \boldsymbol{P} + e_i$

 \hat{s} est une estimation de l'ensemble de données, **I** la matrice des interférences, et **P** un polynôme prenant en compte les effets de dispersion de la lumière ; ai, bi et ci sont les coefficients de régression associées. Le résidu e_i correspond quant à lui au signal résiduel de cette modélisation.

Le spectre corrigé est alors calculé de la manière suivante :

$$S_{iEMSC} = \mathbf{\hat{s}} + \mathbf{e}_i / \mathbf{a}_i$$

Ainsi la variance due aux effets d'interférences et de dispersions est supprimée et le spectre est normalisé pour avoir une intensité comparable à **ŝ**. Dans le cas d'un ensemble de spectres d'interférences large et complexe, celui-ci peut être modélisé par ACP sur des données non centré sur leur moyenne. Ceci simplifie le modèle d'interférences et améliore l'efficacité et la stabilité des calculs.

Publications associées

IR spectral imaging of secreted mucus: a promising new tool for the histopathological recognition of human colonic adenocarcinomas (*Histopathology, 2010*)

Adrian Travo¹, Olivier Piot¹, Rolf Wolthuis², Cyril Gobinet¹, Michel Manfait¹,

Jacques Bara ³, Marie-Elisabeth Forgue-Lafitte ³ and Pierre Jeannesson ¹

¹ CNRS UMR 6237, Faculté de Pharmacie, Université de Reims Champagne-Ardenne, 51096 Reims, France.

² Rivers Diagnostics, 3029 AK Rotterdam, The Netherlands.

³ INSERM UMR 893 EQ 8, Hôpital Saint-Antoine, 75571 Paris, France.

Running title: Mucus IR study in colon adenocarcinomas

Key Words: Colon cancer, Mucin, Microspectroscopy, IR imaging.

Corresponding author: Pierre Jeannesson, CNRS UMR 6237, Faculté de Pharmacie, Université de Reims Champagne-Ardenne, 51 rue Cognacq-Jay, 51096 Reims, France.

Abstract

Aims: During the colon carcinogenesis, mucin-type glycoproteins are known to undergo quantitative and qualitative alterations. In this study, we tested the relevance of IR spectral histology for the histopathological recognition of colon adenocarcinomas based on mucin-associated IR spectral markers.

Methods and results: Paraffin-embedded tissue sections of normal human colon and adenocarcinomas were analyzed directly by IR-microspectroscopy (IR-MSP), without prior chemical dewaxing. IR-MSP imaging combined with multivariate analysis permitted to construct IR color-coded images of the tissue sections providing spatially-resolved biochemical information. This allows localization of mucin-rich areas and provides a label-free spectral-based staining of secreted mucus related to the biochemical heterogeneity of its mucin content. IR images of secreted mucus display the same spectral clusters in both normal and adenocarcinomas colon tissues but with significant differences in surface percentages. Such differences allow a distinction between these two tissue types. Our investigations showed that spectral variations associated to changes of mucins secondary structure are the most accurate mucus spectral marker for discriminating normal colon from adenocarcinomas of the sample set.

Conclusion: This study demonstrates the potential of IR-MSP imaging to provide a new type of histology, independent from visual morphology, presenting tremendous possibilities for discovery and clinical monitoring of cancer markers.

Introduction

Mucus is a viscoelastic gel-like secretion which coats, lubricates and protects epithelia in order to form the primary line of defence for the luminal surfaces against external environment aggressions. The structural characteristics of this protective barrier are primary indicators of its physiological function, and changes of its composition have long been identified in gastrointestinal diseases ^{1, 2}. The secreted mucin-type glycoproteins, characterized by the large number of carbohydrate chains O-glycosidically linked to the peptide, represent the main structural component of mucus and are expressed by specialized epithelial cells with a restricted expression pattern within the human body ³. At present, there is a growing interest in mucins as biomarkers for the early detection and prognosis of various carcinomas ⁴. Indeed, in tumor tissues, the expression of mucins is often quantitatively and qualitatively abnormal. Such changes may not only alter the function of cells, their antigenic and adhesive properties, but may also play a role in invasion and metastasis ⁵.

In human colorectal adenocarcinomas, decreased expression of MUC2, the prominent secreted gelforming mucin in the colon, is generally observed. In contrast, the secreted mucin MUC5AC, a product of normal gastric mucosa absent from normal colon, is frequently expressed in precancerous lesions of colon and in some adenocarcinomas. Additionally, for both of these secreted mucins, alterations of the sialic acid content and truncation of the oligosaccharide side chains resulting in new antigenicities have been described ^{4, 6, 7}. These studies mainly relied on histochemical or immunohistochemical methods, with few direct chemical analyses.

During the last decade, Fourier transform infrared microspectroscopy (IR-MSP) emerged as an alternative tool for differentiating tissue structures and for diagnosing tissue pathologies such as tumoral lesions. Indeed. unlike classical histology (mainly morphology-based) or immunohistochemistry (targeting specifically certain proteins), IR vibrational spectroscopy probes the intrinsic biochemical composition of the tissue. This biophotonic technique can detect subtle molecular and structural changes directly from paraffin-embedded (PE) tissue sections, without prior chemical dewaxing⁸⁻¹¹. These characteristics of IR-MSP make this methodology an attractive tool for routine high throughput diagnostic challenges, independent from operator variability inherent to visual morphology.

Taking into account the potential role of mucins as markers for human colon carcinogenesis, the present study was designed to perform an *in situ* analysis of secreted mucus by IR-MSP imaging, directly on alcohol-fixed PE tissue sections. Our objective was to assess the potential of mucin-associated spectral markers for discriminating normal from tumoral colon tissues.

Materials and Methods

Tissue specimens. Sixteen alcohol-fixed PE human colon samples were included in this study. Nine samples were from patients with colon adenocarcinoma and 7 from normal individuals, according to the informed consent of patients. Colon adenocarcinomas were staging from T2N0 to T3N1. Clinical and histological characteristics are available in Table S1 in the Supplementary Data. All samples were cut to 8 μ m thick slices and deposited onto 2 mm thick CaF₂ substrates (Crystran Ltd, Poole, UK) for IR measurements. No attempt was made to remove the paraffin from these tissue sections. In addition, the two previous and two next adjacent slices were deposited onto standard glass slides to perform hematoxylin-eosin (HE) staining for conventional histology and immunohistochemical evaluation of MUC2 and MUC5AC expression.

Mucus and mucin samples. Four different lyophilised mucus samples from human ovarian cysts, previously characterized by Bara *et al.* ¹², and one sample of bovine submaxillary mucin type II (Sigma, St Louis, USA) were used as reference products.

FTIR data collection of mucus and mucin samples. IR absorption spectra were collected using a Tensor 37 FTIR spectrometer (Brüker Optics GmbH, Ettlingen, Germany) equipped with the high throughput extension (HTS-XT) accessory. The HTS-XT is a micro-plate reader developed for automated IR spectroscopic analysis of large sample sets. An aqueous solution of each sample at 30% (w/vol) was prepared and 15 μ L of each solution were deposited on the 96 well HTS-XT micro-plate, with 6 replicates per sample, and air dried overnight at room temperature. Absorption spectra of the samples were measured with 64 scans co-added and pre-processed by baseline correction and normalisation onto the Amide I band (1600-1700 cm⁻¹). Average spectra were then calculated and used as reference spectra for the comparison of spectral signatures between each sample. Data collection, pre-processing and evaluation of spectra were performed using the version 6.5 of Brüker Optics User Software (Brüker Optics GmbH, Ettlingen, Germany).

FTIR data collection of PE tissue sections. IR spectra were collected with a Spectrum Spotlight 300 FTIR Imaging System (Perkin Elmer Life Sciences, Courtaboeuf, France) using the image mode. A specially designed microscope box was purged with dry air, reducing spectral contributions from atmospheric water vapour and CO₂. Prior to acquisition, a visible image of the tissue section was recorded and mucus-rich areas were selected by comparison to the adjacent HE-stained section. Spectral images were recorded with a spatial resolution of 10 μ m (6.25 μ m step size), a spectral resolution of 8 cm⁻¹ and 2 measurements per spectrum. Prior to each data acquisition, a reference spectrum from the CaF₂ window was recorded with a 8 cm⁻¹ spectral resolution and 240 accumulations, and automatically subtracted from the dataset. In addition, for each slice, an area containing an average of 30,000 spectra from paraffin only was measured at the same conditions as the tissue spectra. Finally, water vapour spectra were measured on the Perkin Elmer Spectrum One detector which is part of the Spotlight 300 system. As previously reported ¹⁰, the paraffin and water vapour spectra were used in the extended multiplicative signal correction (EMSC) process to correct the tissue spectra for the presence of these interferences.

Construction of IR color-coded images. Spectra of the data set from all the tissue sections were clustered using a k-means algorithm. The k-means algorithm partitions the data into a given number of clusters, so that the sum of the distances of the data points to their respective cluster centers are minimized. The created clusters contain spectra that present comparable spectral characteristics. Therefore, spectra from different clusters exhibit different spectral features. The k-means clustering method, apart from the number of clusters, is unsupervised. This algorithm is especially suitable to cluster large data sets such as the spectral images of this study. Following the clustering process, distinct color are arbitrary assigned to all spectra belonging to a given cluster. Then, these colours are plotted as pixels at the (x,y) coordinates from which the spectrum was acquired, thereby permitting to construct color-coded images on the basis of cluster analysis. White areas correspond to spectra omitted in the analysis after a quality test based on the rules listed in Wolthuis *et al.* ¹⁰. K-means clustering was processed using 4 images per tissue section, covering an average area of 2 mm² per image.

Software platform. Data were processed using Matlab 2007b (The Mathworks, Natick, USA) based software. Routines were partly written in house and partly written at the Centre for Optical Diagnostics and Therapy of the Erasmus Medical Centre (Rotterdam, NL).

Immunohistochemistry. The expression of MUC5AC and MUC2 was investigated in all normal and adenocarcinomas samples. Immunohistochemical staining was performed by routine laboratory methods, using anti-M1/MUC5AC monoclonal antibodies and anti-MUC2 polyclonal antibody previously described in Forgue-Lafitte *et al.*⁷.

Immunohistochemical staining evaluation. Each slides was assigned a score for intensity and pattern of the MUC5AC and MUC2 staining. Immunohistochemical scores range from 0 to 3 (0=no staining, 1=localized trace staining, 2=widespread staining of moderate intensity, 3=widespread staining of strong intensity). The investigator who scored all specimens was blinded to all clinical and spectral information.

Statistical analysis. Statistical analysis was carried out using a Mann-Whitney test. Results were considered to be statistically significant when P values were determined to be < 0.1.

Results and Discussion

In this study, a label-free IR spectral-based staining of human secreted mucus in normal and adenocarcinomas colon was implemented directly on PE tissue sections without previous chemical dewaxing. This is, to our knowledge, the first IR study focusing on the secreted mucus. Our aim was to show that its spectral signature could be used as a potential spectral marker for discriminating normal from tumoral colon tissues.

Previously to the *in situ* analysis of mucus within tissue sections, we characterized the spectral signatures of mucus and purified mucins which are described as the main structural component of the mucus. For this purpose, we collected IR spectra of lyophilised mucus samples extracted from human ovarian cysts and compared them to spectra of pure submaxillary bovine mucin type II, a member of the mammalian gastrointestinal mucin family ¹³. The samples of ovarian cysts mucus originated from 4 different patients and expressed both MUC2, the major mucin secreted in colon, and MUC5AC, abnormally expressed in precancerous colon and in some adenocarcinomas ¹². They also presented the advantage over crude mucus extracted from colon, to be free of contamination by faecal material and/or blood. IR-MSP analysis of these samples was performed in the 950-1750 cm⁻¹ spectral range, previously reported as the most informative for biological samples ¹¹. As depicted in Fig. 1, all ovarian mucus samples exhibited the same vibration bands, located at precise wavenumbers, but with different absorbances. The main vibrations of these spectra correspond i) to the protein Amide I and II bands at 1656 and 1550 cm⁻¹, attributed to vibrations of the peptide bonds ¹⁴ and ii) to the glycosylation bands ¹⁵ at 1044, 1076 and 1125 cm⁻¹. Such spectral data strongly reflect the glycopeptide nature of the secreted mucus. It should be noted that the marked variation in intensity observed for glycosylation bands can be indicative of variations in carbohydrate content ¹⁵ between the different samples of secreted mucus.

We next evaluated the spectral contribution of mucins in the secreted mucus signal by comparison to the spectral signature of purified bovine submaxillary type II mucin. As shown in Fig. 1, spectra of purified bovine mucin presented vibration bands located at identical wavenumbers as those of the human mucus spectra. These results support that the characteristic IR signal of secreted mucus corresponds mainly to the mucin IR signature and that minor constituents of mucus such as secreted immunoglobulin A, alpha 1-antitrypsin, lysozyme or lactoferrin ¹ have a negligible spectral contribution. Consequently, these results indicate that IR-MSP offers the possibility to analyze secreted mucus and its mucins content without any dyes or extrinsic probes.

In order to assess the potential of the IR analysis of secreted mucus for discriminating normal from adenocarcinoma tissues, we investigated 9 colon adenocarcinomas and 7 normal specimens by IR-MSP imaging. IR-MSP imaging is an innovative approach, not requiring any dyes or extrinsic probes, which permits to construct highly informative spectral images based on the intrinsic molecular

composition of the tissues. It includes a combination of chemometrics and multivariate statistics previously developed in our group ^{10, 11}, allowing to collect spectral information directly from PE tissue sections, without prior chemical dewaxing. Spectral data from normal and tumoral tissues were processed simultaneously by k-means clustering in the 950–1750 cm⁻¹ spectral range. Color-coded images were constructed based on this clustering analysis, as explained in materials and methods. As shown in Fig. 2A, images were built by using varying settings for the number of clusters (2, 4 (not shown), and 6) to localize spectral clusters assigned to the different histological structures in both tissue types, with a particular emphasis on secreted mucus. Fig. 2A also displays the photomicrographs of HE-stained sections corresponding to the IR-color coded images, and Fig. 2B and C display the second derivatives of the average spectra associated to each cluster in the 950-1350 cm-1 spectral range. The second derivatives are shown to enhance spectral features. Concerning the 2 cluster-based classification (Fig. 2A), cluster 1/2 (dark green) presents the 3 vibrations characteristic of the spectral signature of the mucins carbohydrate side chains (1044, 1076, and 1125 cm⁻¹) as previously shown in Fig 1 In contrast, cluster 2/2 (light green) does not present these simultaneous vibrations but displays bands at 1028 and 1080 cm⁻¹ assigned to glycogen ¹⁶, 1055 cm⁻¹ to DNA ¹⁰, and 1200, 1280 and 1335 cm⁻¹ to collagen ⁹. Such spectral characteristics suggest, that unlike cluster 2/2, cluster 1/2 can be associated to mucin-rich areas including secreted mucus and mucus-secreting structures. These results are confirmed by comparing IR color-coded images with conventional histology. Interestingly, the 6 cluster-based classification efficiently discriminates the secreted mucus from its intracellular form and highly correlates to histological structures as assessed by HE staining. Indeed, as shown in Fig. 2C, the average spectrum of cluster 1/6 (blue) exhibits the most intense IR bands of mucin carbohydrate side chains (1044, 1076, and 1125 cm⁻¹) without presenting IR bands of nucleic acids or collagen. This implies that cluster 1/6 is representative of the secreted extracellular mucus found either in the lumen of the intestine and in the central part of the crypts in normal tissue, or in glandular structures of tumoral tissue. The mean spectra of clusters 2/6 and 3/6 (pink and black, respectively) also display the 3 characteristic bands of mucins, but less intense than cluster 1/6. Spectra of the 3 other clusters (4/6, 5/6 and 6/6) does not present mucin bands but marked absorption at wavenumbers attributed to glycogen (1028 and 1080 cm⁻¹), DNA (1055 cm⁻¹) and collagen (1200, 1280 and 1335 cm⁻¹). These characteristics implied that clusters 4/6, 5/6, 6/6 can be attributed to histological structures devoid of mucus as visible in the HE-stained sections (Fig. 2A). Concerning clusters 1/6, 2/6 and 3/6 presenting a progressive decrease of the mucin signal, analysis of the color-coded images permits to identify a spatial ordering of these spectral clusters. Indeed, the clusters are distributed according to the following sequence: cluster 1/6 borders exclusively with cluster 2/6 which also borders with cluster 3/6. Taking into account the histological structure of the crypts in normal colon (Fig. 2A) and the spectral signatures of each cluster (Fig. 2B and C), this observation suggests that cluster 1/6 can be assigned to the secreted mucus whereas cluster 2/6 and 3/6 can be assigned respectively to the apical and basal parts of the mucus-secreting cells. A similar cluster spatial ordering and assignment can be retrieved in adenocarcinomas despite the altered morphological architecture of these tissues. These outcomes show the potential of IR analysis for an automated label-free histological characterization of secreting tissues such as colon directly from PE tissue sections, providing also spatially-resolved molecular information. As proved by the results of the k-means clustering based on 2, 4 (data not shown) or 6 clusters, a minimum number of 6 clusters is required to achieve a clear distinction between secreted mucus, intracellular mucus and surrounding non-secreting structures in both normal and tumoral tissues.

Most studies about IR-MSP imaging on tissue sections focus on reproducing conventional histology⁸, ^{9, 11, 17}. But, as we recently reported ¹⁰, IR-MSP imaging can also provide additional spatially-resolved biochemical information about tissue subtype heterogeneity, undetectable with conventional HE staining. Consequently, our next step was to assess the potential of IR-MSP imaging for studying mucus heterogeneity as a spectral marker for discriminating normal colon from adenocarcinomas colon tissues. For this, a k-means clustering was performed exclusively on secreted mucus spectra previously classified in cluster 1/6 (Fig. 2). This subclustering permitted i) to avoid the interferences of the other histological structures during the classification process and ii) to provide a spectral-based staining of secreted mucus, according to the biochemical heterogeneity of its mucin content. Data indicated that whatever the number of subclusters tested (2 to 6), they could be found in both tissue types, without detectable spatial ordering and independently of the localization of the secreted mucus, i.e. in intestinal or crypt lumen of normal colon or in glandular structures of adenocarcinomas. This is illustrated for a 3 subcluster-based classification in Fig. 3A, which displays representative color-coded images of secreted mucus in a normal colon (top) and a colon adenocarcinoma (bottom) tissue sections. Furthermore, significant differences of average subcluster surface percentages were observed between tissue types for the whole sample set. This is particularly marked for the 3 subcluster-based classification compared to the 2, 4 or 5 subcluster-based ones (Fig. 1S in the Supplementary Data). Indeed, in 3 subcluster-based color-coded images (Fig. 3), subcluster 1 (green) appears in average to be significantly more present in normal tissues as compared to tumoral ones (39.8% and 14.9% respectively, P = 0.001). By contrast, subcluster 2 (red) appears in average to be significantly less present in normal tissues as compared to tumoral ones (34.5% and 54.9% respectively, P = 0.07). Subcluster 3 (black) is retrieved in approximately equal amounts in both tissue types (25.6% and 29.7%, P = 0.79). Therefore, we next investigated spectral differences between subcluster 1 and 2 by calculating the second derivative of their average spectra. As shown in Fig. 3B, major spectral differences between spectra of these clusters are located in the 1000-1200 cm⁻¹, 1500-1700 cm⁻¹ and 1700-1750 cm⁻¹ spectral ranges respectively related to glycosylation, secondary structure and sialylation of mucins. These observed differences in both subcluster percentages and spectral characteristics between the subclusters may reflect biochemical changes between tumoral and normal secreted mucus.

Next, we examined the classification value of these observed differences for discriminating adenocarcinomas from normal colon. The 3 subcluster-based classification of secreted mucus in the 950-1750 cm⁻¹ spectral range does not permit to discriminate the tumoral samples from the normal ones. This can be visualized in the 3D scatter plot constructed from the calculation of the subcluster percentages for each sample (Fig. 3C). Indeed, this 3D plot evidences an overlap of the tissue types, therefore does not allow an unambiguous discrimination of the samples. Consequently, we next investigated variations in restricted spectral ranges, related to changes in glycosylation (1000-1200 cm⁻ ¹), protein structure (1600-1700 cm⁻¹) or sialylation (1700-1750 cm⁻¹). Fig. 4A-C shows the 3D scatter plots of the normal and tumoral samples according to subclusters surface percentages. Subclusterings based on the glycosylation and sialylation spectral ranges (Fig. 4A and B) lead to an overlapping between normal and tumoral tissues. However it is noteworthy that a clear separation between the tissue types is only achieved with a spectral classification in the Amide I band (Fig. 4C). Analysis of the second derivative of the corresponding average spectra was then performed to identify vibrations assignable to the different protein conformation types such as α -helix, β -sheet or random coils ^{18, 19}. In Fig. 5, subcluster 2A (dark blue), significantly more present in tumoral tissues compared to the normal ones (53.9% and 16.6% respectively, P = 0.001), presents a high absorption in vibrations assigned to β -sheet (1628 and 1684 cm⁻¹) and β -turn (1660 cm⁻¹). By contrast, subclusters 1A and 3A (purple and green respectively), present a high absorption in vibrations attributed to unordered (1651 cm⁻¹) and α helix structures (1656 cm⁻¹) and are significantly more present in normal tissues compared to tumoral ones (51.4% and 31.9% in normal tissues; 32.0% and 14.2% in tumoral tissues; P < 0.011). This indicates that secreted mucins from normal tissues contain mainly unordered and α -helix structures whereas secreted mucins from adenocarcinomas present mainly β -sheet and β -turn structures. We next evaluated whether these structural changes can be related to the expression profile of MUC2 and MUC5AC, abnormally expressed in colon adenocarcinomas^{4, 6, 7}. For this, we have compared the expression of these mucins in each sample, as detected by immunohistochemistry, with their corresponding mucus IR color-coded images. No correlation was established between MUC2 and MUC5AC expression and the spectral clustering analysis of mucus (Table S2 and Fig. S3 in the Supplementary data). This result demonstrates that, in the studied adenocarcinoma samples, changes in mucin secondary structure content are not related to mucin expression profile.

The present study confirms our previous results ¹⁰ demonstrating that IR-MSP imaging can evidence biochemical heterogeneity within tissue subtypes, providing a new type of molecular histopathology that is more informative than conventional HE staining procedures. This approach, independent of observer variability, allows, in conjunction with advanced chemometrics, an automated and fast analysis of tissue samples, based on their intrinsic biochemical constitution. It also offers the potential advantage to detect molecular changes that precede any morphological change, thus enabling earlier detection of tumoral lesions. Here, we demonstrate the potential of IR-MSP imaging for the histopathological recognition of colon adenocarcinomas based on a label-free spectral-based staining

related to mucin secondary structure changes in secreted mucus. To our knowledge, such structural alteration of mucin secondary structure in colon adenocarcinomas has never been described before. Further investigations will focus on the IR-MSP analysis of secreted mucus in the early stages of the adenoma-carcinoma sequence. If spectral markers associated to mucin secondary structure changes could be detected before neotransformation, IR-MSP analysis of secreted mucus may represent in the future an attractive tool for improvements in the early diagnosis of epithelial cancers.

References

1. Aksoy N, Akinci OF. Mucin macromolecules in normal, adenomatous, and carcinomatous colon: evidence for the neotransformation. *Macromol. Biosci.* 2004; **4**; 483-96.

2. Jass JR, Walsh MD. Altered mucin expression in the gastrointestinal tract: a review. *J. Cell. Mol. Med.* 2001; **5**; 327-51.

3. Singh AP, Senapati S, Ponnusamy MP, *et al.* Clinical potential of mucins in diagnosis, prognosis, and therapy of ovarian cancer. *Lancet Oncol.* 2008; **9**; 1076-85.

4. Mall AS. Analysis of mucins: role in laboratory diagnosis. J. Clin. Pathol. 2008; 61; 1018-24.

5. Brockhausen I. Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO Rep.* 2006; **7**; 599-604.

6. Byrd JC, Bresalier RS. Mucins and mucin binding proteins in colorectal cancer. *Cancer Metastasis Rev.* 2004; 23; 77-99.

7. Forgue-Lafitte ME, Fabiani B, Levy PP, *et al.* Abnormal expression of M1/MUC5AC mucin in distal colon of patients with diverticulitis, ulcerative colitis and cancer. *Int. J. Cancer* 2007; **121**; 1543-9.

8. Diem M, Romeo M, Boydston-White S, Miljkovic M, Matthaus C. A decade of vibrational microspectroscopy of human cells and tissue (1994-2004). *Analyst* 2004 Oct; **129**; 880-5.

9. Lasch P, Haensch W, Naumann D, *et al.* Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis. *Biochim. Biophys. Acta.* 2004; **1688**; 176-86.

10. Wolthuis R, Travo A, Nicolet C, *et al.* IR spectral imaging for histopathological characterization of xenografted human colon carcinomas. *Anal. Chem.* 2008; **80**; 8461-9.

11. Ly E, Piot O, Wolthuis R, *et al.* Combination of FTIR spectral imaging and chemometrics for tumour detection from paraffin-embedded biopsies. *Analyst* 2008; **133**; 197-205.

12. Bara J, Loisillier F, Burtin P. Correlation between the presence of gastrointestinal antigens and the histological type of human ovarian mucinous cysts. In Peeters H, ed. *Protides of biological fluids*, 27th *colloquium*. New York: Pergamon Press, 1979; 339-42.

13. Roussel P, Lamblin G, Lhermitte M, *et al.* The complexity of mucins. *Biochimie* 1988; **70**; 1471-82.

14. Parker FS. Introduction and brief theory. In Parker FS, ed. *Applications of infrared, Raman, and resonance Raman spectroscopy in biochemistry*. Plenum Press: New York, 1983; 22-3.

15. Khajehpour M, Dashnau JL, Vanderkooi JM. Infrared spectroscopy used to evaluate glycosylation of proteins. *Anal. Biochem.* 2006; **348**; 40-8.

16. Wong PT, Wong RK, Caputo TA, Godwin TA, Rigas B. Infrared spectroscopy of exfoliated human cervical cells: Evidence of extensive structural changes during carcinogenesis. *Proc. Natl. Acad. Sci. USA* 1991; **88**; 10988-92.

17. Bird B, Miljkovic M, Romeo MJ, *et al.* Infrared micro-spectral imaging: distinction of tissue types in axillary lymph node histology. *BMC Clin. Pathol.* 2008; **8**; 8.

18. Petibois C, Déléris G. Chemical mapping of tumor progression by FT-IR imaging: towards molecular histopathology. *Trends Biotechnol.* 2006; **24**; 455-62.

19. Kong J, Yu S. Fourier transform infrared spectroscopic analysis of protein secondary structures. *Acta. Biochim. Biophys. Sin.* 2007; **39**; 549-59.

FIGURES



Figure 1. Average IR spectra of mucus and mucin lyophilisated samples. *Continuous trace*, human mucus extracted from ovarian begnin cysts of 4 patients. *Dotted trace*, purified submaxillary bovine mucin type II.



Figure 2. IR color-coded spectral images of normal colon and colon adenocarcinoma tissue sections and corresponding IR signatures of spectral clusters. *A*, IR images constructed using k-means clustering of the spectral data and adjacent hematoxylin-eosin (HE) stained sections corresponding to the measured IR images; *scale bar*, 200 μ m. *B*, second derivative spectra of the clusters corresponding to the 2 cluster-based images and *C*, to the 6 cluster-based images. The spectra colors correspond to those of the IR images.



Figure 3. Specific IR spectral imaging of secreted mucus in normal colon and colon adenocarcinoma tissue sections. *A*, IR images were constructed by k-means subclustering of spectra previously attributed to secreted mucus uniquely; *scale bar*, 200 μ m. *B*, second derivative spectra of the subclusters 1 (green) and 2 (red) corresponding to the 3 subcluster-based classification; the colors correspond to those of the IR images. *C*, 3D scatter plot of normal colon (blue triangles, n=7 normal individuals) and adenocarcinoma (red circles, n=9 patients) samples according to subcluster percentages quantified from IR images. Spectral range: 950-1750 cm-1. Shaded regions overlaid on the data points serve as visual guides to the eye.



Figure 4. IR spectral-based discrimination of normal colon from colon adenocarcinomas. The 3D scatter plots were constructed in the same conditions as in Fig. 3C, but based on spectral ranges restricted to glycosylation (*A*, 1000-1200 cm-1), sialylation (*B*, 1700-1750 cm-1) and Amide I (*C*, 1600-1700 cm-1) bands. *Blue triangles*, normal colon (n=7 normal individuals); *red circles*, adenocarcinomas (n=9 patients). Shaded regions overlaid on the data points serve as visual guides to the eye.



Figure 5. Modification of mucin secondary structure in colon adenocarcinomas as evidenced by subclustering of secreted mucus IR spectra. Subclustering was processed in the Amide I spectral range (1600-1700 cm-1). Second derivative spectra of the major subclusters in normal colon (#1 and #3 in Fig. 4B) (green and purple traces) and adenocarcinomas (#2 in Fig. 4B) (blue trace).

SUPPLEMENTARY DATA

Sex	Age	Description	TNM
М	72	Lieberkühnian colon adenocarcinoma with mucinous colloide	T3, N0
F	64	Lieberkühnian colon adenocarcinoma with mucinous colloide	T2, N0
F	71	Lieberkühnian colon adenocarcinoma with mucinous colloide	T2, N1, N2
М	74	Lieberkühnian colon adenocarcinoma with mucinous colloide	T3, N0
М	68	Colon adenocarcinoma with mucosal colloide	T3, N1
М	74	Lieberkühnian colon adenocarcinoma with mucinous colloide	T3,N0
F	69	Lieberkühnian colon adenocarcinoma with mucinous colloide	T3, N1
М	45	Lieberkühnian colon adenocarcinoma with mucinous colloide	T3, N0
м	82	Lieberkühnian colon adenocarcinoma with mucinous colloide	T2, N1

Table S1. Clinical and histological characteristics of the human colon adenocarcinomas studied.Abbreviations : M = Male, F = Female, TNM = Histopathological classification.



Figure S1. Percentages of cluster surfaces according to the number of subclusters tested for the classification of the secreted mucus spectra. Spectral range: 950-1750 cm⁻¹. *Blue triangles*, normal colon; *Red circles*, colon adenocarcinomas. Average cluster percentages in normal colon (*Blue bar*) and in adenocarcinomas (*Red bar*) samples.

*, 0.05 <*P* <0.1; **, 0.01 <*P* <0.05; ***, P <0.01; *NS*, not significant.



Figure S2. Percentages of subcluster surfaces according to the number of subclusters and to the different restricted spectral ranges used for classification of secreted mucus spectra. Classification in the glycosylation (A), sialylation (B) and Amide I (C) bands. Normal colon (*blue triangle*); colon adenocarcinomas (*red circle*). Average cluster percentages in normal colon (*blue bar*) and in adenocarcinomas (*red bar*) samples.

*, 0.05 <*P* <0.1; **, 0.01 <*P* <0.05; ***, P <0.01; *NS*, not significant.

1 2 3

Subcluster Number

2

Subcluster Number

1

Adenocarcinomas	Immunohistochemical scores		Spectral subclusters (% pixels)		
n=9	MUC5AC	MUC2	1	2	3
	(a) 3	2	34,1	47,8	18,1
	1	2	24,7	61,2	14,1
	(b) O	2	43	42,2	14,8
	2	2	32	37,5	30,5
	0	2	23,5	70,4	6,1
	1	2	28,6	58,4	13
	3	2	36,1	60,6	3,3
	2	2	47,8	32,6	19,6
	0	2	17,2	74,2	8,6

Table S2. Comparison between MUC2 and MUC5AC immunohistochemical scores and subclusters percentages in corresponding mucus IR color-coded images in colon adenocarcinomas. Immunohistochemical scores range from 0 to 3 (0= no staining, 1=localized trace staining, 2=widespread staining of moderate intensity, 3=widespread staining of strong intensity). Subclustering was processed in the Amide I spectral range (1600-1700 cm⁻¹) using 3 subclusters; percentages of subclusters 1 (purple), 2 (dark blue) and 3 (green) correspond to those plotted in Fig. 4*C*. The case of adenocarcinomas (a) and (b) are illustrated in Fig. S3.



Figure S3. Representative illustration of the comparison between MUC2 and MUC5AC immunohistochemical staining and corresponding mucus IR color-coded images in colon adenocarcinomas (a) and (b), described in Table S2. Subclustering was processed in the Amide I spectral range (1600-1700 cm⁻¹) using 3 subclusters; *scale bar*, 200 μ m. MUC2 expression was observed in both samples; MUC5AC expression was observed only for adenocarcinoma (a). As shown by mucus IR color-coded images and percentages of subclusters 1 (purple), 2 (dark blue) and 3 (green), both samples present close spectral profiles.

IR Spectral Imaging for Histopathological **Characterization of Xenografted Human Colon** Carcinomas

Rolf Wolthuis,[†] Adrian Travo,[†] Céline Nicolet,[‡] Agnès Neuville,^{‡,§} Marie-Pierre Gaub,^{‡,#} Dominique Guenot,* Elodie Ly,* Michel Manfait,* Pierre Jeannesson,*** and Olivier Piot*

MéDIAN UMR CNRS 6237 MEDyC, Université de Reims Champagne-Ardenne, UFR Pharmacie, 51 rue Cognacq-Jay, 51096 Reims Cedex, France, UMR-S682 Inserm, Université Louis Pasteur, Faculté de Médecine, 67200 Strasbourg, France, Laboratoire de Biologie Moléculaire et Biochimie, CHRU Strasbourg, Höpital de Hautepierre, 67200 Strasbourg, France, and Service d'Anatomo-Pathologie, CHRU Strasbourg, Höpital de Hautepierre, 67200 Strasbourg, France

This study aims to develop IR imaging of tumor tissues for generating an automated IR-based histology. Formalinfixed paraffin-embedded xenografts of human colon carcinomas were analyzed. Chemometric and statistical multivariate treatments of spectral data permitted to probe the intrinsic chemical composition of tissues, directly from paraffinized sections without previous dewaxing. Reconstructed color-coded spectral images revealed a marked tumor heterogeneity. We identified three spectral clusters associated to tumoral tissues, whereas HE staining revealed only a single structure. Nine other clusters were assigned to either necrotic or host tissues. This spectral histology proved to be consistent over multiple passages of the same xenografted tumor confirming that intratumoral heterogeneity was maintained over time. In addition, developing an innovative image analysis, based on the quantification of neighboring pixels, permitted the identification of two main sequences of spectral clusters related to the tissue spatial organization. Molecular attribution of the spectral differences between the tumor clusters revealed differences of transcriptional activity within these tumor tissue subtypes. In conclusion, IR spectral imaging proves to be highly effective both for reproducible tissue subtype recognition and for tumor heterogeneity characterization. This may represent an attractive tool for routine high throughput diagnostic challenges, independent from visual morphology.

Histopathological characterization of tissues is a paramount step for the diagnosis of tumoral lesions and as well as for a better understanding of mechanisms associated with the tumor progression. Among the different types of cancer, colon cancer not only represents a high incidence pathology but also one of the most investigated lesions at the tissular level, from mucin analysis^{1,2} to genetic studies.3-3 Unlike classical histology mainly morphology based, or immunohistochemistry specifically targeting certain proteins, vibrational spectroscopy (Raman scattering and Fourier transform infrared (FT-IR) absorption spectroscopy) probes the chemical composition of the tissue and can detect subtle molecular structural changes. It also avoids inherent operator variability. For many different tumors, spectroscopy based histology has proved to allow discrimination between malignant or benign tissues and normal tissues6-12 as well as detection of drug resistance in cancer cell lines.13 Rigas et al.34 studied malignant versus normal colon tissues using IR spectroscopy. Focusing on colonic tissue, they found changes correlated with malignancy in specific vibrational frequencies attributed to the nucleic acids and spectral peak shifts related to loss of hydrogen bonds in cell proteins. Other authors demonstrated the diagnostic potential of FT-IR microspectroscopy applied to formalin-fixed chemically deparaffined tissues; carcinomas, polyps and normal tissues were distinguished with a

- Forgue-Lafitte, M. E.; Fabiani, B.; Levy, P.; Manrin, N.; Flejou, J. F.; Bara, (1) J Int. I. Cancer 2007, 121, 1543-49
- Byrd, J.: Bresalier, R. <u>Cancer Metatami Rev.</u> 2004, 29, 77–99.
 Guenot, D.: Guérin, E.: Aguillon-Humain, S.: Pencreach, E.: Schneider, A.: Neuville, A.; Chenard, M. P.; Dubuc, L. Du Manoir, S.; Brigand, C.; Oudet, P.; Kedinger, M.; Gaub, M. P. J. Pathol. 2006, 208, 643-52.
- (4) Weber, J. C.; Meyer, N.; Pencreach, E.; Schneider, A.; Guerin, E.; Neuville, A.; Stemmer, C.; Brigand, C.; Bachellier, P.; Rohr, S.; Kedlinger, M.; Meyer, C.; Guenot, D.; Oudet, P.; Jaeck, D.; Gaub, M. P. Int. I. Cancer 2006, 120, 524-32
- Kinzler, K. W.; Vogelstein, B. <u>Cell</u> 1996, 87, 129-70.
 De Jong, W. D.; Bakker Schut, T. C.; Maquello, K.; van der Kwast, T.; Bangma, C. H.; Kok, D. J.; Puppels, G. J. <u>Anal. Chem.</u> 2006, 78, 7761-60. (7) Lusch, P.; Haensch, W.; Naumann, D.; Diem, M. Blochim, Biothys. Acta
- 2004, 1688, 175-86, 08 Krishna, C. M.; Sockalingum, G. D.; Ehat, R. A.; Venteo, L.; Kushtagi, P.;
- Plot, M.; Marhait, M. Anal, Hornal, Chem. 2007, 187, 1649–56.
 Steller, W.; Einenkel, J.; Horn, L. C.; Braamaan, U. D.; Binder, H.; Salzer,
- R. Erzafi, C. Anal. Bioanal. Chem. 2006, 384, 145–54.
 Fernandez, D. C. Bhargava, R. Hewitt, S. M.: Levin, I. W. Nat. Historicand.
- 2005: 23, 469-74 Krishna, C. M.; Begelaer, G.; Adt, L.; Bubin, S.; Kartha, V. B.; Manfait, M.; Socialingum, G. D. <u>Honolomera</u> 2006, *82*, 462–70.
 Lasch, P.; Haensch, W.; Lewis, E. N.; Kidder, L. H.; Naumann, D. <u>And.</u>
- Sheepeng 2002, 56, 1-9. (13) Lasch, P.; Chiriboga, L.; Yee, H.; Diem, M. Technol. Cancer Res. Treat. 2002, 1, 1-7.
- (10) Rigas, B.; Morgello, S.; Goldman, I. S.; Wong, P. T. T. Proc. Nucl. Acad. Sei U.S.A. 1990, 87, 8140-44.

Analytical Chemistry, Vol. 80, No. 22, November 15, 2008 8461

^{*} To whom correspondence should be addressed. Phone: 33.3.26 91 15 67. Fax: 33 3 26 91 35 50. E-mail: pierre jeannesson@univ-reims.fr.

MéDIAN UMR CNRS 6237, Université de Reims Champagne-Ardenne. UMR/S682 Inserm, Université Louis Pasteur,

⁶ Laboratoire de Einiogie Moléculaire et Eiochimie, CHRU Strasbourg, Höpital de Hautenierre

Service d'Anatomo-Pathologie, CHRU Strasbourg, Búpital de Hautepierre,

^{10.1021/}ac801191x CCC: \$40.75 © 2008 American Chemical Society Published on Web 10/11/2006

significant over 80% accuracy.¹⁵⁻¹⁷ Lasch et al.^{7,12,18} demonstrated FT-IR microspectroscopy to be a tool for histopathological analysis by combining spectral analysis to tissue section mapping and by exploiting spectral data using statistical multivariate analysis methods. This was carried out on cryosections of colon biopsies, which permitted distinction of non-neoplastic crypt cells from adenocarcinoma cells. These studies demonstrated the strength of vibrational spectroscopy, i.e. the ability to detect, in a fully automated way, the presence of malignancy. But they also illustrated some limitations: the spectral differences, such as found by Rigas et al.,¹⁴ are hard to relate to specific biological processes and events. By choosing to reproduce classical histology with spectral imaging and multivariate data analysis, as in the study of Lasch et al.,⁷ the additional information that might be present in the spectral data, i.e. chemical composition, is not fully exploited.

New tools are required to further analyze the level of intratumor heterogeneity and its prognostic or predictive value. Thus, using a collection of xenografted colon carcinomas, our study aimed at a further exploration of the possibilities of the FT-IR approach for histological recognition, with a particular emphasis on tumor heterogeneity. The tumors analyzed are from xenografted human colon carcinomas from different patients partly described before by Guenot et al.³ Xenografting of human colon tumors in nude mice bridges the gap between cell line based models and patient material, offering the complexity of the original tumor with the experimental freedom of the cell lines. Xenografted colon tumors have been shown to maintain their three-dimensional morphology and their genetic characteristics over multiple passages.³ In addition, as in xenografting only tumor cells survive, all other cells implicitly originate from the host.

In our approach, we develop image analysis computational methods and demonstrate how general tumor growth patterns can be extracted from the IR spectral images, collected directly on paraffin-embedded tissue sections. Our findings allow access to quantitative spatially resolved information on intratumor heterogeneity that cannot be obtained by conventional histology.

EXPERIMENTAL SECTION

Xenografts. The xenografts and the xenografting protocol were described elsewhere.³ The xenograft experiments were approved by the CCPPRB d'Alsace No. 1 Strasbourg, France ethics committee. Twenty-two xenografts of colon carcinomas and liver metastasis with a staging ranging from T2N0M0 to T4N1M1, from 12 male and 7 female patients, were included in this study. The frequency of chromosomal alterations based on the allelotyping on 33 microsatellite markers ranged from 6 to 97%³ From 4 patients, 2 different passages of the xenografts were included. Clinical and histological characteristics are summarized in Table S1 in the Supporting Information. Xenografts from 9 patients (11 samples) were used to generate the predictive models for histology (so-called training set), and the other samples (9 patients, 11 samples) were used to validate the generated models (so-called validation set). The sets were chosen so that both groups are as equivalent as possible based on sex, tumor location, tumor stage and allelotyping.

IR Spectroscopy. Formalin-fixed paraffin-embedded xenografted tumors were cut to 7 µm thick slices and put on 2 mm thick ZnSe substrates (Crystal GmbH, Berlin, Germany). No attempt was made to remove the paraffin from the tissues. The consecutive slice was put on standard glass and hematoxylin cosin (HE) staining performed for conventional histology. IR spectral images were measured on a Spotlight 300 system (Perkin-Elmer Life Sciences, Courtaboeuf, France) with a spatial resolution of approximately 10 µm (6.25 µm step size). The spectral range of the spectra was 648-2000 cm⁻¹, at 8 cm⁻¹ resolution with 2 measurements per spectrum. For each slice, an area containing only paraffin was also measured at the same conditions as the tissue spectra. Water vapor spectra were measured on the Perkin-Elmer Spectrum One detector which is part of the Spotlight 300 system. For that, the sample compartment was first flushed with nitrogen to remove all traces of water and carbon dioxide. A background spectrum was measured under this condition. The flushing was then stopped and air from the surrounding was allowed to diffuse into the sample compartment. During that rise in humidity, measurements were taken continuously until to reach a stable end point; no change was observed any longer. The experiment was repeated several times, leading to a total number of 58 spectra collected. The paraffin and water vapor spectra were used to correct the tissue spectra for the presence of these interferences. In addition, all the experiments were performed at constant room temperature (22 °C).

The total number of spectral images collected was 68, with, in average, an image size of 135,000 spectra, covering a substantial part of each slice. This whole data set comprises more than 9 million spectra.

Data Processing, Extended Multiplicative Signal Correction (EMSC)-Based Preprocessing, Several sources and factors contribute to the measured IR spectra. Not all contributions are equally relevant; some of them even mask the information one is looking for. The most important ones for this study are listed below:

Atmospheric water vapor gives narrow absorption lines in the spectral ranges 1300–1900 cm⁻¹. The intensity depends mainly on the humidity.

The paraffin used to embed the tissue gives bands at 888, 1377 and 1464 cm⁻¹. Each band may be composed of several overlapping bands and the relative intensity of these bands may be influenced by the paraffin embedding protocol or by the presence of other material such as tissue.

Scatter effects, such as Mie scatter, and interference effects give rise to wavelength dependent backgrounds in the spectra.

The sample thickness is directly related to the infrared absorption according to the Lambert-Beer law. But the sample thickness is not necessarily related to the sample composition.

EMSC is an extension of multiplicative scatter correction (MSC) introduced in 1983 for near-infrared spectra processing.^{10,80} MSC, further developed for IR data, offers an integrated approach

⁽¹⁵⁾ Argov, S.; Sahnan, J. R. A.; Goldstein, I. S. J.; Guterman, H.; Mordechai, S. I. Biomed. Opt. 2002, 7, 248–54.

⁽¹⁶⁾ Sahu, P. K. Argov, S.; Bernshtain, F.; Salman, A.; Wallisch, S.; Goldstein, J.: Mordechai, S. Scand, J. Crystroenterol. 2004, 6, 357–66.

⁽¹⁷⁾ Salman, A.; Argov, S.; Ramesh, J.; Goldatein, J.; Sinelnikov, I.; Gaterman, H.; Mordechai, S. Cell Mol. Biol. 2001, 47, 159–64.

⁽¹⁸⁾ Lasch, P.; Wäsche, W.; McCarthy, W. J.; Muller, G.; Naamann, D. in *Infrared apectroscopy: new feel in medicine*, Mantsch H. H., Jackson, M., Ed.: Proc SPIE, Bellinghan, WA, 1998; 3257, pp 187–98.

⁸⁴⁶² Analytical Chemistry, Vol. 80, No. 22, November 15, 2008

⁽¹⁹⁾ Martens, H.; Jonsen, S. A.; Geladi, P. Multivariate linearity transformation for near-infrared reflectance spectrometry; Christie, O. H. J., Ed.: Proc. Nordic. Symp. Applied Statistics, Stokkand Forlag Publishers: Stavanger, 1983; pp 205–34.
to minimize the influence of all these factors.^{21,00} In our study, we have adapted EMSC to the processing of paraffined tissue IR data. Briefly, each spectrum s, considered as a vector, in the data set can be modeled as a linear combination of known interference spectra, estimates of backgrounds and an estimate of the data set:

$$\mathbf{s} = a\mathbf{\hat{s}} + \mathbf{bI} + \mathbf{cP} + \mathbf{e} \tag{1}$$

with a (scalar), **b** and **c** (both vectors) regression coefficients, **š** an estimate of the data set, **I** the matrix of known interference spectra, **cP** a polynomial modeling background and offsets due to scatter effects and interferences and **e** the model error. If there is not a good estimate of the spectrum **ŝ** available, one could choose the average of the whole data set for **ŝ**.

The corrected spectrum is now calculated as

$$s_{EMSC} = e/a + \hat{s}$$
 (2)

Thus, the variance due to interference and scatter effects is removed and the spectra are normalized to have comparable intensity as Å. In the case of a large and complicated set of interference spectra, this set can be modeled using a principal component analysis (PCA) on non-mean-centered data. This simplifies the model of interferences and increases the numerical robustness of the calculation.

Histology by Spectroscopy: k-Means and Linear Discriminant Analysis (LDA). A k-means algorithm²² was applied on a subset, so-called training set, of the measured spectral images to find spectral clusters. This clustering was compared with the HE stained sections to make a histological assignment for each cluster. The number of clusters was chosen so that each cluster has a unique histological assignment; multiple clusters with the same histological assignment were allowed. This k-means classification was subsequently used to generate a predictive LDA model enabling the classification of samples not included in the training set. Prior to the LDA, the dimensionality of the data was reduced by PCA. The LDA model was validated using the samples in the validation set.

Software Platform. Data were processed using Matlab 2006a (Mathworks, Natick, MA) based software. Routines were partly written in house and partly written at the Centre for Optical Diagnostics and Therapy of the Erasmus Medical Centre (Rotterdam, NL). The used LDA routines were from the Discriminant Analysis Toolbox 0.3, an implementation of the algorithms as published by Ripley.²⁴ The PCA routine was from the PLS Toolbox version 1.5.3 of Eigenvector (Wenatchee, WA).

RESULTS

Preprocessing of Spectra of Parafilin-Embedded Sections of Human Colon Xenografis. Spectra were cut to a spectral range of 724–1720 cm⁻¹ which is assumed to be the most

- (20) Geladi, P.: McDougall, D.: Martens, H. <u>Appl. Spectrosc</u> 1985, 39, 401– 500.
- (21) Martens, H.; Stark, E. J. J. Pharm. Biomed. Anal. 1991, 9, 625-33.
- (22) Kohler, A.; Kirschner, C.; Oust, A.; Martens, H. <u>Appl Spectrue</u>, 2005, 59, 707–716.

informative spectral region for biological materials.26 As explained in the materials and methods, paraffin and water vapor are significant undesirable contributions to the spectra measured. As an estimate of the measured spectra, necessary for the EMSC correction, the mean of the spectra in the training set was taken. Paraffin proved to give spectra with a highly variable background, probably due to interferences. Instead of using a single spectrum for paraffin only, it was judged that the first 8 principal components (PC) of a set of spectral images of paraffin would give a better description of the variations encountered. More elaborate discussion about the model construction was given by Ly et al.26 For this experiment, a number of 8 different slices, measured during different measurement sessions, provided a total of 255,000 spectra. This set of spectra was reduced to 8 PC after scaling each spectrum to zero mean and unit variance. This set of 8 PC, obtained from non-mean-centered data, was included in the paraffin model to capture the variation of experimental conditions, Paraffin spectra with high backgrounds (above 0.35 absorbance) and spectra with a maximum absorbance less than 0.3 (low signal) of the main paraffin band or exceeding 1 (saturation of the detector) were considered aberrant and excluded from the calculation. In addition to the paraffin PC, the first 2 PC of a set of 58 water vapor spectra were added as well to take into account the day to day variation of atmospheric water vapor. This set of PC was used as interference spectra in the EMSC.

First results (not shown) indicated that, if paraffin and tissue were found in conjunction, the relative intensity of spectral bands of paraffin may change in a way not covered by the paraffin principal components. These spectral features could be due to molecular interactions between paraffin and tissue.⁴⁷ To compensate for that, a set of Gaussian curves, calculated from a band decomposition of the spectral bands of paraffin, was added to the set of interference spectra. The full set of interferences taken into account is shown in Figure 1A and B. Finally, to model offsets and Mie scatter effects resulting in slowly curving backgrounds, a fourth order polynomial was included as well. This approach was used, instead of that described by Mohlenhoff et al.⁴⁸⁰ or Kholer et al.⁴⁸⁰ because it permits to speed up the calculation, and was proved to be effective on the spectral region of interest (724–1720 cm⁻¹).

Tissue spectra to be included in the data analysis were selected, based on the three following rules: 1. The sum of squares of e, a measure for the fit error, must be smaller than an empirical cut off. Effectively noisy spectra and spectra with very deviant shapes are excluded based on this rule. 2. The first fit coefficient must be larger than a certain value larger than zero. It is the requirement to ensure that a spectrum contains a significant amount of tissue signal. 3. The other fit coefficients must be inside a range \pm 3 times the standard deviation around the mean for

⁽²³⁾ Jain, A. K.: Dubes, R. C. Clustering methods and algorithms. In Algorithms for clustering data: Prentice Hall: Englewood Cliffs, 1988; pp 55–142.

⁽²⁰⁾ Ripley, B. D. Linear discriminant analysis. In Pattern recognition and neural networks: Cambridge University Press: Cambridge, 1996, pp 91–120.

⁽²⁵⁾ Parker, F. S. Applications of infrared. Reman, and resonance Ruman spectruscopy in blockemistry, Plenum Press, New York, 1983.

⁽²⁶⁾ Ly, E. Fiof, O.; Wolthuis, R.; Durlach, A.; Bernard, P.; Manfait, M. <u>Analyst</u> 2008, 153, 197–205.

⁽²⁷⁾ Vrable, V.; Gobinet, C.; Fiot, O.; Tizyli, A.; Bernard, P.; Huez, R.; Manfait, M. Biomed. Signal Process. Control 2007, 2, 40-50.

⁽²⁸⁾ Mohlenhoff, E.: Romen, M.; Diem, M.; Wood, R. R. <u>Biophys. I</u>, 2005, 88, 3835–40.



Figure 1. Preprocessing of IR spectra of formalin-fixed paraffin-embedded tissue sections of xenografted human colon carcinomas. The preprocessing was performed using extended multiplicative signal correction (EMSC) as described in the Experimental Section. A. From top to bottom, the first 8 principal components of paraffin signal used as interference spectra in the EMSC model. B. From top to bottom, 6 additional Gaussian curves modeling paraffin bands and the first 2 principal components of water vapor spectra. C. Untreated raw spectra collected on a tissue section (average spectrum drawn in bold line). D. Spectra after EMSC: backgrounds are removed, variabilities due to paraffin and water vapor are reduced; and outlier spectra without tissue features are removed.



Figure 2. IR color coded spectral images of a tissue section of xenografted colon carcinoma. A. IR image from a tissue sample (sample #212; cf. Table S1 in the Supporting Information) constructed using k-means clustering of the spectral data after EMSC preprocessing as shown in Figure 1. B. Hematoxylin ecsin (HE) stained section corresponding to the measured IR image. C. IR image of the same tissue section constructed by predictive LDA (linear discriminant analysis) model. Note slight differences between k-means clustering (A) and LDA (C). Scale bar, 400 µm.

each coefficient. This rule also excludes spectra with very deviant shapes.

The effect of the preprocessing is shown in Figure 1C and D: Spectra are normalized, backgrounds are corrected for, interferences are minimized and outliers (spectra that fail toward the three rules) are removed. Construction of Spectral Images and Correlation with HE Staining. Spectra of the training set were clustered into 12 clusters using a k-means algorithm. The k-means algorithm partitions the data in the given number of clusters, so that the sum of the distances of the data points to their respective cluster centers are minimized. The method is, apart from the number of clusters, unsupervised. The algorithm is especially suited to cluster large data sets such as the spectral images in this study.

⁽²⁹⁾ Kohler, A.; Suk-Susa, J.; Sockalingum, G. D.; Tobin, M.; Bahrami, F.; Yang, Y.; Pijanka, J.; Dumas, P.; Cotte, M.; Martena, H. <u>Appl. Spectrosc</u>, 2008, 62, 259–266.

⁸⁴⁶⁴ Analytical Chemistry, Vol. 80, No. 22, November 15, 2008



Figure 3. IR signatures of spectral clusters determined by k-means classification. A. Dendrogram obtained by hierarchical classification analysis of the spectral clusters. The dendrogram allows assessment of the interspectral distances (calculated using Ward's algorithm). B. Average spectra of the clusters. Second derivative spectra are shown to enhance the spectral features; the colors correspond to the colors of the cluster images (Figure 2A). C. Histological assignment of the 11 spectral clusters. The percentages indicate the proportions of pixels belonging to each cluster. These percentages were calculated from the set of IR images constructed by k-means classification on the samples of the training set.

Based on the k-means clustering, pseudocolor images were constructed by assigning a different color to each cluster (Figure 2A). White areas correspond to spectra (outliers) omitted in the analysis based on the rules listed above. The pseudocolor images allow a comparison with the HE stained slices (Figure 2B). A specific histological structure could be assigned to each cluster using all images in the training set. A number of 12 clusters allows for a unequivocal reproduction of HE based histology.

In Figure 3, the assignment to histology as identified from HE stained slices is given together with the average spectrum of each cluster, as well as a dendrogram visualizing the distance between the k-means cluster averages. Spectra are plotted as second order derivatives to enhance spectral differences. The dendrogram was calculated on the Euclidean distance between the cluster averages using Ward's algorithm for hierarchical cluster analysis.

Based on the generated images, the following histological assignment was made: 3 clusters for tumor tissue, 4 clusters for stromal tissue, 4 clusters for necrotic tissue and 1 cluster for excreted mucus. At this point, it has to be reminded that all stromal tissues originate from the host. Interestingly, it was found that the tumor clusters 5, 9 and 10 may appear simultaneously in a single sample, thus revealing an intratumor heterogeneity. Branch 1A of the dendrogram indicates that the two tumor clusters 5 and 9 are the closest and linked to cluster 4 assigned to stroma.

These three clusters represent 62% of the measured spectra in the training set of the samples. In addition, inside branch 1A, tumor cluster 10 is separated from the other tumor clusters 5 and 9. Its spectral signature in the range 1000–1100 cm⁻¹ reveals the presence of mucin³⁰ confirming the assignment of this cluster to tumor areas containing mucus producing cells. Branch 1B is composed of cluster 3 (accumulated mucus without cells) and of a subcluster formed by clusters 2 and 12 (both necrotic tissue). Spectra corresponding to the necrotic tissue belong to separate branches (clusters 1 and 6 in branch 1A, clusters 2 and 12 in branch 1B) reflecting different degrees of necrosis. Besides branch 1, the dendrogram is composed of branch 2, containing clusters 8 and 11 both assigned to stroma.

In conclusion, these data demonstrate the presence of several clusters with the same histological assignment. This reveals a marked heterogeneity in the biochemical composition of each tissue subtype, supporting additional levels of tissue organization, that are undetectable by conventional HE staining.

Training and Validation of IR-Imaging-Based Histology. In order to automatically classify spectra from new samples according to the established cluster membership, LDA²⁴ a supervised classification method, was used to transform the k-means results into a predictive model. Besides classification, LDA also gives the probability that a spectrum belongs to a particular class. A spectrum is then assigned to the cluster with the highest probability. In addition, this probability information can be useful to assess the relevancy of the classification obtained. Together with the established histological assignment of the k-means clusters, the LDA model allows for an IR-imaging-based histology for new samples. The LDA model has to be validated using samples from the validation set. This validation on independent data is necessary to prove that the calculated model performs well. For computational reasons, the LDA model was built using 10% of the available spectra in the training set; every tenth spectrum was included. The EMSC preprocessing described before was applied to the spectra. In addition, the dimensionality of the chosen data set was reduced by applying PCA and taking only the first 25 PC into account.

⁽³⁰⁾ Khajehpour, M.: Dashnau, J. L.: Vanderkooi, J. M. <u>Anal. Hockem</u>. 2005, 348, 40–48.



Figure 4. IR color-coded images of two samples of the validation set obtained by the LDA model. A. IR image of the sample #250. B. HE stained section corresponding to the calculated image in A. C. IR image of the sample #314. D. HE stained section corresponding to the calculated image in B. (See Table S1 in the Supporting Information for description of samples #250 and #314). Scale bar, 400 µm.

First, the validation of the LDA model consisted in comparing the outcomes of the LDA with the k-means results. As shown in Figure 2A and C, the LDA predictive model leads to reconstruct a color coded image highly similar to the k-means image. Only slight discrepancies appear mainly on the margins of the clusters, corresponding to pixels with a probability less than 0.75 of belonging to the assigned cluster indicating that these margin pixels have intermediate spectral signatures. Usually, the other considered pixels have a probability higher than 0.9.

Second, the predictive value of the LDA model was assessed by comparing the LDA results of the validation set with the HEbased histology. Data of Figure 4 and of Figures S1–S4 in the Supporting Information indicate that the LDA predicted histology for the validation set was totally consistent with the conventional staining procedure. Repeated measurements on the same sample area with a time interval of a few weeks proved the repeatability of the IR measurements and the data processing (data not shown). Figure S5 in the Supporting Information depicts the relevancy of the spectral clustering by showing the class membership probabilities for the two closest clusters (#5 and #9). This significance of clusters 5 and 9 was demonstrated for three characteristic samples: #241 and #314 presenting mainly the clusters 5 and 9 respectively, and the mixed sample #250 containing both clusters.

In addition, since the validation set included different passages from the same xenografted tumor, the IR analysis of these samples indicate stability over sequential passages (Figure 4). This stability at the level of the spectral-based histology confirms previous results showing that the xenografted tumors are stable, not only in their morphology but also in the chromosomal alterations.³

8466 Analytical Chemistry, Vol. 80, No. 22, November 15, 2008

Indeed, the spectral images show that the relative contribution of the tumor clusters 5, 9 and 10 is similar between the different passages.

Spectral Analysis of the Tissue Organization and Identification of Tissue Patterns. As illustrated by the assignment of two clusters (5 and 9) to tumor tissue, tissue organization not revealed by HE stained images, was observed on IR images. This organization was analyzed quantitatively by applying image analysis methods at the pixel level and by analyzing the spectral differences. The data of the validation set and the training set were pooled for this analysis and the cluster assignment was calculated using the LDA model. White areas in the images were omitted in the calculation. For this approach, histograms (Figure 5A) presenting the chance for a given cluster to find another cluster in its neighborhood were calculated as follows: for each pixel in the images, the cluster membership of the direct neighbors (left, right, above, under) was determined. If a neighboring pixel belonged to the same cluster, the event was omitted. The result per cluster was normalized and plotted in a set of histograms. Each histogram gives the probability distribution for the cluster membership of the neighboring cluster. For example: cluster 5 borders, on average, for 32% with cluster 4 and 63% with cluster 9 but hardly with any other cluster. From these histogram columns, sequences or morphological patterns were characterized by using the following rules: start from the column of a particular cluster (i.e., cluster 4), determine the cluster with the highest probability of bordering (i.e., cluster 5), go to the column of this last cluster and determine in this column the most probable bordering cluster (i.e., cluster 9) as previously (if this cluster is the same as the preceding one, take the second highest) and continue until an end point is reached. For example, starting from this cluster 4, the most abundant stroma cluster, this gives the following sequence: $4 \rightarrow 5 \rightarrow 9 \rightarrow 1 \rightarrow 12$.

Using all clusters as a starting point, 12 different sequences can be found. Comparing all the sequences, it is observed that they all have the sequence $4 \rightarrow 5 \rightarrow 9$ in common. Considering that stroma cluster 8 borders almost exclusively with cluster 4, the full sequence may start with cluster 8: $8 \rightarrow 4 \rightarrow 5 \rightarrow 9 \rightarrow 1 \rightarrow 12$. This pattern suggests that stroma clusters 4 and 8 and tumor cluster 5 are not in direct contact with necrotic tissues 1 and 12, but always buffered by tumor cluster 9. Another sequence of interest, however less probable, is the one involving cluster 10 as it is also a tumor tissue and mucus cluster 3 as it represents the mucus producing tumors: $8 \rightarrow 4 \rightarrow 5 \rightarrow 9 \rightarrow 10 \rightarrow 3$. These schematic patterns can be visualized in simple diagrams (Figure 5B).

The spectral origins of the sequence involving the predominant tumor clusters (5 and 9) and neighboring clusters (1 and 12 assigned by HE staining as necrotic tissue) were also investigated (Figure 6). For this, difference spectra between the successive clusters $5 \rightarrow 9 \rightarrow 1 \rightarrow 12$ were calculated in such a way that no negative spectral features appear. In addition, to better appreciate the differences, second derivatives of these difference spectra are shown as well. The difference spectrum between clusters 5 and 9 (blue spectrum) puts in evidence bands at 996, 1116, 1240 cm⁻¹ (red vertical lines), associated with RNA²⁶ This observation indicates that cluster 5 is richer in RNA, which is consistent with a higher transcriptional activity in cluster 5 than in cluster 9.



Figure 5. Spatial ordering of the spectral clusters. A. Histograms of the cluster bordering. The percentages of neighboring pixels were calculated from IR images constructed by the LDA model and by considering the totality of the analyzed samples (training and validation sets). The histogram colors correspond to the cluster colors, recalled on the color scale. The contribution of each cluster as a percentage of the total number of spectra is also indicated. B. Schematic representation of two main tumor tissue patterns including tumor clusters #5, 9 and 10. The rules used to characterize the sequences of these morphological patterns are indicated in the Results section, paragraph "Spectral Analysis of the Tissue Organization and Identification of Tissue Patterns". The first pattern (left) includes necrotic tissues as observed in Figure S4A in the Supporting Information, the second pattern (right) includes mucinous tissue as observed in Figure 2C.

Concerning DNA, specific vibrations do not appear on this difference spectrum indicating a constant content in DNA between the two tumor clusters 5 and 9. The difference spectra between clusters 9 and 1 (green spectra) and clusters 1 and 12 (red spectra) exhibit vibration bands specific of DNA (bands at 1055, 1088 and 1718 cm⁻¹).³³ The decrease of DNA bands indicates a progressive loss of DNA between the tumor cluster 9 and the necrotic tissue clusters 1 and 12.

In conclusion, both the analyses of cluster sequences and of spectral differences via IR imaging allow measures of intrinsic molecular changes and reveal a marked heterogeneity in the tumor tissue organization, otherwise not identifiable by conventional histology.

DISCUSSION

In our study, an IR-imaging-based histology of human xenografted colon tumors was established using k-means clustering. Based on this k-means clustering, a predictive LDA model was generated allowing the prediction of the histology of new samples. Using this approach, paraffin-embedded tissue sections can be segmented into identifiable regions of differing pseudocolors. This segmentation was obtained by classification of spectral data into 12 clusters, each of them representing a set of spectra with similar vibrational features and assigned to a single histological structure. On this basis, the proposed IR-based histology reflects, at the micrometric scale, the intrinsic chemical composition of the tissue and consequently brings additional information; it does not require dyes or molecular probes unlike the conventional morphologybased histology.

To our knowledge, this spectroscopic approach is the first to carry out direct spectral analysis of formalin-fixed paraffinembedded specimens, without requiring chemical dewaxing. This method of numerical dewaxing avoids the tedious and timeconsuming experimental steps of deparaffining that necessitates highly toxic organic solvents. In addition, compared to spectroscopic investigations led on cryosections," our analysis allows longterm retrospective studies that involve mostly paraffin-embedded archival samples. This opens the way to detection of spectral markers specific of the early stages of disease and with a prognosis value. The correction of paraffin spectral interferences takes into account the high variability of the IR paraffin signal by modeling it by means of several components (principal compo-

⁽³¹⁾ Arakawa, H.; Neault, J. F.; Tajmir-Riahi, H. A. <u>Biopkys. J</u>. 2001, *n1*, 1580– 87.

postdoctoral fellowship from INCa, We thank Sylvie Record for linguistic assistance. A.T. is a recipient of a doctoral fellowship from Ministère de l'Enseignement Supérieur et de la Recherche.

SUPPORTING INFORMATION AVAILABLE

Clinical and histological characteristics of the studied human colon tumors in Table S1. In Figures S1–S4, IR color-coded images, constructed with the predictive LDA model, and corresponding HE staining sections of various samples, including two xenografis from the same tumor after different passages (Figure S1). In Figure S5, class membership probabilities for clusters #5 and #9. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review June 12, 2008. Accepted September 11, 2008. AC801191X

Adrian TRAVO

Caractérisation de marqueurs du développement tumoral par microspectroscopie Infra-Rouge

Th. Pharm. Univ. : Reims : 2009

Résumé :

La caractérisation histopathologique de tissus, étape clef du diagnostique des lésions tumorales ainsi que de leur prise en charge clinique, est principalement basée sur une analyse morphologique présentant d'importantes limitations. Parmi les méthodes développées dans le cadre de l'amélioration de cette caractérisation, la spectroscopie Infra-Rouge (IR) suscite un intérêt certain par sa capacité à sonder globalement la constitution biomoléculaire d'un tissu et d'en détecter de subtils changements. Dans ce manuscrit, nous avons cherché à approfondir les capacités de la spectroscopie IR pour la caractérisation histopathologique de tissus tumoraux paraffinés. Les travaux présentés démontrent que la spectroscopie IR couplée à des méthodes de traitements multivariées des données, peut fournir des informations sur l'hétérogénéité biomoléculaire inter- et intra-tumorale. Ces informations, non visibles par examen morphologique, permettent alors l'obtention d'une nouvelle génération d'histologie potentiellement plus informative que l'histologie conventionnelle. Nous avons également démontré que ce nouveau type d'histologie peut permettre l'identification de nouveaux types de marqueurs moléculaires. En effet, une discrimination entre tissus de colons sains et tumoraux a été mise en évidence sur la base de variations de la structure secondaire des mucines secrétées. Ainsi, ces travaux démontrent le fort potentiel de ce nouveau type d'histologie IR pour l'amélioration de la caractérisation des lésions tumorales et, par conséquent, de la prise en charge des patients.

Mots clés : Colon, adénocarcinomes, histologie, hétérogénéité tumorale, mucus, mucines, imagerie infrarouge, traitements multivariés

JURY

Pr. Gérard DELERIS (Bordeaux)
Pr. Marie-Christine RIO (Strasbourg)
Dr. Dominique GUENOT (Strasbourg)
Pr. Marie-Danièle DIEBOLD (Reims)
Pr. Dieter NAUMANN (Berlin)
Pr. Pierre JEANNESSON (Reims)
Pr. Michel MANFAIT (Reims)

<u>Adresse de l'auteur :</u> Mr. Adrian TRAVO – 24, rue de Bitche - 67000 STRASBOURG

Adrian TRAVO

Caractérisation de marqueurs du développement tumoral par microspectroscopie Infra-Rouge

Th. Pharm. Univ. : Reims : 2009

<u>Résumé :</u>

La caractérisation histopathologique de tissus, étape clef du diagnostic des lésions tumorales ainsi que de leur prise en charge clinique, est principalement basée sur une analyse morphologique présentant d'importantes limitations. Parmi les méthodes développées dans le cadre de l'amélioration de cette caractérisation, la spectroscopie Infra-Rouge (IR) suscite un intérêt certain par sa capacité à sonder la constitution biomoléculaire d'un tissu et d'en détecter de subtils changements. Dans ce manuscrit, nous avons cherché à approfondir les capacités de la spectroscopie IR pour la caractérisation histopathologique de tissus tumoraux paraffinés. Les travaux présentés démontrent que la spectroscopie IR couplée à des méthodes de traitements multivariées des données, peut fournir des informations sur l'hétérogénéité biomoléculaire inter- et intra-tumorale. Ces informations, non visibles par examen morphologique, permettent alors l'obtention d'une nouvelle génération d'histologie potentiellement plus informative que l'histologie conventionnelle. Nous avons également démontré que ce nouveau type d'histologie peut permettre l'identification de nouveaux types de marqueurs tumoraux. En effet, une discrimination entre tissus de colons sains et tumoraux a été mise en évidence sur la base de variations de la structure secondaire des mucines secrétées. Ainsi, ces travaux démontrent le fort potentiel de ce nouveau type d'histologie IR pour l'amélioration de la caractérisation des lésions tumorales et, par conséquent, de la prise en charge des patients.

Mots clés : Colon, adénocarcinomes, histologie, hétérogénéité tumorale, mucus, mucines, imagerie infrarouge, traitements multivariés

JURY

Rapporteurs :	Pr. Gérard DELERIS (Bordeaux)
	Pr. Marie-Christine RIO (Strasbourg)
Examinateurs :	Dr. Dominique GUENOT (Strasbourg)
	Pr. Marie-Danièle DIEBOLD (Reims)
	Pr. Dieter NAUMANN (Berlin)
Directeurs de thèse :	Pr. Pierre JEANNESSON (Reims)
	Pr. Michel MANFAIT (Reims)

<u>Adresse de l'auteur :</u> Mr Adrian TRAVO – 24, rue de Bitche - 67000 STRASBOURG