THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Sciences des Agroressources

Présentée et soutenue par :

M. SIMON DUTHEN

le vendredi 10 décembre 2021

Titre :

Etude de la caractérisation de matières collagéniques pour spectroscopie Infrarouge. Mise au point et développement d'un système d'analyse en mode dynamique par l'industrie de la Gélatine.

Ecole doctorale :

Sciences de la Matière (SDM)

Unité de recherche :

Laboratoire de Chimie Agro-Industrielle (LCA)

Directeurs de Thèse :

M. JEAN DAYDE

MME CHRISTINE RAYNAUD

Rapporteurs :

MME NATHALIE DUPUY, AIX-MARSEILLE UNIVERSITE

MME VERONIQUE SAINTE-LHOUTELLIER, INRA CLERMONT FERRAND

Membres du jury :

M. JEAN-DENIS BAILLY, ECOLE NATIONALE VETERINAIRE DE TOULOUSE, Président

MME CECILE LEVASSEUR-GARCIA, EI PURPAN, Membre

M. DOMINIQUE BERTRAND, INRA NANTES, Membre

MME CHRISTINE RAYNAUD, TOULOUSE INP, Membre



Tout ce qui a un commencement a une fin. Sois en paix avec cette vérité et tout ira bien.

(Bouddha, 566-480 av. JC)

Remerciements

Tout d'abord je souhaite remercier Véronique Lhoutellier et Nathalie Dupuy d'avoir accepté de relire cette thèse et d'en être rapportrices. Je tiens à remercier Jean-Denis Bailly d'avoir accepté d'être Président du jury. Je remercie également M. Dominique Bertrand d'avoir accepté d'assister à la présentation de ce travail en tant qu'examinateur officiel et rapporteur officieux. La version finale de ce mémoire a bénéficié de leur lecture très attentive et d'échanges enrichissants.

Je voudrais remercier Jean Daydé, Directeur de cette thèse, pour cette opportunité de recherche à l'interface entre académique et industrie ainsi que son soutien dans ce projet partenarial. Je remercie Christine Raynaud, d'avoir co-encadré ce travail de thèse avec bienveillance sur des travaux bien plus loin de l'analyse des composés volatils prévue initialement. C'est à partir de la mise confiance qu'elle a impulsée lors de mon stage de Master que j'ai pu me lancer dans l'aventure du doctorat vos côtés. Je tiens à remercier Cécile Levasseur, porteur du projet, pour son approche ignacienne qui m'a permis d'atteindre une grande autonomie dans la gestion de ce projet et plus personnellement une meilleure connaissance de soi. J'exprime toute ma gratitude à Didier Kleiber qui s'est toujours intéressé à l'avancée de mes travaux, pour les nombreuses discussions que nous avons eu ainsi que ses conseils qui m'ont rendu plus endurant et moins sprinteur dans ce travail de thèse.

Je remercie la région Occitanie d'avoir soutenu financièrement ce projet CLE (Contrat Laboratoire Entreprise) et qui a ainsi permis de soutenir la Recherche d'une entreprise régionale. Cette thèse n'aurait jamais été possible sans la problématique industrielle et le financement de l'entreprise Weishardt. Je souhaite exprimer toute ma gratitude à l'entreprise en les personnes de Michel Guitart et Richard Biarrotte, porteurs de ce projet, pour leurs intérêts dans son avancement ainsi que les moyens qu'ils ont mis en œuvre pour ma réussite. Ce fut un plaisir de travailler avec vous et d'avoir donné un sens à la recherche du laboratoire. Je tiens également à remercier l'ensemble des salariés de l'entreprise pour leur accueil enthousiaste à chaque fois que je les ai rencontrés.

Je tiens à remercier ensuite les laboratoires d'accueil. Je remercie principalement le Laboratoire d'agrophysiologie-agromolécules de l'Ecole d'Ingénieurs de Purpan qui m'a accueilli pendant 80% du temps. Je remercie Frédéric Violleau en tant que Directeur de Laboratoire pour avoir soutenu ce projet de recherche et participé à ma formation de chercheur. Je le remercie également dans sa qualité de chercheur pour avoir enrichi ce travail avec son expertise et sa méthodologie AF4, à la pointe de la Chimie Analytique, qu'il a partagée et permis d'intégrer dans ce projet. Je remercie Mireille Gaucher pour sa gestion de la vie au laboratoire de Chimie qui rend le travail bien plus agréable et son soutien technique sans lequel la thèse aurait été bien plus compliquée. Je remercie également les membres de l'équipe des Enseignants-Chercheurs pour avoir partagé leurs expériences et conseils ainsi que Mickaël et Christel pour leurs conseils techniques et leur soutien moral dans cette aventure. Je tiens également tous les thésards (Sophie, Eugénie, Marielle, Eva pour les conseils parfois trop féminins et Romain, Jérémy, Kévin, Ugo et plus récemment Sylvain pour leur soutien dans ce monde de femmes !). Je remercie également l'ensemble du personnel du Laboratoire de Chimie Agro-Industrielle de l'ENSIACET qui m'a accueilli dans leurs locaux. Je remercie plus particulièrement Carlos Vaca-Garcia pour m'avoir permis d'initier un partenariat de recherche avec l'Université de Nagoya et d'avoir ainsi enrichi mon expérience professionnelle. Je remercie Brigitte Dubreuil pour son expertise en Spectroscopie Infrarouge et Raman. Je remercie également Anne Lung pour son expertise et son soutien technique pendant mon Master et mon Doctorat. Enfin je remercie l'ensemble des thésards « du début » pour m'avoir montré la voix !

Je remercie Jean-Michel Roger et le Pr Tsuchikawa avec qui j'ai eu la chance de pouvoir travailler en France comme au Japon. Leurs expertises en spectroscopie et chimiométrie m'ont permis de progresser dans ses domaines. Merci pour leurs accueils dans leurs laboratoires respectifs.

Ce travail de thèse a grandement profité des travaux des stagiaires alors merci à Adil, le tout premier, Céline, Théo et Chloé. Merci pour votre aide, votre implication et vos résultats dans ce projet.

Enfin, on a beaucoup parlé du soutien professionnel mais la thèse, c'est aussi un grand soutien personnel qui permet d'avancer dans les moments difficiles alors je finis par remercier ma famille, mon père (je ne te laisserai pas seul dans la famille avec le grade de docteur), ma mère (ma psy du dimanche), mon frère (pour tout l'ordre que tu apportes dans ma vie ;)), pour mon grand-père et ma grand-mère qui ont surement trouvé ces 3 années longes mais qui m'ont soutenu dans les rares moments où j'ai quitté la paillasse pour venir les voir. Thierry, merci de ta curiosité pour mon travail et ton aide dans les dernières phases de relecture. Merci à la Famille formidable que j'ai.

Et pour finir, la thèse c'est aussi des rencontres et pour moi la plus belle c'est Hélène que je ne remercierais jamais assez pour son soutien notamment dans les derniers moments…

Table des matières

REMERCIEMENTS 5

TABLE DES FIGURES 11

TABLE DES TABLEAUX 15

LISTE DES ABREVIATIONS 17

INTRODUCTION GENERALE 19

CHAPITRE I. ETAT DE L'ART 23

I.1. APPARITION DE LA GELATINE DANS LES INDUSTRIES 25

I.2. LA GELATINE UNE STRUCTURE PROTEIQUE COMPLEXE 26

I.2.1. LE COLLAGENE ET LA GELATINE : DES POLYMERES PROTEIQUES 26

I.2.1.(1). Rappel sur la structure des protéines 26

I.2.1.(2). Localisation du collagène dans les tissus de mammifères 29

I.2.1.(3). Structure du collagène 32

I.2.2. LA GELATINE 34

I.2.2.(1). Les sources de production de gélatine 34

I.2.2.(2). Structure et composition de la gélatine 35

I.2.2.(3). Propriétés physico-chimiques de la gélatine 38

I.3. PRODUCTION INDUSTRIELLE DE GELATINE 40

I.3.1. PROCEDES D'EXTRACTION DE LA GELATINE 40

I.3.1.(1). Le prétraitement de la matière première 40

I.3.1.(2). La phase d'extraction par hydrolyse 42

I.3.1.(3). La phase de raffinage 42

I.3.2. INDUSTRIE DE LA GELATINE 43

I.3.2.(1). Contexte mondial de la production de gélatine 43

I.3.2.(2). Propriétés technologiques de la gélatine 44

I.3.2.(3). Cas de la gélatine dans l'industrie 45

I.4. LA SPECTROSCOPIE INFRAROUGE 46

I.4.1. GENERALITES SUR LA SPECTROSCOPIE INFRAROUGE 46

I.4.2. LES DIFFERENTS TYPES DE SPECTROMETRES INFRAROUGE 47
I.4.3. MODE D'ACQUISITION 48
I.4.4. L'APPLICATION DE LA SPECTROSCOPIE PROCHE INFRAROUGE EN AGRO-ALIMENTAIRE 49
I.4.5. LA CHIMIOMETRIE POUR ANALYSER LES SPECTRES PROCHE INFRAROUGE 52
I.4.5.(1). Classement des méthodes chimiométriques 52
I.4.5.(2). Développement d'un modèle de prédiction infrarouge 56
I.5. ASYMMETRICAL FLOW FIELD FLOW FRACTIONATION 60
I.5.1. ANALYSE DE DISTRIBUTION MOLAIRE DE LA GELATINE 60
I.5.2. LES METHODES DE FRACTIONNEMENT 61
I.5.3. LE CAS DE FLOW FFF 62
I.5.4. LES DETECTEURS MULTI ANGLE LIGHT SCATTERING (MALS) ET ULTRAVIOLET (UV) POUR L'ANALYSE DES PROTEINES 65
I.6. CONCLUSION 66

CHAPITRE II. ETUDE DE LA CARACTERISATION PAR SPECTROSCOPIE PROCHE INFRAROUGE EN MODE DYNAMIQUE DE COUENNES DE PORC 69

II.1. INTRODUCTION 71
II.2. MATERIELS ET METHODES 72
II.2.1. CARACTERISATION PHYSICO-CHIMIQUE EN LIGNE PAR SPIR D'UNE MATRICE HETEROGENE : LA COUENNE DE PORC 72
II.2.1.(1). Echantillonnage des couennes 72
II.2.1.(2). Acquisition de spectres proche infrarouge 73
II.2.1.(3). Analyses de référence 74
II.2.2. ANALYSE DES BASES DE DONNEES 81
II.3. RESULTATS ET DISCUSSION 86
II.3.1.(1). Caractérisation de la variabilité de la composition de couennes 86
II.3.1.(2). Analyse variabilité spectrale 89
II.3.1.(3). Evaluation de la capacité de prédiction des analyses de références à partir des spectres infrarouges collectés en ligne. 98
II.3.1.(4). Sélection des prétraitements : Lambda de Wilks- Performances de l'indice de qualité du signal (SQi) 108
II.3.1.(5). Effet du moyennage des spectres sur la capacité de prédiction des valeurs de taux de matière sèche, de matières grasses et de protéines. 112
II.4. CONCLUSION 113

CHAPITRE III. DETERMINATION EN LIGNE DU POTENTIEL D'EXTRACTION DE GELATINE DE COUENNES DE PORCS PAR SPECTROSCOPIE PROCHE INFRAROUGE SUR SITE DE PRODUCTION 117

III.1. INTRODUCTION 119

III.1.1.(1). La chaîne de production 120

III.1.1.(2). Démarche expérimentale 124

III.2. MATERIELS ET METHODES 125

III.2.1.(1). Positionnement de la sonde du spectromètre proche infrarouge sur la chaîne de production 125

III.2.1.(2). Conditions de température et humidité dans le hall des matières premières 127

III.2.1.(3). Acquisition spectrale en ligne des spectres de couennes de porc, avant chute dans le bassin de remplissage. 128

III.2.1.(4). Tri des spectres de 75 tonnes de couennes 128

III.2.1.(5). Rendement d'extraction en gélatine, à partir de 75 tonnes de couennes 130

III.2.1.(6). Echantillonnage 130

III.2.1.(7). Les spectres des 71 lots de 75 tonnes de couennes 131

III.1. RESULTATS ET DISCUSSION 132

III.1.1.(1). Les modèles de prédiction 132

III.1.1.(2). Les modèles de classification des échantillons vis-à-vis du rendement en gélatine de 75 tonnes de couennes par spectroscopie proche infrarouge. 133

III.2. CONCLUSION 141

CHAPITRE IV. ETUDE DES PROPRIETES PHYSICO-CHIMIQUES DES GELATINES PAR SPECTROSCOPIE PROCHE INFRAROUGE ET PAR ASYMMETRICAL FLOW FIELD-FLOW FRACTIONATION COUPLEE A UN DETECTEUR UV MULTI ANGLE LIGTH SCATTERING 145

IV.1. INTRODUCTION 147

IV.2. MATERIELS ET METHODES 149

IV.2.1. CARACTERISATION DE LA GELATINE PAR SPECTROSCOPIE PROCHE INFRAROUGE 149

IV.2.1.(1). Echantillons de gélatine 149

IV.2.1.(2). Acquisition spectrale des gélatines 149

IV.2.1.(3). Analyses de références 150

IV.2.1.(4). Analyse des données : 151

IV.2.2. CARACTERISATION DE GELATINE PAR ASYMMETRICAL FLOW FIELD-FLOW FRACTIONATION 152

IV.2.2.(1). Echantillonnage de gélatine 152

IV.2.2.(2). Analyses physico-chimiques 152
IV.2.2.(3). Analyse de la distribution molaire par AsFlFFF-UV/MALS 152
IV.2.2.(4). Analyse des données UV des fractogrammes de gélatine 154
IV.3. RESULTATS ET DISCUSSION 155
IV.3.1. CARACTERISATION DE GELATINE PAR SPECTROSCOPIE PROCHE INFRAROUGE 155
IV.3.1.(1). Mise en évidence des zones spectrales d'intérêt dans la différentiation des gélatines 155
IV.3.2. ETUDE DE LA DISTRIBUTION MOLECULAIRE DE LA GELATINE PAR ASFLFFF 159
IV.3.2.(1). Caractérisation des distributions molaires des gélatines. 159
IV.3.2.(2). Etude du rayon de giration des polymères de gélatine 162
IV.3.2.(3). Variabilité des distributions molaire des gélatines 164
IV.3.2.(4). Etudes des corrélations entre la force en gel (Bloom), la viscosité et les masses molaires des fractions séparées par AsFlFFF 166
IV.3.2.(5). Prédiction de la force en gel et de la viscosité, à partir de la distribution molaire. 168
IV.3.2.(6). Conclusion 169

CONCLUSION GENERALE 173

REFERENCES BIBLIOGRAPHIQUES 184

ANNEXES 192

RESUME 195

ABSTRACT 195

Table des figures

Figure 1 : Condensation de deux acides aminés pour former la liaison peptidique (en rouge). 26
Figure 2 : Répartition des 20 acides aminés en fonction des propriétés apportées par le radical. Groupe A : apolaire. Groupe B : polaire non chargé, Groupe C : polaires chargés (positivement ou négativement). (© Pearson Education 2012) 27
Figure 3 : schéma des différentes couches de la peau 32
Figure 4 : Structure d'une fibre de collagène (tropocollagène) (figure 19_43. Molecular Biology of the Cell, 4th Edition) 33
Figure 5 : Liaisons covalentes entre les tropocollagènes pour la formation de fibrilles. (Figure 19-46. Molecular Biology of the Cell, 4th Edition.) 34
Figure 6 : Exemple de liaisons hydrogène inter-chaînes (a), inter-chaînes avec une molécule d'eau (b) et intra-chaînes avec une molécule d'eau (c). D'après Duconseille et al 2015 (Duconseille et al., 2015) 37
Figure 7 : Réorganisation des chaînes de gélatine en solution (repliement des chaînes et nombre de brins impliqués) (A). Schéma du mécanisme de formation de la triple hélice à partir de deux brins. Les triples hélices avec une boucle sont formées en solution diluée et les triples hélices sans boucle sont formées dans des solutions concentrées, les triples hélices avec une seule boucle coexistent avec les triples hélices sans boucle dans les solutions semi-diluées (B). D'après les travaux de Guo et al. (Guo et al., 2003) 38
Figure 8 : Spectre Electromagnétique (image libre de droit), UV pour Ultra-violet, IR pour Infra-Rouge 46
Figure 9 : Trois conformations instrumentales traditionnelles en spectroscopie proche infrarouge. (A) Instrument à filtre. (B) Instrument pré-dispersif à monochromateur. (C) Instrument post dispersif à barette de diodes. 47
Figure 10 : Schéma des utilisations de la spectroscopie proche infrarouge sur une chaîne de production. 48
Figure 11 : Schéma de présentation des méthodes supervisées et non supervisées appliquées aux données analytiques 53
Figure 12 : Evolution de l'erreur standard dans les modèles de prédiction en fonction du nombre de composantes d'après Martens et al. (Martens et Naes, 1992) 58
Figure 13 : Représentation schématique de la cellule AsFlFFF, d'après Rbii et al. (Rbii et al., 2009) 63
Figure 14 : Schéma du principe de séparation de particules de différentes tailles en AsFlFFF. Les particules de petite taille sont représentées par de petits cercles noir (●), et les particules de grande taille sont représentées par de cercle blanc (○). D'après les travaux de Rbii et al. (Rbii et al., 2009) 64
Figure 15 : Couenne face cuir 73
Figure 16 Tapis de convoyage pour l'acquisition spectrale en laboratoire. 74
Figure 17 : Espacement de trois mesures de couenne en flux. 74
Figure 18 : Couennes après les différentes étapes d'homogénéisation. 75
Figure 19 : Principe de l'ASE : les solvants passent par la pompe puis sont pressurisés avant de traverser la cellule ASE, entrainant les matières grasses pour être récupéré dans un flacon de récupération. 79

Figure 20 : Analyse descriptive des échantillons en fonction du type de production (biologique ou conventionnel), de l'objectif d'élevage (les porcs pour la production ou les coches pour la reproduction), de la partie de l'animal (bardière ou couennes autre) et du conditionnement (congelé ou frais). MS Matières Sèches, PROT Protéines, MG masse Grasse, COL Collagène. 86
Figure 21 : Matrice de corrélation entre le taux de matière sèche (Hum), le taux de protéine (Prot), les matières grasses en mg/g (MG) et le taux d'hydroxyproline (Col) 88
Figure 22 : Spectres collectés en ligne en réflectance sur un de tapis de convoyage, pour la face cuir (haut) et la face graisse (bas) de la couenne numéro 5. 89
Figure 23: Spectres et scores de l'ACP des données spectrales de la face graisse de couennes. A : pour la face graisse de la couenne numéro 1. B : pour la face graisse de la couenne numéro 5. En noir les spectres correspondant au tapis, en gris les spectres correspondant à la couenne, en rouge les spectres de couenne à l'interface avec le tapis. 90
Figure 24 : Spectres en réflectance de la couenne numéro 1 ; En magenta les spectres correspondant à la face graisse, en bleu les spectres correspondant à la face cuir. 91
Figure 25 : Intensité des spectres face cuir (A et B) et face graisse (C et D) en fonction de l'intensité du spectre moyen ; en A spectres bruts face cuir, B spectres après prétraitement SNV face cuir, C spectres bruts face graisse et D spectres après prétraitement SNV face graisse. 93
Figure 26 : Spectres en réflectance de muscle (en jaune), de graisse (en rouge) et de peau (en bleu). 94
Figure 27 : Scores des ACP sur les spectres face cuir (points rouge) et face graisse (points bleu). A gauche l'ACP sur les spectres bruts, à droite l'ACP sur les spectres après prétraitement SNVD (ordre 1) 95
Figure 28 : Score des ACP sur les spectres en ligne des 57 couennes non moyennés après prétraitement SNVD pour ; A - pour la face graisse, B - Pour la face cuir. En haut en bleu sont représentées les catégories par quartile pour la matière grasse (décroissant du bleu foncé au bleu clair), en bas en rouge sont représentées les catégories par quartile pour les protéines (croissant du jaune au rouge foncé) 97
Figure 29 : Modèles PLS par validation croisée du taux de protéine. A gauche la prédiction pour l'ensemble des échantillons sur la base de spectres complets, à droite la prédiction sans les échantillons présentant des valeurs aberrantes. En haut pour les spectres face cuir, au centre pour les spectres face graisse, en bas pour les spectres faces cuir et graisse. Les équations de prédiction présentées sont réalisées sur les données brutes. 99
Figure 30 : Calculs des coefficients de corrélation entre les coefficients de détermination des analyses de référence par PLS et les indices (SQI et lambda de Wilks) 109
Figure 31 : Bacs de 500 kilos de couennes stockés dans le hall matières premières. 120
Figure 32 : Bassins de lavage en début de chaîne de production 122
Figure 33 : Réacteur d'extraction 122
Figure 34 : Schéma du circuit des couennes en fonction de l'ordre de fabrication 123
Figure 35 : Gélatine en sortie d'extrusion 123
Figure 36 : Schéma de la démarche expérimentale mise en place. 125
Figure 37 : Schéma de la position du spectromètre dans le hall de répartition des matières premières 125

Figure 38 : Spectromètre Polytech dans le support anti-vibration (photo de gauche), sonde déportée sur le bras de distribution (photo de droite). *126*
Figure 39 : Schéma du positionnement de la sonde déportée sur le bras de distribution *126*
Figure 40 : Evolution de l'humidité relative (a) et de la température (b) dans l'usine de Weishardt (Graulhet), sur 12 mois (mai 2015 à avril 2016) *127*
Figure 41 : Exemple de sélection de spectres sur 3000 spectres acquis en conditions industrielle. En vert les spectres de couenne, en bleu les spectres de vide ou d'interface vide-couenne. *129*
Figure 42 : Spectres proche infrarouge moyens des 71 échantillons de 75 tonnes de couennes *131*
Figure 43 : Analyse en composantes principales des spectres moyens des 71 échantillons *131*
Figure 44 : Modèle PLS sur le rendement pour les 71 échantillons (données brutes) *133*
Figure 45 : Spectres moyens des 4 classes pour les 71 échantillons *134*
Figure 46 : Projection des barycentres des 4 classes à effectif constant sur les axes factoriels 1 et de 2 de l'Analyse Discriminante sur les données brutes (n=71) *137*
Figure 47 : Projection des échantillons sur les variables latentes 1 et 2 de la régression des moindres carrés partiels – Discriminante Analysis sur les données sans prétraitement *139*
Figure 48 : Spectres proche infrarouge des 135 gélatines industrielles ; (A) pour les spectres bruts, (B) pour les spectres après prétraitement SNVD *156*
Figure 49 : Modèle de prédiction de la viscosité par PLSr. En noir les données de calibration (n=102), en rouge les données en validation (n=33) *158*
Figure 50 : A et B décrivent les fractogrammes de l'analyse AsFlFFF-UV/MALS de gélatines de porc : A et B en fonction du temps d'élution, A' et B' le signal UV en fonction de la masse molaire. La gélatine (A, A') possède une force en gel de 322,46 g et une viscosité de 46,71 mP, la gélatine (B, B') possède une force en gel de 98,64 g et une viscosité de 20,54 mP. Le signal UV en intensité relative (ordonnées de droite-point en gris clair). Le signal de diffusion de la lumière de l'angle 90° en intensité relative (ordonnée de droite- points en gris foncés). La masse molaire calculée en g.mol^{-1} (ordonnée de gauche- points en noir). *161*
Figure 51 : RMS conformation plot pour des gélatines de faible et moyen bloom (A) et pour des gélatine de haut bloom (B).Sur le graphique A, il y a trois gélatine, en gris clair (force en gel de 200 g et viscosité 35,2 mp), en gris foncé (force en gel de 97 g et viscosité de 20,2 mp) et en noir (force en gel de 214 g et viscosité de 56,7 mp). Sur le graphique B, il y a deux gélatines, en gris clair (force en gel de 268 g et viscosité de 106,9 mp) et en noir (force en gel de 214 g et viscosité de 56,7 mp) *163*
Figure 52 : Contribution des absorbances UV en fonction du temps d'élution aux composantes principales. En noir la PC1, en gris la PC2. ..... *165*
Figure 53 : Corrélation entre les valeurs prédites et les valeurs mesurées pour la force en gel (g) et la viscosité (mP). En bleu est présenté le jeu de données en calibration et en orange en validation croisée. *169*
Figure 54 : Schéma de l'interaction entre l'industrie de la gélatine porcine et les améliorations potentielles apportées par la SPIR aux différents états du produit *182*

Table des tableaux

Tableau 1 : Types de collagène et localisation chez les mammifères. Les valeurs d'abondance dans la peau correspondent à un humain adulte. (Inspiré des travaux de A.Duconseille et al. (Duconseille et al., 2015). *30*

Tableau 2 : Comparaison de la composition en acides aminés entre le collagène bovin (type I), la gélatine de peau de porc par extraction acide (type A) et de la gélatine d'os de porc par traitement basique (type B). Les acides aminés sont exprimés en nombre d'acides aminés pour 1000. (Babel et al., 2000) *36*

Tableau 3 : Propriétés fonctionnelles de gélatines dans l'agroalimentaire (Handbook, 2012)..... *45*

Tableau 4 : Prédiction des paramètres technologiques et de la composition par spectroscopie proche infrarouge sur du porc..... *51*

Tableau 5 : Matrice de confusion pour 2 classes..... *55*

Tableau 6 : Matrice de confusion..... *83*

Tableau 7 : Moyenne, Gamme, écart-type, coefficient de variation, écart-type de répétabilité et limite de répétabilité biologique (écart-type en trois prises d'échantillons différentes) des 57 couennes. Les résultats sont exprimés en pourcentage par rapport à la masse fraîche de couenne *87*

Tableau 8 : Matrice de confusion du modèles ACP-LDA sur les 4254 spectres bruts de l'essai, en fonction de leurs provenance (spectres acquis face cuir et spectres acquis face graisse) *96*

Tableau 9 : Statistiques de calibration et de validation croisée pour les modèles de prédiction du taux de protéine. Les données ont été traitées par PLSr en validation croisée, dans le cadre de l'analyse des spectres complets les échantillons ont été considérés par bloc. LV est le nombre de variables latentes...... *101*

Tableau 10 : Statistiques de calibration et de validation croisée pour les modèles de prédiction du taux de matière sèche. Les données ont été traitées par PLSr en validation croisée, dans le cadre de l'analyse des spectres complets les échantillons ont été considérés par bloc. LV correspond au nombre de variables latentes. *103*

Tableau 11 : Statistiques de calibration et de validation croisée pour les modèles de prédiction du taux de matières grasses. Les données ont été traitées par PLSr en validation croisée, dans le cadre de l'analyse des spectres complets les échantillons ont été considérés par bloc. LV est le nombre de variables latentes. *104*

Tableau 12 : Statistiques de calibration et de validation croisée pour les modèles de prédiction du taux de collagène. Les données ont été traitées par PLSr en validation croisée, dans le cadre de l'analyse des spectres complets les échantillons ont été considérés par block..... *106*

Tableau 13 : Coefficient de corrélation de Pearson entre les valeurs de SQi, Wilks et les coefficients de corrélation pour le taux de matières sèches, taux de matière grasse, taux de protéine et collagène des PLSr, pour les différents prétraitements...... *110*

Tableau 14 : Valeurs des indices du SQi pour les prétraitements D1, D2, SNV, SNVD et MSC sur les données en réflectance. Ici le SQi a été calculé avec les quatre analyses de référence. *111*

Tableau 15 : Valeurs de SQi sur les données brutes pour les différentes bases de données, avec le taux de collagène et sans le taux de collagène...... *111*

Tableau 16 : Tableau de comparaison des coefficients de détermination entre les valeurs prédites et les valeurs de référence pour les bases avec les spectres complets, les spectres moyennés et la moyenne des prédictions de la base complète. MS = Matière sèche, Prot = protéine. Ici les calculs ont été réalisés sur l'ensemble des échantillons (n=57), sans prendre en compte le taux de collagène 112
Tableau 17 : Répartition des classes pour les analyses réalisées avec 4 classes. 133
Tableau 18 : Performances de l'analyse discriminante en calibration, validation croisée et validation externe pour les classes à amplitude constante (1) ou les classes à effectif constant (2). 136
Tableau 19 : Performances de la PLS couplée à l'analyse discriminante en calibration, validation croisée et validation externe pour 2 variables latentes 138
Tableau 20 : Répartition des classes pour les analyses réalisées avec 3 classes 139
Tableau 21 : Performances de l'analyse discriminante en calibration, validation croisée et validation externe pour 3 classes 140
Tableau 22 : Propriétés physico-chimiques des gélatines échantillonnées. Nombre d'échantillons (n), moyenne, écart-type (SD), Valeur minimale et maximale et coefficient de variation (CV) de la force en gel et de la viscosité des gélatines. (n=135) 149
Tableau 23 : Propriétés physico-chimiques des gélatines échantillonnées. Nombre d'échantillons (n), moyenne, écart-type (SD), Valeur minimale et maximale et coefficient de variation (CV) de la force en gel et de la viscosité des gélatines. (n=49) 152
Tableau 24 : Statistiques de calibration, de validation croisée et de validation pour les modèles de prédiction de la force en gel par SPIR. Les données ont été traitées par PLSr en validation (75% en validation, 25% en validation). Avec LV pour Variables Latentes. 157
Tableau 25 : Statistiques de calibration, de validation croisée et de validation externe pour les modèles de prédiction de la viscosité par SPIR. Les données ont été traitées par PLSr en validation (75% en validation, 25% en validation). Avec LV pour Variables Latentes. 157
Tableau 26 : Statistiques de calibration, de validation croisée et de validation externe pour les modèles de prédiction de l'humidité. Les données ont été traitées par PLSr en validation (75% en validation, 25% en validation). Avec LV pour Variables Latentes. 159
Tableau 27 : Matrice de similarité (coefficient de corrélation de Pearson) entre la force en gel, la viscosité, les Mn (masse moyenne en nombre), les Mw (masse moyenne en masse) et les % de fraction pour les fractions des chaînes primaires (PRI) et des chaînes polymérique (POLY) pour les 49 échantillons. 166
Tableau 28 : Evaluation de la capacité de prédiction par PLS régression des propriétés physico-chimiques (force en gel, viscosité) à partir du signal UV des fractogrammes. Les résultats présentés un modèle construit avec 3 variables latentes. 168

Liste des abréviations

ACP	Analyse en Composantes Principales
ANN	*Artificial Neural Networks,* Réseau de neurone
ASE	*Accelerate Solvant Extraction*
AsFlFFF	*Asymetrical Flow Field-flow Fractionation*
CV	Coefficient de Variation
D	Dérivée
DA	*Discriminante Analysis,* Analyse Discriminante
FFF	Fractionnement par Flux Force
LCA	Laboratoire de Chimie Agro-industrielle
LV	Latent variable, variable latente
MALS	*Multi Angle Light Scattering*
MG	Matière Grasse
MS	Matière Sèche
MSC	*Multiplicative Scatter Correction*
PC	Principal Component
PLSr	*Partial Least Square regression,* régression des moindres carrés partiels
Rg	Rayon de giration
RMSE	*Root Mean Square Error*
RMSEC	*Root Mean Square Error of Calibration*
RMSEP	*Root Mean Square Error of Prediction*
RPD	*Relative Predictive Determinant,* Déterminant Prédictif Relatif
SEC	*Standar Error of Calibration*
SECV	*Standard Error of Cross Validation*
SEP	*Standard Error of Prediction*
SNV	*Standard Normal Variate*
SPIR	Spectroscopie Proche Infrarouge
SQI	*Signal Quality Index,* index de qualité du signal
UV	Ultraviolet

Introduction générale

Dans le domaine de l'agroalimentaire, la qualité est en perpétuelle augmentation depuis le début des années 90 et le consommateur est de plus en plus exigeant. Elle peut se distinguer en trois sous parties ; la qualité nutritionnelle, la qualité hygiénique et la qualité organoleptique et technologique d'un produit. La production de gélatine est une industrie de transformation. Elle transforme les matières premières (couennes de porc, peaux de poisson) en un produit semi-fini, la gélatine, elle-même utilisée par d'autres industriels en tant qu'ingrédient. Il est important de caractériser les produits aux différents stades du processus de transformation. Pour cela, l'industriel a besoin, d'une part, de caractériser les matières premières, afin de les acheter au prix le plus juste en fonction de la qualité proposée par les différents fournisseurs. La maitrise de la qualité des matières premières peut permettre au gélatinier de mieux gérer le processus d'extraction de la gélatine. D'autre part, il a besoin de garantir la qualité hygiénique, organoleptique et technologique des gélatines, afin de répondre aux attentes des clients. Dans cette optique, la spectroscopie proche infrarouge apparait être un outil adapté au suivi qualité de l'industrie de la gélatinerie.

La spectroscopie proche infrarouge est une technique basée sur l'interaction de la lumière sur une matrice, permettant d'obtenir des spectres qui correspondent à l'absorption de la lumière en fonction de la longueur d'onde et ce, en lien avec les propriétés chimiques et physiques de la matrice. Cette méthode permet une analyse rapide et non destructive d'un produit, et est très largement utilisée dans le domaine de l'agroalimentaire. Dans le domaine de la production animale, les applications les plus connues sont, la prédiction de l'humidité, de la graisse et de la graisse intramusculaire, du taux de protéine ou encore de la couleur.

Les présents travaux s'inscrivent dans le cadre d'une collaboration entre un industriel producteur de gélatine de la région Occitanie (Weishardt) et deux laboratoires académiques : le Laboratoire de Chimie Agro-Industrielle (LCA UMR1010-INRA/INPT) et le Laboratoire d'Agrophysiologie, Agromolécules de l'Ecole d'Ingénieurs de Purpan.

L'objectif finalisé de ce projet est de proposer l'implantation en ligne d'un outil analytique basé sur la détection par spectroscopie proche infrarouge afin d'évaluer la capacité d'appréhension de la qualité des matières premières et des produits sur une chaîne de production de gélatine.

L'objectif spécifique de ce projet consiste à faire la preuve de concept de l'utilisation de la spectroscopie proche infrarouge non seulement pour caractériser les matières premières en perpétuel mouvement afin de prédire le rendement potentiel d'extraction mais également pour reconnaitre les qualités fonctionnelles du produit fini.

Afin de répondre à ces problématiques, la thèse est articulée en quatre chapitres :

- Dans le premier chapitre, un bilan bibliographique des connaissances sur la structure protéique complexe de la gélatine sera établi. Les procédés d'extractions industriels seront également abordés. Nous décrirons par ailleurs deux techniques analytiques adaptées à l'étude des biopolymères : la spectroscopie infrarouge et l'« *Asymmetrical Flow Field flow Fractionation* »
- Dans le second chapitre, la composition chimique de la matière première (couenne de porc) sera étudiée par spectroscopie proche infrarouge, fragment par fragment, en mouvement sur un tapis de convoyage. Cette partie permettra, d'une part, d'appréhender les contraintes liées à la présence d'un flux, et d'autre part, de mieux comprendre la nature de l'interaction entre la spectroscopie proche infrarouge et la couenne de porc.
- Le troisième chapitre sera consacré à l'étude de la mise en place d'une méthodologie, permettant d'appréhender le potentiel de rendement d'extraction à l'échelle industrielle. Cette étude permettra d'évaluer la capacité d'appréhension du rendement potentiel en gélatine par spectroscopie proche infrarouge.
- Le quatrième chapitre du manuscrit traitera des propriétés fonctionnelles des gélatines, d'une part, étudiées par spectroscopie proche infrarouge, afin d'évaluer la capacité d'appréhender le taux d'humidité, la force en gel et la viscosité. D'autre part, étudiées par *Asymmetrical Flow Field-Flow Fractionation* couplée à un détecteur UV et MALS, dans le but d'étudier les relations entre la distribution moléculaire de ces gélatines et leur force en gel ou encore leur viscosité.
- La conclusion nous permettra de décrire les conditions qui permettent de prédire le rendement d'extraction des gélatines ainsi que leurs propriétés fonctionnelles.

Chapitre I. Etat de l’art

I.1. Apparition de la gélatine dans les industries

Les premières traces d'utilisation de la gélatine remontent à 6000 ans avant Jésus-Christ. A cette époque, les hommes du néolithique extrayaient la gélatine de tissus animaux dans le but de l'utiliser comme une colle. Bien plus tard, 1000 ans avant Jésus-Christ, c'est la civilisation égyptienne qui l'utilisera comme colle à bois pour assembler les mobiliers. Il a fallu attendre le XVII^{ème} siècle pour que celle-ci soit employée comme denrée alimentaire (Schrieber et Gareis, 2007) .

Le tout premier outil capable d'extraire la gélatine à partir d'os fut inventé par Denis Papin en 1681. Le « digesteur » était une marmite autoclave sous pression et à haute température (120-130°C). Ce projet avait pour but de produire une denrée alimentaire pour subvenir aux besoins nutritifs des plus pauvres en ce temps de guerre. Cependant, il fut refusé par le roi anglais Charles II car les os étaient destinés à nourrir les chiens de chasse et la qualité médiocre de la gélatine obtenue lui donnait un goût prononcé d'ammoniac.

Les travaux de Papin furent repris par de nombreux chimistes et chercheurs, tels que Changeux, Proust, Rumford ou encore Cadet de Vaux pendant le XVIII^{ème} siècle. Mais c'est avec Jean-Pierre-Joseph d'Arcet que l'industrialisation de la gélatine commença réellement. Ce chimiste mit en place des systèmes d'extraction permettant d'obtenir une gélatine de qualité et en quantité. La possibilité d'en faire un aliment viable et économique était un projet essentiel pour pallier aux carences nutritives causées par la disette lors de la Révolution française(De Claubry, 1831; Viel et Fournier, 2006). Cette industrie a pu se développer rapidement grâce au contexte favorable dont elle bénéficiait. Malheureusement, la consommation de gélatine comme base alimentaire provoquait des carences, notamment à cause de l'absence de certains acides aminés essentiels. Depuis cette époque, la gélatine est utilisée comme ressource pour l'industrie agroalimentaire notamment pour ses propriétés fonctionnelles. Ces propriétés ont été par ailleurs mises à profit dans d'autres domaines tels que la photographie, la parapharmacie et même l'industrie.

En 1834, un étudiant en pharmacie, François Mothes, découvre un procédé pour la fabrication de gélule de gélatine et dépose un brevet avec le professeur Dublanc (Podczeck et Jones, 2004). Ce produit devient immédiatement populaire en France et très rapidement les gélules molles de gélatine vont révolutionner la médecine de l'époque. En 1845, c'est le pharmacien Jules César Lehuby qui dépose un brevet pour la fabrication de gélules solides afin de produire des enveloppes médicamenteuses. A partir de ces découvertes, de nombreuses recherches ont été réalisées dans le domaine de la pharmaceutique, pour que ces capsules deviennent un produit contemporain incontournable. A la fin du XIX^{ème} siècle, la gélatine est déposée sur des plaques photographiques, permettant d'obtenir les premiers négatifs souples participant à la démocratisation de la photographie.

I.2. La gélatine une structure protéique complexe

I.2.1. Le collagène et la gélatine : des polymères protéiques

I.2.1.(1). Rappel sur la structure des protéines

Les protéines sont des macromolécules biologiques composées de plusieurs unités structurelles appelées acides aminés. Par définition, un acide aminé possède, sur un carbone asymétrique (sauf pour la glycine), un groupement carboxylique (COOH), un groupement aminé (NH_2) et un résidu (R) différent pour chaque acide aminé. Dans les polymères protéiques, les acides aminés sont liés entre eux par des liaisons peptidiques. La formation de ces liaisons est le résultat de la condensation (perte d'une molécule d'eau) entre la fonction carboxylique d'un premier acide aminé et la fonction amine d'un second acide aminé (Figure 1) (Petsko *et al.*, 2008).

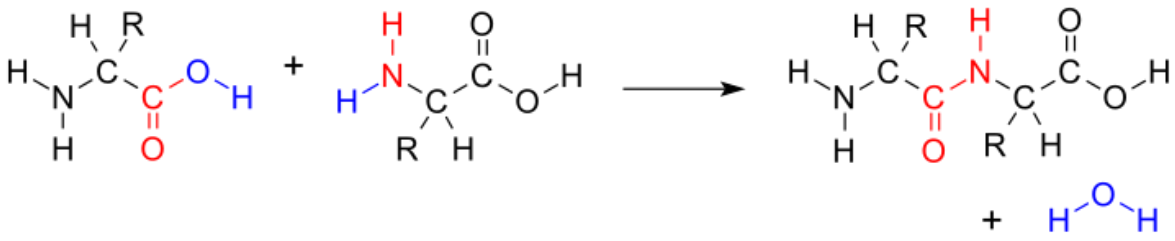


Figure 1 : Condensation de deux acides aminés pour former la liaison peptidique (en rouge).

Sur plus de 300 acides aminés recensés, seulement 20 (21 en considérant l'hydroxyproline qui est dérivée de la proline) sont présents dans les protéines des êtres vivants (Moussard, 2006). Ces acides aminés sont classés en catégories en fonction des propriétés apportées par le résidu (R) : apolaire, polaire (non chargé), polaire chargé positivement et négativement (Figure 2).

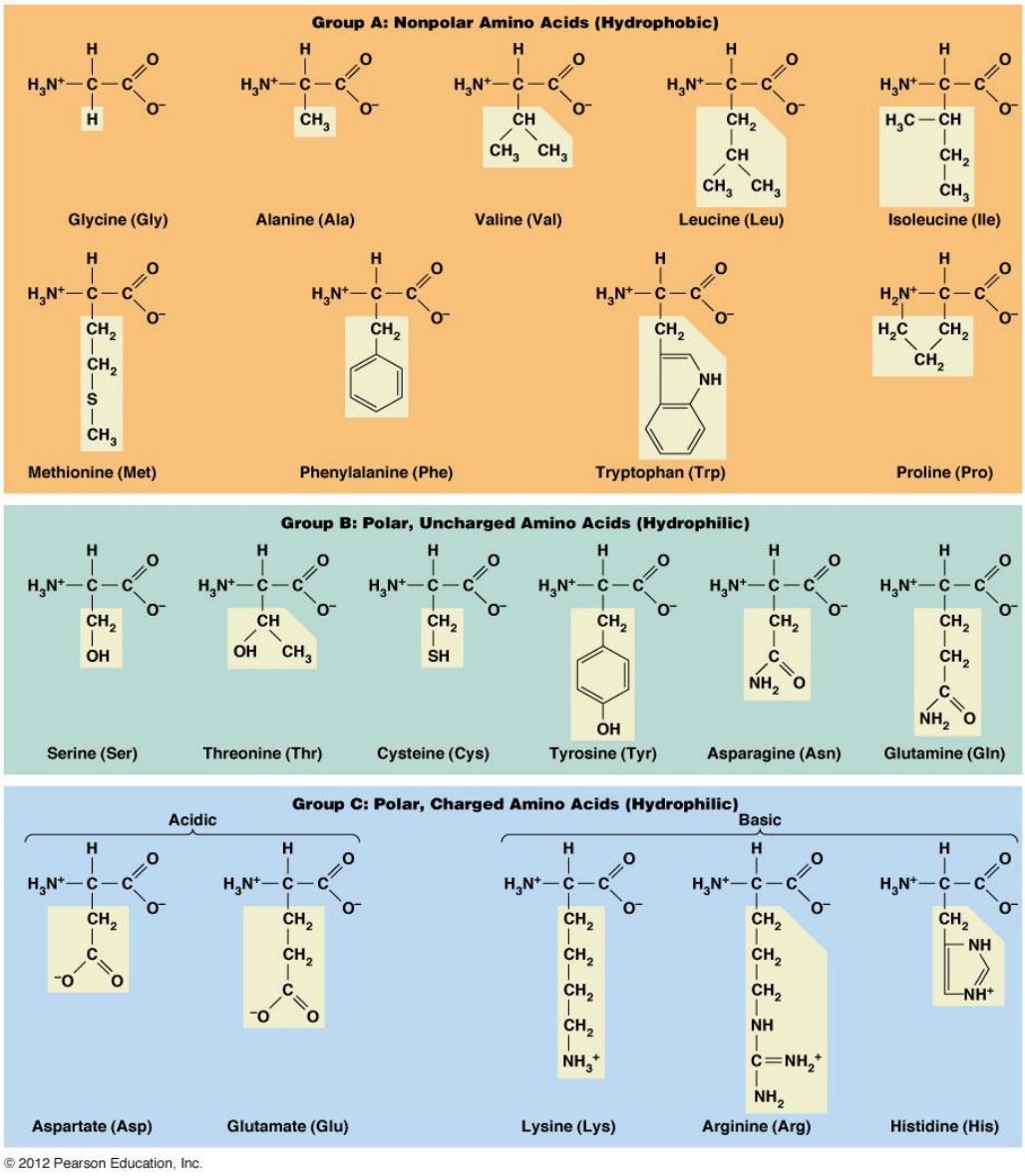


Figure 2 : Répartition des 20 acides aminés en fonction des propriétés apportées par le radical. Groupe A : apolaire. Groupe B : polaire non chargé, Groupe C : polaires chargés (positivement ou négativement). (© Pearson Education 2012)

Les polymères protéiques sont divisés en deux classes en fonction de leur taille ; les « protéines » pour les chaînes supérieures à 100 acides aminés et les « peptides » pour les chaînes inférieures à 100 acides aminés. Par convention, le premier acide aminé d'une chaine est appelé « N-terminal » et possède le groupement aminé libre, le dernier acide aminé est appelé « C-terminal » et présente le groupement carboxylique libre. Il est possible de distinguer quatre niveaux de structure chez les protéines (Moussard, 2006):

- **La structure primaire** : elle correspond à la séquence d'une chaîne polypeptidique, autrement dit à l'enchainement d'acides aminés liés entre eux par une liaison peptidique. La séquence primaire permet de déterminer le nombre, la nature et la position des résidus.
- **La structure secondaire** : cette structure correspond au repliement local des acides aminés d'une chaîne, laquelle est stabilisée par des liaisons hydrogène entre les CO et les NH peptidiques, conférant trois types de conformation : les feuillets, les hélices et les coudes.
- **La structure tertiaire :** elle est aussi appelée structure tridimensionnelle, elle correspond à l'agencement dans l'espace de structures secondaires d'une chaîne polypeptidique. Cette structure est stabilisée par plusieurs types de liaisons ; de Van der Waals, hydrogène, hydrophobe, ionique et covalente (pont disulfure). Généralement, la structure tertiaire est associée à la fonction d'une protéine globulaire et est sensible à son environnement. Elle dépend de la concentration et de la nature d'autres substances dissoutes, du pH et de la température de son milieu. Il n'y a pas de structure tridimensionnelle pour les protéines fibrillaires.
- **La structure quaternaire :** elle correspond à la mise en relation de plusieurs chaînes polypeptidiques différentes, pour former ensemble une protéine globulaire. La structure quaternaire est stabilisée par les mêmes types de liaisons que celles de la structure tertiaire.

I.2.1.(2). Localisation du collagène dans les tissus de mammifères

Le mot collagène provient du grec *kolla* (*κόλλα)* et fait référence à la production de colle. Le collagène est la protéine majoritaire des espèces du règne animal et représente plus d'un quart des protéines (Gorham, 1991). Cette molécule a la particularité de jouer un rôle de structure et permet de résister aux forces de tension dans les tissus (Voet et Voet, 2004). Cette protéine fibreuse est principalement située au niveau de la matrice extracellulaire des tissus conjonctifs (Moussard, 2006). A ce jour, 28 différents types de collagènes ont été identifiés en fonction de leur composition, de leur structure, de leur synthèse et de leur localisation (Tableau 1) (Gómez-Guillén *et al.*, 2011; Duconseille *et al.*, 2015). C'est le collagène de type I qui est le plus largement représenté dans les différents tissus (Bruckner, 2010).

Tableau 1 : Types de collagène et localisation chez les mammifères. Les valeurs d'abondance dans la peau correspondent à un humain adulte. (Inspiré des travaux de A.Duconseille et al. (Duconseille et al., 2015).

Type	**Localisation**	**Abondance dans la peau**	**Références**
I	Peau, muscle, tendon, ligaments, os, dentine, cornée	80–85%	(Bailey et Light, 1989), (Brinckmann *et al.*, 2005), (Gelse *et al.*, 2003)
II	Cartilage, disque intervertébral, noyau pulpeux, yeux	–	(Bailey et Light, 1989), (Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003)
III	Peau, intramusculaire, système vasculaire, intestin, utérus, fibres réticulaires	10–15%	(Bailey et Light, 1989), (Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003)
IV	Jonctions neuromusculaires, membranes basales, poumons	–	(Bailey et Light, 1989), (Gelse, Pöschl, et Aigner, 2003), (Fox, 2008)
V	Peau, intermusculaire, tissus d'embryon, os, cornée	2–4%	(Bailey et Light, 1989), (Brinckmann, Notbohm, et Müller, 2005)
VI	Peau (épiderme), système vasculaire, os, cartilage, cornée, placenta	n/a	(Bailey et Light, 1989), (Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003) (Söderhäll *et al.*, 2007)
VII	Peau, membrane amniotique, vessie, muqueuse buccale, cordon ombilical, fluide amiotique	0.00%	(Bailey et Light, 1989), (Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003), (Chagnot *et al.*, 2012)
VIII	Peau, membranes basales, yeux, vaisseau, os, cerveau, cœur, rein, cartilage	n/a	(Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003)
IX	Cartilage articulaire, cornée, yeux	–	(Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003)
X	Cartilage hypertrophique	–	(Gelse, Pöschl, et Aigner, 2003), (Sweeney *et al.*, 2010)
XI	Cartilage, disque intervertébral	–	(Bailey et Light, 1989), (Gelse, Pöschl, et Aigner, 2003), (Brinckmann, Notbohm, et Müller, 2005)
XII	Peau, tendon, cartilage, ligament	n/a	(Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003)

XIII	Peau, os, structures neuronales, cellules endothéliales, cœur, yeux, muscle squelettique, cheveux, intestin, poumon	n/a	(Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003), (Ylönen *et al.*, 2005)
XIV	Peau, vaisseaux sanguins, os, cartilage, yeux, tendon, utérus, nerfs, foie, poumon	n/a	(Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003)
XV	Peau, capillaire, placenta, reins, cœur, ovaires, testicules, pancréas, muscles lisses	n/a	(Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003)
XVI	Peau, cœur, reins, muscles lisses, kératinocytes	n/a	(Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003)
XVII	Hémidesmosome dans les structures épithéliales et neuronales	–	(Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003), (Has et Kern, 2010)
XVIII	Paroi de la membrane périvasculaire, reins, foie, poumons	–	(Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003)
XIX	Peau, neurones, membrane basale dans le muscle squelettique, rate, prostate, reins, foie, placenta, côlon	n/a	(Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003), (Chagnot et al., 2012), (Su *et al.*, 2010)
XX	Epithélium de la cornée, peau d'embryon, tendon	–	(Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003), (Chagnot et al., 2012)
XXI	Matrice extracellulaire des parois des vaisseaux sanguins, vaisseau, cœur, estomac, reins, muscles squelettique, placenta	–	(Brinckmann, Notbohm, et Müller, 2005), (Gelse, Pöschl, et Aigner, 2003), (Chagnot et al., 2012), (Chou et Li, 2002)
XXII	Jonctions tissulaires	–	(Chagnot et al., 2012), (Koch *et al.*, 2004)
XXIII	Cancer de la prostate et métastase à distance, cœur, rétine	–	(Brinckmann, Notbohm, et Müller, 2005), (Banyard *et al.*, 2007)
XXIV	Os, cornée	–	(Brinckmann, Notbohm, et Müller, 2005), (Chagnot et al., 2012)
XXV	Structures neuronales	–	(Brinckmann, Notbohm, et Müller, 2005)
XXVI	Testicules et ovaires	–	(Chagnot et al., 2012), (Sato *et al.*, 2002)
XXVII	Cartilage chez l'adulte	–	(Chagnot et al., 2012), (Hjorten *et al.*, 2007)
XXVIII	Tissus neuronal	–	(Chagnot et al., 2012), (Grimal *et al.*, 2010)

Dans le cadre de la production de gélatine au niveau industriel, la principale source de collagène est la peau de mammifère, majoritairement la peau de porc (couenne). Trois couches cellulaires structurent la peau de manière générale (Figure 3) :

- L'épiderme : C'est la couche supérieure, directement en contact avec le milieu extérieur. C'est la couche la moins épaisse et elle contient principalement des cellules mortes.
- Le derme : Directement sous l'épiderme, c'est une couche plus épaisse que ce dernier. Le derme est un tissu conjonctif qui comporte majoritairement de la matrice extracellulaire produite par des fibroblastes.
- L'hypoderme : C'est la dernière couche constituant la peau, elle est constituée majoritairement de tissus adipeux (adipocytes). C'est la couche la plus épaisse et la plus variable, elle dépend des individus et de la localisation dans le corps.

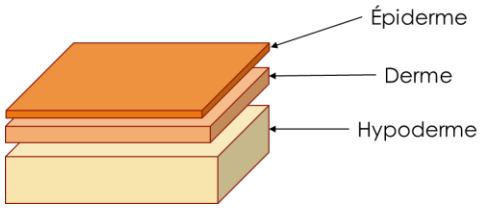


Figure 3 : schéma des différentes couches de la peau

Le collagène est majoritairement présent dans le derme. Il est aussi présent dans l'épiderme, mais cette couche à la particularité de posséder une autre protéine, la kératine, dans les parties pileuses. En revanche le collagène est généralement absent de l'hypoderme (Mélissopoulos et Levacher, 1998).

I.2.1.(3). Structure du collagène

Le collagène est composé d'une unité de base appelée tropocollagène. Cette unité comprend trois chaînes α variant de 662 à 3152 acides aminés. Les chaînes possèdent elles-mêmes une structure primaire répétée, de trois acides aminés, Gly-X-Y, avec Gly qui correspond à la glycine, où X est principalement de la Proline et Y principalement de l'hydroxyproline (Figure 4 A). L'hydroxyproline est un dérivé de la proline spécifique à la protéine de collagène (Ricard-Blum et Ruggiero, 2005; Ricard-Blum, 2011). Les chaînes α de collagène adoptent une configuration en hélice gauche, ces chaînes s'assemblent par trois pour former une triple hélice droite appelée « tropocollagène ».

Cette structure est stabilisée par des liaisons hydrogène inter et intra chaînes, entre les H de la glycine orientés vers le centre de la triple hélice alors que les résidus X et Y sont orientés vers l'extérieur de la structure (Figure 4 B). Aux extrémités de cette triple hélice, les chaînes ne conservent pas cette structure en triple hélice, la zone est appelée télopeptide (Engel et Bächinger, 2005).

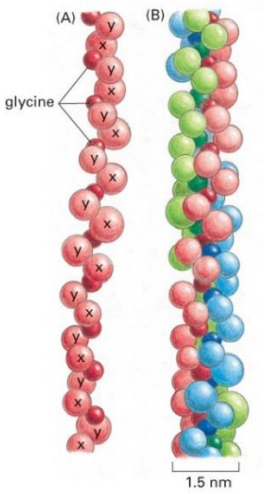


Figure 4 : Structure d'une fibre de collagène (tropocollagène) (figure 19_43. Molecular Biology of the Cell, 4th Edition)

Ces structures s'organisent en fibrilles, lesquelles s'organisent en réseau pour former des fibres de collagène dans la matrice extracellulaire. Ces fibres peuvent interagir entre elles *via* des liaisons faibles, mais aussi avec des liaisons covalentes (spécifiques aux fibres de collagène et d'élastine) (Figure 5), notamment entre les résidus de lysine et d'hydroxyproline de la région télopeptidique. Dans la matrice extracellulaire, le collagène peut former des réseaux avec d'autres polymères : la laminine, l'élastine, les protéoglycanes ou les fibronectines dont les proportions sont variables en fonction du type de tissus, mais aussi de l'âge de l'individu (Ricard-Blum, 2011). Par exemple, le collagène de type III est présent à plus de 50% dans la peau d'un embryon, alors qu'il ne représente plus que 5 à 10% pour la peau de personnes âgés (Babel *et al.*, 2000).

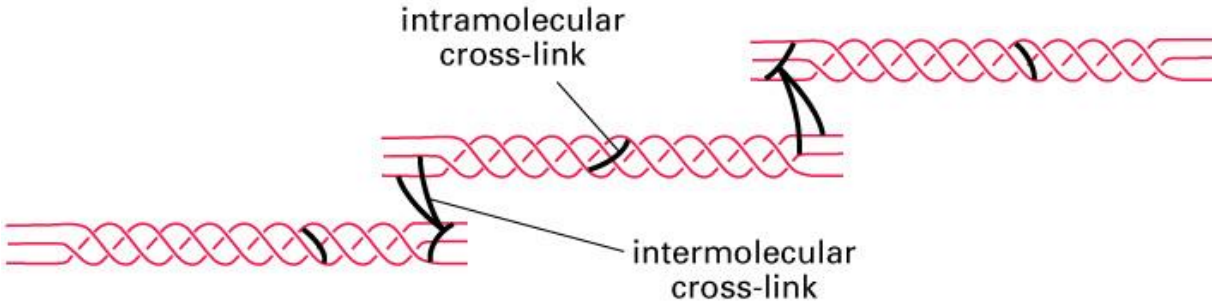


Figure 5 : Liaisons covalentes entre les tropocollagènes pour la formation de fibrilles. (Figure 19-46. Molecular Biology of the Cell, 4th Edition.)

I.2.2. La gélatine

La gélatine est le résultat de l'hydrolyse partielle du collagène, conduisant à une solution de protéine de haut poids moléculaire. C'est un hydrocolloïde, c'est-à-dire un polymère qui se solubilise en présence d'eau et qui va avoir un effet sur la texture de la solution. La composition en acides aminés de la gélatine est très proche de la composition du collagène. Dans le domaine de la production de gélatine, les sources principales de collagène sont les peaux et les os de mammifères, principalement provenant de l'industrie du porc (Babel et al., 2000). Quel que soient les coproduits utilisés, c'est principalement le collagène de type I dans la grande famille des collagènes qui est extrait pour la production de gélatine. Dans une moindre mesure, les collagènes de type II et III peuvent également être extraits (Tableau 2).

I.2.2.(1). Les sources de production de gélatine

Depuis 1930, la première source de matières premières utilisée pour la production de gélatine provient des coproduits de l'industrie porcine. Le porc présente l'avantage de posséder une peau très épaisse (épiderme derme) qui n'est pas bien valorisée dans l'industrie du cuir. La couenne de porc est donc une matière première à moindre coût et présente une capacité d'extraction élevée. Avant les années 1990, les coproduits de l'industrie bovine étaient très largement utilisés, mais la crise de l'encéphalopathie spongiforme bovine en Europe a contraint les industriels à limiter la production de gélatine bovines.

Depuis cette période, de nombreuses sources alternatives de collagène ont été étudiées, principalement pour des raisons socioculturelles telles que la prohibition de la consommation de porc liée au Judaïsme et à l'Islam. Cela a permis de mettre en avant les co-produits de l'industrie de la pêche.

Les gélatines de poisson présentent des températures de gélification et de fusion inférieures aux gélatines provenant de mammifères, ce qui va induire en général des forces en gel plus faibles. La diversité des espèces aquatiques étant plus importante que celle des mammifères, les espèces ont été regroupées en deux classes, celles provenant d'eaux chaudes ou celles d'eaux froides, dont sont extraites des gélatines ayant des propriétés physico-chimiques différentes. Aujourd'hui des gélatines de poulet et de poissons sont produites industriellement, mais ne sont pas favorisées du fait du prix de production nettement supérieur à celui des gélatines provenant de mammifères.

I.2.2.(2). Structure et composition de la gélatine

Il existe deux types de gélatines en fonction du prétraitement réalisé pour l'extraction du collagène. Celles issues d'un prétraitement acide (gélatine de type A) et celles issues d'un prétraitement basique (gélatine de type B) avant de subir une extraction en milieu aqueux à des températures supérieures à 45°C. L'ensemble des conditions d'extraction (type de prétraitement, durée et température d'extraction) confère aux gélatines ainsi extraites des propriétés physico-chimique très différentes (Zhou et Regenstein, 2006; Schrieber et Gareis, 2007; Eysturskarð *et al.*, 2009). Par exemple, même si la composition en acides aminés varie peu pendant la transformation du collagène en gélatine, un traitement basique va enlever la fonction amine (NH2) de la glutamine et de l'asparagine pour former les acides glutamique et aspartique (Weng *et al.*, 2014; Babel et al., 2000; Zhou *et al.*, 2006).

En général, la gélatine de mammifère est composée d'un tiers de glycine et d'un quart d'acides aminés (proline et hydroxyproline). De plus, elle est composée aussi bien d'acides aminés polaires qu'apolaires, ce qui lui confère un caractère amphiphile (Tableau 2).

Tableau 2 : Comparaison de la composition en acides aminés entre le collagène bovin (type I), la gélatine de peau de porc par extraction acide (type A) et de la gélatine d'os de porc par traitement basique (type B). Les acides aminés sont exprimés en nombre d'acides aminés pour 1000. (Babel et al., 2000)

Acide aminé	Collagène bovin Type I	Gélatine de Type A (peau de porc)	Gélatine de Type B (os de porc)
Alanine	114	112	117
Arginine	51	49	48
Asparagine	16	16	-
Acide Aspartique	29	29	46
Glutamine	48	48	-
Acide Glutamique	25	25	72
Glycine	332	330	335
Histidine	4	4	4
Hydroxyproline	109	97	97
Isoleucine	11	10	11
Leucine	24	24	24
Lysine	28	27	28
Méthionine	6	4	4
Phénylalanine	13	14	14
Proline	115	132	124
Sérine	35	35	33
Thréonine	17	18	18
Tyrosine	4	3	1
Valine	22	26	22

Pendant la phase d'extraction à chaud le collagène perd sa structure en triple hélice pour former des pelotes aléatoires. Après l'extraction, les triples hélices ne retrouvent que partiellement leur structure quand la solution refroidit. L'eau est piégée entre les chaînes et une structure en réseau apparaît.

Pour les gélatines extraites à partir de collagène de type I, une chaîne α est composée de 1014 acides aminés pour une masse molaire moyenne de $1\times10^5 g.mol^{-1}$. Les chaînes β (association de deux chaînes α) et les chaînes γ (association de trois chaînes α) ont respectivement une masse molaire moyenne de $2\times10^5 g.mol^{-1}$ et de $3\times10^5 g.mol^{-1}$(Rbii, 2010; Bowen et Mark, 1989). Comme pour le collagène ce sont les liaisons hydrogène intra et inter chaînes qui stabilisent le réseau de gélatine (Figure 6).

La gélatine reste un mélange complexe composé des chaînes α, β, γ, mais aussi de structures à haut poids moléculaire obtenues par l'hydrolyse partielle des fibres de collagène (Eysturskarð et al., 2009).

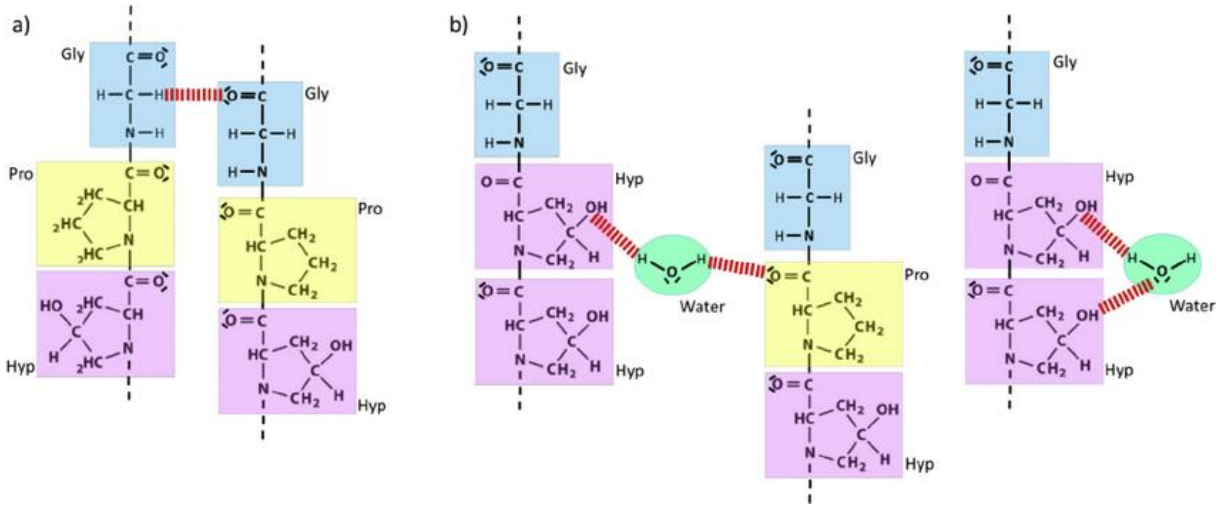


Figure 6 : Exemple de liaisons hydrogène inter-chaînes (a), inter-chaînes avec une molécule d'eau (b) et intra-chaînes avec une molécule d'eau (c). D'après Duconseille et al 2015 (Duconseille et al., 2015)

La formation du gel de gélatine dans l'eau est un phénomène réversible et dépendant de la température (Flory et Weaver, 1960). Une diminution de la température en dessous de la température de fusion va entraîner un réarrangement partiel de la gélatine. Durant ce phénomène de renaturation, l'organisation spatiale des chaînes et leurs interactions sont dépendantes de la vitesse de refroidissement et de la concentration en gélatine.

A faible concentration (< 0,1% en masse), les liaisons intramoléculaires sont favorisées, les brins se replient sur eux-mêmes pour former des triples hélices. Plus le refroidissement est rapide, plus la formation de boucle à l'intérieur de cette triple hélice est favorisée, diminuant ainsi la stabilité de la structure. En revanche, à forte concentration en gélatine (>0,1% en masse), les liaisons inter moléculaires sont favorisées et les chaînes s'associent entre elles pour reformer des triples hélices. Dans ce cas, plus le refroidissement est rapide, plus les chaînes s'agrègent de façon désorganisée, ce qui conduit à la formation d'agrégats (Harrington et Rao, 1970).

Récemment, *Guo et al.* (Guo *et al.*, 2003), ont proposé un modèle de réorganisation prenant en compte à la fois le nombre de brins impliqués et le nombre de repliements (Figure 7, A). Ils classent ensuite ces structures en fonction du nombre de chaînes impliquées pour former ces structures. On parle d'ordre 1 lorsque qu'il n'y a qu'une chaîne α, d'ordre 2 pour deux chaînes et d'ordre trois pour trois chaînes. La stabilité de cette réorganisation sera directement liée à la longueur des chaînes impliquées.

Le rôle de l'hydroxyproline et de la proline est très important pour la stabilité de la structure de la triple hélice, notamment lorsque l'hydroxyproline est positionnée en Y dans la séquence primaire Gly-X-Y (Shoulders et Raines, 2009). La stabilité de la triple hélice est donc en partie liée au taux d'hydroxyproline, c'est pourquoi une gélatine de poisson (pour la morue 5‰ résidus d'hydroxyproline) possède un point de fusion très inférieur à la gélatine de porc (97 ‰ résidus d'hydroxyproline)(Wasswa *et al.*, 2007; Leuenberger, 1991; Jamilah et Harvinder, 2002). Cependant, il n'est pas possible d'estimer les propriétés physico-chimiques d'une gélatine sur la seule base de la séquence primaire en acides aminés, car seul l'énantiomère R (du groupement hydroxyle) du carbone chiral en 4ème position de l'hydroxyproline est capable de créer des liaisons hydrogène par l'intermédiaire de molécules d'eau (Te Nijenhuis, 1997).

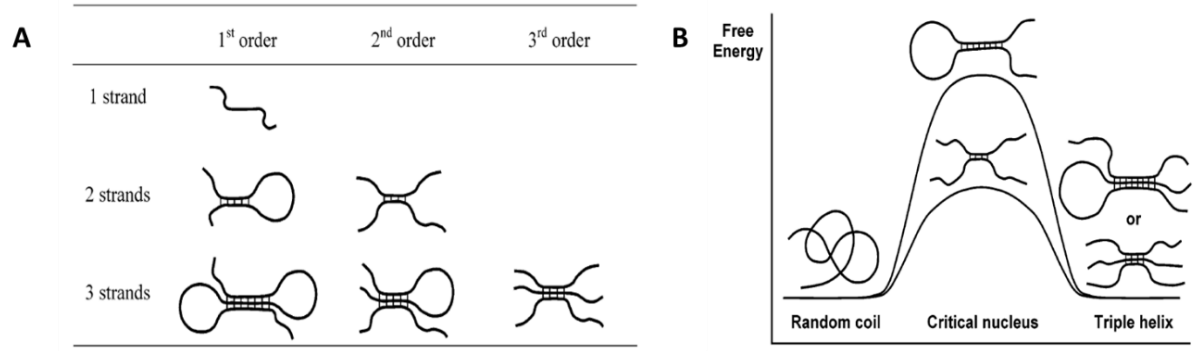


Figure 7 : Réorganisation des chaînes de gélatine en solution (repliement des chaînes et nombre de brins impliqués) (A). Schéma du mécanisme de formation de la triple hélice à partir de deux brins. Les triples hélices avec une boucle sont formées en solution diluée et les triples hélices sans boucle sont formées dans des solutions concentrées, les triples hélices avec une seule boucle coexistent avec les triples hélices sans boucle dans les solutions semi-diluées (B). D'après les travaux de Guo et al. (Guo et al., 2003)

I.2.2.(3). Propriétés physico-chimiques de la gélatine

Les propriétés physico-chimiques des gélatines sont liées à plusieurs paramètres ; ceux directement liés au milieu (température, pH, concentration, force ionique) et ceux liés à la structure (structure primaire, masse molaire). La structure primaire est principalement liée à la matrice utilisée (poisson d'eaux chaudes ou froides, porc, bovin, volaille...). En revanche, les autres paramètres sont directement liés aux conditions d'extraction de la gélatine. Dans l'industrie, la gélatine est commercialisée sous forme de grains (ou poudre) blanchâtres, avec le moins de goût et d'odeurs possibles. En fonction des méthodes de déshydratation, les gélatines sont commercialisées entre 10 et 13% d'humidité, ce qui leur confère une stabilité relative lorsqu'elles sont correctement stockées.

Les gélatines possèdent de nombreuses propriétés fonctionnelles (formation de gel, épaississante, émulsifiante, stabilisante, liante...), qui sont liées à leurs propriétés physico-chimiques. Les industriels caractérisent les gélatines en mesurant deux paramètres rhéologiques qui sont la force en gel et la viscosité. D'après les travaux de *Johnston et al.* (Johnston-Banks, 1990), la force en gel est proportionnelle à la somme des chaînes α et β, alors que la viscosité (et le point de fusion) est fonction de la présence de fractions de haut poids moléculaire (supérieur au poids de la chaîne γ) (Wasswa, Tang, et Gu, 2007; Leuenberger, 1991). De récentes études se sont penchées sur l'évolution de la dissolution de la gélatine en fonction de l'âge de celle-ci (Rbii *et al.*, 2009; Duconseille, Gaillard*, et al.*, 2017; Duconseille, Wien*, et al.*, 2017; Duconseille, Traikia*, et al.*, 2017).

Le paramètre le plus important dans l'industrie est la force en gel, appelée aussi « Bloom » par les professionnels du domaine. Cette mesure est spécifique de la caractérisation des gels. Pour mesurer la force en gel, un gel de gélatine à 6,67% (m/V) est préparé et placé pendant 17 heures à 10 °C dans une verrerie spécifique. Puis la résistance de ce gel est déterminée, à l'aide d'un gélomètre, qui mesure la force nécessaire (en grammes) pour faire pénétrer un piston de 12,7 mm de diamètre sur 4 mm de profondeur dans le gel. La valeur de force en gel varie entre 0 et 350 grammes ; la gélatine industrielle a une valeur généralement comprise entre 50 à 300 grammes. La plupart des applications industrielles sont déterminées par la valeur de force en gel.

La viscosité est le deuxième paramètre le plus important pour les industriels. Elle permet de déterminer les propriétés d'écoulement du milieu. La mesure de viscosité est réalisée en calculant le temps d'écoulement à 60°C d'une solution de gélatine à 6,67% (m/V) dans une pipette viscosimétrique de 100 mL. Le temps d'écoulement est ensuite converti en millipoise (Mps, unité de mesure de la viscosité dynamique). La viscosité est directement liée à la température et à la concentration du milieu ; c'est donc une valeur qui dépend des conditions de son utilisation dans la fabrication de produits transformés. La viscosité des gélatines industrielles (6,67 % (m/V) à 60°C) varie entre 15 et 75 Mps.

En plus de ces deux propriétés, les gélatines peuvent être caractérisées par le pH, le point isoélectrique, la couleur, le taux de cendres ou la turbidité. Le point isoélectrique est aussi un paramètre important pour les industriels. En effet, il y a une différence entre les gélatines obtenues par extraction acide (type A) et celles obtenues par extraction basique (type B). Les gélatines de type A possèdent une gamme de pH isoélectrique comprise entre 7 et 9, alors que les gélatines de type B présentent une gamme de pH isoélectrique comprise entre 4,7 et 5,4 (Schrieber et Gareis, 2007). Les gélatines obtenues par extraction acide sont chargées positivement (H_3O^+), alors que les gélatines obtenues par extraction basique sont chargées négativement (OH^-). Les charges sont portées par les résidus des acides aminés et sont orientées vers l'extérieur de la structure en triple hélice.

I.3. Production industrielle de gélatine

I.3.1. Procédés d'extraction de la gélatine

D'après la pharmacopée des États-Unis et en accord avec la pharmacopée européenne, la gélatine est définie comme une protéine purifiée obtenue, à partir de collagène animal, par hydrolyse partielle acide et/ou basique, par hydrolyse enzymatique ou par hydrolyse à chaud (USP, 2016). Cela permet d'obtenir des hydrolysats avec des propriétés plus ou moins gélifiantes. La gélatine produite industriellement est principalement obtenue par traitement acide et/ou basique. Plus récemment, des procédés ont été développés pour produire de la gélatine par traitement enzymatique.

Quel que soit le procédé utilisé, le processus de production de gélatine peut se décomposer en trois grandes étapes : le prétraitement, l'extraction par hydrolyse et le raffinage.

I.3.1.(1). Le prétraitement de la matière première

Cette phase permet à la matière première d'absorber jusqu'à environ 25% de son poids en eau grâce à la rupture de certaines liaisons intra-moléculaires. Le choix du prétraitement dépend du type d'extraction réalisé et de l'origine de la matière première. Il existe trois types de prétraitements : acide pour obtenir des gélatines de type A, basique pour obtenir les gélatines de type B et enzymatique pour obtenir des hydrolysats de collagène.

- **Le traitement acide** : c'est le plus répandu industriellement car il présente l'avantage d'être plus rapide (entre 10h et 48h) que le traitement alcalin. Il permet d'obtenir des gélatines de type A. Ce traitement est principalement utilisé pour la production de gélatine alimentaire et sur des matières premières peu réticulées et grasses (os déminéralisés (osséine)). Avant cette étape, les matières subissent un pré-nettoyage à l'eau (parfois faiblement acidifiée). Elles sont ensuite traitées à température ambiante (15-20°C) avec une solution d'acide minéral diluée de pH inférieur à 3 (habituellement des acides chlorhydrique, sulfurique ou phosphorique). Ce prétraitement est suffisant pour permettre de rompre les liaisons peptidiques du collagène. Généralement à la fin du processus, l'excès d'acide est éliminé et les matières premières sont rincées à l'eau froide. Après cette étape, le pH des couennes est compris entre 3,5 et 4,5.
- **Le traitement basique** : il est utilisé pour les peaux ou les os déminéralisés. Le traitement est réalisé à pH 12 en présence de chaux ou d'hydroxyde de sodium à température ambiante. Le temps de prétraitement va dépendre de la matière première à extraire et peut durer entre 6 et 20 semaines. Ce traitement est principalement utilisé pour la production de gélatine dans le domaine pharmaceutique. Ce traitement permet aussi d'enlever les impuretés et de casser certaines réticulations pour rendre le collagène soluble dans l'eau. Dans certains cas, la matière première est placée dans un bain d'acide phosphorique pendant un jour avant l'extraction à chaud pour ramener le pH à 7.
- **Les procédés enzymatiques :** leur utilisation est beaucoup plus récente et moins commune que les prétraitements acide et basique. Il existe plusieurs enzymes d'origine animale ou végétale qui permettent l'extraction du collagène. Les plus communes sont : la papaïne, la bromélase, la pepsine ou encore les protéases brutes extraites des viscères de différents poissons. En comparaison avec les traitements acide et alcalin, le traitement enzymatique augmente le rendement d'extraction pour des temps de traitement beaucoup plus courts (quelques jours). Cependant, les méthodes enzymatiques ont tendance à dégrader les chaînes α et β, ce qui réduit la qualité en termes de force en gel de la gélatine (Ahmad *et al.*, 2017; Wasswa, Tang, et Gu, 2007).

De plus, le rendement et la qualité d'extraction de la gélatine peuvent être affectés par l'activité de protéases inhibitrices endogènes, si elles ne sont pas correctement inhibées. Depuis de nombreuses années, les méthodes enzymatiques sont au cœur des projets de recherche des entreprises de production de gélatine.

I.3.1.(2). La phase d'extraction par hydrolyse

C'est la phase de transformation du collagène en gélatine. Cette étape est différente selon les industriels et est fonction des attentes des consommateurs en termes de propriétés. Elle dépend de la température, du pH et du temps d'extraction ainsi que des équipements industriels. En général, les matières premières subissent plusieurs extractions successives (3 à 6 habituellement) à des temps et températures différents. La première extraction est généralement réalisée entre 50 et 60°C, les températures augmentent pour les extractions suivantes jusqu'à environ 100°C à la dernière extraction. Entre les étapes d'extractions, la phase solide est égouttée et la phase liquide, qui contient la gélatine, poursuivra le processus avec la phase de raffinage. Les gélatines obtenues par les extractions successives présentent des propriétés physico-chimiques différentes. Les gélatines issues de la première extraction possèdent une force en gel et des masses moléculaires les plus importantes ainsi qu'une coloration moins importante que les gélatines issues des extractions suivantes. A l'inverse, la dernière extraction, obtenue à haute température, confère aux gélatines des forces en gel faibles et une plus grande quantité de fragments de faible masse molaire ainsi qu'un brunissement plus important (réaction de Maillard).

I.3.1.(3). La phase de raffinage

L'étape de raffinage est plus ou moins complexe en fonction des industriels et du devenir de la production. On retrouve cependant des étapes indispensables et communes à l'ensemble des industriels.

- **La filtration :** elle a pour but d'éclaircir la phase liquide et d'enlever les impuretés. Il peut y avoir plusieurs seuils de filtration pendant l'étape de raffinage.

- **La déminéralisation :** l'objectif de cette étape est d'enlever les ions présents dans la solution. La présence d'ions dépend de la matière première, mais aussi du prétraitement qui utilise des acides ou bases minérales. Pour cette étape la phase liquide passe successivement à travers 2 types de colonne de résine échangeuse d'ions : cationique et anionique.
- **La concentration :** l'objectif de cette étape est de concentrer la phase liquide en gélatine. Les gélatines qui gélifient, sont concentrées de 25% à 45%, afin d'obtenir un gel de gélatine à température ambiante.
- **La stérilisation :** l'objectif de la stérilisation est d'éliminer tous les germes microbiens. Le gel subit une stérilisation Ultra Haute Température (UHT) ce qui correspond à une augmentation de la température du produit entre 120 et 150°C pendant quelques secondes afin de ne pas dégrader le produit.
- **Le séchage :** dans cette phase, le gel va être déshydraté pour atteindre un taux d'humidité compris entre 10% et 13%. Il existe plusieurs méthodes de séchage utilisées par les industriels. En fin de séchage, la gélatine est sous forme solide.

I.3.2. Industrie de la gélatine

I.3.2.(1). Contexte mondial de la production de gélatine

En 2011, la production mondiale de gélatine était estimée à 349 000 000 tonnes pour un marché de 1,77 milliards de dollars (Grand-View-Research, 2013). Entre 2011 et 2015, il a y eu une augmentation de 18% de la production de gélatine soit une production mondiale de gélatine de 413 kilos tonnes. Avec un taux de croissance annuel moyen estimé à 5,3% entre 2016 et 2024, la production mondiale devrait atteindre 652 kilos tonnes en 2024. La gélatine est très majoritairement utilisée dans le domaine de l'agroalimentaire et dans le domaine pharmaceutique.

Le marché de la gélatine est en constante augmentation. Cela s'explique par la diversité des applications dans ces domaines due aux propriétés stabilisatrices et liantes spécifiques au produit. De plus, l'utilisation de nouveaux produits alimentaires dans des pays émergents augmente la demande mondiale (Moyen Orient, Chine, Inde) (Grand-View-Research, 2013).

S'il existe une grande diversité d'hydrocolloïdes, principalement de nature protéique ou glucidique (polysaccarides), dans le domaine de l'agroalimentaire, la gélatine reste l'agent de texture le plus répandu suivi des amidons et amidons modifiés (Taggart et Mitchell, 2009). Dans une moindre mesure, le marché de la gélatine est aussi en concurrence avec ceux des gommes de xanthane, des pectines, des alginates ou encore de l'agar-agar (Sworn, 2009; Armisén et Galatas, 2009; Draget, 2009).

La peau de porc reste la principale source de production, avec environ 41% de la production mondiale en 2015. On retrouve ensuite la peau de bovin (environ 28%), les os de bovin (environ 27%), les autres sources représentant 4% de la production mondiale. Cependant l'augmentation de la demande en nourriture hallal dans le domaine de l'agroalimentaire dans les pays du Moyen Orient devrait favoriser l'utilisation de peau de bovin dans la production de gélatine.

I.3.2.(2). Propriétés technologiques de la gélatine

Dans l'industrie, la gélatine est utilisée pour de nombreuses applications qui vont dépendre de plusieurs paramètres. Ainsi le rôle fonctionnel de la gélatine ne provient pas d'un seul paramètre physico-chimique, mais plutôt d'une synergie complexe de paramètres :

- Structure primaire de la protéine
- Conformation (structures supérieures)
- La charge nette et la distribution de ces charges
- Le ratio d'acides aminés hydrophobes et hydrophiles
- La taille et la conformation
- Liaisons faibles intra et intermoléculaires (liaisons hydrogènes, ioniques, Van der Waal's)

C'est l'ensemble de ces paramètres qui va déterminer le rôle fonctionnel de la gélatine.

Parmi les propriétés d'intérêt de la gélatine on peut citer les fonctionnalités suivantes : gélifiant, épaississant, foisonnant, émulsifiant, filmogène, floculant, stabilisant et liant.

I.3.2.(3). Cas de la gélatine dans l'industrie

La gélatine est majoritairement utilisée dans le domaine de l'agroalimentaire, notamment pour ses propriétés nutritionnelles et fonctionnelles. Ce sont principalement des gélatines de type A qui sont utilisées comme ingrédients à part entière. La gélatine est omniprésente dans ce domaine, elle sert notamment pour la confection de confiseries, de desserts, de produits à base de viande, de produits de boulangerie ou encore dans la production de boissons. Le Tableau 3 présente des exemples d'utilisation de la gélatine dans l'alimentation en fonction de ses propriétés fonctionnelles (Gelatine-Manufacturers-Europe).

Tableau 3 : Propriétés fonctionnelles de gélatines dans l'agroalimentaire (Handbook, 2012)

Fonction	Application
Formation de gel	Dessert gélifié, viandes, confiserie
Formation de mousse	Guimauve, nougats, mousses, soufflés, crème fouettée
Colloïde protecteur	Confiserie, glaçage, crèmes glacés, desserts surgelés
Agent liant	Roulés de viande, viande en conserve, confiserie, fromages, produits laitiers
Agent de clarification	Bière, vin, jus de fruits, vinaigre
Formation de film	Revêtement pour fruits, viandes, articles délicats
Epaississant	Boisson en poudre, bouillon, sauces, soupes, gelées, sirop, produits laitiers
Vecteur dans un process	Microencapsulations de couleurs, de saveurs, huiles, vitamines
Emulsifiant	Soupes à la crème, sauces, arômes, pâtés de viandes, crème fouettée, confiserie, produits laitiers
Stabilisateur	Crèmes de fromages, chocolat au lait, yaourts, glaçage, desserts glacés
Agent adhésif	Fixer des pièces non confectionnées, de la noix de coco et d'autres objets aux confiseries, lier les glaçages aux produits de boulangerie, lier les assaisonnements aux produits à base de viande

Dans l'industrie pharmaceutique la gélatine est aussi omniprésente. Elle sert à la confection de nombreux produits pharmaceutiques tel que : les gélules, les capsules dures et moles, en enrobage de comprimés ou en encapsulation et microencapsulation. Son origine alimentaire lui permet d'être un ingrédient de texture facile à intégrer en formulation pharmaceutique.

I.4. La spectroscopie infrarouge

I.4.1. Généralités sur la spectroscopie infrarouge

La spectroscopie proche infrarouge (SPIR) est une méthode de spectroscopie optique, ce qui correspond à l'interaction d'un rayonnement électromagnétique avec une matière. Le spectre électromagnétique est divisé en plusieurs régions en fonction de la fréquence du rayonnement. En classant les régions par fréquences décroissantes on trouve les rayons γ, les rayons X, les ultraviolets, le domaine du visible, l'infrarouge et les ondes radio (Figure 8). En fonction du type d'interaction entre la matière et le rayonnement, cela peut concerner plusieurs types de niveaux d'énergie de la matière. Ces différents niveaux d'énergie correspondent à différents types de transitions atomiques ou moléculaires. La spectroscopie infrarouge est basée sur les transitions vibrationnelles. Elles correspondent à une augmentation des vibrations des molécules en fonction de la longueur d'onde d'excitation. Cela entraîne une absorption sélective de l'énergie lumineuse (Bertrand et Dufour, 2006b).

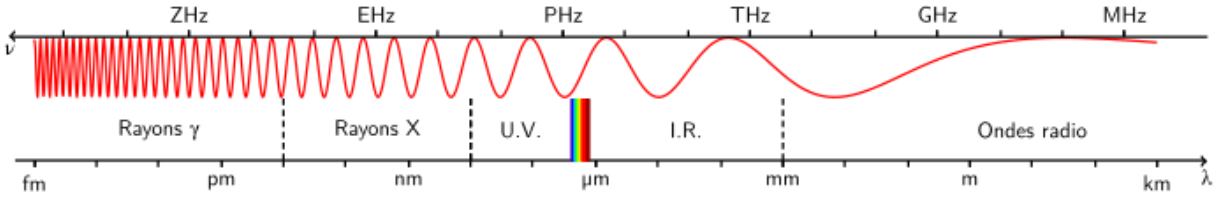


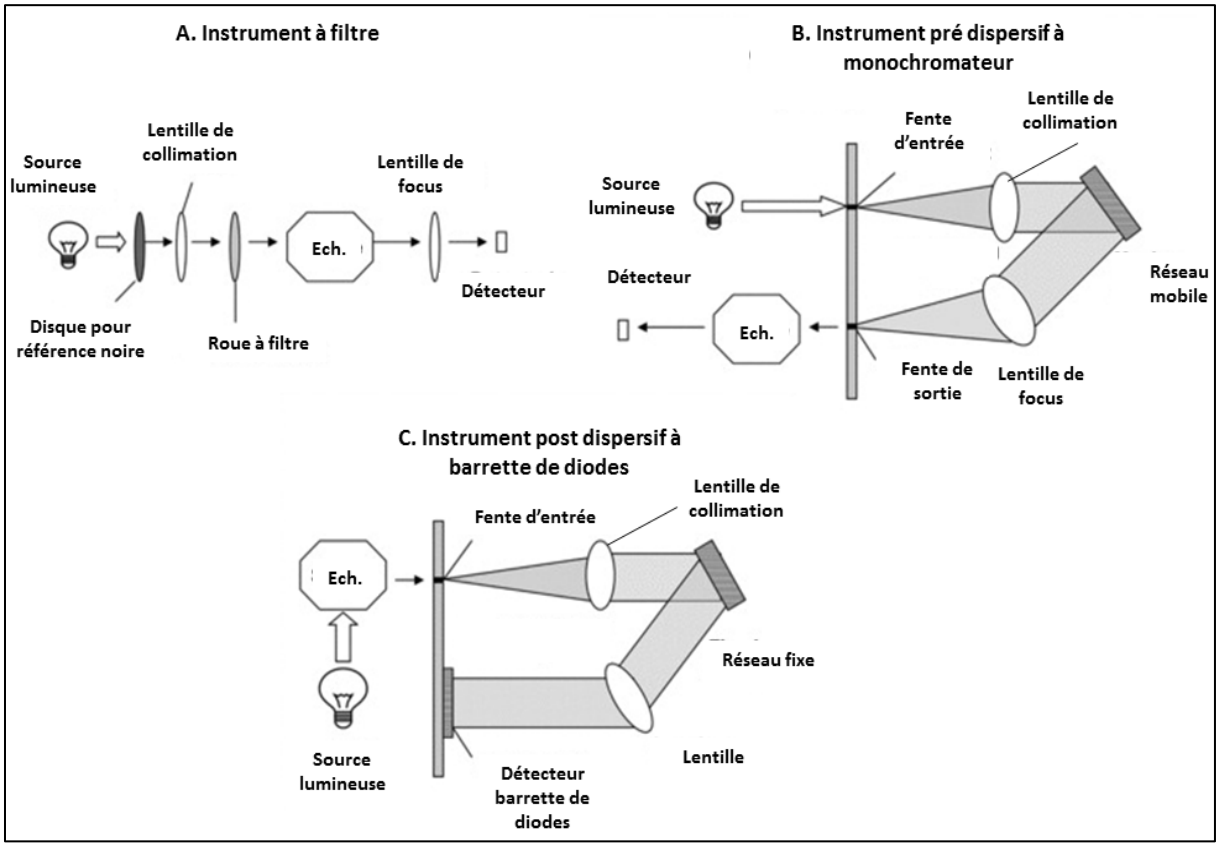
Figure 8 : Spectre Electromagnétique (image libre de droit), UV pour Ultra-violet, IR pour Infra-Rouge

D'après la norme ISO 20473 de 2007 (ISO, 2007), le domaine de l'infrarouge se décompose en trois sous domaines : l'infrarouge lointain entre 50 µm et 50 mm, le moyen infrarouge entre 3 µm et 50 µm et le proche infrarouge entre 780 et 3000 nm. Les bandes les plus importantes sont liées aux harmoniques et aux combinaisons de vibrations fondamentales des groupes fonctionnels CH, NH, OH et dans une moindre mesure SH pour la matière organique. Les molécules ont tendance à absorber plusieurs longueurs d'onde dans la même région. Les pics d'absorption se chevauchent souvent avec ceux d'autres molécules ce qui créé des spectres d'absorption complexes. Les bandes d'absorption SPIR sont généralement larges et 10 à 100 fois plus faibles en intensité que leurs bandes d'absorption fondamentales qui se situent dans le moyen infrarouge. Ces caractéristiques limitent la sensibilité au sens spectroscopique de la SPIR et nécessitent un traitement chimiométrique des données afin de relier l'information spectrale aux propriétés de l'échantillon.

I.4.2. Les différents types de spectromètres infrarouge

Il existe différents types de spectromètre. D'une façon générale, un spectromètre est composé de plusieurs éléments : une source lumineuse, un système de sélection des longueurs d'onde, un système de détection du signal, un système de traitement du signal, et un emplacement pour positionner l'échantillon (Givens et al., 1997; Agelet et Hurburgh, 2010). Nous pouvons distinguer trois principaux types de spectromètres infrarouge détaillés par *Agelet and Hurburgh* (Agelet et Hurburgh, 2010) (Figure 9).

Figure 9 : Trois conformations instrumentales traditionnelles en spectroscopie proche infrarouge. (A) Instrument à filtre. (B) Instrument pré-dispersif à monochromateur. (C) Instrument post dispersif à barette de diodes.



L'industrialisation de la SPIR a amené à certaines adaptations du fait du principe de la chaîne de production. Il existe différentes méthodes d'acquisition pour répondre aux contraintes de l'industrie en termes de suivi des données liées à la production. On parle de « Off-line » lorsque l'échantillon est prélevé pendant le processus de production pour être analysé en laboratoire. Lorsque l'appareil est installé sur la chaîne de production et que les mesures sont réalisées ponctuellement (bien souvent par l'intervention d'un opérateur), on parle alors de « At line ». On parle de « In-line » lorsqu'une partie du flux de la production est dévié, pour être analysé par un appareil de façon continue ou ponctuelle. Enfin, on parle de « On-line » lorsque l'instrument est directement installé sur la chaîne de production et permet d'analyser les échantillons de façon continue ou ponctuelle. L'ensemble de ces méthodes d'acquisition est schématisé sur la Figure 10.

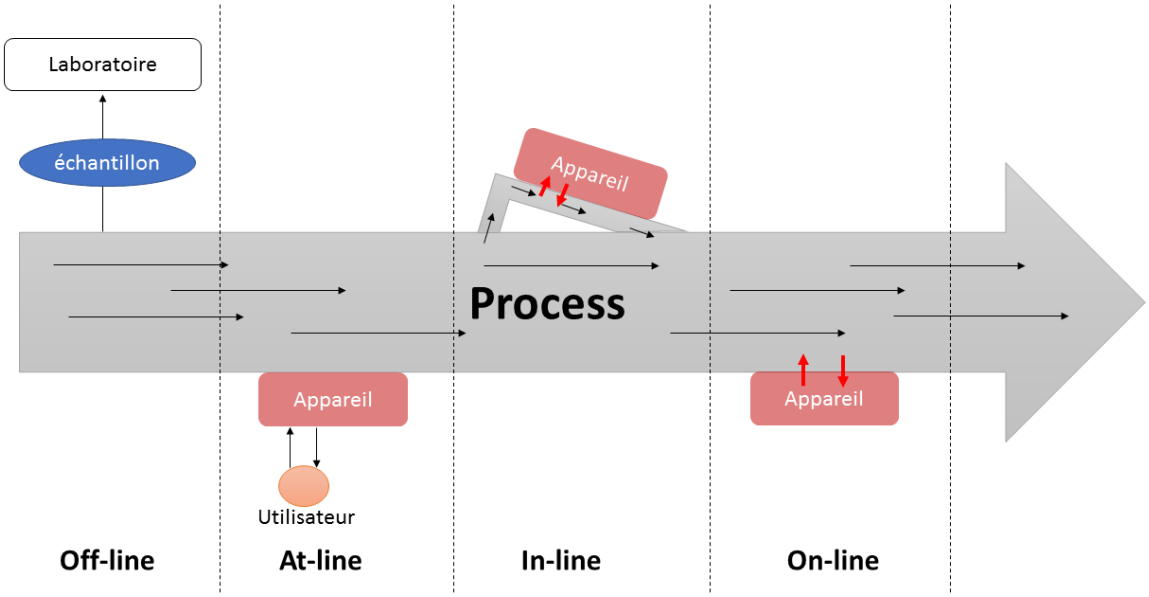


Figure 10 : Schéma des utilisations de la spectroscopie proche infrarouge sur une chaîne de production.

I.4.3. Mode d'acquisition

Il existe deux principaux modes d'acquisition de spectres infrarouge, la transmittance et la réflectance, qui se basent tous deux sur la loi de *Beer-Lambert-Bouguer.* A une longueur d'onde donnée l'absorbance est liée linéairement aux concentrations en constituants de la matrice analysée (Équation 1).

Équation 1

$$A_\lambda = \varepsilon_\lambda \times l \times [C]$$

Avec A qui représente l'absorbance à une longueur d'onde donné, ε_λ ($L \cdot mol^{-1} \cdot cm^{-1}$) qui représente le coefficient d'absorption molaire pour un longueur d'onde, *l* la largeur du milieu (cm) et [C] la concentration en différents constituants ($mol.L^{-1}$).

Dans un système en transmission, la lumière va traverser le milieu à analyser, puis être mesurée par le détecteur. Dans ce cas la loi de *Beer-Lambert-Bouguer* est strictement appliquée et est décrite par l'Équation 2.

Équation 2

$$A_\lambda = -log_{10}(T) = \varepsilon_\lambda \times l \times [C]$$

Avec A qui représente l'absorbance à une longueur d'onde donné, T qui représente la transmittance, ε_λ (L·mol^{-1}·cm^{-1}) qui représente le coefficient d'absorption molaire pour un longueur d'onde, *l* la largueur du milieu (cm) et [C] la concentration en différents constituants (mol.L^{-1}).

Par analogie, dans un système en réflexion, une partie de la lumière va être absorbée par le milieu et l'autre partie renvoyée et mesurée par le détecteur. Dans ce cas, l'absorbance ne suit pas strictement la loi de *Beer-Lambert-Bouguer* et peut être décrite par l'Équation 3

Équation 3

$$A_\lambda = -log_{10}(R) \approx \varepsilon_\lambda \times l \times [C]$$

Avec A qui représente l'absorbance à une longueur d'onde donnée, R qui représente la réflectance, ε_λ (L·mol^{-1}·cm^{-1}) qui représente le coefficient d'absorption molaire pour un longueur d'onde, *l* la largueur du milieu (cm) et [C] la concentration en différents constituants (mol.L^{-1}).

I.4.4. L'application de la spectroscopie proche infrarouge en agro-alimentaire

Une des premières applications a été l'utilisation de la SPIR pour déterminer le taux d'humidité dans des graines (Norris, 1964). Par la suite la SPIR a également été utilisée pour l'analyse des protéines, des glucides ou encore des lipides (Wold *et al.*, 2011; Nicolaï *et al.*, 2007). La spectroscopie présente plusieurs avantages. Elle permet des mesures rapides, peu coûteuses après modélisation, ne demande ni l'utilisation de produit chimique ni de technicité. C'est de plus une méthode non-destructive et bien souvent qui ne nécessite pas de préparation au préalable de l'échantillon (Bertrand et Dufour, 2006a). Ces avantages font qu'elle est très utilisée dans divers domaines : l'agroalimentaire (Olsen *et al.*, 2008; Munera *et al.*, 2017; Chen *et al.*, 2017), la (phyto)pharmacologie (Krämer et Ebel, 2000; Roggo *et al.*, 2007; Lafhal *et al.*, 2016), la chimie , dans la synthèse de polymère (Rohe *et al.*, 1999; Pomerantsev *et al.*, 2017), et plus généralement en médecine (Kasai *et al.*, 2017).

C'est à la fin des années 1960 que les industriels de l'agroalimentaire commencent à s'intéresser à la spectroscopie photonique (l'étude de l'interaction d'une onde électromagnétique avec la matière), dans le but d'améliorer la qualité et la sécurité des aliments et notamment des process de production. Par ailleurs le développement de l'informatique, de la chimiométrie et des technologies optiques a participé au développement des applications en ligne (Siesler *et al.*, 2008). La spectroscopie proche infrarouge est la méthode spectroscopique qui, par l'étendue de ses applications, est la plus représentée pour analyser les composés organiques.

Les produits carnés ont fait rapidement l'objet de développements de cette technologie en raison de leur valeur économique, souvent supérieure à celle des autres types d'aliments. Ainsi de nombreuses applications ont été développées sur les muscles, notamment pour en évaluer la qualité (Prieto *et al.*, 2009; Prieto *et al.*, 2017). Puis les applications ont concerné la capacité de prédiction de la composition chimique : le taux d'humidité, le taux de protéines, les taux d'acides gras, le taux de graisse intramusculaire ou encore le taux de cendre(Balage *et al.*, 2015; Cozzolino *et al.*, 2002; Cozzolino et Murray, 2004; Viljoen *et al.*, 2007). De même les propriétés technologiques des viandes ont été étudiées (Mitsumoto *et al.*, 1991) , notamment le pH, les couleurs dans l'espace CIE L*, a* et b* (CIE, 1978) et plus récemment la capacité de rétention d'eau (Leroy *et al.*, 2004). Il est aussi possible de trouver dans la littérature des travaux sur les propriétés sensorielles, sur la saveur, l'odeur, le goût, la fermeté ou encore la pâleur (Chan *et al.*, 2002). Le Tableau 4 résume les publications scientifiques qui évaluent la capacité de prédiction des différents paramètres technologiques ainsi que la composition chimique par spectroscopie proche infrarouge sur du porc.

Tableau 4 : Prédiction des paramètres technologiques et de la composition par spectroscopie proche infrarouge sur du porc.

Type d'échantillon	Nb d'échantillons	Paramètre étudié	Gamme spectrale (nm)	Analyse multivariée	R^2Cal	SECV	Référence
Viande	39	Taux d'humidité	400-2500	PLSr	0,46	1.13	(Brøndum *et al.*, 2000)
		Graisse intramusculaire			0,70	1.32	
Carcasse	102	Graisse intramusculaire	380-1700	MLR	0,35	3,6	(Hoving-Bolink *et al.*, 2005)
		Couleur (a*)		PLSr	0,31	1,0	
Muscle	44	Graisse intramusculaire	400-2500	PLSr (CV)	0,87	1,8	(Barlocco *et al.*, 2006)
		Taux d'humidité			0,90	1,1	
Muscle	207	Graisse intramusculaire	800-2500	PLSr	0,70-0,86	0,26-0,36	(Savenije *et al.*, 2006)
		pH	400-1100		0,83-0,91	0,45-0,63	
		a*	400-1100		0,76-0,88	0,57-0,59	
		L*	400-800		0,85-0,89	1,39-1,65	
		b*	400-800		0,68-0,81	0,46-0,52	
		Capacité de rétention d'eau	400-800		0,56-0,74	1,24-1,35	
Saucisse (hachée)	100	Taux de graisse	515-1650	PLSr	0,97-0,99	0,94-1,41	(Ortiz-Somovilla *et al.*, 2007)
		Taux d'humidité			0,96-0,99	0,79-0,99	
		Taux de protéine			0,91-0,93	0,6-0,63	
Muscle (Intact/broyé)	44	a*	400-2500	PLSr	0,93/0,96	1,3/1,7	(Cozzolino *et al.*, 2003)
		L*			0,30/0,61	5,3/4,5	
		b*			0,60/0,27	1,1/1,5	
Muscle	96	Taux d'humidité	1000-2500	PLSr	0,51	1.1	(Geesink *et al.*, 2003)
Muscle (Intact/ Haché)	296	pH	400-1100	PLSr	0,59-0,79	0,1	(Čandek-Potokar *et al.*, 2006)
		L*			0,53-0,37	2,9/3,2	
		a*			0,55-0,65	1,3/1,3	
		b*			0,76-0,71	1,0/1,1	
Carcasse	81	pH	400-1495	PLSr	0,80	0,11	(Balage et al., 2015)
	83	L*			0,88	2,02	
	83	a*			0,82	0,61	
	85	b*			0,80	1,07	
	78	Graisse intramusculaire			0,28	1,03	
Muscle	211	Taux de graisse	350-1150	PLSr	0,16-0,85	0,09-0,15	(Liao *et al.*, 2010)
		Taux de protéine			0,14-0,82	0,41-0,74	
		Taux d'humidité			0,42-0,83	0,78-1,34	
		pH			0,78-0,87	0,10-0,16	
Muscle	131	Taux d'humidité	400-2500	MPLS	0,80	0,8	(Kapper, 2012)
		L*			0,76	2,3	
		a*			0,54	1,2	
		b*			0,48	1,3	
		pH			0,39	0,2	

En complément des études sur la capacité de prédiction, d'autres se sont intéressées à une classification de carcasses basée sur le régime alimentaire de l'animal (Zamora-Rojas *et al.*, 2012), de la qualité des muscles (Prieto *et al.*, 2015) ou encore des conditions de stockage (Chen *et al.*, 2011).

Pour les industriels et dans des perspectives économiques, les études se sont principalement focalisées sur la qualité des carcasses de bœuf et de porc qui sont les matériaux bruts des processus (Kapper, 2012; Prieto *et al.*, 2006). Il faut attendre les années 1996 pour avoir les premières applications « *in-line* » et « *on-line* » sur de la viande (Isaksson *et al.*, 1996). Malheureusement, les applications développées sont très vite confrontées à un problème d'hétérogénéité des matières premières. C'est pourquoi la plupart des protocoles développés pour analyser la viande réalisent les mesures soit uniquement sur du tissu musculaire, soit sur des viandes après broyage (Pullanagari *et al.*, 2015; Su *et al.*, 2014; Prieto *et al.*, 2014).

I.4.5. La chimiométrie pour analyser les spectres proche infrarouge

I.4.5.(1). Classement des méthodes chimiométriques

La chimiométrie fait partie des méthodes d'analyses multivariées, c'est-à-dire qui étudient des bases de données avec plus de deux variables. Dans le domaine de la spectroscopie infrarouge, on parle de nombre d'échantillons (individus) et de données spectrales en longueur d'onde (variables). La chimiométrie permet, à l'aide de méthodes mathématiques, statistiques et informatiques, d'extraire de l'information utile provenant de données de mesures chimiques. La spectroscopie proche infrarouge a la spécificité de présenter des recouvrements importants des bandes d'absorption des liaisons moléculaires et est bien souvent utilisée sur des matrices qui sont chimiquement complexes. Afin d'exploiter au maximum le potentiel de cette spectroscopie, il est nécessaire d'utiliser des approches multivariées. Il existe un grand nombre de méthodes chimiométriques que l'on peut classer en deux grandes catégories ; les analyses supervisées et les analyses non-supervisées (Figure 11). Dans le domaine de la SPIR, deux méthodes sont principalement utilisées : les méthodes de régression et les méthodes de classification, chacune faisant intervenir soit des modèles linaires soit des modèles non-linéaires.

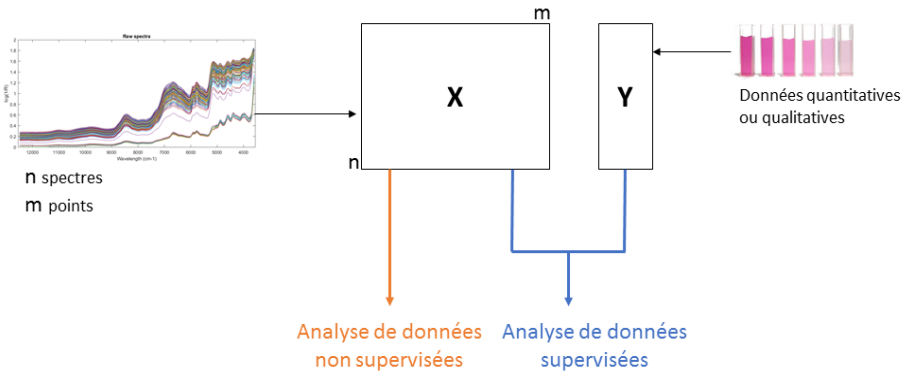


Figure 11 : Schéma de présentation des méthodes supervisées et non supervisées appliquées aux données analytiques

Les méthodes de classification permettent de regrouper les échantillons dans des groupes en se basant sur leurs différences spectrales. Avec une approche non supervisée, les échantillons sont classés en fonction de leurs différences spectrales (similarité, dissimilarité), alors qu'avec une approche supervisée, l'appartenance à un groupe est déterminée par un modèle prenant en compte une ou plusieurs variables explicatives. Dans les méthodes de classification fondées sur un modèle linéaire, on peut citer l'Analyse en Composantes Principales (ACP) pour les analyses non supervisées, ou l'analyse discriminante par les moindres carrés partiels (PLS-DA pour *Partial Least Square Discriminant Analysis*) et l'Analyse Discriminante (DA) pour les analyses supervisées. Pour les méthodes de classification fondées sur un modèle non-linéaire, la plus fréquente est l'analyse par des réseaux de neurone (*Artificial Neural Networks*, ANN) qui peut aussi bien être réalisée par des méthodes supervisées et non supervisées (Tufféry, 2011; Bertrand et Dufour, 2006b).

Les méthodes de régression sont des méthodes dites prédictives, dont l'objectif est de prédire au mieux à partir de données (spectrales), une ou plusieurs variables quantitatives (valeurs chimiques) sur de nouveaux échantillons. Dans le domaine de la spectroscopie, les trois principales méthodes de régressions linéaires sont la Régression Linéaire Multiple (MLR), la Régression sur Composantes Principales (PCR) et la Régression des moindres carrés partiels (PLSr pour *Partial Least Square Régression*). La méthode la plus répandue dans le domaine est la PLSr (Agelet et Hurburgh, 2010).

(a) L'Analyse en Composantes Principales

L'Analyse en Composantes Principales (ACP) est particulièrement adaptée à l'étude exploratoire de bases de données comprenant un grand nombre de variables et d'individus. L'ACP est un des outils les plus puissants pour représenter graphiquement la relation entre des échantillons entre eux en s'appuyant sur un nombre important de variables (les longueurs d'onde). L'objectif général de l'ACP est de réduire les données initiales *p* par une transformation orthogonale en remplaçant les variables initiales par un nombre *q* de facteurs appropriés appelés « composantes principales » non corrélées entre elles. La première composante principale représente la variabilité maximale contenue dans les spectres infrarouge et les composantes principales suivantes représentent une part de la variabilité totale de plus en plus faible (Martens et Naes, 1992; Esbensen *et al.*, 2002). L'ACP peut être définie par l'Équation 4 :

Équation 4

$$X = TP^T + E$$

Avec, **X** la matrice de spectres (variables), **T** qui représente les scores (matrice de coordonnées factorielles) *i.e.* les positions projetées des individus sur les composantes principales, **P** les loadings (matrice de vecteur propre) *i.e.* la corrélation entre les variables (régression de X sur T), **E** les résidus, qui correspondent à l'erreur (bruit de fond).

(b) La régression des moindres carrés partiels (PLSr)

La PLS est née, dans les années 80, de l'association de l'algorithme développé par *Wold* pour l'ACP et de l'approche des moindres carrés ou projection sur des structures latentes (Wold *et al.*, 1983). Elle est similaire en plusieurs points à l'ACP. Cette méthode de régression multivariée est basée sur une régression de la variable à prédire (y) sur des variables latentes (qui sont les combinaisons linéaires des variables explicatives, dans le cas de la SPIR ce sont les absorbances aux différentes longueurs d'onde). La PLS a la particularité de déterminer les variables latentes non seulement en tenant compte des variables explicatrices (spectres), mais aussi de la variable à prédire (y)(Esbensen et al., 2002; Bertrand, 2002).

Le modèle PLSr est défini par l'équation :

Équation 5

$$yi = a_0 + a_1xi_1 + a_2xi_2 + \ldots + a_pxi_p + ei$$

Avec x représentant des variables spectrales, y représentant la variable à prédire, a_i les coefficients et e_i les résidus

(c) L'Analyse Discriminante (DA)

L'Analyse Discriminante est une méthode supervisée. La classe d'appartenance de chaque individu est affectée a priori. Lors du développement du modèle pour un nombre de classes égal à n, (n-1) axes factoriels sont créés. Ces axes sont calculés de telle sorte qu'ils représentent le pouvoir de discrimination entre les classes (minimisation de la variance intra classes et maximisation de la variance inter classes). L'axe 1 représente le maximum de variance, les axes suivants représentant des pourcentages de variance décroissants. La visualisation des groupes d'individus est facilitée par la représentation du plan facteur 1/facteur 2.

La DA permet d'expliquer le regroupement en classes à partir des variables initiales (équivalent des contributions au composantes de l'ACP) (Tuffery, 2007). Enfin, les performances d'un modèle sont évaluées par une matrice de confusion. Le Tableau 5 présente une matrice de confusion pour 2 classes.

Tableau 5 : Matrice de confusion pour 2 classes

		Classe d'appartenance prédite par DA	
		Classe 1	Classe 2
Classe d'appartenance réelle	Classe 1	*a*	*b*
	Classe 2	*c*	*d*

Avec a= nombre de prédictions négatives correctes, b = nombre de prédictions positives fausses, c = nombre de prédictions négatives fausses, d = nombre de prédictions positives correctes

La précision du modèle (pourcentage de bons classements) est calculée selon l'Équation 6 :

Équation 6

$$précision = \frac{a + d}{a + b + c + d}$$

L'erreur du modèle est calculée selon Équation 7:

Équation 7

$$taux\ d'erreur = \frac{b + c}{a + b + c + d}$$

Le meilleur modèle est celui qui présente un taux d'erreur minimal, avec un taux de bon classement maximal (Levasseur-Garcia et Kleiber, 2015).

I.4.5.(2). Développement d'un modèle de prédiction infrarouge

Le développement d'un modèle en spectroscopie proche infrarouge nécessite une calibration entre les données spectrales et les valeurs de références sur un pool d'échantillons. Le développement d'un modèle peut s'organiser en cinq étapes différentes :

1) **L'échantillonnage** : Sélection d'échantillons représentatif de la gamme à prédire.
2) **Références** : Acquisition des caractéristiques physiques ou chimiques des échantillons (taux de protéine, taux de matière grasse, taux d'humidité, concentration en divers composés …)
3) **Acquisition spectrale** : Acquisition des spectres correspondant aux échantillons
4) **Prétraitement des spectres** : Appliqués aux spectres avant de développer le modèle de calibration, ces prétraitements mathématiques permettent de réduire le bruit de fond (lissage) et/ou d'augmenter le signal lié à la composition (différenciation). En effet, les variations de texture de la matrice analysée (taille des particules, hétérogénéité de surface) induisent des effets de diffusion de lumière perturbant la mesure spectroscopique. A cela, s'ajoute le bruit de fond, les variations de la distance d'acquisition, ou les variations de température ou de pression de l'environnement. Ces effets peuvent être à la fois additif et multiplicatif (Esbensen et al., 2002).
    - **Mean Center et autoscale :** La méthode de « *Mean Center* » consiste à centrer les données en soustrayant la moyenne de chaque variable (longueur d'onde). Le centrage des données sur la valeur moyenne permet de réduire la complexité du modèle, en réduisant souvent le nombre de variables à utiliser. Il est possible de faire une mise à l'échelle des spectres en divisant chaque variable (longueur d'onde) par son écart-type, ce qui permet à chaque variable d'avoir le même poids (Haaland et Thomas, 1988). La méthode qui applique un centrage des données, puis divise chaque variable par son écart-type est appelée « *autoscale* ».

- **Standard Normal Variate (SNV) :** La méthode SNV permet d'éliminer les effets de diffusion de lumière. Elle correspond à la normalisation des spectres pour chaque longueur d'onde, c'est-à-dire que chaque spectre est divisé par l'écart type de l'ensemble. Cette méthode permet de ramener chaque spectre à la même échelle, permettant ainsi leur comparaison.
- **Detrend :** Cette méthode consiste en une soustraction pour chaque variable (longueur d'onde) de la valeur de la fonction polynomiale de second degré de ces variables. Autrement dit, cela consiste à supprimer le décalage moyen de chaque échantillon. C'est une méthode de correction de la ligne de base qui est généralement combinée avec l'utilisation d'une méthode SNV.
- **Multiplicative Scatter Correction (MSC) :** Comme la SNV, la MSC est une autre méthode qui permet d'éliminer les effets de diffusion de lumière. Elle consiste à ajuster par la méthode des moindres carrés chaque spectre par rapport au spectre moyen.
- **Lissage et dérivée par la méthode de Savitzky-Golay (Savitzky et Golay, 1964) :** L'utilisation de ce type de prétraitement permet d'accentuer les différences spectrales, d'augmenter la résolution spectrale et de réduire la dérive de la ligne de base. L'utilisation d'une dérivée seule sur des données bruitées va augmenter le signal dû au bruit. C'est pourquoi une étape de lissage est réalisée préalablement. Cette étape permet de réduire le bruit grâce à des moyennes mobiles et des polynômes de lissage le long du spectre. Pour cela, une moyenne est calculée sur une fenêtre de X points et se décale sur l'ensemble du spectre. Il est important de bien définir la taille de la fenêtre, car l'utilisation d'une fenêtre trop large peut entraîner une perte d'information.

5) **Développement du modèle et validation** : Avant de développer les équations d'étalonnage, les bases de données sont divisées en deux parties : un jeu de données de calibration (75% des échantillons) et un jeu de données de validation externe (25% des échantillons). Les échantillons présents dans les différents sets sont choisis de façon aléatoire. Le jeu de données de calibration est utilisé pour développer l'équation de prédiction en utilisant la régression PLS afin de prédire les paramètres de caractérisation. Le jeu de données d'étalonnage est testé en validation croisée : les performances du modèle de prédiction sont calculées à partir des spectres issus du jeu de données d'étalonnage. En validation externe, le jeu de données est utilisé pour évaluer les performances du modèle de prédiction. Une erreur moyenne de prédiction est calculée pour chaque variable latente. Le nombre optimal de variables latentes pour le modèle est déterminé par la valeur minimale de l'erreur de prédiction en validation croisée ou en validation externe (Figure 12).

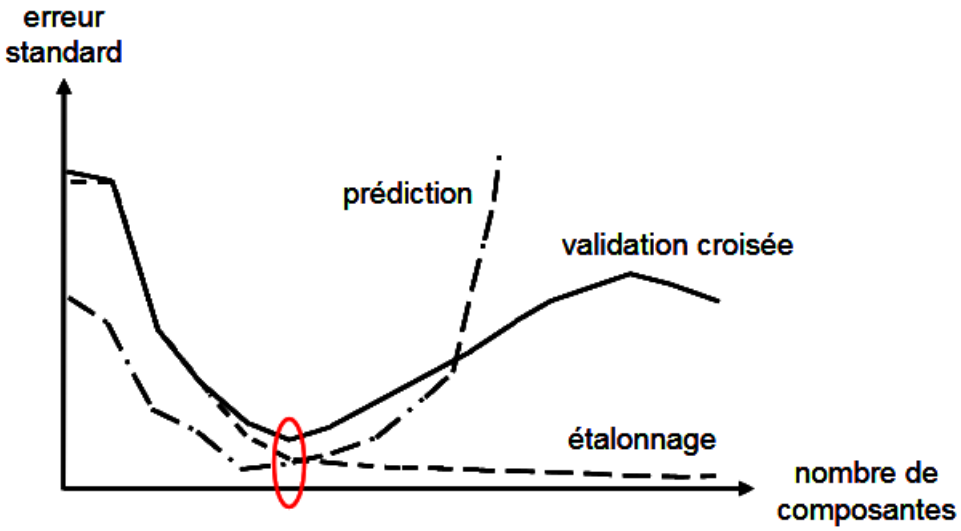


Figure 12 : Evolution de l'erreur standard dans les modèles de prédiction en fonction du nombre de composantes d'après Martens et al. (Martens et Naes, 1992)

6) **Evaluation des performances** : En général, les performances des modèles de prédiction sont principalement évaluées par le coefficient de détermination (R^2) et l'écart-type de prédiction (« *Standard Error of Prediction* » (SEP), ou « *Standard Error of Cross-validation* » (SECV)). Le coefficient de détermination permet d'avoir une estimation de l'écart entre la valeur de référence et la valeur prédite par rapport à la variance totale. L'erreur standard de prédiction est exprimée dans les mêmes unités que les valeurs de référence et permet d'obtenir des informations sur la précision de la calibration (Agelet et Hurburgh, 2010). Un dernier indice est utilisé spécifiquement dans le domaine de la spectroscopie proche infrarouge, c'est le déterminant prédictif relatif (« Relative Predictive Determinant » : RPD). Cet indice permet de déterminer la capacité de prédiction de nouveaux échantillons par rapport à la variabilité des données de calibration. La valeur de RPD peut être considérée comme un ratio entre l'écart-type de la valeur de référence et l'erreur de prédiction du modèle. En général, lorsque le RPD est supérieur à 3, l'équation est considérée comme applicable, un ratio compris entre 2 et 3 est prometteur pour le développement d'une équation de prédiction, mais nécessite encore des améliorations.

D'une manière générale, en validation croisée, les performances du modèle de prédiction sont considérées comme optimistes, les données servant au test d'évaluation étant celles qui ont participé à la constitution du modèle (Næs *et al.*, 2002).

Les paramètres statistiques d'évaluation de performance des modèles sont :

- Coefficient de détermination (R^2) : $r^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$
- Biais : $Biais = \frac{\sum d_i}{n}$
- Erreur Standard de Prédiction : $SEP = \sqrt{\frac{\sum(d_i - biais)^2}{n-1}}$
- Root Mean Square Error of Prediction : $RMSEP = \sqrt{\frac{\sum {d_i}^2}{n}}$

- Relative Predictive Determinant : $RPD = \frac{écart - type_{ref}}{SEP}$

Notons y_i les n valeurs mesurées et $\hat{y}_i$ les n valeurs prédites et $\sum d_i = \sum_{i=1}^{n} (y_i - \hat{y}_i)$ et $\sum x = \sum_{i=1}^{n} x$.

Pendant le développement d'équations de prédiction, les échantillons aberrants (ou « outliers ») sont retirés selon l'Équation 8 :

Équation 8

$$Outlier = Y - (\bar{Y} \pm 2*ET)$$

Avec Y qui correspond à la différence entre la valeur prédite et la valeur de référence du paramètre à prédire, ET correspondant à l'écart-type

I.5. Asymmetrical Flow Field Flow Fractionation

I.5.1. Analyse de distribution molaire de la gélatine

Des méthodes existent pour caractériser la distribution molaire de polymères. Dans certains cas, les masses peuvent être déterminées par l'étude qualitative par électrophorèse sur gel de polyacrylamide (PAGE) (Van den Bosch et Gielens, 2003). Dans le cas de la gélatine, la distribution de masses molaires est principalement étudiée par chromatographie d'exclusion stérique (SEC)(Dupont, 2002; Tromp *et al.*, 2002). La chromatographie d'exclusion stérique présente certaines limites. Elle nécessite un étalonnage de la colonne et ne permet pas d'analyser les molécules qui possèdent une masse molaire supérieure à $7{,}5\times10^7$ g.mol^{-1}(Meyer et Morgenstern, 2003). Or, la gélatine est constituée pour partie d'agrégats de haut poids molaire (entre 3×10^4 et 1×10^9 g.mol^{-1}) (Rbii et al., 2009). La SEC ne permet pas de couvrir l'ensemble de la gamme de masses molaires en une seule analyse. Ces agrégats jouent un rôle important sur les propriétés physico-chimiques(Van Krevelen et Te Nijenhuis, 2009). Récemment une nouvelle méthode de fractionnement, appelée *Asymetrical Flow Field-Flow Fractionation* (AsFlFFF) s'est développée. Elle permet l'analyse de molécules entre 1kDa et quelques micromètres et est appliquée dans de nombreux domaines (Giddings *et al.*, 1977).

Par exemple, elle est utilisée dans l'étude de nanoparticules, de micelles, d'émulsions, de polymères ou encore de protéines complexes (Schimpf *et al.*, 2000; Till *et al.*, 2014; Gibot *et al.*, 2014; Nilsson, 2013; Liu *et al.*, 2006). De plus, en couplant l'AsFlFFF avec un détecteur MALS (Multi Angle Light Scattering), il est possible de déterminer la masse molaire et le rayon de giration de ces fragments, paramètres d'intérêts dans l'analyse de la distribution molaire. Dans le cas de l'analyse de gélatines, les travaux réalisés par *Fraunhofer et al*. (Fraunhofer *et al.*, 2004) et plus récemment par *Rbii et al.* (Rbii, 2010; Rbii et al., 2009), ont permis de montrer la capacité de l'AsFlFFF à fractionner la totalité des gélatines présentant l'ensemble des masses molaires permettant ainsi l'étude d'agrégats de haute masse molaire.

I.5.2. Les méthodes de fractionnement

Les méthodes dites de Fractionnement par Flux Force (FFF) sont des techniques récentes, utilisées principalement pour la caractérisation de macromolécules. Ces méthodes ont été décrites pour la première fois en 1960 par Giddings (Giddings, 1966). Les techniques de fractionnement par flux-force sont basées sur l'action simultanée de l'écoulement d'un liquide dans un canal de faible épaisseur et l'effet d'un champ externe appliqué perpendiculairement au canal. L'écoulement dans le canal est un écoulement laminaire, cela signifie que la vitesse est maximale au centre du canal et pratiquement nulle au niveau des parois. Le champ externe appliqué perpendiculairement au canal, va quant à lui permettre un positionnement des analytes à des hauteurs différentes au niveau du canal en fonction de leurs propriétés physico-chimiques (taille, densité, charge, diffusion thermique...) et de la nature du champ appliqué. Les analytes sont ensuite entrainés par le flux laminaire à des vitesses différentes en fonction de leurs positionnements dans le canal. Cette analyse va permettre de séparer les analytes qui vont ensuite pouvoir être analysés par différents types de détecteurs (ultraviolet (UV), indice de réfraction, diffusion statique (SLS), diffusion dynamique (DLS), spectrométrie de masse (MS)…), pour obtenir un fractogramme.

Différentes techniques de fractionnement ont été décrites, qui diffèrent selon la nature du champ externe appliqué ont été décrites par *Giddings* (Giddings, 1995) :

- **Sedimentation FFF** : Le Fractionnement par couplage avec un Flux de force de Sédimentation (SdFFF) est une technique basée sur la centrifugation comme champ externe. Cette méthode permet de séparer les particules en fonction de leur taille et de leur densité(Contado *et al.*, 2013).
- **Electrical FFF** : Le fractionnement par champ électrique est une méthode qui utilise un champ électrique perpendiculaire au flux. Cette méthode permet de fractionner les molécules en fonction de leur taille, mais aussi de leur nature électrostatique (Caldwell *et al.*, 1972).
- **Thermal FFF** : Cette méthode est basée sur l'application d'un gradient de température sur la hauteur du canal. Elle est principalement utilisée pour l'analyse de polymères synthétiques dans des solvants organiques (Liu et Giddings, 1992).
- **Flow FFF** : Cette méthode est la plus célèbre, comme les autres méthodes elle consiste en l'application d'un flux croisé perpendiculaire de même nature que le solvant utilisé. Cette technique présente l'avantage d'être très adaptable, notamment grâce à la modification de la puissance du flux croisé et par l'influence de la taille de la cellule (Till et al., 2014; Messaud *et al.*, 2009).

I.5.3. Le cas de Flow FFF

Les méthodes de Flow FFF se divisent en deux techniques principales ; la FlFFF symétrique (SyFlFFF) et FlFFF asymétrique (AsFlFFF). La SyFlFFF est historiquement la première technique qui a été étudiée par *Gidding et al.* en 1976 (Giddings *et al.*, 1976). La méthode a ensuite évolué avec le temps pour donner naissance à la AsFlFFF en 1987 (Messaud et al., 2009). Ces méthodes fonctionnent selon le même principe et se régulent selon plusieurs paramètres ; par le flux liquide linéaire horizontal (Vx), le flux d'injection (Vinj), le flux croisé (Vc) et l'épaisseur de la cellule. L'ensemble de ces paramètres va permettre de séparer et d'analyser des molécules de différentes natures (polymères, colloïdes, protéines, nano-objets …) et de tailles différentes (entre 1 nm et 50 µm) (Ratanathanawongs et Giddings, 1992; Benincasa et Giddings, 1992).

La Figure 13 représente les différents constituants d'une cellule AsFlFFF. Pour l'AsFlFFF, le bloc supérieur est imperméable, tandis que l'inférieur est perméable et muni d'un fritté poreux.

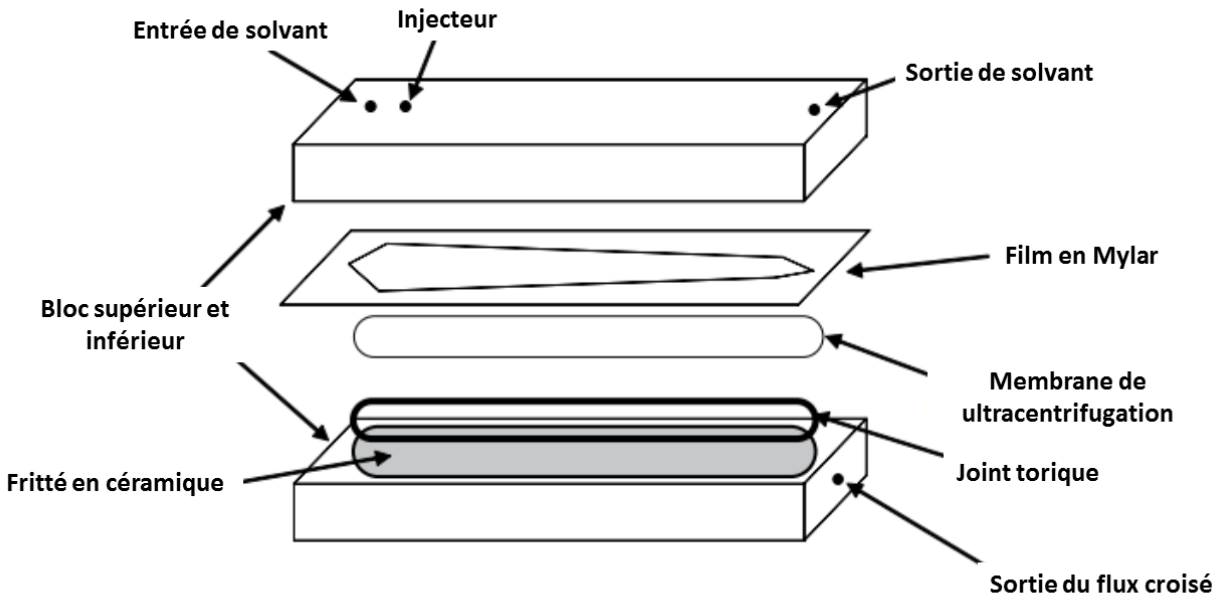


Figure 13 : Représentation schématique de la cellule AsFlFFF, d'après Rbii et al. (Rbii et al., 2009)

Dans le cas de SyFlFFF, les blocs supérieurs et inférieurs sont tous les deux perméables et muni d'un fritté poreux. Une membrane d'ultrafiltration recouvre le bloc du fond pour empêcher l'échantillon de pénétrer dans le canal. Un espaceur (film en Mylar) de 125 à 490 µm d'épaisseur et de forme trapézoïdale est déposé sur la membrane. En mode normal, une fois que le champ de force perpendiculaire est appliqué sur le flux laminaire en écoulement, les analytes sont entraînés vers la couche limite, aussi appelée "mur d'accumulation" du canal (la membrane). Les plus petites particules qui ont des taux de diffusion plus élevés atteignent une position d'équilibre plus haute dans le canal. Les plus grosses molécules atteignent une position d'équilibre plus proche de la membrane. Pendant la phase d'élution, le flux laminaire, dû à la finesse de la membrane va créer un gradient à l'intérieur du canal permettant ainsi de séparer les particules selon leur taille.

Les plus petites molécules, qui possèdent un coefficient de diffusion plus élevé, vont se situer au centre du canal et seront transportées plus rapidement vers la sortie du canal que les grosses, qui possèdent un coefficient de diffusion plus faibles. Ce phénomène est présenté sur la Figure 14 ci-dessous :

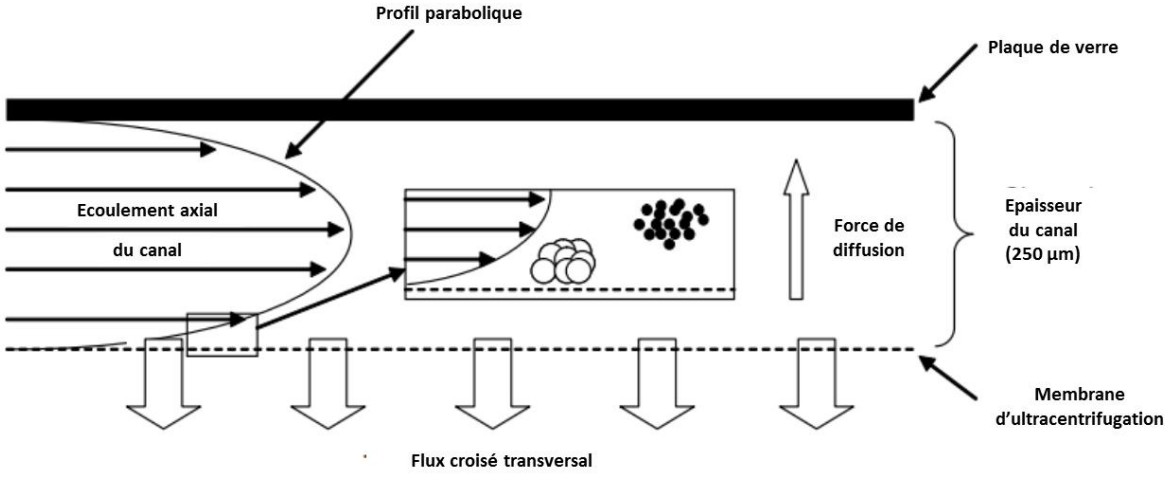


Figure 14 : Schéma du principe de séparation de particules de différentes tailles en AsFlFFF. Les particules de petite taille sont représentées par de petits cercles noir (●), et les particules de grande taille sont représentées par de cercle blanc (○). D'après les travaux de Rbii et al. (Rbii et al., 2009)

L'analyte est donc soumis au flux croisé qui va induire un mouvement Brownien qui est lui-même déterminé en fonction du coefficient de diffusion des molécules dans l'éluant. Le temps de rétention du composé est déterminé en fonction du coefficient de diffusion et est défini par l'Équation 9 :

Équation 9

$$t_r = \frac{t^0 * Vc * w^2}{6 * V_0 * D}$$

Avec t_r pour le temps de rétention de l'analyte, t^0 pour le temps mort (temps nécessaire à l'échantillon pour traverser la cellule avant de la détecter), V^0 pour le volume de la cellule, w pour l'épaisseur de la cellule, D pour le coefficient de diffusion et Vc pour le débit du flux croisé.

D'après l'équation de Stokes-Einstein, le coefficient de diffusion peut être défini en fonction du rayon hydrodynamique de l'analyte d'après l'Équation 10 (Edward, 1970) :

Équation 10

$$D = \frac{k * T}{6 * \pi * \eta * r_h}$$

Avec D le coefficient de diffusion de l'analyte, k la constante de Boltzmann,T la température en Kelvin, η la viscosité de l'éluant et r_h le rayon hydrodynamique de la molécule.

I.5.4. Les détecteurs Multi Angle Light Scattering (MALS) et Ultraviolet (UV) pour l'analyse des protéines

Le détecteur de diffusion de lumière MALS est une technique basée sur la réception de la lumière, diffusée par des particules en suspension sous différents angles. Ces données sont utilisées pour déterminer la masse molaire absolue et la taille moyenne des particules en fonction du temps d'élution. Les molécules sont soumises à un rayonnement électromagnétique, qui lui permettent de se polariser et de diffuser la lumière. Selon les principes de diffusion de la lumière de *Wyatt* et *Zimm* (Zimm, 1948; Wyatt, 1993), une molécule d'un diamètre inférieur à 1/20ème de la longueur d'onde du faisceau lumineux monochromatique, diffuse de manière isotrope. La diffusion est anisotrope pour les molécules plus volumineuses. L'intensité de cette diffusion dépend de plusieurs facteurs, mis en relation par l'équation de *Zimm* et à partir de laquelle il est possible d'en déduire la masse molaire moyenne en masse (Zimm, 1948; Rbii et al., 2009).

Le détecteur UV permet le suivi des mesures quantitatives, et informe sur le nombre de particules présentes tout au long de l'élution. Il permet de déterminer la concentration d'un composé chimique par la loi de *Beer-Lambert-Bouguer.* D'après les travaux de *Tromp et al.* (Tromp, ten Grotenhuis, et Olieman, 2002)et *Rbii et al.* (Rbii et al., 2009), la gélatine possède un incrément d'indice de réfraction *(dn/dc*) de 0,164 mL.g^{-1} et un coefficient d'extinction molaire de l'UV (ε) de 12 700 mL.g^{-1}.cm^{-1}.

Ainsi, l'AsFlFFF, couplé aux détecteurs UV et MALS, permet d'obtenir les valeurs de masse moléculaire moyenne en masse (Mw) et en nombre (Mn), le rayon de giration (Rg) ainsi que l'indice de polydispersité (Mw/Mn). Ils sont calculés à partir des équations suivantes :

Équation 11

$$Mn = \frac{\sum Ci}{\sum \frac{Ci}{Mi}}$$

Équation 12

$$Mw = \frac{\sum Ci \times Mi}{\sum Ci}$$

Équation 13

$$Rg = \sqrt{\frac{\sum niMi < r^2 >}{\sum niMi}}$$

Avec Mi qui représente la masse moléculaire sur la tranche i et de concentration Ci.

I.6. Conclusion

La gélatine, utilisée comme agent de texture dans l'industrie agroalimentaire, est produite à partir de collagène d'origine animale. Son rôle fonctionnel varie en fonction des demandes des différents marchés. Sa production nécessite la mise en place de procédés d'extraction adaptés. Il en résulte donc une large gamme de types de gélatine, aux propriétés physico-chimiques différentes, en fonction des conditions d'extraction et de la nature des matières premières. La caractérisation des matières premières en temps réel, à l'origine de la gélatine, est donc primordiale pour optimiser au mieux sa production et rester concurrentielle sur le marché.

Le développement de la spectroscopie proche infrarouge associée à l'analyse multivariée s'est imposée comme méthode efficace pour répondre aux besoins analytiques de caractérisation des matières premières. En plus d'être une méthode non destructive, non invasive, elle est applicable en mode dynamique, c'est-à-dire sur un flux de matières.

Cependant, l'application de ces outils s'est principalement centrée sur des matières premières homogènes, telles que le tissu musculaire. L'application de telles méthodes sur de la peau nécessite des études complémentaires adaptées aux problématiques industrielles. Elles sont notamment liées à la double hétérogénéité de la matrice elle-même liée aux matières premières (couennes de porc) et à l'acquisition dynamique des données.

Les propriétés physico-chimiques des gélatines, la force en gel et la viscosité, en plus d'être directement liées aux conditions d'extraction et à la nature des matières premières, sont conditionnées par la structure moléculaire de la gélatine. *L'Asymmetrical Flow Field-Flow Fractionation,* couplée à un détecteur UV et MALS, est une nouvelle méthode de fractionnement adaptée à l'étude de la distribution molaire de polymères d'origine protéique. Cette méthode pourrait donc être appliquée sur des échantillons de gélatine. Par ailleurs, la détermination des propriétés physico-chimiques de ces gélatines par la spectroscopie proche infrarouge pourrait permettre une quantification accélérée de ces propriétés.

En conclusion de l'ensemble des connaissances bibliographiques présentées dans ce chapitre, plusieurs interrogations scientifiques et technologiques se sont révélées :

- L'application de la SPIR en ligne, sur une matrice présentant une hétérogénéité intra et inter échantillons, peut-elle permettre de déterminer les propriétés technologiques moyennes d'un échantillon ?
- L'échantillonnage à grande échelle permet-il de s'affranchir de l'hétérogénéité intra et inter échantillons, afin d'obtenir une prédiction robuste d'un lot sur une chaîne de production par SPIR ?
- Les propriétés physico-chimiques des gélatines sont-elles directement liées à la distribution moléculaire et peuvent-elles être prédites par spectroscopie proche infrarouge ?
- Une approche chimiométrique d'un signal quantitatif représentant la distribution moléculaire de gélatine peut-elle permettre la prédiction de ses propriétés physico-chimiques ?

Afin de répondre à ces questions scientifiques couplées aux problématiques industrielles, le travail de thèse s'articulera autour de trois parties :

1- Etude de la caractérisation par spectroscopie proche infrarouge de couennes de porc en mode dynamique en prenant en compte l'hétérogénéité des matières premières.
2- Détermination du potentiel d'extraction de gélatines de couennes de porc par spectroscopie proche infrarouge sur un site de production industriel.
3- Etude des propriétés physico-chimiques de gélatines par spectroscopie proche infrarouge et par *Asymmetrical Flow Field-Flow Fractionation* couplée à un détecteur UV et *Mutli Angle Light Scattering*

Chapitre II. Etude de la caractérisation par spectroscopie proche infrarouge en mode dynamique de couennes de porc

II.1. Introduction

L'industrie agroalimentaire a pour but de transformer des matières premières issues de l'agriculture ou de la pêche, en produits de consommation alimentaire en respectant certains critères de qualité spécifiques à la production (Trienekens et Zuurbier, 2008). En effet, les consommateurs dans les pays industrialisés sont en attente de produits alimentaires ayant une qualité constante et élevée, et ce, à des prix les plus compétitifs possibles. De fait, on comprend que la maîtrise de la variabilité de la matière première soit une préoccupation majeure. Pour cela, les entreprises s'appuient bien souvent sur des standards de matières premières (cahier des charges) afin de s'affranchir de la variabilité dans leurs productions. Il demeure toutefois nécessaire de caractériser les matières premières avant transformation afin d'adapter le processus de transformation.

Dans le cadre de l'industrie de la gélatine, les matières premières sont essentiellement des couennes de porc, issues de l'industrie porcine. Les peaux de porcs utilisées dans l'industrie de la gélatine font l'objet d'importantes variations, qui auront des conséquences dans le processus d'extraction de la gélatine. Ces variations ont pour origine de nombreux facteurs, plus ou moins facilement identifiables par l'industriel : le sexe et l'âge de l'animal, le type d'alimentation et d'élevage (conventionnel ou biologique), la qualité de la découpe des peaux au niveau de l'abattoir, l'origine géographique, l'objectif de l'élevage (production ou reproduction) et des conditions de stockage pendant le transport des matières premières. Ce sont autant de paramètres qui peuvent impacter le rendement ou la qualité de la gélatine extraite.

Généralement, les professionnels de la gélatine estiment la qualité des lots de couennes par une analyse visuelle subjective réalisée par des opérateurs au moment de la réception. La qualité de ces lots est très variable et les rendements peuvent varier de 10% à 23%, pour le même processus de transformation. Ces variations sont directement liées à la qualité de la matière première en début de processus.

En ce qui concerne la caractérisation du produit fini, elle est réalisée en laboratoire par des dosages physico-chimiques. Les méthodes d'analyses en laboratoire sont précises, mais engendrent généralement un coût et un temps d'analyse importants. Il y a, en plus, la question de l'échantillonnage, problématique que l'on retrouve dans de nombreuses disciplines (Qin, 2017).

L'objectif de ce chapitre est d'évaluer la capacité d'une méthode de mesure pouvant se substituer aux analyses de laboratoire, *i.e.* la spectroscopie proche infrarouge, afin d'intégrer une chaîne de production *on-line*. Pour cela, deux hypothèses ont été formulées :

Hypothèse 1 : Une collecte spectrale en ligne dans le Proche infrarouge (PIR), permet de percevoir la variabilité de la composition d'échantillons de couenne de qualité très variable : hétérogénéité intrinsèque (face graisse et face cuir pour chaque couenne) et hétérogénéité entre les couennes.

Hypothèse 2 : Une collecte spectrale en ligne dans le Proche infrarouge sur une chaîne de production, permet d'appréhender le rendement potentiel de 75 tonnes de couenne hétérogènes.

II.2. Matériels et méthodes

II.2.1. Caractérisation physico-chimique en ligne par SPIR d'une matrice hétérogène : la couenne de porc

II.2.1.(1). Echantillonnage des couennes

Cinquante-sept échantillons ont été collectés directement sur l'usine de production de gélatine de Weishardt. La sélection des échantillons a été réalisée de façon à représenter l'approvisionnement en matière première de l'entreprise. Elle est basée sur 5 critères : le pays d'origine (EU), le type de production (biologique (n=6) ou conventionnelle (n=51)), l'objectif d'élevage (porcs charcutiers (n=51) ou coches de reproduction (n=6)), la partie de l'animal (couenne (n=50) ou bardière (n=7)) et le conditionnement (congelé (n=7) ou frais (n=50)) (Figure 20). La base de données complète est présentée en Annexe 1. Les échantillons sont collectés sur 15 jours du fait de la contrainte de l'arrivage de fournisseurs pendant le mois de novembre 2015.

Les couennes prélevées possèdent des masses comprises en 260 g et 560 g. Elles sont placées au congélateur, stockées à -20°C et conservées à cette température jusqu'à la veille de la phase d'acquisition spectrale.

II.2.1.(2). Acquisition de spectres proche infrarouge

La veille de l'acquisition spectrale, les couennes sont mises à décongeler dans une pièce tempérée à 21°C. La température des couennes avant acquisition est comprise entre 15 et 20 °C. Une couenne est par constitution très hétérogène, les différentes faces ont été nommées face cuir et face graisse (Figure 15).



Figure 15 : Couenne face cuir

Les spectres des couennes ont été acquis sur le convoyeur (Gepperd-Band, Figure 17), à une vitesse de 0,365 $m.s^{-1}$, avec un spectromètre Polytec (PSS-2121, Polytec GmbH, 763337 Waldbronn Germany), équipé d'une fibre optique de 2 m et d'une tête de mesure à distance (PSS-H-A03, Polytec GmbH, 763337 Waldbronn Germany). L'acquisition des spectres en réflectance de couennes a été réalisée sur les faces cuir et graisse à une hauteur de 34 cm sur la gamme spectrale de 1100 à 2100 nm avec un pas de 2 nm, un temps d'intégration de 30 ms et sur une surface de 3,5 cm de diamètre, soit un spectre tous les 1,1 cm (Figure 17). Sur chaque face, 75 à 200 spectres ont été collectés en fonction de la taille de la couenne qui varie entre 25 et 70 cm de longueur et entre 15 et 50 cm de largeur.



Figure 16 Tapis de convoyage pour l'acquisition spectrale en laboratoire.

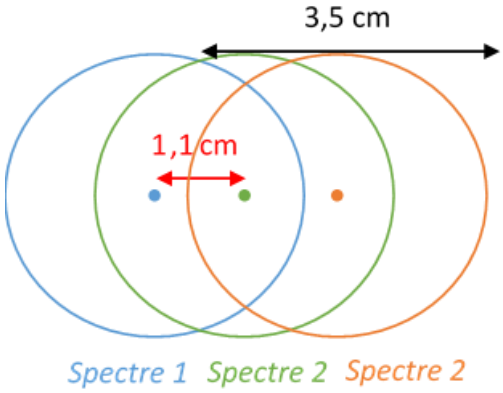


Figure 17 : Espacement de trois mesures de couenne en flux.

II.2.1.(3). Analyses de référence

(a) Préparation des échantillons et détermination du taux de matière sèche

Chaque couenne est intégralement découpée en dés de 1 cm² et pesée puis placée au congélateur à -20°C.

Avant broyage les couennes congelées sont lyophilisées dans un lyophilisateur semi-industriel pendant 72h et pesées à la fin de la lyophilisation.

Le pourcentage de matière sèche est déterminé selon la formule suivante :

Équation 14

$$Taux\ de\ MS = \frac{m\ après\ lyophilisation * 100}{m\ avant\ lyophilisation}$$

avec MS Matière Sèche en % et m masse g

Les dés de couennes lyophilisés sont ensuite broyés au broyeur-mixeur à couteaux GRINDOMIX GM 200 (Retsch©GmbH). A cause des contraintes dues à la résistance physique des couennes, l'opération est réalisée dans un bol en inox et en présence de glace carbonique afin de faciliter le broyage et d'empêcher la dégradation thermique due aux forces de frottement.

À la suite de ces étapes, les couennes sont sous formes de poudre ou de pate plus ou moins homogènes (Figure 18).



Figure 18 : Couennes après les différentes étapes d'homogénéisation.

(b) Dosage des protéines par la méthode de Kjeldahl

(b) i. Principe

Le dosage des protéines par la méthode de Kjeldahl est réglementé par la norme NF EN ISO 5983-1 (ISO, 2005). Le dosage par la méthode de Kjeldahl est composé de deux étapes ; la minéralisation de l'azote organique et le dosage acidobasique de l'ion ammonium du minéralisat.

- Minéralisation

Pendant l'étape de minéralisation, les atomes de carbone, d'azote, d'oxygène et d'hydrogène des matières organiques vont se retrouver sous la forme CO_2, H_20 et NH_3.

L'échantillon est minéralisé par l'acide sulfurique à 400°C, en présence d'un catalyseur afin de limiter l'évaporation de l'acide.

Équation 15

$$\textit{Matière Organique} + H_2SO_4 \rightarrow CO_2 + NH_3 + H_2O$$

$$2NH_3 + H_2SO_4 \rightarrow 2NH_4^+ + SO_4^{2-}$$

- Dosage acidobasique

Les ions ammonium du minéralisat sont alcalinisés afin de les transformer en ammoniac. Pour cette étape, de la soude est ajoutée en excès pour que la réaction soit totale.

Équation 16

$$(2NH_4^+ + SO_4^{2-}) + 2(Na^+ + OH^-) \rightarrow 2NH_3 + (2Na+ + SO_4^{2-}) + 2H_2O$$

L'ammoniac est piégé par un acide faible en excès, l'acide borique (H_3BO_3). L'ammoniac est ensuite neutralisé par une solution d'acide chlorhydrique titrée, en présence d'un indicateur de pH coloré (mélange de rouge de méthyle et de vert de bromocrésol). L'ammoniac dans la solution d'acide borique va faire virer le milieu au vert émeraude, l'acide chlorhydrique va par la suite faire virer le milieu au gris foncé.

Ammoniac dans l'acide borique :

Équation 17

$$NH_3 + H_3BO_3 \rightarrow NH_4^+ + H_2BO^{3-}$$

Dosage en retour :

Équation 18

$$NH_3 + H_3O^+ \rightarrow NH_4^+ + H_2O$$

(b) ii. Dosage des protéines de couennes de porc

- Minéralisation de l'azote organique

La minéralisation est réalisée sur le bloc de minéralisation Tecator Kjeltec 2020. La veille du dosage, les échantillons sont préparés en pesant 0,7g de broyat de couenne lyophilisée. Chaque échantillon a été réalisé en triplicata. Les témoins contiennent une solution d'urée à 10%. Puis 12,5mL d'acide sulfurique concentré (96%) sont ajoutés dans les matras qui sont ensuite laissés pendant environ 12h à température ambiante.

Avant la minéralisation, 2 pastilles de catalyseur (Kjeltabs^{TM}, NH_4^+, $CuSO_4$) sont ajoutées dans chaque matras, qui sont ensuite placés dans un bloc de minéralisation préalablement chauffé à 400°C. Les vapeurs d'acide sont aspirées pendant la minéralisation dont la durée varie de 1h-1h15, avant un refroidissement de 20 minutes.

- Titration de l'ammoniac

La titration de l'ammoniac est réalisée grâce au Tecator-Foss Kjeltec^{TM}84000. La séquence est préparée sur le logiciel Compass^{TM}, puis envoyée et prise en charge par l'automate.

- Expression littérale du pourcentage des protéines

Pourcentage d'azote :

Équation 19

$$n_{azote} = n_{HCl} = (V_{eq}-V_{blanc})*Titre_{HCl}$$

$$m_{couenne} = m_{couenne\ lyoph} *\%MS$$

$$\%\ azote = n_{azote}*M_{azote}/m_{couenne}$$

avec n quantité en mol, V=volume en litre (V_{eq} =volume équivalent et V_{blanc}= Volume blanc), m masse g, M en $g.mol^{-1}$, MS la matière sèche.

Pourcentage de protéine :

Pour calculer le pourcentage de protéine à partir du pourcentage d'azote, un facteur de conversion empirique (α) est utilisé. Ce facteur va dépendre de la matrice analysée et est relatif à la composition en acides aminés des protéines analysées.

Par défaut, le taux d'azote des protéines est estimé à 16%, ce qui donne un facteur de 6,25. Pour la gélatine pure, un facteur empirique de 5,35 est utilisé (Sosulski et Imafidon, 1990).

Équation 20

$$\% \text{ Protéine} = \% \text{ azote} \times \alpha$$

(c) Détermination du taux de matières grasses par Extraction Accélérée par Solvant

La détermination du taux de graisse dans les matières animales est référencée par la norme NF V04-403 (avril 2001)(AFNOR, 2001). Les matières grasses sont extraites à l'éther de pétrole pendant 4h sur un appareil à Soxhlet. Afin de réduire au maximum le temps de l'analyse du taux de graisse, c'est la méthode *Accelerate Solvant Extraction* (ASE) qui a été utilisée. La méthode ASE est basée sur le même principe qu'un appareil d'extraction classique, avec une efficacité de l'extraction augmentée, car le passage du solvant à travers l'échantillon peut être réalisé à des températures variables et des pressions élevées. Cette méthode permet de réduire significativement la consommation de solvant et les temps d'extraction. Les extractions ont été réalisées sur l'ASE (Thermo Scientific™ Dionex™ ASE™ 350 Accelerated Solvent Extractor system) qui permet l'analyse d'une série de 24 cellules.

- Mode opératoire

Les échantillons ont été réalisés en triplicata. Pour l'analyse des broyats de couennes lyophilisées, des cellules d'un volume de 22mL ont été utilisées. Les cellules sont ensuite chargées avec 3 grammes d'échantillon, pesés avec précision (±0,001) et mélangés avec de la Célite 545 (terre diatomée) de façon que le volume de la cellule soit complété. Les conditions opératoires définies pour cette extraction sont : un préchauffage à 120°C de la cellule pendant 3 min avant l'injection sous une pression de 10 bar d'un mélange de solvant hexane-isopropanol (3 :2), le mélange de solvant est injecté à 5 reprises à hauteur de 50% du volume de la cellule, pour finir une purge de 60 secondes à l'azote est réalisée.

Pour chaque cellule, un tube de récupération est pesé à vide, puis placé sur le rack de récupération afin de récolter le mélange de solvant et de matière grasse de la cellule correspondante (Figure 19).

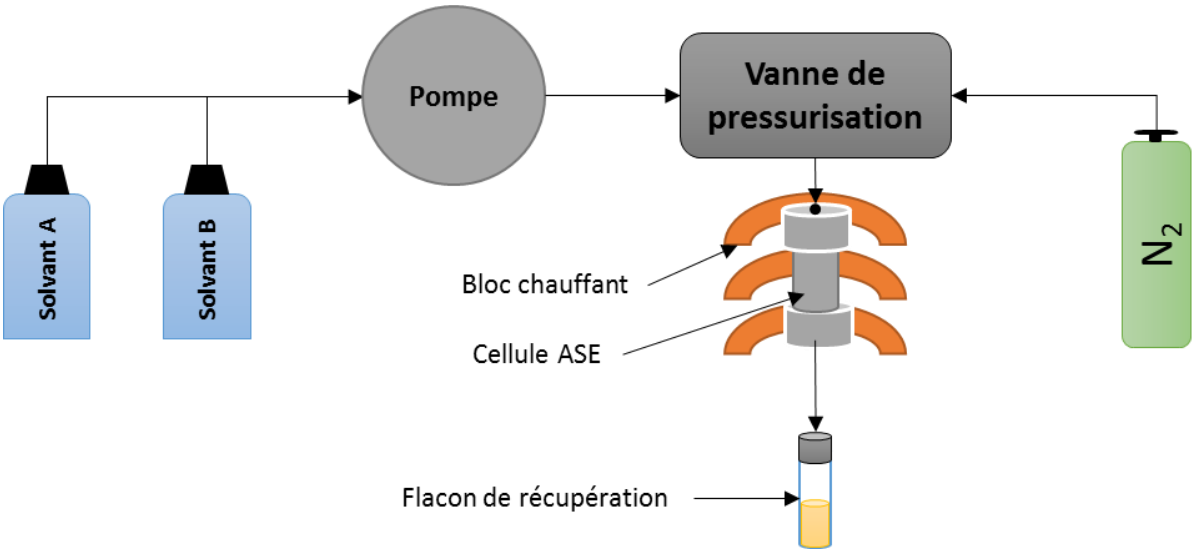


Figure 19 : Principe de l'ASE : les solvants passent par la pompe puis sont pressurisés avant de traverser la cellule ASE, entrainant les matières grasses pour être récupéré dans un flacon de récupération.

Les tubes sont ensuite mis à évaporation (Genevac® EZ-2plus) à 40°C, en utilisant le programme « Low BP mixture ». A la fin de l'évaporation, les tubes avec les matières grasses sont pesés, afin de calculer le taux matière grasse dans la matière première.

Équation 21

$$\% \text{ de graisse de la couenne} = (m_{\text{mat grasse extraite}} * \%MS * 100) / m_{\text{mat initiale du lyophilisat}}$$

avec MS=Matière Sèche en % et m masse g

(d) Détermination du taux de collagène par dosage de l'hydroxyproline

Le protocole du dosage de l'hydroxyproline dans les produits à base de viande et produits de la pêche est décrit par la norme NF V04-415 de 2002 (AFNOR, 2002).

(d) i. Principe

L'hydroxyproline est un acide aminé spécifique de la protéine de collagène, c'est donc le dosage de cet acide aminé qui va permettre de déterminer indirectement le taux de collagène.

Dans un premier temps, l'hydroxyproline est libérée par une hydrolyse acide. L'hydroxyproline est ensuite oxydée par la Chloramine T et est dosée par colorimétrie à 558nm grâce à la présence du *p*-diméthylaminobenzaldéhyde.

(d) ii. Mode opératoire

- Hydrolyse acide

La réaction d'hydrolyse acide est réalisée dans des matras, à l'aide d'un bloc de minéralisation (Kjeldatherm®, Gerhardt) de 20 places. Pour chaque série, 18 matras sont utilisés pour les échantillons à analyser et les 2 matras restant servent de standard. Pour chaque échantillon, des triplicatas sont réalisés, permettant l'analyse de 6 échantillons par série.

Pour l'analyse de broyat de couenne de porc, 500 mg de poudre sont pesés et ajoutés à 10 mL d'acide chlorhydrique 6N. Pour les standards, 100 mg de gélatine Bloom 285 (NF V 59-001)(AFNOR, 1982a) sont pesés. Les matras sont ensuite recouverts avec des entonnoirs et placés sur le bloc de minéralisation à 105°C pendant 24h.

Après hydrolyse, les échantillons sont placés à température ambiante, puis l'intégralité de l'hydrolysat est transférée dans une fiole jaugée de 100 mL en filtrant à l'aide de papier filtre standard plissé 100mm (Fisherbrand, Fisher Scientific). Le volume des fioles jaugées est complété avec de l'eau distillée. La solution filtrée est ensuite diluée au 1/100, en prélevant 0,5 mL versés ensuite dans une fiole jaugée de 50 mL. La dilution finale est alors de 10000.

- Réaction d'oxydo-réduction et dosage spectrophotométrique de l'absorbance

Une gamme étalon est réalisée à partir de solutions d'hydroxyproline de 1 à 4 $\mu g.ml^{-1}$ dans la zone de linéarité. Une solution tampon pH=6, de *p-DiMethylAminoBenzaldehyde* (*p*DMAB) et de chloramine T à 1,4% est préparée selon le protocole présenté en Annexe 2.

- Dosage spectrophotométrique de l'absorbance

Quatre millilitres des solutions à analyser sont ajoutés dans une cuve de spectroscopie. Pour chaque dosage, 3 cuves sont préparées pour chaque solution standard, 2 cuves sont préparées pour chaque échantillon. Deux millilitres de solution de chloramine T sont ajoutées dans chaque cuve. Les cuves sont ensuite placées à température ambiante pendant 20 min à l'obscurité avant d'ajouter 2mL de solution de *p*DMAB, puis placées à 60°C dans un bain marie pendant 15 min.

A la fin de la réaction, les cuves sont placées à température ambiante pendant 60 min. L'absorbance est enfin mesurée dans des cuves de 10 mm de largeur à 558 nm (heλios γ, Thermo Scientific).

(d) iii. Expression des résultats

Les concentrations en µg.mL^{-1} des échantillons à doser sont déterminées à l'aide de l'équation de la courbe d'étalonnage. Le pourcentage d'hydroxyproline est alors déterminé par l'équation suivante.

Équation 22

$$\%d'hydroxyproline = \frac{[C] * d * \%MS}{m\ échantillon * 10}$$

% d'hydroxyproline en µg.g^{-1} de matière fraîche
Soit m masse en mg d'échantillon
Soit [C] concentration en µg.mL^{-1}
Soit d le facteur de dilution, ici de 10000
MS pour matières sèches

Un facteur de conversion est calculé pour déterminer le pourcentage de collagène à partir du pourcentage d'hydroxyproline présent dans l'échantillon : le pourcentage d'hydroxyproline est multiplié par un coefficient, dépendant de la nature du collagène. Conventionnellement un coefficient de 7,5 est utilisé (Sosulski et Imafidon, 1990).

Équation 23

$$\% \text{ de collagène} = \% \text{ hydroxyproline} * 7{,}5$$

II.2.2. Analyse des bases de données

(a) Constitution des bases de données

Deux bases de données ont été constituées. Pour la face cuir, la base de données comprend 57 échantillons, pour la face graisse, la base de données comprend 56 échantillons. Une couenne n'a pu être utilisée que sur la face cuir. Une troisième base de données a été constituée comprenant les spectres faces graisse et cuir.

(b) Préparation des spectres de couennes

(b) i. Tri des spectres

Pour chaque échantillon, les spectres ont été acquis en ligne sur le convoyeur générant des spectres de couennes, des spectres de tapis de convoyage et des spectres à l'interface du tapis et de la couenne. Il est donc nécessaire d'isoler les spectres de couenne pour avoir des bases de données nettoyées. Le tri des spectres a été réalisé à l'aide du logiciel The Unscrambler® X v10.3.0.89 (CAMO ASA, Trondheim, Norway). Les spectres bruts de la face cuir et de la face graisse pour chaque couenne ont été importés sur The Unscrambler®. Soixante-quinze à 200 spectres ont donc été étudiés successivement pour chacune des 56 couennes vues face graisse et les 57 vues face cuir. Les analyses en composantes principales (ACP) avec l'algorithme NIPALS ont été réalisées pour chaque couenne et pour chaque face, permettant d'obtenir au moins deux groupes distincts, correspondant aux spectres de couennes et aux spectres du tapis de convoyage avec parfois un troisième groupe correspondant aux spectres « à cheval ». Une sélection visuelle des spectres de couenne a été réalisée, permettant d'isoler chacun des groupes. Quand cela était nécessaire, une nouvelle ACP a été réalisée sur les spectres de couennes afin d'isoler les observations aberrantes. Après le nettoyage des bases de données, une face de couenne possède en moyenne 37 spectres pour une seule valeur de matière sèche, de protéine, de matière grasse et de collagène.

(b) ii. Prétraitements mathématiques des spectres

Les spectres des couennes ont permis d'obtenir 6 bases de données différentes :

- Face cuir des 57 couennes avec l'ensemble des spectres en ligne : 2146 spectres
- Face graisse des 56 couennes avec l'ensemble des spectres en ligne : 2108 spectres
- Face cuir et graisse des 57 couennes avec l'ensemble des spectres en ligne : 4254 spectres
- Face cuir des 57 couennes moyennées : 57 spectres
- Face graisse des 56 couennes moyennées : 56 spectres
- Face cuir et graisse 57 couennes moyennées : 113 spectres

Pour les analyses de référence (taux de protéine, matières grasse et taux de matière sèche) en complément des valeurs quantitatives, des classes ont été créées, correspondant aux quartiles. Les traitements statistiques ont été réalisés sur XLSTAT Version 2016.04.32980.

Un lissage en 7 points a été appliqué sur tous les spectres infrarouges (Savitzky et Golay). Et un premier prétraitement a ensuite été réalisé parmi les quatre suivants :

- Sans prétraitement
- Passage en absorbance
- Mean centred
- Autoscale

Avant l'application d'un dernier prétraitement :

- Dérivée première
- Dérivée seconde
- Standard Normal Variate (SNV)
- Standard Normal Variate Detrend (SNVD)
- Multiplicative Scatter-Correction (MSC)

(c) Traitement des données

(c) i. L'Analyse Discriminante Linéaire (LDA)

L'analyse discriminante est souvent utilisée en spectroscopie infrarouge (Sun, 2009; Visa *et al.*, 2011; Sankaran *et al.*, 2012). C'est une méthode multivariée supervisée. Les composantes principales de l'ACP serviront d'entrée à ce modèle. Durant l'analyse, des algorithmes sont utilisés pour souligner les relations mathématiques entre les composantes principales résumant les spectres, et la nature du spectre, « face graisse ou cuir » dans notre cas. L'analyse discriminante standard utilisée dans ce projet est linéaire. Elle va donc chercher une combinaison linéaire des variables d'entrée que sont les composantes principales, pour classer les spectres en deux groupes, spectres de face cuir et spectres de face graisse. La qualité du modèle est évaluée à travers une matrice de confusion (Visa et al., 2011). Le Tableau 6 montre une matrice de confusion pour un modèle à deux classes.

Tableau 6 : Matrice de confusion

		Classe d'appartenance prédite par le modèle PCA-LDA	
		cuir	graisse
Classe d'appartenance réelle	cuir	a	b
	graisse	c	d

(c) ii. Lambda de Wilks :

Le lambda de Wilks est ici utilisé sur les spectres bruts ou prétraités pour déterminer quels prétraitements seront les plus efficaces. Le lambda de Wilk's est calculé à l'aide de tests multidimensionnels (test paramétrique, distance de Mahalanobis) et représente la variance interclasse sur la variance totale avec la formule suivante :

Équation 24

$$\boldsymbol{\lambda = 1 - \frac{trace(matrice\ de\ covariance\ intraclasse)}{trace(matrice\ de\ covariance\ totale)}}$$

Ici la matrice de covariance intraclasse correspond à l'ensemble des spectres qui compose une couenne et la matrice de covariance totale correspond à l'ensemble des spectres de la base de données.

Ici l'indice varie entre 0 et 1 ; plus il est proche de 1, plus il est possible de discriminer les spectres correspondant à une couenne de l'ensemble des spectres. En revanche, plus la valeur tend vers 0, plus les centroïdes sont égaux, rendant la discrimination des spectres d'une couenne vis-à-vis de l'ensemble des spectres impossible.

(c) iii. Le signal quality index :

Le signal quality index est un indicateur qui a été développé par l'équipe de Jean Michel Roger (données non publiées) au sein de l'IRSTEA (Institut national de recherche en science et technologies pour l'environnement et l'agriculture). C'est un indicateur qui a pour but d'évaluer la capacité d'un signal pour prédire un ou plusieurs paramètres d'une matrice à analyser.

Pour cela deux matrices sont considérées ; une matrice X qui comporte un nombre n d'échantillons avec p variables (nombre de longueurs d'onde dans le cadre de spectres), une matrice Y avec toujours n échantillons qui possède q caractéristiques à prédire. Ici la matrice X correspond aux spectres d'une base de données et la matrice Y correspond à l'ensemble des analyses de référence associées à cette même base de données. Un sous-espace qui peut être considéré comme l'espace des individus (échantillon) est déterminé en projetant X orthogonalement sur Y :

Équation 25

$$X^{\parallel} = P_y(X) = Y(Y^T Y)^{-1} Y^T X$$

Ce sous espace permet de prendre en compte la part utile de X pour quantifier Y. La covariance de $X^{\|}$ est calculée afin de mesurer la distance des éléments de ce sous-ensemble.

Équation 26

$$Cov(X^{\|}) = X^{\|\mathrm{T}}X^{\|} = X^T Y(Y^T Y)^{-1}Y^T X$$

Le calcul de la trace de cette matrice résume la part commune de X et de Y par un scalaire. Pour construire un indicateur sans dimension, on normalise par la trace de la covariance de X, qui représente la variance totale de X.

Équation 27

$$SQi = \frac{tr(X^T Y(Y^T Y)^{-1}Y^T X)}{tr(X^T X)}$$

Cet indicateur peut permettre de quantifier la part utile des spectres pour prédire les valeurs des références. Il est compris entre 0 et 1 et ne possède pas de dimension.

II.3. Résultats et discussion

II.3.1.(1). Caractérisation de la variabilité de la composition de couennes

Une analyse descriptive des échantillons a été réalisée en fonction du type de production (conventionnel ou biologique), de l'objectif d'élevage (les porcs pour la production ou les coches pour la reproduction), de la partie de l'animal (bardière ou couennes autre) et du conditionnement (congelé ou frais) (Figure 20). Cette analyse ne met pas en évidence de groupe statistiquement différent entre les différentes catégories. Cependant, quel que soit le regroupement utilisé, une catégorie est très fortement majoritaire par rapport à la seconde.

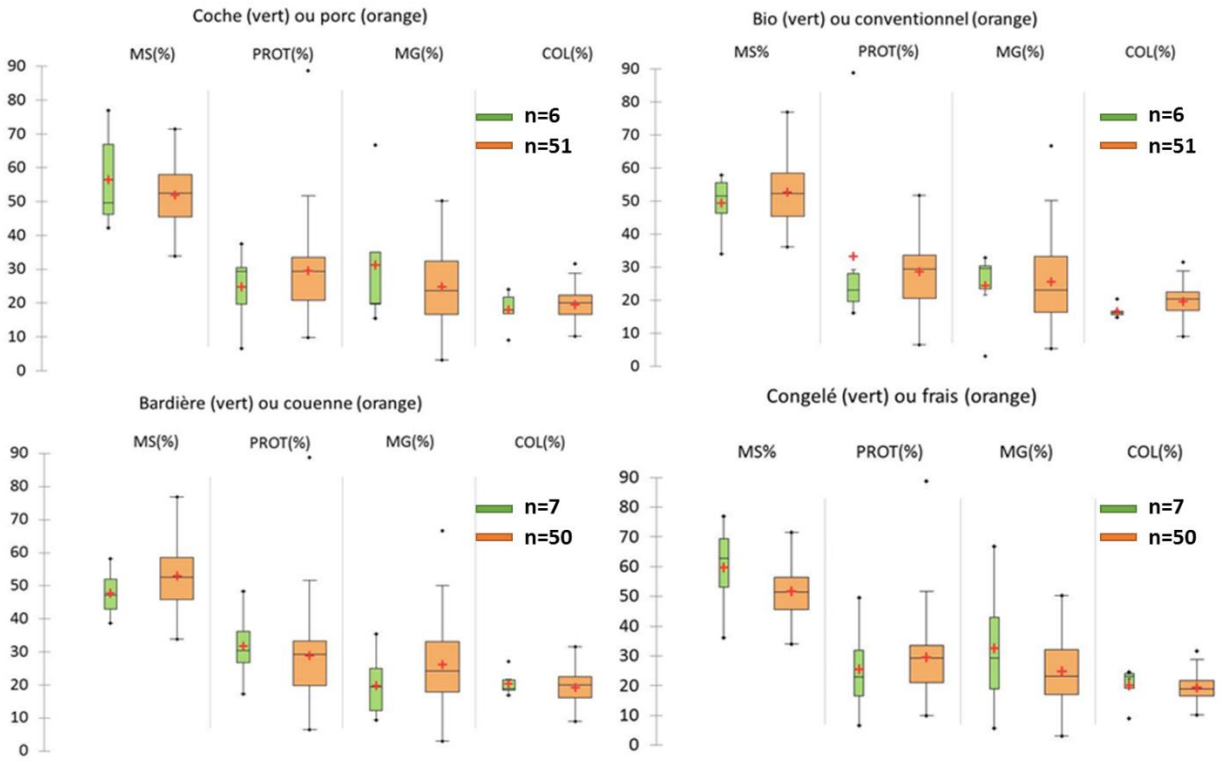


Figure 20 : Analyse descriptive des échantillons en fonction du type de production (biologique ou conventionnel), de l'objectif d'élevage (les porcs pour la production ou les coches pour la reproduction), de la partie de l'animal (bardière ou couennes autre) et du conditionnement (congelé ou frais). MS Matières Sèches, PROT Protéines, MG masse Grasse, COL Collagène.

Une première analyse descriptive a été réalisée sur les 4 critères quantitatifs (le taux de matière sèche, le taux de protéine, le taux de graisse et le taux d'hydroxyproline) à partir de l'analyse des 57 couennes de l'étude (Tableau 7).

Tableau 7 : Moyenne, Gamme, écart-type, coefficient de variation, écart-type de répétabilité et limite de répétabilité biologique (écart-type en trois prises d'échantillons différentes) des 57 couennes. Les résultats sont exprimés en pourcentage par rapport à la masse fraîche de couenne

	Humidité (%)	Protéines (%)	Taux de graisse (%)	Taux de collagène (%)
Moyenne	52.3	29.2	25.4	19.4
Minimum	33.9	6.6	3.1	9.1
Maximum	76.9	88.7	66.7	31.6
Ecart-type	8.8	13.4	12.5	4,6
Coefficient de variation (%)	16.9	45.7	49.1	23.8
Ecart-type de répétabilité	Ø	1.8	6.1	0,6
Limite de répétabilité biologique (3 prises d'essais du même échantillon)	Ø	4.9	17.3	4.8

La mesure du taux de matière sèche a été réalisée par mesure de perte de masse sur l'ensemble de l'échantillon, ce qui ne permet pas de déterminer l'écart-type de répétabilité et la limite de répétabilité biologique. C'est pour le taux de graisse et le taux de protéine que l'on observe les coefficients de variation les plus élevés (45,73% et 49,12%). Une grande plage de teneur est favorable pour le développement d'équation de prédiction. L'amplitude est moins importante pour les matières sèches et le taux d'hydroxyproline (16,90% et 23,77%). Les coefficients de variation confirment l'hétérogénéité des échantillons analysés provenant de l'échantillonnage basé sur une sélection représentative de l'approvisionnement de la chaîne de production. Ici, la variation du taux de matière sèche est cohérente avec la nature des échantillons, provenant d'une matrice biologique animale. L'écart-type de répétabilité et la limite de répétabilité sont relativement élevés pour le taux de graisse (E-T=6,12 et LR=17,3), notamment en comparaison au taux de protéine (E-T=1,75 et LR=4,91). Cela pourrait s'expliquer par l'échantillonnage au moment du broyage sur les couennes lyophilisées. Le broyage a eu pour effet de créer des agglutinats de graisse, ne permettant pas d'avoir un échantillon final homogène. Les coefficients de corrélation entre les analyses de référence prises deux à deux sont présentés dans la Figure 21. La répartition des échantillons selon leurs teneurs en matière sèche, protéines, matières grasses, et collagène, suit une loi normale. La teneur en matière sèche est corrélée positivement avec le taux de graisse (r=0,87) et est corrélée négativement avec le taux de protéine (r=0,80). Le taux d'hydroxyproline, qui correspond au taux de collagène, ne semble pas être corrélé au taux de matière sèche (r=0,18), de graisse (r=0,55) ou de protéine (r=0,46).

Le collagène étant la protéine grandement majoritaire de la peau, l'absence de corrélation avec le taux de protéines interroge sur la représentativité de l'analyse du collagène. En effet, Le taux d'hydroxyproline repose sur une réaction d'oxydo-réduction, il est possible que cette réaction ne réagisse pas spécifiquement avec l'hydroxyproline (Hofman *et al.*, 2011; Stegemann et Stalder, 1967). L'analyse du taux d'hydroxyproline semble être une méthode bien moins fiable que le dosage des protéines par la méthode de Kjeldahl pour des couennes de porc, d'autant que les échantillons n'ont pas été préalablement délipidés.

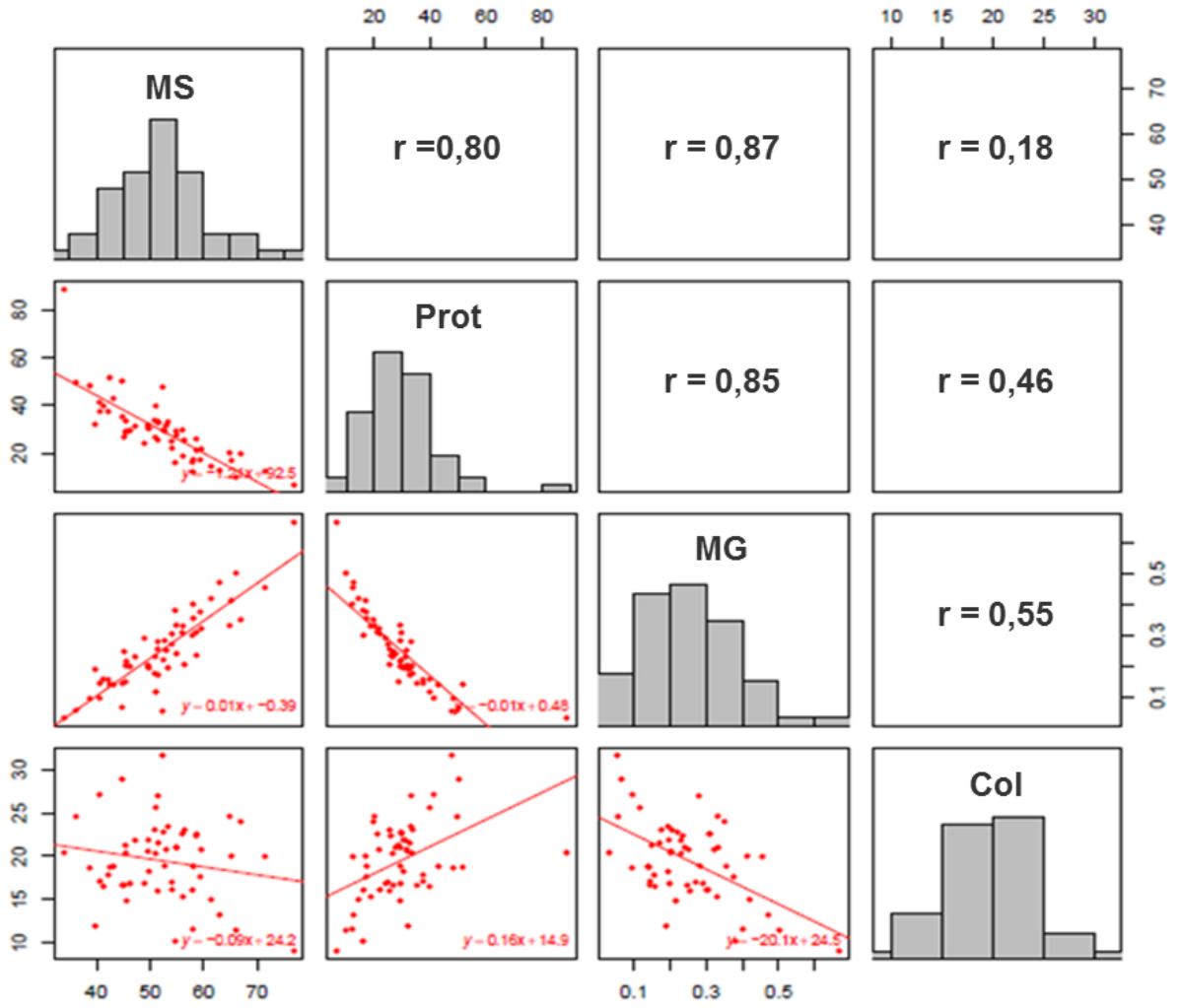


Figure 21 : Matrice de corrélation entre le taux de matière sèche (Hum), le taux de protéine (Prot), les matières grasses en mg/g (MG) et le taux d'hydroxyproline (Col)

II.3.1.(2). Analyse variabilité spectrale

(a) Discrimination des spectres de couenne des spectres parasites

L'analyse en ligne suppose la présence d'un flux continu ou discontinu. Pour l'analyse de matières premières d'une usine de production de gélatine, on est en présence d'un flux discontinu, qui a été simulé en laboratoire. La discontinuité implique de sélectionner les spectres correspondant à l'échantillon en les discriminant des spectres de tapis de convoyage. Il est possible de distinguer deux types de population de spectres. Les spectres possédant une intensité en réflectance importante qui représentent une première population relativement homogène, ce sont les spectres associés au polyéthylène de la bande de convoyage (Figure 22). La seconde population correspond aux spectres de couenne, possédant une intensité en réflectance beaucoup plus faible, représentant une population plus hétérogène.

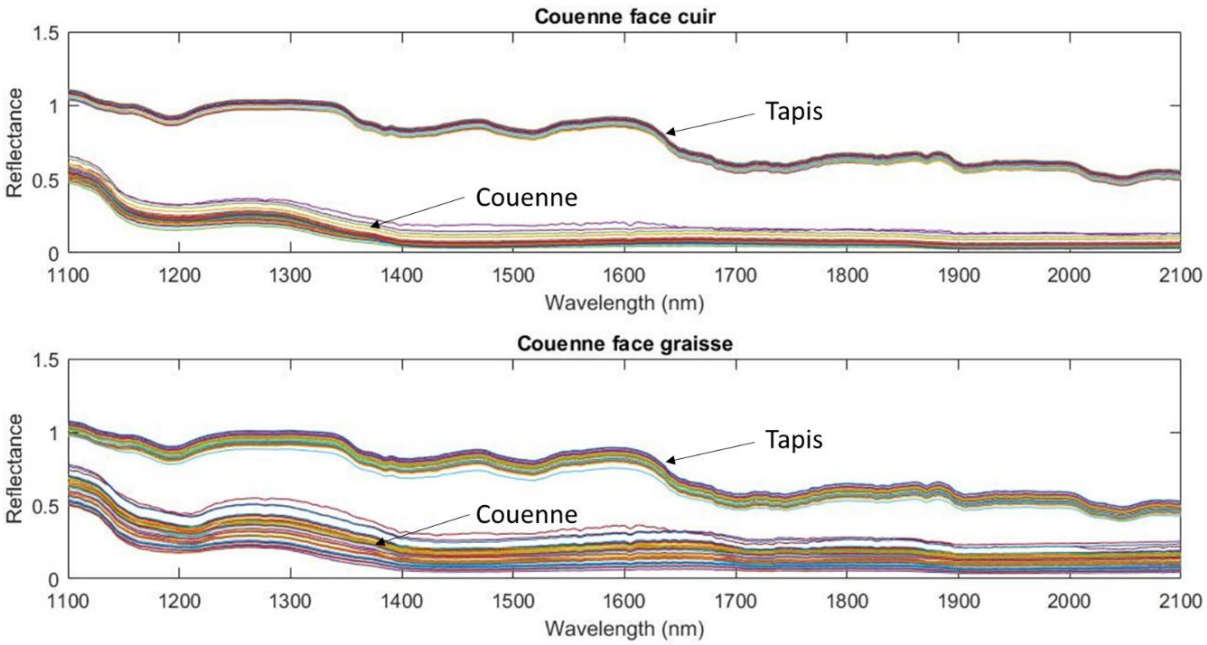


Figure 22 : Spectres collectés en ligne en réflectance sur un de tapis de convoyage, pour la face cuir (haut) et la face graisse (bas) de la couenne numéro 5.

Sur le tapis de convoyage les couennes se différencient par leurs textures de surface, leurs épaisseurs et la nature des tissus exposés. Cette hétérogénéité existe également au sein d'une même couenne. La hauteur de la sonde, la diffusion de lumière à la surface de la couenne ont conduit à une variabilité des spectres. Ainsi les populations de spectres provenant de la face cuir sont plus homogènes que les populations de spectres provenant de la face graisse (Figure 22).

Pour sélectionner les spectres, des modèles d'Analyse en Composantes Principales (ACP) ont été développés sur les spectres de chaque échantillon en mouvement sur le tapis de convoyage. La Figure 23 présente le principe de la sélection à partir des résultats des ACP pour deux couennes différentes sur la face graisse. La composante principale 1 explique 100 % de la variance, probablement du fait de la différence d'intensité entre les différentes populations de spectres. A partir des ACP, les spectres de couennes sont sélectionnés visuellement.

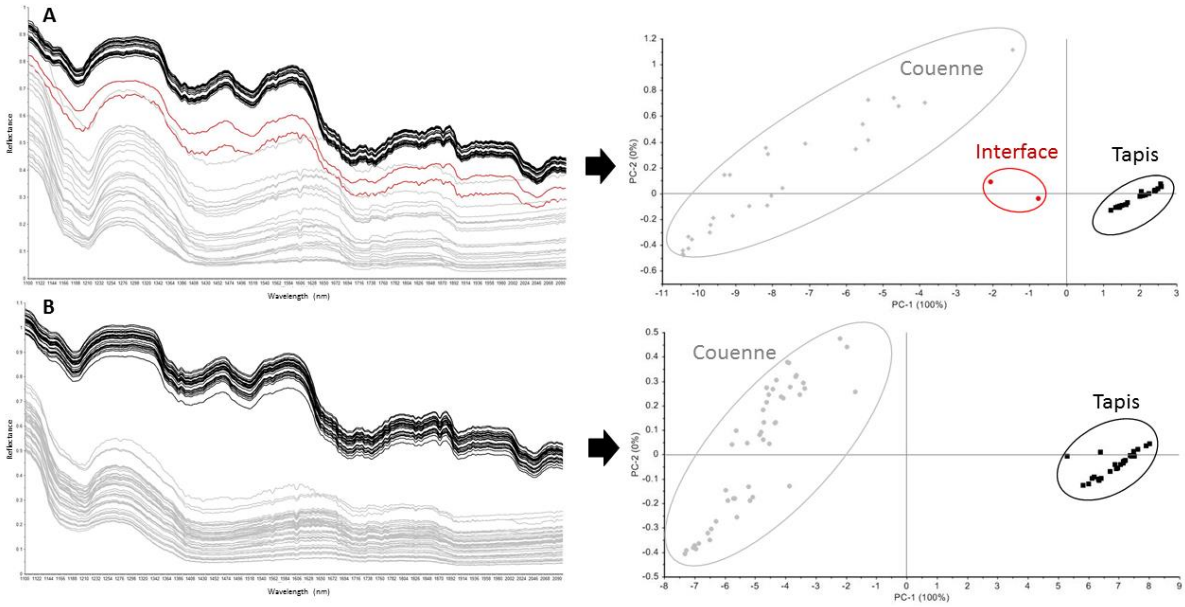


Figure 23: Spectres et scores de l'ACP des données spectrales de la face graisse de couennes. A : pour la face graisse de la couenne numéro 1. B : pour la face graisse de la couenne numéro 5. En noir les spectres correspondant au tapis, en gris les spectres correspondant à la couenne, en rouge les spectres de couenne à l'interface avec le tapis.

La plupart du temps, il est possible de distinguer deux populations sur l'ACP (Figure 23, B), mais dans certains cas il est possible de trouver des spectres à l'interface entre la couenne et le tapis de convoyage (Figure 23, A). Les spectres provenant uniquement de couennes ont été utilisés pour développer les différents modèles (Westad *et al.*, 2004).

Les contraintes d'acquisition industrielle liées essentiellement au tapis de convoyage ainsi qu'aux différents types de tissus peuvent être limitées par le tri des spectres permettant de distinguer les spectres d'intérêt pour les analyses futures.

(b) Différences de variabilités des spectres entre les deux faces d'une couenne

Après sélection des spectres, une couenne possède une série de spectres correspondant à la face cuir (Figure 24, en bleu) et une série correspondant à la face graisse (Figure 24, en magenta) ; c'est l'ensemble des spectres faces cuir et graisse qui caractérisent cette couenne (Figure 24).

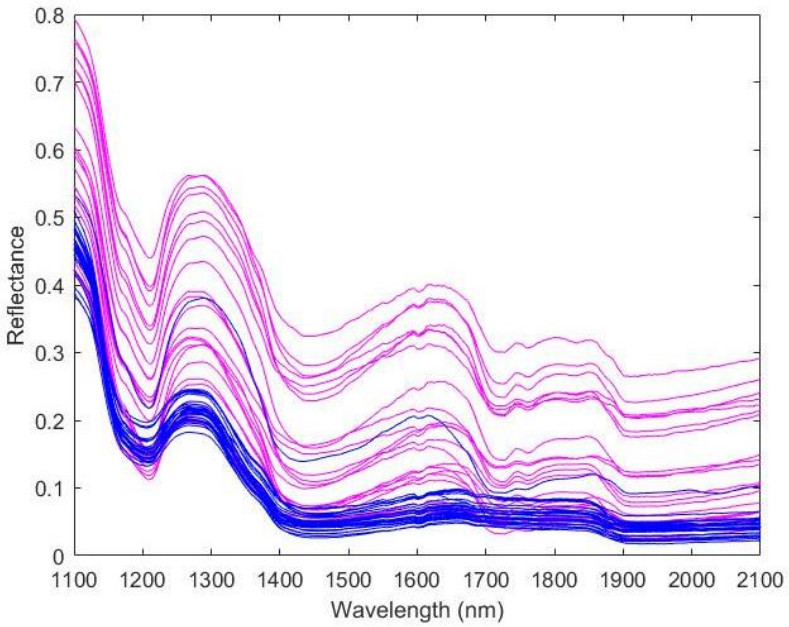


Figure 24 : Spectres en réflectance de la couenne numéro 1 ; En magenta les spectres correspondant à la face graisse, en bleu les spectres correspondant à la face cuir.

Le profil des spectres en réflectance est différent entre les échantillons provenant de la face graisse et ceux provenant de la face cuir. Les spectres sont faiblement bruités et présentent une intensité comprise entre 0 et 1. L'amplitude d'intensité est importante entre les différents spectres d'une même couenne. C'est généralement sur la face graisse que l'on observe les plus grands écarts d'intensité entre les spectres. La surface de la face cuir, plus homogène que celle de la face graisse pourrait expliquer ces écarts. En effet, dans les abattoirs la face graisse n'est pas découpée de manière uniforme, ce qui a pour conséquence d'entraîner une épaisseur variable de l'hypoderme.

Plusieurs éléments participent à l'explication de la différence d'absorption entre la face graisse et la face cuir. La première concerne la nature même de la face externe de la peau, barrière naturelle limitant la pénétration de la lumière. L'absorbance dans le proche infrarouge peut en être limitée.

Le deuxième est lié à l'hétérogénéité de surface, beaucoup plus marquée sur la face graisse, conduisant à une diffusion de lumière supérieure et donc un signal d'absorption apparent plus important. Le troisième élément implique la composition chimique : la face graisse étant plus riche en lipide que la face cuir, elle absorberait plus que cette dernière pour la gamme de longueur d'onde considérée.

(c) Confirmation des forts effets de diffusions de lumière

La Figure 25 A présente, pour une couenne, l'intensité de l'ensemble des spectres pour les deux faces d'une couenne en fonction de l'intensité du spectre moyen. Cette représentation a pour but de visualiser les effets additifs et multiplicatifs liés aux phénomènes de diffusion de lumière. Avant le prétraitement, plus la gamme de répartition de l'intensité des spectres est large plus les effets additifs seront marqués. On observe que pour la face graisse les effets additifs sont nettement plus marqués que pour la face cuir, témoignant d'une hétérogénéité de surface plus importante. Après prétraitement SNV, permettant de corriger les effets additifs, seuls les effets multiplicatifs demeurent visibles (Gobrecht *et al.*, 2014). Les effets multiplicatifs sont également plus importants pour la face graisse, visibles par l'étalement des réponses des spectres individuels aux valeurs les plus élevés d'intensité du spectre moyen.

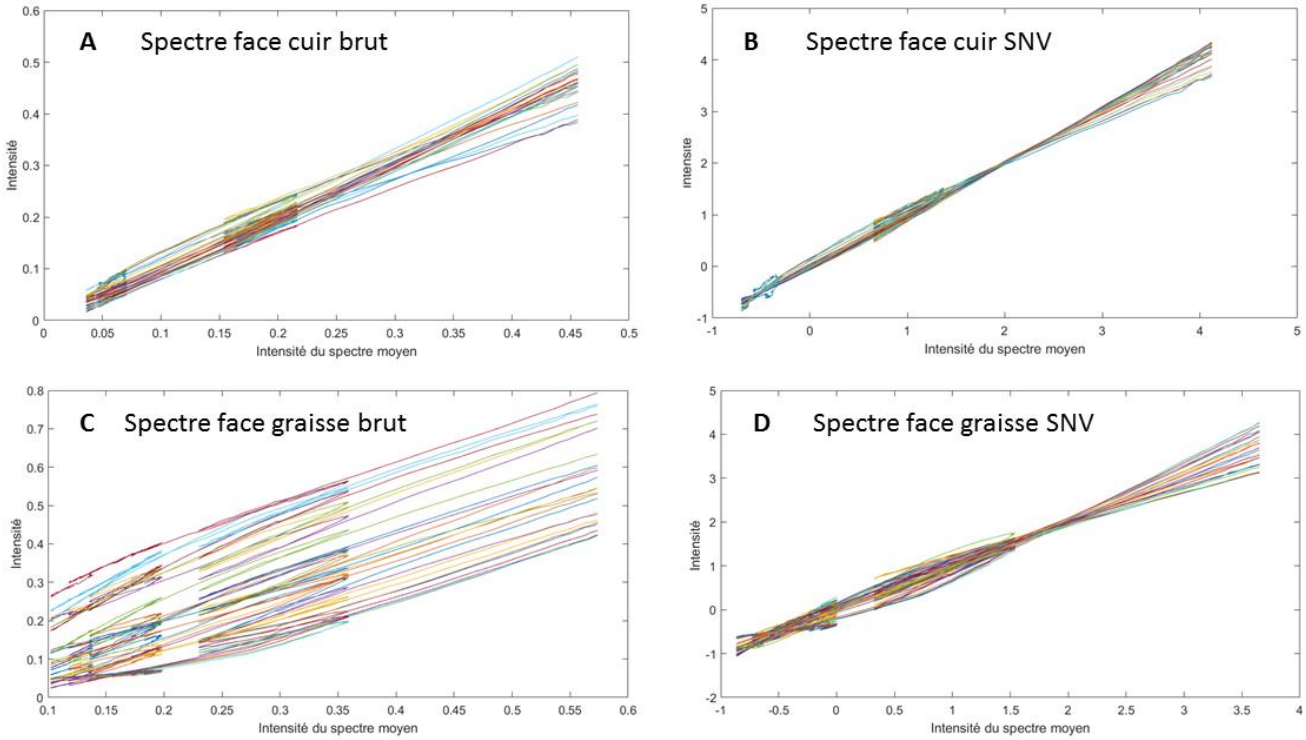


Figure 25 : Intensité des spectres face cuir (A et B) et face graisse (C et D) en fonction de l'intensité du spectre moyen ; en A spectres bruts face cuir, B spectres après prétraitement SNV face cuir, C spectres bruts face graisse et D spectres après prétraitement SNV face graisse.

(a) Spectres des différents constituants d'une couenne.

La Figure 26 représente les spectres bruts en réflectance de graisse, de muscle et de cuir provenant de zones spécifiques de couennes. Il est possible d'observer une différence visuelle entre les spectres provenant des différentes zones. Le spectre provenant du cuir est très similaire au spectre provenant du muscle, hormis une différence d'intensité qui est probablement due à la réflexion importante de la face cuir. En revanche, le spectre de graisse présente un profil différent des deux autres spectres. D'après la bibliographie, les bandes à 1460 nm et à 1960 nm sont généralement attribuées au taux d'humidité et correspondent à des combinaisons de premières et secondes harmoniques des OH (Murray et Williams, 1987; Cozzolino, De Mattos, et Martins, 2002; Cozzolino et Murray, 2004). En ce qui concerne la présence des lipides, en spectroscopie proche infrarouge elle est essentiellement attribuée à la présence de liaisons de type C-H. Dans la gamme spectrale de notre analyse, on peut distinguer deux plages spectrales ; une première gamme entre 1100 nm et 1250 nm et une deuxième gamme entre 1250 nm et 1800 nm. Elles sont attribuées respectivement aux deuxièmes harmoniques et aux premières harmoniques des C-H (Bertrand et Dufour, 2006b).

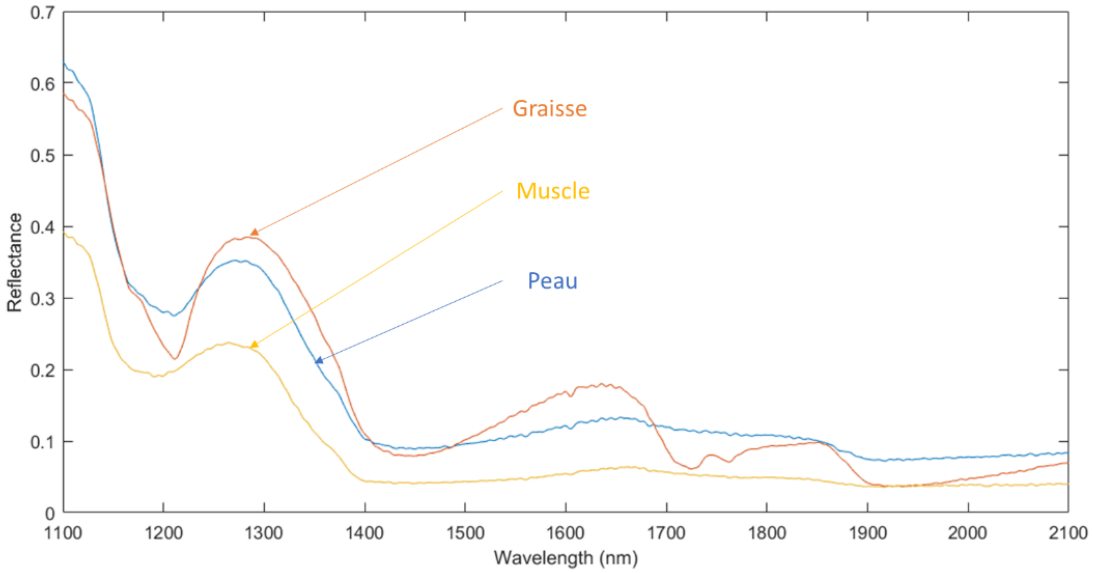


Figure 26 : Spectres en réflectance de muscle (en jaune), de graisse (en rouge) et de peau (en bleu).

Les protéines montrent des bandes caractéristiques de la liaison peptidique qui se situent autour de 2060 nm, 2170 et 2210 nm (Kamishikiryo). Plus généralement, on considère que la liaison peptidique possède une bande d'absorption entre 2100 nm et 2200 nm, non couverte par l'analyse. En revanche, la liaison peptidique comprend aussi une liaison N-H, qui possède des bandes entre 1460 nm et 1570 nm attribuées aux combinaisons et aux premières harmoniques (Sadler *et al.*, 1984).

En comparant les différentes régions d'intérêt sur les trois types de spectre, on remarque que les différences entre les spectres sont principalement liées à l'absorption de l'eau et des lipides. Les lipides semblent être l'élément majeur de discrimination entre les couennes.

(b) ACP pour visualiser la variabilité spectrale

Des analyses non-supervisées (ACP) ont été réalisées sur les spectres bruts ou prétraités afin de faire apparaître des regroupements d'échantillons au sein de la base de données. La Figure 27 représente les scores de l'ACP sur l'ensemble des spectres faces cuir et graisse sur les deux premières composantes. La PC1 explique 92% (données brutes) et 73% (après prétraitement SNVD) de la variance totale. La plus haute valeur observée pour les contributions de la PC1 est observée vers 1350 nm, ce qui correspondrait à la présence de graisse.

La PC2 explique 5% (données brutes) et 16% (après prétraitement SNVD) de la variance totale et présente un maximum vers 1210 nm, pouvant être également associé à la présence de graisse.

Il est possible de distinguer, sur l'ACP des spectres bruts, deux groupes correspondant respectivement aux faces cuir et graisse. En revanche, les groupes sont superposés et présentent une disparité importante. Après prétraitement SNVD, les deux groupes sont plus distincts, avec un groupe plus homogène pour les spectres provenant de la face cuir et un groupe plus hétérogène pour les spectres provenant de la face graisse.

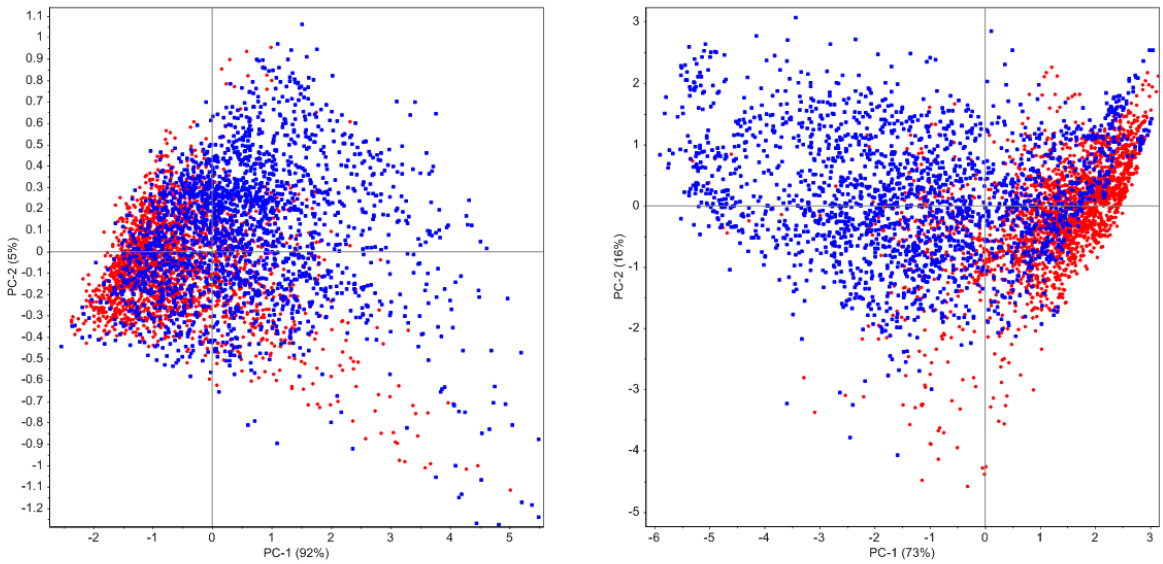


Figure 27 : Scores des ACP sur les spectres face cuir (points rouge) et face graisse (points bleu). A gauche l'ACP sur les spectres bruts, à droite l'ACP sur les spectres après prétraitement SNVD (ordre 1)

Un modèle ACP-LDA a été développé pour prédire l'appartenance des spectres au groupe « cuir » ou « graisse » à partir des 4254 spectres de l'essai. Avec les 10 premières composantes principales, le pourcentage de spectres bien identifiés comme « cuir » ou « graisse » est de 91%. Même si les spectres provenant de la face graisse (95%) sont mieux classés que les spectres provenant de la face cuir (87%), ces résultats montrent qu'il est possible de classer, avec de bonnes performances, les spectres en fonction de la face mesurée (Tableau 8).

Tableau 8 : Matrice de confusion du modèles ACP-LDA sur les 4254 spectres bruts de l'essai, en fonction de leurs provenance (spectres acquis face cuir et spectres acquis face graisse)

Matrice de confusion		Prédit	
		Cuir	Graisse
Vrai	Cuir	2057 (87%)	89 (5%)
	Graisse	295 (13%)	1813 (95%)

De la même façon, les échantillons ont été regroupés sur les différentes ACP par des critères qualitatifs : type d'animal (coche ou porc), partie de l'animal (bardière ou couennes autre), type de production (conventionnel ou biologique), conditionnement (fraîche ou congelée). Les résultats n'ont pas permis de mettre en évidence la présence d'un groupe par rapport à un autre, principalement à cause du faible nombre d'échantillons pour certaines modalités (résultats non présentés dans le document).

Dans le cadre d'analyses de couenne, ces résultats montrent qu'il est possible de discriminer les spectres provenant de la face cuir, des spectres provenant de la face graisse. Dans de nombreux cas, les spectres peuvent appartenir aux deux faces principalement du fait de l'hétérogénéité importante des couennes.

Afin de mettre en évidence le positionnement des couennes sur les graphiques de dispersion des ACP en fonction de leur composition, des classes par quartile ont été créées, relatives aux valeurs d'analyses de référence pour le taux de matière sèche, le taux de protéine, le taux de matière grasse et le dosage du collagène. La Figure 28 représente les scores des ACP sur la base de données spectrales complète après prétraitement SNVD. A droite (A) les spectres provenant de la face graisse, à gauche (B) les spectres provenant de la face cuir. En bleu, les échantillons ont été regroupés par catégorie de taux de matière grasse, le bleu passe du bleu foncé pour le quartile supérieur au bleu clair pour le quartile inférieur. En rouge-orangé, les échantillons ont été regroupés par catégorie de taux de protéine, le rouge foncé est utilisé pour le quartile supérieur et passe au jaune pour le quartile inférieur.

Les résultats des différentes ACP sont relativement proches pour les différents prétraitements, mais la meilleure discrimination a été observée pour le prétraitement SNVD.

Ici la PC1 explique 71% de la variation totale pour la face graisse pour seulement 43% pour la face cuir. Les contributions présentent des maximums à 1210 nm, 1345 nm et 1730 nm, ce qui correspondraient plutôt aux interactions avec la matière grasse.

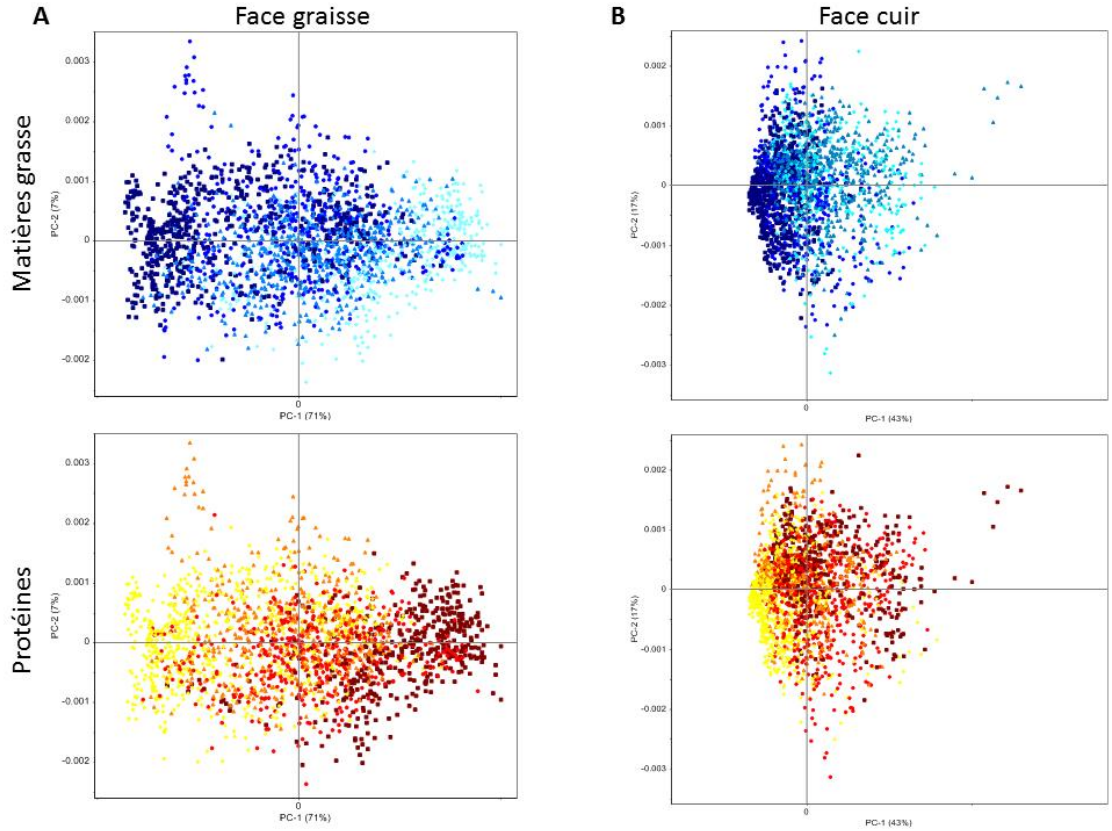


Figure 28 : Score des ACP sur les spectres en ligne des 57 couennes non moyennés après prétraitement SNVD pour ; A - pour la face graisse, B - Pour la face cuir. En haut en bleu sont représentées les catégories par quartile pour la matière grasse (décroissant du bleu foncé au bleu clair), en bas en rouge sont représentées les catégories par quartile pour les protéines (croissant du jaune au rouge foncé)

Conformément aux observations précédentes, la face graisse présente une répartition beaucoup plus hétérogène que la face cuir, avec une répartition des catégories le long de la PC1.

Pour les classes, en fonction du taux de matière grasse, on observe une répartition le long de l'axe de la PC1 allant du quartile supérieur au quartile inférieur. La répartition est strictement inversée pour les classes en fonction du taux de protéine, on observe une répartition de droite à gauche allant du quartile supérieur au quartile inférieur. Cela confirme que les échantillons qui présentent un taux de protéine élevé possèdent un taux de matière grasse faible et réciproquement.

Cependant, il est important de noter qu'il est difficile de distinguer les quartiles entourant la médiane. Pour la face cuir, les groupes sont beaucoup moins distincts, on observe une répartition similaire le long de la PC1. En général, la face cuir présente une moins bonne capacité de répartition des échantillons.

Les mêmes analyses ont été réalisées sur le taux de matière sèche et le dosage du collagène. Pour le dosage du collagène, les résultats sont similaires aux résultats observés pour le taux de protéine, mais la répartition des groupes est beaucoup moins marquée. Pour le taux de matière sèche, l'ACP ne permet pas de mettre en évidence la présence de groupes distincts. En revanche, les échantillons présentant une teneur en matière sèche très faible sont majoritairement répartis sur l'abscisse positive de la PC1, ce qui correspond aux échantillons avec une teneur en protéine élevée et une teneur en matière grasse faible. Cette observation n'a pas été effectuée sur les échantillons présentant une haute teneur en matière sèche ce qui signifie qu'un haut taux de protéine est aussi lié à un faible taux de matière sèche dans l'échantillon (r=-0,8).

II.3.1.(3). Evaluation de la capacité de prédiction des analyses de références à partir des spectres infrarouges collectés en ligne.

Le développement des équations de prédiction a été réalisé sur un total de 57 échantillons pour la face cuir, 56 échantillons pour la face graisse et 113 échantillons pour la face cuir et graisse simultanément. Les équations ont été développées en validation croisée, car le principal objectif est d'étudier la capacité de prédiction. De plus, au vu du faible nombre d'échantillons, l'utilisation d'un jeu de données pour la validation externe ne permettrait pas la création d'un modèle de prédiction robuste.

Un échantillon est en marge sur les différents modèles, ce qui a pour conséquence d'induire un biais (Figure 29, étoiles noires pour les prédictions sur l'ensemble des échantillons à l'extrémité de l'abscisse positive). Cet échantillon a été considéré comme un extrême au moment de l'échantillonnage, principalement dû à son aspect visuel très déshydraté.

Au moment de la construction des modèles, deux échantillons ont été considérés comme valeurs aberrantes aussi bien pour les bases de données avec les spectres complets que pour les bases de données avec les spectres moyennés.

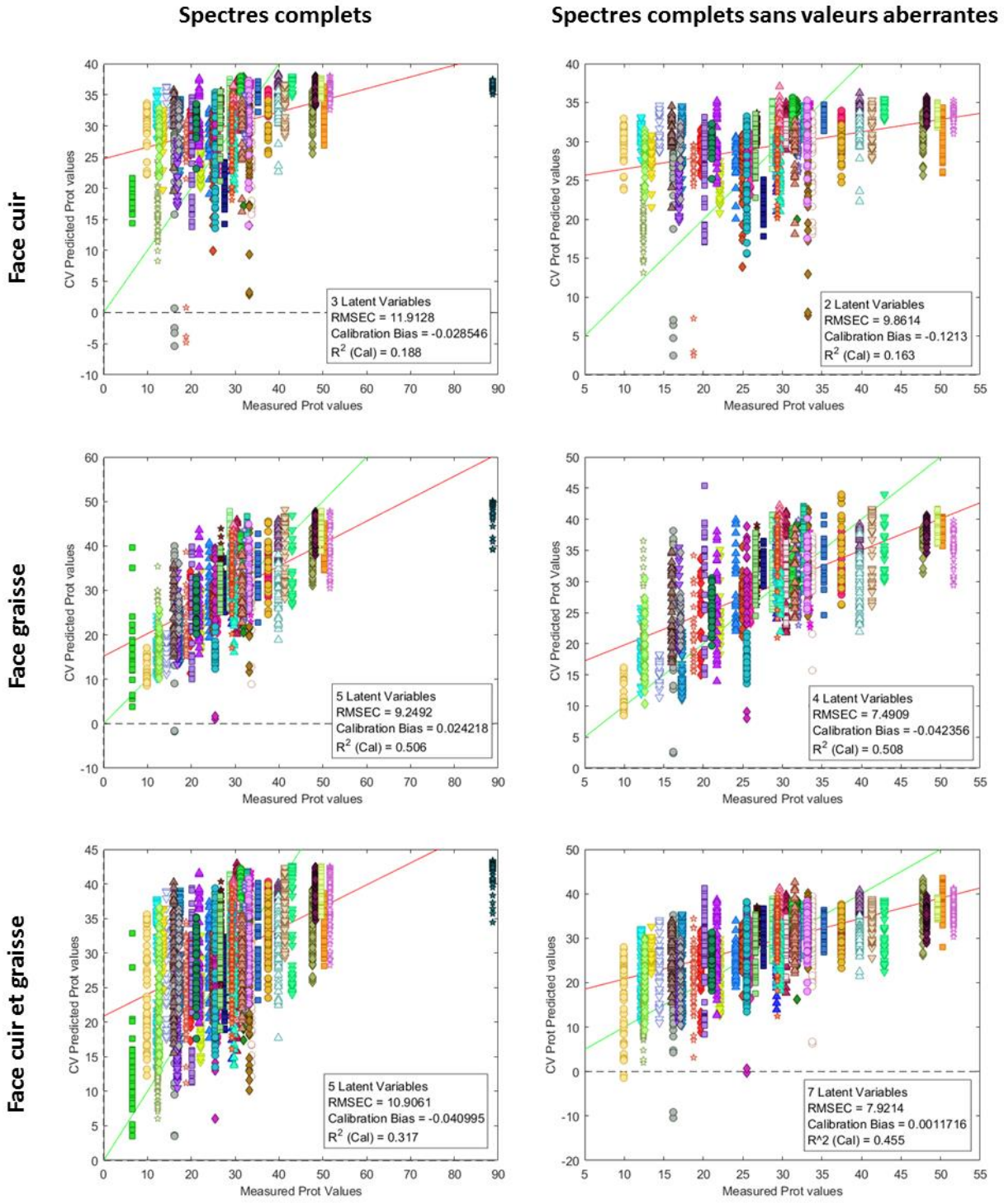


Figure 29 : Modèles PLS par validation croisée du taux de protéine. A gauche la prédiction pour l'ensemble des échantillons sur la base de spectres complets, à droite la prédiction sans les échantillons présentant des valeurs aberrantes. En haut pour les spectres face cuir, au centre pour les spectres face graisse, en bas pour les spectres faces cuir et graisse. Les équations de prédiction présentées sont réalisées sur les données brutes.

Les différents modèles de prédiction développés pour les différents prétraitements possèdent un nombre de variables latentes compris entre 2 et 10. Les meilleures capacités de prédiction sont observées pour la base de données face graisse. A l'inverse, les capacités de prédiction pour la face cuir sont les moins performantes. Les capacités de prédictions pour les faces cuir et graisse ensemble sont intermédiaires. Par exemple, après avoir enlevé les échantillons aberrants sur les données brutes des spectres complets, le coefficient de détermination du taux de protéines est de 0,51 pour la face graisse, 0,16 pour la face cuir, et 0,45 pour la face cuir et graisse. La capacité de prédiction est doublée entre la face cuir et la face graisse. Plus généralement, le taux de matière sèche, le taux de protéine et le taux de graisse présentent des capacités de prédiction similaires, alors que le dosage du collagène présente une capacité de prédiction inférieure.

(a) Prédiction du taux de protéine

La Tableau 9 représente les teneurs en protéine prédites en validation croisée en fonction des valeurs mesurées avec les différents prétraitements, sur les bases de données avec les spectres complets et les spectres moyens. Les coefficients de détermination en calibration pour le taux de protéine varient entre 0,16 et 0,80 avec une RMSEC comprise entre 4,98 % et 9,86 % en fonction des prétraitements et des différentes bases de données. Ces observations sont en accord avec les travaux de Liao (2010) sur les prédictions du taux de protéine à partir d'échantillons de porc frais (Liao, Fan, et Cheng, 2010). Il y a une augmentation de 0,12 à 0,34 pour le coefficient de détermination et une diminution de 0,52 % à 3,29 % pour les RMSEC entre les bases de données complètes (tous les spectres de chaque couenne) et les bases des spectres moyens (spectre moyen de chaque couenne). Les modèles de prédiction présentent un RPD variant entre 1,35 et 2,88, avec un RPD maximum de 1,86 pour les bases de données décomposées. Une équation de prédiction est considérée comme applicable lorsqu'elle possède un $r^2 \geq 0,7$ et un RPD≥1,9 (Kapper, 2012).

Les équations de prédiction obtenues sur les bases de données avec les spectres moyens laissent penser qu'un modèle de prédiction serait applicable pour la prédiction du taux de protéine. Les meilleurs modèles de prédiction ont été observés pour les bases de données comprenant uniquement les spectres provenant de la face graisse. Les protéines sont principalement situées au niveau de l'épiderme et du mésoderme qui se trouve sur la face cuir laissant penser que la prédiction des protéines dépend directement de la proportion de lipide présente.

Tableau 9 : Statistiques de calibration et de validation croisée pour les modèles de prédiction du taux de protéine. Les données ont été traitées par PLSr en validation croisée, dans le cadre de l'analyse des spectres complets les échantillons ont été considérés par bloc. LV est le nombre de variables latentes.

Base de données		Prétraitement	LV	RMSEC	RMSECV	Bias	CV Bias	r^2 cal	r^2 val	RPD
Face cuir n=55	**Spectres complets** n_s=2089	Brut	2	9.861	9.896	-0.12	-0.13	0.16	0.16	1.35
		D1	3	8.295	8.304	0.04	0.03	0.41	0.41	1.61
		D2	4	7.566	7.594	0.00	0.00	0.51	0.50	1.77
		SNVD	5	8.161	8.175	0.02	0.02	0.43	0.42	1.64
		MSC	5	8.273	8.288	0.00	0.00	0.41	0.41	1.61
	Spectres moyens n_s=55	Brut	2	8.784	9.022	0.00	0.00	0.28	0.25	1.52
		D1	3	5.512	6.329	-0.03	-0.03	0.72	0.63	2.42
		D2	4	5.272	7.069	0.01	-0.17	0.74	0.54	2.53
		SNVD	5	6.315	7.244	0.00	-0.15	0.63	0.52	2.12
		MSC	5	6.306	7.177	0.00	-0.20	0.63	0.53	2.12
Face graisse n=54	**Spectres complets** n_s=2052	Brut	4	7.491	7.495	-0.04	-0.04	0.51	0.51	1.78
		D1	3	7.173	7.176	0.01	0.01	0.55	0.55	1.86
		D2	3	7.157	7.161	-0.03	-0.03	0.55	0.55	1.87
		SNVD	3	7.396	7.399	0.00	0.00	0.52	0.52	1.81
		MSC	4	7.336	7.340	0.00	0.00	0.53	0.53	1.82
	Spectres moyens n_s=54	Brut	5	4.870	5.229	-0.01	0.03	0.77	0.74	2.74
		D1	3	4.696	4.975	0.00	0.01	0.79	0.76	2.84
		D2	3	4.665	5.056	-0.04	-0.08	0.79	0.76	2.86
		SNVD	4	4.705	5.004	0.00	-0.02	0.79	0.76	2.84
		MSC	4	4.724	5.008	0.00	0.00	0.79	0.76	2.83
Face cuir et graisse n=55	**Spectres complets** n_s=4141	Brut	7	7.921	7.927	0.00	0.00	0.45	0.45	1.69
		D1	4	7.737	7.741	0.01	0.01	0.48	0.48	1.73
		D2	5	7.512	7.521	0.05	0.05	0.51	0.51	1.78
		SNVD	6	7.866	7.872	0.02	0.02	0.46	0.46	1.70
		MSC	6	7.859	7.856	0.00	0.00	0.46	0.46	1.70
	Spectres moyens n_s=109	Brut	10	4.628	5.409	0.01	0.01	0.80	0.73	2.89
		D1	4	5.724	6.053	0.02	-0.04	0.69	0.66	2.33
		D2	4	6.142	6.525	-0.01	-0.05	0.65	0.60	2.18
		SNVD	6	6.093	6.510	0.00	0.01	0.65	0.60	2.19
		MSC	9	4.967	5.678	0.00	0.06	0.77	0.70	2.69

(b) Prédiction du taux de matières sèches

Le coefficient de détermination (r^2) en calibration du taux de matière sèche varie entre 0,24 et 0,78 avec une RMSEC comprise entre 3,7% et 6,9% et un RPD compris entre 1,28 et 2,41 en fonction des différentes bases de données et des prétraitements (Tableau 10). Entre les bases de données multispectres et les bases de données des spectres moyens, il y a une augmentation du coefficient de détermination en calibration comprise entre 0,14 et 0,39 et une diminution de la RMSEC entre 0,36 et 2,35%. De plus, il y a une différence significative de RPD qui possède un maximum de 1,61 pour les bases multispectres pour un maximum de 2,41 pour les bases avec les spectres moyens. Il y a donc une amélioration conséquente des modèles de prédiction en moyennant les spectres des échantillons. C'est avec les spectres provenant de la face graisse que l'on obtient les meilleurs modèles de prédiction. En comparaison, il y a une diminution du r^2 en moyenne de 0,05 avec la base de données qui contient les spectres de graisse et cuir et une diminution moyenne du r^2 de 0,13 pour la base de données avec les spectres de cuir. Les différents prétraitements permettent d'améliorer jusqu'à 0,32 sur le coefficient de détermination en calibration. Pour la prédiction de la teneur en matière sèche, les meilleurs modèles ont été observés pour les données brutes, ou après un prétraitement de type dérivée seconde. D'après la théorie du RPD il est donc possible d'envisager la prédiction du taux de matière sèche des matières premières, notamment avec les modèles sur les spectres moyens.

Tableau 10 : Statistiques de calibration et de validation croisée pour les modèles de prédiction du taux de matière sèche. Les données ont été traitées par PLSr en validation croisée, dans le cadre de l'analyse des spectres complets les échantillons ont été considérés par bloc. LV correspond au nombre de variables latentes.

Base de données		Prétraitement	LV	RMSEC	RMSECV	Bias	CV Bias	r^2 cal	r^2 val	RPD
Face cuir n=55	Spectres complets n_s=2089	Brut	2	6.869	6.906	-0.19	-0.18	0.24	0.23	1.29
		D1	3	6.560	6.559	-0.10	-0.09	0.31	0.30	1.35
		D2	4	6.171	6.187	-0.09	-0.09	0.38	0.38	1.43
		SNVD	5	6.208	6.219	0.01	0.01	0.38	0.37	1.43
		MSC	5	6.196	6.209	0.00	0.00	0.38	0.38	1.43
	Spectres moyens n_s=55	Brut	3	6.230	6.550	-0.18	-0.27	0.38	0.32	1.42
		D1	5	4.359	5.047	0.01	0.17	0.70	0.60	2.03
		D2	6	4.294	5.708	-0.02	0.03	0.71	0.49	2.06
		SNVD	5	4.947	5.405	0.01	0.08	0.61	0.54	1.79
		MSC	5	4.982	5.530	0.00	0.03	0.60	0.52	1.78
Face graisse n=54	Spectres complets n_s=2052	Brut	4	5.932	5.939	-0.01	-0.01	0.43	0.43	1.49
		D1	3	5.611	5.615	-0.07	-0.07	0.49	0.49	1.58
		D2	3	5.499	5.505	-0.06	-0.06	0.51	0.51	1.61
		SNVD	3	5.888	5.891	0.00	-0.01	0.43	0.43	1.50
		MSC	4	5.681	5.686	0.00	0.00	0.47	0.47	1.56
	Spectres moyens n_s=54	Brut	5	3.706	4.118	-0.05	-0.17	0.78	0.73	2.39
		D1	3	3.798	4.144	-0.05	-0.11	0.77	0.72	2.33
		D2	3	3.676	4.202	-0.02	-0.08	0.78	0.71	2.41
		SNVD	4	3.715	4.138	0.00	-0.08	0.78	0.72	2.38
		MSC	4	3.719	4.112	0.00	-0.08	0.78	0.73	2.38
Face cuir et graisse n=55	Spectres complets n_s=4141	Brut	7	6.017	6.022	-0.11	-0.12	0.41	0.41	1.47
		D1	4	6.033	6.035	-0.09	-0.08	0.41	0.41	1.47
		D2	5	5.953	5.957	-0.06	-0.07	0.42	0.42	1.49
		SNVD	6	6.050	6.054	-0.02	-0.02	0.41	0.41	1.46
		MSC	6	6.023	6.028	0.00	0.00	0.41	0.41	1.47
	Spectres moyens n_s=109	Brut	10	3.667	4.259	-0.03	-0.13	0.78	0.71	2.41
		D1	4	4.755	4.927	-0.08	-0.10	0.64	0.61	1.86
		D2	4	4.820	5.013	-0.03	-0.05	0.63	0.60	1.84
		SNVD	6	4.644	4.892	-0.01	-0.09	0.65	0.62	1.91
		MSC	9	3.897	4.352	0.00	-0.13	0.76	0.70	2.27

(c) Prédiction du taux de matières grasses

En fonction des différentes bases de données et des prétraitements, les coefficients de détermination en calibration varient entre 0,17 et 0,75 avec des RMSEC comprises entre 5,30% et 10,01% et un RPD compris entre 1,24 et 2,35. Cela indique un taux de matière grasse potentiellement prédictible sur des couennes de porc. Il y a une augmentation de coefficient de détermination comprise entre 0,11 et 0,39, une diminution entre 0,53% et 2,56% de la RMSEC et une augmentation du RPD comprise entre 0,11 et 0,73 entre les bases de données multispectres et moyennées. C'est pour la face graisse que les meilleurs modèles de prédiction sont observés.

De plus, les prétraitements améliorent de façon importante les modèles de prédiction. Par exemple, pour les spectres moyens face cuir, le coefficient de détermination est de 0,17 pour les données brutes et de 0,50 après un prétraitement de type dérivée seconde. L'information spectrale du taux de graisse est donc très fortement influencée par la présence des effets additifs et multiplicatifs qui sont supprimés par la dérivée seconde.

Tableau 11 : Statistiques de calibration et de validation croisée pour les modèles de prédiction du taux de matières grasses. Les données ont été traitées par PLSr en validation croisée, dans le cadre de l'analyse des spectres complets les échantillons ont été considérés par bloc. LV est le nombre de variables latentes.

Base de données		Prétraitement	LV	RMSEC	RMSECV	Bias	CV Bias	r^2 cal	r^2 val	RPD
Face cuir n=55	Spectres complets n_s=2089	Brut	2	10.01	10.049	0.06	0.05	0.17	0.16	1.25
		D1	3	8.362	8.362	0.00	0.00	0.42	0.42	1.49
		D2	4	7.773	7.806	0.00	-0.01	0.50	0.50	1.61
		SNVD	5	8.164	8.18	0.00	0.00	0.45	0.45	1.53
		MSC	5	8.269	8.286	0.00	0.03	0.43	0.43	1.51
	Spectres moyens n_s=55	Brut	3	9.223	9.474	0.00	0.00	0.28	0.25	1.35
		D1	5	5.819	6.795	0.00	0.02	0.71	0.62	2.14
		D2	6	5.861	7.548	0.02	0.02	0.71	0.52	2.13
		SNVD	5	6.498	7.256	-0.01	-0.06	0.64	0.56	1.92
		MSC	5	6.358	7.153	0.00	0.00	0.66	0.57	1.96
Face graisse n=54	Spectres complets n_s=2052	Brut	4	7.737	7.741	0.06	0.06	0.49	0.49	1.61
		D1	3	7.446	7.448	0.00	0.00	0.53	0.53	1.68
		D2	3	7.446	7.449	0.00	0.00	0.53	0.53	1.68
		SNVD	3	7.644	7.647	0.00	0.00	0.50	0.50	1.63
		MSC	4	7.546	7.549	0.00	0.00	0.52	0.51	1.65
	Spectres moyens n_s=54	Brut	5	5.368	6.008	0.00	0.03	0.75	0.69	2.32
		D1	3	5.388	5.927	0.00	0.00	0.74	0.69	2.32
		D2	3	5.376	5.986	0.03	0.10	0.75	0.69	2.32
		SNVD	4	5.297	5.889	0.00	0.00	0.75	0.70	2.36
		MSC	4	5.317	5.885	0.00	0.00	0.75	0.70	2.35
Face cuir et graisse n=55	Spectres complets n_s=4141	Brut	7	8.033	8.039	0.00	0.01	0.46	0.46	1.55
		D1	4	7.88	7.884	0.01	0.01	0.48	0.48	1.58
		D2	5	7.714	7.723	-0.03	-0.03	0.50	0.50	1.62
		SNVD	6	7.966	7.972	-0.02	-0.02	0.47	0.47	1.57
		MSC	6	7.973	7.973	0.00	0.00	0.47	0.47	1.57
	Spectres moyens n_s=109	Brut	10	5.475	6.348	0.00	0.00	0.74	0.66	2.28
		D1	4	6.082	6.371	0.00	0.02	0.68	0.65	2.05
		D2	4	6.536	6.955	0.02	0.02	0.63	0.58	1.91
		SNVD	6	6.379	6.796	0.00	-0.06	0.65	0.60	1.96
		MSC	9	5.657	6.358	0.00	0.00	0.72	0.66	2.21

(d) Prédiction du dosage du collagène

Le dosage du collagène présente des capacités de prédiction inférieures aux trois autres analyses de référence avec une gamme de coefficients de détermination comprise entre 0,01 et 0,68, une RMSEC entre 2,50 et 4,60 et un RPD compris entre 1,00 et 1,84 selon les différentes bases de données et les prétraitements. Un modèle de prédiction est considéré comme applicable lorsqu'il possède un $r^2 \geq 0,7$ et un RPD≥1,9. D'après les valeurs indicatrices de la capacité de prédiction, un modèle de prédiction ne serait donc pas applicable pour le collagène. Les meilleures capacités de prédictions sont observées pour les bases de données à partir des spectres de la face graisse et pour les bases de données moyennées. L'analyse du taux d'hydroxyproline est une méthode bien moins fiable que le dosage des protéines par la méthode de Kjeldahl sur des couennes de porc. Notre hypothèse est que le taux d'hydroxyproline repose sur une réaction d'oxydo-réduction, qui ne réagirait pas spécifiquement avec l'hydroxyproline lors de l'analyse de ces couennes (Hofman et al., 2011; Stegemann et Stalder, 1967). Les matières premières devraient donc être délipidés avant de réaliser le dosage.

Tableau 12 : Statistiques de calibration et de validation croisée pour les modèles de prédiction du taux de collagène. Les données ont été traitées par PLSr en validation croisée, dans le cadre de l'analyse des spectres complets les échantillons ont été considérés par block.

Base de données		Prétraitement	LV	RMSEC	RMSECV	Bias	CV Bias	r^2 cal	r^2 val	RPD
Face cuir n=55	Spectres complets n_s=2089	Brut	2	4.605	4.609	-0.02	0.20	0.01	0.01	1.00
		D1	3	3.806	3.809	-0.01	-0.01	0.32	0.32	1.21
		D2	4	3.626	3.631	-0.01	-0.01	0.38	0.38	1.27
		SNVD	5	3.701	3.706	0.00	0.00	0.36	0.36	1.25
		MSC	5	3.755	3.765	0.00	0.00	0.34	0.34	1.23
	Spectres moyens n_s=55	Brut	3	4.401	4.662	-0.13	-0.15	0.01	0.03	1.05
		D1	5	2.576	2.888	-0.01	-0.01	0.66	0.57	1.79
		D2	6	2.498	3.272	0.00	0.00	0.68	0.46	1.85
		SNVD	5	2.691	3.115	0.00	0.03	0.63	0.51	1.71
		MSC	5	2.899	3.322	0.00	-0.02	0.57	0.44	1.59
Face graisse n=54	Spectres complets n_s=2052	Brut	4	3.647	3.647	0.00	0.00	0.35	0.35	1.26
		D1	3	3.599	3.599	-0.01	-0.01	0.36	0.36	1.28
		D2	3	3.543	3.544	-0.01	-0.01	0.38	0.38	1.30
		SNVD	3	3.557	3.632	0.00	0.00	0.38	0.38	1.30
		MSC	4	3.557	3.544	0.00	0.00	0.38	0.38	1.30
	Spectres moyens n_s=54	Brut	5	3.254	3.588	-0.03	-0.14	0.44	0.33	1.42
		D1	3	3.210	3.404	-0.04	-0.08	0.46	0.39	1.44
		D2	3	3.089	3.318	-0.05	-0.13	0.50	0.42	1.49
		SNVD	4	3.144	3.464	-0.01	-0.11	0.48	0.38	1.47
		MSC	4	3.173	3.500	0.00	-0.09	0.47	0.36	1.45
Face cuir et graisse n=55	Spectres complets n_s=4141	Brut	7	3.803	3.806	-0.01	-0.01	0.31	0.31	1.21
		D1	4	3.550	3.552	-0.01	-0.01	0.40	0.40	1.30
		D2	5	3.442	3.446	0.00	0.00	0.43	0.43	1.34
		SNVD	6	3.535	3.538	0.00	0.00	0.40	0.40	1.30
		MSC	6	3.788	3.791	0.00	0.00	0.31	0.31	1.22
	Spectres moyens n_s=109	Brut	10	2.676	3.009	-0.01	-0.09	0.63	0.54	1.72
		D1	4	3.034	3.282	-0.03	-0.09	0.52	0.45	1.52
		D2	4	3.279	3.536	-0.02	-0.06	0.44	0.36	1.41
		SNVD	6	3.109	3.327	-0.01	-0.04	0.50	0.43	1.48
		MSC	9	2.659	2.952	0.00	-0.05	0.63	0.55	1.73

(e) Bilan pour la prédiction des analyses de référence

D'après cette étude, il est possible de caractériser un échantillon de couenne en mouvement sur un tapis de convoyage en laboratoire. Indépendamment des différentes bases de données utilisées, le taux de matière sèche et le taux de matière grasse présentent les meilleures capacités de prédiction. Le taux de protéine présente en général une capacité de prédiction légèrement inférieure au taux de matière sèche et de matières grasse. Pour ces analyses de référence, il est possible d'imaginer une application industrielle en ligne. En revanche le collagène possède une capacité de prédiction inférieure aux autres analyses de références. Ceci s'explique principalement par la robustesse de l'analyse de référence.

Le dosage du collagène repose sur la quantification de l'hydroxyproline par une réaction d'oxydo-réduction, il est possible que la méthode d'analyse ne réagisse pas spécifiquement avec l'hydroxyproline, mais également avec les molécules présentes dans l'hypoderme, car cette méthode est initialement adaptée au dosage dans les muscles. Cette analyse ne sera pas considérée comme fiable pour la quantification du collagène sur des couennes.

Les meilleurs modèles de prédiction ont été observés sur les bases de données avec les spectres face graisse, suivies des bases de données avec les spectres face cuir et graisse. C'est pour la face cuir que les moins bonnes capacités de prédiction ont été observées. Si la base de données avec les spectres cuir et graisse possède généralement une capacité supérieure à la base de données face cuir, c'est principalement dû à l'information apportée par les spectres de graisse. En général, les spectres face cuir présentent des capacités d'absorption très inférieures par rapport aux spectres face graisse. C'est probablement dû au fait que la peau joue un rôle de protection notamment contre la lumière, ce qui aurait pour conséquence de réfléchir une partie importante de la lumière en limitant les phénomènes absorption. Ceci a pour effet de limiter l'information présente sur les spectres issus de la face cuir. En considérant les trois couches qui constituent la couenne de porc (derme, mésoderme et hypoderme), on s'attend à mieux prédire les protéines sur la face cuir et mieux prédire les matières grasses sur la face graisse. Or les protéines sont mieux prédites sur la face graisse et le taux de graisse est aussi bien prédit sur la face cuir que sur la face graisse. Les analyses de références présentent entre elles des corrélations importantes, c'est donc probablement par corrélation indirecte que les modèles arrivent à prédire les différentes analyses de référence.

L'un des objectifs de ce chapitre était d'étudier l'effet des spectres moyennés sur la capacité de prédiction. En effet, les échantillons sont acquis en mouvement, ce qui a pour conséquence d'avoir un ensemble de spectres pour décrire un seul échantillon. Cet ensemble de spectres possède une variation spectrale. Dans un premier temps, on a considéré les échantillons en bloc pour réaliser les modèles de prédiction, puis les modèles ont été recalculés à partir du spectre moyen. Une augmentation significative de la capacité de prédiction a été observée pour les modèles développés à partir des spectres moyens. Cela signifie que la variabilité de la couenne est prise en compte lorsque l'on moyenne les spectres qui lui sont associée.

Dans le cadre d'une analyse en ligne d'un produit présentant une hétérogénéité importante, il est donc plus intéressant de cumuler les informations spectrales d'un lot et de les moyenner afin de prendre en compte cette variabilité.

II.3.1.(4). Sélection des prétraitements : Lambda de Wilks- Performances de l'indice de qualité du signal (SQi)

L'objectif de cette partie est d'évaluer la capacité d'un indicateur permettant de sélectionner les prétraitements des données spectrales qui permettront la meilleure capacité de prédiction. Pour cela, les données spectrales des bases multispectre et des bases moyennées ont été transformées *via* plusieurs prétraitements. Les prétraitements D1, D2, SNV, SNVD, MSC ont été réalisés sur les bases brutes (réflectance), ou après transformation soit par absorbance, par centrage moyen (mean centred) ou par « autoscale » (I.4.5.(2).)

Pour chacune des bases de données issues d'un prétraitement sont calculés le SQi et le coefficient de détermination du modèle de prédiction obtenu par PLSr en validation croisée. Les corrélations entre les deux paramètres ont été calculées afin de déterminer dans quelle mesure le SQi permet d'estimer la qualité du prétraitement ainsi que la pertinence de la base de données. L'objectif est de voir si cet indicateur peut permettre de quantifier la part utile des matrices spectrales pour prédire la matrice de valeurs des références et de choisir le meilleur prétraitement (Gobrecht et Bellon-Maurel, 2014).

Pour les bases multispectres (spectres non moyennés), un indice supplémentaire a été calculé, le lambda de Wilks, afin de quantifier la variation intra et inter échantillons. Plus la valeur de cet indice est élevée (entre 0 et 1), plus il est possible de différencier les spectres appartenant à un même échantillon.

Indirectement, il est supposé, comme pour le SQi, que les corrélations avec les coefficients de détermination des modèles de prédiction calculés sur les bases de données après divers prétraitements, pourront permettre de sélectionner les prétraitements les plus efficaces.

Le Tableau 13 présente les coefficients de corrélation de Pearson entre les indices SQi et Wilks et les r^2 des analyses de références en fonction de différents prétraitements. La démarche de calcul est présentée dans la Figure 30.

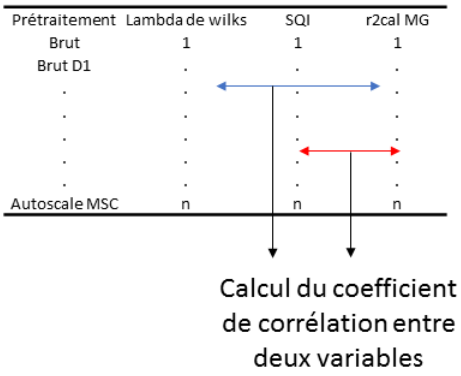


Figure 30 : Calculs des coefficients de corrélation entre les coefficients de détermination des analyses de référence par PLS et les indices (SQI et lambda de Wilks)

(a) Lambda de wilks

Lorsque l'on étudie le lien entre le lambda de Wilks et les r^2 de chacun des quatre modèles (sur les bases multispectre uniquement), les coefficients de corrélation sont relativement élevés pour la face cuir et la face graisse étudiées indépendamment et proche de 0 pour les faces cuir et graisse étudiées en simultanée. Ces résultats étaient prévisibles dans la mesure où la variabilité entre les faces cuir et graisse est telle qu'elle est supérieure à la variabilité entre les échantillons, pénalisant fortement le lambda de Wilks (coefficient de corrélation entre 0,3 et 0,4). Le lien entre le lambda de Wilks et le r^2 du modèle est plus ou moins élevé selon la variable considérée (taux de matière sèche, de matières grasses, de protéine et de collagène). L'homogénéité spécifique de la face cuir entraîne des corrélations avec le Wilks qui sont élevées pour le collagène et les matières grasses.

Le lambda de Wilks est sensible à la nature du prétraitement, passant de 0,06 à 0,69. Les plus hautes valeurs du lambda sont observées pour les prétraitements D2 (0,63 pour la face cuir, 0,69 pour la face graisse) et D1 (0,58 pour la face cuir et 0,67 pour la face graisse) pour les données brutes en réflectance, ce qui correspond aussi aux meilleurs modèles de prédiction.

	Spectres moyen			Spectres décomposés					
	Face cuir	Face graisse	Face cuir et graisse	Face cuir		Face graisse		Face cuir et graisse	
	Sqi	Sqi	Sqi	Sqi	Wilks	Sqi	Wilks	Sqi	Wilks
Wilks	Ø	Ø	Ø	0.753	1	0.848	1	-0.287	1
sqi	1	1	1	1	0.753	1	0.848	1	-0.287
r² MS	0.627	0.808	0.715	0.690	0.257	0.724	0.554	0.687	-0.134
r² MG	0.969	0.597	0.755	0.860	0.567	0.622	0.260	0.720	-0.149
r²Prot	0.860	0.621	0.473	0.587	0.492	0.627	0.445	0.611	0.032
r²Col	0.489	0.355	0.543	0.785	0.803	0.866	0.656	0.329	0.221

Tableau 13 : Coefficient de corrélation de Pearson entre les valeurs de SQi, Wilks et les coefficients de corrélation pour le taux de matières sèches, taux de matière grasse, taux de protéine et collagène des PLSr, pour les différents prétraitements.

(b) L'indice de qualité du signal (SQi)

Les coefficients de corrélation sont calculés entre l'indice SQi et les coefficients de détermination des modèles PLS, pour chaque analyse de référence (taux de matière sèche, de protéine, de matières grasses et de collagène). Ils sont compris entre 0,47 et 0,97 en fonction des différents prétraitements (Tableau 13). Cet indice est donc lié à la performance des modèles de prédiction et ne semble pas être affecté par la variabilité intra échantillon, contrairement à l'indice de Wilks.

Les valeurs du SQi varient entre 0,12 et 0,45 pour les bases avec les spectres complets et entre 0,12 et 0,64 pour les bases moyennées (Tableau 14). Il y une augmentation moyenne du SQi de 0,12 pour la face cuir, 0,17 pour la face graisse et 0,02 pour la face cuir et graisse entre les bases de données multispectres et moyennées. La face graisse présente une augmentation moyenne du SQi de 0,2 par rapport à la face cuir et par rapport à la face cuir et graisse pour les bases multispectres. La face graisse présente également une augmentation moyenne du SQi de 0,25 par rapport à la face cuir et de 0,35 par rapport aux faces cuir et graisse pour les bases moyennées. Les valeurs de SQi permettent donc de mettre en évidence l'amélioration de la capacité de prédiction des modèles entre les données multispectres et les données moyennées, ainsi que l'effet de la face de la couenne lors de l'acquisition spectrale.

Il n'y a pas d'amélioration notable pour les calculs du SQi par les transformations en absorbance, « autoscale » et centrage moyen. Les indices les plus élevés sur les différentes bases ont été observés pour la dérivée première. L'ensemble de ces observations sont en adéquation avec les résultats présentés sur la capacité de prédiction pour les différentes analyses de référence.

Tableau 14 : Valeurs des indices du SQi pour les prétraitements D1, D2, SNV, SNVD et MSC sur les données en réflectance. Ici le SQi a été calculé avec les quatre analyses de référence.

	Spectres moyen			Spectres décomposés		
Prétraitement	**Face cuir**	**Face graisse**	**Face cuir et graisse**	**Face cuir**	**Face graisse**	**Face cuir et graisse**
Brut	0.12	0.23	0.10	0.05	0.12	0.07
D1	0.39	0.64	0.25	0.21	0.45	0.20
D2	0.33	0.60	0.23	0.19	0.42	0.18
SNV	0.34	0.62	0.05	0.19	0.44	0.18
SNVD	0.34	0.62	0.23	0.19	0.44	0.18
MSC	0.35	0.62	0.23	0.20	0.44	0.19

Par ailleurs, le calcul du SQi en présence ou en absence du taux de collagène (Tableau 15), montre que ce dernier n'est pas sensible à la variabilité du dosage du collagène qui pourtant paraissait peu robuste. Il en résulte que soit, l'analyse du collagène est plus robuste que ce qu'elle paraissait, soit le calcul du SQi permet de faire abstraction d'une part non contrôlée de la variabilité du collagène.

Tableau 15 : Valeurs de SQi sur les données brutes pour les différentes bases de données, avec le taux de collagène et sans le taux de collagène.

Sqi		Avec taux de collagène	Sans taux de collagène
Spectres décomposés	Face cuir	0.053	0.049
	Face graisse	0.123	0.120
	Face cuir et graisse	0.066	0.066
Spectres moyen	Face cuir	0.122	0.121
	Face graisse	0.229	0.218
	Face cuir et graisse	0.105	0.104

Les valeurs du SQi permettent donc d'une part d'évaluer les capacités de prédiction des différentes bases de données, mais aussi de déterminer quels prétraitements seront les plus efficaces. Cet indice rassemble les caractéristiques nécessaires pour l'utiliser comme un outil d'exploration des données en chimiométrie notamment pour les travaux en spectroscopie, afin d'évaluer la capacité de prédiction en fonction des différents prétraitements sur les spectres.

Le calcul du SQi nécessite une puissance de calcul bien inférieure à la modélisation classique, ce qui permet de réaliser une exploration des données plus rapide dans le but de cibler les paramètres d'intérêt pour modéliser les données (type de prétraitement, nature des bases de données …).

II.3.1.(5). Effet du moyennage des spectres sur la capacité de prédiction des valeurs de taux de matière sèche, de matières grasses et de protéines.

Les résultats précédents ont permis de mettre en évidence l'amélioration des modèles de prédiction lorsque les spectres appartenant à un échantillon sont moyennés. Une étude a été réalisée dans laquelle les résultats des prédictions des bases multispectres ont été moyennées et comparées aux prédictions des spectres moyens. Pour cela, les prédictions de chaque spectre d'un échantillon ont été moyennées, puis les coefficients de détermination entre les valeurs de prédiction moyennées et les valeurs de référence ont été calculés.

Le Tableau 16 présente les coefficients de détermination des prédictions à partir des bases multispectres, des spectres moyennés et la moyenne des prédictions de la base multispectres. Lorsque l'ensemble des prédictions sont moyennées par échantillon sur les bases décomposées, il y a une augmentation significative (0,23 en moyenne sur l'ensemble des bases) de la valeur du coefficient de détermination par rapport aux bases multispectres. Il est important de noter que ces effets sont indépendants de la base de données et des analyses de référence utilisées. Ces résultats sont cohérents, une couenne est par définition hétérogène, moyenner les prédictions des spectres appartenant à une couenne va permettre de prendre en compte une partie de cette hétérogénéité.

Tableau 16 : Tableau de comparaison des coefficients de détermination entre les valeurs prédites et les valeurs de référence pour les bases avec les spectres complets, les spectres moyennés et la moyenne des prédictions de la base complète. MS = Matière sèche, Prot = protéine. Ici les calculs ont été réalisés sur l'ensemble des échantillons (n=57), sans prendre en compte le taux de collagène

	Face cuir			Face graisse			Face cuir et graisse		
Prétraitement	**MS**	**Prot**	**MG**	**MS**	**Prot**	**MG**	**MS**	**Prot**	**MG**
				Spectres décomposés					
Brut	0.29	0.19	0.22	0.50	0.51	0.53	0.42	0.32	0.36
D1	0.40	0.39	0.51	0.51	0.53	0.51	0.49	0.45	0.52
D2	0.45	0.41	0.51	0.53	0.57	0.55	0.44	0.38	0.44
SNVD	0.39	0.29	0.37	0.38	0.39	0.38	0.37	0.30	0.35
MSC	0.41	0.24	0.31	0.49	0.52	0.52	0.41	0.33	0.36
				Spectres moyennés					
Brut	0.45	0.26	0.33	0.74	0.67	0.67	0.63	0.41	0.51
D1	0.61	0.40	0.65	0.68	0.62	0.57	0.58	0.44	0.58
D2	0.45	0.39	0.55	0.64	0.60	0.58	0.66	0.53	0.67
SNVD	0.51	0.46	0.60	0.52	0.49	0.49	0.61	0.52	0.56
MSC	0.61	0.43	0.68	0.70	0.64	0.65	0.60	0.39	0.45
				Moyenne des prédiction					
Brut	0.47	0.32	0.39	0.78	0.72	0.73	0.71	0.57	0.65
D1	0.63	0.55	0.75	0.76	0.75	0.69	0.76	0.67	0.78
D2	0.66	0.56	0.73	0.77	0.79	0.72	0.74	0.63	0.74
SNVD	0.63	0.47	0.61	0.63	0.63	0.60	0.69	0.58	0.68
MSC	0.67	0.55	0.71	0.77	0.74	0.71	0.70	0.58	0.64

Les prédictions moyennées des bases complètes ont été comparées aux prédictions des spectres moyens, mettant en évidence une augmentation du coefficient de détermination pour les prédictions moyennées (Tableau 16). Sur l'ensemble des bases, cela représente une augmentation moyenne de 0,1 pour le coefficient de détermination. L'effet est donc indépendant des différentes bases de données ainsi que des prétraitements.

Dans le cadre de la prédiction d'échantillons hétérogènes en mouvement, un échantillon (ou un lot d'échantillons) va être caractérisé par un ensemble de spectres. L'objectif est d'obtenir la prédiction la plus représentative de l'échantillon. Pour cela, il est possible de moyenner les spectres de l'échantillon afin de prédire la valeur correspondant à ce spectre, ou de moyenner les prédictions de chaque spectre de l'échantillon. Dans les deux cas, il y a une amélioration significative du modèle de prédiction. Mais, cette étude montre une capacité de prédiction supérieure pour la moyenne des prédictions de chaque spectre de l'échantillon. Cela signifie que lorsque les spectres sont moyennés, une partie de l'information spectrale (ou hétérogénéité) est perdue dans le spectre moyen, en opposition avec la moyenne des prédictions. Pour une problématique de caractérisation d'un échantillon (lot) sur une chaîne de production, il est donc plus intéressant de moyenner les valeurs de prédiction associées à chaque spectre de l'échantillon plutôt que de moyenner les spectres.

II.4. Conclusion

Ce chapitre avait pour but d'étudier la capacité de prédiction en ligne par spectroscopie PIR des propriétés physico-chimiques d'une couenne. Elle a été réalisée dans la perspective de mettre en place un outil proche infrarouge en amont d'une chaîne de production capable de prédire le rendement d'extraction en gélatine à la sortie de la chaîne de production. Deux problématiques principales ont été identifiées : la maîtrise de l'hétérogénéité des couennes et la gestion du flux de matière.

Le rendement d'extraction d'une couenne est principalement lié à la composition de celle-ci, c'est pourquoi la composition chimique des couennes (taux de matière sèche, taux de protéine, taux de matières grasses et taux de collagène) a été mesurée. Les modèles développés pour prédire le taux de matière sèche, le taux de protéine et le taux de matières grasses par infrarouge démontrent une capacité de prédiction intéressante pour ces analyses. Pour le collagène, les résultats des modélisations ne permettent pas d'assurer une bonne capacité de prédiction, principalement en raison de la méthode de dosage, qui n'est probablement pas adaptée à l'analyse de couenne sans délipidation. Toutefois, ces essais ont été réalisés sur un faible nombre d'échantillons et en validation croisée. La performance des modèles est très probablement surestimée ; l'utilisation d'un nombre plus conséquent d'échantillons devrait améliorer les performances.

Ces travaux ont permis de mettre en évidence la variation de la capacité de prédiction en fonction de la surface d'acquisition. C'est sur la face graisse que les meilleurs résultats ont été obtenus, sur la face cuir les capacités de prédiction sont bien inférieures. Les prédictions sur les spectres provenant de la face cuir et graisse présentent des capacités de prédiction intermédiaires. Dans les conditions de la chaîne de production, les couennes sont disposées de façon aléatoire (face cuir, face graisse ou à l'interface) ne permettant pas d'obtenir uniquement des spectres face cuir ou face graisse. Ce sont donc les prédictions de la face cuir et graisse qui correspondent le mieux à la réalité. Dans l'idéal, il serait souhaitable de sélectionner les spectres provenant de la face graisse uniquement. Les résultats de cette étude montrent la capacité de classification des différentes faces, malgré la grande diversité de type d'échantillons et l'hétérogénéité au sein d'un échantillon.

En présence d'une mesure en ligne d'échantillons hétérogènes, il est nécessaire de prendre en compte un maximum de variabilité pour caractériser un échantillon (un lot). Pour un même échantillon, une grande diversité de spectres peut être observée. Les analyses réalisées ont mis en évidence l'intérêt de moyenner les valeurs de prédiction de chacun des spectres de l'échantillon, plutôt que de moyenner les spectres et prédire ce spectre moyen. Indirectement, cela signifie qu'il y a une perte d'information lorsque l'on fait la moyenne des spectres, alors que la moyenne des prédictions est plus proche de la réalité.

L'étude de l'indice SQi présentée ici, met en avant l'intérêt de cet indice dans le domaine de la chimiométrie, notamment pour une première évaluation des données avant la modélisation. Cet indice permet d'évaluer la part utile d'une matrice spectrale (X) pour expliquer une matrice des analyses de référence (Y). Les résultats ont mis en évidence la corrélation entre les performances des modèles et la valeur de cet indice. Cela a permis d'anticiper le choix des bases de données qui auront les meilleures performances (les bases moyennées, face graisse) ainsi que les prétraitements les plus intéressants (dérivée seconde), plus rapidement et sans la puissance de calcul nécessaire à la modélisation.

Chapitre III. Détermination en ligne du potentiel d'extraction de gélatine de couennes de porcs par spectroscopie proche infrarouge sur site de production

III.1. Introduction

Un essai a été réalisé en conditions industrielles sur un site de production de gélatine, exclusivement extraite à partir de peau de porc. Dans la majeure partie des cas en industrie, la qualité des matières premières est évaluée par un contrôle visuel des opérateurs au moment de leur réception (bacs de 500 kilos). Cette analyse visuelle est subjective et ne permet pas de considérer l'ensemble du lot car seule la partie visible en surface des bacs peut être évaluée. L'objectif industriel est de mettre en place un capteur qui analyse les matières premières en amont de la chaîne de production afin de qualifier ces matières en termes de rendement potentiel d'extraction de la gélatine. Une meilleure connaissance de ces matières premières pourrait permettre à l'industriel de prédire le remplissage des bassins, d'avoir une meilleure connaissance des matières premières en fonction des fournisseurs et potentiellement d'adapter le processus.

Le chapitre précédent a permis d'étudier la capacité de la spectroscopie proche infrarouge à appréhender la qualité des couennes de porc scannées individuellement et en mouvement. Il montre que la composition chimique de couennes pourrait être prédite par spectroscopie proche infrarouge en ligne. Lors de cette étude, la principale difficulté rencontrée était due à la caractérisation de couennes présentant une variabilité intra et inter échantillons importante. Certaines propositions ont été faites afin de prendre en compte au maximum cette hétérogénéité. L'étude suivante, s'inspire de l'approche dans laquelle l'ensemble des spectres d'une couenne sont considérés comme un échantillon, mais cette fois l'échantillon est en réalité un lot de plusieurs couennes.

Dans ce chapitre, une démarche expérimentale à l'échelle industrielle a été mise en place. L'objectif a été de caractériser un lot de couennes en termes de rendement potentiel d'extraction de gélatine. Afin de mieux comprendre les problèmes liés à cette échelle d'étude, les travaux ont pris en compte les contraintes industrielles de la société Weishardt.

III.1.1.(1). La chaîne de production

L'usine de production, basée sur le site de Graulhet est équipée d'une chaîne de production qui transforme le collagène de couennes de porc en gélatine par traitement acide. Cette usine fonctionne en continu (24h/24 toute la semaine) et extrait environ 10 lots de 75 tonnes de couennes par semaine pour un rendement d'extraction moyen de 17,1%. Mais actuellement, les valeurs de rendement sur 75 tonnes de matières premières peuvent varier entre 14 et 22%. Ces variations sont considérables quand on considère le volume de matière première traitée.

(a) Conditionnement des matières premières :

Des matières premières en provenance de toute l'Europe sont livrées sous forme fraîche ou congelée quotidiennement sur le site de production. En général, les couennes livrées sous forme fraîche sont directement stockées dans un hall de matières premières dans lequel elles peuvent rester jusqu'à 24h avant de commencer le processus de production (Figure 31). Les couennes conditionnées congelées peuvent soit directement rejoindre le hall de matière première, soit être stockées pendant plusieurs mois dans un hangar de congélation en fonction des arrivages en matières premières.



Figure 31 : Bacs de 500 kilos de couennes stockés dans le hall matières premières.

Au moment de la livraison, les couennes sont conditionnées dans des bacs et des cartons de 500 kilos. Les lots sont identifiés en fonction du fournisseur, du pays de provenance, du conditionnement (« fraîche » ou « congelée »), du type d'animal (coche ou porc), du type de production (conventionnel ou biologique) et de la partie de l'animal (bardière ou couennes autre). Il est difficile pour les industriels de savoir si le type de fournisseur, la provenance, le type de production ou encore le conditionnement ont un effet sur le rendement d'extraction. En revanche, le type d'animal et la partie de l'animal ont un effet sur le rendement d'extraction final. En effet, un porc destiné à la production de viande vit environ 7 mois avant l'abattage, sa peau est donc peu réticulée et présente peu de graisse. La coche est en revanche une truie qui a été mise à la reproduction, elle peut vivre de nombreuses années, sa peau est beaucoup plus réticulée et elle présente une composition en graisse importante, rendant l'extraction du collagène plus difficile.

En plus des variations dues à l'animal, la qualité des couennes peut varier en fonction de la qualité de la découpe dans les abattoirs (plus ou moins de graisse, et dans certains cas de muscle), ou encore de la qualité du conditionnement, du temps l'abattage de l'animal à la livraison.

(b) Gestion de la matière première le long du procédé industriel :

Lorsque les couennes sont destinées à l'extraction, qu'elles soient « réfrigérées » ou « congelées », elles sont stockées dans le hall de matières premières sous forme de lot de 500 kilos. Les blocs de 500 kilos congelés sous forme cubique sont tranchés sur toute la hauteur, les couennes tranchées sont ensuite réceptionnées dans une vis-sans fin. Les couennes sous forme fraîche sont aussi coupées par une trancheuse avant de rejoindre cette même vis-sans fin. Les matières fraîches et congelées ne sont pas introduites simultanément sur la chaîne de production.

En sortie de la vis sans fin, les couennes passent sur un tapis de convoyage pour rejoindre un broyeur et ressortir sur une nouvelle vis-sans fin, puis sur un bras de distribution. Le broyage permet de réduire les couennes en lambeaux de plus petites tailles.

Le bras de distribution permet ensuite de répartir la matière première dans trois bassins de prétraitement (lavage) (Figure 32). Chacun des bassins va être chargé avec 25 tonnes de matières premières, pour une série complète de 75 tonnes.



Figure 32 : Bassins de lavage en début de chaîne de production

Ces 75 tonnes de matières premières correspondent à un ordre de fabrication. Les durées de lavage et prétraitement des différents bassins dépendent des matières premières et des qualités de gélatine attendues par l'industriel. Après ces étapes, les 25 tonnes de matières provenant d'un bassin sont envoyées dans un des trois réacteurs d'extraction de la chaîne (Figure 33). Une première extraction à faible température est réalisée (≈ 55°C), puis la phase liquide est récupérée et la phase solide subit une nouvelle extraction, et ce jusqu'à 3 ou 4 extractions successives (Figure 33). Suite aux extractions successives, la masse des couennes est fortement diminuée, c'est pourquoi après certaines extractions trois réacteurs sont réunis afin de subir une dernière extraction. Les différentes phases liquides sont ensuite traitées en plusieurs étapes de raffinage : filtration, déminéralisation, concentration, stérilisation, séchage et extrusion.



Figure 33 : Réacteur d'extraction

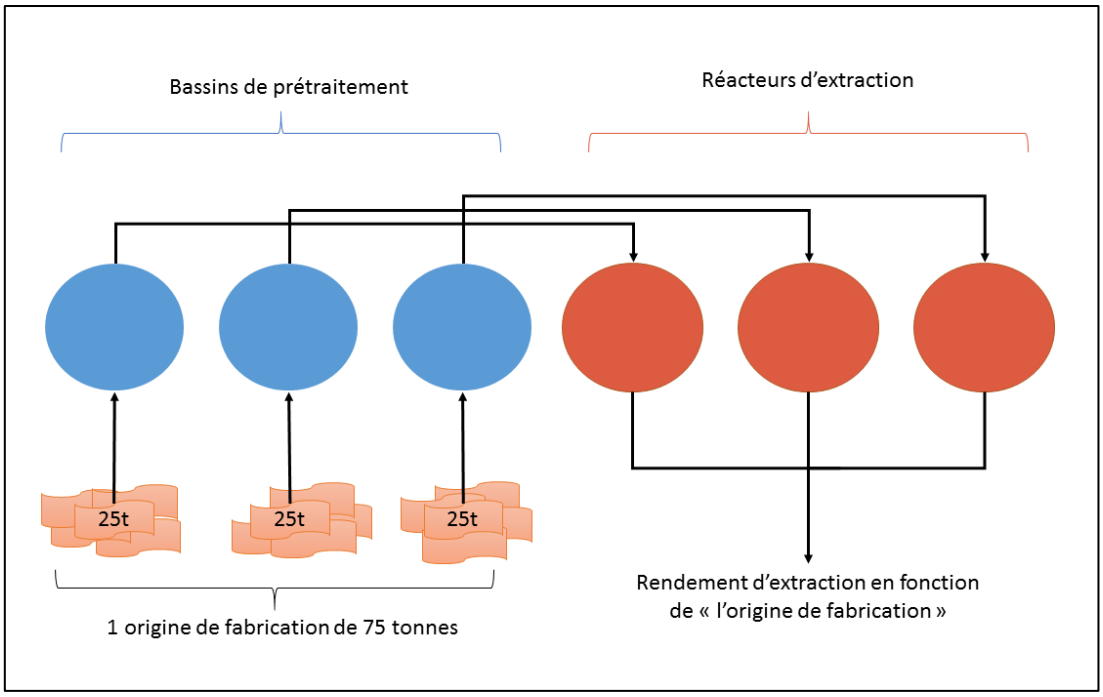


Figure 34 : Schéma du circuit des couennes en fonction de l'ordre de fabrication

Au sein de la chaîne de production il existe différentes technologies de raffinage qui dépendent des propriétés de la gélatine recherchées par l'industriel. A la fin du processus, la gélatine est sous forme de grain (Figure 35) ou de poudre (en fonction du type de séchage).



Figure 35 : Gélatine en sortie d'extrusion

L'entreprise estime le rendement correspondant à un ordre de fabrication (75 tonnes de couennes) de deux façons différentes : soit en déterminant le rendement à l'égouttage, soit en déterminant le rendement en sortie de séchoir. Le rendement à l'égouttage consiste à mesurer l'indice de réfraction de la phase liquide afin de déterminer la concentration en gélatine en fonction du volume de cette solution. Cette mesure est réalisée ponctuellement en sortie du réacteur d'extraction. Le rendement peut aussi être calculé par rapport de masse en sortie de séchoir. Mais comme indiqué précédemment, dans les faits, il y a souvent des mélanges après l'extraction, rendant difficile le suivi du rendement. Le rendement d'un ordre de fabrication est ensuite calculé avec un facteur de correction théorique, qui correspond à un ratio entre le rendement estimé par réfractométrie en sortie de cuve et le rendement en masse déterminé en sortie de séchoir.

III.1.1.(2). Démarche expérimentale

L'objectif est d'évaluer la capacité à estimer le rendement d'extraction de gélatine sur des matières premières en amont de la chaîne de production, par spectroscopie proche infrarouge. Avec les contraintes industrielles, il est impossible d'appliquer les conditions d'échantillonnage de laboratoire. En effet dans le processus industriel, sur le tapis de convoyage, les couennes ne sont pas bien séparées et ne sont pas positionnées à la même hauteur et sur une face. Pour l'application industrielle, l'hypothèse suivante a été faite, le rendement d'extraction est lié à la composition chimique d'une couenne.

Le chapitre précédent met en évidence la capacité d'appréhender la composition d'un échantillon en mouvement en réalisant une moyenne des spectres et en prédisant le spectre moyen. Pour cette application à grande échelle, les spectres d'un flux de 75 tonnes de couennes ont été considérés comme un échantillon (en moyennant les spectres) et le rendement d'extraction de ces couennes comme l'analyse de référence (Figure 36).

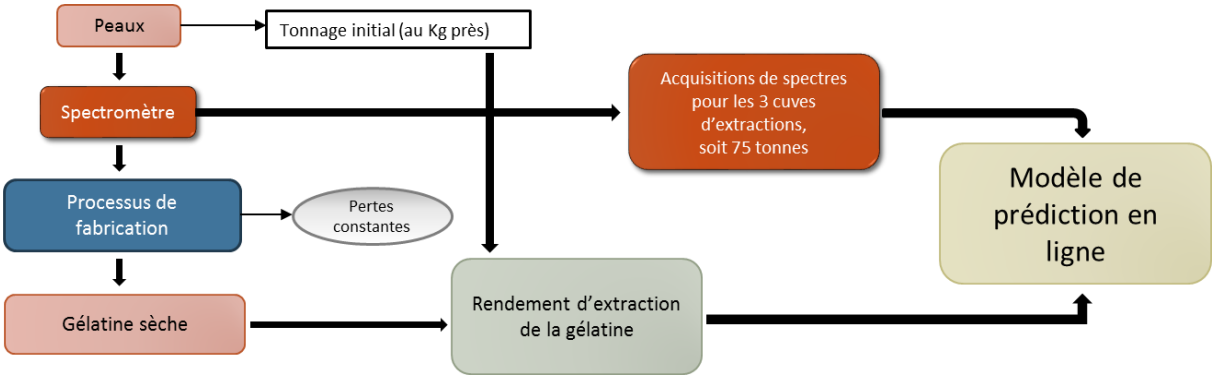


Figure 36 : Schéma de la démarche expérimentale mise en place.

III.2. Matériels et méthodes

III.2.1.(1). Positionnement de la sonde du spectromètre proche infrarouge sur la chaîne de production

La chaîne de production est configurée de telle façon qu'il est uniquement possible d'obtenir la valeur des rendements en gélatine extraite pour un processus complet, ce qui représente 75 tonnes de couennes de porc. Afin de s'affranchir le plus possible des problèmes liés à la température des couennes, le spectromètre est placé au niveau du bras de distribution (Figure 37) des différents bassins de lavage, car les couennes ont subi un broyage à ce stade de la production, et les couennes congelées ou fraîches ont théoriquement atteint une température proche grâce à ce broyage.

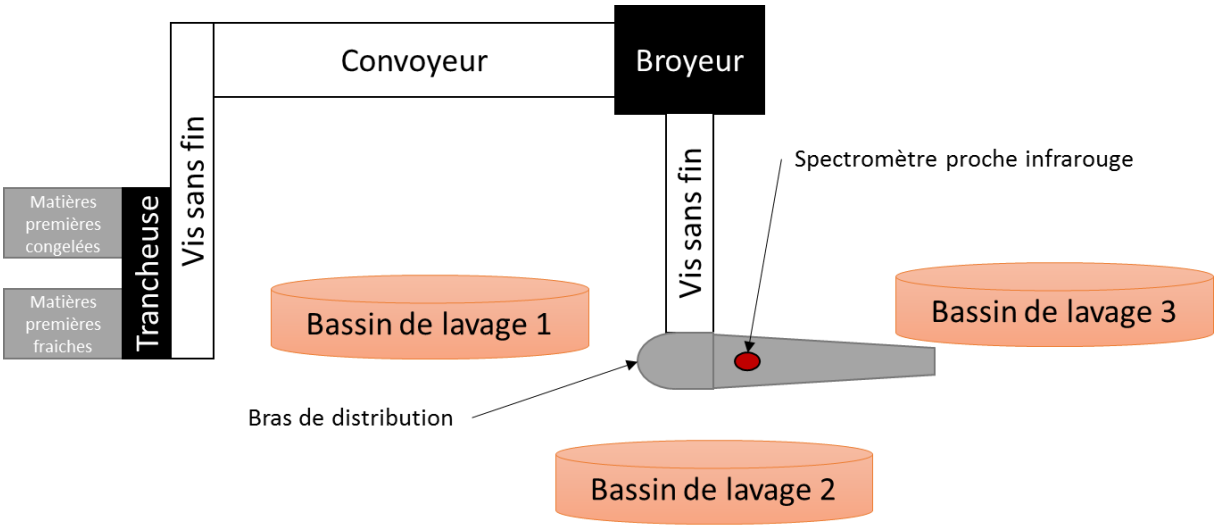


Figure 37 : Schéma de la position du spectromètre dans le hall de répartition des matières premières

Une plateforme a été aménagée au niveau du bras de distribution des matières premières. Elle comprend un support anti-vibration pour le spectromètre (Figure 38), des prises pour l'alimentation électrique des appareils, une connexion internet raccordée au poste de commande et une fenêtre pour positionner la sonde sur le bras de distribution. De plus, une station d'enregistrement de l'humidité relative et de la température a été installée à proximité du lieu d'acquisition.



Figure 38 : Spectromètre Polytech dans le support anti-vibration (photo de gauche), sonde déportée sur le bras de distribution (photo de droite).

Pour des raisons techniques, en lien avec l'accès de l'opérateur à la sonde, la sonde est positionnée au début du bras de distribution. Le bras de distribution est un tube cylindrique métallique de diamètre interne d'environ 50 cm (Figure 39). Une fenêtre carrée adaptée au diamètre de la sonde a été réalisée sur le haut du bras de distribution. La sonde a ensuite été fixée sur des rails coulissants afin de pourvoir la décaler pour la rendre accessible à l'opérateur pour un nettoyage, car quotidiennement la sonde peut recevoir de faibles projections qui peuvent à terme induire un biais. Les nettoyages ont été réalisés quotidiennement par les opérateurs de la chaîne.

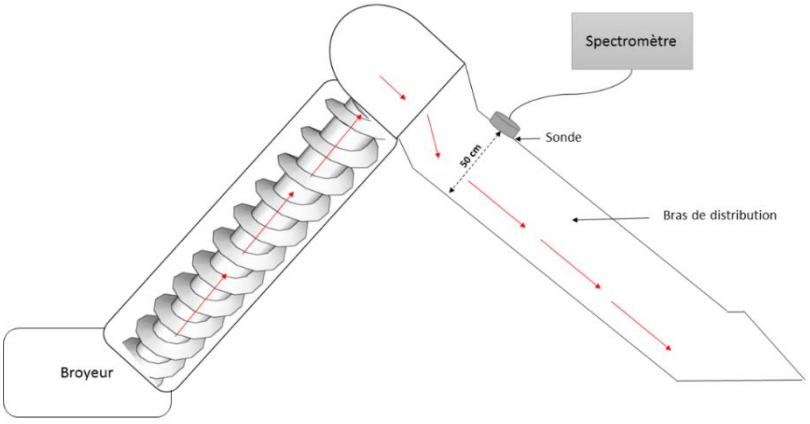


Figure 39 : Schéma du positionnement de la sonde déportée sur le bras de distribution

III.2.1.(2). Conditions de température et humidité dans le hall des matières premières

Une sonde a été placée (Testo© 175 H1) proche du broyeur dans le hall ou les couennes sont réparties dans les différents bassins. Les températures (°C) et l'humidité relative (%) ont été enregistrées toutes les heures durant 12 mois (1er mai 2015 au 31 avril 2016). La Figure 40 présente l'évolution de ces données.

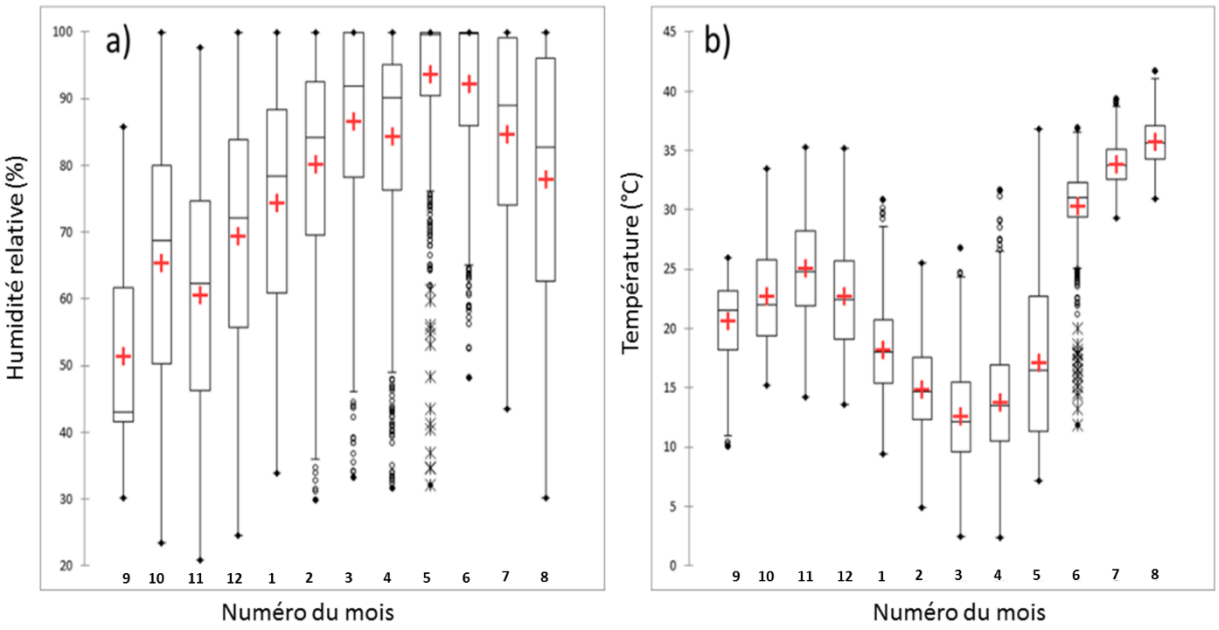


Figure 40 : Evolution de l'humidité relative (a) et de la température (b) dans l'usine de Weishardt (Graulhet), sur 12 mois (mai 2015 à avril 2016)

Le hall de production est un hangar mal isolé, qui subit d'importantes variations de température et d'humidité en fonction des saisons et quotidiennement. Cette année, la température a varié entre 2,4°C et 44°C, pour une moyenne annuelle de 23,3°C. L'humidité relative a varié entre 20,8% et 99,9%, avec une moyenne annuelle de 77%. Mais, le capteur était proche d'une vanne laissant échapper de la vapeur d'eau et il est possible que les valeurs d'humidité soient en réalités inférieures dans cette zone. Les variations sont tout de même très importantes.

III.2.1.(3). Acquisition spectrale en ligne des spectres de couennes de porc, avant chute dans le bassin de remplissage.

Les spectres sont acquis à l'aide du logiciel PSS-SERVER (Polytec GmbH, 763337 Waldbronn Germany) en mode *on-line* et d'un spectromètre Polytec (PSS-2121, Polytec GmbH, 763337 Waldbronn Germany), équipé d'une fibre optique de 2m et d'une tête de mesure à distance (PSS-H-A03, Polytec GmbH, 763337 Waldbronn Germany). Cette tête de mesure a la particularité de posséder un spot d'échantillonnage de 30 mm de diamètre. L'acquisition est réalisée en réflectance entre 1100 et 2100 nm avec un pas de 2 nm et un temps d'intégration de 0,2 ms. Les spectres de couennes sont acquis à une distance variant entre 35 cm et 45 cm (en effet les couennes peuvent être entassées) et les spectres de vide sont acquis à 50 cm.

Le spectromètre utilisé pour cette application en ligne est un spectromètre à barrette de diodes, qui présente l'avantage d'être très rapide et robuste par l'absence de balayage et de pièce mobile au niveau du détecteur.

Les spectres ont été acquis sur la chaîne de production entre le 15 juin et le 8 septembre 2017. Afin de faciliter l'acquisition, un « trigger informatique » a été programmé afin de lancer l'acquisition spectrale au moment où l'opérateur scanne le premier lot de couenne. A la fin du chargement de ces couennes, l'opérateur désactive l'acquisition à l'aide d'un bouton poussoir installé à cet effet.

III.2.1.(4). Tri des spectres de 75 tonnes de couennes

La chaîne de production est alimentée séquentielle en matière première par des opérateurs et les temps de chargement des bassins (25 tonnes) peuvent prendre plusieurs heures. Le flux de matière qui va traverser le bras de distribution alimentant les cuves lavages est discontinu. L'étude menée en laboratoire sur un tapis pilote a montré la capacité de discrimination des spectres provenant du tapis et de la matière première.

Un modèle de filtrage des spectres a été développé afin d'éliminer les spectres provenant de l'espace vide entre deux fragments de couennes. Ce modèle se base sur la comparaison de la similarité par rapport à un spectre moyen, via l'indice de cosinus, qui permet de calculer la similarité entre un spectre de référence et un nouveau spectre en déterminant l'angle du cosinus entre eux. Cela consiste en fait à calculer le coefficient de corrélation entre les deux spectres.

Le modèle de tri des spectres mis en place sur la chaîne de production a été programmé par le logiciel PSS-SERVER (Polytec GmbH, 763337 Waldbronn Germany). Afin de réaliser plus rapidement le tri des spectres, le modèle utilise la longueur d'onde à 1250 nm au lieu de se baser sur l'ensemble du spectre, car cette zone correspond aux secondes harmoniques des liaisons C-H des matières grasses. Le seuil d'acceptabilité du coefficient de corrélation a été fixé à 0,985 (les combinaisons de toutes les longueurs d'onde et de seuil d'acceptabilité ont été testés par itération). La Figure 41, présente un exemple de sélection des spectres par la mesure du cosinus.

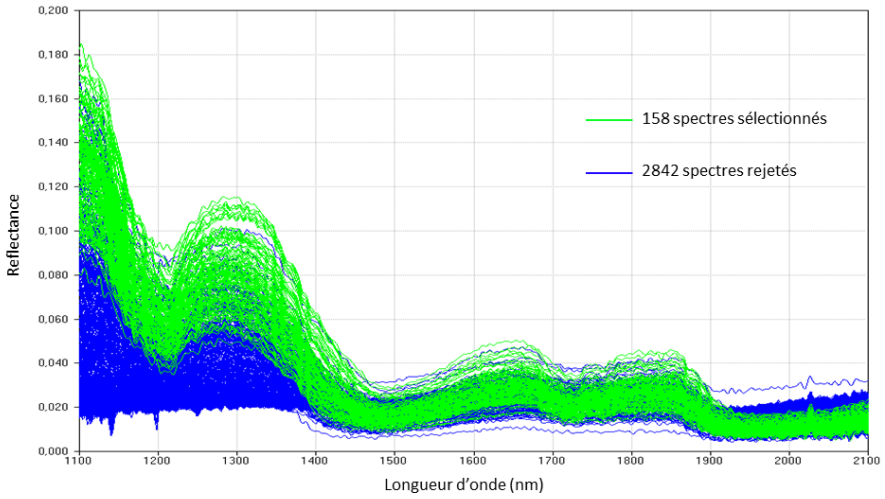


Figure 41 : Exemple de sélection de spectres sur 3000 spectres acquis en conditions industrielle. En vert les spectres de couenne, en bleu les spectres de vide ou d'interface vide-couenne.

Pour l'acquisition des données industrielles, les spectres de couennes sont enregistrés dans des sous fichiers correspondants au jour et à l'heure d'acquisition. Les spectres qui sont rejetés par le modèle sont, quant à eux, supprimés en temps réel.

III.2.1.(5). Rendement d'extraction en gélatine, à partir de 75 tonnes de couennes

Comme décrit précédemment, les rendements calculés par l'entreprise Weishardt sont en réalité calculés avec un facteur de correction pour prendre en compte les biais introduits au moment de l'extraction. Sur les 75 tonnes de matières premières extraites, l'industriel estime par expérience que la précision du rendement est d'environ ±0,5%. Cependant, par expérience, les industriels estiment que les valeurs de rendement inférieures à 18% sont beaucoup plus précises que les valeurs supérieures à 18%. Pour déterminer le rendement en gélatine extraite, la concentration en gélatine est mesurée par réfractométrie en sorti de chaque cuve d'extraction, une mesure du volume de ces cuves permet de remonter à la quantité de gélatine extraite. Cette mesure présente cependant trois niveaux d'imprécision, la représentativité de la mesure de la concentration, l'erreur de la mesure par réfractométrie et l'erreur liée à la mesure du volume.

III.2.1.(6). Echantillonnage

Une campagne d'acquisition de données a été réalisée entre le 15 juin et le 8 septembre 2017. Sur cette campagne les spectres et les rendements d'extraction correspondant à 71 échantillons de 75 tonnes de matières premières ont été mesurés. Près de 410 000 spectres ont été acquis avec une moyenne de 5269 spectres par échantillon (min : 1348, max 18539).

Le rendement moyen d'extraction des 71 échantillons est de 17,8%, le rendement le plus faible est de 14,2% et le rendement le plus élevé de 21,8%. Or, d'après l'industriel, les valeurs de rendement supérieures à 18% sont moins fiables que les valeurs inférieures à 18%. Aussi, deux bases de données seront considérées pour les travaux, la première avec l'ensemble des échantillons, la seconde avec les échantillons dont le rendement est inférieur à 18%. Cette deuxième base comprend 39 échantillons, pour une valeur moyenne de rendement de 16,7%, un rendement maximum de 17,9% et un rendement minimum du 14,2%.

III.2.1.(7). Les spectres des 71 lots de 75 tonnes de couennes

Un lot d'extraction de 75 tonnes de couennes de porc est caractérisé par un ensemble de spectres. Les spectres de chaque lot ont été moyennés, pour obtenir 71 spectres moyens. La Figure 42 représente les spectres moyens collectés pour les 71 lots. En comparaison aux spectres obtenus en laboratoire, avec une sonde placée à une faible distance, les spectres obtenus ici avec une distance entre les couennes et la sonde supérieur, apparaissent légèrement bruitées.

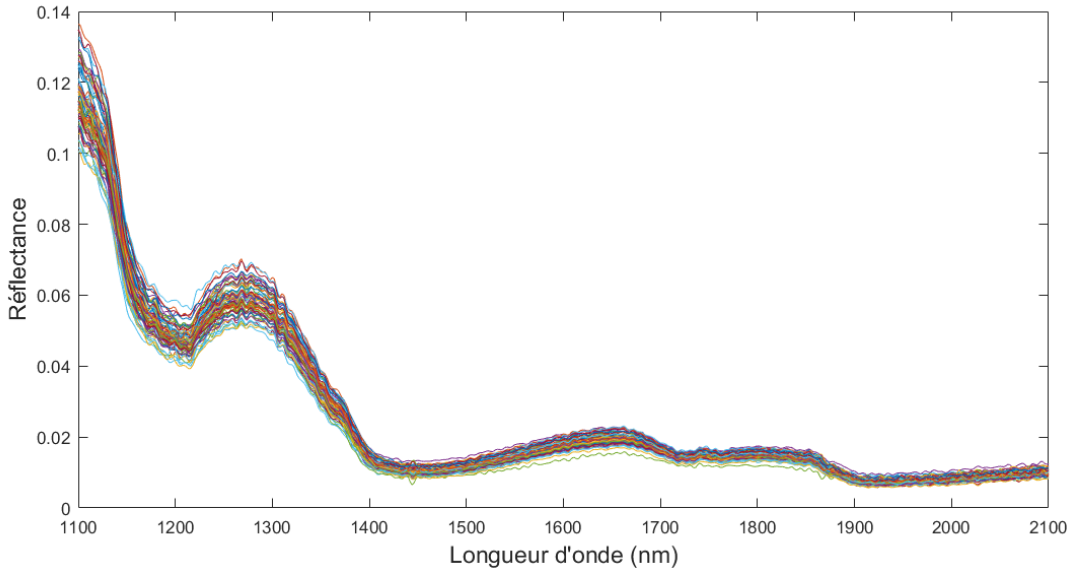


Figure 42 : Spectres proche infrarouge moyens des 71 échantillons de 75 tonnes de couennes

Une analyse en composantes principales a permis de vérifier que les spectres moyens (des 71 lots) ne comportaient pas d'aberrants spectraux (tous les échantillons présents dans un intervalle de confiance de 95%). La Figure 43 représente la répartition des spectres sur les deux premières composantes (95% de la variabilité totale).

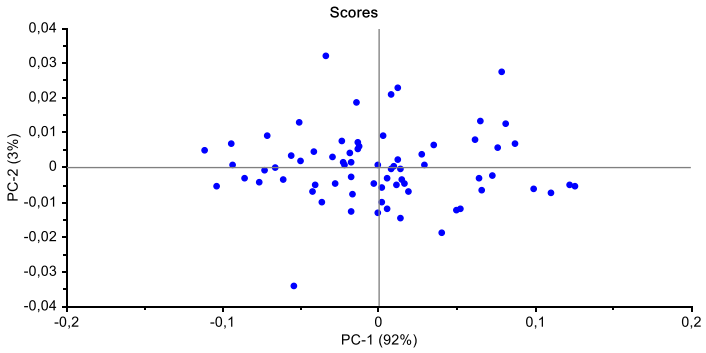


Figure 43 : Analyse en composantes principales des spectres moyens des 71 échantillons

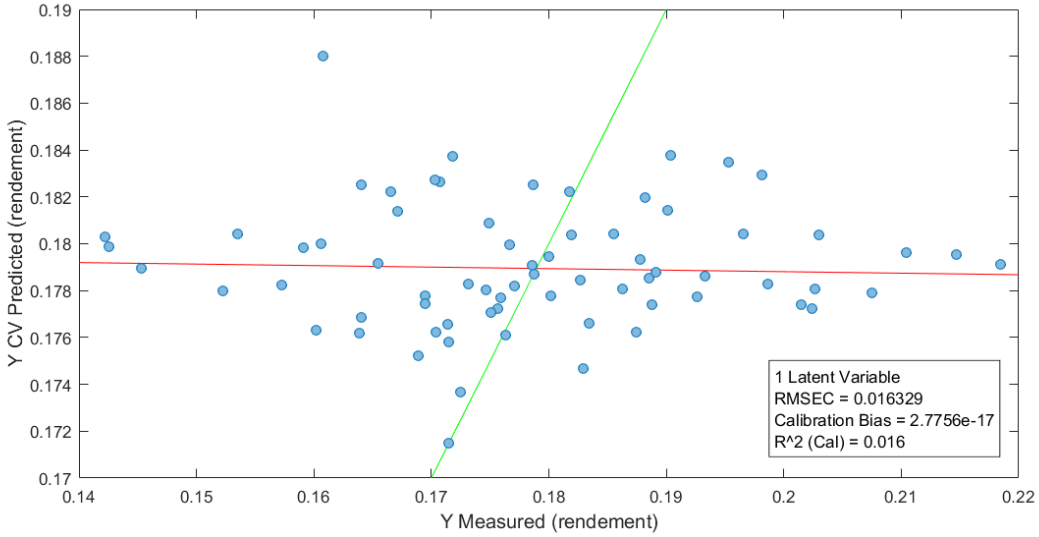
III.1. Résultats et discussion

III.1.1.(1). Les modèles de prédiction

Différents modèles par régression PLS ont été développés à partir des spectres moyens des 71 lots (*i.e.* 71 lots de 75 tonnes de couennes) afin d'évaluer la capacité de prédiction du rendement. Un premier traitement a été réalisé sur les 71 échantillons couvrant la gamme de rendement de 14-22%, puis un deuxième traitement sur les 39 échantillons qui couvrent la gamme de rendement de 14-18%. Les données ont été traitées en validation croisée après application de différents prétraitements : absence de prétraitement, lissage et dérivée première (D1), dérivée seconde (D2) de Savitzky-Golay, SNV, SNVdetrend, et MSC.

La Figure 44 présente le modèle de prédiction du rendement ainsi que les performances sur les données brutes. Quel que soit le prétraitement utilisé aucun modèle ne présente des performances de prédiction satisfaisantes (valeur du coefficient de corrélation en calibration < 0,2). Les modèles ont également été développés sur les échantillons compris entre 14% et 18%, mais les performances ne sont pas meilleures. Ces résultats montrent l'incapacité à prédire quantitativement le rendement industriel de production de gélatine à partir de ce jeu de données. Ces résultats sont en opposition avec les résultats obtenus à partir de la composition de la couenne dans les essais en flux en laboratoire. Ces contradictions peuvent s'expliquer par deux hypothèses : soit les conditions de la mesure spectrale sur site industriel entrainent une perte d'information spectrale, soit la variable de rendement tel que proposée par les industriels manque de précision. Afin de conforter l'une ou l'autre des hypothèses, le travail s'est donc orienté vers l'utilisation de modèles de classification.

Figure 44 : Modèle PLS sur le rendement pour les 71 échantillons (données brutes)



III.1.1.(2). Les modèles de classification des échantillons vis-à-vis du rendement en gélatine de 75 tonnes de couennes par spectroscopie proche infrarouge.

Comme pour les modèles de prédiction, les traitements ont été réalisés sur les échantillons de la gamme 14-22% (n=71). Pour la classification supervisée,3 ou 4 classes ont été créées soit arbitrairement en fonction de la gamme de rendement à couvrir (amplitude constante), soit à partir d'effectifs constants.

Tableau 17 : Répartition des classes pour les analyses réalisées avec 4 classes.

4 Classes		**A**	**B**	**C**	**D**
Par amplitude constante de rendement (**1**)	Gamme de rendement	14-16%	16-18%	18-20%	20-22%
	Nombre d'échantillons	7	32	23	9
Par effectif constant par classe (**2**)	Gamme de rendement	14,2-16,9%	16,9-17,7%	17,7-18,9%	18,9-21,84%
	Nombre d'échantillons	17	18	18	18

Les modèles développés avec 4 classes et une amplitude de rendement constante (Tableau 17 1), permettent d'obtenir des catégories de rendement d'intérêt pour l'industriel.

Cependant le nombre d'échantillons est différent dans chaque classe et les classes intermédiaires (B et C) sont surreprésentées par rapport aux classes extrêmes (A et D). Les modèles développés avec des effectifs constants (Tableau 17 2) présentent l'avantage de posséder le même poids pour chaque classe. En revanche les différentes classes ne couvrent pas une amplitude constante, les classes A et B couvrent plus de 2% de la gamme alors que les classes B et C couvrent respectivement 0,8% et 1,2% de la gamme.

La Figure 45 représente les spectres bruts moyens de chaque classe de rendement A, B, C, et D. C'est principalement dans la zone 1100-1300 nm que les spectres moyens se différencient. Cette zone peut être attribuée aux seconde et troisième harmoniques des liaisons C-H des lipides (Kalinin et Krasheninnikov, 2017). Ainsi, les principales différences observées sur les spectres moyens sont certainement liées à la graisse des couennes.

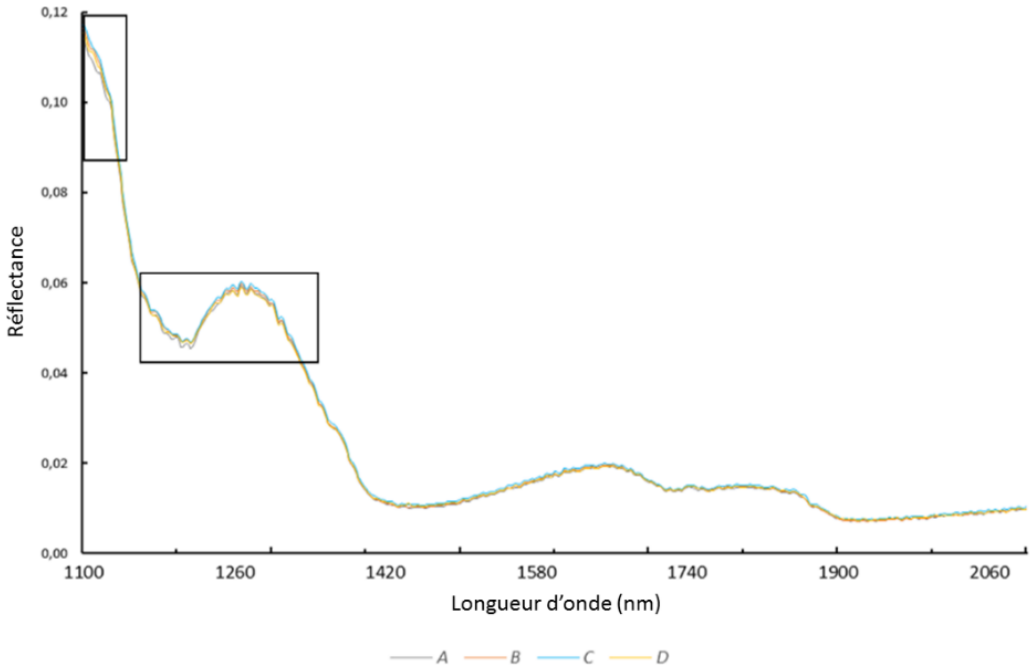


Figure 45 : Spectres moyens des 4 classes pour les 71 échantillons

Afin d'explorer les données, des essais de classification supervisée ont été réalisés par Analyse Discriminante (DA), soit par couplage Analyse en Composantes Principales couplée (ACP-DA), soit par couplage de la méthode des moindres carrés partiels (PLS) à l'Analyse Discriminante (PLS-DA) ou encore les arbres de classifications. Seuls les modèles ACP-DA et PLS-DA seront détaillés.

(a) Les modèles de classification supervisée (4 classes)

(a) i. Analyse en Composante Principale couplée à l'Analyse Discriminante (ACP-DA)

Les ACP-DA ont été réalisées sur les spectres des 71 échantillons après différents prétraitements : sans prétraitements, dérivée première (D1), dérivée seconde (D2), Standard Normal Variate Detrending (SNVD) et Multiplicative Scatter Correction (MSC). Les modèles ont été développés avec 61 échantillons en calibration et 10 échantillons en validation. Les échantillons destinés à la validation ont été choisis aléatoirement dans chacune des classes de façon à représenter la gamme de rendement en gélatine des 61 échantillons. Le Tableau 18 présente les performances des modèles DA pour les différents prétraitements et les deux types de regroupements différents pour les classes (amplitude constante ou effectif constant). Les 7 premières composantes ont été retenues pour la DA.

Tableau 18 : Performances de l'analyse discriminante en calibration, validation croisée et validation externe pour les classes à amplitude constante (1) ou les classes à effectif constant (2).

	Classes par amplitude constante de rendement (1)					**Classes par effectif constant (2)**				
	Brut	D1	D2	SNVD	MSC	Brut	D1	D2	SNVD	MSC
	Calibration (61 éch.)									
Prédiction de A correcte (%)	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Prédiction de B correcte (%)	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Prédiction de C correcte (%)	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Prédiction de D correcte (%)	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Précision (%)	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Validation croisée (61 éch.)									
Prédiction de A correcte (%)	33%	83%	0%	33%	83%	47%	20%	13%	100%	33%
Prédiction de B correcte (%)	63%	96%	37%	26%	26%	47%	53%	13%	93%	7%
Prédiction de C correcte (%)	40%	90%	65%	35%	45%	56%	50%	31%	100%	5%
Prédiction de D correcte (%)	38%	88%	25%	75%	25%	56%	25%	31%	100%	0%
Précision (%)	49%	92%	41%	36%	38%	52%	37%	23%	98%	8%
	Validation externe (10 éch.)									
Prédiction de A correcte (%)	0%	100%	0%	0%	0%	0%	100%	0%	50%	0%
Prédiction de B correcte (%)	40%	60%	0%	0%	40%	67%	67%	67%	33%	0%
Prédiction de C correcte (%)	33%	67%	0%	33%	0%	0%	0%	0%	50%	0%
Prédiction de D correcte (%)	0%	0%	100%	0%	0%	0%	0%	0%	50%	0%
Précision (%)	30%	60%	10%	10%	20%	22%	44%	22%	44%	0%

Les modèles développés en fonction des classes et des différents prétraitements possèdent une précision de 100% en calibration (Tableau 18). Cela met en évidence une capacité d'étalonnage performante des modèles à classer les lots en fonction des rendements d'extraction. En revanche, les performances en validation croisée et en validation externe sont insuffisantes et dépendantes du prétraitement. La précision est comprise entre 8% et 98% d'échantillons bien classés en validation croisée et entre 0% et 60% en validation externe. Les performances apparaissent en général moindres en validation externe qu'en validation croisée, mais cette dernière est par nature souvent plus optimiste que la validation externe.

L'utilisation de classes en effectif constant ne permet pas d'améliorer les modèles en comparaison des classes par amplitude de rendement. De plus, les prétraitements ne conduisent pas les mêmes performances selon les deux types de regroupement. En effet, le meilleur modèle pour les classes d'amplitude constante est obtenu avec la dérivée première (précision de 100% en calibration, 92% en validation croisée et 60% en validation externe), alors que le meilleur modèle pour les classes d'effectif constant est obtenu avec la SNVD (précision de 100% en calibration, 98% en validation croisée et 44% en validation externe). Il est important de noter que les échantillons mal classés sont généralement aux bornes des différentes classes utilisées. Ce constat suggère que l'affectation de ces échantillons a été faussé en raison d'une mauvaise estimation du rendement, les performances des classifications s'en trouvant de fait amenuisées.

La Figure 46 présente la projection des barycentres des 4 classes par effectif constant sur les axes factoriels 1 et 2. L'axe factoriel F1 permet d'expliquer 96,8% de la discrimination entre les échantillons. D'après l'analyse des contributions à l'axe 1, la répartition est expliquée principalement dans la zone 1200 à 1350 nm, zone qui correspond à la présence de graisse. Les travaux réalisés à partir de 4 classes, mettent en évidence la difficulté à séparer les deux classes intermédiaires (B et C) dont les barycentres sont très proches. De plus, le nombre d'échantillons dans les classes extrêmes (A et D) est trop faible.

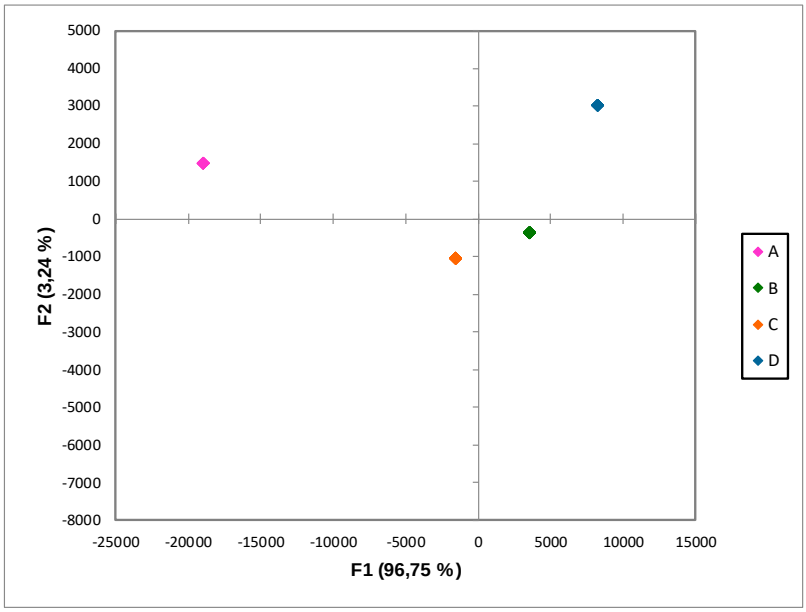


Figure 46 : Projection des barycentres des 4 classes à effectif constant sur les axes factoriels 1 et de 2 de l'Analyse Discriminante sur les données brutes (n=71)

(a) ii. La méthode des moindres carrés partiels (PLS) couplée à l'analyse discriminante (PLS-DA)

Des PLS-DA ont été réalisées sur les spectres des 71 échantillons, en fonction de différents prétraitements : sans prétraitement, dérivée première (D1), dérivée seconde (D2), Standard Normal Variate Detrending (SNVD) et Multiplicative Scatter Correction (MSC). Les modèles ont été développés avec 61 échantillons en calibration et 10 échantillons en validation externe. Le modèle est défini par un nombre très faible de variables latentes (1 à 3 maximum), limitant le risque de surapprentissage (Tuffery, 2007). Le Tableau 19 présente les performances obtenues par PLS-DA sur les données regroupées en 4 classes par amplitude de rendement et pour 2 variables latentes. Les résultats en calibration rendent compte d'une capacité de classification assez faible (précision comprise entre 54% et 71%). Pour la validation croisée et la validation externe, la classification est aléatoire avec une précision voisine de 50%. Le modèle le plus robuste est observé pour le prétraitement MSC avec une précision de 69% en calibration, de 59% en validation croisée et 57% en validation externe.

Tableau 19 : Performances de la PLS couplée à l'analyse discriminante en calibration, validation croisée et validation externe pour 2 variables latentes

	4 classes par amplitude de rendement				
	Brut	D1	D2	SNVD	MSC
	Calibration (61 éch.)				
Prédiction de A correcte (%)	59%	77%	87%	80%	72%
Prédiction de B correcte (%)	45%	53%	57%	62%	71%
Prédiction de C correcte (%)	61%	59%	53%	77%	69%
Prédiction de D correcte (%)	60%	64%	75%	81%	63%
Précision (%)	54%	59%	61%	71%	69%
	Validation croisée (61 éch.)				
Prédiction de A correcte (%)	58%	59%	59%	60%	62%
Prédiction de B correcte (%)	57%	51%	43%	58%	59%
Prédiction de C correcte (%)	56%	46%	51%	56%	61%
Prédiction de D correcte (%)	47%	54%	65%	48%	50%
Précision (%)	55%	51%	50%	56%	59%
	Validation externe (10 éch.)				
Prédiction de A correcte (%)	56%	28%	58%	34%	50%
Prédiction de B correcte (%)	50%	30%	44%	50%	60%
Prédiction de C correcte (%)	62%	55%	33%	60%	62%
Prédiction de D correcte (%)	56%	84%	52%	39%	95%
Précision (%)	53%	49%	54%	54%	57%

La Figure 47 présente la projection des échantillons sur les deux premières variables de la PLS. Il est difficile de visualiser un regroupement des échantillons en fonction de leur appartenance à une classe.

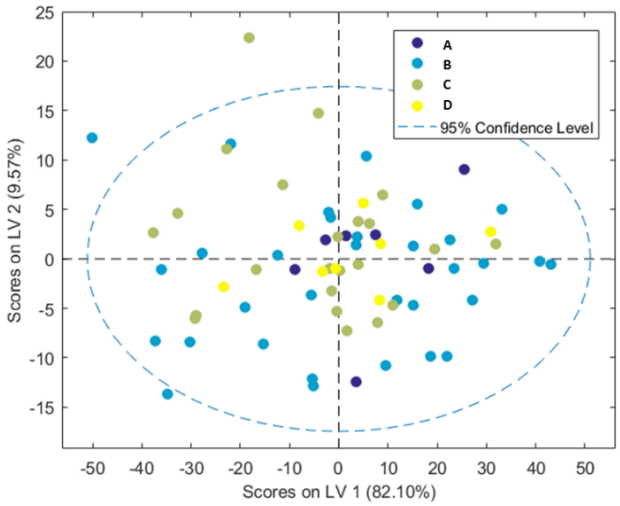


Figure 47 : Projection des échantillons sur les variables latentes 1 et 2 de la régression des moindres carrés partiels – Discriminante Analysis sur les données sans prétraitement

(b) Modèles de classification supervisée réalisés à partir de 3 classes de rendement potentiel en gélatine.

Des modèles ont été développés en créant de nouvelles classes qui regroupent en partie les données des classes intermédiaires. En effet, étant sur une répartition gaussienne les classes centrales présentent peu de différence de rendement. Le Tableau 20 présente la répartition des rendements dans ces nouvelles classes. Cette répartition a l'avantage d'être plus cohérente avec les attentes de l'industriel.

Tableau 20 : Répartition des classes pour les analyses réalisées avec 3 classes

3 Classes		**A**	**B**	**C**
Par amplitude constante de rendement	Gamme de rendement	14-16,5%	16,5-19,5%	19,5-22%
	Nombre d'échantillons	13	45	13

Le Tableau 21 présente les performances des modèles par DA pour 3 classes. Comme pour les modèles réalisés avec 4 classes, les performances en calibration sont excellentes (100% quel que soit le prétraitement utilisé). En calibration les pourcentages de bien classés sont maximum (100%) quel que soit le prétraitement utilisé. En validation croisée la précision est comprise entre 34 et 100% et entre 22 et 56% en validation externe. Les performances sont légèrement améliorées en validation croisée sur 3 clases en comparaison avec les modèles réalisés sur 4 classes. En validation externe, ces performances sont similaires pour trois ou quatre. Au final, la réduction du nombre de classes n'a pas permis d'améliorer les modèles de classification. Pour l'industriel, l'utilisation de 3 classes diminue la précision sur la gamme de rendements rencontrées, d'autant plus que la classe intermédiaire (B) a une amplitude très large (3%).

La précision en validation croisée a été calculée soit à partir de 61 échantillons, soit sur la totalité des échantillons (71 échantillons). L'augmentation du nombre d'échantillons permet en général d'améliorer la précision en validation croisée (données non présentées).

Tableau 21 : Performances de l'analyse discriminante en calibration, validation croisée et validation externe pour 3 classes

	3 classes				
	Brut	D1	D2	SNVD	MSC
	Calibration (61 éch.)				
Prédiction de A correcte (%)	100%	100%	100%	100%	100%
Prédiction de B correcte (%)	100%	100%	100%	100%	100%
Prédiction de C correcte (%)	100%	100%	100%	100%	100%
Précision (%)	100%	100%	100%	100%	100%
	Validation croisée				
Prédiction de A correcte (%)	91%	27%	91%	91%	100%
Prédiction de B correcte (%)	87%	28%	87%	69%	100%
Prédiction de C correcte (%)	83%	58%	83%	83%	100%
Précision (%)	87%	34%	87%	76%	100%
	Validation externe (10 éch.)				
Prédiction de A correcte (%)	0%	50%	50%	0%	50%
Prédiction de B correcte (%)	50%	17%	50%	50%	67%
Prédiction de C correcte (%)	0%	0%	100%	0%	0%
Précision (%)	33%	22%	56%	33%	56%

Pour classer les spectres des 71 lots de 75 tonnes de couennes, d'autres modèles ont été réalisés. Ils sont basés sur l'Analyse en Composantes Principales (ACP) couplée à une Analyse Discriminante (DA). L'ACP est utilisée afin de réduire les données spectrales en de nouvelles variables, les composantes principales, non liées entre elles. Cela permet de réduire le nombre de variables et donc de pouvoir développer des modèles qui n'auraient pas pu l'être (nombre de variables > nombre d'échantillons). Le nombre de composantes principales conservées est de 7, elles remplacent les 501 points (entre 1100 et 2100 nm) d'absorbance en entrées du modèle DA. Trois fonctions discriminantes basées sur les distances euclidiennes (linéaire), quadratique et de Mahalanobis ont été utilisées. Les performances observées sur trois classes par PLS-DA sont comparables aux résultats obtenus par PLS-DA sur 4 classes. Les meilleures performances ont été observées pour les modèles développés en quadratique (précision en calibration comprise entre 79% et 85%), les moins performantes sont les modèles développés en linéaire (précision en calibration entre 49% et 63%). Les données ont été à nouveau traitées en utilisant seulement 39 échantillons avec 13 échantillons dans chacune des trois classes. Cela n'a pas permis d'améliorer significativement les différents modèles en PLS-DA.

III.2. Conclusion

Les attentes initiales de l'entreprise étaient de pouvoir prédire le rendement d'un lot de 75 tonnes avec une précision de 0,5% ou en utilisant 5 classes dans la gamme de rendement d'extraction. Aujourd'hui, compte tenu du fait que l'industriel estime que sa mesure de rendement est d'une précision de ± 1%, atteindre 0,5% de prédiction sur des modèles de prédiction semble impossible, les modèles de prédiction calculés s'appuyant sur les données réelles de rendement.

Nous avons montré que les classifications étaient possibles, sur un nombre de classes restreint, en raison de la plage de variation des rendements des lots sur laquelle s'est basée cette étude. En effet les rendements de cette étude sont compris entre 14,2% et 21,8%, soit pour 5 classes, 1,52 % d'amplitude. Avec 1% de précision de mesure de rendement la répartition en 5 classes n'aurait eu que peu de sens. Par ailleurs, cette plage d'échantillon ne couvrait pas l'ensemble de la gamme de rendement rencontrée par l'industriel (dont l'amplitude couvre de 12 à 22 %)

Pour se rapprocher des réalités de la situation industrielle, l'étude a pris en compte l'unité de base de transformation, correspondant à un ordre de fabrication. Ce travail sur 75 tonnes constituait une contrainte et a limité le nombre des échantillons collectable. Comme cela avait été montré en flux au laboratoire, en situation industrielle il a été possible de discriminer les spectres provenant de couenne, des spectres de vide ou à l'interface entre couenne et vide.

La prédiction du rendement a été envisagée, dans un premier temps par des analyses quantitatives et dans un deuxième temps par des analyses qualitatives. La démarche quantitative n'a pas permis de construire des modèles de prédiction, quel que soit la structure de la base de données ou les prétraitements utilisés. En revanche, les résultats des analyses de classification sont plus prometteurs. Ces méthodes de classification ont fait appel essentiellement à l'ACP-DA et PLS-DA.

Les résultats de classification sont plus optimistes. Les données ont été traitées par des méthodes fondées sur un modèle linéaire, notamment par ACP-DA et PLS-DA. Les résultats obtenus par ACP-DA et PLS-DA ont permis de mettre en avant une tendance pour classer les données. Les résultats obtenus par DA sont plus encourageants, avec des résultats corrects en calibration pour certains prétraitements (60% d'échantillons bien classés en validation avec la dérivée première). La répartition des classes a été définie soit par amplitude de rendement soit par effectif pour les essais réalisés sur 4 classes.

L'utilisation des différents types de classes n'a pas permis d'améliorer les modèles de classification. De plus l'utilisation de classes par effectifs génère des classes qui présentent peu d'intérêt pour les industriels, car la gamme des valeurs extrêmes est large (>2% pour chacune des classes supérieure et inférieure), alors que les deux classes intermédiaires représentent 2% au total. Réduire le nombre de classe de 4 à 3, n'as pas permis d'améliorer les performances des différents modèles. En revanche, l'utilisation de trois classes diminue considérablement l'intérêt pour l'industriel, car la classe intermédiaire possède une amplitude de rendement dans laquelle se retrouve la grande majorité des extractions.

Plusieurs hypothèses peuvent être formulées pour expliquer les écarts entre les hautes performances obtenues en laboratoire pour la prédiction des couennes et celles obtenues en situation industrielle :

- Il est possible que la prise en compte du spectre moyen obtenu sur un lot de production de grande taille (unité minimale de 75 tonnes pour obtenir une valeur de rendement) ait entrainé une perte d'information correspondant à l'hétérogénéité des couennes composants ce lot. Ce constat avait déjà été fait à l'échelle laboratoire, où les modèles de prédiction de la composition des couennes étaient plus performants pour la prédiction issue de la moyenne des prédictions par couenne en comparaison à la prédiction sur des spectres moyens des couennes.
- Les différences observées entre les performances des modèles quantitatifs et des modèles en classification, suggèrent le manque de précision sur l'estimation du rendement fournie par l'industriel. Cette hypothèse est renforcée si l'on considère que les échantillons mal classés dans les modèles de classification sont principalement les échantillons aux bornes de ces classes.
- L'insuffisance du nombre d'échantillons a été un frein à l'évaluation des capacités de prédiction. Pour les modèles quantitatifs, peu d'échantillons étaient disponibles aux rendements extrêmes. Pour les modèles de classification, il faudrait augmenter le nombre d'échantillons et équilibrer les classes afin qu'elles possèdent les mêmes poids.
- Les modèles de prédiction construits ont fait appel essentiellement à des modèles linaires. Or, il a été montré dans la première partie de ce travail, que les reliquats de tissus graisseux sous-jacents à la couenne constituaient le facteur discriminant des couennes. La proportion de ces tissus n'est sans doute pas en progression linéaire d'une couenne à l'autre ou d'un lot à l'autre, mais sera très dépendante de l'origine et des conditions de découpe de l'animal. Ainsi on peut penser que la prise en compte de la seule dimension linéaire est insuffisante pour refléter la réalité de la dispersion de l'hétérogénéité des couennes. Il serait intéressant d'essayer d'autres méthodes basées sur des approches non linéaires du type réseaux de neurones (Gevrey *et al.*, 2003).

- Il a été constaté que les taux d'humidité et de température varient au niveau de la sonde. A terme, l'implémentation de ces informations pourrait permettre d'améliorer les modèles.

Il est clair qu'une plus grande précision de la mesure du rendement pourrait améliorer les performances des différents modèles. L'acquisition spectrale peut, elle-aussi, être optimisée en standardisant le passage des matières premières (hauteur et disposition), mais la chaîne de production telle qu'elle est conçue actuellement ne permet pas cette standardisation. En effet, la mise en place d'un système pour répartir les matières aurait pour conséquence de ralentir fortement le flux en début de processus.

L'option la plus intéressante pour garantir la précision d'estimation du rendement serait de se placer en condition expérimentale. Il s'agirait de raisonner sur 25 tonnes de couennes et de réaliser les extractions successives de ces couennes dans une seule cuve de 25 tonnes avec un processus strictement identique entre les essais. Le rendement de chaque extraction serait déterminé par réfractométrie, afin de calculer le potentiel d'extraction total. Cependant, cette démarche imposerait à l'industriel des contraintes spécifiques.

Chapitre IV. Etude des propriétés physico-chimiques des gélatines par spectroscopie proche infrarouge et par Asymmetrical Flow Field-Flow Fractionation couplée à un détecteur UV Multi Angle Ligth Scattering

IV.1. Introduction

Les travaux et les connaissances acquises au cours des précédents chapitres ont permis de soulever de nouvelles problématiques liées au produit transformé, la gélatine. Ce chapitre se décompose en deux sous-parties ; la première aborde la caractérisation de la gélatine par spectroscopie proche infrarouge en fin de processus, alors que la deuxième partie étudie la distribution moléculaire de gélatines en fonction des propriétés physico-chimiques.

Dans le domaine de l'agroalimentaire, la qualité est omniprésente et n'a cessé de s'améliorer depuis les années 90. La qualité permet de certifier au consommateur l'innocuité microbiologique, physique et chimique d'un produit, ainsi que d'en garantir ses propriétés technologiques. Les règles d'application de la qualité des produits de transformation d'origine animale sont principalement régies par l'application de la norme ISO 22 000 (Arvanitoyannis, 2009). Les gélatines sont également caractérisées en termes de propriétés technologiques, force en gel, viscosité, humidité. En effet, les clients commandent une gélatine avec telle ou telle caractéristique et cela varie en fonction de l'utilisation qu'ils vont en faire. L'entreprise doit donc caractériser ses produits finis pour répondre à cette attente. Pour cela un échantillon pour chaque lot produit est analysé en laboratoire. Les professionnels font ensuite des mélanges pour obtenir un lot final, livrable aux clients. Ces analyses chimiques nécessitent un délai minimum de 24h. Dans certains cas, il arrive que l'industriel réalise de nombreux mélanges successifs avant d'obtenir le produit attendu ce qui engendre un délai d'analyse supplémentaire après chaque mélange.

La gélatine en sortie d'usine est sous la forme de grains, elle est donc relativement homogène. Dans un objectif de gain de temps, la spectroscopie proche infrarouge pourrait donc être un outil adapté pour réaliser une analyse rapide et non destructive de cette gélatine. Des essais ont donc été réalisés afin d'évaluer la capacité à prédire les propriétés technologiques d'intérêt de gélatines produites sur le site de Graulhet. Ce sujet constituera la première sous partie de l'étude.

Un deuxième point d'intérêt pour l'entreprise est lié à la maîtrise du processus de déshydratation de la gélatine. Actuellement, les clients attendent un taux d'humidité des gélatines compris entre 10 et 12%. Afin de répondre à cette attente, l'entreprise Weishardt commercialise les gélatines avec une spécification de 12% d'humidité, en se fixant un objectif de 11,5% en fin de processus de séchage. Or, le taux d'humidité moyen de la production entre 2013 et 2016 sur les trois séchoirs différents est de 11,02% (Annexe 3 - analyse par perte de masse). Cette différence du taux d'humidité représente une perte financière de 0,5% du poids total des gélatines vendues. En effet, le temps nécessaire d'analyse (18h minimum), ne permet pas aux opérateurs responsables des séchoirs de maîtriser en temps réel le processus de séchage. Or, la connaissance en temps réel de ce paramètre pourrait permettre d'adapter le processus afin d'atteindre, plus précisément, la valeur désirée et ainsi diminuer la perte financière liée à la sous-évaluation du taux d'humidité. La détermination du taux d'humidité dans les produits alimentaires est une des applications les plus courantes en spectroscopie proche infrarouge (Tsuchikawa *et al.*, 1996; Büning-Pfaue, 2003) permettant une analyse rapide et non-destructive qui peut aussi bien être appliquée *at-line* et *on-line*. Cet outil semble donc être adapté à cette problématique.

La seconde partie de cette étude sur la gélatine nous permettra de mieux comprendre l'origine de ses propriétés physico-chimiques, force en gel et viscosité, au niveau moléculaire. Généralement, l'étude des proportions de chaines primaires (α, γ, β) est réalisée par chromatographie d'exclusion stérique (SEC). Mais cette technique ne permet pas d'étudier l'ensemble de la gamme de masses molaires de la gélatine. L'apport de l'Asymmetrical Flow Field Field-flow Fractionation (AsFlFFF) permet d'étudier aussi bien les masses molaires des chaines primaires que celles des agrégats (de haute masse molaire). De plus, à concentration égale, les colloïdes qui possèdent des fragments de haute masse molaire apportent de la viscosité (Mewis et Wagner, 2012). Une étude par AsFlFFF nous permet donc, puisqu'elle tient compte de ces hautes masses molaires, d'avoir un outil de compréhension de cette viscosité, outil plus précis que la SEC. Elle permettra également d'appréhender la force en gel. Le premier intérêt de cette étude est de tenir compte, à la fois des chaînes primaires mais également des agrégats de haute masse molaire afin d'expliquer les propriétés physico-chimiques des gélatines.

Le second intérêt concerne le traitement des données chromatographiques. Traditionnellement, les chromatogrammes obtenus par SEC sont étudiés en intégrant les pics chromatographiques. Notre approche est inspirée des techniques de « datamining » appliquées aux spectres infrarouges (Tufféry, 2011; Bertrand et Dufour, 2006a). Il s'agit d'utiliser, non pas les pics pris individuellement, mais d'utiliser le profil du fractogramme (issu de l'AsFlFFF) dans son intégralité.

IV.2. Matériels et méthodes

IV.2.1. Caractérisation de la gélatine par spectroscopie proche infrarouge

IV.2.1.(1). Echantillons de gélatine

Un échantillonnage de 135 gélatines en grains produites sur le site de production de Weishardt à Graulhet, a été réalisé en novembre 2016. Ces gélatines ont été caractérisées en termes d'humidité, de force en gel et de bloom, par l'industriel. L'analyse descriptive de ces gélatines est présentée dans le Tableau 22.

Tableau 22 : Propriétés physico-chimiques des gélatines échantillonnées. Nombre d'échantillons (n), moyenne, écart-type (SD), Valeur minimale et maximale et coefficient de variation (CV) de la force en gel et de la viscosité des gélatines. (n=135)

	Moyenne	**Min**	**Max**	**SD**	**CV (%)**
Humidité (%)	11.5	7.6	13.9	0.98	8.5
Force en gel (g)	256.7	121.0	311.0	47.60	18.5
Viscosité (mPa)	40.4	22.5	56.7	8.05	19.9

IV.2.1.(2). Acquisition spectrale des gélatines

Les spectres Proche Infrarouge ont été acquis sur un spectromètre Proche Infrarouge à transformée de Fourier (MPA, Bruker Optics, Ettlingen, Germany). L'acquisition est réalisée en absorbance sur 100 g de gélatine en grain dans une coupe mise en rotation sur le module sphère d'intégration à température ambiante du laboratoire (22±1°C). L'appareil permet d'acquérir entre 800 nm et 2800 nm avec un pas de 1,3 nm (vitesse de balayage 10 kHz, 32 scans par échantillon et par background). L'acquisition spectrale et le contrôle de l'appareil sont réalisés à l'aide du logiciel OPUS (v. 6.5 Bruker Optics, GmbH 2009).

IV.2.1.(3). Analyses de références

(a) Taux d'humidité

La détermination du taux d'humidité est réalisée selon la norme NF V 59-003 (AFNOR, 1982b). Une première pesée de 5 grammes de gélatine est réalisée dans un creuset, puis celui-ci est placé à 105±1°C pendant 18h. Après étuvage, une contre pesée est réalisée sur les creusets après refroidissement au dessiccateur. Les résultats exprimés sont des moyennes de deux mesures. L'écart-type de répétabilité est de ±0,15 pour cette analyse.

(b) Force en gel (Bloom)

La force en gel est déterminée par la méthode du Bloom, qui correspond à la force nécessaire (en gramme) à un piston de 12,7 mm de diamètre pour pénétrer dans un gel de gélatine de 4 mm de profondeur. Une solution aqueuse de gélatine à 6,67% (m/V) est préparée dans des verres adaptés pour la mesure : hauteur de 60mm, diamètre intérieur de 59mm (SCOTT, 2.102.501). Une solution de gélatine est préparée en pesant 7,5±0,1g de gélatine, qui sont dilués dans 105 ml d'eau distillée (Milli-Q, Millipore, Bedford, MA, USA) à 20°C. Les verres sont placés dans un premier bain à 40±2°C pendant 10 minutes, puis transférés dans un second bain-marie à 65±2°C pendant 20 min pour permettre la dissolution complète de la gélatine. Les verres sont finalement placés dans un bain à 10±0,1°C pendant 17±1h, conduisant à la gélification.

La force en gel (ou degré de Bloom) est mesurée directement à l'aide d'un gélomètre (Stevens-LFRA, C Stevens and Son Ldt, United Kingdom), qui détermine la force maximum nécessaire pour que le cylindre du gélomètre pénètre la surface du gel à une vitesse de 0,5 $mm.s^{-1}$ sur une profondeur de 4 mm.

(c) Viscosité

La viscosité des gélatines est définie par le temps d'écoulement en continu de la gélatine en solution, à une température définie. Le temps nécessaire à cet écoulement continu pour un volume donné de solution de gélatine est ensuite converti afin d'estimer la viscosité (European-norm, 2002). Pour cela, une pipette viscosimétrique (possédant un trait de jauge supérieur et inférieur) de 100mL étalonnée à 60°C, est placée en position verticale dans un thermoplongeur à circulation, réglé à 60±0,1°C. Des solutions de gélatine à 6,67% (m/V) sont préparées en pesant 7,5±0,1g de gélatine pour 105 ml d'eau distillée à 20°C (Milli-Q, Millipore, Bedford, MA, USA).

Les étapes de gonflement et de solubilisation sont identiques à celles décrites précédemment pour la force en gel (Bloom). Immédiatement après solubilisation, le gel de gélatine est transféré dans la pipette viscosimétrique. Les mesures sont réalisées en duplicata. Lorsque la gélatine atteint la température de 61°C, l'analyse est lancée. La viscosité est ensuite déterminée par le temps d'écoulement et convertie en millipoise.

IV.2.1.(4). Analyse des données :

Le traitement chimiométrique des données a été réalisé sur Matlab® avec la PLS Toolbox (Eigenvector Technologies, Manson, Washington, USA). Des modèles de prédiction de la force en gel (Bloom), de la viscosité et du taux d'humidité ont été développés par PLSr. Deux échantillons ont été considérés comme aberrants. Les modèles ont été développés en validation croisée et en validation externe avec un jeu de données en calibration de 102 échantillons (75%) et un jeu de données en validation externe de 32 échantillons (25%). Les échantillons pour la validation ont été sélectionnés par la méthode Onion de la PLS Toolbox. Différents prétraitements ont été utilisés ; les données brutes, dérivée première, dérivée seconde, Standard Normal Variate (SNV), SNV Detrend et Multiplicative Scattering Correction (MSC). La qualité des modèles de calibration a été évaluée en recherchant d'une part le plus haut coefficient de détermination entre les valeurs prédites et les valeurs mesurées en calibration, en validation croisée et en validation externe (r^2_{cal}, r^2_{cv}, r^2_v), et d'autre part, la plus faible variance résiduelle (Root Mean Square Error) en calibration et en validation (RMSEC et RMSEV)(Agelet et Hurburgh, 2010).

IV.2.2. Caractérisation de gélatine par Asymmetrical Flow Field-Flow Fractionation

IV.2.2.(1). Echantillonnage de gélatine

Quarante-neuf échantillons de gélatine industrielle produite durant l'année 2016 et provenant de trois sites de production différents sélectionnés. Les gélatines de type A (extraction acide), extraites à partir de peaux de porcs ont été sélectionnées de façon à représenter la gamme de valeurs de force en gel (0 à 350 Bloom) et de viscosité (0 à 100 millipoise) généralement rencontrées sur les gélatines (Tableau 23).

Tableau 23 : Propriétés physico-chimiques des gélatines échantillonnées. Nombre d'échantillons (n), moyenne, écart-type (SD), Valeur minimale et maximale et coefficient de variation (CV) de la force en gel et de la viscosité des gélatines. (n=49)

	Moyenne	**Min**	**Max**	**SD**	**CV (%)**
Force en gel (g)	231,93	45.05	323.23	75.41	32.51
Viscosité (mPa)	39,70	15.17	97.77	15.62	39.34

IV.2.2.(2). Analyses physico-chimiques

Les analyses de la force en gel et de la viscosité ont été réalisées suivant les même modalités que celles décrites précédemment en page 150.

IV.2.2.(3). Analyse de la distribution molaire par AsFlFFF-UV/MALS

(a) Eluant

L'éluant est réalisé selon la méthode de Fraunhofer (Fraunhofer, Winter, et Coester, 2004), en préparant une solution à 2mM de phosphate de sodium et à 14mM de chlorure de sodium. Le pH de la solution est ajusté à 6 en ajoutant une solution d'acide phosphorique (85% v/v). L'éluant est ensuite filtré à travers deux membranes en PTFE, respectivement de 10 et 40µm (Pall Corporation, Ann Arbor, Michigan) à l'aide d'une pompe à vide (Vacuum/Pressure Pump, Pall Corporation, Ann Arbor, Michigan).

(b) Préparation des solutions de gélatine

La gélatine est solubilisée, à 3,3% (m/V) en prélevant 1,66g de gélatine pour 50mL d'eau distillée (Milli-Q, Millipore, Bedford, MA, USA). Les tubes sont ensuite placés 12h dans un agitateur chauffant (MultiTherm Shaker, Benchmark Scientific Inc, NJ USA) à 37°C et 500rpm, puis centrifugés (Centrifuge MR 23i, JOUAN SA, St Herblain, France) à 13 000g, à une température de 37°C pendant 5min. Après centrifugation, 0,605mL de surnageant sont prélevés et dilués avec de l'éluant, pour obtenir une solution à 0,2% (m/V). Seuls les échantillons ne présentant pas de culot ont été pris en compte dans les analyses. A partir de cette solution, des triplicas sont préparés, afin de réaliser les analyses par AsFlFFF-UV/MALS.

(c) L'analyse par AsFlFFF

L'analyse AsFlFFF est réalisée avec un Dual Tech System (Wyatt Technology Europe, Dernbach, Germany) couplée à un système HPLC Dionex® ultimate 3000 Series (LC-Packings, Dionex, Amsterdam, The Netherlands), connecté à un détecteur 18 angle Multi-Angle Light Scattering (MALS) Heleos II instrument (Wyatt Technology, Santa Barbara, USA), et un détecteur ultraviolet (UV) Ultimate 3000 RS (LC-Packings, Dionex, Amsterdam, The Netherlands) à 216nm. Le détecteur MALS utilise un laser à la longueur d'onde de 690 nm : les détecteurs sont normalisés à l'aide d'un standard de Sérum Albumine Bovine. L'intensité de la diffusion de lumière est calibrée avec une solution pure de toluène (HPLC grade, Sigma-Aldrich, Merck KGaA, Germany). Les échantillons sont injectés dans le canal AsFlFFF par le système HPLC équipé d'un dégazeur sous vide en ligne. Un filtre en ligne de 0,1 µm (VVLP, Millipore, Germany) est placé entre la pompe et le canal. Le canal AsFlFFF possède une forme géométrique de trapèze d'une longueur de 19,5cm, avec une largeur de 1,65cm en entrée de la cellule et de 0,27cm en sortie de la cellule. Au sein de ce canal, l'épaisseur est apportée par un film en Mylar de 250µm, placé sur une membrane d'ultrafiltration avec un maillage de 5kDa (Wyatt Technology Europe, Dernbach, Germany). Les échantillons ont été analysés en triplicas.

Les conditions de fractionnement en AsFlFFF sont décrites par différentes étapes suivantes :

Pendant l'étape de « focus injection », l'éluant est injecté par l'entrée et la sortie du canal et est éliminé entièrement par le « cross-flow ». Le flux du « cross-flow » est alors fixé à 1,5 $mL.min^{-1}$ pendant une minute. Puis, 30 µL de l'échantillon à analyser sont injectés à un débit de 0,2 $mL.min^{-1}$, pendant quatre minutes, le débit du cross-flow étant maintenu à 1,5 $mL.min^{-1}$.

A la fin de l'étape de « focus injection », le flux en sortie de canal, en direction des détecteurs est fixé 1$mL.min^{-1}$. Au début de la phase d'élution, le « cross-flow » est fixé à 4$mL.min^{-1}$ pendant quatre minutes, le flux est réduit ensuite de façon linéaire de 4$mL.min^{-1}$ à 0,1 $mL.min^{-1}$ pendant vingt minutes. Ensuite, le flux reste constant à 0,1 $mL.min^{-1}$ jusqu'à la fin de l'analyse. L'analyse dure 40 min au total.

En accord avec les travaux de *Tromp et al.* (Tromp, ten Grotenhuis, et Olieman, 2002)et *Rbii et al.* (Rbii et al., 2009), pour la gélatine, un incrément d'indice de réfraction *(dn/dc*) de 0,164 $mL.g^{-1}$ et un coefficient d'extinction molaire de l'UV (ε) de 12 700 $mL.g^{-1}.cm^{-1}$ ont été utilisés pour les calculs de masse molaire (masse molaire en nombre (Mn) et masse moyenne en masse (Mw), traduisant la dispersité) et des rayons de giration, selon le modèle de Zimm (Zimm, 1948) (logiciel ASTRA Version 6.1.1.17, Wyatt technology, Santa Barbara, US). Le rayon de giration peut être considéré comme la distance par rapport à un axe d'un point virtuel ayant la masse et l'inertie de la molécule. La masse moyenne en nombre représente la somme des masses molaires de polymères présents pondérée par le nombre de chaînes polymériques de chaque masse, alors que la masse molaire moyenne en masse représente la moyenne des masses molaires pondérée par la masse de chaînes de chaque longueur.

IV.2.2.(4). Analyse des données UV des fractogrammes de gélatine

Le traitement chimiométrique des données a été réalisé sur Matlab® avec la PLS Toolbox (Eigenvector Technologies, Manson, Washington, USA). Les absorbances UV mesurées toutes les 0,49 s durant l'élution de 10 à 40 min, constituent les variables, soit 3665 variables. Le fractogramme moyen des trois répétitions des échantillons est pris en compte pour les 49 échantillons analysés, la matrice résultante est donc de type $n \times t$ avec $n = 49$ et $t = 3665$.

Des modèles de prédiction de la force en gel (Bloom) et de la viscosité ont été développés par PLSr. Les modèles ont été développés en validation croisée. La qualité des modèles de calibration a été évaluée en recherchant d'une part le plus haut coefficient de détermination entre les valeurs prédites et les valeurs mesurées en calibration et en validation croisée (r^2_{cal} et r^2_{cv}), et d'autre part, la plus faible variance résiduelle (Root Mean Square Error) en calibration et en validation croisée (RMSEC et RMSECV)(Agelet et Hurburgh, 2010).

IV.3. Résultats et discussion

IV.3.1. Caractérisation de gélatine par spectroscopie proche infrarouge

IV.3.1.(1). Mise en évidence des zones spectrales d'intérêt dans la différentiation des gélatines

Les spectres proche infrarouge bruts et après transformation SNVD sont présentés dans la Figure 48 A et B. Les spectres obtenus présentent un niveau faible de bruit, sauf sur la zone spectrale comprise entre 2700 nm et 2800 nm. Sur les données brutes, on constate des différences d'intensité probablement dues aux différences de granulométrie entre les gélatines. Le traitement SNVD permet de supprimer ces différences d'intensités. Ce prétraitement permet de mettre en évidence les zones spectrales d'intérêt pour différencier les spectres (zone en rouge). La zone autour de 1500 nm pourrait correspondre à la liaison amide (1ère harmonique) et celle de 1600 nm au N-H libre de cette liaison amide (1ère harmonique et combinaison). La zone comprise entre 2050 nm et 2200 nm, pourrait, elle-aussi, correspondre aux amides I (1ère harmonique) et III (combinaison). Cette zone varie entre les protéines et en fonction de la structure secondaire (Robert *et al.*, 1999; Kamishikiryo). Cependant les bandes autour de 1450 nm (1ère harmonique) et 1950 nm (combinaison) peuvent aussi correspondre au spectre d'absorption de l'eau libre (Büning-Pfaue, 2003; Bertrand et Dufour, 2006b). La dernière zone spectrale qui semble réagir est autour de 2500 nm, elle pourrait correspondre aux liaisons C-H ou C-C caractéristiques des molécules organiques.

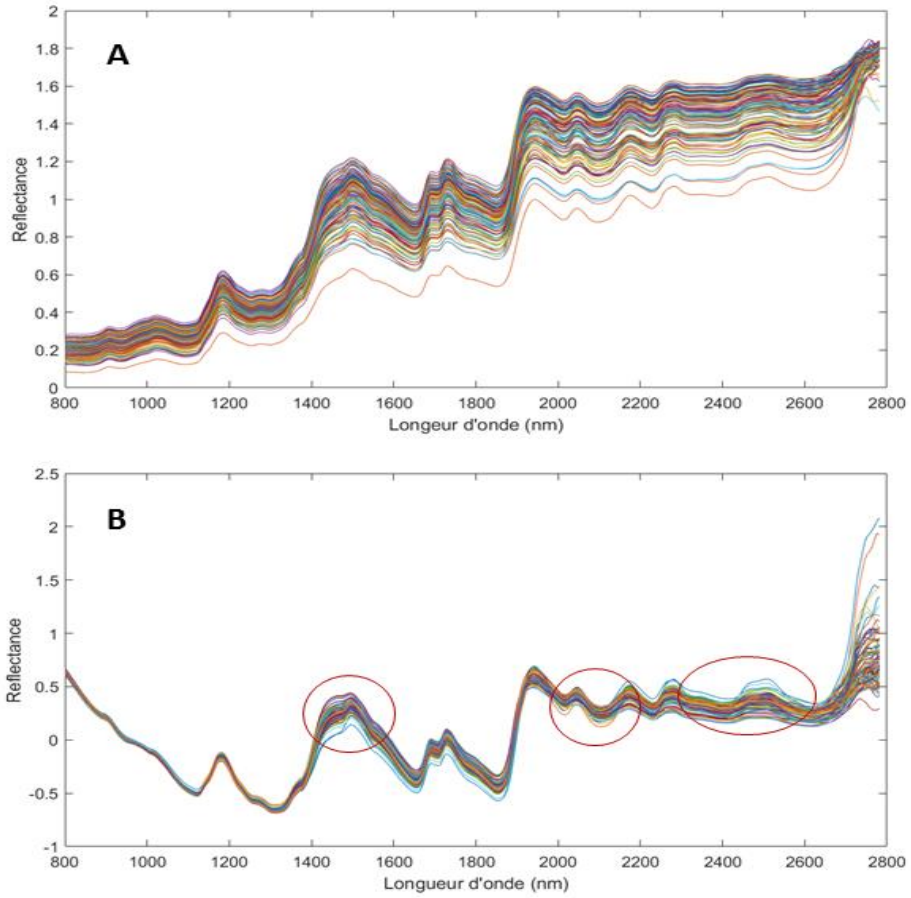


Figure 48 : Spectres proche infrarouge des 135 gélatines industrielles ; (A) pour les spectres bruts, (B) pour les spectres après prétraitement SNVD

(a) Prédiction de la force en gel, de la viscosité et de l'humidité des gélatines par spectroscopie proche infrarouge

(a) i. Prédiction de la force en gel

Les gélatines étudiées proviennent toutes de la même usine et sont exclusivement sous forme de grains. Elles sont représentatives de la majorité de la production de Weishardt, sur le site de Grauhlet. Cependant, l'échantillonnage manque de gélatines avec des forces en gel faibles ce qui explique le coefficient de variation faible (18,5%) pour cette analyse de référence. Le Tableau 24 présente les statistiques de calibration, validation croisée et validation externe des modèles de prédiction PLSr développés pour la force en gel. Le nombre de variables latentes pour les différents modèles de prédiction varie entre 3 et 8, avec donc probablement peu de surinterprétation des modèles (Esbensen et al., 2002).

Tableau 24 : Statistiques de calibration, de validation croisée et de validation pour les modèles de prédiction de la force en gel par SPIR. Les données ont été traitées par PLSr en validation (75% en validation, 25% en validation). Avec LV pour Variables Latentes.

	Prétraitement	LV	r^2 cal	r^2 cv	r^2 val	RMSEC	RMSEV	Ecart-type	RPD
Force en gel	Brut	7	0.80	0.75	0.84	20.57	18.76	54.77	2.66
	D1 (31 points)	8	0.81	0.75	0.78	20.47	22.73	54.77	2.68
	D2 (15 points)	3	0.18	0.13	0.22	42.80	42.20	54.77	1.28
	SNV	5	0.76	0.70	0.82	23.12	20.63	54.77	2.37
	SNVD	**6**	**0.79**	**0.75**	**0.86**	**21.43**	**18.33**	**54.77**	**2.56**
	MSC	5	0.76	0.70	0.82	23.15	20.61	54.77	2.37

De manière générale, les performances des différents modèles sont correctes, sauf pour la dérivée seconde et pour les dérivées premières avec un lissage faible (<21 points). Le modèle le plus adapté pour la prédiction de la force en gel est après prétraitement SNV et detrend. C'est le modèle qui possède les meilleures performances pour le plus faible nombre de variables latentes. Avec un r^2 de calibration et de prédiction respectivement de 0,79 et 0,86, et un RPD de 2,56, il est envisageable d'appliquer un modèle de prédiction sur la gélatine en grains (Nicolaï et al., 2007; Kapper, 2012). Cependant le RMSE reste relativement faible (proche de 18) par rapport à l'erreur de l'analyse de référence (environ 15).

(a) ii. Prédiction de la viscosité

Comme pour la force en gel, la gamme couverte de viscosité est élevée, avec un coefficient de variation relativement faible (19,9%). Le Tableau 25 présente les statistiques de calibration et de validation pour les différents modèles de prédiction de la viscosité. Les performances sont en général non satisfaisantes, les coefficients de corrélation en calibration sont compris entre 0,52 et 0,61 pour les meilleurs prétraitements et les RPD sont inférieurs à 1,9. Un modèle de prédiction de la viscosité, basé sur des analyses par SPIR n'est pas envisageable mais cela pourrait être utilisé pour classer les échantillons.

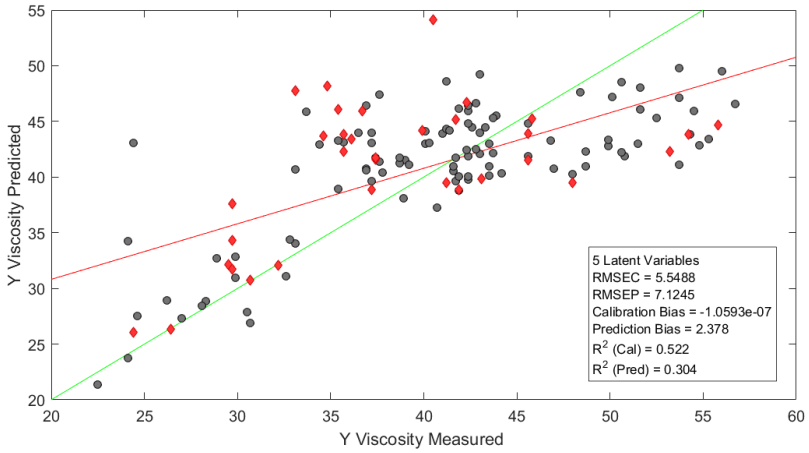
Tableau 25 : Statistiques de calibration, de validation croisée et de validation externe pour les modèles de prédiction de la viscosité par SPIR. Les données ont été traitées par PLSr en validation (75% en validation, 25% en validation). Avec LV pour Variables Latentes.

	Prétraitement	LV	r^2 cal	r^2 cv	r^2 val	RMSEC	RMSEV	Ecart-type	RPD
Viscosité	Brut	7	0.59	0.51	0.37	5.12	6.60	8.83	1.72
	D1 (31 points)	8	0.62	0.49	0.51	4.95	5.82	8.83	1.78
	D2 (15 points)	3	0.12	0.11	0.05	7.55	7.88	8.83	1.17
	SNV	5	0.52	0.46	0.30	5.54	7.13	8.83	1.59
	SNVD	6	0.59	0.51	0.51	5.12	5.84	8.83	1.72
	MSC	5	0.52	0.47	0.30	5.55	7.12	8.83	1.59

La faible performance du modèle pourrait s'expliquer en partie par une perte de sensibilité de l'analyse de référence pour les fortes viscosités entraînant une réponse non proportionnelle (Figure 49).

Ces observations sont en accord avec les observations de *Segtnan et al.* (Segtnan *et al.*, 2003). Une transformation logarithmique a été réalisée sur l'analyse de référence afin de vérifier si cela entraînait une réponse plus linéaire. Ces essais n'ont pas permis d'améliorer significativement la prédiction de la viscosité (r^2cal compris entre 0,63 et 0,54).

Figure 49 : Modèle de prédiction de la viscosité par PLSr. En noir les données de calibration (n=102), en rouge les données en validation (n=33)



(a) iii. Prédiction du taux d'humidité

Les échantillons possèdent un coefficient de variation de 8,8% pour le taux d'humidité, ce qui est faible, mais représentatif de la gamme à analyser. Comme pour la force en gel et la viscosité, les modèles de prédiction ne sont pas performants en dérivée seconde et dérivée première pour les lissages inférieurs à 21 points. En revanche, les autres modèles de prédiction sont très performants. Comme pour la force en gel, le meilleur prétraitement est la SNV detrend, avec un coefficient de corrélation en calibration et validation de 0,95 et un RPD de 4,27 et une RMSEC de 0,23%. Le modèle de prédiction PLSr peut donc être applicable pour prédire le taux d'humidité.

Tableau 26 : Statistiques de calibration, de validation croisée et de validation externe pour les modèles de prédiction de l'humidité. Les données ont été traitées par PLSr en validation (75% en validation, 25% en validation). Avec LV pour Variables Latentes.

	Prétraitement	LV	r^2 cal	r^2 cv	r^2 val	RMSEC	RMSEV	Ecart-type	RPD
Humidité	Brut	7	0.92	0.90	0.90	0.28	0.28	0.98	3.51
	D1 (31 points)	8	0.91	0.88	0.86	0.31	0.34	0.98	3.17
	D2 (15 points)	3	0.03	0.03	0.10	1.14	0.98	0.98	0.86
	SNV	5	0.89	0.87	0.83	0.33	0.35	0.98	2.98
	SNVD	**6**	**0.95**	**0.93**	**0.95**	**0.23**	**0.21**	**0.98**	**4.27**
	MSC	5	0.90	0.88	0.84	0.33	0.35	0.98	2.98

Des essais supplémentaires ont été réalisés pour prédire 6 échantillons inconnus avec une erreur maximale par rapport à la valeur de référence de 0,17% et une erreur moyenne de 0,10%. Ces résultats sont très prometteurs, d'autant plus que l'erreur de l'analyse de référence est estimée à 0,15%.

IV.3.2. Etude de la distribution moléculaire de la gélatine par AsFlFFF

IV.3.2.(1). Caractérisation des distributions molaires des gélatines.

La Figure 50 présente deux fractogrammes caractéristiques de deux gélatines possédant des paramètres technologiques très différents : Gélatine A forte viscosité (46,71 mP) et une force en gel élevée (322,46 g), Gélatine B faible viscosité (20,54mP) et une faible force en gel (98,64 g). Le signal UV répond quantitativement à la présence de protéine ; le signal MALS correspond à la diffusion de lumière mesurée (pour un angle de 90° dans la Figure 50 A et B) et qui est fonction de la taille. Pour les polymères de même nature, le signal MALS permet de calculer la masse molaire dont l'évolution, compte tenu des conditions d'analyse, est quasi linéaire tout au long de l'élution. En final la combinaison de l'UV et du MALS permet une approche globale de la distribution des fractions de gélatine en fonction de la masse molaire. Ainsi la masse molaire augmente régulièrement au court de l'élution couvrant les objets les plus petits jusqu'à 1×10^9 $g.mol^{-1}$.

Les deux gélatines A et B ne présentent pas les mêmes répartitions quantitatives des polymères protéiques tout au long de l'élution. Le signal UV de la gélatine A est compris entre 0,4 et 0,8 de la 10ème à la 35ème minute d'élution témoignant d'une répartition quantitative homogène des polymères sur l'ensemble de la gamme de masses molaires. Il n'en n'est pas de même pour la gélatine B, dès la quinzième minute d'élution, l'intensité du signal UV est inférieure à 0,3 pour s'atténuer progressivement jusqu'à la 35ème minute, montrant ainsi une proportion faible de polymères de masse comprise entre 4×10^5 g.mol^{-1} et 5×10^8 g.mol^{-1}.

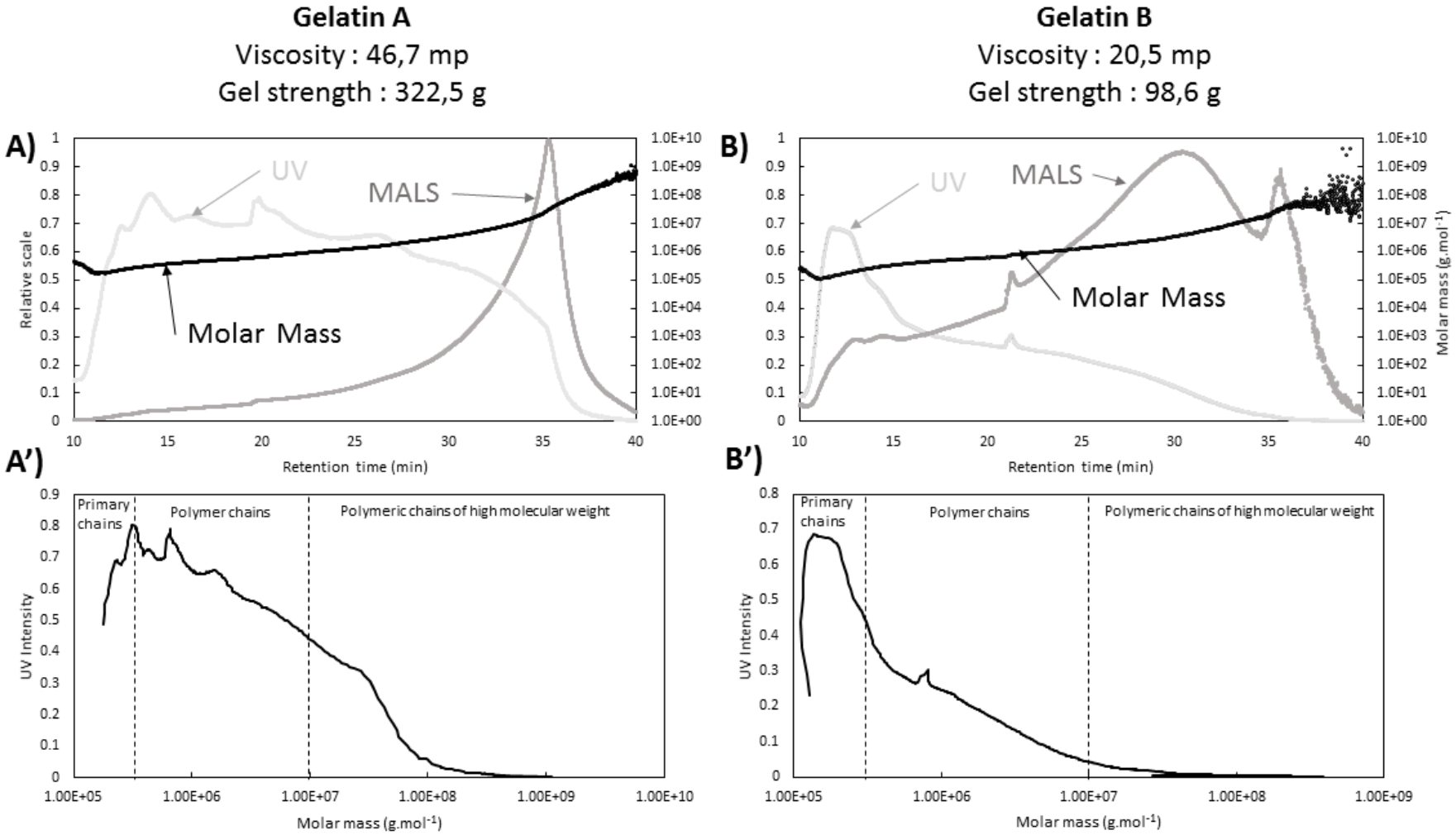


Figure 50 : A et B décrivent les fractogrammes de l'analyse AsFlFFF-UV/MALS de gélatines de porc : A et B en fonction du temps d'élution, A' et B' le signal UV en fonction de la masse molaire. La gélatine (A, A') possède une force en gel de 322,46 g et une viscosité de 46,71 mP, la gélatine (B, B') possède une force en gel de 98,64 g et une viscosité de 20,54 mP. Le signal UV en intensité relative (ordonnées de droite-point en gris clair). Le signal de diffusion de la lumière de l'angle 90° en intensité relative (ordonnée de droite- points en gris foncés). La masse molaire calculée en g.mol^{-1} (ordonnée de gauche- points en noir).

Les diagrammes ont été séparés en deux séquences distinctes (Figure 50, A' et B') : les fractions de faible masse molaire correspondant aux chaînes primaires et les fractions de haute masse molaire se rapportant aux chaînes polymériques. La fraction des chaînes primaires, est constituée des chaînes α de la protéine de gélatine de masse molaire comprise entre 0.5×10^5 et 1×10^5 g.mol^{-1}, le dimère β de masse molaire comprise entre 1×10^5 et 2×10^5 g.mol^{-1} et le trimère γ de masse molaire comprise entre 2×10^5 et 4×10^5 g.mol^{-1}(Meyer et Morgenstern, 2003). La fraction des chaînes primaires comprend des masses molaires entre $0,5\times10^5$ et 4×10^5 g.mol^{-1}. La fraction totale représente l'ensemble des chaînes primaires et des chaînes polymériques, couvrant une gamme de $0,5\times10^5$ à 1×10^9 g.mol^{-1}. Au sein de la fraction protéique on distingue les polymères supérieurs à 1×10^7 g.mol^{-1} et les polymères inférieurs à cette masse molaire. Dans cette étude, la proportion des polymères de plus haute masse, constituent une caractéristique discriminante des gélatines étudiées (Figure 50 A' et B').

IV.3.2.(2). Etude du rayon de giration des polymères de gélatine

L'analyse de la diffusion de la lumière en cours d'élution a permis la détermination des rayons de giration des molécules séparées. La Figure 51 rend compte de l'évolution du logarithme de ce rayon de giration en fonction du logarithme de la masse molaire. En deçà de 1×10^6 g.mol^{-1} et au-delà de 1×10^9g.mol^{-1} les données ne sont pas interprétables. Selon les principes de diffusion de la lumière de Wyatt et Zimm (Zimm, 1948; Wyatt, 1993), une molécule d'un diamètre inférieur à 1/20ème de la longueur d'onde du faisceau lumineux monochromatique, diffuse de manière anisotrope. Ce qui rend la mesure du rayon de giration peu fiable. Pour un laser à 690nm, l'estimation du rayon de giration serait d'environ 15 nm, dans le cas de molécules de conformations régulières. Les données présentées montrent effectivement qu'en deçà de 15 nm de rayon de giration, soit moins de 1×10^6 g.mol^{-1}, les estimations ne sont plus interprétables. Cette figure montre que les réponses de l'évolution du rayon de giration en fonction de la masse molaire sont équivalentes quelle que soit la gélatine considérée (pentes comprises entre 0,35 et 0,50).

Les gélatines présentées ont été choisies pour couvrir la gamme complète de masse molaire jusqu'à 1×10^9 g.mol^{-1}, ce qui correspond à un rayon de giration maximum observé de 250 nm. La Figure 51 A présente des gélatines possédant de faibles et moyennes valeurs de force en gel alors que la Figure 51 B présente deux gélatines avec de hautes valeurs de force

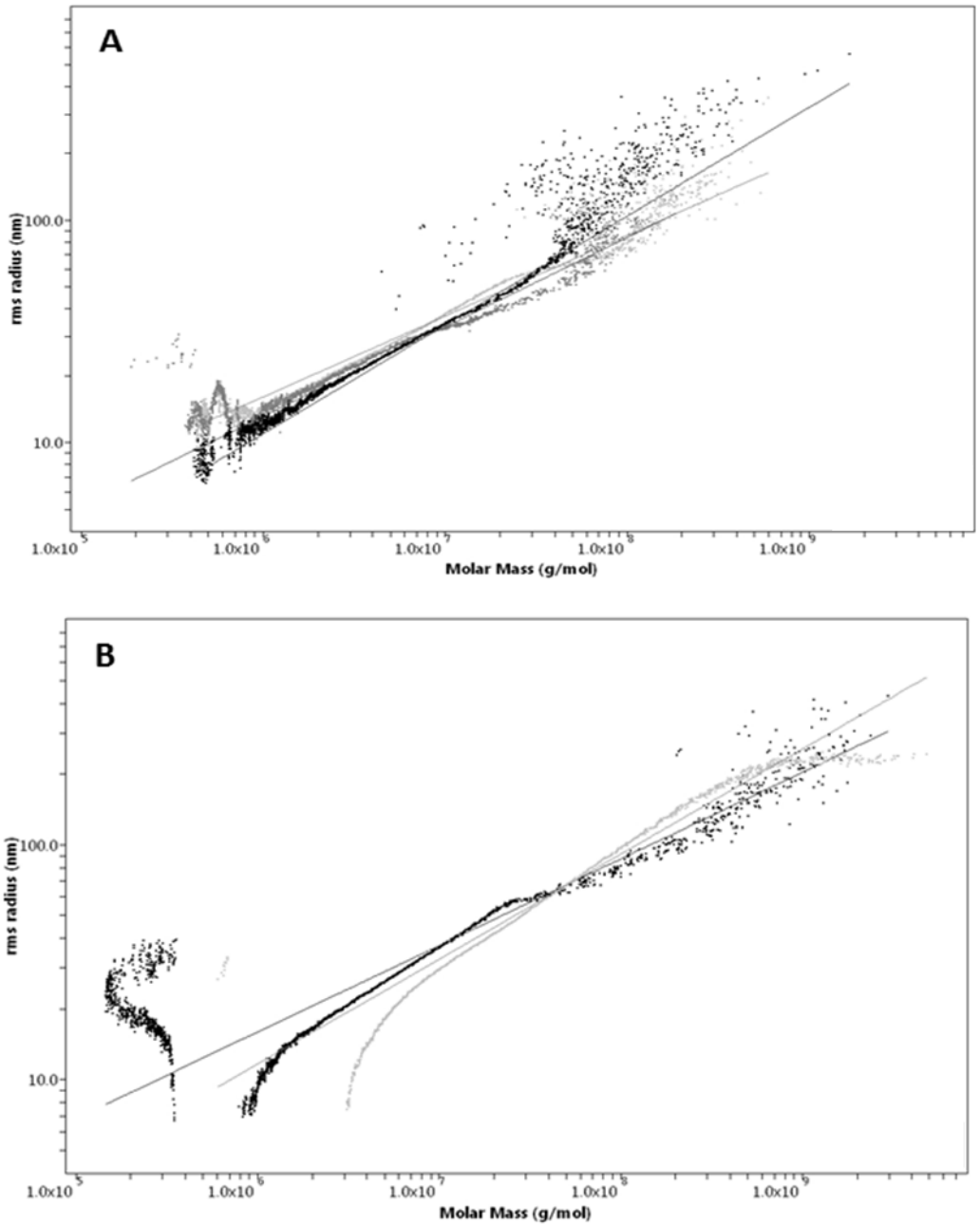


Figure 51 : RMS conformation plot pour des gélatines de faible et moyen bloom (A) et pour des gélatine de haut bloom (B).Sur le graphique A, il y a trois gélatine, en gris clair (force en gel de 200 g et viscosité 35,2 mp), en gris foncé (force en gel de 97 g et viscosité de 20,2 mp) et en noir (force en gel de 214 g et viscosité de 56,7 mp). Sur le graphique B, il y a deux gélatines, en gris clair (force en gel de 268 g et viscosité de 106,9 mp) et en noir (force en gel de 214 g et viscosité de 56,7 mp)

en gel.

Bien que les pentes soient similaires entre les différentes gélatines la gamme de masse molaire couverte par les gélatine n'est pas la même. Selon les travaux de *Wyatt* (Wyatt, 1993), les pentes observées sur la Figure 51 seraient caractéristiques d'un polymère possédant une conformation intermédiaire entre une structure fibrillaire (pente ≥ 0,5) et une structure globulaire (pente ≤ 0,33).

Les gélatines étudiées présentent des gammes de distribution molaire de polymères variables. Ces polymères ont probablement une structure identique, puisque les pentes des réponses du rayon de giration en fonction des masses molaire sont équivalentes.

On observe que l'information du rayon de giration nous indique que les gélatines, quelles qu'elles soient, ont une conformation similaire et intermédiaire entre la structure globulaire et la structure fibrillaire. En revanche, lorsque l'on change de gélatine et que l'on passe d'une gélatine avec une valeur de force en gel élevée à une force en gel faible, la pente reste identique, mais le nuage de points peut se décaler sur la gauche ou la droite du graphique. Ce qui signifie que la gamme de masse molaire varie. Les masses maximales et tous les rayons de girations maximums atteints pour chaque gélatine leurs seront spécifiques et seront à l'origine de propriétés de force en gel et de viscosité différentes.

IV.3.2.(3). Variabilité des distributions molaire des gélatines

Pour l'étude suivante, seul le signal UV est étudié. Une analyse en composante principale (ACP) a été réalisée sur les signaux UV des 49 gélatines. Ceci se fait à travers l'étude des premières composantes principales.

La Figure 52 présente, pour le signal UV mesuré à chaque pas du temps d'élution, la contribution des deux premières composantes de l'Analyse en Composantes Principale (ACP) réalisée sur les données des fractogrammes des 49 échantillons de gélatine. Les composantes principales représentent les sources de variabilités des données.

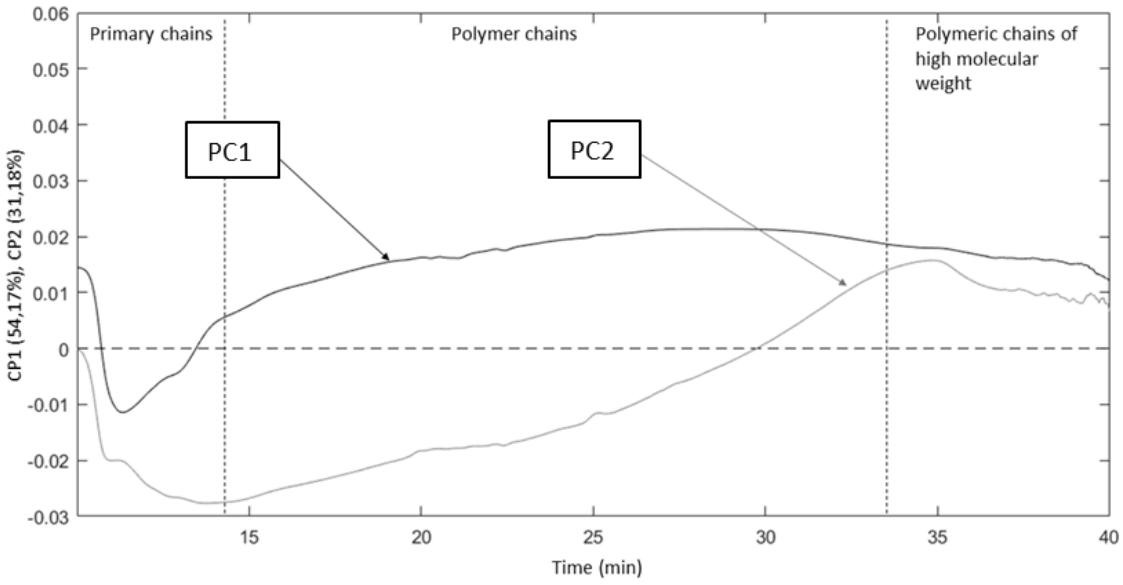


Figure 52 : Contribution des absorbances UV en fonction du temps d'élution aux composantes principales. En noir la PC1, en gris la PC2.

Les trois premières variables prennent en compte 90,13% de la variabilité du signal UV entre les gélatines. La PC1 explique 53,71% de la variation du signal UV. Les absorbances des molécules de faible masse molaire, éluées entre 10 et 15 minutes apparaissent comme peu contributives à cette composante. La contribution augmente progressivement avec l'élution de molécules de masse croissante.

La PC2 explique 31,76% de la variation du signal UV, il y a une corrélation négative avec les molécules de faible masse molaire et une corrélation positive pour les molécules de haute masse molaire (30 à 40 min). Cette variable prend en compte plus spécifiquement la contribution de la fraction des chaînes primaires et une contribution décroissante des chaînes polymériques de faible masse molaire ($<10^7$ g.mol^{-1}). En revanche il y a une contribution positive de la fraction des chaînes polymériques de très haute masse molaire ($>5.10^7$ g.mol^{-1})

Nous observons que ce n'est pas une zone particulière du fractogramme et donc une gamme particulière qui réagit pour modéliser la force en gel et la viscosité. C'est bien l'allure globale du fractogramme qui est utilisée, c'est-à-dire l'ensemble des chaînes primaires et des agrégats. Les molécules quelles que soient leurs tailles, influent pour l'établissement des propriétés physico-chimiques. Notre approche globale chimiométrique permet donc d'exploiter les informations du fractogramme différemment d'une approche classique d'étude de la distribution molaire (qui consisterait à intégrer un ou plusieurs pics).

Ces données rendent compte des deux sources majeures de la variabilité de distribution molaire des échantillons de gélatine analysés. La première, et la plus importante, concerne la variation de l'étendue de la gamme des masses des polymères de gélatine dans la plage de masse molaire de 4.10^5 g.mol^{-1} et 5.10^7 g.mol^{-1}. La seconde met en évidence l'opposition nette entre les échantillons de gélatine qui contiennent des fortes proportions en chaînes primaires et les échantillons comprenant de fortes proportions de chaînes polymériques de haute masse molaire.

IV.3.2.(4). Etudes des corrélations entre la force en gel (Bloom), la viscosité et les masses molaires des fractions séparées par AsFlFFF

Maintenant que cette étude nous a permis de mieux comprendre les différentes configurations moléculaires entre les gélatines, revenons à un autre niveau avec l'étude du lien entre cette configuration, représentée par les masses molaires des différentes fractions et les propriétés physico-chimiques (force en gel et viscosité).

Le Tableau 27 présente les résultats de matrice de similarité entre les valeurs de force en gel et de viscosité des gélatines, les valeurs déterminées de Mn et Mw et le pourcentage de fraction pour les fractions des chaînes primaires et des chaînes polymériques.

Les études dans la littérature ne permettent pas de conclure. En effet, il existe plusieurs sources pour expliquer la variabilité des propriétés physico-chimiques : la composition en acides aminés, la matière première, ou la nature du process (Schrieber et Gareis, 2007; Babel et al., 2000; Eysturskarð et al., 2009; Weng, Zheng, et Su, 2014). La majorité des travaux se sont concentrés sur l'étude des chaînes primaires en lien avec les propriétés physico-chimiques (Tromp, ten Grotenhuis, et Olieman, 2002; Meyer et Morgenstern, 2003)

Tableau 27 : Matrice de similarité (coefficient de corrélation de Pearson) entre la force en gel, la viscosité, les Mn (masse moyenne en nombre), les Mw (masse moyenne en masse) et les % de fraction pour les fractions des chaînes primaires (PRI) et des chaînes polymérique (POLY) pour les 49 échantillons.

	Bloom	**Viscosité**	**Mass fraction PRI (%)**	**Mass fraction POLY (%)**	**Mn PRI (kDa)**	**Mw PRI (kDa)**	**Mn POLY (kDa)**	**Mw POLY (kDa)**
Bloom	1							
Viscosité	0.613	1						
Mass fraction PRI (%)	-0,604	-0,429	1					
Mass fraction POLY (%)	0,503	0,325	-0,935	1				
Mn PRI (kDa)	**0.556**	**0.303**	-0,780	0,780	1			
Mw PRI (kDa)	**0.535**	**0.251**	-0,780	0,822	**0.976**	1		
Mn POLY (kDa)	**0.625**	**0.645**	-0,758	0,148	0.167	0.113	1	
Mw POLY (kDa)	**0.099**	**0.341**	-0,748	-0,003	0.128	0.126	**0.284**	1

Dans cette étude, la proportion de la fraction des chaînes primaires est négativement liée à la force en gel et à la viscosité. Inversement, la proportion de la fraction des chaînes polymériques est positivement liée aux propriétés physico-chimiques. La viscosité est moins directement liée à la proportion de chacune de ces fractions que ne l'est la force en gel.

Pour la population de faible masse molaire, les valeurs de Mn et Mw sont fortement corrélées (R=0.976), ce qui témoigne du fait que les 49 échantillons sont identiques en terme de composition de chaînes primaires. En revanche pour les polymères de haute masse molaire (POLY), les valeurs de Mn et Mw ne sont pas corrélées (R=0.284). En effet, pour deux échantillons différents des valeurs de Mn proches pour des valeurs de Mw très différentes traduit une composition en taille de fragment très différente. Les 49 gélatines sont donc différentes en termes de composition en agrégats de haute masse molaire. Les différences de distribution des polymères de haute masse molaire participent donc grandement à la différentiation de gélatines.

La force en gel a tendance à être reliée aux valeurs des Mn et Mw des chaînes primaires (R> 0,5). Pour cette même fraction, la viscosité n'est pas corrélée avec les Mn et Mw (R≤ 0,3). C'est sensiblement l'inverse que l'on observe pour le Mw de la fraction POLY, il n'y a ni corrélation avec les valeurs de force en gel (R=0,099), ni avec les valeurs de viscosité (R=0,341). En revanche, le Mn des chaînes polymériques présente des corrélations avec la force en gel (R=0.625) et la viscosité (R=0.645). Les chaînes POLY jouent donc un rôle important pour la force en gel et la viscosité. Le Mw est beaucoup plus sensible que le Mn aux valeurs extrêmes et donc moins fiable que la valeur de Mn.

Ces résultats permettent de confirmer le rôle privilégié supposé des fractions polymériques de haute masse molaire dans l'expression de la force en gel et de la viscosité des gélatines. Une forte proportion de chaînes primaires est au contraire défavorable à l'expression de la force en gel et de la viscosité.

IV.3.2.(5). Prédiction de la force en gel et de la viscosité, à partir de la distribution molaire.

Les fractogrammes UV entiers ont été utilisés pour prédire la force en gel et la viscosité. Les modèles de prédiction PLS sont présentés dans la Figure 53 et le Tableau 28. La distribution molaire telle que révélée par le signal UV des diagrammes d'élution est donc très explicative des propriétés technologiques (force en gel et viscosité). Les performances de ces modèles sont bien supérieures à celles des modèles basés uniquement sur certains pics du fractogramme.

Tableau 28 : Evaluation de la capacité de prédiction par PLS régression des propriétés physico-chimiques (force en gel, viscosité) à partir du signal UV des fractogrammes. Les résultats présentés un modèle construit avec 3 variables latentes.

	Calibration		**Validation croisée**	
	r^2cal	**RMSEC**	**r^2cv**	**RMSECV**
Force en gel	0,85	28,69	0,80	33,63
Viscosité	0,66	8,99	0,49	11,29
Log (viscosité)	0,79	0,074	0,69	0,091

La prédiction de la force en gel est meilleure que celle de la viscosité que ce soit en calibration (r^2cal=0,85 *vs* r^2call=0,79) ou en validation croisée (r^2cv=0,79 et r^2cv=0,69). Cela pourrait s'expliquer en partie par une perte de sensibilité de l'analyse de référence pour les fortes viscosités entraînant une réponse non proportionnelle (Figure 50 B). Ces constats sont cohérents avec ceux de Segtnan (Segtnan et al., 2003) qui avait observé une réponse non proportionnelle pour les fortes viscosités dans des modèles PLSr sur des gélatines par SPIR. C'est pourquoi une transformation logarithmique a permis d'améliorer le r^2 de calibration, de 0,66 pour les données brutes à 0,79 pour les données transformées et de 0,49 à 0,69 en validation croisée (Figure 53 C).

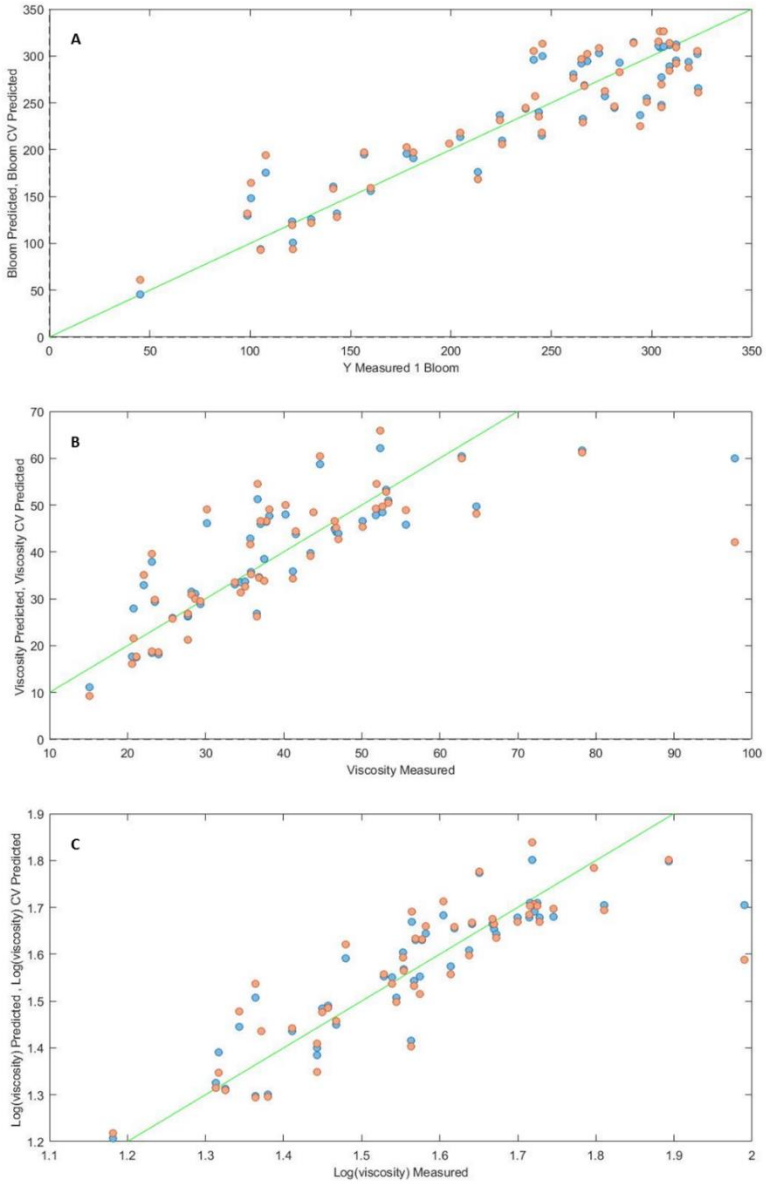


Figure 53 : Corrélation entre les valeurs prédites et les valeurs mesurées pour la force en gel (g) et la viscosité (mP). En bleu est présenté le jeu de données en calibration et en orange en validation croisée.

IV.3.2.(6). Conclusion

Les essais réalisés ont permis de mettre en évidence la capacité de la spectroscopie proche infrarouge à prédire la force en gel et le taux d'humidité de gélatines. En revanche, la prédiction quantitative de la viscosité n'est pas envisageable. Cependant, il serait intéressant d'essayer de prédire la viscosité en utilisant des classes ou en essayant des modèles non linéaire type *Artifical Neural Network* (ANN).

Les performances des modèles pour les prétraitements par dérivée première (pour les lissages < 21 points) et dérivée seconde sont très faibles, probablement dues à une perte d'information à cause de l'amplification du signal. En général, les meilleures performances ont été observées pour la transformation SNV detrend, mais les performances sont similaires pour les traitements SNV, brut et MSC.

Pour la force en gel, un modèle est applicable en fin de production, mais les performances sont moyennes (r^2cal= 0,79, RMSEC de 21,43 et RPD de 2,56 pour le meilleur modèle en SNVD). Cependant, un échantillonnage plus important et l'ajout d'échantillons de gélatine de faible force en gel, devraient améliorer les performances du modèle. C'est résultats sont comparables aux performances obtenues pour la prédiction du bloom (r^2_{cal} = 0,80) par *Segtnan et al (Segtnan et al., 2003)*.

Les modèles de prédiction du taux d'humidité présentent de très hautes performances (r^2cal= 0,95, RMSEC de 0,23 et RPD de 4,27 pour le meilleur modèle en SNVD). Ces essais ont été réalisés sur un appareil de laboratoire (MPA, Bruker Optics, Ettlingen, Germany), qui possède une haute résolution. Un modèle est donc applicable *at-line* avec une précision correcte, qui peut être encore améliorée par l'augmentation du nombre d'échantillons dans les bases de données. De surcroît, les performances globales laissent à penser qu'il est tout à fait envisageable d'appliquer la spectroscopie proche infrarouge avec un appareil de terrain pour suivre le processus *on-line* avec des résultats similaires.

Lors des processus de fabrication de la gélatine, le collagène est dénaturé et perd sa structure native. Ainsi la structure de la gélatine est différente de celle du collagène et elle provient de la reformation partielle des structures initiales du collagène en hélices. Comme nous l'avons montré dans la partie IV.3.2.(4), les gélatines commerciales présentent une grande variabilité de distribution molaire provenant directement du traitement industriel mais aussi sans doute de l'origine du collagène (El Harfaoui, 2007). Le set d'échantillon utilisé pour cet essai AsFlFFF a permis de tester une importante plage de force en gel et de viscosité.

Il existe une relation étroite entre les propriétés physico-chimiques (force en gel et viscosité), la concentration en gélatine et la structure du réseau qui se forme lors de la gélification. Il est notamment montré qu'une faible concentration en gélatine (<1% m/V), favorise une agrégation en pelotes des chaînes de gélatine, alors que pour des concentrations supérieures à 1%, la gélatine a tendance à s'organiser sous forme de nanostructures fibrillaires (Yang et Wang, 2009; Harrington et Rao, 1970). La force en gel et la viscosité des gélatines ont été étudiées à une concentration de 6,67% (m/V). A température ambiante (20°C) et pour ces concentrations la gélatine est sous forme de gel : réseau protéique stabilisé par des liaisons faibles, ce qui lui confère des propriétés viscoélastiques (Duconseille et al., 2015). Or, dans nos travaux la distribution molaire a été étudiée sur des solutions de gélatine à 0,2% (m/V), concentration pour laquelle la gélatine ne peut pas former un réseau structuré et donc se gélifier.

L'étude du rayon de giration a permis de mettre en avant une structure non-linéaire de la gélatine en solution que l'on peut associer à la formation de pelotes. De plus, elle a mis en évidence une forme similaire, quelle que soit la masse molaire pour les 49 échantillons de gélatines. Ces résultats sont en accord avec les travaux de *Guo et al*, qui proposaient une structure en hélice avec la formation de boucles pour la gélatine en solution (Guo et al., 2003). Nous avons observé une gamme de masses molaires très large allant jusqu'à 2000 fois la masse du trimère γ. On peut donc penser que les chaînes polymériques observées résultent soit de l'hydrolyse partielle du collagène et sont des restes de structures ramifiées par des liaisons covalentes, soit de réassociations stabilisées par liaisons faibles.

Par ailleurs, les polymères issus des gélatines avec lesquelles nous avons travaillé peuvent atteindre des niveaux de masse bien supérieurs à ceux observés dans les mêmes conditions par *Rbii et al.* (Rbii et al., 2009) et *Fraunhofer et al*. (Fraunhofer, Winter, et Coester, 2004), de l'ordre de 1×10^{7}g.mol^{-1}. Les distributions de masse molaire sont conformes aux observations de ces auteurs, cependant les gélatines sont très différentes au niveau des chaines polymériques (ou agrégats).

Il semblerait que ces différences de masses molaires maximales résultent des processus d'extraction différents entre ceux utilisés par ces auteurs et ceux utilisés dans nos travaux. En effet certaines gélatines de cette étude correspondent à des étapes différentes de la succession d'extraction conduisant à une dégradation des propriétés physico-chimiques. L'AsFlFFF permet donc de discriminer les échantillons de gélatine entre eux, principalement grâce à l'étude des chaînes de haute masse molaire.

L'analyse des corrélations entre les fractions de la gélatine séparées par AsFlFFF et les propriétés viscoélastiques des gels formés met clairement en évidence l'importance des chaînes polymériques dans l'acquisition de force en gel et de viscosité supérieures.

Conclusion générale

Ces travaux de thèse avaient pour objectif de répondre à des problématiques industrielles liées à la production de gélatine et ont nécessité à la fois des connaissances en chimie, en spectroscopie et en procédés de production. Ils ont abouti à deux types de proposition pour l'industriel visant à supprimer la subjectivité dans la qualification des matières premières *via* l'automatisation des analyses en amont de la chaîne de production et à optimiser le suivi de la production jusqu'au produit fini.

Le chapitre II présente une étude en laboratoire basée sur la caractérisation chimique de matières premières hétérogènes en mouvement par spectroscopie proche infrarouge. Il avait pour objectif de répondre à la première hypothèse portant sur l'acquisition en ligne et l'évaluation de la capacité de discrimination spectrale par spectroscopie proche infrarouge. Ce travail a permis de mettre en place une phase d'expérimentation à l'échelle industrielle (Chapitre III), liée à la caractérisation des matières premières en termes de rendement. Cette phase s'est décomposée en deux niveaux ; une étape d'installation en ligne et une étude de la faisabilité en situation industrielle. Par ailleurs, une étude de la caractérisation de la gélatine sur deux niveaux a été réalisée et a été présentée dans le chapitre IV. Des gélatines ont été caractérisées en termes de force en gel, de viscosité et de taux d'humidité. Ces gélatines ont ensuite été analysées soit par *Asymmetrical Flow Field-Flow Fractionation,* pour étudier la distribution molaire, soit par spectroscopie proche infrarouge afin d'évaluer la capacité de prédiction des propriétés physico-chimiques des gélatines.

Pour l'étude à l'échelle laboratoire sur les matières premières, des couennes de porc ont été considérées comme un solide divisé (fragments provenant d'individus différents). Cette étude introduit la première problématique portant sur la sélection des spectres d'intérêt, ceux associés aux matières premières. Les essais réalisés ont montré la capacité de discrimination des spectres provenant des couennes, des spectres provenant du tapis de convoyage ou de l'interface tapis-couennes par une analyse en composantes principales (ACP). En plus de ces capacités, les spectres provenant de la face cuir ou graisse ont également pu être discriminés par des modèles ACP couplés avec une analyse linéaire discriminante (LDA) (91% de précision).

La caractérisation chimique des couennes a mis en évidence des corrélations entre le taux de matière sèche, le taux de protéine et le taux de matière grasse. Le taux de collagène ne présente pas de corrélation avec les autres analyses de référence.

L'étude spectrale a permis de mettre en évidence que la plupart des bandes spectrales qui réagissent sont liées aux vibrations des liaisons C-H des matières grasses ou des OH de l'eau (Cozzolino et Murray, 2004). C'est principalement la zone 1100 nm - 1250 nm, correspondant aux secondes harmoniques des liaisons C-H des matières grasses (Bertrand, 2002), qui contribue à la discrimination des échantillons. Ces observations sont en adéquation avec les résultats observés pour les différents modèles de prédiction, dans lesquels il y a une augmentation des performances pour les modèles développés sur la face graisse.

Cette étude a également montré une amélioration des performances des modèles de prédiction après avoir réalisé une moyenne spectrale des échantillons. Les modèles développés avec les spectres moyens permettent d'envisager l'utilisation de ces modèles pour prédire le taux de matière sèche, le taux de protéine et le taux de matière grasse. Les meilleurs modèles sont obtenus pour la face graisse après prétraitement dérivée seconde pour la prédiction du taux de matière sèche (r^2_{cal} = 0,78, RMSEC = 3,7% et RPD = 2,41) et du taux de protéine (r^2_{cal} = 0,79, RMSEC = 4,6% et RPD = 2,86). Pour le taux de matière grasse, le meilleur modèle est observé pour la face graisse après prétraitement SNVD (r^2_{cal} = 0,75, RMSEC = 5,3% et RPD = 2,36). Ces performances sont comparables aux performances obtenues en 2010 par *Liao et al.* sur la prédiction du taux de protéine, du taux d'humidité et du taux de matière grasse dans du muscle (Liao, Fan, et Cheng, 2010). Par ailleurs, il y a une amélioration encore plus importante des performances des modèles de prédiction lorsque les valeurs de prédiction de chaque spectre sont moyennées (augmentation moyenne de 0,1 du r^2). Ces observations suggèrent donc de moyenner les prédictions d'un lot de spectres correspondant à un échantillon pour approcher au maximum la valeur réelle, plutôt que de moyenner les informations spectrales, qui auraient tendance à induire une perte d'information.

Au cours de ce travail d'analyse de données, deux indicateurs ont été utilisés, l'indice de qualité du signal (*Signal Quality Index*, SQi) et le lambda de Wilks. Le lambda de Wilks est un indicateur de la variabilité qui permet de mesurer la variabilité intra échantillon par rapport à la variabilité totale. Cette mesure a permis de mettre en évidence la capacité à discriminer un échantillon par rapport à l'ensemble des échantillons. Le SQi, quant à lui, permet de mesurer la part spectrale utile, celle qui serait liée à la composition chimique. Cet indicateur a permis d'observer les mêmes tendances que celles observées par l'analyse des modèles de prédiction. Cela a permis de mettre en évidence une part spectrale utile plus importante pour la face graisse que pour la face cuir et d'indiquer les prétraitements les plus intéressants. Cet indicateur permet donc d'appréhender en général le comportement des données en modélisation en nécessitant moins de puissance de calcul que la modélisation elle-même.

Cette étude de la caractérisation par SPIR de couennes de porc en mouvement a permis de mieux comprendre les problématiques liées à la variabilité des matières premières et à l'acquisition en ligne, afin de pouvoir la transposer à l'échelle industrielle. Cela a permis de sélectionner les spectres d'intérêt sur un flux discontinu, jusqu'à sélectionner uniquement les spectres provenant de la face graisse. L'étude sur le lambda de Wilks a permis de mesurer la variabilité intra échantillon par rapport à la variabilité totale des échantillons, confortant la capacité de discrimination d'un échantillon par rapport à l'ensemble des échantillons. Par ailleurs, ces travaux ont mis en évidence l'importance de l'information spectrale reliée à la présence de matière grasse. La présence de graisse est inversement proportionnelle à la présence de protéine, les protéines étant directement liées au rendement, les matières grasses sont donc elles aussi liées au rendement. Une prédiction du rendement est donc envisageable. Pour l'application industrielle, nous avons fait l'hypothèse que l'ensemble des fragments de couenne, acquis en situation industrielle (75 tonnes) et donc statistiquement acquis autant sur la face cuir que sur la face graisse, peut correspondre à une couenne virtuelle présentant une certaine variabilité.

Par la suite, une application à l'échelle industrielle a été étudiée, rapidement confrontée à la réalité industrielle. Le positionnement en amont de la chaîne de production et le choix du spectromètre proche infrarouge ont été les premiers sujets de réflexion.

Le choix du spectromètre s'est porté sur un spectromètre Polytec (PSS-2121 avec une sonde à distance PSS-H-A03) permettant d'analyser une surface importante (Ø=3,5 cm) et à distance, avec une technologie de barrette à photodiode résistante aux conditions industrielles. Pour le choix du positionnement, deux points ont été envisagés ; sur le tapis de convoyage ou sur le bras de distribution des matières premières. Cependant, le tapis de convoyage est positionné vers une ouverture directement reliée à l'extérieur et est donc sujet à des variations de température et d'humidité plus importantes. De plus, les matières premières sont introduites dans la chaîne de production aussi bien sous forme « fraiche » que « congelée » et souvent sous forme d'amas au niveau du tapis. En revanche, le bras de distribution se trouve à la sortie d'un broyeur qui homogénéise davantage les couennes et est moins sujet aux variations atmosphériques. Le choix s'est donc porté sur le bras de distribution. Les conditions d'acquisition au niveau du bras sont différentes des conditions d'acquisition utilisées en laboratoire, il n'est donc pas possible de discriminer les spectres de la face cuir des spectres de la face graisse.

Le projet a ensuite été confronté à des verrous techniques liés à l'installation de la sonde et au fonctionnement de l'entreprise. Une plateforme a été installée au niveau du bras de distribution, avec un accès au réseau, des alimentations électriques, un boitier pour le spectromètre et un module coulissant pour positionner la sonde en la laissant accessible aux opérateurs. Un protocole d'acquisition lié aux Ordres de Fabrication a été développé en lien avec le logiciel de contrôle du processus industriel, permettant le lancement de l'acquisition automatique au début du chargement des 75 tonnes et un arrêt par bouton (positionné au niveau des opérateurs) à la fin du chargement. A terme, une automatisation de l'arrêt est envisageable. Les spectres sont enregistrés par heure et par jour dans des nouveaux fichiers. L'utilisation de l'Analyse en Composantes Principales pour trier les spectres n'a pas pu être effectuée à cause de la puissance de modélisation pour cette quantité de données. A la place, un modèle de vérification de la similarité, à la longueur d'onde 1250 nm avec un paramètre strict, a été utilisé (seuil de cosinus = 0,985). Il est possible que ce seuil induise une perte de variabilité. A ce jour, la collecte des informations est opérationnelle, l'automatisation de la prédiction est possible, mais n'est pas encore réalisée.

Cette démarche a permis d'évaluer la capacité de prédiction des matières premières à partir d'un jeu de données de 71 échantillons correspondant à 75 tonnes de matières premières. Dans un premier temps, des modèles de prédiction quantitatifs ont été envisagés, mais sans permettre de construire des modèles performants et ce quel que soit la structure des données et les prétraitements utilisés. Par la suite, des analyses de classification basées sur des modèles linéaires ont été réalisées, permettant de mettre en évidence une tendance pour classer les données. Les meilleurs résultats ont été obtenus par ACP couplée à l'analyse discriminante (60% d'échantillons bien classés en validation externe avec la dérivée première), avec une répartition en 4 classes par amplitude de rendement. L'utilisation d'un nombre inférieur de classes n'a pas permis d'améliorer les performances des modèles.

Pour donner suite à ces observations, plusieurs hypothèses ont été développées permettant d'expliquer en partie les performances obtenues en situation industrielle. La première concerne la perte d'informations spectrales. L'essai réalisé en laboratoire met en évidence l'amélioration de la prédiction quand la valeur est prédite par spectre par rapport à la moyenne spectrale. Le spectre moyen généré en condition industrielle sur 75 tonnes de produit entraîne donc lui aussi une perte d'information liée à l'hétérogénéité du lot. La deuxième hypothèse est liée à la fiabilité de la valeur de référence. En effet, les différences de performances entre les analyses en classification et en quantification, laissent suggérer un manque de précision de l'estimation du rendement. Elle peut être liée à l'erreur de la mesure en réfractométrie, ou encore à l'erreur de la mesure du volume des cuves. La dernière hypothèse concerne l'utilisation de modèles linéaires qui pourraient ne pas correspondre à la réalité d'une couenne. Ainsi, on peut penser que la prise en compte de la seule dimension linéaire est insuffisante pour refléter la réalité de la dispersion de l'hétérogénéité des couennes. Il serait intéressant d'étudier les données issues d'un nombre supérieur d'échantillons avec des approches non-linéaires de type réseau de neurone ou analyse discriminante quadratique.

Le dernier volet de cette thèse concerne l'étude du produit fini, la gélatine. Celle-ci est définie principalement par ses propriétés physico-chimiques, la force en gel et la viscosité. Ces propriétés sont déterminées sur un gel à 6,67% (m/V) à une température spécifique (10°C pour le bloom et 60°C pour la viscosité), rendant cette mesure empirique. Dans ce contexte, une approche différente pour étudier 49 gélatines commerciales en termes de distribution molaire a été réalisée. Les gélatines ont été analysées par *Asymmetrical Flow Field-Flow Fractionation*, dans l'objectif d'étudier les relations entre leur distribution molaire et leurs propriétés physico-chimiques.

La caractérisation de la distribution molaire des gélatines par AsFlFFF a permis d'étudier les proportions des chaînes primaires (entre 0.5×10^5 et 4×10^5 $g.mol^{-1}$) et des chaînes polymériques (>4×10^5 $g.mol^{-1}$), en fonction de leurs propriétés physico-chimiques. On observe une augmentation de la proportion de chaînes polymériques pour les gélatines avec des valeurs de force en gel et de viscosité élevées. Les gélatines étudiées couvrent une gamme moléculaire de 0.5×10^5 à 1×10^9 $g.mol^{-1}$ présentant de très gros agrégats en comparaison à ceux étudiés (1×10^7 $g.mol^{-1}$) dans les travaux réalisés par *Rbii et al.* (Rbii et al., 2009). Les corrélations entre les données de masse molaire (Mn et Mw) et les propriétés physico-chimiques (force en gel et viscosité) semblent être liées entre elles. Cela confirme le rôle privilégié des fractions polymériques pour l'expression de la force en gel et de la viscosité. La conformation des gélatines a été étudiée par l'étude du rayon de giration, permettant de révéler une structure intermédiaire entre une structure fibrillaire et une structure en pelote. Les gélatines ont présenté des gammes de masse molaire différentes en fonction des propriétés physico-chimiques, mais quelle que soit la gamme de masse molaire, la structure conserve cette conformation intermédiaire. L'analyse chimiométrique du signal UV des différentes gélatines provenant de l'analyse par AsFlFFF a révélé la relation entre les proportions des différentes masses molaires des gélatines et la force en gel ($r^2_{\text{validation croisée}}$ =0,80). La relation est moins évidente pour la viscosité notamment pour les gélatines qui possèdent des proportions importantes de chaînes polymériques de hautes masses molaires. Or, la viscosité est d'autant plus importante que les molécules qui la composent sont grosses, permettant de négliger le rôle des chaînes primaires. La force en gel est liée à la capacité de l'ensemble des molécules présentes, à former un réseau, alors que pour la viscosité, l'écoulement est freiné par l'interaction plus ou moins intense entre les molécules.

La chapitre IV a abordé par ailleurs la capacité de prédiction de la force en gel, la viscosité et l'humidité des gélatines industrielles. Cette étude a permis de mettre en évidence des performances de prédiction de la force en gel correctes (meilleur modèle r^2_{cal}=0,79, RMSEC de 21,43 g et RPD =2,56), permettant d'envisager la prédiction de ce paramètre. En revanche, les performances des modèles pour la prédiction de la viscosité sont moyennes, ne permettant pas d'envisager la prédiction de ce paramètre par des modèles linéaires. Le modèle de prédiction du taux d'humidité présente de hautes performances (meilleur modèle r^2_{cal}=0,95, RMSEC de 0,23% et RPD =2,56), laissant envisager l'utilisation de la spectroscopie proche infrarouge pour prédire ce paramètre. L'utilisation de la spectroscopie proche infrarouge pour l'humidité et la force en gel peut permettre une analyse plus rapide que les méthodes de références.

Ces travaux de thèse ont finalement permis de répondre à des problématiques industrielles à deux niveaux d'intervention (Figure 54), la première concerne les matières premières et la seconde la caractérisation de la gélatine : (i) la caractérisation des matières premières peut permettre d'avoir un paramètre objectif de la qualité de celles-ci, afin d'adapter le processus de production et d'affiner le cahier des charges des fournisseurs, (ii) la caractérisation par SPIR des gélatines en terme de force en gel peut apporter un gain de temps aux industriels, (iii) la prédiction de l'humidité pourrait permettre un suivi en direct du processus de séchage afin de cibler plus précisément la valeur désirée, et (iv) le développement de l'utilisation de l'AsFlFFF pour analyser la gélatine pourrait contribuer à évaluer l'état structural de cette dernière de façon moins empirique.

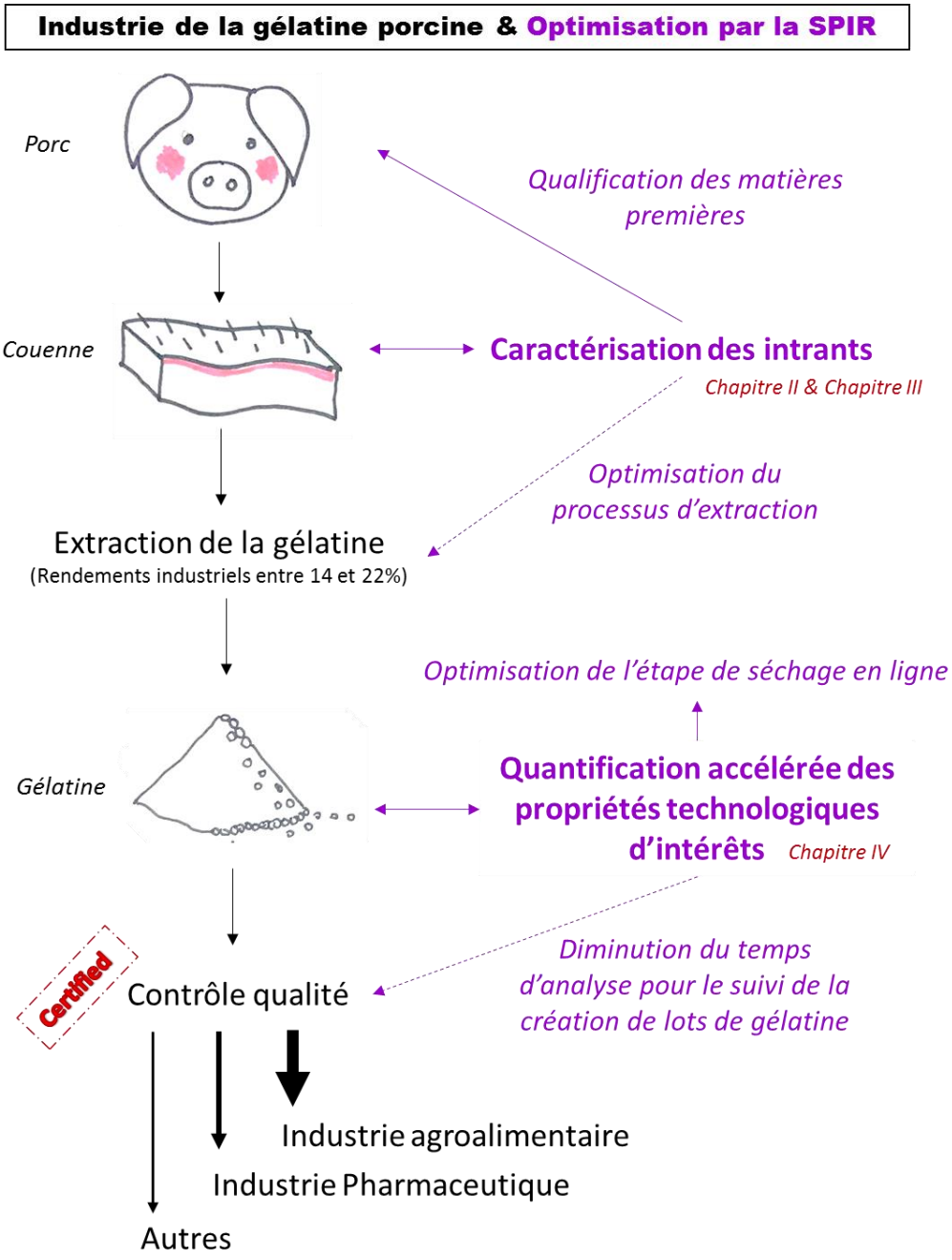


Figure 54 : Schéma de l'interaction entre l'industrie de la gélatine porcine et les améliorations potentielles apportées par la SPIR aux différents états du produit

Références bibliographiques

AFNOR, 1982a. Edible gelatine. Specifications. In *NF V59-001*, edited by AFNOR.

———, 1982b. Gélatine alimentaire - Détermination de l'humidité et des cendres. NF V 59-003

———, 2001. Meat, meat products and fishery products - Determination of free fat content. NF V04-403

———, 2002. Meat, meat product and fishery products - Determination of L-hydroxyproline (collagen) content. NF V04-403

Agelet L.E., Hurburgh C.R., 2010. A Tutorial on Near Infrared Spectroscopy and Its Calibration. Critical Reviews in Analytical Chemistry 40, 246-60.

Ahmad T., Ismail A., Ahmad S.A., Khalil K.A., Kumar Y., Adeyemi K.D., Sazili A.Q., 2017. Recent advances on the role of process variables affecting gelatin yield and characteristics with special reference to enzymatic extraction: A review. Food Hydrocolloids 63, 85-96.

Armisén R., Galatas F., 2009. 4 - Agar. in G. O. Phillips and P. A. Williams (eds.), Handbook of Hydrocolloids (Second edition) (Woodhead Publishing).

Arvanitoyannis I.S., 2009. *HACCP and ISO 22000: Application to foods of animal origin*. (John Wiley & Sons).

Babel W., Schulz D., Giesen-Wiese M., Seybold U., Gareis H., Dick E., Schrieber R., Schott A., Stein W., 2000. Gelatin. Ullmann's Encyclopedia of Industrial Chemistry

Bailey A.J., Light N.D., 1989. *Connective tissue in meat and meat products*. (Elsevier Applied Science).

Balage J.M., da Luz e Silva S., Gomide C.A., Bonin M.d.N., Figueira A.C., 2015. Predicting pork quality using Vis/NIR spectroscopy. Meat Science 108, 37-43.

Banyard J., Bao L., Hofer M.D., Zurakowski D., Spivey K.A., Feldman A.S., Hutchinson L.M., Kuefer R., Rubin M.A., Zetter B.R., 2007. Collagen XXIII expression is associated with prostate cancer recurrence and distant metastases. Clinical Cancer Research 13, 2634-42.

Barlocco N., Vadell A., Ballesteros F., Galietta G., Cozzolino D., 2006. Predicting intramuscular fat, moisture and Warner-Bratzler shear force in pork muscle using near infrared reflectance spectroscopy. Animal Science 82, 111-16.

Benincasa M.A., Giddings J.C., 1992. Separation and molecular weight distribution of anionic and cationic water-soluble polymers by flow field-flow fractionation. Analytical chemistry 64, 790-98.

Bertrand D., 2002. La spectroscopie proche infrarouge et ses applications dans les industries de l'alimentation animale. Productions animales 15, 209-19.

Bertrand D., Dufour E., 2006a. *La spectroscopie infrarouge*. (Lavoisier).

Bertrand D., Dufour E., 2006b. La spectroscopie proche infrarouge et ses applications analytiques. Collection Sciences et Technologiques Agroalimentaires, Tec et Doc, Lavoisier, Paris

Bowen R., Mark H., 1989. *Encyclopedia of polymer science and engineering*. (Sl).

Brinckmann J., Notbohm H., Müller P.K., 2005. *Collagen: primer in structure, processing and assembly*. (Springer Science & Business Media).

Brøndum J., Munck L., Henckel P., Karlsson A., Tornberg E., Engelsen S.B., 2000. Prediction of water-holding capacity and composition of porcine meat by comparative spectroscopy. Meat Science 55, 177-85.

Bruckner P., 2010. Suprastructures of extracellular matrices: paradigms of functions controlled by aggregates rather than molecules. Cell and tissue research 339, 7.

Büning-Pfaue H., 2003. Analysis of water in food by near infrared spectroscopy. Food Chemistry 82, 107-15.

Caldwell K.D., Kesner L.F., Myers M.N., Giddings J.C., 1972. Electrical field-flow fractionation of proteins. Science 176, 296-98.

Čandek-Potokar M., Prevolnik M., Škrlep M., 2006. Ability of near infrared spectroscopy to predict pork technological traits. Journal of Near Infrared Spectroscopy 14, 269-77.

Chagnot C., Listrat A., Astruc T., Desvaux M., 2012. Bacterial adhesion to animal tissues: protein determinants for recognition of extracellular matrix components. Cellular microbiology 14, 1687-96.

Chan D., Walker P., Mills E., 2002. Prediction of pork quality characteristics using visible and near–infrared spectroscopy. Transactions of the ASAE 45, 1519.

Chen J., Zhu S., Zhao G., 2017. Rapid determination of total protein and wet gluten in commercial wheat flour using siSVR-NIR. Food Chemistry 221, 1939-46.

Chen Q., Cai J., Wan X., Zhao J., 2011. Application of linear/non-linear classification algorithms in discrimination of pork storage time using Fourier transform near infrared (FT-NIR) spectroscopy. LWT-Food Science and Technology 44, 2053-58.

Chou M.-Y., Li H.-C., 2002. Genomic organization and characterization of the human type XXI collagen (COL21A1) gene. Genomics 79, 395-401.

CIE, 1978. International Commission on Illumination, Recommendations on uniform color spaces. In, edited by Compagnie Internationale. Supplement.

Contado C., Ravani L., Passarella M., 2013. Size characterization by sedimentation field flow fractionation of silica particles used as food additives. Analytica Chimica Acta 788, 183-92.

Cozzolino D., Barlocco N., Vadell A., Ballesteros F., Gallieta G., 2003. The use of visible and near-infrared reflectance spectroscopy to predict colour on both intact and homogenised pork muscle. LWT-Food Science and Technology 36, 195-202.

Cozzolino D., De Mattos D., Martins D.V., 2002. Visible/near infrared reflectance spectroscopy for predicting composition and tracing system of production of beef muscle. Animal Science 74, 477-84.

Cozzolino D., Murray I., 2004. Identification of animal meat muscles by visible and near infrared reflectance spectroscopy. LWT - Food Science and Technology 37, 447-52.

De Claubry G., 1831. *Notice sur l'extraction de la gélatine des os et sur son application à la nourriture des pauvres: lu à la société univers. pour la propagation des sciences et de l'industrie dans la séance du 4. Mai 1831*.

Draget K.I., 2009. 29 - Alginates. in G. O. Phillips and P. A. Williams (eds.), Handbook of Hydrocolloids (Second edition) (Woodhead Publishing).

Duconseille A., Astruc T., Quintana N., Meersman F., Sante-Lhoutellier V., 2015. Gelatin structure and composition linked to hard capsule dissolution: a review. Food Hydrocolloids 43, 360-76.

Duconseille A., Gaillard C., Santé-Lhoutellier V., Astruc T., 2017. Molecular and structural changes in gelatin evidenced by Raman microspectroscopy. Food Hydrocolloids

Duconseille A., Traikia M., Lagrée M., Jousse C., Pagès G., Gatellier P., Astruc T., Santé-Lhoutellier V., 2017. The impact of processing and aging on the oxidative potential, molecular structure and dissolution of gelatin. Food Hydrocolloids 66, 246-58.

Duconseille A., Wien F., Audonnet F., Traore A., Refregiers M., Astruc T., Santé-Lhoutellier V., 2017. The effect of origin of the gelatine and ageing on the secondary structure and water dissolution. Food Hydrocolloids 66, 378-88.

Dupont A.-L., 2002. Study of the degradation of gelatin in paper upon aging using aqueous size-exclusion chromatography. Journal of Chromatography A 950, 113-24.

Edward J.T., 1970. Molecular volumes and the Stokes-Einstein equation. J. chem. Educ 47, 261.

El Harfaoui N., 2007. Gels et films de gélatine: structure et propriétés fonctionnelles. Thesis, Paris 7.

Engel J., Bächinger H.P., 2005. Structure, stability and folding of the collagen triple helix. Collagen 7-33.

Esbensen K.H., Guyot D., Westad F., Houmoller L.P., 2002. *Multivariate Data Analysis - in Practice: An Introduction to Multivariate Data Analysis and Experimental Design*. (Camo Process AS).

European-norm, 2002. NF EN 12092 : Adhesives - Determination of viscosity. In, ICS : 83.180.

Eysturskarð J., Haug I.J., Elharfaoui N., Djabourov M., Draget K.I., 2009. Structural and mechanical properties of fish gelatin as a function of extraction conditions. Food Hydrocolloids 23, 1702-11.

Flory P.J., Weaver E.S., 1960. Helix [unk] coil transitions in dilute aqueous collagen solutions1. Journal of the American Chemical Society 82, 4518-25.

Fox M.A., 2008. Novel roles for collagens in wiring the vertebrate nervous system. Current opinion in cell biology 20, 508-13.

Fraunhofer W., Winter G., Coester C., 2004. Asymmetrical flow field-flow fractionation and multiangle light scattering for analysis of gelatin nanoparticle drug carrier systems. Analytical chemistry 76, 1909-20.

Geesink G., Schreutelkamp F., Frankhuizen R., Vedder H., Faber N., Kranen R., Gerritzen M., 2003. Prediction of pork quality attributes from near infrared reflectance spectra. Meat Science 65, 661-68.

Gelatine-Manufacturers-Europe, http://www.gelatine.org/infocenter/publications-downloads.html.

Gelse K., Pöschl E., Aigner T., 2003. Collagens—structure, function, and biosynthesis. Advanced drug delivery reviews 55, 1531-46.

Gevrey M., Dimopoulos I., Lek S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecological modelling 160, 249-64.

Gibot L., Lemelle A., Till U., Moukarzel B.a., Mingotaud A.-F.o., Pimienta V.r., Saint-Aguet P., Rols M.-P., Gaucher M., Violleau F.d.r., 2014. Polymeric micelles encapsulating photosensitizer: structure/photodynamic therapy efficiency relation. Biomacromolecules 15, 1443-55.

Giddings J.C., 1966. A new separation concept based on a coupling of concentration and flow nonuniformities. Separation Science 1, 123-25.

———, 1995. Measuring colloidal and macromolecular properties by FFF. Analytical chemistry 67, 592A-98A.

Giddings J.C., Yang F.J., Myers M.N., 1976. Theoretical and experimental characterization of flow field-flow fractionation. Analytical chemistry 48, 1126-32.

———, 1977. Flow field-flow fractionation as a methodology for protein separation and characterization. Analytical biochemistry 81, 395-407.

Givens D.I., De Boever J.L., Deaville E.R., 1997. The principles, practices and some future applications of near infrared spectroscopy for predicting the nutritive value of foods for animals and humans. Nutrition Research Reviews 10, 83-114.

Gobrecht A., Bellon-Maurel V., 2014. Spectroscopie proche infrarouge comptabilité carbone au champ,Irstea, UMR ITAP

Gobrecht A., Roger J.-M., Bellon-Maurel V., 2014. Major issues of diffuse reflectance NIR spectroscopy in the specific context of soil carbon content estimation: a review. Adv. Agron 123, 145-75.

Gómez-Guillén M., Giménez B., López-Caballero M.a., Montero M., 2011. Functional and bioactive properties of collagen and gelatin from alternative sources: A review. Food Hydrocolloids 25, 1813-27.

Gorham S.D., 1991. Collagen. in, Biomaterials (Springer).

Grand-View-Research, 2013. *Gelatin Market analysis*.

Grimal S., Puech S., Wagener R., Ventéo S., Carroll P., Fichard-Carroll A., 2010. Collagen XXVIII is a distinctive component of the peripheral nervous system nodes of ranvier and surrounds nonmyelinating glial cells. Glia 58, 1977-87.

Guo L., Colby R.H., Lusignan C.P., Whitesides T.H., 2003. Kinetics of triple helix formation in semidilute gelatin solutions. Macromolecules 36, 9999-10008.

Haaland D.M., Thomas E.V., 1988. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. Analytical chemistry 60, 1193-202.

Handbook G.G., 2012. Gelatin Manufacturers Institute of America. In.: Inc.

Harrington W.F., Rao N.V., 1970. Collagen structure in solution. I. Kinetics of helix regeneration in single-chain gelatins. Biochemistry 9, 3714-24.

Has C., Kern J.S., 2010. Collagen XVII. Dermatologic clinics 28, 61-66.

Hjorten R., Hansen U., Underwood R.A., Telfer H.E., Fernandes R.J., Krakow D., Sebald E., Wachsmann-Hogiu S., Bruckner P., Jacquet R., 2007. Type XXVII collagen at the transition of cartilage to bone during skeletogenesis. Bone 41, 535-42.

Hofman K., Hall B., Cleaver H., Marshall S., 2011. High-throughput quantification of hydroxyproline for determination of collagen. Analytical biochemistry 417, 289-91.

Hoving-Bolink A., Vedder H., Merks J., De Klein W., Reimert H., Frankhuizen R., Van Den Broek W., 2005. Perspective of NIRS measurements early post mortem for prediction of pork quality. Meat Science 69, 417-23.

Isaksson T., Nilsen B., Tøgersen G., Hammond R., Hildrum K., 1996. On-line, proximate analysis of ground beef directly at a meat grinder outlet. Meat Science 43, 245-53.

ISO, 2005. Animal feeding stuffs - Determination of nitrogen content and calculation of crude protein content - Part 1 : Kjeldahl method. In *5983-1*, edited by ISO NF EN.

———, 2007. Optique et photonique - Bandes spectrales. In *20473*.

Jamilah B., Harvinder K.G., 2002. Properties of gelatins from skins of fish—black tilapia (Oreochromis mossambicus) and red tilapia (Oreochromis nilotica). Food Chemistry 77, 81-84.

Johnston-Banks F., 1990. Gelatine. in, Food gels (Springer).

Kalinin A., Krasheninnikov V., 2017. Detection of fatty product falsifications using a portable near infrared spectrometer. In *EPJ Web of Conferences*, 02009. EDP Sciences.

Kamishikiryo Y., H., Oritani, Y., Takamura, H. and Matoba, T.(1994). Protein content in milk by near infrared spectroscopy. J. Food Sci 59, 313-15.

Kapper C., 2012. Rapid prediction of pork quality. Wageningen University.

Kasai M., Yasuda Y., Mizoguchi H., Soga K., Kaneko K., Takemura H., 2017. In vivo tumor wavelength band selection using Hierarchical clustering and PCA with NIR-Hyperspectral Data. Journal of Biomedical Engineering and Medical Imaging 4, 01.

Koch M., Schulze J., Hansen U., Ashwodt T., Keene D.R., Brunken W.J., Burgeson R.E., Bruckner P., Bruckner-Tuderman L., 2004. A novel marker of tissue junctions, collagen XXII. Journal of Biological Chemistry 279, 22514-21.

Krämer K., Ebel S., 2000. Application of NIR reflectance spectroscopy for the identification of pharmaceutical excipients. Analytica Chimica Acta 420, 155-61.

Lafhal S., Vanloot P., Bombarda I., Kister J., Dupuy N., 2016. Chemometric analysis of French lavender and lavandin essential oils by near infrared spectroscopy. Industrial Crops and Products 80, 156-64.

Leroy B., Lambotte S., Dotreppe O., Lecocq H., Istasse L., Clinquart A., 2004. Prediction of technological and organoleptic properties of beef Longissimus thoracis from near-infrared reflectance and transmission spectra. Meat Science 66, 45-54.

Leuenberger B.H., 1991. Investigation of viscosity and gelation properties of different mammalian and fish gelatins. Food Hydrocolloids 5, 353-61.

Levasseur-Garcia C., Kleiber D., 2015. A method for the allotment of maize contaminated by toxins. Journal of near Infrared Spectroscopy 23, 255-65.

Liao Y.-T., Fan Y.-X., Cheng F., 2010. On-line prediction of fresh pork quality using visible/near-infrared reflectance spectroscopy. Meat Science 86, 901-07.

Liu G., Giddings J., 1992. Separation of particles in aqueous suspensions by thermal field-flow fractionation. Measurement of thermal diffusion coefficients. Chromatographia 34, 483-92.

Liu J., Andya J.D., Shire S.J., 2006. A critical review of analytical ultracentrifugation and field flow fractionation methods for measuring protein aggregation. The AAPS journal 8, E580-E89.

Martens H., Naes T., 1992. *Multivariate calibration*. (John Wiley & Sons).

Mélissopoulos A., Levacher C., 1998. La peau. Structure et physiologie. Techniques et documentation

Messaud F.A., Sanderson R.D., Runyon J.R., Otte T., Pasch H., Williams S.K.R., 2009. An overview on field-flow fractionation techniques and their applications in the separation and characterization of polymers. Progress in Polymer Science 34, 351-68.

Mewis J., Wagner N.J., 2012. *Colloidal suspension rheology*. (Cambridge University Press).

Meyer M., Morgenstern B., 2003. Characterization of gelatine and acid soluble collagen by size exclusion chromatography coupled with multi angle light scattering (SEC-MALS). Biomacromolecules 4, 1727-32.

Mitsumoto M., Maeda S., Mitsuhashi T., Ozawa S., 1991. Near-Infrared Spectroscopy Determination of Physical and Chemical Characteristics in Beef Cuts. Journal of food science 56, 1493-96.

Moussard C., 2006. *Biochimie structurale et métabolique*. (De Boeck Supérieur).

Munera S., Amigo J.M., Blasco J., Cubero S., Talens P., Aleixos N., 2017. Ripeness monitoring of two cultivars of nectarine using VIS-NIR hyperspectral reflectance imaging. Journal of Food Engineering 214, 29-39.

Murray I., Williams P., 1987. Chemical Principles of Near-infrared Technology, Near-infrared Technology in The Agricultural and Food Industries. by Williams PC, Norris K., American Association of Cereal Chemists, Minnesota, USA

Næs T., Isaksson T., Fearn T., Davies T., 2002. *A user friendly guide to multivariate calibration and classification*. (NIR publications).

Nicolaï B.M., Beullens K., Bobelyn E., Peirs A., Saeys W., Theron K.I., Lammertyn J., 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biology and Technology 46, 99-118.

Nilsson L., 2013. Separation and characterization of food macromolecules using field-flow fractionation: A review. Food Hydrocolloids 30, 1-11.

Norris K.H., 1964. Design and development of a new moisture meter. Agric. Eng 45, 370-72.

Olsen E.F., Rukke E.O., Egelandsdal B., Isaksson T., 2008. Determination of omega-6 and omega-3 fatty acids in pork adipose tissue with nondestructive Raman and fourier transform infrared spectroscopy. Applied Spectroscopy 62, 968-74.

Ortiz-Somovilla V., España-España F., Gaitán-Jurado A., Pérez-Aparicio J., De Pedro-Sanz E., 2007. Proximate analysis of homogenized and minced mass of pork sausages by NIRS. Food Chemistry 101, 1031-40.

Petsko G.A., Ringe D., Sanlaville M.C., Charmot M.D., 2008. *Structure et fonction des protéines*. (De Boeck Supérieur).

Podczeck F., Jones B.E., 2004. *Pharmaceutical capsules*. (Pharmaceutical Press).

Pomerantsev A.L., Rodionova O.Y., Skvortsov A.N., 2017. Diffuse Reflectance Spectroscopy of Hidden Objects, Part I: Interpretation of the Reflection–Absorption-Scattering Fractions in Near-Infrared (NIR) Spectra of Polyethylene Films. Applied spectroscopy 0003702817694182.

Prieto N., Andrés S., Giráldez F.J., Mantecón A.R., Lavín P., 2006. Potential use of near infrared reflectance spectroscopy (NIRS) for the estimation of chemical composition of oxen meat samples. Meat Science 74, 487-96.

Prieto N., Juarez M., Larsen I., Lopez-Campos O., Zijlstra R., Aalhus J., 2015. Rapid discrimination of enhanced quality pork by visible and near infrared spectroscopy. Meat Science 110, 76-84.

Prieto N., Lopez-Campos O., Aalhus J., Dugan M., Juarez M., Uttaro B., 2014. Use of near infrared spectroscopy for estimating meat chemical composition, quality traits and fatty acid content from cattle fed sunflower or flaxseed. Meat Science 98, 279-88.

Prieto N., Pawluczyk O., Dugan M.E.R., Aalhus J.L., 2017. A Review of the Principles and Applications of Near-Infrared Spectroscopy to Characterize Meat, Fat, and Meat Products. Applied spectroscopy 0003702817709299.

Prieto N., Roehe R., Lavín P., Batten G., Andrés S., 2009. Application of near infrared reflectance spectroscopy to predict meat and meat products quality: A review. Meat Science 83, 175-86.

Pullanagari R.R., Yule I.J., Agnew M., 2015. On-line prediction of lamb fatty acid composition by visible near infrared spectroscopy. Meat Science 100, 156-63.

Qin J., 2017. General Theory on Biased Sampling Problems. in, Biased Sampling, Over-identified Parameter Problems and Beyond (Springer).

Ratanathanawongs S.K., Giddings J.C., 1992. Dual-field and flow-programmed lift hyperlayer field-flow fractionation. Analytical chemistry 64, 6-15.

Rbii K., 2010. Formation d'agrégats de hauts poids moléculaires dans la gélatine et comportement en solution aqueuse. thesis

Rbii K., Violleau F., Guedj S., Surel O., 2009. Analysis of aged gelatin by AFlFFF-MALS: Identification of high molar mass components and their influence on solubility. Food Hydrocolloids 23, 1024-30.

Ricard-Blum S., 2011. The collagen family. Cold Spring Harbor perspectives in biology 3, a004978.

Ricard-Blum S., Ruggiero F., 2005. The collagen superfamily: from the extracellular matrix to the cell membrane. Pathologie Biologie 53, 430-42.

Robert P., Devaux M.F., Mouhous N., Dufour E., 1999. Monitoring the secondary structure of proteins by near-infrared spectroscopy. Applied spectroscopy 53, 226-32.

Roggo Y., Chalus P., Maurer L., Lema-Martinez C., Edmond A., Jent N., 2007. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. Journal of pharmaceutical and biomedical analysis 44, 683-700.

Rohe T., Becker W., Kölle S., Eisenreich N., Eyerer P., 1999. Near infrared (NIR) spectroscopy for in-line monitoring of polymer extrusion processes. Talanta 50, 283-90.

Sadler A.J., Horsch J.G., Lawson E.Q., Harmatz D., Brandau D.T., Middaugh C.R., 1984. Near-infrared photoacoustic spectroscopy of proteins. Analytical biochemistry 138, 44-51.

Sankaran S., Ehsani R., Inch S.A., Ploetz R.C., 2012. Evaluation of Visible-Near Infrared Reflectance Spectra of Avocado Leaves as a Non-destructive Sensing Tool for Detection of Laurel Wilt. Plant Disease 96, 1683-89.

Sato K., Yomogida K., Wada T., Yorihuzi T., Nishimune Y., Hosokawa N., Nagata K., 2002. Type XXVI collagen, a new member of the collagen family, is specifically expressed in the testis and ovary. Journal of Biological Chemistry 277, 37678-84.

Savenije B., Geesink G., Van der Palen J., Hemke G., 2006. Prediction of pork quality using visible/near-infrared reflectance spectroscopy. Meat Science 73, 181-84.

Savitzky A., Golay M.J., 1964. Smoothing and differentiation of data by simplified least squares procedures. Analytical chemistry 36, 1627-39.

Schimpf M.E., Caldwell K., Giddings J.C., 2000. *Field-flow fractionation handbook*. (John Wiley & Sons).

Schrieber R., Gareis H., 2007. *Gelatine handbook: theory and industrial practice*. (John Wiley & Sons).

Segtnan V., Kvaal K., Rukke E., Schüller R., Isaksson T., 2003. Rapid assessment of physico-chemical properties of gelatine using near infrared spectroscopy. Food Hydrocolloids 17, 585-92.

Shoulders M.D., Raines R.T., 2009. Collagen structure and stability. Annual review of biochemistry 78, 929-58.

Siesler H.W., Ozaki Y., Kawata S., Heise H.M., 2008. *Near-infrared spectroscopy: principles, instruments, applications*. (John Wiley & Sons).

Söderhäll C., Marenholz I., Kerscher T., Rüschendorf F., Esparza-Gordillo J., Worm M., Gruber C., Mayr G., Albrecht M., Rohde K., 2007. Variants in a novel epidermal collagen gene (COL29A1) are associated with atopic dermatitis. PLoS biology 5, e242.

Sosulski F.W., Imafidon G.I., 1990. Amino acid composition and nitrogen-to-protein conversion factors for animal and plant foods. Journal of Agricultural and Food Chemistry 38, 1351-56.

Stegemann H., Stalder K., 1967. Determination of hydroxyproline. Clinica chimica acta 18, 267-73.

Su H., Sha K., Zhang L., Zhang Q., Xu Y., Zhang R., Li H., Sun B., 2014. Development of near infrared reflectance spectroscopy to predict chemical composition with a wide range of variability in beef. Meat Science 98, 110-14.

Su J., Gorse K., Ramirez F., Fox M.A., 2010. Collagen XIX is expressed by interneurons and contributes to the formation of hippocampal synapses. Journal of Comparative Neurology 518, 229-53.

Sun D.-W., 2009. *Infrared Spectroscopy for Food Quality Analysis and Control*. (Academic Press, Elsevier Science).

Sweeney E., Roberts D., Corbo T., Jacenko O., 2010. Congenic mice confirm that collagen X is required for proper hematopoietic development. PLoS ONE 5, e9518.

Sworn G., 2009. 8 - Xanthan gum. in G. O. Phillips and P. A. Williams (eds.), Handbook of Hydrocolloids (Second edition) (Woodhead Publishing).

Taggart P., Mitchell J.R., 2009. 5 - Starch. in G. O. Phillips and P. A. Williams (eds.), Handbook of Hydrocolloids (Second edition) (Woodhead Publishing).

Te Nijenhuis K., 1997. *Gelatin*. (Springer).

Till U., Gaucher-Delmas M., Saint-Aguet P., Hamon G., Marty J.-D., Chassenieux C., Payré B., Goudounèche D., Mingotaud A.-F., Violleau F., 2014. Asymmetrical flow field-flow fractionation with multi-angle light scattering and quasi-elastic light scattering for characterization of polymersomes: comparison with classical techniques. Analytical and bioanalytical chemistry 406, 7841-53.

Trienekens J., Zuurbier P., 2008. Quality and safety standards in the food industry, developments and challenges. International Journal of Production Economics 113, 107-22.

Tromp R.H., ten Grotenhuis E., Olieman C., 2002. Self-aggregation of gelatine above the gelling temperature analysed by SEC–MALLS. Food Hydrocolloids 16, 235-39.

Tsuchikawa S., Torii M., Tsutsumi S., 1996. Application of near infrared spectrophotometry to wood, 4: Calibration equations for moisture content. Journal of the Japan Wood Research Society (Japan)

Tuffery S., 2007. *Data Mining and Statistics for Decision Making*.

Tufféry S., 2011. *Data mining and statistics for decision making*. (Wiley Chichester).

USP, 2016. Gelatin. In *UPS 39 NF 34 -Notice of Adoption of Harmonized Standard*, edited by United States Pharmacopeial.

Van den Bosch E., Gielens C., 2003. Gelatin degradation at elevated temperature. International Journal of Biological Macromolecules 32, 129-38.

Van Krevelen D.W., Te Nijenhuis K., 2009. *Properties of polymers: their correlation with chemical structure; their numerical estimation and prediction from additive group contributions*. (Elsevier).

Viel C., Fournier J., 2006. Histoire des procédés d'extraction de la gélatine et débats des commissions académiques (XIXe siècle). Revue d'histoire de la pharmacie 94, 7-28.

Viljoen M., Hoffman L.C., Brand T., 2007. Prediction of the chemical composition of mutton with near infrared reflectance spectroscopy. Small Ruminant Research 69, 88-94.

Visa S., Ramsay B., Ralescu A., van der Knaap E., 2011. Confusion Matrix-based Feature Selection. In *22nd Midwest Artificial Intelligence and Cognitive Science Conference 2011*, edited by Atsushi Inoue Sofia Visa, Anca Ralescu, 120-27. Cincinnati, USA - 16-17 April 2011.

Voet D., Voet J., 2004. Three-Dimensional Structures of Proteins—Fibrous Proteins—Collagen. Biochemistry 233-39.

Wasswa J., Tang J., Gu X., 2007. Utilization of fish processing by-products in the gelatin industry. Food Reviews International 23, 159-74.

Weng W., Zheng H., Su W., 2014. Characterization of edible films based on tilapia (Tilapia zillii) scale gelatin with different extraction pH. Food Hydrocolloids 41, 19-26.

Westad F., Nilsen B.N., Wahlgren N.M., Hildrum K.I., 2004. Removal of conveyor belt near infrared signals in in-line monitoring of proximal ground beef composition. Journal of Near Infrared Spectroscopy 12, 377-79.

Wold J., O'Farrell M., Høy M., Tschudi J., 2011. On-line determination and control of fat content in batches of beef trimmings by NIR imaging spectroscopy. Meat Science 89, 317-24.

Wold S., Martens H., Wold H., 1983. The multivariate calibration problem in chemistry solved by the PLS method. Matrix pencils 286-93.

Wyatt P.J., 1993. Light scattering and the absolute characterization of macromolecules. Analytica Chimica Acta 272, 1-40.

Yang H., Wang Y., 2009. Effects of concentration on nanostructural images and physical properties of gelatin from channel catfish skins. Food Hydrocolloids 23, 577-84.

Ylönen R., Kyrönlahti T., Sund M., Ilves M., Lehenkari P., Tuukkanen J., Pihlajaniemi T., 2005. Type XIII collagen strongly affects bone formation in transgenic mice. Journal of Bone and Mineral Research 20, 1381-93.

Zamora-Rojas E., Pérez-Marín D., De Pedro-Sanz E., Guerrero-Ginel J.E., Garrido-Varo A., 2012. Handheld NIRS analysis for routine meat quality control: Database transfer from at-line instruments. Chemometrics and Intelligent Laboratory Systems 114, 30-35.

Zhou P., Mulvaney S.J., Regenstein J.M., 2006. Properties of Alaska pollock skin gelatin: a comparison with tilapia and pork skin gelatins. Journal of food science 71,

Zhou P., Regenstein J., 2006. Determination of total protein content in gelatin solutions with the Lowry or Biuret assay. Journal of food science 71,

Zimm B.H., 1948. Apparatus and methods for measurement and interpretation of the angular variation of light scattering; preliminary results on polystyrene solutions. The Journal of Chemical Physics 16, 1099-116.

Annexes

Annexe 1 : Origine des 57 échantillons de couennes de porc prélevés sur le site de Weishardt (fournisseurs, pays, conditionnement, type d'animal, type de production, partie de l'animal)

N° échantillon	N° Fournisseur	Pays	Conditionnement	Type d'animal	Type de production	Partie de l'animal
1	F-1	Belgique	Congelée	Coche	Conventionnel	Couenne
2	F-1	Belgique	Congelée	Coche	Conventionnel	Couenne
3	F-2	Allemagne	Fraiche	Porc	Conventionnel	Couenne
4	F-3	Espagne	Fraiche	Porc	Conventionnel	Couenne
5	F-4	Espagne	Fraiche	Porc	Conventionnel	Couenne
6	F-5	France	Fraiche	Porc	Conventionnel	Couenne
7	F-6	France	Fraiche	Porc	Conventionnel	Couenne
8	F-7	Allemagne	Fraiche	Porc	Conventionnel	Couenne
9	F-8	Espagne	Congelée	Coche	Conventionnel	Couenne
10	F-9	Espagne	Fraiche	Porc	Conventionnel	Bardière
11	F-10	Espagne	Fraiche	Porc	Conventionnel	Couenne
12	F-11	Espagne	Fraiche	Porc	Conventionnel	Couenne
13	F-12	France	Fraiche	Porc	Conventionnel	Couenne
14	F-12	France	Fraiche	Porc	Conventionnel	Bardière
15	F-13	France	Fraiche	Porc	Bio	Couenne
16	F-13	France	Fraiche	Porc	Bio	Couenne
17	F-13	France	Fraiche	Porc	Bio	Couenne
18	F-13	France	Fraiche	Porc	Bio	Couenne
19	F-14	Portugal	Congelée	Porc	Conventionnel	Couenne
20	F-15	France	Fraiche	Porc	Conventionnel	Couenne
21	F-15	France	Fraiche	Porc	Conventionnel	Couenne
22	F-16	France	Fraiche	Porc	Conventionnel	Couenne
23	F-8	Espagne	Congelée	Coche	Conventionnel	Couenne
24	F-8	Espagne	Congelée	Coche	Conventionnel	Couenne
25	F-9	Espagne	Fraiche	Porc	Conventionnel	Bardière
26	F-17	Espagne	Fraiche	Porc	Conventionnel	Couenne
27	F-18	France	Fraiche	Porc	Conventionnel	Couenne
28	F-5	France	Fraiche	Porc	Conventionnel	Couenne
29	F-19	Pologne	Fraiche	Porc	Conventionnel	Couenne
30	F-20	France	Fraiche	Porc	Conventionnel	Bardière
31	F-21	Angleterre	Fraiche	Porc	Conventionnel	Couenne
32	F-22	Espagne	Fraiche	Porc	Conventionnel	Bardière
33	F-15	France	Fraiche	Porc	Conventionnel	Couenne
34	F-13	France	Fraiche	Porc	Bio	Couenne
35	F-13	France	Fraiche	Porc	Bio	Couenne
36	F-2	Allemagne	Fraiche	Porc	Conventionnel	Couenne
37	F-2	Allemagne	Fraiche	Porc	Conventionnel	Couenne
38	F-23	Espagne	Fraiche	Porc	Conventionnel	Couenne
39	F-18	France	Fraiche	Porc	Conventionnel	Couenne
40	F-12	France	Fraiche	Porc	Conventionnel	Bardière
41	F-23	Espagne	Fraiche	Porc	Conventionnel	Couenne
42	F-12	France	Fraiche	Porc	Conventionnel	Couenne
43	F-12	France	Fraiche	Porc	Conventionnel	Bardière
44	F-24	Italie	Congelée	Porc	Conventionnel	Couenne
45	F-12	France	Fraiche	Porc	Conventionnel	Couenne
46	F-25	France	Fraiche	Porc	Conventionnel	Couenne
47	F-26	Espagne	Fraiche	Porc	Conventionnel	Couenne
48	F-6	France	Fraiche	Porc	Conventionnel	Couenne
49	F-10	Espagne	Fraiche	Porc	Conventionnel	Couenne
50	F-6	France	Fraiche	Porc	Conventionnel	Couenne
51	F-27	Espagne	Fraiche	Porc	Conventionnel	Couenne
52	F-10	Espagne	Fraiche	Porc	Conventionnel	Couenne
53	F-28	Allemagne	Fraiche	Porc	Conventionnel	Couenne
54	F-29	Allemagne	Fraiche	Porc	Conventionnel	Couenne
55	F-28	Allemagne	Fraiche	Porc	Conventionnel	Couenne
56	F-11	Espagne	Fraiche	Porc	Conventionnel	Couenne
57	F-11	Espagne	Fraiche	Porc	Conventionnel	Couenne

Annexe 2 : Solution pour la réaction d'oxydoréduction du dosage de l'hydroxyproline.

Solution tampon pH6 :

Dans une fiole jaugée de 1000 mL, ajouter :

- 50 g d'acide citrique monohydrate
- 12 mL d'acide acétique (97%)
- 120 g d'acétate de sodium, trihydraté
- 3,4 g d'hydroxyde de sodium
- Eau distillée jusqu'au trait de jauge

Mélangé les 1000 mL de solution avec 200 mL d'eau distillée et 300 mL de propanol-1

Solution de Chloramine T :

La solution doit être préparée juste avant utilisation. Dissoudre dans un bécher, 1,4 g de chloramine T dans 10 mL d'eau distillée, puis ajouter 10 mL de propanol-1 et 80 mL de solution tampon pH 6. La solution doit être conservée à l'abri de la lumière.

Solution de *p-DiMethylAminoBenzaldehyde* (*p*DMAB) :

Dans un bécher, dissoudre 10 g de *p*DMAB dans 35 mL d'acide perchlorique à 60%, puis ajouter progressivement 65 mL de propanol-2. La solution peut être conservée pendant une journée.

Annexe 3 : Taux d'humidité moyen sur les trois séchoirs de l'entreprise entre 2013 et 2016

Séchoir	2013-2016
A	**10.80%**
B	**11,21 %**
C	**11.23 %**
Moyenne	**11.02%**

Résumé

La gélatine est un biopolymère naturel obtenu après dénaturation et hydrolyse partielle de fibres de collagène, une protéine fibrillaire présente dans les tissus conjonctifs de toutes les espèces du règne animal. Elle est utilisée dans différentes industries, parmi lesquelles l'industrie agroalimentaire, pharmaceutique, photographique et cosmétique. Ses propriétés fonctionnelles dépendent du procédé de fabrication, mais également de l'origine des matières collagéniques. L'objectif de ce travail de thèse est la caractérisation de la matière première (couennes de porc), mais également de la gélatine obtenue lors du processus, par une méthode rapide et non destructive. L'utilisation de la spectroscopie proche infrarouge couplée à des méthodes chimiométriques, a permis de travailler à l'échelle du laboratoire, mais également sur la chaîne de production industrielle. Le premier chapitre porte sur l'étude de l'hétérogénéité des couennes de porc au laboratoire, en termes de teneur en protéines, matières grasses, matières sèches et collagène. Plusieurs modèles ont pu être développés pour prédire ces teneurs, à partir de spectres proche infrarouge collectés sur des couennes en mouvement. Les meilleurs modèles présentent des performances compatibles avec une utilisation dans l'usine. Le second chapitre adapte cette méthode à l'échelle industrielle, c'est-à-dire à partir du rendement en gélatine fabriquée pour un lot de 75t de couennes. L'approche prédictive ne s'est pas avérée concluante, cependant des approches de classification ont montré leur intérêt. Les deux essais suivants ont porté sur la gélatine. Le troisième chapitre porte sur le développement de modèles de prédiction des propriétés physico-chimiques d'échantillons de gélatine à partir de spectres proche infrarouge, mettant en évidence de hautes capacités de prédiction de ces paramètres (r^2>0,9). Enfin, ce chapitre propose de relier les caractéristiques moléculaires de la gélatine à ses propriétés physico-chimiques, par la technique de *Asymmetrical Flow Field-Flow Fractionation* couplée à un détecteur de diffusion de la lumière Multiangulaire (MALS). Les paramètres de caractérisation de l'AFlFFF-MALS permettent de discriminer partiellement des échantillons de gélatine dont les paramètres de bloom et de viscosité sont différents.

Abstract

Gelatin is a natural biopolymer obtained after denaturation and partial hydrolysis of collagen fibers, a fibrillar protein present in the connective tissues of all species of the animal kingdom. It is used in various industries, including the agri-food, pharmaceutical, photographic and cosmetic industries. Its functional properties depend on the manufacturing process, but also on the origin of the collagenous materials. The objective of this thesis is to characterize the raw material (pig rind), but also the gelatin obtained during the process, by a fast and non-destructive tool. The use of near-infrared spectroscopy coupled with chemometric methods allowed us to work on the scale of the laboratory but also on the industrial chain. The first study examines the heterogeneity of pig rinds, under laboratory conditions, in terms of protein, fat, moisture and collagen contents. Several models have been developed to predict these levels, from near infrared spectra collected on moving rinds. The best models have good performances. In the second industrial scale trial, an approach to gelatin yield from 75 tons batches was proposed. The predictive approach has not been conclusive; however, classification approaches have shown interesting results. The next two tests were on gelatin. The third test allowed the development of models for predicting the physico-chemical properties of gelatin samples from near-infrared spectra, demonstrating very good predictive capabilities of these parameters (r^2> 0.9). Finally, a final test allowed to link the molecular characteristics of the gelatin to its physicochemical properties, by the technique of Asymmetrical Field-Flow Fractionation coupled with a Multiangular Light Scattering Detector (MALS). The characterization parameters of AFlFFF-MALS make it possible to partially discriminate gelatin samples with different bloom and viscosity parameters.