



Analyse de la digestion des pains par résonance magnétique nucléaire (RMN) pour mieux comprendre l'hypersensibilité au gluten

Chuyu Ding

► To cite this version:

Chuyu Ding. Analyse de la digestion des pains par résonance magnétique nucléaire (RMN) pour mieux comprendre l'hypersensibilité au gluten. Sciences du Vivant [q-bio]. 2022. hal-03757482

HAL Id: hal-03757482

<https://hal.inrae.fr/hal-03757482>

Submitted on 22 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université de Technologie de Compiègne

Département Génie biologique TN09

Analyse de la digestion des pains par résonance magnétique nucléaire (RMN) pour mieux comprendre l'hypersensibilité au gluten

Du 30 août 2021 au 11 février 2022

Chuyu DING

Encadrement de stage : M. Guilhem PAGES

M. Mounir TRAIKIA

Tutrice de l'UTC : M^{me} Elise PROST



Institut National de Recherche pour l'Agriculture,

l'Alimentation et l'Environnement

Centre Clermont-Auvergne-Rhône-Alpes

Unité QuaPA - AgroResonance

1 REMERCIEMENTS

En premier lieu, je tiens à remercier mon tuteur de stage, **M. Guilhem PAGES**, pour son suivi au cours du stage. Il a su me faire confiance lors de cette aventure dans le monde de la recherche et a partagé ses connaissances de manière très pédagogique. Je le remercie aussi pour sa disponibilité et la qualité de son encadrement en laboratoire.

Je souhaite ensuite adresser mes remerciements à **M. Mounir TRAIKIA**. J'ai beaucoup appris sur la RMN dans son cours et il m'a patiemment enseigné l'analyse des spectres et m'a guidé lorsque j'étais bloqué.

Je remercie également **M. Jean-Marie BONNY**, responsable de la plateforme AgroResonance, pour son accueil chaleureux et ses salutations quotidiennes.

Je voudrai exprimer ma reconnaissance envers tous les membres de la plateforme qui m'ont apporté leur soutien moral et intellectuel tout au long de mon stage, aux conversations quotidiennes, aux échanges interculturels et pour m'avoir fait découvrir cette belle ville de Clermont-Ferrand.

Un grand merci à ma **mère** et mon **père** pour leur soutien inconditionnel, à la fois moral et économique. J'aimerais adresser mes remerciements à mes amis **Lexin XU** et **Yanlin ZHU** qui m'ont conseillé et accompagné virtuellement pendant ces six mois.

Enfin, je tiens à remercier ma tutrice de l'UTC, **Mme Elise PROST**, pour son encadrement, mais aussi pour ses soutiens et son réconfort, son intérêt pour ma vie et mon travail et sa visite au bureau.

2 SOMMAIRE

1	Remerciements	1
3	Introduction	3
4	Présentation de l'INRAE.....	4
4.1	Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE).....	4
4.2	Unité QuaPA	4
4.3	AgroResonance.....	5
5	Contexte.....	7
5.1	Le gluten et les enjeux sur la santé	7
5.2	RMN et métabolomique.....	9
5.3	Mes missions	12
6	Matériels et Méthodes	13
6.1	Plan d'expériences.....	13
6.2	Acquisition des spectres RMN 1D et 2D	14
6.3	Analyse des données	16
7	Résultats et Discussion.....	18
7.1	Analyse préliminaire des spectres RMN.....	18
7.2	Traitement des spectres RMN	20
7.3	Analyse du compartiment gastrique	22
7.4	Identification par RMN 2D.....	29
7.5	Analyse du compartiment intestinal	32
8	Conclusion.....	38
9	Bibliographie	39
10	Liste des tableaux et figures.....	40

3 INTRODUCTION

La résonance magnétique nucléaire (RMN) est une des principales techniques spectroscopiques pour identifier des molécules. Elle est robuste, non-destructive et non-biaisée ; de plus, elle ne requiert que peu de préparation de l'échantillon et l'analyse est quantitative. Elle est donc couramment utilisée dans différents domaines scientifiques dont l'analyse des aliments et la caractérisation des produits de synthèse.

Le blé et ses sous-produits, qui constituent une part importante de l'alimentation humaine, fournissent environ 20 % de l'énergie totale et 22 % de l'apport en protéines. Grâce au réseau de gluten formé par les protéines du blé, la pâte possède d'excellentes propriétés viscoélastiques, propriétés indispensables à la panification du produit. Cependant, le gluten provoque également de nombreuses pathologies, notamment l'allergie, la maladie cœliaque et la sensibilité au gluten non cœliaque (SNCG). Tandis que l'allergie au blé et la maladie cœliaque sont des maladies bien connues, l'origine de la SNCG est encore inconnue et il n'existe pas d'indicateur clinique. En général, les troubles liés à la SNCG disparaissent lorsque le gluten est éliminé de l'alimentation, lorsque la personne choisit un régime dans lequel cet ingrédient est absent.

Une des hypothèses de travail pour expliquer l'apparition de la SNCG est qu'au cours du temps et de la sélection variétale, les blés possédant des réseaux de gluten les plus forts ont été choisis pour leurs propriétés panifiables. Cependant, les produits deviennent moins digestes pour certaines personnes car le réseau ne peut être digéré. Le but de mon stage est d'analyser par RMN les métabolites produits au cours de la digestion de pains préparés avec des farines contrastées. Mon rôle dans le cadre de ce stage a été de traiter et analyser les spectres RMN obtenus au cours de la digestion de ces pains pour générer les tableaux me permettant de faire des analyses statistiques, à réaliser ces analyses statistiques avec les méthodes les plus appropriées et finalement à identifier les métabolites clés à partir de spectres RMN 2D.

Ce stage de 24 semaines (du 30 août 2021 au 11 février 2022) s'est déroulé au sein la plateforme AgroResonance de l'unité QuaPA d'INRAE.

4 PRESENTATION DE L'INRAE

4.1 INSTITUT NATIONAL DE RECHERCHE POUR L'AGRICULTURE, L'ALIMENTATION ET L'ENVIRONNEMENT (INRAE)

L'institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE) est un institut national de recherche français. Créé le 1er janvier 2020, il est issu de la fusion entre l'INRA, Institut national de la recherche agronomique et l'IRSTEA, Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture.

L'institut est composé de 14 départements scientifiques, 267 unités de recherche, service et expérimentales dont les membres sont répartis dans 18 centres régionaux, en métropole et dans les territoires d'outre-mer. Il est l'un des principaux instituts de recherche au monde dans chacun de ses trois domaines scientifiques et le premier français. Chaque année, INRAE publie un très grand nombre d'articles scientifiques, avec un taux de citation moyen dépassant de 49 % la moyenne mondiale.

L'INRAE a pour ambition d'être un acteur clé des transitions nécessaires pour répondre aux grands enjeux mondiaux : l'augmentation de la population, le changement climatique, le déclin de la biodiversité etc. Il compte plus de 10 000 agents, parmi lesquels on retrouve plus de 50% de femmes. Son budget annuel est de plus de 1 000 M€.

4.2 UNITE QUAPA

L'unité Qualité des Produits Animaux, QuaPA, est rattachée au département TRANSFORM et est située sur le site de Theix du Centre INRAE Clermont-Auvergne-Rhône-Alpes. Le Centre est intégré dans un écosystème particulièrement dynamique, au sein de la seconde région française en matière de recherche et de développement, permettant de développer de nombreux partenariats tant au niveau local, national, qu'international.



Figure 1 : Logo de l'unité QuaPA, INRAE

QuaPA est structurée en trois équipes de recherche disciplinaires : Biochimie et Protéines du Muscle (BPM), Imagerie et Transferts (IT), Microcontaminants, Arômes et Sciences Séparatives (MASS) et deux plateformes: AgroResonance et la composante protéomique de la Plate-Forme d'Exploration du Métabolisme (PFEMcp).

La viande et les produits carnés constituent une partie importante de notre alimentation. Cet aliment, qui est riche en protéines et présente un équilibre en acides aminés essentiels, convient idéalement à la consommation humaine. La filière de production de la viande subit une crise sanitaire et de nombreuses critiques. Cependant, il est critique d'évaluer objectivement, sur la base de recherches scientifiques, les forces et faiblesses du secteur. L'unité QuaPA travaille également sur les questions liées à la conservation de la viande, à la consommation de protéines par des populations spécifiques (séniors, sportifs, ...). Plus généralement, l'unité QuaPA travaille sur les produits animaux dans le but d'améliorer la qualité nutritionnelle et la sécurité des aliments tout en maintenant leurs propriétés technologiques et sensorielles. Avec ces équipes et plateformes, l'unité possède une expérience riche sur la caractérisation de la structure des aliments ainsi que de leur digestion *in vitro*.

4.3 AGRORESONANCE

AgroResonance est une Infrastructure Scientifique Collective (ou plateforme) spécialisée en RMN pour l'agronomie, l'agro-alimentaire et la nutrition et certifiée ISO 9001. En outre, AgroResonance est membre fondateur de la plateforme régionale multimodale In Vivo Imaging in Auvergne (IVIA), plateforme labellisée par IBISA, l'instance nationale de référence qui labellise les infrastructures en biologie, santé et agronomie. AgroResonance a donc un accès facilité à la plupart des autres modalités d'imagerie in vivo clermontoise depuis l'animal jusqu'à l'homme.

Elle réunit un haut niveau de compétences et de techniques pour proposer analyses et développements afin de répondre aux questions dans les domaines de l'agroalimentaire, du végétal, de la nutrition et de la santé en utilisant la RMN et l'IRM. Elle collabore aussi bien avec la communauté scientifique académique (CNRS, Facultés, CHU, ...) qu'avec les entreprises privées.

L'ISC AgroResonance est équipée de trois spectromètres de RMN de dernière génération permettant de couvrir une large gamme d'applications en imagerie ou spectroscopie RMN. Le Tableau 1 résume les principales caractéristiques de chacun des

aimants. Plus récemment, AgroResonance s'est équipée d'un aimant mobile à faible champ magnétique le NMR MOUSE.

Tableau 1 : Principales caractéristiques des différents spectromètres RMN à haut champ magnétique présents sur la plateforme

	Biospec 47/40	Avance 400	Biospec 117/16
Champ magnétique	4.7 T	9.7 T	11.7 T
Orientation	Horizontale	Verticale	Horizontale
Fréquence ^1H	200 MHz	400 MHz	500 MHz
Résolution moyenne	~1 mm	~10 μm	~100 μm
Application typique réalisée sur la Plateforme	Imagerie sur un large champ de vue avec éventuellement des contraintes physico-chimiques (saumon, jambon, ...)	Micro-imagerie et spectroscopie <i>in vitro</i> (tissus fixés, plantes, ...)	Imagerie et spectroscopie localisée <i>in vivo</i> chez le petit animal
Taille maximale de l'échantillon	20 cm	3 cm	9 cm

5 CONTEXTE

5.1 LE GLUTEN ET LES ENJEUX SUR LA SANTE

Le gluten est un réseau macromoléculaire qui se forme à partir des protéines présentes dans les grains de céréales comme le blé, l'orge ou le seigle. Ce réseau est principalement constitué de deux types de protéines insolubles dans l'eau : les gliadines et les gluténines qui sont les protéines majoritaires dans le blé. La structure du réseau de gluten est essentielle pour obtenir une pâte possédant les propriétés adaptées pour sa transformation. Le gluten est stable thermiquement et très élastique. Les propriétés fonctionnelles du gluten dépendent, entre autres, du ratio entre les gliadines et la gluténine et de la manière dont ils sont liés les uns aux autres. De plus, le gluten est souvent utilisé comme additif pour améliorer la texture et le goût des produits ou pour les épaissir. Il est donc présent dans les aliments comme les confiseries, les glaces, les beurres et même les médicaments.

Le gluten se retrouve dans tous les aliments préparés à base de farine de blé comme les produits boulangés, les biscuits ou les pâtes. Du gluten est également présent dans la bière puisqu'elle est préparée habituellement à base d'orge. Tandis que la consommation de gluten ne posait pas de problème majeur, elle est devenue aujourd'hui un enjeu de santé publique.

Ces protéines peuvent provoquer, pour une faible proportion de la population, des maladies comme : l'allergie au blé ou la maladie cœliaque (MC). La MC est une réaction auto-immune transmise par les structures compactes et serrées des résidus riches en proline. Les séquences peptidiques de la gliadine sont très résistantes à la digestion protéolytique gastrique, pancréatique et intestinale, ce qui les empêche de se cliver et cause des problèmes de digestion. A la différence de la MC, l'allergie au blé est généralement provoquée par les protéines de la céréale, parmi lesquelles on retrouve les albumines, les globulines, les gliadines et les gluténines. Il y a deux types d'allergies : IgE médiées et non IgE médiées. Le premier type se caractérise par la production d'IgE. Après avoir reconnu (principalement par consommation, mais aussi par inhalation et l'exercice physique) le blé, la libération de molécules comme l'histamine va causer des symptômes allergiques. L'autre type, l'allergie non IgE médiées, se caractérise par une inflammation chronique de l'appareil digestif pouvant alors entraîner des symptômes tels que des nausées, vomissements, douleurs abdominales ou diarrhées. Tandis que

ces deux maladies sont bien connues et possèdent un diagnostic bien établi, il n'en est pas de même pour la sensibilité non cœliaque au gluten (SNCG).

Depuis quelques années, cette nouvelle forme de pathologie est apparue. Les symptômes sont très variés et certains sont proches de la MC. Cependant, le manque de réaction immunologique anormale au gluten élimine le diagnostic de la MC. D'un point de vue médical, le diagnostic est quasiment impossible à poser. Après avoir éliminé comme pathologie la MC et l'allergie au blé, le seul moyen pour conclure que le patient est SNCG est d'éliminer le gluten de l'alimentation et de constater la disparition des symptômes. Actuellement, l'origine de la SNCG est inconnue. Il est donc indispensable de mener des travaux pour comprendre à quoi est due cette pathologie.

5.1.1 Le projet GlutN : Blé et pains spécifiques pour résoudre la sensibilité au gluten

Ce projet est un effort conjoint de diverses équipes, dont les unités QuaPA et de nutrition humaine (UNH) d'INRAE, le CHU Clermont-Ferrand et son service de Médecine Digestive et Hépatobiliaire, et est financé par l'Agence Nationale de la Recherche (ANR) pour étudier la sensibilité au gluten. Il a commencé en décembre 2017 et est porté par des chercheurs de l'unité GDEC d'INRAE. Ce projet cherche à produire des pains qui seraient tolérés par les SNCG. Ces travaux se basent donc sur l'hypothèse que l'origine de la SNCG provient de la structuration du réseau de gluten. En étudiant différentes variétés de blé, il est possible de modifier les propriétés du réseau de gluten et possiblement les symptômes du malade.

Le projet GlutN comporte cinq tâches opérationnelles :

1. L'étude épidémiologique, grâce à la base de données NutriNet-Santé, pour caractériser la relation entre nutrition et prévalence de la SNCG
2. La sélection de variétés de blé pour obtenir des farines qui donneront des réseaux de gluten contrastés
3. Les fabrications des pains spécifiques et l'étude de la structuration de la matrice amidon/protéines
4. La caractérisation des mécanismes de déstructuration du gluten au cours de la digestion
5. Une étude clinique pour découvrir des biomarqueurs de la SNCG

Actuellement, soixante-quinze lignées de blé tendre ont été caractérisées et ont donc permis de sélectionner quelques blés pour panifier selon deux procédés. Les travaux effectués lors de mon stage s'inscrivent dans ce projet et plus particulièrement dans la

tâche 4 et la mise en évidence de la déstructuration de l'aliment au cours de la digestion.

5.2 RMN ET METABOLOMIQUE

5.2.1 Principes de la RMN et de la métabolomique

En plus de la charge et la masse, que tous les noyaux possèdent, de nombreux isotopes possèdent un spin, ou moment angulaire. Le noyau avec un spin non nul, par exemple l'hydrogène (aussi appelé proton) qui possède un spin de $1/2$, peut être comparé à une petite barre aimantée. En insérant des atomes possédant des spins non nuls dans un champ magnétique, une aimantation macroscopique se crée. Il est alors possible de perturber le système puis d'enregistrer le signal lors de son retour à l'équilibre, c'est la résonance magnétique nucléaire (RMN).

La RMN est un outil de recherche et d'analyse très important, largement utilisé en chimie, en biologie et en médecine. Elle est importante pour la résolution structurale des petites molécules organiques et des macromolécules biologiques. La technique de la RMN présente les caractéristiques suivantes :

1. La préparation simple de l'échantillon permet de limiter les modifications chimiques ;
2. C'est une technique non destructive. La solution analysée peut être récupérée et réutilisée pour d'autres analyses. Il est également possible de suivre la dynamique du système en temps réel;
3. Le spectre RMN peut être analysé et interprété pour obtenir des informations structurales et/ou quantitatives. Pour cela, les déplacements chimiques, la multiplicité des signaux et leur intensité doivent être exploités. La large gamme d'expériences disponibles permet de remonter à toutes les informations structurales.

La métabolomique, quant à elle, est un domaine de recherche récent qui consiste à analyser statistiquement les données analytiques pour extraire les métabolites caractérisant des états différents : malade vs non malade, âge d'affinage d'un produit, Elle fournit un aperçu de l'état métabolique et permet en général de déterminer les événements biochimiques globaux associés à un système cellulaire ou biologique. En tant que telle, elle décrit de manière précise et complète l'état

physiologique stable d'une cellule ou d'un organisme et leurs réponses dynamiques aux modulations génétiques, abiotiques et biotiques de l'environnement.

Comme toutes les autres études omiques, la métabolomique est une discipline dépendant des technologies. Les trois principales techniques analytiques utilisées en métabolomique sont la chromatographie liquide (LC) ou gaz (GC) couplée à la spectrométrie de masse (MS) et la RMN [1]. Les avantages et inconvénients de la MS et de la RMN sont décrits dans le Tableau 2 ci-dessous.

Tableau 2 : Comparaisons des avantages et inconvénients de la MS et de la RMN

	MS	RMN
Sensitivité	++ (<1 µM)	-- (~ 1mM)
Sélectivité	++ (>5000 molécules)	- (~ 100 molécules)
Résolution	+ (masse exacte)	- (Seulement $\times 10^{-1}$ Hz, nécessite un champ magnétique très puissant pour augmenter la résolution)
Dommages aux échantillons	Perte de l'échantillon, non réutilisable	Non-destructive, réutilisable
Stabilité (Reproductibilité)	Haute variabilité	Haute Robustesse
Mesure d'échantillon	Méthodes et préparation plus complexes et longue durée	Relativement vite pour 1D 1H-NMR spectroscopie
Analyse quantitative	L'intensité de la ligne MS n'est souvent pas corrélée aux concentrations de métabolites à cause de l'ionisation	Facilement à quantitative car l'intensité de signaux et proportionnelle aux concentrations et nombre de noyaux dans la molécule
Information structurelle	Délicate pour l'analyse des fragments, ne fournit pas de bonnes informations structurelles	Peut fournir de bonnes informations structurelles avec des couplages

5.2.2 La spectroscopie RMN pour l'étude métabolomique

En raison des avantages de la spectroscopie RMN mentionnés ci-dessus, cette technique est particulièrement utilisée dans l'analyse métabolomique, d'autant plus que les atomes ^1H sont présents dans tous les composés organiques. La spectroscopie RMN 1D du ^1H , avec ou sans suppression du signal du solvant, permet d'obtenir rapidement un spectre présentant la totalité de l'information disponible. L'enregistrement de ces spectres est facilement automatisé.

Pour aller plus loin dans l'analyse spectrale, il est en général indispensable d'acquérir des expériences RMN multidimensionnelles pour identifier les métabolites. En règle générale, lors d'applications en métabolomiques, les expériences ont 2 dimensions et les noyaux enregistrés sont le ^1H et le ^{13}C . Les expériences les plus fréquentes sont :

- La COSY (Spectroscopie de Corrélation) qui fournit des informations sur la corrélation homonucléaire entre ^1H - ^1H ;
- La TOCSY (Spectroscopie de Corrélation Totale) qui est une extension de la COSY. Elle crée des corrélations entre tous les protons d'un système tant qu'il existe des couplages entre chaque proton intermédiaire. Ceci est très utile pour identifier les protons sur les anneaux de sucres ou les acides aminés ;
- L'HSQC (Spectroscopie de corrélation quantique hétéro-nucléaire simple) : cette expérience met en corrélation le déplacement chimique du proton avec celui du carbone directement lié. Les pics croisés indiquent le déplacement du proton et du carbone correspondant ;
- L'HMBC (Spectroscopie de corrélation des liaisons multiples hétéro-nucléaires) : cette technique utilise des couplages multiples sur deux ou trois liaisons comme un analogue à l'expérience TOCSY mais entre proton et carbone ;
- J-Res (Spectroscopie J-Résolu 2D) : elle a été présentée pour corréler les couplages J avec les déplacements chimiques.

La RMN 2D peut être utilisée pour surmonter le problème du chevauchement des résonances, sur la base de voisins liés par covalence, de temps de relaxation, de constantes de couplage, etc.

5.3 MES MISSIONS

Mon stage s'inscrit dans le cadre du projet GlutN et se focalise sur l'étude des métabolites produits au cours de la digestion in vitro des pains fabriqués avec des farines contrastées selon deux procédés. Lors de mon arrivée dans l'équipe, l'ensemble des spectres RMN 1D avait déjà été enregistré. Mon travail de stage s'est concentré sur (i) le traitement des spectres RMN pour pouvoir réaliser les analyses statistiques, (ii) les analyses statistiques pour identifier les métabolites d'intérêt et (iii) l'identification des métabolites d'intérêts par RMN multidimensionnelle.

6 MATERIELS ET METHODES

6.1 PLAN D'EXPERIENCES

L'expérience a été conçue comme suit. Un total de quatre farines différentes (w18, w24, w39, w59) a été transformé en pain en utilisant deux méthodes de fermentation : à partir de levure boulanger (notée B) ou de levain (notation L). Les quatre farines possédaient des caractéristiques différentes de tailles de polymères ainsi que d'âge de variété (variété initialement cultivée il y a longtemps ou plus récemment).

Tableau 3 : Description des caractéristiques des différentes farines

	Petits polymères	Gros polymères
Récente variété	w24, w39	w18
Ancienne variété	w59	

Au total, 8 pains différents ont été digérés *in vitro* dans le digesteur artificiel DIDGI® (Figure 2). Cet appareil permet de simuler le processus de digestion dans l'estomac et l'intestin. Des échantillons ont été prélevés au cours du temps dans chacun des compartiments. Ainsi, dans le compartiment gastrique des prélèvements ont été effectués à 5, 30, 60 et 120 min (notés G5, G30, G60 & G120) et à 60, 120, 180 et 240 min dans l'intestinal (notés I60, I120, I180 & I240). Chaque digestion a été répétée 3 fois (R1, R2, R3). Un groupe témoin complète cette étude. Il s'agit d'une digestion effectuée en absence de pain.



Figure 2 : Digesteur artificiel DIDGI®

Après prélèvement dans le digesteur artificiel, les échantillons étaient filtrés sur gaze puis neutralisés avec du carbonate (compartiment gastrique) ou du PEFABLOC (compartiment intestinal) et ensuite centrifugé et filtré à l'aide d'un filtre 0,45 μm . 540 μL de cette solution étaient introduits dans un tube RMN puis 60 μL d'eau deutériée ajoutée pour pouvoir verrouiller le champ magnétique.

Nous avons finalement obtenu un total de 216 spectres 1D RMN.

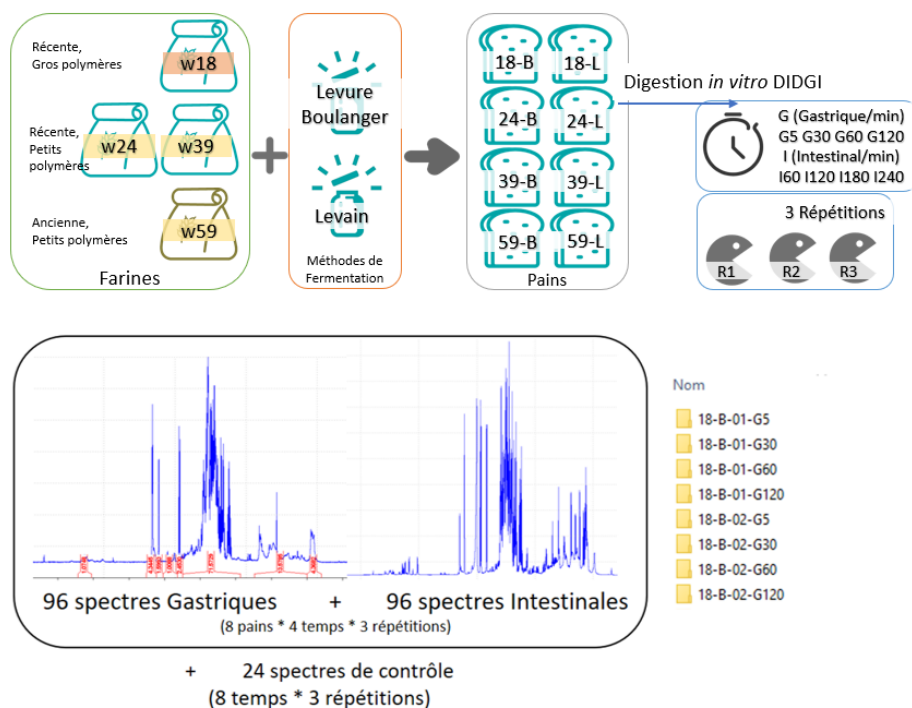


Figure 3 : Représentation schématique du plan d'expériences

6.2 ACQUISITION DES SPECTRES RMN 1D ET 2D

6.2.1 Spectres RMN 1D

Les spectres RMN ont été enregistrés à une température de 25°C sur un aimant Bruker à 11,7T, correspondant une fréquence de résonance pour le ^1H de 500 MHz et de 125 MHz pour le ^{13}C , équipé d'un passeur d'échantillons SampleJet et d'une sonde Prodigy TCI de 5-mm H&F/C/N avec un gradient de champ magnétique aligné le long de l'axe Z. La sonde est refroidie à la température de l'azote liquide pour diminuer le bruit électronique.

Après introduction de l'échantillon dans le spectromètre RMN, un délai de 5 minutes a été laissé pour laisser la température se stabiliser. Ensuite, la sonde était accordée, le champ magnétique verrouillé sur la fréquence de résonance du ^2H du solvant, l'homogénéité du champ magnétique ajustée (shims), la durée de l'impulsion à 90° mesurée (aux alentours de 9,5 μs à 8,5 W) puis la fréquence de résonance du solvant ajustée.

L'expérience RMN a consisté à réaliser une excitation des protons après une présaturation du solvant pendant le temps de relaxation (expérience zgpr). Le temps de saturation du solvant était de 5 s à une puissance d'environ 3,8 μW . L'impulsion radiofréquence était de 90° puis le signal RMN immédiatement enregistré (free

induction decay, FID). Le signal contenait 64 k points avec une bande passante de 20 ppm. Le temps d'acquisition du signal était de 3,2 s. Le gain du préamplificateur (rg) a été gardé constant pour pouvoir comparer les spectres entre eux. Il a été fixé à 8. Le signal est moyenné sur 32 accumulations réalisées après 4 accumulations non enregistrées (dummy scans) pour mettre le système à l'équilibre. Les spectres ont été reconstruits par une transformée de Fourier contenant 64 k points et avec un lissage exponentiel (lb) de 0,3 Hz. Les acquisitions et les premières étapes du traitement des spectres ont été réalisées sous TopSpin 3.5 p15.

6.2.2 Spectres RMN 2D

Suite aux résultats obtenus dans les analyses statistiques, nous avons choisi de nous concentrer sur les différences entre les deux méthodes de fermentation. Pour mener à bien l'identification, l'analyse doit souvent être effectuée avec un pool pour s'assurer que les composants de l'ensemble soient assez concentrés. Par conséquent, quatre pools d'échantillons ont été préparés pour la suite (Tableau 4) :

Tableau 4 : Regroupement des échantillons pour les expériences de spectres 2D

	Partie Gastrique	Partie Intestinale
Levure de Boulanger	Pool 1 (G-B)	Pool 3 (I-B)
Levain	Pool 2 (G-L)	Pool 4 (I-L)

Chaque pool contient 16 échantillons choisis au hasard (4 farines * 4 temps * 1 type de fermentation * 1 répétition), la quantité de prélèvement est aléatoire et de l'ordre de 30~40 µL. (chaque pool contient donc environ 500 µL d'échantillon). Les quatre pools sont ensuite centrifugés et évaporés dans un CentriVap pour concentrer les molécules et éliminer l'eau (concentration trop élevée pour de l'identification). Enfin, chaque échantillon a été repris dans 600µL de D₂O. Les spectres RMN ont été enregistrés avec le même instrument que celui décrit dans la section précédente.

La durée d'acquisition d'une série de spectres 2D (COSY, TOCSY, HMQC, HSQC) est très gourmande en temps pour tous, 6h pour chaque spectre homonucléaire, 11h pour chaque spectre hétéronucléaire. L'analyse complète n'a été réalisée que pour le premier échantillon (G-B). Pour le deuxième échantillon (G-L), seulement les spectres les plus importants, c'est-à-dire COSY et HSQC, ont été réalisés. Après l'analyse, le pool 1 a été mélangé au pool 2 et les spectres ont été refaits pour avoir une meilleure compréhension de la composition de l'échantillon. Les échantillons intestinaux n'ont

pas été analysés par RMN 2D. Les spectres ont été traités et analysés sous TopSpin 3.5pl5.

6.3 ANALYSE DES DONNEES

6.3.1 RMN métabolomique

Les premières opérations de traitement des spectres RMN 1D sont réalisées sous TopSpin 3.5 pl5 : transformée de Fourier, correction de la phase du spectre puis de la ligne de base à l'aide d'une macro spécifique. L'ensemble des spectres 1D est ensuite importé sous NMRProcFlow (<https://nmrprocflow.org/np/>) [2], une solution d'analyse des spectres RMN pour des études de métabolomique en ligne. Les spectres sont manuellement réalignés, et sont ensuite divisés en buckets et la surface totale de chaque buckets est calculée pour représenter le spectre original [3]. Le découpage est réalisé de manière automatique puis optimisé manuellement. A la fin de cette étape, un tableau Excel (la bucket table) est obtenu contenant l'identification des différents échantillons ainsi que l'aire du signal RMN dans chacun des buckets.

Les analyses statistiques ont été réalisées sous R Studio (avec R version 4.1.1). L'importation de données est faite à l'aide des packages(readxl). Pour analyser, nous avons choisi deux packages (mixOmics) [4] et (ropls) [5] pour faire des analyses ACP, PLS-DA et OPLS-DA. Afin d'intégrer les données, tous les rangements sont effectués dans R Studio. Le traitement des paquets nécessite 3 différentes variables : matrice de données, métadonnées d'échantillons et métadonnées de variables. Il est donc nécessaire de séparer les facteurs et les données. Ensuite, les variables peuvent être développées avec des lignes de commandes. Nous avons utilisé deux packages, où mixOmics est plus simple et facile à analyser, mais ropls présente des meilleures opérations pour valider le modèle. Ces deux packages complètent l'analyse. Le regroupement de base a été effectué par des analyses PCA, puis analysé plus en détail par PLS-DA ou OPLS-DA pour nos données d'intérêts, dont les résultats sont discutés au chapitre 8.

6.3.2 Analyse des spectres 2D

Tout d'abord, les spectres RMN des échantillons mélangés ayant servi à enregistrer les spectres RMN 2D sont réalignés avec la série des spectres RMN 1D utilisée pour l'analyse métabolomique. En effet, comme le pH d'échantillon n'est pas contrôlé, il peut exister un léger décalage dans les pics RMN (Figure 4).

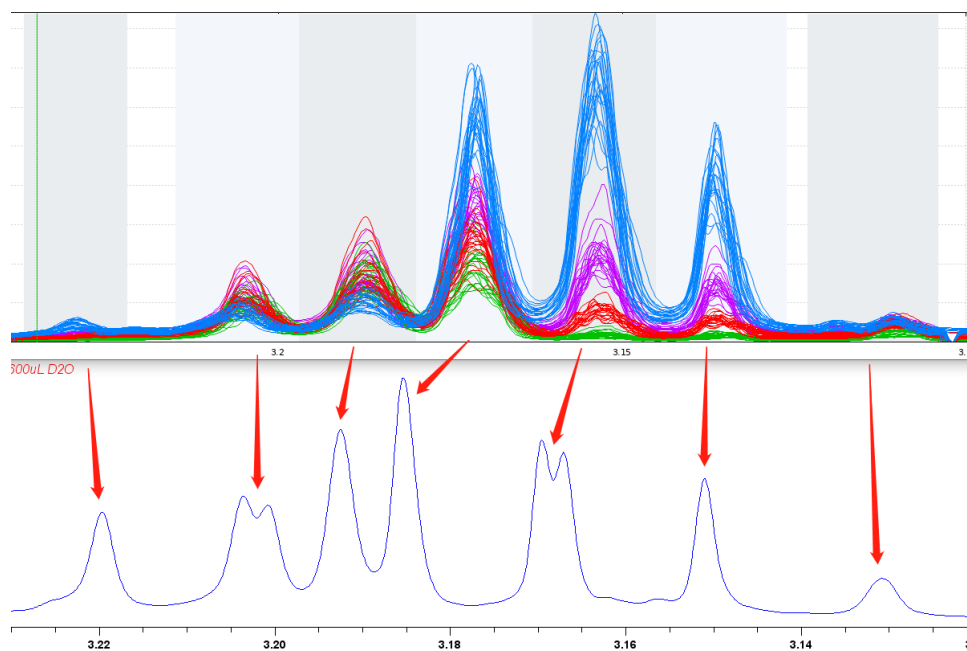


Figure 4 : Spectres RMN ayant servi à l'analyse métabolomique (A) et celui enregistré lors de l'enregistrement des spectres 2D (B). Les flèches rouges illustrent les décalages pouvant exister.

7 RESULTATS ET DISCUSSION

7.1 ANALYSE PRELIMINAIRE DES SPECTRES RMN

Avant d'étudier en détails les spectres RMN, la première étape du travail consiste en une observation globale du jeu de données. En regardant tous les spectres et en les comparant, nous pouvons conclure que :

- Les spectres RMN obtenus dans un même compartiment (gastrique ou intestinal) se ressemblent fortement. C'est principalement l'intensité des signaux qui varie en fonction des conditions expérimentales pour chacun des compartiments ;
- A l'inverse, les spectres de la partie gastrique et ceux de la partie intestinale ne sont pas comparables (Figure 5). Par exemple, nous pouvons remarquer de fortes différences dans la région des aromatiques (7-8 ppm) ou entre 1 et 2 ppm ;
- De plus, nous pouvons également noter une forte différence dans l'intensité des signaux avec un signal élevé pour les spectres intestinaux (facteur 8 environ) ;
- L'analyse des spectres des blancs (Figure 6) montre que le cocktail d'enzymes utilisé dans la partie intestinale donne de nombreux et intenses signaux RMN tandis que le spectre RMN de contrôle de la partie gastrique ne présente que le pic de l'eau.

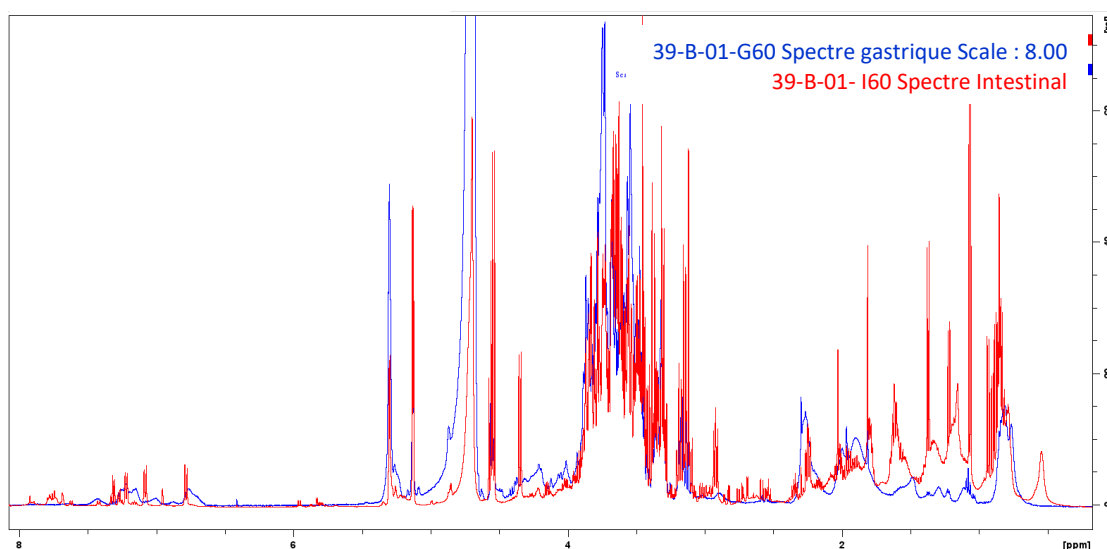


Figure 5 : Exemple de spectres intestinal (rouge) et gastrique (bleu). Les deux spectres ont été obtenus 60 minutes après le début de la digestion. L'intensité du spectre gastrique a été multiplié par 8 pour avoir des signaux comparables à ceux obtenus dans le compartiment intestinal.

La digestion de l'estomac et celle de l'intestin grêle sont complètement différentes dans leur fonction et leur action. Les spectres RMN n'étant pas comparables, il n'est pas possible de les analyser lors d'une même et unique analyse statistique. En conséquence, les spectres ont été divisés en deux groupes : la partie gastrique et la partie Intestinale.

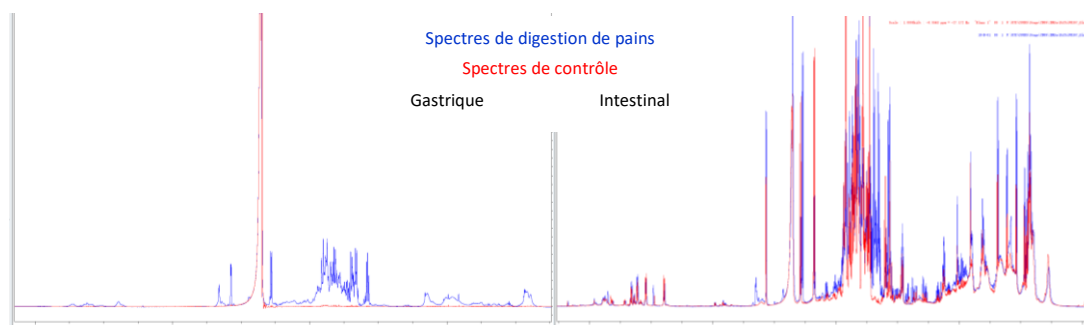


Figure 6 : Comparaison des spectres lors d'une digestion contenant du pain (bleus) ou pas (rouge). Les spectres de la partie gastrique sont représentés à gauche et ceux de la partie intestinale à droite.

Il est ensuite possible d'analyser l'évolution du signal RMN au cours du temps pour tous les spectres d'un même compartiment. Cette opération est plus facilement réalisée dans la solution logicielle en ligne NMRProcFlow. La Figure 7 représente l'ensemble des spectres obtenus dans le compartiment gastrique avec une coloration des spectres en fonction du temps de prélèvement. Nous pouvons voir que l'effet temps est très important. En effet, la plupart de pics augmentent en intensité avec le temps. En choisissant de mettre en évidence les différents types de farines ou le processus de fermentation utilisé (type de levures), aucune séparation claire n'est évidente (figures non montrées).

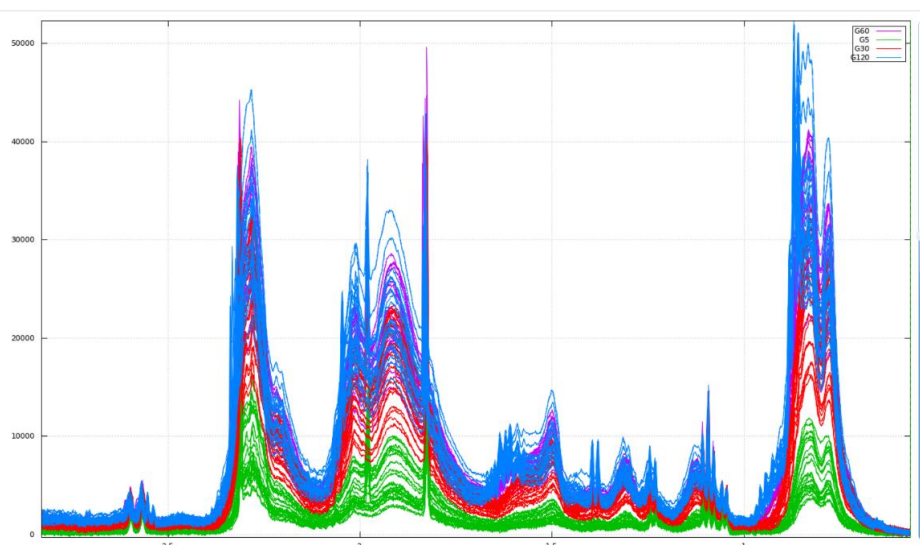


Figure 7 : Superposition de quelques spectres gastriques. La couleur des spectres dépend du temps de digestion : 5 min (vert), 30 min (rouge), 60 min (violet) et 120 min (bleu).

Le même travail a été réalisé sur les spectres intestinaux. Un léger effet du temps a pu être observé et aucun effet pour les deux autres facteurs.

Nous pouvons également constater de légers décalages entre les spectres montrant alors la nécessité de devoir les aligner.

7.2 TRAITEMENT DES SPECTRES RMN

Nous avons vu dans les paragraphes précédents qu'il était indispensable de considérer comme des jeux de données différents les spectres obtenus dans le compartiment gastrique et ceux obtenus dans le compartiment intestinal. A partir de maintenant, ces deux jeux de données sont donc séparés.

L'un des défis les plus importants de l'analyse métabolomique par RMN est l'apparition de décalages entre les pics de différents spectres causés par des fluctuations du pH, de la température, des facteurs de l'instrument, de la teneur en ions, etc. L'alignement est donc souvent une étape cruciale du prétraitement avant l'analyse quantitative. La Figure 8 illustre l'effet de ces traitements sur les spectres de la partie gastrique.

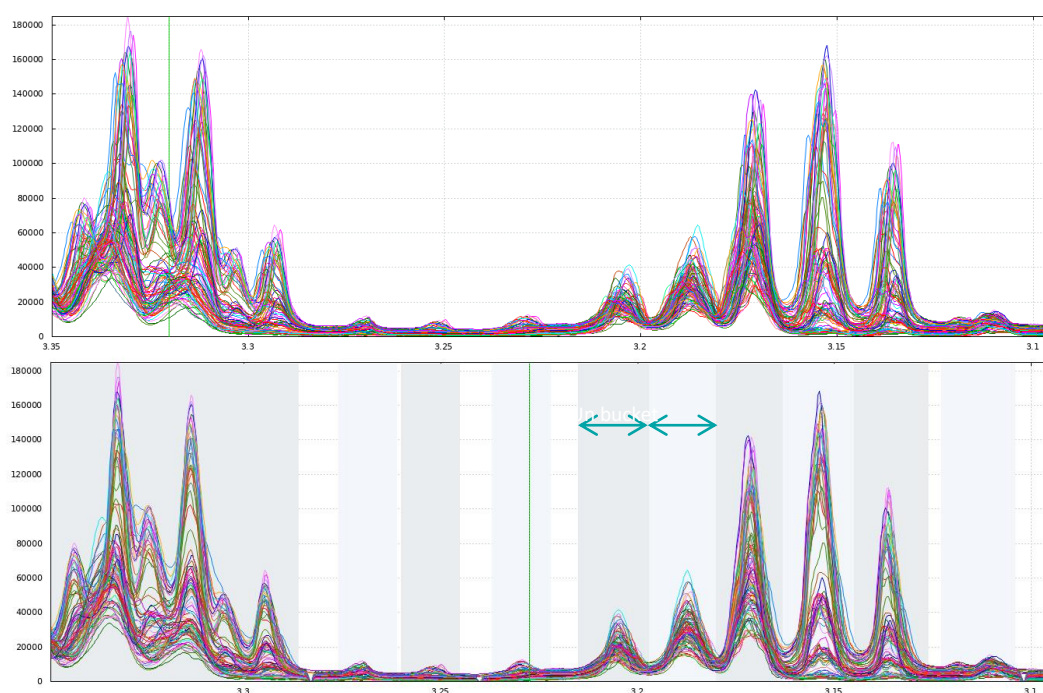


Figure 8 : Exemple d'alignements des spectres et de bucketing sur les spectres de la partie gastrique. Avant (haut) et après (bas) traitements.

Les spectres gastriques présentent moins de décalage que les spectres intestinaux. Ceci est probablement dû à de plus faibles variations du pH. L'alignement des spectres peut entraîner l'apparition d'artéfacts et au final ne pas être la solution la plus adaptée. Pour

éviter ce problème, il existe deux stratégies qui peuvent être mises en œuvre. La première consiste à choisir raisonnablement les zones à aligner pour justement éviter l'apparition de ces artéfacts dans les zones où il y avait des signaux. La seconde option consiste à ne pas (bien) réaligner les signaux mais à prendre cette information en compte lors du choix de la taille du bucket dans cette zone. Ce dernier choix ne doit cependant être réalisé que pour un faible nombre de zones.

Le bucketing est une méthode très utile pour réduire la dimension des données, cette étape sert à intégrer les pics dans une gamme de déplacements chimiques (un bucket) pour visualiser une seule valeur numérique représentant l'intensité du signal RMN dans cette région. Il existe trois types de bucketing : bucketing uniforme, bucketing variable et bucketing dit « intelligent ». Le bucketing uniforme, comme son nom l'indique, consiste à diviser en région de même taille le spectre RMN. En raison de la division arbitraire des pics, un bucket peut contenir des morceaux de deux pics ou plus, ce qui peut affecter l'analyse des données, surtout lorsque l'alignement des spectres n'est pas parfait. Lorsqu'il y a plusieurs métabolites dans un bucket, il est difficile de connaître le rôle de ces métabolites dans le processus métabolique. J'ai donc choisi d'utiliser la méthode plus adaptive, le « bucketing intelligent » qui tente de diviser automatiquement les spectres de façon à ce que chaque zone commune à tous les spectres contienne les mêmes résonances, c'est-à-dire appartenant au même métabolite. Dans cette méthode, la largeur de chaque zone est alors déterminée par la différence maximale de déplacement chimique entre tous les spectres. Cependant, les artéfacts dans les spectres vont interférer avec les calculs du logiciel et donc facilement être ignorés par le système entraînant des erreurs. Pour cette raison, le « bucketing intelligent » est généralement utilisé comme une aide, et éventuellement la zone de bucketing est ensuite optimisée manuellement, ce qui revient à faire du bucketing variable. Une fois la sélection des buckets terminée, le logiciel calcule automatiquement l'aire des pics dans chaque bucket après une normalisation pour éliminer les variations systématiques entre les échantillons et les rendre comparables entre eux. Cette normalisation est effectuée en gardant la somme totale des aires du spectre constant. La sortie de ces opérations est un tableau contenant l'intensité des signaux RMN pour chacun des buckets, l'identification des buckets (déplacement chimique) et celle de l'échantillon (farine, levure, temps). La Figure 9 illustre une partie du tableau obtenu en sorti de ces traitements.

Après avoir obtenu les tableaux de données, il faut utiliser les méthodes statistiques pour faire ressortir les informations cachées dans la masse de données, apparemment désorganisées, et identifier les lois internes des échantillons étudiés.

Dans un premier temps, le tableau est « nettoyé ». Les cellules négatives sont toutes remplacées par la valeur 0. Les valeurs négatives sont probablement dues au fait qu'il n'y a pas de pics dans ce bucket, seulement du bruit, ce qui, après normalisation, rend la valeur inférieure à 0 et n'a donc aucune signification. Ensuite, elles ont été centrées réduites pour réduire les différences qui peuvent exister entre les répétitions. Etant donné que les spectres varient entre les trois répétitions, certaines analyses sont faites avec des données centrées réduites par répétitions car nous nous sommes aperçus que cela permettait de mieux éliminer les différences entre les répétitions.

Samplecode	Fermentation	Temps	Farine	Repetition	B8_4430	B8_3615	B8_2010	B7_8445	B7_4765	B7_3410	B7_2825	B7_2355	B7_1720	B7_1025
18-B-01-G5	B	G5	w18	R01	-28.03104308	21.5282358	-45.45086734	-34.33461098	31.23575184	3.270949542	102.3402385	119.7603651	194.8637915	9.287739469
18-B-02-G5	B	G5	w18	R02	-2.495792374	24.47269963	-1.587188565	25.23671843	42.10175979	8.761989641	133.8557234	144.2501777	221.8850754	12.86391214
18-B-03-G5	B	G5	w18	R03	-16.85304048	24.5816613	-7.504109555	-2.967176739	22.39398057	2.74622558	75.88894039	81.92749556	129.4008261	8.703319157
18-B-01-G30	B	G30	w18	R01	43.93264313	36.83481088	191.7063162	252.517793	414.7132898	39.93323426	295.4098131	302.5346616	522.0560963	44.553037
18-B-02-G30	B	G30	w18	R02	47.56303788	38.72764694	207.4148477	256.5087623	424.6983032	44.28118432	281.099469	292.6480944	499.4770376	41.58508454
18-B-03-G30	B	G30	w18	R03	24.9500422	35.85252001	140.6309488	196.1799701	321.8772083	31.0648097	215.8956657	223.0775866	376.5311054	32.37457016
18-B-02-G120	B	G60	w18	R01	94.73718413	50.30751326	347.0730613	399.7476087	654.5094913	65.61092016	323.8087959	334.4938306	569.7568017	52.88695607
18-B-03-G120	B	G60	w18	R02	126.1001662	62.55205099	446.4721782	374.7701245	803.7427999	67.81889973	292.9713427	309.6318777	517.3345048	55.31550321
18-B-01-G120	B	G60	w18	R03	46.32629008	37.70093951	215.6216571	305.9196094	453.8774807	56.64111297	303.0144915	321.8489501	534.0950992	52.34609168
18-B-02-G120	B	G120	w18	R01	144.2489584	58.92715307	475.1000376	497.7024836	887.205533	105.2031211	345.8472099	364.5513606	576.1385889	71.12161283
18-B-03-G120	B	G120	w18	R02	363.520103	122.0048921	866.1681238	434.9400271	1186.508392	118.4002871	309.1487715	335.9268972	525.3539881	87.76691925
18-B-01-G5	L	G5	w18	R01	288.8520898	106.4370119	734.4085109	392.7909072	1067.684066	108.4867499	289.473274	312.6991294	478.8478661	82.05531743
18-L-02-G5	L	G5	w18	R02	51.69008941	40.83943187	205.4841732	79.20924635	378.7953726	12.33539257	129.0005477	136.8383595	231.3259607	18.72569075
18-L-03-G5	L	G5	w18	R03	17.39644893	30.75812681	113.3364371	81.47293009	282.3845426	16.6673177	124.5760239	137.3120829	221.3504016	19.69634044
18-L-01-G30	L	G30	w18	R01	-7.982089827	15.16745683	-14.27860389	-0.315338732	21.53830205	0.225750084	74.10685671	83.02324783	135.3639211	8.608562086
18-L-02-G30	L	G30	w18	R02	9;					1785826	332.4241032	339.35135	594.7692967	50.47870812
18-L-03-G30	L	G30	w18	R03	5;					5655652	307.501628	314.1497734	543.6448048	47.92121536
18-L-01-G60	L	G60	w18	R01	4.					1333275	175.2797059	194.6309684	309.0921946	28.02441776
18-L-02-G60	L	G60	w18	R02	3;					1461056	309.2151931	322.5616611	576.3117615	71.83170545
18-L-03-G60	L	G60	w18	R03	72.87469178	38.72028957	291.9553137	352.3933599	604.9491777	55.65808985	306.8782957	326.4851921	551.0631473	51.2603941
18-L-01-G120	L	G120	w18	R01	50.91183662	30.51487469	211.11049	275.2774802	425.4225344	55.22616802	266.1927754	283.5706201	474.4073814	47.80842177
18-L-02-G120	L	G120	w18	R02	125.2850148	53.57011063	422.5512681	427.440863	809.2311281	91.20286968	318.7046031	342.4443455	543.2458315	64.25426988
18-L-03-G120	L	G120	w18	R03	108.0718247	46.31459047	381.6187448	469.7650757	756.6070214	101.181826	340.0608214	364.6092932	571.5409666	69.08269461
24-B-01-G5	B	G5	w24	R01	138.5653731	53.6737107	448.5644679	407.342187	788.8610336	100.3881469	283.1585589	302.8612141	462.2722117	70.76600015
24-B-02-G5	B	G5	w24	R02	1.706713128	40.58410725	49.13170966	59.80304413	167.8905969	13.67912679	146.2845973	160.2599873	276.144751	26.98463329
24-B-03-G5	B	G5	w24	R03	0.479513121	37.54336179	77.93503575	58.57379425	200.6604935	27.46249675	154.9031668	171.3460586	279.0251258	25.99063374
24-B-01-G30	B	G30	w24	R01	-5.079178107	46.02832996	15.11132134	18.27435386	50.70302328	9.004760256	84.18181648	96.03605292	160.2553895	16.66170137
24-B-02-G30	B	G30	w24	R02	36.42112426	38.73243863	156.590316	216.0773894	352.0118121	41.44217535	352.5666691	375.1773649	657.7569386	54.0685205
24-B-03-G30	B	G30	w24	R03	36.08966973	39.37166396	162.5006896	220.9040263	372.4257941	42.1310108	342.8328838	367.5541089	648.9014851	53.16735636

Figure 9 : Illustration d'une partie du tableau de données (aussi appelé bucket table) obtenu après tous les traitements réalisés sous NMRProcFlow.

7.3 ANALYSE DU COMPARTIMENT GASTRIQUE

Il y a 4 facteurs permettant de distinguer les données : les farines (w18, w24, w39, w59) ; les méthodes de fermentation (levure de boulanger et levain) ; le temps (G5, G30, G60, G120, I60, I120, I180, I240) et les répétitions (R1, R2, R3). Evidemment, les répétitions sont réalisées pour évaluer la robustesse des résultats et ce paramètre n'est pas présenté plus en détail par la suite.

La première analyse statistique est une analyse en composantes principales (PCA). En effet, la PCA est une méthode statistique multivariée non-supervisée la plus classiquement utilisée. Il s'agit d'une manière d'identifier des modèles dans les données de façon à mettre en évidence leurs similitudes et leurs différences. Comme les modèles de données peuvent être difficiles à trouver dans des données de haute dimension, où

le luxe de la représentation graphique n'est pas disponible, la PCA est un outil puissant pour réduire le nombre de dimensions des données sans grande perte d'informations.

Le résultat de la PCA est représenté sur la Figure 10. Nous pouvons constater que le temps a un impact très significatif sur les données. Au fil du temps, ces points se décalent le long de l'axe x (la composante principale 1) et présentent également un certain décalage sur l'axe de y (la composante principale 2). Deux des ensembles de données, G30 et G60, sont relativement proches l'un de l'autre et sont situés au centre, tandis que G5 et G120 sont excentrés.

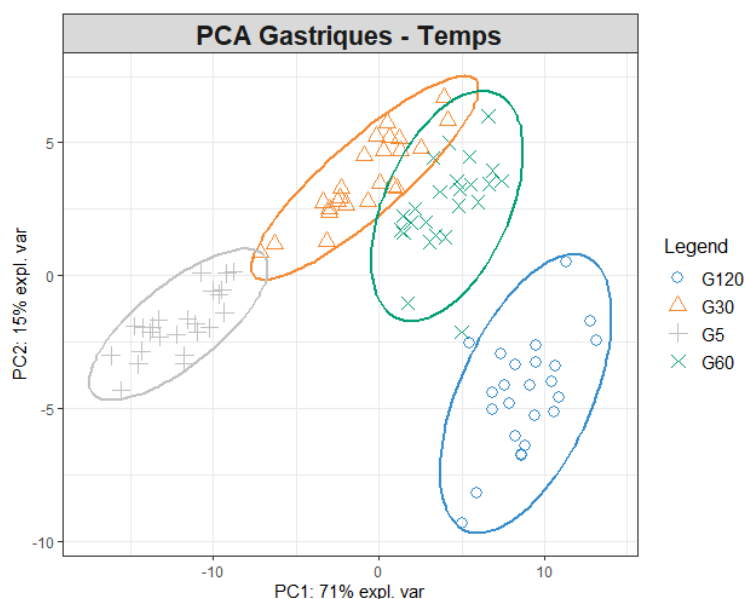


Figure 10 : PCA pour les données gastriques représentée en fonction du temps de digestion. La variance totale expliquée par les 2 premiers axes (PC1 et PC2) est de 86%.

Cependant, si nous étiquetons les données par méthode de fermentation (Figure 11) ou par farine (Figure 12), l'analyse PCA ne montre pas de regroupement en fonction des conditions.

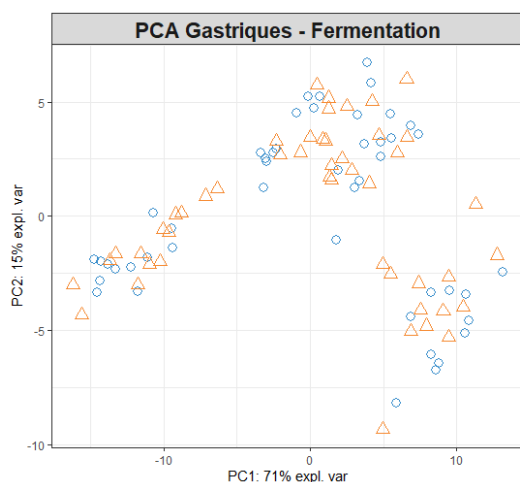


Figure 12 : PCA pour les données gastriques avec un marquage des points en fonction des méthodes de fermentation

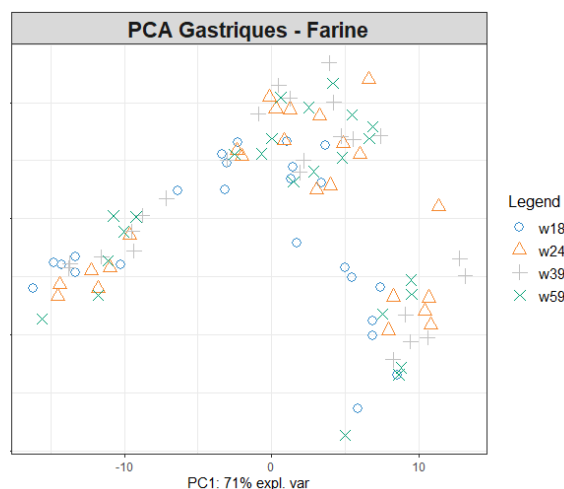


Figure 11 : PCA pour les données gastriques avec un marquage des points en fonction des farines

Nous n'avons pas trouvé d'effet des différentes farines et des différentes levures sur la digestion ; seul le temps de prélèvement est un bon critère de regroupement des échantillons. Malheureusement, il est évident que le temps joue un rôle dominant dans la digestion. Comme l'ACP est une analyse non supervisée et qu'elle ne permet pas de bien distinguer les différents groupes, nous avons décidé d'opter pour une analyse supervisée afin de poursuivre le traitement des données.

La régression des moindres carrés partiels (PLS) est une méthode supervisée pour étudier la relation principale entre deux matrices (X et Y). Contrairement à la PCA, qui cherche un plan de la moindre variance entre la réponse et les variables indépendantes, la PLS se concentre sur la projection de variables prédicteurs et observés dans un nouvel espace pour trouver un modèle de régression. Comme les données X et Y sont projetées dans le nouvel espace, les méthodes de la famille PLS sont connues sous le nom de modèles factoriels bilinéaires. Lorsque Y est une donnée catégorielle, elle devient l'analyse discriminante des moindres carrés partiels (PLS-DA).

Cet ensemble de données gastriques a été analysé plus en détails avec plusieurs analyses PLS-DA. Tout comme avec la PCA, une très bonne séparation des échantillons avec le temps est observée (Figure 13). La représentation obtenue est proche de celle que nous avons eu avec la PCA (Figure 10) mais avec une symétrie le long de l'axe y.

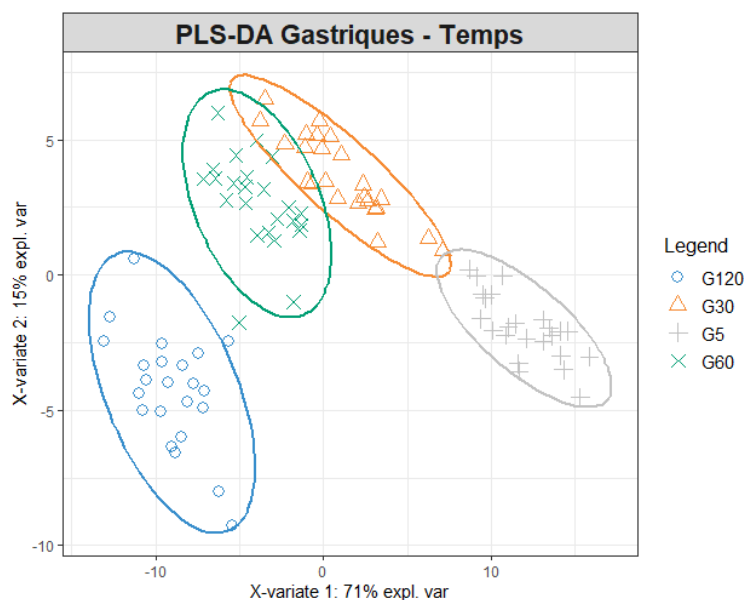


Figure 13 : PLS-DA pour les données gastriques en fonction du temps

Pour la méthode de fermentation (Figure 14), nous constatons une tendance à séparer les deux mais cette séparation ne représente qu'une très faible partie de la variance totale. Les méthodes de fermentation sont séparées le long de l'axe X qui ne représente

que 6 % de la variance. Cela signifie que la différence entre ces deux méthodes de fermentation sur le spectre RMN est faible.

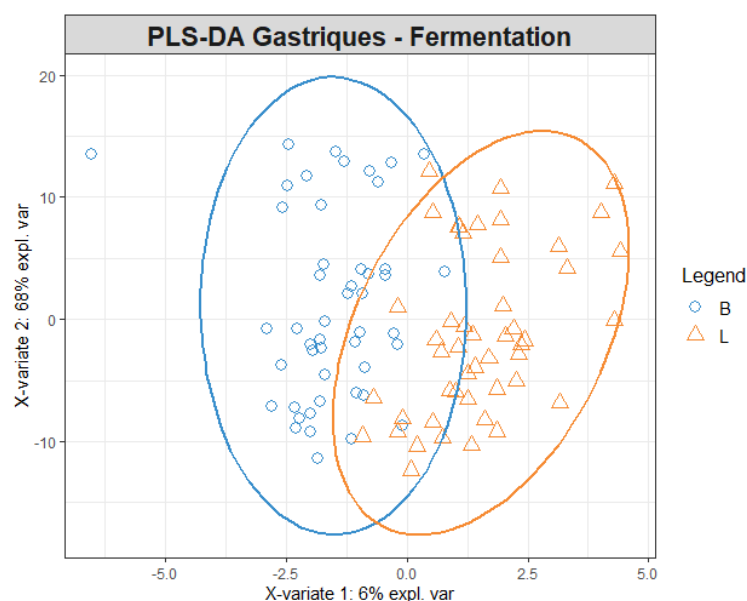


Figure 14 : PLS-DA pour les données gastriques avec une séparation recherchée par les méthodes de fermentation

Concernant les différentes farines (Figure 15), nous pouvons remarquer seulement un petit décalage du groupe de points représentant la farine w18 par rapport aux autres.

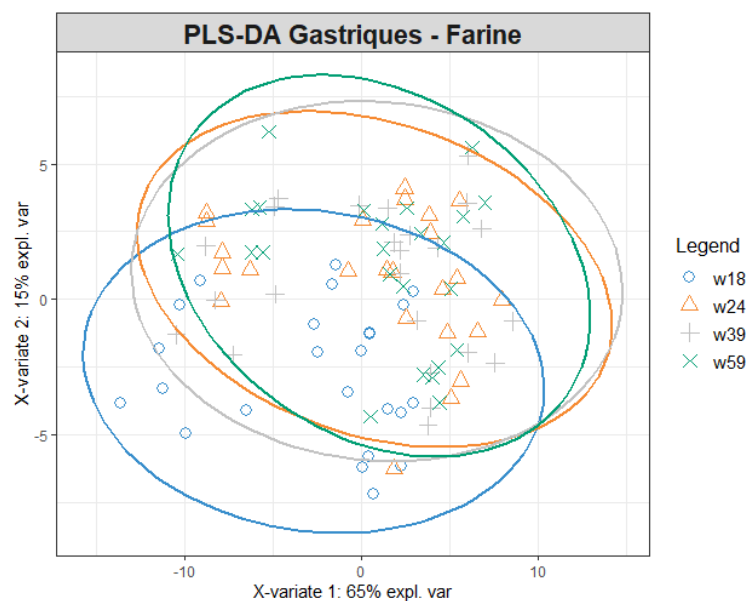


Figure 15 : PLS-DA pour les données gastriques en cherchant à séparer les types de farines.

Afin d'effectuer une analyse plus approfondie, nous avons décidé de regrouper les données G30 et G60 qui possèdent peu de différences entre elles comme démontré dans l'analyse PCA initiale (Figure 10). Cette stratégie nous permet de nous affranchir de l'effet temps de la digestion tout en augmentant le nombre d'échantillons. Nous

commençons par réaliser une analyse non supervisée (PCA) sur ce jeu de données (Figure 16). Il existe un léger décalage des groupes en fonction de la méthode de fermentation choisie. Une représentation des points en fonction du temps montre que ce facteur est toujours dominant dans ce modèle (figure non montrée).

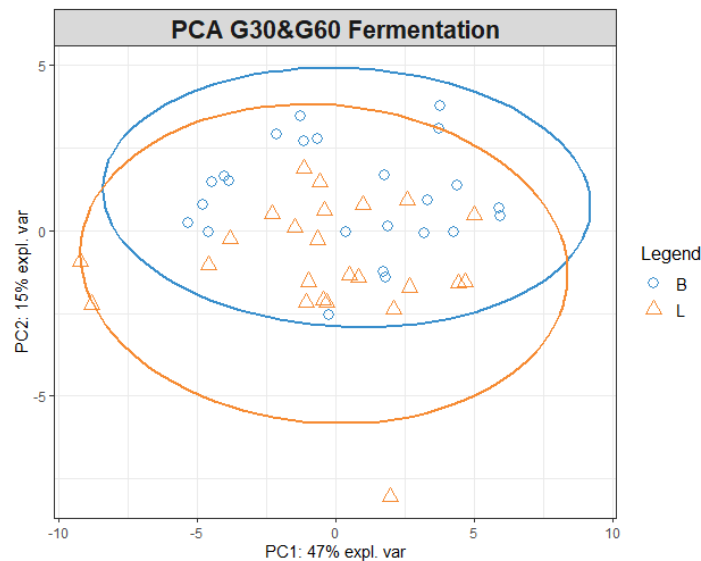


Figure 16 : PCA pour les données gastriques G30 & G60 représentée en fonction de la méthode de fermentation

Nous avons effectué ensuite des analyses supervisées de type PLS-DA. Les modèle concernant la méthode de fermentation est amélioré puisqu'il permet d'expliquer 18% de la variance car les 2 groupes sont séparés le long de l'axe X (Figure 17). Concernant les farines, les résultats restent identiques avec une tendance à séparer w18.

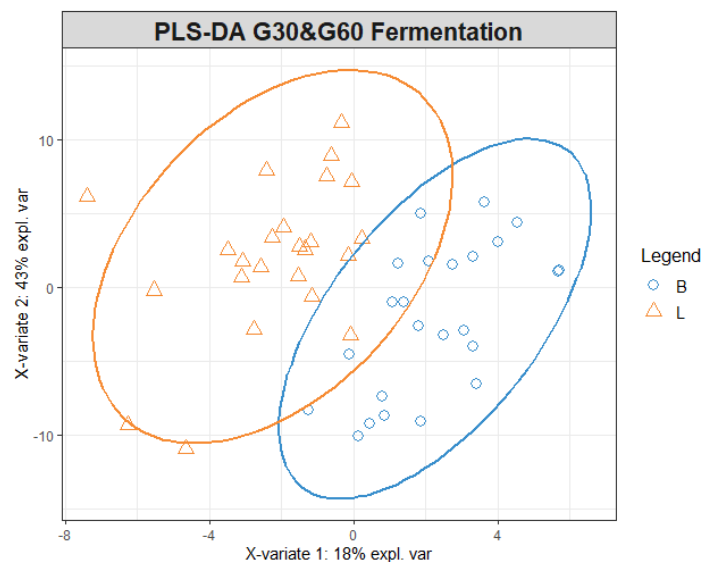


Figure 17 : PLS-DA pour les données gastriques G30 & G60 en fonction des méthodes de fermentation

Vu que le modèle PLS-DA souffre d'un problème de surapprentissage, j'ai essayé d'établir un modèle OPLS-DA (Analyse discriminante PLS orthogonale). L'analyse OPLS-DA a été introduite comme une amélioration de l'approche PLS-DA pour discriminer deux groupes à l'aide de données multivariées. L'avantage évident de l'OPLS-DA par

rapport à la PLS-DA est qu'une seule composante sert de prédicteur pour la classe, tandis que les autres composantes décrivent la variation orthogonale à la première composante prédictive. Cette analyse se focalise sur l'analyse binaire, elle est donc très adaptée à l'analyse entre les deux méthodes de fermentation. L'OPLS-DA pourrait nous permettre de mieux distinguer les deux. Un modèle de OPLS-DA a été fait aussi pour valider ce résultat. Nous pouvons voir sur la Figure 18, graphe en bas à droite, que la variance expliquée sur l'axe t1 n'est que de 6%. Comme précédemment, cette séparation n'explique pas beaucoup des changements qui existent dans les spectres RMN.

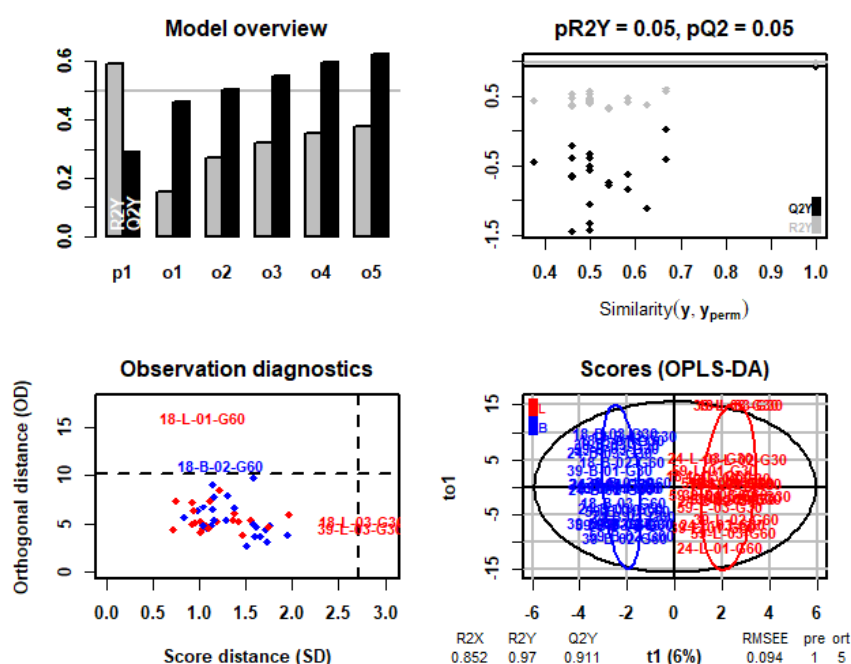


Figure 18 : Modèle OPLS-DA des données gastriques G30 & G60

Vu que l'effet temps était encore prépondérant, les deux types de fermentations ont été analysés pour chaque temps. Les modèles PLS-DA ayant le problème de surapprentissage, les modèles de OPLS-DA ont été effectués (Figure 20 & 21). Cependant, les modèles pour G30 et G120 n'ont pas été effectués, car la 1^{ère} composante prédictive n'aurait pas été significative car négative (Figure 19).

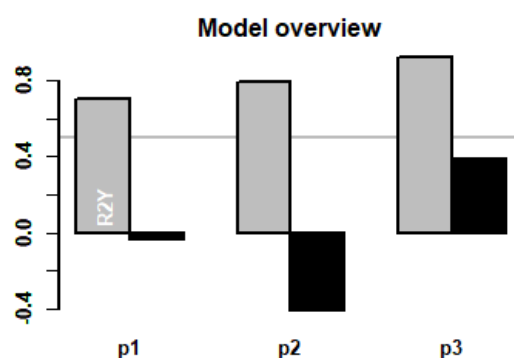


Figure 19 : OPLS-DA modèle pour les données gastriques G30

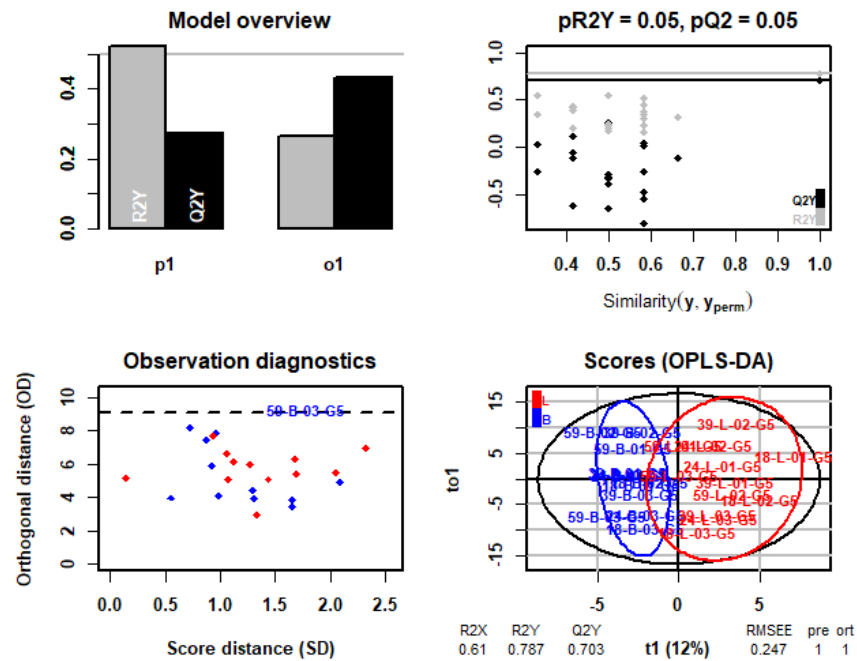


Figure 20 : OPLS-DA modèle pour les données gastriques G5

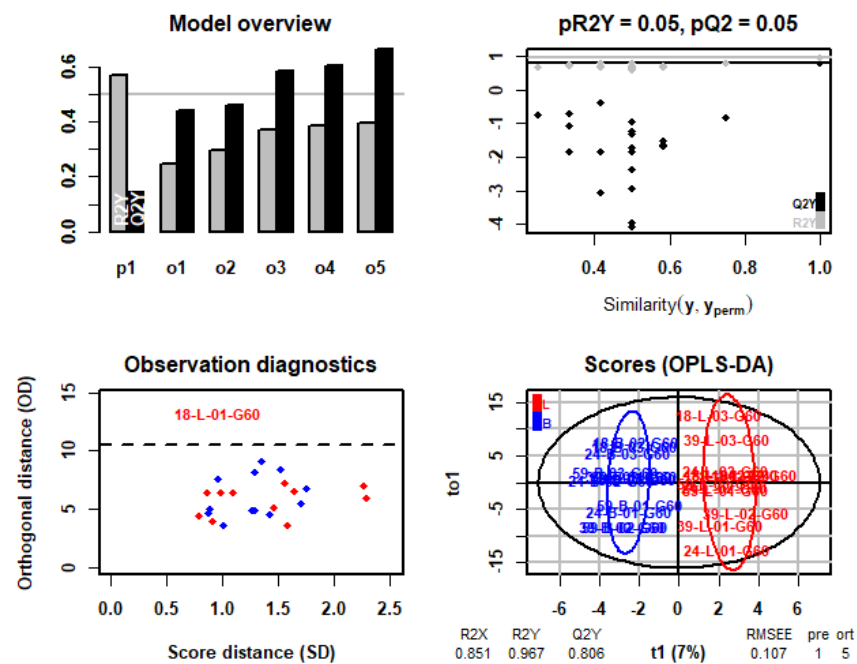


Figure 21 : OPLS-DA modèle pour les données gastriques G60

Bien que ce modèle puisse être construit, nous pouvons constater en regardant le test de permutation dans la figure en haut à droite que le modèle a probablement été sur-ajusté. Les valeurs simulées par le modèle (dispersion, R2Y) sont très proches des valeurs réelles (ligne horizontale). Dans le cas du modèle OPLS-DA pour G60, même si la valeur R2Y n'a pas dépassé la limite, il est évident que ce n'est pas un modèle parfait.

Pour mener à bien l'identification structurale, nous nous sommes basés sur tous les modèles ne mettant pas en évidence l'effet temps (modèles OPLS-DA G5, OPLS-DA G60, OPLS-DA G et PLS-DA G). Nous avons considéré comme significatif les buckets qui possédaient une valeur de VIP (importance des variables pour la projection) supérieure à 0,9 . La valeur VIP signifie la contribution au modèle, ce qui a une valeur VIP élevée contribue davantage au modèle et donc important pour identifier les métabolites. Bien qu'il soit conventionnel de choisir 1 comme valeur de coupure, nous avons décidé de choisir 0,9 car il y avait 2 buckets qui présentaient des valeurs dans l'intervalle 0,9 à 1. De plus, ces buckets sont à proximité d'autres significatifs. En outre, certains buckets contiennent plusieurs pics et peuvent donc masquer les pics les plus importants. Cela nous permet d'obtenir une meilleure taille d'échantillon pour l'identification.

7.4 IDENTIFICATION PAR RMN 2D

Après avoir corrélié les signaux RMN des nouveaux échantillons ayant servi à réaliser les analyses multidimensionnelles avec ceux des spectres utilisés pour les analyses statistiques, le travail d'identification peut être réalisé. Dans ce qui suit, j'ai décidé de me focaliser sur une identification particulière.

Le spectre COSY représente les corrélations directes des protons (en J3), soit les protons séparés par un atome de carbone uniquement. L'expérience HMBC quant à elle est une expérience hétéronucléaire qui, dans notre cas, va montrer des tâches de corrélations entre les protons et le carbone auquel il est directement attaché. Dans la Figure 22, nous pouvons remarquer que le signal à 3,18 ppm est corrélé avec le proton ayant un déplacement chimique de 4,58 ppm et qu'il est rattaché au carbone résonant à 75,28 ppm. Le spectre J-Res est également exploité à ce stade pour nous confirmer les multiplicités des signaux.

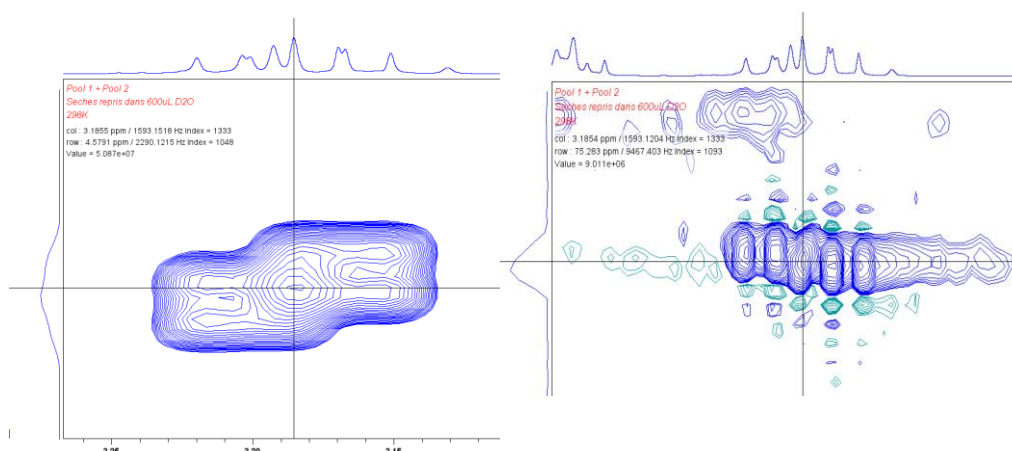


Figure 22 : Exemple d'identification avec les spectres COSY et HMQC.

Ces informations sont utilisées pour rechercher dans des bases de données de possibles molécules [6, 7, 8]. Ensuite, à l'aide des spectres TOCSY et HMBC, nous vérifions si les corrélations entre la molécule candidate et le spectre expérimental sont identiques ou proche. A partir des tâches de corrélations identifiées dans la Figure 21, le glucose est un candidat. La Figure 23 représente la superposition du spectre du glucose pur (tâches vertes) avec le spectre expérimental. Nous remarquons une parfaite superposition et pouvons donc identifier cette molécule.

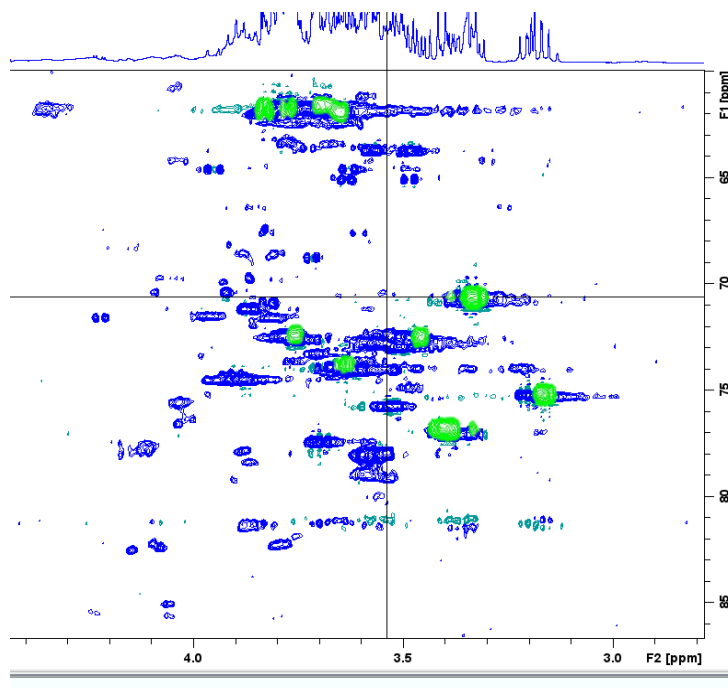


Figure 23 : Exemple de superposition du spectre du pool avec le spectre de glucose.

Cependant, en raison de l'incertitude du pH des échantillons, il peut exister des écarts relativement importants, qui créent alors des incertitudes dans l'identification moléculaire par la seule comparaison numérique. Dans ce cas, nous ne pouvons que proposer la molécule sans pouvoir être plus affirmatif.

Le Tableau 5 présente les différents métabolites qui ont pu être identifiés avec certitude ou pour lesquels nous avons de fortes suspicions. Nous avons identifié 2 sucres, le maltose et le glucose ainsi qu'un marqueur de fermentation, le formiate. Nous pouvons voir que la plupart des métabolites identifiées, qu'ils soient confirmés ou non, sont des sucres. Cela montre que la différence d'un point de vue digestif entre les deux méthodes de fermentation réside dans la vitesse de production des sucres.

Tableau 5 : Tableau d'identification des métabolites (cellules vertes représentent les métabolites identifiés avec confiance, cellules oranges représentent les suspicions des métabolites)

Bucket	VIP	métabolites	1H 2D	multiplicité	COSY	13C (HSQC)
B8_4430	1,344	NH			/	
B8_3615	0,923	Formiate			/	
B8_2010	1,215	NH			/	
B7_4765	1,184		7,52		7,14	137.18 ?
B6_4290	1,240	Fumaric	6,35			136.45 ?
B5_4815	1,103	sucrose	5,41		3,51	100 ?
B5_3025	1,571	maltose	5,34	doublet	3,51	100,7
			5,33	doublet		101
			5,326	doublet	3,56	100
B4_8830	1,117	Pas mannose	4,91		3,53	97,8
			4,88		3,49	99,2
B4_5635	1,157	maltose				
B4_1790	1,252	maltose				
B4_1440	2,784	fructose				
B4_0920	2,861	fructose				
B4_0230	4,218	lactate	4,03	quadruplet	1,25	75,6
B3_9585	3,272		3,983	t?		71,5
			3,966	dd	3.4048/3.8083	64,58
			3,963			
B3_9235	4,928		3,941	dd	3,64/3,617	64,58
			3,936			
B3_8185	0,913		3,916	t	3,587/3,496	74,5
			3,897			
			3,878		3,496	
B3_7230	1,364		3,747			72,64
			3,742			
B3_6570	1,374	melibiose	3,68		3,72	61,74?
		melibiose	3,663		3,33	61,74
		glucose	3,657			73,93
			3,651		3,35	61,93
B3_6135	1,864		3,646			
			3,638	dd?		61,93/73,93
			3,627			61,93
			3,617		3,33/3,34	73,93
			3,612		3,489	
B3_5800	1,442		3,592		3,508	63,7
B3_5485	1,836					
B3_5015	1,123		...			
B3_4570	1,834		?			
B3_4355	1,959		3,452			62,72
			3,445			

B3_4195	1,737		3,432	singulet	3,168	76,93
B3_3970	2,355		3,414	s?		76,98
			3,41			
B3_3770	2,518	glucose	3,399	triplet		77
			3,395			
			3,391			
			3,387	d?		
B3_2060	1,000	maltose	3,2	dd	4,58	75,2
B3_1715	1,422	glucose	3,104		4,56	75,3
			3,097			
B3_1540	2,045		3,0813			
			3,0788			
B2_9740	1,010		?			
B2_3470	1,012	succinate (pyruvate)	2,32			35,4
B1_2415	1,029	lactate	1,1702			21,26
			1,1563			
B1_0940	1,035	éthanol	1.0188?			
B1_0620	1,130	valine	1,07	doublet	3,53	20?
B1_0455	2,019		1,06			
B0_0710	1,173				/	

7.5 ANALYSE DU COMPARTIMENT INTESTINAL

La même stratégie que pour le compartiment gastrique a été mise en œuvre pour analyser les spectres obtenus dans l'intestin. Pendant les analyses, j'ai trouvé que les 3 répétitions présentent un décalage très différent, qui ne réduit pas beaucoup même après avoir centré réduit les données. Pour corriger cet effet, chaque répétition a été centrée réduite individuellement. L'analyse PCA montre un décalage important entre les blancs et les échantillons.

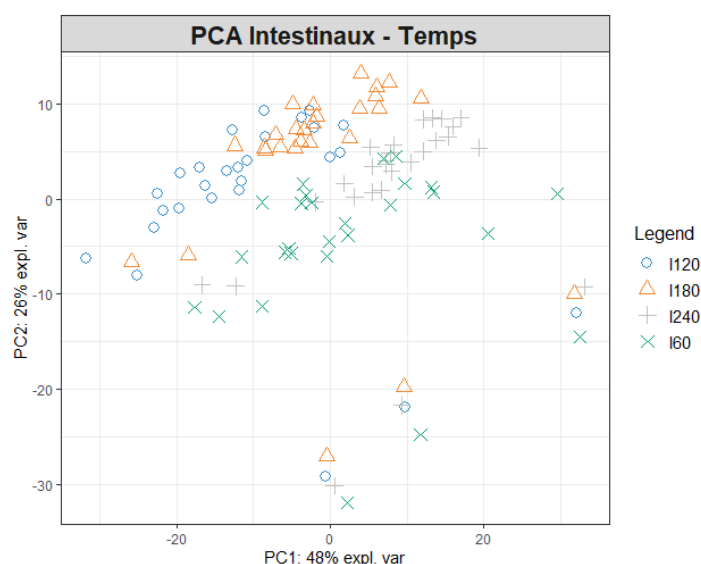


Figure 24 : PCA pour les données intestinales représentée en fonction du temps

Concernant les échantillons, il n'y a aucune séparation évidente ni en fonction du temps (Figure 24), ni pour les farines (Figure 25) ni pour les levures (Figure 26).

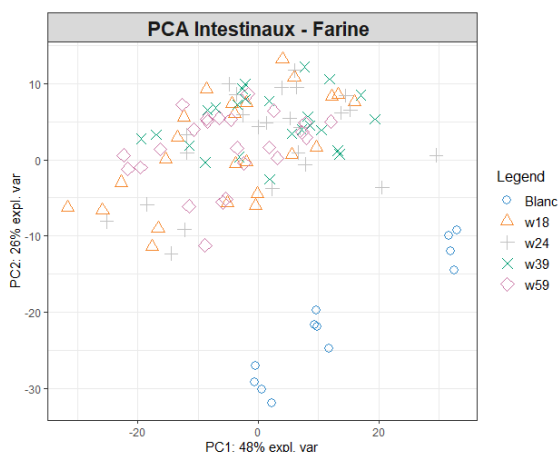


Figure 25 : PCA pour les données intestinales représentée en fonction des farines.

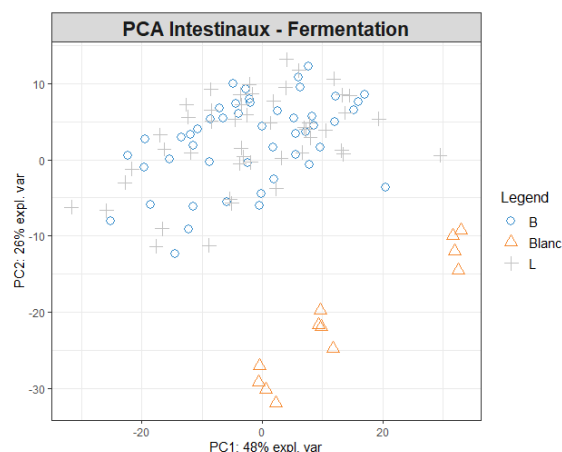


Figure 26 25 : PCA pour les données intestinales représentée pour les différentes méthodes de fermentation.

Nous avons ensuite fait des analyses PLS-DA. Encore une fois, aucun regroupement n'est constaté pour les farines (Figure 27) ou les fermentations (Figure 28).

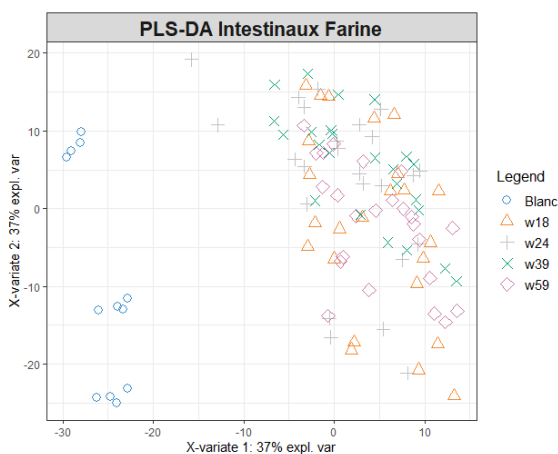


Figure 27 : PLS-DA pour les données intestinales avec un modèle basé sur les farines.

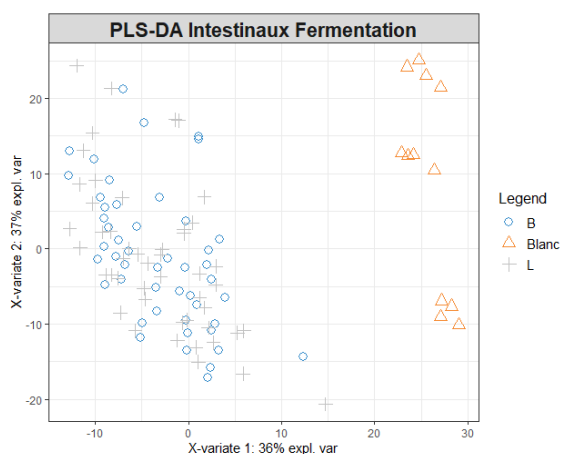


Figure 26 : PLS-DA pour les données intestinales avec un modèle basé sur les méthodes de fermentation.

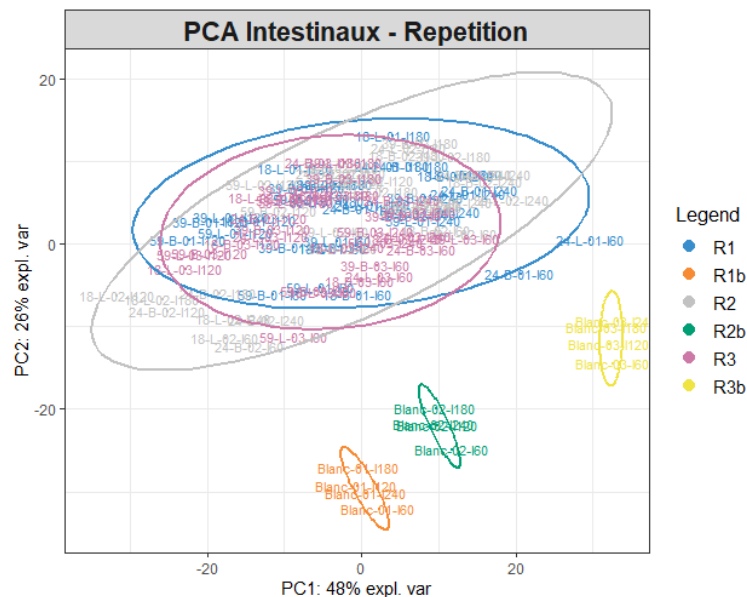


Figure 28 : PCA Plot pour les données intestinales en fonction des répétitions

En étudiant la répartition des points dans le modèle PCA en fonction de la répétition (Figure 29), nous remarquons que dans la deuxième répétition, les données de 18-L et 24-B s'écartaient significativement de l'ensemble. Nous n'avons pas identifié la raison de ce décalage et avons donc décidé de supprimer ces données et de réanalyser les échantillons restants. De plus, comme la différence entre le groupe contrôle et le groupe expérimental était importante, il est probable que cela affectait l'analyse. Nous avons donc retiré le groupe de contrôles des analyses statistiques. Les données restantes sont ensuite centrées réduites.

Les résultats PCA (Figure 30) ont permis de mieux séparer les différents échantillons en fonction du temps de prélèvement.

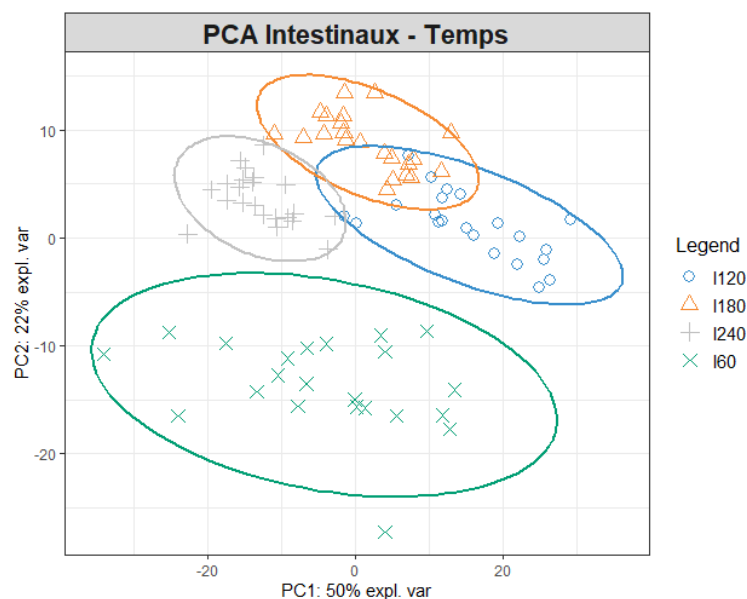


Figure 29 : PCA pour les données intestinales (sans contrôle) en fonction du temps

Pour aller plus loin, nous réalisons donc des analyses supervisées. Les analyses PLS-DA ne montrent pas de séparations évidentes (Figures 31 & 32) même si une tendance pourrait ressortir en fonction de la levure utilisée (Figure 31).

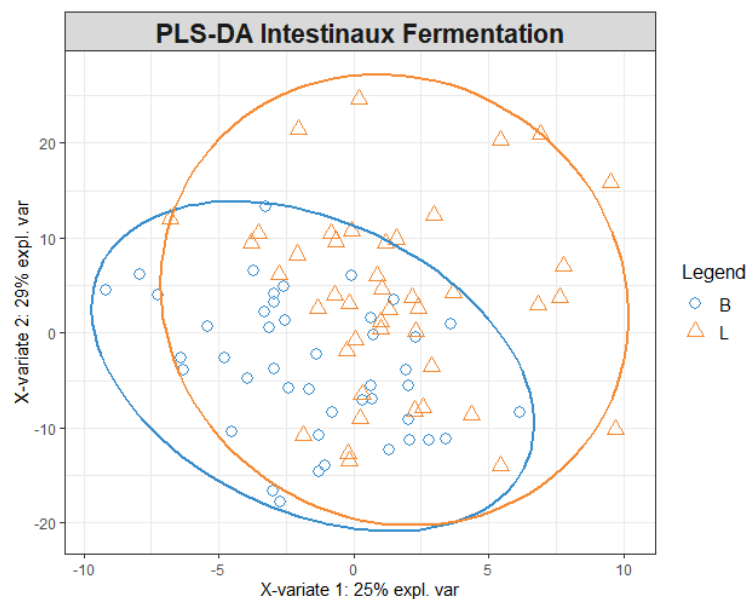


Figure 30 : PLS-DA pour les données intestinales (sans contrôle) en fonction des méthodes de fermentation

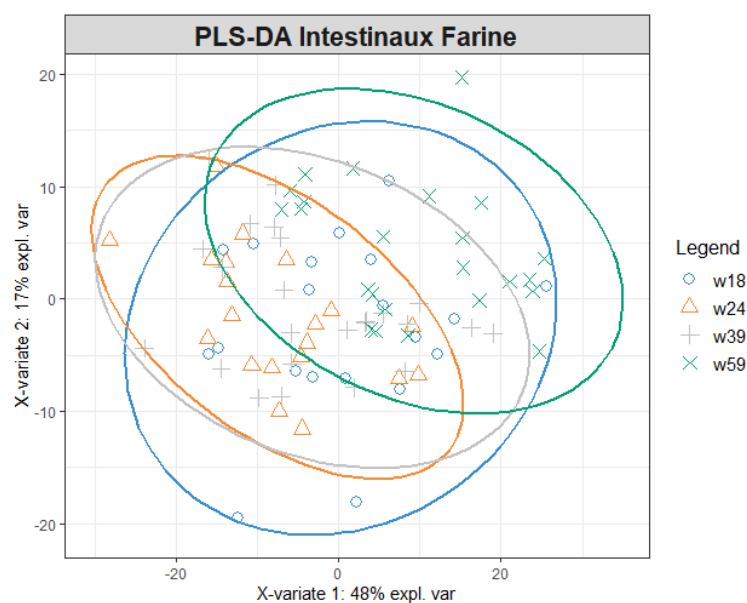


Figure 31 : PLS-DA pour les données intestinales (sans contrôle) en fonction des farines

Afin de réduire l'effet temps, nous avons éliminé les échantillons I60, qui présentent une grande déviation par rapport aux autres. Nous pouvons constater (Figure 33) une meilleure séparation en fonction des différentes levures sans arriver à l'avoir significative.

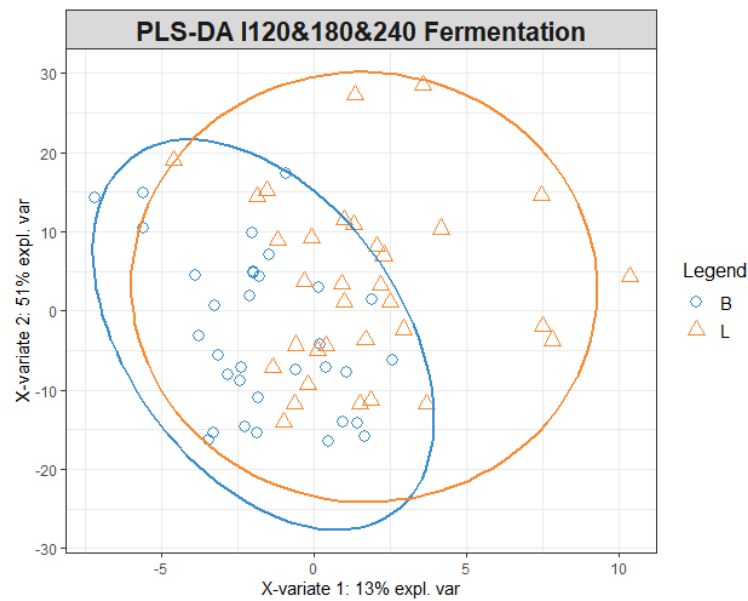


Figure 32 : PLS-DA Plot pour les données intestinales I120 & I180 & I240 (sans contrôle) en fonction des méthodes de la fermentation

En raison de la faible taille de l'échantillon et du fait que ces données n'ont pas pu être ajustées par un bon modèle, une analyse en temps partagé comme celle de la partie gastrique n'a pas été effectuée. Au lieu de cela, les deux méthodes de fermentation ont été analysées séparément pour les farines (Figures 34 & 35). Les résultats ont montré que les différentes farines n'avaient pas d'effet significatif sur le processus de digestion in vitro du pain.

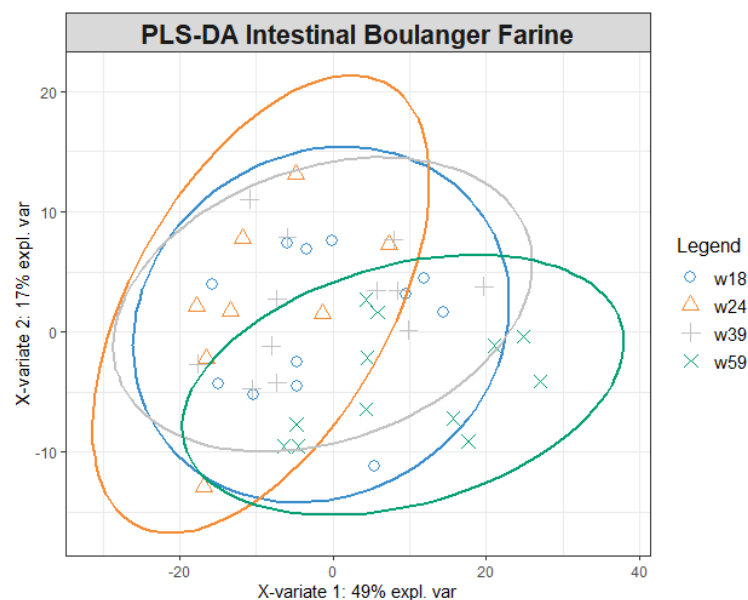


Figure 33 : PLS-DA pour les données intestinales avec levures boulangers (sans contrôle) en fonction des farines

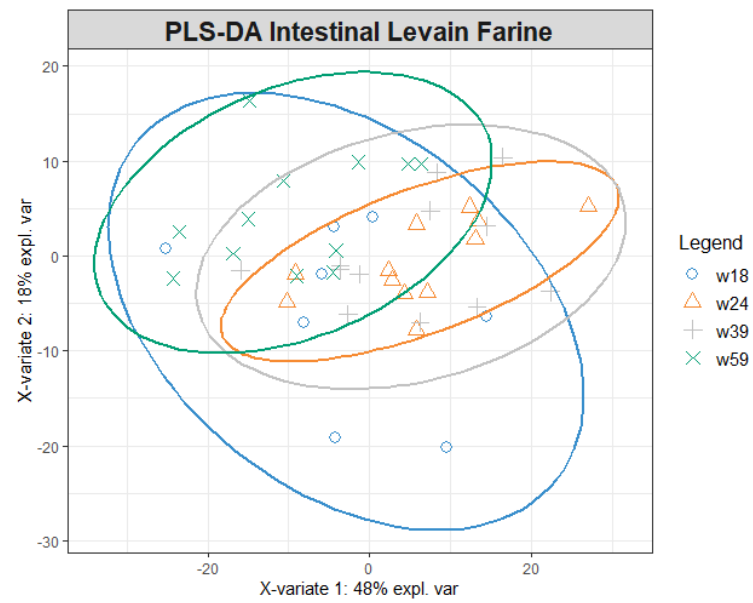


Figure 34 : PLS-DA pour les données intestinales avec levain (sans contrôle) en fonction des farines

8 CONCLUSION

J'ai effectué mon stage d'assistant ingénieur au sein d'INRAE. Lors de ce stage de 6 mois, j'ai pu mettre en pratique mes connaissances théoriques acquises durant ma formation et j'ai également appris des nouvelles connaissances sur l'analyse des données dans le secteur de l'agro-alimentaire. Je tire un bilan très positif de ce stage, qui fut une expérience très enrichissante tant sur le plan professionnel que personnel. Il m'a permis de découvrir le domaine de l'agroalimentaire, et le travail dans un laboratoire.

Grâce à cette expérience, j'ai pu acquérir une grande expérience pratique. Au cours de mon travail, j'ai reçu des formations qui m'ont permises d'approfondir mes connaissances en RMN et de découvrir la métabolomique. J'ai appris à utiliser plusieurs logiciels tels que Topspin, NMRProcFlow, Orange et R pour traiter les spectres RMN et analyser les données, des compétences qui m'aideront dans ma future vie académique et professionnelle.

Pendant ce stage, j'ai eu la chance de pouvoir travailler sur le projet GlutN, en utilisant la technologie RMN métabolomique pour étudier la digestion de différents pains. Au cours de ce travail, j'ai également rencontré des difficultés : manque de connaissances en RMN et métabolomique ; manque de compétences sur les logiciels ; problèmes de communication en français ; le contexte sanitaire ... J'ai surmonté ces difficultés grâce à tous les membres de la plateforme AgroResonance et à plusieurs enseignants de l'UCA qui m'avaient aidé et conseillé avec patience.

Fort de cette expérience, je me suis ainsi orientée vers la filière IAA (Innovation aliments et agro-ressources) pour continuer mes études en vue de mieux connaître les perspectives et les enjeux de l'industrie agroalimentaire.

9 BIBLIOGRAPHIE

- [1] Emwas, Abdul-Hamid, Raja Roy, Ryan T. McKay, Leonardo Tenori, Edoardo Saccenti, G. A.N. Gowda, Daniel Raftery, Fatimah Alahmari, Lukasz Jaremko, Mariusz Jaremko, and David S. Wishart. 2019. "NMR Spectroscopy for Metabolomics Research" *Metabolites* 9, no. 7: 123. DOI 10.3390/metabo9070123
- [2] Jacob, D., Deborde, C., Lefebvre, M., Maucourt, M. and Moing, A. (2017) NMRProcFlow: A graphical and interactive tool dedicated to 1D spectra processing for NMR-based metabolomics, *Metabolomics* 13:36. doi:10.1007/s11306-017-1178-y
- [3] David S. Wishart (2008) Quantitative metabolomics using NMR, *TrAC Trends in Analytical Chemistry*, Volume 27, Issue 3, 228-237, DOI 10.1016/j.trac.2007.12.001
- [4] Rohart F, Gautier B, Singh A, Lê Cao KA (2017) mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology* 13(11): e1005752. DOI 10.1371/journal.pcbi.1005752
- [5] Thévenot E A. (2016) roppls: PCA, PLS (-DA) and OPLS (-DA) for multivariate analysis and feature selection of omics data[J]. R package version, 1(0).
- [6] Teresa W.-M. Fan (1995). Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. *Progress in Nuclear Magnetic Resonance Spectroscopy*, Volume 28, 161-219. DOI 10.3945/jn.110.135657
- [7] Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, et al. (2013) HMDB 3.0 — The Human Metabolome Database in 2013. *Nucleic Acids Res.* 2013. Jan 1;41(D1):D801-7 DOI 10.1093/nar/gks1065
- [8] Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, et al., HMDB 4.0 — The Human Metabolome Database for 2018. *Nucleic Acids Res.* 2018. Jan 4;46(D1):D608-17. 29140435

10 LISTE DES TABLEAUX ET FIGURES

Tableau 1 : Principales caractéristiques des différents spectromètres RMN à haut champ magnétique présents sur la plateforme.....	6
Tableau 2 : Comparaisons des avantages et inconvénients de la MS et de la RMN	10
Tableau 3 : Description des caractéristiques des différentes farines	13
Tableau 4 : Regroupement des échantillons pour les expériences de spectres 2D	15
Tableau 5 : Tableau d'identification des métabolites (cellules vertes représentent les métabolites identifiés avec confiance, cellules oranges représentent les suspicions des métabolites)	31
Figure 1 : Logo de l'unité QuaPA, INRAE	4
Figure 2 : Digesteur artificiel DIDGI®	13
Figure 3 : Représentation schématique du plan d'expériences.....	14
Figure 4 : Spectres RMN ayant servi à l'analyse métabolomique (A) et celui enregistré lors de l'enregistrement des spectres 2D (B). Les flèches rouges illustrent les décalages pouvant exister.....	17
Figure 5 : Exemple de spectres intestinal (rouge) et gastrique (bleu). Les deux spectres ont été obtenus 60 minutes après le début de la digestion. L'intensité du spectre gastrique a été multiplié par 8 pour avoir des signaux comparables à ceux obtenus dans le compartiment intestinal.....	18
Figure 6 : Comparaison des spectres lors d'une digestion contenant du pain (bleus) ou pas (rouge). Les spectres de la partie gastrique sont représentés à gauche et ceux de la partie intestinale à droite.	19
Figure 7 : Superposition de quelques spectres gastriques. La couleur des spectres dépend du temps de digestion : 5 min (vert), 30 min (rouge), 60 min (violet) et 120 min (bleu).....	19
Figure 8 : Exemple d'alignements des spectres et de bucketing sur les spectres de la partie gastrique. Avant (haut) et après (bas) traitements.	20
Figure 9 : Illustration d'une partie du tableau de données (aussi appelé bucket table) obtenu après tous les traitements réalisés sous NMRProcFlow.	22
Figure 10 : PCA pour les données gastriques représentée en fonction du temps de digestion. La variance totale expliquée par les 2 premiers axes (PC1 et PC2) est de 86%.....	23
Figure 11 : PCA pour les données gastriques avec un marquage des points en fonction des farines.....	23
Figure 12 : PCA pour les données gastriques avec un marquage des points en fonction des méthodes de fermentation	23

Figure 13 : PLS-DA pour les données gastriques en fonction du temps	24
Figure 14 : PLS-DA pour les données gastriques avec une séparation recherchée par les méthodes de fermentation	25
Figure 15 : PLS-DA pour les données gastriques en cherchant à séparer les types de farines.	25
Figure 16 : PCA pour les données gastriques G30 & G60 représentée en fonction de la méthode de fermentation.....	26
Figure 17 : PLS-DA pour les données gastriques G30 & G60 en fonction des méthodes de fermentation	26
Figure 18 : Modèle OPLS-DA des données gastriques G30 & G60.....	27
Figure 19 : OPLS-DA modèle pour les données gastriques G30	27
Figure 20 : OPLS-DA modèle pour les données gastriques G5	28
Figure 21 : OPLS-DA modèle pour les données gastriques G60	28
Figure 22 : Exemple d'identification avec les spectres COSY et HMQC.	29
Figure 23 : Exemple de superposition du spectre du pool avec le spectre de glucose.	30
Figure 24 : PCA pour les données intestinales représentée en fonction du temps.....	32
Figure 25 : PCA pour les données intestinales représentée en fonction des farines.	33
Figure 26 : PCA pour les données intestinales représentée pour les différentes méthodes de fermentation.	33
Figure 27 : PLS-DA pour les données intestinales avec un modèle basé sur les farines.....	33
Figure 28 : PLS-DA pour les données intestinales avec un modèle basé sur les méthodes de fermentation.	33
Figure 29 : PCA Plot pour les données intestinales en fonction des répétitions	34
Figure 30 : PCA pour les données intestinales (sans contrôle) en fonction du temps	34
Figure 31 : PLS-DA pour les données intestinales (sans contrôle) en fonction des méthodes de fermentation	35
Figure 32 : PLS-DA pour les données intestinales (sans contrôle) en fonction des farines.....	35
Figure 33 : PLS-DA Plot pour les données intestinales I120 & I180 & I240 (sans contrôle) en fonction des méthodes de la fermentation	36
Figure 34 : PLS-DA pour les données intestinales avec levures boulangers (sans contrôle) en fonction des farines.....	36
Figure 35 : PLS-DA pour les données intestinales avec levain (sans contrôle) en fonction des farines.....	37