THÈSE

Pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITÉ DE POITIERS
UFR des sciences fondamentales et appliquées
Institut de chimie des milieux et matériaux de Poitiers - IC2MP
(Diplôme National - Arrêté du 25 mai 2016)

École doctorale : Rosalind Franklin, Poitiers
Secteur de recherche : Chimie théorique, physique, analytique

Présentée par :
Fabiola Djago

Développement de stratégies pour la détection de marqueurs protéiques du cancer

Directeur(s) de Thèse :
Pauline Poinot, Claude Geffroy, Sébastien Papot

Soutenue le 28 septembre 2023 devant le jury

Jury :

Président	**Raphaël Delépée**	Professeur, Anticipe, Université de Caen Normandie
Rapporteur	**Christine Varon**	Professeure, BRIC, Université de Bordeaux
Rapporteur	**François Fenaille**	Directeur de recherches CEA, Université Paris-Saclay

Pour citer cette thèse :
Fabiola Djago. *Développement de stratégies pour la détection de marqueurs protéiques du cancer* [En ligne]. Thèse Chimie théorique, physique, analytique. Poitiers : Université de Poitiers, 2023. Disponible sur Internet <http://theses.univ-poitiers.fr>

THÈSE

Pour l'obtention du Grade de

DOCTEUR DE L'UNIVERSITE DE POITIERS

(Faculté des Sciences Fondamentales et Appliquées)
(Diplôme National - Arrêté du 25 mai 2016)

École Doctorale 649 :
« *Rosalind Franklin* » - Énergie – Environnement - Biosanté

Secteur de Recherche :
Chimie théorique, physique, analytique

Présentée par :

Fabiola DJAGO

****

Développement de stratégies pour la détection de marqueurs protéiques du cancer

****

Directeur de Thèse :

Pauline POINOT en codirection avec Claude GEFFROY
et Sébastien PAPOT

Soutenue le jeudi 28 septembre 2023

Devant la Commission d'Examen

****

JURY

Christine Varon	Professeure, Université de Bordeaux	Rapportrice
François Fenaille	Directeur de recherche CEA, Université Paris-Saclay	Rapporteur
Raphaël Délépée	Professeur, Université de Caen Basse-Normandie	Examinateur
Pauline Poinot	Maître de conférences, Université de Poitiers	Directrice de thèse
Claude Geffroy	Professeure, Université de Poitiers	Co-Directrice de thèse
Sébastien Papot	Professeur, Université de Poitiers	Co-Directeur de thèse

Abréviations

AcI: Anticorps primaire
AcII: Anticorps secondaire
ADN: Acide désoxyribonucléique
AF: Acide formique
Au: Atome d'or
AIRC: Agence international de la recherche sur le cancer
AP: Phosphatase alcaline
BSA: Albumine de Sérum Bovin
Ca: Atome de calcium
CAR/PDMS: Carboxen/polydimethylsiloxane
CHU: Centre Hospitalier Universitaire
CIC: Centre d'investigation clinique
CID: Dissociation induite par collisions
COV: Composé organique volatile
CT: Chimiothérapie
C-ter: C-terminal
Dh: Déhydroalanine
DO: Densité optique
DTT: Dithiothréitol
EI: Impact électronique
ETOH: Éthanol
ERGIC: Compartiment intermédiaure ER-Golgi
F: Atome de fluor
FASTA: FAST-All
Ga: Atome de gallium
GC: Chromatographie en phase gazeuse
GC-MS/MS : Chromatographie en phase gazeuse liée à la spectrométrie de masse en tandem
GdnCl: Chlorure de guanidine
GERL: Lysosome du réticulum endoplasmique et du golgi
ESI: Ionisation électrospray
ELISA: Dosage d'immunoabsorption enzymatique
H_2O_2: peroxide d'hydrogène
HIV: Virus de l'immunodéficience humaine
HRMS: Spectrométrie de masse à haute résolution
HRP: Peroxydase de raifort
I: Atome d'iode
IAA: Iodoacétamide
In: Atome d'indium
INC: Institut national du cancer
IRM: Imagerie par résonance magnétique
ITZ: Iminothiazolidine

K: Atome de potassium
KB: Cellules du carcinome épithélial humain
LLC: Carcinome pulmonaire de Lewis
LC: Chromatographie liquide
LC-MS/MS : Chromatographie liquide liée à la spectrométrie de masse en tandem
LOD: Limite de détection
MEC: Matrice extracellulaire
Mg: Atome de magnésium
MRM : Monitorage de réactions multiples
Ni: Atome de Nickel
NTCB: acide 2-nitro-5-thiocyanobenzoïque
N-ter: N-terminal
OEPM: Micro réseau de protéine amélioré optiquement
OMS: Organisation Mondiale de la santé
PBS: Tampon phosphate salin).
PCR: Réaction de polymérisation en chaîne
PDA: Détecteur à barrette de diode
PEA: Essai d'extension de proximité
p-ELISA: ELISA plasmonique
PLA: Essai de ligature de proximité
PSA: Antigène spécifique de la prostate
RE: Réticulum endoplasmique
RP-LC: Chromatographie en phase inverse
RPLS: Résonance plasmonique de surface localisée
RT: Radiothérapie
SEC: Chromatographie d'exclusion stérique
SDS: Sodécylsulfate de Sodium
SDS-PAGE : Électrophorèse sur gel de dodécyl-sulfate de sodium et de polyacrylamide
SPME : Microextraction sur phase solide
SRM: Monitorage des réactions sélectionnées
TCA: Acide trichloroacétique
TCEP: tris(2-carboxyéthyl)phosphine
TEMED: Tétraméthyléthylènediamine
TEP-SCAN: tomographie par émission de positons
TIC: Chromatogramme total d'ion
TMB: Tétramethylbenzidine
TMS: Théorie de Mutation Somatique
TNB : ion thionitrobenzoate
Uniprot : Ressource universelle des protéines

Remerciements

Je tiens personnellement à remercier les membres de mon jury de thèse. En premier lieu, je remercie Mme Christine Varon et M. François Fenaille d'avoir accepté d'évaluer ce travail en tant que rapporteurs. C'est avec beaucoup de reconnaissance que je remercie également M. Raphaël Delepee, qui évaluera ma thèse en qualité d'examinateur.

Je souhaite remercier l'IC2MP dirigé par M. Yannick Pouilloux de m'avoir accueillie dans son institut, dans lequel j'ai pu évoluer et acquérir de nombreuses compétences.

J'aimerais remercier mes co-directeurs de thèse de m'avoir accompagnée tout le long de mon parcours. Je remercie Pauline Poinot pour sa joie de vivre et détermination à toute épreuve. J'ai appris avec Pauline Poinot qu'il ne fallait jamais abandonner, ce qui nous a valu une publication dans le grand journal Nature. Je remercie également Claude Geffroy de m'avoir guidée dans les expérimentations qui nécessitaient de fortes connaissances en chimie et en SEC. Je salue également sa positivité. J'adresse mes remerciements à Sébastien Papot d'avoir fait grandir ce projet en créant différentes sondes et molécules thérapeutiques. Je suis obligée d'ajouter que j'ai beaucoup apprécié ta bonne humeur et une phrase que tu as dit « La vie est tellement belle qu'elle vaut ma peine d'être vécue ».

J'envoie mes sincères remerciements à Mehrad Tarighi, qui m'a accompagnée dans l'utilisation des appareils GC-HRMS et GC-MS/MS dont il est responsable. En plus de son aide, j'ai apprécié son amabilité. Merci également à Quentin Blancart-Remaury de la PLATINA, d'avoir contribué aux analyses LC-HRMS et LC-MS/MS et m'avoir formée sur les différents appareils, aussi bien sur la maintenance que sur l'utilisation. En plus d'être un collaborateur, je peux dire que tu es devenu un ami.

Dans ce projet multidisciplinaire, je voudrais remercier tous mes collaborateurs. Merci au chimiste Rémi Châtre d'avoir synthétisé la majorité des sondes que j'ai utilisées dans ma thèse. Je remercie également les collaborateurs du département de biologie Isabelle Petit Paris et Jonathan Clarhault, de m'avoir fournie les biospies et tumeurs, mais aussi permis d'obtenir mon habilitation à pratiquer l'expérimentation animale sur modèle murin.

Au sein de mon équipe qui est l'E-BiCOM, plus précisément à mon étage, j'ai pu sympathiser avec Laurent Grasset dont j'appréciais les blagues le matin, c'était mémorable. Je le remercie aussi pour ses conseils avisés en chimie et biochimie, pour la façon dont il insistait pour que je sois autonome, ce qui est chose faite. Merci Laurent pour le joli livre que tu m'as prêté : « Biochimie structurale » de Claude Audigié et François Zonszain. Je remercie Jean François Chollet, Cécile Marivingt-Mounir et Émilie Dugaro pour leur chaleurosité et la bonne entente qu'il y eu.

Mes nombreux allers-retours à l'ENSIP pour utiliser différents appareillages m'entrainent à remercier Jérôme Labanowski, Leslie Mondamert, Audrey Allavena, Bertrand Gombert.

C'est avec beaucoup de sympathie que je remercie mes collègues étudiants plus particulièrement Justin Lange qui m'a initié et formé pour ce beau projet. Je remercie également Ramzi Timoumi, Estelle Blochouse, Rihab Fkiri, Rony Eid, Ruth Oussou avec qui j'ai pu avoir de beaux échanges.

Je tiens à remercier mes parents tous droit venus du village de Niagbaméko (Côte d'ivoire) qui ont tout sacrifié, 30 ans en arrière, pour que mes sœurs et moi accédions à un haut niveau d'étude. Leur souhait s'est réalisé. Merci encore à mes parents Jean-Patrice et Natacha, mais aussi mes sœurs Grace, Karelle et mon beau-frère Didier, de m'avoir soutenue et conseillé tout au long de cette thèse.

Je remercie aussi celui qui partage ma vie, Jordan, qui a été un réel soutien moral le long de cette thèse.
Merci aussi à mes amis du groupe The more Kongossa Factory, qui m'ont soutenue et m'ont permise de lâcher prise et rigoler un bon coup quand j'en avais besoin.

Enfin, je dédie cette thèse à ma nouvelle nièce Kaïla, tu es arrivée dans ma vie comme un souffle nouveau, une réelle bénédiction. J'espère qu'un jour tu pourras lire cette thèse et voir ce que Tata a fait.

Table des matières

Abréviations 3
Remerciements 5
Introduction générale 15
Partie A. Bibliographie 21
I. Généralités sur le cancer 23
I.1. Définition et chiffres 23
I.2. Les principaux cancers 24
I.2.A Cancer du sein 24
I.2.B. Cancer du poumon 26
I.2.C. Cancer colorectal 27
I.3. Biologie du cancer 29
I.4. Diagnostic et traitement des cancers solides 31
I.4.A. Diagnostic de cancer 31
I.4.B. Traitement des cancers 33
I.4.C. Objectif mondial : établir un diagnostic précoce 38
I.5. Biomarqueurs 38
I.6. Les glycosidases 39
II. Principes de biochimie: Techniques de séparation, de fractionnement et de détection des protéines 45
II.1.SDS-PAGE 45
II.2. Coloration au bleu de Coomassie 46
II.3. Coloration au nitrate d'argent 48
II.4. Electrophorèse 2D 48
II.5. Electrophorèse capillaire 50
II.6. Western blot 50
II.7. Test ELISA 51
II.8. Nouvelles techniques de détection des protéines 53
Méthode ELISA plasmonique 53
Tests de proximité médiés par l'ADN pour l'analyse des protéines ultrasensibles 54
Techniques de miniaturisation 56
Utilisation de microparticules comme surface de réaction 57

III. L'utilisation des sciences multi-omiques pour la cartérisation et détection des marqueurs protéiques. 58
III.1. Analyses « omiques » 58
III.1.A. Protéomique 58
III.1.B. Analyse des protéines et des peptides par spectrométrie de masse 61
Technologie orbitrap 61
QTOF (Time of Flight) 62
Triple Quadripôle (TQ) 63
III.2. Volatolomique induite 64
Objectifs et stratégie 79
Partie B. Matériel et Méthodes 83
Objectif expérimental 85
I. Matériel chimique, biochimique, et biologique 85
Matériel chimique: 85
Matériel biochimique et biologique: 85
II. Détection des glycosidases par volatolomique induite 88
II.1. Recherche de la limite de détection des éthanols marqués : cas de l'éthanol D_5 88
II.2. Étude de la détection des glycosidases par les sondes à COVs sur matrice plasma : cas de la β-Glu 89
II.3. Étude de la détection des glycosidases par une méthode de référence : cas de la β-Glu 89
II.4. Optimisation des conditions de manipulation des prélèvements sanguins : préservation de la stabilité de la β-Glu 90
Effet de l'inhibiteur de protéases 90
Effet de la congélation/décongélation répétées 90
Effet de la durée de congélation 90
II.5. Détection des glycosidases dans les prélèvements de plasma via les sondes à COVs : cas de la β-Glu 90
II.6. Détection des glycosidases sur biopsies de tumeurs et foies via les sondes à COVs : cas de la β-Glu 91
II .7. Développement d'un cocktail de sondes : étude de la spécificité de chaque sonde et leur cible enzymatique 92
II.8. Détection des cibles enzymatiques en utilisant un cocktail de sondes *in vivo* 94
II.9. Paramètres GC-MS/MS 94
III. Extraction, isolation, et détection des glycosidases dans les prélèvements biologiques 95

III.1. Détection de la β-Glu par protéomique *middle-down* et *bottom-up* 95
III.1.A. Recherche des peptides (*in silico*) 95
III.1.B. Protocole digestion trypsique 96
III.1.C. Protocole d'hydrolyse NTCB avec réduction au DTT en atmosphère inerte et clivage à la soude (Crimmins et. al, Srzentiv et.al) 96
III.1.D. Protocole hydrolyse NTCB avec réduction au TCEP(tris(2-carboxyéthyl)phosphine) et clivage au NH4OH en pression atmosphérique (Belghazi et al) 97
III.1.E. Protocole digestion au NTCB avec clivage à la glycine dans un seul mélange réactionnel (Tang et. al) 97
III.2. Optimisation de la séparation des peptides et protéines par chromatographie SEC (exclusion stérique) 97
III.2.A. Séparation des protéines entières 97
III.2.B. Séparation de l'albumine et des protéines commerciales par SEC en conditions dénaturantes analysée par la méthode 1 98
III.3. Analyse des protéines hydrolysées par NTCB 98
III.4. Paramètres LC-UV 98
Méthode 1 98
Méthode 2 99
III.5. Paramètres UPLC-HRMS 99
Partie C. Résultats et discussion 101
I. Volatolomique induite 116
I.1 Détection de la β-Glu suivant un protocole de référence basé sur une sonde de paranitrophénol 116
I.2. Gamme de détection de l'EtOH-D_5 et optimisation de ses transitions 119
I.2.1. Limite de détection de l'EtOH-D_5 119
I.2.2. Limite de détection de la β-Glu en tampon avec la sonde EtGlu-D5 120
I.3. Optimisation des conditions de détection de la β-Glu sur biopsies 121
I.4. Optimisation des conditions de détection de la β-Glu dans le plasma 122
I.5. Étude préliminaire clinique 124
I.5.1. Optimisation des transitions des isotopes de l'éthanol 125
I.5.2. Limite de détection de l'ETOH-$^{13}CD_5$ 126
I.5.3. Étude du croisement des sondes spécifiques de la et β-Glu et la β-GlcNac. 126
I.5.4. Développement d'un cocktail de sondes 128
II. Détection de la β-Glu par protéomique *middle-down* et *bottom-up* 133
II.1. Stratégie expérimentale 133

II.2. Exploitation des résultats d'hydrolyse au NTCB : comparaison des différents protocoles 137
Protocole de digestion trypsique 137
Protocole d'hydrolyse NTCB avec réduction au DTT en atmosphère inerte et clivage à la soude 137
Protocole hydrolyse NTCB avec réduction au TCEP et clivage au NH_4OH SOUS pression atmosphérique 139
Protocole digestion au NTCB avec clivage à la glycine dans un seul mélange réactionnel 142
Conclusion 144
III. Optimisation de la séparation des peptides et protéines par chromatographie SEC 144
III.1. Séparation des protéines entières 144
Stratégie expérimentale 144
III.2. Séparation par SEC de l'albumine et des protéines commerciales par SEC en conditions dénaturantes analysée par la méthode 2 150
Dénaturation des protéines 150
III.3. Analyse des protéines hydrolysées par NTCB 151
Séparation des digestats NTCB par SEC analysée par la méthode 2 151
Conclusion 153
Conclusion générale 185
Références bibliographiques 191
Annexes 201
Résumés 211

Liste des figures

Figure 1: Taux des cancers dans le monde suivant la base de données GLOBOCAN sur 185 pays et 36 cancers pour les deux genres (Source image : Sung et. al. 2021) [10] 23
Figure 2: Mortalité liée aux cancers dans le monde suivant la base de données de GLOBOCAN 185 pays et 36 cancers pour les deux genres (Source image : Sung et. al. 2021) [10] 23
Figure 3: Taux (âges standardisés) et mortalité due au cancer du sein en 2020 en fonction des régions (Source image : Sung et. al. 2021) [10] 25
Figure 4: Mortalité du cancer du sein dans le monde la base de données GLOBOCAN 26
Figure 5: Taux (âges standardisés) du cancer du poumon en 2020 en fonction des régions et du genre (Source image : Sung et. al. 2021) [10] 27
Figure 6: Taux (âges standardisés) du cancer du côlon (A) et du rectum (B) en 2020 en fonction des régions et du genre (Source image : Sung et. al. 2021) [10] 28
Figure 7: Schéma des évènements biologiques et moléculaires lors de la transformation d'une cellule saine à cancéreuse (Source image : Anand et. al. 2022) [19] 30
Figure 8: Tumeur du poumon par radiographie (cercle rouge).(Source image : https://www.santeweb.ch) [38] 32
Figure 9: Clichés de métastases observées par scintigraphie (à gauche) et par TEP-scan (à droite). À gauche, chaque point noir correspond à une métastase. À droite, les métastases sont indiquées par les flèches et correspondent aux zones colorées. 33
Figure 10: Réservoirs des biomarqueurs dans l'organisme 39
Figure 11: Schéma des monomères des N-glycanes (Source image: Yang C. et. al. 2019) [48] 40
Figure 12: Schéma de la diversité des glycanes (Source image : Costa et. al 2020) [78] 40
Figure 13: Composition des glycanes et sucres ciblés pour l'étude (Source image : Ohstubo et Marth, 2006) [67] 41
Figure 14: Rôles des glycanes chez un sujet sain [155–157] 41
Figure 15: Schéma de la voie de sécrétion 43
Figure 16: Schéma du principe de gel SDS-PAGE (Source image. https://www.bio-rad.com/fr-fr) [96] 45
Figure 17: Principe du gel SDS-PAGE 46
Figure 18: Formule chimique du Bleu de Coomassie 47
Figure 19: Mécanisme du complexe du Bleu de Coomassie avec la protéine (Source image : Georgiouet. al. 2008) [119] 47
Figure 20: Schéma de l'électrophorèse 2D (Source image : www.creative-proteomics.com)[110] 49
Figure 21: Schéma de la technique de Western blot (Source image : https://slideplayer.fr/slide/18120366/) [136] 50
Figure 22: Schéma explicatif de la méthode ELISA directe (Source image : www.creative-diagnostics.com) [120] 52
Figure 23: Schéma explicatif de la méthode ELISA indirecte (Source image : www.creative-diagnostics.com) [120] 52
Figure 24: Schéma du principe de la méthode ELISA plasmonique 54

Figure 25: Schéma réactionnel du peroxyde d'hydrogène et la catalase (Source image : Yang et. al 2016)[125] 54
Figure 26: Schéma du principe de la PCR (Source image : https://ib.bioninja.com.au) [135] 55
Figure 27: Schéma du principe de la PLA (Source image : Yang et. al 2016)[125] 55
Figure 28: Schéma de la technique de miniaturisation OEPM (Source image : Zhang et. al 2021)[140] 57
Figure 29: Schéma de la technique multiplexe utilisant des microparticules d'hydrogel fonctionnalisées 58
Figure 30: Schéma du concept de la protéomique (Source image : Kennani et al 2018)[144] 59
Figure 31: Schéma réactionnel de l'hydrolyse au NTCB 60
Figure 32: Représentation du Q-Exactive (Thermo Fisher®) (Source image : https://www.smbp.espci.fr) [153] 62
Figure 33: Représentation de fonctionnement du QTOF (Source image : Mesmin C, 2012) [159] 63
Figure 34: Représentation du fonctionnement du triple quadripôle (Source image: Competency F, 2017) [161] 63
Figure 35: Schéma expérimental de l'analyse des COVs par SPME GC/MS-MS ... 89
Figure 36: Schéma expérimental de la détection de la cible sur biopsies 92
Figure 37: Schéma expérimental du croisement de sondes 93
Figure 38: Réaction d'hydrolyse du glucuronide de paranitrophénol (PNPG) en paranitrophénol (PNP) par la β-Glu 116
Figure 39: Contrôle négatif de la consommation de PNPG sur matrice plasma sans enzyme 117
Figure 40: Contrôle négatif de la libération de PNP sur matrice plasma sans enzyme 117
Figure 41: Suivi de la consommation de PNPG sur matrice plasma avec enzyme 118
Figure 42: Suivi de la libération de PNP en matrice plasma avec enzyme 118
Figure 43: Gamme de détection UV de la β-Glu avec le PNP 119
Figure 44: Gamme de détection de l'ETOH-D_5 par SPME-GC-MS/MS 120
Figure 45: Gamme de détection de la β-Glu *in vitro* 120
Figure 46: Gamme de détection de la β-Glu en matrice plasma 121
Figure 47: Détection de la β-Glu sur les biopsies. 122
Figure 48: Étude de l'effet des protéases sur l'activité enzymatique 123
Figure 49: Étude de l'effet des cycles congélation/décongélation du plasma sur l'activité enzymatique 123
Figure 50: Étude de l'impact de la durée de congélation du plasma sur l'activité enzymatique 124
Figure 51: Résultats de l'étude clinique sur les plasmas du CHU 125
Figure 52: Gamme de détection de l'EtOH-$^{13}CD_5$ 126
Figure 53: Résultats du croisement de sondes 127
Figure 54: Résultats du croisement de sondes 127
Figure 55: Résultats de croisement de sondes 128
Figure 56: Schéma expérimental de l'utilisation du cocktail in vivo (Source image : R. Châtre et al)[7] 130

Figure 57: Détermination de la concentration de sondes EtGlu-D_5 à injecter in vivo 130
Figure 58: Détermination de la concentration de sondes EtGlcNac-$^{13}CD_5$ à injecter in vivo 131
Figure 59: Détermination de la concentration de sondes EtGal-D_2 à injecter in vivo 131
Figure 60: Détermination de la concentration de sondes EtFuc-D_4 à injecter in vivo 132
Figure 61: Quantité d'éthanols marqués exhalée par des souris saines (n=6) et des souris xénogreffées (n=6) après injection du cocktail de sondes 132
Figure 62: Schéma de l'hydrolyse NTCB de l'albumine et des glycosidases. 134
Figure 63: Listes des peptides trouvés de l'hydrolyse de la BSA au NTCB 136
Figure 64: Profil d'un peptide identifié par la plateforme mMass 136
Figure 65: Résultats de digestion de la BSA avec la trypsine 137
Figure 66: Schéma des étapes du protocole du NTCB par Szrentic et. al 138
Figure 67: Schéma du montage de réaction en condition inerte avec un dispositif en cloche 138
Figure 68: Schéma du montage de réaction en condition inerte avec tube scellé 139
Figure 69: Schéma du mécanisme de compétition entre la protéine et le DTT 139
Figure 70: Schéma des étapes du protocole NTCB de Koehn et. al 140
Figure 71: Schéma des étapes du protocole NTCB de Belghazi et. al 140
Figure 72: Résultats de l'hydrolyse NCTB de la BSA avec réduction au TCEP et clivage à l'ammoniaque 141
Figure 73: Schéma des étapes du protocole NTCB de Tang et. al 142
Figure 74: Chromatogrammes de la séparation des glycosidases par SEC-UV à 280 nm en conditions natives 145
Figure 75: Séparation des petites protéines sur SEC-UV à 280 nm en conditions natives 146
Figure 76: Élution du DTT et IAA sur SEC-UV à 280 nm 146
Figure 77: Chromatogrammes des glycosidases en conditions dénaturantes par SEC-UV à 280 nm 147
Figure 78: Chromatogrammes de la gamme détection de la BSA à 280 nm 148
Figure 79: Gamme de détection de la BSA à 280 nm 148
Figure 80: Chromatogrammes de la gamme de détection de la BSA à 220 nm 148
Figure 81: Gamme d'étalonnage de la BSA à 220 nm 149
Figure 82: Zoom du chromatogramme des protéines en conditions natives par SEC-UV avec la Reprosil 200 149
Figure 83:Chromatogramme des protéines en conditions natives par SEC-UV avec la Reprosil 200 150
Figure 84: Chromatogrammes des protéines en conditions dénaturantes par SEC-UV méthode 2 151
Figure 85: Chromatogramme d'analyse des peptides issu de l'hydrolyse de la BSA au NTCB par LC-Orbitrap 151
Figure 86: Chromatogramme des peptides obtenus via le NCTB par SEC-UV méthode 2 152

Figure 87: Zoom 1 du chromatogramme des peptides obtenus via le NCTB par SEC-UV méthode 2 152
Figure 88: Zoom 2 du chromatogramme des peptides obtenus via SEC-UV méthode 2 153
Annexe 89: Fichier FASTA de la séquence d'albumine humaine 203
Annexe 90:Liste des différents types de logiciels utilisés en protéomique1 204
Annexe 91: Présentation de la plateforme du logiciel de protéomique massXpert .. 205
Annexe 92: Présentation du module de conversion de fichiers brutes MSConvert ... 206
Annexe 93: Choix des paramètres de masse sur la plateforme mMass 207
Annexe 94: Importation de la séquence FASTA sur la plateforme mMass 207
Annexe 95: Sélection du type de digestion sur la plateforme mMass 208
Annexe 96: Ajout de l'hydrolyse NTCB parmi les agents de digestion sur la plateforme mMass 208
Annexe 97: Sélection des modifications sur la plateforme mMass 209
Annexe 98: Choix des modifications après hydrolyse NTCB sur la plateforme mMass 209
Annexe 99: Obtention de la liste des peptides théoriques de digestion avec la plateforme mMass 210

Introduction générale

Introduction générale

Le vivant résulte d'un ensemble de processus systémiques se produisant à différentes échelles (moléculaire, cellulaire, tissulaire et organisme). Son étude approfondie a notamment été rendue possible grâce à l'émergence de nouvelles disciplines scientifiques telles que les sciences omiques. Ces approches s'appuient sur des technologies de pointe pour analyser, détecter et explorer les biomolécules intervenant dans la structure et le fonctionnement des systèmes biologiques.

Parmi les technologies employées pour identifier ces métabolites, la spectrométrie de masse représente un outil très intéressant puisqu'elle permet d'identifier un grand nombre de biomarqueurs (métabolites, composés organiques volatils, protéines, ARN) avec un poids moléculaire varié. Depuis maintenant quelques dizaines d'années, des analyseurs de haute résolution ont été développés, permettant d'accéder à la masse exacte des composés, assurant ainsi leur identification. D'autres analyseurs, plus sélectifs, comme le triple quadripôle, ont également été développés dans l'objectif de quantifier les composés.

Bien que ces techniques soient très performantes, l'étude *in vivo* comme *in vitro*, des biomarqueurs fait encore aujourd'hui face à de nombreux verrous liés notamment à la complexité des systèmes analysés. De ce fait, et notamment dans le cadre de travaux expérimentaux *in vitro*, la spectrométrie de masse est souvent placée en aval d'une méthode de séparation telle que la chromatographie, améliorant la sensibilité de l'analyse. Ces méthodes ont permis le screening de nombreuses biomolécules, et ont ainsi pu mettre en évidence différentes signatures moléculaires caractéristiques d'un processus ou phénotype particulier.

Outre les avancées en termes de technologies d'analyse des biomolécules, l'étape de prétraitement des échantillons reste cruciale afin d'améliorer le signal de détection en spectrométrie de masse de ces composés qui sont très souvent en faible quantité.

Ainsi, l'étude moléculaire des systèmes biologiques fait encore aujourd'hui face à de nombreux verrous, liés notamment à la complexité de la matrice biologique (cellule, tissu, organe) voire de l'organisme dans son ensemble. Un des principaux enjeux scientifiques de cette recherche réside dans le développement d'un processus analytique sensible, sélectif et reproductible.

Dans ce contexte, mon travail de thèse à consister à développer de nouvelles approches impliquant des stratégies et outils issus des sciences omiques afin d'explorer au niveau moléculaire des processus biologiques liés au développement de maladies, ou impliqués dans leur régression. Dans ce cadre, nous nous sommes tout particulièrement intéressés à développer des outils moléculaires et/ou stratégies analytiques permettant l'exploration de processus caractéristiques des cancers solides.

La première partie de ce travail consiste en une synthèse bibliographique portant sur des généralités sur le cancer, les chiffres clés et les principaux processus biologiques impliqués. Cette partie sera ensuite suivie d'une description de différents marqueurs du cancer, en appuyant notamment sur les marqueurs enzymatiques ciblés dans le cadre de ce travail, les glycosidases. La description des différentes stratégies de biochimie permettant la détection de biomarqueurs, les principales méthodes et outils d'analyse MS utilisés en sciences omiques seront également présentés.

A l'issue de cette étude bibliographique, les objectifs et la stratégie expérimentale seront exposés.

Dans une première partie, nous présenterons le développement d'un nouveau concept dérivé de la volatolomique, appelée volatolomique induite, qui permet la détection de marqueurs enzymatiques après activation de sondes off/on. Après métabolisation, ces sondes libèrent un composé organique volatil (COV) exogène dont la fonction sera « d'allumer » le phénotype recherché. Ces COVs sont alors utilisés comme traceurs d'une activité métabolique en temps réels puisqu'ils se veulent facilement analysables. Présentés comme des outils moléculaires permettant l'exploration biologique, ces sondes « chémobiologiques » pourraient élargir considérablement l'état de nos connaissances sur le fonctionnement des systèmes biologiques, ainsi que des processus liés à la progression ou à la régression des cancers solides.

Afin de comprendre l'origine biologique de ces marqueurs enzymatiques, nous avons en parallèle travaillé sur une approche de protéomique « middle-down », afin d'extraire ces glycosidases de la matrice biologique, puis de les séparer des protéines majoritaires. L'idée étant ici d'améliorer leur détection par UPLC-HRMS. Ainsi, chaque étape de cette analyse protéomique « middle-down » a été décortiquée et optimisée, afin de purifier au mieux la matrice et ainsi augmenter la sensibilité de détection par UPLC-HRMS ou UPLC-UV des enzymes d'intérêt. Cette partie de résultats se trouvera détaillée en fin de document.

Dans un second temps, nous avons travaillé sur l'optimisation de protocoles analytiques permettant l'étude de nouvelles classes de composés à visée thérapeutique, et modulant des processus biologiques ciblés.

Dans ce cadre, nous avons mis au point une stratégie d'extraction, purification et analyse d'un agent cytotoxique (la doxorubicine) transporté sélectivement au niveau des tissus tumoraux via un système moléculaire dit « programmé » pour le ciblage et la libération d'anticancéreux. Ce travail, réalisé en collaboration avec l'équipe OrgaSynth de l'IC2MP, avait pour objectif de démontrer la présence et l'accumulation de l'agent cytotoxique au niveau de l'ADN des cellules tumorales. Les résultats obtenus ont alors été confrontés aux données in vivo mesurant l'efficacité thérapeutique du traitement ici développé.

Enfin, et en collaboration avec des chercheurs du Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques de l'université Paris Descartes, une stratégie d'analyse d'une classe émergente d'agents thérapeutiques basés sur des petits oligoribonucléotides (conjugués acides aminés-ARN) a été développée.

La conception et la mise au point de l'ensemble de ces approches impliquant la spectrométrie de masse, devrait ainsi permettre à la fois de caractériser, d'appréhender la complexité du vivant au niveau moléculaire et d'évaluer l'état de réponse d'un organisme lorsque soumis à un stress donné.

Partie A. Bibliographie

I. Généralités sur le cancer

I.1. Définition et chiffres

Avec 19,3 millions de nouveaux cas et 10 millions de décès en 2020 selon le GLOBOCAN 2020[9], le cancer est l'une des maladies les plus mortelles au niveau mondial. Néanmoins la répartition des cas de cancers dans le monde est hétérogène (Figure 1 et Figure 2). Il y a une forte prédominance des décès liés au cancer en Asie (58.3 %). En Europe, 19.6 % de décès sont liés au cancer, et aux États Unis, 14.2 %. Enfin, l'Afrique compte 7.2 % de morts liés au cancer.

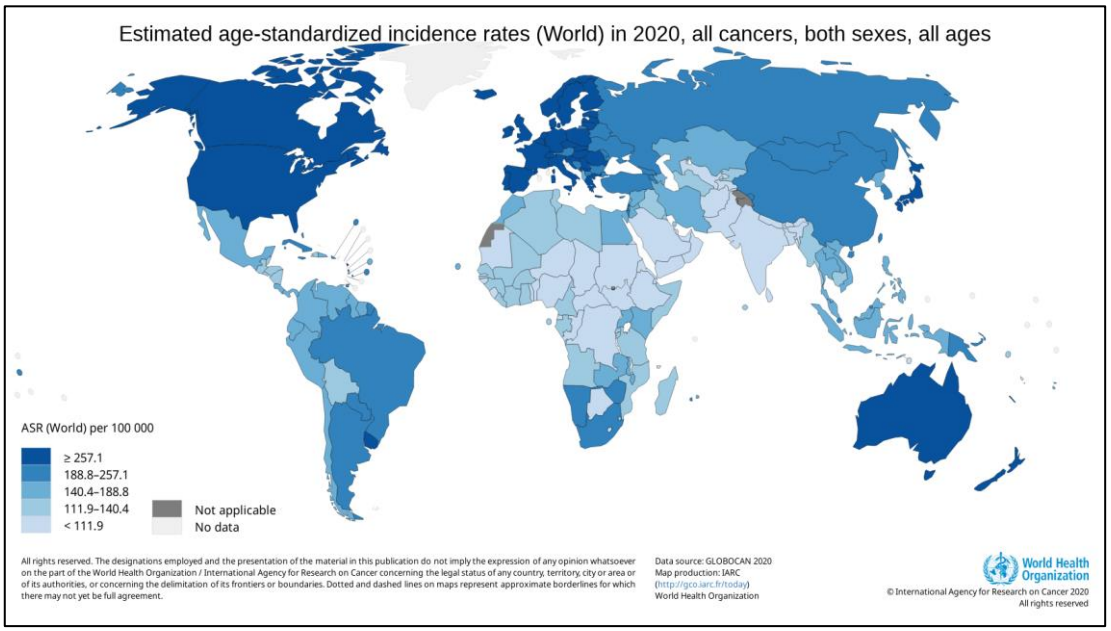


Figure 1: Taux des cancers dans le monde suivant la base de données GLOBOCAN sur 185 pays et 36 cancers pour les deux genres (Source image : Sung et. al. 2021) [10]

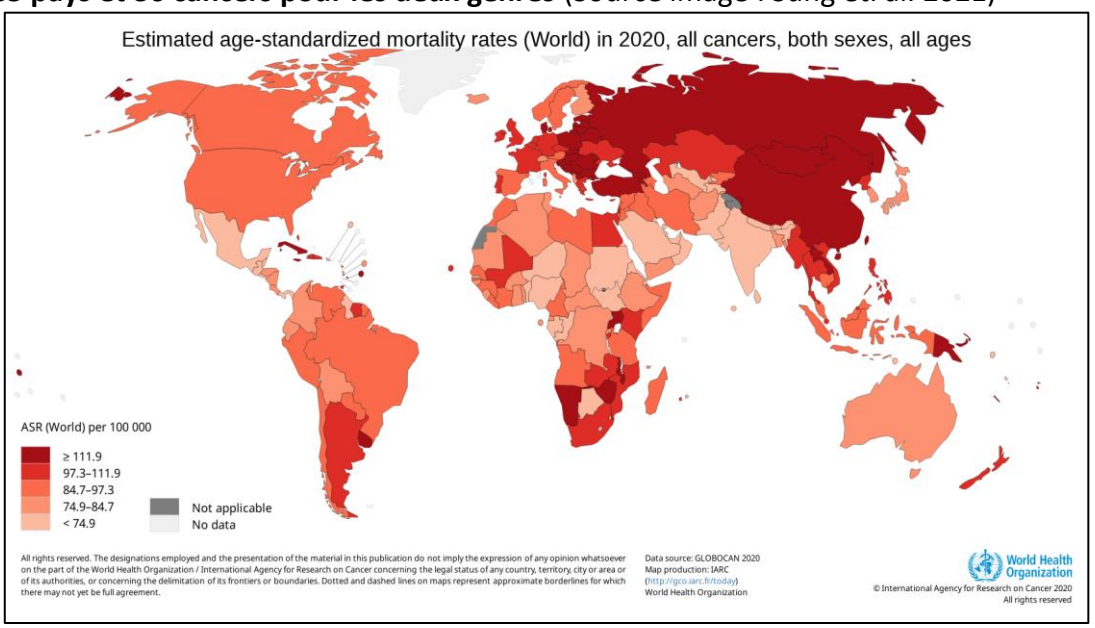


Figure 2: Mortalité liée aux cancers dans le monde suivant la base de données de GLOBOCAN 185 pays et 36 cancers pour les deux genres *(Source image : Sung et. al. 2021)* [10]

Les cancers les plus répandus et plus mortels sont répertoriés dans le Tableau 1et Tableau 2 décrits ci-après.

Classement	Type de cancer	% sur la totalité des cancers en 2020
1	Cancer du sein	11.7
2	Cancer du poumon	11.4
3	Cancer colorectal	10
4	Cancer de la prostate	7.3

Tableau 1: Tableau des cancers les plus répandus dans le monde selon le AIRC *(*Agence Internationale de Recherche du Cancer)

Classement	Type de cancer	Nombre de morts en 2020
1	Cancer du poumon	1.8 millions
2	Cancer colorectal	935 173
3	Cancer du foie	830 180
4	Cancer de l'estomac	768 793
5	Cancer du sein	684 996

Tableau 2: Classement des cancers les plus mortels selon l'AIRC

Selon les prévisions l'agence internationale de la recherche sur le cancer (AIRC), d'ici 2030, 3/4 des décès liés au cancer auront lieu dans les pays à faible revenu. Ceci est dû au manque d'infrastructures de santé, le coût des techniques de diagnostic et de soins non accessibles dans ces pays.[11] De plus, les disparités génétiques entre les pays génèrent des cancers différents à traiter différemment.[12] Les faibles taux annoncés dans certains pays pourraient découler cependant d'un manque de données et d'enregistrements (i.e cas de cancers en Afrique).[11,13] Le nombre de cas de cancer dans ces régions a augmenté parallèlement à l'espérance de vie dans ces pays, pour atteindre en moyenne 60 ans en Afrique,.[14]

I.2. Les principaux cancers

I.2.A Cancer du sein

Le cancer du sein est le plus répandu au niveau mondial. Il est la cause d'1/4 des cancers des femmes et d'1/6 des décès causés par le cancer dans le monde. L'âge moyen des femmes touchées par ce cancer est de 61 ans.[15] La majorité des zones géographiques touchées sont les pays à revenu élevé, avec un taux de cancer de 88 % supérieur aux pays à faible revenu. Comme le montre la Figure 3, les régions les plus touchées par le cancer du sein sont l'Australie/Nouvelle Zélande, l'Europe de l'Ouest, Amérique du Nord et Europe du Nord. Les causes de cette prévalence au cancer du sein dans les pays du Nord sont les suivantes: [10,12]

- Ménopause tardive
- Thérapie hormonale pour la ménopause
- Age avancé de la première grossesse
- Contraceptifs oraux
- Alcool
- Surpoids
- Absence d'activité physique
- Stress
- Alimentation
- Pollution

Bien que le nombre de cas soit plus important dans les pays à revenu élevé, on compte un taux de mortalité supérieur de 17 % en Afrique subsaharienne par rapport aux pays à revenu élevé. Le taux de survie sur 5 ans du cancer du sein en Europe est de 82 %, alors qu'il n'est que de 46 % en Ouganda, moins de 39 % en Algérie et 12 % en Gambie.[13] Selon McCormak et. al[8], la mise en place d'un diagnostic précoce pourrait éviter 28 à 37 % des décès liés au cancer du sein dans les pays Africains.

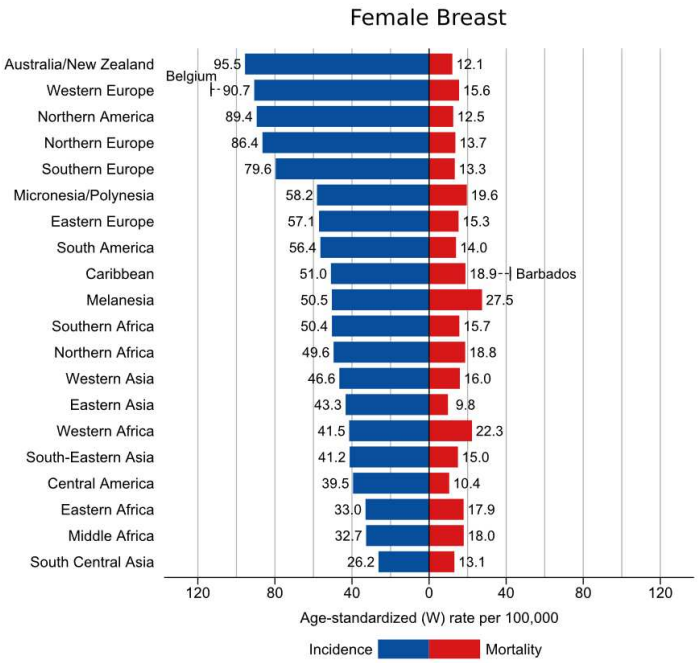


Figure 3: Taux (âges standardisés) et mortalité due au cancer du sein en 2020 en fonction des régions (Source image : Sung et. al. 2021) [10]

Pour les femmes d'Afrique du Nord, le taux du cancer du sein a doublé entre 2011, et 2018, suite également au changement de style de vie. [12] Cette tendance est toujours visible aujourd'hui (Figure 4). Le taux de mortalité, qui renvoie au nombre de morts dans une population, est globalement plus important dans les pays à faible revenu (cas du cancer du sein, Figure 4).

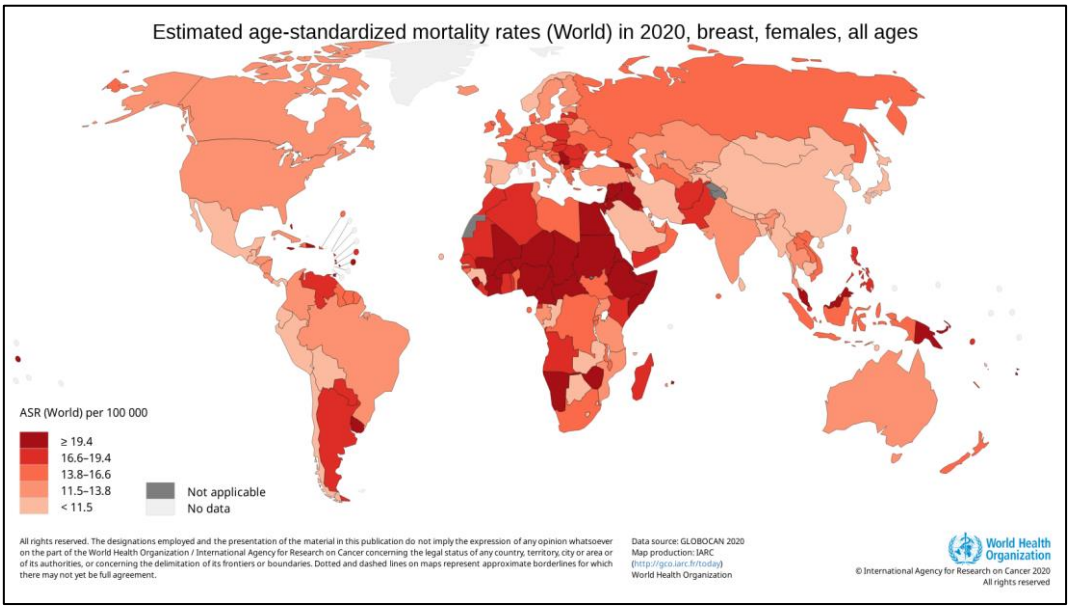


Figure 4: Mortalité du cancer du sein dans le monde la base de données GLOBOCAN (Source image : Sung et. al. 2021) [10]

I.2.B. Cancer du poumon

Selon les **Tableau 1** et **Tableau *2*** le cancer du poumon est le deuxième cancer le plus répandu au monde, et le plus mortel. Le taux de ce cancer est 3 à 4 fois plus important dans les pays à revenu élevé, comparé aux pays à faible revenu.[10] La moyenne d'âge des malades de ce cancer est de 70 ans.[16] Les causes principales de ce cancer sont la pollution de l'air et le tabagisme. La répartition mondiale du cancer du poumon est présentée sur la Figure 5. Au fil du temps, la répartition homme/femme du cancer du poumon tend à s'égaliser en Europe et en Amérique du Nord. Face à cela, l'organisation mondiale de la santé (OMS) a mis en place des mesures de prévention et de contrôle des fumeurs.

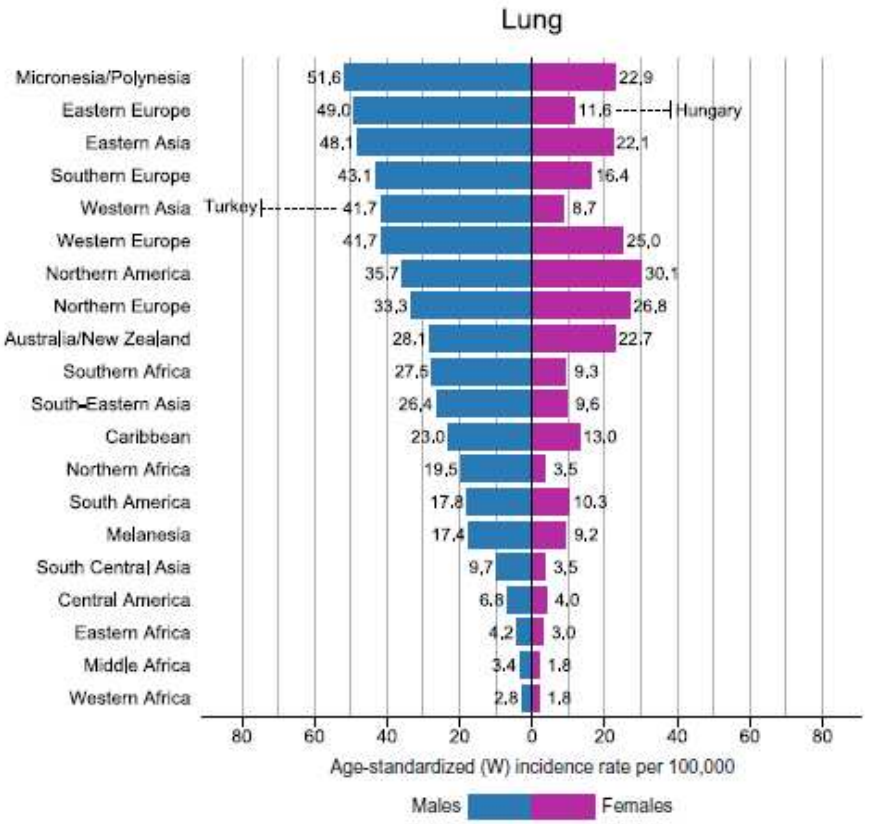


Figure 5: Taux (âges standardisés) du cancer du poumon en 2020 en fonction des régions et du genre (Source image : Sung et. al. 2021) [10]

I.2.C. Cancer colorectal

Le cancer colorectal est troisième au classement des cancers les plus répandus au monde et le deuxième plus mortel (Tableau 1 et Tableau 2). Au niveau de la répartition présentée en Figure 5, les zones géographiques les plus touchées sont aussi les pays à revenu élevé (Australie/Nouvelle Zélande, Amérique du Nord, Hongrie Norvège) contrairement aux pays à faible revenu (pays d'Afrique et d'Asie centre Sud). Les principales causes de ce cancer sont le tabagisme, la consommation d'alcool et une mauvaise alimentation.

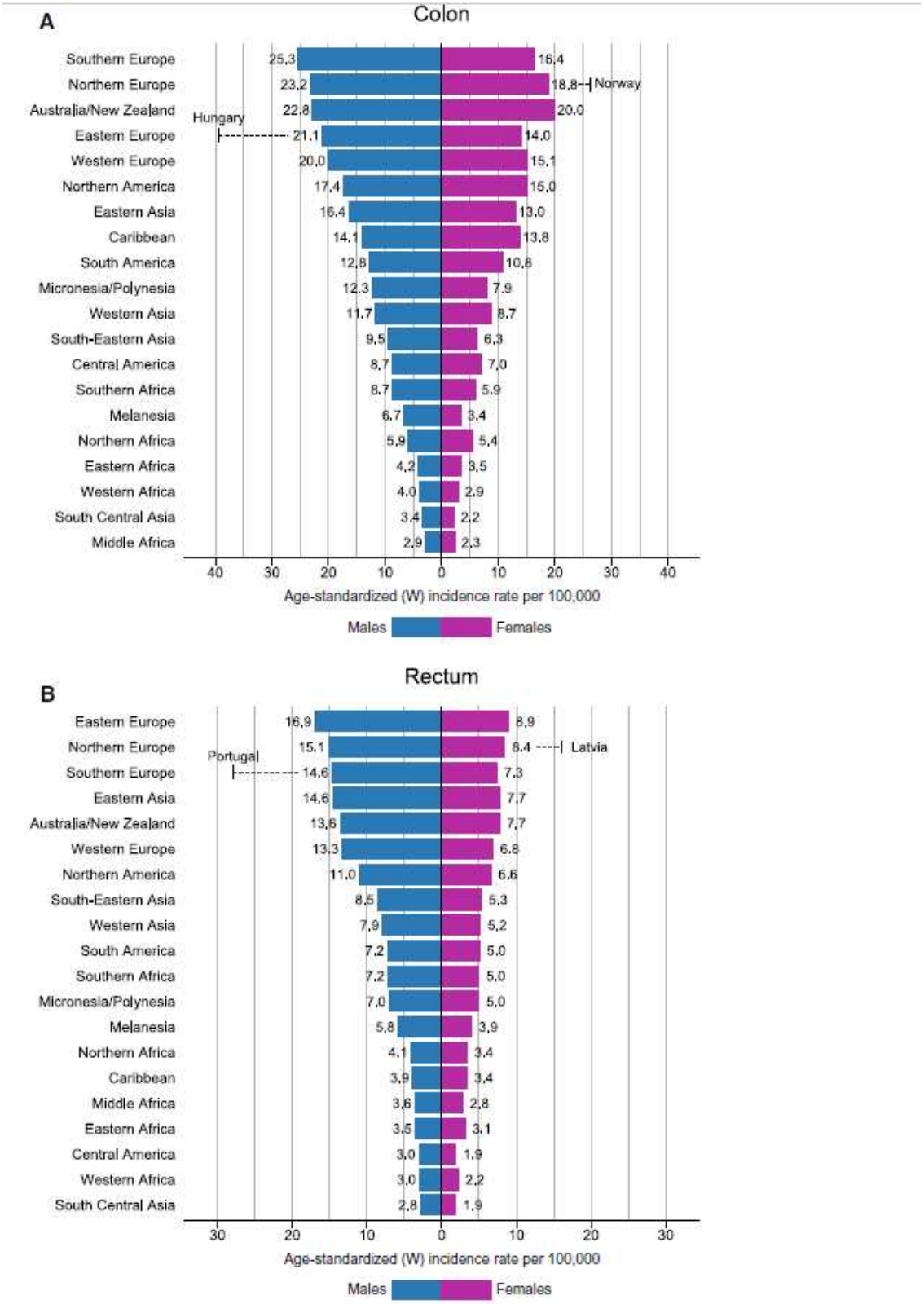


Figure 6: Taux (âges standardisés) du cancer du côlon (A) et du rectum (B) en 2020 en fonction des régions et du genre (Source image : Sung et. al. 2021) [10]

Le cancer est donc une maladie d'incidence mondiale. Comme nous l'avons vu, les chiffres sont conséquents et ne cessent d'augmenter. L'une des stratégies qui permettrait de réduire le taux de mortalité, serait d'établir une technique de diagnostic précoce. Pour se faire il nous faut dans un premier temps comprendre les mécanismes biologiques impliqués.

I.3. Biologie du cancer

L'Homme est composé d'environ 37,2 milliards de cellules. Elles sont les éléments de base d'un organisme et possèdent des caractéristiques spécifiques qui leur permettent d'assurer le bon fonctionnement des tissus et organes. Elles sont par exemple capables de contrôler leur réplication (formation de deux cellules filles possédant le même ADN à partir d'une cellule mère), elles peuvent également entrer dans une mort cellulaire programmée, appelée apoptose, à la suite de dommages cellulaires importants (ex : mutations génétiques non réparées) ou encore se fixer les unes aux autres afin de rester dans leur tissu respectif. Enfin, elles sont capables de se spécialiser en différentes cellules, ayant chacune une fonction bien définie.

Le cancer est une maladie génétique complexe induite par des modifications des gènes de cellules. Ces mutations génétiques représentent la première étape de la cancérogenèse, c'est-à-dire la transformation d'une cellule saine en cellule cancéreuse, appelée l'initiation, et sont favorisées par un agent initiateur mutagène (agent biologique, chimique, UV etc.). Suite à l'initiation, deux autres phases interviennent dans la cancérogenèse : la promotion et la progression (Figure 7). Les deux premières étapes font partie de la phase infra-clinique (aucun symptômes) et la dernière étape correspond à la phase clinique (présence de symptômes).

Suivant l'évolution Darwinienne, plus les cellules cancéreuses mutent, plus elles acquièrent des facultés leur permettant de survivre au détriment des cellules les moins favorisées par la sélection naturelle. Ce phénomène s'appelle la sélection clonale. Toutes ces phases de développement peuvent s'étaler sur plusieurs années, ce qui peut provoquer une déclaration d'un premier cancer chez des personnes relativement âgées, ayant par exemple de plus de 70 ans.
Au terme de cette transformation, la cellule cancéreuse a acquis un certain nombre de caractéristiques dont [17]:

- son indépendance vis-à-vis des signaux qui régulent (favorisent ou freinent) habituellement sa croissance et sa division,
- sa capacité à échapper au processus de mort cellulaire programmée,
- sa capacité à se diviser indéfiniment,
- sa capacité à s'évader et à développer des métastases (tumeur provenant de cellules cancéreuses s'étant détachées, pour circuler et s'installer dans d'autres parties du corps). Au cours de ce trajet, des changements de phénotypes se font en faveur de son adaptation au nouvel environnement.[18]

L’ensemble des mécanismes, causes et effets du cancer sont représentés sur la Figure 7.

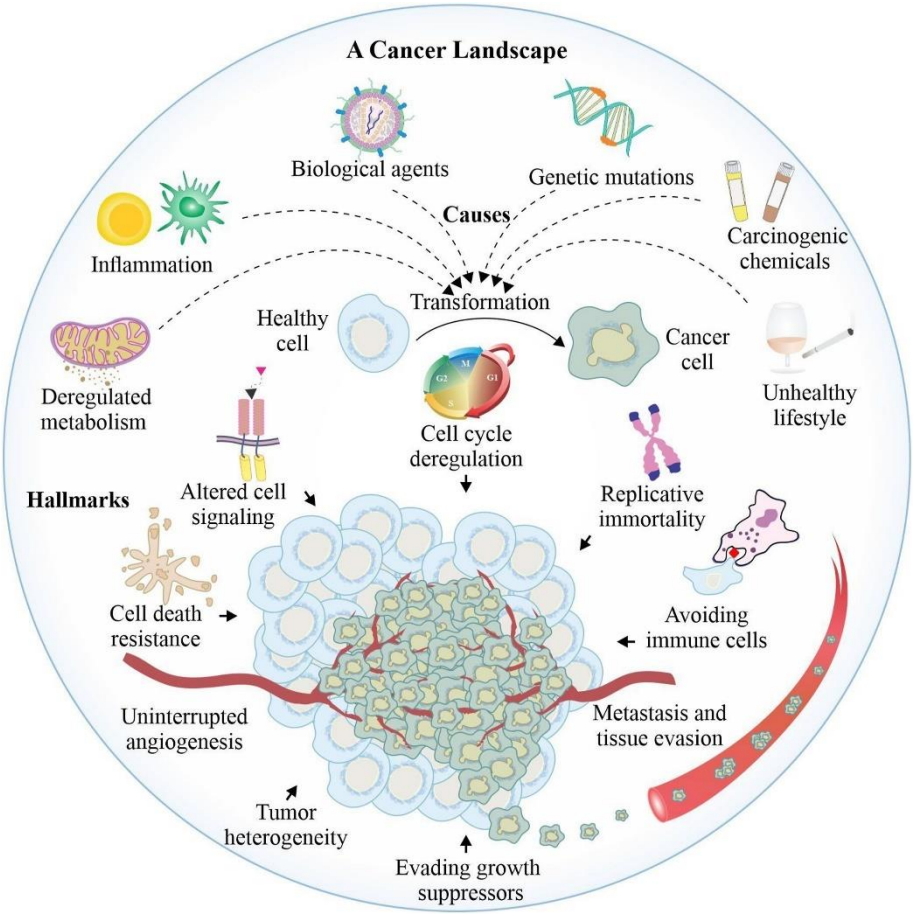


Figure 7: Schéma des évènements biologiques et moléculaires lors de la transformation d'une cellule saine à cancéreuse *(*Source image : Anand et. al. 2022) [19]

Epigénétique et cancer

L’origine du cancer est aussi bien génétique qu’épigénétique.[20] On peut apparenter l’origine du cancer à un changement de l’activité ou de la fonction des gènes. Cette théorie appelée TMS (Théorie de Mutation Somatique) est controversée aujourd’hui, et l’origine des mutations pourrait être différente. Au cours des dernières années, de nombreuses théories sur l’origine des cancers ont été établies, mais aucune n’a été validée .[20]

Selon, Jones et Hanahan, des altérations épigénétiques peuvent influer sur la progression du cancer.[21,22] L’ épigénétique correspond à l’ensemble des mécanismes qui peuvent moduler l’expression des gènes sans opérer de changement sur la séquence d’ADN.[23] Outre l’ADN, la chromatine peut subir des modifications épigénétiques. La chromatine est faite d’ADN et de protéines formant ainsi des granules dans le noyau des cellules. Des modifications covalentes de la chromatine telles que la méthylation, l’acétylation, et la glycosylation peuvent s’opèrer.[24,25] Parmi celles-ci, La glycosylation sera centrale dans notre étude.

Les changements épigénétiques se font principalement au niveau de la cellule somatique.[26] Ils peuvent être causés par des mutations ou par l'âge avancé des cellules. Ils peuvent également toucher les cellules souches, qui deviennent alors plus sensibles aux mutations.[27] L'âge et l'exposition environnementale jouent sur l'épigénome, de même que l'alimentation. Ainsi des déficits en vitamine C peuvent augmenter la méthylation de l'ADN, ce qui provoque l'expression dérégulée des gènes.[28,29]
50% des cancers sont liés à la mutation des protéines de la chromatine.[21,30] Les cellules malignes présentent en effet une altération au niveau des protéines responsables de la méthylation de l'ADN et une modification de la structure de la chromatine. La méthylation de l'ADN est la modification épigénétique la plus étudiée. Les cellules cancéreuses deviennent hyper-méthylées ou hypo-méthylées. L'hyper-méthylation a pour effet de freiner les gènes suppresseurs tumoraux, à la faveur de la progression de la tumeur. En effet, les gènes suppresseurs ont pour rôle d'empêcher la division et la multiplication des cellules tumorales.

Les principaux mécanismes épigénétiques en plus de la méthylation de l'ADN sont la modification des histones, et l'action des ARN (Acide ribonucléique) non codants.[31] Leur dérégulation provoque le développement du cancer.[32–35]

I.4. Diagnostic et traitement des cancers solides

I.4.A. Diagnostic de cancer

La biopsie

La biopsie est l'une des techniques les plus fréquemment utilisées pour le diagnostic du cancer. Il s'agit d'un prélèvement d'un fragment de tissu visant à faire une analyse microscopique. Cependant, cette méthode invasive peut s'avérer douloureuse pour le patient, la biopsie se faisant plusieurs fois avant et au cours du traitement.

L'imagerie médicale

L'imagerie médicale permet l'obtention de clichés de l'organisme. Elle regroupe différentes techniques, à savoir la radiographie, l'imagerie par résonnance magnétique, les ultrasons et les rayons X. Elle permet d'avoir des informations morphologiques, métaboliques, structurales et fonctionnelles.[36] Les avantages de cette technique sont les suivants:

- Contrôle en temps réel
- Pas de destruction des tissus
- Peu, voire non-invasive
- Large champ d'application au niveau temporel (couvre plusieurs phases de la maladie).

L'imagerie médicale peut être divisée en plusieurs sous-domaines, à savoir, la radiographie, l'imagerie en coupe, et l'imagerie fonctionnelle.
La radiographie est une technique d'imagerie qui utilise les rayons X et gamma pour produire une image. C'est la technique d'imagerie la plus ancienne, découverte en 1885 par le scientifique allemand Wilhelm Conrad Röntgen.[37]



Figure 8: Tumeur du poumon par radiographie (cercle rouge).(Source image : https://www.santeweb.ch) [38]

Comme le montre la Figure 8, la radiographie permet d'observer les os, les poumons ainsi que les articulations. Cette technique est la première à avoir été utilisée pour lever un doute dans l'identification d'une masse dans le cadre d'un cancer thoracique. Malheureusement, elle ne permet parfois pas de distinguer une masse non cancéreuse d'une tumeur cancéreuse.[39] De plus, sa sensibilité pour détecter une masse est de l'ordre du millimètre.[40]

L'imagerie en coupe

L'imagerie en coupe permet l'obtention d'une image 3D d'un organe cible. Elle permet de préciser les résultats obtenus par la radiographie. Les trois techniques les plus utilisées en imagerie en coupe sont : l'échographie, l'imagerie par résonnance magnétique (IRM), et la tomodensitométrie.

L'échographie implique l'utilisation des ultrasons. Elle mesure la capacité des tissus à renvoyer les ultrasons émis. Cette méthode qui produit des images en temps réel, est non irradiante, non indolore, en somme, non invasive. Elle est très utilisée pour les cancers du sein, en combinaison à la mammographie.

L'IRM est une technique basée sur la résonnance magnétique nucléaire. Un champ magnétique est appliqué sur un individu. Ceci va générer une rotation des atomes d'hydrogène des tissus du patient. Le retour à une position initiale libère de l'énergie, une couleur noire est générée, nuancée par la quantité d'atomes d'hydrogène.

Cette technique permet d'avoir une image nette, notamment sur les tissus mous contrairement aux rayons X. On obtient également la taille exacte des tumeurs contrairement aux estimations données par la radiographie ou l'échographie.[41] Cette technique est très sensible, car elle permet de détecter une tumeur de 5 mm.[42]

La tomodensitométrie, qui se nomme aussi scanner et tomographie, utilise aussi les rayons X, afin d'obtenir des clichés superposés d'un tissu. La technique est très sensible, car capable de détecter des tumeurs de 5 mm.[43] Cependant, sa sensibilité baisse pour des tumeurs inférieures à 1,5 mm.[44]

L'imagerie fonctionnelle

Il existe deux principales techniques d'imagerie fonctionnelle : la scintigraphie et la tomographie par émission de positons PET-SCAN (Figure 9).

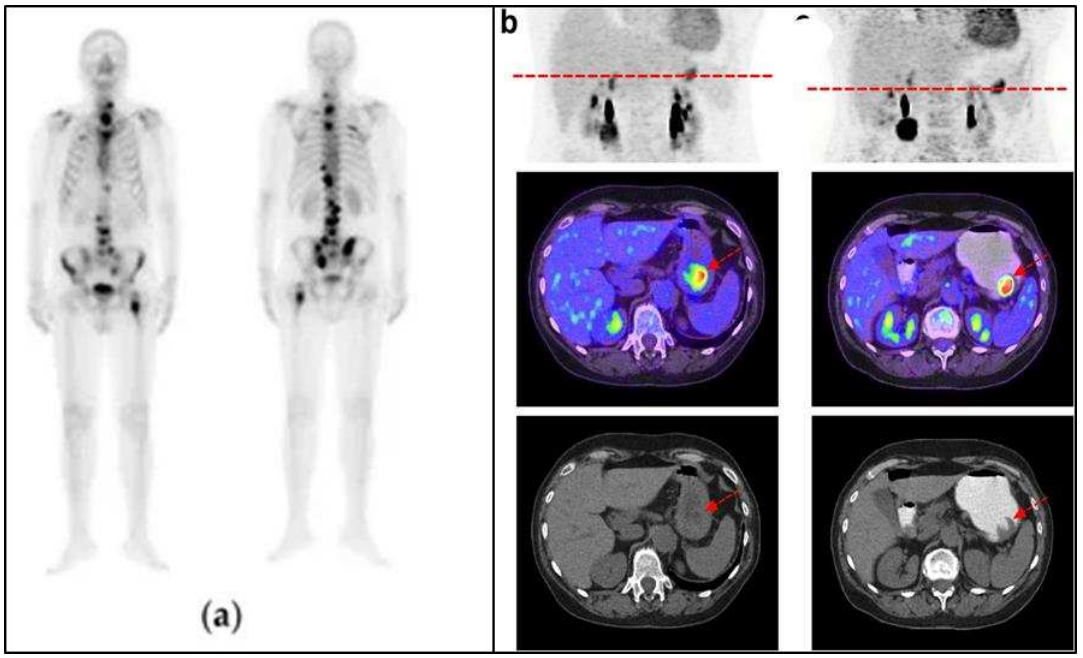


Figure 9: Clichés de métastases observées par scintigraphie (à gauche) et par TEP-scan (à droite). *À gauche, chaque point noir correspond à une métastase. À droite, les métastases sont indiquées par les flèches et correspondent aux zones colorées.*
(Source image: Dudoignon et. al 2020 ; Papandrianos et. al 2020) [45,46]

La scintigraphie consiste en l'injection d'un traceur radioactif par voie intraveineuse. Ce traceur dépend de la nature des tissus ciblés. Après un certain délai, un appareil appelé « caméra gamma » passe au-dessus des tissus ciblés et capte les rayons émis par le traceur. Ensuite, la caméra forme des images fonctionnelles de l'organisme. Cette technique permet détecter une tumeur supérieure à 1 cm.[47]

La technique de PET-SCAN permet le suivi de la métabolisation d'un traceur radioactif (ex : ^{18}F glucose). Cette métabolisation génère des photons, qui seront ensuite détectés par une caméra qui par la suite, seront traduits en image (Figure 9).
Les isotopes utilisés dans cette technique sont les suivants : ^{18}F, ^{68}Ga,^{123}I,^{124}I, ^{111}In.[48] Les tumeurs et les métastases consomment abondement le glucose marqué au ^{18}F.
Utiliser le glucose comme traceur permet donc de révéler les zones spécifiques à étudier.
Enfin, la technique de TEP permet de détecter des tumeurs d'environ 5 mm.[49]

I.4.B. Traitement des cancers

Les traitements disponibles aujourd'hui sont principalement la chirurgie, radiothérapie, et chimiothérapie. Depuis les 10 dernières années, elles sont utilisées seules ou combinées.[50]

Chirurgie

Selon l'INC (Institut National du Cancer), de nombreux cancers sont traités par chirurgie.[51] Elle traite seulement une zone, à savoir celle où la tumeur est retirée. Lorsque le risque est trop important, seulement une partie de la tumeur est retirée. Les inconvénients de la chirurgie sont la douleur post-opératoire et le risque d'infection.

Néanmoins, au fil des années, elle est devenue de moins en moins agressive. En effet, dans le cadre du cancer du sein, la mastectomie, qui consiste à retirer l'entièreté du sein, a été remplacée par la lumpectomie, à savoir, retirer une partie du sein.[52] De plus, la mastectomie entière, technique très invasive a plus d'effets secondaires que la lumpectomie.[53] Le traitement du cancer a progressé et évolué, contribuant à la survie des patients à long terme. En effet, pour le cancer du sein, le taux de survie à 5 ans était de 40% en 1964, et de 87% en 2014. Concernant le cancer colorectal, la survie à 5 ans était de 23% en 1964, et 57% en 2014.[54–56] Ceci n'est pas seulement dû à la chirurgie, mais aussi au diagnostic précoce et les avancées radio- et chimio-thérapeutiques.

Selon une étude comparative de Benjamin et. al [52], l'efficacité de la chirurgie sur les 20 dernières années peut être remise en cause. Selon l'étude, la chirurgie affecterait l'évolution de la maladie, et l'intervention chirurgicale ne serait pas toujours bénéfique. Cependant, elle le serait sur les tissus pré-malins.

Radiothérapie

La radiothérapie (RT) est une thérapie qui a pour but de stopper la multiplication des cellules cancéreuses. En 2012, 50% des patients cancéreux étaient traités par radiothérapie.[57–59] C'est une méthode peu coûteuse qui représente 5% du coût total des traitements liés au cancer.[60]

La RT est l'émission et la propagation d'une énergie dans l'espace ou un milieu solide et la transmission de radiations électromagnétiques. Elle utilise les rayons X et les rayons Gamma, tout comme la radiographie, pour créer une interaction entre les photons et la matière. Elle concourt à l'obtention de modifications génétiques et la suppression irréversible des cellules cancéreuses.[61]

L'envoi de hautes radiations va toucher l'ADN des cellules et bloquer leur division entraînant leur mort.[62] Concernant le mode d'action, soit la radiation détruit directement l'ADN, soit elle génère des radicaux libres à partir de l'eau des cellules (radiolyse), qui vont détruire l'ADN. Ainsi après une radiation, l'ADN double brin est rompu dans la cellule en phase mitotique (division cellulaire), appelée « catastrophe mitotique ».[63] La catastrophe mitotique est la principale cause de la mort des cellules cancéreuses après radiation.[64]

Le succès de la RT dépend de la radiorésistance des cellules tumorales. En effet, la radiorésistance dépend des phases du cycle cellulaire. En phase M (mitotique ; division cellulaire) les cellules sont plus vulnérables car elles mettent leur énergie dans la ségrégation des chromosomes et n'ont pas le temps d'avoir une activité de réparation. Tandis qu'en phase S (synthèse ; multiplication de l'ADN), les cellules sont plus résistantes car l'augmentation de la synthèse de l'ADN est concomitante avec l'augmentation des enzymes de réparation.[65] Ainsi, il se trouve que les cellules tumorales pauvrement différentiées sont plus radiosensibles car ces cellules se divisent en permanence.

Dans le cas d'un traitement en combinaison avec la chirurgie, la RT n'aura pas le même effet curatif en fonction du stade du traitement. Avant l'action chirurgicale, la RT aide à réduire la taille de la tumeur. Après intervention chirurgicale, la radiothérapie sert à supprimer les potentielles cellules tumorales restantes. De plus, la radiothérapie ne détruit pas les cellules de manière instantanée. Cela se fait des heures, jours, semaines après le traitement.[57]

La radiothérapie n'est pas sans risque de toxicité pour les cellules saines. Plus la dose journalière de radiation est basse, plus sa toxicité diminue. L'irradiation avec une faible dose a pu fonctionner uniquement dans le cas du myélome et du lymphome. Une balance est donc à trouver entre envoyer une dose assez faible pour épargner les cellules saines, et assez forte pour détruire les cellules cancéreuses.[32]

Ainsi face à la RT, on distingue plusieurs types d'effets secondaires apparaissant sur les tissus sains au cours du temps:[66]

- Effet précoce : il apparaît en quelques semaines. Il apparaît lorsque la cellule est en phase inter-mitotique et touche la peau et les muqueuses.
- Effet consécutif : il apparaît lorsque les effets précoces ne sont pas traités et que la situation persiste.
- Effet tardif : il apparaît après des mois, années après l'exposition aux radiations. L'effet tardif survient lorsque la cellule en phase post-mitotique et touche le foie, les reins, les muscles ainsi que les os.

Pour exemple, la nature des effets secondaires suite à la RT sont listés dans le Tableau 3.

Organe/Tissu	Effets secondaires
Peau et muqueuses	Desquamation, alopécie, fibrose musculaire
Poumon	Dyspnée[a], hypoxie, fièvre, douleur pulmonaire, inflammation, fibrose
Cœur	Péricardite[b],effusion péricardiale, péricardite constrictive, disfonctionnement valvaire
Système nerveux	Myélopathie[c], maladie cérébro-vasculaire, fatigue, maux de tête, vomissement, nausée
Dents	Sévères carries

Tableau 3: Liste des effets secondaires liés à la RT

[a] La dyspnée est une difficulté respiratoire traduite par un essoufflement.

[b] La péricardite est la libération d'un liquide par la partie du cœur appelée péricarde. Lors d'une inflammation, ce liquide augmente, provoquant alors le gonflement du péricarde, provoquant des douleurs musculaires et thoraciques, ainsi que des difficultés respiratoires.

[c] La myélopathie est la lésion d'une partie de la moelle épinière qui cause des dysfonctionnement nerveux.

Chimiothérapie

Le terme chimiothérapie a été inventé par Paul Ehrlich, qui travaillait sur les traitements de maladies infectieuses.[67] La chimiothérapie (CT) est l'utilisation d'agents médicamenteux pour détruire les cellules cancéreuses. Elle entrave la croissance et la division des cellules cancéreuses, car celles-ci se divisent et se multiplient plus vite que les cellules saines. Les cellules cancéreuses ont aussi un stress métabolique plus important. Ces facteurs sont ciblés par les agents chimio-thérapeutiques pour détruire sélectivement les cellules cancéreuses.[19]

Les agents de la CT peuvent être administrés par voie orale, intraveineuse, sous-cutanée ou intra-musculaire. Ces agents sont métabolisés par le foie et le rein, à la suite de quoi il peut y avoir un effet toxique pouvant provoquer un dysfonctionnement de ces organes. C'est pourquoi il faut trouver une dose de médicament qui soit équilibrée.

Dans les années 70 la bléomycine, vinblastine, et cisplatine étaient les traitements les plus fréquemment utilisés[68], cependant ils provoquaient d'importants vomissements. On peut observer les répercussions de la CT sur l'intestin, le sang, la moelle épinière, la peau et les cheveux.[68] Les organes vitaux comme le cœur, le poumon et le cerveau sont aussi touchés. De plus, à long terme, il peut y avoir certains effets secondaires tels que la résistance aux médicaments, l'infertilité, et la carcinogénicité (capacité à développer une tumeur).

Les effets secondaires les plus fréquents sont représentés sur le Tableau 4.

Système d'organe/tissu	Effet secondaire
Système nerveux	Nausée, vomissement, myelosupression[a]
Peau, muqueuse	Alopécie
Système immunitaire	Immunosuppression
Système reproductif	Stérilité

Tableau 4: Liste des effets secondaires liés à la CT

[a] La myelosuppression est l'incapacité de la moelle osseuse à produire des globules rouges, des plaquettes, et des globule blancs

La chimiothérapie (CT) réduit la morbidité. Malgré tout, la chimiothérapie peut endommager les cellules saines.[69] Cette technique rencontre les problèmes suivants :

- La résistance des cellules cancéreuses au médicament
- La sélection du dosage est difficile
- La spécificité est faible
- La métabolisation du médicament est rapide
- Les effets secondaires sont lourds

Médicament	Cible moléculaire	Mécanisme d'action	Effet secondaire	Organe touché	Type
Altrétamine	Site catalytique de l'ADN polymérase	Inhibition	Nausée Vomissement	Sang	Agents alkylants antinéoplasiques
Cisplatine	ADN-3-methyl glycosylase Macroglobuline alpha 2 Serotransferrine Protéine du transport du cuivre (ATOX1)	Alkylation/ cross-linking de l'ADN	Nausée Vomissement	Testicule Ovaire Cou Toxicité du rein Carence (Mg,K,Ca)	
Vinblastine	Tubuline α,β,ε,γ	Formation d'adduit de l'ADN	Faible numération globulaire Fatigue Fébrilité Réactions sur les sites d'injection	Poitrine Testicule Os Sang	
Doxorobucine	ADN ADN-topoisomérase --α	Intercalation de l'ADN Inhibition de la topoisomérase		Ovaire Poitrine Estomac Sang Vessie Bronche	
Cytarabine	ADN ADN- -polymérase β	Inhibition topoisomérase Alkylation, cross-linking de l'ADN	Maux de tête Nausées Vomissement	Sang	Antimétabolites dérivés de Pyrimidine (antagonistes de pyrimidine)
Fluorouracile	Thymydylate synthase ADN ARN	Intercalation et déstabilisation de l'ARN et l'ADN	Diarrhées Nausées Vomissement Perte d'appétit Réaction de la peau Faible numération globulaire	Poitrine Colon Pancréas Peau Estomac	
Tegafur	Thymydylate synthase	Inhibition	Myelosuppression Neurotoxicité centrale Toxicité gastrointestin-ale	Estomac	
Decitabine	ADN ADN-méthyl-transférases	Inhibition	Fatigue, Nausée Fièvre Toux Faible numération globulaire	Sang	

Tableau 5: Présentation des deux grandes classes d'agents de chimiothérapie

Face aux chiffres alarmant du taux de cancer, comme vu précédemment, il est clair que les traitements actuels ne suffisent pas à réduire le taux de mortalité. Il y a une nécessité de trouver de nouveaux types de traitement et de nouvelles thérapies.

I.4.C. Objectif mondial : établir un diagnostic précoce

Selon l'OMS, un cancer détecté tôt a plus de chances de répondre favorablement aux traitements. Il se trouve que les cancers qui ont une grande chance de rémission après un traitement et une détection précoce sont :

- Le cancer du sein
- Le cancer cervical
- Le cancer oral
- Le cancer colorectal

L'OMS a donc établi un plan d'action global de 2013 à 2020 pour freiner la mortalité liée au cancer. Les engagements suivants ont été soutenus:[70]

- Établir un contrôle rapproché du cancer du sein
- Développer une stratégie de prévention et de contrôle des cancers
- Développer des outils pour favoriser la prévention et le diagnostic précoce, et le traitement des cancers

I.5. Biomarqueurs

Selon l'INC, un biomarqueur est : « une molécule biologique présente dans le sang, d'autres fluides corporels, ou tissus. Il est la preuve d'un processus normal ou anormal, d'une condition, ou d'une maladie telle que le cancer. » [71]

Les biomarqueurs permettent d'établir des diagnostics et pronostics de manière rapide et simple, permettant ainsi d'enclencher une thérapie au plus vite. Ainsi, leur utilisation est très prisée depuis ces dix dernières années. [9,71]

L'utilité du biomarqueur est de pouvoir discriminer une cellule saine d'une cellule malade, selon la présence ou l'absence de ses composants. La première découverte de biomarqueurs date de 1965, faite par Joseph Gold dans le cadre du cancer du côlon.[9] Une entité inconnue présente dans les cellules cancéreuses et non dans les cellules saines a été trouvée. Il nomma l'entité « carcinoembryonic antigen ».

Les biomarqueurs sont localisés dans des réservoirs différents (Figure 10). Dans l'haleine, on trouve des composés organiques volatils (COVs) tels que les hydrocarbures, alcools, cétones, aldéhydes, esters, et les composés aromatiques (ex: heptanal, eicosane)[9]. Dans les urines sont étudiés de nombreux métabolites (ex : urée, acide urique and acide hippurique). [72] Dans la salive, on détecte les biomolécules telles que l'ADN, l'ARN, les protéines, les métabolites (acides aminés, lipides, vitamines). Dans la sueur sont ciblés des métabolites et des protéines (dopamine, progestérone, mélatonine). [73]

Le sang est le réservoir dont l'analyse ouvre un plus large choix de ciblage et d'identification de protéines, des ARNm, et les ADN circulants.

La limite de certains biomarqueurs est qu'ils peuvent être en très faible concentration dans les réservoirs ciblés. De plus, leur concentration présente une grande variabilité interindividuelle, selon l'activité physique et le régime alimentaire du sujet.[9]

C'est dans ce contexte que s'inscrit mon projet de recherche à savoir, améliorer la sensibilité de détection des biomarqueurs par la protéomique et la volatolomique

induite. Étant donné la diversité des biomarqueurs des cancers, nous nous focaliserons sur une classe de marqueurs enzymatiques, les glycosidases.

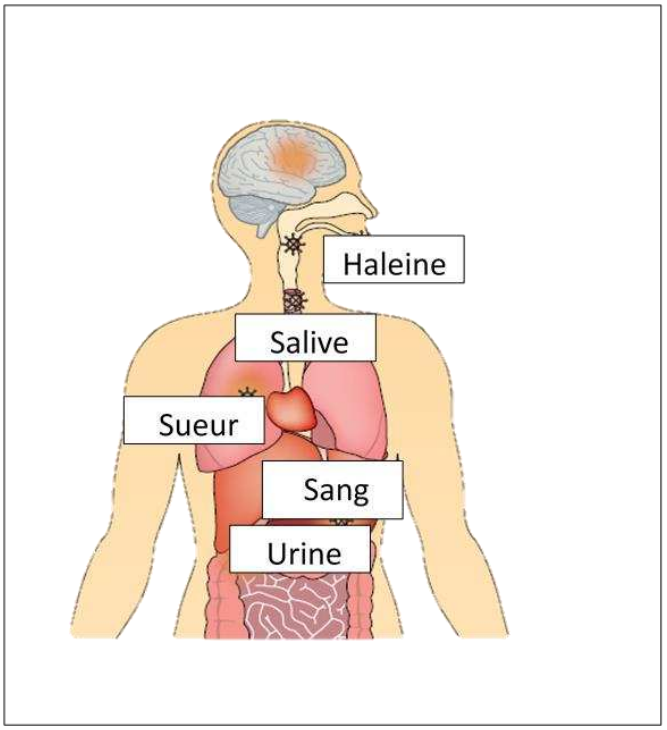


Figure 10: Réservoirs des biomarqueurs dans l'organisme

I.6. Les glycosidases

Les glycosidases sont des enzymes impliquées dans le processus de biosynthèse des glycanes. Un glycane est un biopolymère qui résulte de l'association de plusieurs monosaccharides à des protéines ou des lipides (Figure 11 et Figure 12).[74] La glycosylation est une modification post-traductionnelle majeure, aussi répandue au niveau cellulaire que la phosphorylation, avec 50 % des protéines humaines glycosylées.[25,75] La glycosylation a lieu dans le réticulum endoplasmique et l'appareil de Golgi, compartiments cellulaires.

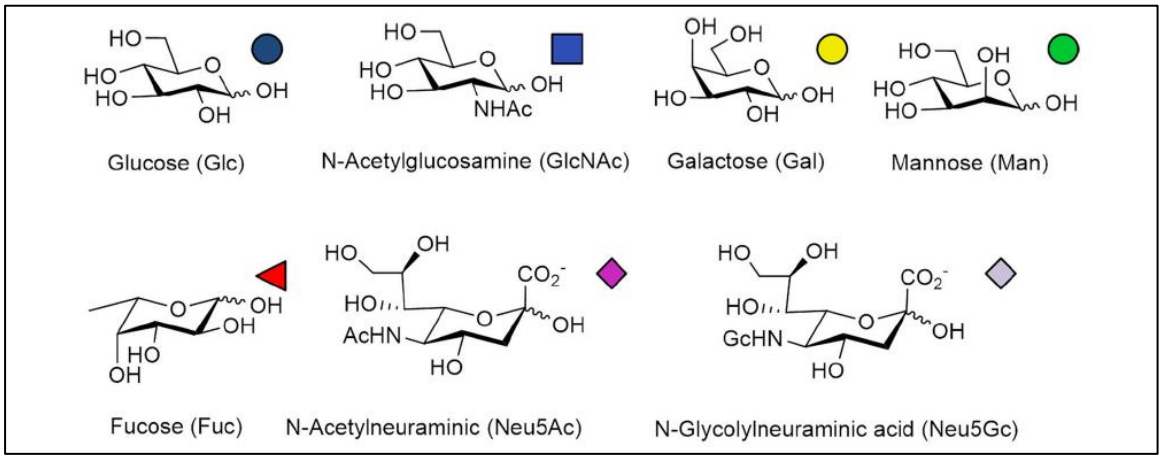


Figure 11: Schéma des monomères des *N*-glycanes (Source image: Yang C. et. al. 2019) [48]

On distingue différents types de glycanes, parmi-eux, les *N*-glycanes et les *O*-glycanes. Les *N*-glycanes sont des glycanes liés aux protéines par les résidus asparagine. Les *O*-glycanes sont liés aux protéines via les chaines latérales de la thréonine et la sérine. A ceux-ci s'apparenteront des enzymes en partie responsables de leur métabolisme appelées glycosidases. Un exemple de glycosidases connu est le lysozyme, responsable de la digestion du biopolymère de peptidoglycane (constituant de la paroi des bactéries).

Mécanismes

Les glycosidases sont responsables de la rupture d'une liaison glycosidique avec rétention ou inversion de l'anomérie du polysaccharide digéré.[76] Cette rupture de liaison C-O entre les sucres est suivie par la libération d'un alcool. Sans ces enzymes, il y aurait accumulation des glycosides pouvant conduire à la mort cellulaire.[77]

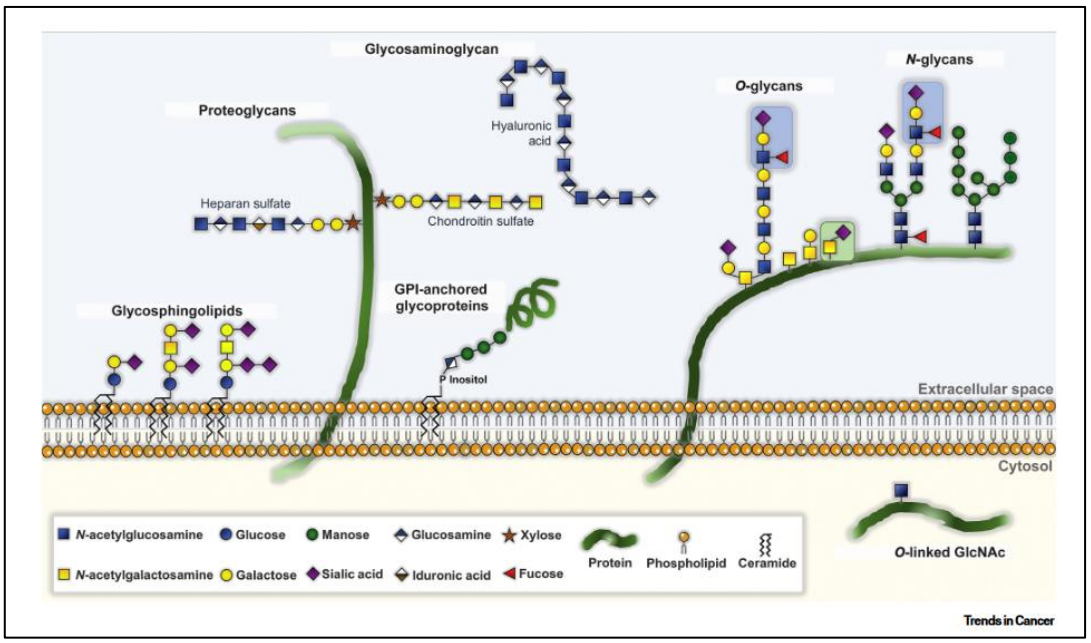


Figure 12: Schéma de la diversité des glycanes (Source image : Costa et. al 2020) [78]

L'activité des glycosidases conjointe avec celles des glycosyltransférases, est de transférer un monosaccharide ou un oligosaccharide à un polypeptide ou un gly*coli*pide. En somme, la glycosyltransférase a pour rôle de polymériser les chaines de glycanes, en action couplée avec la glycosidase, qui dépolymérise les chaines des glycanes.

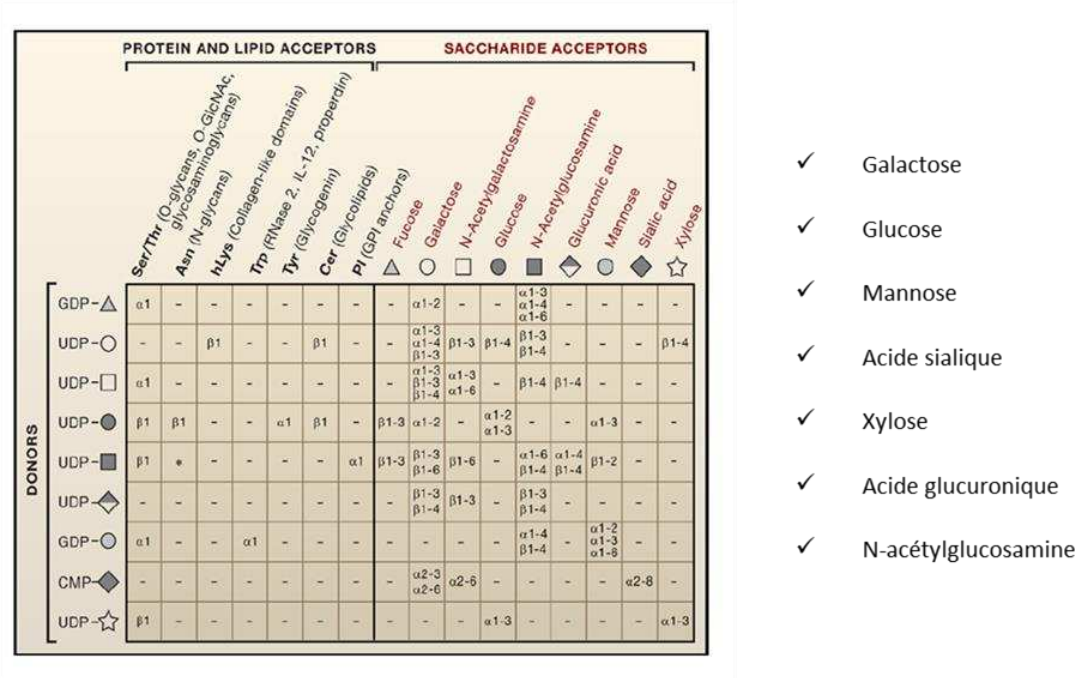


Figure 13: Composition des glycanes et sucres ciblés pour l'étude (Source image : Ohstubo et Marth, 2006) [67]

Chaque sucre cité dans la Figure 13 est pris en charge spécifiquement par une glycosidase. Les glycosidases choisies pour l'étude sont indiquées dans le Tableau 6.

- Repliement des protéines
- Activation des récepteurs
- Trafic membranaire
- Signalisation et communication intracellulaire
- Adhésion cellulaire
- Endocytose
- Clairance
- Reconnaissance cellulaire
- Reconnaissance immunologique
- Dégradation
- Conformation des glycoprotéines
- Interactions dans la matrice extracellulaire (MEC)
- Protection contre les protéases

Figure 14: Rôles des glycanes chez un sujet sain [155–157]

Rôle des glycosidases chez un sujet sain

Dans l'organisme d'un sujet sain, les fonctions des glycanes et donc des glycosidases sont mentionnées dans la Figure 14.[79]

Toutes ces fonctions renvoient principalement à ce qui se passe à l'extérieur de la cellule. Les communications cellulaires semblent s'établir par la détection/reconnaissance des

différents glycanes à la surface des cellules, dépendante de la nature de chaque sucre.[75] Un questionnement se fait alors pour savoir où sont produits les glycanes et comment ils se retrouvent dans le milieu extracellulaire.[80,80]
Il y a deux localisations principales des glycosidases. Les glycosidases présentes dans les lysosomes, ou à la surface des cellules. Chacune d'elles a un rôle différent, comme présenté dans le Tableau7.

Localisation de la glycosidase	Rôle associé
Lysosome	Digestion intracellulaire
Appareil de Golgi	Prise en charge des glycoprotéines
Milieu extracellulaire	Digestion extracellulaire, réparation des tissus, Invasion
Membrane plasmique	Adhésion intracellulaire, hydrolyse extracellulaire de substrats

Tableau 6: Rôle et localisation des glycosidases dans une cellule normale

Dans la cellule, les glycanes sont dégradés dans des lysosomes par deux types de glycosidases : les endoglycosidases et les exoglycosidases. Les endoglycosidases coupent à l'intérieur de la chaîne glycosidique et les exoglycosidases aux extrémités. Ces dernières ne reconnaissent qu'un monosaccharide à retirer. La dégradation se fait essentiellement dans le lysosome mais peut aussi se faire en milieu extracellulaire.
Chez un sujet sain, il n'y a pas de dégradation des glycanes en milieu extracellulaire, cela se fait uniquement dans les lysosomes. En revanche, plusieurs études ont démontré la présence d'exoglycosidases à l'extérieur des cellules, sans que leur rôle ne soit encore clairement élucidé.

Voie de sécrétion

Après glycosylation, les glycoconjugués restent dans l'appareil de Golgi pendant quelques secondes ou minutes. À ce stade, il y a une très forte activité de glycosylation. Le transport des glycoprotéines dans l'appareil de Golgi, ainsi que la localisation des glycosidases et glycotransférases contribuent à la glycosylation normale de la protéine.[78] La voie de sécrétion extracellulaire des glycoconjugués, comme son nom l'indique, est responsable des sécrétions dans le milieu extracellulaire. Elle assure la translocation et le transport des protéines secrétées et lysosomales. Elle assure aussi les modifications post-traductionnelles, le contrôle de la qualité du repliement des glycoprotéines ainsi que l'adressage des glycoprotéines à la membrane plasmique[81].

Dans cette voie de sécrétion sont concernés trois compartiments :

- Le réticulum endoplasmique : les polypeptides sont transloqués dans la membrane du RE, et sont pris en charge par des glucosidases I et II, et mannosidases.[74] C'est dans ce compartiment qu'a lieu la *N*-glycosylation
- Le compartiment intermédiaire RE-Golgi (ERGIC) : joue le rôle de médiateur dans le transport entre le RE et l'appareil de Golgi
- L'appareil de Golgi : dans ce dernier compartiment de la voie de sécrétion les polypeptides sont polymérisés, ramifiés et glycosylés.

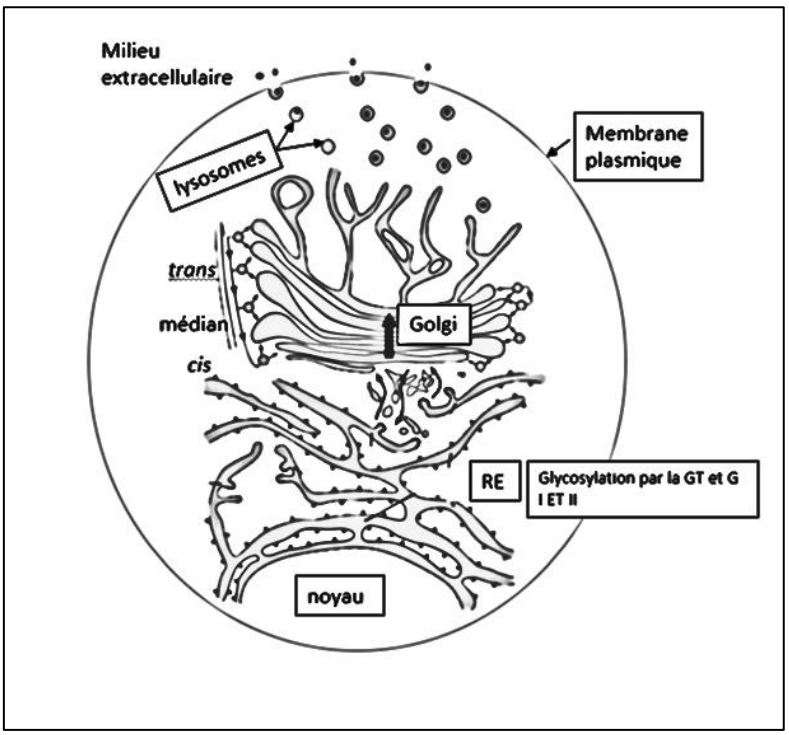


Figure 15: Schéma de la voie de sécrétion

Les glycosidases ancrées dans les membranes du réticulum endoplasmique (RE) et de l'appareil de Golgi jouent un rôle dans le développement et l'élagage des glycoprotéines.[82,83] Dans la biosynthèse des glycoprotéines liées aux *N*-glycanes, les sucres du cœur du glycane sont attachés à un intermédiaire de dolichol, qui par la suite

les transfèrera sur une protéine. La chaîne de sucres obtenue est composée de plusieurs fragments de glucose, qui seront ensuite supprimés ou pris en charge pas les glycosidases du RE. Le complexe GERL (Golgi-Endoplasmic Reticulum Lysosome) est un système membranaire qui forme des vésicules et libère les hydrolases des lysosomes. Avant l'insertion dans la membrane plasmique ou la sécrétion dans le milieu extracellulaire, le mannose des glycoprotéines peut être enlevé par les glycosidases. Dès lors, le N-acétylglucosamine, le galactose, et l'acide sialique peuvent être ajoutés à une chaîne d'oligosaccharide par les glycosyltransférases de l'appareil de Golgi.[84]

Glycosylation dans le cadre cancer

Les glycanes du sérum sont des biomarqueurs pour le pronostic et le suivi des cancers.[85] Depuis quelques années, la glycomique, qui est l'étude des glycanes, est en plein essor dans la recherche médicale. Étant donné la présence de glycanes circulants, la présence de certaines glycosidases dans le sang a été démontrée.

Ainsi, on note des changements de fucosylation et polymérisation des *N*-glycanes dans les tissus malins.[75] La surexpression des *N*-glycanes ramifiés a un impact sur l'adhésion cellule-cellule par une déstabilisation de l'E-cadhérine et amplifie la dissociation de la cellule tumorale et l'invasion[86]. L'expression d'*O*-glycanes non maturés promeut la transformation oncogènique.[78]

Le *N*-glycome du plasma a aussi été mis en lumière dans les maladies humaines. L'altération de la *N*-glycosylation a été trouvée en plus grande partie chez les cellules cancéreuses que chez les cellules saines.[85,87,88] Dans le cadre du cancer, des glycanes hautement fucosylés, appelés antigènes de Lewis s'enrichissent à la surface des cellules.[89] Dans les conditions normales, ils sont à la surface des granulocytes et sur l'épithélium. Leur surexpression joue un rôle dans l'invasion et la propagation des métastases.[90]

La glycosylation favoriserait donc le développement des cellules cancéreuses, et la reconnaissance des glycanes semble jouer un rôle important dans ce processus biologique. L'altération de la glycosylation des protéines est caractéristique des cellules tumorales. Elle affecte les *N*-glycanes et *O*-glycanes des glycoprotéines, gly*coli*pides, ou glysaminoglycanes.[91] Ainsi les glycosidases ont été trouvées dans le fluide interstitiel de tumeur, le sérum de patients cancéreux et dans les tissus malins dans des proportions beaucoup plus élevées que dans les tissus sains.[92]

À titre d'exemple, la β-glucuronidase (β-glu) et la *N*-acetylglucosminidase (β-GlcNac) ont été trouvées dans la MEC.[92] La MEC correspond à une structure qui se trouve dans le milieu extracellulaire d'une ou plusieurs cellules, et fournit un support structurel pour des cellules ou des tissus. Composée de collagène et fibronectine, elle joue le rôle de ciment intercellulaire ou ciment entre cellules. La provenance de la β-glu et de la β-GlcNac la est probablement liée à leur exocytose par des lysozymes des cellules tumorales.[92]

Il s'avère que les glycosidases sont présentes dans le milieu extracellulaire des tumeurs. [92] En effet, elles ont été trouvées dans le sérum en proportions plus importantes chez les animaux atteints d'une tumeur que les animaux sains.
Les glycosidases, si actives dans le cancer contribuent parfois à la résistance contre les agents thérapeutiques. Par exemple, la résistance à la doxorobucine (DOX), agent de chimiothérapie listé précédemment (partie I.1. F.2.) est causée par la glycosylation à la surface des cellules. Celle-ci empêche la fixation de la DOX sur l'ADN.[93] De plus, certains glycanes empêchent la fixation des anticorps thérapeutiques sur les cibles cellulaires .[93–95] Les glycosidases sont à la fois des biomarqueurs, mais aussi des acteurs principaux qui favorisent le développement du cancer.

II. Principes de biochimie: Techniques de séparation, de fractionnement et de détection des protéines

Dans ce paragraphe ne seront pas abordées les techniques d'analyse des protéines basées sur la chromatographie liquide couplée à la spectrométrie de masse. Ces dernières sont présentées ultérieurement.

II.1.SDS-PAGE

La technique très répandue de SDS-PAGE (Électrophorèse sur gel de dodécyl-sulfate de sodium et de polyacrylamide) est très appréciée des biochimistes. Cette technique consiste à séparer les protéines selon leur taille. Les protéines sont diluées dans un tampon SDS, un détergent qui confère une charge négative à l'ensemble des protéines. Les protéines vont par la suite migrer sur un gel de polyacrylamide sous un champ électrique, de l'électrode négative à l'électrode positive (Figure 16). Elles migreront le long du gel selon leur taille. En effet, dans le maillage du gel, les petites protéines n'ont pas de difficulté à migrer vite et loin, contrairement aux grosses protéines. En plus de séparer les petites protéines des grosses, un échantillon étalon avec des composés à la taille définie permet d'estimer la taille de la protéine analysée en comparant son niveau de migration à celle des étalons.

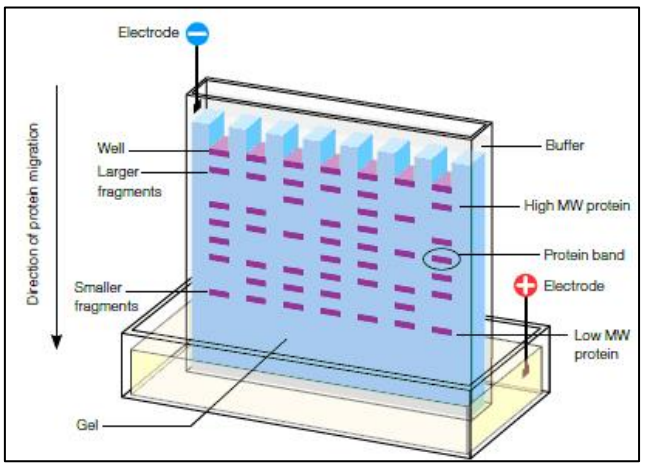


Figure 16: Schéma du principe de gel SDS-PAGE (Source image. https://www.bio-rad.com/fr-fr) [96]

Un gel SDS PAGE est composé de polyacrylamide, un polymère formé de la polymérisation radicalaire de l'acrylamide et du N,N-methylènebisacrylamide (bis). Cette réaction est initiée par le persulfate d'ammonium, et catalysée par la tétraméthyléthylènediamine (TEMED). Le gel est composé de deux parties (Figure 17), le gel de concentration (stacking gel), et le gel de résolution (separating gel). Le gel de concentration permet d'arrimer toutes les protéines au même niveau dans le gel, à savoir, à la ligne de départ. Afin d'arrimer les protéines toutes au même niveau dans le gel de concentration, le tampon de concentration contient de la glycine à pH 6.8. À ce pH, les ions glycinates n'avancent pas car la glycine est peu ionisée. Les protéines sont donc ralenties et arrêtées au niveau de la ligne de départ. Le tampon de migration permet de faire migrer les protéines le plus loin possibles sur la partie du gel de séparation.

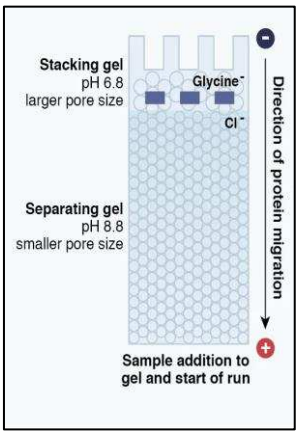


Figure 17: Principe du gel SDS-PAGE
(Source image : Westermeier R. 2016*)* [97]

Le gel de séparation a un pH de 8.8, dans ces conditions, la glycine est chargée négativement, le nombre d'ions électrolytes augmente et les ions migrent vers la l'électrode positive en entrainant les protéines.
Le gel SDS-PAGE peut aussi être utilisé comme une technique de purification des protéines étant donné que chaque bande correspond à une protéine isolée que l'on peut extraire et digérer. C'est ainsi qu'on peut coupler le SDS-PAGE à des techniques plus sensibles comme la spectrométrie de masse. En d'autres termes, l'électrophorèse peut être utilisée en phase préparatrice.
Toutefois, la technique SDS-PAGE a des problèmes de reproductibilité, du fait des nombreuses étapes de manipulation. [98]
Pour améliorer la détection, le gel SDS-PAGE doit subir une coloration spécifique aux protéines pour les détecter.

II.2. Coloration au bleu de Coomassie

La coloration au bleu de Coomassie (Figure 18) est une coloration spécifique des protéines qui permet de révéler leur présence. Elle s'applique aussi bien pour teinter un gel SDS-PAGE, que pour doser des protéines en suspension par la méthode de Bradford.

La forme libre du colorant est rouge, tandis que la forme liée à la protéine est bleue.[99] Ceci s'explique par le fait que la forme libre cationique absorbe à 470 nm, et la forme liée anionique absorbe à 595 nm[1]. L'intensité de la coloration est proportionnelle à la concentration totale en acide aminés. Le complexe colorant-protéine a un haut coefficient d'extinction ce qui a permis d'obtenir une haute sensibilité de détection. De plus la réaction de coloration est rapide (2 min).

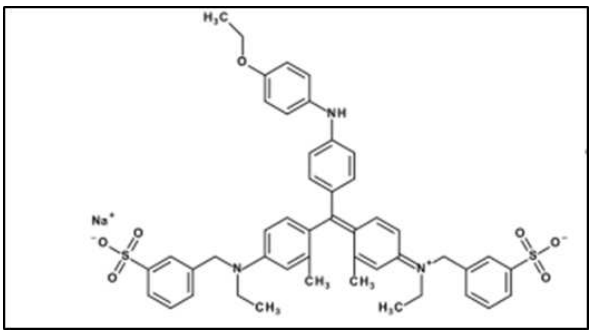


Figure 18: Formule chimique du Bleu de Coomassie

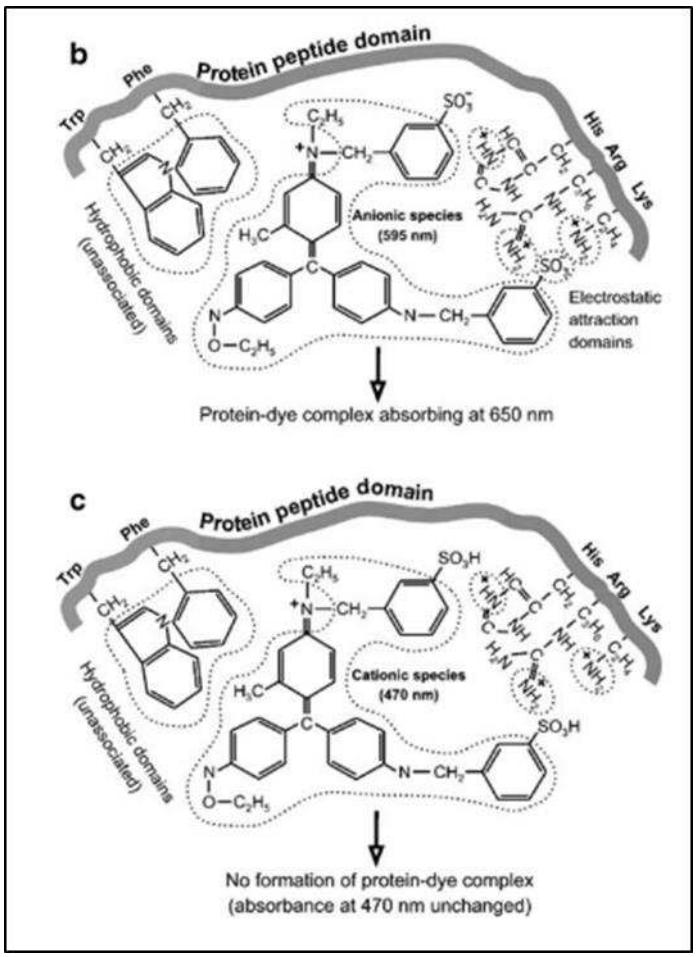


Figure 19: Mécanisme du complexe du Bleu de Coomassie avec la protéine (Source image : Georgiouet. al. 2008) [119]

L'interaction entre le colorant et la protéine résulte de plusieurs interactions avec les acides aminés[100,101], dont l'interaction électrostatique entre les groupements amines de l'arginine et la lysine (Figure 19).Des liaisons hydrogène s'établissent entre ces derniers car ce sont les seuls acides aminés à avoir des groupements amines réactifs à pH physiologique. Il y a aussi des phénomènes de « π-π » stacking entre les cycles aromatiques du bleu de Coomassie et les acides aminés aromatiques tels que la phénylalanine.
La coloration au bleu dye Coomassie a une sensibilité comprise entre 1.5 et 7.5 micromoles de protéine de BSA par bande de gel.[102]

II.3. Coloration au nitrate d'argent

La coloration au nitrate d'argent fait aussi partie des méthodes colorimétriques de détection des protéines. Après avoir fait migrer les protéines sur un gel SDS-PAGE, on immerge le gel dans une solution de nitrate d'argent. S'en suit alors une interaction non spécifique entre les ions nitrates et certains acides aminés des protéines ayant migré. Les fonctions des acides aminés en jeu sont : acide carboxylique (aspartate, glutamate), imidazole (histidine), sulfhydrile (cystéine), amines (lysine). Ces fonctions favorisent l'adhésion des ions argent à la surface de la protéine et sa coloration par précipitation.[103]
La coloration se fait en plusieurs étapes dont le traitement des protéines au formaldéhyde et au nitrate d'argent. Les ions argent sont réduits en argent métallique en présence du formaldéhyde. Cette étape permet la formation d'un complexe coloré (brun sombre) avec l'argent métallique et les protéines.[104] On obtient alors une image de photographie semblable aux photos en noir et blanc faites à l'argentique.[103]
La sensibilité de cette technique est supérieure à celle du bleu de Coomassie, à savoir 3.7-7.5 femtomoles (10^{-15}) de protéines.
En somme, il existe deux techniques majeures de détection basées sur la coloration des protéines. La coloration au Bleu de Coomassie est simple et peu couteuse. En revanche, la coloration au nitrate d'argent est beaucoup plus sensible, mais les étapes sont plus nombreuses et fastidieuses. De plus, la coloration au nitrate d'argent n'est pas compatible avec la spectrométrie de masse contrairement au bleu de Coomassie. En effet, une étape utilise le formaldéhyde qui est incompatible.[105–107]

II.4. Electrophorèse 2D

L'électrophorèse 2D consiste à séparer les protéines selon leur point isoélectrique (pI). Tout premièrement s'opère la focalisation isoélectrique. Les protéines migrent sur une bandelette de gel soumise à un courant électrique. Lorsque le pI est atteint, les protéines ont un état de charge nul, elles arrêtent leur migration sur la bandelette de gel. La bandelette de gel est ensuite placée dans un dispositif classique de SDS-PAGE où les protéines seront séparées selon leur taille (Figure 20).
Les protéines sont dénaturées, et les ponts disulfures des cystéines réduits. Par la suite l'alkylation des thiols à l'iodoacétamide est employée, pour éviter que les ponts disulfures issus des thiols des cystéines se reforment durant le traitement.

En plus d'être utilisée en phase analytique, cette technique peut être utilisée en protéomique, en phase préparative.[108] Ces séparations peuvent permettre de dégrossir les mélanges protéiques complexes. Il en va de même pour le gel SDS-PAGE. Les bandes protéiques sont découpées, puis les protéines sont extraites par des protocoles d'extraction.

Cette technique a moins de problèmes de reproductibilité que la précédente. De plus, on peut combiner l'électrophorèse 2D et la coloration au nitrate d'argent, ce qui permet d'avoir une sensibilité d'approximativement 1 femtomole par protéine et par bande [109].

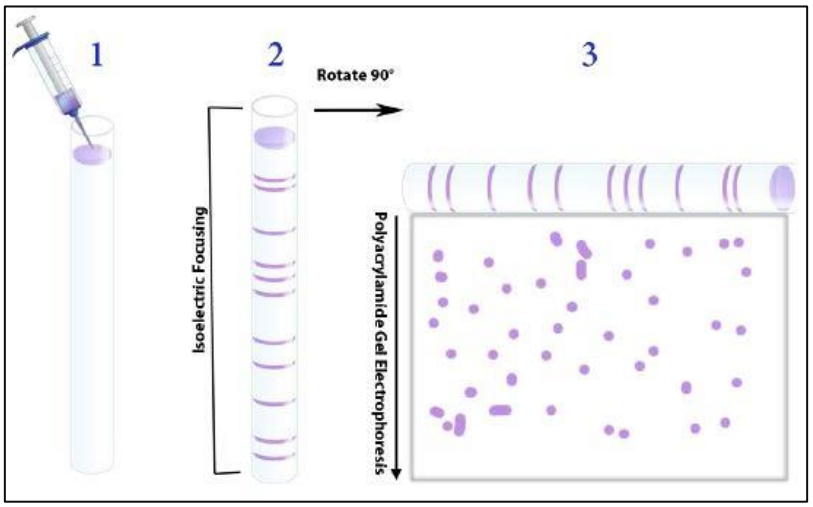


Figure 20: Schéma de l'électrophorèse 2D (Source image : www.creative-proteomics.com)[110]

La focalisation isoélectrique permet aussi de détecter les modifications post-traductionnelles. Par exemple, les acides aminés phosphorylables tels que la sérine, la thréonine et la tyrosine, lorsqu'ils sont phosphorylés, présentent trois tâches qu'on appelle un chapelet correspondant à trois points alignés horizontalement.[111,112] La phosphorylation provoque une différence de pI d'une valeur de 0.2 à 0.3. Cette technique permet donc d'établir quel acide aminé est phosphorylé et permet de donner une estimation sur le nombre d'acides aminés modifiés. Dans cette technique, la modification de la charge des protéines liée aux modifications post-traductionnelles est donc détectable.[113]
En définitive, la technique d'électrophorèse 2D est hautement résolutive, reproductible et permet de séparer une myriade de protéines, cependant la préparation des échantillons et du gel sont délicates. Des étapes de déplétion des protéines majoritaires sont nécessaires, et il est important de bien choisir l'agent chaotrope (détruit la structure tridimensionnelle de la protéine) telle que l'urée ou la thiourée, et d'ajuster le protocole en fonction de la matrice. Cette technique, tout comme les précédentes, est compatible avec la spectrométrie de masse.[114] En effet, les bandes de gel extraites subissent un digestion trypsique avant d'être analysée par la spectrométrie de masse.[113,115]

II.5. Electrophorèse capillaire

L'électrophorèse capillaire est une technique de séparation des protéines basée sur la mobilité électrophorétique. Un capillaire est relié par ses deux extrémités à deux solutions électrolytiques contenant deux électrodes. Sous l'application d'un courant, les protéines vont migrer d'une électrode à une autre selon leur poids moléculaire et leur charge.

La nature du revêtement dans le capillaire doit être bien choisie pour empêcher l'interaction entre la protéine et la phase de du capillaire ainsi que supprimer le flux électroosmotique afin de réduire les problèmes de reproductibilité.[98]Le flux électroosmotique est le mouvement des ions dans un environnement de solvant à travers des canaux très étroits, où un potentiel appliqué à travers les canaux provoque la migration des ions.

L'avantage de cette technique est d'avoir une très haute résolution et une haute définition de la masse moléculaire des protéines. De plus, elle est automatisée et la détection se fait directement sur la colonne de capillaire et non à la sortie.[98] Forte de tous ces avantages l'électrophorèse capillaire est une méthode très utilisée aussi bien dans la biopharmaceutique, que pour des structures plus modestes, car peu coûteuse, avec une instrumentation simple et portable.[116]

II.6. Western blot

La technique de Western blot est la détection de protéines par des interactions immunologiques. Cette méthode hautement spécifique est basée sur la spécificité de reconnaissance d'un anticorps à une protéine.

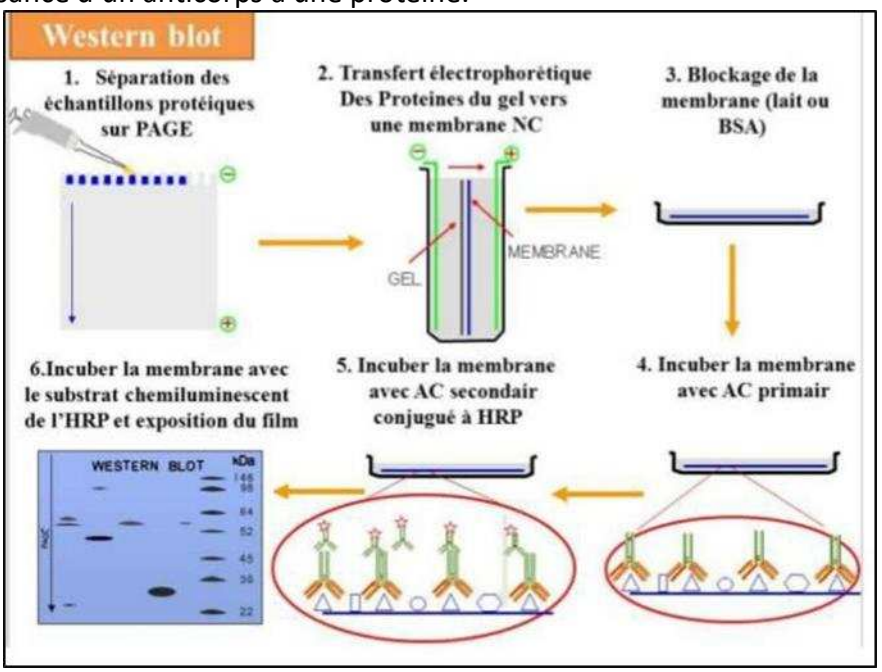


Figure 21: Schéma de la technique de Western blot (Source image : https://slideplayer.fr/slide/18120366/) [136]

La première étape consiste à faire migrer les protéines d'intérêt sur un gel SDS-PAGE. Une fois la migration terminée, les bandes protéiques du gel SDS-PAGE sont transférées sur une membrane de nitrocellulose après l'application d'un courant électrique. On obtient ainsi une membrane de nitrocellulose avec des empreintes qui correspondent aux protéines ayant changé de support (Figure 21). Un anticorps primaire spécifique de la protéine va se fixer à l'épitope, et cette fixation va être révélée par un anticorps secondaire, spécifique du primaire. L'épitope correspond à une partie d'un antigène détectable par un anticorps spécifique. Lorsque l'anticorps secondaire est fixé, une réaction enzymatique (peroxydase HRP) de chimioluminescence se produit, la détection de la fluorescence de la zone fixée se produit par la suite.

Cette méthode laisse place à beaucoup de biais et variabilité d'une expérience à l'autre. Il s'avère qu'il y a de forts problèmes de reproductibilité et de normalisation du procédé.[117]
La sensibilité de cette technique n'est pas normée, mais peut aller jusqu'au femtomole de protéines selon les spécifications commerciales.

II.7. Test ELISA

La méthode d'ELISA (dosage d'immunoabsorption enzymatique) est une méthode immuno-enzymatique basée sur la spécificité anticorps-antigène et la révélation de ce complexe par une réaction enzymatique. Cette méthode est l'une des plus reconnues parmi les méthodes de détection des protéines. Dans cette méthode, la protéine que l'on souhaite détecter est l'antigène. Cette technique fût inventée par Perlmann et Engvall en 1971.[118] Il existe plusieurs types de test ELISA développés dans les années 1980. Parmi les méthodes classiques sont comptées l'ELISA directe, et ELISA indirecte.

La méthode d'ELISA directe peut être divisée en 3 étapes (Figure 22). Premièrement, l'antigène recouvre la multi-plaque par adsorption passive. Deuxièmement, un anticorps conjugué à une enzyme est ajouté. Celui-ci va se fixer à l'antigène car il lui est spécifique. L'ajout de l'albumine de sérum bovin (BSA) va permettre de condamner les sites qui n'ont pas formé de complexe anticorps-antigène.[119] De plus, les étapes alternées avec des séquences de lavages permettent de retirer les composés non liés et les interactions non spécifiques. Enfin, un substrat ou chromophore va être ajouté. L'enzyme conjuguée à l'anticorps va hydrolyser son substrat et permettre une révélation par colorimétrie.

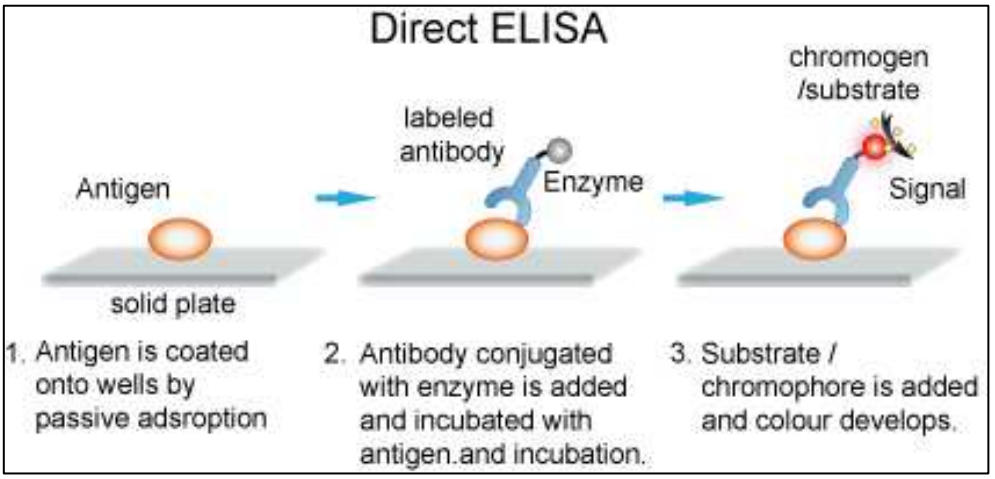


Figure 22: Schéma explicatif de la méthode ELISA directe (Source image : www.creative-diagnostics.com) [120]

La méthode ELISA indirecte fût développée en 1978 par Lindsrtöm et Wager.[121] Cette méthode est semblable à l'ELISA directe excepté qu'il y a une étape supplémentaire avec l'intervention d'un second anticorps. L'anticorps primaire (Ac_I) va se fixer à la protéine d'intérêt (Figure 23). Le second anticorps (Ac_{II}), qui est lui-même conjugué avec une enzyme, va se fixer sur l'anticorps primaire, car ils sont complémentaires.

La méthode d'ELISA directe est de faible sensibilité, tandis que l'indirecte est de haute sensibilité.[122,123] Il se trouve que dans l'étude de Lindsrtöm et Wager la plus faible concentration de protéine détectée avec la méthode ELISA indirecte est de 149 femtomoles.

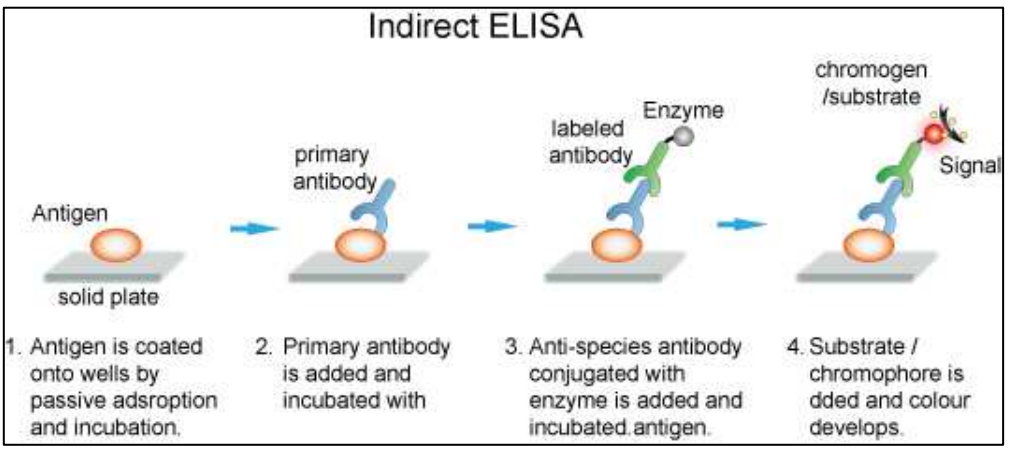


Figure 23: Schéma explicatif de la méthode ELISA indirecte (Source image : www.creative-diagnostics.com) [120]

La méthode ELISA classique présente tout de même quelques inconvénients. Le lecteur optique pour la révélation par colorimétrie est onéreux, donc pas toujours accessible. Il y a un biais expérimental en fonction du manipulateur pour mesurer précisément l'intensité colorimétrique. De plus, la sensibilité va du nanomolaire ou picomolaire alors que, les biomarqueurs des maladies sont au niveau attomolaire à un stade précoce.[124] C'est ce niveau de sensibilité que la recherche clinique vise.

II.8. Nouvelles techniques de détection des protéines

Les techniques développées ces 20 dernières années peuvent maintenant détecter les protéines à un niveau femtomolaire (10^{-15} moles) et attomolaire (10^{-18} moles). Depuis 2007, de nouvelles techniques ultrasensibles sont apparues telles que[125] :

- Les techniques de fabrication à l'échelle micro/nano
- La conception de biosenseurs
- Les nanotechnologies de l'ADN

Ces méthodes semblent, une fois appliquées aux cellules cancéreuses, permettent la détection précoce et le traitement efficace du cancer.[126,127] Les avancées récentes dans les techniques de détection des protéines sont les nanobiosenseurs et les microsenseurs optiques. Elles permettent une détection à plus petites échelles, à savoir, le niveau attomolaire (10^{-18} moles). Ces techniques ultrasensibles sont liées à l'utilisation d'anticorps spécifiques des protéines à étudier.

Méthode ELISA plasmonique

La méthode conventionnelle pour la détection des protéines, ELISA, présente des problèmes de sensibilité et n'atteint pas les 40 picomoles.[128,129] Face à cela, des méthodes d'immuno-essais plasmoniques ont été développées et utilisées dans les domaines du diagnostic médical, de l'environnement, et la santé alimentaire.[130] Contrairement aux autres techniques d'immuno-essais classiques, ces dernières sont très sensibles et permettent une révélation à l'œil nu.

Une des techniques ELISA basées sur l'utilisation des nanomatériaux est l'ELISA plasmonique (p-ELISA), développée par Rica et. al en 2012.[131] Un plasmon est l'ensemble des électrons libres qui oscillent collectivement dans un métal noble.[132] La p-ELISA est plus précisément basée sur le phénomène de résonance plasmonique de surface localisée (RPLS). Dans ce phénomène, les électrons libres sont conduits à la surface du métal noble. Ils oscillent collectivement quand leur fréquence est coordonnée avec celle d'une radiation électromagnétique incidente.[133] Il se trouve que les nanoparticules de métaux plasmoniques sont en or ou en argent, métaux nobles ayant une très bonne RPLS[130]. Ils ont une forte absorption dans le proche-visible. De plus, les nanoparticules d'or et d'argent ont un très haut coefficient d'extinction molaire, qui est 100 000 fois supérieur au colorant traditionnel de tetramethylbenzidine oxidé (TMB), ce qui augmente de surcroît la sensibilité de la technique p-ELISA. En effet, cette technique a pu être utilisée pour détecter l'antigène PSA du cancer de la prostate (Antigène spécifique de la prostate), et la protéine p24 pour le HIV (Virus de l'immunodéficience humaine), et ceci à des quantités extrêmement faibles, de l'ordre de 10^{-18} g.mL^{-1}.[134]

Contrairement à l'ELISA classique, la révélation ne se fait pas par colorimétrie, mais par l'arrêt de la croissance des nanoparticules[125,131]. Un anticorps spécifique (Ac_I) de la protéine d'intérêt est fixé sur un support solide (Figure 24). En présence de la protéine cible, un anticorps secondaire (Ac_{II}) va se fixer dessus. Un complexe « Ac_I-protéine-Ac_{II} » est formé. L'Ac_{II} est conjugué via une ligation chimique à une enzyme qui est la catalase. Cette enzyme a pour substrat le peroxyde d'hydrogène (H_2O_2).[137]

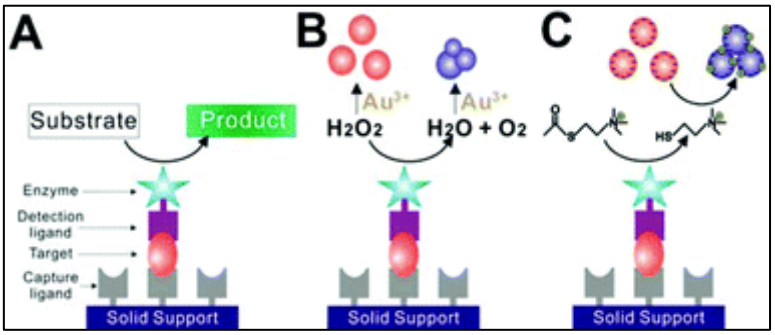


Figure 24: Schéma du principe de la méthode ELISA plasmonique
A. *Principe de l'ELISA directe*. **B**. *En absence de réactivité de la catalase (couleur rouge)*. **C**. *En présence de la catalase (couleur bleue).* **D**. *Principe de l'ELISA plasmonique. Le cercle rouge est la protéine d'intérêt. Le « capture ligand » gris est l'anticorps primaire. Le ligand de détection violet est l'anticorps secondaire. L'étoile verte est l'enzyme (catalase).* (Source image : Yang et. al 2016)[125]

En présence de H_2O_2, les ions d'or sont réduits, le milieu devient rouge (Figure 25). Lorsque le H_2O_2 ainsi que le complexe d'anticorps et l'enzyme sont présents, l'enzyme consomme le H_2O_2. Les ions d'or ne sont donc pas réduits et vont former des agrégats bleus.

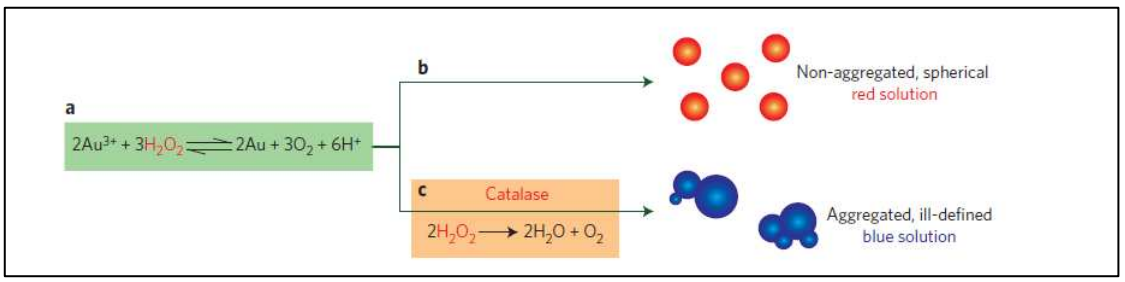


Figure 25: Schéma réactionnel du peroxyde d'hydrogène et la catalase (Source image : Yang et. al 2016)[125]

<u>Tests de proximité médiés par l'ADN pour l'analyse des protéines ultrasensibles</u>
Ces techniques sont basées sur la liaison d'une sonde d'ADN à un ligand d'affinité qui peut se lier à la protéine (anticorps). Un complexe sonde ADN-ligand-protéine est alors formé. La séquence d'ADN de la sonde qui est connue, va être la zone à détecter par PCR (Réaction de polymérisation en chaîne). Le but de la PCR est de détecter la présence ou l'absence d'une certaine séquence d'ADN. Le principe est de synthétiser un nouveau brin d'ADN complémentaire de la séquence de la région d'ADN ciblée puis de l'amplifier par des cycles de réactions de polymérisation en chaîne rendus possible par l'enzyme DNA-polymérase qui a pour rôle la polymérisation de l'ADN. Ceci aboutit à la formation d'une myriade de copies (2^{36}) d'ADN complémentaire de la région cible. La PCR fait ainsi office de révélation, à la place de la colorimétrie. La technique de PCR est schématisée Figure 26.

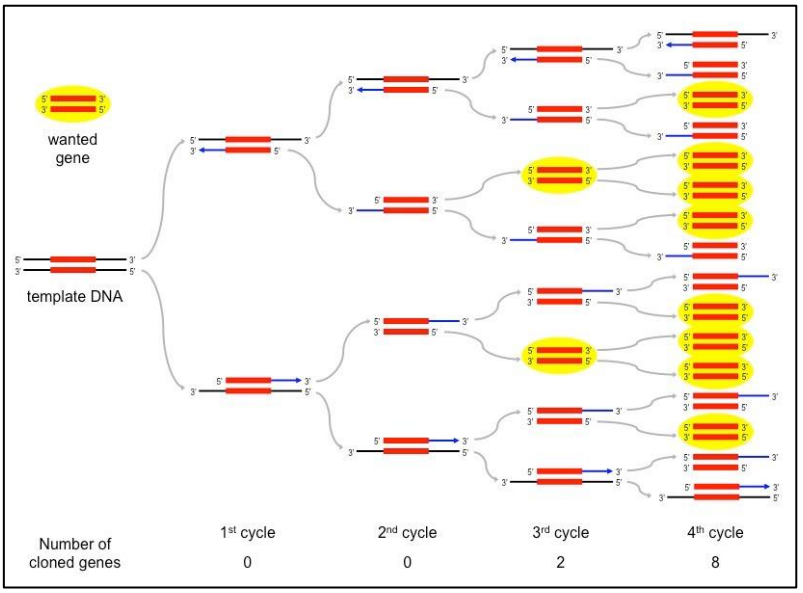


Figure 26: Schéma du principe de la PCR (Source image : https://ib.bioninja.com.au) [135]

La PLA (Essai de ligature de proximité) est basée sur la liaison de proximité entre deux sondes d'ADN, que l'on appelle ligation. La ligation de ces deux sondes se fait au niveau de leurs nucléotides, qui sont des petits fragments d'ADN. Cette méthode passe par la reconnaissance proximale et simultanée d'une molécule cible par des sondes d'affinité, qui concourt à l'apparition d'un signal détectable et amplifié par PCR. Les sondes appartiennent à la classe des aptamères, qui sont des réactifs d'oligonucléotides qui ont une affinité avec la protéine d'intérêt.

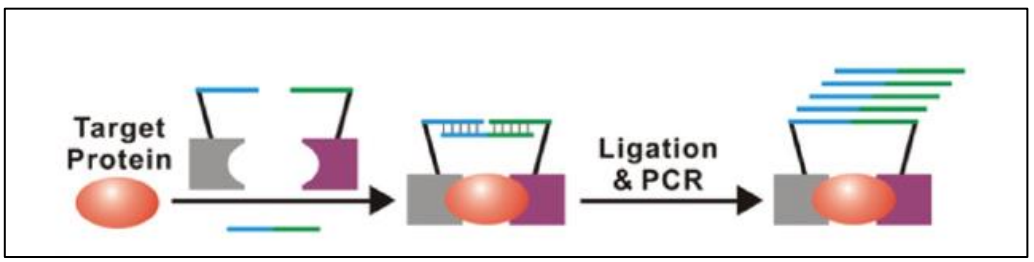


Figure 27: Schéma du principe de la PLA (Source image : Yang et. al 2016)[125]

Un aptamère d'ADN affin de la protéine d'intérêt va être élongé aux extrémités 3' et 5', formant une paire de sondes. Quand les sondes paires se fixent à la protéine d'intérêt, les extrémités libres de leurs séquences sont suffisamment proches pour s'hybrider (s'enchevêtrer sur la longueur par complémentarité de séquence) à un connecteur d'oligonucléotides. Les extrémités devenues proches sont alors liées par une enzyme appelée ADN ligase. La détection de la protéine d'intérêt permet ainsi la réaction de ligation, issue du rapprochement des extrémités libres des sondes paires (Figure 27). Les produits de ligation sont par la suite multipliés par PCR, tandis que les sondes non liées ne présentent aucun signal.[136] La révélation ne se fait pas par colorimétrie mais par PCR fluorométrique. En d'autres termes, elle se fait avec un fluorophore et un quencher présents sur la sonde de PCR, et se produit la PCR en temps réel.[136]

Il a été observé une inhibition de l'activité des ligases, dans le cadre où la PLA a été utilisée sur une matrice de plasma humain.[137] En réponse à cela, Fredriksson et al ont développé en 2011 la PEA (Essai d'extension de proximité), où le nouveau brin d'ADN n'est pas synthétisé par ligation avec une ADN ligase, mais par polymérisation avec une ADN polymérase.[137] La LOD de la technique de PLA est inférieure ou égale à 2 zeptomoles (10^{-21})de protéine.[125,138]

Techniques de miniaturisation

La miniaturisation est le fait d'appliquer l'analyse qui se fait normalement à grande échelle à plus petite échelle, sur une puce. Les avantages de cette technique sont la faible consommation de réactif et d'anticorps. Pour cette technique la LOD est du $ng.mL^{-1}$ au $pg.mL^{-1}$. [139]

La miniaturisation est toujours basée sur l'utilisation d'anticorps, mais on passe à l'utilisation des microplaques 96 puits à des micro-puits dans une micro-puce. Les protéines sont restreintes dans un espace très réduit, ce qui rend le signal plus intense et améliore la sensibilité. Ceci augmente également l'affinité anticorps-antigène. La technique de miniaturisation est aussi une méthode économique, car il y a moins de consommation de solvant, de réactif, et d'anticorps. L'analyse sur puce est une technique biotechnologique à haut débit. Elle permet l'analyse simultanée d'une grande quantité de protéines. Celles-ci sont placées indépendamment les unes des autres dans un réseau de micro-puits.[140]

Parmi les techniques de miniaturisation, on compte les miniaturisations sur surfaces en 2D, ainsi que les miniaturisations sur canaux 3D, sur micro-canaux et micro-puits.

À titre d'exemple de miniaturisation 2D, Zhang et. al ont développé en 2021 une technique appelée OEPM (Micro-réseau de protéine amélioré optiquement). Dans cette technique, des anticorps sont séparés par des espaceurs diélectriques avec un espace de 80 nm, pour renforcer le signal. La technique d'OEPM est une méthode où un sandwich est formé entre l'antigène qui se fixe à l'anticorps Ac_I, lui-même fixé sur une surface métallique (Figure 28). Un anticorps Ac_{II} va reconnaître et se lier à l' Ac_I, la liaison être révélée par fluorescence. En effet, une étiquette fluorescente a été fixée sur l' Ac_{II}. Ainsi, l'antigène (la protéine d'intérêt) sera dosé par mesure fluorométrique avec un microscope à balayage. La sensibilité de l'OEPM est 2 à 10 fois supérieure à celle du kit d'ELISA commercial.[140] Dans l'étude de Zhang et al., la LOD de la méthode OEPM sur 5 biomarqueurs allait de 2.4 à 140 zeptomoles, tandis que le kit ELISA commerciale a une LOD allant de 22 à 370 zeptomoles. De plus, cette technique présente un gain de temps, d'argent et de quantité de réactif. Par exemple, 40 µL d'analyte est nécessaire avec l'OEPM contre 100 µL avec le kit ELISA.[140]

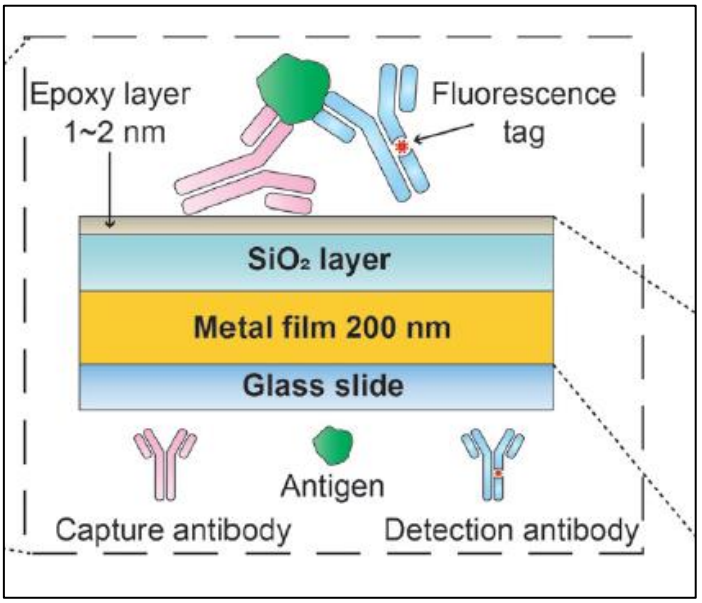


Figure 28: Schéma de la technique de miniaturisation OEPM (Source image : Zhang et. al 2021)[140]

Outre la miniaturisation 2D, il existe aussi la miniaturisation 3D. Dans ce cadre, les interactions sont faites dans des micro-canaux, sur la base de la micro-fluidique. La micro-fluidique correspond à l'étude et la mise en œuvre des écoulements à l'échelle micro dans des réseaux de micro-canaux dont la dimension est de quelques micromètres. En d'autres termes, elle permet le transport de petits volumes à travers une puce. En miniaturisation 3D les anticorps sont aussi utilisés. Les puces utilisées peuvent permettre de faire du « single cell », c'est-à-dire l'analyse des caractéristiques génétiques, transcriptomiques, protéiques, de la métaboliques et des interactions cellule-cellule au niveau d'une cellule unique. De plus, cette technique de miniaturisation réduit le bruit de fond et amoindrit la variabilité entre cellules.[128]

<u>Utilisation de microparticules comme surface de réaction</u>

Ces techniques permettent une plus grande fixation des sondes pour l'extraction des protéines du milieu expérimental.

Dans ce cadre, Lee et.al ont développé une méthode basée sur des microparticules d'hydrogel.[141] Les microparticules de gel poreuses sont fixées aux anticorps primaires. L'échantillon, est mélangé aux microparticules. Un anticorps secondaire va révéler la fixation de la protéine à un anticorps primaire par fluorescence (Figure 29). Cette technique permet l'analyse de différentes protéines d'un même échantillon. La détection se fait par fluorométrie et sa limite permet de détecter un biomarqueur pour une LOD qui va de 1.1 à 537 zeptomoles de protéine détectée. La technique d'ELISA commerciale a une gamme plus large, allant de cette dernière allant de 19.5 à 1250 zeptomoles [128,141].

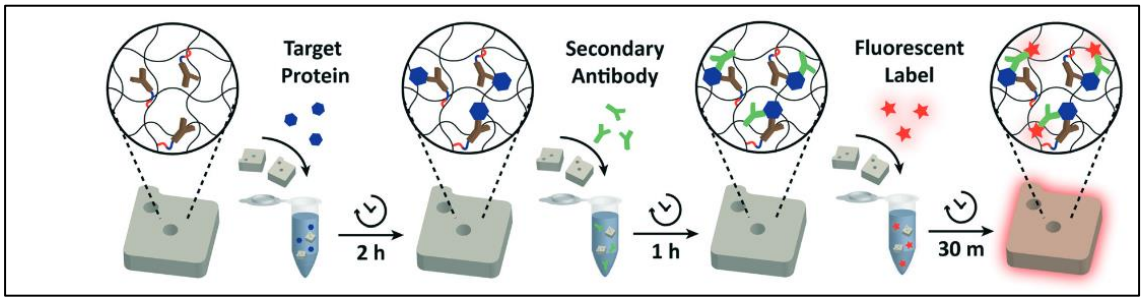


Figure 29: Schéma de la technique multiplexe utilisant des microparticules d'hydrogel fonctionnalisées

III. L'utilisation des sciences multi-omiques pour la cartérisation et détection des marqueurs protéiques.

Les sciences omiques regroupent toutes les disciplines de la biologie qui permettent d'étudier le vivant au niveau moléculaire. Les disciplines de cette science portent toutes le suffixe « -omique », et sont axées sur un type de biomolécules. Parmi elles, on compte :

- La génomique (génome)
- L'épigénomique (l'épigénome)
- La protéomique (protéome)
- La glycomique (glycome)
- La volatolomique (volatolome)
- La métabolomique (métabolome)

L'ensemble des données issues des sciences omiques peut permettre de mieux comprendre le cancer à l'échelle moléculaire.[142] Ces dix dernières années, l'approche multi-omique a permis d'établir des modélisations de comportement des cellules cancéreuses et de l'immunité de la tumeur.[143] Dans ce travail de thèse, protéomique et volatolomique ont été utilisées.

III.1. Analyses « omiques »

III.1.A. Protéomique

La protéomique correspond à l'étude de toutes les protéines présentes dans une cellule, un tissu ou un organisme, à un moment donné et dans des conditions données.

La protéomique permet l'identification et la quantification du contenu global des protéines présentes. La protéomique peut être séparée en trois classes : *bottom-up, middle-down, top down* (Figure 30).

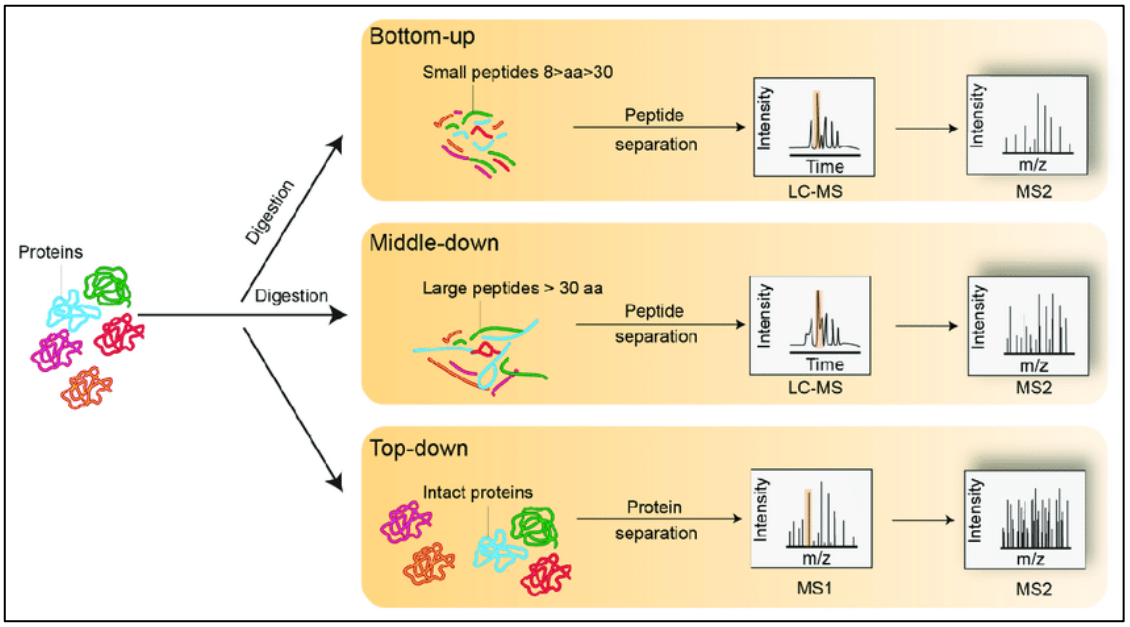


Figure 30: Schéma du concept de la protéomique (Source image : Kennani et al 2018)[144]

Protéomique *bottum up*

La protéomique *bottum up* consiste en la digestion ou hydrolyse d'une protéine pour obtenir des peptides < 3kDa.[145,146] La technique *bottom up* la plus répandue est la digestion trypsique. En effet son prix est peu élevé et les sites de coupure dans les protéines sont nombreux car la trypsine est une endopeptidase qui clive en C-terminal des résidus lysine (Lys – K) ou arginine (Arg – R) s'ils ne sont pas suivis d'une proline (Pro – P). Il y a une fréquence d'environ 6% de ces deux acides aminés dans les chaines protéiques en général.[145] La méthode *bottom up* connaît quelques limites, à savoir le grand nombre de petits peptides générés qui ne sont pas tous analysables dans leur entièreté en spectrométrie de masse.[147]

Protéomique *middle down*

Face aux limites de la méthode *bottum up*, la méthode middle down a été développée. Cette technique aboutie à l'obtention de peptides allant de 13 à 15kD.[147,148] Elle permet d'obtenir une capacité de détection plus importante. Une des techniques *middle down* correspond à l'hydrolyse des protéines au NTCB. La digestion chimique par le NTCB (acide 2-nitro-5-thiocyanobenzoïque) permet de cliver la protéine en *N*-terminal des cystéines (Cys – C). L'intérêt d'utiliser le NTCB est que la taille des peptides générés est plus grande que par digestion avec la trypsine puisque les cystéines sont souvent moins abondantes que les lysines et arginines dans les chaines peptidiques. De plus, cette digestion permet de repérer les cystéines engagées dans des ponts disulfures intra et inter-chaines et donc de détecter les cystéines libres à savoir, les cystéines non engagées dans des ponts disulfure sur la forme native de la protéine.[147]

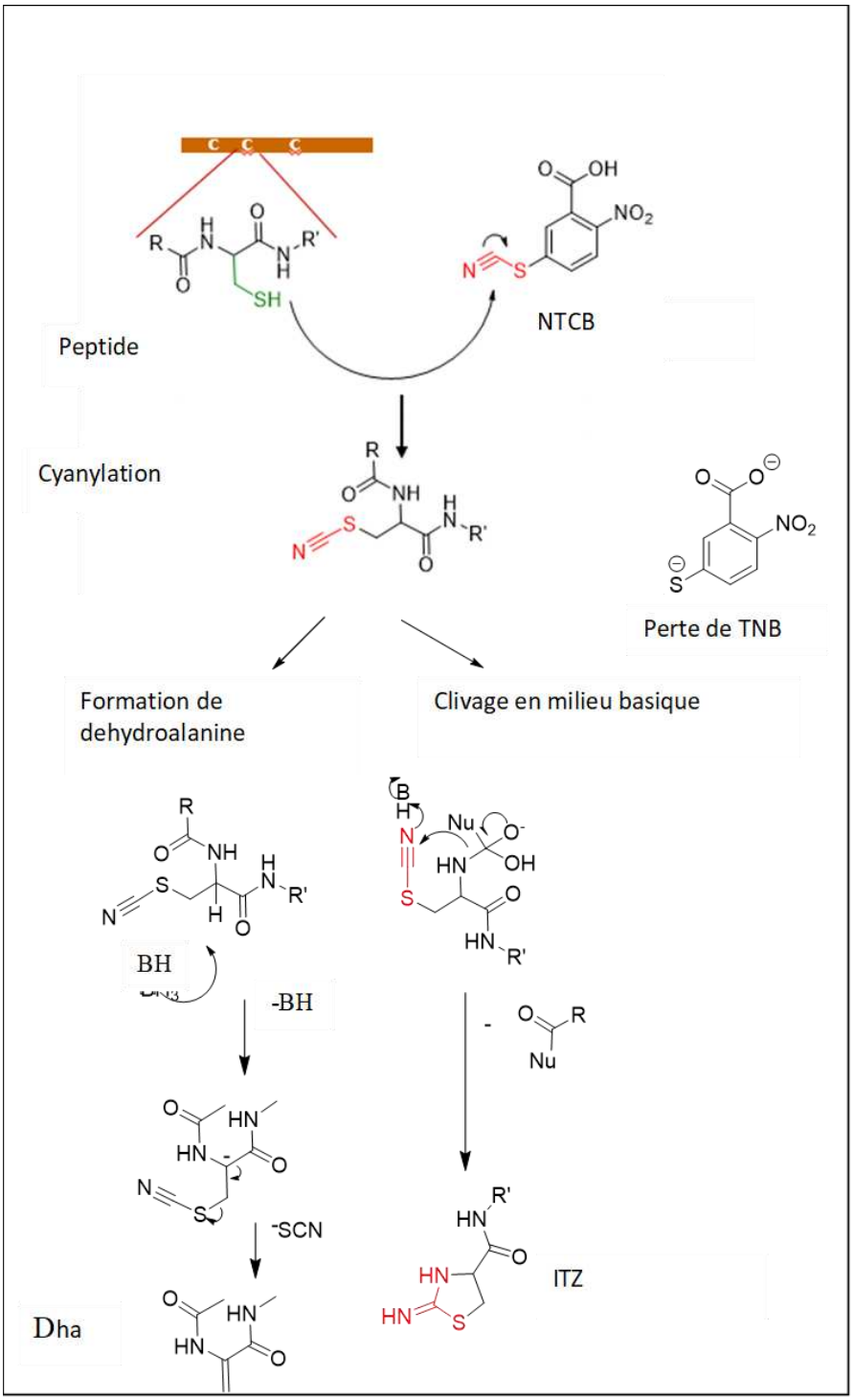


Figure 31: Schéma réactionnel de l'hydrolyse au NTCB

L'hydrolyse NTCB est aussi une méthode de choix pour identifier les cystines (cystéines engagées dans des ponts disulfures) des cystéines libres. En effet, hydrolyser les protéines au NTCB sans faire au préalable une réduction des ponts disulfures, générera des clivages uniquement sur les cystéines libres. Les cystines n'auront pas été réduites, donc ne seront pas clivées.

Concernant le mécanisme de réaction présenté en Figure 31 le thiol du résidu cystéine du peptide réalise une addition nucléophile sur le thiocyanate du NTCB : c'est la cyanylation, causant un départ du thionitrobenzoate (TNB). Ce dernier absorbe le jaune à 412 nm, ce qui permet alors de suivre la réaction.[149]
L'hydrolyse au NTCB aboutit à la formation de deux modifications sur les cystéines clivées après addition nucléophile : l'iminothiazolidine (ITZ) , et la déhydroalanine (Dha). La modification ITZ est celle attendue. Elle résulte d'un réarrangement entre le groupement thiocynanate de l'intermédiaire et le proton d'une base. La déhydroalanination provient de la β-élimination du proton en β du thiocyanate. Cette dernière est non désirée, et son occurrence dépend des acides aminés environnants. Par exemple, les acides aminés avec des chaînes latérales rigides tels que la tyrosine et la proline, connaissent un clivage plus lent et propice à la cyanylation. Une étape peut être rajoutée après le clivage, c'est la suppression du groupement ITZ sur le peptide en utilisant le Nickel de Raney.[150]

Protéomique *top down*
Ce champ de la protéomique est l'analyse de la protéine intacte. Elle permet ainsi l'étude du protéoforme, à savoir, différentes formes d'une protéine produite à partir du génome avec une variété de variations de séquence, d'isoformes d'épissage et de modifications post-traductionnelles.

III.1.B. Analyse des protéines et des peptides par spectrométrie de masse

L'analyse des protéines et peptides par protéomique se fait très régulièrement par LC-MS/MS, c'est-à-dire, chromatographie liquide couplée à la spectrométrie en tandem. Il existe aussi la nano-LC-MS/MS, pour laquelle le diamètre de la colonne LC est réduit, permettant ainsi d'avoir une meilleure sensibilité et efficacité de séparation.[151,152] Différentes combinaisons sont possibles entre la nature de la chromatographie liquide et l'analyseur du spectromètre de masse. Seront présentés présentement les différents types d'analyseurs utilisés dans mon projet de thèse.

Technologie orbitrap

Le fonctionnement du Q-Exactive (Thermo, ville, pays) qui combine un quadripôle, fonctionnant comme un filtre, avec un orbitrap qui assure la détection à haute résolution des ions, disponible au laboratoire, sera plus particulièrement détaillé (Figure 32).

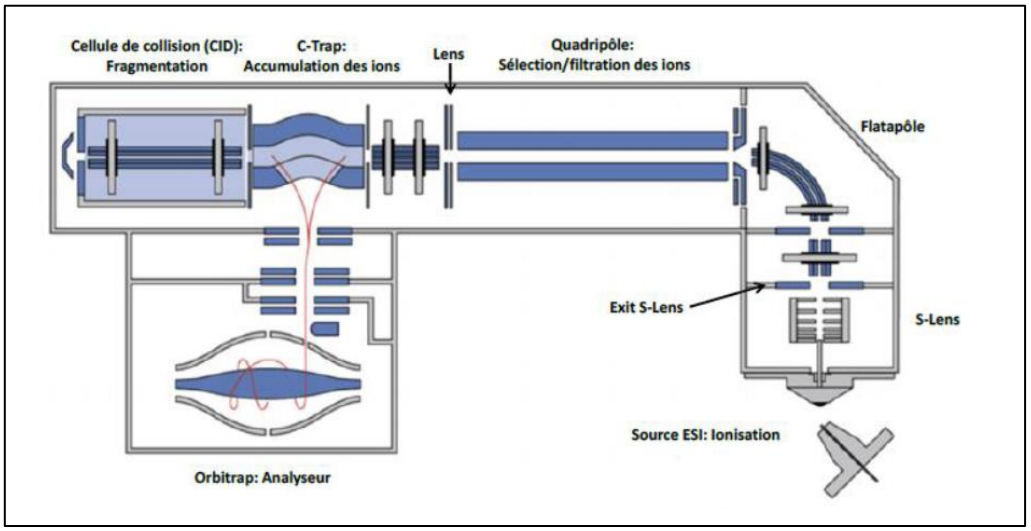


Figure 32: Représentation du Q-Exactive (Thermo Fisher®) (Source image : https://www.smbp.espci.fr) [153]

Dans cet instrument, les ions multichargés produits au niveau de la source ESI (Ionisation électrospray) vont être dirigés via des lentilles (S-lens). Ces lentilles permettent de focaliser et d'accélérer les ions qui sont ensuite transmis au flatapôle via une lentille Exit S-lens. Le flatapôle, constitué de quatre barres plates, permet l'élimination des composés non chargés et des gouttelettes de solvant. Les ions rentrent ensuite dans le quadripôle où ils vont être filtrés en fonction des paramètres de masse.
Ils sont ensuite accumulés dans une C-trap puis acheminés vers l'orbitrap pour y être analysés. L'orbitrap a été développé dans les années 2000 par Makarov.[154] Il est constitué d'une électrode creuse dans laquelle est placée une deuxième électrode. Les ions injectés dans cet analyseur vont tourner autour de l'électrode centrale à une fréquence dépendant du rapport m / z de l'ion.[155,156] Le courant induit par l'oscillation des ions entre les électrodes est amplifié et converti par transformée de Fourrier permettant d'obtenir le m / z et l'intensité de l'ion. Les ions peuvent également être envoyés dans une cellule de collision afin d'être fragmentés. Ce spectromètre peut présenter une résolution de 1000000 voire 2000000[157], et est caractérisé par une précision de masse mesurée inférieure à 1 ppm. Il peut ainsi différencier deux ions avec des m / z de 400,04965 et de 400,04930 (0,00035 Da). La LOD de l'Orbitrap va de 0.3 à 19.2 ng.mL^{1} de protéines.[158]

QTOF (Time of Flight)

Le Q-TOF est un spectromètre de masse hybride constitué de deux analyseurs, un quadripôle (Q) et un analyseur à temps de vol (TOF), séparés par une cellule de collision. Les ions formés dans la source ESI sont d'abord sélectionnés dans le quadripôle avant d'être fragmentés dans la cellule de collision. Les fragments ainsi obtenus sont focalisés au moyen d'un hexapôle et d'une série de lentilles. Les ions sont ensuite impulsés perpendiculairement à leur direction de départ et sont transmis par paquet vers le TOF (Figure 33). Le couplage quadripôle et TOF permet de gagner en termes de résolution (22,500 pour un m/z égale à 956) et de précision sur la mesure de la masse (< 1 ppm).

Le principe du TOF repose sur la mesure du temps de vol d'ions, préalablement accélérés par un champ électrique, dans une zone libre de champ (tube de vol) avant d'atteindre le détecteur. Le TOF est composé de deux régions dans lesquelles règne un vide poussé : une région d'accélération des ions sous l'action d'un champ électrique et une région libre de champ. En mode linéaire, les ions sont transmis dans le tube de vol avec une énergie cinétique qui dépend de la charge (z) de l'ion et de sa masse (m). Cela a pour conséquence de limiter les capacités du TOF à séparer des ions avec des rapports m/z proches. Néanmoins, le développement de l'extraction retardée et du mode réflectron ont permis d'homogénéiser les vitesses des ions lors de leur trajet ce qui se traduit par un gain de la résolution du TOF.

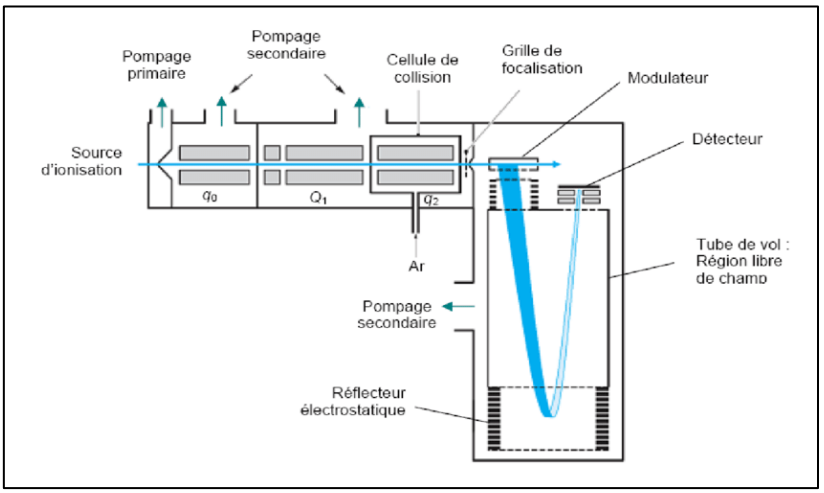


Figure 33: Représentation de fonctionnement du QTOF (Source image : Mesmin C, 2012) [159]

La limite de détection du QTOF va de 28.7 à 881.5 femtomoles de métabolites (glutamate, aspartate, glutamate, phenylalanine, Pooled ribose- and ribulose-5-phosphate).[160]

<u>Triple Quadripôle (TQ)</u>

Des spectromètres de masse en tandem (MSMS), comme le Triple Quadripôle (TQ) sont également très utilisés en protéomique *bottom up* pour quantifier les protéines (Figure 34).

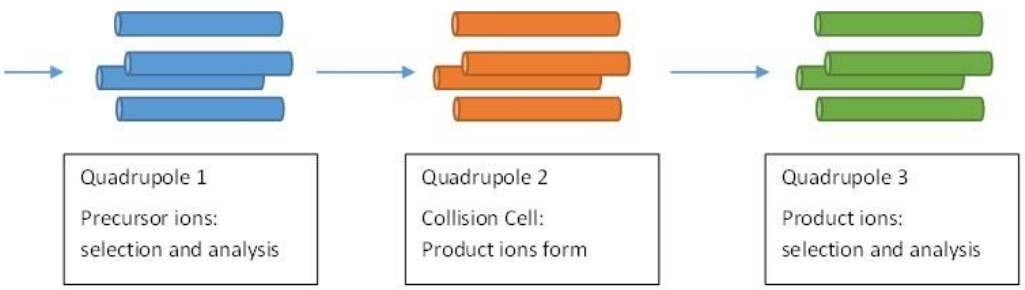


Figure 34: Représentation du fonctionnement du triple quadripôle (Source image: Competency F, 2017) [161]

Les molécules ionisées appelées ions précurseurs, vont être sélectionnées dans un premier quadripôle (Q1) puis entrer dans le second quadripôle (Q2) qui agit comme une cellule de collision. Les ions précurseurs subissent une série de collisions *via* un gaz inerte entrainant la formation d'ions fragments. Les ions fragments produits sont sélectionnés dans un troisième quadripôle (Q3). Ce mode d'acquisition spécifique au TQ est appelé SRM (Monitorage des réactions sélectionnées). Le triple quadripôle est un analyseur hybride de faible résolution (7500 pour un ion à 508 m/z) et faible précision (5 ppm) mais qui peut analyser des ions pour une gamme de masse assez étendue (10 – 3000 m/z). La limite de détection va de 0.02 à 0.58 $ng.mL^{-1}$ des médicaments suivants : ivermectine, doramectine et moxidectine.[162]

III.2. Volatolomique induite

Avant-propos

Alors que nombreuses revues ont synthétisé l'ensemble des travaux réalisés dans le domaine de la volatolomique, aucun article n'avait alors repris les études impliquant les sondes ou métabolites, convertis ou libérant des COVs exogènes.

En effet, après une lecture approfondie de la littérature, nous avions remarqué que nombre d'articles démontraient l'intérêt potentiel des COVs endogènes pour le diagnostic et pronostic des maladies. Revues inclues, ces articles décrivaient essentiellement les différents métabolites volatils marqueurs de maladies telles que le cancer. En outre, ils détaillaient les outils analytiques alors développés ou qui devraient être développés afin de détecter, identifier et quantifier de façon sensible et robuste ces composés.

En revanche, les quelques revues reprenant les travaux sur les COVs exogènes, utilisés comme marqueurs de maladies, ciblaient essentiellement une pathologie et/ou un métabolite volatil donnés. A titre d'exemple, selon la base de données « Web of Science » du 2 Septembre 2023, on compte 258 articles de revues portant sur la détection de l'infection à Helicobacter pylori via la quantification de $^{13}CO_2$ dans l'haleine des patients.

Ainsi, les études développant des sondes et/ou métabolites générant un COV à la suite d'une activité enzymatique ou un processus métabolique, n'avaient jamais été regroupées dans une même publication. Nous avons alors proposé à l'édition de Nature Reviews Chemistry d'écrire une revue reprenant l'ensemble de ces travaux, et de les identifier sous un nouveau terme, la « Volatolomique induite » ou « Induced Volatolomics ».

La revue présentée ci-après, intitulée « Induced Volatolomics of pathologies », définit tout d'abord le concept de volatolomique induite, au regard de la volatolomique dite « classique ». Après un bref repositionnement dans un contexte de diagnostic et de pronostic médical, les différentes sondes, leur mécanisme de conversion en COV exogène, ainsi que les différentes cibles biologiques, y sont présentés.

Check for updates

Induced volatolomics of pathologies

Fabiola Djago, Justin Lange and Pauline Poinot ✉

Abstract | Volatolomics allows us to elucidate cell metabolic processes in real time. In particular, a volatile organic compound (VOC) excreted from our bodies may be specific for a certain disease, such that measuring this VOC may afford a simple, fast, accessible and safe diagnostic approach. Yet, finding the optimal endogenous volatile marker specific to a pathology is non-trivial because of interlaboratory disparities in sample preparation and analysis, as well as high interindividual variability. These limit the sensitivity and specificity of volatolomics and its applications in biological and clinical fields but have motivated the development of induced volatolomics. This approach aims to overcome issues by measuring VOCs that result not from an endogenous metabolite but, rather, from the pathogen-specific or metabolic-specific enzymatic metabolism of an exogenous biological or chemical probe. In this Review, we introduce volatile-compound-based probes and discuss how they can be exploited to detect and discriminate pathogenic infections, to assess organ function and to diagnose and monitor cancers in real time. We focus on cases in which labelled probes have informed us about metabolic processes and consider the potential and drawbacks of the probes for clinical trials. Beyond diagnostics, VOC-based probes may also be effective tools to explore biological processes more generally.

The human body has approximately 40 billion cells, in which innumerable chemical reactions occur[1]. Through integrated processes, cells generate a wide variety of metabolites, ranging from macromolecules like proteins to small molecules like volatile organic compounds (VOCs). VOCs are low-molecular-weight (M < 350 g mol^{-1}) compounds that readily convert from the condensed into the gas phase at 25 °C and 760 mmHg. These molecules can be categorized into different classes depending on their volatility, chemical family, origin and source.

Each of us has an odour fingerprint because we continuously emit more than a thousand VOCs. These biogenic compounds, termed the volatolome, are released in exhaled breath (≈900 VOCs), skin emanations (≈550), urine (≈280), saliva (≈350), faeces (≈400) and blood (≈150)[2,3]. The study of the volatolome is termed volatolomics, and, although the field encompasses all living systems, there has been a particular focus on the human volatolome because it can deepen our understanding of biological processes within us in real time[4,5].

If canine olfaction-based studies[6–8] are anything to go by, volatile molecules can serve as chemical tools for assessing human health status in real time in a non-invasive way. Originally, volatolomics was applied for emergency, rescue and forensic purposes[6,9,10]. Following these studies, large-scale analyses revealed volatile molecules that could potentially serve as clues for detecting and monitoring pathophysiological events in the body (Fig. 1). In comparison with markers based on macromolecules like nucleic acids or proteins, endogenous VOCs are safer and can be sampled non-invasively (at least for breath, sweat, sebum or urine), quickly and as often as necessary, thus facilitating their transfer to the clinic.

No volatile compound has yet proven to be of sufficient diagnostic value so far. This can be explained in terms of both the high interindividual variability[11–14] and individual daily variation[15,16], both of which can afford opposite results from one study to another[17–19]. Indeed, variables such as overall health status, diet, medication intake, exercise and smoking habits can induce qualitative and quantitative changes in the patient volatolome that are unrelated to pathologies[20,21]. In vitro studies have sought to investigate the metabolic pathways that generate VOCs, and the analysis of VOCs from either cultures[22,23] or individual cells[24] has revealed markers for particular phenotypes[25–27] and cancer genotypes[24]. In addition to the variability described above, the field lacks standardized protocols for sample collection, preparation and analysis, which may also influence conclusions in the literature[28–30]. Given the very low concentrations of VOCs in bodily fluids and gases (~10^{-11} M in blood and 0.02 ppb in breath[20]), sampling techniques and analyses must be carefully selected to obtain reliable quantification[31–34]. However, recent developments in colorimetric sensor arrays[35,36], electronic noses[37], array-based 'chemical noses'[38,39], biochemical fluorimetric gas-imaging systems[40] and machine-learning pipelines[41,42] should probably help correlate endogenous

University of Poitiers, UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), E.BiCoM Team, Poitiers, France.

✉*e-mail: pauline.poinot@univ-poitiers.fr*

https://doi.org/10.1038/s41570-020-00248-z

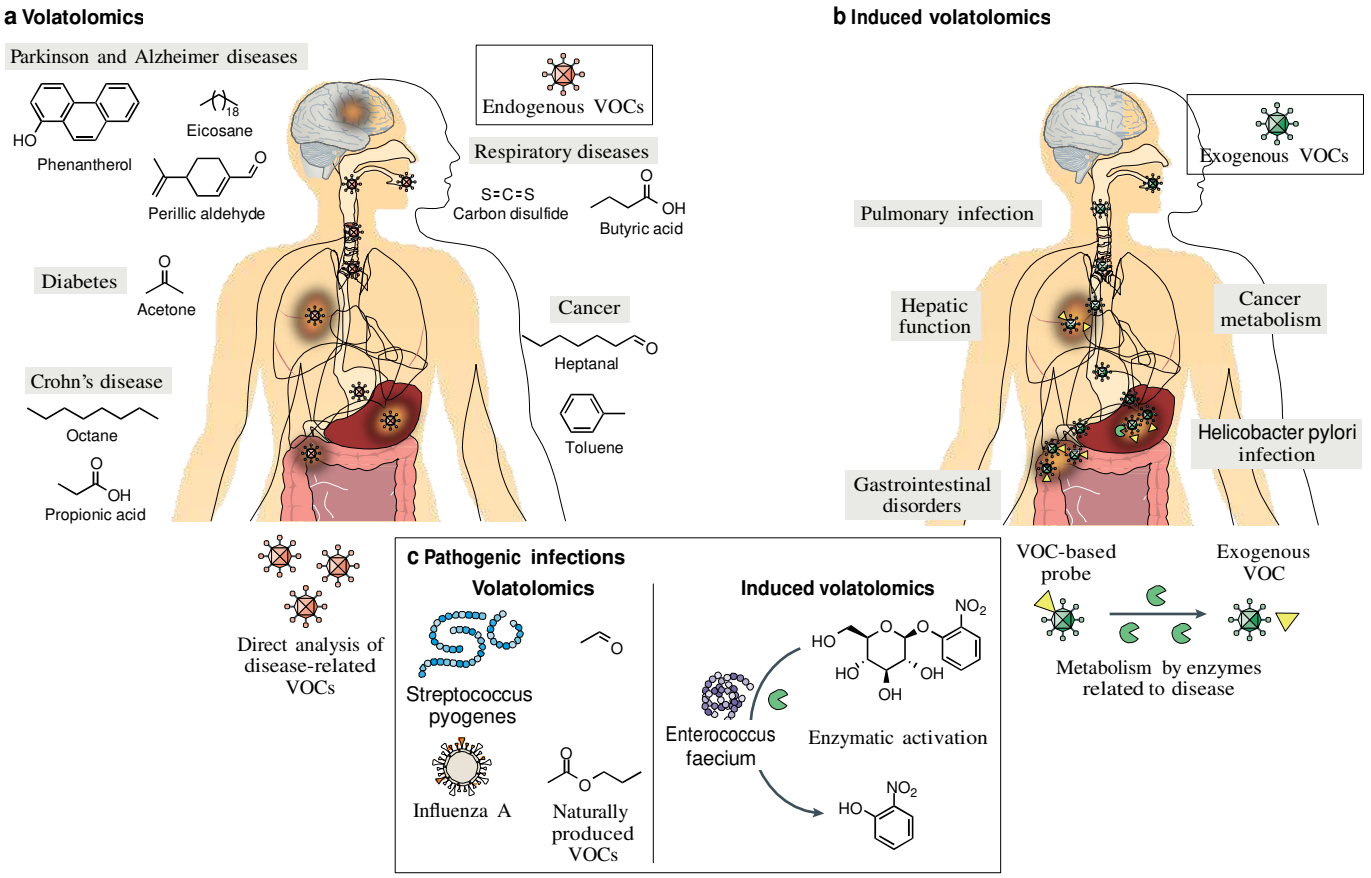


Fig. 1 | **Volatolomics and induced volatolomics. a** | In vivo volatolomics studies involve trapping and analysing each volatile organic compound (VOC) released, to identify the most characteristic volatile markers of a disease. **b** | In vivo-induced volatolomics involves dosing patients with off–on probe molecules. Metabolic processes specific to the suspected pathology convert the molecules into exogenous volatile compounds released in exhaled breath. Most probes exploit the activity of a disease-associated or pathogen-associated enzyme. **c** | In volatolomics, VOCs naturally produced by microorganisms are released into the headspace of cell cultures. VOCs are trapped and analysed. In induced volatolomics, VOC-based probes targeting specific microbial enzymes are added to the culture media. The volatile molecules, released in the sample headspace after probe hydrolysis, are trapped and analysed[57,59,60].

markers with the metabolic disorders responsible for their production. Indeed, volatile compounds specific to numerous pathogenic bacteria such as *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Mycobacterium tuberculosis*[43–45] or viruses like syncytial virus and influenza A virus[46,47] have already been identified. Likewise, endogenous VOCs emitted by patients suffering from a wide range of pathologies have been investigated and potential markers identified for diabetes[48], cancer[31,49–51], respiratory diseases[52,53], Crohn's disease[54], Parkinson disease[55] and Alzheimer disease[56].

In parallel to the study of endogenous VOCs, another novel field within volatolomics has emerged. Termed 'induced volatolomics', this relies on monitoring exogenous volatiles generated by the metabolism of an exogenous probe (Fig. 1). Whereas exogenous compounds are usually considered contaminants that distort the physiological information provided by endogenous markers, here they are used as chemical tracers of biochemical activities characteristic of an organism. This approach has already been applied to detect the pathogenic strains *Enterococcus faecium*[57], *P. aeruginosa*[58], *Salmonella*[59] and *Listeria monocytogenes*[60]. Further, this strategy has been applied to assess liver and gastrointestinal tract functions[61] or monitor pulmonary infection[62], *Helicobacter pylori* infection[63] and cancer metabolism[64].

This Review describes key studies published in volatolomics, first introducing the principle of volatile-compound-based probes. Probes will be categorized into three classes: sugars, peptide derivatives and labelled small molecules. Their reaction mechanisms and medical or fundamental applications are then detailed. From proof-of-concept studies to clinical breakthroughs, we illustrate how VOC-based probes could broaden the scope of volatolomics.

Induced volatolomics

Chemical biology affords us many creatively designed tools to monitor metabolic processes in biological systems[65]. Within this framework, smart off–on imaging probes have substantially enhanced our understanding of how live cells function[66–68]. By imaging a wide range of chemical species — from trace ions to protein complexes — with high spatial resolution, these bioorthogonal

Diabetes
A chronic condition associated with abnormally high levels of sugar (glucose) in the blood. insulin produced by the pancreas lowers blood glucose. Absence or insufficient production of insulin, or an inability of the body to properly use insulin, causes diabetes.

Lactose intolerance
A disease found in approximately 70% of the adult world population that arises because of incomplete intestinal lactose absorption. symptoms include colicky abdominal pain, flatulence and diarrhoea.

Small intestinal bacterial overgrowth
(siBO). siBO is characterized by an increased concentration of bacteria (>10^5 colony-forming units per ml) in the upper small intestine. in addition to the absolute number of organisms, the type of microbial flora seems to play an important role in the appearance of signs and symptoms. Although the majority of microbiota in the small bowel are gram-positive organisms, the flora of patients with siBO mainly comprises coliform bacteria and anaerobes. These microorganisms can ferment carbohydrates and may produce toxins that damage the intestinal mucous membrane, affecting its absorptive function.

Scintigraphy
A diagnostic technique in which a 2D picture of internal body tissue is produced through the detection of radiation emitted by a radioactive substance administered into the body.

tools can inform on cellular injuries even inside deep tissues (>1 mm)[69–78]. Therefore, these probes have offered considerable improvements in disease diagnosis and prognosis. However, their use in mammalian models is limited owing to their toxicity, limited stability and hurdles related to their pharmacokinetics and tissue distribution[66,79].

Off–on probes that can be converted into a small volatile molecule[62,64,80,81] could represent efficient tracers and provide an initial snapshot of pathophysiological events in the human body. Easily processed and monitored, the concept of an exogenous volatile organic compound (EVOC) probe was first proposed by clinical scientists at Owlstone Medical (Cambridge, United Kingdom), an industry leader in breath research and technical innovation[81]. Owlstone administered EVOCs in the form of terpenes to one healthy subject, using these molecules as tracers to investigate their kinetics of gastrointestinal absorption, hepatic conversion and clearance. It was thought that the EVOC probes could constitute efficient tools for monitoring global metabolic processes, while increasing the signal-to-noise ratio and reducing interindividual variation. Moreover, in contrast to volatolomics screening strategies that benefit from high-resolution instruments, advanced computational tools and cross-validation models to highlight potential endogenous markers, the VOC-based probe strategy will benefit more from targeted method development.

Different classes of VOC-based probes following the EVOC paradigm have been described. Most of the probes exploit the activity of a disease-associated or pathogen-associated enzyme belonging to the glycosidase, peptidase, oxidase or urease families. Thanks to this targeted action mechanism, the biochemical events that contribute to the progression or regression of a pathology can be emphasized while attenuating any interindividual variability.

Sugars as volatile-compound-based probes

Numerous non-invasive tests involving sugars as volatile-compound-based probes have been developed for translational medicine. Sugars exhibit clinical efficacy for the diagnosis of gastrointestinal disorders, and these studies have been well documented for more than half a century[82]. In the context of VOC-based probes, administering labelled sugars can help inform us about the activities of specific glycosidases associated with cancers or bacterial infection. Although still in their developmental stages, some of these probes are close to being implemented in the clinic.

Experimental procedures usually see breath samples collected in bags or vacutainer tubes at 15–30-min intervals for 3–5 h. A >10–20 ppm rise in H_2 over basal values within 2–4 h after ingestion is taken to indicate sugar malabsorption[82]. H_2 and CH_4 are analysed by gas chromatography–mass spectrometry (GC-MS) instrumentation, of which the QuinTron BreathTracker is a common specialist example. Gas concentrations are reported as parts per million (ppm)[83]. We now describe some examples of diagnostic applications of sugar-containing probes.

Sugars for clinical diagnosis of gastrointestinal disorders. The first work describing an exogenous volatile as a tracer for a metabolic disorder came in 1969 with the use of lactose as a H_2-based probe to diagnose lactose intolerance[84]. Fructose was similarly employed to indicate fructose malabsorption[85]. Both diseases relate to insufficient activity of the digestive proteins lactase (also called β-d-galactosidase) and fructose transporters GLUT5 or GLUT2 on the surface of epithelial cells in the small intestine (Fig. 2a). Patients who are intolerant present substantially lowered protein functionalities[63,86,87]. In the case of malabsorption, the sugars are transported to the colon where they are fermented by commensal (non-deleterious to the host) bacteria, resulting in CO_2 and H_2 gases responsible for the clinical symptoms (Fig. 2b,c). In that case, H_2 comes from hydrogenases, enzymes that are almost exclusively expressed in archaea and bacteria to catalyse the reversible oxidation of H_2 (refs[88–91]).

Lactulose has emerged as a valuable H_2-based probe to diagnose small intestinal bacterial overgrowth (SIBO)[92]. In contrast to lactose and fructose probes, which evidence the dysfunction of specific enzymes, lactulose was proposed as a nutrient source for bacteria that colonize the small intestine of patients suffering from SIBO. Human cells lack the enzymes required to break lactulose down[93], and the disaccharide is fermented into H_2 gas by the microbiota present in the upper intestine of patients with SIBO[94]. Therefore, an early rise in exhaled H_2 after lactulose ingestion would indicate SIBO[82,95]. The same principles have led to glucose also being developed as a probe for SIBO diagnosis[83].

Being safe, simple, rapid and non-invasive approaches, sugar breath tests have high efficacy and enable rapid testing in humans, enabling their transfer to clinical gastroenterology practice. However, these tests have characteristically variable sensitivity, with the proportion of sick patients detected in the population ranging from 17% to 100%. Further, their specificity is poor because the proportion of healthy people whose absence of disease is confirmed by a negative result ranged from 30% to 100%[85,96–98]. These results do not correlate well with independent data from scintigraphy (measuring a radioactive tracer) or genotyping experiments[99–101]. Hence, in a careful diagnostic routine, sugar breath tests must be confirmed by a subsequent biopsy or endoscopic microbial culture[102].

To improve the accuracy of sugar breath tests, researchers have tried to associate the concentration of CH_4 exhaled in the breath of patients after sugar ingestion with H_2 gas measurements [103]. We exhale CH_4 because of the activity of hydrogenotrophic methanogenic flora — strictly anaerobic bacteria (Fig. 2b) that use H_2 to reduce CO_2 according to $4H_2 + CO_2 \rightarrow CH_4 + 2H_2O$ (refs[91,104]). CO and HCO_2^- can also be reduced to give CH_4. Monitoring both CH_4 and H_2 in a patient's breath has highlighted the false-negative results obtained when only H_2 is measured [105–107]. Nevertheless, this hypothesis remains to be confirmed by measuring the two gases in larger populations[108,109] (Fig. 2c).

Although sugars are useful chemical tools for non-invasive diagnosis, breath tests based on these probes suffer from H_2 and CH_4 naturally being part the breath

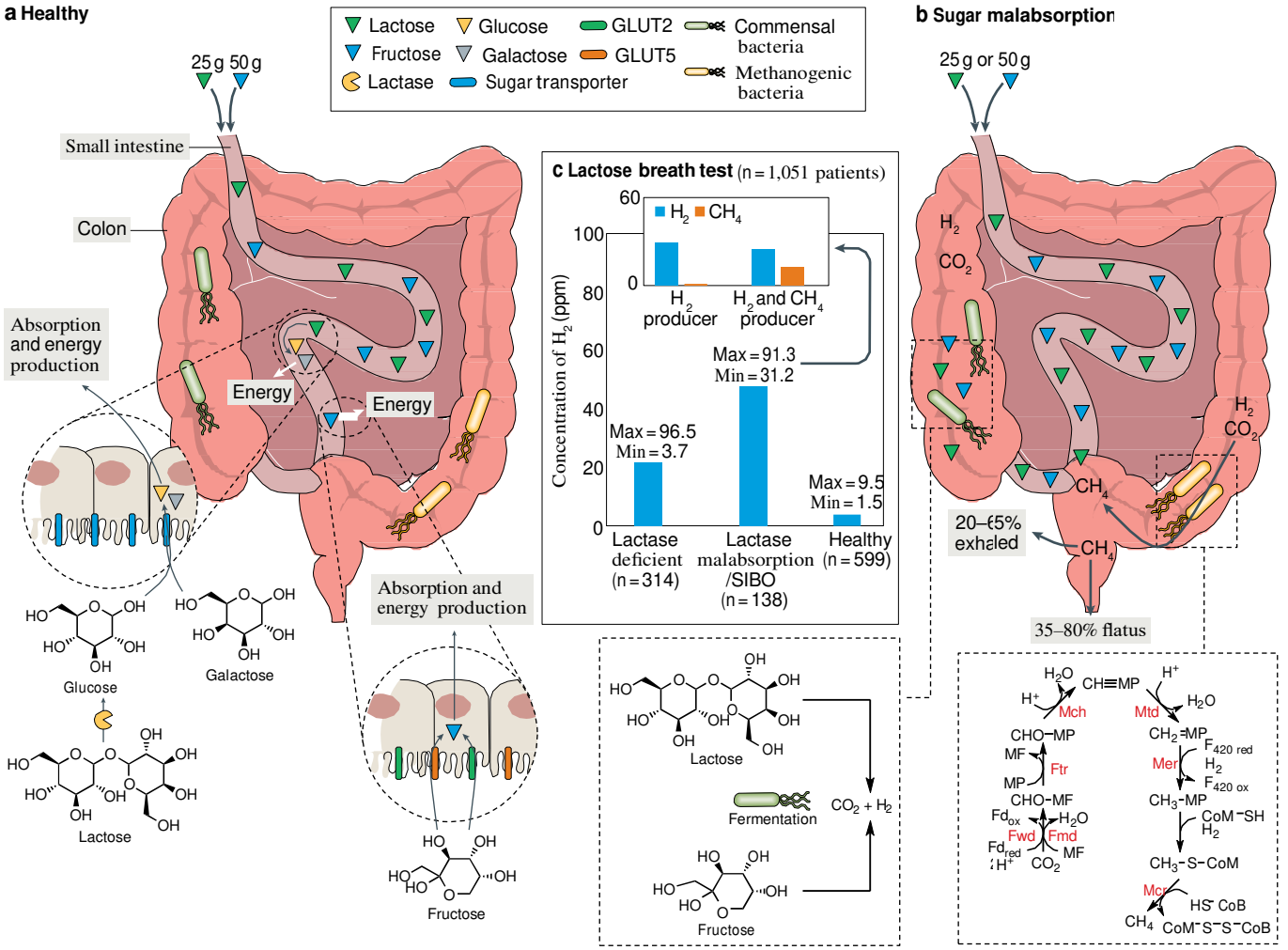


Fig. 2 | **sugar probes for assessing gastrointestinal disorders. a** | H_2-based and CH_4-based probes allow diagnosis of lactose or fructose intolerance in individuals. A dose of 25–50 g of sugars is typically administered to patients. **b** | For intestinal malabsorption, the disaccharide or monosaccharide is transported to the colon, where it is fermented by commensal bacteria into CO_2 and H_2 (refs[91,188]). Both gases could be used by hydrogenotrophic methanogenic flora to produce CH_4 (refs[91,104,189]). As CH_4 is not metabolized in humans, it can serve as an exogenous volatile marker for sugar malabsorption. For both H_2 and CH_4, 20–65% of each gas metabolite is released in the breath [188]. **c** | H_2 and CH_4 levels in the breath samples of healthy patients, patients with lactose intolerance and patients with small intestinal bacterial overgrowth (SIBO) after lactose ingestion[190]. CoB-SH, coenzyme B; CoM-SH, coenzyme M; CoM-S-S-CoB, heterodisulfide; F_{420}, coenzyme F_{420}; Fd_{ox}, ferredoxin; Fmd, molybdenum-dependent formyl-MF dehydrogenase; Ftr, formyl-MF-MP formyltransferase; Fwd, tungsten-dependent formyl-MF dehydrogenase; Mch, methenyl-MP cyclohydrolase; Mcr, methyl-coenzyme M reductase; Mer, F_{420}-dependent methylene-MP reductase; MF, methanofuran; MP, tetrahydromethanopterin; Mtd, F_{420}-dependent methylene-MP dehydrogenase. Part **b** adapted from ref.[91], Springer Nature Limited.

of healthy individuals. Thus, protocols must be standardized to set the optimal dose of probes, with the 50-g lactose dose used in many studies, having since been found to be considered non-physiological[110]. We must also optimize the test duration (2–8 h)[83,100,111] and threshold values (5–20 ppm)[82,95] for a positive test. This latter factor is crucial to discriminate healthy patients from sick patients. Indeed, the habits of either patient class[8] or cross-responses between extra metabolic disorders could increase breath gas baseline values of H_2 and CH_4 (refs[100,112,113]), which is why different recent studies give different reasons for choosing the respective threshold values[107,114]. One way to lower interindividual variability could be to investigate the link between clinical symptoms and patterns of gas production on sugar absorption. For example, recent advances in data-mining techniques[115] or molecular optical probes specific to the dysfunctional proteins[116,117] could be appropriate. Similarly, new preconcentration and analytical methodologies such as fast gas chromatography, which makes use of thinner columns, should help normalize H_2 and CH_4 gas analyses[118]. This instrument features three gas lines that allow exhaled gas to be carried by filtered indoor air through an infrared sensor for CO_2 detection and a chromatography column to separate H_2 and CH_4. Both gases are sequentially detected at a metal-oxide semiconductor. This device enables detection at ppm levels with good chromatographic resolution, thus being highly suitable for the simultaneous

determination of H_2 and CH_4, corrected using the CO_2 concentration in breath.

In contrast to H_2 and CH_4, endogenous $^{14}CO_2$ is undetectable in human breath. Thus, an increase in $^{14}CO_2$ would be easily differentiated from background noise and is the basis for a more consistent breath gas measurement. For these purposes, the efficiencies of radiolabelled sugars such as ^{14}C-lactose and ^{14}C-xylose have been evaluated for their utility in discriminating healthy individuals from patients suffering from either lactose intolerance[119] or SIBO[120,121]. Being less accurate, the ^{14}C-lactose probe has been abandoned in favour of the non-labelled substrate[213], although the ^{14}C-xylose probe for SIBO diagnosis still has proponents. The latter test relies on the fermentation of ^{14}C-xylose by aerobic microbiota present in the small intestine in patients with SIBO. The fermentation gives $^{14}CO_2$, which is collected and easily analysed by scintigraphy. With 95% sensitivity and 100% specificity[120,122], the test has been implemented in the clinic. However, although the ^{14}C-xylose dose was equivalent to 3 weeks of natural radiation exposure, the long decay half-life of ^{14}C and the limited biokinetic data available for humans have prompted hospitals to restrict the routine use of ^{14}C-xylose[123].

Sugars for study of cancer metabolism. Given that radiation concerns are a non-issue with stable isotopes, an induced volatolomics strategy using a deuterated glucuronide as a non-toxic VOC-based probe was proposed for the in vivo diagnosis of cancers[64]. The labelled probe, D_5-ethyl-β-d-glucuronide, was derived from a metabolite of ethanol[124] and designed to target β-glucuronidase, an extracellular enzyme that accumulates in the microenvironment of numerous solid tumours[125–129] (Fig. 3a).

After intravenous administration, D_5-ethyl-β-d-glucuronide converts into D_5-ethanol and glucuronic acid at β-glucuronidases present in the tumour. D_5-ethanol then passes to the bloodstream, is exhaled in the breath, preconcentrated by solid-phase microextraction (SPME) and detected by gas chromatography coupled to high-resolution mass spectrometry. Given the hydrophilic character of the D_5-ethyl-β-d-glucuronide probe, its passive penetration through cell membranes would be limited, preventing its cleavage by β-glucuronidase inside cell lysosomes in healthy tissues. However, to prevent any intracellular activation of the probe, one can thoroughly optimize the amount of probe to be injected into animals. This has let us inject the probe at doses one thousandth ($50\,\mu g\,kg^{-1}$) of those of earlier sugars employed as VOC-based probes.

The present strategy enabled unambiguous discrimination between healthy mice and mice bearing different solid tumours (nasopharyngeal, mammary and pulmonary; Fig. 3b–d). Moreover, this strategy was the first to enable us to monitor either tumour growth or regression during cancer chemotherapy[130] (Fig. 3e), and this convenience and efficacy sees it close to being in clinical pipelines. β-Glucuronidase is present in high concentrations in the microenvironment of almost-solid tumours as soon as they reach 2–3 mm in size[125,126,131,132], making the probe suitable for early detection of tumours. Such sensitivity to small tumours is barely reached with state-of-the-art imaging methods like mammography[133], radiography[134], magnetic resonance imaging[135] and positron emission tomography[136,137]. However, we stress that the VOC-based probe tells us nothing about the spatial distribution of tumours, so this induced volatolomics strategy would have to be only the first diagnostic step before imaging or analysis of blood biomarkers like calcitonin or α-fetoprotein[138,139]. With multimodal and multiscale detection by a radiomic analysis (the data-driven extraction of information from radiographic images)[140], we could address early diagnosis of cancers, many of which have high mortality rates if diagnosed at a later stage.

Aside from the clinical value of D_5-ethyl-β-d-glucuronide, this probe also enables verification of the presence of extracellular β-glucuronidase in tissue samples[64] (Fig. 3f). The results correlate well with the standard immunostaining method, thereby demonstrating the value of this protocol for detecting enzyme activity (Fig. 3g). In the same manner, ^{13}C-glucose has recently been proposed as a VOC-based probe to understand fundamental processes in cancer cells[141]. Treating lung cancer cell cultures with ^{13}C-glucose enables one to track the metabolic pathways of cancer-related volatiles. After the cells reach their metabolic steady state, labelled volatiles released into the sample headspace are trapped by SPME and detected by GC-MS. The 2-pentadecanone synthesis pathway — a metabolic cascade that uses glucose for fatty acid synthesis — is the only process that has been clearly elucidated by this method. However, this bioorthogonal volatile-based probe retains potential as a cutting-edge tool for understanding the metabolic network of airborne metabolites in vivo.

Sugars for in vitro pathogen detection. At the time of writing, culture-based procedures are the state of the art for pathogen detection[142]. Despite their accuracy, these procedures are laborious, taking several days to obtain initial results and up to 1 week to precisely identify the microorganisms responsible. This slowness is a problem because clinical outcomes of patients with infections are strongly related to the early administration of efficient antibiotics[143]. Therefore, there is high demand for alternative methods, such as new strategies combining the polymerase chain reaction and enzyme-linked immunosorbent assays[144,145] with biosensors[142,146]. In parallel, volatolomics approaches to determine pathogen-related VOC signatures have also emerged[43,44]. With the technical and computational advances in volatolomics, metagenomics and metabolomics, we now have a better understanding of what constitutes a 'healthy volatolome'. However, the clinical application of volatolomics to discriminate bacterial infections remains limited because pathogens can borrow similar metabolic pathways to liberate volatiles as other pathogenic or commensal bacteria.

VOC-based probes that target specific microbial enzymes are valuable tools to selectively isolate and differentiate pathogens. Sugar derivatives have been engineered to release an exogenous VOC not naturally emitted by either the bacteria or the host. In this way, one can eliminate analytical background and chemical

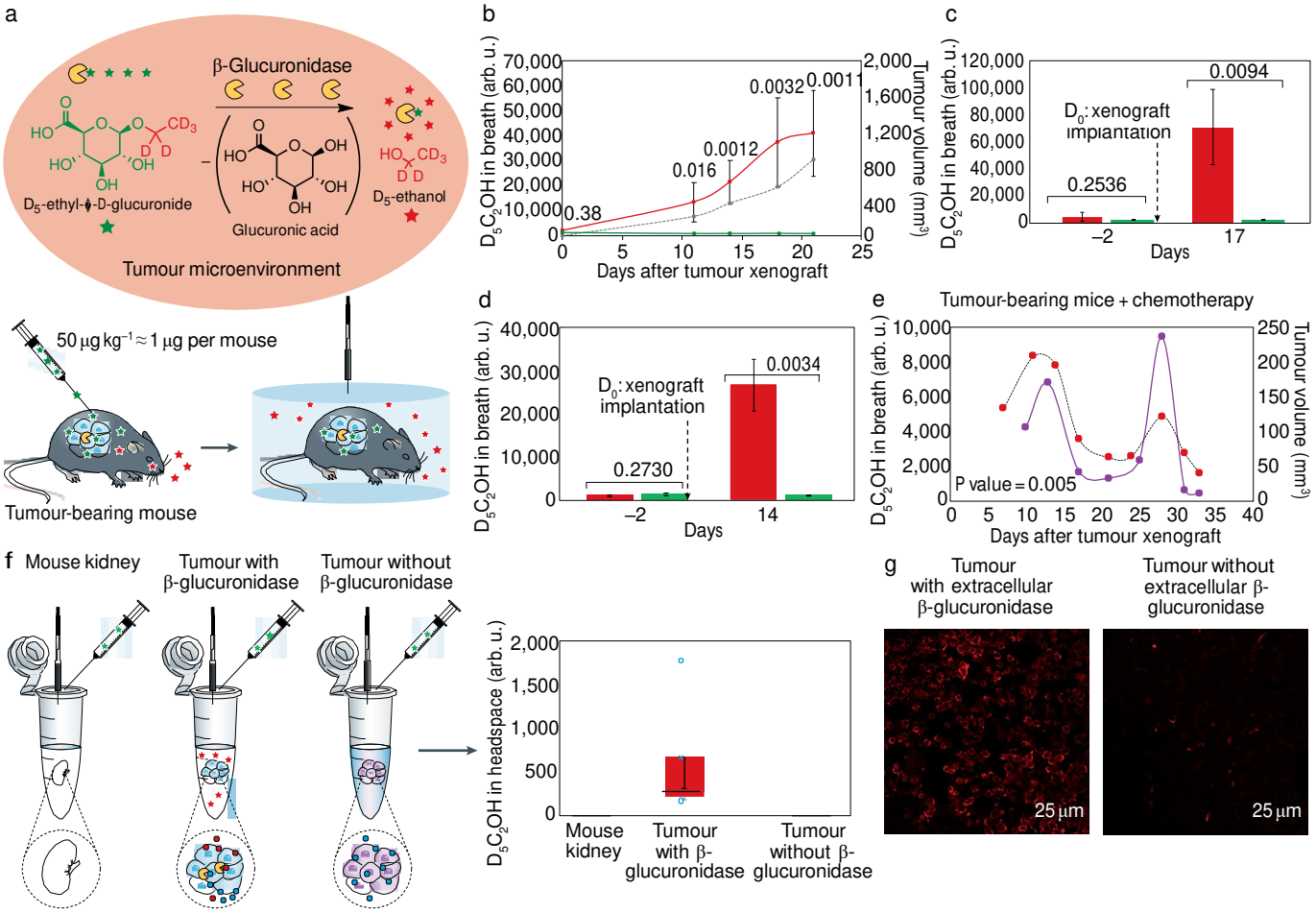


Fig. 3 | **A sugar probe for studying cancer metabolism in vivo and in vitro. a** | D_5-ethyl-β-d-glucuronide probe for monitoring cancer metabolism[64]. Proof of concept in model mice with human nasopharyngeal epidermoid carcinoma (KB subcutaneous xenograft). For discrimination of tumour-bearing animals within a population, a 50 μg kg^{-1} dose of D_5-ethyl-β-d-glucuronide is intravenously administered to the whole mice. In those bearing a tumour, the hydrophilic probe is selectively hydrolysed in the tumour microenvironment by extracellular β-glucuronidase to release D_5-ethanol. D_5-ethanol then passes into the bloodstream and is exhaled in the breath. **b** | Healthy (n = 13) and tumour-bearing animals (n = 13) were discriminated according to the breath signal of D_5-ethanol. Numbers above data points correspond to P values. **c** | Ion intensity for D_5-ethanol exhaled by mice bearing human mammary tumour (MDA-MB-231 subcutaneous xenograft; red bars; n = 8) and healthy mice (green bars; n = 8). Numbers above square brackets correspond to P values. **d** | Ion intensity for D_5-ethanol exhaled by mice with Lewis lung carcinoma subcutaneous xenograft (red bars; n = 6) and healthy mice (green bars; n = 6). Numbers above square brackets correspond to P values. **e** | One can evaluate tumour growth and regression during chemotherapy by treating KB mice with prodrug monotherapy (4 mg kg^{-1})[130] at days 11 and 28. The amount of D_5-ethanol emitted is monitored over time and Spearman's correlation test (P value = 0.005) relates the evolution of tumour volumes (black dotted line with red data points) to the evolution of D_5-ethanol (purple line). **f** | Evidence of β-glucuronidase activity in tissue samples. (o) corresponds to the minimum and maximum signals of D_5-ethanol and (x) to the mean. Four kidneys, four KB xenografts and four tumours of transgenic mice were used. **g** | β-Glucuronidase staining assays. Immunodetection of β-glucuronidase was performed on 7-μm-thick sections with a rabbit anti-β-glucuronidase antibody and a fluorescent secondary antibody. Sections were observed by confocal microscopy. The presence of extracellular β-glucuronidase was revealed by red fluorescence. Parts **a**–**f** adapted with permission from ref.[64], Wiley. Part **g** reprinted with permission from ref.[64], Wiley.

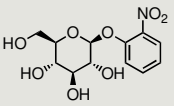
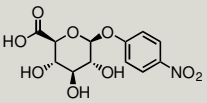
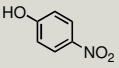
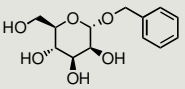
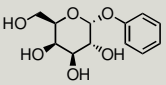
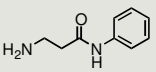
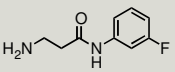
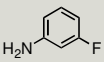
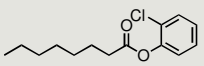
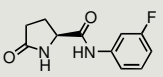
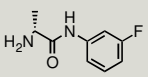
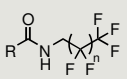
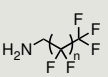
interference. A wide range of these probes have targeted bacterial glycosidases, such as β-glucosidase α-galactosidase, α-mannosidase and β-glucuronidase (Table 1). VOC trapping and analysis methods vary with the main methods being SPME preconcentration, followed by GC coupled to either MS[59] or ion-mobility spectrometry[60]. New methods using agarose gel or functionalized xerogels (dried, shrunken gels) as trapping matrices for volatile compounds have also been proposed. In these methods, VOCs can be detected either with the naked eye (for example, by visualizing yellow 2-nitrophenolate in alkaline solution) or by a colorimetric method (for example, aniline condenses with sodium 1,2-naphthoquinone-4-sulfonate to give a red product)[57,147].

Once a culture sample size has been optimized, a probe tracing the activity of β-glucosidase can confirm the presence of a pathogen and differentiate

it from other species[57]. In particular, the efficacy of 4-nitrophenyl-β-d-glucuronide has been demonstrated on real food samples to indicate the presence of *Salmonella*, a dangerous food-borne pathogen[59]. The exogenous VOC signals here were not perturbed by the sample matrix, which contrasts chromogenic or fluorogenic activity-based probes that may lead to opacity or discolouration. Indeed, the hydrolysis of a chromogenic substrate or the quenching of released fluorophores can interfere with data interpretation[148–150].

Although in vivo studies on animals (including humans) can be envisaged as soon as a probe's safety is

Table 1 | **sugar and peptide probes for pathogen detection**

Assay	Bacterium (Gram) and pathology	VOC-based probes	enzyme target	VOC	trapping and analysis
In vitro Biological samples[57]	Enterococcus faecium (+) Commensal bacteria in the intestinal tract	 2-Nitrophenyl-β-d-glucoside	β-Glucosidase	 2-Nitrophenol	Trapped on xerogels Colorimetric analysis
In vitro Food samples[147]	Escherichia coli (−) Bloody diarrhoea and severe abdominal cramps	 4-Nitrophenyl-β-d-glucuronide	β-Glucuronidase	 p-Nitrophenol	Trapped on functionalized xerogels Colorimetric analysis
In vitro Food samples[60]	Listeria monocytogenes (+) Opportunistic food-borne pathogen	 Benzyl-α-d-mannopyranoside	α-Mannosidase	 Benzyl alcohol	Preconcentrated by static headspace GC-IMS analysis
In vitro Food samples[59]	Salmonella (−) Serious illness in infants, older adults and people with chronic diseases with high mortality	 Phenyl α-d-galactopyranoside	α-Galactosidase	 Phenol	Trapped with HS-SPME GC-MS analysis
In vitro Biological samples[57]	Klebsiella pneumoniae (−) Urinary tract and catheter/urinary tract infections of hospitalized patients	 3-Amino-N-phenylpropanamide	β-Alanyl aminopeptidase	 Aniline	Trapped on xerogels Colorimetric analysis
In vitro Biological samples[58]	Pseudomonas aeruginosa (−) Severe respiratory infections, especially in patients with cystic fibrosis	 3-Amino-N-(3-fluorophenyl) propanamide	β-Alanyl aminopeptidase	 3-Fluoroaniline	Trapped with HS-SPME GC-MS analysis
In vitro Food samples[59]	Salmonella (−) Serious illness in infants, older adults and people with chronic diseases with high mortality	 2-Chlorophenyl octanoate  l-Pyrrollidonyl fluoroanilide	C8-esterase Pyrrolidonyl peptidase	 2-Chlorophenol  3-Fluoroaniline	Trapped with HS-SPME GC-MS analysis
In vitro Food samples[60]	Listeria monocytogenes (+) Opportunistic food-borne pathogen	 d-Alanyl-3-fluoroanilide	β-Alanyl aminopeptidase	 3-Fluoroaniline	Preconcentrated by static headspace GC-IMS analysis
In vivo Mice[62]	Pseudomonas aeruginosa (−) Inflammatory response to respiratory infection	 R = Abu-Oic-Met(O)2-Nle(O-Bzl) PEG4-CKKK(Cy5)-Ac Peptide-coupled hydrofluoroamine	Serine protease neutrophil elastase	 Hydrofluoroamine	Preconcentrated in syringe headspace MS analysis

Each probe undergoes metabolism to release a volatile organic compound (VOC). GC-IMS, gas chromatography–ion-mobility spectrometry; GC-MS, gas chromatography–mass spectrometry; HS-SPME, headspace solid-phase microextraction; MS, mass spectrometry.

demonstrated, this strategy has challenges related to the choice of the enzymatic marker. This is because each of the bacterial glycosidases targeted so far — β-glucosidase, α-galactosidase, α-mannosidase, β-glucuronidase — is present inside human eukaryotic cells. One way to overcome this issue would be to measure enzyme activities belonging to different classes using multiplexed VOC-based probes, molecules that afford VOCs with glycosidases and peptidases. Such a cocktail of probes can provide the much-needed specificity for robust bacterial detection.

Peptides as VOC-based probes

The 5 years up to the time of writing have seen the emergence of a new set of VOC-based probes. These probes feature an exogenous VOC attached to an amino acid or peptide chain through an amide bond. Although only few fundamental studies have described such probes in vitro and in vivo[57–60,62,151], each study highlighted the efficiency of these probes to trace specific peptidases related to bacterial infection. Most of these studies targeted the β-alanyl aminopeptidase enzyme, a biomarker for pathogens such as *P. aeruginosa*, which is involved in cystic fibrosis. Thus, the probes were designed to release an exogenous VOC on hydrolysis by this enzyme (TABle 1), with the measured peptidase activity correlating well with the presence of Gram-negative bacteria[57,58,60]. Since this work, several probes targeting the enzyme have been used to differentiate between species belonging to either Gram-negative or Gram-positive stains[151]. Combined with VOC-based probes tracing the activity of bacterial glycosidases, peptide probes could discriminate between enzyme-producing bacteria[57]. Hence, *P. aeruginosa*[58] and *Salmonella*[59] were detected without ambiguity in biological (for example, sputum) and food (for example, milk) samples.

A new application of the peptide-based VOC approach has allowed us to probe the inflammatory response to respiratory infection[62] (TABle 1). By releasing an exogenous VOC, peptides can trace the activity of serine protease neutrophil elastase, an extracellular protease released from neutrophils in infected lung tissues — white blood cells that contribute to our innate immune response. The protease directly kills microbial pathogens and modulates the inflammatory response by activating and inactivating various cytokines or modulating inflammatory cell recruitment[152]. The VOC-based probes used here consisted of tetrapeptide derivatives with a hydrofluoroamine (HFA; $CF_3(CF_2)_nCH_2NH_2$) appended at their C termini. To prolong peptide retention in the lungs, the probes are incorporated into a volatile-releasing activity-based nanosensor (vABN) with poly(ethylene glycol) arms (Fig. 4a). Hence, once delivered into the diseased lungs, extracellular neutrophil elastase cleaves the peptide and releases a volatile HFA. VOCs were finally exhaled, and breath samples were collected into glass vials. VOC concentrations were quantified using an MS detector.

In a proof-of-concept study, a mouse model with acute lung infection with Gram-negative *P. aeruginosa* was used. Immunohistological staining and confocal imaging enabled the verification of vABN colocalization with extracellular neutrophil elastase in infected lung tissues (Fig. 4b). After intrapulmonary delivery of vABNs, healthy and infected mice were clearly discriminated on the basis of the different levels of HFAs in their breath (Fig. 4c). In addition to the method's diagnostic value, the prognostic efficacy of the bioorthogonal VOC reporters was shown in infected mice treated with an antibiotic such as ciprofloxacin to kill the pathogen (Fig. 4d). A perfect correlation exists between neutrophil elastase levels in lungs and VOC signals in the breath (Fig. 4e,f). The dynamics of both the protein and the volatile markers were further confirmed with the influx of neutrophils to the lungs before and after antibiotic treatment (Fig. 4g). Overall, the specificity of the VOC-based probe was significantly increased, thus avoiding VOC background signals coming from ubiquitous intracellular activation. Although no in vivo toxicity after invasive intrapulmonary injection has been observed, the safety of both the whole nanoparticle and HFA volatiles remains to be studied and must be proven satisfactory before clinical trials can be considered.

Labelled molecules as VOC-based probes

CO_2 and H_2O are the end products of many biochemical reactions. For a long time, CO_2 has represented the optimal volatile marker for in vivo metabolic processes. Isotopic $^{14}CO_2$-based and $^{13}CO_2$-based probes derived from drugs and other small molecules are classic examples in the induced volatolomics field. We now describe some labelled probes that can release labelled CO_2 as part of their metabolism.

Radiolabelled drugs for pharmacology. One of the main reasons to develop radiolabelled drugs has been to investigate liver detoxification processes related to oxidase enzyme activity[153]. Hepatic cells contain cytochrome P450 enzymes (CYPs) in their microsomes (vesicles that form from the endoplasmic reticulum during lab workup) that contribute to almost 75% of drug metabolism[154]. CYPs feature a cysteinate-bound haem active site that is responsible for monooxygenase activity — drugs are converted into metabolites that are often less bioactive and more readily eliminated.

The pharmacological development of a drug, including estimating its clinically relevant dose, requires an investigation of its metabolism, not least by CYPs. In vitro approaches to predict the roles of every CYP have been widely developed but there are often discrepancies between these results and in vivo tests[154]. This motivated the design of non-invasive strategies involving ^{14}C-containing drugs to improve knowledge on drug pharmacokinetics directly in patients. After administration, radiolabelled drugs are either *N*-demethylated or *O*-demethylated by CYPs, affording either $^{14}CO_2$ or intermediates such as $^{14}CH_2O$, which is further converted into $H^{14}CO_2^-$. The latter can be oxidized as part of the methionine-dependent folate pathway to form $^{14}CO_2$, which can be detected by scintillation.

The first biological studies on ^{14}C-containing drugs were performed on animals to evaluate the overall function of CYPs in the presence of inducers or inhibitors. Here, demethylation of ^{14}C-methacethin and ^{14}C-methoxycoumarin was investigated by measuring

Cytochrome P450 (CYP). A CYP is a monooxygenase enzyme with a conserved Phe-$X_{(6-9)}$-Cys-X-gly motif near its C terminus. The Cys thiolate binds a haem cofactor, the Fe site of which splits O_2. One O atom is converted into H_2O, while the other ends up on an organic substrate, with the oxidized product then being released from the enzymatic pocket. High interindividual variation in the activity profiles of these isoforms exist. if an individual is inherently deficient in a CYP or it is inhibited by another drug, toxicity may develop, particularly if drugs accumulate after multiple doses. Drug–drug interactions are a major cause of adverse drug reactions. CYPs themselves can have deleterious effects, converting some chemicals into toxic products that covalently bind enzymes, mitochondrial proteins, rNA, DNA and lipids.

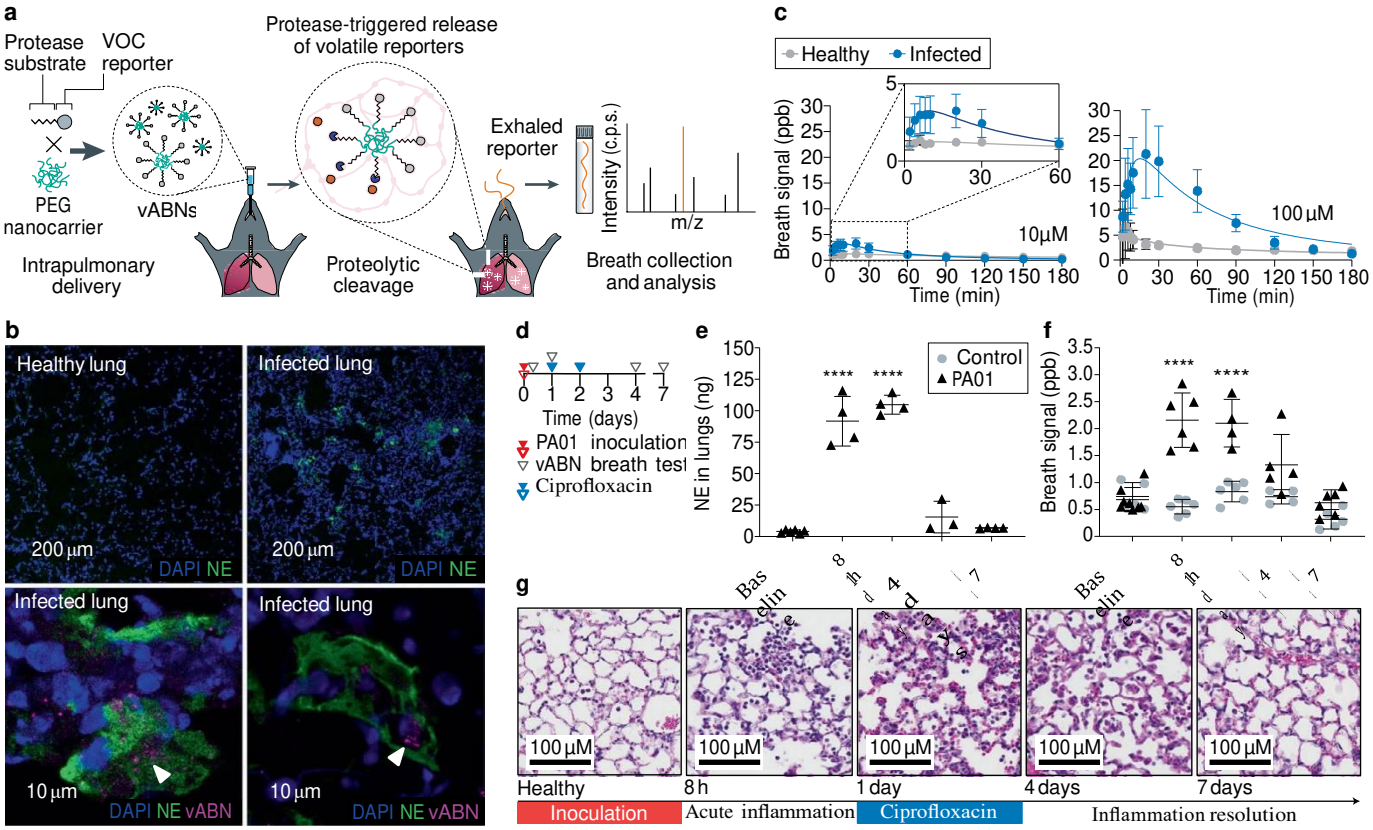
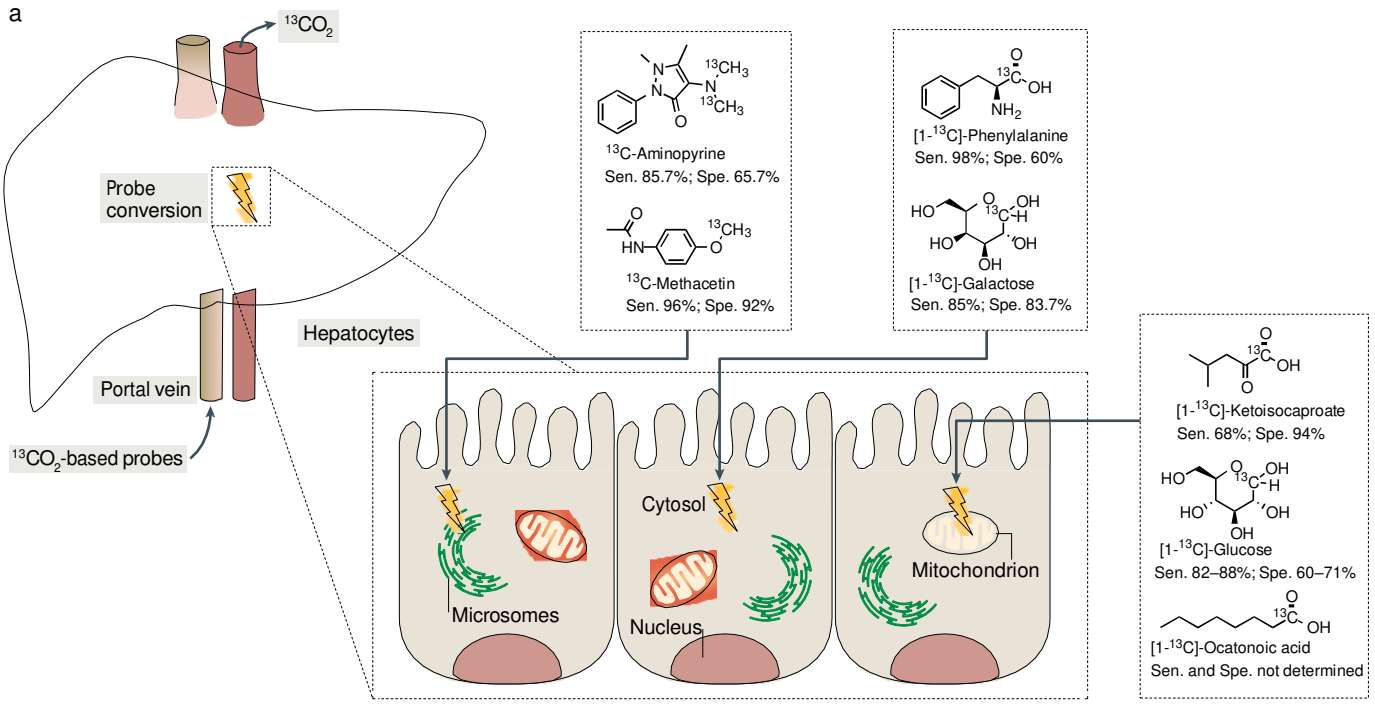


Fig. 4 | **Peptides for monitoring respiratory infections**[62]. **a** | Volatile organic compound (VOC)-based peptides are formulated into volatile-releasing activity-based nanosensors (vABNs), which are administered to model animals by intratracheal instillation. When attached to the vABNs, VOC reporters are undetectable (grey). In the case of lung infection, extracellular serine protease neutrophil elastase (NE) released by inflammatory cells hydrolyses tetrapeptide substrates, thereby releasing VOC reporters (orange). VOCs are further exhaled and detected by mass spectrometry. **b** | Immunofluorescence images of healthy lung (top left) versus lung infected with Pseudomonas aeruginosa (top right). Images show NE in infected lungs (green, top right). After intrapulmonary delivery of vABNs, extracellular NE (green, bottom left) was colocalized with vABNs (magenta, bottom left). vABNs were also detected inside lung cells (magenta, bottom right). White arrows indicate vABNs of interest. **c** | Signals of hydrofluoroamine volatiles after intrapulmonary delivery of vABNs in healthy controls and mouse models of lung infection with P. aeruginosa. vABN was used at 10 and 100 µM by peptide concentration (n = 3 or 4 mice per group). **d** | Kinetics of NE activity before and after antibiotic treatment in mice with P. aeruginosa lung infection. **e** | Amount of NE (ng) in lung homogenates at each experimental time point (n = 3, 4 or 6 mice per group). **f** | Signals of hydrofluoroamines at the same time points (n = 4, 5, 6 or 8 mice per group). **g** | Haematoxylin and eosin stained lung tissues (n = 3) show the recruitment and recession of neutrophils after infection and antibiotic treatment. c.p.s., counts per second; DAPI, 4′,6-diamidino-2-phenylindole; PEG, poly(ethylene glycol). ****$P < 0.0001$. Parts **a** and **c**–**f** adapted from ref.[62], Springer Nature Limited. Parts **b** and **g** reprinted from ref.[62], Springer Nature Limited.

$^{14}CO_2$ in exhaled breath[214]. This work enabled characterization of the activity of two different CYP monooxygenases towards both substrates. Assays in humans were then conducted to determine how the CYP isoforms in us process the substrates. One of the main probes used for this purpose was the labelled antibiotic ^{14}C-erythromycin[61]. Thanks to this probe, the activity of the major drug-metabolizing enzyme CYP3A4, which *N*-demethylates xenobiotics, has been highlighted[155]. The labelled probe is still used to evaluate the pharmacokinetics with which CYP3A4 metabolizes drugs such as docetaxel, a chemotherapeutic[156].

Labelled drugs for clinical diagnosis and prognosis of liver injuries. Numerous ^{13}C-labelled drugs have been used as $^{13}CO_2$-based probes to trace oxidative processes either in the microsomes, cytosol or mitochondria of hepatic cells[80,157–159] (Supplementary Table 1). These probes are orally administered to patients in doses of several dozens of milligrams or on the order of mg kg^{-1}. The labelled functional groups are converted by specific enzymes in the targeted cellular compartment to yield $^{13}CO_2$, which mixes with the body's bicarbonate pool and is exhaled. Breath samples are collected and analysed by isotope ratio mass spectrometry or non-dispersive isotope-selective infrared spectrometry. Because natural abundance of ^{13}C is approximately 1%, excess $^{13}CO_2$ in the breath is generally reported as a delta-over-baseline (DOB = $([^{13}CO_2] - [^{13}CO_2]_{baseline})/[^{13}CO_2]_{baseline}$) value given in per mille (‰). The results are compared with the baseline sample collected before ingestion of the ^{13}C drugs. A low proportion of $^{13}CO_2$ in the breath reflects a rate-limiting enzymatic step in the process, suggesting a reduced liver capacity.

The ^{13}C-aminopyrine breath test was the first protocol proposed for the diagnosis of patients suffering from liver disease[160]. This antipyretic analgesic can be metabolized by microsomal CYP isoforms and analysis of the resulting $^{13}CO_2$ provides a global assessment of liver function[161] (Fig. 5a, Supplementary Table 1). Other ^{13}C-drug probes, namely, ^{13}C-labelled methacetin, pantoprazole, phenacetin, erythromycin, diazepam and caffeine, have also been



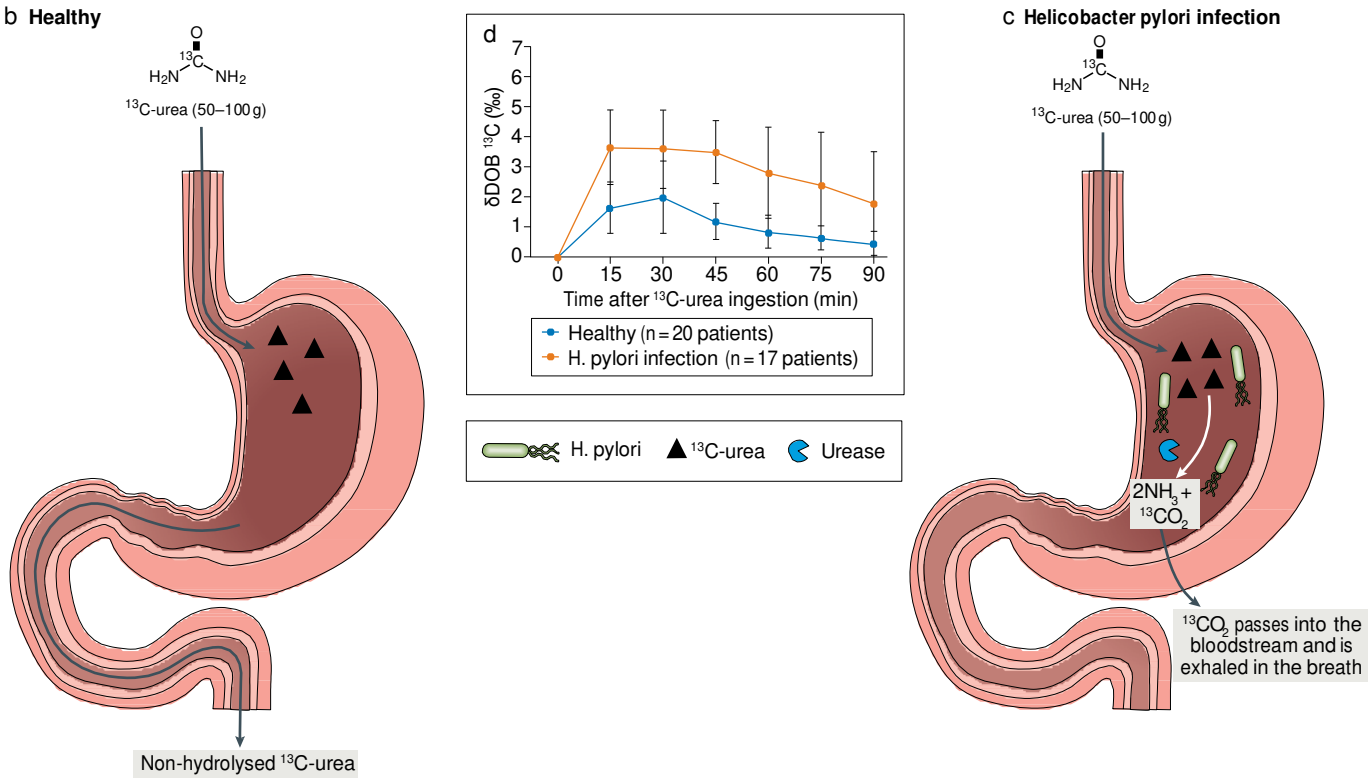


Fig. 5 | **Labelled small molecules in the clinic. a** | ^{13}C-labelled xenobiotics and small metabolites assessing microsomal P450 metabolism, and cytosolic and mitochondrial oxidative pathways. **b** | The ^{13}C-urea probe enables diagnosis of Helicobacter pylori infection in the human stomach[80]. In healthy individuals, ^{13}C-urea is eliminated by the kidneys into urine. **c** | In patients infected with H. pylori, ^{13}C-urea is cleaved by bacterial urease to give $2NH_3$ and $^{13}CO_2$. Absorbed $^{13}CO_2$ is excreted through the lungs. **d** | Kinetics of expired ^{13}C delta-over-baseline (DOB) values in healthy and infected patients. Sen., sensitivity; Spe., specificity. Part **a** drawn using data from ref.[80]. Part **d** drawn using data from ref.[184].

engineered to evaluate the activity of specific CYP isoforms in hepatocellular microsomes[162,163]. As the clinical objective of these ^{13}C drugs was to assess liver clearance, both their biochemical and pharmacological parameters, such as hepatic extraction or intestinal uptake, must be known to set the optimal dosage of probes and avoid any misinterpretation[162]. A variant of the methacetin test (Fig. 5a, Supplementary Table 1), called the LiMAx test, has been proposed to better trace the activity of CYP1A2. In the LiMAx test, the drug is not administered orally but intravenously[164], leading to rapid metabolism such that $^{13}CO_2$ can be exhaled and measured by infrared spectroscopy within a few minutes following injection. The LiMAx test was approved in the United Kingdom in 2017 and is now commercially available in Germany and Austria.

In addition to drugs, a wide range of labelled biological molecules have been tested directly in patients to investigate specific metabolic processes occurring in the cytosol or mitochondria of hepatocytes (Fig. 5a, Supplementary Table 1). Among these non-toxic molecules are ^{13}C-phenylalanine and ^{13}C-galactose, whose metabolism reflects cytosolic oxidative enzyme activity, which is inversely correlated with the severity of liver diseases[165–167]. In other work, ^{13}C-ketoisocaproate targets mitochondrial processes and is a demonstrably efficient tool to assess hepatic liver injury in patients with alcoholism[168] or patients with hepatocellular carcinoma[169] or estimate drug effects on liver mitochondrial function[170]. ^{13}C-methionine has been used as a non-invasive probe for evaluating hepatic mitochondrial oxidative capacity[171], while ^{13}C-glucose can help us monitor liver cell actions in response to glucose ingestion[172]. Thanks to this labelled sugar, glucose-related metabolic disorders such as insulin resistance or diabetes can now be monitored[173–176].

Despite biomedical interest in using labelled VOC-evolving probes to study enzyme activity, they lack specificity and selectivity, partly because they target ubiquitous intracellular enzymatic markers like CYP. Another concern is the kinetics of $^{13}CO_2$ release in exhaled breath, which does not truly reflect the kinetics of enzymatic $^{13}CO_2$ formation in the liver[177]. Indeed, a fraction of the newly formed $^{13}CO_2$ is taken up into several bodily compartments, from which it returns to the plasma only after a delay. Thus, DOB values may either overestimate or underestimate $^{13}CO_2$-based probe metabolism. Consequently, probes are typically administered in very high doses (several dozens or hundreds of mg) to certify an increase in $^{13}CO_2$ in patient breath, which may raise safety questions.

Labelled metabolites for diagnosis of* Helicobacter pylori *infection. From its first description in 1987 (ref.[177]) to its US Food and Drug Administration (FDA) approval in 2010, the development of $(H_2N)_2{}^{13}CO$ is probably the most significant breakthrough in $^{13}CO_2$-evolving probes. This labelled urea is widely used for patients suspected of infection with *H. pylori*[80] (Fig. 5b,c), a microaerophilic pathogen that can colonize mucosal surfaces of the stomach[178] and infects approximately 50% of the world's population, with an incidence of up to 80% in developing countries. This bacterium is the main cause of chronic gastritis and peptic ulcers and is closely linked to the development of gastric cancers[178–180]. During the non-invasive ^{13}C-urea breath test (UBT), patients are asked to ingest tablets of ^{13}C-urea (Fig. 5b,c), which is hydrolysed by the bacterial urease enzyme to give NH_3 and $^{13}CO_2$ (Fig. 5c). Breath samples are typically collected in Tedlar bags or vacutainer tubes 15–30 min after ingestion and analysed by isotope ratio mass spectrometry. Cut-off values between 2‰ and 5‰ for DOB have been widely used to discriminate between positive and negative tests[80]. A baseline breath sample is generally collected and the isotope enrichments compared some specified time after ingestion of the probe.

A positive result is indicated by an increase in expired ^{13}C expressed as DOB. Although there are hundreds of UBT variants (different probe doses and formulations, adjuvants, test meals, time points and isotope detectors)[181], the DOB cut-off values remain the key factor to be optimized[182]. These values are typically set between 2‰ and 5‰[80]. Despite its high sensitivity and specificity (both approximately 97%)[183,184], false-negative and false-positive tests are still obtained, especially for the 5% of samples near the cut-off values. Indeed, significant interindividual variabilities have been observed (reflected in the large error bars in Fig. 5d), owing to geographical location, gender and diet[185,186]. Additionally, false negatives and false positives can be caused by urea hydrolysis by mouth bacteria or by other urease-containing bacteria in the stomach[187]. However, the development of mathematical models such as the finite mixture model, combined with histological validation and serological tests, has allowed optimization of the cut-off point in diagnostic UBTs[182]. Through this translational strategy, the sensitivity of the test for subjects with borderline value of the UBT has been improved. Hence, compared with the gastric biopsy urease activity test, the bacterial culture after endoscopy or the rapid urease test (for example, CLOtest), the UBT is considered the best for therapy control, epidemiological investigation and pharmacological therapy trials.

Conclusions

The past few decades have seen major advances in the field of induced volatolomics. Being simple, rapid and non-invasive, this approach has shown great potential for translational research. Indeed, numerous volatile-compound-based probes have been (or are to be) exploited in the clinic for disease diagnosis and patient stratification for personalized therapy. In parallel, these bioorthogonal tools also provide a fundamental understanding of molecular processes either preceding or arising from diseases, thus helping chemists in the development of new therapeutics.

Belonging to the chemobiological activity-based tools, the probes we have described here work as off–on systems activated by an enzymatic stimulus. The enzymatic targets and volatiles released from probe metabolism are crucial factors affecting the efficiency of probes. Indeed, for induced volatolomics to be efficient, one must select an appropriate enzyme target, a decision that requires a thorough understanding of the molecular pathophysiological mechanisms occurring in the affected organs, tissues and cells. In the same way, the

Insulin resistance
This occurs when a normal or elevated insulin level produces an attenuated biological response. Classically, this refers to impaired sensitivity to insulin-mediated glucose disposal.

VOC reporter must be highly volatile to easily pass into the bloodstream and be exhaled in the breath.

Substantial variations in terms of probe chemistry, selectivity and specificity have been reported, and a wide range of enzymes, including glycosidases, peptidases, urease and oxidases, can now be identified at the molecular, cellular and organismal scale. With these probes in hand, one can envisage a cocktail of probes that targets different enzyme classes to deepen understanding of integrative processes implicated in disease dynamics. In a medical research context, several probes (for example, lactose, fructose, lactulose, ^{13}C-urea and ^{13}C-methacetin) presently find routine use, while others such as D_5-ethyl-β-d-glucuronide are in clinical pipelines. In summary, it is fair to say that induced volatolomics is still in its infancy. However, as part of multimodal and multiscale diagnosis platforms, this emerging chemobiological approach should allow molecular exploration of biological systems in a wide range of applications within a few years.

Published online: 02 February 2021

Bianconi, E. et al. An estimation of the number of cells in the human body. *Ann. Hum. Biol.* **40**, 463–471 (2013).

2. Amann, A. et al. The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J. Breath Res.* **8**, 034001 (2014).
3. de Lacy Costello, B. et al. A review of the volatiles from the healthy human body. *J. Breath Res.* **8**, 014001 (2014).
4. Broza, Y. Y., Zuri, L. & Haick, H. Combined volatolomics for monitoring of human body chemistry. *Sci. Rep.* **4**, 4611 (2014).
This paper reviews the advantages and drawbacks of analysing VOCs for diagnosing and monitoring pathologies.
5. Giannoukos, S., Agapiou, A., Brkić, B. & Taylor, S. Volatolomics: a broad area of experimentation. *J. Chromatogr. B* **1105**, 136–147 (2019).
6. Angle, C., Waggoner, L. P., Ferrando, A., Haney, P. & Passler, T. Canine detection of the volatilome: a review of implications for pathogen and disease detection. *Front. Vet. Sci.* **3**, 47 (2016).
7. Cambau, E. & Poljak, M. Sniffing animals as a diagnostic tool in infectious diseases. *Clin. Microbiol. Infect.* **26**, 431–435 (2020).
8. Pirrone, F. & Albertini, M. Olfactory detection of cancer by trained sniffer dogs: a systematic review of the literature. *J. Vet. Behav.* **19**, 105–117 (2017).
9. Agapiou, A., Amann, A., Mochalski, P., Statheropoulos, M. & Thomas, C. L. P. Trace detection of endogenous human volatile organic compounds for search, rescue and emergency applications. *Trends Anal. Chem.* **66**, 158–175 (2015).
10. Tipple, C. A. et al. Comprehensive characterization of commercially available canine training aids. *Forensic Sci. Int.* **242**, 242–254 (2014).
11. Ubeda, C., Lepe-Balsalobre, E., Ariza-Astolfi, C. & Ubeda-Ontiveros, J. M. Identification of volatile biomarkers of *Giardia duodenalis* infection in children with persistent diarrhoea. *Parasitol. Res.* **118**, 3139–3147 (2019).
12. Dormont, L., Bessière, J.-M. & Cohuet, A. Human skin volatiles: a review. *J. Chem. Ecol.* **39**, 569–578 (2013).
13. Phillips, M. et al. Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. *Lancet* **353**, 1930–1933 (1999).
14. Penn, D. J. et al. Individual and gender fingerprints in human body odour. *J. R. Soc. Interface* **4**, 331–340 (2007).
15. Al-Kateb, H., de Lacy Costello, B. & Ratcliffe, N. An investigation of volatile organic compounds from the saliva of healthy individuals using headspace-trap/GC-MS. *J. Breath Res.* **7**, 036004 (2013).
16. Amann, A. et al. Applications of breath gas analysis in medicine. *Int. J. Mass Spectrom.* **239**, 227–233 (2004).
17. Trefz, P. et al. Exhaled volatile substances in children suffering from type 1 diabetes mellitus: results from a cross-sectional study. *Sci. Rep.* **9**, 15707 (2019).
18. Nelson, N., Lagesson, V., Nosratabadi, A. R., Ludvigsson, J. & Tagesson, C. Exhaled isoprene and acetone in newborn infants and in children with diabetes mellitus. *Pediatr. Res.* **44**, 363–367 (1998).
19. Neupane, S. et al. Exhaled breath isoprene rises during hypoglycemia in type 1 diabetes. *Diabetes Care* **39**, e97–e98 (2016).
20. Mochalski, P. et al. Blood and breath levels of selected volatile organic compounds in healthy volunteers. *Analyst* **138**, 2134–2145 (2013).
21. Mochalski, P. et al. Blood and breath profiles of volatile organic compounds in patients with end-stage renal disease. *BMC Nephrol.* **15**, 43 (2014).
22. Filipiak, W. et al. A compendium of volatile organic compounds (VOCs) released by human cell lines. *Curr. Med. Chem.* **23**, 2112–2131 (2016).
23. Elmassry, M. M. & Piechulla, B. Volatilomes of bacterial infections in humans. *Front. Neurosci.* **14**, 257 (2020).
24. Serasanambati, M., Broza, Y. Y., Marmur, A. & Haick, H. Profiling single cancer cells with volatolomics approach. *iScience* **11**, 178–188 (2019).
25. Lawal, O. et al. Headspace volatile organic compounds from bacteria implicated in ventilator-associated pneumonia analysed by TD-GC/MS. *J. Breath Res.* **12**, 026002 (2018).
26. Filipiak, W. et al. Comparative analyses of volatile organic compounds (VOCs) from patients, tumors and transformed cell lines for the validation of lung cancer-derived breath markers. *J. Breath Res.* **8**, 027111 (2014).
27. He, J. et al. Fingerprinting breast cancer vs. normal mammary cells by mass spectrometric analysis of volatiles. *Sci. Rep.* **4**, 5196 (2014).
28. Martin, A. N., Farquar, G. R., Jones, A. D. & Frank, M. Human breath analysis: methods for sample collection and reduction of localized background effects. *Anal. Bioanal. Chem.* **396**, 739–750 (2010).
29. Kim, K.-H., Jahan, S. A. & Kabir, E. A review of breath analysis for diagnosis of human health. *Trends Anal. Chem.* **33**, 1–8 (2012).
30. Lubes, G. & Goodarzi, M. Analysis of volatile compounds by advanced analytical techniques and multivariate chemometrics. *Chem. Rev.* **117**, 6399–6422 (2017).
31. Schmidt, K. & Podmore, I. Current challenges in volatile organic compounds analysis as potential biomarkers of cancer. *J. Biomark.* **2015**, 981458 (2015).
32. Lan, H., Hartonen, K. & Riekkola, M.-L. Miniaturised air sampling techniques for analysis of volatile organic compounds in air. *Trends Anal. Chem.* **126**, 115873 (2020).
33. Lawal, O., Ahmed, W. M., Nijsen, T. M. E., Goodacre, R. & Fowler, S. J. Exhaled breath analysis: a review of 'breath-taking' methods for off-line analysis. *Metabolomics* **13**, 110 (2017).
34. Shende, P., Vaidya, J., Kulkarni, Y. A. & Gaud, R. S. Systematic approaches for biodiagnostics using exhaled air. *J. Control. Rel.* **268**, 282–295 (2017).
35. Sandlund, J. et al. Development of colorimetric sensor array for diagnosis of tuberculosis through detection of urinary volatile organic compounds. *Diagn. Microbiol. Infect. Dis.* **92**, 299–304 (2018).
36. Rakow, N. A. & Suslick, K. S. A colorimetric sensor array for odour visualization. *Nature* **406**, 710–713 (2000).
37. Hu, W. et al. Electronic noses: from advanced materials to sensors aided with data processing. *Adv. Mater. Technol.* **4**, 1800488 (2019).
38. Geng, Y., Peveler, W. J. & Rotello, V. M. Array-based "chemical nose" sensing in diagnostics and drug discovery. *Angew. Chem. Int. Ed.* **58**, 5190–5200 (2019).
39. Askim, J. R., Mahmoudi, M. & Suslick, K. S. Optical sensor arrays for chemical sensing: the optoelectronic nose. *Chem. Soc. Rev.* **42**, 8649–8682 (2013).
40. Iitani, K., Naisierding, M., Toma, K., Arakawa, T. & Mitsubayashi, K. Evaluation for regional difference of skin-gas ethanol and sweat rate using alcohol dehydrogenase-mediated fluorometric gas-imaging system (sniff-cam). *Analyst* **145**, 2915–2924 (2020).
41. Martinez-Vernon, A. S. et al. An improved machine learning pipeline for urinary volatiles disease detection: diagnosing diabetes. *PLoS ONE* **13**, e0204425 (2018).
42. Palma, S. I. C. J. et al. Machine learning for the meta-analyses of microbial pathogens' volatile signatures. *Sci. Rep.* **8**, 3360 (2018).
43. Tait, E., Perry, J. D., Stanforth, S. P. & Dean, J. R. Use of volatile compounds as a diagnostic tool for the detection of pathogenic bacteria. *Trends Anal. Chem.* **53**, 117–125 (2014).
44. Turner, A. P. F. & Magan, N. Electronic noses and disease diagnostics. *Nat. Rev. Microbiol.* **2**, 161–166 (2004).
45. Walzl, G. et al. Tuberculosis: advances and challenges in development of new diagnostics and biomarkers. *Lancet Infect. Dis.* **18**, e199–e210 (2018).
46. Traxler, S. et al. Volatile scents of influenza A and *S. pyogenes* (co-)infected cells. *Sci. Rep.* **9**, 18894 (2019).
47. Purcaro, G. et al. Volatile fingerprinting of human respiratory viruses from cell culture. *J. Breath Res.* **12**, 026015 (2018).
48. Usman, F. et al. A review of biosensors for non-invasive diabetes monitoring and screening in human exhaled breath. *IEEE Access* **7**, 5963–5974 (2019).
49. Zhou, J., Huang, Z.-A., Kumar, U. & Chen, D. D. Y. Review of recent developments in determining volatile organic compounds in exhaled breath as biomarkers for lung cancer diagnosis. *Anal. Chim. Acta* **996**, 1–9 (2017).
50. Sun, X., Shao, K. & Wang, T. Detection of volatile organic compounds (VOCs) from exhaled breath as noninvasive methods for cancer diagnosis. *Anal. Bioanal. Chem.* **408**, 2759–2780 (2016).
51. Krilaviciute, A. et al. Detection of cancer through exhaled breath: a systematic review. *Oncotarget* **6**, 38643–38657 (2015).
52. Neerincx, A. H. et al. Breathomics from exhaled volatile organic compounds in pediatric asthma. *Pediatr. Pulmonol.* **52**, 1616–1627 (2017).
53. Bos, L. D. J. Diagnosis of acute respiratory distress syndrome by exhaled breath analysis. *Ann. Transl. Med.* **6**, 33 (2018).
54. Smolinska, A. et al. Volatile metabolites in breath strongly correlate with gut microbiome in CD patients. *Anal. Chim. Acta* **1025**, 1–11 (2018).
55. Trivedi, D. K. et al. Discovery of volatile biomarkers of Parkinson's disease from sebum. *ACS Cent. Sci.* **5**, 599–606 (2019).
56. Lau H.-C., Y. J.-B., Lee H. W., Huh J. S. & Lim, J. O. Investigation of exhaled breath samples from patients with Alzheimer's disease using gas chromatography–mass spectrometry and an exhaled breath sensor system. *Sensors* **17**, 1783 (2017).
57. Tait, E., Stanforth, S. P., Reed, S., Perry, J. D. & Dean, J. R. Analysis of pathogenic bacteria using exogenous volatile organic compound metabolites and optical sensor detection. *RSC Adv.* **5**, 15494–15499 (2015).
First application of VOC-based probes to discriminate between enzyme-producing bacteria in biological samples.
58. Thompson, R. et al. Detection of β-alanyl aminopeptidase as a biomarker for *Pseudomonas aeruginosa* in the sputum of patients with cystic fibrosis using exogenous volatile organic compound evolution. *RSC Adv.* **10**, 10634–10645 (2020).
59. Bahroun, N. H. O., Perry, J. D., Stanforth, S. P. & Dean, J. R. Use of exogenous volatile organic compounds to detect *Salmonella* in milk. *Anal. Chim. Acta* **1028**, 121–130 (2018).
This work describes the use of multiplexed VOC-based probes for biomedicine. This strategy enabled clear detection of a pathogenic strain in food.
60. Taylor, C. et al. Analysis of *Listeria* using exogenous volatile organic compound metabolites and their

detection by static headspace–multi-capillary column–gas chromatography–ion mobility spectrometry (SHS–MCC–GC–IMS). *Anal. Bioanal. Chem.* **409**, 4247–4256 (2017).

61. Watkins, P. B. et al. Erythromycin breath test as an assay of glucocorticoid-inducible liver cytochromes P-450. Studies in rats and patients. *J. Clin. Invest.* **83**, 688–697 (1989).
This study describes the use of ^{14}C-erythromycin to investigate CYP demethylase activity by measuring breath $^{14}CO_2$ production after medication. The approach was first tested on rats and then humans.
62. Chan, L. W. et al. Engineering synthetic breath biomarkers for respiratory disease. *Nat. Nanotechnol.* **15**, 792–800 (2020).
This work demonstrates the use of volatile-releasing activity-based nanosensors for assessing in vivo inflammatory response to respiratory bacterial infection.
63. Douard, V. & Ferraris, R. P. The role of fructose transporters in diseases linked to excessive fructose intake. *J. Physiol.* **591**, 401–414 (2013).
64. Lange, J. et al. Volatile organic compound (VOC)-based probe for induced volatolomics of cancers. *Angew. Chem. Int. Ed.* **58**, 17563–17566 (2019).
This study describes a sugar probe used to diagnose cancers and monitor solid tumour responses to chemotherapy.
65. Bucci, M., Goodman, C. & Sheppard, T. L. A decade of chemical biology. *Nat. Chem. Biol.* **6**, 847–854 (2010).
66. Lemke, E. A. & Schultz, C. Principles for designing fluorescent sensors and reporters. *Nat. Chem. Biol.* **7**, 480–483 (2011).
67. Xu, W. et al. In vivo imaging of a novel strain of *Bacteroides fragilis* via metabolic labeling. *Front. Microbiol.* **9**, 2298 (2018).
68. Yang, P. & Liu, K. Activity-based protein profiling: recent advances in probe development and applications. *ChemBioChem* **16**, 712–724 (2015).
69. Sinharay, S., Randtke, E. A., Howison, C. M., Ignatenko, N. A. & Pagel, M. D. Detection of enzyme activity and inhibition during studies in solution, in vitro and in vivo with CatalyCEST MRI. *Mol. Imaging Biol.* **20**, 240–248 (2018).
70. Galas, L. et al. "Probe, sample, and instrument (PSI)": the hat-trick for fluorescence live cell imaging. *Chemosensors* **6**, 40 (2018).
71. Volpe, A., Kurtys, E. & Fruhwirth, G. O. Cousins at work: how combining medical with optical imaging enhances in vivo cell tracking. *Int. J. Biochem. Cell Biol.* **102**, 40–50 (2018).
72. Jayanthi, V. S. P. K. S. A., Das, A. B. & Saxena, U. Recent advances in biosensor development for the detection of cancer biomarkers. *Biosens. Bioelectron.* **91**, 15–23 (2017).
73. Stoddard, E. G. et al. Multifunctional activity-based protein profiling of the developing lung. *J. Proteome Res.* **17**, 2623–2634 (2018).
74. Hewings, D. S., Flygare, J. A., Wertz, I. E. & Bogyo, M. Activity-based probes for the multicatalytic proteasome. *FEBS J.* **284**, 1540–1554 (2017).
75. Chang, L., He, X., Chen, L. & Zhang, Y. A novel fluorescent turn-on biosensor based on QDs@GSH–GO fluorescence resonance energy transfer for sensitive glutathione *S*-transferase sensing and cellular imaging. *Nanoscale* **9**, 3881–3888 (2017).
76. Yoo, B. et al. Detection of in vivo enzyme activity with CatalyCEST MRI. *Magn. Reson. Med.* **71**, 1221–1230 (2014).
77. Sabale, S. et al. Recent developments in the synthesis, properties, and biomedical applications of core/shell superparamagnetic iron oxide nanoparticles with gold. *Biomater. Sci.* **5**, 2212–2225 (2017).
78. Huang, L. et al. Plasmonic silver nanoshells for drug and metabolite detection. *Nat. Commun.* **8**, 220 (2017).
79. Devaraj, N. K. The future of bioorthogonal chemistry. *ACS Cent. Sci.* **4**, 952–959 (2018).
80. Graham, D. Y. & Miftahussurur, M. *Helicobacter pylori* urease for diagnosis of *Helicobacter pylori* infection: a mini review. *J. Adv. Res.* **13**, 51–57 (2018).
81. Gaude, E. et al. Targeted breath analysis: exogenous volatile organic compounds (EVOC) as metabolic pathway-specific probes. *J. Breath Res.* **13**, 032001 (2019).
This review is the first describing exogenous VOCs as probes for monitoring global metabolic processes.
82. Simrén, M. & Stotzer, P.-O. Use and abuse of hydrogen breath tests. *Gut* **55**, 297–303 (2006).
83. Losurdo, G. et al. Breath tests for the non-invasive diagnosis of small intestinal bacterial overgrowth: a systematic review with meta-analysis. *J. Neurogastroenterol. Motil.* **26**, 16–28 (2020).
84. Calloway, D. H., Murphy, E. L. & Bauer, D. Determination of lactose intolerance by breath analysis. *Am. J. Dig. Dis.* **14**, 811–815 (1969).
First clinical work describing an exogenous volatile-based probe for tracing gastrointestinal disorder. The assay highlighted the suitability of lactose as an H_2-based probe sugar for diagnosing patients with lactose-intolerance.
85. Rao, S. S. C., Attaluri, A., Anderson, L. & Stumbo, P. Ability of the normal human small intestine to absorb fructose: evaluation by breath testing. *Clin. Gastroenterol. Hepatol.* **5**, 959–963 (2007).
86. Ebert, K. & Witt, H. Fructose malabsorption. *Mol. Cell. Pediatr.* **3**, 10 (2016).
87. Enattah, N. S. et al. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30**, 233–237 (2002).
88. Kanai, T. et al. Overproduction of the membrane-bound [NiFe]-hydrogenase in *Thermococcus kodakarensis* and its effect on hydrogen production. *Front. Microbiol.* **6**, 847 (2015).
89. Vignais, P. M., Billoud, B. & Meyer, J. Classification and phylogeny of hydrogenases. *FEMS Microbiol. Rev.* **25**, 455–501 (2001).
90. Gevorgyan, H., Trchounian, A. & Trchounian, K. Understanding the role of *Escherichia coli* hydrogenases and formate dehydrogenases in the F_OF_1-ATPase activity during the mixed acid fermentation of mixture of carbon sources. *IUBMB Life* **70**, 1040–1047 (2018). Huang,
91. G., Wagner, T., Ermler, U. & Shima, S. Methanogenesis involves direct hydride transfer from H_2 to an organic substrate. *Nat. Rev. Chem.* **4**, 213–221 (2020).
92. Rhodes, J. M., Middleton, P. & Jewell, D. P. The lactulose hydrogen breath test as a diagnostic test for small-bowel bacterial overgrowth. *Scand. J. Gastroenterol.* **14**, 333–336 (1979).
93. Panesar, P. S. & Kumari, S. Lactulose: production, purification and potential applications. *Biotechnol. Adv.* **29**, 940–948 (2011).
94. Pimentel, M., Chow, E. J. & Lin, H. C. Eradication of small intestinal bacterial overgrowth reduces symptoms of irritable bowel syndrome. *Am. J. Gastroenterol.* **95**, 3503–3506 (2000).
95. Ghoshal, U. C. How to interpret hydrogen breath tests. *J. Neurogastroenterol. Motil.* **17**, 312–317 (2011).
96. Choi, Y. K., Johlin, F. C. Jr, Summers, R. W., Jackson, M. & Rao, S. S. C. Fructose intolerance: an under-recognized problem. *Am. J. Gastroenterol* **98**, 1348–1353 (2003).
97. Erdogan, A., Coss-Adame, E., Yu, S., Rattanakovit, K. & Rao, S. S. C. Optimal testing for diagnosis of fructose intolerance: over-dosage leads to false positive intolerance test. *J. Neurogastroenterol. Motil.* **20**, 560 (2014).
98. Szilagyi, A. et al. Comparison of a real-time polymerase chain reaction assay for lactase genetic polymorphism with standard indirect tests for lactose maldigestion. *Clin. Gastroenterol. Hepatol.* **5**, 192–196 (2007).
99. Lin, E. C. & Massey, B. T. Scintigraphy demonstrates high rate of false-positive results from glucose breath tests for small bowel bacterial overgrowth. *Clin. Gastroenterol. Hepatol.* **14**, 203–208 (2016).
100. Yu, D., Cheeseman, F. & Vanner, S. Combined oro-caecal scintigraphy and lactulose hydrogen breath testing demonstrate that breath testing detects oro-caecal transit, not small intestinal bacterial overgrowth in patients with IBS. *Gut* **60**, 334–340 (2011).
101. Kerber, M. et al. Hydrogen breath testing versus LCT genotyping for the diagnosis of lactose intolerance: a matter of age? *Clin. Chim. Acta* **383**, 91–96 (2007).
102. Amieva-Balmori, M., Coss-Adame, E., Rao, N. S., Dávalos-Pantoja, B. M. & Rao, S. S. C. Diagnostic utility of carbohydrate breath tests for SIBO, fructose, and lactose intolerance. *Dig. Dis. Sci.* **65**, 1405–1413 (2020).
103. Rumessen, J. J., Nordgaard-Andersen, I. & Gudmand-Høyer, E. Carbohydrate malabsorption: quantification by methane and hydrogen breath tests. *Scand. J. Gastroenterol.* **29**, 826–832 (1994).
104. Sahakian, A. B., Jee, S.-R. & Pimentel, M. Methane and the gastrointestinal tract. *Dig. Dis. Sci.* **55**, 2135–2143 (2010).
105. Perelló, A. et al. M1249: Methane and hydrogen breath testing for carbohydrate malabsorption. *Gastroenterology* **138**, S-363 (2010).
106. Materacki, L. et al. PWE-098 Is methane testing a useful adjunct to hydrogen breath testing? *Gut* **67**, A167 (2018).
107. Harvie, R. M., Tuck, C. J. & Schultz, M. Evaluation of lactulose, lactose, and fructose breath testing in clinical practice: a focus on methane. *JGH Open* **4**, 198–205 (2020).
108. Enko, D., Rezanka, E., Stolba, R. & Halwachs-Baumann, G. Lactose malabsorption testing in daily clinical practice: a critical retrospective analysis and comparison of the hydrogen/methane breath test and genetic test (C/T_{-13910} polymorphism) results. *Gastroenterol. Res. Pract.* **2014**, 464382 (2014).
109. Sundin, O. H. et al. Does a glucose-based hydrogen and methane breath test detect bacterial overgrowth in the jejunum? *Neurogastroenterol. Motil.* **30**, e13350 (2018).
110. Ghoshal, U. C., Kumar, S., Misra, A. & Mittal, B. Lactose malabsorption diagnosed by 50-g dose is inferior to assess clinical intolerance and to predict response to milk withdrawal than 25-g dose in an endemic area. *J. Gastroenterol. Hepatol.* **28**, 1462–1468 (2013).
111. Strocchi, A., Corazza, G., Ellis, C. J., Gasbarrini, G. & Levitt, M. D. Detection of malabsorption of low doses of carbohydrate: accuracy of various breath H_2 criteria. *Gastroenterology* **105**, 1404–1410 (1993).
112. Ishibe, A. et al. Detection of gas components as a novel diagnostic method for colorectal cancer. *Ann. Gastroenterol. Surg.* **2**, 147–153 (2018).
113. Wilder-Smith, C. H., Olesen, S. S., Materna, A. & Drewes, A. M. Breath methane concentrations and markers of obesity in patients with functional gastrointestinal disorders. *United European Gastroenterol. J.* **6**, 595–603 (2017).
114. Gottlieb, K. et al. Selection of a cut-off for high- and low-methane producers using a spot-methane breath test: results from a large north American dataset of hydrogen, methane and carbon dioxide measurements in breath. *Gastroenterol. Rep.* **5**, 193–199 (2017).
115. Rubio-Escudero, C. et al. Data mining techniques applied to hydrogen lactose breath test. *PLoS ONE* **12**, e0170385 (2017).
116. Zhang, J., Cheng, P. & Pu, K. Recent advances of molecular optical probes in imaging of β-galactosidase. *Bioconjug. Chem.* **30**, 2089–2101 (2019).
117. Begoyan, V. V. et al. Multicolor GLUT5-permeable fluorescent probes for fructose transport analysis. *Chem. Commun.* **54**, 3855–3858 (2018).
118. Gao, F. et al. Simultaneous detection of hydrogen and methane in breath for the diagnosis of small intestinal bacterial overgrowth by fast gas chromatography. *Anal. Methods* **10**, 4329–4338 (2018).
119. Newman, A. Breath-analysis tests in gastroenterology. *Gut* **15**, 308–323 (1974).
120. King, C. E., Toskes, P. P., Spivey, J. C., Lorenz, E. & Welkos, S. Detection of small intestine bacterial overgrowth by means of a ^{14}C-d-xylose breath test. *Gastroenterology* **77**, 75–82 (1979).
121. Walters, B. & Vanner, S. J. Detection of bacterial overgrowth in IBS using the lactulose H_2 breath test: comparison with ^{14}C-d-xylose and healthy controls. *Am. J. Gastroenterol.* **100**, 1566–1570 (2005).
122. King, C. E. & Toskes, P. P. Comparison of the 1-gram [^{14}C]xylose, 10-gram lactulose-H_2, and 80-gram glucose-H_2 breath tests in patients with small intestine bacterial overgrowth. *Gastroenterology* **91**, 1447–1451 (1986).
123. Gunnarsson, M. et al. Long-term biokinetics and radiation exposure of patients undergoing ^{14}C-glycocholic acid and ^{14}C-xylose breath tests. *Cancer Biother. Radiopharm.* **22**, 762–771 (2007).
124. Wurst, F. M., Skipper, G. E. & Weinmann, W. Ethyl glucuronide — the direct ethanol metabolite on the threshold from science to routine use. *Addiction* **98**, 51–61 (2003).
125. Bosslet, K. et al. Elucidation of the mechanism enabling tumor selective prodrug monotherapy. *Cancer Res.* **58**, 1195–1201 (1998).
126. Bosslet, K., Czech, J. & Hoffmann, D. A novel one-step tumor-selective prodrug activation system. *Tumor Target.* **1**, 45–50 (1995).
127. Fishman, W. H. & Anlyan, A. J. Comparison of the β-glucuronidase activity of normal, tumor, and lymph node tissues of surgical patients. *Science* **106**, 66–67 (1947).
128. Albin, N. et al. Main drug-metabolizing enzyme systems in human breast tumors and peritumoral tissues. *Cancer Res.* **53**, 3541–3546 (1993).
129. Sperker, B. et al. Expression and function of β-glucuronidase in pancreatic cancer: potential role in drug targeting. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **362**, 110–115 (2000).
130. Renoux, B. et al. Targeting the tumour microenvironment with an enzyme-responsive drug

delivery system for the efficient therapy of breast and pancreatic cancers. *Chem. Sci.* **8**, 3427–3433 (2017).
131. Jin, Y. et al. Highly specific near-infrared fluorescent probe for the real-time detection of β-glucuronidase in various living cells and animals. *Anal. Chem.* **90**, 3276–3283 (2018).
132. Awolade, P. et al. Therapeutic significance of β-glucuronidase activity and its inhibitors: a review. *Eur. J. Med. Chem.* **187**, 111921 (2020).
133. Cortadellas, T. et al. Estimation of tumor size in breast cancer comparing clinical examination, mammography, ultrasound and MRI — correlation with the pathological analysis of the surgical specimen. *Gland Surg.* **6**, 330–335 (2017).
134. Wu, M.-H. et al. Features of non-small cell lung carcinomas overlooked at digital chest radiography. *Clin. Radiol.* **63**, 518–528 (2008).
135. Mitchell, D. G. et al. Early invasive cervical cancer: tumor delineation by magnetic resonance imaging, computed tomography, and clinical examination, verified by pathologic results, in the ACRIN 6651/GOG 183 intergroup study. *J. Clin. Oncol.* **24**, 5687–5694 (2006).
136. Cabello, J. & Ziegler, S. I. Advances in PET/MR instrumentation and image reconstruction. *Br. J. Radiol.* **91**, 20160363 (2016).
137. Marom, E. M., Sarvis, S., Herndon, J. E. & Patz, E. F. T1 lung cancers: sensitivity of diagnosis with fluorodeoxyglucose PET. *Radiology* **223**, 453–459 (2002).
138. Peng, L. et al. Tissue and plasma proteomics for early stage cancer detection. *Mol. Omics* **14**, 405–423 (2018).
139. Kałużna-Czaplińska, J. & Jóźwik, J. Current applications of chromatographic methods for diagnosis and identification of potential biomarkers in cancer. *Trends Anal. Chem.* **56**, 1–12 (2014).
140. Lambin, P. et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **14**, 749–762 (2017).
141. Lee, D.-K. et al. In vitro tracking of intracellular metabolism-derived cancer volatiles via isotope labeling. *ACS Cent. Sci.* **4**, 1037–1044 (2018).
142. Alahi, E. E. M. & Mukhopadhyay, C. S. Detection methodologies for pathogen and toxins: a review. *Sensors* **17**, 1885 (2017).
143. Garnacho-Montero, J. et al. Timing of adequate antibiotic therapy is a greater determinant of outcome than are TNF and IL-10 polymorphisms in patients with sepsis. *Crit. Care* **10**, R111 (2006).
144. Kralik, P. & Ricchi, M. A basic guide to real time PCR in microbial diagnostics: definitions, parameters, and everything. *Front. Microbiol.* **8**, 108 (2017).
145. Law, J. W.-F., Ab Mutalib, N.-S., Chan, K.-G. & Lee, L.-H. Rapid methods for the detection of foodborne bacterial pathogens: principles, applications, advantages and limitations. *Front. Microbiol.* **5**, 770 (2015).
146. Chu, Y. W., Wang, B. Y., Engebretson, D. A. & Carey, J. R. Single step, rapid identification of pathogenic microorganisms in a culture bottle. *Analyst* **138**, 5879–5885 (2013).
147. Guillemot, L.-H., Vrignaud, M., Marcoux, P. R., Rivron, C. & Tran-Thi, T.-H. Facile and fast detection of bacteria via the detection of exogenous volatile metabolites released by enzymatic hydrolysis. *Phys. Chem. Chem. Phys.* **15**, 15840–15844 (2013).
148. Bedernjak, A. F. et al. Synthesis and evaluation of novel 7- and 8-aminophenoxazinones for the detection of β-alanine aminopeptidase activity and the reliable identification of *Pseudomonas aeruginosa* in clinical samples. *J. Med. Chem.* **59**, 4476–4487 (2016).
149. Jokerst, J. C. et al. Development of a paper-based analytical device for colorimetric detection of select foodborne pathogens. *Anal. Chem.* **84**, 2900–2907 (2012).
150. Orenga, S., James, A. L., Manafi, M., Perry, J. D. & Pincus, D. H. Enzymatic substrates in microbiology. *J. Microbiol. Methods* **79**, 139–155 (2009).
151. Ramírez-Guízar, S. et al. A chromatographic approach to distinguish Gram-positive from Gram-negative bacteria using exogenous volatile organic compound metabolites. *J. Chromatogr. A* **1501**, 79–88 (2017).
152. Pham, C. T. N. Neutrophil serine proteases: specific regulators of inflammation. *Nat. Rev. Immunol.* **6**, 541–550 (2006).
153. Bircher, J. & Preisig, R. Exhalation of isotopic CO_2. *Methods Enzymol.* **77**, 3–9 (1981).
154. Guengerich, F. P. Cytochrome P450 and chemical toxicology. *Chem. Res. Toxicol.* **21**, 70–83 (2008).
155. Watkins, P. B. Erythromycin breath test and clinical transplantation. *Ther. Drug Monit.* **18**, 368–371 (1996).
156. Michael, M. et al. Docetaxel pharmacokinetics and its correlation with two in vivo probes for cytochrome P450 enzymes: the C^{14}-erythromycin breath test and the antipyrine clearance test. *Cancer Chemother. Pharmacol.* **69**, 125–135 (2012).
157. Modak, A. S. Regulatory issues on breath tests and updates of recent advances on [^{13}C]-breath tests. *J. Breath Res.* **7**, 037103 (2013).
158. Charidemou, E., Ashmore, T. & Griffin, J. L. The use of stable isotopes in the study of human pathophysiology. *Int. J. Biochem. Cell Biol.* **93**, 102–109 (2017).
159. Bonfrate, L., Grattagliano, I., Palasciano, G. & Portincasa, P. Dynamic carbon 13 breath tests for the study of liver function and gastric emptying. *Gastroenterol. Rep.* **3**, 12–21 (2014).
160. Hepner, G. W. & Vesell, E. S. Quantitative assessment of hepatic function by breath analysis after oral administration of [^{14}C]aminopyrine. *Ann. Intern. Med.* **83**, 632–638 (1975).
161. Pijls, K. E. et al. Critical appraisal of ^{13}C breath tests for microsomal liver function: aminopyrine revisited. *Liver Int.* **34**, 487–494 (2014).
162. Armuzzi, A. et al. Breath testing for human liver function assessment. *Aliment. Pharmacol. Ther.* **16**, 1977–1996 (2002).
163. Gorowska-Kowolik, K., Chobot, A. & Kwiecien, J. ^{13}C Methacetin breath test for assessment of microsomal liver function: methodology and clinical application. *Gastroenterol. Res. Pract.* **2017**, 7397840 (2017).
164. Buechter, M., Kersting, S., Gerken, G. & Kahraman, A. Enzymatic liver function measured by LiMAx — a reliable diagnostic and prognostic tool in chronic liver disease. *Sci. Rep.* **9**, 13577 (2019).
A thorough evaluation and comparison of the diagnostic and prognostic performance of different non-invasive tools for detecting chronic liver disease. This led to the LiMAx test, which relies on ^{13}C-methacetin, being clinically approved in different countries.
165. Burke, P. A. et al. l-[1-^{13}C]phenylalanine oxidation as a measure of hepatocyte functional capacity in end-stage liver disease. *Am. J. Surg.* **173**, 270–273 (1997).
166. Ishii, Y. et al. l-[1-^{13}C]phenylalanine breath test reflects phenylalanine hydroxylase activity of the whole liver. *J. Surg. Res.* **112**, 38–42 (2003).
167. Saadeh, S. et al. The utility of the ^{13}C-galactose breath test as a measure of liver function. *Aliment. Pharmacol. Ther.* **18**, 995–1002 (2003).
168. Witschi, A., Mossi, S., Meyer, B., Junker, E. & Lauterburg, B. H. Mitochondrial function reflected by the decarboxylation of [^{13}C]ketoisocaproate is impaired in alcoholics. *Alcohol. Clin. Exp. Res.* **18**, 951–955 (1994).
169. Palmieri, V. O. et al. Liver function as assessed by breath tests in patients with hepatocellular carcinoma. *J. Surg. Res.* **157**, 199–207 (2009).
170. Pessayre, D. et al. Central role of mitochondria in drug-induced liver injury. *Drug Metab. Rev.* **44**, 34–87 (2012).
171. Banasch, M., Ellrichmann, M., Tannapfel, A., Schmidt, W. & Goetze, O. The non-invasive ^{13}C-methionine breath test detects hepatic mitochondrial dysfunction as a marker of disease activity in non-alcoholic steatohepatitis. *Eur. J. Med. Res.* **16**, 258–264 (2011).
172. Savage, D. B., Petersen, K. F. & Shulman, G. I. Disordered lipid metabolism and the pathogenesis of insulin resistance. *Physiol. Rev.* **87**, 507–520 (2007).
173. Mizrahi, M., Lalazar, G., Adar, T., Raz, I. & Ilan, Y. Assessment of insulin resistance by a ^{13}C glucose breath test: a new tool for early diagnosis and follow-up of high-risk patients. *Nutr. J.* **9**, 25 (2010).
174. Hussain, M. et al. [^{13}C]Glucose breath testing provides a noninvasive measure of insulin resistance: calibration analyses against clamp studies. *Diabetes Technol. Ther.* **16**, 102–112 (2013).
175. Maldonado-Hernández, J., Martínez-Basila, A., Rendón-Macías, M. E. & López-Alarcón, M. Accuracy of the ^{13}C-glucose breath test to identify insulin resistance in non-diabetic adults. *Acta Diabetol.* **56**, 923–929 (2019).
176. Tanaka, K. et al. Noninvasive assessment of insulin resistance in the liver using the fasting ^{13}C-glucose breath test. *Transl. Res.* **162**, 191–200 (2013).
177. Graham, D. Y. et al. *Campylobacter pylori* detected noninvasively by the ^{13}C-urea breath test. *Lancet* **329**, 1174–1177 (1987).
First description of the ^{13}C-urea breath test to diagnose *Helicobacter pylori* infection. This breath test was approved in 2010 and is used clinically worldwide.
178. Covacci, A., Telford, J. L., Giudice, G. D., Parsonnet, J. & Rappuoli, R. *Helicobacter pylori* virulence and genetic geography. *Science* **284**, 1328–1333 (1999).
179. Marshall, B. *Helicobacter* connections. *ChemMedChem* **1**, 783–802 (2006).
180. El-Omar, E. M. et al. Interleukin-1 polymorphisms associated with increased risk of gastric cancer. *Nature* **404**, 398–402 (2000).
181. Graham, D. Y. & Klein, P. D. Accurate diagnosis of *Helicobacter pylori*: ^{13}C-urea breath test. *Gastroenterol. Clin. North Am.* **29**, 885–893 (2000).
182. Li, Z.-X. et al. Cut-off optimization for ^{13}C-urea breath test in a community-based trial by mathematic, histology and serology approach. *Sci. Rep.* **7**, 2072 (2017).
183. Gomollón, F. et al. Breath test is very reliable for diagnosis of *Helicobacter pylori* infection in real clinical practice. *Dig. Liver Dis.* **35**, 612–618 (2003).
184. Som, S. et al. Excretion kinetics of ^{13}C-urea breath test: influences of endogenous CO_2 production and dose recovery on the diagnostic accuracy of *Helicobacter pylori* infection. *Anal. Bioanal. Chem.* **406**, 5405–5412 (2014).
185. Moayyedi, P. et al. Do patients need to fast for a ^{13}C-urea breath test? *Eur. J. Gastroenterol. Hepatol.* **9**, 275–277 (1997).
186. Eisdorfer, I., Shalev, V., Goren, S., Chodick, G. & Muhsen, K. Sex differences in urea breath test results for the diagnosis of *Helicobacter pylori* infection: a large cross-sectional study. *Biol. Sex. Differ.* **9**, 1 (2018).
187. Peng, N.-J. et al. Clinical significance of oral urease in diagnosis of *Helicobacter pylori* infection by [^{13}C] urea breath test. *Dig. Dis. Sci.* **46**, 1772–1778 (2001).
188. Cummings, J. H. & Macfarlane, G. T. The control and consequences of bacterial fermentation in the human colon. *J. Appl. Bacteriol.* **70**, 443–459 (1991).
189. Fenchel, T., King, G. M. & Blackburn, T. H. in *Bacterial Biogeochemistry* 3rd edn (eds Fenchel, T., King, G. M. & Blackburn, T. H.) 1–34 (Academic Press, 2012).
190. Houben, E., De Preter, V., Billen, J., Van Ranst, M. & Verbeke, K. Additional value of CH_4 measurement in a combined $^{13}C/H_2$ lactose malabsorption breath test: a retrospective analysis. *Nutrients* **7**, 7469–7485 (2015).

Acknowledgements
The authors thank the Ligue Contre le Cancer (Comité de la Vienne), the Cancéropôle Grand Sud-Ouest and the Région Nouvelle-Aquitaine for their financial support. They also thank Claude Geffroy for her precious help. Finally, the authors gratefully thank Sébastien Papot for his very wise advice.

Author contributions
P.P. and F.D. researched data and literature for the article. All authors contributed to the discussion of content and writing. P.P. wrote the first version of the manuscript. All authors prepared the figures and reviewed and edited the manuscript before submission.

Competing interests
The authors declare no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s41570-020-00248-z.

© Springer Nature Limited 2021

Objectifs et stratégie

Objectifs et stratégie

L'étude bibliographique a mis l'accent sur l'importance de développer de nouveaux outils et stratégies pour mieux comprendre à l'échelle de la molécule, les processus liés au développement des maladies, telles que le cancer. En effet, ces nouvelles stratégies devraient permettre de faciliter le diagnostic comme le pronostic de cette maladie, ou encore de s'assurer de l'efficacité des traitements associés. Suivant cet objectif, les approches dites « omiques » proposent un large choix d'outils qui permettent de suivre en temps réel des particularités biochimiques ou structurales de la tumeur, « allumant » ainsi des processus liés au développement ou la régression de celle-ci.

Dans ce contexte, mon travail de thèse à consister à développer de nouvelles approches impliquant des stratégies et outils basés sur la spectrométrie de masse afin d'explorer au niveau moléculaire des processus biologiques liés au développement de maladies, ou impliqués dans leur régression. Dans ce cadre, nous nous sommes tout particulièrement intéressés à développer des outils moléculaires et/ou stratégies analytiques permettant l'exploration de processus caractéristiques des cancers solides.

Mon travail de thèse s'articule autour de trois objectifs principaux.

Le premier objectif, constituant le plus gros de mon projet, correspond au développement de sondes à base de COVs, ciblant chacune une particularité enzymatique spécifique des tumeurs solides, afin de les détecter précocement. Ce concept de volatolomique induite, décrit dans la revue ci-avant, repose sur l'utilisation de sondes off-on non toxiques, et incluant dans leur structure un COV. Ces sondes seraient métabolisées sélectivement dans les tissus cancéreux *via* l'activité d'enzymes marqueurs de la pathologie. Les COVs exogènes ainsi libérés deviendraient à leur tour des traceurs de la pathologie et seraient détectés, *in vivo*, dans l'haleine des individus, ou *ex vivo*, sur prélèvements biologiques. Une première étude ciblant la β-glucuronidase, marqueur tumoral, a montré la pertinence de l'utilisation d'une sonde à éthanol marquée au deutérium (le D_5-éthyl-β-D-glucuronide) pour détecter *in vivo* différents types de tumeurs et suivre leur évolution au cours d'une chimiothérapie. Afin d'améliorer la robustesse de cette stratégie et limiter le nombre de faux-positifs qui peuvent être induits par le ciblage d'une seule et unique enzyme, nous avons ici développé un cocktail de sondes activables chacune par une glycosidase, marqueur des cancers.
L'objectif de cette approche multi-cibles est d'améliorer la robustesse et la sensibilité du diagnostic, en corrélant l'activation d'une ou plusieurs sonde(s) avec la présence d'une ou plusieurs activité(s) enzymatique(s). Après une première preuve de concept *in vivo*, nous avons commencé à travailler sur un protocole clinique *ex vivo*, basé sur l'ajout du cocktail dans un prélèvement sanguin de patient. Les premières étapes de développement de ce protocole *ex vivo* sont également présentés.

En parallèle, un protocole d'extraction et de purification des glycosidases ici ciblées a été développé. L'objectif était de cartographier ces activités enzymatiques sur tissus. Cette étude devait nous permettre de déterminer l'origine cancéreuse ou immunitaire des enzymes suivant leurs séquences primaires et modifications post-traductionnelles. Dans le cadre de ce travail, une stratégie de purification reposant sur le fractionnement des protéines suivant leur masse moléculaire a été développée.

Ainsi, une approche protéomique de type « middle-down » a été suivie. L'idée était ici d'éliminer les protéines les plus abondantes (ex: albumine, chaînes de collagène, fibronectine…) interférant avec la détection des glycosidases présentes en faible concentration. En parallèle, une méthode d'analyse sur UPLC-HRMS ou UV a été mise au point afin de détecter chaque enzyme, puis déterminer leur origine (cancéreuse ou immunitaire) par séquençage protéomique.

Le deuxième volet de ce travail est consacré au développement d'une approche d'extraction, de purification et d'analyse de la doxorubcine, agent cytotoxique intercalant entre les bases de l'ADN. Cet agent anticancéreux mais efficace quelles que soient les matrices étudiées (environnementales ou biologique).

Enfin, nous avons mis au point une stratégie d'analyse d'une classe émergente d'agents thérapeutiques basés sur des petits oligoribonucléotides (conjugués acides aminés-ARN).

Partie B. Matériel et Méthodes

Matériel et méthodes

Objectif expérimental

Face aux faux-positifs et la variabilité interindividuelle, nous avons cherché à développer une méthode plus robuste et universelle, afin d'améliorer le diagnostic du cancer. Les cibles choisies dans cette étude sont les glycosidases, car potentiellement présentes dans le microenvironnement tumoral. Un nouveau concept de volatolomique induite basée sur détection de sondes à COVs a été développé. Des enzymes marqueurs du cancer et présentes à l'état de trace pourraient être détectées dans les différents réservoirs. Parallèlement à la volatolomique induite pour détecter ces protéines minoritaires, nous avons développé une stratégie d'isolation de ces protéines qui sont très souvent en quantité minime comparées aux protéines majoritaires telles sur l'albumine ou la fibronectine. Une stratégie de discrimination des protéines minoritaires et majoritaires a été établie par protéomique en utilisant une digestion enzymatique et une digestion chimique, suivie d'une analyse des peptides par UPLC-HRMS.

I. Matériel chimique, biochimique, et biologique

Matériel chimique: L'ensemble des produits chimiques provient de Sigma Aldrich (Saint Louis, USA) et Thermo Fischer Scientific (Waltham, USA). Les sondes à COVs ont été synthétisées par l'entreprise @rtMolecule (Poitiers, France) d'une part et l'équipe Orgasynth de l'IC2MP d'autre part. Pour l'analyse des COVs par SPME, des fibres carboxen/polydimethylsiloxane (CAR/PDMS) 75 µm (Supelco, Bellffone, USA) ont été utilisées (Tableau 8 ci-après).

Matériel biochimique et biologique: Plusieurs glycosidases commerciales ont été utilisées : la β-Glu (E.*coli* K12) , la β-GlcNac (foie bovin), β-Gal (E.*coli*), β-glucosidase (amande). L'albumine bovine (BSA) et humaine (HSA), le lysosyme de blanc d'œuf, la Rnase bovine ont aussi été utilisées. Les protéines ont été fournies par Sigma Aldrich (Saint Louis, USA) en plus d'un mélange de peptides standards. Ce mélange également fourni par Sigma Aldrich contient de l'angiotensine II, le dipeptide Gly-Tyr, la Leu-enképhaline, la Met-enképhaline, et le tripeptide Val-Tyr-Val. Ce mélange été utilisé comme contrôle positif du bon déroulement de la digestion trypsique.

Le Tableau 8 répertorie les caractéristiques des protéines commerciales utilisées.

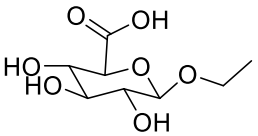
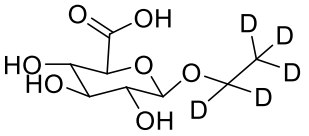
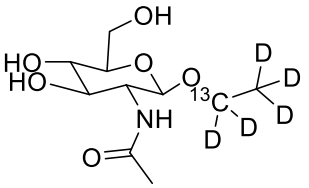
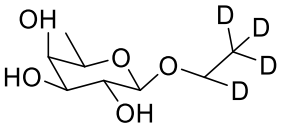
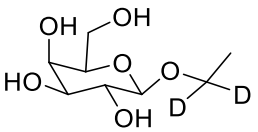
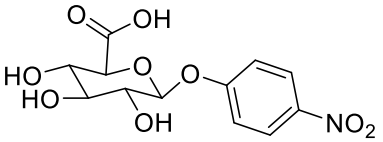
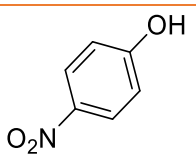
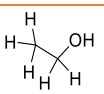
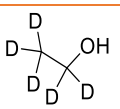
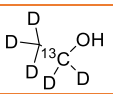
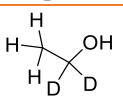
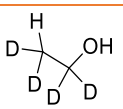
Composé	Formule	Masse molaire ($g.mol^{-1)}$	Numéro CAS	Structure
Éthyl-glucuronide (EtG)	$C_8H_{14}O_7$	222.19	17685-04-0	
Éthyl-beta-D-glucuronide- D_5 (EtGlu-D_5)	$C_8D_5H_9O_7$	227.22	1135070-98-2	
N-acetyl-β-D-glucosamide-$^{13}CD_5$ (EtGlcNac-$^{13}CD_5$)	$C_9{}^{13}CH_{14}D_5NO_6$	255.16	/	
Éthyl-α-D-fucoside- D_4 (.-D_4)	$C_8H_{12}D_4O_5$	196.12	/	
Ethyl-β-D-galactoside- D_2 (EtGal-D_2)	$C_8H_{14}D_2O_6$	210,11	/	
Glucuronide de para-nitrophénol	$C_{12}H_{13}NO_9$	315.23	10344-94-2	
Para-nitrophénol	$C_6H_5NO_3$	139.11	100-02-7	
EtOH	C_2H_6O	46.07	64-17-5	
EtOH-D_5	C_2D_5HO	51.10	1859-08-1	
EtOH-$^{13}CD_5$	$^{13}CC_2D_5OH$	52.09	/	
EtOH-D_2	$C_2H_4D_2O$	48.08	1859-09-2	
EtOH-D_4	$C_2H_2D_4O$	50.09	22544-43-0	

Tableau 7: Formules chimiques des sondes à COVs et COVs correspondants

Protéine	État		Unité/mg de protéine
β-glu(E.*coli* K12)	liquide		80
β-GlcNac (foie bovin),	liquide		10-50
β-Gal (E.*coli*),	liquide		5000
β-glucosidase (amande).	poudre		2
BSA	poudre		/
HSA	poudre		/
Lysosyme	poudre		40 000
RNase	poudre		110-188

Tableau 8: Caractéristiques des protéines commerciales

Enzyme cible du cocktail de sondes	Sonde associée	COV libéré	Cancer associé	Source
β-Glucuronidase (β-Glu)	EtGlu-D_5	Éthanol D_5 (ETOH-D_5)	Colon Poumon Sein Rein	Awolade et.al, 2019[163]
β-N-Acetyl-glucosaminidase (β-GlcNac)	EtGlcNac-$^{13}CD_5$	Éthanol ^{13}C-D_5 (ETOH-$^{13}CD_5$)	Cancer colorectal Cancer du sein	Phueaouan et.al., 2013[164] Chaiyawat et. al., 2014[165]
β-galactosidase (β-Gal)	EtGal-D_2	Éthanol-D_2 (ETOH-D_2)	Ovaire	Asanuma et. al, 2015[166]
α-Fucosidase (α-Fuc)	EtFuc-D_4	Éthanol-D_4 (ETOH-D_4)	Colon Foie	Deugnier et. al, 1984[167], Shuang et.al 2018[168]

Tableau 9: Enzymes cibles associées à leurs sondes

Deux lignées cellulaires cancéreuses différentes ont été étudiées puis implantées sous forme de xénogreffes à des souris :

- une lignée cellulaire humaine KB (cancer de l'utérus, ATCC CCL-17) fournie par LGC Standards (Molsheim, France) ;
- une lignée cellulaire humaine MDA-MB-231 (adénocarcinome du sein, ATCC HTB-26) fournie par Sigma Aldrich

Le modèle de souris suivant a permis l'étude de concept in vivo :

- des souris athymiques nude-Foxn1nu (Envigo, Gannat, France), qui possèdent un système immunitaire déficient ;

Toutes les procédures expérimentales ont été validées par les comités régionaux et nationaux d'éthique animale. Elles sont référencées sous le numéro 2016082411329149 et réalisées sous la responsabilité légale de J. Clarhaut (niveau A d'expérimentation animale R-45GRETA-F1-10) et en accord avec la Directive 2010/63/UE du Conseil Européen et de son décret d'application 2013-118 du Ministère Français de l'Agriculture et de la Forêt

II. Détection des glycosidases par volatolomique induite

L'objectif de cette étude était de pouvoir cibler des enzymes grâce à l'hydrolyse de sondes qui leur sont spécifiques. Après leur hydrolyse les sondes libèrent de l'éthanol marqué, utilisé comme traceur d'activité. La présentation des enzymes figure dans le ***Tableau 9***.

Enzyme cible du cocktail de sondes	Sonde associée	COV libéré	Cancer associé	Source
β-Glucuronidase (β-Glu)	EtGlu-D5	Éthanol D5 (ETOH-D5)	Colon Poumon Sein Rein	Awolade et.al, 2019163
β-N-Acetyl-glucosaminidase (β-GlcNac)	EtGlcNac-13CD5	Éthanol $^{13C\text{-}D5}$ (ETOH-13CD5)	Cancer colorectal Cancer du sein	Phueaouan et.al., 2013164 Chaiyawat et. al., 2014165
β-galactosidase (β-Gal)	EtGal-D2	Éthanol-D2 (ETOH-D2)	Ovaire	Asanuma et. al, 2015166
α-Fucosidase (α-Fuc)	EtFuc-D4	Éthanol-D4 (ETOH-D4)	Colon Foie	Deugnier et. al, 1984167, Shuang et.al 2018168

II.1. Recherche de la limite de détection des éthanols marqués : cas de l'éthanol D_5

Une gamme de concentration de cinq points a été réalisée. Des concentrations allant de 5.10^{-9} M à 10^{-6} M d'EtOH-D_5 ont été réalisées. Pour ce faire, dans un vial SPME de 2 mL, 10 µL de solution d'éthanol à une certaine concentration ont été mélangés avec 90 µL de tampon acétate 0.1 M pH 5. Une extraction SPME a été réalisée pour une durée de 30 min (Figure 35).

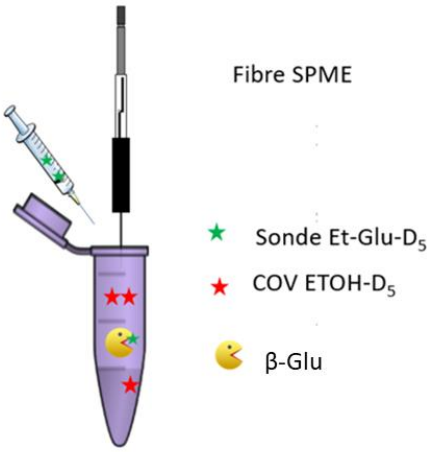


Figure 35: Schéma expérimental de l'analyse des COVs par SPME GC/MS-MS

II.2. Étude de la détection des glycosidases par les sondes à COVs sur matrice plasma : cas de la β-Glu

La LOD de la β-Gluc *E. coli* en matrice plasma en présence de la sonde l'EtGlu-D_5 a été déterminée. Pour cela, des concentrations d'enzyme allant de 1.10^{-11} à 1.10^{-13} M ont été réalisées. La quantité relative d'ETOH-D_5 a été suivie avant ajout de l'EtGlu-D_5 (T0) et 2.5, 4.5 et 7.5 h (T 2.5 h, T 4. 5 h et T 7.5 h) après ajout de la sonde.

Une droite d'étalonnage permettant la détection de d'ETOH-D_5 suite à l'hydrolyse de la sonde l'EtGlu-D_5 par des concentrations croissantes d'enzyme a été réalisée. L'hydrolyse de l'EtGlu-D_5 après 2.5 h, 4.5 h et 7.5 h de réaction a été réalisée. Les points de gamme ont été faits en triplicat.

II.3. Étude de la détection des glycosidases par une méthode de référence : cas de la β-Glu.

La méthode de référence de détection de la β-Glu ici utilisée est la détection du paranitrophénol. Elle est basée sur l'hydrolyse du paranitrophénol de glucuronide (PNPG) en paranitrophénol (PNP) par la β-Glu. Un volume de 25 µL de plasma humain a été mis en présence de 25 µL de glucuronide de paranitrophénol (PNPG) à $0.5.10^{-1}$ M, ainsi que 4 µL de β-Glu *E. coli* à 6.10^{-7} M et 196 µL de tampon acétate 0.1 M pH 5. En parallèle, un blanc contenant 25 µL de plasma, 25 µl de PNG 10^{-2} M et 200 µL de tampon acétate a été préparé. Pour chaque échantillon les protéines ont été précipitées avec 125 µL d'ACN/MeOH (2/1 en v/v). Le surnageant contenant le para-nitrophénol a été centrifugé 5 min à 3180 g, puis concentré à l'évaporateur rotatif pendant 15 min à 40 °C. Aux 50 µL de volume final après évaporation ont été ajoutés 600 µL de NaOH pour augmenter l'intensité d'absorption des ions paranitrophénolate. Le mélange a ensuite été dilué au 8ème avec de l'eau ultrapure. La mesure de l'absorbance a été réalisée avec un spectrophotomètre UV 1280 (Shimadzu, Marlborough, USA). Pour établir la limite de détection du paranitrophénol dans cette étude, des concentrations de β-Glu différentes ont été préparées et ajoutées à la même concentration de sonde, dans le plasma. Une gamme allant de 10^{-10} M à 5.10^{-9} M en enzyme en a été réalisée. La réaction a été faite en triplicat.

II.4. Optimisation des conditions de manipulation des prélèvements sanguins : préservation de la stabilité de la β-Glu

Effet de l'inhibiteur de protéases

Dans 100 µL de tampon acétate pH 5, la β-Glu *coli* a été mise pour une concentration de 10^{-9} M avec 25 µL de plasma. À cela a été ajoutée la sonde d'EtGlu-D_5 à 5.10^{-5} M. La réaction d'hydrolyse d'EtGlu-D_5 conséquente à la libération d'EtOH-D_5 a été suivie par SPME-GC-MS/MS aux points de cinétique suivants : 0 h, 0.5 h, 2.5 h et 4.5h. Trois conditions différentes ont été testées en triplicat, une hydrolyse sans inhibiteur de protéases, une avec inhibiteur en concentration 1X, puis avec inhibiteur en concentration 2X. Dans cette jauge de concentration, X correspond au facteur de dilution nécessaire pour utiliser la solution. L'inhibiteur était fourni par Roche de Bâle en Suisse (Cocktail d'inhibiteurs de protéases sans EDTA Mini complete). Ce cocktail inhibiteur cible les protéases à Sérines et cystéines.

Effet de la congélation/décongélation répétées

25 µL de plasma ont été mis en présence de 25 µL de sonde d'EtGlu-D_5 à 5.10^{-5} M en plus de l'inhibiteur 1X et de la β-Glu *coli* 10^{-9} M. Au premier jour (J0), l'échantillon a été analysé par SPME-GC-MS/MS puis congelé à – 20°C. Au jour suivant (J2), l'échantillon a été décongelé pendant une heure à température puis analysé en SPME sur une cinétique de 7 h: 0 h, 2 h, 4 h, 7 h. La même opération a été réalisée le jour d'après (J3).

Effet de la durée de congélation

25 µL de plasma a été mis en présence de 25 µL de sonde d'EtGlu-D_5 à 5.10^{-5} M en plus de l'inhibiteur 1X et de la β-Glu *coli* 10^{-9} M. Au premier jour (J0), l'échantillon a été analysé par SPME-GC-MS/MS sur un point de 2 h. Il a ensuite été congelé à -20 ° C. Une semaine plus tard, (J7), l'échantillon a été décongelé puis additionné avec la même concentration d'EtGlu-D_5 puis analysé par SPME-GC/MS-MS sur un point de 2 h.

II.5. Détection des glycosidases dans les prélèvements de plasma via les sondes à COVs : cas de la β-Glu

En collaboration avec le Centre d'Investigation Clinique (CIC) du CHU de Poitiers (INSERM 1402, CHU Poitiers), une étude clinique sur des échantillons de plasma de 25 patients a pu débuter. L'objectif était de vérifier si les échantillons de plasma de patients sains pouvaient être discriminés de plasmas de patients avec un cancer, par la présence de β-Glu dans ces derniers. Différents critères d'inclusion ont été pris en compte pour sélectionner les patients : le patient devait être majeur, sans tutelle ou curatelle, il devait bénéficier d'un régime de Sécurité Sociale et ne devait pas s'opposer à participer à l'étude. Les caractéristiques de l'étude sont présentées dans le Tableau 10.

Numéro d'anonymat	Pathologies	Date de prélèvement
12	Cancer du poumon	22/10/2019
13		15/11/2019
14		26/11/2019
15		28/11/2019
16		29/11/2019
27		27/12/2019
29		10/01/2020
61		18/08/2020
62		21/08/2020
63		21/08/2020
17	Cancer du sein	05/12/2019
18		05/12/2019
19		05/12/2019
21		12/12/2019
22		12/12/2019
23		12/12/2019
24		13/12/2019
25		19/12/2019
26		19/12/2019
28		02/01/2020
31		16/01/2020
20	Sain	10/12/2019
48		05/06/2020
51		01/07/2020
56		16/07/2020

Tableau 10: Caractéristiques des échantillons de plasma étudiés pendant l'étude clinique

Deux tubes héparinés contenant environ 1 mL de plasma pour chaque patient ont été congelés à – 20 °C. Dans un vial SPME, 25 µL de plasma ont été additionnés de 50 µL de tampon acétate 0.1M pH 5. Un volume de 25 µL l'ETOH-D_5 à 2.10^{-4} M a été ajouté au plasma. Une fibre SPME a été mise dans le vial après 30 min, 2h, 4h et 7h d'incubation à 37°C. S'en est suivie l'analyse SPME-GC-MS/MS à chaque point de cinétique.

II.6. Détection des glycosidases sur biopsies de tumeurs et foies via les sondes à COVs : cas de la β-Glu

Une gamme de concentration a été réalisée avec des quantités croissantes de sonde ajoutée aux biopsies cancéreuses ou saines pour étudier la pénétration intracellulaire. Premièrement, un tampon PBS a été préparé (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4). Une gamme de concentration a été réalisée avant chaque essai, en allant de 10^{-9} à 10^{-4} M en cinq points. Un blanc de mesure a été effectué pendant 10 min. Il consiste à effectuer un prélèvement SPME dans un vial vide et de détecter en GC-MS/MS des potentielles interférences sur la mesure de l'ETOH-D_5 (Figure 36). Ensuite, une biopsie entière a été mise dans du tampon PBS (Tampon phosphate salin). Afin de s'assurer qu'il n'y ait aucune source d'ETOH-D_5 endogène à ce tissu, un prélèvement SPME de la phase gazeuse au-dessus du tissu était réalisé pendant 30 min.

Un volume de 3 µL de la sonde d'EtGlu-D_5 à 10^{-7}M a été ajouté au vial. Le vial a été aussitôt fermé hermétiquement pour éviter la fuite d'ETOH-D_5 libéré par la réaction enzymatique. Le vial a ensuite été vortexé et placé au bain marie à 37 °C. Une fibre SPME était insérée au niveau de la phase gazeuse du vial pour 30 min d'extraction, à la suite desquelles la fibre était analysée par GC-MS/MS. La même manœuvre a été opérée pour faire les quatre autres points de gamme en duplicat.

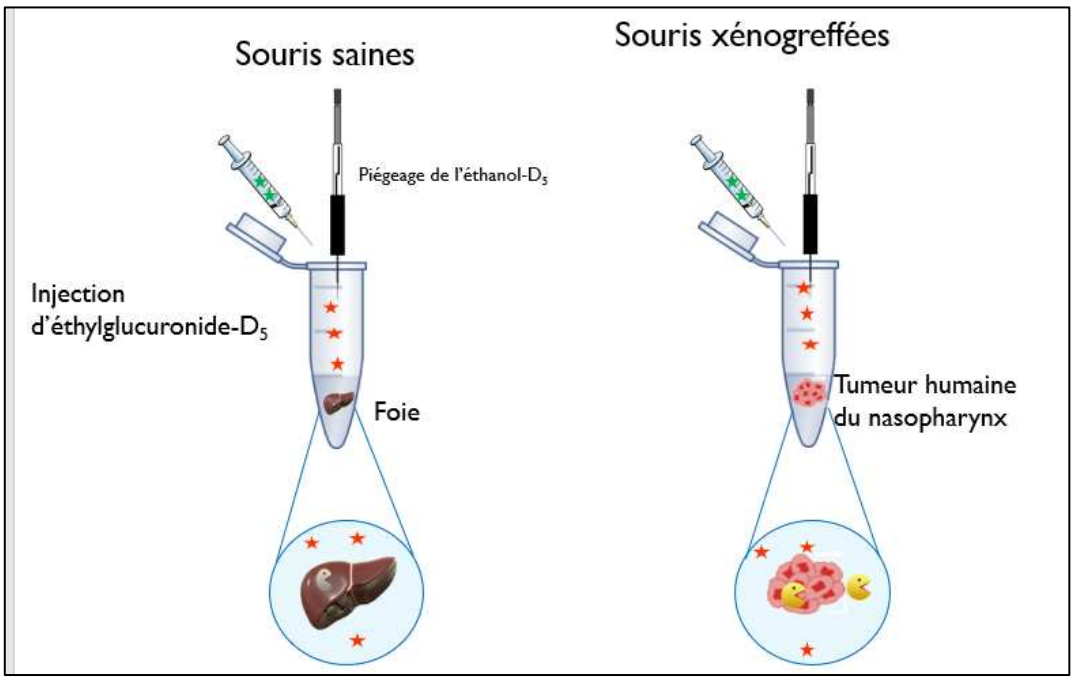


Figure 36: Schéma expérimental de la détection de la cible sur biopsies

II .7. Développement d'un cocktail de sondes : étude de la spécificité de chaque sonde et leur cible enzymatique

Afin d'améliorer la robustesse de cette stratégie et limiter le nombre de faux-positifs qui peuvent être induits par le ciblage d'une seule et unique enzyme, un cocktail de sondes activables chacune par une glycosidase, marqueur des cancers, a été développé. La spécificité des glycosidases pour chaque sonde a été vérifiée en faisant les croisements de sondes (Tableau 12).

	Mélange réactionnel
Test auto-hydrolyse des sondes	EtGlcNac-$^{13}CD_5$ et EtGlu-D_5 à 5.10^{-5}M dans un tampon acétate, $T_{2\ h\text{-}2\ h30}$
Une enzyme + une sonde spécifique (Condition C Figure 37)	• β-GlcNac 10^{-8}M **+** EtGlcNac-$^{13}CD_5$ 5.10^{-5}M, $T_{2\ h\text{-}2\ h30}$ • β-Glu 10^{-8}M **+** EtGlu-D_5, 5.10^{-5}M, $T_{2\ h\text{-}2\ h30,}$
2 enzymes + 1 sonde (Condition D Figure 37)	• β-Glu et β-GlcNac 10^{-8}M **+** EtGlu-D_5 à 5.10^{-5}M, $T_{2\ h\text{-}2\ h30}$ • β-Glu et β-GlcNac 10^{-8}M **+** EtGlcNac-$^{13}CD_5$ à 5.10^{-5}M, $T_{2\ h\text{-}2\ h30}$
1 enzyme +2 sondes (Condition E Figure 37)	• β-GlcNac 10^{-8}M **+** EtGlc-$^{13}CD_5$ et EtGlu-D_5, 5.10^{-5}M • β-Glu 10^{-8}M **+** EtGlu-D_5 et EtGlcNac-$^{13}CD_5$, 5.10^{-5}M, $T_{2\ h\text{-}2\ h30}$
2 enzymes + 2 sondes (Condition F Figure 37)	• β-Glu et β-GlcNac 10^{-8}M **+** EtGlcNac-$^{13}CD_5$ et EtGlu-D_5 à 5.10^{-5}M, $T_{2\ h\text{-}2\ h30}$

Tableau 11: Mode opératoire de l'expérience du croisement de sondes

Pour chaque mélange réactionnel, le volume final était de 100 µL dans du tampon acétate pH 5. Une extraction SPME-GC-MS/MS a été réalisée pour chaque condition (cf. paramètres Tableau 11).

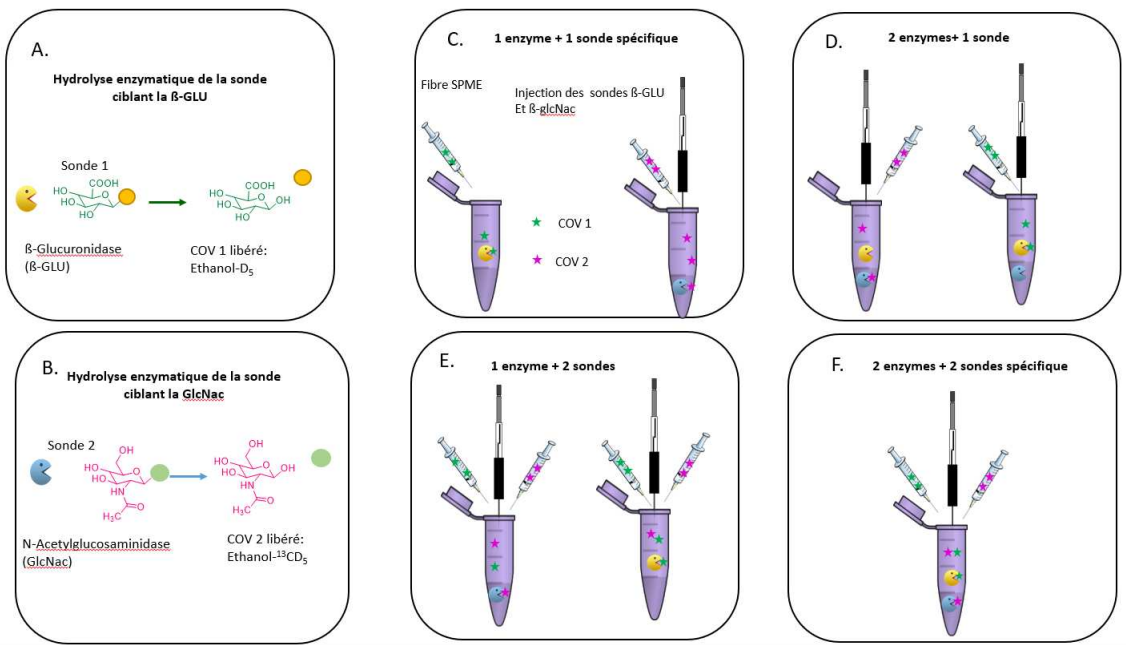


Figure 37: Schéma expérimental du croisement de sondes

***A**. Hydrolyse enzymatique de la ß-Glu; **B**. Hydrolyse de la ß-GlcNac; **C.** afin de vérifier la spécificité de chaque enzyme pour chaque sonde, la ß-Glu a été mise en présence de sa sonde 1, et la ß-GlcNac a été mise en présence de sa sonde 2; **D.** les deux enzymes sont mises en présence de l'une des deux sondes. Par échantillon, une seule réponse de COV est donc attendue, car un seul couple enzyme-sonde est présent, l'enzyme restante ne dois pas hydrolyser une sonde qui n'est pas la sienne; **E.** une seule enzyme est mise en présence des deux sondes, elle hydrolyse alors sa sonde qui libère son COV associé. **F.** les deux enzymes sont en présence des deux sondes, les deux COV 1 et COV2 sont alors libérés.*

II.8. Détection des cibles enzymatiques en utilisant un cocktail de sondes *in vivo*

Afin d'établir le protocole final, les conditions de prélèvement (durée de blancs, durée de pose de la souris dans la cage, durée de prélèvement) des COVs ont tout d'abord été optimisées.

Pour l'étude, un cocktail constitué des sondes ciblant la β-Glu, la β-GlcNac, la β-Gal, et la α-Fuc, a été administré à 6 souris athymiques Nude, préalablement xénogreffées avec des tumeurs KB. La mesure des COVs dans l'haleine des souris a été faite 12 jours après la xénogreffe pour identifier la part de COVs générée par le métabolisme des souris. En mesurant cette activité métabolique basale, on s'affranchit des faux positifs.

Afin de déterminer la dose optimale des sondes à COVs à injecter aux souris, des concentrations croissantes de sondes ont été administrées à des souris saines Nude. Premièrement, un blanc d'expérience a été réalisé en plaçant une souris dans une cage hermétique dont les dimensions étaient de 130 mm x 130 mm x 200 mm, et ceci pendant 10 min. Pendant ce laps de temps, une fibre SPME a été insérée dans l'espace de tête de la cage afin détecter les COVs qui pouvaient interférer avec la détection d'isotopes d'éthanol non désirés dans l'atmosphère. Une fois les blancs effectués, le cocktail de sondes a été injecté aux souris. Au bout d'une heure, un piégeage des COVs par SPME dans l'haleine des souris a été réalisé, pendant 30 min. Chaque dose a été injectée à deux souris.

En raison de la détection d'une molécule dont le signal interférait avec celui de l'ETOH-D_2, les signaux de cet isotope ont été soustraits à ce bruit de fond.

De plus, un délai de trois jours entre une première et une seconde injection a été mis en place pour mettre en évidence l'évolution de l'activité des glycosidases. Lors de l'optimisation d'un autre type de cages de petite taille a été testé. Ces petites cages ont pour dimension 130 x 130 x 100mm.

II.9. Paramètres GC-MS/MS

L'étude des composés volatils a été réalisée avec une chromatographie gazeuse couplée à un triple quadripôle équipé d'une source EI (TSQ 9000, Thermo Fisher Scientific). Le système était contrôlé par le logiciel TSQ 9000 et les résultats ont été analysés avec le logiciel Xcalibur (Thermo Fisher Scientific). La séparation des COVs s'est faite sur une colonne non polaire DB-624 (60 m x 250 µm x 1,4 µm) fournie par Agilent (Santa Clara, USA). L'ensemble des paramètres GC-MS sont répertoriés dans les tableaux suivants (Tableau 13 et Tableau 13).

Composé	Précurseur m/z	Product m/z	Energie de collision (eV)
EtOH-D_5	51.1	33.1	10
EtOH-$^{13}CD_5$	25.1	34.1	6
EtOH-D_2	48.1	33.1	5
EtOH-D_4	50.1	32.1	10

Tableau 12: Transitions des éthanols marqués

	Paramètres GC-MS	Méthode 1
PARAMÈTRES GC	Température de l'injecteur (°C)	280
	Septum purge et gaz saver ($mL.min^{-1}$)	5 - 20
	Débit du gaz - hélium ($mL.min^{-1}$)	1.2
	Split/Splitless	Splitless
	Programme GC	0 – 7 min : 90 °C
PARAMÈTRES MS	Mode	SRM
	Energie d'ionisation (kV)	70
	Vitesse de balayage ($scans.s.^{-1}$)	20
	Température de l'injecteur, de la ligne de transfert et de la source (°C)	280 – 260 – 260

Tableau 13: Paramètres GC et MS de la méthode permettant l'analyse des COVs

III. Extraction, isolation, et détection des glycosidases dans les prélèvements biologiques

Dans les milieux biologiques, parmi les protéines majoritaires se trouvent l'albumine, la globuline et le fibrinogène. Par exemple, dans le plasma l'albumine est majoritaire avec une concentration de 35-50 $g.L^{-1}$, suivie de la globuline avec 20-35 $g.L^{-1}$, puis le fibrinogène (2-4 $g.L^{-1}$).[169–171] Les glycosidases que nous cherchons sont en quantité minime. En raison de ces différences de concentration, une méthode permettant la séparation des protéines majoritaires et minoritaires a été développée.

III.1. Détection de la β-Glu par protéomique *middle-down* et *bottom-up*

III.1.A. Recherche des peptides (*in silico*)

Les logiciels utilisés pour établir la séquence des peptides après digestion sont les suivants :

- MS digest(https://prospector.ucsf.edu/prospector/cgibin/msform.cgi?form=msdigest)
- Expasy (https://web.expasy.org/peptide_cutter/)
- Mascot (https://www.matrixscience.com/)
-Skyline
- MaxQuant
- MMass
- MassXpert

La majorité des logiciels de protéomique étant centrés sur la digestion trypsique, un vrai travail de recherche a été réalisé pour identifier les peptides de l'acide 2-nitro-5-thiocyanatobenzoique (NTCB). L'ajout du mode de coupure, et des modifications selon la nomenclature unimod ont été faits manuellement.

III.1.B. Protocole digestion trypsique

Une quantité de 0.1 nmole de BSA a été mise en suspension dans 100 µL de tampon bicarbonate d'ammonium 50 mM Tris 10 mM pH 8.5. Un volume de 1 µL de mélange de peptides standard à 1 $\mu g.\mu L^{-1}$ a été ajouté (mélange de peptides qui servira de standard de digestion pour l'analyse LC-MS), puis la suspension a été vortexée. Les protéines ont subi une réduction des ponts disulfures avec l'addition 15 µL de DTT (Dithiothréitol) à 15.4 $mg.mL^{-1}$, suivie d'une incubation d'1h sous agitation à 60 °C. Afin d'empêcher la reformation de ponts disulfures, la réduction étant réversible, 15 µL d'agent alkylant, IAA (Iodoacétamide) à 44 $mg.mL^{-1}$, ont été ajoutés à l'échantillon. Le mélange réactionnel a été incubé à température ambiante, à l'obscurité pendant 45 min. Un volume de 100 µL de tampon bicarbonate d'ammonium 50 mM Tris 10 mM pH 8.5 a ensuite été ajouté. L'échantillon a été vortexé. Les protéines ont été digérées avec 11 µL de trypsine 0.1 $\mu g.\mu L^{-1}$. La digestion a été réalisée pendant 24 h à 37°C. L'activité de la trypsine a été stoppée par de l'acide formique à 0,4 % (concentration finale dans l'eppendorf).

III.1.C. Protocole d'hydrolyse NTCB avec réduction au DTT en atmosphère inerte et clivage à la soude (Crimmins et. al, Srzentiv et.al)[147,172]

Une quantité de 0.1 nmole de BSA a été mise en suspension 100 µL de tampon Tris acétate 200 mM pH 8 dans un tube fermé hermétiquement. Une arrivée de gaz N_2 a été connectée aux tubes réactionnels à l'aide d'une seringue pour être en milieu anaérobie/inerte. Pour la réduction de tous les ponts disulfures, les protéines ont été additionnées de 50 µL de DTT
($2.5.10^{-5}$ M) à l'aide d'une seringue. Les échantillons ont été incubés 1 h à température ambiante. S'en est suivie l'addition à la seringue de 50 µL de NTCB ($1.5.\ 10^{-3}$ M) à 40 °C pendant 30 min sur plaque chauffante, en milieu anaérobie/inerte. Dans un second échantillon, l'étape de réduction au DTT n'a pas été effectuée.

Après l'addition du NTCB, le mélange passe de l'incolore au jaune. Les peptides obtenus sont alors précipités dans de l'acétone froid à pH 3, avec une incubation toute une nuit à -20 °C. Les échantillons étaient centrifugés 5 min à 15000 rpm à 4°C. Après pipetage délicat du surnageant, les culots ont été lavés avec 500 µL d'acétone acidifié. Après une seconde centrifugation, les peptides étaient séchés, tubes ouverts, à température ambiante afin d'évaporer tout l'acétone résiduelle. Ensuite, les peptides étaient resuspendus dans 20 µL de tampon de borate de sodium 100 mM (pH 9) et 20 µL d'eau. Le mélange réactionnel était incubé sous agitation 1 h à 50 °C. Après 1 h d'incubation, le pH des échantillons devait être égal à 9, sinon le pH est ajusté avec quelques gouttes de soude 2 mM. Les échantillons étaient à nouveau mis à incuber sous agitation 1 h à 50 °C. Suivant ce protocole, des paramètres ont été optimisés afin d'améliorer l'hydrolyse, qui sont :

- La réaction ne s'est pas faite en atmosphère inerte, mais sous pression atmosphérique
- L'arrivée de gaz N_2 sur le milieu réactionnel s'est effectuée par un flux de gaz dirigé par entonnoir, formant un dispositif en cloche

III.1.D. Protocole hydrolyse NTCB avec réduction au TCEP(tris(2-carboxyéthyl)phosphine) et clivage au NH4OH en pression atmosphérique (Belghazi et al)[149]

Une quantité de 1 nmole de BSA a été suspendue dans 200 µL de tampon bicarbonate d'ammonium 100 mM, TCEP 50 mM pH 8. La réduction au TCEP s'est faite par incubation du mélange sous agitation pendant 45 min à 56 °C. Après avoir laissé refroidir le mélange, 5.2 % (v/v) de NTCB 1.6 $mg.mL^{-1}$ ont été ajoutés. La réaction de cyanylation a été démarrée, et prolongée 1, 2 ou 6 jours sous agitation à 37° C. La cyanylation a été suivie par mesure de DO à 412 nm pour chaque temps. Enfin, 1.5 % (v/v) de NH_4OH 2N ont été ajoutés pour le clivage jusqu'à obtenir un pH d'environ 9.3. S'en est suivie une incubation de 4 h à 37 °C. La réaction a ensuite été neutralisée avec H_2SO_4 2 N.

Suivant ce protocole, des paramètres ont été optimisés afin d'améliorer l'hydrolyse, qui sont :

- La réduction s'est effectuée avec un tampon bicarbonate d'ammonium 200 mM et TCEP 100 mM. Elle a été réalisée 10 min à 60°C, au lieu de 45 min à 56°C
- La quantité de protéines digérée était de 2 nmole au lieu de 1 nmole
- Un dessalage du NTCB par changement tampon après la réduction (tampon 1: TCEP 100 mM et ammonium bicarbonate 200 mM ; tampon 2 : ammonium bicarbonate 50 mM) aves les spins colonnes PD Spin Trap G25 fournies par Cytiva (Marlborough, USA).

III.1.E. Protocole digestion au NTCB avec clivage à la glycine dans un seul mélange réactionnel (Tang et. al)[173]

0.3 mg de protéines ont été resuspendus dans un tampon composé de Tris 20 mM GdnCl (Chlorure de Guanidinium) à 6 M, Glycine 1 M. À ce mélange ont été ajoutés 50 µL d'un tampon de réduction contenant 200 mM de TCEP, Tris 20 mM GdnCl 6M et Gly 1M. Les protéines ont été réduites sous incubation à 37 °C à 500 rpm pendant 15 min. Après réduction 15 µL de NTCB à 1.6 $mg.mL^{-1}$ ont été ajoutés au mélange réactionnel. La réaction s'est poursuivie dans la nuit, sous agitation à 37 °C. Le lendemain, le mélange a été conservé à -20 °C jusqu'à l'analyse.

III.2. Optimisation de la séparation des peptides et protéines par chromatographie SEC (exclusion stérique)

III.2.A. Séparation des protéines entières

- **Séparation de l'albumine et des protéines commerciales par SEC analysée par la méthode 1**

5 µL des protéines à 2 $mg.mL^{-1}$ ont été injectés sur une colonne SEC. Les protéines en question sont : β-Glu (almond), β-Gal (*E.coli*), β-Glu (*E.coli)*, albumine bovine, albumine humaine, et RNase bovine. Les paramètres d'analyse sont figurés partie III.D.2.

- **Séparation de l'albumine et des protéines commerciales par SEC en condition réductrice analysée par la méthode 1**

20 µL de protéine à 2 mg.mL^{-1} ont été dénaturés au SDS 4 % w/v et incubés pendant 10 min à 90 °C sous agitation. Les protéines ont été ensuite réduites au DTT 10 mM, incubées sous agitation pendant 10 min à 90°C. Les thiols des cystéines ont été alkylés avec de l'iodoacétamide 40 mM et incubés 1 h dans l'obscurité à température ambiante.

- **Gamme de détection de la BSA par SEC analysée par la méthode 2**

Une série d'échantillons d'albumine bovine de différentes concentrations été injectée sur une colonne SEC. Les paramètres d'analyse de cette méthode 2 sont présentés partie III.D.2. Les concentrations de la gamme sont 1, 2, 10, 20, 30 mg.mL^{-1}.

- **Séparation des protéines entières SEC en conditions natives analysée par la méthode 2**

500 µL de protéines à 2 mg.mL^{-1} ont été repris dans 700 µL d'eau ultrapure. 100 µL du mélange ont été injectés sur la colonne SEC et analysés selon la méthode 2.

III.2.B. Séparation de l'albumine et des protéines commerciales par SEC en conditions dénaturantes analysée par la méthode 1

Dénaturation des protéines

20 µL de protéines à 2 mg.mL^{-1} ont été mis en présence d'une solution contenant du SDS à 4 %. La suspension a été incubée 10 min sous agitation (500 rpm) à 90°C. Les protéines ont ensuite été réduites avec 10 mM de DTT et incubées 10 min à 90 °C sous agitation puis mises en suspension dans 40 mM d'IAA et une incubation à température ambiantes dans l'obscurité pendant 1 h, pour l'alkylation des thiols.

III.3. Analyse des protéines hydrolysées par NTCB

- **Séparation des digestats NTCB par SEC analysée par la méthode 2**

0.96 mg de protéines ont été digérées au NTCB selon le protocole Tang et.al [173] dans de grands volumes. Le lendemain l'intégralité du mélange réactionnel a été ajouté à 400 µL d'eau ultra pure puis injecté sur la colonne Reprosil avec un débit de 0.2 mL.min^{-1}.

- **Séparation des digestats NTCB par SEC avec les digestats analysée par la méthode 2**

Un échantillon de NTCB (1.6 mg) a été mis en milieu basique dans de la soude 5 M puis dilué 14 fois dans l'eau car la couleur était trop intense (orange foncé), risquant de saturer le détecteur UV. En parallèle les digestats NTCB ont été mélangés avec 5 % (v/v) d'acide formique. La forme acide est un léger jaune ocre.

III.4. Paramètres LC-UV

Méthode 1

L'analyse a été réalisée avec un système de chromatographie liquide type UPLC Acela 1250 pump, couplé à un détecteur PDA (Thermofischer scientific,Massachussets, USA). La séparation des peptides a été effectuée sur une colonne analytique Bio-SEC 3 (4.6x300 mm i.d, de granulométrie de 3 µm, et porosité de 300 Å). L'élution des protéines a été réalisée à température ambiante pendant 30 min, à un débit de 200 µL.min^{-1}, selon un

mode isocratique. La phase mobile était composée de 25 mM acétate d'ammonium, 5 % acétonitrile, acide acétique pH 4.5. La détection a été effectuée en UV à 220 et 280 nm.

Méthode 2

L'analyse des protéines en SEC été réalisée avec un système de chromatographie liquide type 1525 Binary Pump venant de Waters (Milford, USA) couplé à un détecteur UV-visible 2489 en provenace de Waters. La séparation des protéines été effectuée sur une colonne analytique Reprosil 200 (300 X 8 mm i.d de granulométrie de 5 µm, porosité de 200 Å) fournie par Altmann Analytic (Munich, Allemagne). L'élution des peptides a été réalisée à température ambiante pendant 30 min, à un débit de 200 $\mu L.min^{-1}$, selon un mode isocratique. La phase mobile est constituée d'ammonium acétate 10 mM. La durée de l'analyse est de 30 min. La détection UV s'est faite à 220 et 280 nm.

III.5. Paramètres UPLC-HRMS

Les paramètres d'analyse des peptides issus de la digestion trypsique et de l'hydrolyse au NTCB sur l'UPLC Q exactive en utilisant la colonne Hypersil Gold Q (100×2.1 mm i.d, de granulométrie de 1.9 µm, et porosité de 175 Å, Thermo Scientific) sont les suivants donnés tableau 15.

	Paramètres LC-MS	Méthode
PARAMÈTRES LC	Débit ($\mu L.min^{-1}$)	500
	Température colonne (°C)	40
	Éluants	A : EUP + 0.1 % AF (Acide formique) B : ACN + 0.1 % AF
	Programme LC	0 - 1 min : 95 % A ; 5 % B 13 min : 35 % A : 65 % B 14 – 16 min : 10 % A ; 90 % B 16 - 18 min : 95 % A ; 5 % B
PARAMÈTRES MS	Mode	MRM (positif)
	Énergie d'ionisation (kV)	4
	Températures d'interface, du bloc chauffant et de la ligne de désolvatation (°C)	400 – 500 – 250
	Débit du gaz de nébulisation, du gaz de chauffage et du gaz de séchage ($L.min^{-1}$)	180 – 600 – 600

Tableau 14:Paramètres UPLC HRMS de l'analyse des peptides

Partie C. Résultats et discussion

Volatolomique induite

Le cancer est aujourd'hui un problème majeur de santé publique. Malgré des avancées remarquables dans le domaine de la thérapie, les approches actuelles ne sont pas encore totalement efficaces contre de nombreux types de tumeurs solides, en particulier lorsque celles-ci sont détectées à un stade tardif de la maladie. Ainsi, l'une des voies possibles pour diminuer le taux de mortalité lié aux cancers consiste à mettre au point des stratégies de diagnostic précoces de la maladie. L'OMS a alors estimé que la mortalité par cancers serait réduite de 30% grâce à ces dépistages précoces.

Jusqu'à présent, le diagnostic d'un patient atteint d'un cancer résulte soit de l'examen clinique de l'individu lorsque celui-ci présente des symptômes bien particuliers, soit d'une participation à un programme de dépistage national. L'examen clinique peut constituer une méthode efficace pour identifier un cancer à un stade précoce. Il implique néanmoins une détection de la tumeur à la palpation et/ou la connaissance parfaite des symptômes qui lui sont associés, et n'est donc à l'heure actuelle efficace que pour quelques cancers dont l'incidence est élevée. Les programmes de dépistage issus de politiques nationales ont prouvé leur efficacité pour les cancers du sein, côlon, et col utérin, mais ils restent discutables en raison de 1) leur faible accessibilité, 2) leur spécificité, 3) leur coût élevé, et 4) leur sensibilité (les techniques d'imagerie ne peuvent détecter que des lésions cancéreuses ayant une taille de 5 à 10 mm.

L'émergence de la « volatolomique » (i.e. composante de la métabolomique qui s'intéresse à l'analyse du volatolome ; ensemble des composés organiques volatils (COVs) produits par un système biologique) offre de nouvelles perspectives en termes de diagnostic et pronostic non invasif des maladies. En effet, différentes études réalisées à grande échelle, ont révélé que certaines molécules volatiles pourraient servir de marqueurs de détection et de suivi en temps réel d'événements physiopathologiques dans le corps humain. Comparés à l'analyse de marqueurs protéiques ou acides nucléiques, l'étude des COVs endogènes est non invasive (du moins lorsqu'ils sont détectés dans l'haleine, la sueur, le sébum ou l'urine), leur prélèvement est rapide et peut être réalisé aussi souvent que nécessaire. L'ensemble de ces facteurs facilitent donc leur transfert en clinique. Dès lors, de nombreux marqueurs volatils endogènes présents dans l'haleine, la transpiration, l'urine ou les selles des patients, et spécifiques d'un type de cancer donné ont été identifiés. A l'heure actuelle, des combinaisons de plusieurs biomarqueurs volatils endogènes ont démontré un réel intérêt pour le diagnostic de certains cancers. Cependant, derrière l'apparente simplicité de cette méthodologie, aucun marqueur endogène n'est pour l'heure utilisé en clinique. En effet, l'état de santé général de l'individu, son alimentation, ou encore la prise de médicaments, l'exercice physique et le tabagisme peuvent induire des changements qualitatifs et quantitatifs de son volatolome, induisant donc un risque de faux-positif ou faux-négatif important.

Outre cette variabilité inter-individuelle, la volatolomique manque encore de protocoles normalisés pour ce qui est du prélèvement, de la préparation et de l'analyse des échantillons, ce qui peut également influencer les conclusions. De plus, compte tenues des très faibles concentrations de COVs dans les fluides et gaz corporels (~10^{-11} M dans le sang et 0,02 ppb dans l'haleine), les techniques d'échantillonnage et d'analyse doivent être rigoureusement sélectionnées afin d'obtenir une quantification fiable et sensible. Enfin, la méconnaissance des voies métaboliques de formation/dégradation de ces biomarqueurs, pourrait également être la cause des conclusions contradictoires d'une étude à l'autre.

En rupture avec cette recherche classique de COVs endogènes, nous travaillons sur un concept émergent, appelé « *induced volatolomics* » qui repose sur l'utilisation de sondes off-on non toxiques, et incluant dans leur structure un COV (Fig.37 bis). Ces sondes seraient métabolisées sélectivement dans les tissus lésés *via* l'activité d'enzymes marqueurs de la pathologie. Les COVs exogènes ainsi libérés deviendraient à leur tour des traceurs de la pathologie et seraient détectés, *in vivo*, dans l'haleine des patients, ou *ex vivo*, sur prélèvements biologiques.

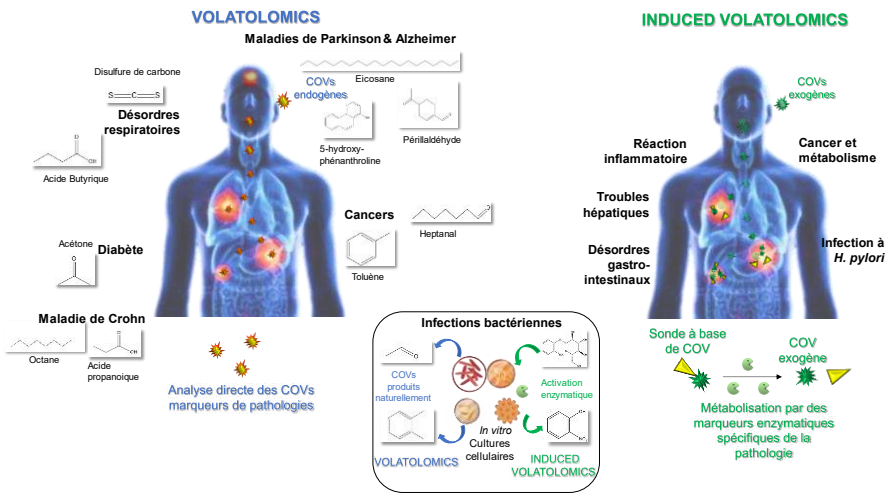


Fig. 37 bis. Volatolomique et approche émergente dérivée « Induced Volatolomics »

Les sondes à base de COVs, appartenant à la classe des sondes off-on, constitueraient donc un outil intéressant pour fournir de façon précoce et sensible, un premier aperçu d'événements oncogéniques se produisant dans le corps d'un individu. Dans ce contexte, nous avons démontré l'efficacité d'une sonde à base d'éthanol deutéré (glucuronide d'éthyle-D_5) permettant de diagnostiquer différents types de tumeurs solides et suivre leur dynamique après chimiothérapie. Cette sonde est actuellement être testée en clinique au Royaume-Uni sur des patients sains et atteints d'un cancer du poumon (Owlstone Medical Ltd, Cambridge).

En ciblant une enzyme extracellulaire, la spécificité de la sonde a été considérablement augmentée.

Cependant, une telle approche pâtit du fait que le marqueur enzymatique ici ciblé, la β-glucuronidase, peut être produit par certains microorganismes du microbiote intestinal (*E. Coli*), ou en cas de pathologies diverses (infection bactérienne ou inflammation). Cette production peut donc conduire à un certain nombre de faux-positifs et faux-négatifs.
Dans ce contexte, nous proposons ici de cibler plusieurs activités enzymatiques avec un cocktail de sondes croisées incluant dans leur structure un COV afin de (1) discriminer des individus sains d'individus atteints d'un cancer, (2) de discriminer les différents types de cancers, (2) de mettre en évidence les activités enzymatiques « up-régulées » dans un cancer donné, (3) de synthétiser un agent chimiothérapeuthique ciblant ces activités, afin de proposer aux individus une chimiothérapie personnalisée et efficace.

Dans le cadre de ce projet pilote, les glycosidases ont constitué nos cibles biologiques principales. Bien que le rôle des glycosidases dans la mise en place et l'établissement des processus cancéreux et des réponses inflammatoires subséquentes n'est pas encore parfaitement défini, le ciblage de ces enzymes est loin d'être trivial.
En effet, certaines sont sur-exprimées dans le micro-environnement (ou éventuellement à l'intérieur) des cellules cancéreuses, ou ont été détectées dans la circulation sanguine de patients présentant différents types de cancers. Ainsi, la β-glucuronidase, la β-galactosidase, la N-acétyl-β-D-glucosaminidase, et l'α-fucosidase semblent être des candidats marqueurs intéressants. Des sondes hétérosidiques substrats de chacune de ces enzymes ont donc été développées. Après hydrolyse, elles libéraient une molécule d'éthanol marqué.

Ce travail est exposé dans le papier « Induced-volatolomics for the design of tumour activated therapy » publié en Avril dernier dans la revue Chemical Science.
Mon implication a consisté à développer la méthode d'analyse ciblée par GC-MS/MS des différents isotopes de l'éthanol, de vérifier la spécificité des enzymes vis à vis de leur sonde substrat. J'ai également établi les conditions d'injection in vivo des sondes, ainsi que les conditions de prélèvement des COVs expirés par les animaux. Enfin, j'ai réalisé l'étude in vivo détaillée dans cet article.

Bien que les sondes à base de COVs permettent de discriminer des individus sains d'individus atteints d'un cancer, leur intérêt médical n'a pour l'heure été démontré qu'in vivo. Or, dans un contexte de transfert clinique, ce protocole, impliquant l'injection d'une sonde, ne pourra être mis en place sans l'obtention d'autorisations de la part l'*Agence* nationale de sécurité du *médicament* et des produits de santé.

Ainsi, nous avons travaillé sur un protocole *in vitro*, à partir de prélèvements sanguins de patients. Les conditions d'optimisation seront ici détaillées. La stabilité des sondes a été évaluée. Plusieurs paramètres (présence de protéases plasmatiques, temps et cycles de congélation des échantillons) qui pourraient avoir un impact soit sur l'activité des enzymes, soit sur la détection des COVs, ont été optimisés afin d'obtenir des réponses enzymatiques maximales.

Enfin, un protocole d'extraction des glycosidases présentes dans la matrice extracellulaire des tumeurs a été développé. Cette étude avait pour objectif de déterminer l'origine cancéreuse ou immunitaire des enzymes suivant leurs séquences primaires et modifications post-traductionnelles.

Dans le cadre de ce travail, une approche protéomique de type « middle down » a été optimisée. Plusieurs stratégies d'hydrolyse et de purification reposant sur le fractionnement des protéines suivant leur masse moléculaire ont été étudiées afin d'éliminer les protéines les plus abondantes (ex: albumine, chaînes de collagène, fibronectine…) interférant avec la détection des glycosidases présentes en faible concentration.

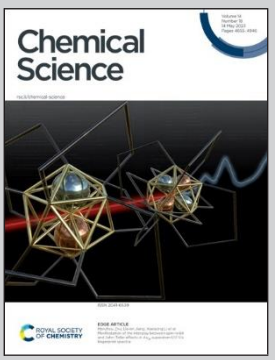


Showcasing research from Professor Poinot's laboratory, Institute of Chemistry, University of Poitiers, France.

Induced-volatolomics for the design of tumour activated therapy

We developed a cocktail of enzyme-responsive probes releasing a volatile tracer upon activation by glycosidases. This cocktail allowed the simultaneous screening of several tumour-associated enzyme activities in living mice, through the detection of volatile tracers in their breath. This induced-volatolomics modality led to the identification of an enzyme target that was exploited in the course of a Tumour Activated Therapy for the treatment of triple-negative breast cancer *in vivo*. This study shows the potential of induced-volatolomics for the exploration of biological processes which could further trigger biomedical and pharmaceutical innovations.

As featured in:



See Sébastien Papot, Pauline Poinot *et al.*, *Chem. Sci.*, 2023, 14, 4697.

Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE



Cite this: Chem. Sci., 2023, 14, 4697

All publication charges for this article have been paid for by the Royal Society of Chemistry

Induced-volatolomics for the design of tumour activated therapy†

Remi Chatre,‡[a] Estelle Blochouse,‡[a] Rony Eid,[a] Fabiola Djago,[a] Justin Lange,[a] Mehrad Tarighi,[a] Brigitte Renoux,[a] Julien Sobilo,[b] Alain Le Pape,[b] Jonathan Clarhaut,[ac] Claude Geffroy,[a] Isabelle Opalinski,[a] Wei Tuo,[a] Sebastien Papot *[ad] and Pauline Poinot *[a]

The discovery of tumour-associated markers is of major interest for the development of selective cancer chemotherapy. Within this framework, we introduced the concept of induced-volatolomics enabling to monitor simultaneously the dysregulation of several tumour-associated enzymes in living mice or biopsies. This approach relies on the use of a cocktail of volatile organic compound (VOC)-based probes that are activated enzymatically for releasing the corresponding VOCs. Exogenous VOCs can then be detected in the breath of mice or in the headspace above solid biopsies as specific tracers of enzyme activities. Our induced-volatolomics modality highlighted that the up-regulation of N-acetylglucosaminidase was a hallmark of several solid tumours. Having identified this glycosidase as a potential target for cancer therapy, we designed an enzyme-responsive albumin-binding prodrug of the potent monomethyl auristatin E programmed for the selective release of the drug in the tumour microenvironment. This tumour activated therapy produced a remarkable therapeutic efficacy on orthotopic triple-negative mammary xenografts in mice, leading to the disappearance of tumours in 66% of treated animals. Thus, this study shows the potential of induced-volatolomics for the exploration of biological processes as well as the discovery of novel therapeutic strategies.

Received 9th December 2022
Accepted 7th April 2023

DOI: 10.1039/d2sc06797h

rsc.li/chemical-science

Enzymes are major targets for drug discovery in oncology.[1,2] The current therapeutic arsenal includes numerous enzyme inhibitors[3] and enzyme-responsive antibody–drug conjugates[4] for the treatment of a wide range of malignancies. Thus, the design of novel strategies dedicated to the exploration of tumour-associated enzymes is of prime importance for cancer therapy and diagnosis.[5,6] Within this framework, activable imaging probes have been developed to detect the dysregulation of some enzyme activities in tumours.[7–17] However, most of these turn-on probes suffer from low signal penetration depth and background noise.[18,19] Recently, we introduced the concept of induced-volatolomics for cancer diagnosis.[20,21] We demonstrated that a volatile organic compound (VOC)-based probe can be activated in vivo by an enzyme overexpressed in solid tumours to trigger the release of an exogenous volatile tracer in the breath of individuals, thereby enabling cancer detection. Herein, we present an original approach allowing real time and simultaneous monitoring of the activity of several tumour-associated enzymes by means of a VOC-based probe cocktail (Fig. 1). This induced-volatolomics modality allowed the identification of upregulated enzymes within the tumour, leading to the development of a new enzyme-targeted drug candidate.

In this proof-of-concept, we chose to target glycolytic activity of tumours since glycosidase dysregulation constitutes a hallmark of many cancers.[22–26] b-Galactosidase (b-Gal), a-L-fucosidase (a-Fuc), b-glucuronidase (b-Glu) and N-acetyl-b-D-glucosaminidase (b-GlcNAc) were selected as the target enzymes since their overexpression has been reported in several solid tumours.[27–34] Thus, the probe cocktail was composed of four ethyl glycosides i.e. D_2-ethyl-b-D-galactopyranoside (D_2-EtGal) D_4-ethyl-a-L-fucopyranoside, (D_4-EtFuc), D_5-ethyl-b-D-glucuronide (D_5-EtGlu) and $^{13}CD_5$-ethyl-N-acetyl-b-D-glucosamine ($^{13}CD_5$-EtGlcNAc), designed to release D_2-, D_4-, D_5- and $^{13}CD_5$-ethanol isotopes, respectively, in the presence of the corresponding glycosidase.

After cocktail administration (Fig. 1, step 1), VOC-based probes can be converted into labelled ethanol exclusively if the corresponding glycosidase is overexpressed in the tumour mass (step 2). The volatile tracers can diffuse in the blood stream and

[a]University of Poitiers, UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Equipe Labellisée Ligue Contre le Cancer, 4 Rue Michel-Brunet TSA 51106, 86073 Poitiers Cedex 9, France. E-mail: pauline.poinot@univ-poitiers.fr; sebastien.papot@univ-poitiers.fr

[b]UAR No. 44 PHENOMIN TAAM-Imagerie In Vivo, CNRS, 3B Rue de la Férollerie, F-45071 Orléans, France

[c]CHU de Poitiers, 2 Rue de la Miléterie, CS 90577, F-86021 Poitiers, France

[d]Seekyo SA, 2 Avenue Galilée, BP 30153, 86961 Futuroscope, France

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2sc06797h

‡ These authors contributed equally.

© 2023 The Author(s). Published by the Royal Society of Chemistry

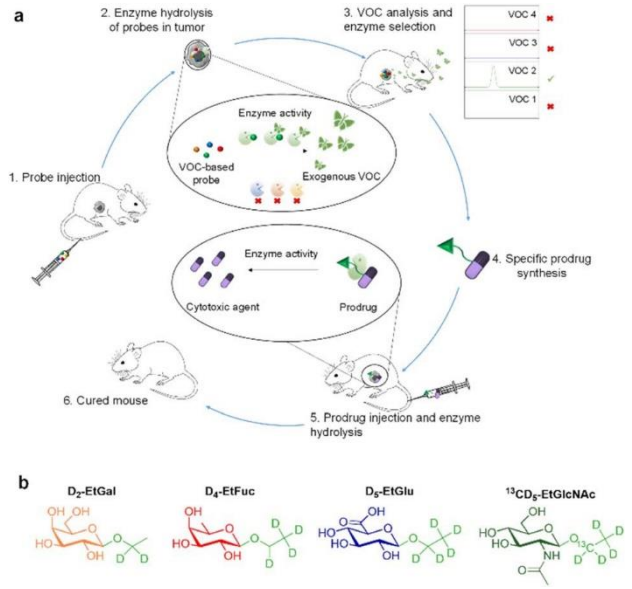


Fig. 1 Volatolomics-based modalities for drug discovery. (a) To highlight the dysregulation of glycosidase activity in tumours, a cocktail of VOC-based probes (D_2-EtGal, D_4-EtFuc, D_5-EtGlu and $^{13}CD_5$-EtGlcNAc) is administered to tumour-bearing mice (1). The probes can be hydrolysed selectively in the tumour microenvironment by the corresponding glycosidases to release D_2, D_4, D_5 or $^{13}CD_5$-ethanol isotopes (2). When released in tumours, labelled VOCs pass in the blood stream and are exhaled in the breath (3). Analysis of VOC signals permits the identification of the most active enzyme in the tumour microenvironment. The corresponding enzyme-responsive drug delivery system is synthesized (4), injected in mice (5) and evaluated for its anticancer efficacy in mice (6). (b) Chemical structure of D_2-EtGal, D_4-EtFuc, D_5-EtGlu and $^{13}CD_5$-EtGlcNAc probes.

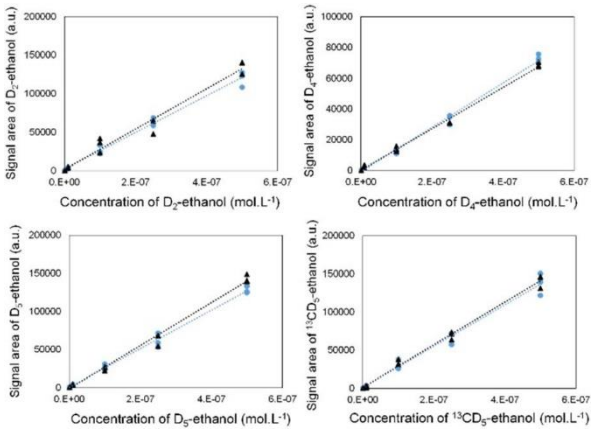


Fig. 2 Calibration curves for solutions of D_2-ethanol, D_4-ethanol, D_5-ethanol and $^{13}CD_5$-ethanol in the absence (black regression lines) or in the presence of the three other isotopes (blue regression lines).

be monitored in an animal's breath as reporters of the glycolytic activity prevailing in the tumour (step 3). This screening phase permitted to identify b-GlcNAc as a target for the development of a tumour activated therapy. On this basis, we designed and evaluated (steps 4 and 5) the first b-GlcNAc prodrug of the potent monomethyl auristatin E (MMAE) that demonstrated a remarkable therapeutic efficacy on orthotopic triple-negative mammary tumour in mice (step 6).

To ensure the selective detection of D_2-, D_4-, D_5- and $^{13}CD_5$-ethanol molecules, we first developed an analytical strategy involving the GC-MRM-MS methodology (see the ESI† for more details).

We validated its performance by drawing calibration curves for each labelled ethanol, in the absence and presence of the three other isotopes (Fig. 2). The superposition of both curves indicates that the isotope signals did not interfere with each other.

Since b-Gal, a-Fuc, b-Glu and b-GlcNAc are lysosomal enzymes widely distributed in healthy tissues, we next evaluated the dose of probes that has to be injected in mice for avoiding unselective activation. Indeed, at too high doses, probes can penetrate across cell membranes and then be hydrolysed by intracellular glycosidases in non-target tissues. Thus, we conducted preliminary experiments in healthy mice to determine the lowest detectable signal (LDS) for each VOC-based probe. For this purpose, we administered increasing probe concentrations to the animals that were then placed for over 1 h 30 in airtight cages equipped with an SPME fibre (Fig. 3a).

During the last 30 minutes, exhaled D_2-, D_4-, D_5- and $^{13}CD_5$-ethanol isotopes were trapped and subsequently analysed by GC-MRM-MS. Under these conditions, injection of 1, 100, 10 and 10 mg kg^{-1} doses of D_2-EtGal, D_4-EtFuc, D_5-EtGlu and $^{13}CD_5$-EtGlcNAc respectively, produced the LDS for the corresponding volatile tracers (see the ESI†). Thus, these probe concentrations were selected for the selective detection of glycosidases overexpressed in the tumour extracellular matrix.

Therea∨er, our experimental protocol was applied to highlight the potential upregulation of glycosidases in malignant tissues. The cocktail of probes was administered to athymic nude mice bearing subcutaneous human KB xenogra∨s (cervical cancer cell line) on days 12 and 15 post-tumour implantation. A control experiment was also carried out with healthy animals. On days 12 and 15, the amounts of D_5- and $^{13}CD_5$-ethanol were signi∥cantly higher in the breath of cancer animals compared to those emitted by their healthy counterparts (Fig. 3b). In contrast, the concentrations of D_2- and D_4-ethanol exhaled by tumour-bearing mice remained in the background noise. These results indicate that both b-Glu and b-GlcNAc exhibit increased catalytic activities in tumours while b-Gal and a-Fuc are not upregulated. Although the overexpression of b-Glu in solid tumours has been widely reported in the literature,[35,36] the presence of elevated concentrations of b-GlcNAc remains weakly documented so far.[37,38] These outcomes emphasize the potential of induced-volatolomics modalities to identify novel enzyme targets for drug discovery. Interestingly, between days 12 and 15, enzymatic activities of b-Glu and b-GlcNAc increased concomitantly with tumour growth (1.9- and 2.5-fold, respectively). Under such circumstances, the VOC-based probe cocktail seems to be a useful tool for monitoring disease progression in real time.[21]

We next investigated our probe cocktail ex vivo on tumour biopsies. The aim of this trial was to compare simultaneously the catalytic activity of the four glycosidases in tumours and

© 2023 The Author(s). Published by the Royal Society of Chemistry

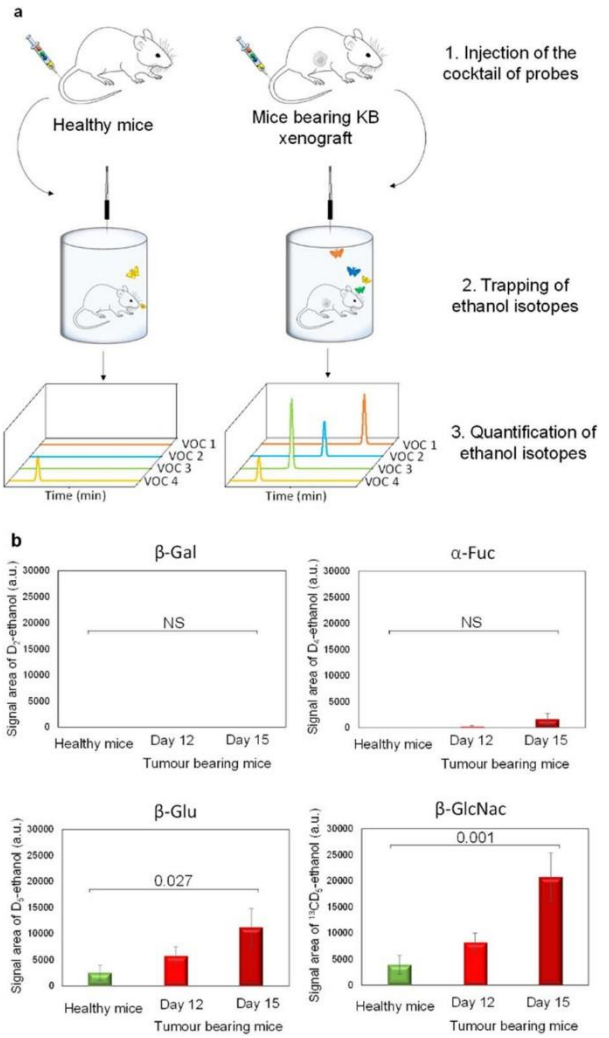


Fig. 3 (a) In vivo protocol to detect the presence of exoglycosidases in the tumour microenvironment. (b) Amount of D_2-ethanol, D_4-ethanol D_5-ethanol and $^{13}CD_5$-ethanol exhaled by athymic nude healthy mice (green bars; n = 6); amount of D_2-ethanol, D_4-ethanol, D_5-ethanol and $^{13}CD_5$-ethanol exhaled by mice with a KB subcutaneous xenograft 12 days post tumour implantation (red bars, n = 6); amount of D_2-ethanol, D_4-ethanol, D_5-ethanol and $^{13}CD_5$-ethanol exhaled by mice with a KB subcutaneous xenograft 15 days post tumour implantation (dark red bars, n = 6). For each isotope, significant differences in ethanol isotope amounts were investigated by ANOVA (confidence interval of 95%). Numbers above square brackets correspond to p-values.

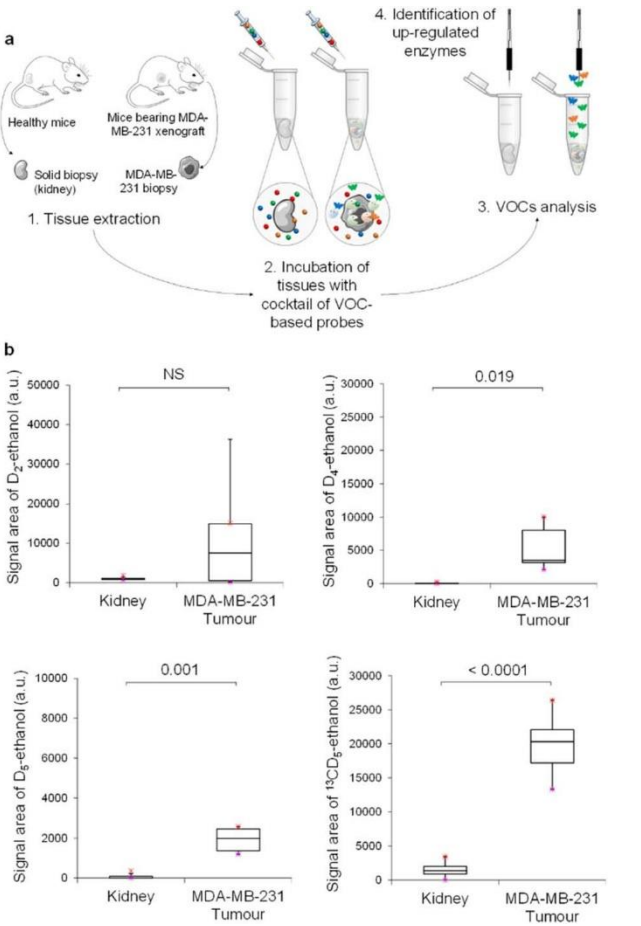


Fig. 4 (a) Ex vivo protocol for the detection of exoglycosidases on solid biopsies. Healthy kidneys and MDA-MB-231 grafts were incubated in the presence of a cocktail of VOC-based probes (D_2-EtGal D_4-EtFuc, D_5-EtGlu and $^{13}CD_5$-EtGlc). The four ethanol isotopes were trapped during 30 min in the sample headspace and analyzed by GC-MRM-MS. (b) Chemical evidence of b-galactosidase, a-L-fucosidase, b-glucuronidase and N-acetyl-b-D-glucosaminidase activities in tissue samples. The results are presented in the form of boxplot representations. Purple and red crosses correspond to the minimum and maximum isotope signals, respectively. Eight kidneys and six MDA-MB-231 xenografts were used in this experiment. Numbers above square brackets correspond to p-values.

healthy tissues such as kidneys. In these experiments, we used D_2-EtGal, D_4-EtFuc, D_5-EtGlu and $^{13}CD_5$-EtGlcNAc probes at 10^{-6}, at 10^{-4}, 10^{-6}, and 10^{-7} mol L^{-1}, respectively, since these concentrations produced the LDS for ethanol isotopes when used on healthy tissues (see the ESI†).

Thus, MDA-MB-231 human mammary tumours harvested from nude mice with orthotopic xenografts were incubated in acetate buffer (pH 5) in the presence of the cocktail of VOC-based probes (Fig. 4a). Volatile tracers released in the sample headspace were monitored by GC-MRM-MS (Fig. 4b). The same experiment was also conducted with mice kidneys. Excepted for D_2-ethanol, concentrations of the volatile tracers released in the headspace were higher for tumour than for kidney samples. These results indicated the presence of an elevated level of a-Fuc, b-Glu and b-GlcNAc in MDA-MB-231 xenografts compared to in healthy tissues. In contrast, b-Gal activity in tumours was highly variable and not significantly different than that of the control group. As was the case for KB tumours, b-GlcNAc was identified once again as the most active glycosidase. Indeed, $^{13}CD_5$-ethanol was the volatile tracer exhibiting the highest GC-RMM-MS signal, although $^{13}CD_5$-EtGlcNAc was the probe incubated at the lowest concentration (10^{-7} mol L^{-1}).

Overall, these results suggest that the overexpression of b-GlcNAc could be a hallmark of solid tumours. Therefore, this glycosidase appeared to be an attractive target for the development of a novel enzyme-activated cancer therapy. With the aim

© 2023 The Author(s). Published by the Royal Society of Chemistry

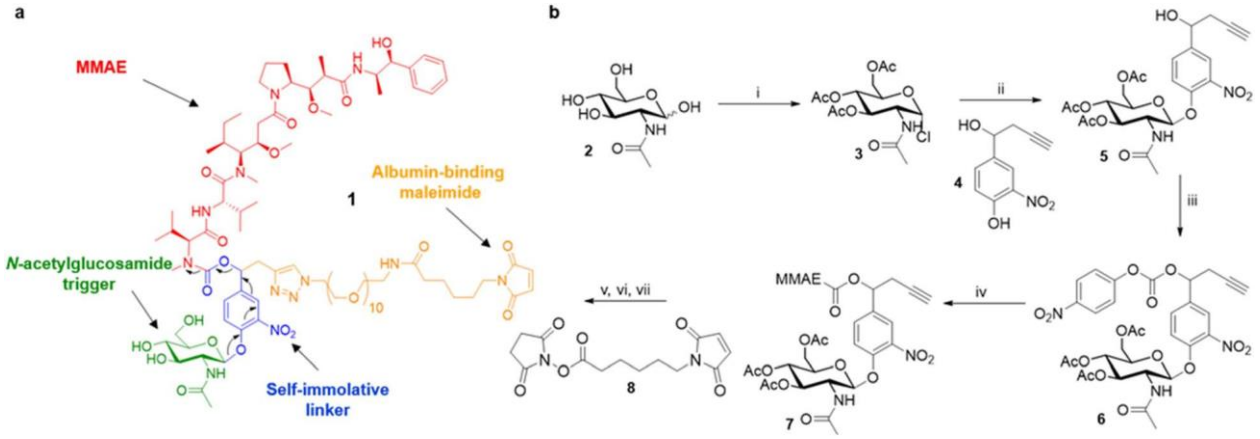


Fig. 5 (a) Structure of the b-GlcNAc-responsive albumin-binding prodrug 1 (b) synthesis of the N-acetylglucosamide prodrug 1 (i) acetyl chloride (14 eq.), RT, 72 h, 61%; (ii) 4 (1.5 eq.), tetrabutylammonium bromide (1 eq.), $CH_2Cl_2/NaHCO_3$ 1 M, RT, 5.5 h, 68%; (iii) 4-nitrophenyl chloroformate (2 eq.), pyridine (2.5 eq.), CH_2Cl_2, 0 °C to RT, 3 h, 72%; (iv) MMAE (1 eq.), HOBt (1 eq.), pyridine/DMF, RT, 72 h, 93%; (v) O-(2-Aminoethyl)-O′-(2-azidoethyl)nonaethylene glycol (1.1 eq.), $Cu(MeCN)_4PF_6$ (1 eq.), CH_2Cl_2, RT, 3 h; (vi) MeONa (0.15 eq.), MeOH, RT, 5 h; (vii) 8 (1.1 eq.), Et_3N (3 eq.) DMSO, RT, 1 h, 24% over 3 steps.

to explore this therapeutic strategy, we designed the b-GlcNAc-responsive albumin-binding prodrug 1 programmed for the selective delivery of MMAE in the tumour microenvironment[39–41] (Fig. 5a). Once in the blood stream, 1 is expected to bind selectively to the free thiol of circulating albumin through Michael addition. The resulting bioconjugate will then accumulate in the tumour microenvironment where the glycosidic bond will be cleaved by extracellular b-GlcNAc, hence triggering the release of MMAE via the self-immolative mechanism[42–44] described in Fig. 5a. It is worth mentioning here that glycosidase-responsive albumin-binding prodrugs are more efficient than their non-binding counterparts since they lead to increased drug deposition in the tumor microenvironment combined with a prolonged plasma half-life.[39,45]

The tumour-activated prodrug 1 was accessible from the commercially available b-N-acetylglucosamide 2 (Fig. 5b). First, this latter one was converted into the glycoside donor 3 (61%) that was coupled subsequently with the nitrophenol 4 (ref. 46 and 47) to afford compound 5 in 68% yield. Treatment of the benzyl alcohol 5 with 4-nitrophenyl chloroformate led to the activated carbonate 6 (72%). MMAE was then introduced via nucleophilic substitution in the presence of hydroxybenzo-triazole to give the carbamate 7 (93%). The three last steps of the synthesis were undertaken without intermediate purification. Thus, alkyne 7 reacted with O-(2-aminoethyl)-O-(2-azidoethyl)-nonaethylene glycol and $Cu(CH_3CN)_4PF_6$ to form the corresponding triazole through copper(I)-catalysed azide–alkyne 1,3-cycloaddition. Deprotection of the glycoside followed by the introduction of the maleimide moiety using the N-hydroxysuccinimide ester 8 finally led to the drug delivery system 1 (24% over 3 steps).

As enzymatic hydrolysis of the b-N-acetylglucosamide trigger is the key step in the process of drug release, we first verified that the carbohydrate moiety was still accessible to b-GlcNAc once bound to albumin. For this purpose, prodrug 1 was incubated with human serum albumin (HSA) at 37 °C in order to form the corresponding bioconjugate through Michael addition. Under these conditions, nearly 90% of 1 was converted in two hours as a result of its rapid binding with the protein (Fig. 6a). The formation of the albumin conjugate was confirmed by trypsin digestion followed by HPLC/HRMS analysis (see the ESI†). When b-GlcNAc was incubated in the medium, the full release of MMAE was achieved in three hours indicating that the b-N-acetylglucosamide was a substrate of the enzyme for the activating enzyme even with the proximity of the bulky albumin (Fig. 6b). In contrast, the release of MMAE was not observed after 22 hours of incubation in the absence of b-GlcNAc (see the ESI†). Thus, these results highlighted that the prodrug 1 possessed the capacity both to bind covalently to HSA and release MMAE following enzymatic activation.

We then examined the antiproliferative activity of N-acetyl-glucosamide 1 against human KB and MDA-MB-231 tumour cell lines. As shown in Fig. 7a, the prodrug 1 exhibited a similar cytotoxicity to free MMAE, with IC_{50} values in the low nanomolar range. This result indicated that 1 was activated by b-GlcNAc produced by tumour cells,[38] leading to the release of the drug. Since the co-incubation of the prodrug with b-GlcNAc did not induce a supplementary cytotoxicity, it appeared that the level of the glycosidase generated by tumour cells in the culture

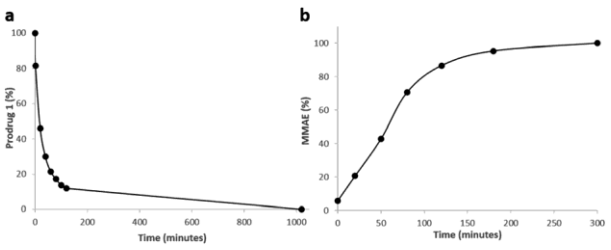


Fig. 6 (a) Disappearance of 1 over time when placed in the presence of HSA at 37 °C. (b) Kinetics of MMAE release from 1 linked to HSA in the presence of b-GlcNAc (0.227 U mL^{-1}).

111

© 2023 The Author(s). Published by the Royal Society of Chemistry

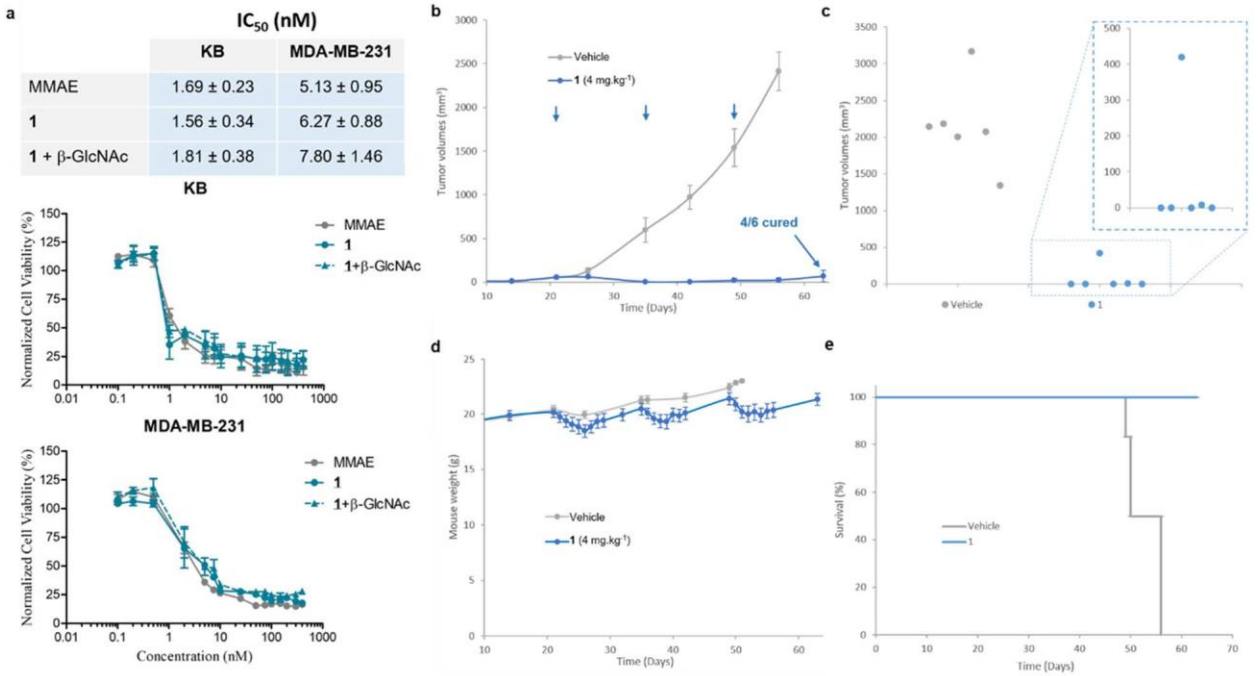


Fig. 7 (a) Antiproliferative activities and IC_{50} of MMAE and 1 with or without b-GlcNAc after 3 days of treatment. Each point shows mean ± s.e.m. from 3 independent experiments in triplicate. (b) MDA-MB-231 tumour growth inhibition under therapy with a vehicle and 1. Each point shows mean ± s.e.m. from 6 tumour volumes. (c) Post-euthanasia tumour volumes of mice bearing MDA-MB-231 xenografts treated with a vehicle and 1. (d) Mean body weights of each group of mice bearing MDA-MB-231 xenografts. Each point shows mean ± s.e.m. from 6 mice. (e) Survival study comparing mice with MDA-MB-231 tumours treated with 1 or untreated.

medium was sufficiently elevated for the full activation of 1. In order to confirm that the cytotoxicity of the b-N-acetylglucosamide 1 was due to its enzymatic activation in the culture medium, we investigated the antiproliferative activities of the galactoside[48] and glucuronide[39] analogues against MDA-MB-231 tumor cells (see the ESI†). In contrast to 1, when these prodrugs were incubated in the absence of the corresponding enzyme (b-Gal or b-Glu), they appeared far less toxic than MMAE. On the other hand, addition of the activating enzyme in the culture medium triggered a strong antiproliferative activity due to the enzyme-catalysed release of the drug. As b-galactoside, b-glucuronide and b-N-acetylglucosamide prodrugs differ exclusively by the nature of their glycoside triggers, these results showed that the cytotoxicity observed on MDA-MB-231 tumor cells is due to the b-GlcNAc-mediated activation of 1.

We next pursued our study by assessing the antitumour activity of the b-GlcNAc-sensitive prodrug 1 in Balb/c athymic mice bearing orthotopic MDA-MB-231 triple-negative mammary tumours (Fig. 7). The animals received i.v. injections of 1 at 4 mg kg^{-1} once every two weeks. Following this protocol, a remarkable reduction of tumour volume was observed in all treated animals (Fig. 7b). At the end of the experiment (day 63), 66% of the mice treated with 1 exhibited complete remission (Fig. 7c). In contrast, a previous study showed that the free MMAE produced only a poor anticancer activity, even when administered at its maximal tolerated dose.[39] The successive i.v. administrations of 1 (1.54 mg per kg per injection MMAE equivalents) were well tolerated without major body weight loss (Fig. 6d) or detectable side effects. These outcomes demonstrated that the derivatization of MMAE in the form of the b-N-acetylglucosamide 1 significantly reduced its toxicity in vivo allowing the administration of at least 3-fold the maximal tolerate dose (MTD) of the drug.[39] Furthermore, the therapeutic efficacy obtained with 1 confirmed the selective and efficient enzyme-mediated release of MMAE in the tumour mass. Thus, this study showed that the targeting of tumour-associated b-GlcNAc by means of the enzyme-responsive albumin-binding prodrug is a promising therapeutic strategy for the treatment of triple-negative breast cancer (TNBC), a clinically aggressive disease for which current drugs are ineffective.

Conclusions

In this study, we demonstrated that a cocktail of VOC-based probes allowed the screening of several tumour-associated enzyme activities in both living mice and tumour biopsies. This induced-volatolomics modality led to the identification of an enzyme target, b-GlcNAc, which was exploited for the selective delivery of an anticancer drug that produced an impressive anticancer activity in mice bearing TNBC xenografts. Hence, our work highlights the potential of induced-volatolomics for the discovery of tumour-associated enzyme targets and the development of novel therapeutic strategies. Numerous VOC-based probe cocktails can be designed for the multimodal exploration of enzymatic dysregulation related to various diseases. Thus, induced-volatolomics modalities could become powerful tools to accelerate the process of drug discovery in the near future.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Ethical statement

Experimental procedures involving animals were conducted according to protocol no. 2017053011069010 approved by the national ethical committee and carried out in accordance with the guidelines of the French Agriculture and Forestry Ministry (decree 2013-118) and the European Communities Council Directive (2010/63/UE).

Data availability

All relevant data supporting this article have been included in the main text and the ESI.†

Author contributions

P. P. and S. P. supervised the whole study. P. P., C. G. and M. T supervised the induced-volatolomics approach and the development of analytical methods. P. P. and J. C. designed in vivo studies to test the efficiency of the cocktail of probes for highlighting highly active enzymes in tumours. J. L., J. C., F. D. and E. B. performed the in vitro and in vivo experiments using VOC-based probes. R. E. completed the VOC-based probe sensitivity study. J. C. and E. B. performed cell viability assays. S. P., I. O., B. R., R. C. and W. T. designed and synthesized the anticancer agent and VOC-based probes. R. C. and R. E. performed biochemical assays to certify the formation of the albumin-prodrug conjugate. J. S. and A. L. P. designed and supervised the experiments in tumour bearing mice treated with the enzyme-responsive prodrug. J. L., E. B. and R. C. collected all the data and reconstructed the ‖gures. J. L., E. B. and R. C. wrote the methods section. P. P. and S. P. wrote the paper with the contribution of all authors.

Conflicts of interest

There are no con/icts to declare.

Acknowledgements

The authors thank La Ligue contre le Cancer (Ligue Nationale, Comités Vienne and Deux-Sèvres), and the “Région Nouvelle Aquitaine” for ‖nancial support. This work has also been funded by the ANR through the Young Researcher grant “Volatolomix” (ANR-21-CE18-0009). We also thank the CNRS which supplied the project through the “AAP Emergence”.

Notes and references

1 G. A. Holdgate, T. D. Meek and R. L. Grimley, Nat. Rev. Drug Discovery, 2018, 17, 115–132.
2 A. C. Rufer, Drug Discov. Today, 2021, 26, 875–886.
3 E. Rajakumara, S. Abhishek, K. Nitin, D. Saniya, P. Bajaj, U. Schwaneberg and M. D. Davari, ACS Chem. Biol., 2022, 17, 266–280.
4 A. Beck, L. Goetsch, C. Dumontet and N. Corvaïa, Nat. Rev. Drug Discovery, 2017, 16, 315–337.
5 N. Khiar-Fernández, J. Macicior, B. Marcos-Ramiro and S. Ortega-Gutiérrez, Eur. J. Org. Chem., 2021, 2021, 1307–1320.
6 M.-J. Blanco and K. M. Gardinier, ACS Med. Chem. Lett., 2020, 11, 228–231.
7 W. Chyan and R. T. Raines, ACS Chem. Biol., 2018, 13, 1810–1823.
8 L. Wu, Z. Armstrong, S. P. Schröder, C. de Boer, M. Artola, J. M. Aerts, H. S. Overkleeν and G. J. Davies, Curr. Opin Chem. Biol., 2019, 53, 25–36.
9 H. Fang, B. Peng, S. Y. Ong, Q. Wu, L. Li and S. Q. Yao, Chem. Sci., 2021, 12, 8288–8310.
10 Y. Kuriki, T. Yoshioka, M. Kamiya, T. Komatsu, H. Takamaru, K. Fujita, H. Iwaki, A. Nanjo, Y. Akagi, K. Takeshita, H. Hino, R. Hino, R. Kojima, T. Ueno, K. Hanaoka, S. Abe, Y. Saito, J. Nakajima and Y. Urano, Chem. Sci., 2022, 13, 4474–4481.
11 H. Singh, K. Tiwari, R. Tiwari, S. K. Pramanik and A. Das, Chem. Rev., 2019, 119, 11718–11760.
12 Y. Wang, J. Weng, X. Wen, Y. Hu and D. Ye, Biomater. Sci., 2021, 9, 406–421.
13 S. Shingo, K. Toru, W. Rikiya, Z. Yi, I. Taiki, K. Mitsuyasu, N. Hidehiko, U. Takaaki, O. Takuji, H. Kazufumi, N. Hiroyuki and U. Yasuteru, Sci. Adv., 2020, 6, eaay0888.
14 R. K. Muir, M. Guerra and M. M. Bogyo, ACS Chem. Biol., 2022, 17, 281–291.
15 S. M. McKenna, E. M. Fay and J. F. McGouran, ACS Chem. Biol., 2021, 16, 2719–2730.
16 Y. Zhang, G. Zhang, Z. Zeng and K. Pu, Chem. Soc. Rev., 2022, 51, 566–593.
17 J. Huang, J. Li, Y. Lyu, Q. Miao and K. Pu, Nat. Mater., 2019, 18, 1133–1143.
18 E. A. Lemke and C. Schultz, Nat. Chem. Biol., 2011, 7, 480.
19 K. M. Dean and A. E. Palmer, Nat. Chem. Biol., 2014, 10, 512–523.
20 F. Djago, J. Lange and P. Poinot, Nat. Rev. Chem., 2021, 5, 183–196.
21 J. Lange, B. Eddhif, M. Tarighi, T. Garandeau, E. Péraudeau, J. Clarhaut, B. Renoux, S. Papot and P. Poinot, Angew. Chem., Int. Ed., 2019, 58, 17563–17566.
22 H. M. Burke, T. Gunnlaugsson and E. M. Scanlan, Chem. Commun., 2015, 51, 10576–10588.
23 A. F. Costa, D. Campos, C. A. Reis and C. Gomes, Trends Cancer, 2020, 6, 757–766.
24 B. N. Vajaria and P. S. Patel, Glycoconj. J., 2017, 34, 147–156.
25 G. A. Kwong, S. Ghosh, L. Gamboa, C. Patriotis, S. Srivastava and S. N. Bhatia, Nat. Rev. Cancer, 2021, 21, 655–668.
26 S. S. Pinho and C. A. Reis, Nat. Rev. Cancer, 2015, 15, 540–555.
27 D. Asanuma, M. Sakabe, M. Kamiya, K. Yamamoto, J. Hiratake, M. Ogawa, N. Kosaka, P. L. Choyke, T. Nagano, H. Kobayashi and Y. Urano, Nat. Commun., 2015, 6, 6463.
28 K. Bosslet, J. Czech and D. Hoffmann, Tumor Target., 1995, 1, 45–50.
29 K. Wang, W. Guo, N. Li, J. Shi, C. Zhang, W. Y. Lau, M. Wu and S. Cheng, Br. J. Cancer, 2014, 110, 1811–1819.

 © 2023 The Author(s). Published by the Royal Society of Chemistry

30 T.-C. Cheng, S.-H. Tu, L.-C. Chen, M.-Y. Chen, W.-Y. Chen, Y.-K. Lin, C.-T. Ho, S.-Y. Lin, C.-H. Wu and Y.-S. Ho, Oncotarget, 2015, 6(25), 21283–21300.
31 C. Slawson and G. W. Hart, Nat. Rev. Cancer, 2011, 11, 678–684.
32 E. F. Petricoin III, A. M. Ardekani, B. A. Hitt, P. J. Levine, V. A. Fusaro, S. M. Steinberg, G. B. Mills, C. Simone, D. A. Fishman, E. C. Kohn and L. A. Liotta, Lancet, 2002, 359, 572–577.
33 H. Kubo, Y. Murayama, S. Ogawa, T. Matsumoto, M. Yubakami, T. Ohashi, T. Kubota, K. Okamoto, M. Kamiya, Y. Urano and E. Otsuji, Sci. Rep., 2021, 11, 10664.
34 R. B. Parker and J. J. Kohler, ACS Chem. Biol., 2010, 5, 35–46.
35 I. Tranoy-Opalinski, T. Legigan, R. Barat, J. Clarhaut, M. Thomas, B. Renoux and S. Papot, Eur. J. Med. Chem., 2014, 74, 302–313.
36 P.-A. Burnouf, S. R. Roffler, C.-C. Wu and Y.-C. Su, J. Controlled Release, 2022, 349, 765–782.
37 B. Woynarowska, H. Wikiel and R. Bernacki, Cancer Res., 1989, 49, 5598–5604.
38 K. T. Ramessur, P. Greenwell, R. Nash and M. V. Dwek, Br. J. Biomed. Sci., 2010, 67, 189–196.
39 B. Renoux, F. Raes, T. Legigan, E. Péraudeau, B. Eddhif, P. Poinot, I. Tranoy-Opalinski, J. Alsarraf, O. Koniev, S. Kolodych, S. Lerondel, A. Le Pape, J. Clarhaut and S. Papot, Chem. Sci., 2017, 8, 3427–3433.
40 B. Renoux, L. Fangous, C. Hötten, E. Péraudeau, B. Eddhif, P. Poinot, J. Clarhaut and S. Papot, MedChemComm, 2018, 9, 2068–2071.
41 R. Châtre, J. Lange, E. Péraudeau, P. Poinot, S. Lerondel, A. Le Pape, J. Clarhaut, B. Renoux and S. Papot, J Controlled Release, 2020, 327, 19–25.
42 I. Tranoy-Opalinski, A. Fernandes, M. Thomas, J.-P. Gesson and S. Papot, Anti-Cancer Agents Med. Chem., 2008, 8, 618–637.
43 A. Alouane, R. Labruère, T. Le Saux, F. Schmidt and L. Jullien, Angew. Chem., Int. Ed., 2015, 54, 7492–7509.
44 A. G. Gavriel, M. R. Sambrook, A. T. Russell and W. Hayes, Polym. Chem., 2022, 13, 3188–3269.
45 T. Legigan, J. Clarhaut, B. Renoux, I. Tranoy-Opalinski, A. Monvoisin, J.-M. Berjeaud, F. Guilhot and S. Papot, J Med. Chem., 2012, 55, 4516–4520.
46 M. Thomas, J. Clarhaut, P.-O. Strale, I. Tranoy-Opalinski, J. Roche and S. Papot, ChemMedChem, 2011, 6, 1006–1010.
47 T. Legigan, J. Clarhaut, I. Tranoy-Opalinski, A. Monvoisin, B. Renoux, M. Thomas, A. Le Pape, S. Lerondel and S. Papot, Angew. Chem., Int. Ed., 2012, 51, 11606–11610.
48 S. Kolodych, C. Michel, S. Delacroix, O. Koniev, A. Ehkirch, J. Eberova, S. Cianférani, B. Renoux, W. Krezel, P. Poinot, C. D. Muller, S. Papot and A. Wagner, Eur. J. Med. Chem., 2017, 142, 376–382.

© 2023 The Author(s). Published by the Royal Society of Chemistry

I. Volatolomique induite

Le cancer est compté parmi les maladies les plus mortelles au monde ayant causé 10 millions de décès en 2020. Durant les 40 dernières années, il a doublé chez les femmes et triplé chez les hommes.[1] Les interventions thérapeutiques fonctionnent d'autant mieux que le cancer est détecté précocement. La nécessité de détecter les tumeurs afin de donner un traitement efficace et augmenter les chances du survie du patient est donc importante.[9] Ainsi, la recherche médicale s'est tournée vers le développement de stratégies de détection précoce, sensible et efficace du cancer.[174] L'objectif de mon travail de thèse est de mettre au point une méthode précoce de diagnostic du cancer, par l'identification de biomarqueurs protéiques. Ce travail fait suite à la thèse de Justin Lange, au cours de laquelle il a travaillé sur le concept de volatolomique induite. Ce concept se base sur la détection de COVs exogènes, pour le diagnostic des maladies. En effet la détection de COVs endogènes peut générer des faux positifs en raison d'une grande variabilité interindividuelle.[6] L'approche développée dans le cadre de la thèse de Justin Lange reposait sur l'utilisation d'une sonde à base de COV ciblant une glycosidase telle que la β-Glu. En présence de cette enzyme, la sonde est hydrolysée, et libère un COV marqué l'éthanol-D_5 (ETOH-D_5). C'est sur ce même principe que s'appuient mes travaux de thèse, appliqués également à la détection d'autres glycosidases (β-GlcNac, β-Gal et α-Fuc). Ainsi, un cocktail de sondes ciblant ces cibles enzymatiques a été établi. Des protocoles de détection de ces glycosidases ont été développés sur des biopsies liquides, solides ainsi que sur des modèles murins.

I.1 Détection de la β-Glu suivant un protocole de référence basé sur une sonde de paranitrophénol

Une méthode standard de détection de la β-Glu a été poursuivie, afin de comparer l'efficacité de cette méthode avec notre stratégie de volatolomique induite. Cette méthode permet la détection de la β-Glu avec le glucuronide de paranitrophénol[175,176] (PNPG). En effet, la β-Glu hydrolyse le PNPG en paranitrophénol (PNP), qui a une absorbance à 400 nm. Le PNPG absorbe à 318 nm. C'est ainsi qu'on peut suivre la réaction d'hydrolyse enzymatique du PNPG en PNP par spectroscopie (Figure 38).

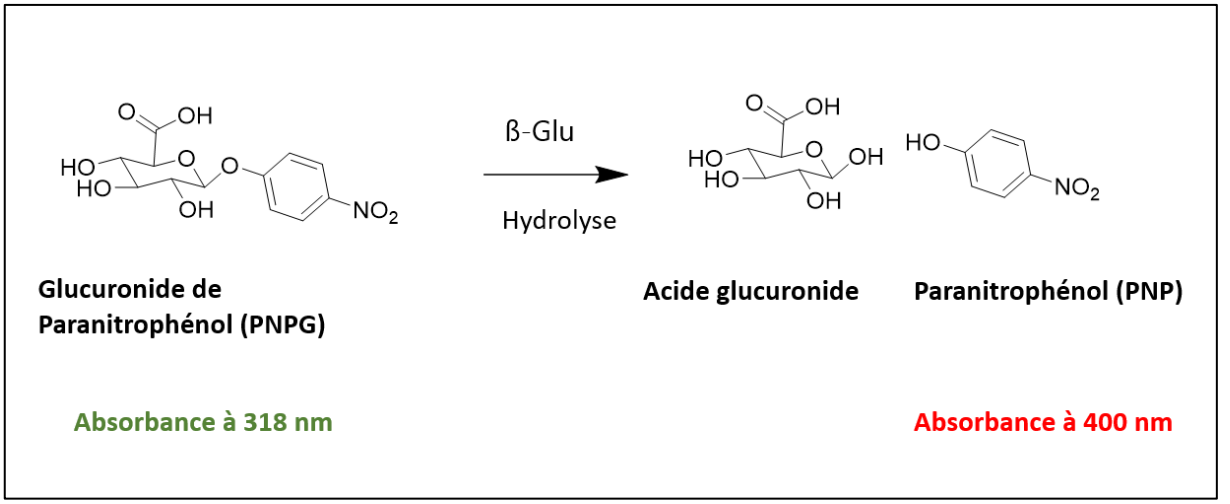


Figure 38: Réaction d'hydrolyse du glucuronide de paranitrophénol (PNPG) en paranitrophénol (PNP) par la β-Glu

La conversion de PNPG en PNP a pu être suivie en additionnant la β-Glu *E.coli* (K.12) à un plasma de patient sain. Il a fallu dépleter le plasma en protéines après réaction. C'est pourquoi une précipitation des protéines avec MeOH/ACN a été réalisée après incubation du PNPG. Le surnageant obtenu était jaune fluorescent et correspondait à l'absorbance du PNP libéré. La réaction a ensuite été suivie par spectroscopie UV-visible.

Dans un premier temps nous avons réalisé une série de tests contrôles en matrice plasma mais sans ajout de β-Glu *E.coli* (K.12). L'objectif de ces expérimentations était de vérifier que l'activité de l'enzyme endogène au plasma était minime. Dans ce cadre, un échantillon de plasma sain a été utilisé, auquel du PNPG a été ajouté. La mesure de l'absorbance a été effectuée 0 min puis 30 min après cet ajout. Les résultats sont présentés Figure 39 et Figure 40.

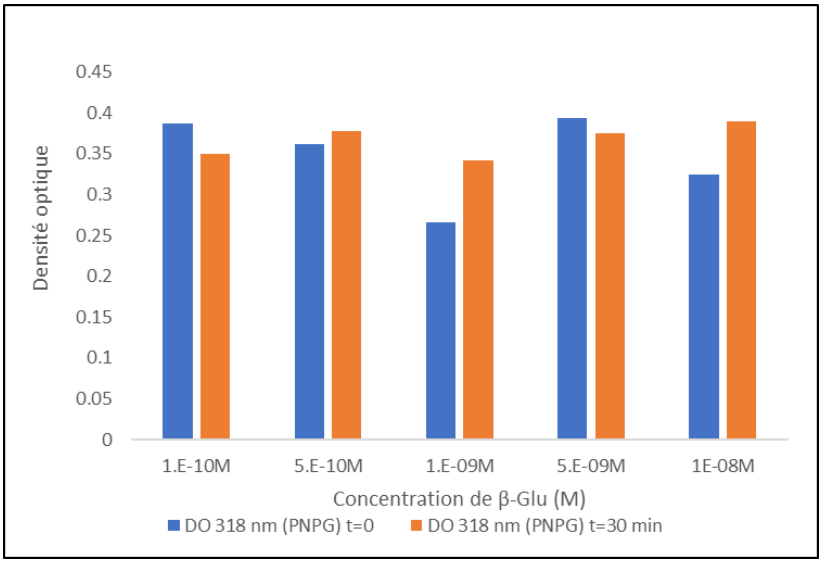


Figure 39: Contrôle négatif de la consommation de PNPG sur matrice plasma sans enzyme

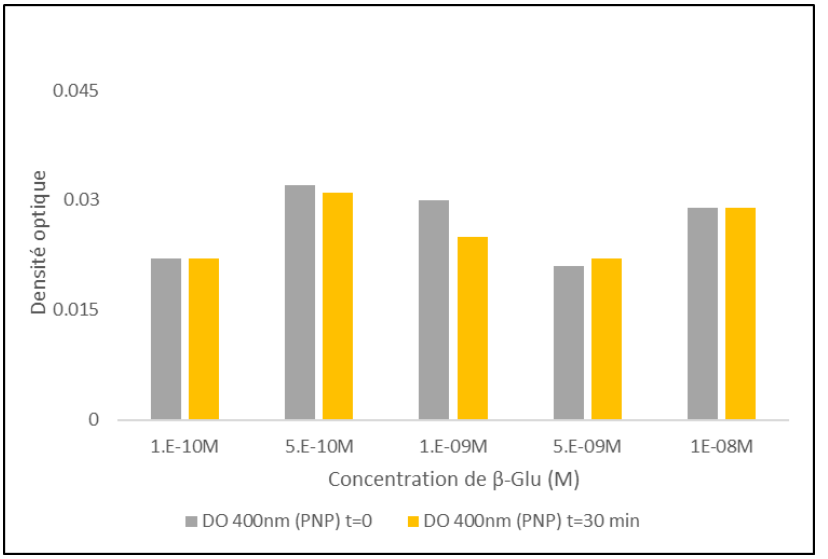


Figure 40: Contrôle négatif de la libération de PNP sur matrice plasma sans enzyme

L'hydrolyse du PNPG par la β-Glu devait induire une baisse de l'absorbance à 318 nm (PNPG) et une augmentation à 400 nm (PNP), selon la loi de Beer Lambert. Sur les Figure 39 et Figure 40, cette tendance n'est pas observée. Le PNPG n'est pas consommé car sa densité optique ne varie pas au cours des 30 min de réaction. De même, la densité du PNP en 30 min reste basale.

Ces résultats indiquent que l'activité de la β-Glu endogène au plasma n'a pas été détectée avec le PNPG. Le test de la réaction d'hydrolyse du PNPG sur matrice plasma avec ajout d'enzyme a donc été mené.
Ainsi le même échantillon de plasma a été additionné de β-Glu *coli* (K12) puis incubé en présence du PNPG. La Figure 41 et la Figure 42 du suivi de l'hydrolyse du PNPG en PNP, illustrent bien la consommation de PNPG au profit de la libération de PNP détecté par spectroscopie, pour différentes concentrations d'enzyme.

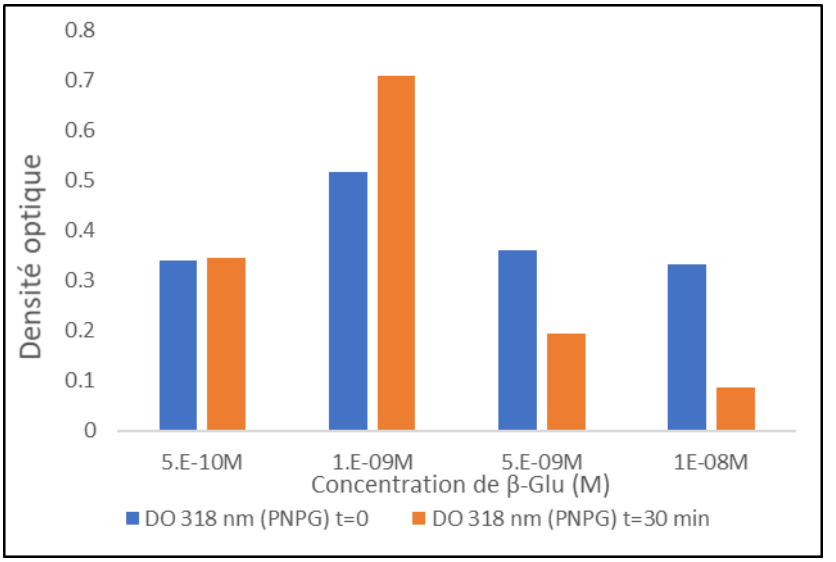


Figure 41: Suivi de la consommation de PNPG sur matrice plasma avec enzyme

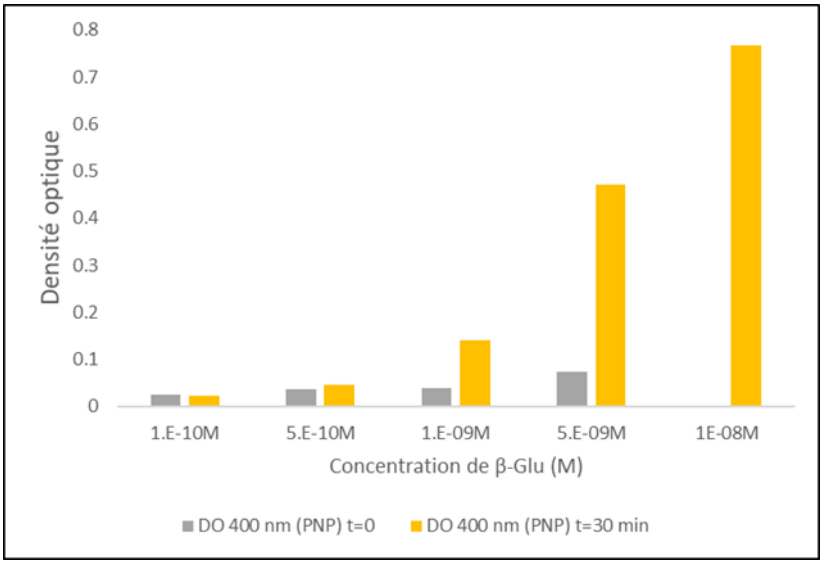


Figure 42: Suivi de la libération de PNP en matrice plasma avec enzyme

Dans ce cadre une gamme de détection de la β-Glu allant de 10^{-10} M à 10^{-8} M a été réalisée en cinq points et en triplicat. Comme le montre la Figure 43, la droite est linéaire. Par ailleurs, la limite de détection de la β-Glu avec la méthode au glucuronide de paranitrophénol est de 10^{-9} M.

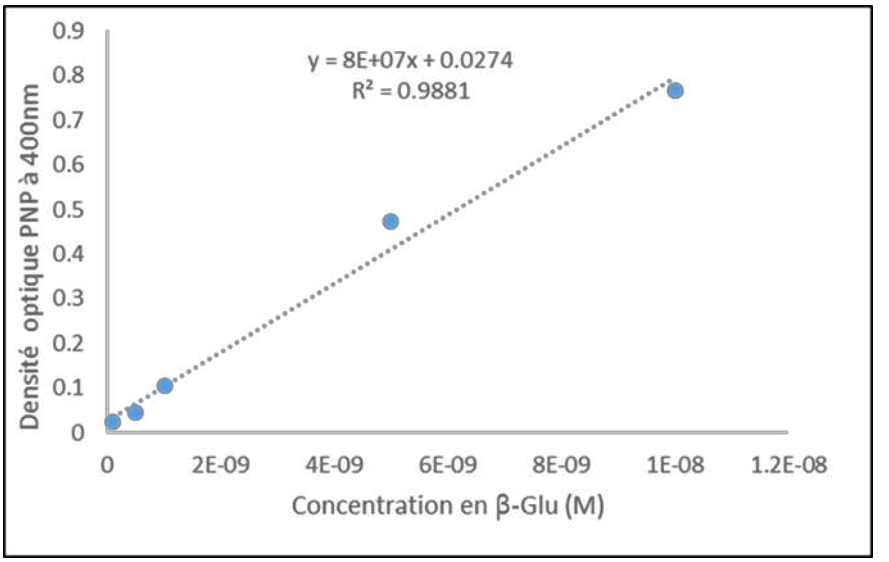


Figure 43: Gamme de détection UV de la β-Glu avec le PNP

La méthode standard de détection de la β-Glu par l'hydrolyse du PNPG en PNP a été validée. Cette technique standard va être comparée à la technique décrite ci-après.

I.2. Gamme de détection de l'EtOH-D_5 et optimisation de ses transitions

L'objectif de notre étude est d'établir une méthode de détection de la β-Glu sur tissu ou plasma qui soit plus sensible que les méthodes standards actuelles telles que l'hydrolyse du PNPG. La détection de l'EtOH-D_5 a été réalisée par MS/MS, il était donc nécessaire de sélectionner les transitions qui pourront au mieux détecter l'éthanol marqué avec une bonne sensibilité et sans interférence avec d'autres composés.

Éthanol marqué	Précurseur (m/z)	Produit (m/z)	Energie de collision (eV)
EtOH-D_5	51.1	30	20
	51.1	**33.1**	**10**
	33.1	30	10

Tableau 15: Transitions sélectionnées lors de l'optimisation de détection de l'EtOH-D_5

L'optimisation de l'EtOH-D_5 a été faite sur GC-TQ (Thermofisher). Différents paramètres ont été évalués, à savoir le rapport m/z du précurseur et de l'ion fragment, et l'énergie de collision permettant de fragmenter l'ion parent en ion fragment (Tableau 15). Ensuite, nous avons choisi la transition qui permettait d'avoir la plus grande intensité en termes d'ionisation, mais aussi la plus grande spécificité. En effet, il a fallu s'assurer que la transition choisie ne répondait que pour l'EtOH-D_5 et pas l'éthanol non marqué. La transition en gras est celle qui a été retenue pour détecter l'EtOH-D_5, car les autres transitions ne permettaient pas de détecter l'EtOH-D_5 de façon spécifique.

I.2.1. Limite de détection de l'EtOH-D_5

Après sélection de la transition la plus pertinente, la gamme de détection de l'éthanol marqué a été réalisée. Des solutions avec des concentrations allant de 5.10^{-9} M à 10^{-6} M ont été préparés et mises dans des vials SPME. Dans chacun d'entre eux, une extraction de l'EtOH-D_5 avec une fibre SPME a été réalisée, ceci pour chaque point de gamme, et en triplicat.

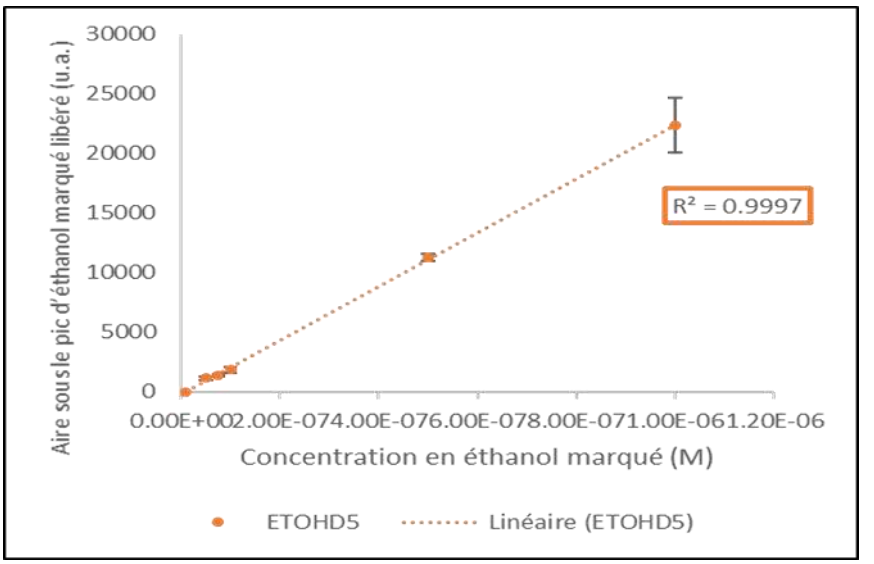


Figure 44: Gamme de détection de l'ETOH-D_5 par SPME-GC-MS/MS

La réponse est linéaire et le R^2 de 0.99. La LOD obtenue était 10^{-9} M. Un étudiant de mon équipe a obtenu la même valeur en faisant également de la SPME-GC-MS/MS. Tout ceci montre la robustesse de la manipulation et l'exactitude du résultat.

I.2.2. Limite de détection de la β-Glu en tampon avec la sonde EtGlu-D5

Après avoir déterminé la LOD de l'ETOH-D_5 et optimisé sa détection, nous avons étudié la limite de détection de notre cible la β-Glu *in vitro* avec la sonde EtGlu-D_5. Pour ce faire, la β-Glu *E. coli* (K12) a été mise en présence de 5.10^{-4} M de sonde EtGlu-D_5, et incubée 7 h avec une concentration variable en β-Glu allant de 5.10^{-12} à 5.10^{-10} M en cinq points avec duplicat. La LOD de β-Glu obtenue *in vitro* en SPME-GC-MS/MS est 10^{-11} M. La gamme de détection est présentée

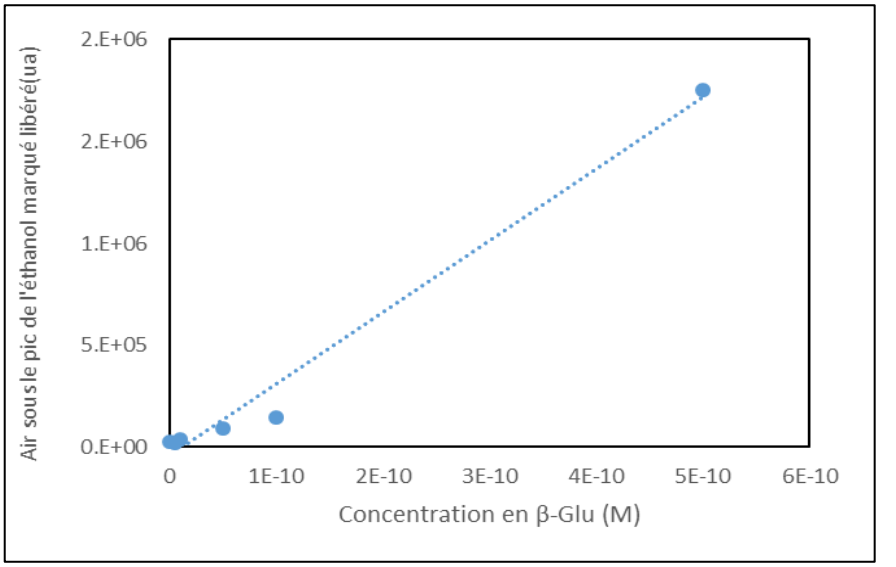


Figure **45**.

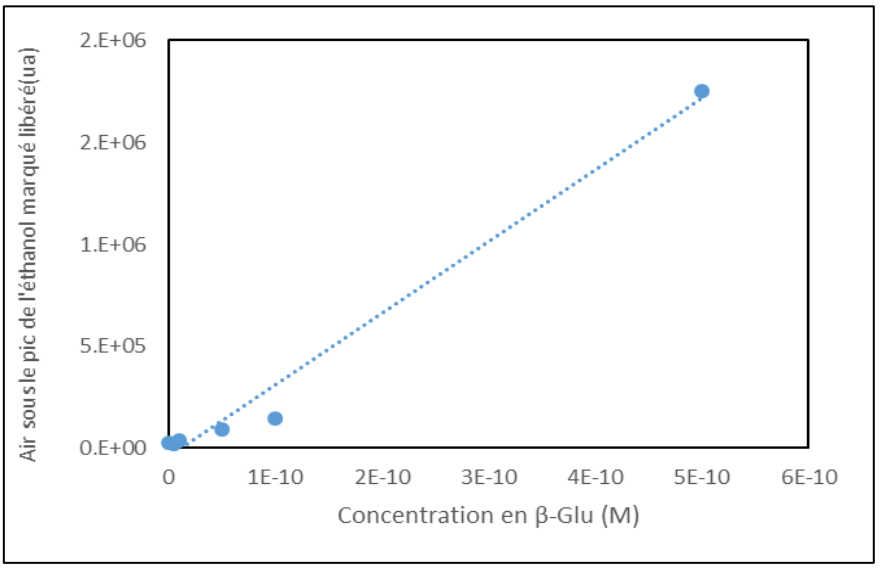


Figure 45: Gamme de détection de la β-Glu *in vitro*

Nous avons ensuite réalisé la même expérimentation en matrice plasma (Figure 46). Pour cela, une gamme de concentration de 5 points en triplicat allant de 1.10^{-13} à 10^{-11} M en β-Glu (K12) a été réalisée. Une cinétique enzymatique composée des temps suivants a été réalisée : T 2.5 h, T 4. 5 h et T 7.5 h. La LOD obtenue de la β-Glu en matrice plasma est comprise entre 1.10^{-13} et 5.10^{-13} M pour le point de cinétique de 7.5 h.

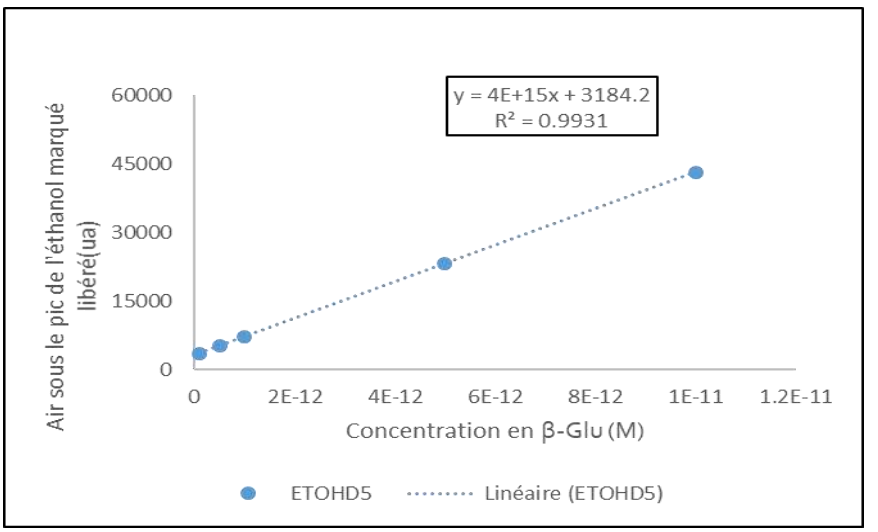


Figure 46: Gamme de détection de la β-Glu en matrice plasma

I.3. Optimisation des conditions de détection de la β-Glu sur biopsies

Afin de confirmer la présence de la β-Glu dans le milieu extracellulaire des tissus tumoraux, une analyse comparative de la réponse extracellulaire et intracellulaire de tissus malins et tissus sains a été réalisée. En d'autres termes, un ajout croissant de sonde d'EtGlu-D_5 a été effectué sur des tumeurs KB et des foies murins. Pour de faibles concentrations, la sonde est potentiellement hydrolysée par l'enzyme extracellulaire. En revanche, passé une concentration, la sonde d'EtGlu-D_5 pénètre les cellules et on obtient alors une réponse de la β-Glu intracellulaire (présente dans les lysosomes). De plus, entre les tissus sains

et malins, seuls les tissus malins devaient donner une réponse extracellulaire. En effet, le foie possède de la β-Glu intracellulaire, mais pas extracellulaire.

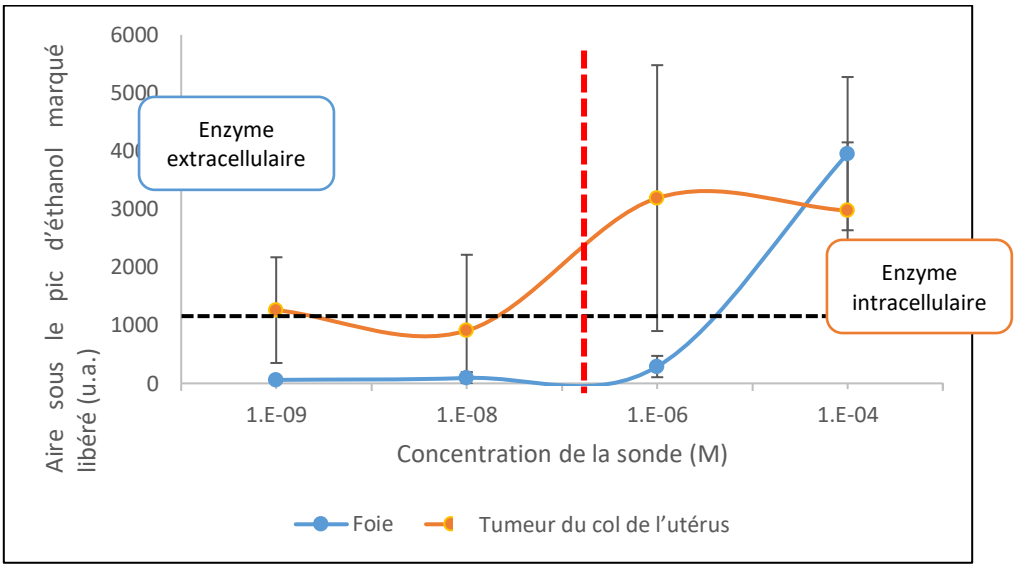


Figure 47: Détection de la β-Glu sur les biopsies.
La ligne en pointillés noirs correspond au seuil de détection d'ETOH-D5 (1000 u.a). La ligne en pointillés rouge correspond à la limite entre la réponse intra- et extra-cellulaire de la β-Glu

La
Figure **47** montre que, pour les faibles concentrations de sonde de 10^{-9} M à 10^{-8} M, le signal en ETOH-D_5 reste dans le bruit de fond, que ce soit pour le foie ou la tumeur.
À partir de 10^{-6} M de sonde, une réponse de la tumeur est observée. Le foie ne répond pas. Ceci correspondrait à la réponse de la β-Glu présente dans la matrice extracellulaire de la tumeur. À partir de 10^{-4} M, la réponse est observée pour les deux tissus. La sonde a pénétré les cellules des deux tissus et a été hydrolysée par la β-Glu intracellulaire. La concentration de sonde à ajouter au milieu pour détecter la réponse de la β-Glu extracellulaire est donc de 10^{-6} M. Ces résultats confirment la présence extracellulaire de la β-Glu dans le microenvironnement tumoral.

I.4. Optimisation des conditions de détection de la β-Glu dans le plasma

L'objectif était ici de préserver au mieux la stabilité des enzymes du plasma dans tout le processus de l'étude, à savoir de son prélèvement en clinique, à sa conservation et analyse dans notre laboratoire. L'étude a été réalisée afin d'avoir un protocole simple et efficace, utilisable en clinique comme au laboratoire. En ce sens, plusieurs points ont dû être soulevés, car il y a un risque que les glycosidases ne soient pas stables dans le plasma, ce qui peut biaiser leur réponse et donc nos résultats.
Les différents paramètres ayant une influence sur la stabilité de l'enzyme dans le plasma sont les suivants :

- L'activité des protéases présentes dans le plasma
- L'effet congélation/décongélation du plasma
- La durée de congélation du plasma

L'altération ou la perte d'activité enzymatique peut être due à l'activité des protéases naturellement présentes dans le plasma. C'est pourquoi détecter des protéines en matrice plasma requiert l'utilisation d'un inhibiteur de protéases. Cet agent chimique va bloquer l'action de protéases et ainsi préserver la stabilité des enzymes et leur activité enzymatique.

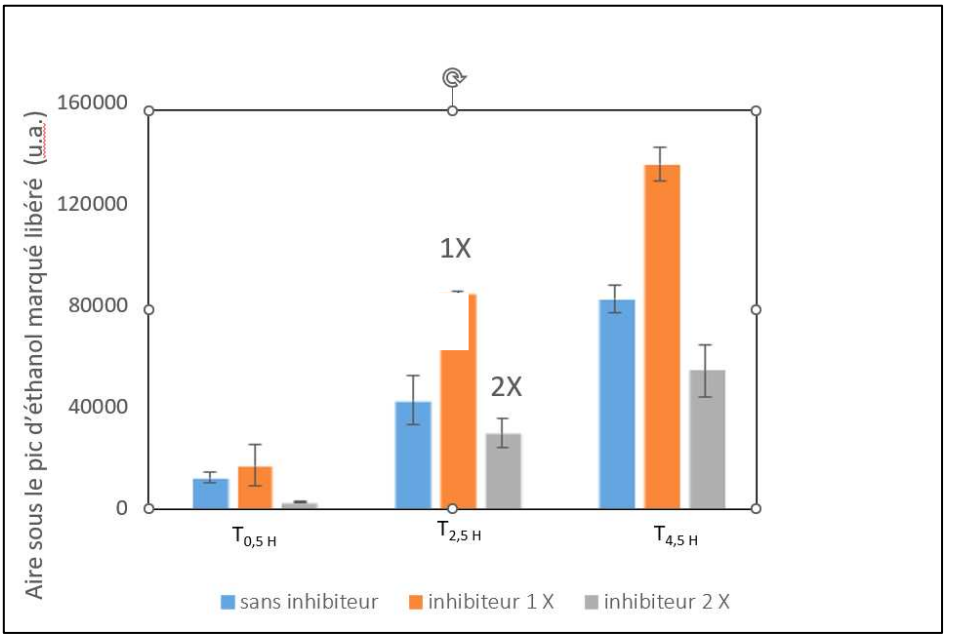


Figure 48: Étude de l'effet des protéases sur l'activité enzymatique

Les résultats présentés sur la Figure 48 montrent qu'une concentration fixe d'inhibiteur de 1X doit être respectée pour conserver une activité enzymatique. Au-delà de 1X en inhibiteur, l'activité enzymatique chute. C'est en cela que la technicienne du CHU, dans le cadre de l'étude clinique, a été informée d'utiliser l'inhibiteur à 1X avant de nous transmettre les plasmas.

L'effet des congélations et décongélations successives du plasma a ensuite été étudié. Un échantillon de plasma (J0) a été additionné à la sonde de β-Glu, puis le plasma a été congelé et décongelé 2 et 3 jours après (J2 & J3). Lorsque les échantillons de plasma sont manipulés, ils sont conservés à -20°C. Ils sont et décongelés à température ambiante lors de l'utilisation. On constate sur la Figure 49, que dès le premier cycle de congélation/décongélation à J2, l'activité enzymatique diminue très fortement, et baisse encore à J3.

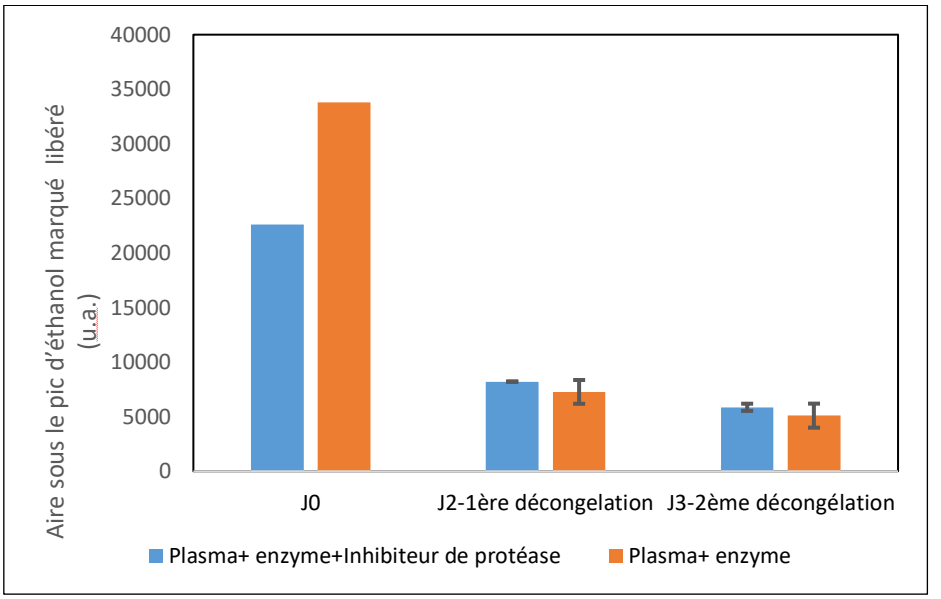


Figure 49: Étude de l'effet des cycles congélation/décongélation du plasma sur l'activité enzymatique

Malgré l'ajout ou non d'inhibiteur de protéases, l'activité enzymatique est très impactée par le cycle congélation / décongélation. Nous avons dû y remédier de la manière suivante. Étant donné qu'à chaque utilisation nous partions d'un même échantillon de plasma que nous décongelions à chaque utilisation, nous avons décidé de faire plusieurs aliquots de ce même plasma avec le volume adéquat. Ainsi, chaque fois que c'était nécessaire, nous sortions un aliquot de plasma, le décongelions et l'utilisions directement.

De la même façon qu'un échantillon de plasma pouvait être congelé et décongelé plusieurs fois, certains restaient plusieurs jours voire semaines au congélateur à -20 °C avant d'être analysés. Nous avons donc étudié l'impact de la durée de congélation du plasma sur l'activité enzymatique (Figure 50). Pour ce faire, un plasma a été mis en présence d'enzyme β-Glu et d'inhibiteur. L'activité de la β-Glu enzymatique a été mesurée au premier jour (J0) puis ré-analysée une semaine après. Il s'avère que l'activité a également baissé malgré la présence de l'inhibiteur de protéases. Cela signifie que la congélation longue durée fait baisser l'activité enzymatique. Cela nous a permis de comprendre et d'expliquer la baisse d'activité observée pour les plasmas qui étaient conservés longtemps.

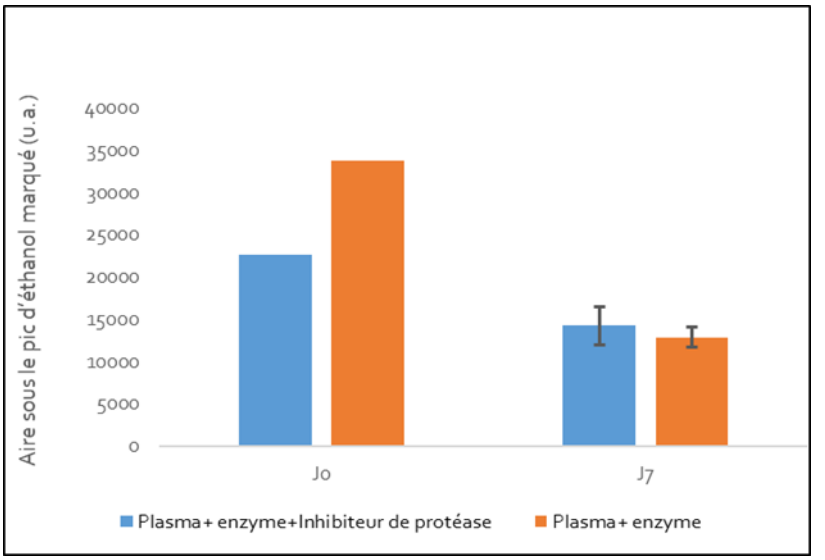


Figure 50: Étude de l'impact de la durée de congélation du plasma sur l'activité enzymatique

I.5. Étude préliminaire clinique

En collaboration avec le Centre d'Investigation Clinique (CIC) du Centre Hospitalier Universitaire (CHU) de Poitiers, une étude préclinique a été menée. Différents lots de plasmas de patients atteints de cancer et de patients sains ont été analysés par SPME-GC-MS/MS après l'addition de la sonde ciblant la β-Glu, à savoir, l'EtGlu-D_5. La quantité d'EtOH-D_5 libérée via l'hydrolyse de la sonde d'EtGlu-D_5 a été mesurée après 2 h, 4 h, et 7 h d'incubation. Les résultats présentés sur la Figure 51 montrent que la réaction d'hydrolyse et la libération d'éthanol marqué ne se fait que pour les plasmas de patients atteints du cancer du sein et du poumon. De plus, les échantillons de plasma sains répondent comme un blanc expérimental. En conclusion, cette étude clinique permet de discriminer des échantillons de plasmas sains des plasmas de patients atteints de cancer. Cela dit, la courbe de réponse des cancers du sein et de cancer du poumon sont identiques. Le ciblage de la β-Glu permet donc de détecter le cancer, mais pas d'identifier la nature de la région touchée. Il est donc nécessaire pour le bien de l'étude, d'augmenter le nombre de cibles enzymatiques pour affiner la technique de diagnostic.

Ainsi nous avons commencé à travailler sur un cocktail de sondes ciblant chacune une activité glycosidase. Une fois activée, chaque sonde libèrera un isotope d'éthanol donné.

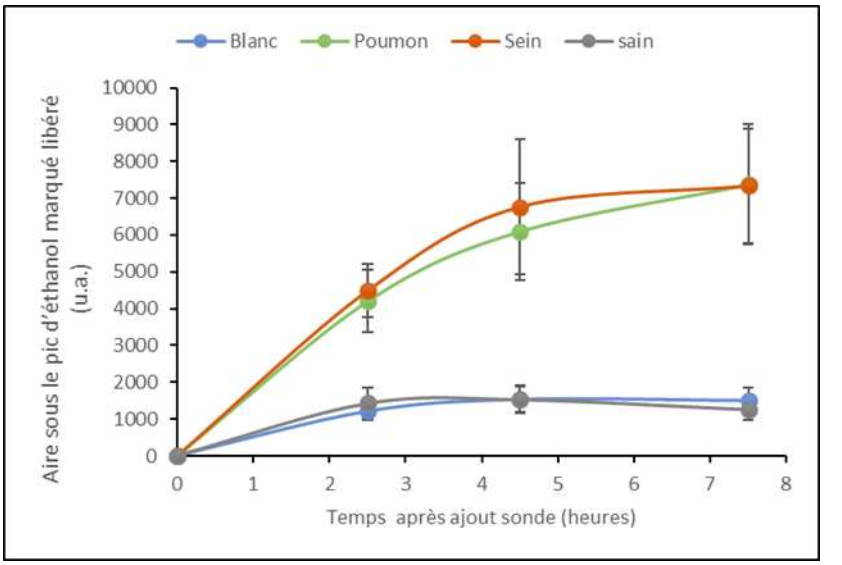


Figure 51: Résultats de l'étude clinique sur les plasmas du CHU

I.5.1. Optimisation des transitions des isotopes de l'éthanol

Étant donné que le projet est basé sur la détection de différents isotopes d'éthanol, il était nécessaire d'établir une stratégie qui permettait de différencier chaque éthanol par MS/MS. C'est pourquoi, l'optimisation des transitions spécifiques à chaque éthanol a été réalisée. Pour cela, 6 isotopes de l'éthanol différents ont été utilisés pour sélectionner uniquement ceux dont les transitions n'interféraient pas avec les autres. Sur ce tableau sont figurés les candidats précurseurs de chaque éthanol et leurs transitions.

Éthanol marqué	Précurseur (m/z)	Produit (m/z)	Energie de collision (eV)
EtOH-^{13}C	47.1	46.1	4
	47	32	6
	47.1	30.1	20
EtOH-$^{13}CD_5$	52.1	50.1	4
	52.1	**34.1**	6
EtOH-D_2	33.1	30	10
	48.1	30	15
	48.1	**33.1**	10
EtOH-D_3	31.1	29.1	10
	49.1	29.1	10
	49.1	31.1	10
	49.1	48	5
EtOH-D_4	49.1	45.1	15
	50.1	**32.1**	10
EtOH-D_5	51.1	30	20
	51.1	**33.1**	10

Tableau 16: Transitions étudiées lors de l'optimisation de détection des éthanols marqués

Parmi les candidats, les isotopes suivants et leurs transitions ont été retenus :

Éthanol marqué	Précurseur (m/z)	Produit (m/z)	Energie de collision (eV)
EtOH-D_2	48.1	33.1	5
EtOH-D_4	50.1	32.1	10
EtOH-D_5	51.1	33.1	10
EtOH-$^{13}CD_5$	52.1	34.1	6

Tableau 17: Tableau des transitions optimales retenues pour la détection des éthanols marqués

En effet, la contrainte était de trouver des transitions spécifiques de chaque éthanol, afin d'éviter la réponse de plusieurs éthanols pour une transition donnée. Au cours de l'optimisation, l'EtOH-D_3 et EtOH-^{13}C ont été retirés du projet. En effet, il s'avère que leurs transitions n'étaient pas spécifiques.

I.5.2. Limite de détection de l'ETOH-$^{13}CD_5$

Parmi ces isotopes, l'EtOH-$^{13}CD_5$ a fait l'objet d'une étude un peu plus approfondie. Ainsi la limite détection de cet éthanol l'EtOH-$^{13}CD_5$ a été déterminée dans les mêmes conditions que l'EtOH-D_5. La gamme de détection a été réalisée de 5.10^{-9} M à 10^{-6} M et chaque point de la gamme a été fait en triplicat. On obtient alors une LOD de 5.10^{-9} M équivalente à celle d l' EtOH-D_5, et une droite linéaire comme le montre la Figure 52.

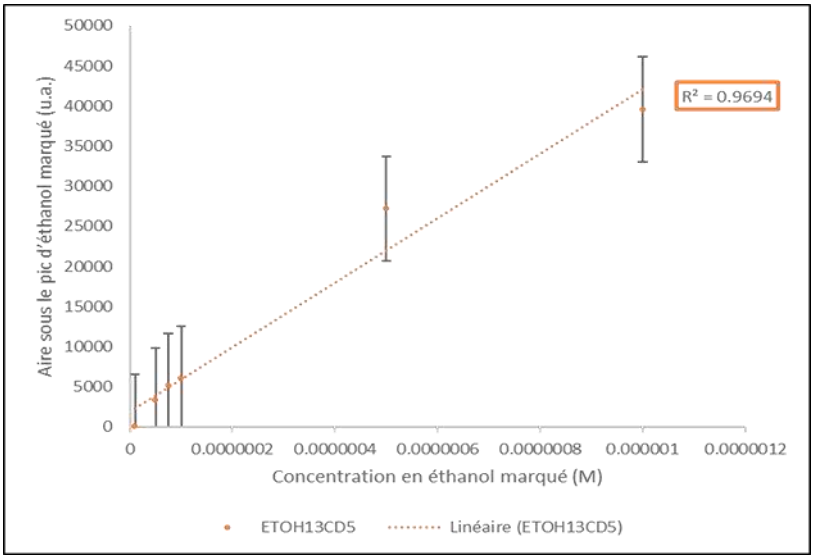


Figure 52: Gamme de détection de l'EtOH-$^{13}CD_5$

I.5.3. Étude du croisement des sondes spécifiques de la et β-Glu et la β-GlcNac

Le projet, qui a pour but la détection de cible enzymatique via l'hydrolyse de sondes à COVs était premièrement fondé sur une première cible, la β-Glu.[5] À terme, le projet a pour vocation de cibler d'avantage de marqueurs.

J'ai donc initié le ciblage d'une deuxième cible enzymatique qui est la β-GlcNac. Cette enzyme présente dans le cancer colorectal et le cancer du sein[164,165] est capable d'hydrolyser la sonde d' EtGlcNac-$^{13}CD_5$ en EtOH-$^{13}CD_5$. Le suivi de l'hydrolyse de cette sonde a été réalisé par SPME-GC-MS/MS.

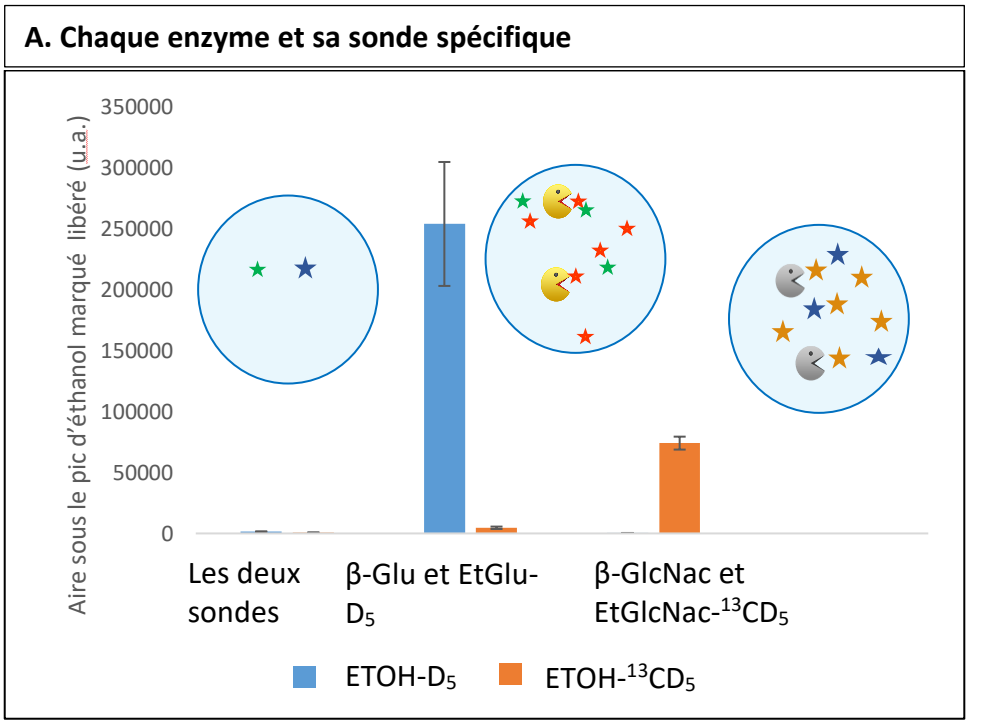


Figure 53: Résultats du croisement de sondes

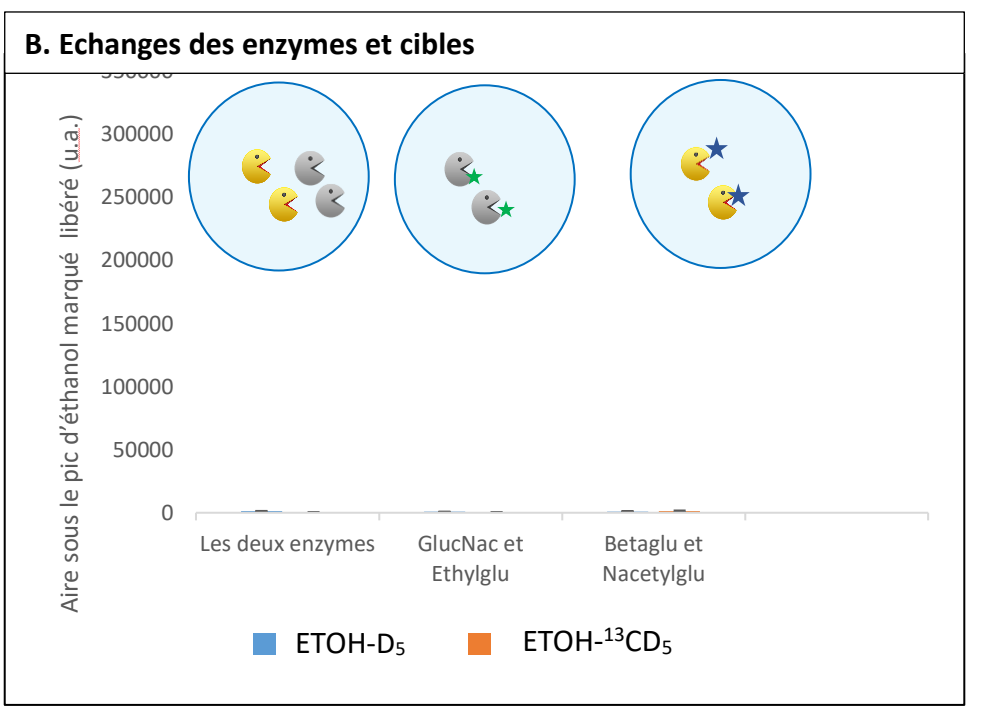


Figure 54: Résultats du croisement de sondes

C. Mélange des enzymes et des sondes

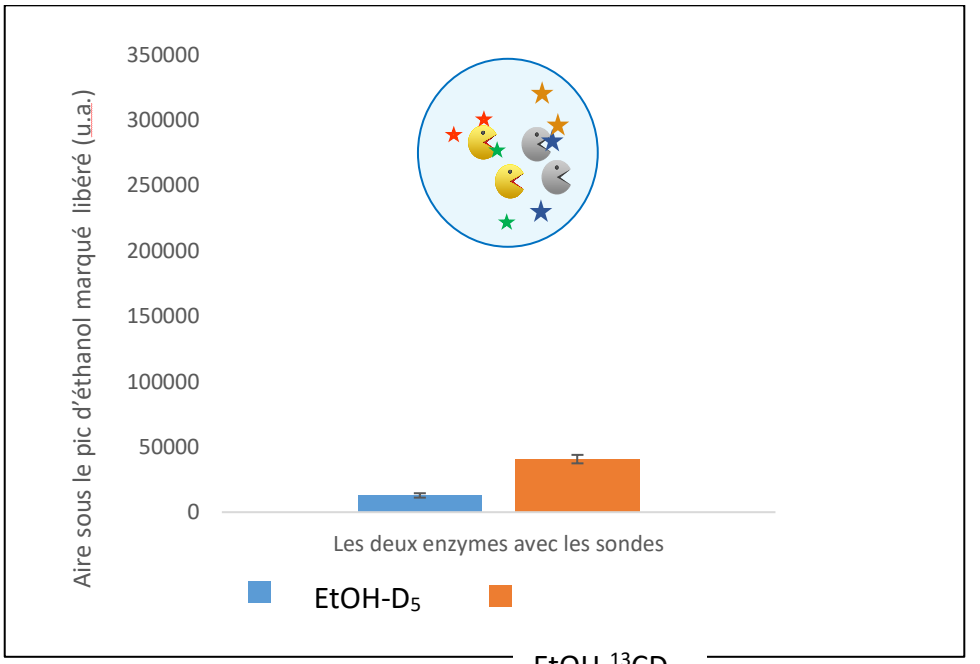


Figure 55: Résultats de croisement de s

Pour s'assurer de la spécificité de chaque enzyme vis-à-vis de sa sonde, une expérience de croisement de sondes a été effectuée. Pour ce faire, les sondes ciblant la β-Glu et la β-GlcNac ont été interverties. La sonde spécifique de la β-Glu a été mise en présence de la β-GlcNac et la sonde spécifique de la β-GlcNac a été mise en présence de la β-Glu. Si l'on compare les réponses de chaque enzyme en présence de sa sonde en Figure 53 , et celle du croisement de sondes en Figure 54, il est clair que les enzymes ne répondent qu'en présence de leur sonde spécifique.
En effet, nous aussi avons mis deux sondes avec une seule des deux enzymes, ou une sonde avec les deux enzymes. Si l'enzyme n'est pas en présence de sa propre sonde, nous n'obtenons aucun signal. En outre, lorsque les enzymes sont mélangées avec les deux sondes (Figure 55), nous obtenons bien une libération d'éthanols marqués qui attestent du bon déroulement des hydrolyses enzymatiques. Dans ce dernier cas, la réponse était 2 à 6 fois moins importante que lorsque les enzymes étaient incubées seules (Figure 53), ce qui est anormal. Ce résultat s'explique par un dysfonctionnement du GC-TQ sur cette période expérimentale. Ce résultat montre également l'importance d'utiliser dorénavant des standards pour qualifier le bon fonctionnement de l'appareil et du bon déroulement de la manipulation (piégeage fibre SPME, température.

I.5.4. Développement d'un cocktail de sondes

Le but de cet essai était de transposer le concept détection des cibles enzymatiques par les sondes à COVs du *in vitro* au *in vivo*, sur modèle murin (Figure 56). C'est en cela que 6 souris Nude ont été xenogreffées avec des tumeurs KB humaines, concourantes à la génération d'un cancer du col de l'utérus. Le suivi du développement du cancer, s'est fait par injection d'un cocktail de sondes à COVs, permettant de tracer l'activité enzymatique de quatre cibles glycosidases et cela pendant 15 jours.

En comparatif, 6 souris saines, non xenogreffées ont subi le même traitement. Le cocktail de sondes associées à chaque enzyme est rappelé dans le tableau suivant :

Enzyme cible du cocktail de sondes	Sonde associée	COV libéré
β-Glucuronidase (β-Glu)	EtGlu-D_5	Éthanol D_5 (ETOH-D_5)
β-N-Acetyl-glucosaminidase (β-GlcNac)	EtGlcNac-$^{13}CD_5$	Éthanol ^{13}C-D_5 (ETOH-$^{13}CD_5$)
β-galactosidase (β-Gal)	EtGal-D_2	Éthanol- D_2 (ETOH-D_2)
α-Fucosidase (α-Fuc)	EtFuc-D_4	Éthanol- D_4 (ETOH-D_4)

Tableau 18: Présentation du cocktail de sondes associées à chaque enzyme

Avant de passer à l'analyse, les conditions de prélèvement des COVs dans l'haleine des souris ont été optimisées. Différents paramètres ont été testés à savoir :

- La taille de la cage
- La durée de position de la souris dans la cage
- Le temps de prélèvement des VOC sur la fibre SPME

Il en est ressorti que pour une souris enfermée dans une grande cage, il était préférable de la laisser pendant 1 h, suivie d'un prélèvement SPME de 30 min. La grande cage a les dimensions suivantes : 130 mm x 130 mm x 200 mm.
Ces paramètres permettent aussi d'éliminer dans le retraitement les signaux des dégagements de COVs dus au métabolisme de la souris.

Une fois les conditions de prélèvement optimisées, les souris saines et cancéreuses ont été placées dans une cage respectivement, et ont subi l'injection du cocktail de sondes (Figure 56).

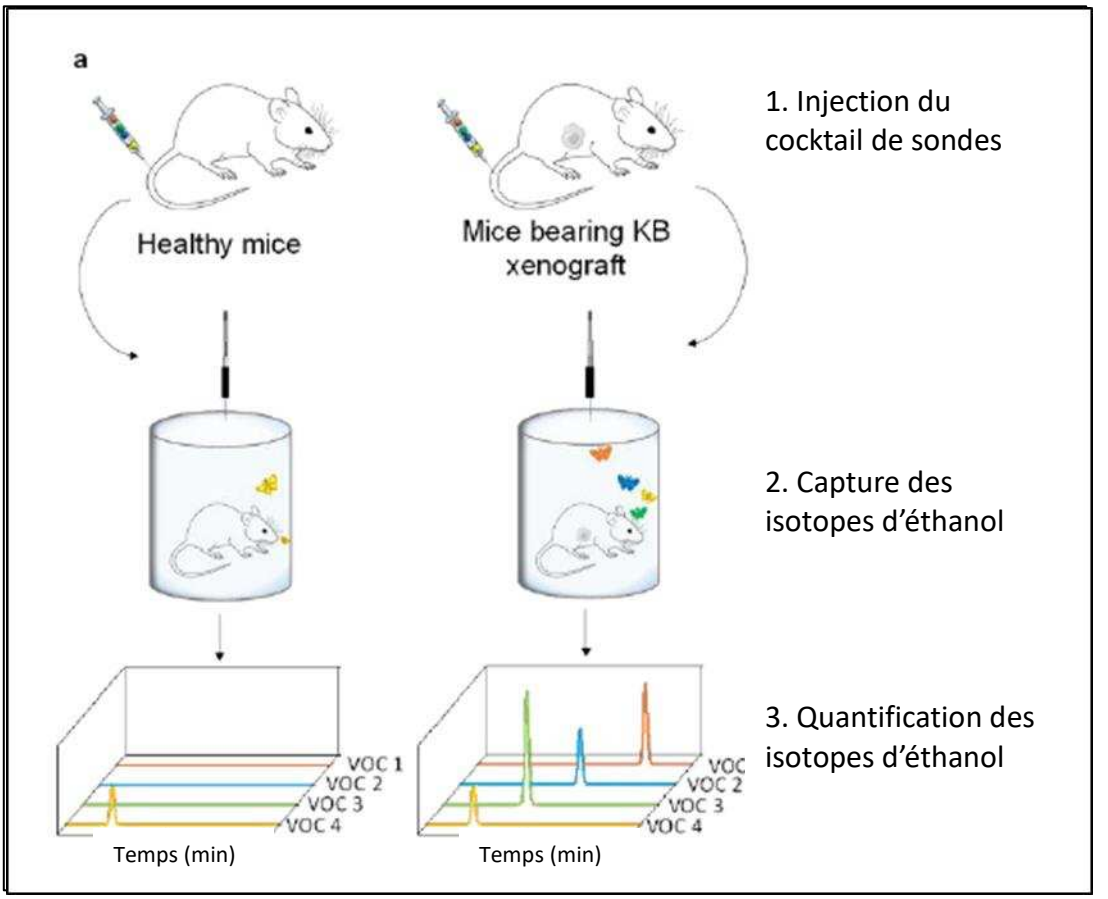


Figure 56: Schéma expérimental de l'utilisation du cocktail *in vivo* (Source image : R. Châtre et al)[7]

Concernant l'injection du cocktail de sondes, il a fallu en amont déterminer les concentrations de chaque sonde. Le but était de trouver la dose minimale de sonde qui permettait d'avoir un signal au-dessus de bruit de fond. Il s'avère que pour les sondes d'EtGal-D_2, EtFuc-D_4, EtGlu-D_5 et le EtGlcNac-$^{13}CD_5$, les doses optimales respectives étaient de 1 µg.kg^{-1}, 100 µg.kg^{-1}, 10 µg.kg^{-1} and 10 µg.kg^{-1} selon les Figure 57, Figure 58, Figure 59, et Figure 60.

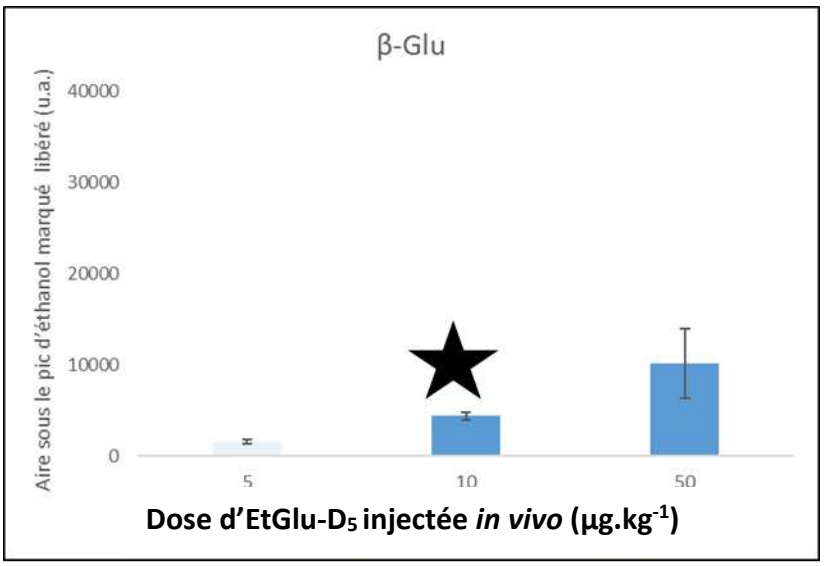


Figure 57: Détermination de la concentration de sondes EtGlu-D_5 à injecter *in vivo*
La concentration optimale de sonde est indiquée par l'étoile noire

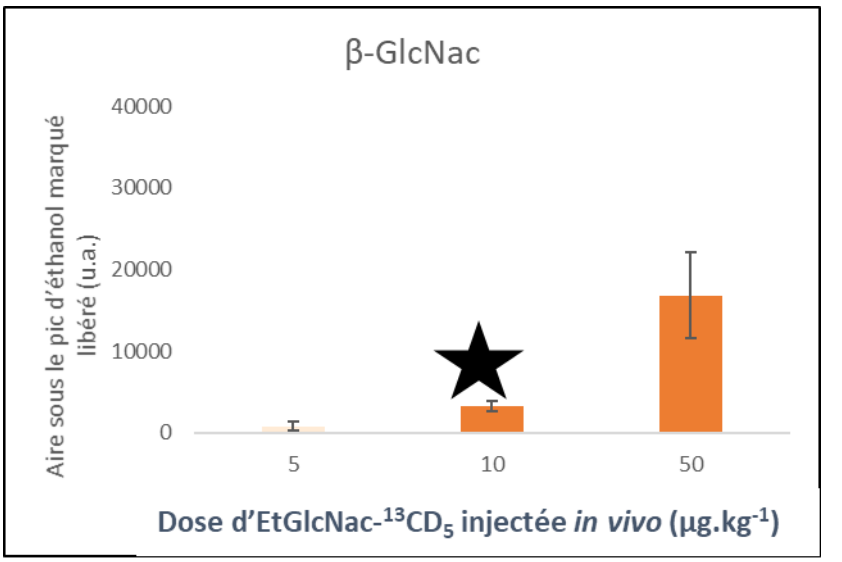


Figure 58: Détermination de la concentration de sondes EtGlcNac-$^{13}CD_5$ à injecter *in vivo*
La concentration optimale de sonde est indiquée par l'étoile noire

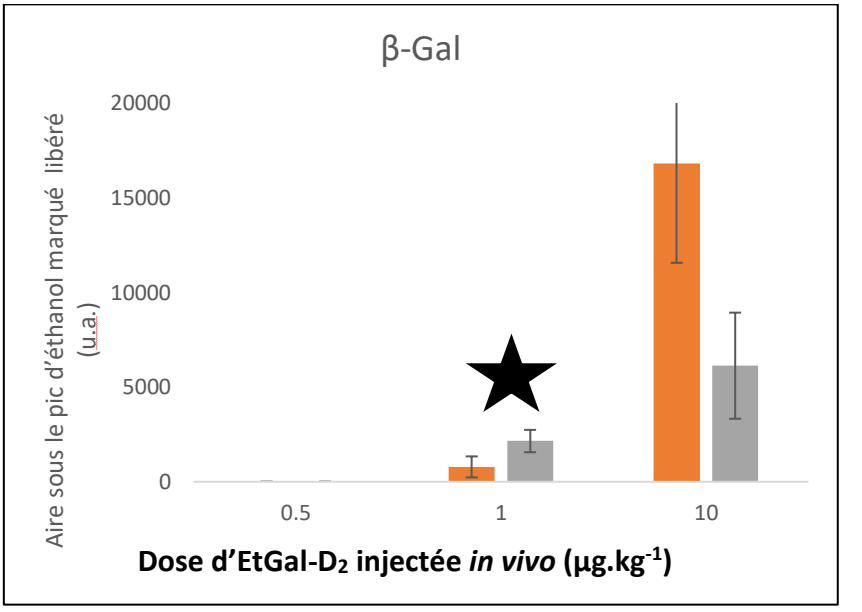


Figure 59: Détermination de la concentration de sondes EtGal-D_2 à injecter *in vivo*
La concentration optimale de sonde est indiquée par l'étoile noire

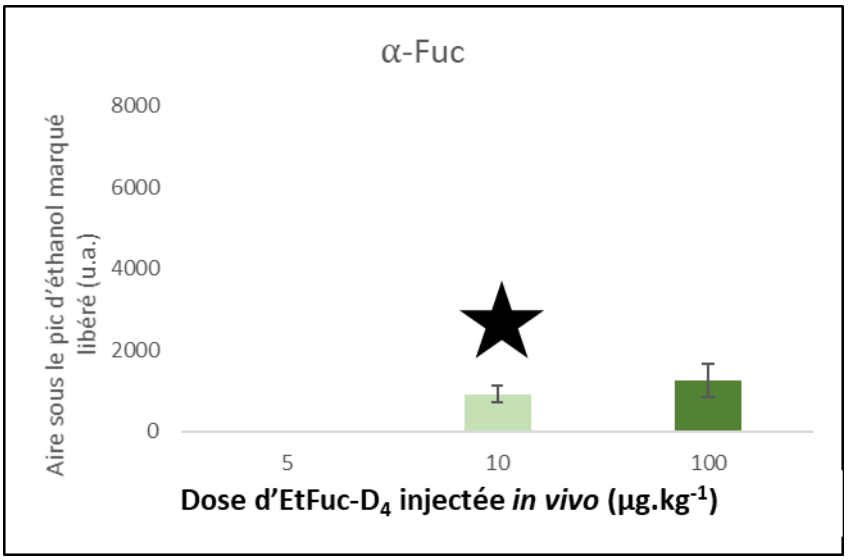


Figure 60: Détermination de la concentration de sondes EtFuc-D_4 à injecter *in vivo*

En ayant établi les conditions de prélèvement et les doses de sondes à injecter aux animaux, des injections du cocktail à 6 souris saines et 6 souris cancéreuses ont pu se faire. Les souris cancéreuses ont été xénogreffées pour générer des tumeurs du col de l'utérus (KB). Leur activité enzymatique a été mesurée via le cocktail de sondes au douzième (J12) et quinzième jour (J15) après la xénogreffe.

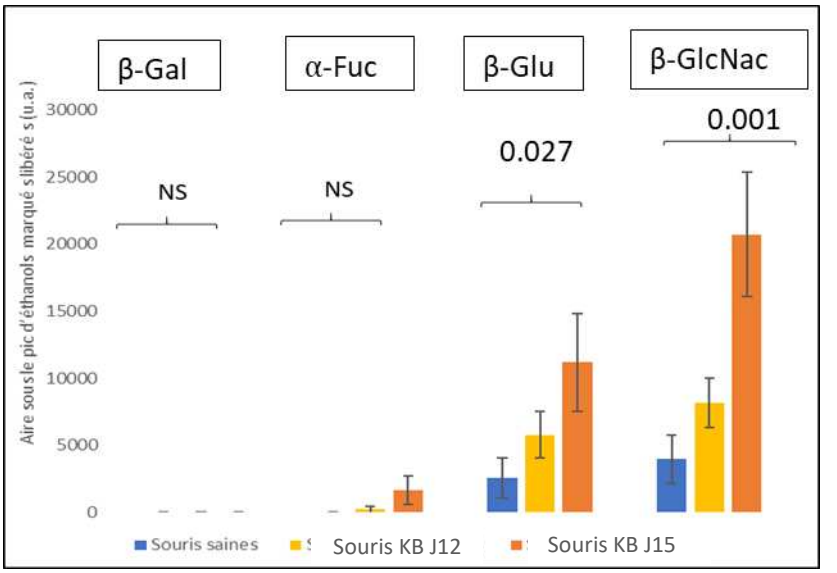


Figure 61: Quantité d'éthanols marqués exhalée par des souris saines (n=6) et des souris xénogreffées (n=6) après injection du cocktail de sondes

Sur la Figure 61, sont représentées les réponses de chaque enzyme après injection du cocktail de sondes et ce, pour des souris saines, et cancéreuses. Il se trouve que pour les souris saines, quelle que soit l'enzyme, la réponse est inférieure à celle des souris cancéreuses. Ces dernières présentent une réelle augmentation du signal entre J12 et J15, caractérisée par une hausse de l'activité enzymatique. Ceci est notable pour la β-Glu et β-GlcNac. Cependant, il n'y a pas d'activité de la β-Gal ou α-Fuc qui soit significative. Cela peut s'expliquer par le fait que dans le cadre de cancer de l'utérus, la β-Glu et β-GlcNac sont sur-régulées, alors que la β-Gal et la α-Fuc sont sous-régulées.

En effet, l'activé enzymatique est différente d'un cancer à un autre. Etablir une combinaison différente et propre à chaque cancer nous permettra à terme de l'identifier via ce cocktail de sondes. Une cartographie des réponses par cancer pourra alors être établie.

II. Détection de la β-Glu par protéomique *middle-down* et *bottom-up*

L'objectif de cette étude est de comprendre l'origine cellulaire des glycosidases. En effet, nous cherchons à savoir si celles-ci proviennent des cellules immunitaires, ou des cellules cancéreuses. De ce fait, si on extrait les glycosidases des cellules immunitaires et les glycosidases des cellules cancéreuses, et qu'on les séquence par protéomique, on connaîtra leurs origines. En effet, dans le cadre de nos expérimentations animales, les souris portaient une xénogreffe de tumeur humaine. Ainsi, si les glycosidases extraites ont une séquence humaine, alors elles sont d'origine cancéreuse. À l'inverse, si les glycosidases de la xénogreffe ont une séquence murine, alors elles proviennent des cellules immunitaires produites par l'animal. En matrice biologique, l'extraction et détection est un défi analytique car ces cibles sont minoritaires. D'autres protéines majoritaires comme l'albumine ou la fibronectine masqueraient le signal des glycosidases.

Au cours du projet, nous avons établi deux stratégies afin d'extraire et d'identifier les glycosidases présentes dans les tissus biologiques : (1) nous avons travaillé sur une stratégie de digestion afin de séparer les protéines minoritaires des protéines majoritaires, (2) nous avons aussi travaillé sur une méthode SEC (Chromatographie d'exclusion stérique) pour séparer les protéines entières.

Dans le cadre de la stratégie (1), La digestion par la trypsine, qui est une technique de protéomique *bottum up* très commune et centrale en protéomique, a été utilisée comme technique de référence. L'hydrolyse au NTCB en revanche, qui est une technique de protéomique *middle down* moins connue et répandue, a été au cœur de notre stratégie de ciblage des glycosidases. L'analyse des résultats a pu être réalisée en faisant en parallèle la digestion de la protéine *in vitro* et *in silico* en utilisant des logiciels ou plateformes de digestion en ligne Avant de faire l'étude sur la β-Glu, celle-ci sera faite sur une protéine standard, qui est l'albumine bovine (BSA).

II.1. Stratégie expérimentale

L'hydrolyse au NTCB permet d'obtenir des peptides plus longs que les peptides trypsiques qui sont de plus petite taille et donc plus difficiles à distinguer les uns des autres, bien qu'ils soient issus de différentes protéines. Le NTCB clive au niveau des résidus cystéines qui sont de surcroît peu abondantes dans les glycosidases ici ciblées. Moins il y a de sites de coupures, plus long seront les peptides.

À l'inverse, l'albumine, protéine majoritaire, compte beaucoup de cystéines et donc génère énormément de petits peptides après hydrolyse au NTCB (Tableau 19, Figure 62).

	Protéine humaine		Protéine murine	
	Cystéines liées	Cystéines libres	Cystéines liées	Cystéines libres
β-Glu	/	6	/	8
β-GlcNac	/	19	/	19
β-Gal	4	4	4	4
α-Fuc	/	5	/	5
Albumine	30	5	34	2

Tableau 19: Liste du nombre cystéines des protéines

Les informations ont été récoltées sur la base de données Uniprot

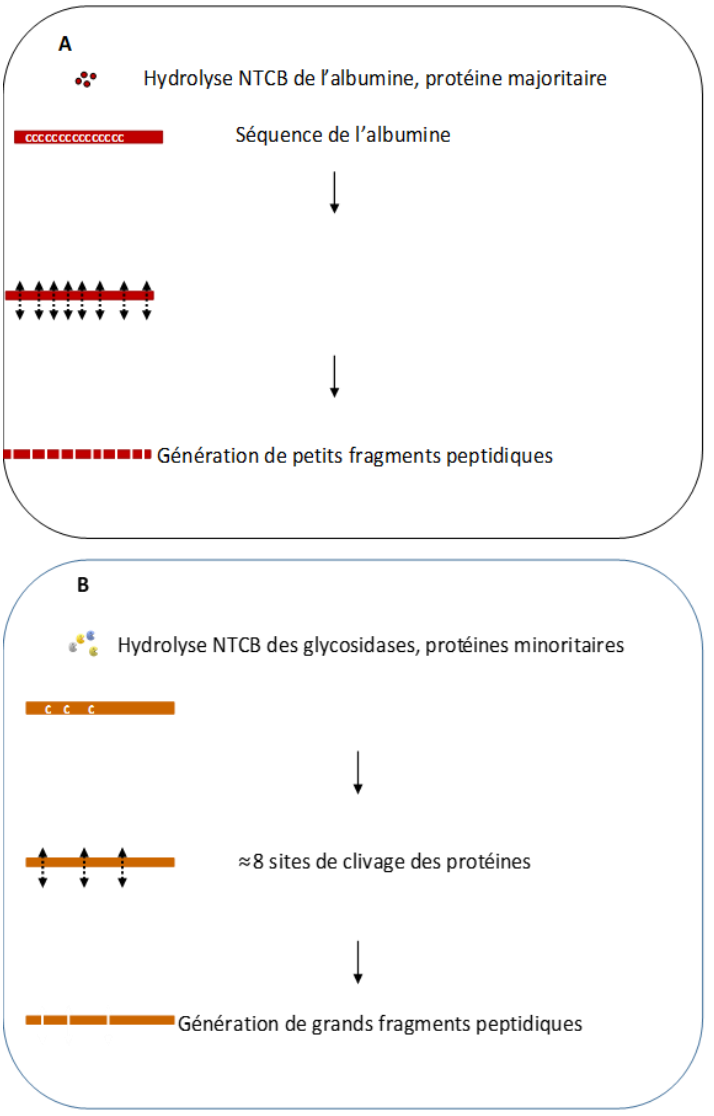


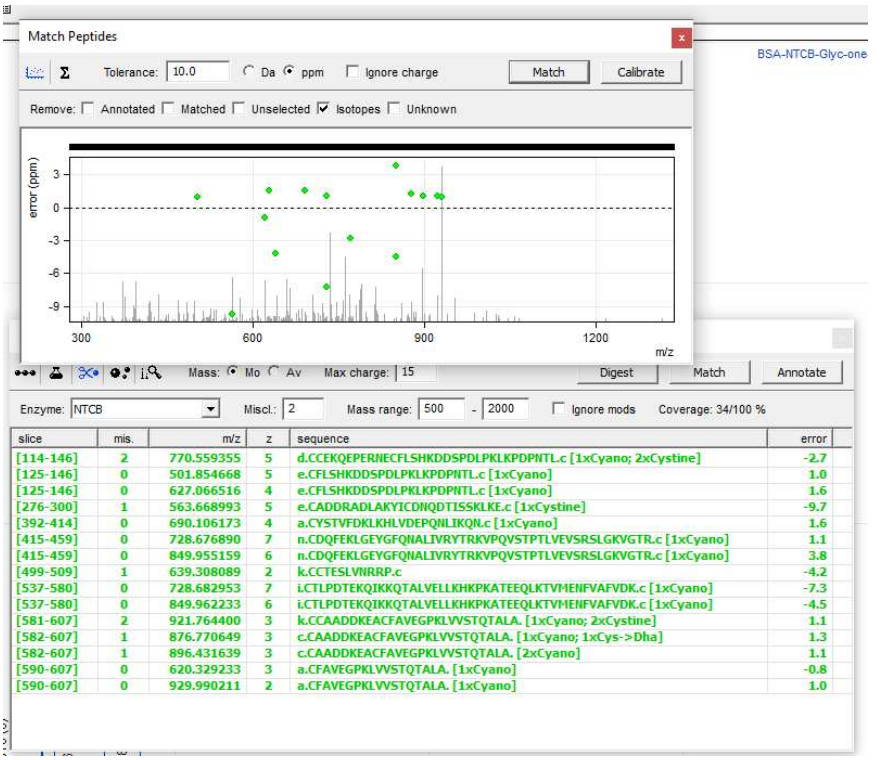
Figure 62: Schéma de l'hydrolyse NTCB de l'albumine et des glycosidases.
A. Hydrolyse de l'albumine**. B.** Hydrolyse des glycosidases

Avant de procéder à la digestion des enzymes au NTCB, il était nécessaire de faire une digestion *in silico* afin de générer la liste des peptides à trouver. Parmi les logiciels de digestion *in silico* à notre disposition, il y a ceux qui génèrent la liste théorique des peptides, que nous pouvons consulter et comparer aux masses expérimentales des peptides que nous avons trouvés. Ce sont les logiciels de production de données théoriques présentés en
Annexe **90:Liste des différents types de logiciels utilisés en protéomique1**

Pour faciliter la manœuvre, il existe un deuxième type de logiciel (ou plateforme en ligne) dans lequel la recherche des peptides se fait de manière automatique. Ce sont des logiciels de traitement de données brutes. Un point à soulever est que la majorité des logiciels de protéomique, qui exploitent les données brutes LC-HRMS, sont basés sur une banque contenant déjà tous les peptides à chercher. Seulement cela ne concerne pas les peptides issus de digestion au NTCB, ce qui fût un grand frein à l'étude.
Il fallait donc trouver un logiciel qui puisse faire la digestion *in silico* du NCTB et traiter les data de manière autonome.

Le premier logiciel qui pouvait correspondre était massXpert, un logiciel d'exploitation des données brutes en protéomique. Il a été découvert à partir du livre « Manuel de spectrométrie de masse à l'usage des biochimistes » de Filippo Rusconi.[177] Le logiciel était plutôt complet, avec les différentes étapes de digestion *in silico* citées en Annexe 96. Cependant la plateforme était difficile à utiliser, malgré le manuel fourni. Le peu qui a pu être obtenu de ce logiciel est la liste des données théoriques. Malheureusement il fut impossible d'exploiter les données brutes. Malgré l'échec d'utilisation de massXpert, qui bloquait donc l'exploitation des premiers résultats d'hydrolyse au NTCB, il semble que ce n'était pas la manipulation qui était problématique, mais plutôt l'exploitation des résultats. Il fallait trouver « la bonne loupe » pour trouver les peptides de l'hydrolyse au NTCB. Nous avons alors choisi le logiciel mMass suivant Strohalm et. al.[178–180]

Pour tester le logiciel de protéomique mMass, nous avons commencé par la digestion au NTCB de la BSA, protéine de référence. Sur la Figure 63, nous pouvons observer les des peptides qui correspondent représentés par les points verts. On voit aussi dans la liste, que l'exactitude en masse est très petite, allant de 1 à 10 ppm.



slice	mis.	m/z	z	sequence	error
[114-146]	2	770.559355	5	d.CCEKQEPERNECFLSHKDDSPDLPKLKPDPNTL.c [1xCyano; 2xCystine]	-2.7
[125-146]	0	501.854668	5	e.CFLSHKDDSPDLPKLKPDPNTL.c [1xCyano]	1.0
[125-146]	0	627.066516	4	e.CFLSHKDDSPDLPKLKPDPNTL.c [1xCyano]	1.6
[276-300]	1	563.668993	5	e.CADDRADLAKYICDNQDTISSKLKE.c [1xCystine]	-9.7
[392-414]	0	690.106173	4	a.CYSTVFDKLKHLVDEPQNLIKQN.c [1xCyano]	1.6
[415-459]	0	728.676890	7	n.CDQFEKLGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTR.c [1xCyano]	1.1
[415-459]	0	849.955159	6	n.CDQFEKLGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTR.c [1xCyano]	3.8
[499-509]	1	639.308089	2	k.CCTESLVNRRP.c	-4.2
[537-580]	0	728.682953	7	i.CTLPDTEKQIKKQTALVELLKHKPKATEEQLKTVMENFVAFVDK.c [1xCyano]	-7.3
[537-580]	0	849.962233	6	i.CTLPDTEKQIKKQTALVELLKHKPKATEEQLKTVMENFVAFVDK.c [1xCyano]	-4.5
[581-607]	2	921.764400	3	k.CCAADDKEACFAVEGPKLVVSTQTALA. [1xCyano; 2xCystine]	1.1
[582-607]	1	876.770649	3	c.CAADDKEACFAVEGPKLVVSTQTALA. [1xCyano; 1xCys->Dha]	1.3
[582-607]	1	896.431639	3	c.CAADDKEACFAVEGPKLVVSTQTALA. [2xCyano]	1.1
[590-607]	0	620.329233	3	a.CFAVEGPKLVVSTQTALA. [1xCyano]	-0.8
[590-607]	0	929.990211	2	a.CFAVEGPKLVVSTQTALA. [1xCyano]	1.0

Figure 63: Listes des peptides trouvés de l'hydrolyse de la BSA au NTCB

Pour s'assurer de la bonne identification d'un peptide, l'expérimentateur a la possibilité, en cliquant sur un peptide parmi la liste, de générer le profil du spectre de masse théorique de ce peptide, en superposition avec le spectre de masse expérimental. Dans la Figure 64 , le spectre théorique est en rouge, et le spectre expérimental en bleu.

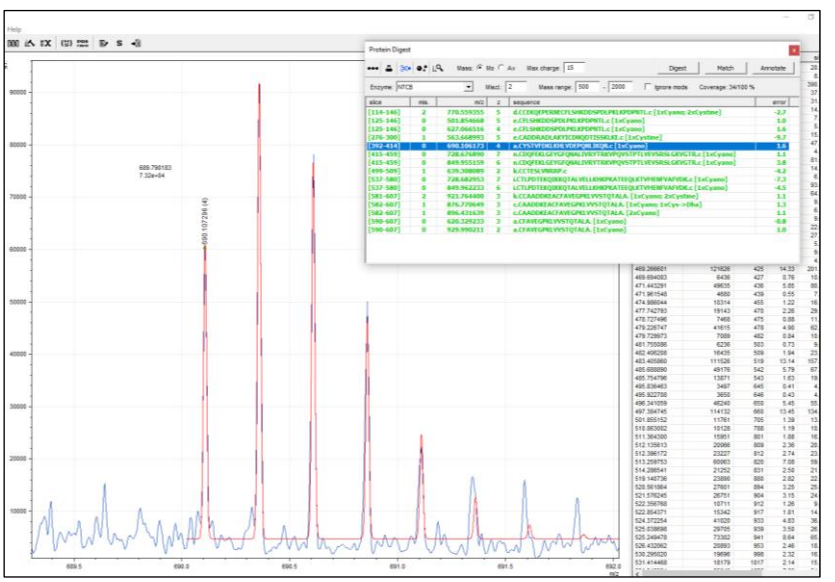


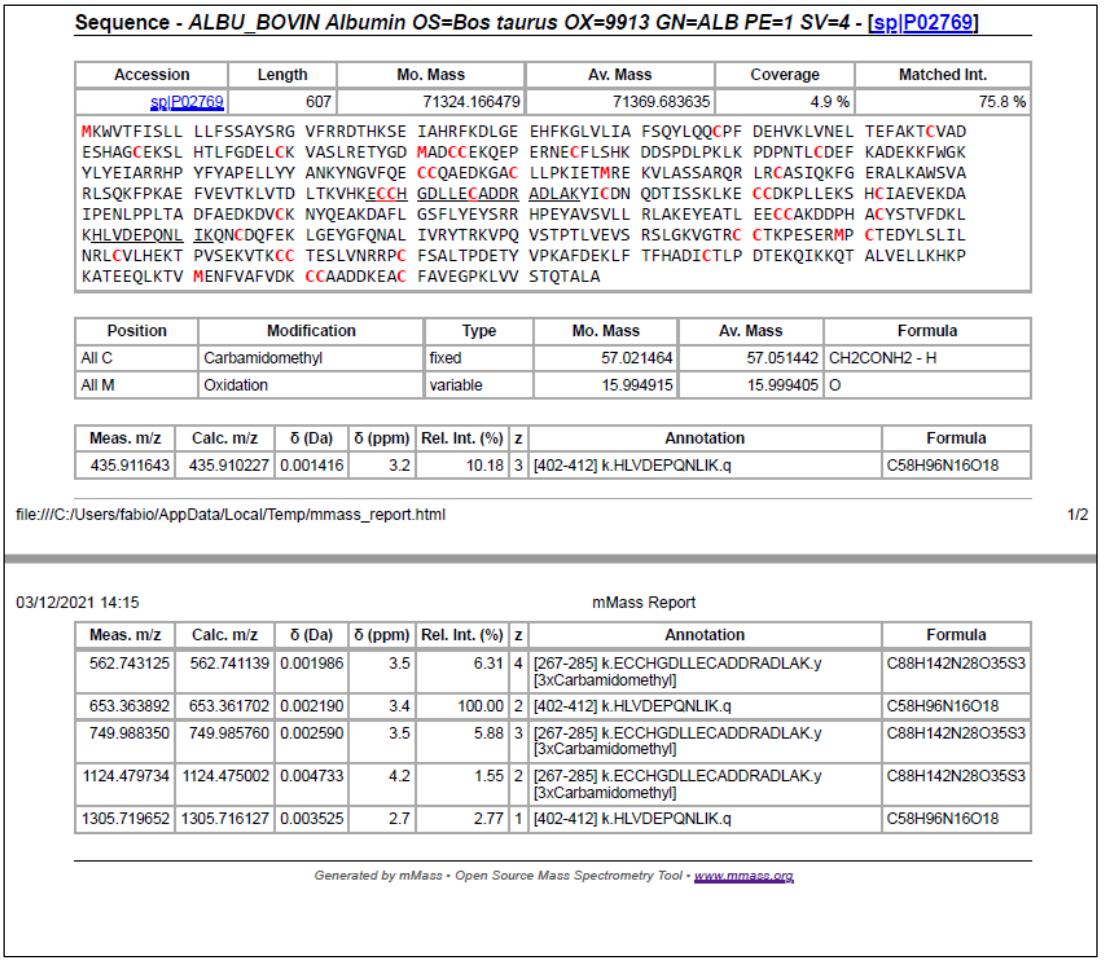
Figure 64: Profil d'un peptide identifié par la plateforme mMass

En conclusion, l'outil mMass a été efficace pour identifier les peptides issus de l'hydrolyse de la BSA au NTCB. De ce fait, toutes les données qui vont suivre ont été traitées avec ce logiciel.

II.2. Exploitation des résultats d'hydrolyse au NTCB : comparaison des différents protocoles

Protocole de digestion trypsique

La trypsine étant une méthode robuste, fiable en protéomique *bottom up,* elle a été utilisée comme méthode de référence. Pour l'utiliser dans mMass, toutes les informations sur les modifications et les sites de clivage étaient déjà renseignées. L'identification des peptides détectés était exacte, avec un écart de masse moyen inférieur à 5 ppm.



Sequence - *ALBU_BOVIN Albumin OS=Bos taurus OX=9913 GN=ALB PE=1 SV=4* - [sp|P02769]

Accession	Length	Mo. Mass	Av. Mass	Coverage	Matched Int.
sp|P02769	607	71324.166479	71369.683635	4.9 %	75.8 %

MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA FSQYLQQCPF DEHVKLVNEL TEFAKTCVAD
ESHAGCEKSL HTLFGDELCK VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK PDPNTLCDEF KADEKKFWGK
YLYEIARRHP YFYAPELLYY ANKYNGVFQE CCQAEDKGAC LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA
RLSQKFPKAE FVEVTKLVTD LTKVHKECCH GDLLECADDR ADLAKYICDN QDTISSKLKE CCDKPLLEKS HCIAEVEKDA
IPENLPPLTA DFAEDKDVCK NYQEAKDAFL GSFLYEYSRR HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL
KHLVDEPQNL IKQNCDQFEK LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS RSLGKVGTRC CTKPESERMP CTEDYLSLIL
NRLCVLHEKT PVSEKVTKCC TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT ALVELLKHKP
KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV STQTALA

Position	Modification	Type	Mo. Mass	Av. Mass	Formula
All C	Carbamidomethyl	fixed	57.021464	57.051442	CH2CONH2 - H
All M	Oxidation	variable	15.994915	15.999405	O

Meas. m/z	Calc. m/z	δ (Da)	δ (ppm)	Rel. Int. (%)	z	Annotation	Formula
435.911643	435.910227	0.001416	3.2	10.18	3	[402-412] k.HLVDEPQNLIK.q	C58H96N16O18

file:///C:/Users/fabio/AppData/Local/Temp/mmass_report.html 1/2

03/12/2021 14:15 mMass Report

Meas. m/z	Calc. m/z	δ (Da)	δ (ppm)	Rel. Int. (%)	z	Annotation	Formula
562.743125	562.741139	0.001986	3.5	6.31	4	[267-285] k.ECCHGDLLECADDRADLAK.y [3xCarbamidomethyl]	C88H142N28O35S3
653.363892	653.361702	0.002190	3.4	100.00	2	[402-412] k.HLVDEPQNLIK.q	C58H96N16O18
749.988350	749.985760	0.002590	3.5	5.88	3	[267-285] k.ECCHGDLLECADDRADLAK.y [3xCarbamidomethyl]	C88H142N28O35S3
1124.479734	1124.475002	0.004733	4.2	1.55	2	[267-285] k.ECCHGDLLECADDRADLAK.y [3xCarbamidomethyl]	C88H142N28O35S3
1305.719652	1305.716127	0.003525	2.7	2.77	1	[402-412] k.HLVDEPQNLIK.q	C58H96N16O18

Generated by mMass - Open Source Mass Spectrometry Tool - www.mmass.org

Figure 65: Résultats de digestion de la BSA avec la trypsine

Protocole d'hydrolyse NTCB avec réduction au DTT en atmosphère inerte et clivage à la soude

Plusieurs protocoles de digestion de la BSA au NCTB ont été testés en amont pour obtenir les conditions idéales d'hydrolyse, en termes de faisabilité, et de robustesse. Ce premier protocole décrit Figure 66 a vu le jour suite à la publication de Srzentic et al 2018.[147] Dans celle-ci, on fait part de la nécessité de réduire les ponts disulfures des protéines en atmosphère inerte. L'optimisation du protocole s'est donc faite sur ce premier point. Afin d'être en atmosphère inerte l'arrivée de gaz N_2 sur le milieu réactionnel s'est effectuée par un flux de gaz dirigé par entonnoir, formant un dispositif en cloche (Figure 67).

1. •Suspension de la BSA dans un tampon Tris acétate 200 mM pH 8
2. •Réduction des ponts disulfures au DTT avec incubation 1 h à température ambiante
3. •Cyanylation par le NTCB 30 min à 40 °C
4. •Précipitation des peptides à l'acétone pH 3 la nuit -20 °C
5. •Centrifugation des peptides et resuspension dans un tampon de borate de sodium 100 mM (pH 9)
6. •Clivage basique à la soude 2 mM
7. •Analyse LC-HRMS

Figure 66: Schéma des étapes du protocole du NTCB par Szrentic et. al

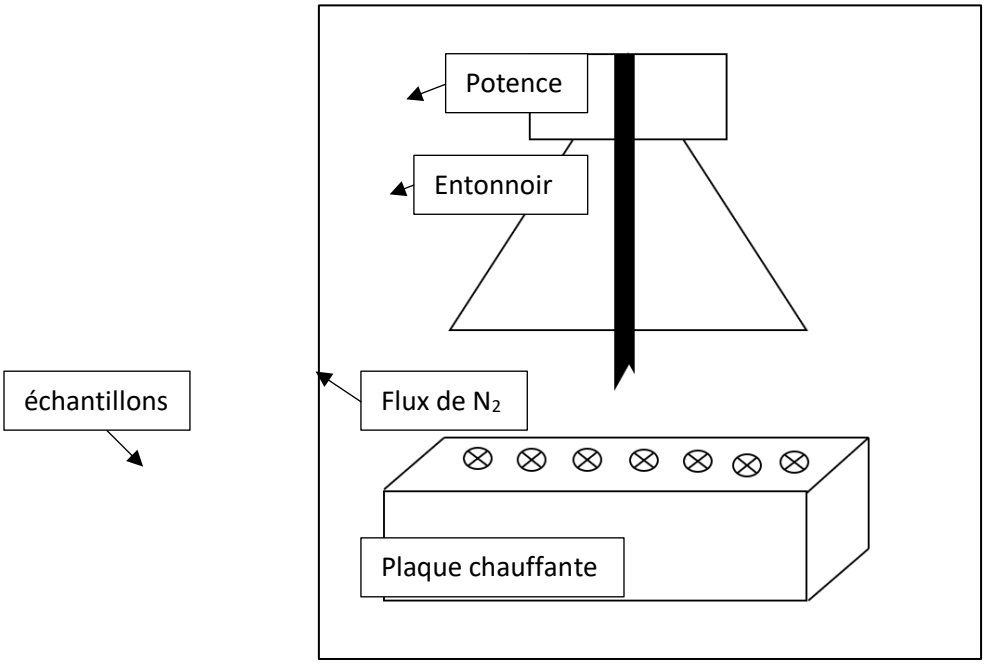


Figure 67: Schéma du montage de réaction en condition inerte avec un dispositif en cloche

La limite de cette technique, est qu'elle ne permettait pas de s'assurer que les tubes réactionnels étaient en présence de N_2, et la mise en place du dispositif était fastidieuse. Avec ce dispositif en cloche, ni l'hydrolyse, ni la détection de peptides n'a été efficiente. Ainsi, nous avons mis en place une arrivée de N_2 directement dans le tube réactionnel scellé. Le montage était plus sophistiqué et plus facile à réaliser (Figure 68).

Cette technique ayant également échoué, la décision de faire la réduction au DTT à pression atmosphérique a été décidée. En effet, dans d'autres protocoles usuels de réduction au DTT, il arrive que la réaction se fasse en pression atmosphérique.

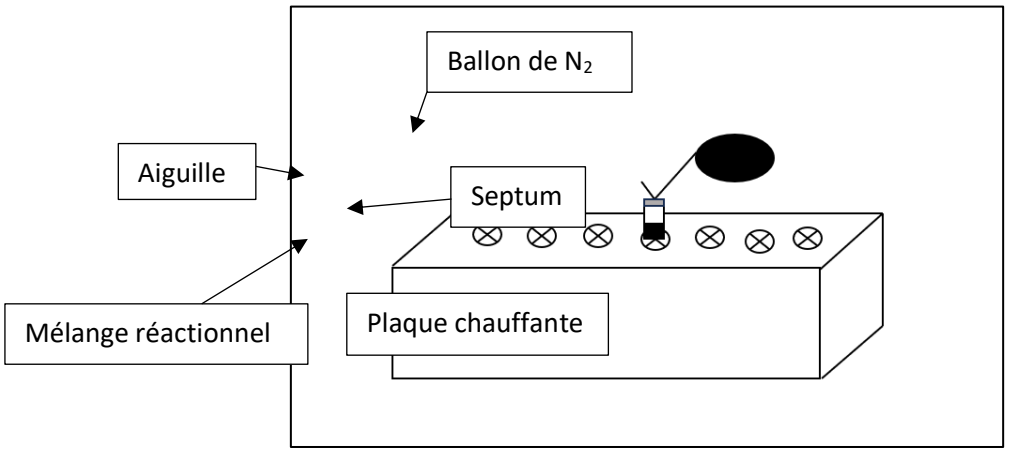


Figure 68: Schéma du montage de réaction en condition inerte avec tube scellé

Malgré une réduction efficace des ponts disulfures avec le DTT, les thiols du DTT rentraient en compétition avec les thiols des cystéines car ces derniers peuvent tous les deux faire l'addition nucléophile sur le NTCB (Figure 69).

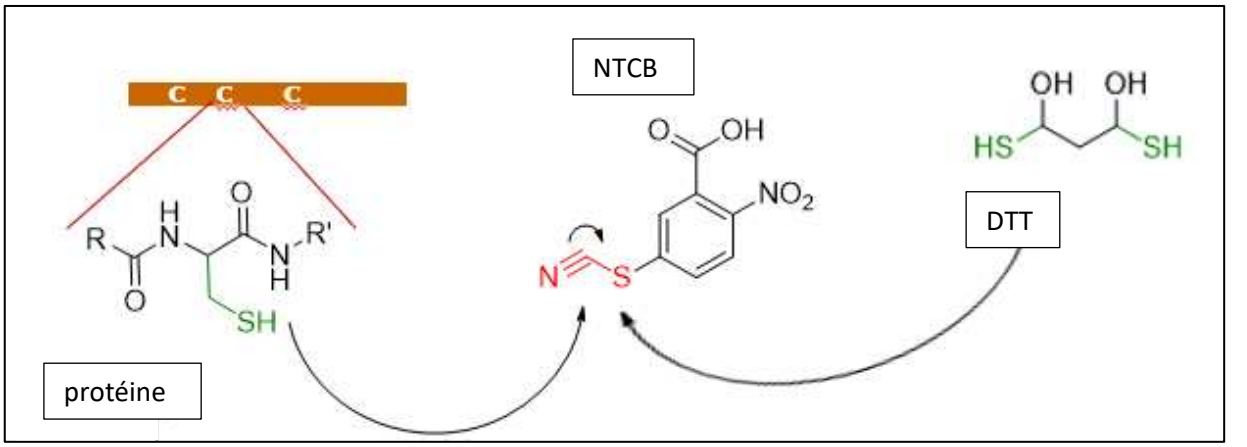


Figure 69: Schéma du mécanisme de compétition entre la protéine et le DTT

Pour s'en affranchir, certains procèdent à la purification des peptides par précipitation pour retirer le DTT avant l'attaque au NTCB.[147,172] Cela a été testé sans succès, tout comme l'utilisation des cartouches PD25 pour faire en changement tampon. Il a donc fallu chercher un autre mode de réduction que le DTT, qui s'est avéré être finalement un inhibiteur de la réaction d'hydrolyse des protéines en raison de sa réactivité avec le NTCB.

<u>Protocole hydrolyse NTCB avec réduction au TCEP et clivage au NH_4OH SOUS pression atmosphérique</u>

À partir de la première expérience, il a été décidé de faire la réduction des ponts disulfure sous pression atmosphérique, et de remplacer le DTT par le TCEP, comme proposé par Koehn et al 2009 (Figure 70).[181]

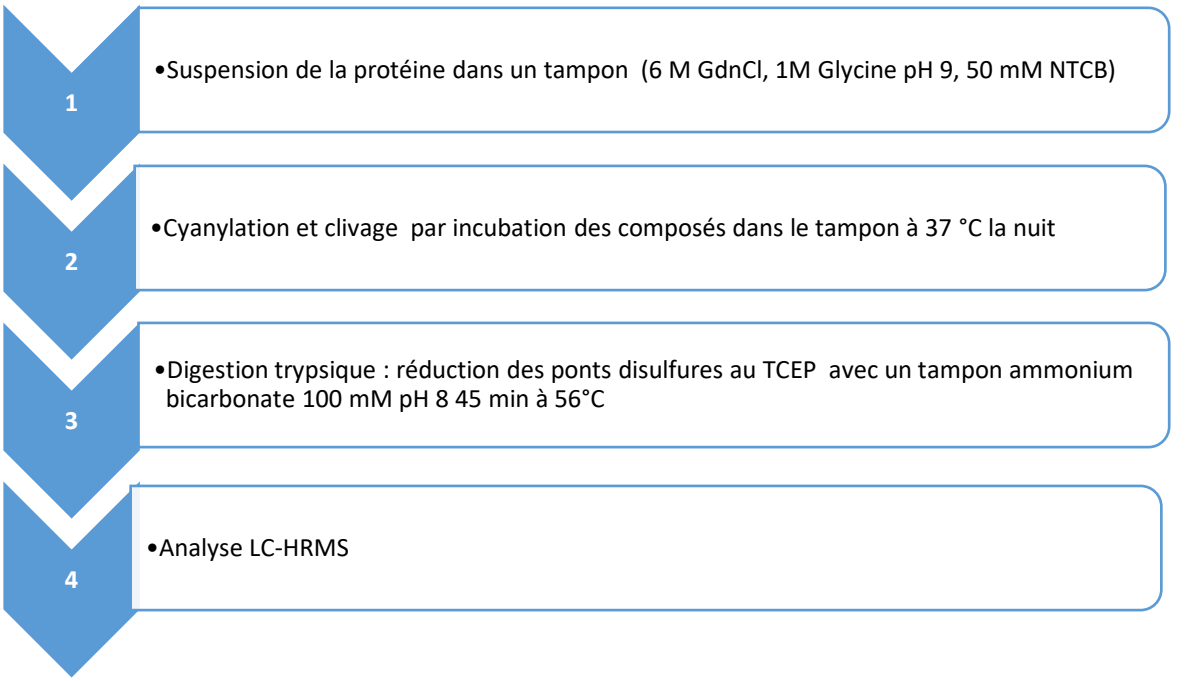


Figure 70: Schéma des étapes du protocole NTCB de Koehn et. al

En effet, le TCEP est un agent réducteur qui ne porte pas de thiols, ce qui n'interférera pas dans la réaction entre la protéine et le NTCB. Il n'y aura donc pas besoin de changer de tampon ou de précipiter les peptides. Ensuite, l'étude de Belghazi et al 2006 (Figure 71) [149] a été suivie pour faire un clivage à l'ammoniaque.

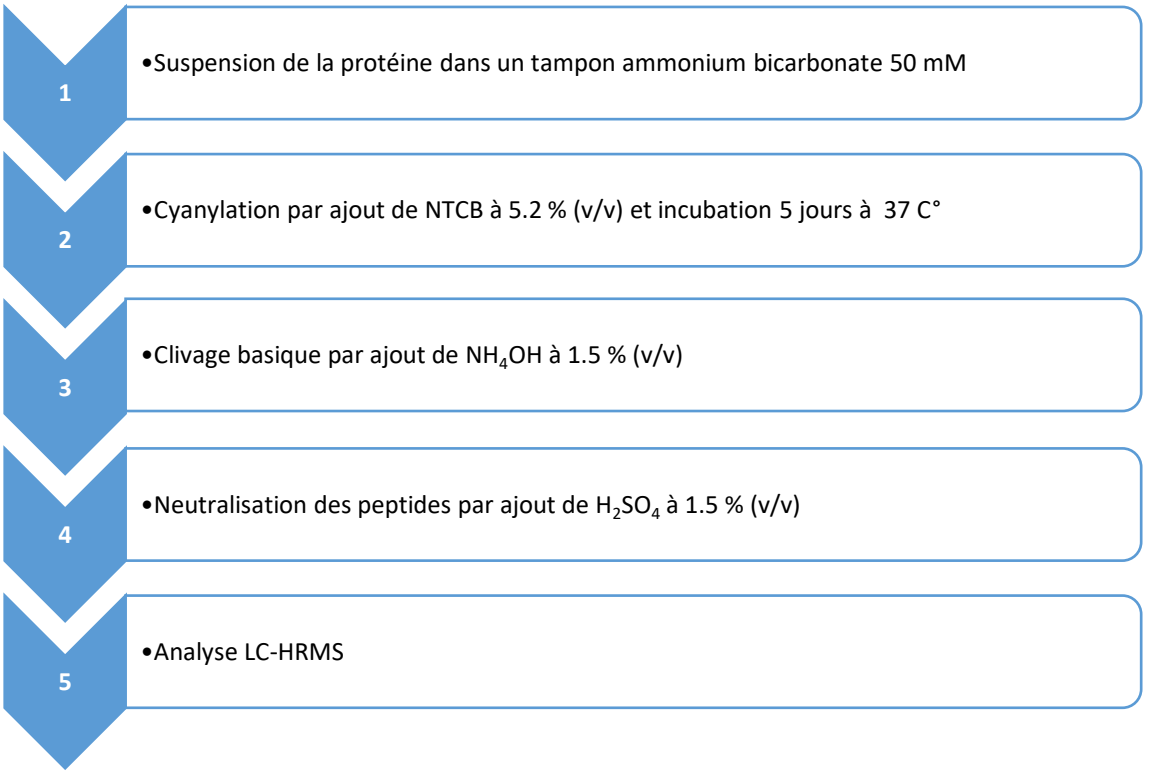


Figure 71: Schéma des étapes du protocole NTCB de Belghazi et. alL'utilisation de la base d'ammoniaque plus douce pour le clivage, a généré moins de réactions secondaires et permet l'obtention de résultats (Figure 72).

Accession	Length	Mo. Mass	Av. Mass	Coverage	Matched Int.

MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA FSQYLQQCPF DEHVKLVNEL
TEFAKTCVAD ESHAGCEKSL HTLFGDELCK VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK
PDPNTLCDEF KADEKKFWGK YLYEIARRHP YFYAPELLYY ANKYNGVFQE CCQAEDKGAC LLPKIETMRE
KVLASSARQR LRCASIQKFG ERALKAWSVA RLSQKFPKAE FVEVTKLVTD LTKVHKECCH GDLLECADDR
ADLAKYICDN QDTISSKLKE CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK NYQEAKDAFL
GSFLYEYSRR HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL KHLVDEPQNL IKQNCDQFEK
LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLCVLHEKT
PVSEKVTKCC TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT ALVELLKHKP
KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV STQTALA

Position	Modification	Type	Mo. Mass	Av. Mass	Formula
All C	Cyano	variable	24.995249	25.009538	CN - H
All C	Cys->Dha	variable	-33.987721	-34.081966	- H2S
All M	Oxidation	variable	15.994915	15.999405	O

Meas. m/z	Calc. m/z	δ (Da)	δ (ppm)	Rel. Int. (%)	z	Annotation	Formula
432.888284	432.850886	0.037398	86.4	0.50	3	[460-470] r.CCTKPESERMP.c [1xOxidation]	C50H85N15O19S3
545.854699	545.864219	-0.009520	-17.4	1.42	5	[288-311] i.CDNQDTISSKLKECCDKPLLEKSH.c [1xCyano; 1xCys->Dha]	C113H185N33O41S2
561.856510	561.873188	-0.016678	-29.7	1.76	5	[460-483] r.CCTKPESERMPCTEDYLSLILNRL.c [1xCyano; 1xCys->Dha]	C119H193N33O39S3
667.322177	667.332713	-0.010536	-15.8	1.03	4	[288-311] i.CDNQDTISSKLKECCDKPLLEKSH.c [2xCys->Dha]	C112H184N32O41S
682.066321	682.078455	-0.012134	-17.8	3.05	4	[288-311] i.CDNQDTISSKLKECCDKPLLEKSH.c [1xCyano; 1xCys->Dha]	C113H185N33O41S2
898.430019	898.447289	-0.017270	-19.2	1.22	3	[461-483] c.CTKPESERMPCTEDYLSLILNRL.c [1xCys->Dha; 1xOxidation]	C115H189N31O39S2
911.114068	911.107044	0.007025	7.7	2.28	3	[581-607] k.CCAADDKEACFAVEGPKLVVSTQTALA. [1xCyano; 1xCys->Dha]	C117H187N31O40S2
921.436404	921.454443	-0.018039	-19.6	1.22	3	[460-483] r.CCTKPESERMPCTEDYLSLILNRL.c [2xCys->Dha; 1xOxidation]	C118H192N32O40S2

Generated by mMass - Open Source Mass Spectrometry Tool - www.mmass.org

Figure 72: Résultats de l'hydrolyse NCTB de la BSA avec réduction au TCEP et clivage à l'ammoniaque

On observe sur la Figure 72, que beaucoup de réactions secondaires ont eu lieu, et la précision de la masse de peptides est majoritairement supérieure à 10 ppm, ce qui n'est pas correct. Cela dit, certaines études ont été publiées avec des écarts similaires voire supérieurs (Koehn et al 2009).[181] Cela est sûrement dû au fait que les peptides ont été obtenus en petite quantité, et que leur intensité ne permet pas une bonne interprétation des résultats. Concernant les modifications, il y a systématiquement la cyanylation, mais aussi la déhydralanination. À la suite de cela, la nécessité de trouver un protocole qui ne favorise que la cyanylation s'est fait sentir.

En conclusion, sur 37 peptides attendus 5 ont été détectés.

Protocole digestion au NTCB avec clivage à la glycine dans un seul mélange réactionnel

Le fait de faire une réaction multi-étapes avec une réduction suivie d'une cyanylation puis d'une coupure, favorise la perte des peptides. C'est pourquoi le protocole avec un seul mélange réactionnel, aussi appelé « one pot » a été réalisé, à partir d'une étude de Tang et al 2004 (Figure 73).[173]

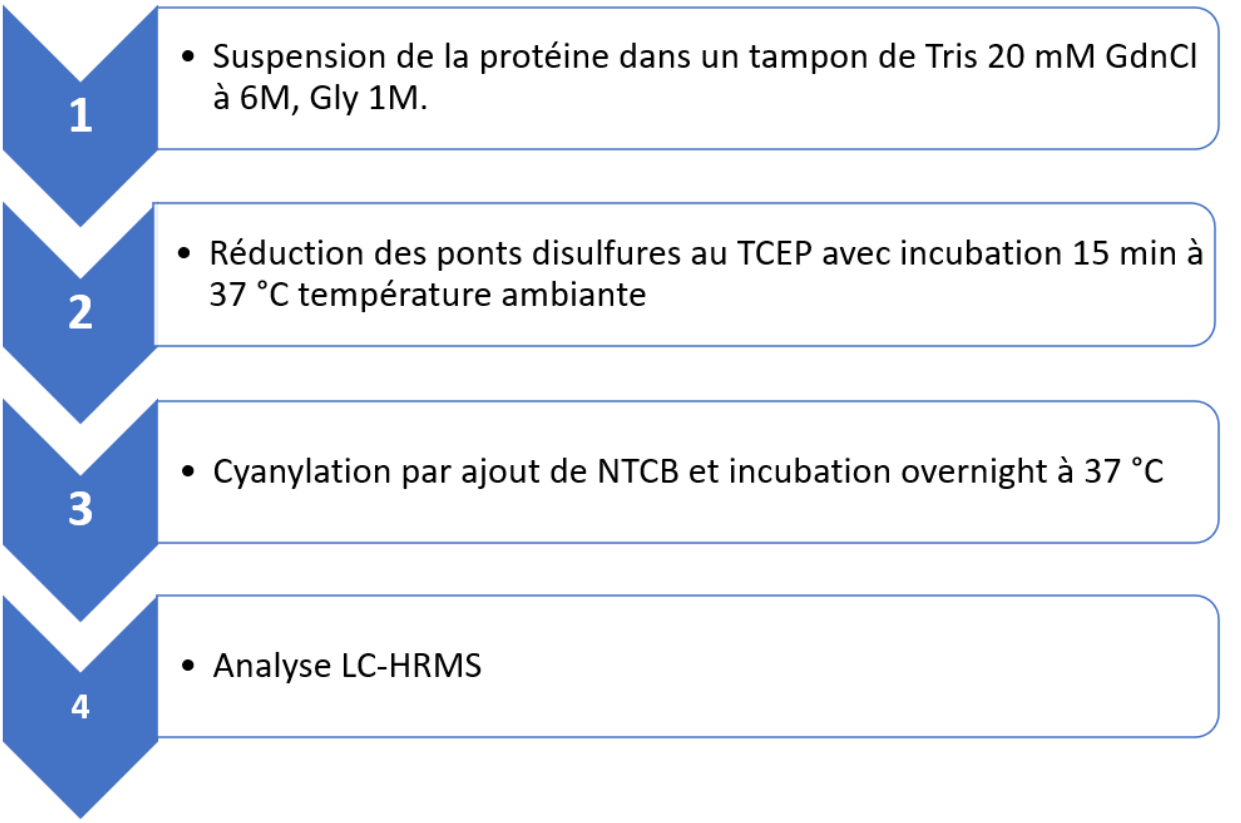


Figure 73: Schéma des étapes du protocole NTCB de Tang et. al

Ce protocole propose un clivage basique encore plus doux que la soude et l'ammoniaque, réduisant les possibilités de réactions secondaires. Comme le montre la Figure , la réaction a bien marché, avec des masses de peptides dont l'erreur est inférieure à 1 ppm, ce qui est très exacte. De plus la modification issue de cyanylation est la plus importante, ce résultat est satisfaisant. Il est intéressant de noter que sur un même peptide, les modifications peuvent être différentes, comme le peptide 582-607, qui porte une modification ITZ sur sa cystéine 582 mais aussi une modification Dha dont les structures sont présentées Figure . Des optimisations seront donc à faire pour pousser l'équilibre vers la modification ITZ. Néanmoins, cette technique a permis de détecter 11 peptides parmi les 37 attendus.

Meas. m/z	Calc. m/z	δ (Da)	δ (ppm)	Rel. Int. (%)	z	Annotation	Formula
501.854746	501.854668	0.000078	0.2	1.31	5	[125-146] e.CFLSHKDDSPDLPKLKPDPNTL.c [1xCyano]	C111H173N29O35S

Meas. m/z	Calc. m/z	δ (Da)	δ (ppm)	Rel. Int. (%)	z	Annotation	Formula
506.241989	506.241536	0.000453	0.9	1.25	3	[86-98] g.CEKSLHTLFGDEL.c [1xCyano]	C66H101N17O22S
562.029908	562.029402	0.000506	0.9	1.79	4	[58-76] q.CPFDEHVKLVNELTEFAKT.c [1xCyano]	C101H153N25O31S
600.301176	600.301121	0.000055	0.1	3.02	2	[500-509] c.CTESLVNRRP.c [1xCyano]	C48H82N18O16S
601.768527	601.768265	0.000261	0.4	1.27	2	[461-470] c.CTKPESERMP.c [1xCyano]	C48H79N15O17S2
620.327922	620.329233	-0.001311	-2.1	11.20	3	[590-607] a.CFAVEGPKLVVSTQTALA. [1xCyano]	C83H135N21O25S
627.066670	627.066516	0.000153	0.2	3.75	4	[125-146] e.CFLSHKDDSPDLPKLKPDPNTL.c [1xCyano]	C111H173N29O35S
637.718297	637.723494	-0.005197	-8.1	0.97	8	[537-580] i.CTLPDTEKQIKKQTALVELLKHKPKATEEQLKTVMENFVAFVDK.c [1xCyano]	C228H377N59O68S2
637.718297	637.718189	0.000109	0.2	0.97	8	[415-459] n.CDQFEKLGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTR.c [1xCyano]	C226H365N65O67S
728.676371	728.682953	-0.006583	-9.0	3.84	7	[537-580] i.CTLPDTEKQIKKQTALVELLKHKPKATEEQLKTVMENFVAFVDK.c [1xCyano]	C228H377N59O68S2
728.676371	728.676890	-0.000519	-0.7	3.84	7	[415-459] n.CDQFEKLGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTR.c [1xCyano]	C226H365N65O67S
728.676529	728.682953	-0.006424	-8.8	13.32	7	[537-580] i.CTLPDTEKQIKKQTALVELLKHKPKATEEQLKTVMENFVAFVDK.c [1xCyano]	C228H377N59O68S2
728.676529	728.676890	-0.000361	-0.5	13.32	7	[415-459] n.CDQFEKLGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTR.c [1xCyano]	C226H365N65O67S
849.954728	849.962233	-0.007505	-8.8	3.70	6	[537-580] i.CTLPDTEKQIKKQTALVELLKHKPKATEEQLKTVMENFVAFVDK.c [1xCyano]	C228H377N59O68S2
849.954728	849.955159	-0.000431	-0.5	3.70	6	[415-459] n.CDQFEKLGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTR.c [1xCyano]	C226H365N65O67S
876.770117	876.770649	-0.000531	-0.6	6.58	3	[582-607] c.CAADDKEACFAVEGPKLVVSTQTALA. [1xCyano; 1xCys->Dha]	C114H182N30O39S
896.431747	896.431639	0.000108	0.1	21.91	3	[582-607] c.CAADDKEACFAVEGPKLVVSTQTALA. [2xCyano]	C115H183N31O39S2
921.764673	921.764400	0.000272	0.3	8.27	3	[581-607] k.CCAADDKEACFAVEGPKLVVSTQTALA. [1xCyano; 2xCystine]	C117H187N31O40S3
929.989756	929.990211	-0.000455	-0.5	85.52	2	[590-607] a.CFAVEGPKLVVSTQTALA. [1xCyano]	C83H135N21O25S

Generated by mMass • Open Source Mass Spectrometry Tool • www.mmass.org

Figure 74: Résultats de l'hydrolyse de la BSA au NTCB en one pot avec clivage à la Glycine

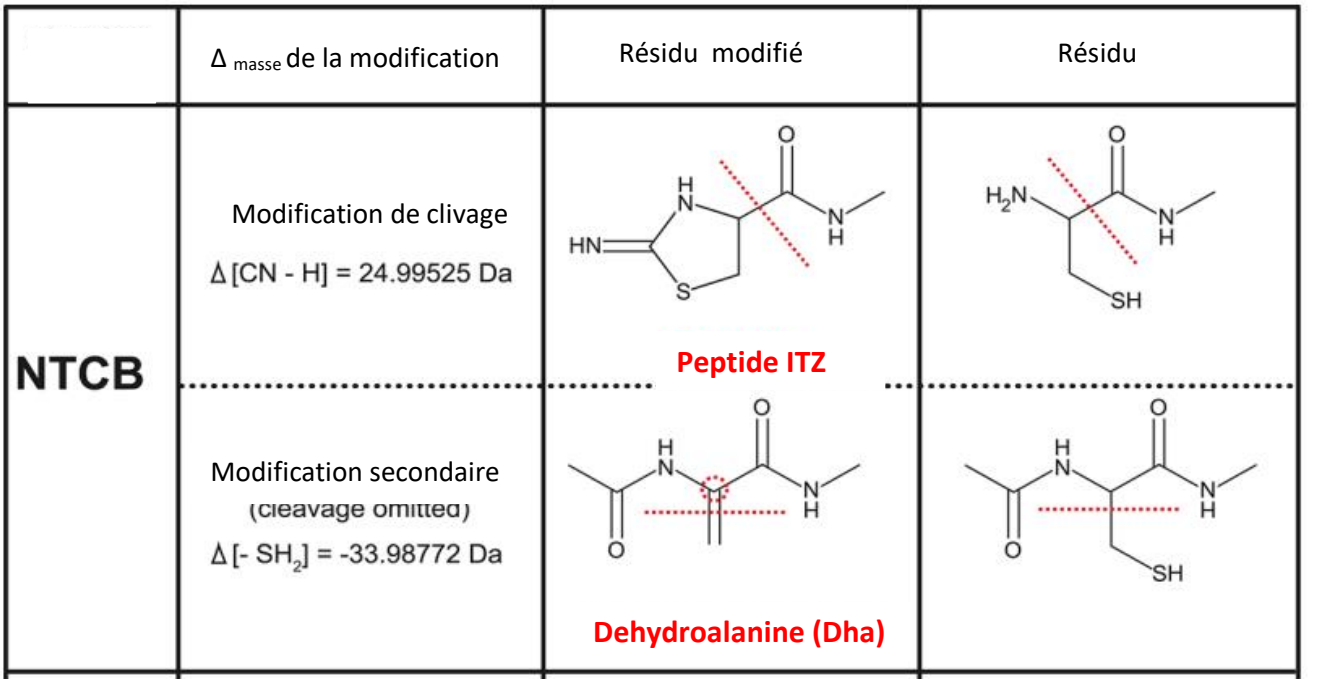


Figure 75: Structure chimique des modifications ITZ et Dha du NTCB

En définitive, la meilleure façon de procéder à l'hydrolyse de la BSA avec le NTCB fût l'hydrolyse one pot avec un clivage à la glycine.

Conclusion

La digestion de la BSA par le NTCB a été un succès grâce à l'hydrolyse one pot. Les prochaines étapes seront de digérer les glycosidases afin d'obtenir de plus grands peptides que la BSA, qui pourront être séparés par la taille par divers procédés de chromatographie liquide. Dans ce cadre, une première approche de la séparation des protéines en fonction de leur taille a été réalisée : la SEC (Chromatographie d'exclusion stérique).

III. Optimisation de la séparation des peptides et protéines par chromatographie SEC

III.1. Séparation des protéines entières

Stratégie expérimentale

Cette partie est consacrée à la séparation des protéines entières, sous leur forme native dans un premier temps. Comme le montre le Tableau 20, il s'avère qu'il y a une grande différence de taille entre les glycosidases natives humaines comme murines, et l'albumine humaine ou murine. Cette différence de taille a été exploitée pour séparer l'albumine majoritaire des glycosidases. La méthode de choix de séparation des protéines selon la taille a été utilisée, la SEC. Dans cette étude, la détection ne se fait plus par spectrométrie de masse, mais par détection UV à l'aide d'un détecteur à barrette de diode (PDA). Les longueurs d'ondes fixées étaient 220 et 280 nm car elles correspondent aux longueurs d'ondes spécifiques des protéines. À 220 nm, c'est la liaison peptidique qui absorbe et à 280 nm, ce sont les cycles aromatiques des résidus.

	Protéine humaine		Protéine murine	
	Masse de la forme native (Da)	Masse de la forme dénaturée (Da)	Masse de la forme native (Da)	Masse de la forme dénaturée (Da)
β-Glu	298 928	74 732	296 780	74 195
β -GlcNac	102 915	102 915	103 102	103 162
β-Gal	152 150	76 075	146 242	73 121
α-Fuc	214 720	53 680	211 240	52 810
Albumine	63 367	69 367	66 093	66 093

Tableau 20: Masses des protéines humaines et murines sous leurs formes natives et dénaturées

Les données présentées ont été obtenues dans la base de données Uniprot. La masse de la forme dénaturée des protéines correspond à la masse du monomère, la protéine étant un assemblage de monomères. Cependant, certaines protéines sont uniquement sous forme monomérique, dénaturées ou non.

- **Séparation de l'albumine et des protéines commerciales par SEC analysée par la méthode 1**

Avec l'élution de la BSA et HSA sur colonne, il devrait y avoir des temps de rétention plus grand que les glycosidases, qui sont de masse plus importante. Ceci s'explique par le fait qu'en SEC, le trajet des petites protéines à travers les pores des billes est plus grand que les grosses protéines qui ont une trajectoire linéaire sans passer par les pores, car leurs tailles ne leur permettent pas
Avec une première colonne SEC Agilent Bio-SEC 3, d'une gamme de masse allant de 5 000 à 1 250 000 Da, les glycosidases ainsi que la HSA et la BSA ont été analysées individuellement. L'objectif était de connaître leurs temps de rétention, pour pouvoir ensuite les analyser en mélange.

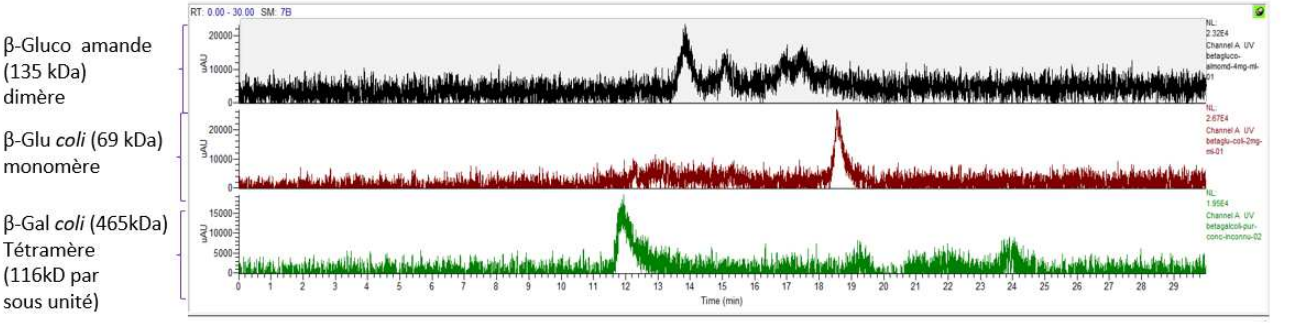


Figure 74: Chromatogrammes de la séparation des glycosidases par SEC-UV à 280 nm en conditions natives

Il est observable sur Figure 74, que les temps de rétention des glycosidases sont compris entre 11 et 18 min. Chacune d'entre elles a un temps de rétention bien spécifique, mais il fallait s'assurer qu'elles soient efficacement séparées de la BSA et HSA. Cependant, on observe Figure 75 que l'albumine sort à 14 min, temps de rétention compris dans la zone d'élution des glycosidases.

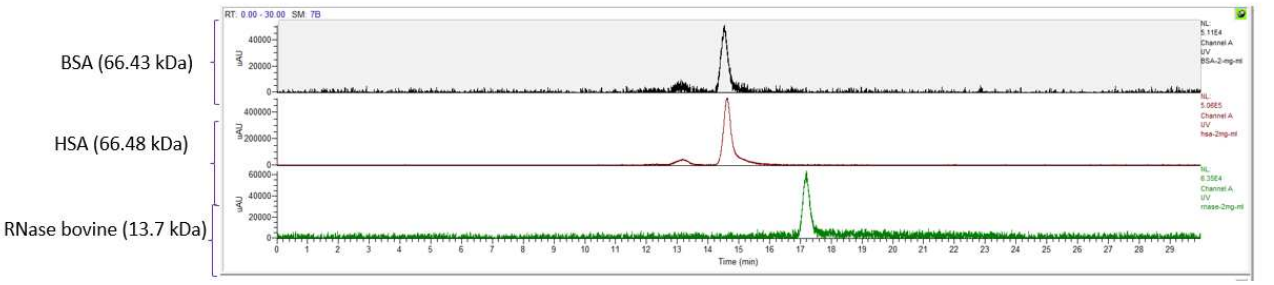


Figure 75: Séparation des petites protéines sur SEC-UV à 280 nm en conditions natives

En conclusion , la séparation sur colonne Agilent Bio-SEC en condition native de l'albumine et des glycosidases n'est pas optimale. Les temps de rétention sont trop rapprochés, signifiant que le domaine de perméation sélective de la colonne n'est pas adapté à la séparation de ces protéines. En effet, la gamme de masse précedement citées de la colonne est trop large, laissant passer trop de protéines et ne retenant l'albumine pas assez longtemps. Pour pallier à ce problème, nous avons décicidé de dénaturer et réduire les glycosidases qui sont multimériques en monomères avec le SDS (Sodécylsulfate de Sodium) et le DTT .

- **Séparation de l'albumine et des protéines commerciales par SEC en condition réductrice analysée par la méthode 1**

L'élution des protéines sur colonne SEC après traitement au SDS, DTT puis IAA été réalisée. Les masses des monomères des protéines attendues à l'issu de ce traitement sont renseignées Tableau 21 .

Protéines commerciales	Masse de la protéine native (Da)	Masse d'un monomère (Da)
β-Glu *coli*	290 000	74 000
β-Gal *coli*	465 000	135 000
β-Gluco amande	135000	75 000

Tableau 21: Masses monomériques et multimériques des glycosidases commerciales

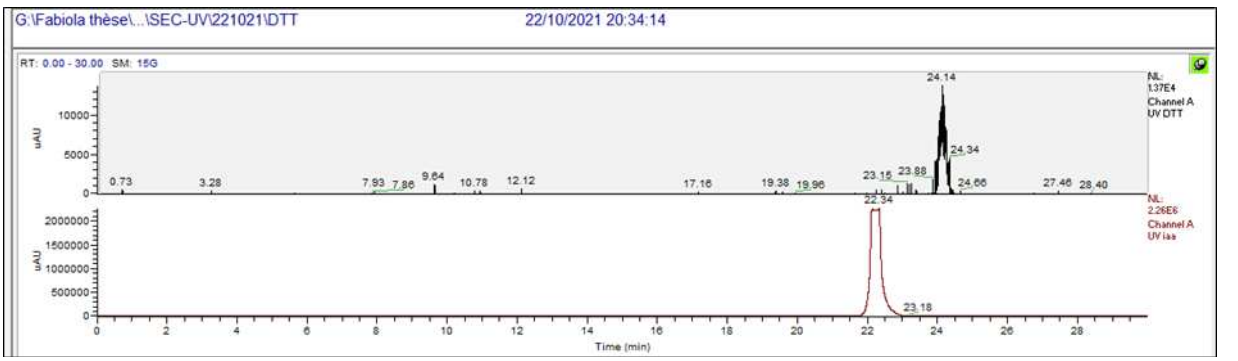


Figure 76: Élution du DTT et IAA sur SEC-UV à 280 nm

Avant de séparer les protéines sur colonne, les temps de rétention des composés de la réaction en excès, l'IAA et le DTT, ont été établis. Selon la Figure 76, ils sont de 22.34 et 24.14 min respectivement, c'est-à-dire hors de la zone de perméation sélective. Il est compréhensible qu'ils sortent très tard dans la colonne, mais pas qu'ils aient des temps de rétention aussi éloignés.
Du fait de leur petite taille, et du maillage de la colonne, ils sont censés sortir en même temps. La phase stationnaire de cette colonne n'est donc pas uniquement un tamis moléculaire, mais une phase hybride avec des groupements qui interagissent avec les composés IAA et DTT.

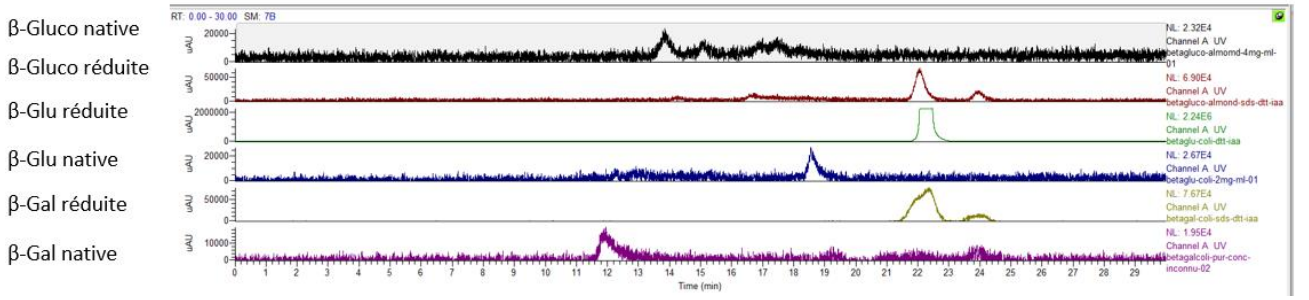


Figure 77: Chromatogrammes des glycosidases en conditions dénaturantes par SEC-UV à 280 nm

Nous pouvons observer Figure 77 que l'analyse des glycosidases en conditions dénaturantes n'a pas fonctionné. On observe bien les pics de DTT à environ 22 et 24 min, mais pas les monomères de protéines. La première explication est que les échantillons sont beaucoup trop dilués et qu'il aurait fallu une première étape de purification avant d'injecter, ou de réduire les volumes réactionnels. La seconde explication est qu'une fois que les protéines sont dénaturées, leur volume hydrodynamique change. Leur nouvelle forme sans repli ne peut pas interagir de la même façon et avoir un passage facilité dans la colonne. On peut supposer que les protéines se sont repliées pour former des complexes et sont restées bloquées en tête de colonne.
En conclusion ni les conditions natives, ni les conditions dénaturantes n'ont permis de séparer les glycosidases de l'albumine sur colonne SEC. Pour optimiser la séparation, nous pouvions soit changer de phase mobile, soit changer la colonne. Cette dernière option a donc été choisie. Cette méthode d'élution est ci-après notée méthode 2.

- **Gamme de détection de la BSA par SEC analysée par la méthode 2**

La partie suivante traite de la séparation des protéines selon la méthode 2, qui a pour phase mobile 10 mM d'acétate d'ammonium, et une autre colonne SEC. Étant donné que la séparation avec la colonne précédente n'a pas fonctionné, une autre colonne a été utilisée, la Reprosil 200 avec une gamme de masse plus restreinte allant de 10 000 – 500 00 Da. Suite à ce changement de colonne, son efficacité de séparation qualité a été évaluée en se basant uniquement sur l'analyse de la BSA native dans un premier temps. Avant de procéder à la séparation des protéines sur cette colonne, la gamme de détection a été effectuée. Les profils chromatographiques de la gamme à 280 nm et à 220 nm sont présentés Figure 78 et Figure 80. Cette gamme de détection a été réalisée avec la BSA native. La limite de détection de la BSA avec cette colonne est entre 2 et 4 $mg.mL^{-1}$ en détection UV à 280 nm.

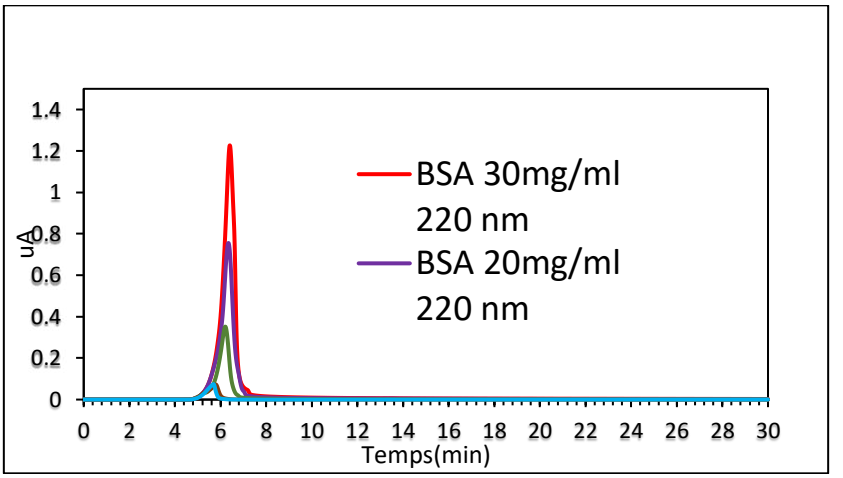


Figure 78: Chromatogrammes de la gamme détection de la BSA à 280 nm

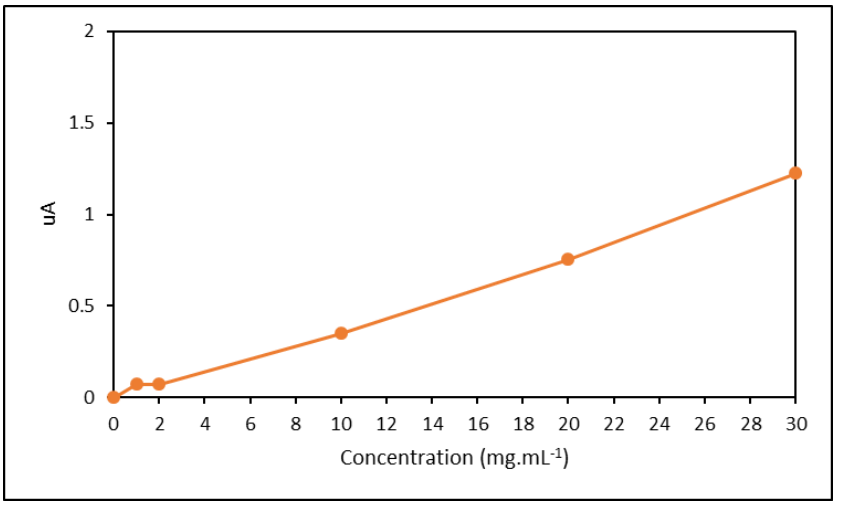


Figure 79: Gamme de détection de la BSA à 280 nm

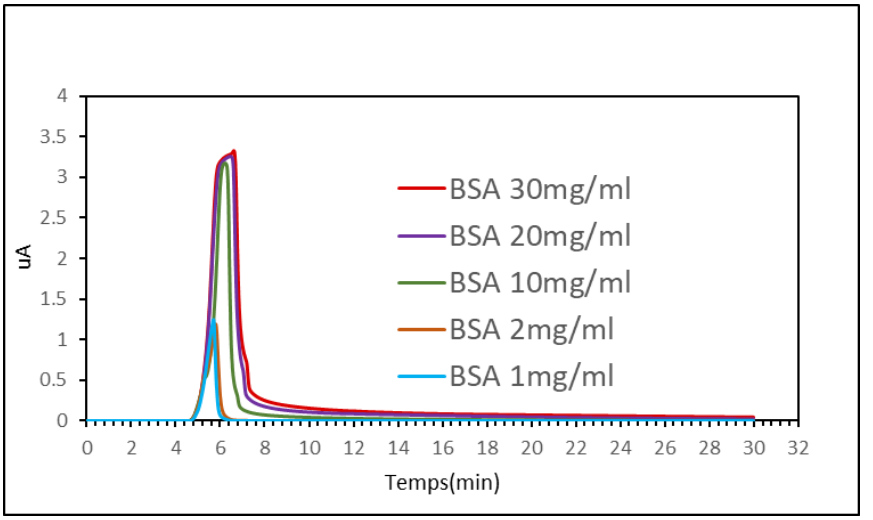


Figure 80: Chromatogrammes de la gamme de détection de la BSA à 220 nm

À 220 nm, la courbe n'est pas linéaire, et le détecteur UV est saturé à partir de 10 $mg.mL^{-1}$ selon les Figure 80 et Figure 81.

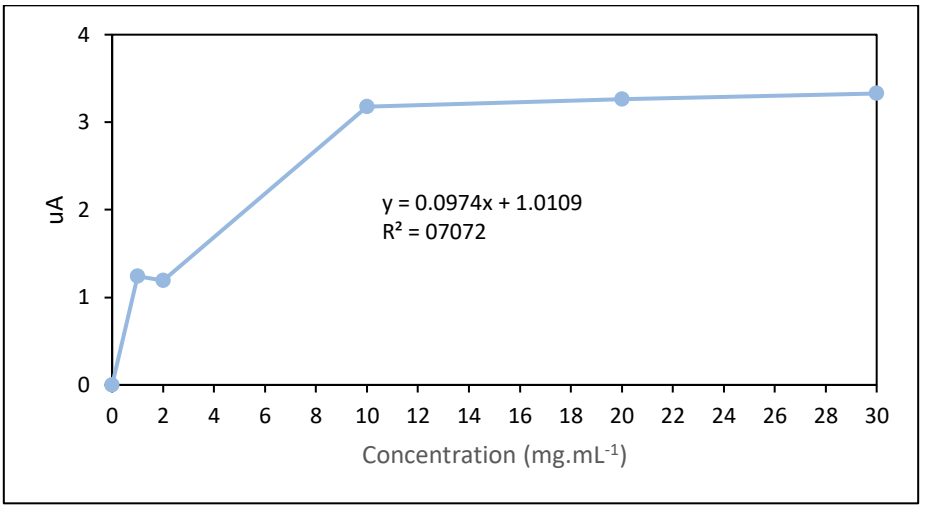


Figure 81: Gamme d'étalonnage de la BSA à 220 nm

- **Séparation par SEC des protéines entières en conditions natives analysée par la méthode 2**

Après avoir étudié les limites de détection de la BSA avec cette colonne, les protéines (BSA et glycosidases) ont été analysées par SEC-UV individuellement tout comme avec la première colonne. Malheureusement, les protéines analysées sortent au même temps de rétention. Il n'y a aucune séparation (Figure 82 et Figure 83).

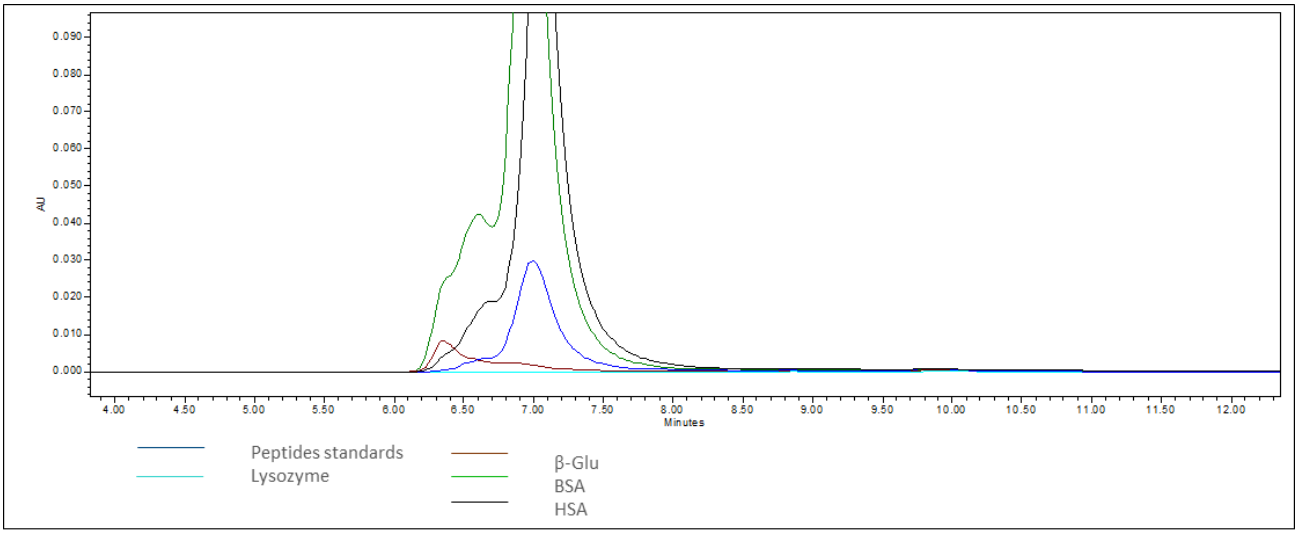


Figure 82: Zoom du chromatogramme des protéines en conditions natives par SEC-UV avec la Reprosil 200

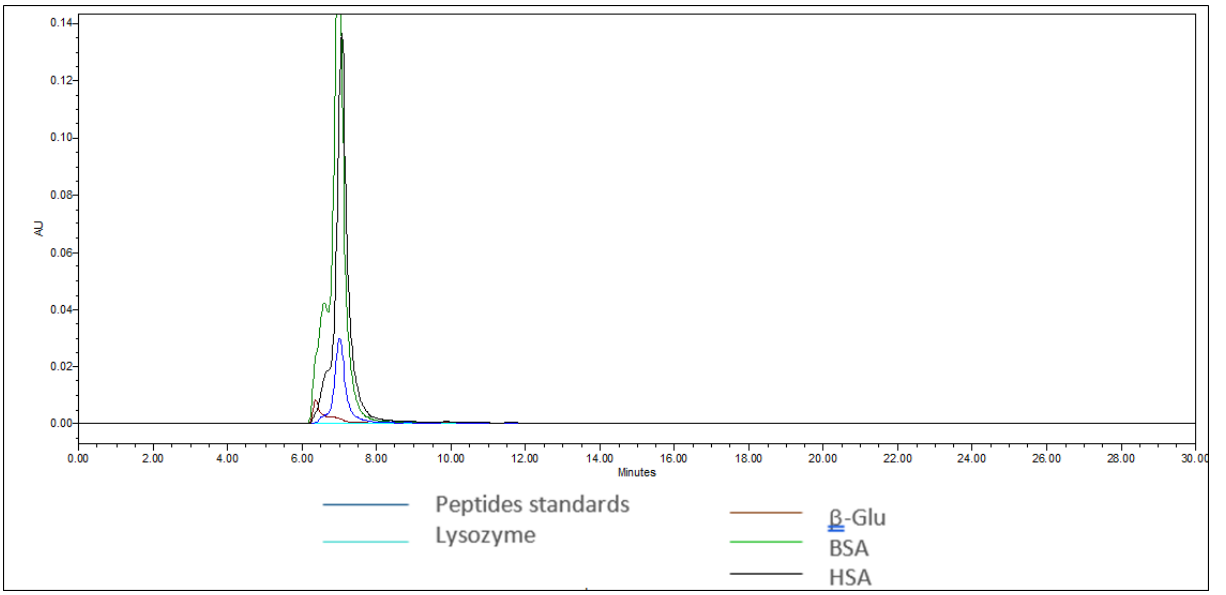


Figure 83:Chromatogramme des protéines en conditions natives par SEC-UV avec la Reprosil 200

III.2. Séparation par SEC de l'albumine et des protéines commerciales par SEC en conditions dénaturantes analysée par la méthode 2

Dénaturation des protéines

Face à la difficulté de séparation, la même idée a été formulée de séparer les protéines en conditions dénaturantes, afin de réduire les protéines multimériques en monomères. Nous observons

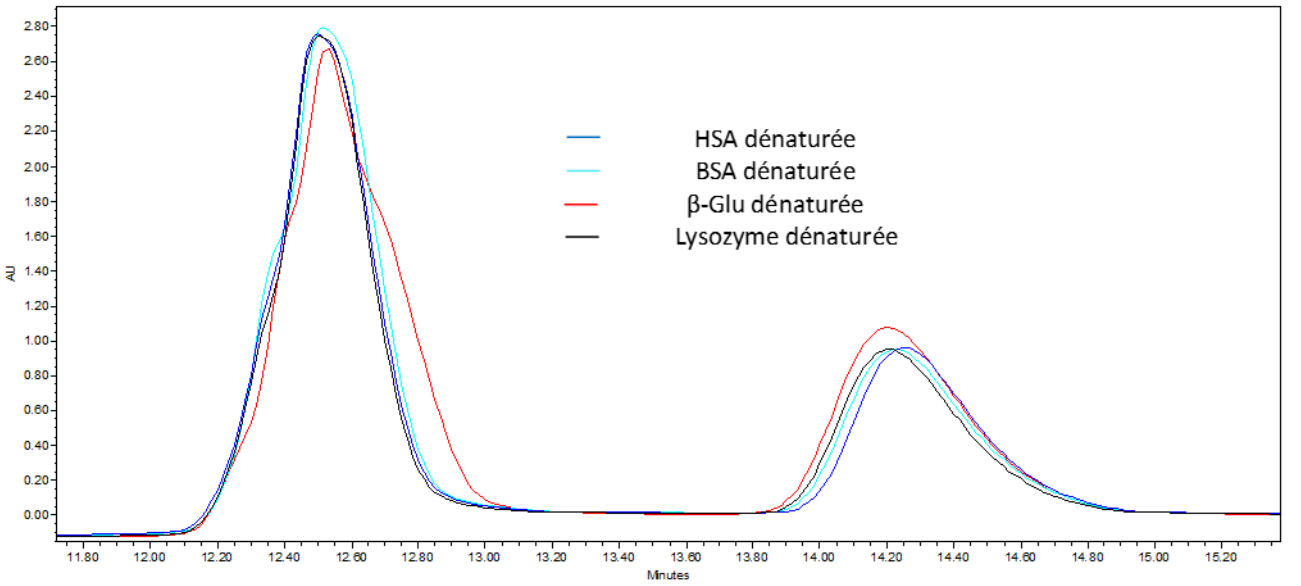


Figure ***84*** que la séparation n'est pas optimale, au contraire il y a des pics communs à toutes les protéines.

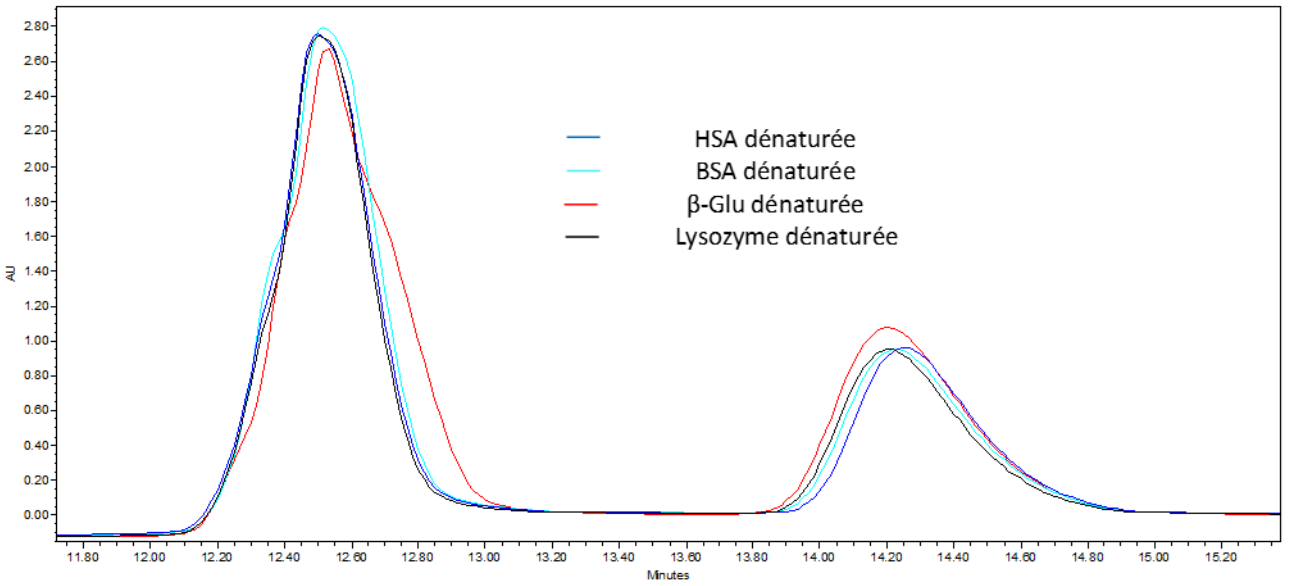


Figure 84: Chromatogrammes des protéines en conditions dénaturantes par SEC-UV méthode 2

III.3. Analyse des protéines hydrolysées par NTCB

Séparation des digestats NTCB par SEC analysée par la méthode 2

É0tant donné que les monomères d'enzyme ont été difficiles à isoler, les protéines ont été réduites en peptide via l'hydrolyse au NTCB. Ainsi, en obtenant des peptides de plus petites tailles, il était attendu d'avoir une bonne séparation comme on peut en obtenir en RP-LC (Reverse Phase Chromatography), Figure 85. Ceci dit, fort est de constater que la séparation des peptides sur la colonne SEC Reprosil est mauvaise comme on le voit sur la Figure 86. Les peptides de la BSA, HSA, et B-Glu sortent dans le même pic.

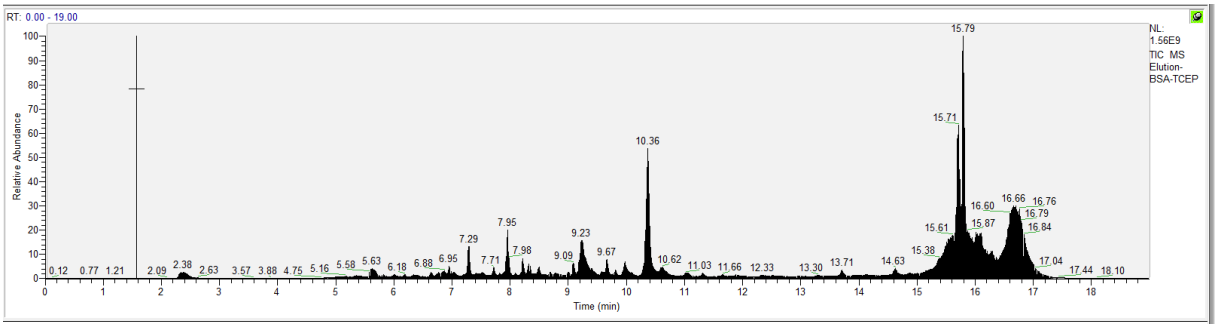


Figure 85: Chromatogramme d'analyse des peptides issu de l'hydrolyse de la BSA au NTCB par LC-Orbitrap

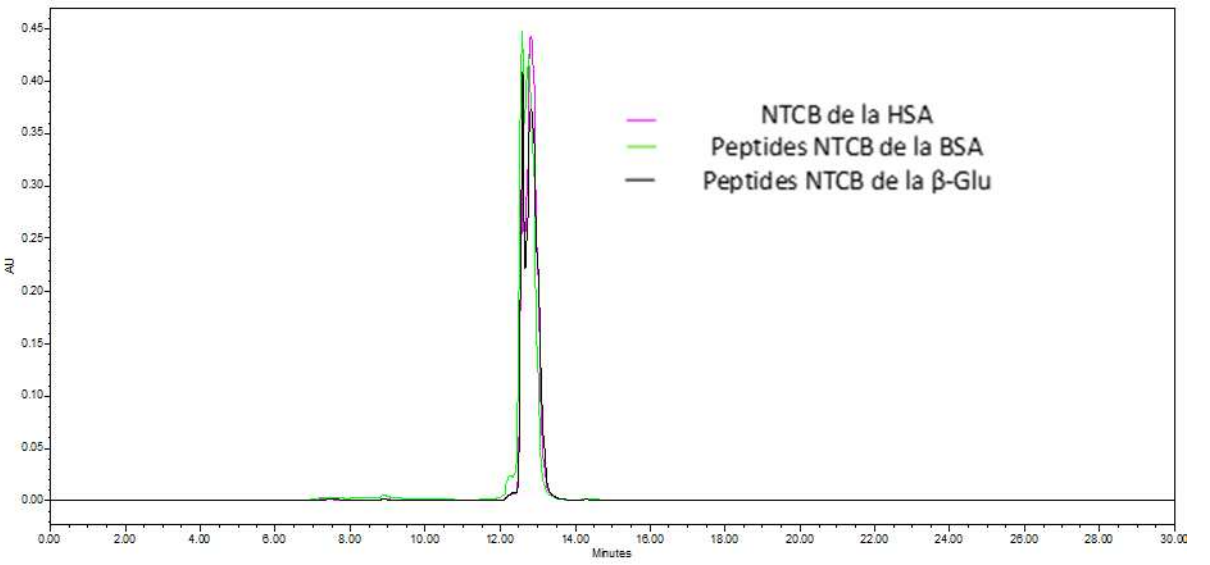


Figure 86: Chromatogramme des peptides obtenus via le NCTB par SEC-UV méthode 2

Ceci est étonnant car la B-Glu a des plus gros peptides que la HSA et BSA, ayant beaucoup moins de cystéines, et donc moins de fréquence de coupure avec le NTCB. En voyant les zooms du chromatogramme sur la Figure 87 et Figure 88, on voit tout de même qu'il y a différentes populations de peptides, mais le profil de séparation est le même pour les 3 protéines.

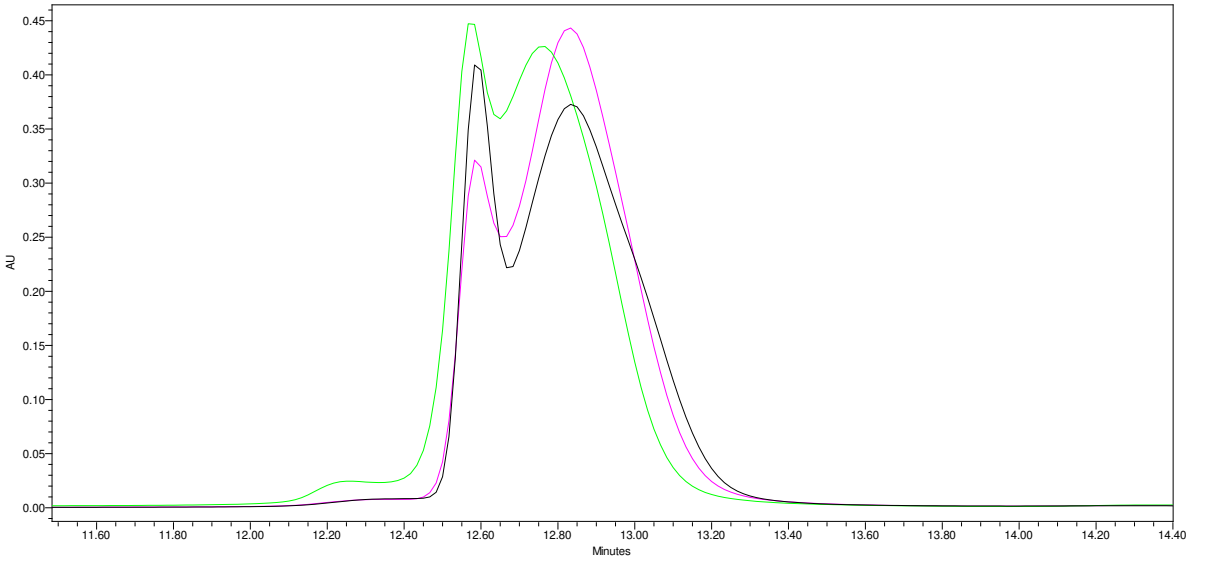


Figure 87: Zoom 1 du chromatogramme des peptides obtenus via le NCTB par SEC-UV méthode 2

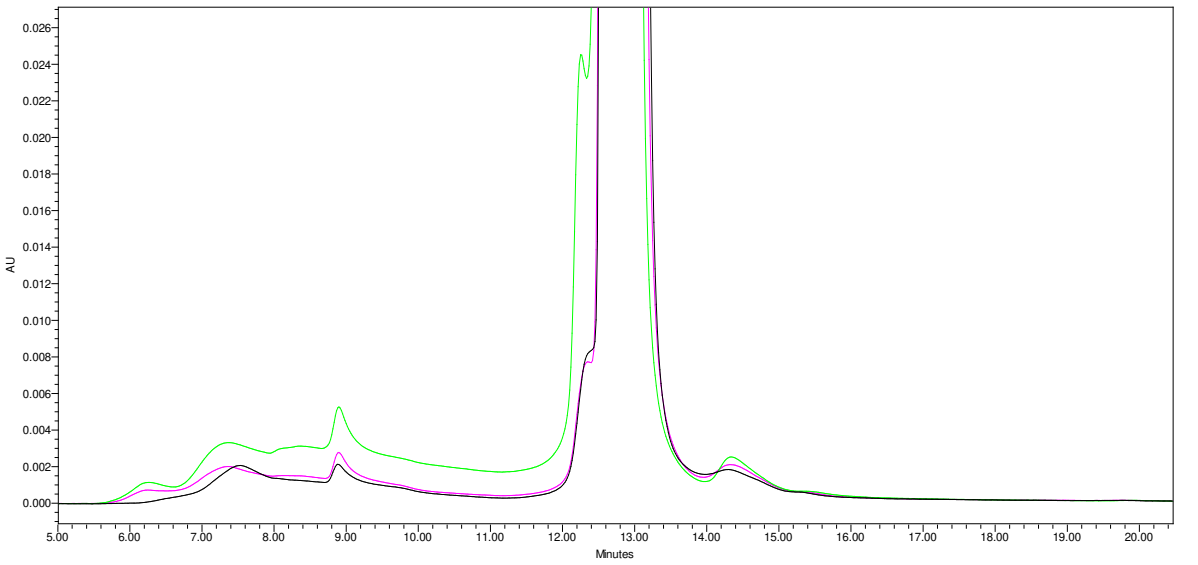


Figure 88: Zoom 2 du chromatogramme des peptides obtenus via SEC-UV méthode 2

Conclusion

En définitive la séparation des glycosidases et de l'albumine est encore en phase de développement. Les deux colonnes utilisées séparément n'ont pas permis de séparer les protéines. En revanche, il se peut qu'utiliser les deux colonnes en série allonge les temps de rétention et donc améliore la séparation. De plus, choisir d'autres phases mobiles est envisageable. La SEC avait pour de séparer les protéines après extraction et purification des protéines du plasma étant donné les limites de détection obtenues.

Etude au niveau moléculaire du fonctionnement d'un agent anticancéreux

La chimiothérapie conventionnelle ne permet pas de guérir la plupart des cancers les plus meurtriers. Les médicaments utilisés cliniquement présentent en effet peu de sélectivité vis-à-vis des cellules tumorales et s'attaquent également aux tissus sains. Cette destruction non-sélective entraîne de sévères effets secondaires et conduit souvent à l'arrêt prématuré du traitement.

Ainsi, la conception de stratégies permettant d'augmenter la sélectivité des agents anticancéreux représente un intérêt majeur dans la lutte contre le cancer. Dans ce cadre, le développement de nouveaux systèmes moléculaires auto-réactifs ciblant les spécificités structurelles ou métaboliques du microenvironnement tumoral offre de nombreux avantages.

Dans le cadre de leurs travaux visant à améliorer l'efficacité des chimiothérapies anticancéreuses, l'équipe OrgaSynth de l'IC2MP a conçu de nouveaux systèmes moléculaires programmés pour le ciblage et la libération sélective de molécules bioactives dans un tissu d'intérêt. Ils ont par exemple étudié une nouvelle approche basée sur l'utilisation de l'albumine plasmatique comme vecteur thérapeutique.

Suivant ce concept, ils ont conçu une plateforme moléculaire non toxique capable de se lier de manière covalente avec l'albumine directement dans le flux sanguin et possédant la capacité d'interagir avec différentes classes d'agents anticancéreux via des interactions supramoléculaires. Leur plateforme comporte un motif maléimide, capable de réagir avec le thiol de l'albumine plasmatique, et trois β-cyclodextrines connues pour leurs propriétés d'encapsulation de principes actifs hydrophobes tel que la doxorubicine. La présence des trois cyclodextrines sur l'albumine pourra alors être exploitée pour le transport et la vectorisation de doxorobucine. Cette stratégie thérapeutique présente l'avantage d'utiliser un adjuvant, non toxique, dont le développement clinique est bien moins contraignant que celui d'un composé cytotoxique. En outre, cette stratégie permet d'augmenter la sélectivité de principes actifs déjà utilisés chez l'Homme (tels que la doxorobucine) sans en modifier la structure.

L'étude présentée ci-après présente les étapes de synthèse de cette plateforme, ainsi que les essais in vitro et in vivo démontrant le fonctionnement et l'efficacité de ce nouvel agent anticancéreux. Ce travail, intitulé « A β-Cyclodextrin-Albumin Conjugate for Enhancing Therapeutic Efficacy of Cytotoxic Drugs » a été publié en Mai 2022 dans Bioconjugate Chemistry.
Mon travail a ici consisté à développer le protocole d'extraction et d'analyse par UPLC-HRMS de la doxorobucine. Du fait de son mécanisme d'action, l'extraction de la doxorubicine est un défi. En effet, il s'agit d'un agent intercalant qui s'insère entre les bases de l'ADN et bloque ainsi la réplication et la transcription. De plus, elle induit une inhibition de la Topoisomérase II. Les deux mécanismes combinés, conduisent à des ruptures dans l'ADN et à la destruction de la cellule.
Suivant les protocoles détaillés dans la littérature, nous avons testé différentes conditions chimiques et mécaniques, d'extraction de la doxorobucine à partir des tissus cancéreux.

Suite Au développement de ce protocole, nous avons démontré que la plateforme moléculaire développée par l'équipe OrgaSynth permettait d'augmenter la « charge » en doxorobucine déposée au sein de la tumeur ; doxorobucine ici déposée via la dissociation équilibrée du complexe d'inclusion. Enfin, ces résultats ont été confrontés aux données de toxicité recueillies après injection de l'agent à des modèles murins greffés avec une tumeur de type LLC.

Bioconjugate Chemistry

pubs.acs.org/bc

Communication

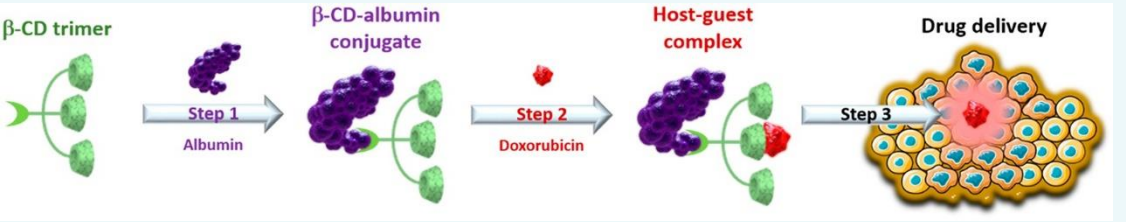
A β-Cyclodextrin-Albumin Conjugate for Enhancing Therapeutic Efficacy of Cytotoxic Drugs

Chad Plumet,[#] Rémi Chatre,[#] Fabiola Djago, Elodie Péraudeau, Quentin Blancart-Remaury, Jonathan Clarhaut, Claude Geffroy, Achmet Said Mohamed, Isabelle Opalinski, Brigitte Renoux, Pauline Poinot, and Sébastien Papot*

Cite This: Bioconjugate Chem. 2022, 33, 1138−1144

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Enhancing the selectivity of anticancer drugs currently used in the clinic is of great interest in order to propose more efficient chemotherapies with fewer side effects for patients. In this context, we developed a β-cyclodextrin trimer that binds to circulating albumin to form the corresponding bioconjugate in the bloodstream. This latter can then entrap doxorubicin following its i.v. administration via the formation of a host−guest inclusion complex and deliver the drug in tumors. In this study, we demonstrate that the β-cyclodextrin trimer improves the therapeutic efficacy of doxorubicin for the treatment of a subcutaneous murine Lewis lung carcinoma (LLC) implanted in C57BL/6 mice. This outcome is associated with an increased deposition of doxorubicin in malignant tissues when used in combination with the β-cyclodextrin trimer compared to the administration of the drug alone.

The selective delivery of potent anticancer drugs in tumors remains a major challenge in cancer chemotherapy. Advances in this field led recently to the approval of ten antibody−drug conjugates (ADCs) for various applications in oncology.[1−3] In parallel to antibodies, albumin was also intensively investigated as a tumor targeting unit for enhancing the selectivity of cancer chemotherapy.[4−12] Indeed, albumin accumulates within malignant tissues through the enhanced permeability and retention (EPR) effect as well as by receptor-mediated active transport. This property conducted to the development of the Abraxane, an albumin paclitaxel nanoparticle that is currently used in the clinic for the therapy of breast, lung, and pancreatic cancer.[13−15] In another approach, endogenous albumin was exploited to transport anticancer drugs toward tumor sites by means of prodrugs possessing the ability to bind in situ to the thiol at the cysteine-34 position of the circulating protein through Michael addition. This strategy was pioneered by Kratz and co-workers[16−20] who developed Aldoxorubicin, an albumin-binding prodrug of doxorubicin that is currently being evaluated in humans for therapy of soft tissue sarcoma.[21] Taking advantage of both the accumulation of albumin in malignant tissues and the presence of a high concentration of β-glucuronidase in the microenvironment of solid tumors,[22−28] we and others designed enzyme-responsive albumin-binding prodrugs that demonstrated significant therapeutic efficacy in numerous animal models.[29−37] Recently, Zelikin and co-workers developed glucuronide prodrugs that associated in situ to endogenous albumin through supramolecular interactions, thereby leading to superior antitumor activity than that of the free drug for the treatment of triple negative breast xenograft implanted in mice.[38]

Herein, we present a novel approach that exploit the tumor targeting properties of albumin for enhancing the therapeutic efficacy of anticancer drugs (Figure 1). Within this framework, we designed the β-cyclodextrin (β-CD) trimer 1 (Scheme 1) that can bind covalently to the free thiol of circulating albumin once in the bloodstream, hence leading to the corresponding bioconjugate 2 (step 1). We hypothesized that the subsequent administration of an anticancer drug should result in the formation of the host−guest inclusion complex 3[39,40] (step 2). The bioconjugate 3 will then accumulate in the tumor microenvironment (step 3) to deliver the anticancer drug via the dissociation of the host−guest interactions, therefore enabling the drug to penetrate within cancer cells to exert its cytotoxic activity (step 4).

Received: April 13, 2022
Revised: May 21, 2022
Published: May 25, 2022



ACS Publications

© 2022 American Chemical Society

https://doi.org/10.1021/acs.bioconjchem.2c00182
Bioconjugate Chem. 2022, 33, 1138−1144

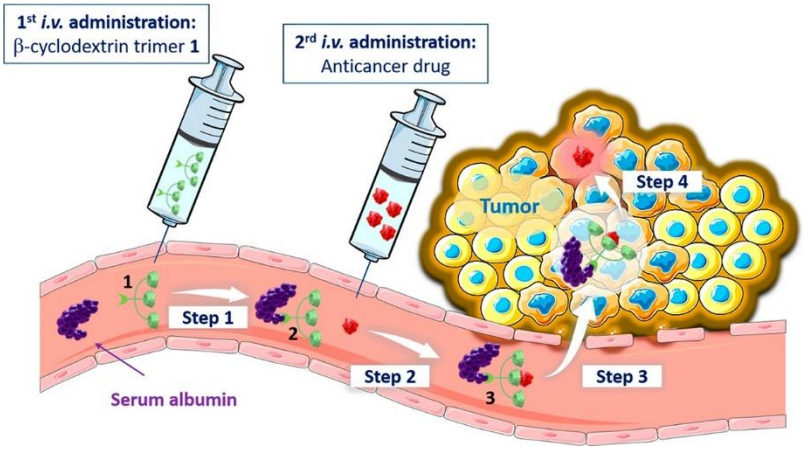
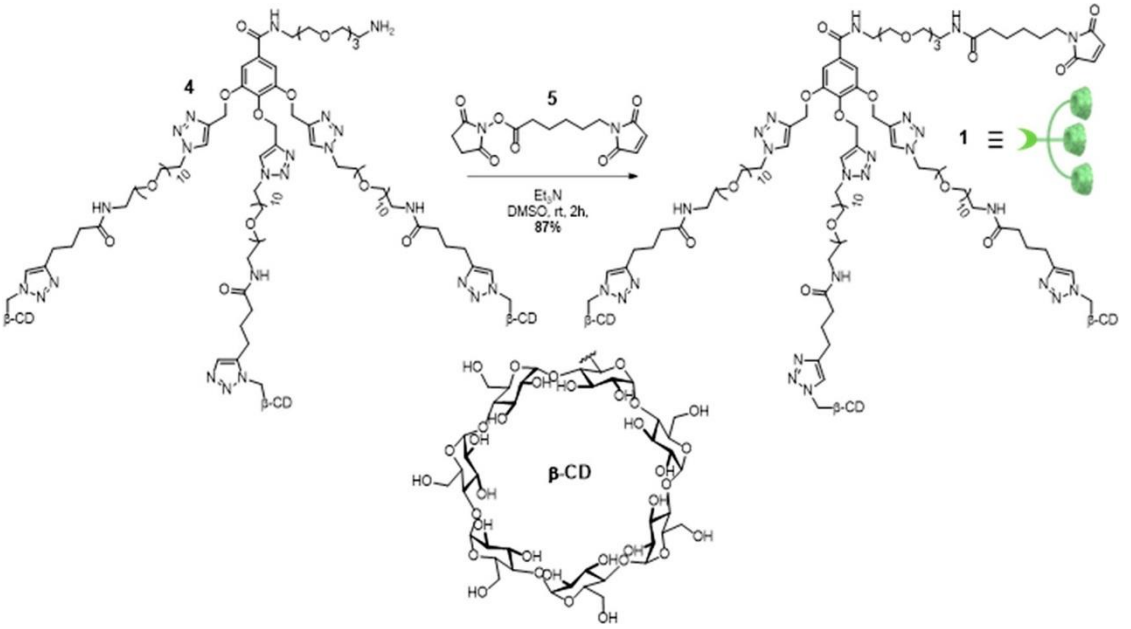


Figure 1. Principle of tumor targeting based on the use of the β-cyclodextrin trimer 1. In the blood, the β-cyclodextrin trimer 1 binds covalently to circulating albumin to generate the bioconjugate 2 (step 1). The i.v. administration of the anticancer drug leads to the formation of the supramolecular complex 3 (step 2). The complex 3 accumulates selectively in the tumor microenvironment (step 3). The dissociation of 3 conducts to the release of the drug that penetrates within tumor cells for inducing the anticancer activity.

Scheme 1. Structure and Synthesis of the β-CD Trimer 1



In this study, we demonstrated the validity of this strategy with doxorubicin (Dox). Indeed, the combination of the β-CD trimer 1 and the anthracycline led to superior therapeutic efficacy in vivo compared to the use of Dox alone. This enhanced antitumor activity was correlated with the increased deposition of Dox at the tumor site, therefore showing the benefit brought by the β-CD derivative 1 to improve the therapeutic index of this clinically used cytotoxic agent.

RESULTS AND DISCUSSION

Our first aim was to prepare the bifunctional trimer 1 that comprises both β-CDs for enabling the encapsulation of doxorubicin and a maleimide allowing its binding to the free thiol of circulating albumin through Michael addition (Scheme 1).

The ability of β-CDs to entrap Dox by the formation of host−guest inclusion complexes was already exploited for the design of several drug delivery systems.[41−44] However, in our approach the formation of such an inclusion complex has to occur in vivo, which remains highly challenging. Under these circumstances, we designed the compound 1 that permits the attachment of three β-CD units per albumin molecule, thereby enhancing the potential of the circulating protein to bind doxorubicin once functionalized by 1 in the bloodstream. Thus, the derivative 1 was synthesized from the primary amine 4[45] already described in the literature via nucleophile substitution with the N-hydroxysuccinimide ester 5 (for the full synthesis of 4 see the Supporting Information).

Next, we determined the stability constant of the Dox/1 complex in phosphate buffer (0.02 M, pH 7) by analyzing the changes in fluorescence intensity of Dox when incubated with various concentration of the β-CD trimer 1. As shown in Figure 2a, the fluorescence intensity of Dox decreased as the concentration of 1 increased, indicating the supramolecular interaction between both compounds.

https://doi.org/10.1021/acs.bioconjchem.2c00182
Bioconjugate Chem. 2022, 33, 1138−1144

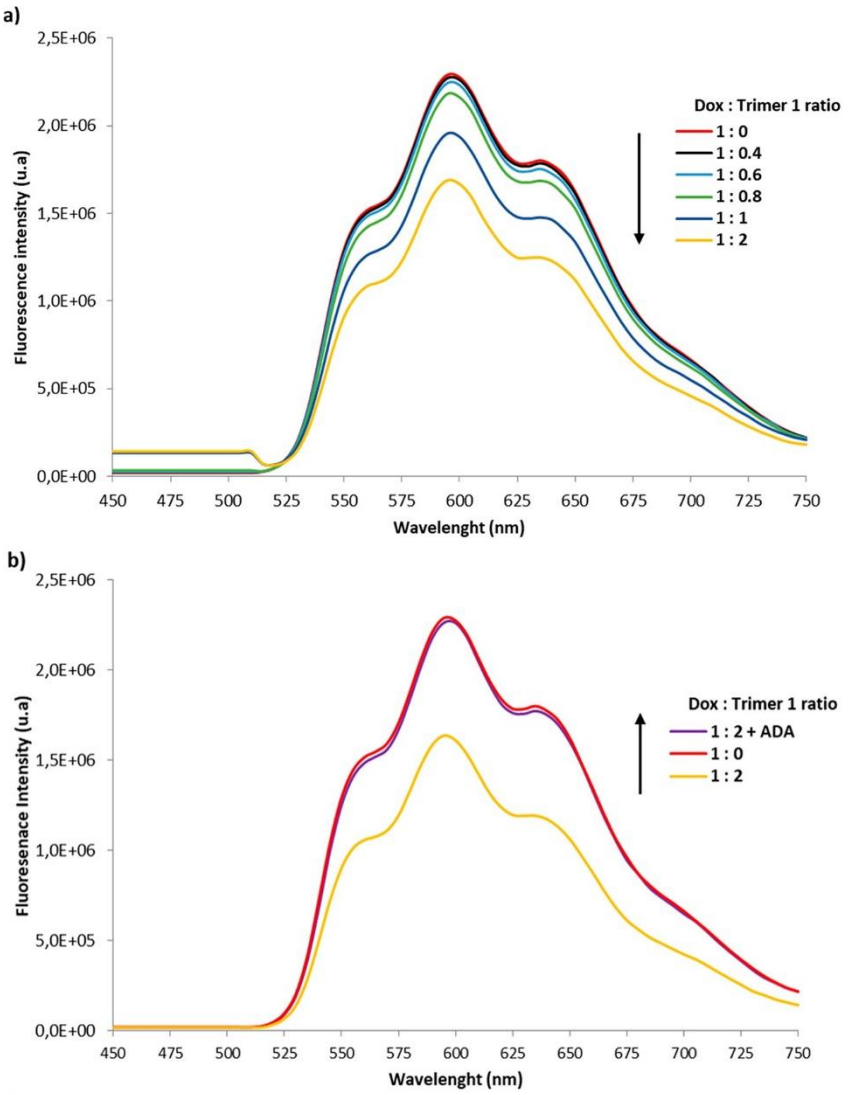


Figure 2. (a) Effects of the β-CD trimer 1 on fluorescence spectrum of Dox in 0.02 M phosphate buffer at pH 7 (λ_{ex} 500 nm). [Dox] = 10^{-5} M; [1] = 4 × 10^{-6} or 6 × 10^{-6} or 8 × 10^{-6} or 10^{-5} or 2 × 10^{-5} M. (b) Effect of ADA (5 equiv per β-CD unit) on fluorescence spectrum of a mixture of Dox and 1 incubated in 0.02 M phosphate buffer at pH 7 at 10^{-5} M and 2 × 10^{-5} M, respectively.

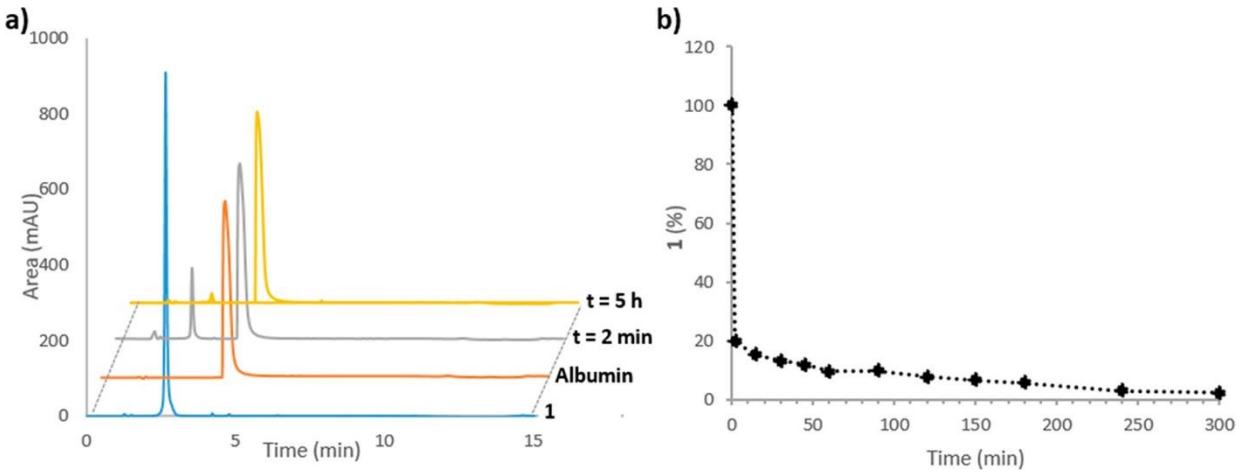


Figure 3. β-CD trimer 1 binds covalently with HSA under physiological conditions (phosphate buffer, pH 7, 37 °C). (a) Chromatograms of 1 (blue) incubated with HSA (orange) after 2 min (gray) and 5 h (yellow). (b) Disappearance of 1 over the time when placed in the presence of HSA.

https://doi.org/10.1021/acs.bioconjchem.2c00182
Bioconjugate Chem. 2022, 33, 1138−1144

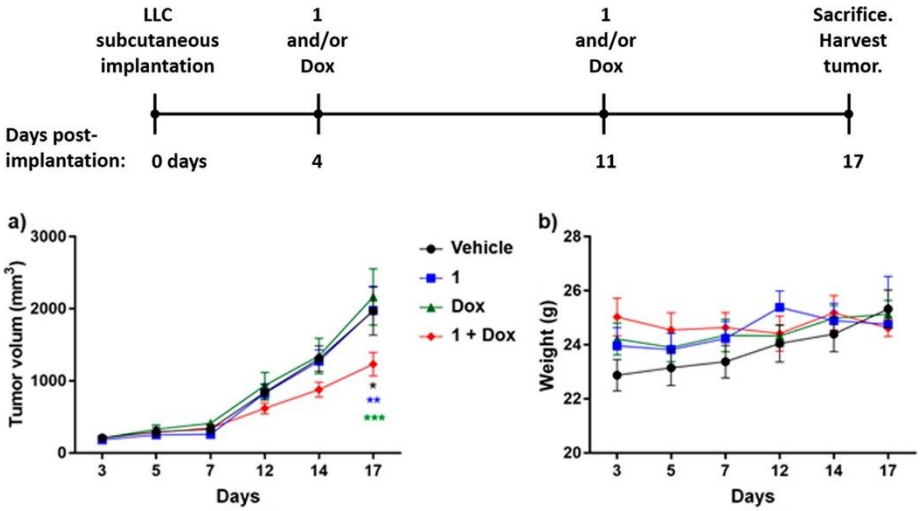


Figure 4. Antitumor activity of the 1/Dox combination on LLC xenograft. (a) LLC tumor growth inhibition under therapy with vehicle, 1, Dox and 1/Dox combination i.v. administered at days 4 and 11. Each point shows the mean ± s.e.m. from 6 tumor volumes. *P < 0.1, **P < 0.01, and *** P < 0.001. (b) Mean body weight of each group bearing LLC xenografts.

In order to confirm the association between 1 and Dox, we conducted additional fluorescence experiments in the presence of an excess of adamantane (ADA, 5 equiv per β-CD unit). In this case, the florescence intensity of Dox increased and became similar to that observed with the free drug (Figure 2b). This reversion of the fluorescence quenching indicated the adamantane-mediated release of Dox from the cavity of β-CD, therefore demonstrating the prior host−guest interactions between the 1 and the drug. The K_c value calculated using the Scott's equation was 6.9×10^5 M^{-1} which was 3 orders of magnitude higher than that of the complex between Dox and β-CD ($K_c = 1.8 \times 10^2$ M^{-1}).[46] This result indicated that the host−guest interactions between 1 and Dox should be strong enough to enable the formation of a relatively stable inclusion complex in the bloodstream.[47]

We then ascertained whether the bioconjugate 2 could be formed spontaneously under physiological conditions. For this purpose, the β-CD trimer 1 and human serum albumin (HSA) were incubated together in phosphate buffer (0.02 M, pH 7) at 37 °C and the evolution of the mixture over the time was monitored by HPLC. Under these conditions, 80% of 1 was converted in 2 min as the result of its rapid covalent binding with the protein (Figure 3). Furthermore, the peak corresponding to compound 1 disappeared almost completely after 5 h of incubation.

Trypsin digestion followed by HPLC/HRMS analysis confirmed the formation of the coupling product 2 with the detection of the HSA peptide fragment including the cysteine-34 linked to 1 (see the Supporting Information). These results suggested that the kinetics of formation of the albumin conjugate 2 is compatible with the use of the β-CD trimer 1 for the targeting of doxorubicin in vivo.[29,30,35]

The efficacy of the combined therapy involving both the β-CD trimer 1 and Dox was assessed in a subcutaneous murine Lewis lung carcinoma (LLC) implanted in C57BL/6 mice. The animals received two i.v. administrations of the 1/Dox combination, 1, or Dox alone at days 4 and 11 after tumor transplantation. In the case of the combined treatment, the Dox was injected 5 h after the trimer 1. It is worth mentioning that neither the 1/Dox ratio nor the delay between the administrations of these two compounds were optimized here.

In this pilot study, the dose of 1 (38.8 $mg.kg^{-1}$, 6.5 $\mu mol.kg^{-1}$, and 19.5 $\mu mol.kg^{-1}$ equivalent β-CD) was chosen in order to administer an excess of β-CDs compared to Dox (5 $mg.kg^{-1}$, 8.6 $\mu mol.kg^{-1}$) for favoring the formation of the inclusion complex in the bloodstream. Furthermore, the 5 h separating the injections of 1 and Dox was selected based on our binding kinetics studies (Figure 3) with the aim to limit the concentration of the free β-CD trimer in the circulation, prior to administering the drug. As shown in Figure 4, the Dox did not exhibit any antitumor activity at the tested dose. In contrast, the combined treatment produced a significant therapeutic effect with a slowdown in tumor growth compared to the control group. At the end of this experiment (day 17), the tumor volumes measured in the group of mice that received the 1/Dox combination were 40% lower than those recorded in the animals treated with Dox. This superior therapeutic efficacy can be attributed to a potential interaction between 1 and Dox in the bloodstream, since the β-CD trimer did not induce antitumor activity when injected alone. Moreover, the combined treatment was well tolerated without any body weight loss or sign of overt toxicity.

To verify that this enhanced efficacy was the result of a higher deposition of the drug within malignant tissues, mice were euthanized at day 17, and the relative concentration of Dox in tumors was quantified by HPLC/HRMS (Figure 5). These analyses revealed that the amount of drug delivered at the tumor site was 1.4-fold higher following the administration of the combined 1/Dox therapy than that of Dox alone.

Overall, these results demonstrate that the β-CD trimer 1 potentiates the therapeutic efficacy of Dox by increasing its concentration within tumor mass without inducing side effects. Interestingly, since compound 1 does not include a cytotoxic drug prior to being administered in vivo, its transfer toward the clinic will be greatly facilitated, rendering the production process as well as regulatory toxicity studies easier than for other drug conjugates. However, while this study represents a proof of principle, the potential of this targeting strategy could be further improved. Indeed, as the bifunctional derivative 1 ensures the interface between the anticancer drug and circulating albumin, the administered dose of the β-CD trimer could be optimized in order to increase the deposition of Dox

https://doi.org/10.1021/acs.bioconjchem.2c00182
Bioconjugate Chem. 2022, 33, 1138−1144

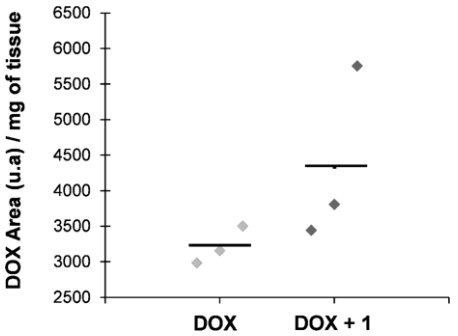


Figure 5. Relative concentration of Dox delivered in tumors when mice were treated with Dox (n = 3) or 1/Dox combination (n = 3). Black lines: mean values.

in the tumor microenvironment. In the same way, the delay between the administrations of the two partners of the combined treatment could be adjusted for injecting the Dox when the plasma concentration of the bioconjugate 2 will be optimal.

CONCLUSION

In summary, we developed a novel targeting approach based on the use of a β-CD trimer that binds in situ to serum albumin via a covalent linkage. The resulting albumin conjugate possesses the ability to entrap the Dox in the bloodstream through the formation of a host−guest inclusion complex, thereby allowing the transport and delivery of the anticancer drug in malignant tissues. Thus, when employed in combination with the β-CD trimer, the Dox exhibits a higher antitumor activity than when used alone as the consequence of its increased deposition in the tumor. As several other anticancer drugs such as paclitaxel, camptothecin, or cisplatin have been successfully incorporated in β-CD, this bifunctional trimer could serve as a versatile drug carrier to enhance the efficacy of cancer chemotherapy. Under such circumstances, this targeting strategy could be very advantageous since the therapeutic index of various cytotoxics could be potentiated thanks to a unique drug delivery system.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.bioconjchem.2c00182.

Synthetic procedures, ^{1}H NMR and ^{13}C NMR spectra for all compounds, materials and methods for biological experiments (PDF)

AUTHOR INFORMATION

Corresponding Author

Sébastien Papot − University of Poitiers, UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Equipe Labellisée Ligue Contre le Cancer, 86073 Poitiers, cedex 9, France; Seekyo SA, 86961 Futuroscope, France; orcid.org/0000-0001-5253-1979; Email: sebastien.papot@univ-poitiers.fr

Authors

Chad Plumet − University of Poitiers, UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Equipe Labellisée Ligue Contre le Cancer, 86073 Poitiers, cedex 9, France

Rémi Chaître − University of Poitiers, UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Equipe Labellisée Ligue Contre le Cancer, 86073 Poitiers, cedex 9, France

Fabiola Djago − University of Poitiers, UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Equipe Labellisée Ligue Contre le Cancer, 86073 Poitiers, cedex 9, France

Elodie Péraudeau − University of Poitiers, UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Equipe Labellisée Ligue Contre le Cancer, 86073 Poitiers, cedex 9, France

Quentin Blancart-Remaury − University of Poitiers, UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Equipe Labellisée Ligue Contre le Cancer, 86073 Poitiers, cedex 9, France

Jonathan Clarhaut − University of Poitiers, UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Equipe Labellisée Ligue Contre le Cancer, 86073 Poitiers, cedex 9, France; CHU de Poitiers, 86021 Poitiers, France

Claude Geffroy − University of Poitiers, UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Equipe Labellisée Ligue Contre le Cancer, 86073 Poitiers, cedex 9, France

Achmet Said Mohamed − University of Poitiers, UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Equipe Labellisée Ligue Contre le Cancer, 86073 Poitiers, cedex 9, France

Isabelle Opalinski − University of Poitiers, UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Equipe Labellisée Ligue Contre le Cancer, 86073 Poitiers, cedex 9, France

Brigitte Renoux − University of Poitiers, UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Equipe Labellisée Ligue Contre le Cancer, 86073 Poitiers, cedex 9, France

Pauline Poinot − University of Poitiers, UMR CNRS 7285, Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Equipe Labellisée Ligue Contre le Cancer, 86073 Poitiers, cedex 9, France; orcid.org/0000-0002-7927-1887

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.bioconjchem.2c00182

Author Contributions

#C.P. and R.C. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank CNRS, La Ligue contre le Cancer (Ligue nationale, comités Vienne and Deux-Sevres) and "Région Nouvelle Aquitaine" for financial support of this study.

REFERENCES

(1) Baah, S.; Laws, M.; Rahman, K. M. Antibody−Drug Conjugates A Tutorial Review. Molecules 2021, 26 (10), 2943.

(2) Chau, C. H.; Steeg, P. S.; Figg, W. D. Antibody−Drug Conjugates for Cancer. Lancet 2019, 394 (10200), 793−804.

(3) Beck, A.; Goetsch, L.; Dumontet, C.; Corvaïa, N. Strategies and Challenges for the next Generation of Antibody−Drug Conjugates. Nat. Rev. Drug Discovery 2017, 16 (5), 315−337.

(4) Yu, L.; Hua, Z.; Luo, X.; Zhao, T.; Liu, Y. Systematic Interaction of Plasma Albumin with the Efficacy of Chemotherapeutic Drugs. Biochim. Biophys. Acta Rev. Cancer 2022, 1877 (1), 188655.
(5) Spada, A.; Emami, J.; Tuszynski, J. A.; Lavasanifar, A. The Uniqueness of Albumin as a Carrier in Nanodrug Delivery. Mol. Pharmaceutics 2021, 18 (5), 1862−1894.
(6) Rahimizadeh, P.; Yang, S.; Lim, S. I. Albumin: An Emerging Opportunity in Drug Delivery. Biotechnol. Bioprocess Eng. 2020, 25 (6), 985−995.
(7) Zhao, P.; Wang, Y.; Wu, A.; Rao, Y.; Huang, Y. Roles of Albumin-Binding Proteins in Cancer Progression and Biomimetic Targeted Drug Delivery. Chembiochem Eur. J. Chem. Biol. 2018, 19 (17), 1796−1805.
(8) Tao, C.; Chuah, Y. J.; Xu, C.; Wang, D.-A. Albumin Conjugates and Assemblies as Versatile Bio-Functional Additives and Carriers for Biomedical Applications. J. Mater. Chem. B 2019, 7 (3), 357−367.
(9) Liu, Z.; Chen, X. Simple Bioconjugate Chemistry Serves Great Clinical Advances: Albumin as a Versatile Platform for Diagnosis and Precision Therapy. Chem. Soc. Rev. 2016, 45 (5), 1432−1456.
(10) Kratz, F. A Clinical Update of Using Albumin as a Drug Vehicle - a Commentary. J. Control. Release Off. J. Control. Release Soc. 2014, 190, 331−336.
(11) Elsadek, B.; Kratz, F. Impact of Albumin on Drug Delivery−New Applications on the Horizon. J. Control. Release Off. J. Control. Release Soc. 2012, 157 (1), 4−28.
(12) Kratz, F. Albumin as a Drug Carrier: Design of Prodrugs, Drug Conjugates and Nanoparticles. J. Control. Release Off. J. Control. Release Soc. 2008, 132 (3), 171−183.
(13) Gradishar, W. J.; Tjulandin, S.; Davidson, N.; Shaw, H.; Desai, N.; Bhar, P.; Hawkins, M.; O'Shaughnessy, J. Phase III Trial of Nanoparticle Albumin-Bound Paclitaxel Compared with Polyethylated Castor Oil-Based Paclitaxel in Women with Breast Cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 2005, 23 (31), 7794−7803.
(14) Chiorean, E. G.; Cheung, W. Y.; Giordano, G.; Kim, G.; Al-Batran, S.-E. Real-World Comparative Effectiveness of Nab-Paclitaxel plus Gemcitabine versus FOLFIRINOX in Advanced Pancreatic Cancer: A Systematic Review. Ther. Adv. Med. Oncol. 2019, 11, 175883591985036.
(15) Zhang, Y.; Xu, J.; Hua, J.; Liu, J.; Liang, C.; Meng, Q.; Ni, Q.; Shi, S.; Yu, X. Nab-Paclitaxel plus Gemcitabine as First-Line Treatment for Advanced Pancreatic Cancer: A Systematic Review and Meta-Analysis. J. Cancer 2019, 10 (18), 4420−4429.
(16) Kratz, F.; Müller-Driver, R.; Hofmann, I.; Drevs, J.; Unger, C. A Novel Macromolecular Prodrug Concept Exploiting Endogenous Serum Albumin as a Drug Carrier for Cancer Chemotherapy. J. Med. Chem. 2000, 43 (7), 1253−1256.
(17) Kratz, F.; Warnecke, A.; Scheuermann, K.; Stockmar, C.; Schwab, J.; Lazar, P.; Drückes, P.; Esser, N.; Drevs, J.; Rognan, D.; et al. Probing the Cysteine-34 Position of Endogenous Serum Albumin with Thiol-Binding Doxorubicin Derivatives. Improved Efficacy of an Acid-Sensitive Doxorubicin Derivative with Specific Albumin-Binding Properties Compared to That of the Parent Compound. J. Med. Chem. 2002, 45 (25), 5523−5533.
(18) Schmid, B.; Chung, D.-E.; Warnecke, A.; Fichtner, I.; Kratz, F. Albumin-Binding Prodrugs of Camptothecin and Doxorubicin with an Ala-Leu-Ala-Leu-Linker That Are Cleaved by Cathepsin B: Synthesis and Antitumor Efficacy. Bioconjugate Chem. 2007, 18 (3), 702−716.
(19) Abu Ajaj, K.; Graeser, R.; Fichtner, I.; Kratz, F. In Vitro and in Vivo Study of an Albumin-Binding Prodrug of Doxorubicin That Is Cleaved by Cathepsin B. Cancer Chemother. Pharmacol. 2009, 64 (2), 413−418.
(20) Pes, L.; Koester, S. D.; Magnusson, J. P.; Chercheja, S.; Medda, F.; Abu Ajaj, K.; Rognan, D.; Daum, S.; Nollmann, F. I.; Garcia Fernandez, J.; et al. Novel Auristatin E-Based Albumin-Binding Prodrugs with Superior Anticancer Efficacy in Vivo Compared to the Parent Compound. J. Control. Release Off. J. Control. Release Soc. 2019, 296, 81−92.
(21) Cranmer, L. D. Spotlight on Aldoxorubicin (INNO-206) and Its Potential in the Treatment of Soft Tissue Sarcomas: Evidence to Date. OncoTargets Ther. 2019, 12, 2047−2062.
(22) Fishman, W. H.; Anlyan, A. J. Comparison of the Beta-Glucuronidase Activity of Normal, Tumor, and Lymph Node Tissues of Surgical Patients. Science 1947, 106 (2742), 66−67.
(23) Boyland, E.; Wallace, D. M.; Williams, D. C. The Activity of the Enzymes Sulphatase and Beta-Glucuronidase in the Urine, Serum and Bladder Tissue. Br. J. Cancer 1955, 9 (1), 62−79.
(24) Albin, N.; Massaad, L.; Toussaint, C.; Morizet, J.; Parise, O.; Gouyette, A.; Chabot, G. G. Main Drug-Metabolizing Enzyme Systems in Human Breast Tumors and Peritumoral Tissues. Cancer Res. 1993, 53, 3541.
(25) Sperker, B.; Werner, U.; Mürdter, T. E.; Tekkaya, C.; Fritz, P.; Wacke, R.; Adam, U.; Gerken, M.; Drewelow, B.; Kroemer, H. K. Expression and Function of Beta-Glucuronidase in Pancreatic Cancer: Potential Role in Drug Targeting. Naunyn. Schmiedebergs Arch. Pharmacol. 2000, 362 (2), 110−115.
(26) Nygren, C.; von Holst, H.; Mansson, J.-E.; Fredman, P. Increased Activity of Lysosomal Glycohydrolases in Glioma Tissue and Surrounding Areas from Human Brain. Acta Neurochir. (Wien) 1997, 139 (2), 146−150.
(27) Bosslet, K.; Czech, J.; Hoffmann, D. A Novel One-Step Tumor-Selective Prodrug Activation System. Tumor Target. 1995, 1, 45−50.
(28) Bosslet, K.; Straub, R.; Blumrich, M.; Czech, J.; Gerken, M.; Sperker, B.; Kroemer, H. K.; Gesson, J. P.; Koch, M.; Monneret, C. Elucidation of the Mechanism Enabling Tumor Selective Prodrug Monotherapy. Cancer Res. 1998, 58 (6), 1195−1201.
(29) Legigan, T.; Clarhaut, J.; Renoux, B.; Tranoy-Opalinski, I.; Monvoisin, A.; Berjeaud, J.-M.; Guilhot, F.; Papot, S. Synthesis and Antitumor Efficacy of a β-Glucuronidase-Responsive Albumin-Binding Prodrug of Doxorubicin. J. Med. Chem. 2012, 55 (9), 4516−4520.
(30) Renoux, B.; Raes, F.; Legigan, T.; Péraudeau, E.; Eddhif, B.; Poinot, P.; Tranoy-Opalinski, I.; Alsarraf, J.; Koniev, O.; Kolodych, S.; et al. Targeting the Tumour Microenvironment with an Enzyme-Responsive Drug Delivery System for the Efficient Therapy of Breast and Pancreatic Cancers †Electronic Supplementary Information (ESI) Available. See DOI: 10.1039/C7sc00472a Click Here for Additional Data File. Chem. Sci. 2017, 8 (5), 3427−3433.
(31) Compain, G.; Oumata, N.; Clarhaut, J.; Péraudeau, E.; Renoux, B.; Galons, H.; Papot, S. A β-Glucuronidase-Responsive Albumin-Binding Prodrug for Potential Selective Kinase Inhibitor-Based Cancer Chemotherapy. Eur. J. Med. Chem. 2018, 158, 1−6.
(32) Renoux, B.; Fangous, L.; Hötten, C.; Péraudeau, E.; Eddhif, B.; Poinot, P.; Clarhaut, J.; Papot, S. A β-Glucuronidase-Responsive Albumin-Binding Prodrug Programmed for the Double Release of Monomethyl Auristatin E. MedChemComm 2018, 9 (12), 2068−2071.
(33) Lange, J.; Eddhif, B.; Tarighi, M.; Garandeau, T.; Péraudeau, E.; Clarhaut, J.; Renoux, B.; Papot, S.; Poinot, P. Volatile Organic Compound Based Probe for Induced Volatolomics of Cancers. Angew. Chem., Int. Ed. 2019, 58 (49), 17563−17566.
(34) Raes, F.; Badiane, S. M.; Renoux, B.; Papot, S.; Lerondel, S.; Le Pape, A. Development of an Embedded Multimodality Imaging Platform for Onco-Pharmacology Using a Smart Anticancer Prodrug as an Example. Sci. Rep. 2020, 10 (1), 2661.
(35) Chatre, R.; Lange, J.; Péraudeau, E.; Poinot, P.; Lerondel, S.; Le Pape, A.; Clarhaut, J.; Renoux, B.; Papot, S. In Vivo Synthesis of Triple-Loaded Albumin Conjugate for Efficient Targeted Cancer Chemotherapy. J. Controlled Release 2020, 327, 19−25.
(36) Huang, Y.; Wang, L.; Cheng, Z.; Yang, B.; Yu, J.; Chen, Y.; Lu, W. SN38-Based Albumin-Binding Prodrug for Efficient Targeted Cancer Chemotherapy. J. Control. Release Off. J. Control. Release Soc. 2021, 339, 297−306.
(37) Cheng, Z.; Huang, Y.; Shen, Q.; Zhao, Y.; Wang, L.; Yu, J.; Lu, W. A Camptothecin-Based, Albumin-Binding Prodrug Enhances Efficacy and Safety in Vivo. Eur. J. Med. Chem. 2021, 226, 113851.
(38) Jarlstad Olesen, M. T.; Walther, R.; Poier, P. P.; Dagnæs-Hansen, F.; Zelikin, A. N. Molecular, Macromolecular, and Supramolecular Glucuronide Prodrugs: Lead Identified for Anticancer

Prodrug Monotherapy. Angew. Chem., Int. Ed. 2020, 59 (19), 7390−7396.
(39) [illegible]
for Che[illegible]
2020, 232, 115805.
(40) Bonnet, V.; Gervaise, C.; Djedaïni-Pilard, F.; Furlan, A.; Sarazin, C. Cyclodextrin Nanoassemblies: A Promising Tool for Drug Delivery. Drug Discovery Today 2015, 20 (9), 1120−1126.
(41) Mohammad, N.; Singh, S. V.; Malvi, P.; Chaube, B.; Athavale D.; Vanuopadath, M.; Nair, S. S.; Nair, B.; Bhat, M. K. Strategy to Enhance Efficacy of Doxorubicin in Solid Tumor Cells by Methyl-β-Cyclodextrin: Involvement of P53 and Fas Receptor Ligand Complex. Sci. Rep. 2015, 5, 11853.
(42) Okamatsu, A.; Motoyama, K.; Onodera, R.; Higashi, T.; Koshigoe, T.; Shimada, Y.; Hattori, K.; Takeuchi, T.; Arima, H. Folate-Appended β-Cyclodextrin as a Promising Tumor Targeting Carrier for Antitumor Drugs in Vitro and in Vivo. Bioconjugate Chem. 2013, 24 (4), 724−733.
(43) Fiorica, C.; Palumbo, F. S.; Pitarresi, G.; Puleio, R.; Condorelli, L.; Collura, G.; Giammona, G. A Hyaluronic Acid/Cyclodextrin Based Injectable Hydrogel for Local Doxorubicin Delivery to Solid Tumors. Int. J. Pharm. 2020, 589, 119879.
(44) Santos, A. C.; Costa, D.; Ferreira, L.; Guerra, C.; Pereira-Silva, M.; Pereira, I.; Peixoto, D.; Ferreira, N. R.; Veiga, F. Cyclodextrin-Based Delivery Systems for in Vivo-Tested Anticancer Therapies. Drug Delivery Transl. Res. 2021, 11 (1), 49−71.
(45) Plumet, C.; Mohamed, A. S.; Vendeuvre, T.; Renoux, B. Clarhaut, J.; Papot, S. Cell−Cell Interactions via Non-Covalent Click Chemistry. Chem. Sci. 2021, 12 (26), 9017−9021.
(46) Bakkour, Y.; Vermeersch, G.; Morcellet, M.; Boschin, F.; Martel, B.; Azaroual, N. Formation of Cyclodextrin Inclusion Complexes with Doxycyclin-Hyclate: NMR Investigation of Their Characterisation and Stability. J. Incl. Phenom. 2006, 54, 109−114.
(47) Stella, V. J.; He, Q. Cyclodextrins. Toxicol. Pathol. 2008, 36 (1), 30−42.

Recommended by ACS

Targeted Drug Delivery Using a Plug-to-Direct Antibody–Nanogel Conjugate
Uyen Huynh, S. Thayumanavan, et al.
JANUARY 13, 2023
BIOMACROMOLECULES
READ

Co-Encapsulation of Hydrophilic and Hydrophobic Drugs into Human H Chain Ferritin Nanocarrier Enhances Antitumor Efficacy
Chen Ning, Yu Zhang, et al.
APRIL 18, 2023
ACS BIOMATERIALS SCIENCE & ENGINEERING
READ

Design and Evaluation of a Carrier-Free Prodrug-Based Palmitic–DEVD–Doxorubicin Conjugate for Targeted Cancer Therapy
Seong-Bin Yang, Jooho Park, et al.
FEBRUARY 03, 2023
BIOCONJUGATE CHEMISTRY
READ

Effective Delivery of Paclitaxel-Loaded Ferritin via Inverso CendR Peptide for Enhanced Cancer Therapy
Yixin Dong, Yu Zhang, et al.
DECEMBER 27, 2022
MOLECULAR PHARMACEUTICS
READ

Get More Suggestions >

Développement d'une méthode d'analyse de conjugués aminoacyl-ARN

Les conjugués aminoacyl- ou peptide-ARN sont des biomolécules naturelles ou synthétiques composés d'une partie ARN liée de manière covalente à un acide aminé ou un peptide. Les ARN naturels les plus connus contenant un résidu d'acide aminé sont les aminoacyl-ARNt (aa-ARNt) dans lesquels l'acide aminé est connecté en position 3' ou 2' du ribose terminal par une liaison ester. Les aa-ARNt sont utilisés pour la synthèse des protéines ribosomales mais également dans la synthèse des peptides non ribosomiques (NRPS).
Les aminoacyl-ARNt participent également au métabolisme secondaire, y compris la synthèse de dipeptides cycliques et d'antibiotiques naturels. Un autre exemple d'aa-ARN est celui des ARN contenant des bases non canoniques modifiées avec des acides aminés. Il a été suggéré qu'une telle modification de l'ARN était représentative de fossiles moléculaires vivants.

Dans tous ces domaines, la synthèse d'ARN fonctionnalisés par des acides aminés ou des peptides est très utile et donne accès à des outils chimiques puissants pour explorer leur fonction et leurs mécanismes biologiques. L'exploration du rôle des conjugués d'acides aminés ou d'ARN peptidiques dans les processus biologiques nécessite des processus de synthèse particuliers (ligation de Staudinger) pour former des molécules stables.

Les deux articles présentés ci-après, publiés en Septembre 2020 dans Organic Letters puis en Janvier 2023 dans ACS Omega, détaillent la synthèse d'ARN fonctionnalisés par des acides aminés ou petits peptides. Mon rôle dans ces travaux a été de développer la méthode d'analyse par UPLC-HRMS de ces composés afin d'aider à la compréhension et l'optimisation des schémas de synthèse.

OL | Organic Letters

pubs.acs.org/OrgLett Letter

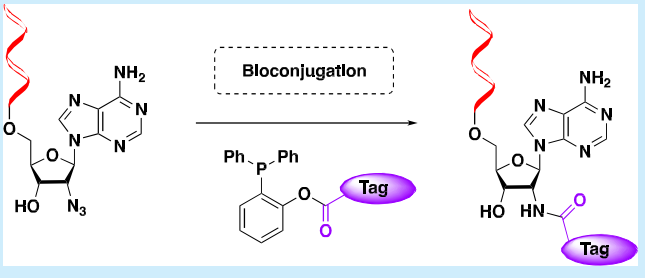
Phosphine-Mediated Bioconjugation of the 3′-End of RNA

Camelia Kitoun, Matthieu Fonvielle, Nicolas Sakkas, Manon Lefresne, Fabiola Djago, Quentin Blancart Remaury, Pauline Poinot, Michel Arthur, Melanie Etheve-Quelquejeu,* and Laura Iannazzo*

Cite This: Org. Lett. 2020, 22, 8034−8038 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Staudinger ligation is an attractive bio-orthogonal reaction that has been widely used to tag proteins, carbohydrates, and nucleic acids. Here, we explore the traceless variant of the Staudinger ligation for 3′-end modification of oligoribonucleotides. An azido-containing dinucleotide was used to study the ligation. Nine phosphines containing reactive groups, affinity purification tags, or photoswitch probes have been successfully obtained. The corresponding modified dinucleotides were synthesized and characterized by LC/MS. Mechanistic interpretations of the reaction are proposed, in particular, the unprecedented formation of an oxazaphospholane nucleotide derivative, which was favored by the vicinal position of 2′-N_3 and 3′-OH functional groups on the terminal ribose has been observed. The post-functionalization of a 24-nt RNA with a photoactivable tag is also reported.



Bioconjugation reactions[1] developed in the past decade have enabled the chemical labeling of biomolecules to enrich their functions and properties. Because of the complexity of the biomolecules, due to the presence of reactive groups and their sensitivity to nonphysiological pH or temperature, the bioconjugation reactions must be chemoselective and regioselective, and compatible with almost physiological conditions in aqueous media. One strategy to enhance the selectivity of a bioconjugation reaction is to first install a unique functional group onto the biomolecule, and then use site-selective type reactions to covalently link the desired tags. Several site-selective reactions that target non-native functional groups of biomolecules have been reported so far:

(i) the condensation between a carbonyl and hydrazines, hydrazides, or alkoxyamines;[2]
(ii) metal-mediated reactions, such as alkyne−azide cycloadditions (CuAAC or NiAAC),[3] sydnone−alkyne cycloaddition (CuSAC),[4] or palladium-catalyzed cross-coupling reactions;[5]
(iii) strained-promoted cycloadditions, such as dipolar cycloadditions with constrainted alkynes (SPAAC),[6] or inverse electron demand Diels−Alder reactions (IEDDA);[7]
(iv) photoclick reactions;[8] and, finally,
(v) chemical ligation such as native chemical ligation (NCL)[9] or Staudinger ligation.[10]

The latter reaction is an attractive bio-orthogonal conjugation reaction, especially for in vitro and in vivo applications, because of the absence of a metal catalyst and to its capacity to form native amide bond between the tag and the biomolecule. The traceless variant of the Staudinger ligation, independently developed by Raines[11] and Bertozzi,[12] is even more interesting, since it leads to the elimination of the phosphine oxide from the final product. These new phosphines, developed by Bertozzi and co-workers,[12] contain an acyl component, which is attached through a cleavable linkage (an ester bond) at the ortho-position of the aryl group of the phenylphosphine. These reagents have been designed to avoid the oxidation of the phosphine under ambient conditions, to introduce conformational rigidity and to promote a six-membered transition state during the acyl transfer. To access a transition state with a smaller ring, Raines and co-workers replaced the O-phenyl group by a single methylene moiety and increased the yield of the traceless ligation to couple peptides and azides.[13] This methodology has been employed for the bioconjugation of polysaccharides,[14] peptides,[15] or proteins.[16] Application of the Staudinger-type ligation to nucleoside or nucleotide modification has been mostly restricted to the synthesis of puromycin analogues,[17] the functionalization with fluorescent probes[18] and the introduction of affinity purification tags.[19] The application of the traceless version of the Staudinger ligation has been even more restricted, and, to the best of our

Received: September 7, 2020
Published: September 30, 2020



ACS Publications © 2020 American Chemical Society

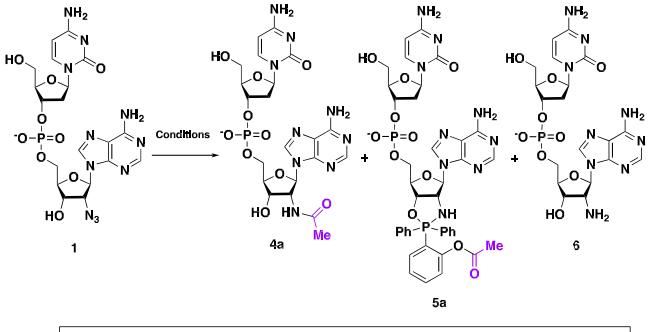
https://dx.doi.org/10.1021/acs.orglett.0c02982
Org. Lett. 2020, 22, 8034−8038

knowledge, only reported for the functionalization of a protected 5′-azido-2′-deoxyadenosine.[12] Since chemistry-driven studies of the biological role of RNAs require new synthetic methodologies to modify their structures, and because no example depicts the modification of RNA by traceless Staudinger ligation, we decided to explore this site-selective reaction as a post-synthetic methodology for the functionalization of RNAs with a variety of functional groups.

Our strategy for accessing functionalized RNA relies on the enzymatic ligation of a 5′-phosphorylated dinucleotides (pdCpA) to an oligoribonuleotide by the T4 RNA ligase. The presence of a 2′-deoxy-cytidine at the 5′-position of the dinucleotide facilitates its synthesis and is compatible with the specificity of the T4 RNA ligase. The presence of a 2′-deoxynucleotide at the penultimate position of the ligation product was previously reported to have little impact on the biological function of the final RNA.[20] To study the bioconjugation of the 3′-terminal RNA position, we selected the dinucleotide 1 containing an azido group at the 2′-position of the terminal adenosine as a model. The second partner of the reaction was the tag-substituted phosphine. Among the modified phosphine reagents developed in the literature for the traceless Staudinger ligation, we chose phenolic ester 2a and alkyl thioester 3, since they were reported to provide the highest ligation yields.[21]

When dinucleotide 1 was treated in the presence of 10 equiv of methylester phenolic phosphine 2a (synthesized as shown in Scheme 2, presented later in this work) and heated for 4 h at 70 °C, followed by the addition of water and 2 h of stirring at 70 °C, full conversion of the starting dinucleotide was observed (Scheme 1, entry 1). These conditions lead to the formation of the ligated compound 4a in 22% yield in competition with the

Scheme 1. Reaction of Dinucleotide 1 with Phenolic Ester 2a and Alkyl Thioester 3 Substituted with a Methyl Group

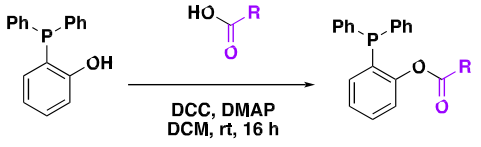


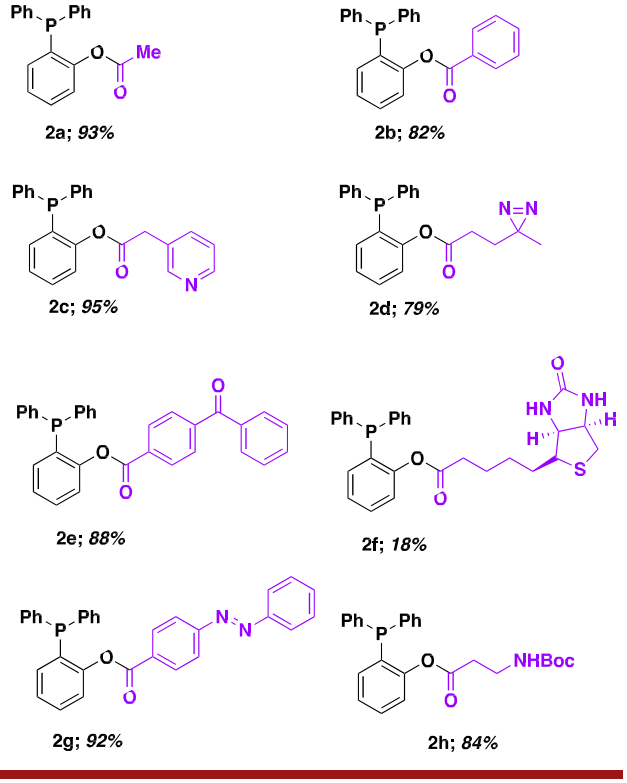
Entry	Phosphine[a]	Conditions	4a[b]	5a[b]	6[b]
1	2a	DMF/H_2O 9/1, 70 °C, 4 h, then H_2O, 70 °C, 2 h	22%	28%	50%
2	3	DABCO, DMF/H_2O 9/1, 40 °C, 16 h	23%	-	77%

[a]Using 10 equiv of phosphine. [b]Product yield based on relative integration in HPLC at 254 nm.

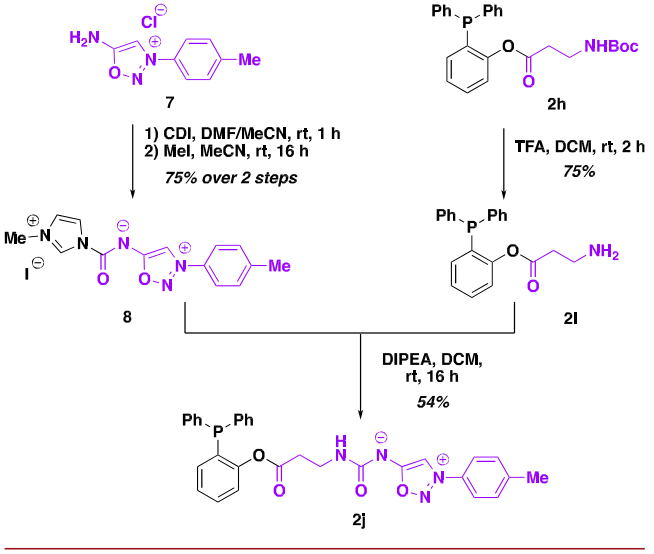
reduced derivative 6. In addition, a cyclic oxazaphospholane derivative 5a was obtained in 28% yield. Evidence for the formation of compound 5a was gathered by monitoring the reaction by ^{31}P NMR spectroscopy (see page S70 in the

Scheme 2. Synthesis of Phosphines 2a−2h





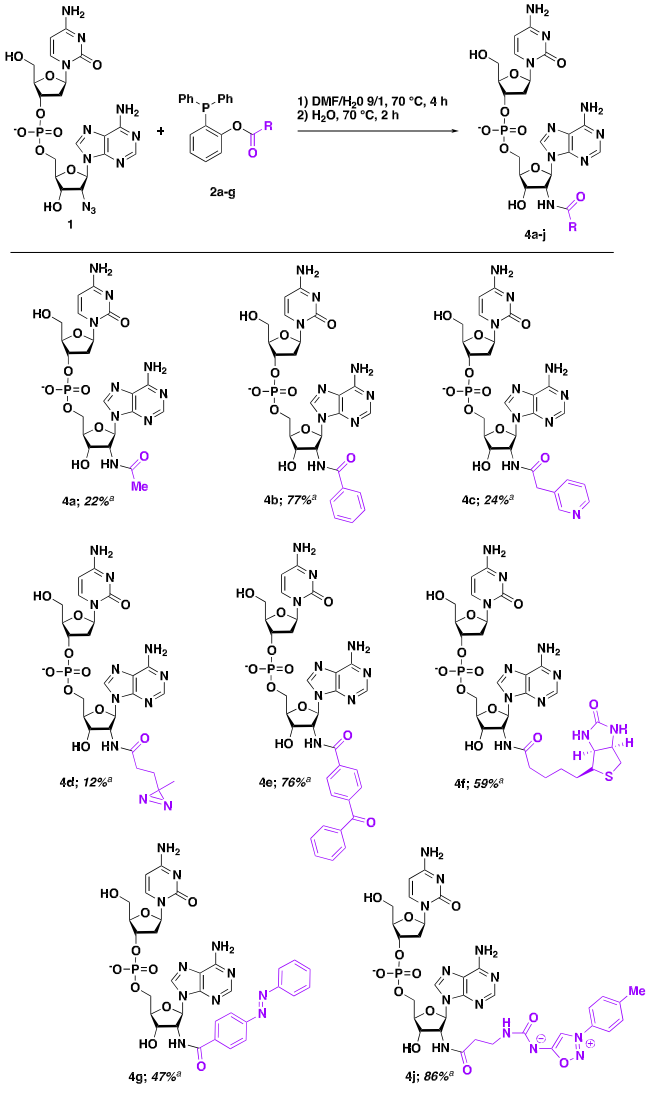
Scheme 3. Synthesis of Phosphine 2j



Supporting Information). After 55 min of reaction in DMF/D_2O (9/1, v/v), a growing signal at +40.6 ppm was observed, compared to the characteristic peak of the phosphine 2a at −15.8 ppm and the occurrence of oxidized species. The formation of such cyclic derivative could be explained by the 1,2-cis position[22] of the 2′-N_3 function and the 3′-OH hydroxyl on the ribose. The modification of the reaction conditions with

https://dx.doi.org/10.1021/acs.orglett.0c02982
Org. Lett. 2020, 22, 8034−8038

Scheme 4. Substrate Scope



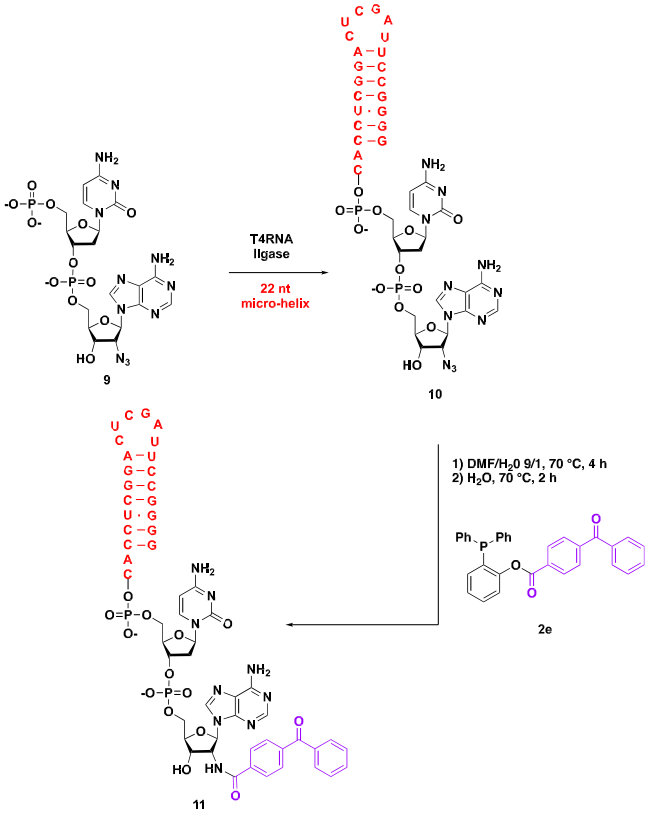
[a]Product yield based on relative integration in HPLC at 254 nm.

a longer reaction time or irradiation under microwaves did not allow an increase in the yield of the ligated compound 4a and dramatically increased the formation of the oxazaphospholane 5a (see page S56 in the Supporting Information).

We next performed the same ligation from the dinucleotide 1 with the alkyl thioester phosphine 3 (the commercially available (diphenylphosphino)methanethiol) described by Raines.[13] Treatment with DABCO was first applied to remove the borohydride group, which prevented oxidation of the phosphine. In this latter condition, the cyclic oxazaphospholane derivative 5a was not observed in the reaction mixture. The reduced compound 6 was obtained as the major product (77% yield), whereas the expected dinucleotide 4a was formed in 23% yield (Scheme 1, entry 2).

To expand the scope of the reaction, we selected the phosphine phenolic ester derivatives due to their stability and easy functionalization. A range of substituted phosphines was synthesized by coupling the commercially available 2-hydroxyphenyl diphenylphosphine with the corresponding carboxylic acid in the presence of DCC and DMAP in DCM at room temperature for 16 h. Eight phosphines were obtained in yields of 18%−95% (see Scheme 2).

Scheme 5. Bioconjugation of 24-nt RNA



Phosphine 2j containing a sydnone substituent, recently reported as a suitable partner for click and release experiments for bioconjugation,[4,23] was synthesized according to the synthetic route described in Scheme 3. Iminosydnone 7 was prepared following standard procedures[23] and activated using a two-step strategy[24] to afford the corresponding carbonyl-imidazolium salt 8 in 75% yield. The sydnone compound 8 was then treated in the presence of DIPEA and unprotected phosphine 2i in DCM to lead to the phosphine 2j in 54% yield.

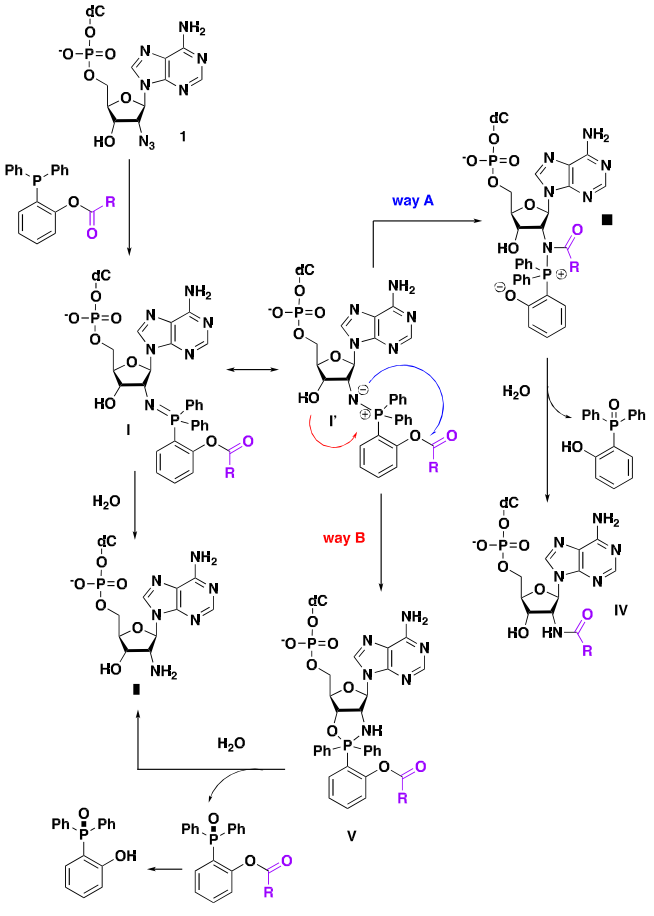
Phosphines 2a−2j containing representative functionalities were subjected to ligation conditions (see Scheme 4).

When phosphine-containing phenyl or pyridine rings (2b, 2c) were used for bioconjugation reaction with dinucleotide 1, compounds 4b and 4c were obtained in yields of 77% and 24%, respectively. Photoactivable dinucleotides 4d and 4e have been prepared in yields of 12% and 76%, respectively, using phosphine 2d[25] and 2e. A biotin moiety, which was used for the analysis and purification of biomolecules, was introduced in the 2′-position of the terminal ribose, in 59% yield. A photoswitch tag has also been ligated and diazobenzene phosphine 2g led to the formation of the corresponding modified dinucleotide 4g in 47% yield. The ligation was then applied to access to the dinucleotide-sydnone conjugate 4j in 86% yield. These results show that the functionalization in 2′-position of the terminal ribose of an unprotected dinucleotide is compatible with ester phenolic phosphine containing a variety of tags.[26]

Furthermore, we applied this reaction to post-functionalize a 24-nt RNA containing an azido group at the 2′-position of the terminal ribose. The 2′-azido-RNA 10 was obtained by enzymatic ligation of dinucleotide 9 (pdCpA-2′-N_3)[27] to a 22-nt RNA, using the T4 RNA ligase RNA molecule (see page

https://dx.doi.org/10.1021/acs.orglett.0c02982
Org. Lett. 2020, 22, 8034−8038

Scheme 6. Proposed Mechanism[a]



[a]Pathways A and B are described in the text.

S14 in the Supporting Information). The sequence of the RNA has been chosen to mimic the acceptor arm of Ala-tRNA^{Ala}. The oligonucleotide 10 was then treated by an excess of phosphine 2e (100 equiv) containing a benzophenone tag. After 4 h of stirring at 70 °C in a 9/1 mixture of DMF and water, followed by the addition of water and 2 h of stirring at 70 °C, the reaction lead to the formation of the functionalized RNA 11 in 7% yield (see Scheme 5). This was confirmed by mass spectrometry (MALDI analysis; see pages S66−S69 in the Supporting Information).

A putative mechanism accounting for our experimental results is proposed in Scheme 6. After formation of the iminophosphorane intermediate I, Staudinger reduction could lead directly to the formation of 2′-NH_2 dinucleotide II. This reaction occurred in competition with the acyl transfer from the phosphine to the nitrogen of the iminophosphorane (pathway A). This step is promoted by the six-membered ring formed during the transition state. Intermediate III could then be hydrolyzed affording the expected ligation product IV. Moreover, the nitrogen of the iminophosphorane and the hydroxyl at the 3′-position were in the cis configuration, which favored the ring-closing step (pathway B) and gave access to the oxazaphospholane intermediate V. This reactivity was reported by Bernardi and co-workers, who investigated the traceless ligation of glycofuranosyl azides.[22] Triaryliminophosphorane moiety[28] could promote the deprotonation of the alcohol at the 3′-position and favor the nucleophilic attack of the phosphorus. Then, the hydrolysis of V could provide reduced compound II and the oxidized phosphine.

In conclusion, we have reported the first application of the traceless Staudinger ligation to the post-functionalization of RNA. We synthesized a series of ester phenylphosphines, which contained a variety of tags, such as a biotin, photoswitch, and photoactivable moieties, and a sydnone. We studied the traceless variant of the Staudinger ligation on a model dinucleotide and showed that the reaction provided moderate to good yields, depending on the nature of the ester−phenyl phosphine. We extended this methodology for the bioconjugation of a 24-nt RNA strand with a photoactivable benzophenone probe.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.orglett.0c02982.

Characterization of phosphines 2a−2i; synthesis and characterization of phosphine 2j; synthesis and characterization of dinucleotide 1; synthesis and characterization of dinucleotide 9; NMR spectra; LCMS analyses (PDF)

FAIR data, including the primary NMR FID files, for compounds 1, 2a, 2b, 2c, 2d, 2e, 2f, 2g, 2h, 2i, 2j, 7, 9, and N-Boc-β-Ala (ZIP)

AUTHOR INFORMATION

Corresponding Authors

Laura Iannazzo − UMR CNRS 8601, Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques, Université de Paris, F-75006 Paris, France; orcid.org/0000-0002-2181-4885; Phone: (33) 1 76 53 42 50; Email: laura.iannazzo@parisdescartes.fr

Melanie Etheve-Quelquejeu − UMR CNRS 8601, Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques, Université de Paris, F-75006 Paris, France; Phone: (33) 1 76 53 42 22; Email: melanie.etheve-quelquejeu@parisdescartes.fr

Authors

Camelia Kitoun − UMR CNRS 8601, Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques, Université de Paris, F-75006 Paris, France

Matthieu Fonvielle − INSERM, Sorbonne Université, Université de Paris, Centre de Recherche des Cordeliers (CRC), F-75006 Paris, France

Nicolas Sakkas − UMR CNRS 8601, Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques, Université de Paris, F-75006 Paris, France

Manon Lefresne − UMR CNRS 8601, Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques, Université de Paris, F-75006 Paris, France

Fabiola Djago − Institut de Chimie des Milieux et Materiaux de Poitiers IC2MP, Université de Poitiers, 86073 Poitiers, France

Quentin Blancart Remaury − Institut de Chimie des Milieux et Materiaux de Poitiers IC2MP, Université de Poitiers, 86073 Poitiers, France

Pauline Poinot − Institut de Chimie des Milieux et Materiaux de Poitiers IC2MP, Université de Poitiers, 86073 Poitiers, France; orcid.org/0000-0002-7927-1887

https://dx.doi.org/10.1021/acs.orglett.0c02982
Org. Lett. 2020, 22, 8034−8038

Michel Arthur − INSERM, Sorbonne Universite, Universite de Paris, Centre de Recherche des Cordeliers (CRC), F-75006 Paris, France

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.orglett.0c02982

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was funded by grants from the Agence National de la Recherche (ANR), Project SyntRNA [No. ANR-17-CE07-0041-02].

REFERENCES

(1) (a) Lang, K.; Chin, J. W. ACS Chem. Biol. 2014, 9, 16−20. (b) Spicer, C. D.; Davis, B. G. Nat. Commun. 2014, 5, 4740.
(2) (a) Mahal, L. K.; Yarema, K. J.; Bertozzi, C. R. Science 1997, 276, 1125−1128. (b) Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R. Nat. Chem. Biol. 2007, 3, 321−322. (c) Kalia, J.; Raines, R. T. Angew. Chem., Int. Ed. 2008, 47, 7523−7526. (d) Flavell, R. R.; Kothari, P.; Bar-Dagan, M.; Synan, M.; Vallabhajosula, S.; Friedman, J. M.; Muir, T. W.; Ceccarini, G. J. Am. Chem. Soc. 2008, 130, 9106−9112.
(3) (a) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. Angew. Chem., Int. Ed. 2002, 41, 2596−2599. (b) Tornøe, C. W.; Christensen, C.; Meldal, J. M. J. Org. Chem. 2002, 67, 3057−3064 (c) Kim, W. G.; Kang, M. E.; Lee, J. B.; Jeon, M. H.; Lee, S.; Lee, J.; Choi, B.; Cal, P. M. S. D.; Kang, S.; Kee, J.-M.; et al. J. Am. Chem. Soc. 2017, 139, 12121−12124.
(4) Decuypere, E.; Plougastel, L.; Audisio, D.; Taran, F. Chem Commun. 2017, 53, 11515−11527.
(5) (a) Chalker, J. M.; Wood, C. S. C.; Davis, B. G. J. Am. Chem. Soc. 2009, 131, 16346−16347. (b) Bisseret, P.; Abdelkafi, H.; Blanchard, N. Chem. Sci. 2018, 9, 5132−5144. (c) Walunj, M. B.; Tanpure, A. A.; Srivatsan, S. G. Nucleic Acids Res. 2018, 46, No. e65. (d) Zhang, C.; Vinogradova, E. V.; Spokoyny, A. M.; Buchwald, S. L.; Pentelute, B. L Angew. Chem., Int. Ed. 2019, 58, 4810−4839.
(6) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. J. Am. Chem. Soc. 2004, 126, 15046−15047.
(7) Blackman, M. L.; Royzen, M.; Fox, J. M. J. Am. Chem. Soc. 2008, 130, 13518−13519.
(8) Wang, J.; Zhang, W.; Song, W.; Wang, Y.; Yu, Z.; Li, J.; Wu, M.; Wang, L.; Zang, J.; Lin, Q. J. Am. Chem. Soc. 2010, 132, 14812−14818.
(9) (a) Dawson, P. E.; Muir, T. W.; Clarklewis, I.; Kent, S. B. H. Science 1994, 266, 776−779. (b) Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. J. Am. Chem. Soc. 1999, 121, 11684−11689. (c) Johnson, E. C. B.; Kent, S. B. H. J. Am. Chem Soc. 2006, 128, 6640−6646.
(10) (a) Saxon, E.; Bertozzi, C. R. Science 2000, 287, 2007−2010. (d) Wang, Z. P A.; Tian, C. L.; Zheng, J. S. RSC Adv. 2015, 5, 107192−107199. (c) Bednarek, C.; Wehl, I.; Jung, N.; Schepers, U.; Brase, S. Chem. Rev. 2020, 120, 4301−4354.
(11) Nilsson, B. L.; Kiessling, L.; Raines, R. T. Org. Lett. 2000, 2, 1939−1941.
(12) Saxon, E.; Armstrong, J. I.; Bertozzi, C. R. Org. Lett. 2000, 2, 2141−2143.
(13) (a) Nilsson, B. L.; Kiessling, L.; Raines, R. T. Org. Lett. 2001, 3, 9−12. (b) Soellner, M. B.; Tam, A.; Raines, R. T. J. Org. Chem. 2006, 71, 9824−9830. (c) Tam, A.; Soellner, M. B.; Raines, R. T. J. Am Chem. Soc. 2007, 129, 11421−11430. (d) Tam, A.; Soellner, M. B.; Raines, R. T. Org. Biomol. Chem. 2008, 6, 1173−1175.
(14) (a) Grandjean, C.; Boutonnier, A.; Guerreiro, C.; Fournier, J M.; Mulard, L. A. J. Org. Chem. 2005, 70, 7123−7132. (b) Bianchi, A.; Bernardi, A. J. Org. Chem. 2006, 71 (12), 4565−4577. (c) Nisic, F.; Andreini, M.; Bernardi, A. Eur. J. Org. Chem. 2009, 2009, 5744−5751.
(d) Liu, S.; Edgar, K. J. Biomacromolecules 2015, 16, 2556−2571. (e) Zhu, S.; Guo, Z. Org. Lett. 2017, 19, 3063−3066.
(15) Yamashita, T.; Kuranaga, T.; Inoue, M. Org. Lett. 2015, 17, 2170−2173.
(16) (a) Tam, A.; Raines, R. T. Methods Enzymol. 2009, 462, 25−44. (b) Bernardes, G. J. L.; Linderoth, L.; Doores, K. J.; Boutureira, O.; Davis, B. G. ChemBioChem 2011, 12, 1383−1386. (c) Wang, Z. A.; Kurra, Y.; Wang, X.; Zeng, Y.; Lee, Y.-J.; Sharma, V.; Lin, H.; Dai, S. Y.; Liu, W. R. Angew. Chem., Int. Ed. 2017, 56, 1643−1647.
(17) Charafeddine, A.; Dayoub, W.; Chapuis, H.; Strazewski, P. Chem. - Eur. J. 2007, 13, 5566−5584.
(18) (a) Kosiova, I.; Janicova, A.; Kois, P. Beilstein J. Org. Chem. 2006, 2, 23. (b) Aigner, M.; Hartl, M.; Fauster, K.; Steger, J.; Bister, K.; Micura, R. ChemBioChem 2011, 12, 47−51. (c) Yu, H.; Zheng, J.; Yang, S.; Asiri, A. M.; Alamry, K. A.; Sun, M.; Zhang, K.; Wang, S.; Yang, R. Talanta 2018, 178, 282−286.
(19) Weisbrod, S. H.; Marx, A. Chem. Commun. 2007, 1828−1830.
(20) Chemama, M.; Fonvielle, M.; Villet, R.; Arthur, M.; Valery, J. M.; Etheve-Quelquejeu, M. J. Am. Chem. Soc. 2007, 129, 12642−12643.
(21) Wang, Z. P. A.; Tian, C. L.; Zheng, J. S. RSC Adv. 2015, 5, 107192−107199.
(22) Nisic, F.; Speciale, G.; Bernardi, A. Chem. - Eur. J. 2012, 18, 6895−6906.
(23) Riomet, M.; Decuypere, E.; Porte, K.; Bernard, K.; Plougastel, L.; Kolodych, S.; Audisio, D.; Taran, F. Chem. - Eur. J. 2018, 24, 8535−8541.
(24) Riomet, M.; Porte, K.; Madegard, L.; Thuery, P.; Audisio, D.; Taran, F. Org. Lett. 2020, 22, 2403−2408.
(25) Ahad, A. M.; Jensen, S. M.; Jewett, J. C. Org. Lett. 2013, 15, 5060−5063.
(26) The use of a 2′-deoxynucleoside is recommended on the penultimate position to avoid breaking the phosphodiester bridge probably triggered by the formation of the oxazaphospholane (see Figure S68 in the Supporting Information).
(27) Fonvielle, M.; Mellal, D.; Patin, D.; Lecerf, M.; Blanot, D.; Bouhss, A.; Santarem, M.; Mengin-Lecreulx, D.; Sollogoub, M.; Arthur, M.; Etheve-Quelquejeu, M. Chem. - Eur. J. 2013, 19, 1357−1363.
(28) Nunez, M. G.; Farley, A. J.; Dixon, D. J. J. Am. Chem. Soc. 2013, 135, 16348−16351.

ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

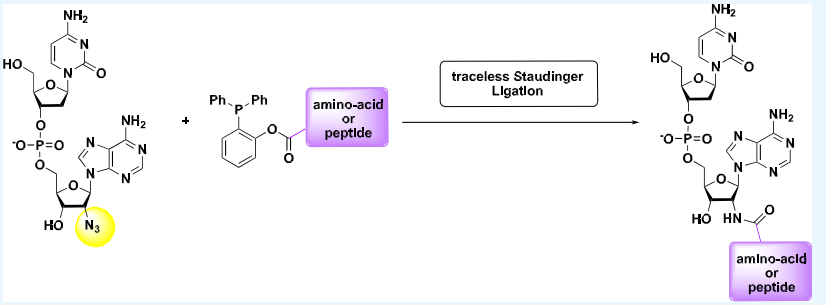
Traceless Staudinger Ligation to Access Stable Aminoacyl- or Peptidyl-Dinucleotide

Camélia Kitoun, Saidbakhrom Saidjalolov, Delphine Bouquet, Fabiola Djago, Quentin Blancart Remaury, Bruno Sargueil, Pauline Poinot, Mélanie Etheve-Quelquejeu,* and Laura Iannazzo*

Cite This: *ACS Omega* 2023, 8, 3850−3860

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Aminoacyl- and peptidyl-tRNA are specific biomolecules involved in many biological processes, from ribosomal protein synthesis to the synthesis of peptidoglycan precursors. Here, we report a post-synthetic approach based on traceless Staudinger ligation for the synthesis of a stable amide bond to access aminoacyl- or peptidyl-di-nucleotide. A series of amino-acid and peptide ester phenyl phosphines were synthetized, and their reactivity was studied on a 2′-N_3 di-nucleotide. The corresponding 2′-amide di-nucleotides were obtained and characterized by LC-HRMS, and mechanistic interpretations of the influence of the amino acid phenyl ester phosphine were proposed. We also demonstrated that enzymatic 5′-OH phosphorylation is compatible with the acylated di-nucleotide, allowing the possibility to access stable aminoacylated-tRNA.

INTRODUCTION

Amino acyl- and peptidyl-tRNAs are natural or synthetic[1] biomolecules composed of an RNA part covalently linked to an amino acid or a peptide. The most well-known natural RNAs containing an amino acid residue are the aminoacyl-tRNAs (aa-tRNA), in which the amino acid is connected at a 3′- or 2′-position of the terminal ribose by an ester link (Figure 1). Aa-tRNAs are used in ribosomal protein synthesis but also in non-ribosomal peptide synthesis (NRPS). This NRPS includes the synthesis of peptidoglycan precursors[2] or tetrapyrrole,[3] modification of bacterial membrane lipids,[4] and *N*-termina labeling of proteins targeted for proteolysis.[5] Aminoacyl-tRNAs also participate in secondary metabolism,[6] including the synthesis of cyclic dipeptides[7] and natural antibiotics.[8]

The exploration of the role of aminoacyl-tRNA in biologica processes requires stable molecules. Stable aa-tRNA analogues have been first designed by replacing the terminal ribose by a deoxy-ribose to avoid the isomerization of the amino acic moiety between the 2′- and 3′-hydroxyl of the ribose. The replacement of the ester bond by a stable function such as a triazole ring,[9] an oxadiazole,[10] or an amide bond[11] could also be used to stabilize the link between the amino acid and the RNA (Figure 1). These stable analogues have been efficiently used to obtain crystallographic structures of complexes between aa-tRNA analogues and the enzyme[12] or to decipher

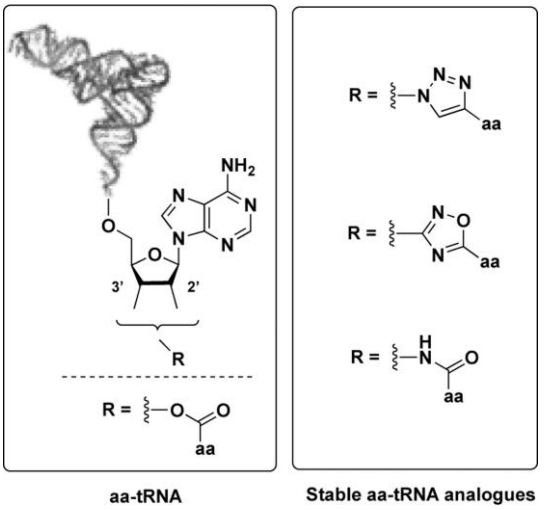


Figure 1. Structure of aa-tRNA and stable analogues.

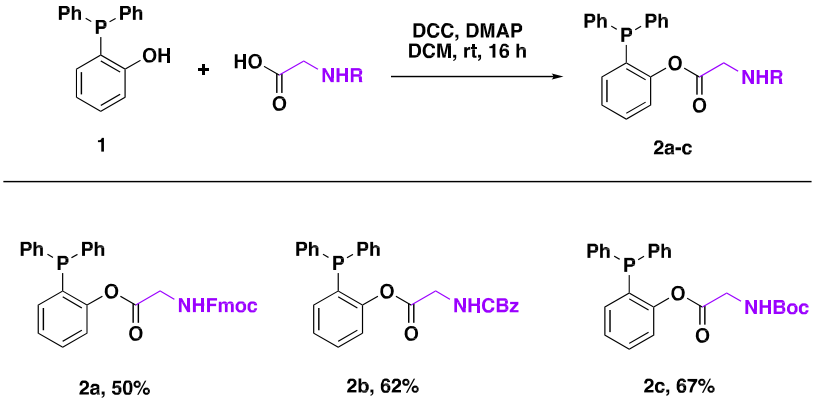
Received: September 22, 2022
Accepted: October 26, 2022
Published: January 17, 2023



© 2023 The Authors. Published by American Chemical Society

https://doi.org/10.1021/acsomega.2c06135
ACS Omega 2023, 8, 3850−3860

Scheme 1. Synthesis of *N*-Protected Glycine Phosphines



the specificity or regiospecificity of the transfer of the amino acid.[13]

For the synthesis of aminoacyl-RNA analogues containing an amide bond, two approaches have been used: (i) the pre-synthetic approach, which involves the formation of the amide bond on the nucleoside and then its incorporation into the RNA by solid phase synthesis or enzymatic ligation, or (ii) the post-synthetic approach, which involves the post-functionalization of an amino-RNA. The first approach was performed using standard peptide coupling[11,14] or Staudinger−Vilarrasa reaction[15] to link activated amino acids to amino- or azido-deoxyadenosine, respectively. This method requires relatively large amounts of modified nucleotide monomers obtained by multi-step syntheses. The second approach has been described in a single report, in which the direct coupling of an activated amino acid to an 18 nt 3′-amino-RNA has been efficiently used only for formyl-methionine.[11b] The challenge in the post-synthetic strategy is to find a chemoselective reaction, allowing site-specific acylation starting from a non-protected RNA, which contains several other nucleophilic groups. Here we propose to use the chemoselective traceless Staudinger ligation[16] in the presence of amino acids or peptide phosphines for the post-functionalization of an azido-di-nucleotide, which is a known precursor for the synthesis of aa-tRNA.[10,17]

The use of amino acids in the traceless variant of the Staudinger ligation has been first introduced for the synthesis of peptides in the presence of phosphinothiol by the work of Raines.[18] Thanks to this Staudinger peptide ligation (SPL) approach, two peptides have been efficiently ligated from glycine-azido derivatives.[19] The ligation with non-glycinic amino-acid turns out to be more complex since an aza-Wittig by-product could be observed. To overcome this issue, an effort has been made to design a novel phosphinothiol in which the phenyl groups have been substituted by an *O*-methoxy group at the *para* position.[20] Long[21] and cyclic[22] peptides were also synthesized by classical traceless Staudinger ligation using phosphino-ester or, more recently, with 2-(diphenylphosphanyl)phenyl)methane.[23] The group of Bernardi[24] applied such a methodology for the synthesis of amide-substituted glycosyl derivatives, while Guo and co-workers described the preparation of glycan-peptide conjugates.[25] Interestingly, all these examples mainly report the use of phosphines substituted with protected amino acids or peptides and no study has ever depicted the reactivity of ester phenyl phosphine carrying unprotected amino acids for traceless Staudinger ligation, which could be useful in a post-synthetic strategy.

RESULTS AND DISCUSSION

To explore the traceless Staudinger ligation to access aminoacyl- or peptidyl-RNA analogues, the dCpA-2′-N_3 di-nucleotide 4 containing an azido group at the 2′ position of the terminal adenosine was used. This dinucleotide has already been used in many studies to mimic the terminal CA motif present in all tRNAs, and the 2′-deoxy-cytidine does not impair the functionality of aa-tRNA.[10] The development of post-functionalization reaction conditions went through the preparation of phosphines substituted with an amino acid or a peptide as a second partner for the ligation step. The reactivity of phosphines was investigated, and a mechanistic interpretation was proposed to rationalize the experimental results.

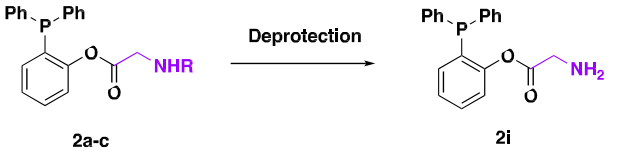
We first started our investigation with the synthesis of NH_2-protected amino acid phenyl ester phosphines with different protecting groups. For this purpose, the commercially available 2-hydroxyphenyl diphenylphosphine 1 was esterified in the presence of protected glycine, DCC, and DMAP in DCM at room temperature for 16 h. Glycine was selected as the model amino acid for our study. This reaction allowed the formation of three glycine phenyl ester phosphines 2a-c containing a variety of protecting groups including Fmoc, benzyloxycarbonyl (CBz), and Boc in 50, 62, and 67% yields, respectively (Scheme 1).

The phosphines 2a−c were then submitted to classical deprotection conditions. Under basic conditions (Scheme 2, entry 1), the degradation of the reaction mixture was observed, which led to the oxidation of phosphine and the cleavage of the ester linkage. Compound 2b was treated with a catalytic amount of Pd/C under hydrogen pressure, and degradation of only the starting material was observed (Scheme 2, entry 2). The synthesis of the NH_2-free glycine phenyl ester phosphine was finally obtained starting from compound 2c treated with a 15% solution of TFA in DCM at room temperature for 2 h. Compound 2i was then obtained in 86% yield as a trifluoroacetate salt (Scheme 2, entry 3). This study shows that the best option to protect the amine to synthesize amino acid-phosphine phenol ester is the *tert*-butoxycarbonyl (Boc).

To expand the scope of NH_2-unprotected amino-acid phosphine, a range of five amino-acids and peptides were selected. *N*-Boc-protected phosphines were obtained by coupling an alanine, a valine, and a phenylalanine in the presence of compound 1 using DCC and DMAP according to the route previously described for the synthesis of glycine phenyl ester phosphine (Scheme 1). The desired amino-acid phosphine esters 2d, 2e, and 2f were isolated in 85, 96, and

https://doi.org/10.1021/acsomega.2c06135
ACS Omega 2023, 8, 3850−3860

Scheme 2. Optimization of Deprotection Conditions



Entry	Phosphine	Conditions	2i
1	2a	Piperidine 20% DMF, 16h, rt	Degradation
2	2b	H_2, Pd/C THF, 16 h, rt	Degradation
3	2c	TFA 15% DCM, 2 h, rt	86%[a]

[a] compound 2i isolated in a trifluoroacetate salt.

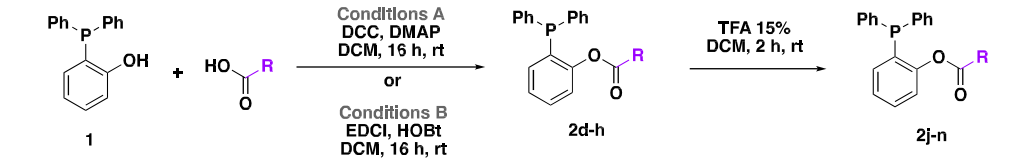
86% yields, respectively (Scheme 3A). Di- and tri-peptides were also engaged in a coupling reaction in the presence of the phenol phosphine 1, with EDCI and HOBt as coupling reagents (Conditions B, Scheme 3). The corresponding phosphines 2g and 2h with a di-glycine and tri-glycine peptide were obtained in 50 and 71% yields, respectively (Scheme 3A). The deprotection conditions using a solution of 15% TFA in DCM were then applied, and five unprotected amino-acid or peptide phenyl ester phosphines 2j−n were isolated in 32 to 80% yields (Scheme 3B).

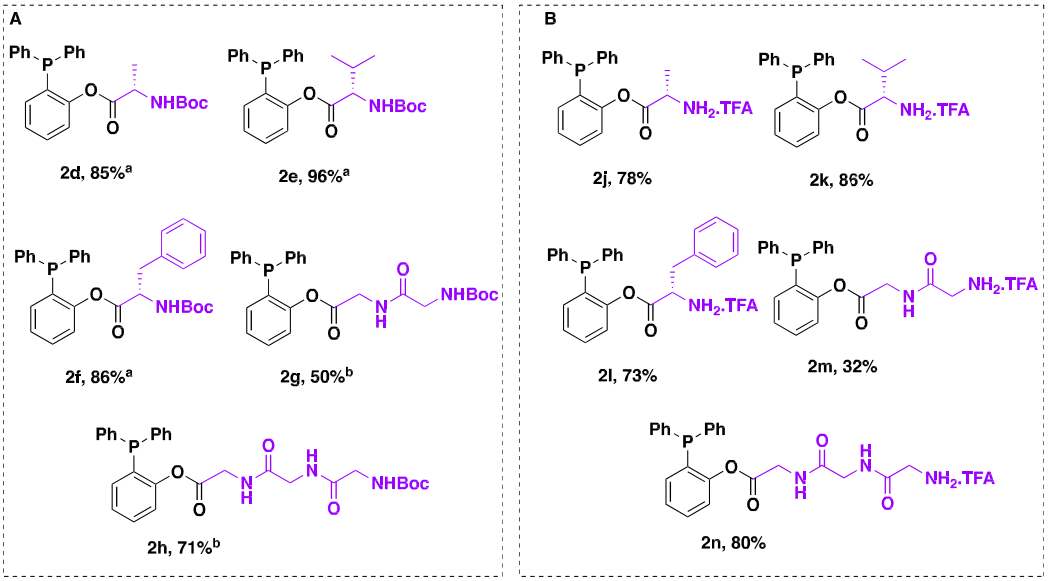
Among the series of previously synthesized phosphines, we next focused our attention on the reactivity of amino acid and peptide phosphines 2i, 2j, 2m, and 2n substituted with Gly and Ala groups. The reactions were carried out on 0.83 μmol of dinucleotide 4 and analyzed by LC-HRMS.

The dinucleotide substrate 4 was first treated in the presence of 10 equivalents of phosphine 2i and heated for 4 h at 70 °C, followed by the addition of water and 2 h of stirring at 70 °C to get a full conversion. Under these conditions, the formation of the ligated compound 5i was observed in 12% yield in competition with the reduced derivative 7 (Scheme 4, entry 1). Phosphine 2j and 2k lead to the same ratio in favor of the reduced dinucleotide 7 (Scheme 4, entries 2 and 3). When dipeptide 2m is used, no compound from the traceless Staudinger ligation was detected by LC-HRMS, and the reduction of dinucleotide 4 afforded compound 7. In addition, a cyclic oxazaphospholane derivative,[26] 6m, was obtained in 28% yield (Scheme 4, entry 4). The formation of such a cyclic derivative could be explained by the 1,2-*cis* position[27] of the 2′-N_3 function and the 3′-OH hydroxyl of the ribose. In the presence of phosphine 2n substituted with a glycine tripeptide, di-nucleotide 4 was converted into the expected ligated compound 2n in 10% yield (Scheme 4, entry 5). The access to the amide function was in competition with the formation of the oxazaphospholane derivative 6n observed in 47% yield and the reduction of the azide (compound 7) observed in 37% yield.

Based on the modest yield of traceless Staudinger ligation obtained with unprotected amino acid or peptide phosphines, we next examined the reactivity of *N*-Boc-protected phosphine derivatives. Traceless Staudinger ligation was then applied to di-nucleotide 4 in the presence of protected phosphines 2c, 2d,

Scheme 3. Synthesis of Phosphine Phenol Esters Containing (A) *N*-Boc-Protected Amine and (B) Unprotected Amine

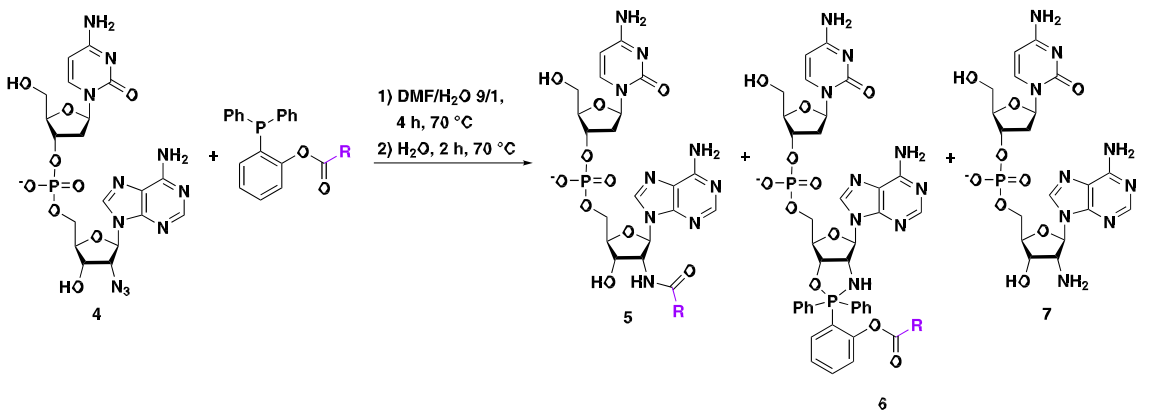


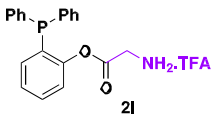
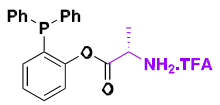
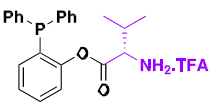
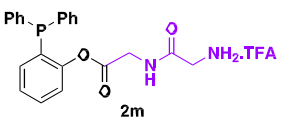
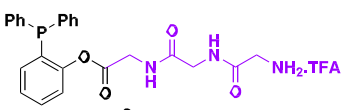
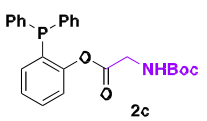
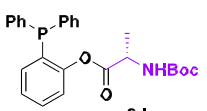
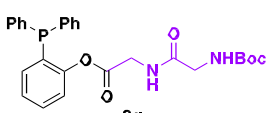
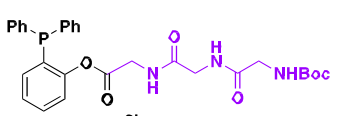


[a] Conditions A, [b] Conditions B

https://doi.org/10.1021/acsomega.2c06135
ACS Omega 2023, 8, 3850−3860

Scheme 4. Substrate Scope

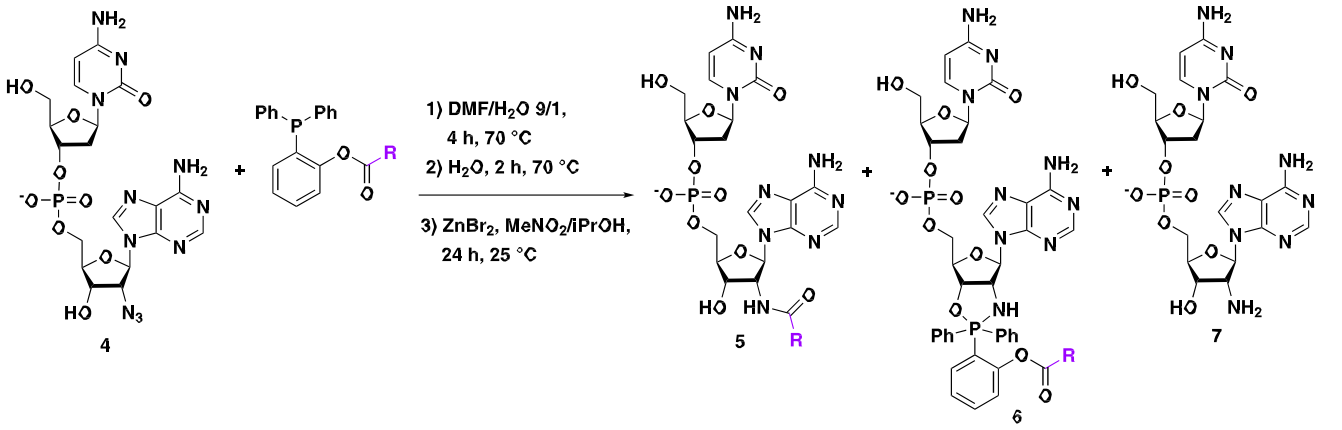


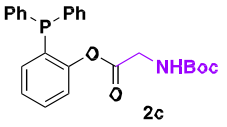
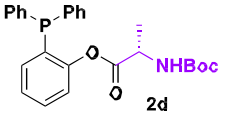
Entry	Phosphine	5[a]	6[a]	7[a]
1	 2l	5l: 12%	6l: -	88%
2	 2j	5j: 18%	6j: -	82%
3	 2k	5k: 4%	6k: 8%	88%
4	 2m	5m: -	6m: 28%	72%
5	 2n	5n: 10%	6n: 47%	37%
6	 2c	5c: 29%	6c: 63%	6%
7	 2d	5d: 35%	6d 46%	19%
8	 2g	5g: 13%	6g: 83%	4%
9	 2h	5h: 21%	6h: 48%	31%

[a] Product yield based on LC-HRMS data by relative quantification of dinucleotides.

https://doi.org/10.1021/acsomega.2c06135
ACS Omega 2023, 8, 3850−3860

Scheme 5. One-Pot Procedure



Entry	Phosphine	5[a]	6[a]	7[a]
1	 2c	5i: 13%	6l: 4%	83%
2	 2d	5j: 30%	6j: -	70%

[a] Product yield based on LC-HRMS data by relative quantification of dinucleotides.

2g, and 2h. By contrast, the protection of the amino group allowed an increase in the yield of the formation of the amide bond up to 35% for compound 2d. Although the reduced dinucleotide 7 was observed in the reaction mixture, the yield drastically decreased with protected phosphines (4 to 31% vs 37 to 88%). The formation of the oxazaphospholane derivatives was also detected with different yields (between 46 and 83% yields), depending on the phosphine.

These experimental results indicate that the use of *N*-Boc-protected phenyl ester phosphines impacts the yield of the reaction in favor of the amide derivative. We then reinvestigated our strategy and developed a three-step procedure to ligate a protected amino acid or peptide phosphine by traceless Staudinger ligation (Scheme 5). After the ligation step of dCpA-2′-N_3 di-nucleotide 4 in the presence of phosphines 2c and 2d at 70 °C for 4 h, followed by an additional 2 h of stirring at 70 °C in water, the crude mixture was treated with $ZnBr_2$ salts[28] for 24 h in a mixture of isopropanol and nitromethane. These conditions enabled the formation of the amide derivatives 5i and 5j in 13 and 30% yields, starting with phosphine 2c and 2d. This three-step procedure allows higher ligation yields than the use of unprotected amino acid or peptide phosphines.

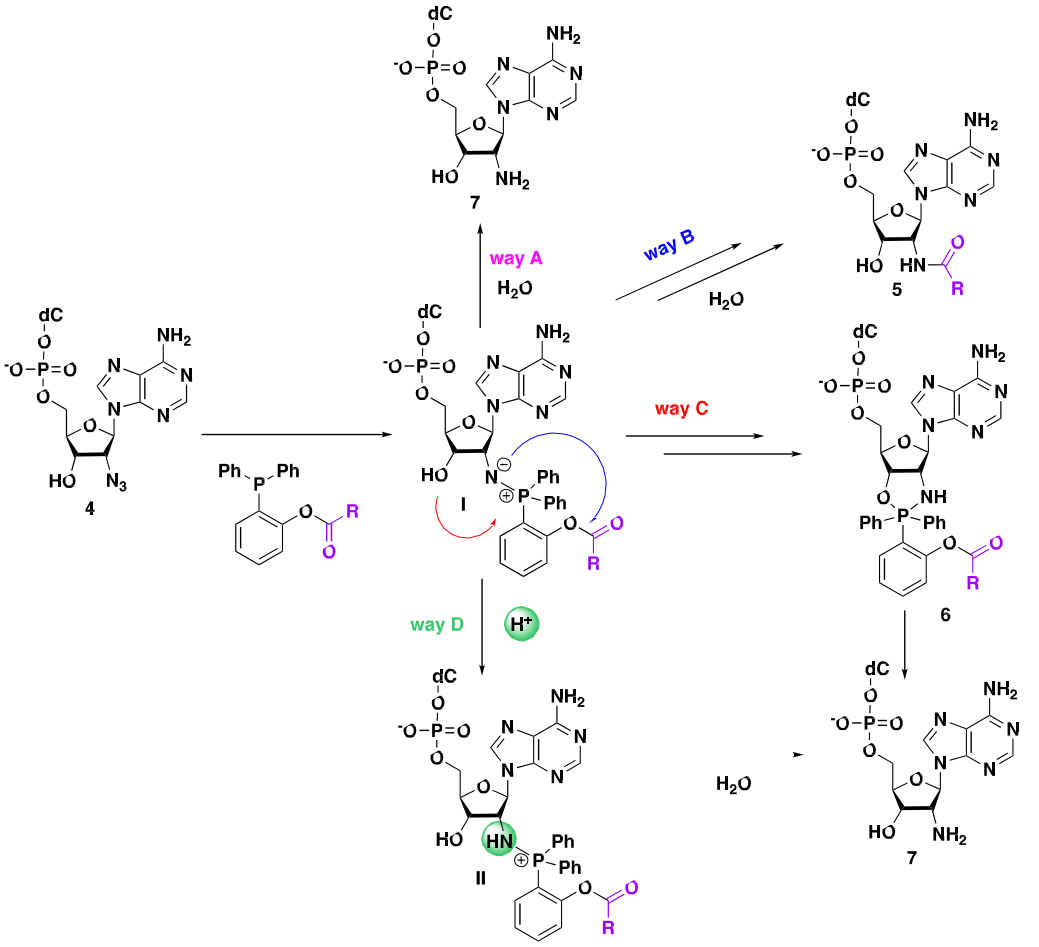
The results observed in this study show that the substituent on the amino group of the amino-acid or peptide linked to the ester phenyl phosphine has a strong impact on the ligation through a traceless Staudinger pathway. When the amino-acid is deprotected, the reduction of azide 4 is favored. In contrast, the yield increased in favor of the desired ligated derivative with *N*-Boc-protected phosphines, but an important amount of cyclic oxazaphospholane could be observed. To rationalize these results, a putative mechanism is proposed in Scheme 6.

After the formation of the iminophosphorane intermediate **I**, Staudinger reduction could lead directly to the formation of the reduced di-nucleotide 7 (way A). In a previous study,[26] we have demonstrated that the reaction occurred in competition with the acyl transfer from the phosphine to the nitrogen on the iminophosphorane through a six-membered ring formed during the transition state (way B) and the formation of the cyclic oxazaphospholane derivative promoted by the vicinal *cis* configuration of the nitrogen of the iminophosphorane and the hydroxyl at the 3′ position (way C). In the presence of *N*-Boc-protected phosphine, the nucleophilic attack of nitrogen on electrophilic carbonyl could be disadvantageous due to the steric hindrance of the Boc group and lead to the formation of compound 6. With NH_2-free amino acid or peptide phenyl ester phosphines, it offers an additional source of proton (coming from the TFA salt) that could easily protonate the nitrogen at the 3′-position, reported as a strong base,[29] affording intermediate **II**. Then, the hydrolysis with water could provide reduced compound 7.

The crude mixture obtained after traceless Staudinger ligation with phosphine 2j was then submitted to enzymatic 5′- phosphorylation using bacteriophage T4 polynucleotide kinase to yield dinucleotide 8 in 24% yield (Scheme 7). This was confirmed by LC-HRMS analyses (see Supporting Information S37-38). An additional enzymatic ligation step, starting from 8, could provide a series of stable aa-tRNAs since T4 RNA ligase is known to efficiently catalyze the ligation of the 3′-OH terminus of an RNA with a 5′-phosphorylated di-nucleotide.[11a]

https://doi.org/10.1021/acsomega.2c06135
ACS Omega 2023, 8, 3850−3860

Scheme 6. Proposed Mechanism



CONCLUSIONS

In conclusion, we have reported the first application of traceless Staudinger ligation to functionalize, in a post-synthetic approach, di-nucleotides with amino-acid or peptide phosphines. We synthesized a series of ester phenyl phosphines containing various amino acids or peptides. Boc was used as the protecting group for the amine since the deprotection step to remove other groups led to the degradation of the phosphine ester. Traceless Staudinger ligation was performed on a 2′-azido di-nucleotide in the presence of the protected and unprotected Gly and Ala phosphines. The use of unprotected phosphine ester provided the direct formation of the amino acid dinucleotide, although with moderate yields (12 and 18% yield). Inversely, the use of the protected amino phosphine ester allows us to obtain a better yield in favor of the amino acid dinucleotide (29 and 35%). We then perform a one-pot three-step procedure in the presence of *N*-BocAla and NBocGly phosphines to form two desired amino acylated di-nucleotides in 13 and 30% yields.

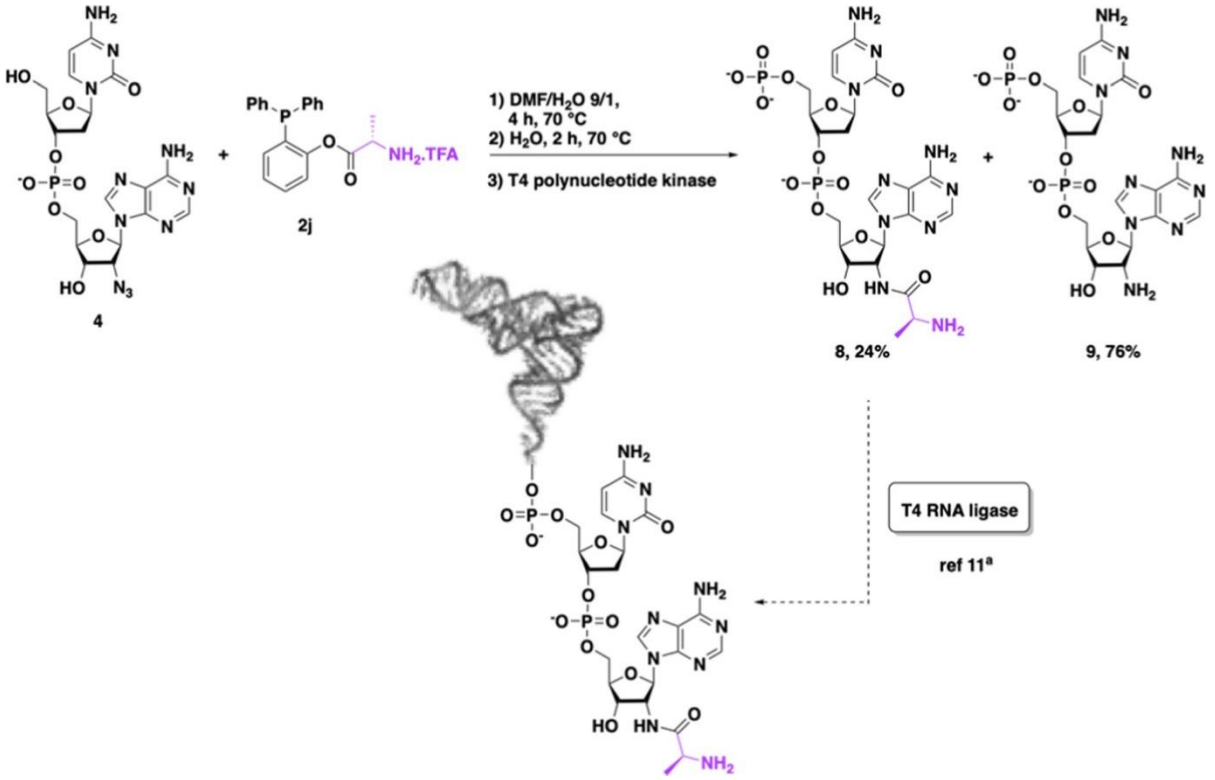
The post-synthetic traceless Staudinger ligation reported here is a versatile method allowing us to obtain a variety of stable amino acyl-tRNAs from a single acylated di-nucleotide. Of note, our synthetic route could also be of interest for studying RNAs containing non-canonical bases modified with amino acids.[30] Indeed, it has recently been suggested that such modifications of RNA could be living molecular fossils, and the group of T. Carell synthesized amino-acyl RNAs to support the idea of a molecular RNA-peptide world.[31]

EXPERIMENTAL SECTION

General Information. Reactions were carried out under an argon atmosphere and performed using freshly distilled solvents. THF was dried over sodium/benzophenone; DCM, DMF, and MeOH were dried over calcium hydride. Di-nucleotide 4 was synthesized according to the procedure previously reported by our group.[26] (2-Hydroxyphenyl) diphenylphosphine was purchased from Merck. The progress of the reactions was monitored by thin layer chromatography (TLC). TLC: precoated silica gel thin layer sheets 60 F_{254}(Merck) and detection by charring with 10% H_2SO_4 in ethanol or phosphomolybdic acid in ethanol, followed by heating. Flash chromatography was carried out over silica gel (60 Å, 180−240 mesh) from Merck. Preparative HPLC was performed using a HPLC system with a reverse phase C-18 NUCLEOSIL column (250 mm × 21.2 mm, 5 μm) using a solvent system consisting of A: 50 mM aqueous NH_4OAc pH 4.5 and B: MeCN (linear gradient from 0% B to 63% B in 30 min) at a flow rate of 15 mL/min and UV detection at 254 nm. NMR spectra were recorded on Bruker spectrometers (Avance II 500 and Avance III HD 4000). Chemical shifts (δ) are reported in parts per million (ppm) and referenced to the residual proton or carbon resonance of the solvents: $CDCl_3$ (δ 7.26), D_2O (δ 4.79), DMSO (δ 2.50), CD_3OD (δ 3.31) or D_2O (δ 4.79) for ^1H and $CDCl_3$ (δ 77.16) or $(CD_3)_2SO$ (δ 39.52), or CD_3OD (δ 49.00) for ^{13}C. Signals were assigned using 1D (^1H and ^{13}C) and 2D (HSQC, COSY, and HMBC) experiments. NMR coupling constants (J) are reported in

https://doi.org/10.1021/acsomega.2c06135
ACS Omega 2023, 8, 3850−3860

Scheme 7. Access to Stable Aa-tRNA by Enzymatic Phosphorylation and Ligation



Hertz (Hz), and splitting patterns are indicated as follows: s (singlet), bs (broad singlet), d (doublet), t (triplet), q (quartet), dd (doublet of doublet), and m (multiplet).

LC/HRMS Conditions: Conditions A. HPLC analysis was performed on a Hypersil GOLD aQ (ThermoFisher Scientific) column (100 mm × 2.1 mm, 1.9 μm) using a solvent system consisting of A: H_2O with 0.02% AcOH and B: MeCN with 0.02% AcOH (linear gradient from 0 to 100% B in 30 min) at a flow rate of 0.2 mL/min at 30 °C and UV detection at 254 nm. Fractions detected using low-resolution mass spectra (HRMS) were obtained on a LC QTOF mass spectrometer. Conditions B: LC-HRMS analyses were performed on a C Exactive mass spectrometer (Thermo Fisher Scientific) equipped with an electrospray ionization source (H-ESI II probe) coupled with an Accela 1250 Pump (Thermo Fisher Scientific). Samples were injected onto a Thermo Fisher Hypersil GOLD aQ chromatography column (100 mm × 2.1 mm, 1.9 μm particle size). The flow rate was set at 0.3 mL/min, and the mobile phase consisted of water containing 0.02% acetic acid (solvent A) and acetonitrile containing 0.02% acetic acid (solvent B). The gradient program was as follows: 0 to 5 min, 2% B; 5 to 20 min, up to 70% B; 20 to 30 min, came back to 98% B and 2 min of equilibration. The column temperature was maintained at 30 °C, and the temperature of the autosampler was set at 4 °C. MS analyses were performed in a full scan negative ion mode with a scan range from 200 to 1500 *m/z*.

General Procedure 1 for the Preparation of Phosphine 2a-f: Conditions A. To a solution of (2-hydroxyphenyl)diphenylphosphine (1 equiv) in DCM were successively added DCC (3 equiv), DMAP (4 equiv), and a *N*-protected amino acid (3 equiv). The mixture was stirred overnight at room temperature, then filtered through a pad of Celite, and concentrated under reduced pressure. The crude product was purified by flash chromatography using cyclohexane/EtOAc as an eluent to afford the desired phosphine.

General Procedure 2 for the Preparation of Phosphine 2g-h: Conditions B. To a solution of *N*-Boc di- or tri-peptide (1 equiv) in DCM were successively added EDCI (2 equiv) and HOBt (2 equiv). The mixture was stirred for 30 min at room temperature, and (2-hydroxyphenyl) diphenylphosphine (2 equiv) was added at 0 °C. The reaction mixture was then stirred at room temperature for 16 h. DCM and brine were added, and the organic phase was washed three times with brine. The organic layer was dried over $MgSO_4$, filtered, and concentrated. The residue was purified by silica gel chromatography using cyclohexane/EtOAc as the eluent to afford the desired phosphine.

Compound (2a). Following general procedure 1 for the preparation of amino acid-phosphine, compound 2a was obtained (150 mg, 50%) as a white foam after purification by flash chromatography using cyclohexane/EtOAc (7/3) as the eluent, starting from (2-hydroxyphenyl) diphenylphosphine (150 mg, 0.54 mmol) and *N*-Fmoc-Gly (480 mg, 1.62 mmol) in DCM (5 mL). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.70 (d, J = 7.5 Hz, 2 H, H^{Ar}), 7.53 (d, J = 7.3 Hz, 2 H, H^{Ar}), 7.34 (d, J = 7.3 Hz, 2 H, H^{Ar}), 7.28−7.24 (m, 13 H, H^{Ar}), 7.11−7.08 (m, 2 H, H^{Ar}), 6.83−6.80 (m, 1 H, H_{Ar}), 4.99 (br s, 1 H, NH), 4.34 (d, J = 7.0 Hz, 2 H, CH_2^{Fmoc}), 4.18−4.15 (m, 1 H, CH^{Fmoc}), 3.85−3.84 (m, 2 H, CH_2^{Gly}). ^{13}C NMR (126 MHz, $CDCl_3$): δ = 168.1 (C=O), 156.1 (C=O), 152.4, 143.9, 141.4, 135.4, 134.2, 134.0, 130.4, 130.2, 129.3, 128.8, 127.8, 127.2, 126.7, 125.2, 122.5, 120.1, 67.3 (CH_2^{Fmoc}), 47.3 (CH^{Fmoc}), 42.8 (CH_2^{Gly}). ^{31}P NMR (202 MHz, $CDCl_3$): δ = −15.7. HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{35}H_{29}NO_4P$ 558.18342; found, 558.18175.

Compound (2b). Following general procedure 1 for the preparation of amino acid-phosphine, compound 2b was

https://doi.org/10.1021/acsomega.2c06135
ACS Omega 2023, 8, 3850−3860

obtained (157 mg, 62%) as a white foam, after purification by flash chromatography using cyclohexane/EtOAc (8/2) as the eluent, starting from (2-hydroxyphenyl) diphenylphosphine (150 mg, 0.54 mmol) and *N*-Cbz-Gly (338 mg, 1.62 mmol) in DCM (10 mL). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.28−7.23 (m, 16 H, 5 H^{ArCbz}, 12 H^{ArPhos}), 7.08−7.05 (m, 2 H, H^{ArPhos}), 6.78−6.76 (m, 1 H, H^{ArPhos}), 5.02 (s, 2 H, CH_2^{Cbz}), 3.80−3.79 (m, 2 H, CH_2^{Gly}). ^{13}C NMR (126 MHz, $CDCl_3$): δ = 168.1 (C=O), 156.2 (C=O), 152.3, 136.3, 135.4, 135.3, 134.1 (2C), 133.9 (2C), 130.5, 130.3, 130.2, 129.3 (2C), 128.8 (2C), 128.7 (2C), 128.6 (2C), 128.5, 128.3, 128.2, 126.7, 122.5, 67.21 (CH_2^{Cbz}), 42.71 (CH_2^{Gly}). ^{31}P NMR (202 MHz, $CDCl_3$): δ = −15.8. HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{28}H_{25}NO_4P$ 470.15212; found, 470.15210.

Compound (2c). Following general procedure 1 for the preparation of amino acid-phosphine, compound 2c was obtained (262 mg, 67%) as a white foam, after purification by flash chromatography using cyclohexane/EtOAc (8/2) as the eluent, starting from (2-hydroxyphenyl) diphenylphosphine (250 mg, 0.89 mmol) and *N*-Boc-Gly (472 mg, 2.67 mmol) in DCM (10 mL). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.40−7.33 (m, 11 H, H^{Ar}), 7.20−7.16 (m, 2 H, H^{Ar}), 6.90−6.88 (m, 1 H, H^{Ar}), 4.70 (br s, 1 H, NH), 3.84 (d, J = 5.0 Hz, 2 H, CH_2^{Gly}), 1.49 (s, 9 H, tBu^{Boc}). ^{13}C NMR (126 MHz, $CDCl_3$): δ = 168.5, 155.6, 152.3, 135.5, 134.1, 134.0, 133.7, 133.5, 130.5, 130.4, 130.2, 129.3, 129.2, 128.8, 128.7, 126.6, 122.6, 80.1 (Cq^{Boc}), 42.4 (CH_2^{Gly}), 28.4 (tBu^{Boc}). ^{31}P NMR (202 MHz, $CDCl_3$): δ = −15.7. HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{25}H_{27}NO_4P$ 436.16777; found, 436.16634.

Compound (2d). Following general procedure 1 for the preparation of amino acid-phosphine, compound 2d was obtained (276 mg, 85%) as a white foam, after purification by flash chromatography using cyclohexane/EtOAc (8/2) as the eluent, starting from (2-hydroxyphenyl) diphenylphosphine (200 mg, 0.7 mmol) and *N*-Boc-Ala (407 mg,2.1 mmol) in DCM (10 mL). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.28−7.17 (m, 11 H, H^{Ar}), 7.09−7.06 (m, 1 H, H^{Ar}), 7.02 (t, J = 7.5 Hz, 1 H, H^{Ar}), 6.71−6.69 (m, 1 H, H^{Ar}), 4.82 (br s, 1 H, NH), 4.26 (br s, 1 H, Hα), 1.35 (s, 9 H, tBu^{Boc}), 1.16 (d, J = 5 Hz, 3 H, CH_3^{Ala}). ^{13}C NMR (126 MHz, $CDCl_3$): δ = 171.2, 152.6, 135.4, 134.0, 133.8, 133.5, 133.3, 131.9, 132.5, 130.0, 129.9, 129.1, 128.9, 128.6, 126.4, 122.2, 121.0, 116.1, 79.8 (Cq^{Boc}), 49.5 (CHα), 28.3 (tBu^{Boc}), 18.1 (CH_3^{Ala}). ^{31}P NMR (202 MHz, $CDCl_3$): δ = −16.8. HRMS (ESI) *m/z*: $[M]^+$ calcd for $C_{26}H_{28}NO_4P$ 449.1555; found, 449.1744.

Compound (2e). Following general procedure 1 for the preparation of amino acid-phosphine, compound 2e was obtained (329 mg, 96%) as a white foam, after purification by flash chromatography using cyclohexane/EtOAc (9/1) as the eluent, starting from (2-hydroxyphenyl) diphenylphosphine (200 mg, 0.7 mmol) and *N*-Boc-Val (390 mg, 2.1 mmol) in DCM (10 mL). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.39−7.32 (m, 7 H, H^{Ar}), 7.31−7.37 (m, 4 H, H^{Ar}), 7.19−7.17 (m, 1 H, H^{Ar}), 7.14 (t, J = 10 Hz, 1 H, H^{Ar}), 6.83−6.81 (m, 1 H H^{Ar}), 4.88 (d, J = 10 Hz, 1 H, NH), 4.29−4.27 (m, 1 H, Hα), 4.18−4.16 (m, 1 H, Hβ), 1.46 (s, 9 H, tBu^{Boc}), 0.94 (d, J = 5 Hz, 3 H, CH_3^{Val}), 0.82 (d, J = 5 Hz, 3 H, CH_3^{Val}). ^{13}C NMR (126 MHz, $CDCl_3$): δ = 170.4, 155.8, 153.1, 152.9, 135.9, 135.8, 135.7, 134.2, 134.1, 134.0, 130.3, 130.2, 129.3, 129.2, 128.9, 128.8, 126.5, 122.3, 79.8 (Cq^{Boc}), 58.8 (Cα), 31.1 (Cβ), 28.5 (tBu^{Boc}), 19.4 (CH_3^{Val}), 17.3 (CH_3^{Val}). ^{31}P NMR (202 MHz, $CDCl_3$): δ = −17.2. HRMS (ESI) *m/z*: $[M]^+$ calcd for $C_{28}H_{32}NO_4P$ 477.2069; found, 477.2057.

Compound (2f). Following general procedure 1 for the preparation of amino acid-phosphine, compound 2f was obtained (324 mg, 86%) as a white foam, after purification by flash chromatography using cyclohexane/EtOAc (9/1) as the eluent, starting from (2-hydroxyphenyl) diphenylphosphine (200 mg, 0.7 mmol) and *N*-Boc-Phe (551 mg, 2.1 mmol) in DCM (10 mL). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.41−7.31 (m, 11 H, H^{Ar}), 7.30−7.26 (m, 2 H, H^{Ar}), 7.24−7.20 (m, 1 H, H^{Ar}), 7.15−7.12 (m, 4 H, H^{Ar}), 6.84−6.81 (m, 1 H, H^{Ar}), 4.78 (br s, 1 H, NH), 4.52 (br s, 1 H, Hα), 3.14−3.10 (m, 1 H, $CH_2\beta$), 2.80−2.76 (m, 1 H, $CH_2\beta$), 1.39 (s, 9 H, tBu^{Boc}). ^{13}C NMR (126 MHz, $CDCl_3$): δ = 169.8, 155.1, 152.7, 136.6, 135.5, 134.2, 134.0, 133.9, 133.6, 133.4, 130.1, 129.6, 129.2, 129.1, 128.8, 128.6, 127.0, 126.5, 122.2, 121.7, 116.2, 80.0 (Cq^{Boc}), 54.7 (CHα), 31.9 (CH_2^{Phe}), 28.4 (tBu^{Boc}). ^{31}P NMR (202 MHz, $CDCl_3$): δ = −16.6. HRMS (ESI) *m/z* $[M]^+$ calcd for $C_{32}H_{32}NO_4P$ 525.2069; found, 525.2051.

Compound (2g). Following general procedure 2 for the preparation of amino acid-phosphine, compound 2g was obtained (260 mg, 50%) as a white foam, after purification by flash chromatography using cyclohexane/EtOAc (6/4) as the eluent, starting from (2-hydroxyphenyl) diphenylphosphine (300 mg, 1.0 mmol), EDCI (412 mg, 2.0 mmol), HOBt (290 mg, 2.0 mmol), and *N*-Boc-GlyGly (505 mg, 2.0 mmol) in DCM (10 mL). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.36−7.11 (m, 13 H, H^{Ar}), 6.80−6.74 (m, 1 H, H^{Ar}), 3.91−3.87 (m, 2 H, CH_2), 3.71−3.67 (m, 2 H, CH_2), 1.40 (s, 9 H, tBu^{Boc}). ^{13}C NMR (126 MHz, $CDCl_3$): δ = 171.5 (C=O), 1697.9 (C=O), 159.2, 156.9 (C=O), 152.8, 136.7, 135.5, 134.4, 133.6, 133.5, 133.4, 133.3, 129.9, 129.8, 128.8, 128.3, 128.2, 127.9, 126.1, 122.2, 119.4, 114.2, 79.3 (Cq^{Boc}), 40.4 (CH_2), 40.3 (CH_2), 27.2 (2 tBu^{Boc}), 24.6(tBu^{Boc}). ^{31}P NMR (202 MHz, $CDCl_3$): δ = −19.4 HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{27}H_{30}N_2O_5P$ 493.18923; found, 493.18743.

Compound (2h). Following general procedure 2 for the preparation of amino acid-phosphine, compound 2h was obtained (140 mg, 71%) as a white foam, after purification by flash chromatography using cyclohexane/EtOAc (3/7) as the eluent, starting from (2-hydroxyphenyl) diphenylphosphine (100 mg, 0.36 mmol), EDCI (138 mg, 0.72 mmol), HOBt (97 mg, 0.72 mmol), and *N*-Boc-GlyGlyGly (208 mg, 0.72 mmol) in DCM (10 mL). ^{1}H NMR (500 MHz, MeOD): δ = 7.36−7.27 (m, 7 H, H^{Ar}), 7.24−7.21 (m, 4 H, H^{Ar}), 7.14−7.11 (m, 2 H, H^{Ar}), 6.81−6.79 (m, 1 H, H^{Ar}), 3.89 (s, 2 H, CH_2^{Gly}), 3.86 (s, 2 H, CH_2^{Gly}), 3.71 (s, 2 H, CH_2^{Gly}), 1.41 (s, 9 H, tBu^{Boc}). ^{13}C NMR (126 MHz, MeOD): δ = 173.0, 122.0, 169.3, 158.3, 155.6, 138.2, 136.9, 136.8, 134.9, 134.7, 134.5, 131.7, 131.6, 131.4, 131.1, 130.2, 129.8, 129.7, 129.6, 129.4, 129.3, 127.5, 123.6, 120.8, 115.9, 80.9 (Cq^{Boc}), 43.1 (CH_2^{Gly}), 41.8 (CH_2^{Gly}), 41.7 (CH_2^{Gly}), 28.7 (tBu^{Boc}). ^{31}P NMR (202 MHz, MeOD): δ = −17.2. HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{29}H_{33}N_3O_6P$ 550.21070; found, 550.20908.

General Procedure 3 for the Deprotection of the Boc-Protecting Group. *N*-Boc amino acid/peptide-phosphine (1 equiv) was suspended in a 2 M TFA solution (11 equiv) in DCM, and the reaction was stirred for 2h at room temperature. The reaction mixture was concentrated under reduced pressure, and the residue was purified by flash chromatography (DCM/MeOH 9/1) to yield the corresponding TFA salt.

Compound (2i). Following general procedure 3 for the deprotection of *N*-Boc amino acid-phosphine, compound 2i was obtained (49 mg, 86%) as a white foam, after the removal

https://doi.org/10.1021/acsomega.2c06135
ACS Omega 2023, 8, 3850−3860

of TFA by co-evaporation (3 × toluene, 3 × DCM), starting from compound 2c (55 mg, 0.12 mmol) and TFA (0.69 mL, 1.38 mmol). ^{1}H NMR (500 MHz, MeOD): δ = 7.49−7.46 (m, 1 H, H^{Ar}), 7.39−7.38 (m, 6 H, H^{Ar}), 7.30−7.24 (m, 6 H, H^{Ar}), 6.91−6.89 (m, 1 H, H^{Ar}), 3.82 (s, 2 H, CH_2^{Gly}). ^{13}C NMR (126 MHz, MeOD): δ = 167.2, 153.7, 136.7, 135.3, 134.9, 134.7, 132.8, 131.6, 130.4, 130.1, 129.8, 129.6, 128.2, 123.3, 40.9 (CH_2^{Gly}). ^{31}P NMR (202 MHz, MeOD): δ = −17.6. ^{19}F NMR (471 MHz, MeOD): δ = −76.9. HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{20}H_{19}NO_2P$ 336.11479; found, 336.11480.

Compound (2j). Following general procedure 3 for the deprotection of *N*-Boc amino acid-phosphine, compound 2j was obtained (212 mg, 78%) as a white foam, after purification by flash chromatography using DCM/MeOH (9/1) as the eluent, starting from compound 2d (262 mg, 0.58 mmol) and TFA (3.2 mL, 6.38 mmol). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.27−7.20 (m, 7 H, H^{Ar}), 7.18−7.14 (m, 4 H, H^{Ar}), 7.06−7.01 (m, 2 H, H^{Ar}), 6.71−6.69 (m, 1 H, H^{Ar}), 3.89−3.82 (m, 1 H, Hα), 1.26 (d, J = 5 Hz, 3 H, CH_3^{Ala}). ^{13}C NMR (126 MHz, $CDCl_3$): δ = 168.5, 152.1, 152.0, 135.1, 135.0, 134.9, 134.0, 133.9, 130.4, 130.0, 129.9, 129.4, 128.9, 128.8, 122.1, 49.3 (Cα), 15.3 (CH_3^{Ala}). ^{31}P NMR (202 MHz, $CDCl_3$): δ = −17.1.^{19}F NMR (471 MHz, $CDCl_3$): δ = −75.8. HRMS (ESI) *m/z*: $[M]^+$ calcd for $C_{21}H_{20}NO_2P$ 349.1231; found 349.1230.

Compound (2k). Following general procedure 3 for the deprotection of *N*-Boc amino acid-phosphine, compound 2k was obtained (239 mg, 86%) as a white foam, after purification by flash chromatography using DCM/MeOH (96/4) as the eluent, starting from compound 2e (271 mg, 0.56 mmol) and TFA (3.1 mL, 6.16 mmol). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.39−7.32 (m, 7 H, H^{Ar}), 7.29−7.22 (m, 5 H, H^{Ar}), 6.84−6.82 (m, 1 H, H^{Ar}), 3.86 (br s, 1 H, Hα),2.30−2.27 (m, 1 H, Hβ), 1.00 (d, J = 5 Hz, 3 H, CH_3^{Val}), 0.92 (d, J = 5 Hz, 3 H, CH_3^{Val}). ^{13}C NMR (126 MHz, $CDCl_3$): δ = 167.8, 152.5, 152.4, 135.4, 135.3, 134.2, 134.1, 134.0, 133.9, 133.8, 130.4, 129.7, 129.3, 128.9, 127.0, 122.1, 58.7 (Cα), 29.4 (Cβ), 16.9 (CH_3^{Val}). ^{31}P NMR (202 MHz, $CDCl_3$): δ = −17.6. ^{19}F NMR (471 MHz, $CDCl_3$): δ = −75.6. HRMS (ESI) *m/z*: $[M]^+$ calcd for $C_{23}H_{24}NO_2P$ 377.1545; found, 377.1544.

Compound (2l). Following general procedure 3 for the deprotection of *N*-Boc amino acid-phosphine, compound 2l was obtained (170 mg, 73%) as a white foam, after purification by flash chromatography using DCM/MeOH (96/4) as the eluent, starting from compound 2f (227 mg, 0.43 mmol) and TFA (2.37 mL, 4.73 mmol). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.41−7.32 (m, 11 H, H^{Ar}), 7.31−7.24 (m, 3 H, H^{Ar}), 7.20−7.13 (m, 4 H, H^{Ar}), 6.85−6.83 (m, 1 H, H^{Ar}), 3.73−3.71 (m, 1 H, Hα), 3.11 (dd, J = 5, 10 Hz, 1 H, Hβ), 2.83−2.78 (m, 1 H, Hβ). ^{13}C NMR (126 MHz, $CDCl_3$): δ = 171.0, 159.2, 152.4, 136.0, 135.3, 134.6, 134.4, 134.0, 133.9, 133.6, 133.5, 131.5, 130.2, 129.6, 129.3, 129.0, 128.9, 128.8, 128.7, 127.3, 126.6, 122.2, 121.3, 121.0, 115.6, 55.4 (Cα), 38.6 (Cβ). ^{31}P NMR (202 MHz, $CDCl_3$): δ = −17.1.^{19}F NMR (471 MHz, $CDCl_3$): δ = −75.5. HRMS (ESI) *m/z*: $[M]^+$ calcd for $C_{27}H_{24}NO_2P$ 425.1544; found, 425.1551.

Compound (2m). Following general procedure 3 for the deprotection of *N*-Boc peptide-phosphine, compound 2m was obtained (20 mg, 32%) as a white foam, after purification by HPLC starting from compound 2g (60 mg, 0.12 mmol) and TFA (0.7 mL, 1.32 mmol). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.27−7.15 (m, 11 H, H^{Ar}), 7.01−6.96 (m, 2 H, H^{Ar}), 6.74−6.71 (m, 1 H, H^{Ar}), 3.74 (m, 2 H, CH_2), 3.6 (s, 2 H, CH_2). ^{13}C NMR (126 MHz, $CDCl_3$): δ = 168.2 (C=O), 167.0 (C=O), 152.4, 135.4, 135.3, 134.1, 133.9, 132.2, 132.1, 130.3, 130.4, 129.3 (2C), 128.9 (2C), 128.8 (2C), 126.7, 122.6, 41.2 (CH_2^{Gly}), 41.6 (CH_2^{Gly}). ^{31}P NMR (202 MHz, $CDCl_3$): δ = −16.2. HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{22}H_{22}N_2O_3P$ 393.13626; found, 393.13669.

Compound (2n). Following general procedure 3 for the deprotection of *N*-Boc peptide-phosphine, compound 2n was obtained (45 mg, 80%) as a white foam, after purification by HPLC, starting from compound 2h (60 mg, 0.1 mmol) and TFA (0.55 mL, 1.1 mmol). ^{1}H NMR (500 MHz, MeOD): δ = 7.44−7.16 (m, 13 H, H^{Ar}), 6.89−6.84 (m, 1 H, H^{Ar}), 3.96 (d, J = 5.25, 2 H; CH_2), 3.74 (s, 2 H, CH_2), 3.26 (s, 2H, CH_2). ^{13}C NMR (126 MHz, MeOD): δ = 171.6 (C=O), 169.4 (C=O), 168 (C=O), 154, 136.9, 136.8, 135.0 (2C), 134.8 (2C), 131.8, 131.7, 131.3, 130.3 (2C), 129.8 (2C), 129.7 (2C), 127.6, 122.6, 41.2 (CH_2^{Gly}), 41.0 (CH_2^{Gly}), 41.1 (CH_2^{Gly}), 31.7 NMR (202 MHz, MeOD): δ = −17.1. HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{24}H_{25}N_3O_4P$ 450.15772; found, 450.15811.

General Procedure 4 for Traceless Staudinger Ligation with Di-nucleotide 4. To a solution of di-nucleotide 4 (0.83 μmol,1 equiv) in an Eppendorf tube in DMF/H_2O (9/1, 1 mL) was added phosphine (10 equiv), and the mixture was stirred for 4 h at 70 °C in a thermo-shaker plate. Then, an excess of water (3mL) was added, and the mixture was stirred for 2 h at 70 °C. After lyophilization, the residue was analyzed by LC/HRMS.

General Procedure 5 for Traceless Staudinger Ligation with Di-nucleotide 4 with *N*-Boc-Protected Amino Acid Phosphine. To a solution of di-nucleotide 4 (0.5 mg, 0.83 μmol) in an Eppendorf tube in DMF/H_2O (9/1, 1 mL) was added phosphine 2c (3,7 mg, 8.3 μmol) or 2d (3,8 mg, 8.3 μmol), and the mixture was stirred for 4 h at 70 °C in a thermo-shaker plate. Then, an excess of water (3mL) was added, and the mixture was stirred for 2 h at 70 °C. After lyophilization, the residue was dissolved in a 5 M solution of $ZnBr_2$ in a 1/1 (v/v) mixture of iPrOH/$MeNO_2$ (69 μL, 344 μmol), and the reaction mixture was stirred at room temperature for 24 h. Water was added, and the mixture was lyophilized and analyzed by UPLC-HRMS.

General Procedure 6 for Traceless Staudinger Ligation and Enzymatic Phosphorylation. To a solution of dinucleotide 4 (1 mg, 1.67 μmol) in DMF/H_2O (9/1, 1 mL) in an Eppendorf tube was added phosphine (7 mg, 16.7 μmol), and the mixture was stirred for 4 h at 70 °C in a thermo-shaker plate (Grant Bio). Then, an excess of water (3 mL) was added, and the mixture was stirred for 2 h at 70 °C. After lyophilization, the residue was resuspended in 770 μL of water. The phosphorylation reaction was then carried out in a 1 mL reaction containing 70 mM Tris−HCl pH7.6, 10 mM $MgCl_2$, 5 mM DTT, 5 mM ATP, and 200 μL of T4 polynucleotide kinase (New England Biolabs ref #M0201L) for 3 h at 37 °C. T4 PNK was then inactivated for 20 min at 65 °C. The crude mixture was then analyzed by LC-HRMS.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c06135.

LC/HRMS analysis for compound 5c−d, 5g−k, 5m−n, 6c−d, 6g−h, 6m−n, 7, 8, and 9NMR spectra for compounds 2a−n FAIR data, including the primary NMR FID files (PDF)

AUTHOR INFORMATION

Corresponding Authors

Mélanie Etheve-Quelquejeu − *Université Paris Cité, CNRS, Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques, Paris F-75006, France;* orcid.org/0000-0002-4105-3243; Phone: (33) 1 76 53 42 50; Email: laura.iannazzo@u-paris.fr

Laura Iannazzo − *Université Paris Cité, CNRS, Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques, Paris F-75006, France;* orcid.org/0000-0002-2181-4885; Phone: (33) 1 76 53 42 22; Email: melanie.etheve-quelquejeu@u-paris.fr

Authors

Camélia Kitoun − *Université Paris Cité, CNRS, Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques, Paris F-75006, France*

Saidbakhrom Saidjalolov − *Université Paris Cité, CNRS, Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques, Paris F-75006, France*

Delphine Bouquet − *Université Paris Cité, CNRS, Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques, Paris F-75006, France*

Fabiola Djago − *Institut de Chimie des Milieux et Matériaux de Poitiers IC2MP, Université de Poitiers, UMR 7285, Poitiers 86073, France*

Quentin Blancart Remaury − *Institut de Chimie des Milieux et Matériaux de Poitiers IC2MP, Université de Poitiers, UMR 7285, Poitiers 86073, France*

Bruno Sargueil − *Université Paris Cité, CNRS, UMR 8038/ CiTCoM, Paris F-75006, France*

Pauline Poinot − *Institut de Chimie des Milieux et Matériaux de Poitiers IC2MP, Université de Poitiers, UMR 7285, Poitiers 86073, France;* orcid.org/0000-0002-7927-1887

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c06135

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was funded by grants from the Agence National de la Recherche (ANR), Project SyntRNA [ANR-17-CE07-0041-02], Project Volatolomix [ANR-21-CE18-0009], and La Ligue contre le cancer.

REFERENCES

(1) Klabenkova, K.; Fokina, A.; Stetsenko, D. Chemistry of peptide-oligonucleotide conjugates: a review. *Molecules* 2021, *26*, 5420.

(2) (a) Shepherd, J.; Ibba, M. Direction of aminoacylated transfer RNAs into antibiotic synthesis and peptidoglycan-mediated antibiotic resistance. *FEBS Lett.* 2013, *587*, 2895−2904. (b) Fonvielle, M.; Chemama, M.; Villet, R.; Lecerf, M.; Bouhss, A.; Valery, J. M.; Etheve-Quelquejeu, M.; Arthur, M. Aminoacyl-tRNA recognition by the FemXWv transferase for bacterial cell wall synthesis. *Nucleic Acids Res.* 2009, *37*, 1589−1601.

(3) Jahn, D.; Verkamp, E.; Söll, D. Glutamyl-transfer RNA: a precursor of heme and chlorophyll biosynthesis. *Trends Biochem. Sci.* 1992, *17*, 215−218.

(4) Roy, H.; Ibba, M. Broad range amino acid specificity of RNA-dependent lipid remodeling by multiple peptide resistance factors. *J. Biol. Chem.* 2009, *284*, 29677−29683.

(5) Mogk, A.; Schmidt, R.; Bukau, B. The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. *Trends Cell Biol* 2007, *17*, 165−172.

(6) (a) Katz, A.; Elgamal, S.; Rajkovic, A.; Ibba, M. Non-canonical roles of tRNAs and tRNA mimics in bacterial cell biology. *Mol. Microbiol.* 2016, *101*, 545−558. (b) Moutiez, M.; Belin, P.; Gondry, M. Aminoacyl-tRNA-Utilizing Enzymes in Natural Product Biosynthesis. *Chem. Rev.* 2017, *117*, 5578−5618.

(7) Gondry, M.; Sauguet, L.; Belin, P.; Thai, R.; Amouroux, R.; Tellier, C.; Tuphile, K.; Jacquet, M.; Braud, S.; Courçon, M.; Masson, C.; Dubois, S.; Lautru, S.; Lecoq, A.; Hashimoto, S.-i.; Genet, R.; Pernodet, J.-L. Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. *Nat. Chem. Biol.* 2009, *5*, 414−420.

(8) (a) Garg, R. P.; Qian, X. L.; Alemany, L. B.; Moran, S.; Parry, R. J. Inestigation of valanimycin biosynthesis: elucidation of the role of seryl-tRNA. *Proc. Natl. Acad. Sci. U.S.A.* 2008, *105*, 6543−6547. (b) Zhang, W.; Ntai, I.; Kelleher, N. L.; Walsh, C. T. tRNA-dependent peptide bond formation by the transferase PacB in biosynthesis of the pacidamycin group of pentapeptidyl nucleoside antibiotics. *Proc. Natl. Acad. Sci. U.S.A.* 2011, *108*, 12249−12253. (c) Bougioukou, D. J.; Mukherjee, S.; van der Donk, W. A. Revisiting the biosynthesis of dehydrophos reveals a tRNA-dependent pathway. *Proc. Natl. Acad. Sci. U.S.A.* 2013, *110*, 10952−10957. (d) Ortega, M. A.; Hao, Y.; Zhang, Q.; Walker, M. C.; van der Donk, W. A.; Nair, S. K. Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature* 2015, *517*, 506−512.

(9) Chemama, M.; Fonvielle, M.; Arthur, M.; Valéry, J. M.; Etheve-Quelquejeu, M. Synthesis of stable aminoacyl-tRNA analogues containing triazole as a bioisoster of esters. *Chem. Eur. J.* 2009, *15*, 1929−1938.

(10) Chemama, M.; Fonvielle, M.; Villet, R.; Arthur, M.; Valéry, J. M.; Etheve-Quelquejeu, M. Stable analogues of aminoacyl-tRNA for inhibition of an essential step of bacterial cell-wall synthesis. *J. Am. Chem. Soc.* 2007, *129*, 12642−12643.

(11) (a) Mellal, D.; Fonvielle, M.; Santarem, M.; Chemama, M.; Schneider, Y.; Iannazzo, L.; Braud, E.; Arthur, M.; Etheve-Quelquejeu, M. Synthesis and biological evaluation of non-isomerizable analogues of Ala-tRNA^{Ala}. *Org. Biomol. Chem.* 2013, *11*, 6161−6169. (b) Steger, J.; Graber, D.; Moroder, H.; Geiermann, A. S.; Aigner, M.; Micura, R. Efficient access to nonhydrolyzable initiator tRNA based on the synthesis of 3'-azido-3'-deoxyadenosine RNA. *Angew. Chem., Int. Ed.* 2010, *49*, 7470−7472.

(12) Fonvielle, M.; Li de La Sierra-Gallay, I.; El-Sagheer, A.; Lecerf, M.; Patin, D.; Mellal, D.; Mayer, C.; Blanot, D.; Gale, N.; Brown, T.; van Tilbeurgh, H.; Etheve-Quelquejeu, M.; Arthur, M. The structure of FemX$_{Wv}$ in complex with a peptidyl-RNA conjugate: mechanism of aminoacyl transfer from Ala-tRNA^{Ala} to peptidoglycan precursors. *Angew. Chem., Int. Ed.* 2013, *52*, 7278−7281.

(13) Fonvielle, M.; Chemama, M.; Lecerf, M.; Villet, R.; Busca, P.; Bouhss, A.; Etheve-Quelquejeu, M.; Arthur, M. Decoding the logic of the tRNA regiospecificity of nonribosomal FemX$_{Wv}$ aminoacyl transferase. *Angew. Chem., Int. Ed.* 2010, *49*, 5115−5119.

(14) Terenzi, S.; Biała, E.; Nguyen-Trung, N. Q.; Strazewski, P. Amphiphilic 3'-peptidyl-RNA conjugates. *Angew Chem Int Ed* 2003, *42*, 2909−2912.

(15) Moroder, H.; Steger, J.; Graber, D.; Fauster, K.; Trappl, K.; Trappl, V.; Marquez, N.; Polacek, D. N.; Wilson, R.; Micura, R. Non-hydrolyzable RNA-peptide conjugates : a powerful advance in the synthesis of mimics for 3'-peptidyl tRNA termini. *Angew. Chem., Int. Ed.* 2009, *48*, 4056−4060.

(16) Saxon, E.; Armstrong, J. I.; Bertozzi, C. R. A "tracless" Staudinger ligation for the chemoselective synthesis of amide bonds. *Org. Lett.* 2000, *2*, 2141−2143.

(17) (a) Robertson, S. A.; Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. The use of 5'-phospho-2-deoxyribicytidylriboadenosine as a facile route to chemical amino-acylation of tRNA. *Nucleic Acids Res.* 1989, *17*, 9649−9660. (b) Fahmi, N. E.; Dedkova, L.; Wang, B.; Golovine, S.; Hecht, S.

M. Site-specific incorporation of glycosylated serine and tyrosine derivatives into proteins. *J. Am. Chem. Soc.* 2007, *129*, 3586−3597.
(18) Nilsson, B. L.; Kiessling, L.; Raines, R. T. Staudinger ligation: a peptide from a thioester and azide. *Org. Lett.* 2000, *2*, 1939−1941.
(19) Nilsson, B. L.; Kiessling, L.; Raines, R. T. High-yielding Staudinger ligation of a phosphinothioester and azide to form a peptide. *Org. Lett.* 2001, *3*, 9−12.
(20) Tam, A.; Soellner, M. B.; Raines, R. T. Electronic and steric effects on the rate of the traceless Staudinger ligation. *Org. Biomol. Chem.* 2008, *6*, 1173−1175.
(21) Nilsson, B. L.; Hondal, J. R.; Soellner, M. B.; Raines, R. T Protein assembly by orthogonal chemical ligation methods. *J. Am. Chem. Soc.* 2003, *125*, 5268−5269.
(22) David, O.; Meester, W. J. N.; Bieräugel, H.; Schoemaker, H. E. Hiemstra, H.; van Maarseveen, J. H. Intramolecular Staudinger ligation: a powerful ring-closure method to form medium-sized lactams. *Angew. Chem., Int. Ed.* 2003, *42*, 4373−4375.
(23) Bajaj, K.; Pillai, G. G.; Sakhuja, R.; Kumar, D. Expansion of phosphane treasure box for Staudinger peptide ligation. *J. Org. Chem.* 2020, *85*, 12147−12159.
(24) (a) Bianchi, A.; Bernardi, A. Traceless Staudinger ligation of glycosyl azides with triaryl phosphines: stereoselective synthesis of glycosyl amides. *J. Org. Chem.* 2006, *71*, 4565−4577. (b) Nisic, F.; Andreini, M.; Bernardi, A. Stereoselective synthesis of N-glycosyl amino acids by traceless Staudinger ligation of unprotected glycosyl azides. *Eur. J. Org. Chem.* 2009, *2009*, 5744−5751.
(25) Zhu, S.; Guo, Z. Chemical synthesis of GPI glycan-peptide conjugates by traceless Staudinger ligation. *Org. Lett.* 2017, *19*, 3063−3066.
(26) (a) Kitoun, C.; Fonvielle, M.; Sakkas, F.; Lefresne, N.; Djago, M.; Blancart Remaury, Q.; Poinot, P.; Arthur, M.; Etheve-Quelquejeu, M.; Iannazzo, L. Phosphine-mediated bioconjugation of the 3'-end of RNA. *Org. Lett.* 2020, *22*, 8034−8038. (b) Kitoun, C.; Fonvielle, M.; Arthur, M.; Etheve-Quelquejeu, M.; Iannazzo, L. Traceless Staudinger ligation for bioconjugation of RNA. *Current Protocols in Nucleic Acid Chemistry* 2021, *1*, No. e42.
(27) Nisic, F.; Speciale, G.; Bernardi, A. Stereoselective synthesis of α- and β-glycofuranosyl amides by traceless ligation of glycofuranosyl azides. *Chem. Eur. J.* 2012, *18*, 6895−6906.
(28) (a) Kaul, K.; Brouillette, Y.; Sajjadi, Z.; Hansford, K. A.; Lubell, W. D. Selective *tert*-butyl ester deprotection in the presence of acid labile protecting groups with use of $ZnBr_2$. *J. Org. Chem.* 2004, *69*, 6131−6133. (b) Ferreira, F.; Morvan, F. Silyl protecting groups for oligonucleotide synthesis removed by a $ZnBr_2$ treatment. *Nucleosides Nucleotides and Nucleic Acids* 2005, *24*, 1009−1013.
(29) Núñez, M. G.; Farley, A. J.; Dixon, D. J. Bifunctional iminophosphorane organocatalysts for enantioselective synthesis application to the ketimine nitro-Mannich reaction. *J. Am. Chem Soc.* 2013, *135*, 16348−16351.
(30) Nainyte, M.; Müller, F.; Ganazzoli, G.; Chan, C. Y.; Crisp, A. Globisch, D.; Carell, T. Amino acid modified RNA bases as building blocks of an early earth RNA-peptide world. *Chem. Eur. J.* 2020, *26*, 14856−14860.
(31) Müller, F.; Escobar, L.; Xu, F.; Węgrzyn, A.; Nainyte, M. Amatov, T.; Chan, C. Y.; Pichler, A.; Carell, T. A prebiotically plausible scenario of an RNA-peptide world. *Nature* 2022, *605*, 279−284.

Recommended by ACS

Steric-Deficient Oligoglycine Surrogates Facilitate Multivalent and Bifunctional Nanobody Synthesis via Combined Sortase A Transpeptidation and Click Chemistry
Eugene M. Obeng, Kylie M. Wagstaff, et al.
AUGUST 03, 2023
BIOCONJUGATE CHEMISTRY
READ

Practical Synthesis of N-Formylmethionylated Peptidyl-tRNA Mimics
Julia Thaler, Ronald Micura, et al.
JULY 11, 2023
ACS CHEMICAL BIOLOGY
READ

Tyrosinase-Catalyzed Peptide Macrocyclization for mRNA Display
Matthew C. Fleming, Albert A. Bowers, et al.
MAY 08, 2023
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
READ

Chemically Acylated tRNAs are Functional in Zebrafish Embryos
Wes Brown, Alexander Deiters, et al.
JANUARY 20, 2023
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
READ

183

Get More Suggestions >

https://doi.org/10.1021/acsomega.2c06135
ACS Omega 2023, 8, 3850−3860

Conclusion générale

Conclusion générale

Les maladies cancéreuses sont multifactorielles et extrêmement complexes. Leur temps de genèse est souvent estimé à plusieurs décennies, nécessaires à l'accumulation de processus moléculaires et cellulaires défaillants.
L'amélioration et le développement d'approches multimodales est nécessaire afin d'envisager une caractérisation multidimensionnelle plus précise et une classification des tumeurs en fonction de leur phénotype, et de leur risque évolutif. Dans ce cadre, l'utilisation d'un pool de sondes enzymo-sensibles programmées pour cartographier, *ex vivo* et *in vivo*, des marqueurs enzymatiques ayant un rôle dans le développement de pathologies malignes, pourrait constituer une avancée majeure en cancérologie.

Nous avons donc élaboré des sondes à bases de COVs, qui, par hydrolyse enzymatique et analyse GC-MS/MS, permettait de détecter des marqueurs enzymatiques que sont les glycosidases. La première cible enzymatique ici étudiée, a été la β-glucuronidase, hydrolysant l'éthyl-β-D-glucuronide-D_5. De prime abord, la limite de détection du traceur volatil, l'éthanol-D5, associé à l'activité de l'enzyme a été déterminée. Puis, la limite de détection de l'enzyme *in vitro* et en matrice plasma a été évaluée. Tout ceci, pour pouvoir détecter l'enzyme en matrice plasma et biopsie, ce qui a été concluant. L'utilisation des sondes a permis de distinguer des échantillons de plasma sains de patients cancéreux. Par ailleurs, notre méthode a permis de détecter la β-glucuronidase sur des biopsies, et de distinguer la réponse de β-glucuronidase intracellulaire de l'extracellulaire. Par cette méthode nous sommes donc capables de différentier des prélèvements biologiques de sujets sains, par rapport à des sujets atteints de cancer.

La validation du concept sur la β-glucuronidase nous a permis de nous tourner vers d'autres cibles enzymatiques : la β-GlcNac, la β-Gal, la α-Fuc. Dans ce cadre, nous avons développé trois sondes à COVs supplémentaires spécifiques de chacune de ces enzymes, qui après activation libéraient un isotope de l'éthanol particulier (respectivement l'éthanol-$^{13}CD_5$, éthanol-D_2 et éthanol-D_4). L'injection à des modèles murins de ces sondes sous forme d'un cocktail a permis de discriminer les animaux sains des animaux porteurs d'une xénogreffe cancéreuse. Dans le cadre de cette étude, nous avons démontré que des sondes différentes étaient activées suivant le type de tumeur cancéreuse. Ceci est un résultat intéressant, car il prouve que nos cibles enzymatiques n'ont pas la même activité selon les cancers. Cela en vient à envisager à terme une cartographie de réponses enzymatiques, propres à chaque cancer. L'absence ou la présence d'activité enzymatique pourra être une piste pour déterminer le type de cancer.

Ainsi, la stratégie exploratoire que nous avons ici développée, fondée sur la simple détection d'isotopes de l'éthanol dans les voies aériennes ou au-dessus de prélèvements biologiques, pourrait constituer le premier test simple et rapide qui orienterait les patients vers une analyse approfondie par imagerie, ou un protocole thérapeutique personnalisé. En effet, les glycosidases ici ciblées peuvent être produites et libérées dans le milieu extracellulaire pour des tailles de tumeurs infracentimètriques (ex. la β-glucuronidase est présente en concentration importante dans le microenvironnement de la plupart des tumeurs solides dès lors qu'elles atteignent une taille de 2-3 mm de diamètre). Une telle sensibilité ne peut être atteinte avec les méthodes d'imagerie que sont la mammographie, la radiographie, l'imagerie par résonance magnétique et la tomographie par émission de positons. En outre, cette stratégie, simple,

peu coûteuse, indolore et accessible à un grand nombre, devrait permettre une surveillance clinique de grande envergure ce qui diminuerait considérablement la déprise médicale.
En revanche, les sondes à base de COVs ne fourniront qu'une information partielle de la distribution spatiale des tumeurs. Elles constitueront donc la première étape de diagnostic avant imagerie ou analyse de biomarqueurs sanguins spécifiques. Une détection multimodale et multi-échelle, couplée aux techniques immunologiques ou encore à des outils de traitement de données type machine ou data learning, devrait améliorer considérablement d'une part la précocité et d'autre par la spécificité du diagnostic, aidant ainsi le personnel médical dans sa prise de décision quant au choix des protocoles thérapeutiques.

En parallèle du ciblage des biomarqueurs par volatolomique induite, nous avons donc travaillé sur le développement d'un protocole de protéomique afin d'extraire et de caractériser nos cibles enzymatiques. Bien qu'il puisse être envisagé comme un protocole de confirmation de diagnostic, le but de ce développement était de séquencer les glycosidases trouvées dans les tissus cancéreux afin de lever le doute sur le phénotype cellulaire qui en est à l'origine. Ainsi une approche de protéomique middle-down a été réalisée via l'utilisation d'un agent d'hydrolyse chimique, le NTCB.
Cet agent permet l'hydrolyse des chaines protéiques au niveau des résidus cystéines, conduisant ainsi à des protéoformes de longueur significative. Après avoir optimisé le protocole d'hydrolyse au NTCB sur la BSA, il a fallu fractionner les peptides obtenus sur colonne SEC. Cette étape a été peu concluante malgré les différents essais.

Si les progrès en termes de diagnostic, pronostic et thérapie pour la lutte contre le cancer sont réels, il est vraisemblable que les succès seraient plus importants si l'on pouvait intervenir aux stades les plus précoces du développement tumoral. Ceci requiert de bien comprendre la biologie de l'état tumoral et de caractériser les facteurs qui influencent la transition vers le cancer ou qui, au contraire, conduisent à la régression de la lésion. Il est donc important de discriminer les lésions à fort potentiel malin des lésions qui ne vont pas évoluer voire disparaître.
Sur l'échelle d'évolution d'une lésion maligne, il apparait dans certains cas un développement métastatique qui lorsqu'enclenché diminue significativement les chances de guérison des patients. Les interactions entre les cellules tumorales - ou en transformation - avec leur microenvironnement sont déterminantes dans l'évolution des processus tumoral et métastatique. En effet, le microenvironnement est un écosystème complexe, spécifique du tissu et de l'organe dans lequel il se trouve. Il englobe les cellules adjacentes appartenant au même tissu et des fibroblastes, cellules vasculaires, cellules dendritiques, cellules du système immunitaire, le tout baignant dans une matrice extracellulaire (MEC), qui est un élément déterminant dans le maintien structurel de la tumeur et son agencement spatial. L'ensemble de ces structures et cellules sécrètent des enzymes (inclues les glycosidases) soit par voie directe soit par l'intermédiaire de microvésicules. Ces enzymes hydrolytiques influencent l'adhésion cellulaire et la reconnaissance cellule-cellule, ainsi que la signalisation intracellulaire et le remodelage de la MEC. Par ce biais, elles ajoutent une autre dimension de complexité à la progression du cancer.
Ainsi décrypter l'évolution de l'activité de ces enzymes via les sondes à COVs dans leur microenvironnement naturel à différents stades de progression tumoral pourrait permettre d'identifier les cancers de mauvais pronostic. Dans ce cadre, à l'instar des glycosidases, nous pourrions envisager de développer des sondes à base de COVs ciblant des estérases ou des oxydoréductases. Marqueurs

pronostiques et prédictifs, ces enzymes pourraient à terme devenir la cible de nouveaux agents chimio thérapeutiques administrés en amont pour prévenir la dissémination des cellules cancéreuses. In fine, ces traitements proposés au patient très précocement suivant son tableau clinique augmenterait de façon considérable ses chances de guérison et pourrait ainsi contribuer à la diminution du nombre de décès par an liés à cette maladie.

Références bibliographiques

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* **71**, 209–249 (2021).
2. Smith, R. *et al.* The randomized trials of breast cancer screening: What have we learned? *Radiologic clinics of North America* **42**, 793–806, v (2004).
3. Hazelton, W. D. & Luebeck, E. G. Biomarker-Based Early Cancer Detection: Is It Achievable? *Sci Transl Med* **3**, 109fs9 (2011).
4. Peixoto, A., Relvas-Santos, M., Azevedo, R., Santos, L. L. & Ferreira, J. A. Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks. *Frontiers in Oncology* **9**, (2019).
5. Lange, J. *et al.* Volatile Organic Compound Based Probe for Induced Volatolomics of Cancers. *Angewandte Chemie International Edition* **58**, 17563–17566 (2019).
6. Djago, F., Lange, J. & Poinot, P. Induced volatolomics of pathologies. *Nature Reviews Chemistry* **5**, 183–196 (2021).
7. Châtre, R. *et al.* Induced-volatolomics for the design of tumour activated therapy. *Chem. Sci.* (2023) doi:10.1039/D2SC06797H.
8. Marzorati, D. *et al.* A review of exhaled breath: a key role in lung cancer diagnosis. *J. Breath Res.* **13**, 034001 (2019).
9. Khan, H., Shah, M. R., Barek, J. & Malik, M. I. Cancer biomarkers and their biosensors: A comprehensive review. *TrAC Trends in Analytical Chemistry* **158**, 116813 (2023).
10. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209–249 (2021).
11. Pramesh, C. S. *et al.* Priorities for cancer research in low- and middle-income countries: a global perspective. *Nat Med* **28**, 649–657 (2022).
12. Hamdi, Y. *et al.* Cancer in Africa: The Untold Story. *Frontiers in Oncology* **11**, (2021).
13. Stefan, D. C. Cancer Care in Africa: An Overview of Resources. *J Glob Oncol* **1**, 30–36 (2015).
14. World Bank Open Data. *World Bank Open Data* https://data.worldbank.org.
15. Alkabban, F. M. & Ferguson, T. Breast Cancer. in *StatPearls* (StatPearls Publishing, 2023).
16. de Groot, P. M., Wu, C. C., Carter, B. W. & Munden, R. F. The epidemiology of lung cancer. *Transl Lung Cancer Res* **7**, 220–233 (2018).
17. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
18. Marx, V. Researchers dig into cancer niches. *Nat Methods* **20**, 484–488 (2023).
19. Anand, U. *et al.* Cancer chemotherapy and beyond: Current status, drug candidates, associated risks and progress in targeted therapeutics. *Genes & Diseases* (2022) doi:10.1016/j.gendis.2022.02.007.
20. Grunt, T. W. & Valent, P. Cancer – A devastating disease, but also an eye-opener and window into the deep mysteries of life and its origins. *Progress in Biophysics and Molecular Biology* **175**, 131–139 (2022).
21. You, J. S. & Jones, P. A. Cancer Genetics and Epigenetics: Two Sides of the Same Coin? *Cancer Cell* **22**, 9–20 (2012).
22. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discovery* **12**, 31–46 (2022).
23. Feinberg, A. P. Phenotypic plasticity and the epigenetics of human disease. *Nature* **447**, 433–440 (2007).
24. Jones, P. A., Issa, J.-P. J. & Baylin, S. Targeting the cancer epigenome for therapy. *Nat Rev Genet* **17**, 630–641 (2016).
25. Stowell, S. R., Ju, T. & Cummings, R. D. Protein glycosylation in cancer. *Annu Rev Pathol* **10**, 473–510 (2015).
26. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**, 245–254 (2003).
27. Kaufman, C. K. *et al.* A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. *Science* **351**, aad2197 (2016).
28. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115 (2013).

29. V, B. *et al.* Changes in DNA methylation patterns in subjects exposed to low-dose benzene. *Cancer research* **67**, (2007).
30. Shen, H. & Laird, P. W. Interplay between the cancer genome and epigenome. *Cell* **153**, 38–55 (2013).
31. Darwiche, N. Epigenetic mechanisms and the hallmarks of cancer: an intimate affair. *Am J Cancer Res* **10**, 1954–1978 (2020).
32. Taby, R. & Issa, J.-P. J. Cancer epigenetics. *CA Cancer J Clin* **60**, 376–392 (2010).
33. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17**, 487–500 (2016).
34. Ferreira, H. J. & Esteller, M. Non-coding RNAs, epigenetics, and cancer: tying it all together. *Cancer Metastasis Rev* **37**, 55–73 (2018).
35. Gutschner, T. & Diederichs, S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol* **9**, 703–719 (2012).
36. Fass, L. Imaging and cancer: A review. *Molecular Oncology* **2**, 115–152 (2008).
37. Tubiana, M. [Wilhelm Conrad Röntgen and the discovery of X-rays]. *Bull Acad Natl Med* **180**, 97–108 (1996).
38. Cancer du poumon, Cancer bronchique. https://www.santeweb.ch/Maladies/Cancer_du_poumon_Cancer_bronchique.html.
39. Choi, G. *et al.* Missed Lung Cancers on Chest Radiograph: An Illustrative Review of Common Blind Spots on Chest Radiograph with Emphasis on Various Radiologic Presentations of Lung Cancers. *J Korean Soc Radiol* **81**, 351 (2020).
40. Wu, M.-H. *et al.* Features of non-small cell lung carcinomas overlooked at digital chest radiography. *Clinical Radiology* **63**, 518–528 (2008).
41. Department of Radiology, Hacettepe University School of Medicine, Ankara, Turkey *et al.* Comparison of MRI and US in Tumor Size Evaluation of Breast Cancer Patients Receiving Neoadjuvant Chemotherapy. *Eur J Breast Health* **15**, 119–124 (2019).
42. Mitchell, D. G. *et al.* Early Invasive Cervical Cancer: Tumor Delineation by Magnetic Resonance Imaging, Computed Tomography, and Clinical Examination, Verified by Pathologic Results, in the ACRIN 6651/GOG 183 Intergroup Study. *JCO* **24**, 5687–5694 (2006).
43. Kaneko, M. *et al.* Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. *Radiology* **201**, 798–802 (1996).
44. Bronstein, Y. L. *et al.* Detection of Small Pancreatic Tumors with Multiphasic Helical CT. *American Journal of Roentgenology* **182**, 619–623 (2004).
45. Dudoignon, D., Pattison, D. A., Legallois, D., Hicks, R. J. & Aide, N. The utility of pharmacological and radiological interventions to optimize diagnostic information from PET/CT. *Cancer Imaging* **20**, 68 (2020).
46. Papandrianos, N., Papageorgiou, E., Anagnostis, A. & Papageorgiou, K. Efficient Bone Metastasis Diagnosis in Bone Scintigraphy Using a Fast Convolutional Neural Network Architecture. *Diagnostics* **10**, 532 (2020).
47. Paganelli, G. *et al.* Optimized sentinel node scintigraphy in breast cancer. *Q J Nucl Med* **42**, 49–53 (1998).
48. Yang, C., Wang, Q. & Ding, W. Recent progress in the imaging detection of enzyme activities in vivo. *RSC Adv.* **9**, 25285–25302 (2019).
49. Marom, E. M., Sarvis, S., Herndon, J. E. & Patz, E. F. T1 Lung Cancers: Sensitivity of Diagnosis with Fluorodeoxyglucose PET. *Radiology* **223**, 453–459 (2002).
50. Debela, D. T. *et al.* New approaches and procedures for cancer treatment: Current perspectives. *SAGE Open Med* **9**, 20503121211034370 (2021).
51. Surgery for Cancer - NCI. https://www.cancer.gov/about-cancer/treatment/types/surgery (2015).
52. Benjamin, D. J. The efficacy of surgical treatment of cancer – 20years later. *Medical Hypotheses* **82**, 412–420 (2014).
53. Veličković, K. *et al.* One-year recovery from breast cancer: Importance of tumor and treatment-related factors, resilience, and sociodemographic factors for health-related quality of life. *Frontiers in Oncology* **12**, (2022).

54. Wyld, L., Audisio, R. A. & Poston, G. J. The evolution of cancer surgery and future perspectives. *Nat Rev Clin Oncol* **12**, 115–124 (2015).
55. Cancer Statistics for the UK. *Cancer Research UK* https://www.cancerresearchuk.org/health-professional/cancer-statistics-for-the-uk (2015).
56. Philip, J. F. Results of treatment of breast cancer at Aberdeen Royal Infirmary 1940-55. *Br Med J* **1**, 323–331 (1967).
57. Baskar, R., Lee, K. A., Yeo, R. & Yeoh, K.-W. Cancer and radiation therapy: current advances and future directions. *Int J Med Sci* **9**, 193–199 (2012).
58. Delaney, G., Jacob, S., Featherstone, C. & Barton, M. The role of radiotherapy in cancer treatment: estimating optimal utilization from a review of evidence-based clinical guidelines. *Cancer* **104**, 1129–1137 (2005).
59. Begg, A. C., Stewart, F. A. & Vens, C. Strategies to improve radiotherapy with targeted drugs. *Nat Rev Cancer* **11**, 239–253 (2011).
60. Ringborg, U. *et al.* The Swedish Council on Technology Assessment in Health Care (SBU) systematic overview of radiotherapy for cancer including a prospective survey of radiotherapy practice in Sweden 2001--summary and conclusions. *Acta Oncol* **42**, 357–365 (2003).
61. Mehta, S. R., Suhag, V., Semwal, M. & Sharma, N. Radiotherapy: Basic Concepts and Recent Advances. *Med J Armed Forces India* **66**, 158–162 (2010).
62. Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071–1078 (2009).
63. Maani, E. V. & Maani, C. V. Radiation Therapy. in *StatPearls* (StatPearls Publishing, 2023).
64. Maier, P., Hartmann, L., Wenz, F. & Herskind, C. Cellular Pathways in Response to Ionizing Radiation and Their Targetability for Tumor Radiosensitization. *Int J Mol Sci* **17**, 102 (2016).
65. Sharda, N., Yang, C.-R., Kinsella, T. & Boothman, D. Radiation Resistance. in *Encyclopedia of Cancer (Second Edition)* (ed. Bertino, J. R.) 1–11 (Academic Press, 2002). doi:10.1016/B0-12-227555-1/00519-0.
66. Majeed, H. & Gupta, V. Adverse Effects Of Radiation Therapy. in *StatPearls* (StatPearls Publishing, 2023).
67. Amjad, M. T., Chidharla, A. & Kasi, A. Cancer Chemotherapy. in *StatPearls* (StatPearls Publishing, 2023).
68. Schirrmacher, V. From chemotherapy to biological therapy: A review of novel concepts to reduce the side effects of systemic cancer treatment (Review). *Int J Oncol* **54**, 407–419 (2018).
69. Moses, M. A., Brem, H. & Langer, R. Advancing the field of drug delivery: taking aim at cancer. *Cancer Cell* **4**, 337–341 (2003).
70. Global action plan for the prevention and control of NCDs 2013–2020. https://www.who.int/publications-detail-redirect/9789241506236.
71. Henry, N. L. & Hayes, D. F. Cancer biomarkers. *Mol Oncol* **6**, 140–146 (2012).
72. Atya, H. B., Ali, S. A., Hegazy, M. I. & El Sharkawi, F. Z. Urinary Urea, Uric Acid and Hippuric Acid as Potential Biomarkers in Multiple Sclerosis Patients. *Indian J Clin Biochem* **33**, 163–170 (2018).
73. Brunmair, J. *et al.* Finger sweat analysis enables short interval metabolic biomonitoring in humans. *Nat Commun* **12**, 5993 (2021).
74. *Essentials of Glycobiology*. (Cold Spring Harbor Laboratory Press, 2022).
75. Ohtsubo, K. & Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **126**, 855–867 (2006).
76. Cerqueira, N., Brs, N., Joo, M. & Alexandrino, P. Glycosidases – A Mechanistic Overview. in *Carbohydrates - Comprehensive Studies on Glycobiology and Glycotechnology* (ed. Chang, C.-F.) (InTech, 2012). doi:10.5772/52019.
77. Resibois, A., Tondeur, M., Mockel, S. & Dustin, P. Lysosomes and storage diseases. *Int Rev Exp Pathol* **9**, 93–149 (1970).
78. Costa, A. F., Campos, D., Reis, C. A. & Gomes, C. Targeting Glycosylation: A New Road for Cancer Drug Discovery. *Trends in Cancer* **6**, 757–766 (2020).

79. Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3–49 (2017).
80. Martins, Á. M., Ramos, C. C., Freitas, D. & Reis, C. A. Glycosylation of Cancer Extracellular Vesicles: Capture Strategies, Functional Roles and Potential Clinical Applications. *Cells* **10**, 109 (2021).
81. Roth, J. Protein N-Glycosylation along the Secretory Pathway: Relationship to Organelle Topography and Function, Protein Quality Control, and Cell Interactions. *Chem. Rev.* **102**, 285–304 (2002).
82. Tabas, I. & Kornfeld, S. Purification and characterization of a rat liver Golgi alpha-mannosidase capable of processing asparagine-linked oligosaccharides. *J Biol Chem* **254**, 11655–11663 (1979).
83. Hubbard, S. C. & Robbins, P. W. Synthesis and processing of protein-linked oligosaccharides in vivo. *J Biol Chem* **254**, 4568–4576 (1979).
84. Schachter, H. & Roseman, S. Mammalian Glycosyltransferases. in *The Biochemistry of Glycoproteins and Proteoglycans* (ed. Lennarz, W. J.) 85–160 (Springer US, 1980). doi:10.1007/978-1-4684-1006-8_3.
85. Adamczyk, B., Tharmalingam, T. & Rudd, P. M. Glycans as cancer biomarkers. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1820**, 1347–1353 (2012).
86. Gu, J. & Taniguchi, N. Potential of N-glycan in cell adhesion and migration as either a positive or negative regulator. *Cell Adh Migr* **2**, 243–245 (2008).
87. Cazet, A. *et al.* Consequences of the expression of sialylated antigens in breast cancer. *Carbohydr Res* **345**, 1377–1383 (2010).
88. Dube, D. H. & Bertozzi, C. R. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat Rev Drug Discov* **4**, 477–488 (2005).
89. Yan, L. *et al.* Lewis (y) Antigen Overexpression Increases the Expression of MMP-2 and MMP-9 and Invasion of Human Ovarian Cancer Cells. *Int J Mol Sci* **11**, 4441–4452 (2010).
90. Zhu, L.-C. *et al.* [Impact of alpha1,2-fucosyl transferase gene transfection on cancer-related gene expression profile of human ovarian cancer cell line RMG-1]. *Ai Zheng* **27**, 934–941 (2008).
91. Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97–130 (1993).
92. Bernacki, R. J., Niedbala, M. J. & Korytnyk, W. Glycosidases in cancer and invasion. *Cancer Metastasis Rev* **4**, 81–101 (1985).
93. Peiris, D. *et al.* Cellular glycosylation affects Herceptin binding and sensitivity of breast cancer cells to doxorubicin and growth factors. *Sci Rep* **7**, 43006 (2017).
94. Movahedin, M. *et al.* Glycosylation of MUC1 influences the binding of a therapeutic antibody by altering the conformational equilibrium of the antigen. *Glycobiology* **27**, 677–687 (2017).
95. Croci, D. O. *et al.* Glycosylation-dependent lectin-receptor interactions preserve angiogenesis in anti-VEGF refractory tumors. *Cell* **156**, 744–758 (2014).
96. Protein Electrophoresis Methods | Bio-Rad. https://www.bio-rad.com/fr-fr/applications-technologies/protein-electrophoresis-methods?ID=LUSOW4GRI.
97. Westermeier, R. *Electrophoresis in Practice: Guide to Methods and Applications of DNA and Protein Separations, A*. (Wiley-VCH Verlag GmbH & Co. KGaA, 2016). doi:10.1002/9783527695188.
98. Zhu, Z., Lu, J. J. & Liu, S. Protein separation by capillary gel electrophoresis: A review. *Analytica Chimica Acta* **709**, 21–31 (2012).
99. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**, 248–254 (1976).
100. Georgiou, C. D., Grintzalis, K., Zervoudakis, G. & Papapostolou, I. Mechanism of Coomassie brilliant blue G-250 binding to proteins: a hydrophobic assay for nanogram quantities of proteins. *Anal Bioanal Chem* **391**, 391–403 (2008).
101. Brunelle, E. *et al.* Coomassie Brilliant Blue G-250 Dye: An Application for Forensic Fingerprint Analysis. *Anal. Chem.* **89**, 4314–4319 (2017).
102. Brunelle, J. L. & Green, R. Chapter Thirteen - Coomassie Blue Staining. in *Methods in Enzymology* (ed. Lorsch, J.) vol. 541 161–167 (Academic Press, 2014).

103. Jiang, Y. *et al.* Modification in Silver Staining Procedure for Enhanced Protein Staining. *Biomed Res Int* **2022**, 6243971 (2022).
104. Hempelmann, E. & Krafts, K. The mechanism of silver staining of proteins separated by SDS polyacrylamide gel electrophoresis. *Biotech Histochem* **92**, 79–85 (2017).
105. Weiss, W., Weiland, F. & Görg, A. Protein Detection and Quantitation Technologies for Gel-Based Proteome Analysis. in *Proteomics: Methods and Protocols* (eds. Reinders, J. & Sickmann, A.) 59–82 (Humana Press, 2009). doi:10.1007/978-1-60761-157-8_4.
106. Chevallet, M., Luche, S. & Rabilloud, T. Silver staining of proteins in polyacrylamide gels. *Nat Protoc* **1**, 1852–1858 (2006).
107. Dyballa, N. & Metzger, S. Fast and Sensitive Coomassie Staining in Quantitative Proteomics. in *Quantitative Methods in Proteomics* (ed. Marcus, K.) 47–59 (Humana Press, 2012). doi:10.1007/978-1-61779-885-6_4.
108. Saraswathy, N. & Ramalingam, P. 11 - Two-dimensional gel electrophoresis of proteins. in *Concepts and Techniques in Genomics and Proteomics* (eds. Saraswathy, N. & Ramalingam, P.) 159–170 (Woodhead Publishing, 2011). doi:10.1533/9781908818058.159.
109. Smith, M. A., Bains, S. K., Betts, J. C., Choy, E. H. S. & Zanders, E. D. Use of Two-Dimensional Gel Electrophoresis To Measure Changes in Synovial Fluid Proteins from Patients with Rheumatoid Arthritis Treated with Antibody to CD4. *Clin Diagn Lab Immunol* **8**, 105–111 (2001).
110. Proteomics, C. 2D Electrophoresis Service. *Creative Proteomics* https://www.creative-proteomics.com/services/2d-electrophoresis-2.htm.
111. Mouzo, D., Bernal, J., López-Pedrouso, M., Franco, D. & Zapata, C. Advances in the Biology of Seed and Vegetative Storage Proteins Based on Two-Dimensional Electrophoresis Coupled to Mass Spectrometry. *Molecules* **23**, 2462 (2018).
112. Bernal, J. *et al. Identification and Mapping of Phosphorylated Isoforms of the Major Storage Protein of Potato Based on Two- Dimensional Electrophoresis*. *Advances in Seed Biology* (IntechOpen, 2017). doi:10.5772/intechopen.70400.
113. Zong, C. *et al.* Two-dimensional electrophoresis-based characterization of post-translational modifications of mammalian 20S proteasome complexes. *Proteomics* **8**, 5025–5037 (2008).
114. Stephens, A. N. *et al.* Post-Translational Modifications and Protein-Specific Isoforms in Endometriosis Revealed by 2D DIGE. *J. Proteome Res.* **9**, 2438–2449 (2010).
115. Baker, M. A., Witherdin, R., Hetherington, L., Cunningham-Smith, K. & Aitken, R. J. Identification of post-translational modifications that occur during sperm maturation using difference in two-dimensional gel electrophoresis. *Proteomics* **5**, 1003–1012 (2005).
116. Van Schepdael, A. Capillary electrophoresis as a simple and low-cost analytical tool for use in money-constrained situations. *TrAC Trends in Analytical Chemistry* 116992 (2023) doi:10.1016/j.trac.2023.116992.
117. Pillai-Kastoori, L., Schutz-Geschwender, A. R. & Harford, J. A. A systematic approach to quantitative Western blot analysis. *Analytical Biochemistry* **593**, 113608 (2020).
118. Engvall, E. & Perlmann, P. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry* **8**, 871–874 (1971).
119. Alhajj, M. & Farhana, A. Enzyme Linked Immunosorbent Assay. in *StatPearls* (StatPearls Publishing, 2023).
120. ELISA Guide - Creative Diagnostics. https://www.creative-diagnostics.com/ELISA-guide.htm.
121. Lindstrom, P. & Wager, O. IgG Autoantibody to Human Serum Albumin Studied by the ELISA-Technique. *Scand J Immunol* **7**, 419–425 (1978).
122. Aydin, S. A short history, principles, and types of ELISA, and our laboratory experience with peptide/protein analyses using ELISA. *Peptides* **72**, 4–15 (2015).
123. Shah, K. & Maghsoudlou, P. Enzyme-linked immunosorbent assay (ELISA): the basics. *Br J Hosp Med* **77**, C98–C101 (2016).

124. Singh, M. M. & Satija, J. Enzyme-assisted metal nanoparticles etching based plasmonic ELISA: Progress and insights. *Analytical Biochemistry* **654**, 114820 (2022).
125. Yang, X., Tang, Y., Alt, R. R., Xie, X. & Li, F. Emerging techniques for ultrasensitive protein analysis. *Analyst* **141**, 3473–3481 (2016).
126. Caswell, D. R. & Swanton, C. The role of tumour heterogeneity and clonal cooperativity in metastasis, immune evasion and clinical outcome. *BMC Medicine* **15**, 133 (2017).
127. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* **15**, 81–94 (2018).
128. Momenbeitollahi, N., Cloet, T. & Li, H. Pushing the detection limits: strategies towards highly sensitive optical-based protein detection. *Anal Bioanal Chem* **413**, 5995–6011 (2021).
129. Hu, R., Sou, K. & Takeoka, S. A rapid and highly sensitive biomarker detection platform based on a temperature-responsive liposome-linked immunosorbent assay. *Sci Rep* **10**, 18086 (2020).
130. Yang, Q. *et al.* Plasmonic ELISA for Sensitive Detection of Disease Biomarkers with a Smart Phone-Based Reader. *Nanoscale Res Lett* **13**, 397 (2018).
131. de la Rica, R. & Stevens, M. M. Plasmonic ELISA for the ultrasensitive detection of disease biomarkers with the naked eye. *Nat Nanotechnol* **7**, 821–824 (2012).
132. Mayer, K. M. & Hafner, J. H. Localized Surface Plasmon Resonance Sensors. *Chem. Rev.* **111**, 3828–3857 (2011).
133. Zeng, J., Duan, W., Li, M. & Xue, Y. Chapter 11 - Plasmonic Metallic Nanostructures as Colorimetric Probes for Environmental Pollutants. in *Novel Nanomaterials for Biomedical, Environmental and Energy Applications* (eds. Wang, X. & Chen, X.) 327–352 (Elsevier, 2019). doi:10.1016/B978-0-12-814497-8.00011-4.
134. Wagner, J. *et al.* Overexpression of the novel senescence marker β-galactosidase (GLB1) in prostate cancer predicts reduced PSA recurrence. *PLoS One* **10**, e0124366 (2015).
135. PCR | BioNinja. https://ib.bioninja.com.au/standard-level/topic-2-molecular-biology/27-dna-replication-transcri/pcr.html.
136. Fredriksson, S. *et al.* Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol* **20**, 473–477 (2002).
137. Lundberg, M., Eriksson, A., Tran, B., Assarsson, E. & Fredriksson, S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res* **39**, e102 (2011).
138. Zhang, H., Li, X.-F. & Le, X. C. Binding-Induced DNA Assembly and Its Application to Yoctomole Detection of Proteins. *Anal. Chem.* **84**, 877–884 (2012).
139. Kalli, M. *et al.* Development of a protein microarray-based diagnostic chip mimicking the skin prick test for allergy diagnosis. *Sci Rep* **10**, 18208 (2020).
140. Zhang, D. *et al.* Controlling the immobilization process of an optically enhanced protein microarray for highly reproducible immunoassay. *Nanoscale* **13**, 4269–4277 (2021).
141. Lee, H. J., Roh, Y. H., Kim, H. U., Kim, S. M. & Bong, K. W. Multiplexed immunoassay using post-synthesis functionalized hydrogel microparticles. *Lab Chip* **19**, 111–119 (2018).
142. Heo, Y. J., Hwa, C., Lee, G.-H., Park, J.-M. & An, J.-Y. Integrative Multi-Omics Approaches in Cancer Research: From Biological Networks to Clinical Subtypes. *Mol Cells* **44**, 433–443 (2021).
143. Chen, S., Xu, H., Guo, C., Liu, Z. & Han, X. Editorial: The role of multi-omics variants in tumor immunity and immunotherapy. *Front Immunol* **13**, 1098825 (2022).
144. El Kennani, S., Crespo, M., Govin, J. & Pflieger, D. Proteomic Analysis of Histone Variants and Their PTMs: Strategies and Pitfalls. *Proteomes* **6**, (2018).
145. Switzar, L., Giera, M. & Niessen, W. M. A. Protein Digestion: An Overview of the Available Techniques and Recent Developments. *J. Proteome Res.* **12**, 1067–1077 (2013).

146. Wiśniewski, J. R., Wegler, C. & Artursson, P. Multiple-Enzyme-Digestion Strategy Improves Accuracy and Sensitivity of Label- and Standard-Free Absolute Quantification to a Level That Is Achievable by Analysis with Stable Isotope-Labeled Standard Spiking. *J Proteome Res* **18**, 217–224 (2019).
147. Srzentić, K. *et al.* Chemical-Mediated Digestion: An Alternative Realm for Middle-down Proteomics? *J. Proteome Res.* **17**, 2005–2016 (2018).
148. Laskay, U. A. *et al.* Practical considerations for improving the productivity of mass spectrometry-based proteomics. *Chimia (Aarau)* **67**, 244–249 (2013).
149. Belghazi, M., Klett, D., Cahoreau, C. & Combarnous, Y. Nitro-thiocyanobenzoic acid (NTCB) reactivity of cysteines 100 and 110 in porcine luteinizing hormone: Metastability and hypothetical isomerization of the two disulfide bridges of its -subunit seatbelt. *Molecular and Cellular Endocrinology* 8 (2006).
150. Otieno, S. Generation of a free alpha-amino group by Raney nickel after 2-nitro-5-thiocyanobenzoic acid cleavage at cysteine residues: application to automated sequencing. *Biochemistry* **17**, 5468–5474 (1978).
151. Sanders, K. L. & Edwards, J. L. Nano-liquid chromatography-mass spectrometry and recent applications in omics investigations. *Anal Methods* **12**, 4404–4417 (2020).
152. Gaspari, M. & Cuda, G. Nano LC-MS/MS: a robust setup for proteomic analysis. *Methods Mol Biol* **790**, 115–126 (2011).
153. Spectrométrie de Masse Biologique et Protéomique : QqOrbitrap (QExactive and Qexactive HF, Thermo). https://www.smbp.espci.fr/Qexactive.html.
154. Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **72**, 1156–1162 (2000).
155. Hu, Q. *et al.* The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry* **40**, 430–443 (2005).
156. Makarov, A., Denisov, E. & Lange, O. Performance evaluation of a high-field orbitrap mass analyzer. *J. Am. Soc. Mass Spectrom.* **20**, 1391–1396 (2009).
157. Denisov, E., Damoc, E. & Makarov, A. Exploring frontiers of orbitrap performance for long transients. *International Journal of Mass Spectrometry* **466**, 116607 (2021).
158. Poojary, M. M. *et al.* Liquid chromatography quadrupole-Orbitrap mass spectrometry for the simultaneous analysis of advanced glycation end products and protein-derived cross-links in food and biological matrices. *J Chromatogr A* **1615**, 460767 (2020).
159. Mesmin, C. Application de la spectrométrie de masse à l'identification et la quantification de peptides biomarqueurs. (2012). doi:10.13140/2.1.3627.0405.
160. Feith, A., Teleki, A., Graf, M., Favilli, L. & Takors, R. HILIC-Enabled 13C Metabolomics Strategies: Comparing Quantitative Precision and Spectral Accuracy of QTOF High- and QQQ Low-Resolution Mass Spectrometry. *Metabolites* **9**, 63 (2019).
161. Competency, F. Review: Rapid Analysis of Drugs of Abuse and their Metabolites in Human Urine Using Dilute and Shoot Liquid Chromatography – Tandem Mass Spectrometry. *Florida Forensic Science* https://www.floridaforensicscience.com/review-rapid-analysis-drugs-abuse-metabolites-human-urine-using-dilute-shoot-liquid-chromatography-tandem-mass-spectrometry/ (2017).
162. Zeleke, G., De Baere, S., Suleman, S. & Devreese, M. Development and Validation of a Reliable UHPLC-MS/MS Method for Simultaneous Quantification of Macrocyclic Lactones in Bovine Plasma. *Molecules* **27**, 998 (2022).
163. Awolade, P. *et al.* Therapeutic significance of β-glucuronidase activity and its inhibitors: A review. *European Journal of Medicinal Chemistry* **187**, 111921 (2020).
164. Phueaouan, T. *et al.* Aberrant O-GlcNAc-modified proteins expressed in primary colorectal cancer. *Oncol Rep* **30**, 2929–2936 (2013).
165. Chaiyawat, P., Netsirisawan, P., Svasti, J. & Champattanachai, V. Aberrant O-GlcNAcylated Proteins: New Perspectives in Breast and Colorectal Cancer. *Frontiers in Endocrinology* **5**, (2014).
166. Asanuma, D. *et al.* Sensitive β-galactosidase-targeting fluorescence probe for visualizing small peritoneal metastatic tumours in vivo. *Nat Commun* **6**, 6463 (2015).

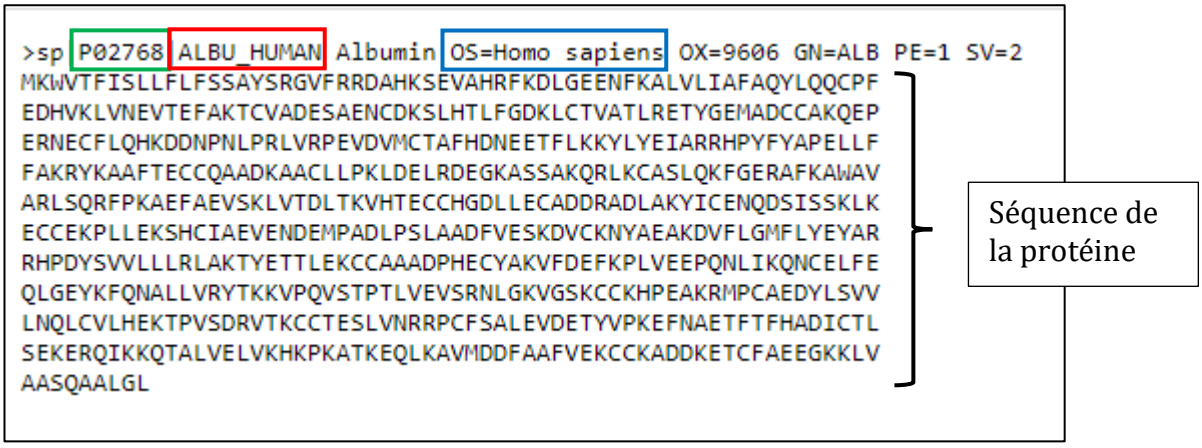
167. Deugnier, Y. *et al.* Serum α-L-Fucosidase: A New Marker for the Diagnosis of Primary Hepatic Carcinoma? *Hepatology* **4**, 889–892 (1984).
168. Shuang, Z. *et al.* Alpha-L-Fucosidase Serves as a Prognostic Indicator for Intrahepatic Cholangiocarcinoma and Inhibits Its Invasion Capacity. *BioMed Research International* **2018**, e8182575 (2018).
169. 5 Human Albumin. *Transfus Med Hemother* **36**, 399–407 (2009).
170. Zhang, L. *et al.* Sodium to globulin ratio as a prognostic factor for patients with advanced gastric cancer. *J Cancer* **11**, 7320–7328 (2020).
171. Tennent, G. A. *et al.* Human plasma fibrinogen is synthesized in the liver. *Blood* **109**, 1971–1974 (2007).
172. Crimmins, D. L., Mische, S. M. & Denslow, N. D. Chemical Cleavage of Proteins in Solution. *Current Protocols in Protein Science* **41**, (2005).
173. Tang, H.-Y. & Speicher, D. W. Identification of alternative products and optimization of 2-nitro-5-thiocyanatobenzoic acid cyanylation and cleavage at cysteine residues. *Analytical Biochemistry* **334**, 48–61 (2004).
174. Dalton, W. S. & Friend, S. H. Cancer Biomarkers—An Invitation to the Table. *Science* **312**, 1165–1168 (2006).
175. Li, Y., Wang, Z. & Cirino, P. C. Design and characterization of new β-glucuronidase active site variants with altered substrate specificity. *Biotechnol Lett* **40**, 111–118 (2018).
176. Little, M. S., Pellock, S. J., Walton, W. G., Tripathy, A. & Redinbo, M. R. Structural basis for the regulation of β-glucuronidase expression by human gut Enterobacteriaceae. 10.
177. Rusconi, F. *Manuel de Spectrométrie de masse à l'usage des biochimistes*. (2011).
178. Strohalm, M., Hassman, M., Košata, B. & Kodíček, M. mMass data miner: an open source alternative for mass spectrometric data analysis. *Rapid Communications in Mass Spectrometry* **22**, 905–908 (2008).
179. Strohalm, M., Kavan, D., Novák, P., Volný, M. & Havlíček, V. mMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data. *Anal. Chem.* **82**, 4648–4651 (2010).
180. Niedermeyer, T. H. J. & Strohalm, M. mMass as a Software Tool for the Annotation of Cyclic Peptide Tandem Mass Spectra. *PLOS ONE* **7**, e44913 (2012).
181. Koehn, H. *et al.* Higher sequence coverage and improved confidence in the identification of cysteine-rich proteins from the wool cuticle using combined chemical and enzymatic digestion. *Journal of Proteomics* **73**, 323–330 (2009).
182. The UniProt Consortium. The Universal Protein Resource (UniProt). *Nucleic Acids Research* **35**, D193–D197 (2007).
183. Comprehensive Analysis of Protein Glycation Reveals Its Potential Impacts on Protein Degradation and Gene Expression in Human Cells | SpringerLink. https://link.springer.com/article/10.1007/s13361-019-02197-4.

Annexes

I. Recherche et présentation des logiciels de digestion *in sllico*

- Étape 1 : Charger la séquence de la protéine en format FASTA (FAST-All)

Toutes les données relatives aux sciences omiques étant répertoriées dans des bases de données internationales, en protéomique, nous utilisons la base de données Uniprot (Universal Protein Resource)[182]. Cette base de données Suisse crée en] recense les séquences de toutes les protéines répertoriées au monde. Pour obtenir une séquence, après avoir rentre le nom ou le code d'une protéine, on obtient un fichier FASTA (Annexe 89). Ce fichier est indispensable car il est compatible avec tous les logiciels de retraitement en protéomique. Le fichier FATSA permet donc de télécharger la s séquence de la protéine issue d'une base de données. Selon les logiciels, avant de prendre le FASTA, on choisit banque de données dont la protéine fait partie, la taxonomie de l'organisme possédant la protéine.



Annexe 89: Fichier FASTA de la séquence d'albumine humaine

Carré vert : code Uniprot de la protéine. **Carré rouge :** nom abrégé de la protéine. **Carré bleu :** Organisme de provenance de la protéine.

- Étape 2 : Choisir les différents paramètres de la digestion

Différents paramètres de digestion doivent être choisis comme la nature de l'enzyme utilisée ou de l'agent chimique de digestion. Il faut aussi renseigner la taille maximale de peptides générés, ainsi que le nombre « miscleavage ». Ce terme correspond au nombre de coupure inattendue qui va se faire pendant la digestion. Il est de 0 avec la trypsine car fait très peu d'erreur. Pour d'autres digestion, quand bien même le nombre d'erreur est plus important, il est préférable de fixer le miscleavage à 0, afin d'avoir une répétabilité des sites de coupures attendus.

Si l'identification des peptides est fastidieuse, on peut tolérer 1 miscleavage par peptides. Dans certains logiciels ou site web, il faut en plus de préciser l'agent de digestion, son site de coupure. Certains associent directement l'enzyme ou l'agent chimique au site de coupure et ses caractéristiques, à savoir, le site de la coupure (choix du ou des aides aminés), et son sens N-ter (N-terminal) ou C-ter(C-terminale).

- Étape 3 : Renseigner les modifications

Il faut savoir que chaque coupure de digestion génère des modifications chimiques à la suite de la réaction d'hydrolyse de la protéine en peptides. Il est donc possible de les renseigner sur les plateformes de digestion *in silico*.

Dans le cadre de la trypsine qui coupe sur la lysine et l'arginine, la modification qui est faite de manière systématique est l'ajout de la fonction carbamidométhyl sur les peptides. Dans la nomenclature unimod, cette modification est nommée (C). À l'origine, cette modification fixe est conséquente à l'addition nucléophile de l'IAA sur le thiol du résidu cystéine du peptide. Le peptide est ainsi modifié avec le groupement carbamidométhyl. Le groupement rajoute 57.0214 Da à sa masse.[183] En plus de la modification fixe lors de la digestion trypsique, il y a aussi une modification variable qui est l'oxydation de la méthionine. (+15.999 Da). Ces deux types de modifications sont essentiels à renseigner dans les logiciels de digestion *in silico*.
Dans le cadre de l'hydrolyse chimique au NTCB, deux modifications sont possibles. La modification concourente à l'addition nucléophile du NTCB en N-ter sur le thiol du résidu cystéine, ajoutant un groupement ITZ(Iminothiazolidine). Ceci rajoute 24.99 Da à la masse du peptide. La deuxième modification est la dehydroalanine (Dha), qui retire 33.98 Da au peptide.

- Étape 4 : Choisir les paramètres de masse

Dans cette étape, il s'agit de renseigner les informations sur le type de spectromètre de masse utilisé et ses caractéristiques comme le type d'analyseur, mode d'ionisation electrospray (ESI) de type ESI+/ESI-, le nombre maximal de charge autorisé lors de l'ionisation des peptides. Dans notre étude, nous avons utilisé principalement le LC-Orbitrap, pour sa haute résolution et ses data compatibles avec les logiciels de retraitement.

- Etape 5 : Générer la liste des peptides issus de la digestion

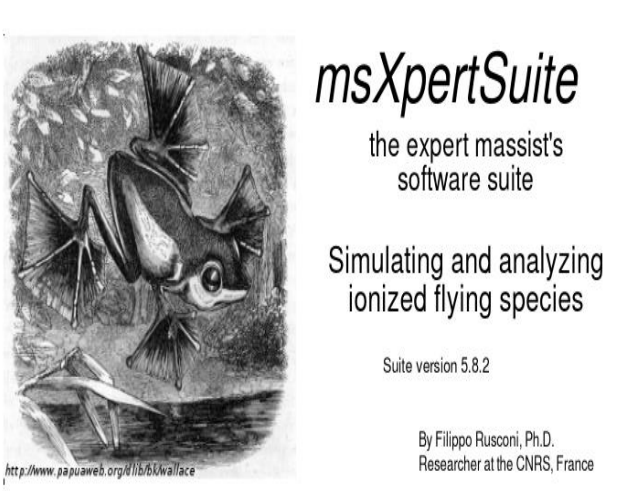
A la fin de la manœuvre, la liste des peptides obtenus et leur rapport masse sur charge (m/z) ainsi que leur masse moyenne et monoisotopique sont renseignées. A partir de ça on peut comparer les données théoriques aux données expérimentales brutes. Ceci peut se faire manuellement ou avec des logiciels de retraitement.

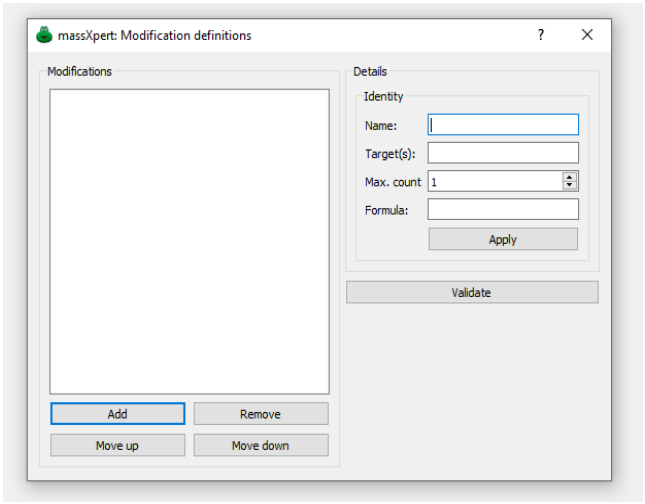
Logiciels et sites usuels de digestion *in silico*

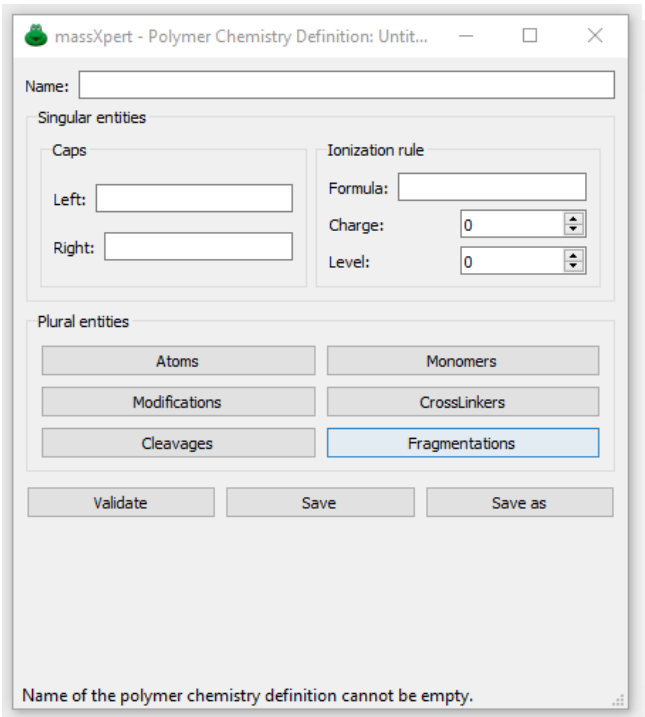
Type de traitement	Logiciels et sites de protéomique
Production de données théoriques	MS digest Expasy
Traitement des données brutes	Skyline MaxQuant MassXpert mMass

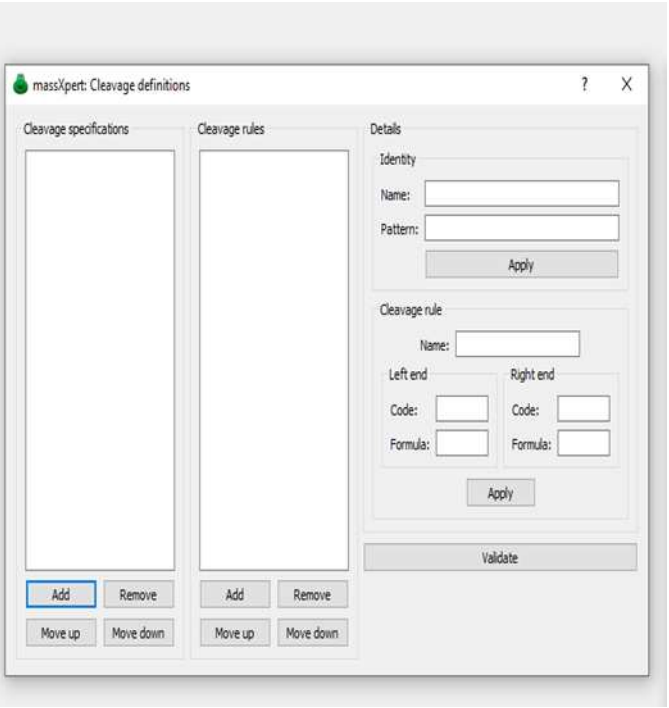
Annexe 90:Liste des différents types de logiciels utilisés en protéomique1

Les logiciels répertoriés sont les logiciels le plus utilisés en protéomique *bottom up*. Ils s'utilisent en suivant les mêmes étapes que celles citées précédemment et permettent d'avoir des résultats fiables et exploitables. Parmi eux, il y a ceux qui font uniquement la production des données théoriques sans exploiter les données brutes, et ceux qui traitent les données brutes peptides. Ces derniers font donc toute l'exploitation des résultats. C'est un avantage car il n'y a plus besoin de comparer les données théoriques et expérimentales manuellement.







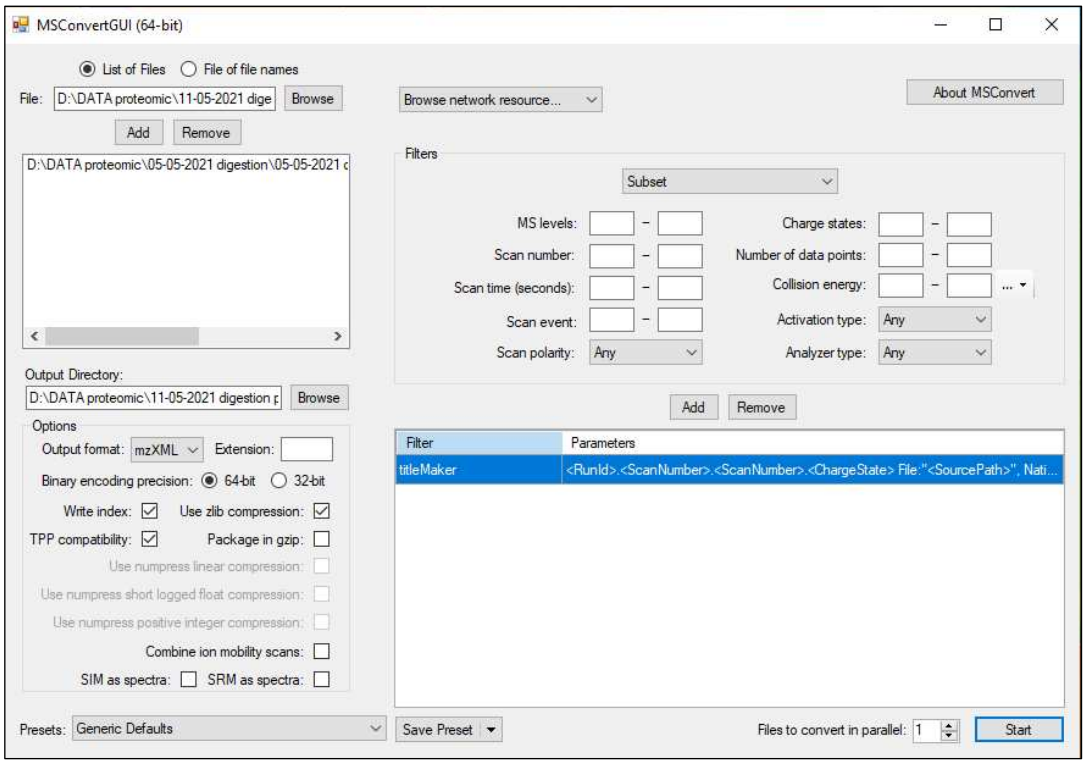


Annexe 91: Présentation de la plateforme du logiciel de protéomique massXpert

II. Utilisation du logiciel mMass pour le traitement des résultats liés à l'hydrolyse au NTCB

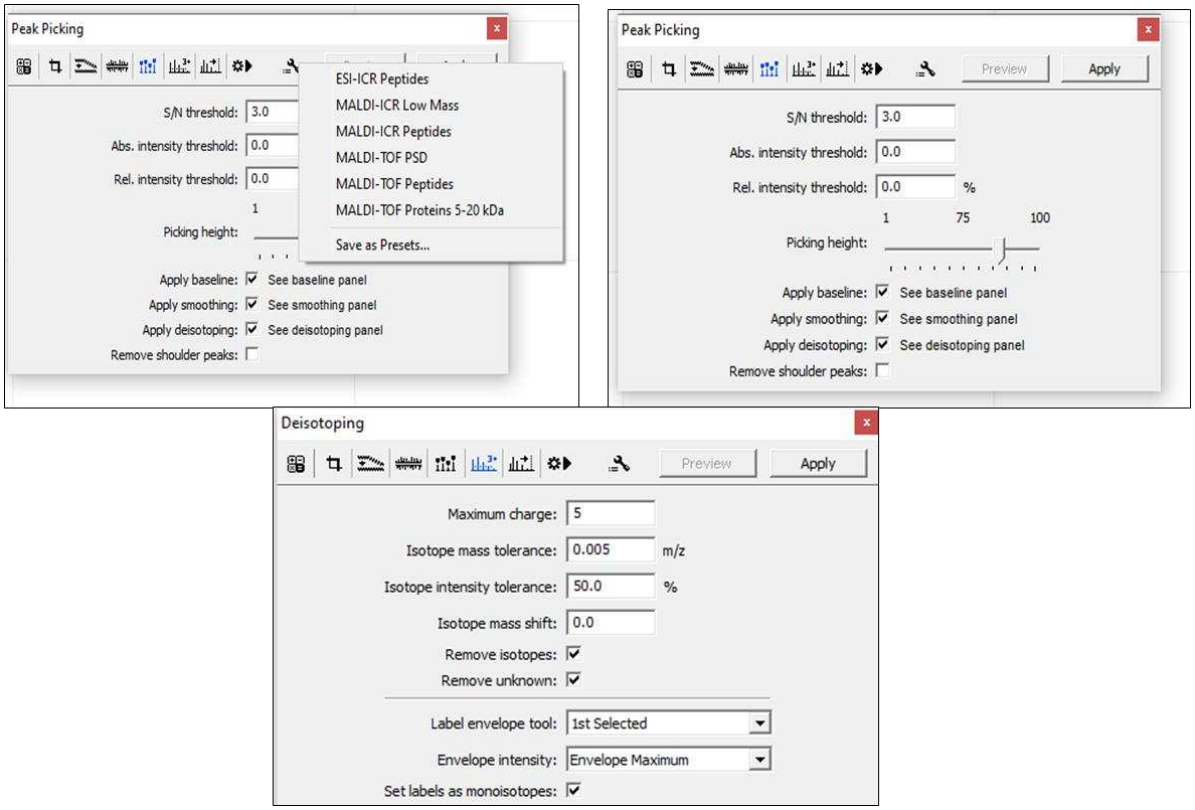
Le logiciel mMass est le dernier logiciel utilisé dans la quête d'utiliser le logiciel le plus adapté au traitement des résultats du NTCB, et surtout il offre plus de choix en termes d'enzyme ou agent chimique utilisé en protéomique. De plus, il n'est basé sur aucune banque de peptides. Premièrement, nous étudierons sont fonctionnement, avant d'exploiter nos résultats avec ce logiciel.

Avant toute chose, pour charger le fichier brut obtenu après analyse des peptides dans mMass, il faut convertir le fichier en extension mzXML (Annexe 92). Ceci peut se faire grâce au logiciel ProteoWizard, avec le module MSConvert, gratuit et téléchargeable sur internet.



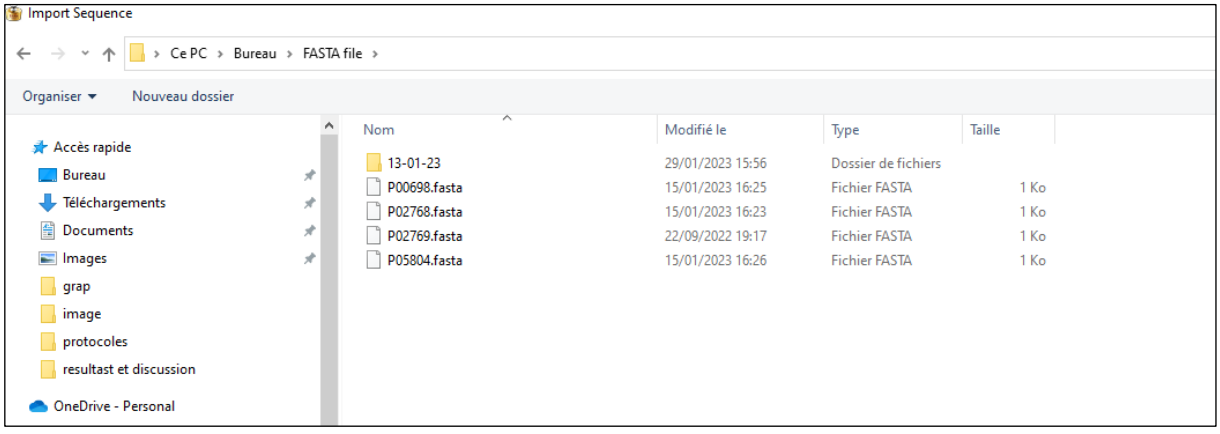
Annexe 92: Présentation du module de conversion de fichiers brutes MSConvert

Lorsque qu'on ouvre la plateforme mMass, on peut charger le fichier converti. Il y a deux types de fichier. Soit c'est un TIC (Total Ion Chromatrogram) associé des milliers de scans, un scan correspondant à un spectromètre de masse à un instant t de l'analyse, soit le fichier correspond à un seul scan. On peut apparenter un scan à un flash photo du passage de peptides dans l'analyseur Orbitrap, et plusieurs scans à plusieurs flash photo. Il est préférable de travailler avec un seul scan, plutôt que de parcourir le TIC et ses milliers de scans à analyser. Nous avons donc rassemblé tous les scans d'un TIC en un seul scan grâce à la fonction « Write to RAW File » du logiciel de l'Orbitrap Xcalibur qui a servi à la séparation et ionisation des peptides, sur lequel notre expérience est basée. Une fois que le fichier est dans le bon format, on peut le charger en cliquant « File » puis « open ».



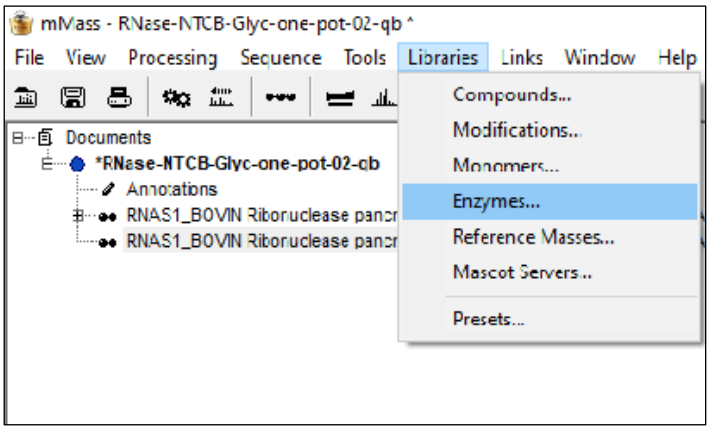
Annexe 93: Choix des paramètres de masse sur la plateforme mMass

L'ensemble des paramètres de masse ont été renseignés comme on le voit Annexe 93. Pour notre étude le mode qui correspond le plus est ESI-ICR Peptides. On a pu également modifier manuellement certains paramètres comme le signal sur bruit (S/N threshold).Ensuite, il a fallu télécharger la séquence de la protéine digérée en format FASTA, en faisant attention à bien retirer l'extension txt lorsque le fichier FASTA est téléchargé. Ceci a pu être fait en utilisant la fonction « Import Sequence » (Annexe 94).



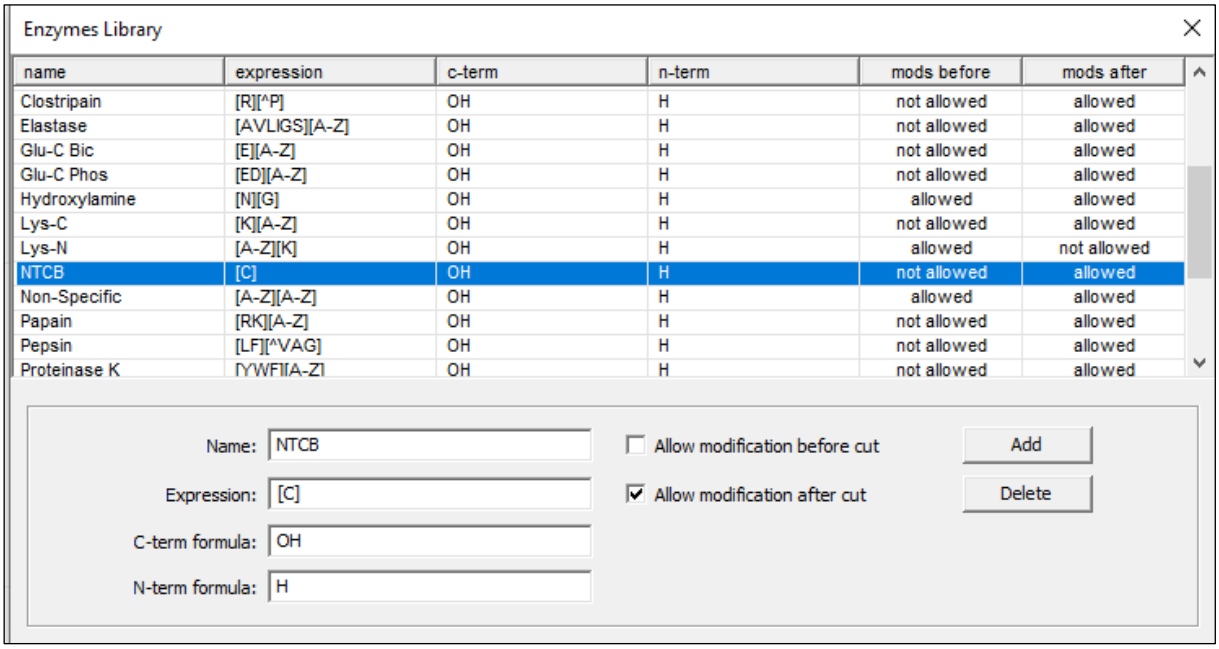
Annexe 94: Importation de la séquence FASTA sur la plateforme mMass

Une fois que le fichier est téléchargé et la séquence aussi l'analyse peut commencer. Comme les autres logiciels, il n'y aura plus qu'à renseigner le type de digestion, le site de clivage, et les modifications. Pour le NTCB, la digestion ne figurait pas dans la liste des enzyme/agent chimiques de digestion. Il a donc fallu l'ajouter en choisissant « Enzymes » dans librairies (Annexe 95).



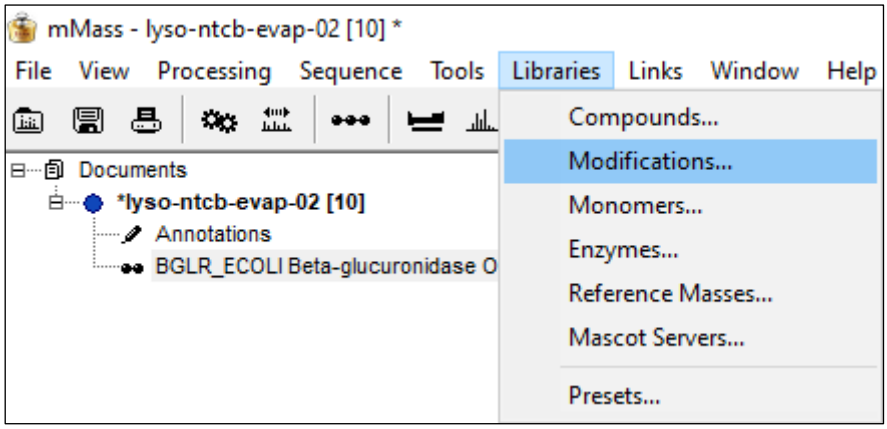
Annexe 95: Sélection du type de digestion sur la plateforme mMass

Manuellement, les informations concernant le NTCB ont donc été ajoutés comme on le voit sur la Annexe 96. Il y a un code spécifique pour renseigner les/le site(s) de clivage dans l'encart « expression ». Les lettres correspondant au code à une lettre des acides aminés. Entre la première paire de crochets et la deuxième, sont renseignés les acides aminés présents avant et après le site de clivage. Le terme [A-Z] signifie que n'importe quel acide aminé peut être présent. Enfin, le terme ^ signifie qu'on exclue le/les résidus présents après ce terme. Pour le NTCB, il a uniquement fallu renseigner [C] qui signifie clivage au niveau de la cystéine.



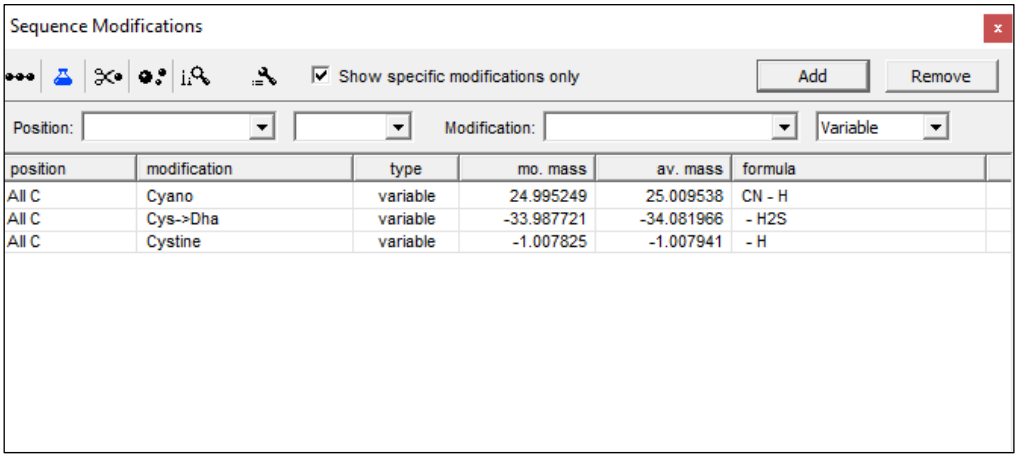
Annexe 96: Ajout de l'hydrolyse NTCB parmi les agents de digestion sur la plateforme mMass

Enfin, les modifications après le clivage ont été autorisées, puisque c'était cas pour les modifications du NTCB à savoir, l'iminothiazolidine (ITZ) et Dha (Dehydroalanination). Par ailleurs, il a aussi fallu renseigner celle-ci sur le logiciel en allant dans « Librairies » puis « Modifications » (Annexe 97).



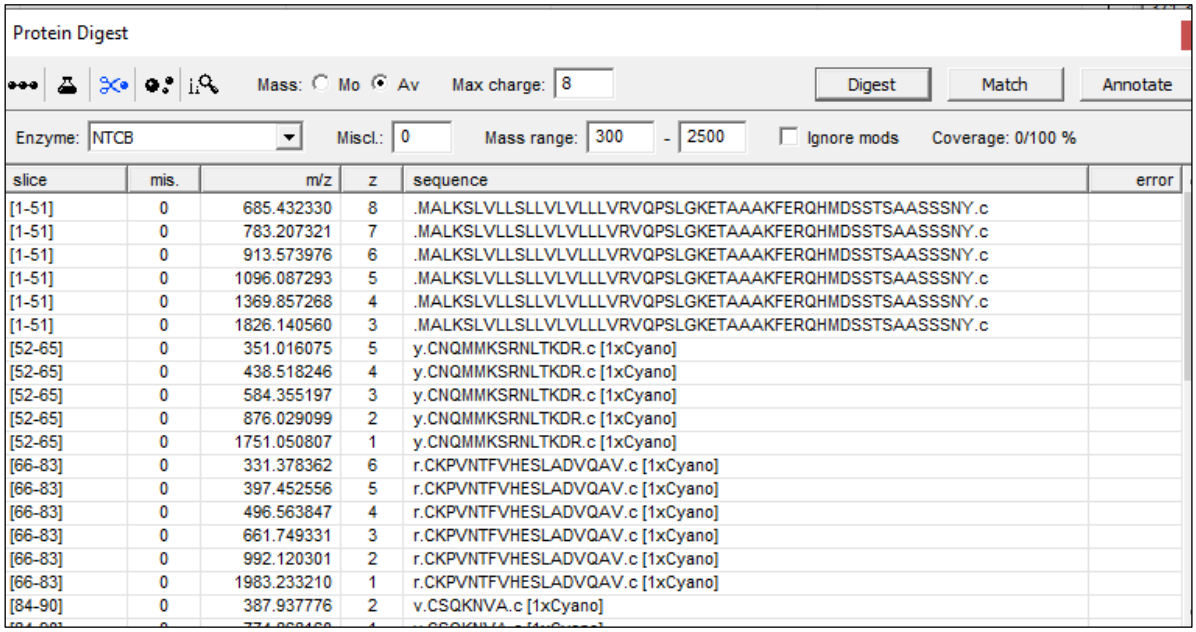
Annexe 97: Sélection des modifications sur la plateforme mMass

Fort heureusement, il n'a pas été nécessaire de créer des modifications pour le NTCB, car elles étaient déjà répertoriées selon le code universel unimod présent sur le site « unimod.org ». La modification issue de la cyanulation rajoutant une fonction ITZ se nomme « cyano, la dehydoalanination « Cys->Dha », et le pont disulfure entre deux cystéines « cystine. Ce sont ces trois modifications variables que l'hydrolyse au NTCB présente, comme on le voit sur la Annexe 98.



Annexe 98: Choix des modifications après hydrolyse NTCB sur la plateforme mMass

Après les modifications choisies, on peut lancer la digestion dans l'onglet « Digest » (Annexe 99). On peut y renseigner comme tout autre logiciel de protéomique le nombre de miscleavage, le nombre maximum de charge du peptide, le choix de la masse monoisotopique ou la masse moyenne du peptide, ainsi que la gamme de masse. En lançant « Digest », on obtient la liste théorique des peptides attendus. Pour croiser ceci aux résultats expérimentaux, il faut cliquer « Match ».



slice	mis.	m/z	z	sequence	error
[1-51]	0	685.432330	8	.MALKSLVLLSLLVLVLLLVRVQPSLGKETAAAKFERQHMDSSTSAASSSNY.c	
[1-51]	0	783.207321	7	.MALKSLVLLSLLVLVLLLVRVQPSLGKETAAAKFERQHMDSSTSAASSSNY.c	
[1-51]	0	913.573976	6	.MALKSLVLLSLLVLVLLLVRVQPSLGKETAAAKFERQHMDSSTSAASSSNY.c	
[1-51]	0	1096.087293	5	.MALKSLVLLSLLVLVLLLVRVQPSLGKETAAAKFERQHMDSSTSAASSSNY.c	
[1-51]	0	1369.857268	4	.MALKSLVLLSLLVLVLLLVRVQPSLGKETAAAKFERQHMDSSTSAASSSNY.c	
[1-51]	0	1826.140560	3	.MALKSLVLLSLLVLVLLLVRVQPSLGKETAAAKFERQHMDSSTSAASSSNY.c	
[52-65]	0	351.016075	5	y.CNQMMKSRNLTKDR.c [1xCyano]	
[52-65]	0	438.518246	4	y.CNQMMKSRNLTKDR.c [1xCyano]	
[52-65]	0	584.355197	3	y.CNQMMKSRNLTKDR.c [1xCyano]	
[52-65]	0	876.029099	2	y.CNQMMKSRNLTKDR.c [1xCyano]	
[52-65]	0	1751.050807	1	y.CNQMMKSRNLTKDR.c [1xCyano]	
[66-83]	0	331.378362	6	r.CKPVNTFVHESLADVQAV.c [1xCyano]	
[66-83]	0	397.452556	5	r.CKPVNTFVHESLADVQAV.c [1xCyano]	
[66-83]	0	496.563847	4	r.CKPVNTFVHESLADVQAV.c [1xCyano]	
[66-83]	0	661.749331	3	r.CKPVNTFVHESLADVQAV.c [1xCyano]	
[66-83]	0	992.120301	2	r.CKPVNTFVHESLADVQAV.c [1xCyano]	
[66-83]	0	1983.233210	1	r.CKPVNTFVHESLADVQAV.c [1xCyano]	
[84-90]	0	387.937776	2	v.CSQKNVA.c [1xCyano]	

Annexe 99: Obtention de la liste des peptides théoriques de digestion avec la plateforme mMass

La fonction match permet d'obtenir les peptides correspondant à l'hydrolyse présents dans la liste théorique, qui ont bien été trouvés dans les données expérimentale. Un match d'entre liste théorique et donnée expérimental est présenté par un point **vert (Annexe 99). Dans** l'onglet match, il est aussi possible de préciser le seuil d'écart de masse autorisée entre la masse théorique du peptide et sa masse expérimentale. Dans notre projet, l'écart autorisé a été de 5 à 10 ppm.

Résumés

Abstract

Targeting biomarkers to improve cancer diagnosis and treatment has become a true medical, even societal challenge today. Indeed, cancer is the second cause of death worldwide, with large disparities in terms of incidence and number of deaths depending on the country. There is an urgent need to develop clinical non-invasive, simple, and accurate diagnosis tools, especially available to a greater number. In this context, we have proposed a new diagnosis strategy, called induced volatolomics. It is based on the use of non-toxic isotopic probes, releasing Volatile Organic Compounds (VOCs). These probes, which can be activated by enzymes, can highlight enzymatic markers associated with pathologies, such as cancer. In a previous study, we showed the efficiency of the first VOC-based probe, i.e. ethylglucuronide-D5 activated by β-glucuronidase, to diagnose solid tumours in vivo and monitor tumour growth during a chemotherapy treatment. In order to improve this strategy's robustness and limit the false-positives number that can be induced by only one enzyme screening, I developed a cocktail made of probes, each activable by one glycosidase, a cancer marker. This strategy validated on a murine model allows discrimination between healthy and cancerous mice after cocktail injection. To facilitate the transfer of this new diagnosis tool to the clinic, I also got involved in the development of a clinical protocol based on adding the probe cocktail to a blood sample. While clinical trials didn't start yet, development and optimization phases were achieved.

In parallel, to confirm the presence of glycosidases in the tumour microenvironment, I implemented a strategy of chemical proteomics that aims to separate abundant protein (albumin, fibronectin) from glycosidases, present in very small quantities in malignant tissue. In this way, a middle-down proteomic approach with a chemical (NTCB) was developed to obtain proteotypic peptides for each glycosidase. This transversal approach permits certification of the presence of targeted proteins in the extracellular matrix of a solid tumour.

Key words: cancer, proteomics, biochemistry, enzyme, protein, biomarkers, VOC, HRMS

Résumé

Cibler des biomarqueurs pour améliorer le diagnostic et les traitements des cancers est devenu un véritable défi médical, voire un challenge sociétal. En effet, le cancer est la seconde cause de mortalité dans le monde, avec des disparités en termes d'incidence et de nombre de décès, fortes selon les pays. Il y a un besoin urgent de développer des outils de diagnostic clinique non-invasif, simple, précis et surtout accessible au plus grand nombre. Dans ce contexte, nous avons proposé une nouvelle stratégie de diagnostic, appelée la volatolomique induite. Elle repose sur l'utilisation de sondes isotopiques non-toxiques, libérant des Composés Organiques Volatils (COVs). Ces sondes activables par des enzymes, peuvent mettre en évidence des marqueurs enzymatiques associés à des pathologies, telles que le cancer.

Dans une précédente étude, nous avons démontré l'efficacité d'une première sonde, i.e. ethylglucuronide-D5, activée par la β-glucuronidase, pour diagnostiquer des tumeurs solides in vivo et suivre leur évolution au cours d'un traitement chimio-thérapeutique. Afin d'améliorer la robustesse de cette stratégie et limiter le nombre de faux-positifs qui peuvent être induit par le ciblage d'une seule et unique enzyme, j'ai développé un cocktail de sondes activables chacune par une glycosidase, marqueur des cancers. Cette stratégie a alors été validée sur modèles murins, et permet la discrimination entre des souris saines et des souris cancéreuses après injection du cocktail. Afin de faciliter le transfert de ce nouvel outil diagnostic en clinique, j'ai également participé au développement d'un protocole clinique basé sur l'ajout de cocktail de sondes dans un prélèvement sanguin. Bien que l'essai clinique n'ait pas encore débuté, les phases de développement et d'optimisation ont, elles, été achevées. En parallèle, afin de confirmer la présence des glycosidases dans le microenvironnement tumoral, j'ai mis au point une stratégie de protéomique chimique afin de séparer les protéines abondantes (albumine, fibronectine) des glycosidases, présentes en très petite quantité dans les tissus malins. De ce fait, une approche de protéomique middle-down avec un agent chimique (NTCB) a été développée afin d'obtenir des peptides protéotypiques pour chacune des glycosidases. Cette approche transversale devrait permettre de certifier la présence des protéines ciblées dans la matrice extracellulaire des tumeurs solides.

Mots clés : cancer, protéomique, biochimie, enzyme, protéine, biomarqueurs, COV, HRMS