



Extension de l'identification des protéines myristoylées chez les Eucaryotes

Frédéric Riviere

► To cite this version:

Frédéric Riviere. Extension de l'identification des protéines myristoylées chez les Eucaryotes. Biochimie [q-bio.BM]. Université Paris-Saclay, 2021. Français. NNT : 2021UPASL105 . tel-03794051

HAL Id: tel-03794051

<https://theses.hal.science/tel-03794051>

Submitted on 3 Oct 2022

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

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

Extension de l'identification des protéines myristoylées chez les Eucaryotes

Extension of the identification of N-acylated proteins in Eukaryotes

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 577 : structure et dynamique des systèmes vivants (SDSV)

Spécialité de doctorat : Sciences de la vie et de la santé

Graduate School : Life Science and Health. Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche : **Institute for Integrative Biology of the Cell (I2BC)** (Université Paris-Saclay, CEA, CNRS),
sous la direction de **Thierry MEINNEL**, Directeur de Recherche

Thèse soutenue à Paris-Saclay, le 08 décembre 2021, par

Frédéric RIVIERE

THESE DE DOCTORAT

NNT : 2021UPASL105

Composition du Jury

Nicolas BAYAN Professeur, Université Paris-Saclay	Président
Christine CARAPITO Chargée de Recherche, HDR, CNRS Université Strasbourg	Rapportrice & Examinatrice
Dominique PADOVANI Chargé de Recherche, HDR, CNRS Université Paris Descartes	Rapporteur & Examineur
Thierry MEINNEL Directeur de Recherche, HDR, CNRS Université Paris-Saclay	Directeur de thèse

Remerciements

Je remercie tout d'abord les membres du jury, Christine Carapito et Dominique Padovani pour m'avoir fait l'honneur de juger mon travail en tant que rapporteurs, ainsi que Nicolas Bayan, qui a accepté de présider ce jury. L'intérêt que vous portez à mon travail me touche particulièrement.

Je remercie aussi les membres de mon comité de suivi de thèse, qui tout au long de mon travail ont été là pour me prodiguer de précieux conseils. Odile Schiltz, pour la partie spectrométrie de masse, et Guy Tran Van Nhieu pour la partie Shigella.

Je tenais à remercier tout particulièrement mon directeur de thèse, Thierry Meinel, de m'avoir confié ce projet et de l'avoir supervisé d'une manière exceptionnelle. Merci d'avoir cru en moi, de m'avoir encouragé, motivé, et d'avoir été toujours disponible. J'ai énormément appris durant ces quatre années de thèse. Tu as rempli ta mission avec excellence.

Je remercie aussi Carmela Giglione, qui m'a suivie aussi tout au long de ce projet en tant que chef d'équipe. Merci pour ta bonne humeur et pour le chauffage dans ton bureau quand on le partageait. Mais surtout merci de nous avoir fait découvrir la Sicile lors de cette retraite scientifique.

Mes remerciements s'adressent également à Cyril Dian, sans qui cette thèse n'aurait jamais vu le jour. Merci de m'avoir pris en stage de Master dans cette équipe, merci d'avoir cru en moi, de m'avoir soutenu, et motivé. Merci pour toutes ces discussions scientifiques ou non que l'on a pu avoir tout le long de mon passage dans ce laboratoire. Toutes ces séries, reportages, ou lectures en commun ont créés des sujets de discussions d'une diversité impressionnante, ce qui nous a permis de mieux supporter ces très longs trajets en RER B.

Je remercie chaleureusement Willy Bienvenu qui a su m'expliquer avec brio et patience le maniement des machines de spectrométrie de masse, et de LC. Qui m'a appris à trouver les pannes et à réparer ce qui est réparable. Merci pour tous ces conseils avisés qui me serviront pour la suite.

Merci aussi à tous les membres de l'équipe toujours présents ou partis : Merci à Thomas pour ces longues parties de jeux vidéo, et pour tout ce que l'on a partagé, merci à Rémi (« le stagiaire ») pour tous ces bons moments. Merci à Frédéric Frottin pour son aide sur les cellules lors de l'écriture de mon manuscrit. Merci à Flaria pour tous ce qu'elle m'a appris sur les cellules eucaryotes au début de ma thèse. Merci au grand Francesco pour tous les moments passés au laboratoire ou à l'extérieur. Merci à tous les autres membres de l'équipe, les JBs, Lucile, et tous les autres qui sont partis.

Je tiens aussi à remercier grandement la plateforme de spectrométrie de masse SiCaPS, qui m'a permis d'utiliser les machines en me faisant confiance.

Il faut surtout que je remercie ma famille, sans qui cette thèse n'aurait jamais vu le jour, grâce au soutien moral et financier tout au long de mes études. Donc un grand merci à mon père, qui malgré tous ses essais pour me dissuader de suivre ses traces, m'a forcément influencé pour choisir ce métier. Merci à ma mère qui a toujours été là à des moments clés de

mes études, qui m'a toujours soutenu et cru en moi. Merci à ma sœur et mon frère pour tous le soutien moral dans les moments difficiles.

Merci à Manon, qui a été la plus belle rencontre dans ce laboratoire. Merci à toi de m'avoir soutenu, motivé et surtout supporté dans les moments difficiles de cette thèse. Merci pour tous ces moments merveilleux passés à tes côtés, et pour le bout de chemin fait ensemble. Le plus beau reste encore à réaliser ensemble, ce n'est que le début. Et je me réjouis en pensant à tout ce l'on va vivre ensemble.

Merci à toutes les autres personnes que j'ai pu rencontrer au cours de cette thèse, Pierre, Lammia, les membres de Yourl2BC, et tous les autres...

Merci aussi à toutes les personnes que j'ai pu rencontrer en dehors du laboratoire, comme Camille et Andres, pour toutes ces sorties escalade à Fontainebleau qui m'ont fait me sentir mieux tout au long de cette thèse.

I. Introduction	1
I.1. La myristoylation chez les eucaryotes	1
I.1.1. Description générale.....	1
I.1.2. Les sites de myristoylation chez les eucaryotes	2
I.1.3. Production de protéines commençant par une glycine N-terminale réactive à la myristoylation	3
I.1.4. La N-myristoylation.....	4
<i>I.1.4.a. Réaction de myristoylation.....</i>	<i>4</i>
<i>I.1.4.b. La NMT, une GNAT particulière.....</i>	<i>6</i>
<i>I.1.4.c. Sélectivité de substrat</i>	<i>8</i>
I.2. Rôle de la N-myristoylation	9
I.3. Implication de la myristoylation dans les pathologies.....	12
I.3.1. L'utilisation de la machinerie de myristoylation de l'organisme hôte par les virus.....	12
I.3.2. La myristoylation chez les protozoaires	13
I.3.3. La myristoylation dans les cancers.....	14
I.3.4. La myristoylation au cours des infections bactériennes.....	15
I.4. Le protéome myristoylé	17
I.4.1. L'analyse <i>in vivo</i>	17
<i>I.4.1.a. Approche à petit débit et moyen débit.....</i>	<i>17</i>
<i>I.4.1.b. Les marquages chimiques pour une étude globale.....</i>	<i>19</i>
<i>I.4.1.c. Analyse protéomique par LC-MS.....</i>	<i>21</i>
I.4.2. L'analyse <i>in vitro</i>	23
<i>I.4.2.a. Test enzymatique.....</i>	<i>23</i>
<i>I.4.2.b. Suivie par MALDI-TOF.....</i>	<i>25</i>
I.4.3. L'analyse <i>in silico</i>	25
I.4.4. L'analyse structurale.....	26
I.5. Défis actuels concernant l'étude de la myristoylation et alternatives	28
I.5.1. Les défis de longue date	28
I.5.2. Les défis abordés dans mon projet de thèse.....	29
II. IpaJ, nouvel outil pour l'étude de la myristoylation	33
II.1. Étude et validation de la protéase IpaJ comme outil d'analyse de la myristoylation	33
II.1.1. Production de la protéase IpaJ	33
II.1.2. Développement d'un test de suivi de l'activité d'IpaJ	34

II.1.2. Développement du test d'activité de myristoylation par la NMT couplé à la démyristoylation par IpaJ	37
II.1.3. Sélectivité d'IpaJ pour des peptides synthétiques N-acétylés	41
II.2. Utilisation d'IpaJ comme outil biotechnologique pour l'étude de la myristoylation.....	42
II.2.1. Fractionnement cellulaire.....	42
II.2.2. Traitement des lysats par IpaJ	46
II.3. Enrichissement des N-terminaux libre issus de la digestion par IpaJ des protéines myristoylées	53
II.4. Exploration de l'effet de la protéase IpaJ <i>in-cellulo</i>	56
<i>III. Des instantanés haute résolution de la N-myristoyltransférase humaine en action éclairent un mécanisme favorisant la myristoylation N-terminale des Glycines et des Lysines</i>	<i>65</i>
<i>IV. Eléments de spécificité contribuant à la lysine myristoylation, mécanisme associé et conception raisonnée d'inhibiteurs de la NMT.....</i>	<i>85</i>
<i>V. Discussion-Conclusions</i>	<i>139</i>
V.1. Validation d'IpaJ comme outil d'assistance à l'étude de la myristoylation.....	139
V.2. Validation de modèles biologiques adaptés.....	140
V.3. Mise en situation d'IpaJ pour l'étude du myristoylome des glycines N-terminales.....	142
V.4. Le mécanisme réactionnel de la NMT révèle son rôle dans la myristoylation des lysines	144
V.5. IpaJ est un outil unique pour distinguer les deux types de myristoylation.....	146
V.6. La protéase IpaJ permet générer de nouveaux sites de K-myristoylation : le cas d'ARF6	147
V.7. Conclusions	150
<i>VI. Matériels et méthodes.....</i>	<i>151</i>
VI.1. Culture cellulaire	151
VI.2. Fractionnement cellulaire	151
VI.3. Digestion échantillons sur gel SDS-Pages	151
VI.4. Extraction des protéines.....	153
VI.5. Préparation du N-acétoxy-[2H3] succinimide.....	153

VI.6. Acétylation N-terminal des protéines ou des peptides.....	153
VI.7. Digestion liquide des protéines par la trypsine	154
VI.8. Chromatographie en phase inversée avec cartouches Sep-Pak.....	154
VI.9. Digestion des échantillons par IpaJ.....	155
VI.10. Marquage des peptides de BSA avec TMPP	155
<i>VII. Références.....</i>	<i>156</i>
<i>VIII. Annexes</i>	<i>164</i>
VIII.1. Publication Dian et al,	164
VIII.1.1. Figures supplémentaires	164
VIII.2. Publication Rivière et al,	189
VIII.2.1. Figures supplémentaires	189
VIII.2.2. Table 1 supplémentaire	197
VIII.2.3. Table 2 supplémentaire	204
VIII.2.4. Table 3 supplémentaire	205
VIII.2.5. Table 4 supplémentaire	207
VIII.2.6. Données supplémentaires 1a.....	210
VIII.2.7. Données supplémentaires 1b	238
VIII.2.8. Données supplémentaires 1c.....	250
VIII.2.9. Données supplémentaires 1d	291
VIII.2.10. Données supplémentaires 1°	311

Chapitre I

I. Introduction

I.1. La myristoylation chez les eucaryotes

I.1.1. Description générale

La cellule correspond à l'unité fondamentale des organismes vivants. Elle est capable de se diviser, d'interagir avec le monde extérieur, et de maintenir le contenu de l'information génétique au cours de ses divisions successives. Pour ce faire, la cellule tire profit des propriétés remarquables de différents types de macromolécules telles que les acides nucléiques, protéines, polymères de sucres (oligosaccharides) ou lipides complexes. La cellule est dotée d'une membrane séparant son contenu de l'extérieur : il s'agit de la membrane plasmique. Cette barrière permet de préserver le contenu de la cellule tout en maintenant les interactions avec le monde extérieur et en la protégeant des menaces extérieures qu'elles soient biotiques (pathogènes par exemple) ou abiotiques (oxydants par exemple). La composition de la membrane plasmique concourt au bon fonctionnement de la cellule. Bien que constituée essentiellement de lipides (40%) et de protéines (>50%), d'autres composants moins abondants, comme les glucides (<10%), sont indispensables au bon fonctionnement cette mégastructure. La membrane est constituée d'une très grande variété de lipides, qui ne sont pas arrangés de façon anarchique mais organisés de façon à compartimenter cette structure. Au sein de cette membrane, on retrouve différents types de protéines, comme des protéines transmembranaires, caractérisées par des domaines hydrophobes directement intégrés dans la bicouche lipidique qui constitue cette dernière. En plus des protéines transmembranaires, les membranes possèdent des protéines dites lipidées, c'est à dire des protéines possédant des lipides, attachés par liaison covalente. On peut citer le cas des acides gras mais aussi des isoprénoïdes voire encore des stérols. Ces modifications servent d'ancre et, lorsqu'elles sont réversibles, permettent non seulement une régulation du positionnement fin de la protéine mais surtout de disposer d'une protéine localisée au niveau d'une membrane tout en proposant une surface entièrement disponible pour l'interaction avec des

petites molécules (substrats ou ligands) ou d'autres protéines par exemple. L'ancrage peut avoir lieu à l'intérieur ou à l'extérieur de la cellule. C'est cette accessibilité qui va concourir à permettre à ces différentes protéines de jouer des rôles très importants au sein de la cellule. C'est le cas par exemple du recrutement d'autres protéines jouant des rôles importants au niveau de la transduction des signaux cellulaires. Les membranes cellulaires ne se limitent pas à la membrane plasmique chez les eucaryotes, car elles forment un grand réseau interne de membrane. Ce réseau d'endomembranes, permet de compartimenter la cellule et de former différents organites aux fonctions diverses et variées.

Je me suis intéressé au cours de ma thèse à une lipidation particulière, la myristoylation plus particulièrement chez l'humain, que ce soit au niveau mécanistique *in vitro* et *in vivo*. Il s'agit d'une modification permettant un ancrage interne des protéines.

I.1.2. Les sites de myristoylation chez les eucaryotes

La myristoylation est une modification connue depuis le début des années 1980. La myristoylation affecte un petit nombre de protéines majeures telles que la sous-unité catalytique de la protéine kinase dépendante de l'AMPc humaine (AMPK (Carr *et al*, 1982)) ou de la calcineurine B (Aitken *et al*, 1982). La découverte concomitante que les protéines p15^{gag} et p27^{nef} du virus HIV, l'agent causatif du SIDA, et que celle d'un marqueur récurrent du cancer (p60^{src}) subissent la modification a déclenché naturellement immédiatement l'engouement pour cette modification et l'identification d'outils pour la bloquer. Depuis, on a progressivement pris conscience que la myristoylation affecte un grand nombre de protéines et est présente chez tous les eucaryotes (Boutin, 1997). La modification est principalement associée aux glycines N-terminales (α) (G-myristoylation), mais il existe d'autres sites où elle peut se produire, bien que plus rarement. C'est le cas des chaînes latérales des lysines (Lys-N- ϵ ; K-myristoylation) ou des cystéines (Cys-S-myristoylation, S-myristoylation) La K-myristoylation a été décrite en 1992 (Stevenson *et al*, 1992) sur le facteur de nécrose tumorale- α (TNF1 α) et le précurseur de l'interleukine- α 1 (Il α 1). Ce précurseur porte

deux myristoylations sur deux lysines internes ; plus récemment, d'autres protéines ont été décrites comme ayant des lysines internes myristoylées (Cao *et al*, 2019). Bien que la découverte de la myristoylation des lysines internes date des années 90, si l'enzyme responsable de cette modification n'a toujours pas été caractérisée, on sait qu'elle utilise le MyrCoA comme donneur (Stevenson *et al*, 1993), comme l'atteste des tests réalisés avec des extraits de monocytes.

En ce qui concerne la S-myristoylation, à l'heure actuelle, trop peu de données ont été collectées à son sujet ; une seule étude robuste montre son existence chez les eucaryotes, mais il n'est pas encore démontré que cette modification corresponde à une acylation spécifique de la longueur de la chaîne en C14. En effet, étant donné que l'étude a été réalisée en présence d'un analogue du myristate, et qu'une grande partie du protéome soit Cys-S-acylé et en particulier Cys-S-palmitoylé (C16), une incorporation non spécifique de l'analogue est possible. Ainsi, il est connu que les enzymes en charge de la palmitoylation sont peu spécifiques de la longueur de chaîne (Greaves *et al*, 2017), ce qui pourrait permettre une incorporation à bas bruit de chaînes plus courtes comme le myristate.

I.1.3. Production de protéines commençant par une glycine N-terminale réactive à la myristoylation

Que ce soit chez les eucaryotes ou chez les procaryotes, la synthèse des protéines est initiée par une méthionine (Giglione *et al*, 2004). La méthionine initiatrice n'est pas systématiquement conservée au cours de la vie de la protéine, car une enzyme est capable de la retirer ; il s'agit de la méthionine aminopeptidase (MetAP). Cette enzyme est une métalloprotéine conservée chez tous les organismes, sous une forme chez les procaryotes et sous deux formes chez les eucaryotes (MetAP1, et MetAP2).

On estime que les deux tiers des protéines sont substrats des MetAPs, et son activité est essentielle pour les cellules. Son rôle est encore mal connu (Giglione *et al*, 2004) mais engendre la grande diversité d'acides aminés identifiés au N-terminus des protéines de chaque protéome. La réaction catalysée par la MetAP est indispensable à

la réalisation d'autres modifications N-terminales. C'est le cas de la N-acétylation par le complexe NatA qui s'exerce sur plus de 50% des protéines (Aksnes *et al*, 2016) du protéome commençant essentiellement par Ala, ou Ser ou encore la N-myristoylation qui s'effectue sur certaines glycines N-terminales.

Le clivage d'une protéine précurseur par des protéases peut aussi contribuer à générer de nouveaux N-terminaux. Parmi les protéases connues, on peut citer les caspases dont le site de clivage est compatible pour générer des séquences commençant par une glycine ; certaines peuvent présenter un contexte favorable à la myristoylation. C'est le cas de certains signaux d'apoptose, comme cela a été décrit pour la protéine BID (Zha *et al*, 2000).

I.1.4. La N-myristoylation

I.1.4.a. Réaction de myristoylation

La N-Myristoylation, qui correspond à l'ajout covalent d'un acide gras saturé de 14 carbones (C14:0) (Johnson *et al*, 1994) sur la Glycine N-terminale d'une protéine, est principalement réalisée de façon co-translationnelle et concerne approximativement 2% du protéome (Castrec *et al*, 2018) (**Figure I-1**). La modification se réalise généralement de manière très précoce au niveau de la chaîne polypeptidique naissante du substrat encore attaché au ribosome (Kramer *et al*, 2009). C'est la N-myristoyltransferase (NMT) qui catalyse cette réaction qualifiée de co-translationnelle. Cette modification correspond au transfert du groupement myristate provenant d'un myristoyl coenzyme A (Myr-CoA) sur l'amine N-terminale d'un résidu glycine nouvellement exposé après excision de la méthionine initiatrice par la méthionine aminopeptidase (MetAP) (Giglione *et al*, 2004). Le ribosome étant un haut lieu d'interactions pouvant recruter plusieurs enzymes modifiant les protéines, d'autres modifications comme la N-acétylation peuvent entrer en compétition avec la N-myristoylation, sans qu'aucun mécanisme de régulation ne soit connu à l'heure actuelle.

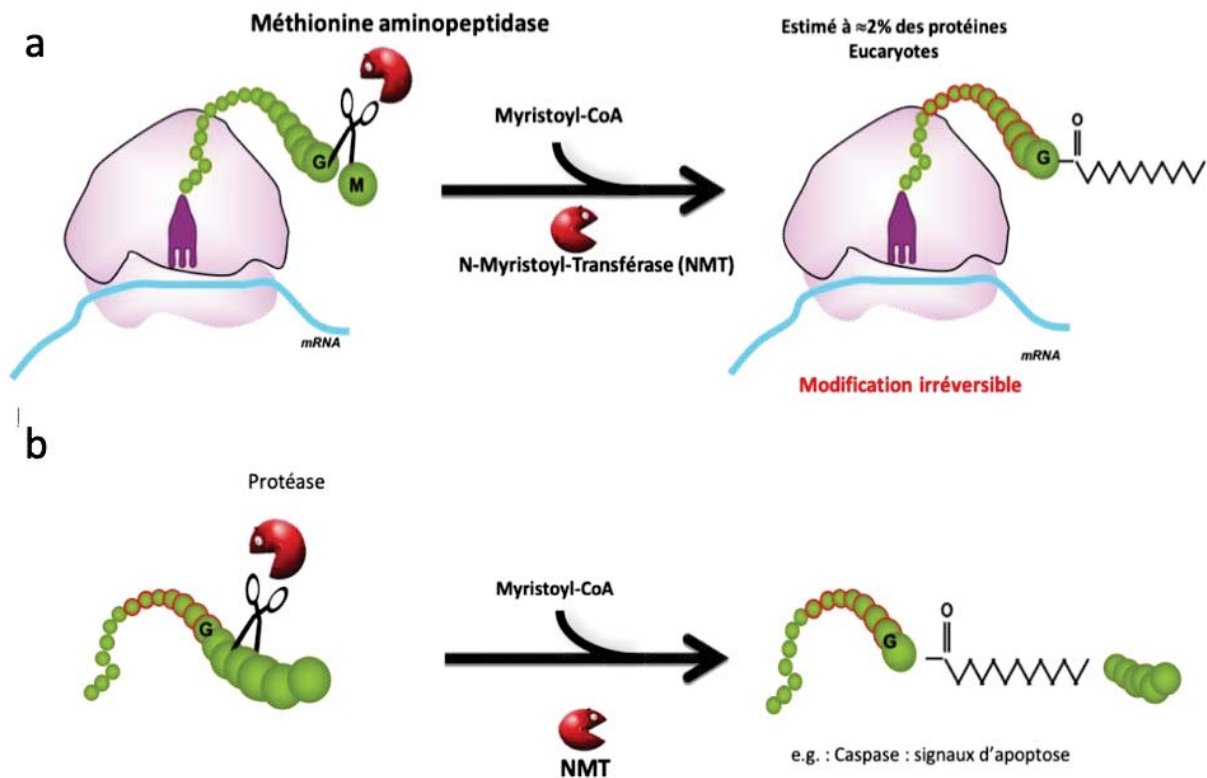


Figure I-1 : Schéma récapitulant les différents types de myristoylations dans la cellule ; (a) cas de la myristoylation co-traductionnelle ; mécanisme le plus fréquent pour cette modification (b) cas de la myristoylation post-traductionnelle ; réalisée dans de plus rares cas, comme lors de l'apoptose.

La réaction de myristoylation se produit *in vivo* uniquement si la méthionine initiatrice est retirée et si et seulement si la séquence comporte une glycine en position deux (cad suivant la Met1) et que la séquence en aval est favorable. Il faut donc, pour que la protéine soit myristoylée de façon co-traductionnelle, que sa séquence subisse successivement l'action de la MetAP puis de la NMT.

Une minorité de protéines peuvent être aussi myristoylées ultérieurement, cette fois de manière post-traductionnelle, si la nouvelle séquence que le clivage protéolytique dévoile est favorable. C'est le cas de certaines protéines clivées par les caspases lors de l'induction du phénomène d'apoptose, la protéine myristoylée pouvant servir elle-aussi de signal d'apoptose (comme cela a été décrit pour la protéine BID (Zha *et al*, 2000).

L'ajout du groupement myristate par la NMT, crée une liaison amide extrêmement stable, qualifiée d'irréversible (Farazi *et al*, 2001). Aucune approche, qu'elle soit

chimique ou enzymatique, n'a en effet été décrite comme permettant de retirer cette modification sans en affecter la séquence polypeptidique. Seule l'enzyme IpaJ a été montrée comme capable d'affecter un substrat myristoylé mais elle agit comme une protéase et clive la protéine ciblée après la première glycine (voir I.3.4).

I.1.4.b. La NMT, une GNAT particulière

La NMT est une enzyme conservée chez tous les eucaryotes et elle composée en moyenne de 500 ± 65 acides aminés; la NMT est présente sous une seule forme chez les eucaryotes unicellulaires et sous deux isoformes chez les organismes multicellulaires (NMT1 et NMT2). Les différences de taille des NMT en fonction des organismes dépendent essentiellement de la partie N-terminale de la protéine. Ce domaine serait impliqué dans la liaison avec le ribosome. En effet, la liaison de la NMT au ribosome implique les 100 premiers acide aminés (Glover *et al*, 1997). Cette extension, mal conservée que ce soit en longueur ou en séquence, n'est pas utile à son activité *in vivo* (Duronio *et al*, 1990). A ce jour, nous disposons d'un grand nombre de structures cristallines à haute résolution (74) de NMTs de différents organismes, pour la plupart dépourvues de l'extension N-terminale, qui n'est probablement pas ordonnée en solution.

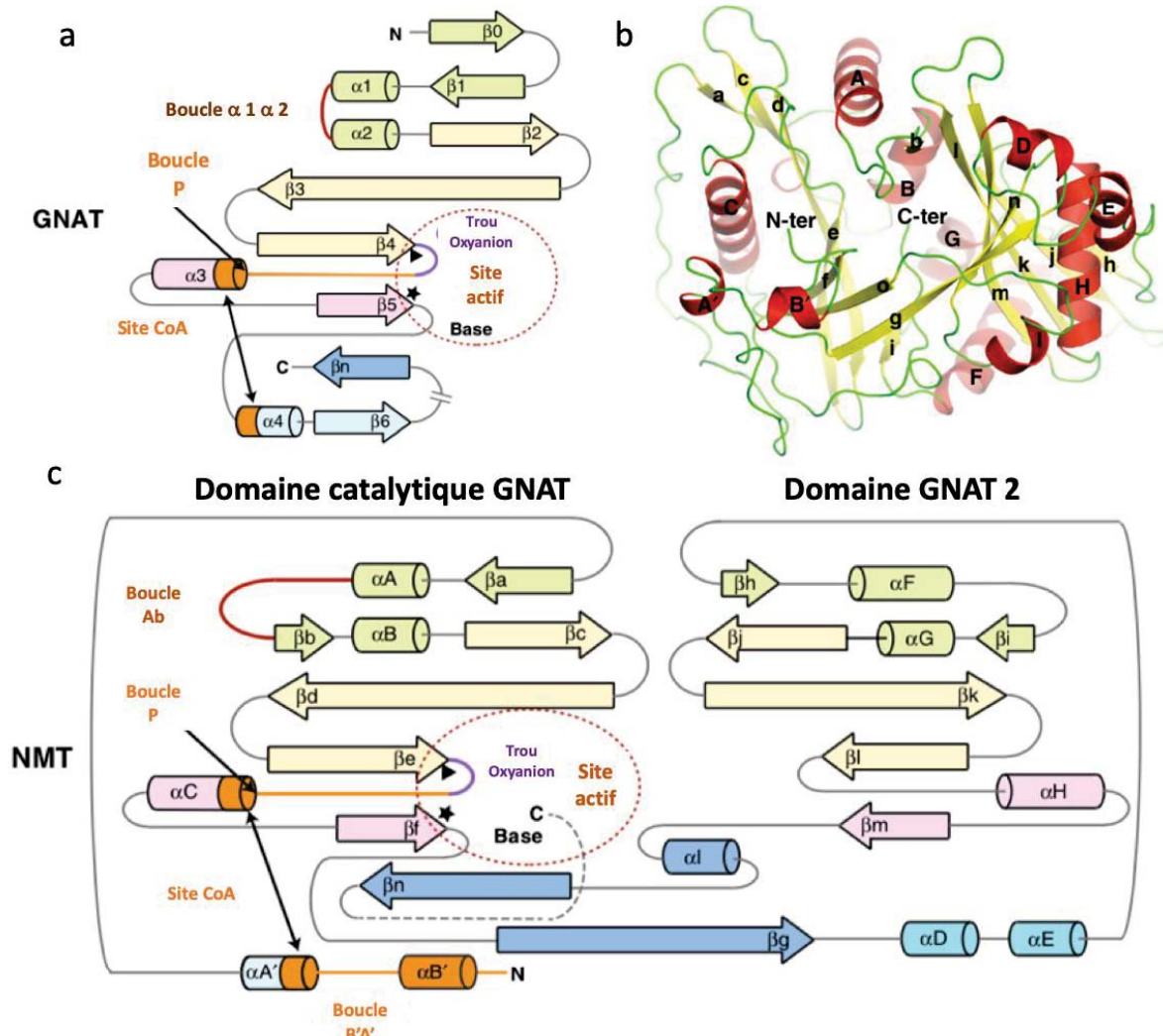


Figure 1-2 : La NMT fait partie de la grande famille des GNATs. (a) Représentation topologique générale des GNAT. Les éléments de structures secondaires conservés sont colorés respectivement en vert pâle, jaune, rose et cyan. Les éléments de structures secondaires non conservés sont colorés en bleu. Les régions importantes pour la liaison du CoA sont indiquées en orange. Le renflement du « trou de l'oxyanion » ou oxyanion hole) est surligné en violet. Un cercle brisé rouge met en évidence la position du site actif. La boucle Ab et la boucle $\alpha_1\alpha_2$ équivalente sont colorées en rouge. Des étoiles et des triangles noirs mettent en évidence des positions cruciales dans le site actif. (b) Vue du repliement de HsNMT:MyrCoA (code pdb 5O9T) chaque élément est représenté dans le même code couleur que celui des panneaux a et c. (c) Représentation topologique de la NMT montrant les deux domaines GNAT en utilisant le même code couleur qu'en a.

Les NMTs appartiennent à la grande famille des N-acétyltransférase (GNAT) toutes reliées au repliement que la structure 3D protéine GCN5 a défini à l'origine (Salah Ud-Din *et al*, 2016; Vetting *et al*, 2005). On retrouve dans cette famille de nombreuses acétyltransférases de petits substrats (antibiotiques, sérotonine, aminoacyl, glucosamine-6-phosphate etc...) mais aussi de protéines, comme les histones acétyltransférases qui modifient des lysines ou les N-acétyltransférases comme Nat1 qui modifient le terminus des protéines. Au sein des GNATs, les NMT sont atypiques du fait de l'existence de deux domaines GNAT fusionnés (**Figure I-2**) ; l'homologie de séquence entre les deux domaines est négligeable. Ces deux domaines GNAT ont des rôles différents dans la catalyse. En effet, le domaine N-terminal contient la majorité des résidus importants du site actif et la poche de fixation du MyristoylCoA tandis que le domaine C-terminal contient les différentes poches de liaison du substrat protéique. Une dernière particularité des NMTs est le rôle du carboxylate du résidu C-terminal qui joue le rôle de base catalytique pour la réaction (Bhatnagar *et al*, 1998).

1.1.4.c. Sélectivité de substrat

La structure tridimensionnelle de la NMT1 humaine en complexe avec les deux produits de la réaction, le CoA et plusieurs Myr-peptides a été résolue récemment au laboratoire. L'analyse du site actif de ces structures en combinaison des tests de N-myristoylation *in vitro* sur une banque de peptides N-terminaux montre que le transfert s'effectue uniquement sur des substrats en cours de synthèse possédant à l'extrémité N-terminale une séquence favorable de type $G_2-[^{\wedge}PRDELTV]_3-X_4-[^{\wedge}DEGR]_5-[SGCA]_6-[KR]_7-X_8-X_9$, où $^{\wedge}$ définit les résidus exclus (Castrec *et al*, 2018). Ce travail correspond à une percée significative dans la compréhension du mécanisme réactionnel de cette enzyme et de façon plus général de cette famille d'enzymes.

Bien que la séquence dite « favorable » ait été découverte, un nombre significatif de substrats *in vivo* possèdent une séquence a priori défavorable, comme la protéine Niban2 (Q96TA1) dont la séquence commence par (M)GDVLSTHLD. Pour cet exemple, l'acide aspartique qui se trouve en position trois, est très fortement

défavorable voire exclut la modification selon le profil et de plus, le peptide n'a pas de lysine en position sept. Néanmoins, la leucine 5 et la serine 6 stimulent la réaction, et cette protéine se retrouve bien myristoylée *in vivo*. Cet exemple illustre bien la difficulté que l'on rencontre avec cette modification, car n'ayant pas de séquence consensus définie, mais qu'une séquence favorable ou non. Le développement d'outils de prédiction fiables nécessite de documenter finement l'espace des possibles afin d'évaluer la pertinence d'éventuels faux-positifs ou faux-négatifs. L'équipe a développé des approches d'apprentissage supervisé à ces fins. Ces outils présentent une fiabilité fortement augmentée (Castrec *et al*, 2018).

I.2. Rôle de la N-myristoylation

Le rôle majeur de la myristoylation, modification quasi-irréversible, est de favoriser l'adressage d'une protéine cible à une membrane cellulaire. En absence d'autres signaux, la protéine acylée par myristoylation est spécifiquement localisée au niveau de la membrane du système endomembranaire, comme c'est le cas pour certaines kinases et petites protéines G impliquées dans la transduction du signal (exemple famille ARF/ARL). Grâce à des signaux supplémentaires et réversibles, la protéine myristoylée est redirigée vers d'autres types de membranes, comme la membrane plasmique. L'ajout d'un groupement myristate, n'est pas suffisant pour induire l'ancrage de la protéine stable à la membrane plasmique, un ou plusieurs autres signaux réversibles sont nécessaires à la stabilisation de cette interaction (Martinez *et al*, 2008; Traverso *et al*, 2013a). Il existe différents types de seconds signaux (**Figure I-3**) :

- L'affinité de la protéine myristoylée à la membrane plasmique est accrue par des liaisons électrostatiques entre un domaine chargé positivement de la protéine située au voisinage du site d'acylation et les charges négatives des phospholipides de la membrane plasmique. Ce mécanisme est utilisé, entre autres, par les protéines MARKS, Src ou encore Gag (McLaughlin & Aderem,

1995). Ce deuxième marquage est donc réversible car un changement conformationnel de la protéine permettra de masquer ce domaine hydrophobe et de casser l'interaction entre ce domaine et les phospholipides de la membrane.

- L'ajout d'un ou plusieurs groupements palmitates sur des cystéines généralement positionnées à proximité du site de myristoylation (protéines $G\alpha$, SFKs, eNOS). L'ajout d'un ou plusieurs acides gras en C16 permet d'augmenter grandement l'hydrophobicité et l'ancrage de la protéine à la membrane. Ici la réversibilité de la modification repose sur le type de liaison entre la palmitoylation et la protéine. En effet, cette modification se lie en règle générale sur des résidus cystéine, et plus rarement sur des résidus serine ou thréonine, formant une liaison thioester ou ester. Cette liaison est donc réversible par réaction chimique ou enzymatique par des enzymes tel que les palmitoyl thioesterase.
- L'interaction protéine-protéine: l'affinité de la protéine myristoylée à la membrane plasmique peut également se faire par interaction entre la protéine myristoylée et une protéine membranaire, ou transmembranaire. Comme c'est le cas de la famille de protéine ARF/ARL. Ce type d'interaction touche souvent des petite protéine G, possédant ce que l'on appelle les interrupteurs (« switch ») « myristoyl-électrostatiques », permettant la régulation de la localisation de la protéine. En effet, les protéines ayant un interrupteur hydrostatique, se trouve sous deux formes, une forme GDP, dont le groupement myristate se retrouve dissimulé dans une cavité hydrophobe de la protéine, et une forme GTP dont le myristate se retrouve à l'extérieur de la protéine. La forme GDP est localisée au niveau du cytosol, alors que la phosphorylation du GDP en GTP induit un changement conformationnel permettant l'ancrage de la protéine à une membrane, et le recrutement d'autres protéines interagissant directement avec elle, pour une localisation à la membrane plasmique.

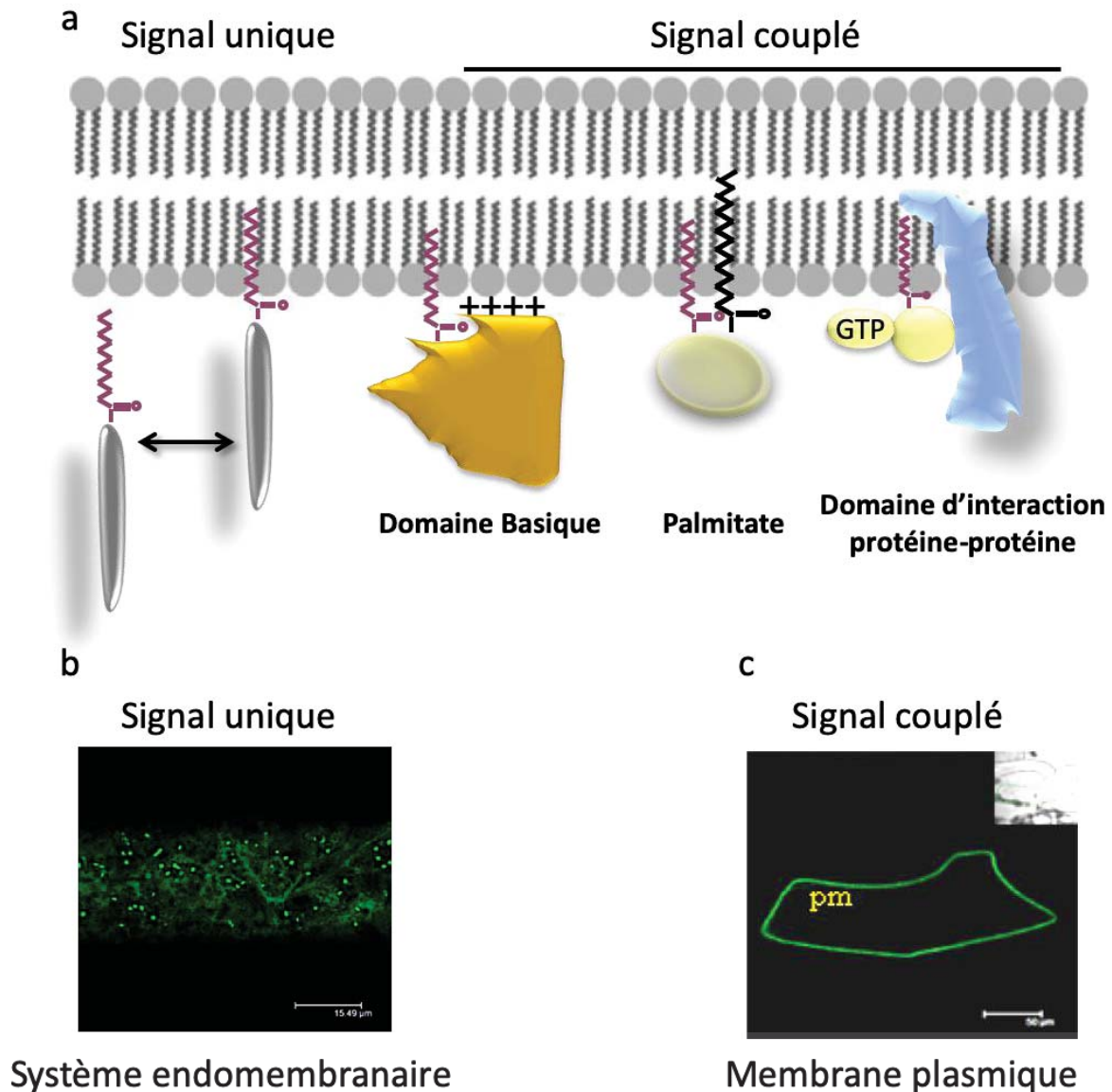


Figure I-3 : représentation des différents types de signaux d'adressage d'une protéine à la membrane utilisant la myristoylation ; (a) le premier cas étant le signal unique, une protéine ne portant que la myristoylation, ne dispose pas d'une constante d'interaction suffisante pour rester ancrée de façon stable sur une membrane. Le second cas correspond aux signaux couplés avec des seconds signaux réversibles comme un domaine basique, un ou plusieurs palmitate (C16) ou une interaction protéine-protéine. Le second signal permet d'ancrer de façon stable les protéines sur les membranes. (b) Localisation d'une protéine portant un signal unique, portant un tag GFP localisé au niveau du système endomembranaire. L'interaction de la protéine avec la membrane est trop faible pour que la protéine soit adressée à la membrane plasmique via les vésicules de sécrétion. (c) Localisation d'une protéine portant un signal couplé, portant un tag GFP localisé au niveau de

la membrane plasmique. L'ajout du second signal stabilise la protéine sur la membrane et permet son exportation du système endomembranaire à la membrane plasmique via les vésicules de sécrétion.

La combinaison des deux signaux, l'un irréversible et le second réversible, facilite la modulation fine de l'interaction entre différents types de membranes, régulant ainsi la localisation et la fonction de la protéine myristoylée (Maurer-Stroh & Eisenhaber, 2004a). Du fait de la grande diversité de membranes dans la cellule, les protéines myristoylées ont l'opportunité de se retrouver dans divers compartiments cellulaires, que ce soit au niveau de la membrane plasmique, de la membrane nucléaire et l'endosome. Les mécanismes permettant la spécificité d'interaction avec tel ou tel type de membrane restent mal connus. Grâce à un mécanisme appelé « myristoyl-switch », certaines protéines transitent via le cytosol et effectuent la navette entre le compartiment d'origine – souvent le réticulum endoplasmique - et le compartiment ciblé. C'est la nature du second signal qui permet la régulation de la localisation de la protéine.

Un grand nombre des protéines myristoylées est impliqué dans des fonctions ou événements clés de la cellule, allant de la réponse immunitaire au contrôle du cycle cellulaire. Ce mécanisme de myristoylation est donc essentiel au fonctionnement cellulaire et l'inactivation de son gène conduit à la mort cellulaire dans tous les organismes eucaryotes où ce type d'expérience a été réalisé (champignons/levures, plantes, mammifères, invertébrés).

I.3. Implication de la myristoylation dans les pathologies

I.3.1. L'utilisation de la machinerie de myristoylation de l'organisme hôte par les virus

Bien qu'il ait été rapporté la présence d'un cadre de lecture (ORF) avec une homologie significative avec les NMTs dans de nombreux génomes viraux, aucune n'a été encore caractérisée pour son activité myristoyltransférase. Pourtant la

myristoylation de nombreuses protéines virales a été décrite, y compris dans de nombreux virus animaux tels que les *arena-*, *lenti-*, *mimi-*, *picorna-*, *pox-*, *polyoma-retro-* et *rota-viridae* (Maurer-Stroh & Eisenhaber, 2004b; Diller *et al*, 2019) mais aussi des virus végétaux dont les *geminiviridae* (Mei *et al*, 2018). On dénombre jusqu'à onze protéine G-myristoylées chez le virus *Herpes simplex*. La G-myristoylation apparaît comme une modification essentielle de la pathogénicité virale (Schrager & Marsh, 1999). L'exemple connu depuis 40 ans est la myristoylation de la protéine Gag (Sengupta *et al*, 2019) qui est indispensable pour le bourgeonnement des particules virales du VIH. En effet pour déclencher la voie de réplication du virus au cours de l'infection, le virus a besoin d'adresser la protéine Gag à la membrane plasmique de la cellule infectée. Pour cela, le virus emploie la myristoylation. Une fois la protéine Gag au niveau de la membrane, elle recrute d'autres protéines et reforme de nouvelles particules virales. Dans l'ensemble de ces virus, il semblerait que ce soit la NMT de l'hôte qui soit responsable de la myristoylation de ces protéines. En détournant la machinerie de l'hôte, le virus dispose d'un moyen d'adresser ses protéines de façon spécifique au(x) compartiment(s) adapté(s).

Récemment, l'importance de la myristoylation dans la progression virale, a été exploitée. Ainsi, l'utilisation d'inhibiteurs spécifiques de la NMT s'avère efficace pour bloquer le rhume causé par les *picornaviridae* (Corbic Ramljak *et al*, 2018). Ce travail a ouvert la voie à l'élaboration de nouveaux inhibiteurs efficaces contre une variété de *picornaviridae* n'exerçant aucune toxicité pour l'hôte (Mousnier *et al*, 2018). Il a été démontré que ces inhibiteurs bloquent la protéine virale d'entrée de capsid VP0 du rhinovirus et empêchent l'assemblage de virions infectieux intacts, et donc bloque la progression de l'infection dans l'organisme hôte.

1.3.2. La myristoylation chez les protozoaires

Les protozoaires sont des eucaryotes unicellulaires, connus pour provoquer un certain nombre de maladies comme le paludisme, ou encore la leishmaniose. Comme tous les eucaryotes, les protozoaires sont des organismes dont la myristoylation est

indispensable à leur survie et à leur reproduction. La différence est que chez ces organismes la NMT n'est codée que par un seul gène, contrairement aux eucaryotes pluricellulaires. Le caractère indispensable de la NMT dans la survie de ces organismes a été démontré dans un premier temps en utilisant des ARN interférent (Price *et al*, 2003), bloquant de façon spécifique la traduction de la NMT du protozoaire et donc la myristoylation de ses protéines. Ces résultats renforcent l'intérêt de l'étude de cette modification, afin d'utiliser la NMT comme cible thérapeutique contre ces organismes pathogènes. En effet, plusieurs protozoaires sont responsables de maladies graves voire mortelles; l'absence de thérapies pour les combattre est flagrant, comme c'est le cas de plusieurs espèces de *Plasmodium* (Gunaratne *et al*, 2000), de *Leishmania major* (Panethymitaki *et al*, 2006) ou de *Trypanosoma brucei* (Panethymitaki *et al*, 2006).

1.3.3. La myristoylation dans les cancers

Comme je l'ai décrit dans les paragraphes précédents, les cibles de la NMT jouent pour la plupart des rôles cruciaux dans divers processus cellulaires, et une dérégulation de cette modification a de graves conséquences pour la cellule. C'est ce que l'on peut observer dans diverses formes de cancer où il a été décrit une surexpression des deux NMT humaines ainsi que de certains de ses substrats agissant en amont de nombreuses voies pro-prolifératives et oncogènes. Ce couplage peut suggérer que la myristoylation des cibles surexprimées nécessite la surexpression de la NMT. Les HsNMT sont d'ailleurs considérées comme biomarqueurs de plusieurs cancers comme ceux du colon, du sein et des tumeurs cérébrales (Selvakumar *et al*, 2007; Wright *et al*, 2010). Ceci suggère que la NMT est limitante dans ces types cellulaires. Mais parmi leurs substrats plusieurs sont eux-mêmes des biomarqueurs de tumeurs. C'est le cas des protéines comme c-Src et d'autres enzymes de la famille des tyrosines kinases (SFK, pour Src Family Kinase) depuis longtemps considérées comme tels et qui s'avèrent être des protéines myristoylées. Ces deux protéines une fois myristoylées et surexprimées sont impliquées dans le processus de tumorigenèse et dans de nombreux cancers (sein, pancréas, poumon, colon, gastrique) (Boutin, 1997; Ducker *et al*, 2005; Selvakumar *et al*, 2007; Wright *et al*, 2010). Il a été observé que les

inhibiteurs des NMT réduisent la prolifération et induit l'apoptose dans les lignées cellulaires de mélanome humain et murin (Bhandarkar *et al*, 2008). Ceci renforce l'intérêt d'utiliser les NMTs comme nouvelles cibles thérapeutiques.

I.3.4. La myristoylation au cours des infections bactériennes

Il existe des bactéries pathogènes possédant une machinerie de K-myristoylation permettant de détourner les fonctions de la cellule hôte eucaryote. Le catalyseur réalisant le myristoylation de la bactérie semble peu spécifique de la longueur de la chaîne acyle (on retrouve des myristates, palmitate,...) comme c'est le cas des eucaryotes lorsqu'elle est catalysée par la NMT (Lim *et al*, 2000; Röttig & Steinbüchel, 2013). En effet un certain nombre de pathogène à Gram négatif expriment des toxines de la classe Repeat in ToXin (RTX), qui sont des facteurs de virulence riches en glycine exporté via le système de sécrétion de type I (Stanley *et al*, 1998). Une fois à la membrane de la cellule hôte, ces toxines RTX forment des pores dans la cellule du système immunitaire pour altérer leur fonction (Stanley *et al*, 1998). Toutes les toxines RTX (y compris RtxA dans *Vibrio*, CyaA dans *Bordetella*, HlyC dans *Escherichia uropathogène* et LtkA dans *Pasteurella*) partagent l'acylation du groupe -amino de deux lysines. Les bactéries ont donc une machinerie de K-myristoylation catalysée par une grande famille de lysine acyltransférases dédiées. L'analyse structurale à haute résolution d'un membre de la famille, ApxC (une acyltransférase activatrice de toxines (TAAT) d'*Actinobacillus pleuropneumoniae* pathogène) a révélé que ces lysine acyltransférases sont membres de la grande famille GNAT (Greene *et al*, 2015). La nature de l'acyle gras sur les lysines semble varier, mais il correspond fréquemment à un C14:0 (myristate), bien qu'il existe également des C15, C16 et C17 saturés (Lim *et al*, 2000; Röttig & Steinbüchel, 2013).

Néanmoins, comme pour les virus, les bactéries pathogènes sont capables de détourner la machinerie de myristoylation de l'hôte eucaryote et de lui faire acyler certaines des protéines qu'elles y exportent. C'est le cas d'un certain nombre de protéines Avr bactériennes que l'on retrouve myristoylées lors de l'infection de

certaines plantes par des bactéries (Nimchuk *et al*, 2000). Pour ces types de protéines deux types de myristoylation est possible : soit sur le N-terminal de la protéine après excision de la méthionine initiatrice, soit sur un N-terminus néoformé après auto-hydrolyse de la protéine (Shao *et al*, 2003; Downen *et al*, 2009). Ce mécanisme rappelle le processus post-traductionnel dépendant de la caspase que l'on retrouve dans les cellules animales pendant l'apoptose.

Certaines bactéries pathogènes utilisent aussi le système de myristoylation de la cellules hôte comme cible de leur virulence. Ainsi, l'agent principal de la dysenterie, la bactérie *Shigella flexneri* injecte, via le système T3SS, plusieurs facteurs de virulence dont une protéase à cystéine nommée IpaJ. Le rôle de cette protéase a été longtemps occulté et n'a été caractérisé que récemment. IpaJ possède une activité protéolytique unique, puisqu'elle est capable d'hydrolyser la liaison peptidique d'une glycine si et seulement si elle porte une myristoylation ; ce mécanisme rend impossible une myristoylation post-traductionnelle de réparation. L'activité IpaJ permet à la bactérie de perturber complètement les différents mécanismes liés à la myristoylation de la cellule infectée, et plus particulièrement le trafic vésiculaire lié au réticulum endoplasmique provoquant sa désorganisation puis la mort cellulaire.

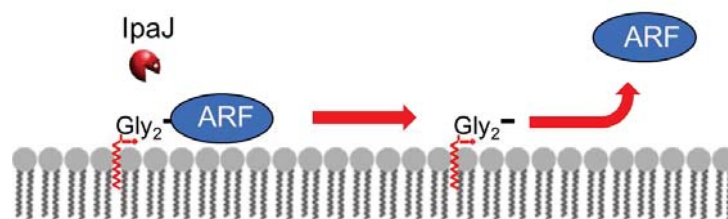


Figure I-4 : Schéma de l'activité d'IpaJ : la démystoylation des protéines par IpaJ cible la famille de petites GTPases ARF/ARL lors de l'infection par *Shigella flexneri*. IpaJ est capable d'hydrolyser la liaison peptidique en C-terminale de la glycine qui porte le groupement myristate, rendant impossible une myristoylation de réparation.

Au cours de l'infection, l'équipe du Pr Alto (Burnaevskiy *et al*, 2013, 2015) ont montré que le produit d'IpaJ agit *in vivo* préférentiellement sur la famille des petites GTPases ARF/ARL associées à l'appareil de Golgi. Les auteurs ont émis l'hypothèse que

IpaJ ne reconnaîtrait que le groupement Myristoyl-Gly2 commun à tous peptides myristoylés mais qu'une interaction directe d'IpaJ avec les protéines ARF/ARL la rendrait spécifique pour cette famille de protéine *in cellulo* (Figure I-4) (Burnaevskiy *et al*, 2013, 2015).

I.4. Le protéome myristoylé

Dans cette partie, j'évoquerai l'ensemble des approches qui ont conduit à définir progressivement les protéomes myristoylés chez différents organismes.

I.4.1. L'analyse *in vivo*

I.4.1.a. Approche à petit débit et moyen débit

Ces approches ne correspondent pas à des approches à grande échelle mais au cas par cas, et nécessitent de disposer d'anticorps spécifiques de chaque protéine cible ou de constructions transgéniques pour l'expression de la protéine fusionnée à une étiquette placée au C-terminus, le N-terminus devant rester disponible à la myristoylation. Les premières analyses mises en place en vue de la recherche de protéines myristoylées ont été réalisées en utilisant un myristate précurseur marqué au tritium, converti ensuite par la cellule en myristoyl-CoA par une acyl-CoA synthétase. Une fois marquées, ces protéines sont ensuite analysées par gel d'électrophorèse, que ce soit 2D ou 1D, car les protéines modifiées présentent une mobilité retardée par rapport à la forme non modifiée. Ce comportement différent résulte de la perte de charge positive et de l'ajout d'un groupement fortement hydrophobe de la taille de deux acides aminés. La radioactivité du tritium permet de repérer la protéine modifiée sur gel d'électrophorèse.

D'autres auteurs ont utilisé des approches alternatives, consistant par exemple à co-exprimer la NMT et la protéine ciblée chez *Escherichia coli*, puis ont confirmé leurs résultats en créant des variants où la glycine N-terminale (G2) est substituée par une alanine (variant G2A) sur la protéine cible.

Une autre approche consiste à observer les changements de localisation subcellulaire résultant de la substitution en Alanine (variants G2A) de protéines recombinantes fusionnées à une protéine fluorescente comme la GFP et exprimées *in cellulo*. Il résulte de cette seule substitution une inhibition de la N-myristoylation et probablement l'insertion d'un groupement acétyle par l'enzyme NatA (Duronio *et al*, 1990; Traverso *et al*, 2013b). Ces analyses permettent de mieux comprendre le rôle de cette modification au sein de la cellule. Le changement du premier acide aminé relocalise le plus souvent la protéine de la membrane plasmique ou du réticulum endoplasmique vers le cytosol. On infère alors que c'est l'absence de myristoylation qui est responsable de ce comportement.

Toutes ces approches ont souvent été combinées pour permettre une meilleure compréhension de l'impact biologique de la modification dans la cellule. Il faut noter qu'elles peuvent introduire des biais liés soit à la surexpression de ces protéines dans la cellule, soit au marquage radioactif suivi de l'analyse sur gel qui ne permet ni la quantification ni de qualifier la stœchiométrie de la réaction. En d'autres termes, il est souvent difficile de savoir si la protéine est minoritairement ou totalement modifiée.

Dans l'ensemble, si ces études ont pu permettre la découverte d'un catalogue non négligeable de plusieurs dizaines de protéines myristoylées, elles n'ont pas permis de disposer d'une vision globale et consolidée du protéome myristoylé.

Plus récemment, grâce aux progrès dans différentes techniques, comme l'amélioration de l'annotation des génomes de différents organismes, ou à l'expression de banques de protéines en cellules d'insectes à partir des ADNc complets, de nouvelles approches permettant le criblage des potentielles protéines myristoylées ont été mises au point. Par exemple, des auteurs ont produit une protéine reportrice (gelsoline) en fusion avec des peptides dérivés de protéines humaines déduites de 2000 ADNc, ce qui a permis de cribler 141 protéines commençant par une glycine, par marquage métabolique radioactif (Suzuki *et al*, 2010). Ceci représente le premier effort significatif de l'analyse à haut débit des protéines myristoylées dans un seul organisme.

1.4.1.b. Les marquages chimiques pour une étude globale

Avec son perfectionnement et sa démocratisation, la spectrométrie de masse s'est progressivement affirmée comme l'outil de référence pour l'études du protéome et l'étude à grande échelle des modifications des protéines. Les nouveaux outils mis au point pour l'étude des lipidations et en particulier de la myristoylation ont considérablement modifié notre compréhension des cibles de cette modification. On ne parlera ici que de N-myristoylation du fait de la découverte encore trop récente de la K-myristoylation dépendante de la NMT ; au moment de la rédaction de ce manuscrit, aucune étude à grande échelle n'a été réalisée pour identifier les cibles K-myristoylées *in vivo*.

Comme il a été décrit dans plusieurs études, la plupart des organismes sont potentiellement capables d'incorporer et d'utiliser les acides gras tels que le C14 présent dans le milieu externe en tant que précurseur de MyrCoA. Dès lors une nouvelle technologie a vu jour : il s'agit de la technologie dite de «chimie click bioorthogonale», reposant sur un précurseur possédant un groupe réactif réputé comme ne réagissant pas avec les molécules biologiques présentes dans les organismes étudiés. C'est ce qui a permis un marquage biomoléculaire hautement sélectif de la modification étudiée et plus particulièrement de la myristoylation. Cette méthode repose sur l'utilisation de lipides avec un groupement azoture ou alcyne pouvant diffuser du milieu de culture vers le cytosol, pour y être transformés en substrat de la NMT. Plusieurs types de rapporteurs ont été étudiés avec soit un groupement azoture (az) ou alcyne (Alk), notamment az-11, az-12, Alk-C13 ou Alk-C14 (**Figure I-5**). Toutes ces nouvelles approches qui utilisent ce type de marquage reposent sur quatre étapes principales qui sont le marquage bioorthogonal par les dérivés azido ou alcynes marqués par une molécule de biotine suivie d'un enrichissement par affinité des réactifs sur des billes de streptavidine, puis clivage protéolytique (typiquement la trypsine) puis analyse par spectrométrie de masse. C'est ce type d'approche qui a permis l'identification de 30 à 90 protéines myristoylées dans différents protéomes.

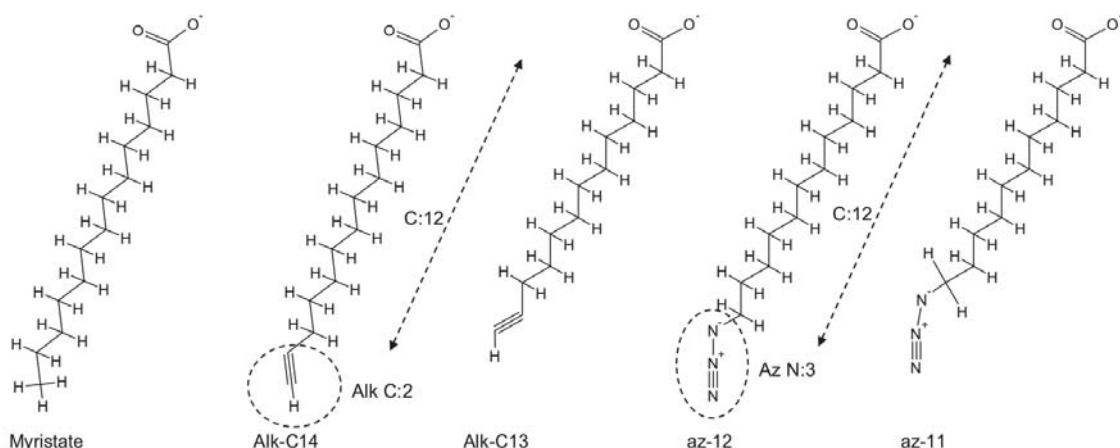


Figure 1-5 : Structures chimiques des différents précurseurs du myristate les plus utilisés. A gauche les analogues Alk-C14 et Alk-C13 avec un groupement alcyne réactif en fin de chaîne acyle, et à droite les analogues az-12 et az-11 avec un groupement azido en fin de chaîne acyle.

En utilisant cette approche, 76 protéines N-myristoylées, dont 35 nouvelles, ont pu être identifiées à partir de cellules humaines en une seule fois (Thinon *et al*, 2014). Ce travail a permis l'identification de 40 autres protéines ayant été myristoylées sur la glycine de façon post-traductionnelle, plus précisément après un clivage de la caspase au cours de l'apoptose.

Bien que puissantes, ces techniques ont leurs limites ; en effet, il a été rapidement rapporté que ces composés réactifs dérivés d'acide gras peuvent également subir une extension de chaîne *in cellulo* par LACS ou une réduction par bêta-oxydation. Le marquage Alk-C14 n'est également pas entièrement spécifique de la myristoylation dirigée par la NMT, car il implique d'autres cibles comme les protéines avec ancres à glycosylphosphatidylinositol (GPI), ou d'autres protéines inconnues, notamment des protéines palmitoylées. Tout cela implique un grand nombre de faux positifs, que l'on peut minimiser en couplant d'autres manipulations comme l'utilisation d'inhibiteurs de la NMT (par exemple l'IMP-1088, un composé commercial, voir <https://pubchem.ncbi.nlm.nih.gov/compound/132274735>) à des approches de normalisation en comparant les protéines piégées en présence ou en absence de

l'inhibiteur utilisé. Mais comme chaque cible de la NMT répond de manière spécifique à l'inhibiteur utilisé, en fonction de sa demi-vie dans la cellule ou autre, il reste difficile d'identifier avec précision les cibles pertinentes. Enfin, ce type d'approche ne permet que de travailler sur des cellules en culture, des tissus ou des cultures organoïdes uniques cultivées en présence d'un milieu de culture prédéfini complémenté par le précurseur et l'inhibiteur choisi.

1.4.1.c. Analyse protéomique par LC-MS

Une autre approche a été mise au point au sein de mon équipe d'accueil. Cette approche repose sur la simplification de la diversité du contenu protéique de l'échantillon par fractionnement des différentes parties de l'échantillon, puis d'une purification spécifique des protéines myristoylées en utilisant le caractère fortement hydrophobe du groupement ajouté (Majeran *et al*, 2018). En effet, le fait de fractionner l'échantillon a permis d'augmenter considérablement la résolution de l'analyse, réalisée par spectrométrie de masse. Dans cette étude réalisée chez *Arabidopsis*, 906 protéines des 8837 identifiées commençaient par MG et 214 appartenaient au myristoylome prédit. Mais le fait de fractionner l'échantillon ne garantit pas l'obtention de la preuve formelle de la présence de cette modification. En effet, pour cela il est nécessaire d'obtenir un spectre de fragmentation non ambigu du peptide N-terminal de la protéine. Pour accéder à ces peptides, qui portent la preuve formelle de la présence de la modification, il a fallu mettre au point un gradient spécifique et optimisé pour les peptides fortement hydrophobes en LC-MS. Afin de maximiser les chances de détection de ce type de peptide le gradient de séparation de la LC a été allongé (**Figure I-6**) déplaçant le temps de rétention de ces peptides vers des concentrations de phase mobile plus élevées que les autres peptides présents dans l'échantillon. En effet, dans des mélanges complexes de ce type ci, on retrouve tous les peptides non modifiés mais aussi ceux portant d'autres type de modification comme les phosphorylations, acétylations etc... Ce déplacement, dans le gradient, des peptides myristoylés vers des régions « moins encombrées » permet de considérablement augmenter la résolution de l'analyse.

Malgré cette optimisation, la grande complexité peptidique de l'échantillon masque une grande partie des peptides myristoylés ce qui empêche leur identification. Pour cela, l'équipe a analysé les échantillons en parallèle sur deux spectromètres de masse possédant chacun des technologies d'analyse des ions différentes. Le premier spectromètre de masse était un Triple-TOF, et le second un LTQ-Orbitrap. Comme attendu, les spectres MS/MS des peptides myristoylés obtenus par Triple-TOF ont montré une meilleure couverture de l'ion diagnostique b_1 (la glycine portant la myristoylation, un ion de faible masse, 267 Da) que ceux observés par LTQ-Orbitrap-Velos, alors que ce dernier, identifie les peptides myristoylés grâce à la présence de l'ion b_2 et de la série d'ions b suivants (**Figure I-6**).

Dans l'ensemble de cette étude 72 protéines myristoylées ont été identifiées chez *Arabidopsis*. L'utilisation des deux appareils a permis d'obtenir une meilleure couverture du myristoylome N-terminal. Il a pu être observé un recouvrement de 33% entre les analyses des deux appareils, ce qui indique leur complémentarité.

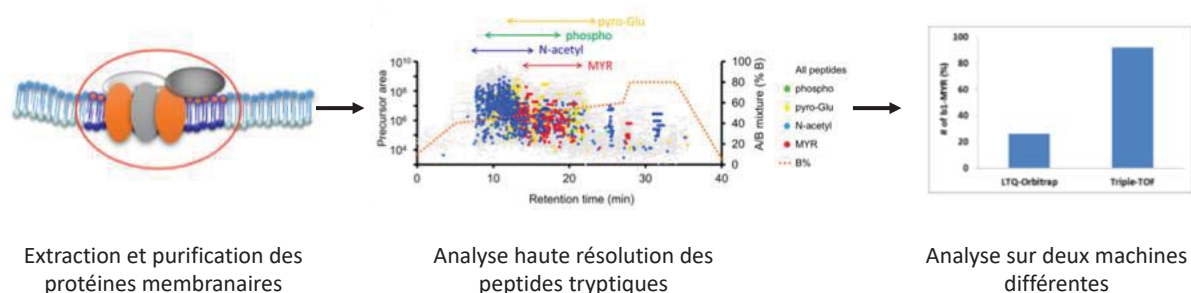


Figure I-6 : Représentation de la stratégie d'analyse par LC-MS : stratégie adaptée à l'étude du myristoylome, reposant sur l'utilisation d'un gradient optimisé pour l'étude des peptides fortement hydrophobes, et analyse sur deux machines d'architectures différentes, avec des analyseurs différent Orbitrap contre Triple TOF.

Malgré cette avancé dans l'étude du myristoylome, le nombre de protéines identifiées reste encore trop faible, compte tenu de la complexité de la préparation de l'échantillon mais aussi de l'analyse et du traitement des données.

1.5.2. L'analyse *in vitro*

1.5.2.a. Test enzymatique

Une autre approche utilisée au sein de mon équipe d'accueil quant à elle ne correspond pas à une analyse *in vivo* mais *in vitro*. Du fait que la myristoylation est réalisée en général de façon co-translationnelle, on s'attend à ce que seule la partie N-terminale de la protéine soit importante pour la réaction de myristoylation. C'est ce que les tests d'activité démontrent, une séquence de 8 acides aminés permettant de reconstituer entièrement de profil d'activité et de spécificité. Les études structurales des NMT ont permis de conclure qu'en effet ce sont les 8 premiers acides aminés qui sont reconnus par la NMT au niveau de leurs chaînes latérales ou peptidiques au sein d'une poche de reconnaissance dédiée qui prolonge le centre actif (Castrec *et al*, 2018). C'est dans ce contexte que l'équipe a construit une banque d'octapeptides correspondant aux séquences N-terminale des protéines commençant par une glycine, que ce soit pour des protéines humaine ou de la plante modèle *Arabidopsis thaliana*. Avec cette banque de données, un test d'activité à haut débit a été développé afin de pouvoir mesurer les constantes catalytiques de la réaction sur les milliers de peptides disponibles au laboratoire. Pour rappel la réaction de la NMT assemble deux substrats, le peptide ainsi que le Myr-CoA et génère deux produits, le peptide myristoylé ainsi que le CoA. Dans cette réaction, si on souhaite être indépendant de la séquence du substrat peptidique, il n'est pas possible de mesurer à moyen débit la disparition des substrat ou l'apparition des produits en continu de manière directe et quantitative, sans utiliser la radioactivité. L'approche classique réside en effet sur des tests radioactifs en discontinu basés sur le marquage du myrCoA et la séparation du peptide modifié du MyrCoA. Il a donc fallu mettre au point un test couplé avec une seconde réaction, quant à elle est mesurable de façon directe, en utilisant les propriétés d'un des produits de la réaction. L'équipe a sélectionné la réaction de myristoylation avec la réaction d'une autre enzyme, la pyruvate déshydrogénase (**Figure I-7**).

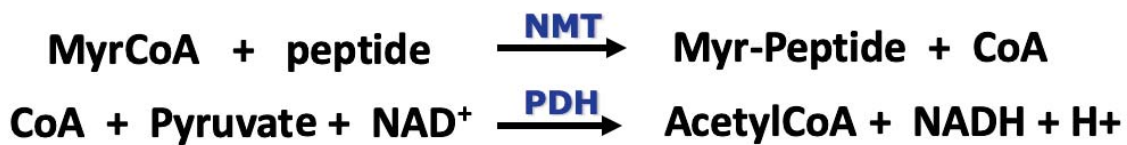


Figure I-7 : Réaction de couplage entre la réaction de myristoylation par la NMT et la réaction de décarboxylation oxydative par la PDH. Ce couplage permet de suivre l'avancé de la réaction de myristoylation en mesurant l'apparition de NADH dans le réacteur par fluorescence ou absorbance. (Traverso et al, 2013a)

La deuxième réaction est catalysée par la PDH. La PDH est une enzyme capable de catalyser la décarboxylation oxydative du pyruvate en acétyl-CoA, en présence de CoA, produit par la réaction de myristoylation, et de NAD^+ . Le produit important de cette seconde réaction est le NADH qui quant à lui peut être mesuré que ce soit par absorbance ou par fluorescence ($\lambda_{\text{abs/ex}}:340\text{nm}$ / $\lambda_{\text{em}}:465\text{nm}$). Ce couplage permet de mesurer indirectement la première réaction, et de déterminer les constantes catalytiques de chaque peptide testé, que ce soit des substrats naturels ou des peptides synthétiques permettant de mieux comprendre la catalyse de la réaction. Grâce à l'étude de plus de 2000 séquences peptidiques testées sur les NMT humaine et d'*Arabidopsis* et la mise en place d'outils d'analyse informatique, ce travail a permis de mieux appréhender la spécificité de substrat de cette enzyme et d'incrémenter les annotations des bases de données (Castrec et al, 2018).

Cette approche dispose également ses limites, comme le coût non négligeable des peptides mais aussi des réactifs ou des enzymes, et la dépendance de la solubilité des peptides, en effet certaines séquences N-terminal de certaines protéines myristoylées dispose d'acides aminés hydrophobes. Bien qu'on se rapproche des conditions physiologiques, ces méthodes *in vitro* peuvent dans certains cas ne pas refléter la réalité physiologique des processus *in vivo*.

I.4.2.b. Suivie par MALDI-TOF

Il est également possible de tester la réaction de myristoylation *in vitro* et d'analyser le résultat de la réaction par spectrométrie de masse MALDI-TOF. Cette approche a l'avantage d'être relativement rapide et permet de pouvoir tester un grand nombre de séquences. De plus, la sensibilité de l'analyse étant très haute, inférieure à la picomole, cela permet de détecter des produits de réaction en très faible quantité. En effet, certaines vitesses de réaction de certains substrats sont si faibles qu'il devient très difficile de mesurer les constantes catalytiques avec des tests enzymatiques comme décrit précédemment. Grâce à ce type d'analyse complémentaire, nous avons la possibilité de confirmer ou d'infirmer un résultat négatif, obtenu par le test de couplage vu précédemment. Un grand avantage de cette méthode est qu'elle permet aussi de localiser précisément le site de myristoylation sur le peptide grâce aux informations liées à la fragmentation de ce dernier. Cependant, le caractère hydrophobe du produit de la réaction induit dans de rares cas quelques difficultés au cours de l'analyse, que ce soit lors de la co-cristallisation du peptide et de la matrice, au niveau de la désorption de certains peptides, ou encore lors de la fragmentation. De plus cette technique ne permet pas de disposer d'une vision quantitative de la réaction. Malgré ces points négatifs cette approche est complémentaire d'autres approches et reste un très bon outil pour permettre une bonne compréhension des mécanismes réactionnels de la NMT.

I.4.3. L'analyse in silico

Au début des années 2000, les études globales du myristoylome utilisaient principalement des outils prédictifs, faute de disposer d'autres outils permettant l'étude *in vivo* de ce type de protéines. Le principal problème alors rencontré était que ces outils utilisaient des outils de calcul entraînés à l'aide d'un trop faible nombre de substrats identifiés. De plus la mauvaise connaissance du N-terminus des protéines à cette époque n'était pas favorable à une découverte extensive de nouveaux substrats. Ce n'est que plus tard, avec l'étude des autres modifications N-terminales et plus

particulièrement des N-acétylations, que ces données se sont affinées. Ce n'est ainsi qu'au cours de la dernière décennie, que grâce à des approches ciblées, que la compréhension du N-terminome et plus particulièrement celui commençant par des glycines a été mieux appréhendé.

Avec les différentes avancées que ce soit des études globales du N-myristoylome, *in vivo* des différentes équipes travaillant avec les différentes approches disponibles, ou les approches *in vitro*, les outils de prédiction se sont de plus en plus perfectionnés. Un de ces outils de prédiction, créé dans mon équipe d'accueil ; TermiNator3, permet en fonction de la séquence N-terminal de la protéine, de déduire le type de modification qui affectera son N-ter. (Martinez *et al*, 2008).

Ces outils sont puissants et suggèrent que le nombre de protéine N-myristoylées restant à identifier *in vivo* est élevé, et qu'il reste donc encore beaucoup à découvrir. Mais ces outils ont aussi leurs limites, comme par exemple le taux très important de faux positifs qui rend délicat l'interprétation de ces prédictions.

I.4.4. L'analyse structurale

Une autre approche, repose sur les analyses structurales de l'enzyme responsable de cette modification. En effet, disposer de structures de ces enzymes permet de mieux comprendre leur fonctionnement mais aussi leur sélectivité de substrat. Des structures obtenues à haute résolution du site actif est d'une très grande aide pour bien comprendre comment le substrat est recruté et comment ce dernier interagit avec l'enzyme. A ce jour, 74 structures cristallines à haute résolution sont disponibles sur la PDB, et sur différents organismes (*Saccharomyces cerevisiae*, *Aspergillus fumigatus*, *Candida albicans*, *Plasmodium vivax*, *Trypanosoma brucei*, *Leishmania major* et *Homo sapiens*). On a pu observer que le cœur catalytique de 400 acides aminés de la NMT est fortement conservé chez les différents organismes. Seule l'extension N-terminale diffère (en longueur et en séquence), mais elle n'est probablement pas ordonnée en solution et n'est pas impliquée dans la sélection du substrat, comme le démontre l'étude de variants de la NMT dépourvus de ce domaine. Le caractère polybasique de

cette extension est en revanche récurrent et pourrait contribuer à l'interaction avec les phosphates des ARN du ribosome. De plus, dans l'équipe, nous avons réussi à cristalliser la HsNMT1 en présence de ses deux substrats (peptide et MyrCoA), ses deux produits ainsi qu'avec un intermédiaire de réaction, ce qui a permis de décortiquer l'ensemble du mécanisme réactionnel (Dian *et al*, 2020). De plus les données obtenus grâce aux différentes structures obtenues à haute résolution avec les substrats ou les produits ont permis de mieux comprendre le positionnement de ces derniers dans le site actif et mieux comprendre les séquences favorables ou non à la myristoylation, acide aminé par acide aminé (**Figure I-8**).

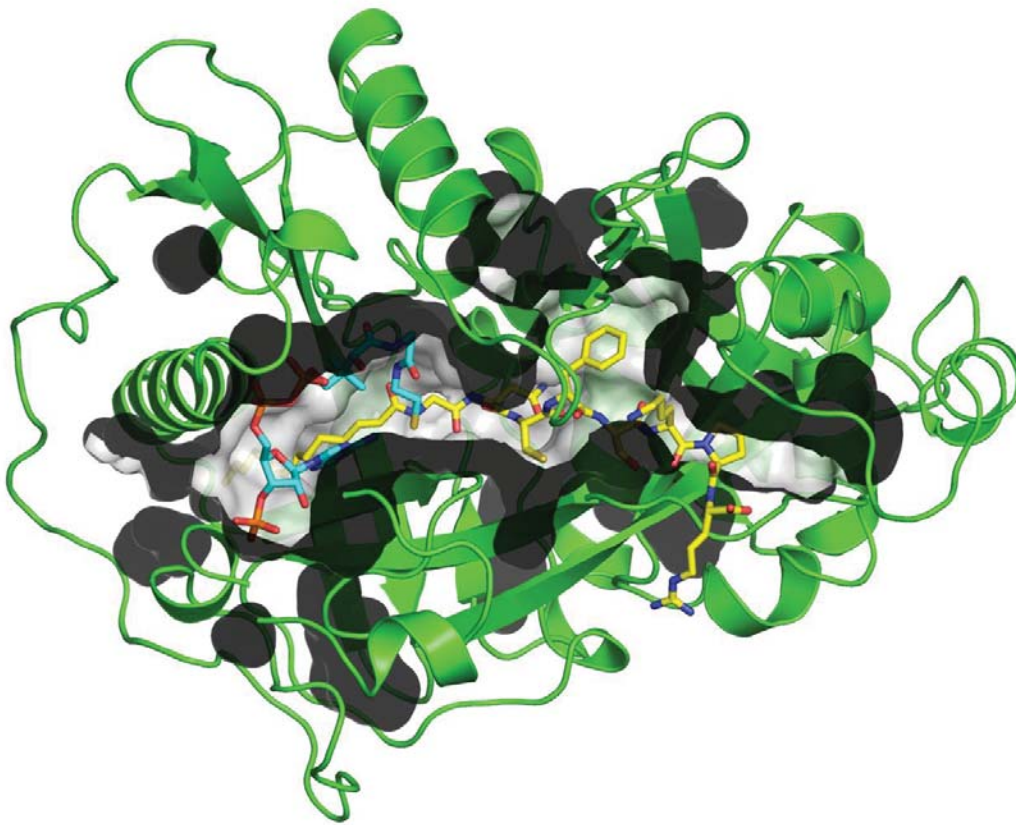


Figure I-8 : Structure haute résolution de la NMT en présence de ses deux produits ; le peptide myristoylé (jaune) et le CoA (bleu), ce qui permet de bien visualiser les différentes « poches » du site actif acceptant des acides aminés de différentes tailles.

I.5. Défis actuels concernant l'étude de la myristoylation et alternatives

I.5.1. Les défis de longue date

Comme nous avons pu le voir tout au long de ce chapitre, l'étude de la myristoylation, quel que soit l'organisme, est un défi certes de longue date mais toujours d'actualité. Que ce soit par la nature du groupement ajouté sur la protéine, augmentant très fortement l'hydrophobicité de cette dernière, mais également du fait de la complexité des N-terminaux dans les cellules. En effet, l'ajout d'un acide gras de 14 carbones sur un peptide change drastiquement ses propriétés physico-chimiques et son comportement en solution. Il sera donc beaucoup plus difficile de l'extraire, de le purifier et de l'analyser. On pense qu'un grand nombre de ses protéines fortement hydrophobes sont perdues au cours de la première étape d'extraction des protéines ; même en utilisant des détergents, il reste compliqué de manipuler ce type de protéine.

Une difficulté supplémentaire concerne la complexité des modifications N-terminales ; en effet le N-terminal des protéines est un haut lieu de modifications. Pour qu'une protéine soit G-myristoylée, il faut dans un premier temps que la protéine soit substrat d'une première enzyme, la méthionine aminopeptidase, puis que la séquence soit favorable à la myristoylation. Une fois la méthionine retirée, cette nouvelle protéine doit avoir une glycine en position deux pour pouvoir être myristoylée, mais il se trouve qu'une autre enzyme peut modifier des protéines de ce type comme c'est le cas de certaines N-acétyl-transférases. Une compétition entre la N-myristoyl-transférase et la N-acétyl-transférase peut se produire dans la cellule, dans certains cas ce qui rend plus complexe l'étude de cette modification (Castrec *et al*, 2018).

Par ailleurs, l'outil le plus utilisé pour l'étude de la myristoylation est la spectrométrie de masse dite bottom-up ; ce type d'analyse - qui permet d'identifier les protéines d'un échantillon grâce à l'analyse des peptides issus de digestion par des protéases (le plus souvent par la trypsine) - est entièrement dépendant de la couverture de séquence de la protéine. En effet, du fait que la modification soit au niveau du N-terminal, la seule façon d'authentifier cette modification en utilisant cette technique

est d'identifier et d'analyser le peptide N-terminal, dont la présence est évidemment moins probable que celle de n'importe quel autre des dizaines de peptides tryptiques associés à une protéine dont la présence d'un seul d'entre eux suffit à caractériser une protéine.

Comme nous avons pu le voir au cours de cette introduction, différentes stratégies ont été mises en œuvre pour l'étude de cette modification ; ce n'est que lors de la dernière décennie que les plus grandes avancées ont été réalisées. Bien que le nombre de protéines myristoylées identifiées soit en nette augmentation dans les différents organismes étudiés, nous sommes encore loin de les avoir identifiées toutes. Rappelons que, chez l'Homme, seule une centaine de protéines myristoylées a été identifiée, alors que selon les estimations cette modification toucherait un peu moins de 2% du protéome eucaryote, ce qui représente un nombre significativement plus fort (4 à 5 fois). Ce différentiel important peut s'expliquer par la faible abondance de certaines protéines portant cette modification ou par le fait que l'expression de certaines des cibles ou même la présence de la modification est probablement tissu- ou cellule-dépendante.

Pour ce qui concerne l'étude de la K-myristoylation, ce sont globalement les mêmes difficultés que l'on rencontre mais s'y ajoute la réversibilité de la modification. En effet, contrairement à la G-myristoylation la K-myristoylation est réversible et on connaît plusieurs protéines capables de retirer cette modification. C'est le cas de certaines sirtuines.

I.5.2. Les défis abordés dans mon projet de thèse

Comme nous avons pu le voir au cours de cette introduction, la myristoylation est une modification indispensable au bon fonctionnement de la cellule, de par son rôle mais aussi par le fait qu'elle touche des protéines impliquées dans diverses voies indispensables à la cellule. De plus, la myristoylation est impliquée dans de nombreuses pathologies, faisant de la NMT une cible thérapeutique particulièrement intéressante. En dépit de cette importance fonctionnelle, et malgré l'existence de quelques analyses

à grande échelles, les outils expérimentaux actuels n'ont permis l'identification que d'un nombre limité de cibles de la NMT. En effet selon les estimations réalisées dans le laboratoire cette modification toucherait aux alentours de 2% du protéome humain, soit plus de 500 protéines, *versus* une centaine découverte chez l'humain.

Au cours de ma thèse, je me suis intéressé à cette modification dans un contexte cancéreux, où il a été remarqué que l'expression de la NMT dans un certain nombre de cancer (comme celui du colon cerveau, vésicule biliaire...) était fortement augmentée en parallèle avec le niveau d'expression d'un certain nombre de ses substrats, impliqué dans des voies majeures du cycle cellulaire, tel que les protéine pp60c-src, ou encore pp60-yes. Dans ces cancers, l'augmentation concomitante des niveaux de protéines myristoylées et de la NMT, suggère donc que les niveaux de myristoylation de l'ensemble du protéome sont altérés. Compte-tenu du rôle crucial que la myristoylation joue au cours du cycle cellulaire et les voies de transduction des signaux, les suivis qualitatif et quantitatif exhaustifs de la modification constituent une approche alternative dans le cadre du diagnostic et du suivi de traitement de certains cancers. Le défi majeur de l'étude de la myristoylation est lié (i) à la nature irréversible de la modification résistant au traitement chimique ou (ii) à la nécessité d'un marquage métabolique avec des analogues et de ce fait incompatible avec une approche de diagnostic ou de suivi thérapeutique.

Dans la première partie de ce mémoire de thèse, je décris comment j'ai élaboré une nouvelle approche protéomique pour l'étude du myristoylome, permettant l'analyse de ce dernier directement dans des échantillons biologiques. Cette nouvelle approche repose sur l'utilisation d'une protéase découverte lors de l'initiation du projet, ayant une activité unique ; l'hydrolyse de la liaison peptidique de la glycine N-terminale si et seulement si elle N-myristoylée. Cette approche a comme objectif de débloquent les protéines modifiées dans les extraits, éliminer le caractère hydrophobe du peptide et générer de nouveaux N-termini. Tous les peptides ainsi modifiés caractériseront seulement les protéines modifiées par IpaJ (un peptide commençant par l'acide aminé 3, après la perte de la glycine 2 myristoylée), et ce dans des conditions

normales et cancéreuses. Pour cela, il a fallu compléter l'étude structure fonction d'IpaJ, dont les données étaient alors très fragmentaires.

Au cours de la deuxième partie de ma thèse, j'ai étudié les mécanismes réactionnels de la NMT, via différentes approches : une étude des structures 3D haute résolution de la NMT en présence de ses substrats et de ses produits, et une étude enzymologique afin de suivre les réactions et d'en déduire ses constantes catalytiques.

Cette étude m'a permis de mettre au jour un deuxième mécanisme réactionnel inédit de cette enzyme, qui correspond à la catalyse de la réaction de myristoylation sur la lysine proche du N-terminal (décrits dans le manuscrit #1 et la première partie du manuscrit #2).

Dans ce cadre, j'ai été amené à mettre au point un protocole de discrimination des produits G-myristoylés et K-myristoylés pour des peptides dont le résultat de la réaction était ambigu (peptide commençant par GK dont la masse G-Myr et K-Myr est identique) grâce à l'utilisation de l'enzyme IpaJ (seconde partie du manuscrit #2).

J'ai enfin mis à profit les résultats obtenus au cours de l'étude de la NMT pour mettre au point et proposer de nouveaux inhibiteurs de la NMT reposant sur le mécanisme de réaction des substrats K-myristoylables (troisième partie du manuscrit #2).

Chapitre II

II. IpaJ, nouvel outil pour l'étude de la myristoylation

II.1. Étude et validation de la protéase IpaJ comme outil d'analyse de la myristoylation

L'objet de ce chapitre est d'évaluer si la protéase IpaJ peut être prise en considération comme un outil pertinent dans le cadre d'une étude de la myristoylation. Dans cette partie, j'ai donc pris le parti de produire, purifier la protéase, de l'étudier *in vitro* sur des modèles pour valider son activité et sa spécificité, d'appliquer ensuite son utilisation à une analyse globale protéomique, puis enfin d'évaluer comment son expression endogène pouvait aussi aider à répondre à mes questionnements.

II.1.1. Production de la protéase IpaJ

Pour rappel (voir introduction), la protéine IpaJ est d'abord sécrétée puis réabsorbée par l'hôte chez lequel la toxine va exercer son action. Les 30 premiers résidus N-terminaux sont prédits par les logiciels de prédiction de structure secondaire (Jpred) pour former une hélice α hydrophobe (séquence 11-28) d'une vingtaine de résidus. L'équipe américaine ayant caractérisé la protéase IpaJ à l'origine en 2013 (Burnaevskiy *et al*, 2013) a montré que ces résidus ne sont pas essentiels à la fonction *in vivo* et *in vitro*. La pré-séquence pourrait correspondre au peptide signal nécessaire à la sécrétion d'IpaJ lors de l'infection. En conséquence, les constructions que j'ai insérées dans un vecteur d'expression plasmidique (pET28M) codent pour la séquence d'IpaJ de *Shigella flexneri* (Identifiant UniProt [Q54150](#)) dépourvue de ses 29 premiers résidus. Après clonage et expression de ces constructions, j'ai rencontré plusieurs problèmes liés à la faible solubilité de la protéine, rendant sa purification compliquée et son stockage à court terme difficile, la protéine ayant une forte tendance à la précipitation. J'ai donc produit une protéine de fusion His(6)-GST-IpaJ pour augmenter la solubilité lors de la purification, mais cette astuce ne s'est pas non plus avérée suffisante pour permettre la solubilité à long terme de la protéine. Une étude bio-informatique a révélé qu'IpaJ correspond à un homologue éloigné du domaine peptidase du transporteur ABC bi-fonctionnel ComA de *Streptococcus* impliqué dans

la détection du quorum (« quorum sensing »). Une équipe japonaise a réussi à cristalliser le domaine peptidase de ComA et à en déterminer la structure en augmentant la solubilité du domaine par substitution du motif hydrophobe de surface potentiellement impliqué dans la reconnaissance du substrat en aval du site de clivage (Ishii *et al*, 2010). En procédant à la mutation des résidus équivalents (L97N, L99N et C102S) d'IpaJ dans le contexte de la séquence IpaJ 30-259, la protéine IpaJ ainsi modifiée (nommée IpaJ 3M) est plus stable en solution après clivage du Tag His-GST, mais aussi au niveau du stockage de la protéine à long terme. Ainsi insérée dans le vecteur d'expression plasmidique inductible pET30M, j'ai cherché à produire de manière soluble les protéines GST-IpaJ-WT (IpaJ-WT) et 3M, afin de développer un test d'activité d'IpaJ. IpaJ étant une cystéine protéase j'ai également produit deux mutants inactifs en mutant la cystéine du site actif sur les variants IpaJ-WT et IpaJ-3M. L'équipe américaine ayant caractérisé IpaJ avait pu déterminer que l'une des bases de la triade catalytique était la cystéine C64, j'ai donc muté le codon par mutagenèse dirigée pour le changer en codon sérine et produire un variant inactif (IpaJ-C64S). Un exemple de préparation type d'IpaJ séparée par gel d'électrophorèse dénaturant (SDS-PAGE) permet d'apprécier la pureté de l'échantillon final (Figure supplémentaire 3c de l'article 2, partie IV).

II.1.2. Développement d'un test de suivi de l'activité d'IpaJ

Aucun test d'activité n'avait été encore développé jusqu'à présent pour permettre d'identifier et évaluer l'activité d'IpaJ. Des études pilotes réalisées par Sarah Ciccone et Benoît Castrec, deux post-doctorants de l'équipe en collaboration avec Jean-Pierre Le Caer (ICSN, Gif sur Yvette) juste avant mon arrivée ont montré que des peptides myristoylés pouvaient être observés facilement en spectrométrie de masse MALDI-TOF. Cette approche est particulièrement adaptée pour détecter la masse de peptides dans un mélange peu complexe. J'ai donc mis au point un test d'activité de la protéine IpaJ en m'assurant que les réactifs utilisés étaient compatibles avec l'analyse en spectrométrie de masse par MALDI-TOF. Le cahier des charges était que l'échantillon à analyser ne devait pas (i) comporter une concentration saline trop

importante et (ii) contenir une haute concentration de produits incompatibles avec la spectrométrie de masse ou pouvant gêner l'analyse (Glycérol, DMSO, détergents ioniques type Triton, polymères, réducteurs...). J'ai ainsi été amené à substituer le Triton par du cholate, aboli l'ajout d'albumine et vérifié par un test d'activité classique comparatif que l'activité n'était pas affectée dans ces conditions.

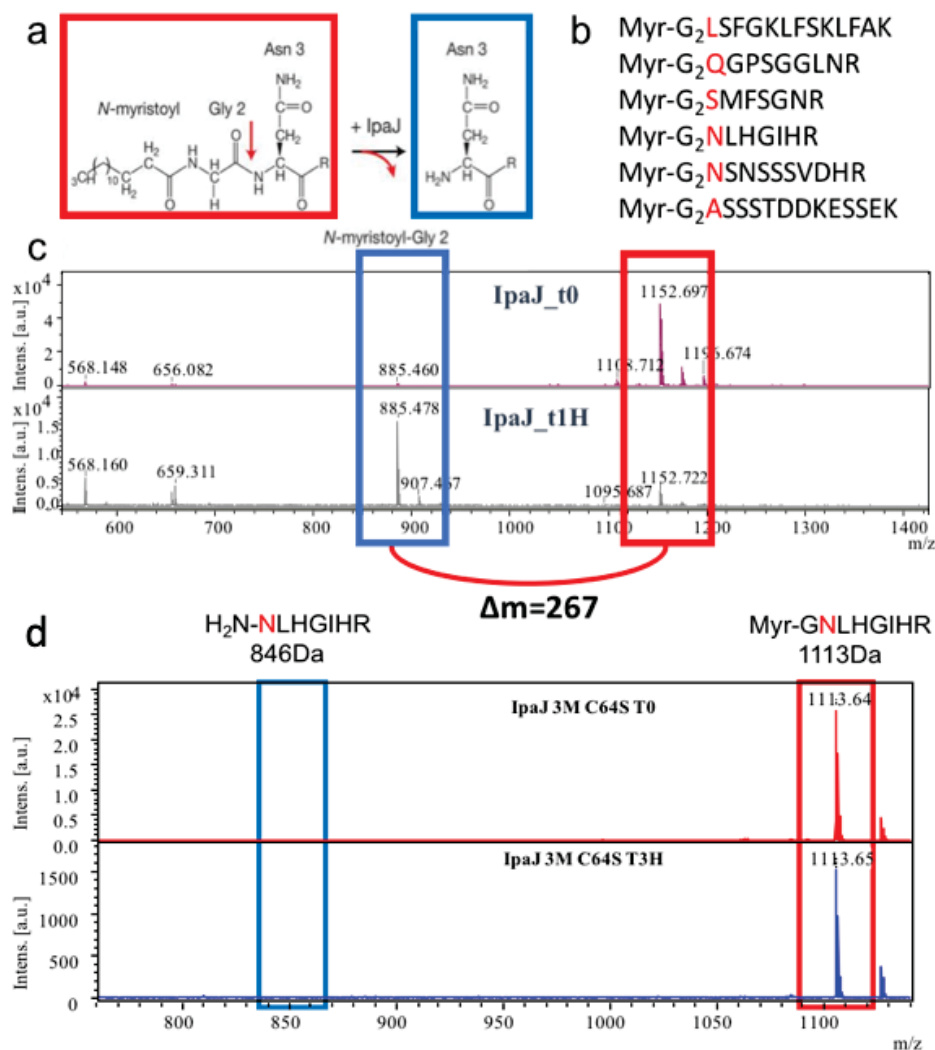


Figure II-1 : Test de démyristoylation de peptides myristoylés synthétiques avec IpaJ. a) réaction d'IpaJ sur un peptide myristoylé. b) liste des peptides myristoylés testés et positifs à la digestion par IpaJ. c) suivi dans le temps de la digestion du peptide Myr-GQGPSGGLNR par spectrométrie de masse MALDI-TOF en présence d'IpaJ WT d) confirmation de l'inactivation du le mutant IpaJ 3M C64S sur un peptide Myr-GNLHGIHR.

La réaction catalysée par IpaJ correspond à l'hydrolyse de la liaison peptidique entre la glycine portant le groupement myristate et l'acide aminé en position 3 (**Figure II-1a**). Si IpaJ est en capacité d'hydrolyser le peptide, on s'attend à une diminution de l'intensité du pic correspondant au peptide myristoylé en faveur de l'apparition et de l'augmentation dans le temps de l'intensité du pic correspondant au peptide sans le groupement N-myristoyl-Gly. J'ai utilisé une série de peptides modèles, tous issus de coupures tryptiques théoriques du N-terminus de protéines naturelles coupées après acétylation chimique (**Figure II-1b**). L'idée était de se rapprocher au maximum des conditions réelles d'un extrait cellulaire après hydrolyse tryptique. C'est ce qui explique la nature du dernier résidu de la série peptidique (R) et leur taille variable (8-14 résidus). Le pic attendu correspond donc à un pic possédant une perte de masse de 267 Da. Dans un premier temps, j'ai vérifié l'activité d'IpaJ-WT dans ces conditions (**Figure II.1c**). Une fois les conditions validées, j'ai testé le mutant IpaJ-3M qui s'est avéré également actif. Par la suite, les expériences ont été réalisées exclusivement avec la construction GST-IpaJ 3M qui s'avère beaucoup plus stable au cours du temps ; j'ai en effet pu vérifier avec ce test qu'au bout de quelques jours, conservée à -20°C avec 55% glycérol, la préparation d'IpaJ-WT perdait totalement son activité, au contraire du variant IpaJ-3M. Une fois l'activité protéolytique des protéines recombinantes purifiées démontrée, j'ai appliqué ce test d'activité aux six autres peptides synthétiques pseudo-tryptiques myristoylés (**Figure II.1b**), me rapprochant ainsi d'une analyse protéomique classique. Ces peptides présentent des séquences variables et toutes éloignées de la séquence N-terminale de la famille ARF1-5 considérée comme les substrats naturels d'IpaJ. Pour l'ensemble des peptides testés, j'ai observé (i) l'apparition d'un pic possédant une différence de masse de 267 Da, correspondant à la perte du groupement Myr-Gly, et (ii) la diminution de l'intensité en parallèle du pic correspondant au peptide myristoylé. Ces résultats préliminaires indiquent que les protéines recombinantes IpaJ-WT et IpaJ-3M purifiées sont actives et capables d'hydrolyser des séquences différentes des séquences N-terminales des substrats décrits (Burnaevskiy *et al*, 2013, 2015).

Dans le but de préciser si la protéine IpaJ est également capable d'hydrolyser d'autres produits issus de la myristoylation par la NMT, j'ai fait évoluer mon test à petite échelle à une plus grande échelle en profitant d'une grande banque de peptides disponibles dans l'équipe. Dans ce contexte, il est important aussi de noter que la myristoylation chimique des peptides est à la fois onéreuse et risquée. J'ai en effet observé qu'il était impossible de disposer en solution de certains des peptides myristoylés acquis dans les tampons compatibles avec la détermination de l'activité d'IpaJ. J'ai observé systématiquement un voile de précipitation dans différentes conditions et l'impossibilité de détecter un signal par spectrométrie MALDI-TOF. Cet effet étant très possiblement lié au fort caractère hydrophobe porté du groupe myristate, nous avons testé l'hypothèse que le couplage des deux réactions pourrait constituer une solution permettant de remédier en partie à ces difficultés.

II.1.2. Développement du test d'activité de myristoylation par la NMT couplé à la démyristoylation par IpaJ

L'équipe dispose d'une banque constituée de plus de 2000 octapeptides synthétiques dérivés des séquences N-terminales naturelles des protéines des protéomes d'*Arabidopsis thaliana* et humaines. 837 peptides de cette banque ont été déjà montrés positifs à la myristoylation *in vitro* par les NMT de ces organismes, comme l'indique une étude antérieure de l'équipe, qui consistait à déterminer les substrats potentiels *in vivo* des NMTs (Castrec *et al*, 2018).

A partir de ces résultats de l'équipe, j'ai réalisé une analyse des séquences peptidiques positives à la myristoylation, que j'ai compilées afin d'obtenir la fréquence des différents résidus à chaque position d'un octapeptide (**Figure II-2**). On remarque dans ce schéma la forte représentation du résidu cystéine, habituellement rare, à toutes les positions 3-9 par rapport aux autres acides aminés possibles. Ceci traduit la probable S-palmytoylation des substrats de la NMT et on retrouve la séquence favorable de type $G_2-[^{\wedge}PRDELTV]_3-X_4-[^{\wedge}DEGR]_5-[SGCA]_6-[KR]_7-X_8-X_9$, (où $^{\wedge}$ définit les résidus exclus). Cette fréquence de cystéines ne traduit donc pas directement le lien

avec la myristoylation mais plutôt la séquence d'évènements que beaucoup de protéines myristoylées subissent, la double acylation permettant de renforcer l'ancrage à la membrane plasmique (voir introduction).

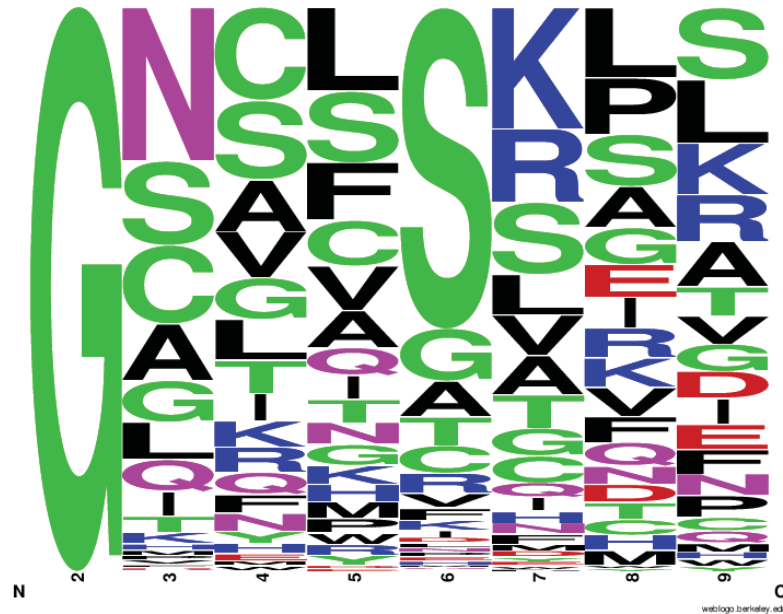


Figure II-2 : Compilation des séquences myristoylables. La compilation est présentée sous forme de logo illustrant la fréquence des acides aminés retrouvés à chaque position des séquences des 837 octapeptides positifs à la myristoylation par la NMT présents dans la mini-banque de l'équipe d'accueil.

J'ai poursuivi la caractérisation de la sélectivité de substrat de la protéase IpaJ, en utilisant la banque d'octapeptides positifs à la myristoylation. J'ai d'abord adapté mon test d'activité en ajoutant le second substrat de la NMT (Myr-CoA) et la NMT dans le mélange de démyristoylation par IpaJ afin de vérifier si IpaJ est capable d'hydrolyser ou non d'autres substrats de la NMT que ARF (**Figure II-3ab**). Cette étape mime la séquence réalisée par la cellule lors de l'infection par Shigella et suivie par l'hydrolyse par la toxine IpaJ.

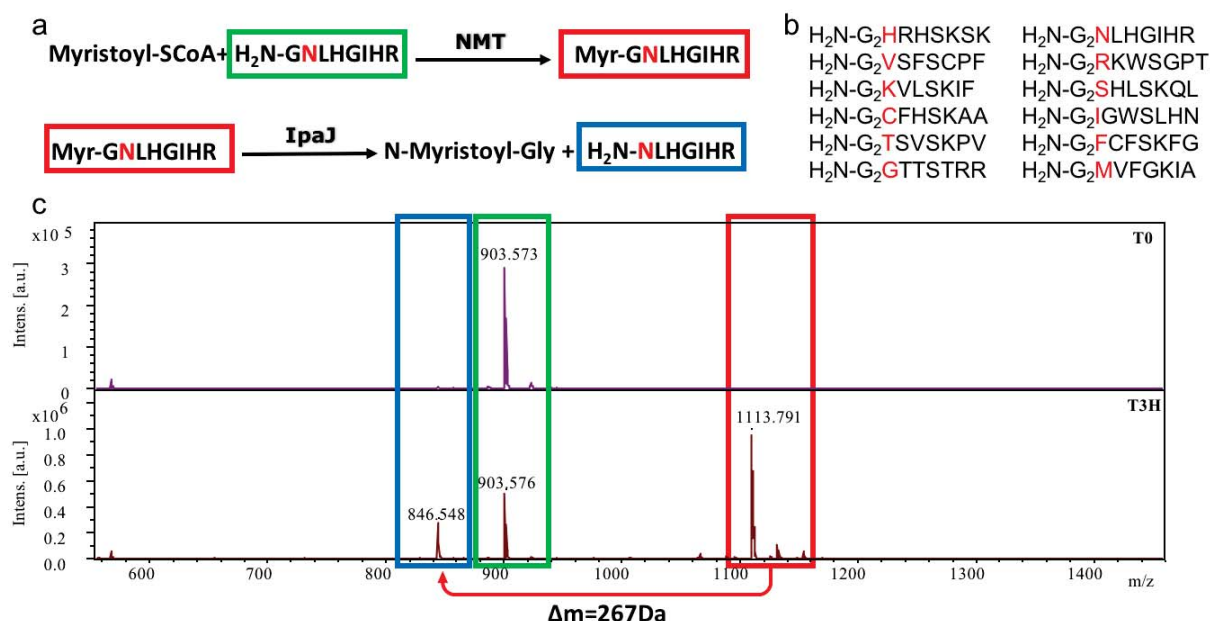


Figure II-3 : Test d'activité d'IpaJ couplé à la myristoylation par la NMT. a) Schéma de la réaction de couplage entre la NMT et IpaJ. b) Liste des peptides synthétiques sélectionnés correspondants à des N-terminaux de protéines, positifs à la myristoylation et démyristoylation testés dans cette étude c) Spectre de masse MALDI-TOF correspondant à la réaction de couplage du peptide GNLHGIHR. On peut observer la présence (encadrement vert) de l'octapeptide et l'apparition des pics correspondants aux peptides Myristoylés et peptide ayant perdu le groupement Myr-Gly (encadrés en rouge et bleu, respectivement).

Si le couplage des deux enzymes fonctionne effectivement, on attend à l'issue de la réaction (3 heures) une diminution de l'intensité du pic m/z correspondant au peptide non-myristoylé, en faveur d'une augmentation du pic m/z correspondant au peptide excisé du groupement Myr-Gly. La **Figure II-3c** illustre les spectres obtenus avec le peptide GNLHGIHR dans ces conditions. On observe que l'on détecte à la fois la présence du peptide myristoylé et encore de celui non modifié, mais dans une intensité moindre (**Figure II-3a,c**). Ce protocole m'a permis de tester 15 nouveaux peptides choisis de manière à évaluer l'ensemble des acides aminés possible en 3^{ème} position en aval de la glycine myristoylé tout en criblant une grande variété de séquence sur les autres positions. Sur ces 15 peptides, douze peptides ont été hydrolysés par GST-IpaJ 3M. Les spectres obtenus pour les trois peptides (GPRCSKLS,

GDVLSTHL et GETMSKRL) testés négatifs montrent que la réaction de myristoylation par la NMT a échoué. Les valeurs de k_{cat}/K_m traduisant l'efficacité de la NMT humaine pour ces deux peptides s'avèrent très faibles par rapport aux autres peptides substrats. En effet, les structures 3D de la NMT en complexe avec plusieurs peptides substrats myristoylés ainsi que les analyses enzymatiques suggèrent que les résidus Aspartate/Glutamate et Proline sont tous deux défavorables en position trois. Pour réaliser la myristoylation de tels substrats répondant faiblement à la NMT, il aurait fallu que j'effectue la réaction dans des conditions de concentrations de NMT bien supérieures. Je ne peux exclure également que ces peptides soient inefficacement désorbés de la matrice par le laser.

AA2	AA3	%AA3 sur positif NMT	AA4	%AA4 sur positif NMT	AA5	%AA5 sur positif NMT	AA6	%AA6 sur positif NMT	AA7	%AA7 sur positif NMT	AA8	%AA8 sur positif NMT	AA9	%AA9 sur positif NMT
G	N	27,12	C	16,01	L	13,14	S	56,75	K	21,51	L	12,43	S	12,78
G	S	14,93	S	14,93	S	11,59	G	9,32	R	13,02	P	10,04	L	11,23
G	C	13,98	A	9,32	F	9,8	A	6,33	S	12,78	S	9,2	K	9,08
G	A	10,04	L	8,84	C	8,6	T	5,14	L	7,17	A	7,29	R	8,36
G	G	7,41	V	8,12	V	8	C	4,9	V	7,17	G	6,57	A	8
G	L	6,57	G	5,85	A	6,45	R	3,58	A	6,93	E	5,85	T	5,14
G	Q	5,73	T	5,5	G	5,26	V	2,51	T	5,97	I	5,5	V	5,02
G	I	4,3	I	4,54	Q	4,9	F	2,15	G	4,9	R	5,26	G	4,9
G	T	2,87	K	4,42	I	4,54	K	1,79	C	4,78	K	5,14	D	4,78
G	K	2,03	Q	4,18	T	4,54	I	1,31	Q	2,63	V	5,14	I	4,54
G	H	1,08	F	3,58	R	4,18	D	0,96	I	2,39	F	4,66	E	4,42
G	M	0,96	N	2,99	N	4,06	L	0,84	H	2,15	Q	4,3	F	4,42
G	P	0,84	H	2,39	K	3,58	N	0,84	N	1,91	N	3,23	N	3,82
G	V	0,84	R	2,27	H	2,51	P	0,84	F	1,79	D	3,11	P	3,82
G	F	0,48	Y	1,91	M	2,51	E	0,72	M	1,19	T	2,99	C	2,39
G	D	0,36	M	1,31	Y	2,03	H	0,72	D	1,08	C	2,75	Q	2,15
G	R	0,36	E	1,19	P	1,67	M	0,48	Y	0,84	H	2,75	M	1,55
G	E	0,12	W	1,19	W	1,67	Q	0,48	E	0,72	M	2,63	H	1,31
G	W	0	P	1,08	E	0,72	W	0,24	W	0,6	W	0,6	W	1,31
G	Y	0	D	0,36	D	0,24	Y	0,12	P	0,48	Y	0,6	Y	0,96

Figure II-4: Tableau récapitulatif des résultats obtenus par les différents tests d'activité NMT. Le tableau représente la fréquence (en %) de chaque acide aminé pour les positions 2 à 9 des 837 peptides positifs à la myristoylation in vitro disponible au laboratoire. En vert sont colorés les acides aminés présents dans les peptides positifs au test de myristoylation/démyristoylation ; en jaune les acides aminés identifiés comme positifs à l'hydrolyse par *IpaI* dans une autre étude (Burnaevskiy et

al, 2013, 2015); en rouge les peptides qui n'ont pas été prouvés comme substrats car la réaction de myristoylation n'a pas été détectée.

Dans le tableau récapitulatif (**Figure II-4**) est représentée la fréquence (en %) de chaque acide aminé à chaque position d'un peptide positif à la myristoylation *in vitro* par les NMTs en utilisant la séquence des 837 N-terminaux positifs à la myristoylation connus dans le laboratoire. Dans ce tableau, j'ai représenté en vert les acides aminés présents dans les séquences testées, indépendamment de la nature de la séquence du peptide. J'ai procédé de la même façon, en colorant en jaune les acides aminés des peptides identifiés *in vitro* et *in vivo* par l'équipe américaine du Pr Alto lors de la caractérisation initiale d'IpaJ. Ce tableau indique que nous avons testé indépendamment la majorité des résidus possibles aux positions 3, 4, 5 et 6 en aval de la glycine.

Je conclus que mon étude de sélectivité d'IpaJ, reposant sur ce nouveau test de couplage de myristoylation/démyristoylation *in vitro*, démontre qu'IpaJ est capable d'hydrolyser une très grande variété de séquences et suggère fortement qu'IpaJ n'est sélective que de la glycine modifiée Myr-Gly du peptide substrat. La protéase IpaJ semble donc adaptée à un usage dédié pour cliver tous les myristates portés par une glycine N-terminale.

II.1.3. Sélectivité d'IpaJ pour des peptides synthétiques N-acétylés

J'ai ensuite testé la capacité d'IpaJ à hydrolyser différents peptides synthétiques comprenant des glycines N-terminales N-acétylées en utilisant le test d'activité développé dans la partie précédente (**Figure II-3**). La séquence en aval de la glycine N-acétylée des peptides utilisés dans ce test fait partie des séquences favorables à la N-myristoylation par la NMT. Ces peptides ont été choisis pour refléter la compétition possible *in vivo* entre la N-acétylation et la N-myristoylation de certains N-terminaux substrats des NMTs et des NATs de type A. Les données obtenues en MALDI-TOF avec le peptide Ac-GNCFSKPR indiquent clairement après incubation en présence de GST-IpaJ mutée 3M la présence d'un unique pic correspondant au peptide N-acétylé non hydrolysé (**Figure II-5**). La même réaction a été reproduite en présence d'une quantité

10 fois supérieure d'enzyme sur une très longue durée de 6 heures d'incubation. Les spectres obtenus sont identiques. Ces données montrent qu'IpaJ est sélective des N-terminaux comprenant une Glycine N-myristoylée et suggèrent qu'IpaJ peut être utilisée pour distinguer des peptides qui seraient partiellement myristoylés (clivage de la Myr-Gly) et N-acétylés (pas de clivage).

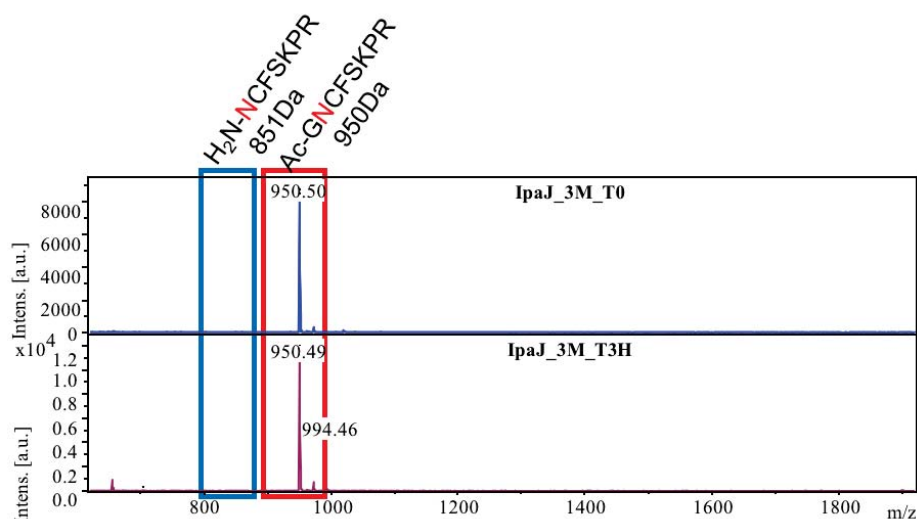


Figure II-5: Test de déacétylation par GST-IpaJ3M suivi par analyse en spectrométrie de masse MALDI-TOF, au cours du temps du peptide Ac-GNCFSKPR. On peut observer la présence (encadrement rouge) du pic m/z du peptide N-acétylé et l'absence du pic m/z correspondant au peptide déacétylé ayant perdu le groupement N-acetyl-Gly (encadré en bleu).

II.2. Utilisation d'IpaJ comme outil biotechnologique pour l'étude de la myristoylation

II.2.1. Fractionnement cellulaire

L'intérêt de cette étape réside sur le fait que dans la cellule les protéines myristoylées ne représentent qu'une très faible partie du protéome, que ce soit en nombre ou en abondance. De ce fait, le fractionnement cellulaire permet d'augmenter la résolution de l'analyse en décomplexant l'échantillon. Je me suis basé sur une méthode bien établie (**Figure II-6**) pour fractionner mes échantillons (Cox & Emili, 2006), cette méthode repose sur un fractionnement par centrifugations successives permettant de séparer les différents compartiments cellulaires. Je me suis focalisé sur

trois compartiments : le cytosol, le microsome, et les mitochondries. Dans le cytosol, on peut retrouver toutes les protéines myristoylées possédant le « myristoyl switch » (cf introduction) ; le microsome concentrera le plus grand nombre de protéine myristoylée. Je me suis aussi intéressé aux mitochondries, que ce soit leurs fractions membranaires mais aussi les fractions solubles afin d'explorer s'il y a bien des protéines myristoylées dans ce compartiment, comme les données le suggèrent (Meinzel *et al*, 2020).

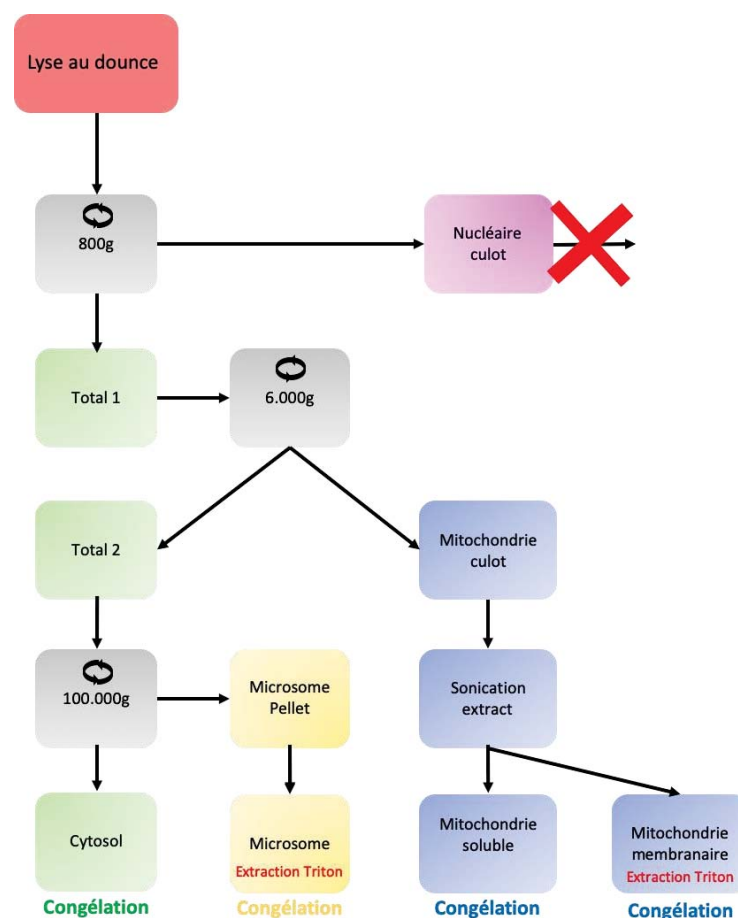


Figure II-6 : schéma de la stratégie de fractionnement. La méthode est basée sur (Cox & Emili, 2006), reposant sur la centrifugation successive de l'échantillon à des vitesses angulaire croissantes afin de séparer les compartiments cellulaire un par un.

J'ai ensuite effectué une électrophorèse SDS-PAGE afin de décomplexer les fractions (Figure II-7a), pour simplifier la complexité protéique de l'échantillon et retirer tous les

produits incompatibles avec la spectrométrie de masse (Triton, etc...). J'ai ensuite hydrolysé à la trypsine les échantillons sur gel pour enfin les analyser par LC-MS. Comme espéré, le fractionnement améliore fortement la résolution de l'analyse (**Figure II-7b**) ; en effet, quand on analyse uniquement la fraction totale on identifie un total de 1872 protéines, alors qu'après fractionnement on obtient 3719 protéines sur le premier fractionnement. De plus, un grand nombre de ces protéines identifiées sont observées uniquement dans une seule fraction, ce qui valide la méthodologie. Parmi les 3719 protéines identifiées, 89 et 79 respectivement, dans le fractionnement 1 et 2, sont décrites comme myristoylées *in vivo* (**Figure II-7c**). De plus, quand on observe la répartition de ces protéines dans les compartiments cellulaires analysés, on se rend compte que la plupart de ces protéines se situent au niveau du microsome et au niveau de la mitochondrie. Malgré ces résultats très encourageants, la myristoylation effective (c'est-à-dire l'identification du peptide N-terminal) d'un nombre faible de protéines a été identifié dans ces échantillons (8 pour le fractionnement 1 et 2 pour le fractionnement 2). Cependant, c'est en proportion ce qu'on attend puisque le peptide N-terminal représente en moyenne environ un vingtième des peptides issus de l'hydrolyse tryptique.

Ce résultat montre bien la difficulté d'une analyse ce type : la grande difficulté à identifier les peptides N-terminaux des protéines myristoylées et ce malgré l'utilisation de gradient de séparation en chromatographie liquide avant analyse en spectrométrie de masse adapté aux peptides hydrophobes.

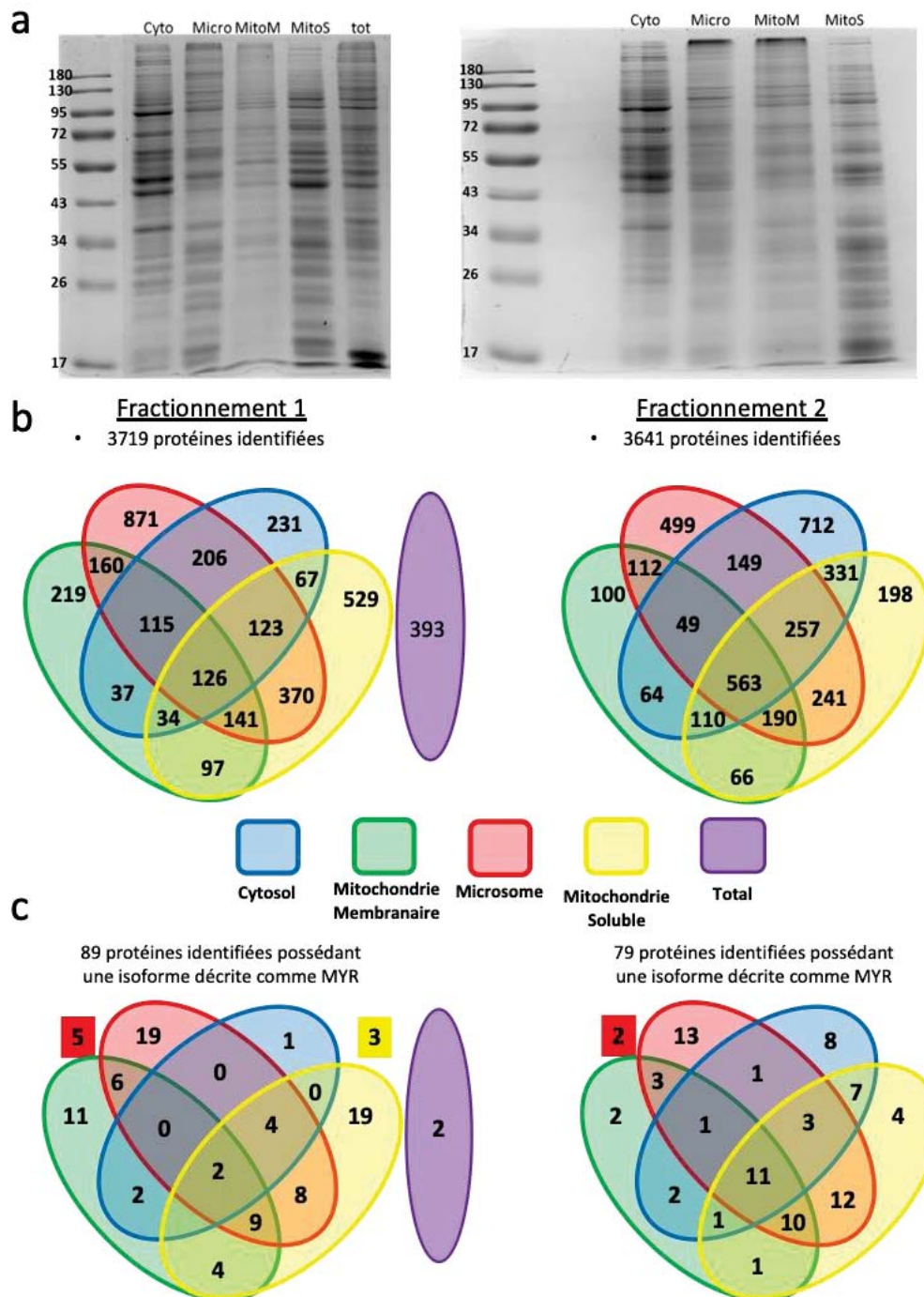


Figure II-7 : Analyse du fractionnement cellulaire et de la myristoylation des protéines dans les fractions issues des différentes étapes. a) analyse par électrophorèse dénaturante, b) identification croisée des protéines, c) analyse des protéines prédites comme myristoylées.

II.2.2. Traitement des lysats par IpaJ

Dans cette seconde section de mon étude, j'ai testé différents lysats protéiques de certaines lignées cellulaires avec IpaJ afin de vérifier la capacité à débloquent le N-terminal en retirant des glycines myristoylées. Pour ce faire, j'ai disposé d'une douzaine de lignées cellulaires humaines d'origine tumorale ou immortalisée. Dans un premier temps, j'ai réalisé des immuno-blots sur les différentes lignées en culture afin de détecter la HsNMT1 – la forme très majoritaire de NMT – et d'évaluer son abondance dans différents types cellulaires (**Figure II-8a**). La NMT se distribue sous deux formes issues d'épissage alternatifs conduisant à des démarrages sur des codons méthionines éloignés. Il en résulte deux formes de taille différente - dites « longue » et « courte » - toutes les deux fonctionnelles car possédant le domaine catalytique entier. Au contraire de la forme longue, la forme courte ne possède pas l'extension N-terminale supposée renforcée la fixation au ribosome. La forme courte permettrait à la NMT d'agir post-traductionnellement – par exemple sur les cibles issues de coupures protéolytiques - alors que la forme longue serait impliquée dans l'activité co-traductionnelle. J'ai pu observer que la distribution n'est pas identique dans tous les types cellulaires (**Figure II-8b**). Dans chaque extrait, j'ai réalisé une analyse protéomique de type « shotgun » (appelée aussi « bottom-up ») afin d'évaluer le nombre de substrats myristoylés détectables. Pour rappel, l'identification d'une protéine myristoylée dépend de l'identification univoque du peptide N-terminal portant la modification et ce avec des scores d'identification pertinents. Ceci nécessite des spectres MS2 présentant une série d'ions b satisfaisante. J'ai choisi les lignées cellulaires pour lesquelles je réussissais à identifier le plus de substrat myristoylé, comme la lignée U87, mais aussi une lignée comme RT4 dont les immuno-blots ont montré une faible expression de la NMT1. Quatre types cellulaires ont été testés : A2780R (carcinome ovarien), JIMT-1 (carcinome du sein), U87 (glioblastome) et RT4 (papillome de la vessie).

J'ai pu observer que le nombre d'identifications des protéines myristoylées évalué par le filtre unique de l'identification du peptide N-terminal acylé dépendait uniquement du type cellulaire et pouvait varier fortement entre deux types (**Figure II-**

8c). Le nombre d'identifications semble sans aucune corrélation avec le niveau d'expression de la NMT1 ; on peut l'observer par exemple pour les lignées RT4 (nombreuses identifications, peu de NMT, **Figure II-8d**). Une analyse plus large des peptides identifiés dans chaque analyse protéomique indique que certaines protéines sont identifiables uniquement dans un seul type cellulaire. Ces éléments traduisent bien que la présence et l'abondance des protéines myristoylées varie significativement d'une ligné à l'autre. Au total, 21 peptides myristoylées issus d'autant de protéines différentes ont été identifiées dans quatre types cellulaires (**Figure II-8d**).

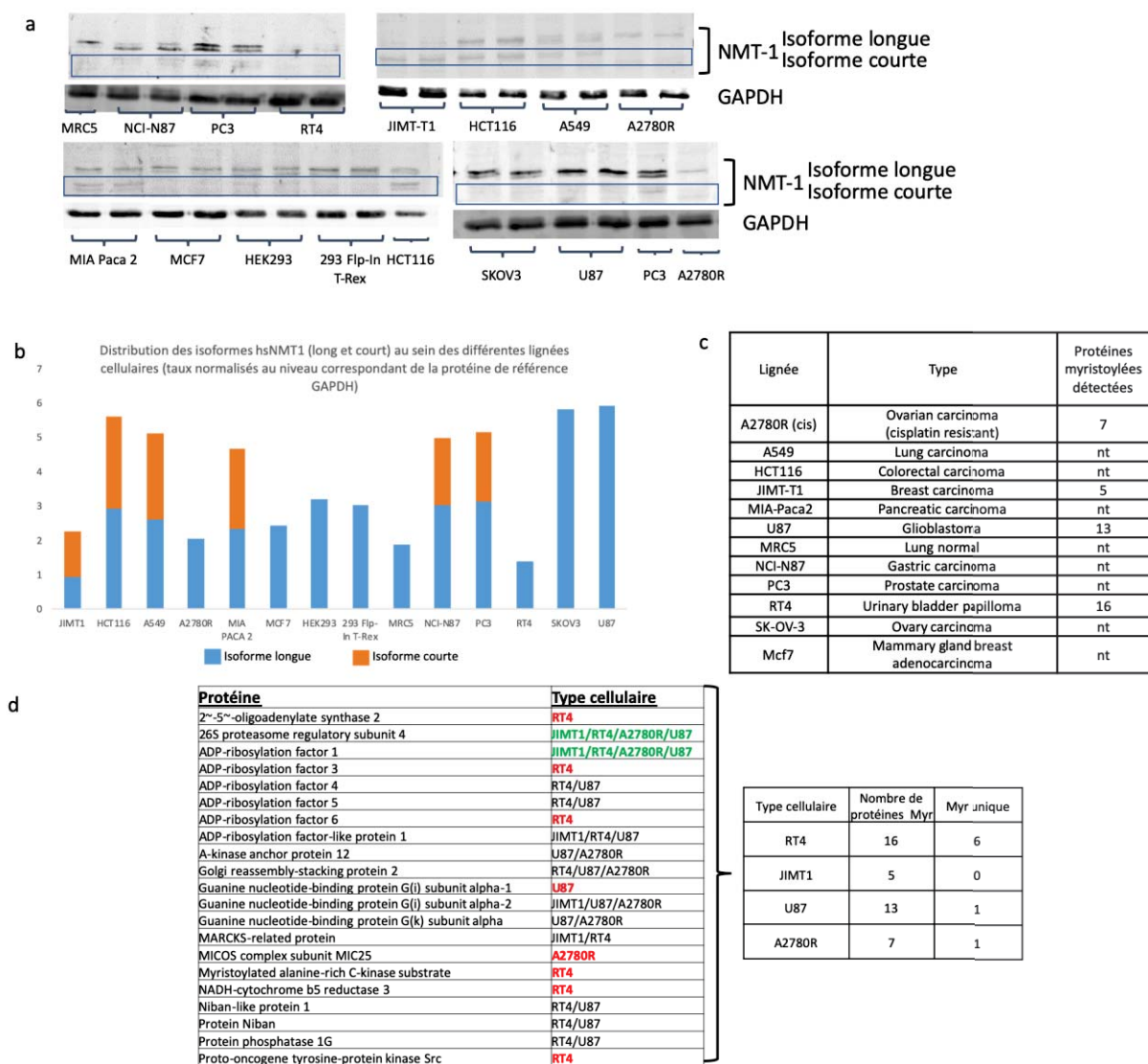


Figure II-8 : Expression de la NMT en fonction du type cellulaire : a) Immunoblot avec détection avec un anticorps polyclonal dirigé contre la HsNMT1. On peut observer deux formes d'HsNMT1, longue et courte (la forme courte est encadrée en bleu), b) quantification relative des isoformes au sein des différentes lignées cellulaires c) tableau des types cellulaires disponibles au laboratoire ; le nombre de protéines myristoylées identifiées dépend du type cellulaire, (nt = non testé) ; d) protéines myristoylées identifiées dans différents types cellulaires et unicité par échantillon de leur identification parmi les quatre types.

Ayant à disposition quatre lignées sélectionnées, la première stratégie que j'ai adoptée (Figure II-9a) pour cette étude a reposé sur l'utilisation du lysat fractionné afin d'augmenter la sensibilité de l'analyse. Une fois ce lysat fractionné je me suis concentré

sur les fractions membranaires et celles du système endomembranaire, où la probabilité d'identifier des protéines myristoylées est plus élevée (Majeran *et al*, 2018). Ces fractions ont ensuite été traitées par IpaJ afin de libérer les protéines du groupement Myr-Gly, et augmenter leur solubilité. A l'issue de cette étape, j'ai procédé à une acétylation chimique des protéines avec de l'acide acétique deutère (« lourd », +3Da par rapport à la forme « légère », naturelle) afin de bloquer tous les groupes N-terminaux libres de l'échantillon. Ensuite, les protéines ont été hydrolysées en présence de trypsine en vue de réaliser un enrichissement par chromatographie d'échange de cation SCX. Ce type d'enrichissement permet la rétention des peptides chargés positivement issus de la digestion tryptique, et l'élution en tête de colonne des peptides N-terminaux ayant été modifiés *in vivo* ou chimiquement avec l'acétylation. Cette approche a pour avantage de permettre d'évaluer la compétition entre la N-acétylation naturelle et la myristoylation. En effet, on retrouverait dans l'analyse les protéines acétylées naturellement, celles non modifiées dans la cellule avec un acétyle lourd sur la méthionine ou sur l'acide aminé deux en cas d'excision de la méthionine par la MetAP, ou encore les protéines initialement myristoylées, portant un acétyle lourd sur l'acide aminé trois (excision de la méthionine plus perte du groupement Myr-Gly2).

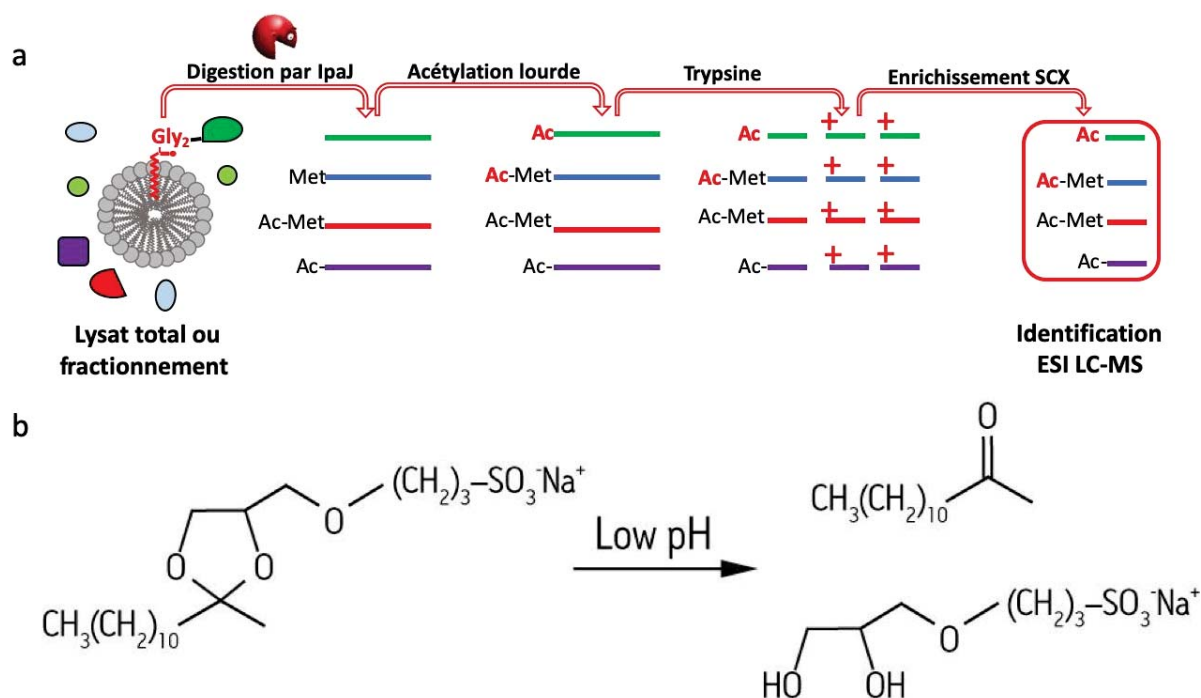


Figure II-9 : Stratégie n°1 utilisation d'IpaJ comme outil biotechnologique pour l'analyse des protéines myristoylées ; a) la stratégie comporte quatre étapes i) traitement des échantillons par IpaJ, ii) acétylation lourde des N-terminaux protéiques ; iii) trypsine des échantillons, iv) enrichissement des N-terminaux par chromatographie de cation SCX. b) structure du détergent RapiGest détergent hydrolysable en condition acide ($pH < 2$).

L'analyse protéomique des échantillons montre que la perte de glycines myristoylées caractéristiques n'est pas obtenue ; l'identification des mêmes peptides myristoylés qu'avant traitement est encore plus troublante. Bien que tous les inhibiteurs de cystéine protéase aient été retirés des tampons de fractionnement pour assurer l'activité d'IpaJ (une protéase à cystéine qui serait sinon inhibée), aucune évaluation de l'activité d'IpaJ dans le mélange n'a été réalisée lors des analyses réalisées avant de passer les échantillons en SCX. Pour l'expliquer, j'ai émis d'abord l'hypothèse que l'inhibition d'IpaJ pouvait provenir du Triton X-100 utilisé dans les tampons, je l'ai donc remplacé par le RapiGest (Min *et al*, 2015) un détergent adapté pour l'utilisation d'enzymes incompatibles avec les détergents classiques (**Figure II-9b**). Le RapiGest présente en effet la particularité d'être hydrolysable à pH acide et est éliminé au cours de l'étape d'acidification avant l'analyse en spectrométrie de masse (ajout du TFA). Cependant, cette approche ne m'a pas permis d'améliorer notre évaluation de l'activité

d'IpaJ, toujours apparemment non fonctionnelle selon les critères définis ci-dessus. Une seconde hypothèse que nous avons alors formulée est qu'il est peut-être plus complexe d'hydrolyser la liaison peptidique de la glycine myristoylée sur une protéine complète. Le repliement tertiaire pourrait en effet bloquer l'accessibilité du site de clivage, ce qui ne serait pas le cas sur un peptide. J'ai donc modifié la stratégie n°1 en déplaçant l'intervention d'IpaJ après la digestion tryptique. L'action sur des peptides a pour objectif d'augmenter la probabilité d'observer une activité d'IpaJ.

Cette nouvelle stratégie (n°2, **Figure II-10**) repose donc sur l'utilisation d'IpaJ sur les peptides issus de la digestion tryptique. La première étape consiste à utiliser la trypsine directement sur les échantillons – comme dans une analyse protéomique « bottom-up » classique - afin d'hydrolyser les protéines en peptides. J'ai ajouté une étape supplémentaire qui consiste à réaliser une acétylation chimique des peptides issus de la digestion tryptique avec un acétyle deutéré dit « lourd ». Le choix d'utiliser de l'acétyle marqué au deutérium repose sur la nécessité de bloquer les N-terminaux et les amines libres (des lysines) des protéines afin de faciliter un éventuel marquage des amines néoformées après l'action d'IpaJ, et pouvoir le différencier des acétylations naturelles (**Figure II-10**). Il faut noter que seuls les peptides N-terminaux naturellement acétylés (la majorité) portent un acétyle « léger », ce qui permet de les identifier. Une fois ce marquage effectué, j'ai incubé l'extrait peptidique acétylé avec IpaJ.

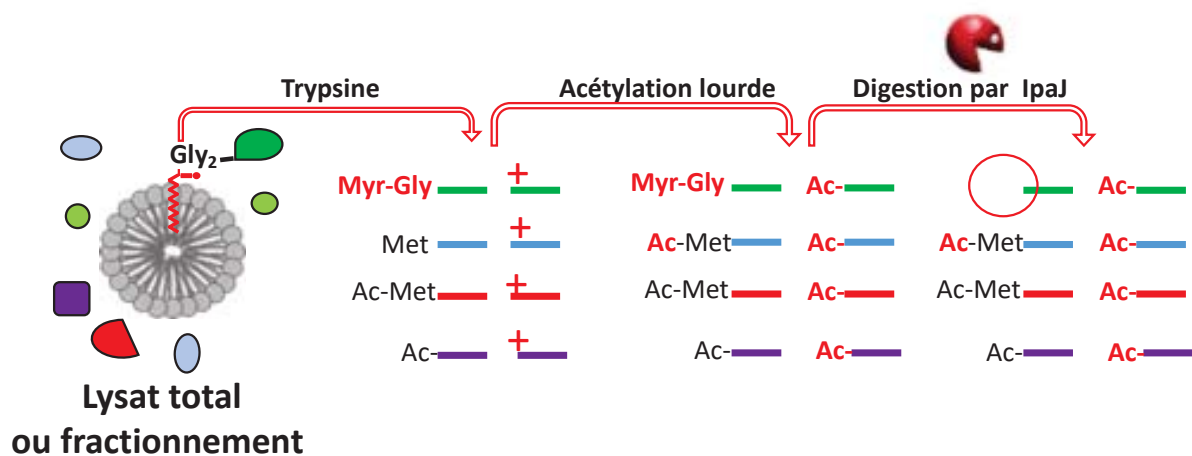


Figure II-10 : Stratégie n°2 pour l'étude du myristoylome via l'utilisation de la protéase IpaJ ; La première étape consiste en la préparation de l'échantillon soit une lyse seule soit par fractionnement cellulaire après la lyse, suivie du traitement de l'échantillon par la trypsine. Une fois les protéines protéolysées, l'acétylation permet de bloquer toutes les amines primaires libres. La dernière étape correspond au traitement de l'échantillon par IpaJ formant un nouveau N-terminus libre

J'ai ensuite analysé par spectrométrie de masse les échantillons traités ou non par IpaJ. J'ai pu observer d'abord que les protéines identifiées comme myristoylées dans l'échantillon non traité avec IpaJ n'étaient plus présentes dans l'échantillon incubé en sa présence. J'ai introduit dans le programme d'identifications MASCOT une nouvelle modification (la perte de glycine N-terminale) afin de reconnaître les substrats potentiels d'IpaJ et les inclure dans la liste des peptides validés. Avec ce protocole, j'ai identifié une protéine ayant subi la perte de sa glycine myristoylée car possédant l'acide aminé successif à son N-terminus (**Figure II-11**). Ces éléments démontrent bien qu'IpaJ est actif non seulement sur des peptides purifiés et issus de la myristoylation *in vitro* mais aussi sur des mélanges complexes de peptides.

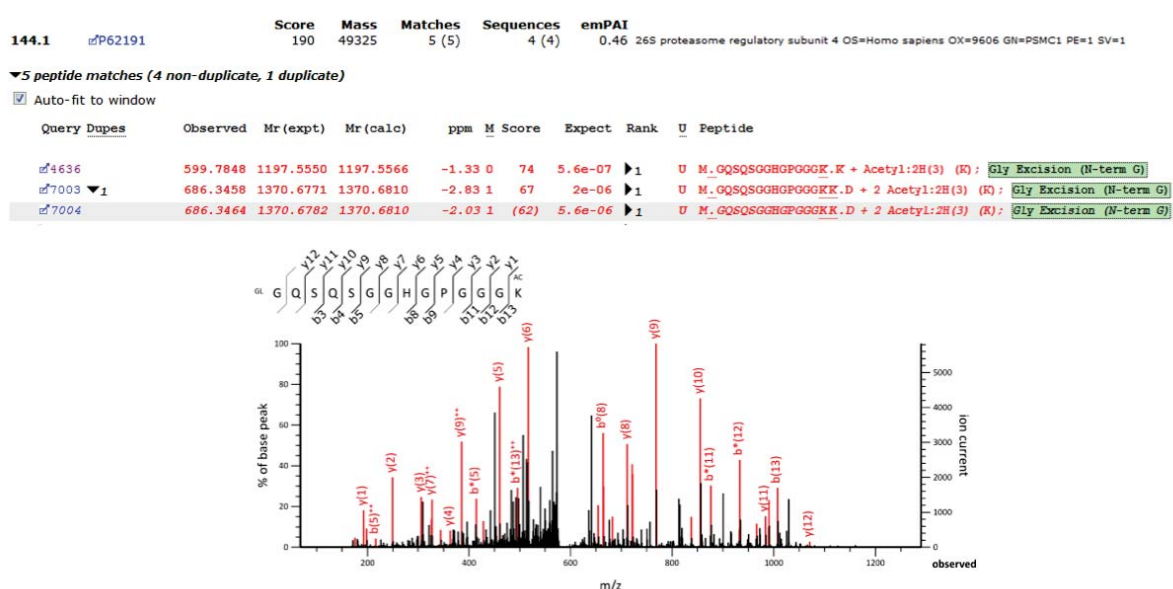


Figure II-11 : Spectre de fragmentation MS2 et données d'identification de la sous unité régulatrice 4 du protéasome 26S démyristoylée par IpaJ dans un extrait cellulaire. Haut de la figure : identification par le logiciel AMSCOT de la sous unité RPT4 du protéasome 26S ayant perdu la glycine dans un lysat

traité par IpaJ. Bas de la figure, spectre MS2, ions issus de la fragmentation identifiés et validés.

Le fait que l'on ait identifié qu'une seule protéine ayant perdu la glycine myristoylée peut s'expliquer par la perte du caractère hydrophobe du peptide. En effet, les peptides ayant cette modification ont tendance à être élués en fin de gradient lors de la chromatographie liquide qui se trouve en amont du spectromètre de masse. En fin de gradient le nombre de peptides élués est moindre ; la perte de ce groupement induit une élution plus précoce et les peptides d'intérêt sont noyés dans la multitude des peptides internes. Comme seuls les produits de la réaction d'IpaJ possèdent une amine primaire réactive et libre en N-terminal, cette particularité peut être utilisée pour réaliser un marquage spécifique N-terminal pour purifier ces peptides. Un avantage remarquable de la stratégie n°2 par rapport à la stratégie n°1 est donc de produire les peptides issus de la digestion d'IpaJ avec une amine réactive libre en N-terminal qui permet l'enrichissement spécifique de ces peptides.

II.3. Enrichissement des N-terminaux libre issus de la digestion par IpaJ des protéines myristoylées

Pour enrichir de façon spécifique les peptides anciennement myristoylés j'ai opté pour l'utilisation d'un marquage chimique reposant sur la réactivité du succinimide. En effet cette molécule réagit avec les amines primaires et permet de transférer différents types de groupements en formant une liaison amide. C'est ce type de dérivé que j'utilise pour les marquages chimiques avec des acétyles deutérés. J'ai donc la possibilité d'utiliser cette propriété pour marquer le N-terminus libre des peptides via cette réaction. Plusieurs choix de marquage sont possibles. J'ai choisi d'abord le marquage au (N-succinimidyl-oxycarbonyl-methyl) tris (2,4,6-trimethoxyphenyl) phosphonium bromide (TMPP-Ac-OSu, TMPP); ce marquage au TMPP est souvent utilisé dans ce cadre (Gallien *et al*, 2009) et a déjà été mis à profit dans ce contexte (Vaca Jacome *et al*, 2015). A pH légèrement alcalin (8,2), seul le N-

terminus qui possède un pK inférieur (7,7) réagit puisqu'il est non protoné, au contraire des amines de lysines qui ne sont pas réactives du fait d'un pK très basique (10,5) conduisant à 100% de protonation (**Figure II-12a**). En effet, l'ajout du groupement au peptide permet de décaler l'élution du peptide dans une zone moins « encombrée » du gradient de la chromatographie liquide (LS) et augmenter d'autant les chances de l'identifier.

J'ai donc d'abord essayé ce marquage sur des peptides issus de la digestion tryptique de l'albumine sérique de bœuf (BSA) afin d'estimer l'efficacité de ce type de marquage (**Figure II-12b**). J'ai pu observer que l'efficacité de ce marquage dans mes conditions était faible malgré l'utilisation de ratios différents de TMPP. Pour l'expliquer, une première hypothèse est que cet effet est lié au tampon de mon échantillon qui serait peu compatible. Une alternative est que ce n'est pas l'efficacité de marquage qui est faible mais plutôt le taux d'identification des peptides puisqu'il a été rapporté que le marquage TMPP modifie le type de fragmentation des peptides lors de l'étape de MS-MS (Bertaccini *et al*, 2013).

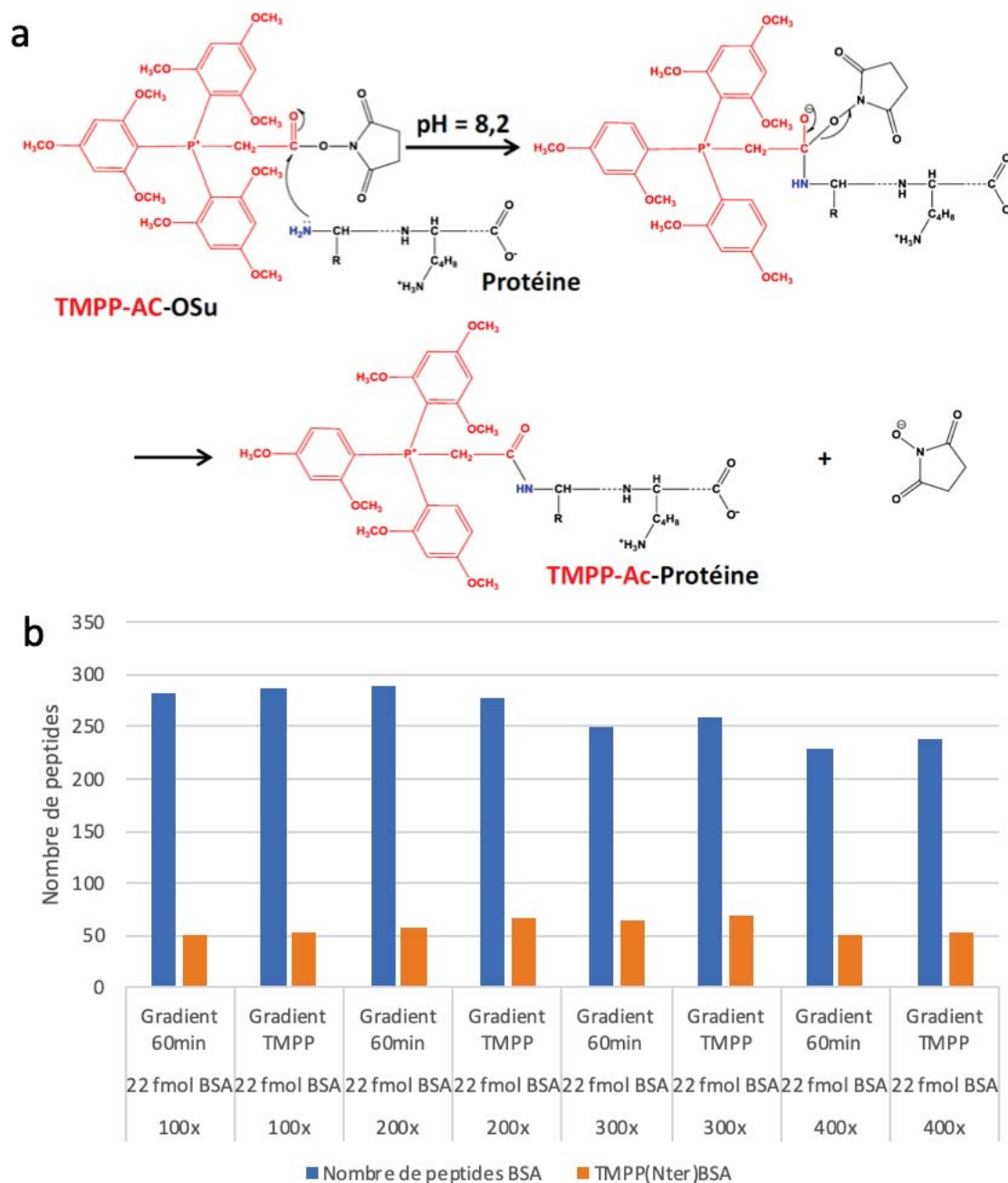


Figure II-12 : Identification des peptides tryptiques de BSA au TMPP. a) principe de marquage au TMPP ; b) Comparaison du marquage au TMPP de 22 fmol de peptides obtenus après digestion tryptique de la BSA à différents ratios TMPP : peptide et avec différents gradients d'élution dans l'étape de chromatographie liquide LC pré-analytique.

Une autre possibilité pour éviter les problèmes survenus lors de l'utilisation du TMPP est l'utilisation d'un autre type de marqueur, la NHS-Biotine (**Figure II-13**), qui a la particularité d'exister sous plusieurs formes. Certaines de ces formes sont hydrolysables ce qui permet de retirer la biotine et facilite l'analyse en spectrométrie

de masse (**Figure II-13b**). Le fait d'ajouter un groupement biotine au N-terminus des peptides d'intérêt permet de purifier sur colonne d'affinité de streptavidine et enrichir fortement et spécifiquement les échantillons en peptide N-terminal. Je n'ai malheureusement pas eu l'opportunité de mettre au point ce protocole mais cette approche me semble très prometteuse et celle que je privilégierais a priori pour valider l'approche que j'ai mise en place.

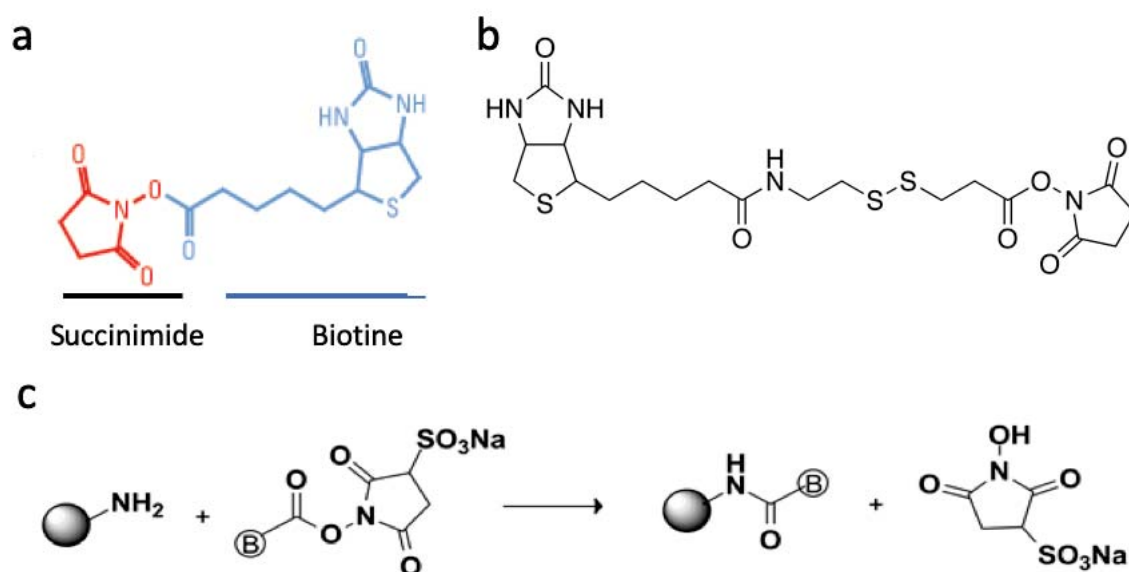


Figure II-13 : Structure et réactivité de la succinimide-biotine. a) structure du réactif avec détail de sa partie succinimide et biotine (b) version NHS biotine clivable, possédant un pont disulfure ; c) Réaction de marquage des N-terminaux des protéines par la NHS-Biotine.

II.4. Exploration de l'effet de la protéase IpaJ *in-cellulo*

Pour aller plus loin dans la caractérisation physiologique de l'action de la protéase IpaJ, j'ai souhaité évaluer sa capacité à retirer la myristoylation des protéines *in cellulo*. Dans ce contexte, j'ai souhaité exprimer la protéine directement dans des cellules humaines en culture pour en évaluer l'impact global. L'effet sur le myristoylome de la cellule a été évalué par spectrométrie de masse en utilisant l'option perte de glycine N-terminale comme modification supplémentaire.

Pour effectuer la transfection, j'ai utilisé un vecteur navette procaryote-eucaryote de type pTracer, ayant la caractéristique d'avoir deux promoteurs CMV pour produire deux protéines différentes. Le premier promoteur sert à produire une protéine de fusion GFP-BleoR. Cette fusion confère la résistance à la bléomycine grâce à la protéine BleoR, et permet d'évaluer rapidement l'efficacité de transfection des cellules grâce à la GFP. Le deuxième promoteur permet de produire la protéine d'intérêt qui dans mon cas correspond à IpaJ. Pour cette expérience j'ai transfecté les différents variants d'IpaJ (IpaJ30-259WT, IpaJ30-259 3M et IpaJ30-259 3M C64S). J'ai choisi d'utiliser le variant sauvage avec le module hydrophobe afin d'évaluer si le domaine hydrophobe muté joue un rôle dans la cellule, comme par exemple pour assurer interaction avec les petites protéines G de la famille ARF/ARL. Le variant C64S sert de contrôle négatif car il s'agit du mutant inactif. J'ai décidé de conserver la fusion GST-His dans ces constructions afin d'augmenter sa solubilité mais aussi pour pouvoir la visualiser par immuno-blot, n'ayant pas d'anticorps anti-IpaJ à ma disposition. In vitro la fusion possède la même activité que la protéine seule (libérée avec la protéase TEV) La transfection des gènes exprimant les variants d'IpaJ a été réalisée sur des cellules humaines immortalisées (HEK293).

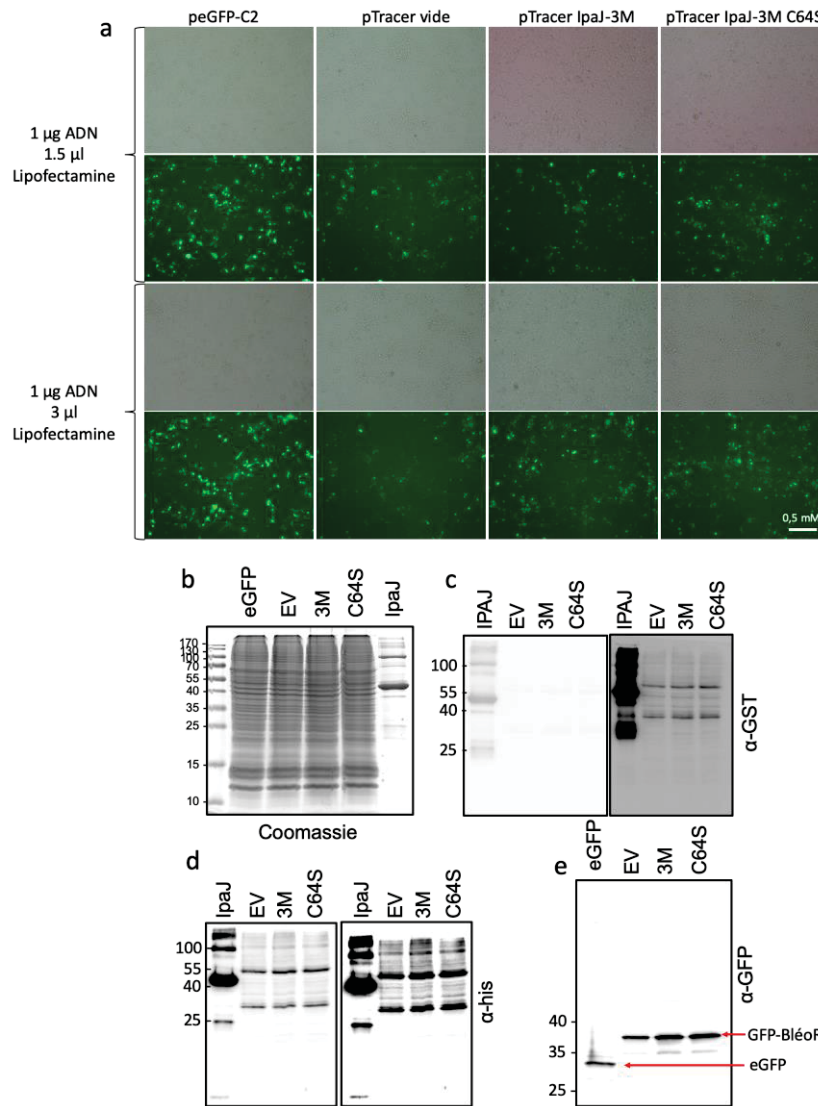


Figure II-14 : Transfection de constructions exprimant *IpaiJ* dans des cellules humaines HEK293 . Les migrations des marqueurs de masse de protéines sont indiquées en kDa sur la gauche de chaque image des gels a) Image par fluorescence des cellules 24h après transfection ; la fluorescence verte de la GFP permet d'observer l'efficacité de transfection, sur les plasmides peGFP-C2, pTracer vide, pTracerIpaiJ3M et pTraverIpaiJC64S b) gel SDS-PAGE des différentes constructions, coloration au bleu de Coomassie c) Immuno-blot du gel en b avec anticorps anti-GST (monoclonal) des cellules ayant été transfectées par les plasmides pTracer vide (EV), pTracer-IpaiJ3M (3M) et pTracer-IpaiJ3MC64S (C64S). Sur la première piste la fusion GST-IpaiJ purifiée est déposée. d) Immuno-blot avec anticorps anti étiquette poly-His (monoclonal) des cellules ayant été transfectées avec les plasmides suivant pTracer vide (EV), pTracer-IpaiJ3M (3M) et pTracer-IpaiJ3MC64S (C64S). Sur la première piste ; IpaiJ purifié. e) Immuno-blot avec anticorps anti-GFP des cellules ayant été transfecté par les plasmides suivant pTracer vide (EV), pTracer-IpaiJ3M (3M) et pTracer-IpaiJ3MC64S (C64S)

Au cours du premier essai, j'ai observé une forte efficacité de transfection (**Figure II-14a**), comparable à un autre plasmide de référence (peGFP-C2). Malheureusement, lors de l'étape qui consiste à vérifier l'expression d'IpaJ effective dans les cellules, je n'ai pas pu observer la GST par immuno-blot (**Figure II-14b-d**) qui indique la présence d'IpaJ, via l'utilisation d'anticorps anti-GST (monoclonal) ou via l'utilisation d'anticorps anti-His (monoclonal). Pour augmenter la sensibilité de détection, j'ai utilisé ensuite des anticorps polyclonaux anti-GST (**Figure II-15**). Cependant, je n'ai pas détecté la présence de la fusion GST-IpaJ dans les extraits cellulaires. En effet, les seules bandes observées sont présente sur toutes les pistes y compris celle du vecteur vide (EV).

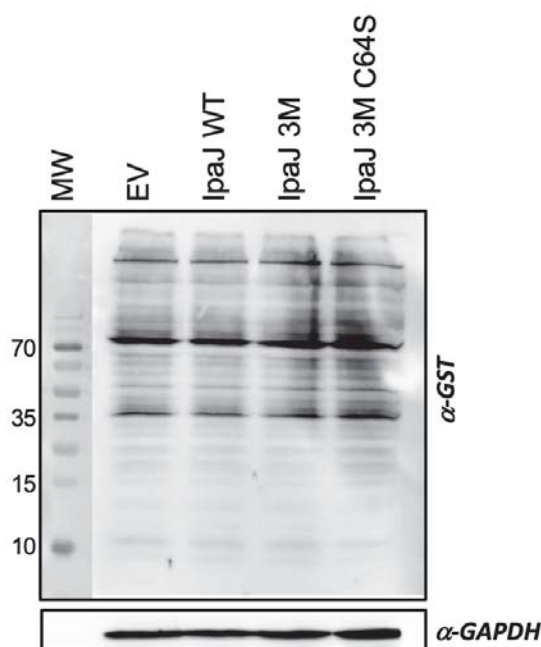


Figure II-15 : Immuno-blot avec anticorps anti-GST (polyclonal) des cellules transfectées par des fusions IpaJ. Les plasmides suivant sont été utilisés : *pTracer vide (EV)*, *pTracer-IpaJ^{WT} (WT)*, *pTracer-IpaJ^{3M} (3M)* et *pTracer-IpaJ^{3MC64S} (C64S)*. Sur la première piste ; marqueur de masse de protéines (en kDa).

Cette expérience sera refaite afin de comprendre pourquoi je ne parviens pas à détecter l'expression d'IpaJ dans les extraits cellulaires. J'ai déjà utilisé un autre vecteur d'expression mais les résultats sont tout autant négatifs. J'ai déjà formulé plusieurs hypothèses pour expliquer mon observation. D'abord, il peut s'agir d'une grande

instabilité de la fusion IpaJ, qui peut ne pas être soluble ou même être dégradée. Je testerai l'expression de la GST seule pour évaluer si je peux la détecter. Je ne peux exclure non plus que l'optimisation du code génétique d'IpaJ que j'ai adopté pour exprimer chez *E. coli* ne soit pas adaptée. Peut-être enfin dois-je exprimer les 30 premiers résidus d'IpaJ qui pourraient avoir une fonction dans l'hôte et ne seraient pas clivés. Enfin, la fusion avec d'IpaJ avec la GST n'est peut-être pas optimale et je devrais exprimer IpaJ complète, avec code génétique optimisée et une étiquette N-terminale de type FLAG pour la détecter par immunoblot avec les anticorps adaptés.

Numéro d'accension	Nom	Total	Mito S		Mito M		Micro		Cyto	
			1	2	1	2	1	2	1	2
P29728	2~-5~-oligoadenylate synthase 2	0	0	0	0	29	0	34	0	0
P62191	26S protease regulatory subunit 4	0	310	0	452	0	623	0	100	945
P62191-2	26S protease regulatory subunit 4 Isoform 2	0	252	0	441	0	607	0	0	0
O43741	5~-AMP-activated protein kinase subunit beta-2	37	0	0	0	0	0	0	0	46
P52209	6-phosphogluconate dehydrogenase, decarboxylating	0	1123	120	470	69	190	0	259	243
P05141	ADP/ATP translocase 2	233	0	0	0	87	51	284	1073	82
P84077	ADP-ribosylation factor 1	28	525	0	0	0	0	0	0	0
P61204	ADP-ribosylation factor 3	20	522	0	0	0	0	0	17	37
P61204-2	ADP-ribosylation factor 3 Isoform 2	0	455	0	0	0	0	0	0	0
P18085	ADP-ribosylation factor 4	0	116	0	0	0	18	0	0	0
P84085	ADP-ribosylation factor 5	0	112	0	62	0	0	0	0	0
P62330	ADP-ribosylation factor 6	0	36	0	0	0	0	0	0	0
P36405	ADP-ribosylation factor-like protein 3	83	49	0	0	0	0	0	0	0
Q02952	A-kinase anchor protein 12	0	0	0	52	0	0	0	17	0
Q02952-3	A-kinase anchor protein 12 Isoform 3	0	0	30	0	0	0	0	0	0
P51572	B-cell receptor-associated protein 31	332	0	0	0	0	383	351	315	189
P51572-2	B-cell receptor-associated protein 31 Isoform 2	0	0	0	44	0	0	0	0	0
P55957	BH3-interacting domain death agonist	32	0	0	0	0	0	0	0	0
P55957-2	BH3-interacting domain death agonist Isoform 2	0	155	0	0	0	0	0	0	0
P80723	Brain acid soluble protein 1	0	48	0	218	70	458	79	338	587
P80723-2	Brain acid soluble protein 1 Isoform 2	0	0	0	174	0	336	0	288	0
Q9UIF8	Bromodomain adjacent to zinc finger domain protein 2B	37	0	0	0	0	0	0	0	0
P17612	cAMP-dependent protein kinase catalytic subunit alpha	34	50	0	49	0	0	0	0	85
P17612-2	cAMP-dependent protein kinase catalytic subunit alpha Isoform 2	0	41	0	37	0	0	0	0	0
P22694	cAMP-dependent protein kinase catalytic subunit beta	26	0	0	0	0	0	0	0	53
P22694-10	cAMP-dependent protein kinase catalytic subunit beta Isoform 10	0	31	0	29	0	0	0	33	0
P22694-4	cAMP-dependent protein kinase catalytic subunit beta Isoform 4	0	0	0	0	0	0	0	60	0
Q13237-2	cGMP-dependent protein kinase 2 Isoform 2	0	0	0	0	0	0	18	0	0
Q8ND76	Cyclin-Y	0	0	0	0	69	0	0	0	20
Q8N7R7	Cyclin-Y-like protein 1	0	0	0	0	0	13	0	0	0
P53701	Cytochrome c-type heme lyase	0	0	0	0	0	0	36	56	0
P53634	Dipeptidyl peptidase 1	0	24	0	24	0	0	0	0	21
Q96EQ8	E3 ubiquitin-protein ligase RNF125	15	0	0	0	0	0	0	0	0
Q14254	Flotillin-2	0	0	0	131	70	43	49	195	65
Q95466-2	Formin-like protein 1 Isoform 2	0	0	17	0	0	0	0	0	0
Q96PY5-3	Formin-like protein 2 Isoform 2	0	0	0	0	0	0	34	22	0
P51116	Fragile X mental retardation syndrome-related protein 2	0	0	0	402	61	0	0	468	0
O00461	Golgi integral membrane protein 4	0	0	0	0	0	0	67	62	0
Q9BQ33	Golgi reassembly-stacking protein 1	0	30	0	34	0	54	0	24	31
Q9H8Y8	Golgi reassembly-stacking protein 2	0	40	0	77	0	73	0	109	0
Q9H8Y8-2	Golgi reassembly-stacking protein 2 Isoform 2	0	0	0	0	0	0	0	46	31
Q9H8Y8-3	Golgi reassembly-stacking protein 2 Isoform 3	0	0	0	56	0	0	0	84	0

Numéro d'accension	Nom	Total	Mito S				Mito M				Micro		Cyto	
			1	2	1	2	1	2	1	2	1	2	1	2
Q9H4G4	Golgi-associated plant pathogenesis-related protein 1	0	0	0	0	0	0	0	107	0	0	0	0	0
P04899	Guanine nucleotide-binding protein G(i) subunit alpha-2	45	0	0	141	49	0	166	0	154	0	0	86	0
P04899-2	Guanine nucleotide-binding protein G(i) subunit alpha-2 Isoform 2	0	0	0	59	0	0	0	0	34	0	0	0	0
P08754	Guanine nucleotide-binding protein G(k) subunit alpha	0	0	0	161	244	0	188	0	168	0	0	78	0
P09471	Guanine nucleotide-binding protein G(o) subunit alpha	0	0	0	100	49	0	139	0	121	0	0	0	0
P38405	Guanine nucleotide-binding protein G(olf) subunit alpha	0	0	0	0	49	0	0	0	69	0	0	0	0
P38405-3	Guanine nucleotide-binding protein G(olf) subunit alpha Isoform 3	0	0	0	33	0	0	0	0	0	0	0	0	0
P11488	Guanine nucleotide-binding protein G(t) subunit alpha-1	0	0	0	0	49	0	0	0	0	0	0	0	0
P19087	Guanine nucleotide-binding protein G(t) subunit alpha-2	0	0	0	0	49	0	0	0	0	0	0	0	0
A8MTJ3	Guanine nucleotide-binding protein G(t) subunit alpha-3	0	0	0	105	49	0	155	184	71	138	0	0	0
P19086	Guanine nucleotide-binding protein G(z) subunit alpha	0	0	0	34	0	0	0	0	36	0	0	0	0
P19367	Hexokinase-1	0	0	0	0	73	0	111	0	0	0	0	0	0
P19367-3	Hexokinase-1 Isoform 3	0	0	0	0	0	0	0	0	631	0	0	0	0
Q13308	Inactive tyrosine-protein kinase 7	0	0	0	0	0	0	20	0	0	0	0	0	0
Q13308-6	Inactive tyrosine-protein kinase 7 Isoform 6	0	0	0	0	0	0	0	0	29	0	0	0	0
Q5JSI4-4	Integrator complex subunit 6-like Isoform 3	15	0	0	0	0	0	18	0	0	0	0	0	0
Q9UQU13	Leucine-rich repeat protein SHOC-2	0	0	0	0	0	0	0	0	57	0	0	0	0
Q9UQU13-2	Leucine-rich repeat protein SHOC-2 Isoform 2	0	0	0	0	0	0	0	0	0	0	0	72	0
A6NIK2	Leucine-rich repeat-containing protein 10B	0	0	0	0	24	0	0	0	0	0	0	0	0
P00338	L-lactate dehydrogenase A chain	1033	1844	223	490	763	514	54	170	324	0	0	0	0
P00338-2	L-lactate dehydrogenase A chain Isoform 2	0	1270	0	339	0	314	0	123	0	0	0	0	0
P00338-3	L-lactate dehydrogenase A chain Isoform 3	0	1436	0	323	0	0	0	0	0	0	0	0	0
P00338-4	L-lactate dehydrogenase A chain Isoform 4	0	1543	0	407	0	233	0	96	0	0	0	0	0
P49006	MARCKS-related protein	82	204	0	212	102	339	155	298	219	0	0	0	0
Q9NXG3	MICOS complex subunit MIC19	117	0	0	0	0	32	171	230	0	0	0	0	0
Q9BRQ6	MICOS complex subunit MIC25	0	0	0	0	0	0	43	0	0	0	0	0	0
Q5JR59-4	Microtubule-associated tumor suppressor candidate 2 Isoform 3	0	0	0	0	0	0	18	0	0	0	0	0	0
Q969Z3	Mitochondrial amidoxime reducing component 2	0	0	0	0	0	0	0	20	0	0	0	0	0
Q5VT66	Mitochondrial amidoxime-reducing component 1	25	0	0	0	0	0	0	65	0	0	0	0	0
Q5VT66-3	Mitochondrial amidoxime-reducing component 1 Isoform 3	0	0	0	0	0	0	0	35	0	0	0	0	0
O96008	Mitochondrial import receptor subunit TOM40 homolog	76	0	0	0	34	0	32	347	0	0	0	0	0
O96008-2	Mitochondrial import receptor subunit TOM40 homolog Isoform 2	0	0	0	0	0	0	0	234	0	0	0	0	0
P29966	Myristoylated alanine-rich C-kinase substrate	61	174	0	380	107	112	0	391	445	0	0	0	0
P17568	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 7	0	0	0	0	0	0	124	31	0	0	0	0	0
P00387	NADH-cytochrome b5 reductase 3	334	0	0	0	583	0	68	0	0	0	0	0	0
P00387-3	NADH-cytochrome b5 reductase 3 Isoform 3	0	0	0	91	0	49	0	42	0	0	0	0	0
Q16798	NADP-dependent malic enzyme, mitochondrial	0	0	0	0	0	0	19	0	0	0	0	0	0
Q86UV6	N-alpha-acetyltransferase 40	0	30	0	0	0	0	0	0	0	0	0	0	0
Q7Z494	Nephrocystin-3	32	0	33	0	48	0	33	0	0	0	0	0	0
Q6ZMN7	PDZ domain-containing RING finger protein 4	0	0	0	0	0	0	30	0	0	0	0	0	0
Q99570	Phosphoinositide 3-kinase regulatory subunit 4	0	0	18	0	0	0	0	0	0	0	0	0	0
Q8N3A8	Poly [ADP-ribose] polymerase 8	0	0	0	0	0	0	0	22	0	0	0	0	0

Numéro d'accension	Nom	Total	Mito S			Mito M			Micro		Cyto	
			1	2	1	2	1	2	1	2	1	2
P02545	Prelamin-A/C	21	0	0	0	0	0	0	0	0	0	0
Q9P2B2	Prostaglandin F2 receptor negative regulator	0	0	0	0	63	0	65	0	81	54	
Q9NR22	Protein arginine N-methyltransferase 8	56	56	0	85	0	0	0	0	0	86	
Q9NUQ9	Protein FAM49B	124	201	0	33	53	0	0	0	32	0	
Q9NUQ9-2	Protein FAM49B Isoform 2	0	168	0	0	0	0	0	0	0	0	
Q96KN4	Protein FAM84A	0	0	0	0	13	0	0	0	0	0	
Q9C0E8	Protein lunapark	0	0	0	0	0	0	20	0	0	0	
Q9BZQ8	Protein Niban	0	0	18	0	0	0	0	27	0	0	
P35813-3	Protein phosphatase 1A Isoform 3	0	27	0	0	0	0	0	0	0	0	
P49593	Protein phosphatase 1F	0	85	63	0	0	0	0	0	0	0	
O15355	Protein phosphatase 1G	0	170	0	104	44	0	0	0	45	103	
O75695	Protein XRP2	0	0	0	48	34	0	0	0	36	42	
P12931	Proto-oncogene tyrosine-protein kinase Src	0	0	0	50	0	72	75	0	51	41	
Q6IAA8	Regulator complex protein LAMTOR1	83	0	0	38	0	334	0	0	0	45	
Q96S59	Ran-binding protein 9	39	0	0	0	0	0	0	0	0	94	
Q13905	Rap guanine nucleotide exchange factor 1	0	0	0	0	16	0	0	0	0	0	
Q9BWE0	Replication initiator 1	0	81	0	0	0	0	0	0	0	0	
Q9NRX5	Serine incorporator 1	0	0	0	0	0	50	23	0	0	0	
Q13177	Serine/threonine-protein kinase PAK 2	0	115	0	0	0	31	0	0	0	137	
Q9Y512	Sorting and assembly machinery component 50 homolog	0	0	0	0	0	0	0	0	93	0	
Q96L92	Sorting nexin-27	0	39	0	0	0	22	0	0	28	51	
Q96L92	Sorting nexin-27	0	0	0	0	0	0	0	0	0	51	
Q6ZUB1	Spermatogenesis-associated protein 31E1	0	0	17	17	0	0	27	0	0	0	
Q96SB4	SRSF protein kinase 1	121	0	0	291	26	0	0	199	300		
Q96SB4-3	SRSF protein kinase 1 Isoform 1	0	0	0	272	0	0	0	176	0		
O60637	Tetraspanin-3	0	0	0	50	33	81	44	36	106		
O60637-2	Tetraspanin-3 Isoform 2	0	0	0	34	0	60	0	26	0		
P68366	Tubulin alpha-4A chain	300	2699	681	1116	217	290	105	1125	1838		
P06241	Tyrosine-protein kinase Fyn	0	0	0	102	0	0	66	75	0		
P08631	Tyrosine-protein kinase HCK	0	0	0	73	0	0	0	0	0		
P08631-2	Tyrosine-protein kinase HCK Isoform 2	0	0	0	0	0	0	36	0	0		
P06239	Tyrosine-protein kinase Lck	0	0	0	73	0	95	53	67	65		
P07948	Tyrosine-protein kinase Lyn	0	0	0	94	0	67	0	83	0		
P07947	Tyrosine-protein kinase Yes	0	0	0	0	0	0	66	0	0		
Q9HB07	UPF0160 protein MYG1, mitochondrial	97	159	0	122	84	0	0	0	0	77	
Q43149	Zinc finger ZZ-type and EF-hand domain-containing protein 1	0	0	0	42	0	0	41	37	0		
Sup table I-1 Identification des protéines ayant une isoforme décrite comme myristoylée par LC-MS des échantillons après fractionnement. Classé par les scores d'identification obtenue sur Mascot.												

Chapitre III

III. Des instantanés haute résolution de la N-myristoyltransférase humaine en action éclairent un mécanisme favorisant la myristoylation N-terminale des Glycines et des Lysines

Dans le premier article auquel j'ai contribué (Dian *et al*, 2020), l'équipe, en collaboration avec celle du Pr Edward Tate (Imperial College, Londres, Royaume Uni), a réalisé une étude structure fonction de la HsNMT1 pour en décrypter le mécanisme. Dans ce contexte, l'équipe a réussi à cristalliser la NMT non seulement en présence de plusieurs de ses substrats naturels mais aussi de toute une série d'intermédiaires réactionnels et de produits de réaction. Jusqu'à présent, les structures disponibles de petites molécules en complexe avec la NMT avaient été obtenues avec des analogues de réaction, comme par exemple un MyrCoA non hydrolysable par la présence d'un groupe méthylène supplémentaire à proximité immédiate du groupe thioester réactif, ou de petites molécules inhibitrices rendues non peptidiques pénétrant dans la poche du substrat peptidique et non myristoylables. Ceci a induit à des biais importants dans la compréhension du mécanisme catalytique fin de l'enzyme (Bhatnagar *et al*, 2001). La nouveauté de l'article consiste d'abord en ce que les structures résolues ont été toutes obtenues en présence des deux substrats (polypeptide et MyrCoA). Les cristallographes de chacune des deux équipes (C. Dian et I. Pérez-Dorado, copremiers auteurs de l'article) ont mis au point des conditions expérimentales permettant de révéler des intermédiaires de réaction - en particulier l'intermédiaire tétramérique - ainsi que des produits de la réaction au sein du site actif, avant et après la libération du CoA, représentant les différentes étapes de la réaction de myristoylation (**Figure 2 de l'article**). L'utilisation de trois peptides avec des séquences très différentes, révèle que ces peptides adoptent une conformation similaire dans le site actif. Le site actif est équipé de de plusieurs poches permettant d'accueillir chaque acide aminé du substrat polypeptidique et plaçant le peptide de façon idéale pour réaliser la réaction, c'est à

dire avec la Gly2 positionnée à distance idéale (3,4Å) du carbonyle réactif du MyrCoA placé dans le trou de l'oxyanion (**Figure 4 de l'article**). Grâce aux différentes structures à haute résolutions que nous avons obtenues, nous avons ainsi pu observer les différentes contraintes que la NMT applique aux deux substrats (MyrCoA et peptide) permettant de la placer dans les conditions optimales de catalyse. Ces contraintes permettent de rapprocher les deux substrats non seulement entre eux mais aussi du canal solvant conservé chez toutes les NMTs (**Figure supplémentaire 3 de l'article**). Le canal solvant permet d'apporter au plus profond du site actif les molécules d'eau nécessaires à la déprotonation du substrat pour l'attaque nucléophile du thioester du MyrCoA. Notre étude a aussi répondu aux incertitudes concernant l'utilité de la boucle Ab de la NMT, en effet dans la littérature certains auteurs ont conclu que sa conformation fermée pouvait être un artéfact de la troncature réalisée au niveau du N-terminal de la NMT, afin de la rendre plus stable pour la cristallisation. Il se trouve que cette boucle joue un rôle très important dans la catalyse de la réaction de myristoylation puisque, dans sa forme fermée, elle induit un changement de conformation du MyrCoA indispensable à la catalyse de la réaction (**Figures 3 et 5 de l'article**). De plus cette boucle Ab en position fermée interagit avec l'hélice N-terminal B'A' lorsque le site actif est occupé par un peptide substrat, et lorsque la réaction est catalysée, cette hélice change de conformation, induisant une relaxation du CoA, afin de le libérer avant le peptide (**Figure 6 de l'article**). L'article permet ainsi de proposer un mécanisme réactionnel nouveau pour la NMT (**Figure 7 de l'article**)

Grâce à ces nouvelles données, nous avons réalisé une analyse fine de l'architecture du site actif recevant le peptide substrat. Ce travail a permis de modéliser que la NMT devrait être aussi capable de réaliser aussi la myristoylation sur le groupement amine de la chaîne latérale (appelé ϵ -NH₂ car l'amine est positionnée sur le carbone 5, le carbone alpha étant le carbone 1) d'une lysine si elle se trouve en position trois (**Figure 8a-c de l'article**). Pour ce faire, une rotation est nécessaire autour du C α de aa3 afin de provoquer l'entrée de l'acide aminé 2 (par exemple la glycine) dans la cavité occupée normalement par la chaîne latérale du résidu 3 (qui devient le résidu réactif dans ce

modèle). La présence d'une douzaine de substrats commençant par Gly dans le protéome humain et disposant d'une lysine en position trois a suggéré que cette réaction devrait théoriquement effectivement être réalisable *in cellulo*. Ma contribution expérimentale à l'article 1 publié dans Nature Communications en 2020 (Dian *et al*, 2020) a donc concerné la démonstration que l'hypothèse selon laquelle la NMT peut réaliser la myristoylation sur une Lysine en position ϵ est bien fondée. Cela m'a conduit à vérifier la faisabilité de cette réaction *in vitro* avec la NMT et à mettre au point en particulier les conditions d'un test par spectrométrie de masse permettant de l'évaluer. J'ai ainsi réalisé la myristoylation de ces substrats et analysé par MALDI-TOF les produits de réaction. Les spectres MS1 permettent d'identifier si la modification a lieu (décalage de masse de 210 Da) et d'évaluer l'efficacité avec laquelle la réaction est réalisée (**Figure supplémentaire 13 de l'article**). Les spectres MS/MS (MS2) des espèces modifiées servent à caractériser le type de modification réalisée (**Figure 9 de l'article**). Les spectres MS2 ont montré sans ambiguïté une myristoylation sur la glycine et non sur la lysine de tous les peptides commençant par GK (**Figure 9a de l'article**). Afin d'empêcher cette réaction qui masque probablement l'autre, j'ai donc choisi de bloquer l'amine alpha N-terminale par acétylation sur des peptides présentant le même contexte, c'est à dire une glycine en position deux et une lysine en position trois ou directement une Lysine en position 1. Dans ce cas la myristoylation de la lysine est observée sans ambiguïté, ce qui indique que cette réaction est bien possible (**Figure 9b-c de l'article**). Ceci suggère qu'elle n'est pas favorisée si l'amine N-terminal de la glycine est disponible. J'ai enfin analysé toute une série de spectres MS/MS de peptides myristoylés observés à partir d'extraits cellulaires que j'ai préparés et n'ai pas pu observer de peptides présentant une myristoylation en Lysine (**Figure 9d de l'article**). Avec l'expertise de Cyril Dian, nous avons ensuite obtenu et résolu la structure haute résolution de la HsNMT1 cristallisée en présence du peptides glycine² lysine³ présentant une acétylation bloquant l'amine N-terminale, et nous avons pu observer à résolution quasi-atomique qu'effectivement la lysine myristoylée sur son amine ϵ . Encore une fois, nous avons pu capturer différentes étapes de la réaction ϵ -

myristoylation, avant et après l'attaque nucléophile, dans les différents monomères d'un monocristal, ce qui nous a permis de visualiser que le mécanisme d' ϵ -myristoylation est légèrement différent du mécanisme de myristoylation des glycines. En effet ce nouveau mécanisme n'est pas médié par une molécule d'eau. L'extrémité amine réactive de la lysine réalise en effet une liaison directe avec la base catalytique de la NMT, le groupement carboxy-terminal du dernier résidu, la Gln496. Ces données sont reportées dans la **Figure 8d-e de l'article**.

ARTICLE



<https://doi.org/10.1038/s41467-020-14847-3>

OPEN

High-resolution snapshots of human *N*-myristoyltransferase in action illuminate a mechanism promoting N-terminal Lys and Gly myristoylation

Cyril Dian^{1,9}, Inmaculada Pérez-Dorado^{2,3,6,9}, Frédéric Rivière¹, Thomas Asensio¹, Pierre Legrand⁴, Markus Ritzefeld^{2,7}, Mengjie Shen^{2,8}, Ernesto Cota³, Thierry Meinzel¹, Edward W. Tate^{2,5} & Carmela Giglione¹

The promising drug target *N*-myristoyltransferase (NMT) catalyses an essential protein modification thought to occur exclusively at N-terminal glycines (Gly). Here, we present high-resolution human NMT1 structures co-crystallised with reactive cognate lipid and peptide substrates, revealing high-resolution snapshots of the entire catalytic mechanism from the initial to final reaction states. Structural comparisons, together with biochemical analysis, provide unforeseen details about how NMT1 reaches a catalytically competent conformation in which the reactive groups are brought into close proximity to enable catalysis. We demonstrate that this mechanism further supports efficient and unprecedented myristoylation of an N-terminal lysine side chain, providing evidence that NMT acts both as N-terminal-lysine and glycine myristoyltransferase.

¹Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette 91198, France. ²Department of Chemistry, Imperial College, Molecular Sciences Research Hub, Wood Lane, London W12 0BZ, UK. ³Department of Life Sciences, Imperial College London, Exhibition Road, South Kensington SW7 2AZ, UK. ⁴Synchrotron SOLEIL, Gif-sur-Yvette, Cedex 91192, France. ⁵The Francis Crick Institute, 1 Midland Rd, London NW 1AT, UK. ⁶Present address: Evotec Ltd, 114 Innovation Dr, Milton Park, Milton, Abingdon OX14 4RZ, UK. ⁷Present address: Evotec SE, Essener Bogen 7, Hamburg 22419, Germany. ⁸Present address: Oakland Innovation, Mill Rd, Harston, Cambridge CB22 7GG, UK. ⁹These authors contributed equally: Cyril Dian, Inmaculada Pérez-Dorado. ✉email: thierry.meinzel@i2bc.paris-saclay.fr; e.tate@imperial.ac.uk; carmela.giglione@i2bc.paris-saclay.fr

Glycylpeptide *N*-tetradecanoyltransferase 1, commonly known as *N*-myristoyltransferases (NMTs, EC 2.3.1.97), are universally conserved eukaryotic enzymes responsible for a major post-translational modification (PTM), myristoylation (MYR)¹. NMTs are thought to transfer a C:14:0 acyl-CoA (MyrCoA) exclusively to the N-terminal glycine (Gly) of specific substrates. NMT-directed *N*-myristoylation is usually co-translational following initiator Met removal from nascent polypeptides by methionine aminopeptidases at the ribosome (Fig. 1a)². MYR is fundamental to diverse cellular processes including recruitment and interaction of myristoylated proteins with membranes³ and other proteins and can contribute to protein stability⁴.

NMT was first characterised in yeast⁵ and later in fungi, protozoa, insects, plants and mammals including humans^{1,6}. Humans express two NMT isoenzymes (HsNMT1 and HsNMT2) in most tissues, accommodating hundreds of protein targets^{7,8}, including some implicated in diseases such as cancer and infection (for reviews see refs. 9–11). NMTs are thought to be essential in all organisms and are validated drug targets, particularly for the treatment of infectious diseases^{12–15}. However, progress in the field of human NMT biology has been undermined by non-selective and now invalidated compounds¹⁶, raising questions about previous studies using these drugs. Further therapeutic development will be greatly enhanced by a complete understanding of NMT catalysis and substrate selectivity.

NMTs are members of the large GCN5-related *N*-acetyltransferase (GNAT) superfamily^{17,18}, which are involved in countless cellular functions and in a wide range of diseases and

resistance mechanisms. GNATs catalyse donor acyl-coenzyme A (CoA) transfer to a variety of acceptor substrates¹. Despite limited sequence homology, GNATs share a common 3D core architecture (Fig. 1b). Atomic-level mechanistic studies of GNATs have been hampered by a lack of structural data on GNAT proteins complexed with their natural ligands. Nonetheless, a common mechanism has been proposed, though definitive proof is still missing. For most GNATs, this Bi:Bi catalytic mechanism involves a general base which deprotonates the protein substrate amino group to generate a tetrahedral intermediate stabilised by a general acid. A proton wire of structured water molecules often connects the catalytic base to surrounding solvent^{17,19}, critically contributing to donor amine deprotonation and proton transfer from a base catalyst.

NMTs are atypical GNAT members featuring two adjacent GNAT domains with reduced sequence homology²⁰ (Fig. 1c, d). Previous studies have suggested that the MYR reaction does not resemble that of the other GNAT superfamily members²¹. The current accepted NMT mechanism, deduced from numerous NMT structures from different organisms in complex with MyrCoA or MyrCoA analogues and inhibitors^{6,22}, posits first the formation of a stable complex with MyrCoA in the N-terminal domain, which contributes to peptide acceptor site formation together with C-terminal GNAT domain residues²³ (Supplementary Fig. 1a–c). From observations arising from the structures of *C. albicans* (CaNMT) and *S. cerevisiae* NMT (ScNMT), MyrCoA binding-induced conformational switch of the conserved flexible ‘Ab-loop’ of NMT (Supplementary Fig. 1a–c), similar to

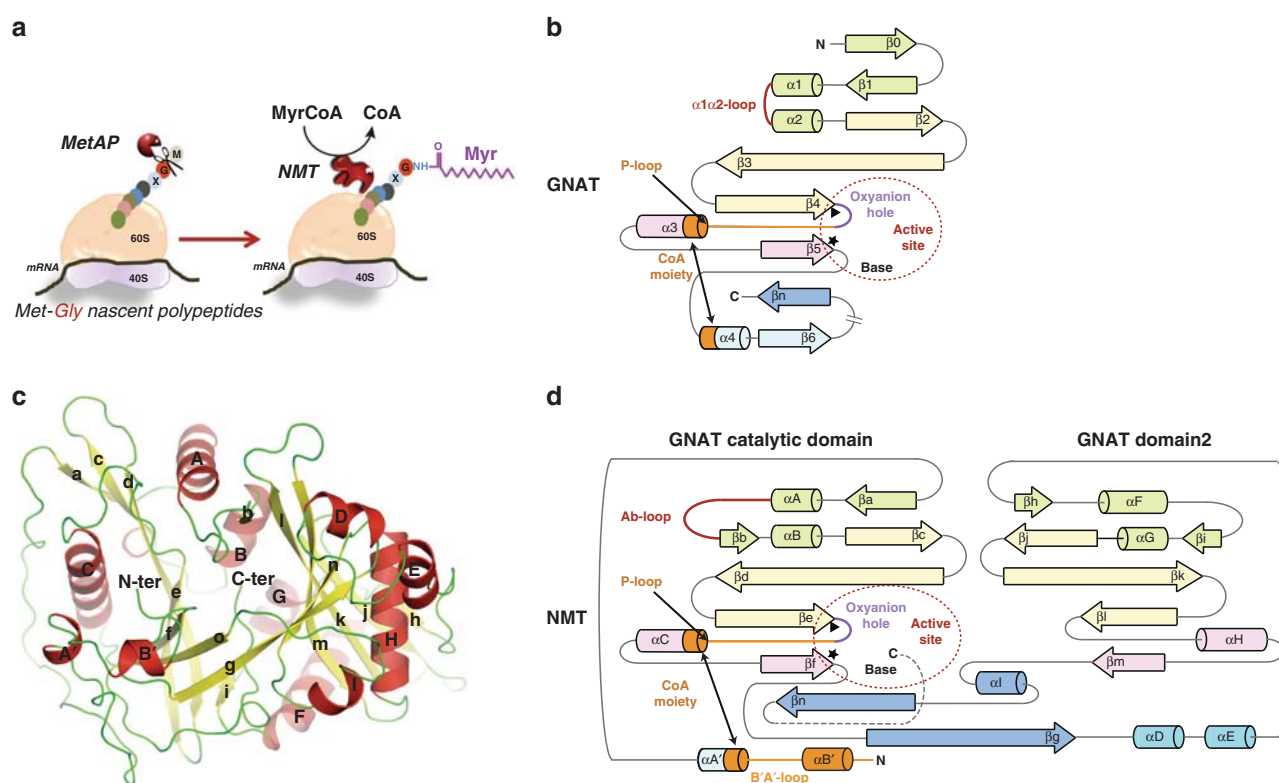


Fig. 1 *N*-myristoylation reaction and organisation of GNAT and NMT. **a** Cartoon representation showing the co-translational myristoylation process of the protein. The Met-Gly nascent polypeptides first undergo removal of initial Met by the action of methionine aminopeptidase (MetAP), and then the unmasked Gly is modified by NMT, which utilises MyrCoA to transfer the Myr moiety to the N-terminal Gly. **b** General topology of GNATs. Conserved secondary structural elements are coloured in pale green, yellow, pink and cyan, respectively. The non-conserved secondary structural elements are coloured in blue. Regions that are important for CoA binding are shown in orange. The oxyanion hole bulge is highlighted in violet. A red broken circle highlights the position of the active site. Ab-loop and equivalent $\alpha 1\alpha 2$ -loop are coloured in red. Black stars and triangles highlight crucial positions in the active site. **c** View of HsNMT:MyrCoA fold (pdb = 5O9T) labelled with secondary structure elements according to sequence alignment of NMTs in Supplementary Fig. 2a. **d** NMT topology showing the GNAT-core fold using the same colour code as in **b**.

the α_1 – α_2 loop of serotonin *N*-acetyltransferase¹⁹ (Fig. 1b), is thought to participate in the formation of a competent peptide-binding site and contribute to overall enzyme efficiency, but a role in catalysis has yet to be defined^{24,25}. However, involvement of the Ab-loop was controversial, as studies of full-length ScNMT suggested that the ordered-closed conformation of the Ab-loop, confining the peptide substrate within its binding site, was an artefact due to NMT missing the N-terminal region including the B'A' loop²⁶.

The structure of *Saccharomyces cerevisiae* NMT (ScNMT; Supplementary Fig. 1d) in complex with a bound natural substrate peptide and a non-hydrolysable MyrCoA analogue (S-(2-oxo)-pentadecyl-CoA, NHM)²³ shows a salt bridge between the distal carboxy base and the primary amine of the Gly (Gly2) acceptor peptide, suggesting, differently from the common mechanism of other GNATs, that deprotonation of this amine is directly mediated by the base. However, in this non-functional ternary complex, the Gly2 amine is 6.3 Å away from the MyrCoA reactive thioester, meaning that the deprotonated Gly2 amine is predicted to rotate and move toward MyrCoA to allow nucleophilic attack of the thioester group²³. This sole motion was suggested to be responsible of NMTs selectivity for Gly over Ala.

Here we report a series of high-resolution HsNMT1 structures co-crystallised with MyrCoA and reactive substrate peptides. These structures reveal snapshots across the entire catalytic mechanism including the reactive intermediate state, thereby providing the high-resolution map of MYR reaction chemistry. These structures highlight a number of important rearrangements of and around the amino-terminal acceptor group and provide an atomic-level deconstruction of each step of NMT reaction catalysis. Specifically, the reactive amino-terminal acceptor group undergoes highly dynamic interactions around a catalyst platform centred on Thr282, revealing a crucial role for a previously unseen water channel. We also show that the NMT long Ab-loop is a mobile structure that plays a crucial role in promoting pre-organisation of MyrCoA in the transition state, in turn triggering the concerted deprotonation of the substrate and nucleophilic attack of MyrCoA. Close examination of these structures provides valuable information about how to improve existing NMT inhibitors, particularly with respect to selectivity. Finally, our studies additionally support the transfer of a myristoyl moiety at the N-terminal epsilon amino group of Lys, decisively expanding the known NMT substrate range beyond the well-established N-terminal alpha moiety of Gly peptides.

Results

Overall structures of HsNMT1 binary and ternary complexes. Previous structural studies have taken advantage of substrate or inhibitor analogues to investigate NMT catalysis. To capture different snapshots of NMT catalysis at high-resolution, we co-crystallised substrates with human NMT1 (HsNMT1) or soaked preformed crystals of HsNMT1/MyrCoA complexes with active peptide substrates featuring an N-terminal Gly residue (Gly2). We therefore solved the crystallographic structure of HsNMT1 in complex with two different peptides (X and Y; see Methods for peptide details) at 2.3 and 2.1 Å resolution, respectively. In addition, we recently reported the crystallographic complex at 2.7 Å resolution of HsNMT1 with a third peptide (Z)⁸, which is further analysed here together in light of the X and Y complexes (see below). Each structure consisted of two HsNMT1 molecules in an asymmetric unit with the same overall fold (rmsd 0.157–0.236 Å). Further, this fold was conserved between HsNMT1 molecules of each solved complex (Supplementary Fig. 2) and between HsNMT1 and HsNMT2 structures previously reported in complex with MyrCoA or non-hydrolysable MyrCoA

NHM². The compact GNAT core was clearly identified (Fig. 1c, d and Supplementary Fig. 2b) and consisted of an internal pseudo two-fold symmetry axis between the N-terminal half building the MyrCoA binding site, and the C-terminal region, shaping the peptide substrate-binding site. The latter comprised seven major pockets accommodating peptide substrate side chains of residues aa3 to aa6⁸.

Our crystallographic structures revealed binary and ternary complexes in substrate, intermediate and product states trapped in crystallo along the *N*-myristoylation (MYR) reaction pathway: (1) peptide X and MyrCoA in a ternary complex with HsNMT1 prior to the reaction (HsNMT1:MyrCoA:X chain B, Fig. 2a, e); (2) the tetrahedral intermediate formed by the reaction of peptide Y N-terminal amine with MyrCoA thioester (HsNMT1:TI-Y chain A, Fig. 2b, f); (3) the myristoylated peptide and CoA products bound prior to release (HsNMT1:CoA:MyrX chain A and HsNMT1:CoA:MyrY chains A and B, Fig. 2c, g); and (4) the binary complex between HsNMT1 and the myristoylated peptide Z product (MyrZ) following ordered CoA release (HsNMT1:MyrZ, Fig. 2d, h⁸). The product complexes with peptides X and Y (HsNMT1:CoA:MyrX chain A and HsNMT1:CoA:MyrY chain B) fully overlapped, indicating that the state was independent of peptide composition (Fig. 3a). Of note, in chain A of the NMT complex with peptide Y, both tetrahedral intermediate and products of the reaction (CoA and MyrY) were observed, exhibiting occupancies of 0.3 and 0.7, respectively.

Snapshots of HsNMT1-mediated MYR catalysis. All three peptide substrates, despite their different sequences, adopted an extended β -strand conformation between aa2–aa7 (Fig. 3a), with the aa3–aa6 side chains located in the four side chain recognition pockets⁸ featuring in strictly identical conformations before, during and after the MYR reaction (Fig. 3a, pockets 2–5).

HsNMT1:MyrCoA:X represents a snapshot of the peptide substrate prior to nucleophilic attack on MyrCoA (Fig. 2a, e). The free N-terminal Gly2 (aa2) of peptide X is hosted in a cavity (pocket 1, Fig. 3a) built by the side chains of Tyr180, Tyr192, Asn246, Thr282, Leu495, the C-terminal carboxylate of Gln496, and the myristoyl thioester of MyrCoA (Fig. 4a). The extended peptide conformation and Gly2 positioning orient the Gly2 peptide bond antiparallel to the MyrCoA thioester, with the amino-terminal group at ideal distance (3.4 Å) from the reactive MyrCoA thioester carbonyl (CM1, Figs. 3b and 4a). This carbonyl lies in the so-called oxyanion hole formed by the amide backbone of Phe247 and Leu248 (equivalent to Phe170 and Leu171 in ScNMT;²⁴ Fig. 4a). At the active site, Gly2 is constrained by three sets of H-bonds. The first set anchors Gly2 at the active site through direct interaction of its amino-terminal group with Thr282 (Fig. 4a). In the second, Asn246, Tyr180 and Tyr192 interact with the aa3 amide nitrogen via a water molecule (wat1) (Fig. 4a), ensuring correct substrate orientation toward the catalytic base C-terminal carboxylate of Gln496. The third set of H-bonds between Gly2 and the Gln496 carboxylate are mediated by a second water molecule, wat2 (Figs. 3b and 4a), at the end of a 22-Å-long, 544 Å³ solvent channel connecting the active site to solvent (Fig. 4 and Supplementary Fig. 3). These three H-bond networks contain strictly conserved residues (Supplementary Fig. 2a), suggesting that the water channel plays an active role in catalysis. These key mechanistic insights were missed in the previous ScNMT structure solved in complex with a peptide and NHM²³, since the latter was non-reactive. Superimposition of the structures reveals steric clashes in the ScNMT structure between Na of Gly2 and the additional carbon atom of NHM that impede effective Gly2 positioning in pocket 1 (Fig. 3b), preventing both the peptide substrate and Ab-loop from adopting

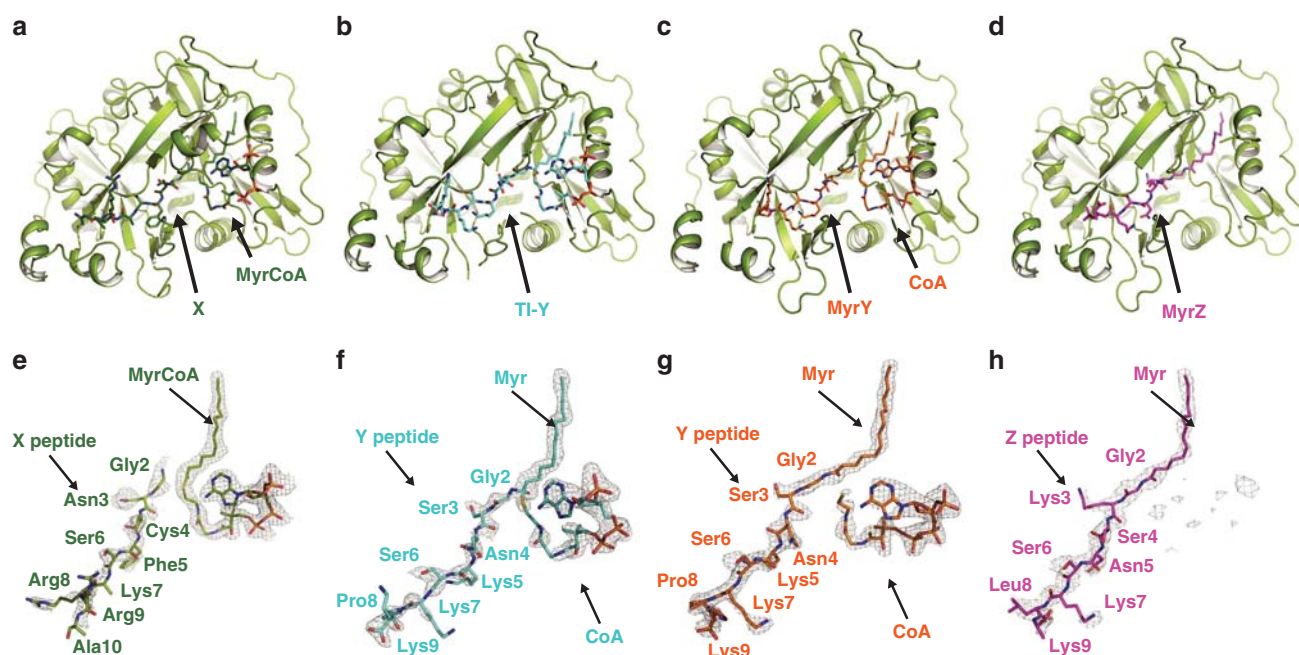


Fig. 2 Overview of the obtained crystallographic NMT complexes. Crystal structures of HsNMT1 in complex with substrates, TI and products obtained with different peptides X, Y and Z. The HsNMT1 main chain is displayed as a green ribbon. Substrates, TI and products are displayed as sticks and are indicated with black arrows. **a** Crystal structure of HsNMT1 in complex with MyrCoA and peptide X (green). **b** Crystal structure of HsNMT1 in complex with TI-Y, in which peptide Y and CoA are coloured in cyan. **c** Crystal structure of HsNMT1 in complex with MyrY and CoA (orange) products. **d** Crystal structure of HsNMT1 in complex with MyrZ product coloured in pink. **e–h** Detail of the “fofc omit” sigmaA-weighted electron density maps ($mF_{obs} - DF_{calc}$, PHI_{calc}), shown in grey, for substrates, TI and products (in sticks coloured as in **a–d**). The omit electron density maps were calculated after a round of refinement in which the occupancy of substrates, TI and products was set to zero. The $e/\text{\AA}^3$ value of each “2fofc omit” sigmaA-weighted electron density map ($2mF_{obs} - DF_{calc}$, PHI_{calc}) at 0.92–1 rmsd level (0.25, 0.33, 0.33 and 0.30 $e/\text{\AA}^3$ for X, TI-Y, Myr-Y and Myr-Z, respectively) was used as reference to set the contoured value of fofc omit electron density maps shown. Peptide, Myr, MyrCoA and CoA moieties are indicated with black arrows.

catalytically competent extended β -strand and closed conformations, respectively (Fig. 3b, c). These findings strongly suggest the ScNMT aa2-aa5 peptide conformation and Gly2-Gln496 carboxylate salt bridge are artefacts caused by the extra methylene of NHM²³.

To challenge the primary importance of the water molecule channel in mediating Gly2 deprotonation, we created two truncated NMT1 variants in which the last 2 or 3 amino acids were removed. The crystal structure of the HsNMT1 deletion variant with truncation of the last two amino acids, HsNMT1 Δ C2, shows the peptide substrate orientated towards Glu244, which becomes the new catalytic base, prior to nucleophilic attack at the carbonyl group of the MyrCoA (Supplementary Fig. 4a, left). The 3D structure also reveals both an intermediate state in which the Gly2 is oriented towards the MyrCoA and a final state with the formed Myr-peptide (Supplementary Fig. 4a, middle and right). Interestingly, the structure of the complex prior to nucleophilic attack shows that Gly2 interacts with the new catalytic base Glu244 via two water molecules of the water channel (Supplementary Fig. 4a, left), confirming their importance for NMT catalysis. Accordingly, using an in vitro MYR assay (see Methods), we established that the truncation does not alter the protein turnover rate (k_{cat}) (Table 1) and observed a significant reduction in the enzyme's relative affinity to the peptide. This suggests that the C-terminus of the enzyme is also directly involved in correct peptide adjustment into the catalytic centre. This hypothesis was confirmed by the biochemical and structural characterisation of the HsNMT1 variant with further truncation of an additional amino acid (HsNMT1 Δ C3). HsNMT1 Δ C3 displays barely detectable enzyme activity (Table 1), and its 3D structure reveals

different conformations of the N-terminal region of the peptide substrate. This indicates that the low measured activity of HsNMT1 Δ C3 is due to the high Gly2 mobility, most likely induced by the large space created by the C-terminal truncation (Supplementary Fig. 4b). The water network shaping the water channel of the NMT:MyrCoA:X complex is unmodified in the various crystal structures described later. Such ordered water molecules ideally bridge the reactive amino terminus next to a catalytic base, usually the carboxy terminal but also alternative carboxylic side chains such as in HsNMT1 Δ C2. This indicates that tight ordering of the water molecules ensured by the water channel is crucial to the catalytic activity of NMT.

In the HsNMT1:TI-Y crystal structure, the tetrahedral intermediate results from nucleophilic attack by the Gly2 amine on the MyrCoA thioester (Fig. 2b, f). Comparison of HsNMT1:MyrCoA:X and HsNMT1:TI-Y reveals that the Gly2 amino-terminal group undergoes a 70° rotation around the Ca-C bond on forming the tetrahedral intermediate (Fig. 4b vs. 4a and Supplementary Fig. 5a), leaving a space filled by a new water molecule (wat3). Wat3 likely arises from the water channel as it extends the network. Wat3 stabilises the TI complex through a new H-bond network with the Ile245 carbonyl and Thr282 amide groups (Fig. 4b vs. 4a). Thr282 appears to play a pivotal role as a catalyst platform by inducing both concerted rotation and nucleophilic attack by the deprotonated Gly2 amine. Importantly, it also offsets the distortion induced by displacement of the newly formed tetrahedral thioester carbon towards the peptide Gly2 amine (Fig. 4b vs. 4a and Supplementary Fig. 5a). The observed tetrahedral intermediate geometry is in agreement with the stereochemistry of the reaction path (Fig. 4b vs. 4a, and Supplementary Fig. 5a; see also Bürgi et al.²⁷).

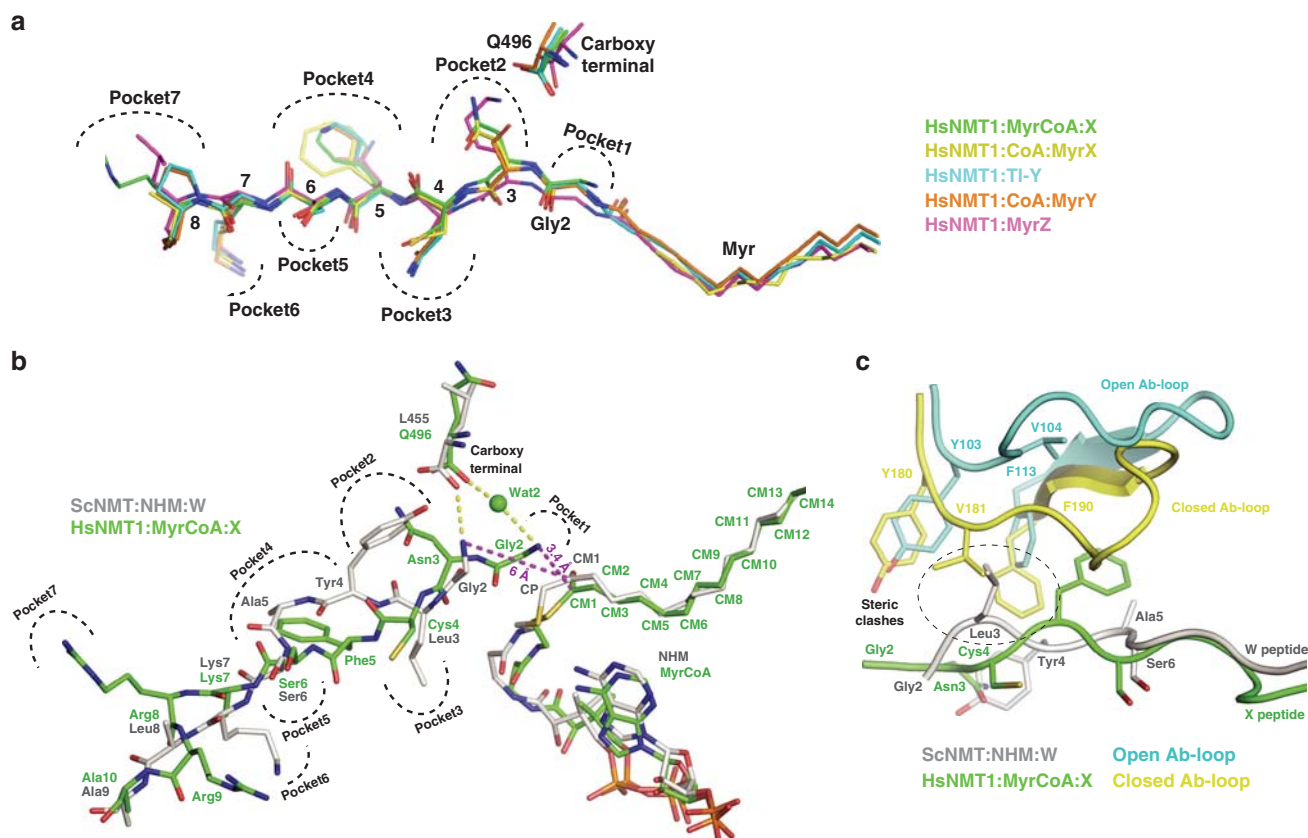


Fig. 3 The β -strand conformation of the substrate peptide is appropriate for the Ab-loop closed conformation. **a** Superimposition of the three peptides X, Y and Z in the different states reveals a conserved β -strand conformation. Gly2 from X peptide is bound into a new cavity, increasing to seven the number of “recognition pockets” of the peptide-binding groove. Pockets allocating each amino acid are indicated with black dashed lines. Peptides X (green), MyrX (yellow), TI-Y (cyan), MyrY (orange) and MyrZ (pink) are displayed as sticks. **b** Superimposed peptides from HsNMT1:MyrCoA:X (green) and ScNMT:NHM:W (grey, PDB:1IID). The 2-Å distance between Gly2 ammonium of peptide X and the extra-methylene (CP) of NHM shows how the analogue prevents Gly2 of peptide W to bind at pocket 1 and forces the C α of Gly2 of W to shift back to occupy the n-1 aa position observed in peptide X. This motion influences the whole peptide W, including aa5, which co-locates with aa4 of peptide X in pocket 4 and illustrates the major consequences induced by the extra methylene group of NHM on peptide conformation, pocket occupancy and amino-terminal reactive group positioning. Magenta dashed lines indicate distances between each Gly2 ammonium and CM1 reactive carbonyl of MyrCoA. **c** Superimposition of HsNMT1:MyrCoA:X (green) and ScNMT:NHM:W (grey, PDB:1IID). The NHM-induced conformation of peptide W favours the Ab-loop open conformation (cyan) to the detriment of the closed conformation (yellow) observed in HsNMT1:MyrCoA:X due to steric clashes between Leu3 and Tyr4 from W peptide and V181 and F190 from HsNMT1:MyrCoA:X Ab-loop, respectively (black broken line circle).

The HsNMT1:TI-Y complex reveals key changes at the active site important for understanding both reaction progression (Fig. 4b vs. 4a) and substrate specificity. Tight packing of the Gly2 C α against the Tyr180 hydroxyl, Asn246 carbonyl and MyrCoA thioester carbonyl groups in the TI structure (Fig. 4b) prevents accommodation of an Ala side chain (Fig. 4b and Supplementary Fig. 6), thereby contributing to specificity for Gly (Supplementary Fig. 6). Moreover, Tyr180 belongs to the conserved and mobile Ab-loop²⁶, which bears residues important for aa7 recognition, and provides a link between upstream substrate recognition and Gly2 selectivity.

HsNMT1-product structures reveal how CoA compacts following tetrahedral intermediate collapse, an event previously thought to be required to enable CoA release from NMT²³. The β -strand conformation of the peptide induces a shift in the myristoyl carbonyl group toward the oxyanion hole (see asterisks in Supplementary Fig. 5b vs. 5a) and away from the sulphur atom of the CoA product (Supplementary Fig. 5b). The “question mark” shape of MyrCoA generates tension in the transition state that is released in the products, which causes the thiolate anion to move to 3.3 Å from the Myr-peptide amide bond (Supplementary Fig. 5b)²⁴ and adopt an optimal position to form an

intramolecular H-bond with the pendant amine of the CoA adenosyl ring (Fig. 4c and Supplementary Fig. 5). No acidic residues are sufficiently close to the thiolate group to permit its protonation; Tyr180 is the closest proton source at 4.0 Å from the thiolate leaving group. This suggests that TI reformation might occur via a 36° rotation around the C2P/C3P bond of CoA (Supplementary Figs. 5b and 4b), an observation that supports CoA release as a mechanism to prevent TI reformation as a key rate-limiting step of the catalytic cycle.

The Ab-loop is a key structural element in the NMT catalysis. The Ab-loop has been proposed to participate in the formation of an accessible peptide-binding site through a disordered-to-open transition induced by MyrCoA binding and is thought to contribute to overall NMT catalytic efficiency^{23–25}. However, parallel studies on a ScNMT:MyrCoA:inhibitor complex²⁶ concluded that the closed Ab-loop conformation may be an artefact of N-terminal truncation in the construct used for crystallography (Supplementary Fig. 2a).

To address these uncertainties, we examined our structures to assess the effects of peptide binding and product formation on the

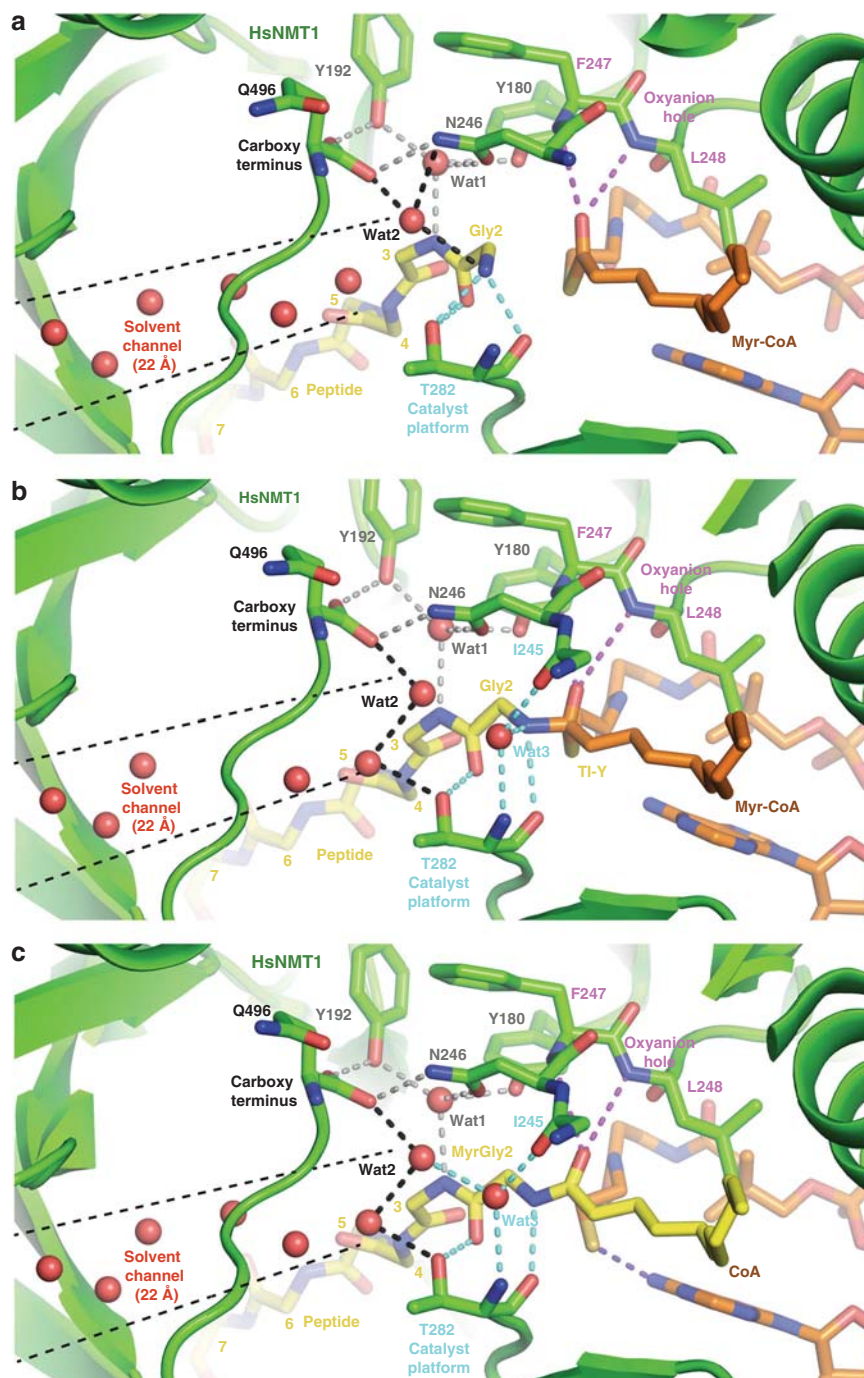


Fig. 4 Three reaction snapshots reveal the importance of the T282 catalyst platform and water solvent channel. View of the HsNMT1 active site showing direct and solvent-mediated hydrogen bonding interactions between the protein and: **a** substrates in HsNMT1:MyrCoA:X, **b** TI in HsNMT1:TI-Y and **c** products in HsNMT1:MyrX:CoA. The HsNMT1 chain is displayed as a green ribbon. Selected amino acids (green), substrate peptides (yellow), and MyrCoA and CoA moieties (both in orange) are shown as sticks. Hydrogen bond set1, linking Gly2 to the T282 catalyst platform, is shown as cyan dashes. Wat1-mediated hydrogen bond set2, linking Q496 to Tyr180/Tyr192 from the Ab-loop, substrate aa3 and Asn246 is shown as grey dashes. Wat1-mediated hydrogen bond set3 linking Gln496 to Gly2 ammonium is shown as black dashes. Water molecules from the solvent channel are shown as red spheres. Hydrogen bonds involved in the “oxyanion hole” are displayed as magenta dashed lines. The H-bond between thiolate and amine from CoA is shown as violet dashes.

Ab-loop conformation (Supplementary Fig. 7). This conformation was closed in all our structures, with the exception of the tetrahedral intermediate complex, in which it was partially disordered (Supplementary Fig. 7e). Our structures revealed six key structural Ab-loop features: (1) the salt bridge between substrate Lys7 and Asp471, which is the only specific interaction stabilising aa7 in the

presence of all peptides (Supplementary Fig. 7); (2) an additional interaction network stabilising the closed Ab-loop conformation in the HsNMT:CoA:MyrX and HsNMT:MyrCoA:X complexes (Fig. 5c vs. 5a, b); (3) the conserved Phe188 residue, which acts as a lid on selectivity pocket 4 in both Ab-loop open and closed conformations (Supplementary Fig. 8); (4) the steric interaction between Val181

Table 1 Catalytic parameters for HsNMT1 and related mutants using MyrCoA.

HsNMT1 variant	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{s}^{-1}\cdot\text{M}^{-1}$)	Relative k_{cat}/K_m
Wild-type	18 ± 4	0.41 ± 0.01	$23,502 \pm 5259$	100
ΔC2	629 ± 127	0.34 ± 0.04	538 ± 122	2.3
ΔC3	2140 ± 134	0.22 ± 0.01	93 ± 5	0.4
Lys107Glu/Lys252Glu	51 ± 15	1.33 ± 0.14	$25,848 \pm 7846$	110
Tyr180Phe/Asn246Ala	33 ± 11	0.68 ± 0.07	$20,924 \pm 7398$	89
Tyr180Ala	765 ± 204	0.16 ± 0.02	206 ± 63	0.9
Tyr180Pro	nm	nm	<2	<0.01
Val181Ala	20 ± 5	0.36 ± 0.03	$18,541 \pm 4950$	78
Val181Leu	54 ± 12	0.33 ± 0.03	6172 ± 1516	26
Tyr192Ala	187 ± 34	0.93 ± 0.10	4960 ± 1056	21

The GlyCysSerValSerLysLysLys (SOS3) peptide was used as reference peptide.
Nm indicates that no reactivity was observed at the lowest measurable rate value performed with the highest NMT concentration (10 μM).

and the peptide aa3 carbonyl, limiting Ab-loop closure (Fig. 5b, c, magenta dotted lines); (5) a positively charged peptide residue at aa7, favouring interaction of Asp471 and the negatively charged Ab-loop in the closed conformation (Fig. 5c and Supplementary Fig. 7); and (6) distortion of the MyrCoA thioester bond when the Ab-loop is closed (Fig. 5c, d).

The charged peptide substrate aa7 (Lys) plays an important role in peptide selectivity and catalysis and is confined within an outer-inner interface cavity (pocket 6) (Fig. 3a). The back and front walls of this cavity are formed by hydrophobic residues surrounding the aliphatic chains of aa7 and Asp471, respectively (Supplementary Fig. 7a–c), whilst the cavity roof is built from three aspartate residues (Asp471 and Asp183/Asp185 of the Ab-loop), salt bridging the aa7 amine (Supplementary Fig. 7a, b). Although this cavity has evolved to accommodate a charged aa7, its side chain is disordered in HsNMT1:MyrCoA:X in contrast to HsNMT1:TI-Y, where the Ab-loop is disordered and the substrate aa7 side chain interacts with Asp471 (Supplementary Fig. 7d, e). These observations suggest that Asp471 initiates aa7 binding in the active cleft and that aa7 does not trigger the closed conformation of the Ab-loop but instead contributes to its transient stabilisation after the peptide is bound.

Comparing ScNMT:MyrCoA (open) and HsNMT1:MyrCoA:X (closed) (Fig. 5b, black circle), the closed switch of the Ab-loop forces MyrCoA to change conformation, shifting its C6 atom by 1.77 Å and distorting the thioester 20° out of sp^2 planarity (Fig. 5d). This observation is supported by reanalysis of reported structures of ternary complexes with inhibitors at the peptide binding site, which show that distortion of the MyrCoA thioester bond only occurs when the Ab-loop is closed (Supplementary Fig. 9). Together, these findings support a role for the Ab-loop in TI formation by pre-organising MyrCoA toward the TI state in a concerted manner with the oxyanion hole (Fig. 4a, pink dotted lines, right hand side), driving thioester bond polarisation when a peptide correctly fills the substrate-binding groove.

To verify the importance of the closed Ab-loop conformation in NMT catalysis, we generated a number of variants targeting key residues of this loop and we assayed their activity. Variants Val181Leu or Tyr192Ala showed reduced NMT catalysis (Table 1). Most interestingly, the variants Tyr180Ala and Tyr180Pro induced reduced and complete loss of MYR activity, respectively (Table 1). Indeed, Tyr180 is likely to be implicated in both MyrCoA conformational change (Fig. 5d) and Gly2 scanning (Supplementary Fig. 6) suggesting a strong link between activity and selectivity in NMTs. As a Tyr180Phe/Asn246Ala substitution did not alter catalytic efficiency with respect to the wild-type, we concluded that the hydrophobic character of residue 180 and its backbone were crucial for the Ab-loop to adopt the required structural constraint on MyrCoA. Tyr180 is positioned at the very

kink of the loop and locks one side of the Ab loop (Fig. 3c). Taken together, these biochemical and structural data highlight the importance of the transient Ab-loop closed conformation in NMT catalysis.

Impact of the B'A' N-terminal α -helix on product release. The HsNMT:MyrCoA:X complex was obtained with a long NMT construct (see Methods) bearing the N-terminal B'A'-loop. Interestingly, the B'A'-loop is fully ordered and partially folds as an α -helix in this complex (Figs. 5b, c and 6a, and Supplementary Fig. 10). Such folding of the B'A'-loop further stabilises both MyrCoA/CoA via several interactions with the CoA moiety (Fig. 5b, c, bottom left and 6a). First, the P3 phosphate of MyrCoA/CoA makes three hydrogen bonds with the backbone nitrogen atoms of Gln118 and Tyr117 and the side chain of Arg115 (Fig. 5e). Second, the B'A'-loop makes interactions with the fg-loop (Fig. 5e, pink), which favour the proper planar orientation of the hydrophobic interactions of the adenine CoA ring provided by both Leu287 and Val285 side chains. The adenosine ring as a result stacks in between this pocket and the panthotenic- β -alanine side of CoA. The side chain of Val286 (part of fg-loop) is settled in a hydrophobic cavity made by the side chains of Ala196, Met109, Ala112 within the helical part of the B'A'-loop region. Stabilisation of the fg-loop, is achieved by direct hydrogen bonding between the Ala106 carbonyl ($\alpha\text{B}'$ helix) and the Val286 backbone nitrogen. Third, Pro105 acts as a top lid to the ribose of the adenosine moiety. Finally, a Lys107-Asp184 salt bridge strengthens the B'A'-loop fold linking B'A'-loop with the closed conformation of Ab-loop (Fig. 5b, c). As a result, reinforced interactions of the B'A'-loop with the MyrCoA are expected to mainly participate in closing of the loop (Fig. 5e). A variant including a Lys107Ala substitution showed a significantly increased k_{cat} value (Table 1), in keeping with CoA release as the limiting step of the reaction. A B'A'-loop can be identified in almost all NMTs (Supplementary Fig. 2a), and despite a low level of length and homology it is likely that similar interactions occur in other NMTs, as exemplified by the replacement of Lys107 and Glu184 by two hydrophobic residues in CaNMT.

The previously reported HsNMT2:MyrCoA complex (PDB code 4C2X) was solved with a construct partially lacking the B'A'-loop (Supplementary Figs. 2 and 10a). In this structure, the residual B'A'-loop is disordered despite the ordered conformation of the Ab-loop, suggesting that the complete B'A'-loop sequence is important for the proper folding of this region. The B'A'-loop conformation reverts to a disordered state in the product complex (HsNMT1:MyrZ) in which CoA has been released (Fig. 6b), driving opening of the distal part of the active site and unmasking the myristoyl group normally buried under

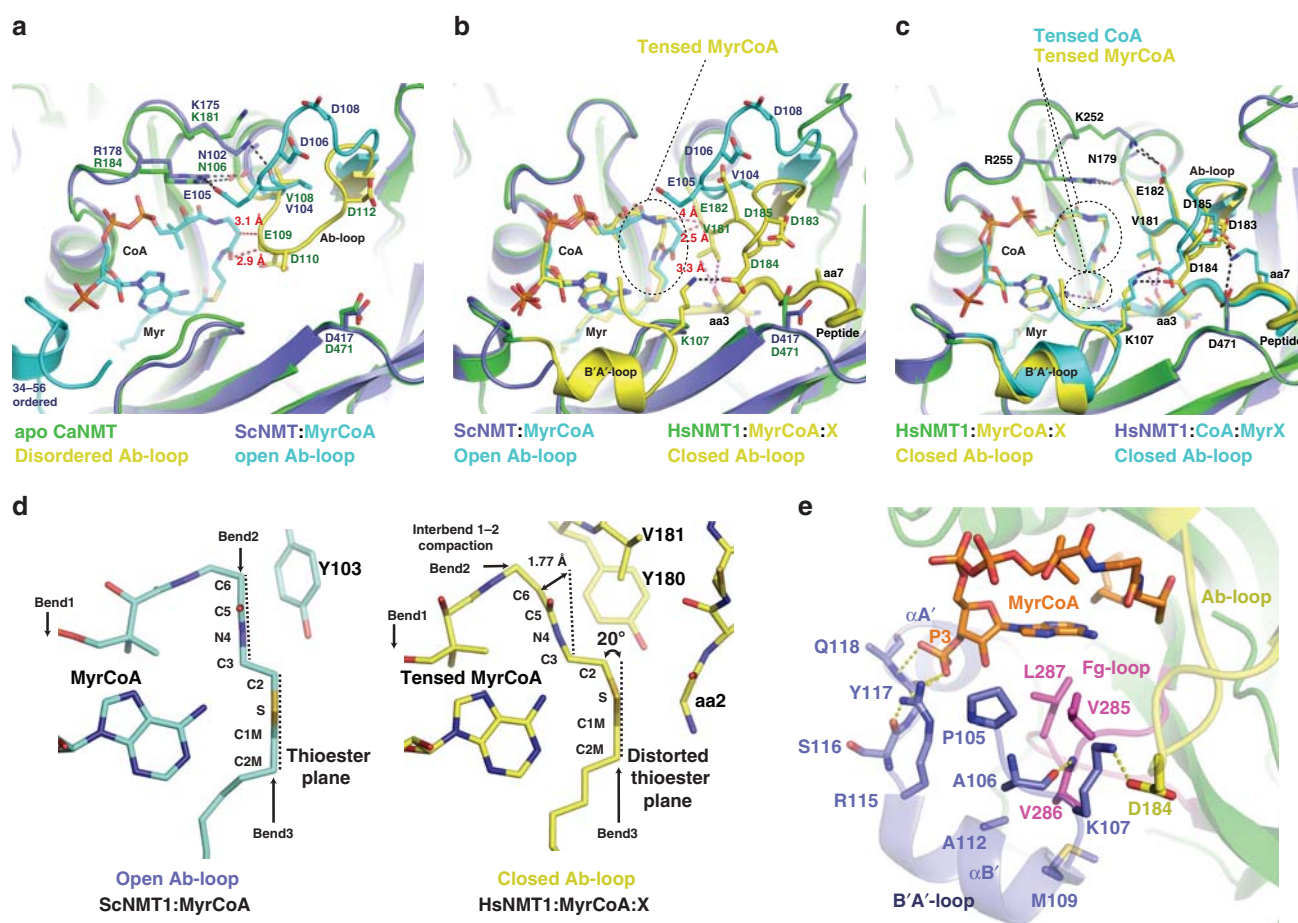


Fig. 5 MyrCoA and Ab-loop relationship and its importance in Myr catalysis. **a** Structural superimposition of Apo-CaNMT (PDB 1NMT, chain B, green) and the binary ScNMT:MyrCoA structure (PDB 1IIC, chain X, blue) showing rearrangements in the N-terminus (34–56, in cyan) and A-loop upon MyrCoA binding. The latter switches the Ab-loop conformation from disordered (Apo-CaNMT, in yellow) to open (in ScNMT:MyrCoA, in cyan) due to steric clashes (red dashed lines) with Glu109. Interactions with Arg178 and Lys181, stabilising the open Ab-loop conformation, are shown as black dashed lines. **b** Superimposition of ScNMT:MyrCoA (PDB 1IIC, blue) and HsNMT1:MyrCoA:X (green) complexes showing MyrCoA compaction induced by the Ab-loop closed conformation observed in the complex with peptide X (yellow), which is induced by steric clashes (red dashed lines) with Ab-loop Val181 (yellow). Detail of the folded B'A'-loop (yellow) induced by the closed conformation of the Ab-loop (yellow) is also shown, as well as its interactions with MyrCoA and Ab-loop (black dashed line). The Ab-loop closed conformation is limited by clashes with aa3 carbonyl from peptide X. **c** Superimposition of HsNMT1:MyrCoA:X (green) and HsNMT1:CoA:MyrX complexes reveals that the Ab-loop-induced compaction of MyrCoA (yellow) is also transmitted into the CoA product (cyan). Interactions stabilising the Ab-loop closed conformation in HsNMT1:CoA:MyrX by salt bridges are highlighted with black dashed lines. **d** Side-view zoom showing MyrCoA thioester planes and relevant Ab-loop residues as sticks, in both ScNMT:MyrCoA (cyan) and HsNMT1:MyrCoA:X (yellow), reveals an interbend 1–2 compaction of MyrCoA caused by the closed Ab-loop conformation in HsNMT1:MyrCoA:X. **e** View of the interactions of B'A'-loop region with MyrCoA moiety in HsNMT1:MyrCoA:X. HsNMT1 is displayed in cartoon, main core, B'A' region, Ab and fg-loops are colour in green, blue, yellow and pink respectively. Selected side chains and MyrCoA are displays as sticks. Hydrogen bonds and salt-bridges are shown as yellow dashes.

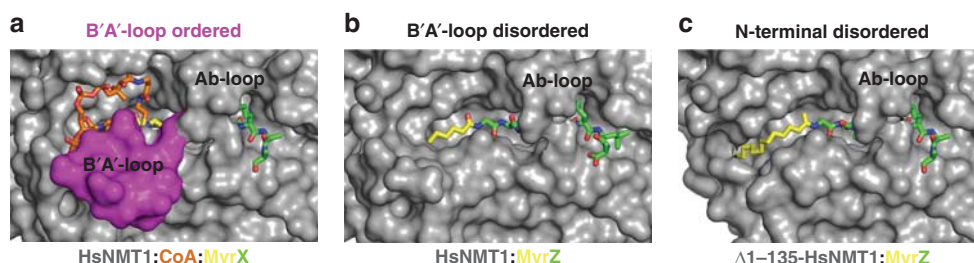


Fig. 6 Role of the B'A'-loop in sequential release of the two products. Solvent accessibility of the products at the HsNMT1 surface (dark grey) with the B'A'-loop surface highlighted in pink. Peptide, Myr and CoA moieties are shown as sticks in green, yellow and orange, respectively. **a** HsNMT1:MyrX:CoA crystal structure shows that an ordered B'A'-loop stabilises the CoA moiety and embeds Myr-peptide from the solvent. **b** HsNMT1:MyrZ crystal structure shows that CoA release, induced by unfolding the B'A'-loop, reveals a partially unmasked Myr moiety. **c** Model of HsNMT1:MyrZ deleted from the first 135 residues, mimicking the reported Apo-NMT crystal structure (pdb 1NMT), shows that the Myr moiety would be fully accessible to the solvent favouring Myr-peptide release or MyrCoA competition.

the CoA, facilitating Myr-peptide product release (Fig. 6b). Myr-peptide is a competitive inhibitor of MyrCoA binding²⁸, while CoA acts as a poor non-competitive inhibitor, stressing the importance of the B'A'-loop disordered conformation in Myr-binding site opening and product release^{1,28}. These observations highlight a dual role for the B'A'-loop in stabilising the CoA and a closed Ab-loop (Fig. 6a) and in unmasking the myristoyl moiety to facilitate Myr-peptide product release.

We next considered how product formation influences B'A'-loop unfolding. Compaction of CoA in HsNMT1:CoA:MyrX (Fig. 5c, d) by the Ab-loop is further amplified by an intramolecular H-bond between the amine and thiol groups of NMT-bound CoA (Fig. 4c, violet dotted lines, in right hand-side). This contrasts with the elongated CoA conformation observed in acetyl-CoA (AcCoA) or CoA-bound structures from members of the GNAT family (Supplementary Fig. 11), in which the 3'-phosphoadenosine moiety is exposed to solvent with no specific interactions identified at the active site, thereby allowing diverse conformations (Supplementary Fig. 11). These observations are consistent with the hypothesis that relaxation of the tensed CoA in NMT could occur from the adenine ring, inducing N-terminal B'A'-loop disorder and CoA release from the binding site.

Taken together, these analyses establish: (i) a causal relationship between the ordering of the N-terminal B'A'-loop with the closed conformation of the Ab-loop when the HsNMT1 active site is loaded with a substrate peptide, and (ii) the disordering of the N-terminal B'A'-loop induced by CoA relaxation, which in turn releases first CoA and next Myr-peptide products.

Reconstructing catalysis dynamics from crystal snapshots.

Interpolation of HsNMT1 ternary complexes suggests a putative trajectory for the NMT mechanism at high-resolution (Fig. 7) and identifies the conformational changes required for catalysis and apo-NMT recycling (Supplementary Movie 1). First, a nascent polypeptide chain with a positively charged Gly aa2 combined with an aa3-aa7 sequence compatible with pockets 2–6 enters and binds in the peptide-binding groove of HsNMT1:MyrCoA in an extended β -strand conformation (Fig. 7a). In this conformation, Gly2 is at an ideal distance and orientation from the carbonyl group of MyrCoA thioester, which is polarised in the oxyanion hole. A match between substrate and the substrate-binding groove switches the Ab-loop to a closed conformation (Figs. 4a, 5b and 7b), which translates into distortion of the MyrCoA thioester bond, leading to concerted water-mediated deprotonation of substrate Gly2 and nucleophilic attack on the thioester carbonyl guided by the Thr282 catalyst platform. The Ab-loop switch thus ensures tight and specific substrate selectivity and a β -strand conformation for the peptide (Fig. 7c). Reinforcement of the water-mediated interaction network around the peptide combined with thioester plane distortion enable NMT to lower the activation energy, while increasing its affinity for the tetrahedral transition state (Figs. 4b and 7d).

This detailed NMT mechanism is fully supported by existing site-directed mutagenesis quench-flow pre-steady state studies, which indicate that tetrahedral intermediate collapse is rate-limiting and that both ammonium deprotonation and nucleophile attack are ~ 140 -fold faster than the steady-state rate²¹, in agreement with Figs. 4b and 7c and the potential reversibility of Figs. 4b and 7d prior to release of CoA. Unlike previous structures, our data are also fully consistent with retention of activity in C-terminal truncations, since the water channel can bridge the missing residues^{26,29} (Supplementary Fig. 4). Similarly, the decreased activity reported for ScNMT[Thr205Ala] and ScNMT[Asn169Leu] variants agrees with the importance of

analogous Thr282 and Asn248 in HsNMT1 (Supplementary Fig. 12) for the catalytic platform and Gly2 positioning, respectively. Furthermore, the importance of the N-terminal B'A'-loop in stabilising the ordered-closed Ab-loop is supported by a 10-fold reduction in peptide K_m when this loop is deleted in ScNMT.

In the final steps of the mechanism, the new tensed CoA products and Myr-peptide are still bound to the active site (Fig. 7e), and subsequent relaxation of tensed CoA favours their sequential release (Fig. 7f). Spontaneous deprotonation of the Gln496 C-terminus may occur during this step; in the HsNMT1:MyrZ crystal structure, two rotamers of the Lys3 side chain suggest different protonation states for the C-terminal Gln496 carboxyl in each HsNMT1 molecule of the asymmetric unit. Finally, a closed-to-open Ab-loop switch and CoA-induced N-terminal NMT1 unfolding may induce Myr-peptide product release or alternatively allow competition by MyrCoA substrate to displace the Myr-peptide product.

NMT catalyses MYR on the ϵ -NH₂ of Lys mimicking the α -NH₂ of Gly.

NMTs are thought to exclusively acylate N-terminal Gly residues. Nevertheless, the ϵ -amino groups of internal Lys of several proteins including tumour necrosis factor- α ³⁰, interleukin- $\alpha 1$ precursor³¹ and SHMT2³² have been shown to undergo MYR. Our structural data strongly suggest that both Gly- and Lys-MYR could be substrates for NMTs, a previously unprecedented class of substrates for this enzyme class. We recently reported the crystal structure of a complex of HsNMT1, MyrCoA, and a peptide substrate devoid of its amino-terminus i.e., the N-terminal Gly2 residue substituted with an *N*-acetyl group attached to an N-terminal Asn⁸. Compared to the complexes made with genuine peptide substrates, this 3D complex showed rotation around the Ca of aa3 causing the *N*-acetylated amide to enter the cavity normally occupied by the side-chain of residue 3⁸. As a result of this motion, the Asn side-chain also rotated around its own Ca backbone toward the position usually occupied by the Gly2 of genuine NMT substrates⁸. Models based on this crystal revealed that the Lys (the only natural amino acid with a side-chain featuring such a reactive amino group) at position 3 of the substrate peptide may position its ϵ -amino group for potential MYR (Fig. 8a, b), while being compatible with the constraints imposed by side chains of Tyr180 and Asn246 in charge of Gly2 selectivity.

A dozen NMT substrates featuring a Lys3 have been identified in humans including Arf6 and several calcium-binding proteins of the hippocalcin (HPCA) family⁸. We therefore investigated whether peptides featuring Lys3 could be NMT substrates through either or both Gly and Lys moieties. To challenge this hypothesis, we first performed MALDI-ToF-ToF analysis of the HsNMT1-catalysed products of two Gly-Lys starting peptides derived from Arf6 and HPCA. The MS/MS spectra unambiguously revealed only MYR-Gly and no Lys-MYR (Fig. 9a and Supplementary Fig. 13a, b), fully in keeping with crystallographic data of the complex between HsNMT1 and the HPCA peptide (see Supplementary Fig. 3 in Castrec et al.⁸).

We next challenged whether blocking the alpha amino group of a Gly-Lys starting peptide with an *N*-acetylation, by analogy with the crystal structure featuring Ca3 rotation, could render the ϵ -amino group of Lys3 reactive for MYR. MS analysis of the reaction in the presence of HsNMT1 revealed that Lys-MYR occurred (Fig. 9b and Supplementary Fig. 13c). We further assessed whether shortening the acetyl-Gly-Lys to acetyl-Lys would also permit NMT-driven ϵ -MYR. MS analysis of the product of HsNMT1 catalysis using an acetyl-Lys starting peptide clearly demonstrated conversion to N-terminal acetyl-Lys-MYR (Fig. 9c and Supplementary Fig. 13d).

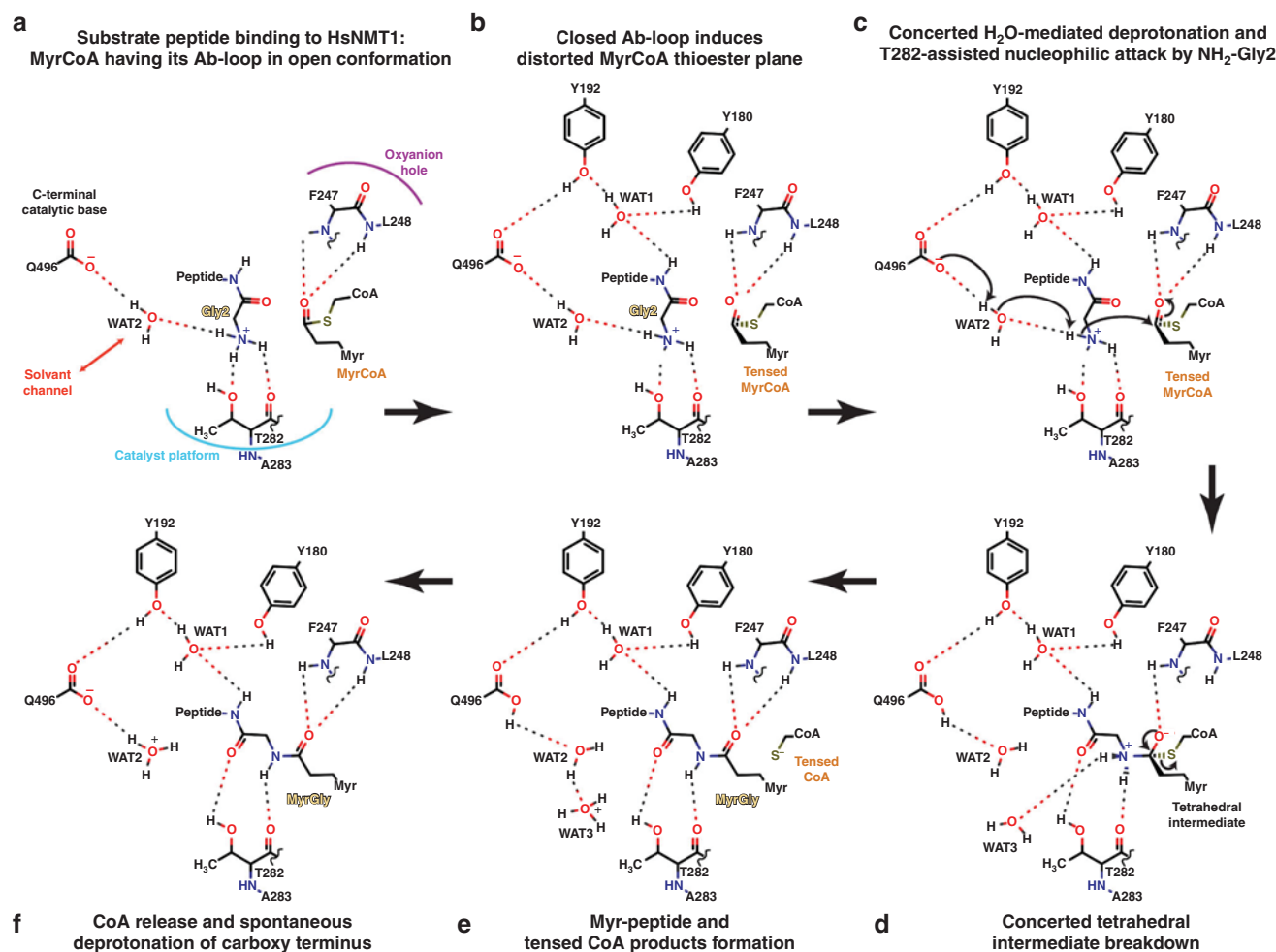


Fig. 7 Detailed N-Myristoylation mechanism deduced from reaction snapshots. **a** Substrate peptide binding to HsNMT1:MyrCoA having its Ab-loop in open conformation. **b** Closed Ab-loop induces distorted MyrCoA thioester plane. **c** Concerted H₂O-mediated deprotonation and T282-assisted nucleophilic attack by NH₂-Gly2. **d** Concerted Tetrahedral Intermediate breakdown. **e** Myr-peptide and tensed CoA products formation. **f** CoA release and spontaneous deprotonation of carboxy terminus. Hydrogen bonds between wat3 and backbone atoms of I495 and T282 were not shown for clarity. The reaction dynamics are also detailed in Supplementary Movie 1.

To challenge whether only unavailability of the free amino group was responsible for Ca3 rotation leading to ϵ -MYR, we determined the crystal structure of HsNMT1 in complex with an acetyl-Gly-Asn starting peptide (Fig. 8c). As expected, there was similar rotation to the N-acetyl-Asn peptide. Occurrence of a free amino group by either Gly or Lys is crucial for proper orientation of both chains towards the pocket for aa3 or the active site (Fig. 8c). Finally, determination of the crystal structure of HsNMT1 in complex with an acetyl-Gly-Lys starting peptide clearly showed Lys- ϵ -MYR (Fig. 8d, e). Again, we could capture different steps of the ϵ -MYR reaction, prior and after nucleophilic attack, in the different monomers of a single crystal. We show that the N ϵ of Lys3 interacts directly with the C-terminus catalytic base, the N ϵ of Lys3 occupying the position of wat2 observed in the HsNMT1:MyrCoA:X complex (Fig. 4a vs 8d). This suggests that Lys- ϵ -MYR, in contrast to Gly-MYR, is not water-mediated and consequently highlights that the solvent channel is not necessary for N ϵ deprotonation.

Discussion

The NMT catalytic mechanism described here provides an astonishing level of mechanistic sophistication involving motion

in the N- and C-terminal domains together with mobile loops and provides a unified explanation for the importance of key conserved residues.

Examination of the structures of GNAT enzymes bound to different ligands or inhibitors suggests some common conserved features with NMT (Fig. 1a–c, and Supplementary Fig. 14a, b). The carbonyl group of the β 5 C-terminal residue of some GNAT enzymes is likely to serve as a catalytic platform, placing the acceptor ammonium at an ideal distance and geometry relative to the thioester carbonyl. In the same vein, the last residue of GNAT β 4 promotes the water molecule network (wat1/wat2/base) required for the ammonium group deprotonation, but only when AcCoA and acceptor are both bound. Identification of the water molecule network comprising wat1 and wat2 would likely help to track the position of the catalytic base in different GNAT superfamily enzymes. The GNAT α 1 α 2-loop equivalent to the Ab-loop in NMT also seems to be required for acceptor ligand selectivity. Future determination of ternary complexes of other GNAT enzymes with both bound ligands will help to address the involvement of this loop in preforming the tetrahedral intermediate.

NMT is now well established as a valid target in parasitic and fungal diseases in which MYR is essential for the pathogen³³. The

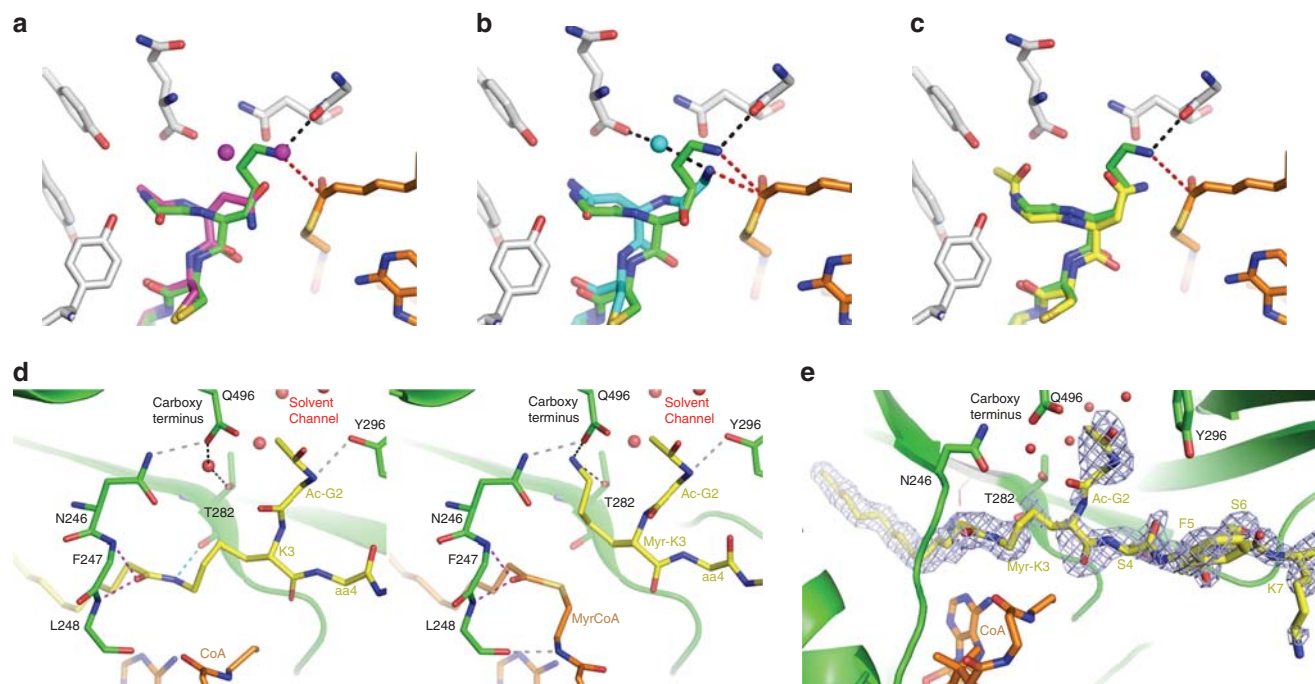


Fig. 8 Structural analysis and characterisation of the binding of a Lys side chain into the active site reactive cavity. The crystal structures 5O9T (HsNMT1/Ac-N₃CFSKPR), 6QRM (HsNMT1/ GN₃CFSKRRRA, this work) and 6SJZ (HsNMT1/Ac-GN₃CFSKPR, this work) were used in panels **a–c** to model the position of the side chain with the reactive epsilon amino group of a Lys in the active site instead of the alpha amino of a Gly2. **a** Gly-Lys modelled based on peptide Ac-Asn₃ of 5O9T (green) with 5O9T (peptide in magenta). **b** Same as **a**, but with Gly₂-Asn₃ from 6QRM (cyan). **c** Same as **a**, with peptide Ac Gly₂-Asn₃ (U) from 6SJZ (yellow). Selected amino acids of HsNMT1 are shown as grey sticks. MyrCoA moiety are shown as orange sticks. Wat1 linking the carboxy terminus of Gln496 to Gly2 ammonium is shown as a cyan sphere. Selected hydrogen bonds are displayed as black dashed line. Distance between either N α (Gly₂) or N ϵ (Lys₃) and thioester carbonyl are displayed as red dashed line. **d** Snapshots of ϵ -Myr reaction trapped in crystal structure of HsNMT1 (green; 6SK2, V) in complex with MyrCoA (orange) and peptide V (Ac-GlyLysSerPheSerLysProArg; yellow) showing substrates (left; chain B) and reaction products (right; chain A). **e** Detail of the "2foc omit" sigmaA-weighted electron density map (mF_{obs} - DF_{calc}, PHI_{calc}), shown in blue, for Myr-V peptide in 6SK2 structure. The omit electron density maps were calculated after a round of refinement in which the occupancy of the Myr-V peptide was set to zero. The e/Å³ value of the "2foc omit" sigmaA-weighted electron density map (2mF_{obs} - DF_{calc}, PHI_{calc}) at 1 rms (0.31 e/Å³) level was used as reference to set the contoured of the fofc omit electron density map shown.

dynamic catalytic mechanism revealed here opens up new opportunities for the design of inhibitors that exploit NMT plasticity. For example, there are now several examples of highly potent NMT inhibitors against both human and malaria parasite NMTs, but achieving a high level of selectivity has proven challenging to date due to high homology between NMTs at the substrate-binding site (Supplementary Fig. 14c–e). Indeed, this similarity was sufficient to develop IMP-1088, a 0.2 nM K_d human NMT inhibitor with potent antiviral properties, starting from fragments originally identified as *Plasmodium* NMT inhibitors³³. Alignment of the recently reported HsNMT1:IMP-1088 complex with the structures solved here highlights that the inhibitor occupies aa2–aa5 pockets, thus mimicking key aspects of peptide binding (Supplementary Fig. 14c–e). NMT amino acids involved in the interactions with both substrate and inhibitor molecules are conserved (Supplementary Fig. 2a), and a comparison of substrate-bound NMT and inhibitor-bound NMT suggests that more selective potency might be achieved through remodelling the interaction with the Gln496 C-terminal carboxyl group to maintain a water molecule at position +1 and repositioning a primary amine to react with MyrCoA, allowing the generation of a myristoylated inhibitor. Interestingly, examination of the closed Ab-loop intermediate structure reveals that the four residues forming the N-terminal helix starting from Pro105 in HsNMT1 or Leu20 in PfNMT of the B'A'-loop (see magenta box in Supplementary Fig. 14f) are not conserved in parasites (Supplementary Fig. 2a). No parasite NMT inhibitor design has

yet exploited this region of the binding pocket, and since this helix is critical for catalysis in HsNMT1 inhibitors, making specific interactions with the B'A'-loop of PfNMT may significantly increase selectivity. For instance, extension of the N-methyl moiety found in IMP-1088 could provide a means to probe this region of the pocket (Supplementary Fig. 14c–e).

We show that the uncovered NMT catalytic mechanism is fully compatible with ϵ -MYR of Lys residues located at the N-terminal side. When the N-terminus next to the Lys residue is blocked by an acylation event compatible with the size of the aa3 pocket, then Lys-MYR takes place. N-acetylation of Gly-Lys nascent peptides can be disfavoured since NatA – which may compete with NMT in vivo – generally selects negatively charged Gly-starting substrates⁸. Our data indicate that Gly-MYR is favoured over Lys-MYR when the alpha and epsilon amino groups are both available on the same peptide. The pK_a of the conjugate acid of the alpha amino group is below that of epsilon (7.7 ± 0.5 vs. 10.5 ± 1.1 ³⁴) but close to both physiological pH (7.2 ± 0.2) and the pH of the assay performed at the optimal value for catalysis (pH 8.0; Towler et al.⁵). This might suggest that when both groups are available at the same N-terminal polypeptide, binding of the fully-protonated ϵ -amino group is favoured over the α -amino in the negatively charged environment of the aa3 pocket (see Fig. 1d in Castrec et al.⁸). This imbalance in favour of Gly-MYR could also result from the slightly longer side chain of Lys compared to Gly, which fits less well in the active site (Fig. 8d). Occurrence of Lys-MYR on an alpha amino blocked peptide indicates that the

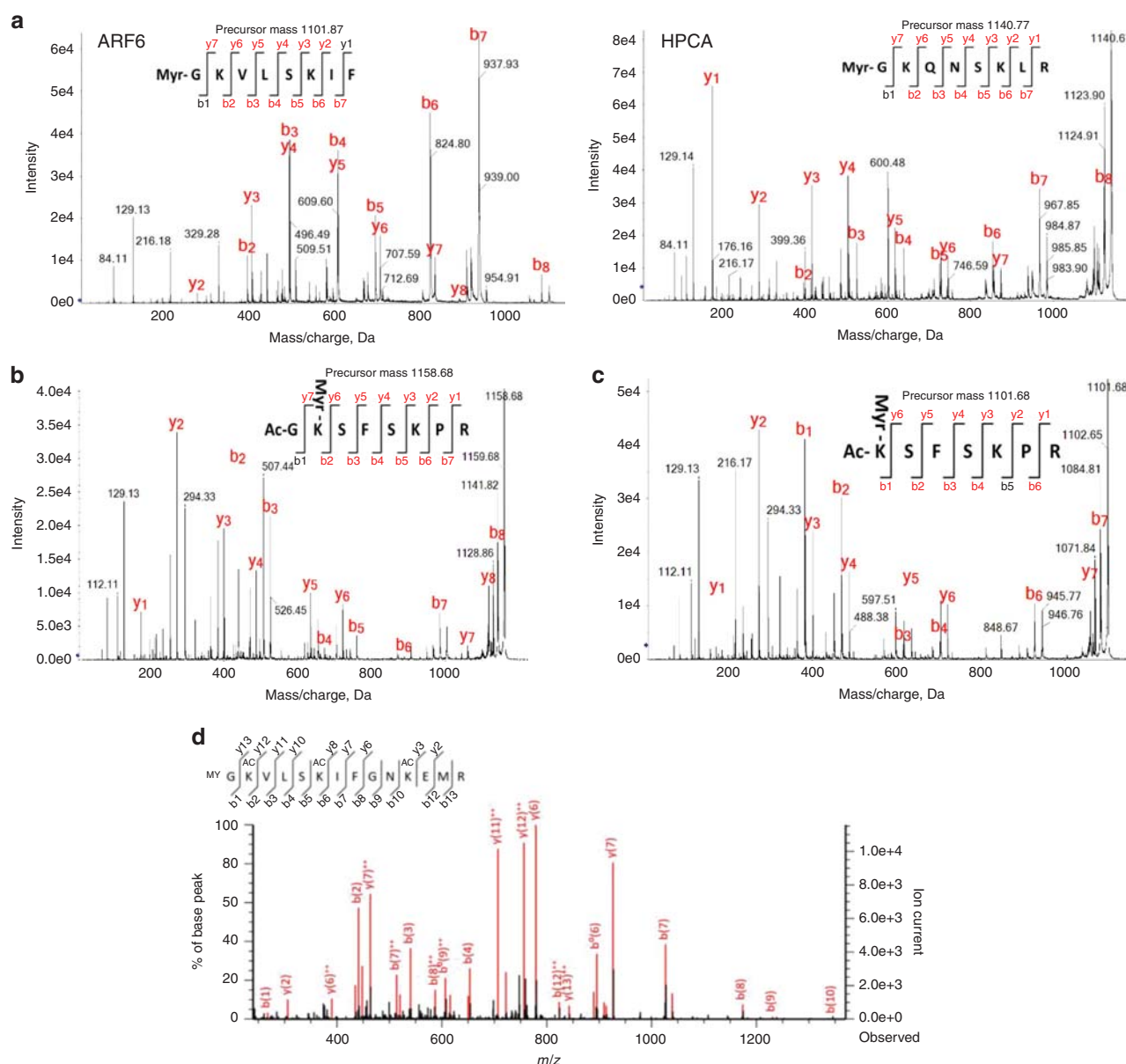


Fig. 9 NMT catalyses Lys-MYR when the NH₂ amino group is blocked. **a–c** MS/MS spectra resulting from MALDI-ToF-ToF analysis performed as described in the Methods after NMT incubation with various octapeptides are shown. The control MS spectra in the absence or presence of NMT are displayed in Supplementary Fig. 13. MS/MS spectra from: **a** ARF6 (left) and HPCA-derived peptides (right); **b** AcGKSFSPK (U peptide); and **c** AcKSFSPK (V peptide). **d** Example of annotated MS/MS spectrum of the Myristoylated peptide GKVLSKIFGNKEMR from ARF6 (P62330); mass spectrometry of RT4 urinary bladder cells is as in Castrec et al.⁸. Acetylation of the Lys side chains of the protein sample was chemically induced in the presence of deuterated acetate. Both y_{13} and b_1 ions favour Gly- over Lys-MYR.

modification is plausible *in vivo* on N-terminal Lys- or potentially Gly-Lys N-terminal peptides resulting, for instance, from post-translational proteolytic cleavage. However, we observed no such event in thousands of MS spectra produced in our labs over recent years or in a recent large-scale survey of human protein modifications³⁵. This included Arf6, for which MYR was frequently observed in ESI-MS experiments after chemical acetylation of the sample to enhance detection of all free amino groups. MS/MS spectra indeed systematically favoured Gly- over Lys-MYR (Fig. 9d).

Why has no N-terminal Lys-MYR yet been identified *in vivo*? Two possible reasons include the likely reversibility of this modification and its putative role in an N-degron pathway. A number of active acyl hydrolases targeting Lys-MYR have been

described over recent years including Sirtuin6, HDAC8 and HDAC11^{32,36–38}, making Lys-MYR a reversible modification dependent on the activity of these specific hydrolases, in contrast to the irreversible Gly-MYR. This implies that Lys-MYR, as with other reversible modifications (i.e., phosphorylation and ϵ -acetylation), might be intimately related to signal transduction pathways and tightly regulated by as yet unknown biotic or abiotic signals. Alternatively, it is tempting to speculate that N-terminal Lys-MYR may be part of an N-degron, and that an important role of Lys-MYR deacylases would be to ensure protein quality control by preventing intracellular Lys-MYR accumulation. Previous studies suggested an extension of the N-degron pathway³⁹ to Gly-starting proteins⁸, which was supported by recent models in which conditional exposure of this degron

allows degradation of aberrant proteins that have escaped N-terminal MYR^{4,40}. The Lys-MYR N-degron would further participate in this proofreading in a process dedicated to cellular MYR checking. This might be crucial at the physiological level given the number of major pathways in which known Gly-MYR proteins participate, preventing competition with irrelevant MYR proteins. Finally, our data do not support a mechanism or role for NMT in ϵ -MYR of internal Lys residues and further suggest that other GNAT members, including N-alpha acetyl transferases such as NatA or lysine aminotransferases, might also acetylate both Lys epsilon and alpha amino-terminal residues.

Methods

Protein production and crystallisation. We used two different HsNMT1 constructs containing either 99–496 residues (Long-HsNMT1) or 109–496 residues (Short-HsNMT1). Long-HsNMT1 was used to solve the HsNMT1 complexes with peptides U, V, X and Z, while Short-HsNMT1 was used to solve the HsNMT1 complex with peptide Y. The open reading frames encoding two forms were both amplified by PCR using the full-length cDNA as a template. The PCR products were cloned into a modified pET28 vector featuring an N-terminal hexahistidine tag followed by Tobacco Etch Virus (TEV) cleavage site. Site-directed variants were obtained as described in Supplementary Methods. All primers used in this study are reported in Supplementary Table 3.

The recombinant plasmid encoding short-HsNMT1 was introduced into *Escherichia coli* BL21(DE3) pRareS. Cells were grown for 20 h at 18 °C in Overnight Express Instant TB auto-induction medium (Novagen). Cells were sonicated in 20 mM Tris-HCl (pH 7.5), 0.5 M NaCl (buffer A) supplemented with 10 mM imidazole, 1 mM MgCl₂, protease inhibitors, bovine pancreatic DNase I and Triton X-100 (0.1% (v/v)). The lysate was loaded directly onto a HisTrap FF Crude column (GE Healthcare, 1 mL) and protein eluted by a linear gradient of 10–500 mM imidazole in buffer A. The NMT pool was diluted 20-fold in 20 mM Tris-HCl (pH 8.9), 1 mM DTT (buffer B) before loading onto an anion-exchange column (GE Healthcare, 1 mL, HiTrap Q HP) and elution by a linear gradient of 25–500 mM NaCl in buffer B. Gel filtration on a Superdex 75 16/60 column (GE Healthcare) in 250 mM NaCl, 1 mM EGTA and 1 mM EDTA, 20 mM Tris-HCl (pH 8.5) buffer was used as final purification step. Short-HsNMT1 purified samples for crystallisation were concentrated to 20 mg/ml in 10 mM Tris-HCl (pH 8.5), 125 mM NaCl, 1 mM EGTA and 1 mM EDTA.

The long-HsNMT1 plasmid was introduced into *E. coli* BL21-PLys Rosetta2 cells (Novagen). After growth at 22 °C and induction, cells were lysed by sonication at 4 °C in 20 mM Tris-HCl (pH 8.0), 0.5 M NaCl, 5 mM 2-mercaptoethanol and 5 mM imidazole (buffer C). The lysate was loaded onto an immobilised nickel ion affinity chromatography (HisTrap Crude FF, 5 mL, GE Healthcare) and washed with buffer C with 25 mM imidazole. Elution was carried out at a 2-mL/min with buffer C plus 0.5 M imidazole with a linear gradient. The pool of purified protein was next dialysed overnight against 20 mM Tris-HCl (pH 8.9), 50 mM NaCl, 25 mM imidazole and 5 mM 2-mercaptoethanol in the presence of TEV protease (1 mg per 15 mg purified protein). His-tag free HsNMT1 was purified with a HisTrap Crude FF. The flow-through with HsNMT1 was diluted five times in buffer B and applied to an anion exchange chromatography column (HiTrap Q FF, 5 mL, GE Healthcare). HsNMT1 was eluted with buffer B supplemented with 0.2 M NaCl. Highly purified long-HsNMT1 fractions were pooled, concentrated to 15–30 mg/mL with an Amicon centrifugal filtration device (Merck Millipore) in 20 mM Tris-HCl (pH 8.0), 0.2 M NaCl, 1 mM DTT and stored at –80 °C before crystallisation trials. Protein samples for biochemical analysis were stored at –20 °C in the same buffer plus 55% glycerol.

Short-HsNMT1 was crystallised in complex with MyrCoA (HsNMT1:MyrCoA crystals) using the sitting-drop vapour diffusion method at 20 °C⁴¹. Crystals of HsNMT1:MyrCoA in complex with the N-terminal substrate peptide of the tyrosine-protein kinase Src (p60-Src) protein (peptide Y with sequence “Gly-Ser-Asn-Lys-Ser-Lys-Pro-Lys”) were obtained by soaking HsNMT1:MyrCoA crystals in a cryosolution containing the precipitant components (16–18% (v/w) PEG 4K, 5 mM NaCl, 100 mM sodium citrate (pH 4.5) and 2.5% (v/v) glycerol), peptide Y at 150 mM and 25% (v/v) glycerol for 15 min. HsNMT1:MyrCoA:Y crystals were then flashed cooled in liquid nitrogen and stored for data collection. A complete X-ray data set at 2.10 Å resolution was collected at 100 K using synchrotron radiation at Diamond Light Source (Oxford, UK).

Crystals of long-HsNMT1 (HsNMT1:MyrCoA:X, HsNMT1:MyrCoA:U and HsNMT1:MyrCoA:V) were obtained by co-crystallisation using the hanging-drop vapour diffusion method at 20 °C. Crystallisation droplets were formed by mixing 2 μ L of the HsNMT1:MyrCoA:peptide complex (ratio 1:1.5:1.5) at 7.5 mg/mL (i.e., ~160/250/250 μ M) with 2 μ L of the precipitant solution containing 0.1 M MgCl₂, 0.2 M NaCl, 0.1 M sodium citrate pH 5.6 and 18–24% (w/v) PEG 6 K or 8 K. Crystals were cryoprotected in the reservoir solution supplemented with 15% (v/v) glycerol and flash cooled in liquid nitrogen. Complete X-ray datasets of HsNMT1:MyrCoA:X and HsNMT1:MyrCoA:U were collected at $\lambda = 0.96770$ Å and $\lambda = 0.96600$ Å at 100 K in the ID30a3 and ID30a1 beamlines, respectively, at the European Synchrotron Radiation Facility (ESRF, France). Complete X-ray

dataset of HsNMT1:MyrCoA:V was collected at $\lambda = 0.98400$ Å from a single crystal at 100 K in the Proxima1 beamline, at the French National Synchrotron Facility (SOLEIL).

Datasets were integrated with XDS⁴² and scaled and reduced using AIMLESS from the CCP4 package⁴³. Crystals belonged to the space group P2₁2₁2 with similar unit cell parameters (Supplementary Table 1) and contained two NMT molecules per asymmetric unit. Structure resolution was accomplished in all cases using the molecular replacement method. The HsNMT1:MyrCoA:Y structure was solved using PHASER⁴⁴ and the HsNMT1:MyrCoA binary complex (PDB entry 4C2Y) as a search model. The structure of the HsNMT1:MyrCoA:X complex was solved using MOLREP⁴⁵ and protein coordinates of a Long-HsNMT1 model (PDB entry 509V) as a search model. The HsNMT1:MyrCoA:V structure was solved using PHASER⁴⁴ with PDB entry 509V as a search model. Structures were subjected to alternating refinement cycles using REFMAC-5 and PHENIX and manual model building using COOT^{46–48}. NCS restrictions were applied. In both structures, the electron density maps allowed us to model most of both HsNMT1 molecules of the asymmetric unit, with the exception of the first 6 residues (residues 109–114) and some flexible loops (A182–184, A316–317, A408–413 and B313–318) in Short-HsNMT1:MyrCoA; and the first 6 residues (residues 99–104) in HsNMT1:MyrCoA:X, HsNMT1:MyrCoA:U and HsNMT1:MyrCoA:V, which were flexible and thus could not be traced. The good quality of the electron density maps also enabled the refinement of substrate peptide, reaction intermediate, and reaction product molecules bound to HsNMT1 in each complex. In the HsNMT1:MyrCoA:Y complex, linkers and cif files for the refinement of the MyrCoA-Y and the MyrY molecules were generated with JLIGAND as part of the CCP4i package⁴⁹. In the refinement of the HsNMT1:MyrCoA:X, HsNMT1:MyrCoA:U and HsNMT1:MyrCoA:V complexes, chemical compound libraries were generated using the PRODRG server⁵⁰ in combination with eLBOW from the PHENIX suite. Difference electron density maps were calculated in PHENIX. The geometry of the final models was validated using MOLPROBITY⁵¹. Figures were generated using PYMOL (DeLano Scientific LLC, <http://pymol.sourceforge.net/>). X-ray data collection and refinement statistics are summarised in Supplementary Table 1 and 2.

Fluorescence-based measurement of NMT activity. NMT activity was assayed at 30 °C in a coupled assay as previously described⁸. Briefly, a reaction mixture containing MyrCoA and different concentrations of peptide acceptors was pre-incubated for 3 min at 30 °C before starting the reaction with HsNMT1. MYR kinetics were followed continuously for 10 min in triplicate, and the data were fitted over the first 2 min to obtain the initial velocity. The kinetic parameters (k_{cat} and K_m) were obtained with the Enzyme Kinetics module 1.2 of Sigma Plot (version 9.0) by nonlinear Michaelis–Menten equation fitting.

Mass spectrometry. For MALDI-ToF/ToF analysis, 300 μ L of a mixture containing 50 mM Tris (pH 8), 0.193 mM EGTA, 1 mM MgCl₂, 1 mM DTT, 5 μ M sodium cholate, 0.04 mM Myr-CoA solution (stock solution 0.2 mM Myr-CoA, 10 mM sodium acetate, 2.5 μ M sodium cholate), 0.5 μ M NMT and 100 μ M of synthetic peptide (Genscript, Piscataway, NJ) were incubated at 30 °C. The MYR reaction was followed over time by collection of 10 μ L samples further diluted in 90 μ L of water/acetonitrile solution. The different samples were then diluted five times in the matrix solution made of 5 mg/mL of α -cyano-4-hydroxycinnamic acid solubilized in water/formic acid/acetonitrile (50/50/0.1%). In all, 1 μ L of each dilution was spotted on a metal target and dried. MS and MS/MS spectra of each sample were acquired with an AB SCIEX 5800 instrument in positive ion mode.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The 3D structures of NMT in complex with MyrCoA and the peptides reported here have been deposited at the PDB under codes 6QRM (peptide X), 6EHJ (peptide Y), 5O9S (peptide Z), 6SJZ (peptide U), 6SK2 (peptide V), 6SKJ (peptide S), 6SK8 (peptide T) and 6SK3 (peptide S). Other data are available from the corresponding authors upon reasonable request.

Received: 17 September 2019; Accepted: 4 February 2020;

Published online: 28 February 2020

References

- Bhatnagar, R. S., Ashrafi, K., Futterer, K., Waksman, G. & Gordon, J. I. In *The enzymes*, Vol. XXI (Protein lipidation) (eds F. Tamanoi & D. S. Sigman) 241–286 (Academic Press, 2001).
- Thinon, E. et al. Global profiling of co- and post-translationally N-myristoylated proteomes in human cells. *Nat. Commun.* **5**, 4919 (2014).

3. Traverso, J. A. et al. Roles of N-Terminal fatty acid acylations in membrane compartment partitioning: Arabidopsis h-type thioredoxins as a case study. *Plant Cell* **25**, 1056–1077 (2013).
4. Timms, R. T. et al. A glycine-specific N-degron pathway mediates the quality control of protein N-myristoylation. *Science* <https://doi.org/10.1126/science.aaw4912> (2019).
5. Towler, D. A. et al. Purification and characterization of yeast myristoyl CoA: protein N- myristoyltransferase. *Proc. Natl Acad. Sci. USA* **84**, 2708–2712 (1987).
6. Ritzefeld, M., Wright, M. H. & Tate, E. W. New developments in probing and targeting protein acylation in malaria, leishmaniasis and African sleeping sickness. *Parasitology* **145**, 157–174 (2018).
7. Tate, E. W., Bell, A. S., Rackham, M. D. & Wright, M. H. N-Myristoyltransferase as a potential drug target in malaria and leishmaniasis. *Parasitology* **141**, 37–49 (2014).
8. Castrec, B. et al. Structural and genomic decoding of human and plant myristoylomes reveals a definitive recognition pattern. *Nat. Chem. Biol.* **14**, 671–679 (2018).
9. Udenwole, D. I. et al. Myristoylation: an important protein modification in the immune response. *Front. Immunol.* **8**, 751 (2017).
10. Wright, M. H., Heal, W. P., Mann, D. J. & Tate, E. W. Protein myristoylation in health and disease. *J. Chem. Biol.* **3**, 19–35 (2010).
11. Das, U., Kumar, S., Dimmock, J. R. & Sharma, R. K. Inhibition of protein N-myristoylation: a therapeutic protocol in developing anticancer agents. *Curr. Cancer Drug Targets* **12**, 667–692 (2012).
12. Wright, M. H. et al. Validation of N-myristoyltransferase as an antimalarial drug target using an integrated chemical biology approach. *Nat. Chem.* **6**, 112–121 (2014).
13. Wen, Z. et al. N-myristoyltransferase deficiency impairs activation of kinase AMPK and promotes synovial tissue inflammation. *Nat. Immunol.* **20**, 313–325 (2019).
14. Corpas-Lopez, V. et al. Pharmacological Validation of N-Myristoyltransferase as a Drug Target in Leishmania donovani. *ACS Infect. Dis.* <https://doi.org/10.1021/acscinfdis.8b00226> (2018).
15. Bayliss, T. et al. Design and synthesis of brain penetrant trypanocidal N-myristoyltransferase inhibitors. *J. Med. Chem.* **60**, 9790–9806 (2017).
16. Kallemijn, W. W. et al. Validation and invalidation of chemical probes for the human N-myristoyltransferases. *Cell Chem. Biol.* **26**, 892–900 e894 (2019).
17. Salah Ud-Din, A. I., Tikhomirova, A. & Roujeinikova, A. Structure and functional diversity of GCN5-related N-acetyltransferases (GNAT). *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms17071018> (2016).
18. Abe, J. et al. A fumagillin derivative angiogenesis inhibitor, AGM-1470, inhibits activation of cyclin-dependent kinases and phosphorylation of retinoblastoma gene product but not protein tyrosyl phosphorylation or protooncogene expression in vascular endothelial cells. *Cancer Res.* **54**, 3407–3412 (1994).
19. Dyda, F., Klein, D. C. & Hickman, A. B. GCN5-related N-acetyltransferases: a structural overview. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 81–103 (2000).
20. Weston, S. A. et al. Crystal structure of the anti-fungal target N-myristoyl transferase. *Nat. Struct. Biol.* **5**, 213–221 (1998).
21. Farazi, T. A., Manchester, J. K. & Gordon, J. I. Transient-state kinetic analysis of Saccharomyces cerevisiae myristoylCoA:protein N-myristoyltransferase reveals that a step after chemical transformation is rate limiting. *Biochemistry* **39**, 15807–15816 (2000).
22. Olaleye, T. O. et al. Peptidomimetic inhibitors of N-myristoyltransferase from human malaria and leishmaniasis parasites. *Org. Biomol. Chem.* <https://doi.org/10.1039/c4ob01669f> (2014).
23. Farazi, T. A., Waksman, G. & Gordon, J. I. Structures of Saccharomyces cerevisiae N-myristoyltransferase with bound myristoylCoA and peptide provide insights about substrate recognition and catalysis. *Biochemistry* **40**, 6335–6343 (2001).
24. Bhatnagar, R. S. et al. Structure of N-myristoyltransferase with bound myristoylCoA and peptide substrate analogs. *Nat. Struct. Biol.* **5**, 1091–1097 (1998).
25. Farazi, T. A., Waksman, G. & Gordon, J. I. The biology and enzymology of protein N-myristoylation. *J. Biol. Chem.* **276**, 39501–39504 (2001).
26. Wu, J. et al. Crystal structures of Saccharomyces cerevisiae N-myristoyltransferase with bound myristoyl-CoA and inhibitors reveal the functional roles of the N-terminal region. *J. Biol. Chem.* **282**, 22185–22194 (2007).
27. Bürgi, H. B., Dunitz, J. D., Lehn, J. M. & Wipff, G. Stereochemistry of reaction paths at carbonyl centres. *Tetrahedron* **30**, 1563–1572 (1974).
28. Rudnick, D. A. et al. Kinetic and structural evidence for a sequential ordered Bi Bi mechanism of catalysis by Saccharomyces cerevisiae myristoyl-CoA: protein N-myristoyltransferase. *J. Biol. Chem.* **266**, 9732–9739 (1991).
29. Farazi, T. A., Manchester, J. K., Waksman, G. & Gordon, J. I. Pre-steady-state kinetic studies of Saccharomyces cerevisiae myristoylCoA:protein N-myristoyltransferase mutants identify residues involved in catalysis. *Biochemistry* **40**, 9177–9186 (2001).
30. Stevenson, F. T., Bursten, S. L., Locksley, R. M. & Lovett, D. H. Myristyl acylation of the tumor necrosis factor alpha precursor on specific lysine residues. *J. Exp. Med.* **176**, 1053–1062 (1992).
31. Stevenson, F. T., Bursten, S. L., Fantom, C., Locksley, R. M. & Lovett, D. H. The 31-kDa precursor of interleukin 1 alpha is myristoylated on specific lysines within the 16-kDa N-terminal propeptide. *Proc. Natl Acad. Sci. USA* **90**, 7245–7249 (1993).
32. Cao, J. et al. HDAC11 regulates type I interferon signaling through defatty-acylation of SHMT2. *Proc. Natl Acad. Sci. USA* **116**, 5487–5492 (2019).
33. Mousnier, A. et al. Fragment-derived inhibitors of human N-myristoyltransferase block capsid assembly and replication of the common cold virus. *Nat. Chem.* **10**, 599–606 (2018).
34. Pace, C. N., Grimsley, G. R. & Scholtz, J. M. Protein ionizable groups: pK values and their contribution to protein stability and solubility. *J. Biol. Chem.* **284**, 13285–13289 (2009).
35. Devabhaktuni, A. et al. TagGraph reveals vast protein modification landscapes from large tandem mass spectrometry datasets. *Nat. Biotechnol.* **37**, 469–479 (2019).
36. Jiang, H. et al. SIRT6 regulates TNF-alpha secretion through hydrolysis of long-chain fatty acyl lysine. *Nature* **496**, 110–113 (2013).
37. Aramsangtienchai, P. et al. HDAC8 catalyzes the hydrolysis of long chain fatty acyl lysine. *ACS Chem. Biol.* **11**, 2685–2692 (2016).
38. Moreno-Yruela, C., Galleano, I., Madsen, A. S. & Olsen, C. A. Histone deacetylase 11 is an epsilon-N-myristoyllysine hydrolase. *Cell Chem. Biol.* **25**, 849–856 e848 (2018).
39. Varshavsky, A. N-degron and C-degron pathways of protein degradation. *Proc. Natl Acad. Sci. USA* **116**, 358–366 (2019).
40. Tran, A. The N-end rule pathway and Ubr1 enforce protein compartmentalization via P2-encoded cellular location signals. *J. Cell Sci.* **132**, jcs231662 (2019).
41. Goncalves, V. et al. A fluorescence-based assay for N-myristoyltransferase activity. *Anal. Biochem.* **421**, 342–344 (2012).
42. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800 (1993).
43. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
44. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
45. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1225 (1997).
46. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
47. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
48. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
49. Lebedev, A. A. et al. JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 431–440 (2012).
50. Schüttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355–1363 (2004).
51. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

Acknowledgements

This work was supported by MMV (grant 15–0054 to EWT), Cancer Research UK (Programme Foundation Award C29637/A20183 to EWT), the 7th Framework Programme of the European Union (PIEF-GA-2013-623648 to MR), and ANR PalMyProt grant to TM (ANR-2010-BLAN-1611-01). This work has benefitted from the support of the Labex Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS), from the I2BC crystallisation platform supported by FRISBI ANR-10-INSB-05-01 and from the facilities and expertise of the I2BC proteomic platform SICaPS, supported by IBSA, Ile de France Region, Plan Cancer, CNRS and Paris-Sud University. FR is supported by the PhD fellowship ARDoc program from Region Ile-de-France (17012695). We thank the staff at the I03 beamline, Diamond Light Source (Harwell Science and Innovation Campus, Oxfordshire, UK); the staff at the X-ray Crystallography Facility, Centre for Structural Biology at Imperial College London; Soleil for provision of synchrotron radiation facilities and the staff of Proxima1 beamline; the European Synchrotron Radiation Facility (ESRF) staff, particularly Matthew Bowler and Christoph Mueller-Dieckmann, for help with data collection; Remi Serwa for sequence analysis; Andy Bell for discussion and input on inhibitor development based on the structures here reported; and Yvain Nicolet for fruitful discussion concerning the NMT1 catalysis mechanism.

Author contributions

C.G., E.W.T. and T.M. conceived the project and supervised the experiments. C.D. performed cloning, purification and structural analysis of 509S, 6QRM, 6SJZ, 6SK2, 6SKJ, 6SK8 and 6SK3. I.P.D. carried out the purification of 6EHJ and structural characterisation was performed by I.P.D. and C.D.; F.R. participated in cloning and protein purification and performed the MS experiments. T.A. cloned, purified all HsNMT mutants and participated in HsNMT1 biochemical studies and crystallisation of the majority of complexes. P.L. participated in structural characterisation of HsNMT1ΔC2 and HsNMT1ΔC3 complexes. M.R. synthesised and purified the Src N-terminal peptide used in the study. M.S. assisted in Short HsNMT1 protein production and crystallisation. E.C. provided support in structural characterisation and analysis of 6EHJ. C.G., C.D., T.M., E.W.T. and I.P.D. analysed the data; C.G., T.M. and C.D. wrote the manuscript with involvement of E.W.T. and I.P.D.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-14847-3>.

Correspondence and requests for materials should be addressed to T.M., E.W.T. or C.G.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

Chapitre IV

IV. Éléments de spécificité contribuant à la lysine myristoylation, mécanisme associé et conception raisonnée d'inhibiteurs de la NMT

Dans le chapitre précédent, j'ai montré que la NMT est capable de myristoyler les lysines en plus des glycines. Dans un second article pour lequel je suis premier auteur, je me suis focalisé sur cette observation initiale et j'ai cherché à mieux comprendre et décrire le phénomène de la myristoylation sur le groupement amine de la chaîne latérale de lysines (ϵ -myristoylation) par la NMT. J'ai ainsi cherché à déterminer les mécanismes associés et les règles qui en régissent la spécificité. En particulier, j'ai essayé aussi de déterminer l'issue de la compétition entre les deux types de modifications, lorsque les deux groupes accepteurs (α -Gly ou ϵ -Lys) sont proposés à la NMT sur le même substrat. De telles situations sont pertinentes physiologiquement, comme j'y reviendrai au cours de ce chapitre.

Dans un premier temps, j'ai défini et utilisé un peptide modèle possédant une lysine en position trois (G₂KSFSKPR ; référence, GK). Ce peptide est caractérisé par une séquence « idéale » issue du motif construit à partir de l'ensemble des séquences connues comme myristoylées (Castrec *et al*, 2018; Meinnel *et al*, 2020), à l'exception de la lysine en position 3 que j'ai insérée ; il en résulte une compétition possible entre les deux types de myristoylation. Il faut noter qu'à priori une lysine en position trois est un élément très défavorable à la myristoylation du fait de la longueur et de l'encombrement stérique de la chaîne latérale ; seuls des éléments très favorisés (S₆, K₇) permettent de compenser cet effet et permettent la modification de ce type de substrat, comme c'est le cas dans le cadre du peptide GK. J'ai analysé ce peptide en spectrométrie de masse après myristoylation *in vitro* en présence de NMT. Les spectres obtenus en MALDI-TOF-TOF incluant les analyses MS₂ indiquent la myristoylation fortement majoritaire sur la glycine en position deux. Aucun indice d' ϵ -Lys n'est décelable, à la sensibilité du test près. Les données cinétiques que j'ai obtenues

montrent également que ce peptide se comporte comme un très bon substrat de la NMT (**Figure 1b** et **Tableau supplémentaire 1a**). J'ai ensuite déterminé le type de myristoylation de toutes les protéines humaines commençant par GK y compris les six membres étroitement liés de la famille HPCA/NCS1 et ARF6 et dans tous les cas seuls la G-myristoylation a été observé sans ambiguïté et ce avec des efficacités catalytiques dans la même gamme que celles rapportées pour la G-myristoylation (**Figure 1b, jeu de données supplémentaire 1** et **Tableau supplémentaire 1a**). J'ai pu remarquer que la plupart des protéines humaine commençant par GK dispose du motif S₆K₇ favorisant la myristoylation, comme nous l'avons déterminé dans l'article lié au chapitre précédent (Dian *et al*, 2020). Une nouvelle structure cristalline obtenue à haute résolution de la HsNMT1 co-cristallisée en présence d'un peptide correspondant à HPCA révèle que la chaîne latérale de la lysine en position trois se positionne très profondément au sein de la poche numéro trois (**Figure 1c**). Ce placement de la lysine induit un déplacement de la glycine deux pour le rapprocher du groupement thioester du MyrCoA (**Figure 1c,d**) mais conserve l'interaction avec la molécule d'eau numéro deux (W2) qui est requise pour la G-myristoylation en médiant l'interaction avec la base catalytique, le groupe carboxy-terminus de la NMT. Le léger déplacement du groupe réactif induit par la présence de la chaîne latérale de la lysine en position trois rend le peptide moins favorable à la myristoylation qu'un peptide avec une chaîne latérale non chargée et plus courte en position 3. J'ai donc étudié si les peptides GK possédait bien les mêmes règles de reconnaissance que celles d'un peptide G^XK (où ^X signifie tout acide aminé autre que X). Pour cela, j'ai analysé la capacité de la HsNMT1 à acyler une série de peptides dérivés du N-terminal de NCS1, en variant les acides aminés six et sept. Avec les variants de la lysine sept changée en glycine ou de la serine six en alanine, j'ai observé que seule la G-myristoylation est toujours la seule modification produite quand bien même j'ai pu vérifier que l'efficacité catalytique est considérablement diminuée (**Tableau supplémentaire 1b** et **jeu de données supplémentaire 1**). J'en conclus que le motif constitué d'une serine en position six et de la lysine en position sept est tout aussi important pour les peptides commençant par GK que les peptides

commençant par G^K (^= « exclu ») comme vu dans (Dian *et al*, 2020). J'ai ensuite étudié si la longueur de la chaîne latérale de la lysine sept avait une influence sur la myristoylation, pour cela j'ai utilisé des peptides avec des acides aminés non naturels en position sept, qui diffère de la lysine seulement par la longueur de la chaîne latérale. Lorsque l'on utilise des acides aminés comme l'ornitine (Orn, réduit d'un CH₂) ou le di-amino butyrate (Dab, réduit de deux CH₂), j'ai pu observer peu d'influence au niveau des efficacités catalytiques de la G-myristoylation, suggérant le maintien du pont salin avec la boucle Ab précédemment décrite. Un raccourcissement supplémentaire en di-amino propionate (Dap, chaîne basique courte CH₂NH₂) ou l'élimination du caractère basique avec de l'homocystéine (Hcy) réduit considérablement l'efficacité catalytique mais favorise toujours uniquement la G-myristoylation (**Tableau supplémentaire 1b** et **jeu de données supplémentaire 1**). Pour aller plus loin sur les règles régissant la myristoylation, j'ai remplacé la glycine 2 par une alanine sur un peptide de séquence A2NCFSKPR. Une alpha myristoylation est observée sur ce peptide en spectrométrie de masse et cristallographie (**Données étendues 1a, jeu de données supplémentaire 1** et **Tableau supplémentaire 1a**), mais les données catalytiques ont montré que ce peptide était très inefficacement modifié (**Figure 1b**). Pour conclure l'ensemble des données obtenues indique que les peptides GK conduisent préférentiellement à la G-myristoylation malgré le fait que la lysine trois soit moins favorable. Pour la myristoylation de ce type de peptide le contexte du reste du peptide devient très important, comme la serine 6 et la lysine 7.

Pourtant, il a été démontré que la protéine ARF6 acceptait la K-myristoylation *in vivo* lorsque le N-terminal est bloqué, par une myristoylation soit par une acétylation (Kosciuk *et al*, 2020). J'ai donc cherché à comprendre les règles de cette myristoylation ; pour cela, j'ai myristoylé différents peptides, HPCA, ARF6 et G₂KSFSKPR possédant un N-terminal réactif bloqué par une acétylation. Comme prévu, ces peptides ont été myristoylés sur la chaîne latérale de la lysine (**Figure 2a, Tableau supplémentaire 1c** et **jeu de données supplémentaires 1**) et même lorsque l'on remplace la lysine par une ornithine (**Figure 2a**). En revanche, j'ai pu observer que les efficacités catalytiques

(k_{cat}/K_m) des peptides acétylés étaient inférieures d'un ordre de grandeur à celles de la G-myristoylation (**Tableau supplémentaire 1a,c**) et c'est surtout au niveau du k_{cat} que l'on observe la majeure diminution, suggérant que la catalyse soit l'étape limitante. Une série de structure haute résolution de HsNMT1 en présence de différents peptides (acG[Orn], acGN, acGK, acGN), nous a permis de mieux comprendre ce phénomène. En effet, on peut observer sur les structures des variants acG[Orn] et acGK une rotation autour du C $_{\alpha}$ de la chaîne latérale de l'acide aminé trois. Cette rotation positionne à la fois le groupement acG dans la poche 3 et la chaîne latérale de l'acide aminé 3 en contact direct avec la base catalytique (**Données supplémentaires 2a-d**), sans médiation par une molécule d'eau comme c'est le cas pour la G-myristoylation (**Figure 1c,d**). Cette liaison directe du groupement amino réactif de la chaîne latérale de la lysine à la base catalytique explique très probablement une K-myristoylation efficace mais plus lente par rapport à la G-myristoylation. Sachant que la G-myristoylation est préférée à la K-myristoylation quand l'amine réactif de la glycine est disponible, j'ai étudié comment se comportait un peptide avec une lysine en position trois et un acide aminé moins favorable qu'une glycine en position deux, comme une alanine. Pour cela, j'ai testé des mutants G₂A de HPCA, ARF6 et AKSFSKPR et dans tous les cas, que ce soit par cristallographie ou par MS, ce sont les lysines qui sont myristoylées (**Figure 2b-e**, **Données étendues 3** et **Tableau supplémentaire 1c**). La structure cristalline de AKSFSKPR en complexe avec HsNMT1 (**Figure 2b,c**) a montré que ces types de peptide réagissent de la même façon que les peptides ac-GK, mais les analyses enzymatiques montrent qu'ils ont un K_m inférieur ce qui est probablement dû à des interactions de l'alanine dans la poche numéro trois (**Figure 2b**). De plus pour aller plus loin, j'ai testé d'autres résidus que la glycine et l'alanine (**Figure 2b**). Collectivement, nos données révèlent que la K-myristoylation ne dépend pas du premier acide aminé mais repose sur des résidus optimaux en position 5-7. Le motif [\wedge G]KX[FL]SK apparaît comme le déterminant de la K-myristoylation (où \wedge X signifie tout résidu sauf X). En présence d'une lysine en position 3, tous les acides aminés hors glycine en position deux semblent privilégier le positionnement dans la poche trois à la fois de leur chaîne latérale et de l'extrémité α -

amino ; ceci empêche la réactivité et assure la K-myristoylation. De plus, il apparaît que cette poche trois est assez profonde pour accepter des résidus plus volumineux voire un enchainement de résidus. Ceci est conditionné à la présence d'un espaceur de deux acides aminés entre la lysine acceptrice et le motif S₆K₇. C'est ce que j'ai pu observer pour des peptides AcGGK et AcGGGK dont la myristoylation a pu être observée par MALDI-TOF. Sachant que la lysine peut être myristoylée en position plus avancée, comme en position quatre, une compétition entre la G-myristoylation et la K-myristoylation est possible sur des peptides de type GGK en respectant l'espaceur entre la lysine réactive et le motif SK. C'est ce que j'ai vérifié sur un peptide de séquence GGKFSSKPR, où une myristoylation a pu être observée en MS1. Cependant, la MS2 n'a pas pu permis de démontrer formellement la K-myristoylation ; en effet, l'intensité des ions b₁ ou y₇ caractéristiques n'est jamais suffisamment significative pour distinguer l'une ou autre des deux types de myristoylation. Ceci m'a conduit à mettre au point un protocole spécifique pour différencier les deux types de myristoylation.

Étant donné que la G-myristoylation semble systématiquement priorisée par rapport à la K-myristoylation, il est peu probable que la K-myristoylation se produise de façon co-traductionnelle. C'est donc de façon post-traductionnelle que cette modification a le plus de chance d'être réalisée. Pour cela, deux alternatives physiologiques sont possibles, soit elle se forme après action d'une protéase (comme une caspase, voir introduction) sur une protéine, libérant un nouveau N-terminal, soit par la protéase IpaJ qui retire la glycine myristoylée. Le facteur de virulence IpaJ de *Shigella flexneri* est une cystéine protéase qui reconnaît spécifiquement le motif (Myr)-Glycine (**Figure 3a supplémentaire**) de plusieurs protéines humaines *in vivo* (voir introduction et (Burnaevskiy *et al*, 2013). Cette protéase étant spécifique à la G-myristoylation, elle pourrait être utilisée pour différencier les sites de modification ambigus (**Figure 3a**). Grâce à cet outil que j'ai appris à maîtriser (**Figure 3b-d supplémentaire**), j'ai pu mettre en évidence la compétition qu'il peut se produire entre les deux types de myristoylation (G- et K-) sur un peptide GGKFSSKPR. En effet, au temps un (incubation du peptide avec la NMT, T1) on peut observer en MS une myristoylation sur la glycine alors qu'au

temps deux (après incubation du peptide avec la NMT et IpaJ, T2) on peut observer un pic ayant subi une perte de masse de 267Da (correspondant à une perte du groupement Myr-Gly) ainsi qu'un reste du peptide myristoylé (**Figure 3b,c,d,e**) dont la myristoylation se situe sur la lysine.

Ces expériences avec IpaJ ont révélé un nouveau site de myristoylation inattendu. En effet comme on peut le voir sur les spectres de la **Figure 3**, le peptide Myr-GKFSSKPR conduit après l'action d'IpaJ au peptide KFSSKPR, qui à son tour peut être de nouveau myristoylé. Bien qu'il soit impossible de déterminer la localisation de la myristoylation (que ce soit sur l'amine N-terminale ou sur l'amine de la chaîne latérale de la lysine) les modélisations montrent que c'est la myristoylation en lysine qui est la plus probable. Pour le démontrer, j'ai mesuré les constantes catalytiques des peptides KFSSKPR et ac-KFSSKPR, et elles sont similaires (**Figure 2a, Tableau supplémentaire 1c**). J'ai ensuite essayé de re-myristoyler les quinze peptides correspondant au N-terminaux de protéine commençant par G2K3 et j'ai pu observer une re-myristoylation systématique de ce peptides (**Ensemble de données supplémentaires 1**).

Au cours de cette étude, j'ai pu observer que les peptides commençant par ac[Orn], ac[Dap], A[Orn] et acG[Dab] ne subissent pas de myristoylation aux concentrations les plus élevées de NMT (**Figure 2a et jeu de données supplémentaires 1**), contrairement à leurs homologues proches, acK, AK et acGK. J'ai émis l'hypothèse que de tels composés pourraient agir comme des inhibiteurs de la NMT. Alors que le dérivé ac[Dap] s'avère un inhibiteur médiocre avec une constante d'inhibition de $11 \pm 1 \mu\text{M}$, le dérivé ac[Orn] a montré une affinité d'un ordre de grandeur plus élevé (**Figure 5a**), suggérant que le groupe amino réactif N-terminal - s'il est correctement aligné avec le carbone via 2-3 liaisons carbone comme dans Orn ou Dab - peut contribuer à la liaison sans favoriser la catalyse. Cela m'a encouragé à poursuivre la conception et l'étude de composés présentant des variations des dérivés Orn et Dab N-terminaux basés sur la séquence GK. J'ai modifié l'orientation de l'Orn en utilisant la D-Orn - le stéréoisomère énantiomère de la L-Orn. (**Figure 5b**; branchement symétrique de R₁ et R₂ par rapport à C α 3) soit avec ac ou acG pour remplir la poche numéro trois. Les deux dérivés D-Orn

étaient de puissants inhibiteurs, la version ac étant la plus puissante ($K_D = 50$ nM). Lorsque ac[D-Orn] a été installé sur le châssis NCS1 - c'est-à-dire sans position hydrophobe à la position 5 - une myristoylation significative, bien que faible, a été observée (tableau complémentaire 1c). Cela indique qu'une inhibition puissante n'est possible qu'avec un ajustement optimal pour la K-myristoylation, y compris une poche numéro 5 intégrant un résidu fortement hydrophobe comme Leu ou Phe. Enfin, j'ai vérifié par spectrométrie de masse si les inhibiteurs étaient myristoylés après incubation avec la NMT, et j'ai pu observer la présence en faible quantité du peptide myristoylé, indiquant que ces inhibiteurs étaient des inhibiteurs suicides, restant bloqués dans le site actif.

Proteome-wide probing of the dual NMT-dependent myristoylation tradeoff unveils potent, mechanism-based suicide inhibitors

Frédéric Rivière^{iD,1}, Cyril Dian^{iD,1}, Rémi F. Dutheil^{iD,1}, Carmela Giglione^{iD,@,1,*}, Thierry Meinel^{iD,@,1,*}

¹ Université Paris Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC),
91198 Gif-sur-Yvette cedex, France

Running title: NMT-mediated lysine myristoylation rules and mechanism

***Correspondence:** carmela.giglione@i2bc.paris-saclay.fr (C. Giglione) or
thierry.meinel@i2bc.paris-saclay.fr (T. Meinel)

^{iD} ORCID 0000-0001-6366-1904 (F. Rivière); ORCID iD 0002-6349-3901 (C. Dian); 0002-
2910-6648 (R.F. Dutheil), 0002-7475-1558 (C. Giglione); 0001-5642-8637 (T. Meinel)
@twitter @giglionelab (C. Giglione); @meinel (T. Meinel)

Keywords: Acylation; lysine; myristoylation; *N*-myristoyltransferase; N-terminal
modification

ABSTRACT

N-myristoyltransferases (NMTs) catalyze protein myristoylation, a major and ubiquitous lipid modification. Originally thought to modify only N-terminal glycine α -amino groups (G-myristoylation), NMTs are now known to modify lysine ϵ -amino groups (K-myristoylation), the significance of which is uncertain. Here we exploited systematic structural proteomics analyses and a novel pipeline involving the *Shigella* IpaJ protease to discriminate K- and G-myristoylation with unprecedented accuracy and identify the specific features driving each modification. NMT-dependent K-myristoylation occurs post-translationally and only on lysines 1, 2, or 3 following G-myristoylation or caspase cleavage. Direct interactions between the substrate's reactive amino group and the NMT catalytic base slow K-myristoylation catalysis. IpaJ unmasked novel K-myristoylation sites in a dozen human proteins. The unique properties of NMT-driven K-myristoylation allowed us to design potent, mechanism-based suicide NMT inhibitors. These analyses unravel the respective paths towards K-myristoylation, G-myristoylation, or NMT inhibition, which rely on a very subtle tradeoff embracing the chemical landscape around the reactive group.

SIGNIFICANCE STATEMENT

We report the specific and unique elements guiding N-myristoyltransferase to either alpha or epsilon myristoylation, allowing us to establish the post-translational nature of N-myristoyltransferase-dependent lysine myristoylation and design novel, potent N-myristoyltransferase inhibitors.

INTRODUCTION

Myristoylation is an essential lipidation that adds a C:14:0 fatty acid to proteins in all eukaryotes including pathogens¹⁻⁴. The lipid moiety anchors soluble proteins to membranes, where they interact with partners to initiate signal transduction⁵⁻⁷. N-myristoyltransferases (NMTs; glycylpeptide N-tetradecanoyltransferases) are the only enzyme class known to catalyze myristoylation and, given their central role in pathobiology, are promising therapeutic targets⁸⁻¹². NMTs were long thought to exclusively target proteins with an N-terminal glycine (G-myristoylation) usually arising from co-translational methionine excision^{1,13} or, less frequently, post-translation cleavage exposing new N-terminal glycines^{14,15}. Proteome-wide approaches have delivered an exhaustive list of substrates undergoing co-translational G-myristoylation - the so-called G-myristoylome - covering approximately 2% of the human proteome (~600 proteoforms¹⁶). Additionally, high-resolution co-crystallography of human NMT1 with a number of reactive substrates displaying an N-terminal Gly¹⁷ has established a specific water channel in NMT that provides a water-mediated bond (Wat2) between the N-terminal group of the G-starting substrate and the C-terminal carboxy group (Q496) of NMT, which acts as the catalytic base for deprotonation and further reactivity^{1,17}. These data have been instrumental in revealing that NMTs further support myristoylation of lysine side chains (K-myristoylation), significantly expanding the range of known substrates^{17,18}. The mammalian small G-protein ARF6, which features a K3 following an N-terminal G2, is the only protein currently known to undergo both N-terminal G- and K-myristoylation¹⁸, allowing it to attach to plasma membranes when GTP is bound and permitting K-myristoylation by sirtuin deacylases upon GTP hydrolysis. In this cycle, NMT ensures the novel ARF6 K-myristoylation after GTP refueling.

NMT-dependent G-myristoylation occurs on the α -amino group of the target protein via an amide bond, creating an extremely stable attachment *in vivo* and resistance to chemical cleavage *in vitro*¹⁹⁻²¹. Accordingly, G-myristoylation was considered to be irreversible. However, the discovery of IpaJ, a pathogenic cysteine-dependent protease produced by several disease-causing bacteria including *Shigella flexneri*, challenged this four decade-old dogma²². IpaJ cleaves the N-terminal myristoylated G from several proteins. IpaJ relies on the presence of a G-myristoylated group embedded within a small dipeptide as a minimal chassis²³, with strong specificity for C:14 chains over longer (C:16) or shorter (C:10) fatty acyl chains²³. Since this study predated the discovery of NMT-driven K-myristoylation, known IpaJ substrates are limited *in vivo* to the small G-proteins of the ADP-ribosylation factor family (ARF and ARL) with the exception of ARF6^{23,24}. Critically, it is unknown whether IpaJ contributes to the NMT-dependent K-myristoylome. One major challenge to establishing this lies in the difficulty in identifying and proving the NMT-dependent K-myristoylome and the type of myristoylation taking place, the latter usually relying on mass spectrometry (MS). Unequivocal MS/MS (MS2) spectra of K-myristoylation profiles are not available and, because several acyl hydrolases targeting K-myristoylation have been described including SIRT2, SIRT6, HDAC8, and HDAC11, their action might prevent the identification of K-myristoylated substrates *in vivo*. New approaches for characterizing the K-myristoylome are urgently required.

To this end, here we combine mass spectrometry, kinetic studies, *in silico* analysis, and crystallography of human NMT1 with synthetic peptides to assess the molecular basis of NMT-dependent K-myristoylation. In doing so, we provide an overview of its coverage in the human proteome. We reveal that K-myristoylation results from post-translational events as illustrated by IpaJ-induced cleavage triggering K-myristoylation of a number of targets. Additionally, we show that G-myristoylation outcompetes K-myristoylation, since G- and K-myristoylation

share a similar recognition chassis at the NMT active site. Nevertheless, K-myristoylation features a number of unique recognition elements, based on which we design novel, selective, and potent mechanism-based suicide NMT inhibitors.

RESULTS

Co-translational G-myristoylation predominates over K-myristoylation on N-terminal GK-starting peptides

We previously reported unforeseen NMT-binding clefts accommodating peptide substrate side-chains (pockets 2-8 in **Fig. 1a**) and the N-terminal myristoylome based on the hypothesis that NMT-dependent myristoylation only occurs on glycines¹⁶. We and others later reported NMT-dependent K-myristoylation on position 3 lysines^{17,18}. An updated NMT reaction scheme with the two myristoylation types is displayed in **Supplementary Fig. 1**.

As a model peptide with a K3 (G₂KSFSKPR; reference, GK) was an efficient NMT substrate (**Fig. 1b** and **Supplementary Table 1a**), we assessed which type of myristoylation (G- and/or K) *Homo sapiens* (Hs)NMT1 catalyzed on this peptide by MALDI-Tof-Tof including MS-MS (MS2). Together with kinetic analysis, K3 favored G-myristoylation in this model context (**Supplementary Table 1a, Supplementary Dataset 1** and **Fig. 1b**). We next examined the myristoylation status of all human proteins starting with N-terminal GKs, including the six closely related members of the HPCA/NCS1 family and ARF6. In all cases, only G-myristoylation was unambiguously retrieved (**Fig. 1b, Supplementary Dataset 1** and **Supplementary Table 1a**), with efficiencies in the same range as reported for the G-myristoylome, as exemplified by the SOS3 variant (**Fig. 1b**). Remarkably, most N-terminal GK-starting proteins harbored the S₆K₇ motif, which is known to facilitate myristoylation²⁵.

Indeed, the salt bridges of K7 with the three contiguous Ds of the NMT Ab-loop were recently shown to promote the movement of the Ab-loop from an open to a closed conformation (SuppFig.7 in ¹⁷), favoring catalysis through distortion of the MyrCoA thioester bond in the closed conformation and facilitating transition to the tetrahedral state.

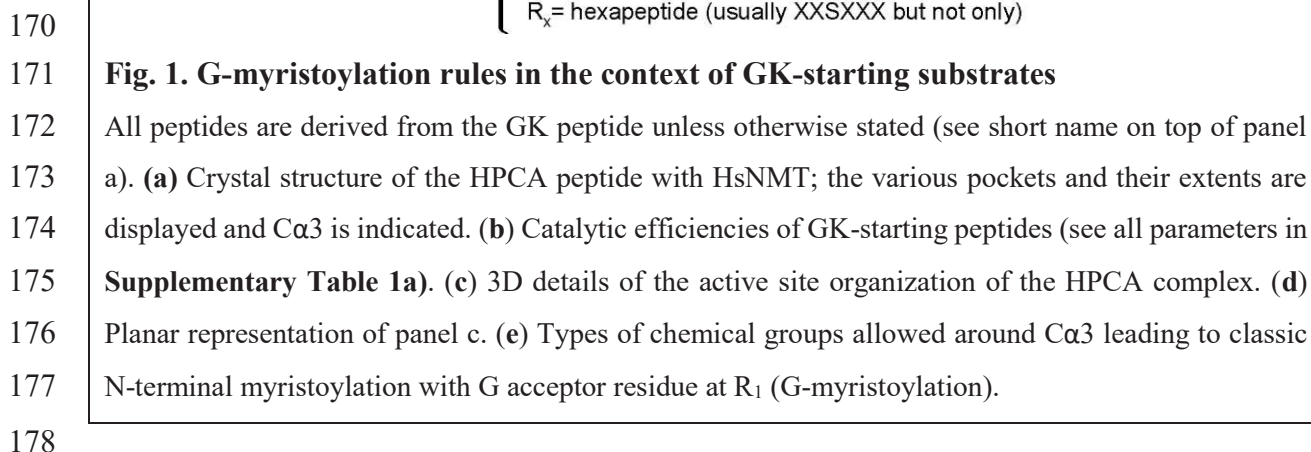
A new crystal structure of HsNMT1 with a peptide corresponding to the N-terminus of HPCA revealed that the K3 side chain buries tightly in pocket 3 of NMT and that its positive charge bonds the catalytic base (**Fig. 1c**). Full occupation of pocket 3 with K3 slightly displaces the reactive alpha amino group of G2 from the MyrCoA thioester group whilst retaining the Wat2 interaction of the reactive N-terminal α -amino group required for G-myristoylation¹⁷ (**Fig. 1c,d**). This K3-induced displacement of G2 confirms that, at position 3, a K residue is less favorable to myristoylation than any amino acid with a smaller, uncharged side chain.

To examine whether GK-starting proteins have the same recognition rules as those in the G[[^]K] context (where [^]X means any amino acid other than X)¹⁶, we analyzed HsNMT1's capacity to acylate a series of changes in the N-terminal peptide derived from NCS1. Only G-myristoylation occurred when K7 was changed into G or S6 into A, although HsNMT1's catalytic efficiency significantly decreased (**Supplementary Table 1b** and **Supplementary Dataset 1**). Therefore, the S6K7 motif is equally important for GK- and G[[^]K]-starting proteins, i.e., the role of S6 in substrate binding and K7 in the MyrCoA binding-induced conformational switch of the conserved flexible acidic Ab-loop of NMT is equal to that already reported and starting with G[[^]K]¹⁷. We next assessed whether modifying the side-chain length but retaining the basic character at position 7 impacted the myristoylation type. Shortening the basic side chain to ornithine (Orn, reduced by one CH₂) or di-amino butyrate (Dab, reduced by two CH₂s) had little impact on G-myristoylation (**Supplementary Table 1b**), suggesting

maintenance of the salt bridge with the Ab-loop. Further shortening to di-amino propionate (Dap, short basic chain CH_2NH_2) or removal of the basic character with homocysteine (Hcy) significantly reduced catalytic efficiency but still promoted only G-myristoylation (**Supplementary Table 1b** and **Supplementary Dataset 1**).

To further investigate the rules governing myristoylation by NMT, we substituted the G2 with an A in the GK-context variant devoid of K3 ($\text{A}_2\text{NCFSKPR}$; AN). There was unambiguous α -myristoylation by both MS and crystallographic analyses (**Extended Data 1a, Supplementary Dataset 1** and **Supplementary Table 1a**). Catalytic data indicated that the substrate was very inefficiently modified by NMT (**Fig. 1b**), and the crystal structure of AN confirmed that the active site was filled with a methyl addition, improperly orientating the reactive amino group. To decipher whether an A resulted in significant myristoylation in another sequence context, we substituted G2 with A in $\text{G}_2\text{SSVSKKK}$, which was not myristoylated (**Supplementary Dataset 1**), clearly indicating that the S_6K_7 motif is insufficient for reactivity in the A_2 context. The same was also true for an AGK variant (**Fig. 1b, Supplementary Table 1a**). In all cases, the G2A substitution dramatically reduced myristoylation but eventually led to G-myristoylation.

We therefore wondered whether the strong hindrance to the active site promoted by the additional methyl chain of A was similar if this group was grafted directly on the N-terminus. The N-methyl-GN peptide (meGN, meG is an isomer of A) led to very poor efficiency G-myristoylation (**Fig. 1b, Supplementary Table 1a**). Furthermore, the crystal structure of HsNMT1 in complex with meGN revealed that the methyl rotated the side-chain of N3 around $\text{C}\alpha_3$, which was orientated towards the catalytic base, whereas the methylated N-terminus was



Together, these data indicate that GK-starting peptides lead to G-myristoylation, which relies on G2 and the S6K7 motif. In the less favorable K3 context, S₆K₇ becomes even more crucial for G-myristoylation, consistent with previous reports suggesting strong dependence of G-myristoylation on K3 and K7^{18,26}. However, considering that the N-termini of T106A/B (GKSLSHLP, GKTFSQLG) are unambiguously NMT substrates (**Fig. 1b** and **Supplementary Table 1a**) despite the absence of K7, the deduced motif for G-myristoylation with K3 is GKXXSX (**Fig. 1e**). We noticed that both T106A/B harbor a highly hydrophobic residue at position 5 (L and F, respectively), another feature favoring G-myristoylation^{16,25}, which probably compensates for the absence of K7 and leads to efficient G-myristoylation. In addition, both ARF6 and GK possess a hydrophobic residue at position 5 (**Fig. 1b**) so, consistent with its positive role, permutation of positions 4 and 5 in the GK background (GKFS) significantly reduced myristoylation efficiency (**Fig. 1b**).

K3 myristoylation of N-acetylated peptides with the S6K motif

In apparent contradiction to the above results, the GK-starting protein ARF6 was recently shown to accept K-myristoylation *in vivo*¹⁸, but only if the G2 α -amino group was unreactive due to previous N-acetylation (*ac*) or G-myristoylation. We therefore examined any modification around the reactive G2 amino group allowing K-myristoylation of either the HPCA, ARF6, and/or GK-starting peptides. As expected, *ac* of the three G-starting peptides led to side-chain myristoylation, provided that K occupied position 3 (**Fig. 2a, Supplementary Table 1c** and **Supplementary Dataset 1**). When the K side chain was shortened to Orn, in the context of N-acetylated peptides, side chain myristoylation was still observed at position 3 (**Fig. 2a**). The catalytic efficiencies of the *ac* series were one order of magnitude lower than those of G-myristoylation (**Supplementary Table 1a,c**), with the catalytic constant (k_{cat})

accounting for most of the k_{cat}/K_m decrease in both cases. This suggested that the affinity provided by the peptide core was similar but that catalysis had a limiting step.

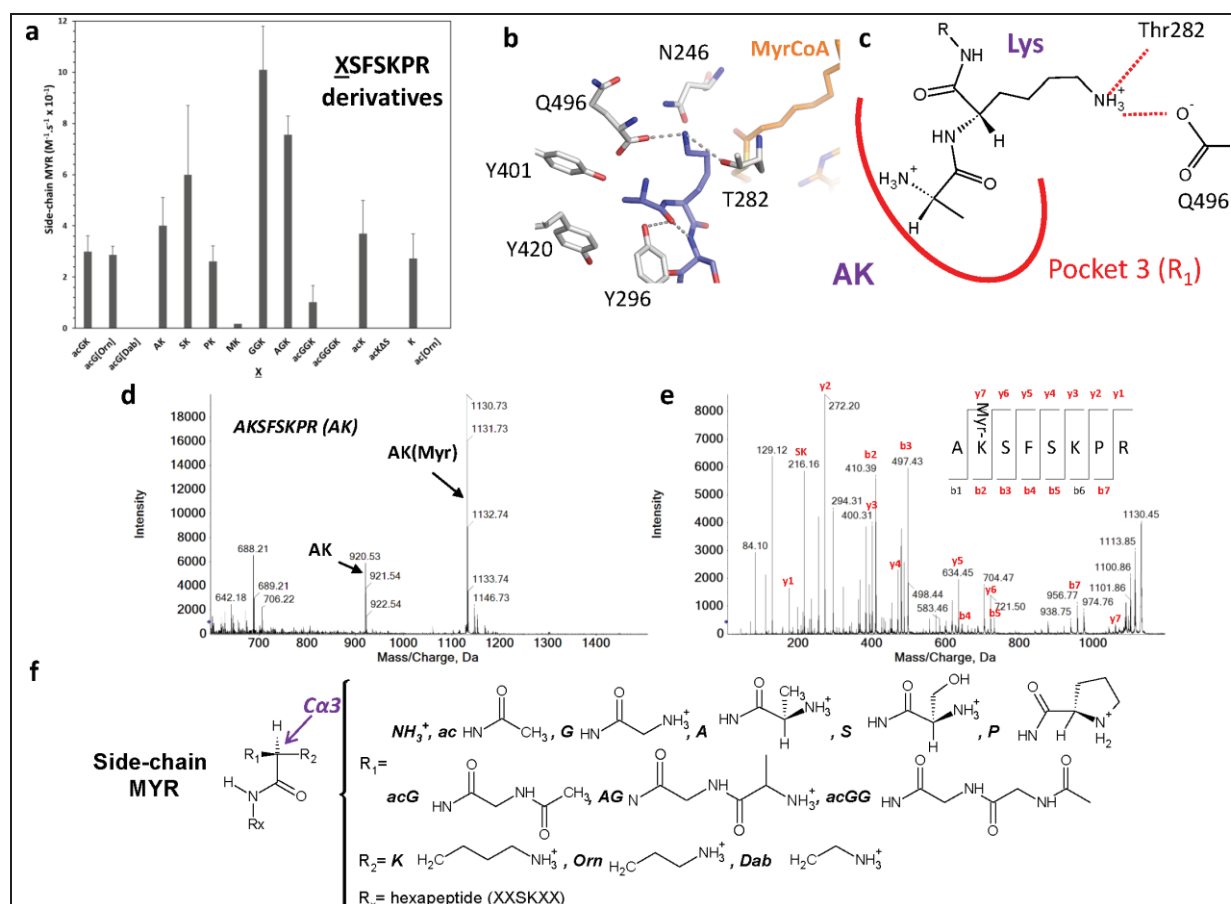


Fig. 2. K-myristoylation and G-myristoylation rules

The peptides are derived from the GK peptide. (a) Main K-myristoylation catalytic efficiencies measured (see details in **Supplementary Table 1a**). (b) Crystal structures of HsNMT1 in complex with MyrCoA and the AK variant. The panel displays a close-up of the structural environment around the catalytic base, the C-terminus of the C-terminal residue Q496. (c) corresponds to plane representations of panel b. (d) MS1 analysis of the AK derivative after incubation with HsNMT1. (e) MS2 analysis of the 1,131 Da myristoylated peptide in panel d showing K-myristoylation. (f) New knowledge on the type of acceptor amino group leading to K-myristoylation. The panel shows that if a lysine side chain or a non-natural basic side-chain with reduced length occurs at R₂, and depending on the substitutions made around the N-terminus, i.e., while blocking it (ac derivatives) or slightly modifying its plane orientation (e.g., if G is substituted for A, S, or P), then K-myristoylation (ϵ -myristoylation) is favored over N-terminal myristoylation (α -myristoylation)^{17,18}.

The crystal structures of HsNMT1 in complex with acG[Orn] or an acGN non-reactive derivative were obtained and compared with the acGK and acGN variants. In both cases, there was rotation around the C α of the aa3 side chain (Ca3), as expected from the *ac* addition. The acG group moved to pocket 3, and the aa3 side chain contacted the catalytic base (**Extended Data 2a-d**). The K side chain of acGK bonded directly with the C-terminal catalytic base of NMT, unlike the G-myristoylation mechanism, which involves a water-mediated bond (**Fig. 1c,d**). acG[Orn] led to a weaker salt bridge with the catalytic base than acGK, as the distance was increased from 3.1 to 3.6 Å (**Extended Data 2e**). Nevertheless, the associated K-myristoylation mechanism still involved direct interaction between the catalytic base and the ammonium group of Orn, like with acGK (**Extended Data 2c-f**). Further reduction to Dab dramatically decreased K-myristoylation (**Fig. 2a, Supplementary Dataset 1 and Supplementary Table 1c**). Therefore, tight direct binding of the K side chain reactive amino group to the catalytic base most likely explains effective but slower K-myristoylation compared to G-myristoylation.

Side chain myristoylation at K3 is preferred over N-terminal α -myristoylation when the N-terminal residue is not G

Therefore, Ca3 tends to rotate and embed the N-terminal group into pocket 3 provided any change in G2 (**Fig. 1e**) and, if aa3 displays an amino group (e.g., K3), it becomes reactive and promotes K-myristoylation. This suggests that the strong negative effect of the G2A substitution might favor K-myristoylation in the presence of K3. To test this hypothesis, we examined G2A HPCA (A-HPCA), ARF6 (A-ARF6), and GK variants (AK), all featuring a K3. In all cases, crystallographic and/or MS analysis unambiguously showed K-myristoylation (**Fig. 2b-e, Extended Data 3 and Supplementary Table 1c**). The catalytic efficiencies with

AK and A-ARF6 were (i) significantly greater than with A-HPCA and (ii) of the same order of magnitude as the acetylated versions of the corresponding variants. Therefore, the K side chain was preferred over the α -amino provided that a side chain (e.g., methyl of A) was grafted onto the reactive residue instead of G2. The crystal structure of AK complexed with HsNMT1 (**Fig. 2b,c**) showed that the K3 side chain bonded to the C-terminal catalytic base (Q496) and to T282 and the A2 side chain buried in pocket 3. Again, direct bonding to the catalytic base likely explains the reduced catalysis compared to the G-myristoylation water-mediated bonding of the GK or HPCA peptide. The lower K_m of the acG derivatives may be related to additional interactions of the entire A2 chain in pocket 3 (**Fig. 2b**).

We next interrogated the human proteome to discover possible K-myristoylation targets featuring AKXXSK immediately following an M start residue (M1). Only tuberin (TCS2; P49815; **Supplementary Table 2a**) was identified, but no myristoylation was observed with the corresponding peptide derivative (AKPTSKDS; **Supplementary Table 1c**), perhaps due to the P4 inducing a local conformational kink disfavoring both G- and K-myristoylation by misaligning S6K7 with the reactive amino upstream group. The absence of a hydrophobic side chain at residue 5 (similar to A-HPCA vs A-ARF6 or AK) might also explain this phenomenon, as noted above for G-myristoylation.

We next examined whether non-G N-termini other than A might also lead to K-myristoylation by replacing the N-terminal G2 of the GK peptide with M, S, and P. K-myristoylation occurred in all cases with catalytic efficiencies inversely proportional to side chain length (**Fig. 2a**, **Supplementary Table 1c** and **Supplementary Dataset 1**). We concluded that the nature of the N-terminal residue was a major determinant guiding N-terminal K-myristoylation, provided that aa3 harbored a proximal acceptor group (**Fig. 2f**). Due to improper positioning

of the N-terminal amino group, any non-G-starting peptides with the S6K7 motif may be modified by NMT but with extremely poor catalytic efficiencies several orders of magnitude lower than the corresponding G derivative. A K3 downstream of any amino acid at position 2 but G permits K-myristoylation with dramatically increases associated catalytic efficiencies (**Supplementary Table 1a**). Higher efficiencies are possible when the peptide context is favorable, as with AK and A-ARF6 peptides (this work and ¹⁸). Remarkably, both peptides feature a bulky, hydrophobic, favorable residue at position 5 (F or L, respectively), already shown to be important for G-myristoylation (see ¹⁶). Finally, interrogation of data libraries indicated that no protein with A2 (i.e., originating from M1 excision) undergoes K-myristoylation in the human proteome.

Taken together, our data reveal that K-myristoylation does not require a given side chain of the first amino acid but relies on optimal residues at position 5-7. The [^G]KX[FL]SK motif emerges as the determinant of K-myristoylation (where ^X means any residue but X). Any non-G2 side chains appear to privilege pocket 3 binding, preventing their reactivity and ensuring high specificity for G2 undergoing G-myristoylation.

In the human proteome, 7548 hits harbored a KXXSK motif (**Supplementary Table 2a**) corresponding to 2831 unique entries with 3615 sites in 6764 proteoforms (14-15% of the proteome). 572 sequences displayed F or L at position 3, like in ARF6 or the GK peptide. Among the 29 sequences with translation start sites at positions 1 or 2 (i.e. arising from cotranslational N-Met removal rules), 10 of the 15 known G-myristoylated N-termini starting with GK were found ¹⁶ plus a few additional sequences, most not comprising F or L3, like TSC2. The other 3595 sequences corresponded to internal sequences starting beyond codon 3, and 285 were preceded by a G suggesting large scope for post-translational G-myristoylation

in this subset. Post-translational myristoylation, however, requires prior unmasking by proteolytic cleavage provided before the first XK. Indeed, a number of K-myristoylation sites among the 543 [^G]KX[FL]SK internal sequences were preceded by a classic caspase motif (**Supplementary Table 2b-e**). Therefore, unlike G-myristoylation, K-myristoylation should solely arise from post-translational myristoylation.

Myristoylation may occur on Lys4

We next wondered whether extending the amino acid sequence upstream of the reactive K might lead to K-myristoylation, provided that a two-amino acid spacer between the accepting K and the S6K7 motif was retained. We therefore grafted an extra N-terminal A or G to the GK peptide (A- or G-GKSFSKP; AGK or GGK) and checked whether HsNMT1 could acylate the K at position 4. By MS and kinetic analysis, K-myristoylation was as efficient with these A/G-grafted peptides as with the acG peptides (**Fig. 2a; Supplementary Table 1c and Extended Data 3**). We concluded that the pocket usually hosting residue 3 could host a dipeptide moiety as large as AG provided that a close and available K-reactive side chain was properly spaced from the crucial S6K7 motif. We noted that, with five bonds and a positive charge, a K side chain was the most bulky moiety tolerated in NMT pocket 3 for G-myristoylation. With seven bonds, the AG backbone is much bulkier than K but still promoted K-myristoylation (**Fig. 2f**). To establish whether larger chains would make K-myristoylation possible, we examined large, poorly-branched N-terminal grafts larger than GG or AG (AcGGK, AcGGGK, AcGGGGK, and AcGGGGGK with 9, 12, 15, and 18 bonds, respectively) upstream of K3 in the GK series. With AcGGK and AcGGGK, there was unequivocal K-myristoylation by MALDI analysis (**Supplementary Dataset 1**). K-myristoylation catalytic efficiency progressively decreased with increasing chain length to very

low values, and K-myristoylation was not detected for AcGGGGK or AcGGGGGK (Supplementary Table 1c). The crystal structure obtained with the GGK derivative showed that the GG chain is fully buried in pocket 3 and propels the main backbone peptide chain into a more remote location at Ca3 (Supplementary Fig. 2a,b); there is also a greater distance between the reactive K amino group and both the catalytic base and T282. The 7-bond GG or AG chain corresponds to the bulkiest chains accepted by pocket 3 for efficient K-myristoylation (Fig. 2f). When longer than 4 residues, as in the case of K, the backbone chain tends to not be perfectly aligned, but this is not responsible for the lower catalytic rates observed.

We concluded that to allow K-myristoylation, the [^G]KX[FL]SK motif may be associated with a one amino acid extension or *ac* at the N-terminus, provided that an *ad hoc* proteolytic cellular machinery can produce such extremities. In addition, about two dozen internal XXXXXSK K-myristoylation motifs were noted in the human proteome displaying an upstream classic caspase cleavage site (Supplementary Table 2a,b,c).

With the knowledge that K4 may be reactive in the context of a G₂XXK₄ derivative, we next assessed whether G₂- and K₄-myristoylation could compete within the same protein. For this, we first altered the G-myristoylation S₆K₇ motif of the GK peptide by introducing an additional S between S₆ and K₇, pushing the K back to position 8 (G₂G₃K₄F₅S₆S₇K₈PR). The resulting peptide therefore featured two overlapping myristoylation motifs separated by a two amino acid spacer, one for modification of the α-group of G₂, relying on motif S₆S₇, and a second on the ε-amino of K₄, relying on motif S₇K₈. This modification reduced the catalytic efficiency compared to the GK peptide (Supplementary Table 1a). Furthermore, the MS2 spectra of the

1,173 Da myristoylated product favored G-myristoylation over K-myristoylation (**Extended Data 4a**), with evidence of the major G-myristoylation product as unambiguous characteristic b_2 , y_7 , and y_8 ions (**Extended Data 4b**). Nevertheless, we could not exclude the possibility of K-myristoylation, as the associated prototypic K-myristoylation ions could only be identified at low intensity. Therefore, other experimental approaches were required to assess for simultaneous K-myristoylation.

The IpaJ protease as a tool to unravel ϵ -myristoylation

Whatever the context, G-myristoylation is favored over K-myristoylation when in competition. Since only an N-terminal G appears to promote myristoylation, K-myristoylation is unlikely to originate from a co-translational event. However, because K-myristoylation might be further added through post-translational insertion, identifying the conditions favoring this event would be useful. To date, two physiological proteolytic states have been shown to impact myristoylation: caspase and IpaJ proteolytic cleavage, and indeed several internal K-myristoylation sites may arise from caspase cleavage.

The virulence factor IpaJ from *Shigella flexneri* is a cysteine protease that specifically recognizes the myristoyl (Myr)-G moiety of several human proteins *in vivo*²⁴. As a result of its activity on Myr-G-peptides (**Supplementary Fig. 3a**), IpaJ is predicted to induce a 267 Da shift (Myr-Gly product) in MALDI MS spectra. Since it cleaves off G-myristoylation sites, we anticipated that IpaJ - provided it specifically processes any G-myristoylated peptide – could be used to unravel unambiguous and undiscovered K-myristoylation modification sites.

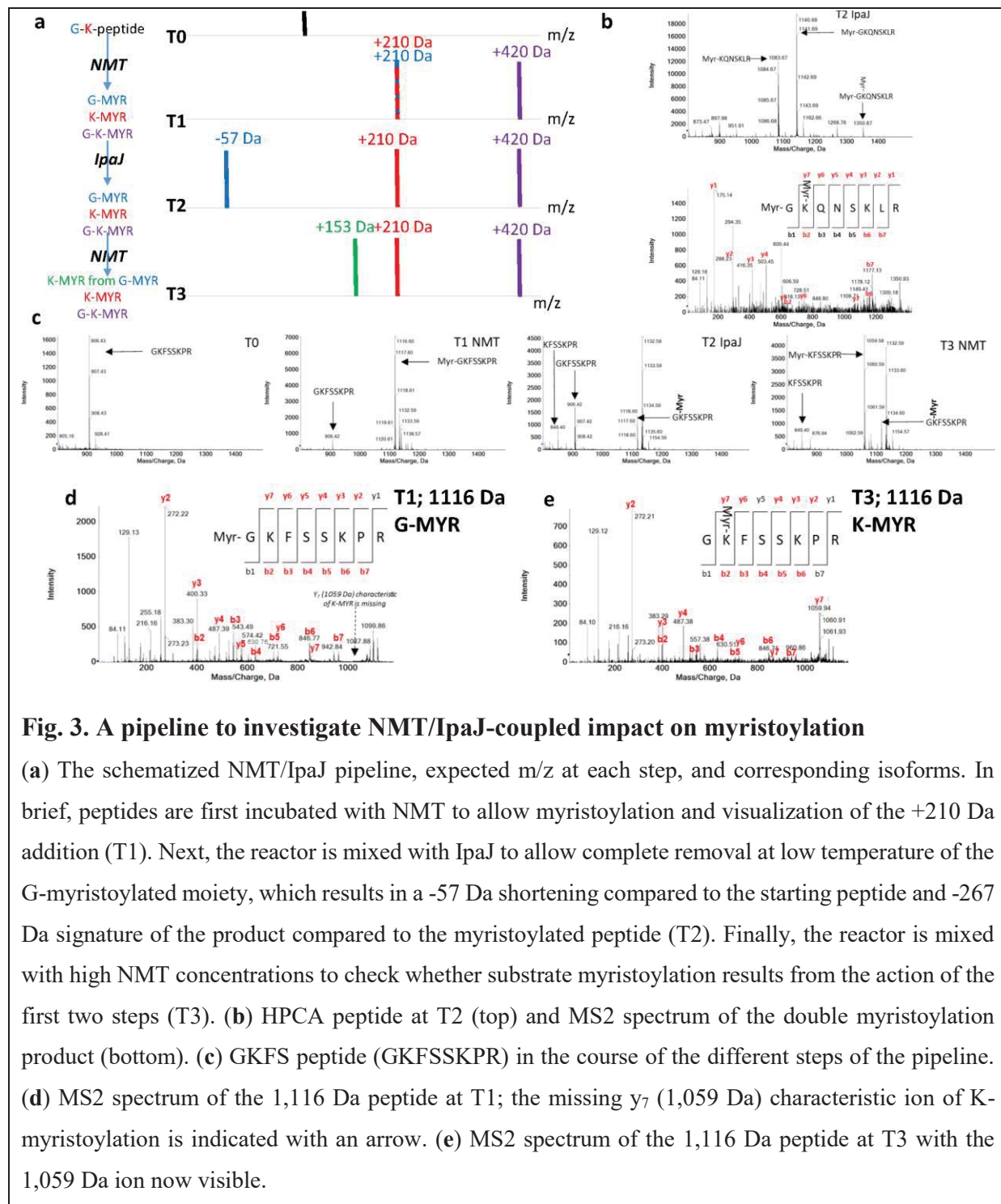


Fig. 3. A pipeline to investigate NMT/IpaJ-coupled impact on myristoylation

(a) The schematized NMT/IpaJ pipeline, expected m/z at each step, and corresponding isoforms. In brief, peptides are first incubated with NMT to allow myristoylation and visualization of the +210 Da addition (T1). Next, the reactor is mixed with IpaJ to allow complete removal at low temperature of the G-myristoylated moiety, which results in a -57 Da shortening compared to the starting peptide and -267 Da signature of the product compared to the myristoylated peptide (T2). Finally, the reactor is mixed with high NMT concentrations to check whether substrate myristoylation results from the action of the first two steps (T3). (b) HPLC peptide at T2 (top) and MS2 spectrum of the double myristoylation product (bottom). (c) GKFS peptide (GKFSSKPR) in the course of the different steps of the pipeline. (d) MS2 spectrum of the 1,116 Da peptide at T1; the missing y₇ (1,059 Da) characteristic ion of K-myristoylation is indicated with an arrow. (e) MS2 spectrum of the 1,116 Da peptide at T3 with the 1,059 Da ion now visible.

The *S. flexneri* *IpaJ* gene was cloned and the protein overexpressed and purified to homogeneity (Supplementary Fig. 3b,c). An inactive variant (IpaJ-C64S) was also produced as a control. Under specific conditions (Supplementary Fig. 3d), IpaJ was sufficient to cleave

off the G-myristoylated moiety from any tested peptide *in vitro*, unlike IpaJ-C64S (Supplementary Fig. 4a,b,c). Additionally, IpaJ was inactive on an N-acetylated G-peptide, an A2-myristoylated peptide, and a K3-myristoylated peptide derived from ARF6 (Supplementary Fig. 4d and Supplementary Table 1d,e). Due to its unique and high specificity for G-myristoylation and the observed complete cleavage of any G-myristoylated peptide, we concluded that IpaJ could be used as powerful tool to distinguish G-myristoylation from K-myristoylation and, additionally, investigate IpaJ-dependent post-translational K-myristoylation.

We therefore established an IpaJ pipeline that determined the myristoylation type of any peptide subjected to NMT (Fig. 3a). The workflow faithfully reproduced the sequence of events occurring in the context of cellular IpaJ activity. As proof-of-concept, a dozen peptides were carefully chosen for their sequence diversity and ability to undergo G-myristoylation (Supplementary Fig. 5a,b; see ARF6 in Supplementary Fig. 6). Combined with previous data²³, IpaJ efficiently cleaved any G-myristoylated peptide regardless of the sequence downstream of the G residue (Supplementary Fig. 5c).

Taking the double myristoylation of ARF6 at both the N-terminus of G2 and the side chain of K3 into account¹⁸, we used the IpaJ pipeline to identify double myristoylation of other GK-starting NMT substrates. We identified dual myristoylation of a peptide derived from HPCA/HPCL1/NCALD in the presence of IpaJ (Fig. 3b; Supplementary Fig. 7). Finally, double myristoylation was not observed with other closely related substrates of HPCA in the same NCS1 family of calcium sensors sharing the GK motif (Supplementary Dataset 1). We cannot exclude that the extremely high hydrophobicity of such acylated products prevented their crystallization in the matrix and consequently their identification.

Competition between G- and K-myristoylation

With the IpaJ/NMT pipeline in hand, we next examined the ambiguous myristoylation state of the GGKFSSKPR peptide. At T1, MS2 analysis unambiguously revealed G-myristoylation as a product (**Extended Data 4b**). At T2 (i.e., after incubation with NMT and IpaJ), both an uncleaved myristoylated peptide and a shorter, 267 Da peptide resulting from G-myristoylation cleavage were identified (T2, **Extended Data 4c**). There was also unambiguous and unique K-myristoylation at K4 of the GGKFSSKPR peptide at T3 (**Extended Data 4c,d**).

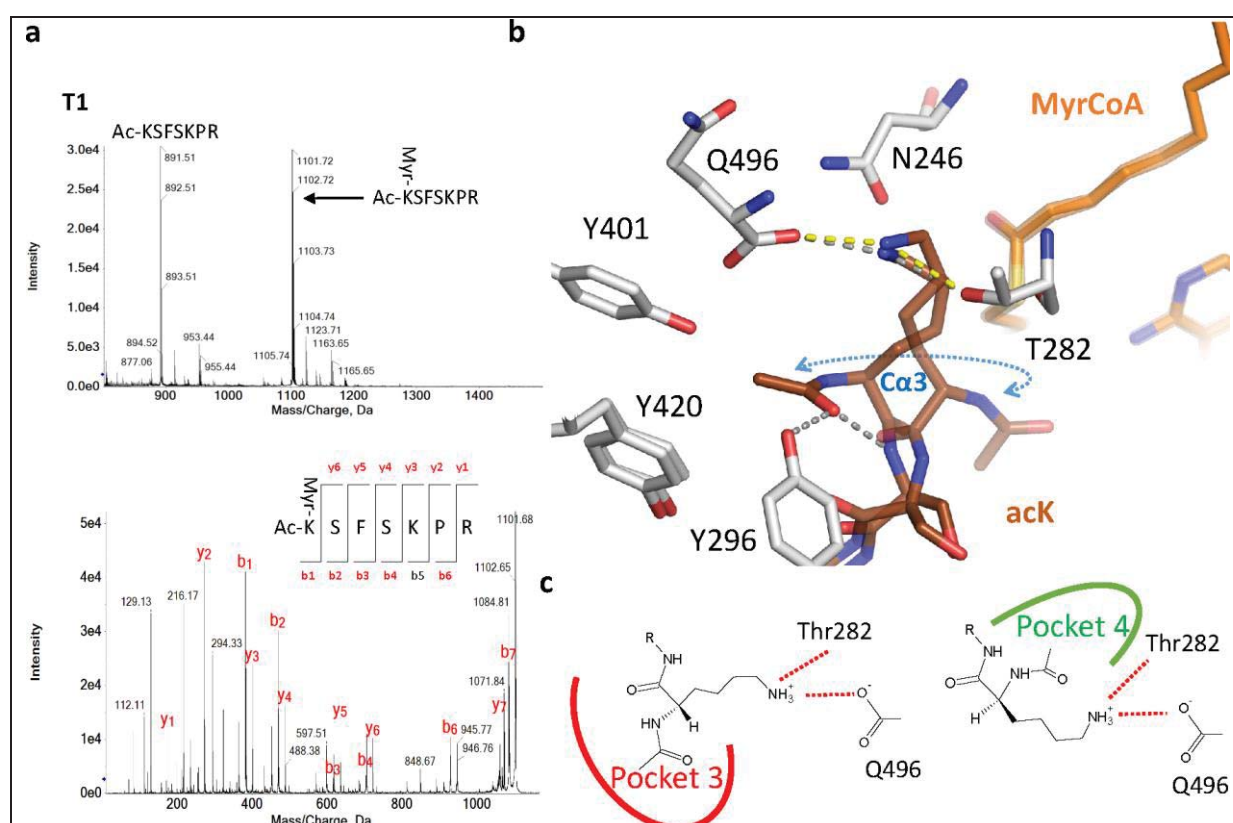


Fig. 4. K-myristoylation mechanism on an N-terminal side chain

The crystal structure of the complex made between acK and HsNMT1 is displayed. (a) MS analysis showing myristoylation of the acK; top MS1, bottom MS2. (b) The two conformations are displayed, and the rotation point around Ca3 is indicated. (c) Planar representation showing how each conformation may position the ac chain within pocket 3 (conformation A, left hand side) or pocket 4 (conformation B, right hand side).

Interestingly, the IpaJ product of G-myristoylation of this peptide started with GK to form GKFSSKPR (GKFS), which displayed reduced myristoylation efficiency (**Fig. 1b**). This peptide could be further myristoylated at T3 to produce a 1,116 Da form. MS2 analysis revealed K-myristoylation at the N-terminal K3 of this peptide (**Extended Data 4d**). To obtain a simpler set of products for the pipeline, we further characterized the myristoylation profile of GKFS in the IpaJ workflow (**Fig. 3c,d,e**). This peptide underwent both G- and K-myristoylation. GKFS features a bulky residue at position 4 and a small residue at position 5. We hypothesized that while a spacer is mandatory for both myristoylation types, either myristoylation might be favored depending on as yet unknown amino acid requirements occurring in that spacer (i.e., aa4-5, the two residues switched in the GK series). This is likely, as both pockets 4 and 5 are large enough to accommodate all side chains¹⁶, so some combinations could favor K-myristoylation in a GK-starting context.

We found no straightforward natural chassis in human sequences starting with a G2 that leads to competition between K- and G-myristoylation and a systematic preference for G-myristoylation in sequences originating from co-translational myristoylation. When we applied the IpaJ pipeline to GK-starting peptides derived from the human myristoylome and exploited its enhanced sensitivity to detect K-myristoylation, we observed that most displayed the myristoylated peptide at T2 after IpaJ cleavage (**Supplementary Table 1e**). Therefore, competition exists between G- and K-myristoylation on GK-starting peptides. Furthermore, none of these peptides displayed a hydrophobic residue at position 4, explaining why the K-myristoylation/G-myristoylation ratio was smaller than with the GKFS peptide and why prototypic K-myristoylation ions were not detected in MS2 spectra. Nevertheless, as myristoylation can also arise from post-translational addition provided that a cleavage generates a new K-accepting site, competition favoring K-myristoylation is likely. Indeed, as

noted above, hundreds of internal proteins display an F/L4 profile (**Supplementary Table 2a**; see columns aa3, F3, and L3 corresponding to this position, as all sequences start at the cleavage site and not the unprocessed M1).

In conclusion, a G at the N-terminus of peptides with GKFXSK motifs sustains competition between K- and G-myristoylation. These data also indicate that the KXXSK motif – not only [G]KX[FL]SK – is indicative of K-myristoylation independent of the residue preceding the K.

IpaJ reveals that K-myristoylation may also occur on free or acetylated K1

When the IpaJ/NMT pipeline was applied to either the myristoylated GGKFSSKPR or the GKFSKPR peptides (see above and **Fig. 4a**), IpaJ-induced cleavage of this reaction product also led to myristoylation of an unexpected K-starting peptide resulting from reiterated G-myristoylation IpaJ cleavage of each new G-myristoylation product. Although we could not distinguish alpha and epsilon myristoylation, side-chain modification was expected (K-myristoylation). Taking into account the S₆K₇ clamping of the peptide at the peptide-binding site, 3D molecular modeling indicated that the N-terminal free amino group would be too distant from the catalytic base for myristoylation. To test the hypothesis of K-myristoylation, we first prepared an acetylated (i.e., N-α-blocked derivative) of the GK peptide starting with K3 (acK). In the presence of MyrCoA and NMT, this peptide was myristoylated with catalytic efficiencies similar to the previously tested peptides (**Fig. 2a, Supplementary Table 1c**). When the N-terminal acK was installed at the N-terminus of a peptide with the aa4-5 spacer shortened to only one residue, myristoylation was no longer observed (acKΔS, **Supplementary Table 1c**). Finally, we checked that such K-starting peptides derived from the GK (K) and

HPCA chassis (K-HPCA) underwent K-myristoylation (**Fig. 2a, Supplementary Table 1c**). The crystal structure of the complex between NMT and an acK derivative revealed two alternative structures, both of which supported a K-myristoylation mechanism involving direct interaction between the K side chain and the catalytic base (**Fig. 4b**). The group around C α 3 could be retrieved in two opposite directions: the first showed the *ac* group entering either pocket 3 mimicking an N side chain (**Fig. 4c**) while the overall active site positioning mimicked that observed with GK; and the second revealed the *ac* moiety occupying part of pocket 4, which is large enough to host two moieties (i.e., *ac* and S4 in acK) from the same peptide (**Fig. 4b,c, right**). We concluded that side chain myristoylation can also occur on an N-terminal K (K1) regardless of acetylation status and that might result from post-translational proteolytic cleavages such as those induced by IpaJ during the course of *Shigella* infection.

We next interrogated the 15 GK starting peptides derived from different human proteins using the IpaJ pipeline (**Supplementary Table 1e**). As noted above, because they mostly resulted in G-myristoylation, they are likely to generate K1 N-termini after IpaJ cleavage. **Supplementary Fig. 7** summarizes such behavior with HPCA. As expected, there was evidence of K1-myristoylation at T3, further indicating that all such peptides can undergo K-myristoylation after IpaJ cleavage (**Supplementary Dataset 1**). We conclude that all GK-starting proteins likely undergo post-translational K-myristoylation after IpaJ cleavage.

K-myristoylation-inspired design of potent NMT inhibitors

The data above use synthetic K mimics to assess the effects of the side-chain length, and in doing so we noticed that ac[Orn], ac[Dap], A[Orn], and acG[Dab] do not undergo myristoylation even at the highest NMT concentrations (**Fig. 2a and Supplementary Dataset**

1), unlike their close homologs, acK, AK, and acGK. We hypothesized that such compounds might act as NMT inhibitors. While the ac[Dap] derivative was a poor inhibitor with an inhibition constant of $11 \pm 1 \mu\text{M}$, the ac[Orn] derivative showed an order of magnitude higher affinity (**Fig. 5a**), suggesting that the N-terminal reactive amino group – if properly aligned with the α -carbon through 2-3 carbon bonds like in Orn or Dab – might contribute to binding while not promoting catalysis. This encouraged us to further design and study compounds with variations around N-terminal Orn and Dab derivatives based on the GK sequence.

We first assessed acG[Dab], [Dab], and A[Orn], all of which acted as potent competitive inhibitors with IC_{50} values of the order of the NMT concentration used in the assay, indicating a tight binding constant in the low micromolar range (i.e., $<0.4 \mu\text{M}$; **Fig. 5a**). Dab and Orn but not Lys clearly led to strong inhibition.

To modify the amino group orientation towards the catalytic base and to mimic the G-myristoylation productive mechanism, we built two derivatives with D-Orn - the enantiomeric stereoisomer of L-Orn (**Fig. 5b**; symmetric branching of R_1 and R_2 with respect to $C\alpha_3$) - either with *ac* or acG to optimize the filling of pocket 3. Interestingly, the ac[D-Orn] derivative was a remarkably close mimic of a GN3 variant (**Extended Data 5a**), while our previous work revealed that N is favored in pocket 3^{1,16}. Both D-Orn derivatives were potent inhibitors, with the *ac* version the most potent ($K_D=50 \text{ nM}$). A similar result was observed with ac[D-Dab] or A[D-Orn] derivatives (**Fig. 5a**), and acG[D-Orn], G[D-Orn], or [D-Orn] derivatives were also strong inhibitors but to a lesser extent than ac[D-Orn]. When ac[D-Orn] was installed on the NCS1 chassis – i.e., without a hydrophobic position at position 5 – significant, though weak, myristoylation was observed (**Supplementary Table 1c**). This indicated that potent inhibition

was only possible with an optimal fit for K-myristoylation including a hydrophobic residue filling pocket 5.

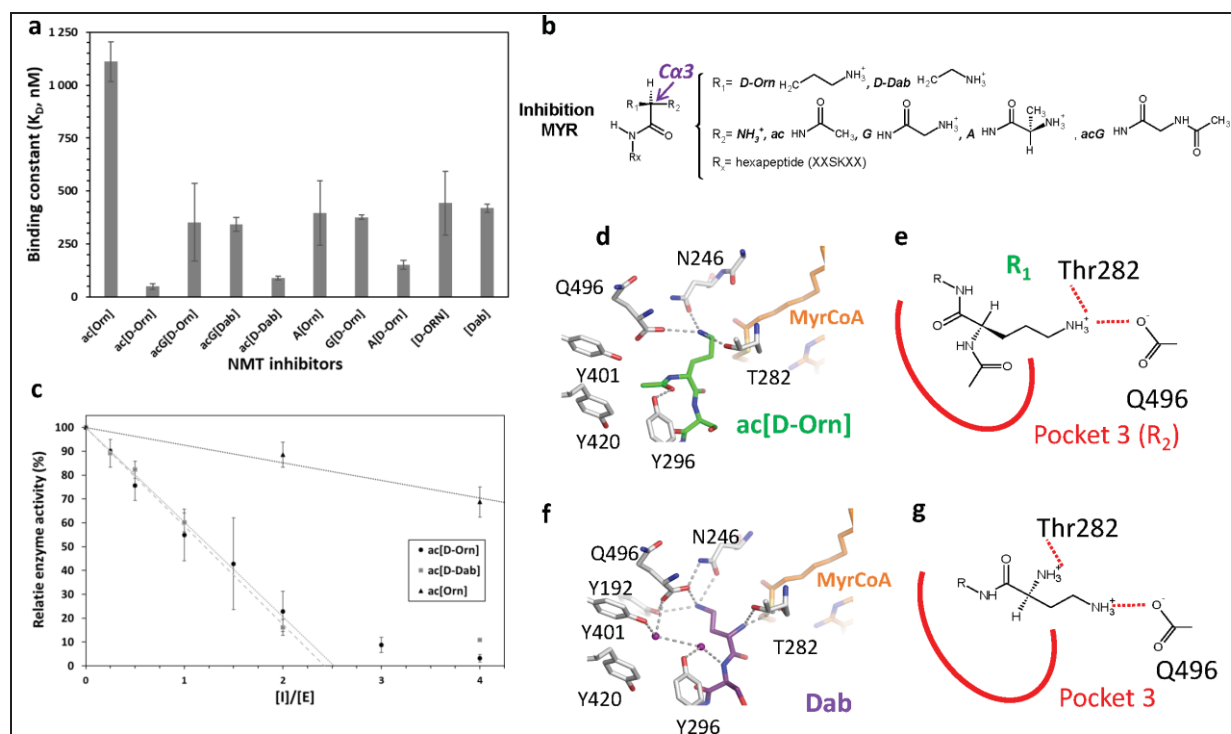


Fig. 5. Characterization of K-myristoylation-inspired suicide inhibitors of NMT

The peptides are derived from the GK peptide. **(a)** Apparent binding constants were derived from IC_{50} value measurements according to ⁴⁶. All reactants but the reactive peptide SOS3 were pre-incubated in the presence of the two peptides for 30 minutes. The reaction was triggered by the addition of SOS3 and followed over time for 60 minutes. Because of the bi:bi mechanism, MyrCoA is required to form the complex^{5,50}. **(b)** 3D and **(c)** plane representations of the interaction networks around the ac[D-Orn] derivative bound to HsNMT1. **(d)** Plot showing that the ac[D-Orn] (black circles) and ac[D-Dab] (grey squares) derivatives display low partition ratios. The ac[L-Orn] inhibition pattern is indicated (black triangles). The reaction was triggered by the addition of the productive substrate after pre-incubation of the inhibitor for 60 minutes in the presence of MyrCoA and HsNMT1. Time-course analysis was next achieved to determine the remaining enzyme activity. A value of 100 was given to the value obtained in the absence of inhibitor. 3D **(e)** and plane **(f)** representations of the interaction network around Dab bound to HsNMT1. **(g)** Representation in a model similar to that of **Fig. 1b and 2a** for G- or K-myristoylation, respectively substrates of the chemical requirement leading to NMT inhibition.

Prolonged incubation with NMT followed by MS analysis of the reaction products of the above inhibitors revealed product formation, though to a much lower level than the substrate. Furthermore, these peptides acted as suicide inhibitors (i.e., the unreleased product was the effective inhibitor). Indeed, pre-incubation of the inhibitors with the enzyme and MyrCoA rapidly inactivated the enzyme, indicating conversion of the inhibitors bound to the enzyme into a one- to two-order of magnitude more potent inhibitor. We concluded that these inhibitors behave as single-turnover substrates²⁷. There was also a partition ratio of 1.5 of either D-Dab or D-Orn derivatives of the *ac* series (**Fig. 5c**), indicating very high inactivation efficiency with an inhibition on-rate similar to that of the off-rate.

Finally, we obtained the crystal structure of ac[D-Orn] and [Dab] in complex with NMT and retrieved the complex before conversion of the substrate into a product (**Fig. 5d-g**). The complex featured direct binding of the D-Orn and Dab amino groups to the carboxy terminal base. The reactive amino group of the Dab side chain bonded directly with the carboxy-terminal base of NMT through two alternative orientations flipped around Cα3, one of which revealed a bond between the Nα ammonium and Thr282 (**Fig. 5f-g, Extended Data 5b,c**). Such direct bonding of the reactive amino group of both inhibitors is unlike the efficient GK peptide, which requires water-mediated binding, instead being more similar to the AK substrate derivative binding (**Fig. 3**). According to our data, this direct salt bridge created in the context of the Orn or Dab side chains, which leads to optimal sizing, is expected to be stronger than the one made with K and at least 100-fold less prone to reaction as a result. This likely explains the lesser reactivity and the inhibitory behavior. Both compounds bound to Thr282 (**Fig. 5d-g**). The non-acetylated Dab compound featured a water network within pocket 3 indicative of less bonding compared to both acetyl derivatives, perhaps explaining its lower binding constant.

In conclusion, (i) the chain length bringing the reactive amino group, (ii) the chiral conformation of the substituents R_1 and R_2 of the $C\alpha$ group facing pocket 3 ($C\alpha 3$; compare **Figs. 1e, 2f and 5b**), and (iii) optimal occupation of pocket 3 by a moderately bulky N-terminal moiety mimicking N were mandatory for effective K-myristoylation mechanism-based design of NMT inhibitors.

DISCUSSION

N-myristoyltransferases (NMT), also known as glycylpeptide N-tetradecanoyltransferases (EC 2.3.1.97; see BRENDA resource²⁸), are the only enzymes known to catalyze N-myristoylation in eukaryotes. It was long believed that NMT only co-translationally transferred a lipid (myristate) to the α -amino group of the N-terminal amino group of G-starting proteins mostly exposed by the prior action of methionine aminopeptidases (MetAP). Recent crystal structures of NMTs co-crystallized with G-starting substrates have revealed how the side chains of these substrates are recognized and lie within dedicated cavities (pockets 3-8; ^{1,16,17}), characteristics that contribute to the G-starting recognition pattern. Although this G-starting recognition pattern appeared to clearly define a global specificity, it did not explain the entire NMT-dependent myristoylation landscape. This was even more evident when we and others reported the possibility that NMTs catalyze K-myristoylation^{17,18}. Although this K-myristoylation function should expand the range of NMT substrates, the extent and clues ruling NMT-dependent K-myristoylation are unknown.

Here we disclose the unique and common specificities underlying NMT-dependent G- and K-myristoylation. We reveal that the main specificity rules for NMT-dependent G- and K-myristoylation partially overlap, leading to competition between the two types of

myristoylation in GK-starting contexts. The myristoylation rules are summarized in **Figs. 1e and 2f**. The very narrow shape of pocket 2 ensures reactivity of a primary amino borne only on unbranched moieties such as N-terminal glycine or side chains like lysine, which facilitates the 70° rotation of the amino-reactive group from the catalytic base towards the thio-reacting group of MyrCoA. This motion allows Wat3 to enter from the water channel and play a crucial role at the transition state¹⁷. The addition of a methyl group such as in A2 or meG variants makes this motion unlikely and dramatically decreases catalysis. The rules directing G- or side chain-myristoylation are predominantly determined by the R₁ and R₂ chemical groups branched at Cα3. NMT usually favors G- over K-myristoylation on proteins starting with G, while K-myristoylation dominates on any other N-terminal chain.

Overall, NMT-dependent K-myristoylation is less efficient than G-myristoylation whatever the context. Although the two myristoylation types may directly act on an N-terminal residue, K-myristoylation appears less strict than G-myristoylation in this regard; indeed, K-myristoylation does not require a given side chain of the first amino acid but relies on optimized residues at positions 5-7. A major difference between the two myristoylation types is that K-myristoylation needs the KXXSK-restricted pattern. This is unlike G-myristoylation, which allows the modification of many proteins not exhibiting the S₆K₇ motif by NMT on their N-terminal G (see ¹⁶ and **Supplementary Table 1a**). The KXXSK motif of K-myristoylation can be associated with a one amino acid extension or acetylation at the N-terminus provided that an *ad hoc* proteolytic machinery can produce these extremities. The specificity of MetAPs - which remove the first M from proteins - favors new N-terminal residues with short side chains like A, G, P, or S. A survey of the human proteome reveals, however, no K-myristoylation on either K2, K3, or K4 and 15 G-myristoylation substrates in the subset of proteins arising from MetAP-guided co-translational cleavage rules. Among MetAP unprocessed substrates, five

MK-starting proteins were identified, but we ruled out that they might undergo K-myristoylation. A number of internal KXXSK K-myristoylation motifs are preceded by caspase cleavage sites in the human proteome suggesting that K-myristoylation - in contrast to mostly co-translational G-myristoylation - results from post-translational addition, for instance initiated by cleavage by specific caspases during apoptosis^{15,29}.

We show that post-translational myristoylation may occur either after initial G-myristoylation, causing double myristoylation, or upon protease cleavage. This includes the action of bacterial protease IpaJ, which we used as a case study and as a unique tool to discriminate between G- and K-myristoylation. Our data confirm that ARF6 is myristoylated in both states^{17,18}, in contrast to the other ARFs (ARF1/3/4/5) or ARL1, which do not feature a K3 like ARF6. This insensitivity of ARF6 to IpaJ was initially proposed to be due to ARF6 not being located at the ER/Golgi, unlike the other members of the ARF/ARL family, but rather at the plasma membrane. ARF6 translocates to endosomes, where it acts as a major regulator of endocytosis^{30,31}. Our data indicate that ARF6 resistance to IpaJ might also result from the unmasking of a myristoylatable K3 upon G-myristoylation cleavage. As a result, ARF6 would still be myristoylated and retained on myristoylation affinity azido-biotin prior to and after IpaJ cleavage.

We noticed that K-myristoylation or any other side chain myristoylation systematically reduced k_{cat} values, making the reaction poorly efficient. The associated k_{cat} values were significantly lower than G-myristoylation and in the low range of reported values³². These data, together with the observation of direct, non-water mediated bonding of the reactive amino group of the side chain to the catalytic base, suggested the occurrence of long-lasting enzyme-ligand or enzyme-product intermediary complexes prior to product release. Taking advantage

of this observation, we designed a series of NMT inhibitors based on our new knowledge of the determinants of side chain myristoylation. Specifically, a K-myristoylation substrate could be turned into a very potent suicide inhibitor through simple manipulation around the acceptor side chain, i.e., the stereo-isomeric D version of a K-mimicking, shorter side chain, including Orn or Dab. This reflects the tight fit of the active site and mechanism of both K- and G-myristoylation and highlights that even very small modifications (such as methylene variations, e.g., G2 to A or me-Gly for G-myristoylation, acK to ac[Orn] for K-myristoylation) produce distinct and significant effects on the reaction outcome.

To conclude, this study extends our knowledge of the subproteome, which is sensitive to NMT-guided myristoylation, and provides new understanding about the subtle requirements for specific NMT inhibition.

ACKNOWLEDGMENTS

This work was supported by French National Research Agency (ANR) DynaMYT (ANR-20-CE44-0013) and Fondation ARC (ARCPJA32020060002137) grants to TM. This work has benefited from the support of a French State grant (ANR-17-EUR-0007, EUR SPS-GSR) managed by the ANR under an Investments for the Future program (ANR-11-IDEX-0003-02), from the I2BC crystallization platform supported by FRISBI (ANR-10-INSB-05-01), and from the facilities and expertise of the I2BC proteomic platform SICaPS, supported by IBiSA, Ile de France Region, Plan Cancer, CNRS and Paris-Sud University. FR is supported by grants from Région Ile-de-France (17012695) and Fondation pour la Recherche Médicale (FDT202001010779). We warmly acknowledge Odile Schiltz (IPBS, Toulouse), Virginie Redeker, Jean-Pierre Le Caer, Laila Sago and David Cornu (all at SICaPS, Gif/Yvette) for their

extensive support with mass spectrometry analyses. We thank the French National Synchrotron Facility (SOLEIL) for providing synchrotron radiation facilities and the staff of the Proxima 1 & 2 beamlines.

AUTHOR CONTRIBUTIONS

FR conducted all MS experiments and kinetic analyses, characterized IpaJ variants, and set up the IpaJ/NMT pipeline. CD undertook cloning, mutagenesis, and purification of IpaJ and performed all NMT structural analyses. RFD completed and consolidated the kinetic analysis. TM and CG conceived the project, supervised the experiments and analyzed the data. FR, CD, CG and TM, wrote the manuscript.

DATA AVAILABILITY

The nine crystal structures of NMT in complex with MyrCoA and the peptides reported here have been deposited at the PDB under codes 7OWM, 7OWN, 7OWO, 7OWP, 7OWQ, 7OWR, 7OWS, 7OWT and 7OWU.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper.

METHODS

Chemicals

All peptides (see all sequences in **Supplementary Tables 1**) were purchased at 95% purity (Genscript, Leiden, Netherlands). NAD was purchased from Roche (Basel, Switzerland). All other chemicals were purchased from Sigma-Aldrich (St. Quentin, France). Stock solutions of myristoyl-CoA (0.2 mM) were prepared in sodium acetate, pH 5.6, and 1% Triton X-100, except for MALDI analysis, where cholate was used to reduce background.

Because the N-terminal myristoylated G residue is usually the second in the original open reading frame (ORF) after the initial M, we call it aa2 as the extreme N-terminal residue throughout this work for consistency with previous work from ourselves and others.

IpaJ cloning

Wild-type full-length IpaJ numbering is used throughout the text. The nucleotide sequence encoding residues including the final TGA stop codon of *S. flexneri IpaJ* ORF (UniProt code Q54150) was optimized for *E. coli* expression (**Supplementary Table 3**) and synthesized *in vitro* (Genscript). For protein expression, the 30-259 fragment was cloned between the *NcoI* and *HindIII* restriction sites of expression vector pETM30 as a C-terminal fusion with glutathione S-transferase (GST, **Extended Data 2a**). A His-tag was placed at the N-terminus of the fusion to facilitate protein purification (see below). A TEV cleavage sequence inserted between the two ORFs allowed release of IpaJ from GST with only an additional N-terminal G-Ala-Met-Ala tetrapeptide sequence upstream of R30, the first residue of IpaJ. Two variants were created by site-directed mutagenesis using a QuikChange site-directed mutagenesis kit (Stratagene, San Diego, CA) with the primer pairs displayed in **Supplementary Table 3**. The first variant was inactive (C64S), while the other (3M) was fully active with improved solubility and stability over time (L97N/L99N/C103S).

Enzyme production and purification

HsNMT1l. The *HsNMT1l* isoform containing residues 81-496 corresponds to the major isoform³. *HsNMT1l* was cloned into pET16b as an N-terminal His-tag fusion³, and the recombinant protein was expressed and purified as described¹⁶. Isoform *HsNMT1s* containing residues 99-496 was cloned into pET28 and purified as described¹⁷.

The three soluble 30-259 variants of the protease IpaJ were produced as follows. Rosetta2pLysS cells expressing the pETM30 derivative were grown in 1 L of 2xTY medium supplemented with kanamycine (50 µg/mL) and chloramphenicol (34 µg/mL) at 37°C under vigorous shaking. Protein overexpression was induced with 0.5 mM IPTG at OD₆₀₀=0.4. Cells were transferred at 20°C and grown for 20 hours. Cells were centrifuged at 10,000 x g, and the pellet was resuspended (10 mL/g) in lysis buffer (20 mM Tris pH=8.0, 0.2 M NaCl, 5 mM 2-mercaptoethanol, 5 mM imidazole, 5% glycerol). Cells were lysed with a sonicator Q700 (amplitude 50%, 10 s on and 20 s off) for 3 min at 4°C. The sample was then centrifuged at 40,000 x g (Rotor JA20) for 30 min and the pellet discarded. The supernatant was loaded on a Ni-IMAC-HisTrapTM FF 5 mL column (GE Healthcare, Chicago, IL) at 3.5 mL/min; elution was achieved with a linear 0-0.5M imidazole gradient run over 100 ml. Purified proteins were dialyzed in Spectra-Por7 semi-permeable dialysis membranes (8 kDa cut-off; Thermo Fisher Scientific, Waltham, MA) for 48 hours against conservation buffer (20 mM Tris pH=8.0, 0.2 M NaCl, 5 mM dithiothreitol, 55% glycerol) at 4°C. The sample was stored at -20°C. Protein concentrations were measured with the Bio-Rad Protein Assay Kit using bovine serum albumin (Sigma-Aldrich) as reference.

Protein crystallization and structure determination

HsNMT1:MyrCoA:X structure. *HsNMT1s* was used to solve the structures of the complexes with nine different peptides in the course of this study (HPCA, AK, AcK, AcG[Orn], AN,

meGN, GGK, Dab and ac[D-Orn] with PDB entries 7OWM, 7OWN, 7OWO, 7OWP, 7OWU, 7OWQ, 7OWR, 7OWS and 7OWT, respectively. Suitable crystals of HsNMT1:MyrCoA:peptide substrates or inhibitors were obtained by co-crystallization using the hanging-drop vapor diffusion method at 20°C in the crystallization conditions previously described^{16,17}. Briefly, crystallization droplets were formed by mixing 2 µL of the of HsNMT1:MyrCoA:peptide complex (ratio 1:1.5:1.5) at 6-9 mg/mL with 2 µL of the precipitant solution containing either 0.1 M MgCl₂, 0.2 M NaCl, 0.1 M sodium citrate pH 5.6, and 18-24% (w/v) PEG 6K or 0.1 M sodium acetate pH 4.6, and 18-24% (w/v) PEG 6K. Crystals were cryoprotected in the reservoir solution supplemented with 15% (v/v) glycerol and flash cooled in liquid nitrogen. Complete X-ray datasets of complex were collected at 100K a single wavelength from a single crystal at the French National Synchrotron Facility (SOLEIL) PROXIMA1 (for HPCA and AcG[Orn] at $\lambda=0.98400$ Å and AcK at $\lambda=0.97856$ Å) or PROXIMA2 (for AK and GGK at $\lambda=0.98012$ Å and Dab and Ac[D-ORN] at $\lambda=0.97563$ Å) beamlines and at European Radiation Synchrotron Facility (ESRF) ID30A1 (for AN at $\lambda=0.96600$ Å) and ID30A3 (for me-GN $\lambda=0.96775$ Å) beamlines. Datasets were integrated with XDS³³ and scaled and reduced using AIMLESS from the CCP4 package³⁴. For crystals that suffered from anisotropic diffraction (HPCA, AcK and GGK), data were processed with STARANISO on unmerged data³⁵ before AIMLESS data reduction. Crystals belonged essentially to the space group P2₁2₁2 with similar unit cell parameters with the exception of GGK and meGN belonging both to C2 with similar unit cell parameters (summarized in **Supplementary Table 4**). In both space groups identified, unit cells contained two NMT molecules per asymmetric unit. Structure resolution was accomplished in all cases using the molecular replacement method and solved using PHASER³⁶ and the HsNMT1:MyrCoA:peptide ternary complex (PDB entry 5O9T or 6SK2) as a search model.

The structure of AN was solved using MOLREP³⁷ and protein coordinates of HsNMT1:MyrCoA (PDB entry 5O9T) as a search model. Structures were subjected to alternating refinement cycles using PHENIX and manual model building using COOT³⁸⁻⁴⁰. The good quality of the electron density maps also enabled the refinement of substrate peptide, reaction intermediate, and reaction product molecules bound to HsNMT1 in each complex. Chemical compound libraries were generated using PRODRG server⁴¹ in combination with eLBOW from the PHENIX suite. The geometry of the final models was validated using MOLPROBITY⁴². Figures were generated using PYMOL (DeLano Scientific LLC, <http://pymol.sourceforge.net/>). X-ray data collection and refinement statistics are summarized in **Supplementary Table 4**.

Measurements of activity and inhibition parameters

HsNMT1 activity was assayed at 30°C in a coupled assay using an updated version of the previously described protocol⁴³. The reaction mixture contained 50 mM Tris-HCl (pH 8.0), 1 mM MgCl₂, 0.193 mM EGTA, 0.32 mM DTT, 0.2 mM thiamine pyrophosphate, 2 mM pyruvate, 0.1 mg/mL of BSA, 0.1% Triton X-100, 2.5 mM NAD⁺, 0.125 units/mL of porcine heart PDH (0.33 units/mg), 40 μM myristoyl-CoA, and 1-2000 μM peptides. Unless otherwise stated for tight binding studies, the reaction mixture was pre-incubated for 3 min at 30°C before starting the reaction by adding MyrCoA. A final volume of 200 μL was used in 96-well black plates (Grenier Bio One and Dominique Dutscher, Brumath, France; the optical path for 0.2 mL is 0.55 cm). A value of 6300 M⁻¹.s⁻¹ was used as the molecular extinction coefficient of NADH at 340 nm. An Infinite M Nano⁺ plate reader equipped with micro-injectors (Tecan, Lyon, France) was set at 340 nm to monitor the absorbance over time at 30°C. Briefly, a reaction mixture containing HsNMT1 at different concentrations of peptide acceptors was pre-incubated for 3 min at 30°C before starting the reaction with MyrCoA.

Myristoylation kinetics were monitored continuously for 30 min, and the data were fitted as to obtain the initial velocities associated to each peptide concentration. Curve fits to obtain kinetic parameters were achieved by non-linear regression with GraphPad Prism 9.1 (GraphPad Software, La Jolla, CA). Parameters with standard errors were computed for all parameters using the complete dataset including replicates. Both k_{cat} and K_m kinetic parameters were obtained by fitting to the Michaelis–Menten equation. k_{cat}/K_m values and the associated standard deviations were obtained by taking advantage of the k_{SP} approach⁴⁴ with $k_{SP}=k_{cat}/K_m$ and $v_0/[E]=k_{SP}[S]/(1+[S]/K_m)$, where v_0 is the reaction rate measured at NMT concentration $[E_0]$.

For inhibition assays, we used the same reaction mixture for 30 min at 30°C incubation in the presence of the inhibitor peptide and MyrCoA. The reaction was triggered by the addition of SOS3 peptide and monitored for 30 min. Inhibition curves and associated IC_{50} values were fitted with the absolute IC_{50} module with baseline set at 0. Inhibition constants were calculated as described in⁴⁵⁻⁴⁷ with $K_I=(IC_{50}-[E]/2)/(1+[S]/K_m)$. The HsNMT1 concentration was 0.25 μ M, and the substrate (G₂CSVKKK; SOS3; $K_m=18\pm3$ μ M^{43,48}) was used at a concentration of 20 μ M.

The NMT/IpaJ pipeline described in **Fig. 5b** first involved incubation of the peptide (100 μ M) for 1 hour in the presence of 0.5 μ M HsNMT1 or HsNMT2 (T1). The buffer was the same as for NMT activity measurement but contained cholate (5 μ M) instead of Triton X-100. Full IpaJ cleavage conditions involved further incubation at 20°C for 1 hour in the presence of 10 μ M IpaJ-3M (T2). 5 μ M HsNMT1 was finally added for 60 min at 30°C (T3). T0 corresponded to the T3 time (3 hours) with the peptide diluted in the incubation buffer but in the absence of any enzyme.

Mass spectrometry

300 μ L of a mixture containing 50 mM Tris (pH 8), 0.193 mM EGTA, 1 mM $MgCl_2$, 1 mM DTT, 5 μ M sodium cholate, 40 μ M Myr-CoA solution (stock solution 0.2 mM Myr-CoA, 10 mM sodium acetate, 2.5 μ M sodium cholate), 0.5 μ M NMT, and 100 μ M of synthetic peptide were pre-incubated at 30°C. The myristoylation reaction was followed over time by collection of 10 μ L samples further diluted in 90 μ L of water/acetonitrile (90/10) solution. The different samples were then diluted five times in the matrix solution made of 5 mg/mL of α -cyano-4-hydroxycinnamic acid solubilized in water/formic acid/acetonitrile (50/50/0.1%). 1 μ L of each dilution was spotted on a metal target and dried. MS and MS/MS spectra of each sample were acquired with an AB SCIEX 5800 MALDI-Tof-Tof instrument in positive ion mode as reported in SI. Survey scans were performed using delayed extraction (390 ns) in reflector mode for a total of 15,000 shots. MS/MS scans were operated with a collision energy of 1 kV. Peptide and fragment mass tolerances were set at 10 ppm and 0.8 Da, respectively. Mass spectra were analyzed with Peakview software (AB Sciex, Macclesfield, UK), and MS2 analysis was performed with ProteinProspector v6.2.2 (<https://prospector.ucsf.edu/prospector/mshome.htm>)⁴⁹. All spectra are available in **Supplementary Dataset 1**. Theoretical mass values were calculated at https://web.expasy.org/peptide_mass/.

REFERENCES

1. Meinel, T., Dian, C. & Giglione, C. Myristoylation, an ancient protein modification mirroring eukaryogenesis and evolution. *Trends Biochem Sci* **45**, 619-632 (2020).
2. Lodge, J.K., Johnson, R.L., Weinberg, R.A. & Gordon, J.I. Comparison of myristoyl-CoA:protein N-myristoyltransferases from three pathogenic fungi: *Cryptococcus neoformans*, *Histoplasma capsulatum*, and *Candida albicans*. *J Biol Chem* **269**, 2996-3009 (1994).

- 825 3. Pierre, M. et al. N-Myristoylation regulates the SnRK1 pathway in Arabidopsis. *Plant*
826 *Cell* **19**, 2804-2821 (2007).
- 827 4. Price, H.P. et al. Myristoyl-CoA:protein N-myristoyltransferase, an essential enzyme
828 and potential drug target in kinetoplastid parasites. *J Biol Chem* **278**, 7206-14 (2003).
- 829 5. Bhatnagar, R.S., Ashrafi, K., Futterer, K., Waksman, G. & Gordon, J.I. Biology and
830 enzymology of protein N-myristoylation. in *The enzymes*, Vol. XXI (Protein lipidation) (eds.
831 Tamanoi, F. & Sigman, D.S.) 241-286 (Academic Press, San Diego, 2001).
- 832 6. Resh, M.D. Trafficking and signaling by fatty-acylated and prenylated proteins. *Nat*
833 *Chem Biol* **2**, 584-90 (2006).
- 834 7. Traverso, J.A. et al. Roles of N-Terminal fatty acid acylations in membrane
835 compartment partitioning: Arabidopsis h-type thioredoxins as a case study. *Plant Cell* **25**,
836 1056-77 (2013).
- 837 8. Frearson, J.A. et al. N-myristoyltransferase inhibitors as new leads to treat sleeping
838 sickness. *Nature* **464**, 728-32 (2010).
- 839 9. Wright, M.H. et al. Validation of N-myristoyltransferase as an antimalarial drug target
840 using an integrated chemical biology approach. *Nat Chem* **6**, 112-21 (2014).
- 841 10. Wright, M.H., Paape, D., Price, H.P., Smith, D.F. & Tate, E.W. Global profiling and
842 inhibition of protein lipidation in vector and host stages of the sleeping sickness parasite
843 *Trypanosoma brucei*. *ACS Infect Dis* **2**, 427-441 (2016).
- 844 11. Mousnier, A. et al. Fragment-derived inhibitors of human N-myristoyltransferase block
845 capsid assembly and replication of the common cold virus. *Nat Chem* **10**, 599-606 (2018).
- 846 12. Beauchamp, E. et al. Targeting N-myristoylation for therapy of B-cell lymphomas. *Nat*
847 *Commun* **11**, 5348 (2020).

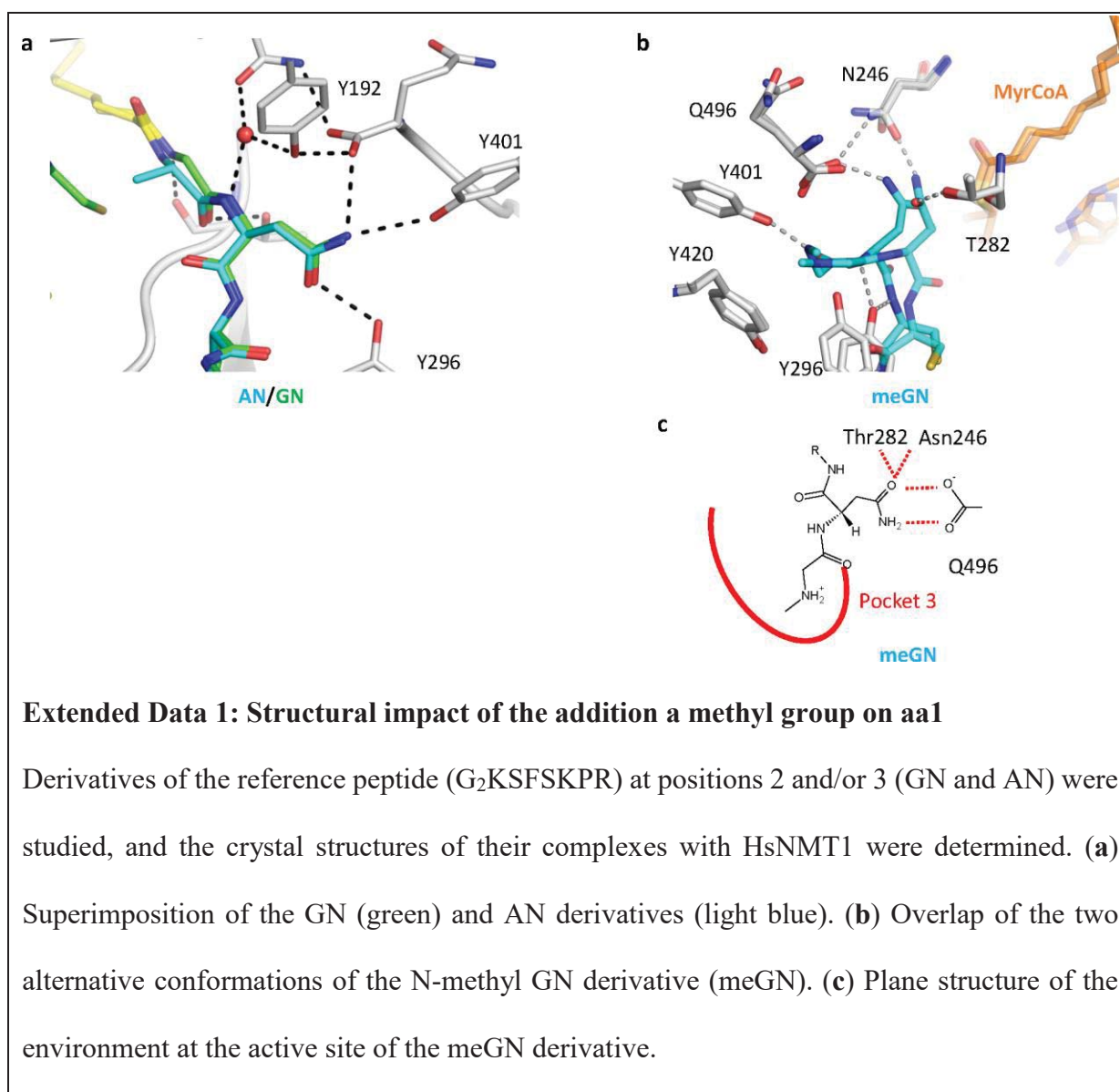
- 848 13. Kosciuk, T. & Lin, H. N-Myristoyltransferase as a glycine and lysine
849 myristoyltransferase in cancer, immunity, and infections. *ACS Chem Biol* **15**, 1747-1758
850 (2020).
- 851 14. Martin, D.D., Beauchamp, E. & Berthiaume, L.G. Post-translational myristoylation:
852 Fat matters in cellular life and death. *Biochimie* **93**, 18-31 (2011).
- 853 15. Thinon, E. et al. Global profiling of co- and post-translationally N-myristoylated
854 proteomes in human cells. *Nat Commun* **5**, 4919 (2014).
- 855 16. Castrec, B. et al. Structural and genomic decoding of human and plant myristoylomes
856 reveals a definitive recognition pattern. *Nat Chem Biol* **14**, 671-679 (2018).
- 857 17. Dian, C. et al. High-resolution snapshots of human N-myristoyltransferase in action
858 illuminate a mechanism promoting N-terminal Lys and Gly myristoylation. *Nat Commun* **11**,
859 1132 (2020).
- 860 18. Kosciuk, T. et al. NMT1 and NMT2 are lysine myristoyltransferases regulating the
861 ARF6 GTPase cycle. *Nat Commun* **11**, 1067 (2020).
- 862 19. Carr, S.A., Biemann, K., Shoji, S., Parmelee, D.C. & Titani, K. n-Tetradecanoyl is the
863 NH₂-terminal blocking group of the catalytic subunit of cyclic AMP-dependent protein kinase
864 from bovine cardiac muscle. *Proc Natl Acad Sci U S A* **79**, 6128-31 (1982).
- 865 20. Gheorghe, M.T. & Bergman, T. Deacetylation and internal cleavage of polypeptides
866 for N-Terminal sequence analysis. in *Methods in Protein Structure Analysis* (eds. Atassi, M.Z.
867 & Appella, E.) 81-86 (Springer, Boston, MA, 1995).
- 868 21. Leone, J.W. et al. Removal of N-terminal blocking groups from proteins. *Curr Protoc*
869 *Protein Sci* **63**, 11.7.1-11.7.20 (2011).
- 870 22. Mattock, E. & Blocker, A.J. How do the virulence factors of Shigella work together to
871 cause disease? *Front Cell Infect Microbiol* **7**(2017).

- 872 23. Burnaevskiy, N., Peng, T., Reddick, L.E., Hang, H.C. & Alto, N.M. Myristoylome
873 profiling reveals a concerted mechanism of ARF GTPase deacylation by the bacterial protease
874 IpaJ. *Mol Cell* **58**, 110-22 (2015).
- 875 24. Burnaevskiy, N. et al. Proteolytic elimination of N-myristoyl modifications by the
876 Shigella virulence factor IpaJ. *Nature* **496**, 106-9 (2013).
- 877 25. Boisson, B., Giglione, C. & Meinnel, T. Unexpected protein families including cell
878 defense components feature in the N-myristoylome of a higher eukaryote. *J Biol Chem* **278**,
879 43418-43429 (2003).
- 880 26. Utsumi, T. et al. Vertical-scanning mutagenesis of amino acids in a model N-
881 myristoylation motif reveals the major amino-terminal sequence requirements for protein N-
882 myristoylation. *Eur J Biochem* **271**, 863-74 (2004).
- 883 27. Houglund, J.L. et al. Identification of novel peptide substrates for protein
884 farnesyltransferase reveals two substrate classes with distinct sequence selectivities. *J Mol Biol*
885 **395**, 176-90 (2010).
- 886 28. Chang, A. et al. BRENDA, the ELIXIR core data resource in 2021: new developments
887 and updates. *Nucleic Acids Res* **49**, D498-D508 (2021).
- 888 29. Zha, J., Weiler, S., Oh, K.J., Wei, M.C. & Korsmeyer, S.J. Posttranslational N-
889 myristoylation of BID as a molecular switch for targeting mitochondria and apoptosis. *Science*
890 **290**, 1761-1765 (2000).
- 891 30. Donaldson, J.G. Multiple roles for Arf6: sorting, structuring, and signaling at the
892 plasma membrane. *J Biol Chem* **278**, 41573-41576 (2003).
- 893 31. Donaldson, J.G. & Jackson, C.L. ARF family G proteins and their regulators: roles in
894 membrane transport, development and disease. *Nat Rev Mol Cell Biol* **12**, 362-75 (2011).
- 895 32. Bar-Even, A. et al. The moderately efficient enzyme: evolutionary and
896 physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402-10 (2011).

- 897 33. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially
898 unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795-800 (1993).
- 899 34. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution?
900 *Acta Crystallogr D Biol Crystallogr* **69**, 1204-14 (2013).
- 901 35. Tickle, I.J. et al. STARANISO ([http://staraniso.globalphasing.org/cgi-](http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi)
902 [bin/staraniso.cgi](http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi)). (Global Phasing Ltd. , Cambridge, United Kingdom, 2018).
- 903 36. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674
904 (2007).
- 905 37. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular
906 replacement. *J. Appl. Cryst.* **30**, 1022-1225 (1997).
- 907 38. Murshudov, G.N. et al. REFMAC5 for the refinement of macromolecular crystal
908 structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-67 (2011).
- 909 39. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for
910 macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
- 911 40. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of
912 Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
- 913 41. Schuttelkopf, A.W. & van Aalten, D.M. PRODRG: a tool for high-throughput
914 crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **60**, 1355-63
915 (2004).
- 916 42. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular
917 crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
- 918 43. Boisson, B. & Meinnel, T. A continuous assay of myristoyl-CoA:protein N-
919 myristoyltransferase for proteomic analysis. *Anal Biochem* **322**, 116-123 (2003).
- 920 44. Johnson, K.A. New standards for collecting and fitting steady state kinetic data.
921 *Beilstein J Org Chem* **15**, 16-29 (2019).

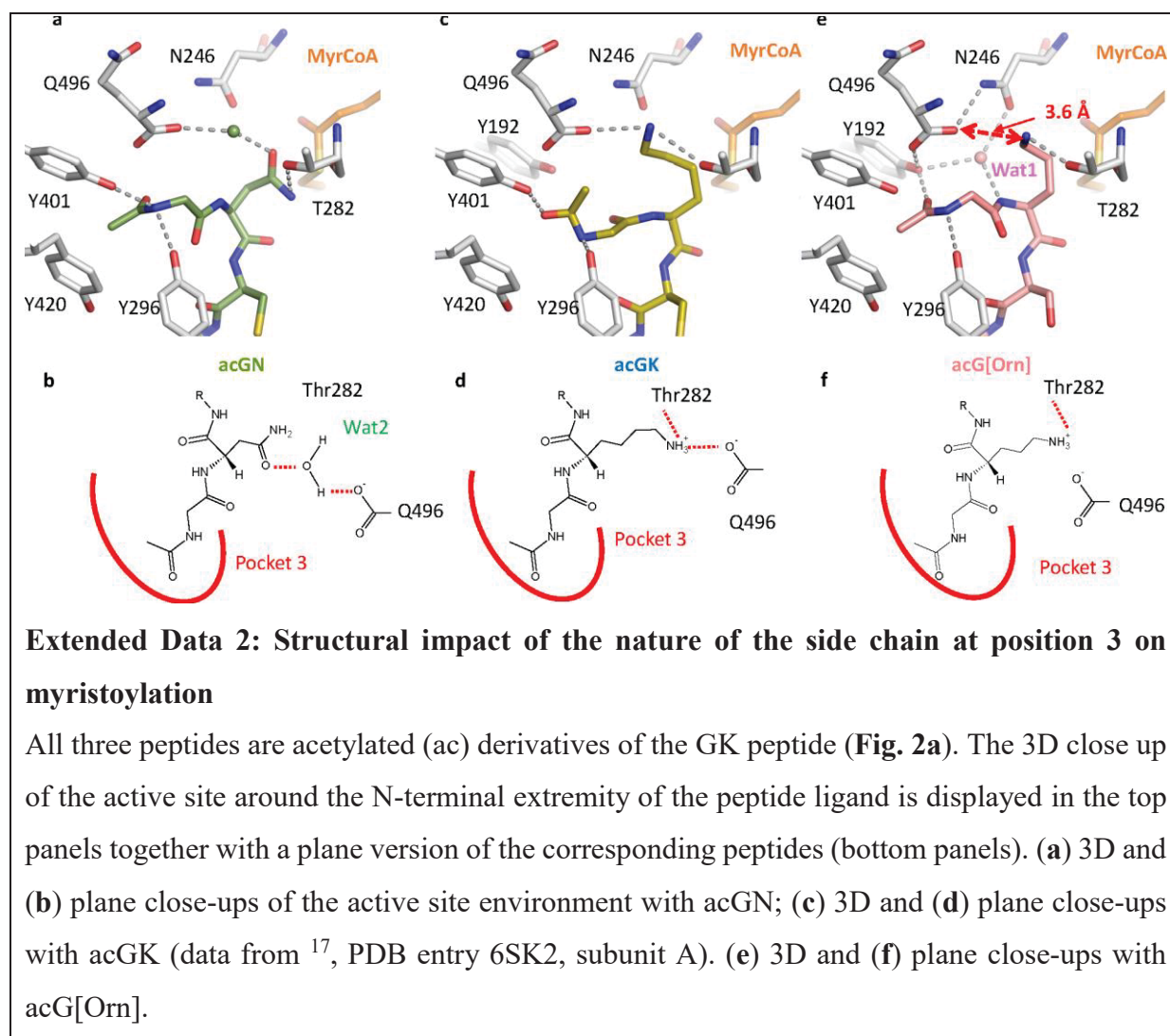
45. Copeland, R.A. *Evaluation of enzyme inhibitors in drug discovery: a guide for medicinal chemists and pharmacologists*, 296 (John Wiley & Sons, New Jersey, 2005).
46. Copeland, R.A., Lombardo, D., Giannaras, J. & Decicco, C.P. Estimating KI values for tight binding inhibitors from dose-response plots. *Bioorg Med Chem Lett* **5**, 1947-1952 (1995).
47. Fieulaine, S. et al. Trapping conformational states along ligand-binding dynamics of peptide deformylase: the impact of induced fit on enzyme catalysis. *PLoS Biol* **9**, e1001066 (2011).
48. Traverso, J.A., Giglione, C. & Meinnel, T. High-throughput profiling of N-myristoylation substrate specificity across species including pathogens. *Proteomics* **13**, 25-36 (2013).
49. Chalkley, R.J., Baker, P.R., Medzihradszky, K.F., Lynn, A.J. & Burlingame, A.L. In-depth analysis of tandem mass spectrometry data from disparate instrument types. *Mol Cell Proteomics* **7**, 2386-98 (2008).
50. Johnson, D.R., Bhatnagar, R.S., Knoll, L.J. & Gordon, J.I. Genetic and biochemical studies of protein N-myristoylation. *Annu Rev Biochem* **63**, 869-914 (1994).
51. Kumar, S., van Raam, B.J., Salvesen, G.S. & Cieplak, P. Caspase cleavage sites in the human proteome: CaspDB, a database of predicted substrates. *PLoS One* **9**, e110539 (2014).

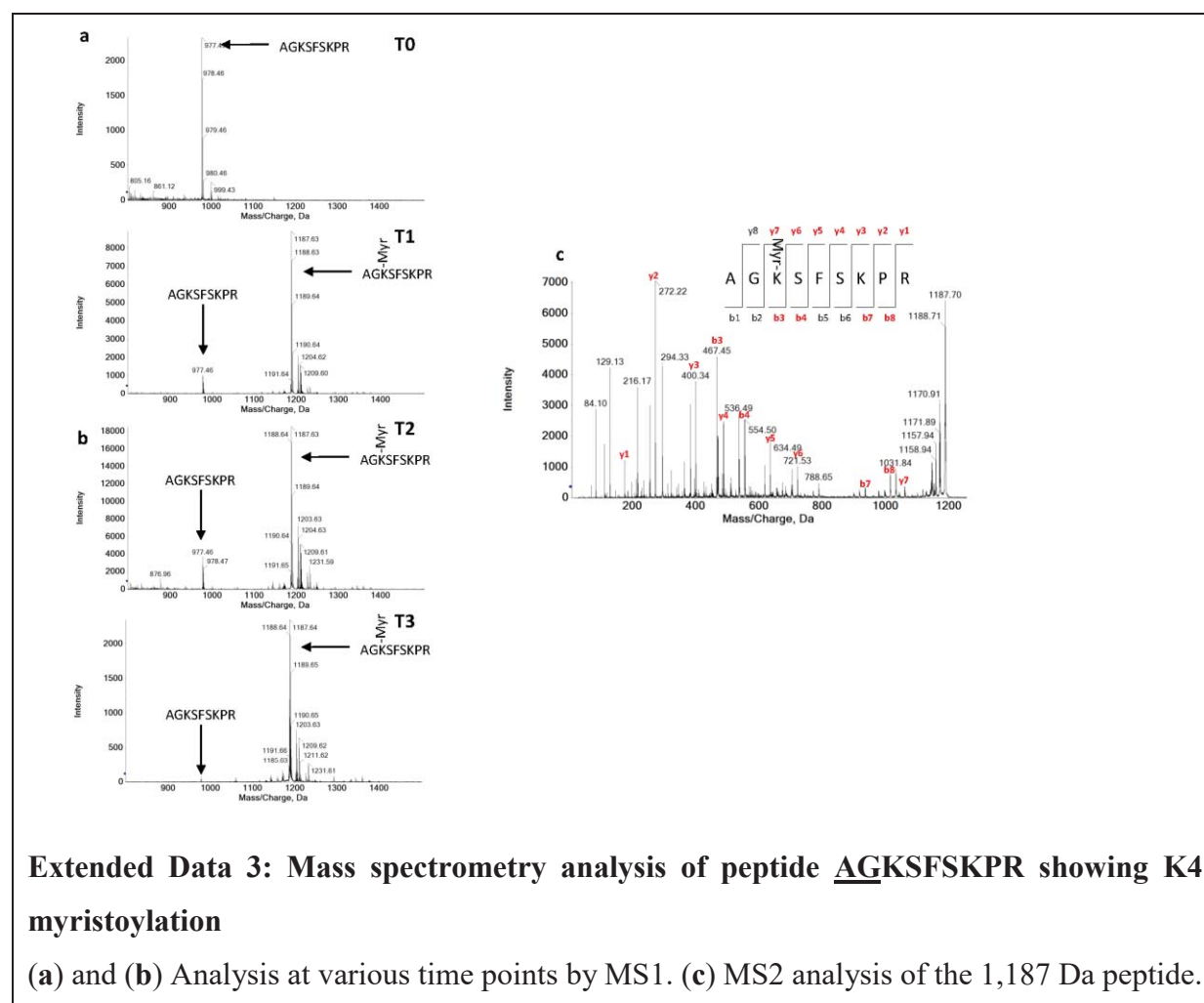
EXTENDED DATA



Extended Data 1: Structural impact of the addition a methyl group on aa1

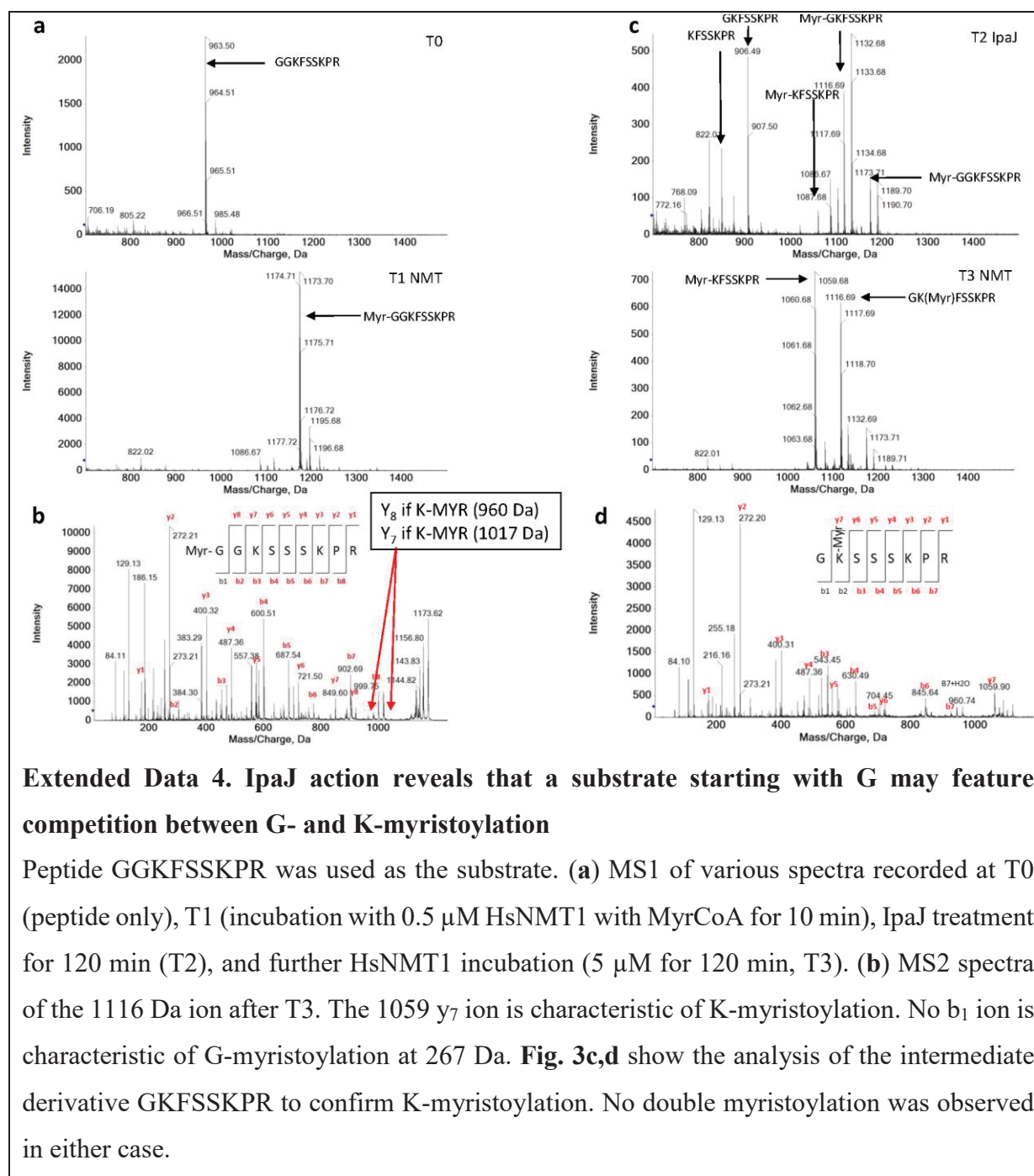
Derivatives of the reference peptide (G₂KSFSKPR) at positions 2 and/or 3 (GN and AN) were studied, and the crystal structures of their complexes with HsNMT1 were determined. **(a)** Superimposition of the GN (green) and AN derivatives (light blue). **(b)** Overlap of the two alternative conformations of the N-methyl GN derivative (meGN). **(c)** Plane structure of the environment at the active site of the meGN derivative.





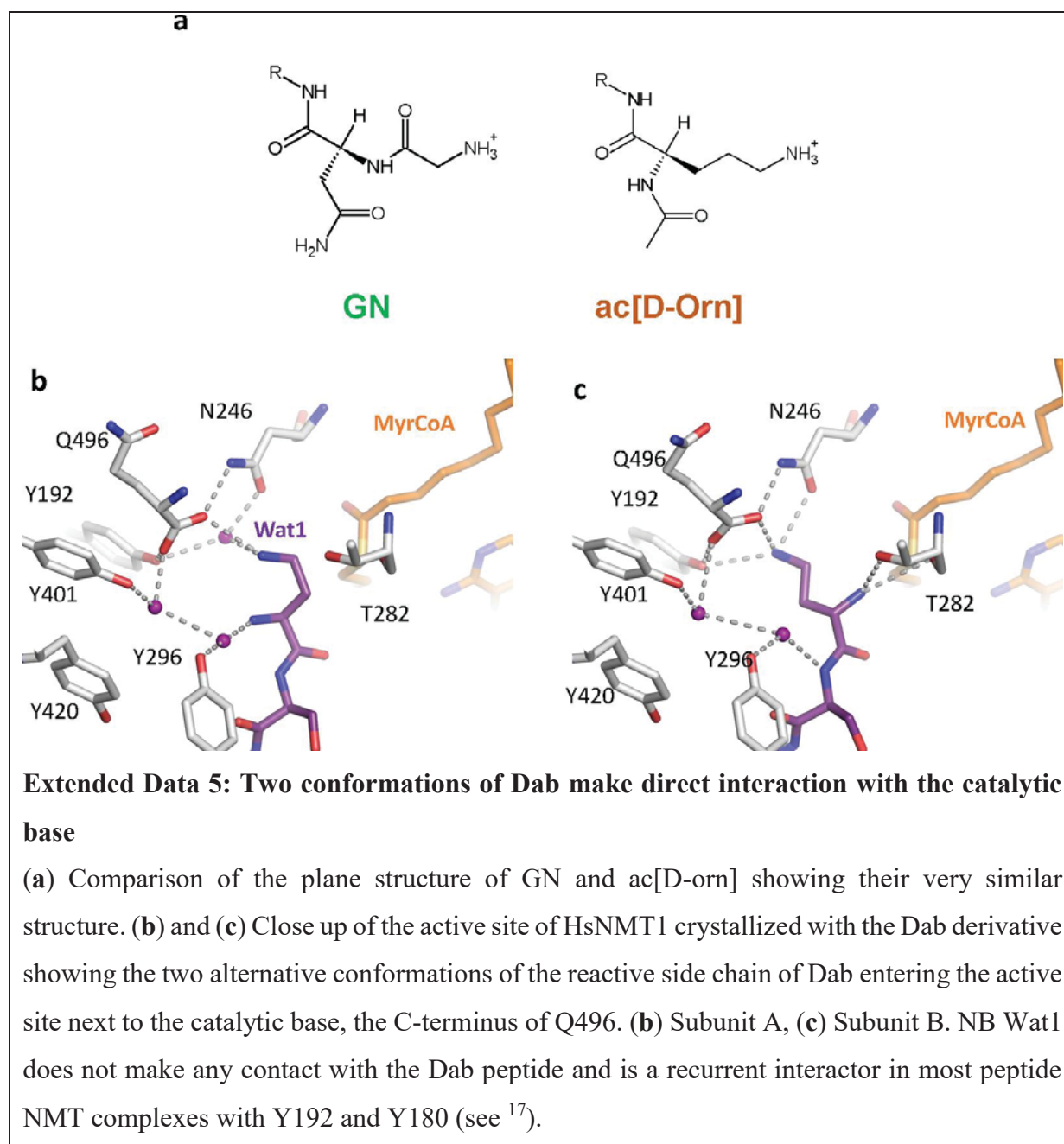
Extended Data 3: Mass spectrometry analysis of peptide AGKSFSKPR showing K4 myristoylation

(a) and (b) Analysis at various time points by MS1. (c) MS2 analysis of the 1,187 Da peptide.



Extended Data 4. IpaJ action reveals that a substrate starting with G may feature competition between G- and K-myristoylation

Peptide GGKFSSKPR was used as the substrate. (a) MS1 of various spectra recorded at T0 (peptide only), T1 (incubation with 0.5 μ M HsNMT1 with MyrCoA for 10 min), IpaJ treatment for 120 min (T2), and further HsNMT1 incubation (5 μ M for 120 min, T3). (b) MS2 spectra of the 1116 Da ion after T3. The 1059 y_7 ion is characteristic of K-myristoylation. No b_1 ion is characteristic of G-myristoylation at 267 Da. Fig. 3c,d show the analysis of the intermediate derivative GKFSSKPR to confirm K-myristoylation. No double myristoylation was observed in either case.



Chapitre V

V. Discussion-Conclusions

L'objectif premier de ma thèse était d'augmenter significativement la couverture d'identification des protéines subissant la myristoylation eucaryote grâce à l'utilisation d'une enzyme capable de retirer non seulement cette modification mais laissant aussi une preuve de son activité avec l'excision de la glycine. Cette enzyme – IpaJ - a été caractérisée comme un effecteur principal fortement lié à la toxicité de l'agent principal de la dysenterie, la bactérie *Shigella flexneri* (Burnaevskiy *et al*, 2013, 2015; Slagowski *et al*, 2008). Un deuxième aspect de mon objectif de thèse était lié à une meilleure compréhension de la myristoylation en lysine, un domaine peu abordé jusqu'alors. Le constat initial était qu'on ignorait le catalyseur de la réaction de modification de certaines protéines comme le TNF1 α . Dans ce contexte, l'enzyme IpaJ s'est avérée un outil précieux pour caractériser son impact. L'utilisation d'IpaJ comme outil d'assistance à l'étude de la myristoylation en glycine et en lysine constitue donc le fil conducteur des approches que j'ai menées.

V.1. Validation d'IpaJ comme outil d'assistance à l'étude de la myristoylation

Dans ce contexte, j'ai produit de manière hétérologue et purifié des variant actifs et stables de l'enzyme de *Shigella flexneri* et mis au point un test rapide par spectrométrie de masse Maldi-TOF permettant de qualifier l'activité de la protéine et d'en déterminer la spécificité de substrat. J'ai ainsi pu montrer que cette enzyme était capable d'hydrolyser des peptides N-myristoylés *in vitro* dont les séquences différaient fortement des N-terminaux des cibles déjà connues d'IpaJ. Les données obtenues avec différents peptides myristoylés m'ont conduit à compléter l'étude de sélectivité de substrats d'IpaJ. Pour ce faire, une étape de couplage *in vitro* de la myristoylation par la NMT à l'activité IpaJ a facilité l'étude de la sélectivité de substrats d'IpaJ et m'a donné accès à la grande quantité de peptides positifs à la N-myristoylation *in vitro* disponible dans l'équipe. La mise en place de ce nouveau test a permis de démontrer la démyristoylation par IpaJ d'un plus grand nombre de substrats de la NMT présentant

une grande variation de séquences. Ces résultats mettent en lumière la faible importance de la séquence située en aval de la glycine N-myristoylée dans l'activité de démyristoylation de cette protéase et suggère qu'IpaJ est effectivement capable de cliver tous les substrats de la NMT *in vitro*. Ce résultat est différent du contexte infectieux puisqu'IpaJ a été décrite comme spécifique d'une seule famille de protéines, les petites protéines G ARF/ARL. L'équipe du Pr N. Alto a proposé que cette spécificité restreinte repose sur une interaction avec des partenaires de la famille. Par ailleurs, j'ai vérifié la capacité de la protéine IpaJ à discriminer entre une protéine myristoylée et une protéine G-acétylée. Cette spécificité nous est apparue importante pour éviter des erreurs d'interprétation entre les deux types de modifications, les protéines commençant par G et acétylées étant deux fois plus abondantes que les protéines myristoylées et certaines peuvent subir chacune des deux modifications (Castrec *et al*, 2018). Une fois tous les prérequis pour l'utilisation d'IpaJ comme outil biotechnologique pour l'étude du myristoylome tous vérifiés, j'ai mis au point le protocole pour cette étude. J'ai dans un premier temps utilisé une méthode bien établie de fractionnement cellulaire afin d'augmenter la profondeur de l'analyse en décomplexant l'échantillon (**Figure II-6**). J'ai été amené à l'adapter afin de la rendre compatible avec l'activité d'IpaJ et à l'analyse en spectrométrie de masse.

V.2. Validation de modèles biologiques adaptés

L'outil IpaJ et sa mise en application étant presque validés, il m'a fallu disposer d'un modèle biologique adapté. J'ai ensuite sélectionné une douzaine de lignées cellulaires d'origine tumorale ou immortalisées. L'intérêt des types cellulaires tumoraux, repose sur le fait qu'un grand nombre de tumeurs présente une surexpression soit de la NMT soit de certains de leurs substrats. J'ai pu observer que, si l'expression de la HsNMT1 sous ces différentes formes n'est pas identique dans tous les types cellulaires, la surexpression d'une de ces formes n'est pas corrélée avec l'expression de l'autre forme. Ces deux formes, issues d'un épissage alternatif différent au niveau de la longueur de leur N-terminal, et donc joueraient un rôle sur la fonction de la NMT. La partie N-

terminale de la NMT est supposée interagir avec le ribosome, la forme courte aurait donc une fonction de myristoylation post-traductionnelle (acylation sur des substrats non liés au ribosome, déjà repliés et issus de coupure protéolytique), alors que la version longue serait majoritairement impliquée dans l'activité co-traductionnelle. De plus, après une analyse protéomique de quatre de ces lignées, ayant des niveaux d'expression différentes, j'ai pu observer que le nombre d'identification de protéines myristoylées n'est pas corrélé au taux d'expression de la NMT1 dans ces modèles. D'autre part, le nombre et le type d'identification s'est montré cellule-dépendant, montrant bien que ce ne sont pas les mêmes voies de transduction des signaux qui sont affectées en fonction du type de cancer. Mon analyse démontre l'intérêt de mieux étudier le myristoylome et de disposer d'outils adaptés permettant – comme c'est l'objectif avec lpaJ – de l'étudier en profondeur et d'en observer les variations selon le type cellulaire ou le tissu. Les cibles de la myristoylation étant des protéines majeures des voies de transduction (tyrosines kinases de la famille SRC par exemple) et leur myristoylation étant essentielle à leur activité, l'identification de voies dérégulées par l'intermédiaire de la caractérisation de la myristoylation est une approche qui peut paraître pertinente.

Cependant, même si on identifie jusqu'à 16 protéines myristoylées dans les cellules RT4, cette valeur reste encore plus faible par rapport à ce qui a été identifié lors des études utilisant la chimie « click » (50-80 identification), ou encore des études sur des tissus comme le foie ou le cerveau (Castrec *et al*, 2018). Cela peut aussi s'expliquer par le fait que la technique de chimie « click » repose sur l'utilisation d'un analogue réactif du myristate. L'apport en acides gras des cellules repose sur la nutrition et les acides gras sont véhiculés dans le sang par l'albumine. En culture cellulaire, c'est l'ajout d'albumine de veau fœtal (l'albumine lie naturellement les acides gras, avec 7 sites distincts) dans le milieu nutritif qui permet de le supplémenter en ce type de molécule essentielle. Pour réaliser cette chimie, on procède à un marquage dit métabolique en complétant le milieu par l'analogue réactif (dérivé azoture ou alcyne d'un myristate, voir **Figure I-5**). Pour garantir le marquage, les études utilisent un milieu de

culture caractérisé par une albumine préalablement appauvrie en acide gras et dont la seule source provient de l'intermédiaire réactif du myristate. Comme plusieurs études rapportent que l'apport de myristate est limitant – en particulier lors de certains cancers - il est donc possible lors de mon étude que le myristate était limitant alors que dans les études en chimie « click », son apport ne l'était plus. Ceci favoriserait l'identification d'un plus grand nombre de cibles myristoylées mais biaiserait leur validation physiologique. Lors de mon analyse, j'ai ainsi pu observer un nombre important de peptides internes de protéines annotées comme myristoylées, sans que je puisse obtenir les spectres de masses permettant de valider la myristoylation de leur peptide N-terminal. Cependant, je n'ai pas identifié non plus ces peptides sous forme non modifiée. Ceci démontre bien l'une des grandes difficultés que l'on rencontre dans l'étude de cette modification, qui est l'obtention de la preuve formelle de la présence de la modification sur la protéine. En effet, cette modification, très hydrophobe, rend difficile l'extraction, la purification mais surtout l'analyse.

V.3. Mise en situation d'IpaJ pour l'étude du myristoylome des glycines N-terminales

Les données rapportées ci-dessus mettent en relief tout l'intérêt de l'approche consistant à tirer profit des propriétés catalytiques remarquables d'IpaJ. L'utilisation d'IpaJ permet donc de passer outre les difficultés en retirant la glycine myristoylée. Le retrait de cette glycine myristoylée à deux avantages ; i) retirer le groupement hydrophobe rendant l'analyse difficile, ii) réaliser un marquage spécifique pour l'identification de la protéine après analyse, en recherchant dans l'échantillon la perte de la glycine qui caractérise l'activité de la protéase. J'ai ensuite dans un premier temps traité les lysats ou les fractionnements par IpaJ selon la stratégie n°1 (**Figure II-9,a**) et cherché les preuves de son activité, c'est à dire la perte de glycine. Cependant, aucune preuve de l'activité de cette protéase n'a pu être détectée lors de cette analyse. Ceci peut être expliqué par le fait les protéines myristoylées, une fois extraites des membranes où elles sont le plus souvent localisées pourraient subir un changement

de conformation enfouissant le groupement myristate et éviter un contact avec le milieu aqueux comme c'est le cas de certaines protéines caractérisé par le « myristoyl-switch ». C'est le cas des petites protéines G de la famille ARF dont la forme GTP seule expose le myristate, alors que la forme GTP internalise l'acide gras, le rend accessible à IpaJ et empêche sa coupure sous cette forme.

Dans une approche alternative (stratégie n°2), j'ai pu démontrer qu'IpaJ est cependant bien active sur des mélanges complexes de peptides ; j'ai en effet détecté des preuves de son activité sur des lysats cellulaires hydrolysés au préalable (au contraire de la stratégie n°1) par la trypsine à la première étape (**Figure II-10**). Dans le cadre de cette stratégie, après action de la trypsine, j'ai préalablement à l'utilisation d'IpaJ acétylé chimiquement avec un acétate « lourd » toutes les extrémités N-terminales, naturelles et néo-formées après la trypsine. Après action d'IpaJ, ce marquage permet de reconnaître de façon spécifique les N-terminaux libres issus du clivage des peptides myristoylés par IpaJ. En effet, ce sont les seuls à posséder une extrémité amino réactive, toutes les autres étant N-acétylées chimiquement ou naturellement. Cette singularité permet théoriquement d'isoler ces peptides tous marqueurs de myristoylation et d'augmenter fortement la concentration de ces peptides dans un échantillon complexe. Pour ce faire une méthode de « pêche » des peptides présentant ces étiquettes réactives est nécessaire. J'ai opté pour l'utilisation d'un marquage chimique reposant sur la réactivité du succinimide et la possibilité de transférer différents types de groupements sur les amines libres. Dans un premier temps, j'ai choisi le marqueur TMPP (**Figure II-12,a**) dont je ne suis pas parvenu à identifier les conditions idéales de marquage ; cet échec peut s'expliquer soit par la présence d'un produit dans les tampons conduisant à une inhibition de la réaction, soit au niveau de l'identification des peptides marqués. En effet, il a été rapporté que ce type de marquage change radicalement les profils de fragmentation peptidique, il se peut donc que les résultats obtenus sur des peptides de l'albumine soient sous-estimés. Mon approche nécessite cependant une forte efficacité de marquage car les protéines myristoylées que je recherche sont faiblement exprimées dans la cellule. Il me reste

plus que cette dernière étape à optimiser pour finaliser cette étude et enfin pouvoir estimer et évaluer le rôle de la myristoylation dans les différents types de cellules cancéreuses à ma disposition. L'objectif est d'observer les protéines myristoylées de faible expression puisque ce sont ces dernières qui ne sont jamais observées par les approches traditionnelles dites de « shotgun » que j'ai pratiquées (**Figure II-7** et **Sup Tableau II-1**) ou celles tirant profit de marquage métabolique et de chimie « click ».

V.4. Le mécanisme réactionnel de la NMT révèle son rôle dans la myristoylation des lysines

Afin de mieux comprendre la spécificité de la myristoylation, nous avons également réalisé une étude du mécanisme catalytique de la NMT1 humaine. La NMT est dotée d'un mécanisme réactionnel sophistiqué impliquant le mouvement de deux domaines conservés en N- et C-terminal de la protéine. Lorsque l'on observe différentes structures de protéines de la famille GNAT - la grande famille d'acyl-transférase utilisant un donneur le plus souvent lié au CoA dont la NMT fait partie - en complexe avec différents ligands ou inhibiteurs, on peut observer certaines caractéristiques conservées commune avec la NMT (**Figure III-1a-c** et **III-14a, b**). En effet, le groupement carbonyle de certaines enzymes GNAT est susceptible de servir de plateforme catalytique, permettant de placer l'ammonium accepteur à distance et géométrie idéales par rapport au lien thioester du donneur CoA. On retrouve également la présence d'un réseau de molécules d'eau assurant la déportation du groupe ammonium lorsque AcCoA et accepteur sont tous deux dans leur site respectif. Cette information est importante car elle pourrait permettre d'identifier la base catalytique dans différentes enzymes de cette famille GNAT. De même, les boucles $\alpha1\alpha2$ des GNATs - équivalentes à la boucle Ab de la NMT - semblent aussi requises pour la sélectivité du ligand accepteur.

Grâce à des structures tridimensionnelles à haute résolutions de la NMT en présence des substrats, produits et même d'intermédiaire de réaction, nous avons pu déterminer toutes les étapes de la réaction de G-myristoylation et mieux comprendre la sélectivité

de substrat de cette enzyme. Cette information est indispensable pour l'élaboration de nouveau inhibiteur de cette famille d'enzyme et plus particulièrement de la NMT. En effet la NMT est maintenant établie comme une cible de choix pour différentes maladies comme les pathologie parasitaires et fongiques ou les cancers, pour lesquels la myristoylation est connue comme indispensable. Ces résultats ouvrent de nouvelles possibilités de création d'inhibiteurs utilisant la plasticité de la NMT. Les inhibiteurs actuels des NMT humaine ou parasitaire, bien que très puissants ne sont encore pas assez sélectifs pour une NMT donnée, en raison de la forte homologie de séquence et de conformation du centre actif. Ces inhibiteurs sont dans la majorité des cas des molécules interférant avec le substrat protéique. C'est le cas de l'IMP-1088 qui occupe les poches spécifiques des acides aminés 2 à 5 (**Figure supplémentaire III-14c-e, article 1**). Bien que les acides aminés de la NMT impliqués dans les interactions avec les molécules de substrat et d'inhibiteurs soient conservés (**Figure III-2a supplémentaire, article 1**), une comparaison de la NMT en complexe avec un substrat et de la NMT avec l'inhibiteur suggère que la création d'inhibiteurs plus sélectif est possible. Ainsi, en remodelant l'interaction avec le groupe carboxyle C-terminal du Gln496 pour maintenir une molécule d'eau en position +1 et repositionner une amine primaire pour réagir avec MyrCoA permettant la génération d'un inhibiteur myristoylé. De plus, il est intéressant de constater que l'examen de la structure intermédiaire fermée de la boucle Ab a pu révéler que les quatre résidus formant l'hélice N-terminale à partir de Pro105 dans HsNMT1 ou Leu20 de la NMT de *Plasmodium falciparum* - l'agent causal du paludisme - de la boucle B'A' (voir l'encadré magenta dans la **Figure III-14f supplémentaire, article 1**) ne sont pas conservés chez les parasites (**Figure III-2a supplémentaire, article 1**). Comme nous avons pu le voir lors de cette étude, l'hélice B'A' est critique pour la catalyse. Cet aspect n'a pas encore exploité pour la conception des inhibiteurs et pourrait permettre de concevoir des inhibiteurs spécifiques à la NMT du plasmodium.

V.5. IpaJ est un outil unique pour distinguer les deux types de myristoylation

L'étude de ce mécanisme a mis en évidence un deuxième mécanisme insoupçonné de la NMT, en effet j'ai démontré que l'activité de la NMT ne se limite pas à la myristoylation sur des glycines N-terminales (G-myristoylation) mais permet d'autres types de myristoylation sur des amines de chaînes latérales de lysine par exemple (K-myristoylation). J'ai étudié les règles qui régissent cette nouvelle activité de la NMT. J'ai pu observer que la G-myristoylation est favorisée par rapport à la K-myristoylation si les deux amines sont disponibles sur un même peptide. Le pKa de l'acide conjugué du groupe alpha amino de la glycine est très inférieur à celui de l'amine epsilon de la chaîne latérale de la lysine ($7,7 \pm 0,5$ vs $10,5 \pm 1,1$) mais proche du pH physiologique et du pH du dosage réalisé au cours de l'étude. A ces pH, les deux amines sont protonées et le mécanisme de la réaction indique clairement que la déprotonation est la première étape que permet la base catalytique, le carboxy-terminus de l'enzyme.

Le fait que la G-myristoylation soit favorisée lorsque les deux amines acceptrices sont en compétition peut suggérer que lorsque les deux groupes sont disponibles sur le même polypeptide N-terminal, la liaison du groupement ϵ -aminé entièrement protoné est préférée à celle de l'amine α dans l'environnement chargé négativement de la poche aa3 (Castrec *et al*, 2018). Cela placerait l'amine de la glycine dans une géométrie plus favorable à la catalyse de la réaction que celle de la lysine. Ce déséquilibre entre ces deux myristoylations peut aussi s'expliquer par le fait que, par rapport à l'axe autour duquel le carbone α de l'acide aminé 3 de la lysine peut pivoter, l'amine ϵ de la lysine est plus éloignée d'une liaison que celle de l'amine α de la glycine, et elle s'intègre moins bien dans le site actif. On observe ainsi que le lien est direct avec la base catalytique alors qu'elle est médiée par une molécule d'eau pour la glycine. Malgré cela j'ai pu démontrer que les principales règles de spécificité de la K- et G-myristoylation se chevauchent fortement, conduisant à une compétition entre les deux types de myristoylation dans les contextes de protéines commençant par GK. Il existe une

quinzaine de substrats connus de ce type de modification, parmi lesquels la peptide protéine G ARF6. Dans ce contexte, il est remarquable de constater que l'équipe du Pr H. Lin est parvenu en parallèle de notre étude à démontrer qu'ARF6 subit une double myristoylation en G et en K (Kosciuk *et al*, 2020).

Lors de la caractérisation de la myristoylation en lysine et du constat de la compétition potentielle, j'ai rapidement compris qu'il me fallait me doter d'un outil adapté pour conclure sur le type de modification effectuée. J'ai mis au point un protocole permettant de discriminer les deux types de myristoylation. J'ai mis à profit ma maîtrise de la protéase IpaJ dont j'ai montré qu'elle est spécifique de la glycine myristoylée. Les règles de myristoylation sont résumées dans les **Figures IV-1e et IV-2f**. La forme très étroite de la poche numéro deux du site actif de la NMT assure la réactivité de l'amine primaire réactif porté uniquement sur des chaînes non ramifiées telles que celles de la glycine N-terminale ou de la chaîne latérale de la lysine. Ceci facilite la rotation de 70° du groupe amino-réactif de la base catalytique vers le groupement thioester du MyrCoA.

V.6. La protéase IpaJ permet générer de nouveaux sites de K-myristoylation : le cas d'ARF6

Au cours de mes caractérisations, j'ai mesuré les efficacités de réaction de plusieurs dizaines de substrats potentiels des deux réactions. J'ai observé que la K-myristoylation par la NMT est toujours moins efficace que la G-myristoylation. Cette plus faible réactivité repose sur des modes de reconnaissance identiques qui impose des contraintes plus fortes à la K-myristoylation. Ainsi, la différence majeure est que la K-myristoylation nécessite une reconnaissance de substrat beaucoup plus stricte au niveau des acides aminés 6-7. En effet contrairement à la G-myristoylation qui se caractérise par un champ de substrat plus large est permet la myristoylation de substrats de séquences assez diversifiées, la K-myristoylation nécessite le châssis restreint KXXSK à condition d'être associé à un acide aminé N-terminal autre que la glycine ou une N-acétylation de l'acide aminé N-terminal. Dans le contexte cellulaire,

la création de telles extrémités ne peut dépendre que d'une machinerie protéolytique post-traductionnelle. En effet, les MetAPs les enzymes cotraductionnelles qui retirent les méthionines initiatrices des protéines (**Figure I-1a**) n'agissent que si les acides aminés en position deux possèdent des chaînes latérales peu encombrantes comme celles des A, G, P, S, T ou V. Les protéines commençant par MK mais ne sont pas substrats de la MetAP et j'ai montré que ce contexte, même associé à la présence de SK en positions 6 et 7 est très peu favorable en présence d'une méthionine N-terminale à la K-myristoylation de la lysine deux. J'ai montré que la K-myristoylation peut être catalysée lorsque la lysine se situe non seulement en position 2 mais aussi en position 1 ou 3. Ainsi, les protéomes possèdent un certain nombre important de motifs de K-myristoylation KXXSK internes aux protéines. Dans le protéome humain, certains sont précédés par des sites de clivage de caspases. Ceci suggère que la K-myristoylation, contrairement à la G-myristoylation, pourrait ne résulter que d'un mécanisme post-traductionnel. Ceci justifierait l'existence des formes courtes de NMT. Ce mécanisme pourrait donc être impliqué dans des signaux d'apoptose issus de clivages par des caspases.

Un autre évènement peut conduire à des protéines commençant par une lysine. En effet, au cours de l'infection par *Shigella*, la protéase IpaJ permet de retirer la glycine myristoylée et générer des protéines commençantes par une lysine. De plus quand on observe les substrats de la NMT une dizaine de substrats présentent un contexte GKXXSK, comme c'est le cas d'ARF6. Le groupe du Pr N. Alto (Burnaevskiy *et al*, 2013, 2015), a démontré que, parmi les protéines ARF (5 isoformes ARF1/3/4/5/6 chez l'homme), seule ARF6 était insensible à la protéase IpaJ lors de l'infection. Les auteurs ont proposé que la localisation particulière d'ARF6 serait à l'origine de cette résistance ; ARF6 est en effet la seule ARF à ne pas être localisée au niveau du système du réticulum endoplasmique/appareil de Golgi. ARF6 se localise en effet sur les endosomes et la membrane plasmique, afin y jouer un rôle de régulation majeur de l'endocytose (Donaldson, 2003; Donaldson & Jackson, 2011). Je propose une explication alternative à cette observation. En effet, les auteurs ont utilisé la chimie « click » pour marquer les

peptides myristoylés avec la biotine puis les ont immobilisés sur des colonnes d'affinité streptavidine.

Les protéines retenues sur la colonne d'affinité azido-biotine avant et après le clivage IpaJ ont été analysées. Ainsi les auteurs ont pu identifier quelles protéines n'étaient plus retenues par la colonne (car démyristoylées après clivage par IpaJ), et constater qu'ARF6 faisait toujours partie des protéines retenues par la colonne. L'élution de la colonne étant réalisée avec de la trypsine, l'identification des protéines retenues a été obtenu uniquement pour les peptides internes et non les N-terminaux. Nos données indiquent qu'une fois démyristoylée, ARF6 présente une lysine N-terminale dans un contexte lui permettant de subir encore la K-myristoylation et donc d'être toujours retenue par la colonne. La résistance apparente d'ARF6 à IpaJ qu'a constaté l'équipe du Pr. N. Alto pourrait donc être le fruit de cette activité cachée. Ainsi, le démasquage de la lysine trois lors du clivage de la glycine myristoylée conduirait à une seconde myristoylation au niveau de la lysine. Avec cette nouvelle K-myristoylation ARF6 serait retenue sur la colonne d'affinité streptavidine. Mon hypothèse est donc qu'ARF6 est bien substrat d'IpaJ mais qu'elle subit une seconde myristoylation qui compense l'action protéolytique d'IpaJ. ARF6 résisterait donc bien à IpaJ mais de manière très indirecte. J'ai noté aussi que, bien que les efficacités catalytiques de K-myristoylation soient plus faibles de celle de la G-myristoylation des peptides identiques, les valeurs restent dans la fourchette basse des valeurs rapportées pour la G-myristoylation de l'ensemble des substrats pour lesquels ces valeurs ont été déterminées. La K-myristoylation *in cellulo* est donc très probable *in vivo*. En fin de thèse, je me suis donc naturellement orienté vers la recherche de preuves *in vivo* de la K-myristoylation. Dans ce contexte, j'ai axé ma stratégie sur la protéine ARF6 au cours de l'infection par *Shigella*. Une preuve de l'existence cette K-myristoylation démontrerait que la cellule dispose d'un dispositif de sauvegarde permettant de contourner en partie le caractère délétère de l'action d'IpaJ. Ces recherches n'ont pas encore abouti mais restent prometteuses.

V.7. Conclusions

Mes approches valident l'utilisation d'IpaJ pour l'étude de la myristoylation, qu'elle soit en glycine ou en lysine. Mes études se sont focalisées sur des modifications affectant essentiellement le N-terminus des protéines. Une question qui reste ouverte concerne les K-myristoylations internes qui ont été décrites, comme dans le cas du TNF- α qui possède deux sites. Mes données indiquent qu'il est fort peu probable que la NMT soit responsable de ces K-myristoylations. Ces modifications sont connues comme instables et peuvent être retirées par une famille d'enzymes récemment caractérisée et connues sous le nom de sirtuines (Jiang *et al*, 2018). Malgré des recherches répétées dans les différents protéomes que j'ai analysés en spectrométrie de masse, je ne les ai jamais identifiées. L'utilisation d'inhibiteurs de sirtuines pourrait s'avérer un outil supplémentaire pour la découverte de nouveaux sites de myristoylation. Je devrai vérifier que ces inhibiteurs n'interfèrent pas avec l'activité d'IpaJ. Si c'est le cas, les deux outils combinés devraient contribuer significativement à l'objectif que je m'étais fixé d'étendre les connaissances et la caractérisation de protéines myristoylées. Il va de soi, comme je l'ai montré, que ces études doivent être réalisées sur différents types cellulaires ou tissus. Les conditions de culture et l'apport exogène en myristate devront certainement faire partie des conditions de croissance à contrôler.

VI. Matériels et méthodes

La grande majorité du matériels et méthodes étant disponible dans l'article Rivière, et al 2021, je ne développerai que les parties non détaillées dans cet article.

VI.1. Culture cellulaire

Les cellules ont été cultivées avec les milieux suivants ; RPMI (A2780R, A549, MCF7, NCI-N87, PC3), Mc Coy's 5A (HCT116, RT4, SK-OV3), DMEM (JIMT-1, MIA-Paca2, MRC5, U87) complémenté avec 10% SVF +/- antibiotique (Zéocine et Blasticidine).

VI.2. Fractionnement cellulaire

Cette expérience est basée sur le protocole établi (Cox & Emili, 2006), en remplaçant le triton X-100 par 0.01% de RapiGest™ (Waters) dissous dans 10% méthanol (recommandation du fabricant et (Min *et al*, 2015)). Les inhibiteurs de cystéines protéases ont été également retirés pour ne pas inhiber l'activité d'IpaJ. Le matériel cellulaire de départ correspond à une flasque de culture de 10 ml dont la croissance est proche de la confluence.

VI.3. Digestion échantillons sur gel dénaturants « SDS-PAGE »

La séparation des protéines est réalisée par migration des échantillons sur gel dénaturant de polyacrylamide de type SDS-PAGE (12%). Chaque piste est découpée en six portions, elles-mêmes redécoupé en cubes de 1 mm³. Des lavages successifs sont réalisés sur les cubes de gel afin de retirer le bleu de Coomassie et extraire les peptides après la digestion par la trypsine. Pour cela, il faut placer les bandes de gels découpées dans 10 % d'éthanol. Cette étape est suivie de la déshydratation des bandes dans 20-100 µl d'acétonitrile à 100 % pendant 10 minutes ; le liquide est ensuite retiré. Cette étape peut être répétée si la bande de gel nécessite une décoloration plus importante. La réduction des protéines est obtenue en ajoutant 20-100 µl de dithiothreitol (DTT) 10mM dans une solution de bicarbonate d'ammonium 50mM frais suivie d'une incubation à 37°C pendant 1 heure en agitant de temps en temps mais vigoureusement. Le liquide est ensuite jeté. L'alkylation est réalisée en ajoutant le même volume que l'étape précédente d'une solution d'iodoacétamide 55 mM dans du

bicarbonate d'ammonium 50 mM, cette étape est suivie d'une incubation à l'obscurité pendant 30 à 60 min avec des agitations occasionnelles mais vigoureuses. Le liquide est alors retiré et jeté. La déshydratation des morceaux de gel se fait dans 20-100 µl d'acétonitrile à 100 % pendant 10 minutes avant de retirer le surnageant. L'étape de réhydratation des morceaux de gel est réalisée pendant 20 min en secouant occasionnellement mais vigoureusement dans 20-100 µl d'une solution de bicarbonate d'ammonium 50 mM. Pour finir, la déshydratation des morceaux de gel est réalisée dans 50 µl d'une solution d'acétonitrile 100% pendant 10 minutes, avant de retirer le surnageant. Le lavage des morceaux de gel est réalisé dans 50-200 µl d'une solution de bicarbonate d'ammonium 50 mM en secouant occasionnellement mais vigoureusement avant de retirer le surnageant. Il faut ensuite déshydrater les morceaux de gel dans 50-200 µl d'acétonitrile à 100 % pendant 10 minutes, puis jeter le liquide et retirer l'acétonitrile restant dans les morceaux de gel avec une centrifugeuse sous vide pendant 30 minutes à 1 heure (l'acétonitrile restant dans le gel pourrait inhiber la trypsine instantanément et nuire à votre digestion). Il faut ensuite diluer la trypsine (20 µg par flacon) dans 400 µl de solution de réhydratation (Promega) et agiter vigoureusement. La trypsine reconstituée peut être conservée à -20°C. On doit alors ajouter 2-20 µl ou 1/10 du volume de gel. A cette étape, le gel ne doit pas être complètement réhydraté. Il faut laisser le temps de se réhydrater à 0°C. Lorsque toute la solution de trypsine est adsorbée, 10-100 µl de bicarbonate d'ammonium 47,5 mM, 5% d'acétonitrile sont ajoutés aux morceaux de gel pour les réhydrater complètement. La digestion se fait pendant 1h à 37°C. Il faut alors ajouter 2-10 µl d'une solution à 1% FA (acide formique) ou 0,1% TFA (acide trifluoroacétique) dans l'eau pour arrêter la réaction de digestion. Pour l'extraction, on doit prélever le surnageant et le recueillir dans un tube avant d'ajouter 10-50 µl (selon le volume de gel) de 1% FA. Une incubation pendant 20-30 min à température ambiante est réalisée avec des secousses occasionnelles et vigoureuses. Le surnageant est recueilli puis évaporé à l'aide d'une centrifugeuse sous vide. L'échantillon est solubilisé dans 10-50 µl de solvant A (5 % acétonitrile, 0,1 % FA). Selon le volume de gel, le gel doit être correctement réhydraté

et un peu de liquide doit rester.

VI.4. Extraction des protéines

Les protéines des échantillons issus du fractionnement sont extraites par précipitation à l'acétone en ajoutant quatre fois le volume d'échantillon d'acétone froid (-20°C). Après agitation au vortex, le tube est incubé au moins 60 min ou sur la nuit à -20°C. La solution est ensuite centrifugée à 13000 g pendant 30 minutes au moins à 4°C (ou moins si possible). Le surnageant est éliminé et le culot est séché à l'air pour diminuer la quantité d'acétone restante. Cette étape est suivie par l'alkylation et réduction des protéines. Les échantillons de protéines sont dénaturés à l'aide de guanidine/HCl 6 M, 50 mM Tris-HCl pH 8, 4 mM DTT (1-10 mg/ml), puis réduites pendant 15 min à 95°C. Les protéines sont alkylées par l'ajout d'iodacétamide (55 mM) pendant 1 heure à température ambiante. Les protéines sont précipitées par l'acétone pour éliminer le sel et les réactifs en ajoutant quatre fois le volume d'échantillon d'acétone froide. Après vortex, le tube est incubé au moins 60 min à -20°C ou pendant la nuit. La solution est ensuite centrifugée à 13'000 x g pendant 10 minutes (4°C ou moins).

VI.5. Préparation du N-acétoxy-[²H₃] succinimide

Pour la préparation de l'N-acétoxy-[²H₃] succinimide il faut mélanger dans un premier temps 1 mL d'anhydride D3-acétique avec 373 mg de N-hydroxy-succinimide. NB ; il peut être difficile de solubiliser les cristaux de N-hydroxy-succinimide, mais il est possible de chauffer modérément le tube pour aider à la solubilisation des cristaux. Le mélange est incubé alors à température ambiante pendant 12 à 15 heures. Les cristaux sont retirés du surnageant puis séchés dans une centrifugeuse à vide (réactif dit de 1^{ère} qualité). La partie liquide est également séchée avec une centrifugeuse à vide mais celle-ci peut également contenir du N-hydroxy-succinimide n'ayant pas réagi (réactif dit de 2^{ème} qualité). Les cristaux sont lavés deux fois avec 200 µl d'hexane et séchés à nouveau. Le N-acétoxy-[²H₃] succinimide est conservé à -20°C.

VI.6. Acétylation N-terminal des protéines ou des peptides

L'acétylation des protéines ou des peptides se fait avec une solution d'N-acétoxy-[²H₃] succinimide (32 mg (200 µmol) de N-acétoxy-[²H₃] succinimide remis en suspension

dans 100 μ L de DMSO (solution à 2 μ mol/ μ L). Un excès molaire de 50 fois (qui représente en moyenne 25 μ mol de réactif par mg de protéine dans l'échantillon) de solution de N-acétoxy- $^{[2]}\text{H}_3$ succinimide est ajouté à l'échantillon (en tampon phosphate 50 mM à pH 7,5). La réaction est effectuée pendant 90 min à 30°C, après quoi l'O-acétylation des chaînes latérales Ser, Thr et Tyr a été inversée en ajoutant 10 μ L de 50% d'hydroxylamine à la solution de protéine et incubation pendant 20 min à température ambiante. Cette réaction est très sensible au temps ; 20 minutes précises évitent la désacétylation du groupe amino tout en permettant d'éliminer l'acétylation artificielle sur les groupes hydroxyles. Les protéines/peptides sont précipitées par l'acétone comme décrit précédemment. Le culot est séché à l'aide du concentrateur pour éliminer un maximum de l'acétone restant. Il s'agit d'une étape critique pour empêcher la dénaturation.

VI.7. Digestion liquide des protéines par la trypsine

Le culot obtenu lors de l'étape d'acétylation précédente est remis en suspension dans du bicarbonate d'ammonium 50 mM. Les protéines sont hydrolysées en présence de 1/100ème (p/p) de trypsine porcine modifiée TPCK (poudre conservée au congélateur pour remise en suspension dans HCl 10⁻³ (pH = 3) ou solution mère au congélateur à 1 mg/ml) pendant 1-1,5 heures à 37°C deux fois. L'échantillon est séché dans une centrifugeuse sous vide (« SpeedVac ») avant la prochaine étape.

VI.8. Chromatographie en phase inversée avec cartouches Sep-Pak

Cette étape est cruciale pour la préparation de l'échantillon pour la chromatographie liquide avant l'étape de spectrométrie de masse. Il faut réhydrater dans un premier temps la phase stationnaire avec cinq à dix volumes phase avec de l'acétonitrile à 80 %, 0,1 % de TFA, par exemple 5 à 10 ml pour une cartouche de 1 ml. Il faut ensuite rincer la cartouche avec cinq à dix fois le volume de rétention avec une solution de 0,1 % de TFA diluée dans de l'eau sans laisser sécher la cartouche. L'échantillon, qui aura été préalablement séché en centrifugeuse à vide lors de l'étape précédente, est resuspendu alors dans un tampon TFA 0,1% avant de le charger dans la colonne. L'élution des composants indésirables est réalisée avec une solution de 0,1% de TFA

dilué dans l'eau (au moins 10 fois le volume de la colonne). L'élution des peptides est réalisée avec 3 fois le volume de colonne avec une solution composée de 80% ACN, 0,1% TFA. Les fractions collectées (regroupées) sont séchées à l'aide d'une centrifugeuse à vide.

VI.9. Digestion des échantillons par IpaJ

Les échantillons de nature protéique ou peptidique (10 μ M) sont incubés en présence d'IpaJ (concentration variable) pendant une heure à 20°C.

VI.10. Marquage des peptides de BSA avec TMPP

Dans un premier temps, j'ai cherché à caractériser l'influence du type de tampon utilisé lors de la réaction de marquage (Hepes VS Tris) et l'influence du glycérol provenant de l'ajout d'IpaJ sur les échantillons (IpaJ est stocké en 55% glycérol) sur la réaction. La réaction a été réalisée dans 4 conditions différentes : - 50mM Tris pH 8.2 - 50mM Hepes pH 8.2 - 50mM Tris pH 8.2, 27,5% glycérol - 50mM Hepes pH 8.2, 27,5% glycérol. Le marquage a été réalisé selon le protocole standard, avec de l'albumine bovine sérique (BSA) à une concentration de 110 fmol/ μ L dans un volume de 100 μ L. La réaction est bloquée avec 10 μ L d'hydroxylamine pendant 40 minutes. L'analyse est réalisée par spectrométrie de masse LC-MS en injectant 22 fmol ; l'élution est réalisée avec deux gradients différents : le gradient optimisé pour le marquage TMPP (Vaca Jacome *et al*, 2015) et un gradient optimisé pour les peptides fortement hydrophobe du type myristoylée (Castrec *et al*, 2018).

VII. Références

- Aitken A, Cohen P, Santikarn S, Williams DH, Calder AG, Smith A & Klee CB (1982) Identification of the NH₂-terminal blocking group of calcineurin B as myristic acid. *FEBS Lett.* **150**: 314–318
- Aksnes H, Drazic A, Marie M & Arnesen T (2016) First Things First: Vital Protein Marks by N-Terminal Acetyltransferases. *Trends Biochem. Sci.* **41**: 746–760 Available at: <http://dx.doi.org/10.1016/j.tibs.2016.07.005>
- Bertaccini D, Vaca S, Carapito C, Arsène-Ploetze F, Van Dorsselaer A & Schaeffer-Reiss C (2013) An improved stable isotope N-terminal labeling approach with light/heavy TMPP to automate proteogenomics data validation: DN-TOP. *J. Proteome Res.* **12**: 3063–3070
- Bhandarkar SS, Bromberg J, Carrillo C, Selvakumar P, Sharma RK, Perry BN, Govindarajan B, Fried L, Sohn A, Reddy K & Arbiser JL (2008) Tris (dibenzylideneacetone) dipalladium, a N-myristoyltransferase-1 inhibitor, is effective against melanoma growth in vitro and in vivo. *Clin. Cancer Res.* **14**: 5743–5748
- Bhatnagar RS, Ashrafi K, Fütterer K, Waksman G & Gordon JI (2001) 9 Biology and enzymology of protein N-myristoylation. *Enzymes* **21**: 241–290
- Bhatnagar RS, Fütterer K, Farazi TA, Korolev S, Murray CL, Jackson-Machelski E, Gokel GW, Gordon JI & Waksman G (1998) Structure of N-myristoyltransferase with bound myristoylCoA and peptide substrate analogs. *Nat. Struct. Biol.* **5**: 1091–1097
- Boutin JA (1997) Myristoylation. *Cell. Signal.* **9**: 15–35
- Burnaevskiy N, Fox TG, Plymire D a, Ertelt JM, Weigele B a, Selyunin AS, Way SS, Patrie SM & Alto NM (2013) Proteolytic elimination of N-myristoyl modifications by the Shigella virulence factor IpaJ. *Nature* **496**: 106–9 Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3722872&tool=pmcentrez&rendertype=abstract>
- Burnaevskiy N, Peng T, Reddick LE, Hang HC & Alto NM (2015) Myristoylome

- profiling reveals a concerted mechanism of ARF GTPase deacylation by the bacterial protease IpaJ. *Mol. Cell* **58**: 110–122 Available at: <http://dx.doi.org/10.1016/j.molcel.2015.01.040>
- Cao J, Sun L, Aramsangtienchai P, Spiegelman NA, Zhang X, Huang W, Seto E & Lin H (2019) HDAC11 regulates type I interferon signaling through defatty-acylation of SHMT2. *Proc. Natl. Acad. Sci. U. S. A.* **116**: 5487–5492
- Carr SA, Biemann K & Shoji S (1982) n-Tetradecanoyl is the NH₂-terminal blocking group of the catalytic subunit of cyclic AMP-dependent protein kinase from bovine cardiac muscle. *Proc. Natl. Acad. Sci. U. S. A.* **79**: 6128–6131
- Castrec B, Dian C, Ciccone S, Ebert CL, Bienvenut W V, Le Caer JP, Steyaert JM, Giglione C & Meinnel T (2018) Structural and genomic decoding of human and plant myristoylomes reveals a definitive recognition pattern. *Nat. Chem. Biol.* **14**: 671–679 Available at: <http://dx.doi.org/10.1038/s41589-018-0077-5>
- Corbic Ramljak I, Stanger J, Real-Hohn A, Dreier D, Wimmer L, Redlberger-Fritz M, Fischl W, Klingel K, Mihovilovic MD, Blaas D & Kowalski H (2018) Cellular N-myristoyltransferases play a crucial picornavirus genus-specific role in viral assembly, virion maturation, and infectivity
- Cox B & Emili A (2006) Tissue subcellular fractionation and protein extraction for use in mass-spectrometry-based proteomics. *Nat. Protoc.* **1**: 1872–1878
- Dian C, Pérez-Dorado I, Rivi re F, Asensio T, Legrand P, Ritzefeld M, Shen M, Cota E, Meinnel T, Tate EW & Giglione C (2020) High-resolution snapshots of human N-myristoyltransferase in action illuminate a mechanism promoting N-terminal Lys and Gly myristoylation. *Nat. Commun.* **11**: 1–15 Available at: <http://dx.doi.org/10.1038/s41467-020-14847-3>
- Diller JR, Parrington HM, Patton JT & Ogden KM (2019) Rotavirus Species B Encodes a Functional Fusion-Associated Small Transmembrane Protein. *J. Virol.* **93**:
- Donaldson JG (2003) Multiple Roles for Arf6: Sorting, Structuring, and Signaling at the Plasma Membrane. *J. Biol. Chem.* **278**: 41573–41576
- Donaldson JG & Jackson CL (2011) ARF family G proteins and their regulators: Roles

- in membrane transport, development and disease. *Nat. Rev. Mol. Cell Biol.* **12**: 362–375 Available at: <http://dx.doi.org/10.1038/nrm3117>
- Dowen RH, Engel JL, Shao F, Ecker JR & Dixon JE (2009) A family of bacterial cysteine protease type III effectors utilizes acylation-dependent and -independent strategies to localize to plasma membranes. *J. Biol. Chem.* **284**: 15867–15879
- Ducker CE, Upson JJ, French KJ & Smith CD (2005) Two N-myristoyltransferase isozymes play unique roles in protein myristoylation, proliferation, and apoptosis. *Mol. Cancer Res.* **3**: 463–476
- Duronio RJ, Jackson-Machelski E, Heuckeroth RO, Olins PO, Devine CS, Yonemoto W, Slice LW, Taylor SS & Gordon JI (1990) Protein N-myristoylation in *Escherichia coli*: Reconstitution of a eukaryotic protein modification in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **87**: 1506–1510
- Farazi TA, Waksman G & Gordon JI (2001) The Biology and Enzymology of Protein N-Myristoylation. *J. Biol. Chem.* **276**: 39501–39504 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11527981> [Accessed May 4, 2017]
- Gallien S, Perrodou E, Carapito C, Deshayes C, Reytrat JM, Van Dorsselaer A, Poch O, Schaeffer C & Lecompte O (2009) Ortho-proteogenomics: Multiple proteomes investigation through orthology and a new MS-based protocol. *Genome Res.* **19**: 128–135
- Giglione C, Boularot A & Meinnel T (2004) Protein N-terminal methionine excision. *Cell. Mol. Life Sci.* **61**: 1455–1474 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15197470> [Accessed May 4, 2017]
- Glover CJ, Hartman KD & Felsted RL (1997) Human N-Myristoyltransferase amino-terminal domain involved in targeting the enzyme to the ribosomal subcellular fraction. *J. Biol. Chem.* **272**: 28680–28689
- Greaves J, Munro KR, Davidson SC, Riviere M, Wojno J, Smith TK, Tomkinson NCO & Chamberlain LH (2017) Molecular basis of fatty acid selectivity in the zDHHC family of S-acyltransferases revealed by click chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **114**: E1365–E1374

- Greene NP, Crow A, Hughes C & Koronakis V (2015) Structure of a bacterial toxin-activating acyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **112**: E3058–E3066
- Gunaratne RS, Sajid M, Ling IT, Tripathi R, Pachebat JA & Holder AA (2000) Characterization of N-myristoyltransferase from plasmodium falciparum. *Biochem. J.* **348**: 459–463
- Ishii S, Yano T, Ebihara A, Okamoto A, Manzoku M & Hayashi H (2010) Crystal structure of the peptidase domain of Streptococcus ComA, a bifunctional ATP-binding cassette transporter involved in the quorum-sensing pathway. *J. Biol. Chem.* **285**: 10777–10785
- Jiang H, Zhang X, Chen X, Aramsangtienchai P, Tong Z & Lin H (2018) Protein Lipidation: Occurrence, Mechanisms, Biological Functions, and Enabling Technologies. *Chem. Rev.* **118**: 919–988
- Johnson D, Bhatnagar RS, Knoll LJ & Gordon JI (1994) Genetic and biochemical studies of protein N-myristoylation. *Annu. Rev. Biochem.* **63**: 869–914 Available at: <http://www.annualreviews.org/doi/10.1146/annurev.bi.63.070194.004253> [Accessed May 4, 2017]
- Kosciuk T, Price IR, Zhang X, Zhu C, Johnson KN, Zhang S, Halaby SL, Komaniecki GP, Yang M, DeHart CJ, Thomas PM, Kelleher NL, Christopher Fromme J & Lin H (2020) NMT1 and NMT2 are lysine myristoyltransferases regulating the ARF6 GTPase cycle. *Nat. Commun.* **11**: Available at: <http://dx.doi.org/10.1038/s41467-020-14893-x>
- Kramer G, Boehringer D, Ban N & Bukau B (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* **16**: 589–597
- Lim KB, Bazemore Walker CR, Guo L, Pellett S, Shabanowitz J, Hunt DF, Hewlett EL, Ludwig A, Goebel W, Welch RA & Hacketta M (2000) Escherichia coli α -hemolysin (HlyA) is heterogeneously acylated in vivo with 14-, 15-, and 17-carbon fatty acids. *J. Biol. Chem.* **275**: 36698–36702
- Majeran W, Le Caer JP, Ponnala L, Meinnel T & Giglione C (2018) Targeted profiling of

- Arabidopsis thaliana* subproteomes illuminates co- and posttranslationally N-terminal myristoylated proteins. *Plant Cell* **30**: 543–562
- Martinez A, Traverso JA, Valot B, Ferro M, Espagne C, Ephritikhine G, Zivy M, Giglione C & Meinel T (2008) Extent of N-terminal modifications in cytosolic proteins from eukaryotes. *Proteomics* **8**: 2809–2831 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18655050> [Accessed May 11, 2017]
- Maurer-Stroh S & Eisenhaber F (2004a) Myristoylation of viral and bacterial proteins. *Trends Microbiol.* **12**: 178–185 Available at: <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=23543785&retmode=ref&cmd=prlinks%5Cnpapers3://publication/doi/10.1105/tpc.112.106849>
- Maurer-Stroh S & Eisenhaber F (2004b) Myristoylation of viral and bacterial proteins. *Trends Microbiol.* **12**: 178–185 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15051068> [Accessed May 11, 2017]
- McLaughlin S & Aderem A (1995) The myristoyl-electrostatic switch: a modulator of reversible protein-membrane interactions. *Trends Biochem. Sci.* **20**: 272–276 Available at: <http://www.sciencedirect.com/science/article/pii/S0968000400890428%5Chttp://www.sciencedirect.com/science/article/pii/S0968000400890428/pdf?md5=40ce66e4f40f90a9f834ee07c644f580&pid=1-s2.0-S0968000400890428-main.pdf>
- Mei Y, Wang Y, Hu T, Yang X, Lozano-Duran R, Sunter G & Zhou X (2018) Nucleocytoplasmic Shuttling of Geminivirus C4 Protein Mediated by Phosphorylation and Myristoylation Is Critical for Viral Pathogenicity. *Mol. Plant* **11**: 1466–1481 Available at: <https://doi.org/10.1016/j.molp.2018.10.004>
- Meinel T, Dian C & Giglione C (2020) Myristoylation, an Ancient Protein Modification Mirroring Eukaryogenesis and Evolution. *Trends Biochem. Sci.* **45**: 619–632 Available at: <https://doi.org/10.1016/j.tibs.2020.03.007>
- Min L, Choe LH & Lee KH (2015) Improved protease digestion conditions for membrane protein detection. *Electrophoresis* **36**: 1690–1698

- Mousnier A, Bell AS, Swieboda DP, Morales-Sanfrutos J, Pérez-Dorado I, Brannigan JA, Newman J, Ritzeveld M, Hutton JA, Guedán A, Asfor AS, Robinson SW, Hopkins-Navratilova I, Wilkinson AJ, Johnston SL, Leatherbarrow RJ, Tuthill TJ, Solari R & Tate EW (2018) Fragment-derived inhibitors of human N-myristoyltransferase block capsid assembly and replication of the common cold virus. *Nat. Chem.* **10**: 599–606
- Nimchuk Z, Marois E, Kjemtrup S, Leister RT, Katagiri F & Dangl JL (2000) Eukaryotic fatty acylation drives plasma membrane targeting and enhances function of several type III effector proteins from *Pseudomonas syringae*. *Cell* **101**: 353–363
- Panethymitaki C, Bowyer PW, Price HP, Leatherbarrow RJ, Brown KA & Smith DF (2006) Characterization and selective inhibition of myristoyl-CoA:protein N-myristoyltransferase from *Trypanosoma brucei* and *Leishmania major*. *Biochem. J.* **396**: 277–285
- Price HP, Menon MR, Panethymitaki C, Goulding D, McKean PG & Smith DF (2003) Myristoyl-CoA:Protein N-myristoyltransferase, an essential enzyme and potential drug target in kinetoplastid parasites. *J. Biol. Chem.* **278**: 7206–7214 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12488459> [Accessed May 11, 2017]
- Röttig A & Steinbüchel A (2013) Acyltransferases in Bacteria. *Microbiol. Mol. Biol. Rev.* **77**: 277–321
- Salah Ud-Din AIMS, Tikhomirova A & Roujeinikova A (2016) Structure and functional diversity of GCN5-related n-acetyltransferases (GNAT). *Int. J. Mol. Sci.* **17**:
- Schrager JA & Marsh JW (1999) HIV-1 Nef increases T cell activation in a stimulus-dependent manner. *Proc. Natl. Acad. Sci. U. S. A.* **96**: 8167–8172
- Selvakumar P, Lakshmikuttyamma A, Shrivastav A, Das SB, Dimmock JR & Sharma RK (2007) Potential role of N-myristoyltransferase in cancer. *Prog. Lipid Res.* **46**: 1–36
- Sengupta P, Seo AY, Pasolli HA, Song YE, Johnson MC & Lippincott-Schwartz J (2019) A lipid-based partitioning mechanism for selective incorporation of proteins into membranes of HIV particles. *Nat. Cell Biol.* **21**: 452–461 Available at: <http://dx.doi.org/10.1038/s41556-019-0300-y>

- Shao F, Golstein C, Ade J, Stoutemyer M, Dixon JE & Innes RW (2003) Cleavage of Arabidopsis PBS1 by a bacterial type III effector. *Science (80-.)*. **301**: 1230–1233
- Slagowski NL, Kramer RW, Morrison MF, LaBaer J & Lesser CF (2008) A functional genomic yeast screen to identify pathogenic bacterial proteins. *PLoS Pathog.* **4**: 0096–0107
- Stanley P, Koronakis V & Hughes C (1998) Acylation of Escherichia coli Hemolysin: A Unique Protein Lipidation Mechanism Underlying Toxin Function. *Microbiol. Mol. Biol. Rev.* **62**: 309–333
- Stevenson FT, Bursten SL, Fanton C, Locksley RM & Lovett DH (1993) The 31-kDa precursor of interleukin 1 α is myristoylated on specific lysines within the 16-kDa N-terminal propiece. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 7245–7249
- Stevenson FT, Bursten SL, Locksley RM & Lovett DH (1992) Myristyl acylation of the tumor necrosis factor α precursor on specific lysine residues. *J. Exp. Med.* **176**: 1053–1062
- Suzuki T, Moriya K, Nagatoshi K, Ota Y, Ezure T, Ando E, Tsunasawa S & Utsumi T (2010) Strategy for comprehensive identification of human N-myristoylated proteins using an insect cell-free protein synthesis system. *Proteomics* **10**: 1780–1793
- Thinon E, Serwa RA, Broncel M, Brannigan JA, Brassat U, Wright MH, Heal WP, Wilkinson AJ, Mann DJ & Tate EW (2014) Global profiling of co- and post-translationally N-myristoylated proteomes in human cells. *Nat. Commun.* **5**: 1–13
Available at: <http://dx.doi.org/10.1038/ncomms5919>
- Traverso JA, Giglione C & Meinnel T (2013a) High-throughput profiling of N-myristoylation substrate specificity across species including pathogens. *Proteomics* **13**: 25–36
- Traverso JA, Micalella C, Martinez A, Brown SC, Satiat-Jeunemaître B, Meinnel T & Giglione C (2013b) Roles of N-terminal fatty acid acylations in membrane compartment partitioning: Arabidopsis h-type thioredoxins as a case study. *Plant Cell* **25**: 1056–1077

- Vaca Jacome AS, Rabilloud T, Schaeffer-Reiss C, Rompais M, Ayoub D, Lane L, Bairoch A, Van Dorsselaer A & Carapito C (2015) N-terminome analysis of the human mitochondrial proteome. *Proteomics* **15**: 2519–2524
- Vetting MW, Luiz LP, Yu M, Hegde SS, Magnet S, Roderick SL & Blanchard JS (2005) Structure and functions of the GNAT superfamily of acetyltransferases. *Arch. Biochem. Biophys.* **433**: 212–226
- Wright MH, Heal WP, Mann DJ & Tate EW (2010) Protein myristoylation in health and disease. *J. Chem. Biol.* **3**: 19–35
- Zha J, Weiler S, Oh KJ, Wei MC & Korsmeyer SJ (2000) Posttranslational N-myristoylation of BID as a molecular switch for targeting mitochondria and apoptosis. *Science (80-.).* **290**: 1761–1765

VIII. Annexes

VIII.1. Publication Dian et al,

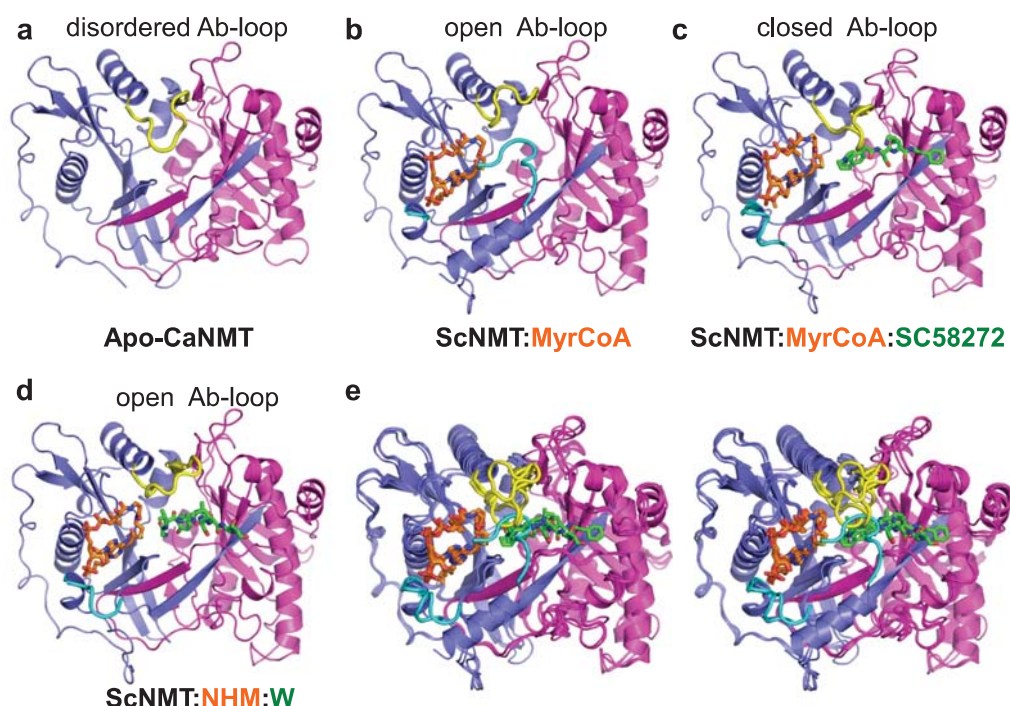
VIII.1.1. Figures supplémentaires

SUPPLEMENTARY INFORMATION

High-resolution snapshots of *N*-myristoyltransferase in action illuminate a mechanism promoting N-terminal Lys and Gly myristoylation

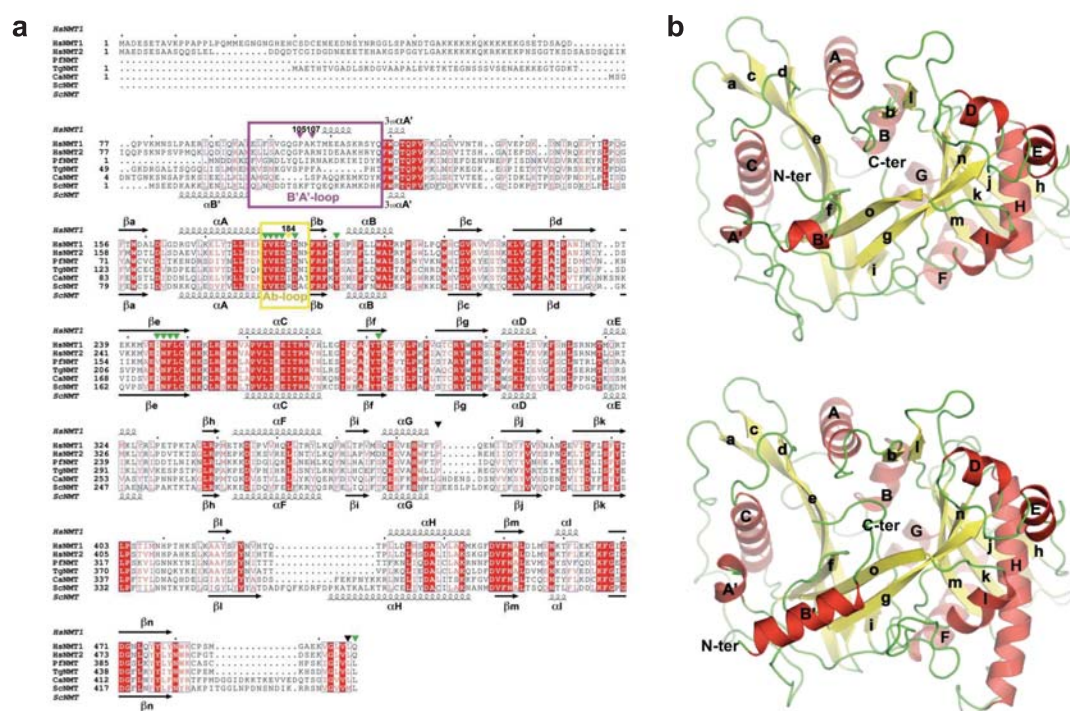
C. Dian, I. Pérez-Dorado et al.

Supplementary Figures

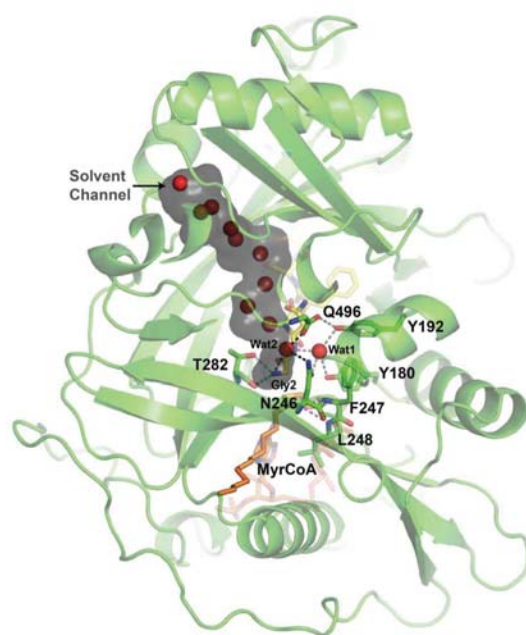


Supplementary Figure 1. Structures used to build the previously-accepted NMT catalytic mechanism

Overview of previously reported NMT crystal structures revealing the different conformations of both the B'A'- and Ab-loops. NMT main chains are displayed as ribbons. GNAT subdomains of NMTs are coloured in blue and pink, respectively. B'A' and Ab-loops are coloured in cyan and yellow, respectively. MyrCoA or NHM moieties are shown as orange sticks. Peptide substrate W and peptidomimetic inhibitor SC58272 are shown as green sticks. **a)** Crystal structure of apo CaNMT (PDB 1NMT, chain B)¹. **b)** Crystal structure of ScNMT:MyrCoA binary complex (PDB 26PE, chain A)². **c)** Crystal structure of ScNMT:MyrCoA:SC58272 ternary complex (PDB 2NMT)³. **d)** Crystal structure of ScNMT:MyrCoA:peptide W ternary complex (PDB 1IID)⁴. **e)** Stereoview of superimposed crystal structures shown in a, b, c and d.

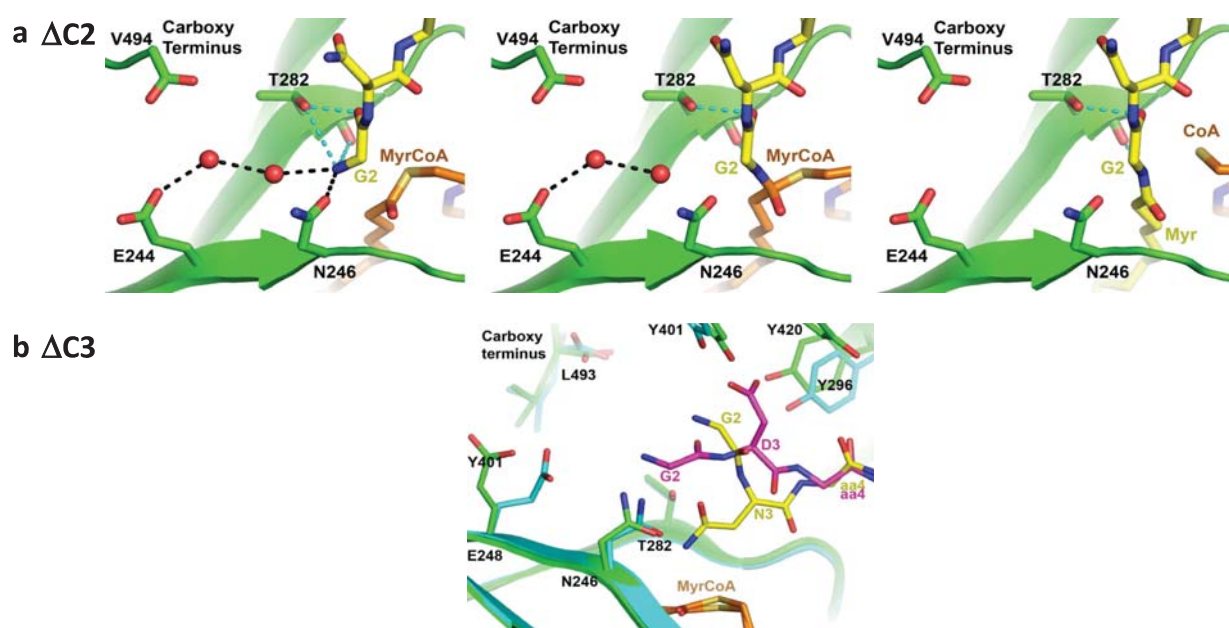


Supplementary Figure 2. Structural conservation of complexes reported here with published structures
a) Sequence alignment of NMTs with the corresponding secondary structure of HsNMT1:MyrCoA:X (top) and ScNMT:MyrCoA crystal structures (bottom, pdb code 1IID)⁴ obtained with ESPrnt3 (<http://esprnt3.ibcp.fr>). Magenta and yellow boxes indicate the positions of the B'A'- and Ab- loops, respectively. Residues identified in this study as important for the MYR mechanism are indicated with triangles. Green triangles define conserved residues, whereas yellow and magenta triangles denotes non-conserved ones, using B'A'- and Ab-loops colour coding. **b)** View of HsNMT1:MyrCoA:X (top) and ScNMT:MyrCoA (bottom)⁴ folds labelled with secondary structure elements according to **a)**. Both crystal structures are displayed as ribbons. a-helix, b-strand, and loop are coloured in red, yellow and green, respectively.



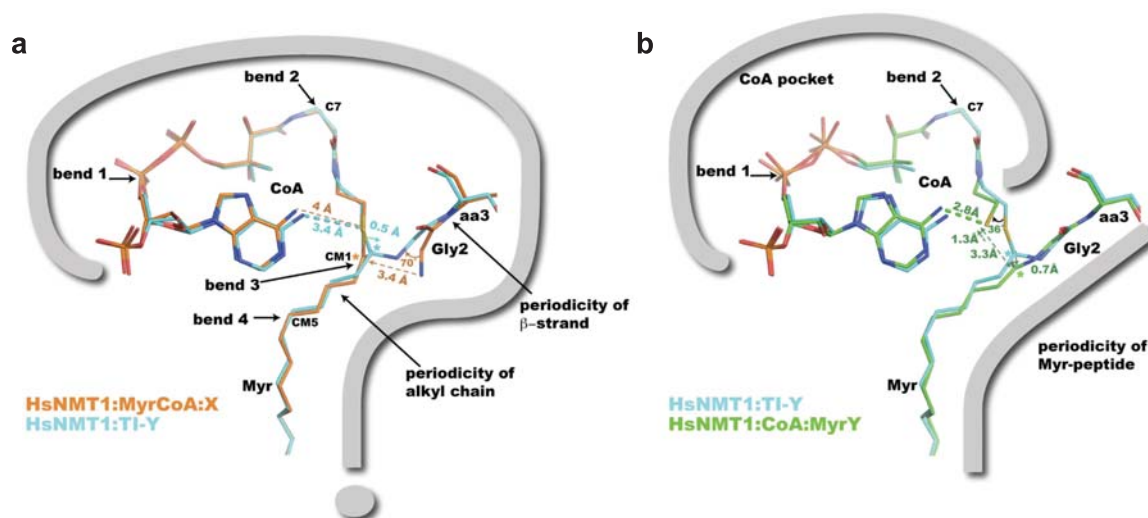
Supplementary Figure 3. Space filling cartoon representation of the water channel

View of the specific channel feeding the active site of HsNMT1 with water molecules in the HsNMT1:MyrCoA:X crystal structure. The solvent channel surface is shown as a grey surface. The HsNMT1 chain is displayed as a green ribbon, and selected amino acids are shown as green sticks. Substrate peptides (yellow) and MyrCoA and CoA moieties (both in orange) are shown as sticks. Hydrogen bond set1, linking Gly2 to Thr282 catalyst platform, are shown as cyan dashes. Wat1-mediated hydrogen bond set2, linking the carboxy terminus of Q496 to the hydroxyl groups of Tyr180/Tyr192 from the Ab-loop, amide group of substrate aa3, and amide of Asn246 are shown as grey dashes. Wat1-mediated hydrogen bond set3, linking the carboxy terminus of Gln496 to Gly2 ammonium, is shown as black dashes. Water molecules from the solvent channel are shown as red spheres. Hydrogen bonds involved in the "oxanion hole" are displayed as a magenta dashed line. The H-bond between thiolate and amine from CoA is shown with violet dashes.



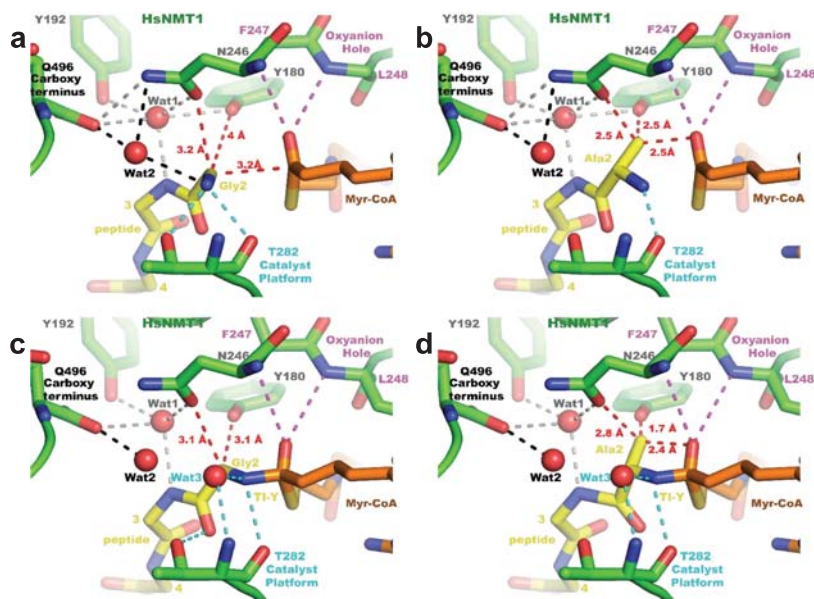
Supplementary Figure 4. Crystal structures of short C-terminal truncations of HsNMT1

View of C-terminal HsNMT1 truncations' active site. **a**) Snapshots of the MYR reaction trapped in crystal structure of HsNMT1 Δ C2 (green) in complex with MyrCoA (orange) and peptide S (GlyAsnCysPheSerLysProArg; yellow) showing: substrates (left; conformation A and chain A), T1 (middle; conformation B and chain A or conformation A and chain B), and products (right; conformation B and chain B) of the reaction. **b**) Superimposition of HsNMT1 Δ C3 (cyan) in complex with MyrCoA (orange) and peptide S (GlyAsnCysPheSerLysProArg) with HsNMT1 Δ C2 (green) in complex with MyrCoA (orange) and peptide T (GlyAspCysPheSerLysProArg) revealing a different conformation of the N-terminal extremity of substrate peptides in the enlarged cavity. The HsNMT1 chain is displayed as a ribbon, and selected amino acids are shown as sticks. Substrate peptides are shown as sticks. MyrCoA and CoA moieties shown as orange sticks. Hydrogen bond set1, linking Gly2 to T282 catalyst platform, is shown as cyan dashes. Wat-mediated hydrogen bond set linking the carboxylic group of E244 to Gly2 ammonium are shown as black dashes.



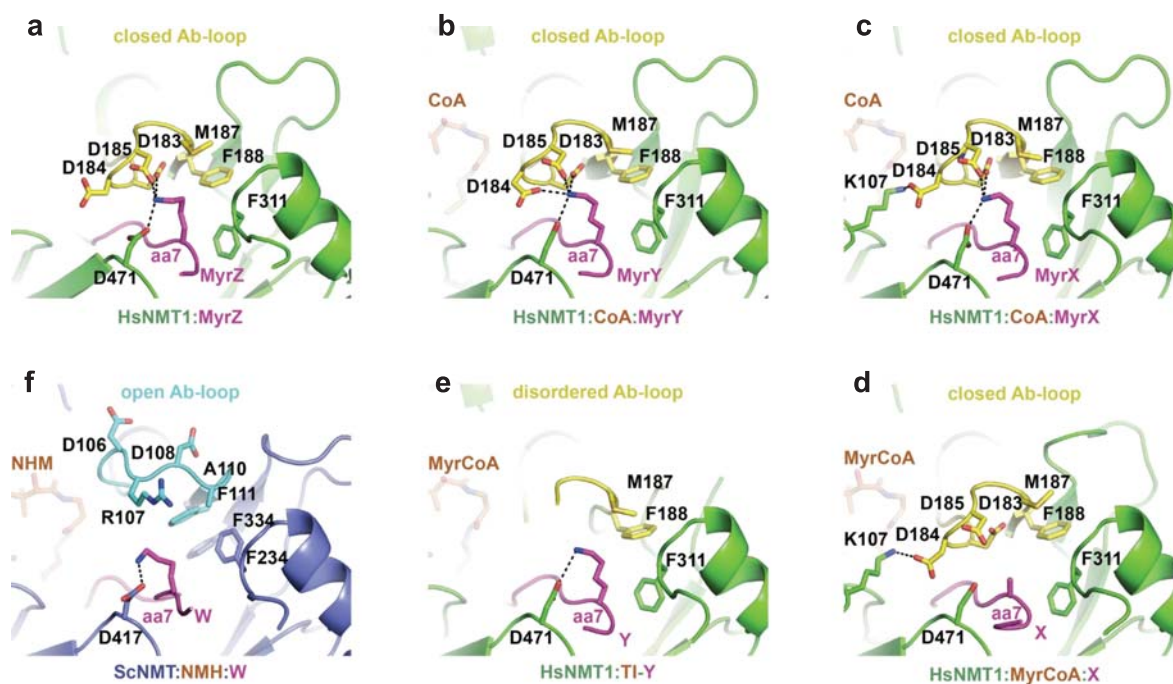
Supplementary Figure 5. Substrates and products changes along the catalytic reaction

a) Superimposition of HsNMT1:MyrCoA:X (orange carbons) and HsNMT1:TI-Y (cyan carbons) focusing on both substrate peptide and MyrCoA moieties reveals the 70° rotation of the N-terminal amino group from substrate peptide Gly2 towards MyrCoA CM1 carbonyl and the induced displacement of the latter to produce the tetrahedral intermediate. Carbon CM1 from the MyrCoA carbonyl is marked with a star. Bound MyrCoA presents a question mark-like shape with four previously reported bends bringing tension to the MyrCoA structure³. **b)** Superimposition of HsNMT1:TI-Y (cyan carbons) and HsNMT1:CoA:MyrY (cyan carbons) focused on TI-Y and CoA moieties reveals two motions at the level of the carbon of the newly formed amide group in the myristoylated product: (i) a back motion of this carbon in the amide plane that allows proper positioning of the carbonyl oxygen within the oxyanion hole, and (ii) a 0.7 Å shift of this electrophilic carbon away from the sulphur atom of CoA induced by the peptide β -strand conformation. This suppresses the third bend involved in the “question mark” shape of MyrCoA at its binding site. The newly formed thiolate anion moves 1.3 Å away from the amide bond of MyrY due to van der Waals repulsion with the newly formed N-C bond. This shift places the thiolate at an optimal distance to form an intramolecular H-bond with the amine group of CoA. However, despite rearrangements occurring upon breakdown of the reaction intermediate, the polarized carbonyl of the Myr-Gly2 amide group is 3.3 Å away from the leaving thiolate anion of CoA, a distance sufficient to lead to reformation of the TI via a 36° rotation around C2P/C3P bond of CoA.



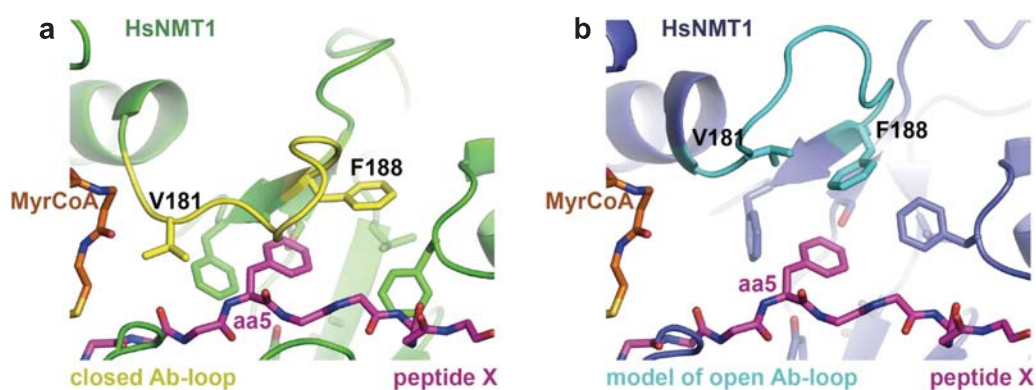
Supplementary Figure 6. Gly to Ala substitution prevents amino terminal rotation due to steric clashes

View of the HsNMT1 active site highlighting the Gly selectivity at position 2 of the substrate peptide constrained by both Asn246 and Tyr180 side chains together with the thioester bound **a)** substrates in HsNMT1:MyrCoA:X, **b)** substrates in HsNMT1:MyrCoA:X, in which Gly2 was replaced by Ala, **c)** the transition state in HsNMT1:TI-Y and **d)** the transition state in HsNMT1:TI-Y, in which Gly2 was replaced by Ala. The HsNMT1 chain is displayed as a green ribbon, and selected amino acids are shown as green sticks. Substrate peptides are shown as yellow sticks. MyrCoA and CoA moieties shown as orange sticks. Distances between either Ca of Gly2 or Cb of modelled Ala2 with both Asn246 and Tyr180 side chains and MyrCoA thioester are shown as red dashes. Hydrogen bond set1, linking Gly2 to the T282 catalyst platform, is shown as cyan dashes. Wat1-mediated hydrogen bond set2, linking the carboxy terminus of Q496 to hydroxyl groups of Tyr180/Tyr192 from the Ab-loop, amide group of substrate aa3, and amide of Asn246 are shown as grey dashes. Wat1-mediated hydrogen bond set3 linking the carboxy terminus of Gln496 to Gly2 ammonium are shown as black dashes. Water molecules from the solvent channel are shown as red spheres. Hydrogen bonds involved in the "oxyanion hole" are displayed as a magenta dashed line. The H-bond between thiolate and amine from CoA is shown with violet dashes.



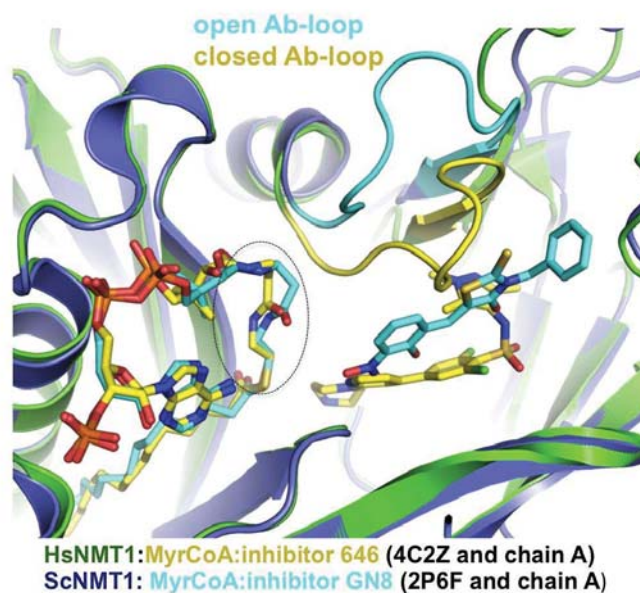
Supplementary Figure 7. Link between Ab-loop and aa7 substrate position

Outside view of the peptide substrate cavity allowing the accommodation of a Lys7 from the substrate peptide in the function of Ab-loop conformation. **a)** Products in HsNMT1:MyrZ, **b)** products in HsNMT1:CoA:MyrX, **c)** products in HsNMT1:CoA:MyrX, **d)** substrates in HsNMT1:MyrCoA:X, **e)** transition state in HsNMT1:TI-Y and, **f)** substrates in HsNMT1:NHM:W⁴. The HsNMT1 (green) and ScNMT (blue) main chains are in ribbons, and selected residues are in sticks. Main chain and selected residues of HsNMT1 Ab-loop (yellow) and ScNMT Ab-loop (cyan) are displayed as ribbon and sticks respectively. Substrate and product peptides are displayed as magenta ribbon, and selected amino acids are shown as magenta sticks. MyrCoA, CoA and NHM moieties are shown as orange sticks. Salt bridges are shown as black dashes.



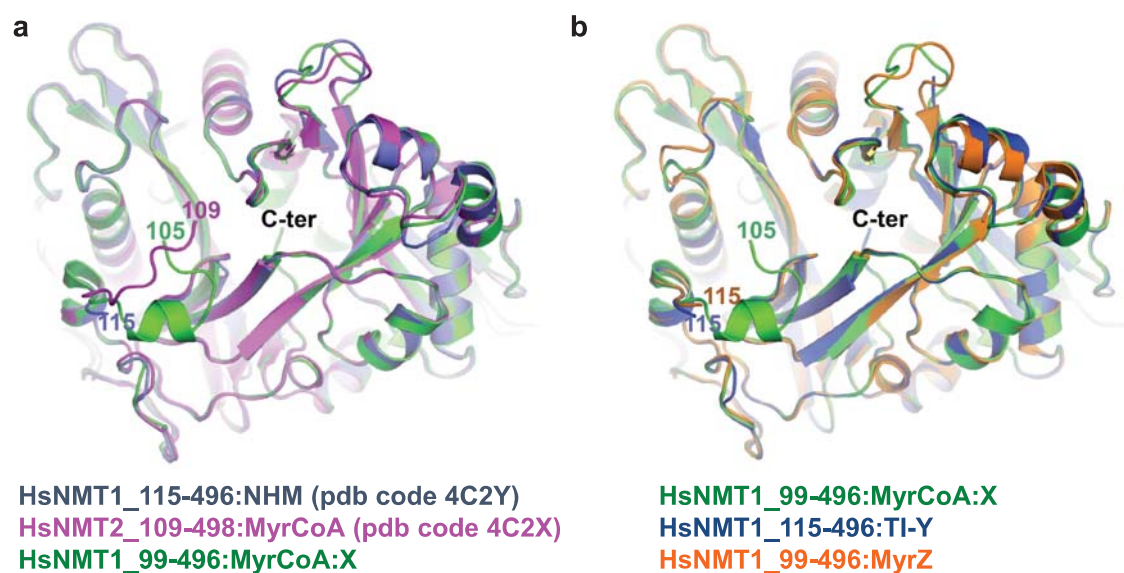
Supplementary Figure 8. Ab-loop conformation does not favour substrate aa5

Outside view of the peptide substrate cavity that allows accommodation of aa5 from the substrate peptide in the function of Ab-loop conformation. **a)** Substrates in HsNMT1:MyrCoA:X harbouring a closed conformation of the Ab-loop. The HsNMT1 chain is displayed as a green ribbon, and selected amino acids are shown as green sticks. The closed conformation of the Ab-loop is displayed as a yellow ribbon, and selected amino acids are shown as yellow sticks. **b)** Substrates in HsNMT1:MyrCoA:X, in which the Ab-loop was modelled in an open conformation based on the ScNMT:NHM:W crystal structure⁴. The HsNMT1 chain is displayed as a blue ribbon, and selected amino acids are shown as blue sticks. The model of the open conformation of the Ab-loop is displayed as a cyan ribbon, and selected amino acids are shown as cyan sticks.



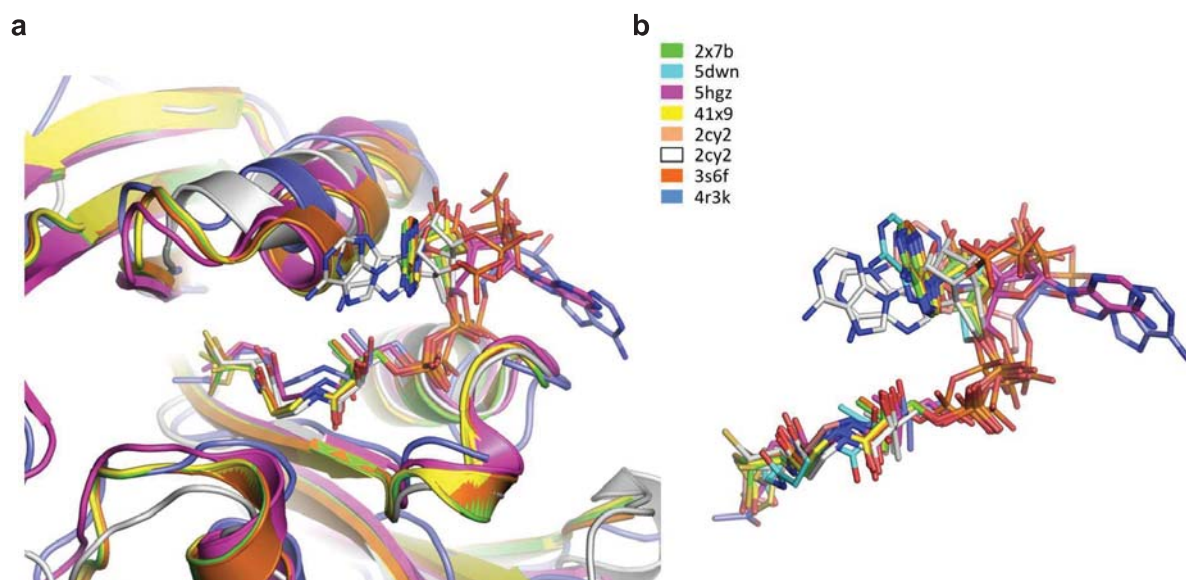
Supplementary Figure 9. NMT:MyrCoA:inhibitor complexes show that the MyrCoA compaction depends only on the Ab-loop conformation

Structural superimposition of the ternary ScNMT:MyrCoA:inhibitor GN8 complex (PDB 2P6F, blue)² and the ternary HsNMT1:MyrCoA:inhibitor 646 (PDB 4C2Z, green)⁶ complex shows that the Ab-loop closed conformation (yellow) observed in the latter induced MyrCoA compaction (black dashed line circle) in comparison to MyrCoA (cyan) from ScNMT:MyrCoA. GN8 and 646 inhibitors are shown in cyan and yellow, respectively.



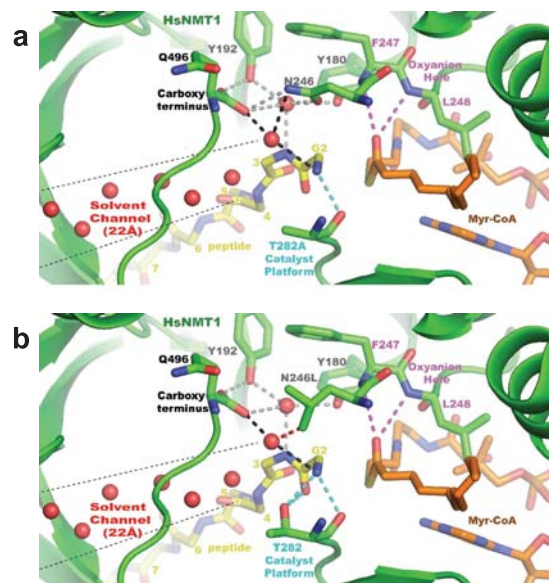
Supplementary Figure 10. Comparison of solved structures and previous structures of HsNMT

a) Structural superimposition of the previously solved binary HsNMT1:NHM complex (PDB 4C2Y, blue ribbon)⁶ and the binary HsNMT2:MyrCoA complex (PDB 4C2Y, magenta ribbon)⁶ with the ternary HsNMT1:MyrCoA:X complex solved in this study (green ribbon) shows the importance of residues 100-108 in the formation of a folded N-terminal B'A'-loop. **b)** Structural superimposition of the ternary HsNMT1:MyrCoA:X complex (green ribbon), ternary HsNMT1:TI-Y complex (blue ribbon), and binary HsNMT1:MyrZ complex (orange ribbon) shows that CoA release induced disordering of B'A'-loop.

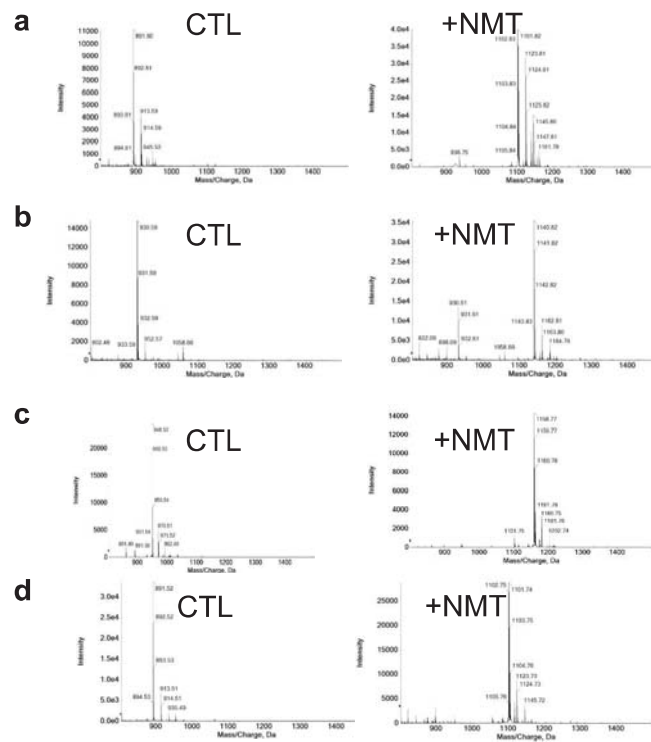


Supplementary Figure 11. Structural alignment of AcCoA/CoA-bound NAT structures suggesting that adenosine moiety flexibility could be important for CoA release from NMT active site

a) The different NATs structures were structurally aligned based on their conserved GNAT domains. Chains are displayed as ribbons with their bound AcCoA or CoA bound shown as sticks coloured according to their PDB codes. **b)** Focus on the CoA/AcCoA moieties reveals the flexibility of the adenine ring.

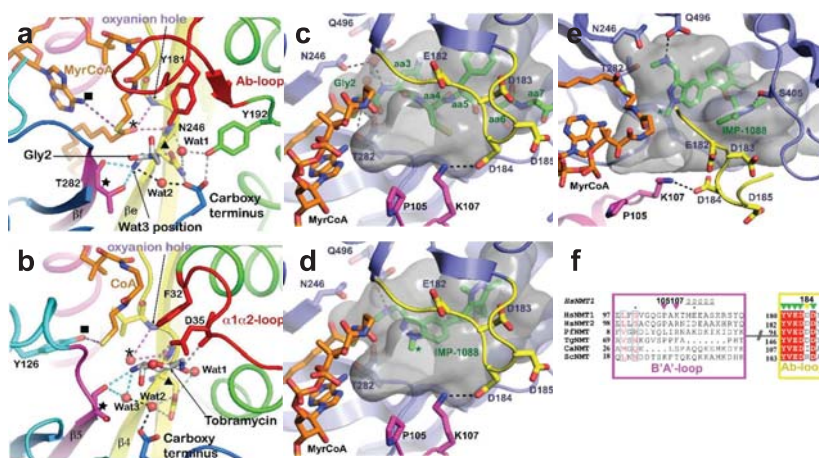


Supplementary Figure 12. Structural impact of either T282A or N246L substitution modelled on HsNMT1 active site. View of the HsNMT1 active site of the HsNMT1:MyrCoA:X ternary complex in which **a)** Thr282Ala substitution was modelled **b)** Asn246Leu substitution was modelled. In ScNMT, single amino acid mutants Thr205Ala (T282 in HsNMT1) and Asn169Leu (Asn246 in HsNMT1) targeting the catalysis platform and the first H-bond set 1 are consistent with our newly proposed catalytic mechanism⁴. Both two-fold k_{burst} and three-fold K_m decreases in ScNMT-Thr205Ala are in agreement with the moderate effect expected from the analogous Thr282Ala substitution, which was modelled with HsNMT1. This substitution indeed removes two out of three H-bonds made by Gly2 of X peptide. Similarly, the Asn248Leu substitution in HsNMT1 mimicking Asn169Leu in ScNMT reveals that this mutation would disturb both the water-mediated interaction of the carboxy terminus of Gly2 in peptide X and H-bond set 1, involved in optimal positioning of the C-terminus of the enzyme and peptide X in the catalytic centre. These effects cause both a 10-fold reduction of the k_{burst} and of the K_m value measured with ScNMT.



Supplementary Figure 13. Mass spectrometry analysis of NMT-catalysed MYR of N-Lys-containing peptides

Both control MS1 spectra of the four peptides described in Fig. 8 in the absence (left) or presence of NMT (right) are displayed as follows: **a)** GKVLSKIF (Arf6), **b)** GKQNSKLR (HPCA), **c)** AcGKSFSKPR, and **d)** AcKSFSKPR. The MYR mass shift is 210 Da.



Supplementary Figure 14. Transposition of the NMT mechanism to GNAT proteins and inhibitor selectivity.

Comparison of **a**) HsNMT1:MyrCoA:X and **b**) aminoglycoside 2'-N-acetyltransferase from *M. tuberculosis* (AgNAT) bound to CoA and the inhibitor tobramycin reveals conservation of the catalytic mechanism of HsNMT1 with other GNAT enzymes. The carbonyl of the residue positioned at the end of $\beta 5$ of the aminoglycoside 2'-N-acetyltransferase (black star) is likely to act as a catalytic platform, while the residue positioned at the end of $\beta 4$ of the same enzyme is likely to organize the wat1/wat2 network involved in ammonium group deprotonation. The oxyanion hole bulge and thioester carbonyl from AcCoA position are also conserved, as shown by the pink dashes and black star, respectively. The $\alpha 1\alpha 2$ -loop in the GNAT superfamily is the equivalent of the Ab-loop in HsNMT1 and is coloured in red. The hydroxyl group of the Tyr126 residue in AgNAT, suggested to act as an acid to reprotonate the thiolate anion, is placed at an equivalent position of the identified CoA amine group to stabilise the leaving thiolate group as shown by the violet dashes. Secondary structure colour code of the GNAT catalytic domain is as in **Figure 1b**. Peptide X and tobramycin are shown partially as grey sticks, for clarity. MyrCoA and CoA moieties are shown as orange sticks. Hydrogen bond set1, set2, and set3 identified in HsNMT1 and AgNAT are shown as cyan, grey, and dark dashes. Wat1, 2 and 3 molecules are shown as red spheres. A detailed comparison of the HsNMT1 active site around the B'A'-loop and Ab-loop interaction in **c**) HsNMT1:MyrCoA:X, **d**) and **e**) HsNMT1:MyrCoA:X, in which the X substrate peptide has been replaced by IMP-1088, showing that the IMP-1088 inhibitor binds the highly conserved peptide-binding site shown in two different views. Grey surface of the peptide-binding site shows that there is a space not occupied by the inhibitor close to the B'A'-loop and Ab-loop interaction. A green asterisk highlights the moiety of IMP-1088 that might be used for selectivity improvement. **f**) Sequence alignment of NMT proteins obtained with ESPript3 (<http://esprict.ibcp.fr>,⁵) with the corresponding secondary structure of the HsNMT1:MyrCoA:X, reveals that the B'A'-loop/Ab-loop interaction motif is not conserved among NMTs. Residues in white and red indicate identical and conserved residues, respectively. Magenta and yellow broken boxes indicate the residues participating in the B'A'-loop defined in the ScNMT:MyrCoA structure and Ab-loop as defined in the ScNMT and CaNMT crystal structures, respectively. Residues identified in this study important for the MYR mechanism are indicated with triangles. Green triangle defines conserved residues, whereas yellow and magenta triangles denote non-conserved ones using B'A' and Ab-loop colour coding.

Supplementary Tables

Supplementary Table 1. Data collection and refinement statistics concerning wild type NMT1 in complex with peptides X, Y, U and V.

PDB accession code	6QRM	6EHJ	6SJZ	6SK2
Crystal molecules	HsNMT1: MyrCoA:X	HsNMT1: MyrCoA:Y	HsNMT1: MyrCoA:U	HsNMT1: MyrCoA:V
Chain A ligands	MyrX	TI-Y/MyrY	U	MyrV
Chain B ligands	X	MyrY	U	V
Peptide sequence	GNCSFSKRRA	GSNKS KPK	AcGKSFSKPR	AcKSFSKPR
Space group	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Number of complexes in ASU	2	2	2	2
Unit cell (Å)	a= 79.44 b=178.93 c= 58.41	a=80.25 b=177.54 c=58.14	a= 79.91 b=178.90 c=58.38	a= 78.82 b= 178.44 c= 58.27
Wavelength (Å)	0.96770	0.97625	0.96600	0.98400
Beamline	ID30a3 (ESRF)	I03 (DLS)	ID30a1 (ESRF)	PX1 (SOLEIL)
Resolution range (Å)	47.77-2.30 (2.38-2.30)	47.08-2.1 (2.15-2.10)	48.89-1.99 (2.05-1.99)	48.79-1.90 (1.94-1.90)
R _{merge}	0.143 (0.669)	0.112 (0.968)	0.103 (0.968)	0.102 (0.941)
R _{pim}	0.051 (0.226)	0.047 (0.533)	0.038 (0.203)	0.028 (0.252)
Number of total reflections	348468 (33889)	330247 (14866)	461333 (26772)	930363 (61061)
Number of unique reflections	37748 (3619)	49357 (3595)	57674 (3937)	65710 (4177)
Mean (I /SD[I])	12.8 (3.6)	12.3 (1.3)	14.4 (4)	15.4 (3.2)
CC _{1/2}	0.996 (0.875)	0.998 (0.508)	0.998 (0.508)	0.999 (0.934)
Completeness (%)	99.7 (99.8)	99.8 (99.3)	99.5 (93.6)	99.9 (99.7)
Multiplicity	9.2 (9.4)	6.7 (4.1)	8 (6.8)	14.2 (14.6)
Wilson plot B (Å ²)	29.09	10.4	21.32	26.6
Number of atoms	7003	6609	7209	7411
Protein	6334	5982	6347	6422
Peptide	67		114	57
MyrCoA	63		126	63
MyrCoA-peptide		122		
Myr-peptide	69	148		76
CoA	48	96		48
Ions / Glycerol	2 Cl ⁻ , 24	24	2 Cl ⁻ , 24	0, 18
Water	396	237	596	727
R _{work} (%)	17.20	18.43	17.16	17.20
R _{free} (%)	20.33	23.86	20.16	22.00
 (Å ²)	30.7	37.83	27.1	31.6
rms bond length deviation (Å)	0.009	0.008	0.004	0.006
rmsd angle deviation (°)	0.989	0.979	0.810	0.858

Ac is for N-acetyl.

Values in parentheses correspond to the highest resolution shell.

$$R_{\text{merge}} = \Sigma(|I_{\text{hl}} - \langle I_{\text{h}} \rangle|) / \Sigma \langle I_{\text{h}} \rangle$$

$$R_{\text{meas}} = \Sigma \sqrt{(n_{\text{h}}/n_{\text{h}} - 1)} (|I_{\text{hl}} - \langle I_{\text{h}} \rangle|) / \Sigma \langle I_{\text{h}} \rangle$$

$$R_{\text{pim}} = \Sigma \sqrt{(1/n_{\text{h}} - 1)} (|I_{\text{hl}} - \langle I_{\text{h}} \rangle|) / \Sigma \langle I_{\text{h}} \rangle$$

Supplementary Table 2. Data collection and refinement statistics of NMT1 variants

PDB accession code	6SKJ	6SK8	6SK3
Crystal molecules	Δ C2:MyrCoA:S	Δ C3:MyrCoA:T	Δ C3:MyrCoA:S
Chain A	S/TI-S	T	S
Chain B	TI-S/Myr-S	T	S
Peptide sequence	GNCFSKPR	GDCFSKPR	GNCFSKPR
Space group	C121	C121	C121
Number of complexes in	2	2	2
Unit cell (Å)	$a=92.97$ $b=58.27$ $c=154.13$ $\beta=90.92^\circ$	$a=92.62$ $b=58.26$ $c=154.79$ $\beta=91.26^\circ$	$a=91.81$ $b=58.07$ $c=148.21$ $\beta=92.25^\circ$
Wavelength (Å)	0.978565	0.978565	0.978565
Beamline	PX1 (Soleil)	PX1 (Soleil)	PX1 (Soleil)
Resolution range (Å)	49.37-2.8 (2.95-2.80)	49.31-1.87 (1.91-	49.07-2.70 (2.83-
R_{merge}	0.122 (0.885)	0.159 (1.219)	0.171 (0.707)
R_{pim}	0.053 (0.393)	0.063 (0.467)	0.074 (0.322)
Number of total reflections	114615 (4601)	507596 (33696)	129478 (4104)
Number of unique	19085 (706)	68387 (4360)	21519 (630)
Mean ($ I /SD[I]$)	10.8 (2.0)	7.3 (1.6)	7.4 (2.0)
$CC_{1/2}$	0.997 (0.796)	0.994 (0.665)	0.994 (0.842)
Completeness (%)	92.5 (79.4)	99.9 (100)	99.1 (95.8)
Multiplicity	6.0 (5.4)	7.4 (7.7)	6.0 (5.0)
Wilson plot B (Å ²)	36.12	25.54	38.84
Number of atoms	6673	7493	6514
Protein	6286	6338	6234
Peptide	31	126	126
MyrCoA	63	126	126
MyrCoA-peptide	122		
Myr-peptide	38		
CoA	24		
Ions / Glycerol	1 Mg ²⁺ , 18	1 Cl ⁻ , 36	1 Mg ²⁺ , 12
Water	90	843	15
R_{work} (%)	18.90	20.78	24.59
R_{free} (%)	21.88	24.39	26.86
$\langle B \rangle$ (Å ²)	29.6	30.2	39
rms bond length deviation	0.003	0.003	0.004
rmsd angle deviation (°)	0.738	0.691	0.768

Values in parentheses correspond to the highest resolution shell.

$$R_{\text{merge}} = \sum (|I_{hl}| - \langle I_h \rangle) / \sum \langle I_h \rangle$$

$$R_{\text{meas}} = \sum \sqrt{(n_h/n_h - 1)} (|I_{hl}| - \langle I_h \rangle) / \sum \langle I_h \rangle$$

$$R_{\text{pim}} = \sum \sqrt{(1/n_h - 1)} (|I_{hl}| - \langle I_h \rangle) / \sum \langle I_h \rangle$$

Supplementary Table 3. Oligonucleotide primers used in this study.

Variant		Primer sequence
DeltaC2	For	5'-CGAGAATCTTTA TTTTCAGGGCGGATCCGAATTCTCAGTGGGTCAGGGACCT-3'
	Rev	5'-GGTGGTGCTCGAGTTATTGTTACACCAGTCCAACCTTCTCTGC-3'
DeltaC3	For	5'-CGAGAATCTTTA TTTTCAGGGCGGATCCGAATTCTCAGTGGGTCAGGGACCT-3'
	Rev	5'-GTGGTGCTCGAGTTATTGTAGTTACAGTCCAACCTTCTCTGCCC-3'
Y180A	For	5'-CTCCTGAATGAGAACGCTGTGGAAGATGATGA-3'
	Rev	5'-TCATCATCTTCCACAGCGTTCTCATT CAGGAG-3'
Y180P	For	5'-CACCTCCTGAATGAGAACCCGGTGGAAGATGATGACACCAT G -3'
	Rev	5'-GTTGTCATCATCTTCCACCGGGTTCTCATT CAGGAGGGTG-3'
V181A	For	5'-TGAATGAGAACTATGCGGAAGATGATGACAA-3'
	Rev	5'-TTGTCATCATCTTCCGCATAGTTCTCATTCA-3'
Y181L	For	5'-CTGAATGAGAACTATCTGGAAGATGATGACA-3'
	Rev	5'-TGTCATCATCTTCCAGATAGTTCTCATT CAG-3'
Y192A	For	5'-ATGTTCCGATTTGATGCTTCCCCGGAGTTTCT-3'
	Rev	5'-AGAAACTCCGGGGAAGCATCAAATCGGAACAT-3'
K107E	For	5'-GGTCAGGGACCTGCCGAAACCATGGAGGAGGC-3'
	Rev	5'-GCCTCCTCCATGGTTTCGGCAGGTCCCTGACC-3'
N246A	For	5'-GAAGAAGATGGTAGAGATCGCGCTTCTTGTGTGTCCACAAG-3'
	Rev	5'-CTTGTGGACACACAGGAAGGCGATCTCCACCATCTTCTTC-3'
K252E	For	5'-CAACTTCCTGTGTGTCCACGAGAAGCTGCGTTCCAAGAG-3'
	Rev	5'-CTCTTGGAACGCAGCTTCTCGTGGACACACAGGAAGTTG-3'

Supplementary Methods

Production of site-directed NMT1 variants

C-terminal DeltaC2 and DeltaC3 truncation of Long-HsNMT1 were generated by PCR using primers reported in **Supplementary Table 3**. The resulting PCR products were subcloned in the pET28 derivative plasmid using Fast digest BamH1 and XhoI restriction enzymes (Thermo Fisher Scientific, Waltham, MA). Point mutants of long-HsNMT1 were made using a QuikChange site-directed mutagenesis kit (Stratagene, San Diego, CA) using primers pairs displayed in **Supplementary Table 3**.

C-terminal truncation and points mutants of long-HsNMT1 were expressed, purified and crystallized following the same protocol as the WT protein. Truncations of Long-HsNMT1 were expressed and purified as previously described⁷. Suitable crystals of $\Delta C2:MyrCoA:S$, $\Delta C3:MyrCoA:S$ and $\Delta C3:MyrCoA:T$ were obtained by co-crystallization using the hanging-drop vapor diffusion method at 20°C. Crystallization droplets were formed by mixing 2 μ L of the of HsNMT1:MyrCoA:peptide complex (ratio 1:1.5:1.5) at 7.5 mg/mL (i.e., ~160/250/250 μ M) with 2 μ L of the precipitant solution containing 0.1 M MgCl₂, 0.2 M NaCl, 0.1 M sodium citrate pH 5.6, and 18-24% (w/v) PEG 6K or 8K. Crystals were cryoprotected in the reservoir solution supplemented with 15% (v/v) glycerol and flash cooled in liquid nitrogen. Complete X-ray dataset of $\Delta C2:MyrCoA:S$, $\Delta C3:MyrCoA:S$ and $\Delta C3:MyrCoA:T$ were collected at $\lambda = 0.978565$ from a single crystal at 100K in the Proxima1 beamline, at the French National Synchrotron Facility (SOLEIL).

Datasets were integrated with XDS⁸ and scaled and reduced using AIMLESS from the CCP4 package⁹. Crystals belonged to the space group C121 with similar unit cell parameters (summarized in **Supplementary Table 2**) and contained two NMT molecules per asymmetric unit. Structure resolution was accomplished in all cases

using the molecular replacement method. The different structure were solved using PHASER¹⁰ and protein coordinates of a Long-HsNMT1 model (PDB entry 509V) as a search model. Structures were subjected to alternating refinement cycles using PHENIX and manual model building using COOT^{11, 12, 13}. NCS restrictions were applied. In both structures, the electron density maps allowed us to model most of both HsNMT1 molecules of the asymmetric unit. The good quality of the electron density maps also enabled the refinement of substrate peptide, reaction intermediate, and reaction product molecules bound to HsNMT1 in each complex. In the refinement of the $\Delta C2:MyrCoA:S$, $\Delta C3:MyrCoA:S$ and $\Delta C3:MyrCoA:T$ complexes, chemical compound libraries were generated using PRODRG server¹⁴ in combination with eLBOW from the PHENIX suite. The geometry of the final models was validated using MOLPROBITY¹⁵. X-ray data collection and refinement statistics are summarized in **Supplementary Table 2**.

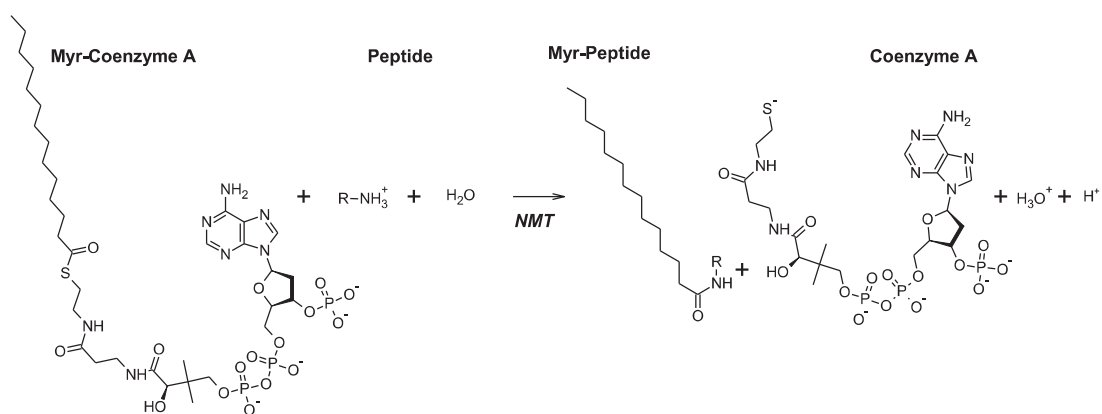
Supplementary References

1. Weston SA, *et al.* Crystal structure of the anti-fungal target N-myristoyl transferase. *Nat Struct Biol* **5**, 213-221 (1998).
2. Wu J, Tao Y, Zhang M, Howard MH, Gutteridge S, Ding J. Crystal structures of *Saccharomyces cerevisiae* N-myristoyltransferase with bound myristoyl-CoA and inhibitors reveal the functional roles of the N-terminal region. *J Biol Chem* **282**, 22185-22194 (2007).
3. Bhatnagar RS, *et al.* Structure of N-myristoyltransferase with bound myristoylCoA and peptide substrate analogs. *Nat Struct Biol* **5**, 1091-1097 (1998).
4. Farazi TA, Waksman G, Gordon JI. Structures of *Saccharomyces cerevisiae* N-myristoyltransferase with bound myristoylCoA and peptide provide insights about substrate recognition and catalysis. *Biochemistry* **40**, 6335-6343 (2001).
5. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324 (2014).
6. Thinon E, *et al.* Global profiling of co- and post-translationally N-myristoylated proteomes in human cells. *Nat Commun* **5**, 4919 (2014).
7. Castrec B, *et al.* Structural and genomic decoding of human and plant myristoylomes reveals a definitive recognition pattern. *Nat Chem Biol* **14**, 671-679 (2018).
8. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* **26**, 795-800 (1993).
9. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214 (2013).
10. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
11. Murshudov GN, *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367 (2011).
12. Adams PD, *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).
13. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
14. Schuttelkopf AW, van Aalten DM. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **60**, 1355-1363 (2004).

15. Chen VB, *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).

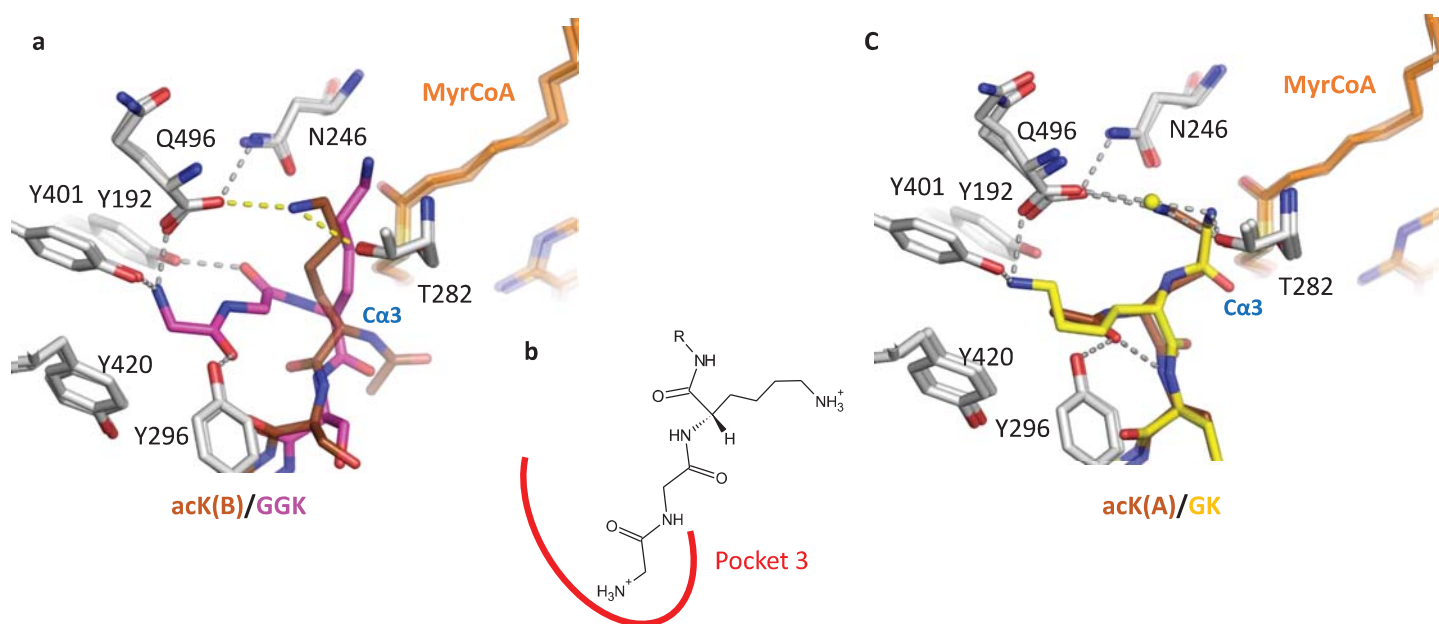
VIII.2. Publication Rivière et al,

VIII.2.1. Figures supplémentaires



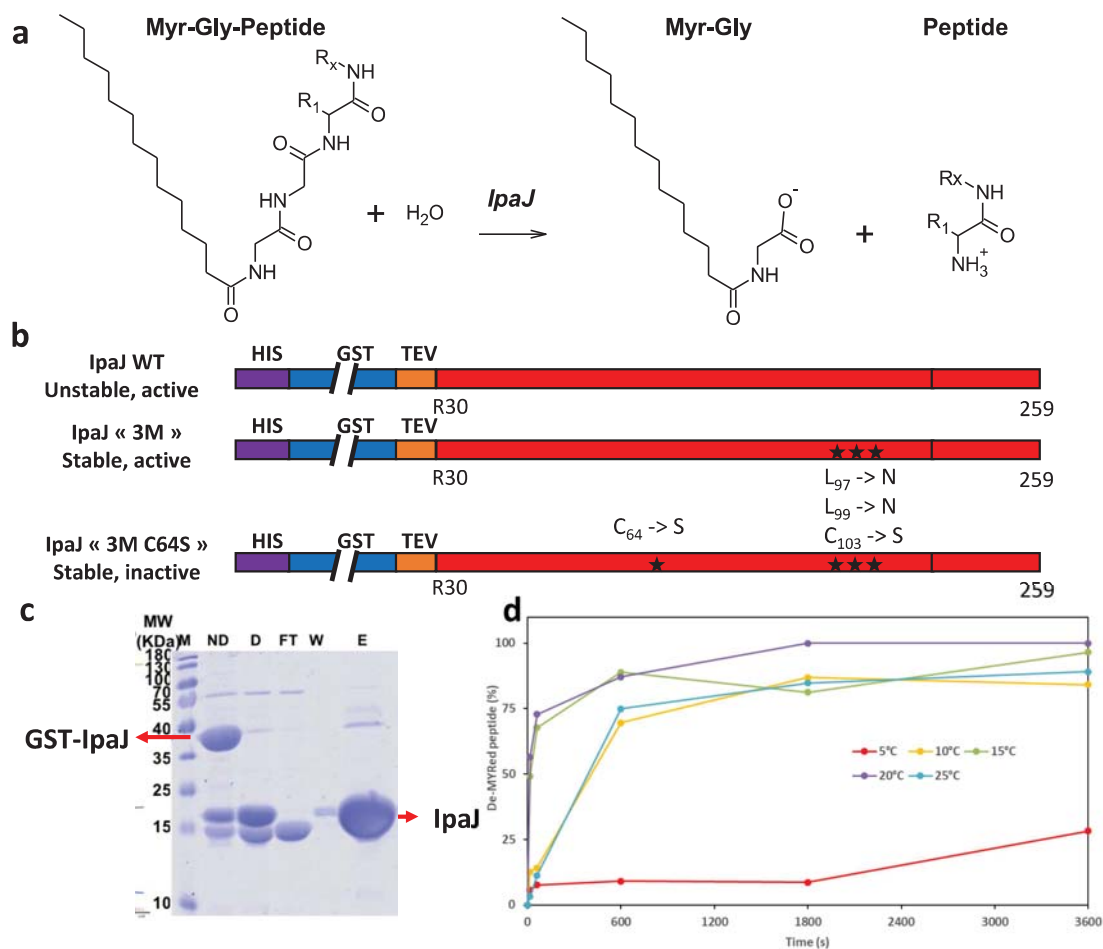
Supplementary Figure 1. The reaction catalyzed by NMT

NMT-catalyzed G-myristoylation. The use of a water molecule for the reaction is indicated.

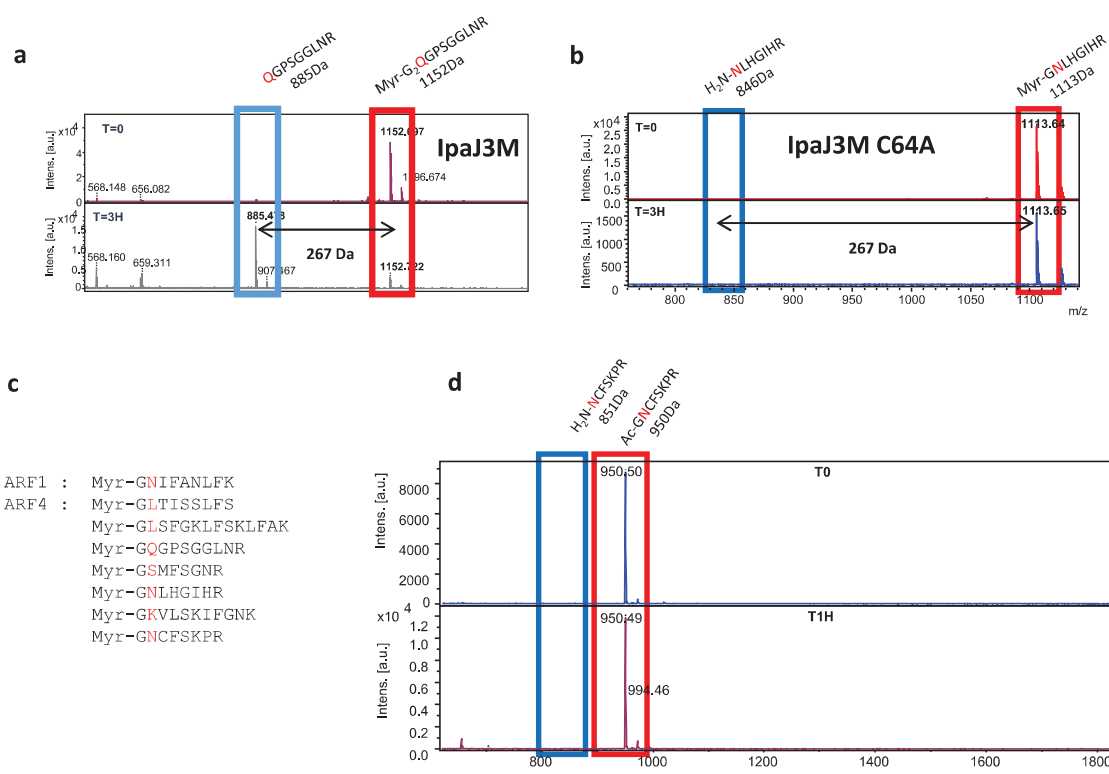


Supplementary Figure 2: Two alternative paths involving rotation around Ca3 of the main chain may lead to K-myristoylation

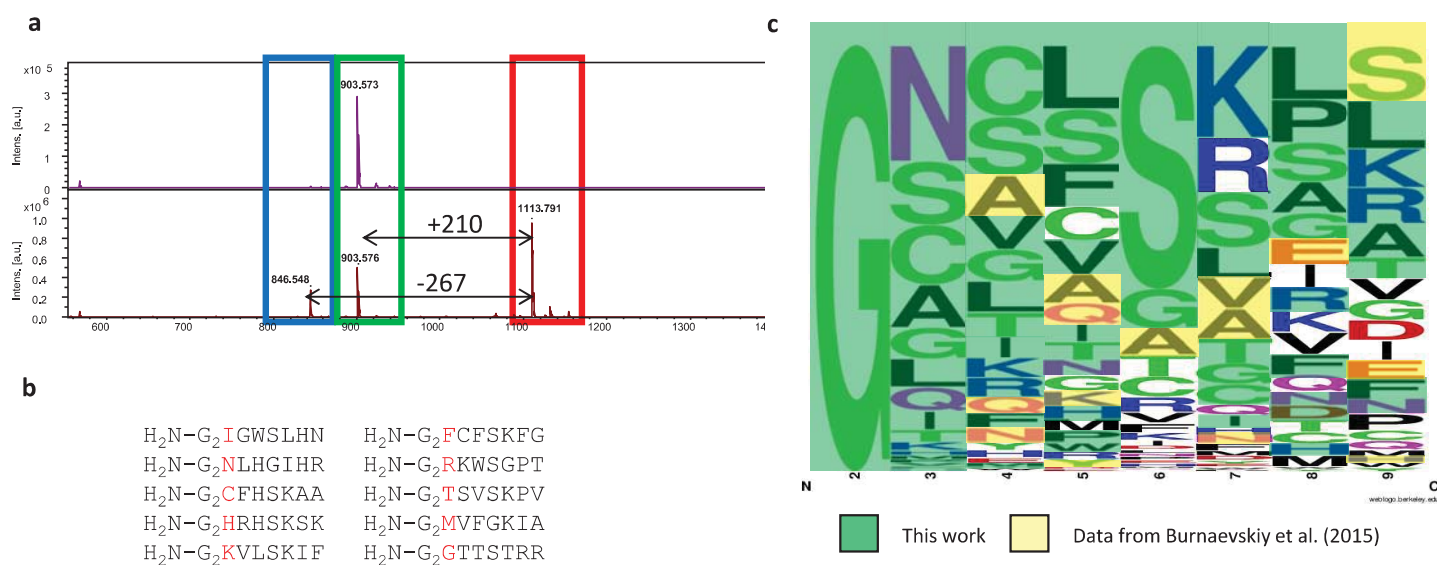
Close up of the 3D structure of the active site of variants of the reference peptide. (a) Overlap of acK (B subunit) with GGK. (b) Overlap of GK (HPCA) with acK (A subunit).



Supplementary Figure 3: A stable variant of IpaJ allows efficient peptide cleavage at low temperatures
 (a) Scheme summarizing the IpaJ-catalyzed chemical reaction. (b) Cartoon summarizing the three constructs used in this study. (c) Purification of IpaJ as a GST fusion and GST cleavage with protease. (d) Impact of temperature incubation on the activity of IpaJ-3M; peptide Myr-GQGPGGLNR was used as the substrate. IpaJ was used at 1 μ M with 100 μ M peptide and incubated for 60 min to allow visualization of the various species by MS. The cleaved/uncleaved peak intensity yield is plotted as a function of time.

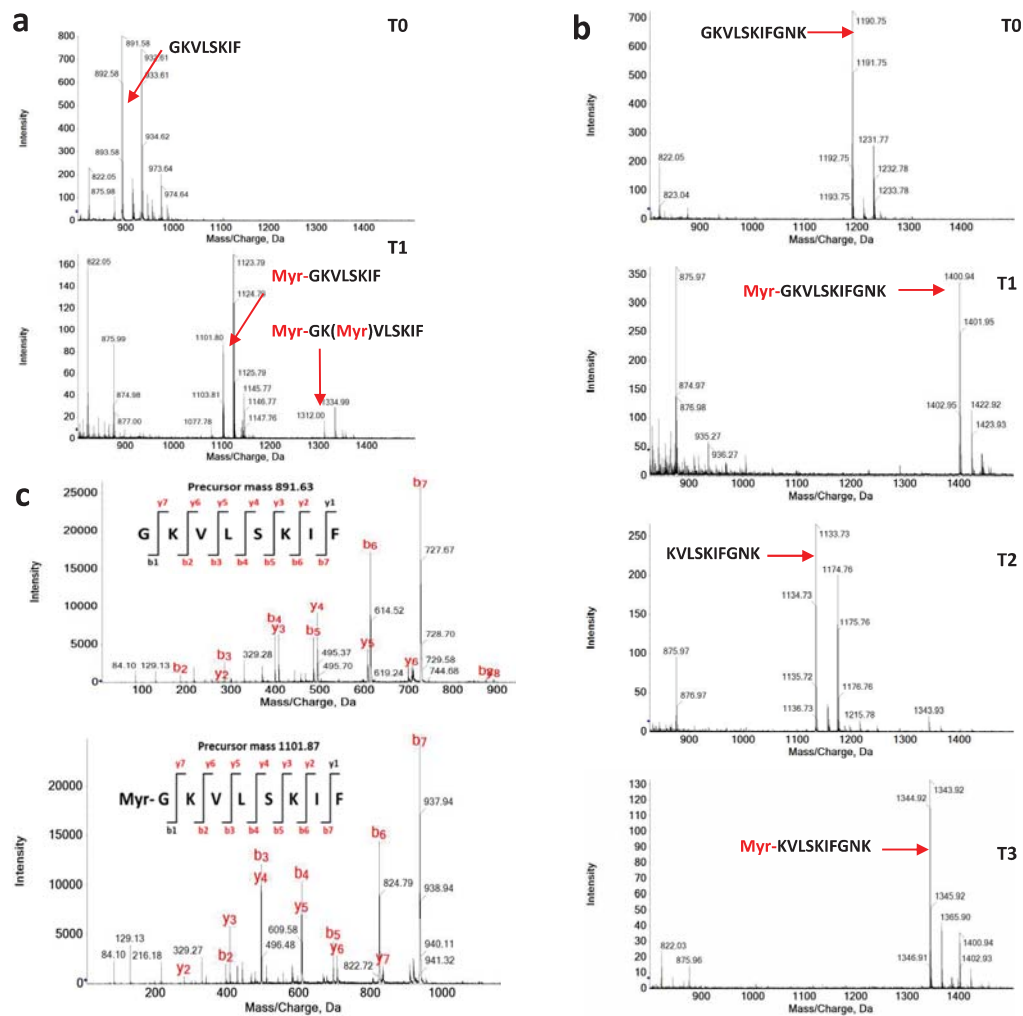


Supplementary Figure 4: The tight substrate selectivity of IpaJ
 (a) Cleavage of a peptide by IpaJ; IpaJ was used at a low (1 μ M) concentration and incubated for 10 min to allow visualization of the various species. (b) Demonstration that variant IpaJ C64A is inactive. Enzyme concentration was 10 μ M and incubation time was 180 min. (c) Various tryptic peptides used to challenge IpaJ *in vitro*. (d) IpaJ-3M does not cleave either Myr-Ala or acetyl-Gly derivatives. IpaJ concentration was 10 μ M and incubation time was 180 min.



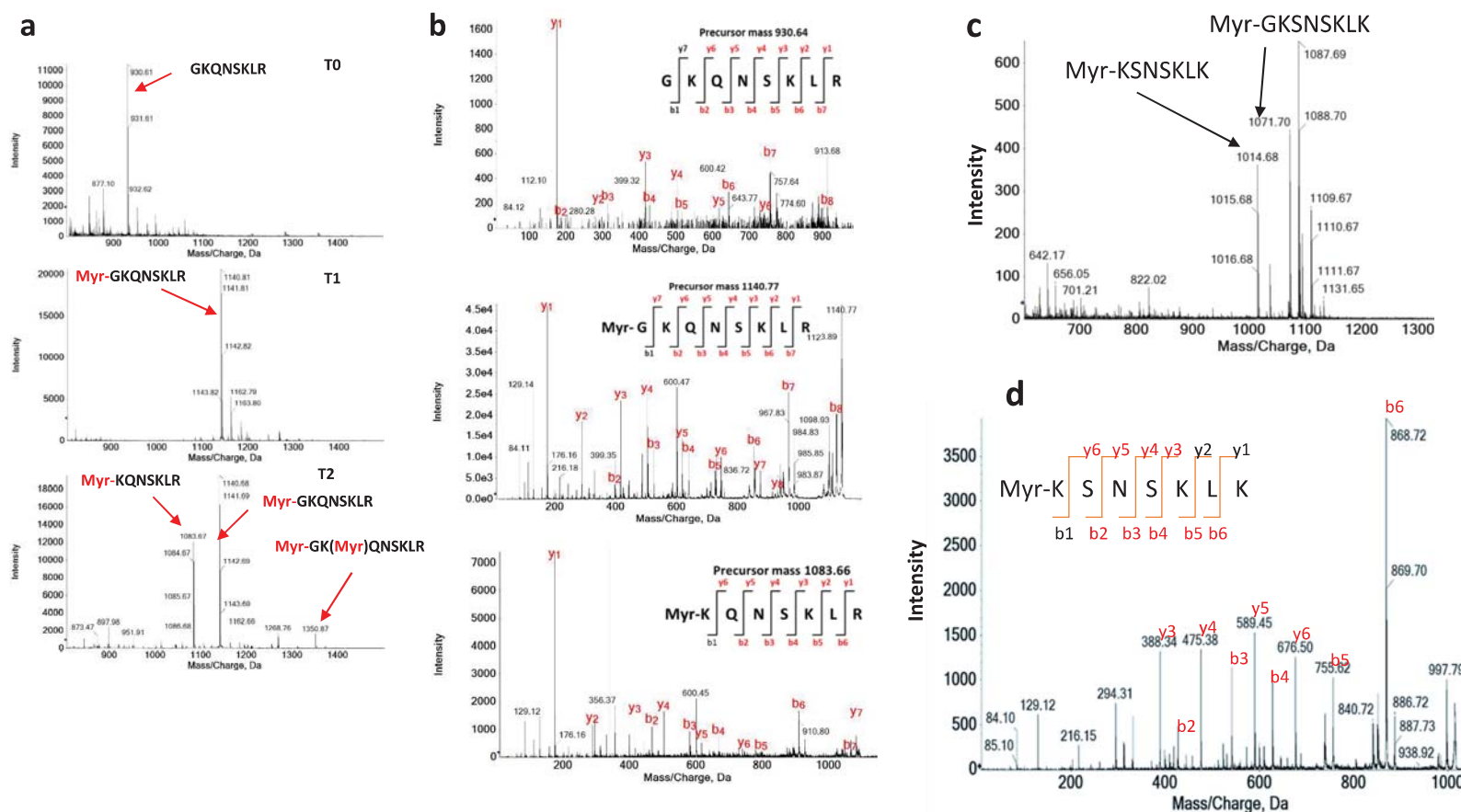
Supplementary Figure 5: MALDI-TOF analyses with the NMT-IpaJ workflow

(a) Example of the behavior of an NMT substrate at T0 and T2. (b) Various peptide sequences verified in the current analysis. (c) Logo recapitulating the sequence space covered by our (colored in green) or other (yellow; see ²³ experiments).



Supplementary Fig. 6. The many ARF6 myristoylation products as revealed by the IpaJ pipeline

(a) The indicated octapeptide derived from ARF6 was incubated in the presence of HsNMT2. (b) An 11-mer version of the ARF6 peptide was subjected to the IpaJ pipeline; MS1 MALDI spectra are displayed at T0, T1, T2, and T3 to reveal four isoforms. (c) MS2 spectra of the unmodified and mono modified short ARF6 peptide, from data in panel a.



Supplementary Figure 7. Neuron-specific calcium-binding proteins may undergo all the many different interconversions before and upon IpaJ cleavage

Two octapeptides were derived from neuron-specific calcium-binding protein hippocalcin isoforms (HPCA/HPCL1/NCALD, ID P84074/P37235/P61601) and neuronal calcium sensor 1 (NCS1; P62166), respectively. The NMT/IpaJ pipeline was applied in the presence of HsNMT1, and the various products were identified. **(a)** HPCA peptide at T0, T1, and T2. **(b)** MS2 spectrum of the double myristoylation product from T2 (panel A). **(c)** MS2 spectrum of the mono myristoylation products from T0, T1, and T2 (panel a). **(d)** NCS1 MS1 spectrum at T2 **(e)** NCS1 MS2 spectrum of the 1014 Da species at T2.

VIII.2.2. Table 1 supplémentaire

Supplementary Table 1: List of all peptides and diagnoses

MALDI and kinetic data of peptides used in this study. MS spectra are displayed in **Supplementary Dataset 1**. Reference peptide sequence is GKSFSKPR. The page number is indicated in one column of each table. **(a)** Peptides displaying G-myristoylation, including those from the human proteome. **(b)** All peptides with K-myristoylation. **(c)** Series with variations on the side chains at positions 6 and 7. **(d)** IpaJ cleavage on various peptides. **(e)** myristoylation after IpaJ cleavage (T2 and T3).

Supplementary Table 1a: GK-starting peptides derived from the human proteome reveal N-terminal alpha-MYR as the major modification made by NMT

Series ^a	Peptide sequence	Short name (text & Figures)	MS data page in Supplementary Dataset 1a	Crystal structure	k_{cat} (s ⁻¹)	K_m (μM)	k_{cat}/K_m (s ⁻¹ .M ⁻¹)	Relative k_{cat}/K_m (%) ^c
Reference GK-peptide variants	GKSF ^S SKPR	GK	2	-	0.050 ± 0.0002	10 ± 2	4 931 ± 626	165
	G{Orn}SFSKPR		3	-	0.018 ± 0.001	6 ± 2	2808 ± 951	94
	G{Dab}SFSKPR		4	-	0.015 ± 0.002	16 ± 7	935 ± 362	31
	ANCFSKPR	AN	5	2.16, ED1a	0.0004 ± 0.0001	>300	1.2 ± 1.1	0.05
	AGKFSKPR		6	-	NM	NM	<0.02	<0.01
	me-GNCFSKPR	meGN	7	3, ED1b	0.0018 ± 0.0001	1 068 ± 51	1.65 ± 0.05	0.06
	GKS ^W SKGR		8	-	0.090 ± 0.002	9 ± 1	10 000	334
	GK ^F SSKPR ^b	GKFS	9/10	-	0.11 ± 0.02	346 ± 75	330 ± 33	11
Human GKX	GKVL ^S KIF	ARF6	11	-	1.01 ± 0.04	194 ± 14	5 215 ± 436	174
	GKQNSKLR	HPCA	12	1.5, Fig. 1a,c	1.03 ± 0.03	204 ± 11	5 055 ± 320	169
	GKTNSKLA	HPCL4	13	-	0.58 ± 0.06	48 ± 18	12 026 ± 3 560	401
	GKQNSKLA	VIS	14	-	0.092 ± 0.004	58 ± 11	1 589 ± 265	53
	GKSNSKLK	NCS1	15	509S, Ref.Castrec	0.047 ± 0.001	15 ± 1	3 157 ± 277	105
	GKSL ^S HLP	T106B	16	-	0.31 ± 0.04	811 ± 224	386 ± 66	13
	GKTFS ^Q LG	T106A	17	-	0.43 ± 0.03	200 ± 53	2 147 ± 467	72
	GKSASKQF	GAPP	18		0.065 ± 0.04	69 ± 17	952 ± 198	32
	GKLH ^S KPA	NKD1	19	-	0.10 ± 0.01	40 ± 2	2 500 ± 200	83
	GKLQSKHA	NKD2	20	-	0.10 ± 0.01	41 ± 2	2 501 ± 200	83
GKX	GAQSGPA	ZNRF2	21	-			103	3
	GGKQSTAA	ZNRF1	22		1.45 ± 0.23	2 320 ± 573	623	21
	GGK ^F FSSKPR		23/24	-	0.14±0.01	85 ± 32	1 647	21
	GGKLSKKK	BASP	25	-			3 900	130
SOS3	GCSVSKKK	SOS3	26	-	0.055 ± 0.002	18 ± 3	2 997 ± 415	100
	AS ^S SVSKKK		27	-	NM	NM	<1	<0.01

^a X means any aminoacid or N-acetyl (ac), which corresponds to a Gly devoid of NH₂, the reactive moiety for G-MYR

^b both G- and K-MYR

Supplementary Table 1b: Impact of the S6K7 motif on G-MYR

Peptide sequence	MS data page in SupDataset 1b	WT		
		k_{cat} (s ⁻¹)	K_m (μM)	k_{cat}/K_m (s ⁻¹ .M ⁻¹)
GKSNSKLLK	2	0.047±0.001	15 ± 2	3133
GKSNSGLKP	3	NM	NM	2.2±0.2
GKSNS {DAP} LKP	4	0.280±0.002	1042±113	269
GKSNS {DAB} LKP	5	0.096±0.003	23±4	4174
GKSNS {ORN} LKP	6	0.16±0.005	29±4	5517
GKSNS {HCY} LKP	7	NM	NM	40
GKSNS {bhLYS} LKP	8	0.10±0.01	92±26	1087
GKVWSQGV	9	0.123	329	373
GKSNAKLKP	10	NM	NM	69

Supplementary Table 1c: Substitutions around the N-terminus of peptides promoting side chain MYR or NMT inhibition

Series ^a	Peptide sequence	Shortname (text & Figures)	Series sequence ^d	MS data page in Supplementary Dataset 1c	Crystal structure ^e	Kinetic parameters with HsNMT1 ^f		
						k_{cat} (s ⁻¹)	K_m (μM)	k_{cat}/K_m (s ⁻¹ .M ⁻¹)
acGK	ac-GKSFSPKPR	acGK		2	6SK2, ED2e,f, Ref.	0.0017 ± 0.0002	58 ± 15	30 ± 6
	ac-G{Orn}SFSKPR	acG[Orn]		3	1.81, ED2,cd	0.0006 ± 0.0001	22 ± 3	29 ± 3
	ac-G{Dab}SFSKPR	acG[Dab] ^c		4	-	NM	NM	<<0.01
	ac-G{D-Orn}SFSKPR	acG[D-Orn] ^c		5	-	NM	NM	<<0.01
	ac-GNCFSPKPR	acGN		6	ED2a,b	NM	NM	<<0.01
	ac-GKVLISKIF	acGK	ARF6	7	-	0.0019 ± 0.0001	37 ± 4	51 ± 6
XK	AKQNSKLR	A-HPCA	HPCA	8	-	>0.0004	>1000	0.4
	AKVLSKIF	A-ARF6	ARF6	9	-	0.0005 ± 0.0001*	8 ± 3*	64 ± 18
	AKSFSKPR	AK		10	Fig.2b,c	0.0006 ± 0.0002*	14 ± 5*	40 ± 11
	AKPTSKDSGLK	TCS2	TCS2	11	-	NM	NM	<<0.01
	A{Orn}SFSKPR	A{Orn} ^c		12	-	NM	NM	<<0.01
	A{D-Orn}SFSKPR	A{D-Orn} ^c		13	-	NM	NM	<<0.01
	SKSFSKPR	SK		14	-	0.0025 ± 0.0003	43 ± 23	60 ± 27
	MKSFSKPR	MK		15	-	>0.0001	>100	1.7
	PKSFSKPR	PK		16	-	0.00068 ± 0.00005	26 ± 8	26 ± 6
	G{D-ORN}SFSKPR	G{D-Orn} ^c		17	-	NM	NM	<<0.01
	ac-KSFSKPR	acK		18	1.7, Fig.4a,b	0.0009 ± 0.0001	24 ± 10	37 ± 13
	ac-{Orn}SFSKPR	ac[Orn] ^c		19	-	NM	NM	<<0.01
	ac-{D-Orn}SFSKPR	ac[D-Orn] ^c		20	Fig.5d,e	NM	NM	<<0.01
	ac-{D-Dab}SFSKPR	ac[D-Dab] ^c		21	-	NM	NM	<<0.01
	ac-{Dap}SFSKPR	ac[Dap]		22	-	NM	NM	<<0.01
K	K	K		23	-	0.00043 ± 0.00004	16 ± 6	27 ± 6
	{Orn}SFSKPR	[Orn]		24	-	>0.0002		2
	{D-Orn}SFSKPR	[D-Orn] ^c		25	-	NM	NM	<<0.01
	{Dab}SFSKPR	[Dab] ^c		26	ED5	NM	NM	<<0.01
	KQNSKLRP	K-HPCA	HPCA	27	?	0.0013 ± 0.0002	133 ± 49	9
	ac-KFSKPR	acKΔS		28	-	NM	NM	<<0.01
	KSFS{Orn}PR			29	-	0.0029 ± 0.0002	59 ± 15	49
	KSFS{Dab}PR			30	-	0.0024 ± 0.0009	145 ± 109	17
	KSFS{Dap}PR			31	-	0.0012 ± 0.00003	39 ± 4	31
	ac-[D-Orn]SNSKLK		NCS1	32	-			1.5
ZXGK	KSNAKLKP		NCS1	33	-	NM	NM	<<0.01
	GGKSFSKP	GGK		34/35	SupFig.2	0.089 ± 0.012	882 ± 263	101 ± 17
	AGKSFSKPR	AGK		36	-	0.0024 ± 0.0004	31 ± 3	76 ± 7
	ac-GGKSFSKPR	acGGK		37	-	0.00062 ± 0.00005	82 ± 24	8 ± 2
	ac-GGGKSFSKPR	acGGGK		38	-	NM	NM	<<0.01
	ac-GGGGKSFSKPR			39	-	NM	NM	<<0.01
	ac-GGGGKSFSKPR			40	-	NM	NM	<<0.01

* only concentrations lower than 50 μM were taken into account as substrate concentration inhibition was observed at higher concentrations

Supplementary Table 1d: IpaJ acts as a cysteine protease with strict specificity for alpha MYR-Gly/peptide cleavage

Sequence	Cleavage by IpaJ	MS data page in Supplementary Dataset 1e
Myr-GLSFGKLFSKLFAK	+	2
Myr-GSMFSGNR	+	3
Myr-GQGPSGGLNR	+	4
Myr-GNLHGIHR	+	5
Myr-GKVLSKIFGNK	+	6
Myr-GNCFSKPR	+	7
ac-GNCFSKPR	-	8
Myr-ATNGSKVA	-	9
	IpaJ cleavage after MYR	
GIGWSLHN	+	10
GNLHGIHR	+	11
GCFHSKAA	+	12
GHRHSKSK	+	13
GMVFGKIA	+	14
GGTTSTRR	+	15
GKVLSKIF	+	16
GFCFSKFG	+	17
GRKWSGPT	+	18
GTSVSKPV	+	19

Supplementary Table 1e: Epsilon myristoylation is induced after IpaJ cleavage

Peptide sequence	Delta mass (Da) after	Myr after cleavage	Myr-peptide after IpaJ cleavage	MS data page in Supplementary Dataset 1e
met-GNCFSKPR	Nd	No	-	2
GKSFSKPR	267,2	Yes	+	3
G{ORN}SFSKPR	267,26	Yes	+	4
GKVLISKIFGNK	267,21	Yes	+	5/6/7/8
GKVLISKIFGN	267,22	Yes	+	5/6/7/8
GKVLISKIF	267,41	Yes	+	5/6/7/8
Myr-GKVLISKIFGNK	267,21	Yes	+	9
G-K (ε-MYR) VLSKIFGNK	Nd	No	+	10
GKFSSKPR	267,2	Yes	+	11
GKQNSKLR	267,35	Yes	+	12
KQNSKLRP	Nd	No	-	13
GKQNSKLA	Nd	Yes	+	14
GKSNSKLLK	Nd	Yes	+	15
GKTNSKLA	Nd	Yes	+	16
GKSASKQF	267,21	Yes	+	17
GAQSGPA	Nd	No	-	18
GGKQSTAA	Nd	No	-	19
GGKLSKKK	Nd	Nd	-	20
GGKWSKLS	267,21	Yes	-	21
AKSFSKPR	No	No	-	22
GKSWSKGR	267,2	Yes	+	23
AKQNSKLR	Nd	No	-	24
GGKQSKAA	Nd	No	-	25
GGKSFSKP	No	No	-	26
GKGFSSKPR	267,21/324,23	Yes/Yes	+/+	27/28
AKPTSKDSGLK	Nd	No	-	29
AGKFSKPR	No	No	-	30
AGKSFSKPR	No	No	-	31

VIII.2.3. Table 2 supplémentaire

Présent en fichier annexe séparé.

VIII.2.4. Table 3 supplémentaire

Supplementary Table 3: Primers used in this study

Cloning and site-directed mutagenesis primers for IpaJ.

Sequence	Use
TGGCCGCACCGCTAGAAATTATCATACATTGTTTGC	C64S-for
GCAAACAAATGTATGATAATTCTAGCGGTGCGGCCA	C64S-rev
AGATACAGATCGCGTTCACTACGGTTGTCATTATCATTAGAGCTTTACGGGTCATTTCCGACATTG	L97N-L99N-C102S-for
CAATGTCGGAAATGACCCGTAAAAGCTCTAATGATAATGACAACCGTAGTGAACGCGATCTGTATCT	L97N-L99N-C102S-rev
AATCTTTATTTTCAGGGCGCCATGGCGCGTGAATATATGATTAGCCATCAG	IpaJ-Nt30-for
TCGAGTGCGGCCCAAGCTTTCACAGTGCTTCGTTGGTAATAA	IpaJ-Ct259-rev
ATGTCCGAACAACGCAAAACCGTGTAACGTGGCTGTATTCATACGGGTGTGATGCTGTATGGTGTCTGCTGCAAGGTGC TATTCCGCGTGAATATATGATTAGCCATCAGACCGATGTCCGTGTGAACGAAAAATCGCGTCAACGAACAGGGCTGCTTTC TGGCGCGCAAAACAAATGTATGATAATTCTTGGCGTGCGGCCAGTCTGCTGTGTGCAGCTAAAGAACTGGGCGTTGACAAA ATTCCGCAATACAAAGGTTCAATGTCGGAAATGACCCGTAAAAGCTCTCTGGATCTGGACAACCGTTGTGAACGCGATCT GTATCTGATCACGTCCGGCAACTACAATCCGCGTATTCACAAAGATAATATCGCGGACGCCGTTACTCTATGCCGGACA AAATTGTTATGGCCACCCGTCTGCTGGGTCTGAACGCCTATGTGGTTGAAGAATCAAACATCTTTTCGCAGGTTATCAGC TTCATCTACCCGGATGCCCGTGACCTGCTGATTGGCATGGGTTGCAACATCGTCCATCAGCGTGATGTGCTGAGTTCCAA TCAACGCGTGCTGGAAGCAGTTGCTGTACGCTTTATCGGCGTGCCGTTGGTCTGCACTGGGTTCTGTGCCGCCCGGATG GCAGTTATATGGACCCGGCAGTCGGTGAAAACTACAGTTGTTTCTCCACGATGGAACCTGGGTGCTCGTCGCAGCAATTCT AAGTTTATCGGCTACACCAAAATCGGCATCAGCATTGTTATTACCAACGAAGCACTGTGA	Full synthetic DNA sequence for <i>S.</i> <i>flexneri</i> IpaJ (1-259)

VIII.2.5. Table 4 supplémentaire

Supplementary Table 4: Data collection and refinement statistics of HsNMT1 in complex with peptide substrates and inhibitors

PDB accession code	HPCA		AK		AcK		AcG[ORN]		AN		me-GN		GGK	
	7OWM		7OWN		7OWO		7OWP		7OWU		7OWQ		7OWR	
Data collection														
Space group	<i>P</i> 2 ₁ 2 ₁ 2		<i>P</i> 2 ₁ 2 ₁ 2		<i>P</i> 2 ₁ 2 ₁ 2		<i>P</i> 2 ₁ 2 ₁ 2		<i>P</i> 2 ₁ 2 ₁ 2		<i>C</i> 121		<i>C</i> 121	
Cell dimensions														
<i>a</i> , <i>b</i> , <i>c</i> (Å)	79.88, 178.4, 58.3		79.1, 178.1, 58.2		79.32, 178.3, 58.2		79.32, 178.2, 58.2		78.74 178.0 58.3		93.65, 58.88, 154.21		90.4, 61.4, 158.9	
α , β , γ (°)	90.0, 90.0, 90.0		90.0, 90.0, 90.0		90.0, 90.0, 90.0		90.0, 90.0, 90.0		90.0, 90.0, 90.0		90.0, 90.0, 90.0		90.0, 90.0, 90.0	
Resolution (Å)	48.8–1.5		48.90–2.01		48.7–1.70		48.7–1.81		46.90–2.04		47.69–3.0		48.4–2.39	
	(1.54–1.50)*		(2.06–2.01)		(1.74–1.70)		(1.85–1.81)		(2.08–2.04)		(3.18–3.00)		(2.44–2.39)	
<i>R</i> _{sym} or <i>R</i> _{merge}	0.115 (1.532)		0.171 (1.486)		0.11 (1.401)		0.100 (0.562)		0.060 (0.317)		0.134 (0.601)		0.090 (0.481)	
<i>I</i> / σ <i>I</i>	12.2 (1.4)		15.9 (2.0)		14.4 (1.9)		16.9 (4.8)		13.0 (4.3)		11.2 (3.0)		12.7 (3.12)	
Completeness S(%)	96.1 (77.5)		99.4 (92.1)		86.6 (48.7)		99.1 (98.2)		99.6 (99.9)		98.6 (98.7)		99.4 (91.3)	
Completeness E(%)	98.1 (97.8)				97.2 (96.1)								99.4 (91.3)	
Redundancy	13.9 (8.9)		14.9 (15.7)		13.2 (12.3)		15.0 (14.7)		3.4 (3.3)		6.3 (5.9)		5.71 (4.9)	
Refinement														
Resolution (Å)	48.8–1.50		48.8–2.10		48.7–1.70		47.6–1.81		46.86–2.08		47.7–3.00		48.4–2.39	
No. reflections	34542		49330		79532		75344		49769		32321		34536	
<i>R</i> _{work} / <i>R</i> _{free}	15.44/19.58		16.63/21.50		18.29/22.32		15.52/19.10		17.86/22.36		21.05/25.51		17.34/21.87	
No. atoms														
Protein	6433		6411		6489		6538		6517		6330		6387	
Ligand/ion	219		150		207		144		156		126		144	
Water	1006		537		676		790		525				366	
B-factors														
Protein	23.98		31.88		25.65		31.01		29.75		45/17		30.48	
Ligand/ion	25.97		36.67		29.69		40.45		36.23		53.26		26.83	
Water	32.68		35.25		31.85		37.61		35.30				30.37	
R.m.s. deviations														
Bond lengths (Å)	0.006		0.007		0.007		0.006		0.004		0.003		0.003	
Bond angles (°)	0.874		0.913		0.899		0.862		0.793		0.700		0.823	

The datasets were collected from a single crystal for each hsNMT1 complex.

*Highest-resolution shell is shown in parentheses.

Supplementary Table 4 (following): Data collection and refinement statistics of HsNMT1 in complex with peptide substrates and inhibitors

Ac[D-ORN]		DAB	
PDB accession code	7OWS	7OWT	
Data collection			
Space group	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2	
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	79.8, 179.1, 58.4	78.74 178.0 58.3	
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	
Resolution (Å)	48.9–1.90 (1.94–1.90)*	48.8–1.90 (1.94–1.90)	
<i>R</i> _{sym} or <i>R</i> _{merge}	0.080 (0.678)	0.057 (0.584)	
<i>I</i> / σ <i>I</i>	21.6 (5.0)	25 (4.5)	
Completeness S(%)	99.9 (99.7)	100 (99.9)	
Completeness E(%)			
Redundancy	14.9 (15.7)	13.0 (13.1)	
Refinement			
Resolution (Å)	48.9–1.90	48.8–1.90	
No. reflections	66831	65500	
<i>R</i> _{work} / <i>R</i> _{free}	15.78/19.48	17.19/21.50	
No. atoms			
Protein	6553	6484	
Ligand/ion	179/1	158/2	
Water	616	487	
<i>B</i> -factors			
Protein	30.52	37.70	
Ligand/ion	29.64/43.62	41.73/49.53	
Water	37.57	40.78	
R.m.s. deviations			
Bond lengths (Å)	0.007	0.007	
Bond angles (°)	0.940	0.943	

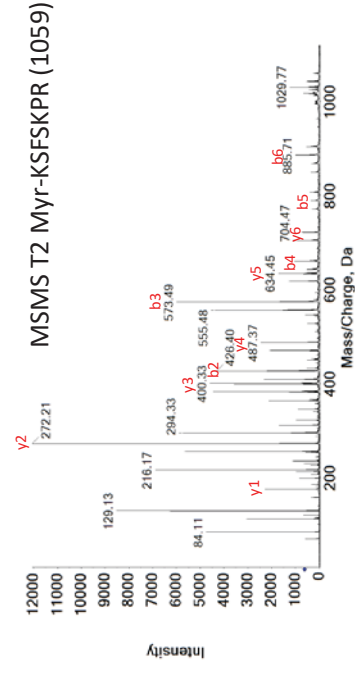
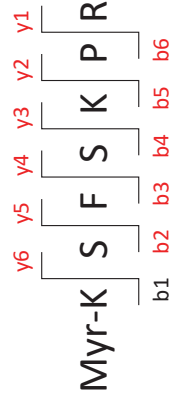
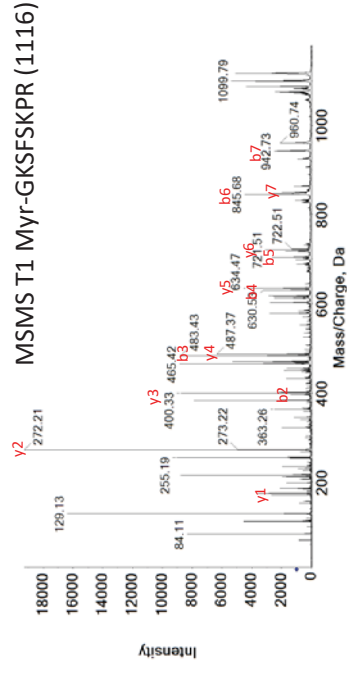
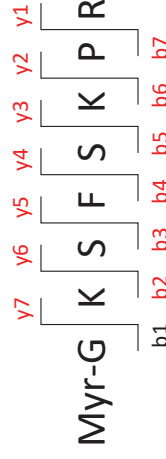
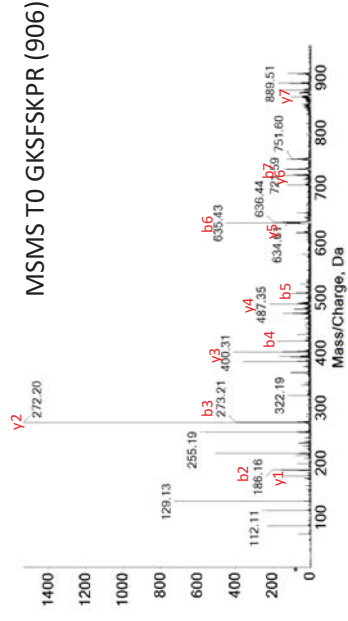
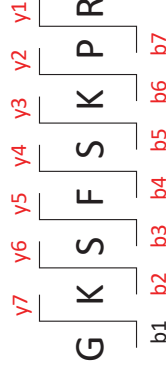
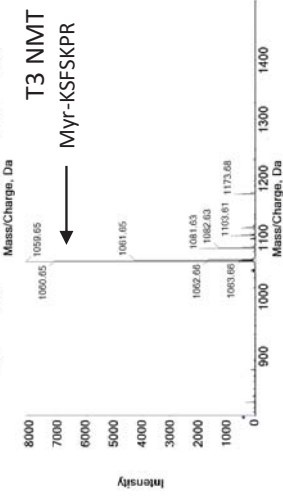
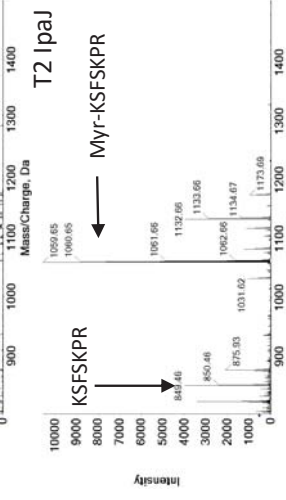
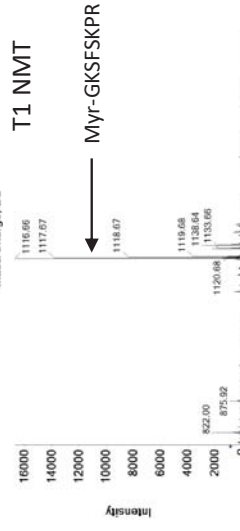
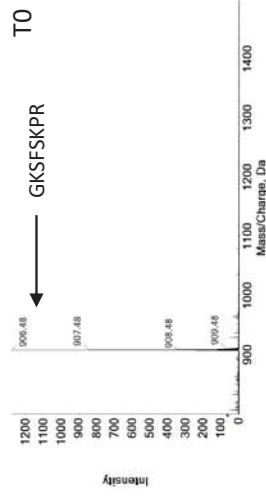
The datasets were collected from a single crystal for each hsNMT1 complex.
*Highest-resolution shell is shown in parentheses.

VIII.2.6. Données supplémentaires 1a

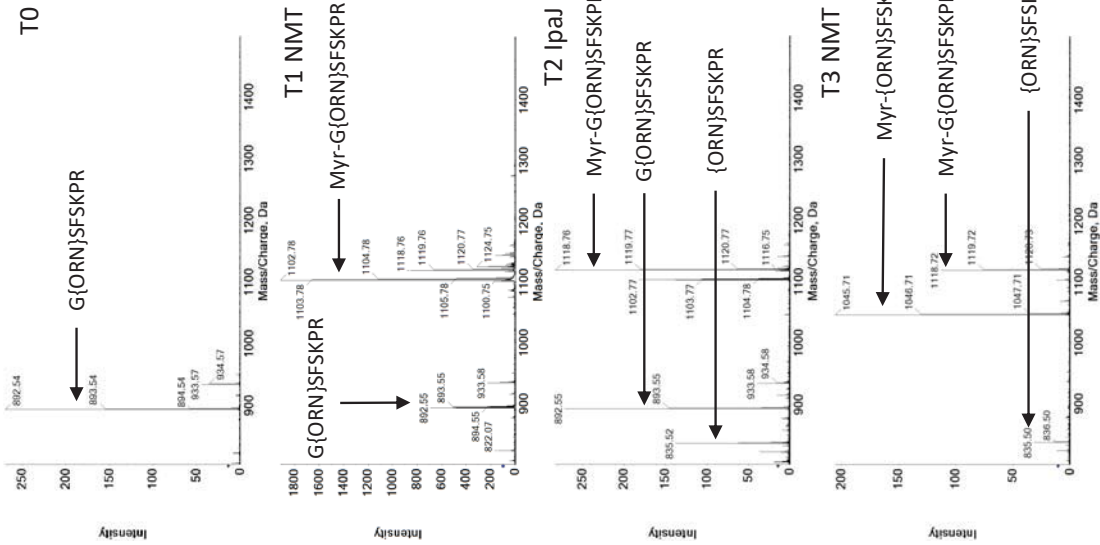
Sup Dataset 1a

Series ^a	Peptide sequence	Short name (text & Figures)	MS data in Supplementary Dataset 1a
Reference GK-peptide variants	GKFSKPR	GK	2
	G{Orn}SFSKPR		3
	G{Dab}SFSKPR		4
	ANCFSKPR	AN	5
	AGKFSKPR		6
	me-GNCFSKPR	meG	7
	GKSWSKGR		8
	GKFSKPR ^c	GKFS	9/10
Human GKX	GKVLSKIF	ARF6	11
	GKQNSKLR	HPCA	12
	GKTN SKLA	HPCL4	13
	GKQNSKLA	VIS	14
	GKSNSKLR	NCS1	15
	GKSLSHLP	T106B	16
	GKTFSQLG	T106A	17
	GKSASKQF	GAPP	18
	GKLHSPAP	NKD1	19
	GKLQSKHA	NKD2	20
GKX	GAQSGPA	ZNRF2	21
	GGKQSTAA	ZNRF1	22
	GGRFSSKPR		23/24
	GGKLSKKK	BASP	25
SOS3	GCSVSKKK	SOS3	26
	ASVSKKK		27

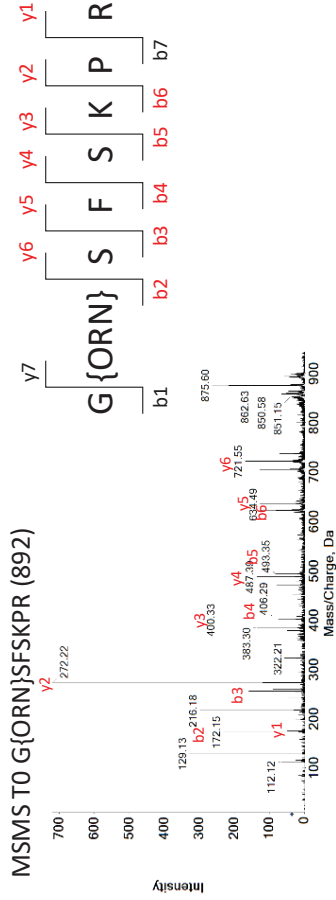
GKSFSKPR



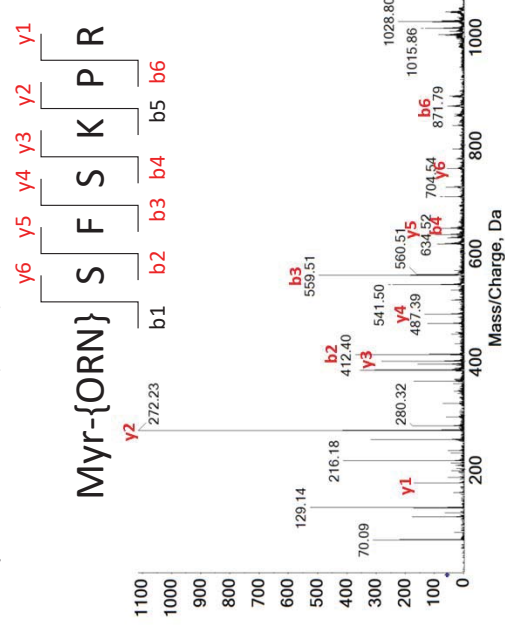
G{ORN}SFSKPR



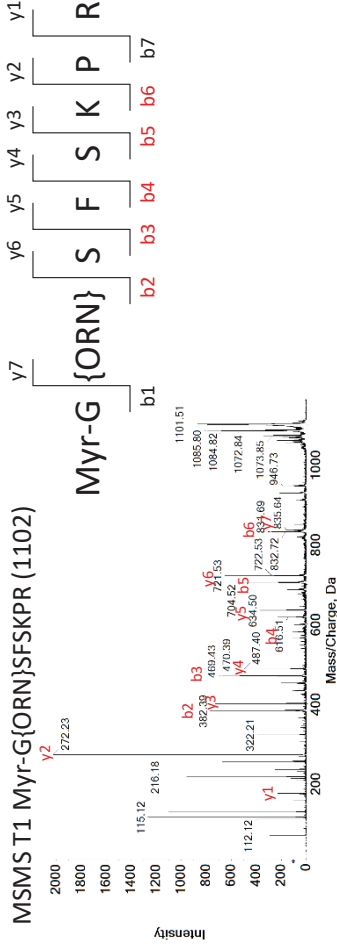
MSMS T0 G{ORN}SFSKPR (892)



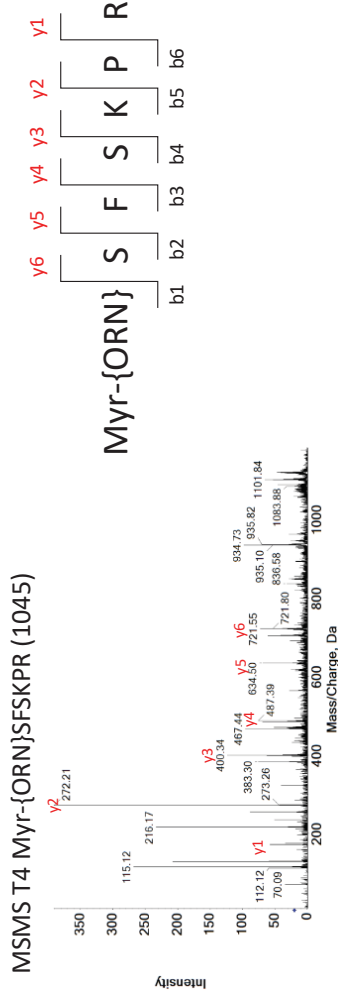
MSMS T3 Myr-{ORN}SFSKPR (1045)



MSMS T1 Myr-G{ORN}SFSKPR (1102)

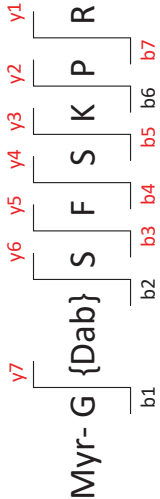
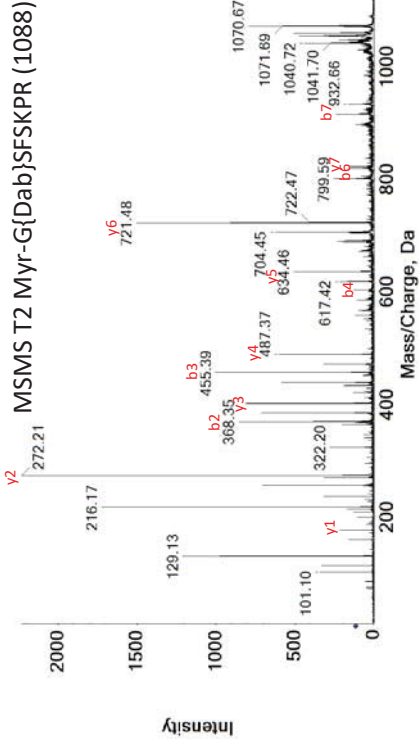
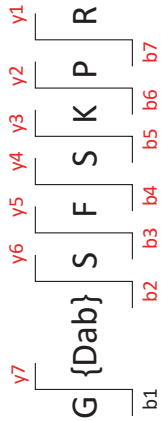
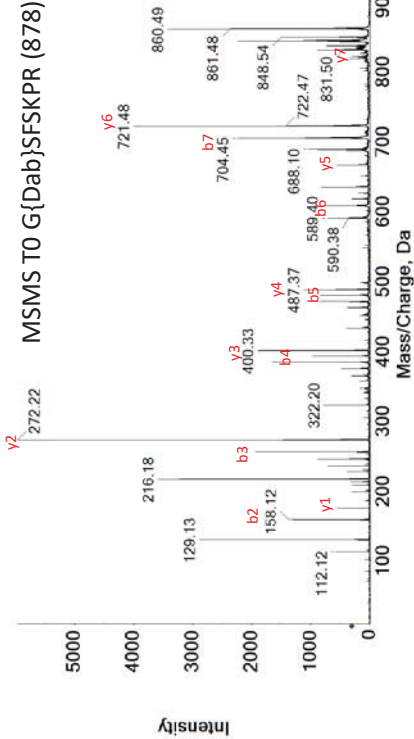
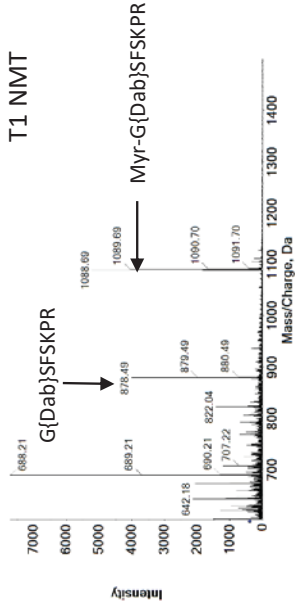
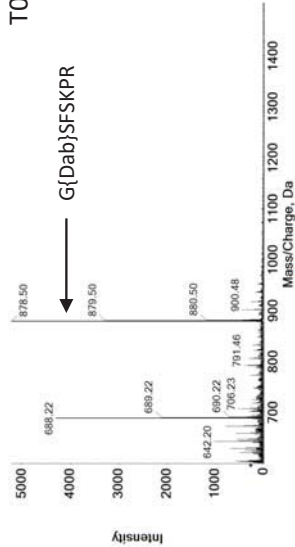


MSMS T4 Myr-{ORN}SFSKPR (1045)

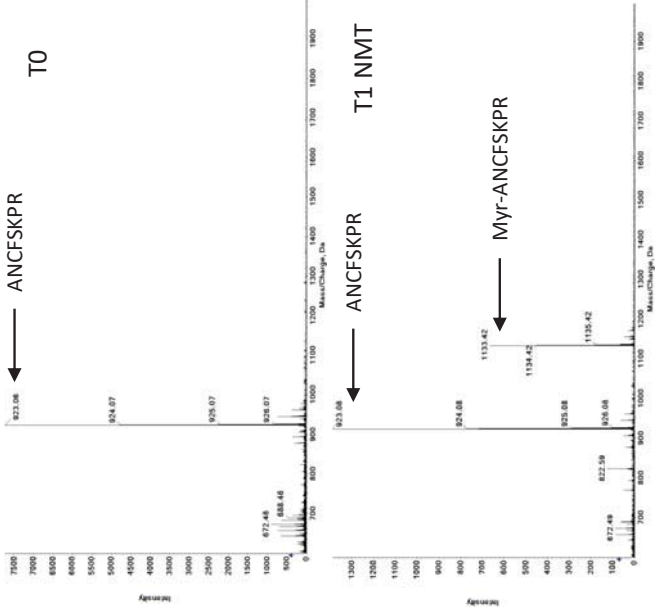


Mass/C harge (Da)	Intensity	ion
875	175,15	170,50 y1
1983	272,22	1113,88 y2
4152	400,33	282,50 y3
4407	412,37	371,60 b2
5781	487,39	134,58 y4
6965	559,47	496 b3
8430	634,48	97,41 y5
8686	646,51	76,39 b4
9878	721,51	61,96 y6
11989	871,78	56,62 b6

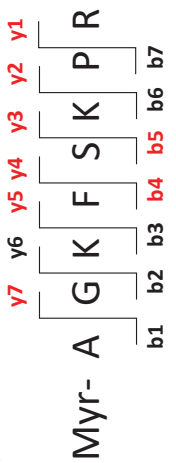
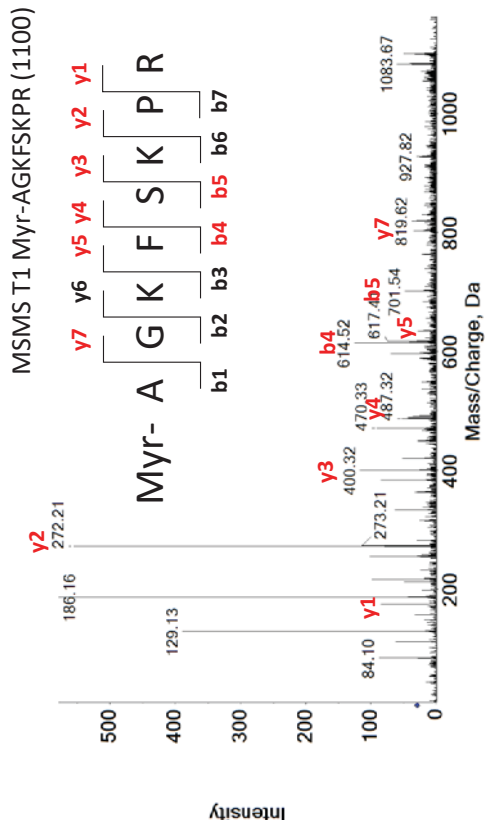
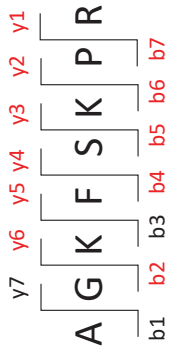
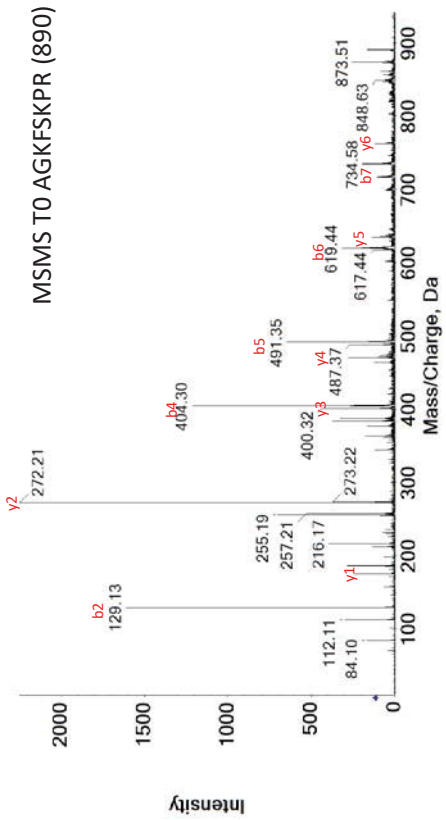
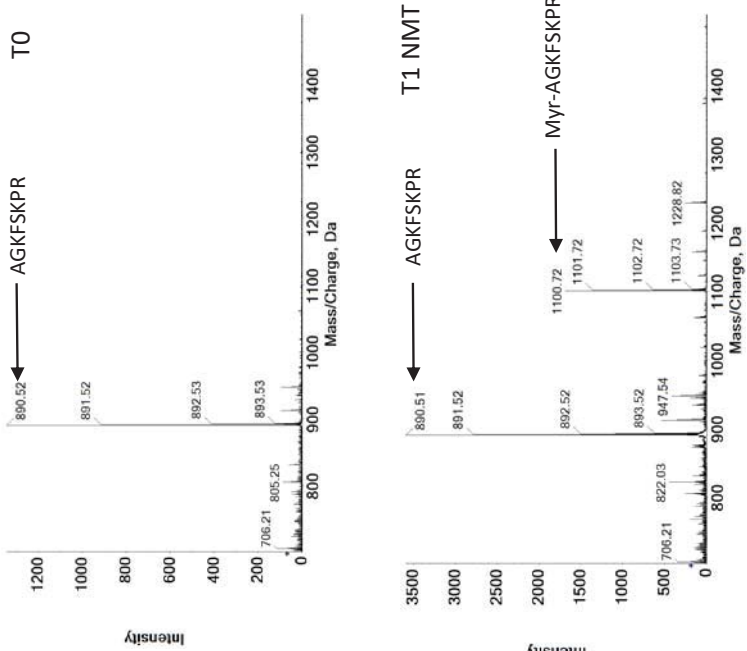
G{Dab}SFSKPR



ANCFKPR

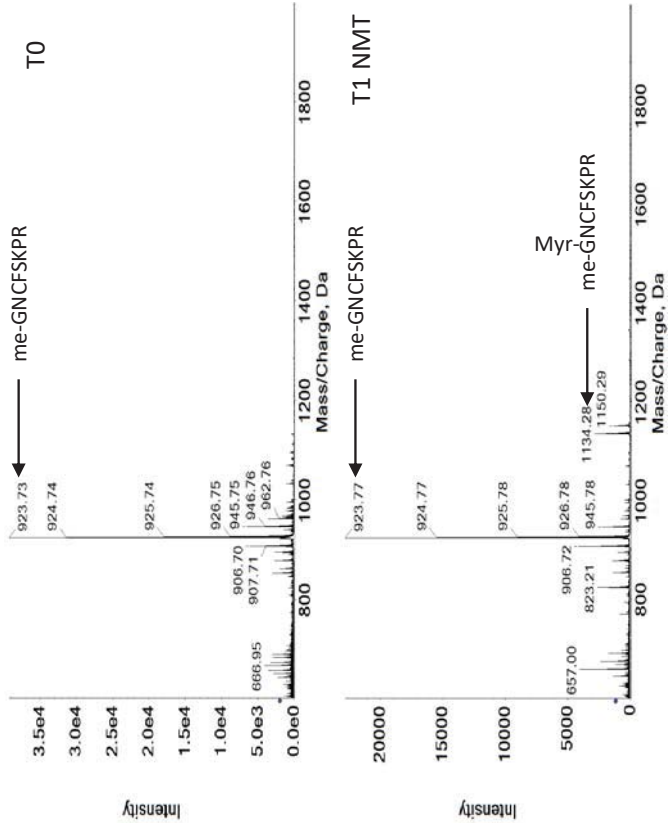


AGKFSKPR

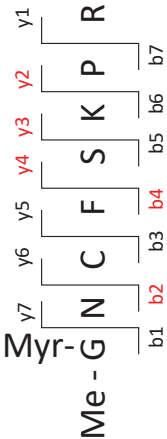
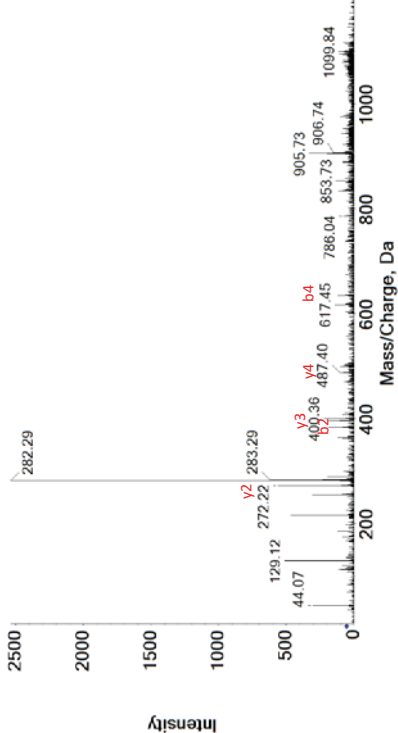


L4 1100	Mass/Charge (Da)	Intensity
542	129.138	388,8628 K
801	175,1467	84,86275 Y1
894	186,1513	578,353 GK
1149	216,181	98,66667 SK
1590	272,2057	555,9216 Y2
2024	333,2545	63,21569 GKF
2715	400,3234	116,8627 Y3
2896	420,3111	51,45098 GKFS
3333	470,3304	90,82353 Y4-NH3
3503	487,3036	52,07843 Y4
4591	614,5147	126,1177 b4
4836	634,4205	28,07843 Y5
5421	701,5098	47,68628 b5
6483	819,5577	37,17647 Y7

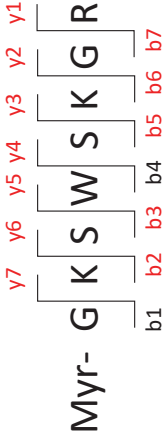
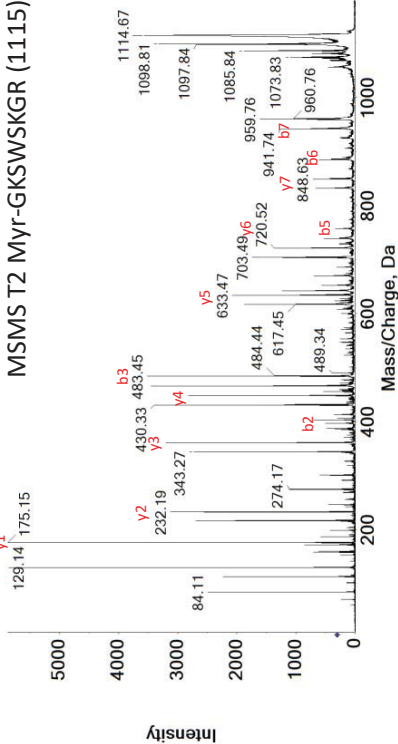
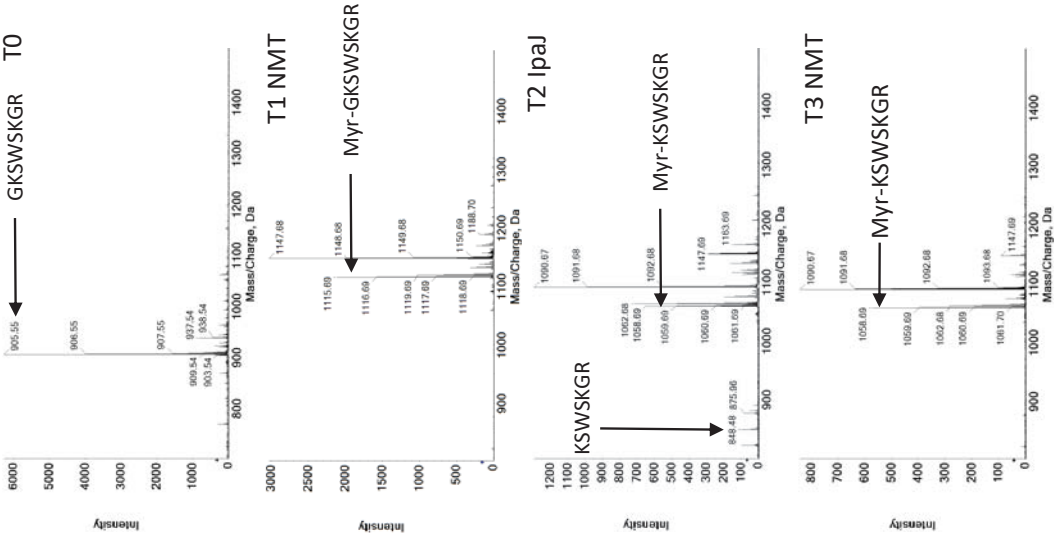
me-GNCFSKPR



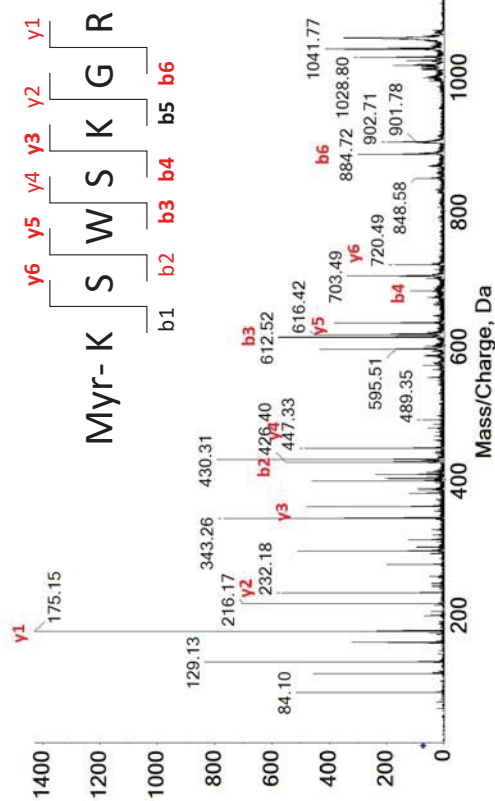
MSMS T1 me-(Myr)GNCFSKPR (1134)



GKSWSKGR

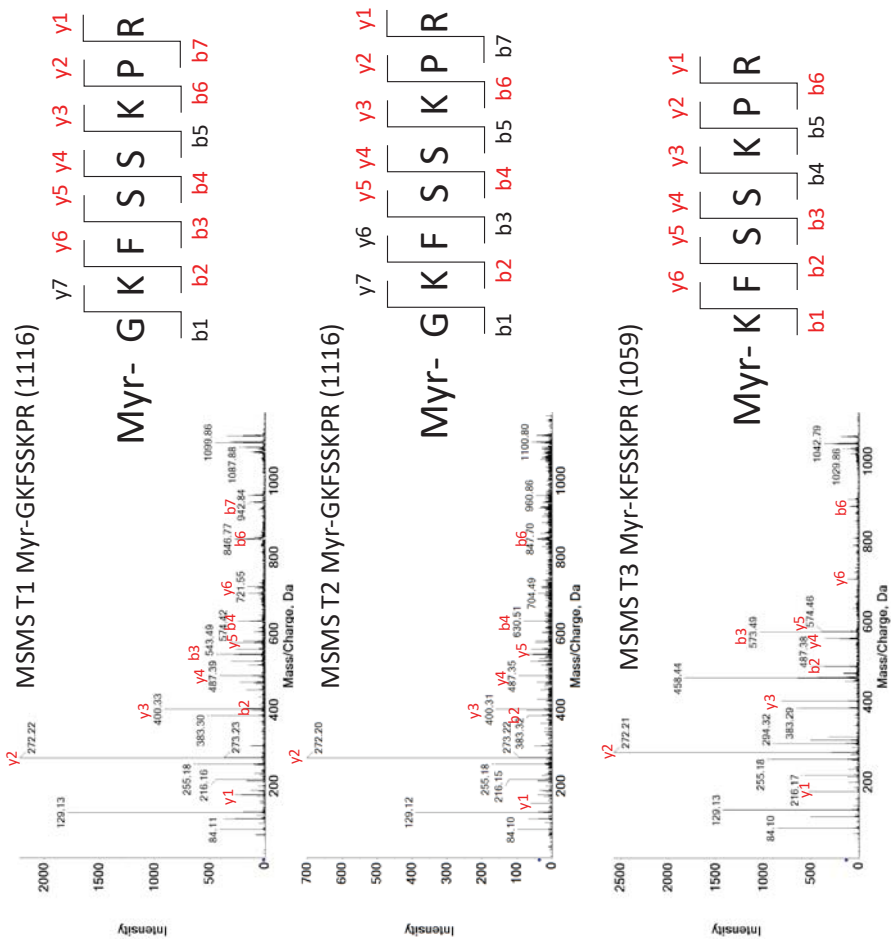
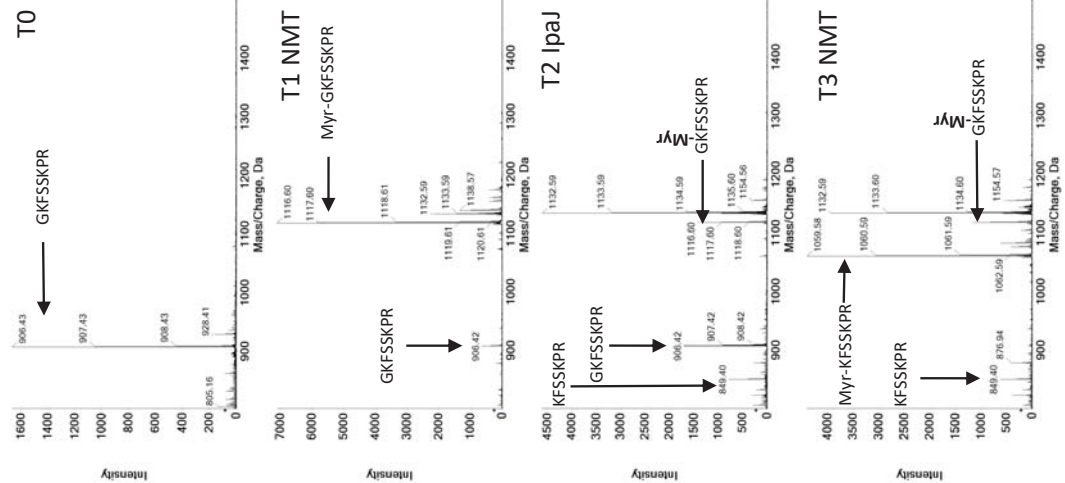


MSMS T3 Myr-KSWSKGR (1058)

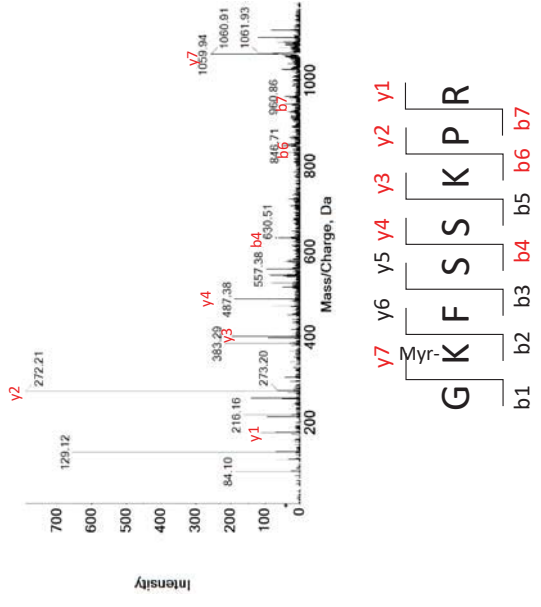


Mass/Ch	arge (Da)	Intensity
1520	175,1429	1428,392 y1
2460	232,1888	496,1569 y2
5185	360,3024	479,3726 y3
7074	426,405	552,1569 b2
7655	447,3432	485,4902 y4
11575	612,537	578,0392 b3
12102	633,4509	383,6863 y5
13648	699,5988	61,80392 b4
14137	720,4839	172,0784 y6
17256	884,7047	301,9608 b6

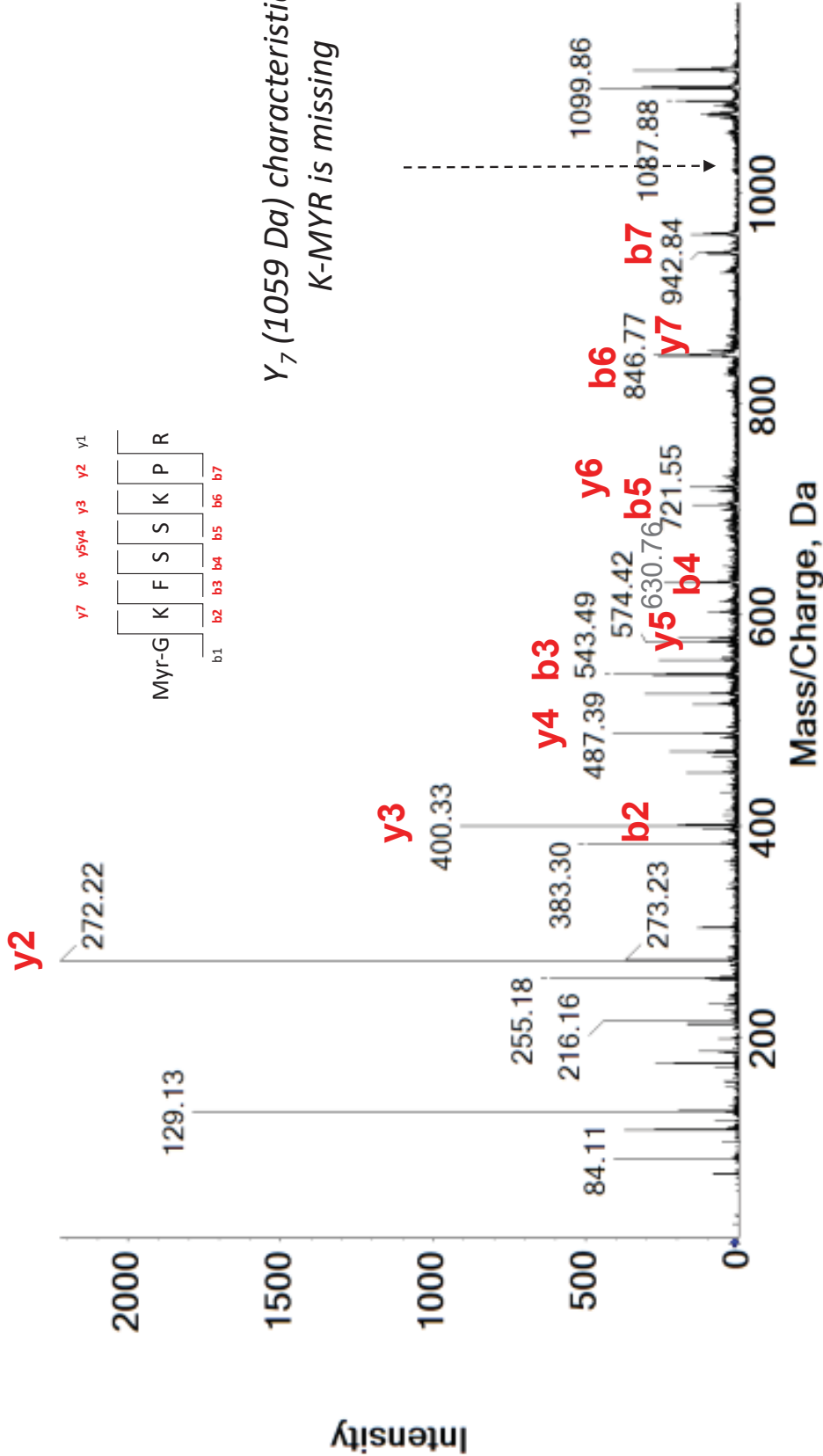
GKFSSKPR



MSMS T3 GK(Myr)FSSKPR (1116)



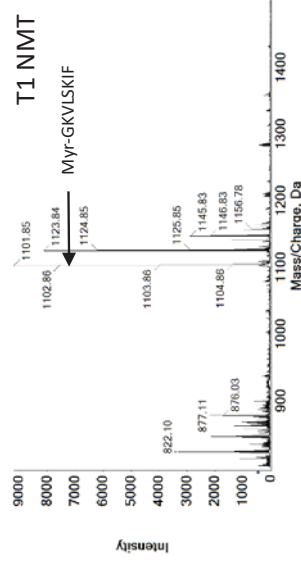
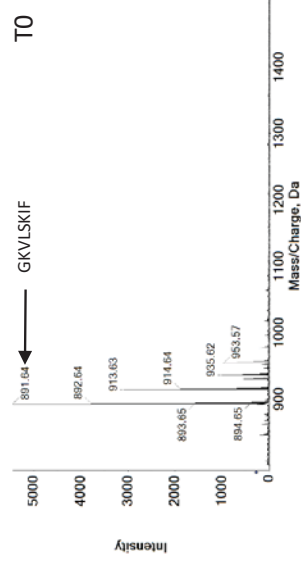
MSMS T1 Myr-GKFFSKPR (1116)



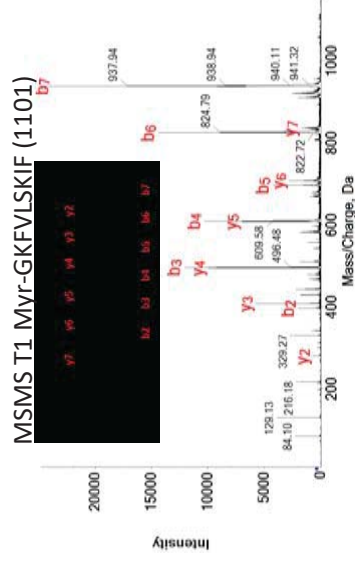
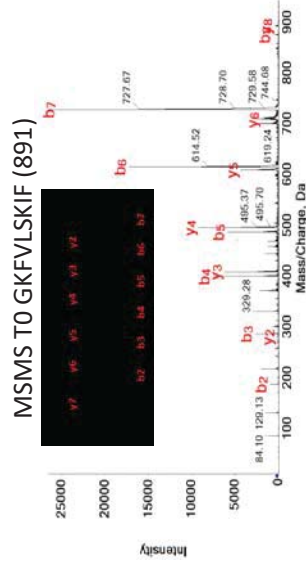
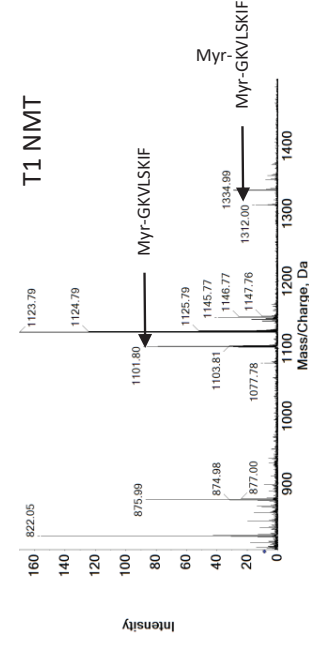
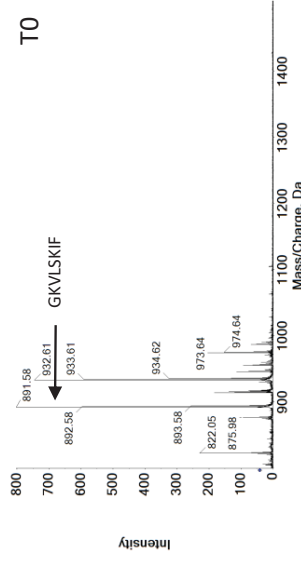
*Y₇ (1059 Da) characteristic of
K-MYR is missing*

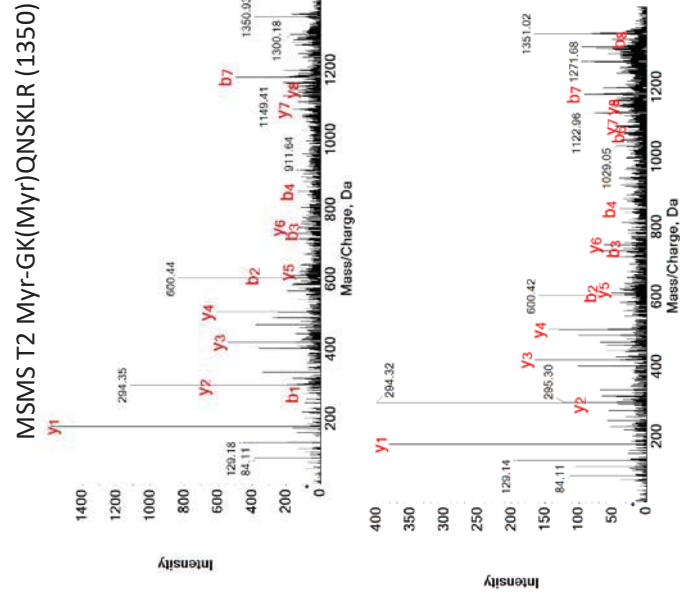
L9	1116.6	Mass/Charge (Da)	Intensity
	1118	175,1541	270,5883y1
	2531	272,2059	2221,647y2
	4646	396,3899	116,549b2
	4763	400,3088	912,1569y3
	6619	487,3806	411,9216y4
	7923	543,4726	409,5686b3
	8670	574,7245	101,1765y5
	9901	630,7576	151,5294b4
	11807	717,5554	91,76471b5
	11897	721,5382	156,7059y6
	14106	845,9461	136b6
	14189	849,7706	87,52941y7
	15804	942,8001	135,0588b7

GKVL SKIF NMT1

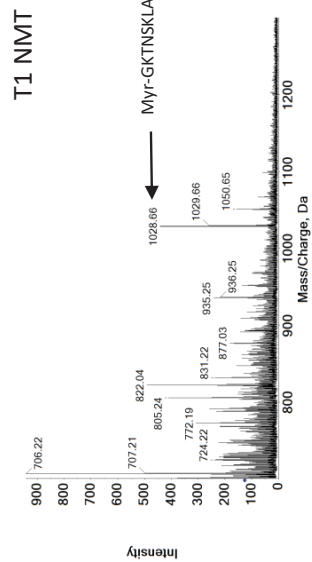
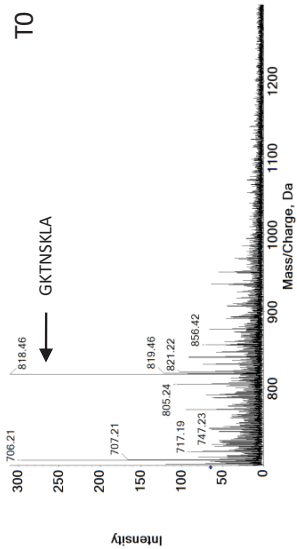


GKVLKIF NMT2

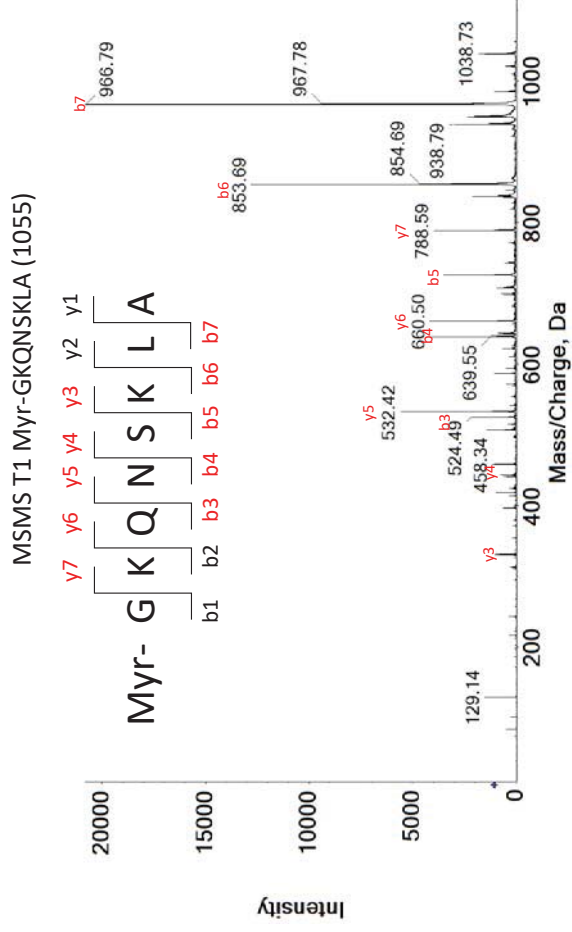
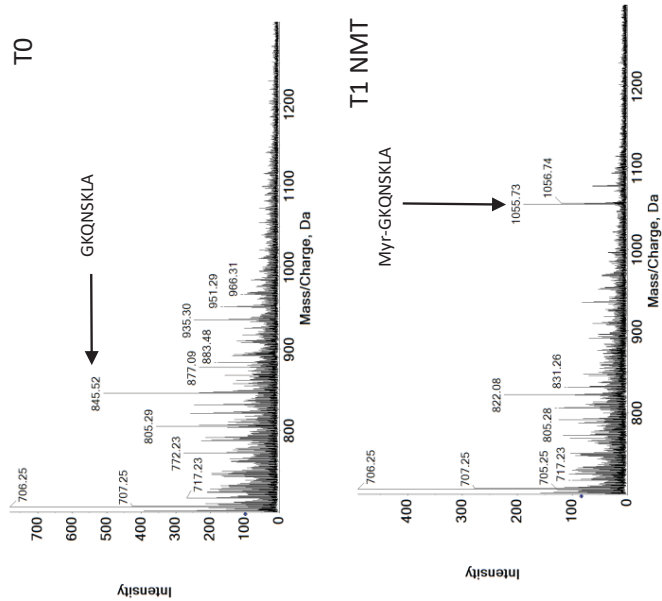




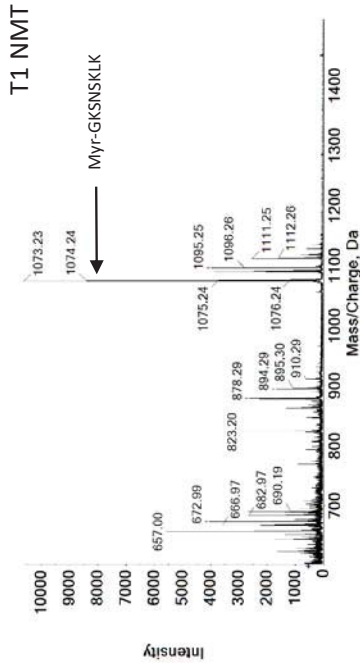
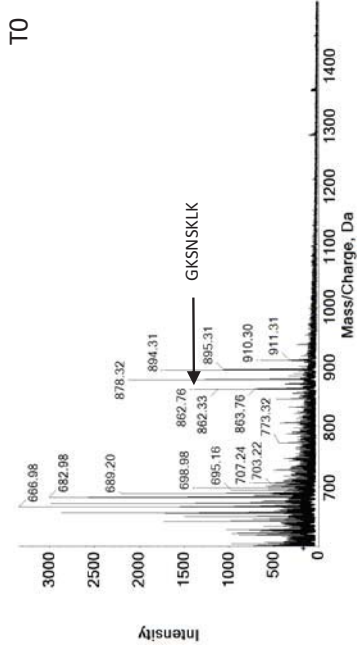
GKTN SKLA



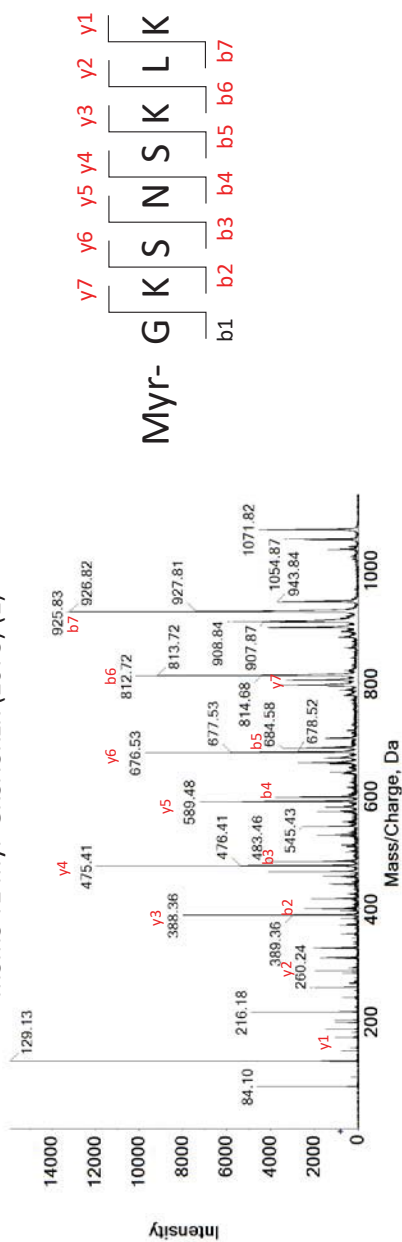
GKQNSKLA



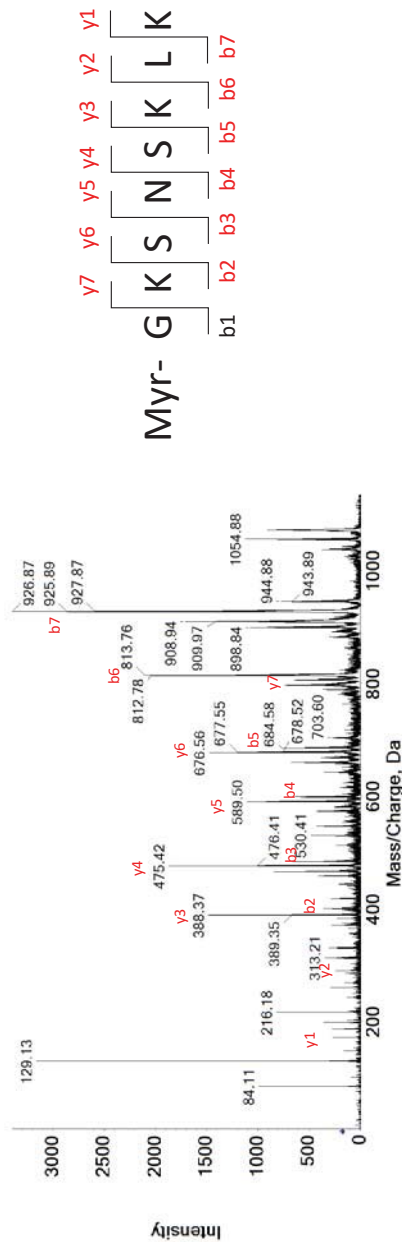
GKSNSKLIK



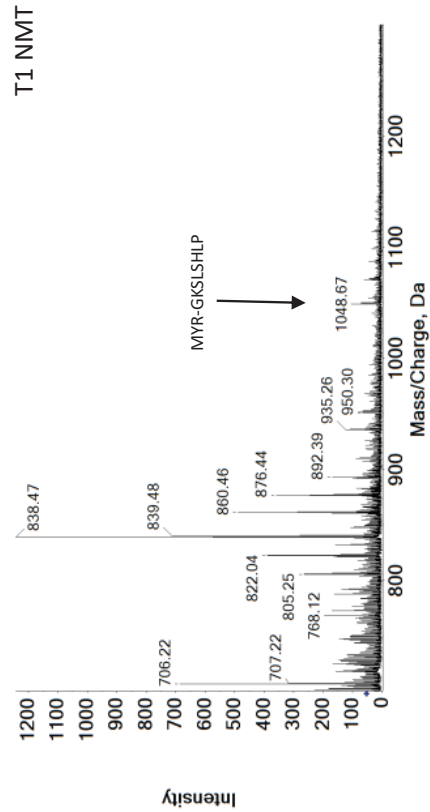
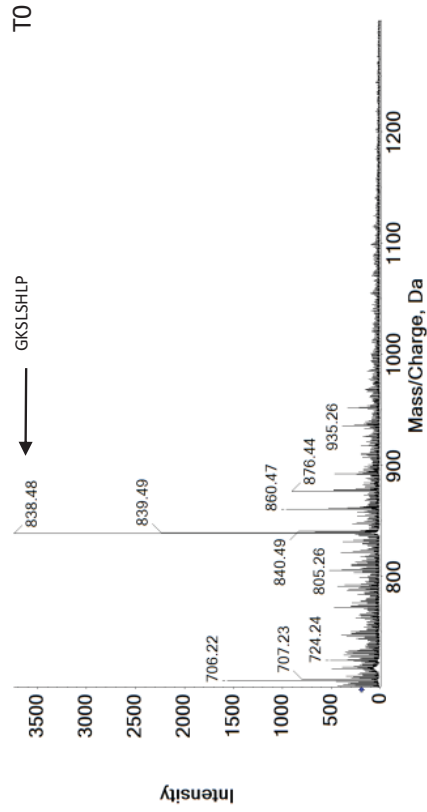
MSMS T1 Myr-GKSNSKLIK (1073) (1)



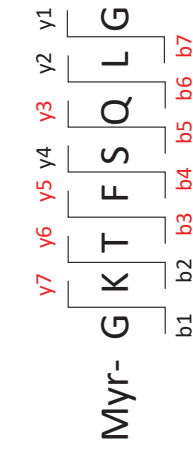
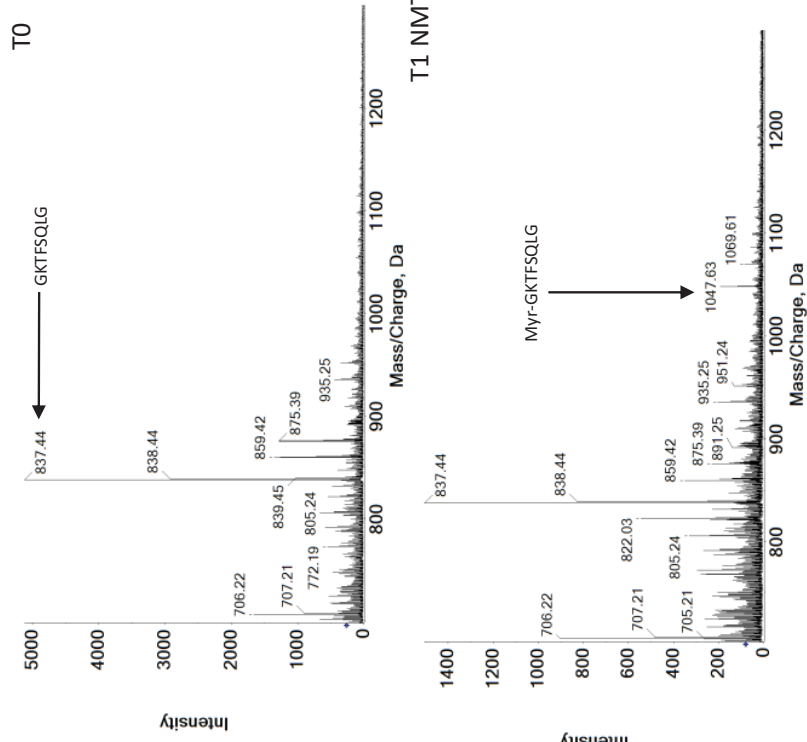
MSMS T1 Myr-GKSNSKLIK (1073) (2)



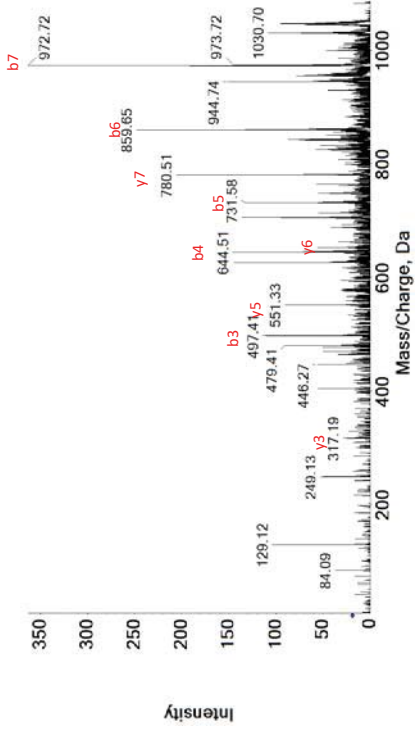
GKSLSHLP



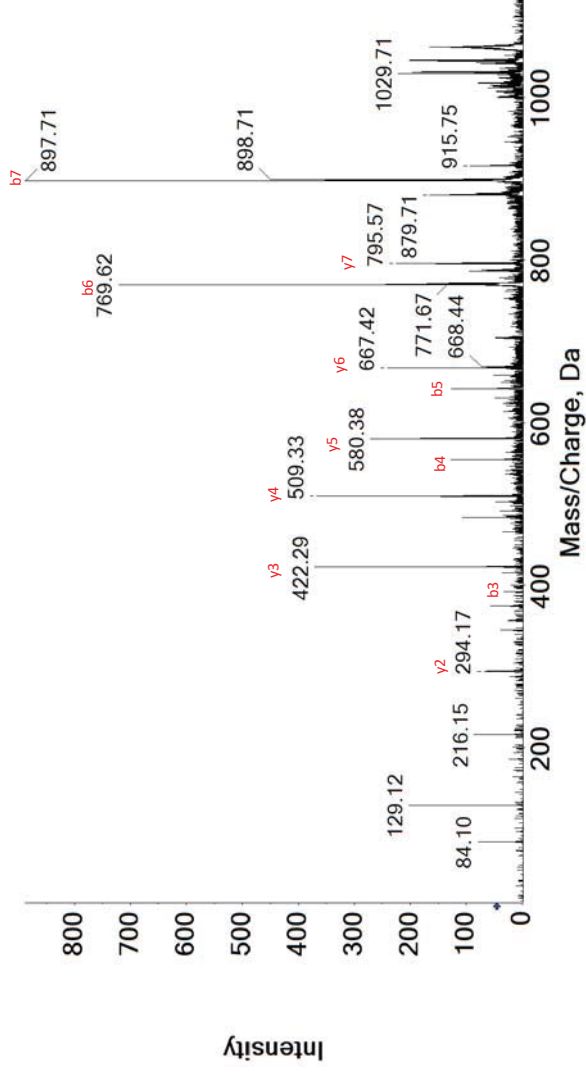
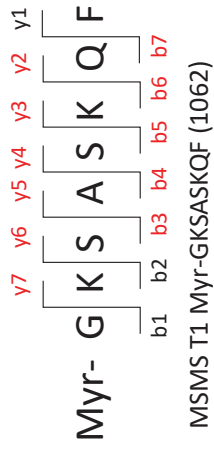
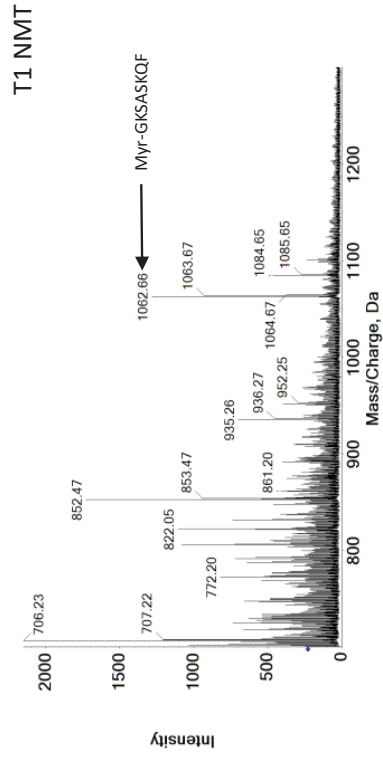
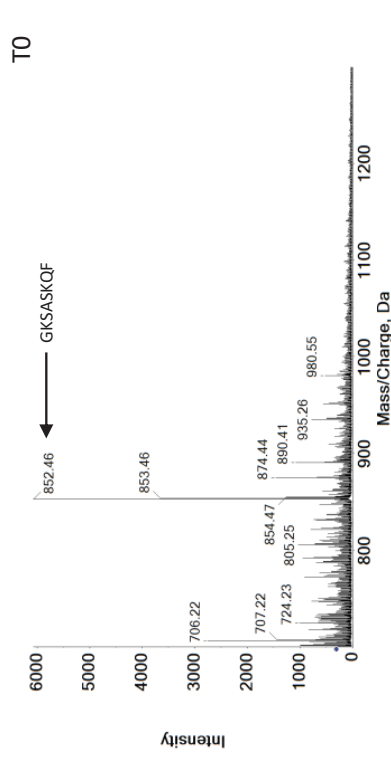
GKTFSQLG



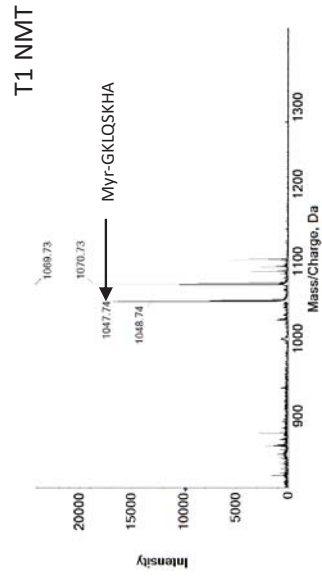
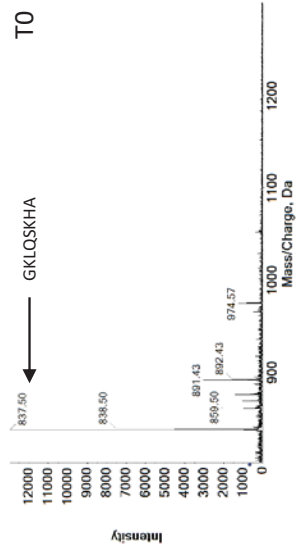
MSMS T1 Myr-GKTFSQLG (1047)



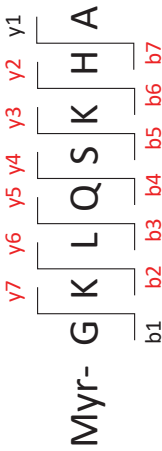
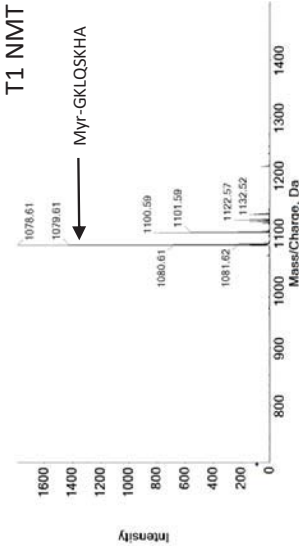
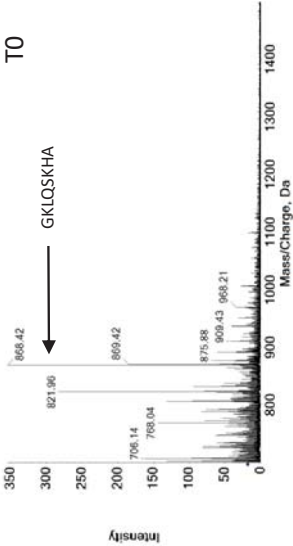
GKSASKQF



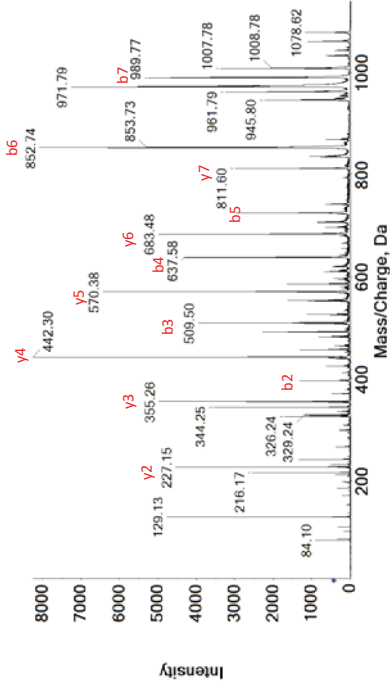
GKLOQSKHA



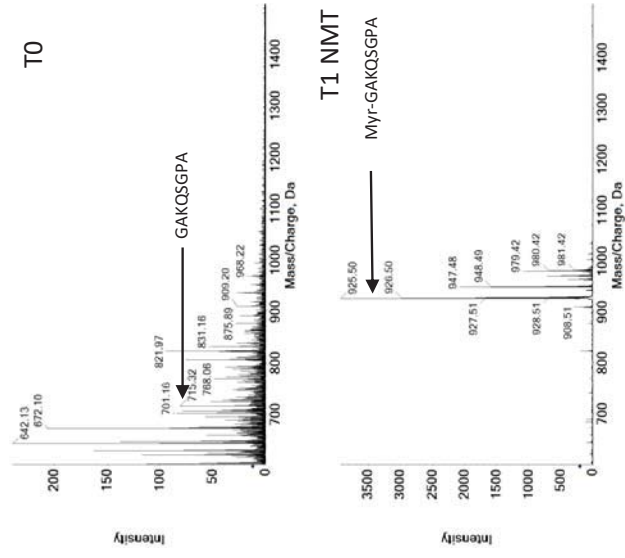
GKLQSKHA



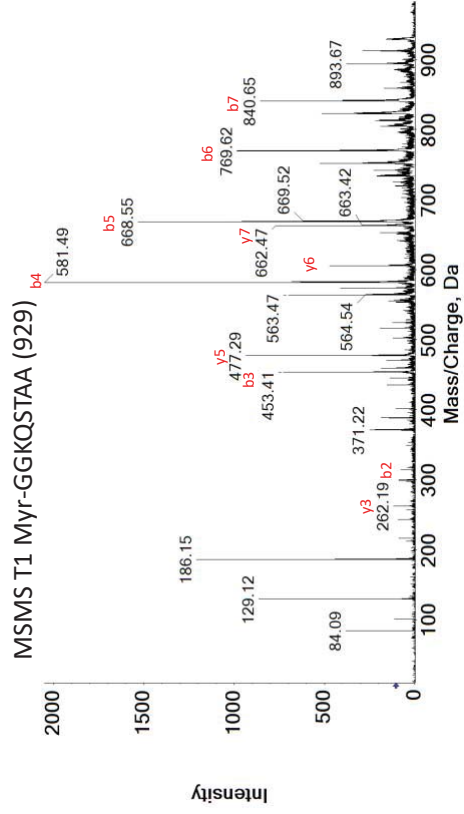
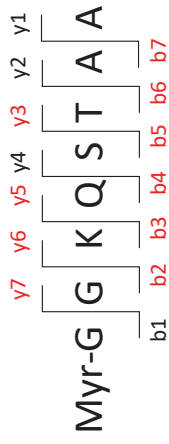
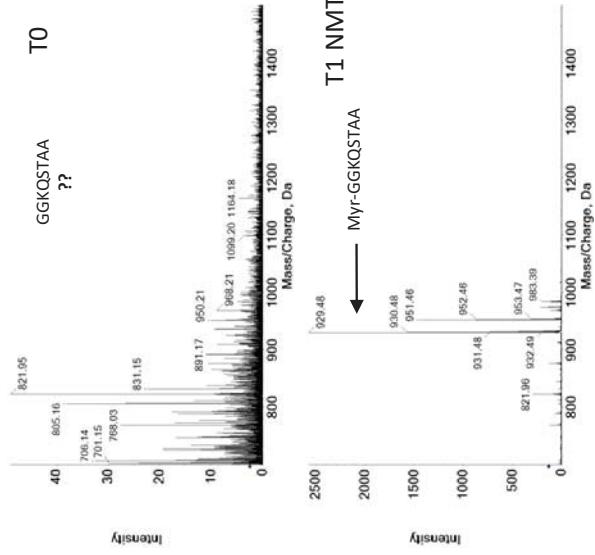
MSMS T1 Myr-GKLQSKHA (1078)



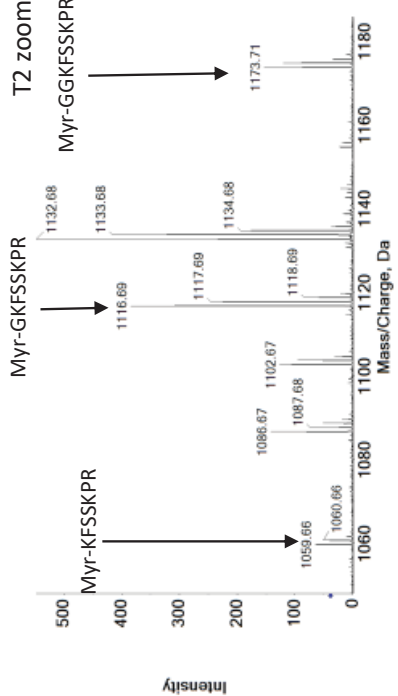
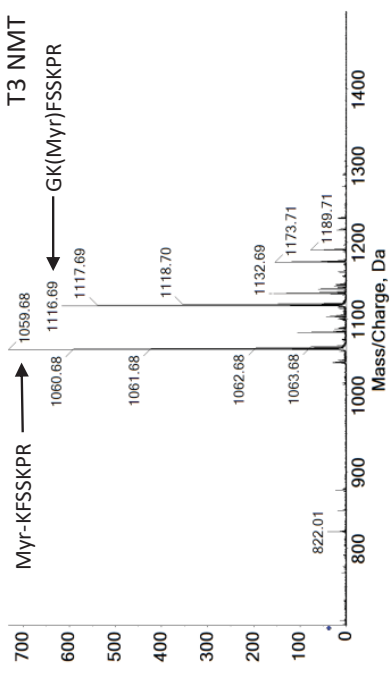
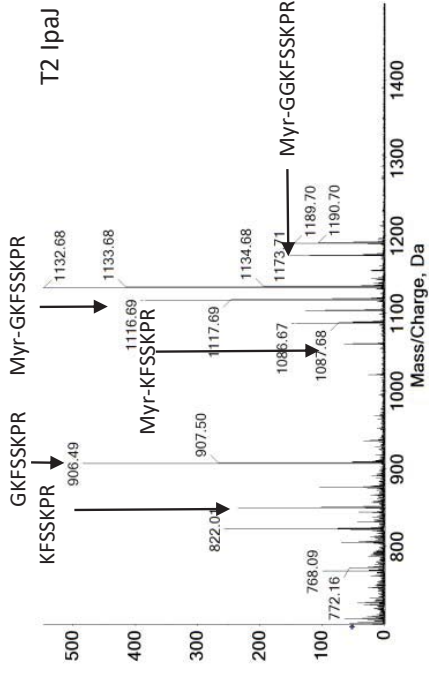
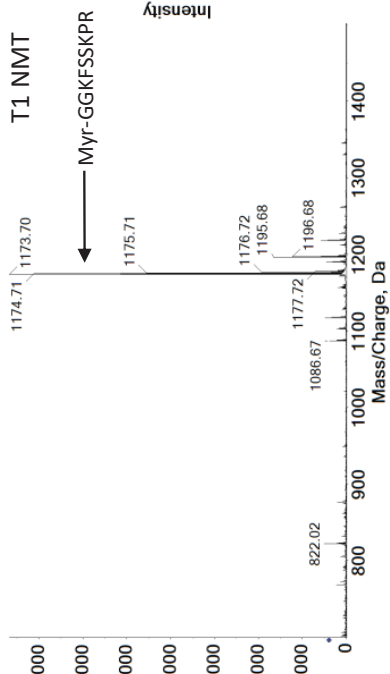
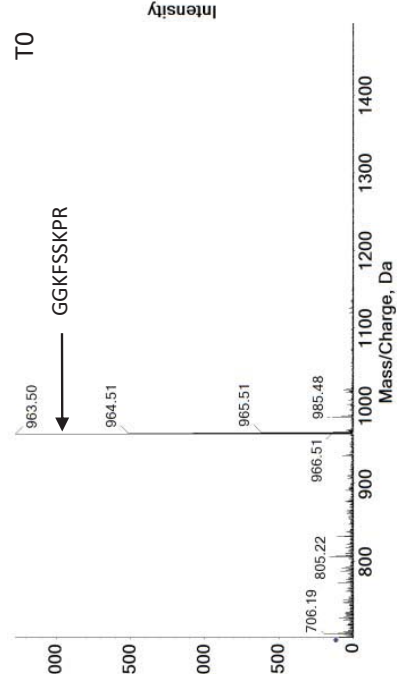
GAKQSGPA



GGKQSTAA

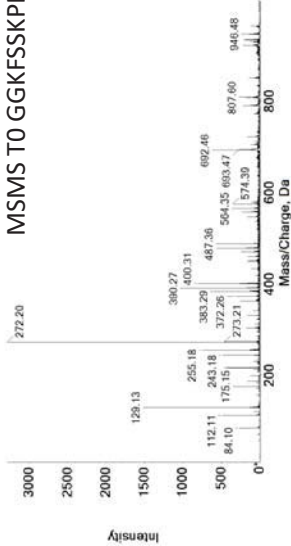


GGKFSSKPR

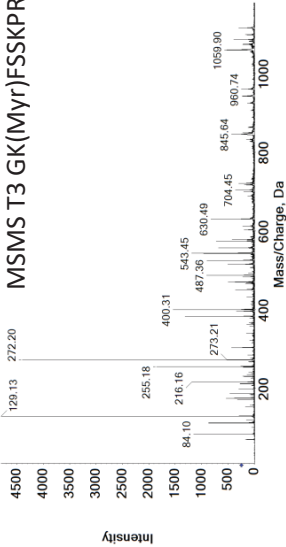


GGKFSSKPR

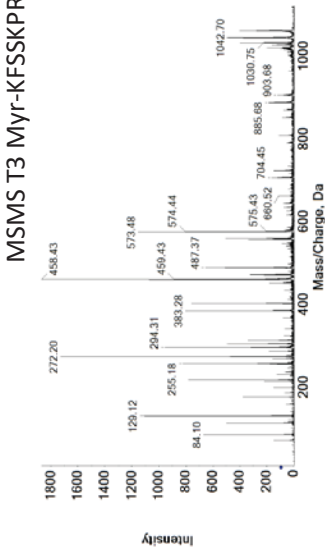
MSMS TO GGKFSSKPR (963)



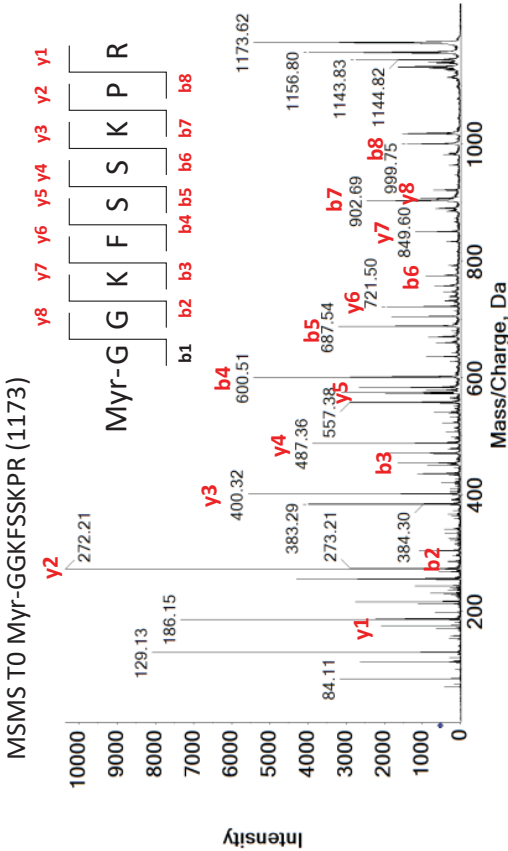
MSMS T3 GK(Myr)FSSKPR (1116)



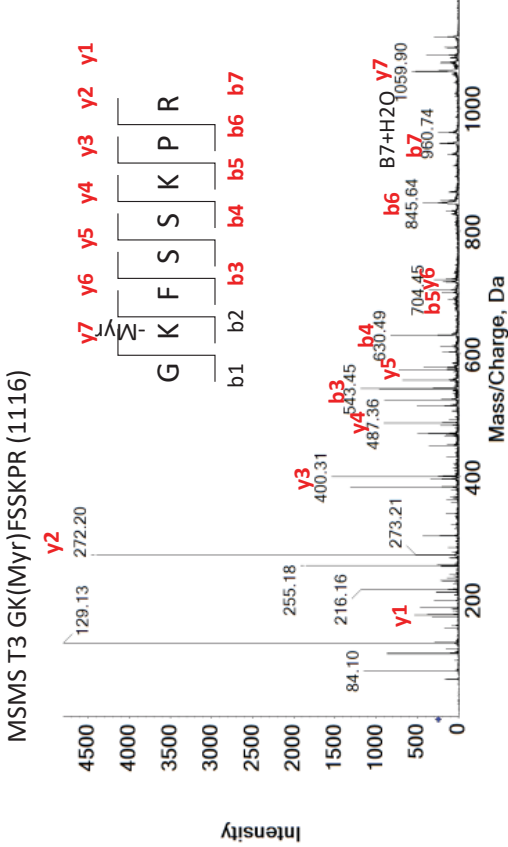
MSMS T3 Myr-KFSSKPR (1059)



MSMS TO Myr-GGKFSSKPR (1173)



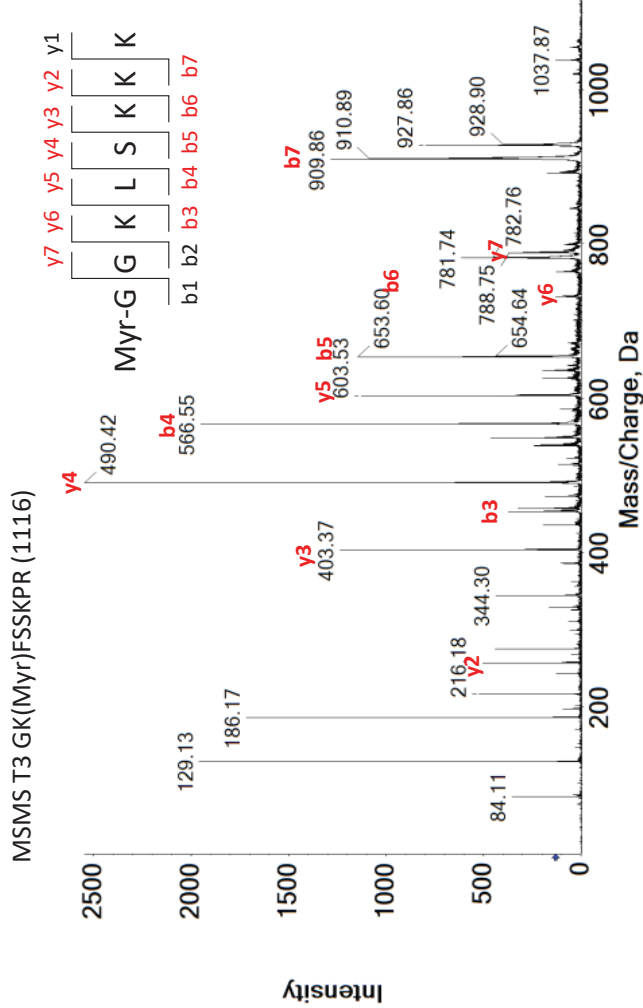
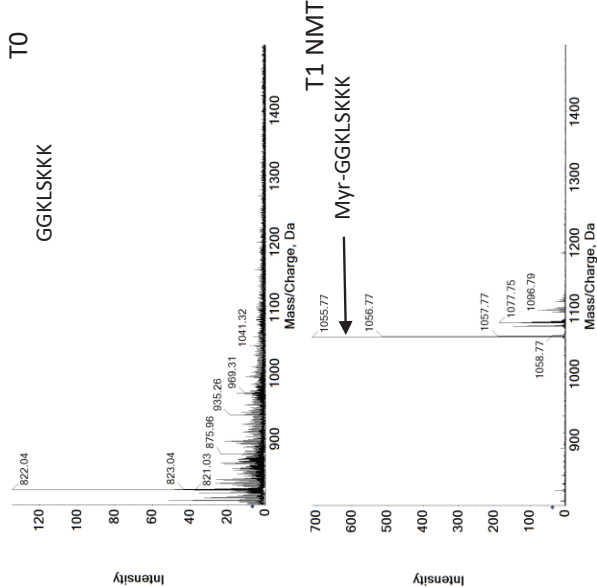
MSMS T3 GK(Myr)FSSKPR (1116)



K20 1173	Mass/Charge (Da)	Intensity
4372	175,1503	2072 y1
8355	272,1954	10360 16 y2
10293	325,2772	320 b2
12791	400,3083	5570,51 y3
14427	453,4173	1645,961 b3
15423	487,3767	3883,765 y4
17825	574,4145	3119,059 y5
18508	600,4936	5425,726 b4
20684	687,5204	3206,745 b5
21495	721,4933	1923,922 y6
22724	774,5721	912,7844 b6
24390	849,6013	1189,02 y7
25523	902,6548	2460,706 b7
25606	906,606	1046,588 y8
27511	999,7244	1402,039 b8

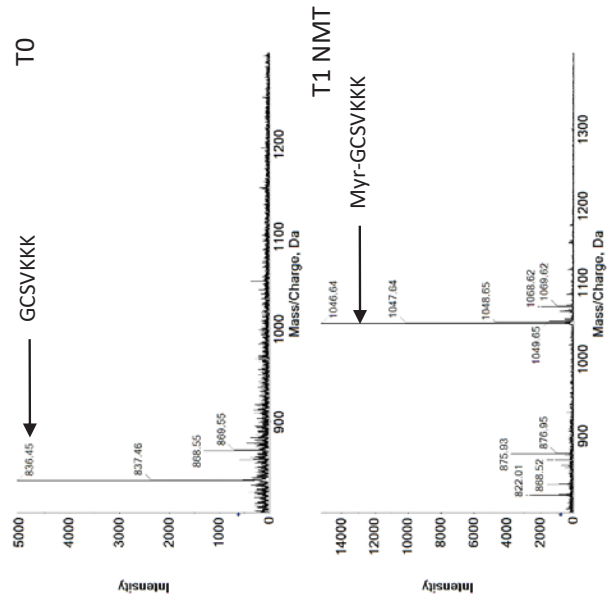
K22 1116	Mass/Charge (Da)	Intensity
2008	175,1541	538,35 y1
2891	216,1627	1182,90 SK
3999	255,1691	1852,24 y2-NH3
4553	272,2057	4394,67 y2
5462	303,1855	437,33 SKK
7699	383,2735	1310,43 SSKP-NH3
8218	400,3085	1541,02 y3
9657	450,2645	359,53 FSSK
10244	470,3054	500,24 y4-NH3
10729	487,3455	905,41 y4
11494	515,4305	506,04 a3
12266	543,4355	1191,69 b3
13097	574,384	724,55 y5
14532	630,4798	822,75 b4
16634	717,5125	190,59 b5
16727	721,4527	298,67 y6
19445	845,6232	432,31 b6
21453	942,7505	231,69 b7
21812	960,7414	246,90 b7+H2O
23625	1059,864	569,10 y7

GGKLSKKK

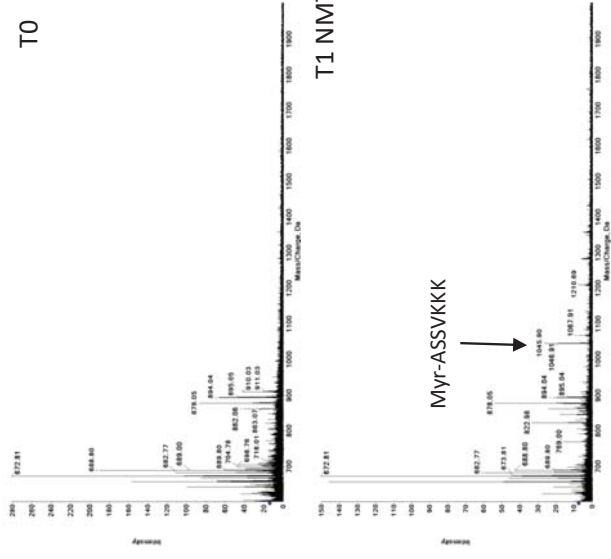


y2	2642	275,2397	400,9412
y3	5363	403,3621	1237,961
b3	6526	453,4303	375,8431
y4	7591	490,8696	179,6078
b4	9598	566,5328	1952,314
y5	10561	603,5288	1129,098
b5	11824	653,5715	1144,941
y6	13686	731,6654	131,2941
b6	14812	781,6852	615,8431
y7	14969	788,6453	359,6863
b7	17555	909,8929	1265,882

GCSVKKK



ASSVKKK

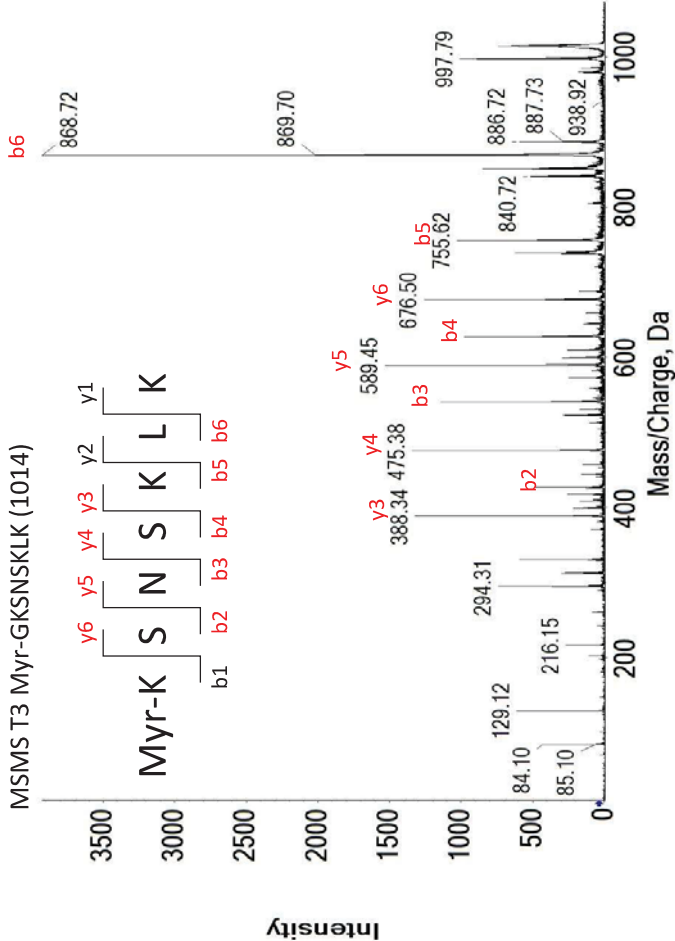
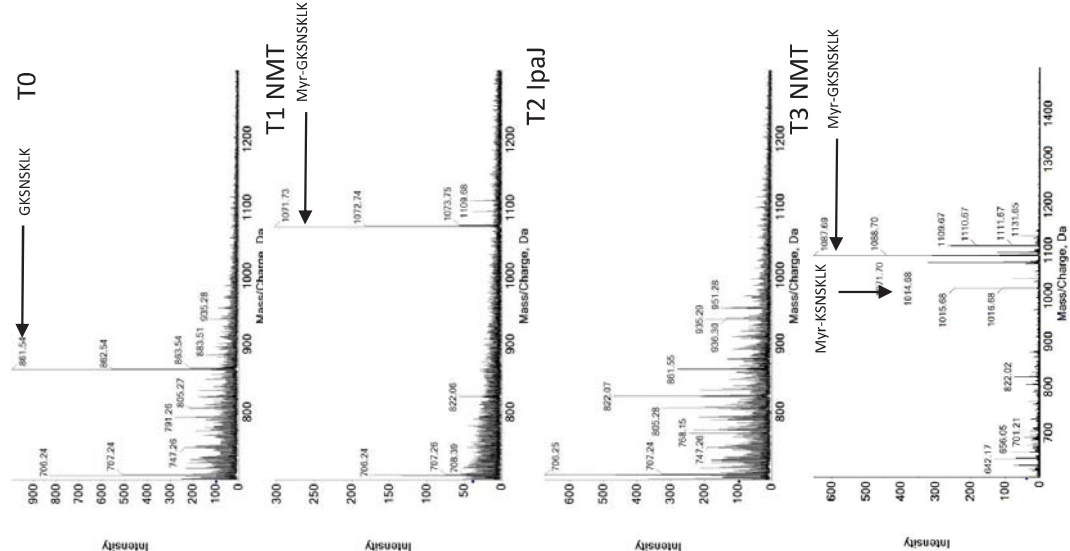


VIII.2.7. Données supplémentaires 1b

Sup Dataset 1b

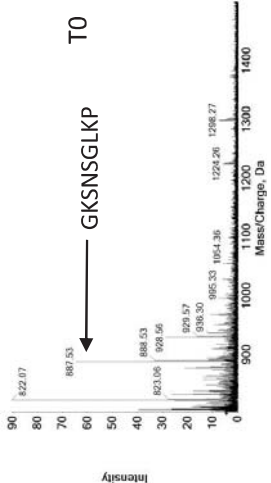
Peptide sequence	MS data page in SupDataset 1b
GKNSK L K	2
GKNS G LKP	3
GKNS { DAP } LKP	4
GKNS { DAB } LKP	5
GKNS { ORN } LKP	6
GKNS { HCY } LKP	7
GKNS { bhLYs } LKP	8
GKVWS Q GV	9
GKSN A KLKP	10

GKSN SKLK

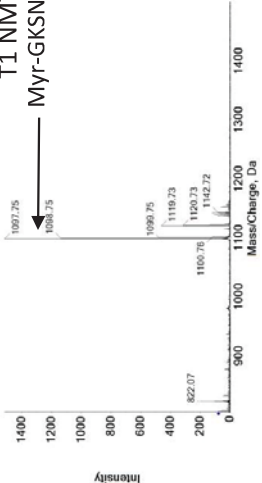


F9_MSMS_1014.6000_2	ion		
pic	Mass/Charge (Da)	Intensity	ion
4036	388,3268	1320,784	y3
5032	426,3984	484,2353	b2
6101	475,3693	1344,157	y4
7622	540,4544	1142,118	b3
8831	589,4373	1535,059	y5
9779	627,5093	978,5098	b4
10957	676,5035	1258,196	y6
12771	755,6361	1029,804	b5
15221	868,7002	3926,588	b6

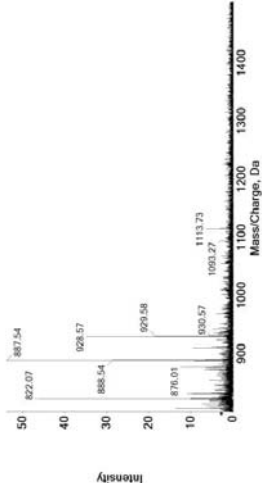
GKNSGSLKP



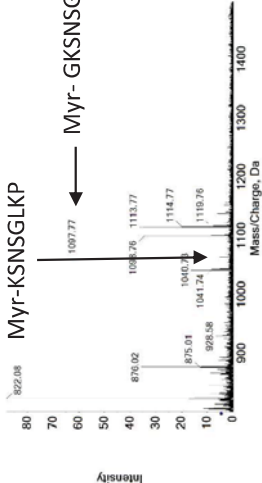
T1 NMT
Myr-GKNSGSLKP



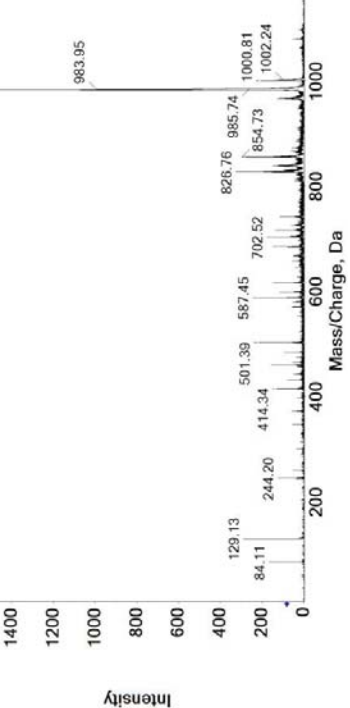
T2 IpaJ



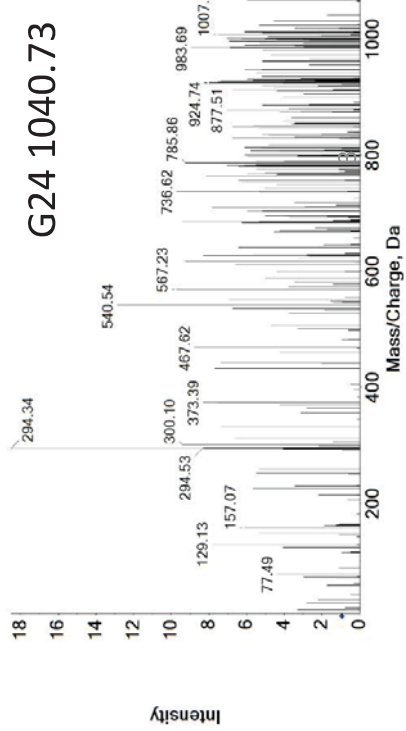
T3 NMT



MSMS T3 Myr- GKNSGSLKP (1097,74)



MSMS T3 Myr- KNSGSLKP (1040)



G24 1040.73

T0

Intensity

Mass/Charge, Da

916.52

917.52

822.03

706.20

717.18

822.21

847.20

954.48

955.48

1074.67

1294.76

1426.86

GKSNS[DAP]LKP

T1 NMT

Intensity

Mass/Charge, Da

1126.71

1127.72

1154.68

1165.67

1166.68

1228.72

1252.72

874.96

881.96

952.23

706.21

706.11

875.96

822.03

Myr-GKSNS[DAP]LKP

T2 IpaJ

Intensity

Mass/Charge, Da

1126.72

1127.72

1148.70

1149.70

1154.68

1165.67

1009.70

897.25

875.97

706.11

875.96

874.96

709.11

822.04

Myr-KSNS[DAP]LKP

T3 NMT

Intensity

Mass/Charge, Da

1126.72

1127.72

1128.72

1148.70

1069.70

875.96

822.04

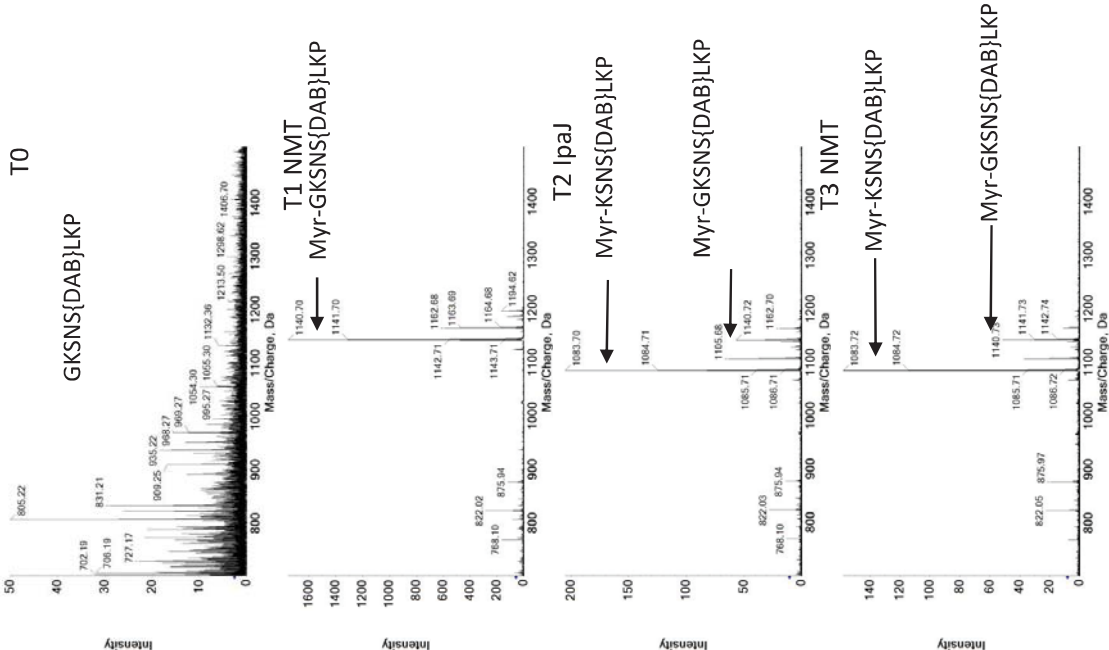
820.04

Myr-GKSNS[DAP]LKP

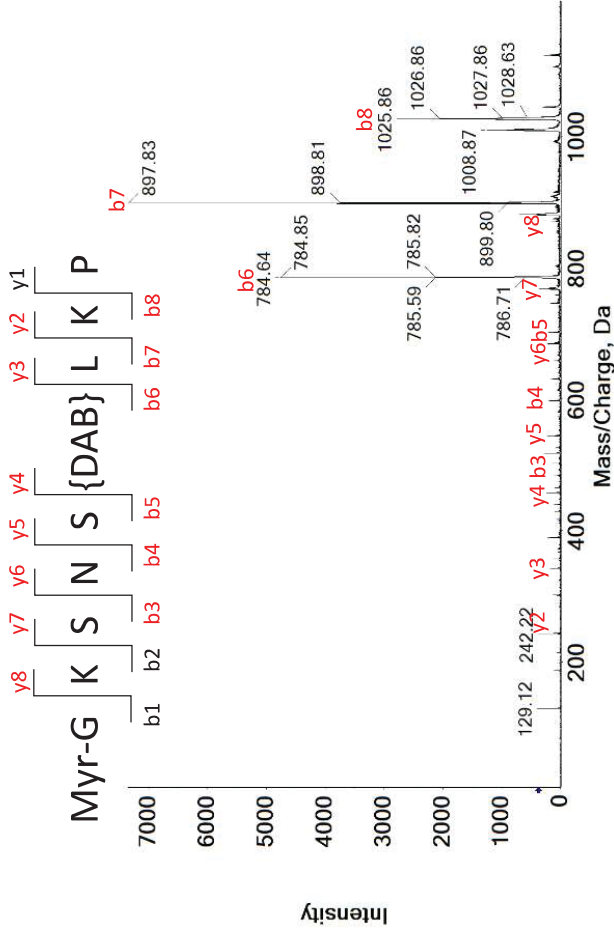


H24_MS
MS_1126
.7100_2

GKSNS{DAB}LKP

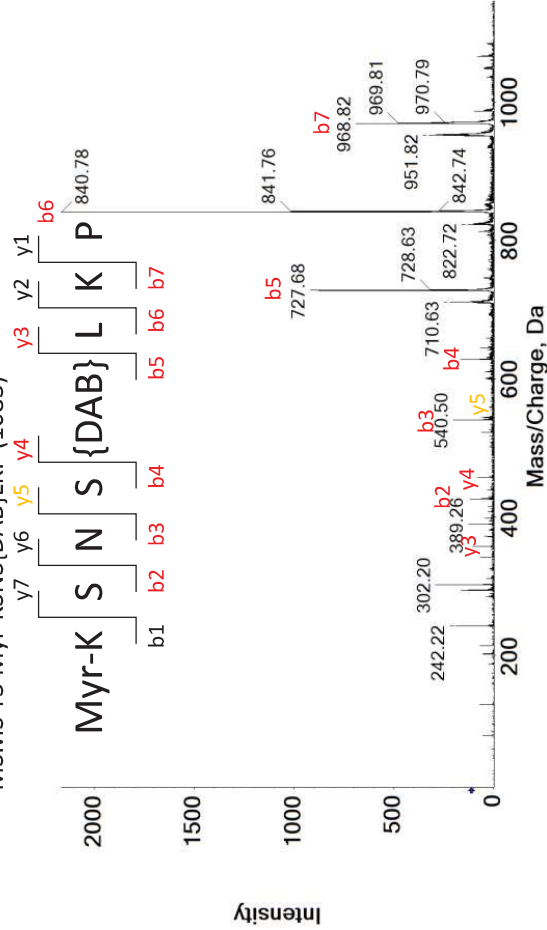


MSMS T1 Myr-GKSNS{DAB}LKP (1140)



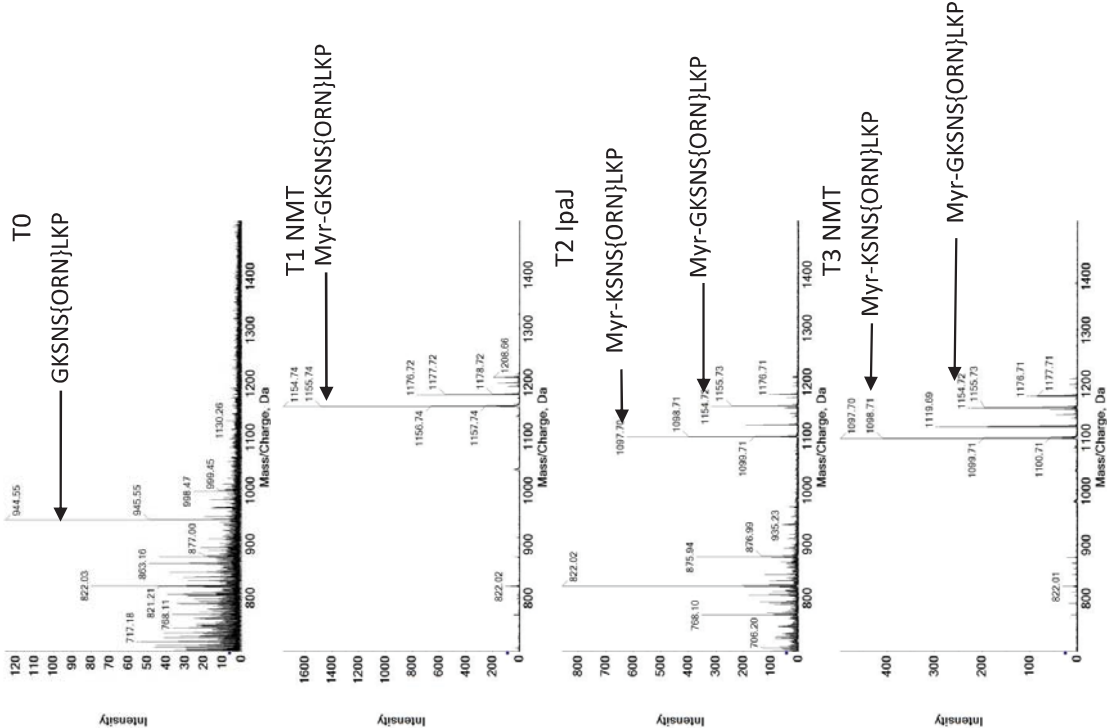
H19_MSMS_1140_6900_2	Mass/Charge (Da)	Intensity
2456	242,2261	376,9412 LK
2532	244,2031	115,1373 y2
4318	357,2703	131,6078 y3
6462	457,3843	234,8235 y4
7032	483,4075	116,8627 b3
7891	517,3342	283,6078 KNSK(-28)
8561	544,3931	224,3137 y5
9868	597,4956	180,2353 b4
11358	658,4525	103,8431 y6
11985	684,5401	220,3922 b5
12393	701,5362	204,3922 c5
13419	745,4974	157,6471 y7
14318	784,8348	4759,687 b6
16184	869,7394	146,8235 a7
16271	873,8047	141,1765 y8
16781	897,8309	7348,549 b7
19047	1008,825	1253,177 b8-NH3
19384	1025,869	2782,431 b8

MSMS T3 Myr-KSNS{DAB}LKP (1083)

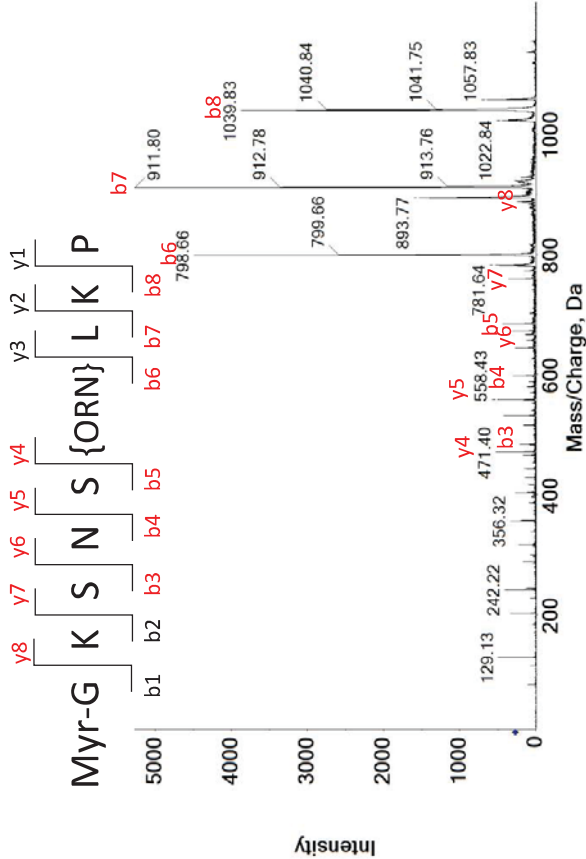


H21_MSMS_1083_7100_2	Mass/Charge (Da)	Intensity
1610	357,2901	105,4118 y3
2513	426,3815	116,2353 b2
2908	457,3836	81,09805 y4
3906	540,4713	202,1961 b3
4011	544,3975	57,72549 y5
5351	627,5062	164,7059 b4
6940	710,5687	219,9216 b5-NH3
7330	727,6011	920,1569 b5
9182	822,6997	130,353 b6-H2O
9577	840,8204	2164,549 b6
11624	951,7894	324,0784 b7-NH3
11970	968,7888	688,3137 b7

GKSNS{ORN}LKP

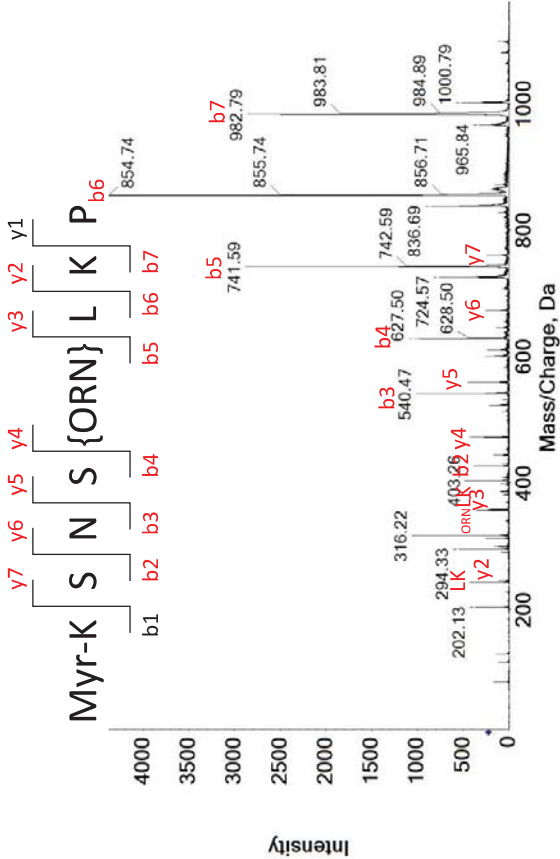


MSMS T1 Myr-GKSNS{ORN}LKP (1154)



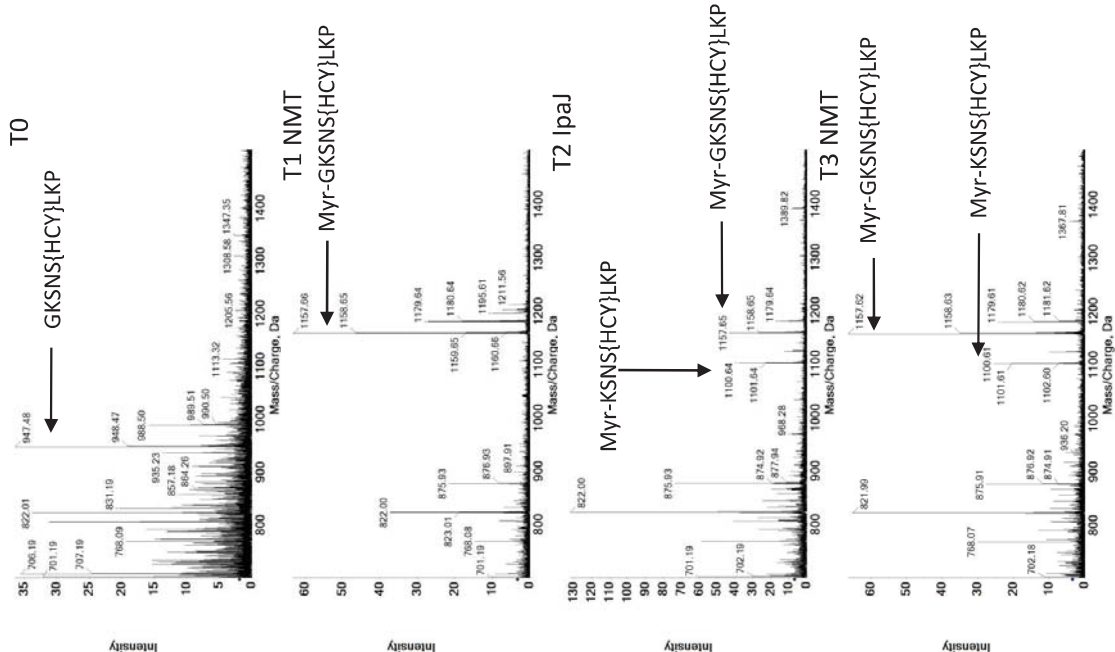
I3_MSMS	Mass/Ch	Intensity
_1154.73	arge (Da)	5743 471,3982 528,3137 y4
00_2		6033 483,4253 213,0196 b3
		7833 558,4199 504,7843 y5
		8842 597,5019 298,8235 b4
		10683 672,4924 301,6471 y6
		10976 684,5358 438,7451 b5
		12735 759,5251 359,6863 y7
		13621 798,6034 4492,079 b6
		15564 887,8157 240,3137 y8
		15689 893,7206 1531,765 b7-H2O
		16070 911,8423 5269,333 b7
		18657 1039,824 3872,628 b8

MSMS T3 Myr-KSNS{ORN}LKP (1097)

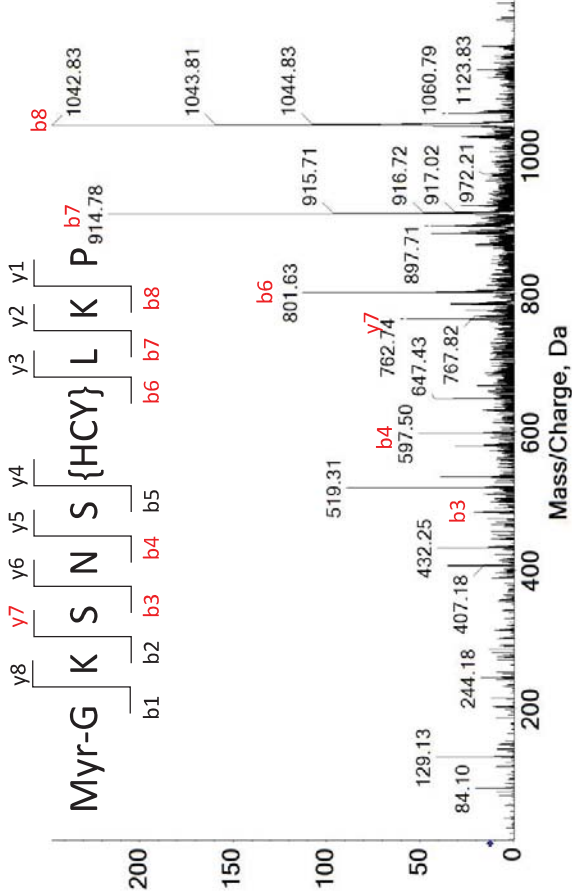


I5_MSMS	Mass/Cha	Intensity
_1097.700	arge (Da)	921 242,2105 432,4706 LK
0_2		975 244,1879 116,2353 y2
		2121 316,2183 1062,902 NSK(-14)
		2755 356,3063 358,902 K(-14)LK
		2788 357,2847 229,6471 y3
		3759 403,2611 479,8431 SNSK(-14)
		4264 426,4007 383,8431 b2
		4651 443,383 167,5294 SK(-14)LK
		5194 471,4 426,8235 y4
		6713 540,4425 1013,804 b3
		7165 558,4264 452,2353 y5
		8818 627,5032 1036,235 b4
		9919 672,5188 246,7451 y6
		11532 741,5885 2886,588 b5
		11946 759,515 245,1765 y7
		14059 854,6556 4378,196 b6
		16703 982,7454 2862,275 b7

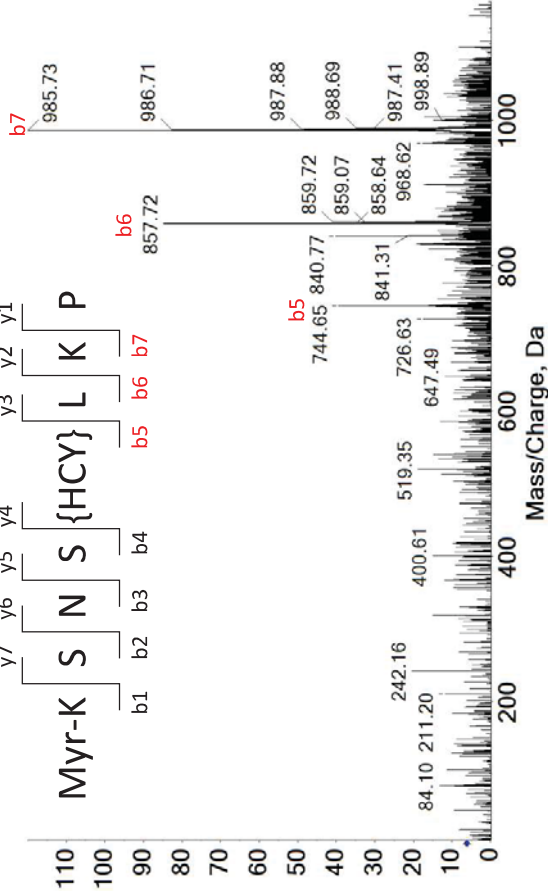
GKSNS{HCY}LKP



MSMS T1 Myr-GKSNS{HCY}LKP (1157)



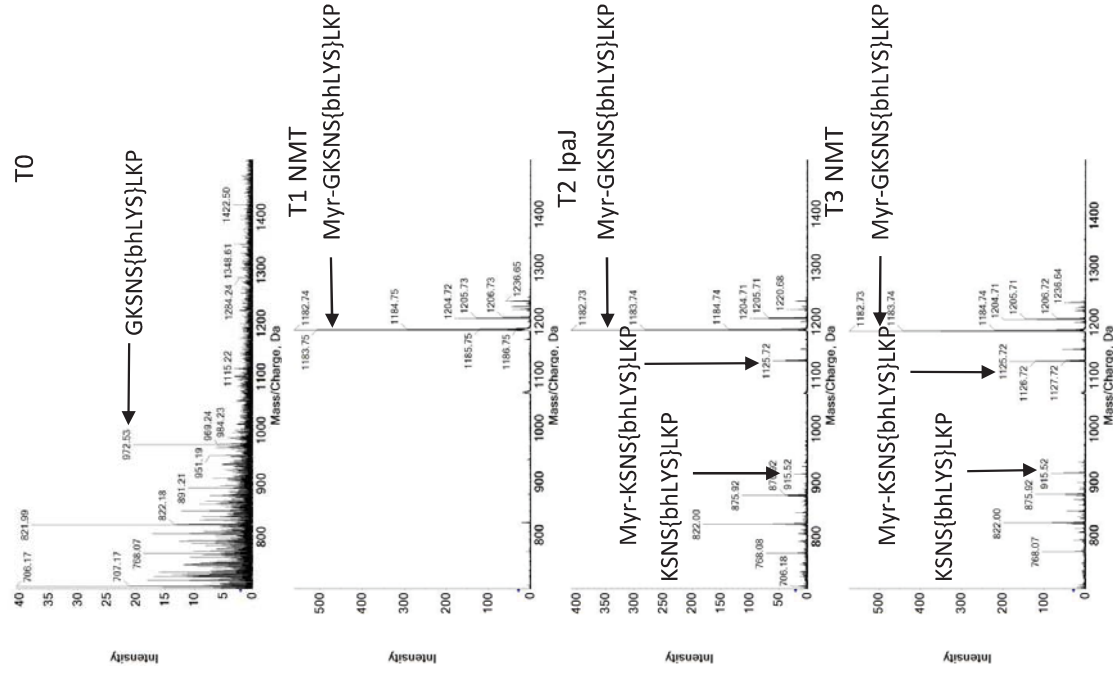
MSMS T3 Myr-KSNS{HCY}LKP (1100)



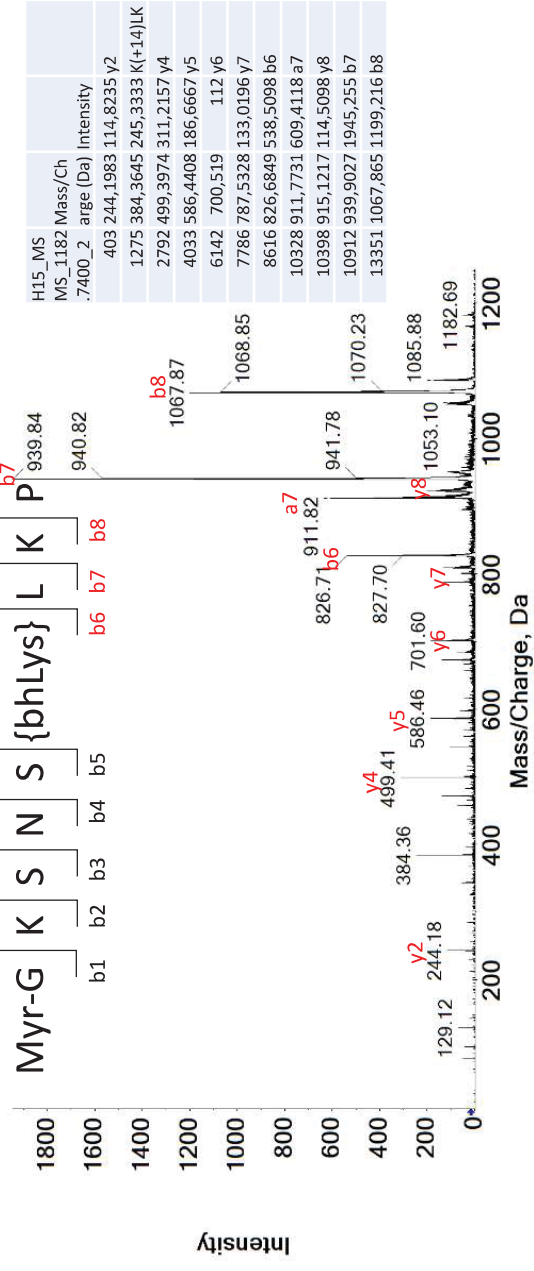
H6_MSMS_1157.6500_2	Mass/Charge (Da)	Intensity
820.00	406.15	35.29 SNSK(-11)
947.00	432.24	40.94 NSK(-11)L
1128.00	483.31	20.86 b ₃
1393.00	519.31	89.10 SNSK(-11)L
1490.00	534.45	39.06 KSNSK(-11)
1882.00	597.49	50.67 b ₄
2269.00	647.39	42.20 KSNSK(-11)L
2270.00	647.43	34.82 SNSK(-11)LK
3429.00	762.73	57.25 y ₇
3602.00	775.53	25.57 KSNSK(-11)LK
3744.00	784.48	33.88 b ₆ -NH ₃
4004.00	801.62	112.63 b ₆
4985.00	886.75	43.92 a ₇
5180.00	897.72	44.24 b ₇ -NH ₃
5489.00	914.78	216.63 b ₇
7034.00	1042.81	246.75 b ₈

H7_MSMS_1100.6000_2	Mass/Charge (Da)	Intensity
1983	726.6255	17.41177 b5-H2O
2148	744.6686	39.37255 b ₅
2803	840.7583	40.15686 b6-NH3
2999	857.7784	84.86275 b6
4161	985.7468	120 b7

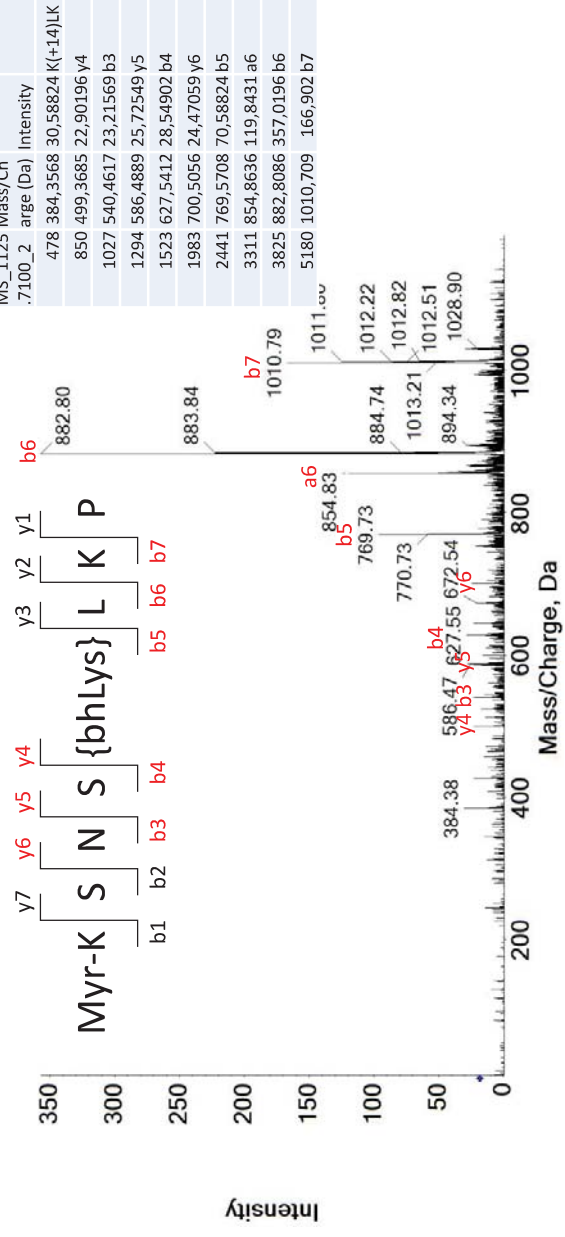
GKSNS{bhLYS}LKP



MSMS T1 Myr-GKSNS{bhLYS}LKP (1182)

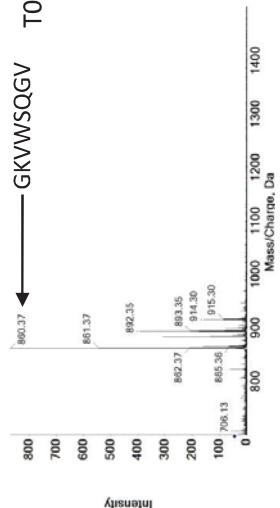


MSMS T3 Myr-KSNS{bhLYS}LKP (1125)

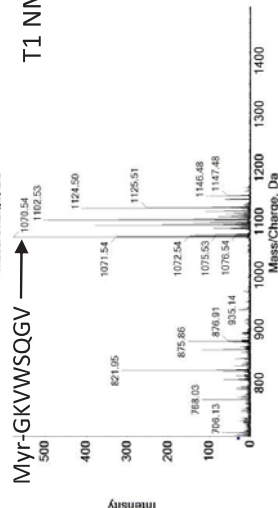


GKVWSQGV

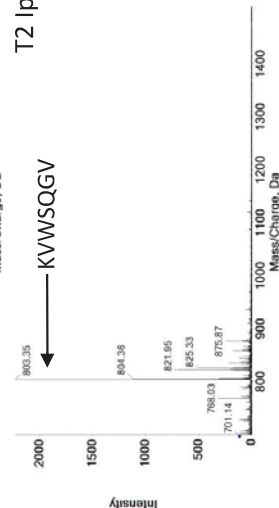
Intensity



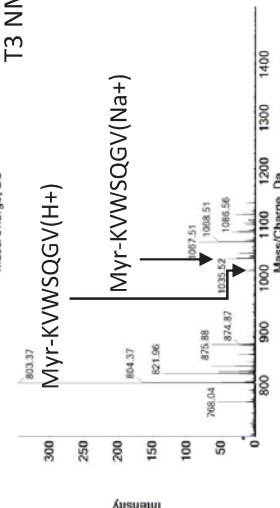
Myr-GKVWSQGV T1 NMT



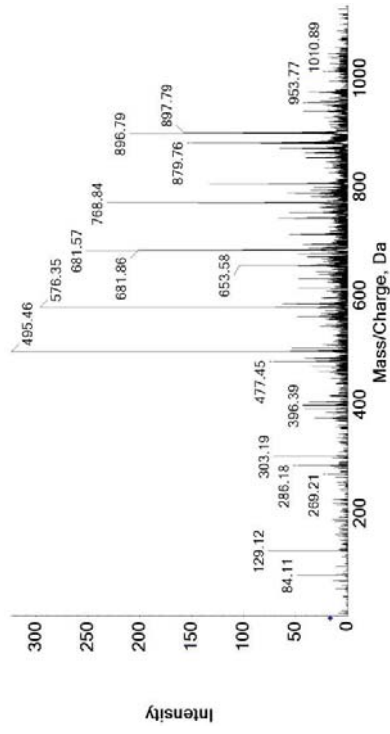
T2 IpaJ



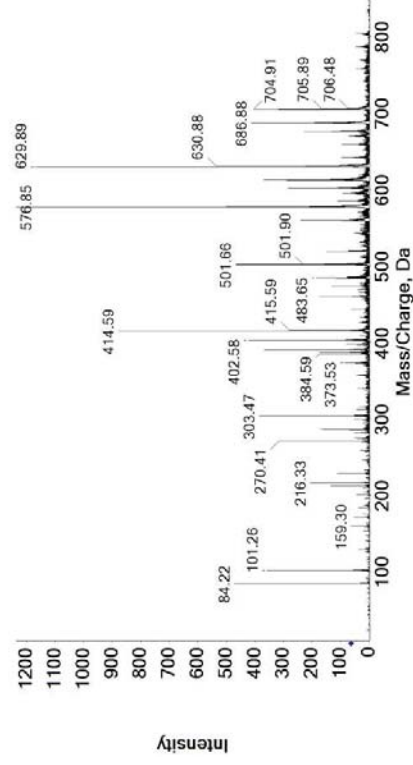
T3 NMT



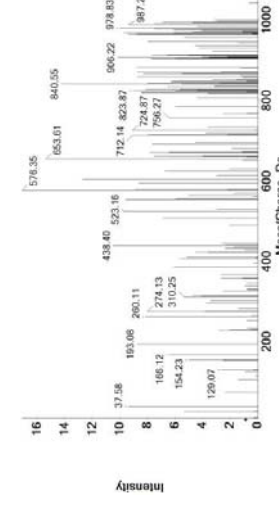
MSMS T1 Myr-GKVWSQGV (1070)



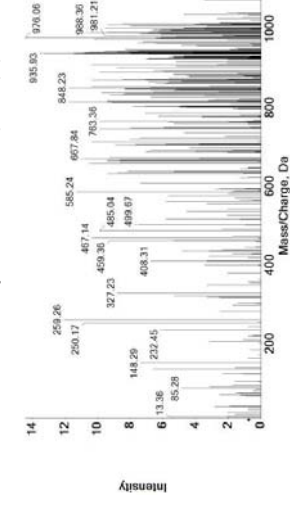
MSMS T2 KVWSQGV (803)



MSMS T2 Myr-KVWSQGV(H+) (1013)

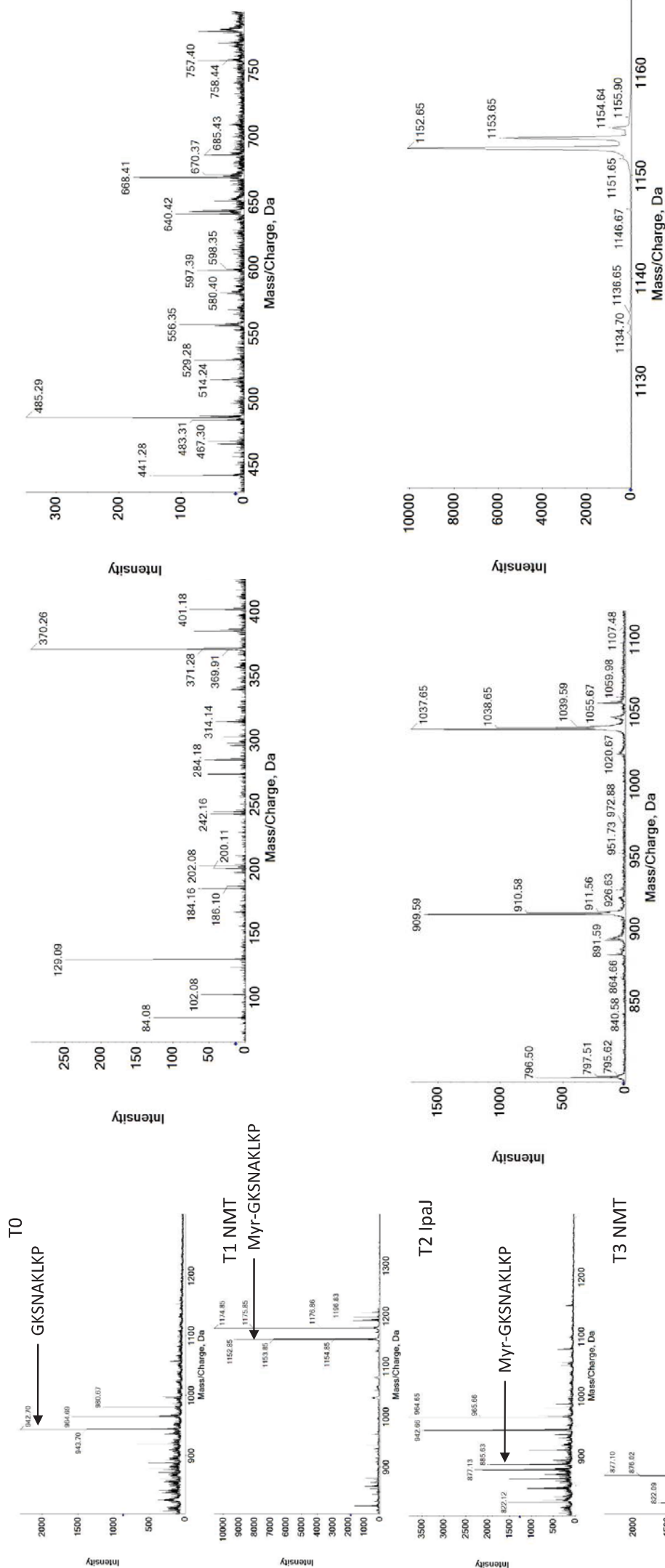


MSMS T2 Myr-KVWSQGV(Na+) (1035)



GKSNAKLKP

MSMS T1 Myr-GKSNAKLKP (1152)

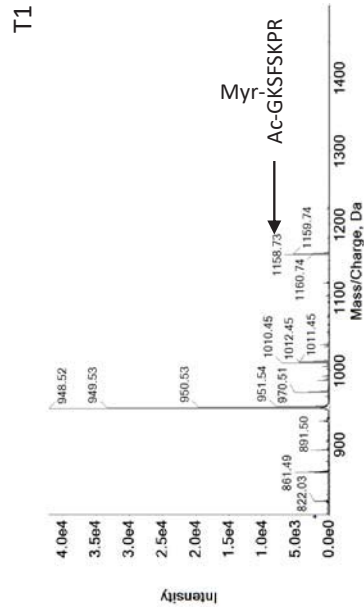
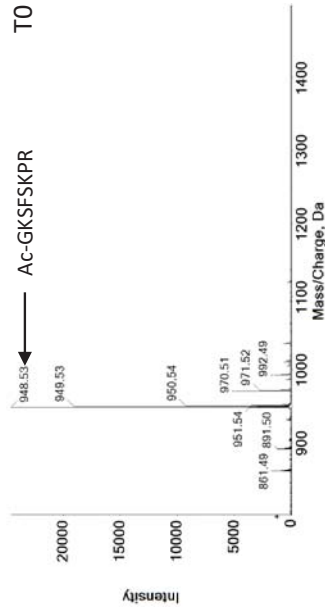


VIII.2.8. Données supplémentaires 1c

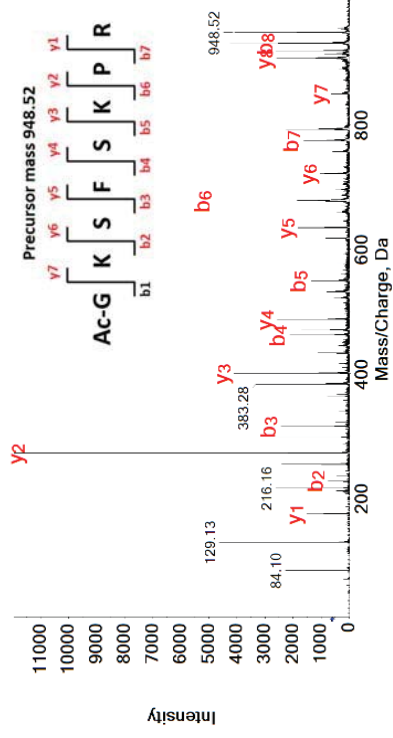
Sup Dataset 1c

Series ^a	Peptide sequence	Shortname (text & Figures)	Series sequence ^d	MS data page in Supplementary Dataset 1c
acGK	ac-GKSF ^{SKPR}	acGK		2
	ac-G{Orn}SF ^{SKPR}	acG[Orn]		3
	ac-G{Dab}SF ^{SKPR}	acG[Dab] ^c		4
	ac-G{D-Orn}SF ^{SKPR}	acG[D-Orn] ^c		5
	ac-GNCSF ^{SKPR}	acGN		6
	ac-GKVL ^{SKIF}	acGK	ARF6	7
XK	AKQNSKLR	A-HPCA	HPCA	8
	AKVL ^{SKIF}	A-ARF6	ARF6	9
	AKSF ^{SKPR}	AK		10
	AKPTSKDSGLK	TCS2	TCS2	11
	A{Orn}SF ^{SKPR}	A{Orn} ^c		12
	A{D-Orn}SF ^{SKPR}	A[D-Orn] ^c		13
	SKSF ^{SKPR}	SK		14
	MKSF ^{SKPR}	MK		15
	PKSF ^{SKPR}	PK		16
	G{D-Orn}SF ^{SKPR}	G[D-Orn] ^c		17
	ac-KSF ^{SKPR}	ack		18
	ac-{Orn}SF ^{SKPR}	ac[Orn] ^c		19
	ac-{D-Orn}SF ^{SKPR}	ac[D-Orn] ^c		20
	ac-{D-Dab}SF ^{SKPR}	ac[D-Dab] ^c		21
	ac-{Dap}SF ^{SKPR}	ac[Dap]		22
	K	K		23
	{Orn}SF ^{SKPR}	[Orn]		24
	{D-Orn}SF ^{SKPR}	[D-Orn] ^c		25
K	{Dab}SF ^{SKPR}	[Dab] ^c		26
	KQNSKLRP	K-HPCA	HPCA	27
	ac-KF ^{SKPR}	ackΔS		28
	KSFS{Orn}PR			29
	KSFS{Dab}PR			30
	KSFS{Dap}PR			31
	ac-[D-Orn]SNSKLR		NCS1	32
	KSNAKLRP		NCS1	33
	GGKSF ^{SKP}	GGK		34/35
	AKSF ^{SKPR}	AKK		36
ZXGK	ac-GGKSF ^{SKPR}	acGGK		37
	ac-GGGKSF ^{SKPR}	acGGGK		38
	ac-GGGKSF ^{SKPR}			39
	ac-GGGGKSF ^{SKPR}			40

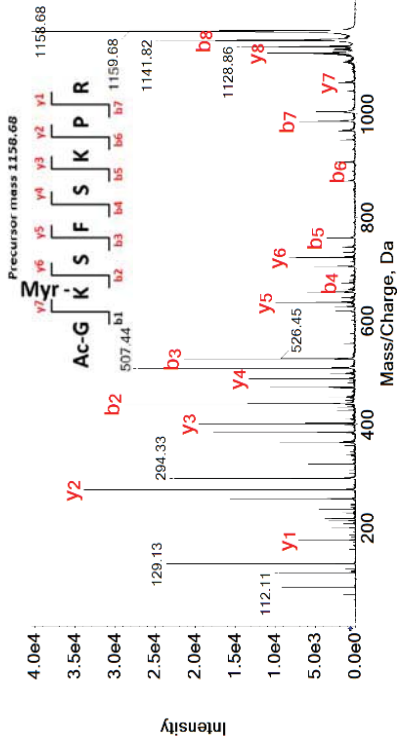
Ac-GKSF⁺SKPR



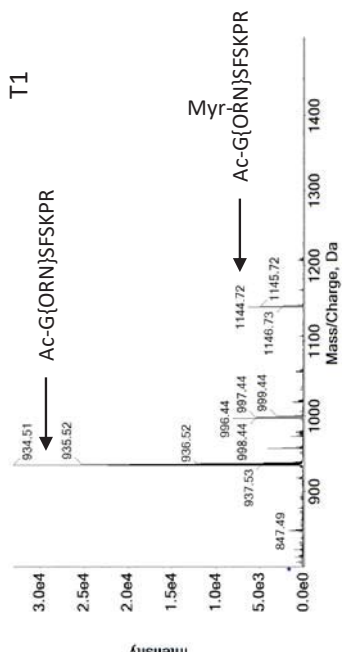
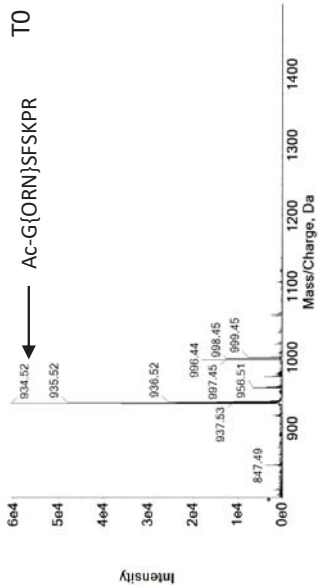
MSMS T0 Ac-GKSF⁺SKPR (948)



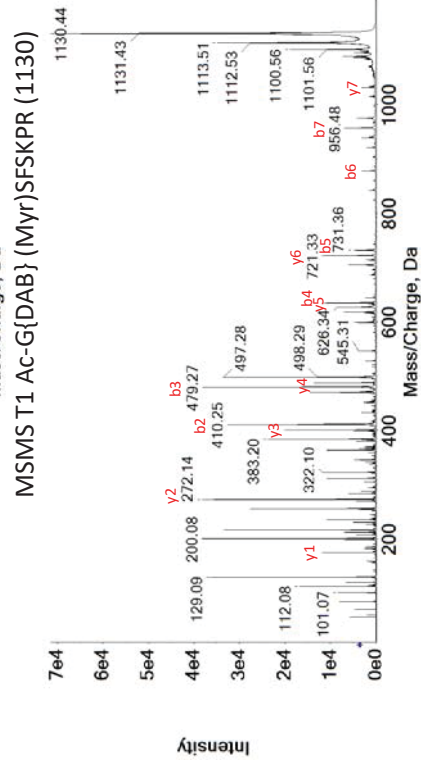
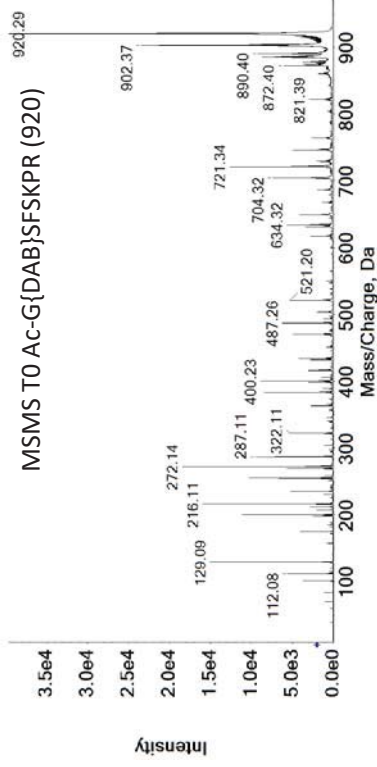
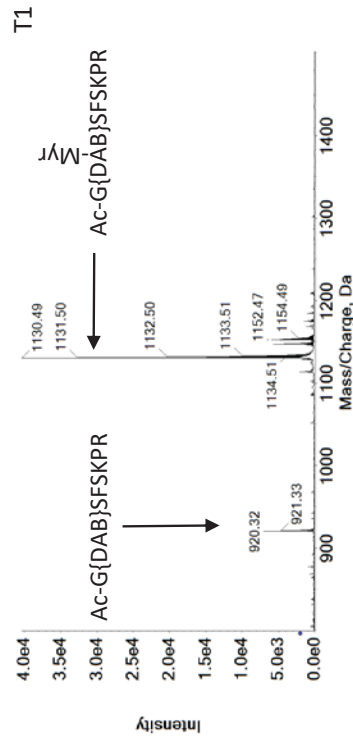
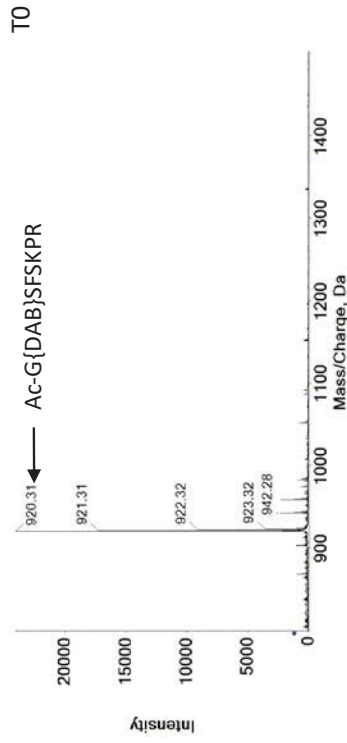
MSMS T1 Ac-GK(Myr)SFSKPR (1158)



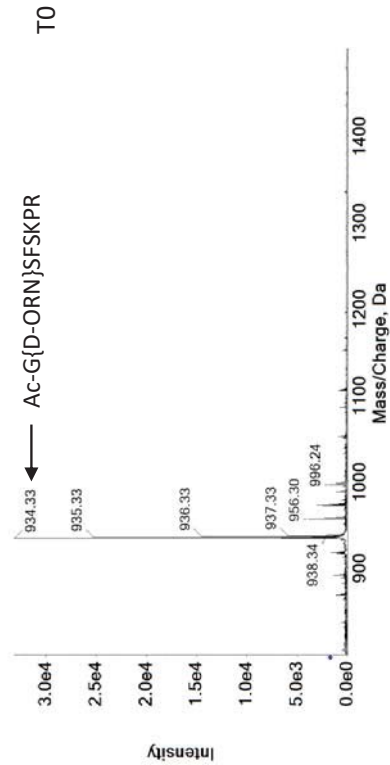
Ac-G{ORN}SFSKPR



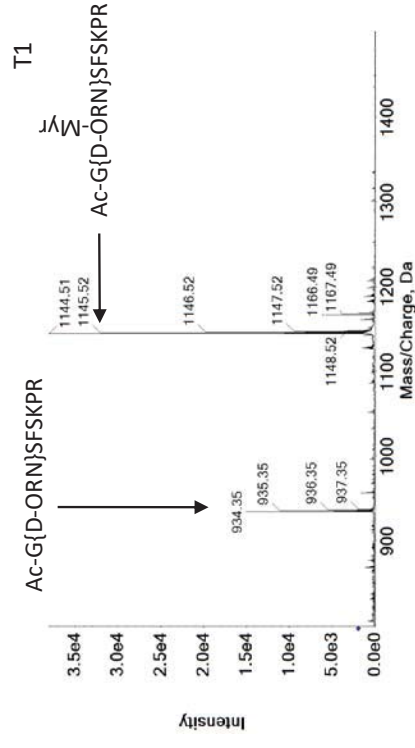
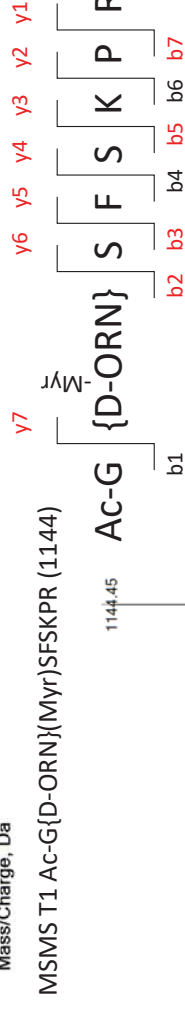
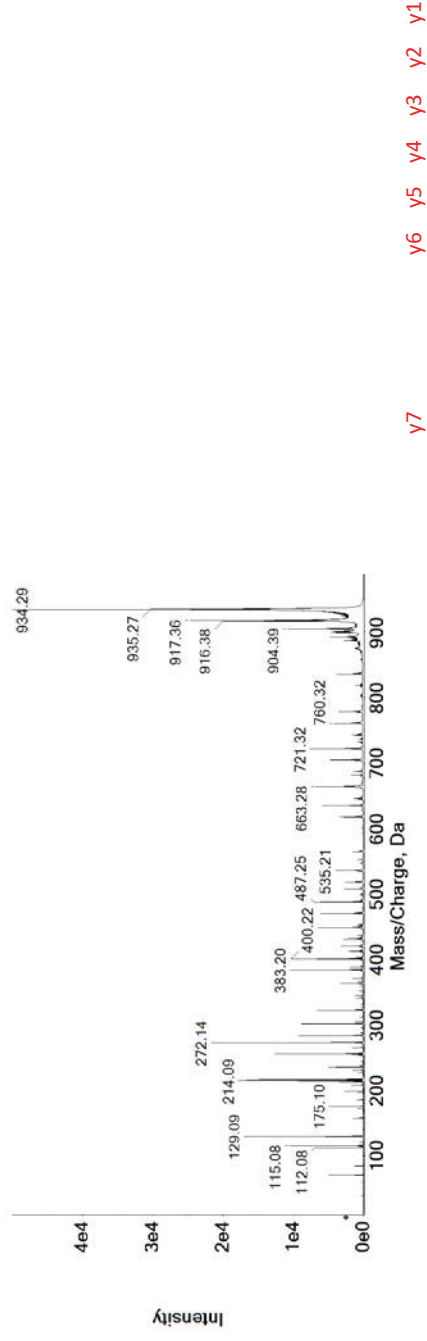
Ac-G{DAB}SFSKPR



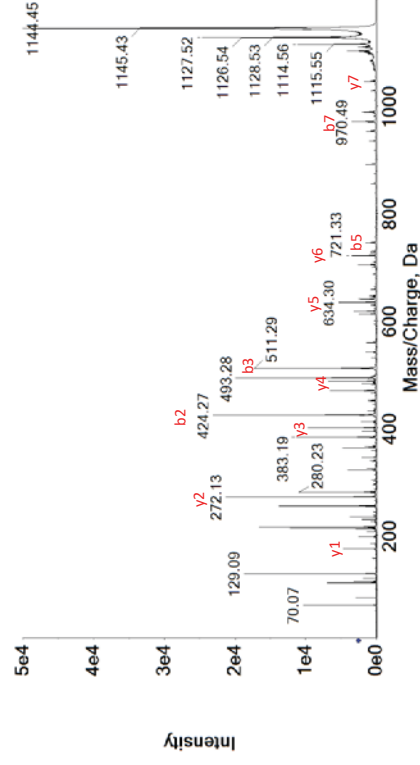
Ac-G{D-ORN}SFSKPR



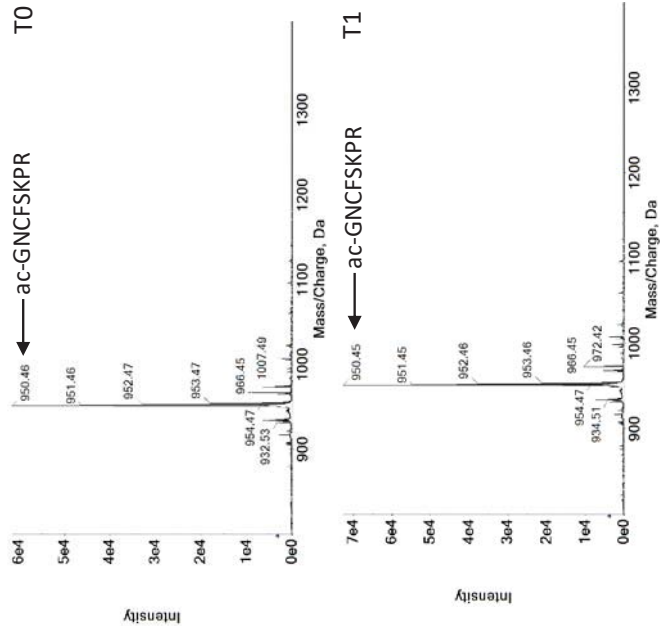
MSMS T0 Ac-G{D-ORN}SFSKPR (934)



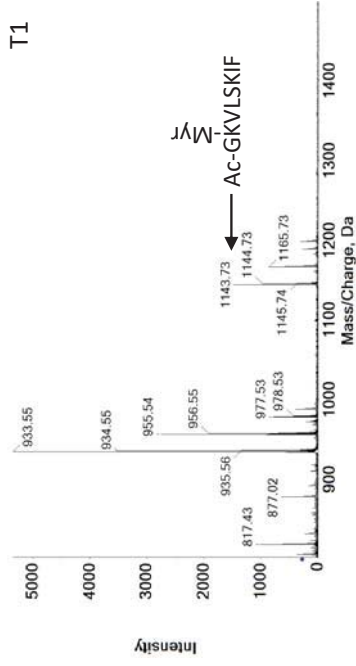
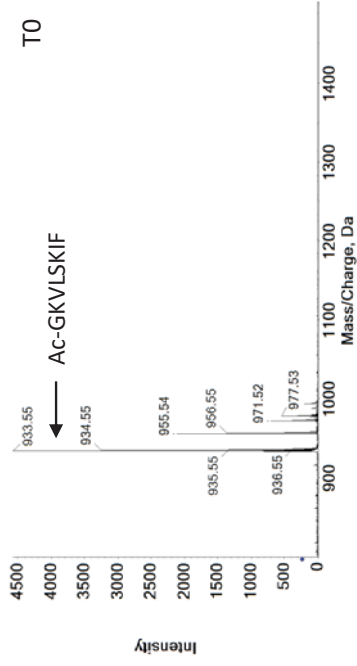
MSMS T1 Ac-G{D-ORN}{M_{Yr}}SFSKPR (1144)



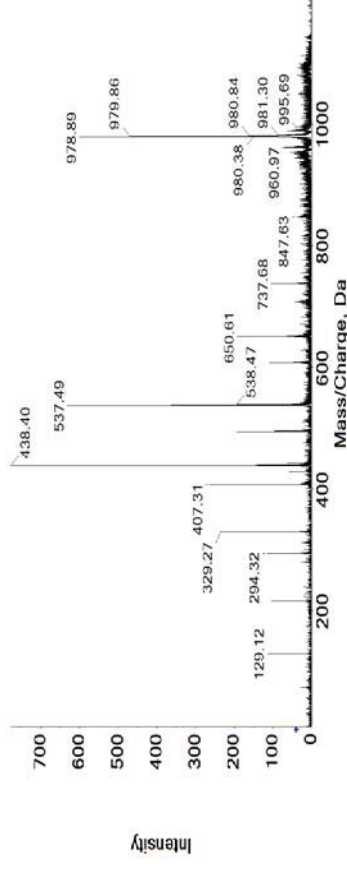
Ac-GNCFSKPR



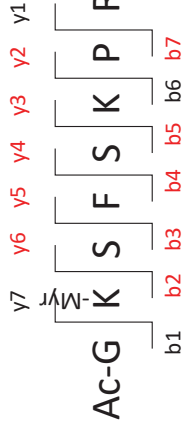
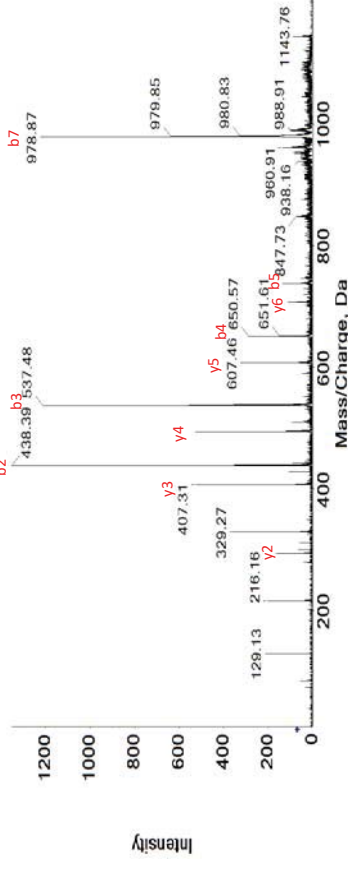
Ac-GKVLISKIF



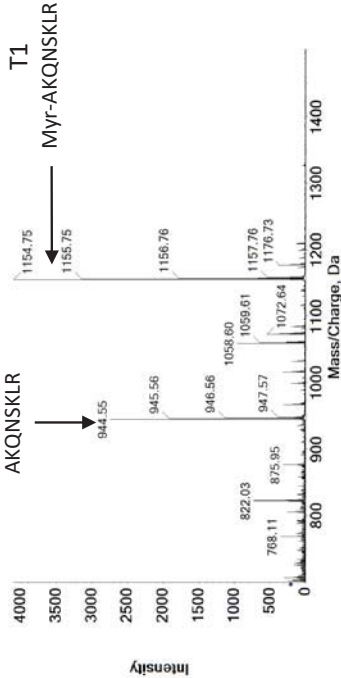
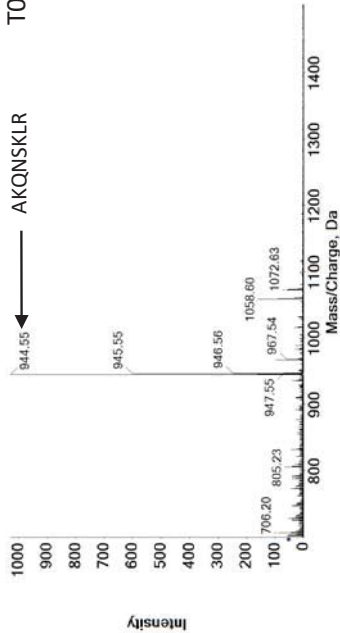
MSMS T1 ac-GK(My^r)VLISKIF (1143)



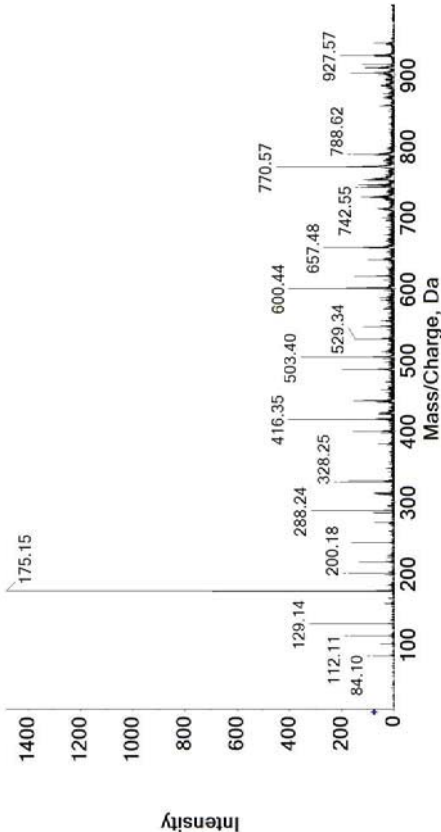
MSMS T1 ac-GK(My^r)VLISKIF (1143) (2)



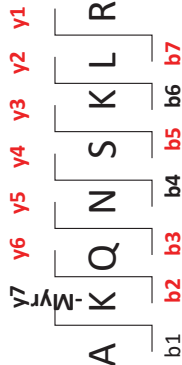
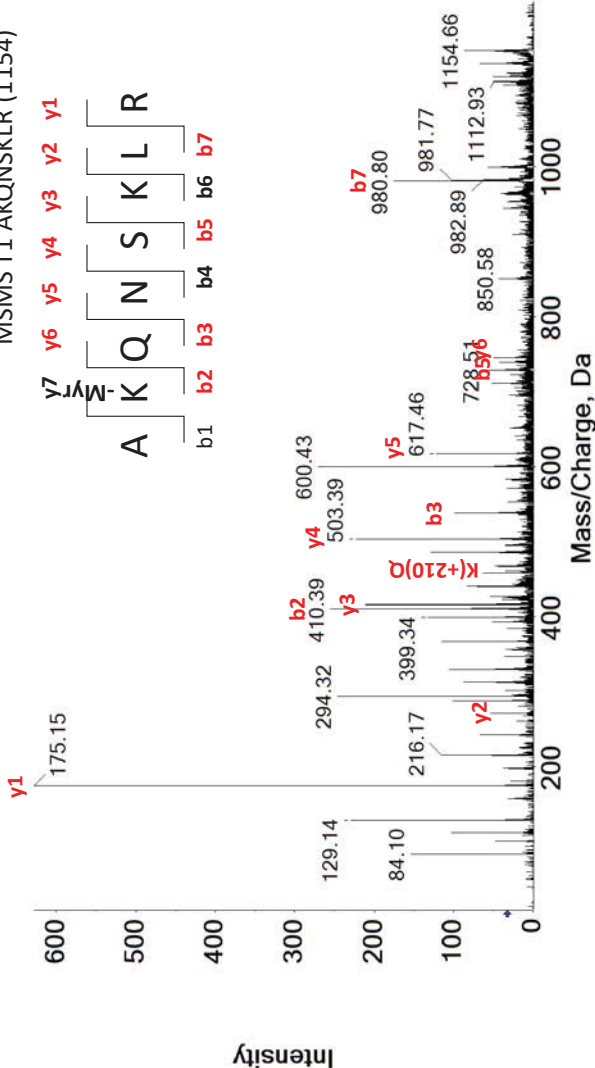
AKQNSKLR



MSMS TO AKQNSKLR (944)

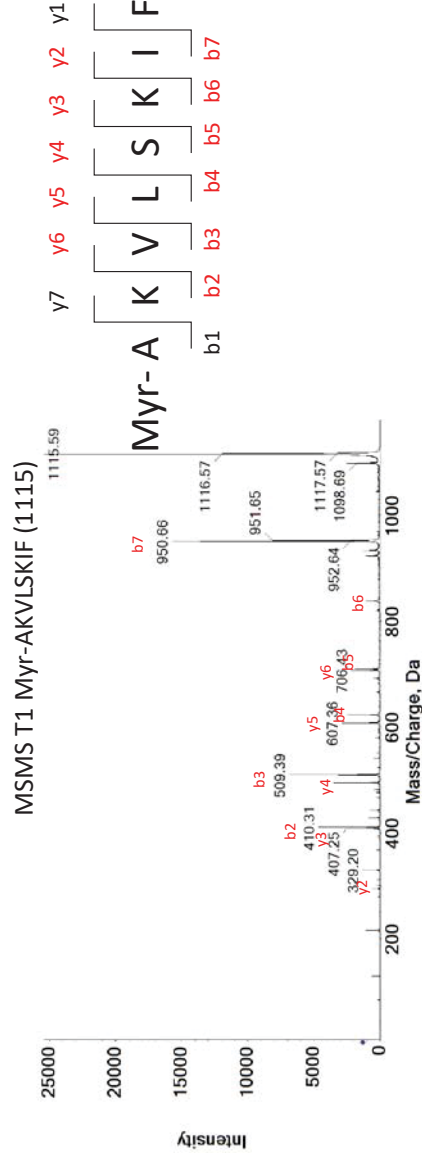
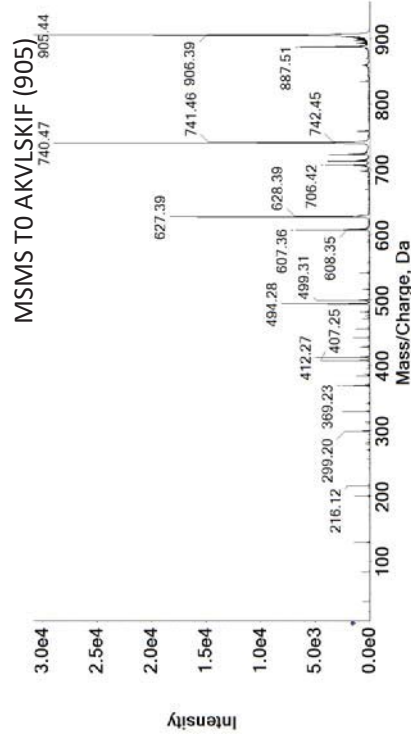
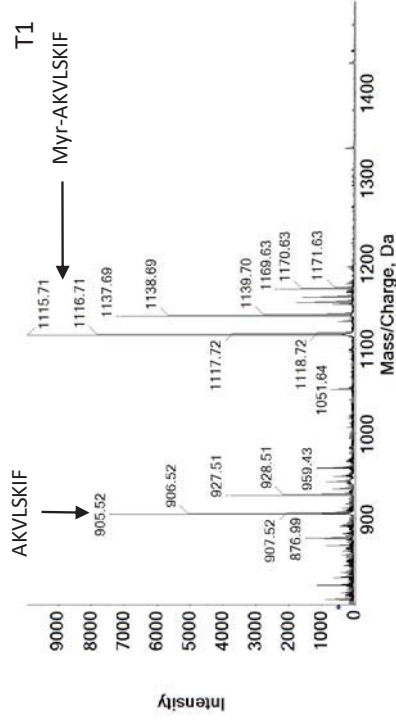
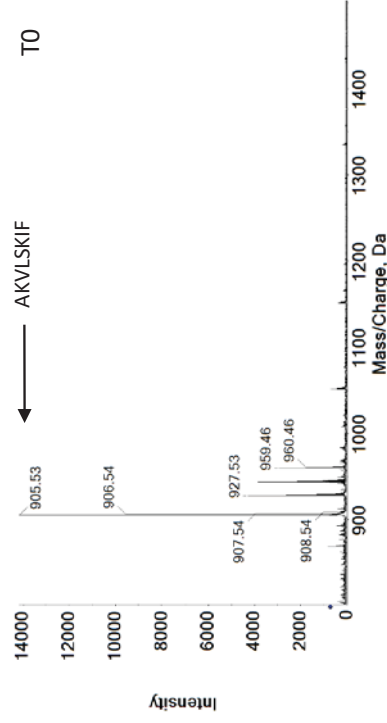


MSMST1 AKQNSKLR (1154)

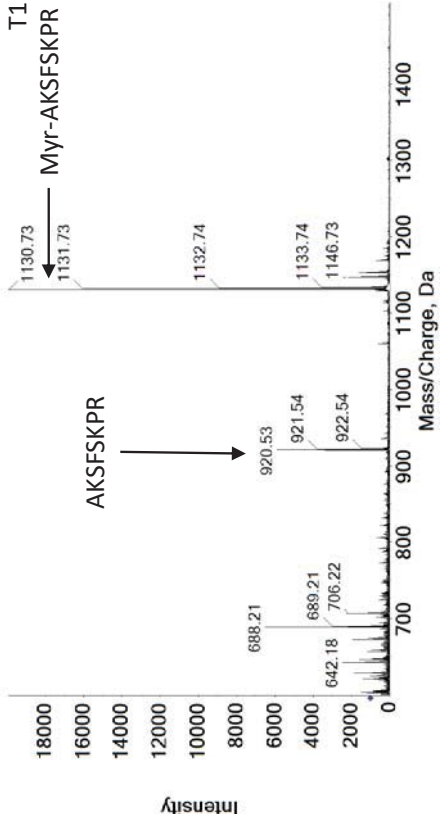
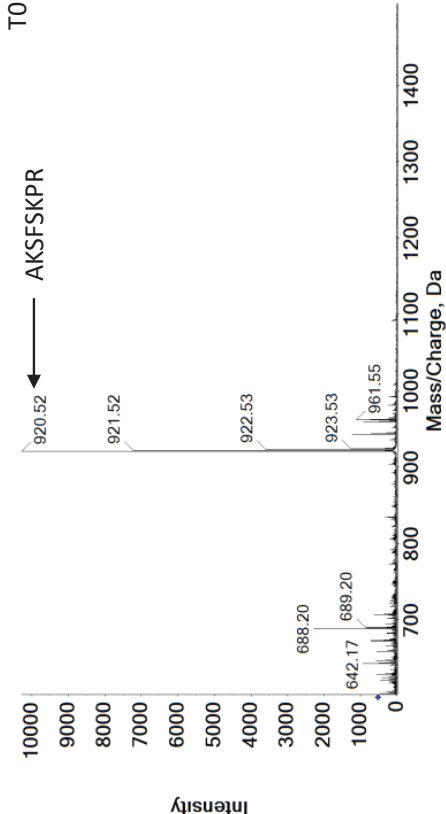


Mass/Ch	Mass/Ch	Intensity
K24 1154	arge (Da)	535 175,0203 35,45098 y1
1231	288,1639	31,52941 y2
2914	410,3834	254,7451 b2
3049	416,3236	210,8235 y3
3905	467,378	43,60785 ?
4432	503,3777	223,3726 y4
4875	538,4333	98,98039 b3
5859	600,4333	249,8824 y5-NH3
6146	617,4308	121,4118 y5
8017	739,5452	42,5098 b5
8118	745,4847	49,09804 y6
9226	850,5908	42,5098 b6-NH3
10884	980,7767	175,5294 b7
12877	1137,707	67,45098 v8

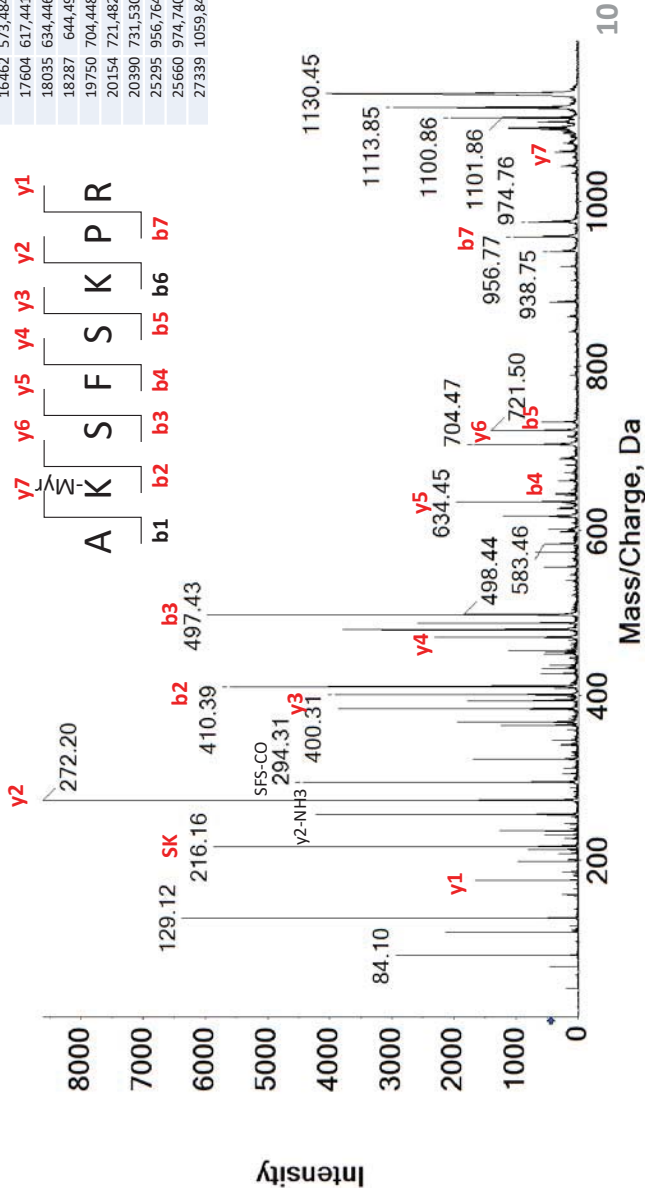
AKVLSKIF



AKSFSKPR

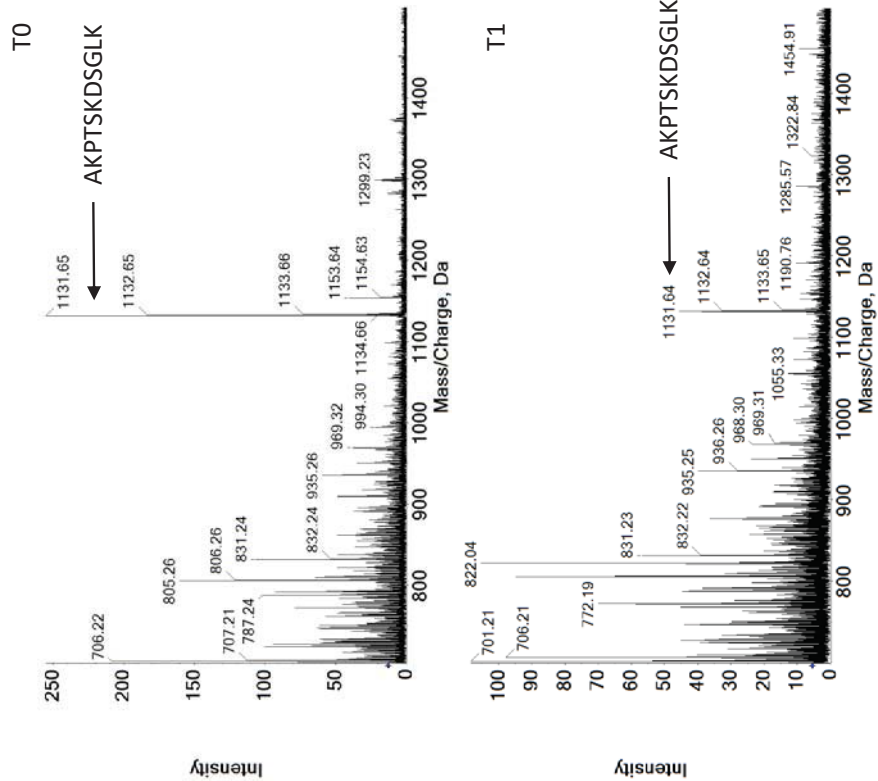


MSMS T1 Myr-AKSFSKPR (1130)

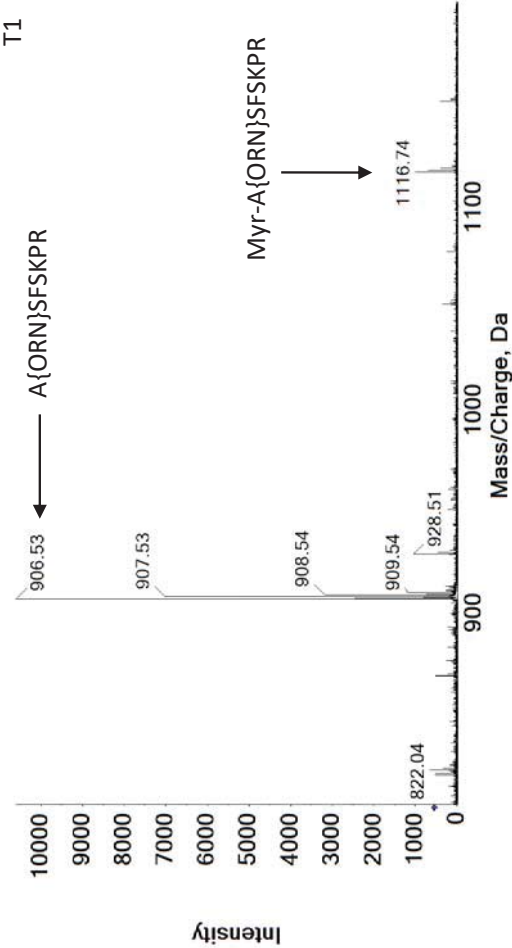
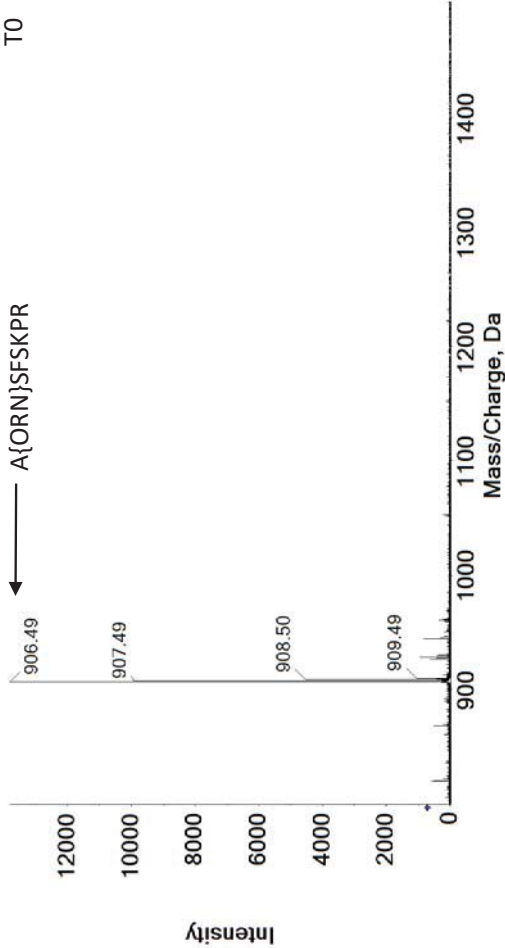


766	84,09495	2935,84 K
1481	112,1094	2126,90 R
2056	129,1225	6378,35 K
3482	175,1427	1653,96 Y1
4180	198,1471	973,80 KP-CO
4900	216,147	5869,18 SK
5668	235,1309	1267,45 F5
6392	255,174	4213,96 Y2-NH3
7039	272,2352	6608,47 Y2
7848	294,3016	4431,22 SFS-CO
8846	322,1745	1693,18 SFS
9640	345,2368	414,43 FSK-H2O
10247	363,2515	1236,71 FSK
10905	383,2669	3857,57 Y3-NH3
11231	393,3545	1791,69 b2-NH3
11453	400,3001	3915,61 Y3
11709	408,3862	384,78 H2O
11772	410,3888	5604,39 b2
12270	426,3936	597,18 K(+210)S
12996	450,2837	557,65 SFSK
13590	470,3228	2309,80 Y4-NH3
13855	479,4061	3797,80 b3-H2O
13884	480,4055	1171,45 b3-NH3
14085	487,3613	2589,18 Y4
14373	497,4166	5962,98 b3
16462	573,4848	685,96 K(+210)SF
17604	617,4413	1208,78 Y5-NH3
18035	634,4463	1964,86 Y5
18287	644,498	369,88 b4
19750	704,4481	1667,14 Y6-NH3
20154	721,4829	1395,76 Y6
20390	731,5301	590,59 b5
25295	956,7646	1045,80 b7
25660	974,7401	803,45 b7+H2O
27339	1059,846	375,53 Y7

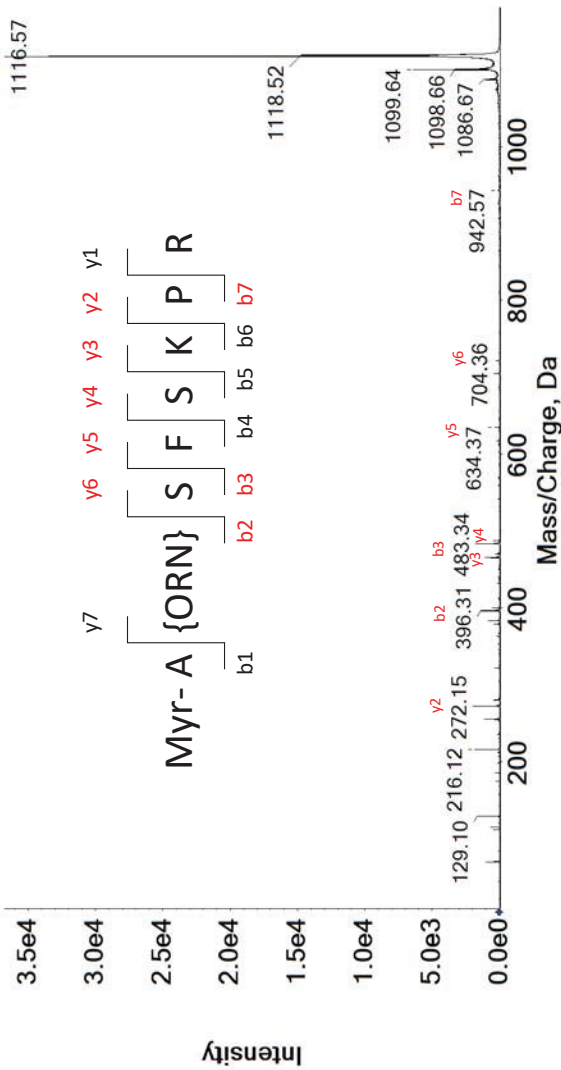
AKPTSKDSGLK



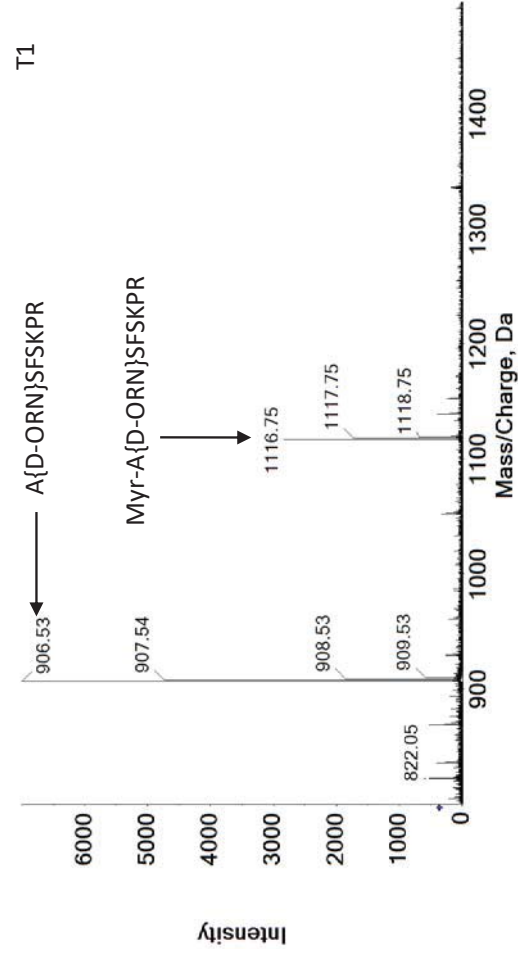
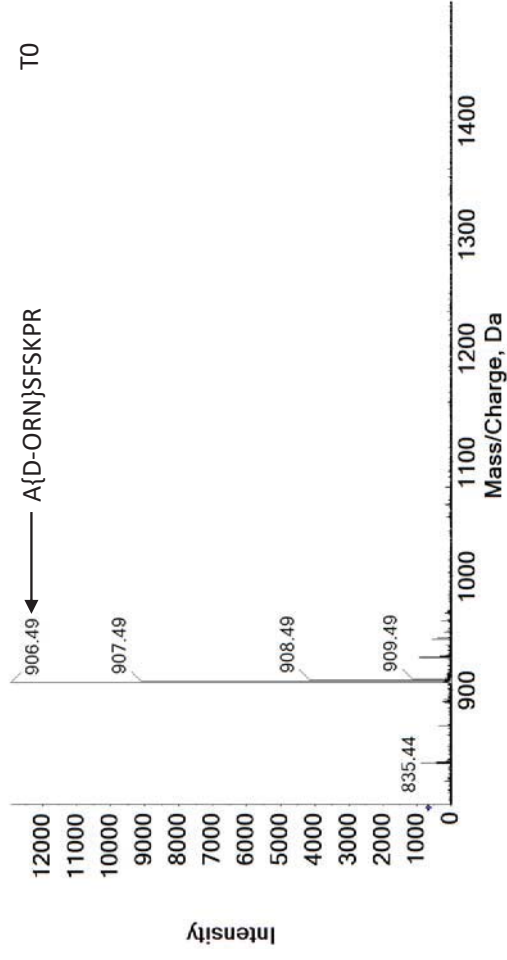
A{ORN}SFSKPR



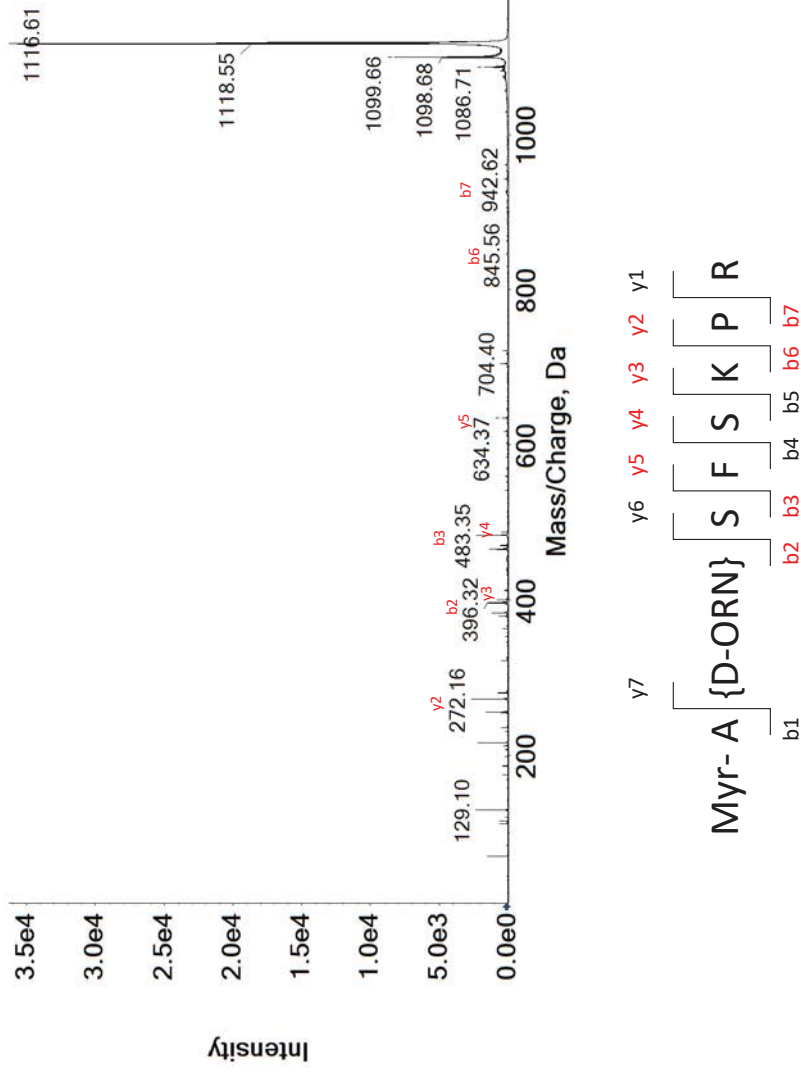
MSMS T1 Myr-A{ORN}SFSKPR (1116)



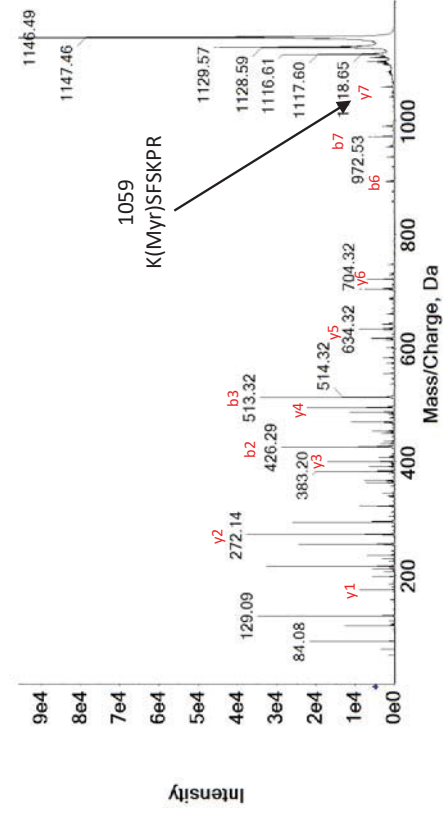
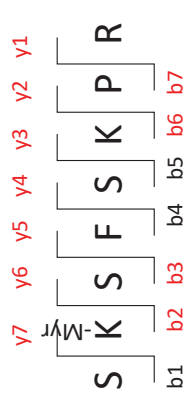
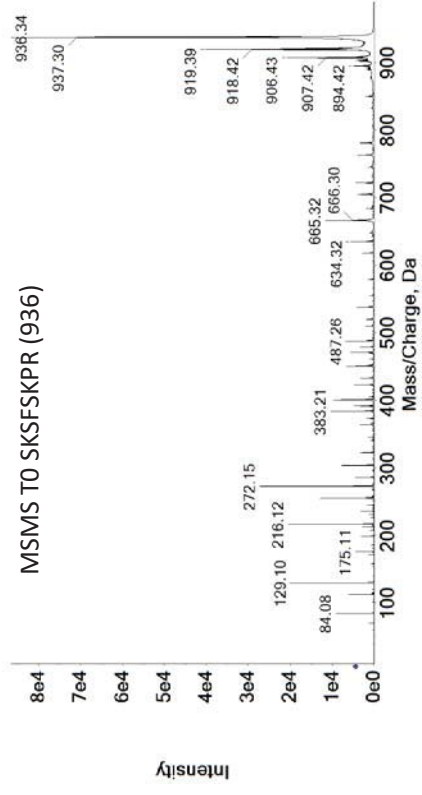
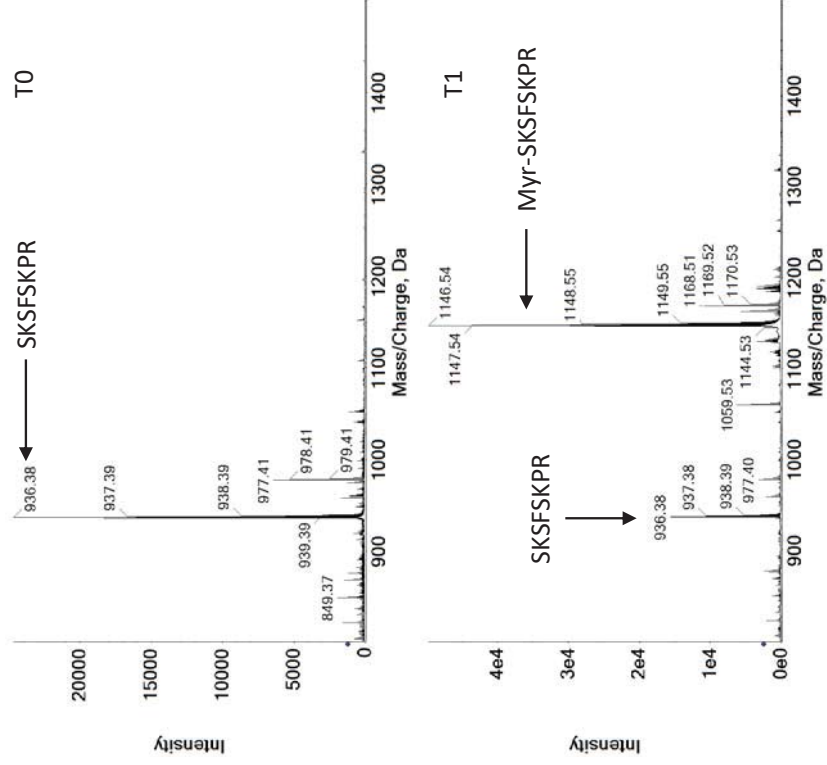
A{D-ORN}SFSKPR



MSMS T1 Myr-A{D-ORN}SFSKPR (1116)

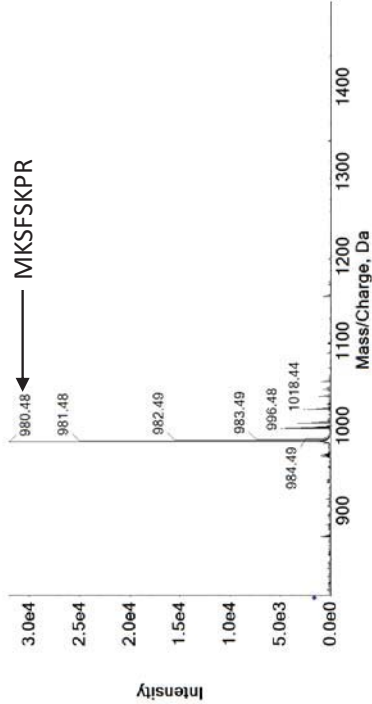


SKFSKPR

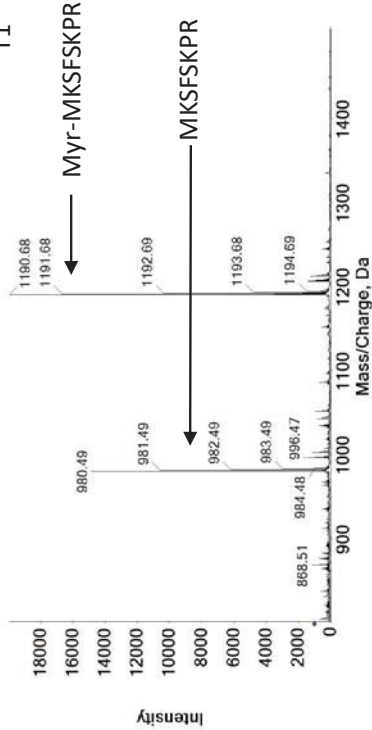


MKSFSKPR

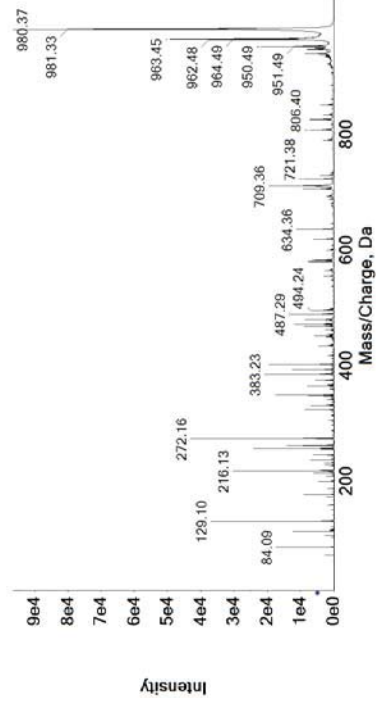
T0



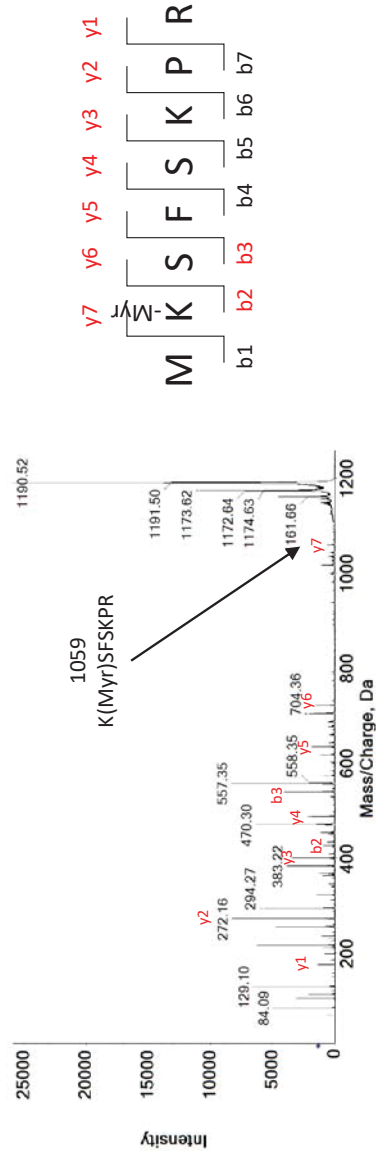
T1



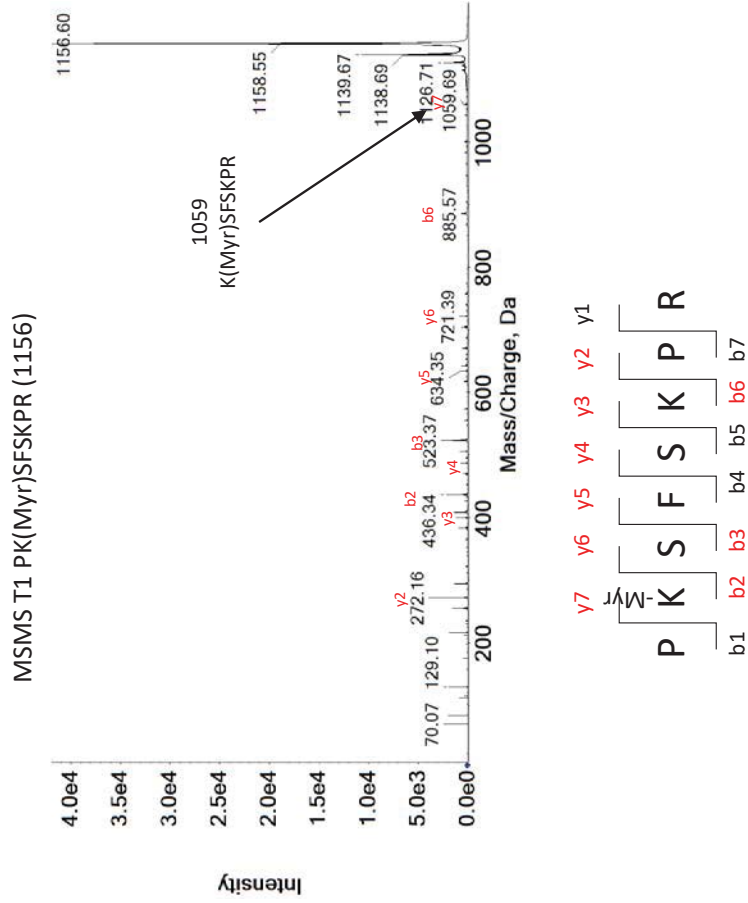
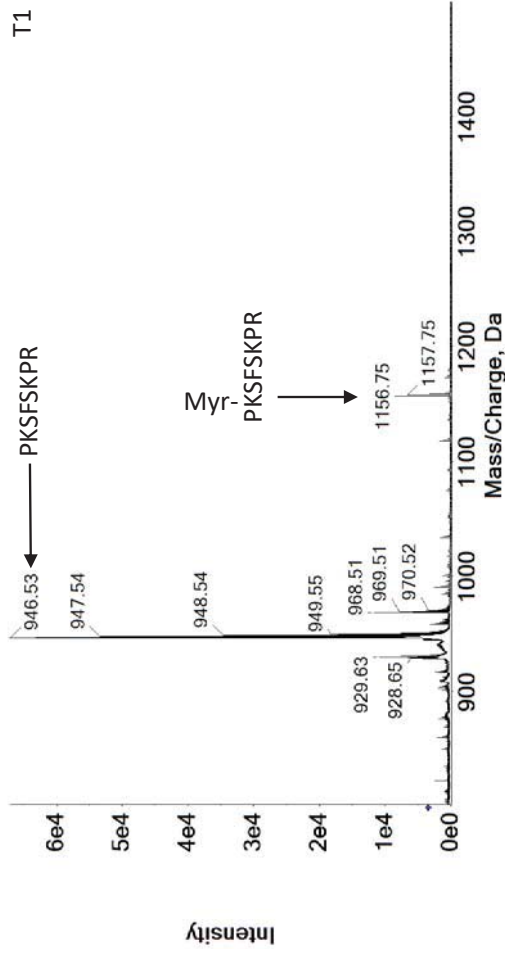
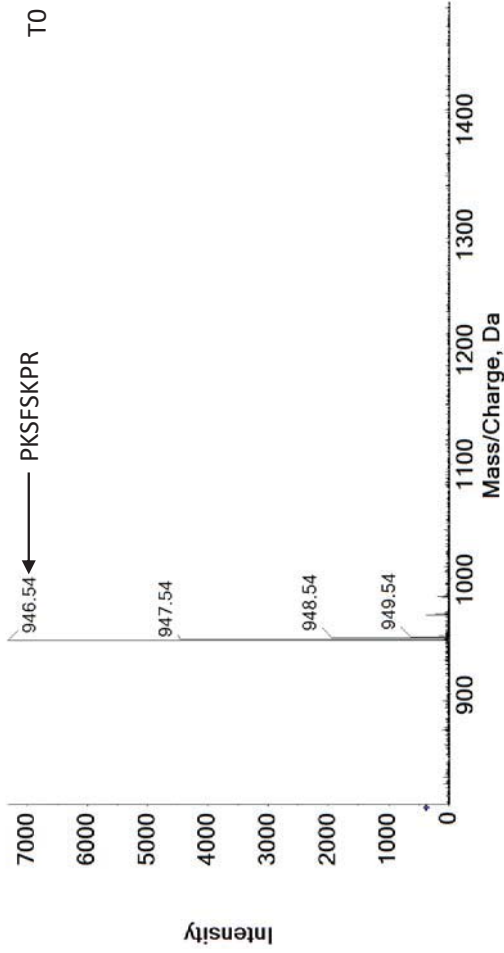
MSMS T0 MKSFSKPR (980)



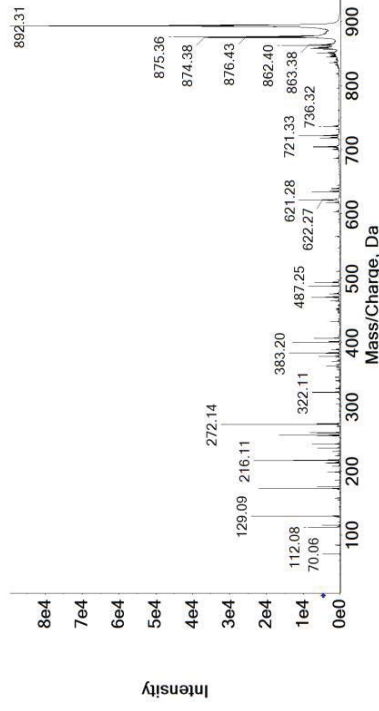
MSMS T1 MK(Myr)SFSKPR (1158)



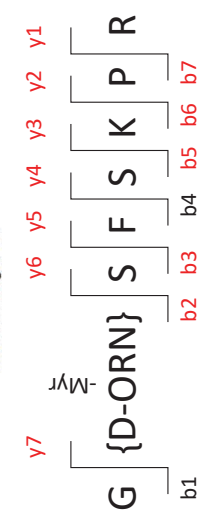
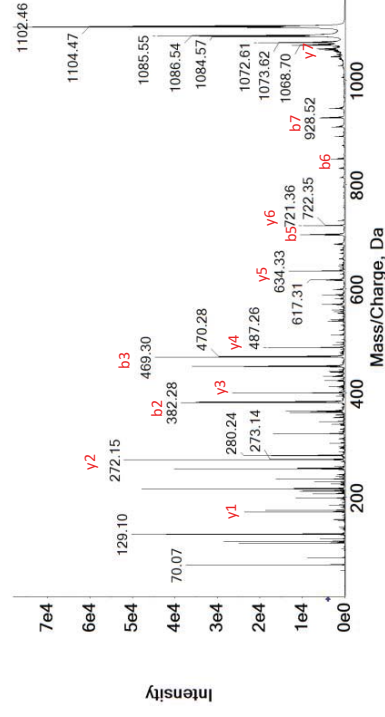
PKSFSKPR



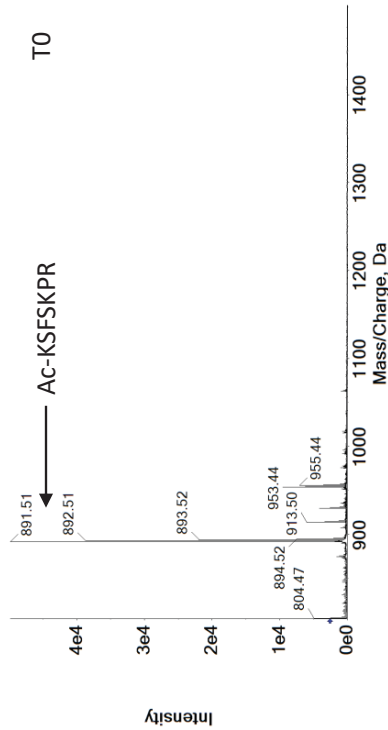
MSMS TO G{D-ORN}SFSKPR (892)



MSMS T3 G{D-ORN}MyrSFSKPR (1102)

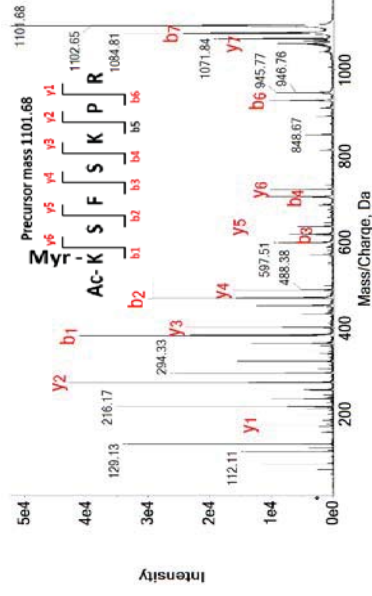


Ac-KSFSKPR



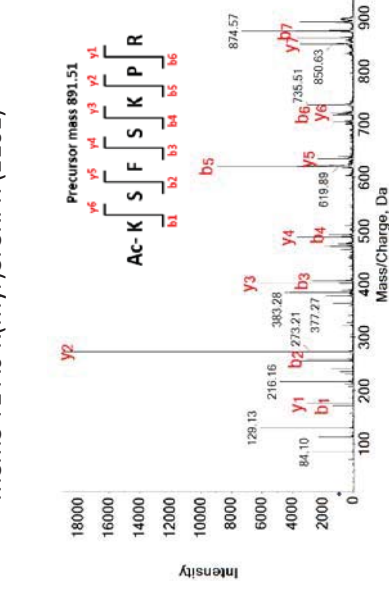
T0

MSMS T0 Ac-KSFSKPR (891)



T1

MSMS T1 Ac-K(Myr)SFSKPR (1101)

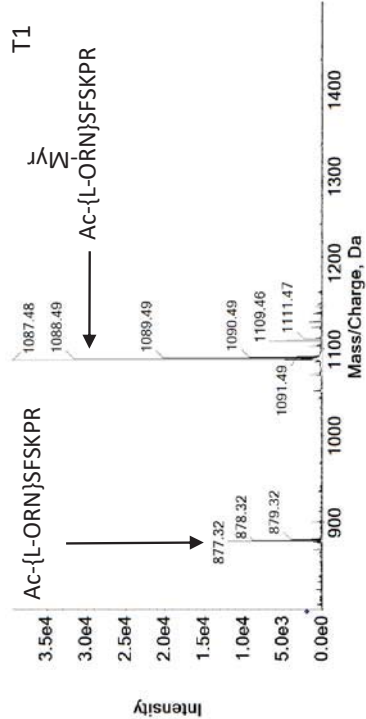
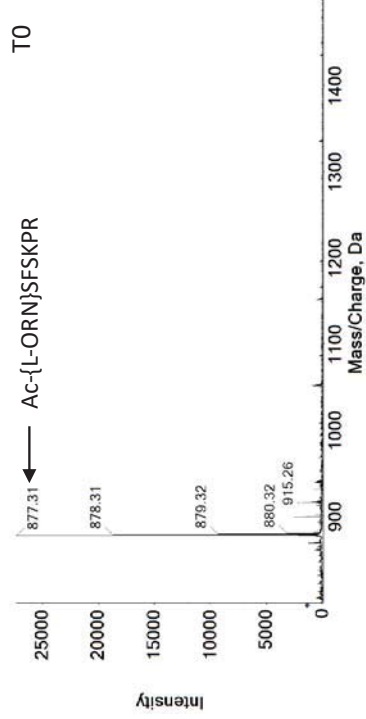


Myr

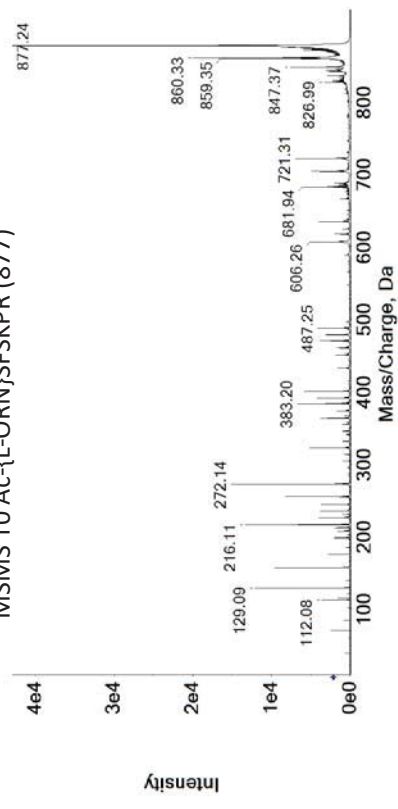
Ac-KSFSKPR

Ac-KSFSKPR

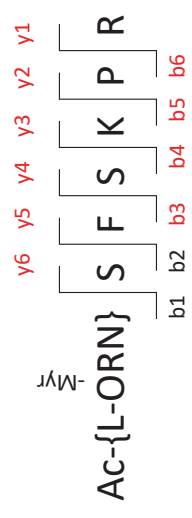
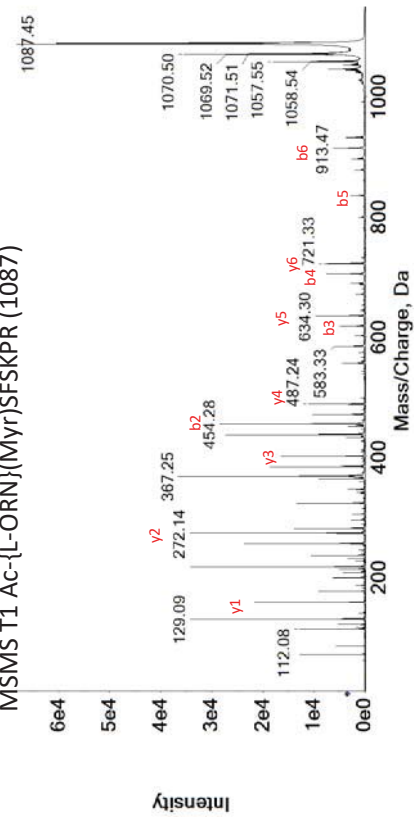
Ac-{L-ORN}SFSKPR



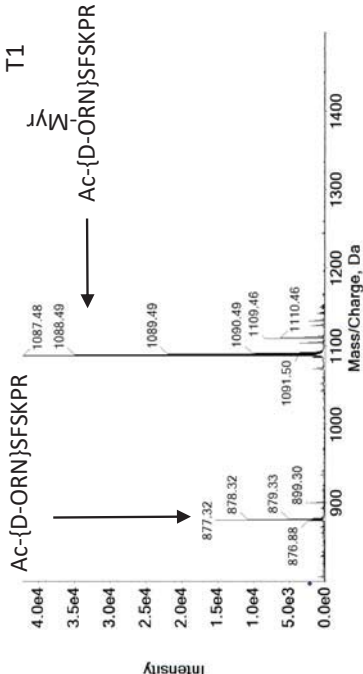
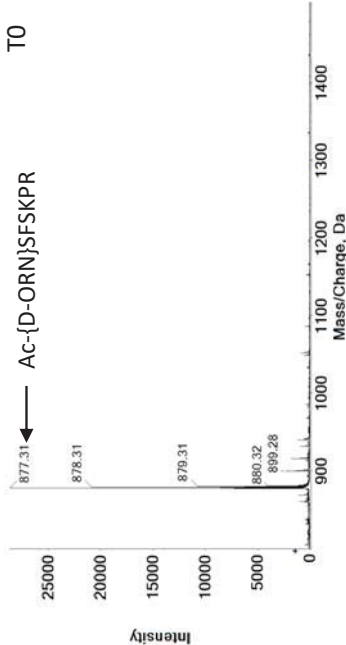
MSMS T0 Ac-{L-ORN}SFSKPR (877)



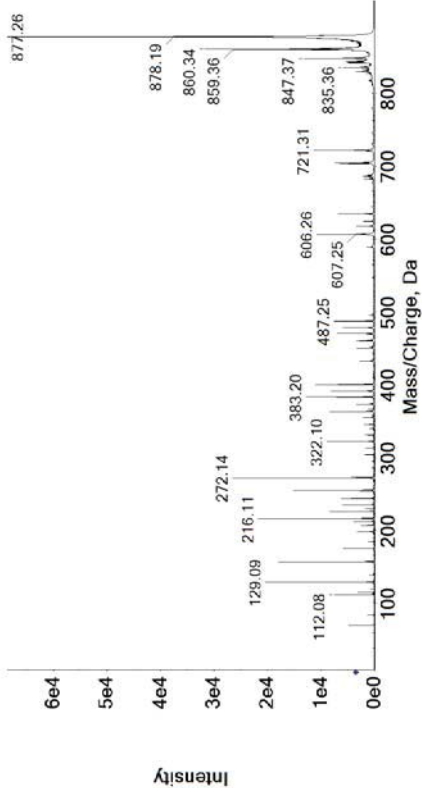
MSMS T1 Ac-{L-ORN}(M)yr}SFSKPR (1087)



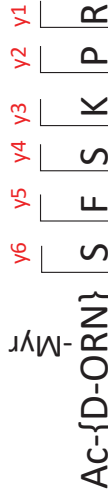
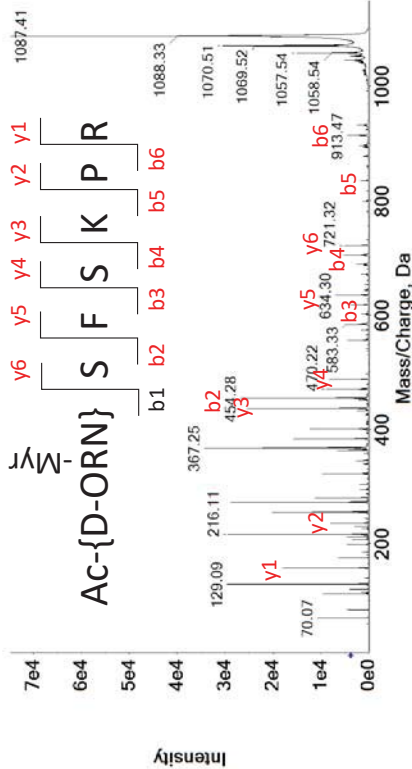
Ac-{D-ORN}SFSKPR



MSMS TO Ac-{D-ORN}SFSKPR (877)

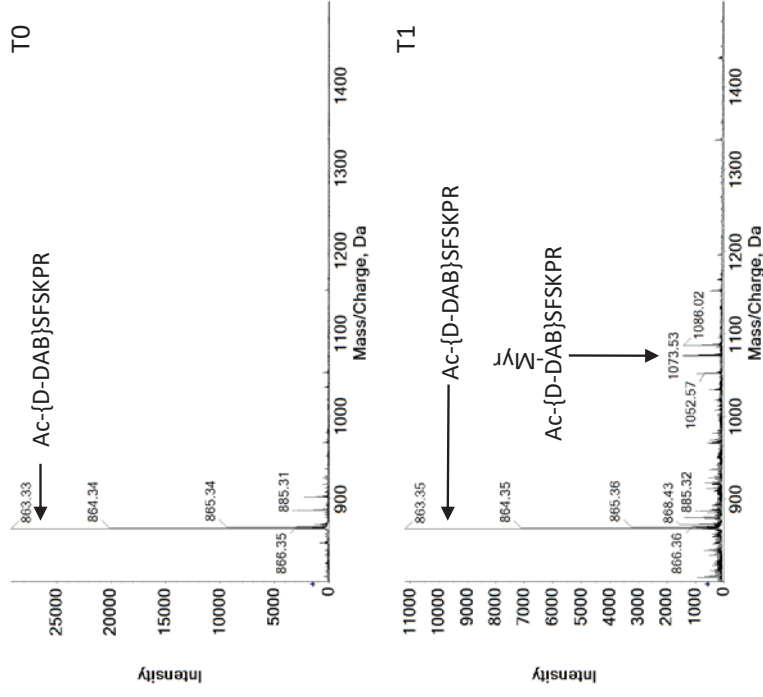


MSMS T1 Ac-{D-ORN}(Myr)SFSKPR (1087)

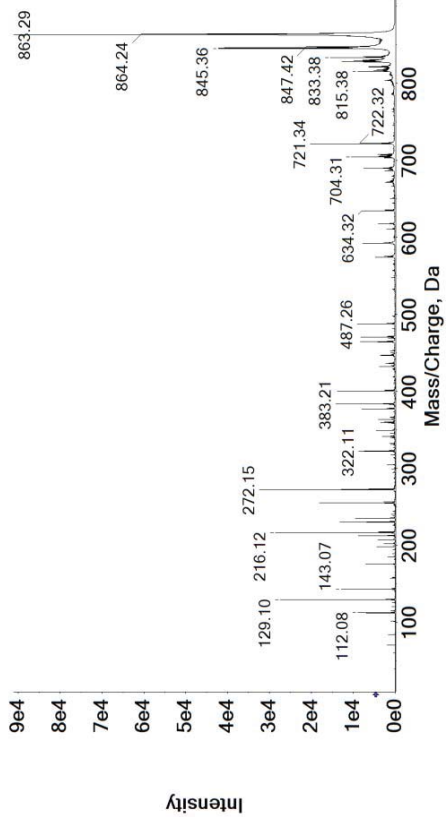


Mass/Ch	arge (Da)	Intensity	ion
N16	1087	6416,418	y1
9780	175,1064	8455,216	y2
13956	272,2247	12488,37	y3
18426	400,2117	26812,13	b2
20091	454,2782	8257,673	y4
21057	487,2445	1466,144	b3
24166	601,3394	2666,876	b4
25006	634,266	5124,183	y6
26339	688,3574	361,3072	b5
27125	721,3122	4516,706	b6
29297	816,4788		
31381	913,469		

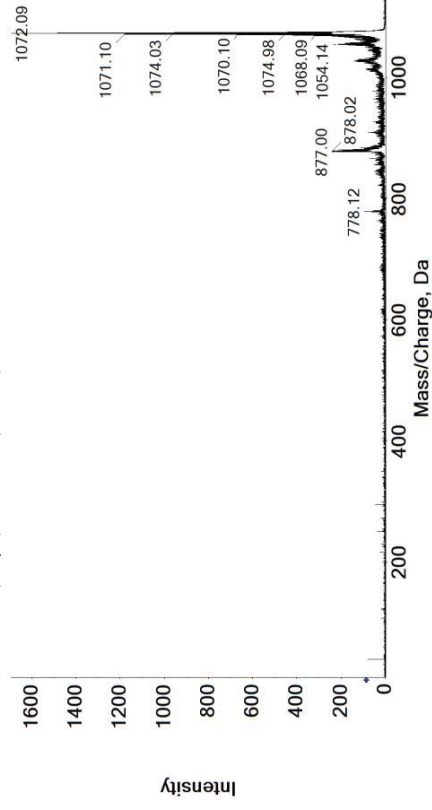
Ac-{D-DAB}SFSKPR



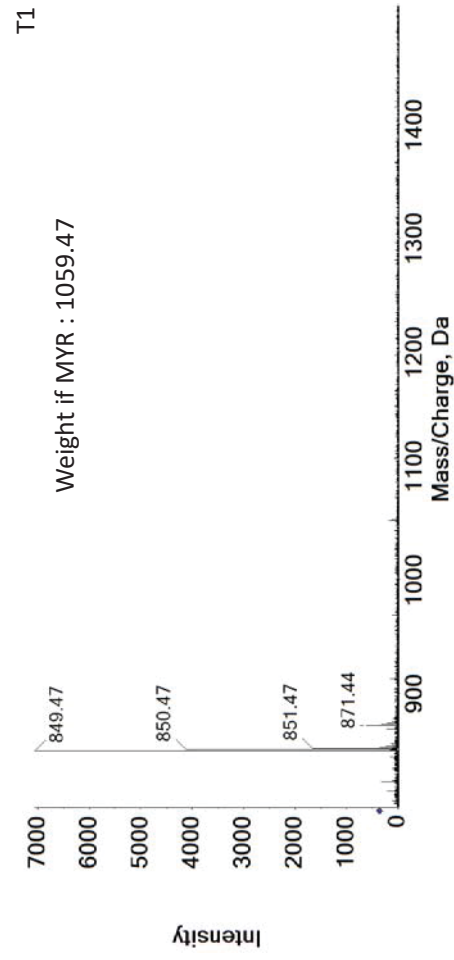
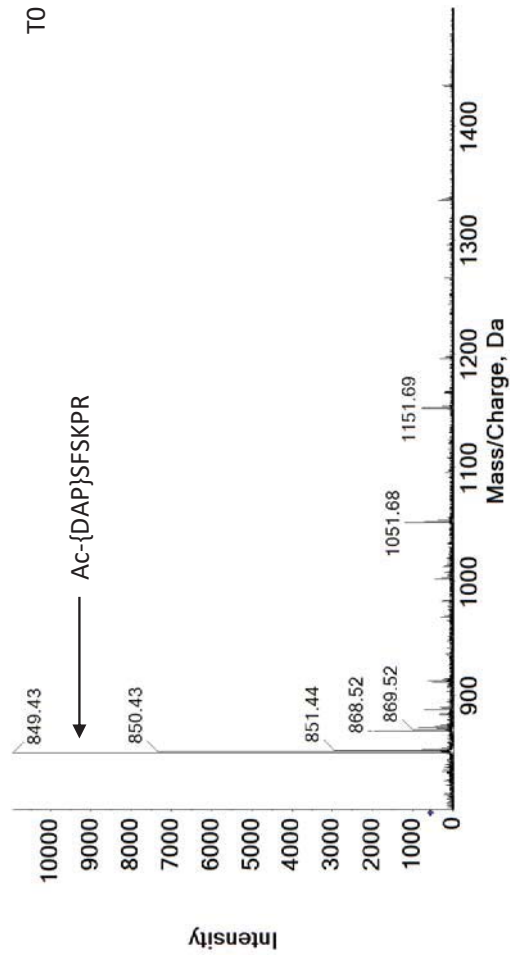
MSMS T0 Ac-{D-DAB}SFSKPR (863)



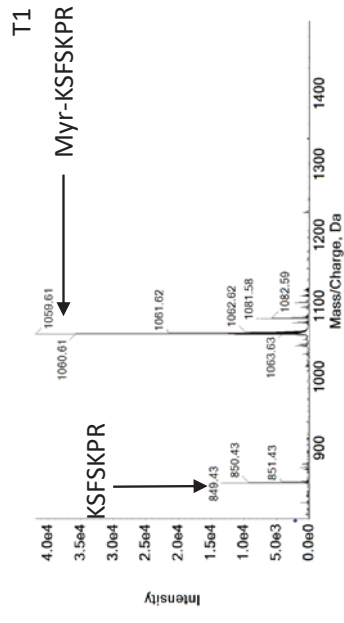
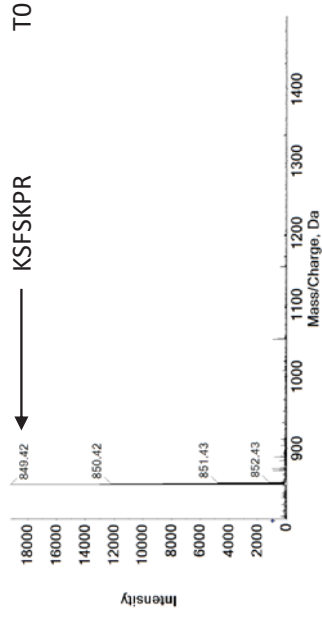
MSMS T1 Ac-{D-DAB}{Myr}SFSKPR (1073)



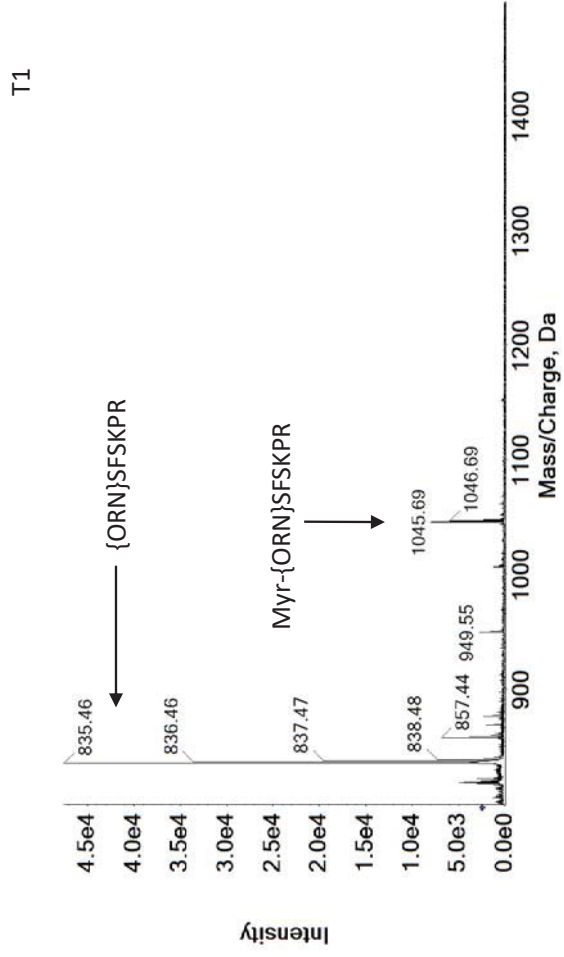
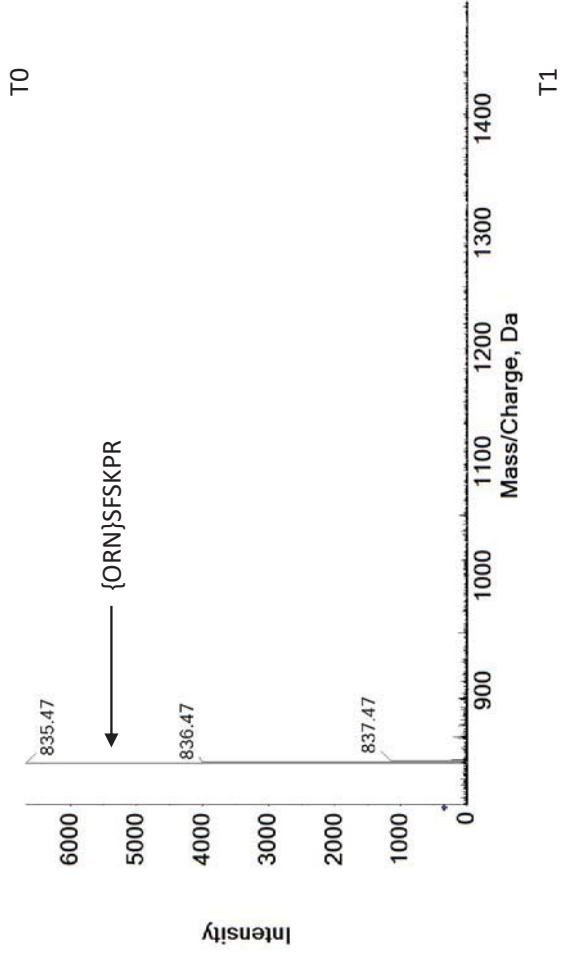
Ac-{DAP}SFSKPR



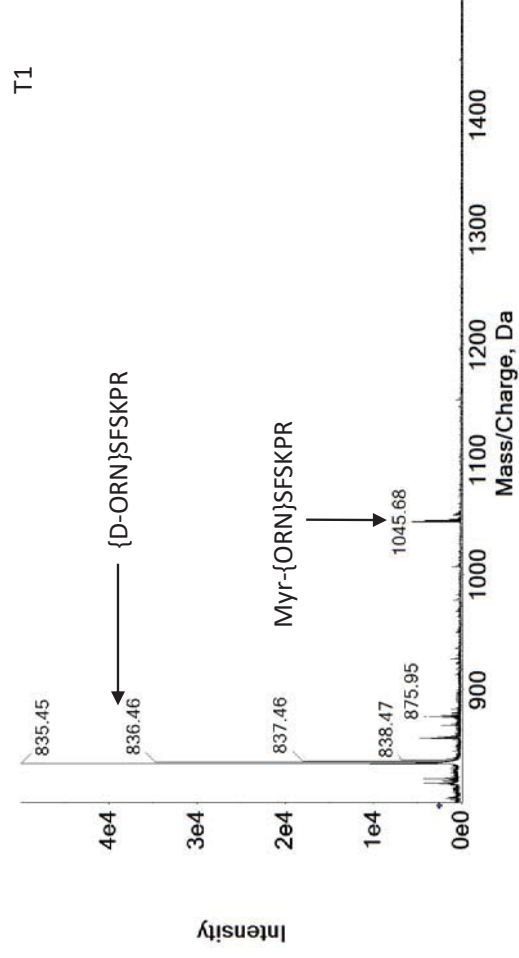
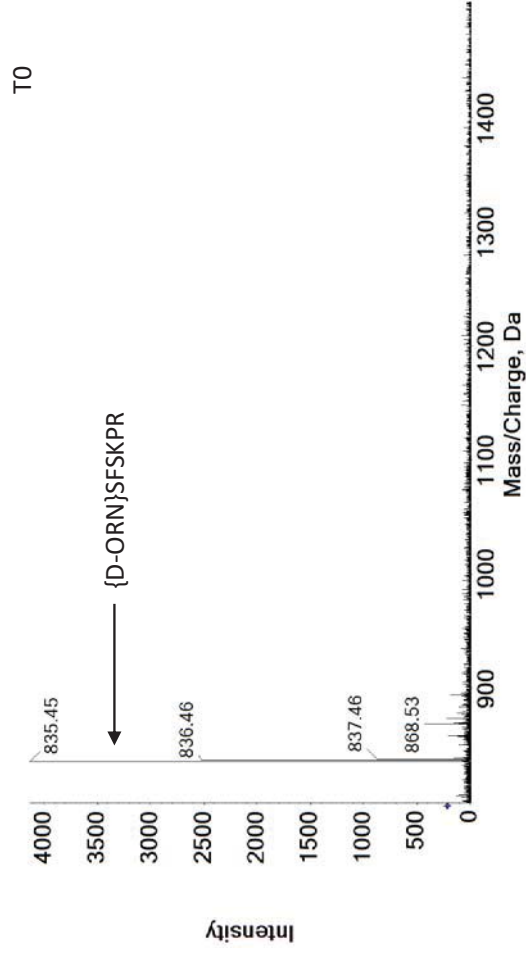
KSFSKPR



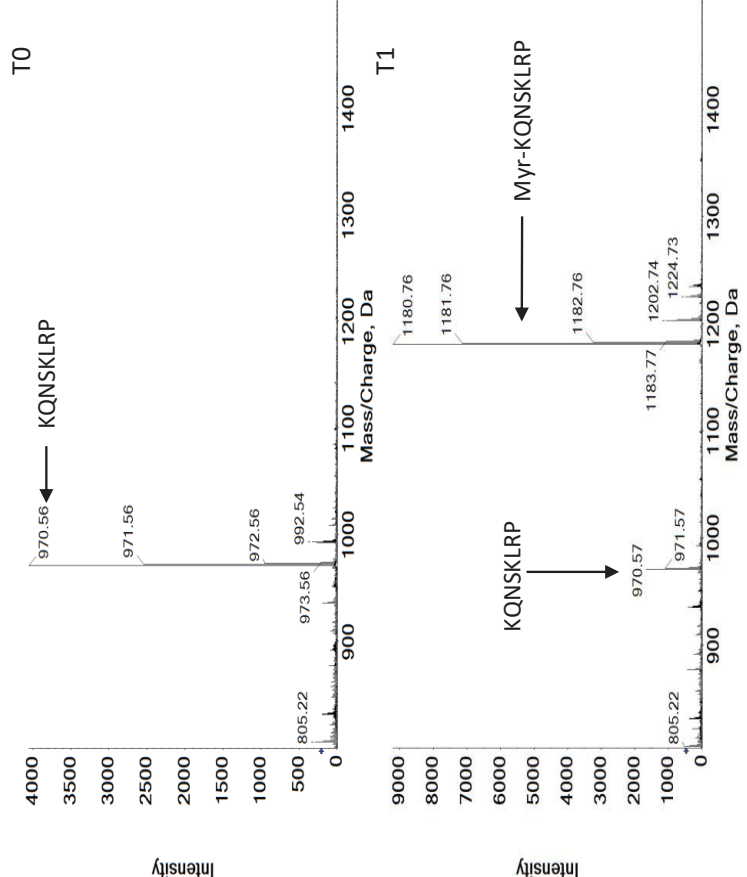
{ORN}SFSKPR



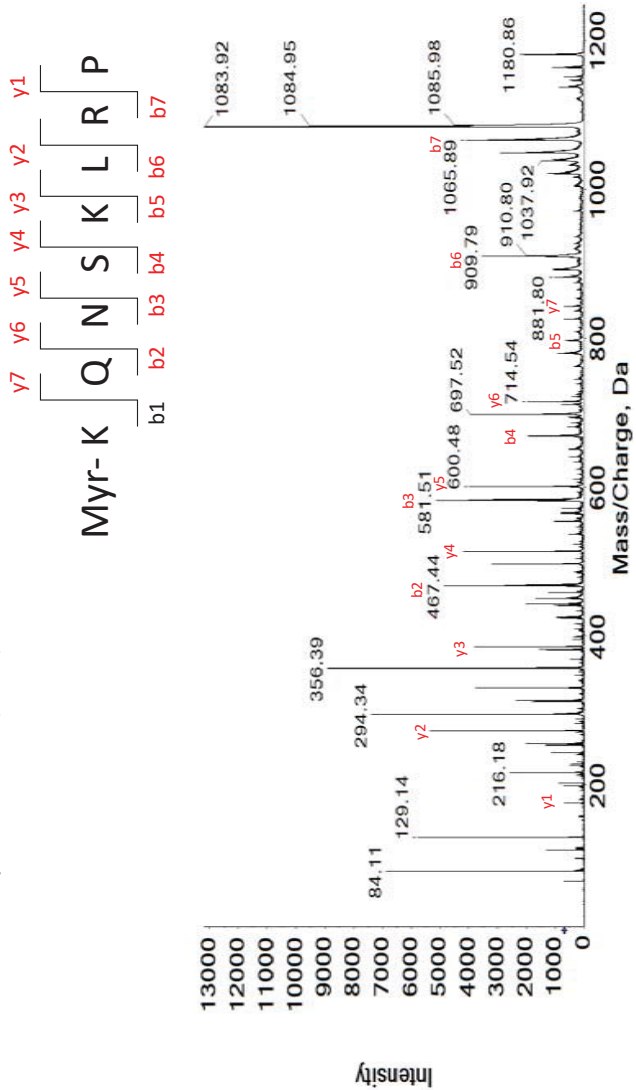
{D-ORN}SFSKPR



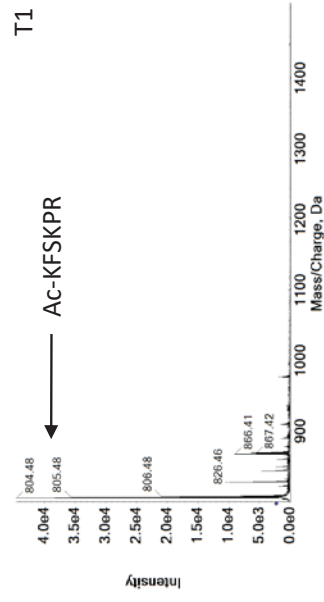
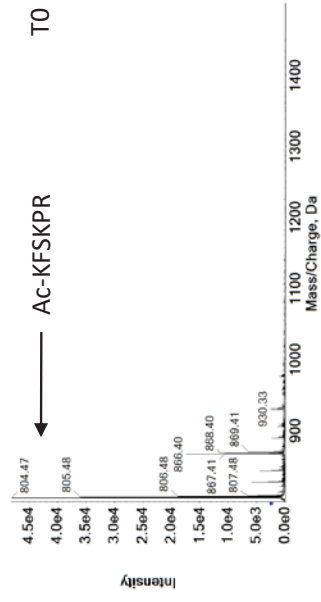
KQNSKLRP



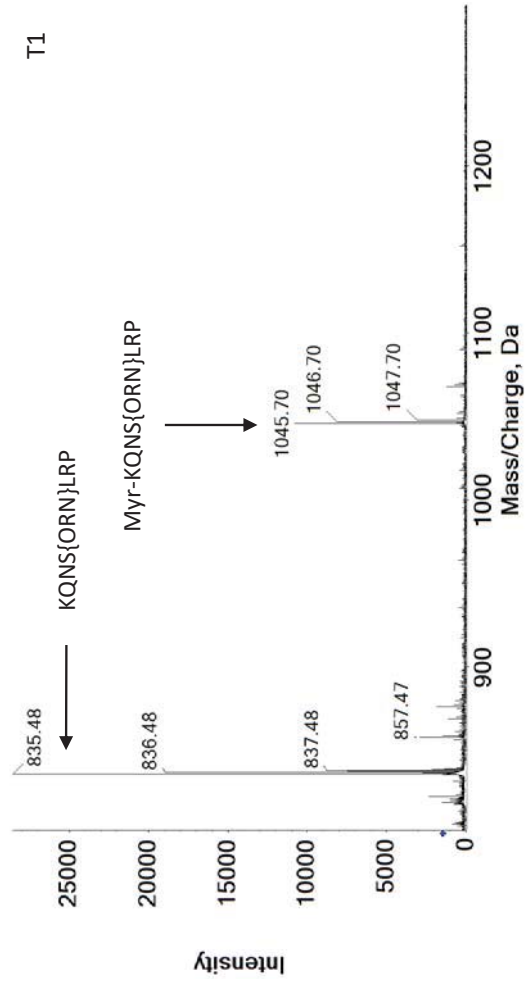
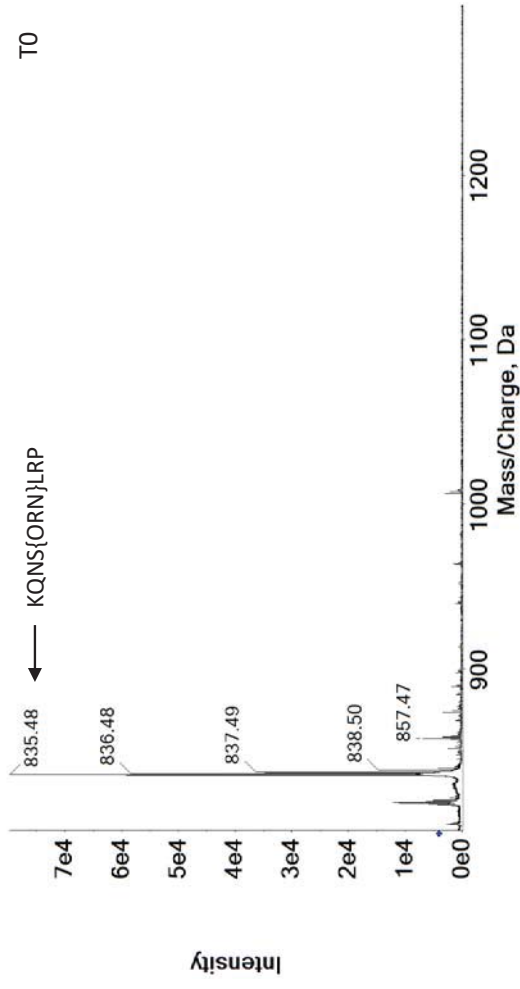
MSMS T1 Myr-KQNSKLRP (1180)



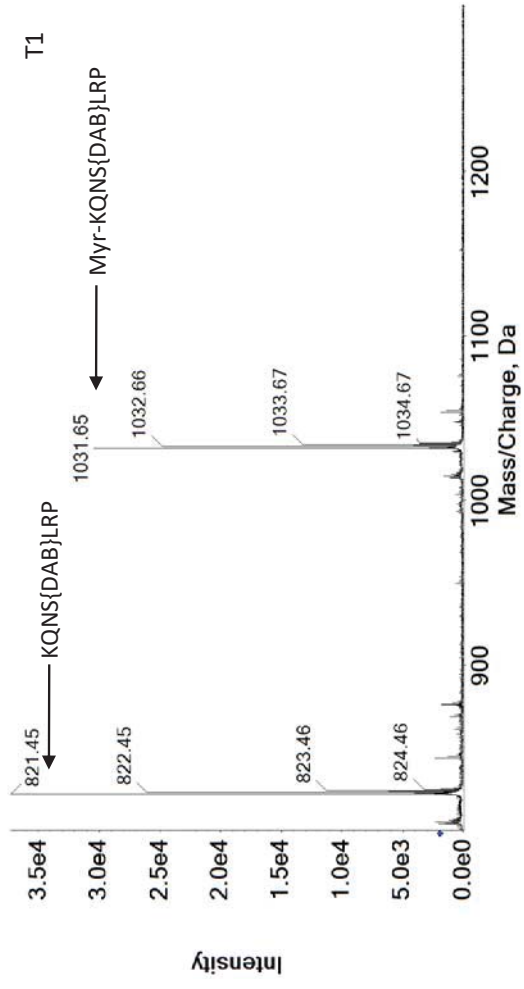
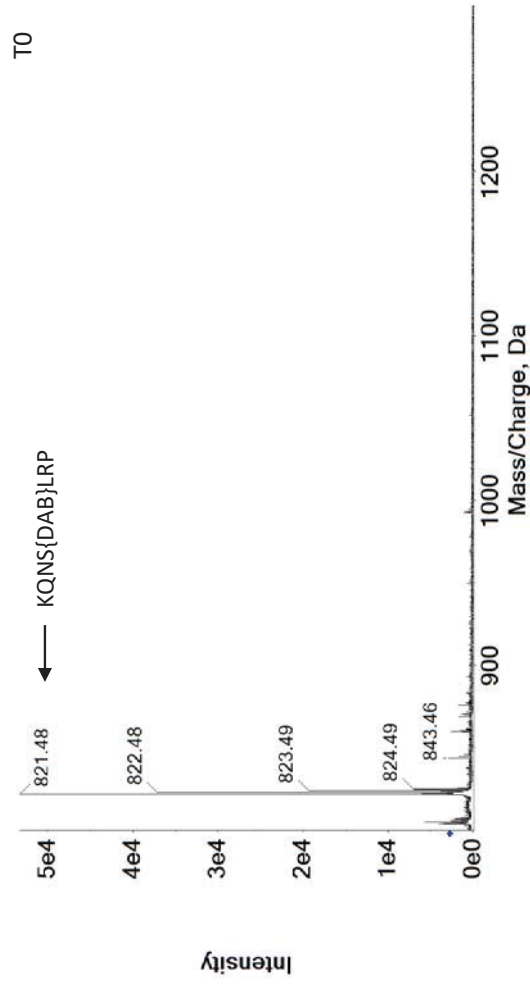
Ac-KFSKPR



KQNS{ORN}LRP

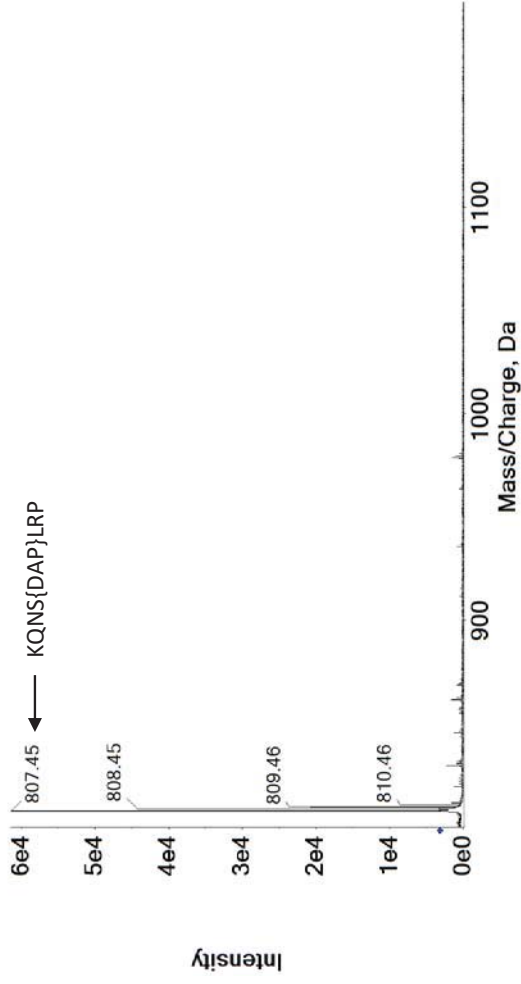


KQNS{DAB}LRP

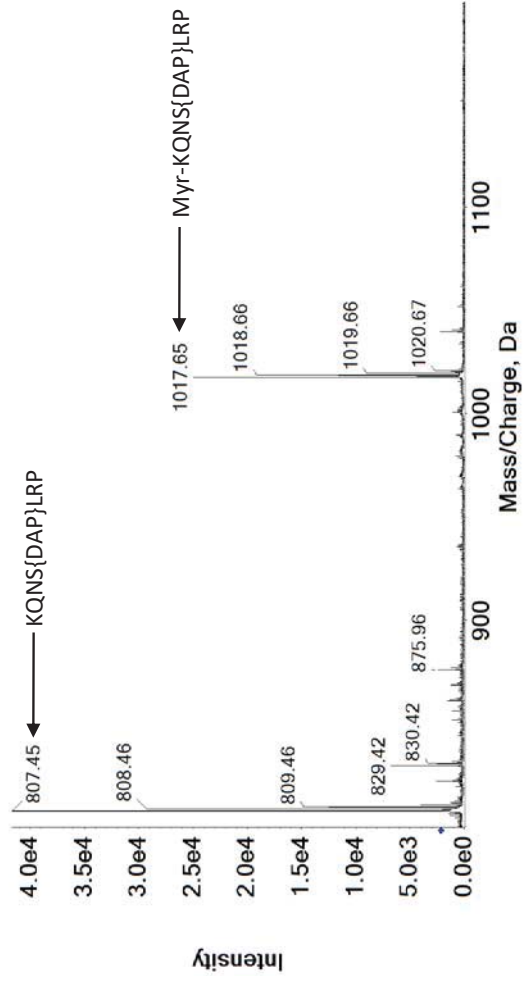


KQNS{DAP}LRP

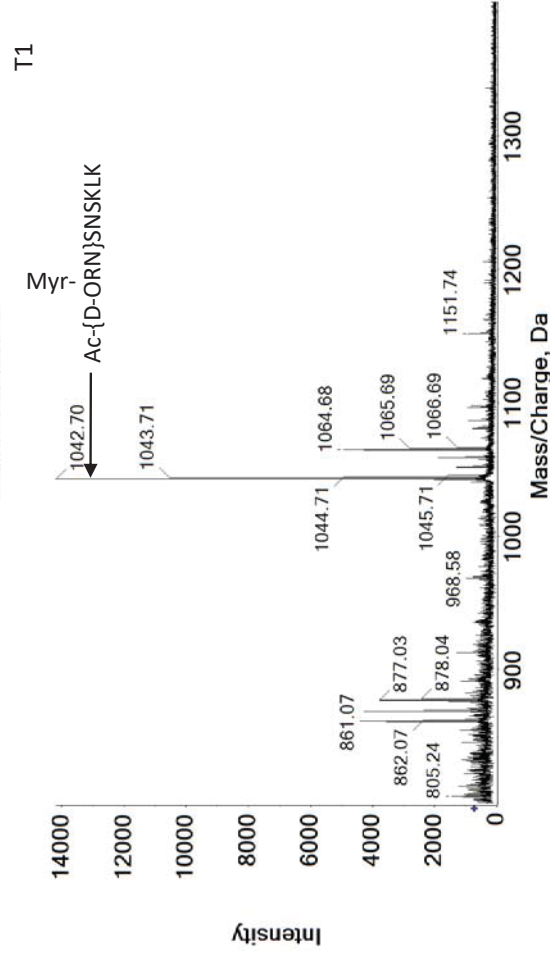
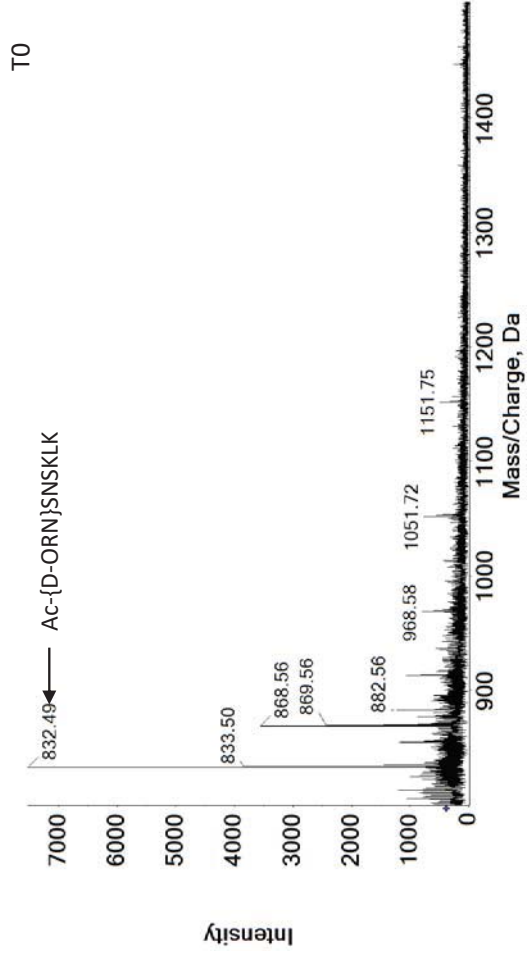
T0



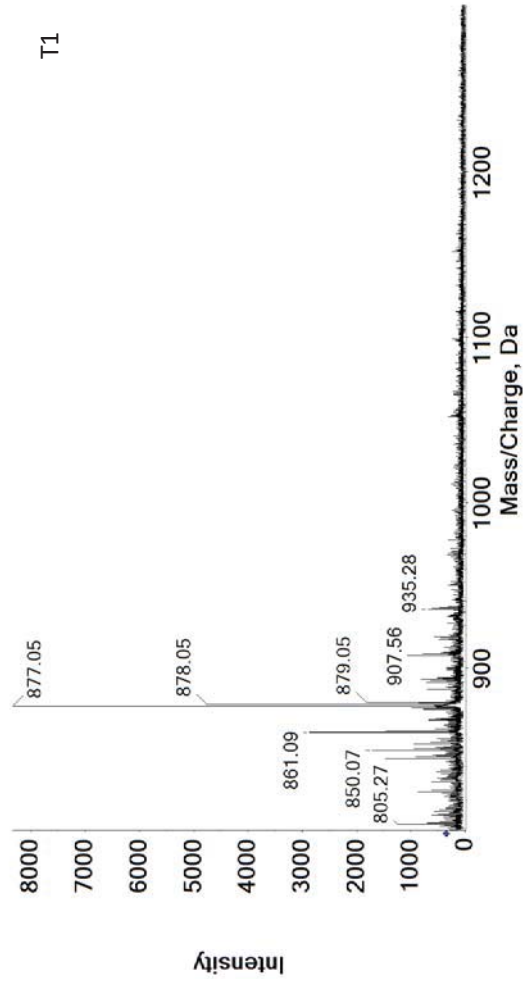
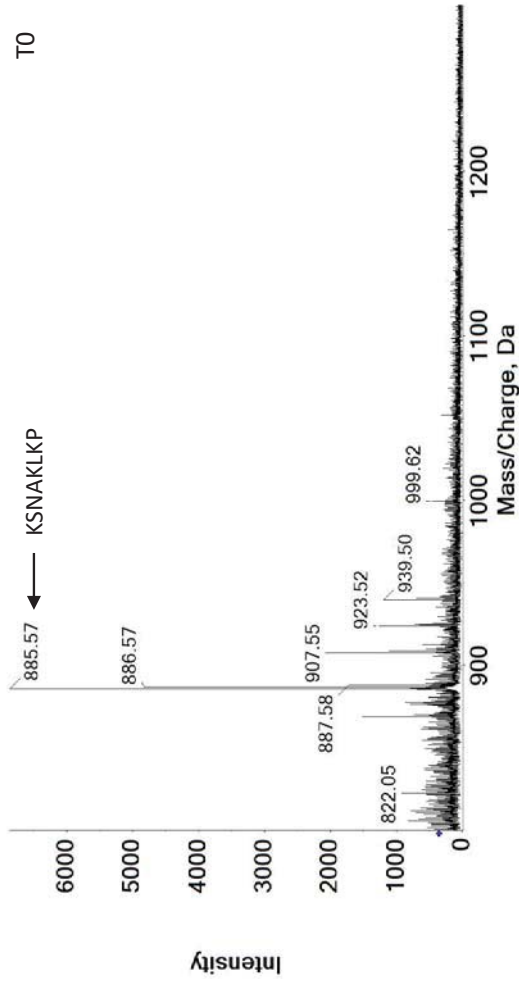
T1



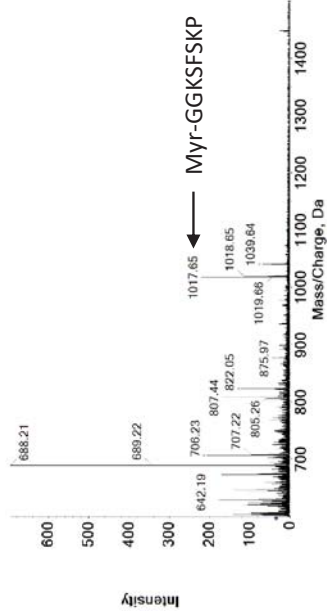
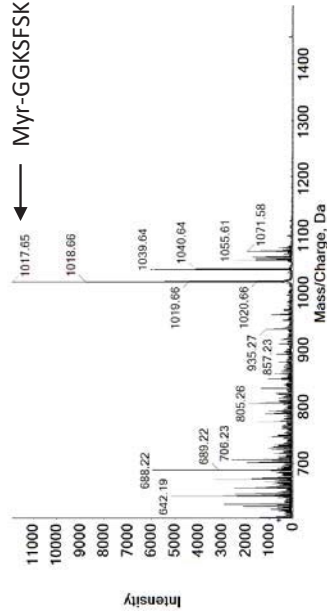
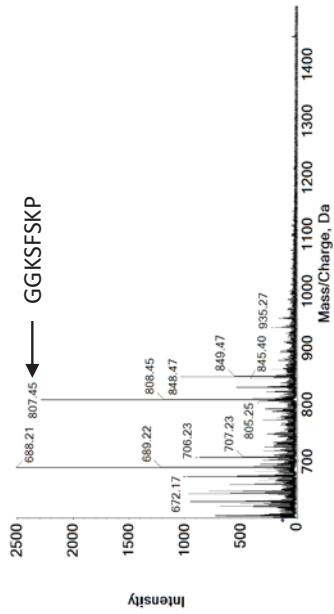
Ac-{D-ORN}SNSKCLK



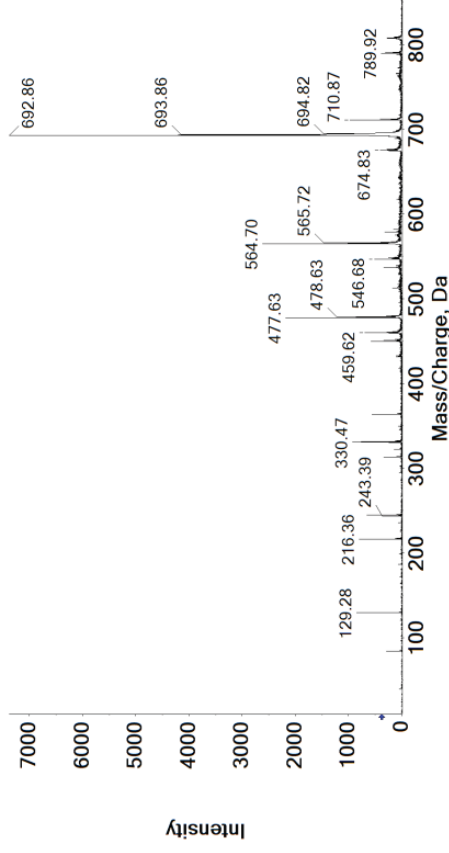
KSNAKLKP



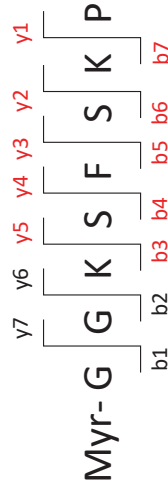
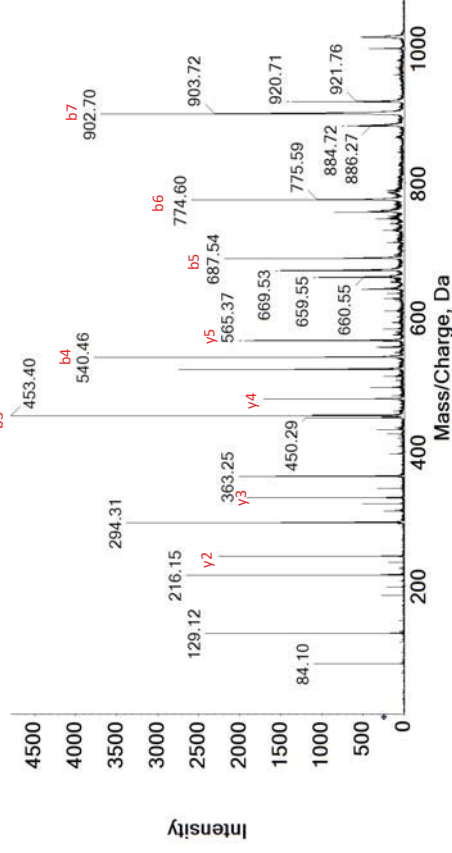
GGKSFSKP

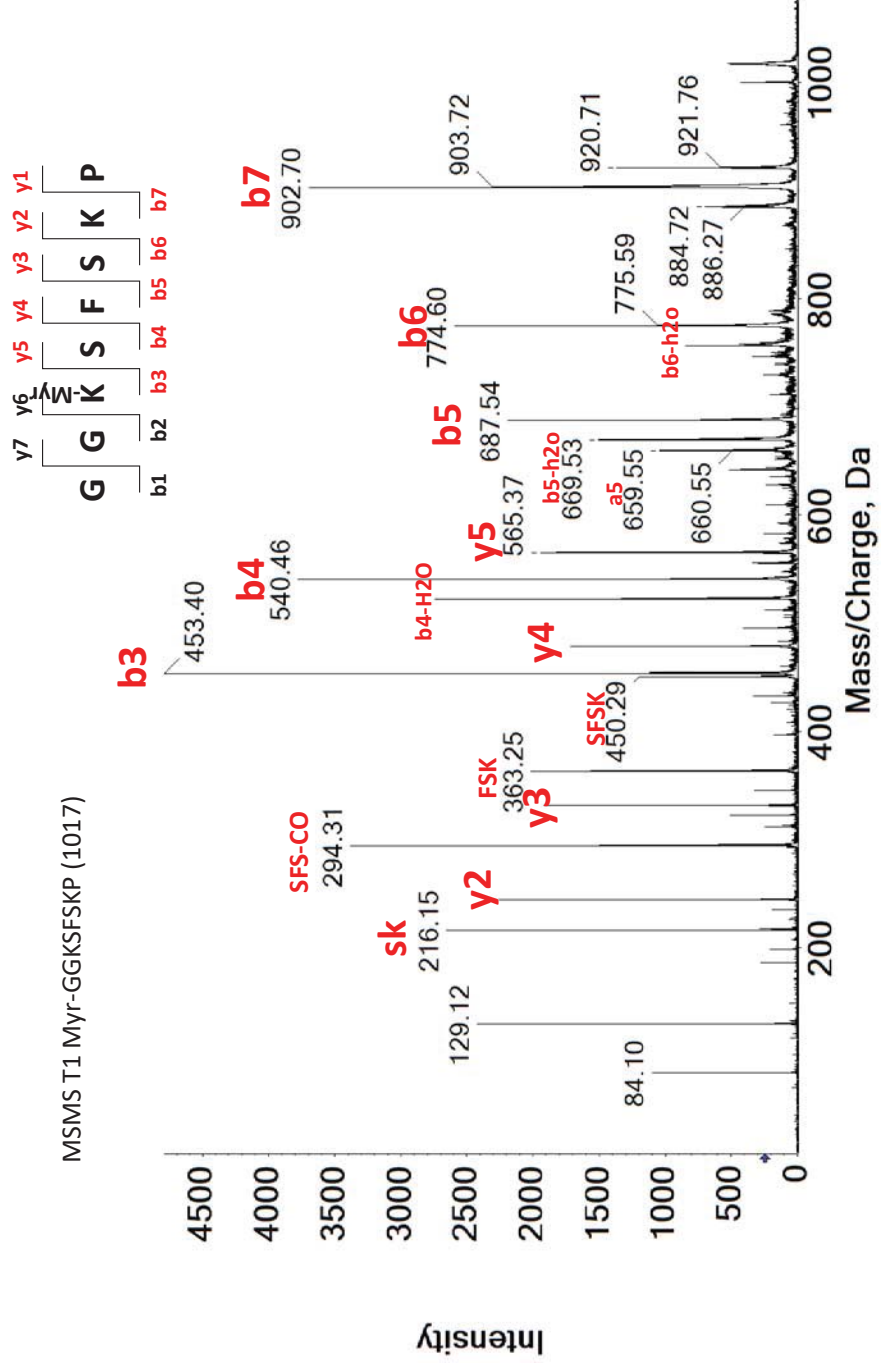


MSMS TO GGKSFSKP (807)

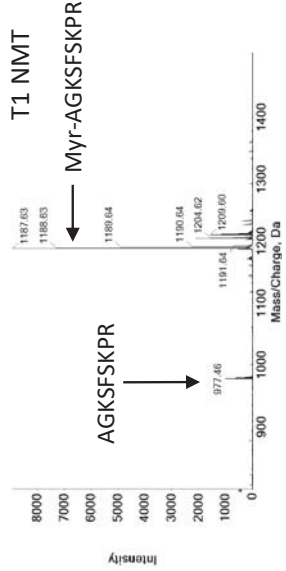
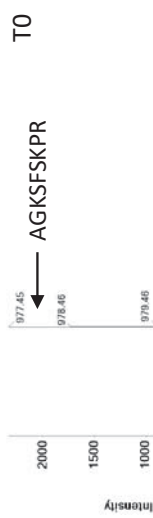


MSMS T1 Myr-GGKSFSKP (1017)

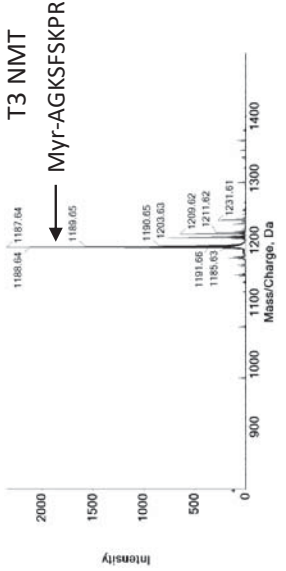
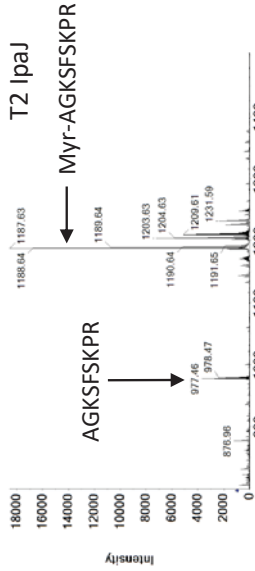




AGKFSFKPR



T2 Ipai



L12

MSMS T1 AGK(Myr)SFSKPR (1187)



L13

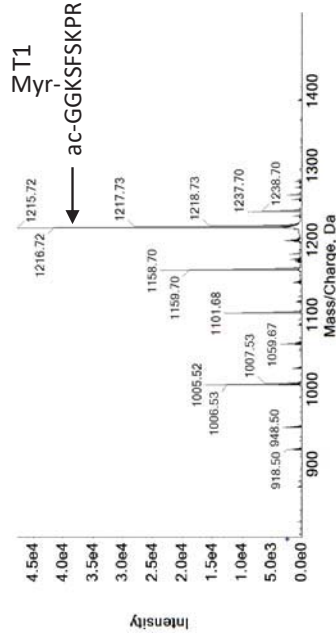
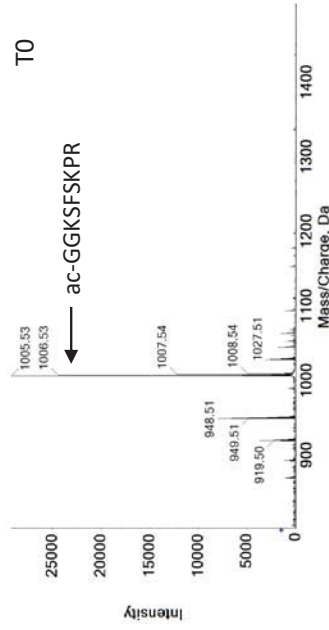
L14

MSMS T2 AGK(Myr)SFSKPR (1187)

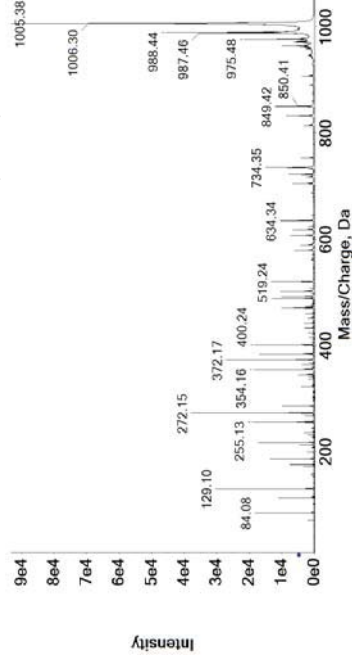


L15

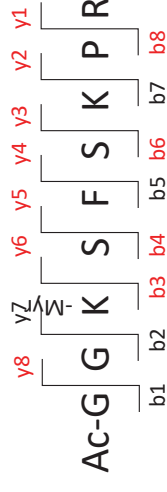
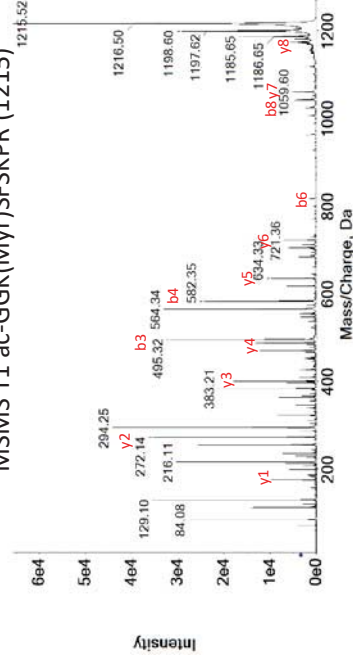
Ac-GGKSFSPKPR



MSMS T0 ac-GGKSFSPKPR (1005)

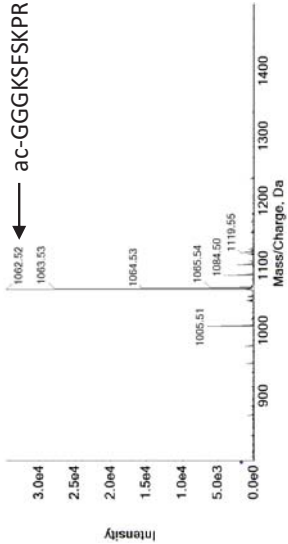


MSMS T1 ac-GGK(MYr)SFSKPR (1215)

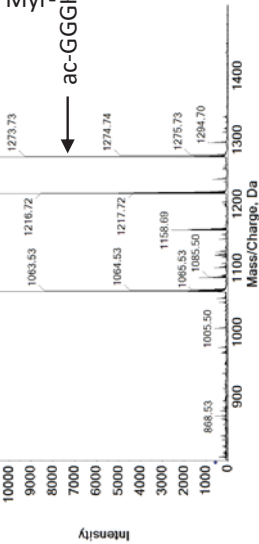


Ac-GGGKSFSPKPR

T0



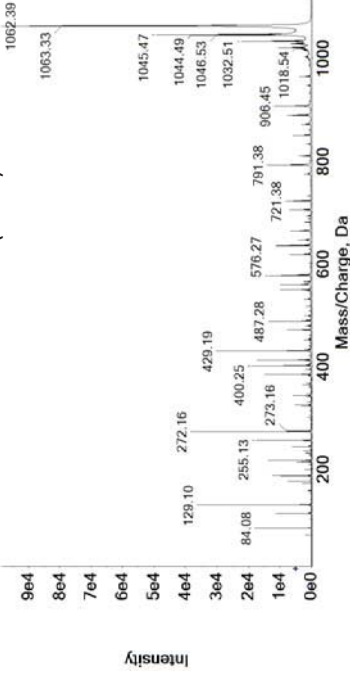
T1



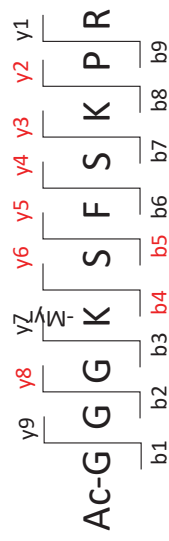
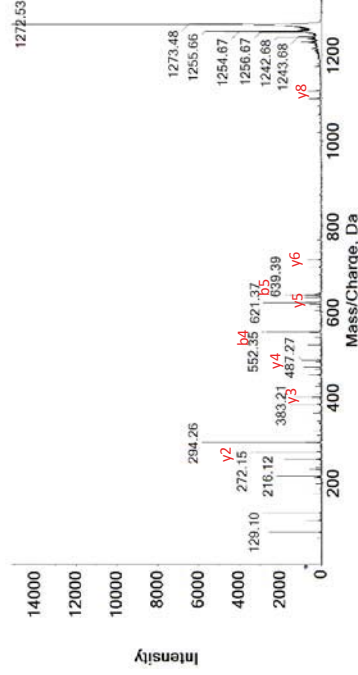
Myr

ac-GGGKSFSPKPR

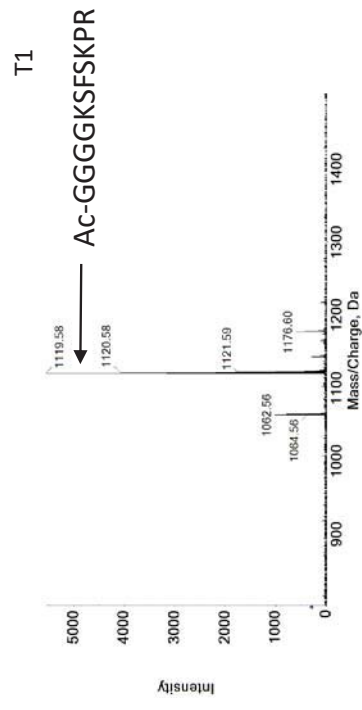
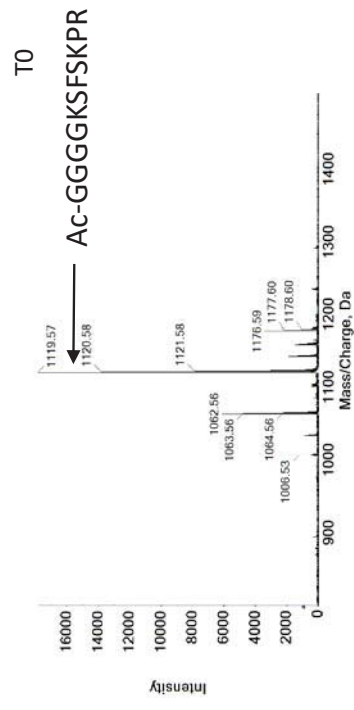
MSMS T0 ac-GGGKSFSPKPR (1062)



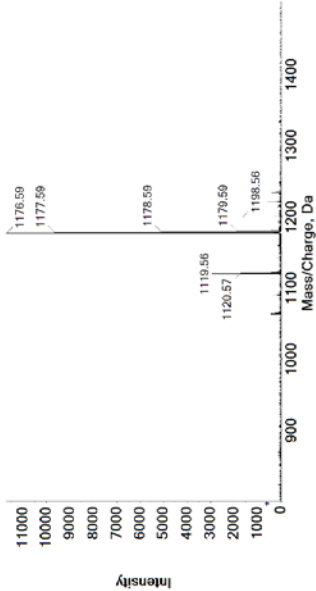
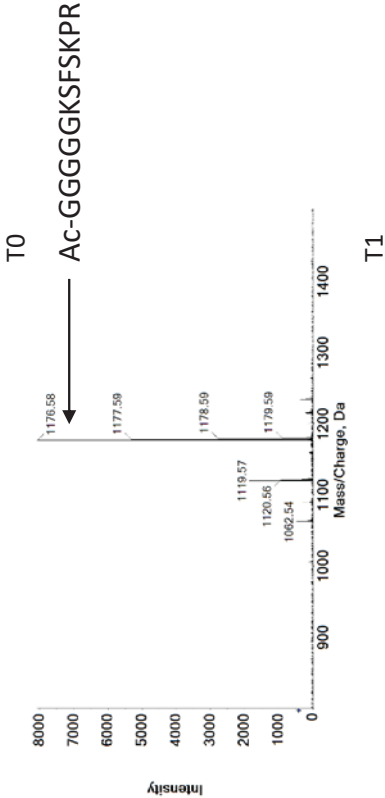
MSMS T1 ac-GGGK(Myf)SFSPKPR (1272)



Ac-GGGGKSFSKPR



Ac-GGGGGKSFSPR

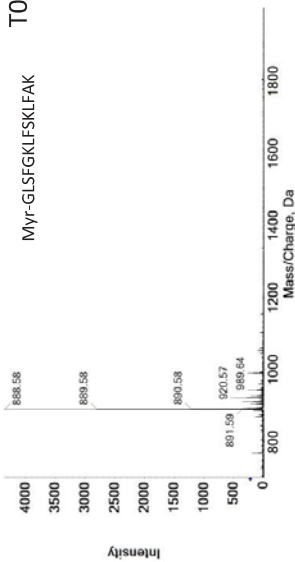


VIII.2.9. Données supplémentaires 1d

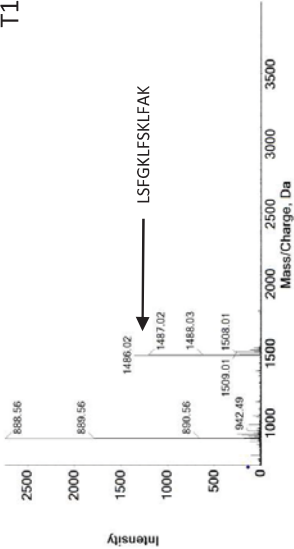
Sup Dataset 1d

Sequence	Cleavage by IpaJ	MS data in Supplementary Dataset 1e
Myr-GLSFGKLF SKLFAK	+	2
Myr-GSMFSGNR	+	3
Myr-GQGPSGGLNR	+	4
Myr-GNLHGIHR	+	5
Myr-GKVLSKI FGNK	+	6
Myr-GNCF SKPR	+	7
aC-GNCF SKPR	-	8
Myr-ATNGSKVA	-	9
	IpaJ cleavage after MYR	
GIGWSLHN	+	10
GNLHGIHR	+	11
GCFHSKAA	+	12
GHRHKS SK	+	13
GMVFGKIA	+	14
GGTSTRR	+	15
GKVLSKI F	+	16
GFCFSKFG	+	17
GRKWSGPT	+	18
GTSVSKPV	+	19

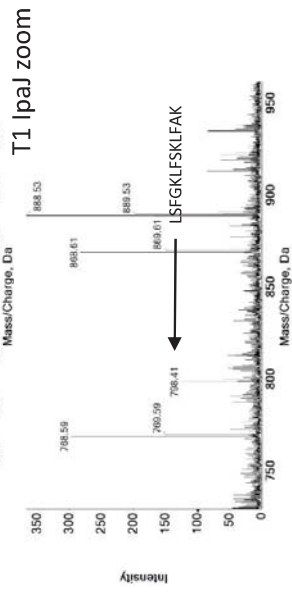
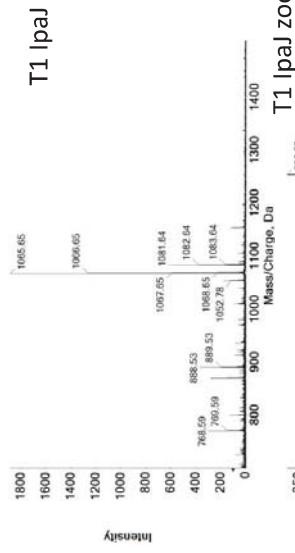
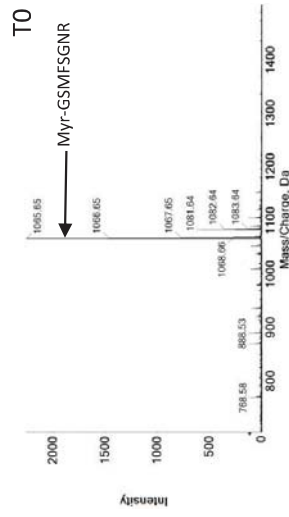
Myr-GLSFGKLFSLFAK



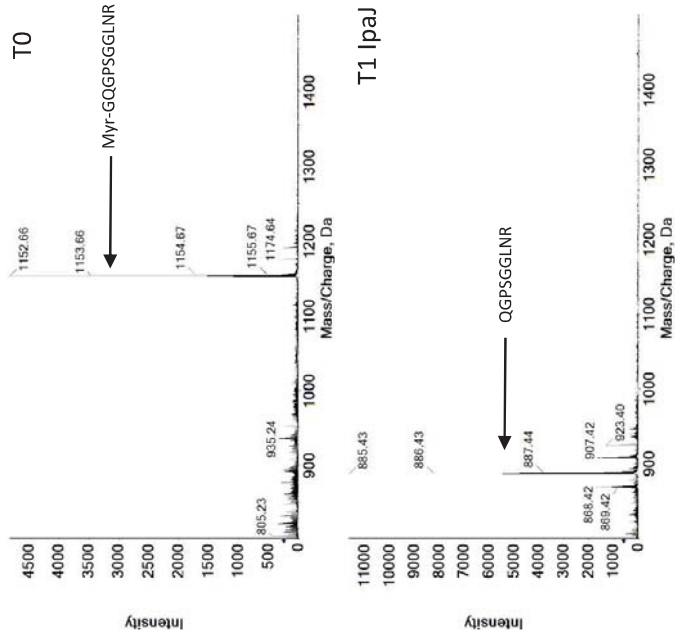
T1 IpJ



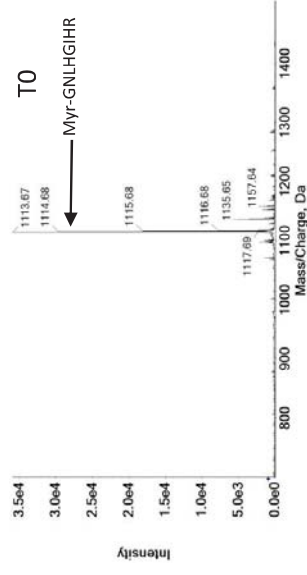
Myr-GSMFSGNR



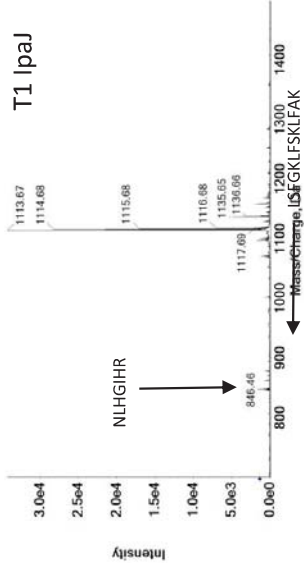
Myr-GQGPGGLNR



Myr-GNLHGIHR

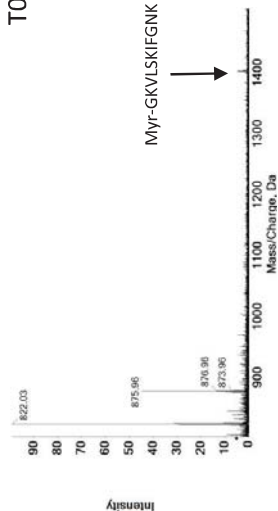


T1 Ipaj

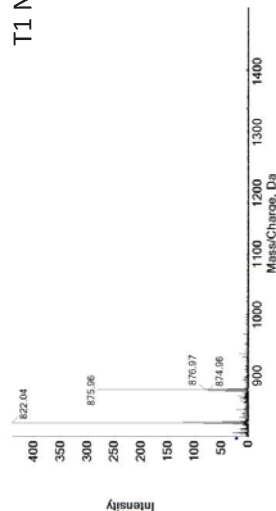


Myr-GKVLISKIFGNK

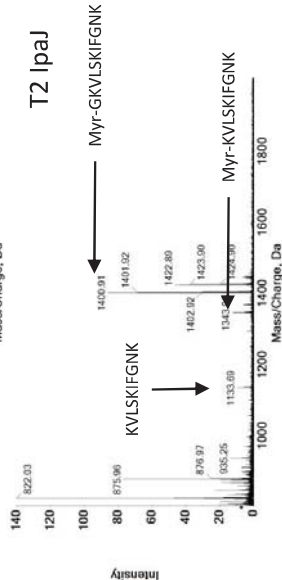
T0



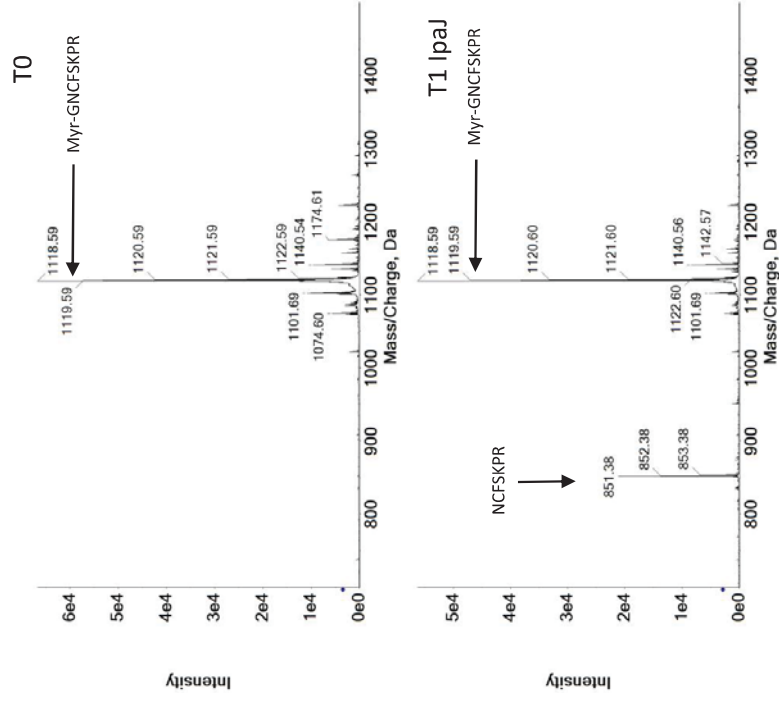
T1 NMT



T2 Ipaj

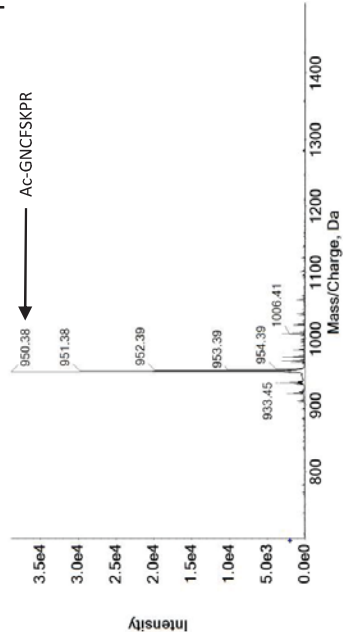


Myr-GNCFSKPR

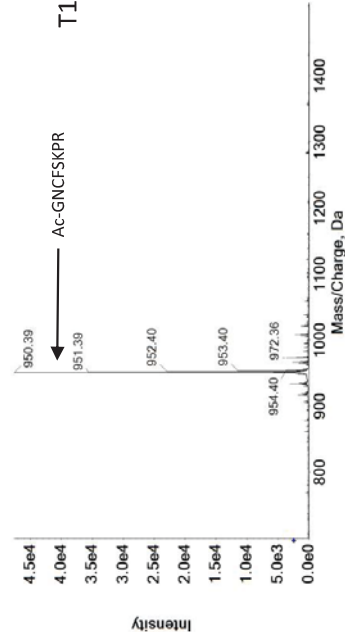


Ac-GNCFSKPR

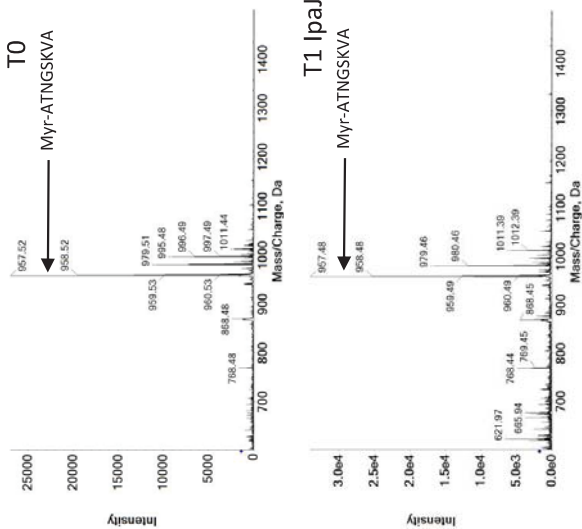
T0



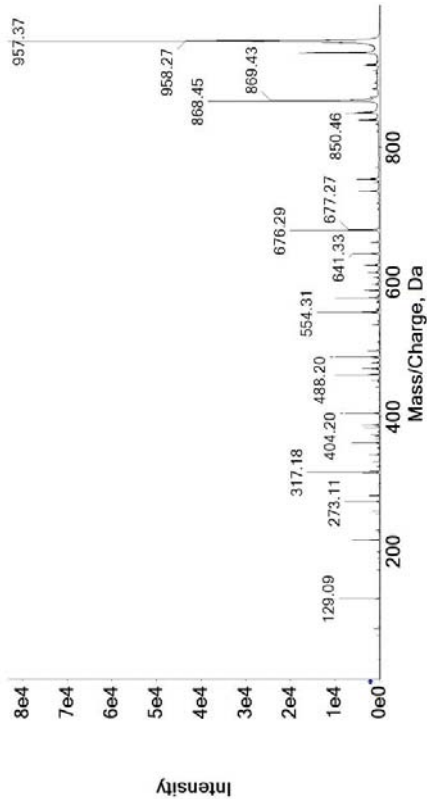
T1 lpaJ



Myr-ATNGSKVA



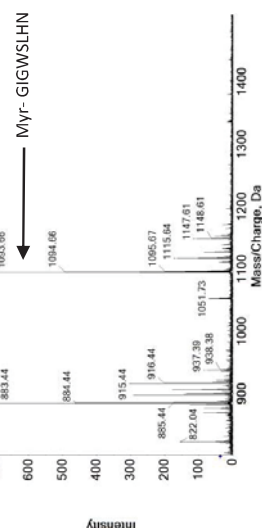
MSMS TO Myr-ATNGSKVA (957)



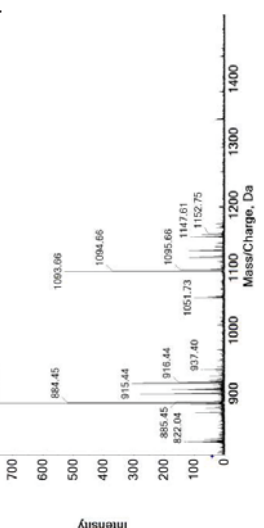
GIGWSLHN



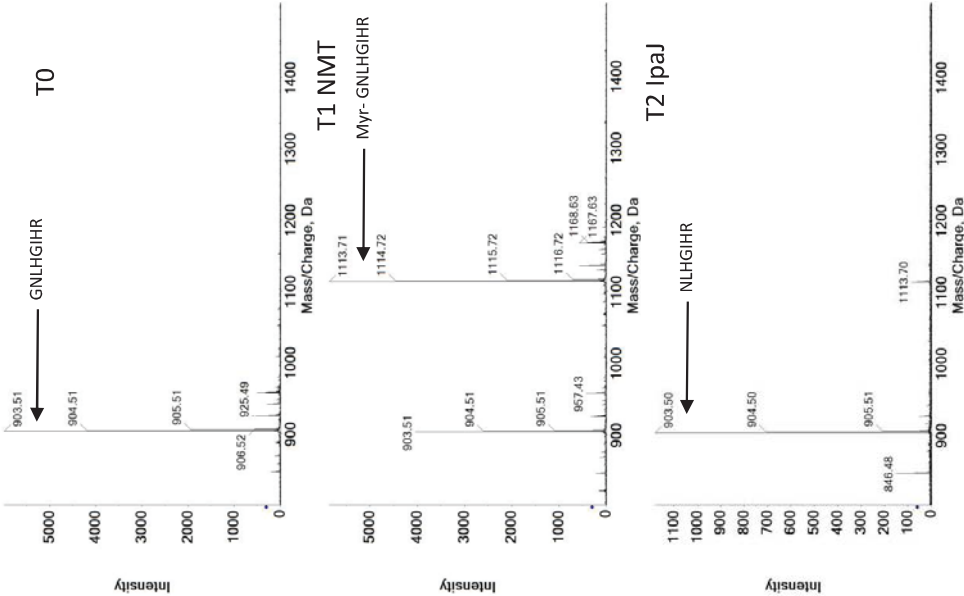
T1 NMT



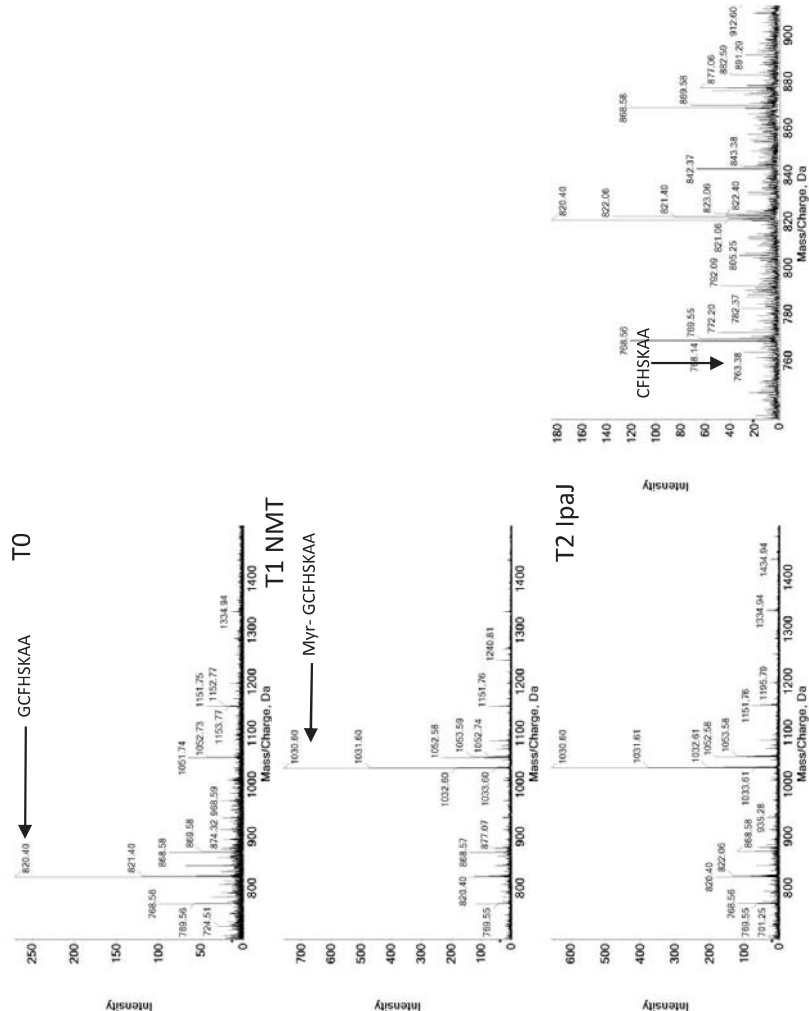
T2 IpaJ



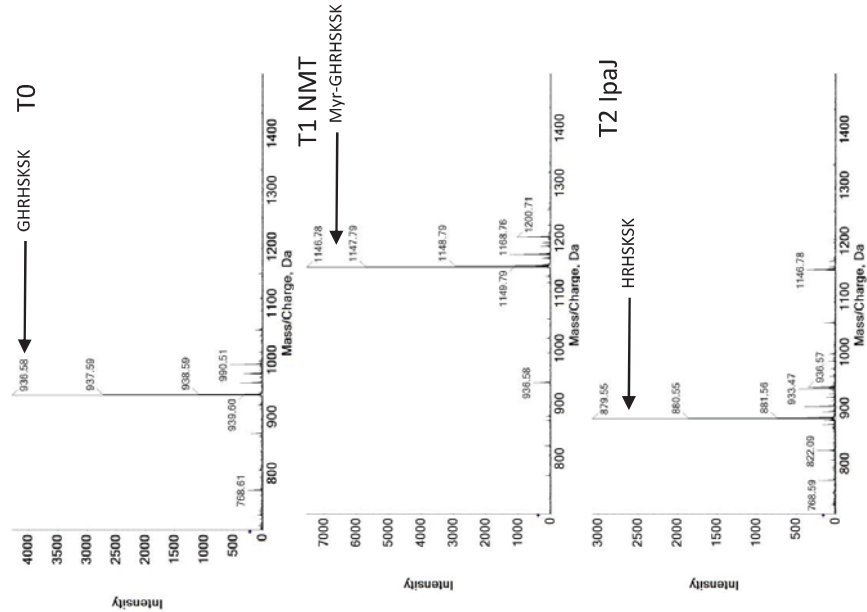
GNLHGIHR



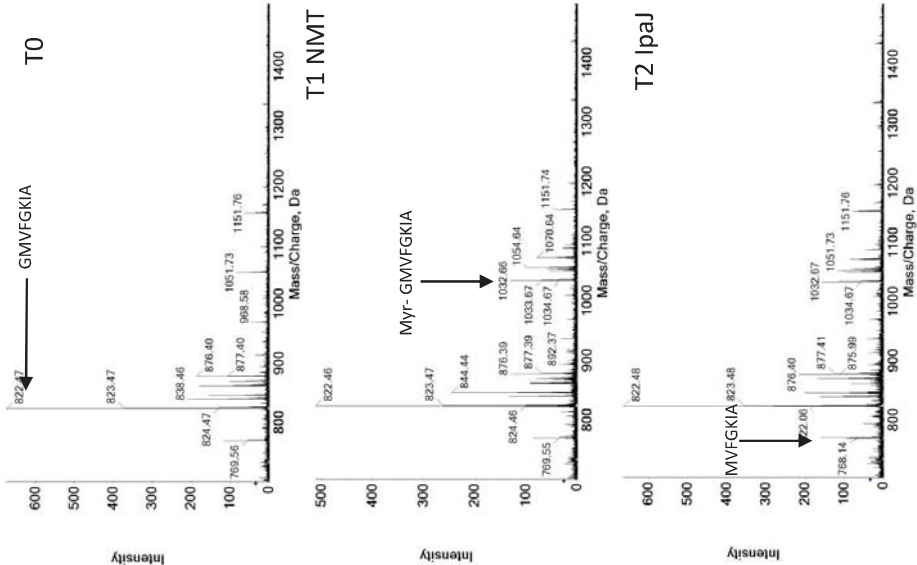
GCFHSKAA



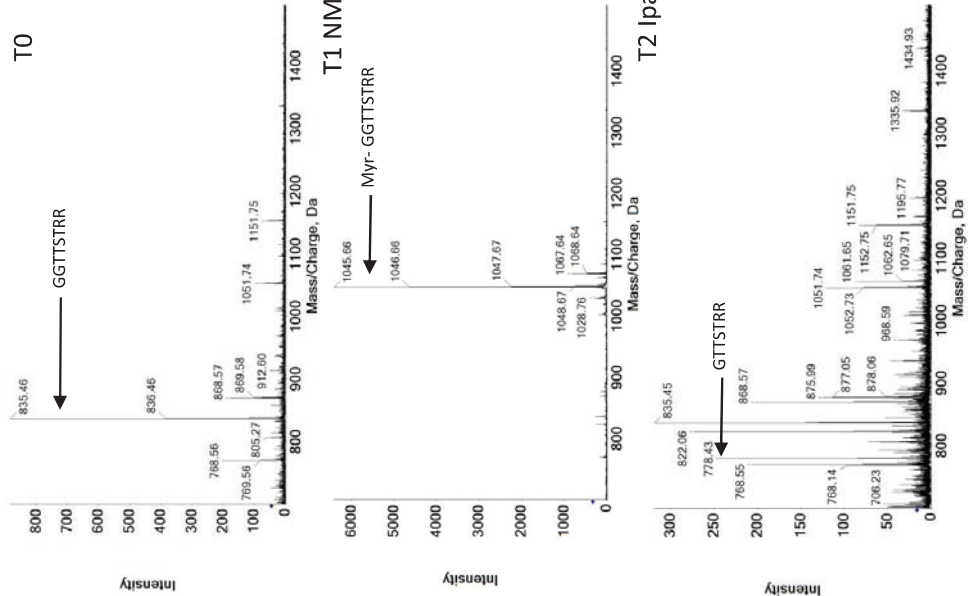
GHRHSKSK



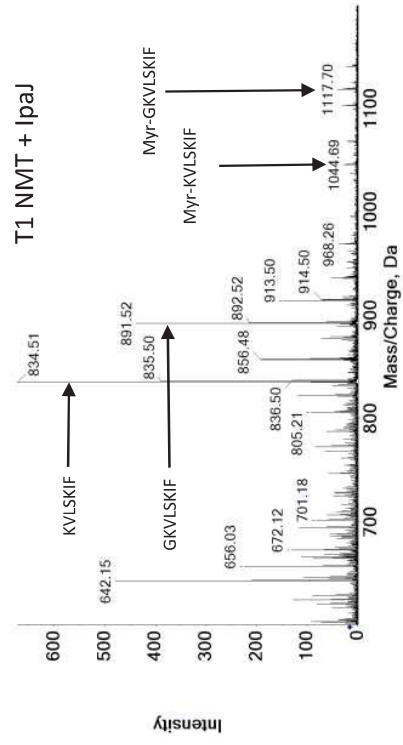
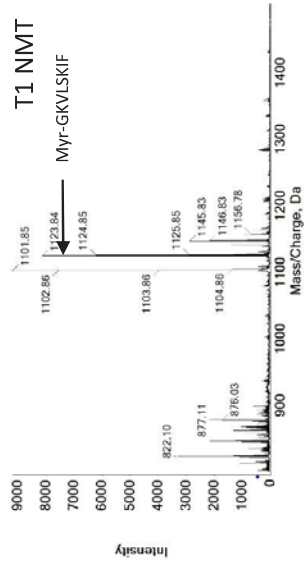
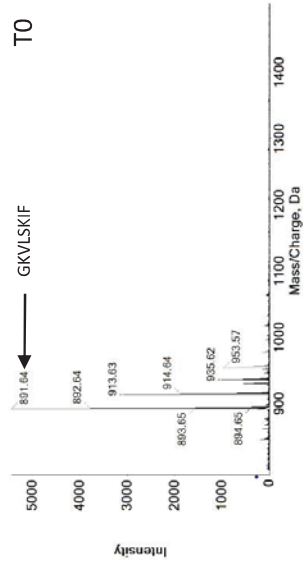
GMVFGKIA



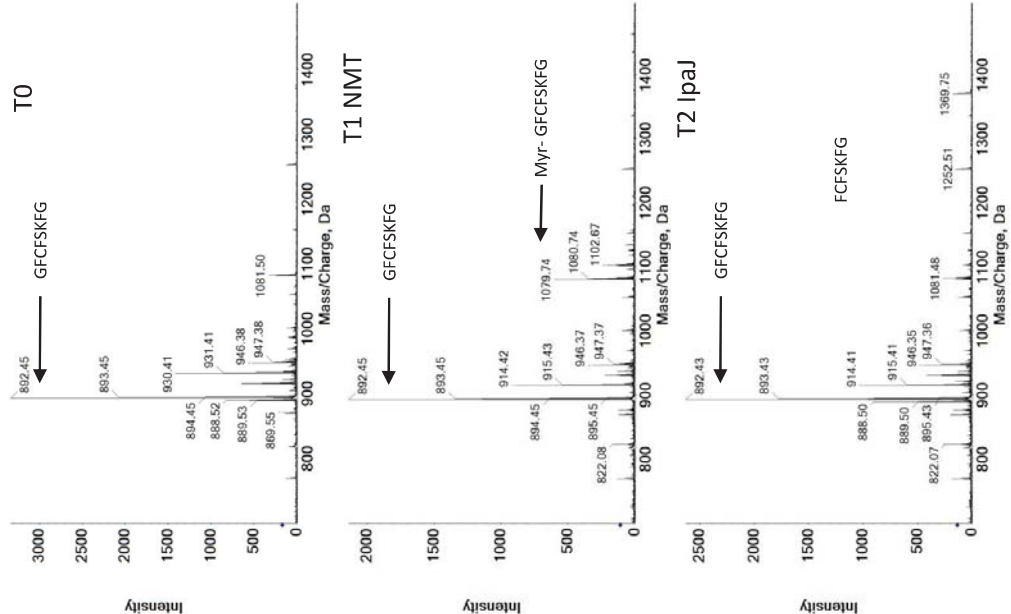
GGTTSTRR



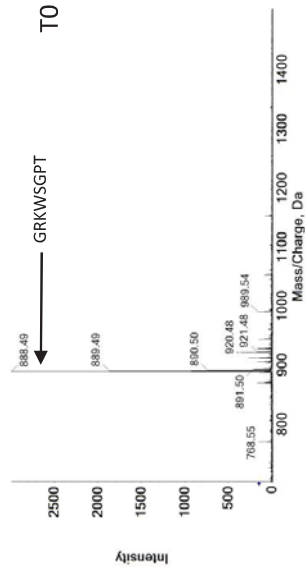
GKVLSKIF



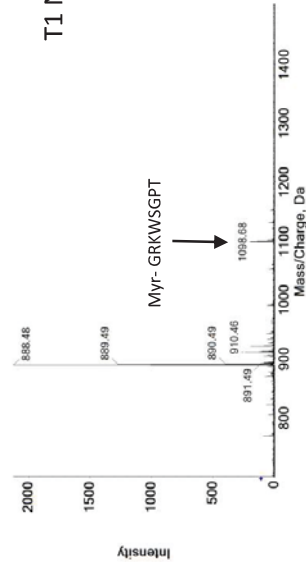
GFCFSKFG



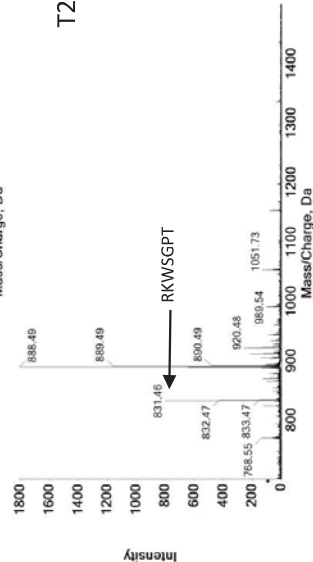
GRKWSGPT



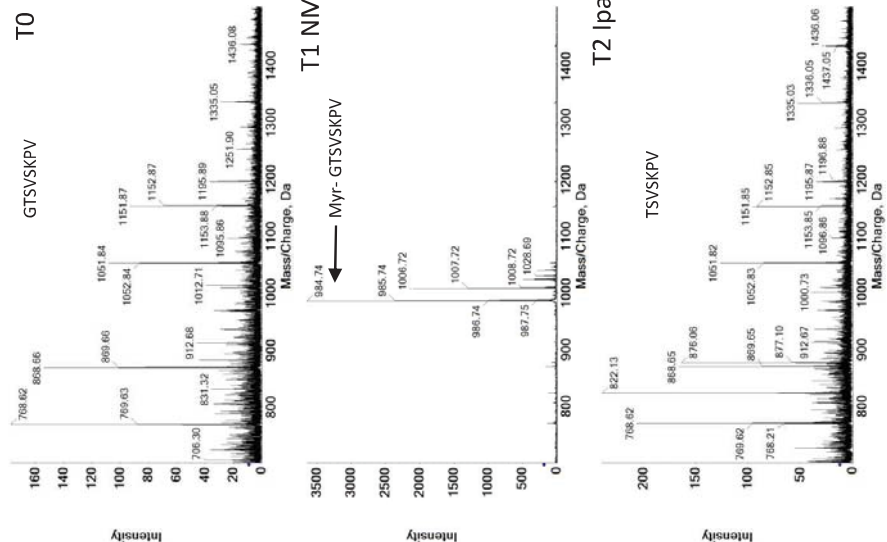
T1 NMT



T2 Ipaj



GTSVSKPV

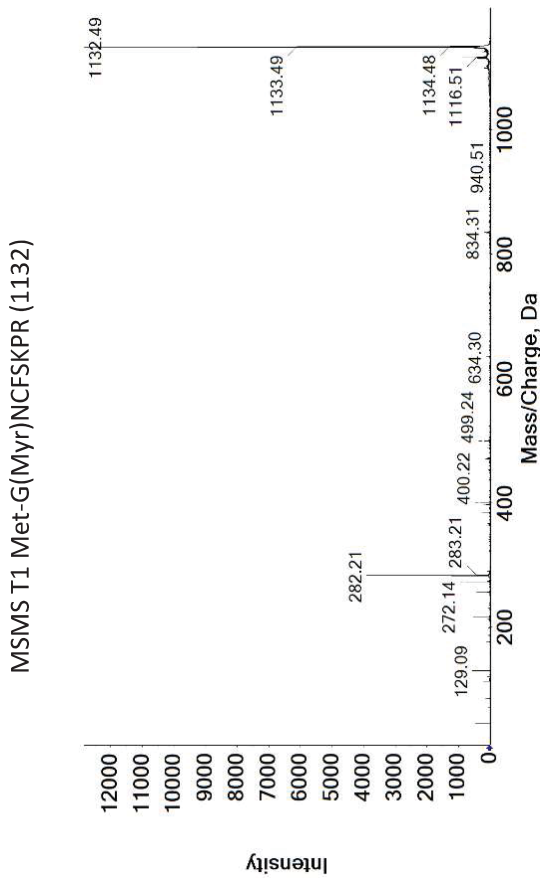
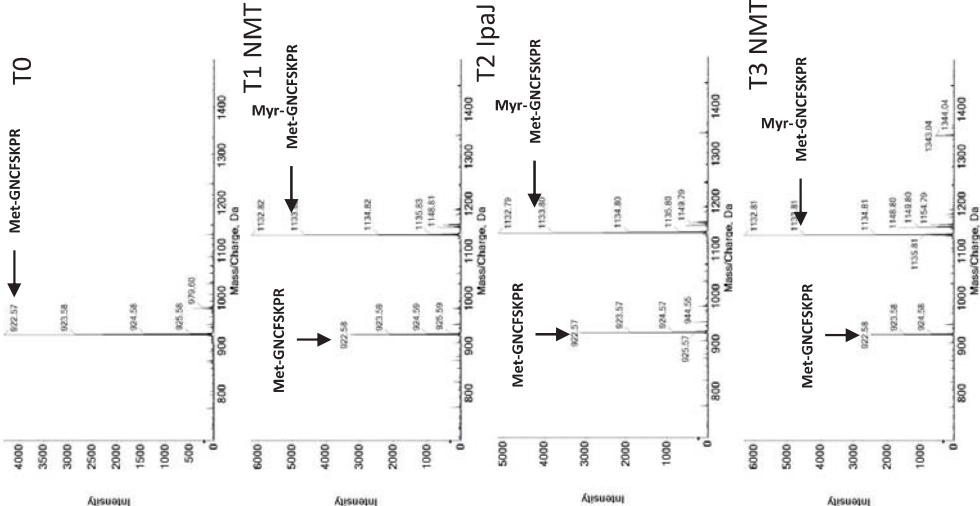


VIII.2.10. Données supplémentaires 1^e

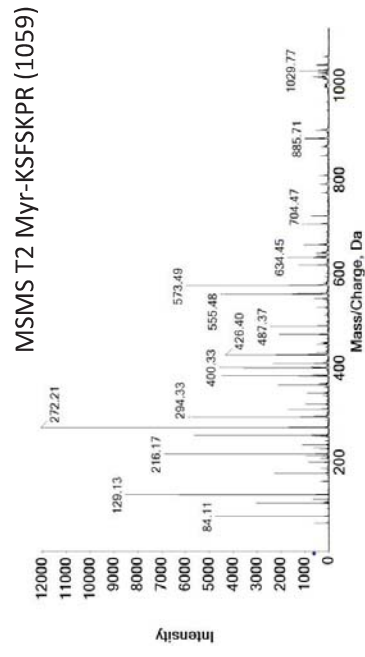
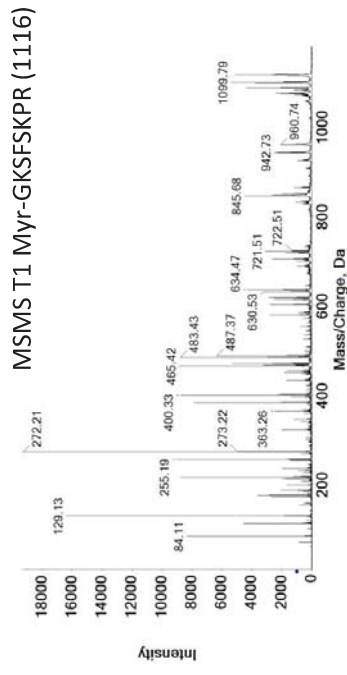
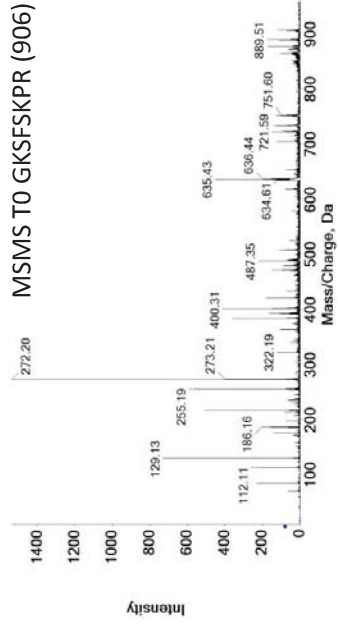
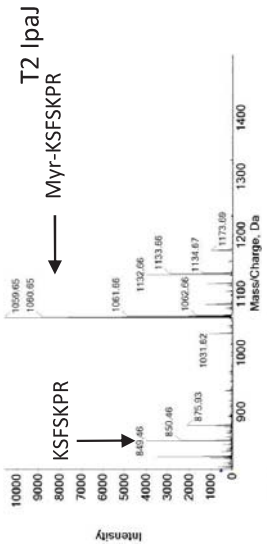
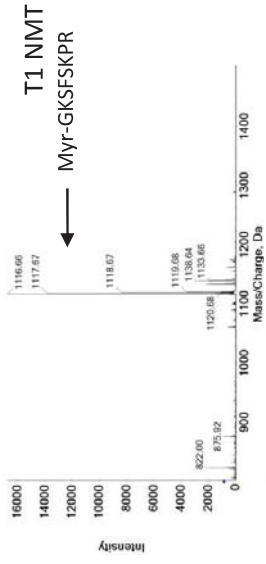
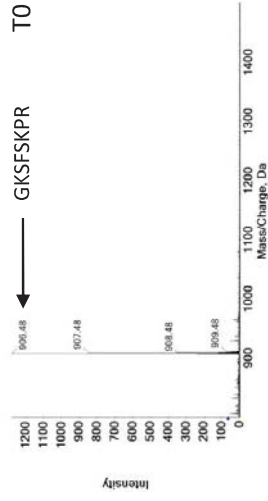
Sup Dataset 1e

Peptide sequence	MS data in Supplementary
met-GNCFSKPR	2
GKSF SK PR	3
G{ORN}SFSKPR	4
GKVL SK IFGNK	5/6/7/8
GKVL SK IFGN	5/6/7/8
GKVL SK IF	5/6/7/8
Myr-GKVL SK IFGNK	9
G-K(ϵ -MYR)VLSKIFGNK	10
GKFSSKPR	11
GKQNSKLR	12
KQNSKLRP	13
GKQNSKLA	14
GKSNSKLG	15
GKTN SK LA	16
GKSASKQF	17
GA KQSG PA	18
GGKQSTAA	19
GGKLSKKK	20
GGKWSKLS	21
AKSFSKPR	22
GKSWSKGR	23
AKQNSKLR	24
GGKQSKAA	25
GGKSFSKP	26
GGKFSKPR	27/28
AKPTSKDSGLK	29
AGKFSKPR	30
AGKSFSKPR	31

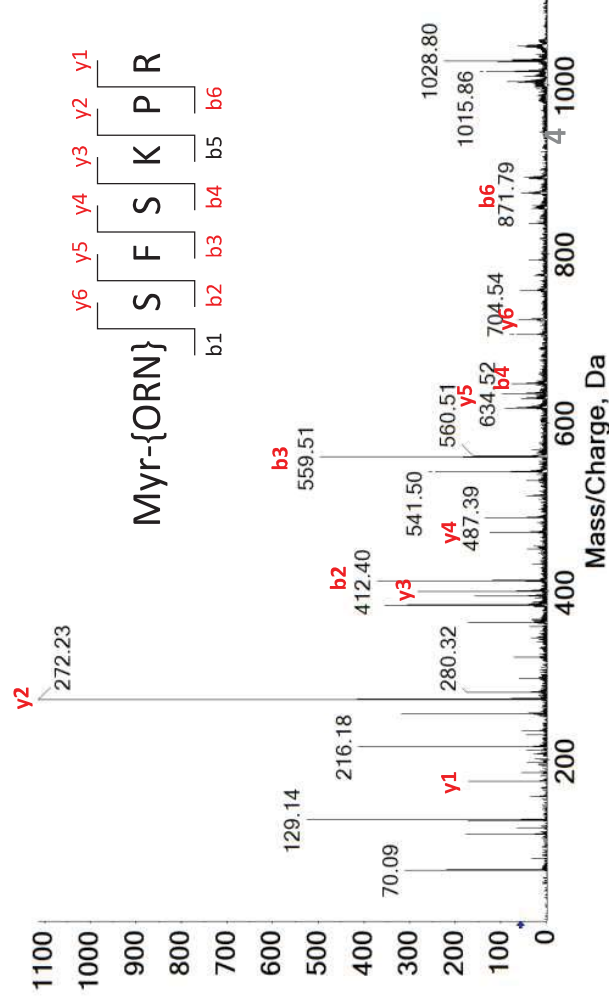
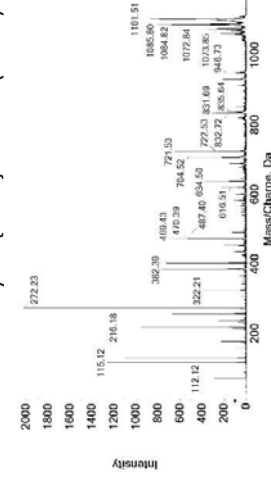
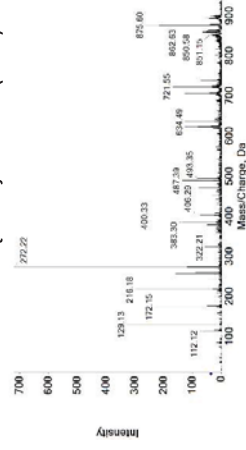
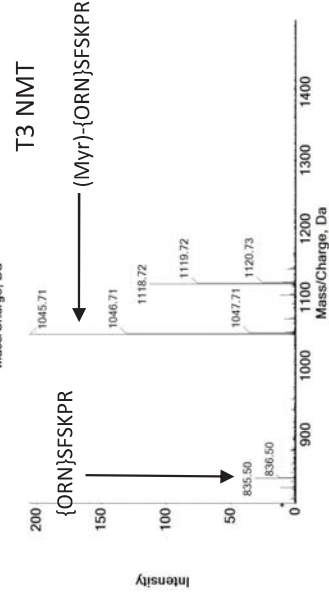
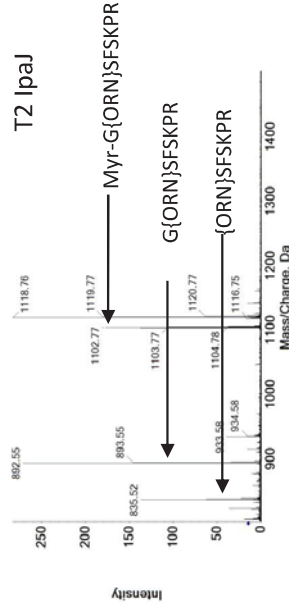
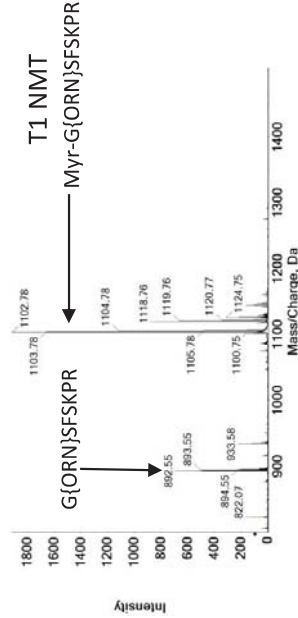
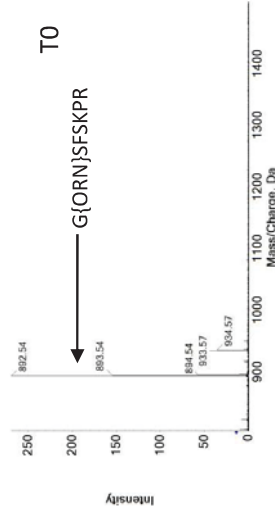
Met-GNCFSKPR



GKSFSKPR



G{ORN}SFSKPR

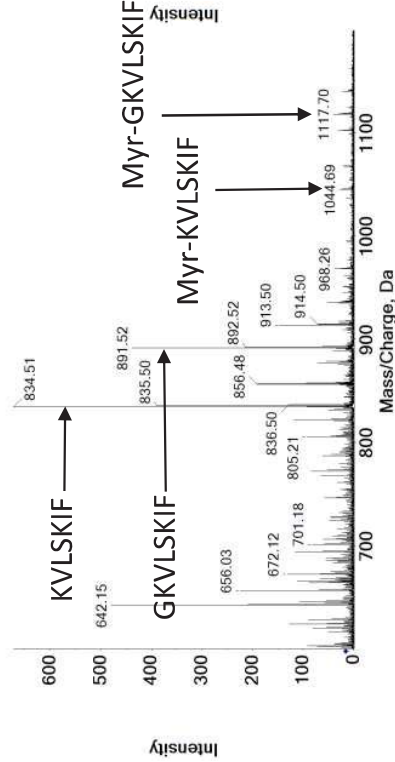


ARF6 +/- Long NMT 1

Final point after NMT IpaJ NMT

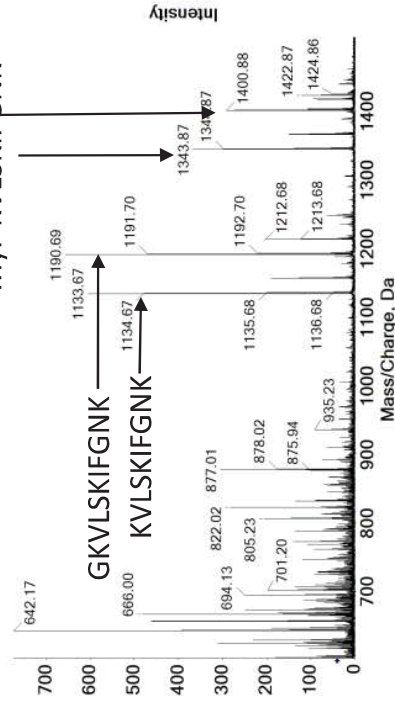
KVLISKIF

T1 NMT + IpaJ + NMT



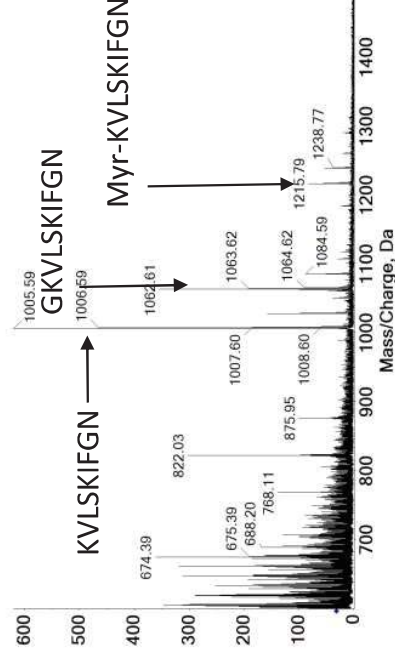
GKVLSKIFGNK

T1 NMT + IpaJ + NMT

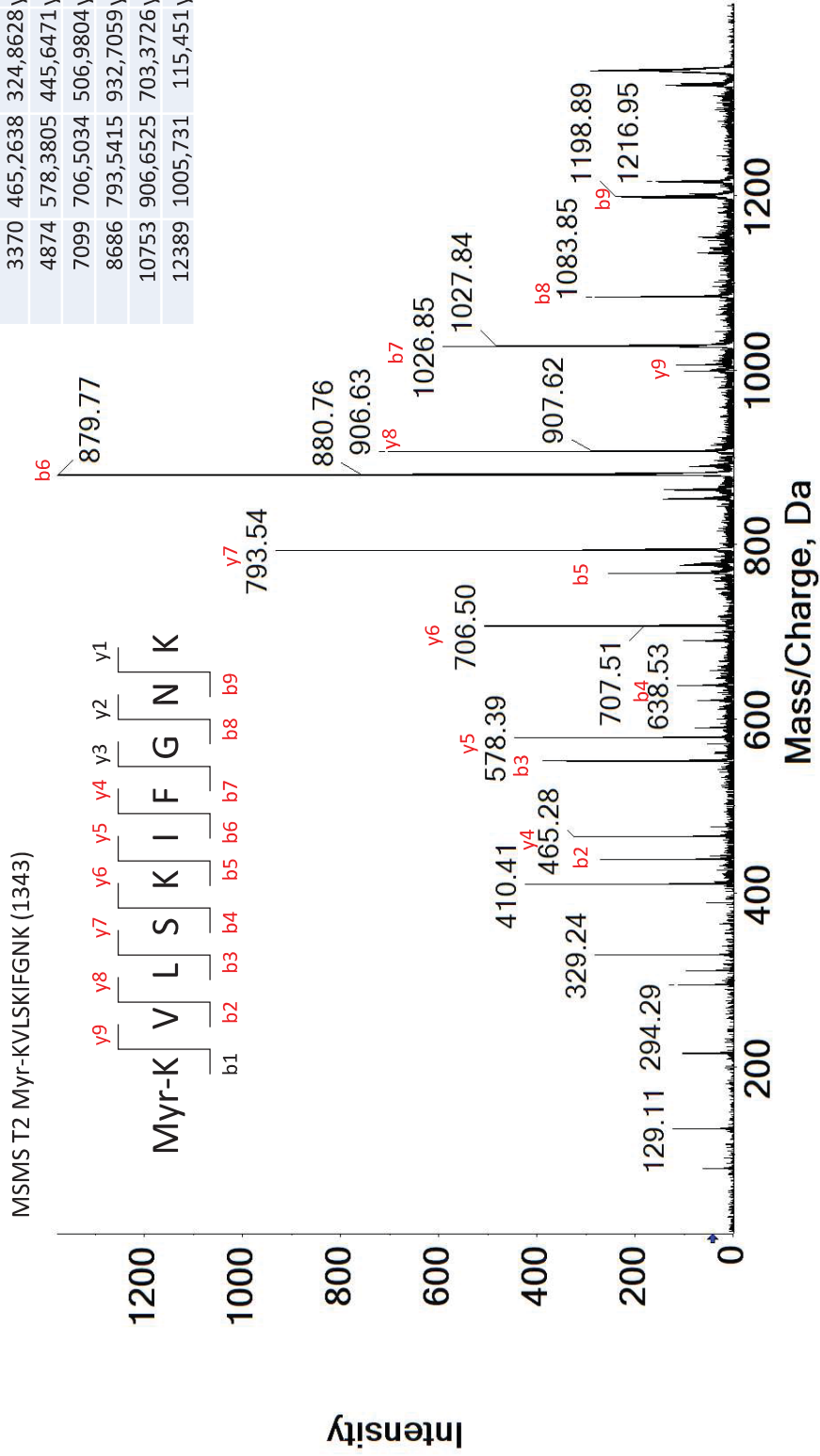


GKVLSKIFGN

T1 NMT + IpaJ + NMT



E24_MSMS_1343.9000_2			ion
pic	Mass/Charge (Da)	Intensity	ion
2965	438,4012	270,4314	b2
4415	551,5016	387,9216	b3
5943	638,5012	113,7255	b4
8104	766,6613	253,8039	b5
10210	879,7369	1377,098	b6
12772	1026,816	591,8431	b7
13697	1083,844	281,4118	b8
15474	1197,918	197,8039	b9
3370	465,2638	324,8628	y4
4874	578,3805	445,6471	y5
7099	706,5034	506,9804	y6
8686	793,5415	932,7059	y7
10753	906,6525	703,3726	y8
12389	1005,731	115,451	y9



ARF6 +/- Long NMT 2

Final point after NMT IpaJ NMT

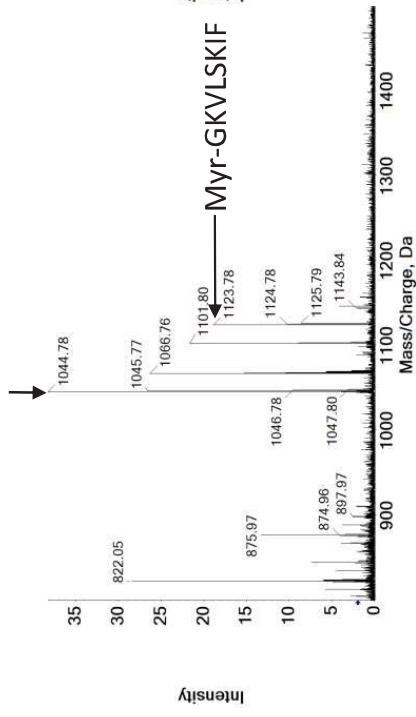
KVLISKIF

GKVLISKIFGNK

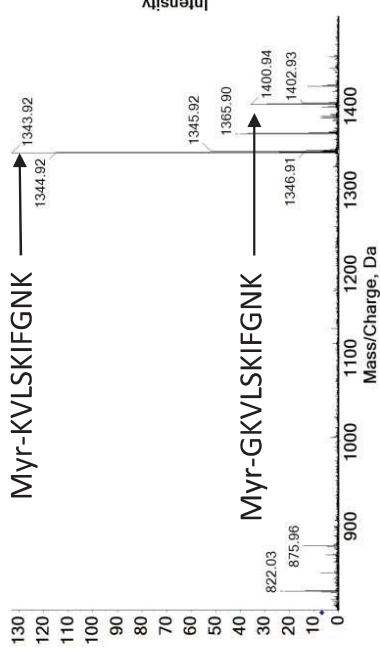
GKVLISKIFGN

T1 NMT + IpaJ + NMT

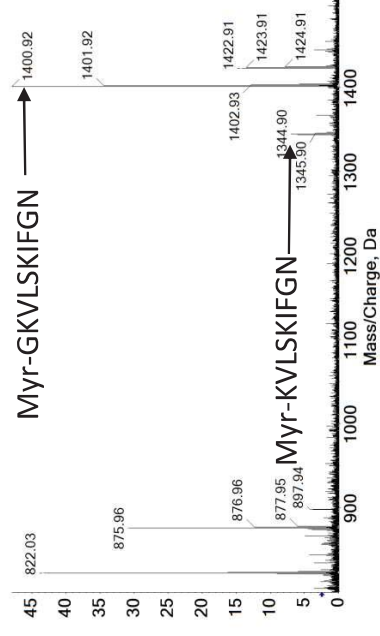
Myr-KVLISKIF



T1 NMT + IpaJ + NMT

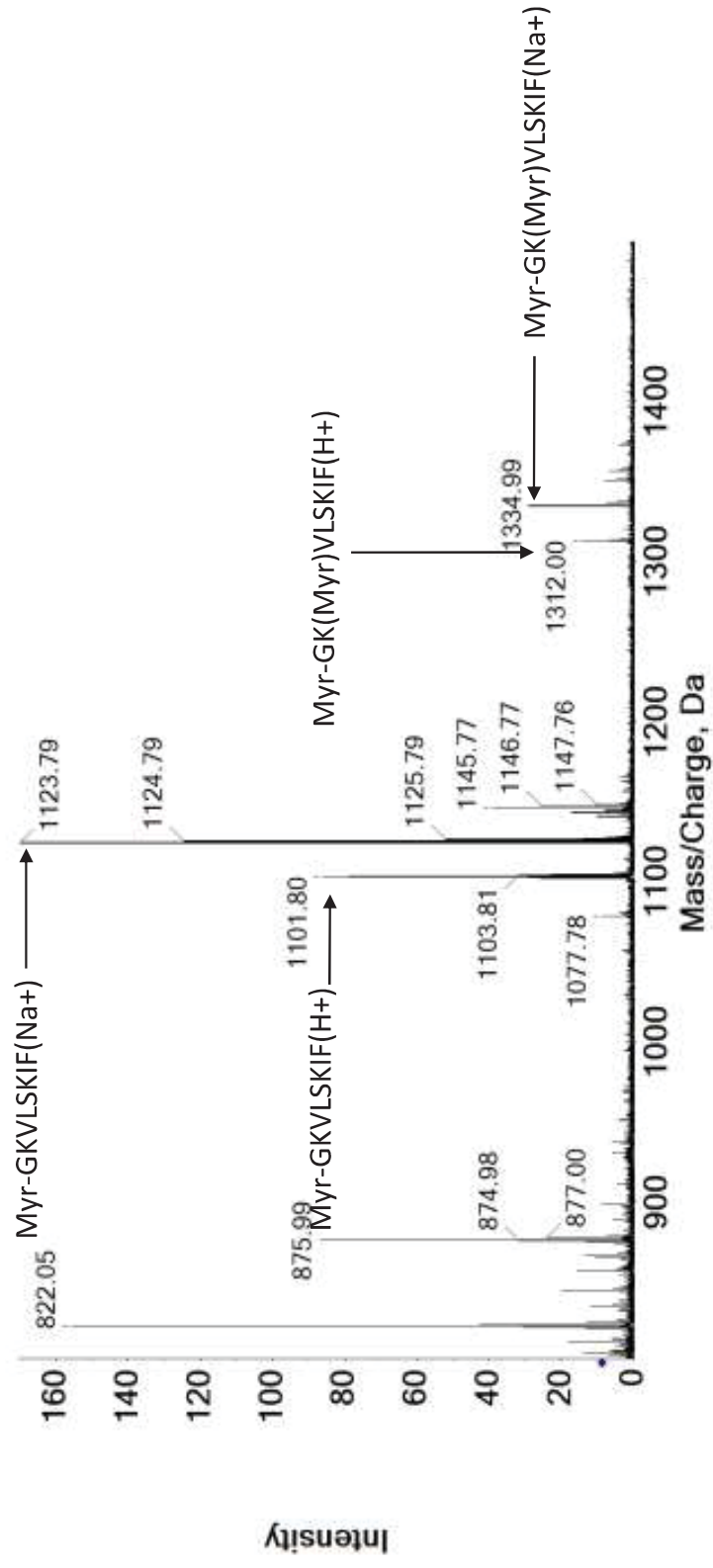


T1 NMT + IpaJ + NMT

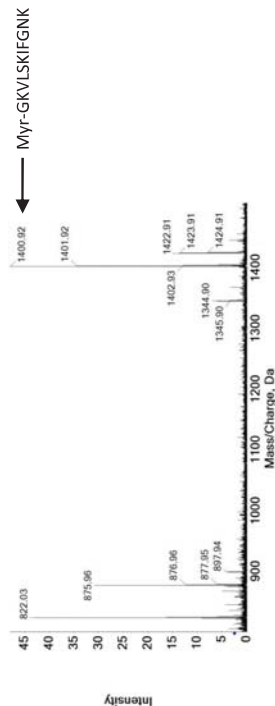
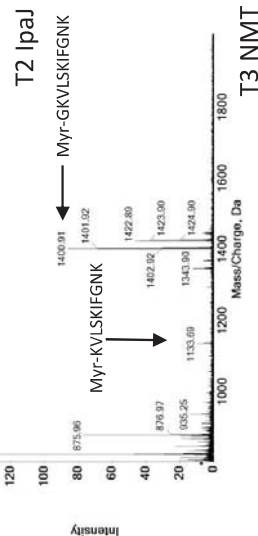
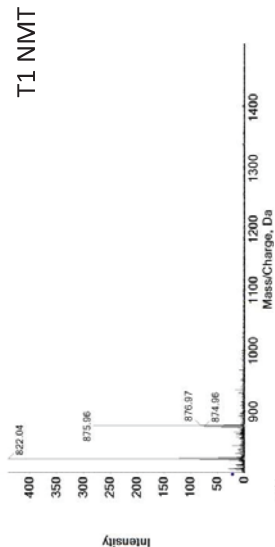
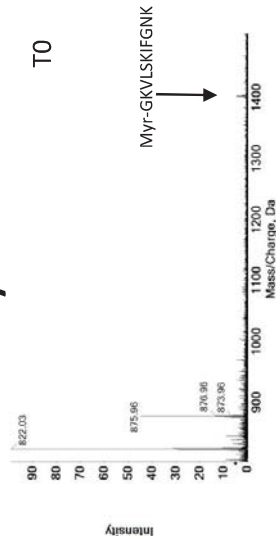


T1 (1h NMT2 30°C)

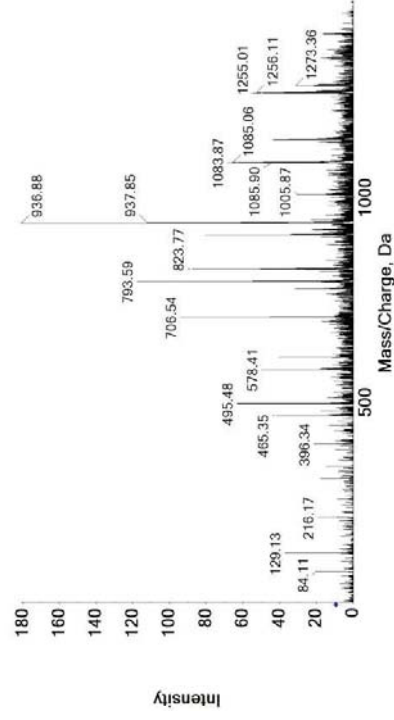
ARF6 short



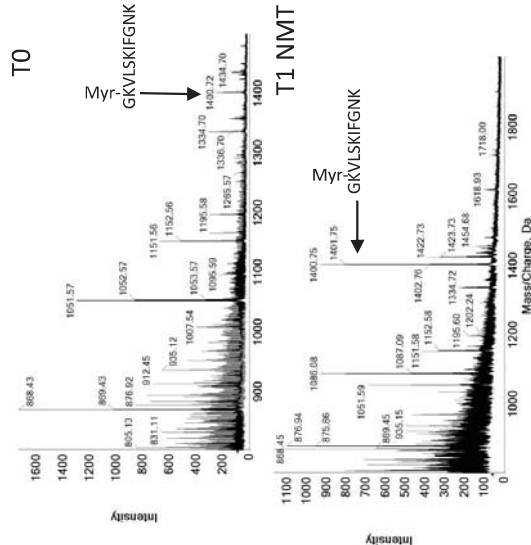
Myr-GKVLISKIFGNK



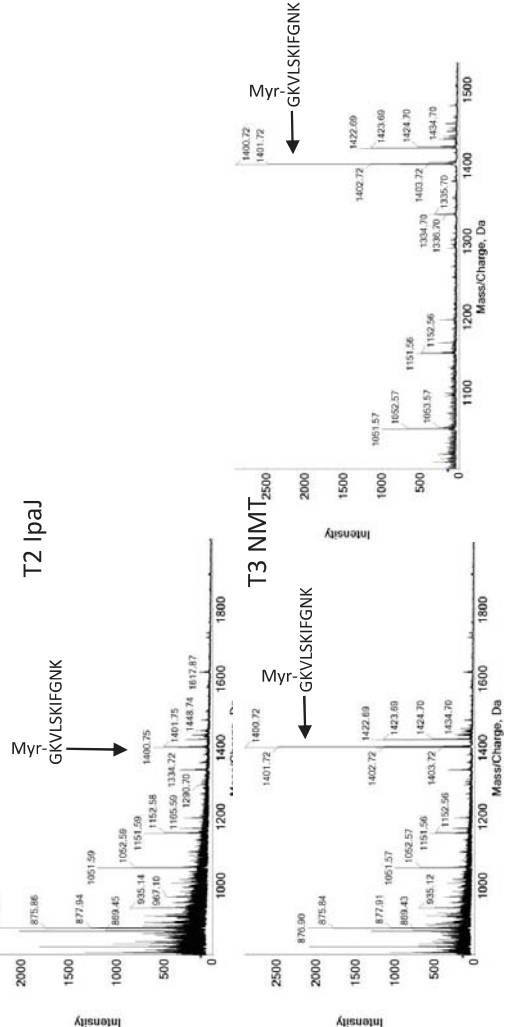
MSMS T2 Myr-GKVLISKIFGNK (1400)



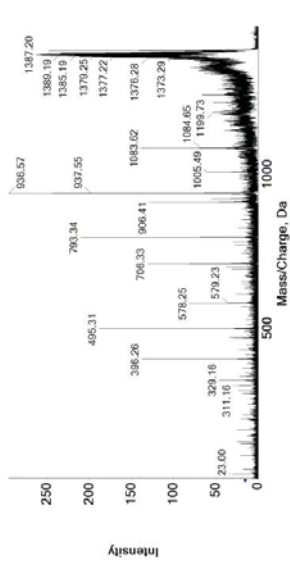
G-K(ϵ -MYR)VLSKIFGNK



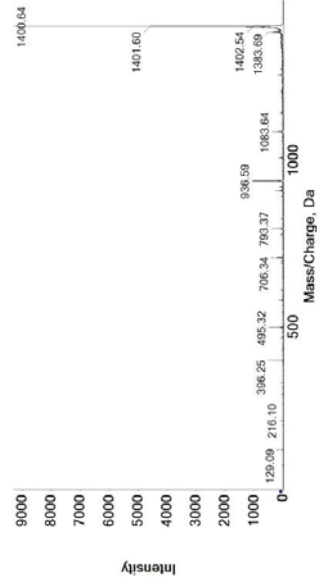
T2 Ipal

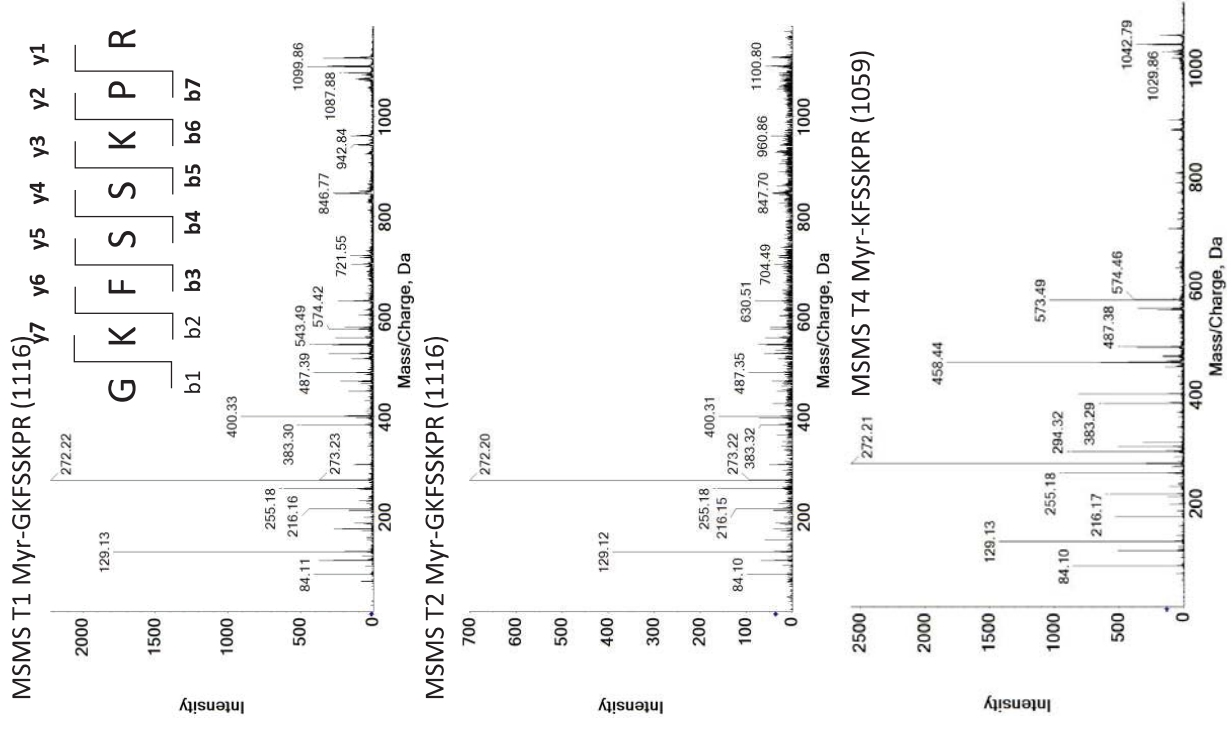
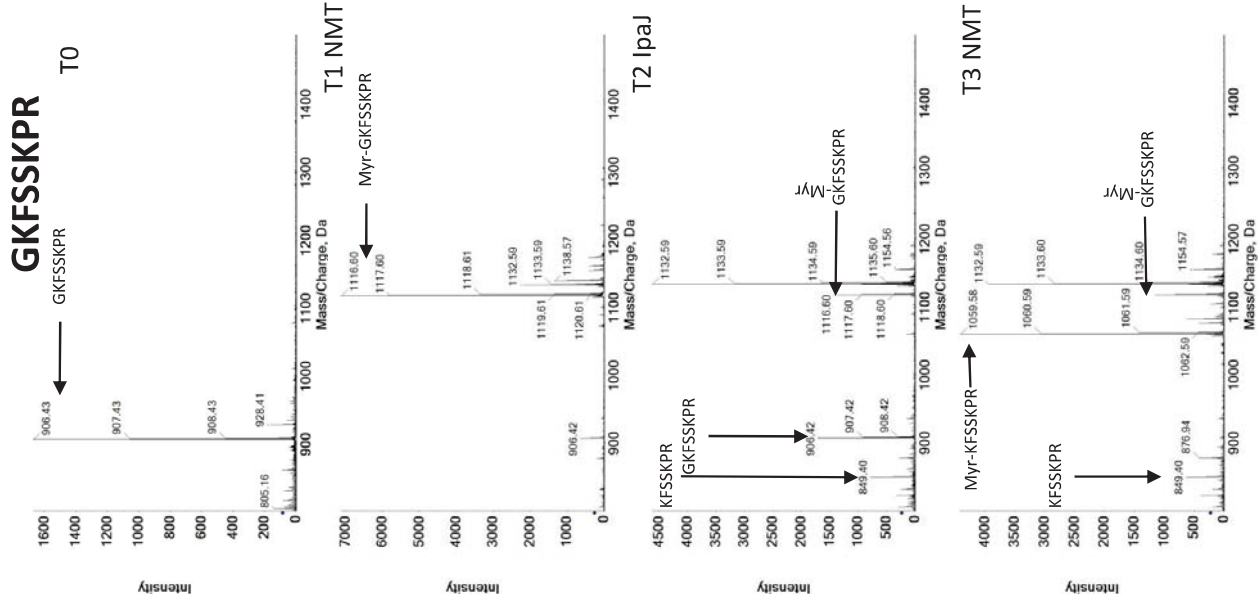


MSMS T1 Myr-GKVLKIFGNK (1400)

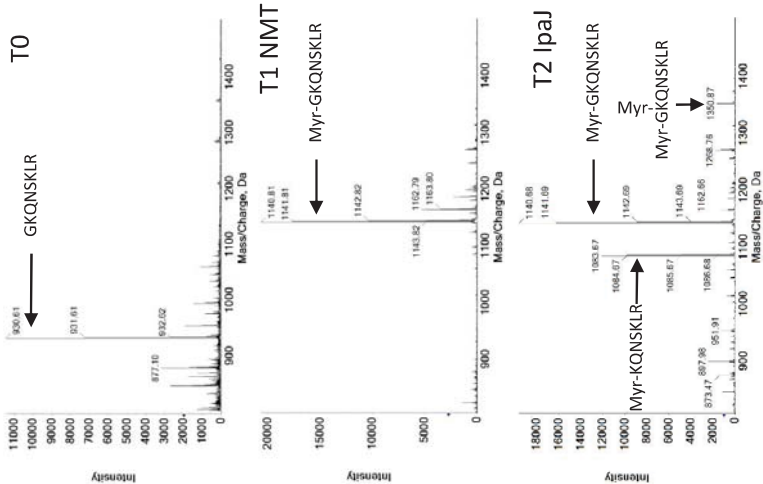


MSMS T3 Myr-GKVLKIFGNK (1400)

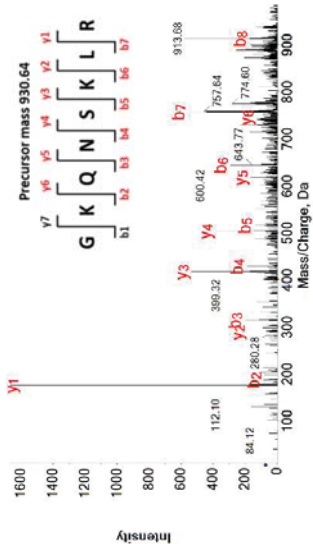




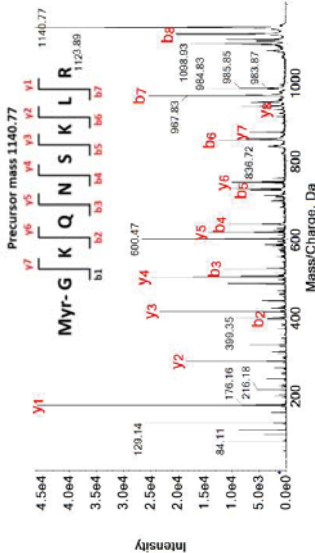
GKQNSKLR



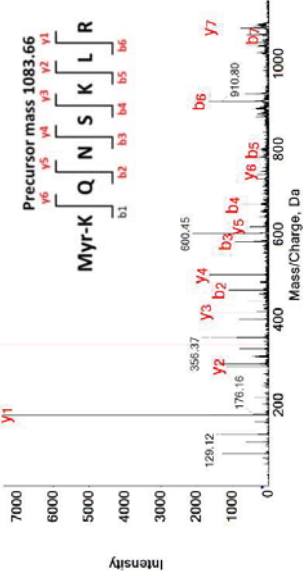
MSMS T0 GKQNSKLR (930)



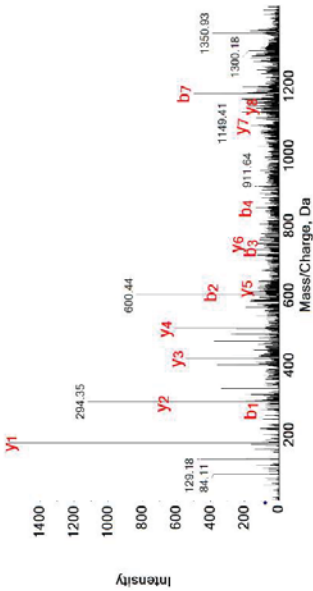
MSMS T1 Myr-GKQNSKLR (1140)



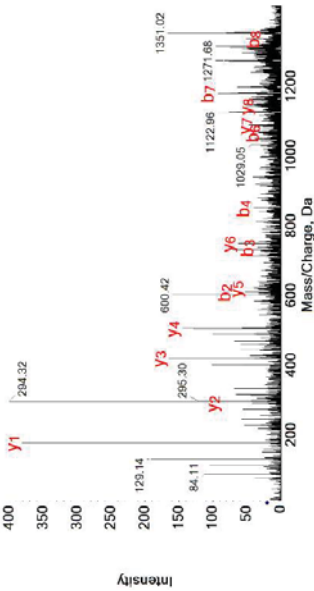
MSMS T2 Myr-KQNSKLR (1083)



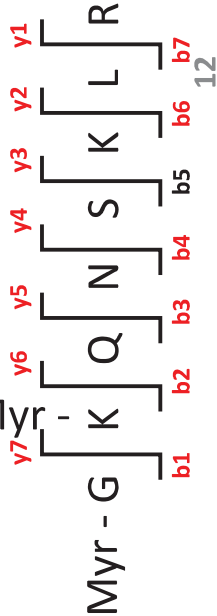
MSMS T2 Myr-GK(Myr)QNSKLR (1350)



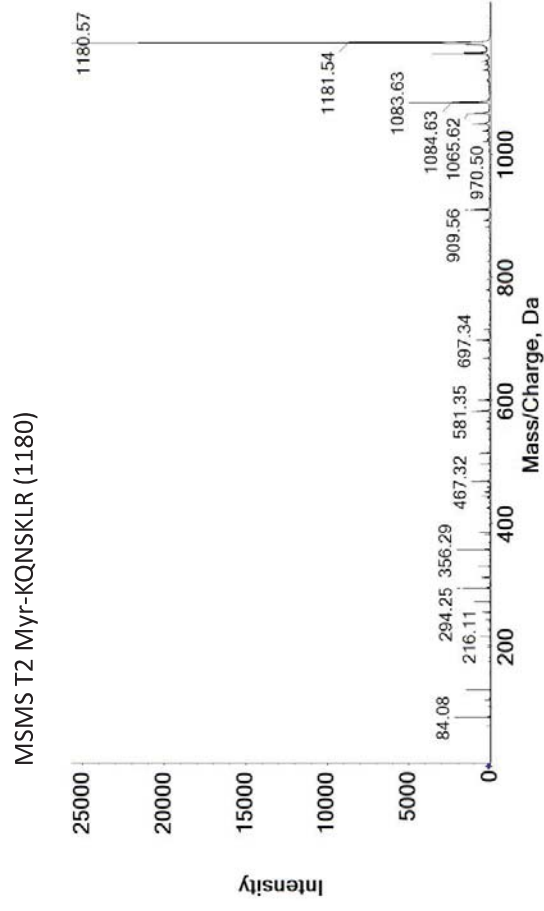
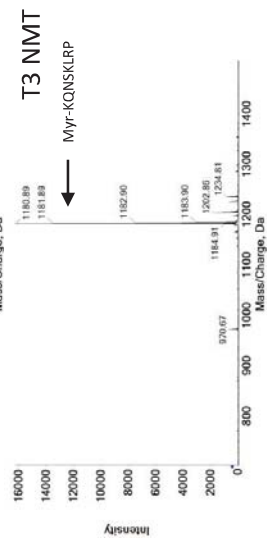
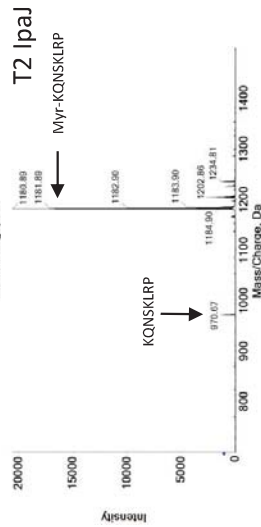
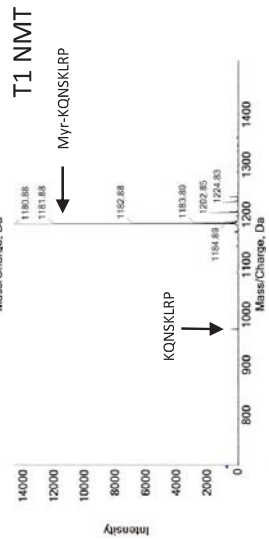
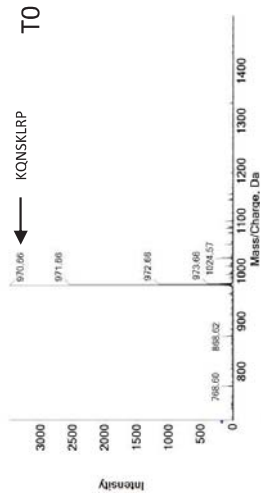
MSMS T2 Myr-GK(Myr)QNSKLR (1350)



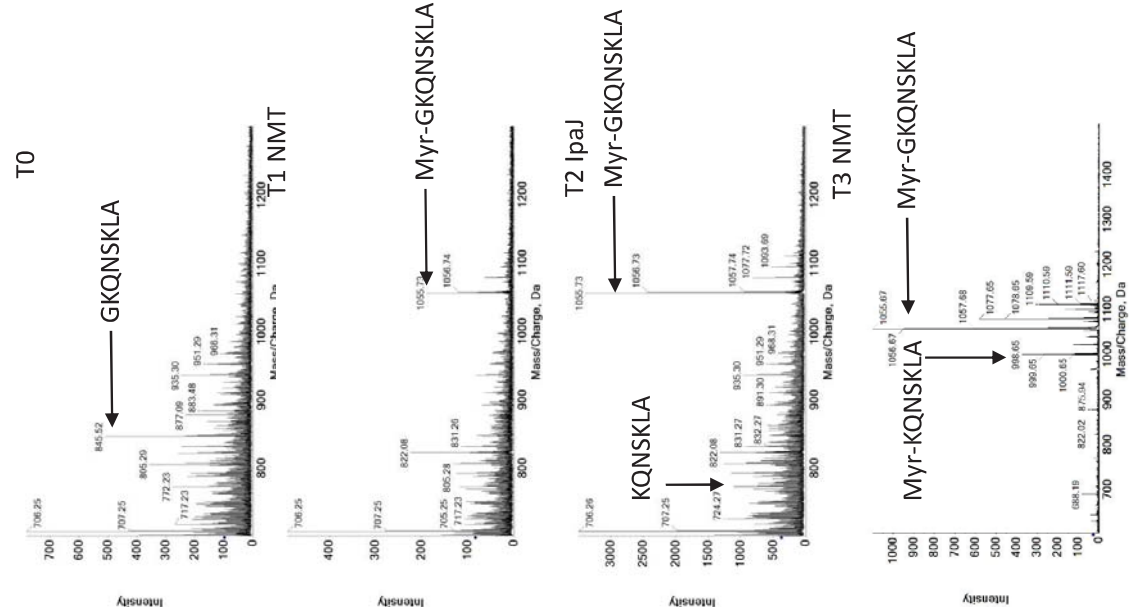
Myr-GK(Myr)QNSKLR



KQNSKLRP

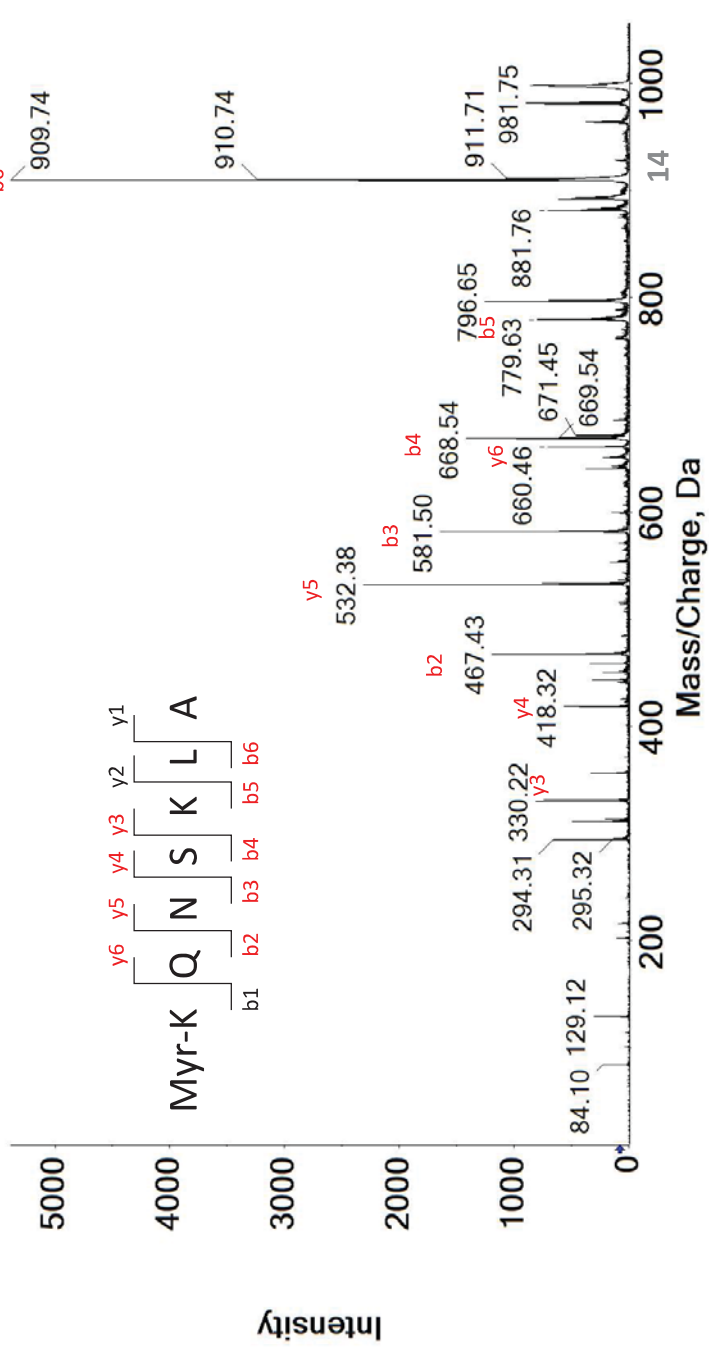


GKQNSKLA

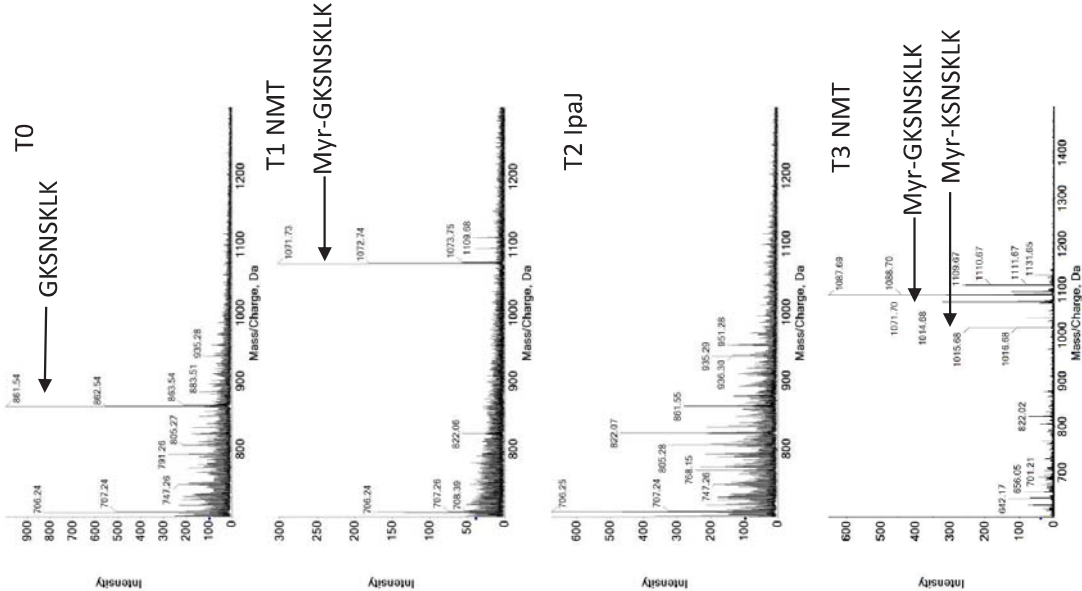


F10_MSMS_998.6400_3		ion
pic	Mass/Charge (Da)	Intensity
	2155	331,2692
	3523	418,3104
	4794	467,4386
	6248	532,3957
	7446	581,477
	9310	660,4426
	9508	668,5073
	12015	779,6317
	14810	909,7243

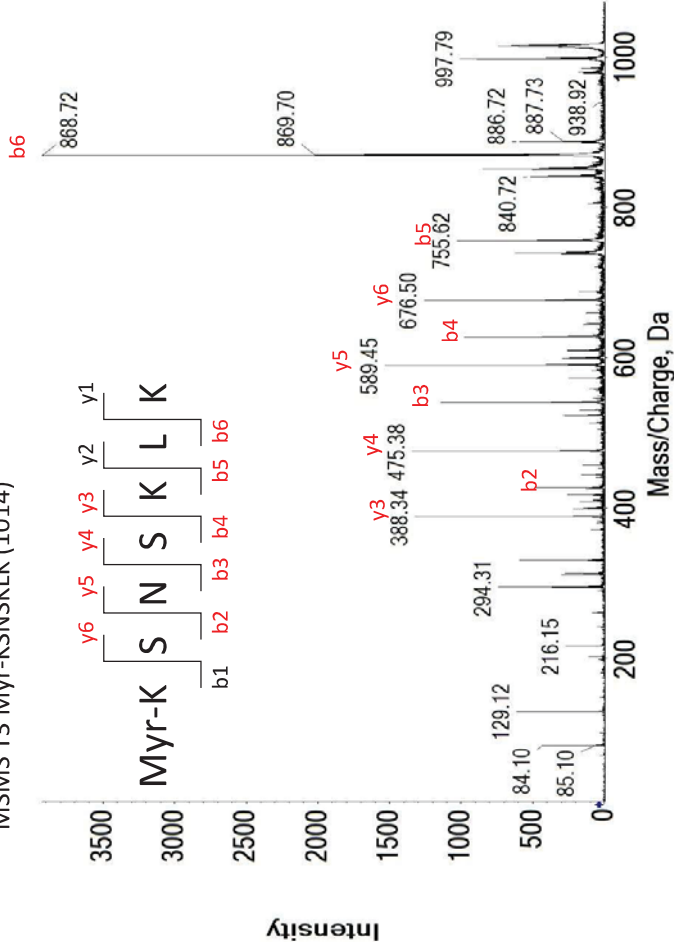
MSMS T2 Myr-KQNSKLA (998)



GKSNSKCLK

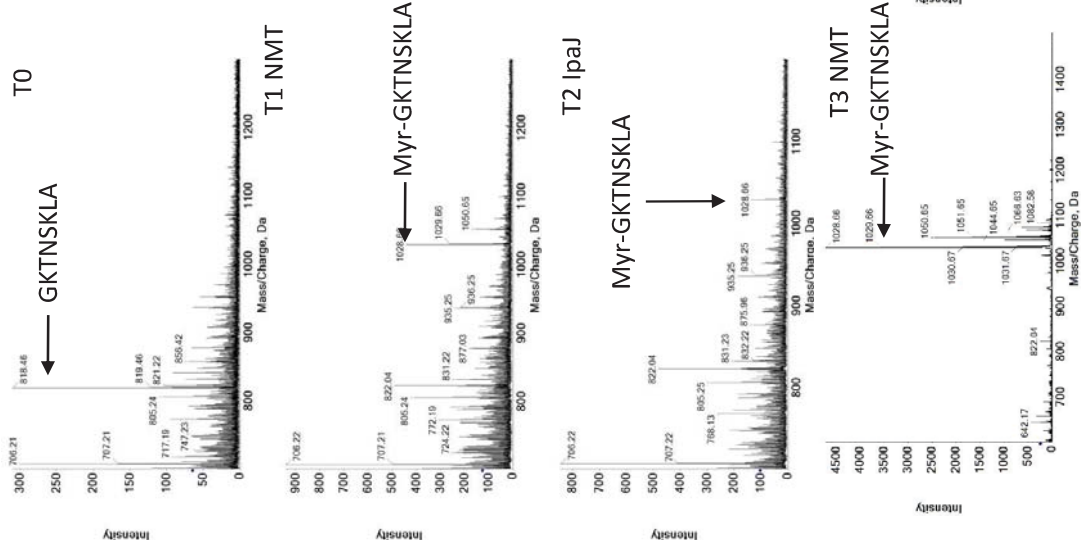


MSMS T3 Myr-KSNSKCLK (1014)

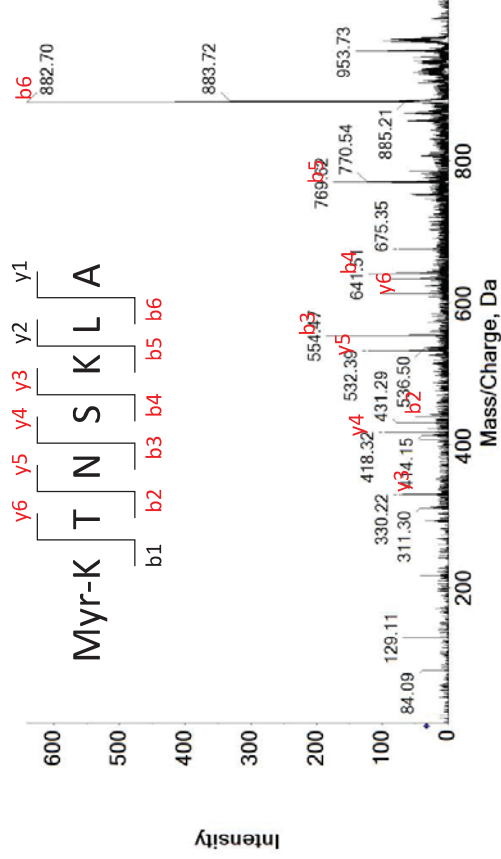


F9_MSMS_1014.6000_2	Mass/Charge (Da)		Intensity	ion
pic				
4036	388,3268	1320,784	y3	
5032	426,3984	484,2353	b2	
6101	475,3693	1344,157	y4	
7622	540,4544	1142,118	b3	
8831	589,4373	1535,059	y5	
9779	627,5093	978,5098	b4	
10957	676,5035	1258,196	y6	
12771	755,6361	1029,804	b5	
15221	868,7002	3926,588	b6	

GKTN SKLA

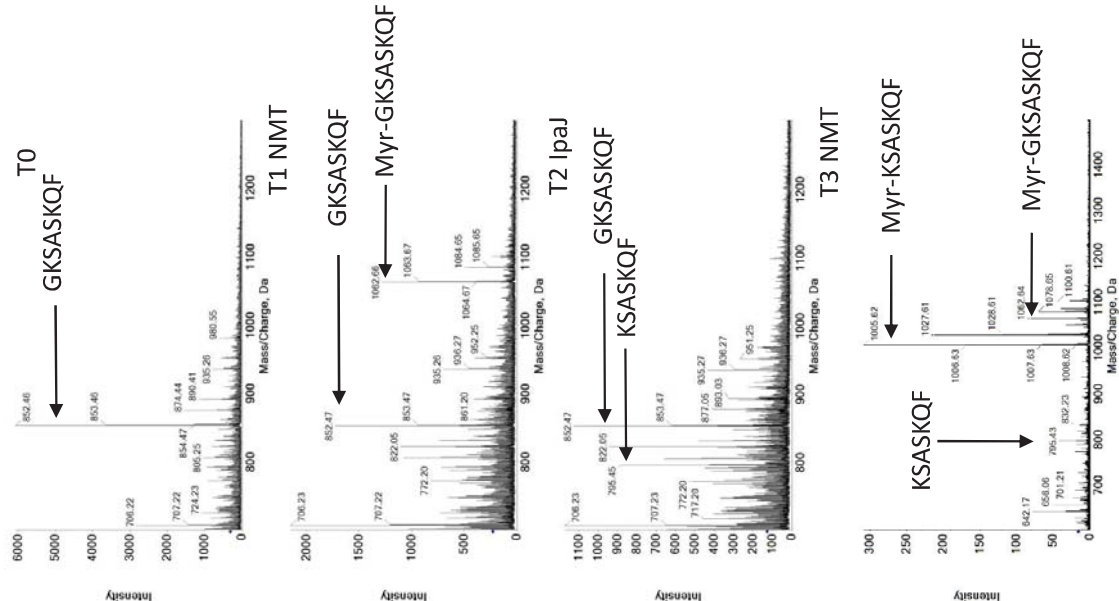


MSMS T3 Myr-KTNSKLA (971)

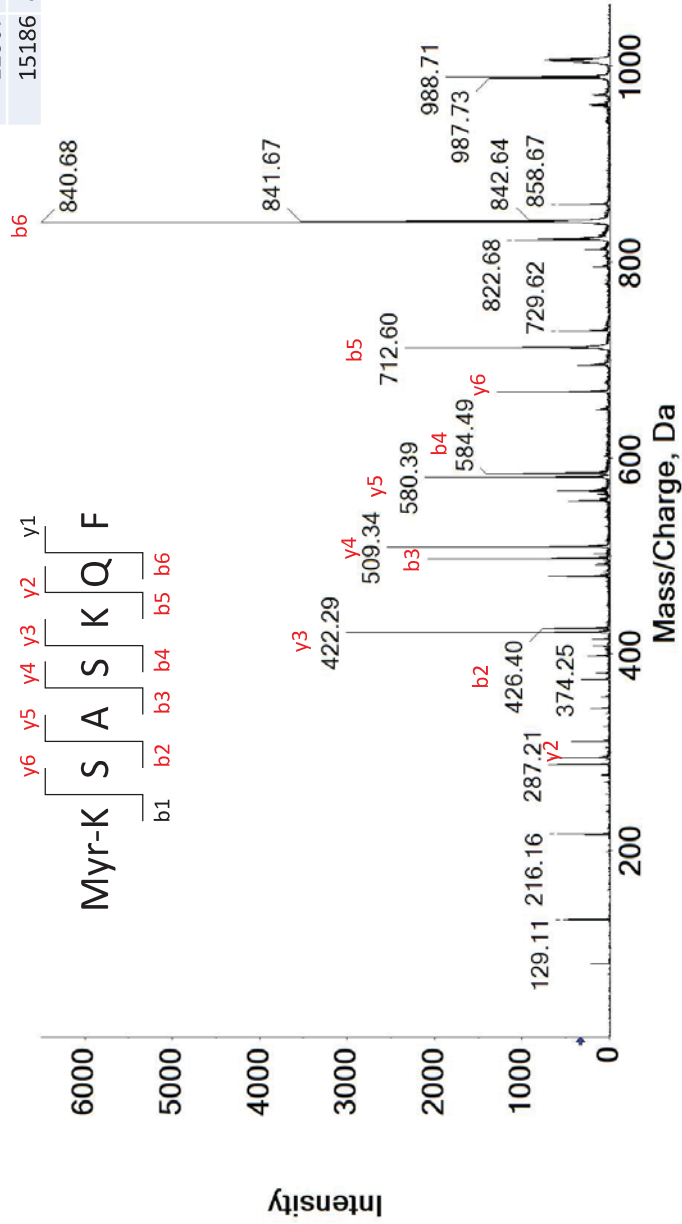


F7_MSMS_971.6000_2		ion
pic	Mass/Charge (Da)	Intensity
1835	331,2539	69,4902 y3
2576	418,3141	97,09805 y4
2865	440,4206	49,56863 b2
3905	532,3945	121,8824 y5
4239	554,4536	185,5686 b3
5292	633,4332	87,84314 y6
5428	641,4906	121,4118 b4
7520	769,2246	62,27451 b5
9452	882,458	73,56863 b6

GKSASKQF

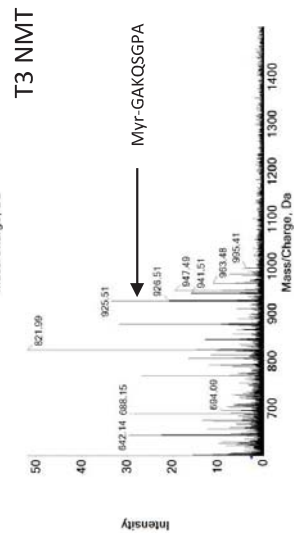
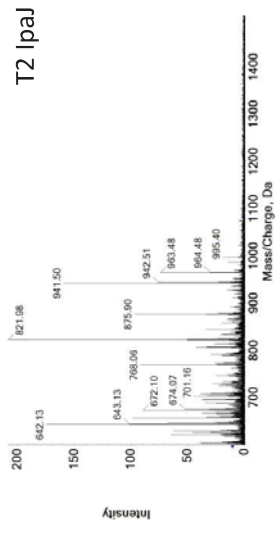
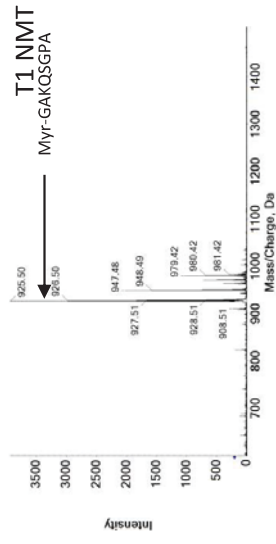
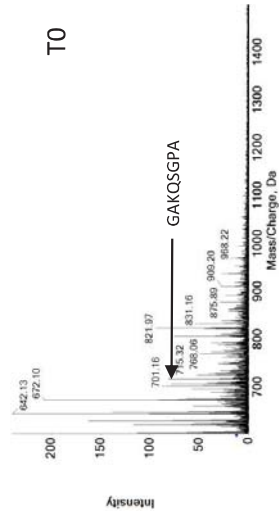


MSMS T3 Myr-KSASKQF (1005)



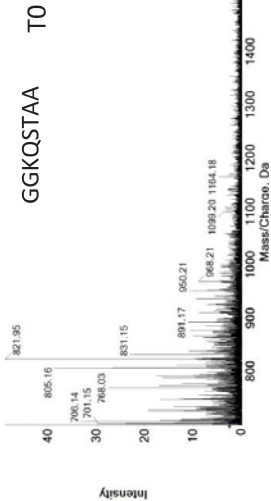
F8_MSMS_1005.6000_3		ion
pic	Mass/Charge (Da)	Intensity
2238	294,1691	387,451 y2
5108	422,2837	3010,667 y3
5235	426,3989	758,902 b2
6967	497,4295	2081,255 b3
7300	509,3415	2547,294 y4
9172	580,4089	2050,039 y5
9279	584,4832	1414,902 b4
11315	667,4761	958,5883 y6
12367	712,6038	2342,745 b5
15186	840,6764	6497,726 b6

GAQSGPA



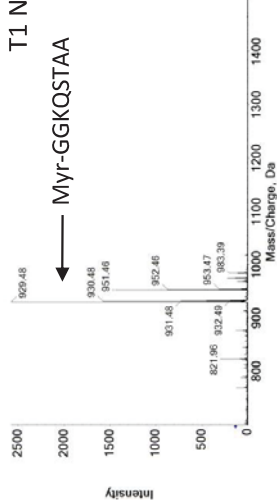
GGKQSTAA

GGKQSTAA T0



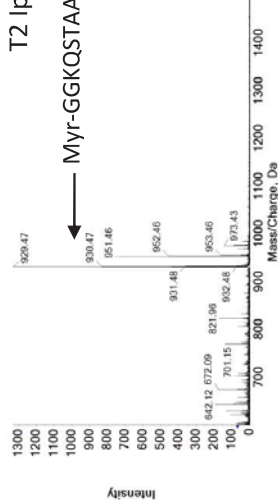
T1 NMT

Myr-GGKQSTAA



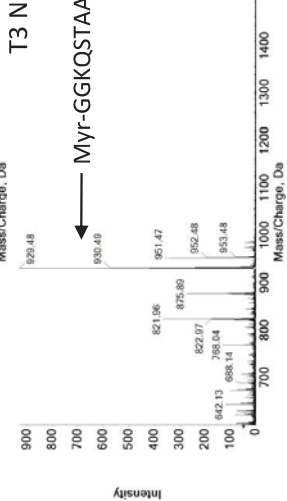
T2 IpaJ

Myr-GGKQSTAA



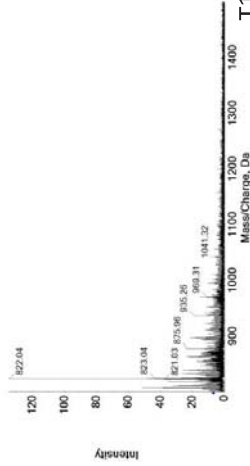
T3 NMT

Myr-GGKQSTAA

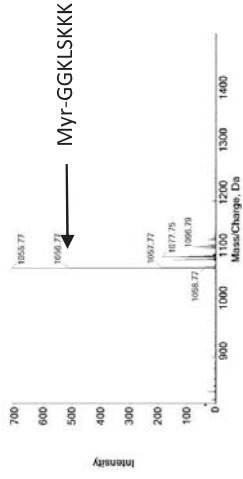


GGKLSKKK

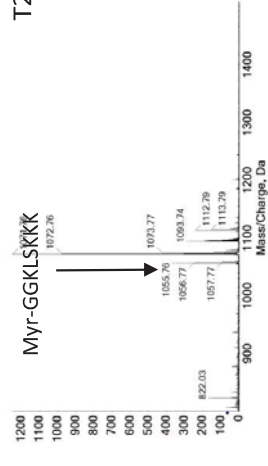
T0



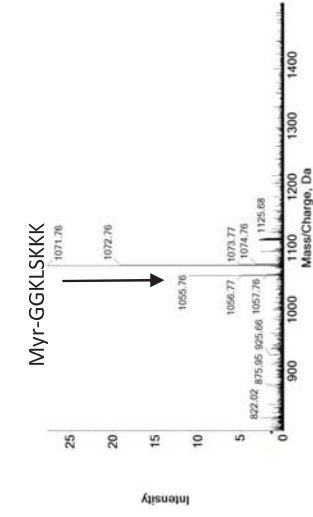
T1 NMT



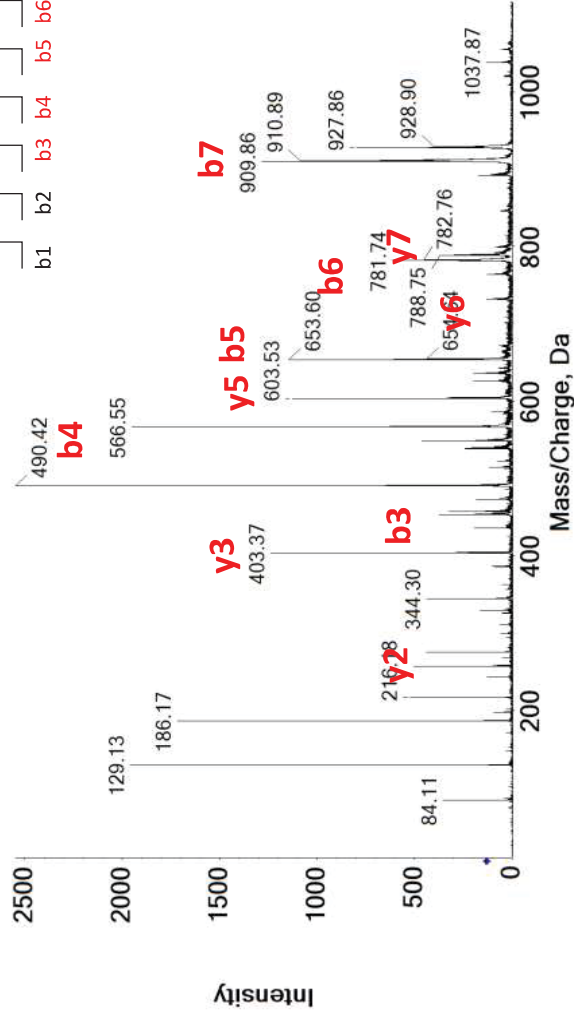
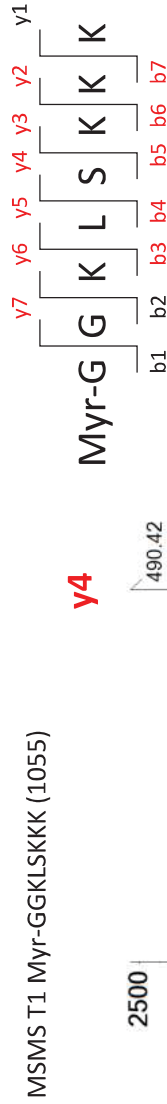
T2 IpaJ



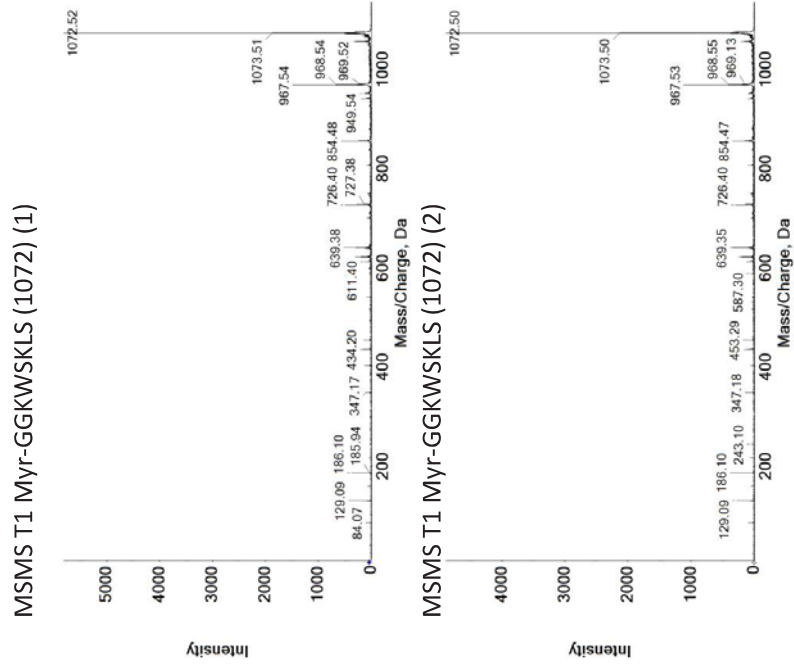
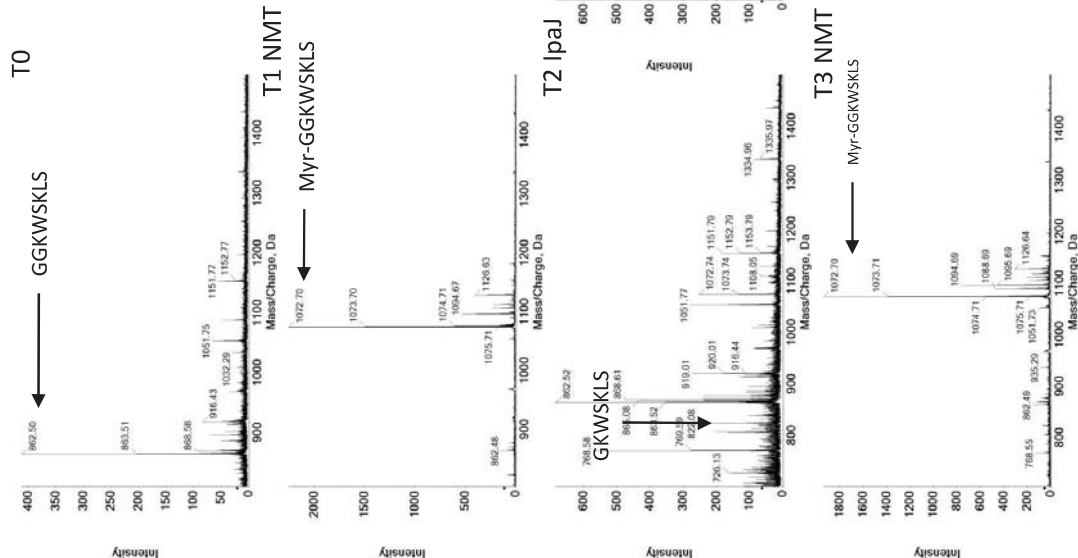
T3 NMT



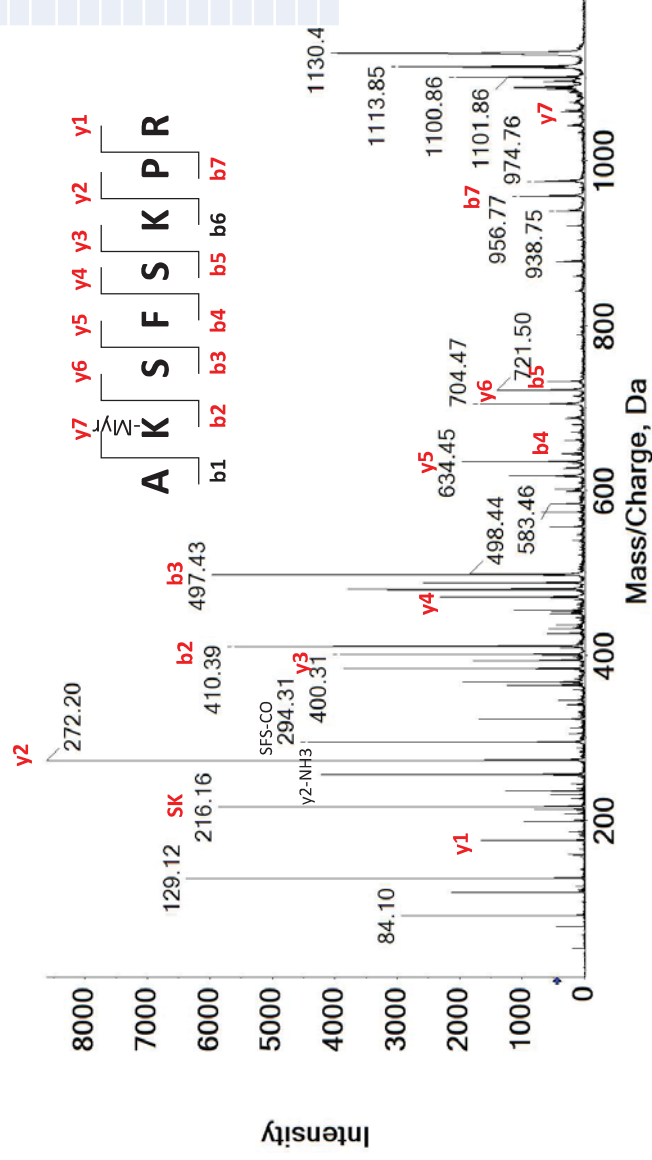
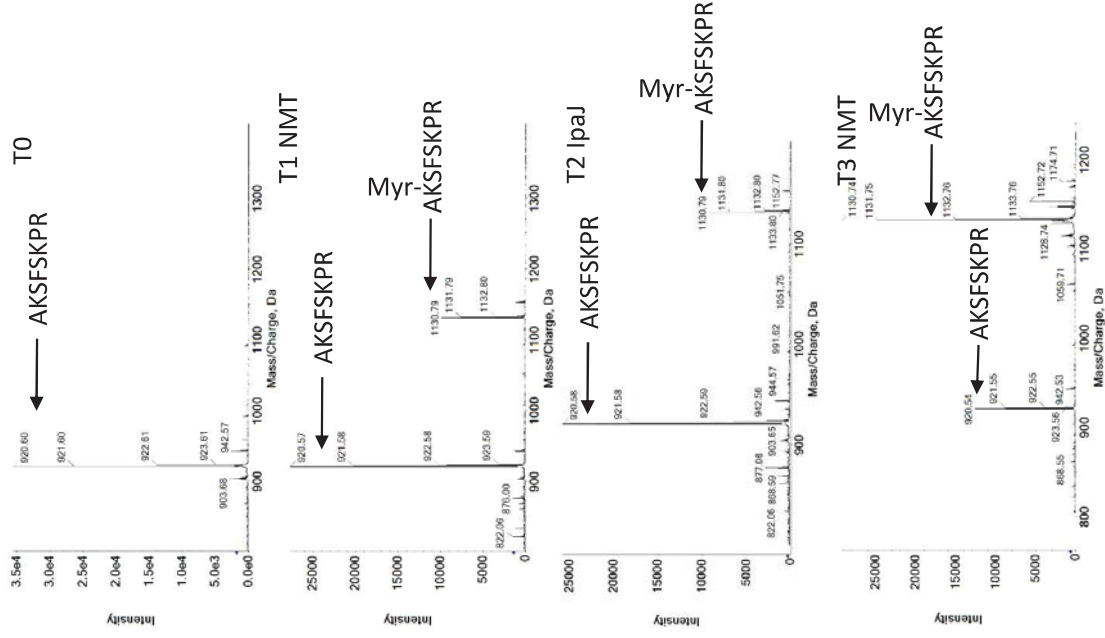
y2	2642	275,2397	400,9412
y3	5363	403,3621	1237,961
b3	6526	453,4303	375,8431
y4	7591	490,8696	179,6078
b4	9598	566,5328	1952,314
y5	10561	603,5288	1129,098
b5	11824	653,5715	1144,941
y6	13686	731,6654	131,2941
b6	14812	781,6852	615,8431
y7	14969	788,6453	359,6863
b7	17555	909,8929	1265,882



GGKWSKLS



AKFSKPR



766	84,09495	2935.84 K
1129	112,1094	2126.90 R
2056	129,1259	6378.35 K
3482	175,1427	1653.96 y1
SK-H2O/		
4180	198,1471	973.80 KP-CO
4900	216,147	5869.18 SK
5668	235,1309	1267.45 FS
6392	255,174	4213.96 v2-NH3
7039	272,2352	6608.47 v2
7846	294,3016	4431.22 SF5-CO
8846	322,1745	1693.18 SF5
9640	345,2368	4114.43 FK-H2O
10247	363,2515	1236.71 FSK
10945	383,2669	5877.57 v3-NH3
11231	393,3545	1791.69 v2-NH3
11453	400,3001	3915.61 v3
K(+210)S-		
11709	408,3862	384.78 H2O
11770	410,3888	5604.39 b2
12272	426,3936	597.18 K(+210)S
12996	450,2837	557.65 SF5K
13590	470,3228	2309.80 v4-NH3
13855	479,4061	3797.80 b3-H ₂ O
13884	480,4055	1171.63-NH3
14085	487,3613	2589.18 v4
14373	497,1166	5667.98 b3
16462	573,4848	695.96 K(+210)SF
17604	617,4413	1208.78 v5-NH3
18035	634,4463	1664.86 v5
18287	644,498	369.88 b4
19750	704,4481	1667.14 v6-NH3
20154	721,4829	1395.76 v6
20390	731,5301	590.59 b5
25295	956,7646	1045.80 b7
25660	974,7401	803.45 b7+H2O
27339	1059,846	735.53 v7

[illegible]

Mass spectrum of Myr-K peptide. The x-axis represents the mass-to-charge ratio (m/z) from 0 to 2300, and the y-axis represents relative intensity from 0 to 1000. The spectrum shows numerous peaks, with several labeled in red and black. A fragmentation map at the top indicates the positions of b and y ions.

Fragmentation map (top):

- Myr-K S W S K G R
- b1 b2 b3 b4 b5 b6
- y1 y2 y3 y4 y5 y6

Mass spectrum peaks (m/z):

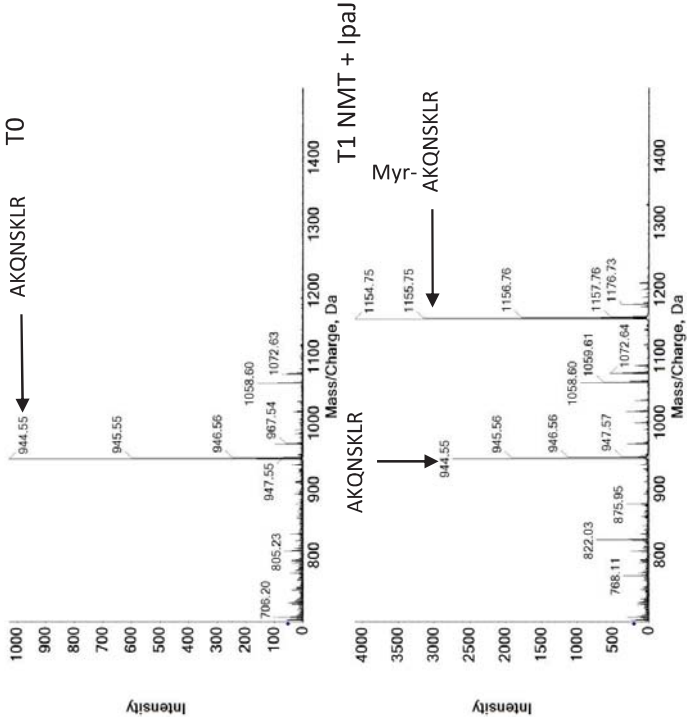
- 129.13
- 84.10
- 216.17
- 232.18
- 343.26
- 430.31
- 426.40
- 447.33
- 595.51
- 612.52
- 616.42
- 703.49
- 720.49
- 848.58
- 884.72
- 901.78
- 902.71
- 1028.80
- 1041.77

Peak labels (red and black):

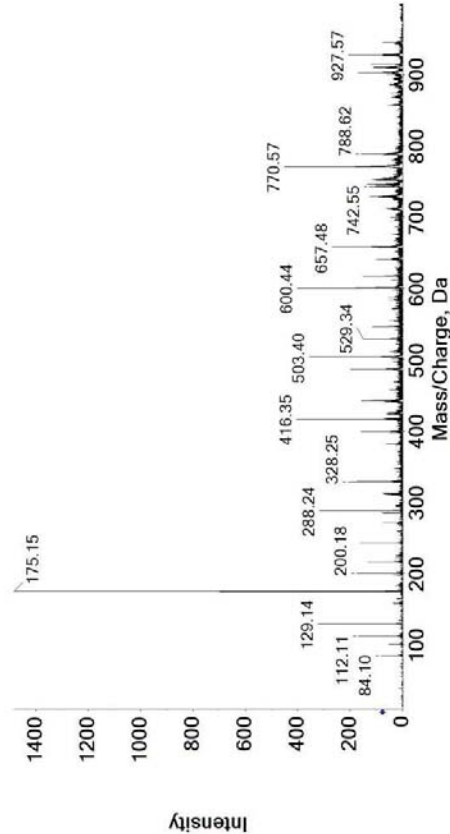
- y2
- b2
- y3
- b3
- y4
- b4
- y5
- b5
- y6
- b6



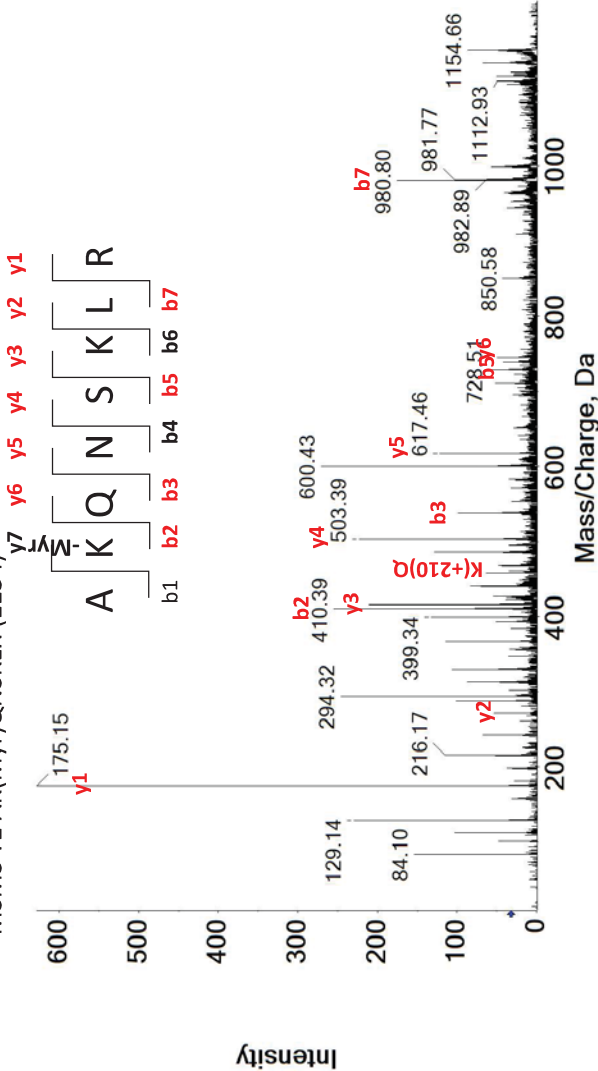
AKQNSKLR



MSMS TO AKQNSKLR (944)



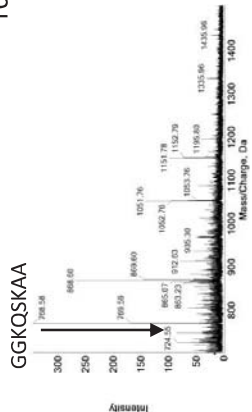
MSMS T1 AK(Myr)QNSKLR (1154)



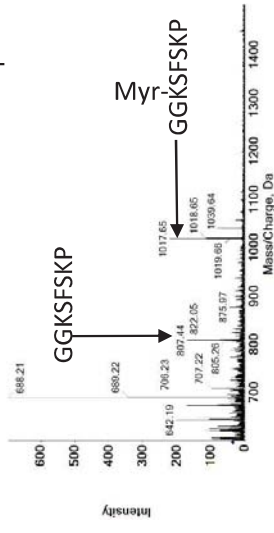
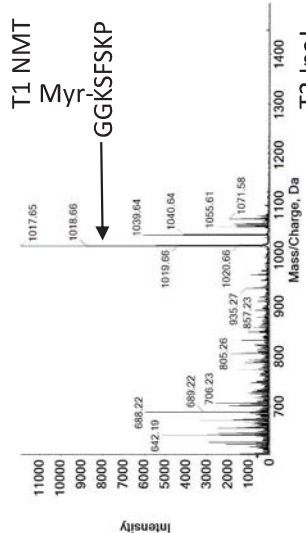
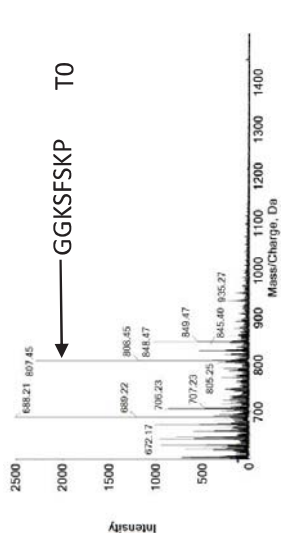
Mass/Ch	arge (Da)	Intensity
K24 1154	535	175,0203
	535	35,45098
	1231	288,1639
	1231	31,52941
	2914	410,3834
	2914	254,7451
	3049	416,3236
	3049	210,8235
	3905	467,378
	3905	43,60785
	4432	503,3777
	4432	223,3726
	4875	538,4333
	4875	98,98039
	5859	600,4333
	5859	249,8824
	6146	617,4308
	6146	121,4118
	8017	739,5452
	8017	42,5098
	8118	745,4847
	8118	49,09804
	9226	850,5908
	9226	42,5098
	10884	980,7767
	10884	175,5294
	12877	1137,767
	12877	57,45098

GGKQSKAA

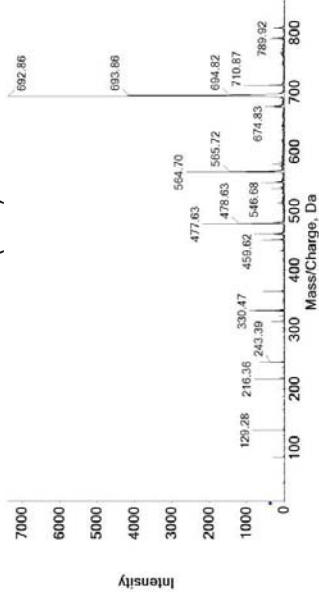
T0



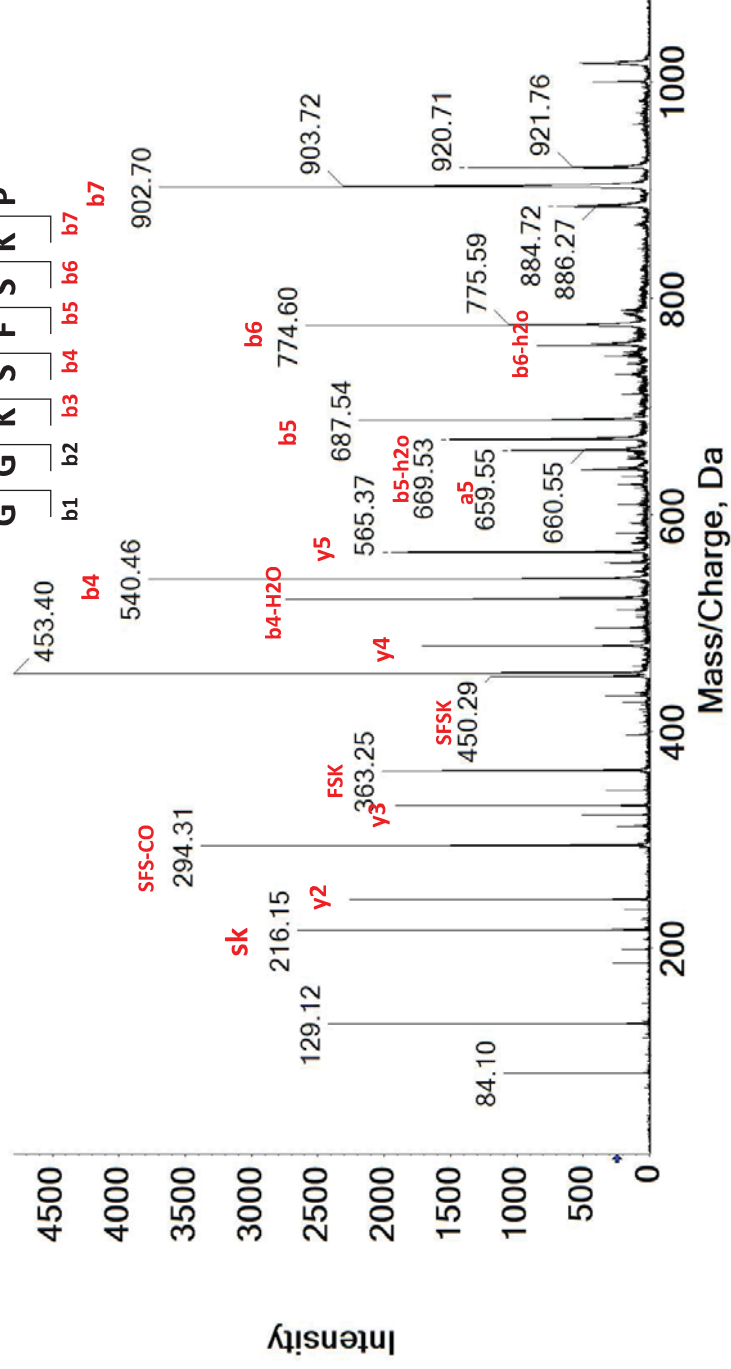
GGKSFSKP



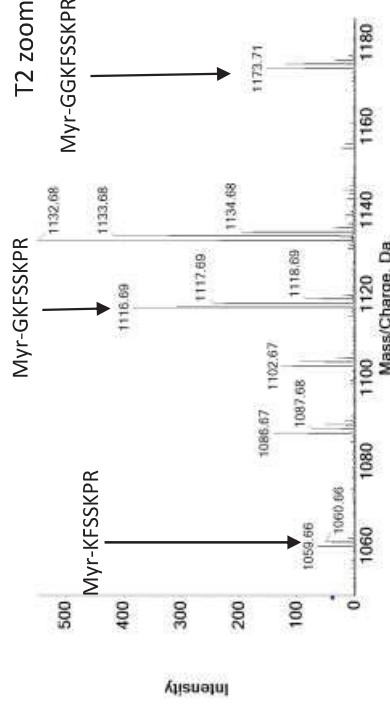
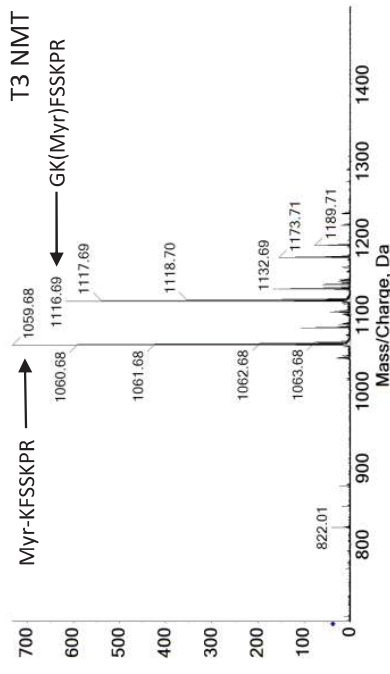
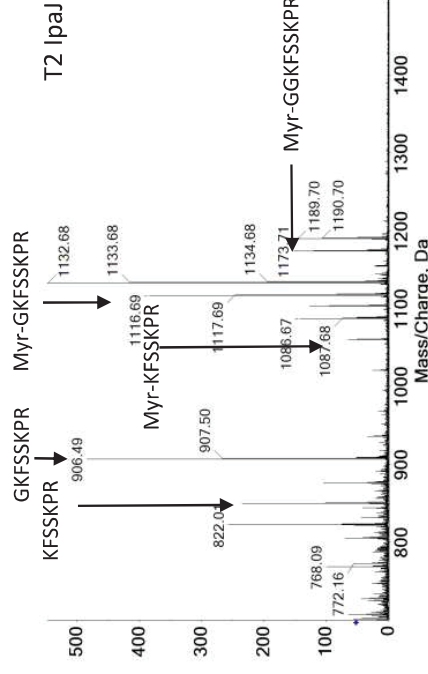
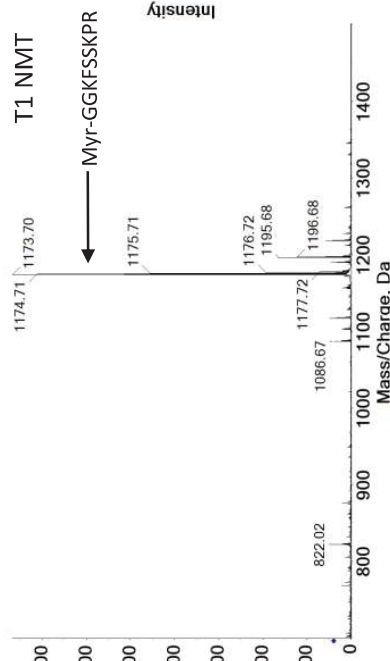
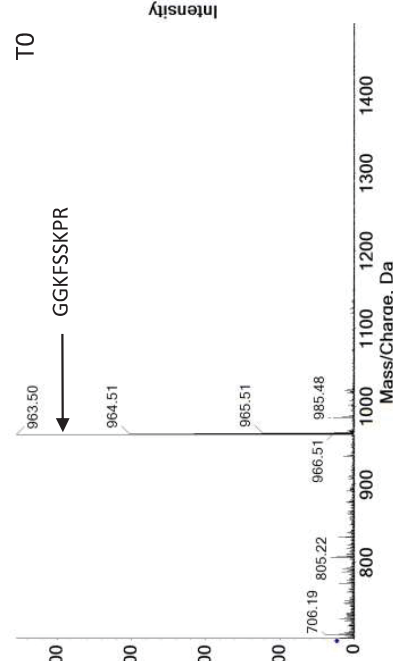
MSMS TO GGKSFSKP (807)



MSMS T1 GGK(Myr)SFSKP (1017)

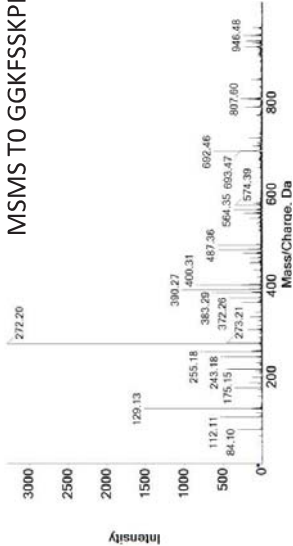


GGKFSSKPR

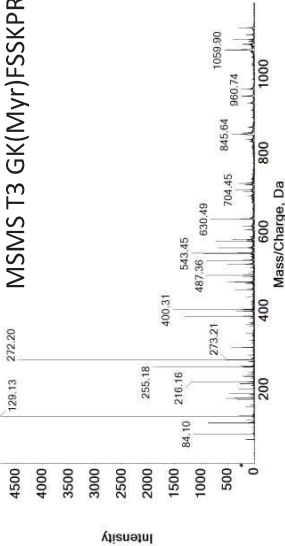


GGKFSSKPR

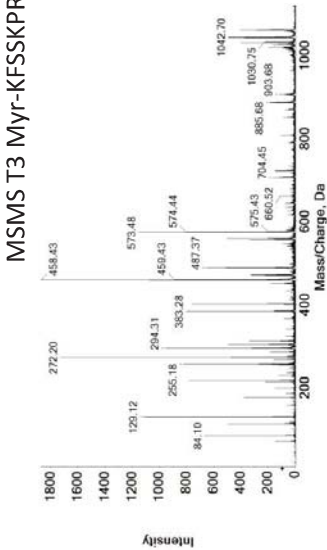
MSMS TO GGKFSSKPR (963)



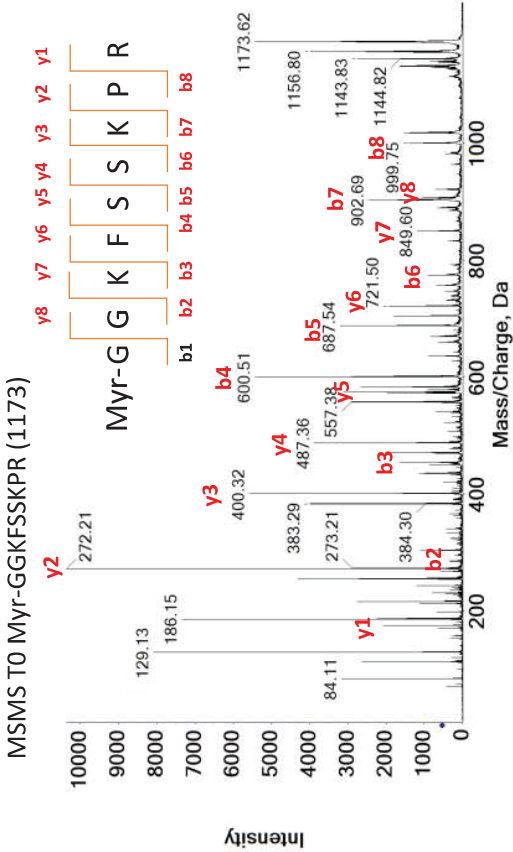
MSMS T3 GK(Myr)FSSKPR (1116)



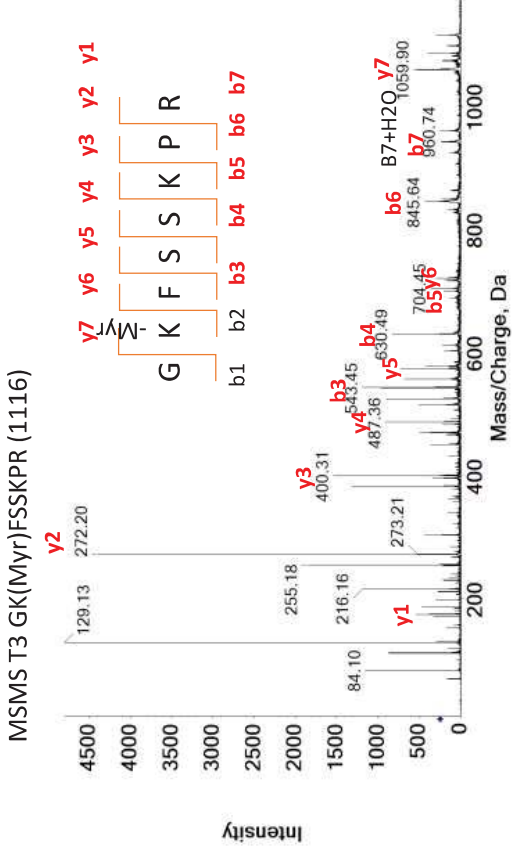
MSMS T3 Myr-KFSSKPR (1059)



MSMS TO Myr-GGKFSSKPR (1173)



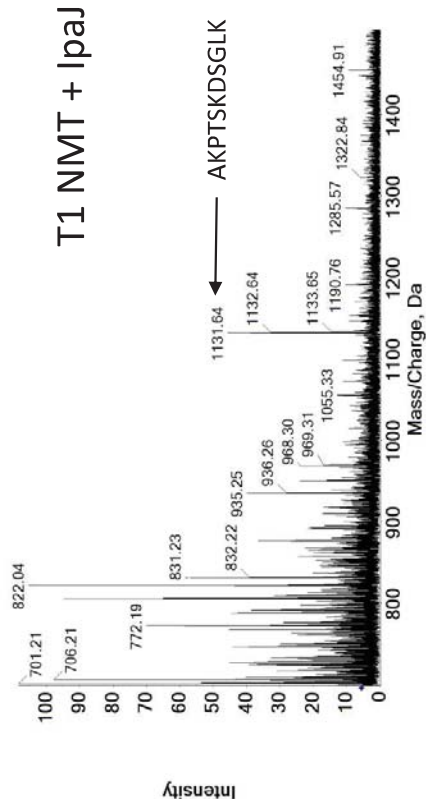
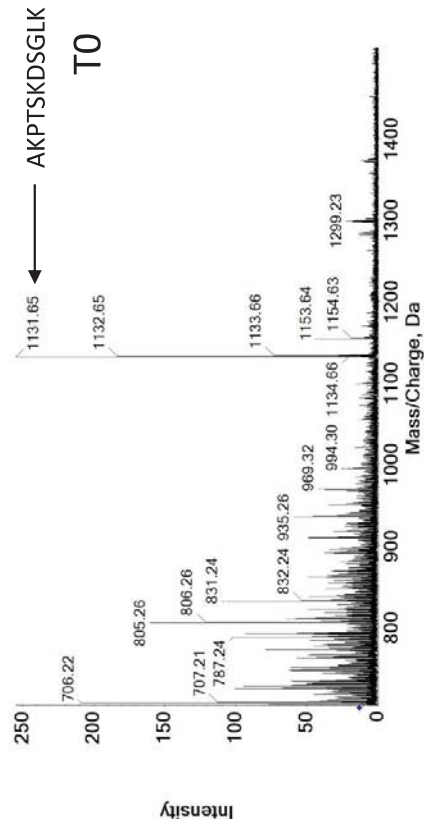
MSMS T3 GK(Myr)FSSKPR (1116)



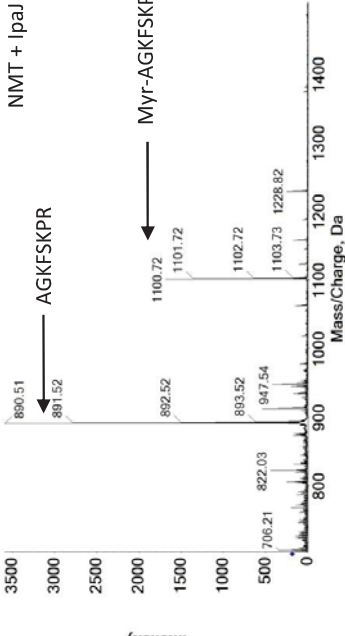
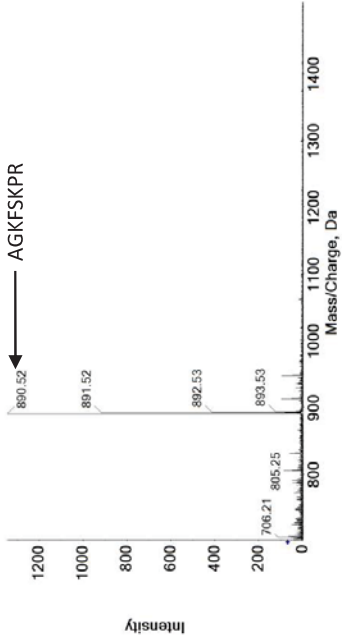
K20 1173	Mass/Charge (Da)	Intensity
4372	175,1503	2072 y1
8355	272,1954	10360,16 y2
10293	325,2772	320 b2
12791	400,3083	5570,51 y3
14427	453,4173	1645,961 b3
15423	487,3767	3883,765 y4
17825	574,4145	3119,059 y5
18508	600,4936	5425,726 b4
20684	687,5204	3206,745 b5
21495	721,4933	1923,922 y6
22724	774,5721	912,7844 b6
24390	849,6013	1189,02 y7
25523	902,6548	2460,706 b7
25606	906,606	1046,588 y8
27511	999,7244	1402,039 b8

K22 1116	Mass/Charge (Da)	Intensity
2008	175,1541	538,35 y1
2891	216,1627	1182,90 SK
3999	255,1691	1852,24 y2-NH3
4553	272,2057	4394,67 y2
5462	303,1855	437,33 SSK
7699	383,2735	1310,43 SSKP-NH3
8218	400,3085	1541,02 y3
9657	450,2645	359,53 FSSK
10244	470,3054	500,24 y4-NH3
10729	487,3455	905,41 y4
11494	515,4305	506,04 a3
12266	543,4355	1191,69 b3
13097	574,384	724,55 y5
14532	630,4798	822,75 b4
16634	717,5125	190,59 b5
16727	721,4527	298,67 y6
19445	845,6232	432,31 b6
21453	942,7505	231,69 b7
21812	960,7414	246,90 b7+H2O
23625	1059,864	569,10 y7

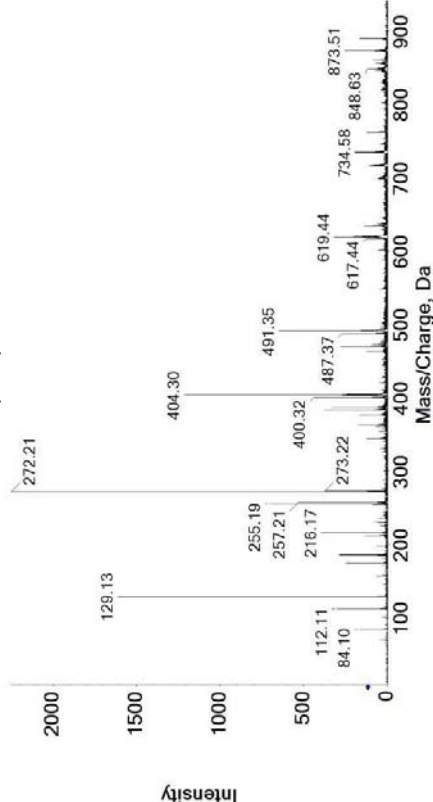
AKPTSKDSGLK



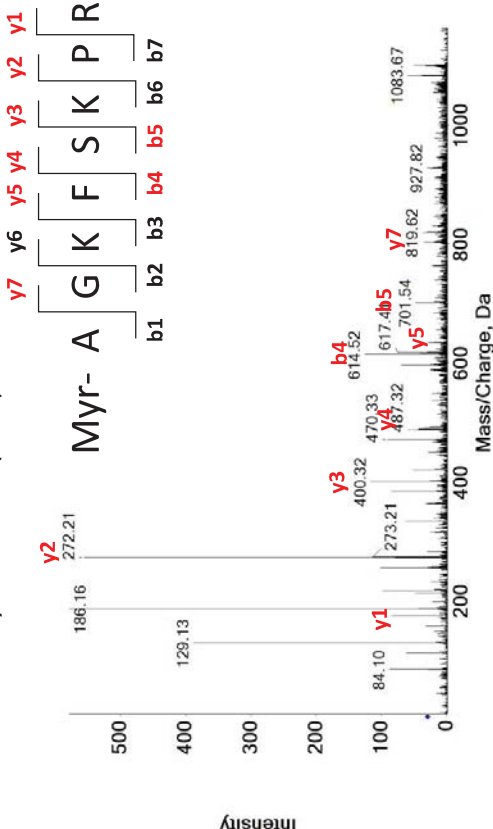
AGKFSKPR



MSMS TO AGKFSKPR (890)

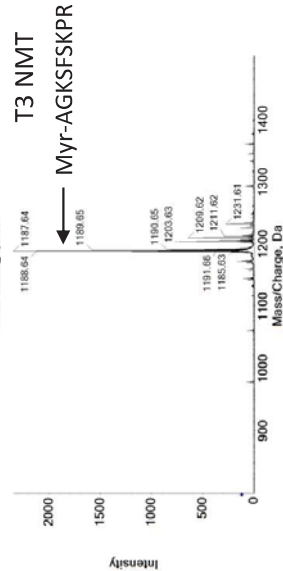
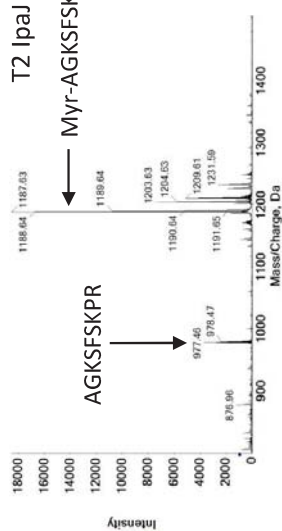
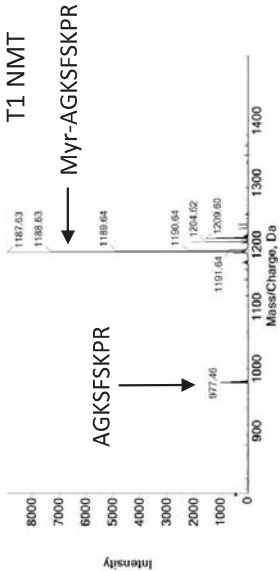
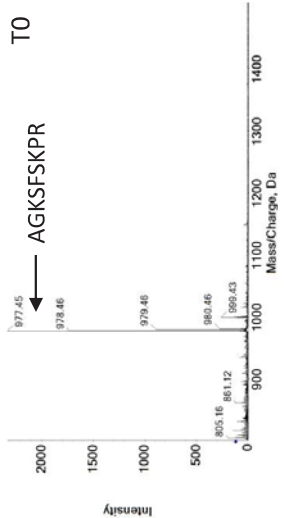


MSMS T1 Myr-AGKFSKPR (1100)



L4 1100	Mass/Charge (Da)	Intensity
542	129.138	388,8628 K
801	175,1467	84,86275 y1
894	186,1513	578,353 GK
1149	216,181	98,66667 SK
1590	272,2057	555,9216 y2
2024	333,2545	63,21569 GKF
2715	400,3234	116,8627 y3
2896	420,3111	51,45098 GKFS
3333	470,3304	90,82353 y4-NH3
3503	487,3036	52,07843 y4
4591	614,5147	126,1177 b4
4836	634,4205	28,07843 y5
5421	701,5098	47,68628 b5
6483	819,5577	37,17647 y7

AGKFSFSKPR

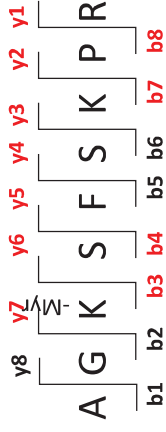
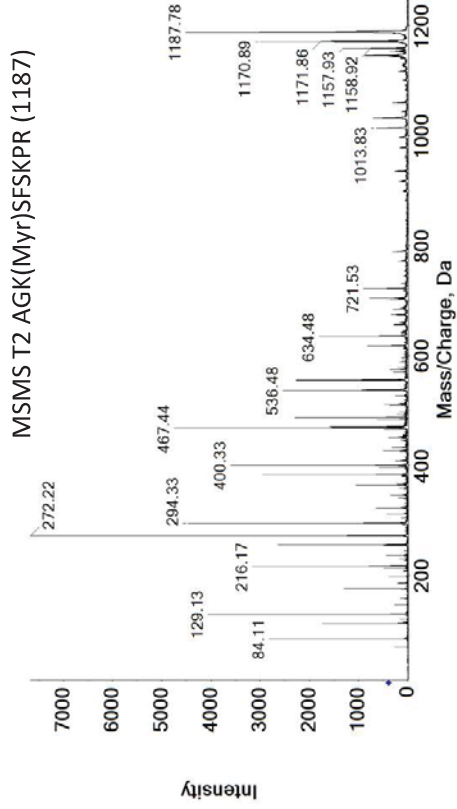
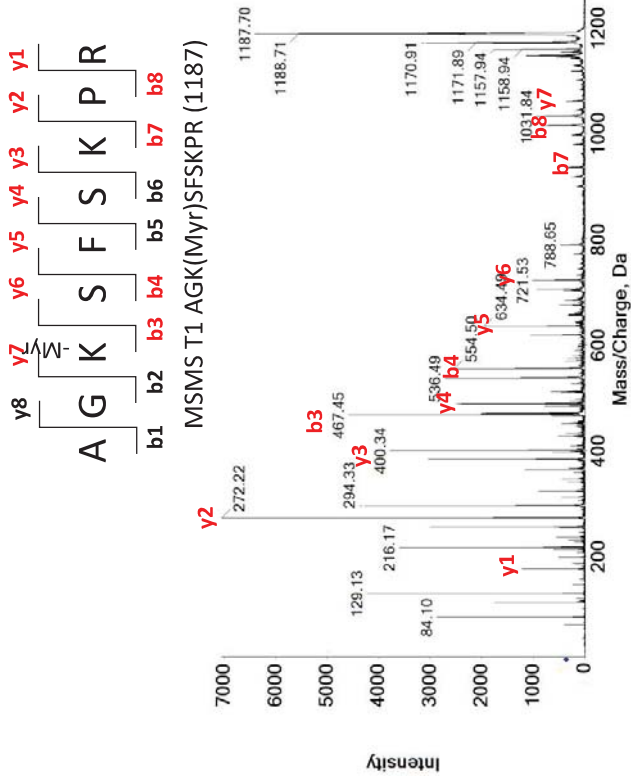


L12

L13

L14

L15



Titre : Extension de l'identification des protéines myristoylées chez les Eucaryotes

Mots clés : Acylation ; Myristoylation ; N-myristoyltransferase ; N-terminal ; Lysine ; Protéomique

Résumé : La N-myristoylation correspond à l'ajout d'un groupement myristate - un acide gras saturé de 14 carbones - sur une glycine N-terminale d'une protéine cible. Cette modification essentielle des Eucaryotes est catalysée en grande partie par la N-myristoyltransférase (NMT) et permet aux protéines ciblées d'être orientées vers des membranes spécifiques. Ainsi localisées, les protéines, via l'interaction avec d'autres partenaires, peuvent contribuer activement au fonctionnement de voies de signalisation indispensables à la cellule. La dérégulation de ce système est associée à de nombreuses pathologies humaines. Cependant, les cibles de la modification ne sont généralement pas identifiées dans les analyses protéomiques classiques du fait de limitations techniques inhérentes aux approches disponibles. Il en résulte une perte d'information importante dans la compréhension de mécanismes fondamentaux du fonctionnement cellulaire normal et pathologique. Dans ce contexte, l'objectif de ma thèse a été de développer un nouvel outil biotechnologique permettant d'approfondir et

combler ce manque d'information concernant la myristoylation. Ainsi, j'ai cloné et purifié la protéase IpaJ de la bactérie pathogène *Shigella flexneri*. IpaJ est capable de retirer la glycine N-terminale myristoylée. J'ai démontré la très grande spécificité d'IpaJ pour ces deux seuls groupes grâce à un test en spectrométrie de masse de suivi d'activité que j'ai mis au point. Ma caractérisation de l'activité d'IpaJ indique qu'elle est bien adaptée à l'étude du myristoylome, l'ensemble des protéines myristoylées d'un protéome. J'ai alors introduit l'utilisation d'IpaJ dans une étape d'analyse protéomique de type « shotgun » et démontré sa pertinence pour l'étude du myristoylome. J'ai montré dans une analyse parallèle que la NMT peut réaliser aussi la myristoylation de lysines, ce qui était inconnu jusqu'alors. J'ai utilisé IpaJ pour distinguer les deux modifications lorsqu'elles sont en compétition sur la même protéine. Globalement, mon étude démontre qu'IpaJ constitue un nouvel outil de valeur pour l'étude de la myristoylation des glycines et des lysines.

Title : Extension of the identification of N-acylated proteins in Eukaryotes

Keywords : Acylation ; Myristoylation ; N-myristoyltransferase ; N-terminal ; Lysine ; Proteomics

Abstract : N-myristoylation is the addition of a myristate group - a 14-carbon saturated fatty acid - on an N-terminal glycine of a target protein. This essential modification of eukaryotes is largely catalyzed by N-myristoyltransferase (NMT) and allows targeted proteins to be directed to specific membranes. Thus located, proteins, through the interaction with other partners, can actively contribute to the functioning of signaling pathways essential to the cell. Deregulation of this system is associated with many human pathologies. However, the targets of the modification are generally not identified in classical proteomic analyzes due to technical limitations inherent in the available approaches. This results in a loss of important information in the understanding of fundamental mechanisms of normal and pathological cell function. In this context, the objective of my thesis was to develop a new biotechnological tool allowing to deepen and fill this lack of information concerning myristoylation.

Thus, I cloned and purified the IpaJ protease from the pathogenic bacterium *Shigella flexneri*. IpaJ is able to remove the N-terminal myristoylated glycine. I have demonstrated the very high specificity of IpaJ for these two groups alone through an activity monitoring mass spectrometry test that I have developed. My characterization of the activity of IpaJ indicates that it is well suited to the study of myristoylome, the set of myristoylated proteins in a proteome. I then introduced the use of IpaJ in a shotgun-type proteomic analysis step and demonstrated its relevance for the study of myristoylome. I showed in a parallel analysis that NMT can also perform lysine myristoylation, which was not previously known. I used IpaJ to distinguish the two modifications when they compete on the same protein. Overall, my study demonstrates that IpaJ is a valuable new tool for the study of myristoylation of glycines and lysines.