



HAL
open science

Nouvelles méthodologies pour la synthèse totale de protéines

Marine Cargoët

► **To cite this version:**

Marine Cargoët. Nouvelles méthodologies pour la synthèse totale de protéines. Médecine humaine et pathologie. Université du Droit et de la Santé - Lille II, 2017. Français. NNT : 2017LIL2S025 . tel-02441694

HAL Id: tel-02441694

<https://theses.hal.science/tel-02441694v1>

Submitted on 16 Jan 2020

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

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

Université de Lille – Droit et Santé
École Doctorale Biologie – Santé de Lille (EDBSL)

THÈSE

Pour obtenir le titre de
Docteur de l'Université de Lille

Spécialité : Chimie organique

Présentée par

Marine CARGOËT

<p>NOUVELLES MÉTHODOLOGIES POUR LA SYNTHÈSE TOTALE DE PROTÉINES</p>
--

Soutenue le 13 octobre 2017

Devant le jury composé de :

Président :	Pr. Anthony ROMIEU
Rapporteur :	Dr. Florine CAVELIER
Examineur :	Dr. Stéphanie DEVILLE-FOILLARD
Directeur de thèse :	Dr. Oleg MELNYK

Institut de Biologie de Lille, CNRS UMR 8161
Institut Pasteur de Lille
1 Rue du Professeur Calmette 59019 Lille Cedex France

REMERCIEMENTS

Je tiens à remercier l'ensemble des membres de mon jury qui ont accepté de juger ce travail. Je remercie le Pr. Annemieke Madder et le Dr. Florine Cavelier d'avoir accepté d'être mes rapporteurs. Merci au Pr. Anthony Romieu d'avoir accepté de présider mon jury et merci au Dr. Stéphanie Deville-Foillard d'être mon examinateur.

Je remercie le Pr. Yvan de Launoit de m'avoir accueillie au sein de l'UMR 8161.

Je remercie tout particulièrement le Dr. Oleg Melnyk pour m'avoir encadré durant ma thèse. Merci de m'avoir guidé durant ces trois années et de m'avoir poussé toujours plus loin pour donner le meilleur de moi-même. Je vous remercie également pour votre disponibilité, vos conseils et le partage de vos connaissances scientifiques.

J'aimerais également remercier le Dr. Jérôme Vicogne pour toute l'aide qu'il m'a apportée en biologie du tout début jusqu'à la fin de ma thèse. Merci pour tes conseils, ta patience et ton soutien.

Je remercie également toutes les personnes qui m'ont apporté une aide précieuse dans mes travaux de recherches : Hervé (l'ourson) le maître des machines analytiques et le roi du mot fléchés ; Annick avec qui j'ai partagé de nombreuses joies et déceptions concernant « notre » robot ; Julien & Reda pour les précieux conseils scientifiques mais surtout les pauses déjeuners.

Je remercie ensuite Nathalie (BZH power), Magalie, Vangelis, Rémi, Vincent et Jennifer (bon courage avec SUMO !) pour leurs conseils, leur aide, leur soutien et leur écoute. Mais également la team biologie Alex, Claire, Bérénice, Lu et Roland pour leur aide et leur disponibilité.

Je souhaite également remercier Emmanuelle Boll. Tu as été un soutien sans faille tant au niveau professionnel que personnel durant ces trois années. Depuis ce tout premier jour où tu m'as accueilli dans ton bureau jusqu'à la fin où tu as accueilli mes amis (une grande épreuve ^^). Nous partageons ce sale caractère et cette détermination et nous en sommes fières !

Je remercie la team BioAddoct and co : Clément x 2, Cindy et Juju. Un merci tout particulier à Mickaël pour ces beaux graphiques statistiques et les gâteaux de l'Impertinente livrés à « domicile ».

Enfin je terminerai par remercier mes proches qui ont su me soutenir durant ces trois années et bien plus : mes parents, ma sœur, Lucie (Mlbs), Florent (3B forever), mes coloc de toujours Sylvain et Marine (saucisse, saucisse ♪).

TABLE DES MATIÈRES

Remerciements	3
Table des matières	5
Liste des abréviations	9
Résumé.....	13
Abstract	15
ÉTUDE BIBLIOGRAPHIQUE.....	17
Chapitre I : Synthèse totale de protéines par voie chimique	19
1. Émergence de la synthèse de peptides et de protéines par voie chimique	19
2. Les débuts de la synthèse de protéines par voie chimique	20
2.1 Synthèse peptidique sur support solide.....	20
2.2 Assemblage de fragments peptidiques synthétisés par SPPS.....	22
2.2.1 Approche séquentielle et approche convergente.....	22
2.2.2 Fragments peptidiques protégés.....	22
3. La ligation chimique.....	23
3.1 Principe.....	23
3.2 Ligation chimique native (NCL).....	24
3.2.1 Les travaux de T. Wieland.....	24
3.2.2 La ligation chimique native.....	25
3.2.3 La ligation native de type Staudinger	26
3.2.4 La ligation "α-ketoacid-hydroxylamine" (KAHA)	27
3.2.5 La ligation SEA	29
3.2.6 La ligation SEALide.....	30
4. Accès à de grandes protéines.....	30
4.1 Assemblage de C vers N	31
4.1.1 Groupements protecteurs.....	31
4.1.2 Les procédés one pot de C vers N	33
4.2 Assemblage N vers C	34
4.2.1 Les peptides hydrazides.....	35
4.2.2 Les systèmes de transfert d'acyle de O→S.....	36
4.2.3 Les systèmes de transfert d'acyle de N→S.....	37
4.2.4 Les procédés one pot de N vers C	38
5. Conclusion/Bilan.....	40
Chapitre II : Aspects mécanistiques de la ligation chimique native	43
1. Mécanisme général	43

2.	Dissection des étapes individuelles de la NCL.....	43
2.1	Échanges thiol-thioester.....	44
2.1.1	Généralité.....	44
2.1.2	Catalyse.....	48
2.2	Transfert d'acyle S→N intramoléculaire.....	53
3.	Rôle des agents dénaturants, détergents et solvants en NCL.....	56
3.1	Les agents dénaturants.....	56
3.1.1	L'urée.....	57
3.1.2	Le chlorhydrate de guanidine (Gdn.HCl).....	57
3.2	Les détergents ou tensioactifs.....	59
3.3	Les solvants organiques.....	60
4.	Réactions secondaires au niveau des cystéines/sélocystéines.....	61
4.1	Réaction d'oxydation des thiols/sélocols.....	61
4.2	L'acide ascorbique : inhibiteur d'oxydation.....	63
4.3	Agent réducteur de disulfures ou disélocures : la TCEP.....	64
4.3.1	Propriétés générales.....	64
4.3.2	Stabilité en solutions aqueuses.....	65
4.3.3	Déchalcoénéation.....	68
5.	Bilan/Conclusion.....	69
	Chapitre III : Vision statistique de la synthèse chimique de protéines.....	71
1.	Fréquence des acides aminés protéinogéniques.....	71
2.	Le processus de désulfurisation.....	72
3.	L'introduction de nouveaux précurseurs de thioester.....	73
4.	Les types de jonctions Xaa-Cys.....	75
5.	Taille des protéines produites par synthèse chimique.....	77
6.	Bilan/Conclusion.....	79
	RÉSULTATS ET DISCUSSION.....	81
	Chapitre IV : Stratégie d'assemblage « one-pot » utilisant des peptides sélocnoester de type bis(2-sélocnyléthyl)amido (SeEA).....	83
1.	La problématique d'assemblage de multiples segments peptidiques.....	83
2.	L'émergence des sélocnoesters comme donneurs d'acyle dans la réaction de NCL.....	83
3.	La synthèse des peptides bis(2-sélocnyléthyl)amino (SeEA).....	86
4.	Les propriétés oxydoréductrices du groupement bis(2(sélocnyléthyl)amine (SeEA).....	88
5.	La réactivité des peptides bis(2(sélocnyléthyl)amido (SeEA).....	90
6.	Assemblage de la protéine NK1.....	94
6.1	La protéine NK1 : un modèle de protéine thérapeutique.....	94
6.2	Synthèse totale de la protéine NK1.....	96

6.2.1	Synthèse d'un sélénoester latent.....	97
6.2.2	La ligation cinétiquement contrôlée SeEA/SEA.....	99
Chapitre V : Développement de nouveaux catalyseurs		103
1.	Problématique de la ligation aux jonctions difficiles.....	103
2.	Transthioestérification	103
2.1	Aspects mécanistiques	103
2.2	Émergence de nouveaux catalyseurs à base de sélénium	106
2.2.1	Résultats préliminaires.....	106
2.2.2	Synthèse des catalyseurs N-alkyl disélénures	110
2.2.3	Catalyse de la formation de jonctions difficiles.....	112
3.	Ligation SEA	114
4.	La synthèse totale de la granulysine (23-145).....	117
4.1	Intérêt biologique.....	117
4.2	Synthèse totale de la granulysine 15 kDa	120
4.2.1	Assemblage de N vers C.....	121
4.2.2	Assemblage de C vers N.....	122
Chapitre VI : Développement d'un procédé automatisé de synthèse de protéines en phase solide.		125
1.	La problématique d'assemblage de segments peptidiques en solution.....	125
2.	Émergence de la ligation native en phase solide (SPCL)	125
3.	Conception d'une stratégie d'assemblage en phase solide	126
3.1	Procédé chimique : la SEA SPPS	126
3.2	Développement d'un bras espaceur chimiosélectif.....	129
3.3	Contraintes du procédé.....	130
4.	Synthèse supportée et automatisée de protéines.....	131
4.1	Caractéristiques techniques du prototype.....	131
4.2	Mise en place de la synthèse.....	134
4.2.1	Synthèse du support solide	134
4.2.2	Procédure d'élongation	134
4.3	Résultat préliminaire/ Premier résultat	137
5.	Optimisation.....	138
5.1	Processus chimique	139
5.1.1	L'étape d'activation : l'échange SEA-thiol	139
5.1.2	L'étape de ligation (NCL)	141
5.2	Support solide.....	144
CONCLUSION ET PERSPECTIVES		147
RÉFÉRENCES BIBLIOGRAPHIQUES		151
ANNEXES.....		161

LISTE DES ABRÉVIATIONS

Ac	Acétyle
Acm	Acétamidométhyle
ACN	Acétonitrile
ADN	Acide désoxyribonucléique
Alk	Alkyle
Ar	Aryle
Bn	Benzyle
Boc	<i>tert</i> -butyloxycarbonyle
Bz	Benzoyle
CMC	Concentration micellaire critique
COSY	Correlation spectroscopy (RMN 2D)
CPE	Ester de cystéinyl proline
CuAAC	<i>Copper-Catalyzed Azide-Alkyne Cycloaddition</i>
DCE	Dichloroéthane
DCM	Dichlorométhane
DIEA	<i>N,N</i> -diisopropyléthylamine
DMF	<i>N,N</i> -diméthylformamide
DMS	Diméthylsulfure
DMSO	Diméthylsulfoxyde
DTT	Dithiothréitol
DST	Disélénothréitol
EDTA	Acide éthylène diamine tétraacétique
EPL	Ligation de protéines exprimées (<i>Expressed Protein Ligation</i>)
ESI	Ionisation par électronébulisateur (<i>Electrospray Ionization</i>)
Et	Ethyle
Et₂O	Ether diéthylique
FDA	Agence américaine des produits alimentaires et médicamenteux (<i>Food and Drug Administration</i>)
Fmoc	9-fluorénylméthoxycarbonyle
Gdn.HCl	Chlorhydrate de guanidine
Gdn.HSCN	Thiocyanate de guanidine
GFP	Protéine fluorescente verte
GLNY	Granulysine
HATU	<i>N</i> -[(diméthylamino)-1 <i>H</i> -1,2,3-triazolo[4,5- <i>b</i>]pyridino-1-ylméthylène]- <i>N</i> -méthylmethanaminium hexafluorophosphate <i>N</i> -oxide
HBTU	<i>N</i> -[(1 <i>H</i> -benzotriazol-1-yl)(diméthylamino)méthylène]- <i>N</i> -méthylmethanaminium hexafluorophosphate <i>N</i> -oxide
HGF/SF	Facteur de croissance des hépatocytes/Facteur de dispersion
Hnb	<i>N</i> -hydroxynitrobenzyle
HRMS	Spectrométrie de masse haute résolution
HSQC	Heteronuclear single quantum coherence (RMN 2D)
KAHA	Ligation cétoacide-hydroxylamine (<i>Ketoacid-Hydroxylamine ligation</i>)

KCL	Assemblage cinétiquement contrôlé (<i>Kinetically Controlled Ligation</i>)
LC-MS	Chromatographie en phase liquide couplée à la spectrométrie de masse
LFER	Relation linéaire d'énergie libre
LG	Groupement partant (<i>Leaving Group</i>)
MALDI-TOF	<i>Matrix-assisted laser desorption ionisation-time of flight</i>
Mapoc	(4-diméthylamino)phénacyloxy-carbonyle
MDCK	Cellules épithéliales rénales canines (<i>Madin-Darby Canine Kidney</i>)
Me	Méthyle
Me-MPA	Méthyl ester de l'acide 3-mercaptopropionique
MESNa	2-mercaptoéthanesulfonate de sodium
MPA	Acide 3-mercaptopropionique
MPAA	Acide 4-mercaptophénylacétique
MPAA-Na	4-mercaptophénylacétate de sodium
MPS	Acide 3-mercaptopropiosulfonique
MS	Spectrométrie de masse
Msc	2-méthylsulfonyléthoxy-carbonyle
MTG	Méthylthioglycolate
Nbz	<i>N</i> -acetylbenzimidazolone
NCL	Ligation chimique native (<i>Native Chemical Ligation</i>)
NMM	<i>N</i> -méthyl-morpholine
NMP	<i>N</i> -méthyl-2-pyrrolidone
Nvoc	6-nitroveratryloxy-carbonyle
Pac	Phénacyle
PCM	Modèle polarisable continu (<i>Polarizable Continuum Model</i>)
PCS	Base de données de protéines obtenues par synthèse chimique (<i>Protein Chemical Synthesis Database</i>)
PEG	Polyéthylène glycol
PG	Groupement protecteur (<i>Protecting group</i>)
Ph	Groupement phényle
PyBOP	Benzotriazol-1-yloxytri(pyrrolidino) phosphonium hexafluorophosphate
RMN	Résonance magnétique nucléaire
RP-HPLC	Chromatographie liquide à haute performance en phase inverse
RTK	Récepteur tyrosine kinase
SDS	Dodécylsulfate de sodium
SEA	<i>N,N</i> -Bis(2-sulfanyléthyl)amide
SeEA	<i>N,N</i> -Bis(2-sélényléthyl)amide
SEAlide	<i>N</i> -sulfanyléthylanilide
Sec	Sélénocystéine
SPCL	Ligation chimique en phase solide (<i>Solid Phase Chemical Ligation</i>)
SPH	Domaine homologue aux sérines protéases (<i>serine proteinase homology domain</i>)
SPPS	Synthèse peptidique en phase solide (<i>Solid Phase Peptide Synthesis</i>)
tBu	<i>tert</i> -butyle
TCEP•HCl	Chlorohydrate de <i>tris</i> (2-carboxyéthyl)phosphine
T_{eb}	Température d'ébullition

TFA	Acide trifluoroacétique
TFE	Trifluoroéthanol
TFET	Trifluoroéthanethiol
THF	Tétrahydrofurane
Thz	Cysteine thiazolidine
TIS	Triisopropylsilane
Trt	Trityle
TS	Etat de transition ou état transitoire
UV	Ultra-violet
VA-044	Dihydrochlorhydrate de 2,2'-azobis[2-(2-imidazolin-2-yl)propane]
Xaa	Acide aminé quelconque

RÉSUMÉ

Les protéines jouent un rôle essentiel dans le fonctionnement des organismes vivants et sont au cœur de nombreux mécanismes biologiques. La synthèse totale chimique des protéines permet un contrôle précis de leur composition et constitue donc un outil puissant d'investigation des processus biologiques impliquant ces molécules. La synthèse peptidique en phase solide (SPPS) décrite par B. Merrifield en 1963 a révolutionné le domaine. Quelques décennies plus tard, les méthodes de ligation chimiques et le développement de stratégies d'assemblage de segments peptidiques ont permis la synthèse de nombreuses protéines par voie chimique. Cependant, la difficulté des chimistes à produire de grandes protéines traduit l'absence de méthodes robustes pour la synthèse en routine de tels macromolécules.

De nouvelles méthodologies de synthèse totale basées sur la *Native Chemical Ligation* et les ligations SEA et SeEA ont été développées dans le cadre de cette thèse. En particulier, l'utilisation de sélénioesters latents ou formés *in situ* a été étudiée pour faciliter l'accès à des protéines de grande taille.

La première partie de cette thèse porte sur le potentiel du groupement *bis*(2-sélényléthyl)amido SeEA, à savoir l'analogue sélénié du groupement *bis*(2-sulfanyléthyl)amido SEA, pour accélérer la formation de liaisons peptidiques par formation d'intermédiaires sélénioesters. Cette méthode a permis la synthèse de la protéine NK1 (variant naturel du facteur de croissance des hépatocytes HGF) constituée de 180 acides aminés.

La seconde partie décrit la formation *in situ* de sélénioesters à partir de segments peptidiques SEA grâce à la conception de nouveaux catalyseurs à base de sélénium. Ces catalyseurs ont été utilisés pour accélérer deux réactions essentielles dans les procédés de synthèse totale développés dans cette thèse, à savoir la ligation SEA ainsi que la synthèse de peptides thioesters à partir de peptides SEA. L'efficacité de l'un de ces catalyseurs a été illustrée par la synthèse totale de la granulysine (123 acides aminés).

La dernière partie de cette thèse décrit une stratégie d'assemblage séquentielle de segments peptidiques en phase solide, qui constitue une alternative aux méthodes d'assemblage en phase liquide. Ce procédé a été automatisé et présente un grand potentiel pour la synthèse automatisée de protéines par voie chimique.

ABSTRACT

Proteins play a crucial role in living organisms and in almost all biological mechanisms. The total chemical synthesis of proteins allows an atom by atom control of their structure and thus constitutes a powerful tool for investigating the biological processes involving these molecules. The solid phase peptide synthesis (SPPS) introduced by B. Merrifield in 1963 has revolutionized the field. Few decades later, the discovery of chemical ligation methods and the development of peptide segment assembling strategies has been applied to the synthesis of many proteins. However, the difficulty in accessing large proteins reflects the absence of robust methods for the routine synthesis of such macromolecules.

Novel synthetic methods based on Native Chemical Ligation and SEA/SeEA ligations have been developed in the frame of this thesis. In particular, I explored the interest of latent or *in situ* formed selenesters for facilitating the access to large proteins.

The first part of this thesis describes the chemical properties of the *bis*(2-selenylethyl)amido (SeEA) group, i.e. the selenium analog of the *bis*(2-sulfanylethyl)amido (SEA) group, and its usefulness for accelerating peptide bond formation. This method was used for the synthesis of the NK1 protein (natural variant of the hepatocyte growth factor, HGF) constituted of 180 amino acids.

The second part describes the *in situ* formation of selenesters through the design of novel selenium catalysts. These catalysts were used to accelerate the SEA ligation as well as the synthesis of thioester peptides from SEA peptides. Both reactions are central in the total synthesis processes developed in this thesis. Their usefulness is illustrated by the total synthesis of granulysin (123 amino acids).

The last part describes a method for the sequential ligation of peptide segments on a water compatible solid support which is complementary to the solution methods discussed above. This process has been automated and has a great potential for the automated chemical synthesis of proteins.

ÉTUDE BIBLIOGRAPHIQUE

CHAPITRE I : SYNTHÈSE TOTALE DE PROTÉINES PAR VOIE CHIMIQUE

1. Émergence de la synthèse de peptides et de protéines par voie chimique

En 1835, le chimiste Gerardus Johannes Mulder découvre que le blanc d'œuf est une substance constituée d'une base de carbone, d'hydrogène, d'azote et d'oxygène associés à de petites quantités de soufre et de phosphore¹. En raison de la forte abondance de substances similaires dans des tissus aussi bien animaux que végétaux, son confrère Jons Jacob Berzelius lui suggère en 1838 de les nommer protéines (du grec *protos* signifiant premier). C'est ainsi que le blanc d'œuf devint le matériel pionnier dans les découvertes des protéines.

Quelques années plus tard, les chimistes Franz Hofmeister (1902) et Emil Fisher (1906) introduisirent la notion de liaison peptidique, d'acide aminée et de chaîne latérale. Fisher fut le premier à synthétiser des polypeptides. Les techniques de synthèse peptidique progressèrent au fil des années et en 1953, Vincent du Vigneaud réussit la synthèse de plusieurs hormones peptidiques, l'ocytocine² et la vasopressine, constituées de 9 acides aminés. Il obtint le prix Nobel en 1955 pour ces travaux. En parallèle, F. Sanger détermina la structure primaire de l'insuline humaine³ qui comporte 51 acides aminés. En 1958, il obtint lui aussi le prix Nobel de chimie pour ses travaux sur la structure des protéines. La synthèse chimique des protéines connut une avancée considérable en 1963 avec les travaux de B. Gutte et R. B. Merrifield portant sur la synthèse sur support solide de peptides. En 1969, grâce à ce procédé, ils réalisèrent la première synthèse totale d'une enzyme active de 124 acides aminés, la ribonucléase A⁴. Au-delà de l'exploit d'avoir produit une protéine de cette taille par voie chimique, ce travail fut la démonstration que la séquence primaire d'une protéine peut contenir toute l'information permettant à la chaîne polypeptidique de se conformer correctement dans l'espace.

Dès lors, l'intérêt des scientifiques de toutes disciplines pour les protéines de synthèse n'a cessé de croître. Les protéines jouent un rôle essentiel dans le fonctionnement des organismes vivants et sont la clé de nombreux mécanismes biologiques. La synthèse chimique des protéines constitue donc un outil puissant d'investigation des processus biologiques impliquant ces molécules. L'accès à des peptides ou protéines de synthèse ouvre également de nombreuses possibilités dans des domaines de recherche très divers comme la chimie, l'agronomie, la médecine et la science des matériaux⁵. La recherche de méthodes de synthèse et de purification de ces macromolécules est toujours en plein essor. Elles induisent aujourd'hui un effort de recherche considérable pour la mise au point de méthodes biologiques, chimiques ou hybrides pour accéder à ces molécules complexes. Les retombées attendues notamment dans le domaine de la santé sont importantes.

Depuis quelques années, la synthèse chimique est perçue comme une alternative à l'utilisation de systèmes vivants pour accéder à des protéines d'intérêt biologique. L'une des applications les plus intéressantes de la synthèse chimique est la production de protéines modifiées que les systèmes vivants peuvent difficilement produire ou ne peuvent pas produire du tout⁶. Cela comprend des protéines toxiques ou des protéines qui ont des architectures inhabituelles telles que des protéines cycliques ou des protéines ramifiées. Les méthodes chimiques permettent également la production de protéines marquées sur des sites précis. On peut ainsi imaginer l'introduction de modifications non natives comme des acides aminés non naturels^{7,8}, des isotopes stables^{9,10} ou des fluorophores^{11,12}. Enfin, la synthèse chimique est intensivement utilisée pour l'étude des modifications post-traductionnelles, qu'elles soient simples comme la phosphorylation¹³, l'acétylation^{13,14} ou plus complexes comme l'ubiquitination^{15,16} ou la sumoylation^{17,18}. En effet, ces modifications sont souvent difficiles à introduire à l'aide des techniques recombinantes.

La chimie apparaît donc comme une science complémentaire à la biologie pour l'obtention de protéines de grande taille. Elle peut permettre de comprendre comment les différentes modifications contribuent à moduler les propriétés des protéines et leurs modes d'action.

2. Les débuts de la synthèse de protéines par voie chimique

2.1 Synthèse peptidique sur support solide

L'introduction par R.B. Merrifield de la synthèse peptidique en phase solide (SPPS) a permis de produire de petits peptides par couplages successifs d'acides aminés sur un support solide. L'assemblage est réalisé par répétition d'un cycle de déprotection/couplage (étape 2 et 3, Figure 1). Une étape d'acylation permet de masquer les fonctions amine libres résultant d'un couplage incomplet, on parle alors d'une étape de *capping*. Habituellement, il s'agit d'une réaction d'acétylation (étape 4, Figure 1). Finalement, le peptide final est obtenu par déprotection des chaînes latérales et coupure du lien entre le peptide et le support solide (étape 5, Figure 1). Ce procédé a très vite été automatisé pour permettre un accès simple, robuste, flexible et rapide à de petites protéines fonctionnelles¹⁹. Aujourd'hui, il existe une multitude de résines, d'agents de couplage, de groupements protecteurs de chaînes latérales permettant d'améliorer les procédés de synthèse. Toutes ces optimisations permettent l'accès à des peptides allant jusqu'à 50 acides aminés environ. Il existe toutefois quelques exceptions dans la littérature résultant d'optimisations avancées et conduisant à des protéines d'une centaine d'acides aminés¹⁴. Mis à part ces cas particuliers, la synthèse de peptides de plus de 50 acides aminés reste un défi. En effet, des couplages incomplets lors de l'élongation et des réactions secondaires se traduisent par l'accumulation d'une grande quantité d'impuretés qui

compliquent l'étape de purification finale et conduisent à de faibles rendements globaux. Ces limitations ont stimulé la recherche de nouveaux procédés chimiques d'obtention de longs peptides et notamment de différentes techniques de ligation chimique.

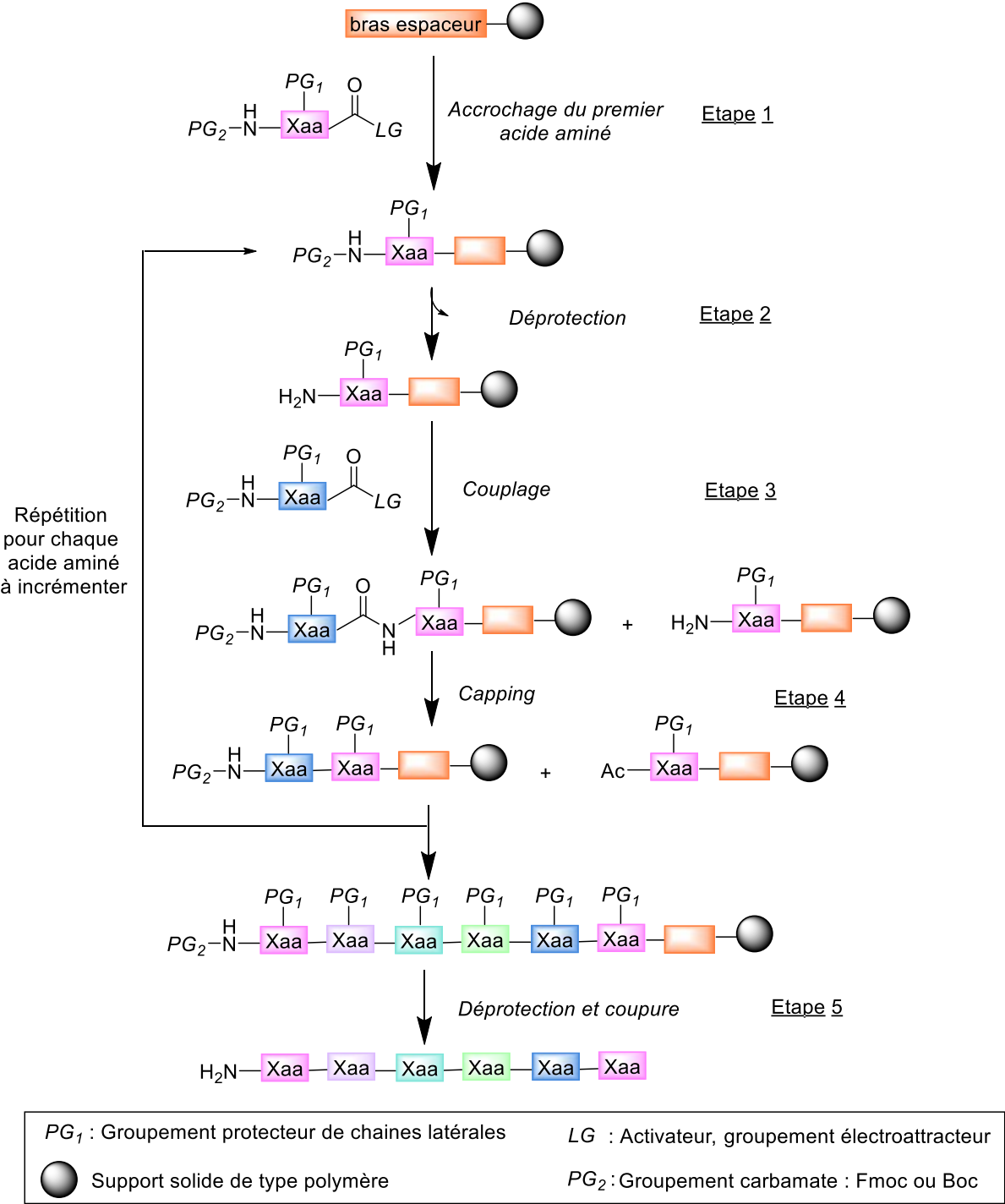


Figure 1 : Principe de la synthèse peptidique sur support solide (SPPS).

2.2 Assemblage de fragments peptidiques synthétisés par SPPS

2.2.1 Approche séquentielle et approche convergente

L'assemblage de segments peptidiques préparés par SPPS permet logiquement d'augmenter la taille des protéines synthétiques. Il existe deux approches différentes pour atteindre cet objectif, une première dite séquentielle et une seconde dite convergente. Comme son nom l'indique, la première stratégie consiste à assembler de manière séquentielle des fragments peptidiques. C'est un principe similaire à celui utilisé en SPPS, à la différence que l'on assemble des segments peptidiques au lieu d'acides aminés. La seconde approche consiste en l'assemblage de bloc de tailles similaires. Cette stratégie est dite convergente car elle permet de concevoir des blocs qui par combinaisons différentes permettent d'obtenir une librairie de protéines. Finalement, pour choisir l'une ou l'autre des stratégies d'assemblage, il est important de tenir compte des propriétés intrinsèques des segments et des intermédiaires générés, notamment en termes de solubilité.

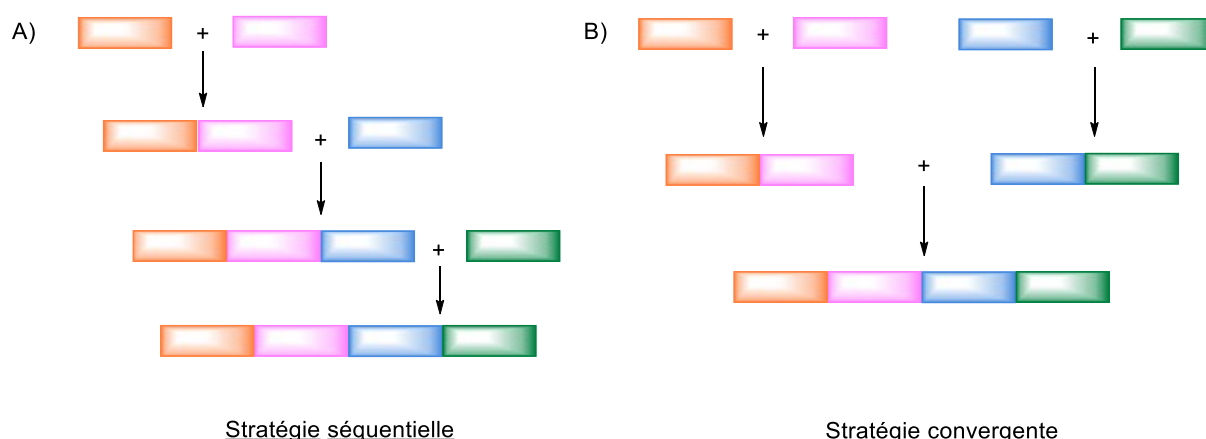


Figure 2 : Stratégie d'assemblage A) Séquentielle (linéaire) B) Convergente.

2.2.2 Fragments peptidiques protégés

Initialement, les peptides synthétisés par SPPS ont été assemblés en milieu organique selon les principes utilisés en solution ou en SPPS pour le couplage d'acides aminés (activation de la fonction acide carboxylique, couplage, déprotection de l'amine). L'assemblage a donc lieu entre deux fragments protégés pour éviter les réactions secondaires avec les chaînes latérales des peptides lors du couplage.

À titre d'exemple, le groupe de F. Albericio a décrit en 2003 la synthèse d'une protéine de 66 acides aminés par condensation convergente en solution de fragments peptidiques protégés synthétisés par Boc-SPPS²⁰. Cette stratégie nécessite des groupements protecteurs de chaînes latérales stables dans les conditions de coupure du peptide de la résine. De plus, les fonctions acides

carboxyliques des segments ont dû être protégées de façon temporaire grâce aux groupements phénacyle (Pac) ou benzyle (Bn).

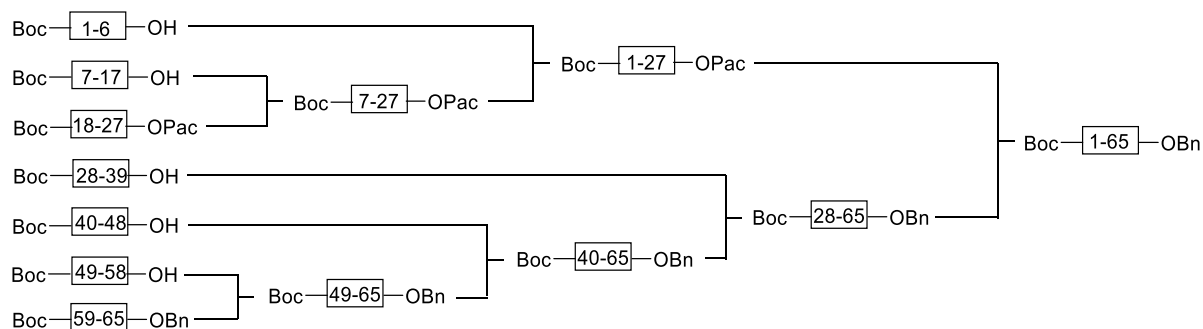


Figure 3 : Stratégie d'assemblage convergente pour la synthèse de P41icf.

Ce type de procédé a permis la synthèse d'objets remarquables comme la *green fluorescence protein* (GFP) composée de 238 acides aminés²¹. Cependant, l'utilisation de segments peptidiques totalement ou partiellement protégés présente un certain nombre d'inconvénients (difficulté de solubilisation des fragments, temps de réaction longs, risque d'épimérisation). Ces limitations en font une méthode peu utilisée à l'heure actuelle et ont stimulé le développement de techniques de ligation de fragments peptidiques totalement déprotégés, qui sont plus facile à produire, à purifier et à solubiliser.

3. La ligation chimique

3.1 Principe

Le terme de ligation chimique a été introduit en 1992 par M. Schnölzer et S. Kent pour décrire une réaction régio- et chimiosélective entre deux fragments peptidiques non-protégés²². Le principe consiste à utiliser des fonctions chimiques particulières aux extrémités des peptides à assembler. Ces fonctions doivent être spécifiques et avoir une réactivité adaptée l'une à l'autre (fonctions X et Y, Figure 4). Les réactions de ligation chimique permettent d'utiliser des segments peptidiques non protégés qui présentent une meilleure solubilité que leur analogue protégé. Différentes réactions chimiosélectives conduisant à des liens non natifs ont été décrites dans la littérature et ne seront pas discutées dans le cadre de cette thèse (on pourra consulter d'excellentes revues à ce sujet^{23,24}).

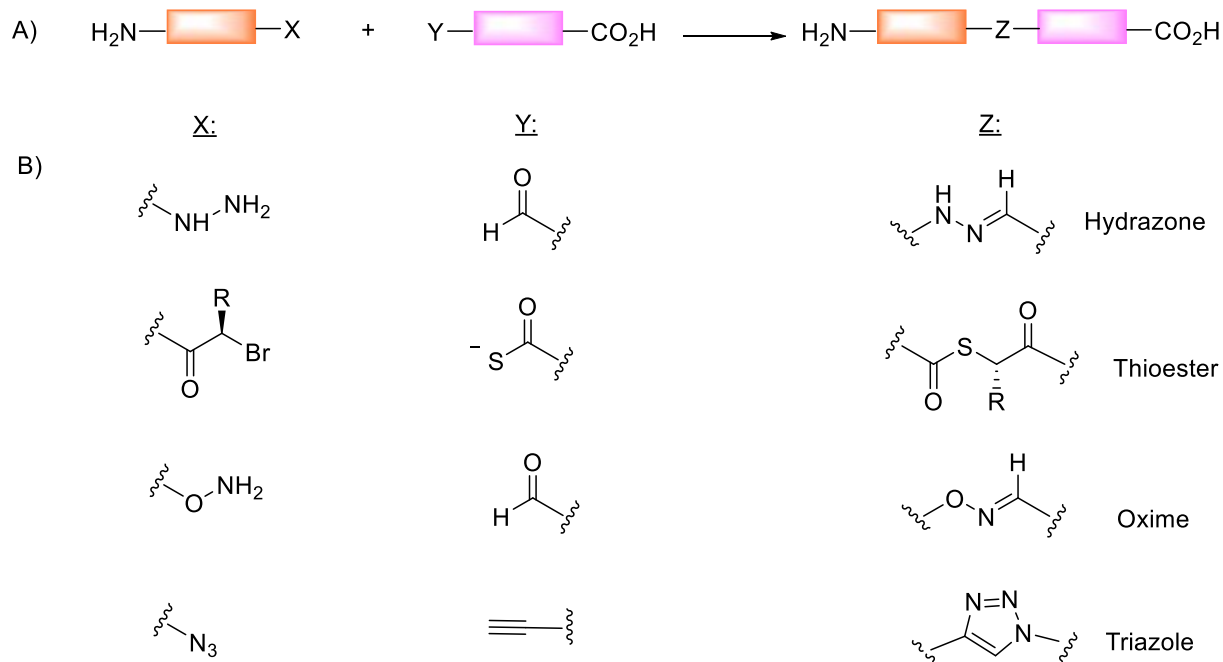


Figure 4 : A) Principe de la ligation chimique. B) Exemples de ligations non natives : hydrazone²⁵, thioester²², oxime^{26,27}, triazole²⁸.

L'utilisation de ces méthodes de ligations chimiques non natives a permis la synthèse de nombreuses macromolécules et mimes de protéines. Cependant, la nécessité d'obtenir des protéines natives par voie chimique notamment pour l'étude de leurs fonctions biologiques a stimulé la recherche de méthodes conduisant à un lien peptidique natif.

3.2 Ligation chimique native (NCL)

3.2.1 Les travaux de T. Wieland

En 1953, T. Wieland a décrit la réaction entre un thioester de valine et une cystéine à pH neutre et en milieu aqueux²⁹. Il a observé une cinétique de réaction particulièrement rapide avec la cystéine au regard d'autres acides aminés. Il a expliqué cette particularité par la formation d'un intermédiaire thioester par échange thiol-thioester, suivi d'un réarrangement intramoléculaire par transfert d'acyle entre le soufre et l'azote (Figure 5).

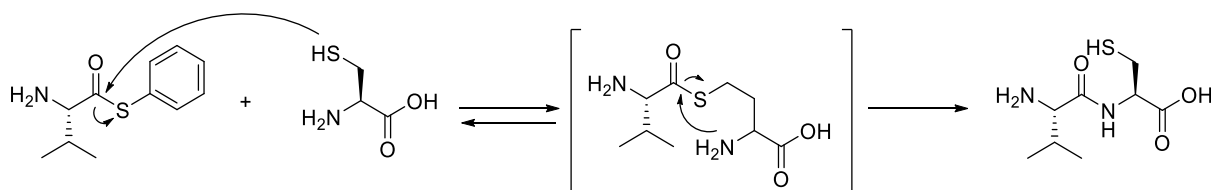


Figure 5 : Réaction d'un thioester de valine avec une cystéine conduisant à la formation d'un lien natif.

Inspiré par les travaux de T. Wieland et M. Brenner³⁰, D. S. Kemp a décrit en 1981 une nouvelle stratégie de ligation pour la formation de liens peptidiques natifs. Cette méthode met en jeu une étape de capture par échange thiol-disulfure (étape 1, Figure 6) qui est suivie d'une migration d'acyle intramoléculaire de l'oxygène vers l'azote (étape 2, Figure 6)³¹. Le produit final est obtenu par réduction du pont disulfure liant l'auxiliaire à la chaîne latérale de la cystéine (étape 3, Figure 6). Ce procédé a été appliqué à la synthèse de deux peptides, l'un composé de 25 acides aminés et l'autre de 49 acides aminés³². La méthode de Kemp n'est pas totalement chimiosélective puisque les cystéines internes doivent être protégées. Cependant, ces travaux ont permis de mettre en place les principaux concepts à la base des méthodes modernes de ligation chimique native : une première étape de capture chimiosélective permet de rapprocher les extrémités réactives puis un réarrangement intramoléculaire conduit à la formation d'un lien peptidique natif.

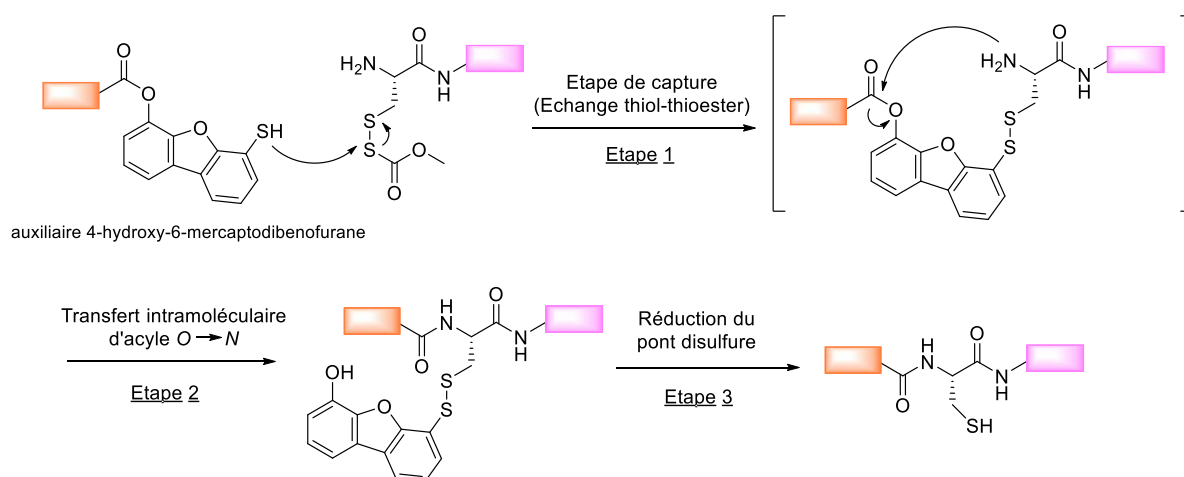


Figure 6 : Ligation chimique native par capture de thiol développée par D. S. Kemp.

3.2.2 La ligation chimique native

En 1994, près de quarante ans après l'observation de T. Wieland, le groupe de S. Kent introduisit une réaction clé dans le domaine de la synthèse totale de protéines, à savoir la *Native Chemical Ligation* (NCL)³³. Cette méthode se définit par la réaction d'un peptide thioester (fragment A) avec un peptide cystéinyl (fragment B) conduisant à un lien amide natif de type Xaa-Cys. Plus précisément, la fonction thiol de la cystéine du fragment B attaque la fonction thioester du fragment A. Cette réaction dite de transthioestérification (échange thiol-thioester) conduit à la formation d'un intermédiaire thioester transitoire. Celui-ci se réarrange spontanément par migration du groupement acyle entre le soufre et l'azote.

Depuis, le principe de la ligation chimique native a été étendu à l'utilisation de sélénocystéinyl peptides (Figure 7, Y = Se) et de sélénioesters (Figure 7, X = Se). À noter qu'il est également possible d'assembler des segments issus du génie biologique, on parle alors d'*Expressed Protein Ligation* (EPL).

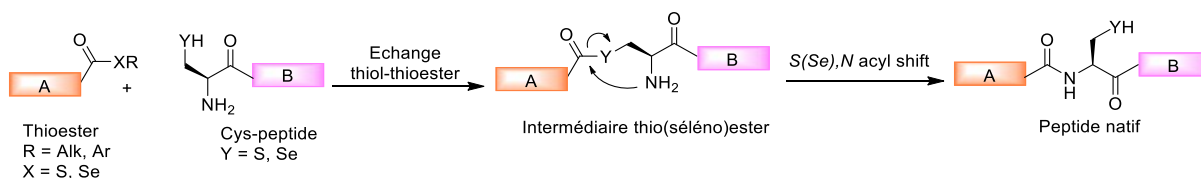


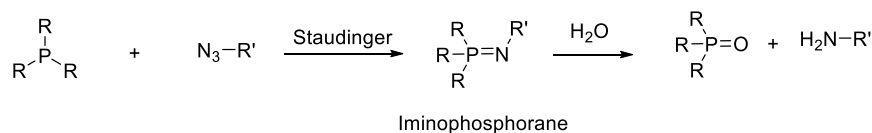
Figure 7 : Principe de la ligation chimique native.

La NCL est à l'heure actuelle la méthode de ligation chimiosélective la plus utilisée pour la synthèse totale de protéines.

3.2.3 La ligation native de type Staudinger

En 2000, deux groupes distincts, ceux de C. Bertozzi³⁴ et de R. T. Raines³⁵, ont décrits indépendamment l'application de la réaction de Staudinger³⁶ à la synthèse de peptides par ligation chimiosélective. La ligation Staudinger fait intervenir un azoture qui réagit avec une phosphine pour former un iminophosphorane dans l'étape de capture. Dans le cas de la réaction de Staudinger, cet intermédiaire est hydrolysé pour former une amine (Figure 8A). Dans le cas de la ligation dite de Staudinger, un peptide portant une fonction phosphinothioester réagit avec un peptide comportant un azoture en N-terminal. Cette réaction produit un aza-ylure intermédiaire qui, par transfert d'acyle intramoléculaire du soufre vers l'azote, conduit à la formation d'un sel d'amidophosphonium. Ce dernier s'hydrolyse spontanément pour former un lien peptidique amide (Figure 8B).

A) Réaction de Staudinger



B) Ligation de Staudinger

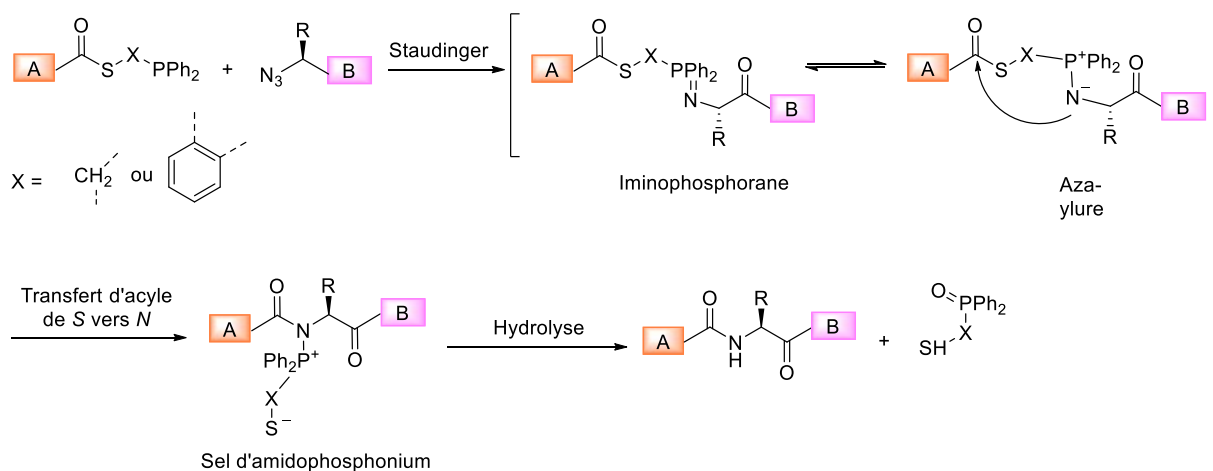


Figure 8 : A) Principe de la réaction de Staudinger B) Ligation de Staudinger.

Contrairement à la NCL, cette réaction peut en théorie être appliquée à la formation de n'importe quelle jonction peptidique (Xaa-Xaa). Cependant, l'obtention, la manutention et le stockage de dérivés azotures sont complexes et limitent fortement l'utilisation de la ligation Staudinger. De plus, les esters (ou thioesters) de phosphines utilisés pour former les peptides phosphinothioesters peuvent réagir avec les acides aminés nucléophiles des peptides non protégés et conduire à des réactions secondaires.

3.2.4 La ligation "α-ketoacid-hydroxylamine" (KAHA)

En 2006, le groupe de Bode a décrit une réaction entre un peptide comportant une hydroxylamine en N-terminal et un peptide portant un α-cétoacide en C-terminal (Figure 9A)³⁷. Cette réaction appelée *α-ketoacid-hydroxylamine ligation* (ligation KAHA) conduit à la formation d'un lien peptidique natif. Elle ne nécessite aucun ajout de réactif et produit uniquement de l'eau et du dioxyde de carbone comme co-produits de la réaction. La version décrite initialement s'effectue en milieu organique (DMF, DMSO ou MeOH) et implique des hydroxylamines non substituées (KAHA de type 1, R=H, Figure 9A). Par la suite, Bode et ses collaborateurs développèrent les conditions pour l'utilisation de dérivés hydroxylamines O-substitués qui réagissent bien plus rapidement en milieu aqueux³⁸ (KAHA de type 2, R= alkyl).

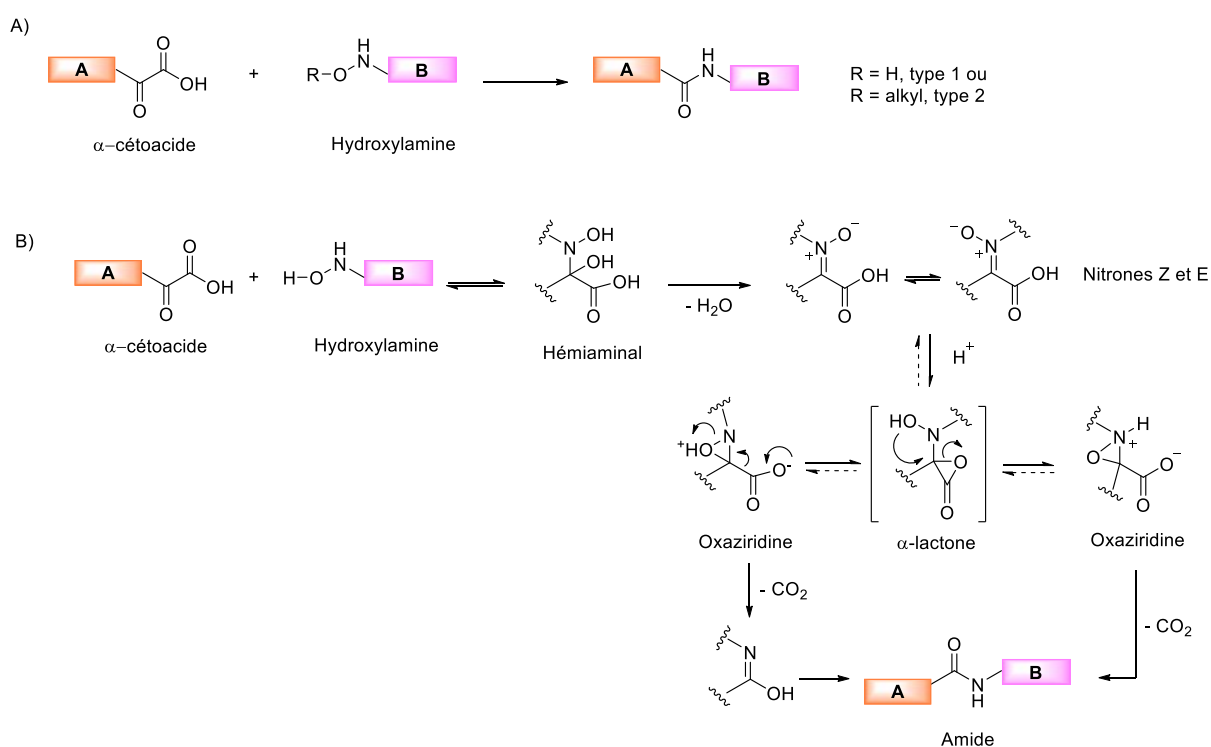


Figure 9 : A) Principe de la ligation KAHA. B) Mécanisme dans le cas où R = H³⁸.

Le mécanisme de la ligation KAHA n'est décrit que pour l'utilisation de composés de type 1 (R=H, Figure 9B). Il implique la formation d'un héli-aminal suivie d'une perte d'une molécule d'eau

conduisant à la formation d'une nitronne. Finalement, plusieurs réarrangements suivis d'une décarboxylation conduisent à la formation du lien amide.

Les hydroxylamines *O*-substituées sont très intéressantes pour l'obtention de protéines puisqu'elles impliquent des conditions aqueuses pour la ligation. Cependant leur synthèse nécessite des procédés longs et multi-étapes. Quelques années plus tard, Bode introduisit un résidu facilement accessible, le motif 5-oxaproline (Figure 10A)³⁹. Ce composé est facilement synthétisé et incorporé sur le peptide. De plus, il est stable dans les conditions de déprotection/coupage des peptides. La réaction impliquant ce résidu conduit à la formation d'une homosérine après réarrangement intramoléculaire à pH basique.

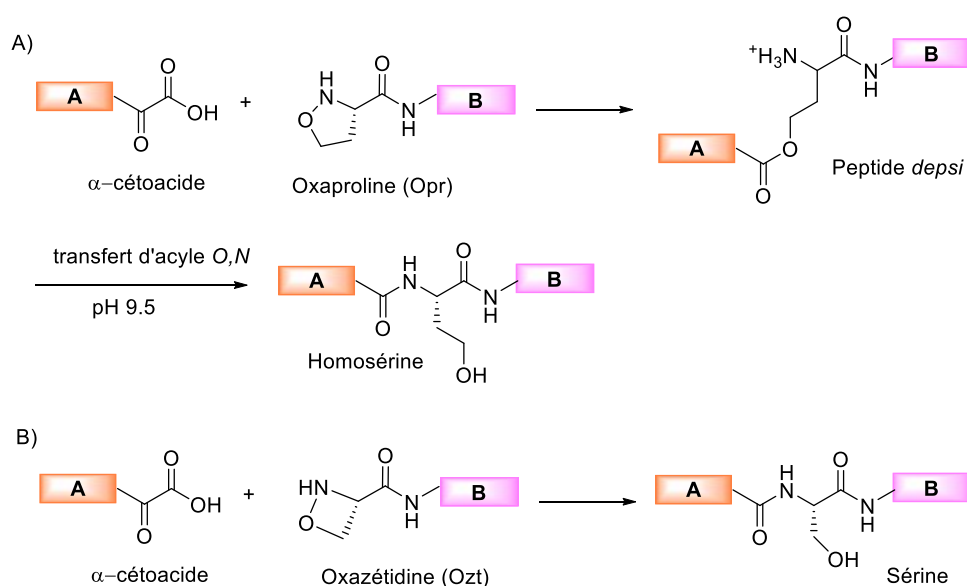


Figure 10 : Optimisation de la ligation KAHA en utilisant A) un motif oxaproline. B) un motif oxazétidine⁴⁰.

L'utilisation de la ligation KAHA impliquant des 5-oxaprolyl peptides a permis la synthèse totale de petites protéines comme un analogue de l'ubiquitine procaryotique (protéine Pup, 66 acides aminés)³⁹ mais également la synthèse de protéines de taille plus importante comme les domaines SUMO 2 et 3 humains (>90 acides aminés)⁴¹. Cependant, l'utilisation de motifs hydroxylamines substitués, quels qu'ils soient, requière de fortes concentrations en peptides (10 à 20 mM). Ceci représente une limitation non négligeable à l'utilisation de la ligation KAHA. Une optimisation très récente de la méthode (Figure 10B) a permis d'utiliser de plus faibles concentrations en peptides (100 μ M à 5 mM) et des cinétiques de réaction plus favorables grâce à la tension de cycle élevée de l'hétérocycle oxazétidine qui augmente sa réactivité. Cette variante de la ligation KAHA a permis la synthèse totale de la protéine de liaison au calcium S104 (101 acides aminés), dont la séquence primaire rend difficile sa synthèse par NCL⁴⁰.

3.2.5 La ligation SEA

La ligation *bis*(2-sulfanyléthyl)amido (SEA) a été décrite de façon indépendante par les groupes d'O. Melnyk (2010)⁴² puis de C. F. Liu (2011)⁴³. Elle consiste essentiellement à assembler un peptide équipé d'un groupement SEA à l'extrémité C-terminale avec un cystéinyl peptide. Cette réaction conduit à la formation d'une liaison peptidique native (Figure 11). Les peptides SEA sont obtenus par SPPS sous une forme de dithiol en équilibre avec une forme thioester (transfert d'acyle de N→S).

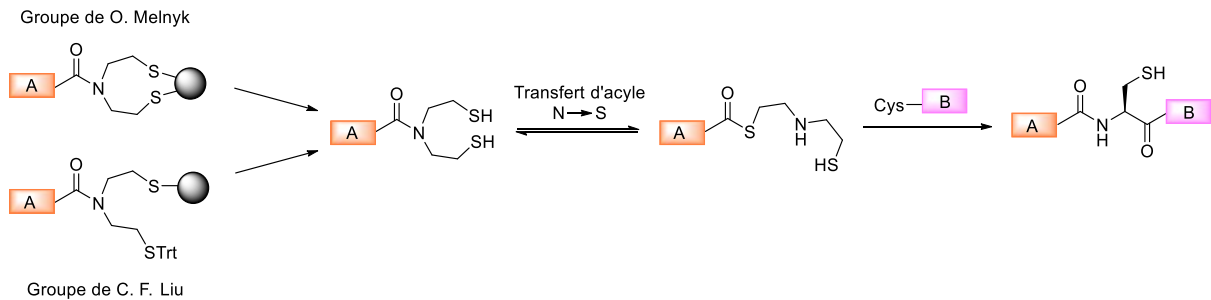


Figure 11 : Principe de la ligation SEA.

Le groupe d'O. Melnyk a démontré que le groupement *bis*-(2-sulfanyléthyl)amide peut se présenter sous deux formes, une première dite active que l'on notera SEA^{on} et une seconde dite inactive ou latente que l'on notera SEA^{off}. L'une des propriétés les plus intéressantes du groupement SEA^{on} est qu'il se présente sous la forme d'un dithiol qui peut être réversiblement inactivé par simple oxydation en disulfure cyclique. À l'inverse, la forme SEA^{off} peut être activée très facilement en utilisant des conditions réductrices. Cette propriété unique permet une autoprotection réversible du système (Figure 12A). De plus, le groupement SEA^{off} est très stable au cours des purifications par HPLC et lors du stockage.

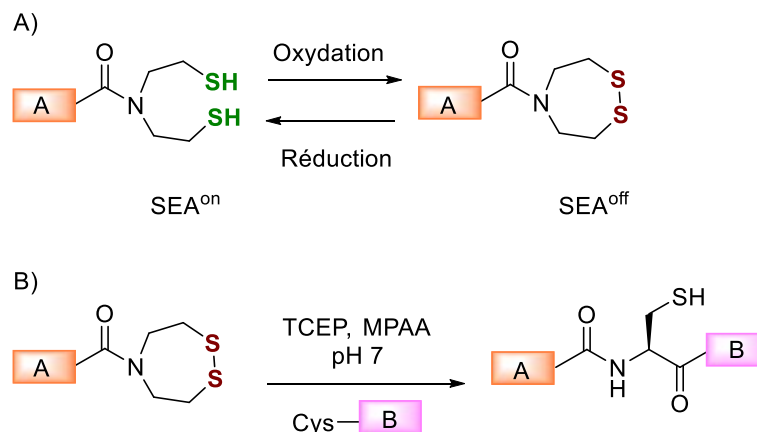


Figure 12 : A) Concept SEA^{on}/SEA^{off} B) Utilisation directe du SEA^{off} en NCL en présence de TCEP.

Par ailleurs, un peptide SEA^{off} peut être utilisé directement pour des procédés de ligation. En effet, il peut être activé *in situ* sous la forme SEA^{on} en utilisant un agent réducteur comme la TCEP ou

le DTT. Le groupement SEA^{on} se réarrange *in situ* par transfert d'acyle de N→S et réagit avec un cystéinyl peptide dans une étape de ligation native. Enfin, cette technique permet de travailler dans une gamme de pH assez large (pH 3-7,5) et donc de s'adapter à la solubilité des segments utilisés⁴⁴.

3.2.6 La ligation SEALide

Le groupement *N*-sulfanyléthylanilide (SEALide) a été décrit en 2009 par le groupe de A. Otaka⁴⁵. Initialement, ce motif a été utilisé pour produire des peptides thioester par échange du groupe SEALide par un thiol en solution. La transthioestérification était réalisée par transfert d'acyle N→S en milieu acide en présence d'un alkylthiol exogène. Cependant cette étape d'activation sous forme d'un thioester en milieu acide conduit à l'épimérisation partielle de l'acide aminé en C-terminal.

En 2011, Otaka et ses collaborateurs ont décrit l'utilisation directe des peptides SEALide dans la réaction de NCL⁴⁶. Cette réaction est catalysée par les ions hydrogénophosphate puisqu'en l'absence de sels de phosphate la ligation est très lente. Les auteurs ont proposé que la catalyse du transfert d'acyle N→S fasse intervenir une bicatalyse acide-base par l'ion hydrogénophosphate selon le mécanisme proposé ci-dessous (Figure 13).

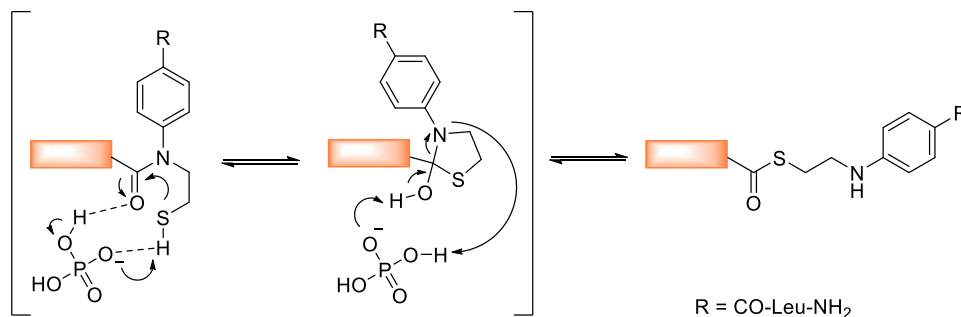


Figure 13 : Hypothèse de la catalyse du réarrangement du SEALide par les sels de phosphate.

4. Accès à de grandes protéines

La NCL et les méthodes apparentées permettent de réaliser l'assemblage de segments peptidiques préparés par SPPS. La ligation de deux segments peptidiques conduit à une protéine d'une certaine d'acide aminés maximum au regard des limitations actuelles de la SPPS (discuté en 2.1). L'obtention de protéines de tailles supérieures requière l'assemblage de plus de deux fragments. Deux stratégies sont possibles en fonction du sens d'élongation utilisé, de l'extrémité N terminale vers l'extrémité C terminale (N vers C) ou l'inverse (C vers N) (Figure 14).

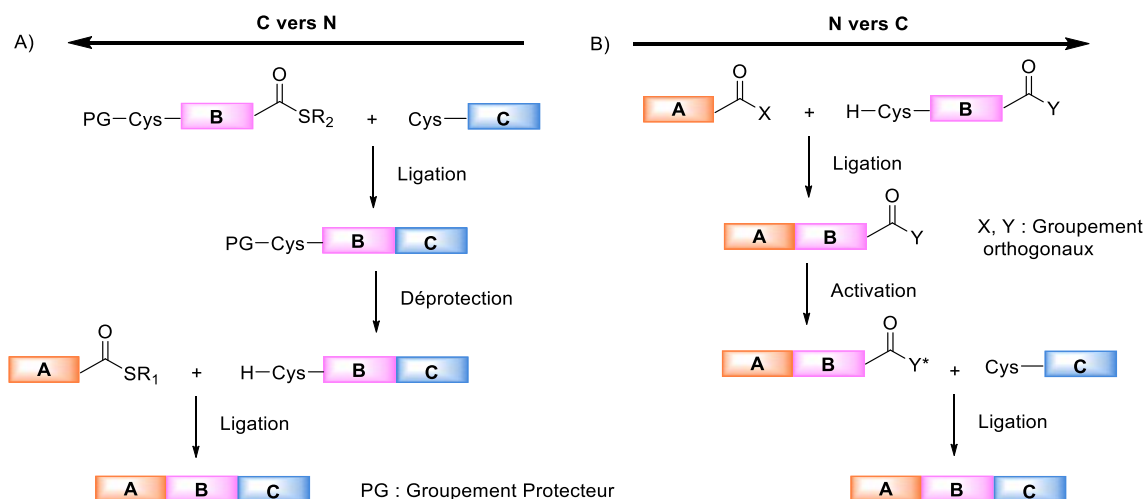


Figure 14 : Stratégie d'assemblage de C vers N (A) ou N vers C (B).

4.1 Assemblage de C vers N

4.1.1 Groupements protecteurs

L'assemblage de peptides de l'extrémité C vers N requière l'utilisation de groupements protecteurs temporaires pour la cystéine du coté N-terminal du fragment à incrémenter (segment B, Figure 14A). En effet, le masquage de la cystéine est nécessaire pour éviter des phénomènes d'oligomérisation ou de cyclisation. Ce groupement doit être stable dans les conditions de ligation et s'éliminer dans des conditions douces. Il peut se situer sur la fonction thiol de la chaîne latérale (Acm^{47,48}, Figure 15), sur la fonction amine (Msc^{49,50}, Figure 15) ou sur les deux (Thz^{51,52}, Figure 15). La Figure 15 présente un échantillon non exhaustif des groupements protecteurs utilisés (pour une revue récente, consulter la référence⁵³).

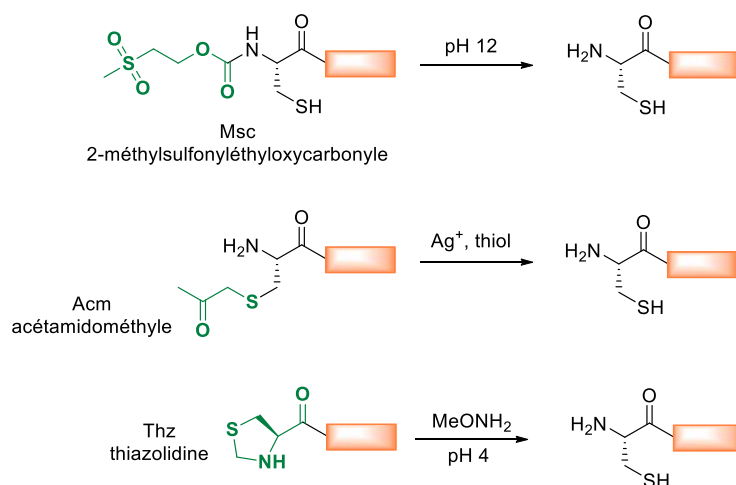


Figure 15 : Exemples de groupements protecteurs pour l'assemblage de C vers N.

À titre d'exemple, l'équipe de Wintermann a décrit en 2013 la synthèse d'une protéine de 286 acides aminés par assemblage séquentiel de 6 segments peptidiques de l'extrémité C vers N⁵¹. Ils ont utilisé pour cela la ligation chimique native et la protection thiazolidine (Thz) comme groupement protecteur des cystéines. Ce travail comprend 5 étapes de ligation et au global 9 étapes chimiques en comptant les étapes de déprotection des cystéines.

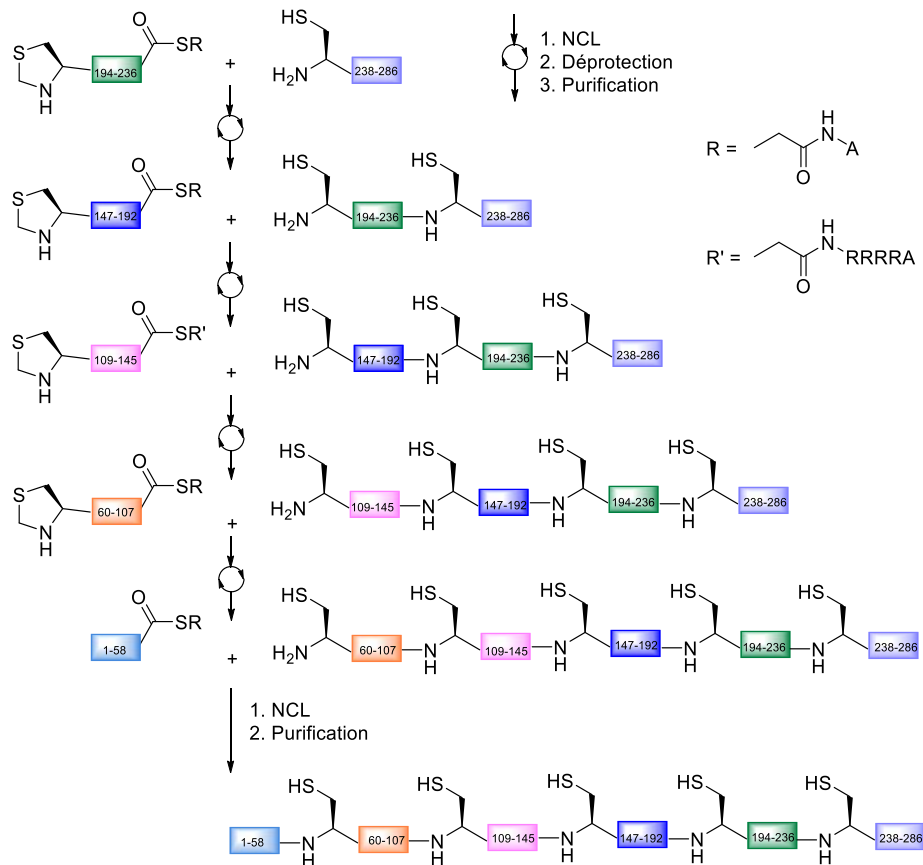


Figure 16 : Synthèse totale de la sous unité F-ATPase d'*Escherichia coli*.

Ici, la multiplication des étapes chimiques a conduit à une perte de rendement significative (Figure 17) au cours de l'élongation. Ce phénomène résulte en partie de la purification par chromatographie RP-HPLC des intermédiaires réactionnels, qui conduit à d'importantes pertes de matière. Cet exemple illustre l'importance de simplifier les stratégies d'assemblage, de réduire le nombre d'étapes chimiques et de réduire le nombre d'étapes de purification. Une solution pour éviter la multiplication des étapes chimiques et de purification est d'envisager des stratégies en phase solide ou des procédés *one-pot*.

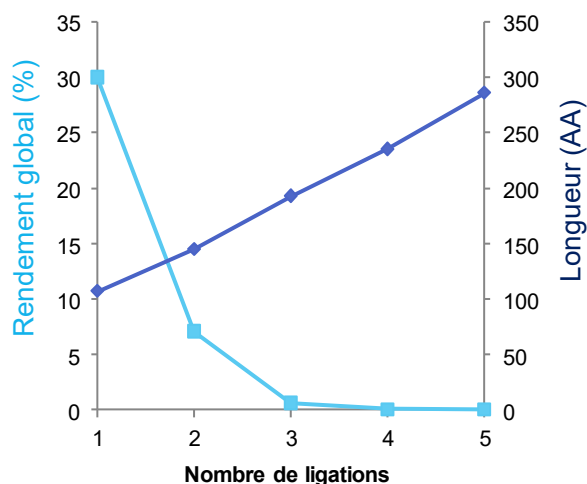


Figure 17 : Évolution du rendement en fonction du nombre d'étapes de ligation pour la synthèse de la sous unité F-ATPase d'*Escherichia coli*. Les données présentées sont issues de la référence⁵¹.

4.1.2 Les procédés *one pot* de C vers N

Les procédés *one pot* permettent de réaliser plusieurs réactions dans un même réacteur sans avoir à isoler les intermédiaires formés. Pour une élongation de C vers N, il est donc nécessaire que les conditions de déprotection des groupements protecteurs des cystéines soient compatibles avec les conditions de ligation mises en œuvre après ces étapes chimiques.

En 2004, Bang et ses collaborateurs ont décrit pour la première fois l'assemblage *one pot* de trois segments peptidiques en utilisant la thiazolidine comme groupement protecteur⁵². Dans cette stratégie, les segments B et C ont été assemblés en premier (étape 1, Figure 18). La thiazolidine a ensuite été déprotégée par ajout de *O*-méthyl-hydroxylamine à pH 4 (étape 2, Figure 18). Enfin, la seconde ligation a été déclenchée en ajustant le pH du milieu à 7 et en ajoutant le segment peptidique A (étape 3, Figure 18). Cette stratégie a permis la synthèse de plusieurs protéines^{52,54} et a également été étendue à l'assemblage de quatre segments peptidiques⁵⁵.

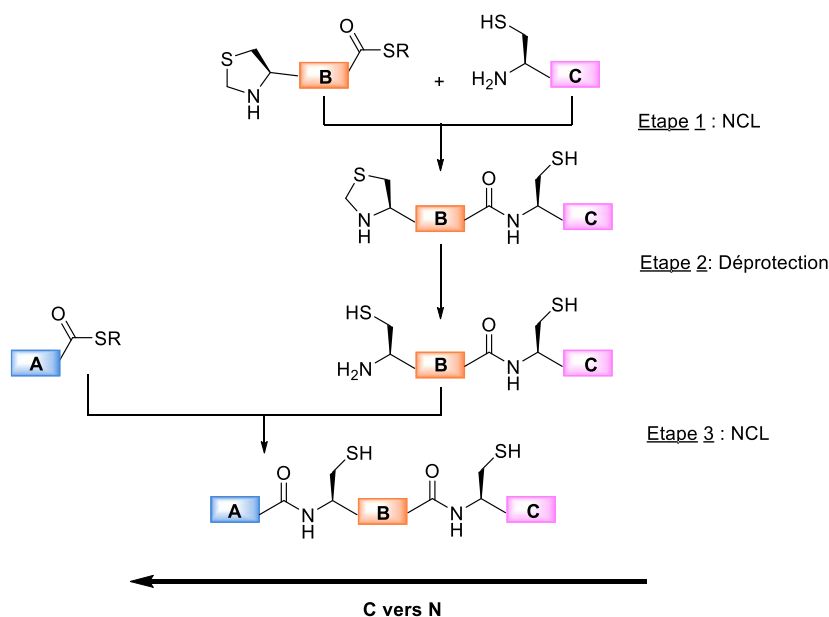


Figure 18 : Synthèse totale de la crambine en procédé *one-pot*⁵².

À la suite des travaux de Bang et de ses collaborateurs, de nombreux groupes se sont intéressés au développement de nouveaux groupements protecteurs pour des assemblages *one-pot*. Par exemple, des groupements protecteurs photo-labiles comme le Mapoc (4-(diméthylamino)phényloxyloxycarbonyl)⁵⁶ ou le Nvoc (6-nitroveratryloxycarbonyl)⁵⁷ ont été introduits. Ces groupements se déprotègent sous irradiation UV ($\lambda > 300$ nm) et ne nécessitent donc pas l'ajout de réactif. Parmi les nombreux outils disponibles aujourd'hui, on retrouve également le groupement acétoacétamide⁵⁸, la trifluoroacétyl thiazolidine⁵⁹ ou le trifluoroacétamidométhyle⁶⁰.

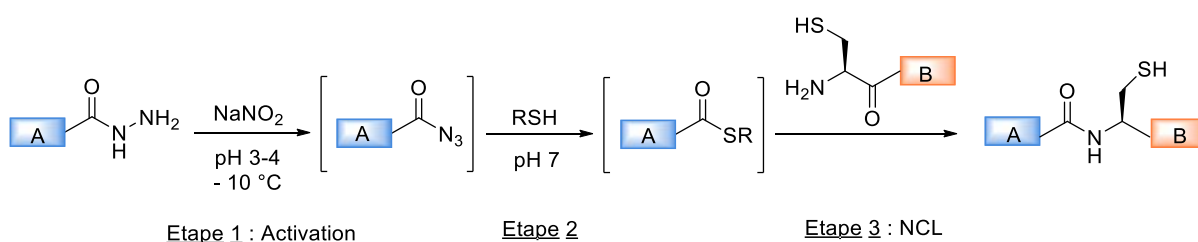
Le développement et l'optimisation de procédés *one pot* ont permis de faciliter la synthèse de nombreuses protéines de C vers N. Cependant, chaque protéine possède des propriétés uniques (solubilité des segments, jonctions difficiles, stabilité des thioesters). Il faut donc développer des outils permettant de s'adapter à ces particularités. Ce constat a stimulé le développement de stratégies d'assemblage de N vers C.

4.2 Assemblage N vers C

L'assemblage des peptides des extrémités N vers C requière l'utilisation de fonctions thioester latentes du côté C-terminal du fragment à incrémenter (segment B, Figure 14B). En effet, la fonction thioester doit être soit non réactive, soit très peu réactive pour permettre l'assemblage en évitant les phénomènes d'oligomérisation ou de cyclisation. Les conditions pour activer ou générer la fonction thioester doivent être douces et compatibles avec l'utilisation de peptides non protégés. Quelques précurseurs de thioesters ont ainsi été développés dans ce but au cours des dernières années.

4.2.1 Les peptides hydrazides

En 2011, L. Liu et ses collaborateurs ont introduit l'utilisation directe des peptides hydrazide en NCL⁶¹. La fonction hydrazide à l'extrémité N terminale peut être convertie en fonction thioester par nitrosation (étape 1, Figure 19) suivie d'une réaction de thiolysse (étape 2, Figure 19). Le peptide thioester n'est pas isolé et utilisé directement dans la réaction de NCL (étape 3, Figure 19). Bien que les conditions d'oxydation utilisées puissent entraîner des réactions secondaires⁶², les peptides hydrazide sont les précurseurs de thioesters les plus utilisés à l'heure actuelle. En effet, ils sont facilement accessibles par Fmoc-SPPS^{63,64} et les conditions d'oxydation sont relativement douces. Ils permettent également de masquer une fonction thioester et donc de concevoir des assemblages par ligations successives en activant le groupe hydrazide au moment voulu et de façon sélective⁶⁵.



R : groupement alkyle ou aryle

Figure 19 : Les peptides hydrazide précurseurs de thioesters pour la NCL.

En parallèle, l'introduction de nouveaux systèmes de réarrangement de type *O,S*,⁶⁶⁻⁶⁸ *N,S*⁶⁹ ou *N,Se*⁷⁰ a permis d'envisager la génération *in situ* de thioesters à partir de précurseurs produits par Fmoc-SPPS. Ces systèmes exploitent la capacité de β-mercapto-esters ou de β-mercapto-amides à se réarranger spontanément par migration d'acyle intramoléculaire (étape 1, Figure 20). L'intermédiaire transitoire formé peut ensuite réagir avec un thiol exogène pour former un thioester stable qui peut soit être purifié soit engagé directement dans une étape de ligation NCL (étape 2 et 3, Figure 20).

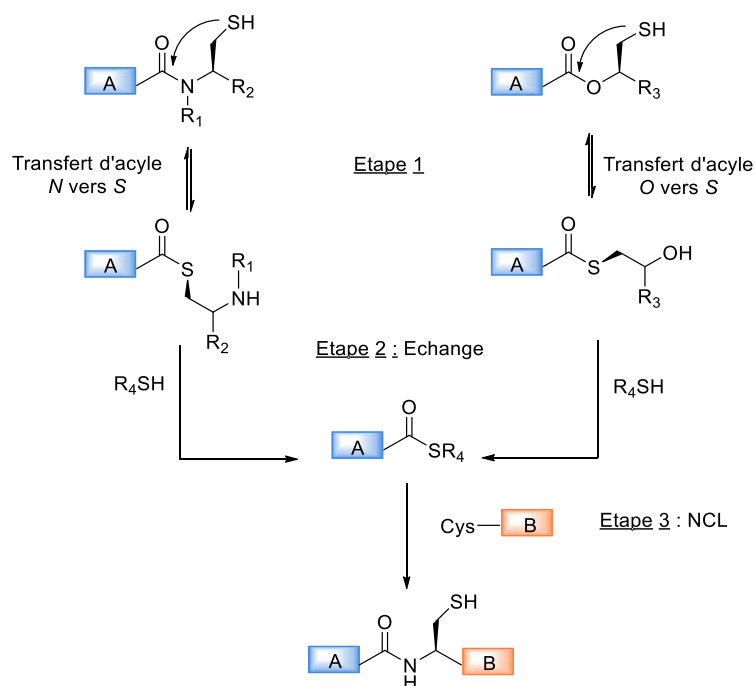


Figure 20 : Formation de thioesters par transfert d'acyle $N \rightarrow S$ et $O \rightarrow S$ et ligation de type NCL.

4.2.2 Les systèmes de transfert d'acyle de $O \rightarrow S$

Le transfert d'acyle O, S permet de générer un thioester *in situ* qui peut réagir directement dans les conditions de la NCL (étape 3, Figure 20). Le premier système de transfert d'acyle de l'oxygène vers le soufre a été introduit par l'équipe de S. J. Danishefsky en 2004⁶⁶ (Figure 21). Cependant, ce système doit être introduit après l'étape de SPPS car il n'est pas compatible avec les conditions d'élongation. Il est à noter que les esters de phénol ont l'inconvénient d'être sensibles à l'hydrolyse. Le groupe de Gaertner décrit la même année un système d'accroche générant un peptide portant un ester de 2-mercapto-carboxamide en C-terminal après coupure de la résine⁶⁷ (Figure 21). Une fois encore, les esters générés ne sont pas stables dans les conditions de NCL et subissent une réaction secondaire d'hydrolyse. Quelques années plus tard, L. Liu optimisa ce dernier système en substituant le groupement amide par un groupement plus électro-donneur permettant ainsi de limiter le phénomène d'hydrolyse⁶⁸ (Figure 21) qui malgré ces améliorations reste une limitation majeure de ces systèmes.

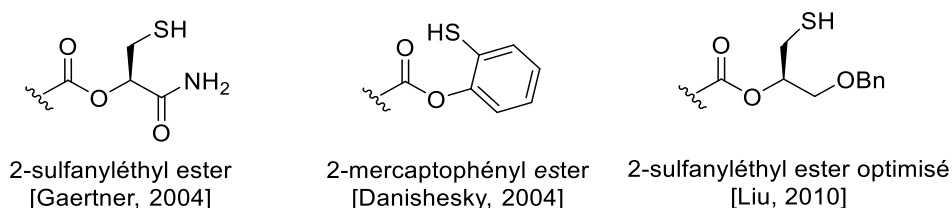


Figure 21 : Système de transfert d'acyle intramoléculaire de O vers S .

4.2.3 Les systèmes de transfert d'acyle de $N \rightarrow S$

Le développement de systèmes de transfert $N \rightarrow S$ a permis de concevoir des thioesters dits bloqués ou crypto-thioesters particulièrement performants et, contrairement aux systèmes de transfert $O \rightarrow S$, peu sensibles à l'hydrolyse. Il existe un certain nombre de systèmes de transfert d'acyle intramolécule de type $N \rightarrow S$, un échantillon non exhaustif est présenté dans la Figure 22 : *N*-alkyl cystéine⁷¹, ester de cystéinyl proline (CPE)^{72,73}, *N*-sulfanyléthylanilide (SEAlide)⁴⁵, bis-(2-sulfanyléthyl)amide (SEA)⁴², α -méthylcystéine⁷⁴ et *N*-hydroxynitrobenzyl cystéine⁷⁵. Plus récemment, des systèmes de $N \rightarrow Se$ ont été introduits pour permettre la formation de sélénoesters^{76,77}. Les sélénoesters sont de très bons donneurs d'acyle et présentent donc un fort potentiel pour la NCL.

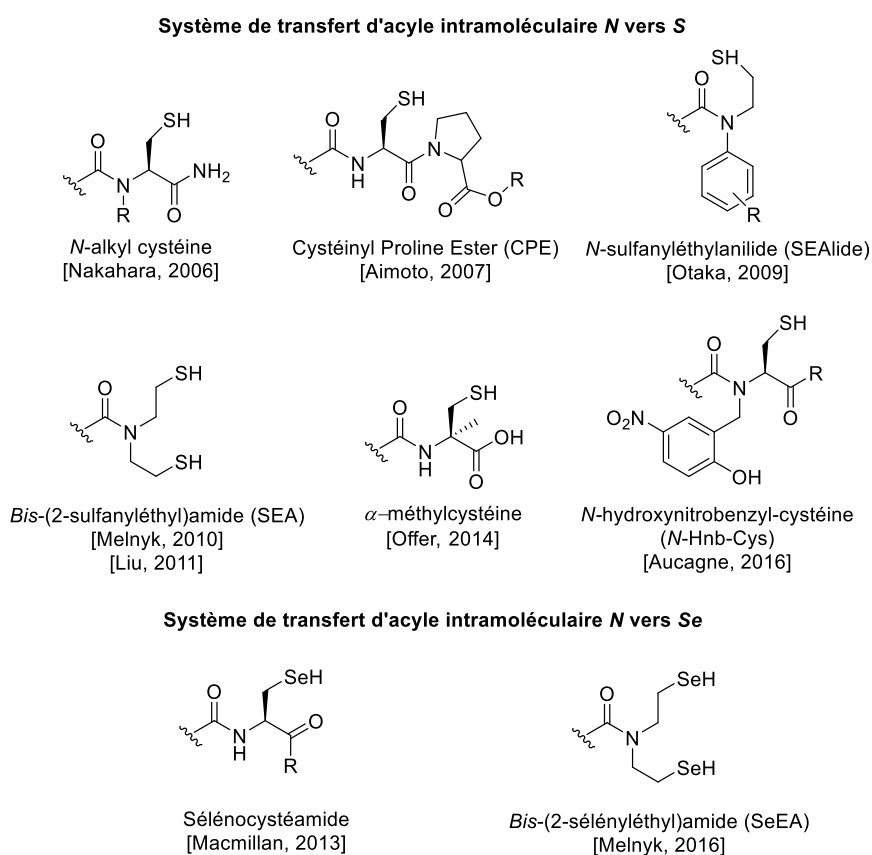


Figure 22 : Système de transfert d'acyle de type $N \rightarrow S$ et $N \rightarrow Se$.

Il est important de noter qu'en général l'utilisation de tels systèmes de transfert conduit à des cinétiques de ligation plus lentes qu'à partir d'un thioester classique en raison de l'énergie d'activation nécessaire pour le transfert d'acyle de l'azote vers le soufre (par exemple 31.1 kcal/mol pour la cystéine)⁷⁸. Cependant, ils sont nettement plus stables que les thioesters (hydrolyse) et plus faciles à produire par SPPS. L'utilisation de tels systèmes est donc en plein essor depuis les années 2010. Par ailleurs, certains de ces systèmes présentent des propriétés particulières permettant d'envisager

l'assemblage consécutifs de N vers C par réaction de NCL de segments peptidiques. C'est le cas du groupement SEAride de l'équipe d'Otaka et du groupement SEA du groupe de Melnyk.

4.2.4 Les procédés *one pot* de N vers C

a) La ligation cinétiquement contrôlée (KCL)

La ligation cinétiquement contrôlée (KCL) a été introduite en 2006 par le groupe de Kent⁷⁹. Cette méthode exploite la grande différence de réactivité entre les thioesters d'alkyle et les thioesters d'aryle en l'absence de catalyseur de type arylthiol (MPAA), pour réaliser un assemblage successif^{80,81}. En l'absence de catalyseur, le thioester d'aryle réagit plus rapidement que le thioester d'alkyle pour former le premier produit de ligation AB (étape 1, Figure 23). Ce dernier peut ensuite réagir avec un autre cystéinyl peptide après ajout d'un thiol exogène dans le milieu permettant d'activer le thioester d'alkyle par échange thiol-thioester (étapes 2 et 3, Figure 23).

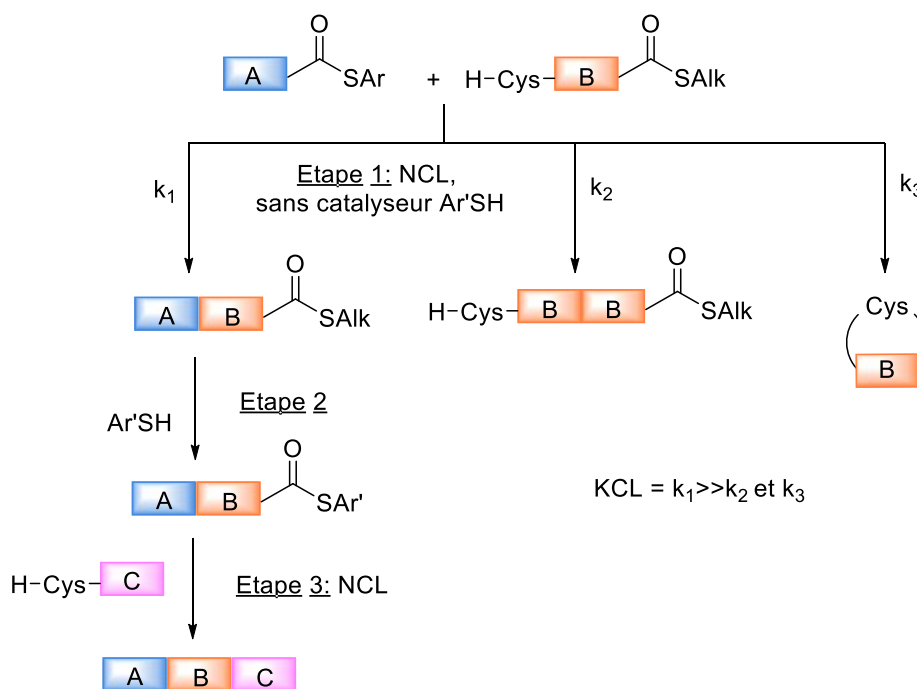
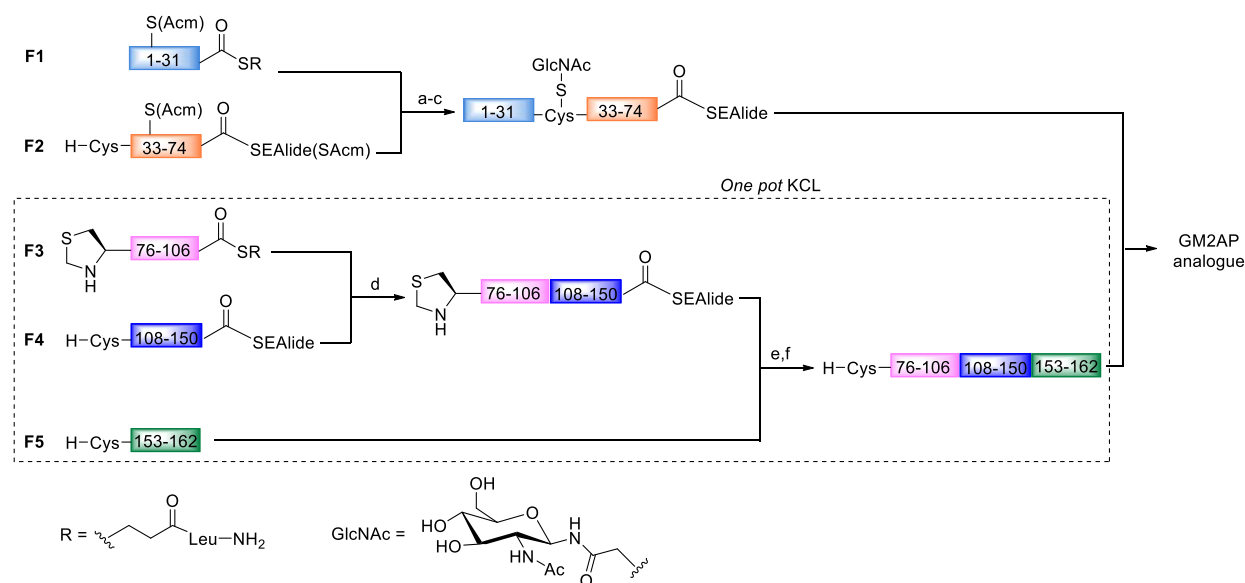


Figure 23 : Principe de la ligation chimique cinétiquement contrôlée KCL ($k_1 \gg k_2 \text{ et } k_3$).

Il est important de noter que l'efficacité de la KCL dépend de la nature des acides aminés en C-terminal qui portent les fonctions thioester. L'utilisation d'acides aminés très réactifs (glycine, histidine, cystéine) pour le segment alkylthioester (segment B, Figure 23) ou d'acides aminés très peu réactifs (valine, isoleucine, proline) pour le segment arylthioester (segment A, Figure 23) compliquent la stratégie d'assemblage par KCL. En règle générale, les KCL sont réalisées en utilisant des alkylthioester dérivés d'acides aminés encombrés en combinaison avec des arylthioesters dérivés d'acide aminés peu encombrés. Par exemple, Mandal et ses collaborateurs ont utilisé un arylthioester

porté par une alanine et un alkylthioester porté par une thréonine pour la synthèse d'un analogue de la crambine⁸².

La différence de réactivité entre les fonctions de type thioester peut être réalisée en jouant sur les conditions opératoires. À titre d'exemple, la ligation SEALide est catalysée par les ions hydrogénophosphate (décrit 3.2.6). En absence de phosphate le groupement SEALide n'est pas un thioester latent, puisqu'une réactivité résiduelle est toujours présente. La ligation est donc très lente mais possible. En jouant sur cette propriété, Otaka a pu envisager une ligation en conditions cinétiquement contrôlées (conditions d, Figure 24). Cette particularité du groupement SEALide lui a permis de réaliser la synthèse totale d'un analogue de l'activateur protéique GM2 (162 acides aminés) par assemblage convergent de 5 fragments peptidiques (Figure 24)⁸³. À noter que les ligations des fragments F3, F4 et F5 (Figure 24) ont été réalisées en one pot grâce à l'utilisation ou non de tampon phosphate en fonction de la nature du donneur d'acyle et de la sélectivité recherchée.



a) NCL. b) Glycosylation par réaction d'alkylation sur la cystéine. c) Elimination du groupement Ac. d) NCL en absence de sels de phosphate. e) NCL en présence de sels de phosphate. f) Elimination du groupement Thz. g) NCL en présence de sels de phosphate.

Figure 24 : Synthèse convergente d'un analogue de la GM2AP en utilisant le SEALide⁸³.

b) Le système SEA^{on}/SEA^{off}

Nous avons mentionné précédemment (section 3.2.5) que le groupement SEA^{off} est très stable au cours des purifications par HPLC, du stockage ou en solution en présence d'agents nucléophiles forts ou de réducteurs faibles tels que l'acide 4-mercaptophénylacétique (MPAA) à pH neutre. Il est important de noter que le MPAA est le catalyseur standard utilisé pour accélérer la NCL réalisée à partir de peptides alkylthioesters.

Grâce à ses propriétés particulières le groupement SEA^{off} peut être utilisé comme un thioester latent dans les conditions de NCL en absence d'agents réducteurs forts dans le milieu. Ceci permet d'assembler deux segments (étape 1, Figure 25) par NCL en gardant le SEA^{off} intact à l'extrémité C-terminale (segment B, Figure 25). Une fois que la ligation est terminée, le groupe SEA^{off} peut être activé par réduction en utilisant par exemple du DTT ou de la TCEP, déclenchant ainsi la seconde ligation avec le segment C (étape 3, Figure 25). En utilisant cette méthode, Melnyk et ses collaborateurs ont assemblé en *one pot* trois fragments les domaines K1⁸⁴ et N⁸⁵ du facteur de croissances des hépatocytes (HGF) ainsi que d'autres protéines.

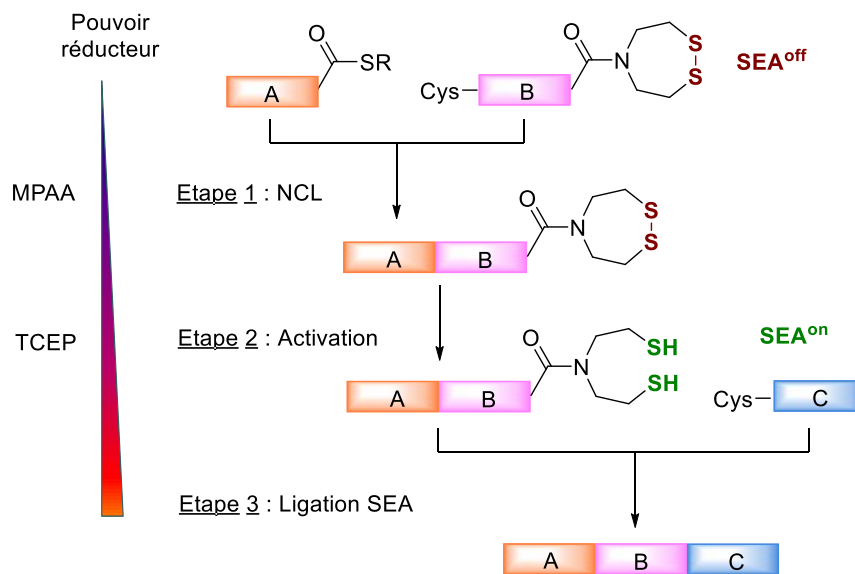


Figure 25 : Assemblage "one pot" de trois segments en utilisant le système SEA^{off}/SEA^{on}.

Les méthodes de ligation chimique et le développement de méthode d'élongation de C vers N ou de N vers C ont permis d'accroître la taille des protéines synthétisées par voie chimique à partir de segments produits par SPPS. Par exemple, l'assemblage de trois segments peptidiques en utilisant des techniques de ligation chimiosélectives permet d'accéder à une protéine allant jusqu'à 150 acides aminés, compte-tenu des limitations actuelles de la SPPS. En 2017, le groupe de L. Liu a décrit la synthèse totale d'un mutant de l'ADN polymérase virale P2 de *Sulfolobus solfataricus*⁸⁶. À l'heure actuelle, cette protéine de 352 acides aminés est la plus grande protéine synthétisée par voie chimique.

5. Conclusion/Bilan

La synthèse de peptides et protéines par voie chimique s'est imposée comme une alternative intéressante et complémentaire à l'utilisation de systèmes vivants. Elle permet un contrôle précis de la composition des protéines et constitue donc un outil puissant d'investigation des processus biologiques impliquant ces molécules. Deux avancées majeures ont révolutionné le domaine et

contribué à son succès : la synthèse peptidique en phase solide (SPPS) puis la ligation chimique native (NCL). Depuis son introduction, la NCL a fait l'objet de nombreuses améliorations. L'émergence des peptides hydrazides et des systèmes de transfert d'acyles de type $O \rightarrow S$, $N \rightarrow S$, $N \rightarrow Se$ a notamment facilité l'accès aux peptides thioester ou sélénioester. L'introduction de nouveaux groupements protecteurs de cystéine a stimulé le développement de stratégies *one pot*. Ces nouvelles méthodologies ont permis seules ou en combinaison avec la NCL d'étendre la portée des protéines accessibles par synthèse chimique. Cependant, des limitations importantes subsistent et la synthèse de protéines de haut poids moléculaire reste un défi.

Pour comprendre les difficultés rencontrées lors de la synthèse de protéines de grande taille, il est important d'appréhender les mécanismes mis en jeu. En effet, la NCL et les méthodes apparentées font intervenir un certain nombre de concepts fondamentaux et processus chimiques complexes. Nous proposons dans la suite de ce manuscrit un examen approfondi de ces méthodologies. Les aspects mécanistiques fondamentaux, les réactions secondaires ainsi que le rôle des différents additifs, solvants et tampons seront discutés.

CHAPITRE II : ASPECTS MÉCANISTIQUES DE LA LIGATION CHIMIQUE NATIVE

1. Mécanisme général

La NCL se définit par la réaction d'un peptide thioester avec un cystéinyl peptide. En général, une première étape d'échange thiol-thioester entre un peptide alkylthioester et un thiol exogène (souvent un aryle thiol) permet de former un nouveau peptide thioester plus réactif (étape 1, Figure 26). La seconde étape de capture par échange thiol-thioester conduit à la formation d'un intermédiaire thioester transitoire (étape 2, Figure 26). La chaîne acyle du peptide migre spontanément du soufre vers l'azote formant ainsi un intermédiaire à 5 chaînons qui conduit finalement à la formation d'un lien peptidique natif (étape 3, Figure 26). Les deux premières étapes d'échange thiol-thioester sont réversibles tandis que le réarrangement conduisant à la liaison amide est généralement irréversible dans les conditions de la réaction.

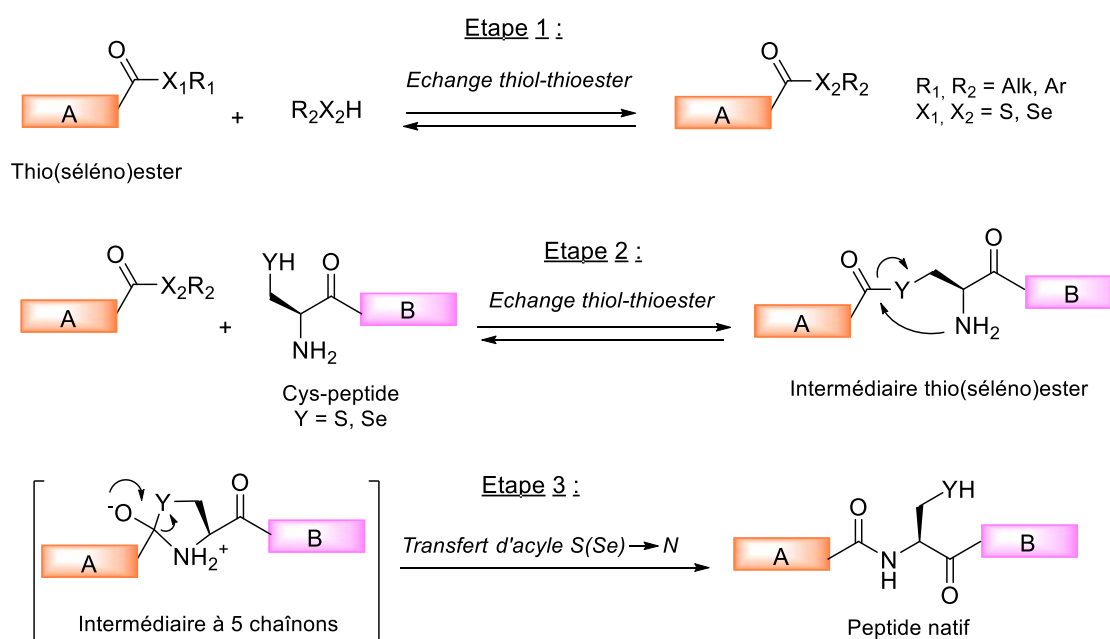


Figure 26 : Mécanisme détaillé de la réaction de NCL.

2. Dissection des étapes individuelles de la NCL

Les échanges thiol-thioester et le transfert d'acyle S→N sont les deux étapes clés de la NCL. Afin d'améliorer l'efficacité des réactions de NCL, il est important de comprendre les mécanismes mis en jeu dans ces réactions.

2.1 Échanges thiol-thioester

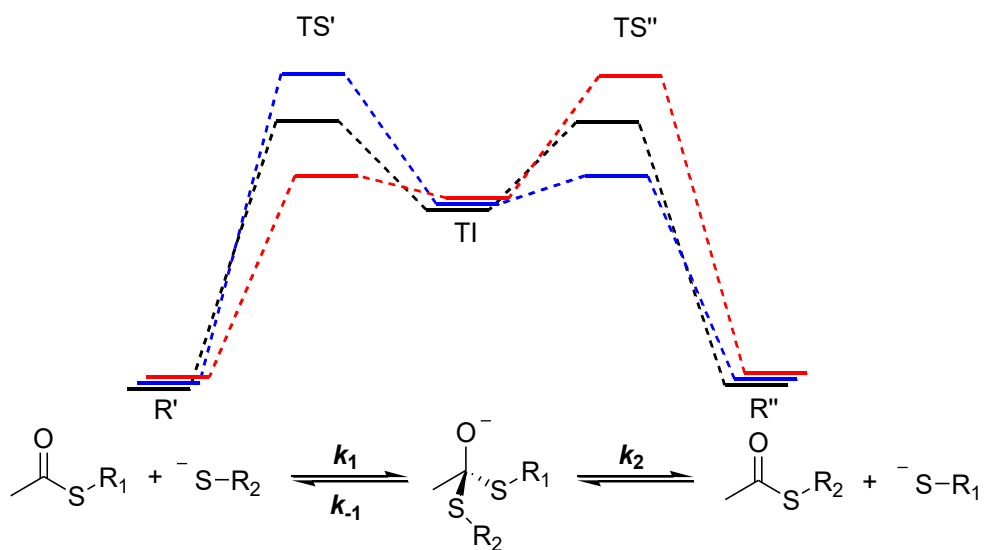
2.1.1 Généralité

L'échange thiol-thioester correspond à un transfert d'acyle entre deux thiols nucléophiles. Les aspects mécanistiques et cinétiques impliqués dans les réactions des thioesters avec les nucléophiles ont été le sujet de nombreux travaux⁸⁷⁻⁹⁰.

De façon générale, il est admis que les réactions de transferts d'acyle procèdent par étapes via la formation d'un intermédiaire tétraédrique. Ce procédé implique deux étapes distinctes, la première consiste en l'addition du nucléophile sur le carbonyle, la seconde consiste au départ du groupement partant. Dans certains cas, l'intermédiaire tétraédrique peut être très instable mais exister ou être tellement instable qu'il n'existe pas. Dans ces conditions, on parle de mécanisme concerté c'est-à-dire un processus réactionnel au cours duquel la rupture d'une liaison et la formation d'une autre sont plus ou moins simultanées. D'un point de vue thermodynamique, cela signifie qu'il n'y a pas formation d'un minimum local pour l'énergie et que la réaction a lieu en une seule étape. La stabilité de l'intermédiaire tétraédrique est donc un paramètre déterminant pour observer l'un ou l'autre de ces mécanismes. Ces aspects mécanistiques ont notamment été étudiés et décrits par E. Castro⁹⁰.

Hupe et Jencks ont étudié le transfert S→S d'un groupement acétyle en utilisant une relation linéaire d'énergie libre (LFER) de type Brønsted⁹¹. En d'autres termes, ils ont étudié la relation entre le logarithme de la constante de vitesse de la réaction (noté k) et la constante d'acidité K_a du groupement nucléophile ou du groupement partant. Les auteurs ont observé un profil biphasique avec deux portions linéaires pour ces réactions de transfert, l'une pour des pK_a faibles et l'autre pour des pK_a élevés. Ceci traduit selon eux le passage par un intermédiaire tétraédrique.

Dans un premier temps, ils ont étudié la réaction d'un thiolate et d'un thioester où le thiol nucléophile et le thiol partant sont identiques ($R_1=R_2$, Figure 27). Le profil thermodynamique de cette réaction est totalement symétrique, l'intermédiaire tétraédrique doit s'il existe conduire de façon équivalente aux réactifs et aux produits (cinétiques de réaction similaires, $k_{-1}=k_2$). Tout changement de nucléophile (R_2) ou groupement partant (R_1) induit une asymétrie dans le profil, l'une des étapes devient alors limitante.



if $R_1 = R_2$, then $k_{-1} = k_2$

if $R_1, R_2 = \text{alkyl, aryl with } pK_a(R_2SH) > pK_a(R_1SH)$, then $k_{-1} < k_2$

if $R_1, R_2 = \text{alkyl, aryl with } pK_a(R_2SH) < pK_a(R_1SH)$, then $k_{-1} > k_2$

Figure 27 : Profil réactionnel théorique pour l'échange thiol-thioester. En noir : profil symétrique. En bleu : l'étape 1 est cinétiquement limitante. En rouge : l'étape 2 est cinétiquement limitante. R' : réactifs, R'' : produits; TI : intermédiaire tétrahédrique; TS : état de transition.

Hupe et Jencks ont étudié la réaction nucléophile entre différents thiols et l'acétate de nitrothiophényle (PNTPA). La Figure 28B représente le logarithme des constantes de vitesses pour la réaction de différents alkyles et aryles thiols avec le PNTPA en fonction de leur pK_a . On observe un profil biphasique avec deux portions linéaires (courbes de tendance en pointillé). Pour les thiols avec un pK_a inférieur à celui du *p*-nitrothiophénol la pente de la courbe $\beta_2 = 1$, ce qui implique que le caractère nucléophile reflète complètement tout changement de basicité. Pour des thiols avec des pK_a supérieurs à celui du *p*-nitrothiophénol, on observe un changement de pente, $\beta_1 = 0,27$, l'effet du changement de basicité sur la nucléophilie est nettement diminué.

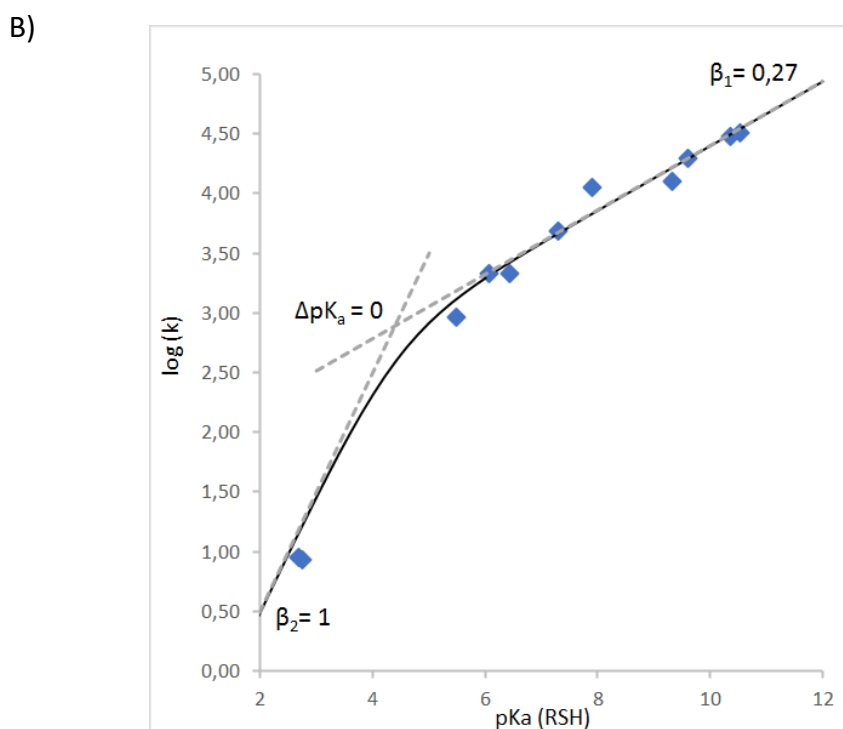
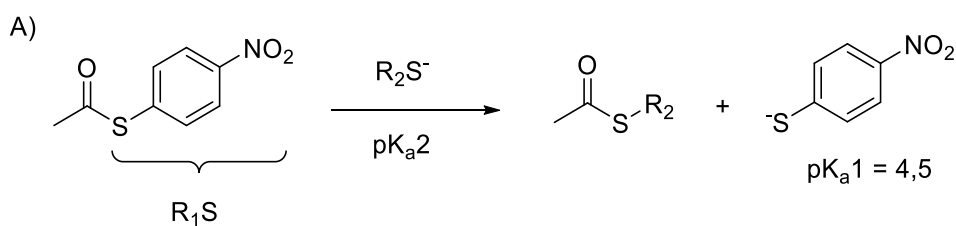


Figure 28 : A) Réaction d'échange thiol-thioester du PNTPA avec différents thiols. B) Logarithme de la constante de vitesse de la réaction d'échange en fonction du pK_a du thiol R_2SH . Le graphique est reproduit à partir des données de la référence Jencks 1977.

Par ailleurs, Hupe et Jencks ont déterminé l'équation (1) décrivant l'état d'équilibre de la réaction. Les équations (2) et (3) décrivent la cinétique de réaction observée lorsque l'attaque nucléophile et le départ du groupement partant sont respectivement les étapes cinétiquement limitantes. La courbe continue noire de la Figure 28 représente la courbe théorique décrite par l'équation 1 (avec $\beta_1 = 0,27$; $\log C_1 = 1,70$ et $\beta_2 = 1$; $\log C_2 = -1,5$). Les données expérimentales sont en accord avec la courbe calculée. De plus, la rupture de pente observée est nette ce qui implique bien un changement d'étape cinétiquement limitante et non un processus progressif.

$$(1) k_{\text{observé}} = k_1 \cdot \frac{k_2}{k_{-1} + k_2} = \frac{C_1 \cdot C_2 \cdot 10^{(\beta_1 + \beta_2) pK_a}}{C_1 \cdot 10^{\beta_1 pK_a} + C_2 \cdot 10^{\beta_2 pK_a}}$$

$$(2) k_{\text{observé}} = k_1 = C_1 \cdot 10^{\beta_1 pK_a}$$

$$(3) k_{\text{observé}} = \frac{k_1 \cdot k_2}{k_{-1}} = C_1 \cdot 10^{\beta_1 pK_a}$$

Quand le thiolate R_2S^- est plus basique et donc un moins bon groupement partant que le thiolate du thioester R_1S^- ($pK_{a2} > pK_{a1}$), l'intermédiaire se décompose préférentiellement en faveur du produit. L'attaque nucléophile est l'étape cinétiquement limitante ($k_2 > k_{-1}$). La cinétique de réaction est alors dépendante du $\Delta pK = pK_{a2} - pK_{a1}$, plus celui-ci est grand plus la réaction est rapide (courbe en pointillée avec $\beta = 0,27$, Figure 28). Ce type de mécanisme est sans doute observé dans les réactions de NCL entre des peptides de thioesters activés de types arylthioesters et le thiol du cystéinyl peptide (typiquement l'étape 2 de la Figure 26).

À l'inverse, lorsque R_2S^- est une base faible et un meilleur groupement partant ($pK_{a2} < pK_{a1}$), l'intermédiaire se décompose en faveur des réactifs de départ. L'étape cinétiquement limitante est l'expulsion du groupement partant ($k_2 < k_{-1}$). Dans les réactions de NCL, c'est le cas de la réaction des peptides alkylthioesters avec des additifs de type arylthiols qui sont fréquemment utilisés pour catalyser la réaction (typiquement l'étape 1 de la Figure 26). Finalement, Hupe et Jencks concluent que la différence de réactivité est corrélée à la différence entre le pK_a du nucléophile et celui du groupement partant et non pas à la nature (aryle ou alkyle) du thiol.

Plus récemment, Wang et ses collaborateurs ont réalisé une étude théorique du mécanisme de la NCL⁷⁸. Ils ont utilisé pour cela la théorie de la fonctionnelle de la densité électronique (*density functional theory* en anglais) qui est une méthode de calcul quantique permettant l'étude de la structure électronique. Les calculs ont été réalisés à 298,5 K au niveau B3LYP/6-31+G* de la théorie et le modèle polarisable continu (PCM) a été utilisé pour modéliser les effets de solvants (ici, l'eau). La réaction entre le méthylthioester de *N*-formyl glycine, le thiophénol et la cystéine a été utilisée comme modèle d'étude de l'étape 1 du mécanisme de la NCL (Figure 29 voir aussi Figure 26).

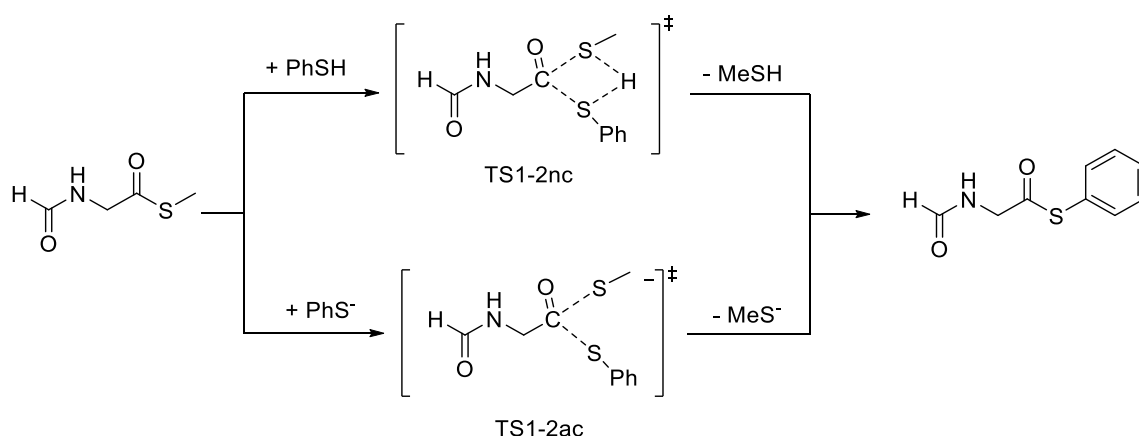


Figure 29 : Mécanisme hypothétique pour l'échange thiol-thioester entre le méthylthioester de *N*-formyl glycine et le thiophénol.

Dans leur hypothèse, les auteurs décrivent quatre voies possibles pour l'échange thiol-thioester : deux mécanismes par étapes (neutres ou anioniques) et deux mécanismes concertés (neutres ou anioniques). Wang et ses collaborateurs n'ont pas pu isoler d'état de transition (TS) compatibles avec un mécanisme par étape. Ils en ont conclu que le carbone du carbonyle ne tend pas à se lier avec les deux atomes de soufre simultanément à cause du grand rayon atomique de ces derniers. Selon les auteurs, l'échange thiol-thioester procède donc selon un mécanisme concerté pour lequel ils ont pu isoler un état de transition neutre (TS1-2nc) et un autre anionique (TS1-2ac). Ils ont également déterminé les barrières énergétiques de TS1-2nc et TS1-2ac, respectivement de 49,6 et 29,4 kcal.mol⁻¹. L'énergie d'activation nécessaire pour atteindre l'état de transition anionique étant plus faible et l'anion thiophénolate étant l'espèce majoritaire en solution à pH 7, les auteurs en ont conclu que l'échange thiol-thioester procède préférentiellement selon un mécanisme concerté anionique via TS1-2ac.

De façon intéressante, la structure du TS1-2ac présente des différences de longueur de liaison au carbonyle entre le groupement partant méthylthiolate ((C=O)-S : 2,845 Å) et le nucléophile thiophénolate ((C=O)-S : 1,863 Å). Cette observation est en corrélation avec la différence de pK_a entre les deux acides conjugués correspondants et est également cohérente avec les expériences réalisées par Hupe et Jencks, bien que ceux-ci décrivent un mécanisme par étape⁹¹. En 2015, Sun et ses collaborateurs ont également observé cette différence de liaison en utilisant un tout autre système de calcul computationnel (MX06-2X/6-31+G*) en phase gaz⁹². Il apparaît donc clairement que le mécanisme et la cinétique de ligation dépend de la nature des groupements thiols mis en jeu.

Wang et ses collaborateurs ont également modélisé l'échange entre le thiophénylester de *N*-formyl glycine et la cystéine, ce qui correspond fondamentalement à l'étape de transthioestérification de la NCL (ou typiquement l'étape 2 de la Figure 26)⁷⁸. Les auteurs ont décrit les mêmes résultats que pour l'échange entre le méthylthioester de *N*-formyl glycine et le thiophénol (décrit dans la section précédente), à savoir que la transthioestérification suit un mécanisme concerté anionique. Il est important de noter que l'énergie d'activation pour l'échange avec la cystéine (24,9 kcal.mol⁻¹) est plus faible que pour l'échange avec le thiophénol (29,4 kcal.mol⁻¹), ce qui implique que le premier échange (étape 1, Figure 26) est l'étape cinétiquement limitante.

2.1.2 Catalyse

a) Catalyse par des arylthiols

Plusieurs types de peptides thioester ont été utilisés dans des réactions de NCL. Le thioester de benzyle est le premier décrit par le groupe de Kent en 1994³³. De façon plus générale, les peptides

alkylthioester sont souvent utilisés comme réactifs de départ pour les réactions de NCL. En effet, ils sont facilement accessibles et sont relativement stables à l'hydrolyse en comparaison à des thioesters présentant de meilleurs groupements partants comme les arylthioesters. Cependant, l'utilisation d'alkylthioesters dans des réactions d'échange thiols-thioesters conduit à des cinétiques plus lentes que les arylthioesters⁹³. Les aspects mécanistiques évoqués précédemment indiquent qu'il est donc nécessaire de trouver des catalyseurs efficaces qui permettent à la fois de produire des thioesters très réactifs comme les arylthioesters pour la réaction de transthioestérification mais également qui permettent un échange thiol-thioester avec l'alkylthioester initial efficace (ΔpK_a élevé).

Dès l'introduction de la NCL, Kent et ses collaborateurs ont anticipé que l'addition d'un arylthiol comme le thiophénol permettrait de garder les cystéines sous formes réduites mais surtout de convertir le thioester de benzyle en thioester de phényle bien plus réactif pour la réaction d'échange avec la cystéine. En 1996, ils utilisèrent ce concept pour réaliser la synthèse d'un analogue d'une ribonucléase bactérienne, la barnase, composée de 110 acides aminés⁸¹. Le benzylmercaptan et le thiophénol ont été comparés ce qui a permis de montrer un net avantage cinétique pour le thiophénol comme additif exogène.

Depuis, plusieurs arylthiols ont été décrits pour catalyser la réaction de NCL grâce à la formation *in situ* d'espèces arylthioester beaucoup plus réactives. Ces additifs sont présentés dans la Figure 30. Les arylthiols sont de modestes nucléophiles, ils sont donc utilisés en général en large excès durant les étapes de ligations (1-2 % volume pour le thiophénol, 100-200 mM pour le MPAA, entrées 1 et 2 de la Figure 30). Dans la plupart des cas, l'arylthioester produit par échange thiol-thioester n'est pas observé puisque le produit d'échange réagit très rapidement une fois formé. Ce constat est en accord avec toutes les études expérimentales et théoriques montrant que la réaction d'échange thiol-thioester entre le peptide alkylthioester et l'arylthiol exogène est l'étape cinétiquement limitante.

Le premier arylthiol à avoir été utilisé pour catalyser la réaction de NCL est le thiophénol (entrée 1, Figure 30)⁸¹. Comme mentionné précédemment, il a été utilisé pour la première fois en 1996 par le groupe de Kent. Par la suite, le thiophénol a été beaucoup utilisé en combinaison avec le benzylmercaptan ou l'éthanethiol qui sont des thiols plus réducteurs (entrées 3 et 4, Figure 31). Le thiophénol est compatible avec le repliement de certaines protéines qui nécessitent la présence de ponts disulfures. Par exemple, Bang et Kent ont décrit la synthèse totale de la crambine, qui est stabilisée par trois ponts disulfures, en utilisant la NCL suivie de l'étape de repliement en *one-pot*⁵². L'étape de ligation a été réalisée dans 6 M Gdn.HCl 0,1 M tampon phosphate à pH 7 en présence de thiophénol (1 % volume). Puis, l'étape de repliement a été déclenchée par dilution du milieu réactionnel avec du tampon Tris (0,1 M, pH final de 8) contenant 8 mM Cys.HCl/1 mM cystine.

Cependant, le thiophénol est très toxique, présente une odeur forte et n'est que très partiellement soluble dans l'eau. En 2006, Johnson et Kent réalisèrent une étude sur plusieurs additifs thiol (aryle et alkyle) pour la catalyse de la réaction de NCL⁸⁰. Dans ces travaux, l'acide 4-mercaptophénylacétique (MPAA) apparaît comme une bonne alternative à l'utilisation de thiophénol. Le MPAA est très soluble dans l'eau (à pH > 5,5), n'a pas l'odeur désagréable du thiophénol et est commercialement disponible. C'est également un excellent groupement partant pour la réaction de transthioestérification avec la cystéine et la cinétique d'échange thiol-thioester avec l'alkylthioester est similaire en principe à celle obtenue avec le thiophénol. Aujourd'hui, le MPAA est le catalyseur le plus utilisé pour réaliser les réactions de NCL et est considéré comme le catalyseur de référence. Le MPAA est un faible réducteur, il est donc souvent utilisé en combinaison avec la TCEP pour maintenir les cystéines sous forme réduite. L'un des désavantages du MPAA est qu'il complique les étapes de purifications par HPLC lorsque la protéine cible élue à un temps de rétention proche de celui-ci. Il est donc préférable de l'éliminer du milieu réactionnel avant purification. Ceci est réalisé par acidification du milieu (pH < 5,5), ce qui entraîne la précipitation du MPAA qui est ensuite extrait par de l'éther diéthylique^{17,42,94}. Ce protocole n'est cependant pas applicable à toutes les protéines, certaines ne sont pas stables en milieu acide.

Très récemment, le sulfonate de 3-mercaptobenzyle (3-MBSA, entrée 3, Figure 30) a été introduit comme nouveau catalyseur de la NCL pour surpasser les limitations du MPAA⁹⁵. En effet, ce composé est plus hydrophile et élue plus tôt en HPLC que le MPAA. Cette propriété a permis à Cowper et ses collaborateurs d'isoler directement après ligation un analogue du glucagon-like peptide (GLP-1) en présence de 0,1 M 3-MBSA. Cependant, c'est un moins bon catalyseur que le MPAA et sa synthèse est longue (5 étapes).

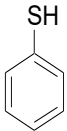
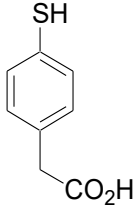
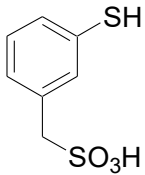
Entrée	Structure	Nom	pK _a	Propriétés	Référence
1		Thiophénol ⁹⁶	6,52 ⁹⁷ 6,2 ⁹⁸ 6,43 ⁹⁹ 6,62 ¹⁰⁰	Partiellement soluble dans l'eau (470 mg/L at 15 °C ¹⁰¹). Odorant. Souvent utilisé en combinaison avec le benzylmercaptan.	NCL ⁸¹ EPL ¹⁰²
2		MPAA	6,6 ¹⁰³	Soluble dans l'eau jusqu'à pH 5,5. Non odorant.	80
3		3-MBSA	Non référencé	Soluble dans l'eau et significativement plus hydrophile que le MPAA. Moins efficace que le MPAA comme catalyseur.	95

Figure 30 : Les arylthiols utilisés pour catalyser la réaction de NCL.

b) Catalyse par des alkylthiols

Le benzylmercaptan ou benzylthiol a été utilisé dès l'introduction de la NCL pour ses propriétés réductrices³³ (entrée 3, Figure 31). L'ajout de benzylthiol permet de diminuer la formation de ponts disulfures entre les cystéines. Le pK_a du benzylmercaptan est très élevé, il est donc peu probable que celui-ci participe à la catalyse de la réaction de NCL. C'est pourquoi, il a souvent été utilisé en combinaison avec le thiophénol^{81,104}. Comme il a été mentionné précédemment le thiophénol catalyse la réaction de NCL en formant un intermédiaire très réactif mais ce n'est pas un bon agent réducteur (le benzylmercaptan est plus réducteur que le thiophénol¹⁰⁵). L'éthanethiol a été utilisé comme alternative au benzylthiol pour la synthèse de protéines hydrophobes (entrée 4, Figure 31)¹⁰⁶.

Le 2-mercaptoéthane sulfonate MESNa est un très bon nucléophile mais un modeste catalyseur pour la réaction de NCL. En effet, pour des alkylthiols comme le MESNa et le benzylthiol la réaction d'échange thiol-thioester n'est pas limitante (ΔpK_a élevé). Ils entraînent des échanges rapides avec le peptide thioester initial mais comme ce sont de mauvais groupements partants, la réaction de transthioestérification devient limitante. En d'autres termes, les alkylthioesters formés par échange avec le MESNa ou le benzylthiol sont beaucoup moins réactifs que ceux formés avec le thiophénol ou le MPAA, par exemple. Néanmoins, le MESNa et l'un de ses analogues l'acide 3-

mercaptopropiosulfonique (MPS) ont été utilisés pour la synthèse de plusieurs protéines d'intérêt biologique^{102,107,108}. Par exemple en 2004, l'équipe de Gentle a décrit la réaction entre un thioester de 2-mercaptoéthane-sulfonate et un cystéinyl peptide dérivé d'une protéine homologue à un domaine SH2 (pour l'anglais *Scr homology region 2*)¹⁰⁷. La ligation a été réalisée en tampon phosphate (pH 7,5) et en présence de MESNa (~10 mM).

Récemment, l'utilisation d'alkylthiols comme catalyseurs de la réaction de NCL a connu un regain d'intérêt avec l'introduction des procédés de désulfuration (discuté au chapitre III section 2). En effet, les alkylthiols sont de modestes inhibiteurs de la formation de radicaux thiyls comparés aux arylthiols. Par conséquent, l'utilisation d'alkylthiols pour catalyser la réaction de NCL permet d'envisager des procédés de ligation-désulfuration *one-pot* et ainsi de simplifier l'accès à certaines protéines.

En 2014, l'équipe de Paynes a décrit un procédé de ligation-désulfuration basé sur l'utilisation du trifluoro-éthanethiol TFET (entrée 6, Figure 31)¹⁰⁹. Hupe et Jencks avaient d'ores et déjà notifié que le TFET est un meilleur nucléophile que le thiophénol pour la réaction d'échange avec l'acétate de nitrothiophényle PNTPA (constante de vitesse à 25 °C, TFET $4,9 \cdot 10^3 \text{ M}^{-1} \text{ min}^{-1}$, thiophénol $2,16 \cdot 10^3 \text{ M}^{-1} \text{ min}^{-1}$)⁹¹. En effet, le pK_a du TFET (7,30) est légèrement supérieur à celui du thiophénol (6,43). Concernant la réaction de NCL, Paynes et ses collaborateurs ont montré que le TFET est un catalyseur légèrement moins efficace que le MPAA à concentrations identiques. Cependant, le caractère très volatile du TFET permet de le retirer facilement du milieu réactionnel et donc de déclencher *in situ* une réaction de désulfuration. La désulfuration s'effectue par ajout de TCEP, de glutathion et d'un précurseur de radicaux comme le VA-044.

Plus récemment, Huang et ses collaborateurs ont introduit le méthylthioglycolate (MTG) comme catalyseur de la NCL (entrée 5, Figure 31)⁵⁹. La cinétique de ligation en présence de MTG est plus lente qu'avec le MPAA (~1,8 fois) mais l'utilisation de ce composé permet également de réaliser des procédés de ligation-désulfuration *one-pot*. Le MTG possède un point d'ébullition beaucoup plus élevé ($T_{eb} = 152 \text{ °C}$) que le TFET ce qui permet de travailler dans une gamme de température plus large. Initialement, ce composé a été utilisé pour catalyser la ligation SEA à pH 5, tandis que le MESNa ne présentait pas de résultats satisfaisants au regard de la faible réactivité des thioesters de MES formés *in situ*¹¹⁰.

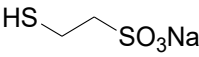
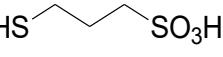
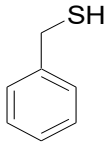
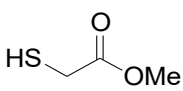
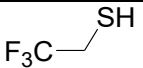
Entrée	Structure	Nom	pK _a	Propriétés	Référence
1		MESNa	9,53 (méthode de titrage) 9,08 (méthode spectrophotométrique)	Non odorant.	EPL ¹⁰² NCL ¹¹¹
2		MPS	Non référencé		108
3		Benzylthiol	9,43 ⁹⁷	Odorant. En majorité insoluble dans l'eau. Utilisé en combinaison avec le thiophénol.	33,81,104
4	EtSH	Éthanethiol	10,5	Odorant. Utilisé en combinaison avec le thiophénol.	106
5		MTG	7,9 ⁹⁹	Bonne solubilité dans l'eau ¹¹² . Faible temps de rétention en RP-HPLC.	NCL ⁵⁹ Ligation SEA ¹¹⁰
6		TFET	7,3 ¹¹³	T _{eb} = 35-37 °C	109

Figure 31 : Les alkylthiols utilisés pour la catalyse de la réaction de NCL.

2.2 Transfert d'acyle S→N intramoléculaire

Le réarrangement de l'intermédiaire thioester par transfert d'acyle du soufre vers l'azote (S→N) est l'étape de la réaction de NCL la moins discutée dans la littérature (étape 3, Figure 26).

Dans les années 1950, le réarrangement de la cystéamine ou de dérivés de la cystéine ont fait l'objet de plusieurs études^{29,114-116}. Wieland et ses collaborateurs ont montré que la S-glycylcystéamine Gly-SCH₂CH₂NH₂ se réarrange complètement en 2 min à pH 5,2. Trudelle et Caille ont également montré que la S-acétylcystéine se réarrange par migration d'acyle S→N jusqu'à pH 3¹¹⁷.

Barnett et Jencks ont étudié le transfert d'acyle du soufre vers l'azote $S \rightarrow N$ à travers l'hydrolyse de la thiazolidine T^{118} (Figure 32). Cette réaction conduit en conditions acides à la formation d'un intermédiaire (I) qui peut se réarranger soit en *S*-acétyl-mercaptoéthylamine (S) soit en *N*-acétyl-mercaptoéthylamine (N). Les auteurs ont montré que la réaction implique un changement d'étape cinétiquement limitante en fonction du pH. L'attaque de l'eau est cinétiquement limitante au-dessus de pH 3. En dessous de pH 2, la rupture de l'intermédiaire tétraédrique devient limitante et conduit à des proportions équivalentes de *S*-acétyl-mercaptoéthylamine (S) et de *N*-acétyl-mercaptoéthylamine (N). Barnett et Jencks ont montré qu'entre pH 2 et pH 4, la proportion d'espèce thioester S résultant de l'hydrolyse de la thiazolidine décroît de 0,37 à 0,01. Cependant, dans cette gamme de pH le taux d'hydrolyse varie très peu. Pour expliquer ce phénomène, les auteurs ont décrit un mécanisme impliquant la formation d'intermédiaires I^{+} et I^+ (Figure 32A). Ils postulent qu'au-dessus de pH 2, la protonation de I^{+} en I^+ est l'étape limitante du processus d'aminolyse du thioester S en amide N.

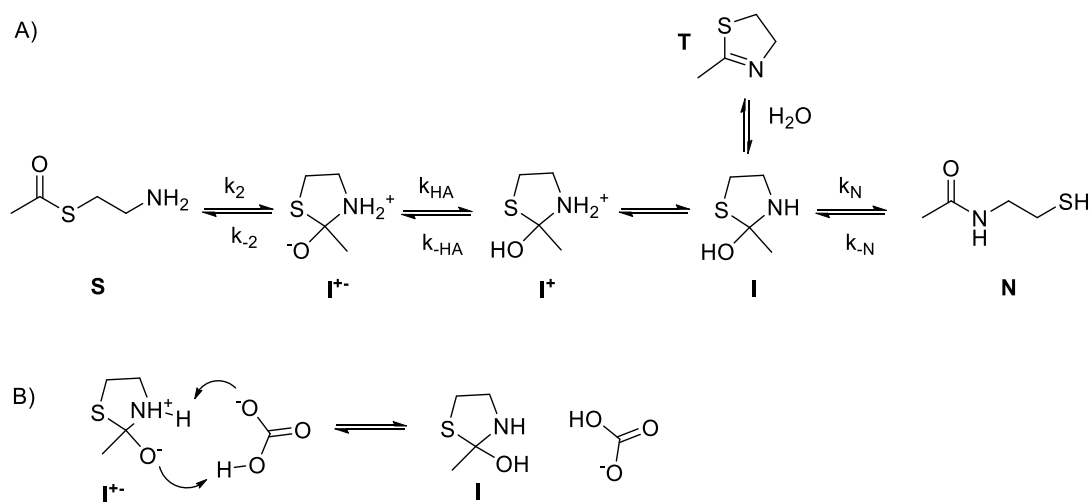


Figure 32 : Mécanisme proposé par Barnett et Jencks pour la formation de la *S*-acétyl-mercaptoéthylamine (S) et de la *N*-acétyl-mercaptoéthylamine (N) à partir de l'intermédiaire I produit par hydrolyse acide de la thiazolidine (T).

Les auteurs ont étudié la catalyse acide générale de la réaction de transfert d'acyle de l'espèce (S) vers l'espèce (N). La Figure 33 présente la relation entre le logarithme de la constante de vitesse de la réaction de transfert de proton et le pK_a de différents acides. La courbe en pointillés rouges correspond à un régime où la pente de la courbe α est égale à zéro c'est-à-dire que la protonation de I^{+} est contrôlée par la diffusion et donc que la vitesse de transfert de proton est indépendante de la force de l'acide. L'ion H_3O^+ se situe au-dessus de cette ligne, probablement parce que ce transfert de proton est assisté par les molécules d'eau adjacentes ce qui induit un mécanisme par « saut de protons » (« *proton-jump* » d'après Barnett et Jencks). Pour des acides plus faibles ($pK_a > 7$), le transfert de proton devient thermodynamiquement non favorable et la vitesse de ce transfert devient

proportionnel à la force de l'acide (courbe en pointillés bleus, Figure 33). On remarque que l'eau et le bicarbonate sont en dehors de la courbe de prédiction (courbe en pointillés bleus, Figure 33) et que la vitesse de réaction est plus rapide que prévue par la courbe de Brønsted. En effet, ces deux acides induisent un transfert de proton par catalyse bifonctionnelle qui implique un mécanisme concerté intramoléculaire (Figure 32B).

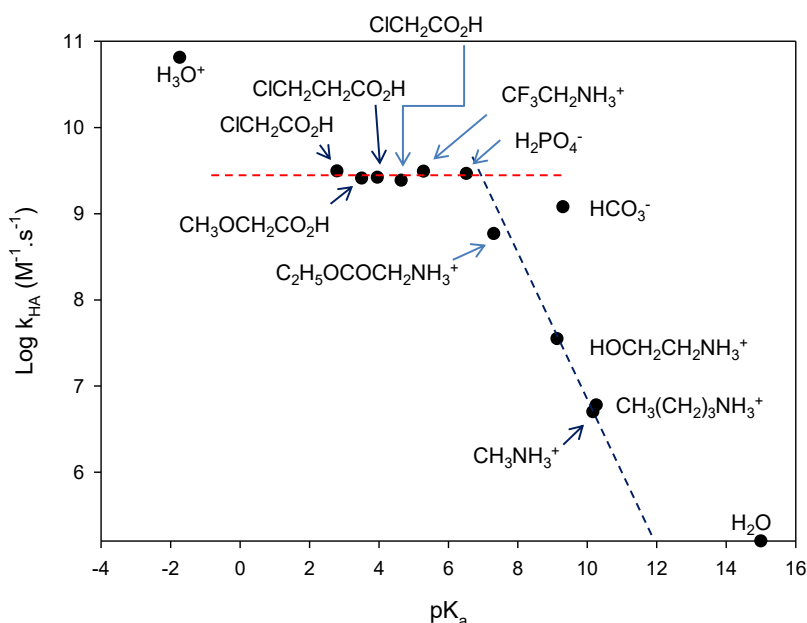


Figure 33 : Logarithme de la constante de vitesse de la réaction de transfert d'acyle de l'espèce (S) vers l'espèce (N) en fonction du pK_a de divers acides. La courbe en pointillés rouges correspond à α est égale à zéro tandis que la courbe en pointillés bleus correspond à $\alpha \sim 1$. Le graphique est reproduit à partir des données de la référence Barnett, 1969¹¹⁸.

Plus récemment, Wang et ses collaborateurs ont réalisé une étude théorique du transfert d'acyle de S vers N dans le cadre de leur étude du mécanisme de la NCL⁷⁸. Les calculs ont été réalisés en utilisant le niveau B3LYP/6-31+G* de la théorie et le modèle polarisable continu (PCM) pour modéliser les effets de solvants (ici, l'eau). Ils ont montré que le réarrangement intramoléculaire suivait un mécanisme par étape comme décrit dans la Figure 34. Dans un premier temps, l'amine attaque le carbonyle pour former un intermédiaire cyclique à 5 chaînons I_1^+ via l'état de transition TS1. Puis, la liaison carbone-soufre se rompt pour former le lien amide via l'état de transition TS2. Les auteurs ont calculé une énergie d'activation relative de 12,7 kcal.mol⁻¹ pour TS1 par rapport au réactif de départ S. La rupture de la liaison C-S (TS2) implique une énergie d'activation relative de 7,1 kcal.mol⁻¹ par rapport à I_1^+ . La formation du produit final N implique des énergies d'activations significativement plus faibles que pour les autres étapes de la NCL. Les auteurs en déduisent que le transfert d'acyle intramoléculaire de S→N n'est pas l'étape significativement limitante de la réaction de NCL. Au global, le réarrangement est exogernique de 7,4 kcal.mol⁻¹. Par ailleurs, l'énergie

d'activation calculée pour le réarrangement inverse de l'azote vers le soufre ($N \rightarrow S$) est de 31,1 kcal.mol⁻¹, ce qui a amené les auteurs à conclure que le réarrangement est irréversible.

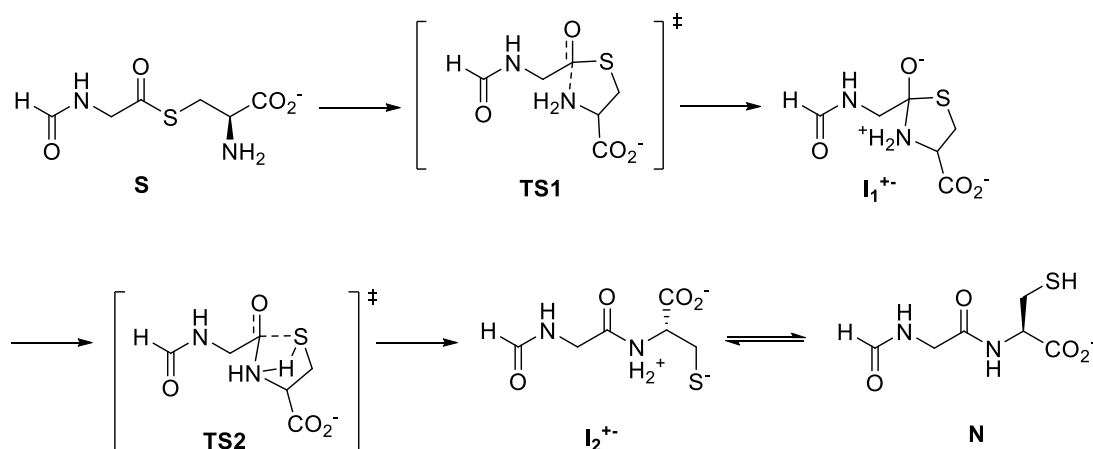


Figure 34 : Mécanisme calculé par Wang et ses collaborateurs pour le transfert intramoléculaire de $S \rightarrow N$.

Barnett et Jencks ont également montré que la vitesse de transfert d'acyle $S \rightarrow N$ de la S-acétylmercaptoéthylamine (S) est insensible à l'addition d'une forte concentration en sel ou de co-solvant comme le *n*-BuOH. À l'inverse, Cuccovia et ses collaborateurs ont montré que la vitesse de transfert d'acyle $S \rightarrow N$ de la S-acétylmercaptoéthylamine est fortement ralentie (100 fois) par l'ajout de dodécylsulfate de sodium (SDS), un détergent chargé négativement¹¹⁹. Les détergents chargés positivement ou neutres comme le bromure d'hexadécyltriméthylammonium (CTAB) ou le Brij n'ont quant à eux pas d'effets. Les auteurs ont assigné l'effet du SDS à une augmentation du pK_a de l'amine primaire de la S-acétylmercaptoéthylamine. Dans l'eau, le pK_a de celle-ci est de 9,1¹²⁰, Cuccovia et ses collaborateurs ont estimé un pK_a de 11 en présence de SDS.

Cette observation met en exergue l'influence d'additifs sur la cinétique de réaction. Au cours des dernières années de nombreux agents dénaturants, détergents, co-solvants ont été utilisés dans les réactions de NCL notamment pour augmenter la solubilité des segments peptidiques. Au vu de l'importance de la réaction de NCL pour l'accès à des grandes protéines et de la complexité des mécanismes mis en jeu, il est important d'identifier et de comprendre le rôle et les effets d'additifs sur la réaction de NCL.

3. Rôle des agents dénaturants, détergents et solvants en NCL

3.1 Les agents dénaturants

Les réactions de NCL sont souvent réalisées en présence d'agents dénaturants comme le chlorhydrate de guanidine ou l'urée. De façon générale, les agents chaotropiques destabilisent la

structure tridimensionnelle des protéines en interférant avec les interactions faibles (liaisons hydrogènes, forces de van der Waals et effet hydrophobe). Dans le cas de la NCL, les agents dénaturants permettent donc d'éviter l'agrégation des peptides en solution et d'augmenter leur solvataion dans les conditions de réaction.

3.1.1 L'urée

Bien qu'elle soit utilisée depuis de nombreuses années, le mécanisme d'action de l'urée pour la dénaturation des protéines n'a pas été totalement élucidé. En 2009, le groupe de Berne suggère que l'urée forme des interactions plus fortes avec les chaînes latérales et le squelette des protéines que l'eau¹²¹. Ceci entraînerait le dépliement de la protéine et permettrait à l'urée de se rapprocher du cœur hydrophobe et d'agir comme un tensioactif par liaison hydrogène avec l'eau et la protéine. La même année, un autre groupe décrit un tout autre processus¹²² démontrant ainsi la controverse au sujet du mode d'action de l'urée. Il n'en reste pas moins que l'urée est très utilisée en biologie pour dénaturer les protéines notamment pour les techniques d'électrophorèse.

Cependant, l'utilisation d'urée peut entraîner des réactions secondaires. En effet, en solution l'urée est en équilibre avec le cyanate d'ammonium. Ce composé peut réagir avec l'amine terminale des protéines ou les chaînes latérales des lysines¹²³. Dans les conditions de NCL classique (pH neutre, 37 °C), l'isomérisation est très lente et l'équilibre est atteint après quelques jours¹²⁴. Par conséquent, ce phénomène ne doit pas en principe gêner l'utilisation d'urée dans des procédés de ligation dès lors que le tampon de ligation est fraîchement préparé.

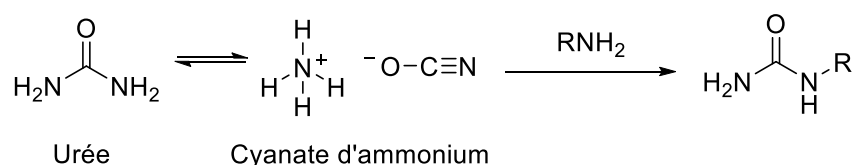


Figure 35 : Réaction de l'isocyanate d'ammonium avec les amines.

3.1.2 Le chlorhydrate de guanidine (Gdn.HCl)

La série de Hofmeister définit les effets de différents ions sur les propriétés physiques (solubilité, stabilité structurale) de molécules organiques ou de protéines¹²⁵. En général, les propriétés d'un sel sont guidées par la nature de l'anion. Cependant, l'ion guanidinium (Gdn⁺) apparaît comme une exception à cette observation. L'ion Gdn⁺ est constitué d'une partie hydrophobe mais peut induire des interactions hydrogènes grâce aux trois groupements NH₂. La structure particulière du Gdn⁺ lui confère des propriétés physicochimiques uniques. L'action de l'ion guanidinium est très dépendante du contre-ion avec lequel il est apparié. En particulier, seuls des sels formés avec des partenaires peu hydratés

(SCN⁻, Cl⁻) induisent la dénaturation des protéines^{126,127}. À l'inverse, les sels formés avec des anions très hydratés (SO₄²⁻) induisent un effet stabilisateur^{128,129}.

En particulier, le mécanisme d'action du Gdn.HCl a été intensivement étudié. Il apparaît que le Gdn.HCl interagit avec la plupart des chaînes latérales des acides aminés^{130,131}. L'ion guanidinium induit des interactions fortes avec les chaînes anioniques¹³² et participe également à des liaisons hydrogènes avec les groupements carbonyles du squelette protéique¹³³. Le rôle du contre-ion a été étudié par le groupe de Cremer. Ils ont montré que lorsque l'on travaille à un pH supérieur au point isoélectrique d'une protéine, il est préférable d'utiliser le Cl⁻ comme contre ion. Dans ce cas, le Gdn⁺ interagit avec l'interface négative de la protéine et entraîne sa dénaturation en stabilisant la forme non repliée. Il est à noter que les processus de dénaturation requièrent de grandes concentrations en sels. Ici, une faible concentration peut entraîner un effet stabilisateur. Par ailleurs, pour des pH de travail inférieurs ou proches du point isoélectrique, il est préférable d'utiliser le thiocyanate de guanidine (Gdn.SCN). En effet, l'ion guanidinium n'interagit pas ou peu avec des surfaces neutres ou chargées positivement. Dans ce cas, l'anion SCN⁻ est nécessaire pour participer à des interactions avec la protéine et ainsi rendre la surface suffisamment négative pour permettre le recrutement de l'ion guanidinium et dénaturer la protéine.

À ce jour, le thiocyanate de guanidine n'a jamais été utilisé pour réaliser des réactions de NCL. L'effet de l'ajout d'un tel composé sur la cinétique de réaction ou sur l'introduction de réactions secondaires est difficile à prévoir. En effet, la présence potentielle d'une fonction thiol (résultant du thiocyanate) pourrait interférer dans la réaction de ligation.

À l'inverse, le chlorhydrate de guanidinium est un additif classiquement utilisé pour la NCL. La plupart des protéines sont dénaturées dans une solution à 6 M de Gdn.HCl. Ceci permet d'éviter le repliement (*foldng*) des segments peptidiques qui pourrait masquer les extrémités réactives au cours de la réaction. Par ailleurs, le Gdn.HCl est beaucoup plus efficace que l'urée en tant que dénaturant : 2,8 fois pour le dépliement de ribonucléase et 1,7 fois pour le lysozyme¹³⁴.

Cependant, bien que les thioesters soient stables en présence de Gdn.HCl à pH neutre ou acide, l'utilisation de conditions basiques peut entraîner l'aminolyse de ces derniers. On observe ainsi la formation de produits secondaires de type *N*-peptidyl guanidine (Figure 36).

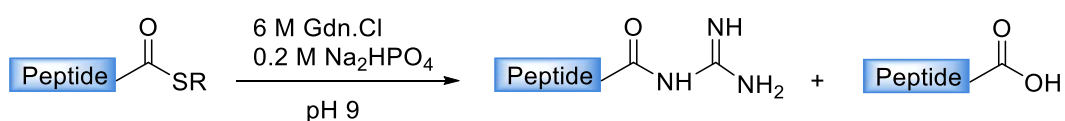


Figure 36 : Réaction d'hydrolyse de thioester en présence de Gdn.HCl à pH basique.

Finalement, il est important de noter que l'utilisation de fortes concentrations en sels peut avoir un effet sur la valeur réel du pH, à opposer avec la valeur mesurée par une électrode de verre. En 2001, le groupe de Sanchez-Ruiz a étudié ce phénomène et établi qu'un facteur de correction de ~0,8 devait être ajouté à la valeur de pH mesurée dans 6 M de Gdn.HCl¹³⁵.

3.2 Les détergents ou tensioactifs

Une grande variété de détergents a été utilisée dans des réactions de ligation pour aider à solubiliser à la fois les fragments peptidiques et le produit de ligation.

Le β -*N*-octylglucoside (OG) est l'un des premiers détergents à avoir été utilisé dans une réaction de ligation. C'est un détergent non ionique¹³⁶ qui présente l'avantage d'être facilement éliminé par dialyse¹³⁷. Il a initialement été utilisé pour solubiliser des protéines membranaires tout en conservant leurs activités enzymatiques^{137,138}. En 1999, Bianchi et ses collaborateurs l'ont utilisé pour la synthèse d'un cofacteur de la protéase HCV (virus de l'hépatite C) par ligation chimique native¹³⁹. Ils ont assemblé deux fragments peptidiques dans 25 mM de tampon HEPES (pH 7) en présence de 1 % de thiophénol et 2 % d'OG. Il est intéressant de noter que l'utilisation de différentes concentrations de chlorohydrate de guanidine (2-6 M) en présence ou non de co-solvants (ACN, dioxane, DMF) n'a pas permis de solubiliser la protéine cible.

La dodécylphosphocholine (DPC) a également été utilisée dans des procédés de ligation. En 2004, Clayton l'a utilisé pour la synthèse totale de deux protéines membranaires d'*E. coli* (Ec-MscL) et de *Mycobacterium tuberculosis* (Tb-MscL)¹⁴⁰. Les ligations ont été réalisées dans 8 M d'urée, 0,25 M de tampon phosphate (pH 7,5) en présence de 48 mM de DPC et 0,5 % thiophénol. Plus récemment, Kwon a décrit l'utilisation de DPC et d'urée pour solubiliser des fragments hydrophobes de la protéine M2 du virus de la grippe⁹. Cependant, il rapporte que la DPC contamine le produit final de ligation et ne peut être complètement éliminée. À noter, l'utilisation d'OG n'a pas permis de conduire la réaction de ligation dans le cas décrit.

Le dodécylsulfate de sodium (SDS) est un tensioactif ionique également très utilisé pour réaliser des réactions de NCL avec des segments hydrophobes. En 2002, Valiyaveetil a montré l'utilité du SDS pour la synthèse du canal potassique KcsA¹⁴¹. Dans cette étude, il a testé différentes conditions de ligation, 50 % TFE + 4,8 M Gdn.HCl, 1 % SDS ou 1 % SDS + 8 M urée, pour augmenter la solubilité des segments mis en jeu. En présence de 1 % SDS, les segments restent solubles même après des temps de réactions longs tandis que dans les autres conditions ils observent une précipitation lente. Des études complémentaires ont montré qu'une lyophilisation préalable des peptides en présence de SDS permettait d'obtenir des concentrations peptidiques plus élevées dans le milieu de ligation.

Ces différents exemples montrent l'importance d'adapter les combinaisons détergent/catalyseur thiol/agent dénaturant au cas par cas pour obtenir des ligations efficaces.

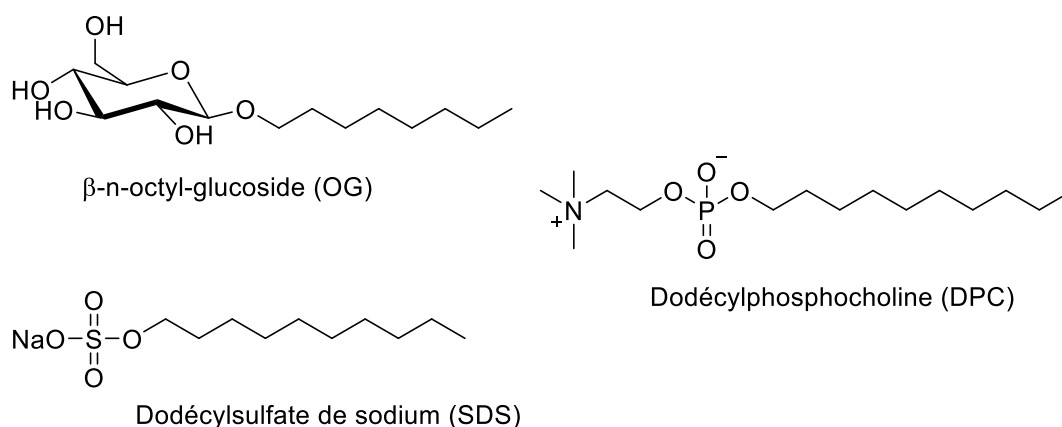


Figure 37 : Exemples de détergents utilisés pour la ligation de peptides hydrophobes.

3.3 Les solvants organiques

Les réactions de NCL ont été réalisées en présence de différents solvants organiques. Dittman et ses collaborateurs ont même décrit des réactions de NCL dans des milieux totalement organiques^{142,143}. En 2010, ils ont réalisé une étude sur quatre solvants polaires : DMF, NMP, DMSO et TFE¹⁴². Ils notent que la présence d'une base, ici la triéthylamine, est nécessaire pour induire la ligation. Le DMF a conduit aux cinétiques de réaction les plus rapides comparés à la NMP ou au TFE, tandis qu'aucun produit de ligation n'a été observé dans le DMSO. Ils ont également noté que l'utilisation d'une combinaison de thiophénol et de benzylmercaptan catalyse plus efficacement la réaction de NCL comparativement au MPAA. Dans les deux cas, ils ont observé la formation de précipités de réactifs et de produits de ligation après de long temps de réaction (12 h). Le groupe de Dittman a montré que l'ajout de LiCl au milieu réactionnel permettait d'éviter la précipitation des segments hydrophobes, même après de longs temps réactionnels. L'utilisation de sels de lithium a été décrite en premier lieu par Seebach pour augmenter la solubilité de peptides dans des solvants aprotiques et non polaires¹⁴⁴.

Plus tard, Dittman et ses collaborateurs ont vérifié si l'utilisation de DMF n'induisait pas de réactions secondaires¹⁴³. Dans leur étude, ils montrent que l'utilisation de triéthylamine (20 mM), de thiophénol (20 mM), de LiCl (0,18 M) dans du DMF conduit au minimum à 3 % d'épimérisation. De plus, ils démontrent que la réaction de NCL dans le DMF est chimiosélective. Il n'y a pas d'interactions avec les groupements fonctionnels des chaînes latérales des peptides mis en jeu. L'utilisation de DMF ne nécessite donc pas l'utilisation de groupements protecteurs pour les chaînes latérales. Inspiré par ces travaux, Sohma a décrit l'utilisation de DMF/Et₃N pour la synthèse d'O-acyl peptides par ligation¹⁴⁵. L'utilisation de DMF a permis de prévenir la migration d'acyle de l'oxygène vers l'azote conduisant au

réarrangement des *O*-acyl peptides. En effet, en absence de solvant protique (eau, alcool) l'aminolyse des esters est probablement très lente¹⁴⁶.

Bien qu'aucun produit de ligation n'ait été observé par Dittmann dans le DMSO, ce dernier a été utilisé en tant que co-solvant par le groupe de Herbst. En 2016, il a utilisé 7,5 % de DMSO dans du tampon phosphate salin (PBS) pour éviter l'agrégation des fragments peptidiques comportant des fluorophores¹⁴⁷. Il est important de noter que l'utilisation de DMSO peut avoir un effet sur la valeur réel du pH. En effet, Mukerjee et Ostrow ont montré que le DMSO a un effet significatif sur les valeurs de pH notamment dans les tampons acétate et phosphate¹⁴⁸.

Enfin, les solvants organiques peuvent également être utilisés en combinaison avec les dénaturants classiques pour augmenter la solubilité de fragments peptidiques hydrophobes. En 1999, Kochendoerfer et Salom réalisèrent la synthèse de la protéine M2 de la grippe A en assemblant deux segments dans un mélange 6 M Gdn.HCl (pH 7,5) TFE (70/30 v/v)¹⁴⁹. En 2015, Kwon et ses collaborateurs ont décrit la synthèse de la protéine M2 (22-71) de la grippe⁹. La ligation entre les fragments a été réalisée dans 60 % TFE (v/v) et 5,4 M d'urée. Il est intéressant de noter que dans un mélange TFE/Gdn.HCl la ligation n'a pas eu lieu.

Cette observation montre bien que chaque protéine a ses particularités et qu'il n'y a pas de méthodes et de conditions universelles pour assembler des segments qui posent de sérieux problèmes de solubilité et d'agrégation.

4. Réactions secondaires au niveau des cystéines/sélénocystéines

4.1 Réaction d'oxydation des thiols/séléniols

La réaction de NCL et les méthodes apparentées nécessitent la présence de fonctions thiols ou séléniols libres dans les conditions de ligation.

Les thiols sont connus pour réagir spontanément avec l'oxygène. Ce phénomène a été intensivement étudié depuis de nombreuses années. En 1968, Cavallini a décrit le mécanisme de la réaction entre les composés thiols et l'oxygène en solution aqueuse¹⁵⁰. Cette réaction conduit à la formation de disulfures, d'eau ou de peroxyde d'hydrogène (équations 1 et 2, Figure 38). Cependant, l'oxydation des thiols peut conduire à différents états d'oxydation du soufre en fonction des conditions opératoires. À pH neutre (pH 7-9), l'oxydation des thiols conduit à la formation de disulfure selon le mécanisme décrit par la Figure 38¹⁵¹. Ce mécanisme met en jeu la production d'espèces radicalaires thiyles, qui par combinaison avec les espèces thiolates conduit à la formation de disulfures. Les réactions de formation ou de décomposition des radicaux RSSR^\bullet sont en général les étapes limitantes du processus (équations 4 et 5, Figure 38).

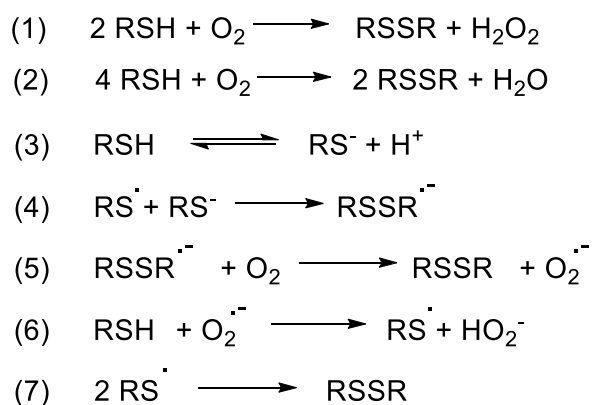


Figure 38 : Réaction d'oxydation des thiols par l'oxygène.

La cinétique d'oxydation des thiols est corrélée à la concentration en espèces thiolates et donc au pH du milieu aqueux (ou à la quantité de base en milieu organique¹⁵²). Ainsi, l'oxydation est plus rapide quand le pH de la solution est proche ou au-dessus du pK_a de l'espèce thiol impliquée.

Par ailleurs, plusieurs scientifiques ont remarqué que la présence d'ions métalliques catalyse la réaction d'oxydation¹⁵³⁻¹⁵⁵. En 2003, Bagiyan et ses collaborateurs ont étudié plusieurs métaux et décrivent l'ordre d'efficacité catalytique suivant dans des tampons borate-phosphate : Cu > Mn > Fe > Ni >> Co¹⁵¹. Ils ont également décrit que la catalyse par ces métaux augmente de 100 fois la cinétique d'oxydation en milieu neutre ou légèrement basique (pH 6-12). En milieu acide, les thiols s'oxydent selon des cinétiques similaires avec ou sans catalyseur métallique. Le fait que les radicaux thiyles peuvent être produits par réaction entre un thiol et l'ion ferrique est bien connu. Ce procédé est une source importante de radicaux thiyles et d'espèces réactives de l'oxygène dans les organismes vivants¹⁵⁶.

La plupart des réactifs, solutions, tampons utilisés en laboratoires contiennent des traces de fer ou d'autres métaux¹⁵⁷. La stabilité de différents thiols (DTT, 2-mercaptoéthanol, MPA, glutathion) a été étudié par Stevens et ses collaborateurs¹⁵⁸. En jouant sur différents paramètres (température, pH, présence Cu²⁺), ils ont montré que le temps de demi-vie de ces thiols pouvait décroître considérablement. En absence de précautions particulières, ils sont de l'ordre de quelques heures voire inférieurs à une heure en présence d'ions métalliques du cuivre (exemple du DTT et du MPA, Figure 39). À l'inverse, l'ajout d'acide éthylène diamine tétraacétique (EDTA) permet de diminuer la cinétique d'oxydation. L'EDTA est un chélateur de métaux très utilisé dans les tampons biologiques mais peu utilisé lors des ligations chimiques^{159,160}.

Conditions	Temps de demi-vie (h)	
	DTT	MPA
pH 6,5 à 20 °C	40	7
pH 8,5 à 20 °C	1,4	4,5
pH 8,5 à 40 °C	0,2	1,6
pH 8,5 à 20 °C + 0,1 M Cu ²⁺	0,6	0,4
pH 8,5 à 20 °C + 1 M EDTA	4	> 100

Figure 39 : Étude de l'effet de différents paramètres sur le temps de demi-vie du DTT dans 0,1 M tampon phosphate. Tableau reproduit à partir des données de R. Stevens¹⁵⁸.

La réaction de NCL et les méthodes apparentées sont généralement réalisées à 37 °C, en milieu neutre ou légèrement acide. Dans ces conditions, les thiols réagissent rapidement avec l'oxygène que ce soit la cystéine participant à la jonction ou les thiols exogènes (catalyseurs). Il est donc important d'empêcher ce phénomène en travaillant en conditions inertes et/ou en présence d'agents réducteurs forts comme le chlorhydrate de tris(2-carboxyéthyl)phosphine (TCEP). Si la réaction a lieu sous atmosphère inerte, la présence d'aryle thiol comme catalyseur de la réaction est suffisante pour maintenir les cystéines sous forme réduite. En effet, les aryle thiols sont des réducteurs faibles mais ils sont généralement utilisés en excès.

4.2 L'acide ascorbique : inhibiteur d'oxydation

Les procédés d'oxydation des thiols ou des sélénols sont principalement dus à la formation de radicaux thiyles ou sélényles dans les conditions de ligation. L'acide ascorbique (vitamine C) est un diacide très utilisé pour prévenir l'oxydation des thiols en disulfure. En effet, celui-ci réagit avec les radicaux formés pour régénérer des thiols selon le mécanisme présenté par la Figure 40. Le radical $AscS\dot{H}^{\cdot}$ formé par réaction d'un radical thiyle avec l'ascorbate se dismute et conduit à la formation d'ascorbate et d'acide dehydroascorbique (DHA). La réaction de dismutation est un équilibre, ce qui signifie que la présence de DHA en solution peut conduire à la formation de $AscS\dot{H}^{\cdot}$. Cependant, l'électron non apparié de ce composé est très délocalisé ce qui en fait un faible réducteur ou oxydant. Il n'est pas capable de réagir avec un thiol pour former de nouveaux radicaux thiyles¹⁶¹. Il est important de noter que l'acide ascorbique peut également inhiber la réaction de désélénsation¹⁶².

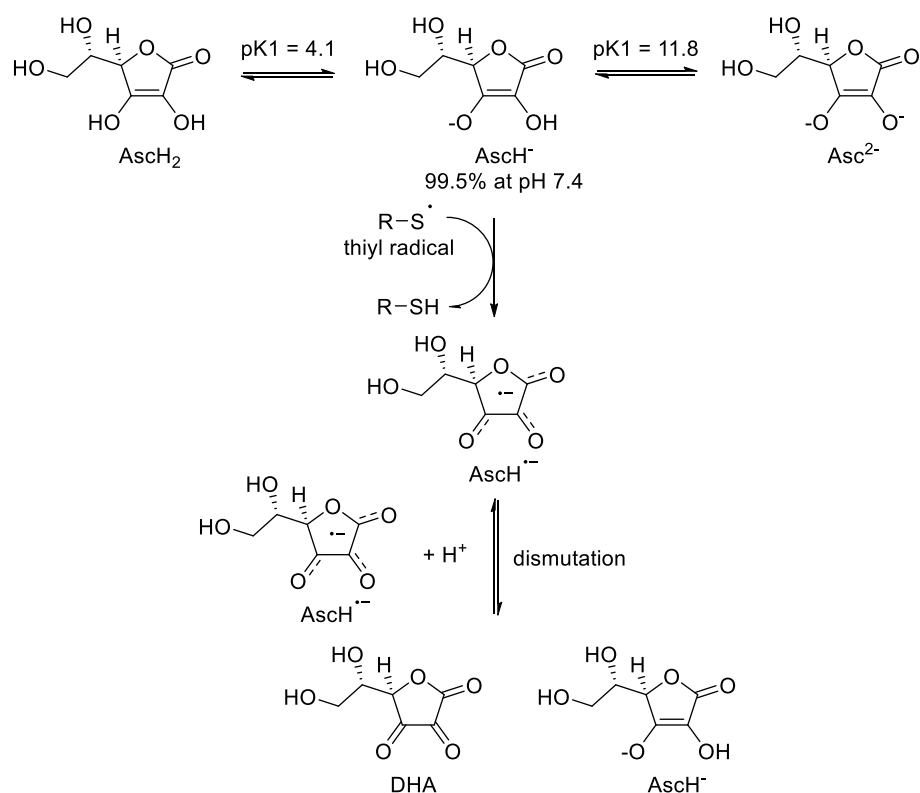


Figure 40 : Réaction de l'acide ascorbique avec les radicaux thiyles.

Par ailleurs, l'acide ascorbique n'est pas capable de réduire les ponts disulfures de protéines bien qu'il puisse réduire des disulfures activés comme l'acide 5,5'-dithiobis(2-nitrobenzoïque) (DTNB)¹⁶³. D'autres agents réducteurs de ponts disulfures ou disélénures sont donc généralement utilisés dans les réactions de NCL.

4.3 Agent réducteur de disulfures ou disélénures : la TCEP

4.3.1 Propriétés générales

La TCEP est un agent réducteur très utilisé dans les réactions de NCL. Ce composé est inodore et est disponible commercialement sous forme de sel de chlorhydrate¹⁶⁴. La TCEP présente une bonne solubilité dans l'eau (310 g/L) et est stable sous forme solide ou en conditions aqueuses¹⁶⁵. Le pK_a de la phosphine est de 7,66 tandis que les pK_a des fonctions acides carboxyliques sont de 2,99 ; 3,67 et 4,36¹⁶⁶. En 1969, Levison et ses collaborateurs ont étudié pour la première fois la TCEP pour sa capacité à réduire des ponts disulfures dans les protéines¹⁶⁷. Dans leur étude, ils ont comparé les propriétés de réduction du mercaptoéthanol, de la tris(hydroxyméthyl)phosphine et de la TCEP pour la réduction de ponts disulfures d'immunoglobulines G (IgG). Ils montrent que la TCEP présente un fort pouvoir réducteur. Par ailleurs, Gray et ses collaborateurs ont montré que la TCEP était efficace dans des conditions acides, jusqu'à pH 3¹⁶⁸. En effet, ils notèrent une chute de la cinétique de réduction entre pH 3 et pH 2. Cette observation suggère l'implication d'un groupement carboxylate dans le mécanisme

de réduction. Par exemple, le carboxylate de l'espèce décrite Figure 41B pourrait déprotoner la phosphine intramoléculairement, et ainsi permettre une attaque nucléophile de la phosphine sur le disulfure.

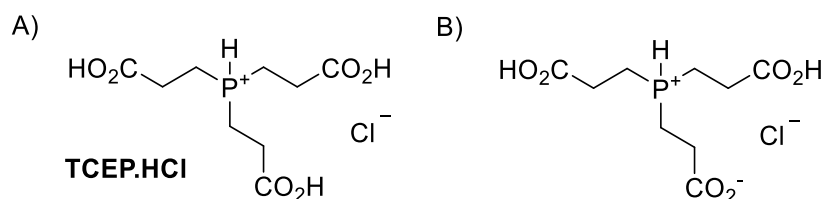
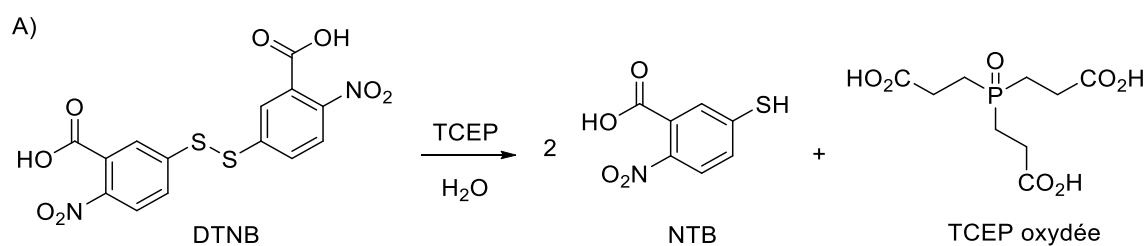


Figure 41 : A) Chlorhydrate de tris(2-carboxyéthyl)phosphine B) Le groupement carboxylate participe à la réduction des ponts disulfures à $pH > 3$.

4.3.2 Stabilité en solutions aqueuses

La stabilité de la TCEP en solution a été étudiée par le groupe de Han¹⁶⁹ à l'aide d'une méthode permettant de déterminer la concentration de TCEP en solution. Cette méthode mesure par spectrométrie UV ($\lambda=412$ nm) la quantité de 2-nitro-5-thiobenzoate (NTB) formée après réduction de l'acide 5,5'-dithiobis(2-nitrobenzoïque) par la TCEP (Figure 42A). En utilisant cette méthode, il est facile de déterminer la stabilité de la TCEP dans différents tampons (Figure 42B).



B)

Entrée	Solution/tampon d'incubation	% TCEP oxydée	
		24 h	3 semaines
1	100 mM HCl	< 3 %	< 3 %
2	100 mM NaOH	< 3 %	15,4±0,4
3	50 mM Tris-HCl, pH 7,5	< 3 %	18,7±0,6
4	50 mM Tris-HCl, pH 8,5	< 3 %	16,8±0,5
5	50 mM Tris-HCl, pH 9,5	< 3 %	14,5±0,5
6	50 mM Borate, pH 8,2	< 3 %	6,6±0,2
7	50 mM Borate, pH 10,2	< 3 %	5,3±0,1
8	50 mM HEPES, pH 6,8	< 3 %	14,8±0,4
9	50 mM HEPES, pH 8,2	< 3 %	13,6±0,3
10	50 mM CAPS, pH 9,7	< 3 %	9,8±0,3
11	50 mM CAPS, pH 11,1	< 3 %	4,3±0,1

Figure 42 : A) Réduction du DTNB par la TCEP B) Stabilité de la TCEP (2mM, 25-28 °C, 5 mL dans des flacons scellés) dans différents tampons.

Il apparaît que la TCEP est stable durant 24 heures dans des conditions acides ou basiques (entrées 1 et 2, Figure 42B) mais également dans des tampons Tris, Hepes et Caps couvrant une gamme de pH 6,8-11,1 (entrées 3-11, Figure 42B). De plus, après 3 semaines dans des flacons scellés contenant les différentes solutions, le pourcentage d'oxydation est inférieur à 20 %, traduisant la grande stabilité de la TCEP dans ces conditions.

Il est cependant évident que la TCEP est moins stable en présence de tampon phosphate. Ce phénomène a été décrit en premier lieu par l'équipe de Burns¹⁶⁵. En 1991, ils ont montré que la TCEP s'oxyde rapidement en tampon phosphate (0,4 M) à des pH basiques. Ils ont noté également que

l'oxydation est beaucoup plus lente à des pH inférieurs à celui du pKa de la phosphine (pKa 7,66). Pour leur étude, les solutions ont été agitées vigoureusement à l'air et à température ambiante et les analyses ont été réalisées par RMN ^1H . Après 23 heures, ils ont observé 80 % de forme oxydée et l'oxydation était complète après 72 heures pour un pH de 7,4 tandis qu'à pH de 11,6, l'oxydation complète était atteinte après 30 min seulement.

Han et ses collaborateurs ont décrit également une oxydation accélérée en présence de tampon phosphate. Ils ont montré que la TCEP est particulièrement sensible à l'oxydation pour des pH proches de la neutralité (pH 7-8) : après 72 heures ils ont observé 100 % d'oxydation en présence de 0,35 M de tampon phosphate à pH 7.5 (Figure 43). Cependant, leur étude est en contradiction avec celle de Burns pour des pH plus basiques (pH 11,6).

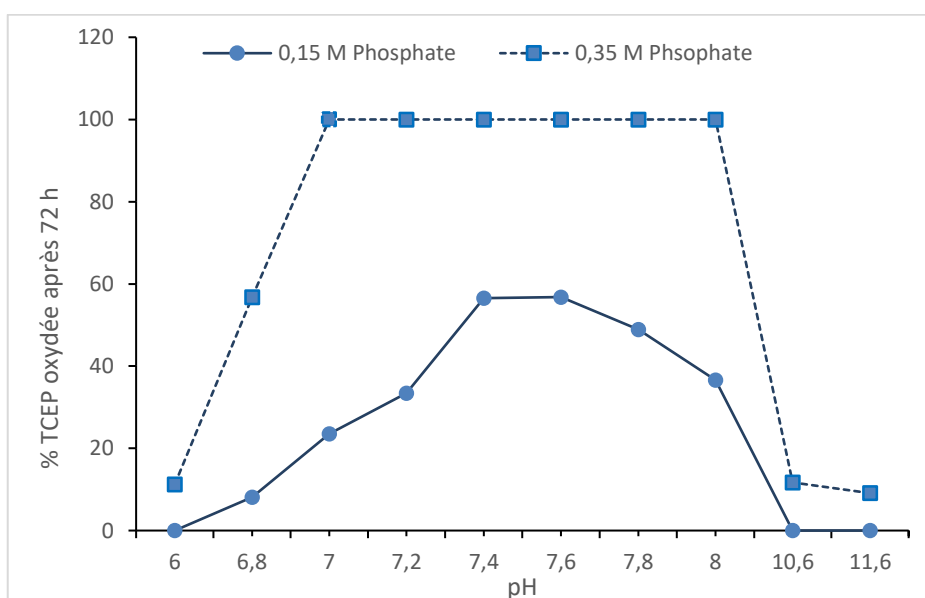


Figure 43 : Stabilité de la TCEP en tampon phosphate (0,15 M et 0,35 M phosphate, 5 mL de solution dans des flacons scellés, 25-28 °C). La concentration en TCEP est déterminée en utilisant la méthode DTNB (quantification de NTB produit par réduction à 412 nm).

Il est important de noter que les réactions de NCL sont généralement conduites dans une gamme de pH allant de 6,5 à 8. De plus, les ligations à des jonctions difficiles induisent des temps de réactions prolongés (plusieurs jours). L'oxydation de la TCEP dans ces conditions pourrait poser problème.

Enfin, la méthode développée par le groupe de Han a été utilisée par le groupe de Getz pour déterminer l'influence d'ions métalliques (Mg^{2+} , Ni^{2+} , Fe^{3+}) et d'un chélateur de métaux sur la stabilité de la TCEP¹⁷⁰. Dans leur étude, ils ont montré que la TCEP (0,1-1 M) est relativement stable en présence d'ions métallique à 4 °C (≤ 5 % d'oxydation après une semaine) et à 25 °C (≤ 11 % d'oxydation après une semaine). À l'inverse, l'ajout d'acide éthylène-*bis*(oxyéthylènenitrilo)tétracétique (EGTA) favorise significativement l'oxydation de la TCEP (68-72 % d'oxydation après une semaine).

4.3.3 Déchalcogénéation

L'utilisation de TCEP peut induire la déchalcogénéation de thiols ou de sélénols et notamment de la cystéine¹⁷¹ et de la séléncystéine¹⁷²⁻¹⁷⁴.

La désulfuration d'alkylthiols par des phosphites a été observée pour la première fois par le groupe de Hoffman en 1956¹⁷⁵. Ils notèrent que la réaction est catalysée par des températures élevées ou par irradiation UV à température ambiante. Les conditions photochimiques induiraient la scission homolytique de la liaison S-H. Un an après, le groupe de Walling a décrit l'utilisation d'un initiateur de radical, l'AIBN (azobisisobutyronitrile), pour accélérer la désulfuration¹⁷⁶. Il a également proposé un mécanisme impliquant la combinaison de radicaux thiyles avec l'atome de phosphore conduisant à la formation d'un radical phosphoryle. Ce dernier se décompose par coupure homolytique de la liaison C-S en radical alkyle et en sulfure de phosphine. Finalement, le radical d'alkyle adsorbe un proton du solvant ou d'un thiol du milieu réactionnel pour former un groupement alkyle. Ce mécanisme s'applique à la désulfuration de peptides par la TCEP (X=S, Figure 44).

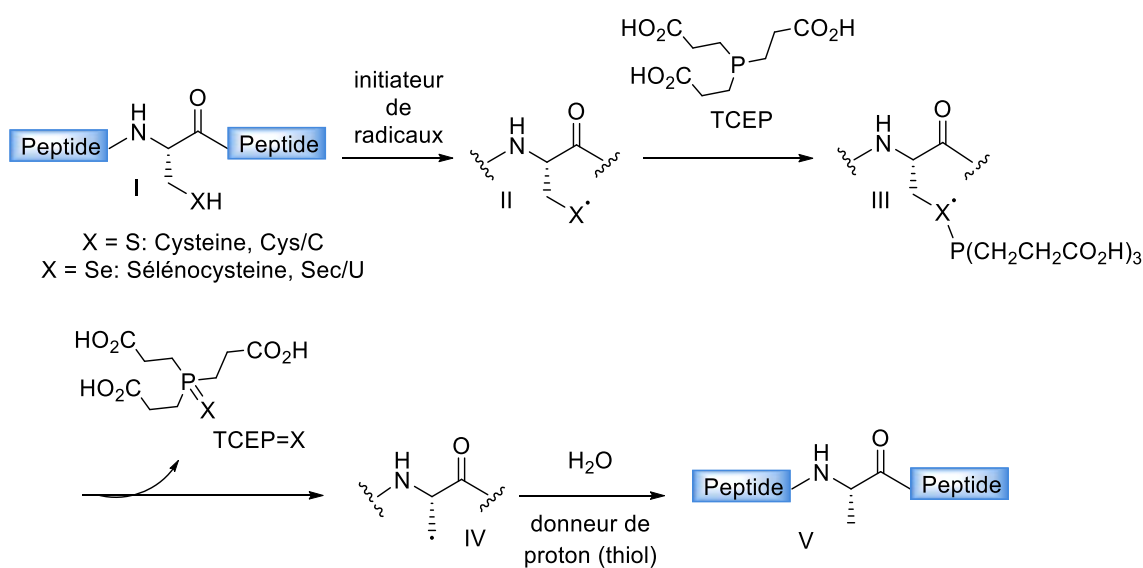


Figure 44 : Mécanisme radicalaire de déchalcogénéation.

Le procédé de désélénsation de disélénures par les phosphines a également été intensivement étudié. En 1975, le groupe de Millington a décrit l'abstraction du sélénium du disélénuire de diéthyle par des phosphines tertiaires sous radiation UV¹⁷⁷. Il propose un mécanisme impliquant la formation photochimique de radicaux sélényles suivie de la réaction avec une phosphine (équations 1 et 2, Figure 45). Le radical $[\text{Ph}_2\text{MePSeEt}]^*$ se décomposerait ensuite en séléno-phosphine et radical éthyle (équation 3, Figure 45). Finalement, le radical éthyle réagirait avec le disélénuire de diéthyle pour produire le sélénuire de diéthyle et EtSe^* (équation 4, Figure 45). Un an plus tard, le groupe de Chu a décrit un

mécanisme similaire pour la désélénisation du diséléniure de benzyle par la triphénylphosphine¹⁷⁸. Ce mécanisme s'applique également à la désélénisation de sélénopeptides par la TCEP (X=Se, Figure 44).

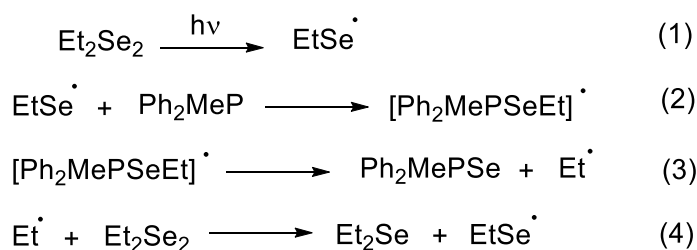


Figure 45 : Mécanisme de désélénisation du diséléniure de diéthyle par une phosphine tertiaire.

Il est à noter que les énergies de dissociation des liaisons X-H (BDE) de la cystéine (X=S, BDE = 367 kJ/mol)¹⁷⁹ et de la sélénocystéine (X=Se, BDE = 314 kJ/mol)¹⁸⁰ ont été déterminées par des calculs *ab initio*. Il apparaît que la liaison séléniol (Se-H) est plus faible que la liaison thiol (S-H). Par conséquent, la formation d'un radical thiyle à partir d'un thiol nécessite un initiateur de radicaux, une température élevée ou une irradiation UV tandis que la formation de radicaux sélényles est facile et rapide à température ambiante¹⁸¹. Finalement, la différence de réactivité de ces composés permet d'envisager la désélénisation sélective d'une sélénocystéine par la TCEP en présence d'une cystéine¹⁷².

La déchalcogénation de cystéines ou sélénocystéines en alanine a été beaucoup étudiée. En 2007, le groupe de Danishefsky a décrit une méthode de désulfuration par voie radicalaire en utilisant un analogue hydrosoluble de l'AIBN, le VA 044¹⁸². Depuis, cette méthode a été étendue à la déchalcogénation de nombreux dérivés thiolés d'acides aminés ce qui permet d'envisager des ligations à tout type de jonction (discuté au chapitre III section 2).

Dans le cas où elle ne serait pas souhaitée, la réaction de déchalcogénation par la TCEP peut être inhibée par l'ajout d'acide ascorbique. Comme décrit dans la section précédente (4.2), l'acide ascorbique réagit avec les radicaux thiyles ou sélényles, cette réaction doit probablement avoir lieu avant que les radicaux n'aient pu réagir avec la phosphine. L'ajout d'acide ascorbique permet donc de prévenir la désulfuration¹⁸³ ou la désélénisation¹⁶² lors de réaction de NCL. Finalement, ce mécanisme étant catalysé par la présence d'une phosphine tertiaire, une solution potentielle est de remplacer la TCEP par du DTT.

5. Bilan/Conclusion

En résumé, la NCL et les méthodes apparentées sont devenues des outils courants et indispensables à la synthèse de protéines par voie chimique. Ces méthodologies ont fait l'objet de beaucoup de revues concernant les aspects synthétiques et les applications biologiques. Ici, nous avons présenté une étude approfondie des concepts fondamentaux et processus chimiques impliqués

dans ces réactions. L'étude des aspects mécanistiques des réactions mises en jeu a permis de mettre en exergue l'importance du développement de catalyseurs. En effet, les cinétiques des réactions sont très importantes afin d'éviter l'hydrolyse et l'agrégation des segments et prévenir les réactions secondaires. Le rôle de différents additifs, solvants et détergents sur la réactivité et la solubilité des protéines a été discuté. Nous avons vu qu'il n'y a pas de stratégie synthétique universelle et que chaque protéine possède des propriétés qui lui sont propres et qui impliquent des conditions réactionnelles particulières. Finalement, l'étude et la compréhension de ces concepts fondamentaux nous permet de mieux appréhender les améliorations et développements nécessaires afin de surmonter les limitations actuelles des méthodes de ligation. Nous proposons un état des lieux de ces limitations dans la suite de ce manuscrit, ainsi que des pistes de réflexion pour y remédier.

CHAPITRE III : VISION STATISTIQUE DE LA SYNTHÈSE CHIMIQUE DE PROTÉINES

La NCL et les méthodes dérivées ont permis d'apporter de nouveaux outils pour l'obtention de protéines synthétiques. Cependant, plus de 20 ans après l'introduction de la NCL se pose la question de l'impact réel de cette chimie sur le domaine et sur l'évolution des méthodes utilisées. Nous avons souhaité réaliser une étude statistique de l'utilisation de la NCL et des méthodes apparentées pour la synthèse totale de protéines par voie chimique dans le but d'identifier les limitations actuelles dans le domaine. Les données présentées sont tirées de la base de données de synthèse de protéines (*Protein chemical synthesis database, PCS*) (note^a). J'ai participé à la création de cette base qui regroupe aujourd'hui près de 700 protéines produites de 1994 à juillet 2017, publiées dans plus de 500 articles^b. Elle ne contient que des protéines d'intérêt biologique et exclut les peptides et protéines modèles, les polymères ou molécules hydrides. Seules les protéines produites par NCL ou en utilisant des méthodes apparentées ont été incluses dans l'étude qui suit. Les méthodes apparentées impliquent ici soit l'utilisation de précurseurs de thioesters (sélénioester, hydrazide, Nbz, systèmes de transfert d'acyle) soit l'utilisation d'acides aminés thiols autre que la cystéine. À noter que les autres types de ligation native comme la ligation de Staudinger ou la ligation KAHA ont été récemment incluses dans la base de données mais ne sont pas pris en compte dans les statistiques ci-après.

1. Fréquence des acides aminés protéinogéniques

Initialement, la mise en œuvre de la réaction de NCL nécessite la présence d'une cystéine ou d'une sélénocystéine au niveau du site de ligation. La fréquence des acides aminés protéinogènes peut être estimée à partir de la base de données UniRef50 qui fournit un échantillon important de protéines et leurs caractéristiques (Figure 46). Il ressort que la fréquence du résidu cystéine est faible (1,48 %) et que celle de l'alanine est 5,7 fois supérieure à celle-ci (8,42 %). Le génome humain ne codant que pour seulement 25 sélénoprotéines¹⁸⁴, la fréquence du résidu sélénocystéine est extrêmement faible (1,19 x 10⁻⁵ %). Par ailleurs, O. Carugo a montré que la fréquence du résidu cystéine diminue avec la taille des protéines. À l'inverse celle de l'alanine croît (Figure 46). À titre de comparaison pour une protéine de 50 acides aminés, la fréquence des cystéines est de 2,25 % contre 6,68 % pour l'alanine. Lorsque l'on s'intéresse à une protéine de 200 acides aminés, les valeurs passent à 1,45 % pour la cystéine et 8,75 % pour l'alanine. Ce constat a stimulé le développement de nouvelles méthodes pour étendre la

^a Cette base est consultable à l'adresse suivante <http://pcs-db.fr/>.

^b Mise à jour du 05/07/2017

ligation chimique native à d'autres types de jonctions, notamment Xaa-Ala. Aujourd'hui, la formation de jonctions Xaa-Ala est rendue possible par des procédés de ligation couplés à des techniques de désulfuration^{182,185} ou désélémentation^{162,172} (Figure 47A).

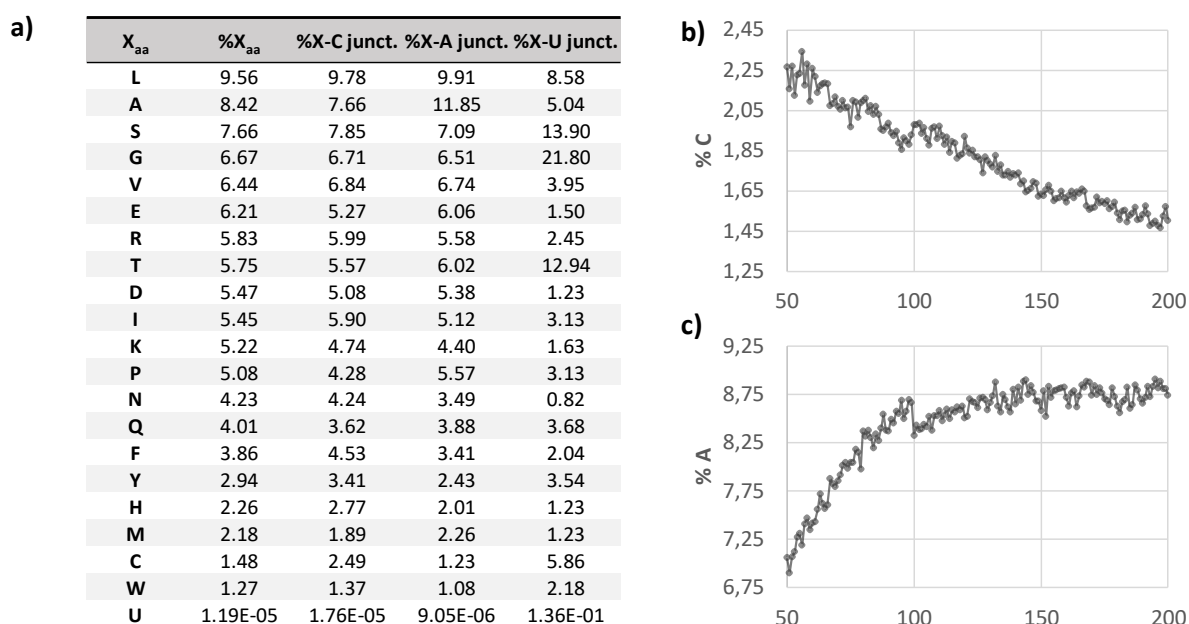


Figure 46 : A) Statistique issues de la base de données UniRef50 : fréquence des acides aminés. B-C) : Evolution de la fréquence des résidus Cys (a) ou Ala (b) en fonction de la taille des protéines (données reproduites à partir de Carugo, O. *Protein Science* 2008, 17, 2187-2191).

2. Le processus de désulfuration

En 2001, P.E Dawson réalise une étude décrivant la réaction de désulfuration d'une cystéine en alanine après ligation chimique native¹⁸⁵. Il établit ainsi le premier procédé de ligation désulfuration avec la synthèse d'un peptide antimicrobien. Ce travail a stimulé le développement de nouveaux procédés de ligation-désulfuration (ou désélémentation¹⁷²) que ce soit de la cystéine (ou sélélocystéine) ou de tout autre dérivé thiol d'acide aminé. En effet, on observe aujourd'hui la conception de nombreux dérivés thiol d'acide aminés (Figure 47) qui permettent, par la combinaison de la réaction de NCL et de désulfuration, d'étendre la NCL à la formation de pratiquement tout type de jonction¹⁸⁶. Le premier acide aminé non naturel synthétisé fût la β -mercapto phénylalanine (Phe) en 2007. Puis, en 2011, le groupe de S. J. Danishefsky a réalisé la synthèse d'une protéine ne contenant aucune cystéine, l'hormone parathyroïde humaine Hpth, en utilisant des dérivés thiols de la valine et de la leucine¹⁸⁷.

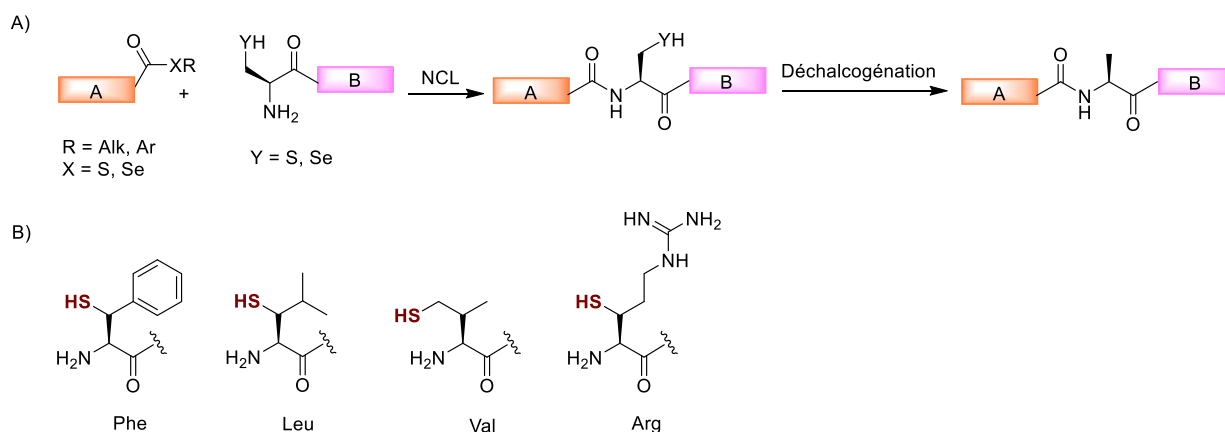


Figure 47 : A) Principe de la déchalcogénéation. B) Exemples d'acides aminés non naturels équipés d'une fonction thiol : Phe¹⁸⁸, Leu¹⁸⁷, Val¹⁸⁷ et Arg¹⁸⁹.

Ces procédés permettent d'une part de surpasser la faible abondance du résidu cystéine dans les protéines naturelles mais également le problème de positionnement de celle-ci dans la séquence primaire des protéines cibles. En effet, il est préférable d'assembler des segments de tailles similaires pour accéder à des protéines synthétiques. C'est pourquoi l'utilisation de méthodes de ligation combinées à des processus de désulfuration est en évolution constante depuis les années 2010 (Figure 48). Le développement de procédés de ligation/déchalcogénéation en un pot a fait tout particulièrement l'objet de nombreux travaux¹⁰⁹.

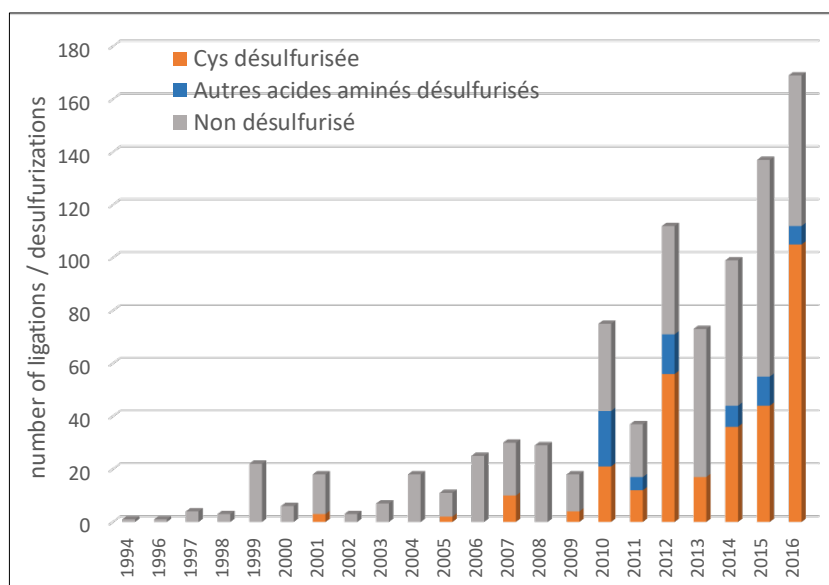


Figure 48 : Nombre d'acides aminés de type thiol (Cys ou dérivés thiols d'acides aminés) désulfurisés après ligation et nombre de ligations à la cystéine sans désulfuration.

3. L'introduction de nouveaux précurseurs de thioester

Depuis 2010, le nombre de protéines synthétisées par année à l'aide de la NCL n'a cessé d'augmenter (Figure 49). Cette évolution traduit le développement de nouvelles méthodes,

notamment celles permettant la synthèse de peptides thioester¹⁹⁰. Initialement, les peptides thioester ont été obtenus par synthèse sur support solide en utilisant la stratégie Boc/benzyle¹⁹¹. La Boc-SPPS nécessite l'utilisation de HF anhydre pour détacher le peptide du support solide et déprotéger les chaînes latérales. Ce réactif est très dangereux et limite donc l'utilisation de cette stratégie. Malheureusement, les thioesters ne sont généralement pas stables dans les conditions classiques de la Fmoc-SPPS, qui est aujourd'hui la technique de synthèse de peptides en phase solide la plus utilisée. Ceci est dû au fait que les thioesters subissent une aminolyse en présence d'agents de déprotection nucléophiles comme la pipéridine. La puissance de la NCL pour l'obtention de protéines synthétiques, a stimulé la recherche de nouveaux précurseurs de thioesters compatibles avec la Fmoc-SPPS par exemple les peptides *N*-acetylbenzimidazolinone (Nbz)¹⁹² et les peptides hydrazides⁶¹. En parallèle, l'introduction de nouveaux systèmes de réarrangement de type *O,S*,⁶⁶⁻⁶⁸ *N,S* ou *N,Se*⁷⁰ permet d'envisager la formation *in situ* de ces thioesters ou sélénioesters à partir de précurseurs également produits par Fmoc-SPPS⁶⁹ (vu au Chapitre I section 4.2).

Il est clair que l'introduction de nouveaux précurseurs de thioesters a permis de stimuler l'utilisation de la NCL. Si l'on observe en détail l'année 2016, on peut voir que près de 30 % des ligations impliquent des peptides hydrazides comme précurseurs de peptides thioester. Si l'on combine ce pourcentage à celui des systèmes de migration d'acyle et peptides Nbz, on atteint plus de la moitié des ligations réalisées sur l'année (54%), le reste correspondant à l'utilisation de thioesters classiques.

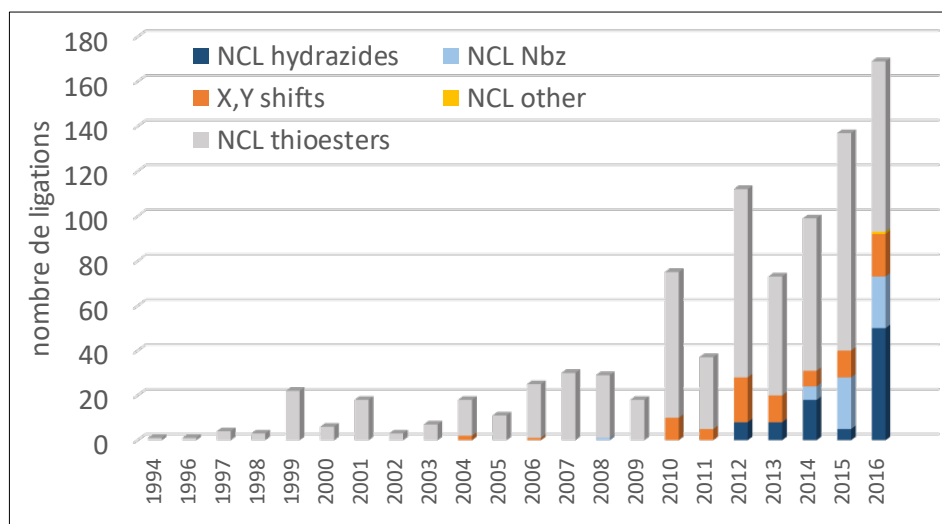


Figure 49 : Nombre de ligations via des thioesters classiques, Nbz, hydrazides ou de type transfert d'acyle X,Y.

On peut retenir de cette analyse que le développement de systèmes de réarrangement de type *N,Se* est encore très peu développé, alors que les sélénioesters constituent des donneurs d'acyle

extrêmement prometteurs. Une partie de mon travail de thèse a donc consisté à explorer l'intérêt de ces systèmes pour la synthèse totale de protéines.

4. Les types de jonctions Xaa-Cys

Lors d'une réaction de ligation, la nature du donneur d'acyle (peptide thioester ou équivalent) est très importante. Les cinétiques des réactions dépendent de l'encombrement stérique et des propriétés électroniques de la chaîne latérale de l'acide aminé en C-terminal du thioester. Les acides aminés β branchés (Val, Thr, Ile) sont connus pour induire un encombrement stérique important, les vitesses de réaction pour de tels acides aminés sont de plus de vingt-quatre heures pour une NCL réalisée dans des conditions classiques¹⁹³. Par ailleurs, les acides aminés asparagine (Asn), acide aspartique (Asp), acide glutamique (Glu) et glutamine (Gln) peuvent entraîner des réactions secondaires (hydrolyse¹⁹⁴, racémisation, formation de succinimide¹⁹⁵, formation d'isopeptide¹⁹⁶) et nécessiter des précautions particulières comme la protection de la chaîne latérale. Enfin, la proline est peu réactive en raison de facteurs stéro-électroniques et peut entraîner la formation de produits secondaires présentant une délétion d'acides aminés (réaction secondaire impliquant la formation d'intermédiaires de type dicétopipérazine¹⁹⁷).

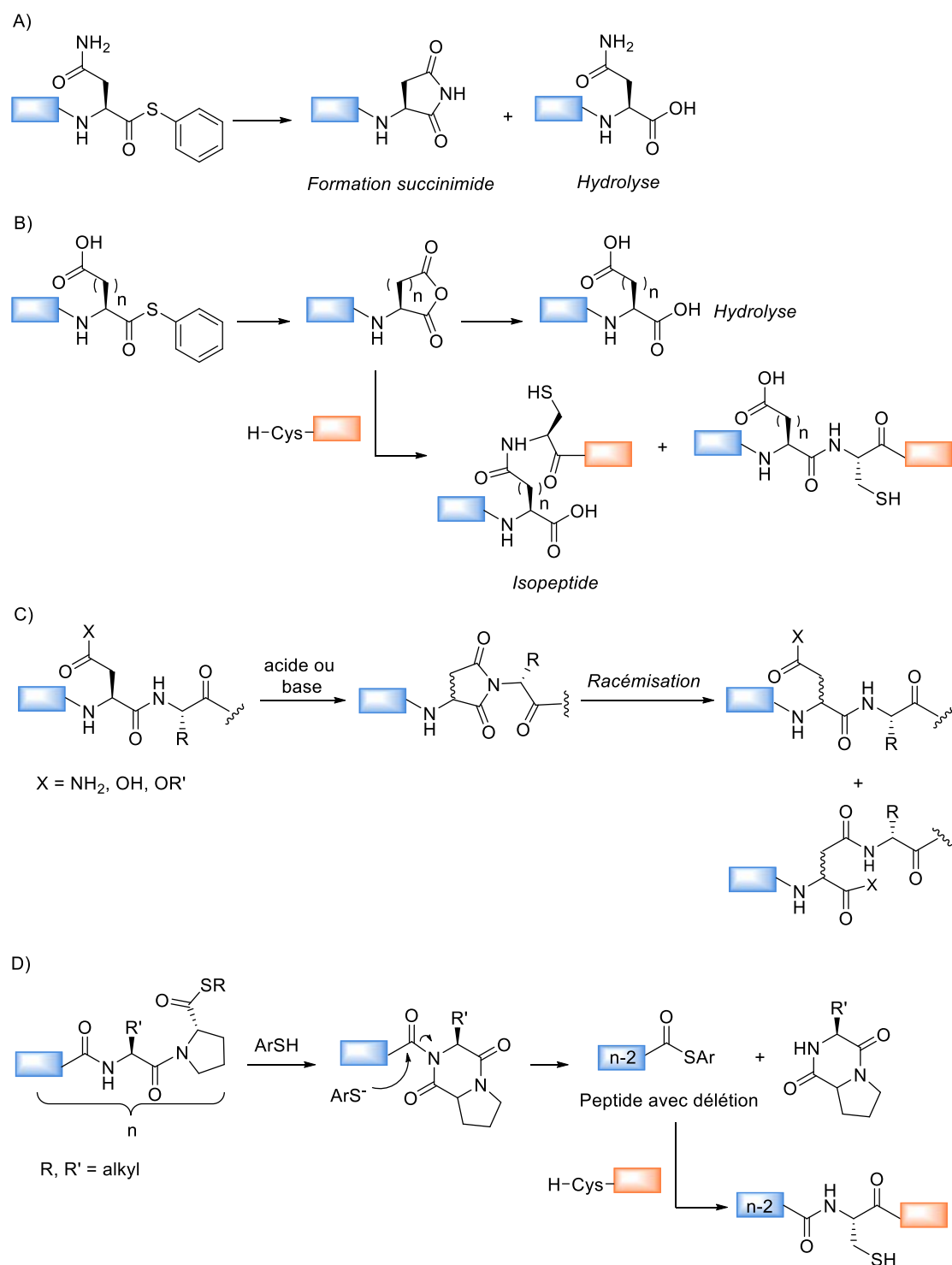


Figure 50 : Réactions secondaire A) Cas d'Asn/Gln. B) Cas d'Asp/Glu. C) Cas d'Asp/Asn. D) Cas de la proline.

Bien que certains travaux aient permis des avancées dans la formation de jonctions difficiles ou problématiques^{7,198,199}, ce type de ligation reste une limitation importante dans le domaine. En effet, si l'on observe le nombre de ligations réalisées à des jonctions de type Xaa-Cys en fonction de la fréquence naturelle des acides aminés (Figure 51), on note une déviation fortement négative pour les

acides aminés cités précédemment (Xaa encadrés en rouge dans la Figure 51). À l'inverse, les acides aminés alanine et glycine présentent une déviation nettement positive. D'après les travaux de T. M. Hackeng¹⁹³, ces deux acides aminés présentent les cinétiques de réactions les plus favorables.

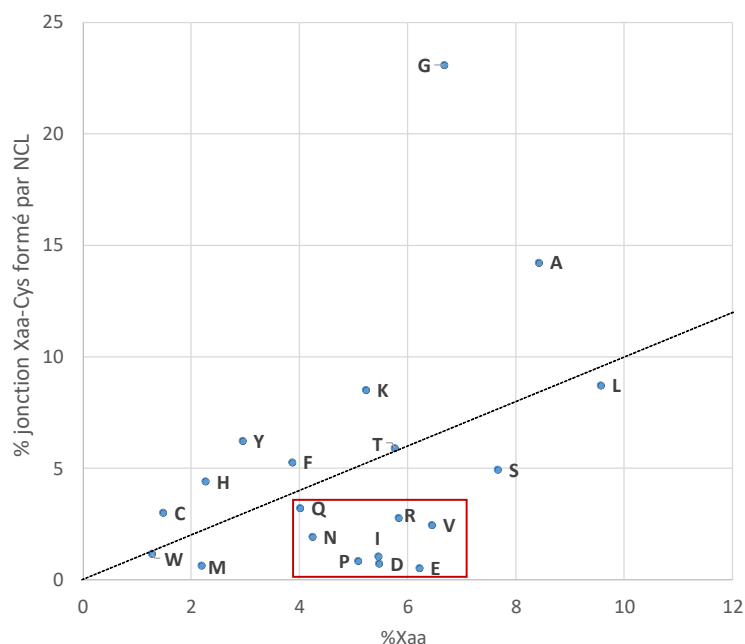


Figure 51 : Représentation de la fréquence de jonctions Xaa-Cys formées par ligation NCL (et méthodes apparentées) en fonction de la fréquence naturelle de l'acide aminé Xaa concerné.

Cette analyse révèle que des avancées méthodologiques sont nécessaires pour surmonter les difficultés liées à la formation de jonctions difficiles. C'est pourquoi je me suis attachée pendant ma thèse à développer de nouveaux catalyseurs pour accélérer la formation de ce type de liaisons peptidiques, notamment celles impliquant des acides aminés encombrés stériquement (Thr, Ile, Val).

5. Taille des protéines produites par synthèse chimique

Un autre aspect important au regard des travaux qui ont été réalisés pendant ma thèse concerne l'évolution de la taille des protéines produites par synthèse chimique au cours des dernières décennies (Figure 53). On peut remarquer que la moyenne et la médiane de taille varient très peu. Elles se situent aux alentours de 80 acides aminés. Il est clair que la taille des plus grandes protéines synthétisées a augmenté significativement au cours des dernières années comme le montre l'évolution du 3ème quartile Q3 (donnée de la série qui sépare les 25 % supérieurs des données). Ceci traduit le développement de nombreuses méthodes permettant de surpasser les limitations de la NCL (par exemple procédé de déchalcogénéation) ou de faciliter l'accès aux segments mis en jeu (précurseurs de thioesters). Cependant, leur impact sur l'ensemble du domaine n'est pas suffisant au regard de la faible variation de médiane dans la taille des protéines produites de 1994 à 2017. En effet, ces nouvelles technologies permettent de résoudre le problème de découpe des protéines en segments de taille

raisonnable mais ne permettent pas de résoudre la problématique d'assemblage. De ce fait, la majorité des protéines sont synthétisées en utilisant une seule étape de ligation. La Figure 53 présente le nombre de ligations utilisées pour synthétiser des protéines en fonction de l'année de publication. On remarque qu'avant 2010 la médiane et la moyenne étaient proche de 1. Ces dernières années, bien que la médiane soit égale à 1, la moyenne s'approche de la valeur 2 indiquant que le nombre de publications utilisant deux ligations ou plus a considérablement augmenté. Ceci est probablement dû à l'introduction de procédés *one-pot*, cinétiquement contrôlés ou en phase solide (pour une revue récente voir référence²⁰⁰) qui permettent de faciliter l'assemblage consécutif de segments peptidiques.

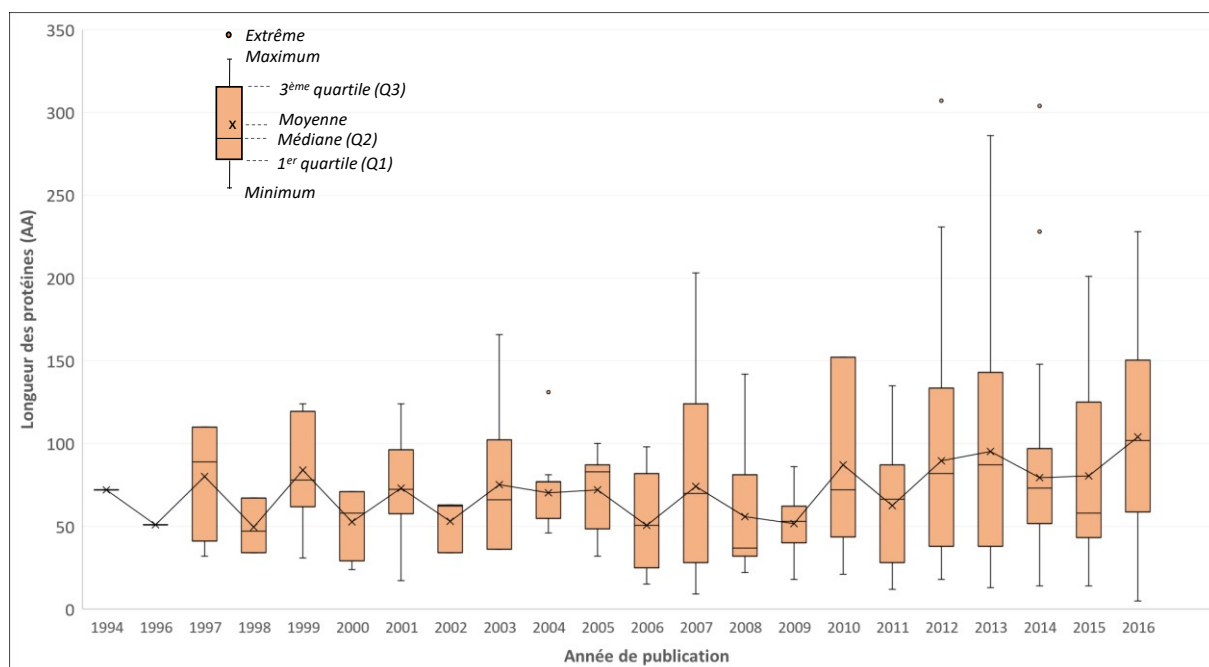


Figure 52 : Taille des protéines produites par synthèse chimique en fonction de l'année de publication.

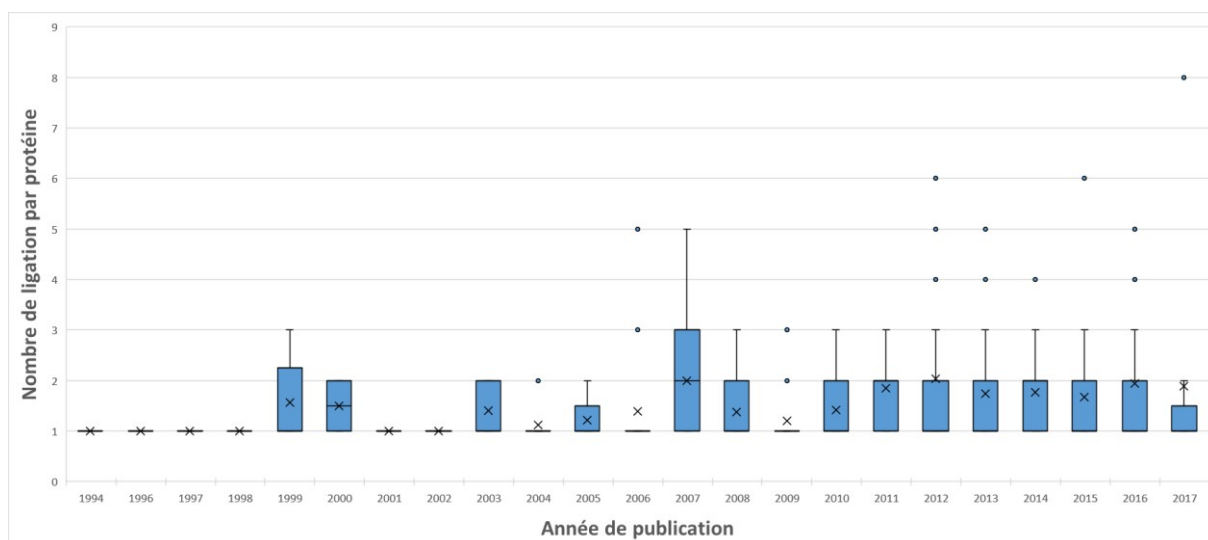


Figure 53 : Nombre de ligations utilisées pour synthétiser des protéines en fonction de l'année de publication. Pour les années 1998, 2003, 2005, 2008, et 2010-2017, la médiane est égale à 1 et n'est donc pas visible sur le graphe.

6. Bilan/Conclusion

L'étude statistique présentée dans ce chapitre a pour but de fournir un état de l'art de la synthèse par voie chimique des protéines. Ce travail sur l'utilisation de la NCL et des méthodes apparentées a été publié dans le journal *Bioorganic & Medicinal Chemistry* (Annexe B).

La base de données PCS fournit un certain nombre d'informations sur la taille, le nombre et le type de protéines produites par synthèse chimique. Elle permet également de mettre en avant les tendances récentes et d'identifier les outils chimiques les plus utilisés pour accéder aux protéines cibles. L'analyse statistique réalisée met également en évidence les limitations actuelles du domaine.

Malgré les progrès significatifs réalisés au cours des dernières années (substituts de thioester, désulfuration, etc.), il y a peu d'exemples dans la littérature de l'utilisation de la NCL pour des protéines de tailles significativement supérieures à 100-200 acides aminés. De fait, les problématiques d'assemblage demeurent une limitation importante pour l'accès à des protéines de grandes tailles. La multiplication des étapes chimiques et de purification conduit notamment à une baisse significative de rendement. Le développement de méthodes permettant de réaliser des assemblages consécutifs de segments peptidiques est au cœur de la problématique de ma thèse.

Une partie de mon travail a porté sur le développement de procédés *one-pot* utilisant des systèmes de transfert *N,Se* pour la synthèse totale de protéines. Les sélénioesters sont des donneurs d'acyles très réactifs et très prometteurs pour la conception de nouvelles méthodologies d'assemblage. Nous avons notamment démontré leur utilité en réalisant la synthèse de la protéine NK1-B (Chapitre IV).

En parallèle, les propriétés particulières des sélénioesters ont été exploitées pour faciliter la formation de jonctions difficiles liées à l'encombrement stérique (Val, Thr, Ile). Nous présentons dans le **Chapitre V**, la conception de nouveaux catalyseurs à base de sélénium permettant d'accélérer la conversion des peptides SEA en alkylthioesters et la ligation SEA.

Enfin, nous nous sommes intéressés au développement d'une stratégie d'assemblage séquentielle de segments peptidiques en phase solide. Le principe de ligation séquentielle de segments peptidiques en phase solide (SPCL) a été introduit par l'équipe de Kent en 1999²⁰¹. Malgré le fort potentiel de la phase solide, nous n'avons recensé que douze protéines synthétisées en utilisant des méthodes de ligation supportée, dont trois en 2016²⁰² et trois en 2017^{203,204}. La mise en place méthode de SPCL nécessite la mise au point de supports solides qui soient stables dans les conditions d'élongation et qui permettent de libérer la protéine dans des conditions douces. Ce type de stratégie nécessite également de trouver un moyen simple et efficace de générer des peptides thioesters en phase solide. À l'heure actuelle, peu de méthodes de ligation utilisées en solution peuvent répondre à ces critères. C'est pourquoi, nous avons choisi d'utiliser les propriétés uniques du groupement SEA pour développer un nouveau procédé d'assemblage séquentielle de segments peptidiques en phase solide. Ce nouveau procédé est décrit dans le **Chapitre VI**.

RÉSULTATS ET DISCUSSION

CHAPITRE IV : STRATÉGIE D'ASSEMBLAGE « ONE-POT » UTILISANT DES PEPTIDES SÉLÉNOESTER DE TYPE *BIS(2-SÉLÉNYLETHYL)AMIDO (SeEA)*

1. La problématique d'assemblage de multiples segments peptidiques

Dans la première partie de cette thèse, nous avons vu que l'introduction de nouveaux précurseurs de thioesters ou le développement de procédés de désulfuration ont permis de surmonter des limitations importantes de la réaction de NCL. Cependant, l'assemblage successif de segments peptidiques reste un défi. En particulier, nous avons observé que la taille moyenne des protéines synthétisées et le nombre de ligations utilisées pour y parvenir ont très peu variés depuis les années 1990. Ceci traduit l'absence de méthodes d'assemblage simples et robustes.

La multiplication des étapes chimiques et des étapes de purification est au cœur de la problématique d'assemblage. C'est pourquoi, une partie de mon travail de thèse a consisté au développement de procédés *one-pot* basés sur l'utilisation de systèmes de transfert *N,Se* pour la synthèse totale de protéines. En effet, des travaux très récents ont introduits les sélénosters comme donneurs d'acyles pour la réaction de NCL. Ces sélénosters présentent une très grande réactivité et un fort potentiel pour le développement de nouvelles méthodologies d'assemblage.

2. L'émergence des sélénosters comme donneurs d'acyle dans la réaction de NCL

En 2011, Durek et Alewood sont les premiers à introduire les sélénosters comme donneurs d'acyle dans la réaction de NCL¹⁹⁸. Les auteurs ont étudié la réactivité de peptides thio- ou sélénosters LYRAX (X= Phe) avec un cystéinyl peptide CFRANK dans une réaction de NCL (Figure 54). Le sélénoster réagit 1,9 fois plus rapidement que le thioester dans les mêmes conditions de réaction (Condition A). L'échange thiol-sélénoster avec l'arylthiol MPAA est plus rapide que l'échange thiol-thioester. Les auteurs notent d'ailleurs que cet échange est tellement rapide que la réaction de transthioestérification de l'arylthioester de MPAA avec le cystéinyl peptide devient cinétiquement limitante. Face à ce constat, les auteurs ont remplacé le MPAA par le sélénophénol, ceci, afin d'introduire un catalyseur qui permette de conserver la grande réactivité des sélénosters. L'utilisation de sélénophénol conduit à une augmentation de la vitesse de réaction d'un facteur 9 (Condition B). Cependant, dans ces conditions, la réaction s'accompagne de l'accumulation de thioesters improductifs impliquant les cystéines internes. Cette réaction secondaire est due à la faible nucléophilie du sélénophénol, qui est incapable de réagir avec des thioesters d'alkyles et donc d'inverser la formation de thioesters improductifs.

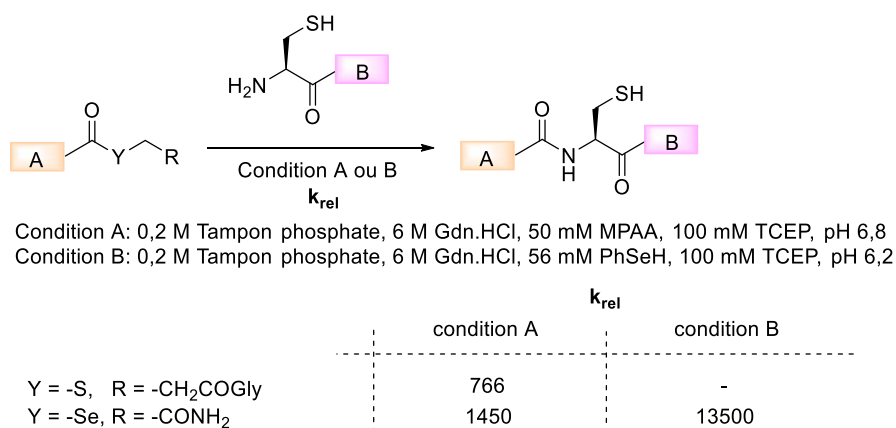


Figure 54 : Ligation de peptides thio- ou sélénosters LYRAF-S(Se)R avec CFRANK.

La grande réactivité des sélénosters en comparaison avec les thioesters est probablement due au fait que les sélénosters sont de meilleurs donneurs d'acyle que les thioesters²⁰⁵. En effet, les sélénols ont des pK_a significativement plus bas que leurs analogues thiols²⁰⁶. Le groupement sélénoste est un donc un meilleur nucléofuge que le groupement thiolate correspondant. Par conséquent, l'échange thiol ou sélénol/sélénoste avec un catalyseur de type arylthiol (MPAA) ou arylsélénol (PhSeH) doit procéder plus rapidement que pour les thioesters. Les principales propriétés de ces catalyseurs sont présentées dans la Figure 55.

Structure	Nom	pK _a	Propriétés	Référence
	Sélénophénol	5,9 ²⁰⁷ : voir note ^c 4,6 ²⁰⁹	Généré <i>in situ</i> à partir du diphényldisélénide DPDS et TCEP. DPDS est faiblement soluble dans l'eau.	198
	MPAA	6,6 ¹⁰³	Soluble dans l'eau jusqu'à pH 5,5. Non odorant.	198

Figure 55 : Catalyseurs utilisés pour la NCL avec des sélénosters.

Les sélénosters présentent un fort potentiel pour permettre la synthèse totale de protéines par NCL. Cependant, il n'existe que très peu de méthodes pour permettre leur synthèse et ils sont de

^c Le pK_a du PhSeH a été estimé par Fagioli et ses collaborateurs en utilisant une méthode électrochimique²⁰⁸. La valeur déterminée dans cette étude est probablement sujette à une erreur expérimentale importante.

surcroit très sensible aux nucléophiles²¹⁰. Durek et Alewood ont développé une méthode en solution dans laquelle des peptides thioesters sont d'abord convertis en sélénio-acides par traitement avec de l'hydrogénosélénure de sodium NaHSe à pH 7²¹¹. Cette étape est directement suivie par une réaction d'alkylation avec un halogénure d'alkyle à pH 4 pour obtenir les sélénioesters correspondants. Cette synthèse doit être réalisée en *one-pot* et sous atmosphère inerte en raison de l'instabilité des intermédiaires sélénio-acides.

Pour surmonter cette difficulté, les auteurs ont développé une stratégie en phase solide par Boc-SPPS. Ils ont introduit pour cela une résine polystyrène fonctionnalisée par un linker HF-labile de type HSe-(CH₂)₂-CONH-CH(R)-COO-CH₂-PAM-polystyrène. La résine est ensuite réduite en présence de DTT et acylée par l'introduction du premier acide aminé dont l'amine α est protégée par le groupement Boc. La suite de l'élongation est réalisée par Boc-SPPS classique. Les auteurs ont utilisé ce type de résine pour réaliser la synthèse d'un peptide sélénioester issu d'une toxine d'araignée (*Phoneutria nigriventer*, PnTx2-6). Ce peptide a été obtenu avec un rendement de 71 % démontrant ainsi l'efficacité de la méthode. L'utilisation de la stratégie Boc-SPPS implique l'utilisation de fluorure d'hydrogène qui est un composé dangereux et très toxique. Pour étendre l'utilisation de sélénioesters dans des réactions de NCL, le développement de méthodes compatibles avec la Fmoc-SPPS, qui est la méthode de synthèse peptidique la plus utilisée, est donc essentiel.

En réponse à cette problématique, nous nous sommes intéressés à l'utilité du groupement *bis*(2-sélényléthyl)amido (SeEA), à savoir l'analogue sélénié du groupement SEA, pour accélérer la formation de liaisons peptidiques par formation d'intermédiaires sélénioesters. Dans un premier temps, nous avons développé un procédé chimiosélectif d'échange pour convertir les peptides SEA en peptides SeEA. La réactivité particulière de ces nouveaux dérivés nous a permis de développer de nouveaux procédés d'assemblage, en particulier des stratégies *one-pot* et cinétiquement contrôlées. La force de ces nouveaux concepts a été illustrée par la synthèse de la protéine NK1 biotinylée (**NK1-B**, 180 acides aminés), qui a été réalisée par assemblage de 6 segments peptidiques. Ces travaux sont présentés dans la suite de ce chapitre.

3. La synthèse des peptides *bis(2-sélényléthyl)amino (SeEA)*

Les peptides SEA sont facilement accessibles par Fmoc-SPPS grâce à des résines SEA de type polystyrène (PS) ou polyéthylène glycol (PEG) mises au point au laboratoire^{17,212}. C'est pourquoi, nous nous sommes demandé s'il était possible de convertir ces peptides SEA en leurs analogues sélénés. En effet, les peptides amide SEA **1** sont en équilibre avec la forme thioester de SEA **2** dans l'eau (Figure 56). Nous avons pensé que ce dernier pourrait participer à une réaction d'échange sélénol-thioester avec un excès de *bis(2-sélényléthyl)amine* **3** pour produire, après un transfert d'acyle $N \rightarrow Se$, le peptide SeEA cible **5**. Ce procédé de conversion correspond en fait à une réaction de transamidation chimiosélective.

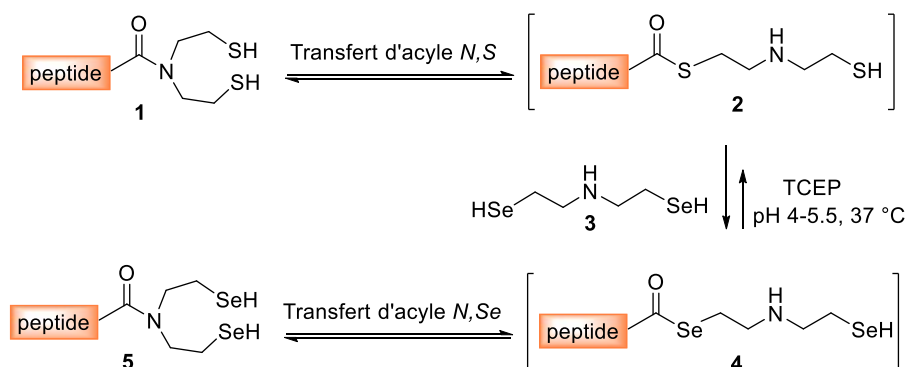


Figure 56 : Conversion des peptides SEA en peptides SeEA par transamidation.

Le sel d'acide trifluoroacétique du composé cyclique disélénié **7a** a été envisagé comme précurseur pour la *bis(2-sélényléthyl)amine* **3**. La synthèse de ce dernier a été décrite par le laboratoire dans des travaux antérieurs²¹³ et est présentée dans la Figure 57. Dans un premier temps, la réduction de sélénium métallique par du borohydrure de sodium conduit à la formation majoritaire de disélénuire de disodium Na_2Se_2 . Les ions disélénures Se_2^{2-} ainsi formés réagissent avec la *bis(2-chloroéthyl)amine* **6** pour conduire par substitution nucléophile et après purification par RP-HPLC au sel d'acide trifluoroacétique du composé cyclique disélénié **7a**. Cette réaction conduit également à la formation du composé cyclique trisénié **8**.

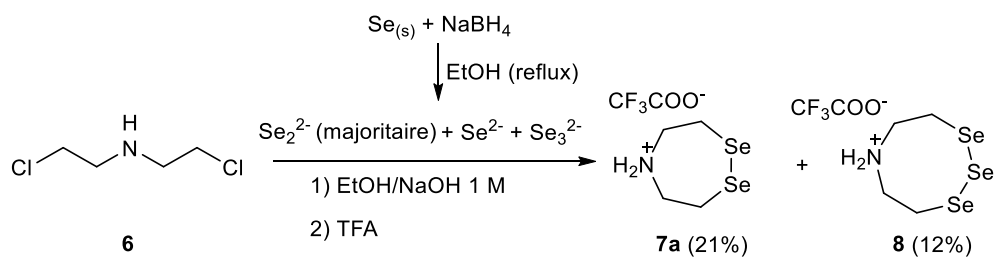


Figure 57 : Synthèse des composés cycliques di- ou triséniés **7a** et **8**.

Dans un premier temps, nous avons réalisé la réaction d'échange avec le composé disélénié **7a** à pH 4,0 et en présence de TCEP (Figure 58A). Cette réaction a conduit à une faible proportion de peptide SeEA **10** (Figure 58B). Nous avons également observé une réaction secondaire qui conduit à la formation majoritaire d'un composé dont la masse correspondant à la perte d'un sélénium. Nous supposons par conséquent que le peptide obtenu correspond à la structure décrite pour le peptide **11** (Figure 58A). Cette réaction d'extrusion d'un atome de sélénium de composés diséléniures conduisant à la formation de sélénoéthers a été intensivement étudiée par le passé. Comme nous l'avons vu précédemment (Chapitre II, section 4.3.3), ce mécanisme est catalysé par la présence d'une phosphine tertiaire comme la TCEP. L'utilisation de DTT comme agent réducteur devrait être une solution à ce problème. Malheureusement, le DTT n'est pas capable de réduire des diséléniures cycliques de type SeEA, il faut donc trouver des conditions expérimentales permettant l'utilisation de TCEP tout en évitant le phénomène d'extrusion du sélénium.

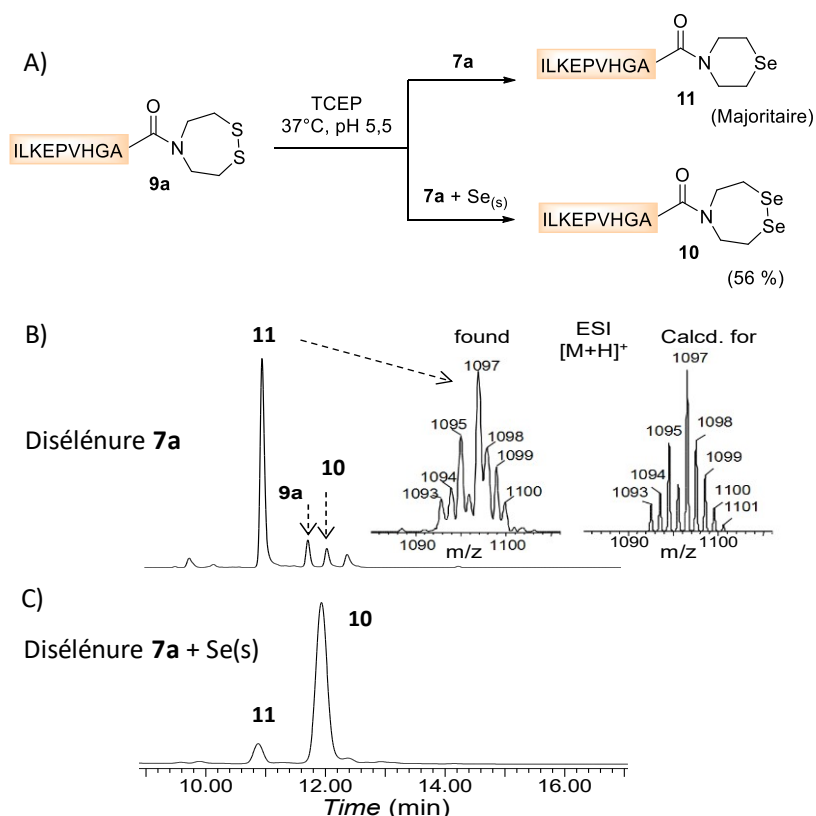


Figure 58 : A) Conversion du peptide SEA en dérivés séléniés. HPLC du milieu réactionnel avec le diséléniure **7a** (B), le diséléniure **7a** et du sélénium métallique (C).

Des travaux récents du laboratoire ont montré que la Se=TCEP permet de prévenir la déséléniation de la sélénocystéine par la TCEP²¹⁴. Par conséquent l'utilisation de Se=TCEP dans le milieu réactionnel devait permettre d'inhiber la formation du peptide **11**. Pour vérifier cette hypothèse, nous avons réalisé la réaction d'échange en utilisant le composé **7a** et du sélénium

métallique. Ce dernier permet la formation *in situ* de Se=TCEP par réaction avec la TCEP. Ce protocole a permis d'isoler le peptide **10** avec un bon rendement (56 %) (Figure 58A et C).

L'échange chimiosélectif entre un peptide SEA et la *bis*(2-sélényléthyl)amine **3** générée *in situ* par la réduction du disélénaire **7a** constitue un moyen simple de synthèse de sélénioesters latents. La synthèse des peptides SEA est compatible avec la Fmoc-SPPS et la conversion de ceux-ci en peptide SeEA s'effectue dans des conditions douces. De plus, les peptides SeEA sont obtenus sous forme cyclique et sont donc très stables durant les purifications HPLC et le stockage.

Finalement, les différences de propriétés décrites dans la littérature entre des composés sélénés et soufrés nous amènent à nous poser la question de la réactivité des peptides SeEA^{215,216}. Dans la suite de ce manuscrit, nous allons nous intéresser en particulier aux propriétés oxydo-réductrices du groupement *bis*(2-sélényléthyl)amine SeEA.

4. Les propriétés oxydoréductrices du groupement *bis*(2(sélényléthyl)amine (SeEA)

De façon générale, les séléniols sont connus pour être beaucoup plus réducteurs que les thiols²¹⁷. En d'autres termes, la réduction de disélénaires nécessite l'utilisation d'agents réducteurs plus forts par rapport aux disulfures. Par exemple, le potentiel redox de la cystine ($E^\circ = -238$ mV) est plus élevé que celui de la sélélocystine ($E^\circ = -383$ mV). De façon intéressante, Iwoka et ses collaborateurs ont montré que le disélénaire cyclique disélénothréitol (DST^{ox}) n'est pas réduit en présence d'un large excès de DTT (Figure 59)²¹⁸.

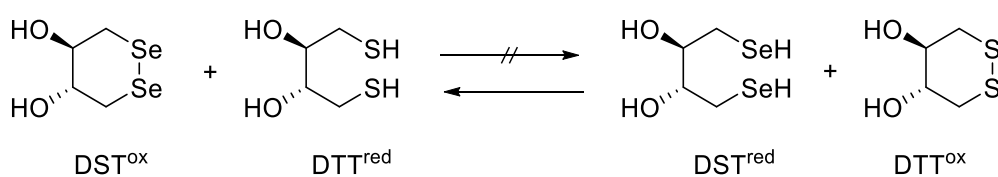


Figure 59 : Équilibre de réduction entre les espèces dithiothréitol (DTT) et disélénothréitol (DST).

Face à ces observations, nous nous sommes demandé si nous pourrions réduire et donc activer sélectivement le groupement SEA en présence de son analogue sélénié en utilisant du DTT. Lors de travaux antérieurs, Raibaut et ses collaborateurs ont étudié la stabilité de ce groupement SeEA vis-à-vis des différents réducteurs utilisés lors des réactions de NCL et de ligation SEA²¹³. Pour cela, ils ont réalisé l'expérience décrite dans la Figure 60. Ils ont mélangé trois segments peptidiques : un peptide SEA (A), un peptide SeEA (B) et un cystéinyl peptide (C). En présence de MPAA, il n'y a pas de réactions observées (étape 1, Figure 60). Les deux peptides SEA et SeEA sont stables et se comportent comme

des thio(sélén)esters latents. Après l'ajout de DTT au milieu réactionnel, on observe la formation exclusive du produit AC due à la ligation entre le peptide SEA (A) et le cystéinyl peptide (C) (étape 2, Figure 60). Le segment SeEA reste totalement stable dans ces conditions. Finalement, l'ajout de TCEP déclenche la réaction entre le peptide SeEA (B) et le cystéinyl peptide (C) pour former le produit de ligation BC (étape 3, Figure 60). Par conséquent, seule la TCEP, qui est un réducteur plus fort que le MPAA et le DTT, permet la réduction du pont disélenure et donc la ligation SeEA.

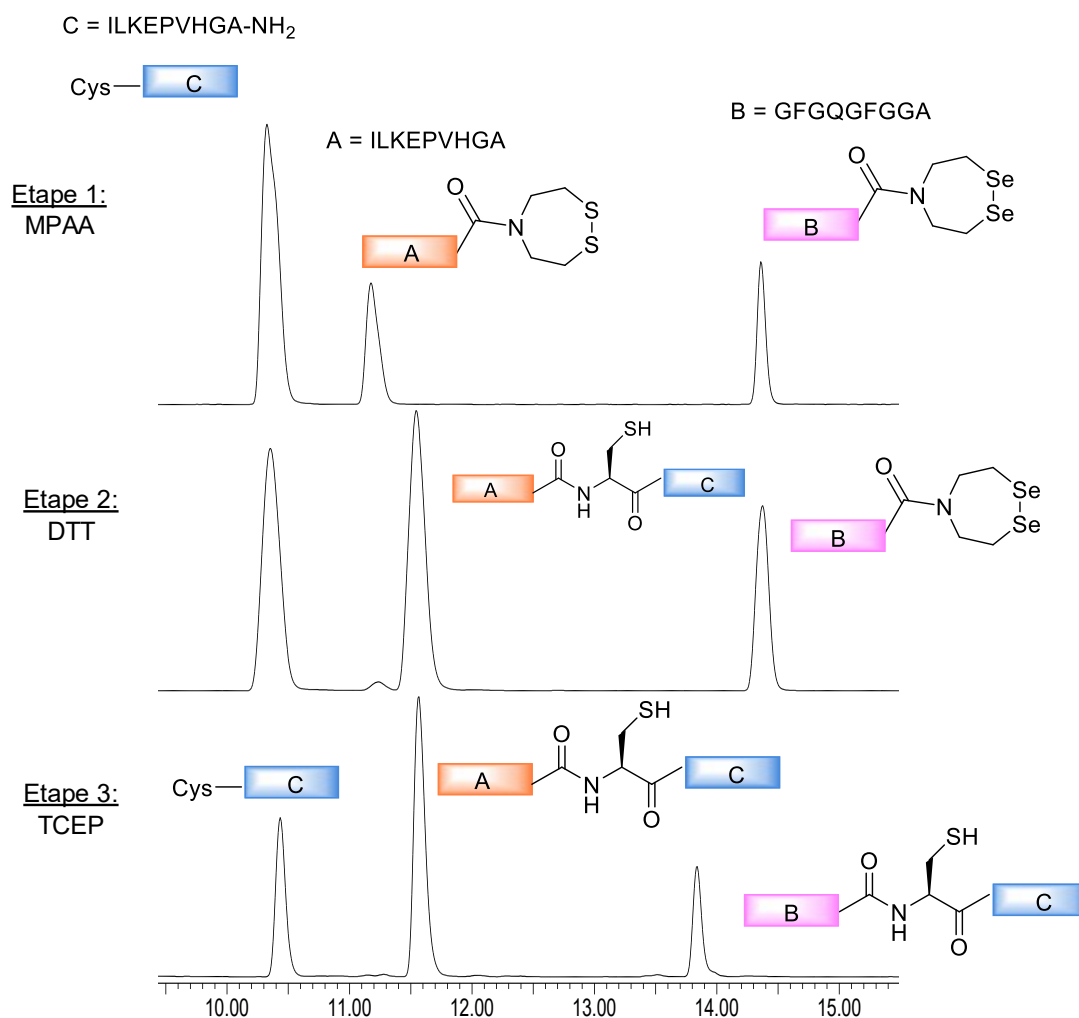


Figure 60 : Concept SEA/SeEA : activation sélective et séquentielle par ajout de différents réducteurs. Suivi HPLC ($\lambda = 215 \text{ nm}$).

Cette expérience, nous montre qu'il est possible de réduire sélectivement l'entité soufrée par le DTT en présence du dérivé sélénié. Les propriétés latentes des thio(sélén)esters de SEA ont permis à Raibaut et ses collaborateurs de réaliser l'assemblage *one-pot* de quatre segments peptidiques en utilisant séquentiellement la NCL, la ligation SEA et la ligation SeEA. Nous verrons plus tard dans ce document comment ces propriétés particulières nous ont permis de synthétiser la protéine NK1.

5. La réactivité des peptides *bis*(2(sélényléthyl)amido (SeEA)

Nous avons vu précédemment que les sélénioesters réagissent plus vite que les thioesters en particulier dans la réaction de NCL. Au cours de notre travail, nous nous sommes donc intéressés à évaluer la réactivité des peptides SeEA lors de réaction de ligation. Pour cela, nous avons réalisé une étude cinétique des réactions des peptides SEA, SeEA et thioester de MPA avec un cystéinyl peptide (Figure 61A). Les réactions ont été réalisées dans les conditions classiques de NCL c'est-à-dire à pH 7,2 en présence de MPAA (100 mM) et de TCEP (200 mM). Dans le cas du peptide SeEA, nous avons ajouté de la Se=TCEP (80 mM) afin d'empêcher le processus de désélénisation. Le suivi des réactions a été réalisé par HPLC et les données obtenues ont été corrélées avec une loi d'ordre 1 (voir note^d) (Figure 61B). Les temps de demi-réaction $t_{1/2}$ qui correspondent au temps nécessaire pour atteindre 50 % de conversion ont été déterminés. Ainsi, on observe que le peptide SeEA **10** réagit beaucoup plus rapidement que le peptide SEA **9a**. Dans le cas présenté par la Figure 61, les temps de demi-réaction pour les peptides SEA et SeEA sont respectivement de 3,3 h et 0,24 h soit un facteur d'accélération de la constante cinétique d'environ 14. Il est intéressant de noter que le peptide SeEA **10** réagit également plus rapidement que le thioester de MPA **12a**. Des résultats similaires ont été obtenus pour des peptides comportant un acide aminé encombré en C-terminal.

^d La procédure pour vérifier si la cinétique suit bien une loi du premier ordre est décrite dans la partie expérimentale (Annex A, Chapter IV, Section 3.3).

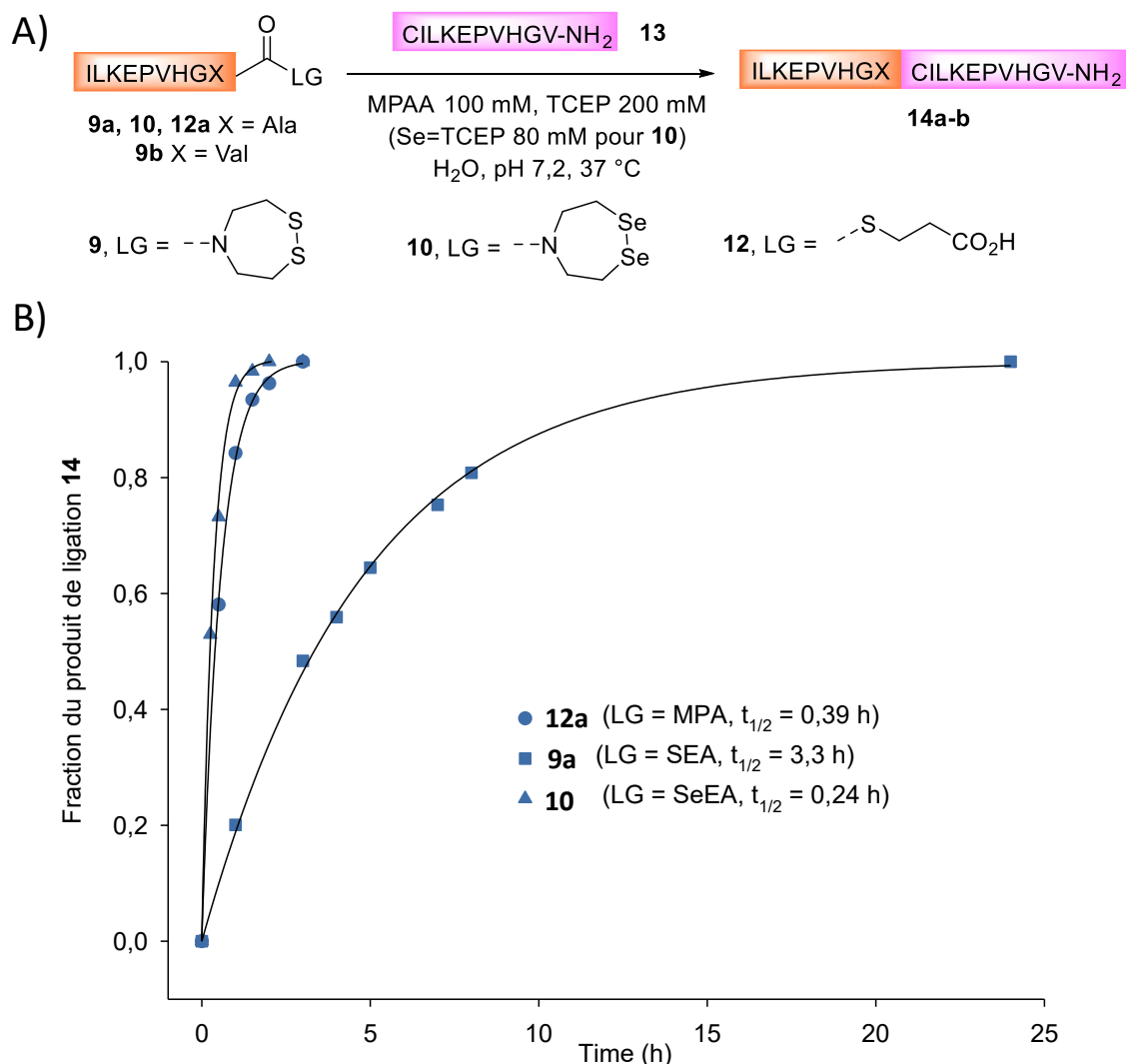


Figure 61 : A) Réaction de ligation des peptides SEA **9a**, SeEA **10** et thioester **12a** avec le cystéinyl peptide **13**. B) Données cinétiques (HPLC, $\lambda = 215$ nm). Les données ont été corrélées par une loi d'ordre 1, $t_{1/2} = \ln(2)/k$ avec k constante de vitesse en s^{-1} .

Pour expliquer cette différence de réactivité, nous nous sommes intéressés au mécanisme de la ligation SEA. De nombreuses expériences réalisées au laboratoire nous laissent penser que le transfert d'acyle de $N \rightarrow S$ est l'étape cinétiquement limitante de la ligation SEA. De plus, nous avons observé que la cinétique de la ligation SEA est accélérée avec la diminution du pH. Ce phénomène de catalyse acide a été observé pour des acides aminés faiblement encombrés (Ala) ou au contraire fortement encombrés (Val).

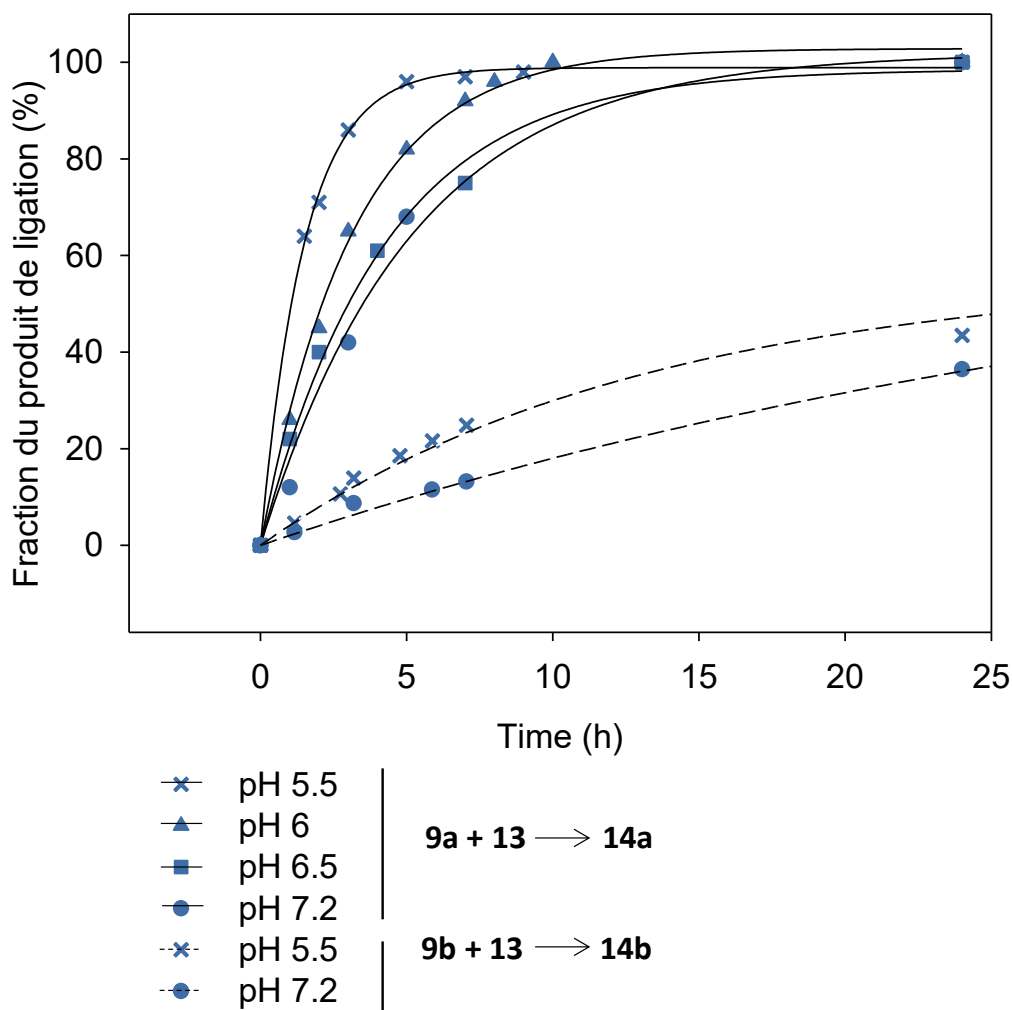


Figure 62 : Effet du pH sur la cinétique de ligation SEA (peptide **9a/9b** 7mM, peptide **13** 10,5 mM, MPAA 200 mM, TCEP 200 mM, 37°C. Suivi HPLC ($\lambda = 215$ nm), les données sont corrélées par une loi d'ordre 1.

Ces résultats expérimentaux sont en accord avec des expériences préliminaires de modélisation moléculaire réalisées en collaboration avec J-C. Monbaliu de l'Université de Liège visant à identifier un état de transition pour la réaction de migration d'acyle *N,S*. La *N*-formyl glycine SEA **15a** a été utilisée comme modèle pour l'étude de modélisation. Comme pour les travaux de Wang sur la NCL (discutés au chapitre II section 2.1.1), deux états de transition concertés l'un neutre et l'autre monoanionique ont été identifiés. L'état de transition concerté monoanionique (Figure 63A) présente la barrière d'activation la plus faible. Cet état de transition montre deux phénomènes s'opérant simultanément : la protonation de l'azote de l'amide par l'un des groupements thiols, et une attaque nucléophile du carbonyle par le groupement thiolate du second bras. Le transfert intramoléculaire d'un proton du soufre vers l'azote est en adéquation avec les résultats expérimentaux d'accélération de la ligation SEA à pH acide. De plus, cette modélisation nous permet d'expliquer la plus grande réactivité des peptides SeEA. En effet, les sélénols sont plus acides que les thiols. Par exemple, le pK_a du groupement thiol de la cystéine ($pK_a \sim 8,3$) est de trois unités de pH supérieur à celui du groupement sélénol ($pK_a \sim 5,5$) de la

sélocystéine (Sec). Les pK_a pour des groupements S(Se)EA ont été estimés à l'aide du logiciel ACD/Labs (Figure 63B) et présentent également une grande différence de valeurs. Par conséquent, la protonation intramoléculaire de l'azote devrait être plus marquée dans le cas du sélénium. De plus, à pH neutre, les groupements sélénols du résidu SeEA sont majoritairement sous forme d'anion sélénoate tandis que les groupements thiols du SEA sont principalement sous forme RSH. Il y a donc une concentration d'espèces actives nucléophiles plus importantes pour les peptides SeEA et donc une attaque nucléophile du carbonyle plus efficace. Nous pensons que la combinaison de ces deux phénomènes permet d'expliquer la grande réactivité du groupement SeEA en comparaison de l'analogie soufré SEA.

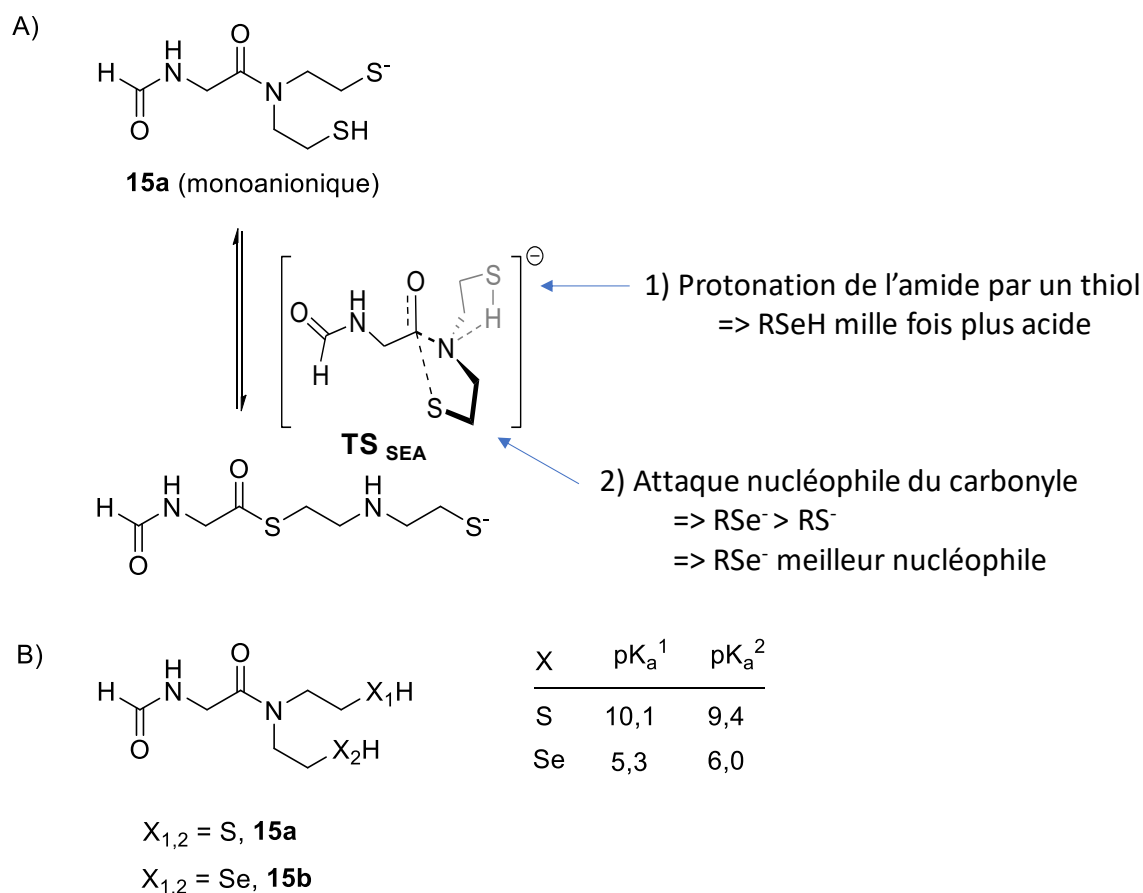


Figure 63 : A) Structure d'un état de transition monoanionique hypothétique pour le transfert d'acyle intramoléculaire N→S. B) pK_a estimés pour les groupements thiols (SEA) et sélénols (SeEA) en utilisant le logiciel ACD/Labs.

De façon intéressante, cette grande différence de réactivité nous permet d'envisager des processus cinétiquement contrôlés combinant les groupements SeEA et SEA dans cet ordre (KCL, Figure 64). Nous pouvons imaginer réduire et donc activer les deux segments **A** et **B** grâce à la TCEP (étape 1, Figure 64). Le segment SeEA **A** devrait réagir plus vite que le segment SEA **B** pour former l'intermédiaire **AB** avant que le segment **B** ne cyclise ou ne polymérise (étape 2, Figure 64). Finalement,

l'ajout du segment **C** permet de déclencher la réaction de ligation SEA conduisant à l'assemblage *one-pot* de trois segments peptidiques (étape 3, Figure 64).

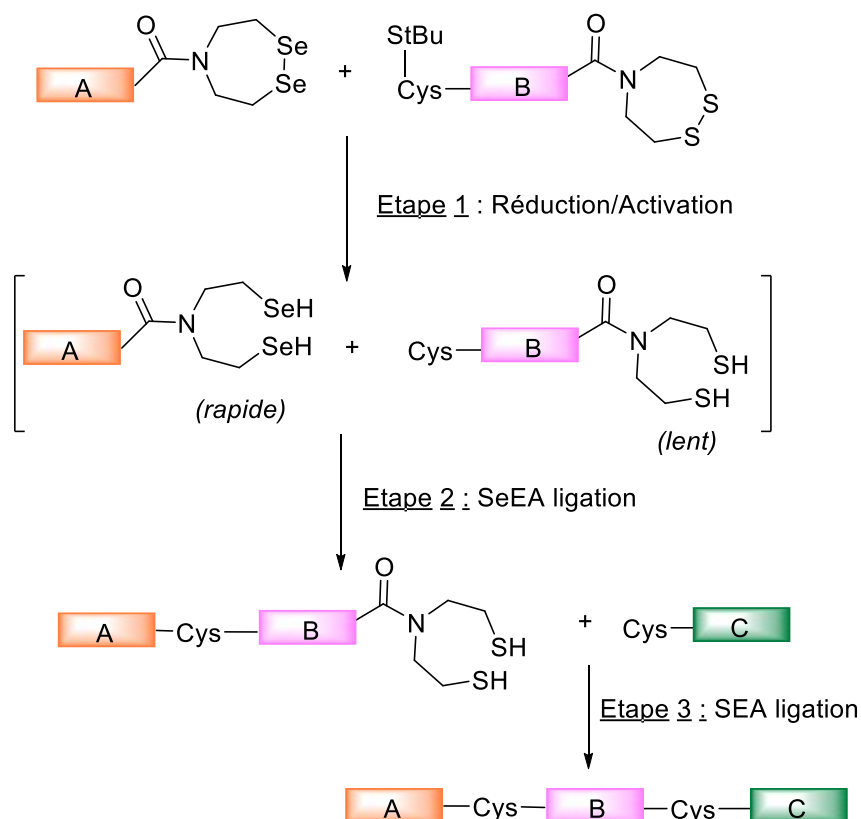


Figure 64 : Ligations cinétiquement contrôlées en utilisant la ligation SEA et SeEA.

Finalement, en utilisant des conditions cinétiquement contrôlées et les propriétés latentes des fonctions S(Se)EA, nous avons mis au point un nouveau procédé d'assemblage en solution. Ce procédé a été appliqué à la synthèse de la protéine NK1 et est décrit ci-après.

6. Assemblage de la protéine NK1

6.1 La protéine NK1 : un modèle de protéine thérapeutique

La protéine NK1 est un variant naturel du facteur de croissance des hépatocytes (HGF/SF, *Hepatocyte Growth Factor/Scatter Factor*). L'HGF/SF est le ligand du récepteur à tyrosine kinase MET. La signalisation HGF/MET induit de nombreux phénotypes cellulaires comme la prolifération, la migration, la morphogénèse ou la survie (Figure 65). D'autre part sa dérégulation peut conduire au développement tumoral et à la prolifération métastatique. Ce couple joue un rôle important dans le développement embryonnaire et la régénération tissulaire chez l'adulte. De fait, des analogues de l'HGF/SF possèdent un fort potentiel en thérapie régénérative. C'est pourquoi le couple HGF/SF-MET est intensivement étudié au laboratoire et a fait l'objet de plusieurs publications^{85,219}.

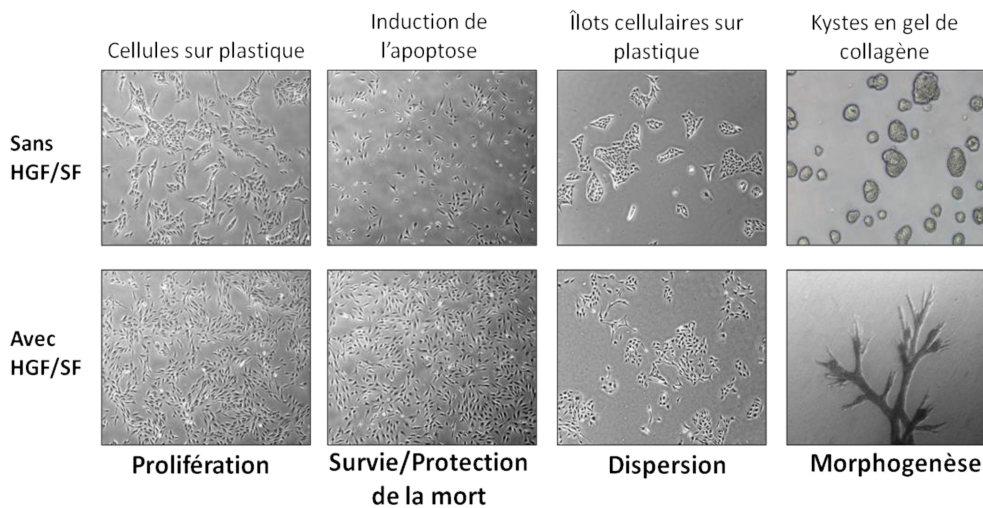


Figure 65 : Principales réponses biologiques associées au couple HGF/SF-MET sur des cellules épithéliales (MDCK) en culture.

L'HGF/SF est un hétérodimère de 90 kDa constitué d'une chaîne α et d'une chaîne β reliées entre elles par un pont disulfure. La sous-unité α est composée d'un domaine N-terminal (N) suivi par quatre domaines dits *Kringle* (K1-K4). La sous-unité β est un domaine homologue aux sérines protéases (SPH) (Figure 66). L'HGF/SF est faiblement actif *in vivo* par voie systémique. En effet, la présence du domaine N qui a la capacité de lier les héparanes sulfate avec une haute affinité, limite sa diffusion dans les tissus. De plus, l'HGF/SF peut être dégradé rapidement par les enzymes protéolytiques.

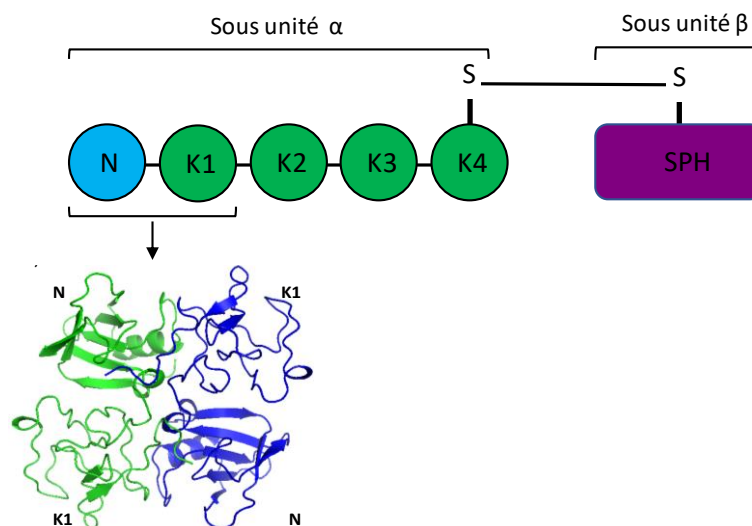


Figure 66 : Organisation de l'HGF/SF et structure tridimensionnelle du domaine NK1.

La protéine NK1 qui est constituée des domaines N et K1, porte le site de haute affinité de l'HGF pour MET et constitue un bon agoniste du récepteur MET. Si l'on observe la Figure 67, on remarque que la concentration d'HGF/SF nécessaire pour observer un phénotype est nettement plus faible que pour la protéine NK1. Bien qu'elle soit moins active que l'HGF/SF, des analogues de la protéine NK1

représentent de bons modèles thérapeutiques pour des applications en régénération tissulaire. En effet, la protéine NK1 induit des phénotypes similaires à l'HGF/SF *in vitro* et *in vivo* et représente la zone d'interaction minimale nécessaire et suffisante pour la liaison et l'activation du récepteur MET. Il est à noter que NK1 possède le domaine N qui est responsable de la faible diffusion de l'HGF/SF *in vivo*. Une tendance actuelle est à la conception d'analogues de la protéine NK1 présentant des mutations au niveau du site d'interaction avec les héparanes sulfate. Il est clair que dans ce contexte, le développement d'une méthode chimique permettant d'accéder très rapidement et facilement à des dérivés de la protéine NK1 présente un grand intérêt. Nous proposons, ici, de combiner les propriétés latentes des groupements SEA et SeEA pour assembler la protéine NK1.

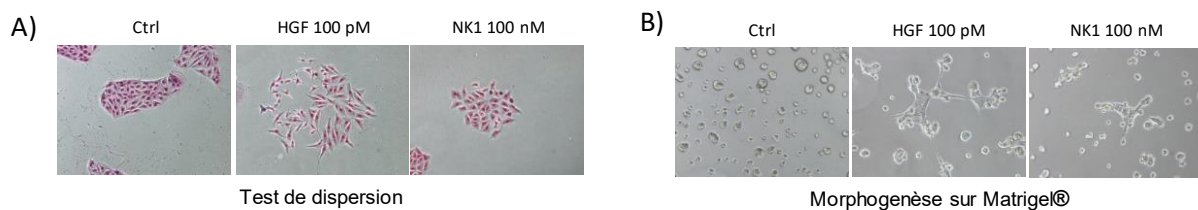


Figure 67 : Phénotypes cellulaires induits par l'HGF/SF et NK1²¹⁹. A) Test de dispersion sur des îlots cellulaires isolés de cellules MDCK incubés pendant 18 h. Les cellules sont ensuite colorées et observées sous microscope (40x). B) Test de morphogénèse sur des cellules de MDCK ensemencées sur une couche de Matrigel® pendant 18 h. Les cellules sont observées sous microscope (40x).

6.2 Synthèse totale de la protéine NK1

La protéine NK1 contient dix cystéines et est donc une cible de choix pour réaliser un assemblage par voie chimique en utilisant des méthodes de ligation de type NCL. Cette abondance en cystéines nous a permis de découper la protéine en six segments peptidiques de taille similaire. À noter que le dernier segment est équipé d'une biotine pour permettre des tests biochimiques. Dans un premier temps, les segments **16, 17, 20 et 22** ont été assemblés pour conduire à la formation d'un sélénoester latent SeEA de 118 acides aminés **23**. Cet assemblage a été réalisé en utilisant de façon séquentielle la NCL et la ligation SEA.

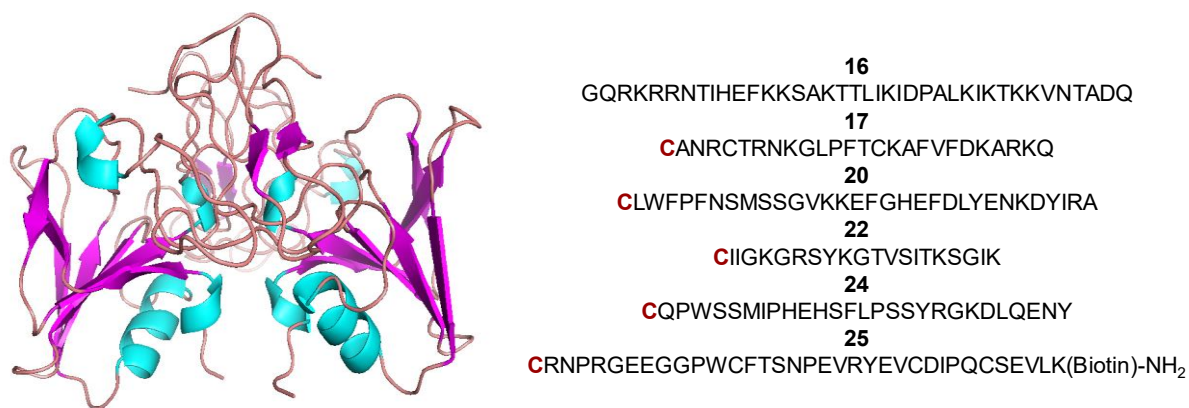


Figure 68 : Structure tridimensionnelle NK1 (PDB : 1BHT, Ultsch, M. Structure 1998, 6, 1383-1393) et les six segments peptidiques utilisés pour la synthèse totale de NK1.

6.2.1 Synthèse d'un sélénoester latent

Dans un premier temps, les segments **16** et **17** sont assemblés en utilisant une réaction de NCL. Cette ligation a lieu en présence du catalyseur MPAA mais en absence de réducteur (TCEP/DTT) afin de conserver le groupement SEA sous sa forme inactive. Une fois la réaction de NCL terminée, la fonction SEA est activée par ajout de TCEP. Le pH est ensuite ajusté à 4 afin de déclencher l'échange *in situ* entre l'intermédiaire thioester de SEA **18** et l'acide 3-mercaptopropionique (MPA). Ce procédé nous a permis de produire en une seule étape le peptide **19** avec une fonction thioester en C-terminal (*one-pot 1*, Figure 69). Après purification, le peptide **19** est engagé dans une seconde réaction de NCL avec le segment **20** en utilisant la fonctionnalité latente du SEA. Celui-ci est directement activé par ajout de DTT au milieu réactionnel pour réaliser la ligation SEA avec le segment **22** en une seule étape (*one-pot 2*, Figure 69).

La synthèse du segment SeEA **22** a également été réalisée en utilisant un procédé *one-pot* (*one-pot 3*, Figure 69). L'échange SEA → SeEA décrit précédemment a été optimisé et étendu à un peptide thioester de MPA, typiquement le segment **21**. Le segment **21** est équipé d'un groupement protecteur acétoacétyle à l'extrémité N-terminale. Ce groupement a été utilisé au laboratoire pour réaliser des processus de ligation/cyclisation *one-pot*⁵⁸. Très récemment, il a également servi comme linker pour une stratégie d'assemblage en phase solide²⁰³. La conversion du thioester en peptide SeEA procède par une réaction d'échange séléno-thioester en utilisant un excès de *bis*(2-sélényléthyl)amine **7a**. Cet échange est directement suivi par une réaction de transfert d'acyle *Se,N* qui permet la production du peptide cible **22**. Comme indiqué précédemment, la réaction a lieu en présence de sélénium métallique pour éviter la formation d'impuretés déséléniées. Finalement, la formation du diséléno-cyclique est réalisée par oxydation avec du DMSO tandis que le groupement acétoacétyle est éliminé par ajout d'hydroxylamine après acidification du milieu⁵⁸.

Finalement, on obtient le peptide sélénioester latent SeEA **23** composé de 118 acides aminés avec un contrôle précis de la séquence atome par atome (Figure 70). En effet, si l'on compare le spectre de masse haute résolution déconvolué obtenu avec le profil théorique on remarque qu'ils sont parfaitement identiques (Figure 70C). L'obtention du peptide SeEA **23** nous a permis d'envisager la suite de la synthèse de la protéine NK1 en utilisant des conditions cinétiquement contrôlées.

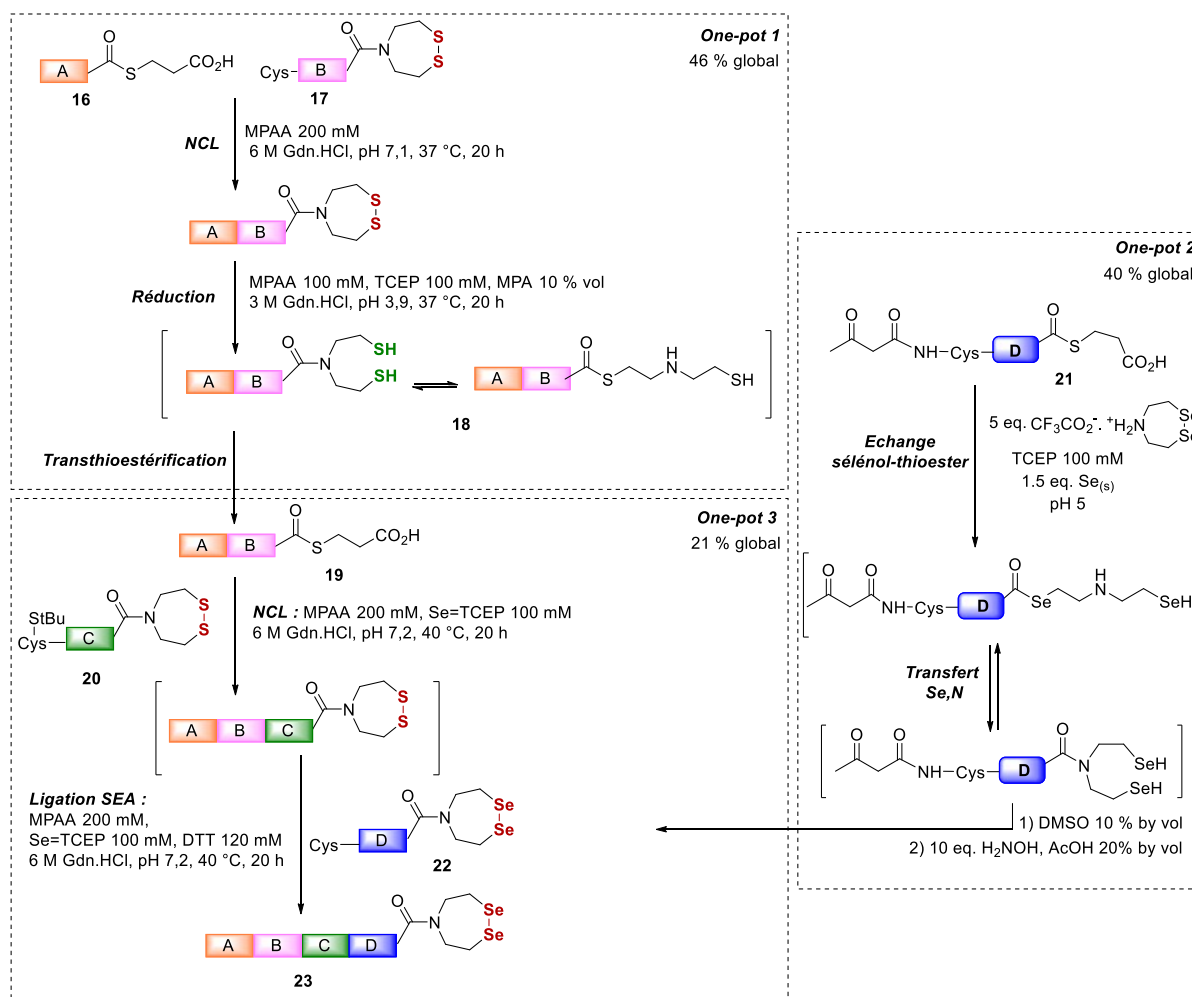


Figure 69 : Synthèse du sélénioester latent **23** (118 acides aminés) par procédés one-pot.

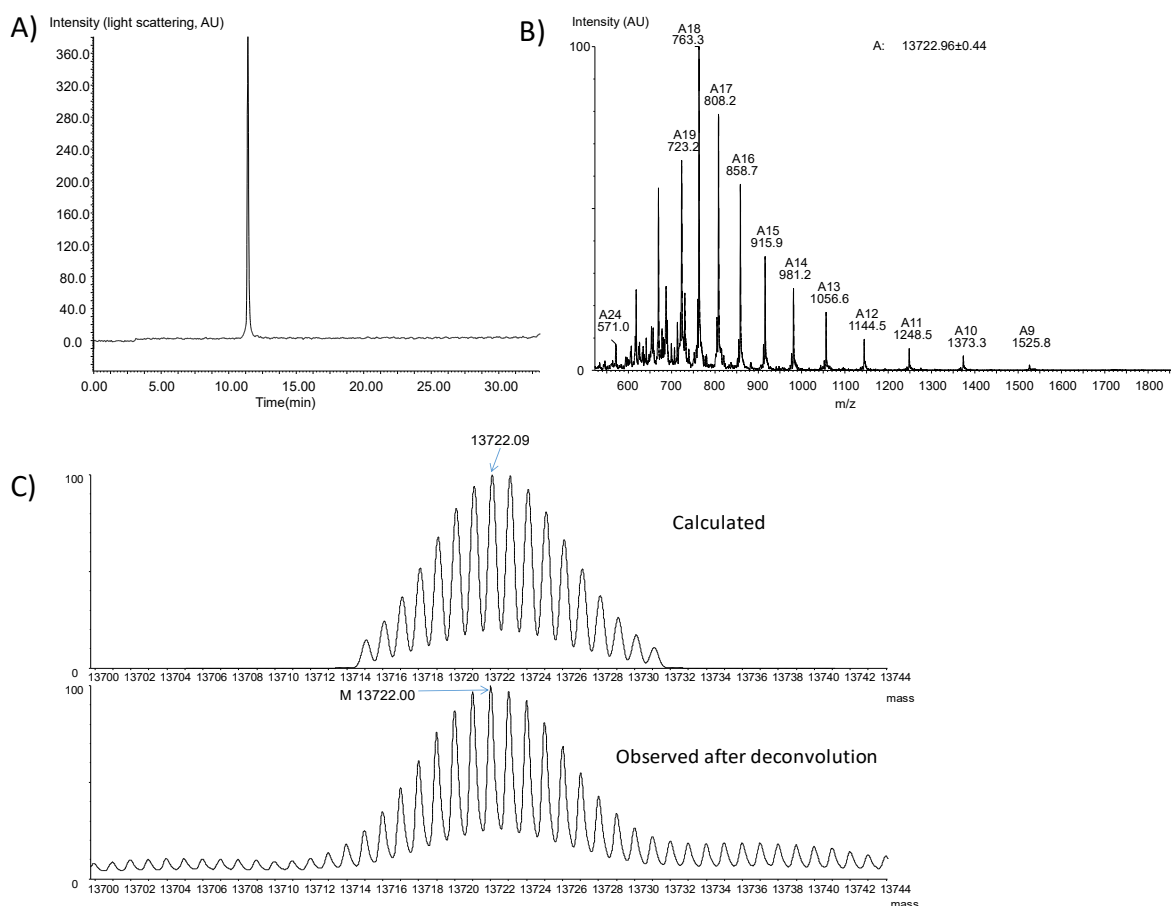


Figure 70 : Caractérisation du sélénoester **23** correspondant à l'HGF/SF (31-148). A) Analyse LC. B) Spectre haute résolution ESI MS. C) Spectre déconvolué et comparaison avec le profil théorique.

6.2.2 La ligation cinétiquement contrôlée SeEA/SEA.

La conception de procédés de ligation cinétiquement contrôlés nécessite l'utilisation de fonctions thioesters (ou substitués de thioesters) présentant une différence de réactivité significative.

Ici, nous proposons de réaliser une ligation SeEA en présence d'un fragment SEA grâce à la grande différence de réactivité entre les groupements SeEA et SEA au regard des cystéinyl peptides. Suivant le principe décrit par la Figure 64, il est possible d'assembler séquentiellement trois segments peptidiques en combinant les groupements SeEA et SEA dans cet ordre (KCL, Figure 64). Ici, cette méthode a été appliquée à la synthèse de la protéine NK1. Cet exemple implique l'assemblage des segments SeEA **23**, SEA **24** et **25**. La réduction du peptide SeEA **23** nécessite la présence de TCEP. Ici encore, la Se=TCEP a été ajoutée au milieu réactionnel afin de minimiser le phénomène de désélénsation. Une fois réduit le peptide **23** réagit très rapidement avec le cystéinyl peptide **24**. Lorsque la ligation SeEA est terminée, l'ajout du cystéinyl peptide **25** permet de déclencher la ligation SEA et de conduire à la formation du polypeptide cible (*one-pot* 4, Figure 71). Après purification, nous avons obtenu la protéine NK1 biotinylée avec un contrôle précis de la séquence (Figure 72), un

rendement de 31 % à partir du segment **23** (KCL) et un rendement global de 3 % à partir du segment **16**.

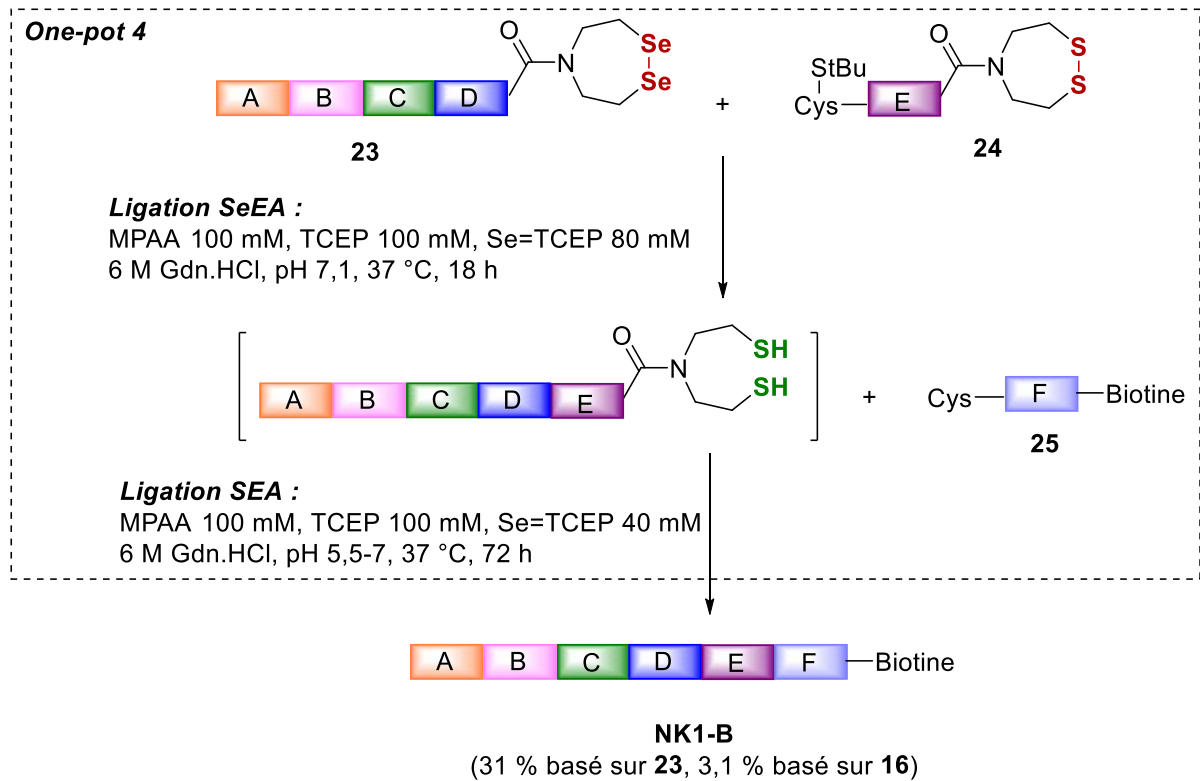


Figure 71 : Synthèse totale de la protéine **NK1-B** (180 acides aminés).

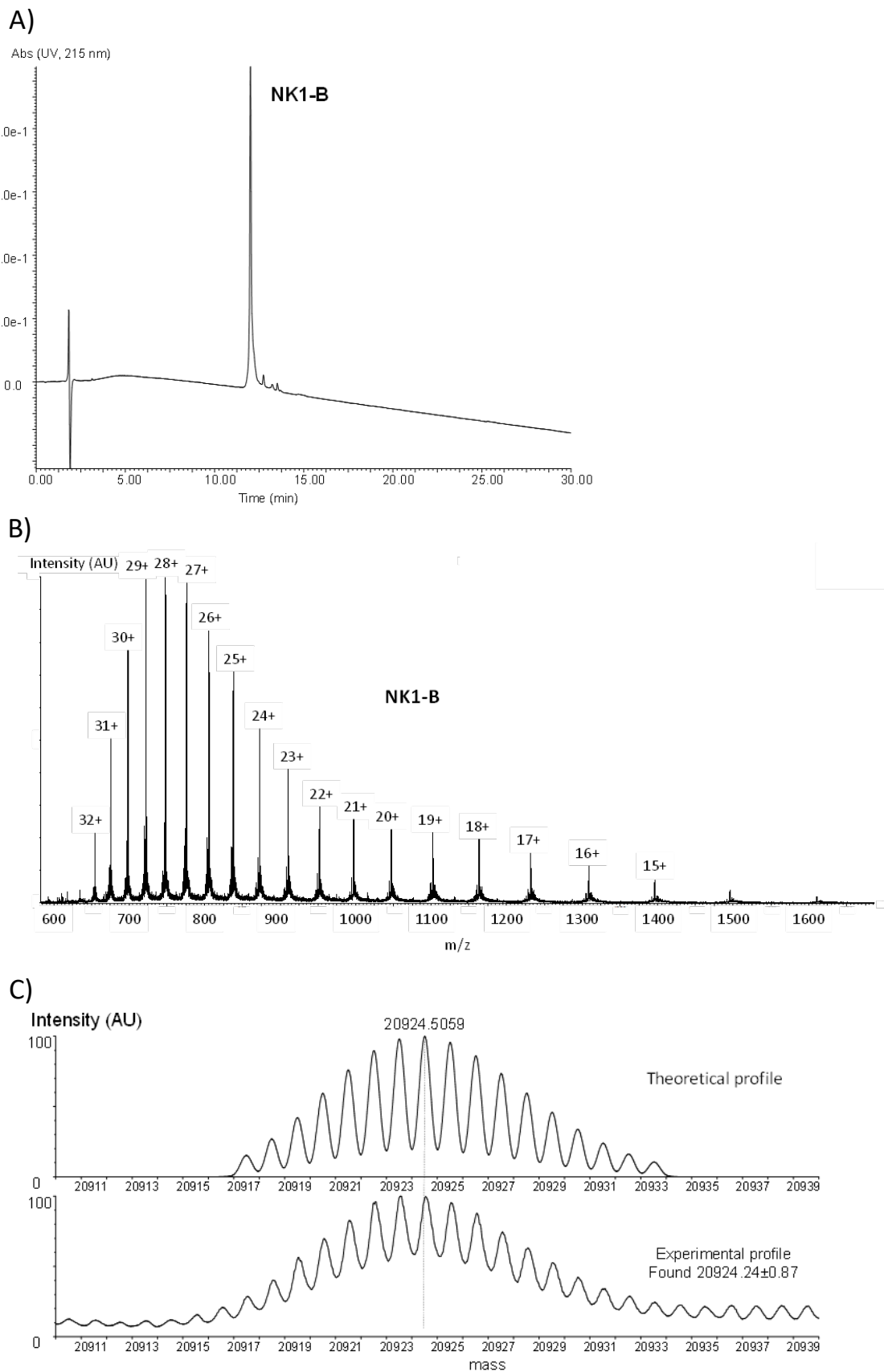


Figure 72 : Caractérisation de la protéine **NK1-B**. A) Analyse LC. B) Spectre haute résolution ESI MS. C) Spectre déconvolué et comparaison avec le profil théorique.

Pour résumer, la protéine **NK1-B** a été synthétisée en utilisant trois processus *one-pot* séquentiels, ce qui permet de diminuer le nombre d'intermédiaires et de purifications (3 procédés *one-pot* et 3 purifications en tout). L'assemblage de six segments peptidiques a été réalisé en combinant la KCL SeEA/SEA et les propriétés latentes des groupements S(e)EA révélant ainsi le grand potentiel du dérivé *bis*(2-sélényléthyl)amido pour faciliter l'accès à de grandes protéines.

En conclusion, nous avons décrit dans ce chapitre une nouvelle méthode pour accéder à des peptides sélénioesters de type SeEA. La synthèse de ceux-ci à partir des peptides SEA présente l'avantage d'être simple, rapide et compatible avec la Fmoc-SPPS. Les propriétés de ces nouveaux dérivés ont été étudiées et leur réactivité particulière a permis d'envisager de nouvelles stratégies pour faciliter et accélérer l'accès à de grandes protéines. La grande différence de réactivité entre les peptides SEA et SeEA a permis de concevoir un processus de ligation cinétiquement contrôlé tandis que la réduction sélective des groupements SEA et SeEA a été exploitée pour établir des stratégies séquentielles *one-pot*. L'utilisation de ces sélénioester latents a notamment permis la synthèse de la protéine **NK1-B** constituée de 180 acides aminés. Cette protéine est, à ce jour, la plus grande protéine synthétisée en utilisant un procédé KCL. Ce travail a fait l'objet d'une publication dans le journal *Chemical Science*, la version finale est annexée à ce manuscrit (Annexe B).

1. Problématique de la ligation aux jonctions difficiles

Dans le chapitre IV, nous avons vu que les propriétés uniques des peptides SeEA permettent de réaliser des procédés *one-pot* et donc de faciliter l'accès à des protéines de grande taille. Une autre limitation des procédés de ligation résulte de la difficulté à former des jonctions encombrées de type Val-Cys, Thr-Cys, Ile-Cys. L'encombrement stérique induit par les acides aminés β -branchés entraîne des temps de réactions prolongés qui peuvent entraîner des réactions secondaires. En partant des travaux de Durek et Alewood¹⁹⁸, nous avons cherché à explorer l'intérêt des peptides SeEA pour faciliter la formation de liaisons peptidiques encombrés.

Au lieu de préparer des peptides sélénioester ou sélénioester latents de façon préalable, ce qui peut demander un travail conséquent avec le risque de voir ces dérivés très réactifs se dégrader rapidement, nous avons souhaité les former *in situ* à l'aide de catalyseur. Plus précisément, nous avons mis au point une méthode pour la formation *in situ* de sélénioesters à partir de segments peptidiques SEA grâce à la conception de nouveaux catalyseurs adaptant une structure de type N-alkyl-perhydro-1,2,5-disélenazépine. Ces catalyseurs ont été utilisés pour accélérer deux réactions essentielles dans les procédés de synthèse totale développés dans cette thèse, à savoir la ligation SEA ainsi que la synthèse de peptides thioesters à partir de peptides SEA. Finalement, l'efficacité de l'un des catalyseurs a été illustrée par la synthèse totale de la granulysine 15 kDa (123 acides aminés).

2. Transthioestérification

2.1 Aspects mécanistiques

La réaction d'échange SEA/thiol est un moyen d'accès simple et très utile pour l'obtention de peptides thioester. La conversion d'un peptide SEA en thioester d'alkyle est réalisée de manière optimale à pH 4 et en présence d'un réducteur fort comme la TCEP qui permet d'activer la fonction SEA. Un réarrangement d'acyle de N vers S conduit à la formation d'un intermédiaire thioester transitoire qui réagit ensuite avec un alkylthiol tel que le MPA pour former un alkylthioester isolable.

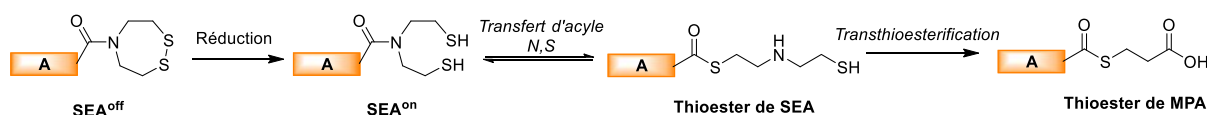


Figure 73 : Principe de conversion d'un peptide SEA en alkylthioester de MPA.

Nous savons qu'entre pH 7 et 5,5, le réarrangement intramoléculaire $N \rightarrow S$ est cinétiquement limitant pour la ligation SEA. Par exemple, le pré-réarrangement d'un peptide SEA amide en peptide SEA thioester permet d'accélérer considérablement la réaction de ligation. De plus, les travaux présentés dans le chapitre IV ont montré que la ligation SEA en présence de MPAA était significativement accélérée en abaissant le pH de 7 à 5,5⁷⁷. Cette dépendance au pH est la conséquence d'un transfert de proton dans l'état de transition de l'étape limitante, le transfert d'acyle $N \rightarrow S$. Logiquement, abaisser le pH en dessous de 5,5 devrait permettre d'accélérer davantage la ligation SEA car le réarrangement intramoléculaire $N \rightarrow S$ devrait être encore plus rapide. En pratique ce n'est pas le cas. La ligation SEA à pH 4,0 en présence de MPAA est plus lente qu'à pH 5,5 et suggère qu'un changement de l'étape limitante de cette réaction s'opère en baissant le pH. Il est donc probable qu'à pH 4,0, le pH utilisé pour l'échange du groupement SEA par un thiol, la situation soit similaire.

Comme nous l'avons vu dans la partie bibliographique (chapitre II), les échanges thiol/thioester s'effectuent préférentiellement via des espèces thiolate⁹¹. Or, à pH 4,0 la proportion d'espèces thiolates de MPA ($pK_a \sim 10,84$ ²²⁰) ou de MPAA ($pK_a \sim 6,6$) est très faible. Il est donc probable qu'à pH 4,0, les échanges thiol-thioester deviennent cinétiquement limitants. Nous nous sommes donc intéressés à l'utilisation d'additifs permettant de restaurer un régime d'échange thiol/thioester rapide à pH acide.

Les conditions classiques d'échange d'un groupement SEA avec le MPA ont été appliquées à un peptide SEA **9a** portant une alanine en C-terminal (Figure 74A). L'évolution de la réaction d'échange conduisant au peptide thioester **12a** est suivie par HPLC (Figure 74B). Les données cinétiques sont corrélées à une loi du premier ordre avec un temps de demi-réaction ($t_{1/2}$) de 7,28 h (entrée 1, Figure 74C). L'utilisation de MPAA en tant qu'additif permet d'accélérer la réaction. L'effet de celui-ci sur le $t_{1/2}$ est cependant modéré (accélération $\sim 1,5$ fois, entrée 2, Figure 74C). L'ajout de ce dernier permet de former un thioester transitoire plus réactif mais également d'augmenter la proportion d'espèces thiolates en solution puisque le pK_a du MPAA est nettement inférieur au pK_a du MPA. À un pH de 4,0 qui est nettement inférieur au pK_a du MPAA ($pK_a = 6,6$), la concentration en espèce thiolate reste cependant faible. De surcroît, le MPAA est peu soluble aux pH inférieurs à 5,5 et son utilisation dans ces conditions est difficile.

Nous nous sommes ensuite intéressés au sélénophénol qui a été utilisé avec succès par Durek et Alewood pour catalyser la réaction de NCL à partir de sélénioesters. Le sélénophénol est produit *in situ* par réduction du diphényldisélnure par la TCEP. Le sélénophénol possède un pK_a de 4,6²⁰⁹ de deux unités inférieur au pK_a du MPAA. Par conséquent, la concentration d'espèces nucléophiles à pH 4,0 doit être nettement plus élevée avec le sélénophénol que le MPAA. Malgré cette propriété intéressante, les données expérimentales montrent un faible gain cinétique (entrée 3, Figure 74C). Les

esters de sélénophényles sont des espèces acylantes plus puissantes que les thioesters dérivés du MPAA²²¹. Le faible effet catalytique du sélénophénol suggère un processus d'échange sélénophénol-alkylthioester défavorable. Ce phénomène pourrait être dû à la grande stabilité thermodynamique des alkylthioesters par rapport au sélénophénylesters dont la formation est défavorisée. Dans une étude antérieure, Durek et ses collaborateurs ont observé que le sélénophénol n'était pas capable d'activer les peptides alkylthioesters lors de la NCL ce qui est en accord avec nos résultats.

De façon intéressante, nous avons également noté un effet significatif du chlorhydrate de guanidine sur la cinétique de réaction (diminution d'un facteur $\sim 3,5$, entrées 1 et 5, Figure 74C). En pratique, la préparation d'un tampon à base de Gdn.HCl (6 M) induit une réduction de la concentration d'eau d'un facteur ~ 2 . En s'appuyant sur des expériences antérieures, nous émettons l'hypothèse que la diminution de la vitesse de réaction est due à la diminution de la quantité d'eau dans le système plus qu'au Gdn.HCl lui-même. En effet, lorsqu'un peptide SEA^{on} est solubilisé dans du DMF ou du TFA, il n'y a pas de réarrangement intramoléculaire $N \rightarrow S$ conduisant à la formation du thioester de SEA. A l'inverse, lorsqu'un peptide SEA^{on} est dissout dans de l'eau à pH acide, on observe un réarrangement avec un équilibre en faveur de la forme thioester à pH très acide. Nous supposons donc que l'eau joue un rôle important dans la réactivité du groupement SEA bien que les mécanismes mis en jeu ne soient pas encore élucidés. Le Gdn.HCl étant très souvent utilisé pour augmenter la solubilité des segments peptidiques, il est d'autant plus important de trouver de nouveaux moyens d'accélérer les vitesses des réactions dans un tampon aqueux contenant un agent chaotropique comme le Gdn.HCl.

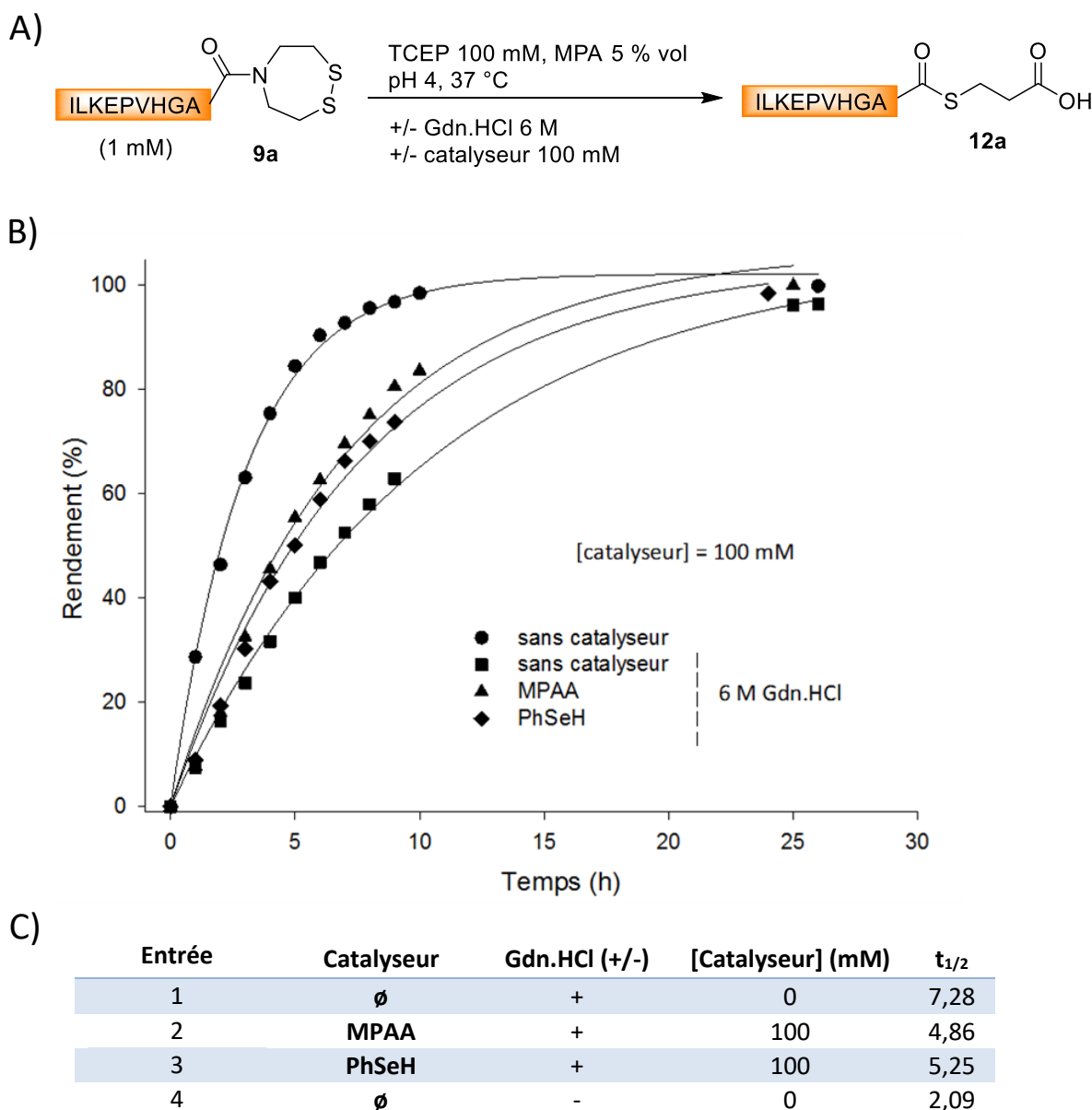


Figure 74 : A) Réaction d'échange du peptide **9a** avec le MPA. B) Suivi cinétique HPLC ($\lambda = 215$ nm) de la réaction d'échange. Les données sont corrélées par une loi d'ordre 1. C) Tableau des temps de demi-réaction pour les différentes conditions.

2.2 Émergence de nouveaux catalyseurs à base de sélénium

2.2.1 Résultats préliminaires

Durant notre recherche de nouveaux catalyseurs, nous nous sommes intéressés aux travaux de Makriyannis et ses collaborateurs. Ils ont montré qu'un mélange de benzoyle sélénocholine $\text{PhCO-SeCH}_2\text{CH}_2\text{N}^+(\text{Me})_3$ et de choline thiol $\text{HSCH}_2\text{CH}_2\text{N}^+(\text{Me})_3$ s'équilibre pour produire deux nouvelles espèces, le benzoyle thiocholine $\text{PhCO-SCH}_2\text{CH}_2\text{N}^+(\text{Me})_3$ et la sélénocholine $\text{HSeCH}_2\text{CH}_2\text{N}^+(\text{Me})_3$. Bien que l'équilibre soit significativement déplacé vers le thioester à pH 7 ($K_{\text{eq}} = 140$), la situation est plus

équilibrée à pH 5 où K_{eq} est égal à 3. Cette diminution nette de K_{eq} lorsque le pH est abaissé a été attribuée au pK_a de la sélénocholine ($pK_a = 4,7$) qui est de 3 unités inférieur aux pK_a de la thiocholine ($pK_a = 7,7$). Par conséquent, la concentration d'espèces sélénates de choline à pH 5 dépasse largement celle des espèces thiolates de choline et favorise la formation du sélénioester.

En s'inspirant de cette étude, nous avons testé le potentiel catalytique de la *bis*(2-sélényléthyl)amine **7a** pour la conversion d'un peptide SEA **9a** en peptide alkylthioester **12a**. Nous avons vu dans le chapitre précédent que l'utilisation de *bis*(2-sélényléthyl)amine **7a** permet la synthèse de peptides SeEA. Les peptides SeEA sont beaucoup plus réactifs que leurs analogues soufrés à pH 7 en présence de MPAA. Dans ces conditions, les échanges thiol/thioester ne sont pas limitants, les dérivés séléniums sont plus réactifs dû à l'accélération du transfert azote/chalcogène.

À l'inverse, il s'agit, ici d'accélérer la réaction d'échange thiol/thioester à pH acide. En théorie, il est possible de produire les peptides SeEA *in situ* et donc de générer des espèces sélénates très réactives en utilisant la *bis*(2-sélényléthyl)amine **7a**. Nous avons calculé les pK_a de la *bis*(2-sélényléthyl)amine à l'aide du logiciel ACD/Lab (Figure 75A). Les valeurs obtenues sont cohérentes avec les pK_a mesurés pour des alkylsélénols ou celui de la sélénocystéine (pK_a 5,5) qui est très proche structurellement. Ces valeurs nous indiquent qu'à pH 4, la moitié des fonctions sélénols du composé **7a** se trouvent sous forme de sélénate.

En accord avec ces hypothèses, la *bis*(2-sélényléthyl)amine **7a** induit une accélération significative de la réaction de transthioestérification (entrée 4, Figure 76C). Cependant, la conversion n'est pas complète et atteint un plateau à 80 % de conversion. L'analyse du milieu réactionnel révèle la formation d'un produit secondaire qui a été identifié par spectrométrie de masse comme étant le peptide SeEA **10** (Figure 75B).

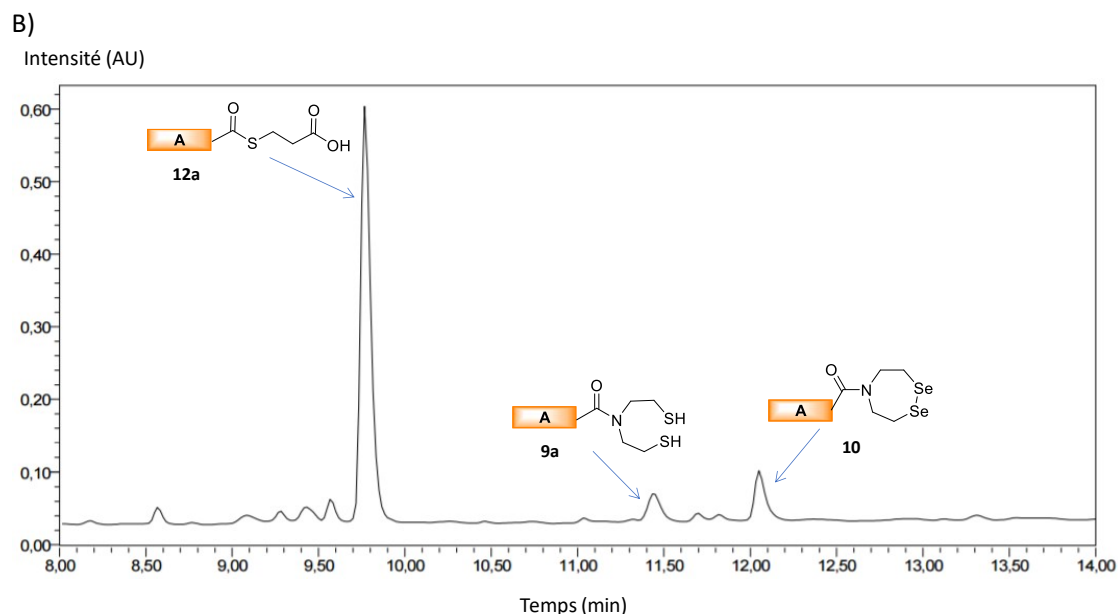
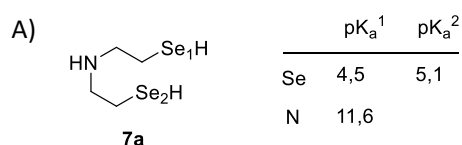
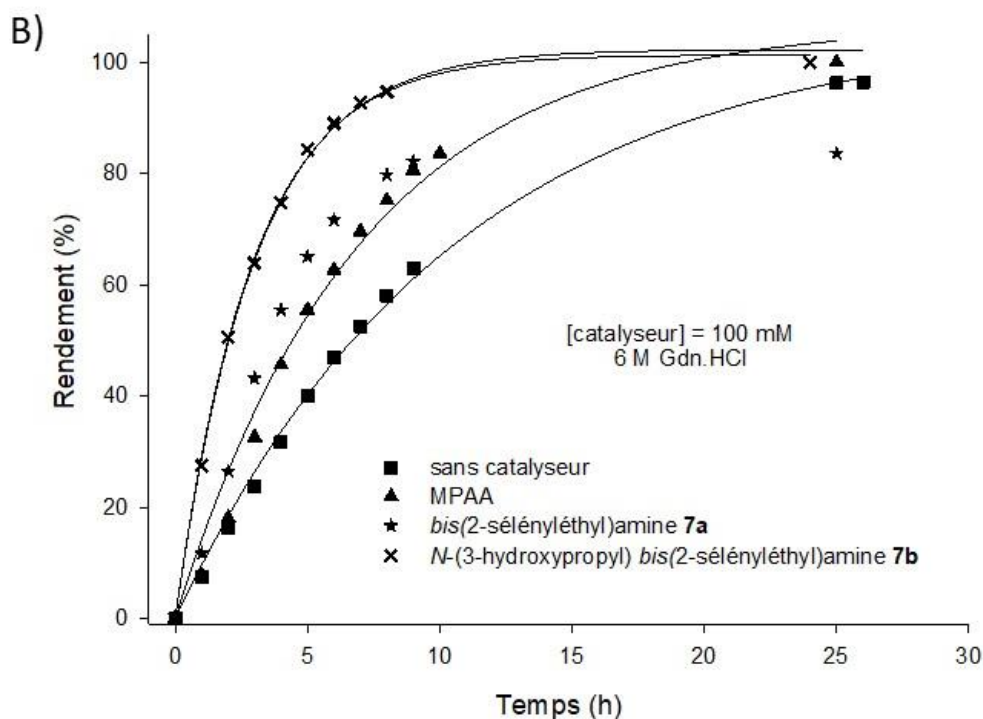
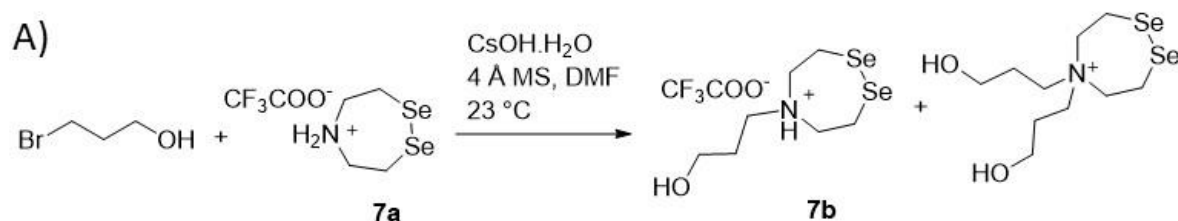


Figure 75 : A) Estimation des pK_a de l'amine **7a** à l'aide du logiciel ACD/Labs. B) Chromatogramme HPLC ($\lambda=215$ nm) à $t=24$ h pour la conversion du peptide SEA **9a** en peptide thioester MPA **12a** en présence de **7a**.

Encouragés par les résultats obtenus avec la bis(2-sélénylethyl)amine **7a**, nous avons cherché à prévenir l'accumulation du produit secondaire tout en conservant les mêmes propriétés catalytiques. L'étude de l'influence du pH sur la formation du produit **10** n'a pas donné de résultats satisfaisants. Une alternative consiste à substituer le proton de l'amine par un autre groupement ce qui permettrait d'empêcher la formation d'amide SeEA tout en conservant la réactivité des fonctions sélénol. De plus, la présence d'une amine tertiaire garantit une bonne solubilité à pH acide.

Dans un premier temps, nous avons utilisé le 3-bromopropanol en présence d'hydroxyde de césium pour alkyler **7a**. En effet, Salvatore et ses collaborateurs ont montré que l'hydroxyde de césium permet de favoriser la monoalkylation et de minimiser la polyalkylation. Dans notre cas, la réaction a conduit à la formation des produits mono et di-alkylés. De plus, des problèmes lors de la purification ont conduit à de très faibles rendements de synthèse (4 %, Figure 76A). Néanmoins, nous avons testé ce dérivé et observé une conversion totale du peptide SEA en thioester (Figure 76B). De plus, ce nouveau catalyseur **7b** présente un potentiel catalytique supérieur au composé **7a**. On observe une accélération d'un facteur 3,6 en comparaison avec la réaction en absence de catalyseur (entrée 5, Figure 76C). Il est important de noter que les deux groupements sélénol peuvent participer aux propriétés catalytiques du dérivé **7b** ce qui peut induire un biais dans la comparaison avec le MPAA.

Pour comparer ces catalyseurs de façon objective, nous avons réalisé une étude du potentiel catalytique du diséléniure **7b** (50 et 100 mM) et du MPAA (100 et 200 mM). Dans tous les cas, l'échange thiol/thioester est significativement plus rapide avec le diséléniure **7b** qu'avec le MPAA. La supériorité du composé **7b** n'est pas due à une concentration double en espèce séléniol.



C)

Entrée	Catalyseur	[catalyseur] (mM)	t _{1/2} (h)
1	∅	0	7,28
2	MPAA	100	4,86
3	PhSeH	100	5,25
4	7a	100	3*
5	7b	100	2,04

* Dans ce cas, les données ne corrélaient pas avec une loi d'ordre 1. La valeur indiquée a été déterminée graphiquement et correspond au temps nécessaire pour atteindre 50 % de conversion.

Figure 76 : A) Schéma de synthèse du composé N-propanol bis(2-sélényléthyl)amine **7b**. B) Suivi cinétique HPLC ($\lambda = 215$ nm) de la réaction d'échange. Les données sont corrélées par une loi d'ordre 1, tableau des temps de demi-réaction pour les différentes conditions.

Au regard de ces résultats très encourageants, nous avons cherché à optimiser la synthèse de composés N-alkylés dérivés du diséléniure **7a**.

2.2.2 Synthèse des catalyseurs *N*-alkyl disélénures

La synthèse de composés à base de sélénium est difficile et complexe en raison de leur grande réactivité et sensibilité à différents réactifs oxydants ou réducteurs. Dans un premier temps, nous avons essayé d'alkyler l'amine par réaction d'amination réductrice. Pour cela, nous avons utilisé le triacétoxyborohydrure de sodium $\text{NaBH}(\text{OAc})_3$ et le butyraldéhyde. Le $\text{NaBH}(\text{OAc})_3$ est un agent réducteur doux et très efficace pour l'amination réductrice avec des aldéhydes ou des cétones. Malgré des conditions opératoires particulières (atmosphère inerte et absence de rayonnement lumineux), on observe l'extrusion d'un atome de sélénium conduisant à la formation majoritaire du composé mono-sélénié.

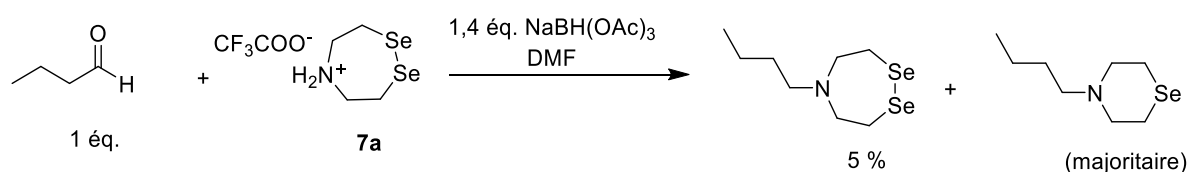


Figure 77 : Schéma de la réaction d'amination réductrice par le $\text{NaBH}(\text{OAc})_3$.

Devant ces résultats négatifs, nous nous sommes intéressés à la réaction d'addition de Michael sur l'azote autrement dit une réaction d'*aza*-Michael. Pour cela, nous avons utilisé l'acrylamide qui sert d'accepteur de Michael et une base. Initialement, nous avons étudié l'utilisation de la triéthylamine ou du 1,8-diazabicyclo[5.4.0]undéc-7-ène (DBU) pour neutraliser le sel TFA de la *bis*(2-sélényléthyl)amine **7a**. Il est apparu qu'en présence de Et_3N ou de DBU, le composé **7a** tend à précipiter ce qui limite la conversion. De plus, l'utilisation du DBU induit des problèmes de séparation lors des étapes de purification.

En s'appuyant sur les travaux de You et de ses collaborateurs²²², nous avons testé l'effet de l'ajout de gel de silice au milieu réactionnel. Dans leur étude, les auteurs ont montré que l'utilisation de gel de silice permet d'accélérer la réaction d'*aza*-Michael. Dans notre cas, nous émettons l'hypothèse que la silice permet d'adsorber le composé **7a** ce qui l'empêche de précipiter en présence de Et_3N . Finalement, les conditions opératoires utilisées sont décrites dans la Figure 78. Grâce à ces conditions et à l'utilisation de différents accepteurs de Michael, nous avons pu isoler deux nouveaux dérivés **7c** et **7d** avec des rendements satisfaisants.

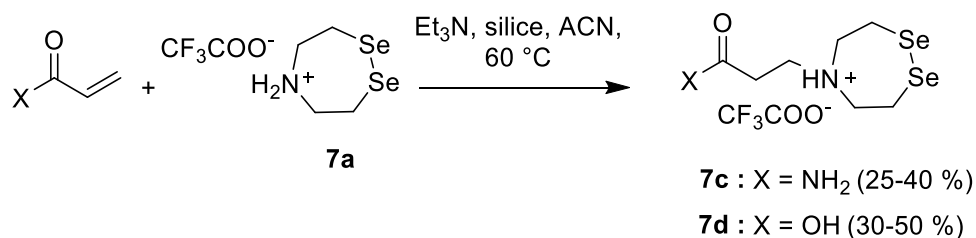


Figure 78 : Schéma de synthèse des composés disélénures N-alkylés **7c** et **7d**.

Le composé **7d** présente l'avantage de permettre une modification supplémentaire grâce à sa fonction acide carboxylique libre. Il est envisageable de greffer ce type de composé sur une résine et de réaliser une catalyse en phase hétérogène ce qui permettrait de recycler le catalyseur. Devant le potentiel de ces nouveaux composés, nous avons développé une autre approche synthétique afin de permettre une montée en échelle. Cette nouvelle stratégie a été initiée par V. Diemer dans le cadre d'un autre projet mené au laboratoire. Je l'ai étendue par la suite à la synthèse du disélénuire **7d** (Figure 79).

La synthèse du benzylséléno aldéhyde est décrite dans la littérature²²³ (Figure 79A). Cet aldéhyde est assez stable pour être purifié par chromatographie sur colonne de silice, ce qui permet de travailler sur une échelle de synthèse plus grande comparativement à l'approche basée sur la réaction d'*aza*-Michael (échelle de synthèse de 4 mmoles). Une réaction amination réductrice avec la β -alanine *tert*-butyl ester (commerciallement disponible) suivie d'une étape de déprotection/cyclisation permet de produire le disélénuire **7d** (Figure 79B). Cette dernière réaction est réalisée en utilisant la 2,2'-dithio-*bis*(5-nitropyridine) (DTNP) qui permet à la fois de déprotéger le groupement méthoxybenzyle (Mob) et de former le pont disélénuire²²⁴. Seul le composé **7d** final est purifié par chromatographie en phase inverse et les rendements par étape sont relativement satisfaisants. À noter que le rendement global est de 26 % ce qui est nettement supérieur au rendement global de la stratégie impliquant la synthèse et l'amination réductrice du composé **7a** (10 %). Cette nouvelle méthode de synthèse représente une voie d'accès simple et rapide aux nouveaux catalyseurs.

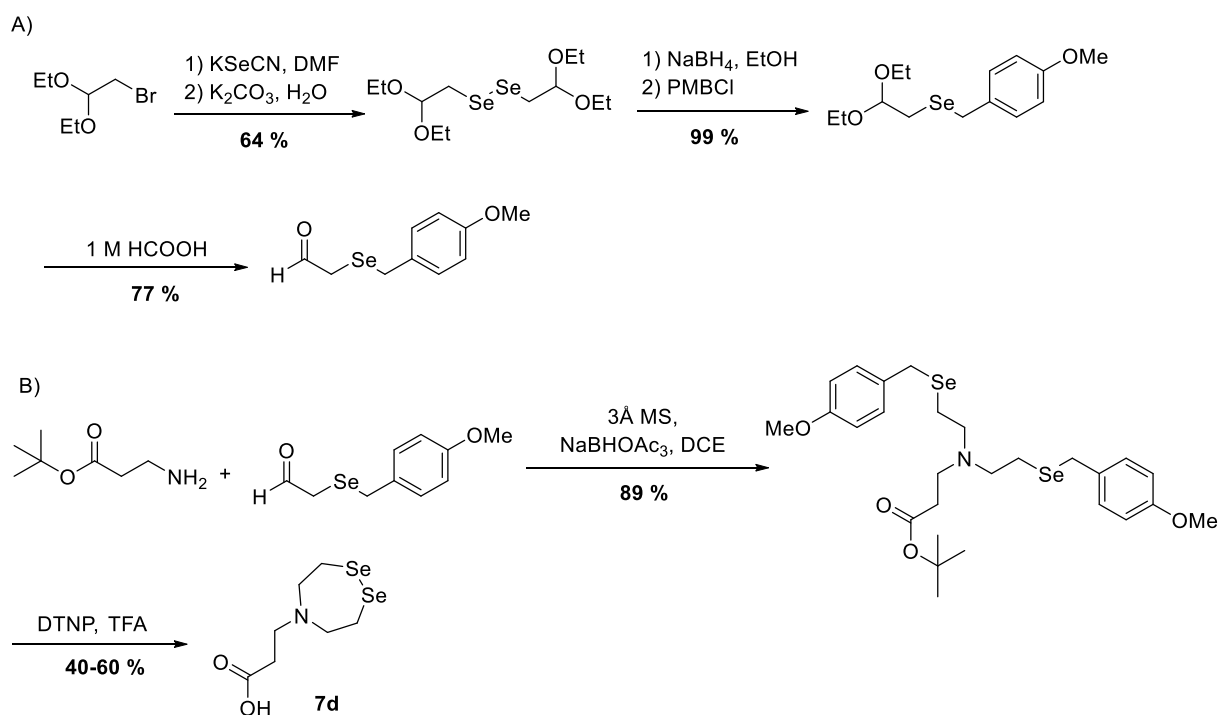


Figure 79 : Nouvelle stratégie de synthèse de **7d**. A) Synthèse du benzylséléno-aldéhyde. B) Réaction d'amination réductrice puis déprotection/cyclisation conduisant au composé **7d**.

Les nouveaux composés **7b-d** ont été testés pour catalyser la conversion de peptides SEA **9b-d** en peptides thioesters **12b-d** (Figure 80). Ils présentent des acides aminés encombrés en C-terminal pour tester l'intérêt des nouveaux catalyseurs dans des cas difficiles.

2.2.3 Catalyse de la formation de jonctions difficiles

Les études préliminaires présentées en 2.2.1 ont été réalisées sur un peptide comportant une alanine en C-terminal. Or, l'une des limitations majeures des méthodes de ligation est la difficulté à former des liaisons peptidiques encombrées de type Val-Cys, Thr-Cys et Ile-Cys. Pour démontrer le potentiel catalytique de nos nouveaux composés à base de sélénium, nous nous sommes donc intéressés à des réactions d'échange thiol/thioester impliquant ces acides aminés encombrés (Figure 80).

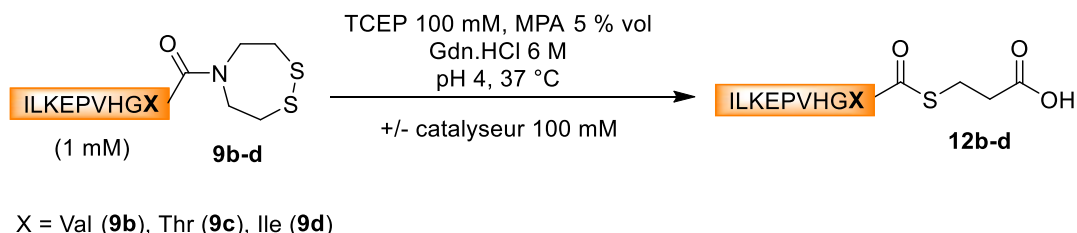
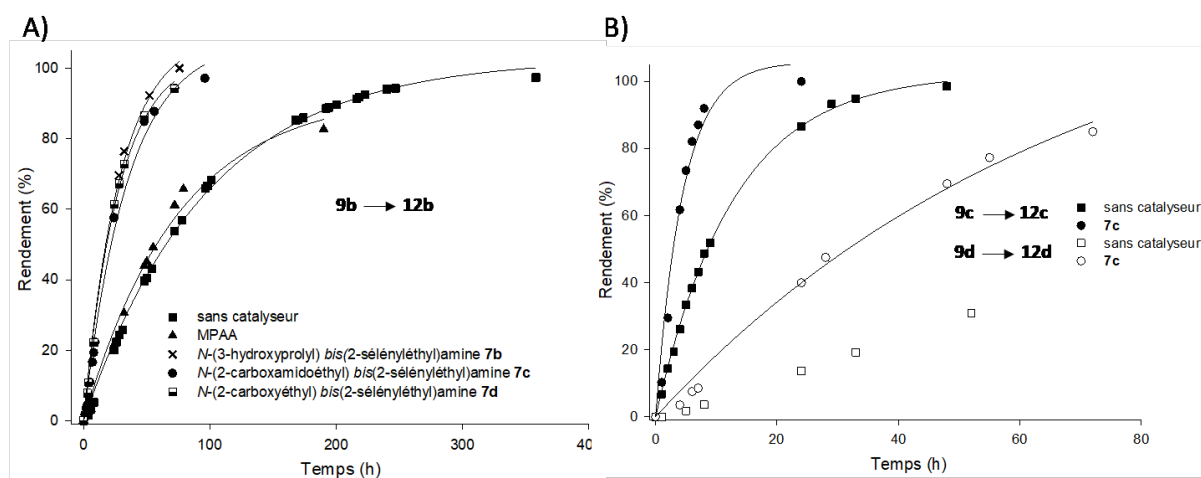


Figure 80 : Réaction de transthioestérification de peptides comportant des acides aminés β -branchés en C-terminal.

Dans un premier temps, nous avons étudié la réaction de transthioestérification du peptide **9b** portant une valine en C-terminal en présence du MPAA ou des composés diséléneures. On observe que le MPAA n'a quasiment aucun effet sur la cinétique de réaction tandis que les composés **7b-d** entraînent une accélération significative de la vitesse de réaction (~3,7 fois). On note également que la nature de la chaîne alkyle n'a pas d'influence sur l'effet catalytique des composés séléniés (Figure 81A). La *N*-(2-carboxamidoéthyl) bis(2-sélényléthylamine) **7c** a également été testée pour la catalyse de la réaction d'échange des peptides **9c** et **9d** comportant respectivement une thréonine et une isoleucine en C-terminal (Figure 81B). Le même effet d'accélération a été observé.



Entrée	X	Catalyseur	[Catalyseur] (mM)	t _{1/2} (h)
1	9b	∅	0	67,96
2	9b	MPAA	100	51,73
3	9b	7b	100	20,03
4	9b	7c	100	24,32
5	9b	7d	100	19,36
6	9c	∅	0	8,86
7	9c	7c	100	3,12
8	9d	7c	100	47,5

Figure 81 : Suivi cinétique HPLC ($\lambda = 215$ nm) de la réaction d'échange des peptides **9b** (A), **9c-d** (B). Les données sont corrélées par une loi d'ordre 1, tableau des temps de demi-réaction pour les différentes conditions.

L'utilité du catalyseur **7d** pour la synthèse préparative de peptides thioester a été confirmée par la synthèse du peptide thioester **12b** avec un rendement de 63 %. Il est important de noter que l'analyse par GC-MS chirale du peptide après hydrolyse acide montre que le taux de racémisation des résidus valine est inférieur à la limite détectable (<0,1 %). Dans un second exemple, un peptide SEA de

46 acides aminés comportant une thréonine en C terminal a été converti en thioester en utilisant le même protocole (36 %, Figure 82).

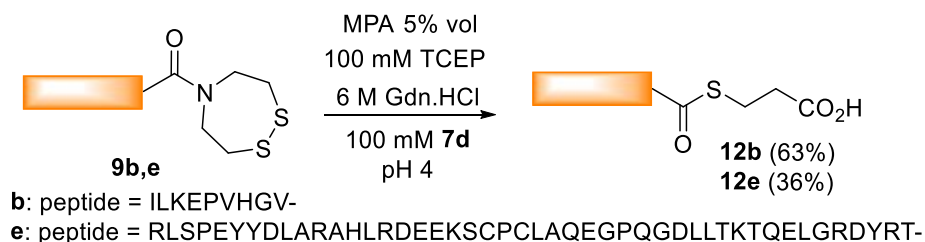


Figure 82 : Synthèse préparative de peptides thioesters en utilisant **7d**.

3. Ligation SEA

Nous avons montré l'utilité des composés **7c,d** pour la catalyse des échanges SEA/thiol à pH 4, nous proposons d'étudier dans cette section leur potentiel pour accélérer la ligation SEA aux jonctions difficiles. Pour cela, nous avons comparé l'effet catalytique du disélényure **7c** et du MPAA pour la réaction de ligation SEA entre le peptide SEA **9b** et le cystéinyl peptide **13** conduisant à la formation d'une jonction Val-Cys. La Figure 83 présente les résultats cinétiques obtenus.

Comme nous l'avons précédemment mentionné, la ligation SEA en présence de MPAA est significativement accélérée en abaissant le pH de 7,0 à 5,5. Bien que le transfert d'acyle soit probablement encore plus rapide en dessous de pH 5,5, on remarque que la ligation à pH 4,0 en présence de MPAA est plus lente qu'à pH 5,5. Cette observation suppose un changement d'étape cinétiquement limitante en dessous de pH 5,5 (Figure 83). De façon intéressante, Asahina et ses collaborateurs ont décrit un effet du pH similaire pour la ligation impliquant la *N*-méthylcystéine en présence de MPAA. Ils ont observé une augmentation de la cinétique en abaissant le pH de 8,0 à 5,5 puis une diminution en dessous de pH 5,5. A l'inverse, la cinétique de ligation SEA est considérablement accélérée en présence du disélényure **7c** en abaissant le pH à 4,0 (Figure 83). Ce résultat démontre l'utilité d'introduire de nouveaux catalyseurs qui permettent d'accélérer les échanges thiol-thioester et ainsi d'étendre la gamme de pH de travail.

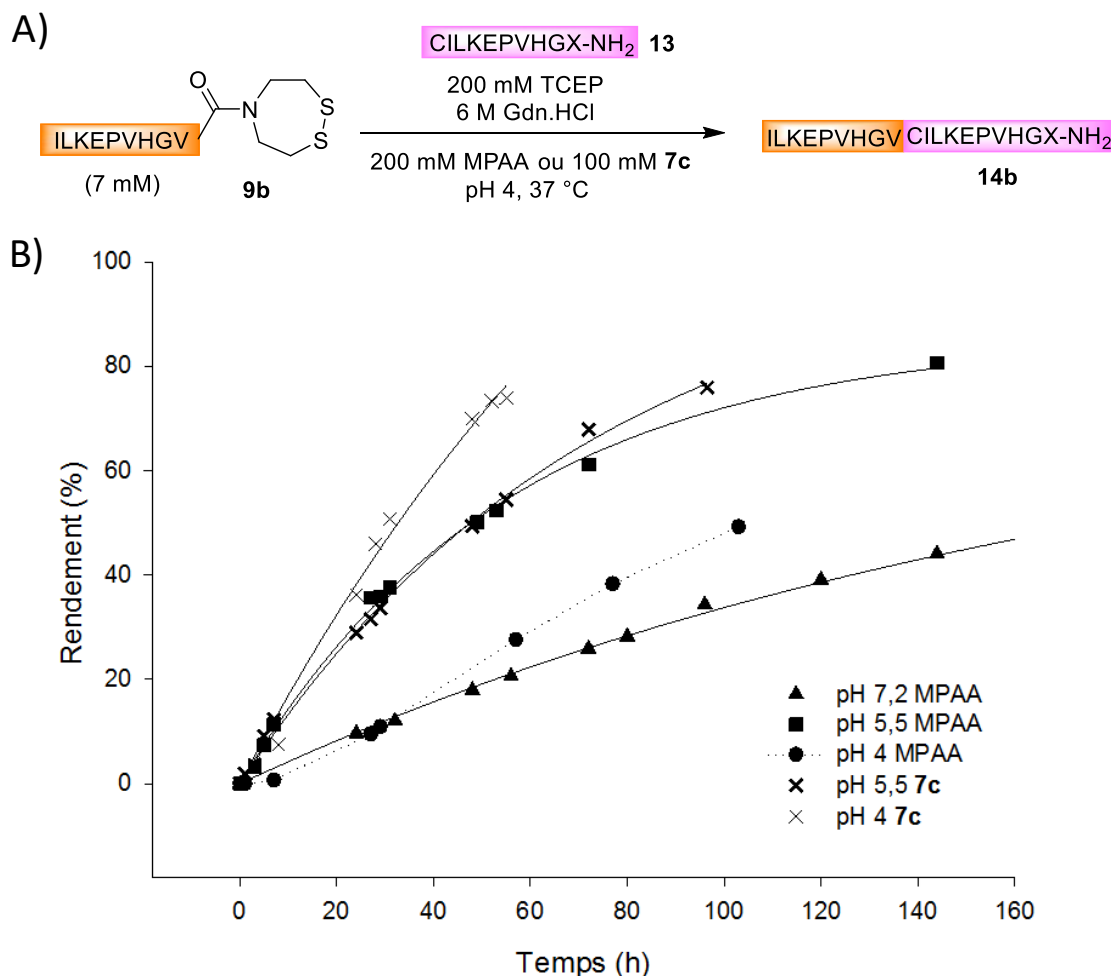


Figure 83 : Catalyse de la ligation SEA par le MPAA ou le disélényure **7c**. Les courbes continues correspondent aux courbes de tendance d'ordre 1. La courbe en pointillé indique une réaction qui ne peut pas être corrélée par une loi d'ordre 1.

De façon intéressante, les deux catalyseurs étudiés suivent des cinétiques comparables à pH 5,5. Cette observation est en accord avec le fait que le transfert d'acyle $N \rightarrow S$ soit l'étape limitante du processus à pH 5,5. Dans ces conditions, le choix du catalyseur n'impacte pas la vitesse globale de la réaction tant que les échanges thiol/thioester sont maintenus dans un régime rapide. Des expériences complémentaires viennent renforcer cette hypothèse (Figure 84A). En effet, l'utilisation de 100 mM MPAA au lieu de 200 mM n'a pas d'effet sur la ligation à pH 5,5. Cependant, si l'on diminue considérablement la concentration de catalyseur (10 mM) on observe une diminution significative de la vitesse de réaction due aux échanges thiol/thioester qui deviennent limitants dans ces conditions (Figure 84B).

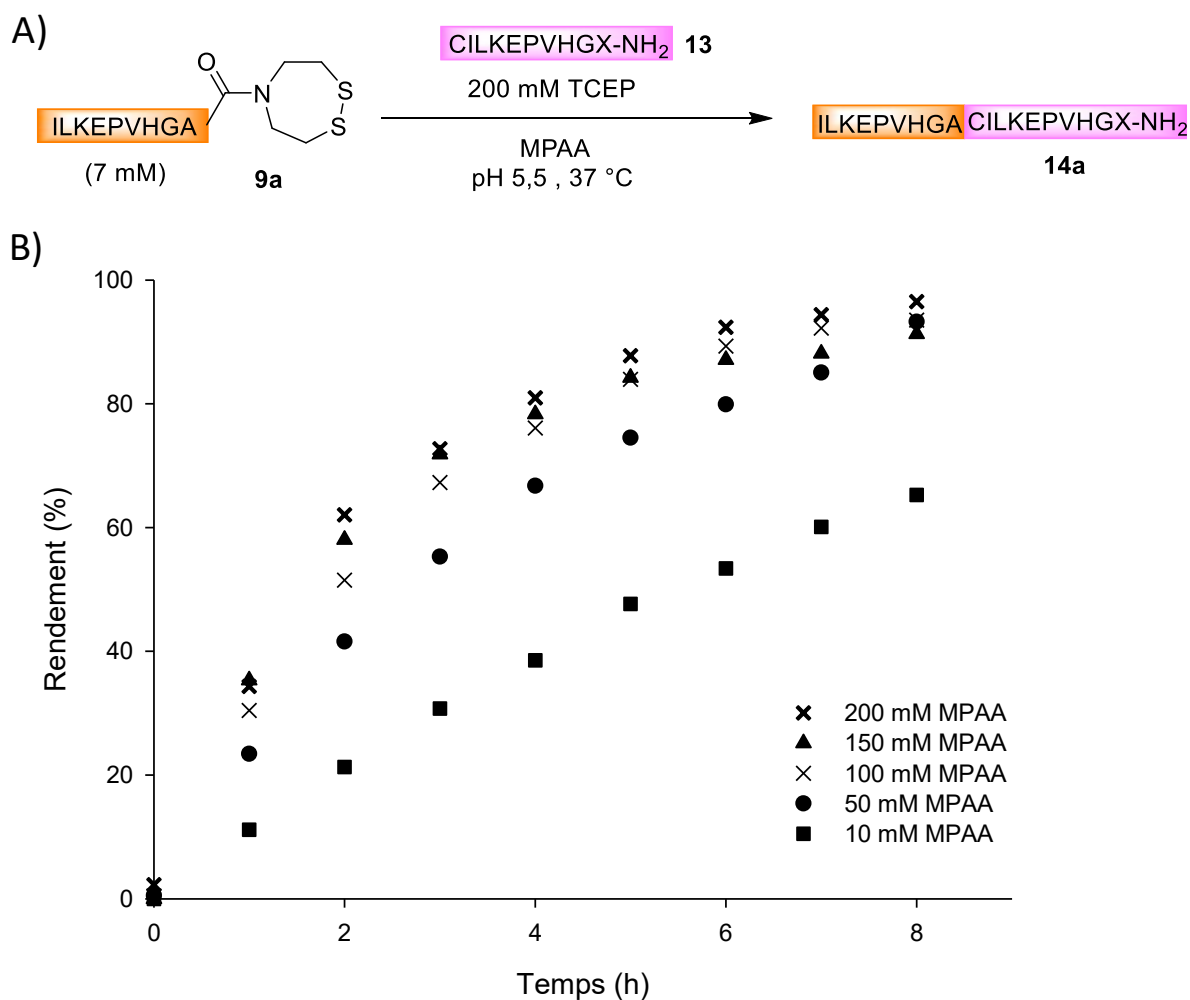


Figure 84 : Effet de la concentration de MPAA sur la cinétique de ligation SEA.

Par la suite, le catalyseur **7d** a été utilisé pour réaliser la ligation SEA entre les peptides **9b** et **13**. Le produit de ligation **14b** a été obtenu avec un rendement de 51 %. L'analyse chirale par GC-MS de celui-ci indique la présence de seulement 0,19 % de résidus D-Val. La racémisation de l'acide aminé en C-terminal du peptide **9b** est donc négligeable. Le catalyseur **7d** a également permis de réaliser la cyclisation par ligation du peptide **26**. Le peptide SEA **26** est composé de 38 acides aminés et comporte une valine en C-terminal. Dans cet exemple, nous avons utilisé un détergent non-ionique, le *N*-octylglucoside, pour éviter la gélification du milieu réactionnel de ligation qui a été observée en utilisant 6 M de Gdn.HCl. Le produit cyclique a été obtenu avec un rendement de 26 %. Des analyses par spectrométrie de masse MALDI-TOF après digestion par la trypsine ont permis de confirmer la structure cyclique du produit isolé (voir Annexe A, Chapter V, Section 4.2.2).

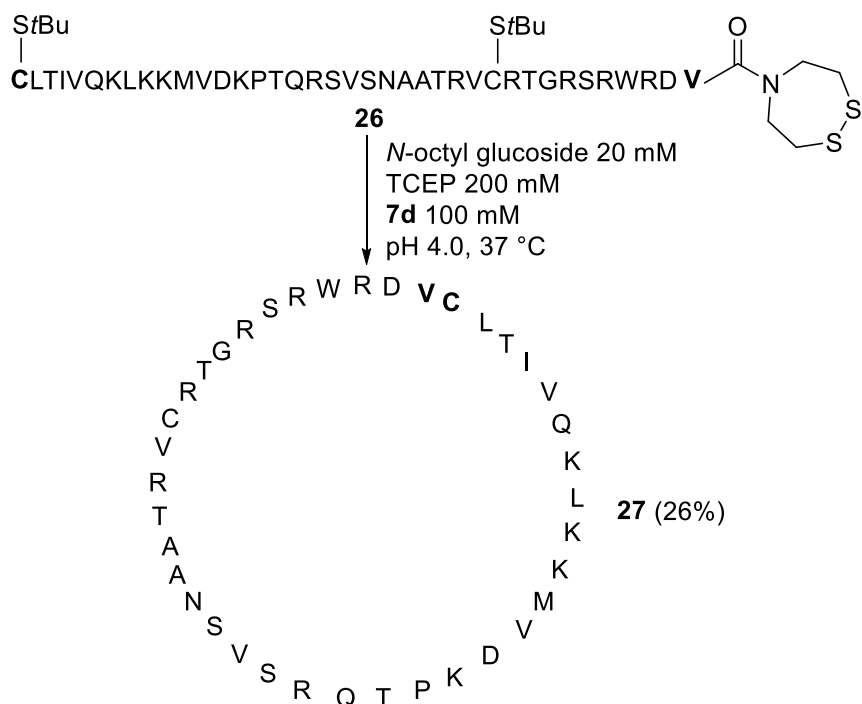


Figure 85 : Cyclisation par ligation SEA du segment peptidique 26 en présence du catalyseur 7d.

Ces expériences sur des peptides modèles nous ont permis de montrer que les disélénures **7c,d** catalysent la réaction d'échange SEA-thiol et la ligation SEA à pH 4,0. Les cinétiques observées sont significativement accélérées en comparaison avec celles obtenues avec le MPAA comme catalyseur. Ces résultats montrent l'intérêt d'utiliser les catalyseurs disélénures pour faciliter les réactions aux jonctions difficiles. Fort des résultats obtenus, nous avons entrepris la synthèse totale de la granulysine pour démontrer le potentiel de nos catalyseurs pour accéder à des protéines complexes.

4. La synthèse totale de la granulysine (23-145)^e

4.1 Intérêt biologique

La granulysine est une protéine cytolytique et pro-inflammatoire. Elle est présente dans les granules cytolytiques des lymphocytes T cytotoxiques (réponse immunitaire adaptative) et des cellules tueuses naturelles, *NK cells* (réponse immunitaire non spécifique, innée)²²⁵. La granulysine est produite sous deux isoformes stables de 9 kDa et 15 kDa. La forme à 9kDa résulte du clivage en N et C terminal de la forme 15 kDa²²⁵.

^e Numérotation des acides aminés en fonction de la nomenclature d'UniProtKB : P22749 (GNLY_HUMAN)

La granulysine recombinante de 9 kDa a été intensivement étudiée. Elle possède une activité cytolytique sur une grande variété de microbes comme les bactéries gram-positif (*S. aureus*) et gram-négatif (*E. coli*), les champignons et les parasites²²⁶. De façon intéressante, il a été démontré que la granulysine peut directement tuer la bactérie *M. tuberculosis* extracellulaire par dégradation de la membrane du bacille. En combinaison avec la perforine, elle diminue également la viabilité de *M. tuberculosis* intracellulaire²²⁷. La granulysine tue les bactéries en altérant la perméabilité de leurs membranes grâce à des interactions directes de type électrostatiques et liaisons hydrogène²²⁸. Au regard des nombreuses études menées sur la granulysine, on remarque que celle-ci présente une large pertinence clinique pour diverses maladies incluant les cancers, le syndrome de Stevens-Johnson et la transplantation (maladie du greffon contre l'hôte)^{226,229,230}. La granulysine possède donc un grand potentiel thérapeutique. Elle pourrait contribuer à comprendre les mécanismes biologiques impliqués dans de nombreuses maladies et contribuer à l'élaboration de traitements pour celles-ci.

Toutefois, aucune des deux formes n'a été produite par synthèse totale et leurs productions par des méthodes biologiques classiques est complexe. Seule la granulysine de 15 kDa est commercialement disponible. Elle est exprimée dans *E. coli* et contient des *tags*. Des études comparatives ont montré que la présence de ces *tag* altère les fonctions biologiques de la granulysine^{231,232}. La synthèse de la granulysine 15 kDa native permettrait d'étudier le véritable potentiel thérapeutique de celle-ci via des études fonctionnelles et structurales.

En effet, seule la structure tridimensionnelle de la forme 9 kDa est connue à ce jour. Une étude cristallographique a montré pour cette isoforme la présence de cinq hélices α et de deux ponts disulfures (Cys69-Cys132 et Cys96-Cys107). La granulysine appartient à la famille *saposin-like* (SAPLIP).²²⁶ Les membres de cette famille partagent plusieurs propriétés structurales. Ils possèdent six cystéines localisées de façon similaire formant trois ponts disulfures avec le même appariement. Plusieurs études ont montré une homologie entre la granulysine et la NK-lysine porcine. Elles possèdent 35 % de résidus identiques (55 % similitude, Figure 86) et présentent tous deux une activité cytolytique sur une grande variété d'organismes.



Figure 86 : Alignement des séquences de la granulysine humaine (9 kDa) et de la NK-lysine porcine à l'aide du logiciel DNASTAR Megalign Pro v12.3 matrice Clustal Omega. Le code couleur est basé sur les propriétés des acides aminés.

Bien que la granulysine 9 kDa ne possède que quatre cystéines et ne forme donc que deux ponts disulfures, sa structure tridimensionnelle est très proche de celle de la NK-lysine qui, elle, est stabilisée par trois ponts disulfures (Figure 87).

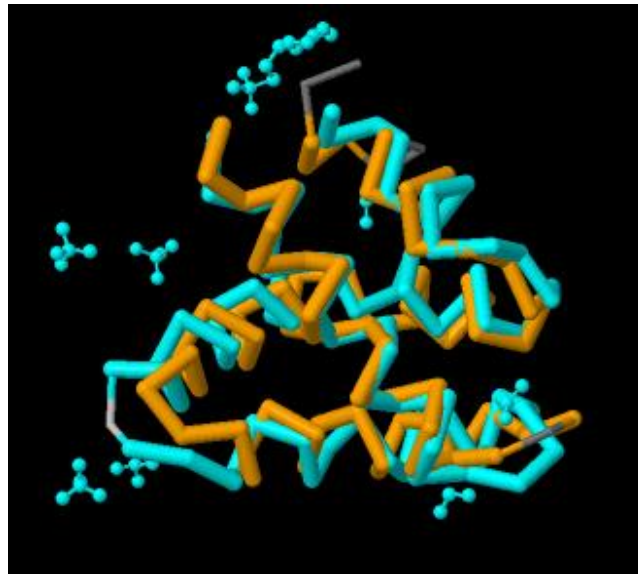


Figure 87 : Alignement de structure entre la NK-lysine (couleur orange, PDB ID : 1NKL) et la GLNY-9 kDa (couleur cyan, PDB ID : 1L9L) réalisé sur le site RSCB PDB (méthode jFATCAT-rigid).

La granulysine 15 kDa possède 7 résidus cystéines. Quatre d'entre eux forment probablement deux ponts disulfures identiques à ceux de la forme 9 kDa (Cys69-Cys132 et Cys96-Cys107). Cependant, aucune information n'est connue concernant l'état d'oxydation des trois autres cystéines (Cys43, Cys45, Cys140). Nous avons généré un modèle structural en utilisant l'outil I-TASSER²³³ qui est présenté dans la Figure 88. Cette structure hypothétique présente les deux ponts disulfures classiques du domaine *saposin-like* et placent la cystéine en C-terminal (Cys140) très loin des deux autres cystéines Cys 43 et 45. Nous prévoyons que la présence d'un noyau peu compact, de régions intrinsèquement non structurées et d'un nombre impair de résidus de cystéine rendent le repliement de cette protéine très complexe.

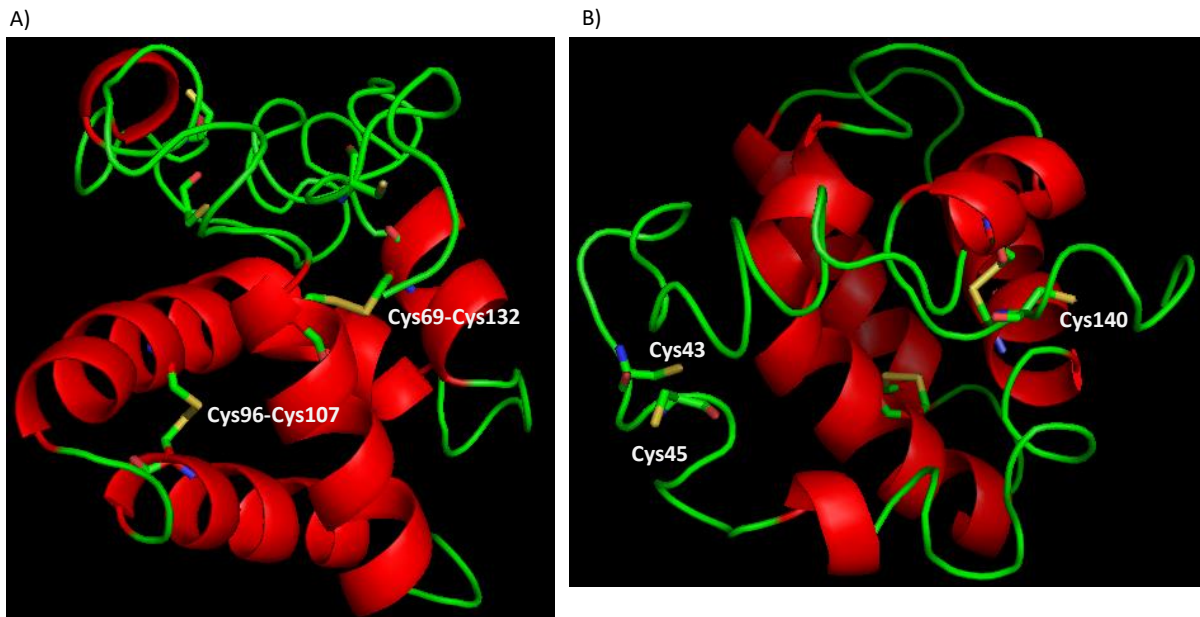


Figure 88 : Le modèle structural de la granulysine 15 kDa généré en ligne grâce à I-TASSER suite. A) Vue de côté montrant les deux ponts disulfures selon le domaine saposin-like. B) Vue de haut montrant les trois résidus Cys pour lesquelles il n’y a pas d’informations structurales disponibles.

Afin d’aider à la caractérisation structurale mais également d’étudier la fonction et donc le potentiel thérapeutique de la granulysine 15 kDa, nous avons réalisé sa synthèse par voie chimique.

4.2 Synthèse totale de la granulysine 15 kDa

L’assemblage de cette protéine par ligation chimique est complexe puisque la plupart des cystéines sont impliquées dans des jonctions difficiles (Figure 90). Nous avons choisi d’assembler la granulysine par ligation de trois segments peptidiques de tailles similaires : segment **A** (23-68), segment **B** (69-106) et segment **C** (107-123) présentés dans la Figure 89. Cette stratégie implique la formation de jonctions Thr-Cys et Val-Cys qui sont difficiles en raison de l’encombrement stérique. Initialement, nous avons décidé d’assembler ces segments peptidiques de l’extrémité N-terminale vers l’extrémité C-terminale.

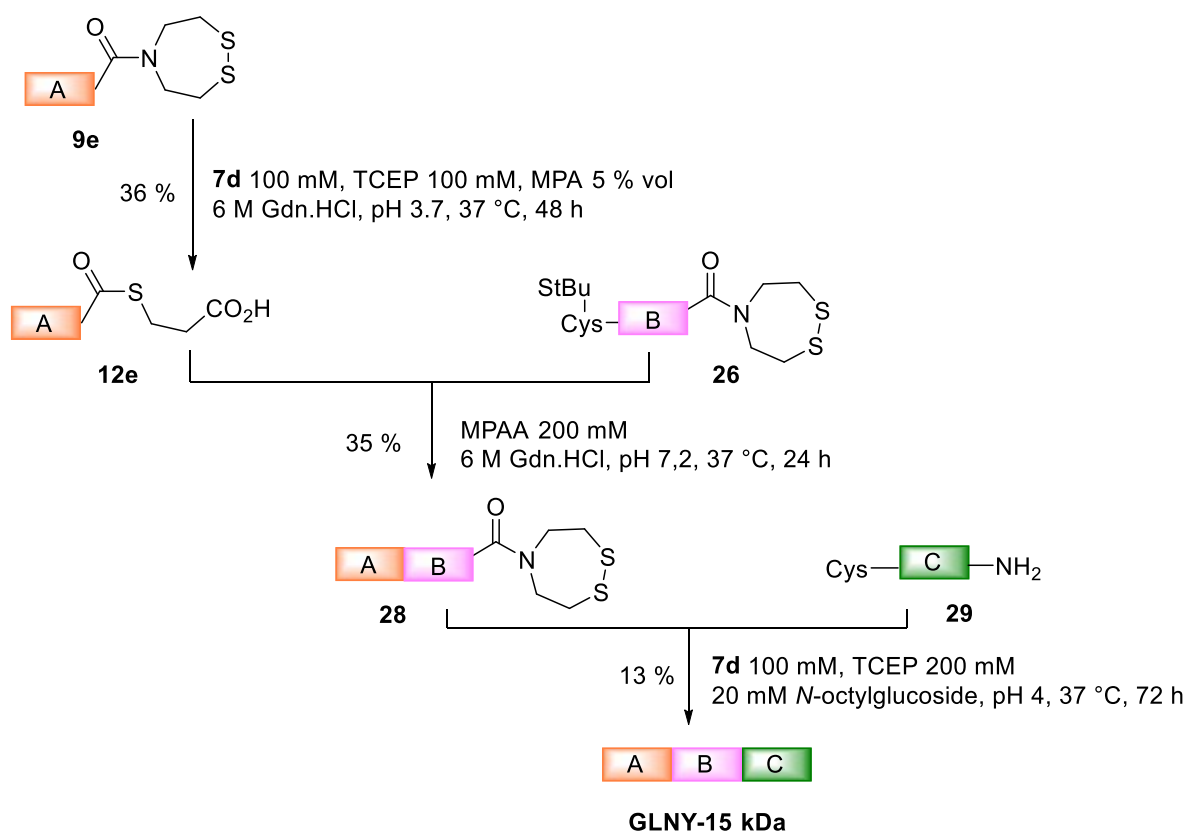
A	RLSPEYYDLARAHLRDEEKSCPCLAQEGPQGDLLTKTQ ELGRDYRT
B	CLTIVQKLKMKMVDKPTQRSVSNAATRVCRTGRSRWRDV
C	CRNFMRRYQSRVTQGLVAGETAQQICEDLRLCIPSTGPL-NH ₂

Figure 89 : Les trois segments peptidiques utilisés pour la synthèse totale de la granulysine 15 kDa.

4.2.1 Assemblage de N vers C

Dans un premier temps, le segment SEA **9e** a été converti en peptide thioester de MPA **12e** en présence du catalyseur **7d**. Cette réaction a été décrite dans la section 2.2.3 de ce chapitre. Puis, les segments **12e** et **26** ont été assemblés en utilisant une réaction de NCL. Cette ligation a été effectuée en présence du catalyseur MPAA mais en absence de réducteur (TCEP/DTT) afin de conserver le groupement SEA sous sa forme inactive.

Après purification, le peptide **28** a été engagé dans une réaction de ligation SEA avec le segment **29** en présence de *N*-octylglucoside. Lors d'une première expérience, l'utilisation de 6 M de Gdn.HCl avait conduit à la gélification du milieu réactionnel. Nous avons également testé l'utilisation de SDS mais ce détergent n'a pas permis d'éviter la gélification contrairement au *N*-octylglucoside. Finalement, nous avons obtenu la granulysine 15 kDa avec un rendement global de 1,6 % après purification HPLC.



RLSPEYYDLARAHLRDEEKSCPCLAQEGPQGDLLTKTQELGRDYRTCLTIVQKLKMKMVDKPT
 QRSVSNAATRVCRTGRSRWRDVCRNFMRRYQSRVTQGLVAGETAQQICEDLRLCIPSTGPL-NH₂

Figure 90 : Synthèse totale de la granulysine 15 kDa de N vers C. Les jonctions formées sont soulignées dans la séquence.

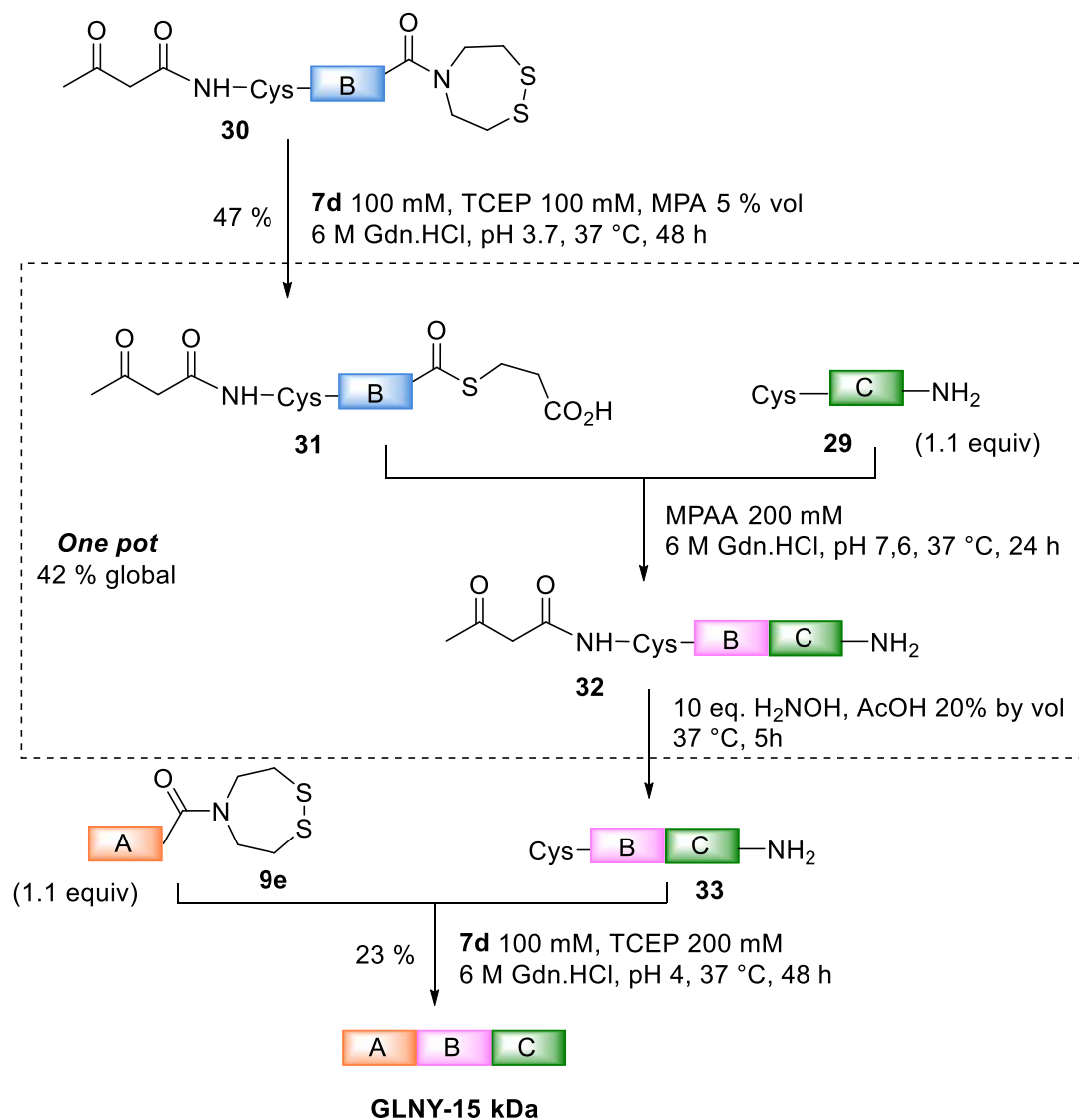
La difficulté de synthèse est en partie due à la faible solubilité des intermédiaires. Même en utilisant le *N*-octylglucoside le milieu réactionnel est trouble et non homogène ce qui limite la conversion. Dans l'idée de prévenir ces problèmes de solubilité, nous avons envisagé de produire la granulysine en assemblant cette fois-ci les segments de C vers N.

4.2.2 Assemblage de C vers N

Un assemblage dans cette direction nécessite d'équiper le segment **B** avec une fonction SEA en C-terminal mais également de protéger la cystéine en N-terminal pour prévenir la cyclisation ou polymérisation de celui-ci. Pour cela, nous avons utilisé le groupement acétoacétyle (AcA) qui a également été utilisé dans le chapitre IV lors de la synthèse de NK1. Cette fois-ci, le groupement acétoacétyle est introduit sur support solide après l'élongation par Fmoc-SPPS du peptide et la déprotection de l'amine en N-terminal⁵⁸. Les premiers essais de synthèse de ce fragment sur résine SEA ChemMatrix n'ont pas été concluants en raison de couplages incomplets après la Thr93. Afin de minimiser l'agrégation du peptide au cours de l'élongation, nous avons introduit deux unités *O*-acyl dipeptides au niveau des enchainements Ala92-Thr93 et Val98-Ser99 (Figure 91). Malgré un couplage toujours difficile entre les deux *O*-acyl dipeptides conduisant à la formation d'un fragment acétylé (*capped peptide*), nous avons obtenu le segment SEA **30** souhaité, protégé par l'AcA en C-terminal.

Le segment SEA **30** a ensuite été traité avec du MPA en présence du catalyseur **7d** à pH 3,7 pour produire le thioester correspondant **31** avec un rendement de 47 % après purification HPLC. De façon intéressante, le peptide **31** produit contient toujours les deux liens *O*-acyl isopeptide puisque le transfert d'acyle *O*→*S* n'a pas lieu dans les conditions d'échange. La présence de ces unités *O*-acyl dipeptides permet de maintenir une bonne solubilité du peptide **31**. Le peptide thioester **31** a ensuite été engagé dans une réaction de NCL avec le cystéinyl peptide **29** en présence de MPAA et à pH 7,6. Dans ces conditions, les liaisons *O*-acyl isopeptide se réarrangent rapidement pour former des liens amide natifs. Le groupement AcA a été retiré en *one-pot* par ajout d'acide acétique et d'hydroxylamine au milieu réactionnel. À noter que le MPAA a été extrait à l'aide de diéther. Nous avons ainsi obtenu le segment **32** avec un rendement global de 42 %.

Finalement, les segments **9e** et **32** ont été assemblés par ligation SEA à pH 4 en utilisant le disélényre **7d** comme catalyseur. Cette nouvelle stratégie, nous a permis d'isoler la granulysine 15 kDa avec un rendement global de 4,5 %, ce qui est nettement supérieur à la précédente approche de N vers C. La Figure 92 présente la caractérisation par LC-MS de la protéine granulysine 15 kDa obtenue. Une analyse par masse haute résolution est en cours ainsi que des tests de repliement. Nous espérons pouvoir tester son activité biologique et notamment antibactérienne très prochainement.



B CLTIVQKLKMKMVDKPTQRS **VSNAATR**VCRTGRSRWRDV

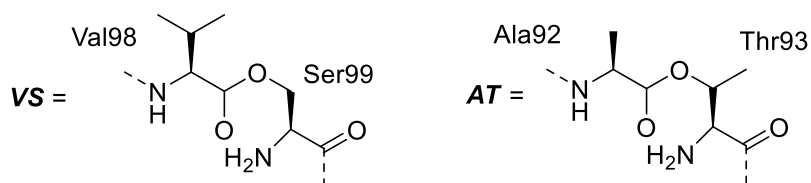


Figure 91 : Synthèse totale de la granulysine 15 kDa de C vers N.

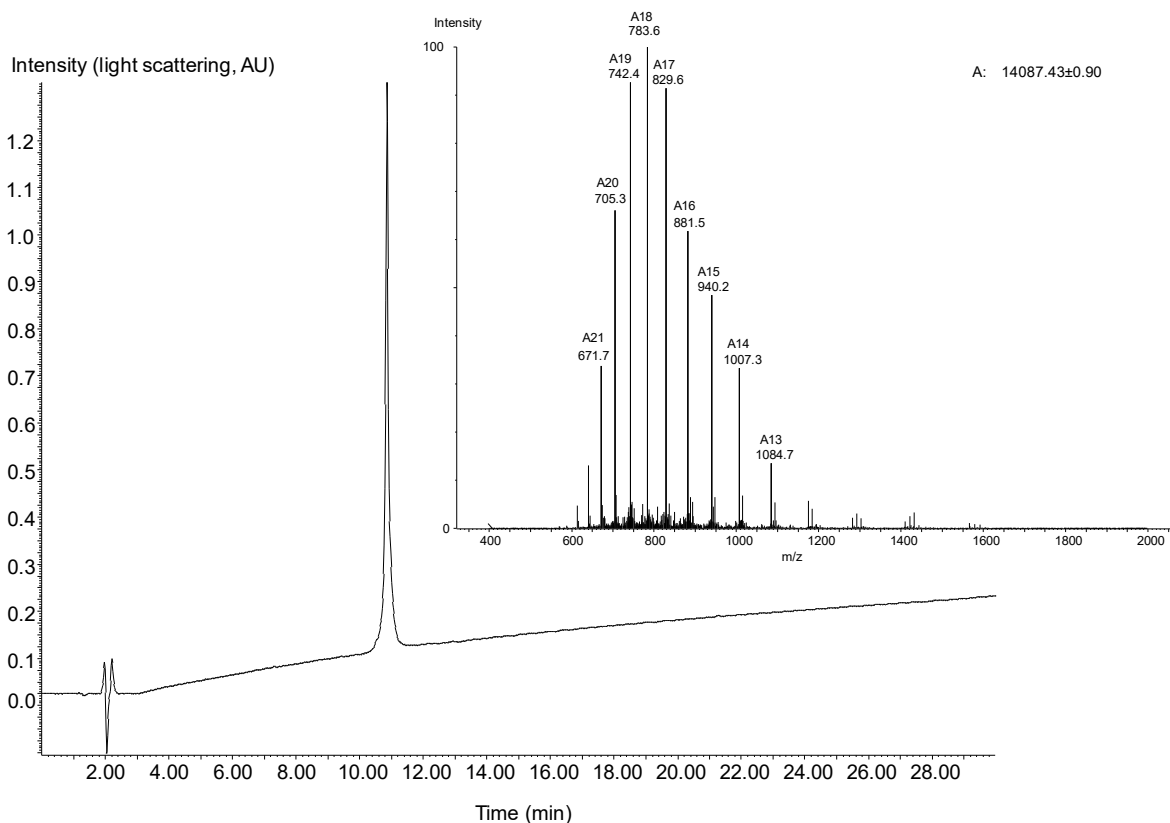


Figure 92 : Analyse LC-MS de la granulysine 15 kDa.

En résumé, nous avons développé de nouveaux catalyseurs pour la réaction d'échange SEA-thiol et la ligation SEA. Ces composés nous permettent de bénéficier de l'influence de la baisse du pH sur les cinétiques de transfert d'acyle $N \rightarrow S$ en maintenant les échanges thiol-thioester rapides à pH acide. Nous avons montré que ces catalyseurs sont bien plus efficaces que le MPAA ou le séléno-phénol et qu'ils facilitent la formation de jonctions difficiles. L'utilité de ces catalyseurs a été démontrée par la synthèse totale de la granulysine 15 kDa dont l'étude fonctionnelle et structurale est limitée à cause d'une production difficile par des méthodes biologiques. Nous pensons que la synthèse par voie chimique de la granulysine 15 kDa permettra d'aider à résoudre sa structure tridimensionnelle et d'approfondir l'étude de ses fonctions biologiques. En effet, la synthèse par voie chimique de la granulysine 15 kDa nous a permis de réaliser une étude préliminaire sur sa structure tridimensionnelle. Les premiers essais de repliement ont montré deux formes : une première avec deux ponts disulfures et une seconde avec trois ponts disulfures. L'analyse par spectrométrie de masse de ces formes est toujours en cours au laboratoire et nous espérons apporter prochainement de plus amples informations sur la structure tridimensionnelle de cette protéine. L'ensemble de ce travail est en cours de valorisation sous forme d'un article qui sera soumis au journal *Chemical Science*.

CHAPITRE VI : DÉVELOPPEMENT D'UN PROCÉDÉ AUTOMATISÉ DE SYNTHÈSE DE PROTÉINES EN PHASE SOLIDE

1. La problématique d'assemblage de segments peptidiques en solution

Jusqu'à présent, la majorité des protéines produites par synthèse chimique ont été assemblées en solution en utilisant la réaction NCL et les méthodes apparentées. La synthèse de grandes protéines nécessite inévitablement un nombre important d'étapes chimiques. Outre la ligation chimique elle-même, les stratégies de synthèse peuvent inclure plusieurs étapes d'activation (conversion de thioesters latents en thioesters actifs), étapes de déprotection (déprotection de résidus cystéines) ou une étape de désulfuration pour convertir des cystéines ou d'autres acides aminés thiols en résidus protéiniques. Les conditions expérimentales utilisées pour une transformation chimique donnée ne sont pas toujours compatibles avec l'étape suivante. En conséquence, un grand nombre de transformations chimiques implique d'isoler en général un ou plusieurs intermédiaires. De plus, chaque étape chimique ou de purification peut être compliquée par la faible solubilité des segments ou leur tendance à former des agrégats.

Dans le chapitre IV, nous avons vu que les stratégies basées sur des réactions en *one-pot* permettent de minimiser le nombre d'étapes chimiques et de purifications nécessaires pour obtenir la protéine cible. On économise ainsi du temps et du rendement. Ici, nous proposons d'utiliser une nouvelle stratégie d'assemblage séquentielle de segments peptidiques en phase solide pour faciliter la synthèse de grandes protéines par voie chimique.

2. Émergence de la ligation native en phase solide (SPCL)

Le concept de ligation chimique native en phase solide (SPCL) a été introduit par l'équipe de Kent en 1999²⁰¹. Cette méthode combine les avantages de la NCL et de la SPPS. La NCL permet l'assemblage chimiosélectif de segments peptidiques non protégés et conduit à la formation d'un lien natif. La synthèse sur une phase solide permet l'élimination des réactifs en excès par filtration, ce qui permet de ne pas isoler les intermédiaires. Le peptide immobilisé ne peut pas précipiter et, en ajustant la charge du support solide, l'agrégation du peptide peut être minimisée pour assurer une plus grande accessibilité des sites réactifs.

Compte tenu du fort potentiel de la phase solide, il est surprenant de constater que le nombre de publications sur le sujet est très faible (12 protéines dans 7 publications). L'une des principales raisons au manque de références dans le domaine est la difficulté de concevoir un support solide qui soit compatible avec l'ensemble des étapes chimiques, et en particulier qui permette l'accrochage

chimiosélectif du premier segment de manière efficace et simple et plus tard le décrochage de la protéine cible. Une autre limitation importante est due à la difficulté à adapter les méthodes développées en solution sur la phase solide. La formation des fonctions thioester est notamment un point crucial.

Dans ce chapitre, nous proposons de décrire la conception d'un processus automatisé d'élongation par ligation chimique basé sur les propriétés du groupement SEA. Ce travail comprend le développement d'un bras espaceur, l'adaptation des conditions réactionnelles en phase solide et l'optimisation des caractéristiques techniques du prototype. Nous pensons que ce procédé présente un grand potentiel pour la synthèse de grandes protéines par voie chimique.

3. Conception d'une stratégie d'assemblage en phase solide

3.1 Procédé chimique : la SEA SPPS

Le procédé de ligation en phase solide mis en place au laboratoire est basé sur les propriétés du groupement SEA. Comme il a déjà été mentionné, ce dernier est stable dans les conditions de la NCL en absence de réducteur fort et agit donc comme un thioester latent. On peut ainsi imaginer un cycle d'élongation composé d'une étape d'activation permettant de convertir le groupement SEA en alkylthioester (en rouge, Figure 93), suivie d'une étape de ligation NCL (en orange, Figure 93). L'utilisation d'un support solide permet d'éliminer facilement les réactifs et segments peptidiques qui n'auraient pas réagi. On peut en principe répéter le cycle d'élongation autant de fois que nécessaire pour atteindre une protéine cible. Cette technique est appelée SEA SPPS.

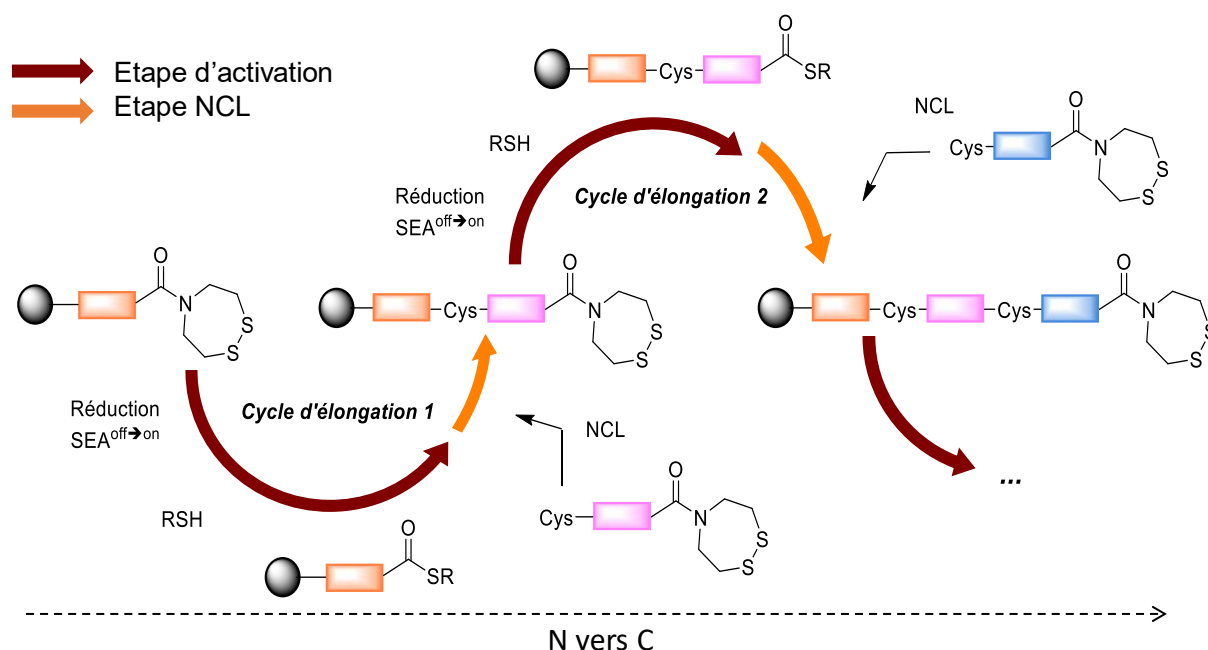


Figure 93 : Principe général du procédé de ligation en phase solide développé à partir de segments SEA (SEA SPPS).

La mise en place d'un tel procédé implique l'utilisation d'un support solide qui soit compatible avec les conditions réactionnelles (milieux aqueux) et stable en présence des différents réactifs utilisés (thiols, nucléophiles, réducteurs). La conception d'un bras espaceur (*linker* en anglais) est également nécessaire. Celui-ci doit permettre l'attachement chimiosélectif du premier fragment. Le bras espaceur doit également être stable dans les conditions d'accrochage et d'élongation du processus. Enfin, les conditions de coupure de ce dernier doivent être simples pour libérer la protéine du support après assemblage.

Des travaux antérieurs du laboratoire ont décrit l'utilisation d'un bras espaceur baso-labile, le 2-[2-(2-(azido-éthoxy)éthyl sulfonyle)-2-éthoxycarbonyl]-2-éthoxycarbonyl (N_3 -Esoc). De fait, l'extrémité N-terminale du premier segment SEA^{off} est fonctionnalisée par le groupement N_3 -Esoc en utilisant la méthode développée par le groupe d'Aucagne (Figure 94)²³⁴. Le support solide est modifié par une fonction alcyne qui permet l'attachement du premier segment par réaction de cycloaddition 1,3 de Huisgen catalysée au cuivre (CuAAC).

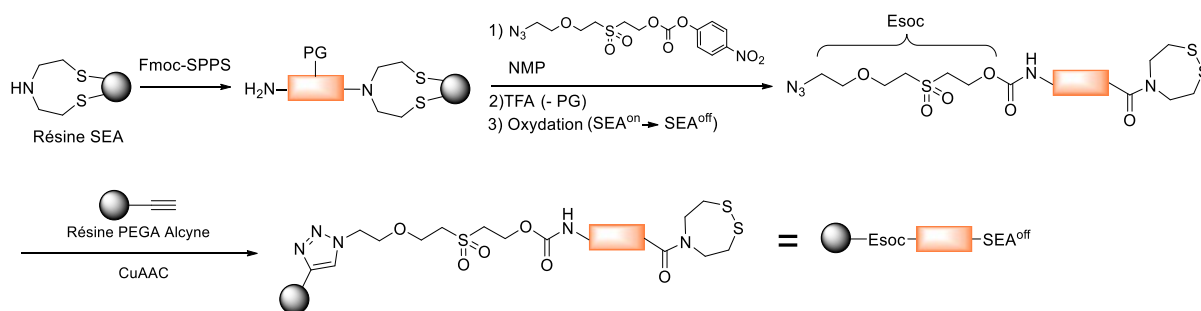


Figure 94 : Principe de l'attachement chimiosélectif du premier segment sur support solide grâce au bras espaceur Esoc.

Ce bras espaceur a notamment permis la synthèse d'un peptide de 136 acides aminés par assemblage séquentiel de 5 segments peptidiques suivant le procédé présenté à la Figure 93²³⁵. Néanmoins, ce bras espaceur implique des conditions de coupure fortement basiques (pH > 11) qui ne sont pas compatibles avec un certain nombre de protéines. De plus, la formation d'une vinylsulfone par β -élimination du groupement Esoc peut conduire à l'alkylation des cystéines²³⁶.

Plus récemment, un autre bras espaceur de type acétoacétamide (AcA) a été développé au sein du laboratoire. Ici, le premier segment peptidique est modifié après Fmoc-SPPS par l'acétoacétyl N-succinimidyl ester (AcA-OSu) puis attaché au support solide par formation d'un lien oxime (Figure 95). La coupure de ce bras espaceur est réalisée par réaction de transoximation avec l'hydroxylamine à pH acide, puis cyclisation spontanée de l'intermédiaire obtenu.

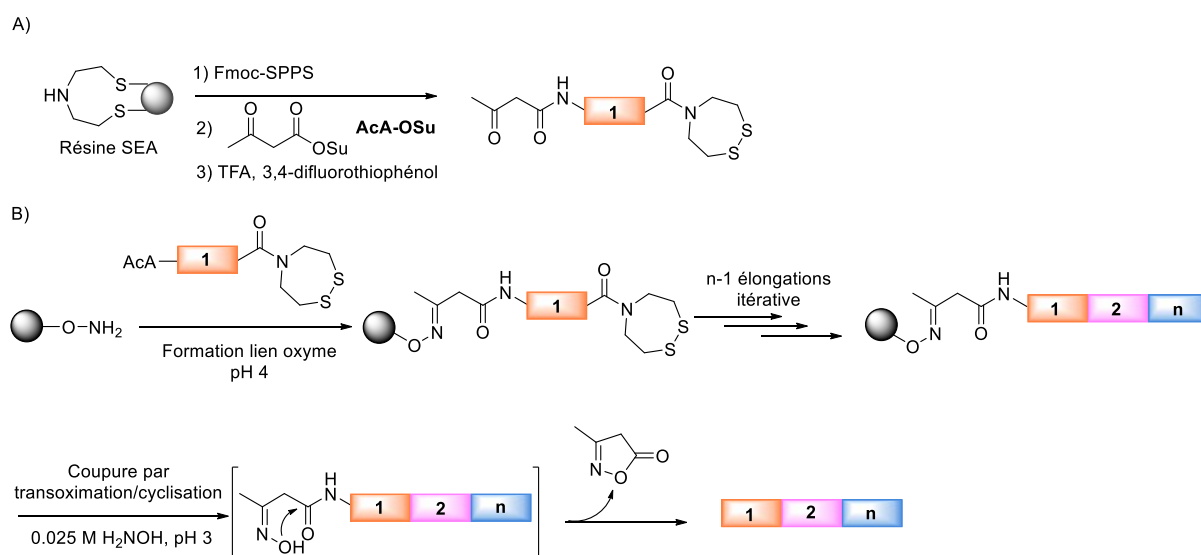


Figure 95 : Synthèse de protéines par assemblage séquentielle de N vers C en utilisant le linker AcAO et la réaction de NCL. A) Synthèse des segments peptidiques AcA par SPPS. B) Stratégie d'assemblage.

Ce procédé permet d'obtenir le polypeptide sans groupement résiduel provenant du linker et a notamment permis la synthèse de la protéine NK1-B. Cependant, une réaction secondaire a été

observée lors de l'étape de coupure. En effet, l'hydroxylamine qui est très nucléophile peut réagir avec les amides primaires et en particulier avec les chaînes latérales de Asn et Gln. Les conditions de coupure ont été optimisées pour minimiser cette réaction secondaire mais elle n'a pas pu être inhibée totalement. De plus, les réactifs permettant la synthèse des peptides AcA qui étaient à l'origine commercialement disponibles ne le sont plus à ce jour.

Pour conclure, le nombre de bras espaceurs décrit dans la littérature qui soient compatibles avec les conditions réactionnelles du procédé de SEA SPPS est très faible. Ce constat nous a poussé à nous intéresser à la synthèse d'un nouveau bras espaceur.

3.2 Développement d'un bras espaceur chimiosélectif

Il est important de noter que nous souhaitons que notre procédé de synthèse de protéines permette la production de protéines natives. L'idée est de présenter une alternative aux méthodes recombinantes qui puisse permettre d'accéder rapidement à une librairie de protéines d'intérêt biologique. Une fois une protéine d'intérêt identifiée, celle-ci pourra être produite à plus grande échelle en utilisant des systèmes vivants qui sont moins coûteux. Par conséquent, nous ne souhaitons pas utiliser de bras espaceurs qui introduisent après coupure un élément non naturel comme par exemple un polymère.

Pour la conception de notre nouveau bras espaceur, nous avons utilisé des réactions chimiques simples et des réactifs disponibles commercialement (Figure 96). Cette stratégie permet une certaine modularité du point de vue de la résine utilisée et de la charge finale de celle-ci.

Dans un premier temps, l'acide aminé Fmoc-Glu(*t*Bu)-OH commercial est couplé à la *bis*(2-(tritylthio)éthyl)amine dont la synthèse est décrite dans la littérature²¹². Les groupements protecteurs trityles (Trt) sont éliminés par traitement à l'iode et les thiols sont oxydés pour conduire à la formation du groupement SEA^{off}. C'est ce groupement qui permettra l'attachement chimiosélectif du premier fragment en utilisant le cycle d'élongation transthioestérification/NCL décrit plus haut (Figure 93). Pour permettre l'attachement de l'acide aminé à la résine équipée d'un linker de Rink, la chaîne latérale de l'acide aminé est déprotégée en utilisant un cocktail de coupure classique. Les conditions de couplage à la résine seront détaillées dans la suite de ce manuscrit. Il est important de noter que la présence d'un groupement Fmoc nous permet de déterminer la charge de notre support solide en utilisant un dosage par spectroscopie UV²¹². Finalement, l'amine en N-terminal est totalement déprotégée et acétylée mais naturellement d'autres modifications sont possibles.

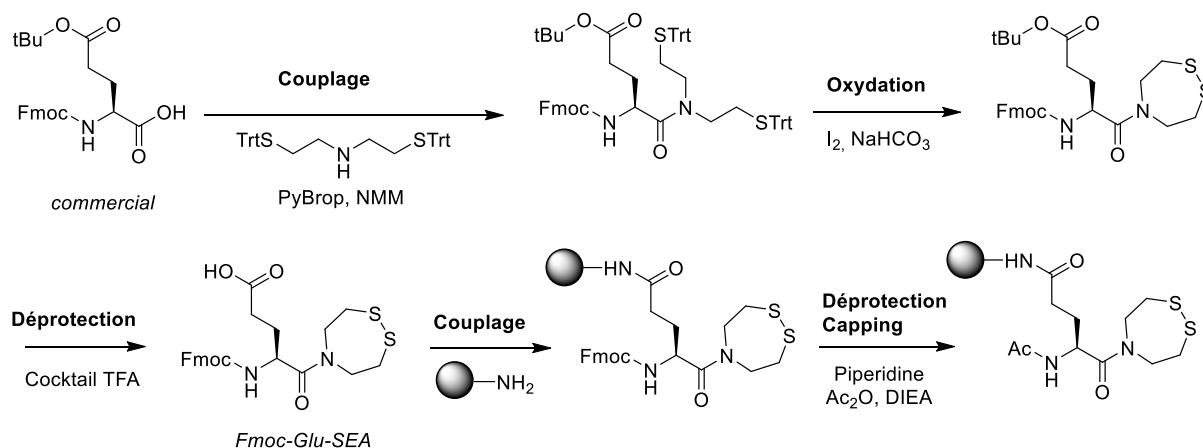


Figure 96 : Schéma de synthèse du linker Fmoc-Glu-SEA et couplage sur le support solide.

Les conditions de coupure pour un tel bras espaceur sont identiques à celles utilisées en SPPS classique. Il s'agit d'un cocktail à base de TFA et d'un certain nombre de pièges à carbocation (*scavenger*) suivant la séquence de la protéine synthétisée. Il est important de noter que cette méthode entraîne après coupure du linker de Rink l'ajout d'un acide aminé Gln à l'extrémité N terminale du polypeptide cible. Cette méthode est simple, peu coûteuse et à priori compatible avec de nombreux polypeptides. Néanmoins, la recherche d'un linker universel est toujours d'actualité au laboratoire et fait partie des perspectives de ce projet.

3.3 Contraintes du procédé

L'assemblage en phase solide est soumis à un certain nombre de contraintes.

La solubilité des segments peptidiques est un point critique pour toutes stratégies d'assemblages que ce soit en solution ou sur support solide. Dans le cadre d'un procédé automatisé, il est d'autant plus important que les segments soient bien solubles. En effet, le mécanisme implique un certain nombre de vannes, tuyaux, seringues qui pourraient être obstrués si les segments ou les réactifs n'étaient pas complètement solubles. Pour éviter cela, il est important de pré-solubiliser les segments peptidiques manuellement. L'une des forces du procédé que nous développons résulte de la grande stabilité des peptides SEA qui restent stables en solution sur des temps prolongés, ce qui n'est pas le cas des thioesters classiques (hydrolyse rapide). Ainsi, les segments peptidiques sont solubilisés en amont dans un tampon à pH 7 contenant du Gdn.HCl. La pré-solubilisation des segments implique que toutes les cystéines des segments à incrémenter soient protégées avec le groupement StBu afin d'éviter l'oxydation de celles-ci au cours du temps.

Les réactifs (thiols, nucléophiles, réducteurs) doivent également être solubles dans les conditions réactionnelles. Les différents réactifs utilisés sont donc mis en solution au préalable. Pour

l'étape de conversion du peptide SEA en thioester d'alkyle, le thiol exogène tel que le MPA et la TCEP sont conditionnés au préalable dans un tampon à pH 4. Pour la réaction de NCL, nous avons opté initialement pour une solubilisation du MPAA de façon extemporanée. En effet, en solution au pH de ligation (pH 7), le MPAA s'oxyde rapidement. Cette stratégie implique de trouver un moyen de solubiliser le MPAA solide juste avant son utilisation et de façon automatisée. Cette méthode ne s'est avérée satisfaisante et nous verrons comment nous avons surmonté cette difficulté.

Enfin, les processus en phase solide impliquent en général des temps réactionnels plus long qu'en phase liquide. De plus, nous avons vu dans le chapitre V que l'utilisation de Gdn.HCl entraîne une augmentation significative des cinétiques de l'étape d'activation du SEA en thioester. Il est donc important de trouver des conditions (température, pH) et des additifs (catalyseurs) qui permettent d'accélérer cette transformation. C'est pourquoi une grande partie de mon travail de thèse a porté sur la mise au point de nouveaux catalyseurs pour la conversion des peptides SEA en peptides alkylthioesters et la réaction de ligation SEA (chapitre V). Nous verrons plus tard comment ces travaux ont été appliqués à la phase solide.

4. Synthèse supportée et automatisée de protéines

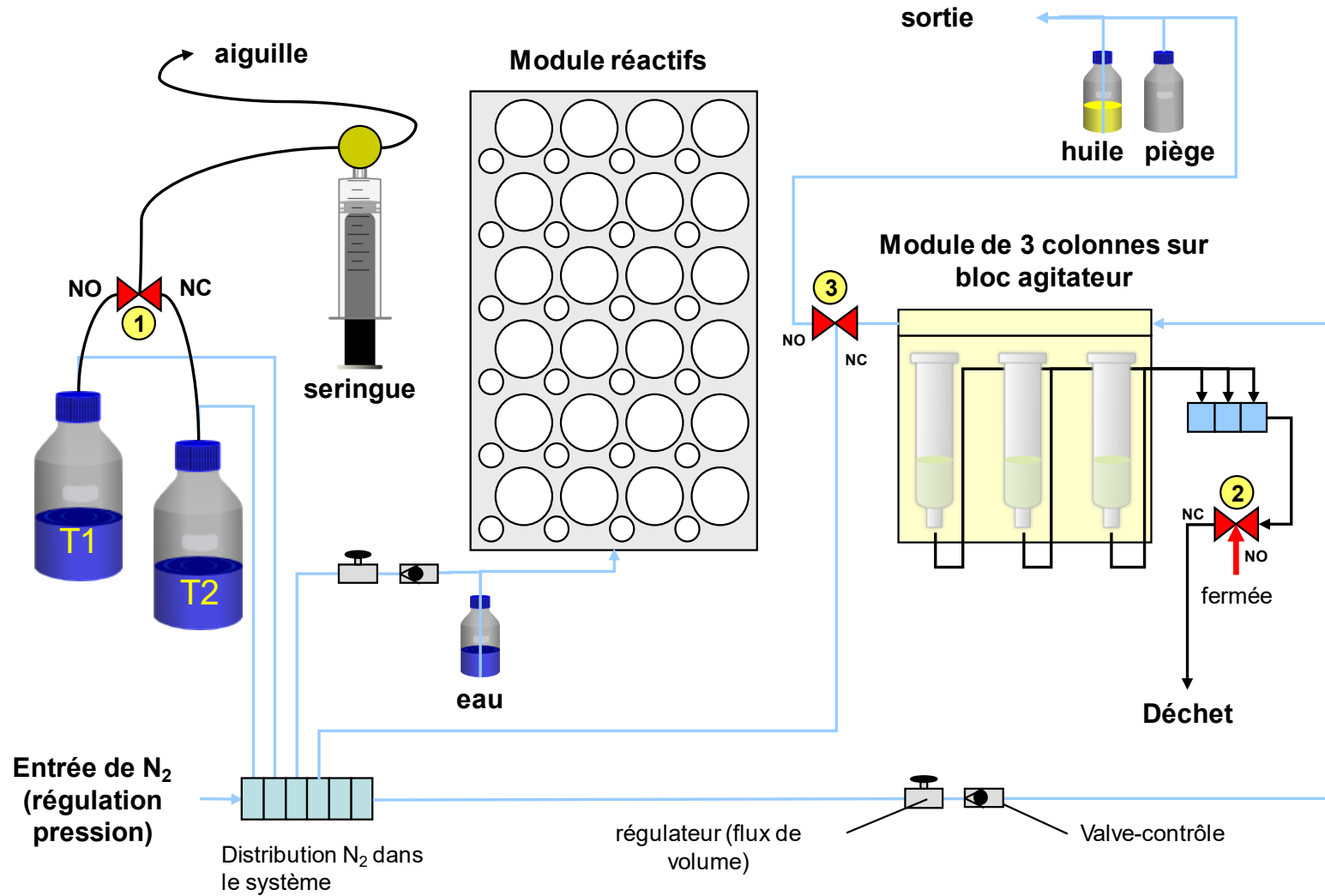
4.1 Caractéristiques techniques du prototype

Le prototype mis au point au laboratoire est basé sur le synthétiseur de peptide ResPep SL de la société Intavis. L'idée ici est d'utiliser le même appareil pour réaliser la synthèse de segments peptidiques et la synthèse de protéines par assemblage de ces segments. Il suffit pour cela de changer les réactifs de SPPS (activateurs, acides aminés, etc.) par les réactifs de la SEA SPPS (thiols, réducteurs, segments peptidiques).

Un premier prototype a été développé au sein du laboratoire avant ma thèse. Sans entrer dans les détails, celui-ci était composé d'un bloc pour les réactifs, d'un module pour les colonnes de réaction composé d'un régulateur de température et d'un bras dispenseur. Afin d'éviter les problèmes d'oxydation le prototype présentait également un système de distribution d'azote. Ainsi, le bloc contenant les réactifs était sous atmosphère inerte en continu. Ce flux d'azote permettait également de vider les colonnes lors des lavages en exerçant une pression au sein de celles-ci. Toutefois, il nous est rapidement apparu primordial d'optimiser et de simplifier le prototype. C'est de cette dernière configuration dont nous allons discuter dans la suite de ce manuscrit.

Le nouveau prototype est plus modulaire, il présente trois configurations possibles mais nous allons discuter ici de la version permettant un balayage sous azote continu à la fois du bloc réactif mais également des colonnes (Figure 97). La présence de deux régulateurs d'azote l'un pour le bloc réactif

et l'autre pour le module colonne nous permet une régulation précise du volume d'azote au sein du système et donc de minimiser les problèmes d'oxydation. De façon pratique, l'accès aux différents éléments et notamment aux valves composant le prototype a été facilité pour un contrôle de l'état de celles-ci plus régulier et facilité. De plus, le nombre de valves a été réduit significativement (3 valves contre 6 initialement) ce qui permet de diminuer le risque d'échec de synthèse dû à l'obstruction ou à une panne de l'une de ces valves. Enfin, un module d'agitation a été ajouté au système pour permettre une meilleure homogénéisation du milieu réactionnel. Ce bloc a également été conçu pour permettre une meilleure diffusion de la chaleur au sein des colonnes.



NO : Normally Open - NC : Normally Closed

Figure 97 : Schéma technique du prototype développé au laboratoire.

4.2 Mise en place de la synthèse

4.2.1 Synthèse du support solide

Le développement du prototype a nécessité de nombreuses expériences préliminaires. En amont des synthèses effectuées sur l'appareil, des expérimentations « manuelles » ont été réalisées afin de mettre au point les conditions opératoires initiales. Cette phase préparatoire n'est pas présentée en détail dans ce manuscrit qui se consacre exclusivement au développement de l'automatisation du processus. Toutefois grâce à ce travail, nous avons notamment déterminé le nombre de lavages nécessaires pour éliminer les réactifs entre deux étapes chimiques. Par analyse HPLC des solutions de rinçage, nous avons conclu que dix lavages du support solide sont nécessaires pour éliminer les réactifs qui n'ont pas réagi. À partir des résultats obtenus durant ce travail préliminaire nous avons mis en place une première expérience de synthèse sur le prototype.

Préalablement, le bras espaceur Fmoc-Glu-SEA (Figure 96) est greffé sur une résine de type H-Rink Amide ChemMatrix®. Cette résine est exclusivement composée de polyéthylène glycol et présente une bonne solvatisation dans l'eau. Les conditions de couplage sont présentées dans la Figure 98. Nous utilisons le Boc-Glu(OtBu) comme compétiteur afin de diminuer significativement la charge initiale (0,18 mmol/g) et le PyBop comme agent de couplage. Puis, en utilisant une méthode de dosage par spectrométrie UV (voir Annex A, Chapter VI, Section 2.2), nous déterminons une charge finale après couplage de 0,073 mmol/g. Finalement, l'amine en N-terminal est déprotégée et acétylée. Ce support solide est ensuite utilisé pour réaliser l'assemblage de segments peptidiques à l'aide du prototype décrit précédemment.

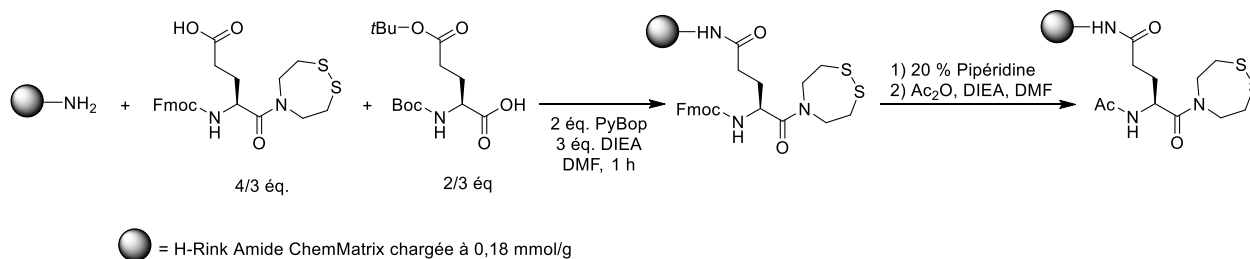


Figure 98 : Couplage du Fmoc-Glu-SEA sur la résine H-Rink Amide ChemMatrix®.

4.2.2 Procédure d'élongation

La procédure d'élongation permettant l'assemblage séquentielle de segments peptidiques en phase solide et de façon automatisée est décrite ci-après. À noter que pour toutes les synthèses décrites dans la suite du manuscrit nous travaillons sur une échelle de synthèse de 2,5 μ moles.

Dans un premier temps, la résine est conditionnée dans le tampon utilisé pour l'étape d'activation à savoir 5 M de Gdn.HCl 0,1 M tampon acétate à pH 4. Puis, la fonction SEA du bras espaceur est activée en présence de TCEP (100 mM) et de MPA (5 %). En pratique, la TCEP (200 mM) et le MPA (10 % volume) sont mis en solution séparément dans le tampon d'activation afin d'éviter d'éventuelles réactions secondaires au cours du temps comme la désulfuration du MPA. À noter que ces solutions serviront pour toutes les étapes activation nécessaires à l'obtention de la protéine cible. L'appareil prélève 250 µL de chacune des solutions, les mélange et ajoute 500 µL de ce mélange d'activation sur la résine. Il est important de noter que la présence des billes de résine induit un volume mort. Il est donc primordial de renouveler plusieurs fois la solution d'activation pour atteindre les concentrations souhaitées en réactif. La Figure 99B présente le cas de l'évolution de la quantité de MPA dans la colonne lors de l'étape d'activation. On remarque qu'au moins six renouvellements de solution sont nécessaires pour atteindre le pourcentage classique de MPA utilisé lors de l'échange SEA-thiol, qui est de 5%. Usuellement, le prototype est programmé pour renouveler la solution dix fois durant les 12 heures d'étape d'activation.

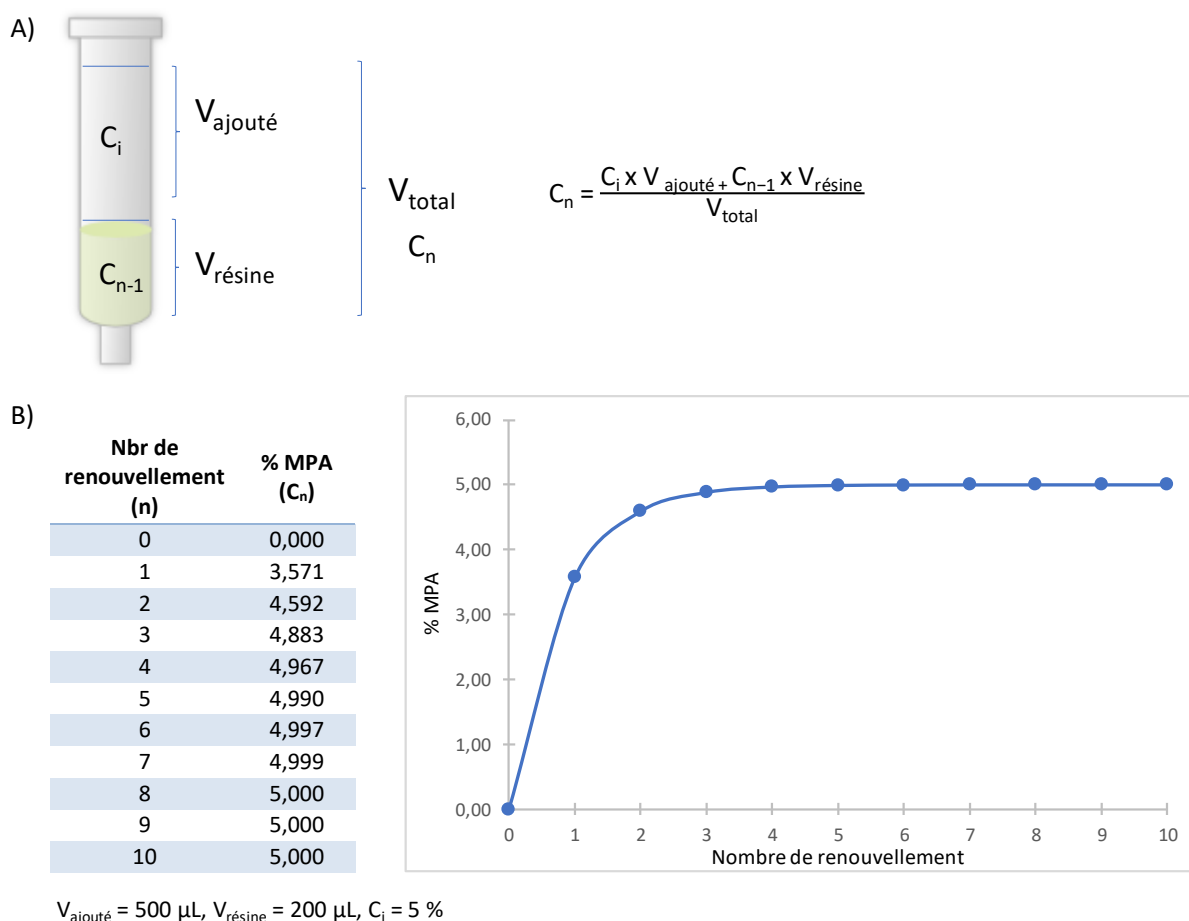


Figure 99 : A) Principe général du calcul de quantité réelle de réactif au sein de la colonne. B) Exemple de l'évolution de la quantité de MPA en fonction du nombre de renouvellements de la solution d'activation.

De la même façon, une étude de l'évolution de la température au sein des colonnes a été réalisée. On a observé un écart de 2 °C entre la valeur de consigne et la température réelle au sein des colonnes. Ainsi, la température consigne lors de l'étape d'activation est fixée à 39 °C puisque la température visée est de 37 °C.

Une fois l'étape d'activation terminée, les colonnes sont lavées avec le tampon d'activation (5 M de Gdn.HCl 0,1 M tampon acétate à pH 4) afin d'éliminer le MPA et la TCEP qui n'auraient pas réagi. Comme nous l'avons mentionné précédemment, dix lavages sont nécessaires pour laver correctement les colonnes. Puis, celles-ci sont conditionnées pour l'étape de ligation. Comme pour l'étape précédente, dix lavages avec 5 M de Gdn.HCl 0,1 M tampon phosphate à pH 7,5 sont réalisés dans ce but.

Pour l'étape de ligation, la solubilisation du MPAA est un point essentiel. Pour éviter son oxydation, une option est de le solubiliser juste avant son utilisation. Si l'on utilise le MPAA sous forme solide et que l'on ajoute du tampon phosphate à pH 7,5, celui-ci va précipiter en raison de la chute de pH dû à sa fonction acide carboxylique. Même en réajustant le pH par ajout de soude, une agitation vigoureuse est nécessaire pour solubiliser le MPAA correctement. Ce protocole classique en synthèse manuelle n'est pas applicable à notre prototype. C'est pourquoi, nous avons cherché un autre moyen de pré-solubiliser le MPAA tout en évitant le processus d'oxydation.

En amont, le MPAA est solubilisé dans un mélange de 0,2 M NaOH et *t*-BuOH. Cette solution est congelée et lyophilisée. On obtient ainsi le sel de 4-mercaptophénylacétate de sodium, MPAA-Na. Ce composé est introduit sous forme solide dans le prototype. Lors de l'étape de ligation, 55 µL de NaOH 6 N, 445 µL de 5 M de Gdn.HCl 0,1 M tampon phosphate à pH 7,5 sont utilisés dans cet ordre pour solubiliser le MPAA-Na et obtenir un pH final de 7,5. Les segments peptidiques sont quant à eux préalablement solubilisés avec 5 M de Gdn.HCl 0,1 M tampon phosphate à pH 7,5 (280 µL pour 5 µmol de peptide soit 2 équivalents). Finalement, l'appareil prélève 250 µL de la solution du segment à incrémenter, 250 µL de solution de MPAA et ajoute le tout sur la résine. La température de consigne est fixée à 39 °C et le milieu réactionnel est mis sous agitation pendant 12 heures. La synthèse de segments peptidiques peut être complexe et représenter un coût considérable. C'est pourquoi il ne nous semble pas judicieux de renouveler l'étape de couplage pour forcer la conversion. Seul un excès de 2 équivalents de segment peptide par rapport au support solide est appliqué. Une fois la ligation effectuée, les colonnes sont lavées avec 5 M de Gdn.HCl 0,1 M tampon phosphate à pH 7,5 afin d'éliminer le MPAA et les segments qui n'auraient pas réagi. Puis la résine est à nouveau conditionnée dans le tampon d'activation et le cycle d'élongation est répété pour chaque segment à incrémenter.

Ce procédé nous a permis de réaliser l'assemblage de trois segments peptidiques pour former le peptide modèle **37** composé de 23 acides aminés.

4.3 Résultat préliminaire/ Premier résultat

La synthèse du peptide **37** implique la formation d'une jonction Gly-Cys et Ala-Cys. La Figure 100 présente les trois fragments utilisés pour réaliser l'assemblage de celui-ci en utilisant la procédure d'élongation détaillée dans la section précédente.

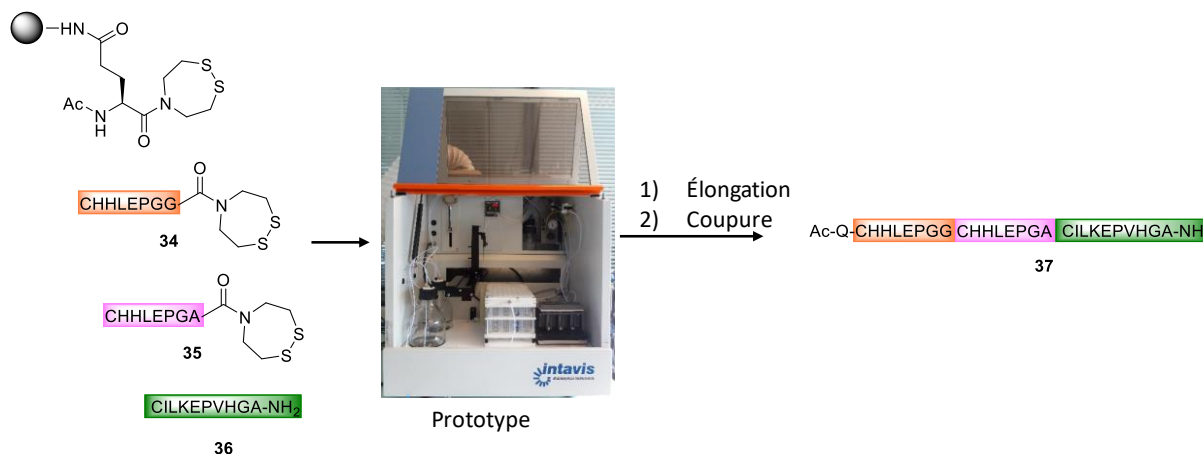


Figure 100 : Synthèse du peptide **37**.

Une fois l'élongation sur le prototype terminée, la résine a été lavée avec de l'eau puis du méthanol pour éliminer les réactifs et la Gdn.HCl résiduels. La résine a ensuite été lavée à l'éther et séchée. La nature du bras espaceur permet l'utilisation de conditions de coupure classiques avec un cocktail de TFA /scavengers adapté à la séquence de la protéine synthétisée. Ici, nous avons utilisé le mélange TFA/TIS/Anisole/Thiophénol/H₂O (85/5/5/2,5/2,5). Finalement, nous avons obtenu 2,40 mg de peptide **37** brut soit un rendement de 27 % pour sept étapes chimiques (élongation + coupure). L'analyse par LC-MS montre une pureté de 63,7 % (pic E, Figure 101). L'impureté majoritaire résulte de la délétion du premier fragment **34** qui peut être due soit à une mauvaise activation du bras espaceur soit à une ligation incomplète du premier segment (pic D ~ 20 %, Figure 101). Nous observons également la délétion du second fragment **35** dans une plus faible proportion (pic C ~ 5,4 %, Figure 101). Enfin, l'un des produits secondaires résulte de l'hydrolyse du thioester de MPA lors du dernier cycle d'élongation (pic A ~ 7,5 %, Figure 101).

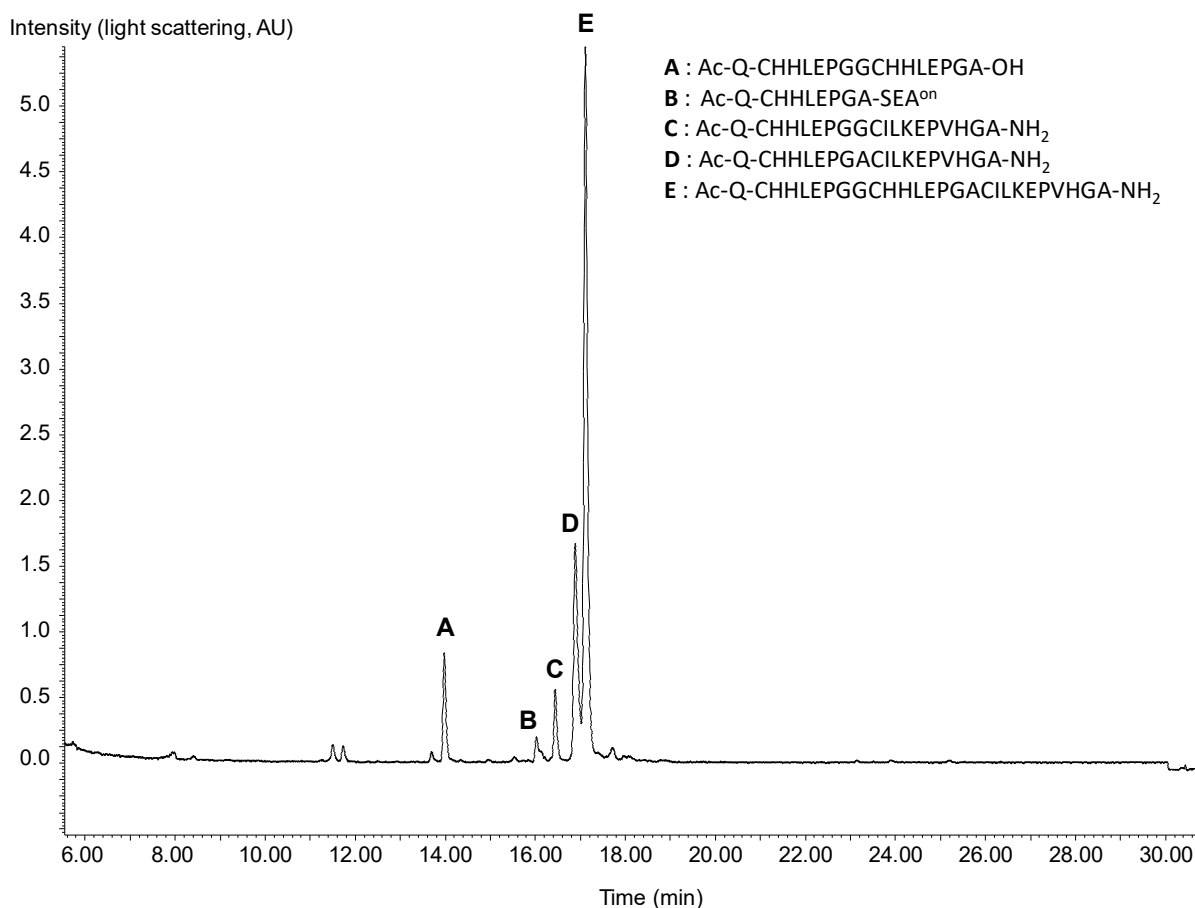


Figure 101 : Analyse LC du brut après élongation/coupure. Le pic E correspond au peptide **37**.

Finalement après purification HPLC, nous obtenons 1,2 mg du peptide **37** soit un rendement de 13 % au global. Bien que le modèle étudié ne présente pas de difficultés, ce premier résultat est très encourageant. Par ailleurs, si l'on souhaite atteindre des protéines de grandes tailles, il faut multiplier les cycles d'élongation. Une réaction secondaire infime mais récurrente peut entraîner une baisse significative du rendement final. Il est donc indispensable d'améliorer chaque étape du cycle d'élongation et en particulier l'activation du bras espaceur.

5. Optimisation

Pour améliorer le procédé, nous pouvons jouer sur plusieurs paramètres. La suite de ce chapitre ne présente qu'une partie des nombreuses expériences réalisées afin d'optimiser le procédé en prévenant la délétion du premier fragment mais également en optimisant chaque étape du processus pour augmenter le rendement global.

5.1 Processus chimique

5.1.1 L'étape d'activation : l'échange SEA-thiol

Le MPA est le thiol exogène classiquement utilisé pour la conversion d'un peptide SEA en alkylthioester. Cependant, dans le cas de la phase solide, nous pensons que la fonction acide carboxylique du MPA pourrait jouer un rôle dans le détachement acido-catalysé du bras espaceur au cours de l'élongation. Cette hypothèse est en accord avec certains résultats expérimentaux que nous avons obtenu lors de la phase de test manuelle. En effet, nous avons observé une chute significative du rendement avec l'augmentation du nombre de cycle d'élongation. Cette observation pourrait traduire un détachement progressif du bras espaceur au cours de la synthèse. Pour vérifier notre hypothèse, nous avons testé le méthyl ester de l'acide 3-mercaptopropionique, Me-MPA. De façon intéressante, nous avons observé une augmentation du rendement d'un facteur ~ 2 , que ce soit pour le brut réactionnel ou pour la fraction purifiée. Néanmoins, la qualité du brut réactionnel reste identique et l'impureté majoritaire correspond toujours à la délétion du premier fragment.

Pour renforcer les conditions d'activation, nous avons essayé d'augmenter la quantité de Me-MPA dans le milieu réactionnel. Malheureusement, le Me-MPA n'est pas soluble dans les conditions d'activation à plus de 6 % en volume. En s'appuyant sur les travaux présentés dans le chapitre V, nous nous sommes ensuite intéressés à l'utilisation de catalyseurs. Les composés disélénures **7c-d** ont permis de considérablement accélérer les cinétiques de réactions et ce, même à des jonctions difficiles. Cependant, leur synthèse est couteuse et les quantités nécessaires pour réaliser le procédé en phase solide sont considérables. C'est pourquoi, nous nous sommes intéressés au MPAA qui a également démontré un effet catalytique, certes modeste, mais qui a l'avantage d'être disponible commercialement. À noter que la production des composés **7c-d** à plus grande échelle est en cours d'étude notamment grâce à la voie de synthèse présentée dans la Figure 79. Leur utilisation pour la SEA SPPS pourra être réservée à des synthèses plus complexes que celles présentées dans ce manuscrit.

Bien que le MPAA ait été solubilisé à pH 4 pour les études réalisées dans le chapitre V, il nous a été impossible de le solubiliser dans le tampon d'activation utilisé pour le procédé SEA SPPS. Nous avons donc entrepris une série d'études de solubilité et stabilité du MPAA dans divers solvants organiques. Après une longue optimisation, la pré-solubilisation du MPAA dans la *N*-méthyl-2-pyrrolidone NMP nous a permis de mettre au point de nouvelles conditions d'activation : 100 mM MPAA, 10 % vol. NMP, 50 mM acide acétique, 10 % vol. Me-MPA, 50 mM TCEP, 6 M Gdn.HCl dans l'eau ajusté à pH 4. Il est important de noter que la présence de NMP nous a également permis de solubiliser le Me-MPA jusqu'à 10 % par volume. Par contre, des études cinétiques en phase liquide ont montré que la présence de NMP induit une baisse de la cinétique qui peut être due à la diminution de la

quantité d'eau (discuté au chapitre V section 2.1). Malgré tout, la cinétique en présence de MPAA/NMP est supérieure à celle sans catalyseur ni NMP.

Les travaux présentés au chapitre V ont permis de montrer que la présence de Gdn.HCl induisait une baisse significative des cinétiques d'échange SEA-thiol. Ici, il n'y a pas de segments peptidiques en solution dans le milieu d'activation puisqu'ils sont accrochés au support solide. Par conséquent, l'absence de Gdn.HCl dans le milieu d'activation ne devrait pas entraîner de problèmes de précipitation des segments peptidiques et permettre d'augmenter les cinétiques d'échanges. Nous avons donc essayé d'appliquer de nouvelles conditions sans Gdn.HCl pour améliorer l'étape d'activation. Malheureusement, nous avons observé la précipitation du Me-MPA que ce soit en absence de Gdn.HCl ou en présence d'une concentration réduite à 3 M de celle-ci. En s'appuyant sur de précédents travaux du laboratoire, nous avons testé un autre thiol, le 2-mercaptoéthanesulfonate de sodium MESNa. Celui-ci est soluble dans les nouvelles conditions d'activation à savoir 100 mM MPAA, 10 % vol. NMP, 50 mM acide acétique, 10 % massique MESNa, 50 mM TCEP dans l'eau ajusté à pH 4. L'utilisation de ce nouveau thiol devrait permettre d'accélérer la cinétique d'échange. L'étude comparative réalisée en phase liquide est présentée par la Figure 102. On observe un gain cinétique significatif grâce à l'utilisation du MESNa (accélération d'un facteur 3, Figure 102C). Au regard de ce résultat très satisfaisant, nous avons adopté le couple MESNa/MPAA comme nouveau thiol exogène et catalyseur pour l'étape d'activation de notre stratégie d'assemblage en phase solide.

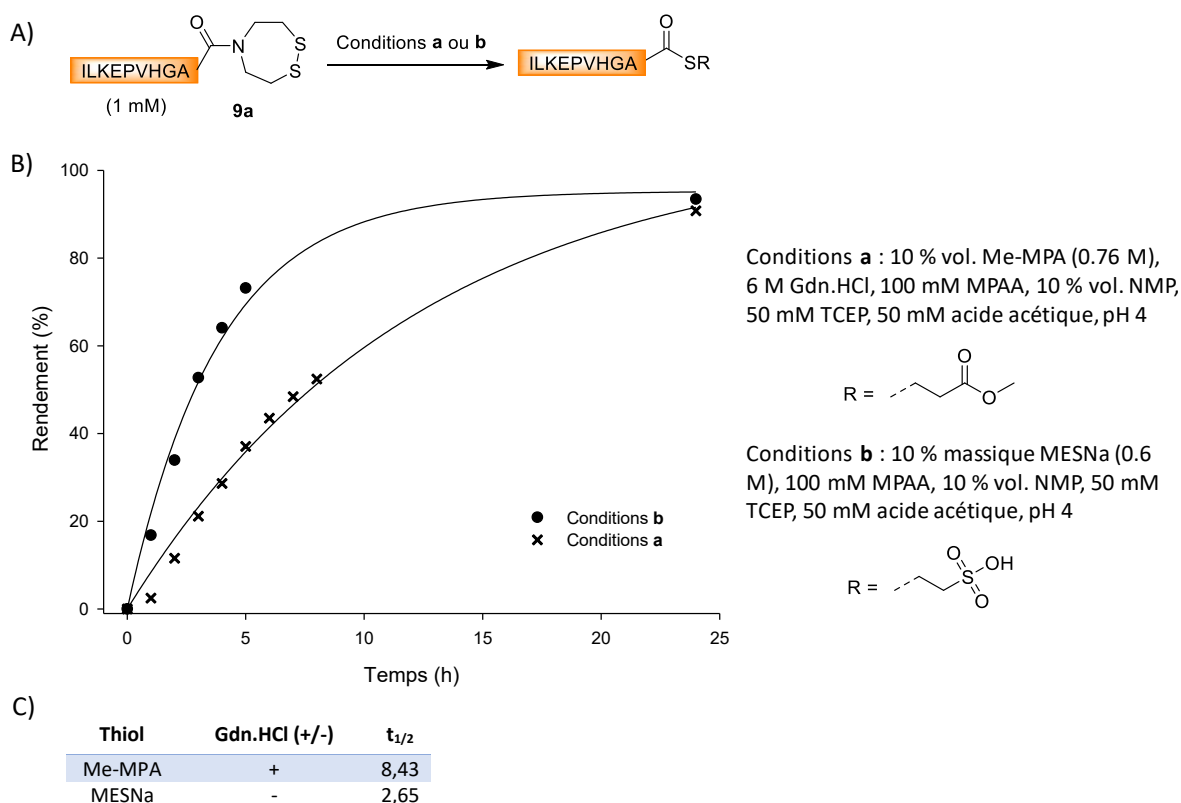


Figure 102 : A) Réaction d'activation par échange SEA-thiol. B) Suivi cinétique HPLC ($\lambda = 215$ nm) de la réaction d'échange. Les données sont corrélées par une loi d'ordre 1. C) Tableau des temps de demi-réaction pour les différentes conditions.

5.1.2 L'étape de ligation (NCL)

Puisque la présence de NMP facilite la solubilité du MPAA à des pH acides, nous avons décidé d'utiliser cette propriété pour l'étape de ligation. Pour rappel, le MPAA pour l'étape de ligation était préparé sous forme de sel MPAA-Na introduit sous forme solide dans le bloc réactifs. Nous avons remarqué que la solubilisation du MPAA-Na par l'appareil n'était pas reproductible. Dans certains cas, la solubilisation n'était pas homogène et du MPAA-Na solide persistait. Si la quantité de MPAA lors de la ligation n'est pas suffisante, cela peut entraîner l'agrégation du peptide en extension par formation de ponts disulfures dû à un manque de réducteur. Cette mauvaise solubilisation peut également induire une ligation incomplète mais également obstruer l'aiguille du prototype et empêcher celui-ci de délivrer les volumes adéquats. Une solution simple à ce problème a été de solubiliser le MPAA dans la NMP, et de l'introduire tel quel dans le bloc réactif. Nous avons montré que l'oxydation du MPAA est négligeable dans ces conditions sur de longues périodes (plusieurs jours).

Les nouvelles conditions d'élongation sont présentées dans la Figure 103A. Hormis les modifications mentionnées ci-dessus, la température de l'étape d'activation a été ajustée à 41 °C, ce facteur ayant un impact positif sur la cinétique de réaction. De plus, nous avons augmenté la quantité

de TCEP (100 mM) afin d'avoir un léger excès dans le cas où le MESNa serait en partie oxydé. Grâce à ces nouvelles conditions, nous avons considérablement augmenté le rendement de synthèse du peptide **37**. En effet, nous avons obtenu 3,67 mg de peptide **37** brut soit un rendement de 41 % pour sept étapes chimiques (élongation + coupure). L'analyse par LC-MS montre une pureté de 81,7 % (pic D, Figure 103B). L'impureté majoritaire résulte une fois encore de la délétion du premier fragment **34** (pic C ~ 12,4 %, Figure 103B). Nous observons également la délétion du second fragment **35** (pic B ~ 4,6 %, Figure 103B) et la délétion cumulée des deux premiers fragments **34** et **35** (pic A ~ 1,19 %, Figure 103B).

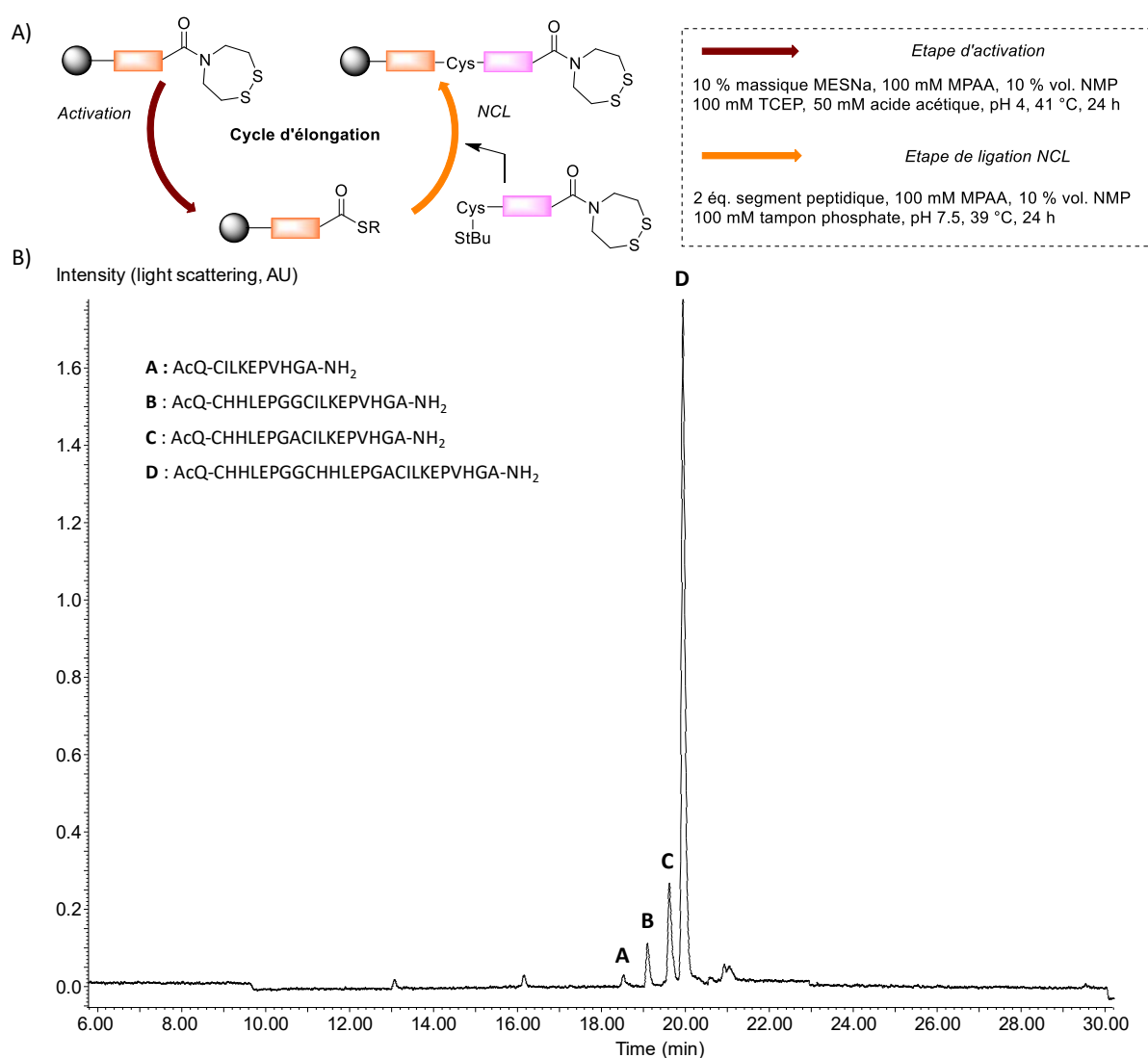


Figure 103 : A) Nouvelles conditions d'élongation. B) Analyse LC du brut après élongation/coupure. Le pic D correspond au peptide **37**.

Les modifications apportées aux deux étapes impliquées dans le processus d'élongation ont permis de diminuer les réactions secondaires et d'augmenter le rendement de synthèse. Afin de montrer l'efficacité de notre méthode, nous avons réalisé la synthèse d'un second modèle. Le peptide

41 est composé de 52 acides aminés et implique la formation de deux jonctions Ser-Cys (Figure 104A). En utilisant les nouvelles conditions d'élongation (Figure 103A), nous avons obtenu 5,2 mg de peptide **41** brut soit un rendement global de 37 % (soit 87 % par étape). L'analyse par LC-MS montre une pureté de 89,4% (Figure 104B). Les peptides **38** et **39** ayant une masse similaire, l'impureté majoritaire correspond à la fois à la délétion du premier fragment **38** et à celle du fragment **39** (~10,5 %, Figure 104B).

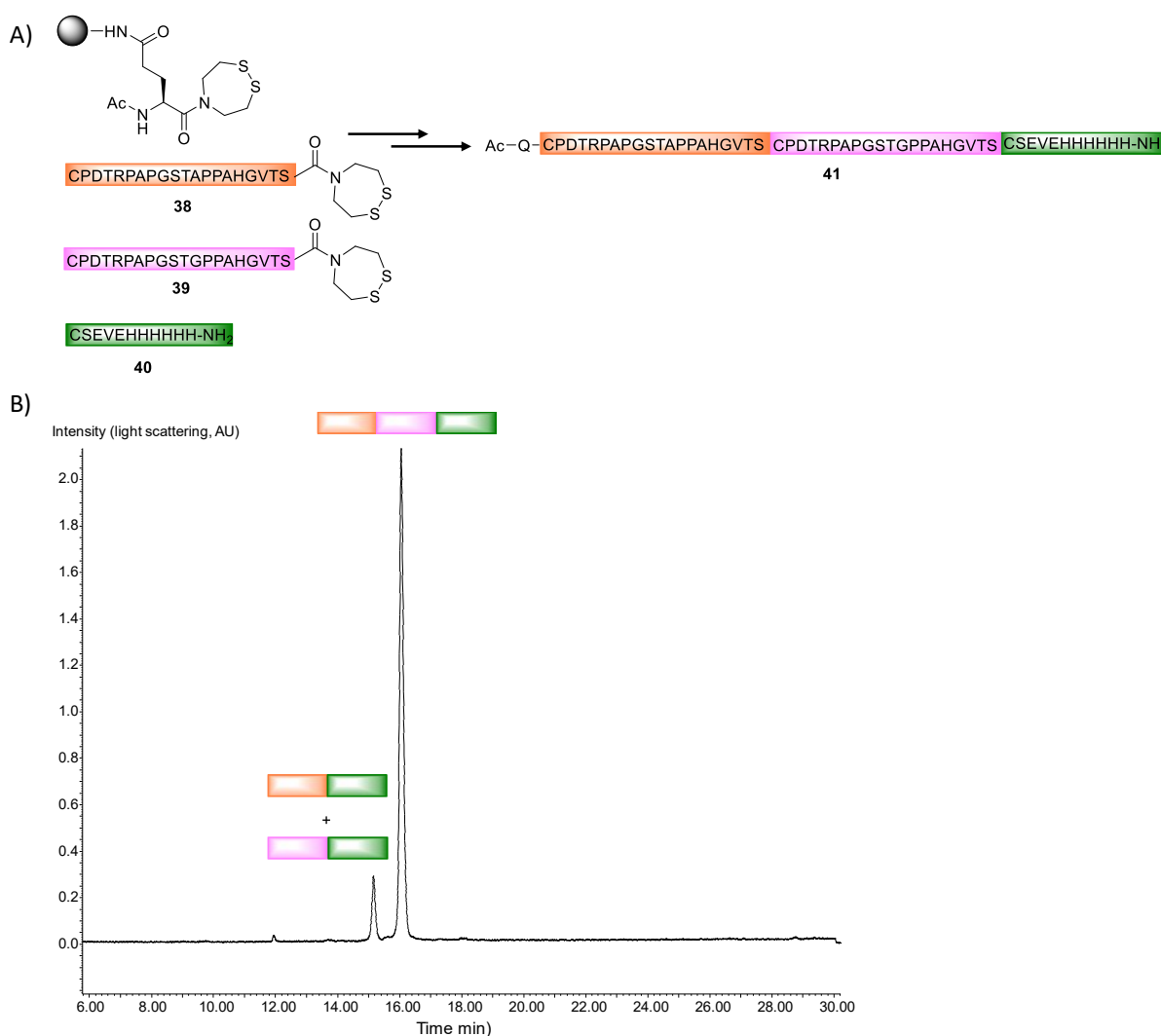


Figure 104 : A) Synthèse du peptide **41** par assemblage de 3 segments peptidiques. B) Analyse LC du brut après élongation/coupeure.

Les résultats obtenus grâce à l'optimisation des conditions d'élongation sont très encourageants et montrent le potentiel de notre procédé pour la synthèse automatisée de protéines. À ce stade, nous pensons que les modifications chimiques apportées aux étapes d'activation et ligation sont suffisantes et surtout compatibles avec une grande majorité de protéines (température, pH). Cependant, nous pensons qu'en jouant sur la nature du support solide et la charge de celui-ci nous pourrions accroître davantage les rendements de synthèse.

5.2 Support solide

Les résines utilisées pour la synthèse de peptides ou protéines sont à base de polymère. Chaque polymère possède des propriétés qui lui sont propres. En général pour la synthèse de longs peptides ou protéines, il est préférable d'utiliser des supports plus polaires à base de polyéthylène glycol (PEG) ou polyacrylamide plutôt que du polystyrène. En outre, les conditions d'élongation que nous utilisons nécessitent une résine qui soit compatible avec les milieux aqueux. Nous avons identifié trois types de résines qui répondent à ces critères : PEG-polystyrène (Novasyn® TGR), polyéthylène réticulé à du PEG (ChemMatrix®) ou PEG-acrylamide (PEGA).

La résine Novasyn® TGR est souvent utilisée pour la synthèse de peptides par Fmoc-SPPS. Nous l'avons essayé pour notre procédé SEA SPPS mais nous n'avons pas obtenu le peptide cible. Cette résine présente un faible taux de gonflement dans l'eau (4 mL/g) ce qui pourrait expliquer ce résultat. À l'inverse, la résine PEGA est composée de polymères hydrophiles et présente donc un excellent taux de gonflement dans l'eau (16 mL/g). La résine ChemMatrix® est composée uniquement de polyéthylène glycol et possède les mêmes propriétés de gonflement que la résine PEGA. Contrairement à cette dernière, elle se présente sous la forme d'une poudre sèche, ce qui la rend plus facile à manipuler. La nature amphiphile de cette résine en fait un excellent support pour la synthèse de peptides difficiles qui ont tendance à s'agréger.

Nous avons comparé ces deux résines en réalisant deux synthèses du peptide **37** en simultanée sur le prototype. À noter que la charge des deux résines étaient identiques après couplage du bras espaceur dans les conditions mentionnées précédemment (section 3.2). Ainsi, nous pouvons observer de façon objective l'effet de la nature du polymère utilisé sur le rendement et la qualité du brut réactionnel. Finalement, il est apparu que la résine ChemMatrix® conduit à de meilleurs résultats de synthèse (pureté et rendement supérieurs).

Une fois la nature du polymère définie, nous nous sommes intéressés à l'effet de la charge. En effet, en ajustant la charge il est possible de minimiser l'agrégation des peptides et d'assurer une plus grande accessibilité des sites réactifs. En jouant sur les proportions de Fmoc-Glu-SEA et de Boc-Glu(OtBu) lors du couplage, nous avons obtenus des résines de différentes charges. Le tableau ci-dessous présente les résultats obtenus.

Entrée	Fmoc-Glu-SEA	Boc-Glu(OtBu)	Charge
1	4/3	2/3	0,073
2	1/3	2/3	0,051
3	1/10	9/10	0,012
4	1,5/10	8,5/10	0,018

Figure 105 : Tableau des charges de résines obtenues en fonction des proportions d'acides aminés utilisés lors du couplage.

Les résines des entrées 1 à 3 ont été testées simultanément pour la synthèse du peptide modèle **37**, en utilisant les conditions d'élongation « optimisées » décrites dans la section précédente^f. Il est apparu que l'utilisation d'une résine moins chargée conduit à de meilleurs rendements. Cependant, la résine très peu chargée (entrée 3, Figure 105) présente un volume important. Ceci impliquerait d'utiliser de plus grands volumes (> 500 µL) pour les étapes d'activation et la ligation et donc d'utiliser des quantités très importantes de réactifs et notamment de segments peptidiques. Pour éviter cela, nous n'avons utilisé que 0,625 µmol de résine 3 comparé aux deux autres (2,5 µmol) en gardant la quantité de segment peptidique constante. Il y a donc 8 équivalents de peptides qui sont délivrés sur cette résine contre 2 pour les autres. La synthèse des segments peptidiques étant en général couteuse en moyen et en temps, un tel excès représente une perte considérable qui n'est pas acceptable. Pour éviter cela, nous avons mis au point une résine peu chargée nécessitant l'utilisation de 4 équivalents de fragments (entrée 4, Figure 105). Cependant, en raison de problèmes techniques liés en partie à l'appareil, nous n'avons pas été en mesure de tester cette nouvelle résine. En effet, nous sommes face à un problème de drainage des colonnes durant l'élongation. Nous menons actuellement une étude pour résoudre ce problème afin de pouvoir présenter un procédé qui puisse être utilisé en routine et avec des résultats reproductibles.

Malgré tout, les résultats obtenus jusqu'à présent démontrent un fort potentiel du procédé SEA SPPS pour faciliter l'accès à des protéines de grandes tailles. L'utilisation de la chimie SEA, nous a permis de concevoir une stratégie robuste. En effet, les peptides SEA^{off} sont stables en solution ce qui n'est pas le cas des peptides thioesters classiques soumis à l'hydrolyse. La conversion des peptides SEA en alkythioester permet de réaliser ensuite une réaction de ligation chimique. Le fait que les peptides SEA^{off} soient stables dans les conditions de NCL permet de répéter ce cycle d'élongation pour atteindre

^f Activation : 100 mM MPAA, 10 % vol. NMP, 50 mM acide acétique, 10 % massique MESNa, 50 mM TCEP, pH 4, 41 °C. Ligation : 2 éq. Segments, 100 mM MPAA, 10 % NMP, 100 mM tampon phosphate, pH 7,5, 39 °C.

la protéine cible souhaitée. De plus, l'automatisation permet un gain de temps considérable. L'utilisation d'un tel procédé devrait permettre de produire une collection de protéines en simultané.

CONCLUSION ET PERSPECTIVES

De nos jours, la plupart des médicaments présents sur le marché sont basés sur de petites molécules qui sont plus faciles à produire à l'échelle industrielle que les peptides. Cependant, ces petites molécules sont en général peu spécifiques et ne permettent pas de reproduire certaines interactions biologiques essentielles. Les peptides et protéines sont quant eux omniprésents dans les systèmes biologiques et présentent une grande spécificité. Leur capacité à interagir sélectivement avec leur cible au sein des systèmes vivants permet de réduire les problèmes de toxicité et d'atteindre une meilleure tolérance. Aujourd'hui, plus de 60 peptides médicaments (approuvés par la FDA) sont sur le marché. Il y a près de 140 médicaments à base de peptides en essais cliniques et plus de 500 peptides thérapeutiques en développement pré-clinique²³⁷. L'émergence des peptides et protéines dans le domaine pharmaceutique a stimulé le développement de méthodes permettant leur synthèse notamment par voie chimique.

Deux découvertes majeures ont révolutionné le domaine de la synthèse par voie chimique de protéine : la SPPS (Merrifield, 1963) et la NCL (Kent, 1994). Depuis son introduction, la NCL a fait l'objet de nombreuses optimisations avec notamment l'introduction de divers précurseurs de thioesters (hydrazide, Nbz, système de transfert d'acyle) et de groupements protecteurs de cystéines. Ces améliorations ont permis de développer des stratégies d'assemblages multiples et ainsi d'augmenter la taille des protéines synthétisées par voie chimique. En utilisant la base de données PCS (*Protein Chemical Synthesis Database*), nous avons réalisé une analyse statistique sur la NCL et les méthodes apparentées qui nous a permis d'identifier les limitations actuelles du domaine et d'établir des pistes de réflexion pour y répondre.

Nous avons constaté que l'une des principales limitations du domaine résulte de la difficulté à réaliser l'assemblage successif de segments peptidiques. Je me suis donc intéressée dans un premier temps au développement de procédés *one-pot* en utilisant des systèmes de transfert *N,Se*. Nous avons décrit dans le chapitre IV, l'utilisation des sélénioesters latents basés sur le groupement *bis*(2-sélényléthyl)amido SeEA. Ces composés présentent d'une part une grande réactivité et d'autre part des propriétés d'oxydo-réduction différentes de leurs analogues soufrés SEA. La combinaison des propriétés latentes des groupements SEA et SeEA a permis de réaliser des ligations *one pot* par activation successive des fonctions SEA et SeEA. La grande réactivité des sélénioesters de SeEA a rendu possible la ligation *one-pot* de trois segments en conditions cinétiquement contrôlées, ce qui nous a permis d'isoler la protéine NK1-B (180 acides aminés, 20 kDa).

Au regard de la grande réactivité des peptides SeEA, nous avons envisagé la formation *in situ* de sélénioesters à partir de segments peptidiques SEA grâce à la conception de nouveaux catalyseurs à base de sélénium. La synthèse et l'utilisation de ces nouveaux catalyseurs pour accélérer la ligation

SEA ainsi que la synthèse de peptides thioester à partir de peptides SEA (échange SEA-thiol) ont été présentées dans le chapitre V. En constatant que la formation de jonctions difficiles dues à l'encombrement stérique (Thr, Val, Ile) est l'une des limitations des méthodes de ligation, nous avons choisi de démontrer le grand potentiel des nouveaux catalyseurs disélénures sur ce type de jonctions. Le fort potentiel de ces catalyseurs a été démontré par la synthèse de la granulysine 15 kDa impliquant la formation des jonctions Cys-Thr et Cys-Val. Cette protéine est un bon exemple de la nécessité de développer des méthodes chimiques de synthèse car sa production par les méthodes recombinantes est très complexe et ne conduit pas à l'heure actuelle à une protéine native du fait de la présence de tags. La granulysine présente des fonctions biologiques très intéressantes. L'une des perspectives de ce projet est de réaliser le repliement de cette protéine et de réaliser des études structurales et fonctionnelles pour évaluer son potentiel thérapeutique.

Finalement, un autre moyen de faciliter l'assemblage consécutif de segments peptidiques est d'utiliser des procédés en phase solide. Nous avons décrit dans le chapitre VI une stratégie d'assemblage séquentielle de segments peptidiques en phase solide basée les propriétés du groupement SEA. Ce processus tire parti des avantages de la phase solide qui permet l'élimination des réactifs en excès par simple lavage/filtration, ce qui permet de ne pas isoler les intermédiaires et donc d'augmenter les rendements de synthèse. L'automatisation du processus induit un gain de temps considérable en accord avec les problématiques économiques actuelles. En effet, les organismes de financements et les sociétés privées en particulier souhaitent identifier rapidement le potentiel thérapeutique de leur cible (petite molécule ou peptide). Le développement de méthodes de synthèse permettant d'accéder rapidement à un ensemble de protéines même en faible quantité est donc primordial. Le procédé SEA SPPS qui a été développé durant ma thèse permet de produire en parallèle plusieurs protéines de façon simple. Une fois synthétisée ces protéines peuvent être testées soit directement sur support solide ou après décrochage. Cette démarche devrait permettre de réaliser rapidement les premiers essais biologiques qui détermineront s'il est intéressant de débloquent des financements plus conséquents pour poursuivre le projet. Les méthodes recombinantes, permettant de produire de plus grandes quantités à plus faible coût, peuvent ensuite être mises au point pour accéder à la protéine cible. À l'heure actuelle, l'optimisation du prototype est toujours en cours au laboratoire. Nous souhaiterions obtenir à terme un processus simple d'utilisation, conduisant à des résultats reproductibles et permettant la synthèse de toutes sortes de protéines. Pour cela, nous devons encore résoudre les problématiques techniques et développer un bras espaceur dit universel, ne laissant aucune trace sur la protéine finale. En parallèle, la conception d'une nouvelle méthode de synthèse des catalyseurs disélénures (chapitre V) conduisant à une montée en échelle et à l'isolement de quantités plus importantes devrait permettre leur utilisation pour le processus en phase solide.

RÉFÉRENCES BIBLIOGRAPHIQUES

- (1) Mulder, G. J. In *Bulletin des Sciences Physiques et Naturelles en Néerlande*; 1838; pp 104–119.
- (2) du Vigneaud, V.; Ressler, C.; Swan, C. J. M.; Roberts, C. W.; Katsoyannis, P. G.; Gordon, S. J. *Am. Chem. Soc.* **1953**, *75* (19), 4879–4880.
- (3) Sanger, F.; Thompson, E. O. P. *Biochemistry (Mosc.)* **1953**, *53*, 353–366.
- (4) Gutte, B.; Merrifield, R. B. *J. Am. Chem. Soc.* **1969**, *91* (2), 501–502.
- (5) Naughton, K. L.; Phan, L.; Leung, E. M.; Kautz, R.; Lin, Q.; Van Dyke, Y.; Marmioli, B.; Sartori, B.; Arvai, A.; Li, S.; Pique, M. E.; Naeim, M.; Kerr, J. P.; Aquino, M. J.; Roberts, V. A.; Getzoff, E. D.; Zhu, C.; Bernstorff, S.; Gorodetsky, A. A. *Adv. Mater.* **2016**, *28* (38), 8405–8412.
- (6) Pellois, J.; Muir, T. *Curr. Opin. Chem. Biol.* **2006**, *10* (5), 487–491.
- (7) Dang, B.; Kubota, T.; Mandal, K.; Bezanilla, F.; Kent, S. B. H. *J. Am. Chem. Soc.* **2013**, *135* (32), 11911–11919.
- (8) Dong, S.; Shang, S.; Li, J.; Tan, Z.; Dean, T.; Maeda, A.; Gardella, T. J.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2012**, *134* (36), 15122–15129.
- (9) Kwon, B.; Tietze, D.; White, P. B.; Liao, S. Y.; Hong, M. *Protein Sci.* **2015**, *24* (7), 1087–1099.
- (10) Dhayalan, B.; Fitzpatrick, A.; Mandal, K.; Whittaker, J.; Weiss, M. A.; Tokmakoff, A.; Kent, S. B. H. *ChemBioChem* **2016**, *17* (5), 415–420.
- (11) von Eggelkraut-Gottanka, R.; Klose, A.; Beck-Sickinger, A. G.; Beyermann, M. *Tetrahedron Lett.* **2003**, *44* (17), 3551–3554.
- (12) Becker, C. F.; Hunter, C. L.; Seidel, R. P.; Kent, S. B.; Goody, R. S.; Engelhard, M. *Chem. Biol.* **2001**, *8* (3), 243–252.
- (13) Chiang, K. P.; Jensen, M. S.; McGinty, R. K.; Muir, T. W. *ChemBioChem* **2009**, *10* (13), 2182–2187.
- (14) Shimko, J. C.; North, J. A.; Bruns, A. N.; Poirier, M. G.; Ottesen, J. J. *J. Mol. Biol.* **2011**, *408* (2), 187–204.
- (15) El Oualid, F.; Merckx, R.; Ekkebus, R.; Hameed, D. S.; Smit, J. J.; de Jong, A.; Hilkmann, H.; Sixma, T. K.; Ovaa, H. *Angew. Chem. Int. Ed.* **2010**, *49* (52), 10149–10153.
- (16) Seenaiah, M.; Jbara, M.; Mali, S. M.; Brik, A. *Angew. Chem. Int. Ed.* **2015**, *54* (42), 12374–12378.
- (17) Boll, E.; Drobecq, H.; Ollivier, N.; Blanpain, A.; Raibaut, L.; Desmet, R.; Vicogne, J.; Melnyk, O. *Nat. Protoc.* **2015**, *10* (2), 269–292.
- (18) Drobecq, H.; Boll, E.; Sénéchal, M.; Desmet, R.; Saliou, J.-M.; Lacapère, J.-J.; Mougel, A.; Vicogne, J.; Melnyk, O. *Bioconjugate. Chem.* **2016**, *27* (6), 1540–1546.
- (19) Merrifield, R. B. *Science* **1965**, *150* (3693), 178–185.
- (20) Chiva, C.; Barthe, P.; Codina, A.; Gairí, M.; Molina, F.; Granier, C.; Pugnière, M.; Inui, T.; Nishio, H.; Nishiuchi, Y.; Kimura, T.; Sakakibara, S.; Albericio, F.; Giralt, E. *J. Am. Chem. Soc.* **2003**, *125* (6), 1508–1517.
- (21) Nishiuchi, Y.; Inui, T.; Nishio, H.; Bódi, J.; Kimura, T.; Tsuji, F. I.; Sakakibara, S. *Proc. Natl. Acad. Sci.* **1998**, *95* (23), 13549–13554.
- (22) Schnolzer, M.; Kent, S. B. H. *Science* **1992**, *256* (5054), 221–225.
- (23) Tam, J. P.; Xu, J.; Eom, K. D. *Pept. Sci.* **2001**, *60* (3), 194–205.
- (24) Thapa, P.; Zhang, R.-Y.; Menon, V.; Bingham, J.-P. *Molecules* **2014**, *19* (9), 14461–14483.
- (25) Cabezas, E.; Wang, M.; Parren, P. W. H. I.; Stanfield, R. L.; Satterthwait, A. C. *Biochemistry (Mosc.)* **2000**, *39* (47), 14377–14391.
- (26) Shao, J.; Tam, J. P. *J. Am. Chem. Soc.* **1995**, *117* (14), 3893–3899.
- (27) Dirksen, A.; Hackeng, T. M.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2006**, *45* (45), 7581–7584.
- (28) Valverde, I. E.; Lecaille, F.; Lalmanach, G.; Aucagne, V.; Delmas, A. F. *Angew. Chem. Int. Ed.* **2012**, *51* (3), 718–722.
- (29) Wieland, T.; Bokelmann, E.; Bauer, L.; Lang, H. U.; Lau, H. *Eur. J. Org. Chem.* **1953**, *583* (1), 129–149.
- (30) Brenner, M.; Zimmermann, J. P.; Wehrmüller, J.; Quitt, P.; Photaki, I. *Experientia* **1955**, *11*, 397–399.
- (31) Kemp, D. S. *Biopolymers* **1981**, *20* (9), 1793–1804.
- (32) Kemp, D. S.; Carey, R. I. *J. Org. Chem.* **1993**, *58* (8), 2216–2222.

- (33) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 773–773.
- (34) Saxon, E.; Armstrong, J. I.; Bertozzi, C. R. *Org. Lett.* **2000**, *2* (14), 2141–2143.
- (35) Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. *Org. Lett.* **2000**, *2* (13), 1939–1941.
- (36) Staudinger, H.; Meyer, J. *Helv. Chim. Acta* **1919**, *2* (1), 635–646.
- (37) Bode, J. W.; Fox, R. M.; Baucom, K. D. *Angew. Chem. Int. Ed.* **2006**, *45* (8), 1248–1252.
- (38) Pusterla, I.; Bode, J. W. *Angew. Chem. Int. Ed.* **2012**, *51* (2), 513–516.
- (39) Pattabiraman, V. R.; Ogunkoya, A. O.; Bode, J. W. *Angew. Chem. Int. Ed.* **2012**, *51* (21), 5114–5118.
- (40) Pusterla, I.; Bode, J. W. *Nat. Chem.* **2015**, *7* (8), 668–672.
- (41) Wucherpennig, T. G.; Pattabiraman, V. R.; Limberg, F. R. P.; Ruiz-Rodríguez, J.; Bode, J. W. *Angew. Chem. Int. Ed.* **2014**, *53* (45), 12248–12252.
- (42) Ollivier, N.; Dheur, J.; Mhida, R.; Blanpain, A.; Melnyk, O. *Org. Lett.* **2010**, *12* (22), 5238–5241.
- (43) Hou, W.; Zhang, X.; Li, F.; Liu, C.-F. *Org. Lett.* **2011**, *13* (3), 386–389.
- (44) Desmet, R.; Pauzuolis, M.; Boll, E.; Drobecq, H.; Raibaut, L.; Melnyk, O. *Org. Lett.* **2015**, *17* (13), 3354–3357.
- (45) Tsuda, S.; Shigenaga, A.; Bando, K.; Otaka, A. *Org. Lett.* **2009**, *11* (4), 823–826.
- (46) Sato, K.; Shigenaga, A.; Tsuji, K.; Tsuda, S.; Sumikawa, Y.; Sakamoto, K.; Otaka, A. *ChemBioChem* **2011**, *12* (12), 1840–1844.
- (47) Veber, D.; Milkowski, J.; Varga, S.; Denkwalter, R.; Hirschmann, R. *J. Am. Chem. Soc.* **1972**, *94* (15), 5456–5461.
- (48) Bang, D.; Chopra, N.; Kent, S. B. H. *J. Am. Chem. Soc.* **2004**, *126* (5), 1377–1383.
- (49) Tesser, G. I.; Balvert-Geers, I. C. *Chem. Biol. Drug Des.* **1975**, *7* (4), 295–305.
- (50) Camarero, J. A.; Cotton, G. J.; Adeva, A.; Muir, T. W. *J. Pept. Res.* **1998**.
- (51) Wintermann, F.; Engelbrecht, S. *Angew. Chem. Int. Ed.* **2013**, *52* (4), 1309–1313.
- (52) Bang, D.; Kent, S. B. H. *Angew. Chem. Int. Ed.* **2004**, *43* (19), 2534–2538.
- (53) Isidro-Llobet, A.; Alvarez, M.; Albericio, F. *Chem. Rev.* **2009**, *109*, 2455–2504.
- (54) Mandal, K.; Kent, S. B. H. *Angew. Chem. Int. Ed.* **2011**, *50* (35), 8029–8033.
- (55) Boerema, D. J.; Tereshko, V. A.; Kent, S. B. H. *Biopolymers* **2008**, *90* (3), 278–286.
- (56) Ueda, S.; Fujita, M.; Tamamura, H.; Fujii, N.; Otaka, A. *ChemBioChem* **2005**, *6* (11), 1983–1986.
- (57) Aihara, K.; Yamaoka, K.; Naruse, N.; Inokuma, T.; Shigenaga, A.; Otaka, A. *Org. Lett.* **2016**, *18* (3), 596–599.
- (58) Boll, E.; Ebran, J.-P.; Drobecq, H.; El-Mahdi, O.; Raibaut, L.; Ollivier, N.; Melnyk, O. *Org. Lett.* **2015**, *17* (1), 130–133.
- (59) Huang, Y.-C.; Chen, C.-C.; Gao, S.; Wang, Y.-H.; Xiao, H.; Wang, F.; Tian, C.-L.; Li, Y.-M. *Chem. - Eur. J.* **2016**, *22* (22), 7623–7628.
- (60) Tang, S.; Si, Y.-Y.; Wang, Z.-P.; Mei, K.-R.; Chen, X.; Cheng, J.-Y.; Zheng, J.-S.; Liu, L. *Angew. Chem.* **2015**, *127* (19), 5805–5809.
- (61) Fang, G.-M.; Li, Y.-M.; Shen, F.; Huang, Y.-C.; Li, J.-B.; Lin, Y.; Cui, H.-K.; Liu, L. *Angew. Chem. Int. Ed.* **2011**, *50* (33), 7645–7649.
- (62) Mong, S. K.; Vinogradov, A. A.; Simon, M. D.; Pentelute, B. L. *ChemBioChem* **2014**, *15* (5), 721–733.
- (63) Chang, J. K.; Shimizu, M.; Wang, S.-S. *J. Org. Chem.* **1976**, *41* (20), 3255–3258.
- (64) Camarero, J. A.; Hackel, B. J.; de Yoreo, J. J.; Mitchell, A. R. *J. Org. Chem.* **2004**, *69* (12), 4145–4151.
- (65) Fang, G.-M.; Wang, J.-X.; Liu, L. *Angew. Chem. Int. Ed.* **2012**, *51* (41), 10347–10350.
- (66) Warren, J. D.; Miller, J. S.; Keding, S. J.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2004**, *126* (21), 6576–6578.
- (67) Botti, P.; Villain, M.; Manganiello, S.; Gaertner, H. *Org. Lett.* **2004**, *6* (26), 4861–4864.
- (68) Zheng, J.-S.; Cui, H.-K.; Fang, G.-M.; Xi, W.-X.; Liu, L. *ChemBioChem* **2010**, *11* (4), 511–515.
- (69) Mende, F.; Seitz, O. *Angew. Chem. Int. Ed.* **2011**, *50* (6), 1232–1240.
- (70) Melnyk, O.; Agouridas, V. *Curr. Opin. Chem. Biol.* **2014**, *22*, 137–145.

- (71) Hojo, H.; Onuma, Y.; Akimoto, Y.; Nakahara, Y.; Nakahara, Y. *Tetrahedron Lett.* **2007**, *48* (1), 25–28.
- (72) Kawakami, T.; Aimoto, S. *Tetrahedron Lett.* **2007**, *48* (11), 1903–1905.
- (73) Kawakami, T.; Aimoto, S. *Tetrahedron* **2009**, *65* (19), 3871–3877.
- (74) Burlina, F.; Papageorgiou, G.; Morris, C.; White, P. D.; Offer, J. *Chem. Sci.* **2014**, *5* (2), 766–770.
- (75) Terrier, V. P.; Adihou, H.; Arnould, M.; Delmas, A. F.; Aucagne, V. *Chem. Sci.* **2016**, *7* (1), 339–345.
- (76) Adams, A. L.; Macmillan, D. J. *Pept. Sci.* **2013**, *19* (2), 65–73.
- (77) Raibaut, L.; Cargoët, M.; Ollivier, N.; Chang, Y. M.; Drobecq, H.; Boll, E.; Desmet, R.; Monbaliu, J.-C. M.; Melnyk, O. *Chem. Sci.* **2016**, *7* (4), 2657–2665.
- (78) Wang, C.; Guo, Q.-X.; Fu, Y. *Chem. - Asian J.* **2011**, *6* (5), 1241–1251.
- (79) Bang, D.; Pentelute, B. L.; Kent, S. B. H. *Angew. Chem. Int. Ed.* **2006**, *45* (24), 3985–3988.
- (80) Johnson, E. C. B.; Kent, S. B. H. *J. Am. Chem. Soc.* **2006**, *128* (20), 6640–6646.
- (81) Dawson, P. E.; Churchill, M. J.; Ghadiri, M. R.; Kent, S. B. H. *J. Am. Chem. Soc.* **1997**, *119* (19), 4325–4329.
- (82) Mandal, K.; Pentelute, B. L.; Bang, D.; Gates, Z. P.; Torbeev, V. Y.; Kent, S. B. H. *Angew. Chem. Int. Ed.* **2012**, *51* (6), 1481–1486.
- (83) Sato, K.; Shigenaga, A.; Kitakaze, K.; Sakamoto, K.; Tsuji, D.; Itoh, K.; Otaka, A. *Angew. Chem. Int. Ed.* **2013**, *52* (30), 7855–7859.
- (84) Ollivier, N.; Vicogne, J.; Vallin, A.; Drobecq, H.; Desmet, R.; El Mahdi, O.; Leclercq, B.; Goormachtigh, G.; Fafeur, V.; Melnyk, O. *Angew. Chem. Int. Ed.* **2012**, *51* (1), 209–213.
- (85) Raibaut, L.; Vicogne, J.; Leclercq, B.; Drobecq, H.; Desmet, R.; Melnyk, O. *Bioorg. Med. Chem.* **2013**, *21* (12), 3486–3494.
- (86) Xu, W.; Jiang, W.; Wang, J.; Yu, L.; Chen, J.; Liu, X.; Liu, L.; Zhu, T. F. *Cell Discov.* **2017**, *3*, 17008.
- (87) Hengge, A. C.; Hess, R. A. *J. Am. Chem. Soc.* **1994**, *116* (25), 11256–11263.
- (88) Robins, L. I.; Meisenheimer, K. M.; Fogle, E. J.; Chaplan, C. A.; Redman, R. L.; Vacca, J. T.; Tellier, M. R.; Collins, B. R.; Duong, D. H.; Schulz, K.; Marlier, J. F. *J. Org. Chem.* **2013**, *78* (23), 12029–12039.
- (89) Kovach, I. M.; Elrod, J. P.; Schowen, R. L. *J. Am. Chem. Soc.* **1980**, *102* (25), 7530–7534.
- (90) Castro, E. A. *Chem. Rev.* **1999**, *99* (12), 3505–3524.
- (91) Hupe, D. J.; Jencks, W. P. *J. Am. Chem. Soc.* **1977**, *99* (2), 451–464.
- (92) Sun, X.-H.; Yu, H.-Z.; Pei, S.-Q.; Dang, Z.-M. *Chin. Chem. Lett.* **2015**, *26* (10), 1259–1264.
- (93) Bracher, P. J.; Snyder, P. W.; Bohall, B. R.; Whitesides, G. M. *Orig. Life Evol. Biospheres* **2011**, *41* (5), 399–412.
- (94) van de Langemheen, H.; Brouwer, A. J.; Kemmink, J.; Kruijtzter, J. A. W.; Liskamp, R. M. J. *J. Org. Chem.* **2012**, *77* (22), 10058–10064.
- (95) Cowper, B.; Sze, T. M.; Premdjee, B.; Bongat White, A. F.; Hacking, A.; Macmillan, D. *Chem. Commun.* **2015**, *51* (15), 3208–3210.
- (96) Campopiano, O.; Minassian, F. *E-EROS Encycl. Reag. Org. Synth.* **1996**.
- (97) Jencks, W. P.; Regenstein, J. Lundbad, R. L., Macdonald, F. M.: New York, 2010; pp 595–635.
- (98) Dalby, K. N.; Jencks, W. P. *J. Chem. Soc. Perkin Trans. 2* **1997**, No. 8, 1555–1564.
- (99) Jencks, W. P.; Salvesen, K. *J. Am. Chem. Soc.* **1971**, *93* (18), 4433–4436.
- (100) De Maria, P.; Fini, A.; Hall, F. M. *J. Chem. Soc. Perkin Trans* **1973**, 1969–1971.
- (101) Ayres, D. C.; Hellier, D. G. *Dictionary of environmentally important chemicals*, Blackie Academic & Professional.; London, 1997.
- (102) Evans, T. C.; Benner, J.; Xu, M.-Q. *Protein Sci.* *7*, 2256–2264.
- (103) DeCollo, T. V.; Lees, W. J. *J. Org. Chem.* **2001**, *66* (12), 4244–4249.
- (104) Lu, W.; Qasim, M. A.; Kent, S. B. H. *J. Am. Chem. Soc.* **1996**, *118*, 8518–8523.
- (105) Lees, W. J.; Whitesides, G. M. *J. Org. Chem.* **1993**, *58* (3), 642–647.
- (106) Lahiri, S.; Brehms, M.; Olschewski, D.; Becker, C. F. W. *Angew. Chem. Int. Ed.* **2011**, *50* (17), 3988–3992.
- (107) Gentle, I. E.; De Souza, D. P.; Baca, M. *Bioconjug. Chem.* **2004**, *15* (3), 658–663.

- (108) Pool, C. T.; Boyd, J. G.; Tam, J. P. *J. Pept. Res.* **2004**, *63* (3), 223–234.
- (109) Thompson, R. E.; Liu, X.; Alonso-García, N.; Pereira, P. J. B.; Jolliffe, K. A.; Payne, R. J. *J. Am. Chem. Soc.* **2014**, *136* (23), 8161–8164.
- (110) Yang, R.; Hou, W.; Zhang, X.; Liu, C.-F. *Org. Lett.* **2012**, *14* (1), 374–377.
- (111) Siman, P.; Blatt, O.; Moyal, T.; Danieli, T.; Lebendiker, M.; Lashuel, H. A.; Friedler, A.; Brik, A. *ChemBioChem* **2011**, *12* (7), 1097–1104.
- (112) Gastaldi, S.; Routaboul, L. *E-EROS Encycl. Reag. Org. Synth.* 1–2.
- (113) Gregory, M. J.; Bruce, T. C. *J. Am. Chem. Soc.* **1967**, *89* (9), 2121–2127.
- (114) Wieland, T.; Lambert, R.; Lang, H. U.; Schramm, G. *Eur. J. Org. Chem.* **1955**, *597* (3), 181–195.
- (115) Wieland, T.; Hornig, H. *Justus Liebigs Ann. Chem.* **1956**, *600* (1), 12–22.
- (116) Stirling, C. J. M. *J. Chem. Soc.* **1958**, 4524–4530.
- (117) Trudelle, Y.; Caille, A. *Chem. Biol. Drug Des.* **1977**, *10* (4), 291–298.
- (118) Barnett, R. E.; Jencks, W. P. *J. Am. Chem. Soc.* **1969**, *91* (9), 2358–2369.
- (119) Cuccovia, I. M.; Schroeter, E. H.; De Baptista, R. C.; Chaimovich, H. *J. Org. Chem.* **1977**, *42* (21), 3400–3403.
- (120) Martin, R. B.; Hedrick, R. I.; Parcell, A. *J. Org. Chem.* **1964**, *29* (11), 3197–3206.
- (121) Zangi, R.; Zhou, R.; Berne, B. J. *J. Am. Chem. Soc.* **2009**, *131* (4), 1535–1541.
- (122) Das, A.; Mukhopadhyay, C. *J. Phys. Chem. B* **2009**, *113* (38), 12816–12824.
- (123) Stark, G. R. *Biochemistry (Mosc.)* **1965**, *4* (6), 1030–1036.
- (124) Dirnhuber, P.; Schütz, F. *Biochem. J.* **1948**, *42* (4), 628.
- (125) Hofmeister, F. *Arch Exp Pathol Pharmacol* **1888**, *24*, 247–260.
- (126) Dooley, K. H.; Castellino, F. J. *Biochemistry (Mosc.)* **1972**, *11* (10), 1870–1874.
- (127) Arakawa, T.; Timasheff, S. N. *Biochemistry (Mosc.)* **1984**, *23* (25), 5924–5929.
- (128) Dempsey, C. E.; Mason, P. E.; Brady, J. W.; Neilson, G. W. *J. Am. Chem. Soc.* **2007**, *129* (51), 15895–15902.
- (129) Graziano, G. *Phys. Chem. Chem. Phys.* **2011**, *13* (25), 12008.
- (130) Zheng, W.; Borgia, A.; Buholzer, K.; Grishaev, A.; Schuler, B.; Best, R. B. *J. Am. Chem. Soc.* **2016**, *138* (36), 11702–11713.
- (131) Mason, P. E.; Dempsey, C. E.; Neilson, G. W.; Kline, S. R.; Brady, J. W. *J. Am. Chem. Soc.* **2009**, *131* (46), 16689–16696.
- (132) O'Brien, E. P.; Dima, R. I.; Brooks, B.; Thirumalai, D. *J. Am. Chem. Soc.* **2007**, *129* (23), 7346–7353.
- (133) Shao, Q.; Fan, Y.; Yang, L.; Gao, Y. Q. *J. Chem. Theory Comput.* **2012**, *8* (11), 4364–4373.
- (134) Greene, R. F.; Pace, C. N. *J. Biol. Chem.* **1974**, *249* (17), 5388–5393.
- (135) Garcia-Mira, M. M.; Sanchez-Ruiz, J. M. *Biophys. J.* **2001**, *81* (6), 3489–3502.
- (136) Noller, C. R.; Rockwell, W. C. *J. Am. Chem. Soc.* **1938**, *60* (9), 2076–2077.
- (137) Baron, C.; Thompson, T. E. *Biochim. Biophys. Acta BBA-Biomembr.* **1975**, *382* (3), 276–285.
- (138) Womack, M. D.; Kendall, D. A.; MacDonald, R. C. *Biochim. Biophys. Acta BBA-Biomembr.* **1983**, *733* (2), 210–215.
- (139) Bianchi, E.; Ingenito, R.; Simon, R. J.; Pessi, A. *J. Am. Chem. Soc.* **1999**, *121* (33), 7698–7699.
- (140) Clayton, D.; Shapovalov, G.; Maurer, J. A.; Dougherty, D. A.; Lester, H. A.; Kochendoerfer, G. *G. Proc. Natl. Acad. Sci.* **2004**, *101* (14), 4764–4769.
- (141) Valiyaveetil, F. I.; MacKinnon, R.; Muir, T. W. *J. Am. Chem. Soc.* **2002**, *124* (31), 9113–9120.
- (142) Dittmann, M.; Sauermann, J.; Seidel, R.; Zimmermann, W.; Engelhard, M. *J. Pept. Sci.* **2010**, *16* (10), 558–562.
- (143) Dittmann, M.; Sadek, M.; Seidel, R.; Engelhard, M. *J. Pept. Sci.* **2012**, *18* (5), 312–316.
- (144) Seebach, D.; Thaler, A.; Beck, A. K. *Helv. Chim. Acta* **1989**, *72* (5), 857–867.
- (145) Sohma, Y.; Kitamura, H.; Kawashima, H.; Hojo, H.; Yamashita, M.; Akaji, K.; Kiso, Y. *Tetrahedron Lett.* **2011**, *52* (52), 7146–7148.
- (146) Gordon, M.; Miller, J. G.; Day, A. R. *J. Am. Chem. Soc.* **1949**, *71* (4), 1245–1250.
- (147) Herbst, E.; Shabat, D. *Org. Biomol. Chem.* **2016**, *14* (15), 3715–3728.
- (148) Mukerjee, P.; Ostrow, J. D. *Tetrahedron Lett.* **1998**, *39* (5–6), 423–426.

- (149) Kochendoerfer, G. G.; Salom, D.; Lear, J. D.; Wilk-Orescan, R.; Kent, S. B. H.; DeGrado, W. F. *Biochemistry (Mosc.)* **1999**, *38* (37), 11905–11913.
- (150) Cavallini, D.; De Marco, C.; Dupré, S. *Arch. Biochem. Biophys.* **1968**, *124*, 18–26.
- (151) Bagiyan, G. A.; Koroleva, I. K.; Soroka, N. V.; Ufimtsev, A. V. *Russ. Chem. Bull.* **2003**, *52* (5), 1135–1141.
- (152) Wallace, T. J.; Schriesheim, A.; Bartok, W. *J. Org. Chem.* **1963**, *28* (5), 1311–1314.
- (153) Wallace, T. J.; Schriesheim, A.; Hurwitz, H.; Glaser, M. B. *Ind. Eng. Chem. Process Des. Dev.* **1964**, *3* (3), 237–241.
- (154) Hanaki, A.; Kamide, H. *Chem. Pharm. Bull.* **1975**, *23*, 1671–1676.
- (155) Pecci, L.; Montefoschi, G.; Musci, G.; Cavallini, D. *Amino Acids* **1997**, *13*, 355–367.
- (156) Halliwell, B.; Gutteridge, J. M. C. *Methods Enzymol.* **1990**, *186*, 1–85.
- (157) Gutteridge, J. M. *FEBS Lett.* **1987**, *214* (2), 362–364.
- (158) Stevens, R.; Stevens, L.; Price, N. C. *Biochem. Educ.* **1983**, *11*, 70.
- (159) Rao, C.; Tam, J. P. *J. Am. Chem. Soc.* **1994**, *116* (15), 6975–6976.
- (160) Spetzler, J. C.; Tam, J. P. *Int. J. Pept. Protein Res.* **1995**, *45* (1), 78–85.
- (161) Bors, W.; Buettner, G. R. In *Vitamin C in Health and Disease*; Marcel Dekker: New York, 1997; pp 75–94.
- (162) Dery, S.; Reddy, P. S.; Dery, L.; Mousa, R.; Dardashti, R. N.; Metanis, N. *Chem. Sci.* **2015**, *6* (11), 6207–6212.
- (163) Giustarini, D.; Dalle-Donne, I.; Colombo, R.; Milzani, A.; Rossi, R. *Nitric Oxide* **2008**, *19* (3), 252–258.
- (164) Yost, J. M.; Knight, J. D.; Coltart, D. M. *E-EROS Encycl. Reag. Org. Synth.* **2001**.
- (165) Burns, J. A.; Butler, J. C.; Moran, J.; Whitesides, G. M. *J. Org. Chem.* **1991**, *56* (8), 2648–2650.
- (166) Podlaha, J.; Podlahova, J. *Collect. Czechoslov. Chem. Commun.* **1973**, *38* (6), 1730–1736.
- (167) Levison, M. E.; Josephson, A. S.; Kirschenbaum, D. M. *Experientia* **1969**, *25*, 126–127.
- (168) Gray, W. R. *Protein Sci.* **1993**, *2* (10), 1732–1748.
- (169) Han, J. C.; Han, G. Y. *Anal. Biochem.* **1994**, *220*, 5–10.
- (170) Getz, E. B.; Xiao, M.; Chakrabarty, T.; Cooke, R.; Selvin, P. R. *Anal. Biochem.* **1999**, *273* (1), 73–80.
- (171) Hamm, M. L.; Nikolic, D.; van Breemen, R. B.; Piccirilli, J. A. *J. Am. Chem. Soc.* **2000**, *122* (49), 12069–12078.
- (172) Metanis, N.; Keinan, E.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2010**, *49* (39), 7049–7053.
- (173) Gorlatov, S. N.; Stadtman, T. C. *Proc. Natl. Acad. Sci.* **1998**, *95* (15), 8520–8525.
- (174) Muttenthaler, M.; Nevin, S. T.; Grishin, A. A.; Ngo, S. T.; Choy, P. T.; Daly, N. L.; Hu, S.-H.; Armishaw, C. J.; Wang, C.-I. A.; Lewis, R. J.; Martin, J. L.; Noakes, P. G.; Craik, D. J.; Adams, D. J.; Alewood, P. F. *J. Am. Chem. Soc.* **2010**, *132* (10), 3514–3522.
- (175) Hoffmann, F. W.; Ess, R. J.; Simmons, T. C.; Hanzel, R. S. *J. Am. Chem. Soc.* **1956**, *78*, 6414–6414.
- (176) Walling, C.; Rabinowitz, R. *J. Am. Chem. Soc.* **1957**, *79* (19), 5326–5326.
- (177) Cross, R. J.; Millington, D. *J. Chem. Soc. Chem. Commun.* **1975**, No. 11, 455–456.
- (178) Chu, J. Y.; Marsh, D. G. *J. Org. Chem.* **1976**, *41* (19), 3204–3205.
- (179) Rauk, A.; Yu, D.; Armstrong, D. A. *J. Am. Chem. Soc.* **1998**, *120* (34), 8848–8855.
- (180) Kaur, D.; Sharma, P.; Bharatam, P. V.; Kaur, M. *Int. J. Quantum Chem.* **2008**, *108* (5), 983–991.
- (181) Nauser, T.; Dockheer, S.; Kissner, R.; Koppenol, W. H. *Biochemistry (Mosc.)* **2006**, *45* (19), 6038–6043.
- (182) Wan, Q.; Danishefsky, S. J. *Angew. Chem. Int. Ed.* **2007**, *46* (48), 9248–9252.
- (183) Rohde, H.; Schmalisch, J.; Harpaz, Z.; Diezmann, F.; Seitz, O. *ChemBioChem* **2011**, *12* (9), 1396–1400.
- (184) Kryukov, G. V. *Science* **2003**, *300* (5624), 1439–1443.
- (185) Yan, L. Z.; Dawson, P. E. *J. Am. Chem. Soc.* **2001**, *123* (4), 526–533.
- (186) Malins, L. R.; Payne, R. J. *Curr. Opin. Chem. Biol.* **2014**, *22*, 70–78.
- (187) Shang, S.; Tan, Z.; Danishefsky, S. J. *Proc. Natl. Acad. Sci.* **2011**, *108* (15), 5986–5989.

- (188) Crich, D.; Banerjee, A. *J. Am. Chem. Soc.* **2007**, *129* (33), 10064–10065.
- (189) Malins, L. R.; Cergol, K. M.; Payne, R. *J. ChemBioChem* **2013**, *14* (5), 559–563.
- (190) Zheng, J.-S.; Tang, S.; Huang, Y.-C.; Liu, L. *Acc. Chem. Res.* **2013**, *46* (11), 2475–2484.
- (191) Hojo, H.; Aimoto, S. *Bull. Chem. Soc. Jpn.* **1991**, *64* (1), 111–117.
- (192) Blanco-Canosa, J. B.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2008**, *47* (36), 6851–6855.
- (193) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci.* **1999**, *96* (18), 10068–10073.
- (194) Creech, G. S.; Paresi, C.; Li, Y.-M.; Danishefsky, S. J. *Proc. Natl. Acad. Sci.* **2014**, *111* (8), 2891–2896.
- (195) Gross, C. M.; Lelievre, D.; Woodward, C. K.; Barany, G. *J. Pept. Res.* **2005**, *65* (3), 395–410.
- (196) Villain, M.; Gaertner, H.; Botti, P. *Eur. J. Org. Chem.* **2003**, *2003* (17), 3267–3272.
- (197) Raibaut, L.; Seeberger, P.; Melnyk, O. *Org. Lett.* **2013**, *15* (21), 5516–5519.
- (198) Durek, T.; Alewood, P. F. *Angew. Chem. Int. Ed.* **2011**, *50* (50), 12042–12045.
- (199) Nakamura, T.; Shigenaga, A.; Sato, K.; Tsuda, Y.; Sakamoto, K.; Otaka, A. *Chem. Commun.* **2014**, *50* (1), 58–60.
- (200) Raibaut, L.; Ollivier, N.; Melnyk, O. *Chem. Soc. Rev.* **2012**, *41* (21), 7001.
- (201) Canne, L. E.; Botti, P.; Simon, R. J.; Chen, Y.; Dennis, E. A.; Kent, S. B. H. *J. Am. Chem. Soc.* **1999**, *121* (38), 8720–8727.
- (202) Yu, R. R.; Mahto, S. K.; Justus, K.; Alexander, M. M.; Howard, C. J.; Ottesen, J. J. *Org. Biomol. Chem.* **2016**, *14* (9), 2603–2607.
- (203) Ollivier, N.; Desmet, R.; Drobecq, H.; Blanpain, A.; Boll, E.; Leclercq, B.; Mougél, A.; Vicogne, J.; Melnyk, O. *Chem. Sci.* **2017**, *8* (8), 5362–5370.
- (204) Zitterbart, R.; Krumrey, M.; Seitz, O. *J. Pept. Sci.* **2017**, *23* (7–8), 539–548.
- (205) Mautner, H. G.; Chu, S.-H.; Gunther, W. H. H. *J. Am. Chem. Soc.* **1963**, *85*, 3458–3462.
- (206) Huber, R. E.; Criddle, R. S. *Arch. Biochem. Biophys.* **1967**, *122*, 164–173.
- (207) Reich, H. J.; Cohen, M. L. *J. Org. Chem.* **1979**, *44* (18), 3148–3151.
- (208) Fagioli, F.; Pulidori, F.; Bigli, C.; De Battisti, A. *Gazzetta Chim. Ital.* **1974**, *104*, 639.
- (209) Huang, J.; Li, C.; Nolan, S. P. *Organometallics* **1998**, *17*, 3516–3521.
- (210) Chu, S.-H.; Mautner, H. G. *J. Org. Chem.* **1966**, *31* (1), 308–312.
- (211) Ghassemian, A.; Vila-Farrés, X.; Alewood, P. F.; Durek, T. *Bioorg. Med. Chem.* **2013**, *21* (12), 3473–3478.
- (212) Ollivier, N.; Raibaut, L.; Blanpain, A.; Desmet, R.; Dheur, J.; Mhidia, R.; Boll, E.; Drobecq, H.; Pira, S. L.; Melnyk, O. *J. Pept. Sci.* **2014**, *20* (2), 92–97.
- (213) Raibaut, L.; Drobecq, H.; Melnyk, O. *Org. Lett.* **2015**, *17* (14), 3636–3639.
- (214) Ollivier, N.; Blanpain, A.; Boll, E.; Raibaut, L.; Drobecq, H.; Melnyk, O. *Org. Lett.* **2014**, *16* (15), 4032–4035.
- (215) Moroder, L. *J. Pept. Sci.* **2005**, *11* (4), 187–214.
- (216) Steinmann, D.; Nauser, T.; Koppenol, W. H. *J. Org. Chem.* **2010**, *75* (19), 6696–6699.
- (217) Beld, J.; Woycechowsky, K. J.; Hilvert, D. *Biochemistry (Mosc.)* **2007**, *46* (18), 5382–5390.
- (218) Iwaoka, M.; Takahashi, T.; Tomoda, S. *Heteroat. Chem.* **2001**, *12*, 293–299.
- (219) Simonneau, C.; Bérénice Leclercq, B. L.; Mougél, A.; Adriaenssens, E.; Paquet, C.; Raibaut, L.; Ollivier, N.; Drobecq, H.; Marcoux, J.; Cianférani, S.; Tulasne, D.; de Jonge, H.; Melnyk, O.; Vicogne, J. *Chem. Sci.* **2015**, *6* (3), 2110–2121.
- (220) Forlano, P.; Olabe, J. A.; Magallanes, F. F.; Blesa, M. A. *Can. J. Chem.* **1997**, *75*, 9–13.
- (221) Jakubke, H. D. *Z. Chem.* **1963**, *3*, 65–66.
- (222) You, L.; Feng, S.; An, R.; Wang, X.; Bai, D. *Tetrahedron Lett.* **2008**, *49* (35), 5147–5149.
- (223) Abbas, M.; Bethke, J.; Wessjohann, L. A. *Chem. Commun.* **2006**, No. 5, 541–543.
- (224) Schroll, A. L.; Hondal, R. J.; Flemer, S. *J. Pept. Sci.* **2012**, *18* (3), 155–162.
- (225) Peña, S. V.; Hanson, D. A.; Carr, B. A.; Goralski, T. J.; Krensky, A. M. *J. Immunol.* **1997**, *158* (6), 2680.
- (226) Krensky, A. M.; Clayberger, C. *Tissue Antigens* **2009**, *73* (3), 193–198.
- (227) Stenger, S. *Science* **1998**, *282* (5386), 121–125.

- (228) Ernst, W. A.; Thoma-Uszynski, S.; Teitelbaum, R.; Ko, C.; Hanson, D. A.; Clayberger, C.; Krensky, A. M.; Leippe, M.; Bloom, B. R.; Ganz, T.; Modlin, R. L. *J. Immunol.* **2000**, *165* (12), 7102.
- (229) Chung, W.-H.; Hung, S.-I.; Yang, J.-Y.; Su, S.-C.; Huang, S.-P.; Wei, C.-Y.; Chin, S.-W.; Chiou, C.-C.; Chu, S.-C.; Ho, H.-C.; Yang, C.-H.; Lu, C.-F.; Wu, J.-Y.; Liao, Y.-D.; Chen, Y.-T. *Nat. Med.* **2008**, *14* (12), 1343–1350.
- (230) Dotiwala, F.; Mulik, S.; Polidoro, R. B.; Ansara, J. A.; Burleigh, B. A.; Walch, M.; Gazzinelli, R. T.; Lieberman, J. *Nat Med* **2016**, *22* (2), 210–216.
- (231) Finn, M. W.; Clayberger, C.; Krensky, A. M. *Protein Expr. Purif.* **2011**, *75* (1), 70–74.
- (232) Wei, H.-M.; Lin, L.-C.; Wang, C.-F.; Lee, Y.-J.; Chen, Y.-T.; Liao, Y.-D. *PLoS One* **2016**, *11* (6), e0156321.
- (233) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. *Nat. Methods* **2015**, *12* (1), 7–8.
- (234) Aucagne, V.; Valverde, I. E.; Marceau, P.; Galibert, M.; Dendane, N.; Delmas, A. F. *Angew. Chem. Int. Ed.* **2012**, *51* (45), 11320–11324.
- (235) Raibaut, L.; Adihou, H.; Desmet, R.; Delmas, A. F.; Aucagne, V.; Melnyk, O. *Chem. Sci.* **2013**, *4* (10), 4061.
- (236) Canne, L.; Kent, S. B. H.; Simon, R. *Solid phase native chemical ligation of unprotected or n-terminal cysteine protected peptides in aqueous solution*; US 6326468 B1, 2001.
- (237) Fosgerau, K.; Hoffmann, T. *Drug Discov. Today* **2015**, *20* (1), 122–128.

ANNEXES

Table of Contents

ANNEX A: EXPERIMENTAL PART.....	165
General Methods.....	167
1. Reagents and solvents.....	167
2. Analyses.....	167
3. HPLC purification.....	168
4. Peptide synthesis.....	168
4.1 General protocol for automated peptide synthesis.....	168
4.2 Synthesis of SEA ^{off} peptide segments.....	168
4.3 Synthesis of MPA thioesters.....	177
4.4 Synthesis of Cys peptides (13, 25 and 29).....	179
Accelerating chemoselective peptide bond formation using bis(2-selenylethyl)amido peptide selenolester surrogates (Chapter IV).....	183
1. Synthesis of diselenide 7a and 8.....	183
2. Synthesis of SeEA ^{off} peptide 10.....	184
3. Kinetic measurements.....	186
3.1 General procedure for kinetic measurements for the ligation of 9a, 10 and 12a with 13.....	186
3.2 General procedure for kinetic measurements for the SEA ligation of 9a-b with 13 at different pH.....	186
3.3 First-order reaction.....	187
4. Synthesis of SeEA peptide 23.....	187
4.1 One-pot synthesis of MPA peptide thioester 19.....	187
4.2 One-pot synthesis of SeEA peptide 22.....	189
4.2.1 Synthesis of AcA-MPA.....	189
4.2.2 Synthesis of peptide 21.....	192
4.2.3 Preparation of SeEA ^{off} peptide 22.....	195
4.3 One-pot assembly of peptide 23.....	197
4.3.1 Step 1.....	197
4.3.2 Step 2.....	197
4.4 One-pot synthesis of NK1-B.....	199
5. Computational analysis.....	200
Catalysis of SEA-thiol exchange and SEA ligation by a water soluble alkylselenol catalysts (Chapter V).....	203
1. Synthesis of diselenide catalysts.....	203
1.1 Synthesis of N-(3-hydroxypropyl) bis(2-selenylethyl)amine 7b.....	203
1.2 Synthesis of N-(2-carboxamidoethyl) bis(2-selenylethyl)amine 7c.....	204
1.3 Synthesis of N-(2-carboxyethyl) bis(2-selenylethyl)amine 7d.....	209
1.3.1 First strategy.....	209

1.3.2	Second strategy	213
2.	Kinetic measurements.....	217
2.1	SEA-thiol exchange. General procedure illustrated with peptide 9a and MPAA as catalyst	217
2.2	SEA ligation. General procedure illustrated with peptide 9b, Cys peptide 13 and MPAA as catalyst	217
3.	Preparative synthesis of peptide thioester 12b by SEA-thiol exchange	218
4.	Preparative SEA ligation	219
4.1	Synthesis of the peptide 14b by reaction of the SEA ^{off} peptide 9b with Cys peptide 13 ..	219
4.2	Cyclisation of peptide 26. Synthesis of cyclic peptide 27	221
4.2.1	Procedure	221
4.2.2	Proteomic analysis.....	222
5.	Synthesis of 15 kDa granulysin in the N to C direction	225
5.1	Synthesis of peptide 12e by SEA-thiol exchange of 9e.....	225
5.2	Synthesis of peptide 28	227
5.3	Synthesis of 15 kDa GLNY.....	229
6.	Synthesis of 15 kDa granulysin in the C to N direction.....	229
6.1	Synthesis of peptide 30	229
6.2	Synthesis of peptide 31	231
6.3	Synthesis of peptide 33	232
6.4	Synthesis of 15 kD granulysin	235
	Automated SEA solid phase protein synthesis (Chapter VI)	237
1.	Synthesis of the linker Fmoc-Glu-SEA ^{off}	237
1.1	Synthesis of tert-butyl 5-(bis(2-tritylsulfanylethyl)amino)-4-(9H-fluoren-9-ylmethoxycarbonylamino)-5-oxo-pentanoate A.....	237
1.2	Synthesis of tert-butyl 5-(1,2,5-dithiazepan-5-yl)-4-(9H-fluoren-9-ylmethoxycarbonylamino)-5-oxo-pentanoate B	241
1.3	Synthesis of 5-(1,2,5-dithiazepan-5-yl)-4-(9H-fluoren-9-ylmethoxycarbonylamino)-5-oxo-pentanoic acid : Fmoc-Glu-SEA ^{off}	244
2.	Synthesis of the solid support	247
2.1	General procedure for the coupling of the linker on the resin using a ratio 1/3 Fmoc-Glu-SEA ^{off} and 2/3 Boc-Glu(OtBu)-OH	247
2.2	Determination of Fmoc loading by UV spectrophotometry.....	247
2.3	Deprotection and acetylation.....	248
	References.....	249
	ANNEX B: ARTICLES	251
	A statistical view of protein chemical synthesis using NCL and extended methodologies.....	253
	Accelerating chemoselective peptide bond formation using bis(2-selenylethyl)amido peptide selenoester surrogates.....	261

ANNEX A: EXPERIMENTAL PART

General Methods

1. Reagents and solvents

2-(1*H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium fluorophosphate (HBTU) and *N*-Fmoc protected amino acids were obtained from Iris Biotech GmbH. Side-chain protecting groups used for the amino acids were Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Pro-OH, Fmoc-Phe-OH, Fmoc-Ser(*t*Bu)-OH, Fmoc-Thr(*t*Bu)-OH, Fmoc-Tyr(*t*Bu)-OH, Fmoc-Val-OH, Fmoc-Cys(*St*Bu)-OH or Fmoc-Cys(Trt)-OH. Synthesis of *bis*(2-sulfanylethyl)aminotriyl polystyrene (SEA PS) resin was carried out as described elsewhere¹.

4-mercaptophenylacetic acid (MPAA), 3-mercaptopropionic acid (MPA), *tris*(2-carboxyethyl)phosphine hydrochloride (TCEP), guanidine hydrochloride were purchased from Sigma-Aldrich. All other reagents were purchased from Acros Organics or Merck and were of the purest grade available.

Peptide synthesis grade *N,N*-dimethylformamide (DMF), dichloromethane (CH₂Cl₂), diethylether (Et₂O), acetonitrile (CH₃CN), heptane, LC–MS-grade acetonitrile (CH₃CN, 0.1% TFA), LC–MS-grade water (H₂O, 0.1% TFA), *N,N*-diisopropylethylamine (DIEA), acetic anhydride (Ac₂O) were purchased from Biosolve and Fisher-Chemical. Trifluoroacetic acid (TFA) was obtained from Biosolve. Water was purified with a Milli-Q Ultra-Pure Water Purification System.

2. Analyses

The reactions were monitored by analytical LC–MS (Waters 2695 LC/ZQ 2000 quadripole) on a reversed phase column XBridge BEH300 C18 (3.5 μm, 300 Å, 4.6 × 150 mm) unless otherwise stated. The elutions were carried out at 30 °C using a linear gradient of 0–100% of eluent B in eluent A over 30 min at a flow rate of 1 mL/min (eluent A = 0.1% TFA in H₂O; eluent B = 0.1% TFA in CH₃CN/H₂O: 4/1 by vol). The column eluate was monitored by UV at 215 nm and by evaporative light scattering (ELS, Waters 2424 detector). The peptide masses were measured by on-line LC–MS: Ionization mode: ES⁺, *m/z* range 350–2040, capillary voltage 3 kV, cone voltage 30 V, extractor voltage 3 V, RF lens 0.2 V, source temperature 120 °C, desolvation temperature 350 °C. Calculated masses were based on average isotope composition.

MALDI-TOF mass spectra were recorded with a BrukerAutoflex Speed using α-cyano-4-hydroxycinnamic acid, sinapinic acid or 2,5-dihydroxybenzoic acid (DHB) as matrix. The observed *m/z* corresponded to the monoisotopic ions, unless otherwise stated.

^1H and ^{13}C NMR spectra were recorded on a Bruker Advance-300 spectrometer operating at 300 MHz and 75 MHz respectively. The spectra are reported as parts per million (ppm) down field shift using tetramethylsilane or dimethylselenide as internal references. The data are reported as chemical shift (δ), multiplicity, relative integral, coupling constant (J Hz) and assignment where possible.

The determination of optical purity of the C-terminal amino acid was done by chiral GC-MS following total acid hydrolysis in deuterated aqueous acid (C.A.T. GmbH & Co. Chromatographie und Analysetechnik KG, Heerweg 10, D-72070 Tübingen, Germany)².

3. HPLC purification

Preparative reversed phase HPLC of crude peptides were performed with a preparative HPLC Waters system using a reversed phase column XBridge BEH300 Prep C18 (5 μm , 300 \AA , 10 \times 250 mm) and appropriate linear gradient of increasing concentration of eluent B in eluent A (flow rate of 6 mL/min). Selected fractions were then combined and lyophilized.

Preparative reversed phase HPLCs of crude selenium compounds were performed with an Autopurification prep HPLC–MS Waters system using a reversed phase column XBridge ODB prep C-18 (5 μm , 300 \AA , 19 \times 100 mm) and appropriate gradient of increasing concentration of eluent B in eluent A (flow rate of 25 mL/min). The fractions containing the purified target peptide were identified on-line using MS (ZQ 2000 quadripole). Selected fractions were combined and lyophilized.

4. Peptide synthesis

4.1 General protocol for automated peptide synthesis

Peptide elongation was performed using standard Fmoc/*tert*-butyl chemistry on an automated peptide synthesizer (0.2 mmol scale). Couplings were performed using 5-fold molar excess of each Fmoc-L-amino acid, 4.5-fold molar excess of HBTU, and 10-fold molar excess of DIEA. A capping step was performed after each coupling with Ac_2O /DIEA in DMF. At the end of the synthesis, the resin was washed with CH_2Cl_2 , diethylether (3 \times 2 min) and dried in vacuo.

4.2 Synthesis of SEA^{off} peptide segments

Peptide elongation was performed on SEA PS resin (0.2 mmol, 0.16 mmol/g) using standard Fmoc/*tert*-butyl chemistry on an automated peptide synthesizer. Typical procedures for the synthesis of SEA^{off} peptide segments were described in previous papers^{1,3}. For a detailed protocol see the protocol article⁴. The characterization of the purified SEA^{off} peptide segments are shown below.

Yields for the HPLC purified SEA^{off} peptides

-**Peptide 9a**: 50 mg (35% yield, 0.1 mmol scale), MALDI-TOF calcd. for $[M+H]^+$: 1080.6, observed mass: 1080.6 (monoisotopic).

-**Peptide 9b**: 56 mg (39% yield, 0.1 mmol scale), MALDI-TOF calcd. for $[M+H]^+$: 1108.6, observed mass: 1108.6 (monoisotopic).

-**Peptide 9c**: 75 mg (10% yield, 0.1 mmol scale), MALDI-TOF calcd. for $[M+H]^+$: 1110.6, observed mass: 1110.8 (monoisotopic).

-**Peptide 9d**: 70 mg (25% yield, 0.1 mmol scale), MALDI-TOF calcd. for $[M+H]^+$: 1122.6, observed mass: 1122.7 (monoisotopic).

-**Peptide 9e**: 26 mg (8% yield, 0.05 mmol scale), MALDI-TOF calcd. for $[M+H]^+$: 5618.44, observed mass: 5617.2 (average).

-**Peptide 17**: 42 mg (10% yield, 0.1 mmol scale), MALDI-TOF calcd. for $[M+H]^+$: 3383.6, observed mass: 3383.7 (monoisotopic).

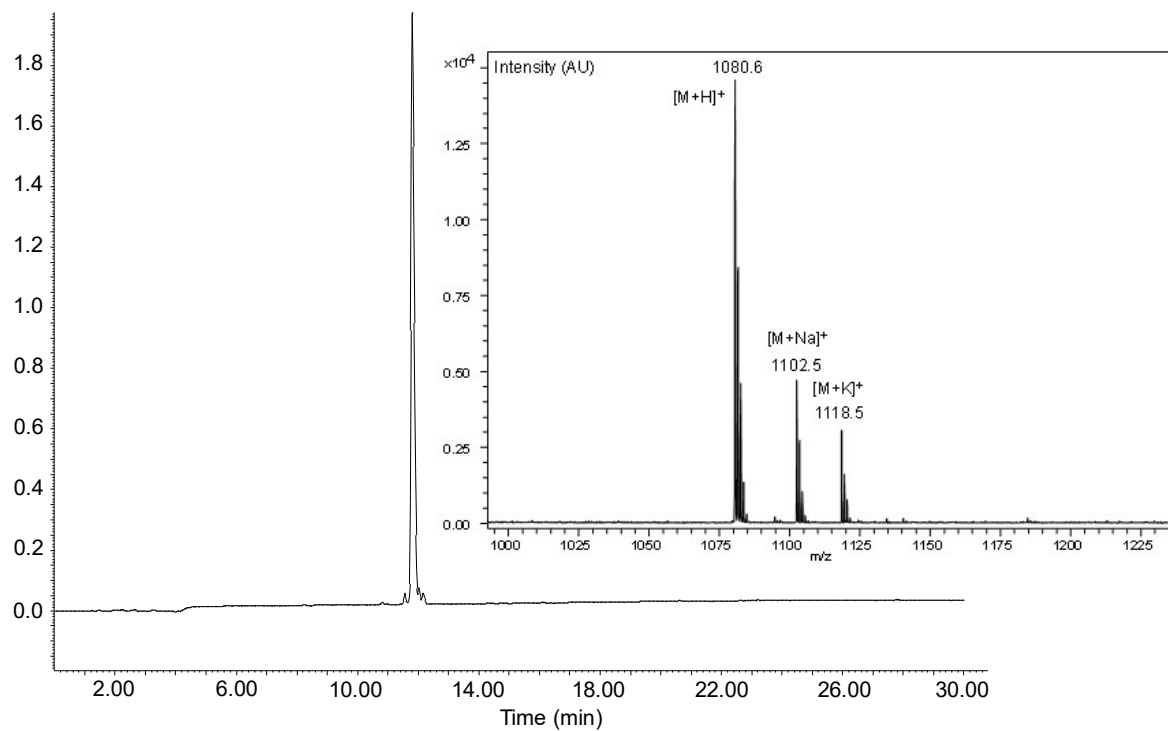
-**Peptide 20**: 48 mg (4% yield, 0.25 mmol scale), MALDI-TOF calcd. for $[M+H]^+$: 4062.8, observed mass: 4062.5 (monoisotopic).

-**Peptide 24**: 47 mg (23% yield, 50 μ mol scale), MALDI-TOF calcd. for $[M+H]^+$: 3541.6, observed mass: 3541.5 (monoisotopic).

-**Peptide 26**: 11.5 mg (4% yield, 0.05 mmol scale), MALDI-TOF calcd. for $[M+H]^+$: 4683,74 observed mass: 4683.3 (average).

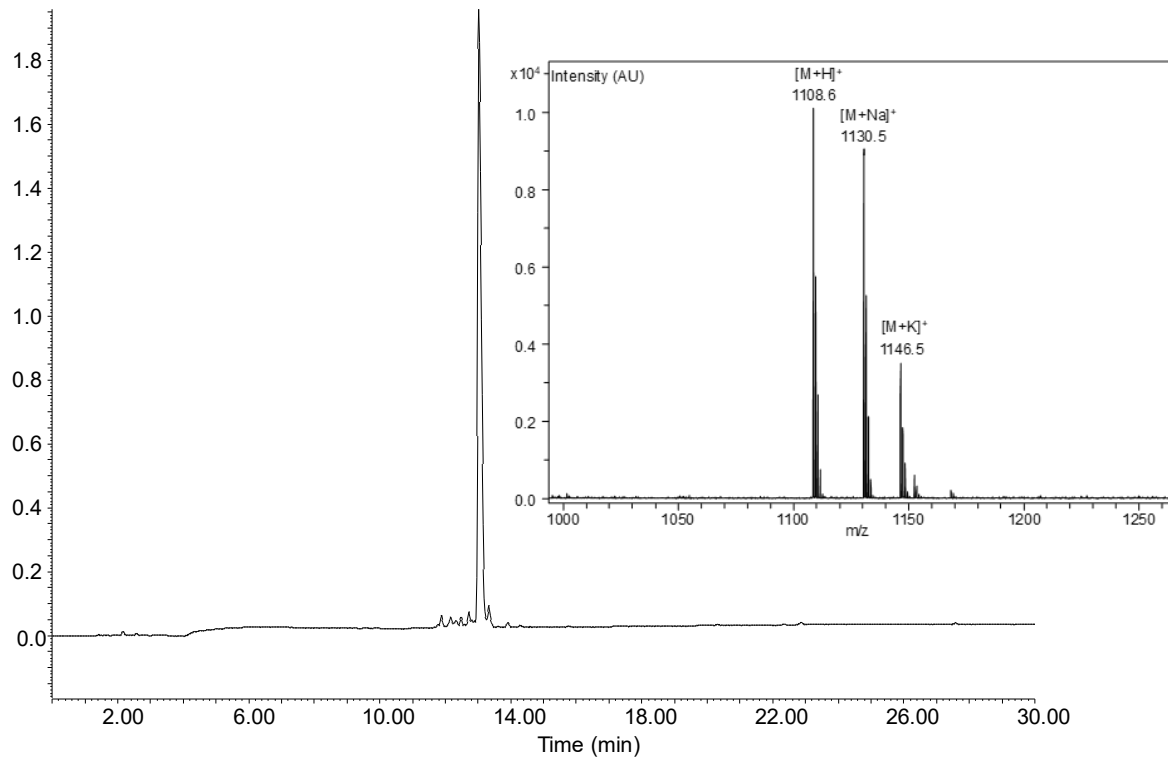
-ILKEPVHGA-SEA^{off} (9a)

Intensity (light scattering, AU)

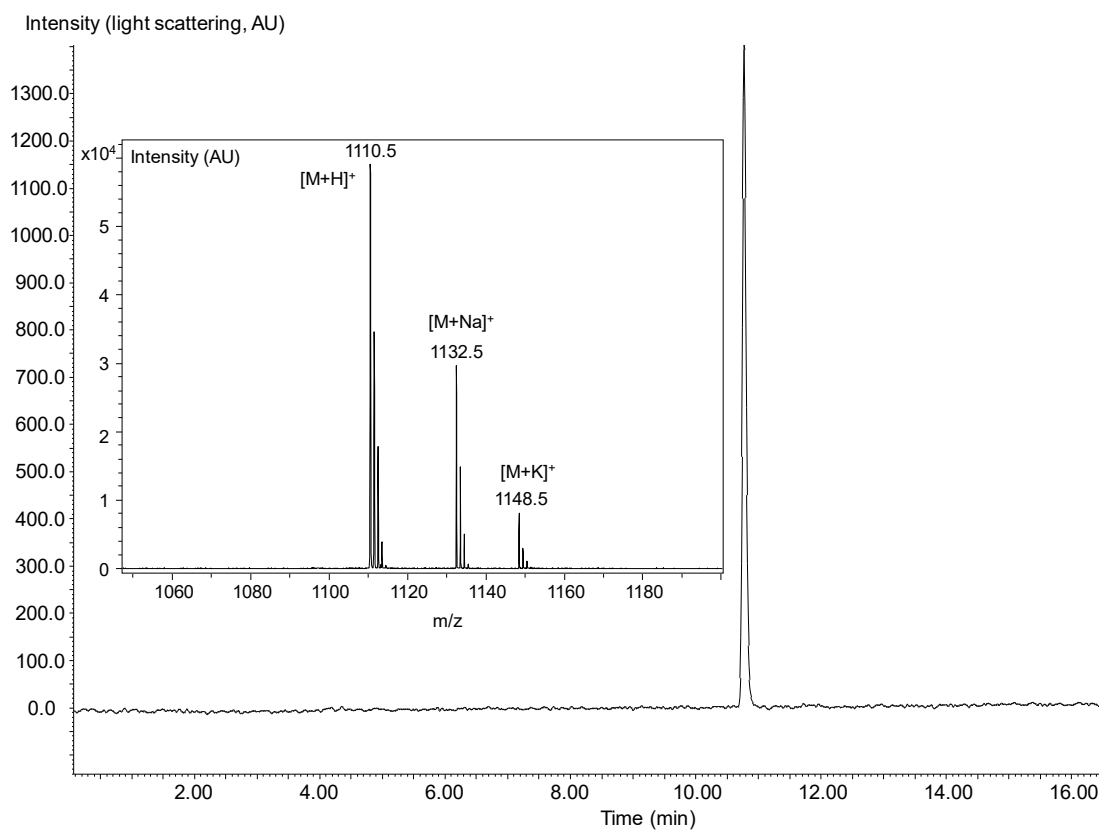


-ILKEPVHGV-SEA^{off} (9b)

Intensity (light scattering, AU)



-ILKEPVHGT-SEA^{off} (**9c**)



-ILKEPVHGI-SEA^{off} (**9d**)

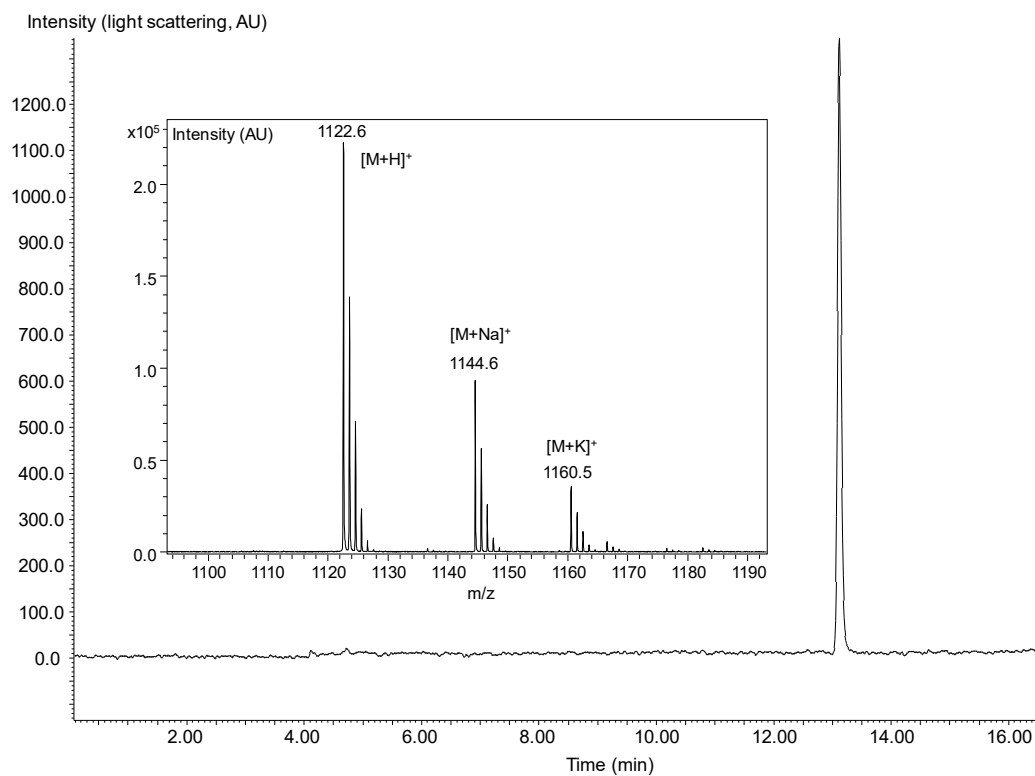


Figure S1. HPLC analysis of peptide **9a-d** and MALDI-TOF data corresponding to each product, matrix α -cyano-4-hydroxycinnamic acid, positive detection mode.

- RLSPEYYDLARAHLRDEEKSCPCLAQEGPQGDLTKTQELGRDYRT-SEA^{off} (**9e**)

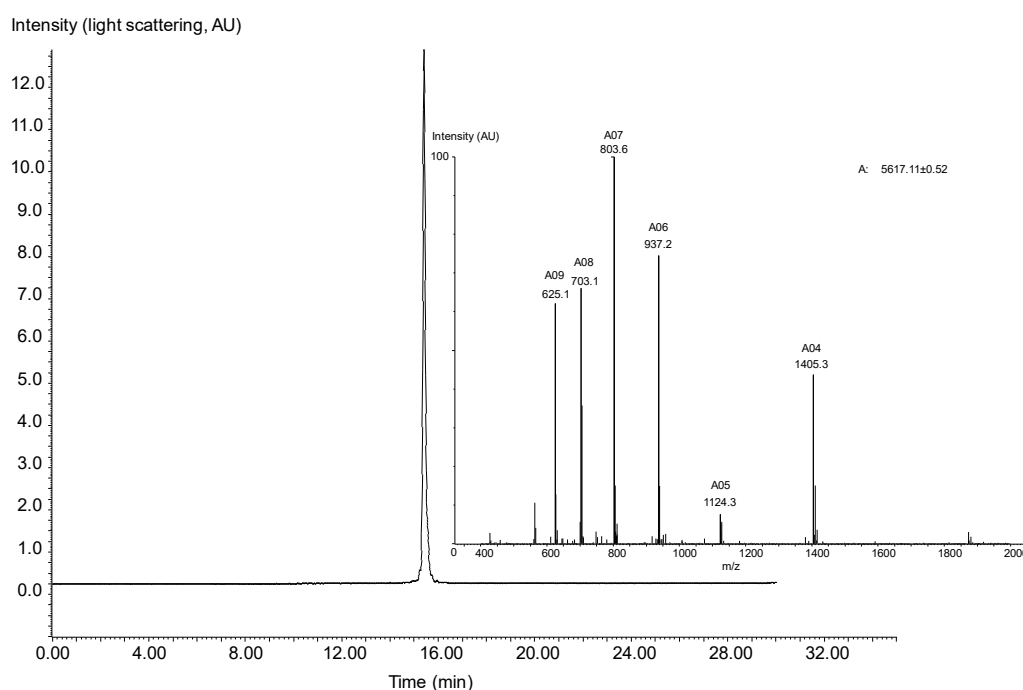


Figure S2. LC-MS analysis of peptide **9e**. LC trace, eluent A 0.1% TFA in water, eluent B 0.1% TFA in CH₃CN/water: 4/1 by vol. C3 Zorbax 300SB Å 3.5 µm (4.6 × 150 mm) column, gradient 0-100% B in 30 min (60 °C, 1 mL/min, detection 215 nm). MS trace. [M+H]⁺ m/z calcd. (monoisotopic) 5618.44, found 5618.11.

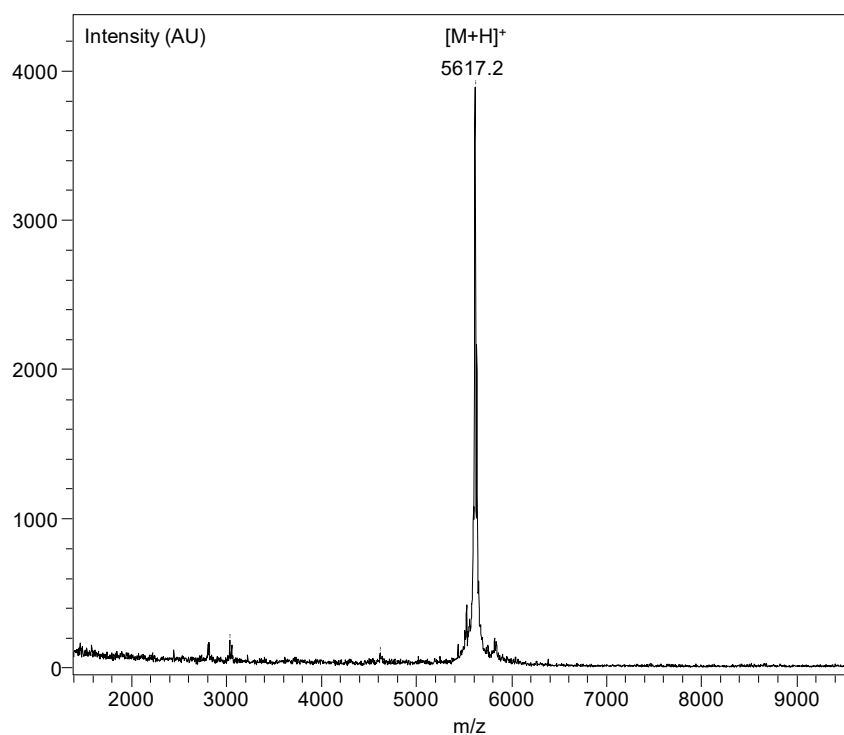


Figure S3. MALDI-TOF of peptide **9e**. Matrix sinapinic acid, positive detection mode, [M+H]⁺ calcd (monoisotopic) 5618.44, observed mass: 5617.2 (average).

-C(StBu)ANRC(StBu)TRNKGLPFTC(StBu)KAFVFDKARKQ-SEA^{off} (**17**)

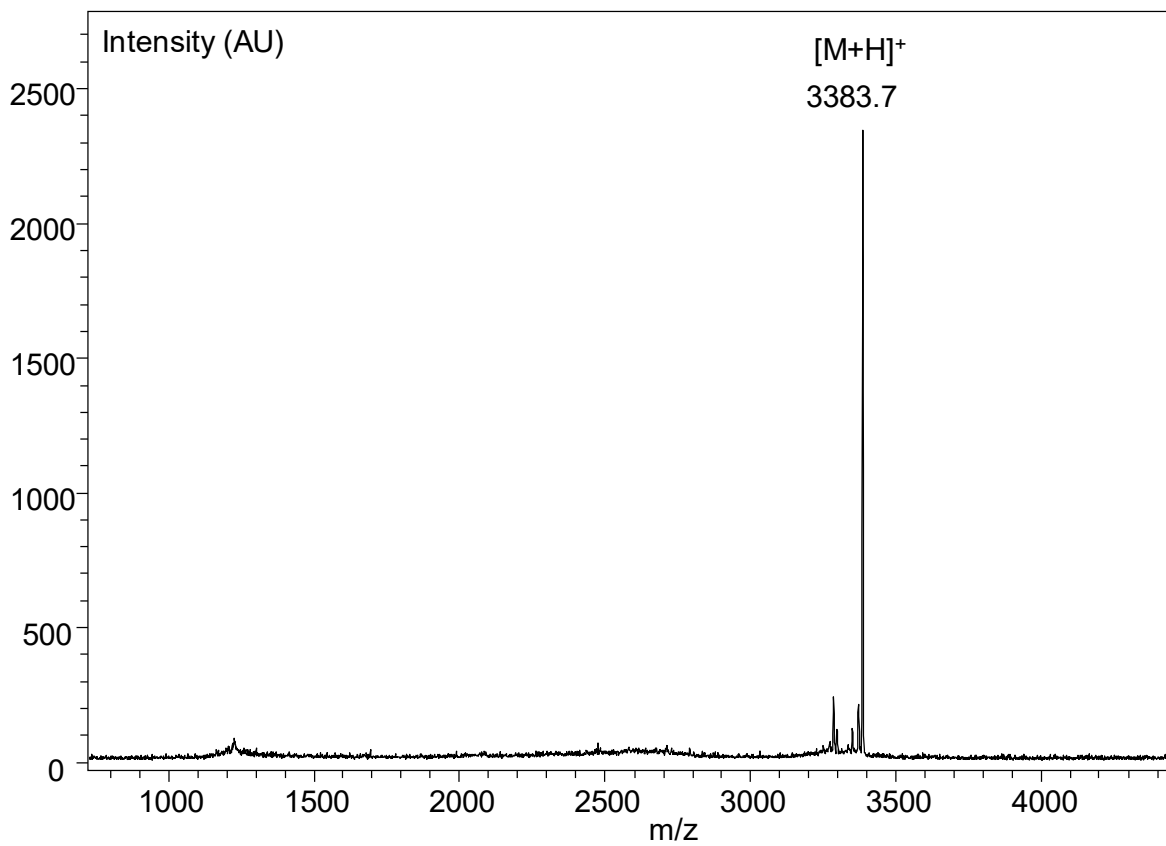
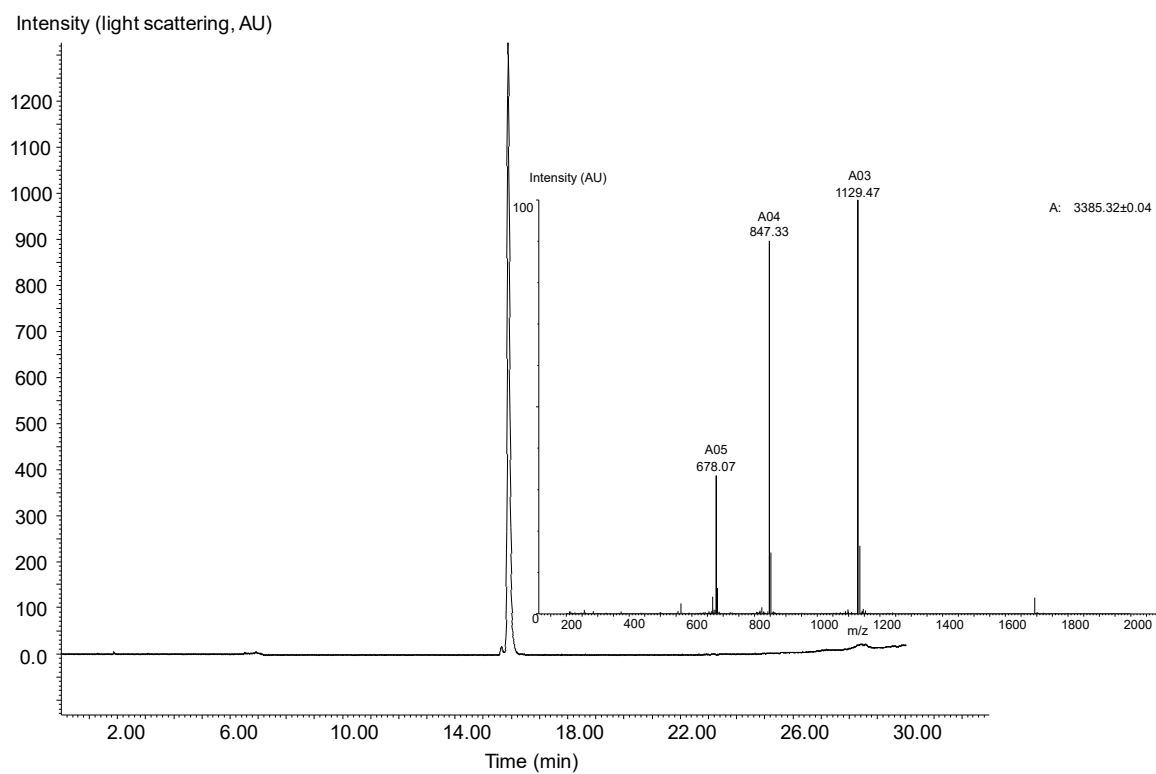


Figure S4. LC-MS analysis and MALDI-TOF of peptide **17**. Matrix α -cyano-4-hydroxycinnamic acid, positive detection mode.

- C(StBu)LWFPFNSMSSGVKKEFGHEFDLYENKDYIRA-SEA^{off} (**20**)

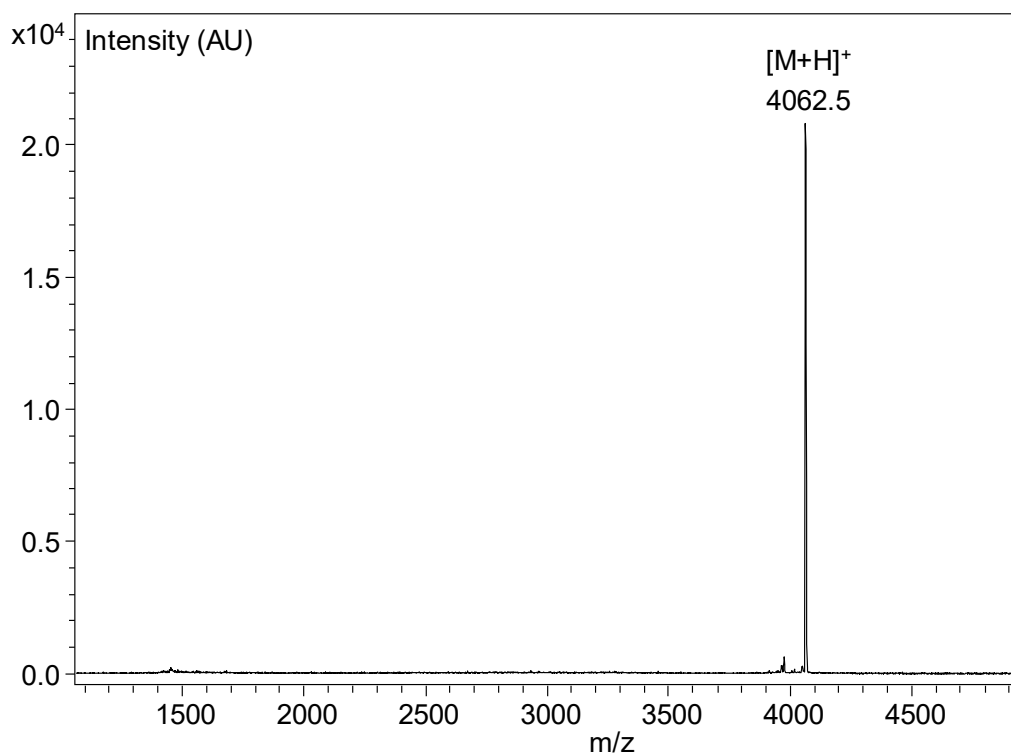
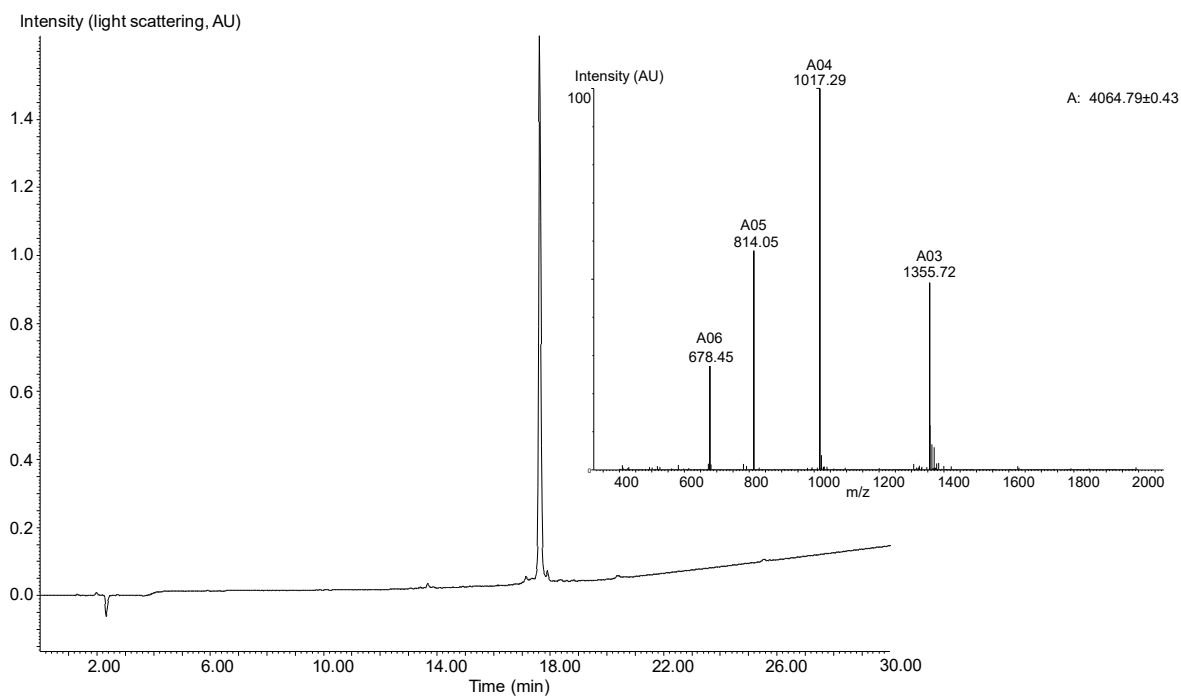


Figure S5. LC-MS analysis and MALDI-TOF of peptide **20**. Matrix α -cyano-4-hydroxycinnamic acid, positive detection mode.

-C(StBu)QPWSSMIPHEHSFLPSSYRGKDLQENY-SEA^{off} (**24**)

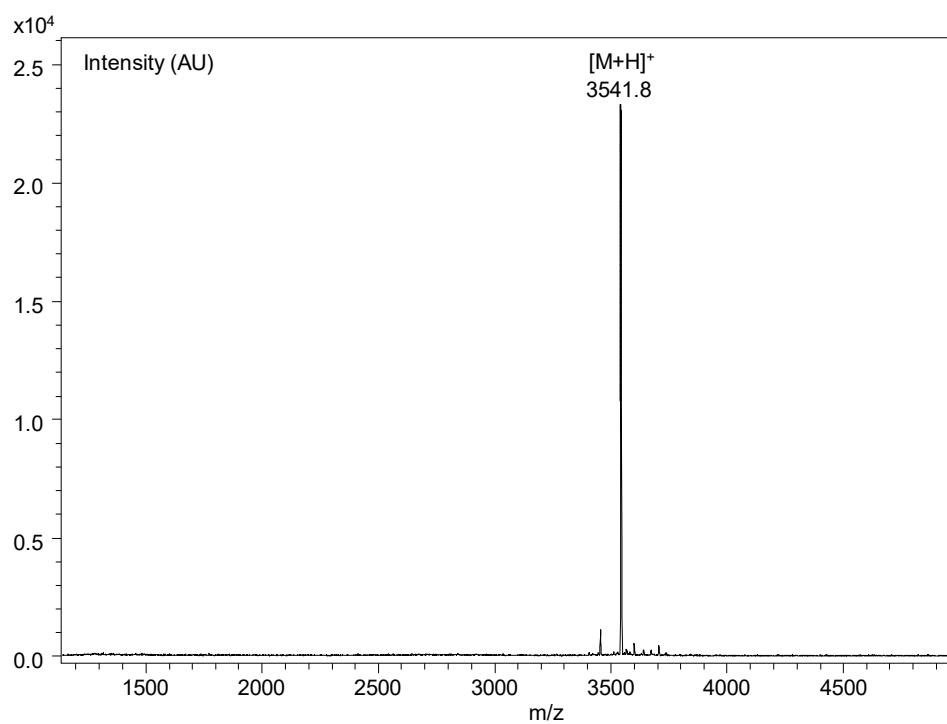
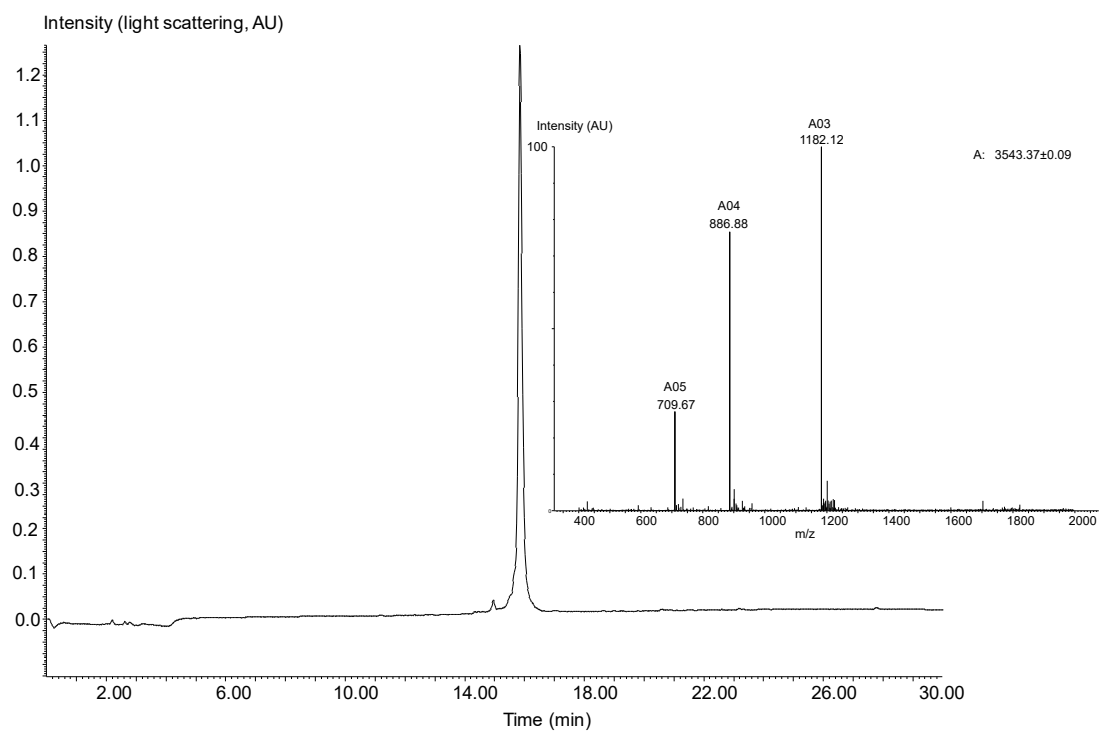


Figure S6. LC-MS analysis of peptide **24** and MALDI-TOF. Matrix α -cyano-4-hydroxycinnamic acid, positive detection mode.

- C(StBu)LTIVQKLKMKMVDKPTQRSVSNAATRVC(StBu)RTGRSRWRDV-SEA^{off} (**26**)

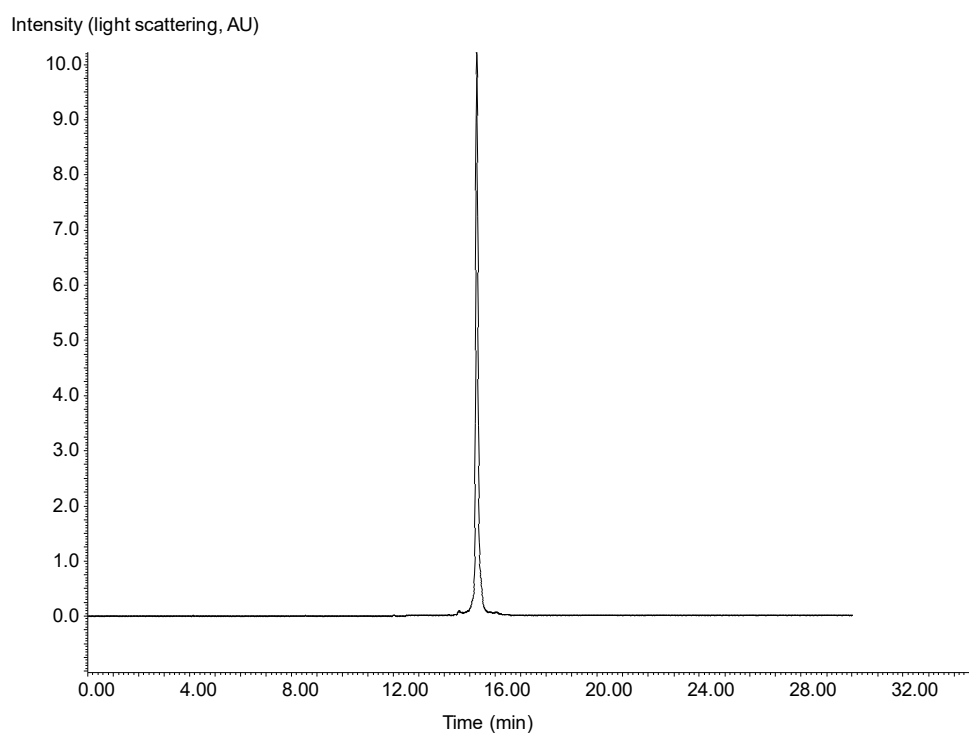


Figure S7. LC analysis of peptide **26**. LC trace, eluent A 0.10% TFA in water, eluent B 0.10% TFA in CH₃CN/water: 4/1 by vol. C3 Zorbax 300SB Å 3.5 µm (4.6 × 150 mm) column, gradient 0-100% B in 30 min (60 °C, 1 mL/min, detection 215 nm).

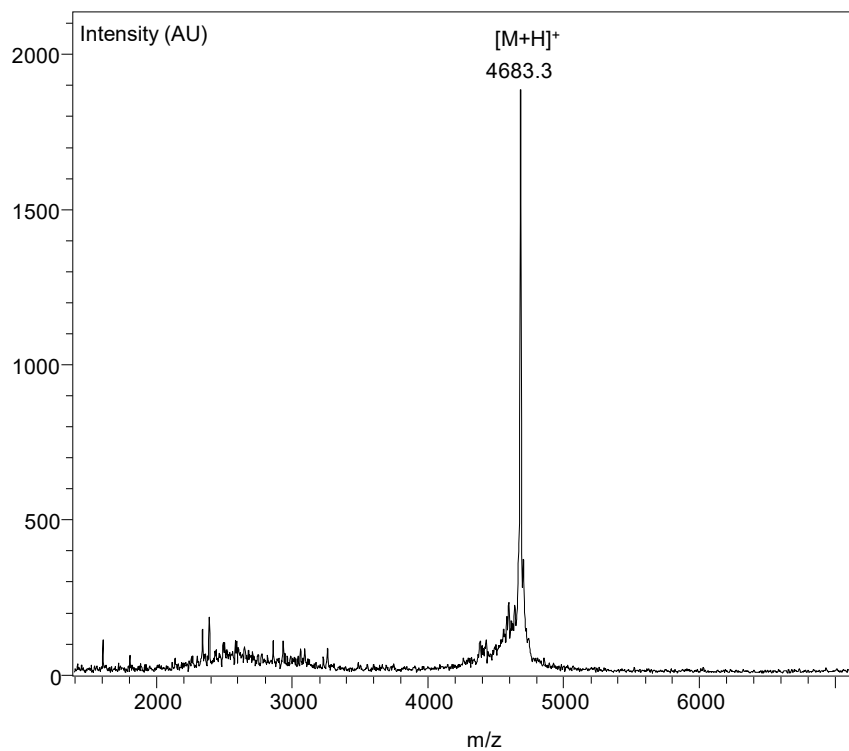


Figure S8. MALDI-TOF of peptide **26**. Matrix sinapinic acid, positive detection mode, [M+H]⁺ calcd (average) 4683.74, observed mass: 4683.3 (average).

4.3 Synthesis of MPA thioesters

Typical procedures for the synthesis of MPA thioester peptides using SEA^{off} peptides were described in detail elsewhere^{4,5}. The analytical HPLC and MS analyses of the purified MPA thioester peptides are shown below.

Yields for the HPLC purified MPA peptide thioesters

-**Peptide 12a**: 9.5 mg (68% yield, 10 μ mol scale), MALDI-TOF calcd. for $[M+H]^+$: 1051.6, observed mass: 1051.5 (monoisotopic).

-**Peptide 16**: 10 mg (71% yield, 2.2 μ mol scale), MALDI-TOF calcd. for $[M+H]^+$: 4594.4, observed mass: 4594.5 (average).

-ILKEPVHGA-S(CH₂)₂COOH (**12a**)

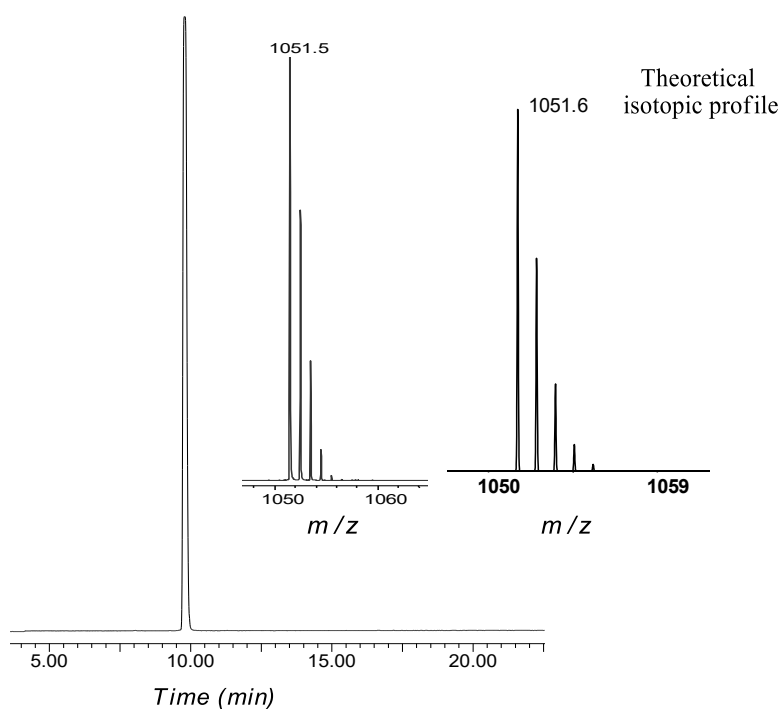


Figure S9. HPLC analysis and MALDI-TOF of peptide **12a**. Matrix α -cyano-4-hydroxycinnamic acid, positive detection mode.

-GQRKRRNTIHEFKKSAKTTLIKIDPALKIKTKKVNTADQ-S(CH₂)₂COOH (**16**)

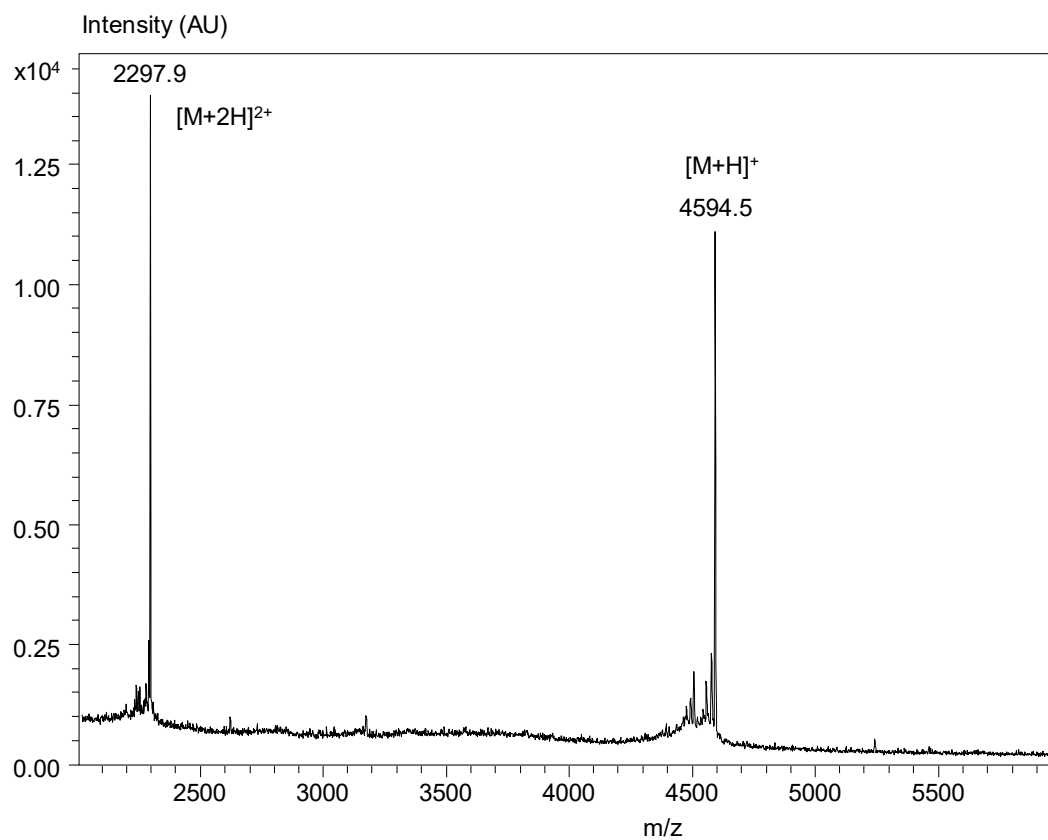
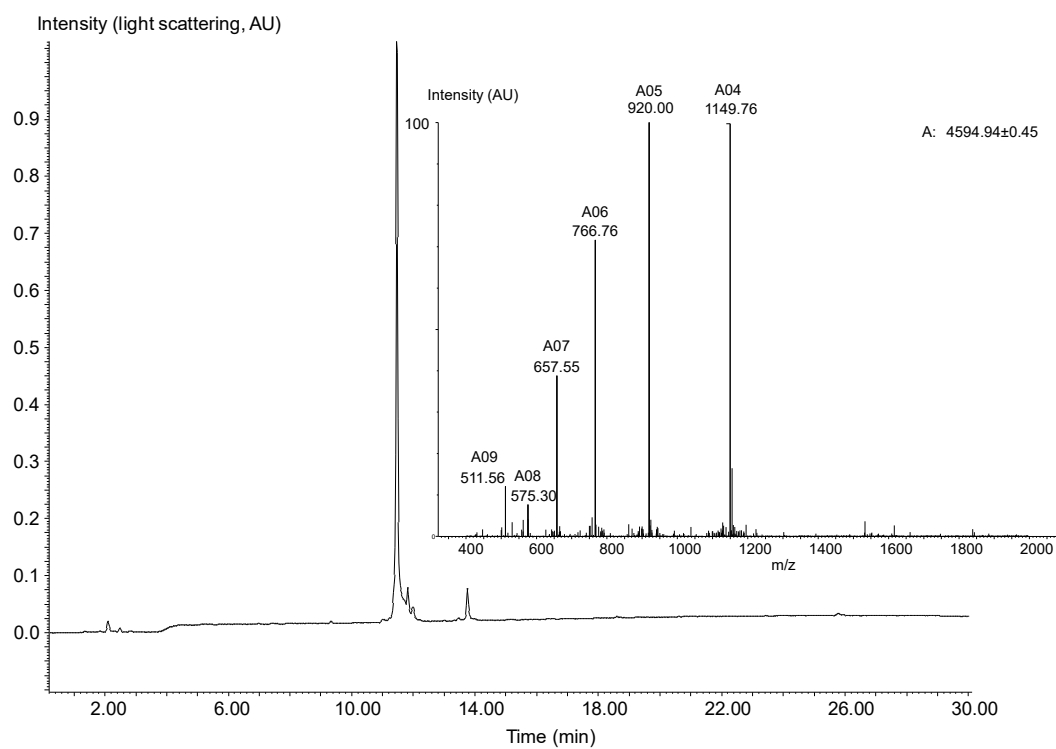


Figure S10. HPLC analysis and MALDI-TOF of peptide **16**. Matrix α -cyano-4-hydroxycinnamic acid, positive detection mode.

4.4 Synthesis of Cys peptides (13, 25 and 29)

Peptide elongation were performed on Rink-PEG-PS resin (NovaSyn TGR, 0.25 mmol, 0.25 mmol/g) or on PAL-Chemmatrix resin (0.05 mmol, 0.18 mmol/g) by using standard Fmoc/tert-butyl chemistry on an automated peptide synthesizer. The analytical HPLC and MS analyses of the purified peptide **13**, **25** and **29** are shown below.

Yield for the HPLC purified peptide segments

-**Peptide 13**: 89 mg (62% yield, 0.1 mmol scale), MALDI-TOF calcd. for $[M+H]^+$: 1093.6, observed mass: 1093.6 (monoisotopic).

-**Peptide 25**: 242 mg (21% yield, 0.25 mmol scale), MALDI-TOF calcd. for $[M+H]^+$: 4110,8 observed mass: 4109.9 (average).

-**Peptide 29**: 17.6 mg (7% yield, 0.05 mmol scale), MALDI-TOF calcd. for $[M+H]^+$: 4412,13 observed mass: 4412.2 (average).

- CILKEPVHGV-NH₂ (**13**)

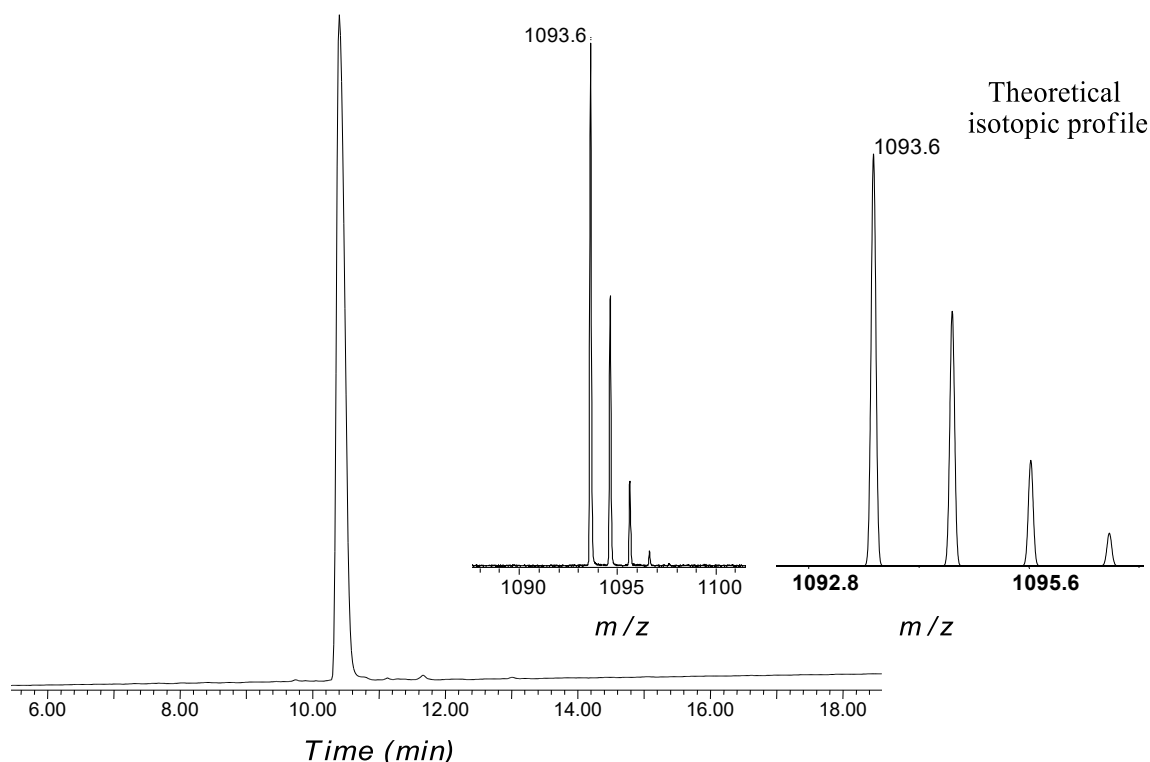


Figure S11. HPLC analysis and MALDI-TOF of peptide **13**. Matrix α -cyano-4-hydroxycinnamic acid, positive detection mode.

- CRNPRGEEGPPWCFTSNPEVRYEVCDDIPQCSEVLK(Biotin)-NH₂ (**25**)

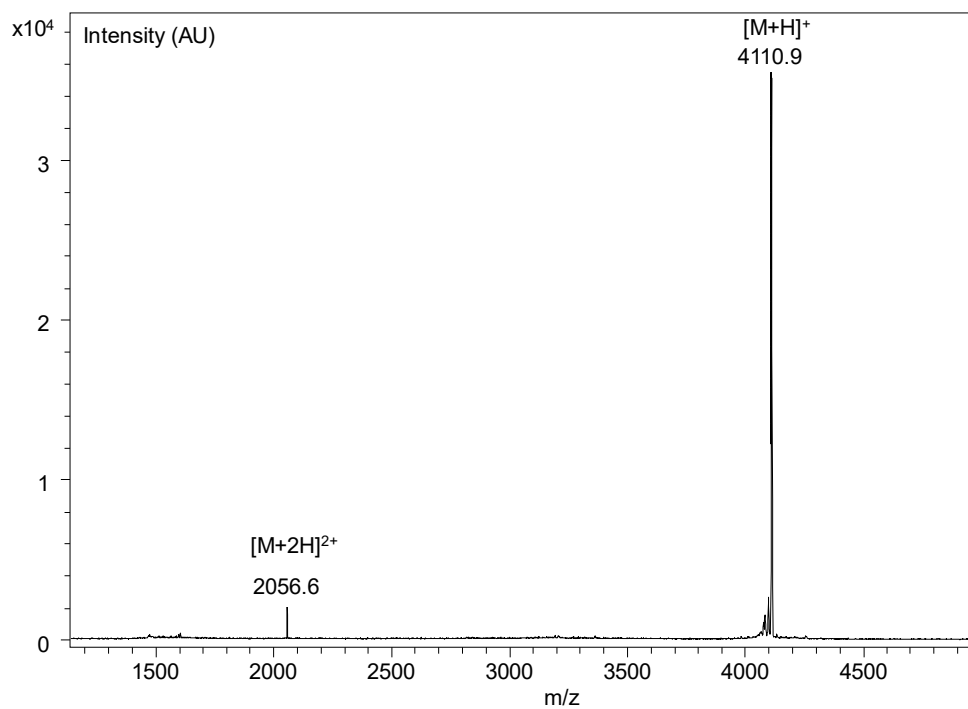
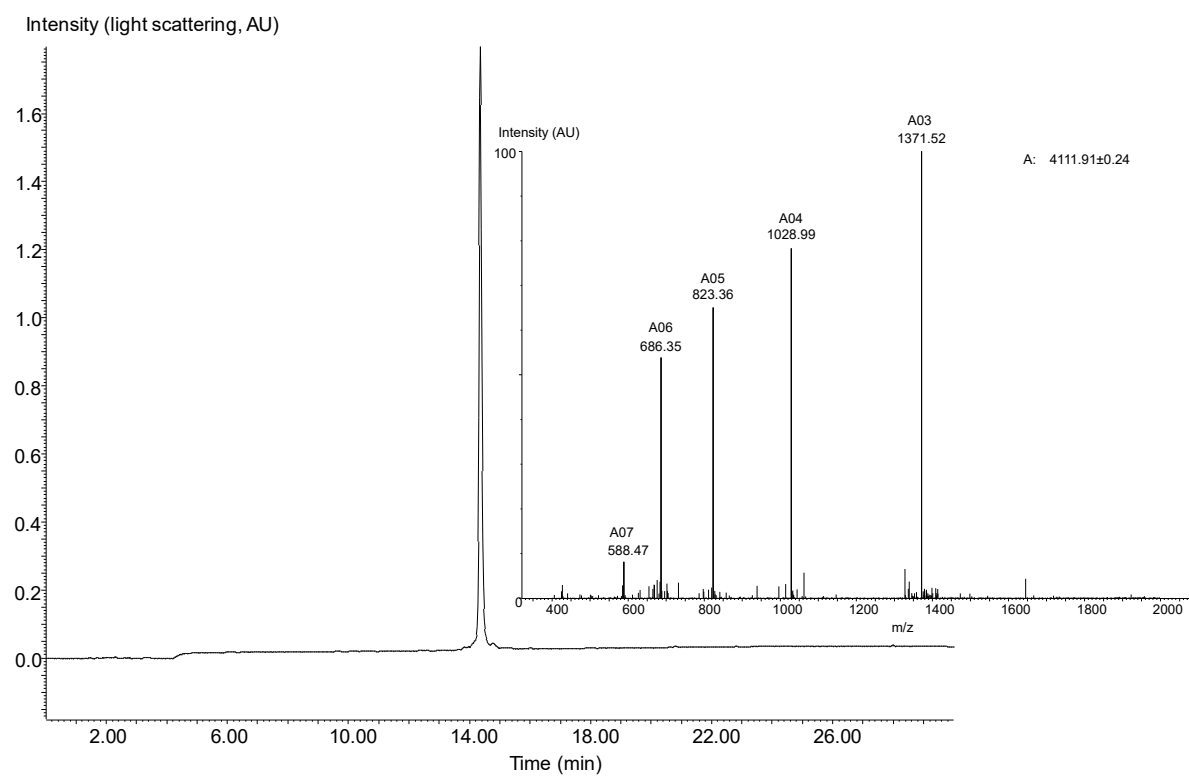


Figure S12. LC-MS analysis and MALDI-TOF of peptide **25**. Matrix α -cyano-4-hydroxycinnamic acid, positive detection mode.

-CRNFMRRYQSRVTQGLVAGETAQQICEDLRLCIPSTGPL-NH₂ (**29**)

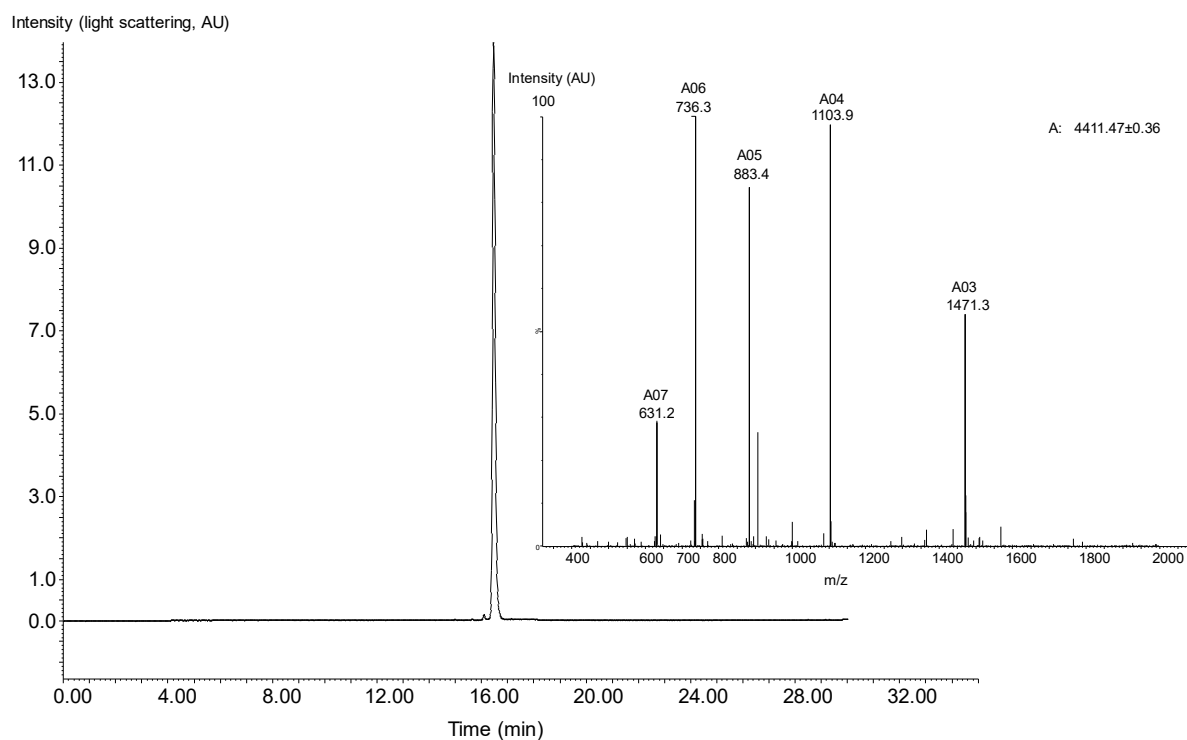


Figure S13. LC-MS analysis of peptide **29**. LC trace, eluent A 0.1% TFA in water, eluent B 0.1% TFA in CH₃CN/water: 4/1 by vol. C3 Zorbax 300SB Å 3.5 µm (4.6 × 150 mm) column, gradient 0-100% B in 30 min (60 °C, 1 mL/min, detection 215 nm). MS trace. [M+H]⁺ m/z calcd. (monoisotopic) 4412.13, found 4412.47.

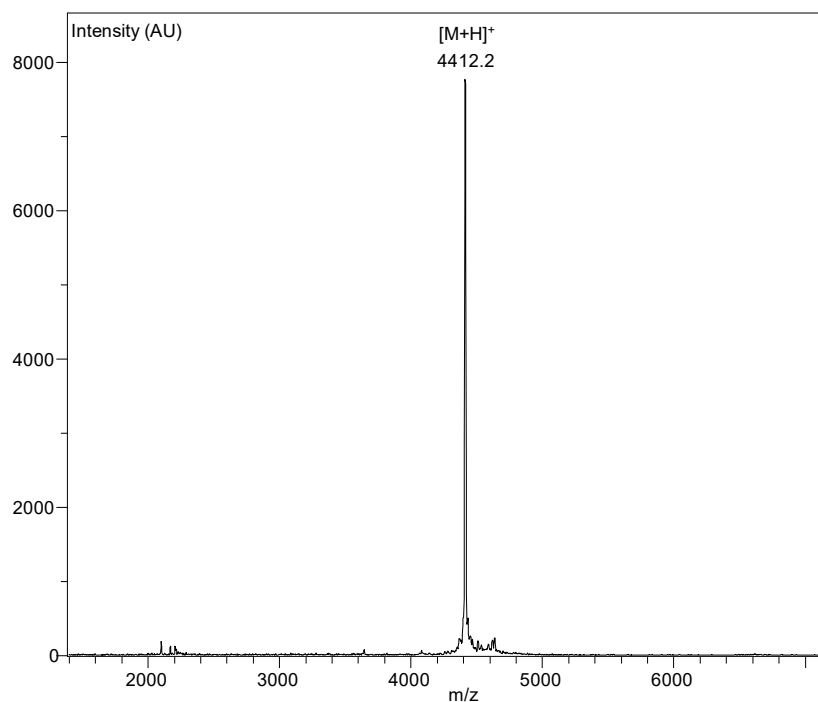
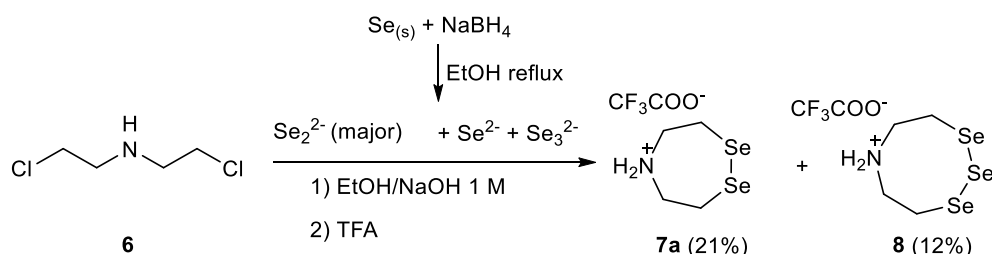


Figure S14. MALDI-TOF of peptide **29**. Matrix sinapinic acid, positive detection mode, [M+H]⁺ calcd (monoisotopic) 4412.13, observed mass: 4412.2 (average).

Accelerating chemoselective peptide bond formation using *bis*(2-selenylethyl)amido peptide selenolester surrogates (Chapter IV)

1. Synthesis of diselenide **7a** and **8**



Scheme S1. Synthesis of diselenide **7a** & triselenide **8**.

CAUTION: H_2Se is highly toxic. The reaction must be performed in an efficient fume hood with appropriate protection (glasses, lab coat and gloves).

1-Preparation of sodium diselenide: Absolute ethanol (27 ml) was added dropwise with magnetic stirring to metallic selenium powder (1.5 g, 19.8 mmol, 1.5 eq) and sodium borohydride (0.5 g, 13.2 mmol) cooled in an ice bath. After the vigorous exothermic reaction had occurred, the mixture was further stirred and heated at reflux for 1.5 h with N_2 passing into the liquid in order to expel H_2Se . The nitrogen flow containing H_2Se and going out of the reaction vessel is passed through NaOH and NaOCl traps respectively. The resulting brown/red colored solution was then cooled down at room temperature and used immediately in the next step.

2- Preparation of compounds **7a and **8**:** The *bis*(2-chloroethyl)amine hydrochloride **6** (1.2 g, 6.6 mmol, 0.6 eq) was dissolved in 1 M NaOH/EtOH: 1/1 by volume (5 mL) and then was added dropwise over a period of 20 min to the above solution of Na_2Se_2 or Na_2Se_3 . The solution was then stirred for 1 hour under argon at room temperature. The reaction mixture was diluted with 1 M NaOH (15 mL) and extracted with CH_2Cl_2 (3 \times 20 mL). The organic phase was dried over solid Na_2SO_4 and evaporated to dryness in vacuo. The crude product was purified by reversed-phase HPLC using a linear water-acetonitrile gradient to give 495 mg (21%) of 1,2,5-diselenazepane as the trifluoroacetate salt **7a** and 80 mg (12%) of 1,2,3,6-triselenazocane as the trifluoroacetate salt **8** (yellow powders).

The full characterization of diselenide **7a** and triselenide **8** can be found elsewhere⁶.

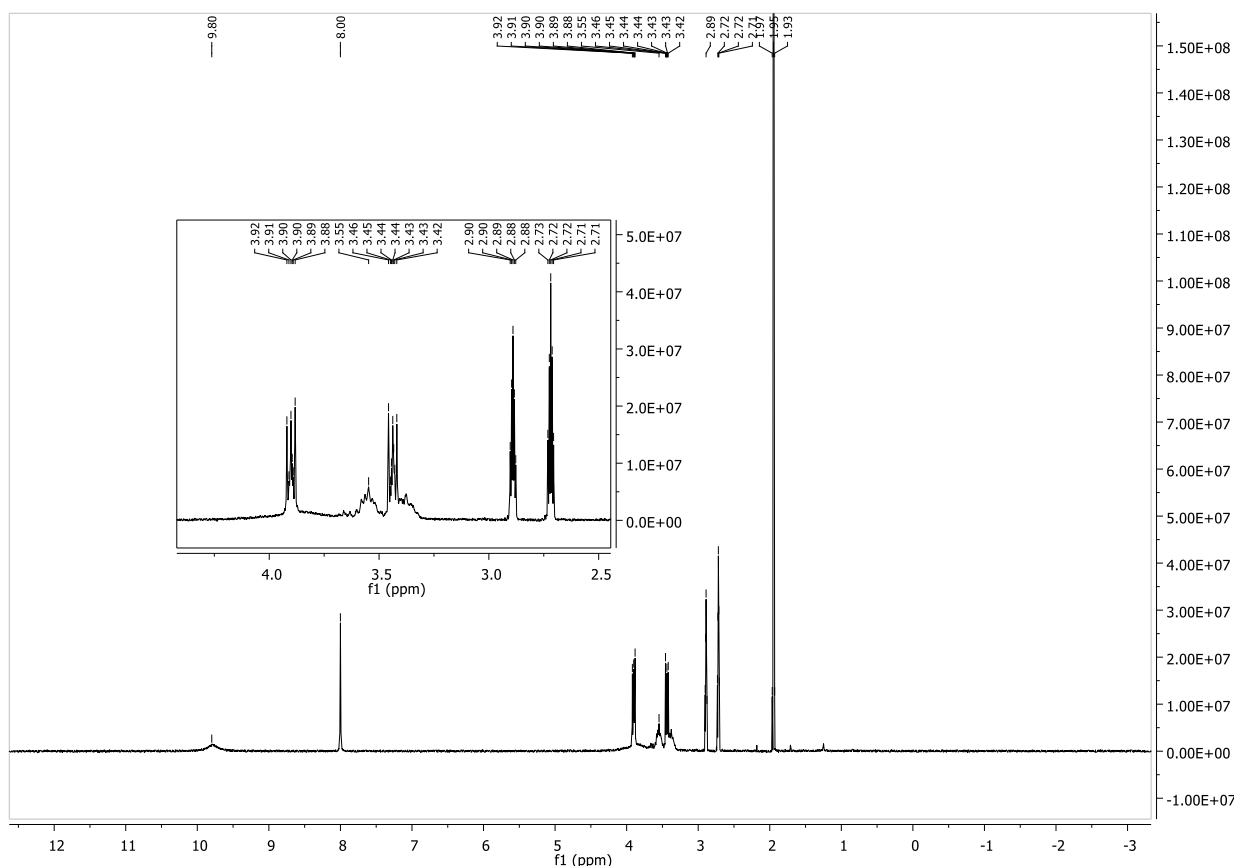


Figure S15. ^1H NMR (300 MHz) spectrum for compound **7a** (DMF- d_7).

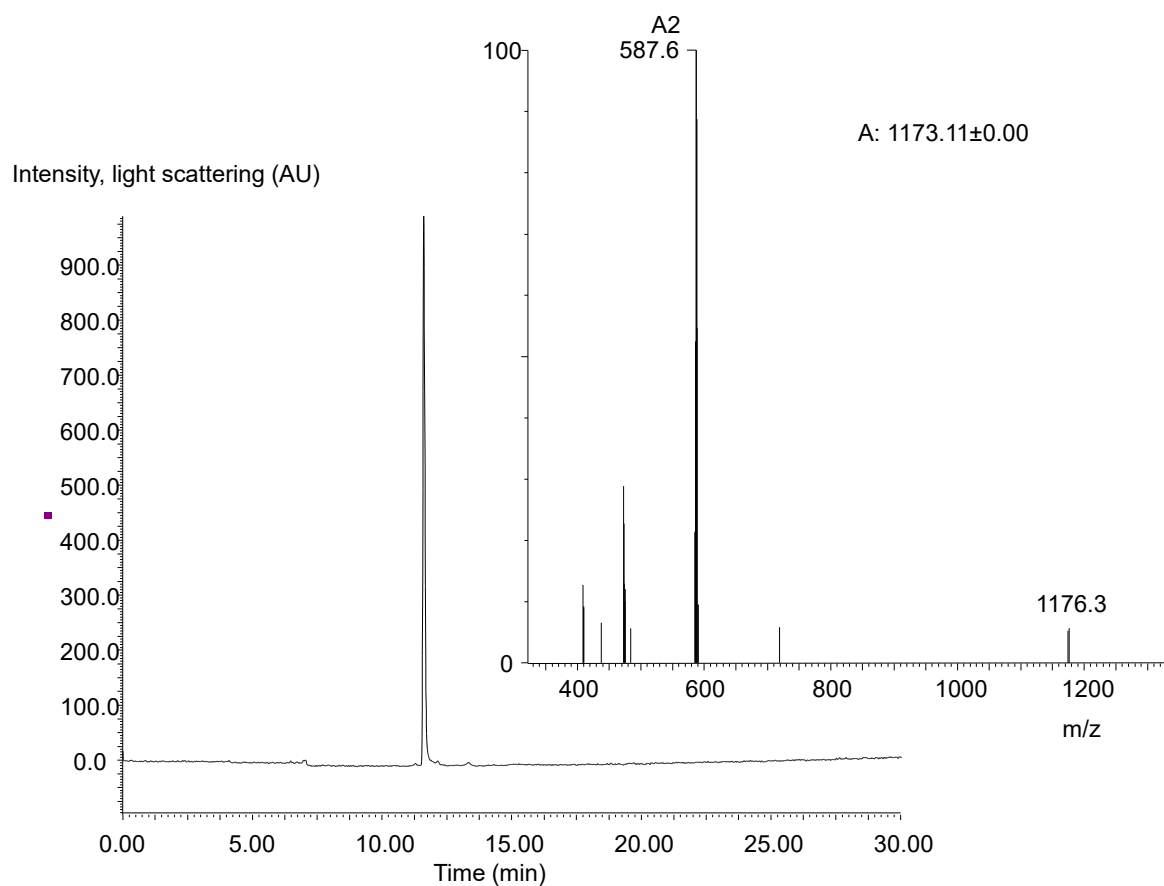
2. Synthesis of SeEA^{off} peptide **10**

*Synthesis of SeEA^{off} peptide **10** starting from SEA^{off} peptide **9a** using diselenide **7a** + Se(s)*

TCEP-HCl (28.7 mg, 0.1 mmol), was dissolved in 0.2 M pH 4.2 sodium acetate buffer (1 mL). Diselenide compound **7a** (24.1 mg, 70 μmol , 10 equiv) and metallic selenium (1.6 mg, 3 equiv) were suspended in the above solution (703 μL). NaOH (5 M) was then added to adjust the pH to 5.5.

Peptide **9a** (10 mg, 7 μmol) was dissolved in the above solution (703 μL , final peptide concentration 10 mM). The reaction mixture was shaken at 37 $^\circ\text{C}$ under nitrogen atmosphere and monitored by LC-MS. After 20 h, the mixture was diluted with water-TFA 1% (2 mL) and purified by reversed-phase HPLC (eluent A = water containing 0.1% TFA, eluent B=acetonitrile in water 4/1 containing 0.1% TFA, 50 $^\circ\text{C}$, detection at 215 nm, 6 mL/min, 0 to 10% eluent B in 10 min, then 10 to 25% eluent B in 45 min, C18 XBridge column) to give 5.92 mg of pure product (56%).

A)



B)

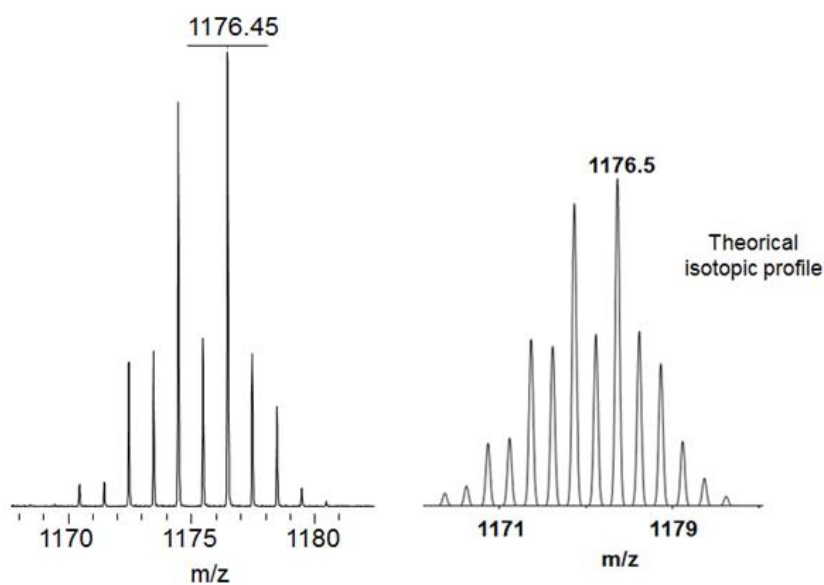


Figure S16. A) LC-MS analysis of peptide **10**. LC trace, eluent A 0.10% TFA in water, eluent B 0.10% TFA in CH₃CN/water: 4/1 by vol. C18 Xbridge BEH 300 Å 5 μm (4.6 x 250 mm) column, gradient 0-100 % B in 30 min (1 mL/min, detection 215 nm). MS trace: [M+2H]²⁺ m/z calcd. 589.3, obs. 587.6. B) MALDI-TOF analysis and comparison with the theoretical profile.

3. Kinetic measurements

Table S1 : Peptides used for the kinetic study

SEA ^{off} , SeEA ^{off} or thioester peptide	
9a	H-ILKEPVHGA-SEA ^{off}
9b	H-ILKEPVHGV-SEA ^{off}
10	H-ILKEPVHGA-SeEA ^{off}
12a	H-ILKEPVHGA-S(CH ₂) ₂ COOH
13	CILKEPVHGV-NH ₂

3.1 General procedure for kinetic measurements for the ligation of 9a, 10 and 12a with 13

For SEA peptides **9a** and thioester peptides **12a**: TCEP-HCl (29 mg, 0.1 mmol) and MPAA (17 mg, 0.1 mmol) were dissolved in 0.1 M pH 7.2 sodium phosphate buffer (1 mL). NaOH (5 M) was then added to adjust the pH to 7.1.

For SeEA peptides **10**: TCEP-HCl (29 mg, 0.1 mmol), MPAA (17 mg, 0.1 mmol) and Se=TCEP (12 mg, 35 μ mol) were dissolved in 0.1 M pH 7.2 sodium phosphate buffer (1 mL). NaOH (5 M) was then added to adjust the pH to 7.1. Se=TCEP is used to inhibit the deselenization of SeEA peptides by TCEP⁷.

Peptides **9a or 10 or 12a** (1 μ mol) and peptide **13** (1.5 μ mol, 1.5 equiv) were dissolved in the above solution (300 μ L, final peptide concentration 3.5 mM, pH 7.1). The ligations were performed at 37°C under nitrogen atmosphere and monitored by RP-HPLC. For each time point, 10 μ L aliquots were withdrawn, quenched by adding 90 μ L of 1% aqueous TFA and extracted with Et₂O to remove MPAA. The samples were stored at -20°C until analysis.

3.2 General procedure for kinetic measurements for the SEA ligation of 9a-b with 13 at different pH

TCEP-HCl (57.3 mg, 0.2 mmol) and MPAA (33,6 mg, 0.2 mmol) were dissolved in 0.1 M pH 7.2 sodium phosphate buffer (1 mL). NaOH (5 M) was then added to adjust the pH to 5.5-7.2.

Peptides **9a-b** (7 mM) and peptide **13** (10.5 mM) were dissolved in the above solution. The reaction mixture was shaken at 37 °C under nitrogen atmosphere and monitored by HPLC. For each time point, 10 μ L aliquots were withdrawn, quenched by adding 90 μ L of 1% aqueous TFA and extracted with Et₂O to remove MPAA.

3.3 First-order reaction

To check the kinetic order of the reaction, we used the integrated form: $\ln\left(\frac{C_A}{C_{A0}}\right) = -kt$, where C_A is the concentration at time t , C_{A0} is the concentration at time 0 and k is the first order rate constant. Using a graphical method, we have to plot $\ln(C_A)$ with respect to time, if we obtain a straight line the reaction followed a first order. The slope of the line equal to $-k$ and the half-life $t_{1/2} = \frac{\ln(2)}{k}$.

We can also use the Pearson correlation coefficient, r , which is a measure of the linear correlation between two variables. As an example, the values calculated for the Pearson correlation coefficient and the p -value for the ligation between peptides **9a**, **10**, **12a** and **13** are shown in the table below.

Table S 2 : Table of the Pearson correlation coefficient r and p -value for the ligation describe in 3.1.

Peptide	r	p-value
9a	-0,9992	<0,01
10	-0,9889	<0,01
12a	-0,9961	<0,01

4. Synthesis of SeEA peptide 23

4.1 One-pot synthesis of MPA peptide thioester 19

6 M Gdn.HCl 0.1 M pH 7.2 sodium phosphate buffer was degassed during 30 min. MPAA was dissolved in this solution to a final concentration of 200 mM. NaOH (5 M) was then added to adjust the pH to 7.1.

Peptide thioester **16** (13.09 mg, 2.12 μ mol) and SEA^{off} peptide **17** (8.84 mg, 2.12 μ mol) were dissolved in the above solution (650 μ L, final peptide concentration 3.3 mM). The mixture was stirred at 37 °C under nitrogen atmosphere. After 20 h, the exchange of the SEA group by MPA was started by adding a 20 % MPA solution in water containing 0.2 M TCEP (pH=3.9, 650 μ L). After 20 h, the reaction mixture was diluted with 10% aqueous acetic acid (22 mL), extracted with Et₂O (3 x 15 mL) to remove MPAA and purified by RP-HPLC (eluent A = water containing 0.1% TFA, eluent B=acetonitrile in water 4/1 containing 0.1% TFA, 50 °C, detection at 215 nm, 6 mL/min, 0 to 10 % eluent B in 5 min, then 10 to 40 % eluent B in 60 min, C18 XBridge column) to give 9.84 mg of thioester peptide **19** (46.5%).

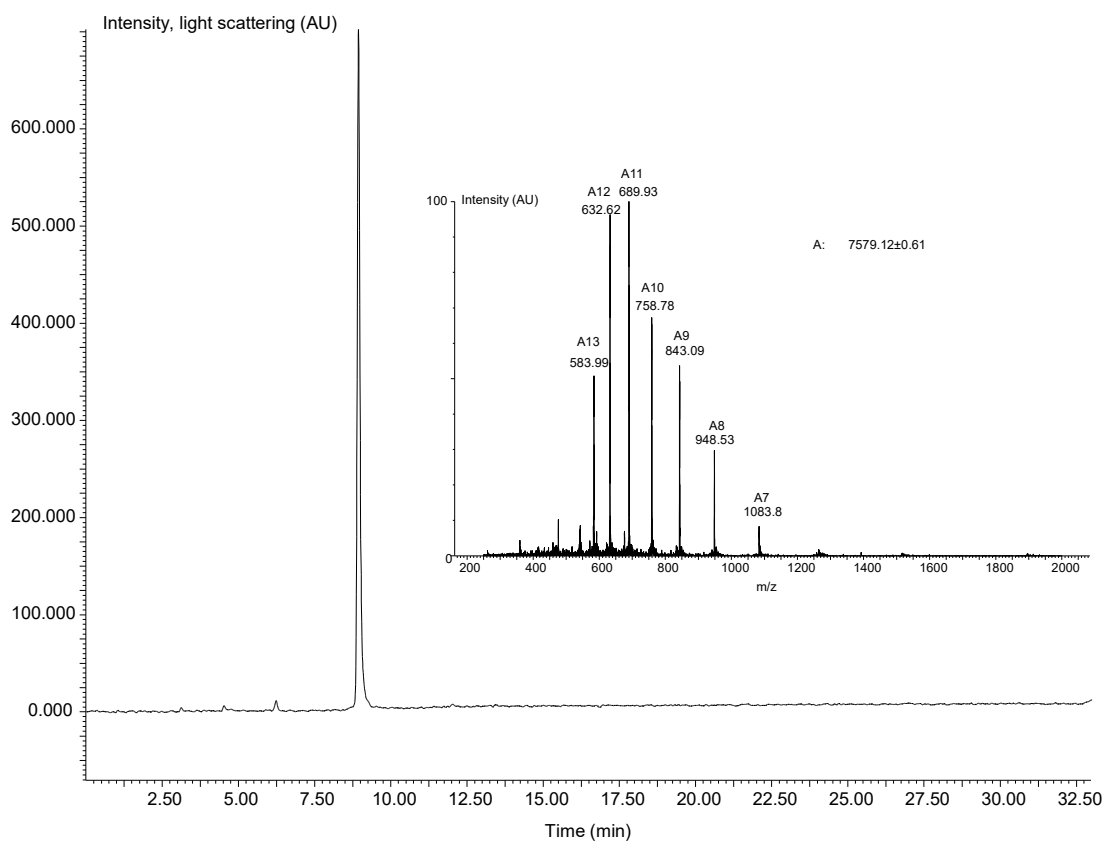


Figure S17. LC-MS analysis of thioester peptide **19**. LC trace, eluent A 0.10% TFA in water, eluent B 0.10 % TFA in CH₃CN/water: 4/1 by vol. C18 Xbridge BEH 300 Å 5 μm (4.6 x 250 mm) column, gradient 0-100 % B in 30 min (1 mL/min, detection 215 nm). MS trace: M calculated (mean) 7580.01, found 7579.12±0.61.

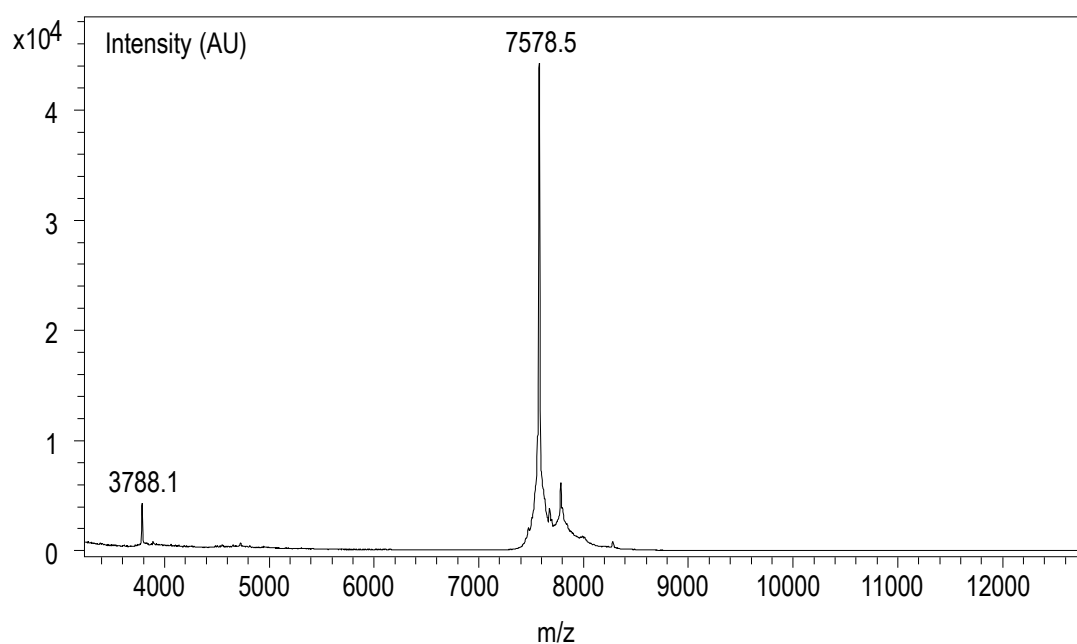
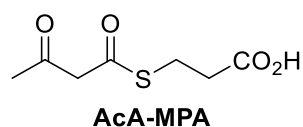


Figure S18. MALDI-TOF analysis of thioester peptide **19**. Matrix : 2,5-dihydroxybenzoic acid (DHB), positive mode. [M+H]⁺ calcd. (mean) 7580.01, found 7578.5.

4.2 One-pot synthesis of SeEA peptide 22

4.2.1 Synthesis of AcA-MPA



N-hydroxysuccinimidyl acetoacetate (265.4 mg, 1.33 mmol) was dissolved in CH₂Cl₂ (4 mL) at rt under argon. 3-Mercaptopropionic acid (116 μL, 1.33 mmol) and *N*-methylmorpholine (292.8 μL, 2.66 mmol) were added in one portion and the reaction medium was stirred overnight. Then, the solvent was evaporated and the resulting yellow oil was purified by silica gel chromatography (ethyl acetate/cyclohexane : 4/6 v/v containing 1% acetic acid) to give 49.0 mg of **AcA-MPA** (19.4%).

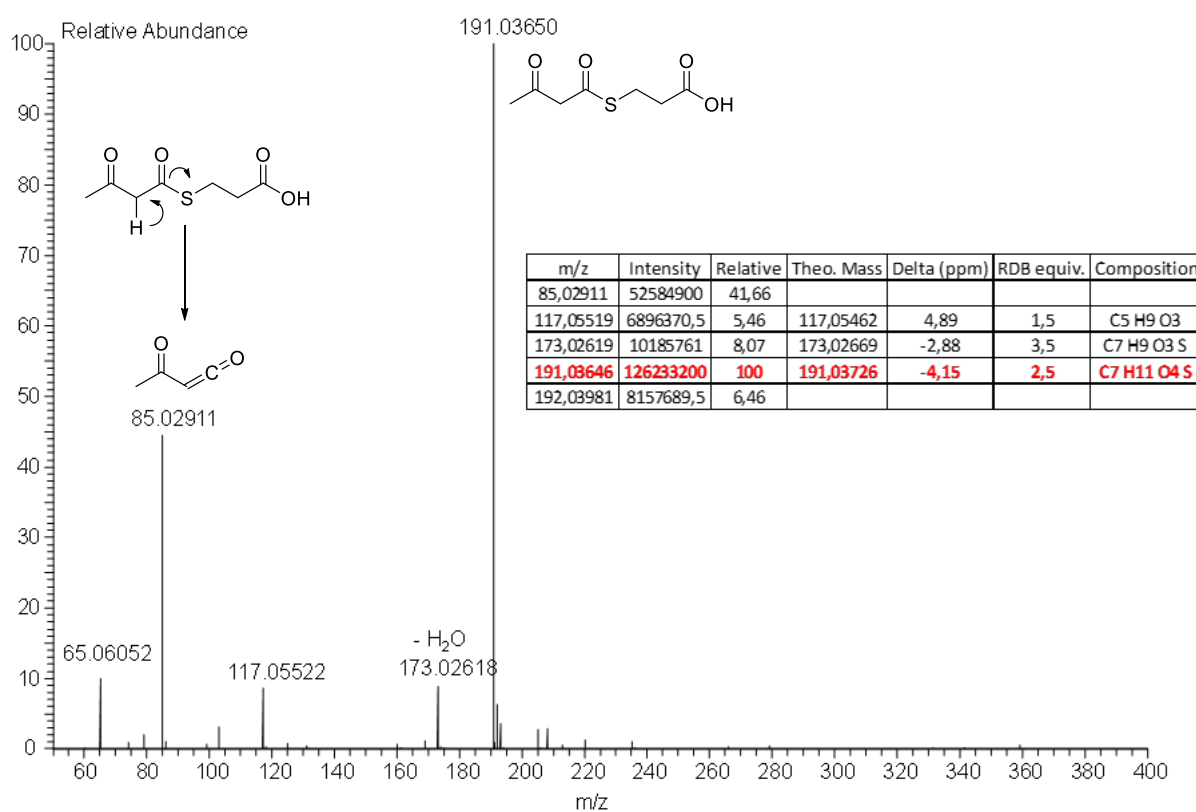


Figure S19. High resolution mass spectrometry of AcA-MPA (positive ion mode).

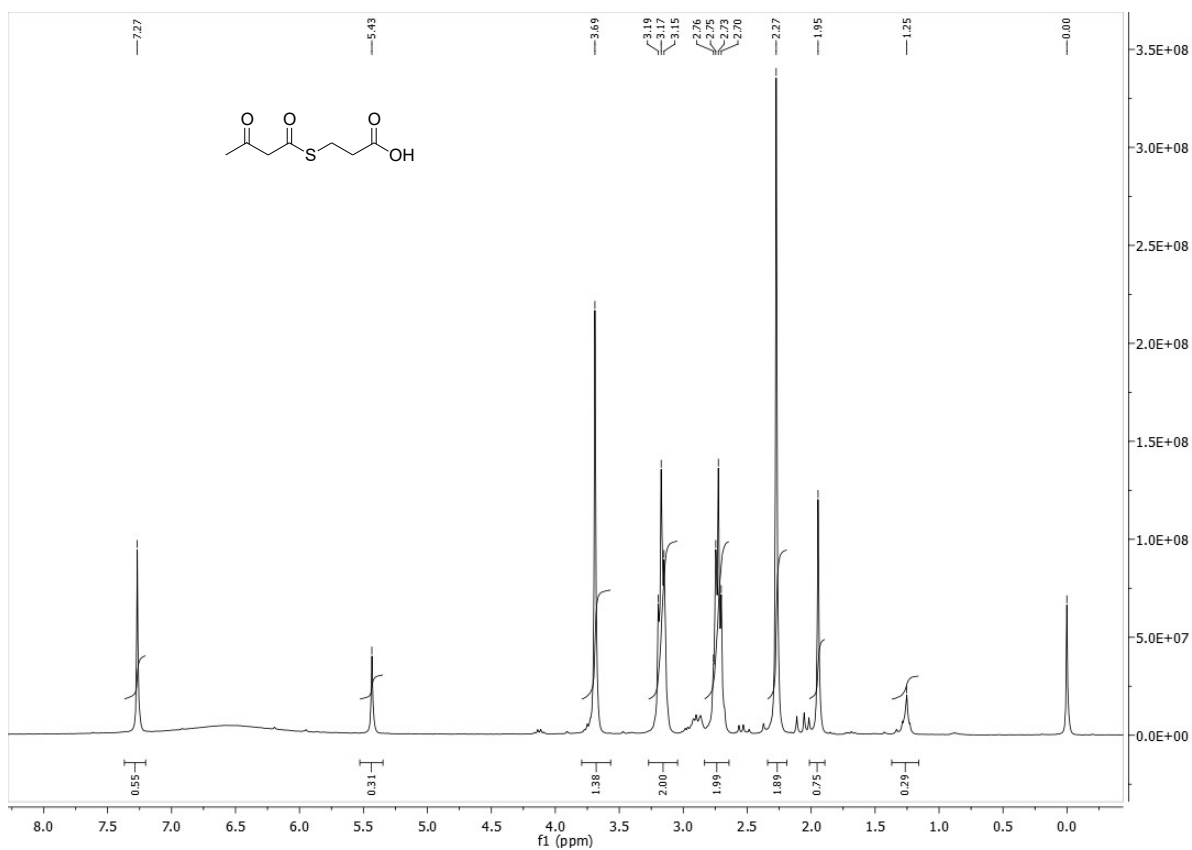


Figure S20. ^1H NMR spectrum (300.0 MHz, CDCl_3) of AcA-MPA.

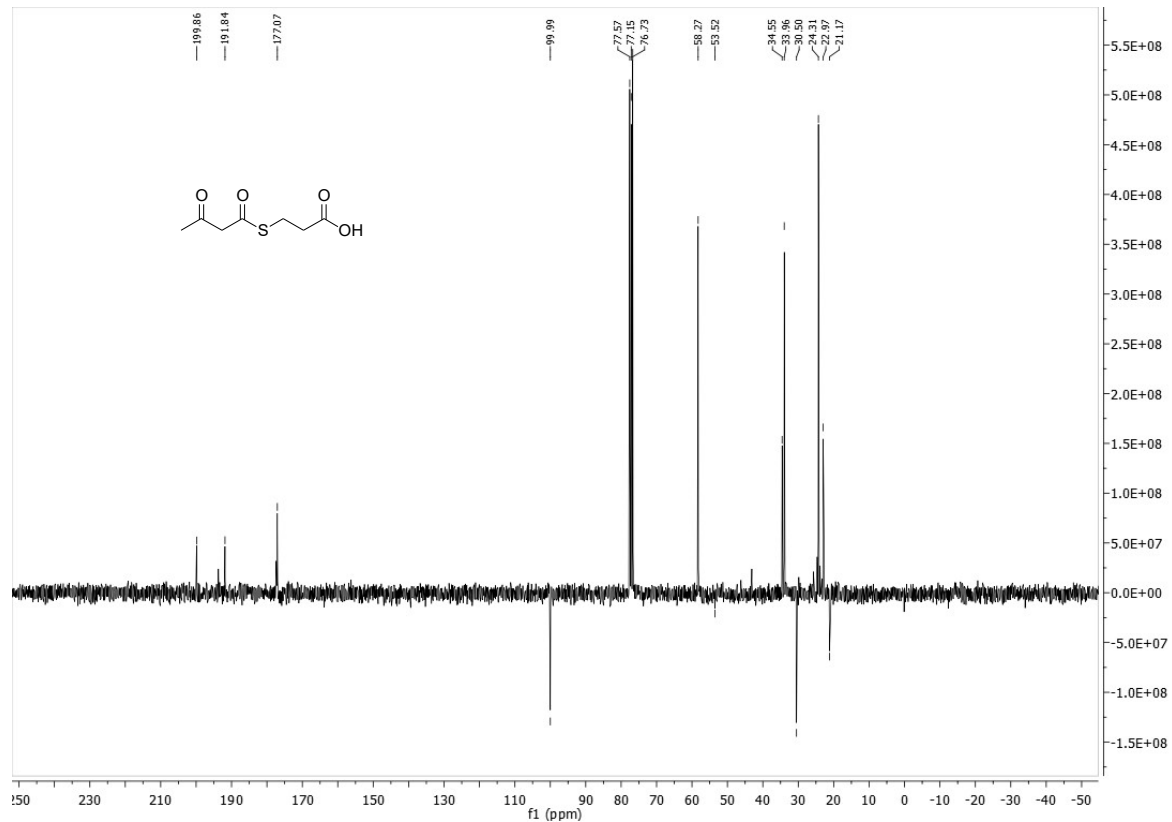


Figure S21. ^{13}C JMOD spectrum (75 MHz, CDCl_3) of AcA-MPA.

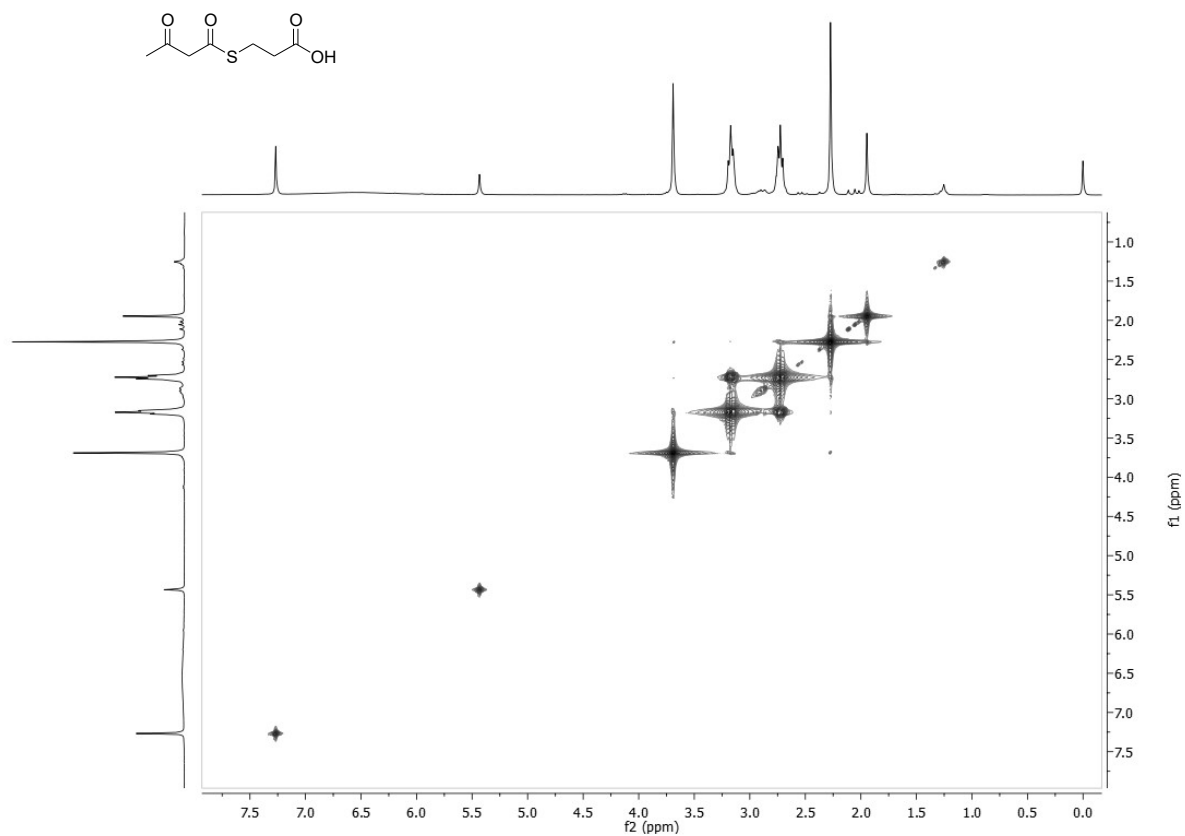


Figure S22. ^1H - ^1H COSY NMR spectrum of AcA-MPA.

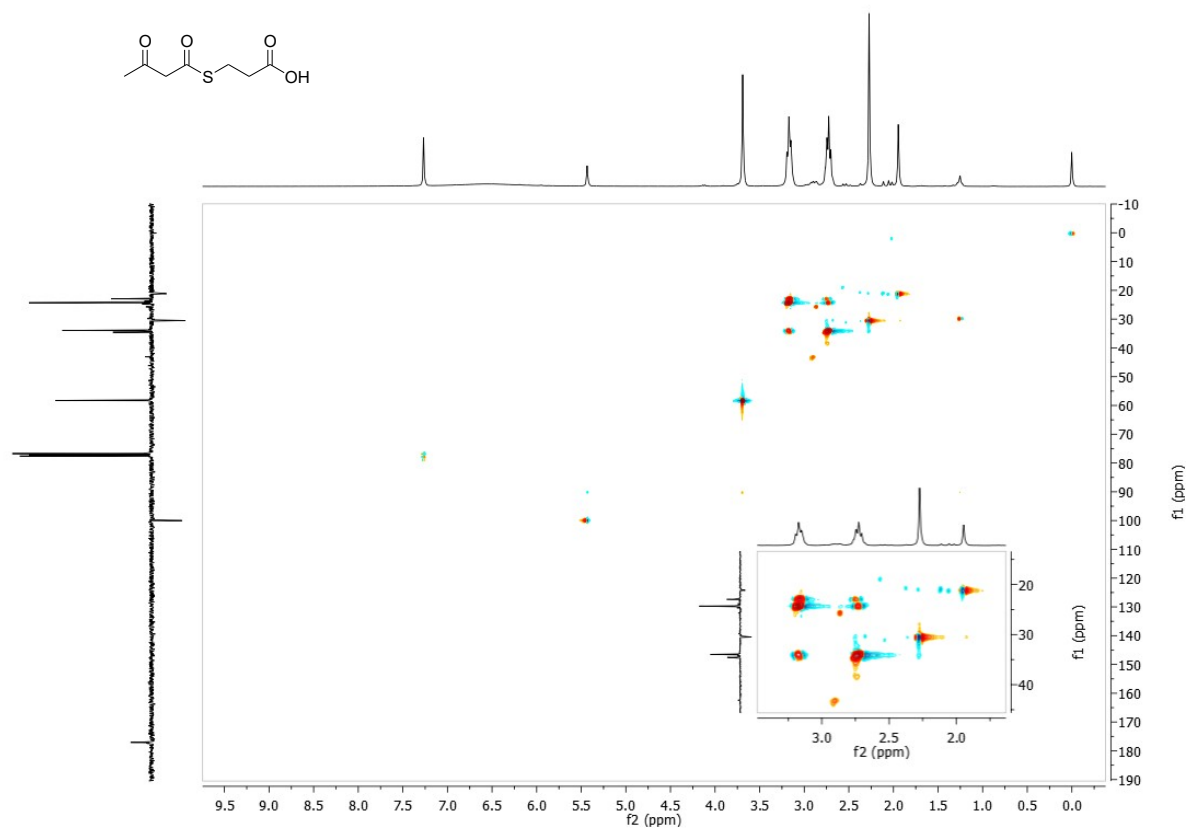


Figure S23. ^1H - ^{13}C HSQC NMR spectrum of AcA-MPA.

4.2.2 Synthesis of peptide 21

Synthesis of AcA-HGF 128-149-SEA^{off}

HGF 128-149: CIIGKGRSYKGTVSITKSGIK

0.1 M pH 7.2 sodium phosphate buffer was degassed during 30 min. Gdn.HCl (2.86 g, 0.03 mol) and MPAA (168.3 mg, 1 mmol) were dissolved in this buffer (5 mL qsp). NaOH (6 M) was then added to adjust the pH to 7-7.5.

Peptide CIIGKGRSYKGTVSITKSGIK-SEA^{off} (19.84 mg, 6.43 μ mol) and AcA-MPA (1.47 mg, 7.72 μ mol) were dissolved in the above solution (1.5 mL). The mixture was stirred at 37 °C under nitrogen atmosphere overnight.

The reaction mixture was then acidified with glacial acetic acid (400 μ L), diluted with water (6.1 mL) and extracted with Et₂O (3 x 2 mL) to remove MPAA. The crude peptide was purified by RP-HPLC (eluent A = water containing 0.1% TFA, eluent B = acetonitrile in water 4/1 containing 0.1 % TFA, 30 °C, detection at 215 nm, 6 mL/min, 0 to 20% eluent B in 5 min, then 20 to 40% eluent B in 60 min, C18 XBridge column) to give 11.9 mg of peptide AcA-CIIGKGRSYKGTVSITKSGIK-SEA^{off} (62.4%).

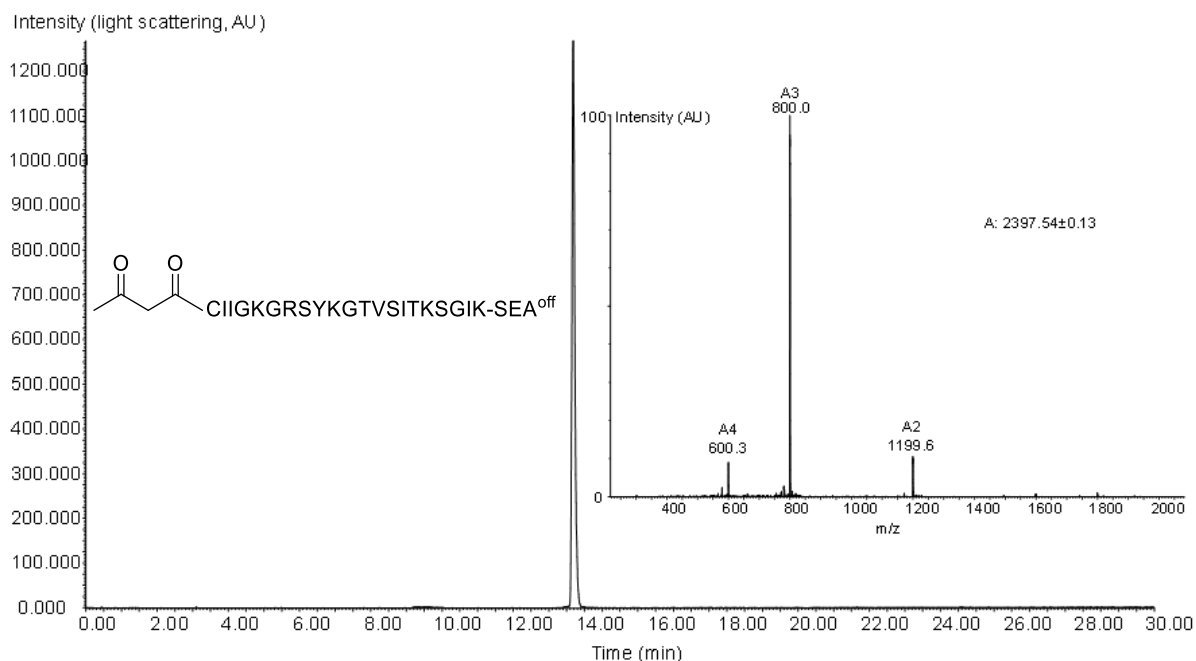


Figure S24. LC-MS analysis of peptide AcA-HGF 128-149-SEA^{off}. LC trace, eluent A 0.10% TFA in water, eluent B 0.10% TFA in CH₃CN/water: 4/1 by vol. C18 Xbridge BEH 300 Å 5 μ m (4.6 x 250 mm) column, gradient 0-100% B in 30 min (1 mL/min, detection 215 nm). MS trace: [M+2H]²⁺ m/z calcd. 1199.98, obs. 1199.6, [M+3H]³⁺ m/z calcd. 800.32, obs. 800.0, [M+4H]⁴⁺ m/z calcd. 600.49, obs. 600.3.

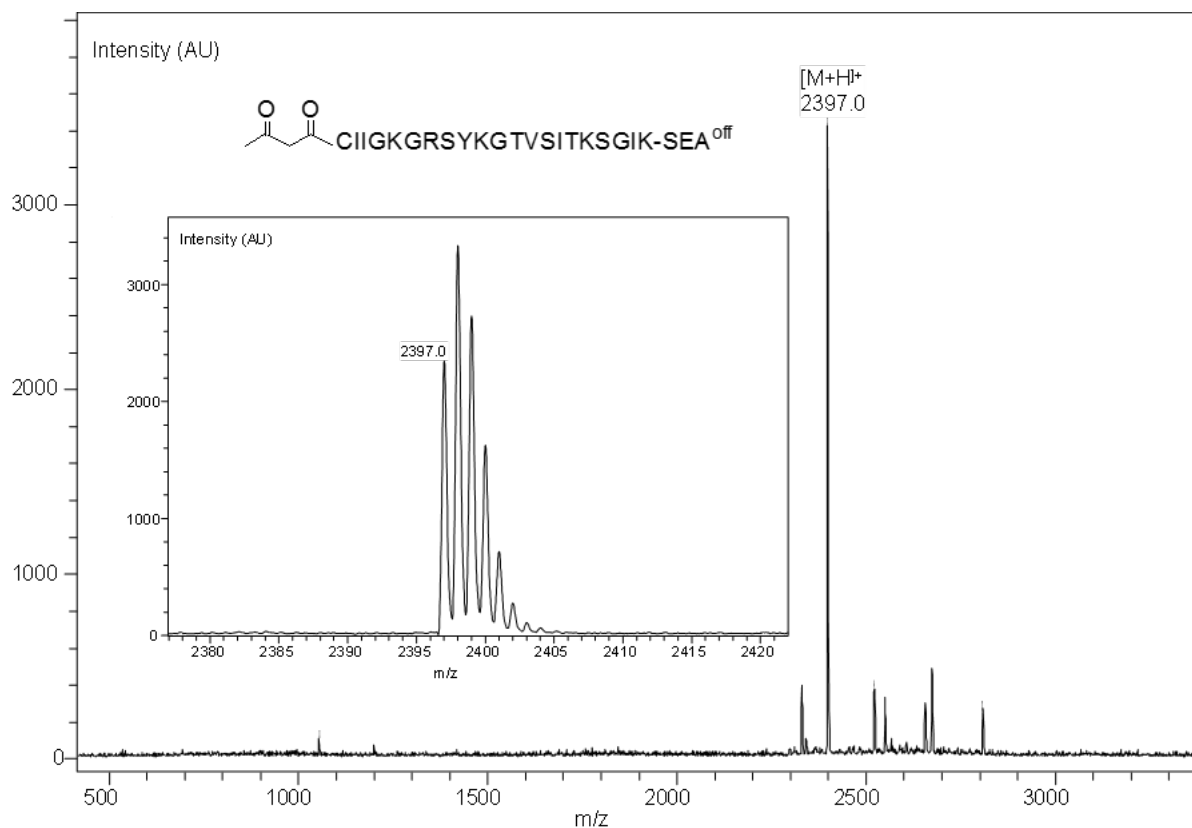


Figure S25. MALDI-TOF analysis of peptide **Aca-HGF 128-149-SEA^{off}**. Matrix, 2,5-dihydroxybenzoic acid (DHB) positive mode, [M+H]⁺ calcd. (monoisotopic) 2397.28, found 2397.0.

Exchange of the SEA group by MPA

A typical procedure for the exchange of the SEA group by MPA was described in detail elsewhere.^{4, 5} The analytical HPLC and MS analyses of the purified thioester peptide **21** are shown below.

Yields for the HPLC purified MPA peptide thioester 21

Peptide 21: 6.41 mg (40% yield, 5.6 μmol scale), MALDI-TOF calcd. for [M+H]⁺: 2368.3, observed mass: 2368.3 (monoisotopic).

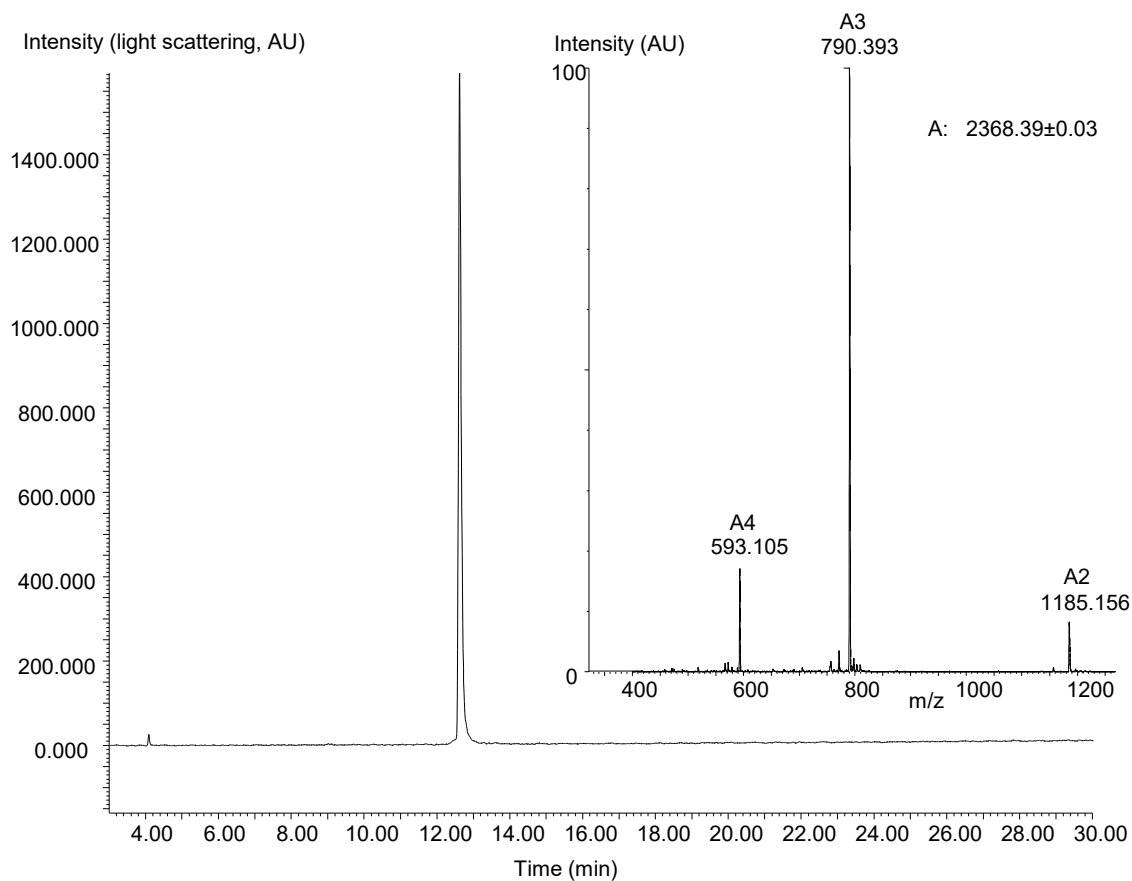


Figure S26. LC-MS analysis of peptide **21**. LC trace, eluent A 0.10 % TFA in water, eluent B 0.10% TFA in CH₃CN/water: 4/1 by vol. C18 Xbridge BEH 300 Å 5 μm (4.6 x 250 mm) column, gradient 0-100% B in 30 min (1 mL/min, detection 215 nm). MS trace: M calculated (mean) 2368.84, found 2368.39±0.03.

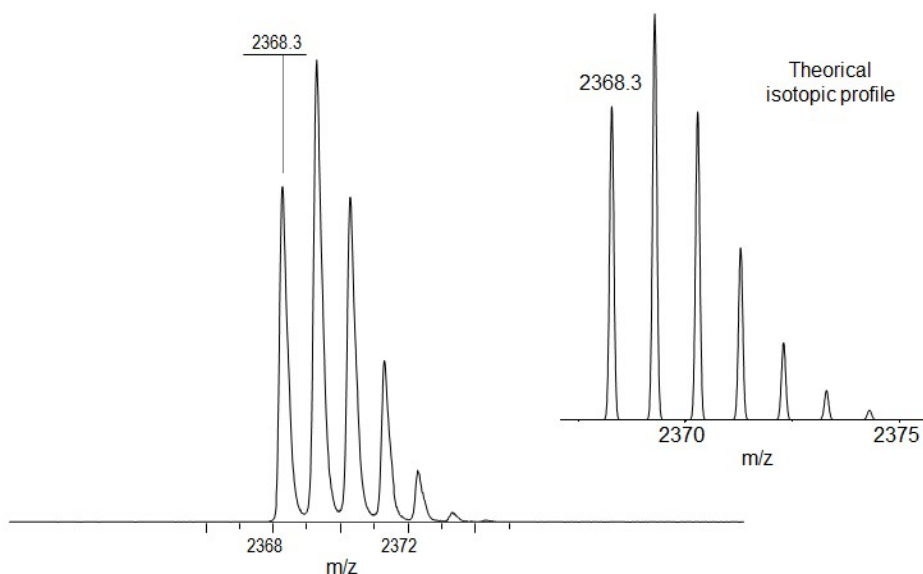


Figure S27. MALDI-TOF analysis of thioester peptide **21**. Matrix, 2,5-dihydroxybenzoic acid (DHB) positive mode. $[M+H]^+$ calcd. (monoisotopic) 2368.27, found 2368.3.

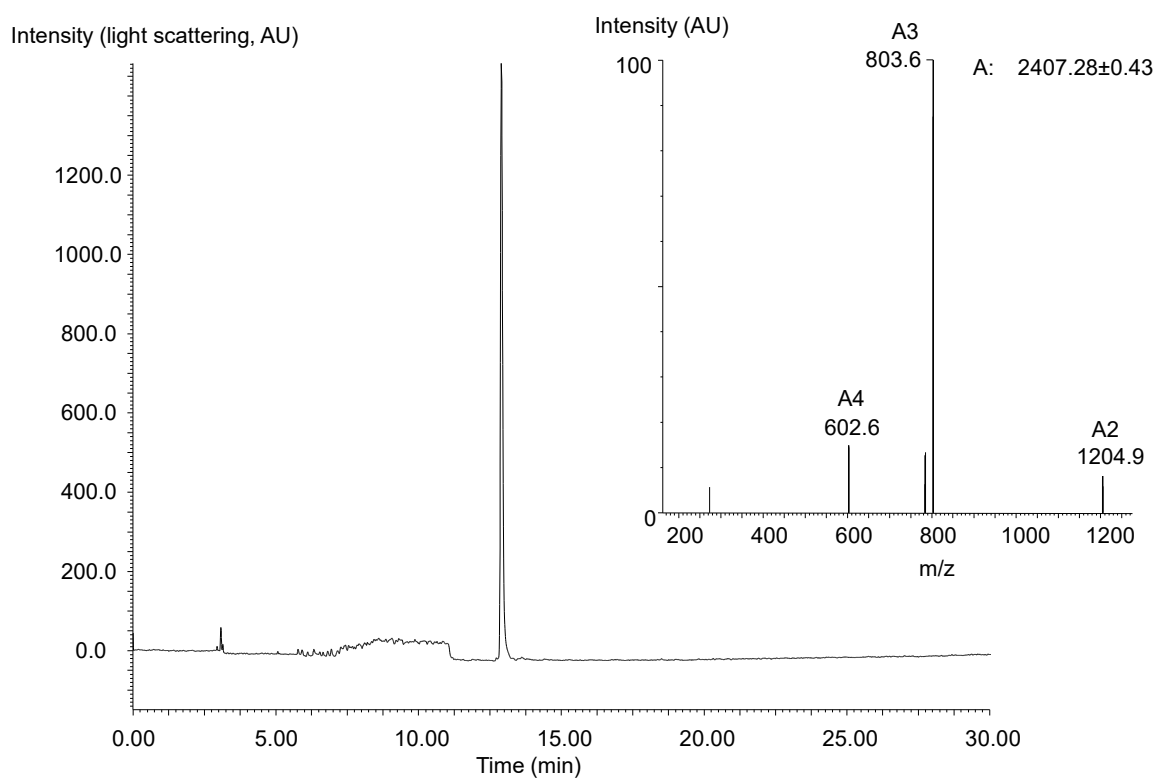
4.2.3 Preparation of SeEA^{off} peptide **22**

TCEP-HCl (2.89 mg, 0.1 mmol), selenide compound **7a** (3.42 mg, 0.01 mmol, 5 eq) and metallic selenium (0.237 mg, 1.5 equiv) were dissolved in 0.2 M pH 4.2 sodium acetate buffer (1 mL). NaOH (5 M) was then added to adjust the pH to 5-5.5.

Peptide **21** (5.65 mg, 1.94 μ mol) was dissolved in the above solution (972 μ L, final peptide concentration 2 mM). The reaction mixture was shaken at 37°C under nitrogen atmosphere and monitored by LC-MS.

After 20 h of exchange, DMSO (105.8 μ L) was added (10% of DMSO by vol) to inactivate the SeEA group by formation of a diselenide bond. The reaction mixture was stirred 30 min at 37°C. Then, acetic acid (275.7 μ L) was added to have a final 20% acetic acid solution. 0.43 M hydroxylamine solution in water (45 μ L) was then added to remove AcA protecting group. The reaction medium was heated at 37°C for 1 h 30 min and then diluted with water (11 mL final volume). The aqueous solution was extracted with diethyl ether (3 x 3 mL) and then purified by RP-HPLC (eluent A = water containing 0.1% TFA, eluent B=acetonitrile in water 4/1 containing 0.1% TFA, 30°C, detection at 215 nm, 6 mL/min, 0 to 20% eluent B in 5 min, then 20 to 40% eluent B in 60 min, C18XBridge column) to give 2.4 mg of SeEA^{off} peptide **22** (40%).

A)



B)

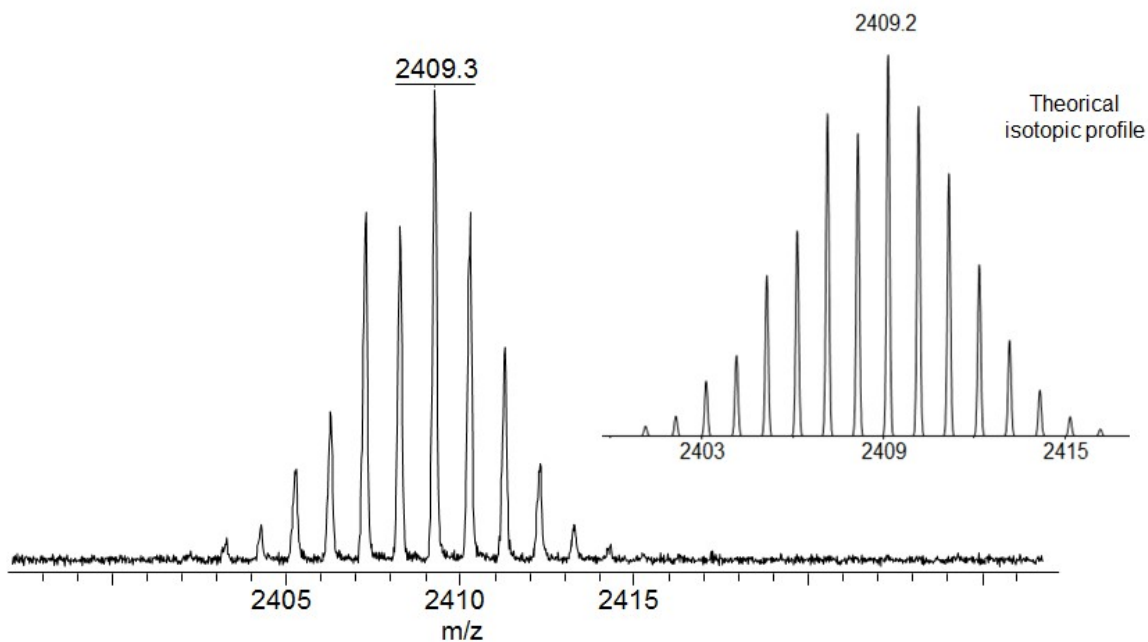


Figure S28. A) LC-MS analysis of peptide **22**. LC trace, eluent A 0.10% TFA in water, eluent B 0.10% TFA in CH₃CN/water: 4/1 by vol. C18 Xbridge BEH 300 Å 5 μm (4.6 x 250 mm) column, gradient 0-100% B in 30 min (1 mL/min, detection 215 nm). B) MALDI-TOF analysis of selenoester peptide **22**. Matrix, 2,5-dihydroxybenzoic acid (DHB) positive mode.

4.3 One-pot assembly of peptide 23

4.3.1 Step 1

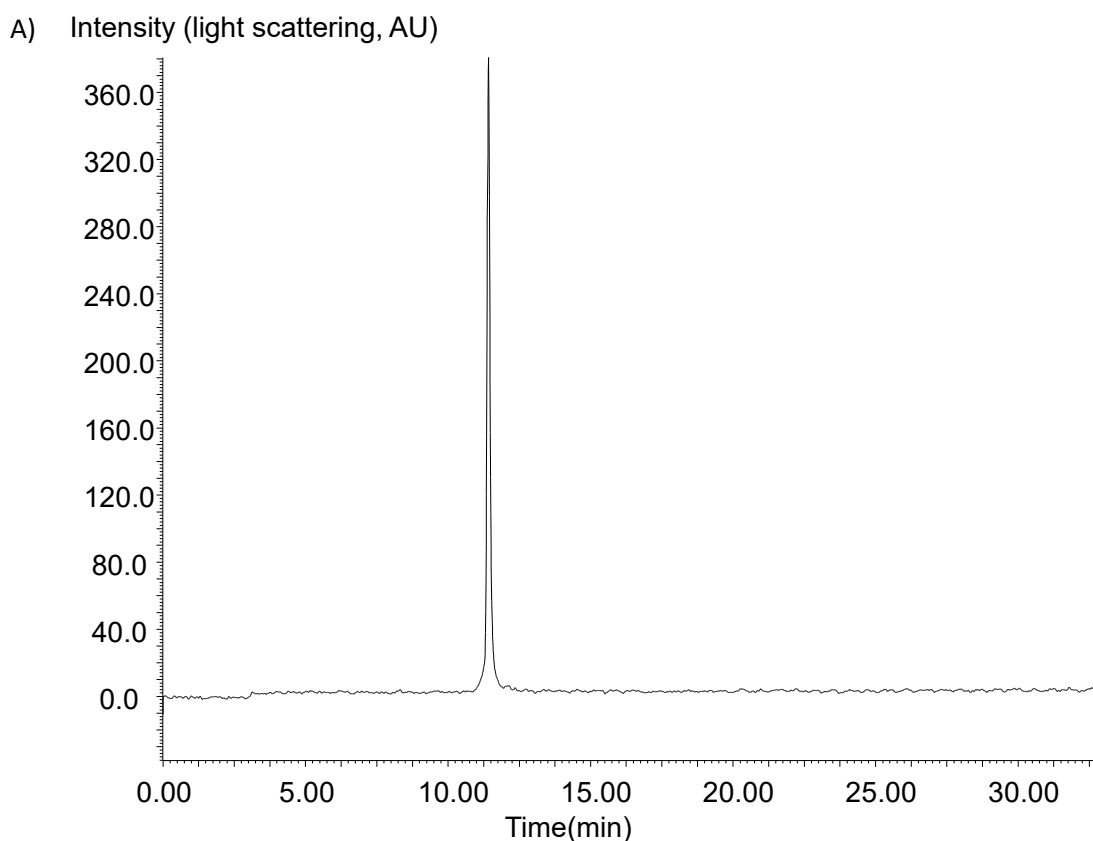
MPAA (3.34 mg, 0.2 mmol) and Se=TCEP (6.58 mg, 0.02 mmol) were dissolved in 6 M Gdn.HCl, 0.1 M pH 7.2 sodium phosphate buffer (100 μ L). NaOH (5 M) was then added to adjust the pH to 7-7.5.

Peptide **19** (5.27 mg, 0.53 μ mol) and peptide **20** (2.51 mg, 0.53 μ mol) were dissolved in the above solution (75 μ L, final peptide concentration 7 mM). The reaction mixture was shaken at 37 °C under nitrogen atmosphere and monitored by LC-MS.

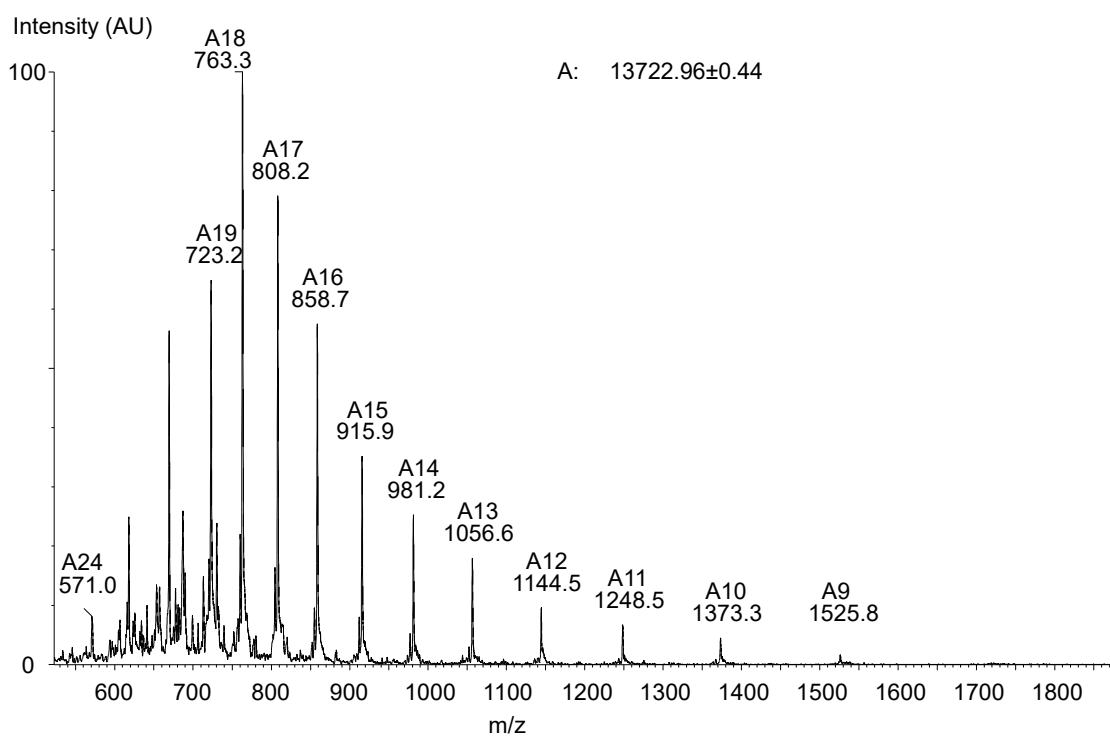
4.3.2 Step 2

After 20 h, the reaction mixture was divided in two equal portions. To one of these, DTT (0.59 mg, 100 mM) and SeEA^{off} peptide **22** (1.04 mg, 0.3 μ mol) were added in this order. The reaction mixture was shaken at 37 °C under nitrogen atmosphere and monitored by LC-MS.

After 42 h, the reaction mixture was diluted with 5% aqueous acetic acid (4 mL), extracted with Et₂O (5 x 2 mL) to remove MPAA and purified by RP-HPLC (eluent A = water containing 0.1% TFA, eluent B=acetonitrile in water 4/1 containing 0.1% TFA, 50 °C, detection at 215 nm, 6 mL/min, 0 to 20% eluent B in 5 min, then 20 to 50% eluent B in 90 min, C18 XBridge column) to give 0.94 mg of SeEA^{off} peptide **23** (21% overall starting from peptide **19**).



B)



C)

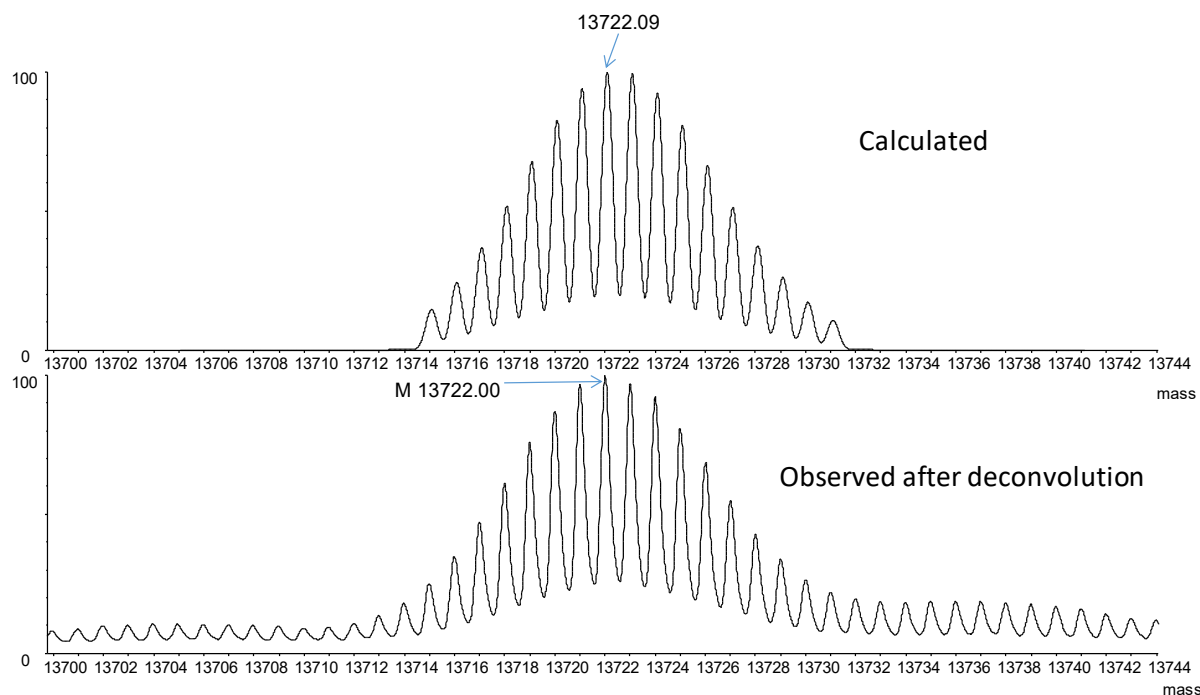


Figure S29. Analysis of thioester peptide **23**. A) LC trace, eluent C 0.1% FA in water, eluent D 0.1% FA in CH₃CN/water: 4/1 by vol. C3 Zorbax 300SB 3.5 μ m (4.6 x 150 mm) column, gradient 0-100% D in 30 min (1 mL/min, detection 215 nm). B) ESI+ MS trace: M calculated (mean) 13722.89, found 13722.96. C) High resolution ESI+ MS with experimental and calculated profile after deconvolution.

4.4 One-pot synthesis of NK1-B

TCEP.HCl (2.87 mg, 0.01 mmol), MPAA (1.29 mg, 0.01 mmol) and Se=TCEP (2.63 mg, 0.008 mmol) were dissolved in 6 M guanidine-HCl, 0.1 M pH 7.2 sodium phosphate buffer (100 μ L). NaOH (5 M) was then added to adjust the pH to 7-7.5.

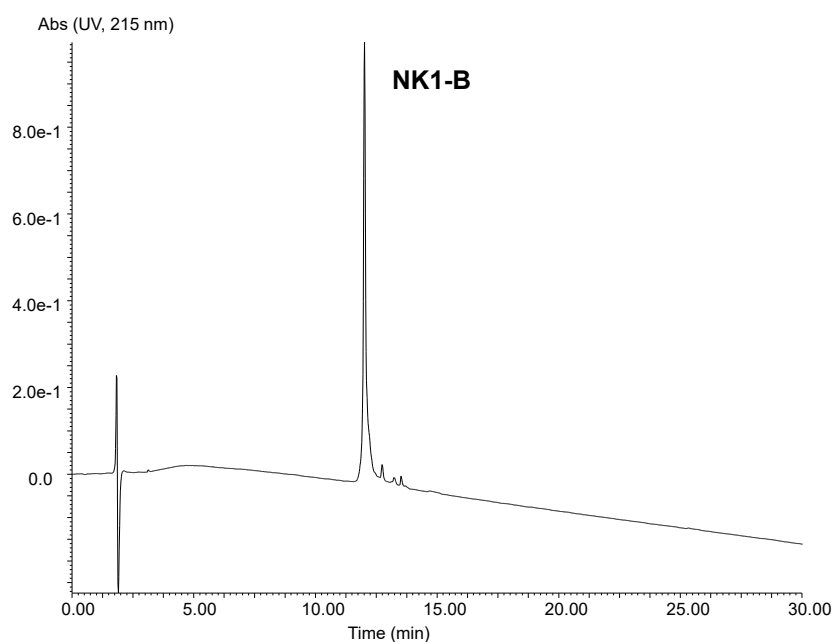
Peptide **23** (0.82 mg, 0.047 μ mol) and peptide **24** (0.23 mg, 0.057 μ mol, 1.2 equiv) were dissolved in the above solution (12 μ L, final peptide concentration 4 mM). The reaction mixture was shaken at 37°C under nitrogen atmosphere and monitored by LC-MS.

After 18 h, TCEP.HCl (2.87 mg, 0.01 mmol) and MPAA (1.29 mg, 0.01 mmol) were dissolved in 6 M Gdn.HCl, 0.1 M pH 7.2 sodium phosphate buffer (100 μ L). NaOH (5 M) was then added to adjust the pH to 5.

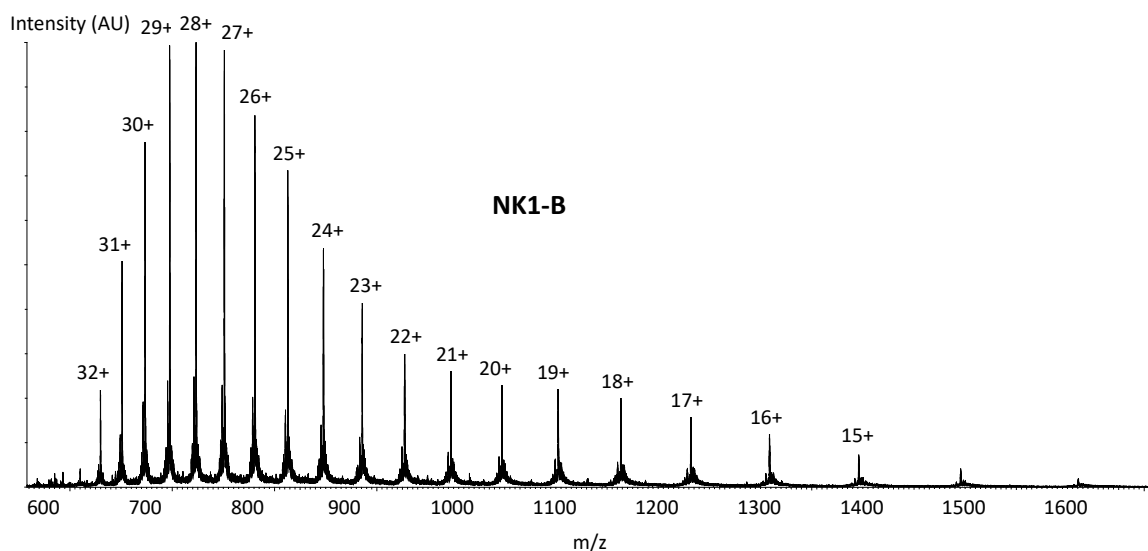
Peptide **25** (0.43 mg, 0.095 μ mol, 2 equiv) was dissolved in the above solution (12 μ L) and then added to the reaction mixture which was shaken at 37°C under nitrogen atmosphere.

After 72 h, the reaction mixture was diluted with 5% aqueous acetic acid (4 mL), extracted with Et₂O (3 x 2 mL) to remove MPAA and purified by RP-HPLC (eluent C = water containing 0.1% formic acid (FA), eluent D=acetonitrile in water 4/1 containing 0.1 %FA, 50 °C, detection at 215 nm, 6 mL/min, 0 to 20% eluent B in 5 min, then 20 to 50% eluent B in 90 min, C3 Zorbax column) to give 335 μ g of **NK1-B** (32% overall starting from peptide **23**).

A)



B)



C)

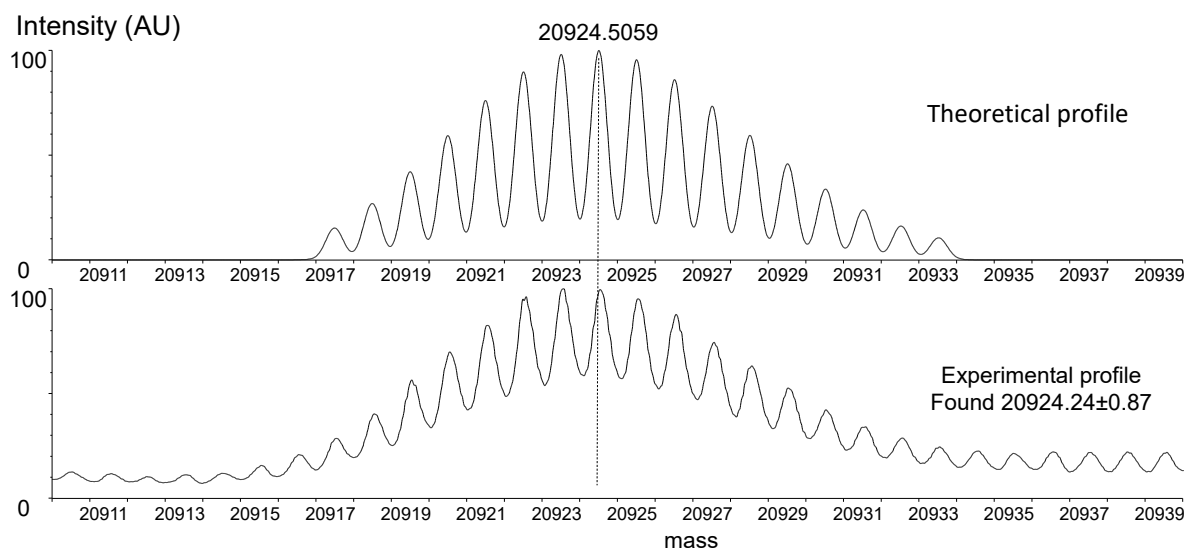


Figure S30. Analysis of **NK1-B**. A) LC trace, eluent A 0.1% formic acid (FA) in water, eluent B 0.1% FA in CH₃CN/water: 4/1 by vol. C3 Zorbax 300SB 3.5 μ m (4.6 x 150 mm) column, gradient 0-100% B in 30 min (1 mL/min, detection 215 nm). B) ESI+ MS high resolution spectrum. C) Deconvoluted spectrum and comparison with the theoretical profile.

5. Computational analysis

Quantum chemical calculations were performed using the Gaussian 09 package of programs⁸. DFT computations were carried out using the B3LYP hybrid functional employing the 6-31+G* basis set with 5 pure d functions. Gradient techniques using internal coordinates with very tight optimization convergence criteria (each component of the first energy derivative below 2.0×10^{-6} Hartree/Bohr or radian) were used for both geometry optimization and computation of vibrational properties. The transition states were localized using the Newton-Raphson algorithm, and the nature of the stationary

points was determined by analysis of the Hessian. The activation and reaction energies were calculated from the thermochemical output (298.150 Kelvin, 1 atm) computed for the reagents, transition states and products, using standard thermochemistry as implemented in Gaussian 09. Intrinsic reaction coordinate (IRC) calculations were performed in the gas phase to localize the nearest local minima on the reactant and product sides of the reaction coordinate⁹. Solvent effects (water) were taken into account using Tomasi's polarizable continuum model (PCM)¹⁰.

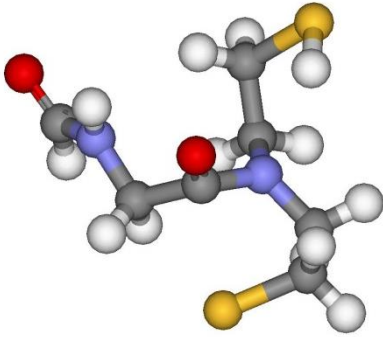
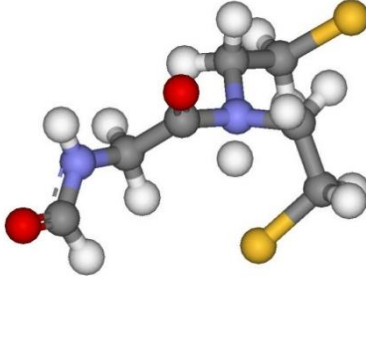
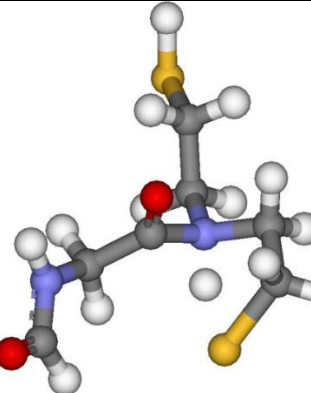
TS SEA	TS SEA_2	TS SEA_3
		
H = -1330.731085 Hartree G = -1330.789404 Hartree	H = -1330.703070 Hartree G = -1330.763103 Hartree	H = -1331.221528 Hartree G = -1331.281239 Hartree

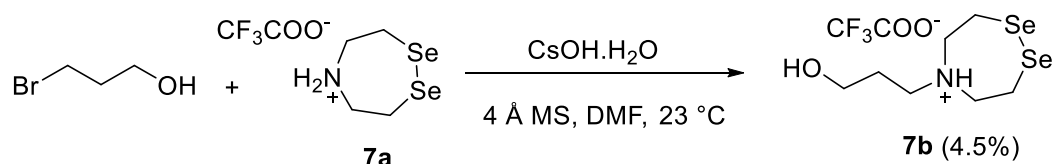
Figure S31. Structures and absolute energies for anionic (**TS SEA**, **TS SEA_2**) and neutral (**TS SEA_3**) transition states (gas phase). **TS SEA** is discussed in the main manuscript (one neutral 2-mercaptoethyl limb protonates the amide nitrogen, while the other is anionic and attacks the amide carbonyl). In the other anionic transition state **TS SEA_2**, one neutral 2-mercaptoethyl limb protonates the amide nitrogen, while the other is anionic and spectator. The third transition state **TS SEA_3** is neutral (one neutral 2-mercaptoethyl limb protonates the amide nitrogen, while the other is neutral too and spectator). The activation barrier for **TS SEA** is about 20 kcal mol⁻¹ less than for **TS SEA_2** and **TS SEA_3**.

Catalysis of SEA-thiol exchange and SEA ligation by a water soluble alkylselenenol catalysts (Chapter V)

1. Synthesis of diselenide catalysts

Diselenide **7a** has been described in the previous chapter (section 1).

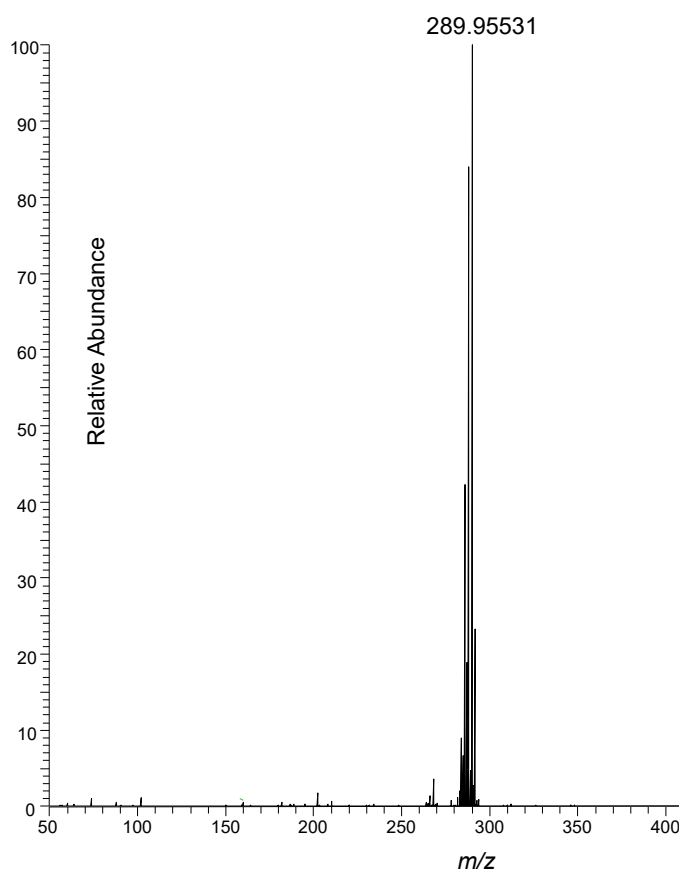
1.1 Synthesis of *N*-(3-hydroxypropyl) bis(2-selenylethyl)amine **7b**



Scheme S2. Synthesis of *N*-alkyl diselenide **7b**.

Cesium hydroxide monohydrate (509 mg, 3 mmol, 2 equiv) was added to activated powdered 4 Å molecular sieves (400 mg) in anhydrous *N,N*-dimethylformamide (1 mL). The white suspension was vigorously stirred for 10 min at rt. Then, the bis(2-selenylethyl)amine trifluoroacetate salt **7a** (350 mg, 1.5 mmol) was dissolved in DMF (2.5 mL) and added to the reaction mixture. The reaction was stirred for 30 min. Then, 3-bromopropanol (402 μL, 4.55 mmol, 3 equiv) was added to the suspension and the mixture was stirred for 20 h at rt.

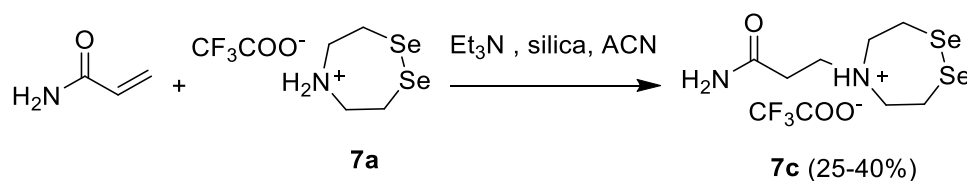
Finally, the reaction mixture was diluted with water containing 0.1% TFA (20 mL) and centrifuged. The supernatant was collected, filtered and purified by RP-HPLC (XBridge C18 OBD column, eluent A : water containing 0.1% TFA by vol, eluent B: water/acetonitrile : ¼ by vol containing 0.1% TFA by vol, detection at 215 nm, flow rate 25 mL/min, 0 - 15% eluent B in 10 min). The purified fractions were collected, frozen and lyophilized to give 19.53 mg (4.5%) of **7b** as the trifluoroacetate salt.



m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	RDB equiv.	Composition
285,95789	7890467	44,41	285,95838	-1,72	0,5	C7 H16 O N [76]Se Se
286,95883	3286462	18,5	286,95908	-0,85	0,5	C7 H16 O N [77]Se Se
287,95615	14460140	81,39	287,95647	-1,13	0,5	C7 H16 O N [78]Se Se
289,9552	17765790	100	289,95568	-1,67	0,5	C7 H16 O N Se2
291,95529	4317527	24,3	291,95586	-1,95	0,5	C7 H16 O N Se [82]Se

Figure S32. HR-MS analysis for purified **7b**. Calcd. for $[M+H]^+$: 289.955, observed mass: 289.955 (monoisotopic).

1.2 Synthesis of *N*-(2-carboxamidoethyl) bis(2-selenylethyl)amine **7c**



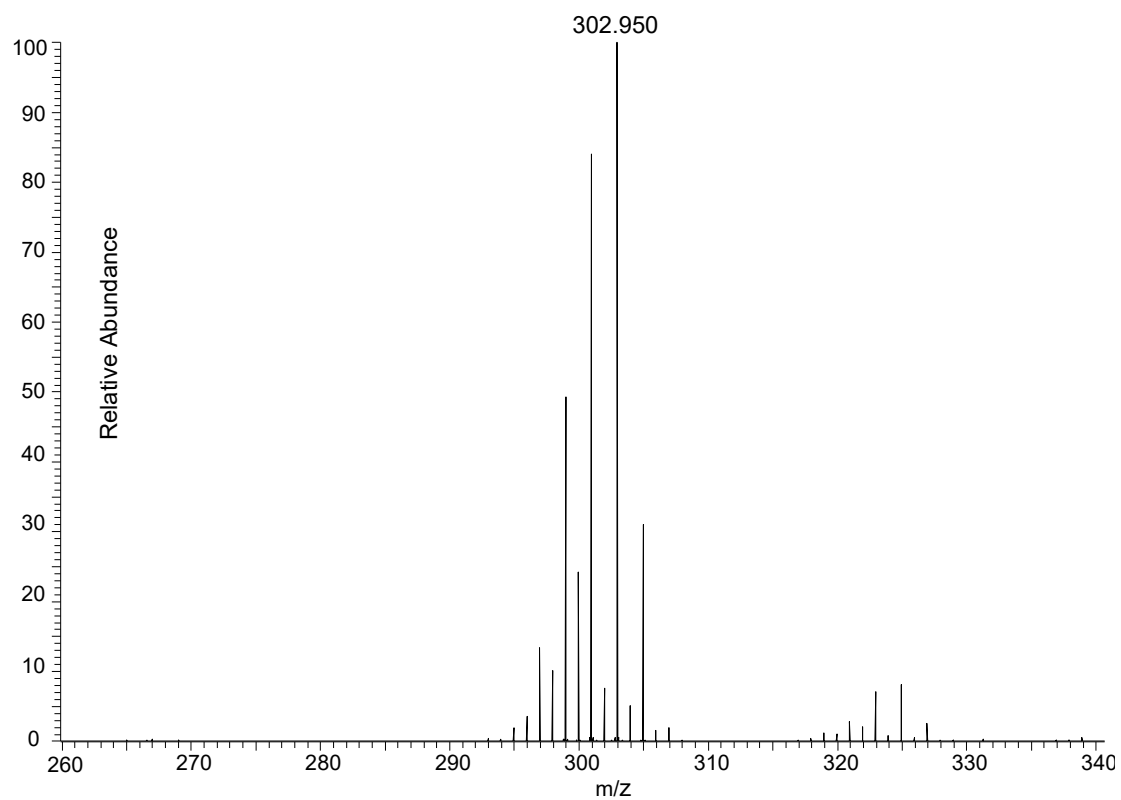
Scheme S3. Synthesis of *N*-(2-carboxamidoethyl) bis(2-selenylethyl)amine **7c**.

The bis(2-selenylethyl)amine trifluoroacetate salt **7a** (43.2 mg, 0.125 mmol, 1 equiv) and acrylamide (44.4 mg, 0.625 mmol, 5 equiv) were dissolved separately in acetonitrile (1 mL for each). The two

solutions were combined and silica gel (50% w/w relative to compound **7a**) was added to the mixture. Triethylamine (17 μ L, 0.125 mmol, 1 equiv) was then added to the reaction. Note that the *bis*(2-selenylethyl)amine **7a** precipitates partially upon Et₃N addition but after few time the precipitate disappears probably due to adsorption of **7a** on the silica gel.

The suspension was then stirred at 60 °C and monitored by LC-MS. After 24 h, 5 equiv of acrylamide were added to the reaction mixture. The reaction was stirred at 60 °C for further two days. When the reaction was completed, the supernatant was collected and 2 mL of water-TFA 0.1 % were added to the residue of silica gel. The mixture was stirred for 30 min and the supernatant was collected. This washing procedure was repeated twice. All supernatants were combined and concentrated in vacuo to remove the acetonitrile. This operation must be carefully controlled to avoid evaporation to dryness.

The solution was filtered and purified by RP-HPLC (XBridge C18 OBD column, eluent A : water containing 0.1% TFA by vol, eluent B: water/acetonitrile : ¼ by vol containing 0.1% TFA by vol, detection at 215 nm, flow rate 25 mL min, 0 - 15% eluent B in 20 min). The purified fractions were collected, frozen and lyophilized to give 30.9 mg (40%) of **7c** as the trifluoroacetate salt.



m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	RDB equiv.	Composition
298,9531	61619380	50,61				
299,9541	30622960	25,15				
300,9513	103122168	84,7				
302,9503	121743136	100	302,9509	-1,98	1,5	C7 H15 O N2 Se2
304,9504	38553024	31,67				

Figure S33. HR-MS analysis for purified **7c**. Calcd. for $[M+H]^+$: 302.951, observed mass: 302.950 (monoisotopic).

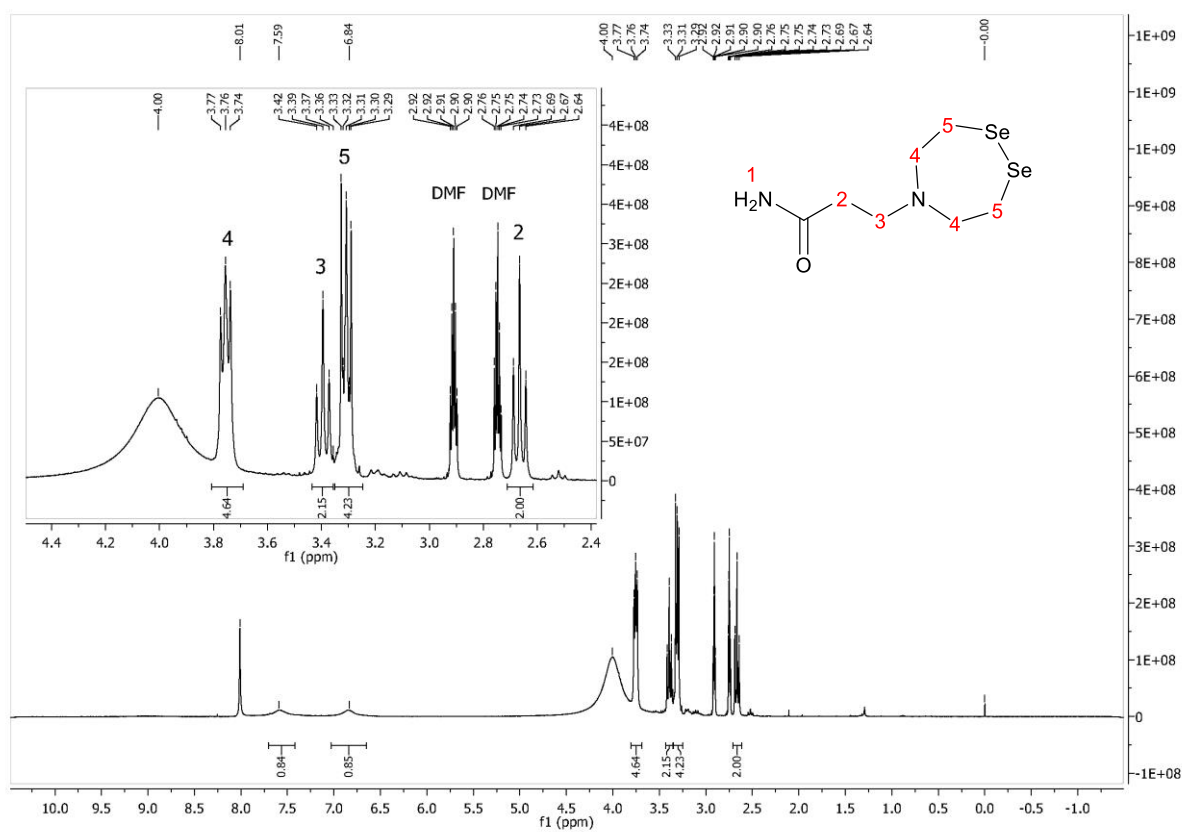


Figure S34. ^1H NMR (300 MHz) spectrum for compound **7c** (DMF-d_7 , 333 K). δ 7.59 (s, 1H), 6.84 (s, 1H), 3.77 – 3.74 (m, 4H), 3.42 – 3.36 (t, 2H), 3.33 – 3.29 (m, 4H), 2.73 – 2.64 (t, 2H).

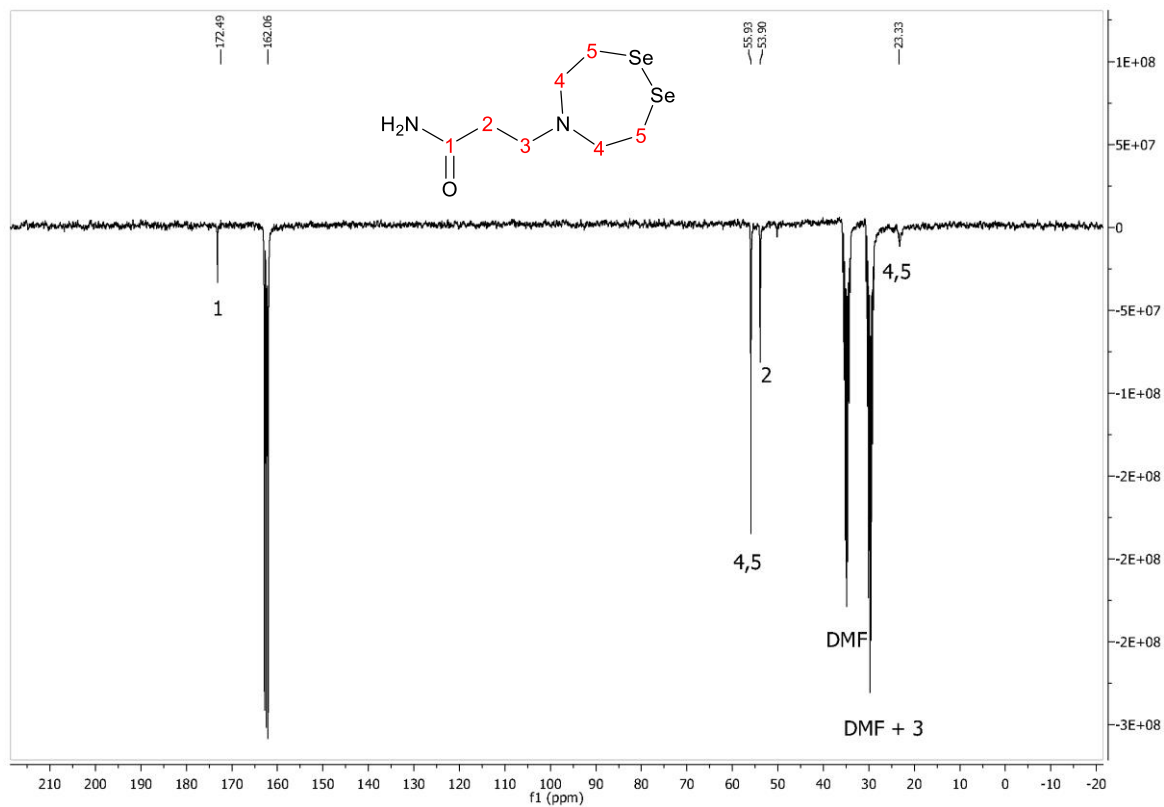


Figure S35. ^{13}C JMOD (75 MHz) spectrum for compound **7c** (DMF-d₇, 291 K). δ (ppm) = 172.50 (m), 55.93 (s), 53.90 (s), 23.33 (s).

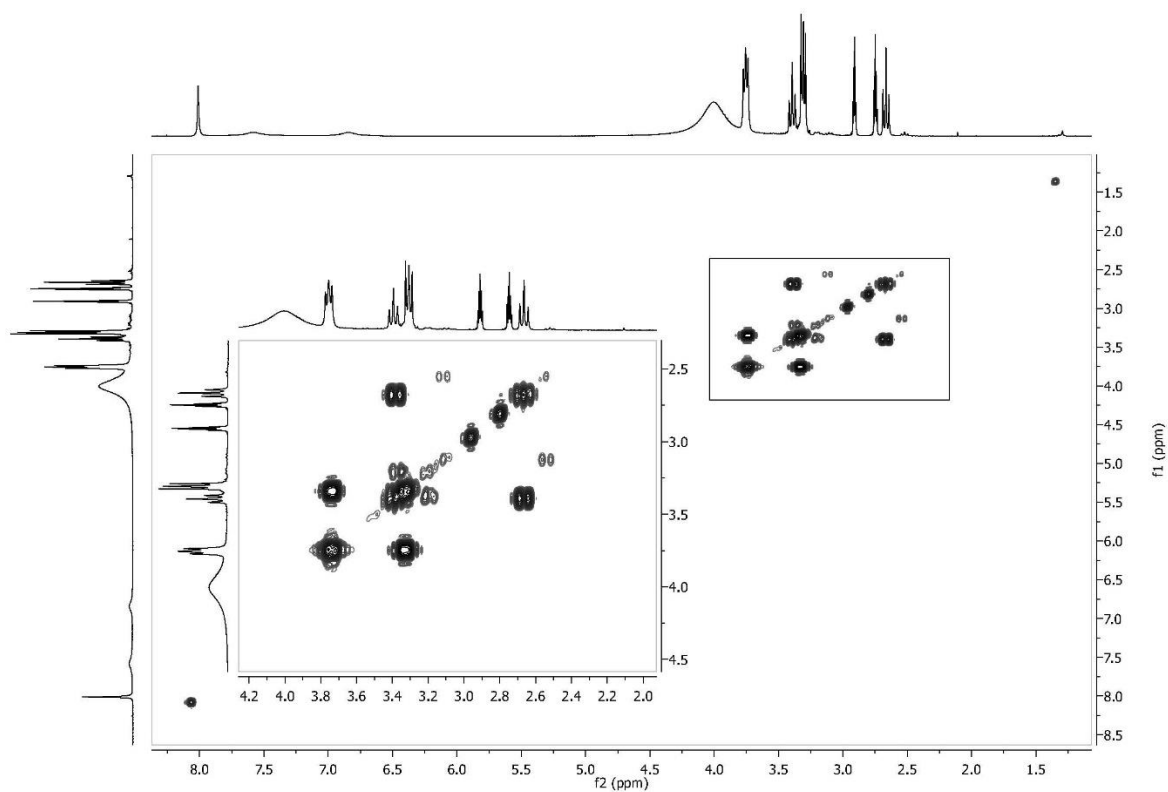


Figure S36. ^1H - ^1H COSY spectrum for compound **7c** (DMF-d₇, 333 K).

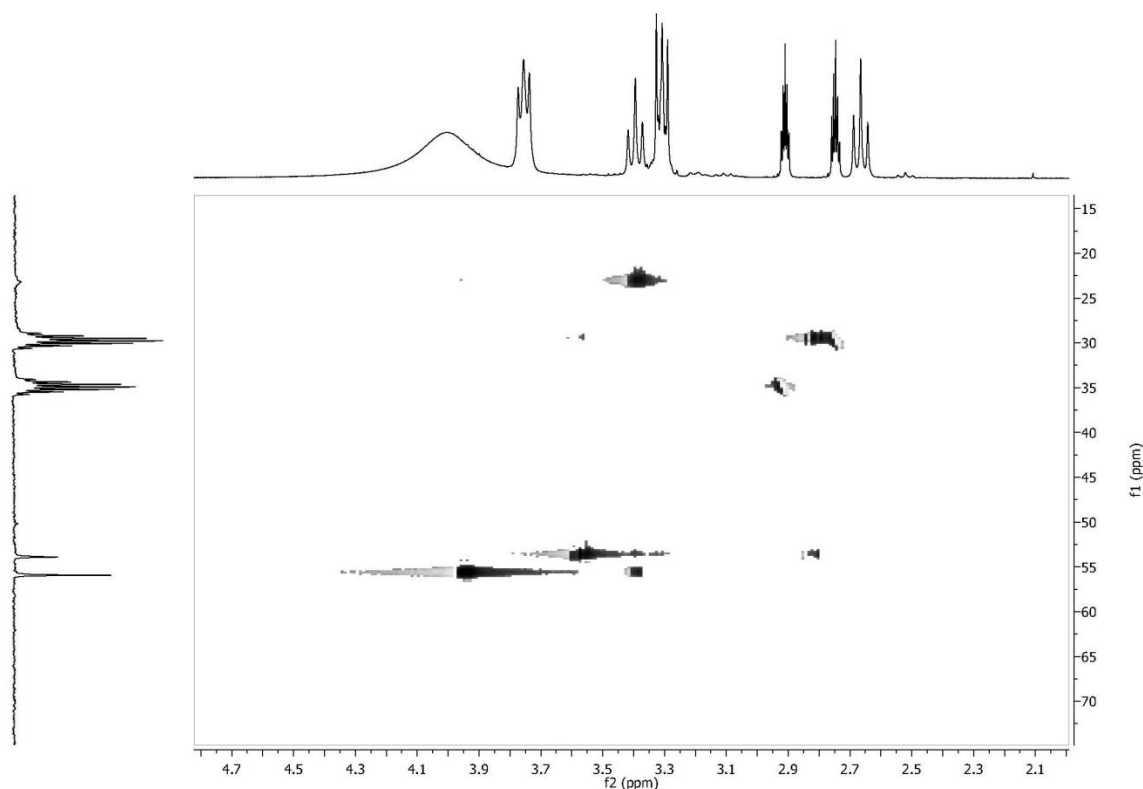
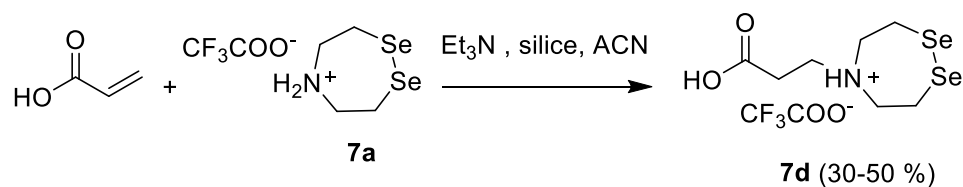


Figure S37. ^1H - ^{13}C HSQC spectrum for compound **7c** (DMF- d_7 , 291K).

1.3 Synthesis of *N*-(2-carboxyethyl) *bis*(2-selenylethyl)amine **7d**

1.3.1 First strategy



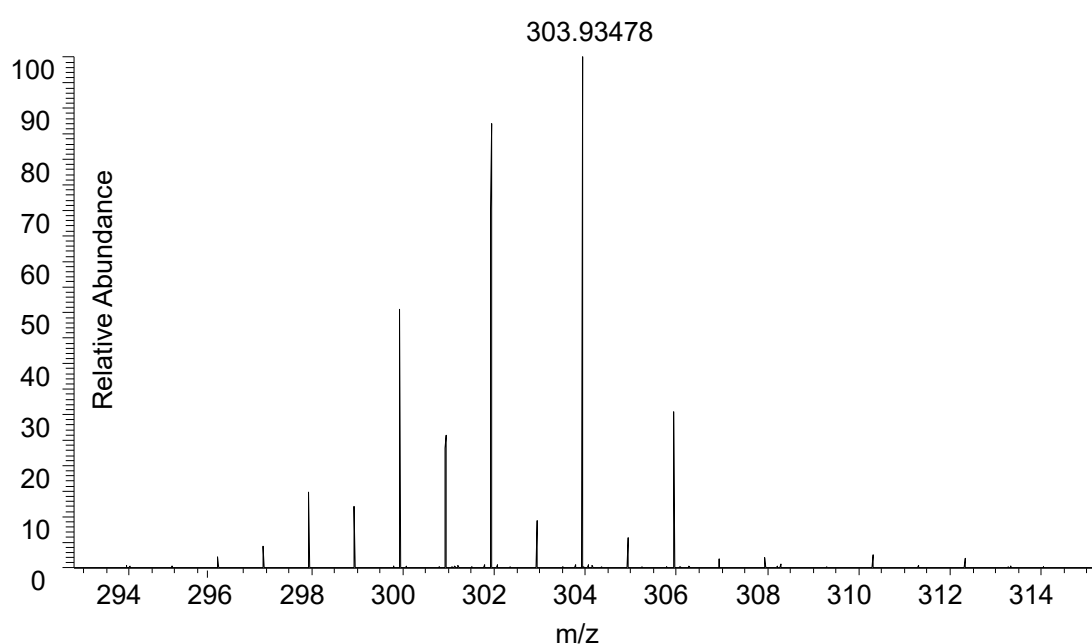
Scheme S4. Synthesis of *N*-(2-carboxyethyl) *bis*(2-selenylethyl)amine **7d**.

The *bis*(2-selenylethyl)amine trifluoroacetate salt **3a** (53.5 mg, 0.155 mmol, 1 equiv) and acrylic acid (53 μL , 0.733 mmol, 5 equiv) were dissolved separately in acetonitrile (2.5 mL for each). The two solutions were combined and silica gel (50 % w/w relative to amine **3a**) was added to the mixture. Triethylamine (125 μL , 0.93 mmol, 6 equiv) was then added to the reaction. Note that the *bis*(2-selenylethyl)amine **3a** precipitates partially upon Et_3N addition but after a few time the precipitate disappears probably due to adsorption of **3a** on the silica gel.

The solution was then stirred at 60 $^\circ\text{C}$ and monitored by LC-MS. After 24 h, 5 equiv of acrylamide and 6 equiv of Et_3N were added again to the reaction mixture. The reaction was further stirred at 60 $^\circ\text{C}$ for 24 h. When the reaction was completed, the supernatant was collected then 2 mL of water were added

to the residue of silica gel. The reaction was stirred for 30 min and then the supernatant was collected. This operation was repeated twice. All supernatants were combined and the resulting solution was diluted with water (6 mL). Then, the pH was adjusted to 4-5 by adding water containing 10% TFA (100 μ L). The solution was concentrated in vacuo to remove the acetonitrile. This operation must be carefully controlled to avoid evaporation to dryness.

The solution was filtered and purified by RP-HPLC (XBridge C18 OBD column, eluent A: water containing 0.1% TFA by vol, eluent B: water/acetonitrile : ¼ by vol containing 0.1% TFA by vol, detection at 215 nm, flow rate 25 mL/min, 0 - 15% eluent B in 20 min). The purified fractions were collected, frozen and lyophilized to give 26 mg (40%) of **7d** as the trifluoroacetate salt.



m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	RDB equiv.	Composition
299,93747	2139153,3	52,61				
300,93848	1105033	27,18				
301,93576	3534280,3	86,92				
303,93481	4066237,8	100	303,93495	-0,13	1,5	C7 H14 O2 N Se2
305,93481	1315523,8	32,35				

Figure S38. HR-MS analysis for purified **7d**. Calcd. for $[M+H]^+$: 303.934, observed mass: 303.934 (monoisotopic).

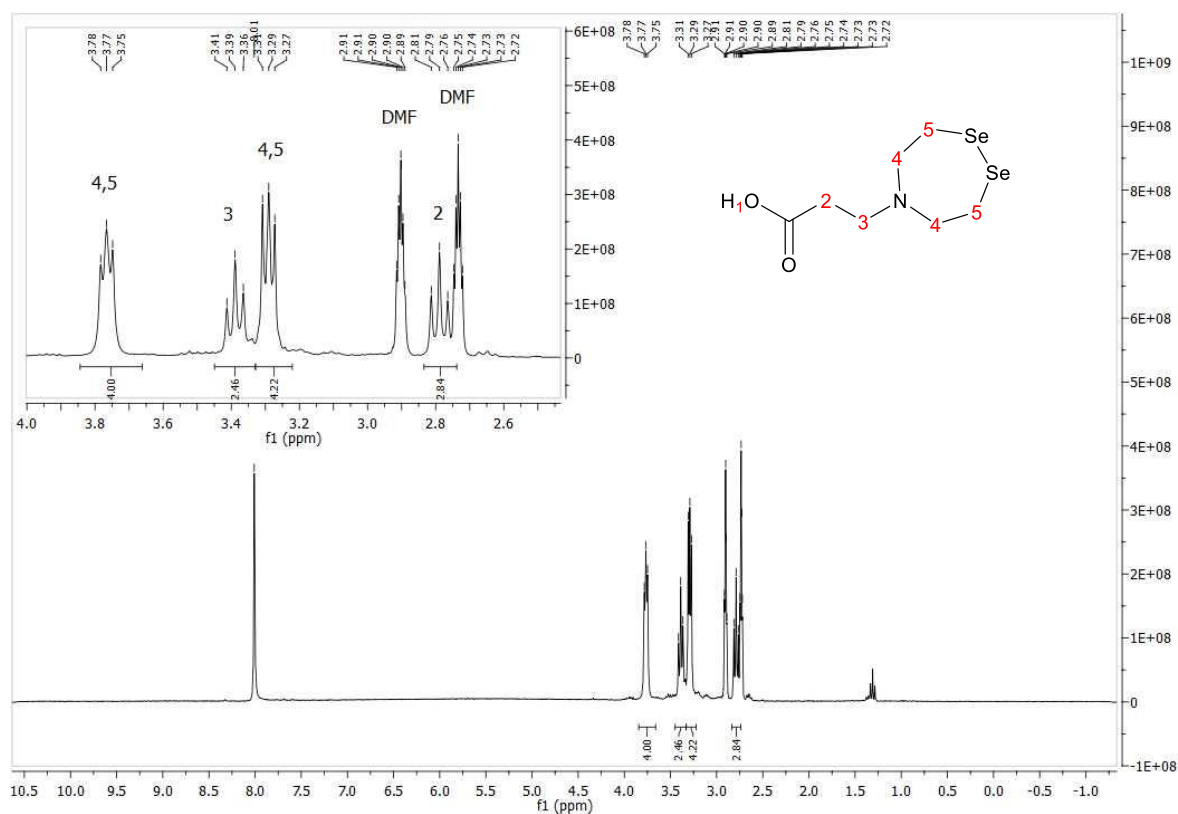


Figure S39. ^1H NMR (300 MHz) spectrum for compound **7d** (DMF- d_7 , 310 K). δ 8.01 (s, 1H), 6.84 (s, 1H), 3.78 – 3.75 (t, 4H), 3.41 – 3.36 (t, 2H), 3.31 – 3.27 (t, 4H), 2.81 – 2.76 (t, 2H).

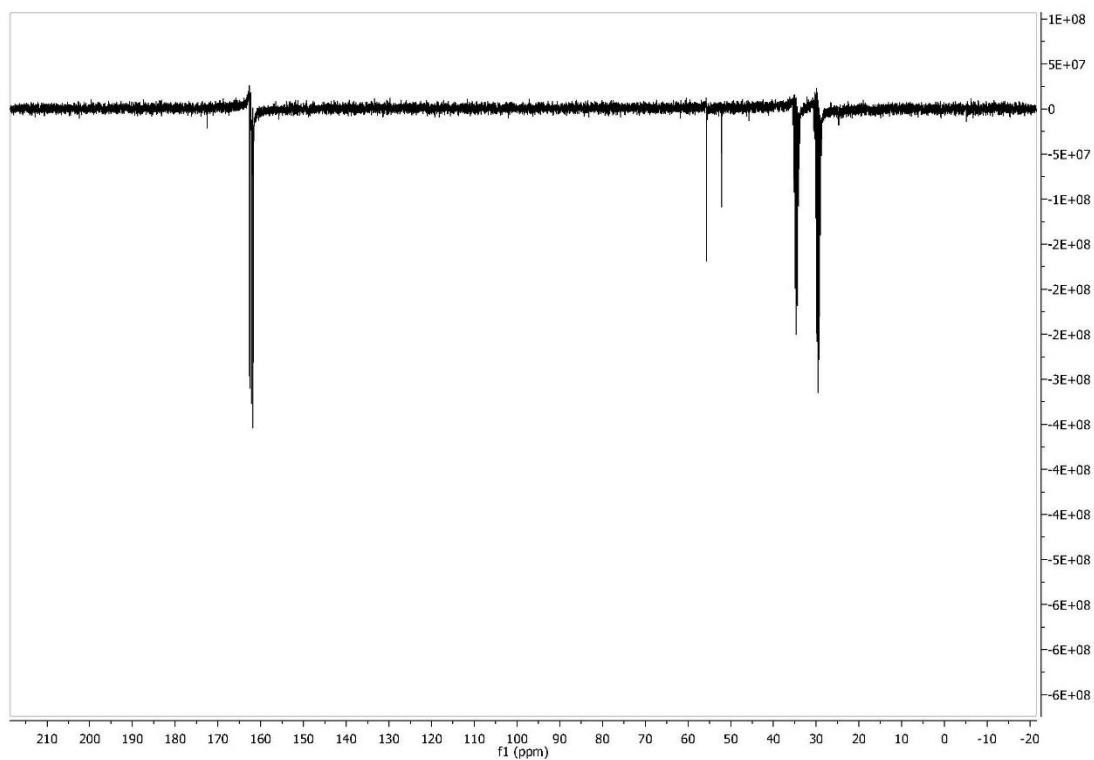


Figure S40. ^{13}C JMOD (75 MHz) spectrum for compound **7d** (DMF- d_7 , 310 K). δ (ppm) = 172.50 (m), 55.66 (s), 52.16 (s).

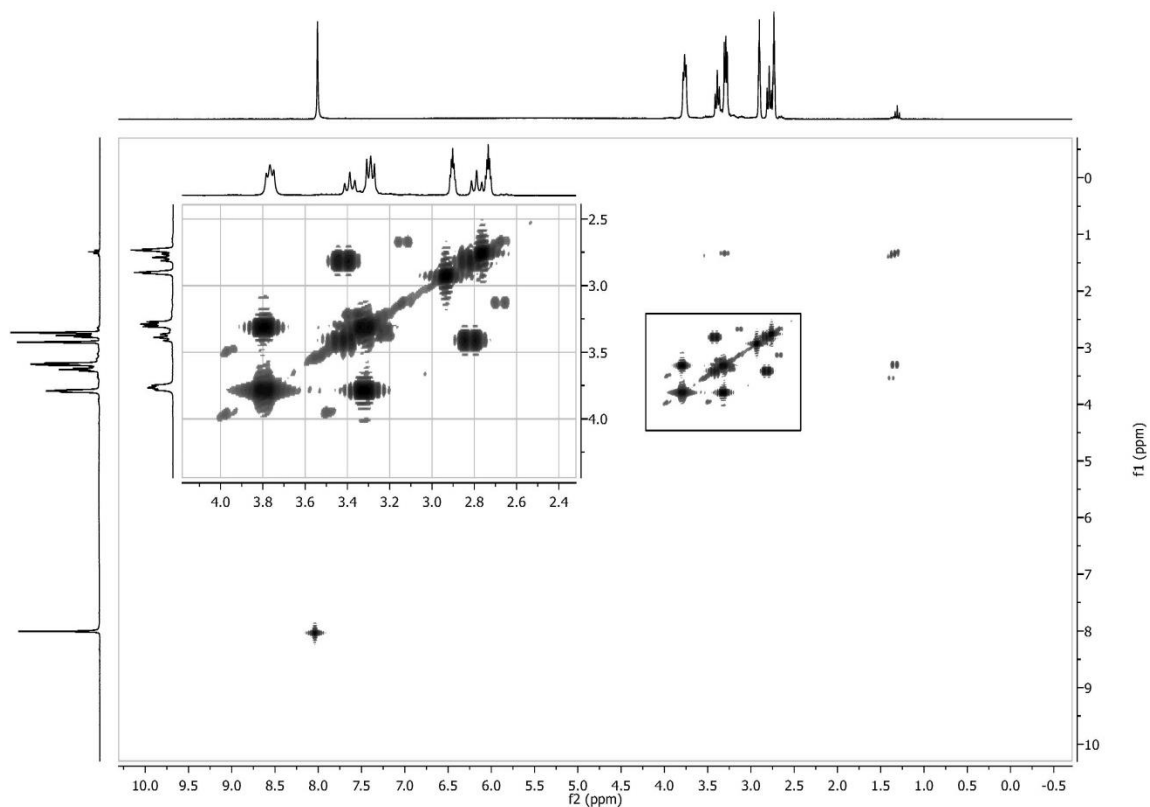


Figure S41. ^1H - ^1H COSY spectrum for compound **7d** (DMF-d₇, 310 K).

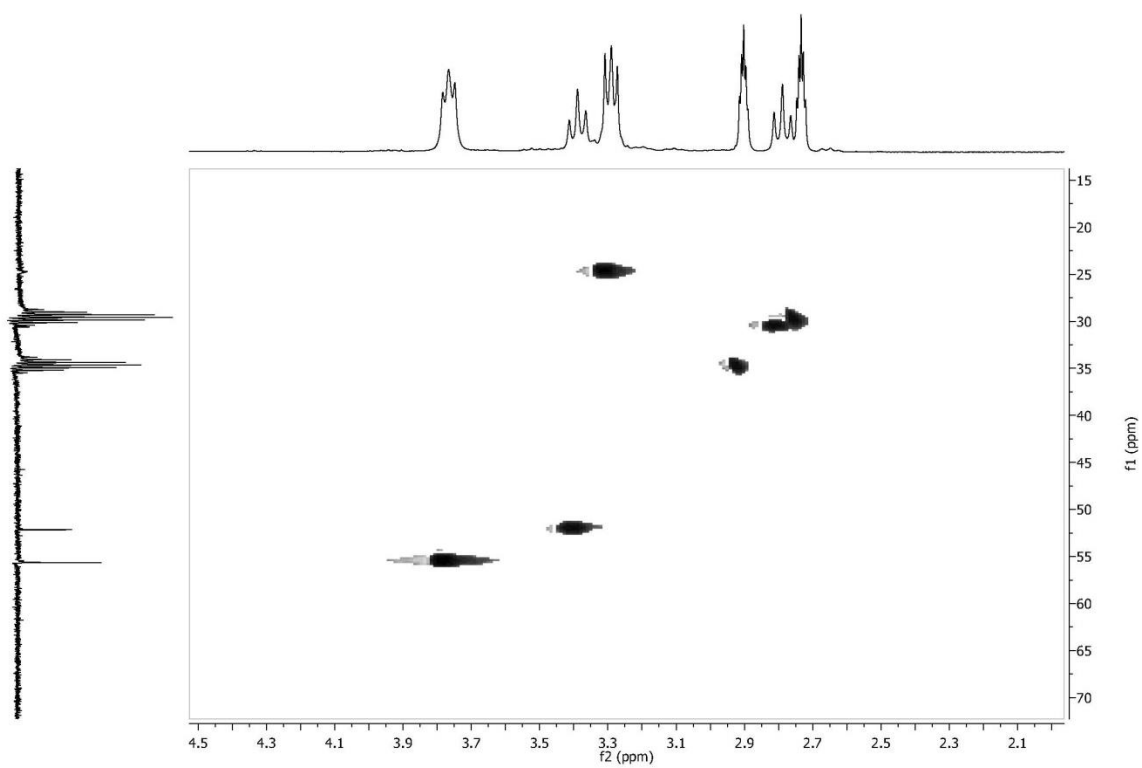
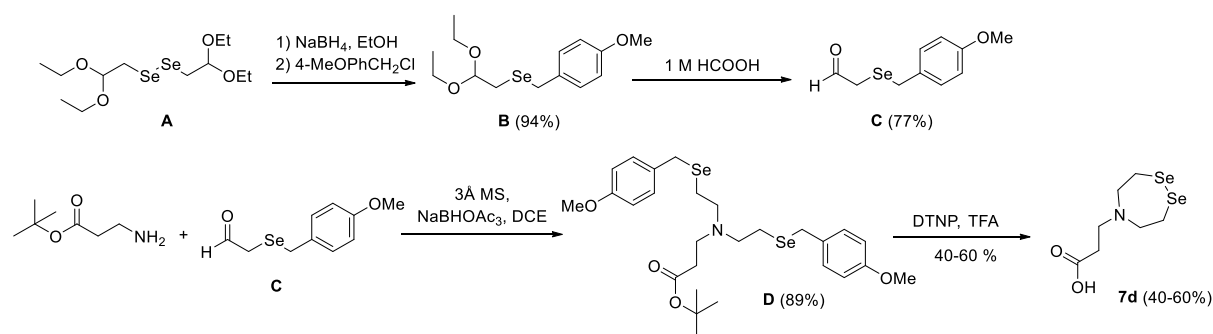


Figure S42. ^1H - ^{13}C HSQC spectrum for compound **7d** (DMF-d₇, 310K).

1.3.2 Second strategy



Scheme S5. Synthesis of *N*-(2-carboxyethyl) bis(2-selenylethyl)amine **7d**.

The synthesis of compound **A** is described in the literature¹¹.

Synthesis of (2,2-diethoxyethyl)(4-methoxybenzyl)selane B

To a solution of diselenide **A** (1.94 g, 4.94 mmol) in ethanol (34 mL) was added portionwise (30 min) at 0 °C NaBH₄ (500 mg, 13.2 mmol, 2.7 equiv). Once the emission of H₂ stopped, the reaction mixture was warmed to 25 °C for 30 min. Then the reaction mixture was cooled to 0 °C and 4-methoxybenzyl chloride (1.24 mL, 9.19 mmol, 1.9 equiv) was added. The reaction mixture was stirred at room temperature for 2 h. Water (150 mL) was added afterwards and the obtained aqueous layer was extracted with Et₂O (2 × 100 mL). The combined organic layers were dried over MgSO₄ and the solvent was evaporated under reduced pressure. Purification of the crude by column chromatography (cyclohexane/ethyl acetate 95:5) provided the expected selenide **B** (2.75 g, 94%) as an oil.

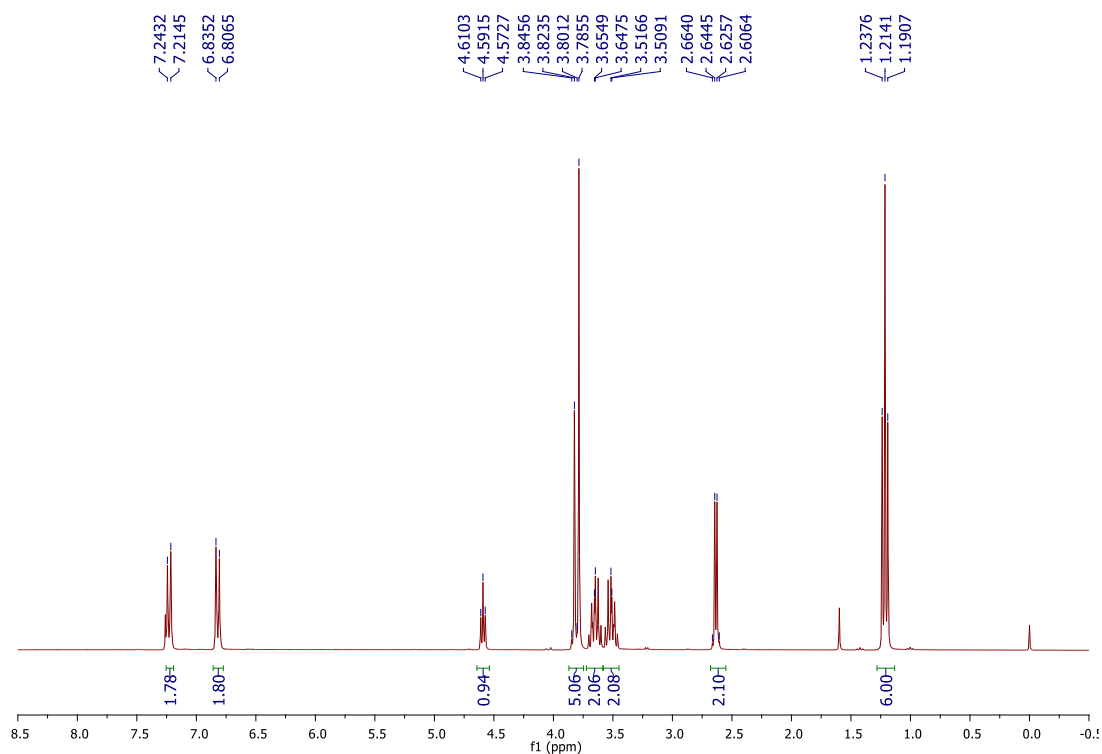


Figure S43. ^1H NMR (300 MHz) spectrum for compound **B** (CDCl_3 , 273 K). δ 7.23 (d, $J = 8.6$ Hz, 2H), 6.82 (d, $J = 8.6$ Hz, 2H), 4.59 (t, $J = 5.6$ Hz, 1H), 3.82 (s, 2H), 3.78 (s, 3H), 3.59-3.71 (m, 2H), 3.45-3.57 (m, 2H), 2.63 (d, $J = 5.6$ Hz, 2H), 1.21 (d, $J = 7.0$ Hz, 6H).

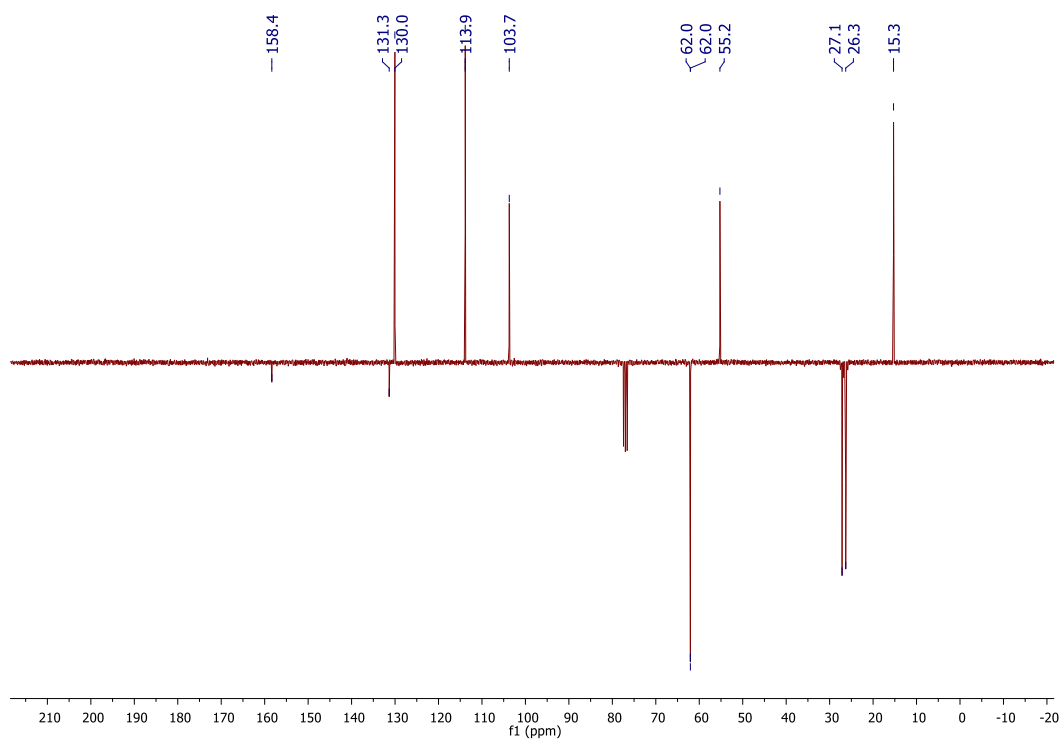


Figure S44. ^{13}C JMOD (75 MHz) spectrum for compound **B** (CDCl_3 , 273 K). δ 158.4 (C), 131.3 (C), 130.0 ($2 \times \text{CH}$), 113.9 ($2 \times \text{CH}$), 103.7 (CH), 62.0 ($2 \times \text{CH}_2$), 55.2 (CH_3), 27.1 (CH_2), 26.3 (CH_2), 15.3 ($2 \times \text{CH}_3$) ppm.

Synthesis of 2-((4-methoxybenzyl)selenyl)acetaldehyde **C**

The selenide **B** (1.24 g, 3.90 mmol) was suspended in formic acid (1M, 16 mL). The reaction mixture was stirred at 50 °C overnight. Water (50 mL) was added and the obtained aqueous layer was extracted with Et₂O (2 × 50 mL). The combined organic layers were dried over MgSO₄ and the solvent was evaporated under reduced pressure. The compound **C** is directly used for the next step without further purification because isn't very stable. However, to characterized it, we performed a purification by column chromatography (cyclohexane/ethyl acetate 95:5) provided the expected compound **C** (730 mg, 77%).

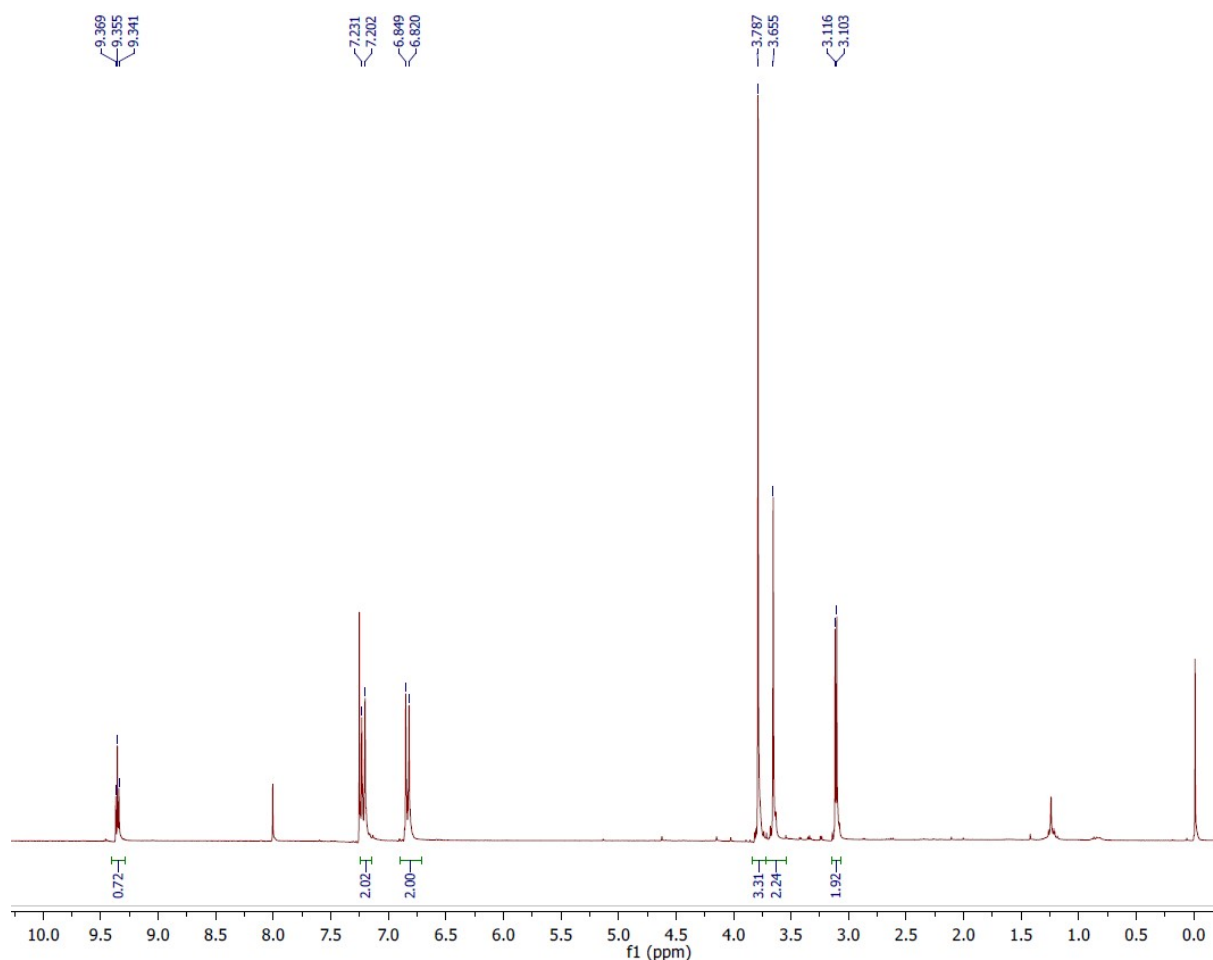


Figure S45. ¹H NMR (300 MHz) spectrum for compound **C** (CDCl₃, 273 K). δ 9.35 (t, *J* = 4.2 Hz, 1H), 7.21 (d, *J* = 8.67 Hz, 2H), 6.83 (d, *J* = 8.67 Hz, 2H), 3.78 (s, 3H), 3.65 (s, 2H), 3.11 (d, *J* = 3.9 Hz, 2H).

Synthesis of tert-butyl 3-(bis(2-((4-methoxybenzyl)selenyl)ethyl)amino)propanoate **D**

Compound **C** (730 mg, 3.90 mmol, 3 equiv) was added to activated powdered 3 Å molecular sieves (1 g) in anhydrous 1,2-dichloroethane (8 mL) under argon atmosphere. Then, sodium triacetoxyborohydride (700 mg, 4.3 mmol, 3.3 equiv) and β-Alanine *t*-butyl ester hydrochloride (236 mg, 1.3 mmol) dissolved in DCE (5 mL) were added. The reaction mixture was stirred at r.t under argon

atmosphere. After 24 h, the reaction mixture was filtrated on Büchner funnel and washed with DCM. The solvent was evaporated under reduced pressure. Then, 10 % K_2CO_3 (25 mL) and brine (30 mL) were added and the obtained aqueous layer was extracted with DCM (3 × 50 mL). The combined organic layers were dried over $MgSO_4$ and the solvent was evaporated under reduced pressure. Purification of the crude by column chromatography (cyclohexane/ethyl acetate 95:5) provided the expected diselenide **D** (695 mg, 89%).

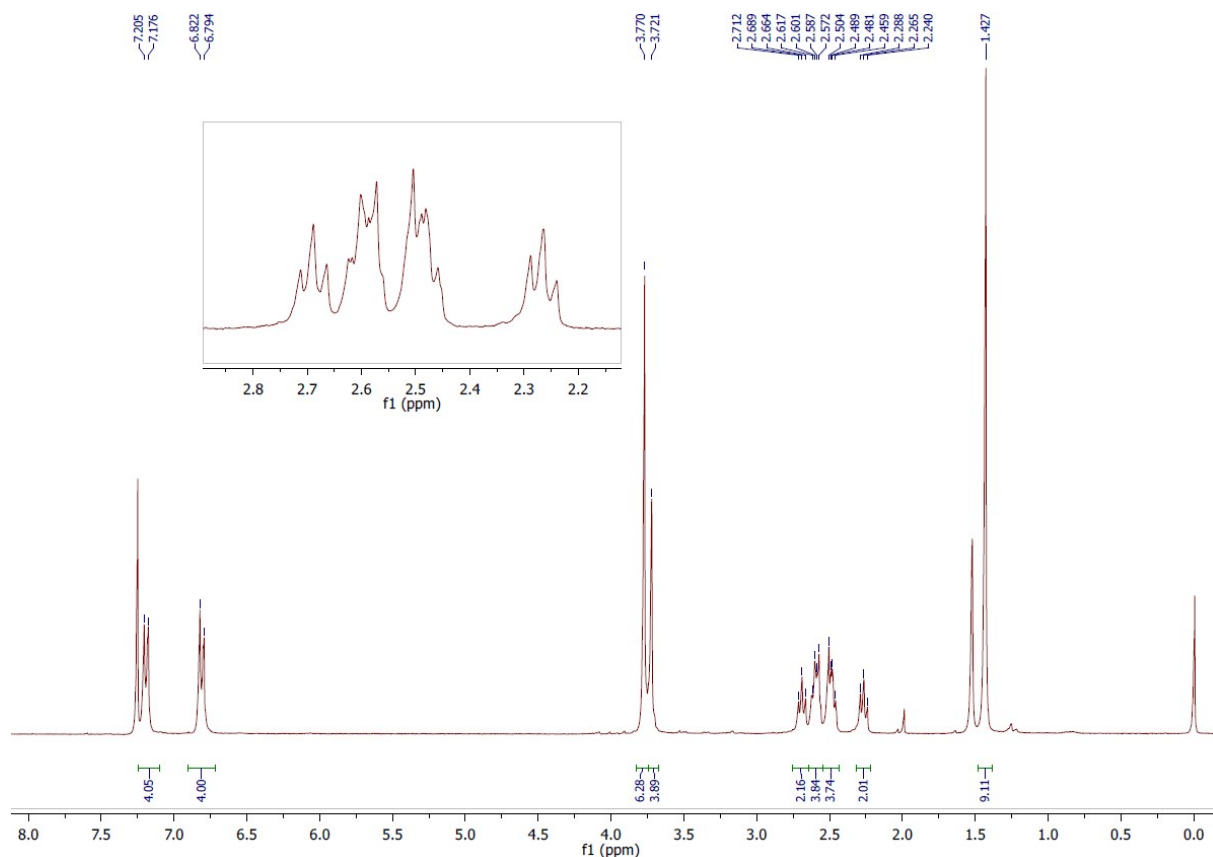


Figure S46. 1H NMR (300 MHz) spectrum for compound **D** ($CDCl_3$, 305 K). δ 7.20 (d, J = 8.7 Hz, 4H), 6.82 (d, J = 8.4 Hz, 4H), 3.77 (s, 6H), 3.72 (s, 4H), 2.69 (t, J = 6.9 Hz, 2H), 2.60 (m, 4H), 2.48 (m, 4H), 2.26 (t, J = 6.9 Hz, 2H), 1.42 (s, 9H).

Cyclization/deprotection of **D**

Diselenide **D** (50 mg, 0.08 mmol) and DTNP (51.4 mg, 0.16 mmol, 2 equiv) were dissolved separately in TFA (0.5 mL for each). The solution of DTNP was added dropwise in the solution of diselenide **D**. The reaction mixture was stirred during 1 h at r.t. Then, cold Et_2O was added (40 mL) and was evaporated under reduced pressure in order to eliminate TFA. Water (10 mL) was added and the solution was purified by RP-HPLC (XBridge C18 OBD column, eluent A : water containing 0.1% TFA by vol, eluent B: water/acetonitrile : ¼ by vol containing 0.1% TFA by vol, detection at 215 nm, flow rate 25 mL/min, 0 -

15% eluent B in 15 min). The purified fractions were collected, frozen and lyophilized to give 19.8 mg (57%) of **7d** as the trifluoroacetate salt.

2. Kinetic measurements

Table S3 : Peptides used for the kinetic study

SEA ^{off} , Cys-peptides	
9a	H-ILKEPVHGA-SEA ^{off}
9b	H-ILKEPVHGV-SEA ^{off}
9c	H-ILKEPVHGT-SEA ^{off}
9d	H-ILKEPVHGI-SEA ^{off}
13	CILKEPVHGV-NH ₂

2.1 SEA-thiol exchange. General procedure illustrated with peptide **9a** and MPAA as catalyst

TCEP-HCl (28.67 mg, 0.1 mmol, 100 mM final concentration) was dissolved in 6 M guanidine-HCl, 0.2 M pH 7.2 sodium phosphate buffer (1 mL). Then, MPAA (16.82 mg, 0.1 mmol, 100 mM final concentration) was dissolved in 20 μ L NaOH (6 M) to achieve complete dissolution prior to addition to the above solution.^a

The SEA peptide **9a** (0.994 mg, 0.6 μ mol) was dissolved in the above solution (final peptide concentration 1 mM) and MPA (5% v/v) added in one portion. NaOH (5 M) was added to adjust the pH to 4.0. The reaction mixture was shaken at 37 °C under nitrogen atmosphere and monitored by HPLC. For each time point, 2 μ L aliquots were withdrawn, quenched by adding 50 μ L of 0.1% aqueous TFA and in case of MPAA extracted with Et₂O to remove it.

2.2 SEA ligation. General procedure illustrated with peptide **9b**, Cys peptide **13** and MPAA as catalyst

TCEP-HCl (11.47 mg, 40 μ mol, 200 mM final concentration) was dissolved in 6 M guanidine-HCl, 0.1 M pH 7.2 sodium phosphate buffer (200 μ L). Then, MPAA (6.72 mg, 40 μ mol, 200 mM final concentration) was dissolved in the above solution. MPAA was first dissolved in 20 μ L NaOH to achieve complete dissolution prior to addition to the above solution.^a

The SEA peptide **9b** (1.498 mg, 1 μ mol, 7 mM final concentration) and Cys peptide **13** (2.22 mg, 1.5 μ mol, 10.5 mM, 1.5 equiv) were dissolved in the above solution. The reaction mixture was shaken at 37 °C under nitrogen atmosphere and monitored by HPLC. For each time point, 1 μ L aliquots were

^a Presolubilization in NaOH was not required for diselenide catalysts

withdrawn, quenched by adding 50 μL of 0.1% aqueous TFA and in case of MPAA extracted with Et_2O to remove it.

3. Preparative synthesis of peptide thioester **12b** by SEA-thiol exchange

TCEP-HCl (14.33 mg, 50 μmol , 100 mM final concentration) was dissolved in 6 M guanidine-HCl, 0.2 M pH 7.2 sodium phosphate buffer (0.5 mL). Then, diselenide **7d** (18.64 mg, 50 μmol , 100 mM final concentration) was dissolved in the above solution (448 μL).

Peptide **9b** (4.55 mg, 3.1 μmol) was dissolved in the above solution (448 μL , final peptide concentration 7 mM) and MPA (22 μL , 5% v/v) was added in one portion. The pH was adjusted to 4.0 using NaOH (5 M). The reaction mixture was shaken at 37 $^\circ\text{C}$ under nitrogen atmosphere. After completion of the reaction, the mixture was diluted with water-TFA 0.1% (5 mL) and purified by reversed-phase HPLC using a linear water -acetonitrile gradient containing 0.1% TFA to give the peptide **12b** (2.8 mg, 63% yield).

The determination of the optical purity for the C-terminal Val residue was performed by chiral GC-MS analysis after acid hydrolysis in deuterated acid. The analysis was done by C.A.T. GmbH & Co. company (Chromatographie und Analysen technik KG, Heerweg 10, D-72070 Tübingen, Germany). The analysis indicated a D-Val content < 0.1%.

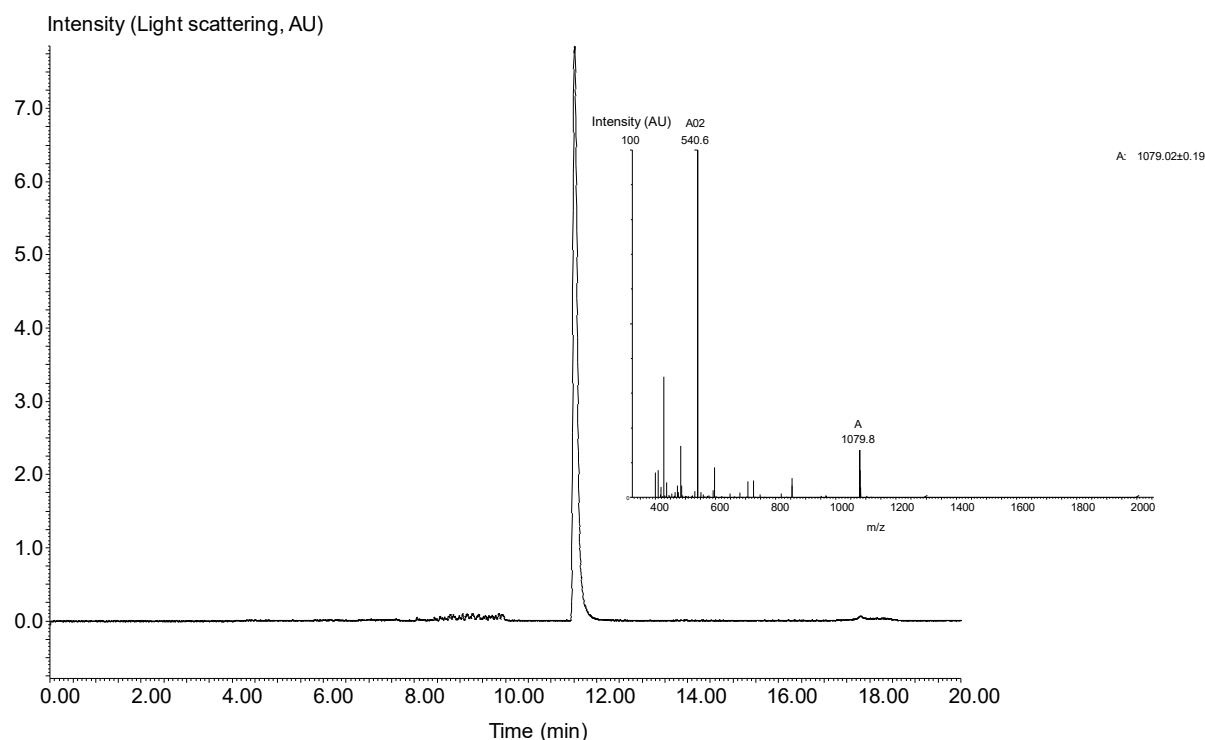


Figure S47. LC-MS analysis of peptide **12b**. LC trace, eluent A 0.10% TFA in water, eluent B 0.10% TFA in $\text{CH}_3\text{CN}/\text{water}$: 4/1 by vol. C18 Xbridge BEH 300 \AA 5 μm (4.6 \times 250 mm) column, gradient 0-50%

B in 15 min (1 mL/min, detection 215 nm). MS trace. $[M+H]^+$ m/z calcd. (monoisotopic) 1079.6, found 1079.02.

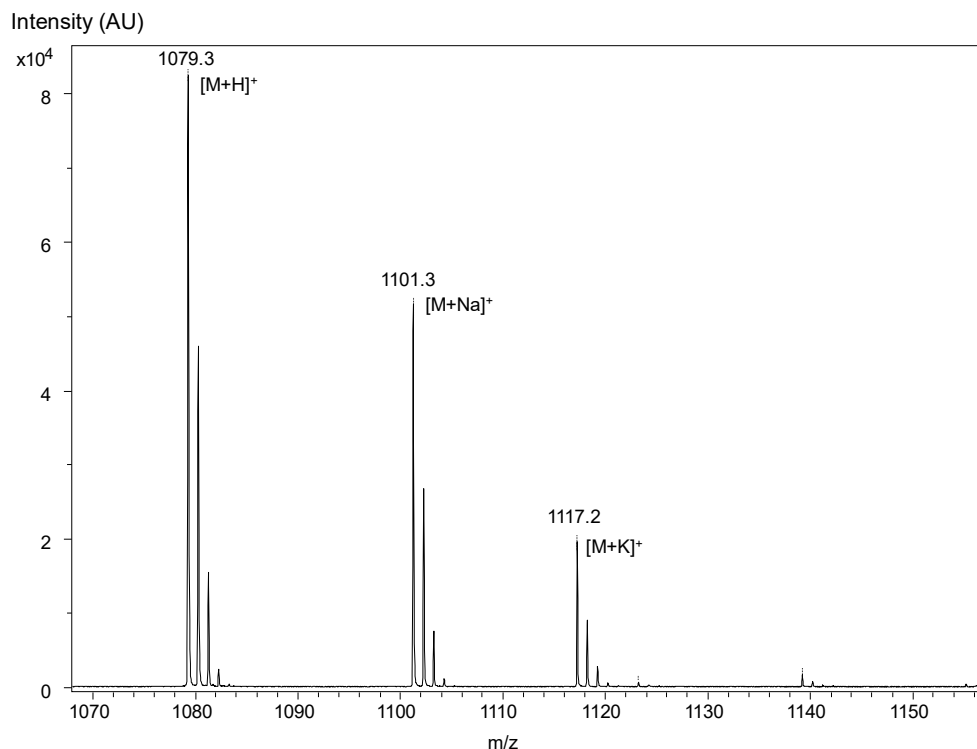


Figure S48. MALDI-TOF of peptide **12b**. Matrix α -cyano-4-hydroxycinnamic acid, positive detection mode, $[M+H]^+$ calcd (monoisotopic) 1079.6, observed mass: 1079.3 (monoisotopic).

4. Preparative SEA ligation

4.1 Synthesis of the peptide **14b** by reaction of the SEA^{off} peptide **9b** with Cys peptide **13**

TCEP-HCl (28.66 mg, 100 μ mol, 200 mM final concentration) was dissolved in 6 M guanidine-HCl, 0.1 M pH 7.2 sodium phosphate buffer (0.5 mL). Then, diselenide **7d** (20.76 mg, 50 μ mol, 100 mM) was dissolved in the above solution (499 μ L). The pH was adjusted to 4.0 using NaOH (6 M).

The SEA peptide **9b** (5.06 mg, 3.5 μ mol, 7 mM) and Cys peptide **13** (7.52 mg, 10.5 mM, 1.5 equiv) were dissolved in the above solution (499 μ L). The reaction mixture was shaken at 37 °C under nitrogen atmosphere. After completion of the ligation, the mixture was diluted with water-TFA 0.1% (5 mL) and purified by reversed-phase HPLC using a linear water-acetonitrile gradient containing 0.1% TFA to give the purified ligation product **14b** (4.69 mg, 51% yield).

The determination of the optical purity for the C-terminal amino acid residue within peptide **14b** was performed by chiral GC-MS analysis after acid hydrolysis in deuterated acid. The analysis was done by C.A.T. GmbH & Co. company (Chromatographie und Analysen technik KG, Heerweg 10, D-72070 Tübingen, Germany). The analysis indicated a D-Val content of 0.19%.

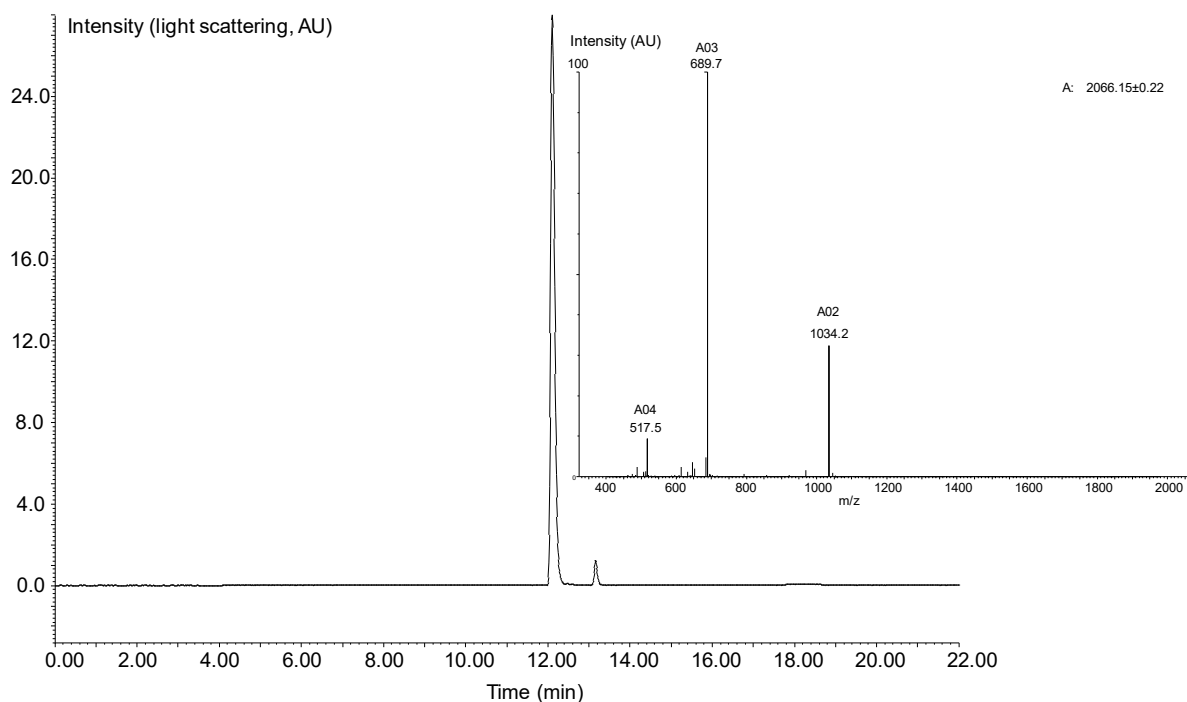


Figure S49. LC-MS analysis of peptide **14b**. LC trace, eluent A 0.10% TFA in water, eluent B 0.10% TFA in CH₃CN/water: 4/1 by vol. C18 Xbridge BEH 300 Å 5 μm (4.6 × 250 mm) column, gradient 0-50% B in 15 min (1 mL/min, detection 215 nm). MS trace. [M+H]⁺ m/z calcd. (monoisotopic) 2066.19, found 2067.15.

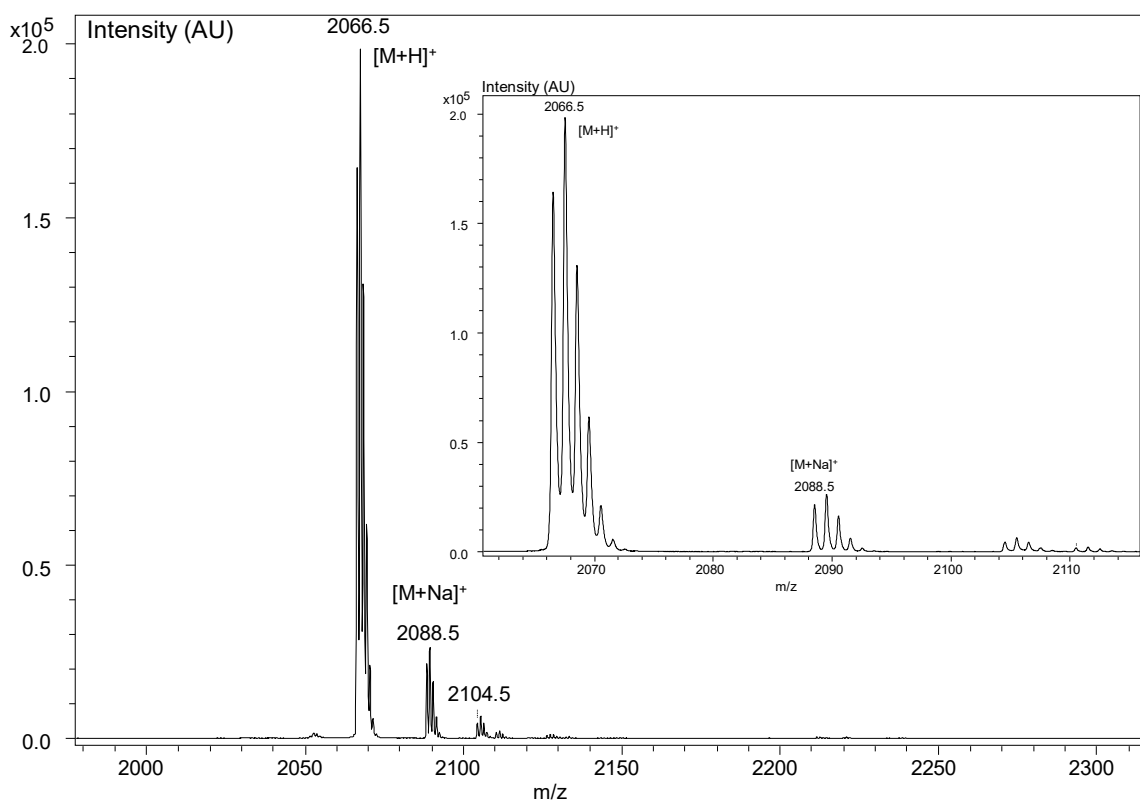


Figure S50. MALDI-TOF of peptide **14b**. Matrix α -cyano-4-hydroxycinnamic acid, positive detection mode, [M+H]⁺ calcd (monoisotopic) 2066.19, observed mass: 2066.5 (monoisotopic).

4.2 Cyclisation of peptide 26. Synthesis of cyclic peptide 27

4.2.1 Procedure

N-octyl glucoside (1.17 mg, 4 μ mol, 20 mM) was dissolved in 0.1 M pH 7.2 sodium phosphate buffer (200 μ L). Then, TCEP (11.47 mg, 40 μ mol, 200 mM) and diselenide **7d** (6.45 mg, 20 μ mol, 100 mM) were dissolved in the above solution. The pH was adjusted to 4.0 using NaOH (6 M).

Peptide **26** (2.75 mg, 0.5 μ mol, 3 mM) was dissolved in the above solution (154 μ L). The reaction mixture was shaken at 37 $^{\circ}$ C under nitrogen atmosphere and monitored by HPLC.

After completion of the cyclisation, the mixture was diluted with 6 M guanidine-HCl in water to achieve 500 μ L final volume. The crude peptide was purified by RP-HPLC (eluent A = water containing 0.1% TFA, eluent B = acetonitrile in water 4/1 containing 0.1% TFA, 60 $^{\circ}$ C, detection at 215 nm, 6 mL/min, 0 to 25% eluent B in 10 min, then 25 to 45% eluent B in 60 min, C3 Zorbax column) to give 0.67 mg of cyclic peptide **27** (26%).

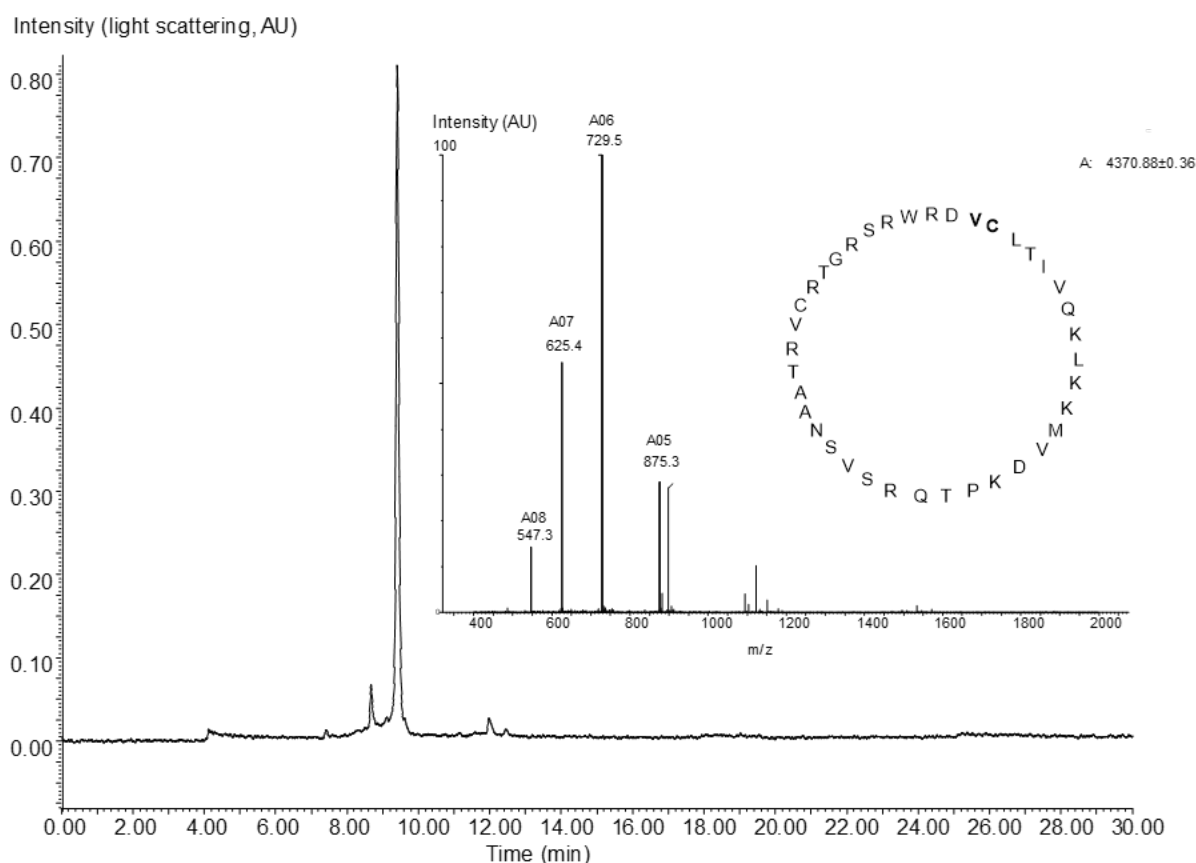


Figure S51. LC-MS analysis of peptide **27**. LC trace, eluent C 0.10% FA in water, eluent D 0.10% FA in CH₃CN/water: 4/1 by vol. C3 Zorbax 300SB \AA 3.5 μ m (4.6 \times 150 mm) column, gradient 0-100% D in 30 min (60 $^{\circ}$ C, 1 mL/min, detection 215 nm). MS trace. [M+H]⁺ m/z calcd. (monoisotopic) 4370.38, found 4370.88.

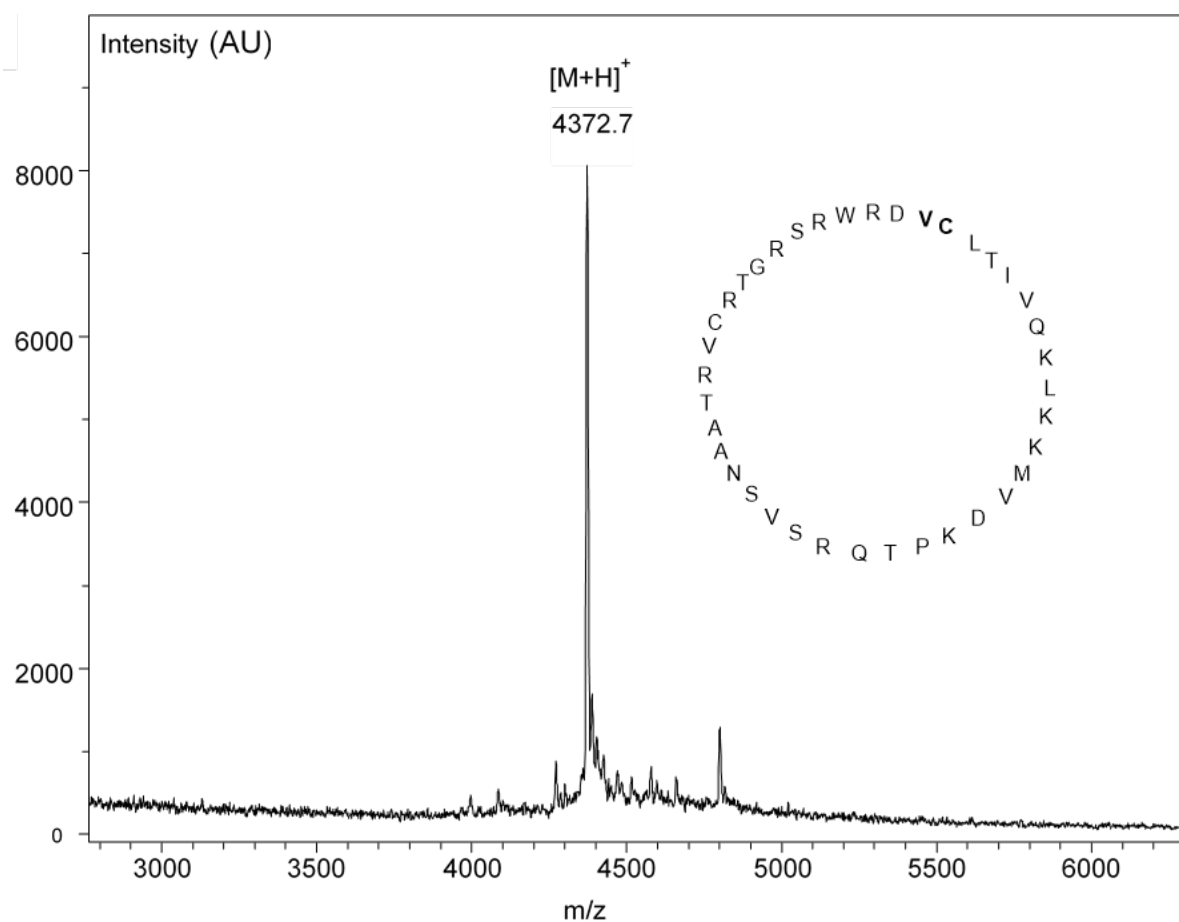


Figure S52. MALDI-TOF of peptide **27**. Matrix sinapinic acid, positive detection mode, $[M+H]^+$ calcd (average) 4373.2, observed mass: 4372.7 (average).

4.2.2 Proteomic analysis

Reduction-alkylation:

Peptide **27** (10 μg) was diluted with dithiothreitol (1 mg/mL in 0.025 M ammonium bicarbonate, 10 μL) and treated with iodoacetamide (10 mg/mL in 0.025 M ammonium bicarbonate, volume 10 μL) for 10 min. The alkylation step was monitored by MALDI-TOF mass spectrometry as shown below.

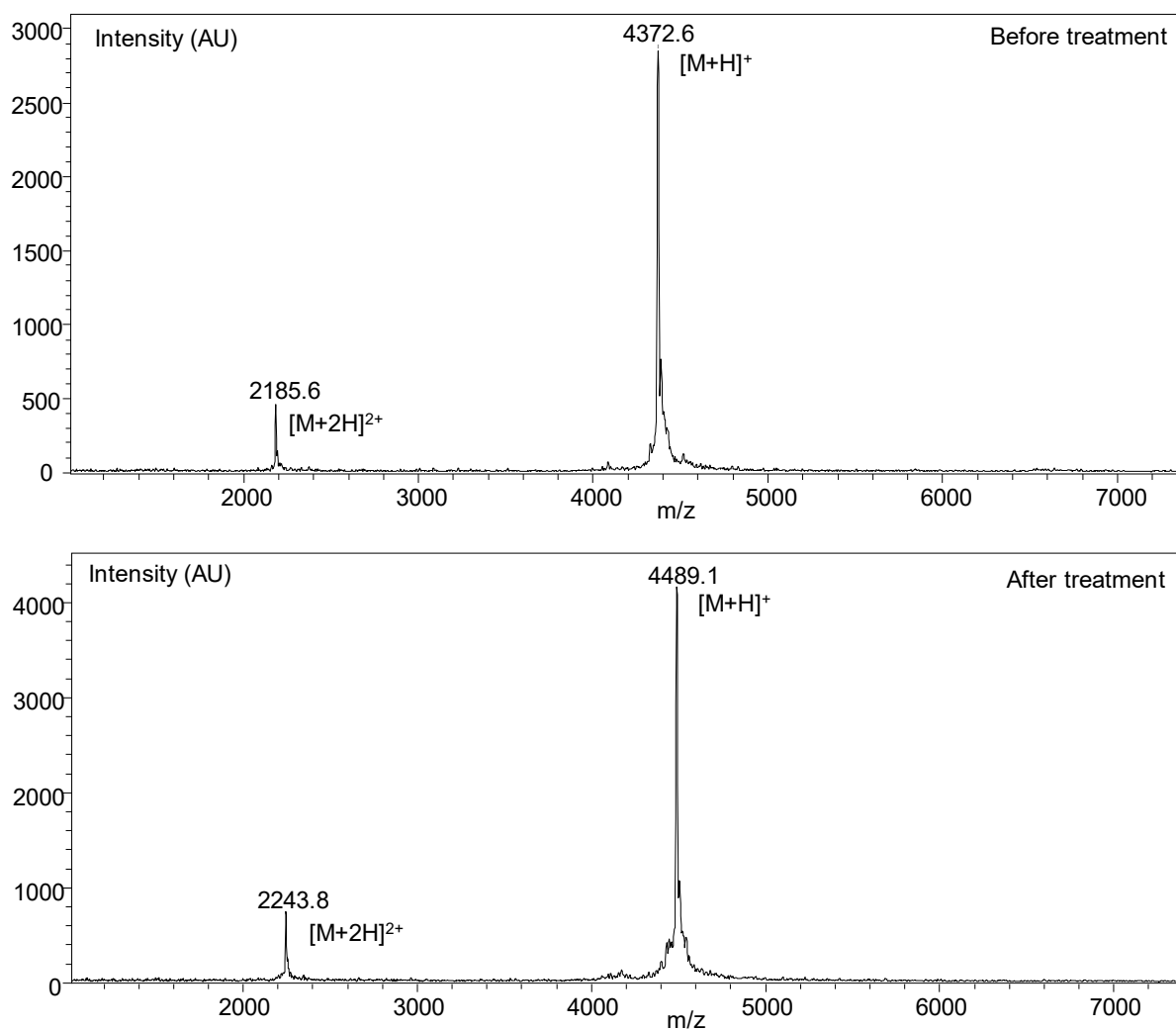
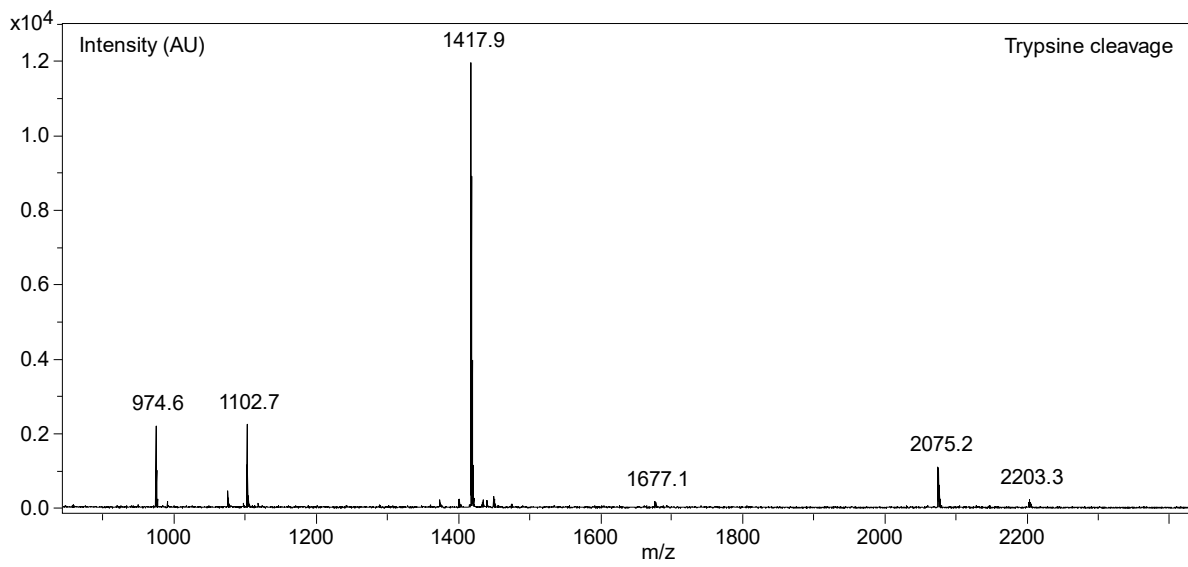


Figure S53. Monitoring of the alkylation step by MALDI-TOF mass spectrometry (matrix: α -cyano-4-hydroxycinnamic acid).

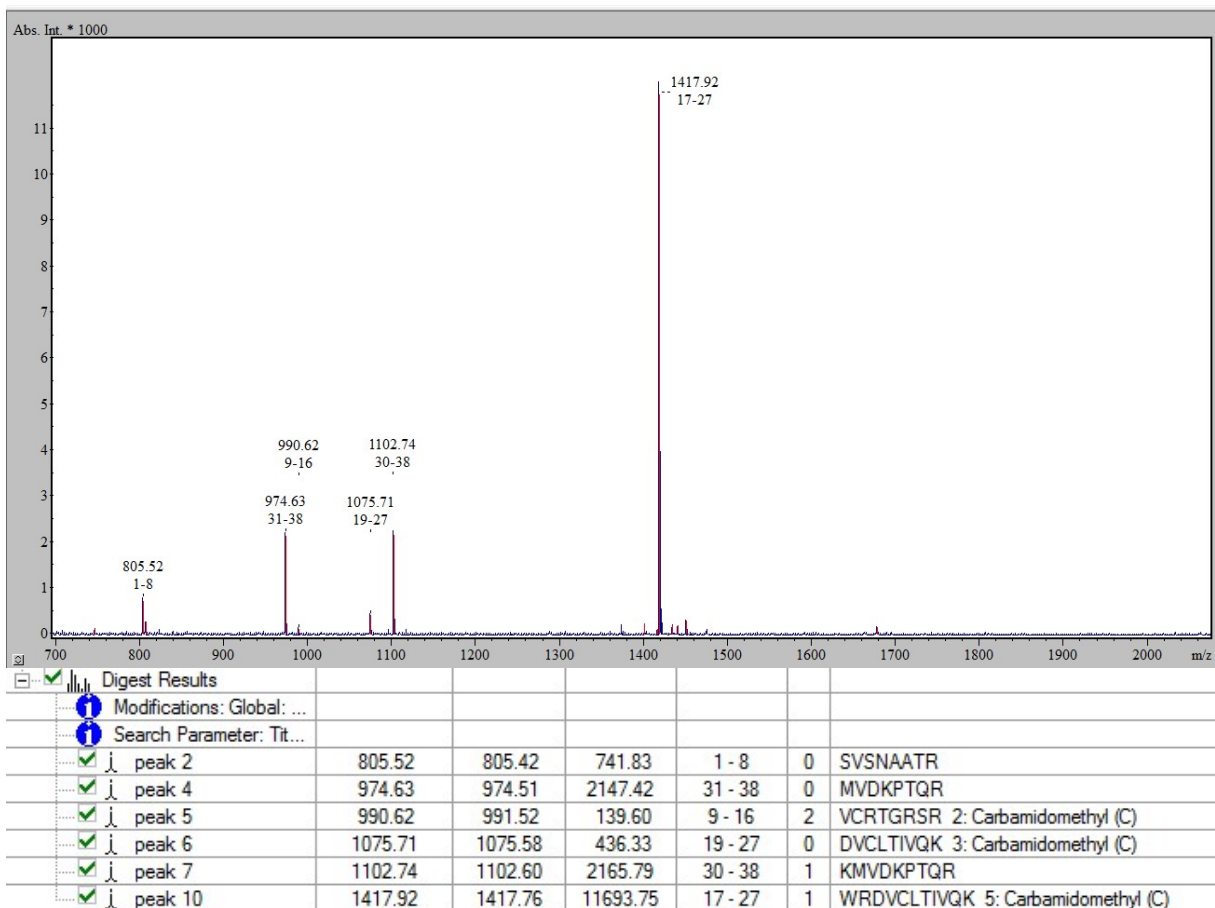
Trypsine cleavage:

Then, trypsin (0.1 mg/mL, 1 μ L) was added to the mixture to cleave the alkylated cyclic peptide. The formed peptide fragments were identified by MALDI-TOF MS-MS.

A)



B)



C)

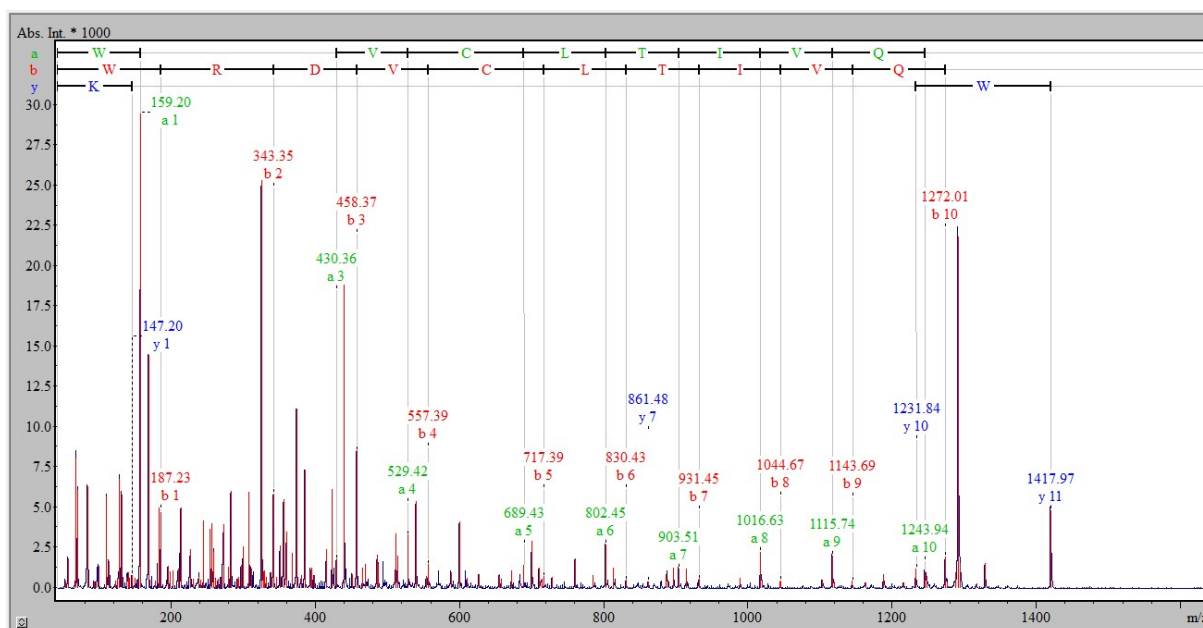


Figure S54. In source MALDI-TOF sequencing of peptide **27** after alkylation and trypsin cleavage using α -cyano-4-hydroxycinnamic acid as matrix (positive reflector mode). A) Global spectrum, B) MS-MS sequencing of the peptides, C) MS-MS sequencing of the ion at m/z 1417.97 shows the presence of the Val-Cys peptide bond (...RDVCLT...).

5. Synthesis of 15 kDa granulysin in the *N* to *C* direction

5.1 Synthesis of peptide **12e** by SEA-thiol exchange of **9e**

TCEP-HCl (8.6 mg, 30 μ mol, 100 mM final concentration) was dissolved in 6 M guanidine-HCl, 0.2 M pH 7.2 sodium phosphate buffer (300 μ L). Then, diselenide **7d** (12.48 mg, 30 μ mol, 100 mM final concentration) was dissolved in the above solution and MPA (6 μ L, 5% v/v) was added in one portion. The pH was adjusted to 4.0 using NaOH (6 M).

Peptide **9e** (13.32 mg, 2 μ mol) was dissolved in the above solution (287 μ L, final peptide concentration 7 mM). The reaction mixture was shaken at 37 $^{\circ}$ C under nitrogen atmosphere and monitored by HPLC.

After completion of the reaction, the mixture was diluted with water-TFA 0.1% (5 mL). The crude peptide was purified by RP-HPLC (eluent A = water containing 0.1% TFA, eluent B = acetonitrile in water 4/1 containing 0.1% TFA, 60 $^{\circ}$ C, detection at 215 nm, 6 mL/min, 0 to 25% eluent B in 10 min, then 25 to 35% eluent B in 60 min, C3 Zorbax column) to give 4.65 mg of peptide **9e** (36%).

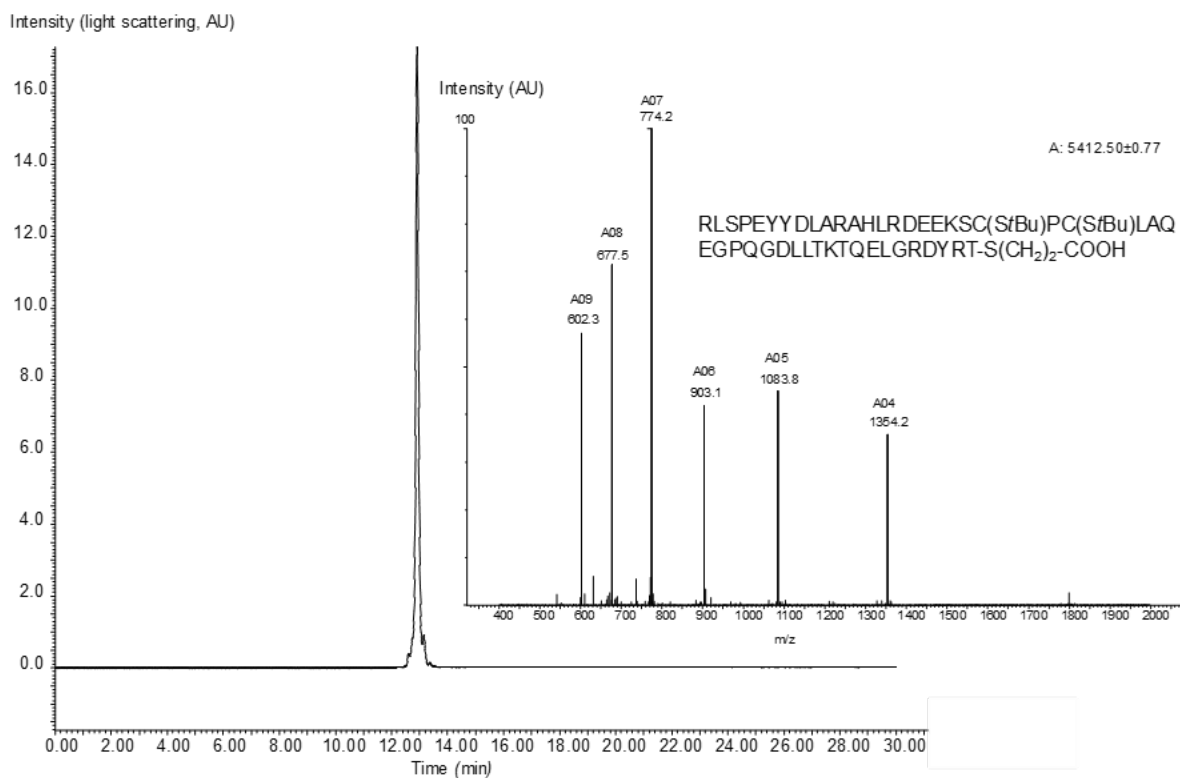


Figure S55. LC-MS analysis of peptide **12e**. LC trace, eluent A 0.10% TFA in water, eluent B 0.10% TFA in CH₃CN/water: 4/1 by vol. C3 Zorbax 300SB Å 3.5 µm (4.6 × 150 mm) column, gradient 0-100% B in 30 min (60 °C, 1 mL/min, detection 215 nm). MS trace. [M+H]⁺ m/z calcd. (monoisotopic) 5413.04, found 5413.5.

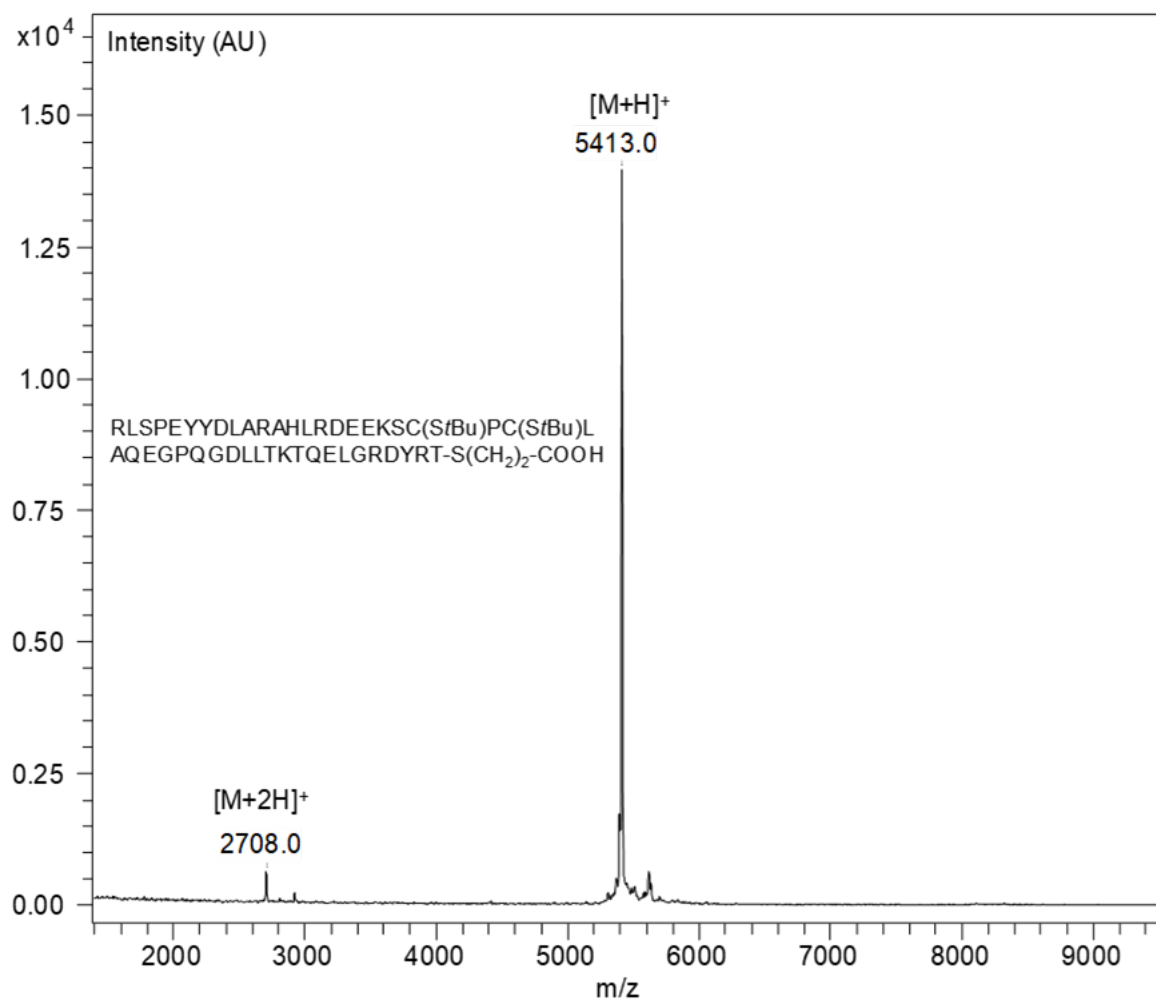


Figure S56. MALDI-TOF of peptide **12e**. Matrix sinapinic acid, positive detection mode, $[M+H]^+$ calcd (monoisotopic) 5413.04, observed mass: 5413.0 (average).

5.2 Synthesis of peptide **28**

MCAA (6.73 mg, 40 μ mol, 200 mM final concentration) was first dissolved in NaOH (6 M, 10 μ L) to achieve complete dissolution. Then, this solution was combined with 6 M guanidine-HCl, 0.1 M pH 7.2 sodium phosphate buffer (190 μ L). The pH was adjusted to 7.7 using NaOH (6 M).

Peptide **12e** (4.64 mg, 0.7 μ mol, 7 mM) and peptide **26** (4.28 mg, 0.7 μ mol, 7 mM) were dissolved in the above solution (103 μ L). The reaction mixture was shaken at 37 $^{\circ}$ C under nitrogen atmosphere.

After 24 h of reaction, the mixture was diluted with 6 M guanidine-HCl, 0.1 M pH 7.2 sodium phosphate buffer (2 mL) and acidified with water-TFA 10 % (200 μ L). The solution was extracted with Et₂O (3 x 3 mL) to remove MCAA. The crude peptide was purified by RP-HPLC (eluent A = water containing 0.1% TFA, eluent B = acetonitrile in water 4/1 containing 0.1% TFA, 60 $^{\circ}$ C, detection at 215 nm, 6 mL/min, 0 to 25% eluent B in 10 min, then 25 to 45% eluent B in 60 min, C3 Zorbax column) to give 2.97 mg of **28** (35%).

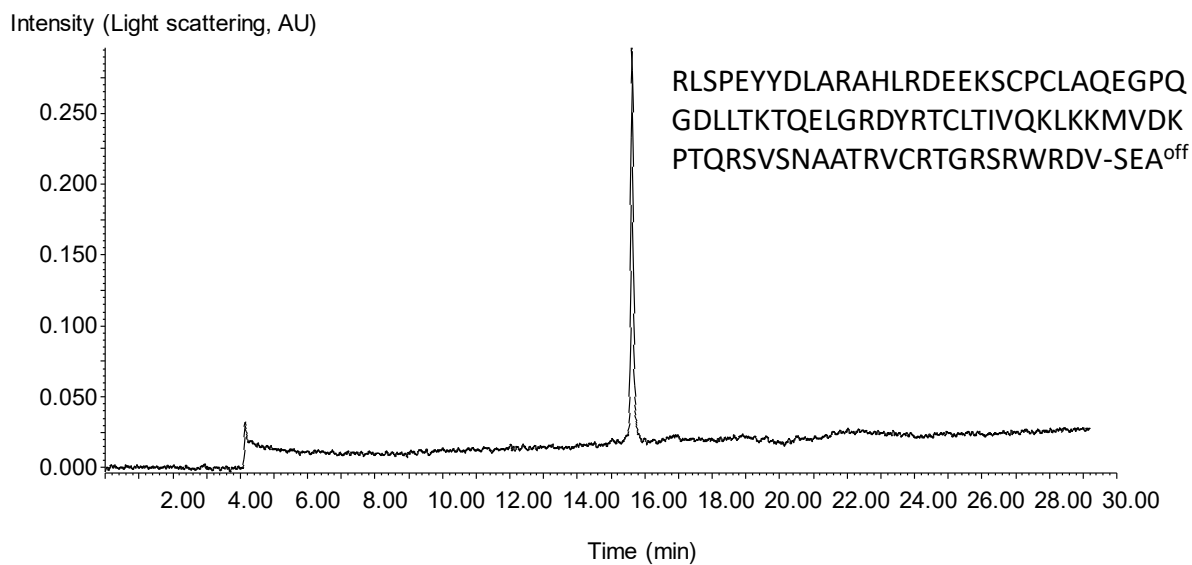


Figure S57. LC-MS analysis of peptide **28**. LC trace, eluent A 0.10% TFA in water, eluent B 0.10% TFA in CH₃CN/water: 4/1 by vol. C3 Zorbax 300SB Å 3.5 µm (4.6 × 150 mm) column, gradient 0-100% B in 30 min (60 °C, 1 mL/min, detection 215 nm).

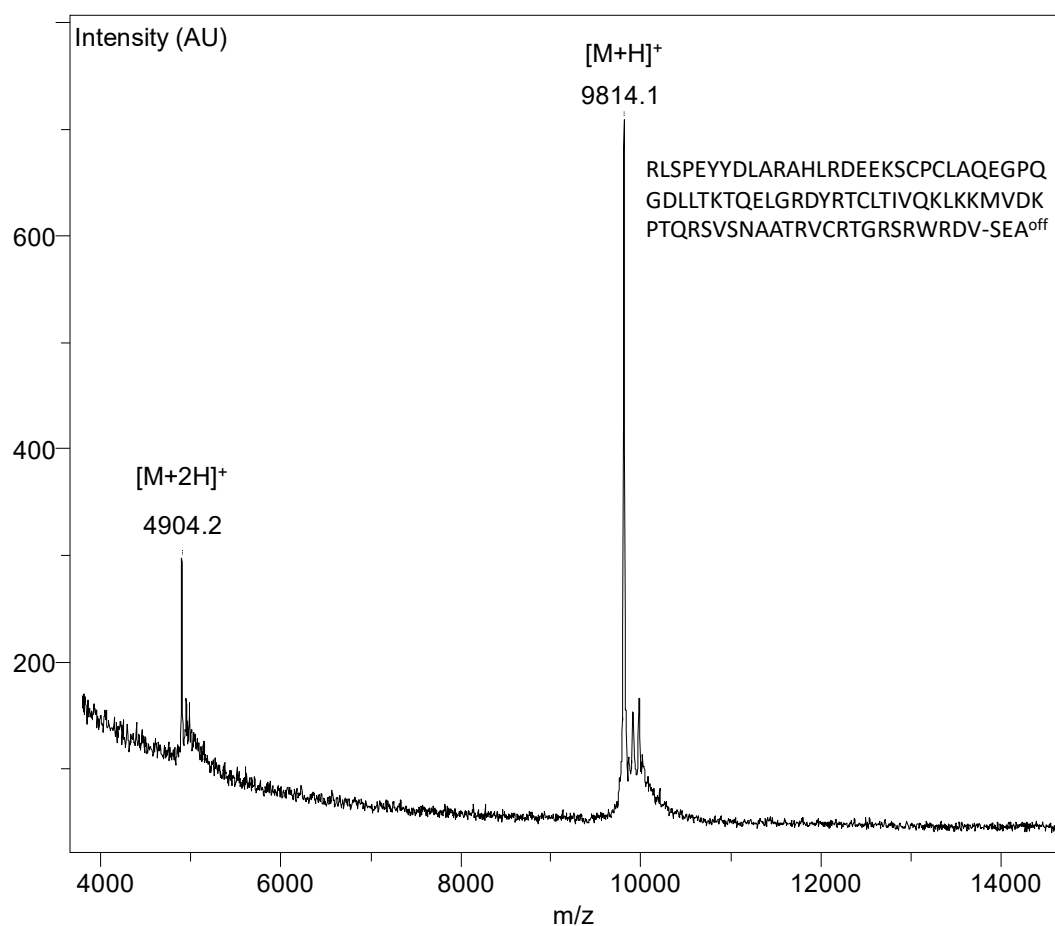


Figure S58. MALDI-TOF of peptide **28**. Matrix sinapinic acid, positive detection mode, [M+H]⁺ calcd (monoisotopic) 9813.33, observed mass: 9814.1 (average).

5.3 Synthesis of 15 kDa GLNY

N-octyl glucoside (0.88 mg, 3 μmol, 20 mM) was dissolved in 0.1 M pH 7.2 sodium phosphate buffer (150 μL). Then, TCEP (8.6 mg, 30 μmol, 200 mM) and diselenide **7d** (6.24 mg, 15 μmol, 100 mM) were dissolved in the above solution. The pH was adjusted to 4.0 using NaOH (6 M).

Peptide **28** (2.97 mg, 0.4 μmol, 2 mM) and peptide **29** (2.57 mg, 0.5 μmol, 1.2 equiv) were dissolved in the above solution (137 μL). The reaction mixture was shaken at 37 °C under nitrogen atmosphere and monitored by HPLC.

After 72 h, the mixture was diluted with 6 M guanidine-HCl, 0.1 M pH 7.2 sodium phosphate buffer (500 μL final volume). The crude protein was purified by RP-HPLC (eluent A = water containing 0.1% TFA, eluent B = acetonitrile in water 4/1 containing 0.1% TFA, 60 °C, detection at 215 nm, 6 mL/min, 0 to 35% eluent B in 10 min, then 35 to 55% eluent B in 60 min, C3 Zorbax column) to give 0.92 mg of linear 15 kDa granulysin polypeptide (13%).

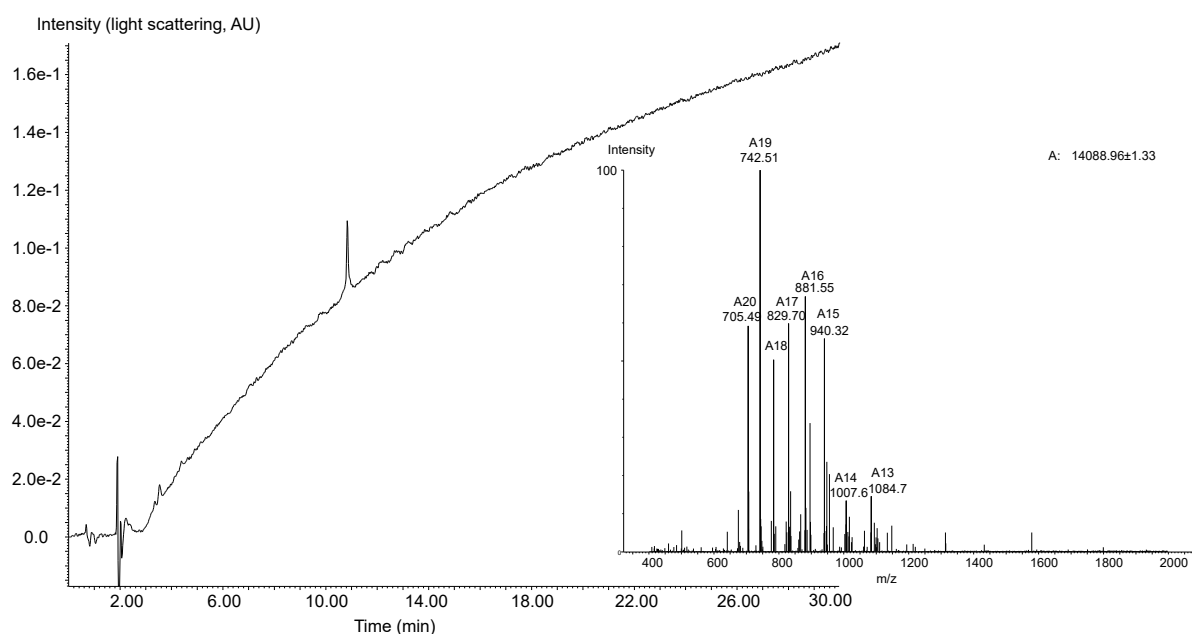


Figure S59. LC-MS analysis of peptide 15 kDa GLNY LC trace, eluent A 0.1% TFA in water, eluent B 0.1% TFA in CH₃CN/water: 4/1 by vol. C3 Zorbax 300SB Å 3.5 μm (4.6 × 150 mm) column, gradient 0-100% B in 30 min (60 °C, 1 mL/min, detection 215 nm).

6. Synthesis of 15 kDa granulysin in the C to N direction

6.1 Synthesis of peptide 30

Peptide elongation was performed on SEA ChemMatrix resin (0.1 mmol, 0.34 mmol/g) using standard Fmoc/*tert*-butyl chemistry on an automated peptide synthesizer. We used typical procedures for the synthesis of SEA^{off} peptide segments^{1,3}. However, two *O*-acyl isodipeptide units Boc-Thr(Fmoc-Ala)-OH

and Boc-Ser(Fmoc-Val)-OH (in bold in the sequence below) were incorporated in the peptide backbone as described in a previous article¹².

Granulysin 69-106: CLTIVQKLKMKMVDKPTQRS**VSNAATR**VCRTGRSRWRDV

Coupling of N-acetoacetyl group (AcA) on the solid phase

After elongation of the peptide on SEA PS resin and terminal Fmoc deprotection (0.1 mmol scale), *N*-hydroxysuccinimidyl acetoacetate (10 eq, 199.16 mg) and *N*-methyilmorpholine (5 eq, 54.97 μ L) were dissolved in the minimal volume of DMF and the resulting solution was added to the peptidyl resin. After stirring 40 min, the resin was washed with DMF (3 \times 2 min). The completion of the acylation step was verified using the TNBS test for primary amines. Then, resin was washed with DCM (3 \times 2 min), diethyl ether (3 \times 2 min), and dried *in vacuo*.

Final peptide deprotection and cleavage

Peptides were deprotected and cleaved from the resin using TFA/DMS/thioanisol/3,4-difluorothiophenol/water: 90/2.5/2.5/2.5/2.5 by vol for 2 h (10 mL). The crude peptides were precipitated in cold diethyl ether/heptane: 1/1 by vol (200 mL), solubilized in deionised water and lyophilized. The crude peptide was purified by RP-HPLC (eluent A = water containing 0.1% TFA, eluent B = acetonitrile in water 4/1 containing 0.1% TFA, 50 $^{\circ}$ C, detection at 215 nm, 6 mL/min, 0 to 20% eluent B in 5 min, then 20 to 60% eluent B in 45 min, C18 XBridge column) to give 18.48 mg of peptide AcA- CLTIVQKLKMKMVDKPTQRS**VSNAATR**VCRTGRSRWRDV-SEA^{off} (3.2%).

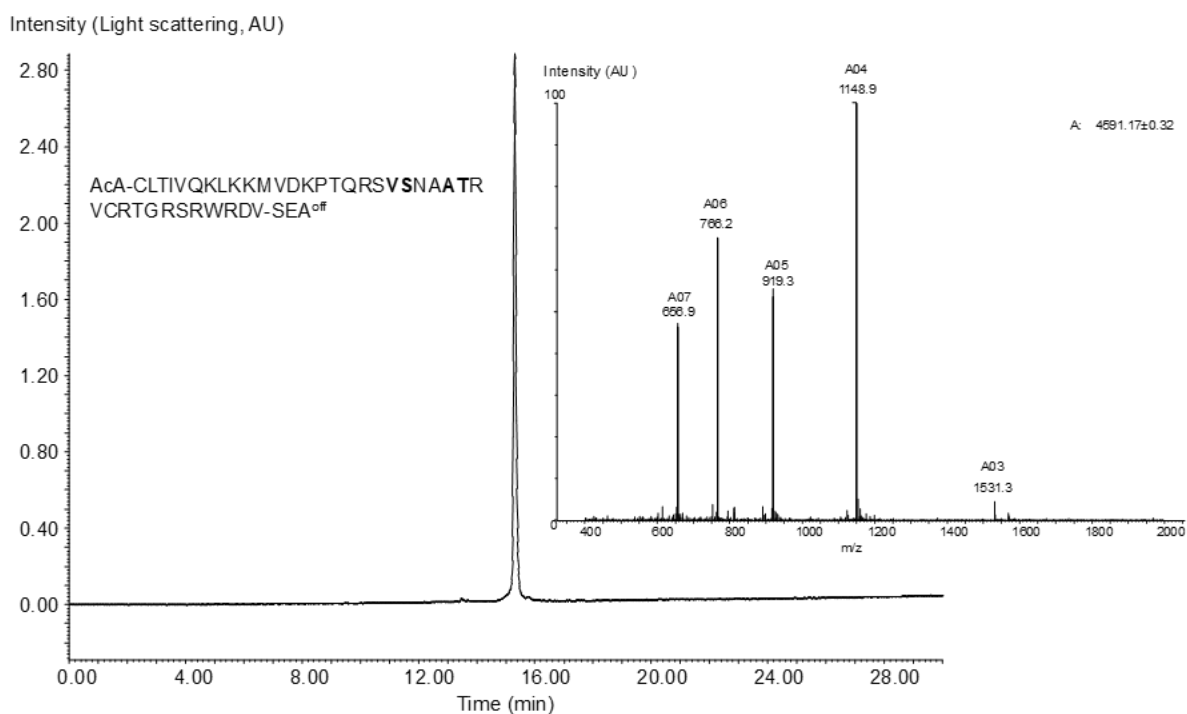


Figure S60. LC-MS analysis of peptide **30**. LC trace, eluent A 0.10% TFA in water, eluent B 0.10% TFA in CH₃CN/water: 4/1 by vol. C18 Xbridge BEH 300 Å 5 µm (4.6 × 250 mm) column, gradient 0-100% B in 30 min (60 °C, 1 mL/min, detection 215 nm). MS trace. [M+H]⁺ m/z calcd. (average) 4592.5, found 4591.2.

6.2 Synthesis of peptide **31**

TCEP-HCl (17.77 mg, 60 µmol, 200 mM final concentration) was dissolved in 6 M Gn•HCl, 0.1 M pH 7.2 sodium phosphate buffer (311 µL). Then, the diselenide **7d** (12.90 mg, 31 µmol, 100 mM final concentration) was dissolved in the above solution and MPA (16 µL, 5% v/v) was added. The pH was adjusted to 3.7 using NaOH (6 M).

Peptide **30** (6.08 mg, 1 µmol) was dissolved in the above solution (303 µL, final peptide concentration 3.5 mM). The reaction mixture was shaken at 37 °C under nitrogen atmosphere and monitored by HPLC.

After completion of the reaction (48 h), the mixture was diluted with water-TFA 0.1% (5 mL). The crude peptide was purified by RP-HPLC (eluent A = water containing 0.1% TFA, eluent B = acetonitrile in water 4/1 containing 0.1% TFA, 60 °C, detection at 215 nm, 6 mL/min, 0 to 25% eluent B in 10 min, then 25 to 45% eluent B in 60 min, C18 XBridge BEH300 Prep column) to give 2.86 mg of peptide **31** (47%).

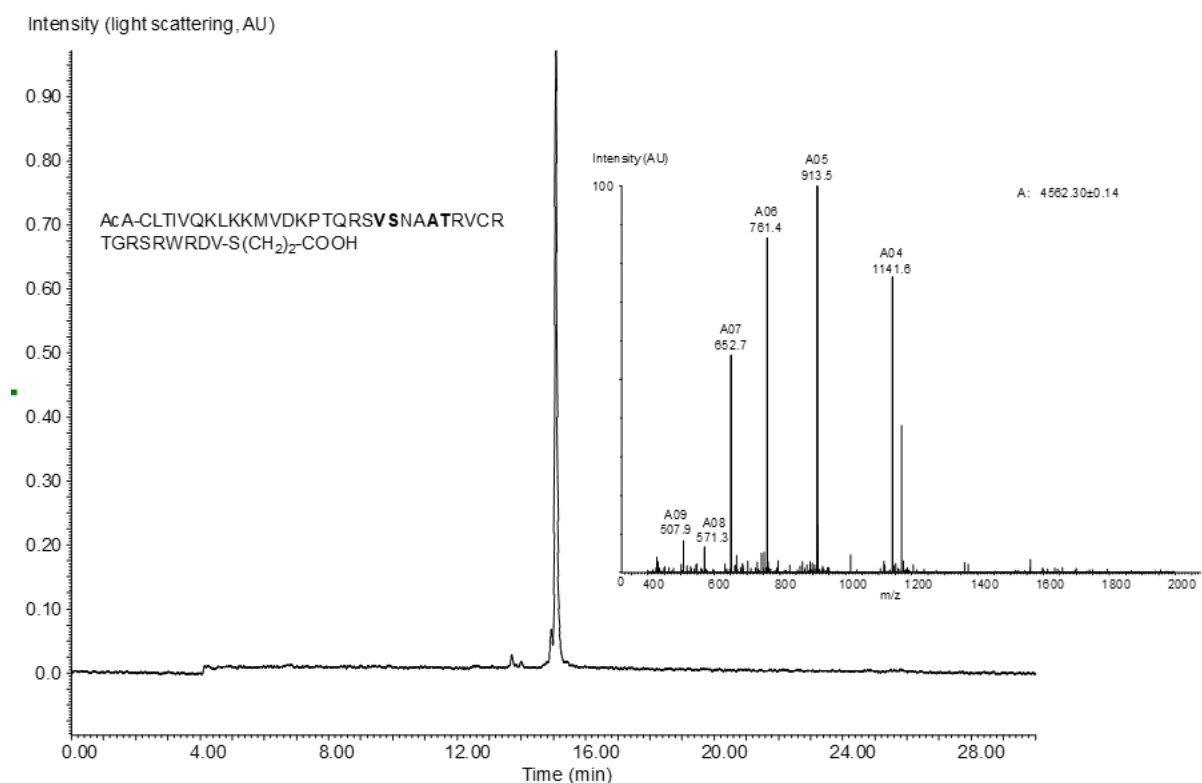


Figure S61. LC-MS analysis of peptide **31**. LC trace, eluent A 0.10% TFA in water, eluent B 0.10% TFA in CH₃CN/water: 4/1 by vol. C18 Xbridge BEH 300 Å 5 µm (4.6 × 250 mm) column, gradient 0-100% B in 30 min (60 °C, 1 mL/min, detection 215 nm). MS trace. [M+H]⁺ m/z calcd. (average) 4561.4, found 4562.3.

6.3 Synthesis of peptide 33

MCAA (6.72 mg, 40 µmol, 200 mM final concentration) was first dissolved in NaOH (6 M, 10 µL) to achieve complete dissolution. Then, this solution was combined with 6 M Gn•HCl, 0.1 M pH 7.2 sodium phosphate buffer (200 µL). The pH was adjusted to 7.6 using NaOH (6 M).

Peptide **31** (2.86 mg, 0.5 µmol, 5 mM) and peptide **29** (2.87 mg, 0.55 µmol, 5.5 mM, 1.1 equiv) were dissolved in the above solution (100 µL). The reaction mixture was shaken at 37 °C under nitrogen atmosphere. The result of the NCL reaction is shown in Figure S62.

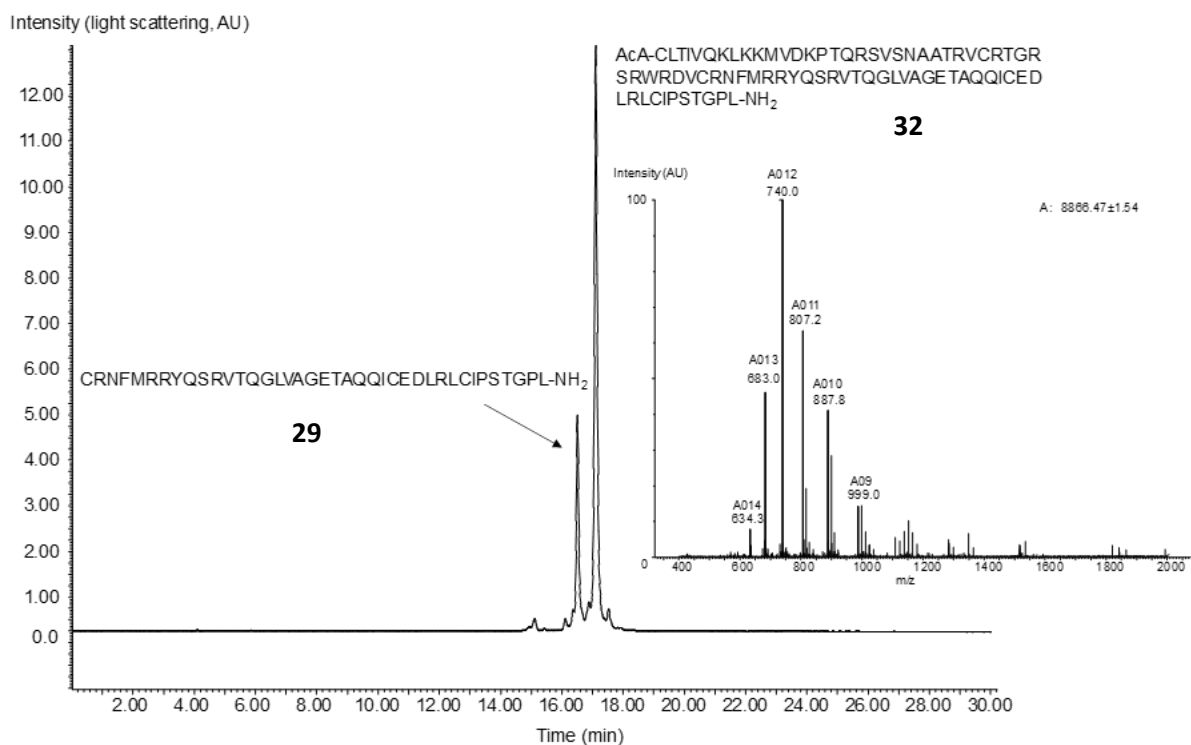


Figure S62. LC-MS analysis of the crude reaction mixture (synthesis of peptide **32** by the ligating peptides **31** and **29**) after 22 h of reaction. LC trace, eluent A 0.1% TFA in water, eluent B 0.1% TFA in CH₃CN/water: 4/1 by vol. C3 Zorbax 300SB Å 3.5 µm (4.6 × 150 mm) column, gradient 0-100% B in 30 min (60 °C, 1 mL/min, detection 215 nm). MS trace. [M+H]⁺ m/z calcd. (average) 8866.4, found 8866.5.

After 24 h of reaction, acetic acid (24 µL) was added to have a final 20% acetic acid solution. The solution was extracted with Et₂O (3 × 3 mL) to remove MPAA. Then, 0.43 M hydroxylamine solution in water (24 µL) was added to remove AcA protecting group. The reaction medium was heated at 37 °C for 5 h and monitored by MALDI-TOF (Figure S63). After complete deprotection, TCEP (3.75 mg, 5 equiv/cysteine) was diluted in 0.1 M pH 8 phosphate buffer (100 µL) and added to the reaction mixture to reduce any disulfide bonds. The reaction medium was stirred for 1 h. The mixture was diluted with water-TFA 0.1% (5 mL). The crude peptide was purified by RP-HPLC (eluent A = water containing 0.25% TFA, eluent B = acetonitrile in water 4/1 containing 0.25% TFA, 60 °C, detection at 215 nm, 6 mL/min, 0 to 35% eluent B in 10 min, then 35 to 55% eluent B in 60 min, C3 Zorbax column) to give 2.19 mg of **33** (42%, corrected yield 49%).

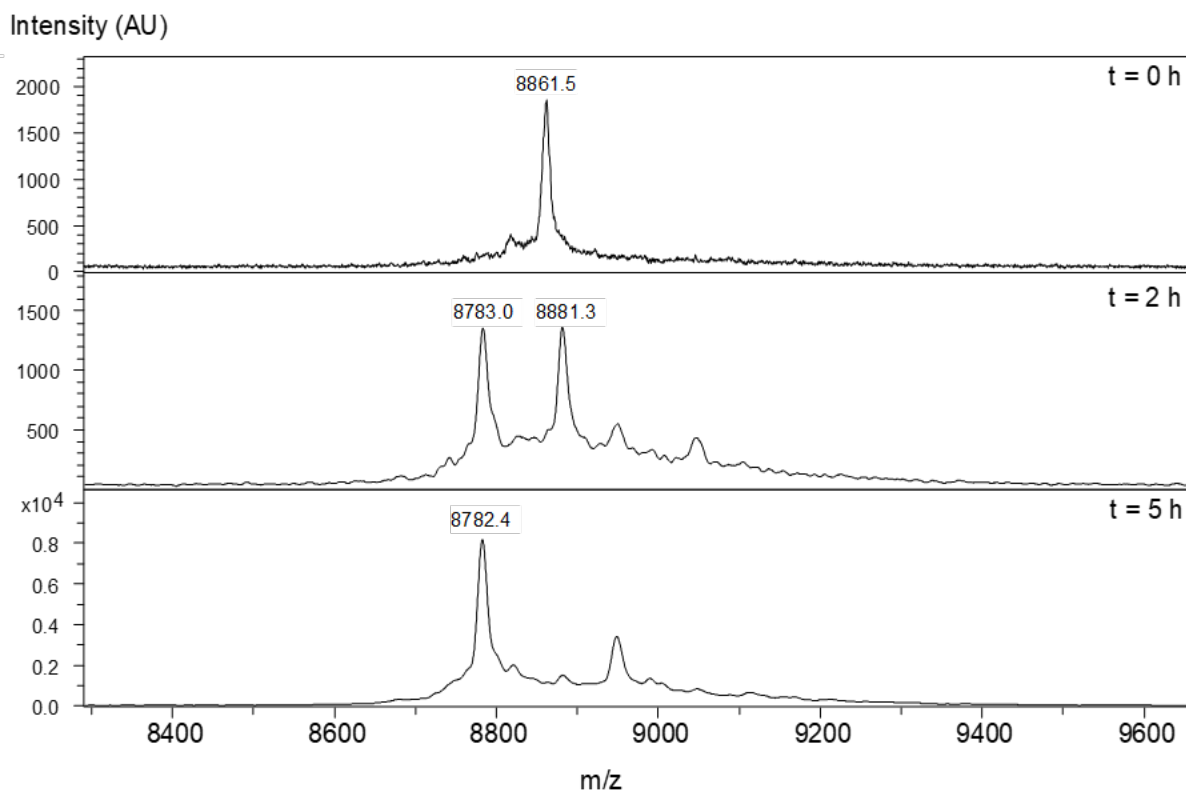


Figure S63. MALDI-TOF monitoring of the AcA removal (**32** → **33**). Matrix sinapinic acid, positive detection mode.

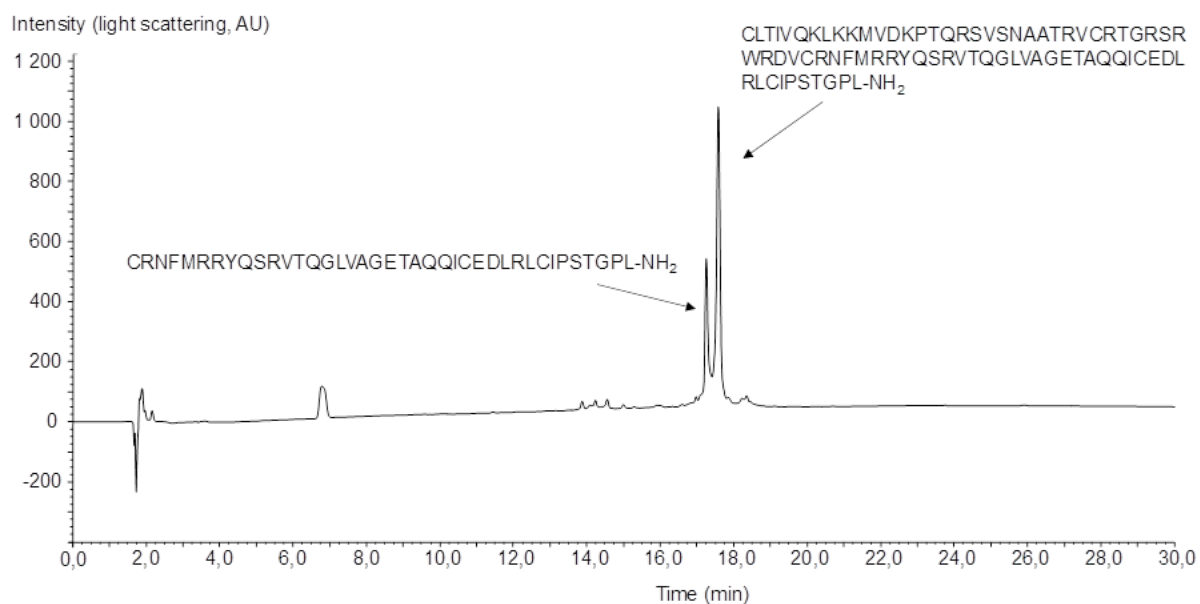


Figure S64. HPLC analysis of the crude peptide **33**. LC trace, eluent A 0.25% TFA in water, eluent B 0.25% TFA in CH₃CN/water: 4/1 by vol. C3 Zorbax 300SB Å 3.5 µm (4.6 × 150 mm) column, gradient 0-100% B in 30 min (60 °C, 1 mL/min, detection 215 nm).

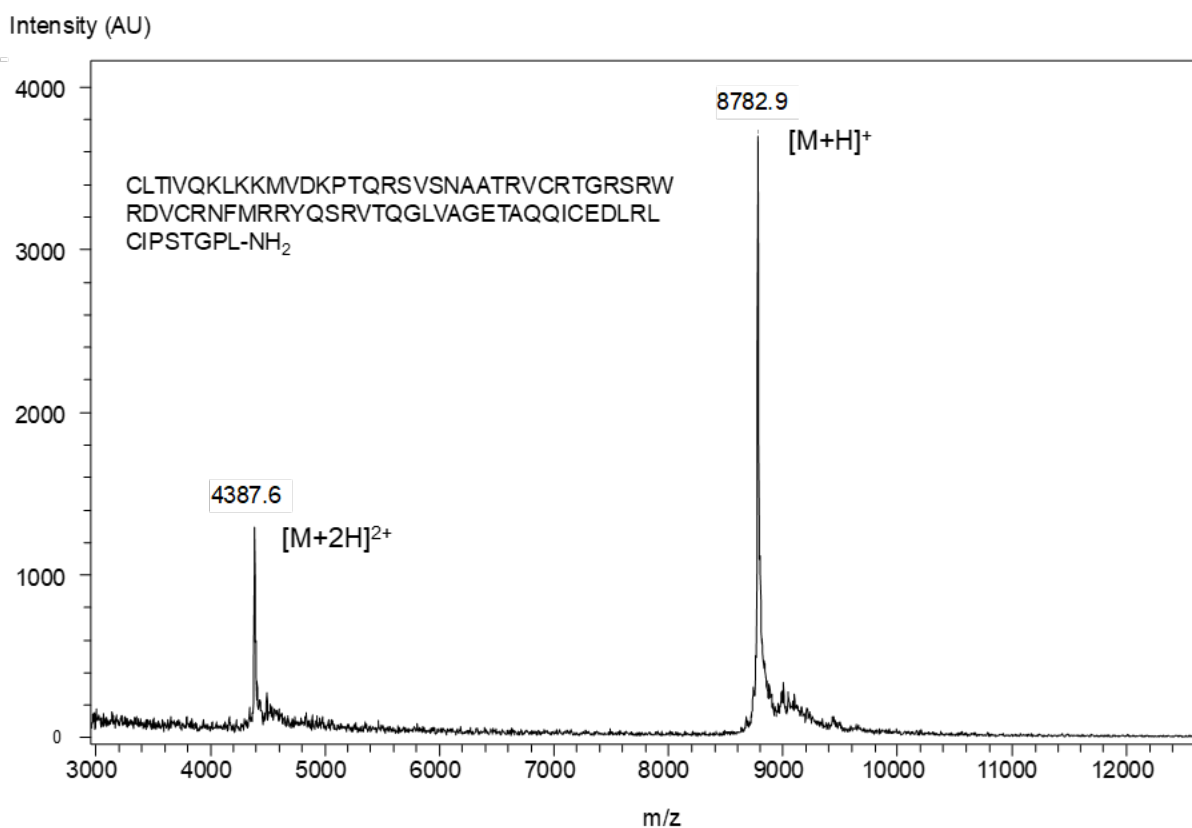


Figure S65. MALDI-TOF of purified peptide **33**. Matrix sinapinic acid, positive detection mode, [M+H]⁺ calcd (average) 8783.3, observed mass: 8782.9 (average).

6.4 Synthesis of 15 kD granulysin

TCEP-HCl (4.30 mg, 15 μmol, 200 mM final concentration) was dissolved in 6 M Gn•HCl, 0.1 M pH 7.2 sodium phosphate buffer (75 μL). Then, the diselenide **7d** (3.12 mg, 7.5 μmol, 100 mM final concentration) was dissolved in the above solution. The pH was adjusted to 3.9 using NaOH (6 M).

Peptide **9e** (1.48 mg, 0.2 μmol, 3.5 mM, 1.1 equiv) and peptide **33** (2.12 mg, 0.2 μmol, 3 mM) were dissolved in the above solution (58 μL). The reaction mixture was shaken at 37 °C under nitrogen atmosphere and monitored by HPLC.

After 48 h, the mixture was diluted with water-TFA 0.1% (4 mL). The crude peptide was purified by RP-HPLC (eluent A = water containing 0.1% TFA, eluent B = acetonitrile in water 4/1 containing 0.1% TFA, 60 °C, detection at 215 nm, 6 mL/min, 0 to 30% eluent B in 10 min, then 30 to 50% eluent B in 60 min, C3 Zorbax column) to give 0.791 mg of linear 15 kDa granulysin polypeptide (23%).

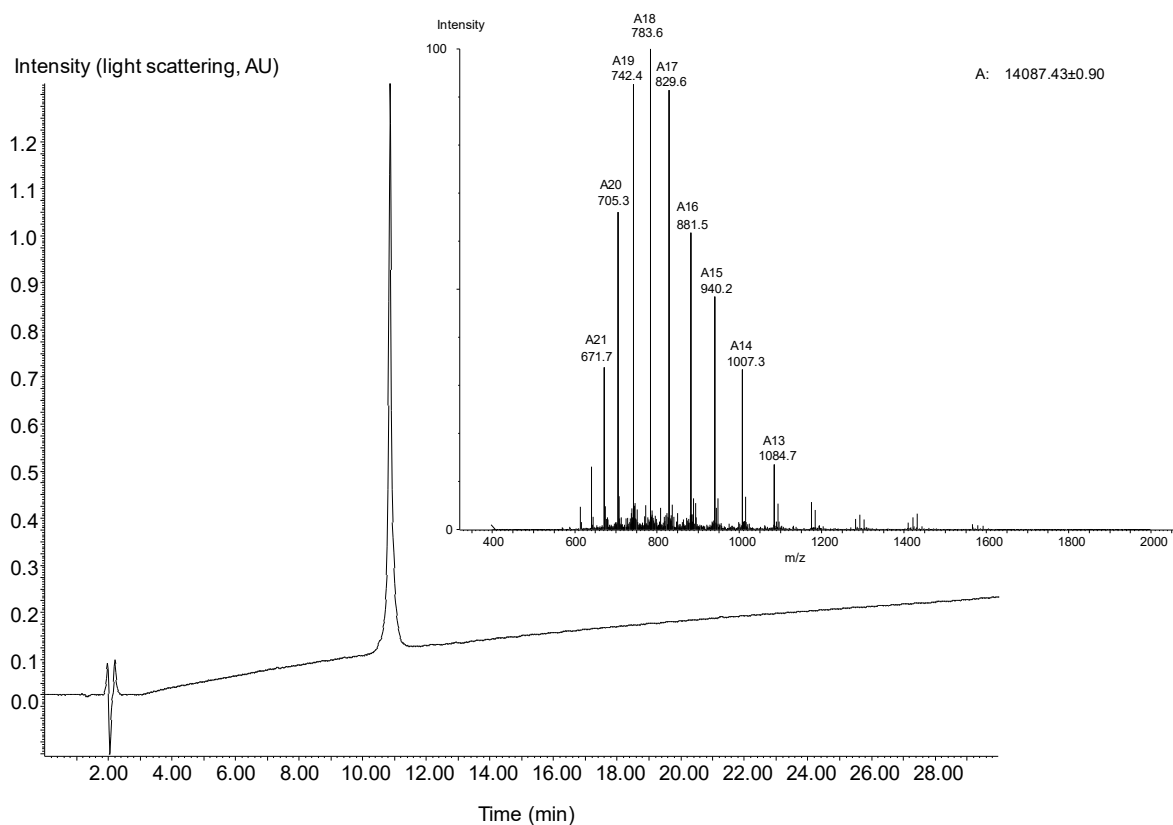
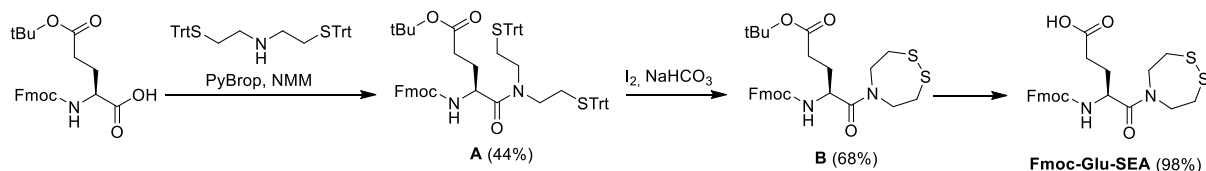


Figure S66. LC-MS analysis of peptide **15 kDa GLNY**. LC trace, eluent A 0.10% TFA in water, eluent B 0.10% TFA in CH₃CN/water: 4/1 by vol. C3 Zorbax 300SB Å 3.5 µm (4.6 × 150 mm) column, gradient 0-100% B in 30 min (60 °C, 1 mL/min, detection 215 nm). MS trace. [M+H]⁺ m/z calcd. (average) 14088.22, found 14087.43.

Automated SEA solid phase protein synthesis (Chapter VI)

1. Synthesis of the linker Fmoc-Glu-SEA^{off}

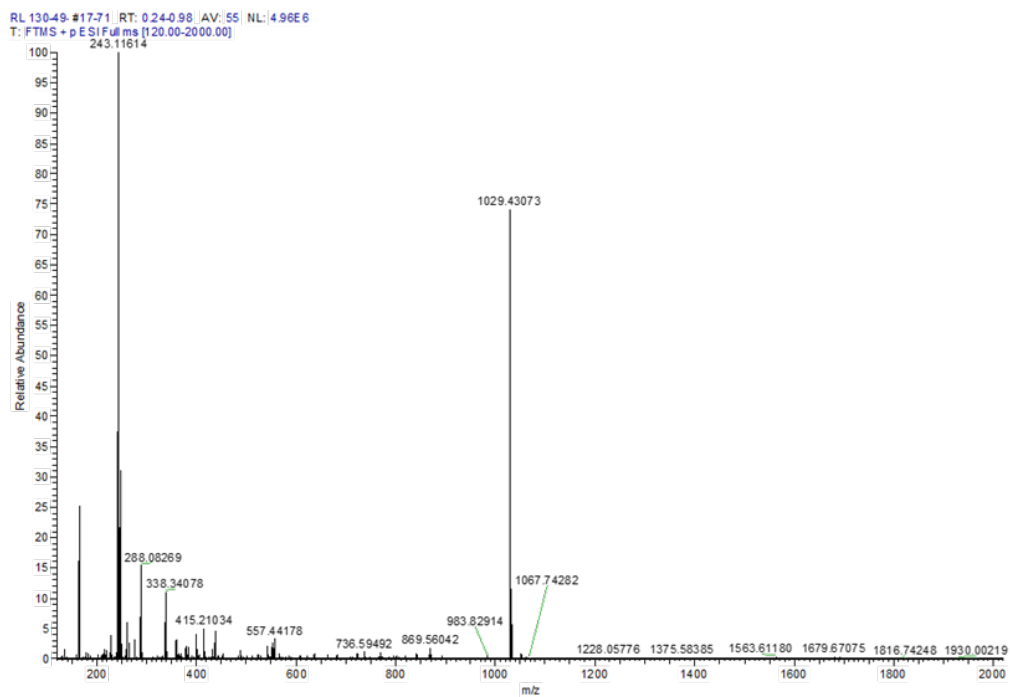


Scheme S6. Synthesis of Fmoc-Glu-SEA.

1.1 Synthesis of *tert*-butyl 5-(bis(2-tritylsulfanylethyl)amino)-4-(9H-fluoren-9-ylmethoxycarbonylamino)-5-oxo-pentanoate A

Fmoc-Glu(OtBu)-OH (2.43 g, 5.72 mmol, 1.5 equiv) was dissolved in DMF (5 mL) under argon atmosphere. The *bis*(2-sulfanylethyl)amine (2.38 g, 3.82 mmol, 1 equiv) previously dissolved in DCM (10 mL) was added to the solution of amino acid. Then, NMM (1.3 mL, 11.4 mmol, 3 equiv) and PyBrop (2.7 g, 5.72 mmol, 1.5 equiv) were added and the reaction mixture was stirred overnight at room temperature.

DCM was added (20 mL) afterwards and the obtained organic layer was washed with KH₂PO₄ (10% massique). The combined organic layers were dried over anhydrous Na₂SO₄ and the solvent was evaporated under reduced pressure. Purification of the crude by column chromatography (cyclohexane/ethyl acetate/acetic acid: 80/20/0.1%) provided the expected compound **A** (1.73 g, 44%).



m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	RDB equiv.	Composition
165,0696	1248012,4	24,47				
243,1161	5099771,5	100				
247,0562	1546787,1	30,33				
1029,431	3727155,5	73,08	1029,433	-2,17	35,5	C66 H65 O5 N2 S2
1030,434	2366470,3	46,4	1030,435	-1	35,5	C66 H64 O8 N S

Figure S67. HR-MS analysis for purified **A**. Calcd. for $[M+H]^+$: 1029.433, observed mass: 1029.431 (monoisotopic).

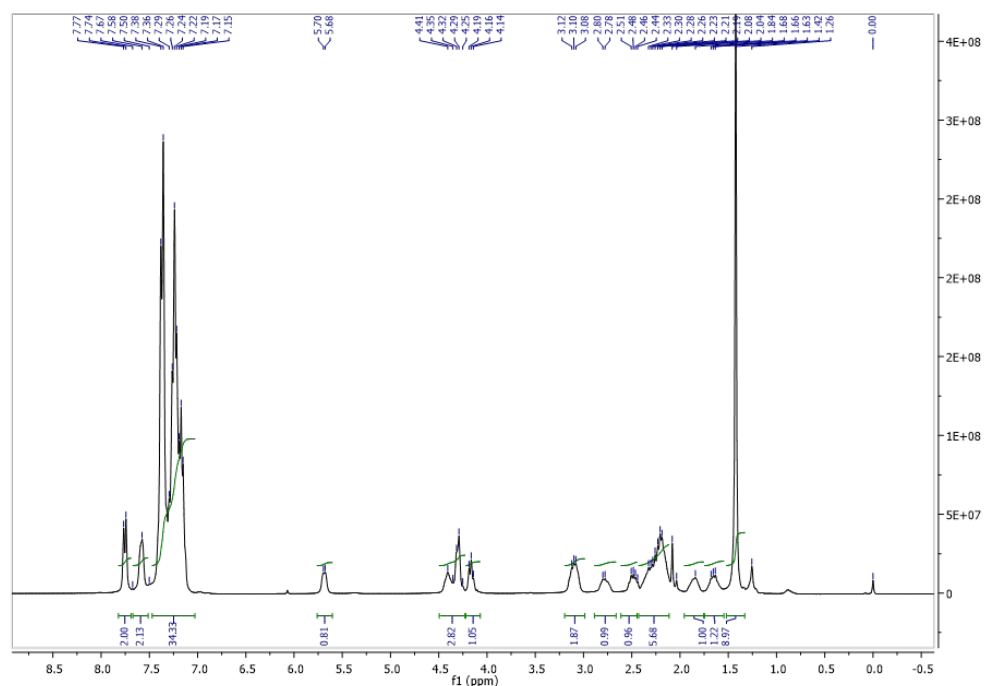


Figure S68. ^1H NMR (300 MHz) spectrum for compound **A** (CDCl_3 , 291 K). δ 7.75 (d, $J = 7.3$ Hz, 2H), 7.58 (s, 2H), 7.47 – 7.03 (m, 35H), 5.69 (d, $J = 5.0$ Hz, 1H), 4.50 – 4.23 (m, 3H), 4.16 (t, $J = 6.7$ Hz, 1H), 3.20 – 2.99 (m, 2H), 2.89 – 2.66 (m, 1H), 2.61 – 2.45 (m, 1H), 2.43 – 2.11 (m, 6H), 1.95 – 1.75 (m, 1H), 1.75 – 1.55 (m, 1H), 1.42 (s, 9H).

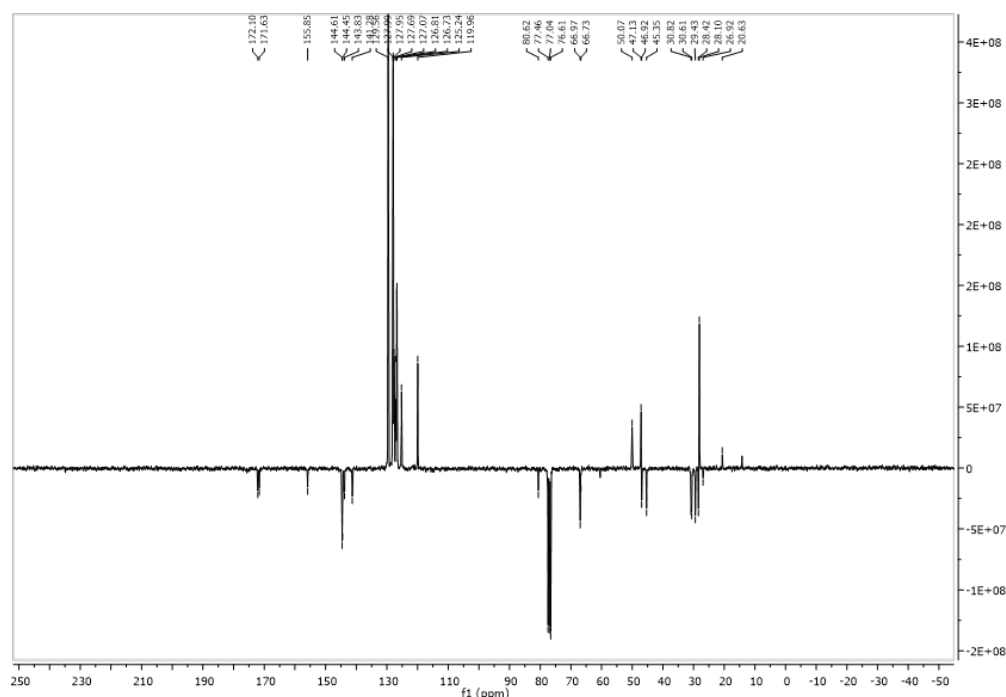


Figure S69. ^{13}C JMOD (75 MHz) spectrum for compound **A** (CDCl_3 , 291 K). δ 172.10 (s), 171.63 (s), 155.85 (s), 144.61 (s), 144.45 (s), 144.01 (s), 143.83 (s), 141.28 (s), 129.56 (s), 127.97 (d, $J = 2.8$ Hz), 127.69 (s), 127.07 (s), 126.77 (d, $J = 5.7$ Hz), 125.24 (s), 119.96 (s), 80.62 (s), 66.85 (d, $J = 18.4$ Hz), 50.07 (s), 47.13 (s), 47.13 (s), 45.35 (s), 30.72 (d, $J = 15.5$ Hz), 29.43 (s), 28.10 (s), 26.82 (d, $J = 15.5$ Hz), 20.63 (s).

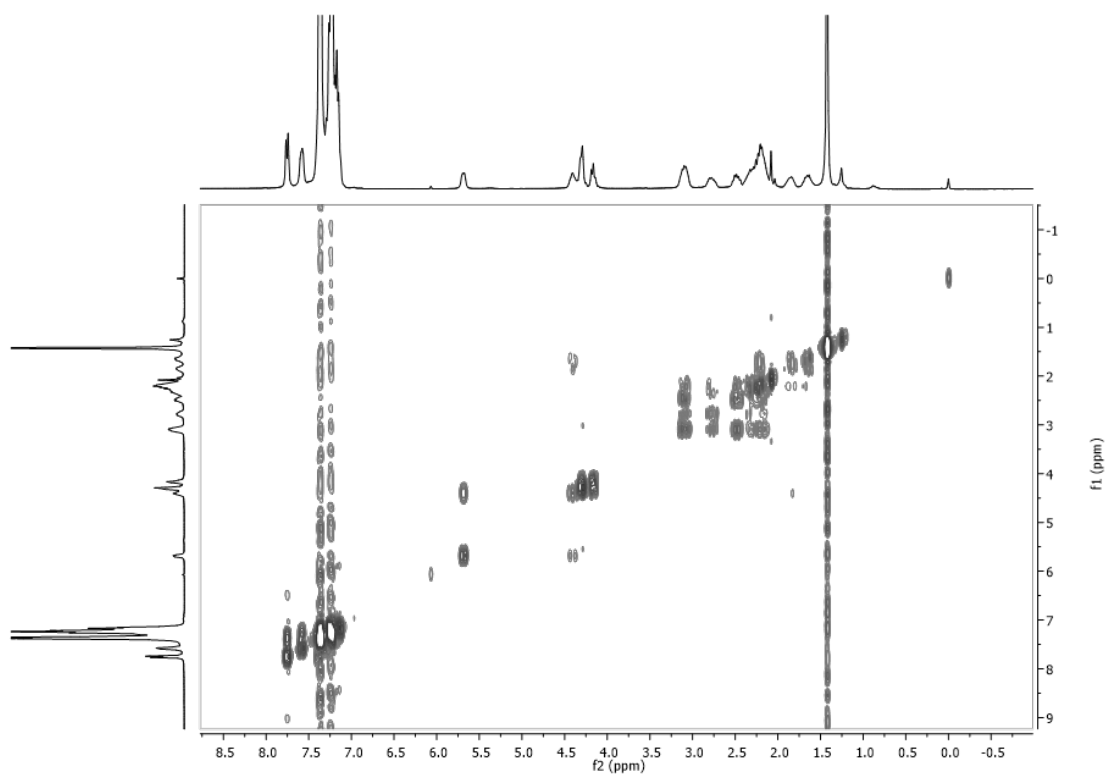


Figure S70. ^1H - ^1H COSY spectrum for compound **A** (CDCl_3 , 291 K).

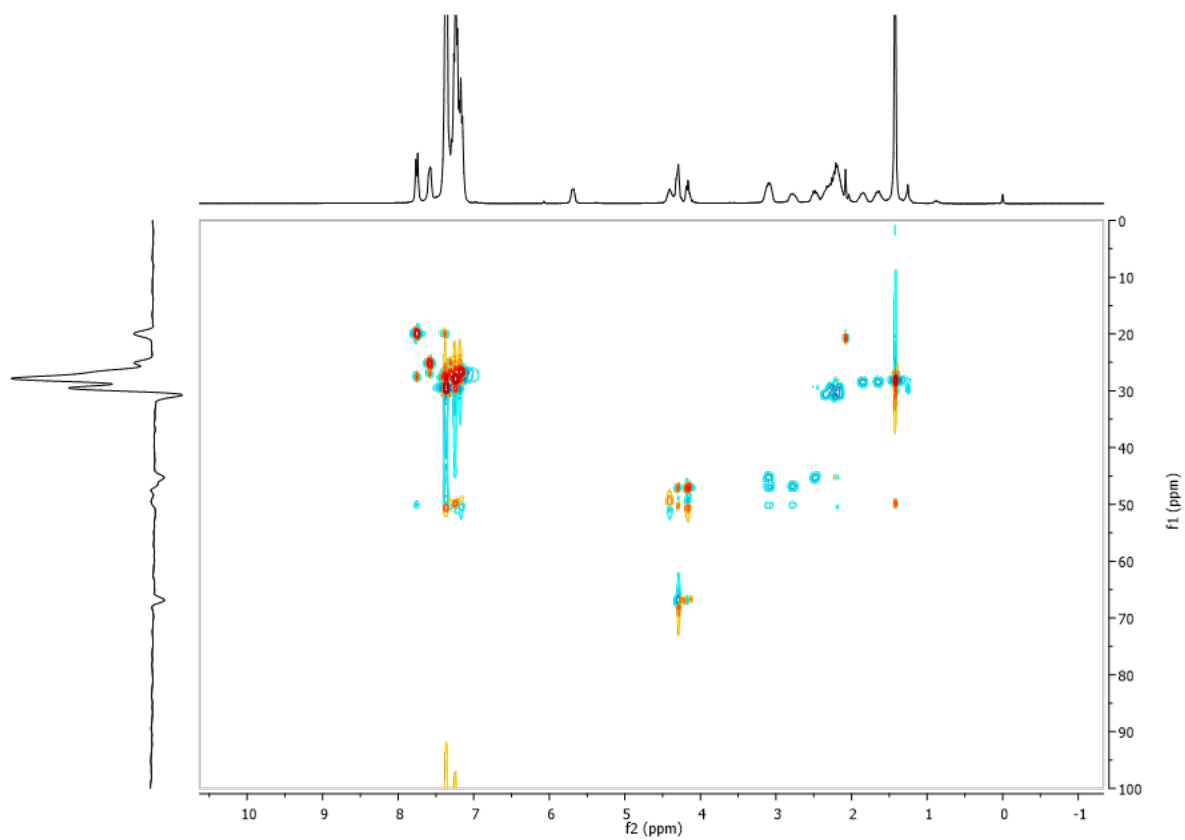
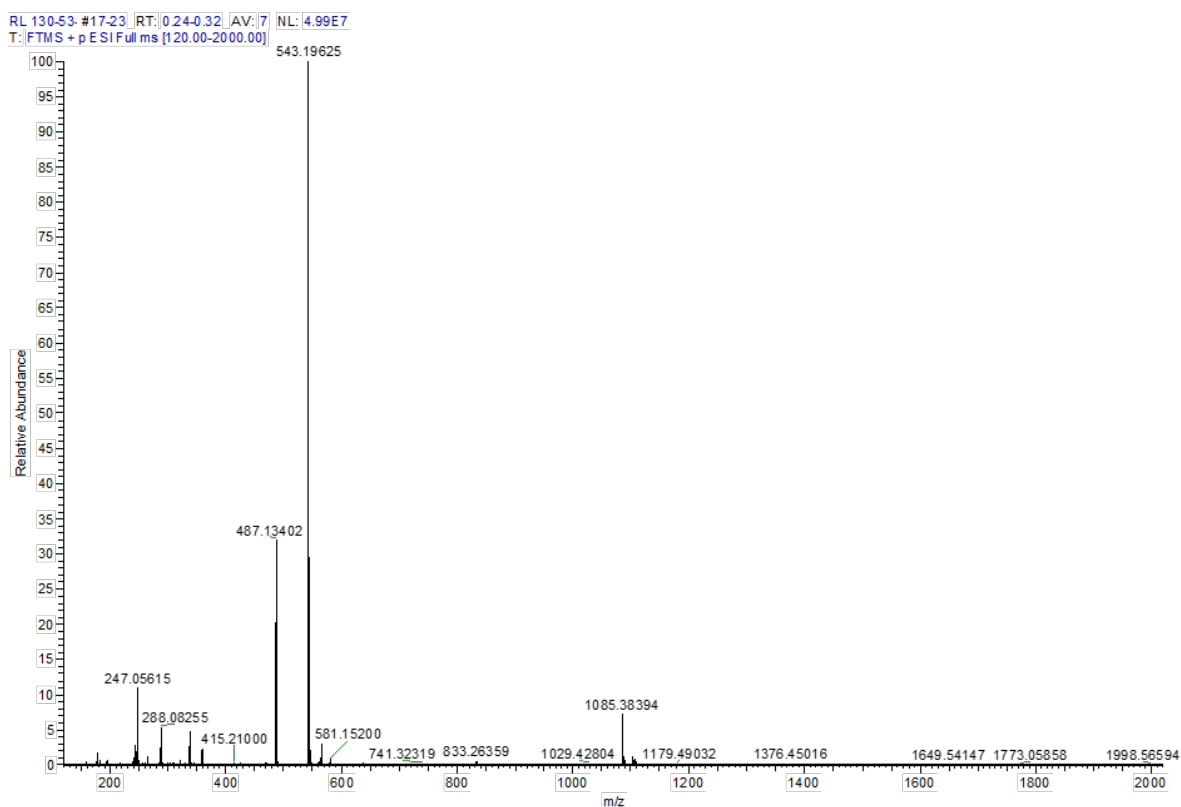


Figure S71. ^1H - ^{13}C HSQC spectrum for compound **A** (CDCl_3 , 291 K).

1.2 Synthesis of *tert*-butyl 5-(1,2,5-dithiazepan-5-yl)-4-(9H-fluoren-9-ylmethoxycarbonylamino)-5-oxo-pentanoate **B**

Compound **A** (695 mg, 0.657 mmol, 1 equiv) was dissolved in DCM (10 mL), I₂ (688 mg, 2.7 mmol, 4 equiv) and NaHCO₃ (234 mg, 2.7 mmol, 4 equiv) were dissolved separately in DCM (25 mL for each). The three solutions were combined and the reaction mixture was stirred during 1 h.

Afterwards, Na₂S₂O₃ (3 M) was added until decoloration to quench I₂. Water (15 mL) and DCM (10 mL) were added. The obtained aqueous layer was extracted with DCM (3 × 10 mL). The combined organic layers were dried over anhydrous Na₂SO₄ and the solvent was evaporated under reduced pressure. Purification of the crude by column chromatography (cyclohexane/ethyl acetate 95:5) provided the desired product **B** (249 mg, 68%).



m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	RDB equiv.	Composition
247,05618	5692899	10,6	247,05695	-3,09	3,5	C9 H15 O2 N2 S2
543,19635	53730064	100	543,19819	-3,39	12,5	C28 H35 O5 N2 S2
1085,38428	4070724	100	1085,3891	-4,45	24,5	C56 H69 O10 N4 S4

Figure S71. HR-MS analysis for purified **B**. Calcd. for [M+H]⁺: 543.198, observed mass: 543.196 (monoisotopic).

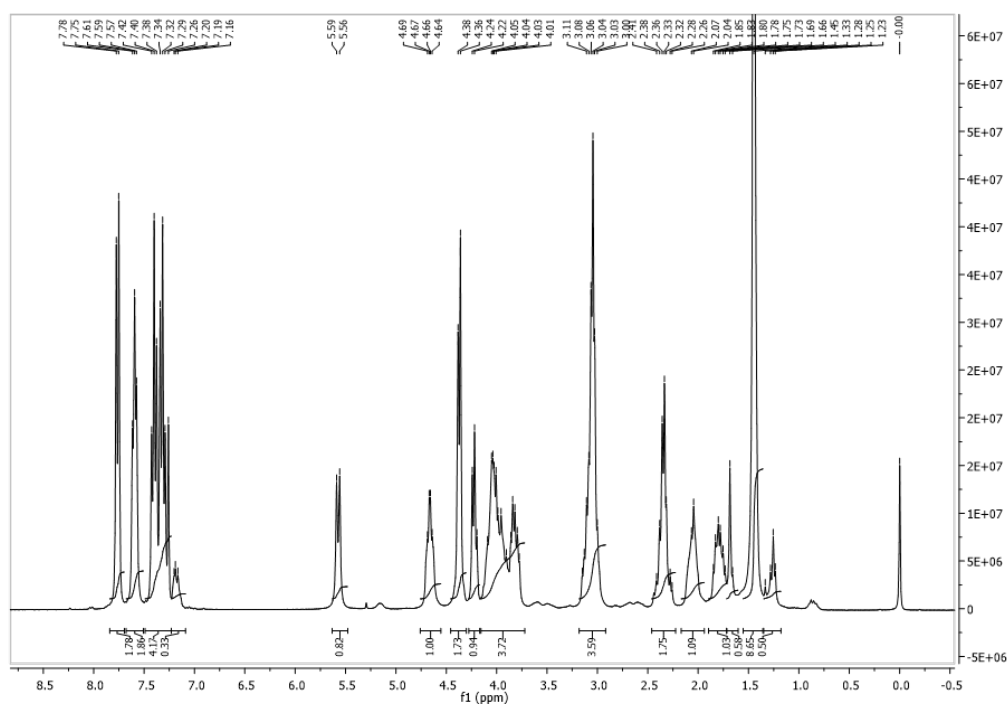


Figure S73. ^1H NMR (300 MHz) spectrum for compound **B** (CDCl_3 , 291 K). δ 7.76 (d, $J = 7.3$ Hz, 2H), 7.59 (t, $J = 6.0$ Hz, 2H), 7.49 – 7.23 (m, 4H), 7.23 – 7.09 (m, 1H), 5.57 (d, $J = 9.1$ Hz, 1H), 4.76 – 4.55 (m, 1H), 4.37 (d, $J = 7.0$ Hz, 2H), 4.22 (t, $J = 6.9$ Hz, 1H), 4.15 – 3.72 (m, 4H), 3.18 – 2.92 (m, 4H), 2.46 – 2.22 (m, 2H), 2.17 – 1.94 (m, 1H), 1.90 – 1.72 (m, 1H), 1.72 – 1.60 (m, 1H), 1.45 (s, 9H), 1.35 – 1.18 (m, 1H).

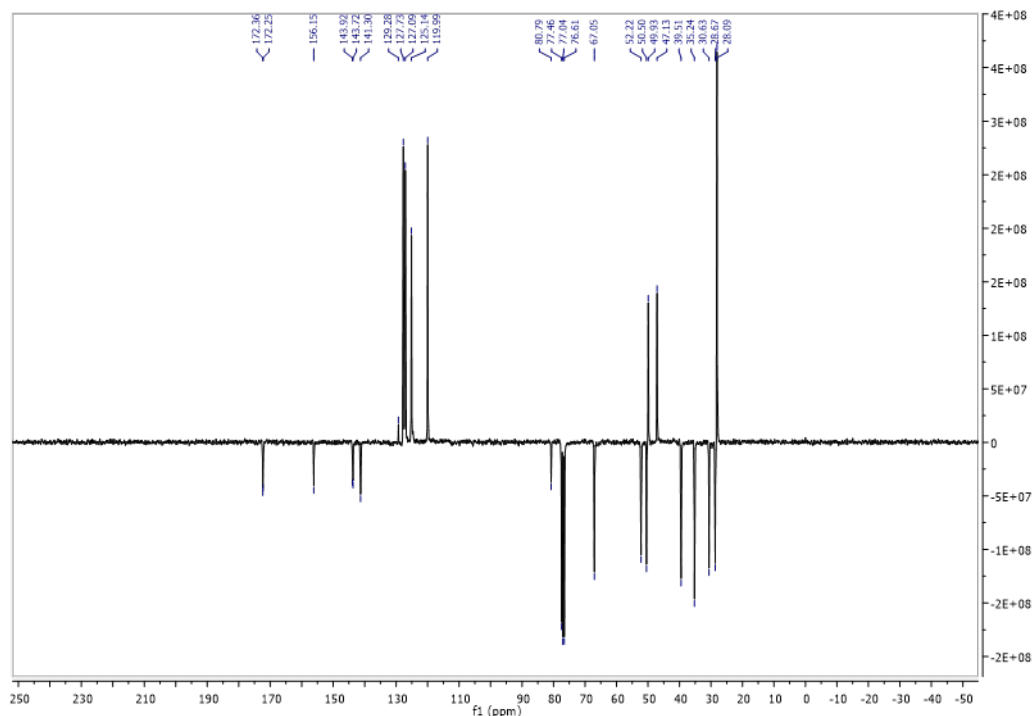


Figure S74. ^{13}C JMOD (75 MHz) spectrum for compound **B** (CDCl_3 , 291 K). δ 172.30 (d, $J = 8.5$ Hz), 156.15 (s), 143.82 (d, $J = 15.5$ Hz), 141.30 (s), 129.28 (s), 127.73 (s), 127.09 (s), 125.14 (s), 119.99 (s), 80.64 (d, $J = 22.6$ Hz), 67.05 (s), 52.22 (s), 50.50 (s), 49.93 (s), 47.13 (s), 39.51 (s), 35.24 (s), 30.63 (s), 28.67 (s), 28.09 (s).

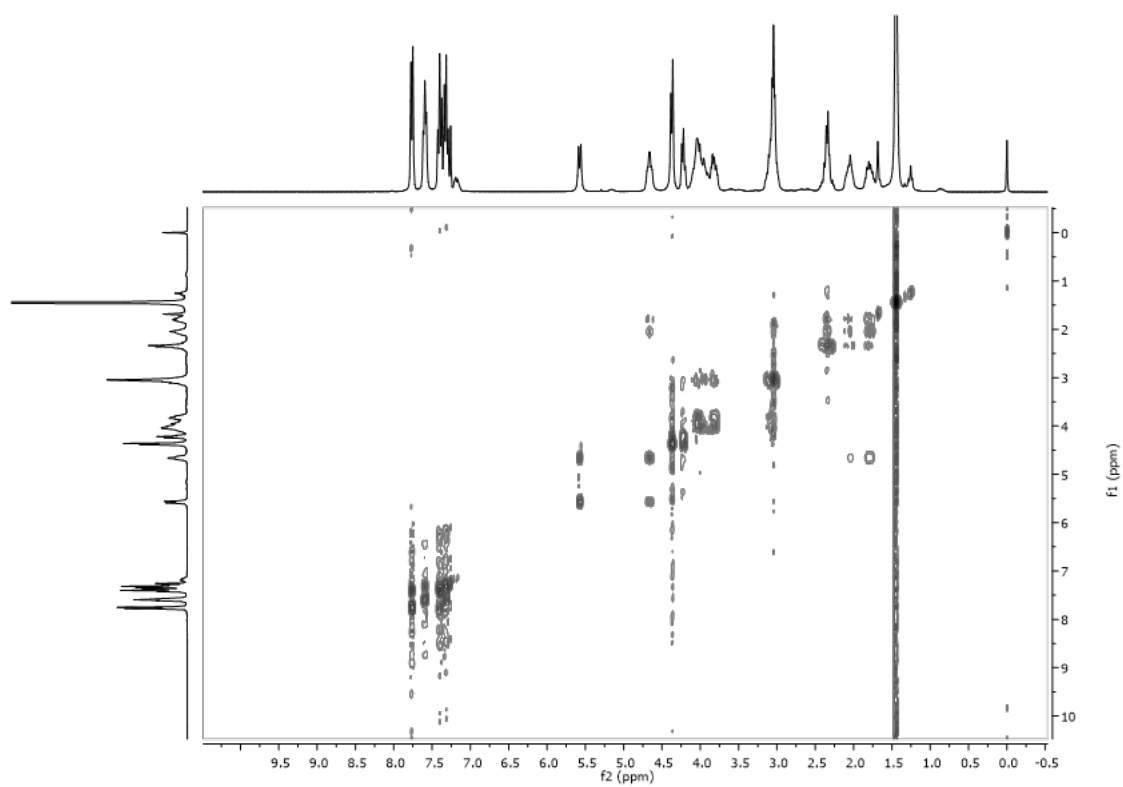


Figure S75. ^1H - ^1H COSY spectrum for compound **B** (CDCl_3 , 291 K).

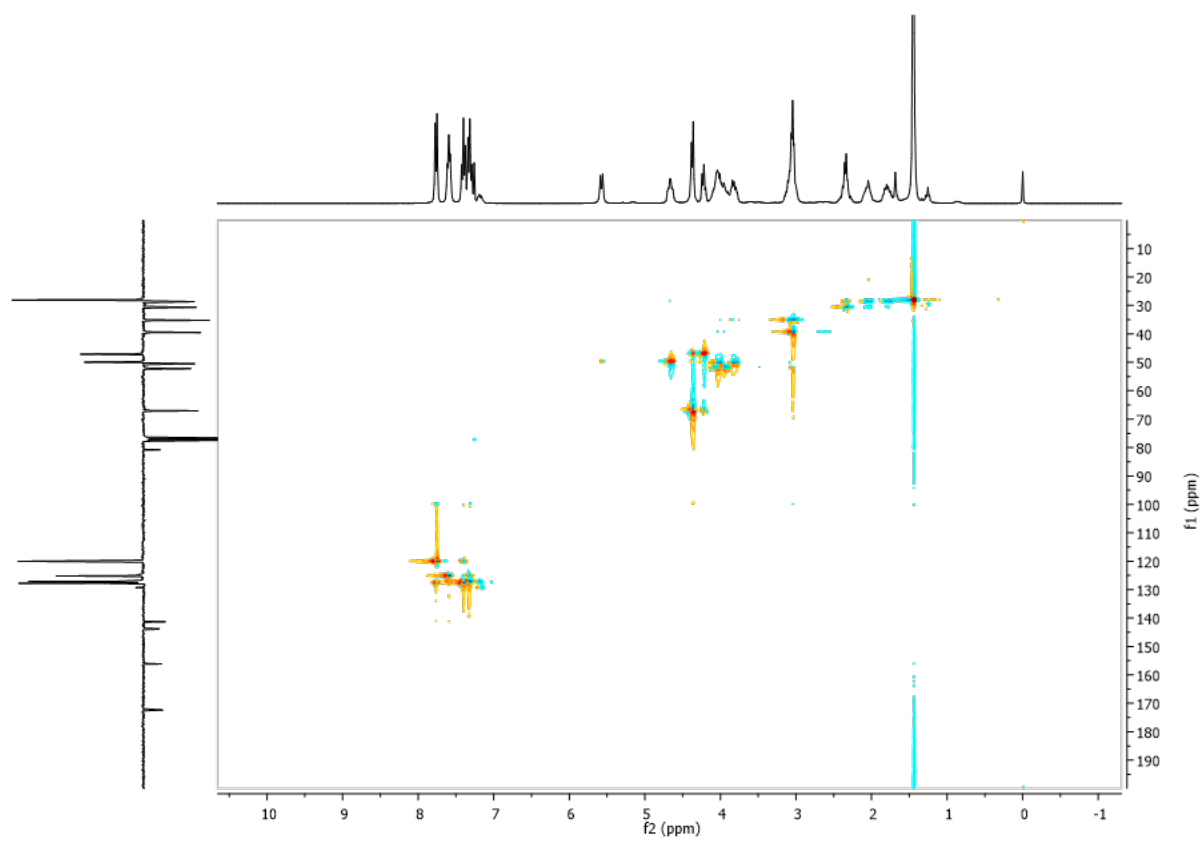
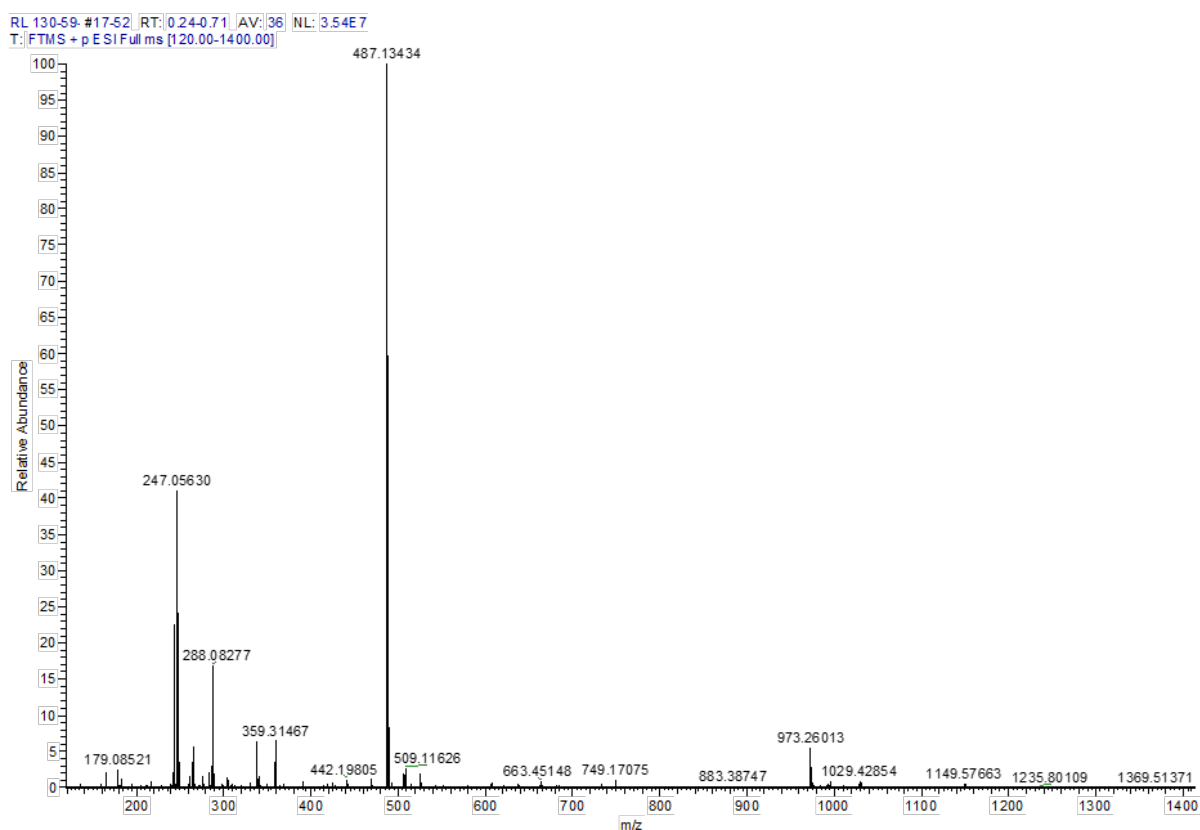


Figure S76. ^1H - ^{13}C HSQC spectrum for compound **B** (CDCl_3 , 291 K).

1.3 Synthesis of 5-(1,2,5-dithiazepan-5-yl)-4-(9H-fluoren-9-ylmethoxycarbonylamino)-5-oxo-pentanoic acid : Fmoc-Glu-SEA^{off}

Compound **B** (700 mg, 1.3 mmol, 1 equiv) was deprotected using TFA/H₂O/DMS: 95/2.5/2.5 by vol for 30 min (5 mL). TFA was evaporated under reduced pressure and water (20 mL) was added. The obtained aqueous layer was extracted with DCM (3 × 30 mL). The combined organic layers were dried over anhydrous Na₂SO₄ and the solvent was evaporated under reduced pressure. **Fmoc-Glu-SEA^{off}** was obtained without further purification step (619 mg, 98%).



m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	RDB equiv.	Composition
243,1164	8297674,5	22,03				
247,0563	14740088	39,14				
288,0828	6223051	16,52				
487,1343	37658580	100	487,1356	-2,59	12,5	C24 H27 O5 N2 S2
973,2602	2057939,3	100	973,2639	-3,84	24,5	C48 H53 O10 N4 S4

Figure S77. HR-MS analysis for purified **Fmoc-Glu-SEA^{off}**. Calcd. for [M+H]⁺: 487.1135, observed mass: 487.134 (monoisotopic).

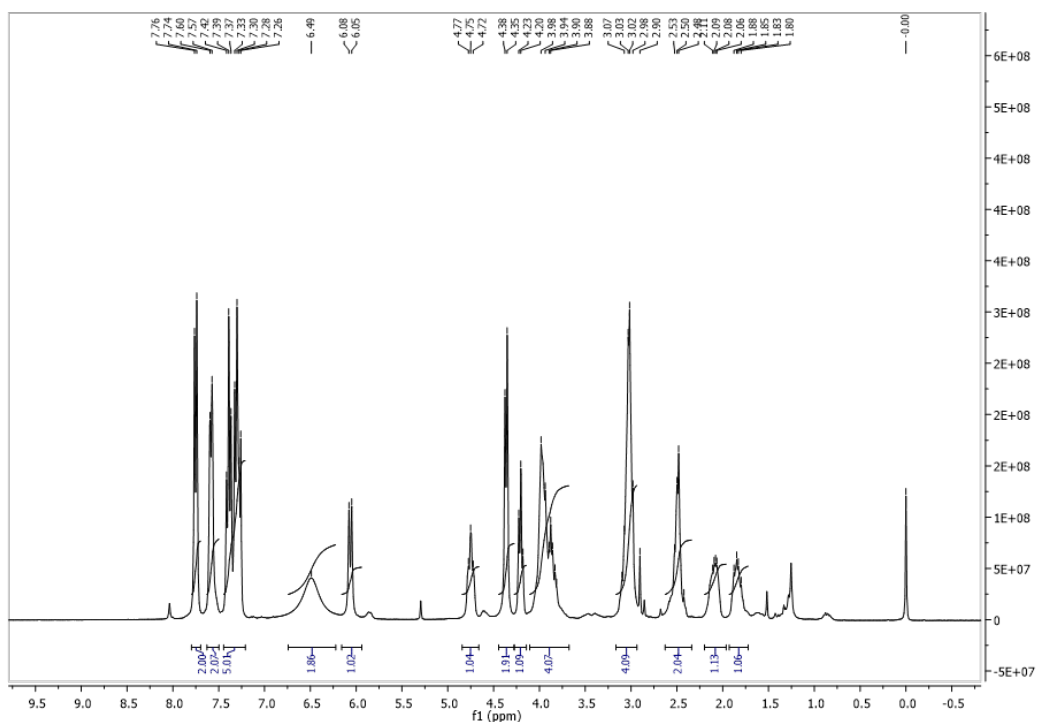


Figure S78. ^1H NMR (300 MHz) spectrum for compound **Fmoc-Glu-SEA^{off}** (CDCl_3 , 291 K). δ 7.76 (d, J = 7.3 Hz, 2H), 7.59 (t, J = 6.0 Hz, 2H), 7.49 – 7.23 (m, 4H), 7.23 – 7.09 (m, 1H), 5.57 (d, J = 9.1 Hz, 1H), 4.76 – 4.55 (m, 1H), 4.37 (d, J = 7.0 Hz, 2H), 4.22 (t, J = 6.9 Hz, 1H), 4.15 – 3.72 (m, 4H), 3.18 – 2.92 (m, 4H), 2.46 – 2.22 (m, 2H), 2.17 – 1.94 (m, 1H), 1.90 – 1.72 (m, 1H), 1.72 – 1.60 (m, 1H), 1.45 (s, 9H), 1.35 – 1.18 (m, 1H).

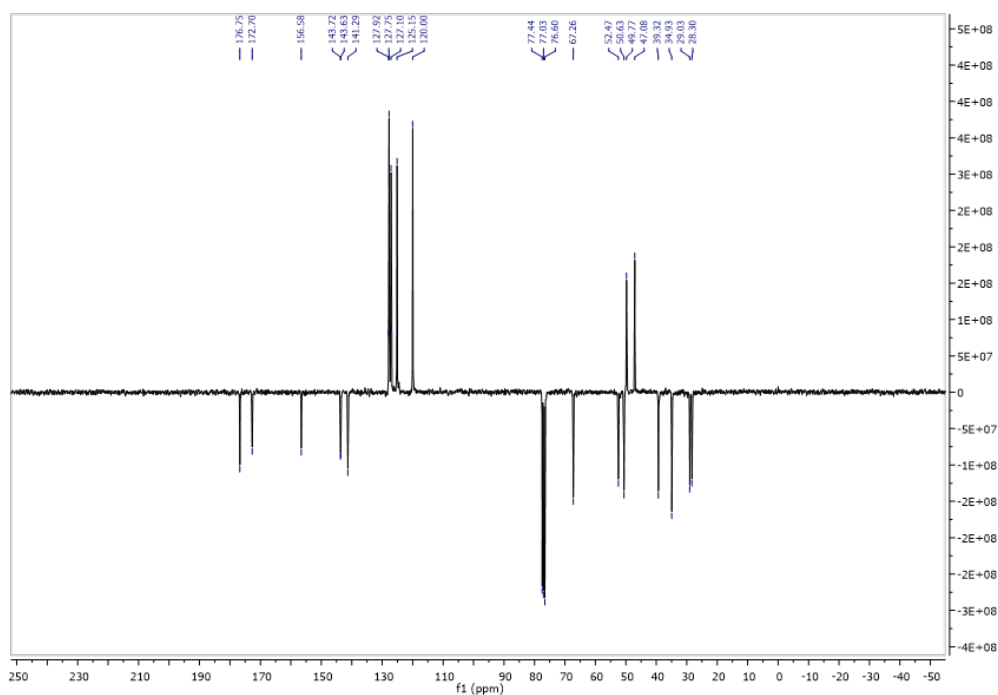


Figure S79. ^{13}C JMOD (75 MHz) spectrum for compound **Fmoc-Glu-SEA^{off}** (CDCl_3 , 291 K). δ 172.30 (d, J = 8.5 Hz), 156.15 (s), 143.82 (d, J = 15.5 Hz), 141.30 (s), 129.28 (s), 127.73 (s), 127.09 (s), 125.14 (s),

119.99 (s), 80.64 (d, $J = 22.6$ Hz), 67.05 (s), 52.22 (s), 50.50 (s), 49.93 (s), 47.13 (s), 39.51 (s), 35.24 (s), 30.63 (s), 28.67 (s), 28.09 (s).

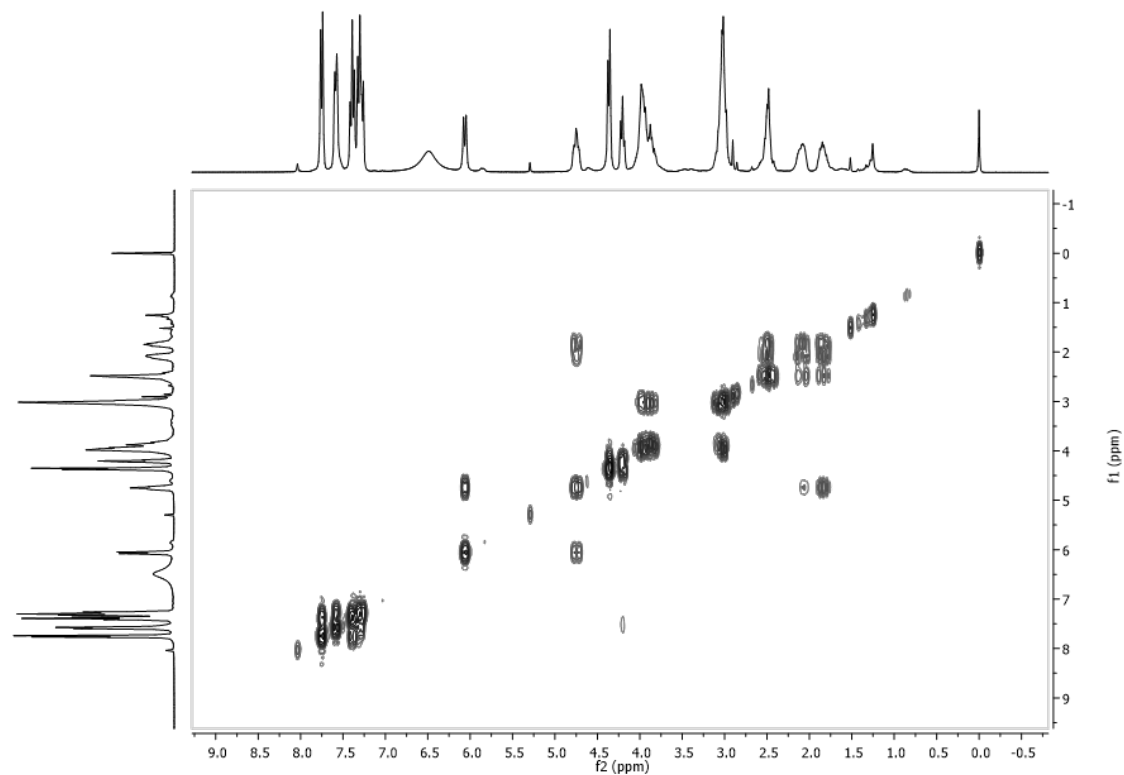


Figure S80. ^1H - ^1H COSY spectrum for compound **Fmoc-Glu-SEA^{off}** (CDCl_3 , 291 K).

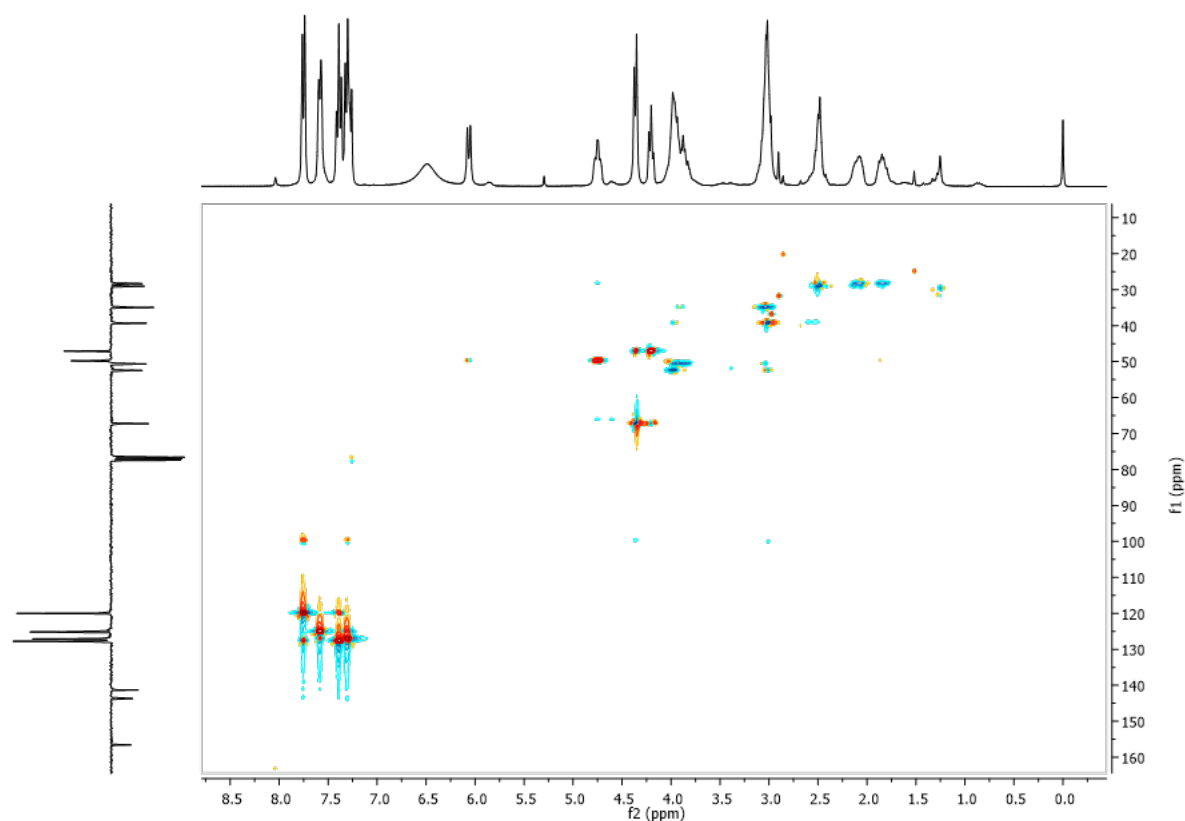


Figure S81. ^1H - ^{13}C HSQC spectrum for compound **Fmoc-Glu-SEA^{off}** (CDCl_3 , 291 K).

2. Synthesis of the solid support

2.1 General procedure for the coupling of the linker on the resin using a ratio 1/3 Fmoc-Glu-SEA^{off} and 2/3 Boc-Glu(OtBu)-OH

Rink amide ChemMatrix LL (278 mg, 0.05 mmol, 0.18 mmol/g) was conditioned in CH_2Cl_2 (3 x 2 min) and DMF (3 x 2 min). Fmoc-Glu-SEA^{off} (16.22 mg, 0.03 mmol) was dissolved in the minimal volume of DCM. Boc-Glu(OtBu)-OH (20.22 mg, 0.06 mmol) and PyBop (52.04 mg, 0.1 mmol, 2 equiv) were dissolved in the minimal volume of DMF. The two amino acid solutions were combined and DIEA (34.25 μL , 0.15 mmol, 3 equiv) was added. The aforementioned solution was agitated at room temperature during 1 min and then added to the beads. Then Pybop was added. The resin was agitated during 1 h and washed with DMF (3 x 2 min). The coupling step was controlled using the acetaldehyde/chloranil¹³ and TNBS¹⁴ assays.

2.2 Determination of Fmoc loading by UV spectrophotometry

Three samples of the Fmoc-Glu-SEA ChemMatrix resin (3–5 mg for each) were weighed in glass test tubes. Piperidine/DMF solution (20 % by vol, 1 mL) were added to each tube. The solutions were agitated for 10 min at room temperature. The supernatants were transferred to 10 mL volumetric flask. The

process was repeated three times and piperidine/DMF solution was added to obtain a final volume (V) of 10 mL. The absorbance (Abs) at 290 nm of each solution was immediately measured with a UV spectrophotometer. The piperidine/DMF solution was used as the blank. The loading of each sample was calculated using the following equation :

$$\text{Loading (mmol/g)} = \frac{\text{Abs} \times V}{\epsilon \times l}$$

With ϵ = molar extinction coefficient = $4.04 \text{ L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$ and l = length of optical path = 1 cm.

The mean of the three values obtained for the three samples give the loading of the resin.

2.3 Deprotection and acetylation

The resin was conditioned in DMF (3 x 2 min). The N_α -Fmoc protecting group was deprotected using 20% piperidine solution in DMF (2 x 2 min). The resin was then treated with $\text{Ac}_2\text{O}/\text{DIEA}/\text{CH}_2\text{Cl}_2$: 10/5/85 by volume (1 x 1 min then 1 x 10 min) to cap the N_α amine. Finally, the resin was washed with DMF (3 x 2 min), DCM (3 x 2 min) and Et_2O (3 x 2 min) and dried in *vacuo*.

References

- (1) Ollivier, N.; Dheur, J.; Mhidia, R.; Blanpain, A.; Melnyk, O. *Org. Lett.* **2010**, *12* (22), 5238–5241.
- (2) Amelung, W.; Brodowski, S. *Anal. Chem.* **2002**, *74* (13), 3239–3246.
- (3) Ollivier, N.; Vicogne, J.; Vallin, A.; Drobecq, H.; Desmet, R.; El Mahdi, O.; Leclercq, B.; Goormachtigh, G.; Fafeur, V.; Melnyk, O. *Angew. Chem. Int. Ed.* **2012**, *51* (1), 209–213.
- (4) Ollivier, N.; Raibaut, L.; Blanpain, A.; Desmet, R.; Dheur, J.; Mhidia, R.; Boll, E.; Drobecq, H.; Pira, S. L.; Melnyk, O. *J. Pept. Sci.* **2014**, *20* (2), 92–97.
- (5) Dheur, J.; Ollivier, N.; Vallin, A.; Melnyk, O. *J. Org. Chem.* **2011**, *76* (9), 3194–3202.
- (6) Raibaut, L.; Drobecq, H.; Melnyk, O. *Org. Lett.* **2015**, *17* (14), 3636–3639.
- (7) Ollivier, N.; Blanpain, A.; Boll, E.; Raibaut, L.; Drobecq, H.; Melnyk, O. *Org. Lett.* **2014**, *16* (15), 4032–4035.
- (8) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H. et al. *Gaussian 09. Revision A.02*, Gaussian, Inc.; Wallingford (CT, USA), 2009.
- (9) Fukui, K. *Acc. Chem. Res.* **1981**, *14* (12), 363–368.
- (10) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94* (7), 2027–2094.
- (11) Abbas, M.; Bethke, J.; Wessjohann, L. A. *Chem Commun* **2006**, No. 5, 541–543.
- (12) Desmet, R.; Pauzuolis, M.; Boll, E.; Drobecq, H.; Raibaut, L.; Melnyk, O. *Org. Lett.* **2015**, *17* (13), 3354–3357.
- (13) Vojkovsky, T. *Pept. Res.* **1995**, *8*, 236–237.
- (14) Hancock, W. S.; Battersby, J. E. *Anal. Biochem.* **1976**, *71*, 260–264.

ANNEX B: ARTICLES

A statistical view of protein chemical synthesis using NCL and extended methodologies

Vangelis Agouridas^{*a}, Ouafâa El Mahdi^b, Marine Cargoët^a and Oleg Melnyk^{*a}

^a Univ. Lille, CNRS, Institut Pasteur de Lille, UMR 8161 - M3T – Mechanisms of Tumorigenesis and Target Therapies, F-59000 Lille, France.

^b Université Sidi Mohamed Ben Abdellah, Morocco.

Native chemical ligation and extended methodologies are the most popular chemoselective reactions for protein chemical synthesis. Their combination with desulfurization techniques can give access to small or challenging proteins that are exploited in a large variety of research areas. In this report, we have conducted a statistical review of their use for protein chemical synthesis in order to provide a flavor of the recent trends and identify the most popular chemical tools used by protein chemists. To this end, a Protein Chemical Synthesis (PCS) database (<http://pcs-db.fr>) was created by collecting a set of relevant data from more than 450 publications covering the period 1994-2017. A preliminary account of what this database tells us is presented in this report.

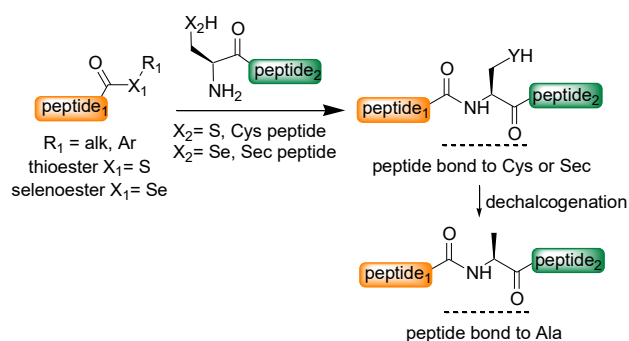
1. Introduction

Proteins can be synthesized from short unprotected peptide segments by performing chemoselective peptide bond forming reactions in water.¹ Among the few chemoselective peptide bond forming reactions that have been designed so far,²⁻⁶ the native chemical ligation (NCL) is certainly the most popular ligation technique in the field.⁷ NCL involves the chemoselective reaction of a C-terminal peptide thioester with an N-terminal cysteinyl (Cys) peptide in water at neutral pH. The reaction produces a native peptide bond to cysteine, and has been extended to the synthesis of peptide bonds to selenocysteine.⁸⁻¹¹ The low frequency of Cys residues in proteins has stimulated the development of novel chemistries for extending NCL to the formation of nearly all types of peptide junctions. This has been achieved through desulfurization^{12,13} or deselenization¹⁴ of thiol or selenol derivatives of amino acids, and by the design of various thiol auxiliary approaches.^{15,16}

In parallel with these developments, several groups have reported novel thioester surrogates aimed at facilitating the production of C-terminal thioesters and thus the assembly of the target proteins. In this respect, the usefulness of *N*-acyl-benzimidazolinone (Nbz)^{17,18} or hydrazide^{19,20} methods is noteworthy. Another significant advance in the field of protein chemical synthesis is the recent design of several *O,S,N,S* or

N,Se-acyl shift systems which can act as thioester surrogates *in*

situ (for recent reviews see^{21,22}). The combination of some of these systems with NCL has been shown to significantly facilitate the assembly of challenging proteins.



Scheme 1. Principle of native chemical ligation (NCL).

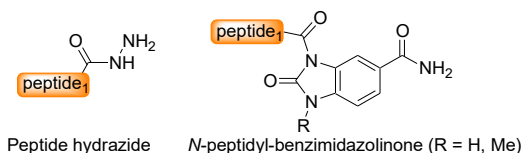


Figure 1. Structure of peptide hydrazides and *N*-peptidyl-benzimidazolinones used as thioester precursors.

* Corresponding author. Tel.: +33-3-20-87-12-14; e-mail: oleg.melnyk@ibl.cnrs.fr

* Corresponding author. Tel.: +33-3-20-87-15; e-mail: vangelis.agouridas@ibl.cnrs.fr

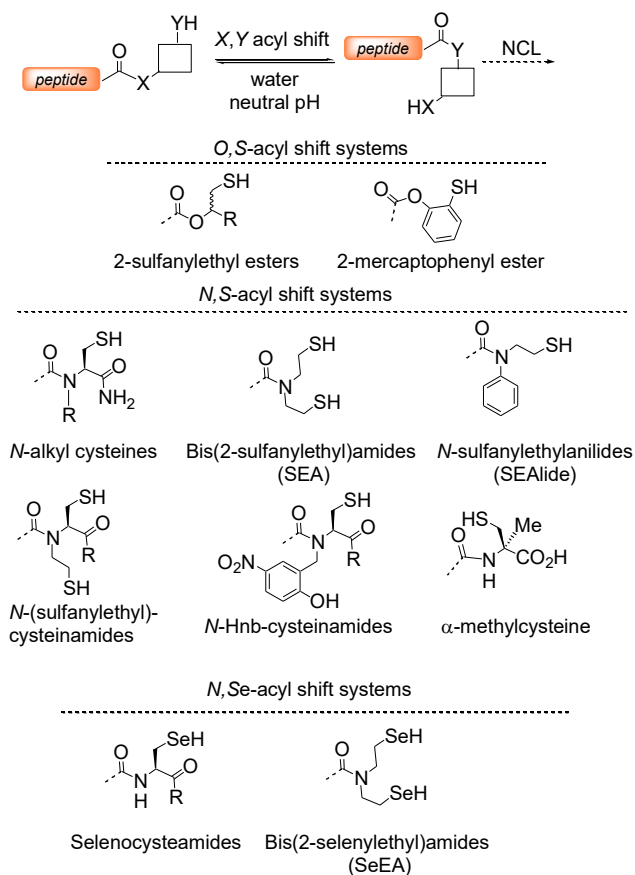


Figure 2. Structure of some *O,S*, *N,S* and *N,Se*-acyl shift systems. *O,S*: 2-sulfanylethyl esters²³⁻²⁵; 2-mercaptophenyl ester²⁶; *N,S*: *N*-alkyl cysteines^{27,28}; *bis*(2-sulfanylethyl)amido^{4,29}; *N*-sulfanylethylanilides (SEAlide)^{5,30}; *N*-(sulfanylethyl)-cysteinamides³¹; *N*-Hnb-cysteinamides³²; α -methylcysteine³³; selenocysteamides³⁴; *bis*(2-selenylethyl)amides (SeEA)^{35,36}.

More than 20 years after the introduction of NCL, we wanted to conduct a statistical review of its use for protein chemical synthesis in order to provide a flavor of the recent trends and identify the most popular chemical tools used by protein chemists. In other words, how are proteins made today? On average, is there a significant increase in the size of the proteins made by chemical synthesis? What are the main limitations encountered by protein chemists? To answer these questions we created the Protein Chemical Synthesis (PCS) database by collecting a set of relevant data from more than 450 publications covering the period 1994-2017. We determined the number and size distribution of the proteins produced by chemical synthesis per year and the average number of ligation steps used to obtain the target protein. We estimated the frequency of use of Nbz, hydrazide and *O,S*, *N,S* or *N,Se*-acyl shift systems for peptide bond formation. We also determined the frequency of use of desulfurization or deselenization techniques as well as the frequency of use of mercapto or seleno derivatives of proteinogenic amino acids in protein synthesis. We also provide the frequency of each amino acid at the junctions formed (X-Cys, X-Sec, X-Ala), which can be compared to the natural distribution. We also provide useful statistics on the occurrence of Cys, Sec and Ala residues in proteins, as well as the preferences for the preceding residue X, thereby allowing some comparisons to be made.

These statistics give a precise idea of the importance of protein chemical synthesis in fundamental and applied research today, and highlight the important challenges that remain to be addressed.

2. Statistical analysis of the UniRef50 protein database

The UniProt Reference Clusters (UniRef) are comprehensive and non-redundant secondary databases computed from clustered sets of protein sequences from UniProtKB and UniParc.³⁷⁻³⁹ With about 12,000,000 clusters, the UniRef50 dataset provides an extensive sample of proteins from which useful statistics can be extracted. We and others used this database to determine the occurrences of amino acids and other useful information discussed below.

2.1. Occurrence of amino acids

The occurrence of proteinogenic amino acids in the UniRef50 database can be seen in Fig. 3a. These data confirm what is often mentioned in the literature, namely the low frequency of Cys residues in proteins (1.48%). In comparison, the occurrence of Ala is 5.7 fold higher (8.42%), making it important in methods that enable the formation of X-Ala junctions by combining NCL with desulfurization or deselenization reactions. Also unsurprising is the very low frequency of Sec residues in proteins (1.19 E-5%). The human genome encodes only 25 selenoproteins, which are also rare in all other living organisms.⁴⁰

What is less discussed and worth considering is how these frequencies vary as a function of protein size. Variation in the occurrence of Cys and Ala residues as a function of protein size was determined by Carugo using the UniRef50 database (Fig. 4).⁴¹ This analysis shows that the chance of finding a Cys residue in a protein decreases with increasing protein size. The frequency of Cys in small proteins comprising 50 amino acids is 2.25% and drops to ~1.45% for proteins of 200 amino acids. Interestingly, the opposite is observed for Ala (Fig. 4b). The frequency of Ala in small proteins (50 AA) is ~6.8% and increases rapidly to reach a plateau at ~8.75% in proteins of 100 aa or more.

It is also interesting to compare the occurrence of amino acids involved in a peptide bond with Cys (X-Cys, Fig. 3b), Ala (X-Ala, Fig. 3c) or Sec (X-Sec, Fig. 3d) to the global frequencies (X-Y, Y any amino acid). For X-Cys pairs, there is a good correlation between global frequencies and occurrence for X when bound to Cys. Cys is an exception to this rule since Cys-Cys pairs are clearly more frequent than expected from global frequencies. The Cys residue is shown in red in Fig. 3b to highlight this fact. A similar picture can be drawn for X-Ala pairs (Fig. 3c). While the occurrence of most amino acids in X-Ala pairs is very close to the general frequencies, Ala is clearly overrepresented by +41%. As seen for Cys-Cys pairs, the occurrence of Ala-Ala pairs is more than expected from global frequencies.

Sec is obviously a case apart, since most of the X-Sec frequencies deviate significantly from global frequencies (Fig. 3D). The most striking feature of X-Sec pairs is the overrepresentation of Gly (+227%), Ser (81%), Thr (125%) and Cys (+297%) residues.

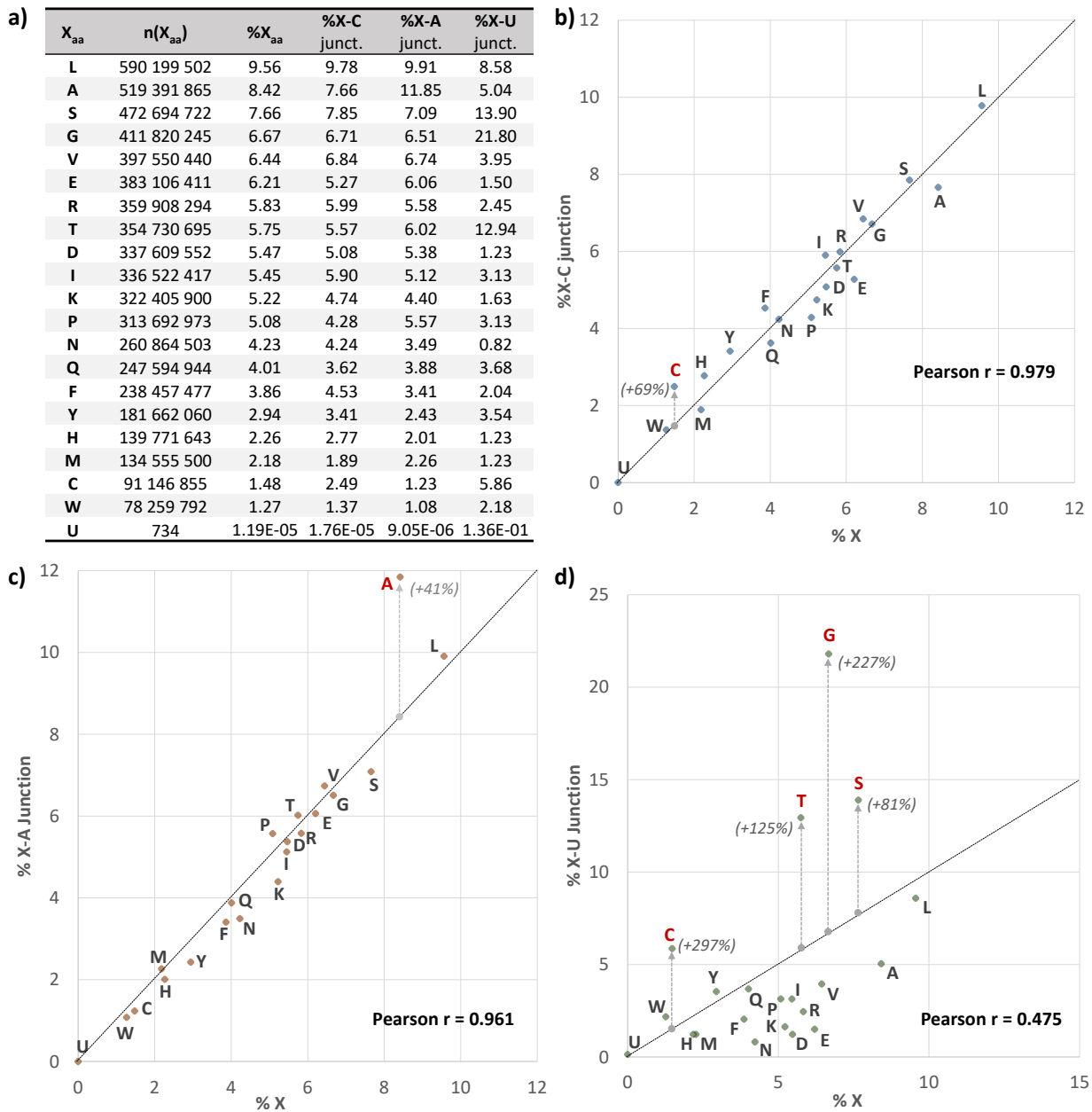


Figure 3. Statistics from the UniRef50 database. A) Occurrence of amino acids globally or when involved in a peptide bond with Cys, Ala or Sec. B-D) The occurrence of amino acids linked to Cys, Ala or Sec is compared with the global occurrence of amino acids; B) X-Cys, C) X-Ala and D) X-Sec. Amino acids shown in red show an occurrence significantly different from the global occurrence.

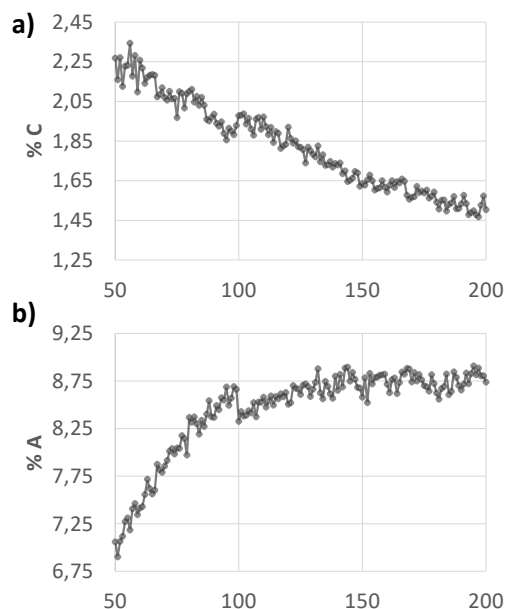


Figure 4. Occurrence of Cys (a) or Ala (b) as a function of protein size (data reproduced from Carugo⁴¹).

2.2. The protein chemical synthesis database

The protein chemical synthesis (PCS) database was created by collecting a set of data from approximately 450 unique references in the literature published over the period 1994-February 2017. We considered only reports describing the total synthesis of peptides or proteins of biological significance, and excluded from the database reports describing the synthesis of model peptides, polymers, hybrid materials, etc. Here, we have considered only the proteins produced by NCL and extended methodologies. By extended methodologies we mean the methods for which the thioester segment is substituted by a thioester surrogate (selenoester, hydrazide, Nbz, *O,S*, *N,S*, and *N,Se* acyl shift systems), or for which the Cys segment is substituted by a thiol or selenol amino acid residue (thiol-based auxiliaries linked to the α -amino group, mercapto or seleno amino acid surrogates). Some proteins produced using EPL were also included. Other ligation techniques such as serine/threonine ligation (STL)⁶ or alpha-ketoacid-hydroxylamine (KAHA) ligation³ will be included in the future. The PCS database will be updated regularly, and can be consulted through a dedicated website (<http://pcs-db.fr>).

The name, date of publication, length and origin of more than 600 peptide/proteins and analogs prepared by total chemical synthesis or semisynthetic approaches were collected. Importantly, details about the synthetic design, such as the type of ligation chemistry used, nature of the junctions formed, use of amino acid surrogates or thiol auxiliaries and/or use of dechalcogenation reactions were compiled. For statistics regarding the nature of junction residues or the type of ligation used, semisynthetic methods using EPL were taken into consideration but were excluded from the sample for information about protein length. In addition, the database contains some information regarding the nature of target proteins, for example if they are post-translationally modified (by acetylation, phosphorylation, methylation, ubiquitination or a ubiquitin-like modifier, cyclization, etc.). The PCS database establishes a broad overview of the use of native chemical and extended methodologies for protein synthesis and the type of targets pursued by protein chemists.

2.3. The size of proteins produced by chemical synthesis

The first interesting information that can be gathered from the PCS database is the size of the proteins produced by NCL or extended methodologies as a function of the year of publication from 1994 to 2017 (Fig. 5A). This figure gives the mean or median size, the first and third quartiles, minimum and maximum. For some years, the statistical analysis generates outliers that are shown as open circles in Fig. 5. The sizes of these proteins are exceptional in relation to the other syntheses carried out the same year. The current year is a case apart due to the limited number of entries, but has nevertheless been included in Fig. 5.

An important conclusion that can be drawn from Fig. 5 is that the mean or median size of the proteins produced by chemical synthesis varies little over the period under consideration, and is around 80 amino acids. In contrast, the size of the largest proteins produced by chemical synthesis increased significantly from 1994 to 2016 as shown by the increase in Q3 and in the maximum and outlier values. These reports represent outstanding achievements, and in a sense reflect the progress made in the field, in particular the ability of protein chemists to set up and perform complex multistep syntheses allowing access to challenging proteins. However, although these publications logically have a strong impact on the protein chemist community, due to their rarity they do not significantly displace the median size in this statistical analysis.

A complementary way to discuss these points is to look at the mean number of ligation steps used by protein chemists to produce their proteins (Fig. 5B). The majority of the proteins are prepared by a single ligation step. Since the median size of synthetic proteins is around 80 amino acids, the ligation of two peptide segments of 40 amino acids in length meets the majority of needs. The synthesis of large peptide segments is possible thanks to the dramatic improvements in the SPPS methods used to obtain large peptides (for a recent review see Raibaut et al.⁴²). Moreover, for a long time the production of long peptide thioesters was a serious bottleneck. Today, the protein chemist has an array of efficient Boc and Fmoc-SPPS methods at his/her disposal for obtaining long peptide thioesters or thioester surrogates such as Nbz, hydrazides or some *N,S*-acyl shift systems, usually making the preparation of the peptide segments straightforward. It is also interesting to compare the median and mean number of ligation steps per protein and per year (Fig. 5B). Before 2010 and with some exceptions, the mean and the median were usually close to 1. For the last 6 years, although the median is equal to 1, the mean approaches the value of 2. This means that although the majority of the proteins are synthesized by a single ligation step, the proportion of publications using two or more ligations has increased significantly during the past 6 years. The number of outliers has also risen significantly during the same period. Overall, these data indicate that the synthesis of small proteins (~80 amino acids, one ligation step) represents an important application of NCL and covers many needs. However, the trend is clearly toward the use of multistep synthetic schemes for accessing larger proteins, starting with the use of three peptide segments that are joined by two ligation steps. The fact that a large variety of one-pot three peptide segment assembly schemes has been designed in the past 6 years supports this view. Indeed, several kinetically controlled ligation schemes (KCLs) have been developed using *O,S*²⁴ or *N,S*^{5,43-46} or *N,Se*-acyl shift systems.³⁵ Complementarily, the design of one-pot assembly schemes has also been facilitated by the design of several latent thioester^{47,48} or selenoester surrogates^{35,36} and by Cys protection strategies.⁴⁹⁻⁵⁶ The mean size of the proteins produced by chemical synthesis

increases significantly on going from one to four ligation steps (Fig. 5C). In contrast, there is no increase in the size of the proteins prepared by chemical synthesis using five rather than four ligation steps, and the majority of these are below 200 aa.

Another important point is that although the median size of the proteins prepared by chemical synthesis has not changed

significantly since the introduction of NCL, the number of proteins synthesized per year (or the number of ligations performed per year) has increased dramatically since 2010. This is obvious from Fig. 6.

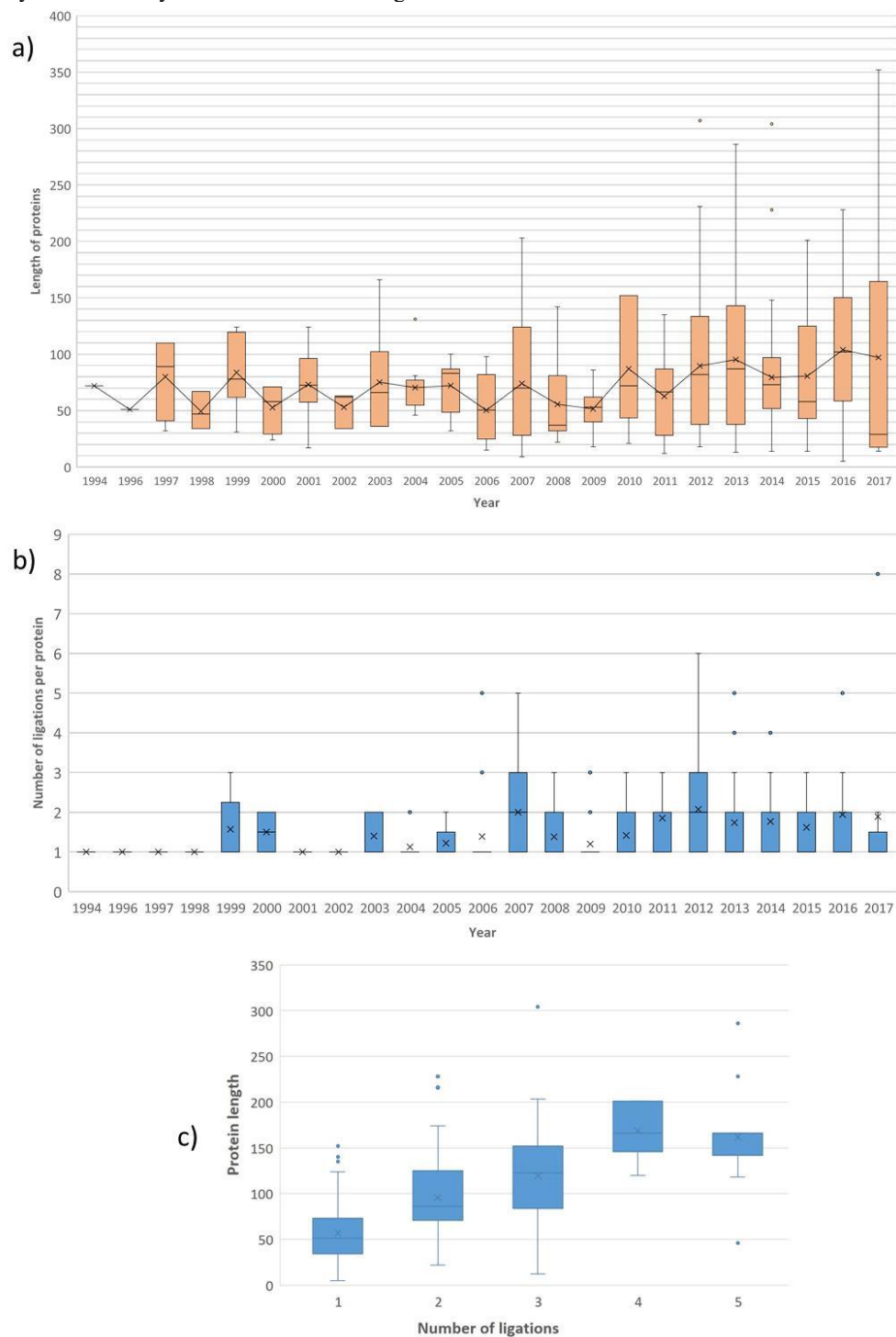


Figure 5. A) The size of the proteins produced by chemical synthesis plotted against the year of publication. B) Number of ligation steps used to synthesize proteins plotted against the year of publication. For years 1998, 2003, 2005, 2008 and 2010-2017, the median equals 1 and is therefore not visible on the graph. C) The size of proteins produced by chemical synthesis plotted against the number of ligation steps. The data were analyzed using standard non-parametric statistics. The number of proteins synthesized in a given year is shown in Fig. 6.

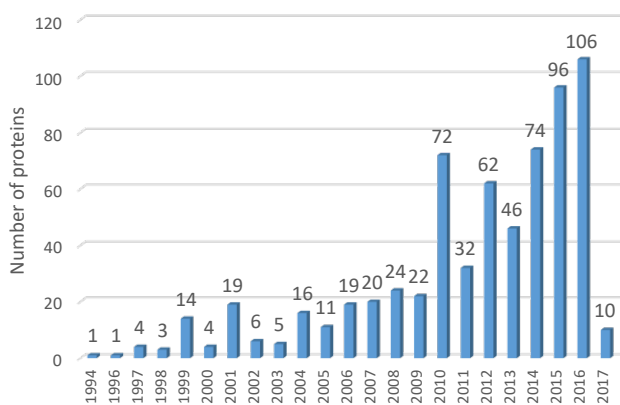


Figure 6. Number of proteins produced by chemical synthesis and included in the PCS database as a function of the year of publication.

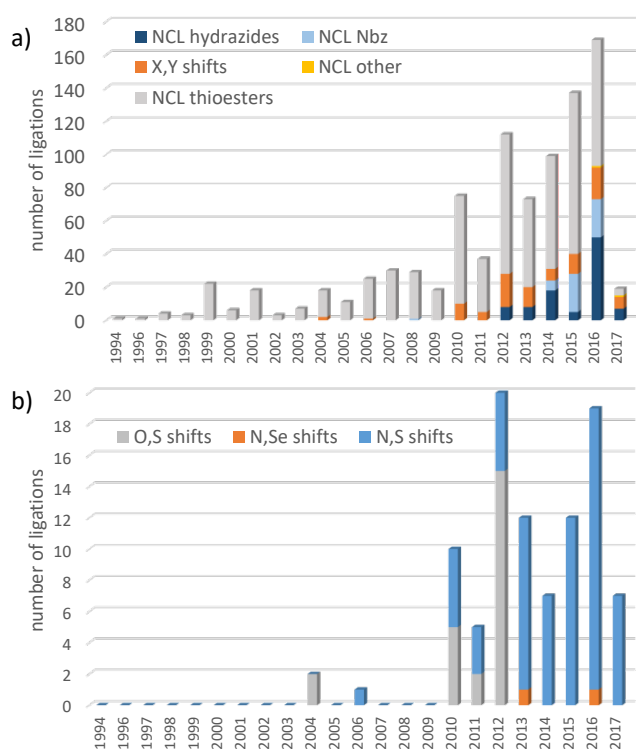


Figure 7. Type of chemistry used for protein chemical synthesis. A) Number of ligations carried out using thioesters, Nbz, hydrazide or *X,Y*-acyl shift systems as acyl donors. B) Type of *X,Y*-acyl shift system used for protein chemical synthesis.

2.4. Methods used for protein chemical synthesis

2.4.1 The acyl donor

Next, we used the PCS database to extract the information regarding the principal methods used for protein chemical synthesis. We will first discuss the nature of the acyl donor in the ligation. The number of ligations performed using thioester, Nbz, hydrazide or *X,Y*-acyl shift systems as acyl donors is shown in Fig. 7a. The data show the increased use of Nbz, hydrazide and *X,Y*-acyl shift chemistries compared with traditional thioesters since 2010. In 2016, they represented more than half (54%) of the ligations performed for protein chemical synthesis. In the same year, the number of ligations carried out using *X,Y*-acyl shift

systems and Nbz was similar and represented 11 and 13% of the total respectively, while hydrazide chemistry represented 30%. The type of *X,Y*-acyl shift system used in the ligations is detailed in Fig. 7b. It is interesting to note that both *O,S* and *N,S*-acyl shift systems co-existed in the period 2010-2012. However, the *O,S*-acyl shift systems are rarely used for protein chemical synthesis since 2013. Finally, the design of *N,Se*-acyl shift systems is clearly in its infancy with only two reports, one in 2013 and one in 2016.

Overall, these data reveal that traditional thioesters are gradually being replaced by other acyl donors, primarily the peptide hydrazides introduced by Liu and coworkers.^{19,20}

2.4.2 Desulfurization reactions

We further examined the use of Cys desulfurization techniques to produce peptide bonds involving alanine. As mentioned above, Cys is one of the rarest amino acids in proteins. Its occurrence in mammals is 2.26% (in human, bovine and murine proteins), while it is only 0.5% in proteins from some members of the domain *Archaea* (0.5%).⁵⁷ It is important to note that Cys is absent from as many as 8% of the human proteins. Moreover, even if a protein contains one or several Cys residues, their location is not always optimal in relation to the length of the peptide segments being prepared or the type of junction required. For this reason, setting up an efficient synthetic strategy for a given protein is often complex if it relies entirely on Cys residues.

Fig. 8 makes it possible to compare the number of ligations that are followed by a desulfurization step for Cys and thiol amino acid derivatives to the number of ligations intended to produce a peptide bond with Cys. The data clearly show that ligation/desulfurization approaches have gained ground since 2010. In 2016, the number of ligation/desulfurization steps was far larger than the number of ligations that were not combined with a desulfurization step.

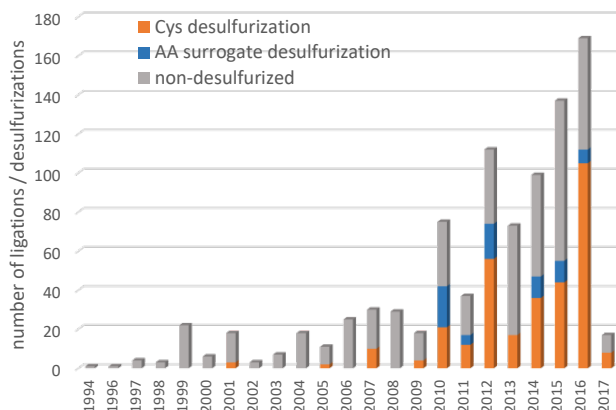


Figure 8. Number of thiol amino acids (Cys, thiol amino acid derivatives) desulfurized after ligation and number of ligations at Cys without desulfurization.

2.4.3 Types of junctions formed

The last point discussed in this statistical study concerns the type of junctions formed using NCL or related reactions (Fig. 9). We extracted the type of Xaa-Cys junctions produced during the period 1994-2017 from the PCS database and plotted the frequencies obtained against the natural abundance. Fig. 9 shows a large positive deviation for Gly and Ala residues. Conversely, a

group of amino acids shows a marked negative deviation from the natural abundance, namely proline, isoleucine, aspartic acid and glutamic acid.

The large deviation for Gly might be due to the large number of publications that describe the synthesis of proteins modified by ubiquitin or ubiquitin-like modifiers such as small ubiquitin-like modifiers (SUMO). These small proteins are actually attached to their targets through a Gly-Lys isopeptide bond. The positive deviation for Gly and Ala might also be due to the favorable kinetic rate observed with these amino acids, while the opposite can be inferred for Pro, Ile and Val.⁵⁸ Finally, C-terminal aspartyl and glutamyl thioesters are subject to side-reactions during NCL, such as hydrolysis, succinimide, macrolactamization or isopeptide bond formation, which can explain the reluctance to assemble proteins through these types of junctions.⁵⁹⁻⁶³

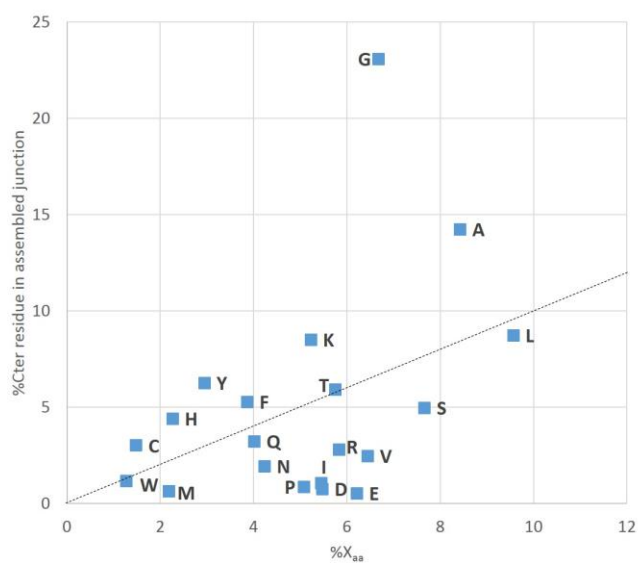


Figure 9. The frequency of Xaa-Cys junctions formed using NCL and related reactions plotted against the natural abundance.

3. Conclusion

The PCS database provides useful information regarding the number and type of proteins produced by chemical synthesis, and provides a valuable insight into the chemical methods used to obtain target proteins. It also highlights the synthetic difficulties encountered during syntheses and makes it possible to identify important trends in the field. Some of the collected data have not been discussed here, such as the type of post-translational modifications pursued by protein chemists, the importance of peptide or protein cyclization or the frequency of mutating some residues for overcoming some synthetic difficulties. The PCS database will be updated regularly and discussed in more detail in future reports.

Acknowledgments

We warmly thank Dr Oliviero Carugo from the University of Pavia (Italy) for providing the data reproduced in Fig. 4. We also thank Arnaud Jolly for helping us to write the Java programs used to mine the UniRef 50 database.

References

- (1) Raibaut, L.; Ollivier, N.; Melnyk, O. *Chem. Soc. Rev.* **2012**, *41*, 7001-7015.
- (2) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776-779.
- (3) Bode, J. W.; Fox, R. M.; Baucom, K. D. *Angew. Chem. Int. Ed.* **2006**, *45*, 1248-1252.
- (4) Ollivier, N.; Dheur, J.; Mhida, R.; Blanpain, A.; Melnyk, O. *Org. Lett.* **2010**, *12*, 5238-5241.
- (5) Sato, K.; Shigenaga, A.; Tsuji, K.; Tsuda, S.; Sumikawa, Y.; Sakamoto, K.; Otaka, A. *ChemBioChem* **2011**, *12*, 1840-1844.
- (6) Zhang, Y.; Xu, C.; Lam, H. Y.; Lee, C. L.; Li, X. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6657-6662.
- (7) Kent, S. B. *Chem. Soc. Rev.* **2009**, *38*, 338-351.
- (8) Quaderer, R.; Sewing, A.; Hilvert, D. *Helv. Chim. Acta* **2001**, *84*, 1197-1206.
- (9) Hondal, R. J.; Nilsson, B. L.; Raines, R. T. *J. Am. Chem. Soc.* **2001**, *123*, 5140-5141.
- (10) Gieselman, M. D.; Xie, L.; van Der Donk, W. A. *Org. Lett.* **2001**, *3*, 1331-1334.
- (11) Metanis, N.; Hilvert, D. *Curr. Opin. Chem. Biol.* **2014**, *22*, 27-34.
- (12) Yan, L. Z.; Dawson, P. E. *J. Am. Chem. Soc.* **2001**, *123*, 526-533.
- (13) Wan, Q.; Danishefsky, S. J. *Angew. Chem. Int. Ed.* **2007**, *46*, 9248-9252.
- (14) Metanis, N.; Keinan, E.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2010**, *49*, 7049-7053.
- (15) Botti, P.; Carrasco, M. R.; Kent, S. B. H. *Tetrahedron Lett.* **2001**, *42*, 1831-1833.
- (16) Offer, J.; Boddy, C. N.; Dawson, P. E. *J. Am. Chem. Soc.* **2002**, *124*, 4642-4646.
- (17) Blanco-Canosa, J. B.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2008**, *47*, 6851-6855.
- (18) Blanco-Canosa, J. B.; Nardone, B.; Albericio, F.; Dawson, P. E. *J. Am. Chem. Soc.* **2015**, *137*, 7197-7209.
- (19) Fang, G.-M.; Li, Y.-M.; Shen, F.; Huang, Y.-C.; Li, J.-B.; Lin, Y.; Cui, H.-K.; Liu, L. *Angew. Chem. Int. Ed.* **2011**, *50*, 7645-7649.
- (20) Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L. *Nat. Protocols* **2013**, *8*, 2483-2495.
- (21) Zheng, J.-S.; Tang, S.; Huang, Y.-C.; Liu, L. *Acc. Chem. Res.* **2013**, *46*, 2475-2484.
- (22) Melnyk, O.; Agouridas, V. *Curr. Opin. Chem. Biol.* **2014**, *22*, 137-145.
- (23) Botti, P.; Villain, M.; Manganiello, S.; Gaertner, H. *Org. Lett.* **2004**, *6*, 4861-4864.
- (24) Zheng, J. S.; Cui, H. K.; Fang, G. M.; Xi, W. X.; Liu, L. *ChemBioChem* **2010**, *11*, 511-515.
- (25) Tofteng, A. P.; Jensen, K. J.; Hoeg-Jensen, T. *Tetrahedron Lett.* **2007**, *48*, 2105-2107.
- (26) Warren, J. D.; Miller, J. S.; Keding, S. J.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 6576-6578.
- (27) Asahina, Y.; Nabeshima, K.; Hojo, H. *Tetrahedron Lett.* **2015**, *56*, 1370-1373.
- (28) Kang, J.; Richardson, J. P.; Macmillan, D. *Chem. Commun.* **2009**, 407-409.
- (29) Hou, W.; Zhang, X.; Li, F.; Liu, C. F. *Org. Lett.* **2011**, *13*, 386-389.
- (30) Eto, M.; Naruse, N.; Morimoto, K.; Yamaoka, K.; Sato, K.; Tsuji, K.; Inokuma, T.; Shigenaga, A.; Otaka, A. *Org. Lett.* **2016**, *18*, 4416-4419.

- (31) Ruff, Y.; Garavini, V.; Giuseppone, N. *J. Am. Chem. Soc.* **2014**, *136*, 6333-6339.
- (32) Terrier, V. P.; Adihou, H.; Arnould, M.; Delmas, A. F.; Aucagne, V. *Chem. Sci.* **2016**, *7*, 339-345.
- (33) Burlina, F.; Papageorgiou, G.; Morris, C.; White, P. D.; Offer, J. *Chem. Sci.* **2014**, *5*, 766-770.
- (34) Adams, A. L.; Macmillan, D. *J. Pept. Sci.* **2013**, *19*, 65-73.
- (35) Raibaut, L.; Cargoet, M.; Ollivier, N.; Chang, Y. M.; Drobecq, H.; Boll, E.; Desmet, R.; Monbaliu, J.-C. M.; Melnyk, O. *Chem. Sci.* **2016**, *7*, 2657-2665.
- (36) Raibaut, L.; Drobecq, H.; Melnyk, O. *Org. Lett.* **2015**, *17*, 3636-3639.
- (37) Suzek, B. E.; Huang, H.; McGarvey, P.; Mazumder, R.; Wu, C. H. *Bioinformatics* **2007**, *23*, 1282-1288.
- (38) Suzek, B. E.; Yang, Y. Y.; Huang, H.; McGarvey, P. B.; Wu, C. H. *The UniProt Consortium Bioinformatics* **2015**, *31*, 926-932.
- (39) Suzek, B. E.; Wang, Y.; Huang, H.; McGarvey, P. B.; Wu, C. H. *Bioinformatics* **2015**, *31*, 926-932.
- (40) Kryukov, G. V.; Castellano, S.; Novoselov, S. V.; Lobanov, A. V.; Zehntab, O.; Guigo, R.; Gladyshev, V. N. *Science* **2003**, *300*, 1439-1443.
- (41) Carugo, O. *Protein Sci* **2008**, *17*, 2187-2191.
- (42) Raibaut, L.; El Mahdi, O.; Melnyk, O. *Top. Curr. Chem.* **2015**, *363*, 103-154.
- (43) Yang, R.; Hou, W.; Zhang, X.; Liu, C.-F. *Org. Lett.* **2012**, *14*, 374-377.
- (44) Ding, H.; Shigenaga, A.; Sato, K.; Morishita, K.; Otaka, A. *Org. Lett.* **2011**, *13*, 5588-5591.
- (45) Nakamura, T.; Shigenaga, A.; Sato, K.; Tsuda, Y.; Sakamoto, K.; Otaka, A. *Chem. Commun.* **2014**, *50*, 58-60.
- (46) Raibaut, L.; Seeberger, P.; Melnyk, O. *Org. Lett.* **2013**, *15*, 5516-5519.
- (47) Ollivier, N.; Vicogne, J.; Vallin, A.; Drobecq, H.; Desmet, R.; El-Mahdi, O.; Leclercq, B.; Goormachtigh, G.; Fafeur, V.; Melnyk, O. *Angew. Chem. Int. Ed.* **2012**, *51*, 209-213.
- (48) Boll, E.; Drobecq, H.; Ollivier, N.; Blanpain, A.; Raibaut, L.; Desmet, R.; Vicogne, J.; Melnyk, O. *Nat. Protoc.* **2015**, *10*, 269-292.
- (49) Bang, D.; Kent, S. B. *Angew. Chem. Int. Ed.* **2004**, *43*, 2534-2538.
- (50) Li, J.; Li, Y.; He, Q.; Li, Y.; Li, H.; Liu, L. *Org. Biomol. Chem.* **2014**, *12*, 5435-5441.
- (51) Boll, E.; Ebran, J. P.; Drobecq, H.; El-Mahdi, O.; Raibaut, L.; Ollivier, N.; Melnyk, O. *Org. Lett.* **2015**, *17*, 130-133.
- (52) Ueda, S.; Fujita, M.; Tamamura, H.; Fujii, N.; Otaka, A. *Chembiochem* **2005**, *6*, 1983-1986.
- (53) Aihara, K.; Yamaoka, K.; Naruse, N.; Inokuma, T.; Shigenaga, A.; Otaka, A. *Org. Lett.* **2016**, *18*, 596-599.
- (54) Tang, S.; Si, Y.-Y.; Wang, Z.-P.; Mei, K.-R.; Chen, X.; Cheng, J.-Y.; Zheng, J.-S.; Liu, L. *Angew. Chem. Int. Ed.* **2015**, *19*, 5713-5717.
- (55) Jbara, M.; Maity, S. K.; Seenaiah, M.; Brik, A. *J. Am. Chem. Soc.* **2016**, *138*, 5069-5075.
- (56) Maity, S. K.; Jbara, M.; Laps, S.; Brik, A. *Angew. Chem. Int. Ed.* **2016**, *55*, 8108-8112.
- (57) Miseta, A.; Csutora, P. *Mol. Biol. Evol.* **2000**, *17*, 1232-1239.
- (58) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 10068-10073.
- (59) Villain, M.; Gaertner, H.; Botti, P. *Eur. J. Org. Chem.* **2003**, *2003*, 3267-3272.
- (60) Creech, G. S.; Paresi, C.; Li, Y. M.; Danishefsky, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 2891-2896.
- (61) Arnold, U.; Hinderaker, M. P.; Nilsson, B. L.; Huck, B. R.; Gellman, S. H.; Raines, R. T. *J. Am. Chem. Soc.* **2002**, *124*, 8522-8523.
- (62) Gross, C. M.; Lelièvre, D.; Woodward, C. K.; Barany, G. *J. Pept. Res.* **2005**, *65*, 395-410.
- (63) Johnson, E. C.; Kent, S. B. *J. Am. Chem. Soc.* **2006**, *128*, 6640-6646.

CrossMark
click for updatesCite this: *Chem. Sci.*, 2016, 7, 2657

Accelerating chemoselective peptide bond formation using bis(2-selenylethyl)amido peptide selenoester surrogates†

Laurent Raibaut,^{‡a} Marine Cargoët,^{‡a} Nathalie Ollivier,^a Yun Min Chang,^a Hervé Drobecq,^a Emmanuelle Boll,^a Rémi Desmet,^a Jean-Christophe M. Monbaliu^{*b} and Oleg Melnyk^{*a}

Given the potential of peptide selenoesters for protein total synthesis and the paucity of methods for the synthesis of these sensitive peptide derivatives, we sought to explore the usefulness of the bis(2-selenylethyl)amido (SeEA) group, *i.e.* the selenium analog of the bis(2-sulfanylethyl)amido (SEA) group, for accelerating peptide bond formation. A chemoselective exchange process operating in water was devised for converting SEA peptides into the SeEA ones. Kinetic studies show that SeEA ligation, which relies on an initial *N*,*Se*-acyl shift process, proceeds significantly faster than SEA ligation. This property enabled the design of a kinetically controlled three peptide segment assembly process based on the sequential use of SeEA and SEA ligation reactions. The method was validated by the total synthesis of hepatocyte growth factor K1 (85 AA) and biotinylated NK1 (180 AA) domains.

Received 14th September 2015

Accepted 8th January 2016

DOI: 10.1039/c5sc03459k

www.rsc.org/chemicalscience

Introduction

The field of protein chemical synthesis has flourished substantially during the last decades to a point where the use of chemical methods is considered as a useful alternative to recombinant techniques for accessing small functional protein domains. Chemical synthesis is especially useful when proteins having unusual architectures^{1–6} or well-defined modifications at specific sites are desirable. An application of protein chemical synthesis which is of utmost importance in biological sciences is the study of some complex post-translational modifications such as glycosylation,⁷ ubiquitination^{8–10} or sumoylation.^{11,12}

Today, synthetic proteins are usually produced by assembling unprotected peptide segments in water. The native chemical ligation (NCL) is a popular chemoselective amide bond forming reaction in the field,^{13,14} although other ligations show great promise as well.^{15–19} The NCL reaction between a C-terminal peptide thioester and an N-terminal cysteinyl (Cys) peptide yields a native peptide bond to cysteine. Recently, C-terminal peptide selenoesters were also shown to be useful reactants for amide bond formation with the advantage of being significantly more reactive than peptide thioesters.^{20,21}

Unfortunately, the application of peptide selenoesters for protein synthesis is difficult due to the limited access to these sensitive peptide derivatives, which were produced up to now using Boc solid phase peptide synthesis (SPPS) methods.^{20,22}

The powerfulness of the NCL reaction has stimulated the development of several methods for accessing peptide thioesters using the Fmoc SPPS.²³ In particular, several *N*,*S*-acyl shift systems relying on the *N*-(2-sulfanylethyl)amide structure were shown to rearrange into thioesters in acidic media.^{24–29} The fact that *N*-(2-sulfanylethyl)amides rearrange spontaneously into thioesters in strong aqueous acids has been noticed in the 1950s by Martin and coworkers³⁰ using *N*-acetyl- β -mercaptoethylamine as a model system. This property is also shared by the peptide bond to cysteine.³¹ In contrast, the capacity of some *N*-(2-sulfanylethyl)amides^{32–35} such as the bis(2-sulfanylethyl)amido (SEA) group^{32,36} to rearrange into thioesters in water at neutral³² or mildly acidic^{37–39} pH is a recent observation (Scheme 1). The reaction of a SEA peptide with an N-terminal cysteinyl peptide yields a native peptide bond to cysteine and proceeds efficiently in the presence of 4-mercaptophenylacetic acid (MPAA⁴⁰). The usefulness of the SEA ligation has been demonstrated by the total synthesis of several functional proteins.^{11,12,41–45}

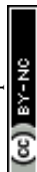
Given the promise of peptide selenoester chemistry for protein total synthesis, we sought to explore the potential of the bis(2-selenylethyl)amido (SeEA) amido group,⁴⁶ *i.e.* the selenium analog of the SEA group (Scheme 1), for accelerating the chemoselective formation of the peptide bond to Cys. We first developed a mild and chemoselective exchange process for converting SEA peptides into SeEA derivatives in water.

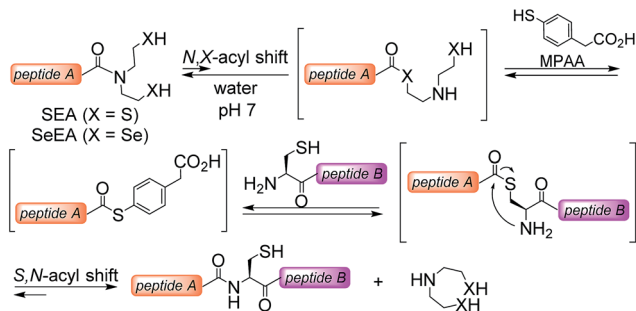
^aUMR CNRS 8161 CNRS, Université de Lille, Institut Pasteur de Lille, 1 rue du Pr Calmette, 59021 Lille Cedex, France. E-mail: oleg.melnyk@ibl.cnrs.fr

^bCenter for Integrated Technology and Organic Synthesis, Department of Chemistry, University of Liège, Building B6a, Room 3/16a, Sart-Tilman, B-4000 Liège, Belgium

† Electronic supplementary information (ESI) available: Calculations, synthetic protocols and characterization for all compounds. See DOI: 10.1039/c5sc03459k

‡ Laurent Raibaut and Marine Cargoët contributed equally to this work.





Scheme 1 Principle of SEA (X = S) and SeEA (X = Se) chemoselective ligations.

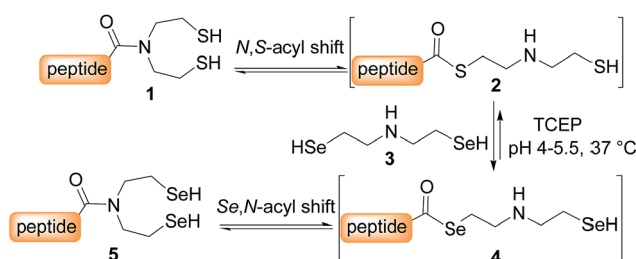
Importantly, SeEA peptides proved to be more reactive than classical peptide alkylthioesters and significantly more reactive than SEA peptides in model ligation experiments conducted at neutral pH. This property enabled designing a one-pot kinetically controlled⁴⁷ three peptide segment assembly process involving the sequential use of SeEA and SEA ligations.

Synthesis of SeEA peptides

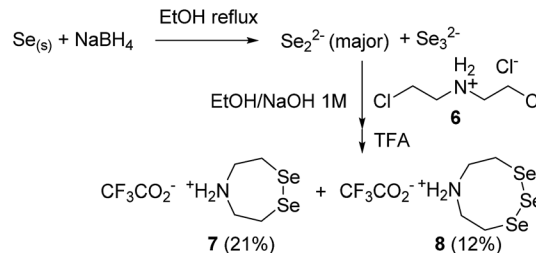
SEA peptides are easily accessible by the Fmoc SPPS using SEA polystyrene^{32,48} or PEG-based resins.^{11,12} We therefore sought to convert SEA peptides **1** into SeEA derivatives **5** by using the exchange reaction shown in Scheme 2, which takes advantage of the capacity of SEA amide form **1** to equilibrate in water with the SEA transient thioester form **2**.^{29,32,49} We reasoned that the latter might participate to a selenol–thioester exchange reaction with an excess of the bis(2-selenylethyl)amine **3** to produce, after an *Se,N*-acyl shift, the target SeEA peptide **5**. Overall, the exchange process shown in Scheme 2 corresponds to a chemoselective transamidation reaction³⁶ which is applied to unprotected SEA peptide segments.

The trifluoroacetate salt of cyclic diselenide **7**, *i.e.* 1,2,5-diselenazepane, was envisioned as a good precursor for diselenol **3**. Its synthesis has been described elsewhere⁴⁶ by reducing metallic selenium with sodium borohydride and reacting the resulting diselenide dianion Se_2^{2-} with bis(2-chloroethyl)amine hydrochloride **6** (Scheme 3). The reaction yielded also the trifluoroacetate salt of triselenide **8**, *i.e.* 1,2,3,6-triselenazocane, which was isolated and characterized in this work.

The first attempts to transamidate model peptide **9a** with diselenide **7** in the presence of TCEP yielded only trace amounts



Scheme 2 Conversion of SEA peptides into SeEA derivatives by transamidation.



Scheme 3 Preparation of cyclic di or triselenides **7** and **8**.

of the target SeEA peptide **10a** (Fig. 1A and B). Instead, we observed the formation of a compound whose molecular mass corresponded to peptide **10a** minus one selenium atom and to whom the structure of peptide **11a** was assigned on the basis of the following arguments.

The deselenization of selenocysteine into alanine by TCEP, which is reminiscent of the desulfurization of thiols by phosphites,^{50,51} is a well-known process which has been noticed as a serious side-reaction or utilized as a useful synthetic transformation.^{52–54} However, a similar deselenization process when

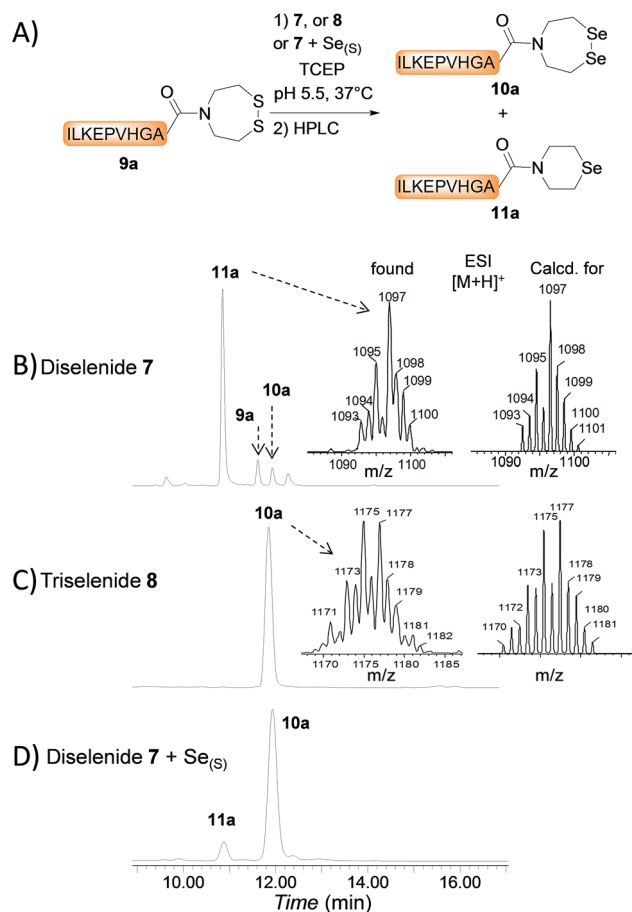


Fig. 1 Conversion of SEA peptides into SeEA derivatives by transamidation (A). HPLC of the crude exchange reaction with diselenide **7** (10 equiv) (B), triselenide **8** (10 equiv) (C) or diselenide **7** (10 equiv) plus metallic selenium (2 equiv) (D). Peptide **9a** 10 mM, TCEP 100 mM, pH 5.5, 37 °C, 24 h.



applied to SeEA peptide **10a** would yield a peptide featuring a C-terminal *N*-(2-selenylethyl)-*N*-ethyl amide moiety, whose molecular weight is 2 units greater to those measured for **11a**. This mode of deselenization can thus be ruled out.

In contrast, the extrusion of selenium from diselenides to produce selenoethers has received much attention in the past and is a simple explanation to the formation of peptide **11a** from diselenide **10a**.⁵⁵ Since this free-radical process is catalysed by tertiary phosphines,⁵⁶ a potential solution to this problem was to replace TCEP by dithiothreitol (DTT) in the exchange reaction. However, while DTT is capable of reducing acyclic diselenides such as selenocystine bonds,⁵⁷ it is unable to reduce cyclic diselenide **7**.⁴⁶ This fact imposed the use of TCEP in the exchange reaction and thus to find solution for avoiding the extrusion of selenium. Gratefully, the use of triselenide **8** as a precursor for diselenol **3** yielded exclusively the target SeEA peptide **10a** (51%, Fig. 1C and Scheme 4). The exchange procedure using triselenide **8** and TCEP enabled also the successful preparation of several SeEA peptides in good yield such as SeEA peptide **10b** which features a sterically demanding valine residue at the C-terminus or the 24 amino acids SeEA peptide **10c** which includes an internal cysteine residue.

The triselenide **8** is reduced *in situ* by TCEP into diselenol **3** with the concomitant production of the selenophosphine derived from TCEP, *i.e.* Se = TCEP, which is indeed easily detected by LC-MS (see ESI†). We have recently shown that Se = TCEP can inhibit the free-radical process leading to the deselenization of Sec by TCEP.⁵⁸ Therefore, the presence of Se = TCEP in the exchange mixture might inhibit the formation of peptide **11a** as well. In support of this hypothesis, the combined use of diselenide **7** and of metallic selenium as an alternative for the *in situ* formation of Se = TCEP led to a significant decrease in the proportion of peptide **11a** in the exchange mixture and allowed the isolation of **10a** in good yield (**10a** 56%, Fig. 1D and Scheme 4).§

Interestingly, SeEA peptides could also be prepared from C-terminal peptide thioesters by using triselenide **8** (**12a** → **10a**

Scheme 4) or diselenide **7** and metallic selenium (**29** → **30**, Scheme 6). This exchange reaction is an interesting alternative for accessing SeEA peptides, since a large variety of synthetic methods are now available for producing peptide thioesters.²³ Note that although SeEA peptides are probably produced in the exchange mixture in their reduced diselenol form due to the excess of TCEP, they eluted during analytical or preparative HPLC only as cyclic diselenides such as **10**. This is probably due to the rapid air oxidation of the diselenols into the corresponding cyclic diselenides during workup.

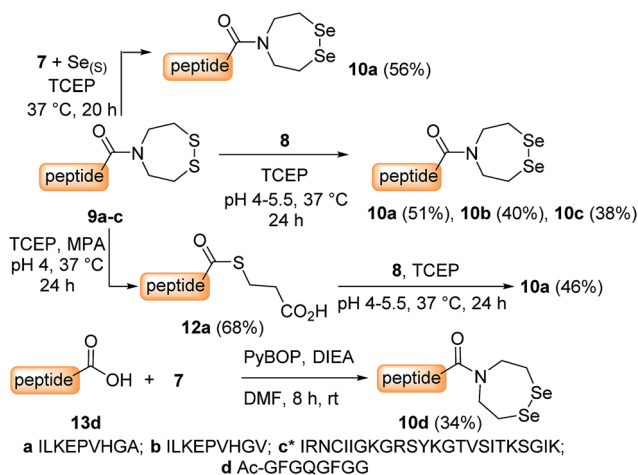
The last synthetic access to SeEA peptides which is depicted in Scheme 4 is based on the coupling of diselenide **7** to a peptide acid using PyBOP/DIEA activation (**13d** → **10d**, Scheme 4). This method which has been described elsewhere⁴⁶ is not compatible with peptides featuring functional residues. It is thus of limited scope in comparison with the exchange processes discussed above.

To summarize at this stage, the chemoselective exchange of SEA peptides or peptide thioesters by bis(2-selenylethyl)amine **3** generated *in situ* by reduction of diselenide **7** or triselenide **8** constitutes a mild access to SeEA peptides. The high stability of SeEA peptides during the HPLC purification step and upon storage thanks to their tertiary amide structure is also worth mentioning.

Kinetic studies

We next determined the kinetic rates of the reaction of SEA peptides **9a,b**, SeEA peptides **10a,b** or MPA peptide thioesters **12a,b** with Cys peptide **14** (Fig. 2A). The ligations with SeEA peptides **10a,b** were performed in the presence of Se = TCEP to avoid the deselenization of the SeEA group by TCEP.⁵⁸ The fraction ligated was quantified by HPLC (Fig. 2B). The data were nicely fitted to a first order kinetic law from which the $t_{1/2}$ were determined (see the insert in Fig. 2B). Interestingly, the SeEA peptides proved to be significantly more reactive than SEA peptides. For example, the $t_{1/2}$ for SEA peptide **9a** with a C-terminal alanine residue was 3.3 h, while the $t_{1/2}$ for the SeEA analog **10a** was 0.24 h only, that is ~14 fold lower. Peptide **15a** produced by reacting SeEA peptide **10a** with Cys peptide **14** was isolated by HPLC with a 56% yield. No hydrolysis of the SeEA peptide nor side-reactions with nucleophilic residues such as lysine could be detected by LC-MS. Moreover, analysis of peptide **15a** by chiral GC-MS after total acid hydrolysis showed a D-Ala content <0.43%.⁵⁹ Similarly, an ~8 fold decrease of the $t_{1/2}$ was observed for valine as C-terminal residue (SEA peptide **9b** $t_{1/2}$ 23 h, SeEA peptide **10b** $t_{1/2}$ 2.9 h). Interestingly, SeEA peptides **10a,b** were also more reactive than the corresponding MPA peptide thioesters **12a,b** (Fig. 2). Therefore, the observed order of reactivity at pH 7 was SeEA peptides > MPA peptide thioester ≫ SEA peptide.

We further performed the experiment depicted in Fig. 3 with the aim to identify the rate limiting step of the SEA ligation process, and propose an explanation for the significant impact of the S → Se substitution on the reactivity of the bis(2-chalcogenoethyl)amido group. The first experiment corresponds to a typical SEA ligation process (Fig. 3A, condition A). In these



Scheme 4 Synthesis of SeEA peptides. *The Cys residue is temporarily protected by a *tert*-butylsulfenyl (tBuS) group in **9c**. It is removed during the exchange process.



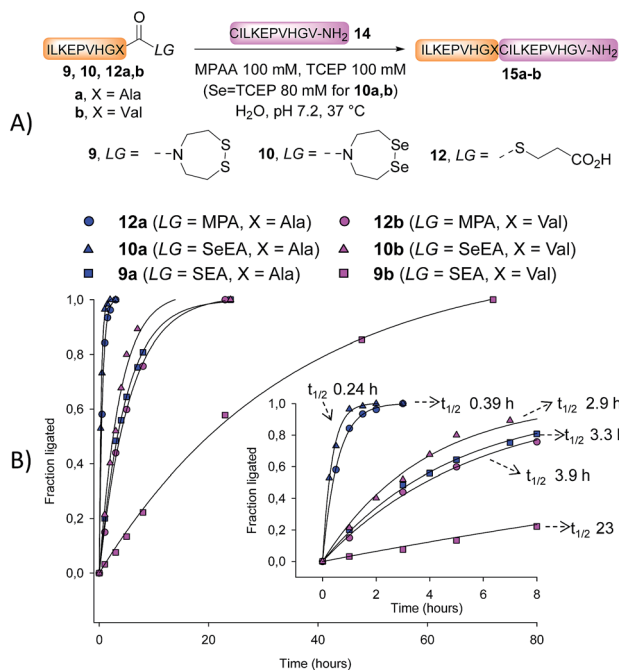


Fig. 2 Kinetic rates for the reaction of SEA peptides **9a,b**, SeEA peptides **10a,b** or MPA peptide thioesters **12a,b** with Cys peptide **14**. The data were fitted to a first order kinetic law (continuous curves). The half-times ($t_{1/2}$) of the reactions are indicated in the insert. Peptide concentration was 3.5 mM.

conditions, the SEA amide dithiol form is the only species observed for the reduced SEA peptide **9b** during ligation. In the second experiment (Fig. 3A, condition B), the SEA peptide **9b** was first reduced and incubated at pH 1 and 37 °C for 8 h to convert the SEA amide form into the SEA thioester form which represents $\sim 90\%$ at equilibrium.²⁹ Then, peptide **14** was added together with MPAA and the pH was quickly adjusted to 5.5. Fig. 3B shows the kinetic rate for these two ligation reactions as determined by HPLC. The data obtained using condition A obeys to a first-order kinetic law with a reaction half-life ($t_{1/2}$) of 27.8 h. In contrast, the kinetic data obtained using condition B clearly show an initial burst of ligation product **15b** formation followed by a slower phase. The first phase is due to the rapid consumption of the SEA thioester form produced during the pre-equilibration step. However, it is also competitively converted back to the SEA amide form, which is significantly less reactive and at the origin of the slower phase. The significant increase of the ligation rate which is observed when the SEA amide form is pre-equilibrated into the SEA thioester form supports the N,S -acyl shift of the SEA group to be the rate limiting step of the SEA ligation process.

Another interesting observation is that the rate of SEA ligation is accelerated while the pH is decreased from pH 7.2 to 5.5 (Fig. 4). The significant increase of the SEA ligation rate upon acidification, which has also been observed with peptidyl prolyl SEA peptides³⁸ or with peptides featuring a SEA group on an aspartic or glutamic side-chain,³⁷ shows that the transfer of a proton occurs in the transition state of the rate limiting step, that is the N,S -acyl shift.

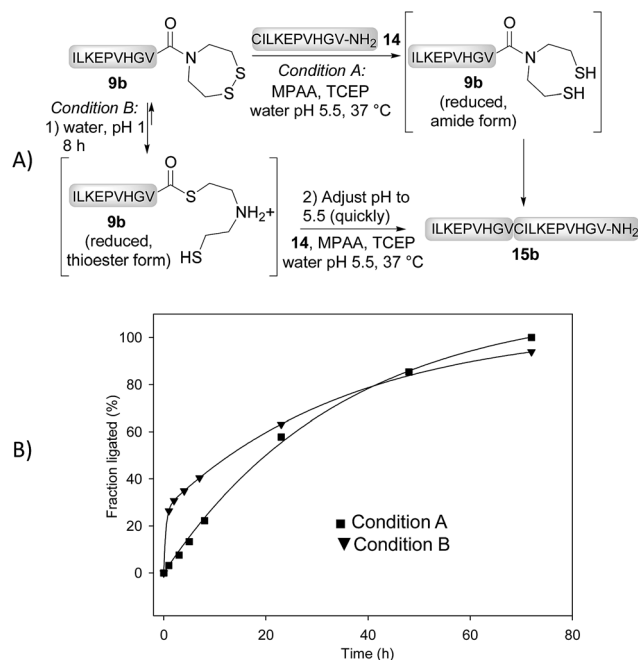


Fig. 3 SEA ligation without (condition A) or with pre-equilibration of the SEA amide form into the SEA thioester form (condition B).

Preliminary DFT computations using N -formyl Gly-SEA **16a** as a model for SEA peptides revealed that amongst the different plausible neutral or anionic transition states which were identified up to now for the N,S -acyl shift, the anionic and concerted transition state TS_{SEA} shown in Fig. 5A is by far the one displaying the lowest barrier of activation. In TS_{SEA} , the neutral 2-sulfenylethyl limb is involved in an intramolecular S–H–N interaction. The intramolecular transfer of the thiol proton to the amide nitrogen in TS_{SEA} is fully consistent with the experimentally observed pH-dependency of the SEA ligation

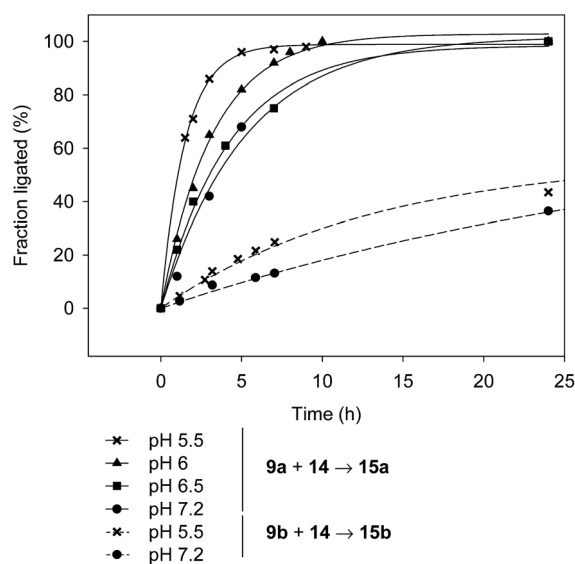


Fig. 4 The effect of the pH on the rate of SEA ligation (peptide **9a** or **9b** 7 mM, peptide **14** 10.5 mM, MPAA 200 mM, TCEP 200 mM, 37 °C).



reaction. The N-protonation of amides is known to weaken the $n_N \rightarrow \pi_{C=O}$ conjugation, to disrupt the amide bond planarity and thus to increase the electrophilicity of the carbonyl toward nucleophiles.⁶⁰ N-protonation is often used by biocatalysts for catalyzing biochemical processes such as amide bond isomerization,⁶¹ hydrolysis or cleavage.⁶² The N-protonation of amides ($pK_a^N \sim -7$ (ref. 63)), although thermodynamically unfavourable over O-protonation ($pK_a^O \sim 0$), can be facilitated within constrained systems which favour a hydrogen bond donation to the amide nitrogen.⁶⁴

Importantly, the intramolecular N-protonation/nucleophilic attack mechanism permitted by the presence of two N-mercaptoethyl groups in close proximity to the amide group might explain the capacity of the SEA group to act as a good thioester surrogate at neutral pH,³² in comparison with related *N,S*-acyl shift systems which feature only one mercaptoethyl limb such as *N*-methyl cysteine²⁵ or *N*-butyl-*N*-(2-sulfanylethyl)amido²⁷ thioester surrogates. It also allows proposing an explanation for the greater reactivity of the SeEA group compared to the SEA one. Indeed, the pK_a of selenols such as selenocysteine or selenocysteine ($pK_a \sim 5.5$ (ref. 64)) are about 3 pK_a units lower than the pK_a of the sulfur analogs, cysteamine or cysteine ($pK_a \sim 8.2$ (ref. 65 and 66)). This difference in pK_a , which is reminiscent of the higher acidity of hydrogen selenide (pK_a 3.8) compared to hydrogen sulfide (pK_a 7.0), is also found in S(Se)EA models **16a,b** those pK_a was estimated using ACD/Labs software (Fig. 5B).⁶⁷ Therefore, the intramolecular protonation of the amide nitrogen and thus the activation of the amide carbonyl might be more pronounced for SeEA peptides. Moreover, the SEA group is mostly present at neutral pH as the dithiol form,

i.e. the productive monoanionic species of type **16a** (Fig. 5A) represents less than 1% of the total. In contrast, the monoanionic species for the SeEA analog is significantly more populated ($\sim 7\%$). Thus, the higher acidity of selenols relative to thiols might contribute to the higher reactivity of the SeEA group by both increasing the concentration of the productive monoanionic species and lowering the barrier of activation for the *N,S*-acyl shift. DFT computations aiming at identifying the transition states for the *N,S*-acyl shift are in progress and will be reported in due course.

Kinetically controlled ligation using SEA and SeEA latent thio(seleno)ester peptides

The availability of fast and slow reacting thioester functionalities or thioester surrogates can be useful for designing kinetically controlled ligations (KCLs), which involve an NCL reaction between a Cys peptide and a fast reacting thioester in the presence of a slow reacting one. The field has been pioneered by Kent and coworkers who reacted a peptide arylthioester in the presence of a peptide alkylthioester.⁴⁷ Later on, other KCLs have been developed using *O,S*-⁶⁸ or *N,S*-acyl shift systems.^{33,38,44,69,70} Here we show that SeEA ligation can be performed in the presence of a SEA moiety thanks to the large difference of reactivity between SeEA and SEA groups towards Cys peptides. In particular, the sequential use of SeEA and SEA ligations enabled to develop the three peptide segment assembly method depicted in Scheme 5.

The reduction of the SeEA peptide segment **10c** (or **10d**) required the presence of TCEP at the first stage of the process. Here again, Se = TCEP (80 mM) proved to be a useful additive

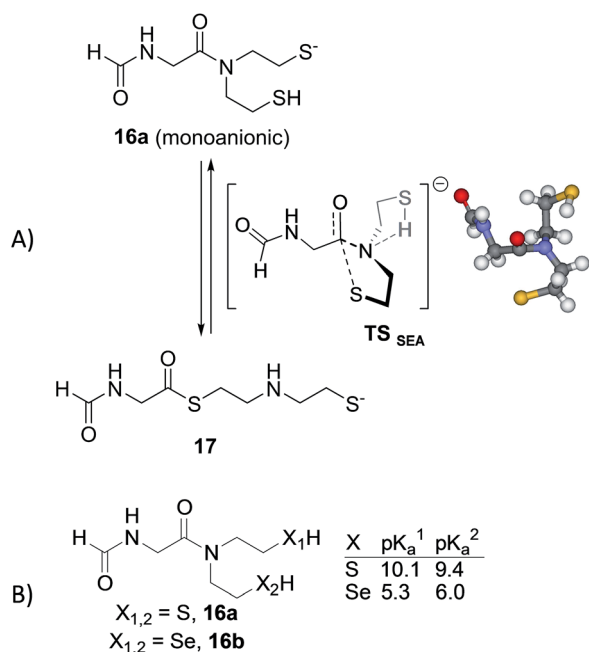
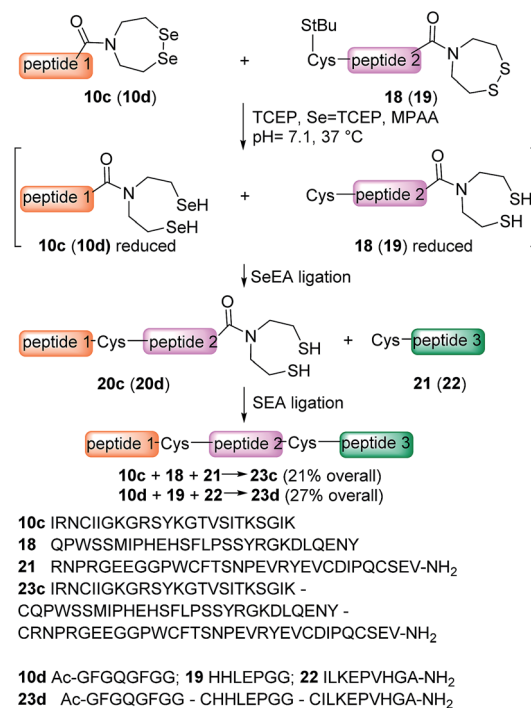


Fig. 5 (A) Hypothetical anionic transition state structure for the intramolecular *N*-to-*S* acyl transfer of the SEA group. (B) pK_a s for the SEA thiol or SeEA selenol groups were estimated using ACS/Labs software.



Scheme 5 Kinetically controlled ligation using sequential SeEA and SEA ligations.



for minimizing the deselenization of the SeEA group by TCEP. Once the SeEA ligation was complete, the addition of the Cys peptide **21** (or **22**) triggered the SEA ligation step and the formation of the target polypeptide. The assembly of the model peptide segments **10d**, **19** and **22** involved the formation of two Gly–Cys junctions and yielded the 26 amino acids peptide **23d** (27% overall after HPLC purification). The method was also applied to the synthesis of the 85 amino acids Kringle 1 domain (K1) of hepatocyte growth factor/scatter factor (HGF/SF).^{42,71} This example involved the assembly of peptide segments **10c** (HGF/SF 125–148), **18** (HGF/SF 149–176) and **21** (HGF/SF 177–209) through the sequential formation of Lys–Cys and Tyr–Cys junctions. Fig. 6 highlights the efficiency of the one-pot process which yielded the K1 polypeptide **23c** with 21% overall yield after HPLC purification. The K1 polypeptide produced in this work using the SeEA/SEA KCL method was found to be identical by LC-MS to a sample obtained by another route (see ESI†).⁴²

Synthesis of large SeEA peptides and total synthesis of biotinylated NK1 domain of hepatocyte growth factor

So far, KCL methods were used for accessing proteins composed of less than 100 amino acids. One reason for this is the difficulty in accessing large peptide segments (>50 AA) equipped with a reactive C-terminal thio or selenoester group. Usually, large peptides are prepared by coupling

chemoselectively several unprotected peptide segments together. However, such processes are hardly compatible with the presence of highly reactive functional groups such as selenoesters. A potential solution to this problem is to introduce the selenoester moiety after the chain assembly.

An appealing alternative is to use a latent selenoester during the assembly of the peptide. With such a strategy, the formation of side-products during the assembly of the latent selenoester is minimized, while the KCL process can be triggered in a subsequent step when needed.

In a previous communication we showed that the SeEA group is compatible with the NCL and SEA ligations in the presence of reducing thiols such as MPAA or DTT.⁴⁶ We show here that the sequential use of NCL and SEA ligations in a one-pot process permitted the synthesis of the 118 amino acids SeEA peptide **31** as illustrated in Scheme 6 (21% based on **26**, one-pot process III). Peptide **31** corresponds to HGF 31–148 and was assembled using MPA thioester peptide **26**, SEA peptide **27** and SeEA peptide **30**.

Peptide **26** was produced using one-pot process I by ligating MPA thioester **24** with SEA peptide **25** at pH 7 in the presence of MPAA, followed by the *in situ* exchange of the SEA group by MPA (46% for the two steps). SEA peptide **27** was obtained by standard Fmoc SPPS starting from SEA polystyrene resin.^{32,48} The third segment, *i.e.* SeEA peptide **30**, was produced starting from

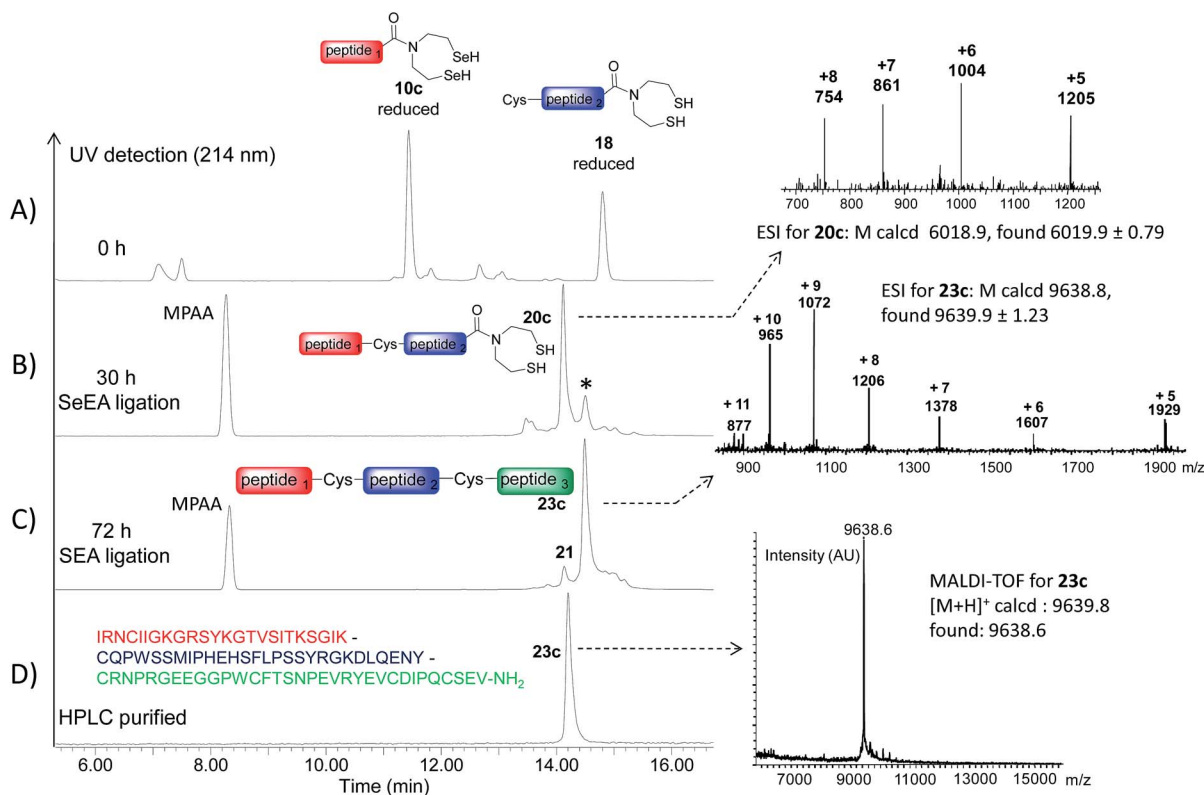
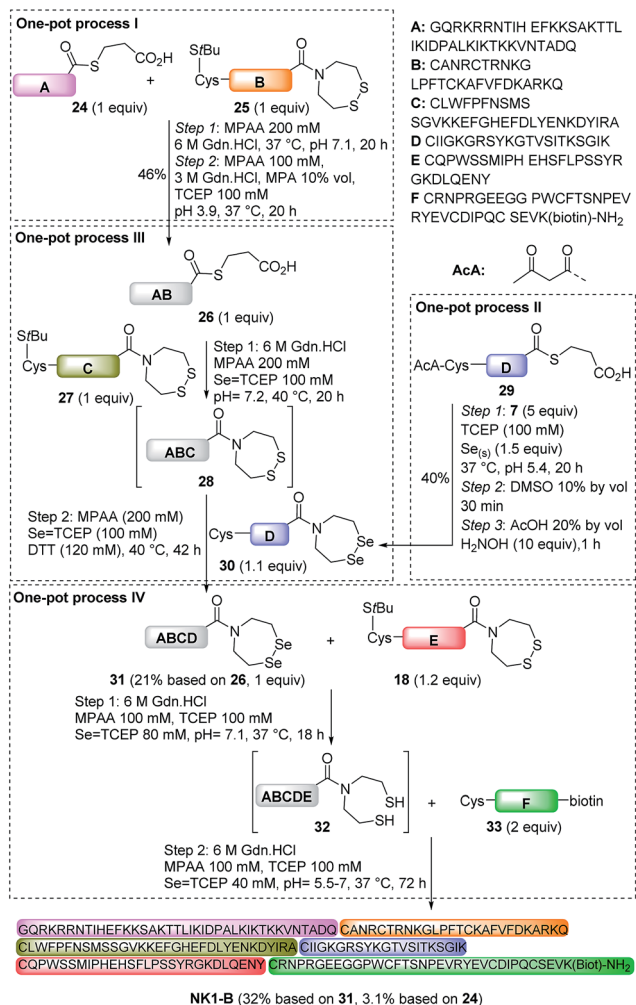


Fig. 6 Monitoring of the one-pot assembly of K1 polypeptide using HPLC. (A) Few minutes after dissolving peptide segments **10c** and **18** (4 mM each). (B) SeEA ligation step after 30 h. The peak marked by an asterisk corresponds to the partial cyclization of peptide **18**. (C) SEA ligation step 72 h after the addition of Cys peptide **21**. For (A–C), the analysis were run on a C3 Zorbax column with a linear gradient of 0% eluent B (ACN/H₂O (4/1 by vol) containing 0.1% TFA) in eluent A (water containing 0.1% TFA) to 100% eluent B in eluent A over 30 min, UV detection at 214 nm. (D) Purified K1 polypeptide. The analysis was run on a C18 Xbridge column using the above conditions.





Scheme 6 Total synthesis of SeEA peptide **31** by a sequential NCL/SEA ligation process (118 amino acids, one-pot process III) and of **NK1-B** protein by a SeEA/SEA KCL process (180 amino acids, one-pot process IV).

MPA thioester **29** using one-pot process II. One-pot process II involved first the introduction of the SeEA group by treatment with diselenide **7**, metallic selenium and TCEP. The second step allowed the formation of the cyclic diselenide by oxidation with DMSO, while the N-terminal acetoacetyl group was removed in the final step by addition of hydroxylamine.⁷² The successful preparation of SeEA peptide **31** set the stage for the assembly of the biotinylated NK1 domain of HGF (**NK1-B**), a protein composed of 180 amino acids (Scheme 6, one-pot process IV). **NK1** protein was modified by a C-terminal biotinylated lysine residue to allow the development of binding assays in the future. **NK1-B** was assembled using the SeEA/SEA KCL method depicted in Scheme 6 (one-pot process IV), which is almost identical to the SeEA/SEA KCL discussed above (see Scheme 5). Gratifyingly, the sequential assembly of SeEA peptide **31**, SEA peptide **18** and Cys peptide **33** furnished successfully **NK1-B** polypeptide (32% based on **31**, 3.1% based on **24**, Fig. 7), showing the usefulness of S(Se)EA chemistry for accessing challenging proteins.

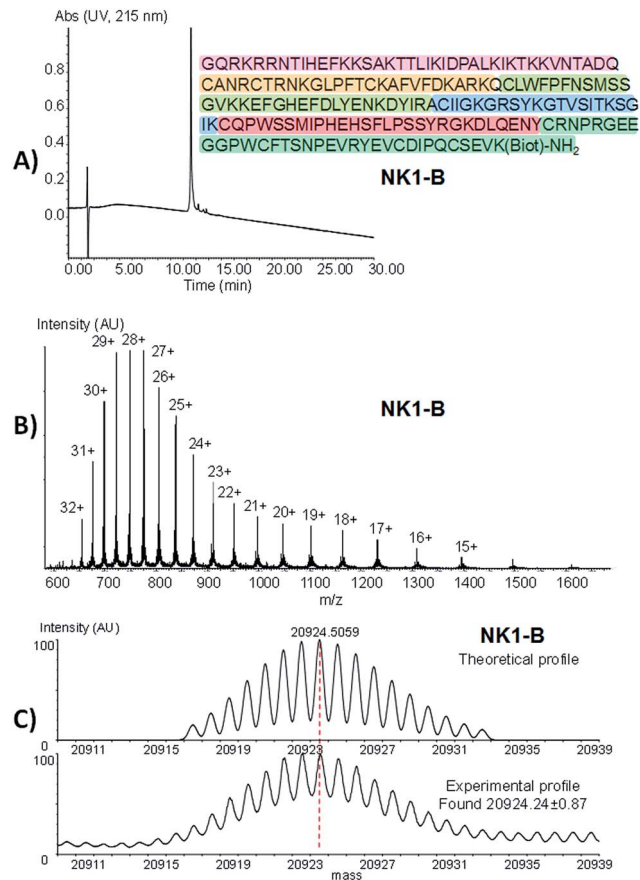


Fig. 7 Characterization of **NK1-B** protein. (A) LC analysis of **NK1-B**. (A) LC trace, eluent A 0.10% formic acid (FA) in water, eluent B 0.10% FA in CH₃CN/water: 4/1 by vol. C3 Zorbax 300SB 3.5 μm (4.6 × 150 mm) column, gradient 0–100% B in 30 min (1 mL min⁻¹, detection 215 nm). (B) ESI MS high resolution spectrum. (C) Deconvoluted spectrum and comparison with the theoretical profile.

Conclusions

In conclusion, SeEA peptides are highly stable peptide derivatives which behave as latent selenoesters upon reduction of the cyclic diselenide. They are easily accessible through the chemoselective transamidation of SEA peptides or the selenol-thioester exchange of thioester peptides in water by an excess of bis(2-selenylethyl)amine. This reagent is formed *in situ* by reduction of 1,2,5-diselenazepane or 1,2,3,6-triselenazocane by TCEP. The presence of Se = TCEP in the exchange mixture minimizes the extrusion of one selenium atom from SeEA diselenide. SeEA peptides react faster than peptide alkylthioesters or SEA peptides with cysteinyl peptides. The large difference of reactivity between SeEA and SEA functionalities enabled designing a one-pot kinetically controlled three peptide segment assembly process working in the N-to-C direction. The method was validated by the total synthesis of hepatocyte growth factor K1 (85 AA) and biotinylated NK1 (**NK1-B**, 180 AA) domains. **NK1-B** is by far the largest polypeptide produced by using a KCL process. The synthesis of a latent SeEA selenoester with a size >100 amino acids and the assembly of a 180 amino



acids protein by a series of one-pot processes show the great potential of S(Se)EA chemistries for accessing challenging proteins.

Acknowledgements

LR and MC thank the Région Nord Pas-de-Calais and the Région Centre councils for a PhD fellowship. YMC thanks the HHMI-Institut Pasteur de Lille program for financial support. We thank the Chemistry Systems Biology (CSB) facility for technical help and advice. JCOMM acknowledges the University of Liège for financial support (WG-13/03) and the F.R.S.-FNRS for the access to the HPC facilities (CéCI). We thank Jean-Michel Saliou for HR MS analyses (P3M facility, Institut Pasteur de Lille).

Notes and references

§ Note that triselenide **8** was found to decompose partially into diselenide **7** plus selenium upon storage (see ESI†). Therefore, the combined use of diselenide **7** and metallic selenium is a useful alternative to triselenide **8** for accessing SeEA peptides.

- 1 K. Mandal, B. L. Pentelute, D. Bang, Z. P. Gates, V. Y. Torbeev and S. B. H. Kent, *Angew. Chem., Int. Ed.*, 2012, **51**, 1481–1486.
- 2 R. J. Clark and D. J. Craik, *Methods Enzymol.*, 2012, **503**, 57–74.
- 3 R. J. Clark and D. J. Craik, *Biopolymers*, 2010, **94**, 414–422.
- 4 G. T. Dolphin, *J. Am. Chem. Soc.*, 2006, **128**, 7287–7290.
- 5 J. A. Camarero, J. Pavel and T. W. Muir, *Angew. Chem., Int. Ed.*, 1998, **37**, 347–349.
- 6 J. A. Camarero and T. W. Muir, *Chem. Commun.*, 1997, 1369–1370.
- 7 C. Unverzagt and Y. Kajihara, *Chem. Soc. Rev.*, 2013, **42**, 4408–4420.
- 8 K. S. Kumar, L. Spasser, L. A. Erlich, S. N. Bavikar and A. Brik, *Angew. Chem., Int. Ed.*, 2011, **49**, 9126–9131.
- 9 K. S. Kumar, S. N. Bavikar, L. Spasser, T. Moyal, S. Ohayon and A. Brik, *Angew. Chem., Int. Ed.*, 2011, **50**, 6137–6141.
- 10 B. Fierz, C. Chatterjee, R. K. McGinty, M. Bar-Dagan, D. P. Raleigh and T. W. Muir, *Nat. Chem. Biol.*, 2011, **7**, 113–119.
- 11 E. Boll, H. Drobecq, N. Ollivier, L. Raibaut, R. Desmet, J. Vicogne and O. Melnyk, *Chem. Sci.*, 2014, **5**, 2017–2022.
- 12 E. Boll, H. Drobecq, N. Ollivier, A. Blanpain, L. Raibaut, R. Desmet, J. Vicogne and O. Melnyk, *Nat. Protoc.*, 2015, **10**, 269–292.
- 13 P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. Kent, *Science*, 1994, **266**, 776–779.
- 14 S. B. Kent, *Chem. Soc. Rev.*, 2009, **38**, 338–351.
- 15 J. W. Bode, R. M. Fox and K. D. Baucom, *Angew. Chem., Int. Ed.*, 2006, **45**, 1248–1252.
- 16 E. Saxon, J. I. Armstrong and C. R. Bertozzi, *Org. Lett.*, 2000, **2**, 2141–2143.
- 17 B. L. Nilsson, L. L. Kiessling and R. T. Raines, *Org. Lett.*, 2000, **2**, 1939–1941.
- 18 V. R. Pattabiraman and J. W. Bode, *Nature*, 2011, **480**, 471–479.
- 19 Y. Zhang, C. Xu, H. Y. Lam, C. L. Lee and X. Li, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 6657–6662.
- 20 T. Durek and P. F. Alewood, *Angew. Chem., Int. Ed.*, 2011, **50**, 12042–12045.
- 21 N. J. Mitchell, L. R. Malins, X. Liu, R. E. Thompson, B. Chan, L. Radom and R. J. Payne, *J. Am. Chem. Soc.*, 2015, **137**, 14011–14014.
- 22 A. Ghassemian, X. Vila-Farres, P. F. Alewood and T. Durek, *Bioorg. Med. Chem.*, 2013, **21**, 3473–3478.
- 23 F. Mende and O. Seitz, *Angew. Chem., Int. Ed.*, 2011, **50**, 1232–1240.
- 24 F. Nagaike, Y. Onuma, C. Kanazawa, H. Hojo, A. Ueki, Y. Nakahara and Y. Nakahara, *Org. Lett.*, 2006, **8**, 4465–4468.
- 25 H. Hojo, Y. Onuma, Y. Akimoto, Y. Nakahara and Y. Nakahara, *Tetrahedron Lett.*, 2007, **48**, 25–28.
- 26 S. Tsuda, A. Shigenaga, K. Bando and A. Otaka, *Org. Lett.*, 2009, **11**, 823–826.
- 27 M. Taichi, X. Hemu, Y. Qiu and J. P. Tam, *Org. Lett.*, 2013, **15**, 2620–2623.
- 28 J.-S. Zheng, S. Tang, Y.-C. Huang and L. Liu, *Acc. Chem. Res.*, 2013, **46**, 2475–2484.
- 29 J. Dheur, N. Ollivier, A. Vallin and O. Melnyk, *J. Org. Chem.*, 2011, **76**, 3194–3202.
- 30 R. B. Martin, S. Lowey, E. L. Elson and J. T. Edsall, *J. Am. Chem. Soc.*, 1959, **81**, 5089–5095.
- 31 J. Kang, J. P. Richardson and D. Macmillan, *Chem. Commun.*, 2009, 407–409.
- 32 N. Ollivier, J. Dheur, R. Mhidia, A. Blanpain and O. Melnyk, *Org. Lett.*, 2010, **12**, 5238–5241.
- 33 K. Sato, A. Shigenaga, K. Tsuji, S. Tsuda, Y. Sumikawa, K. Sakamoto and A. Otaka, *ChemBioChem*, 2011, **12**, 1840–1844.
- 34 F. Burlina, G. Papageorgiou, C. Morris, P. D. White and J. Offer, *Chem. Sci.*, 2014, **5**, 766–770.
- 35 V. P. Terrier, H. Adihou, M. Arnould, A. F. Delmas and V. Aucagne, *Chem. Sci.*, 2016, **7**, 339–345.
- 36 O. Melnyk and V. Agouridas, *Curr. Opin. Chem. Biol.*, 2014, **22**, 137–145.
- 37 E. Boll, J. Dheur, H. Drobecq and O. Melnyk, *Org. Lett.*, 2012, **14**, 2222–2225.
- 38 L. Raibaut, P. Seeberger and O. Melnyk, *Org. Lett.*, 2013, **15**, 5516–5519.
- 39 R. Desmet, M. Pauzuolis, E. Boll, H. Drobecq, L. Raibaut and O. Melnyk, *Org. Lett.*, 2015, **17**, 3354–3357.
- 40 E. C. Johnson and S. B. Kent, *J. Am. Chem. Soc.*, 2006, **128**, 6640–6646.
- 41 L. Raibaut, J. Vicogne, B. Leclercq, H. Drobecq, R. Desmet and O. Melnyk, *Bioorg. Med. Chem.*, 2013, **21**, 3486–3494.
- 42 N. Ollivier, J. Vicogne, A. Vallin, H. Drobecq, R. Desmet, O. El-Mahdi, B. Leclercq, G. Goormachtigh, V. Fafeur and O. Melnyk, *Angew. Chem., Int. Ed.*, 2012, **51**, 209–213.
- 43 F. Ancot, C. Leroy, G. Muharram, J. Lefebvre, J. Vicogne, A. Lemiere, Z. Kherrouche, B. Foveau, A. Pourtier, O. Melnyk, S. Giordano, A. Chotteau-Lelievre and D. Tulasne, *Traffic*, 2012, **13**, 1261–1272.



- 44 R. Yang, W. Hou, X. Zhang and C.-F. Liu, *Org. Lett.*, 2012, **14**, 374–377.
- 45 W. Hou, X. Zhang, F. Li and C. F. Liu, *Org. Lett.*, 2011, **13**, 386–389.
- 46 L. Raibaut, H. Drobecq and O. Melnyk, *Org. Lett.*, 2015, **17**, 3636–3639.
- 47 D. Bang, B. L. Pentelute and S. B. Kent, *Angew. Chem., Int. Ed.*, 2006, **45**, 3985–3988.
- 48 N. Ollivier, L. Raibaut, A. Blanpain, R. Desmet, J. Dheur, R. Mhidia, E. Boll, H. Drobecq, S. L. Pira and O. Melnyk, *J. Pept. Sci.*, 2014, **20**, 92–97.
- 49 J. Dheur, N. Ollivier and O. Melnyk, *Org. Lett.*, 2011, **13**, 1560–1563.
- 50 F. W. Hoffmann, R. J. Ess, T. C. Simmons and R. S. Hanzel, *J. Am. Chem. Soc.*, 1956, **78**, 6414.
- 51 C. Walling and R. Rabinowitz, *J. Am. Chem. Soc.*, 1957, **79**, 5326.
- 52 M. D. Gieselman, L. Xie and W. A. van der Donk, *Org. Lett.*, 2001, **3**, 1331–1334.
- 53 N. Metanis, E. Keinan and P. E. Dawson, *Angew. Chem., Int. Ed.*, 2010, **49**, 7049–7053.
- 54 S. Dery, P. S. Reddy, L. Dery, R. Mousa, R. N. Dardashti and N. Metanis, *Chem. Sci.*, 2015, **6**, 6207–6212.
- 55 L. Henriksen, in *The Chemistry of Organic Selenium and Tellurium Compounds*, ed. S. Patai, John Wiley & Sons, Chichester, 1987, vol. 2, pp. 393–420.
- 56 J. Y. C. Chu and D. G. Marsh, *J. Org. Chem.*, 1976, **41**, 3204–3205.
- 57 T. Koide, H. Itoh, A. Otaka, H. Yasui, M. Kuroda, N. Esaki, K. Soda and N. Fujii, *Chem. Pharm. Bull.*, 1993, **41**, 502–506.
- 58 N. Ollivier, A. Blanpain, L. Raibaut, H. Drobecq and O. Melnyk, *Org. Lett.*, 2014, **16**, 3032–4035.
- 59 W. Amelung and S. Brodowski, *Anal. Chem.*, 2002, **74**, 3239–3246.
- 60 R. Szostak, J. Aube and M. Szostak, *Chem. Commun.*, 2015, **51**, 6395–6398.
- 61 C. Cox and T. Lectka, *Org. Lett.*, 1999, **1**, 749–752.
- 62 D. G. A. Johansson, G. Wallin, A. Sandberg, B. Macao, J. Aqvist and T. Härd, *J. Am. Chem. Soc.*, 2009, **131**, 9475–9477.
- 63 A. R. Fersht, *J. Am. Chem. Soc.*, 1971, **93**, 3504–3515.
- 64 A. P. Arnold, K. S. Tan and D. L. Rabenstein, *Inorg. Chem.*, 1986, **25**, 2433–2437.
- 65 L. Riauba, G. Niaura, O. Eicher-Lorka and E. Butkus, *J. Phys. Chem. A*, 2006, **110**, 13394–13404.
- 66 S. G. Taje, B. S. Tolbert, R. Basavappa and B. L. Miller, *J. Am. Chem. Soc.*, 2004, **126**, 10508–10509.
- 67 M. Meloun and S. Bordovska, *Anal. Bioanal. Chem.*, 2007, **389**, 1267–1281.
- 68 J. S. Zheng, H. K. Cui, G. M. Fang, W. X. Xi and L. Liu, *ChemBioChem*, 2010, **11**, 511–515.
- 69 H. Ding, A. Shigenaga, K. Sato, K. Morishita and A. Otaka, *Org. Lett.*, 2011, **13**, 5588–5591.
- 70 T. Nakamura, A. Shigenaga, K. Sato, Y. Tsuda, K. Sakamoto and A. Otaka, *Chem. Commun.*, 2014, **50**, 58–60.
- 71 C. Simonneau, L. Berenice, A. Mougel, E. Adriaenssens, C. Paquet, L. Raibaut, N. Ollivier, H. Drobecq, J. Marcoux, S. Cianferani, D. Tulasne, H. de Jonge, O. Melnyk and J. Vicogne, *Chem. Sci.*, 2015, **6**, 2110–2121.
- 72 E. Boll, J. P. Ebran, H. Drobecq, O. El-Mahdi, L. Raibaut, N. Ollivier and O. Melnyk, *Org. Lett.*, 2015, **17**, 130–133.

