

UNIVERSITÉ D'ORLÉANS

ÉCOLE DOCTORALE SANTE, SCIENCES BIOLOGIQUES ET CHIMIE DU VIVANT

Centre de Biophysique Moléculaire, Orléans

THÈSE

présentée par :

Skander ABOUD

soutenue le : 17 Décembre 2020

pour obtenir le grade de : **Docteur de l'Université d'Orléans**

Discipline/ Spécialité : Chimie

**Development of chemo-enzymatic
methodologies for the total synthesis of
proteins through solid-supported
chemical ligation**

THÈSE dirigée par :

Dr. Vincent AUCAGNE

Directeur de recherche CNRS, CBM, Orléans

RAPPORTEURS :

Dr. Vladimir TORBEEV

Maître de conférences, CNRS, ISIS, Strasbourg

Dr. Muriel AMBLARD

Directrice de recherche, CNRS, IBMM, Montpellier

JURY :

Dr. Muriel AMBLARD

Directrice de recherche CNRS, IBMM, Montpellier

Dr. Vincent AUCAGNE

Directeur de recherche CNRS, CBM, Orléans

Prof. Chrystel LOPIN-BON

Professeur des Universités, ICOA, Orléans

Prof. Sébastien PAPOT

Professeur des Universités, IC2MP, Poitiers

Dr. Vladimir TORBEEV

Maître de conférences, ISIS, Strasbourg

Acknowledgements

Le travail de thèse présenté dans ce manuscrit a été réalisé au Centre de Biophysique Moléculaire (CBM), unité propre du Centre National de la Recherche Scientifiques (CNRS).

Je remercie l'école doctorale SSBCV pour l'octroi d'une bourse de thèse, et l'agence nationale de la recherche (ANR) pour le financement des coûts associés. Je souhaite également remercier la directrice du laboratoire, le Dr. Eva Jakab Tóth pour m'avoir permis d'effectuer cette thèse au CBM.

J'adresse mes remerciements aux Drs. Vladimir Torbeev et Muriel Amblard d'avoir accepté d'être les rapporteurs de ce manuscrit. Je remercie également le Pr Sébastien Papot et le Pr Chrystel Lopin-Bon d'avoir accepté de faire partie du jury soutenance de ma thèse.

Je remercie très chaleureusement le Dr. Vincent Aucagne, mon directeur de thèse, de m'avoir fait confiance pour mener à bien ce projet. Je tiens aussi à le remercier pour la riche formation qu'il m'a dispensée, pour sa disponibilité quotidienne, son soutien continu, son encadrement efficace et pour tous les conseils judicieux qu'il m'a prodigués. J'ai beaucoup appris à ses côtés durant ces trois années, et je repars avec le sentiment d'avoir grandi humainement et professionnellement.

Je souhaiterais aussi exprimer ma gratitude au Dr. Agnès Delmas, pour m'avoir fait bénéficier de sa grande expérience scientifique.

Je remercie le Dr. Mehdi Amoura pour avoir fait tout son possible pour faciliter mon installation à Orléans et mon intégration au sein du laboratoire, ainsi que pour son soutien scientifique.

Je remercie également Florian Kersaudy pour sa complicité et sa disponibilité pendant ces trois années, que ce soit pour discuter de mes travaux, préparer mes présentations ou même pour m'aider à réparer une machine.

Je tiens également à remercier les autres membres de l'équipe, et tout particulièrement le Dr. Véronique Piller pour son aide et ses conseils sur les enzymes. Je tiens aussi à préciser qu'un grand nombre de peptides utilisés pour ce travail ont été synthétisés par Jean-Baptiste Madinier, Ingénieur d'Etude CNRS dans l'équipe.

Un grand merci au Dr. Cyril Colas pour l'acquisition des spectres de masse haute résolution, au Dr. Franck Coste pour m'avoir initié au Dichroïsme circulaire, à Michel Doudeau et au Dr. Hélène Bénédicti pour la caractérisation biochimique de la protéine SUMO synthétisée dans le chapitre 4.

Je souhaite également remercier Laura Lefevre, Mathilde Seinfeld et El Hadji Cisse, les trois stagiaires que j'ai encadrés durant ma thèse et qui ont grandement contribué à l'avancée de mes travaux.

J'adresse mes remerciements chaleureux à tous mes amis orléanais (Julien, Saida, Anton, Justine, Cédric, Lauren, Vinni, Héloïse) pour tous les moments passés ensemble (mais aussi pour avoir gardé Chelsea pendant mes absences), ce fut un plaisir de partager ces trois années à Orléans avec vous, je garde de merveilleux souvenirs de nos repas, sorties, soirées et voyages. Merci encore pour tous ces moments.

Je souhaite aussi remercier mes amis montpelliérains : ni les années ni les kilomètres n'ont pu altérer notre amitié et notre complicité.

Je remercie aussi mes collègues de mon mandat de président au sein du bureau de l'association des doctorants du CBM ainsi que tous les membres du CBM que j'ai eu plaisir à côtoyer pendant ma thèse.

Je remercie aussi le Dr. Fabienne Medecin ainsi que Corine Galindo pour m'avoir permis d'enseigner la Chimie en solution.

Acknowledgments

Je souhaiterais adresser un remerciement tout particulier à ma femme Laura pour son soutien inconditionnel tout au long de ces 3 ans et bien avant. À nos futures aventures ensemble.

Enfin je souhaite témoigner de ma profonde gratitude à ma famille, pour tous les sacrifices qu'ils ont consentis ainsi que pour leur soutien tant matériel que moral durant toutes ces années.

Foreword

This manuscript was prepared using the original publications associated with my thesis work. Except for the first chapter which is a literature review published as a book chapter, each publication is accompanied with a specific bibliographic introduction in order to offer to the reader an opportunity to make his own idea about the novelty of the described work as compared with the published literature. Note that we do not intend to submit the last chapter for publication as it is, as it represents preliminary data that will hopefully be a solid basis for a future work and future publications: it is therefore presented in a more classical thesis format, less condensed than a publication style.

For ease of reading, bibliographical references and figures are specific to one chapter. Therefore, their numbering is re-initialized at the beginning of each chapter. Due to this format, many key references are repeated several times. Compound's referencing is also specific to each chapter, and a few compounds are used in different chapters with different numbering. We however tried to make this as clear as possible, with also referring to the alternate numbering when possible. Finally, the publications as well as the corresponding "supporting information" documents are inserted at the end of each chapter in their original form. This leads to repetition of the general methods and synthetic protocols, but we preferred this option rather than the compilation of a global experimental part at the end of the manuscript, that we thought to be more difficult to read.

Abbreviations

aa: Amino acid

Ac: Acetyl

Aca: Acetoacetyloxime

Acm: Acetamidomethyl

ADC: Antibody-drug conjugate

Ades: 2-Amino-2,2-dimethylsulfanyl

Aloc: Allyloxycarbonyl

AzF: Azidophenylalanine

BAL: Backbone amide linker

Boc: *tert*-Butyloxycarbonyl

Bn: Benzyl

CAACU: Cysteine aminoethylation-assisted chemical ubiquitination

CAM: Carboxamidomethyl ester

CAN: Canaline

CatB: Cathepsin B

CD: Circular Dichroism

CIP: Calf intestinal alkaline phosphatase

CLEAR: Cross-linked ethoxylate-acrylate resin

CM: ChemMatrix

CPE: Cysteiny l prolyl ester

CPG: Controlled pore glass

CPP: Cell penetrating peptide

CuAAC: Copper(I)-atalyzed azide/alkyne cycloaddition

Cy5: Cyanine5 fluorophore

DBCO: Dibenzocyclooctyne

DBU: 1,8-Diazabicyclo[5.4.0]undec-7-ene

Dde: (1-(4,4-Dimethyl-2,6-dioxocyclo-hexylidene)-3-[2-(2-aminoethoxy) ethoxy]-propyl)]

Dbz: 3,4-Diaminobenzoyl

DCC: *N,N'*-Dicyclohexylcarbodiimide

Ddap: 2-(7-Amino-1-hydroxyheptylidene)-5,5-dimethylcyclohexane-1,3-dione

Abbreviations

DIC: *N,N'*-Diisopropylcarbodiimide

DIEA: *N,N*-Diisopropylethylamine

DCM: Dichloromethane

DFT: Density-functional theory

DMAP: 4-(Dimethylamino)pyridine

Dmmb: *N*-4,5-Dimethoxy-2-mercapto-benzyl

DMF: *N,N*-Dimethylformamide

DMSO: Dimethyl sulfoxide

DNA: Deoxyribonucleic acid

DODT: 3,6-Dioxa-1,8-octanedithiol

Dox: Doxorubicin

DTT: Dithiothreitol

EDT: Ethanedithiol

EDTA: Ethylenediaminetetraacetic acid

EETI-II: Ecballium elaterium trypsin inhibitor II

EPL: Expressed protein ligation

EPR: Enhanced permeability and retention effect

Equiv.: Equivalent

ESI: Electrospray ionization

Et: Ethyl

FDA: U.S. food and drug administration

Fmoc: 9-Fluorenylmethoxycarbonyl

FRET: Förster resonance energy transfer

Glp: Pyroglutamyl amide

Gn.HCl: Guanidine hydrochloride

GV-PLA2: Group V phospholipase A2

HATU: 1-*bis*(Dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxide hexafluorophosphate

HEPES: 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid

HFIP: Hexafluoroisopropanol

HGF: Human growth factor

Hmb: 2-Hydroxy-4-methoxybenzyl

Abbreviations

HMBA: 4-Hydroxymethylbenzoic acid

HMBC: Heteronuclear multiple bond correlation

Hmp: Hydroxy-3-mercaptopropionamide

HMPA: 4-(Hydroxymethyl)phenoxyacetic

HMPP: 4-(Hydroxymethyl)phenoxypropionic

HOA: Hydroxyoctanoic acid

HOBT: 3*H*-[1,2,3]-Triazolo[4,5-*b*]pyridin-3-ol

HPLC: High-pressure liquid chromatography

HRMS: High-resolution mass spectrometry

HSQC: Heteronuclear single quantum coherence

HMPA: *N*-(2-Hydroxypropyl)-methacrylamide

***i*Bu:** Isobutyl

ICL: Isopeptide chemical ligation

***i*Pr:** Isopropyl

KAHA: α -Ketoacid-hydroxylamine ligation

LCMS: High-pressure liquid chromatography coupled with mass spectrometry

LipB: Lipase B

M: Molecular mass

MALDI: Matrix-assisted laser desorption ionization

Me: Methyl

MEGA: *N*-Mercaptoethoxyglycinamide

MESNa: Sodium mercaptoethanesulfonate

MIF: Macrophage migration inhibitory factor

MMAE: Monomethyl auristatin E

MMP: Matrix metalloproteinase

MPA: 3-Mercaptopropionic acid

MPAA: 4-Mercaptophenylacetic acid

MPAL: 3-Mercaptopropionamide linker

MUC1: Mucin1

MRI: Magnetic resonance imaging

MS: Mass spectrometry

Msc: Methylsulfonylethyloxycarbonyl

Abbreviations

MSNPs: Silica nanoparticle nanocarriers

***m/z*:** Mass-to-charge ratio

N₃-ESOC: 2-[2-(2-Azido-ethoxy)-ethyl-sulfonyl]-2-ethoxycarbonyl

N₃-Dtpp: 1-Azido-5-[1,3-dimethyl-2,4,6(1H,3H,5H)-trioxopyrimidine-5-ylidene]pentyl

NAC: *N*-Alkylcysteine

NBS: *N*-Bromosuccinimide

NDMBA: *N,N*-Dimethyl barbituric acid

Nbz: *N*-Acyl benzimidazolone

n.d: Non determined

Nf1: Neurofibromin 1

***N-Hnb-Cys*:** *N*-(2-Hydroxy-5-nitrobenzyl)cysteine

NCL: Native chemical ligation

NEP: Neutral endopeptidase

NMP: 1-Methyl-2-pyrrolidone

NMR: Nuclear magnetic resonance

NOESY: Nuclear Overhauser effect spectroscopy

Nvoc: 6-Nitroveratryloxycarbonyl

Oxyma: Ethyl (hydroxyimino)cyanoacetate

***p*:** Para

PA: Propargylamide

Pbf: 2,2,4,6,7-Pentamethyl-2H-benzofuran-5-sulfonyl

Phacm: Phenylacetamidomethyl

PEG: Poly(ethyleneglycol)

PEGA: Acrylamide-PEG copolymer

PGA: Penicillin G acylase

PMP: Protein metallophosphatases

PNA: Peptide nucleic acid

ppm: Parts per million

PS: Polystyrene

PTM: Post-translational modification

PTL: Peptidomimetic triazole ligation

PyAOP: 7-Azabenzotriazol-1-yl-oxyltripyrrolidinophosphonium hexafluorophosphate

Abbreviations

RNA: Ribonucleic acid
rSAP: Recombinant shrimp alkaline phosphatase
ROESY: Rotating Overhauser effect spectroscopy
ROS: Reactive oxygen species
SCAL: Safety catch acid-labile linker
SEA: *bis*(2-Sulfanylethyl)amino
SEAlide: *N*-Aminoacyl-*N*-sulfanylethyl-anilides
SEAOxy: *N*-Sulfanylethylaminooxybutyramides
SPAAC : Strain-promoted alkyne/azide cycloaddition
SPCL: Solid phase chemical ligation
SPOCC: Superpermeable organic combinatorial chemistry resin
SPPS: Solid phase peptide synthesis
SUMO: Small ubiquitin-like modifier
tBu: *tert*-Butyl
TCEP: *tris*(2-Carboxyethyl)phosphine
TEA: Thioethylalkylamido
TEV: Tobacco etch virus
TFA: Trifluoroacetic acid
TFE: 2,2,2-Trifluoroethanol
Thz: Thiazolidine-4-carboxylic acid
THPTA: *tris*(3-Hydroxypropyltriazolylmethyl)amine
TIS: Triisopropylsilane
TIPS: Tri-*isopropylsilyl*
Trt: Trityl
TRIS: *tris*(Hydroxymethyl)aminomethane
UV: Ultraviolet
VA.044: 2,2'-Azobis[2-(2-imidazolin-2-yl)propane] dihydrochloride
vMIP: Viral macrophage inflammatory protein
v/v : Volume per volume percentage concentration

General table of contents

Acknowledgements	3
Foreword	6
Abbreviations	7
General table of contents	12
Résumé de thèse	14
General Introduction	25
Chapter 1: Solid Phase Chemical Ligation (Book chapter published in Wiley)	28
I. Introduction	30
II. SPCL in the C-to-N direction.....	34
III. SPCL in the N-to-C direction.....	42
IV. Post-ligations solid-supported transformations	49
V. Solid support	51
VI. References.....	56
Chapter 2: Enzyme-cleavable linkers for the total synthesis of proteins through SPCL	64
I. Bibliographic study	65
II. Exploration of the use of enzyme-cleavable linkers for SPCL.....	87
III. References for the bibliographic study part.....	91
IV. Publication 2 (Published in Angewandte Chemie International Edition)	101
V. Supporting Information	110
Chapter 3: An improved method for the solid-phase synthesis of <i>N</i>-Hnb-Cys crypto thioester peptides ...	203
I. Bibliographic study	204
II. Aim of this work: Improved synthesis of <i>N</i> -Hnb-Cys peptide crypto-thioesters.....	226
III. References for the bibliographic study part.....	228
IV. Publication 3 (Published in Organic & Biomolecular Chemistry)	236
V. Supporting Information	247
Chapter 4: Solubilizing tags for N-terminal cysteinyl peptide segments	298
I. Bibliographic study.....	299
II. Aim of this work	315
III. References for the bibliographic study part.....	316
IV. Publication 4 (Published in Chemical Science)	322
V. Supporting Information.....	331
Chapter 5: Synthesis of a SUMO proteases-resistant SUMOylated peptide through SPCL	362
I. Bibliographic study	363
II. Aim of this work: towards chemical biology tools to decipher neurofibromin SUMOylation....	377

Table of contents

III.	Results and discussion.....	379
IV.	Experimental part	387
V.	References.....	402
Conclusion and perspectives		410

Résumé de thèse

La production de protéines par chimie de synthèse est un champ de recherche en plein essor. Cette approche est très complémentaire des techniques recombinantes pour des applications de décryptage de mécanismes biologiques. Les technologies actuelles se focalisent sur des approches d'assemblage modulaire de segments peptidiques de plusieurs dizaines d'acides aminés, par le biais de réactions très sélectives de « ligation chimique ». Au sein de notre équipe, un accent particulier a été mis depuis plus de dix ans sur le développement de techniques permettant un assemblage sur support solide (*Solid Phase Chemical Ligation*, SPCL) afin de s'affranchir des étapes de purifications intermédiaires. Le premier chapitre de ce manuscrit constitue une revue exhaustive des méthodes développées pour la synthèse chimique des protéines par ligation chimique sur support solide au cours des 25 dernières années. Ces travaux ont été publiés sous le format d'un chapitre d'un livre traitant la SPCL : S. A. Abboud, V. Aucagne, A. F. Delmas. Solid-phase Chemical Ligation, in *Total Chemical Synthesis of Proteins*, (Eds.: A. Brik, P. Dawson, L. Liu), Wiley, 2021, pp. 259–284.

Dans ce chapitre d'ouvrage, nous nous sommes particulièrement intéressés à la nature des supports solides utilisés, aux différentes stratégies de ligation employées pour l'assemblage des protéines mais aussi à la structure des bras (*linkers*) utilisés pour immobiliser spécifiquement un premier segment peptidique pur et non-protégé sur un support solide. Notre équipe a par le passé introduit une gamme de nouveaux bras utilisables en SPCL. Le principal enjeu est de maîtriser les conditions nécessaires à la coupure de ces *linkers*, afin de décrocher du support solide la protéine assemblée par ligations successives. Cependant, les *linker* développés jusqu'ici font appel à des conditions acides, basiques ou nucléophiles qui, quoique de plus en plus douces, sont encore difficilement compatibles avec certaines cibles et limitent donc leur champ d'application. Le deuxième chapitre de ce manuscrit se focalise sur le développement d'une nouvelle génération de *linker* en explorant une approche enzymatique plutôt que purement chimique dans le but d'induire une coupure rapide et efficace dans des conditions extrêmement douces et sélectives.

Dans un premier temps, nous avons exploré l'utilisation d'un *linker 3* développé par le Prof. Sébastien Papot dans le cadre d'applications en ciblage thérapeutique. Les conditions d'incorporation de ce *linker* sur un peptide modèle ont été optimisées ainsi que les conditions

de la déprotection des fonctions acétates du β -galactoside et enfin le greffage sur un support solide d'un peptide équipé de ce *linker* (fig. 1).

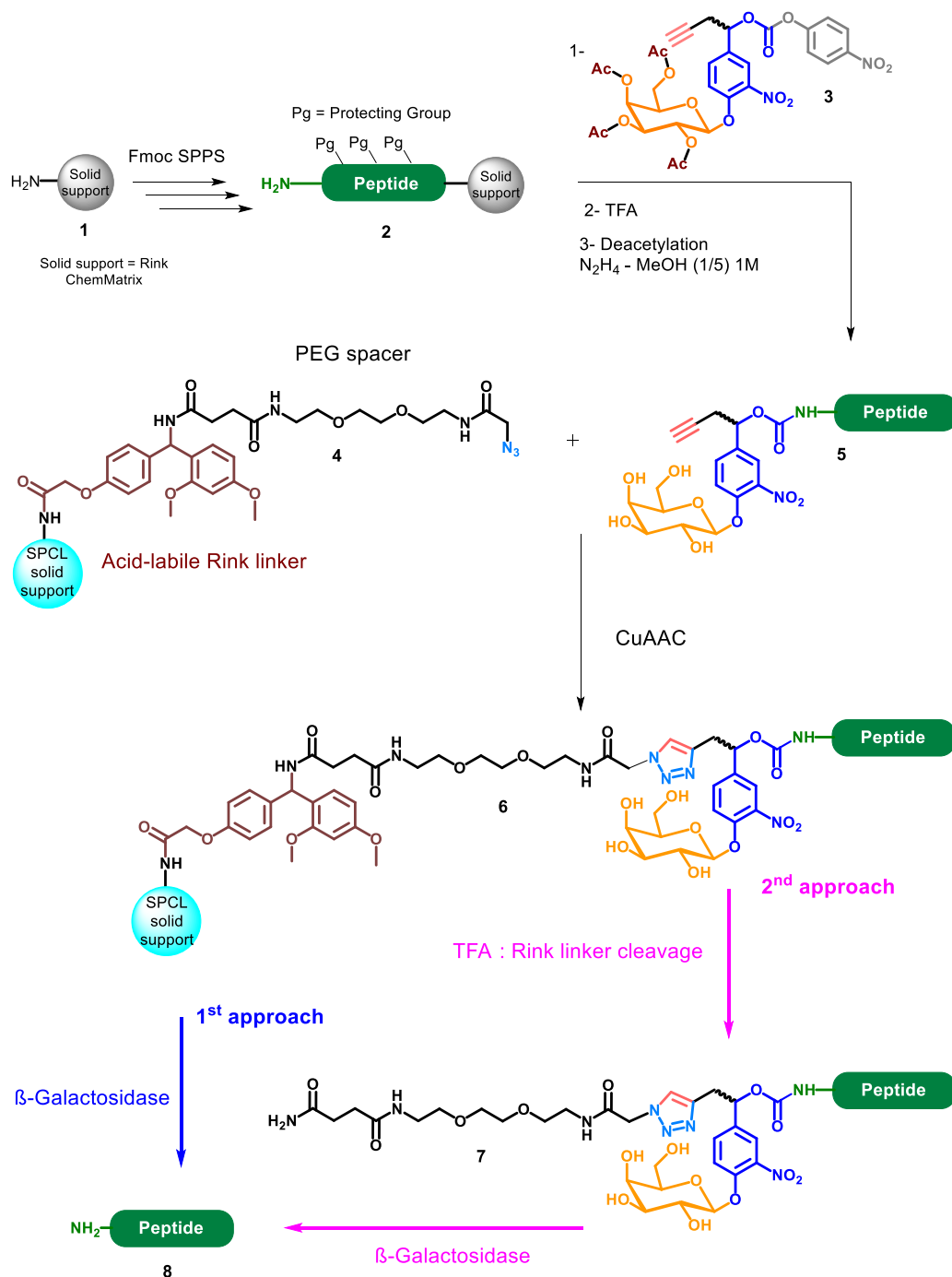


Figure 1

Des essais préliminaires d'hydrolyse enzymatique par la β -galactosidase ont été réalisés en solution (avec le composé 5). Les résultats ont été très concluants, le *linker* est coupé au bout de quelques minutes, libérant le peptide modèle.

Le peptide fonctionnalisé **5** a ensuite été immobilisé par le biais d'une réaction de cycloaddition catalysée par le cuivre (composé **6**). Plusieurs supports solides ont été testés afin d'évaluer leur compatibilité avec la réaction enzymatique. Les cinétiques de coupures sont beaucoup plus lentes qu'en solution et la nature du support solide joue un rôle prépondérant sur la cinétique de la réaction.

Nous avons évalué cette différence de vitesse en solution et sur le support ayant donné les meilleurs résultats (*control pore glass* CPG2000), en mesurant les constantes cinétiques : le rapport $k_{\text{solution}}/k_{\text{support}}$ est d'environ 60 000, soit un ralentissement considérable. Une deuxième approche a été explorée : dans un premier temps la coupure du *linker* de Rink nous permet de détacher le peptide du support solide, ensuite la β -galactosidase va couper le *linker* enzymolabile en solution pour libérer la protéine: cette approche est facile à réaliser, beaucoup plus rapide et ne nécessite qu'une très faible quantité d'enzyme comparé à la première approche. Néanmoins la limite de cette stratégie réside dans les conditions de coupures relativement dures utilisées pour couper le *linker* de Rink (95% TFA). Pour remédier à cela, nous avons envisagé de remplacer le *linker* de Rink par un autre *linker* acidolabile qui serait idéalement coupé dans l'eau et dans des conditions plus douces, tout en étant stable dans les conditions de réaction de ligations chimiques natives (NCL), la réaction qui sera utilisée par la suite pour l'élongation de la protéine sur le support solide. Dans un premier temps, le *linker* trityle a été exploré, néanmoins ce dernier s'est avéré instable dans les conditions de NCL. Ensuite, le *linker* de Sieber a été testé, ce dernier est stable dans les conditions de NCL et est efficacement coupé avec 10% TFA dans l'eau.

Bien qu'efficace, cette méthode requiert l'utilisation de 10% de TFA ce qui est incompatible avec des protéines sensibles. Une deuxième génération de *linker* enzymolabiles a donc été synthétisée afin de pallier aux limites du premier *linker*, qui sont (1) la complexité de sa synthèse, (2) le greffage sur le support solide par une réaction catalysée par le cuivre, (3) les groupements acétates qui nécessitent une étape supplémentaire de déprotection en solution, (4) la présence d'un centre asymétrique qui donne des diastéréoisomères séparés en HPLC après la dérivatisation du peptide et (5) la cinétique très lente de coupure enzymatique sur support solide, que nous avons imputé à la grande taille de la β -galactosidase (environ 20 nm par 5 nm) résultant en une diffusion lente à travers les pores du support (200 nm).

Un *linker* de deuxième génération **10** a été conçu et synthétisé en 4 étapes avec un rendement global de 38% (fig. 2). Il est pensé pour être coupés par des phosphatases, des enzymes de taille bien moins importante que les β -galactosidases. Le groupement phosphate est protégé par des groupements benzyle acidolabiles. Ce *linker* est conçu comme un « *building block* », à la différence du premier *linker* qui était fonctionnalisé par une fonction alcyne, celui-ci comprend une amine protégée par un groupement Fmoc, ce qui permet d'introduire de façon modulaire un groupement pour le greffage sur le support solide par simple couplage. Nous avons opté dans un premier temps pour l'introduction d'une fonction azoture, afin de réaliser le greffage sur le support solide *via* une cycloaddition de type SPAAC (*Strain-Promoted Azide / Alkyne Cycloaddition*) avec un cyclooctyne, ne nécessitant pas de catalyseur étant donné la tension de cycle rendant l'alcyne très réactif dans le cadre de cycloadditions 3+2.

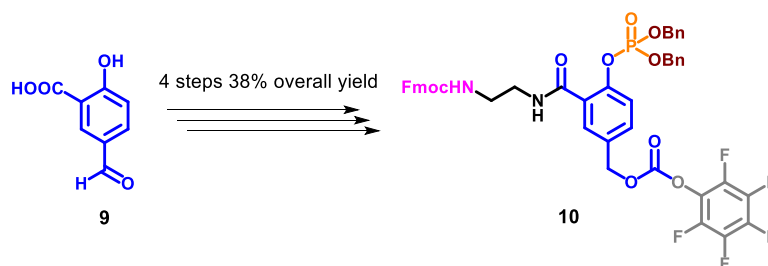


Figure 2

Ensuite nous avons procédé comme avec le *linker* de 1^{ère} génération : le *linker* a été incorporé sur un peptide modèle, puis des essais enzymatiques ont été effectués en solution pour vérifier la viabilité de l'approche. Quatre phosphatases différentes ont été évaluées, la *Shrimp Alkaline Phosphatase*, l'*Antartic Phosphatase*, la *Calf Intestinal Alkaline Phosphatase* et la *Lambda Protein Phosphatase*. Les meilleurs résultats ont été obtenus avec la *Shrimp Alkaline Phosphatase* et la *Lambda Phosphatase*. Ces deux dernières ont donc été sélectionnées pour les essais enzymatiques sur support solide. Pour effectuer ceci, le peptide contenant le *linker* de 2^{ème} génération a été greffé sur le support solide (CPG2000) fonctionnalisé avec un cyclooctyne, par une réaction de SPAAC. Pour évaluer l'efficacité de la coupure enzymatique sur support solide, le rapport $k_{\text{solution}}/k_{\text{support}}$ a été mesuré pour les deux enzymes. La vitesse de coupure relative avec la *shrimp phosphatase* a été multipliée par un facteur 7 ($k_{\text{solution}}/k_{\text{support}} \sim 8\,000$) et celle avec la *lambda phosphatase* par un facteur 300 ($k_{\text{solution}}/k_{\text{support}} \sim 200$), semblant valider notre hypothèse de l'influence de la taille de l'enzyme sur la cinétique de la réaction supportée (fig. 3).

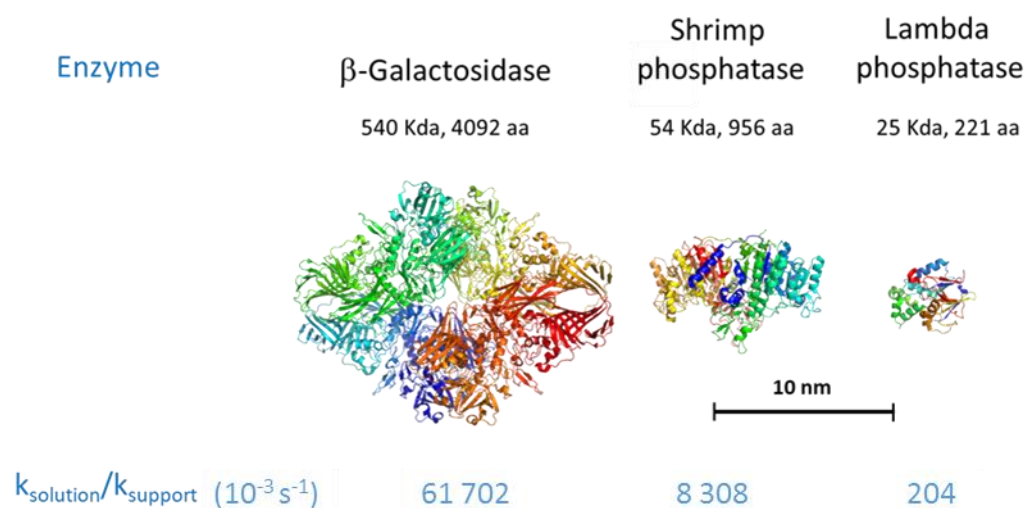


Figure 3

Pour démontrer l'utilité de notre méthode nous avons décidé de l'utiliser pour la synthèse par SPCL d'un polypeptide dérivé de la protéine mucine 1 que nous avons synthétisée par deux ligations successives (fig. 4). Les différents segments peptidiques nécessaires pour réaliser l'élongation du polypeptide ont donc été synthétisés et purifiés. Le premier segment **11** est équipé sur son coté N-terminal par le *linker* de 2^{ème} génération et par un crypto-thioester (voir le chapitre 3 pour plus des détails concernant la synthèse des crypto-thioesters) sur son coté C-terminal. Le segment médian **12** quant à lui était équipé d'un thioester masqué par un groupement AcM qui peut être déprotégé *via* un traitement au Pd(II). Le troisième segment **15** est équipé d'une cystéine N-terminale. Ces segments ont été engagés dans des réactions de ligation chimique pour obtenir le polypeptide de 160 acides aminées **16**. Une dernière étape de coupure enzymatique du *linker* a permis de libérer **17** en solution. Ce dernier a été obtenu avec un rendement total de 13 % (correspondant à 66% par étape) après une purification par HPLC. Ces travaux ont été acceptés pour publication dans **Angewandte Chemie International Edition** : S. A. Abboud, M. Amoura, B. Renoux, S. Papot, V. Piller, V. Aucagne, Enzyme-cleavable linkers for the total synthesis of proteins through solid-supported ligations.

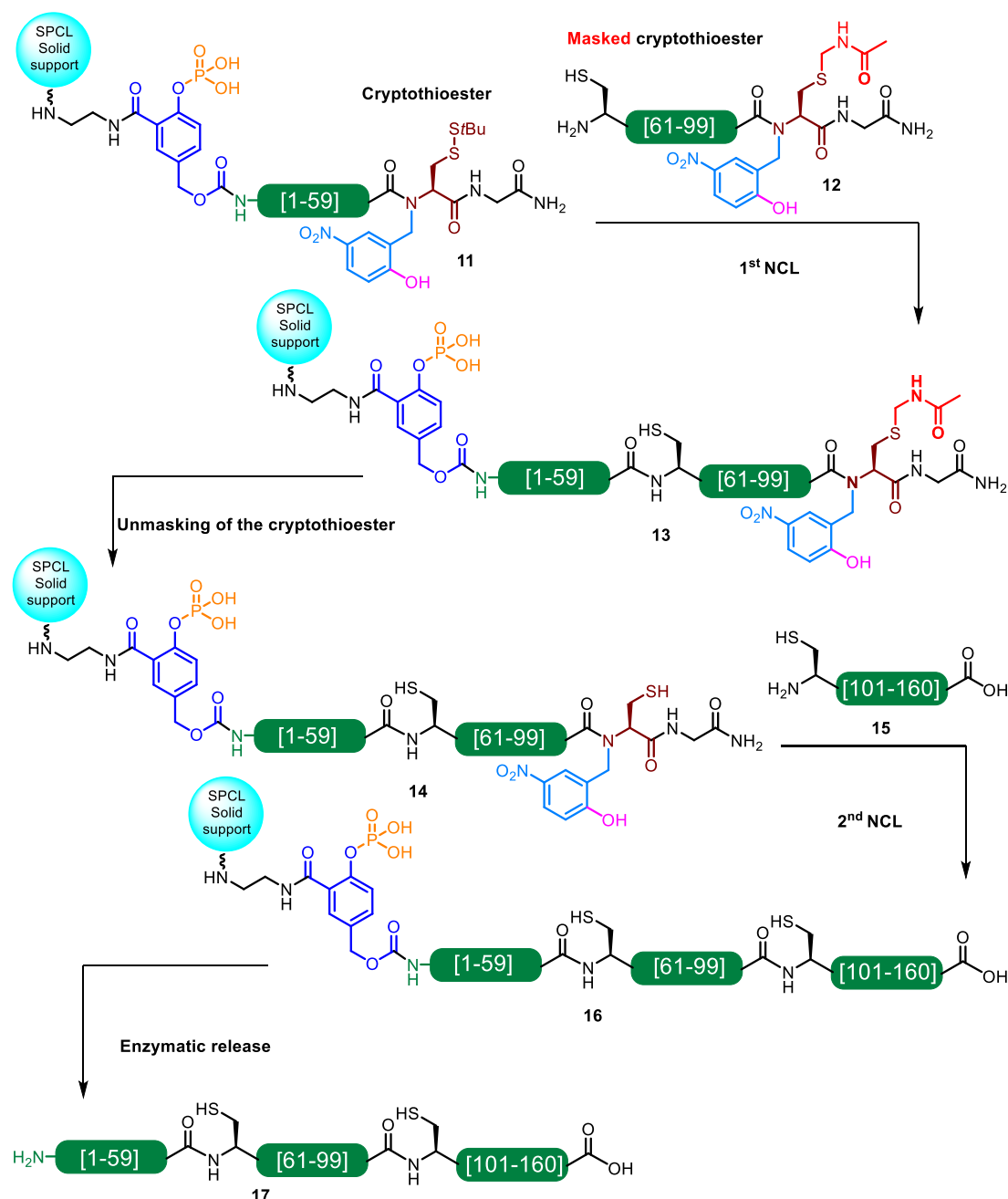


Figure 4

Comme mentionné plus haut, l'élargissement de la protéine sur le support solide a été réalisée *via* des réactions de NCL qui consistent à coupler un peptide muni d'une fonction thioester en position C-terminale avec un cystéinyl-peptide. Notre équipe a conçu un dispositif moléculaire de type *N*-(2-hydroxy-4-nitrobenzyl) cystéine (ou *N*-Hnb-Cys) programmé pour former un thioester *in situ* dans les conditions réactionnelles de la NCL : on parle alors de peptides cryptothioester (composé **22** fig. 5). Lors de l'analyse HPLC de ces dérivés crypto-thioesters, nous avons observé un pic satellite au pic majoritaire possédant la même masse moléculaire que le

produit désiré (composé **21** fig. 5). Une étude systématique de cette réaction secondaire parasite a été réalisée. Dans un premier temps nous avons synthétisé des standards HPLC qui nous ont permis d'attribuer ce pic satellite à l'épimérisation du résidu cystéine. Ensuite nous avons varié systématiquement tous les paramètres de la réaction d'amination réductrice pour comprendre le mécanisme et minimiser cette réaction secondaire. Une méthode permettant de diminuer considérablement la proportion du produit épimérisé a été développée. Par ailleurs, deux autres produits secondaires, issus eux aussi de la réaction d'amination réductrice, ont pu être identifiés et caractérisés. Il s'agit d'imidazolidinones produites suite à une réduction incomplète de l'imine. Cette méthode optimisée a pu être automatisée, et validée par la synthèse d'une dizaine de séquences peptidiques différentes avec de très faibles taux d'épimérisation. C'est donc cette méthode qui a été utilisée pour la synthèse des segments crypto-thioesters utilisées pour la synthèse du polypeptide **17**.

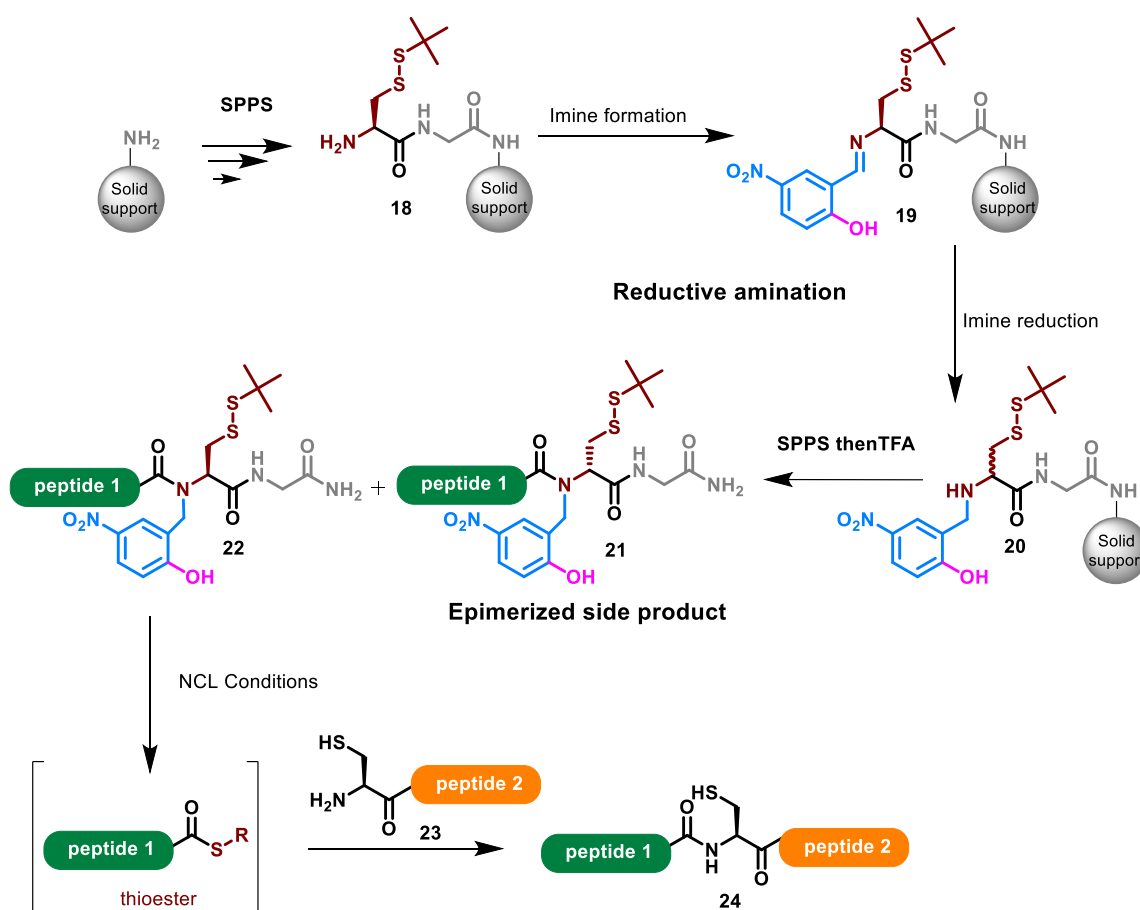


Figure 5

Tout ce travail d'optimisation est développé dans le chapitre 3 de ce manuscrit et a été publié dans *Organic & Biomolecular Chemistry*: S. A. Abboud and V. Aucagne (2020) An optimized

protocol for the synthesis of *N*-2-hydroxybenzyl-cysteine peptide crypto-thioesters. *Org. Biomol. Chem.*, 18, 8199–8208.

Par la suite, nous nous sommes intéressés à une protéine d'intérêt biologique qui est SUMO-2 (*small ubiquitin-like modifier 2*), une protéine de 93 acides aminés. Cette dernière peut se lier à la chaîne latérale de certaines lysines d'autres protéines, par formation d'une liaison isopeptidique avec son extrémité C-terminale : on appelle SUMOylation cette modification post-traductionnelle (PTM). Il est à noter que la protéine SUMO-2 possède une lysine SUMOylable, pouvant conduire à des chaînes oligoSUMO. La SUMOylation est médiée par des SUMO ligases, et cette PTM est réversible : des SUMO protéases sont capables d'hydrolyser la liaison isopeptidique.

Notre idée était d'assembler un peptide SUMOylé **31** par SUMO-2 en alternant deux types de ligations orthogonales : la NCL et la PTL (*peptidomimetic triazole ligation*), de façon à obtenir des cibles qui ne pourront pas être coupées par les SUMO protéases, étant donné la nature non-hydrolysable de la liaison triazole mimant une liaison amide (fig 7).

Les deux segments de la protéine SUMO-2 ont donc été synthétisés et nous avons constaté un problème de faible solubilité du segment SUMO-2 [48-93], ce qui rend extrêmement difficile sa manipulation et sa purification. La faible solubilité et l'agrégation de segments peptidiques utilisés pour l'assemblage de protéines par ligation est un problème très courant, et de nombreuses stratégies ont été développées pour introduire de façon temporaire des groupements (*tags*) solubilisant. Une des stratégies les plus faciles à mettre en œuvre, est l'introduction d'un petit peptide constitué de plusieurs lysines ou arginines (chargées positivement et très hydrophiles) sur la chaîne latérale d'une des lysines du segment, *via* un *spacer* qui est coupé après la ligation. Nous avons développé une nouvelle stratégie, en introduisant le *tag* solubilisant *via* un pont disulfure avec la cystéine N-terminale pour rendre soluble ce segment (fig. 6).

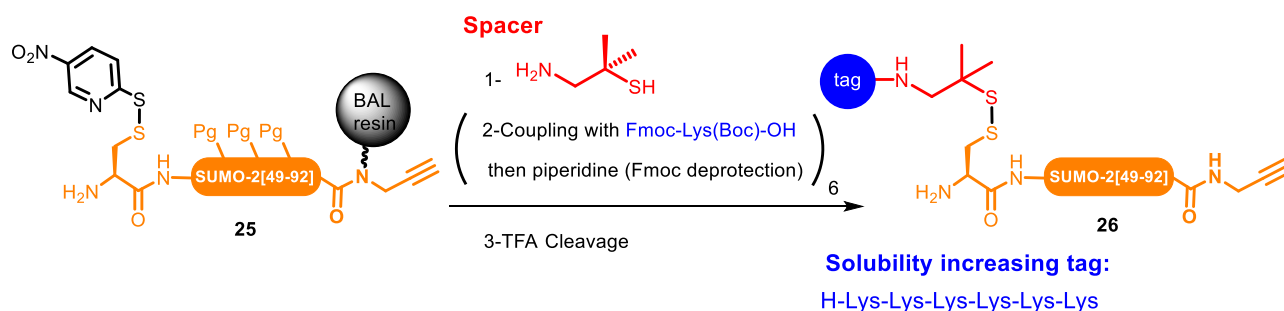


Figure 6

Ce *tag* a aussi été utilisé pour la synthèse de la protéine SUMO-2 **29** *via* une ligation native en solution (fig. 7). Ces travaux ont été publiés dans **Chemical Science**: S. A. Abboud, E.H. Cisse, M. Doudeau, H. Bénédicti, V. Aucagne. A straightforward methodology to overcome solubility challenges for N-terminal cysteinyl peptide segments used for native chemical ligation. **Chem. Sci.**, 2021, 12, 3194-3201.

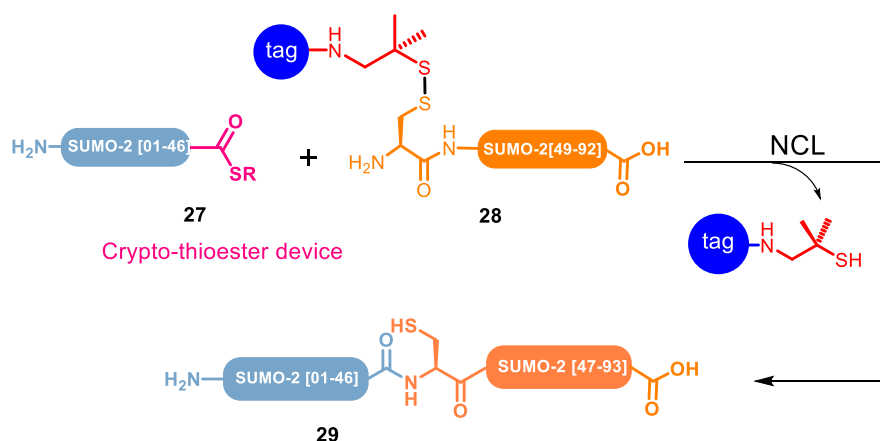


Figure 7

Une fois le problème de solubilité résolu, le peptide SUMOylé **34** a été synthétisé par deux ligations orthogonales sur support solide comme indiqué plus haut. L'avantage de cette méthode est qu'elle ne nécessite pas des étapes de protections et de déprotections intermédiaires étant donné que lors de la première NCL entre le segment **30** (équipé sur son côté N-terminal par le *linker* de 2^{ème} génération et par un crypto-thioester sur son côté C-terminal) et le segment **31** (équipé du *tag* solubilisant sur la chaîne latérale de la cystéine N-terminale ainsi que d'un groupement alcyne sur son côté C-terminal), le groupement alcyne ne va pas réagir et peut donc rester déprotégé durant cette ligation. Ceci va permettre d'engager directement le produit obtenu **32** dans la ligation suivante avec le peptide **33** qui

contient sur sa chaîne latérale un groupement réactif azoture permettant la réalisation de la CuAAC (fig. 8) et obtenir le peptide SUMOylé **34**.

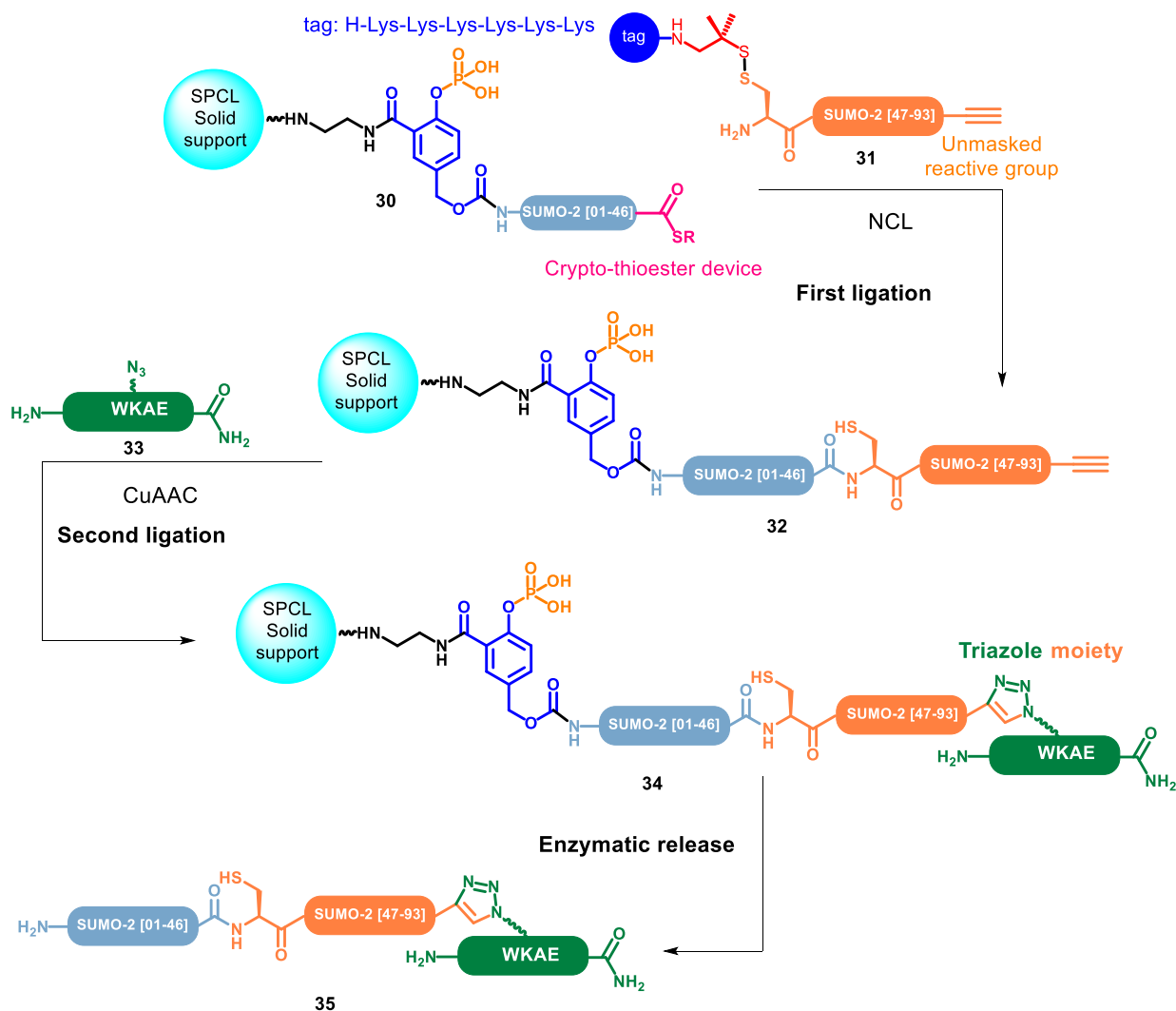


Figure 8

Ce travail de thèse a donc permis de proposer des solutions innovantes pour plusieurs verrous technologiques majeurs liés à la synthèse de protéines par ligation sur support solide. Nous avons ainsi constitué une boîte à outils moléculaires polyvalente potentiellement applicable à de nombreuses cibles.

General Introduction

The Investigation of natural proteins and the synthesis of unnatural ones require the ability to easily access and manipulate proteins. Nevertheless, accessing sufficient quantities from biological sources for a thorough study is not always possible. The recombinant production of proteins can provide access to long proteins in large quantities. However, this approach is limited by the restrictions of the genetic code, which seriously limits the possible structural modifications. Moreover, the introduction of post-translational modifications could be tremendous and impossible to control precisely.

An alternative to protein expression is chemical synthesis, which overcomes the limitations of molecular biology. Indeed, this approach allows studying biological functions of proteins that living systems can hardly produce in addition to enabling a greater level of control on protein composition to afford unusual scaffolds and to label proteins without any limitations. Solid-support synthesis, introduced by Merrifield half a century ago, has been a fundamental shift in the field of protein synthesis. The development of native chemical ligation in 1994 by Kent's group was another major breakthrough. This method has become an often used and powerful method for connecting unprotected peptide segments and is gradually being democratized for the synthesis of small proteins (50-100 amino acids). However, access to more ambitious targets in terms of size remains a real challenge due to repeated purifications steps, ultimately leading to very low overall yields. An early-recognized solution to this problem is the assembly of the protein on a solid support. Although extremely appealing, solid phase chemical ligation (SPCL) has been so far essentially limited to proof-of-concepts, one of the main reason being the difficulty to immobilize a first peptide segment on a water-compatible solid support through a dedicated SPCL linker. A wide range of linkers have been developed but their cleavage requires chemical conditions that proved to be incompatible with some sensitive protein targets

My PhD project is focused on the development of new chemo-enzymatic methodologies for the total synthesis of proteins through SPCL.

Chapter 1 is an exhaustive review of the methods developed in the last 25 years for chemical proteins synthesis through SPCL. A particular emphasis was put on the nature of compatible

solid supports, the structure of the linkers used to link the first segment to the solid support and on the different chemoselective ligation strategies investigated.

This part was published as a chapter of a **Wiley** book: Abboud S. A., Delmas A. F., Aucagne V., Solid-phase Chemical Ligation, in *Total Chemical Synthesis of Proteins*, (Eds.: A. Brik, P. Dawson, L. Liu), **Wiley**, 2021, pp. 259–284.

Chapter 2 focuses on the development of a new generation of linkers by exploring an enzymatic rather than purely chemical approach, in order to induce cleavage under extremely mild and selective conditions. A first β -galactosidase cleavable linker initially designed for drug delivery purpose was explored however, cleavage kinetics on different solid support were extremely slow as compared to solution phase. We hypothesized that the large size of the enzyme (540 kDa) slowed down its diffusion inside the solid support. Hence, a second linker designed to be cleaved by smaller enzymes was synthesized. In the case of lambda-phosphatase (25 kDa), the relative cleavage rate solid/solution phase was increased by almost three logs, validating our hypothesis. The utility of this approach was demonstrated by the native chemical ligation (NCL)-based synthesis of a 160 residues polypeptide through two successive ligations. This chapter is essentially based on a manuscript that is published in **Angewandte Chemie International**: Abboud S.A., Amoura M., Renoux. B., Papot S., Piller V., Aucagne V. (2021) Enzyme-cleavable linkers for protein chemical synthesis through solid-phase ligations. *Angew. Chem. Int. Ed.* 2021. It is also completed by a bibliographic introduction on enzyme cleavable linkers.

Chapter 3 is devoted to the amelioration of a method previously reported by our group for the synthesis of *N*-2-hydroxy-4-nitrobenzyl-cysteine peptide “crypto-thioesters” (*N*-Hnb-Cys), advantageously used as segments for the NCL-based assembly of proteins. A thorough study of the synthesis key step, a reductive amination reaction used to introduce the *N*-Hnb group, allowed us to identify several side products (in particular due to Cys epimerization) and to understand their formation mechanism, which ultimately led to the development of an improved synthetic protocol that can be easily automated on a peptide synthesizer. This chapter is essentially based on a manuscript that was published in **Organic & Biomolecular Chemistry**: Abboud, S.A., and Aucagne, V. (2020) An optimized protocol for the synthesis of *N*-2-hydroxybenzyl-cysteine peptide crypto-thioesters. *Org. Biomol. Chem.*, 18, 8199–8208,

completed by a bibliographic introduction on the synthesis of peptide thioesters, with a special focus on *N*-S-acyl shift-based peptide crypto thioesters.

Chapter 4 focuses on the development of a straightforward methodology to overcome solubility challenges for N-terminal cysteinyl peptide segments used for NCL assembly of proteins that can be easily automated on a peptide synthesizer. The broad utility of this method was exemplified through the NCL-based synthesis of the well-known difficult target SUMO-2. [This chapter is essentially based](#) on a manuscript that was published in **Chemical Science**: Abboud S. A., Cisse E. Dodeau M. Bénédicti H. and Aucagne V., *Chem. Sci.*, 2021, 12, 3194-3201, completed by an exhaustive bibliographic introduction on temporary solubilizing tags used for chemical protein synthesis.

Chapter 5 is dedicated to the exploration of innovative chemical tools to overcome the major difficulty in the study of SUMOylated proteins: the very high sensitivity of isopeptide bond linking SUMO and the protein to specific proteases called SUMO proteases, ubiquitous in cells. We focused on the synthesis of a SUMOylated peptide, where the isopeptide bonds was replaced by a non-hydrolyzable 1,4-disubstituted 1,2,3-triazole, recognized as an excellent mimics of amide bonds. An original synthetic route was used, since it combines two types of orthogonal chemical ligation, NCL and peptidomimetic triazole ligation (PTL) developed in our team based on CuI-catalyzed azide/alkyne cycloaddition (CuAAC). In this way, there is no need for masking/unmasking of chemoselectively reactive groups to perform successive ligations, therefore limiting the number of steps and intermediate purifications, and this strategy is potentially amenable to the synthesis of biologically-relevant oligo-SUMOylated peptides and proteins through multiple successive solid-supported NCL/PTL cycles.

Chapter 1: Solid Phase Chemical Ligation

Skander A. Abboud, Agnès F. Delmas and Vincent Aucagne*

Centre de Biophysique Moléculaire, CNRS UPR 4301, Rue Charles Sadron 45071
Orléans, France.

DOI : 10.1002/9783527823567.ch9

* : corresponding author

Authors emails : skander.abboud@cnrs-orleans.fr ; delmas@cnrs-orleans.fr ;
aucagne@cnrs-orleans.fr

Abstract

This is an exhaustive review of the methods developed in the last 25 years for chemical proteins synthesis through solid-supported ligation-based assembly of peptide segments. A particular emphasis was put on the nature of solid supports, the structure of the linkers used to link the first segment to the solid support, the different chemoselective ligation strategies investigated, and on post-ligation solid-phase transformations.

Keywords

SPCL; Solid phase; Solid support; Linker; Native Chemical ligation; Peptidomimetic triazole ligation; Oxime ligation

Acknowledgments:

We thank the French Agenc.e Nationale de la Recherche for funding (ANR-15-CE07-022 EasyMiniProt grant).

Table of contents

I.	Introduction	30
1)	The promises of solid phase peptide chemical ligation (SPCL)	30
2)	Chemical ligation reactions used for SPCL	31
3)	Key requirements for a SPCL strategy.....	32
II.	SPCL in the C-to-N direction	34
1)	Temporary masking groups to enable iterative ligations	34
2)	Linkers for C-to-N SPCL.....	37
	Use of same linker and solid support for SPPS and SPCL.....	38
	Re-immobilization of the C-terminal segment.....	39
III.	SPCL in the N-to-C direction	42
1)	Temporary masking groups to enable iterative ligations	43
2)	Linkers for N-to-C SPCL.....	44
3)	Case study	47
4)	SPCL with concomitant purifications	49
IV.	Post-ligations solid-supported transformations.....	49
1)	Chemical transformations	49
2)	Biochemical transformations	50
V.	Solid support	51
VI.	References.....	56

I. Introduction

1) The promises of solid phase peptide chemical ligation (SPCL)

The concept of solid-supported synthesis introduced by Merrifield half a century ago [1] has been a paradigm shift in the field of peptide research. The development of ligation reactions [2] was another major breakthrough: building blocks are no longer protected amino acids but unprotected peptide segments equipped with functional groups able to react regio- and chemoselectively in an aqueous environment.

As largely discussed throughout this book, the intensive optimization of the chemical toolbox of the peptide/protein chemist made considerably easier the chemical synthesis of proteins, and has led to tremendous achievements in terms of target size (several hundred amino acid residues). Notable achievements in the field include the synthesis of the enzyme DapA (31 kDa, 312 residues) and its D-version through the convergent ligation-based assembly of six segments [3] or, more recently, the synthesis of tetra-ubiquitinated α -globulin (53 kDa, 472 residues) from six segments [4,5]. A key point for the ligation of such impressive syntheses is the requirement for one of the chemoselectively-reactive moieties of the internal segments to be temporarily masked. This calls for a careful optimization of each unmasking reactions in addition to the strictly speaking chemical ligations steps, as well as for the use of peptide segments as long as possible to minimize the overall number of steps.

As a general statement, multiple ligation-based protein synthesis is often associated with very low overall yields, essentially due to the laborious intermediate chromatographic purifications. Having in mind the idea to minimize such purifications, native chemical ligation (NCL) [6] was extended soon after its discovery to a solid-phase version by its inventors [7-9], aiming to simplify the multiple successive ligation process. Later on, despite its appealing advantages, solid phase peptide ligation (SPCL) was barely considered [10] until we and others [11-15] started to reinvestigate the field. In this review, we exhaustively describe the efforts done in the SPCL field in the past twenty years. The low-hanging fruits have already been harvested and published as proofs of concept, but many further efforts and methodology developments will probably be needed to transform SPCL into a mature technology and fulfill its promises. In the following sections, we will try to stress the main challenges associated with SPCL and to identify the remaining bottlenecks.

Note that the term “solid phase chemical ligation” was introduced by Muir in 1998 [16]. However it referred in this case to the use of a solid supported thioester that is released into solution during the ligation, rather than to a ligation-based assembly on a solid support. Later on, SPCL referred only to ligation-based assembly on a solid support, on which this review is concentrated.

2) Chemical ligation reactions used for SPCL

The great potential of SPCL was initially applied to NCL [6], and not to other amide bond-forming ligation reactions so far. It was later extended to peptidomimetic triazole ligation (PTL) [17], which exploits the amide mimicking features of the 1,4-disubstituted 1,2,3-triazole products generated by of the copper-catalyzed alkyne/azide cycloaddition (CuAAC) [18,19]. Moreover, besides application to chemical protein synthesis, SPCL has also been applied to oxime ligation for the modular synthesis of complex peptide-based architectures [13] (figure 1).

These three types of chemical ligation reactions typically take place in aqueous media at room temperature or slightly above (37°C), and their kinetics rates in solution [20,21] allows to reach completion within a few hours at millimolar concentrations using a slight excess of one of the peptide segments. Note however that the solid support can have a dramatic influence on kinetics, as discussed later in part V.

From a more practical point of view, reaction conditions for solid-supported ligations do not really differ from solution phase ones. Typical conditions for NCL includes a pH 6.5-7.5 phosphate buffer and 6 M guanidinium chloride (to increase solubility of segments), in conjunction with TCEP to keep a reducing environment, and an arylthiol catalyst such as 4-mercaptophenylacetic acid (MPAA). CuAAC proceeds well at neutral pH and various co-solvents, such as NMP, HFIP or methanol can be used in case of low solubility of segments. It requires a copper(I) catalyst which can lead to Fenton-type oxidation in the presence of oxygen, which is particularly detrimental for methionine and histidine side chains. Therefore, the reaction mixtures should be carefully deoxygenated, and use of copper(I) ligands such as *tris*(3-hydroxypropyltriazolylmethyl)amine (THPTA) in conjunction with radical scavengers [22] is sometimes necessary. Acidic sodium acetate buffer (pH 4.5) is well suited for oxime

ligations, which can proceed without any additives. However, aniline has been shown to considerably accelerate oxime formation kinetics in solution [20].

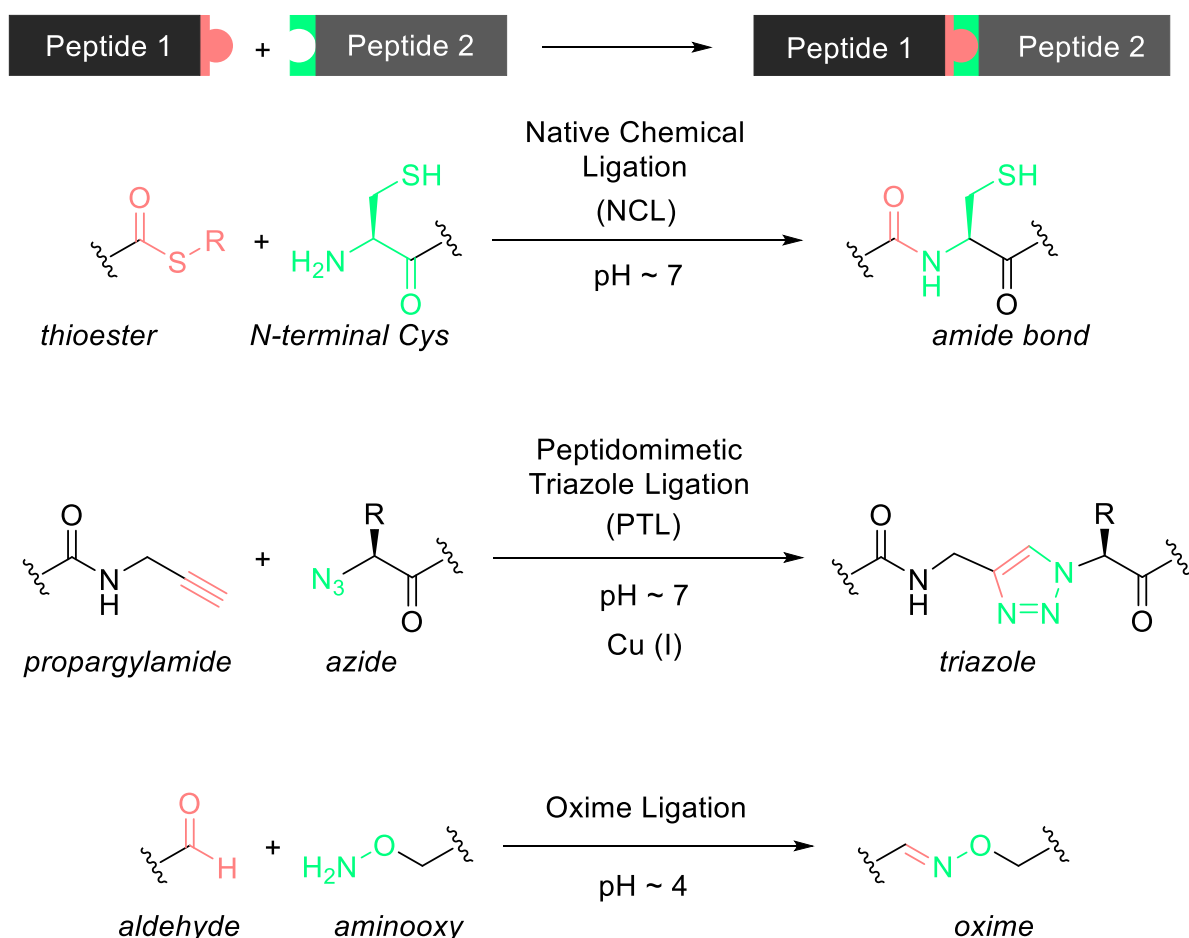


Figure 1 – Chemoselective ligation reactions that have been used for SPCL

3) Key requirements for a SPCL strategy

If SPPS can be conducted only in one way, from the C-terminus to the N-terminus, for reasons of epimerization during amide formation, solid-supported ligations of unprotected peptide segments can be conducted either in the N-to-C or C-to-N directions (figure 2) as illustrated in the seminal publication of Kent and coworkers introducing the SPCL concept in 1999 [8].

In this review, we will illustrate pros and cons of both methods, which share a number of common requirements:

- the use of masking groups for one of the chemoselective reactive moieties to enable successive ligation, by preventing segment cyclization and oligomerization, such as for solution phase multiple segments assembly;
- a strategy to immobilize the starting material, a pure and unprotected peptide segment;
- a SPCL linker cleavable under conditions compatible with a large spectrum of proteins, while being stable to ligation and unmasking conditions;
- a SPCL solid support compatible with the aqueous environment used for ligations, with the unmasking reactions and the SPCL linker cleavage.

In addition, we will highlight post-ligation solid phase synthetic transformations, taking further advantage of the solid support. A table recapitulating all the publications related to SPCL can be found at the end of this manuscript.

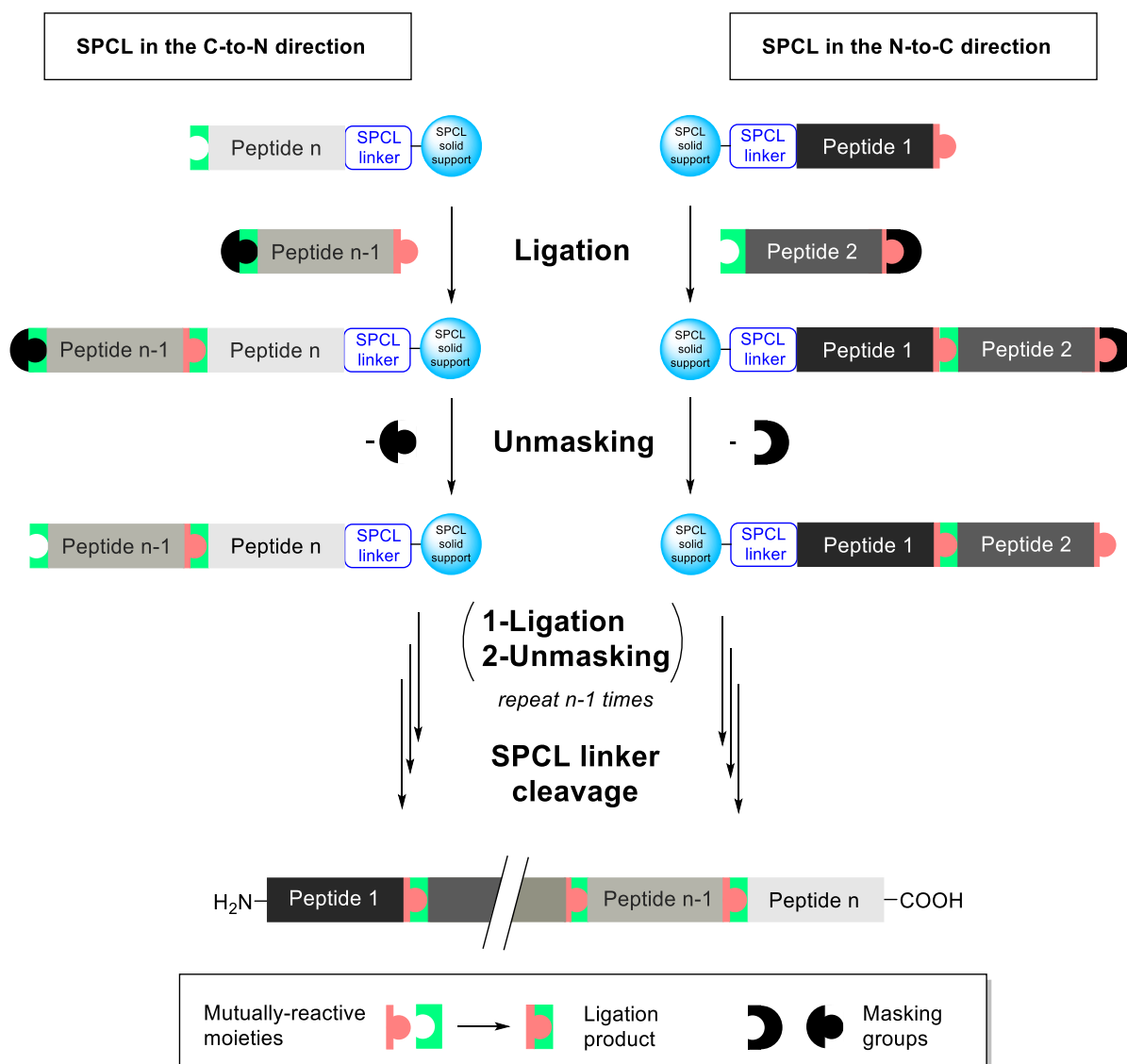


Figure 2 – General principles of solid phase chemical ligation

II. SPCL in the C-to-N direction

1) Temporary masking groups to enable iterative ligations

As discussed earlier, the key tactic for multiple successive ligations is the temporary masking of one of the reactive ends of internal peptide segments. Importantly for the SPCL strategies, the unmasking conditions have to be compatible with the stability of the SPCL linker.

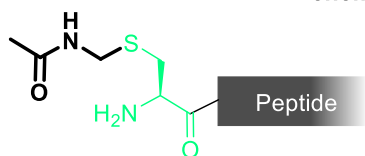
The masking groups that have been used for solid phase ligations in the C-to-N direction are listed in Figure 3. Most reports of C-to-N SPCL are based on NCL reactions between N-terminal cysteine and C-terminal thioesters segments: as a consequence, a Cys-masking strategy is necessary. Three different Cys protecting groups have been exploited for this purpose, all

three being commonly used in solution-phase multi-segment ligation strategies. The methylsulfonylethyloxycarbonyl (Msc) [23] amine protecting group can be removed under aqueous basic conditions [9]. The nowadays most used strategy is *N,S* bis-protection as a thiazolidine ring (Thz) that can be opened using aqueous methoxyamine [10,15, 24]. The acetamidomethyl group (Acm) [25] thiol protecting group used by Kent in the first SPCL report can be removed on resin by treatment with mercury(II) acetate [4]. Despite its wide applicability, one of the drawbacks of the Acm protecting group is its removal requiring extended reaction times and harsh conditions. Recently, Brik reported the use of Pd(II) salts to remove this protecting group within minutes in aqueous media [26]. This method is very efficient for Acm deprotection in solution and looks promising for SPCL, although this has not been demonstrated yet.

As an alternative to protecting groups, Muir employed an N-terminal appendage of a 17-mer propeptide sequence (MASSRVDGGRSDLIEGR) after the cysteine residue, this sequence being selectively removed by the protease factor Xa [7].

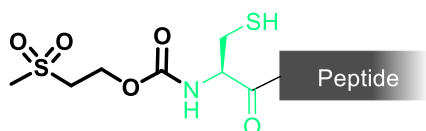
Besides NCL, Aucagne and Delmas reported multiple solid supported oxime ligation [13] using segments bearing an N-terminal aminoxy function protected by an *N*-allyloxycarbonyl (Aloc) group [27]. Aloc can be removed by a Tsuji-Trost reaction using a Pd(0) complex [28] in combination with *N,N*-dimethyl barbituric acid (NDMBA). Note that NDMBA was preferred to classical hydride donors as they caused partial cleavage of the *N-O* bond.

N-terminal cysteine masking groups to enable multiple successive solid-supported N-to-C native chemical ligations



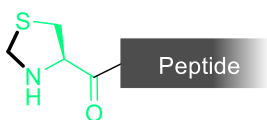
Deprotection : Hg(II) salts

S-Acetamidomethyl group (Acm)



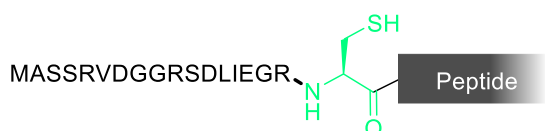
Deprotection : pH = 13

N-Methylsulfonylthioethoxycarbonyl group (Msc)



Deprotection : MeONH₂

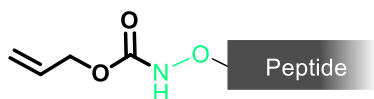
Thiazolidine (Thz)



Propeptide sequence cleavage : factor Xa

Propeptide sequence

Aminoxy protecting group to enable multiple successive solid-supported N-to-C oxime ligations



Deprotection : Pd(0) / NuH

N-Allyloxycarbonyl group (Alloc)

Figure 3 – Temporary masking groups used for SPCL in the C-to-N direction

2) Linkers for C-to-N SPCL

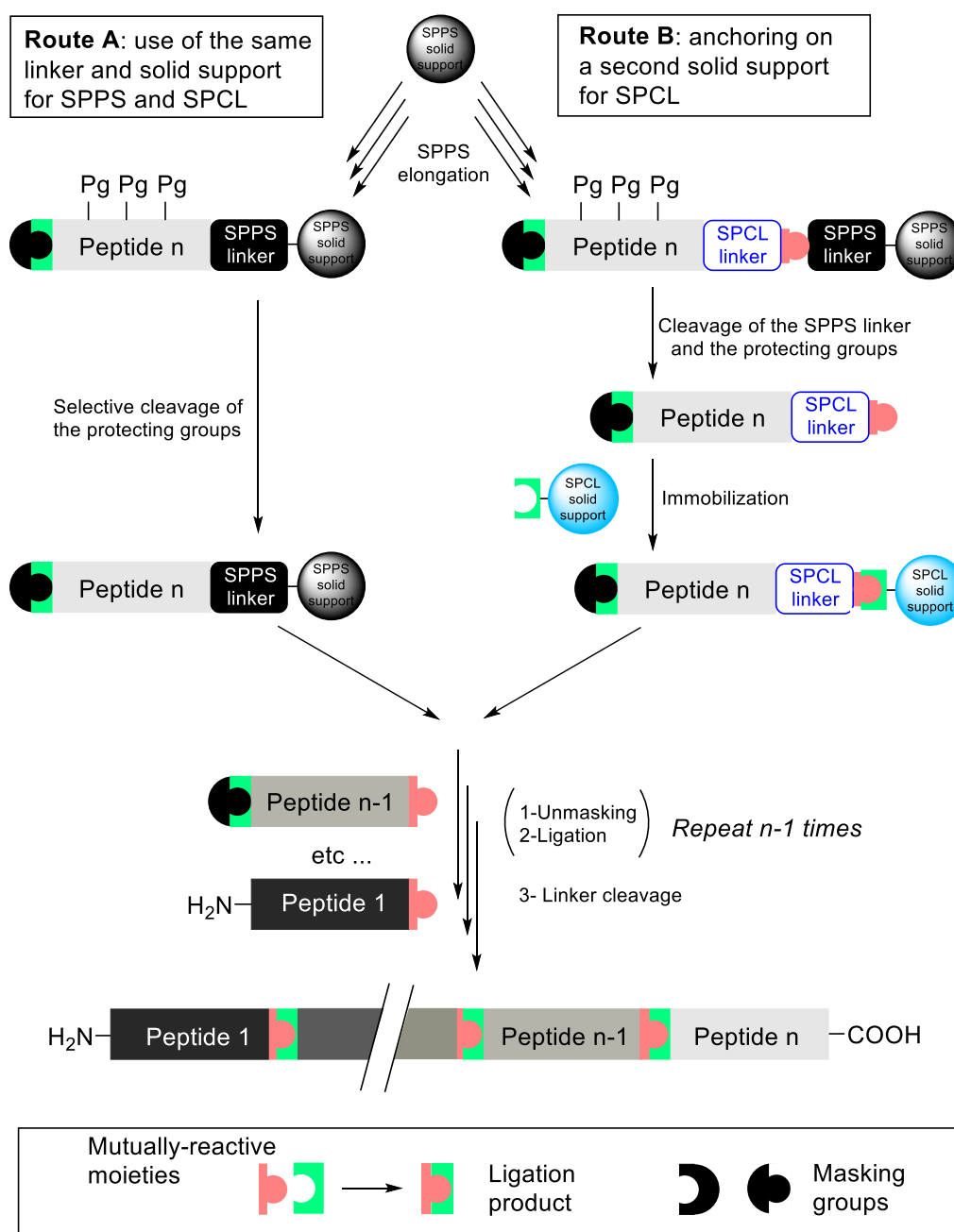


Figure 4 – Two possible routes for SPCL in the C-to-N direction

Two different strategies have been explored for SPCL in the C-to-N direction, calling for different SPCL linkers design. The same linker and solid support employed for the SPPS elongation can be also used for SPCL after selective side chains deprotection (Figure 4, route A), or the C-terminal segment can be re-immobilized on a second solid support through a chemoselective reaction, after its cleavage from the SPPS support (route B).

a) Use of same linker and solid support for SPPS and SPCL

In this case, the C-terminal segment is synthesized on a solid support which is compatible to both SPPS and SPCL conditions: organic solvents and aqueous buffers. The other crucial point is the use of side-chain protecting groups that can be cleaved without affecting the SPPS linker, in order to obtain a fully deprotected peptide still attached to the solid support. To date, all C-to-N SPCL strategies exploiting the same support as SPPS have been developed on Fmoc SPPS. They differ in the nature of the linker, side-chain protecting groups and ligation type.

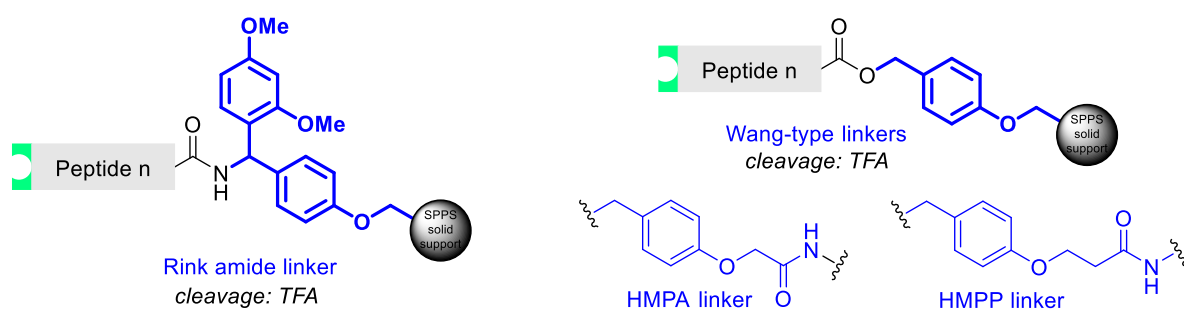


Figure 5 – Linkers exploited for SPCL in the C-to-N direction using the same support for both SPPS and SPCL

This strategy was introduced by Kent [10] for the synthesis of the 28 amino acid knottin EETI-II [29] on the superpermeable organic combinatorial chemistry (SPOCC) resin [30,31] equipped with a TFA-labile Wang-type linker. The first segment was here limited to a Cys-Gly dipeptide, the cysteine being protected as a Thz. Two successive native chemical ligations gave the expected cysteine-rich peptide with a 80 to 90 % yield based on the initial loading of the resin.

Brik [15] extended the approach by the synthesis of a 135 amino acid HA-tagged human histone 2B (H2B) derivative, using a Rink amide-PEGA solid support, through four successive NCL reactions using Thz as the N-terminal Cys masking group. In this case, the first segment was a single cysteine residue, protected as a Thz. On-resin desulfurization of Cys residues into Ala, followed by Rink linker cleavage, gave the target polypeptide (10% isolated yield), which slightly differs from the native sequence by the appendage of a C-terminal alaninamide residue.

Besides applications to chemical proteins synthesis, it is worth mentioning that Cys-Rink amide PEGA resin was recently exploited by Olsen [32] for the chemoselective trapping and enrichment of cyclic peptide thioesters from a complex mixture of natural products secreted by bacteria. Release through TFA treatment allowed the identification of previously unknown auto-inducing peptides that regulate population-wide behavior by acting on quorum sensing.

With the objective to increase the amino acid diversity in the C-terminal segments, Aucagne and Delmas [13] explored the use of highly acid-labile trityl-type side chain protecting groups. The key point was to use acidic deprotection conditions that would not cleave the TFA-labile linker. The nonapeptide PPAHGVSTA derived from the human mucin protein MUC1 was synthesized on a PEGA resin equipped with a Wang-type linker, with the appendage of a N-terminal aminoxyacetic (Aoa) residue. Trityl (Trt) groups were used to protect Aoa, Ser and Thr, and methoxytrityl (Mmt) for His. Mild treatment using dilute TFA in dichloromethane in the presence of $i\text{Pr}_3\text{SiH}$ was used to cleave the protecting groups without affecting the linker. Optimization of the linker structure revealed that the 4-(hydroxymethyl)phenoxypropionic linker (HMPP) was cleaved 3.5 time slower than 4-(hydroxymethyl)phenoxyacetic (HMPA) under these conditions. Notably, Mmt was preferred over Trt for His protection, as the later required prolonged 2% TFA treatment that is detrimental to the linker integrity. Two successive oxime ligations using Aloc-protected Aoa as a masked aminooxy group gave a 61 amino acid polypeptide construct in 52% yield without the need of any further purification after TFA cleavage of the SPPS/SPCL resin. Note that, this strategy could not be generalized to peptide segments containing Arg residues, due to the lack of compatible highly-acid labile protecting groups for the side chain guanidine.

b) Re-immobilization of the C-terminal segment

The single support C-to-N SPCL strategy is inherently limited by the necessity to obtain a C-terminal segment devoid of any co-products arising from side-reactions or incomplete couplings (truncations or deletions). A more versatile, though more sophisticated, strategy implies the synthesis of the C-terminal segment on a SPPS resin, followed by re-immobilization on a second, water-compatible SPCL resin. In this case, a SPCL linker is introduced during SPPS and is functionalized by a group allowing its chemoselective grafting on the SPCL support (red

moiety, Figure 6). The SPCL linker obviously has to be stable to SPPS elongation and cleavage conditions, in addition to the ligation and unmasking steps used for SPCL.

This double support strategy is compatible with the twenty proteinogenic amino acids, and allows the intermediate purification of the C-terminal segment prior to re-immobilization.

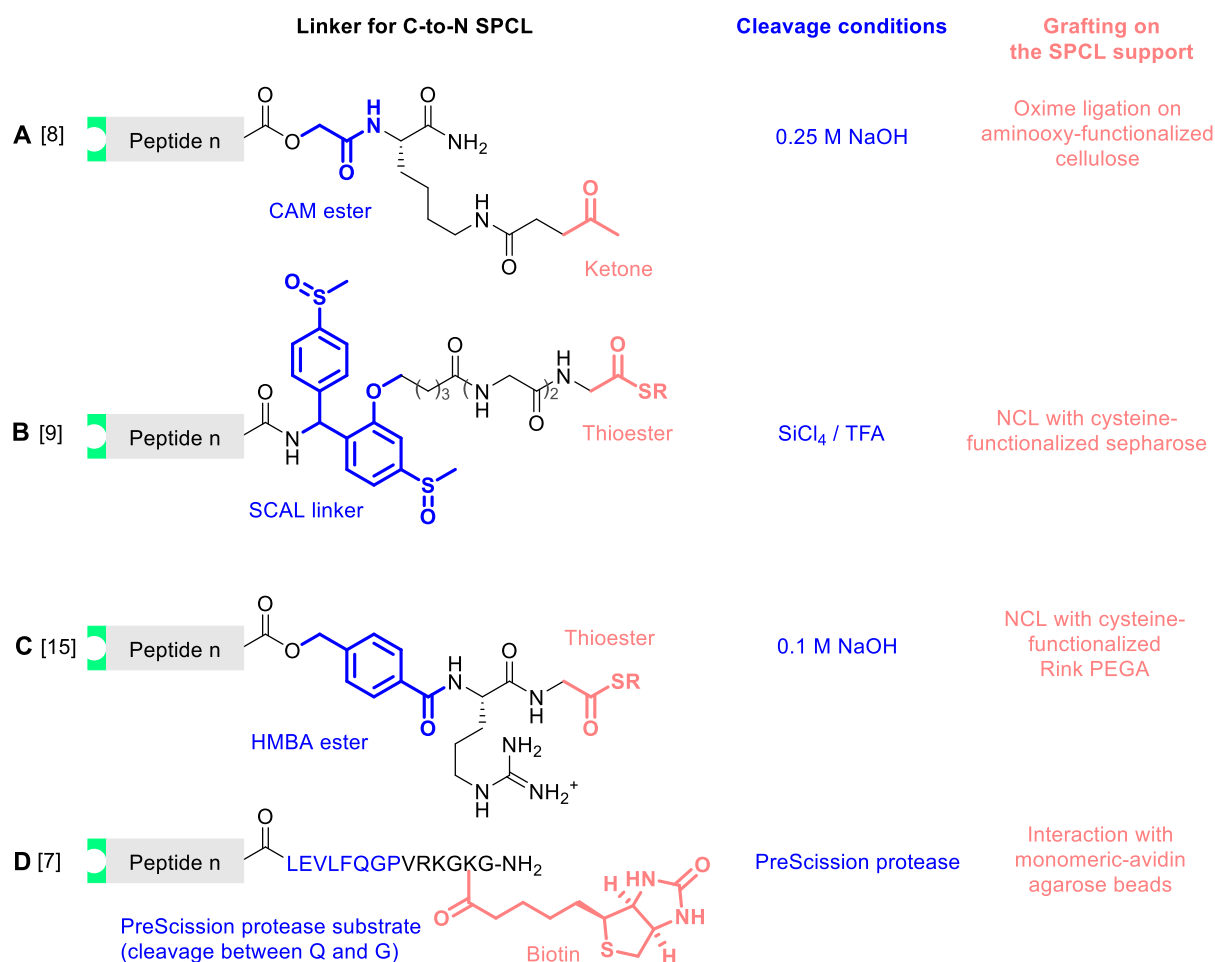


Figure 6 – Different cleavable linkers used to immobilize the C-terminal segment on a SPCL solid support

In the first-ever SPCL report, Kent [8] proposed the use of a base-labile linker derived from the carboxamidomethyl ester (CAM) [-OCH₂CONH₂] protecting group for carboxylic acids (Figure 6A) [33]. Esterification was carried out in solution to provide building blocks for Boc-SPPS. After elongation of the peptide segment, a ketone functional group was introduced through solid-supported coupling of levulinic acid on the side chain of a Fmoc-protected Lys. After cleavage from the SPPS resin, the segment was immobilized by oxime ligation onto an aminoxy-functionalized cellulose solid support. The 118 amino acids human group V

secretory phospholipase A2 (GV-PLA2) was synthesized through three successive NCL, using AcM as Cys-masking groups. Cleavage from the SPCL resin was achieved through a short treatment with aqueous sodium hydroxide and the crude product was then folded to give a six disulfide bonds-containing protein with full enzymatic activity.

Dawson [9] later extended the double support C-to-N SPCL approach to C-terminally-amidated peptides, exploiting the safety catch acid-labile linker (SCAL). [34] This linker is stable to Boc-SPPS elongation and cleavage conditions in the form of *bis*-sulfoxide, and TFA-labile after reduction into sulfides. A thioester moiety located on the SCAL linker (figure 6B) mediated NCL immobilization of the segment onto a Cys-functionalized sepharose solid support. The 71 amino acids vMIP I chemokine was assembled through two consecutive NCL, the use of either S-Acm or N-Msc as Cys-masking group demonstrating comparable efficiencies. Quantitative cleavage of the SCAL linker was achieved through a treatment with 1 M SiCl₄ in TFA in the presence of scavengers. The crude polypeptide product was purified in a 21% recovered yield.

More recently, Ottesen [24] extended the double support C-to-N SPCL approach to Fmoc-SPPS segment synthesis, making use of the base-labile 4-hydroxymethylbenzoic acid (HMBA) linker [35], functionalized with a thioester moiety (figure 6C). In order to allow simple monitoring of the different ligation and Cys unmasking steps, a Rink amide linker was incorporated between the Cys and the PEGA support. Two histone proteins, H4 (102 aa) and CENP-A (139 aa), *N*-acetylated on some of their Lys residues, were synthesized through NCL-based assembly on a Cys-PEGA resin, using Thz-protected segments. In a first approach, both proteins were synthesized through solid-supported NCL using four and five segments, respectively, followed by on resin desulfurization and HMBA cleavage by a short aqueous sodium hydroxide treatment. Even though both products were obtained in a good purity, recovery yields were extremely low. Based on the observation that the recovery severely dropped down after the ligation of the fourth segment, a hybrid phase approach was designed, in which the first three segments of each protein were assembled on resin, and the last ones in solution.

Muir applied C-to-N SPCL to expressed protein ligation (EPL) [36–38], to assemble synthetic C- and N-terminal segments with recombinant ones [7]. C-terminal segment consisted in a 19-mer cysteinyl peptide incorporating the LEVLFQGP sequence, referred to as “PreSciss linker”, and a biotinylated Lys residue (figure 6D). This segment was immobilized on avidin-

functionalized sepharose beads, followed by EPL with a recombinant thioester with an N-terminal MASSRVDGGRSDLIEGR_C appendage. Selective cleavage of the R/C peptide bond by factor Xa generated an N-terminal Cys, which was further ligated with a synthetic thioester. Release of the semisynthetic protein in solution was finally achieved through treatment with PreScission protease that cleaves the Q/G peptide bond of the PreSciss linker. This process was used to dual-label the amino and carboxyl termini of the 304 amino acids Crk-II adapter protein with a FRET pair for application as a protein–protein sensor.

The double support N-to-C SPCL approach was also exploited by Seitz for the NCL-based assembly of proteins on microtiter plates without using any cleavable linker, with the aim to generate small libraries of proteins that can be subsequently screened on-surface [39,40].

III. SPCL in the N-to-C direction

SPCL can also be accomplished in a N-to-C direction. In this approach, the SPCL linker is introduced at the N-terminus of the N-terminal segment after SPPS elongation, followed by linker cleavage and re-immobilization on a SPCL solid support (figure 7) [41].

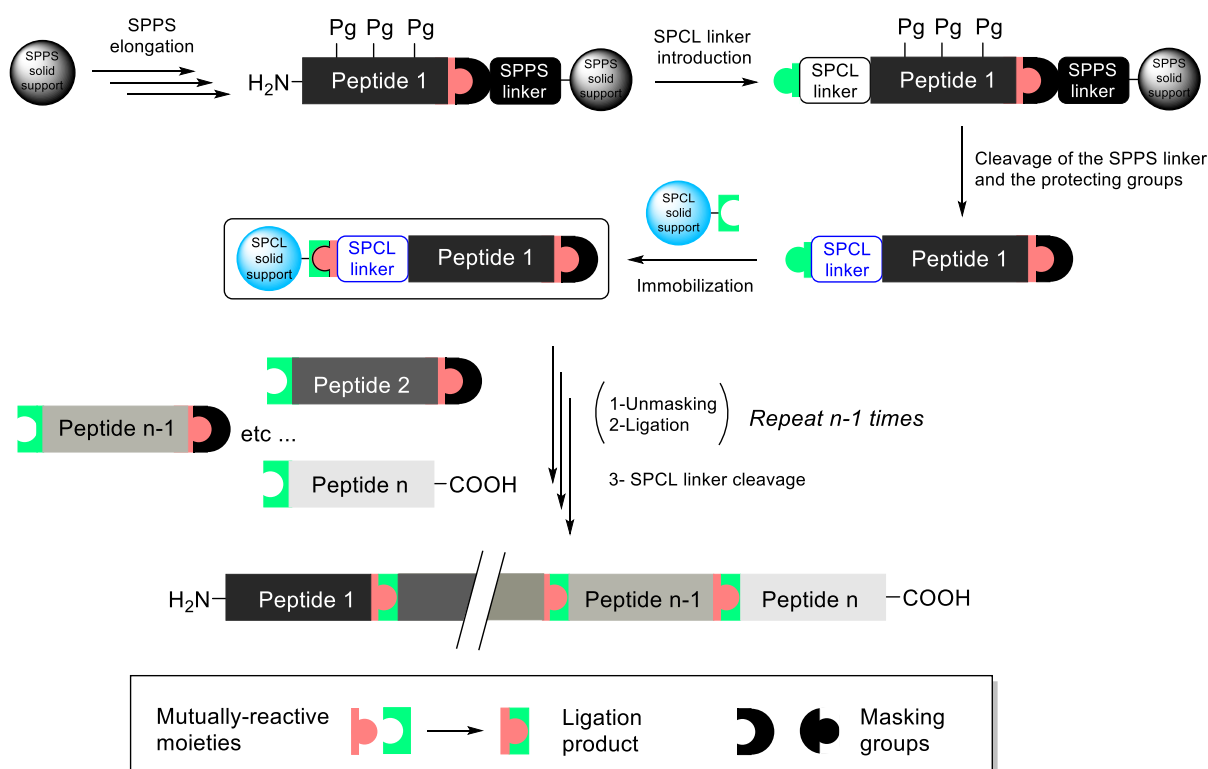


Figure 7 – SPCL in the N-to-C direction

1) Temporary masking groups to enable iterative ligations

A challenge in N-to-C NCL assembly of proteins is the design of masked thioesters, and many groups investigated such thioesters precursors that could be activated on demand. The first report of masked thioesters for N-to-C SPCL was the use of peptide thiocarboxylate segments by Kent [8] : the absence of a proton can be considered as a protecting group for the thioacid, because the ionized thiocarboxylate is almost unreactive under NCL conditions. The thiocarboxylate can be easily converted into a reactive thioester, through chemoselective S-alkylation with bromoacetic acid at pH 4-5 (figure 8A). At this low pH, other potentially nucleophilic functional groups in the unprotected peptide segments are unreactive.

In the last decade, a new strategy for the Fmoc-SPPS-based generation of thioesters has been developed that exploits the spontaneous conversion of α -mercato amides into β -amino thioester through a $N \rightarrow S$ acyl shift. The equilibrium between the two species can be shifted towards a thioester by trans-thioesterification with an exogenous thiol used in large excess. Melnyk and Liu exploited these properties for N-to-C SPCL, through a latent thioester surrogate the *bis*-(2-sulfanylethyl)amido group (SEA) [42,43]. The disulfide form of SEA (SEA^{off}) is unreactive under NCL conditions [44], providing that no strong reducing agent is used, but can be converted on demand into an active thioester by treatment with a combination of TCEP and mercaptoacetic acid as the exogenous thiol, through an intermediate SEA^{on} species (figure 8B).

N-to-C SPCL also applicable to the peptidomimetic triazole ligation strategy developed by Aucagne and Delmas, based on the protection of the alkyne moiety by a tri-isopropylsilyl group (TIPS) [45]. This masking group can be cleanly removed under aqueous conditions through protidesilylation upon treatment with silver nitrate, to generate a terminal alkyne suitable for CuAAC (figure 8C) [14].

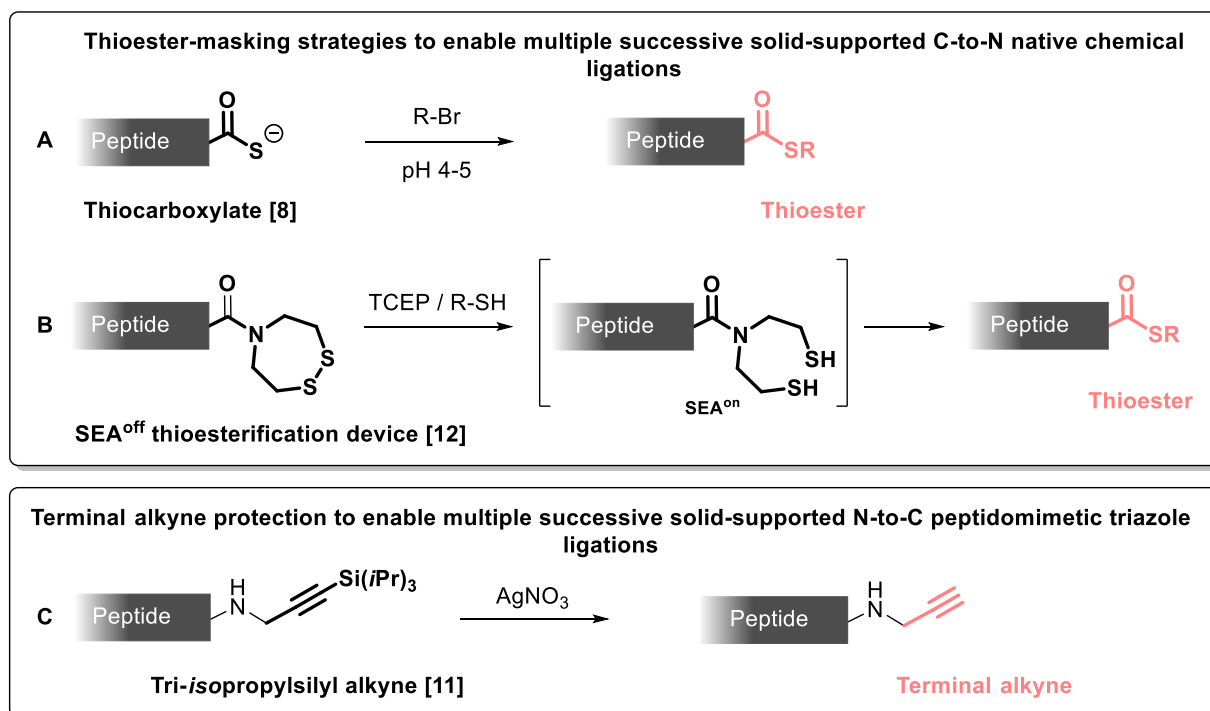


Figure 8 – Temporary masking strategies used for solid-supported iterative ligations in the N-to-C direction

2) Linkers for N-to-C SPCL

The N-to-C SPCL strategy requires the introduction of the SPCL linker at the N-terminus, which is more straightforward for peptide chemists than the functionalization at the C-terminus. Kent on his first report [8] on SPCL derivatized the N-terminal peptide segment at its N-terminus by a base-labile SPCL linker derived from the Msc amine protecting group. The linker featured a ketone functionality for oxime ligation-mediated immobilization on a cellulose-based solid support containing an aminoxy group (figure 9A). The linker was incorporated in three steps after Boc-SPPS elongation, making use of a *N*-Boc protected amine-containing sulfonyl ethyl carbonate building block [46] followed by coupling of levulinic acid. This linker was applied to the synthesis of several polypeptides (68 to 115 amino acids) through the NCL-based solid phase assembly of up to three peptide segments, exploiting thiocarboxylates as thioester precursors. The linker is cleaved by aqueous base (NaOH, pH 12-14 for 2 min) to release the polypeptide into solution.

More than ten years later, Aucagne and Delmas introduced the (2-[2-(2-azido-ethoxy)-ethyl-sulfonyl]-2-ethoxycarbonyl linker (N₃-Esoc) derived from the same Msc protecting group, and

functionalized with an azide moiety (figure 9B) [11]. N₃-Esoc can be incorporated on the N-terminal amine of peptide segments using an activated carbonate derivative. Peptide equipped with this linker can be immobilized on an alkyne-functionalized solid support by either CuAAC or strain-promoted azide/alkyne cycloaddition (SPAAC) [47]. This linker was first applied to the synthesis of a triazole-containing analogue of the extracellular domain of the human mucin 1 (MUC1), composed of a densely *O*-glycosylated twenty amino acids tandem repeat. Three successive peptidomimetic triazole ligations using segments equipped with TIPS-masked alkynes, followed by linker cleavage at pH 11.5 afforded an unglycosylated 160 residue polypeptide in a 60 % overall yield. This constitutes the longest sequence that has been assembled through SPCL to date.

The same linker has been exploited for NCL-based SPCL in collaboration with Melnyk [12], using segment incorporating a SEA^{off} moiety as masked thioester. Assembly of five peptide segments gave a 136 amino acids polypeptide derived from hepatocyte growth factor (HGF) in 6.5 % yield.

The basic conditions required to cleave the sulfonyl ethyl carbamate are relatively harsh, and were shown to be incompatible with *O*-glycosylated serine residues, known to be sensitive to base-mediated beta-elimination. With the objective to extend N-to-C SPCL to the synthesis of *O*-glycoproteins, Aucagne and Delmas developed another azide-functionalized linker, 1-azido-5-[1,3-dimethyl-2,4,6(1*H*,3*H*,5*H*)-trioxypyrimidine-5-ylidene]pentyl (N₃-Dtp), [14] (figure 9C) derived from the enamine-based amine protecting group DTPM [48,49] which can be cleaved by hydrazinolysis under mild conditions. N₃-Dtp was introduced on a peptide segment through the coupling of a derivative of the last amino acid *N*-protected with this linker. Assembly of a 160 residues triazole-containing MUC1 polypeptide on a controlled pore glass (CPG) solid support was realized through four successive PTL. Linker cleavage can be achieved by aqueous hydrazine at neutral pH, but these conditions were shown to partially reduce TIPS-protected C-terminal alkyne groups, which complicated the monitoring of each ligation and unmasking step through analytical scale cleavages. Optimized conditions using aqueous hydroxylamine allowed clean and efficient cleavage of the linker.

More recently, Melnyk investigated the use of an acetoacetyloxime (AcA) linker (figure 9D) [50] which was introduced by solution phase NCL of a N-terminal cysteinyl segment with an

AcA thioester. The segment can then be grafted onto an aminoxy-functionalized SPCL solid support by oxime ligation. This strategy was used to synthesize polypeptides derived from the K1 domain of HGF on a PEGA solid support. Treatment with a mixture of hydroxylamine and aniline at a mildly acidic pH liberates an acetoacetyloxime in solution, which undergoes spontaneous cyclizative amide bond cleavage affording the native protein. Hydroxamate formation resulting from the reaction with Asn/Gln was observed during the linker cleavage, however decreasing the pH and hydroxylamine concentration was found to nearly abolish this side reaction.

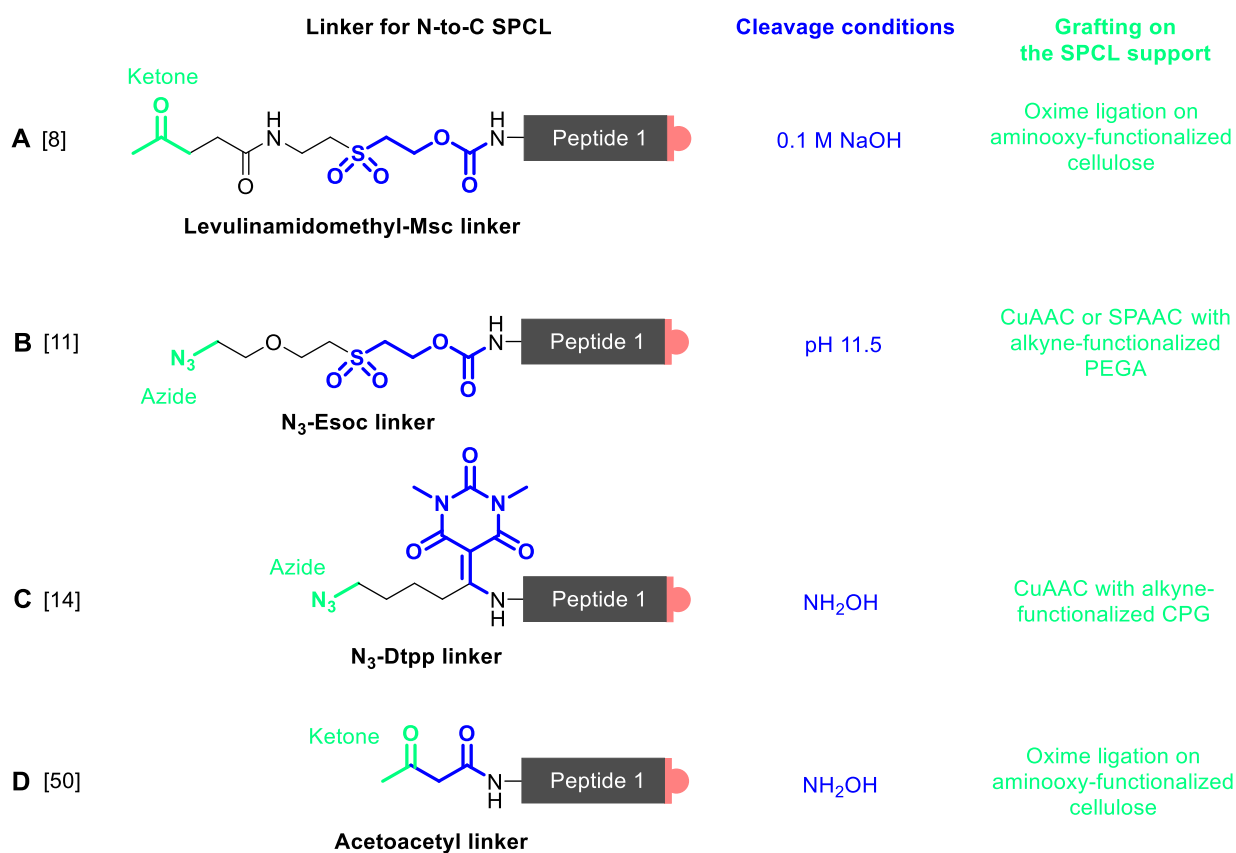


Figure 9 – Different cleavable linkers used to immobilize the N-terminal segment on a SPCL solid support

3) Case study

In order to illuminate SPCL from a more practical point of view, we have chosen one example of N-to-C assembly of a protein to illustrate the performance of the process by showing the HPLC analysis of the intermediates and final products (Figure 10). This case study is based on the 2013 report by Melnyk and our group [12] exploiting the N₃-Esoc linker along with the blocked thioester properties of the SEA^{off} group for the assembly of five segments to give a 15 kDa polypeptide derived from human hepatocyte growth factor 1 (HGF1). To date, this constitutes the only example of more than three successive ligations conducted on a solid support.

The first segment was immobilized onto PEGA1900 resin through a SPAAC reaction, then the N-to-C elongation was based on a cycle of (1) conversion of the C-terminal SEA^{off} group into a thioester and (2) NCL reactions. For the coupling of the last segments, these two steps were realized concomitantly. The target polypeptide was obtained in an overall 6.5 % yield after a final HPLC purification. The remarkable HPLC purity of all the different intermediates highlights the efficiency of the SPCL process.

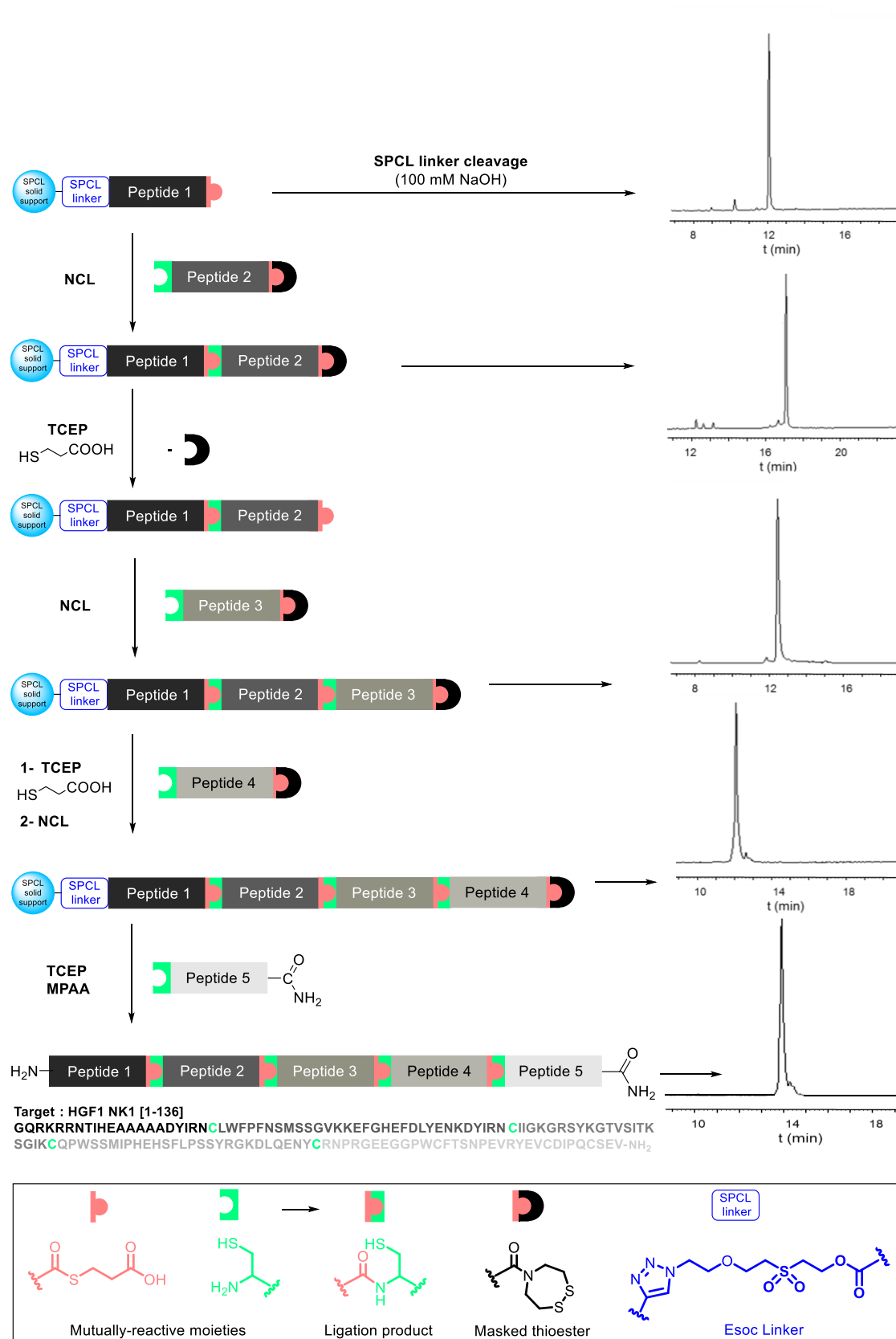


Figure 10 – Case study of the synthesis of HGF1 NK1 [1-136] through five successive solid-supported ligations (adapted from ref [12] with permission from The Royal Society of Chemistry)

4) SPCL with concomitant purifications

Protein elongation in the N-to-C direction provides an important advantage over the alternative C-to-N direction, which is the “self-purification” of the incoming peptide segments, provided that a capping step is used after each amino acid coupling during SPPS. The side products are normally truncated sequences rather than deleted ones. This self-purification effect is based on the fact that only the full-length peptide segments contain the N-terminal reactive group needed for immobilization or segment ligation, *i.e.* the linker, a cysteine for NCL or an azide for PTL. Truncated peptides, which constitutes the major co-products from Fmoc-SPPS, are unreactive and are simply washed off after the immobilization and ligation steps, allowing to work with crude, unpurified peptide segments providing that no other side reaction occurred during SPPS (e.g. aspartimide formation, oxidation etc...). This self-purification effect has been largely exploited for the non-chromatographic purification of peptides from the 70's to the 90's, [46, 51–56] and the extension of this concept for SPCL with self-purification has been pioneered by Aucagne and Delmas in 2012 [11], then exploited by other groups [50,57].

IV. Post-ligations solid-supported transformations

Taking significant advantage of the solid support used for SPCL, different chemical or biochemical transformations can be implemented after ligation-based protein elongation, in order to simplify the work-up and to limit purification steps.

1) Chemical transformations

To bypass the limitation of the requirement of the rare Cys residue at the ligation disconnection, desulfurization reactions after ligation has been developed, but generate extra handling and purification steps. The metal-free desulfurization approach developed by Danishefsky [58] using a radical initiator in combination with *tris*-(2-carboxyethyl)phosphine (TCEP) is the most popular method. It was successfully applied to SPCL by Brik [15] and then Ottesen [25] to convert cysteine and penicillamine residues into Cys and Val, respectively.

Probes can be chemoselectively conjugated to the synthesized supported-protein, as demonstrated by Aucagne and Delmas with the CuAAC-based conjugation of a biotin [14].

Photochemical processes can also be conducted on the solid support. Brik used thioester segments incorporating a C-terminal Lys protected with 6-nitroveratryloxycarbonyl (Nvoc) group on side-chain to prevent lactamization during the ligation step [15]. The Nvoc group was successfully removed before SPCL linker cleavage through exposure to UV.

2) Biochemical transformations

Formation of disulfide bonds is a post-translational modification crucial for the structure stabilization and function of many proteins. The total chemical synthesis of disulfide-containing proteins involves a final oxidative folding step, i.e. formation of the disulfide bond concomitantly with the folding of the peptide backbone. Solid-supported oxidative folding would benefit from the pseudo-dilution effect [59], which favors intramolecular reactions, thus potentially avoiding intermolecular aggregation, a common problem during protein folding. In 2006, Kent [10] explored this idea to fold the disulfide rich miniprotein EETI-II (28 amino acids, three disulfide bonds) by air oxidation after SPCL, while it was still covalently attached to the solid support. Note that in this case the oxidative folding processed slower on solid phase than in solution.

In order to provide a simple access to long glycoproteins, Aucagne and Delmas combined SPCL of peptide segments with enzymatic glycosylation [14]. In this work, they performed a solid-supported enzymatic glycosylation by using recombinant polypeptide α -N-acetyl-galactosaminyl transferase 1 (pp-GalNAc-T1). In order to facilitate monitoring of enzymatic glycosylation and the different ligation and unmasking steps, a TFA-labile PAL linker [71] was incorporated between the alkyne used for the CuAAC-mediated immobilization and the solid support. Different solid supports were screened, and the best results were obtained using controlled pore glass (CPG) beads with pore size around 100 nm. The enzymatic reaction was much slower than in solution, and kinetics were improved by the introduction of a PEG spacer (3000 Da) between the polypeptide chain and the solid support. The expected product incorporating 160 amino acids and 20 carbohydrate residues was purified in a rather moderate 6% yield, due to the presence in the crude product of small amounts of under-glycosylated sequences (18 or 19 GalNAc residues) difficult to separate from the target.

V. Solid support

The ideal support for solid-phase ligation must be compatible with aqueous conditions required for the solvation of unprotected peptides, ideally including a large range of buffers and pH, and to both ligations and unmasking conditions. A solid support does not behave as an inert carrier but rather acts as a co-solvent [60, 61] and this feature will affect any parameters of a chemical reaction, yield, purity, speed and cost.

Resin bead size has an effect on the reaction rate as well. Many studies conducted on solid phase organic chemical reactions showed that the diffusion rate increased with decreasing bead size.[62–64] What was observed in organic solvents is certainly also true in aqueous media that are used for SPCL. The degree of cross-linking can also have a large impact on reaction yield, as shown by Meldal [62] and Lippens [65], who observed that increasing cross-linking slowed down reactions by lowering the diffusion rates. Diffusion of peptide segments could be a limiting process in SPCL likewise.

The nature of the SPCL resin can impair the ligation reaction as observed by Dawson [9] with the PEGA resin. He attributed poor NCL yields to peptide aggregation on the support due to the high loading of the PEGA resin used. Yields were not improved by adding a large excess of peptide segment or increasing the temperature and ligation time. Aggregation on PEGA was also hypothesized by Ottesen [25] who found that a tenfold reduction of the resin loading significantly improved kinetic reaction rates. In this special case, this is however proved to be insufficient for long polypeptide chains: a first series of three solid-supported ligations were successfully carried out whereas the last one had to be performed in solution to avoid presumable aggregation issues.

The nature of the solid support can also affect the SPCL linker cleavage kinetics as demonstrated by Aucagne and Delmas in their work on N₃-Esoc linker [11]. Base-mediated cleavage of this linker was found to be considerably slower when the peptide was attached to polyethylene glycol-based water-compatible resins than in solution phase. TRIS buffer at pH = 8.5 was sufficient to cleave the linker in solution ($t_{1/2}$ = 7 h), whereas pH has to be raised up to 10.7 (50 mM arginine) on solid support, with polymer-dependant rates: $t_{1/2}$ = 13.7 h, 5.4 h and 2.1 h for ChemMatrix [66], PEGA₈₀₀ and PEGA₁₉₀₀, respectively.

Besides non-covalent effects, the chemical structure of the polymer has obviously to be compatible with the chemical reaction used. For example Dawson [9] observed that the CLEAR solid support (Cross-Linked Ethoxylate-Acrylate Resin) developed by Barany [67] completely dissolved upon treatment with mercury(II) acetate in acetic acid used for AcM deprotection, probably resulting from ester hydrolysis. Furthermore, this resin is known to be unstable under the basic conditions used to remove the cysteine-masking *N*-Msc group [9].

Up to now, three families of solid support have been used for SPCL (Figure 11): i) biosourced oligosaccharide-based polymers, ii) PEG-based synthetic polymers, iii) inorganic matrixes (i.e. control pore glass, CPG).

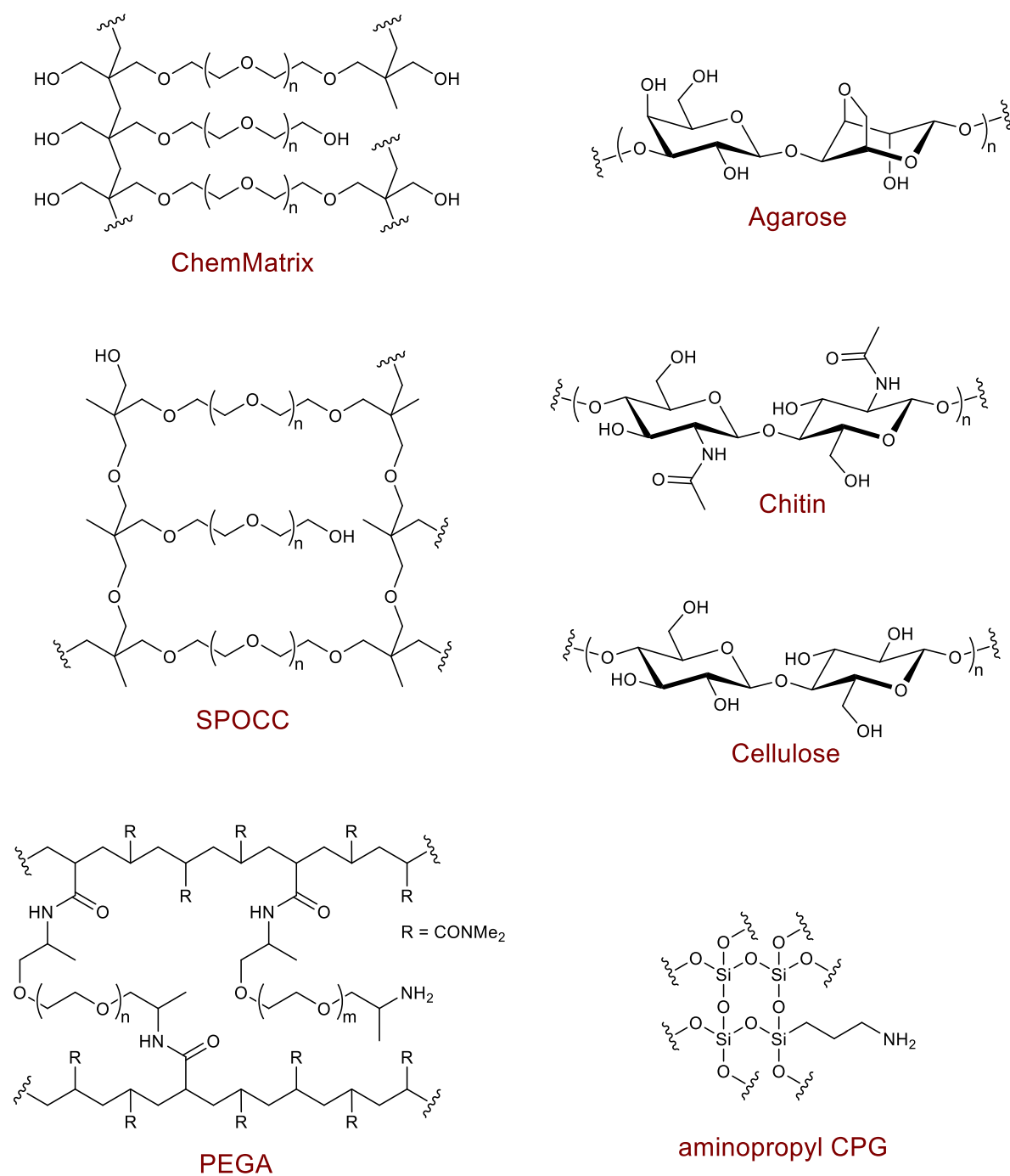


Figure 11 – Different supports that have been employed for SPCL

Kent employed a spherilose resin [8], which is a cellulose-based solid support originally developed for affinity chromatography of proteins. Similarly, Dawson reported the use of a sepharose resin consisting of cross-linked agarose [9]. Muir [7] used a chitin-based solid support, which has physicochemical properties similar to cellulose. These three matrices classically used in biochemistry, have been used in the first reports on SPCL [7,8,9,] and more recently by Seitz [57]. These polymers have been shown to be very suitable for the aqueous

conditions used in SPCL. However, their main drawback concerns their limited chemical stability.

Polar PEG-based polymers emerged in the late 90s with the search for resins compatible with both solid phase organic chemistry and enzymatic reactions. Kent [10] reported the use of a PEG-poly(oxetane)copolymer, SPOCC [30,31], which was commercially-available at that time. In more recent works, PEG-polyacrylamide PEGA resins [68] have widely been used. PEGA are characterized by high swelling properties in both non polar and hydrophilic solvents and by high functional group reactivity facilitating the gel phase reaction to go to completion. Several PEGA resins have been developed by Meldal [69], differing in the length of the PEG chain. Two different ones have been previously used for SPCL: PEGA₈₀₀ (800 Da PEG) [11,13,15, 25, 32, 50] and the no longer commercially available but very SPCL-suitable PEGA₁₉₀₀ (1900 Da) [12]. Note that the size of the PEG chain was omitted as in many reports on works done with a PEGA resin, and we suppose that PEGA₈₀₀ was used because of its commercial availability. Important to note, the industrial production processes of PEGA polymers may have been changed as a result of successive licence transfers to different companies, as we have observed with the evolution of physical behavior of PEGA₈₀₀ in the past decade.

An alternative to organic polymers is the use of controlled pore glass CPG 1000 Å [70] as reported by Aucagne and Delmas [14]. CPG is a non-swelling inorganic matrix, a silica glass forming porous beads with pores of 75 to 3000 angstroms Å in diameter. This solid support has been originally developed for solid-phase oligonucleotide synthesis. CPG is chemically stable and is compatible with both aqueous and organic conditions, but not well suited for SPPS.

VI. Conclusion and perspectives

About 20 years after the first report on the solid-supported version of chemical ligation *aka* SPCL, the method is still a compelling and highly challenging field of research. The adaptation of ligation chemistry to the solid phase may have the potential to considerably extend the limits of protein chemical synthesis. However, despite the significant gains made in advancing the technology, there are many remaining limitations. SPCL is still typically limited to less than three successive solid-supported ligation reactions, this being probably –at least in part– due to aggregation-prone segments or intermediates. From a technical perspective, monitoring SPCL reactions using on-resin-analysis would facilitate the optimization of the ligation and unmasking reactions, but that remains a very tricky task and physical analytic tools are not available in routine. For example FT-IR is not resolutive enough to give information for SPCL, and instrumental developments would be needed to use MAS-NMR [65] as a routine tool to analyse SPCL reactions. Alternatives are based on destructive approaches such as analytical scale “micro-cleavage” followed by LC-MS analysis of the released peptide. For SPCL reaction optimization purposes, working on sub-micromole analytical scale is still facing technical limitations related to resin handling, available glassware or maintaining oxygen-free conditions necessary for NCL and CuAAC.

Thus, there is still room for major technical breakthroughs such as development of new polymers, new linkers, new ligations reactions with improved kinetics, new masking/unmasking strategies to broaden the scope of SPCL applications to more ambitious targets. New instrumentation would probably be also needed in order to allow the routine use of SPCL by non-specialists in a comparable way as SPPS. Why not one day even come to an automated version and bring to the market proteins synthesizers to complement peptide synthesizers?

VI. References

1. Merrifield, R.B. (1963) Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.*, **85** (14), 2149–2154.
2. Schnolzer, M., and Kent, S. (1992) Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease. *Science*, **256** (5054), 221–225.
3. Weinstock, M.T., Jacobsen, M.T., and Kay, M.S. (2014) Synthesis and folding of a mirror-image enzyme reveals ambidextrous chaperone activity. *Proc. Natl. Acad. Sci. USA*, **111** (32), 11679–11684.
4. Sun, H., and Brik, A. (2019) The Journey for the Total Chemical Synthesis of a 53 kDa Protein. *Acc. Chem. Res.*, **52** (12), 3361–3371.
5. Sun, H., Mali, S.M., Singh, S.K., Meledin, R., Brik, A., Kwon, Y.T., Kravtsova-Ivantsiv, Y., Bercovich, B., and Ciechanover, A. (2019) Diverse fate of ubiquitin chain moieties: The proximal is degraded with the target, and the distal protects the proximal from removal and recycles. *Proc Natl Acad Sci USA*, **116** (16), 7805–7812.
6. Dawson, P., Muir, T., Clark-Lewis, I., and Kent, S. (1994) Synthesis of proteins by native chemical ligation. *Science*, **266** (5186), 776–779.
7. Cotton, G.J., and Muir, T.W. (2000) Generation of a dual-labeled fluorescence biosensor for Crk-II phosphorylation using solid-phase expressed protein ligation. *Chem. Biol.*, **7** (4), 253–261.
8. Canne, L.E., Botti, P., Simon, R.J., Chen, Y., Dennis, E.A., and Kent, S.B.H. (1999) Chemical Protein Synthesis by Solid Phase Ligation of Unprotected Peptide Segments. *J. Am. Chem. Soc.*, **121** (38), 8720–8727.
9. Brik, A., Keinan, E., and Dawson, P.E. (2000) Protein Synthesis by Solid-Phase Chemical Ligation Using a Safety Catch Linker. *J. Org. Chem.*, **65** (12), 3829–3835.
10. Johnson, E.C.B., Durek, T., and Kent, S.B.H. (2006) Total Chemical Synthesis, Folding, and Assay of a Small Protein on a Water-Compatible Solid Support. *Angew. Chem. Int. Ed.*, **45** (20), 3283–3287.
11. Aucagne, V., Valverde, I.E., Marceau, P., Galibert, M., Dendane, N., and Delmas, A.F. (2012) Towards the Simplification of Protein Synthesis: Iterative Solid-Supported Ligations with Concomitant Purifications. *Angew. Chem. Int. Ed.*, **51** (45), 11320–11324.
12. Raibaut, L., Adihou, H., Desmet, R., Delmas, A.F., Aucagne, V., and Melnyk, O. (2013) Highly efficient solid phase synthesis of large polypeptides by iterative ligations of bis(2-sulfanylethyl)amido (SEA) peptide segments. *Chem. Sci.*, **4** (10), 4061–4066.

13. Decostaire, I.E., Lelièvre, D., Aucagne, V., and Delmas, A.F. (2014) Solid phase oxime ligations for the iterative synthesis of polypeptide conjugates. *Org. Biomol. Chem.*, **12** (29), 5536–5543.
14. Galibert, M., Piller, V., Piller, F., Aucagne, V., and Delmas, A.F. (2015) Combining triazole ligation and enzymatic glycosylation on solid phase simplifies the synthesis of very long glycoprotein analogues. *Chem. Sci.*, **6** (6), 3617–3623.
15. Jbara, M., Seenaiiah, M., and Brik, A. (2014) Solid phase chemical ligation employing a rink amide linker for the synthesis of histone H2B protein. *Chem. Commun.*, **50** (83), 12534–12537.
16. Camarero, J.A., Cotton, G.J., Adeva, A., and Muir, T.W. (1998) Chemical ligation of unprotected peptides directly from a solid support. *J. Pept. Res.*, **51** (4), 303–316.
17. Valverde, I.E., Lecaille, F., Lalmanach, G., Aucagne, V., and Delmas, A.F. (2012) Synthesis of a Biologically Active Triazole-Containing Analogue of Cystatin A Through Successive Peptidomimetic Alkyne-Azide Ligations. *Angew. Chem. Int. Ed.*, **51** (3), 718–722.
18. Rostovtsev, V.V., Green, L.G., Fokin, V.V., and Sharpless, K.B. (2002) A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective ‘ligation’ of azides and terminal alkynes. *Angew. Chem. Int. Ed. Engl.*, **41** (14), 2596–2599.
19. Tornøe, C.W., Christensen, C., and Meldal, M. (2002) Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.*, **67** (9), 3057–3064.
20. Dirksen, A., Hackeng, T.M., and Dawson, P.E. (2006) Nucleophilic Catalysis of Oxime Ligation. *Angew. Chem. Int. Ed.*, **45** (45), 7581–7584.
21. Saito, F., Noda, H., and Bode, J.W. (2015) Critical Evaluation and Rate Constants of Chemoselective Ligation Reactions for Stoichiometric Conjugations in Water. *ACS Chem. Biol.*, **10** (4), 1026–1033.
22. Presolski, S.I., Hong, V.P., and Finn, M.G. (2011) Copper-Catalyzed Azide–Alkyne Click Chemistry for Bioconjugation. *Curr. Protoc. Chem. Biol.*, **3** (4), 153–162.
23. Tesser, G.I., and Balvert-Geers, I.C. (1975) The methylsulfonylethoxycarbonyl group, a new and versatile amino protective function. *Int. J. Pept. Protein Res.*, **7** (4), 295–305.
24. Yu, R.R., Mahto, S.K., Justus, K., Alexander, M.M., Howard, C.J., and Ottesen, J.J. (2016) Hybrid phase ligation for efficient synthesis of histone proteins. *Org. Biomol. Chem.*, **14** (9), 2603–2607.

25. Veber, D., Milkowski, J., Varga, S., Denkwalter, R., and Hirschmann, R. (1972) Acetamidomethyl. A Novel Thiol Protecting Group for Cysteine. *J. Am. Chem. Soc.*, **94** (15), 5456–5461.
26. Maity, S.K., Jbara, M., Laps, S., and Brik, A. (2016) Efficient Palladium-Assisted One-Pot Deprotection of (Acetamidomethyl)Cysteine Following Native Chemical Ligation and/or Desulfurization To Expedite Chemical Protein Synthesis. *Angew. Chem. Int. Ed.*, **55** (28), 8108–8112.
27. Stevens, C.M., and Watanabe, R. (1950) Amino Acid Derivatives. I. Carboallyloxy Derivatives of α -Amino Acids. *J. Am. Chem. Soc.*, **72** (2), 725–727.
28. Kunz, H., and Unverzagt, C. (1984) The Allyloxycarbonyl (Aloc) Moiety: Conversion of an Unsuitable into a Valuable Amino Protecting Group for Peptide Synthesis. *Angew. Chem. Int. Ed. Engl.*, **23** (6), 436–437.
29. Favel, A., Mattras, H., Coletti-Previero, M.A., Zwilling, R., Robinson, E.A., and Castro, B. (1989) Protease inhibitors from Ecballium elaterium seeds. *Int. J. Pept. Protein Res.*, **33** (3), 202–208.
30. Rademann, J., Grøtli, M., Meldal, M., and Bock, K. (1999) SPOCC: A Resin for Solid-Phase Organic Chemistry and Enzymatic Reactions on Solid Phase. *J. Am. Chem. Soc.*, **121** (23), 5459–5466.
31. Grøtli, M., Gotfredsen, C.H., Rademann, J., Buchardt, J., Clark, A.J., Duus, J.Ø., and Meldal, M. (2000) Physical Properties of Poly(ethylene glycol) (PEG)-Based Resins for Combinatorial Solid Phase Organic Chemistry: A Comparison of PEG-Cross-Linked and PEG-Grafted Resins. *J. Comb. Chem.*, **2** (2), 108–119.
32. Gless, B.H., Bojer, M.S., Peng, P., Baldry, M., Ingmer, H., and Olsen, C.A. (2019) Identification of autoinducing thiodepsipeptides from staphylococci enabled by native chemical ligation. *Nat. Chem.*, **11** (5), 463–469.
33. Martinez, J., Laur, J., and Castro, B. (1983) Carboxamidomethyl esters (CAM esters) as carboxyl protecting groups. *Tetrahedron Lett.*, **24** (47), 5219–5222.
34. Pátek, M., and Lebl, M. (1991) Safety-catch anchoring linkage for synthesis of peptide amides by Boc/Fmoc strategy. *Tetrahedron Lett.*, **32** (31), 3891–3894.
35. Atherton, E., Logan C. J., and Sheppard, R. C. (1981) Peptide synthesis. Part 2. Procedures for solid-phase synthesis using N^{α} -fluorenylmethoxycarbonylamino-acids on polyamide supports. Synthesis of substance P and of acyl carrier protein 65–74 decapeptide. *J. Chem. Soc., Perkin Trans. 1*, 538–546.

36. Muir, T.W., Sondhi, D., and Cole, P.A. (1998) Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci. USA*, **95** (12), 6705–6710.
37. Evans, T.C., Benner, J., and Xu, M.-Q. (1999) The Cyclization and Polymerization of Bacterially Expressed Proteins Using Modified Self-splicing Inteins. *J. Biol. Chem.*, **274** (26), 18359–18363.
38. Berrade, L., and Camarero, J.A. (2009) Expressed protein ligation: a resourceful tool to study protein structure and function. *Cell. Mol. Life Sci.*, **66** (24), 3909–3922.
39. Zitterbart, R., and Seitz, O. (2016) Parallel Chemical Protein Synthesis on a Surface Enables the Rapid Analysis of the Phosphoregulation of SH3 Domains. *Angew. Chem. Int. Ed.*, **55** (25), 7252–7256.
40. Zitterbart, R., Krumrey, M., and Seitz, O. (2017) Immobilization methods for the rapid total chemical synthesis of proteins on microtiter plates: Rapid Chemical Synthesis of Proteins on Microtiter Plates. *J. Pept. Sci.*, **23** (7–8), 539–548.
41. Note that Huang and coworkers described a single-support N-to-C SPCL strategy, where the N-terminal segment is synthesized on a polystyrene resin by side-chain anchoring of a C-terminal Asn residue through a Rink amide linker. Side-chain protections are kept during SPCL, or simple non-trifunctional amino acids are used, and only model peptides up to 15 amino acids have been synthesized through a single NCL reaction. The approach is inherently limited by the use of a polystyrene resin that is not compatible with aqueous buffers, and is thus not detailed herein. See: Zhou, B., Faridoon, Tian, X., Li, J., Guan, D., Zheng, X., Guo, Y., and Huang, W. (2018) On-resin peptide ligation via C-terminus benzyl ester. *Chin. Chem. Lett.*, **29** (7), 1123–1126.
42. Ollivier, N., Dheur, J., Mhidia, R., Blanpain, A., and Melnyk, O. (2010) Bis(2-sulfanylethyl)amino Native Peptide Ligation. *Org. Lett.*, **12** (22), 5238–5241.
43. Hou, W., Zhang, X., Li, F., and Liu, C.-F. (2011) Peptidyl *N*, *N*-Bis(2-mercaptoethyl)-amides as Thioester Precursors for Native Chemical Ligation. *Org. Lett.*, **13** (3), 386–389.
44. Ollivier, N., Vicogne, J., Vallin A, Drobecq, H., Desmet, R., El Mahdi, O., Leclercq, B., Goormachtigh, G., Fafeur, V., and Melnyk, O. (2012) A one-pot three-segment ligation strategy for protein chemical synthesis. *Angew. Chem. Int. Ed. Engl.*, **51** (1), 209–213.
45. Valverde, I.E., Delmas, A.F., and Aucagne, V. (2009) Click à la carte: robust semi-orthogonal alkyne protecting groups for multiple successive azide/alkyne cycloadditions. *Tetrahedron*, **65** (36), 7597–7602.
46. Canne, L.E., Winston, R.L., and Ken, S.B.H. (1997) Synthesis of a versatile purification handle for use with Boc chemistry solid phase peptide synthesis. *Tetrahedron Lett.*, **38** (19), 3361–3364.

47. Agard, N.J., Prescher, J.A., and Bertozzi, C.R. (2004) A Strain-Promoted [3 + 2] Azide–Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.*, **126** (46), 15046–15047.
48. Dekany, G., Bornaghi, L., Papageorgiou, J., and Taylor, S. (2001) A novel amino protecting group: DTPM. *Tetrahedron Lett.*, **42** (17), 3129–3132.
49. da Silva, E.T., and Lima, E.L.S. (2003) Reaction of 1,3-dimethyl-5-acetyl-barbituric acid (DAB) with primary amines. Access to intermediates for selectively protected spermidines. *Tetrahedron Lett.*, **44** (18), 3621–3624.
50. Ollivier, N., Desmet, R., Drobecq, H., Blanpain, A., Boll, E., Leclercq, B., Mougel, A., Vicogne, J., and Melnyk, O. (2017) A simple and traceless solid phase method simplifies the assembly of large peptides and the access to challenging proteins. *Chem. Sci.*, **8** (8), 5362–5370.
51. Funakoshi, S., Fukuda, H., and Fujii, N. (1991) Chemoselective one-step purification method for peptides synthesized by the solid-phase technique. *Proc. Natl. Acad. Sci. USA*, **88** (16), 6981–6985.
52. del Solar, G., Albericio, F., Eritja, R., and Espinosa, M. (1994) Chemical synthesis of a fully active transcriptional repressor protein. *Proc. Natl. Acad. Sci. USA*, **91** (11), 5178–5182.
53. Hara, T., Tainosho, A., Nakamura, K., Sato, T., Kawakami, T., and Aimoto, S. (2009) Peptide purification by affinity chromatography based on α -ketoacyl group chemistry. *J. Pept. Sci.*, **15** (5), 369–376.
54. Vizzavona, J., Villain, M., and Rose, K. (2002) Covalent capture purification of polypeptides after SPPS via a linker removable under very mild conditions. *Tetrahedron Lett.*, **43** (48), 8693–8696.
55. Krieger, D.E., Erickson, B.W., and Merrifield, R.B. (1976) Affinity purification of synthetic peptides. *Proc. Natl. Acad. Sci. USA*, **73** (9), 3160–3164.
56. Merrifield, R.B., and Bach, A.E. (1978) 9-(2-Sulfo)fluorenylmethyloxycarbonyl chloride, a new reagent for the purification of synthetic peptides. *J. Org. Chem.*, **43** (25), 4808–4816.
57. Loibl, S.F., Harpaz, Z., Zitterbart, R., and Seitz, O. (2016) Total chemical synthesis of proteins without HPLC purification. *Chem. Sci.*, **7** (11), 6753–6759.
58. Wan, Q., and Danishefsky, S.J. (2007) Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew. Chem. Int. Ed.*, **46** (48), 9248–9252.

59. Jayalekshmy, P., and Mazur, S. (1976) Pseudodilution, the solid-phase immobilization of benzyne. *J. Am. Chem. Soc.*, **98** (21), 6710–6711.
60. Czarnik, A. W., (1998) Solid-phase synthesis supports are like solvents. *Biotechnol. Bioeng.*, **61** (1), 77–79.
61. Feng, Y., and Burgess, K. (2000) Resin effects in solid phase S(n)Ar and S(n)2 macrocyclizations. *Biotechnol. Bioeng.*, **71** (1), 3–8.
62. Groth, T., Grøtli, M., and Meldal, M. (2001) Diffusion of Reagents in Macrobeads. *J. Comb. Chem.*, **3** (5), 461–468.
63. Li, W., Xiao, X., and Czarnik, A.W. (1999) Kinetic Comparison of Amide Formation on Various Cross-Linked Polystyrene Resins. *J. Comb. Chem.*, **1** (2), 127–129.
64. Kress, J., Zanaletti, R., Rose, A., Frey, J.G., Brocklesby, W.S., Ladlow, M., and Bradley, M. (2003) Which Sites React First? Functional Site Distribution and Kinetics on Solid Supports Investigated Using Confocal Raman and Fluorescence Microscopy. *J. Comb. Chem.*, **5** (1), 28–32.
65. Rousselot-Pailley, P., Ede, N.J., and Lippens, G. (2001) Monitoring of Solid-Phase Organic Synthesis on Macroscopic Supports by High-Resolution Magic Angle Spinning NMR. *J. Comb. Chem.*, **3** (6), 559–563.
66. García-Martín, F., Quintanar-Audelo, M., García-Ramos, Y., Cruz, L.J., Gravel, C., Furic, R., Côté, S., Tulla-Puche, J., and Albericio, F. (2006) ChemMatrix, a Poly(ethylene glycol)-Based Support for the Solid-Phase Synthesis of Complex Peptides. *J. Comb. Chem.*, **8** (2), 213–220.
67. Kempe, M., and Barany, G. (1996) CLEAR: A Novel Family of Highly Cross-Linked Polymeric Supports for Solid-Phase Peptide Synthesis^{1,2}. *J. Am. Chem. Soc.*, **118** (30), 7083–7093.
68. Renil, M., and Meldal, M. (1995) Synthesis and application of a PEGA polymeric support for high capacity continuous flow solid-phase peptide synthesis. *Tetrahedron Lett.*, **36** (26), 4647–4650.
69. Auzanneau, F.-I., Meldal, M., and Bock, K. (1995) Synthesis, characterization and biocompatibility of PEGA resins. *J. Pept. Sci.*, **1** (1), 31–44.
70. Damha, M.J., Giannaris, P.A., and Zabarylo, S.V. (1990) An improved procedure for derivatization of controlled-pore glass beads for solid-phase oligonucleotide synthesis. *Nucleic Acids Res.*, **18** (13), 3813–3821.

71. Albericio, F., and Barany, G. (1987) An acid-labile anchoring linkage for solid-phase synthesis of C-terminal peptide amides under mild conditions. *Int. J. Peptide Protein Res.*, **30** (2) 206–216.

Ref	year	Direction	Immobilization	Solid support	Cleavable linker(s)	Linker cleavage	Ligation type	Nb of solid phase ligations	Overall yield (%)	Prominent targets
[8]	1999	C-to-N	Oxime	Cellulose	ester (CAM)	pH 12-14	NCL	3	<i>nd</i>	GV-PLA2 (118 aa)
[8]	1999	N-to-C	Oxime	Cellulose	Msc	pH 12-14	NCL	2	<i>nd</i>	MIF (115 aa)
[9]	2000	C-to-N	NCL	Sepharose	SCAL	SiCl ₄ in TFA	NCL	2	20	vMIP I (71 aa)
[7]	2000	C-to-N	Biotin/avidin	Agarose	PreScission protease substrate	PreScission protease	NCL (EPL)	1	55	Crk-II (304 aa)
[10]	2006	C-to-N	SPCL on SPPS resin	SPOCC	PHB	TFA	NCL	2	80-90*	EETI-II (28 aa)
[11]	2012	N-to-C	CuAAC	PEGA ₈₀₀	Esoc / PAL	pH 11.5 / TFA	PTL	3	60	MUC1 (160 aa)
[12]	2013	N-to-C	CuAAC or SPAAC	PEGA ₁₉₀₀	Esoc	pH 11.5	NCL	4	6.5	HGF fragment (136 aa)
[15]	2014	C-to-N	SPCL on SPPS resin	PEGA	Rink	TFA	NCL	3	10	HA-tagged histone H2B (135 aa)
[13]	2014	C-to-N	SPCL on SPPS resin	PEGA800	HMPA	TFA	Oxime	2	33	Padre-MUC1 (61 aa)
[14]	2015	N-to-C	CuAAC	CPG100	Dtp	hydroxylamine	PTL	3	6	Glycosylated MUC1 (160 aa)
[24]	2016	C-to-N	Oxime	PEGA	MBHA / Rink	pH 13 / TFA	NCL	3	>90% and 16%	CpA (70 aa) and histone H4 (65 aa) fragments
[39]	2016	C-to-N	Hydrazone	Polypropylene	No cleavable linker	-	NCL	1	-	SH3 domain (61 aa)
[57]	2016	C-to-N	His tag / Ni-NTA	Agarose	No cleavable linker	-	NCL	1	-	MUC1 (126 aa)
[50]	2017	N-to-C	Oxime	PEGA	acetoacetyloxime	hydroxylamine	NCL	3	23	HGF fragment (87 aa)
[40]	2017	C-to-N	His tag / Ni-NTA or CuAAC or hydrazone	Polypropylene	No cleavable linker	-	NCL	1	-	SH3 domain (61 aa)
[41]	2018	N-to-C	SPCL on SPPS resin	Polystyrene	Rink linker on Asn side chain	TFA	NCL	1	>90**	Protected model peptides (max 15 aa)

Appendix Table – Recapitulation of the SPCL strategies developed to date (*nd*: not determined *recovered yield **based on HPLC peak integration)

Chapter 2: Enzyme-cleavable linkers for the total synthesis of proteins through solid-supported ligations

Table of contents

I. Bibliographic study	65
1) Enzymes: biological catalysts proven to be very useful for synthetic chemists	65
2) Enzyme-cleavable linkers structure	66
3) Enzyme-cleavable linker applications	69
A. Targeted drug delivery	70
B. Imaging	75
C. Enzyme-labile linkers for solid phase synthesis	79
a) Solid support enzymatic accessibility	79
b) Solid phase synthesis of peptides, glycopeptides and peptide nucleic acid	82
c) Solid phase synthesis of oligosaccharides	83
d) Solid phase synthesis of organic molecules	85
II. Exploration of the use of enzyme-cleavable linkers for SPCL	87
1) Limits of chemically-cleavable linkers in SPCL	87
2) Thesis project aim	87
3) Enzyme-cleavable linkers investigated in this work	88
III. References for the bibliographic study part	91
IV. Publication 2 (Published in Angewandte Chemie International Edition)	101
V. Supporting Information	110

I. Bibliographic study

1) Enzymes: biological catalysts proven to be very useful for synthetic chemists

Enzymes are generally proteins but also include deoxyribonucleic acids (DNA) and ribonucleic acids (RNA). They are responsible for catalyzing reactions in a variety of biological processes in all living organisms by lowering reaction's activation energy barrier, hence increasing the rate of the reaction and improving its specificity. Chemists took advantage of this exquisite selectivity for applications to various synthetic processes, and in particular developed enzyme-cleavable protecting groups. The proper introduction and removal of protecting groups is one of the most important and widely carried out transformations in chemical synthesis, and enzyme-labile protecting group techniques are increasingly finding their use in almost all areas of synthetic organic chemistry. The use of such protecting groups in combination with classical methods offers the opportunity to carry out highly chemo- and regioselective transformations and a number of interesting enzyme-labile blocking moieties have been reported for the protection of amines, thiols, carboxylic acids and alcohols [1,2]. For a few representative examples in the field of peptide synthesis, we can mention the phenylacetamidomethyl (PhAcM) [3,4] protecting group for thiols, deprotected through penicillin G acylase (PGA)-catalyzed hydrolysis of the amide incorporated in an acylated thioaminal (fig. 1-a) as well as the pyroglutamyl amide (Glp) protecting group for the lysine side chain, [5] readily removed with pyroglutamate aminopeptidase from calf liver (fig. 1-b).

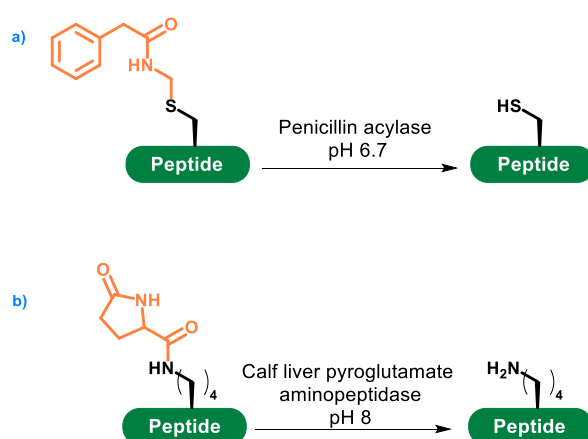


Figure 1: Examples of enzyme-labile peptide side chain protecting groups

2) Enzyme-cleavable linkers structure

A substantial number of enzyme-cleavable linkers has been inspired by these protecting groups, for a wide range of applications that are overviewed in this bibliography study.

A simple and general definition applicable to most cleavable linker could be: a chemical entity that links a compound of interest to a “carrier” (including, targeting biomolecules, nanoparticles, solid-supports etc...), and that can be cleaved to liberate the compound. An enzyme-cleavable linker is thus constituted of three parts: a moiety for attachment to the compound of interest, another for the attachment of the carrier, and an enzyme substrate, which enzymatic hydrolysis induces the liberation of the compound of interest and its separation from the carrier.

Two main categories of enzyme-cleavable linkers design can be defined. The first one involves the direct connection of an enzyme substrate, which acts like a protecting group, to a functional group of the compound of interest *via* an enzyme-labile bond (fig. 2-a). This strategy works very well when the linkage is easily accessible. However, when the linker and/or the compound of interest are sterically hindered, their spatial proximity can slow down or totally prevent the enzymatic reaction. [6,7]

Thus, to overcome this issue, a second strategy was developed, which is the most widely used approach. It consists in the incorporation of an additional “self-immolative” spacer designed to bridge the enzyme substrate and the enzyme-labile bond with the compound of interest. In this strategy, the enzymatic cleavage of a bond will generate a cascade of reactions leading to the disassembly of the spacer and the release of the compound of interest (fig. 2-b).

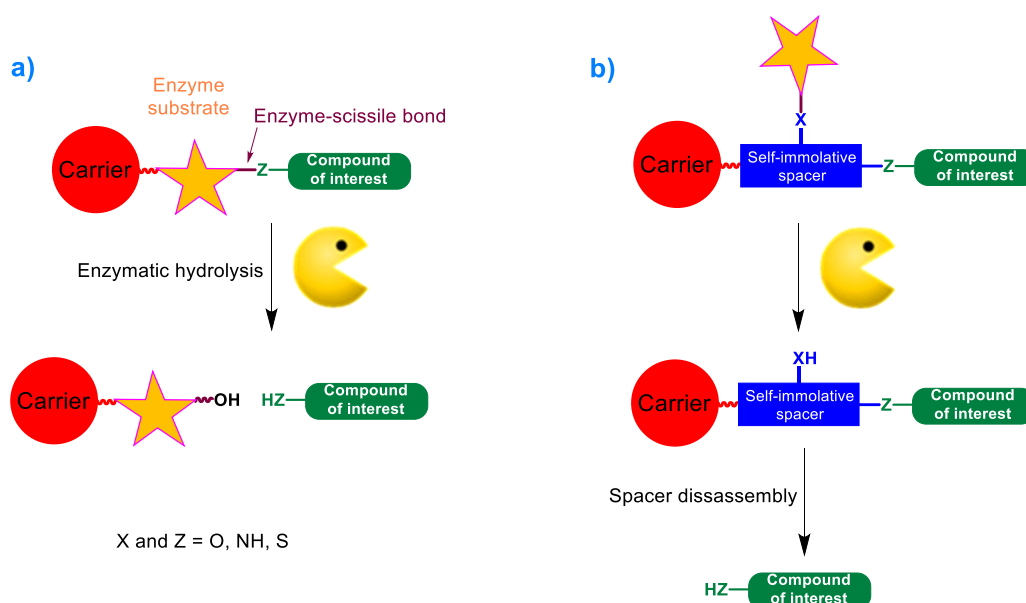


Figure 2: The two types of enzyme-cleavable linker

Self-immolative spacers function through two types of processes. In the first one, the release of the desired compound results from an electronic cascade, in the second one, it occurs through a cyclization.

The first class is generally based on (but not limited to) an aromatic structure bearing a phenol, [8–10], aniline, [11,12] or thiol [13] group substituted with an enzyme substrate (Fig. 3). The enzymatic hydrolysis frees electron-donating substituents and initiates a spontaneous and irreversible self-immolation through an 1,4- or 1,6-elimination processes, liberating the target compound in addition to the formation of a quinone, thioquinone or azaquinone methide intermediate, as well as CO₂ in the case of carbonates or carbamates. Quinone methides are electrophilic, and often further react with water. [14] In some cases, it can however react with a nucleophilic group of the compound of interest or the enzyme, thus being problematic. Environmental pH variations can lead to variation of kinetics of the cleavage, through protonation or deprotonation of the thiol, phenol or aniline group.

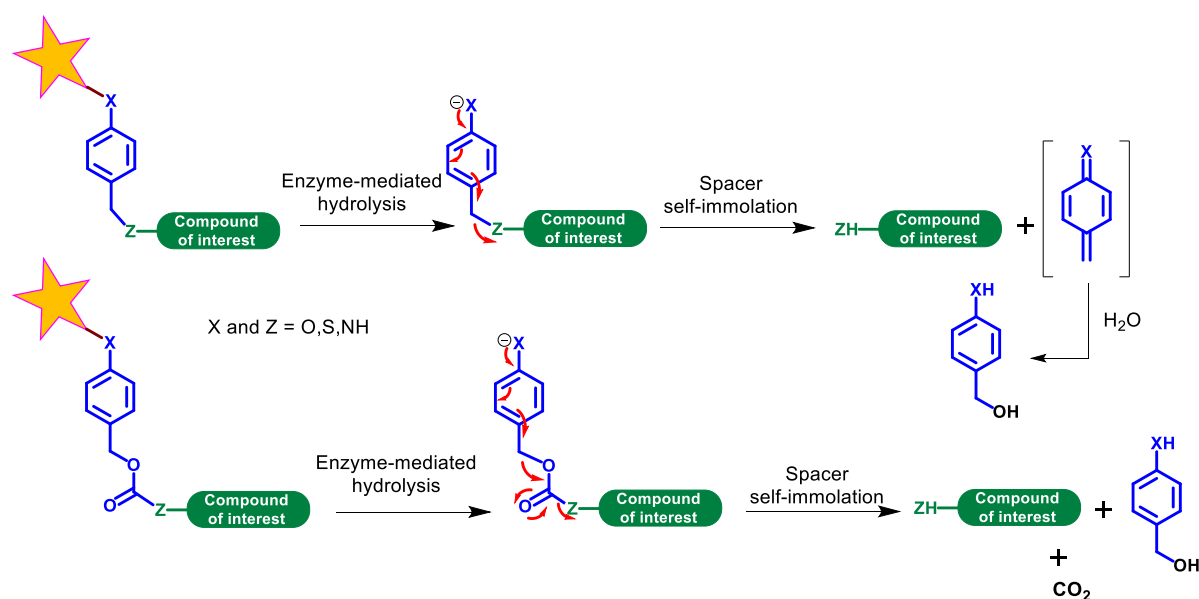
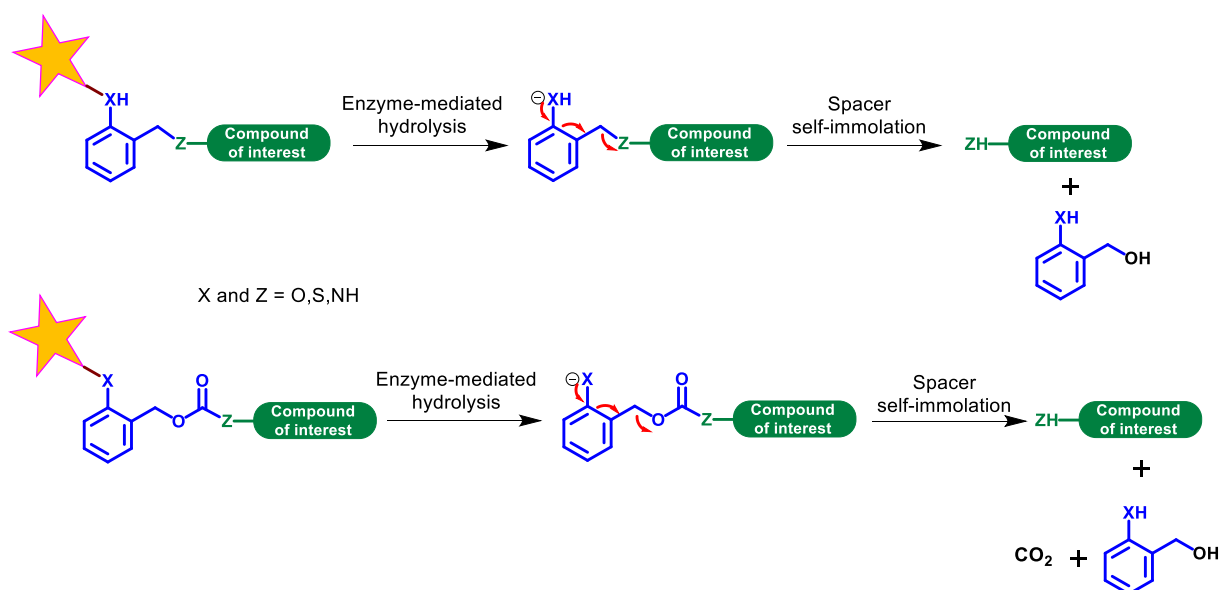
1,6-elimination**1,4-elimination**

Figure 3: Self-immolation through electronic cascades (carrier molecules omitted for clarity)

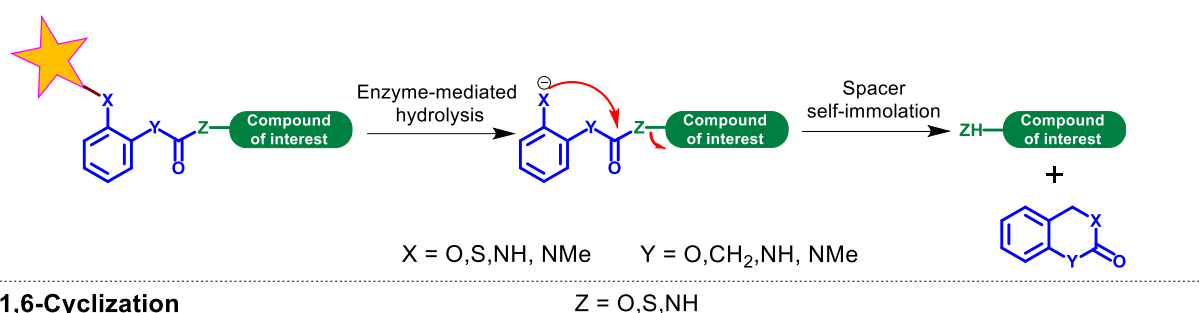
It is worth noting that several electronic cascades and spacers can be combined, in order to release multiple compounds of interest upon a single enzyme-mediated cleavage. [15]

In the second class of self-immolative spacers, a nucleophilic group protected by an enzyme-labile moiety is linked to a functional group of the molecule of interest through an alkyl chain [16–21] or an aromatic structure. [20,22–26]

Once the protection enzymatically removed, spacer disassembly occurs through a nucleophilic attack on a carbonyl group leading to the release of the desired compound along with the

formation of thermodynamically stable 5- and 6-membered rings such as imidazolidinones, oxazolidinones, or 1,3-oxathiolan-2-ones (Fig. 4). In most cases, such cyclizations show slower kinetics than self-immolation by electronic cascade. This is mainly due to its dependence on a series of proton transfers making it more influenced by the environmental pH. [27,28] However, the cleavage can be accelerated by the addition of substituents such as bulky aromatic or gem-dimethyl substituents (Thorpe-Ingold effect). [29]

1,5-Cyclization



1,6-Cyclization

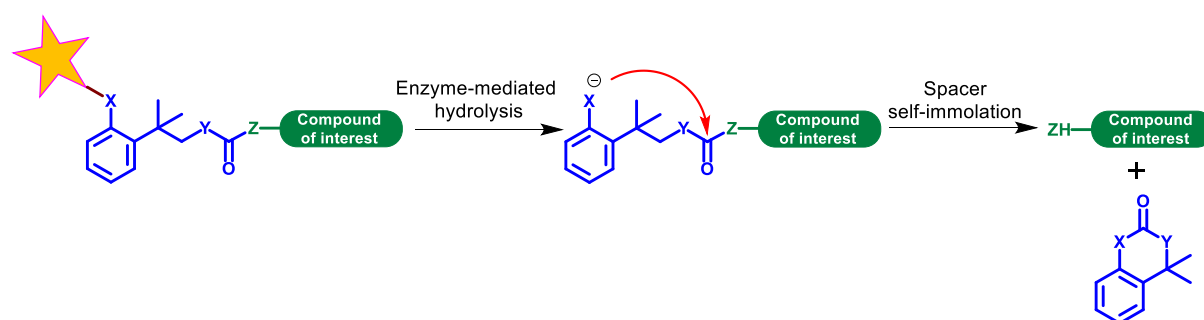


Figure 4: Self immolation through cyclisation (carrier molecules omitted for clarity)

3) Enzyme-cleavable linker applications

Enzyme-cleavable linkers find a number of applications in biochemistry and molecular biology, but this aspect will not be treated in details, as the focus on this thesis project is chemically-synthesized linkers. To mention only one widely used application, affinity purification of recombinant proteins makes use of specific non-covalent interactions between a peptide or a protein (the “affinity tag”) fused to the sequence of the protein of interest, and a solid support (one of the mostly used example being the binding of His₆ tag to Ni-NTA columns). Introduction of a “cleavage site” consisting in a short peptide sequence between the tag and the protein of interest allows removing the tag after purification. The most frequently employed proteases for cleavage are thrombin, factor Xa, enterokinase,

rhinovirus 3C protease (called PreScission protease) and the nuclear inclusion protease of tobacco etch virus (TEV protease). [30]

The flexibility in the design and chemical synthesis of enzyme-cleavable linkers has led to their application in many different disciplines including drug delivery, imaging and solid supported synthesis, which are overviewed in the next pages. Note that we have limited this bibliography study to actual linkers bridging two chemical entities, and did not mention applications of related enzyme-mediated “triggers” only linked to a compound of interest, but not to a carrier.

A. Targeted drug delivery

Targeted drug delivery systems aim to administer a pharmaceutical compound to a patient in a manner that increases the concentration of the medication in a specific part of the body (organ, solid tumor, type of cells) relative to others in order to achieve an optimal therapeutic effect when minimizing side-effects.

In this context, antibody-drug conjugates (ADCs) are now considered as a major class of therapeutics, especially for cancer treatment. [31] ADCs are hybrid molecules composed of antibodies and drug payloads, conjugated by chemically-synthesized linkers (fig. 5). Payloads are most often an extremely cytotoxic drug for anticancer therapy. The administration of such drugs not conjugated to a targeting unit is not suitable due to their high toxicity for healthy cells and organs. ADCs are usually prodrugs requiring the release of their toxic agent for its activation that commonly happens after ADC internalization into the target cells by endocytosis after antibody–antigen binding at the target cell surface.

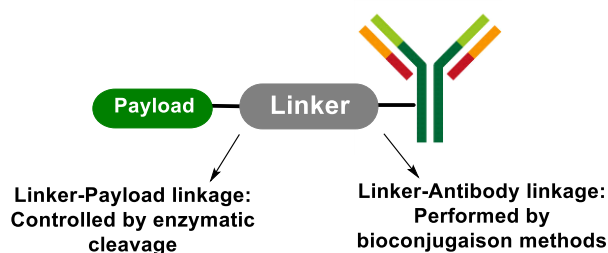


Figure 5: General structure of an ADC bearing an enzyme-cleavable linker

The properties of the linker between the antibody and the drug payload are proven to be critical to the success of an ADC, and enzyme-cleavable linkers are to date the most efficient and the most widely used ones. Since the first dipeptide-containing protease-cleavable linker developed in 1998, a cathepsin B-cleavable linker for doxorubicin (Dox) incorporating a self-

immolative para-amino benzyloxy carbamate spacer (fig. 6-a), [32,33] similar dipeptide linkers can be found in many ADCs that have reached clinical trials including brentuximab vedotin (trade name: Adcertis) approved by the U.S. Food and Drug Administration (FDA) in 2011 (fig.6-b) [34] and three other ADCs approved in 2019, all incorporating a valine-citrulline dipeptide linker cleaved by cathepsins.

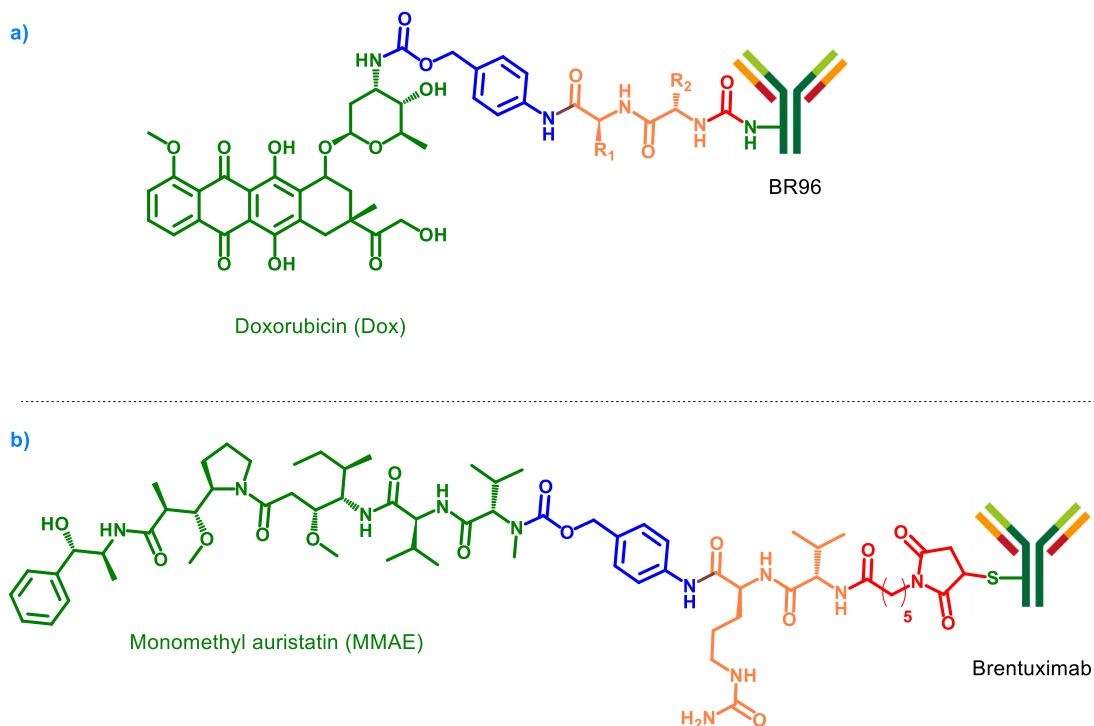


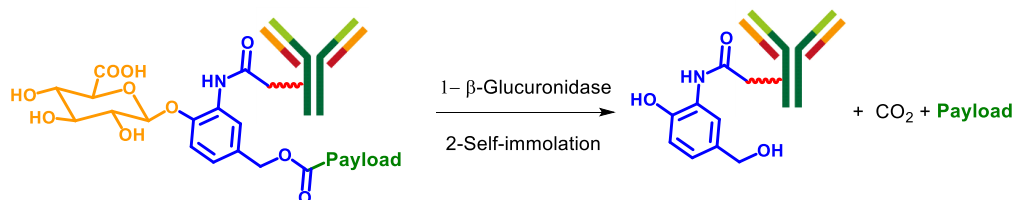
Figure 6: Structures of the immunoconjugate BR96-DOX (a) and brentuximab vedotin (Adcertis) (b)

The vast majority of research in the field of such dipeptide-based cleavable linkers-containing ADCs has focused on a mechanism of action where cleavage is performed after ADC internalization, in the lysosomes where a variety of enzymes are specifically sub-localized in cells, including many proteases such as cathepsin B [34–36]. However, such lysosomal enzymes can also be present extracellularly in a wide range of solid tumors (often through release upon cell necrosis), in these cases, extracellular cleavage around the tumor can be further beneficial to allow selective drug delivery near the tumor. [37,38]

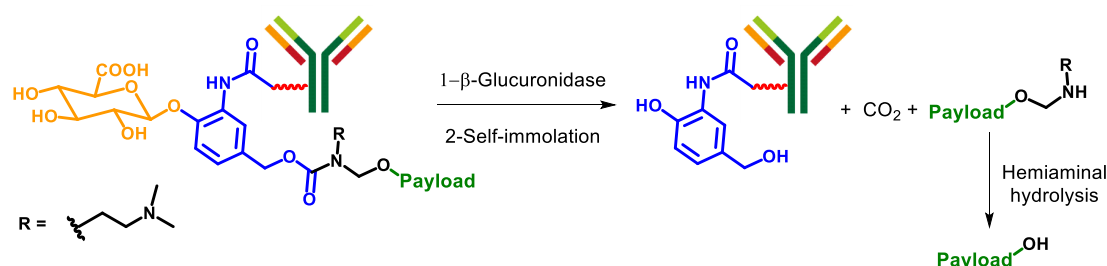
Numerous other lysosomal protease-cleavable dipeptide and tripeptide-based linkers [39] have been developed. Nevertheless, peptidic linkers have a number of drawbacks, including instability in rodent plasma in addition to high hydrophobicity. New type of ADC linkers have emerged (fig. 7) to overcome these limitations, like those cleaved by lysosomal glycosydases such as β -glucuronidase (fig. 7a-b) [40–42] and β -galactosidase (fig. 7b-c). [43] We can also

cite the work of Spring [44] who reported the use of sulfatase-cleavable linker which takes advantage of the stability, solubility and synthetic tractability of its arylsulfate structure (Fig. 7-d).

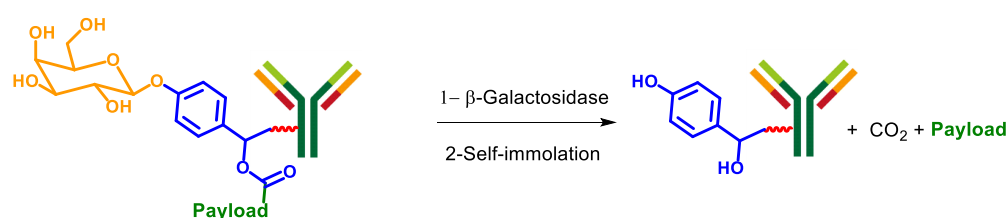
a)-Structure and cleavage of β -glucuronic acid containing ADCs



b)-Structure and cleavage of a methylene-alkoxy β -glucuronic acid containing ADCs



c)-Structure and cleavage of a β -galactose containing ADCs



d)-Structure and cleavage of a sulfate containing ADCs

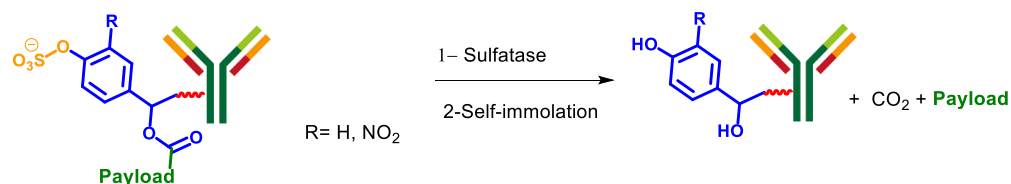


Figure 7: ADCs glycosidase and sulfatases based cleavable linkers

The presence of acid pyrophosphatase and acid phosphatase in the lysosome was also exploited (fig. 8) with the aim of developing new ADCs. [45]

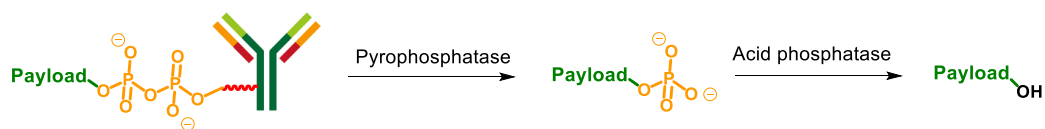


Figure 8: Structure and cleavage of pyrophosphate containing ADCs

Besides ADCs, other drug delivery systems based on enzyme-cleavable linkers have been developed where the cleavage is performed either intra- or extracellularly, depending on the nature of the carrier replacing the antibody in these systems. Numerous examples were reported in the literature, and just a few selected ones are given in the next section to illustrate the variety of these different drug delivery systems.

In an remarkable example of intracellular delivery, Shabat and his co-workers [15] developed a copolymers-drug conjugate with a self-immolative dendritic linker designed to release a triple payload of the paclitaxel drug as a result of cleavage by lysosomal cathepsin B upon internalization of the conjugate in cells. Here, the *N*-(2-hydroxypropyl)-methacrylamide (HPMA) copolymers act as a water-soluble carrier which increases the solubility of the drug-delivery system equipped with three hydrophobic paclitaxel molecules and enables specific delivery into tumor tissue (fig. 9). These macromolecules indeed do not diffuse through normal blood vessels but rather accumulate selectively in tumors due to the enhanced permeability and retention (EPR) effect. [46]

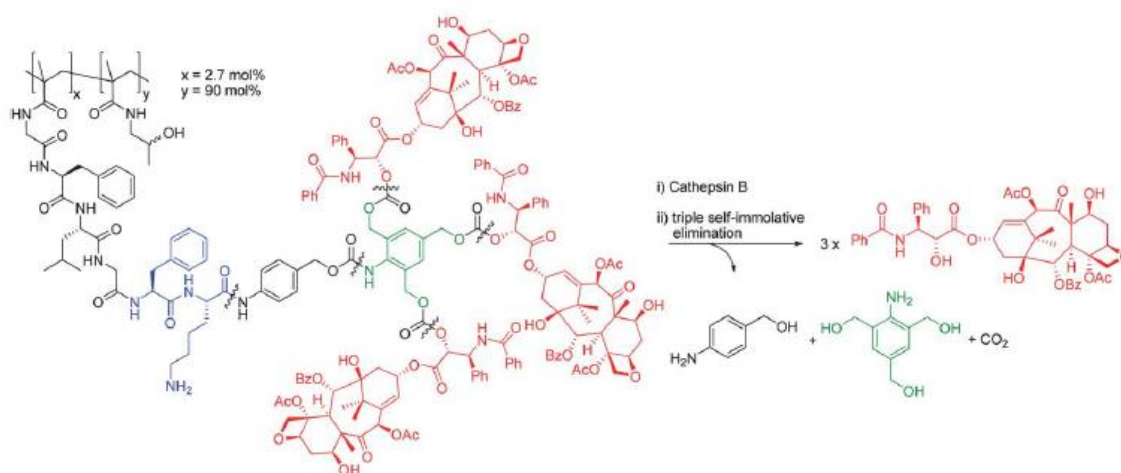


Figure 9: Cathepsin B-mediated self-immolative disassembly of HPMA – 2,4,6-tris(hydroxymethyl)aniline based drug delivery system (Figure taken from ref. [47])

The vast majority of targeted drug delivery systems (including ADCs) have been designed to recognize a specific cell surface marker allowing them to penetrate inside target cells to release their payload. However, the extent of such targeting devices is restricted to the treatment of tumors expressing a high level of the targeted cell surface marker. Therefore, extracellular drug delivery systems were also investigated, relying on enzymes present in the microenvironment of solid tumors.

As an illustration we can cite the work of Papot and his co-workers [48] who reported a drug delivery system (fig. 10) programmed for the selective release of the potent monomethylauristatin E (MMAE). This compound includes a glucuronide trigger, MMAE payload, and a self-immolative linker equipped with a poly(ethylene glycol) side chain terminated by a maleimide functional group. Upon administration, the *in vivo* reaction of this system to the thiol at the cysteine 34 position of the highly abundant plasmatic albumin *via* Michael addition produces the conjugated payload carrier that will transport the prodrug to targeted malignant tissues through EPR effect, where the cleavage of the glucuronide by extracellular β -glucuronidase triggers the release of MMAE.

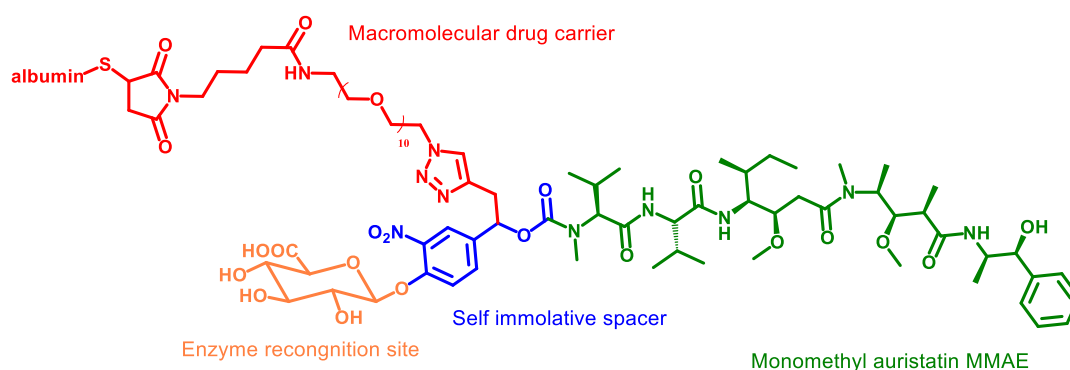


Figure 10: β -glucuronidase-responsive drug extracellular delivery system

When combining high bio-recognition capabilities of enzymes with the unique physical properties of nanomaterials, resulting enzyme-responsive nanoparticles can be designed to perform functions efficiently and with high specificity for the triggering stimulus. This powerful concept has been successfully applied to the fabrication of drug delivery where nanoparticles were covalently modified or non-covalently loaded with drugs through an enzyme-cleavable linker so that the tissue of interest is targeted *via* the cleavage of the linker by a specific enzyme.

An example was the strategy for enzyme-triggered release of drugs from mesoporous silica nanoparticle nanocarriers (MSNPs) that was proposed by Kim and co-workers. They designed a drug delivery system responsive toward two key-enzymes involved in the pathology of acute pancreatitis since the drug release could be triggered by both lipase and α -amylase enzymes. [49] The cyclodextrin covered MSNPs were functionalized at their surface with cyclodextrin “gatekeepers” that prevent the release of guest molecules from the porous reservoir. The enzymatic hydrolysis of either an ester linkage of the linker (lipases) or the cyclodextrin itself (amylase) remove these caps and unleashes the drugs (fig. 11).

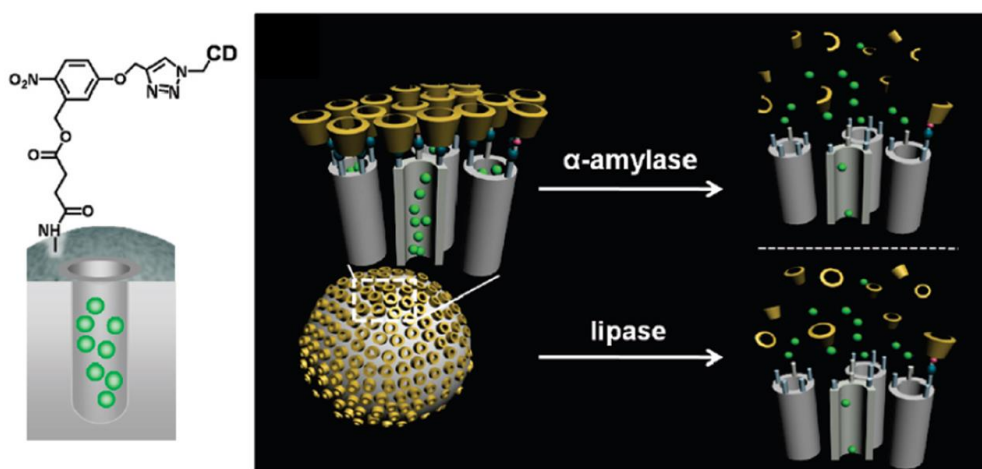


Figure 11: Mesoporous silica nanoparticle nanocarrier- based drug delivery system (Figure taken from ref. [49]). Green dots: guest drugs. CD and orange cylinders: cyclodextrin

B. Imaging

Among many potential cancer biomarkers, enzymes are gaining increased attention due to the fact that abnormal enzymatic activities are usually related to the development of cancers. In fact, many studies have shown that abnormal activities of some specific enzymes are related to a specific type on cancer. For instance, alkaline phosphatase is overexpressed in some bone cancers, [50] and high expressions levels of β -galactosidase are usually observed in primary ovarian cancers. [51]

Therefore, identifying the location and expression levels of these enzymes *in vivo* could be very crucial in early-stage cancer diagnoses and monitoring the efficacy of therapies. Several imaging techniques have been applied to detect or image enzymes, such as magnetic resonance imaging (MRI) and fluorescence imaging. Numerous examples have been reported

in the literature, and just a very few examples of enzyme-cleavable linker-based imaging tools will be mentioned in the next section to briefly illustrate this field of research.

MRI often uses contrast agents (generally based on gadolinium complexes) to increase the relaxivity of neighboring water molecules, resulting in an improvement of the visibility of internal body structures upon nuclear magnetic resonance detection. Since the 2010s, many “smart” MRI contrast agents containing enzyme-cleavable linkers were reported for the functional imaging of an enzyme activity. For example, they were used to enable the *in vivo* detection of MMP-2 and MMP-9 enzymes involved in malignancy and metastasis, based on a protease-cleavable linker. [52] In this work, the authors designed dendrimeric nanoparticles coated with “masked” cell penetrating peptides (CPPs), and labeled with the gadolinium complexes (Gd). Each nanoparticle consists of a dendrimer (fig. 11, gray circle) covalently attached to the polycationic segments (blue) of several CPPs. In tumors, MMP-2 and MMP-9 cleave the pentapeptide linkers (sequence: PLGLAG) (green), releasing polyanions (red) that were masking the cationic CPPs through electrostatic interactions, and thus mediate nanoparticle internalization into cells in the immediate vicinity of the protease. Once activated, the Gd-labeled nanoparticles deposit high levels (30–50 μM) of Gd in tumors, resulting in useful contrast lasting several days after injection. These nanoparticles can also be used as a multimodal imaging agent, for both MRI and fluorescence imaging (in fig. 12, yellow circles indicate imaging agents, which can be gadolinium, the cyanine5 fluorophore (Cy5), or both).

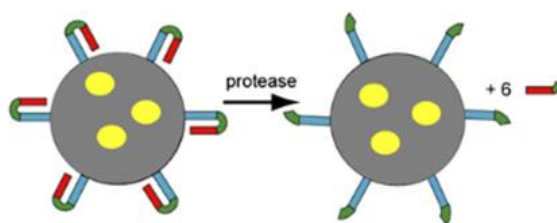


Figure 12: Protease-activatable CPP-coated nanoparticles for MRI and fluorescence imaging of tumors (figure taken from ref [52])

However, MRI generally lacks sensitivity, and enzyme-activatable fluorescence imaging is a widely used alternative approach which offers higher sensitivity, fast analysis, and real-time detection. An option is to use Förster resonance energy transfer (FRET) probes. [53,54] FRET occurs due to the interaction between the electronic excited states of two light-sensitive

molecules, a donor and an acceptor. This mechanism is distance-dependent meaning that the donor and the acceptor must be in close proximity. A donor molecule may transfer energy to an acceptor molecule through non-radiative dipole–dipole coupling, resulting in the fluorescence at acceptor emission wavelengths when the donor is sensitized at its excitation wavelength. Another option is the combination of a fluorescent donor with a quencher acceptor, not able to fluoresce, thus resulting in the turn-off of the fluorescent properties of the system.

The two strategies have been utilized to design enzyme-cleavable linkers-based FRET probes (fig- 13). In both cases, the two dyes are attached one to each other *via* an enzyme-cleavable linker.

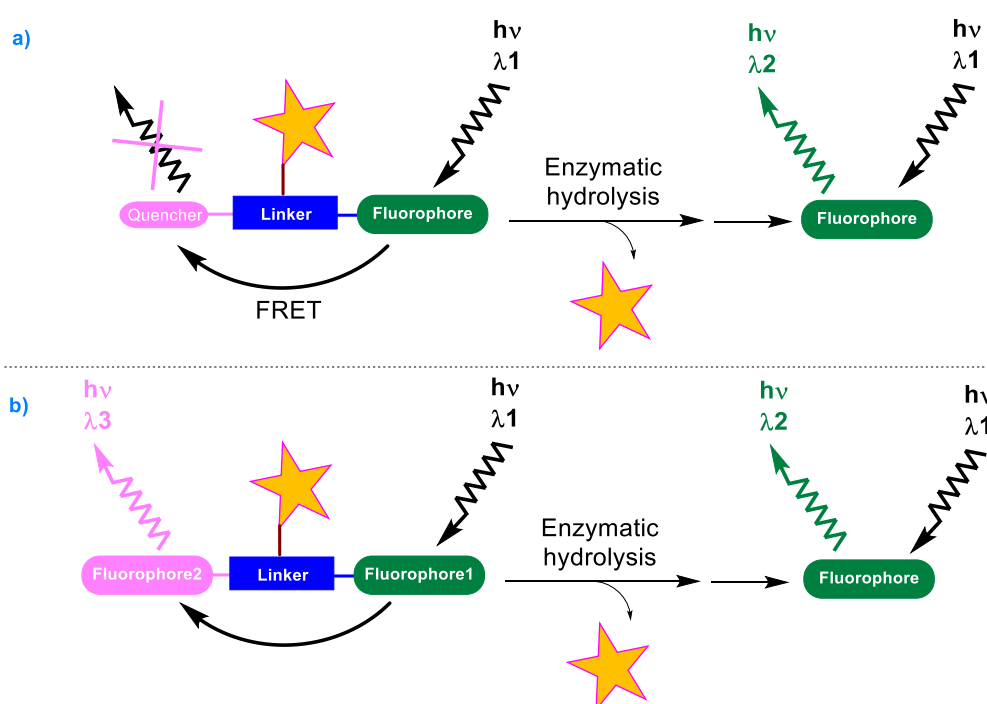


Figure 13: Mechanism of FRET-based fluorescent probes based on enzyme-cleavable linker

As an illustration among the numerous examples in literature, Cui and co-workers reported a FRET-based theranostic design using a fluorophore/quencher pair. This probe is composed of doxorubicin, which is playing both roles of drug payload and fluorophore, a quencher (black hole quencher 2, BHQ2), and a CPP covalently linked together *via* a peptidic linker (sequence: GFLG) (fig. 14-a). [55]

The probe is designed to stay at an OFF-state prior to cellular uptake. Following cellular entry mediated by CPP, lysosomal cathepsin B (CatB) cleaves the GFLG linker, and doxorubicin is released thus emitting fluorescence (ON-state) while acting as a therapeutic agent (fig. 14-b).

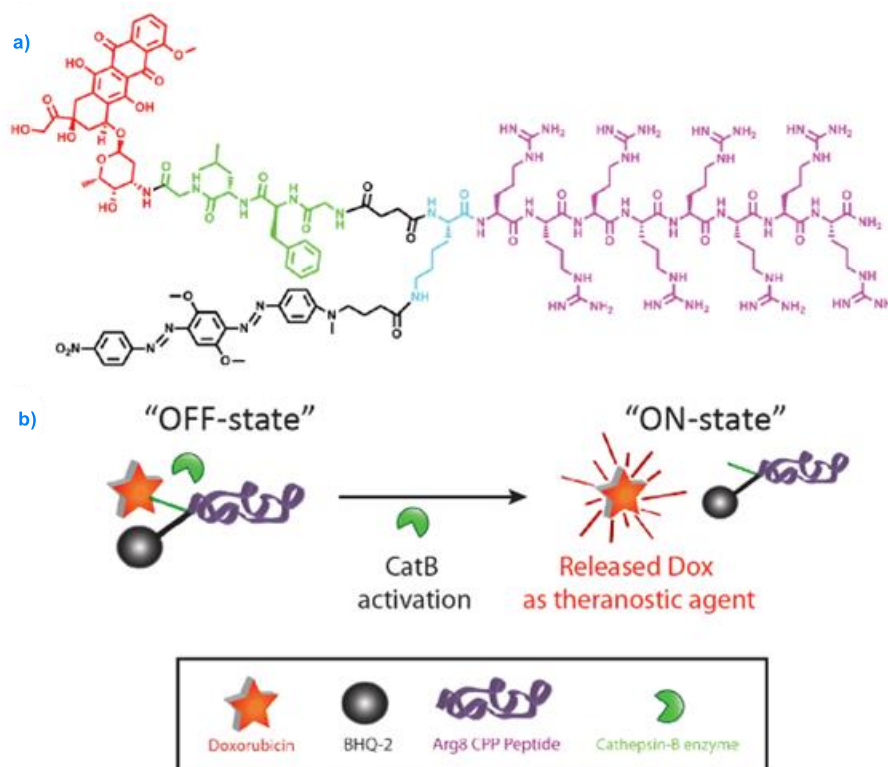


Figure 14: Mechanism of a FRET-based fluorescent probe based on an enzyme-cleavable linker (figure adapted from ref. [55])

The same dye molecule can also act both as a fluorophore and quencher (self-quenching). A good illustration of this approach is the work of Shabbat [56] who designed FRET-based probes equipped with a self-immolative linker that releases two identical Cy5 fluorophores upon enzymatic cleavage of the linker (phenylacetamide) by PGA (fig. 15).

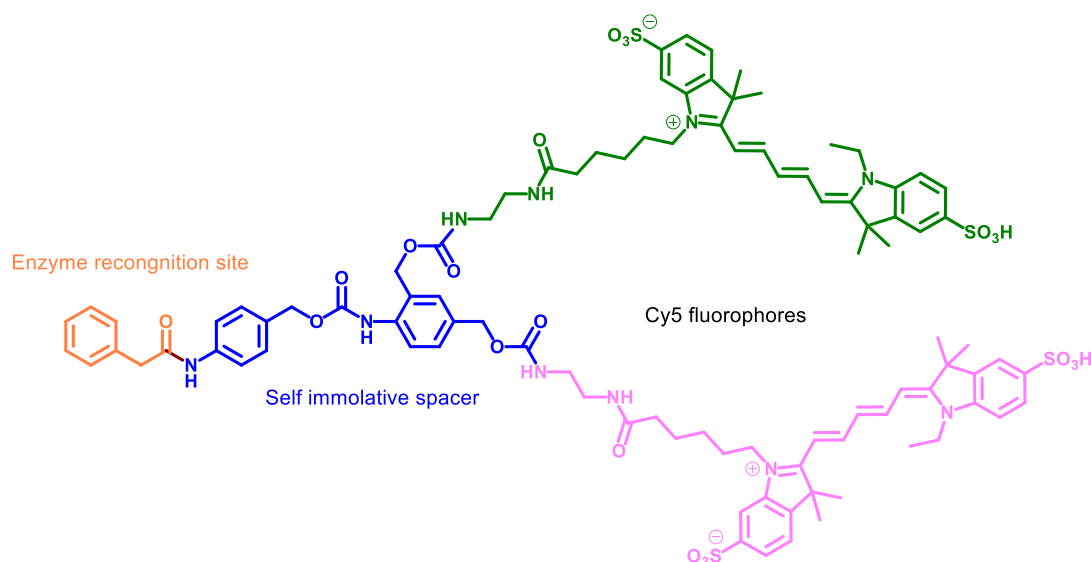


Figure 15: Structure of homo-FRET probe equipped with Cy5 fluorophores

C. Enzyme-labile linkers for solid phase synthesis

Since development of the concept of solid-supported synthesis by Merrifield half a century ago, [57] the use of solid supports in organic synthesis has been steadily increasing. The choice of the linker bridging the solid support and the compound to be synthesized is very crucial. A wide range of linkers have been developed for this purpose but their cleavage requires chemical conditions that proved to be incompatible with some sensitive targets [as it is the case for the synthesis of proteins through solid phase chemical ligation, see chapter 1]. Thus, an alternative solution based on enzymatic cleavage has been developed. Therefore, different enzyme-cleavable linkers have been investigated and applied for the synthesis of numerous different classes of compounds.

a) Solid support enzymatic accessibility

A crucial point that needs to be taken into account when performing any enzymatic reactions on solid support –including but not only limited to linker cleavage– is accessibility of the enzyme to its substrates inside the selected solid support.

There is evidence that very large enzymes are not able to penetrate solid supports and hence reactions are limited to bead surfaces and proceed in low overall yields. [96] Therefore, the nature and the properties of solid supports intended to achieve enzyme-mediated reactions should be thoughtfully selected. The most successfully used solid supports in enzymatic reactions are organic polymers [66] and inorganic rigid porous supports.[61]

Advantageous polymers need to show large swelling volumes in water thus providing better enzyme accessibility. Concerning porous rigid supports, they should have a large internal surface area and the pore diameter of the materials should effectively be chosen as large as possible so that there are no limitations on mass transfer through diffusion.

The solid support should also not contain any groups or counter-ions which can react with, strongly bind to, inhibit or denature the enzyme used for cleaving the linker. In the other hand, some surface treatment or polymer modification can result in increasing the rate of the enzymatic cleavage, as it was established for the introduction of permanent positive charges at the solid support surface probably by substantially increasing the swelling of the resin beads in the case of the reaction of PGA on polymer-based resins like the polyethylene glycol (PEG)-polyacrylamide copolymer (PEGA) resins [58] and by presenting the immobilized enzyme substrate in a more accessible microenvironment in the case of the reaction of α -chymotrypsin on control pore glass (CPG) beads. [59]

The utility of introducing a spacer between the support and the enzyme-cleavable linker have also been demonstrated, such as in the study by Halling and co-workers, who investigated the effect of adding an oligoglycine spacer on the enzymatic cleavage from different solid supports and showed that in the absence of the spacer chain, the rate of release was relatively slow. [59] This tells us that the manner of presentation of the immobilized substrate to the enzyme has a crucial role and will dictate the ability of the enzyme to perform a reaction on a solid support. In another example, Yamada and Nishimura showed the usefulness of a hexaglycyl phenylalanine spacer to enzymatically release synthesized oligosaccharides. [60]

Many studies about the enzyme compatibility of some commonly used supports were carried out. This allowed us to establish a ranking of the best suitable solid support for enzymatic reactions:

The water-compatible PEG-grafted polystyrene (PS) TentaGel resin was demonstrated to be incompatible with enzymatic cleavage. For example, in a systematic study of many supports and enzymes, Bradley and co-workers [61] found that none of the investigated enzymes (MMP-12 (22 kDa), thermolysin (35 kDa), MMP-13 (42.5 kDa), clostridium collagenase (68 kDa), and NEP (90 kDa)) could enter the polymer matrix of TentaGel. Another example was the illustration of the low efficiency of the protease papain (23 kDa) with Tentagel by Lowe

and his co-workers where a model peptide was hydrolysed in very low yields. [62] A similar work of Waldmann showed that no cleavage at all was obtained when TentaGel was used with PGA (80kDa). [63] TentaGel was proven to be also inefficient with a 33 kDa lipase by Smith and Diederich, [64] and with PFTase (80–90 kDa) by Wang and Distefano. [65]

Other water-compatible PS-PEG solid support related to TentaGel such as ArgoGel, [62] were also proven to be incompatible with enzymatic reactions and efforts aimed at finding a polymeric solid support more hydrophilic and highly accessible by enzymes have focused on PEGA resins. This class of support developed by Meldal and co-workers is characterized by the size of the PEG used to crosslink polyacrylamide chains: PEGA⁸⁰⁰ (800 Da PEG) has been optimized for solid phase peptide synthesis, whereas PEGA¹⁹⁰⁰ (1900 Da PEG) has been optimized for solid-supported enzymatic reactions. PEGA was proven to be compatible with papain (23 kDa) [62] as well as MMP-12 (22 kDa) and thermolysin (35 kDa). [61] The same conclusion was obtained with 33 kDa lipases by Hailes and his co-workers. [66]

However, when it came to larger enzymes, PEGA solid support have shown their limitations. This showed that the relationship between the molecular weight of the enzyme and the type of the solid support plays a key role in the yield of the cleavage reaction. For example, PGA (80kDa) which was evaluated by Waldmann and Grether did not have access to the interior of the polymer matrix yielding less than 15% cleavage of model compound with PEGA¹⁹⁰⁰. [63] Within the same context, PFTase enzyme (80–90 kDa) was reported to be incompatible with PEGA resins. [65]

Therefore, to circumvent the limitations of polymeric solid supports, studies focused on the use of inorganic porous rigid supports like CPG which form porous beads with pores of 75 to 3000 Å in diameter and which has been widely used in the field of enzyme immobilization since the first time in 1976. [67]

For instance, Bradley and co-workers [61] obtained conclusive results when enzymes up to 90kDa using were used on CPG1000 (1000 Å pores) supports thus demonstrating that the diffusion of enzymes inside the solid support strongly depends on their size. In another example, Halling and co-workers [59] reported that the pore diameter of CPG solid support has an effect on the levels of enzyme accessibility, with the number of enzyme accessible

molecules per nm² surface increasing by a factor of 16 when the average pore diameter of CPG is increased from 500 to 2900 Å.

b) Solid phase synthesis of peptides, glycopeptides and peptide nucleic acid

In the next section, we will give some representative examples of enzymatically-cleavable linkers used for the solid phase synthesis of peptides, glycopeptides and peptide nucleic acids (PNAs).

Protease-cleavable linkers have been frequently used. For example, α -chymotrypsin was used by Wong and co-workers to release glycopeptides from a solid support by selectively cleaving a Phe-Xaa bond (fig. 16). [68]

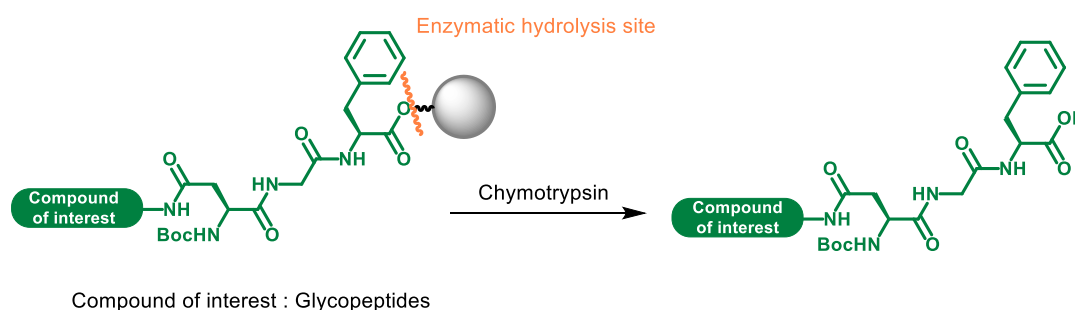


Figure 16: Chymotrypsin-based enzymatic cleavage for the synthesis of glycopeptides

Hoheisel used a polymer membrane as a solid support in the synthesis of various PNA oligomers where a trypsin-sensitive Glu-Lys linker was introduced between the membrane and the PNA and used to release PNA in very mild conditions (fig. 17). [69]

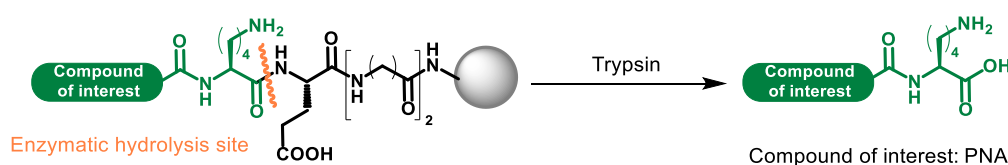


Figure 17: Trypsin based enzymatic cleavage for the synthesis of peptide nucleic acids

Chymotrypsin was also used by Flitsch and co-workers [70] to cleave a hydroxymethylphenoxy linker (HMPA) and to release peptides and glycopeptides from solid support (fig. 18-a). This linker was surprisingly found to be susceptible to efficient cleavage by chymotrypsin and to not being limited to peptides containing C-terminal aromatic amino acids normally associated with chymotrypsin recognition motifs. It is important to mention that the hydroxymethylbenzoic acid linker (HMBA) as well as the hydroxyoctanoic acid linker (HOA) were also readily cleaved by the enzyme (fig. 18-b).

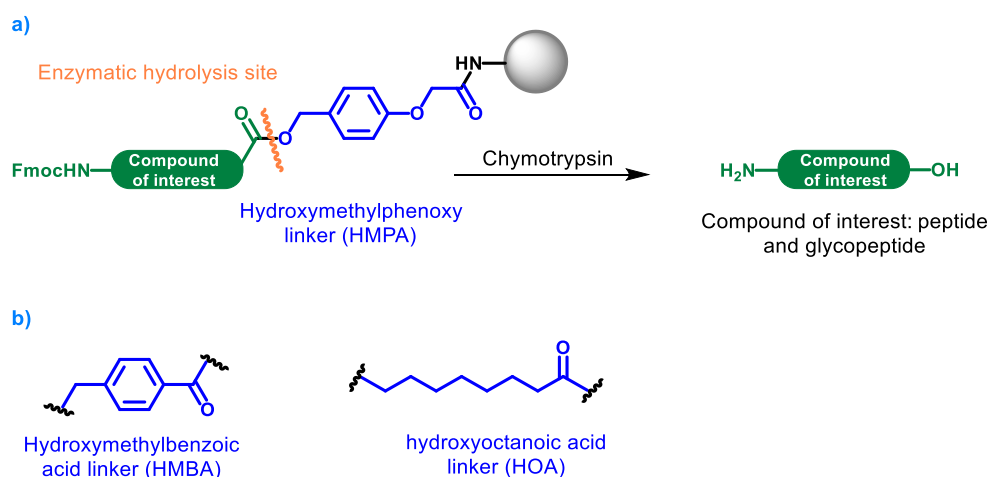


Figure 18: a) HMPA based enzyme-cleavable linker for the synthesis of peptides and glycopeptides

b) Structures of HMBA and HOA linkers

In 1992, Bates and co-workers described a linker containing a phosphodiester group for solid-phase peptide synthesis (fig. 19). The later was cleaved by calf spleen phosphodiesterase to release a collagenase substrate. [71]

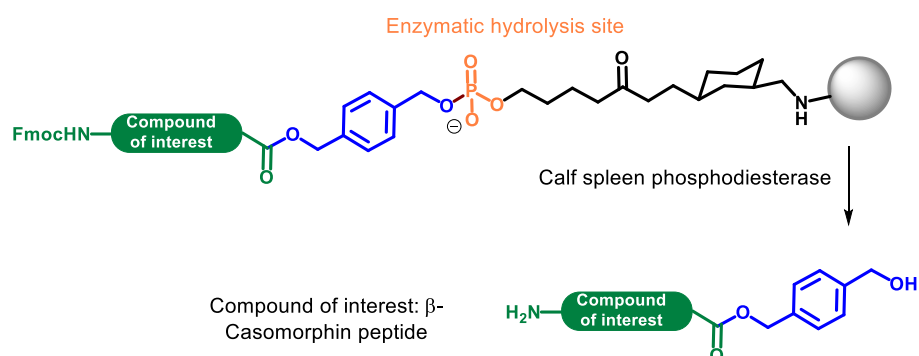


Figure 19: Phosphodiester based enzyme-cleavable linker for the synthesis of peptides

c) Solid phase synthesis of oligosaccharides

Enzyme-assisted approach for the synthesis of oligosaccharides is recognized as one of the most appealing practical alternatives to chemical synthesis mainly due to its highly selective reactions with no tedious protection/deprotection steps. There has been a concerted effort to combine this concept with solid phase synthesis, involving in some cases enzymatic cleavage from the solid support.

As an illustration, Nishimura and Yamada [60,72] synthesized a trisaccharide on a water-soluble solid support by an enzymatic transformation with galactosyl- and sialyl-transferases. An L-phenylalanine residue was used as a-chymotrypsin-sensitive structure to release the

saccharide from the solid support. We can also cite the similar work of Wong and Schuster [68] who used a α -chymotrypsin-sensitive linker for the solid-phase enzymatic synthesis of oligosaccharides (fig. 20).

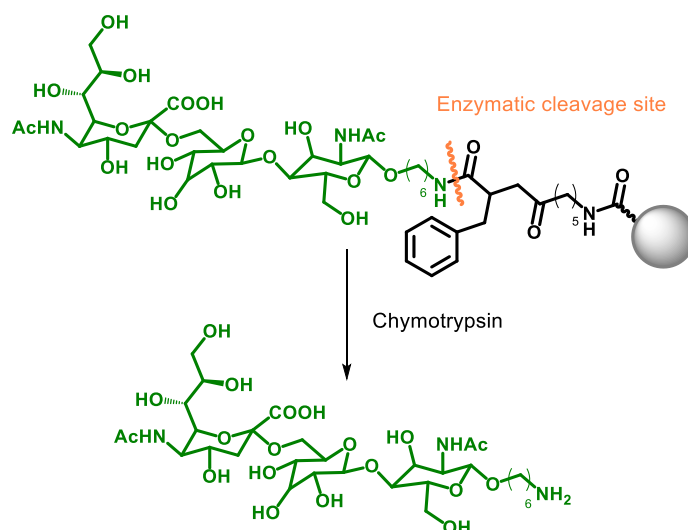


Figure 20: Example of Chymotrypsin-catalyzed release of oligosaccharides from a solid support

Within the same context, solid-phase enzymatic synthesis of carbohydrates derivatives was also reported: Clark and co-workers [73] reported the synthesis of bergenin-derived C-glycosides that was performed on a controlled pore glass solid support and where the release of compound of interests was accomplished using both chymotrypsin and lipase B (LipB) with better cleavage yields obtained with lipB (fig. 21).

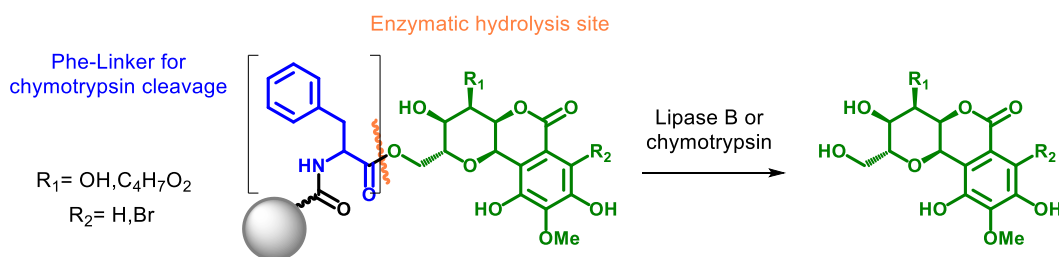


Figure 21: Enzymatic release of derived bergenin glycosides

As another related example, we can mention the use of ceramide glycanase by Nishimura and Yamada for the synthesis of the lipo-trisaccharide ganglioside GM3 (fig. 22). [74]

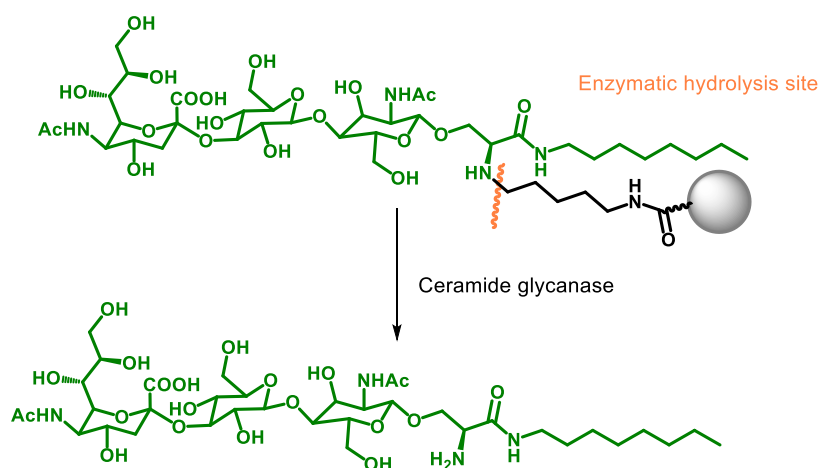


Figure 22: Enzymatic release of GM3

d) Solid phase synthesis of organic molecules

The field of small molecules organic synthesis has also benefited greatly from the development of enzyme-cleavable linkers. The next section gives some representative examples of their use.

Lipases and esterases [75] were used to fragment the 4-acetoxybenzyloxy group used as a linker for the synthesis of organic molecules on solid supports. This enzyme-initiated fragmentation proceeds under very mild conditions to release chemical libraries of various amines, alcohols or carboxylic acids constructed by combinatorial chemistry (fig. 23).

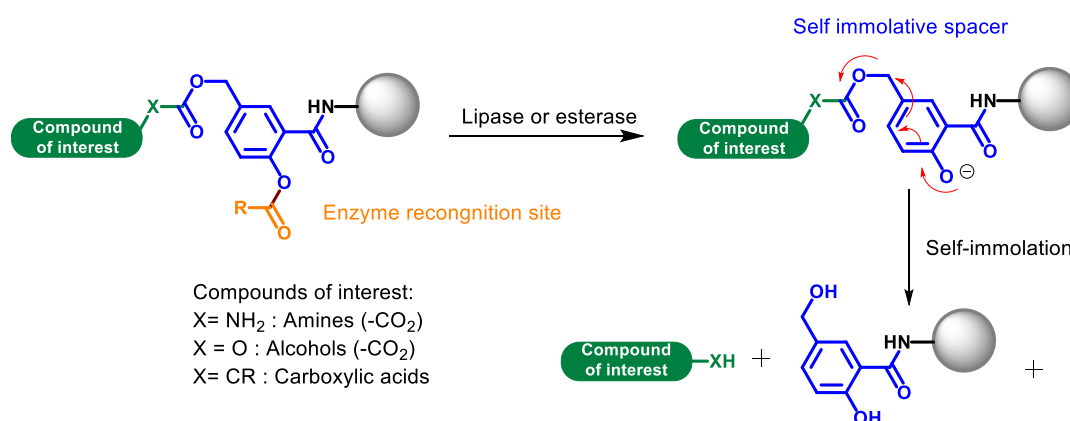


Figure 23: Enzymatic cleavage of 4-acetoxybenzyloxy-based linkers developed for combinatorial chemistry

Furthermore, a PGA-cleavable linker was developed by Grether and Waldmann for the release of alcohols and amines through enzymatic cleavage of an amino group and subsequent lactam formation (fig. 24). [24,63]

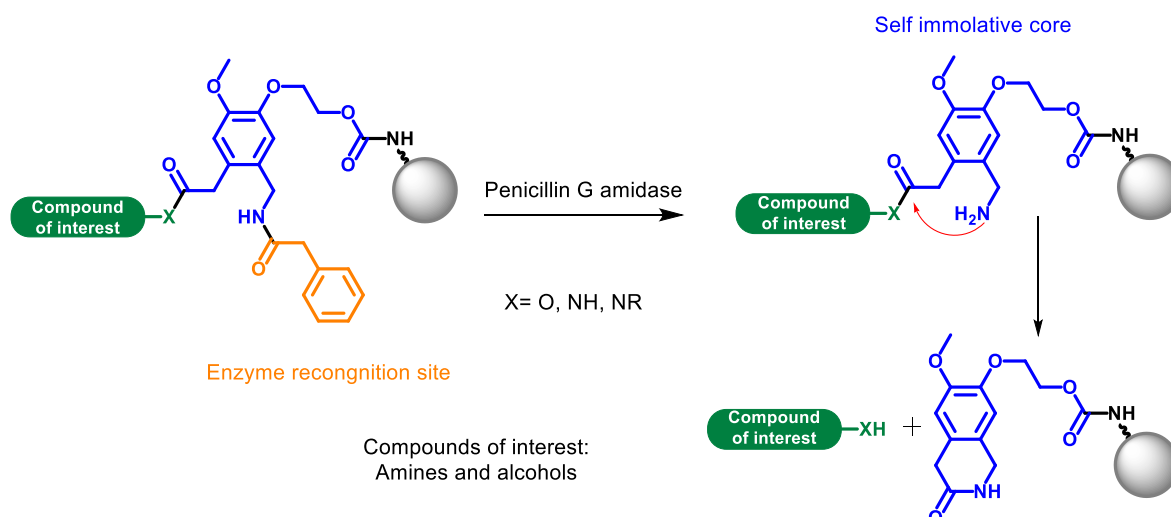


Figure 24: Waldmann's PGA-labile linker cleavage

Flitsch and Turner [76] have developed another PGA-labile linker for the synthesis of various molecules (fig. 25).

Enzymatic hydrolysis site

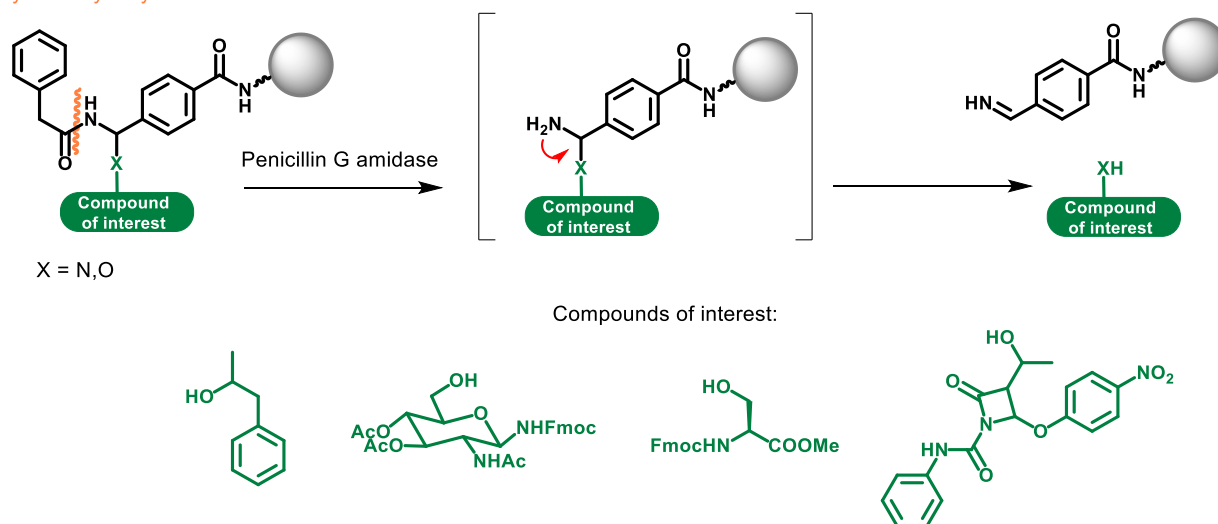


Figure 25: Flitsch-Turner's PGA-labile linker cleavage

II. Exploration of the use of enzyme-cleavable linkers for SPCL

1) Limits of chemically-cleavable linkers in SPCL

The total synthesis of long proteins requires the assembly of multiple segments through successive ligations. The need of numerous intermediate purifications is a strong limitation, as this generally leads to very low overall yield.

As detailed in the [chapter 1](#), particular emphasis has been placed on the development of techniques allowing a solid support assembly in order to avoid intermediate purifications. Although extremely appealing, solid phase chemical ligation (SPCL) has been so far essentially limited to proof-of-concepts, one of the main reasons being the difficulty to immobilize a first peptide segment on a water-compatible solid support through a dedicated SPCL linker. The main challenge is to control the conditions demanded for the cleavage of these linkers in order to release from the solid support the protein assembled by successive ligations.

The choice of the linker is a keystone in SPCL strategies: it should be stable along the protein assembly then be efficiently cleaved without affecting the chemical integrity of the synthesized protein. An extensive range of SPCL linkers have been developed, introduced on either the C- [77–85] or N-terminal [77,86–90] parts of the first peptide segment depending on the directional strategy chosen for the successive ligations: from the C- to the N-terminus of the target protein (C-to-N), or N-to-C. Cleavage conditions include treatments with acids, [78–81,90] bases, [77,82,86,87] and nucleophiles. [88,89] However, such conditions are incompatible with many sensitive targets, and the development of alternative linkers represents a major challenge to expand the applicability of SPCL and thus further push away the limits of chemical protein synthesis.

2) Thesis project aim

The main objective of this thesis work was the development of a new generation of linkers by exploring an enzymatic rather than purely chemical approach. We intended to induce a rapid and effective cleavage under extremely mild and selective conditions (fig. 26).

Besides considerations on the chemical integrity of the target protein upon linker cleavage, enzymatic conditions are also non-denaturant, which represents a significant additional potential advantage, to exploit the “pseudo-dilution” effect [91] for the folding of the immobilized synthesized target, while avoiding intermolecular aggregation.

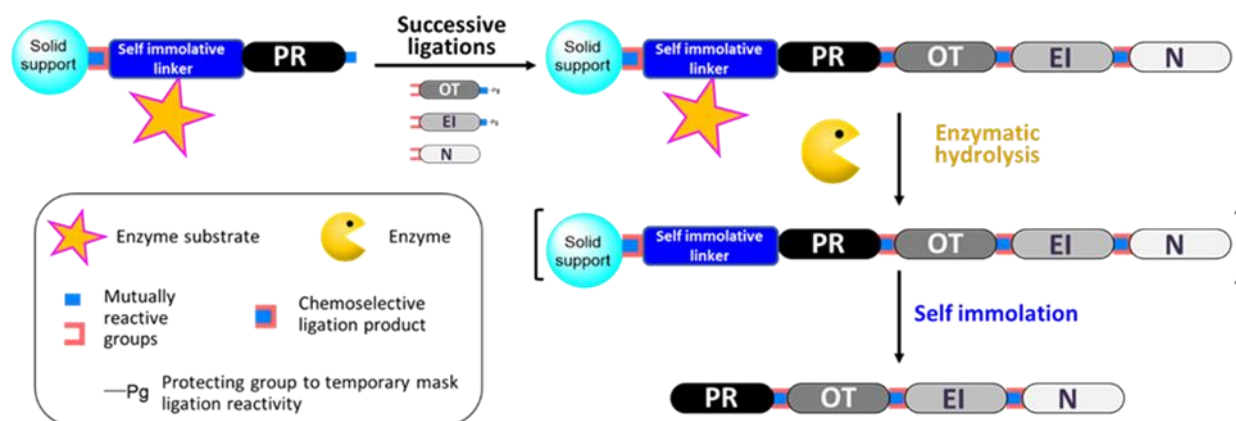


Figure 26: General approach for the application of enzyme-cleavable linkers in SPCL

Linkers selected for this work had to meet the following requirements:

- Their cleavage should free an amine moiety since we chose to conduct protein elongation in an N-to-C direction. This choice was motivated by the fact that this approach provides two major advantage over the alternative C-to-N direction: it allows the re-immobilization of the purified peptide on the SPCL solid support of our choice in addition to the self-purification of the incoming peptide segments (see section “SPCL with concomitant purifications” in [chapter 1](#)).
- They should not be dependent on the nature of the N-terminal amino acid, meaning that they should bear a self-immolative core that will be removed regardless of the nature of the neighboring amino acid.
- They should be equipped with an enzyme substrate compatible with a wide range of target proteins.
- The linker should allow the immobilization of deprotected peptide segments on solid support through a chemoselective ligation.
- Finally, the linker should be compatible with ligation reactions used for the assembly of the protein.

3) Enzyme-cleavable linkers investigated in this work

The first generation linker precursor used in this work contains four different functionalities: The central linker spacer or core, the enzyme substrate site (a β -galactoside) which is protected with acetate moieties to avoid side reactions during SPPS, a leaving group to incorporate the linker on the peptide and a reactive group (alkyne) for chemoselectively grafting the product on the solid support through a spacer (fig. 27).

The central linker core consists of an aromatic structure, with a *p*-substituted leaving group. Para substitution was selected since some studies [92] showed that the *p*-substituted linker eliminates rapidly while the *o*-substituted decomposition could require up to several hours after enzymatic hydrolysis. This enzymatic cleavage initially results in a phenolate which spontaneously decomposes into a quinone derivate. It's worth noting that since the first generation linker was developed by our collaborator Prof. Sébastien Papot [43] for targeted drug delivery, the linker core is bearing a nitro group because of previous studies reporting the advantage of prodrugs bearing chloro, fluoro, methoxy and nitro substituents in terms of lower toxicity of the prodrug, rapid self-elimination and plasma stability. [93–95]

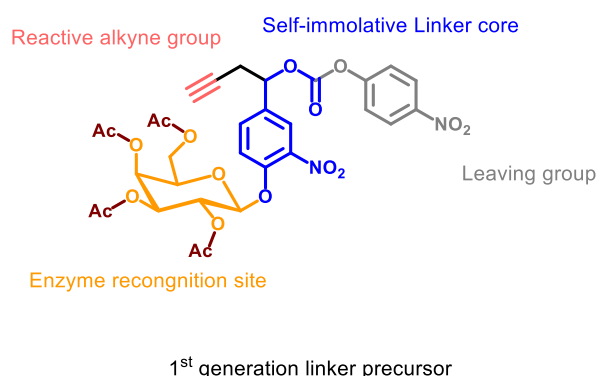


Figure 27: Structure of the 1st generation enzyme-cleavable linker precursor

This linker is grafted into solid support using the Copper (I)-catalyzed azide-alkyne cycloaddition (CuAAC) reaction. In order to increase the distance between the enzyme recognition site and the solid support, a small PEG spacer was introduced. As discussed earlier, such spacers are known to substantially preclude steric impairment of the enzymatic reaction by the substrate and thus increases the rate of the enzymatic reaction.

However, as detailed in the following paper, this first generation linker had many limitations with the main limitation being the slow enzymatic solid support cleavage. Hence, we developed a second generation linker designed to address these limitations. This linker was synthesized in 4 steps and 39% yield (fig. 28). It was grafted into solid support using the Strain Promoted Azide Alkyne Cycloaddition (SPAAC) reaction and phosphatases were used to cleave this linker.

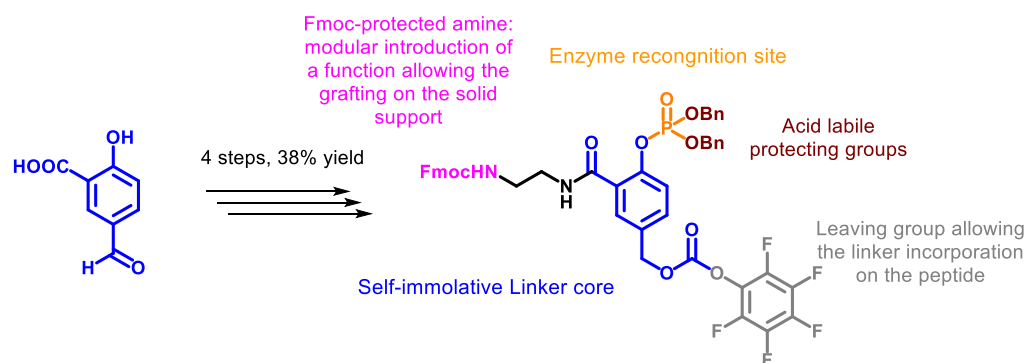


Figure 28: Structure of the 2nd generation enzyme-cleavable linker precursor

III. References for the bibliographic study part

1. Kadereit, D., and Waldmann, H. (2001) Enzymatic Protecting Group Techniques. *Chem. Rev.*, **101** (11), 3367–3396.
2. Pathak, T., and Waldmann, H. (1998) Enzymes and protecting group chemistry. *Curr Opin Chem Biol*, **2** (1), 112–120.
3. Tulla-Puche, J., Góngora-Benítez, M., Bayó-Puxan, N., Francesch, A.M., Cuevas, C., and Albericio, F. (2013) Enzyme-Labile Protecting Groups for the Synthesis of Natural Products: Solid-Phase Synthesis of Thiocoraline. *Angew. Chem. Int. Ed.*, **52** (22), 5726–5730.
4. Royo, M., Alsina, J., Giralt, E., Slomczynska, U., and Albericio, F. (1995) S-Phenylacetamidomethyl (Phacm): an orthogonal cysteine protecting group for Boc and Fmoc solid-phase peptide synthesis strategies. *J. Chem. Soc., Perkin Trans. 1*, (9), 1095.
5. Glass, J.D., and Pande, C.S. (1983) In *Peptides: Structure and Function*.
6. Chakravarty, P.K., Carl, P.L., Weber, M.J., and Katzenellenbogen, J.A. (1983) Plasmin-activated prodrugs for cancer chemotherapy. 2. Synthesis and biological activity of peptidyl derivatives of doxorubicin. *J. Med. Chem.*, **26** (5), 638–644.
7. de Groot, F.M.H., de Bart, A.C.W., Verheijen, J.H., and Scheeren, H.W. (1999) Synthesis and Biological Evaluation of Novel Prodrugs of Anthracyclines for Selective Activation by the Tumor-Associated Protease Plasmin. *J. Med. Chem.*, **42** (25), 5277–5283.
8. Alouane, A., Labruère, R., Silvestre, K.J., Le Saux, T., Schmidt, F., and Jullien, L. (2014) Disassembly Kinetics of Quinone-Methide-Based Self-Immolative Spacers that Contain Aromatic Nitrogen Heterocycles. *Chem. Asian J.*, **9** (5), 1334–1340.
9. Lee, H.Y., Jiang, X., and Lee, D. (2009) Kinetics of Self-Immolation: Faster Signal Relay over a Longer Linear Distance? *Org. Lett.*, **11** (10), 2065–2068.
10. Alouane, A., Labruère, R., Le Saux, T., Aujard, I., Dubruille, S., Schmidt, F., and Jullien, L. (2013) Light Activation for the Versatile and Accurate Kinetic Analysis of Disassembly of Self-Immolative Spacers. *Chem. Eur. J.*, **19** (35), 11717–11724.
11. de Groot, F.M.H., Albrecht, C., Koekkoek, R., Beusker, P.H., and Scheeren, H.W. (2003) “Cascade-Release Dendrimers” Liberate All End Groups upon a Single Triggering Event in the Dendritic Core. *Angew. Chem. Int. Ed.*, **42** (37), 4490–4494.

12. Erez, R., and Shabat, D. (2008) The azaquinone-methide elimination: comparison study of 1,6- and 1,4-eliminations under physiological conditions. *Org. Biomol. Chem.*, **6** (15), 2669.
13. Senter, P.D., Pearce, W.E., and Greenfield, R.S. (1990) Development of a drug-release strategy based on the reductive fragmentation of benzyl carbamate disulfides. *J. Org. Chem.*, **55** (9), 2975–2978.
14. Bosslet, K., Czech, J., and Hoffmann, D. (1994) Tumor-selective prodrug activation by fusion protein-mediated catalysis. *Cancer Res.*, **54** (8), 2151–2159.
15. Erez, R., Segal, E., Miller, K., Satchi-Fainaro, R., and Shabat, D. (2009) Enhanced cytotoxicity of a polymer–drug conjugate with triple payload of paclitaxel. *Bioorganic & Medicinal Chemistry*, **17** (13), 4327–4335.
16. Shan, D., Nicolaou, M.G., Borchardt, R.T., and Wang, B. (1997) Prodrug Strategies Based on Intramolecular Cyclization Reactions. *J. Pharm. Sci.*, **86** (7), 765–767.
17. DeWit, M.A., and Gillies, E.R. (2009) A Cascade Biodegradable Polymer Based on Alternating Cyclization and Elimination Reactions. *J. Am. Chem. Soc.*, **131** (51), 18327–18334.
18. DeWit, M.A., and Gillies, E.R. (2011) Design, synthesis, and cyclization of 4-aminobutyric acid derivatives: potential candidates as self-immolative spacers. *Org. Biomol. Chem.*, **9** (6), 1846.
19. Dewit, M.A., Beaton, A., and Gillies, E.R. (2010) A reduction sensitive cascade biodegradable linear polymer. *J. Polym. Sci. A Polym. Chem.*, **48** (18), 3977–3985.
20. Greenwald, R.B., Choe, Y.H., Conover, C.D., Shum, K., Wu, D., and Royzen, M. (2000) Drug Delivery Systems Based on Trimethyl Lock Lactonization: Poly(ethylene glycol) Prodrugs of Amino-Containing Compounds. *J. Med. Chem.*, **43** (3), 475–487.
21. De Gracia Lux, C., McFearn, C.L., Joshi-Barr, S., Sankaranarayanan, J., Fomina, N., and Almutairi, A. (2012) Single UV or Near IR Triggering Event Leads to Polymer Degradation into Small Molecules. *ACS Macro Lett.*, **1** (7), 922–926.
22. Milstien, S., and Cohen, L.A. (1972) Stereopopulation control. I. Rate enhancement in the lactonizations of o-hydroxyhydrocinnamic acids. *J. Am. Chem. Soc.*, **94** (26), 9158–9165.
23. Cain, B.F. (1976) 2-Acyloxymethylbenzoic acids. Novel amine protective functions providing amides with the lability of esters. *J. Org. Chem.*, **41** (11), 2029–2031.

24. Grether, U., and Waldmann, H. (2000) An Enzyme-Labile Safety Catch Linker for Combinatorial Synthesis on a Soluble Polymeric Support This work was supported by the Bundesministerium für Bildung und Forschung and BASF AG. *Angew. Chem. Int. Ed. Engl.*, **39** (9), 1629–1632.
25. Zheng, A., Shan, D., and Wang, B. (1999) A Redox-Sensitive Resin Linker for the Solid Phase Synthesis of C- Terminal Modified Peptides. *J. Org. Chem.*, **64** (1), 156–161.
26. Zheng, A., Shan, D., Shi, X., and Wang, B. (1999) A Novel Resin Linker for Solid-Phase Peptide Synthesis Which Can Be Cleaved Using Two Sequential Mild Reactions. *J. Org. Chem.*, **64** (20), 7459–7466.
27. Saari, W.S., Schwering, J.E., Lyle, P.A., Smith, S.J., and Engelhardt, E.L. (1990) Cyclization-activated prodrugs. Basic carbamates of 4-hydroxyanisole. *J. Med. Chem.*, **33** (1), 97–101.
28. Saari, W.S., Schwering, J.E., Lyle, P.A., Smith, S.J., and Engelhardt, E.L. (1990) Cyclization-activated prodrugs. Basic esters of 5-bromo-2'-deoxyuridine. *J. Med. Chem.*, **33** (9), 2590–2595.
29. Carpino, L.A., Triolo, S.A., and Berglund, R.A. (1989) Reductive lactonization of strategically methylated quinone propionic acid esters and amides. *J. Org. Chem.*, **54** (14), 3303–3310.
30. Waugh, D.S. (2011) An overview of enzymatic reagents for the removal of affinity tags. *Protein Expression and Purification*, **80** (2), 283–293.
31. Chudasama, V. (2018) Antibody – Drug conjugates (ADC) – Drug discovery today: Technologies. *Drug Discov. Today Technol.*, **30**, 1–2.
32. Dubowchik, G.M., and Firestone, R.A. (1998) Cathepsin B-sensitive dipeptide prodrugs. 1. A model study of structural requirements for efficient release of doxorubicin. *Bioorg. Med. Chem. Lett.*, **8** (23), 3341–3346.
33. Dubowchik, G.M., Mosure, K., Knipe, J.O., and Firestone, R.A. (1998) Cathepsin B-sensitive dipeptide prodrugs. 2. Models of anticancer drugs paclitaxel (Taxol®), mitomycin C and doxorubicin. *Bioorg. Med. Chem. Lett.*, **8** (23), 3347–3352.
34. Beck, A., Goetsch, L., Dumontet, C., and Corvaia, N. (2017) Strategies and challenges for the next generation of antibody–drug conjugates. *Nat Rev Drug Discov*, **16** (5), 315–337.

35. Dubowchik, G.M., Firestone, R.A., Padilla, L., Willner, D., Hofstead, S.J., Mosure, K., Knipe, J.O., Lasch, S.J., and Trail, P.A. (2002) Cathepsin B-Labile Dipeptide Linkers for Lysosomal Release of Doxorubicin from Internalizing Immunoconjugates: Model Studies of Enzymatic Drug Release and Antigen-Specific In Vitro Anticancer Activity. *Bioconjugate Chem.*, **13** (4), 855–869.
36. Doronina, S.O., Toki, B.E., Torgov, M.Y., Mendelsohn, B.A., Cervený, C.G., Chace, D.F., DeBlanc, R.L., Gearing, R.P., Bovee, T.D., Siegall, C.B., Francisco, J.A., Wahl, A.F., Meyer, D.L., and Senter, P.D. (2003) Development of potent monoclonal antibody auristatin conjugates for cancer therapy. *Nat Biotechnol*, **21** (7), 778–784.
37. Gébleux, R., Stringhini, M., Casanova, R., Soltermann, A., and Neri, D. (2017) Non-internalizing antibody-drug conjugates display potent anti-cancer activity upon proteolytic release of monomethyl auristatin E in the subendothelial extracellular matrix: Non-internalizing antibody-drug conjugates and their anti-cancer activity. *Int. J. Cancer*, **140** (7), 1670–1679.
38. Dal Corso, A., Cazzamalli, S., Gébleux, R., Mattarella, M., and Neri, D. (2017) Protease-Cleavable Linkers Modulate the Anticancer Activity of Noninternalizing Antibody–Drug Conjugates. *Bioconjugate Chem.*, **28** (7), 1826–1833.
39. Anami, Y., Yamazaki, C.M., Xiong, W., Gui, X., Zhang, N., An, Z., and Tsuchikama, K. (2018) Glutamic acid–valine–citrulline linkers ensure stability and efficacy of antibody–drug conjugates in mice. *Nat Commun*, **9** (1), 2512.
40. Jeffrey, S.C., Andreyka, J.B., Bernhardt, S.X., Kissler, K.M., Kline, T., Lenox, J.S., Moser, R.F., Nguyen, M.T., Okeley, N.M., Stone, I.J., Zhang, X., and Senter, P.D. (2006) Development and Properties of β -Glucuronide Linkers for Monoclonal Antibody–Drug Conjugates. *Bioconjugate Chem.*, **17** (3), 831–840.
41. Kolakowski, R.V., Haelsig, K.T., Emmerton, K.K., Leiske, C.I., Miyamoto, J.B., Cochran, J.H., Lyon, R.P., Senter, P.D., and Jeffrey, S.C. (2016) The Methylene Alkoxy Carbamate Self-Immolative Unit: Utilization for the Targeted Delivery of Alcohol-Containing Payloads with Antibody-Drug Conjugates. *Angew. Chem. Int. Ed.*, **55** (28), 7948–7951.
42. Lyon, R.P., Bovee, T.D., Doronina, S.O., Burke, P.J., Hunter, J.H., Neff-LaFord, H.D., Jonas, M., Anderson, M.E., Setter, J.R., and Senter, P.D. (2015) Reducing hydrophobicity of homogeneous antibody-drug conjugates improves pharmacokinetics and therapeutic index. *Nat Biotechnol*, **33** (7), 733–735.

43. Kolodych, S., Michel, C., Delacroix, S., Koniev, O., Ehkirch, A., Eberova, J., Cianférani, S., Renoux, B., Krezel, W., Poinot, P., Muller, C.D., Papot, S., and Wagner, A. (2017) Development and evaluation of β -galactosidase-sensitive antibody-drug conjugates. *Eur. J. Med. Chem.*, **142**, 376–382.
44. Bargh, J.D., Walsh, S.J., Isidro-Llobet, A., Omarjee, S., Carroll, J.S., and Spring, D.R. (2020) Sulfatase-cleavable linkers for antibody-drug conjugates. *Chem. Sci.*, **11** (9), 2375–2380.
45. Kern, J.C., Cancilla, M., Dooney, D., Kwasnjuk, K., Zhang, R., Beaumont, M., Figueroa, I., Hsieh, S., Liang, L., Tomazela, D., Zhang, J., Brandish, P.E., Palmieri, A., Stivers, P., Cheng, M., Feng, G., Geda, P., Shah, S., Beck, A., Bresson, D., Firdos, J., Gately, D., Knudsen, N., Manibusan, A., Schultz, P.G., Sun, Y., and Garbaccio, R.M. (2016) Discovery of Pyrophosphate Diesters as Tunable, Soluble, and Bioorthogonal Linkers for Site-Specific Antibody–Drug Conjugates. *J. Am. Chem. Soc.*, **138** (4), 1430–1445.
46. Maeda, H., Wu, J., Sawa, T., Matsumura, Y., and Hori, K. (2000) Tumor vascular permeability and the EPR effect in macromolecular therapeutics: a review. *Journal of Controlled Release*, **65** (1–2), 271–284.
47. Blencowe, C.A., Russell, A.T., Greco, F., Hayes, W., and Thornthwaite, D.W. (2011) Self-immolative linkers in polymeric delivery systems. *Polym. Chem.*, **2** (4), 773–790.
48. Renoux, B., Raes, F., Legigan, T., Péraudeau, E., Eddhif, B., Poinot, P., Tranoy-Opalinski, I., Alsarraf, J., Koniev, O., Kolodych, S., Lerondel, S., Le Pape, A., Clarhaut, J., and Papot, S. (2017) Targeting the tumour microenvironment with an enzyme-responsive drug delivery system for the efficient therapy of breast and pancreatic cancers. *Chem. Sci.*, **8** (5), 3427–3433.
49. Park, C., Kim, H., Kim, S., and Kim, C. (2009) Enzyme Responsive Nanocontainers with Cyclodextrin Gatekeepers and Synergistic Effects in Release of Guests. *J. Am. Chem. Soc.*, **131** (46), 16614–16615.
50. Wolf, P.L. (1994) Clinical significance of serum high-molecular-mass alkaline phosphatase, alkaline phosphatase–lipoprotein-X complex, and intestinal variant alkaline phosphatase. *J. Clin. Lab. Anal.*, **8** (3), 172–176.
51. Chatterjee, S.K., Bhattacharya, M., and Barlow, J.J. (1979) Glycosyltransferase and glycosidase activities in ovarian cancer patients. *Cancer Res*, **39**, 1943–1951.

52. Olson, E.S., Jiang, T., Aguilera, T.A., Nguyen, Q.T., Ellies, L.G., Scadeng, M., and Tsien, R.Y. (2010) Activatable cell penetrating peptides linked to nanoparticles as dual probes for in vivo fluorescence and MR imaging of proteases. *Proc Natl Acad Sci. USA*, **107** (9), 4311–4316.
53. Qian, L., Li, L., and Yao, S.Q. (2016) Two-Photon Small Molecule Enzymatic Probes. *Acc. Chem. Res.*, **49** (4), 626–634.
54. Chen, L., Li, J., Du, L., and Li, M. (2014) Strategies in the Design of Small-Molecule Fluorescent Probes for Peptidases: Small-Molecule Fluorescent Probes For Peptidases. *Med. Res. Rev.*, **34** (6), 1217–1241.
55. Lock, L.L., Tang, Z., Keith, D., Reyes, C., and Cui, H. (2015) Enzyme-Specific Doxorubicin Drug Beacon as Drug-Resistant Theranostic Molecular Probes. *ACS Macro Lett.*, **4** (5), 552–555.
56. Redy, O., Kisin-Finfer, E., Sella, E., and Shabat, D. (2012) A simple FRET-based modular design for diagnostic probes. *Org. Biomol. Chem.*, **10** (4), 710–715.
57. Merrifield, R.B. (1963) Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.*, **85** (14), 2149–2154.
58. Basso, A., Martin, L.D., Gardossi, L., Margetts, G., Brazendale, I., Bosma, A.Y., Ulijn, R.V., and Flitsch, S.L. (2003) Improved biotransformations on charged PEGA supports. *Chem. Commun.*, (11), 1296.
59. Deere, J., De Oliveira, R.F., Tomaszewski, B., Millar, S., Lalaoui, A., Solares, L.F., Flitsch, S.L., and Halling, P.J. (2008) Kinetics of Enzyme Attack on Substrates Covalently Attached to Solid Surfaces: Influence of Spacer Chain Length, Immobilized Substrate Surface Concentration and Surface Charge. *Langmuir*, **24** (20), 11762–11769.
60. Yamada, K., and Nishimura, S.-I. (1995) An efficient synthesis of sialoglycoconjugates on a peptidase-sensitive polymer support. *Tetrahedron Lett.*, **36** (52), 9493–9496.
61. Kress, J., Zanaletti, R., Amour, A., Ladlow, M., Frey, J.G., and Bradley, M. (2002) Enzyme accessibility and solid supports: which molecular weight enzymes can be used on solid supports? An investigation using confocal Raman microscopy. *Chemistry*, **8** (16), 3769–3772.
62. Leon, S., Quarrell, R., and Lowe, G. (1998) Evaluation of resins for on-bead screening: A study of papain and chymotrypsin specificity using pega-bound combinatorial peptide libraries. *Bioorg. Med. Chem. Lett.*, **8** (21), 2997–3002.

63. Grether, U., and Waldmann, H. (2001) An enzyme-labile safety catch linker for synthesis on a soluble polymeric support. *Chemistry*, **7** (5), 959–971.
64. Smith, D.K., and Diederich, F. (1998) functional Dendrimers: Unique Biological Mimics. *Chem. Eur. J.*, **4** (8), 1353–1361.
65. Wang, Y.-C., and Distefano, M.D. (2012) Solid-phase synthesis of C-terminal peptide libraries for studying the specificity of enzymatic protein prenylation. *Chem. Commun.*, **48** (66), 8228.
66. Chiva, A., Williams, D.E., Tabor, A.B., and Hailes, H.C. (2010) Screening of polymeric supports and enzymes for the development of an endo enzyme cleavable linker. *Tetrahedron Lett.*, **51** (20), 2720–2723.
67. Weetall, H. (1976) *Chimia*, 429–430.
68. Schuster, M., Wang, P., Paulson, J.C., and Wong, C.-H. (1994) Solid-Phase Chemical-Enzymic Synthesis of Glycopeptides and Oligosaccharides. *J. Am. Chem. Soc.*, **116** (3), 1135–1136.
69. Weiler, J. (1997) Hybridisation based DNA screening on peptide nucleic acid (PNA) oligomer arrays. *Nucleic Acids Res.*, **25** (14), 2792–2799.
70. Maltman, B.A., Bejugam, M., and Flitsch, S.L. (2005) Enzyme-cleavable linkers for peptide and glycopeptide synthesis. *Org. Biomol. Chem.*, **3** (14), 2505.
71. Elmore, D.T., Guthrie, D.J.S., Wallace, A.D., and Bates, S.R.E. (1992) An enzyme-scissile linker for solid-phase peptide synthesis. *J. Chem. Soc., Chem. Commun.*, (14), 1033.
72. Yamada, K., Fujita, E., and Nishimura, S.-I. (1997) High performance polymer supports for enzyme-assisted synthesis of glycoconjugates. *Carbohydrate Research*, **305** (3–4), 443–461.
73. Akbar, U., Shin, D.-S., Schneider, E., Dordick, J.S., and Clark, D.S. (2010) Two-step enzymatic modification of solid-supported bergenin in aqueous and organic media. *Tetrahedron Lett.*, **51** (8), 1220–1225.
74. Nishimura, S.-I., and Yamada, K. (1997) Transfer of Ganglioside GM3 Oligosaccharide from a Water Soluble Polymer to Ceramide by Ceramide Glycanase. A Novel Approach for the Chemical-Enzymatic Synthesis of Glycosphingolipids. *J. Am. Chem. Soc.*, **119** (43), 10555–10556.

75. Sauerbrei, B., Jungmann, V., and Waldmann, H. (1998) An Enzyme-Labile Linker Group for Organic Syntheses on Solid Supports. *Angew. Chem. Int. Ed. Engl.*, **37** (8), 1143–1146.
76. Flitsch, S.L., and Turner, N.J. (1997) Solid phase preparation and enzymic and non-enzymic bond cleavage of sugars and glycopeptides. *CAN*, **127** (81736).
77. Canne, L.E., Botti, P., Simon, R.J., Chen, Y., Dennis, E.A., and Kent, S.B.H. (1999) Chemical Protein Synthesis by Solid Phase Ligation of Unprotected Peptide Segments. *J. Am. Chem. Soc.*, **121** (38), 8720–8727.
78. Brik, A., Keinan, E., and Dawson, P.E. (2000) Protein Synthesis by Solid-Phase Chemical Ligation Using a Safety Catch Linker. *J. Org. Chem.*, **65** (12), 3829–3835.
79. Johnson, E.C.B., Durek, T., and Kent, S.B.H. (2006) Total Chemical Synthesis, Folding, and Assay of a Small Protein on a Water-Compatible Solid Support. *Angew. Chem. Int. Ed.*, **45** (20), 3283–3287.
80. Jbara, M., Seenaiah, M., and Brik, A. (2014) Solid phase chemical ligation employing a rink amide linker for the synthesis of histone H2B protein. *Chem. Commun.*, **50** (83), 12534–12537.
81. Decostaire, I.E., Lelièvre, D., Aucagne, V., and Delmas, A.F. (2014) Solid phase oxime ligations for the iterative synthesis of polypeptide conjugates. *Org. Biomol. Chem.*, **12** (29), 5536–5543.
82. Yu, R.R., Mahto, S.K., Justus, K., Alexander, M.M., Howard, C.J., and Ottesen, J.J. (2016) Hybrid phase ligation for efficient synthesis of histone proteins. *Org. Biomol. Chem.*, **14** (9), 2603–2607.
83. Loibl, S.F., Harpaz, Z., Zitterbart, R., and Seitz, O. (2016) Total chemical synthesis of proteins without HPLC purification. *Chem. Sci.*, **7** (11), 6753–6759.
84. Zitterbart, R., and Seitz, O. (2016) Parallel Chemical Protein Synthesis on a Surface Enables the Rapid Analysis of the Phosphoregulation of SH3 Domains. *Angew. Chem. Int. Ed.*, **55** (25), 7252–7256.
85. Zitterbart, R., Krumrey, M., and Seitz, O. (2017) Immobilization methods for the rapid total chemical synthesis of proteins on microtiter plates: Rapid Chemical Synthesis of Proteins on Microtiter Plates. *J. Pept. Sci.*, **23** (7–8), 539–548.

86. Valverde, I.E., Lecaille, F., Lalmanach, G., Aucagne, V., and Delmas, A.F. (2012) Synthesis of a Biologically Active Triazole-Containing Analogue of Cystatin A Through Successive Peptidomimetic Alkyne-Azide Ligations. *Angew. Chem. Int. Ed.*, **51** (3), 718–722.
87. Raibaut, L., Adihou, H., Desmet, R., Delmas, A.F., Aucagne, V., and Melnyk, O. (2013) Highly efficient solid phase synthesis of large polypeptides by iterative ligations of bis(2-sulfanylethyl)amido (SEA) peptide segments. *Chem. Sci.*, **4** (10), 4061.
88. Galibert, M., Piller, V., Piller, F., Aucagne, V., and Delmas, A.F. (2015) Combining triazole ligation and enzymatic glycosylation on solid phase simplifies the synthesis of very long glycoprotein analogues. *Chem. Sci.*, **6** (6), 3617–3623.
89. Ollivier, N., Desmet, R., Drobecq, H., Blanpain, A., Boll, E., Leclercq, B., Mougel, A., Vicogne, J., and Melnyk, O. (2017) A simple and traceless solid phase method simplifies the assembly of large peptides and the access to challenging proteins. *Chem. Sci.*, **8** (8), 5362–5370.
90. Zhou, B., Faridoon, Tian, X., Li, J., Guan, D., Zheng, X., Guo, Y., and Huang, W. (2018) On-resin peptide ligation via C-terminus benzyl ester. *Chin. Chem. Lett.*, **29** (7), 1123–1126.
91. Jayalekshmy, P., and Mazur, S. (1976) Pseudodilution, the solid-phase immobilization of benzyne. *J. Am. Chem. Soc.*, **98** (21), 6710–6711.
92. Papot, S., Tranoy, I., Tillequin, F., Florent, J., and Gesson, J. (2002) Design of Selectively Activated Anticancer Prodrugs: Elimination and Cyclization Strategies. *CMCACA*, **2** (2), 155–185.
93. Florent, J.-C., Dong, X., Gaudel, G., Mitaku, S., Monneret, C., Gesson, J.-P., Jacquesy, J.-C., Mondon, M., Renoux, B., Andrianomenjanahary, S., Michel, S., Koch, M., Tillequin, F., Gerken, M., Czech, J., Straub, R., and Bosslet, K. (1998) Prodrugs of Anthracyclines for Use in Antibody-Directed Enzyme Prodrug Therapy. *J. Med. Chem.*, **41** (19), 3572–3581.
94. Andrianomenjanahary, S., Dong, X., Florent, J.-C., Gaudel, G., Gesson, J.-P., Jacquesy, J.-C., Koch, M., Michel, S., Mondon, M., Monneret, C., Petit, P., Renoux, B., and Tillequin, F. (1992) Synthesis of novel targeted pro-prodrugs of anthracyclines potentially activated by a monoclonal antibody galactosidase conjugate (part 1). *Bioorg. Med. Chem. Lett.*, **2** (9), 1093–1096.
95. Gesson, J.P., Jacquesy, J.C., Mondon, M., Petit, P., Renoux, B., Andrianomenjanahary, S., Dufat-Trinh Van, H., Koch, M., Michel, S., and Tillequin, F. (1994) Prodrugs of

anthracyclines for chemotherapy via enzyme-monoclonal antibody conjugates.

Anticancer Drug Des., **9** (5), 409–423.

IV. Publication 2 (Published in Angewandte Chemie International Edition)

These results have been submitted (16 March 2021) and published (07 June 2021) for publication in Angewandte Chemie International. In this thesis manuscript, we have chosen to reproduce the original paper.

Abboud S.A., Amoura M., Renoux. B., Papot S., Piller V., Aucagne V. (2021) Enzyme-cleavable linkers for protein chemical synthesis through solid-phase ligations. *Angew. Chem. Int. Ed.* 2021. DOI : 10.1002/anie.202103768

Enzyme-cleavable linkers for protein chemical synthesis through solid-phase ligations

Skander A. Abboud,^[a] Mehdi Amoura,^[a] Brigitte Renoux,^[b] Sébastien Papot,^[b] Véronique Piller,^[a] and Vincent Aucagne^{*[a]}

[a] Dr Skander A. Abboud, Dr Mehdi Amoura, Dr Véronique Piller and Dr Vincent Aucagne
Centre de Biophysique Moléculaire, CNRS UPR 4301
Rue Charles Sadron, 45071 Orléans cedex 2, France
E-mail: vincent.aucagne@cnrs-orleans.fr

[b] Dr Brigitte Renoux and Prof Sébastien Papot
Institut de Chimie des Milieux et des Matériaux de Poitiers, UMR-CNRS 7285
4 rue Michel Brunet, 86073 Poitiers cedex 9, France

Supporting information for this article is given via a link at the end of the document.

Abstract: The total synthesis of long proteins requires the assembly of multiple fragments through successive ligations. The need for intermediate purifications is a strong limitation, as this generally leads to low to very low overall yields. An early-recognized solution to this problem is the assembly of the protein on a solid support, but no mature solid-supported chemical ligation (SPCL) technology has yet emerged. One of the main reasons lies in the difficulty to immobilize a first peptide segment on a SPCL-compatible solid support through a linker that can be cleaved under very mild conditions to release the assembled protein. A wide range of SPCL linkers have been developed but their cleavage requires chemical conditions sometimes incompatible with sensitive protein targets. Here, we describe the exploration of an alternative enzymatic approach in order to trigger cleavage under extremely mild and selective conditions. Optimization of the linker structure and use of a small enzyme able to diffuse into the solid support was a key for the success of the strategy. We demonstrated its utility by the native chemical ligation-based assembly of three peptide segments to afford a 15 kDa polypeptide.

Introduction

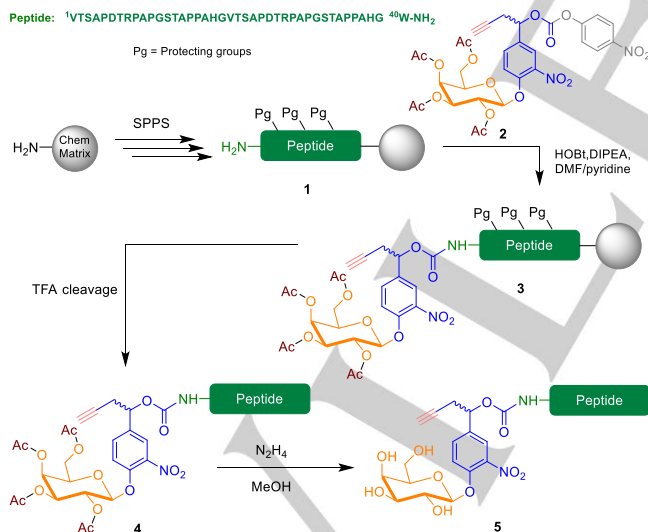
In the last decades, the production of proteins by total synthesis became highly complementary to recombinant techniques for investigating biological mechanisms and for drug discovery. Current technologies focus on the modular assembly of unprotected peptide segments in aqueous solutions, through highly chemoselective “chemical ligation” reactions. The ligation approach revolutionized the field some thirty years ago^[1] and has gradually been democratized for the synthesis of small proteins. However, the access to more ambitious targets in terms of size remains a challenge^[2] since multiple successive ligations are often associated with very low overall yields, due to the laborious intermediate chromatographic purifications. Reminiscent to the solid phase peptide synthesis (SPPS) concept developed by Merrifield^[3] for assembling protected amino-acids (aa), solid phase chemical ligation (SPCL)^[4] has been proposed to obviate this difficulty. Sequential ligations of unprotected peptide segments on an insoluble support potentially lead to the rapid synthesis of proteins in highest yields, as purifications after each synthetic step are limited to simple washes. Although extremely appealing, SPCL has been so far essentially limited to proof-of-concepts. The choice of the linker is a keystone in SPCL strategies: it should be stable along the protein assembly then be

efficiently cleaved to release the synthesized protein into solution without affecting its chemical integrity. An extensive range of SPCL linkers have been developed and introduced on either the C-^[4a-c,f,g,i] or N-terminal^[4a,d,e,h,i,n] part of the first peptide segment, depending on the directional strategy chosen for the successive ligations: from the C- to the N-terminus of the target protein (C-to-N), or N-to-C. Cleavage conditions include treatments with acids,^[4b,c,f,g,n] bases,^[4a,d,e,i] and nucleophiles,^[4h,i] sometimes incompatible with sensitive targets,^[4h,5] and the development of alternative linkers represents a major challenge to expand the applicability of SPCL and thus further push away the limits of chemical protein synthesis, as illustrated very recently by the development of a transition metal-cleavable linker.^[6] Within this context, we thought exploiting an enzymatic rather than purely chemical approach, in order to induce cleavage under extremely soft and selective conditions. Indeed, enzymatic transformations often occur at pH 6–8, at mild temperatures and with a high chemo-, regio- and stereoselectivity for the corresponding substrates. Besides considerations on the chemical integrity of the target protein, enzymatic cleavage conditions are also non-denaturing, a potential significant advantage to exploit the “pseudo-dilution” effect^[7] for the folding of the immobilized synthesized target, while avoiding intermolecular aggregation.^[8] Herein we describe the exploration of a new generation of SPCL linkers, cleavable by β -galactosidases or phosphatases, and demonstrate their utility by the solid supported native chemical ligation (NCL)-based synthesis of a 160 residues polypeptide, the longest sequence ever synthesized through solid-supported NCLs.

Results and Discussion

Enzyme-cleavable linkers have been developed for applications to the solid-supported synthesis of peptides,^[9] glycopeptides,^[9b] oligosaccharides,^[10] oligonucleotides,^[11] peptide nucleic acids^[12] and glycosphingolipids,^[13] and small molecules combinatorial libraries.^[14] They exploit phosphodiesterases,^[9a] lipases,^[12b] glycosidases,^[10a,11,13] amidases^[14a,14c,13] and peptidases,^[10b] to release alcohols,^[10,14] amines,^[11,13,14b-d] or carboxylic acids^[9,11,14b] from the solid support after multistep synthesis of the target compounds. Nevertheless, none of these linkers appears as directly applicable to SPCL.

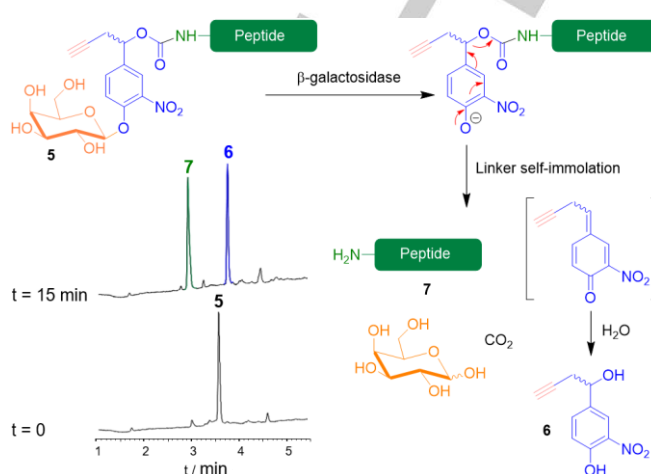
An ideal SPCL linker should allow re-immobilization of the first peptide segment on a solid support different from the one used for SPPS, due to different requirements in terms of physicochemical properties. The SPCL support has to be tolerant to the aqueous environment used for the successive ligations and the enzymatic cleavage, and be compatible with the diffusion of the enzyme. Re-immobilization has other valuable advantages, such as offering the possibility to screen different solid supports, and to allow a purification step after the SPPS in order to provide a clean and pure segment as the starting material for SPCL. We thus privileged an N-terminal linker, well adapted to a re-immobilization strategy, introduced on the N-terminal amino group of the segment after SPPS elongation. The linker precursor should thus be equipped with both an amine-reactive moiety and a functional group adapted for chemoselective immobilization of the unprotected peptide segment, and linker cleavage should release the unmodified N-terminal amine upon enzymatic treatment. Having these requirements in mind, we first evaluated a β -galactosidase-sensitive linker originally developed for targeted drug delivery.^[15] Linker precursor **2** is synthesized in a few steps as a diastereomeric mixture,^[15a] and bears a per-*O*-acetylated β -galactoside, an alkyne suitable for chemoselective immobilization, and an amine-reactive activated carbonate. **2** was reacted with a solid-supported model peptide **1** which sequence is based on the human mucin MUC1 variable number tandem repeat (VNTR) region, made of a duplicated 20 aa sequence and frequently used as a model for peptide ligation and bioconjugation.^[4d,h,g,k,16,17] Resulting peptidyl resin **3** was cleaved from the SPPS solid support, and galactoside deacetylation was carried out in solution through treatment with hydrazine^[18,19] to afford compound **5** as a mixture of two diastereoisomers easily separable by HPLC (scheme 1 and SI p S7-S8).



Scheme 1. Introduction of the β -galactosidase-labile linker on a model peptide

We first investigated the enzymatic cleavage of the linker in solution. As expected, treatment of **5** by *E. coli* β -galactosidase resulted in a clean release of peptide **7** in less than 15 minutes, through galactoside hydrolysis followed by self-immolation of the resulting *p*-hydroxybenzyloxy carbamate moiety (scheme 2).^[20] Subsequently, **5** was immobilized through copper-catalyzed

cycloaddition (CuAAC)^[21] on three different azide-functionalized solid supports: polyethylene glycol (PEG)-crosslinked polyacrylamide resin PEGA¹⁹⁰⁰ **8a**,^[22] and controlled pore glasses CPG1000 **8b** and CPG2000 **8c**, silica-based inorganic matrixes forming porous beads with pores of 1000 and 2000 Å in diameter, respectively.^[23]



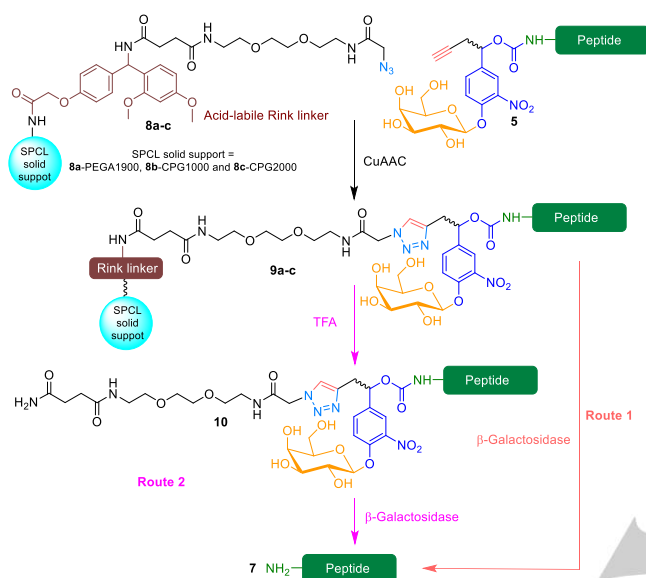
Scheme 2. Solution phase enzymatic cleavage of linker-functionalized model peptide **5**. HPLC monitoring ($\lambda = 280$ nm) shown for one diastereomer of **5**, the other one giving similar results

These supports were chosen for their superior efficiency for enzyme-mediated transformations as compared to other water-compatible resins.^[4h,14d,24] For analytical purposes, the supports were functionalized with a trifluoroacetic acid (TFA)-labile Rink amide linker, separated from the azide moiety by an oligo(ethylene glycol)-based spacer to ensure accessibility of the enzyme to the substrate^[25] (scheme 3).

Very disappointingly, β -galactosidase-mediated linker cleavage of immobilized peptides **9a-c** proceeded considerably slower than for **5** in solution (see SI fig. S8): after more than 5 h reaction using a ten-fold increase in enzyme concentration as compared to solution phase conditions, only 3%, 4% and 16% of peptide **7** was released from PEGA, CPG1000 and CPG2000, respectively. The superior efficacy of CPG towards PEGA is in accordance with Bradley's systematic comparison of different supports with a range of enzyme of increasing sizes,^[24b] showing that enzymatic reaction rates are directly related to the ability of enzyme to diffuse inside the beads. This correlates well with the significant differences we observed between CPG1000 and 2000. The relative cleavage kinetics rate in solution and on the latter support was estimated to $\sim 60\,000$ (SI p S59), from the ratio of apparent first order kinetic constants $k_{\text{solution}} / k_{\text{CPG2000}}$. Such a slow reaction is prohibitive for SPCL applications, and all our attempts to improve cleavage efficiency were unsuccessful: increasing enzyme concentration, using a long PEG spacer (3 kDa), or prior β -galactosidase-mediated "shaving"^[26] of the CPG beads in order to install azide groups only at sites easily accessible by the enzyme (see SI p S19-S26) only moderately affected cleavage kinetics. We explored an alternative approach taking advantage of the rapid enzymatic cleavage in solution, consisting in TFA-mediated cleavage of the Rink linker of **9c** followed by β -galactosidase treatment (scheme 3).^[27] Although this approach gave clean results and was easy to perform, it requires one

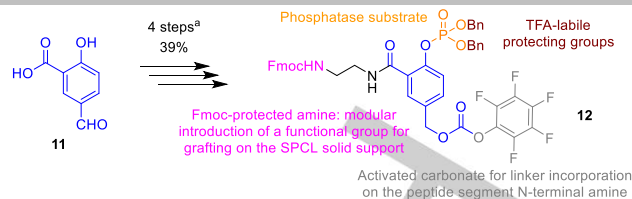
additional step, and uses denaturing conditions that would preclude future potential applications to on-resin folding following the SPCL-based assembly of ambitious targets.^[28]

We assumed that the large size of β -galactosidase (a 540 kDa homo-tetramer) hampered its diffusion inside the solid supports. Hence, we developed a second-generation tailor-made linker cleavable by phosphatases, enzymes of much smaller size than β -galactosidases. Linker precursor **11** was easily synthesized in



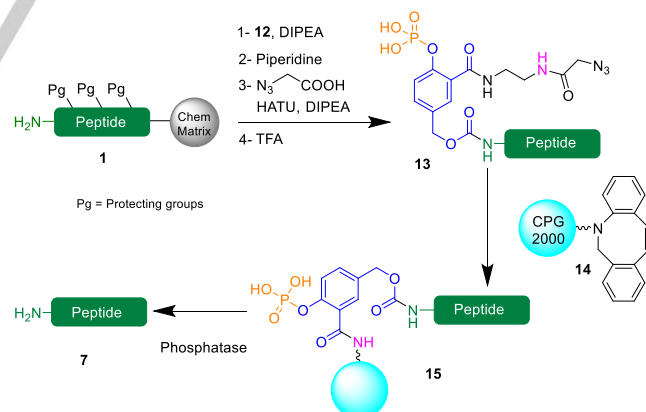
Scheme 3. The two routes explored for the release in solution of peptide **7**.

four steps from commercially-available materials (39% overall yield). (scheme 4, see SI p S29-S40 for the detailed synthesis). **11** was designed to address all the other limitations we identified with the β -galactosidase-labile linker. Indeed, the latter requires an additional step in solution after SPPS to cleave the acetate groups of the galactoside, Cu(I) salts used for CuAAC-based immobilization can lead to the oxidation of peptide segment side-chains if conditions are not well controlled,^[29] and presence of a non-controlled chiral center leads to a diastereomeric mixture that complicates peptide segment purification and characterization. Our second-generation linker is achiral, and precursor **12** was designed as a *N*-fluorenylmethyloxycarbonyl (Fmoc) SPPS building block to provide optimal modularity for the further introduction of any functional group suited for chemoselective grafting on a solid support. The phosphate group is protected by TFA-labile *O*-benzyl groups, compatible with standard Fmoc-SPPS protocols.^[30] Also note that the *N*-Fmoc amino group is branched through an electron-withdrawing carboxamide group in order to lower the pKa of the phenol generated upon phosphate cleavage and thus promote a fast 1,6-elimination-mediated self-immolation at neutral pH, this process being greatly accelerated by phenol deprotonation.^[31]



Scheme 4. Synthesis of the second generation linker precursor. ^a: 1) *N*-Fmoc ethylene diamine, HATU, DIPEA, NMP; 2) dibenzyl phosphite, CCl₄, DIPEA, MeCN, 75% over 2 steps; 3) NaBH₄, THF, 97%; 4) bis-pentafluorophenyl carbonate, Et₃N, MeCN, 53%.

12 was reacted with solid-supported model peptide **1**, followed by *N*-Fmoc group cleavage and subsequent coupling with azidoacetic acid in order to introduce an azide group suitable for copper-free strain-promoted alkyne/azide cycloaddition (SPAAC).^[32] As expected, SPPS resin cleavage cleanly afforded compound **13**. We next screened linker cleavage in solution by four commercially available enzymes: alkaline phosphatase extracted from calf intestine, and the recombinant antarctic phosphatase, shrimp alkaline phosphatase and bacteriophage lambda protein phosphatase, using the respective standard manufacturer recommendations in terms of reaction buffers and enzyme concentrations (SI, pS51). All enzymes proved efficient to mediate a clean release of peptide **1** (see SI, p S51-S53) through phosphoester hydrolysis followed by instantaneous self-immolation of the resulting *p*-hydroxybenzyloxy carbamate, not observed in LC-MS analyses as for the first generation linker. Best results in terms of kinetics were obtained with the shrimp and lambda phosphatases and these two enzymes were further evaluated for solid supported cleavage. Grafting of **13** on the CPG2000-Rink solid support **14** bearing an azadibenzocyclooctyne (DBCO) moiety^[33] through a chemoselective SPAAC reaction gave solid-supported **15** (scheme 5).^[34]

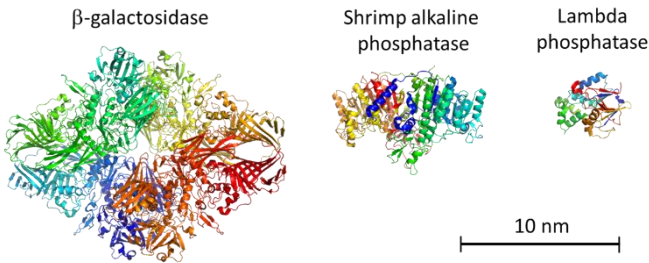


Scheme 5. 2nd generation linker introduction on model peptide and solid supported enzymatic cleavage.

To quantify the relative efficiency of the enzymatic cleavage on solid support as compared to solution phase, the $k_{\text{solution}} / k_{\text{CPG2000}}$ ratio was evaluated for the two enzymes (table 1, SI p S58 and p S61-S62). As compared with the results obtained for β -galactosidase, the relative ratio was enhanced by a moderate factor of 6 for shrimp phosphatase and by an impressive factor of

300 for lambda phosphatase. These results perfectly correlate with the molecular weight and size of the three enzymes: shrimp phosphatase is a homodimeric 108 kDa protein, whereas lambda phosphatase is a much smaller 25 kDa monomer (fig. 1). Gratifyingly, these results validate our hypothesis on the influence of the size of the enzyme on the kinetics of the solid-supported linker cleavage. Despite the 200-fold slow down compared to the reaction in solution, lambda phosphatase cleavage kinetics are perfectly compatible with our SPCL purpose since the reaction can be completed in a few hours using low enzyme amounts.^[35] Importantly, we never detected any peak consistent with the enzyme in LC-MS analyses of the enzymatic cleavages. This suggests a very low molar enzyme amount as compared to the substrate, thus being probably easily removed with a final HPLC “polishing” purification after SPCL assembly of a protein target.

Table 1. Comparison of enzymatic cleavage efficiency.

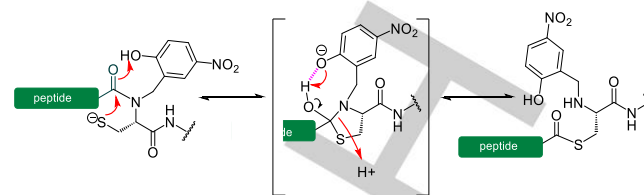


Entry	Enzyme	MW	aa	$k_{\text{solution}} / k_{\text{CPG2000}}$
1	β -galactosidase	540 kDa	4092	~ 60 000
2	Shrimp phosphatase	108 kDa	956	~ 10 000
3	Lambda phosphatase	25 kDa	221	~ 200

Figure 1. Comparison of the size of enzymes used in this work. PDB IDs: β -galactosidase: 4V40, shrimp phosphatase: 1SHQ, lambda phosphatase: 1G5B.

Having in hand a robust method for the enzymatic release from CPG2000 solid support, we then implemented this approach for the synthesis of a 160 aa polypeptide derived from MUC1 VNTRs, through two successive solid-supported NCLs. NCL is the reaction between the C-terminal C α thioester function of one unprotected peptide segment and the N-terminal terminal cysteine of another segment. Consequently, a keystone for N-to-C successive NCLs is the use of dormant precursors of thioesters that can be activated on demand, to prevent cyclization and oligomerization of segments. Such precursors include thioacids,^[4a] hydrazides,^[36] and *N*→*S* acyl shift-based^[4e,] crypto-thioesters.^[37] The latter are based on the rearrangement of a β -mercapto amide into a β -amino-thioester under NCL conditions, through intramolecular attack of the thiol on the amide. Conveniently, simple protection of the thiol afford unreactive species, keeping the crypto-thioester in the dormant state until *S*-deprotected.^[4e,41,38] In the present work, we used *N*-(2-hydroxy-4-nitrobenzyl)cysteine-based devices (*N*-Hnb-Cys)^[39] that are straightforward to synthesize through automated Fmoc-SPPS

and show fast rearrangement and NCL kinetics owing to internal catalysis by a judiciously placed phenol group (Scheme 6).^[39a,40]



Scheme 6. Putative *N*-(Hnb)Cys self-catalyzed *N*→*S* acyl shift mechanism.

Segment **16** was equipped with both our second generation linker functionalized by an azide group and an *N*-Hnb-Cys device which thiol is protected by an *S*-*t*Bu disulfide, labile under reductive NCL conditions. This segment was immobilized onto CPG2000 solid support **14** through SPAAC, in a similar fashion to model peptide **13** (scheme 7). The second segment (**18**) bears a N-terminal cysteine and a *N*-Hnb-Cys device where the thiol is protected by a *S*-acetamidomethyl (Acm) group,^[41] that is stable under NCL conditions.^[38b] NCL between **17** and **18** gave a clean product as judged by an analytical-scale cleavage. The Acm group was then deprotected through treatment with Pd(II) salts^[42] to afford an active crypto-thioester (**20**) which was immediately engaged into a second NCL with the C-terminal segment **21**. Finally, enzymatic cleavage with lambda phosphatase released the desired 160 aa polypeptide **23**, recovered in a 60% overall yield as determined by HPLC peak integration-based quantification. Final semi-preparative HPLC purification gave pure **23** in a 13% isolated yield.^[43,44]

Conclusion

In summary, we introduced a new SPCL strategy based on an enzyme cleavable linker. The major bottleneck identified for the success of the strategy is the difficulty for high molecular weight enzymes to diffuse into the solid support, therefore resulting in extremely slow kinetics. Design of a tailor-made linker cleavable by a small enzyme, lambda phosphatase, gratifyingly solved the problem, and the usefulness of strategy was exemplified for the synthesis of a long polypeptide from three fragments through N-to-C NCL reactions.

Acknowledgements

We warmly thank Jean-Baptiste Madinier for his help in peptide synthesis, as well as Florian Kersaudy for technical assistance and Dr Cyril Colas for LC/HRMS analyses. We also thank Mathilde Seinfeld and Laura Lefèvre for their participation to the optimization of the synthesis of different building blocks. The University of Orléans and doctoral school SSBCV (ED549) are gratefully acknowledged for a PhD fellowship for SA, and ANR (EasyMiniprot, ANR-15-CE07-0022) for financial support.

Keywords: SPCL • NCL • Linkers • Enzyme • Solid support



11. a) M. Schnolzer, S. Kent, *Science* **1992**, 256, 221; b) P. Dawson, T. Muir, I. Clark-Lewis, S. Kent. *Science* **1994**, 266, 776.

- Chem.* **2012**, *124*, 11482; e) L. Raibaut, H. Adihou, R. Desmet, A. F. Delmas, V. Aucagne, O. Melnyk, *Chem. Sci.* **2013**, *4*, 4061; f) M. Jbara, M. Seenayah, A. Brik, *Chem. Commun.* **2014**, *50*, 12534; g) I. E. Decostaire, D. Lelièvre, V. Aucagne, A. F. Delmas, *Org. Biomol. Chem.* **2014**, *12*, 5536; h) M. Galibert, V. Pillier, F. Pillier, V. Aucagne, A. F. Delmas, *Chem. Sci.* **2015**, *6*, 3617; i) R. R. Yu, S. K. Mahto, K. Justus, M. M. Alexander, C. J. Howard, J. J. Ottesen, *Org. Biomol. Chem.* **2016**, *14*, 2603; j) S. F. Loibl, Z. Harpaz, R. Zitterbart, O. Seitz, *Chem. Sci.* **2016**, *7*, 6753; k) R. Zitterbart, O. Seitz, *Angew. Chem. Int. Ed.* **2016**, *55*, 7252; *Angew. Chem.* **2016**, *128*, 7368; l) N. Ollivier, R. Desmet, H. Drobecq, A. Blanpain, E. Boll, B. Leclercq, A. Mougél, J. Vicogne, O. Melnyk, *Chem. Sci.* **2017**, *8*, 5362; m) R. Zitterbart, M. Krumrey, O. Seitz, *J. Pept. Sci.* **2017**, *23*, 539; n) B. Zhou, Faridoon, X. Tian, J. Li, D. Guan, X. Zheng, Y. Guo, W. Huang, *Chin. Chem. Lett.* **2018**, *29*, 1123; l) S. A. Abboud, A. F. Delmas, V. Aucagne, Solid-phase Chemical Ligation, in *Total Chemical*

- Synthesis of Proteins*, (Eds.: A. Brik, P. Dawson, L. Liu), Wiley-VCH, Weinheim, **2021**, pp. 259–284.
5. R. Zitterbart, N. Berger, O. Reimann, G. T. Noble, S. Lüdtko, D. Sarma, O. Seitz, *Chem. Sci.* **2021**, *12*, 2389.
 6. G. B. Vamiseti, G. Satish, P. Sulkshane, G. Mann, M. H. Glickman, A. Brik, *J. Am. Chem. Soc.* **2020**, *142*, 19558.
 7. P. Jayalekshmy, S. Mazur, *J. Am. Chem. Soc.* **1976**, *98*, 6710.
 8. The appealing approach of post-SPCL solid-supported folding has only been reported once, for the synthesis of EETI-II through oxidative folding. Note that in this case, the disulfide-reticulated miniprotein is not sensitive to (non-reductive) denaturing conditions used for cleavage. See ref. [4c]. Solid-phase NCL followed by spontaneous folding (as judged by functional assays) has also been applied for the generation of small libraries of SH3 domains (61 aa) on microtiter plates, even though no cleavable linker was incorporated to release the synthesized peptides in solution. See references [4a] and [4m].
 9. a) D. T. Elmore, D. J. S. Guthrie, A. D. Wallace, S. R. E. Bates, *J. Chem. Soc. Chem. Commun.* **1992**, *14*, 1033; b) B. A. Maltman, M. Bejugam, S. L. Flitsch, *Org. Biomol. Chem.* **2005**, *3*, 2505.
 10. a) G. Corrales, A. Fernández-Mayoralas, E. García-Junceda, Y. Rodríguez, *Biotransform.* **2000**, *18*, 271; b) K. Yamada, S. I. Nishimura, *Tetrahedron Lett.* **1995**, *36*, 9493; c) K. Yamada, E. Fujita, S. I. Nishimura, *Carbohydrate Res.* **1997**, *305*, 443.
 11. C. Schmitz, M. T. Reetz, *Org. Lett.* **1999**, *1*, 1729.
 12. J. Weiler, *Nucleic Acids Res.* **1977**, *25*, 2792.
 13. S. I. Nishimura, K. Yamada, *J. Am. Chem. Soc.* **1997**, *119*, 10555.
 14. a) G. Böhm, J. Dowden, D. C. Rice, I. Burgess, J. F. Pilard, B. Guilbert, A. Haxton, R. C. Hunter, N. J. Turner, S. L. Flitsch, *Tetrahedron Lett.* **1998**, *39*, 3819; b) B. Sauerbrei, V. Jungmann, H. Waldmann, *Angew. Chem. Int. Ed.* **1998**, *37*, 1143; *Angew. Chem.* **1998**, *110*, 1187; c) U. Grether, H. Waldmann, *Angew. Chem. Int. Ed.* **2000**, *39*, 1629; *Angew. Chem.* **2000**, *112*, 1688; d) U. Grether, H. Waldmann, *Chem. Eur. J.* **2001**, *7*, 959.
 15. a) M. Thomas, J. Clarhaut, P. O. Strale, I. Tranoy-Opalinski, J. Roche, S. Papot, *ChemMedChem*, **2011**, *6*, 1006; b) T. Legigan, J. Clarhaut, B. Renoux, I. Tranoy-Opalinski, A. Monvoisin, J. M. Berjeaud, F. Guilhot, S. Papot, *J. Med. Chem.* **2012**, *55*, 4516; c) T. Legigan, J. Clarhaut, I. Tranoy-Opalinski, A. Monvoisin, B. Renoux, M. Thomas, A. Le Pape, S. Lerondel, S. Papot, *Angew. Chem. Int. Ed.* **2012**, *51*, 11606; *Angew. Chem.* **2012**, *124*, 11556; d) B. Renoux, F. Raes, T. Legigan, E. Péraudeau, B. Eddhif, P. Poinot, I. Tranoy-Opalinski, J. Alsarraf, O. Koniev, S. Kolodych, S. Lerondel, A. Le Pape, J. Clarhaut, S. Papot, *Chem. Sci.* **2017**, *8*, 3427; e) R. Châtre, J. Lange, E. Péraudeau, P. Poinot, S. Lerondel, A. Le Pape, J. Clarhaut, B. Renoux, S. Papot, *J. Control. Release*, **2020**, *327*, 19.
 16. D. Montoir, M. Amoura, Z. E. A. Ababsa, T. M. Vishwanatha, E. Yen-Pon, V. Robert, M. Beltramo, V. Piller, M. Alami, V. Aucagne, S. Messaoudi, *Chem. Sci.* **2018**, *9*, 8753.
 17. An additional Trp residue was introduced at the C-terminus of this model peptide to facilitate HPLC monitoring and quantification of the different steps.
 18. H. Kunz, S. Birnbach, *Angew. Chem. Int. Ed. Engl.* **1986**, *25*, 360; *Angew. Chem.* **1986**, *98*, 354.
 19. Acetate deprotection was at first attempted on solid support using either N_2H_4 and MeONa. However, we could not drive the reaction to completion, leading to a mixture of acylated and deacylated hydroxyl moieties in addition to numerous unidentified co-products. For an example of on resin deacetylation, see: M. Liu, G. Barany, D. Live, *Carbohydrate Res.* **2005**, *340*, 2111.
 20. To further optimize the enzymatic hydrolysis, the buffer and the amount of enzyme used were optimized to obtain faster and more economic cleavage conditions (see SI p S11).
 21. a) V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, *41*, 2596; *Angew. Chem.* **2002**, *114*, 2708; b) C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057.
 22. F. I. Auzanneau, M. Meldal, K. Bock, *J. Pept. Sci.* **1995**, *1*, 31.
 23. M. J. Damha, P. A. Giannaris, S. V. Zabarylo, *Nucleic Acids Res.* **1990**, *18*, 3813.
 24. a) S. Leon, R. Quarrell, G. Lowe, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2997; b) J. Kress, R. Zanaletti, A. Amour, M. Ladlow, J. G. Frey, M. Bradley, *Chem. Eur. J.* **2002**, *8*, 3769; c) A. Chiva, D. E. Williams, A. B. Tabor, H. C. Hailes, *Tetrahedron Lett.* **2010**, *51*, 2720.
 25. J. Deere, R. F. De Oliveira, B. Tomaszewski, S. Millar, A. Lalaoui, L. F. Solares, S. L. Flitsch, P. J. Halling, *Langmuir* **2008**, *24*, 11762.
 26. a) J. Vagner, G. Barany, K. S. Lam, V. Krchnak, N. F. Sepetov, J. A. Ostrem, P. Strop, M. Lebl, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 8194; b) Z. P. Gates, A. A. Vinogradov, A. J. Quartararo, A. Bandyopadhyay, Z. N. Choo, E. D. Evans, K. H. Halloran, A. J. Mijalis, S. K. Mong, M. D. Simon, E. A. Standley, E. D. Styduhar, S. Z. Tasker, F. Touti, J. M. Weber, J. L. Wilson, T. F. Jamison, B. L. Pentelute, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E5298.
 27. As an alternative to Rink linker, we systematically evaluated two highly acid-labile linkers classically used in Fmoc-SPPS, but whose susceptibility to acid cleavage under aqueous conditions has never been reported to our knowledge (See SI p S26–S28): a trityl-based linker for carboxylic acids and Sieber's amide linker. The trityl linker was found to be unstable under NCL conditions, thus inappropriate, and Sieber's linker required 10% aqueous TFA for an efficient cleavage. [28] One could imagine that solid-supported folding could not be achieved for some targets (in particular oligomeric proteins), and that the non-denaturing conditions used for linker cleavage could, in some cases, lead to misfolded species that could potentially hinder the accessibility of the phosphatase to its phosphoester substrate, and thus prevent or slow down cleavage. In such case, denaturing TFA cleavage followed by in solution folding prior to enzymatic linker cleavage in solution would be a valuable option. This is out of the scope of the present article, and will be explored in due course.
 29. V. Hong, S. I. Presolski, C. Ma, M. G. Finn, *Angew. Chem. Int. Ed.* **2009**, *48*, 9879; *Angew. Chem.* **2009**, *121*, 10063.
 30. J. W. Perich, E. C. Reynolds, *Synlett*, **1991**, *8*, 577.
 31. A. Alouane, R. Labruère, T. Le Saux, I. Aujard, S. Dubruille, F. Schmidt, L. Jullien, *Chem. Eur. J.* **2013**, *19*, 11717.
 32. N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126*, 15046.
 33. M. F. Debets, S. S. van Berkel, S. Schoffelen, F. P. J. T. Rutjes, J. C. M. van Hest, F. L. van Delft, *Chem. Commun.* **2010**, *46*, 97.
 34. Various conditions were evaluated to carry out the SPAAC reaction before we succeeded in driving it to completion. (see SI p S54).
 35. It is important to note that the “enzyme units” (U) value used to characterize the activity of a given enzyme, and herein used as a basic comparison between the enzyme used, is not comparable from one enzyme to another, and does not correlate to the enzyme

concentration, which is extremely difficult to evaluate precisely. For example, commercial lambda phosphatase activity is 400,000 U/mL whereas *E. coli* β -galactosidase is only 500–7,500 U/mL.

36. G. M. Fang, Y. M. Li, F. Shen, Y. C. Huang, J. B. Li, Y. Lin, H. K. Cui, L. Liu, *Angew. Chem. Int. Ed.* **2011**, *50*, 7645; *Angew. Chem.* **2011**, *123*, 7787.

37. K. Sato, A. Shigenaga, K. Tsuji, S. Tsuda, Y. Sumikawa, K. Sakamoto, A. Otaka, *ChemBioChem.* **2011**, *12*, 1840.

38. a) L. A. Erlich, K. S. A. Kumar, M. Haj-Yahya, P. E. Dawson, A. Brik, *Org. Biomol. Chem.* **2010**, *8*, 2392; b) K. Sato, A. Shigenaga, K. Kitakaze, K. Sakamoto, D. Tsuji, K. Itoh, A. Otaka, *Angew. Chem. Int. Ed.* **2013**, *52*, 7855; *Angew. Chem.* **2013**, *125*, 8009; c) K. Aihara, K. Yamaoka, N. Naruse, T. Inokuma, A. Shigenaga, A. Otaka, *Org. Lett.* **2016**, *18*, 596.

39. a) V. P. Terrier, H. Adihou, M. Arnould, A. F. Delmas, V. Aucagne, *Chem. Sci.* **2016**, *7*, 339; b) D. Lelièvre, V. P. Terrier, A. F. Delmas, V. Aucagne, *Org. Lett.* **2016**, *18*, 920; c) G. Martinez, J. P. Hograindleur, S. Voisin, R. Abi Nahed, T. M. Abd El Aziz, J. Escoffier, J. Bessonnat, C. M. Fovet, M. De Waard, S. Hennebicq, V. Aucagne, P. F. Ray, E. Schmitt, P. Bulet, C. Arnould, *Mol. Hum. Reprod.* **2016**, *23*, 116; d) V. P. Terrier, A. F. Delmas, V. Aucagne, *Org. Biomol. Chem.* **2017**, *15*, 316; e) L. De Rosa, R. Di Stasi, L. D. D'Andrea, *Tetrahedron*, **2019**, *75*, 894; f) K. Loth, A. Vergnes, C. Barreto, S. N. Voisin, H. Meudal, J. Da Silva, A. Bressan, N. Belmadi, E. Bachère, V. Aucagne, C. Cazevielle, H. Marchandin, R. D. Rosa, P. Bulet, L. Touqui, A. F. Delmas, D. Destoumieux-Garzón, *mBio*, **2019**, *10*, e01821; g) S. A. Abboud, V. Aucagne, *Org. Biomol. Chem.* **2020**, *18*, 8199; h) S. A. Abboud, E. Cisse, M. Doudeau, H. Bénédetti, V. Aucagne, *Chem. Sci.* **2021**, *12*, 3194.

40. S. W. Bi, P. Liu, B. P. Ling, X. A. Yuan, Y. Y. Jiang, *Chinese. Chem. Lett.* **2018**, *29*, 1264.

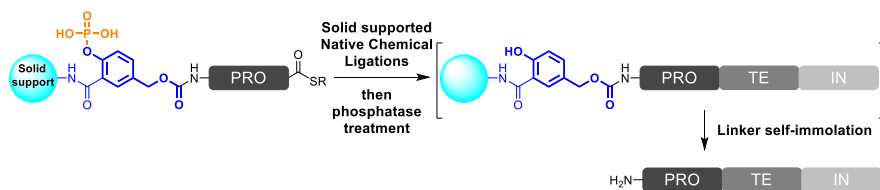
41. D. Veber, J. Milkowski, S. Varga, R. Denkwalter, R. Hirschmann, *J. Am. Chem. Soc.* **1972**, *94*, 5456.

42. S. K. Maity, M. Jbara, S. Laps, A. Brik, *Angew. Chem. Int. Ed.* **2016**, *55*, 8108; *Angew. Chem.* **2016**, *128*, 8240.

43. The overall yield is calculated based on the five synthetic steps (grafting, NCL, AcM deprotection, NCL, and enzymatic cleavage) and the final HPLC purification. The reduction of the isolated yield can be attributed to the small-scale purification, where in loss of product due to non-specific binding can be significant. In addition, we employed a strict signal level-based peak collection method, for maximizing purity over yield.

44. Note that we found important to conduct all the SPCL assembly process (including SPAAC-based immobilization, NCL reactions, AcM deprotection and enzymatic release) without delay between synthetic steps in order to provide with a reproducible enzymatic release yield, this being attributed to unwanted polypeptide aggregation upon storage as advocated in other SPCL-based protein chemical syntheses. See for example references [4b] and [4i].

Entry for the Table of Contents



We herein introduce a chemo-enzymatic approach for the synthesis of proteins through solid-supported chemical ligation, based on a tailor-made phosphatase-cleavable linker. We show that the size of the enzyme used directly correlates with the efficiency of the release of the synthesized protein. The method was applied to the synthesis of a 15 kDa polypeptide, the longest sequence ever synthesized through solid phase native chemical ligation.

Institute and/or researcher Twitter usernames:
@CBM_UPR4301 / @vvinz

V. Supporting Information

These results have been submitted (16 March 2021) and accepted (12 Mai 2021) for publication in *Angewandte Chemie International*. Here, we have reproduced the corresponding supporting information file.

Enzyme cleavable linkers for the total synthesis of proteins through Solid-Supported ligations

Skander A. Abboud,^[a] Mehdi Amoura,^[a] Brigitte Renoux,^[b] Sébastien Papot,^[b] Véronique Piller,^[a] and Vincent Aucagne

a : Centre de Biophysique Moléculaire, CNRS UPR 4301, Rue Charles Sadron 45071 Orléans cedex 2, France.

b : Institut de Chimie des Milieux et des Matériaux, UMR-CNRS 7285, 4 rue Michel Brunet 86073 Poitiers Cedex 9, France.

aucagne@cnrs-orleans.fr

Supporting Information

Table of contents

1-	General information	113
2-	General procedures for solid phase peptide synthesis and NCL	114
3-	First generation, β-galactosidase-labile linker	116
3-1-	Linker introduction on model peptide	116
3-2-	Acetate groups deprotection	117
3-3-	Solution phase β -galactosidase-mediated linker cleavage.....	119
3-4-	Optimization of the solution phase enzymatic cleavage.....	121
3-5-	Solid phase enzymatic cleavage	122
4-	Optimization of first generation linker enzymatic cleavage on CPG2000-supported peptide 9c	127
4-1-	Addition of DTT and $MgCl_2$ salts in the reaction buffer.....	127
4-2-	Use of larger amount of the enzyme	128
4-3-	Use of a larger spacer: PEG3000	128
4-4-	β -galactosidase-mediated “shaving” of the CPG beads in order to install azide groups only at sites easily accessible by the enzyme	129
4-5-	Dual linker strategy	136
5-	Determination of the molar extinction coefficient of the 1st generation linker	139
6-	Second generation enzyme-cleavable linker	140
6-1-	Linker precursor synthesis	140
6-2-	Synthesis of a derivative to determine the molar extinction coefficient of the second generation linker.....	153
6-3-	Linker introduction on model peptide	160
6-4-	Solution phase enzymatic assays.....	163
6-5-	Solid supported enzymatic cleavage	166
7-	Kinetics studies	169
8-	Optimization of a strategy to temporarily mask the reactivity of N-Hnb-Cys crypto-thioesters in order to enable successive N-to-C ligations.	176
8-1-	Synthesis of model crypto-thioesters S32a and S32b	176
8-2-	Screening of different conditions for the deprotection of the Ac μ group in S32a	178
8-3-	PGA-mediated deprotection of the PhAc μ group in S32b	183
9-	Synthesis of the different segments needed for SPCL.....	185
9-1-	N-terminal segment 16	185
9-2-	Compound 18 : Median segment.....	189
9-3-	Compound 21 : C-terminal segment	191
10-	Synthesis of polypeptide 23 through two successive solid supported NCL	193
10-1-	Grafting of the first segment 16 on solid support via SPAAC reaction to obtain 19	193
10-2-	NCL between solid-supported crypto-thioester 19 and cysteinyl peptide 17	195
10-3-	Removal of Ac μ protecting group and second NCL between solid supported crypto-thioester 20 and cysteinyl peptide 21	197

10-4- Enzymatic cleavage and HPLC purification	199
11- References.....	201

1- General information

All reagents and solvents were used without further purification. Protected amino acids were purchased from Gyros Protein Technology (Uppsala, Sweden). DIEA was purchased from Carlo ERBA (Val-de-Reuil, France). Rink ChemMatrix resin was purchased from Biotage (Uppsala, Sweden). Peptide synthesis grade DMF was obtained from VWR (Fontenay-sous-Bois, France). Fmoc-NH-PEG3000-COOH was purchased from Iris Biotech GmbH (Marktredwitz, Germany). Activated carbonate **2** and alcohol **S7** were prepared as described elsewhere [1]. All other chemicals were purchased from Sigma Aldrich (St-Quentin-Fallavier, France) and solvents from Carlo Erba. Ultrapure water was obtained using a Milli-Q water system from Millipore (Molsheim, France). Calf Intestinal Alkaline Phosphatase (CIP) was purchased from Thermo Fisher Scientific (Carlsbad, USA), Antarctic Phosphatase, Lambda Protein Phosphatase (Lambda PP) and CutSmart buffer were purchased from New England Biolabs (Évry, France), Shrimp Alkaline Phosphatase (rSAP) was purchased from Promega (Madison, USA) and *E. coli* β -galactosidase was purchased from Sigma Aldrich (St-Quentin-Fallavier, France). Polypropylene syringes fitted with polypropylene frits were obtained from Torviq (Niles, MI, USA) and were equipped with PTFE stopcock bought from Biotage. ^1H and ^{13}C NMR spectra were recorded on a Bruker AVANCE III 600 instrument, at a constant temperature of 25 °C. Chemical shifts are reported in parts per million from low to high field and referenced to tetramethylsilane (TMS). Coupling constants (J) are reported in hertz (Hz). Standard abbreviations indicating multiplicity were used as follows: s = singlet, d = doublet, dd = doublet of doublets, t = triplet, m = multiplet, b = broad signal. HPLC analyses were carried out on a Chromaster system equipped with a 5160 pump, a 5430 diode array detector and a 5260 auto sampler. Semi-preparative purifications were carried out on a LaChromElite system equipped with a Hitachi L-2130 pump, a Hitachi L-2455 diode array detector and a Hitachi L-2200 auto sampler. Chromolith High Resolution RP-18e (150 Å, 10 × 4.6 mm, 3 mL/min flow rate) columns were used for analysis, Nucleosil C18 (300 Å, 5 μm , 250 × 10 mm, 3 mL/min flow rate) and Jupiter C4 (300 Å, 5 μm , 250 × 10 mm, 3 mL/min flow rate) for purification. As mobile phases, mixtures of 0.1% TFA in H_2O (A) and 0.1% TFA in MeCN (B) were used. Each gradient was followed by a column washing step to elute any compound not eluted during the gradient (up to 95% B over 0.5 min, then isocratic 95% B over 0.5 min for the HR Chromolith). ESI-MS analyses were carried out on an Agilent 1260 Infinity HPLC system, coupled with an Agilent 6120 mass spectrometer (ESI + mode), and fitted with an Aeris Widepore XB-C18 2 (3.6 μm , 150 × 2.1 mm, 0.5 mL/min flow rate, 60 °C) column. The reported m/z values correspond to the monoisotopic ions if not specified otherwise. As mobile phases, mixtures of 0.1% formic acid in H_2O (A') and 0.1% formic acid in MeCN (B') were used. Gradient: 3% B' for 1 min, then 3 to 50% B' over 15 min. Low resolution MS of pure compounds were obtained using this system. When needed, the multiply-charged envelope was deconvoluted using the charge deconvolution tool in Agilent OpenLab CDS ChemStation software to obtain the average [M] value. High resolution ESI-MS analyses were performed on a maXisTM ultra-high-resolution Q-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany), using the positive mode. The multiply-charged envelope was deconvoluted using the charge deconvolution algorithm in Bruker Data Analysis 4.1 software to obtain the monoisotopic [M] value. For yield calculations purposes, the quantities of purified peptides were determined by weight, taking into account a molecular weight including trifluoroacetate counter-ions (one per Arg, His, Lys and N-terminal amine of the peptide sequence) but not water content.

2- General procedures for solid phase peptide synthesis and NCL

Protocol PS1 - Peptide elongation: automated Fmoc-based solid phase peptide syntheses (SPPS) were carried out on a Prelude synthesizer from Protein technologies. Manual couplings were performed on polypropylene syringes fitted with polypropylene frits using rotation stirring. The side-chain protecting groups used were Arg(Pbf), Asp(OtBu), Cys(Trt), His(Trt), Ser(tBu), Thr(tBu), and Trp(Boc), and, as special building blocks, Cys(StBu), Cys(Acm) and Cys(PhAcm). Syntheses were performed on a 0.025 mmol-per-reactor scale. Protected amino acids (0.25mmol, 10 equiv.) were coupled using HATU (90 mg, 0.238 mmol, 9.5 equiv.) and DIEA (87 μ L, 0.5 mmol, 20 equiv.) in NMP (3 mL) for 30 min. Coupling on *N*-Hnb-cysteine secondary amine were performed for 18 h. Capping of potential unreacted amine groups was achieved by treatment with acetic anhydride (143 μ L, 1.51 mmol, 60 equiv.), DIEA (68 μ L, 0.39 mmol, 15.5 equiv.) and HOBt (6 mg, 0.044 mmol, 1.8 equiv.) in NMP (3 mL) for 7 min (4 x 7 min in the case of *N*-Hnb-cysteine secondary amine). Fmoc groups were deprotected by three successive treatments with 20% piperidine in NMP (3 mL) for 3 min.

Protocol PS2 - Reductive amination: introduction of the Hnb group: 25 μ mol of H-Cys(StBu)-Gly-Rink-resin (**1a-c**) was washed with 1:1 DMF/MeOH (4 x 3 mL, 30 s). 2-Hydroxy-5-nitrobenzaldehyde (42 mg, 10 equiv.) in 2 mL 44.5:44.5:1 DMF/MeOH/AcOH (125 mM aldehyde concentration) was then added, and the reactor was stirred for 5 min. The reactor was drained and the resin was washed with 1:1 DMF/MeOH (3 x 3 mL, 5 s) then DMF (3 x 3 mL, 5 s). Without delay, a fresh solution of sodium borohydride (19 mg, 20 equiv.) in 2 mL DMF (250 mM borohydride concentration) was added and the reactor was stirred for 20 min. The reactor was drained and the resin was washed with DMF (4 x 3 mL, 30 s), 20% v/v piperidine in NMP (3 x 3 mL, 3 min), NMP (3 x 3 mL, 30 s), dichloromethane (3 x 5 mL, 30 s) and NMP (3 x 3 mL, 30 s).

Protocol PS3 - Peptide cleavage: peptidyl resin was extensively washed with DCM then dried under an air flow, and the crude peptide was deprotected and cleaved from the resin through a treatment with TFA/H₂O/*i*Pr₃SiH/phenol, 88:5:2:5 for 2 h. The peptide was precipitated by dilution into an ice-cold diethyl ether/petroleum ether 1:1 mixture, recovered by centrifugation and washed twice with diethyl ether.

Protocol PS4 - Small peptides cleavage (less than 4 residues): peptidyl resin was extensively washed with DCM then dried under an air flow, and the crude peptide was deprotected and cleaved from the resin through a treatment with TFA/H₂O/*i*Pr₃SiH, 93:5:2 for 30 min following by concentration under reduced pressure.

Protocol PS5 - Selective Hnb ester cleavage to allow UV titration of Fmoc deprotection: as a consequence of the formation of variable amount (5-90%) of *O*-acylated Hnb during each coupling, this ester being cleaved upon piperidine treatment during Fmoc deprotection, standard UV titration of the fluorenylmethyl-piperidine adduct after Fmoc deprotection is useless unless using a prior treatment for selective ester cleavage before piperidine treatment. Ester cleavage mixture was prepared as follows: 1.25 g (1.80 mmol) of NH₂OH·HCl and 0.918 g (1.35 mmol) of imidazole were suspended in 5 mL of NMP; the mixture was sonicated until complete dissolution. This solution can be stored for few months at -20 °C. 5 volumes of this this solution is diluted with 1 volume of DCM prior to utilization, and the resin is treated for 3x20 min for quantitative ester cleavage. The fluorenylmethyl-piperidine

adduct is quantified by UV spectroscopy at $\lambda = 301 \text{ nm}$ ($\epsilon = 7800 \text{ mol}^{-1} \text{ cm}^{-1}$) in order to evaluate the Fmoc SPPS elongation yield of crypto-thioester peptides.

Protocol PS6 - Native chemical ligation:

Preparation of the NCL buffer (100 mM MPAA, 50 mM TCEP, 6 M guanidinium chloride, 200 mM sodium phosphate, pH = 7.2*):

16.8 mg 4-mercaptophenylacetic acid (MPAA, 0.18 mmol), 14.3 mg tris-carboxyethylphosphine hydrochloride (TCEP, 0.057 mmol) and 10.3 mg of dry NaOH powder (0.26 mmol) were weighted into a vial, which was sealed with a septum and purged with an argon flow.

A 200 mM disodium hydrogen phosphate / 6 M guanidinium chloride aqueous solution was prepared by dissolving 356 mg disodium hydrogen phosphate dihydrate and 5.73 g guanidinium chloride in water (10 mL final volume). This solution can be kept at 4 °C for several months. Just before use, this solution was thoroughly deoxygenated through successive vacuum/argon cycles.

Under argon, 1 mL of this solution was added to the MPAA/TCEP/NaOH vial, followed by sonication upon complete dissolution to give the NCL buffer which was immediately used.

NCL reaction: The cysteinyl peptide and crypto-thioester peptidyl resin were weighted in two separate centrifuge tubes, which were sealed with a septum and purged with an argon flow. The volume of the NCL buffer appropriate to reach the desired final peptide segments concentration was added to the cysteinyl peptide under argon and then transferred to the tube containing the crypto-thioester peptidyl resin. The reaction vessel was sealed with Parafilm, and the resulting yellow solution was allowed to stir at 37 °C for 18 h. The resin was washed with the 200 mM disodium hydrogen phosphate, 6 M guanidinium chloride deoxygenated aqueous solution (3 × 1 mL, 30 s), 200 mM deoxygenated disodium hydrogen phosphate solution (3 × 1 mL, 30 s) and 6 M guanidinium chloride deoxygenated aqueous solution (3 × 1 mL, 30 s) then drained and finally extensively washed with deoxygenated water.

*: apparent pH (measured with a pHmeter) = 6.5, corrected according to M.M. Garcia-Mira and J.M. Sanchez-Ruiz, *Biophys. J.* (2001), **81** (6): 3489–3502.

3- First generation, β -galactosidase-labile linker

SI-p S5

A model peptide derived from the human MUC1 mucin variable number tandem repeat region was used for the study of this first generation linker (two tandem repeats). A Trp residue was added at the C-terminal position of the peptide to facilitate UV monitoring of different reactions. Peptidyl resin **1** was prepared following general protocol PS1, and an analytical TFA cleavage was performed following protocol PS4, giving **7**.

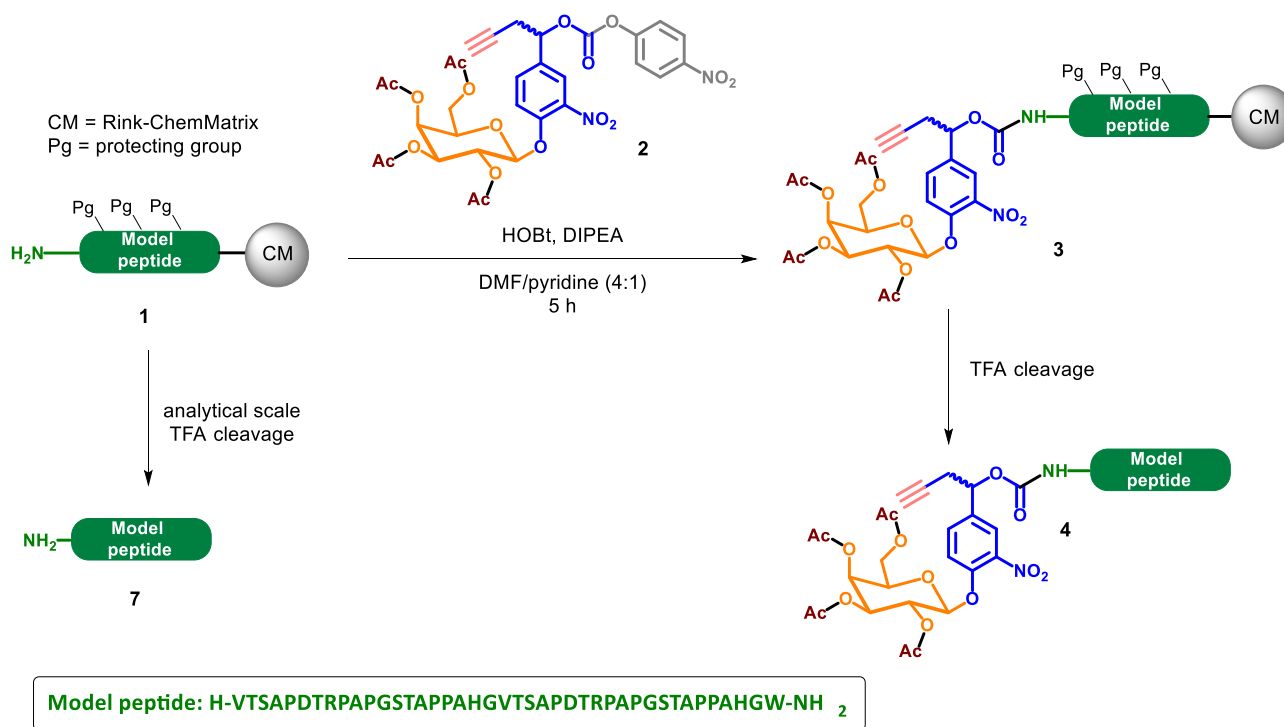
- Compound **7**

Sequence: H-APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSW-NH₂

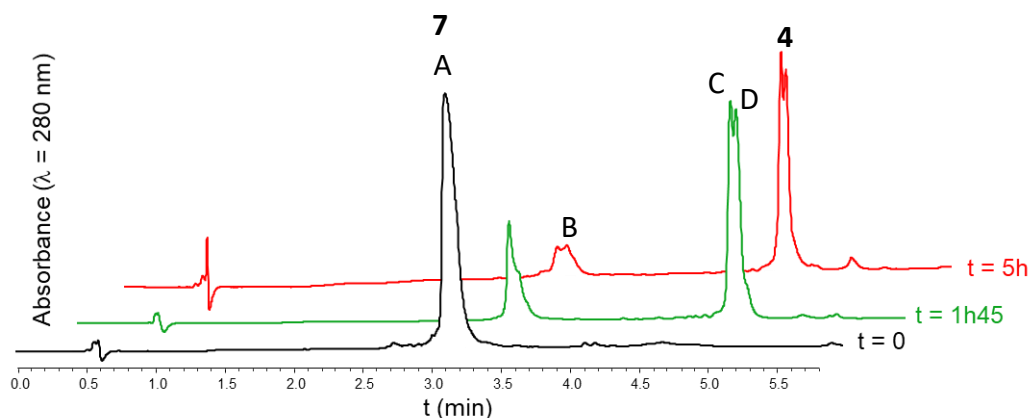
ESI-MS (*m/z*): [M] calcd. for C₁₇₁H₂₆₃N₅₃O₅₅: 3941.3, found: 3940.7 (average mass, deconvoluted)

HPLC analysis: t_R = 3.17 min (Chromolith, gradient: 5-50% B over 5 min)

3-1- Linker introduction on model peptide



To peptidyl resin **1** (15 μ mol) was added a solution of carbonate **2** (15.8 mg, 1.5 equiv.) in 4:1 DMF/pyridine (1.42 mL, 10.5 mM peptide concentration) then DIEA (5.6 μ L 1.8 equiv.) and hydroxybenzotriazole hydrate (HOBt) (3.1 mg, 1.5 equiv.). The mixture was stirred at room temperature for 5 h then the resin was extensively washed with NMP and DCM. Capping of potential unreacted amine groups was achieved through treatment with acetic anhydride as described in protocol PS1. The reaction was monitored by Kaiser Test [2] and analytical scale TFA cleavage (protocol PS3).



Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (3.16)	3941.2	3940.7	7
B (3.18)	3983.3	3982.7	Ac-(Muc1) ₂ -W-NH ₂
C (4.72)	4504.8	4504.1	4-1
D (4.77)	4504.8	4504.1	4-2

Supplementary figure S1: HPLC traces (Chromolith, gradient: 5-50% B over 5 min, $\lambda = 280$ nm) and MS analysis monitoring of linker introduction on **1** (average masses, deconvoluted)

- **4** (mixture of two diastereoisomers)

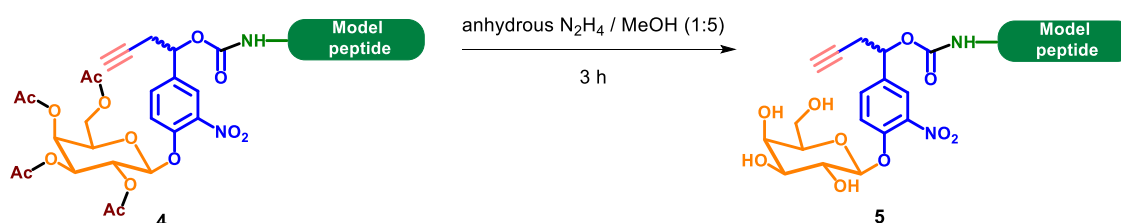
ESI-MS (m/z): [M] calcd. for C₁₉₆H₂₈₈N₅₄O₆₉: 4504.8, found: 4504.1 (for both **4-1** and **4-2**) (average mass, deconvoluted)

HPLC analysis: $t_R = 4.70$ (**4-1**) and 4.74 min (**4-2**) (Chromolith, gradient: 5-50% B over 20 min)

3-2- Acetate groups deprotection

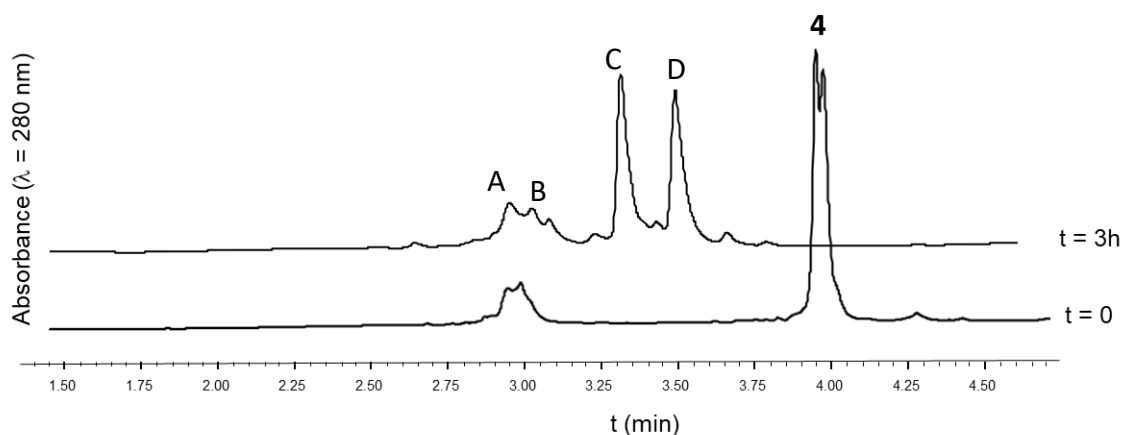
Acetate deprotection was first attempted on solid support using different protocols: N₂H₄ and MeONa were used along with different solvents (DMF, MeOH, mixture of both) and at different concentrations (6 mM for MeONa, and from 1 M to 5 M for hydrazine). However, the reaction was never complete, leading to a mixture of partially and completely deacetylated compounds in addition to numerous unidentified products.

Contrastingly, the reaction was complete when performed in solution after TFA cleavage.



TFA cleavage was performed on peptidyl resin **3** following protocol PS3 affording **4**. Note that we found that prolongation of the treatment for more than 2 h led to by-products attributed to the reduction of the nitro group into amine (average M calcd. 4474.8 for C₁₉₆H₂₉₀N₅₄O₆₇, found 4473.9) and beta-galactoside hydrolysis (average M calcd. 4174.5 for C₁₈₂H₂₇₀N₅₄O₆₀, found 4173.8), respectively.

A 20% anhydrous hydrazine solution in methanol was then added to crude **4** (1 mg/mL). The reaction was stirred for 3 h at room temperature. pH was then adjusted to 2 by dropwise addition of 10% aqueous TFA before semi-preparative HPLC purification. **5-1** and **5-2** were well separated by RP-HPLC, and each diastereoisomer was purified independently



Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (2.92)	3941.2	3940.7	7
B (3.03)	3983.3	3982.7	Ac-(Muc1) ₂ -W-NH ₂
C (3.33)	4336.6	4336.1	5-1
D (3.51)	4336.6	4336.0	5-2

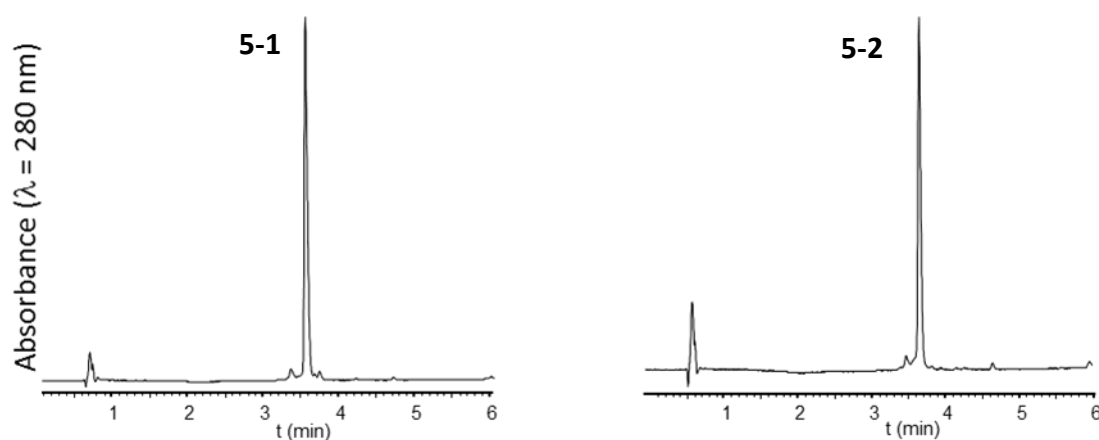
Supplementary figure S2: HPLC traces (Chromolith, gradient: 5-50% B over 5 min) and MS analysis monitoring of deacetylation of **4** to give **5** (average masses, deconvoluted)

- Compound **5** (mixture of two diastereoisomers)

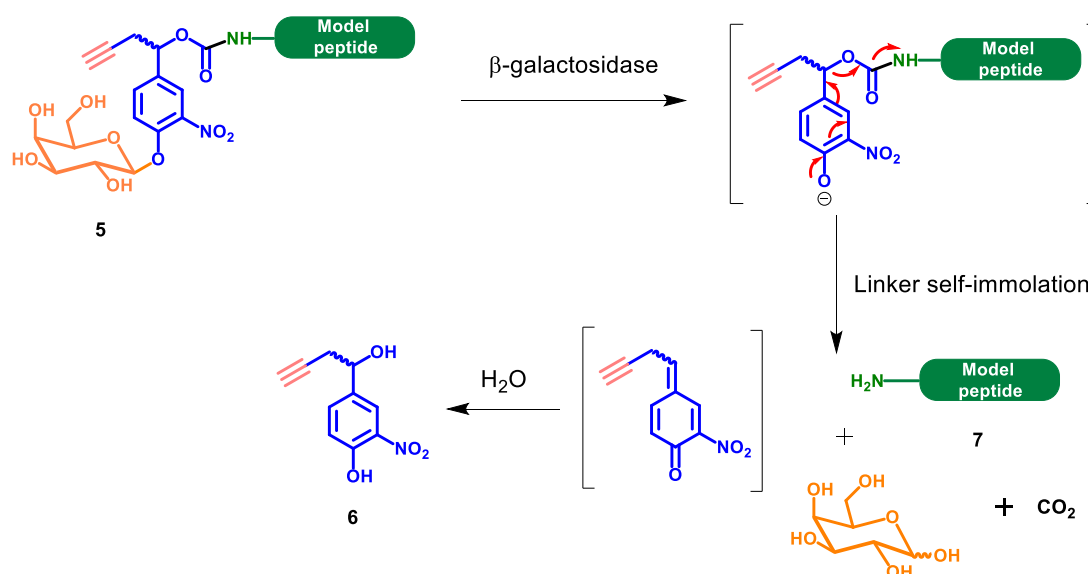
ESI-MS (m/z): [M] calcd. for C₁₈₈H₂₈₀N₅₄O₆₅: 4336.6, found: 4336.1 (**5-1**) and 4336.0 (**5-2**) (average masses, deconvoluted)

HPLC analysis: t_R = 3.52 min (**5-1**) and 3.71 min (**5-2**) (Chromolith, gradient: 20-50% B over 5 min)

HPLC purification: Nucleosil C18, gradient: 18-28% B over 20 min, 11% yield for **5-1**, 13% for **5-2**

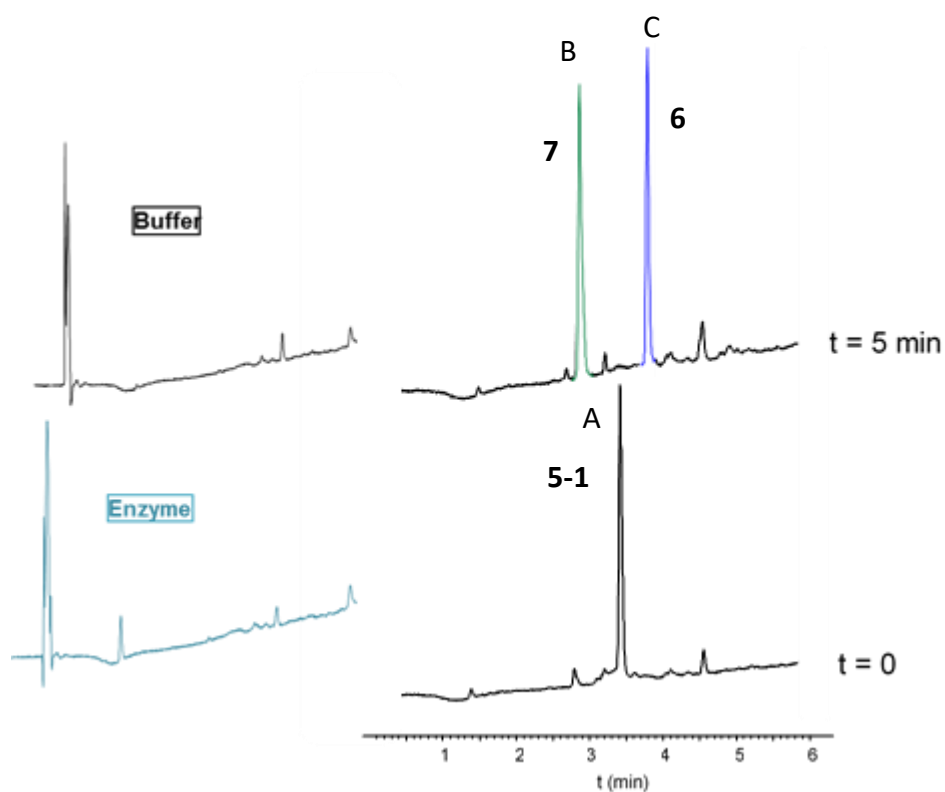


Supplementary figure S3: HPLC traces of purified **5-1** and **5-2**

3-3- Solution phase β -galactosidase-mediated linker cleavage

Cleavage of the linker of **5** was carried out as described [1] with commercial β -galactosidase from *Escherichia coli* (E.C. 3.2.1.23, 766 units/mg protein (biuret), suspension in 50% glycerol, 10 mM Tris buffer and 10 mM magnesium chloride, pH 7.0). **5** (100 μM) was incubated at 37 $^\circ\text{C}$ with the enzyme (0.4 U/10 nmol) in 20 mM phosphate buffer, pH 7.3. Linker cleavage was monitored by analytical HPLC.

Note that the enzymatic assay was at first tested on one of the isomers (**5-1**) then tested on the second isomer (**5-2**) giving identical reaction in term of kinetics and nature of the released compounds (data not shown).



Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (3.42)	4336.6	4336.1	5-1
B (2.94)	3941.2	3940.7	7
C (3.81)	207.1	207.1	6

Supplementary figure S4: HPLC traces (Chromolith, gradient: 5-50% B over 5 min, $\lambda = 280$ nm) and MS monitoring of enzymatic hydrolysis of **5-1** using optimized conditions (average masses, deconvoluted, for peaks A and B)

- Compound **7**

ESI-MS (m/z): [M] calcd. for $C_{171}H_{263}N_{53}O_{55}$: 3941.2, found: 3940.7 (average mass, deconvoluted)

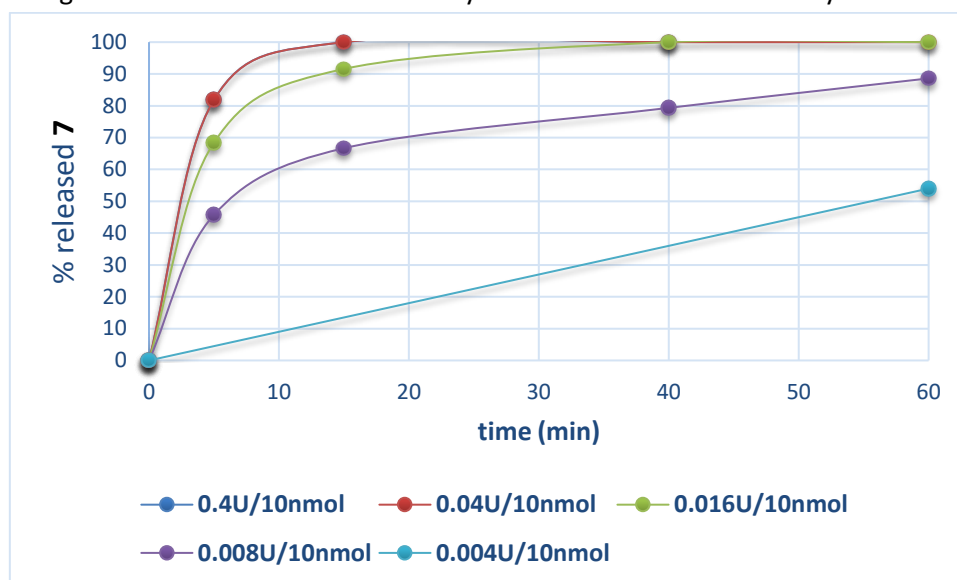
HPLC analysis: $t_R = 2.94$ min (Chromolith, gradient: 5-50% B' over 5 min)

3-4- Optimization of the solution phase enzymatic cleavage

Relative rates were determined by HPLC peak integration at $\lambda = 280$ nm, taking into account the molar absorption coefficient of Trp [3] and the 1st generation linker (see p S28) at 280 nm: $\epsilon_{\text{Trp}} = 5500 \text{ L.mol}^{-1}.\text{cm}^{-1}$; $\epsilon_{\text{linker}} = 6938 \text{ L.mol}^{-1}.\text{cm}^{-1}$.

- Reduction of enzyme amount**

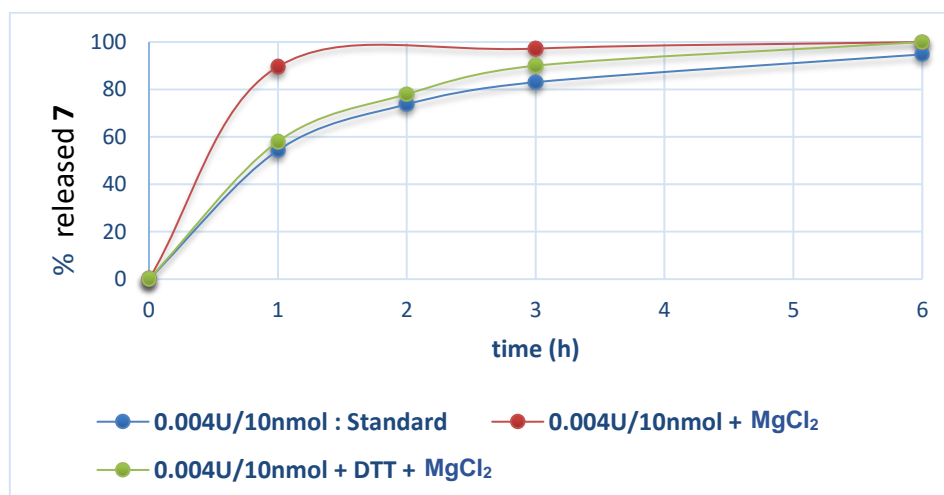
Enzyme cleavage in solution still works effectively even with 100 times less enzyme.



Supplementary figure S5: Enzymatic release of model peptide **7** from compound **5-1** using different concentrations of β -galactosidase

- Buffer optimization**

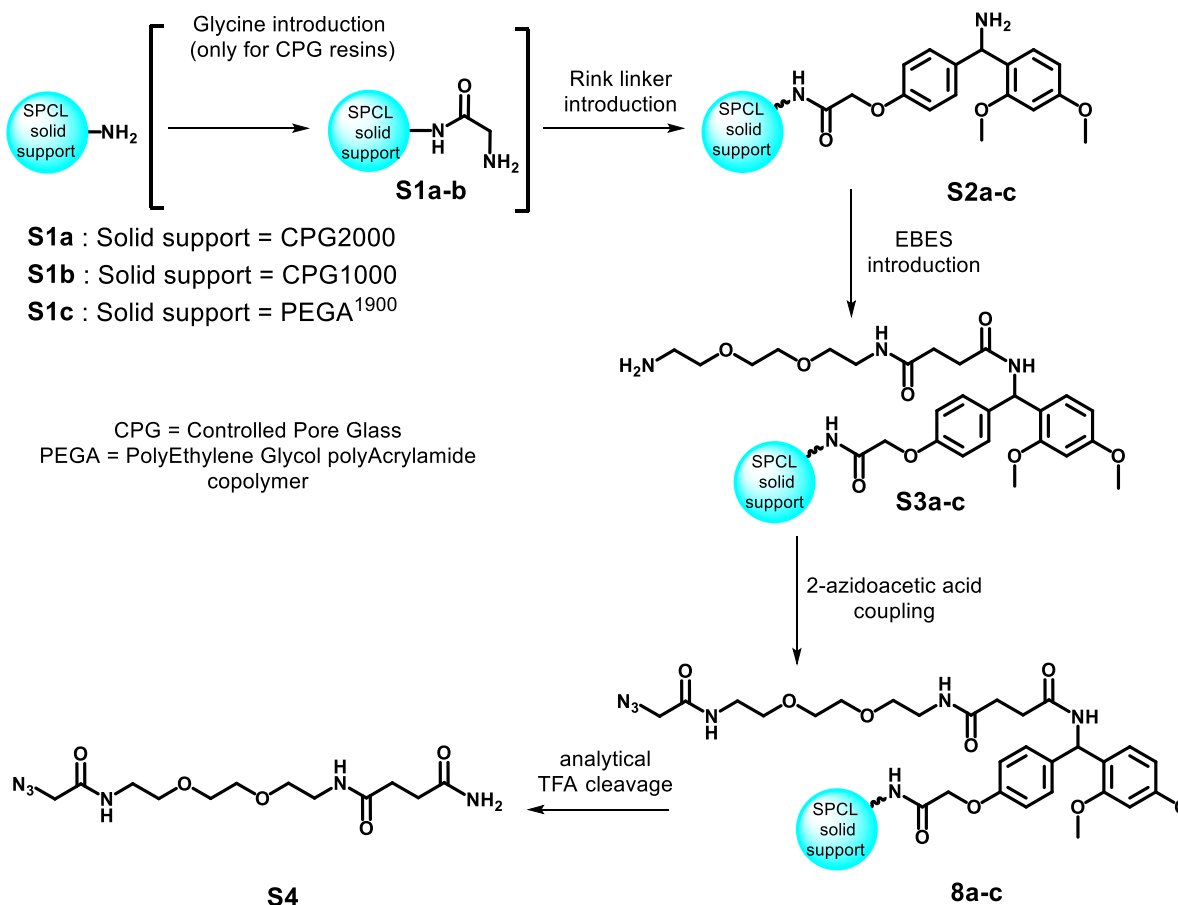
Most *in vitro* applications of β -galactosidase are implemented using 0.1 mM dithiothreitol (DTT) and 0.1 mM MgCl_2 salts in the reaction buffer. Therefore, we tried two different buffers: one with both and a second only with magnesium. The latter has been found to be the most effective.



Supplementary figure S6: Enzymatic release of model peptide **7** from compound **5-1** using different buffers

3-5- Solid phase enzymatic cleavage

Synthesis of azide-functionalized solid supports 8a-c



- PEGA resin:

25 μmol of aminomethyl PEGA¹⁹⁰⁰ (63 mg, 400 $\mu\text{mol/g}$) was washed with NMP (5 mL, 1 min). Fmoc-Rink-OH (135 mg, 10 equiv.) in NMP (1.5 mL), HATU (90 mg, 9.5 equiv.) in NMP (2 mL) and DIEA (87 μL , 20 equiv.) were added and the reactor was stirred for 6 h followed by extensive washing with NMP. Capping of potential unreacted amine groups was achieved by treatment with acetic anhydride as described in protocol PS1. Fmoc group was deprotected by three successive treatments with 20% piperidine in NMP for 3 min. The same procedure was repeated for the coupling of Fmoc-EBES-OH (117.6 mg, 10 equiv.) and $\text{N}_3\text{-CH}_2\text{-COOH}$ (19 μL , 10 equiv.) to obtain **8c** in 91% elongation yield. A small amount of the resin was cleaved for analytical purpose affording **S4** (protocol PS3 small peptides cleavage).

- CPG solid supports:

25 μmol of aminopropyl CPG solid support (CPG2000: 641 mg, 39 $\mu\text{mol/g}$; CPG1000: 281 mg, 89 $\mu\text{mol/g}$) were washed with NMP (5 mL, 1 min). Fmoc-Gly-OH (112 mg, 15 equiv.) in NMP (1.5 mL), PyAOP (261 mg, 20 equiv.) in NMP/DCM 65/35 (2 mL, final NMP/DCM ratio 8:2) were added and the reactor was stirred for 6 h followed by extensive washing with NMP. Capping of potential unreacted amine groups was achieved by treatment with acetic anhydride as described in protocol PS1. Fmoc

group was deprotected by three successive treatments with 20% piperidine in NMP for 3 min. The same procedure was repeated for the coupling of Fmoc-Rink-OH (202 mg, 15 equiv.) Fmoc-EBES-OH (176 mg, 15 equiv.) and $\text{N}_3\text{-CH}_2\text{-COOH}$ (28 μL , 15 equiv.) **8a** was obtained in 92% elongation yield (CPG2000) and **8b** in 75% elongation yield (CPG1000). A small amount of the resin was cleaved for analytical purpose affording **S4** (protocol PS3 small peptides cleavage).

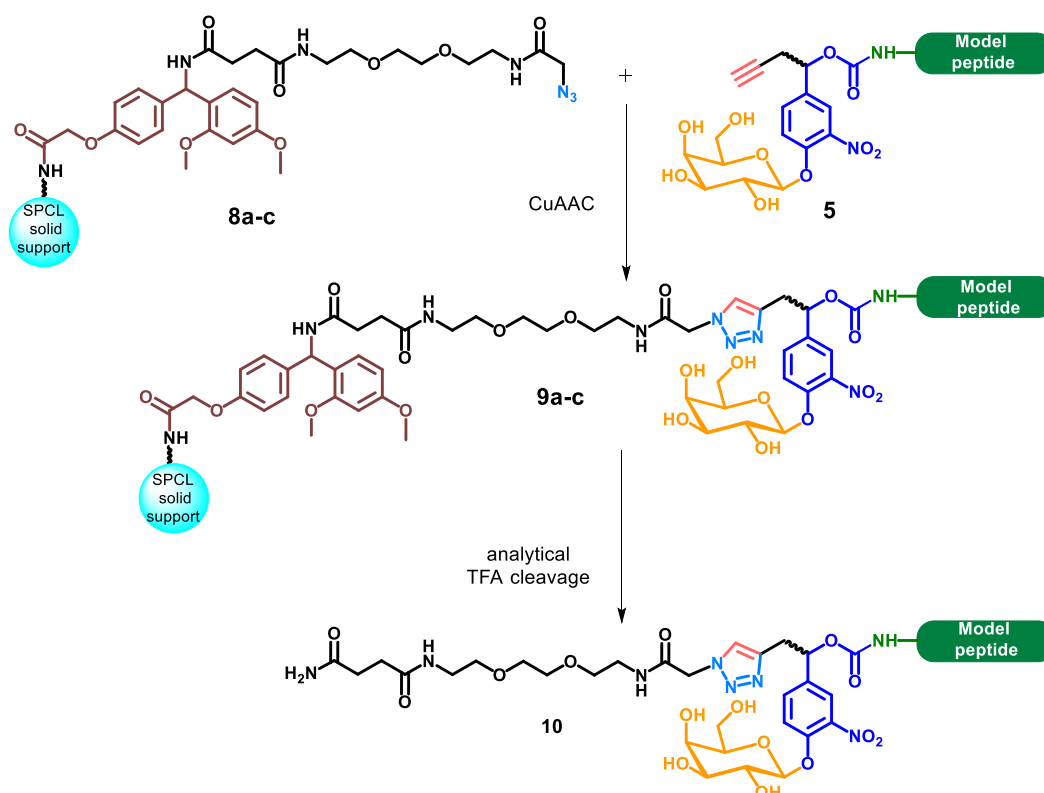
Elongation yields were calculated by UV titration of the fluorenylmethyl-piperidine adduct after Fmoc deprotection of **S3a-c** compared to the initial resin loading. In the case of CPG solid supports, this yield was highly increased by using excess of PyAOP and 20% DCM in NMP as the reaction's solvent, and by coupling an extra Gly residue before the Rink amide linker.

- Compound **S4**

ESI-MS (m/z): [M] calcd. for $\text{C}_{12}\text{H}_{22}\text{N}_6\text{O}_5$: 330.2, found: 330.2

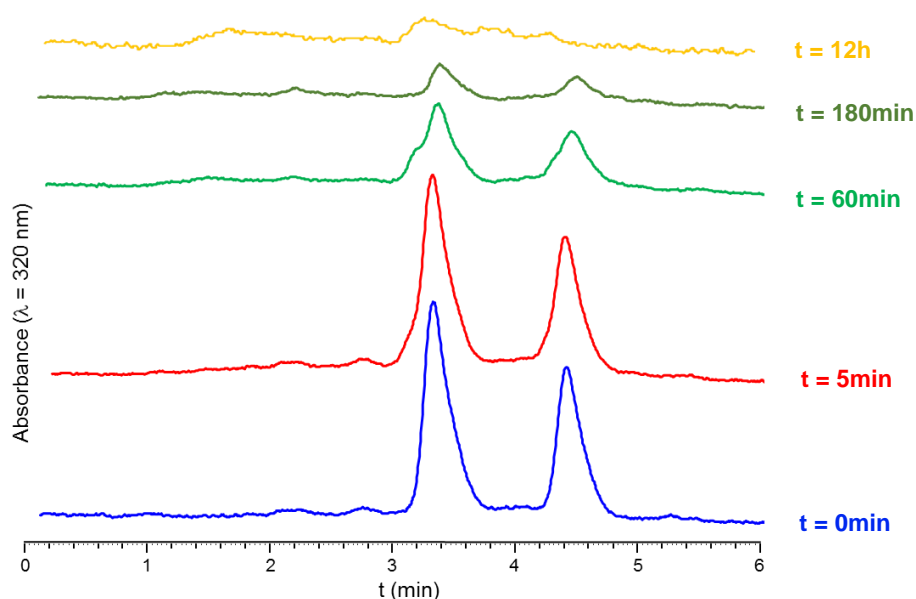
HPLC analysis: t_R = 2.10 min (Chromolith, gradient: 5-50% B over 5 min)

Copper(I)-catalyzed azide/alkyne cycloaddition to graft linker-functionalized peptide (5) on solid supports (8a-c)



A solution of alkynyl linker-functionalized peptide **5** (0.35 μmol) dissolved in 280 μL of a 1:1 mixture of NMP and 200 mM HEPES buffer pH 7.5 was added to a 2 mL microcentrifuge tube containing the azide-functionalized solid support **8a-c** (6 equiv.). To the suspension were subsequently added a 1 M aqueous aminoguanidine solution (10.5 μL , 30 equiv.) and a 1 M aqueous *tert*-butanol solution (10.5 μL , 30 equiv.). The tube was sealed with a rubber septum and the resulting mixture was further deoxygenated through several successive vacuum (15mbar) / argon cycles. Then, a mixture containing copper(I)

bromide-dimethylsulfide complex (30 equiv.) and THPTA (40 equiv.) dissolved under argon in 35 μL of deoxygenated NMP was added, the resulting suspension was further deoxygenated and was stirred overnight at 37 °C. The supernatant was analysed by HPLC to check the total consumption of the alkynyl linker-functionalized peptide **5**. Capping of remaining solid-supported azide groups was performed by adding propargylic alcohol (100 equiv.) and stirring the mixture for an additional hour at 37 °C. The resin was then extensively washed with NMP, de-ionized water, then repeatedly treated (3x) with a pH 7 aqueous buffer containing 6 M guanidinium chloride, 0.1 M EDTA and 0.1 M sodium dihydrogenophosphate for a few minutes then drained and finally extensively washed with deoxygenated water. A small amount of the peptide-resin was cleaved for analytical purpose affording **10** (protocol PS3).

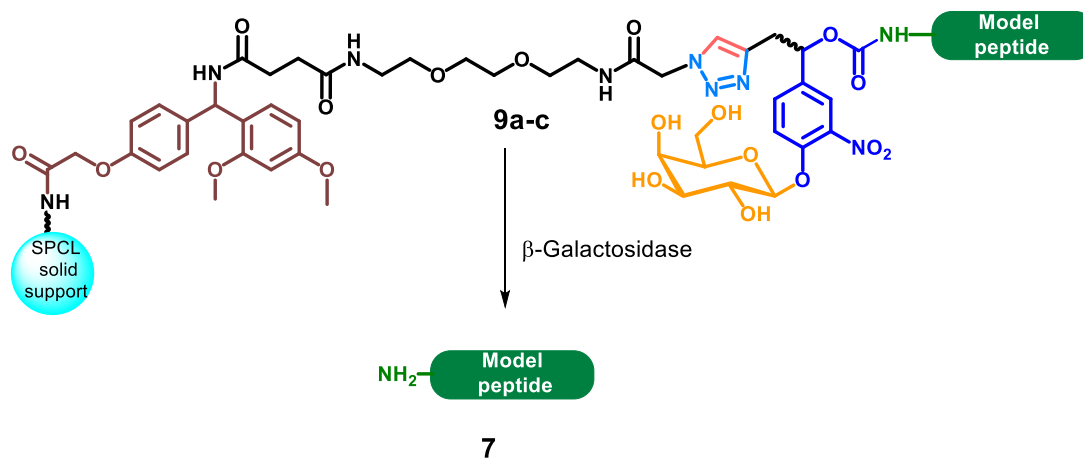


Supplementary figure S7: HPLC traces (Chromolith, gradient: 5-50% B over 5 min) of CuAAC reaction monitoring showing the disappearance of **5-1** and **5-2** in the supernatant

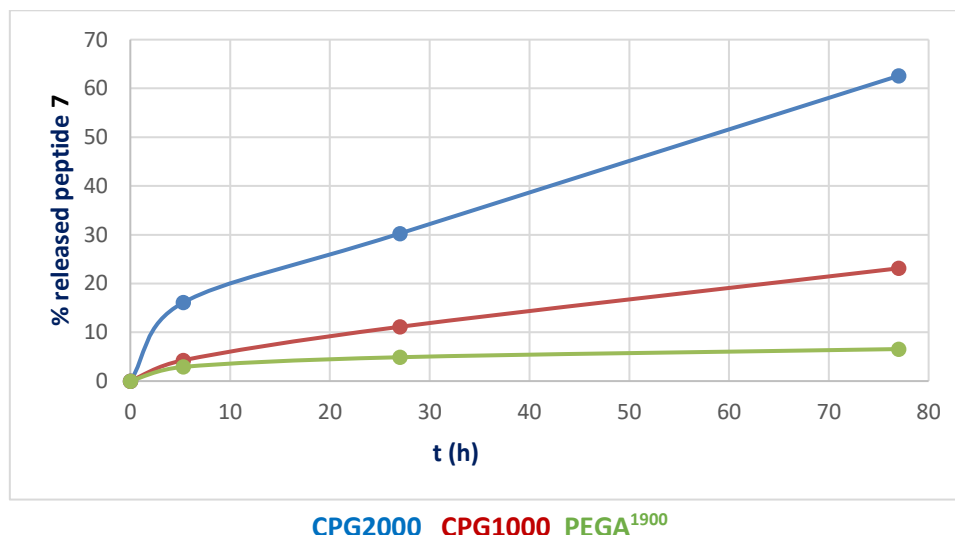
- Compound **10** (mixture of two diastereoisomers)

ESI-MS (m/z): [M] calcd. for $\text{C}_{200}\text{H}_{302}\text{N}_{60}\text{O}_{70}$: 4667.0, found: 4665.9 (average mass, deconvoluted)

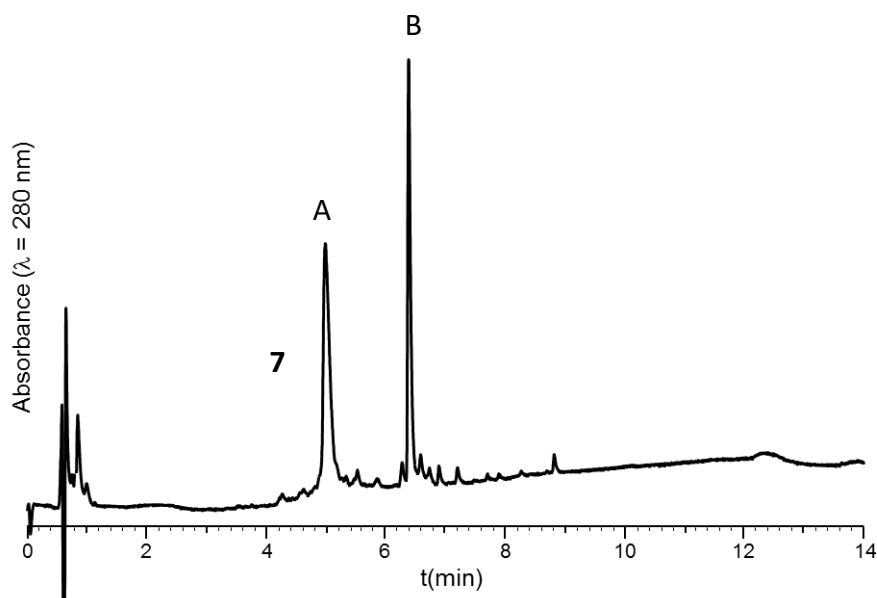
HPLC analysis: t_R = 7.26 min (Aeris Widespore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C). The two diastereoisomers were indistinguishable under these conditions.

Solid phase β -galactosidase-mediated linker cleavage

Solid supported peptides incorporating the enzyme-labile linker **9a-c** (CPG2000, CPG1000, PEGA¹⁹⁰⁰) were incubated at 37 °C with the enzyme (4 U /10 nmol) in 20 mM phosphate buffer at pH 7.3 (100 μ M solid-supported peptide concentration). Release of model peptide **7** was monitored by analytical HPLC. Amounts of released **7** were determined by HPLC peak integration at $\lambda = 280$ nm, using tryptophan as an external calibrant, and compared to the initial quantity of peptide **5** grafted on the solid supports, as also determined by HPLC peak integration at $\lambda = 280$ nm before the start of the CuAAC reaction. Molar absorption coefficients of **5** and **7** were determined from the molar absorption coefficient of Trp and of the 1st generation linker (see p S28) at 280 nm: $\epsilon_{\text{Trp}} = 5500 \text{ L.mol}^{-1}.\text{cm}^{-1}$ [3]; $\epsilon_{\text{linker}} = 6938 \text{ L.mol}^{-1}.\text{cm}^{-1}$.



Supplementary figure S8: Enzymatic release of model peptide **7** from solid supports **9c** under non-optimized conditions



Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (5.80)	3941.2	3940.7	7
B (6.74)	-	883.3	Non peptidic compound*

*: Note that peak B is also observed when treating the starting azide solid support **8** under the same conditions, and thus does not correspond to a peptide.

Supplementary figure S9: Typical HPLC trace and MS analysis (average masses, deconvoluted, for peak A) of solid supported enzymatic hydrolysis with β -galactosidase (corresponding to 77 h cleavage of CPG2000-supported peptide **9c**)

- Compound **7**

ESI-MS (m/z): [M] calcd. for $C_{171}H_{263}N_{53}O_{55}$: 3941.2, found: 3940.7 (average mass, deconvoluted)

HPLC analysis: t_R = 5.80 min (Aeris Widepore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C)

4- Optimization of first generation linker enzymatic cleavage on CPG2000-supported peptide 9c

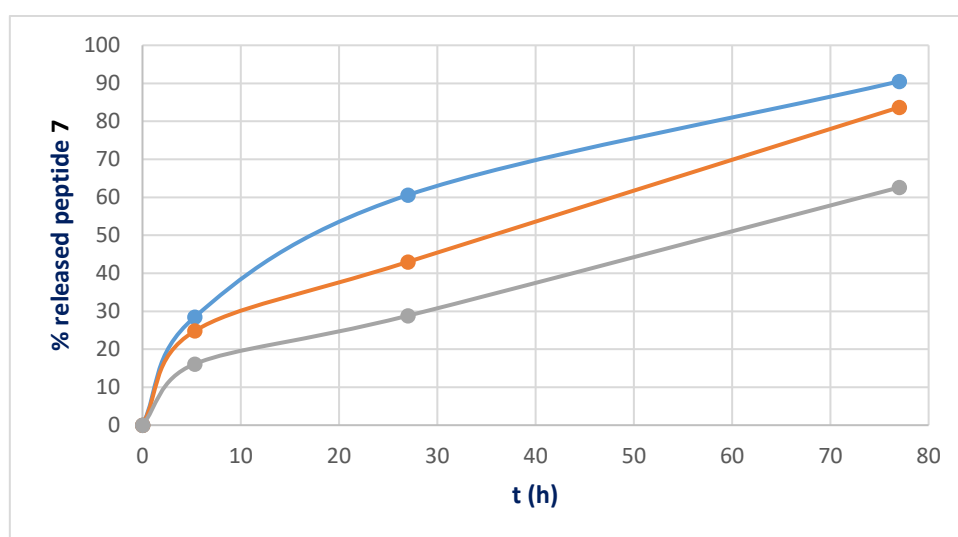
4-1- Addition of DTT and MgCl₂ salts in the reaction buffer

As in the case of solution phase cleavage study, three buffers have been investigated, with CPG2000 solid support **9c** which gave the best results in the initial solid support screening.

-20 mM phosphate buffer, pH 7.3

-20 mM phosphate buffer, pH 7.3 + 0.1 mM DTT + 0.1 mM MgCl₂

-20 mM phosphate buffer, pH 7.3 + 0.1 mM MgCl₂



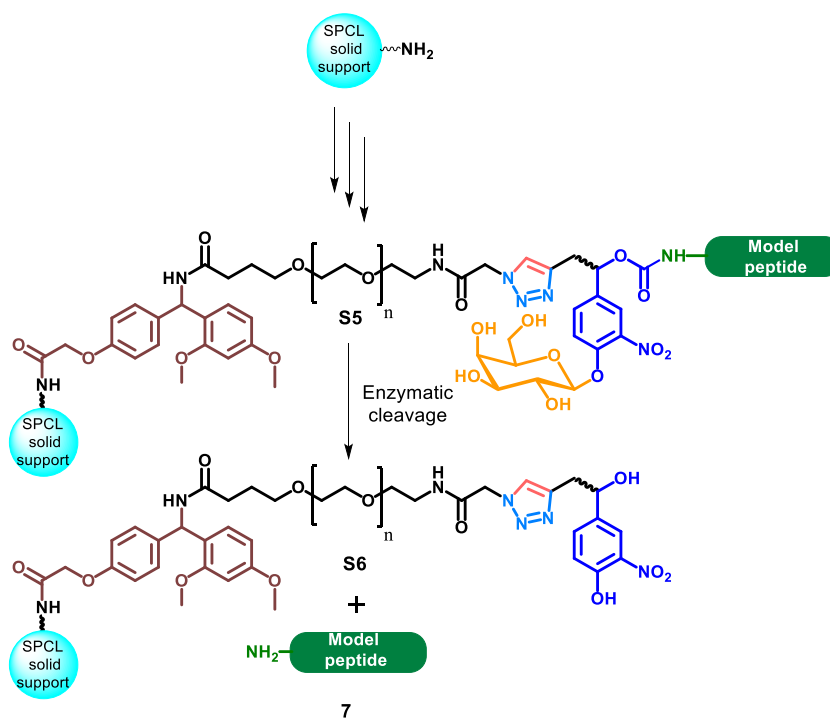
Supplementary figure S10: Enzymatic release of model peptide **7** from CPG2000 solid support **9c** using different buffers

Similar to what was observed in solution, the buffer including 0.1 mM MgCl₂ gave significantly faster reactions, and was kept in the following experiments.

4-2-Use of larger amount of the enzyme

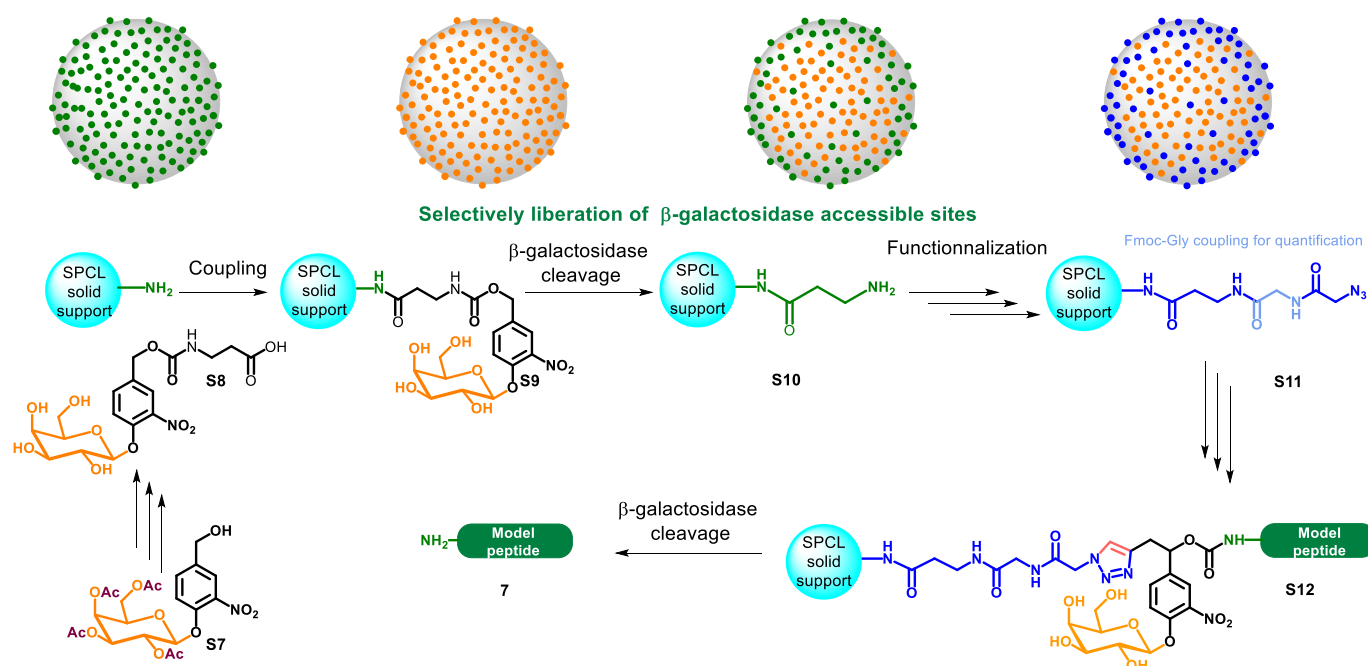
Conditions used: 20 mM phosphate buffer pH 7.3 + 0.1 mM MgCl_2 , 40 enzyme units/10 nmol of peptide (100 μM initial concentration), 37 °C. However, no significant increase in enzymatic cleavage rate was noticed.

4-3- Use of a larger spacer: PEG3000



Coupling of Fmoc-NH-PEG3000-COOH (225 mg, 15 equiv., polydisperse) on NH₂-Rink-Gly-CPG2000 (5 μmol) was performed as described for introducing the EBES linker in solid support **9c**. Again, no significant increase in enzymatic cleavage rate was noticed.

4-4- β -galactosidase-mediated “shaving” of the CPG beads in order to install azide groups only at sites easily accessible by the enzyme



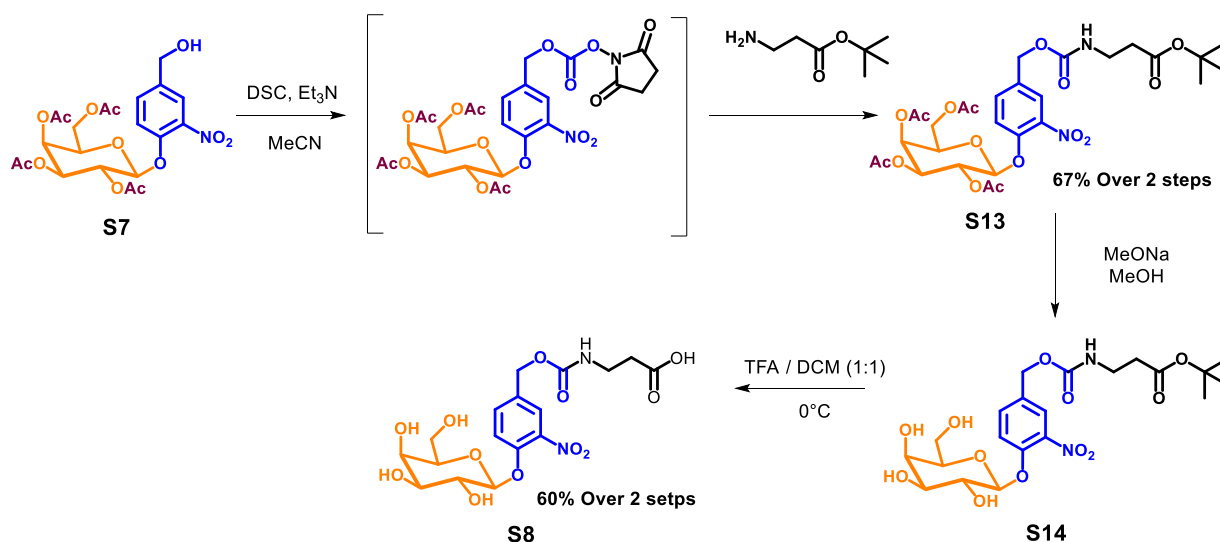
Supplementary figure S11: β -galactosidase-mediated “shaving” of the CPG beads in order to install azide groups only at sites easily accessible by the enzyme. Top: schematic representation of beads, small dots representing possible grafting sites (color code corresponds to the ones of chemical drawing below).

We attempted to selectively graft the peptide only on sites of the solid support easily accessible by the enzyme. To do so, we synthesised a β -alanine derivative **S8** which amine was protected by a β -galactosidase-labile group derived from the first generation enzyme cleavable linker and then introduced it on the CPG2000 solid support. Next, the resulting solid support was treated with β -galactosidase. This liberated amine groups only on the sites where the enzyme had an easy access (c.a.~ 30% of the total sites*), leaving *N*-protected the less accessible sites. The freed sites were functionalized with an azide group through successive couplings of Fmoc-Gly-OH then azidoacetic acid, next, the model peptide bearing the alkynyl linker-functionalized peptide **S12** was introduced by CuAAC on the sites freed by the first enzymatic cleavage, and a second enzymatic cleavage was performed to release the model peptide. Disappointingly, this did not significantly improve the linker cleavage kinetics.

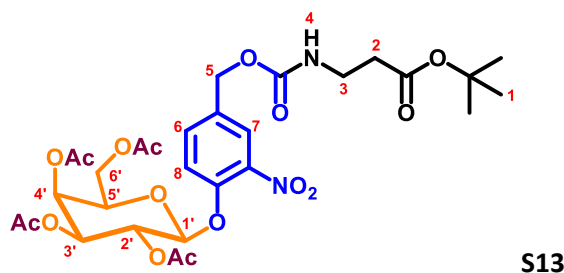
*: Fmoc-Gly-OH was coupled on free amine groups after the first enzymatic treatment, allowing to quantify the amount of liberated amines groups through UV titration of the Fmoc deprotection.

- **Synthesis of S8**

S8 was synthesized in four steps from previously described **S7** [1].



➤ **Synthesis of S13**



Alcohol **S7** (130 mg, 0.26 mmol) was dissolved in MeCN (0.65 mL, 0.4 M alcohol concentration) and Et₃N (176 μ L, 5 equiv.) was added. Then, *N,N'*-disuccinimidyl carbonate (130 mg, 2 equiv.) in MeCN (1.3 mL, 0.4 M carbonate concentration) was added dropwise and the reaction was stirred for 1.5 h at room temperature. Then, a mixture of H- β -Ala-OtBu (47 mg, 3.5 equiv.) and Et₃N (176 μ L, 5 equiv.) was added and the reaction was stirred again for 1.5 h at room temperature. Next EtOAc (60 ml) was added and the reaction was washed successively with saturated NaHCO₃ (x 3), 1 M HCl (x 2) and saturated NaCl aqueous solutions, dried over MgSO₄, and concentrated under reduced pressure. The product was purified by flash column chromatography (eluent: 50:50 petroleum ether/AcOEt) affording **S13** as a white amorphous solid (117 mg, 67% over two steps).

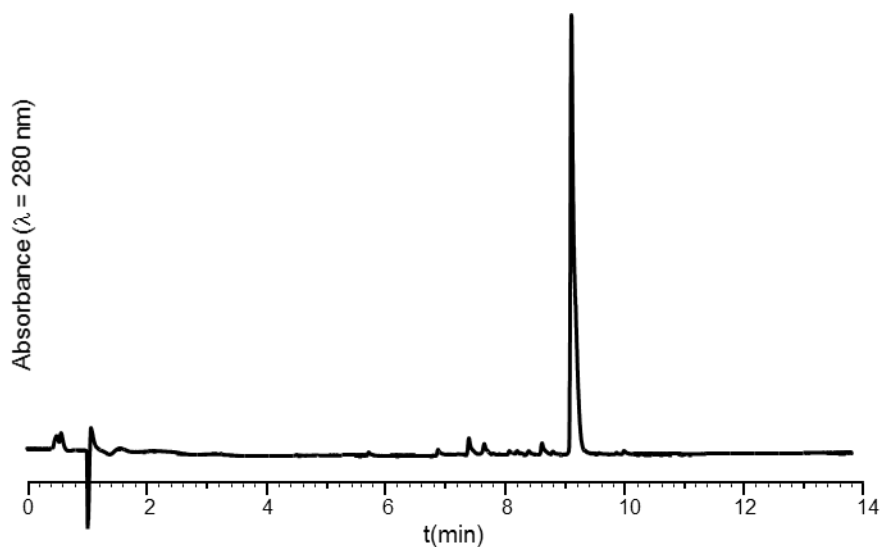
¹H NMR (600 MHz, CDCl₃): δ 7.77 (d, *J* = 2.2 Hz, 1H, H₇), 7.49 (dd, *J* = 8.5, 2.2 Hz, 1H, H₆), 7.32 (d, *J* = 8.5 Hz, 1H, H₈), 5.52 (dd, *J* = 10.4, 7.9 Hz, 1H, H_{2'}), 5.45 (dd, *J* = 3.4, 1.2 Hz, 1H, H_{4'}), 5.36 (bt, *J* = 5.3 Hz,

^1H , H_4), 5.09 (dd, $J = 10.5, 3.4$ Hz, 1H, $\text{H}_{3'}$), 5.03-5.07 (m, 3H, $\text{H}_{1'}$ and H_5), 4.23 (dd, $J = 11.4, 7.0$ Hz, 1H, $\text{H}_{6'a}$), 4.15 – 4.19 (m, $J = 11.4, 6.1$ Hz, 1H, $\text{H}_{6'b}$), 4.05 (ddd, $J = 7.0, 6.1, 1.2$ Hz, 1H, $\text{H}_{5'}$), 3.35 – 3.44 (m, 2H, H_3), 2.39-2.46 (m, 2H, H_2), 2.17 (s, 3H, OAc), 2.10 (s, 3H, OAc), 2.05 (s, 3H, OAc), 1.99 (s, 3H, OAc), 1.43 (s, 9H, H_1).

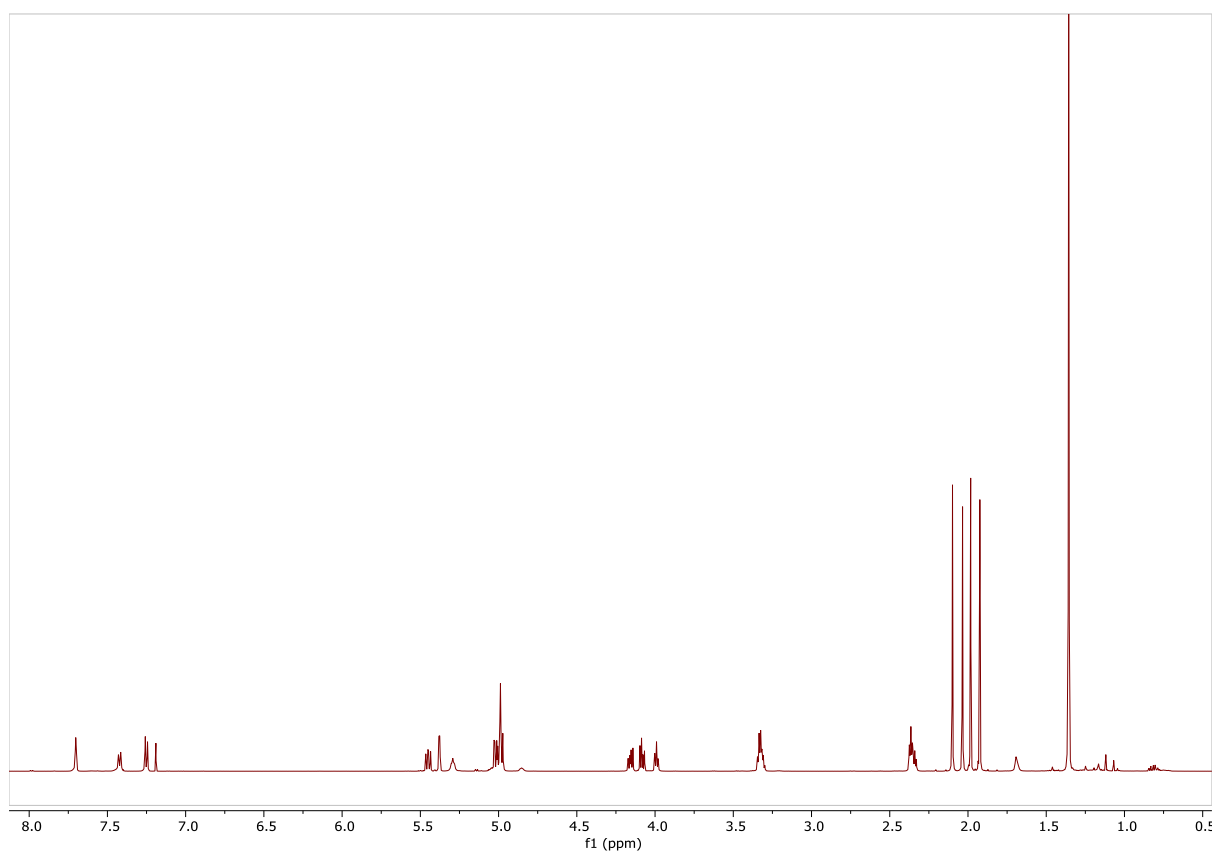
^{13}C NMR (151 MHz, CDCl_3): δ 171.7, 170.3, 170.2, 170.1, 169.4, 155.8, 148.9, 141.2, 133.2, 133.0, 124.6, 119.8, 100.8, 81.2, 71.4, 70.5, 67.8, 66.7, 64.7, 61.4, 36.8, 35.4, 28.1, 20.66, 20.63 (2C), 20.55.

ESI-MS (m/z): [M] calcd. for $\text{C}_{29}\text{H}_{38}\text{N}_2\text{O}_{16}$: 670.2 found: 670.2

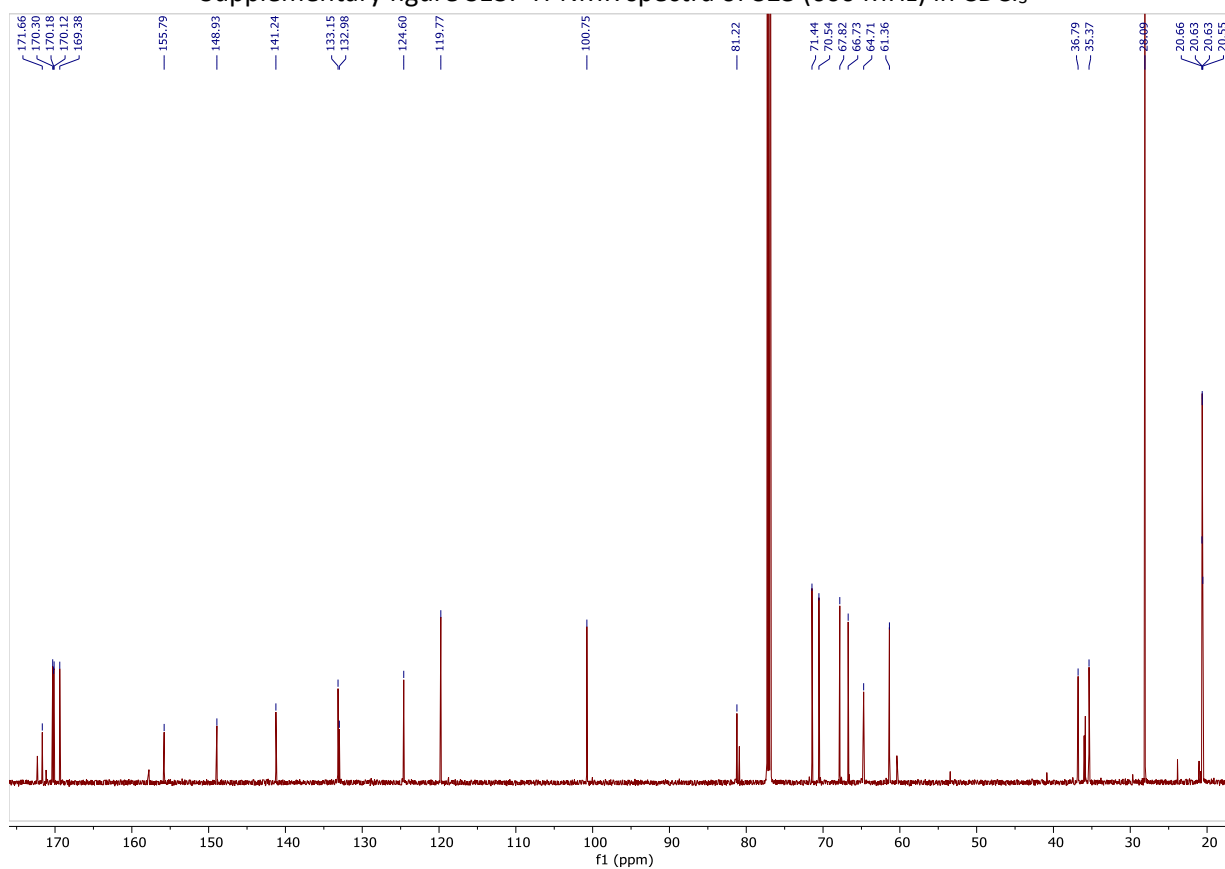
HPLC analysis: $t_R = 9.22$ min (Aeris Widespore XB-C18 2, gradient: 3-90% B' over 15 min, 60 °C)



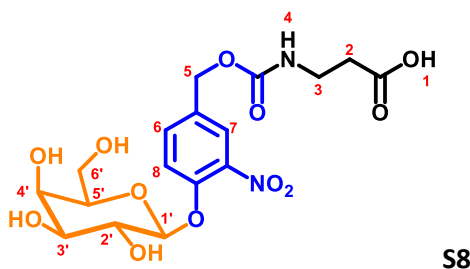
Supplementary figure S12: HPLC trace of purified **S13**



Supplementary figure S13: ^1H -NMR spectra of **S13** (600 MHz) in CDCl_3



Supplementary figure S14: ^{13}C -NMR spectrum of **S13** (151 MHz) in CDCl_3

➤ Synthesis of **S8**

S13 (88 mg, 0.131 mmol) was dissolved in 1.31 mL anhydrous MeOH (0.1 M galactoside concentration) under an argon atmosphere. 6 μ L of a 4.4 M MeONa solution in MeOH (0.2 equiv.) was then added and the reaction was stirred for 1 h at room temperature. The reaction was then quenched by the addition of a Dowex 50WX8 H⁺ resin (treated with 1 M HCl for 48 h and washed thoroughly with water and methanol prior to use) until pH was dropped to \sim 7, as judged by spotting aliquots on a wet pH paper, and corresponding to a change of reaction colour from yellow to colourless. Next the reaction was filtrated and concentrated affording 60 mg of crude **S14** which was not further purified nor characterized. **S14** was immediately dissolved in 1.1 mL of 1:1 TFA/DCM (0.12 mM galactoside concentration) and the reaction was stirred at 0 °C for 30 min. Solvents were evaporated under reduced pressure, and the resulting yellowish solid was purified by semi-preparative HPLC and lyophilized to give **S8** as a white amorphous solid (35 mg, 60% over two steps).

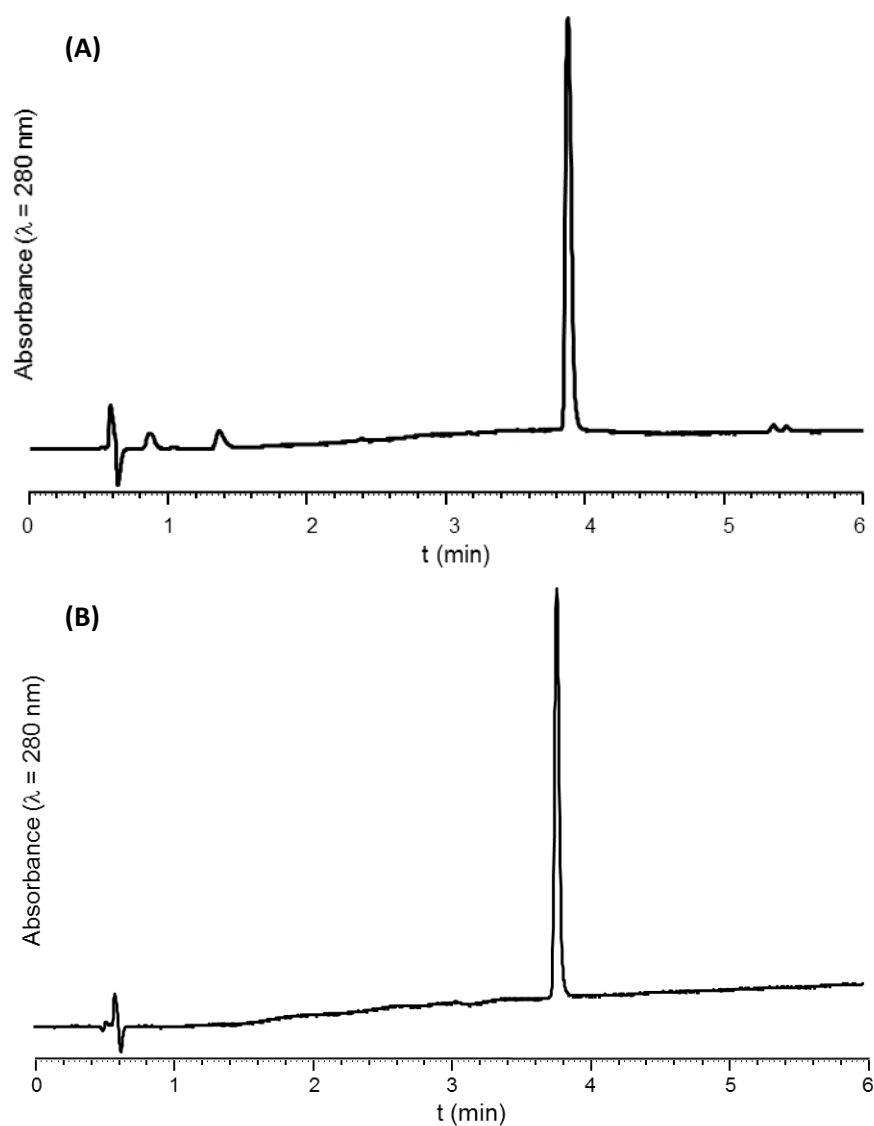
¹H NMR (600 MHz, methanol-*d*₄): δ 7.71 (bs, 1H, H₇), 7.48 (bd, *J* = 8.6 Hz, 1H, H₆), 7.35 (d, *J* = 8.6 Hz, 1H, H₈), 5.07 (s, 2H, H₅), 5.04 (d, *J* = 7.7 Hz, 1H, H_{1'}), 3.90 (d, *J* = 3.4 Hz, 1H, H_{4'}), 3.84 (dd, *J* = 9.7, 7.6 Hz, 1H, H_{2'}), 3.80 – 3.72 (m, 2H, H_{5'} and H_{6'}), 3.59 (dd, *J* = 9.7, 3.4 Hz, 1H, H_{3'}), 3.30 – 3.23 (m, 2H, H₃), 2.45 – 2.34 (m, 2H, H₂).

¹³C NMR (151 MHz, methanol-*d*₄): δ 175.2, 158.3, 150.9, 141.8, 134.2, 132.7, 125.2, 118.7, 102.9, 77.2, 74.7, 71.8, 70.0, 65.7, 62.3, 37.7, 35.0.

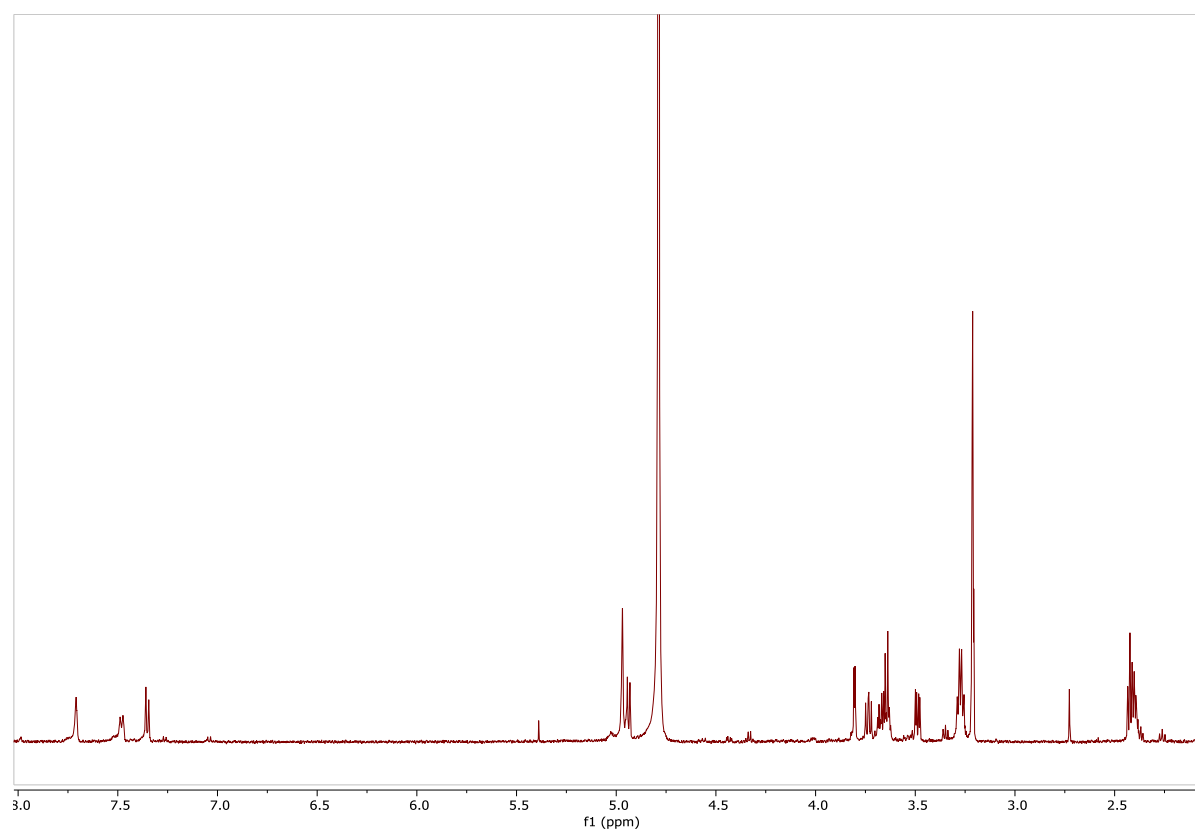
ESI-MS (*m/z*): [M] calcd. for C₂₉H₃₈N₂O₁₆: 446.1 found: 446.1

HPLC analysis: *t*_R = 3.90 min (Chromolith, gradient: 20-50% B over 5 min)

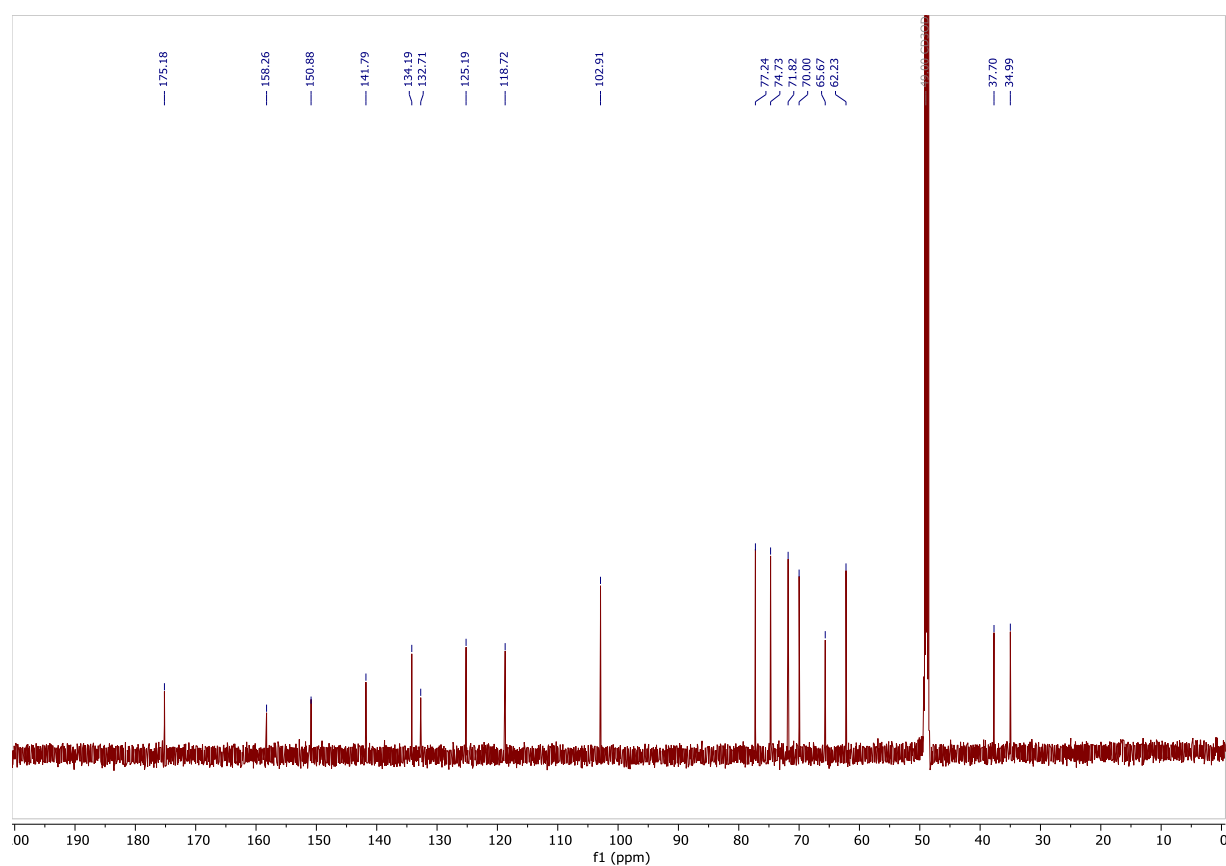
HPLC purification: [5mg/mL] Jupiter C4, gradient: 5-40% B over 20 min



Supplementary figure S15: HPLC trace of crude (A) and purified (B) **S8**

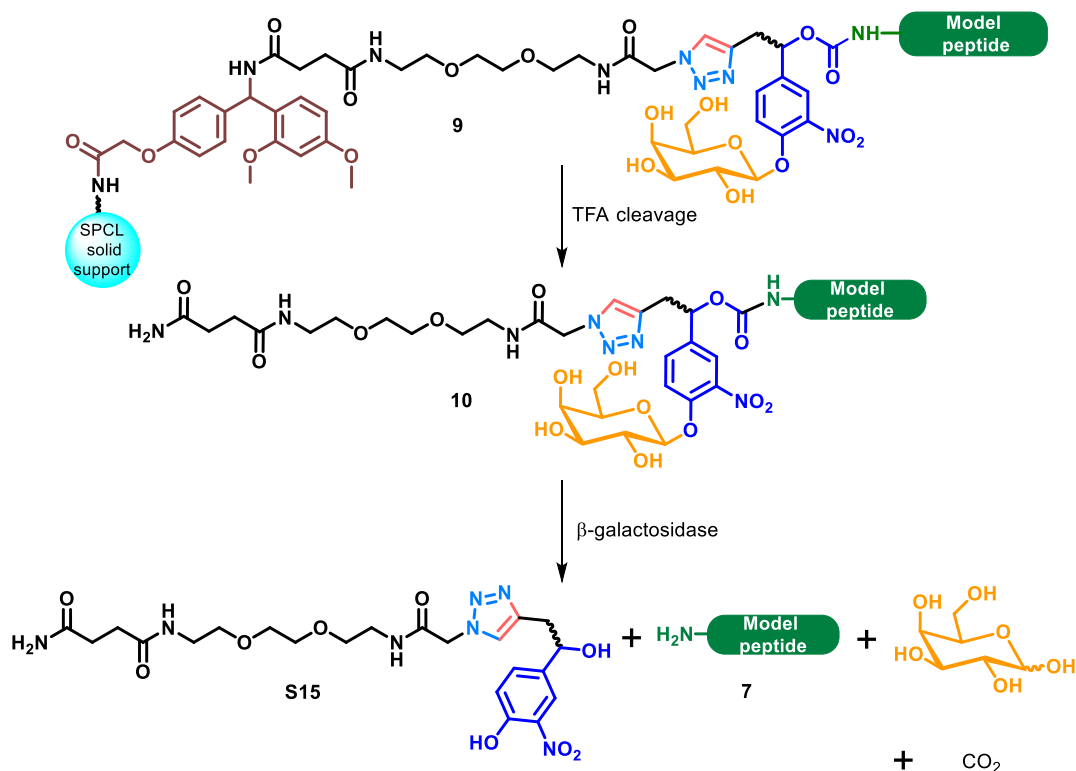


Supplementary figure S16: ¹H-NMR spectra of **S8** (600 MHz) in CD₃OD

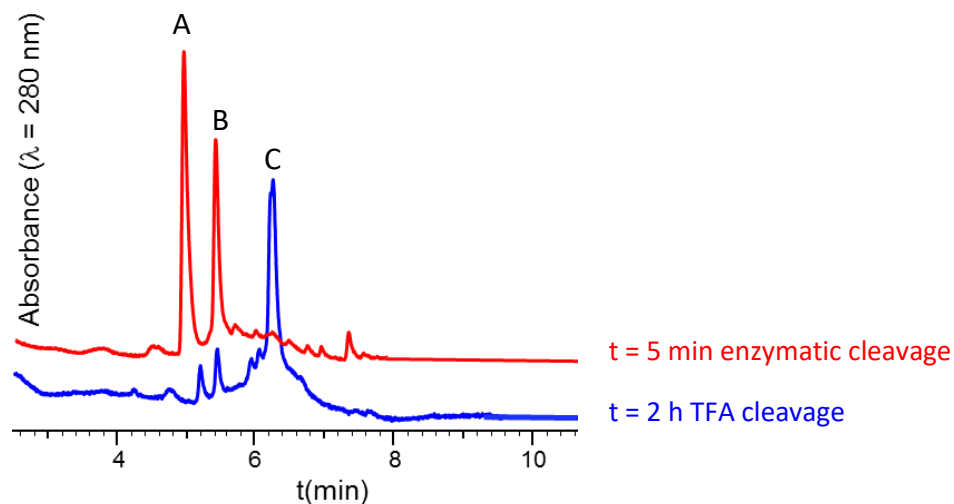


Supplementary figure S17: ¹³C-NMR spectra of **S8** (151 MHz) in CD₃OD

4-5- Dual linker strategy



A dual linker (Rink / β -galactosidase-labile) strategy was next explored. The peptide was cleaved from the resin following protocol PS2. Then enzymatic hydrolysis was carried out in solution, in 20 mM phosphate buffer pH 7.3 + 0.1 mM MgCl_2 , using 4 enzyme unit / 10 nmol substrate for 5 min.



Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (5.01)	537.2	537.1	S15
B (5.42)	3941.2	3940.7	7
C (6.22)	4667.0	4665.9	10

Supplementary figure S18: HPLC traces and MS analyses of TFA and β -galactosidase-mediated linker cleavages of **9** to give **7** (average masses, deconvoluted)

- **Compound 7**

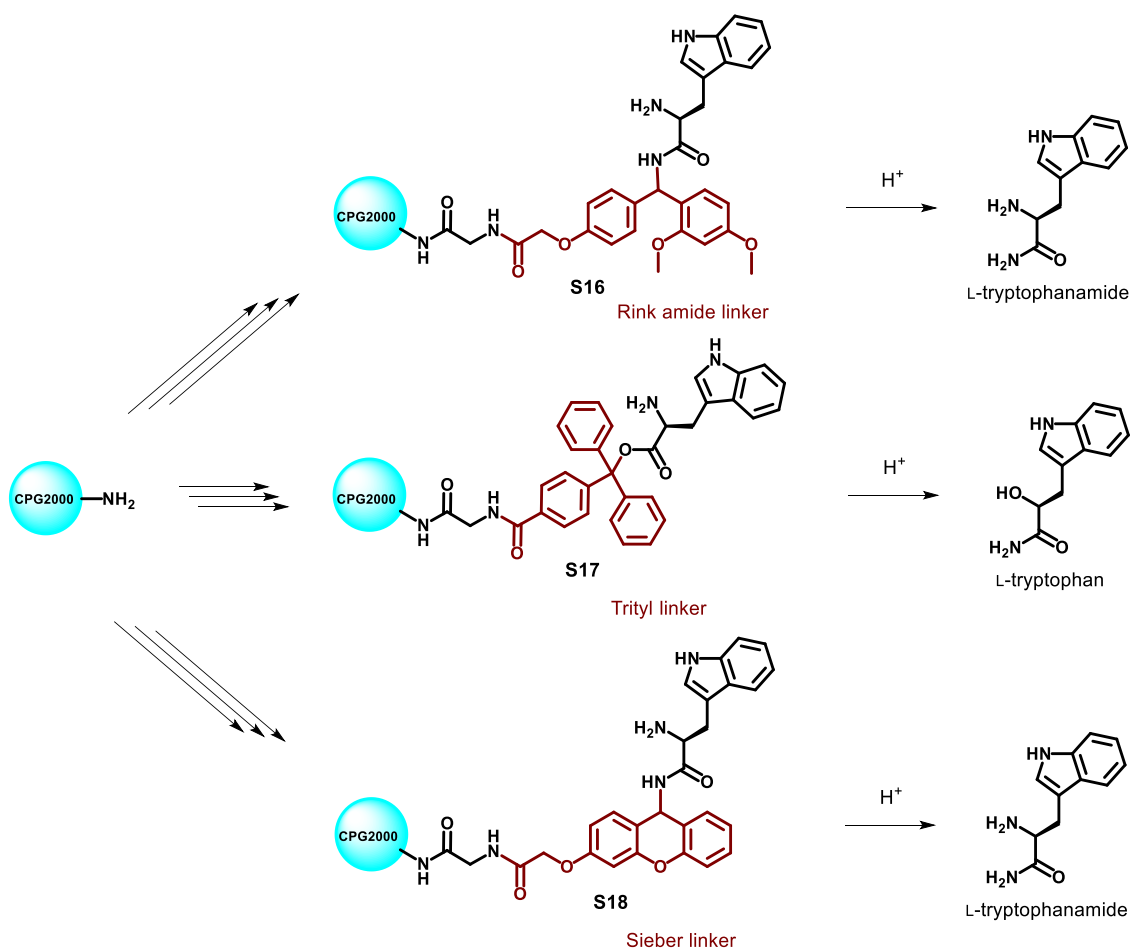
ESI-MS (m/z): [M] calcd. for $C_{171}H_{263}N_{53}O_{55}$: 3941.2, found: 3940.7 (average mass, deconvoluted)

HPLC analysis: t_R = 5.42 min (Aeris Widepore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C)

- **Exploration of different acid-labile linkers**

This dual linker strategy gave satisfactory results, the protocol is rapid (15 min for the enzymatic cleavage and 2 h for TFA cleavage and this strategy does not require large amounts of the enzyme. However, the Rink amide linker cleavage requires the use of harsh acidic conditions, which is not compatible with our aim of developing a mild cleavage strategy. Thus, two others linkers were considered to replace the Rink linker: trityl and Sieber linkers.

In order to study the stability of these linkers and compare it with the Rink amide linker under different conditions (TFA cleavage, ligations), they were grafted on CPG2000 and a tryptophan residue was coupled. Next, the release of Trp was monitored by HPLC and quantified by peak integration at λ = 280 nm (ϵ_{Trp} = 5500 L.mol⁻¹.cm⁻¹) [3]. NB: Fmoc-Trp-OH was used as building block, and not Fmoc-Trp(Boc)-OH classically used for the Fmoc-SPPS of Trp-containing peptides.



➤ **Introduction of the Rink linker and Trp coupling**

The same procedure for the introduction of $\text{N}_3\text{-CH}_2\text{-EBES-NH}_2$ on CPG2000 and CPG1000 resins (p S13) was used for the introduction of the Rink linker. Then Fmoc-Trp-OH (160 mg, 15 equiv.) in NMP (1.5 mL), PyAOP (261 mg, 20 equiv.) in NMP/DCM 65:35 (2 mL, final NMP/DCM ratio 8:2) and DIEA (131 μL , 30 equiv.) were added and the reactor was stirred for 6 h followed by extensive washing with NMP. The Fmoc group was finally deprotected by three successive treatments with 20% piperidine in NMP for 3 min affording **S16**.

➤ **Introduction of the trityl linker and Trp coupling**

25 μmol of aminopropyl CPG2000 resin was washed with NMP (5 mL, 1 min). Fmoc-Gly-OH (112 mg, 15 equiv.) in NMP (1.5 mL), PyAOP (261 mg, 20 equiv.) in NMP/DCM 65:35 (2 mL, final NMP/DCM ratio 8:2) and DIEA (131 μL , 30 equiv.) were added and the reactor was stirred for 6 h followed by extensive washing with NMP. Capping of potential unreacted amine groups was achieved by treatment with acetic anhydride as described in protocol PS1. The Fmoc group was deprotected by three successive treatments with 20% piperidine in NMP for 3 min. Then trityl linker (23 mg, 3 equiv.) in NMP (1.5 mL), PyAOP (79 mg, 6 equiv.) in NMP/DCM 65:35 (2 mL, final NMP/DCM ratio 9:1) and DIEA (26 μL , 6 equiv.) were added and the reactor was stirred for 12 h followed by extensive washing by NMP and capping with acetic anhydride (see protocol PS1). Then acetyl chloride (53 μL , 30 equiv.) in 1.5 mL of a 1:1 toluene/DCM mixture was added under an argon atmosphere and the reaction stirred for 12 h. Then Fmoc-Trp-OH (160 mg, 15 equiv.) in 3.5 mL DCM and DIEA (131 μL , 30 equiv.) were added and the mixture was stirred for 12 h. The Fmoc group was finally deprotected by three successive treatments with 20% piperidine in NMP for 3 min affording **S17**.

➤ **Introduction of the Sieber linker and Trp coupling**

25 μmol of aminopropyl CPG2000 resin was washed with NMP (5 mL, 1 min). Fmoc-Gly-OH (112 mg, 15 equiv.) in NMP (1.5 mL), PyAOP (261 mg, 20 equiv.) in NMP/DCM 65:35 (2 mL, final NMP/DCM ratio 8:2) and DIEA (131 μL , 30 equiv.) were added and the reactor was stirred for 6 h followed by extensive washing with NMP. Capping of potential unreacted amine groups was achieved by treatment with acetic anhydride as described in protocol PS1. The Fmoc group was deprotected by three successive treatments with 20% piperidine in NMP for 3 min. Then Fmoc-Sieber linker (37 mg, 3 equiv.) in NMP (1.5 mL), PyAOP (78 mg, 6 equiv.) in NMP/DCM 65:35 (2 mL, final NMP/DCM ratio 9:1) and DIEA (26 μL , 6 equiv.) were added and the reactor was stirred for 12 h followed by extensive washing with NMP and capping with acetic anhydride (see protocol PS1). Then Fmoc-Trp-OH (160 mg, 15 equiv.) in NMP (1.5 mL), PyAOP (261 mg, 20 equiv.) in NMP/DCM 65:35 (2 mL, final NMP/DCM ratio 8:2) and DIEA (131 μL , 30 equiv.) were added and the reactor was stirred for 6 h followed by extensive washing with NMP. The Fmoc group was finally deprotected by three successive treatments with 20% piperidine in NMP for 3 min affording **S18**.

➤ **Linkers stability under NCL conditions**

An NCL buffer was prepared according to protocol PS6. This solution was then added to 1 μmol of supports **S16**, **S17** and **S18** for 12 h. Then reactors were respectively washed thoroughly with NCL solution, water/MeCN 8:2 and water and then cleaved with TFA and analysed by HPLC to quantify the released tryptophan (for trityl linker) and tryptophanamide (for Rink and Sieber linkers). Linker cleavage was determined by comparing the released amount to the initial quantity of solid support.

Table S1: Linkers stability under NCL conditions

	Rink linker	Trityl linker	Sieber linker
% linker cleavage	0	100	0

➤ **Linkers cleavage from solid support under different acidic conditions**

Cleavage of Rink and Sieber linkers under various acidic conditions was quantified using a similar procedure as above. 1 μ mol of supports **S16** and **S18** were treated with different acidic conditions, followed by washing of the solid support with 1:1 water/MeCN, concentration under reduced pressure and HPLC analysis to quantify the released tryptophanamide. Linker cleavage was determined by comparing the released amount to the initial quantity of solid support.

❖ Sieber linker

Table S2: Sieber linker cleavage under various acidic conditions

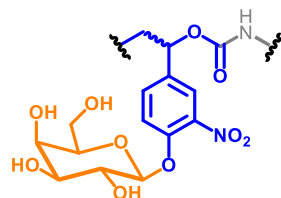
Conditions	0.5 h	1.5 h	3 h
10% TFA, water	31%	46%	100%
0.1% TFA, 1:1 water/MeCN	- ^a	- ^a	< 1%
5% TFA, water	2.7%	5.6%	12.3%
5% TFA, HFIP	12.6%	33%	49%
5% TFA, 4.5:5 HFIP/water	1.3%	4.2%	7.4%

^a: not determined

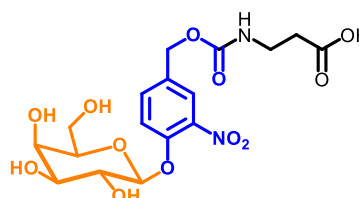
❖ Rink amide linker

As expected, this linker was stable toward all the cleavage conditions tested with Sieber linker (table S2).

5- Determination of the molar extinction coefficient of the 1st generation linker



First generation linker



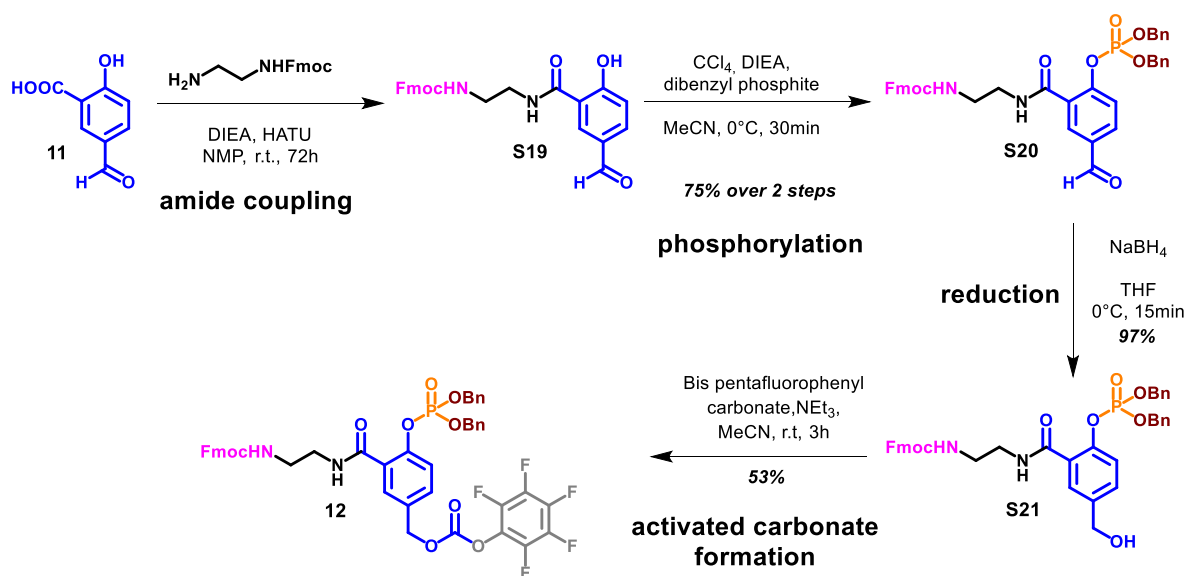
S8

In order to facilitate product quantification, 280 nm molar extinction coefficients of the first generation linker was determined. **S8** was used for this purpose since it does not contain the 4-nitrophenol leaving group which also absorbs at 280 nm.

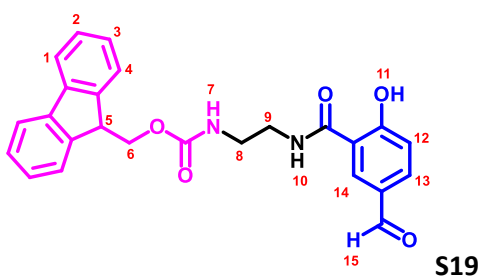
$\epsilon^{280} = 6938 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$ in 8:2:0.01 H₂O/MeCN/TFA.

6- Second generation enzyme-cleavable linker

6-1- Linker precursor synthesis



- Synthesis of compound S20**



To an ice-cold solution of mono-Fmoc ethylene diamine (5 g, 15.7 mmol) in NMP (79 mL, 0.2 M amine) was added 5-formylsalicylic acid (4.45 g, 31.4 mmol, 2 equiv.), HATU (7.75 g, 18.8 mmol, 1.2 equiv.) and DIEA (5.45 mL, 31.4 mmol, 2 equiv.). The resulting solution was stirred for 20 h at 25 °C. Then a second batch of HATU was added (3.25 g, 7.85 mmol, 0.5 equiv.). After 4 h the mixture was diluted with EtOAc (50 mL) and washed successively with saturated NaHCO_3 (x 3), 1 M HCl (x 3) and saturated NaCl aqueous solutions, dried over MgSO_4 , and concentrated under reduced pressure to give 6.05 g of a yellowish solid. The crude compound **S19** was directly used in the next step without further purification.

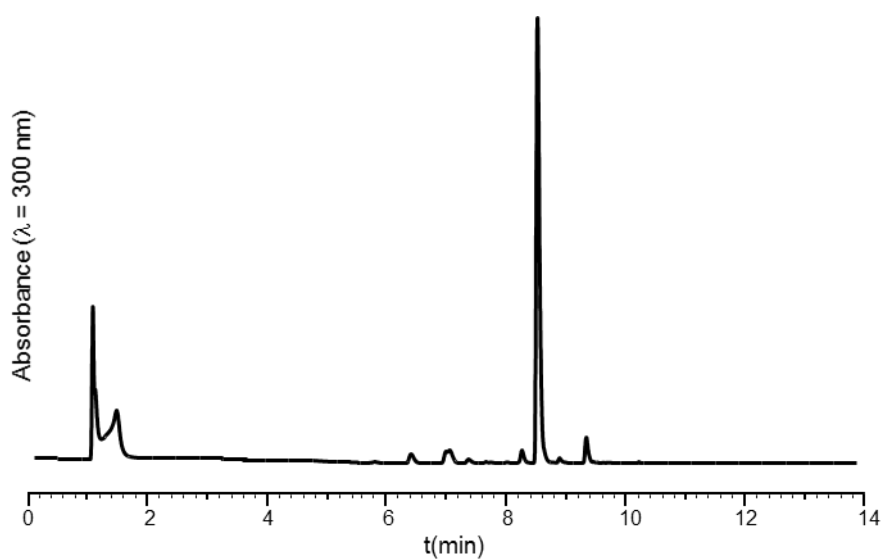
¹H NMR (600 MHz, DMSO-*d*₆): δ 13.48 (s, 1H, H₁₁), 9.83 (s, 1H, H₁₅), 9.13 (t, *J* = 5.7 Hz, 1H, H₇), 8.46 (d, *J* = 2.0 Hz, 1H, H₁₄), 7.94 (dd, *J* = 8.5, 2.0 Hz, 1H, H₁₃), 7.88 (d, *J* = 7.6 Hz, 2H, H₁), 7.67 (d, *J* = 7.5 Hz, 2H, H₄), 7.46 (t, *J* = 5.8 Hz, 1H, H₁₀), 7.40 (dd, *J* = 7.5, 7.5 Hz, 2H, H₄), 7.30 (dd, *J* = 7.5, 7.5 Hz, 2H, H₃), 7.08 (d, *J* = 8.5 Hz, 1H, H₁₂), 4.31 (d, *J* = 6.9 Hz, 2H, H₆), 4.21 (t, *J* = 6.9 Hz, 1H, H₅), 3.41 – 3.38 (m, 2H, H₈), 3.20 – 3.21 (m, 2H, H₉).

¹³C NMR (151 MHz, DMSO-*d*₆): δ 190.7, 167.9, 165.4, 160.7, 146.1, 138.7, 134.3, 131.3, 128.9, 128.4, 127.7, 127.4, 127.0, 118.5, 116.0, 42.5, 39.2*, 39.1*, 38.2.

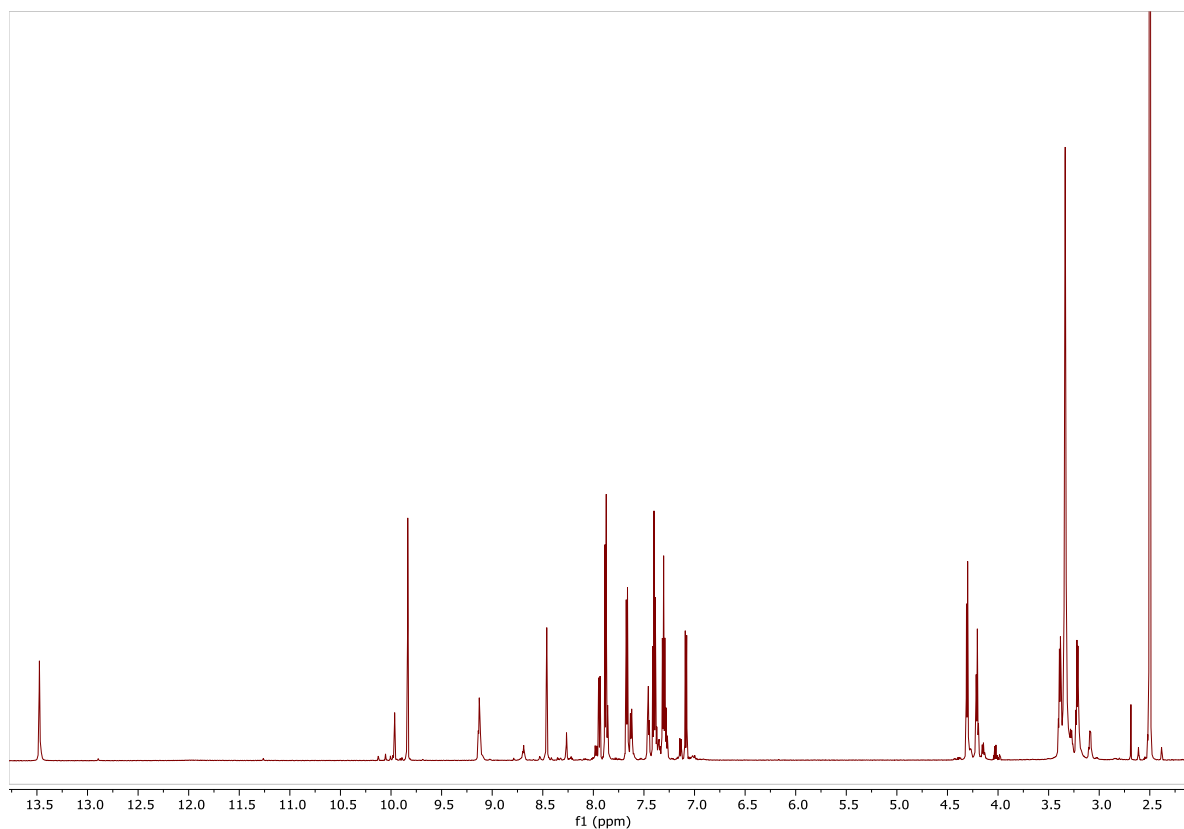
*These two peaks are hidden below the (CD₃)₂SO peak, identified by 2D ¹³C-¹H HSQC analysis.

ESI-MS (*m/z*): [M] calcd. for C₂₅H₂₂N₂O₅: 430.2 found: 430.0

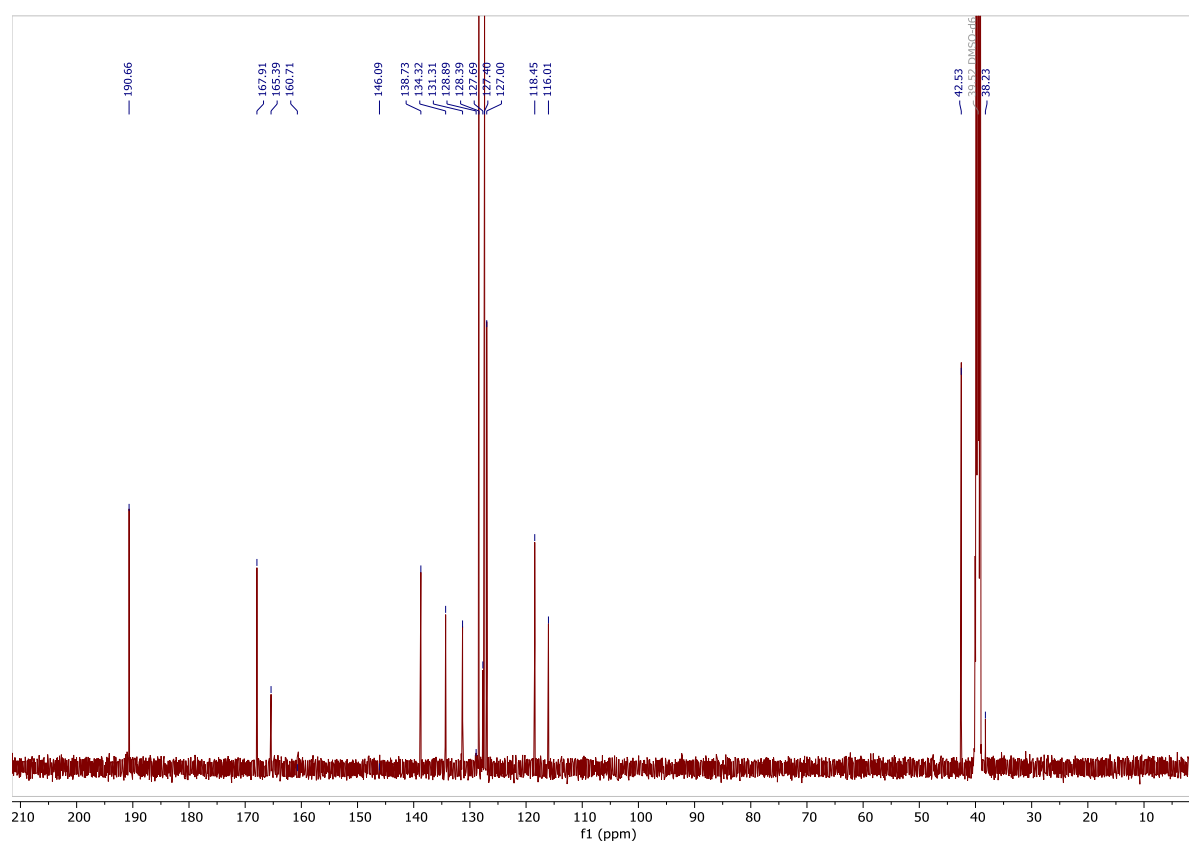
HPLC analysis: *t*_R = 8.74 min (Aeris Widespore XB-C18 2, gradient: 3-90% B' over 15 min, 60 °C)



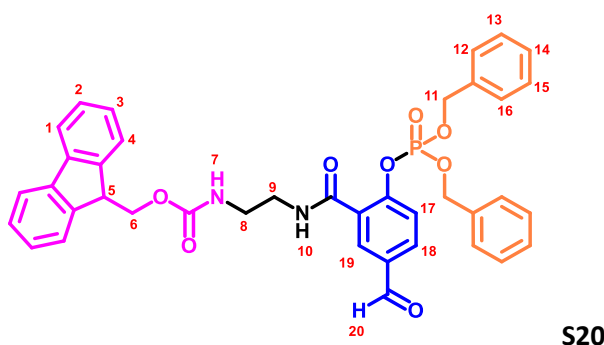
Supplementary figure S19: HPLC trace of crude **S19**



Supplementary figure S20: ^1H -NMR spectra of crude **S19** (600 MHz) in $\text{DMSO}-d_6$



Supplementary figure S21: ^{13}C -NMR spectrum of crude **S19** (151 MHz) in $\text{DMSO}-d_6$



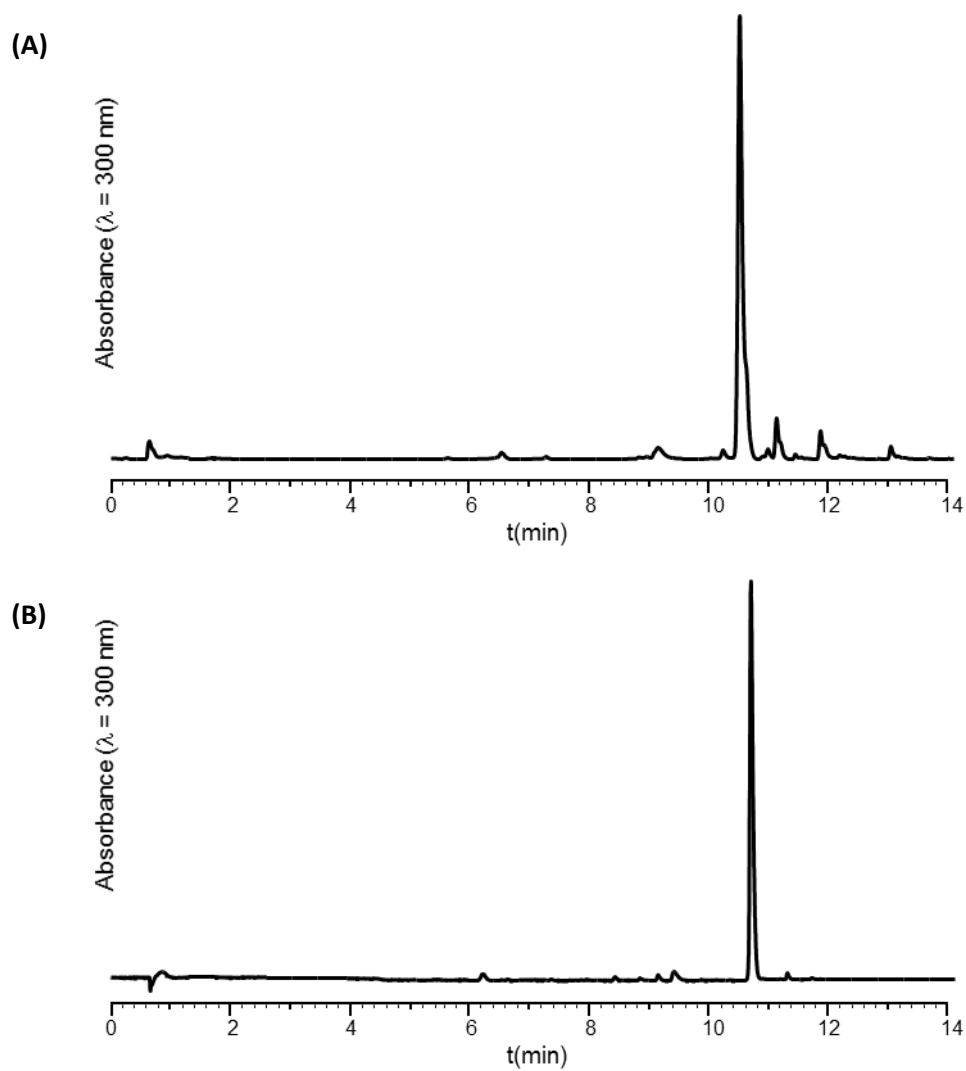
Crude **S19** (6.05 g, 14.06 mmol if it was pure) was dissolved in anhydrous MeCN (281 mL, 0.05M **S19**) by refluxing the mixture for a few minutes. The resulting solution is then cooled in an ice bath and CCl_4 (9.48 mL, 98.4 mmol, 7 equiv.), DIEA (7.1 mL, 40.3 mmol, 2.9 equiv.) and DMAP (178 mg, 1.42 mmol, 0.1 equiv.) were successively added. One minute later, dibenzylphosphite (4.98 mL, 22.5 mmol, 1.6 equiv.) was added dropwise and the reaction was stirred for 30 min at 0 °C. Then, the mixture was diluted with EtOAc and washed successively with saturated Na_2CO_3 (x 3), 1 M HCL (x 3) and saturated NaCl aqueous solutions, dried over MgSO_4 , and concentrated under reduced pressure. The product was purified by flash column chromatography (eluent: 70:30 DCM/AcOEt) affording **S20** a white amorphous solid (8.13 g, 75% yield over two steps).

^1H NMR (600 MHz, $\text{DMSO}-d_6$): δ 9.98 (s, 1H, H_{20}), 8.58 (t, $J = 5.7$ Hz, 1H, H_7), 8.01 (bs, 2H, H_{19}), 7.88 (d, $J = 7.6$ Hz, 2H, H_1), 7.67 (d, $J = 7.5$ Hz, 2H, H_4), 7.53 (bd, $J = 8.5$ Hz, 1H, H_{18}), 7.40 – 7.37 (m, 2H, H_2), 7.41–7.30 (m, 11H, $\text{H}_{10+12+13+14+15+16}$), 7.31 – 7.28 (m, 3H, H_{17+3}), 5.16 – 5.20 (m, 4H, H_{11}), 4.30 (d, $J = 7.0$ Hz, 2H, H_6), 4.20 (t, $J = 7.0$ Hz, 1H, H_5), 3.31 – 3.28 (m, 2H, H_8), 3.18–3.16 (m, 2H, H_9).

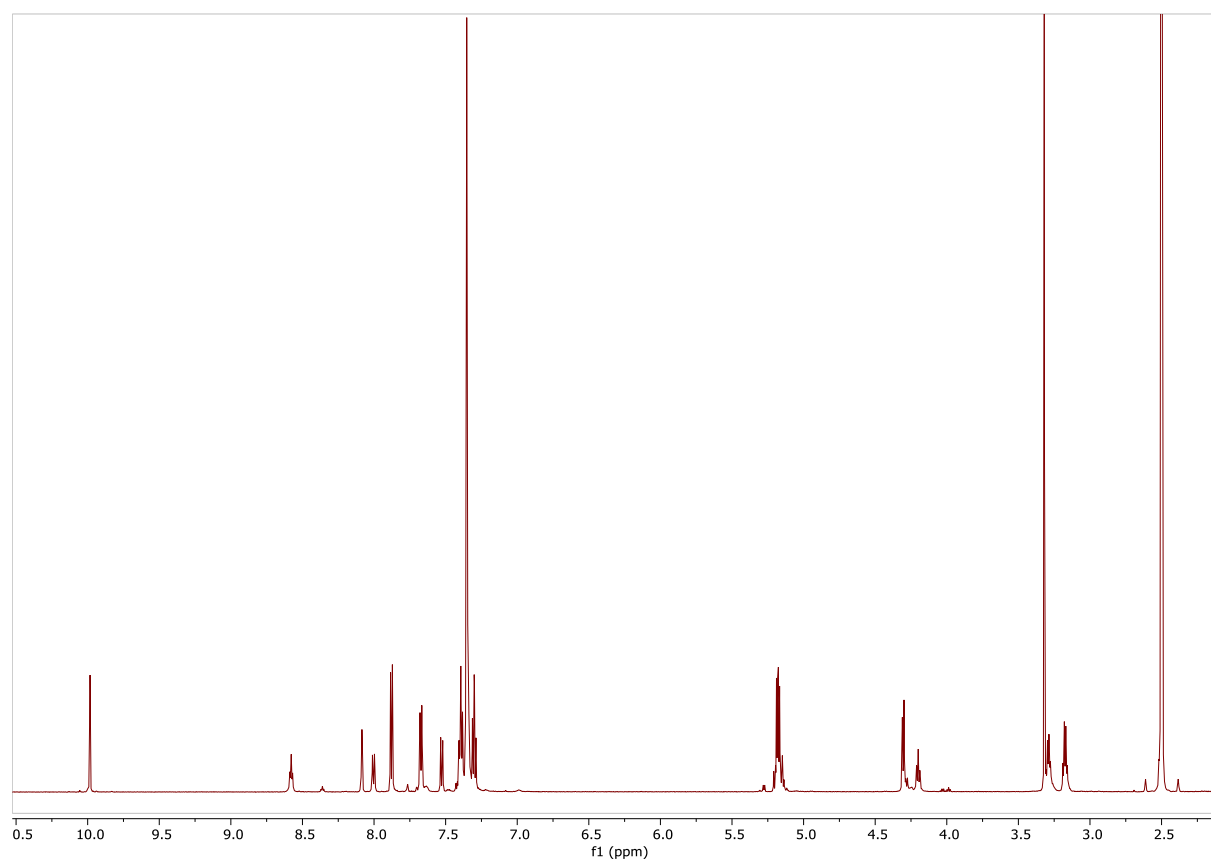
^{13}C NMR (151 MHz, $\text{DMSO}-d_6$): δ 191.5, 164.6, 156.3, 143.9, 140.7, 135.41, 135.37, 132.7, 132.6, 132.5, 130.7, 128.6, 128.5, 128.09, 128.06, 127.6, 127.0, 125.1, 120.1, 69.79, 69.75, 65.4, 46.7, 40.1.

ESI-MS (m/z): [M] calcd. for $\text{C}_{39}\text{H}_{35}\text{N}_2\text{O}_8\text{P}$: 690.2 found: 690.2

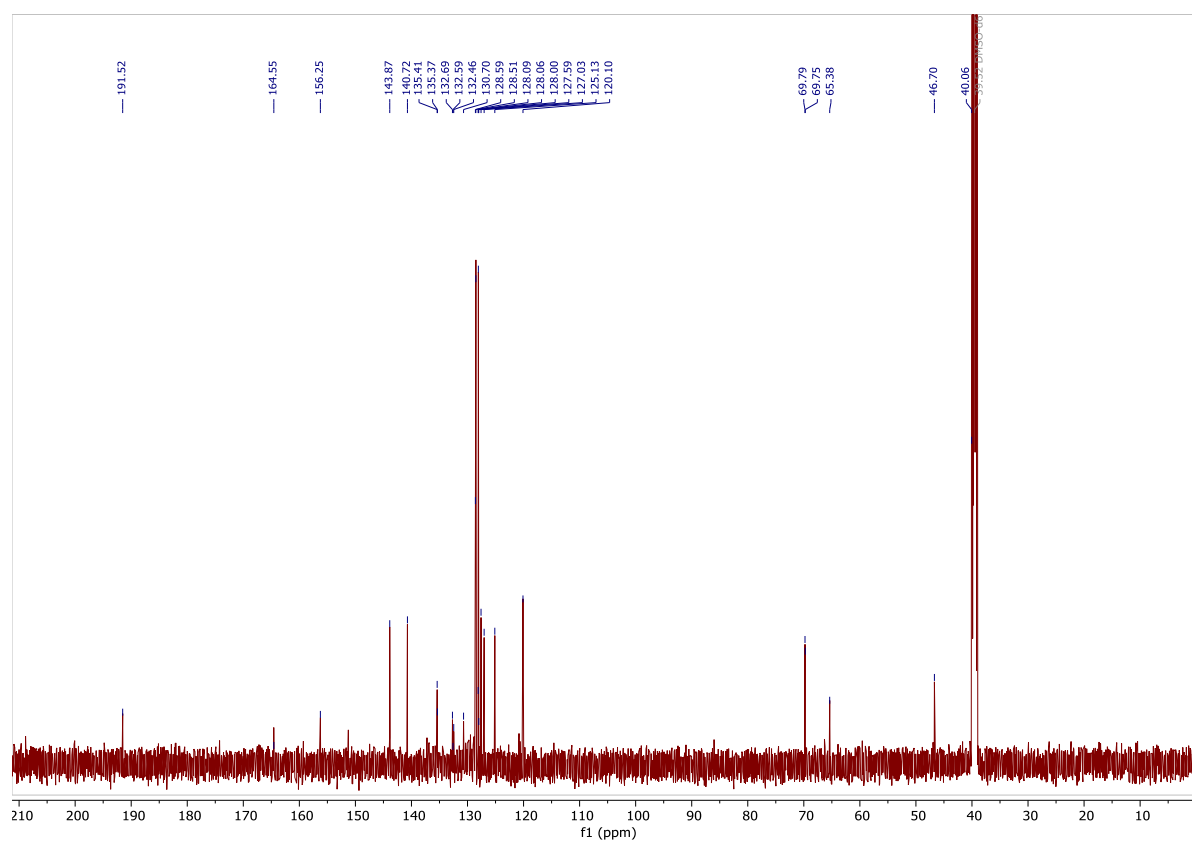
HPLC analysis: $t_R = 10.79$ min (Aeris Widedpore XB-C18 2, gradient: 3-90% B' over 15 min, 60 °C)



Supplementary figure S22: HPLC traces of crude (A) and purified (B) **S20**

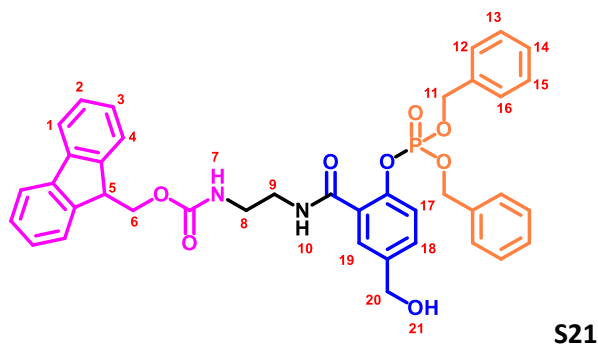


Supplementary figure S23: ^1H -NMR spectra of **S20** (600 MHz) in $\text{DMSO-}d_6$



Supplementary figure S24: ^{13}C -NMR spectrum of **S20** (151 MHz) in $\text{DMSO-}d_6$

- **Synthesis of compound 21**



A stirred solution of **S20** (5.78 g, 8.37 mmol) in anhydrous THF (42 mL, 0.2 M aldehyde) was cooled to 0° C and NaBH₄ (330 mg, 8.37 mmol, 1 equiv.) was added. The resulting reaction mixture was stirred at 0° C for 15 min. Then the mixture was diluted with EtOAc (266 ml) and washed successively with 1 M HCl (x 3), saturated NaHCO₃ (x 3), and saturated NaCl aqueous solutions, dried over MgSO₄, and concentrated under reduced pressure. The resulting crude white solid (5.62 g) was directly engaged in the next step without further purification.

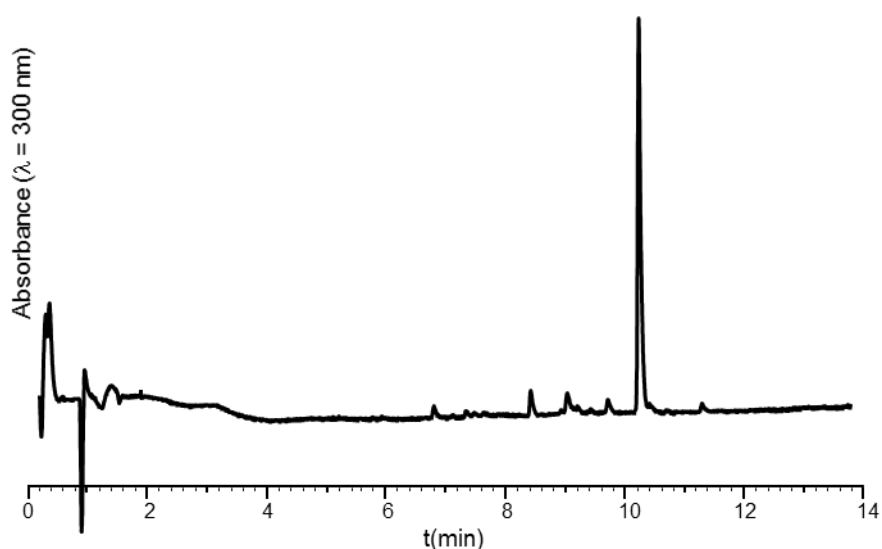
¹H NMR (600 MHz, DMSO-*d*₆): δ 8.34 (t, *J* = 5.7 Hz, 1H, H₇), 7.88 (d, *J* = 7.5 Hz, 2H, H₁), 7.67 (d, *J* = 7.5 Hz, 2H, H₄), 7.48 (bs, 1H, H₁₉), 7.41 – 7.36 (m, 3H, H₂₊₁₈), 7.37 – 7.31 (m, 11H, H₁₀₊₁₂₊₁₃₊₁₄₊₁₅₊₁₆), 7.28 – 7.31 (m, 2H, H₃), 7.25 (d, *J* = 8.4 Hz, 1H, H₁₇), 5.09 – 5.18 (m, 4H, H₁₁), 4.49 (d, *J* = 5.7 Hz, 2H, H₂₀), 4.28 (d, *J* = 7.0 Hz, 2H, H₆), 4.19 (t, *J* = 7.0 Hz, 1H, H₅), 3.25–3.28 (m, 2H, H₈), 3.14 – 3.16 (m, 2H, H₉).

¹³C NMR (151 MHz, DMSO-*d*₆): δ 165.4, 156.1, 143.7, 140.6, 139.2, 135.5, 135.46, 130.1, 128.7, 128.3, 127.9, 127.8, 127.5, 127.3, 126.9, 125.0, 120.0, 119.9, 69.3, 69.2, 65.2, 61.8, 46.6, 39.9*.

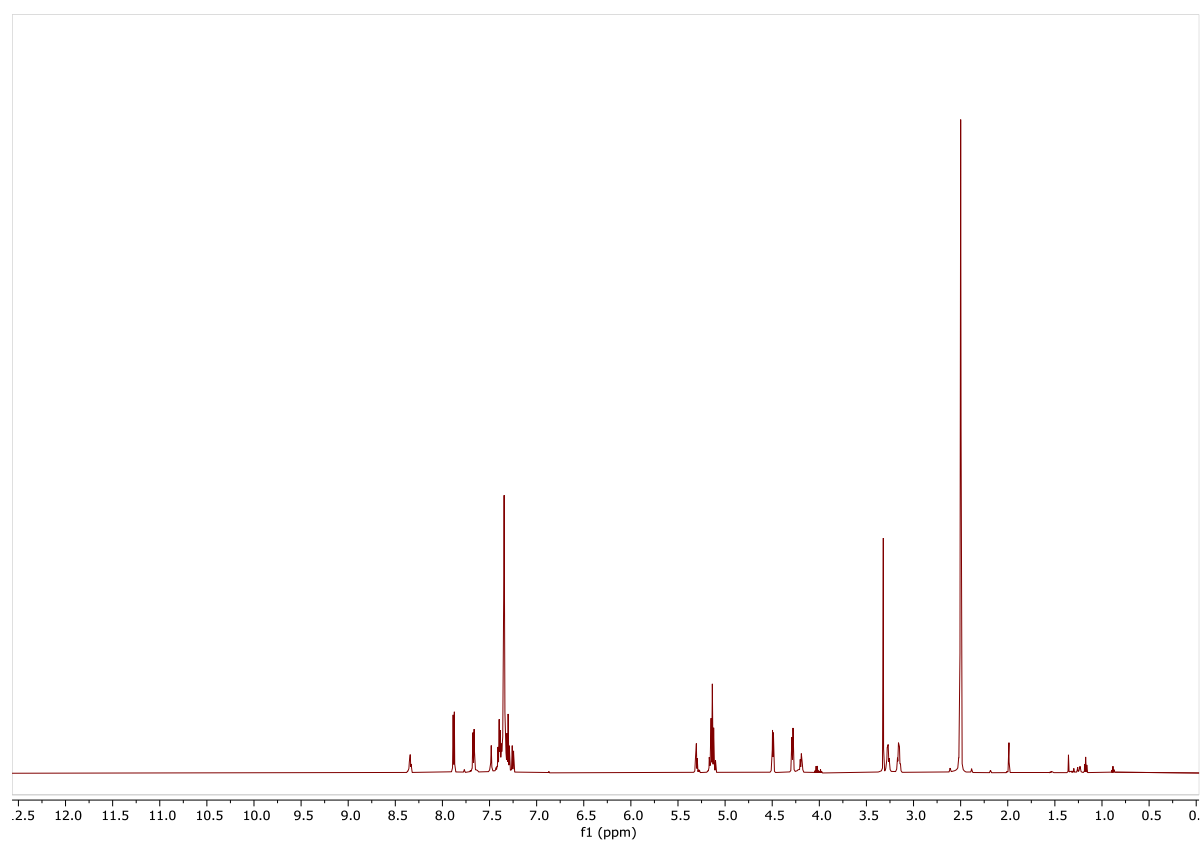
*This peak is hidden below the (CD₃)₂SO peak, identified by 2D ¹³C-¹H HSQC analysis.

ESI-MS (m/z): [M] calcd. for $C_{39}H_{37}N_2O_8P$: 692.2 found: 692.2

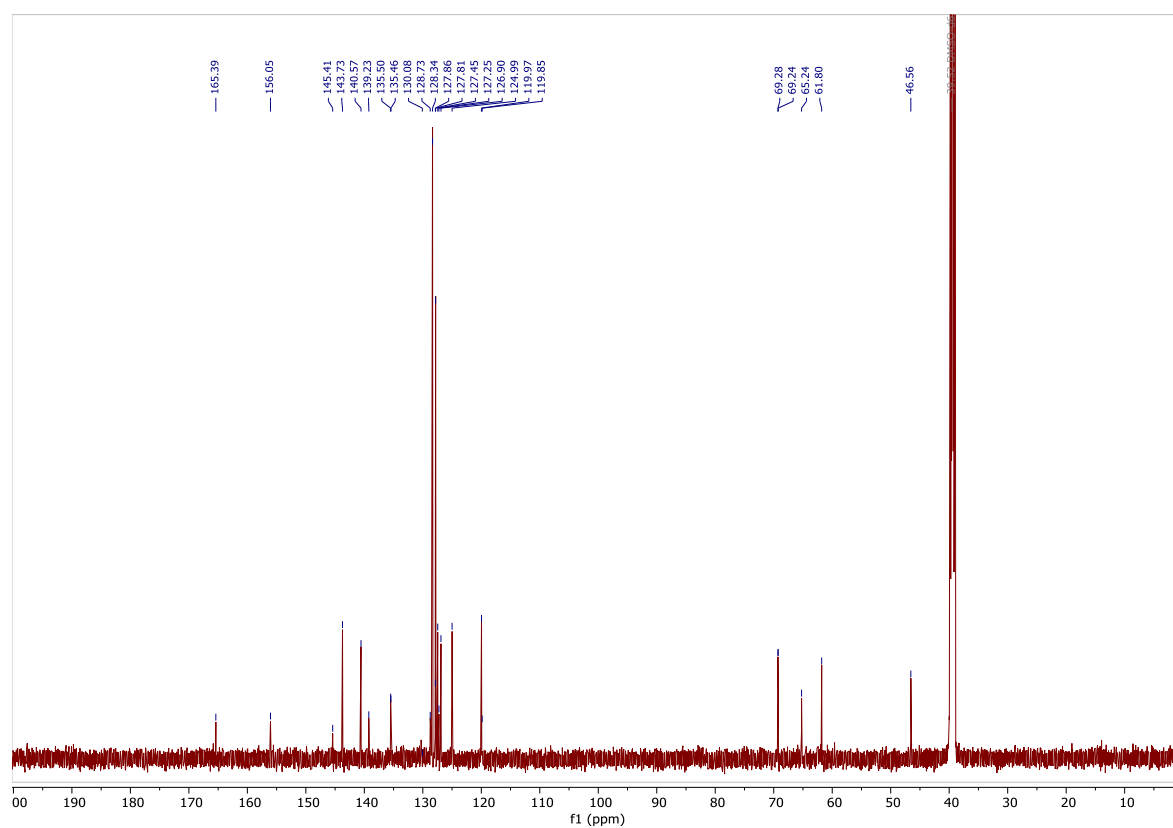
HPLC analysis: t_R = 10.11 min (Aeris Widepore XB-C18 2, gradient: 3-90% B' over 15 min, 60 °C)



Supplementary figure S25: HPLC trace of crude **S21**



Supplementary figure S26: ^1H -NMR spectra of crude **S21** (600 MHz) in $\text{DMSO-}d_6$



Supplementary figure S27: ^{13}C -NMR spectrum of crude **S21** (151 MHz) in $\text{DMSO-}d_6$

- Synthesis of **12**

Optimization of the introduction of the leaving group

Four different reagents were tested for the transformation of benzylic alcohol **S21** into an activated carbonate or carbamate. Optimization results are listed in table S3.

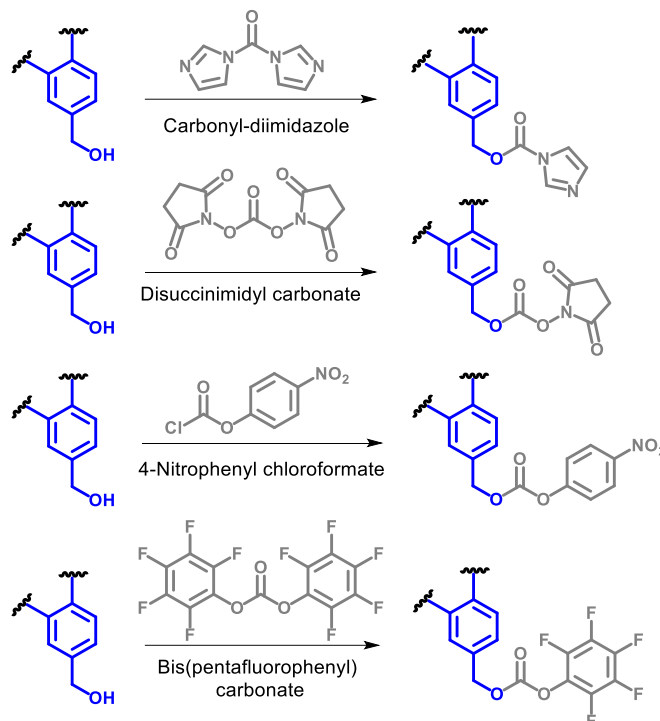
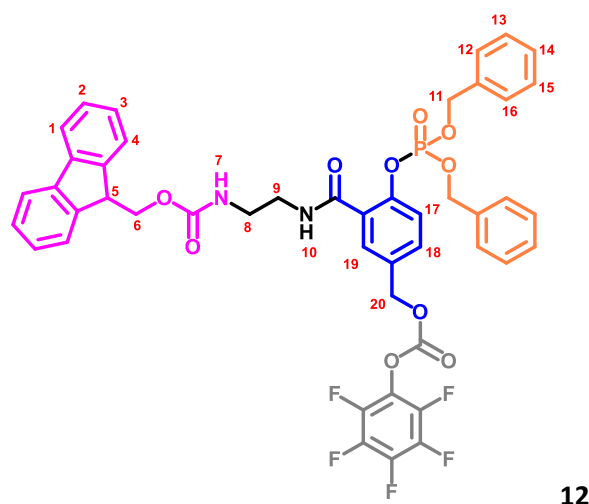


Table S3: Different reagents and conditions used for the introduction of the leaving group on the second-generation linker

Entry	Reagent	Conditions	Results
1	Carbonyl-diimidazole	1.5 equiv. in toluene [0.3 M]	Very low conversion rates
2		3 equiv. in THF [0.3 M]	
3	Disuccinimide carbonate	1.5 equiv. + 0.3 equiv. NEt ₃ in dry MeCN [0.2 M]	Very low conversion rates
4		3 equiv. + 4.5 equiv. DIEA in DMF [0.2 M]	
5	4-Nitrophenyl chloroformate	6 equiv. + 6.5 equiv. DIEA in DCM [2.5 M]	Best results were obtained with entry 10 conditions; however, conversion was low. Column purification was challenging due to product degradation on silica, and purified compound decomposed upon storage at -20 °C
6		6 equiv. + 6.5 equiv. pyridine in DCM [2.5 M]	
7		6 equiv. + 6.5 equiv. lutidine in DCM [2.5 M]	
8		6 equiv. + 6.5 equiv. NaHCO ₃ in DCM [2.5 M]	
9		6 equiv. without base in DCM [2.5 M]	
10		20 equiv. + 6.5 equiv. NaHCO ₃ in DCM [2.5 M]	
11	Bis-pentafluorophenyl carbonate	20 equiv., 21 equiv. Na ₂ CO ₃ in DCM [2.5 M]	Conditions 14 were selected for the synthesis of the activated carbonate derivative of the linker precursor. More than 97% conversion yield was obtained and the resulting product was stable during column purification and upon storage at -20 °C.
12		1.5 equiv., 0.3 equiv. NEt ₃ in dry MeCN [0.2 M] then addition of 10 equiv. of NEt ₃ after 24 h	
13		0.5 equiv. NEt ₃ in dry MeCN [0.4 M] was added. After 30 min, dropwise addition of 1.5 equiv. of reagent in dry MeCN [0.4 M]	
14		0.5 equiv. NEt ₃ was added in dry MeCN [0.3 M]. After 30 min, dropwise addition of 1.5 equiv. of reagent in dry MeCN [0.4 M]. After 1 h addition of additional 0.75 equiv. of reagent and stirring for additional 1.5 h.	

**12**

Benzylic alcohol **S21** (200 mg, 0.29 mmol if it was pure) was dissolved in MeCN (1 mL, 0.3 M alcohol) and then NEt₃ (6.8 μ L, 0.29 mmol, 1 equiv.) was added and the mixture was stirred for 30 min. Next bis-pentafluorophenyl carbonate (226 mg, 0.435 mmol, 1.5 equiv.) in MeCN (1 mL) was added dropwise and the mixture stirred for 1 h. Then, additional bis-pentafluorophenyl carbonate (113 mg, 0.217 mmol, 0.75 equiv.) was added and the mixture stirred for 1.5 h. Then, EtOAc (50 mL) was added and the mixture washed successively with saturated NaHCO₃ (x 3), 1 M HCl (x 3) and saturated NaCl aqueous solutions, dried over MgSO₄, and concentrated under reduced pressure. The crude product was dissolved in DCM and purified by flash chromatography (eluent: 50/50 petroleum ether/AcOEt) affording a white amorphous solid (102 mg, 38% over two steps).

¹H NMR (600 MHz, DMSO-*d*₆): δ 8.47 (t, *J* = 5.7 Hz, 1H, H₇), 7.87 (d, *J* = 7.6 Hz, 2H, H₁), 7.70 – 7.65 (m, 2H, H₄), 7.58 (d, *J* = 8.5, 1H, H₁₉), 7.41 – 7.33 (m, 14H, H₂₊₁₀₊₁₂₊₁₃₊₁₄₊₁₅₊₁₆₊₁₈), 7.32 – 7.27 (m, 3H, H₃₊₁₇), 5.41 (bs, 2H, H₂₀), 5.21 – 5.13 (m, 4H, H₁₁), 4.30 (d, *J* = 7.0 Hz, 2H, H₆), 4.20 (t, *J* = 7.0 Hz, 1H, H₅), 3.32 – 3.28 (m, 2H, H₈), 3.19 – 3.17 (m, 2H, H₉).

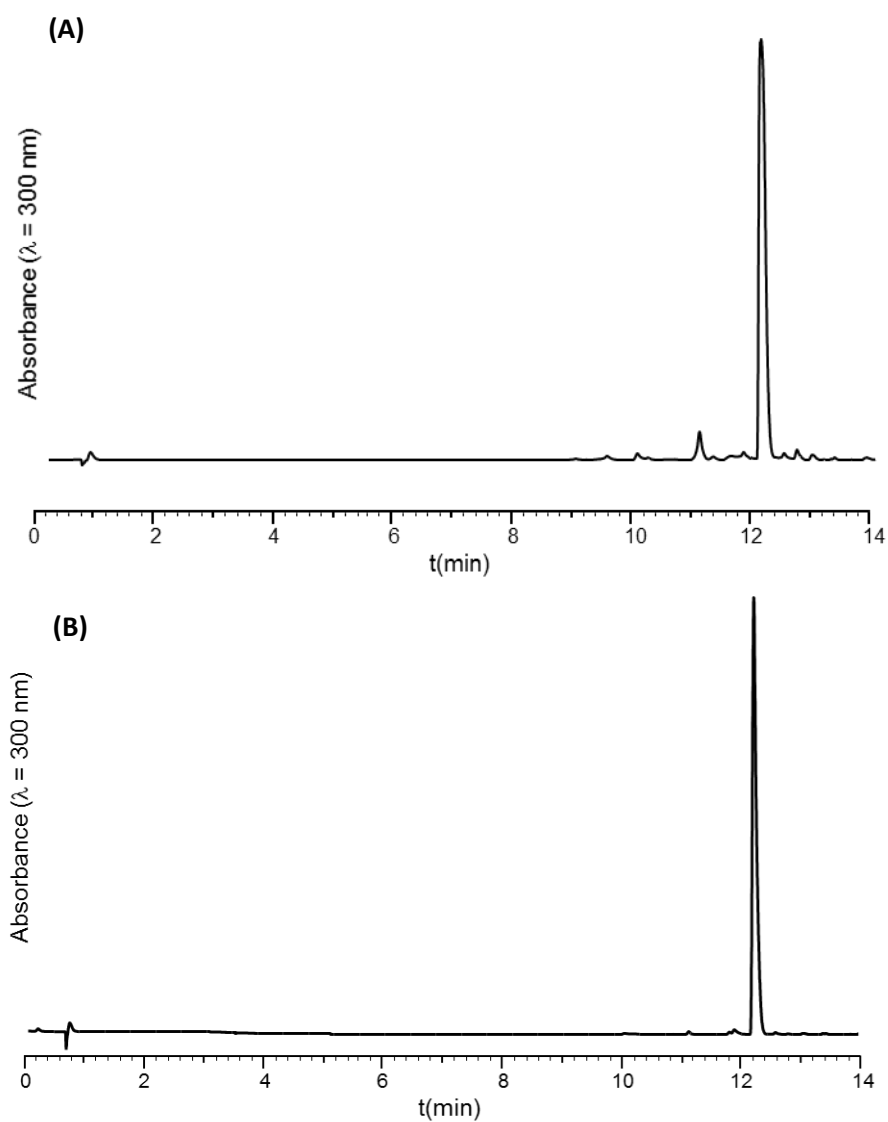
¹³C NMR (151 MHz, DMSO-*d*₆): δ 165.3, 156.5, 150.7, 147.6, 144.1, 141.8, 141.0, 138.6, 136.9, 135.8, 131.9, 131.1, 130.4, 129.7, 128.8, 128.76, 128.3, 127.9, 127.3, 125.4, 120.9, 120.4, 71.2, 69.9, 69.8, 65.7, 47.0, 39.9*.

*This peak is hidden below the (CD₃)₂SO peak, identified by 2D ¹³C-¹H HSQC analysis.

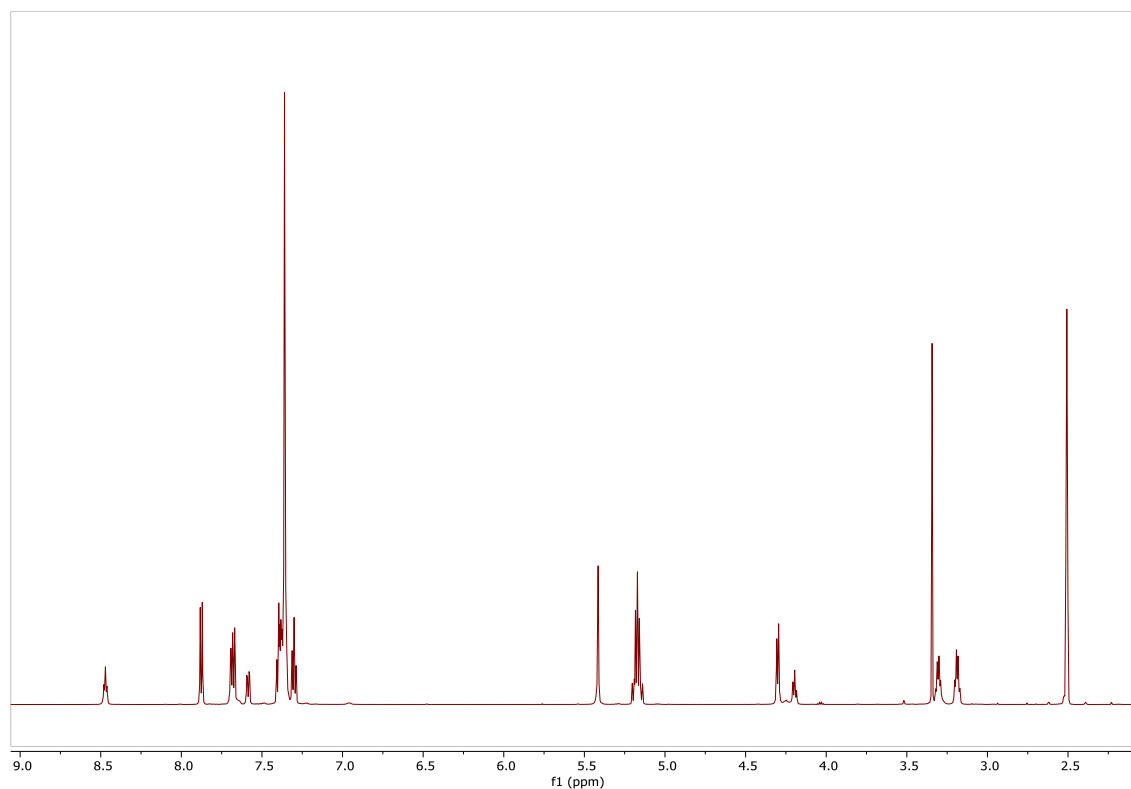
¹⁹F NMR (564 MHz, DMSO-*d*₆): δ -152.9 (d, *J* = 17.4 Hz, 2F), -157.1 (t, *J* = 21.8 Hz, 1F), -161.7 (dd, *J* = 21.8, 17.4 Hz, 2F).

ESI-MS (*m/z*): [M] calcd. for C₄₆H₃₆ F₅N₂O₁₀P: 902.2, found: 902.1

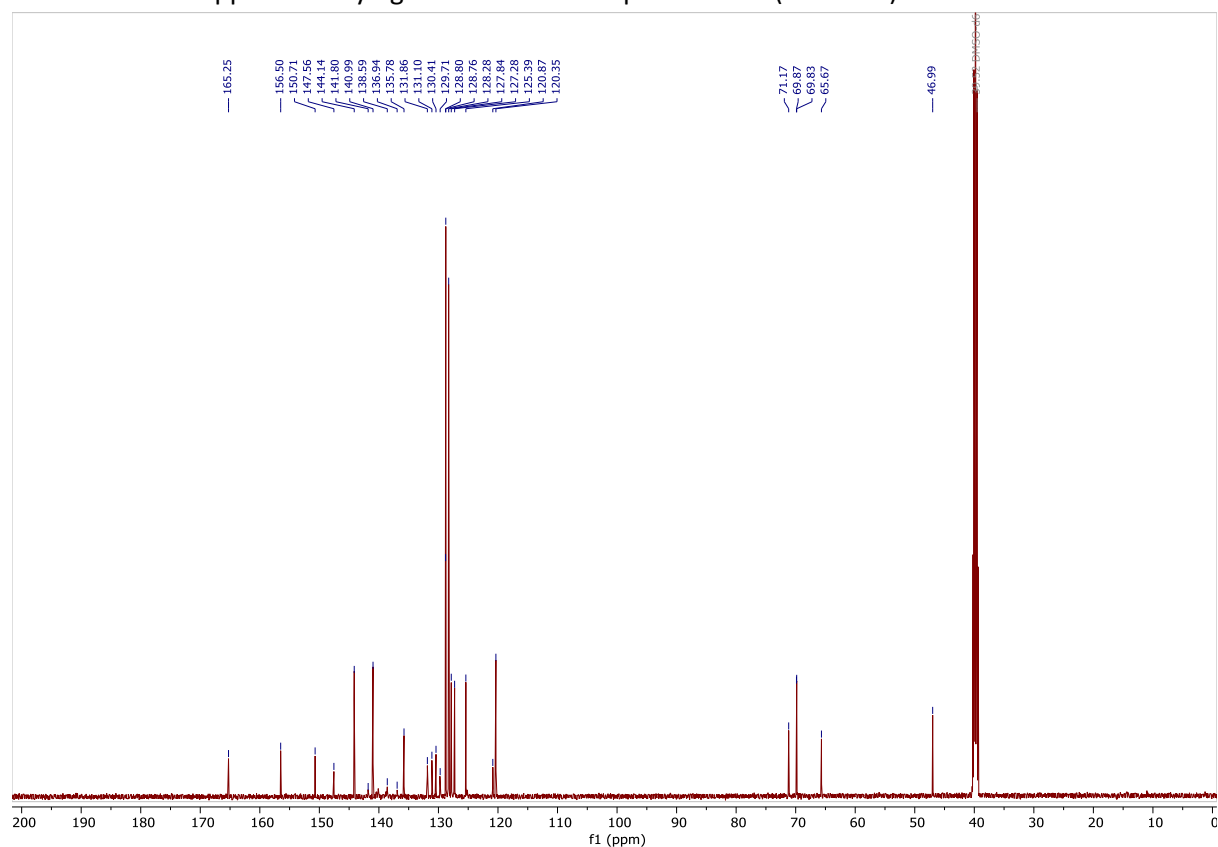
HPLC analysis: *t*_R = 12.11 min (Aeris Widespore XB-C18 2, gradient: 3-90% B' over 15 min, 60 °C)



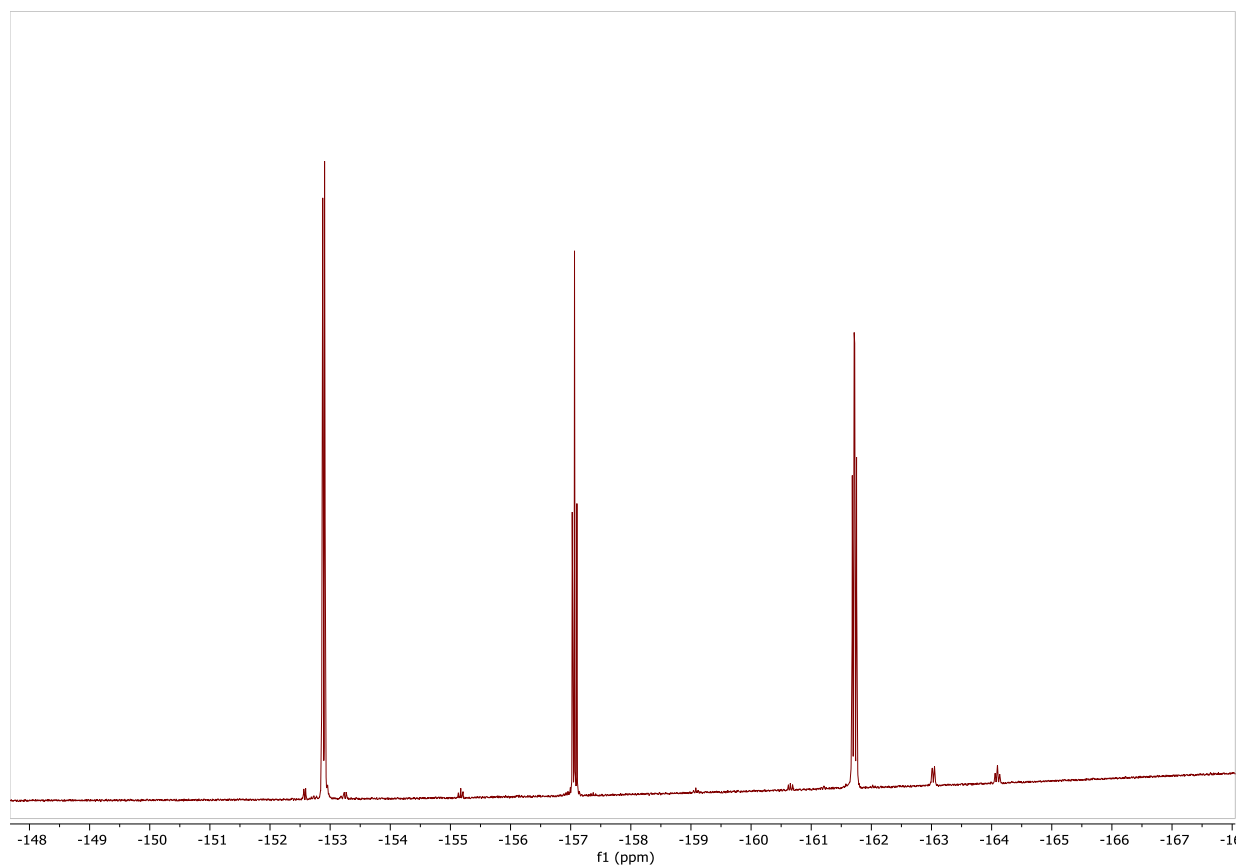
Supplementary figure S28: HPLC traces of crude (A) and purified (B) **11**



Supplementary figure S29: ^1H -NMR spectra of **11** (600 MHz) in $\text{DMSO-}d_6$



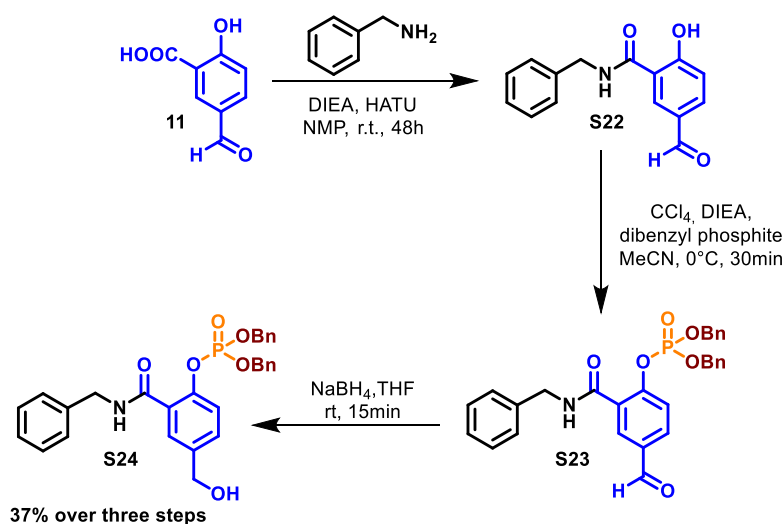
Supplementary figure S30: ^{13}C -NMR spectrum of **11** (151 MHz) in $\text{DMSO-}d_6$



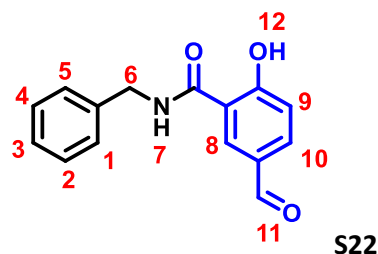
Supplementary figure S31: ^{19}F -NMR spectrum of **11** (564 MHz) in $\text{DMSO-}d_6$

6-2- Synthesis of a derivative to determine the molar extinction coefficient of the second generation linker

In order to determine the molar extinction coefficient of the second generation linker, a related compound was synthesised. A three-step synthesis was performed to afford product **S24**.



➤ Synthesis of **S24**



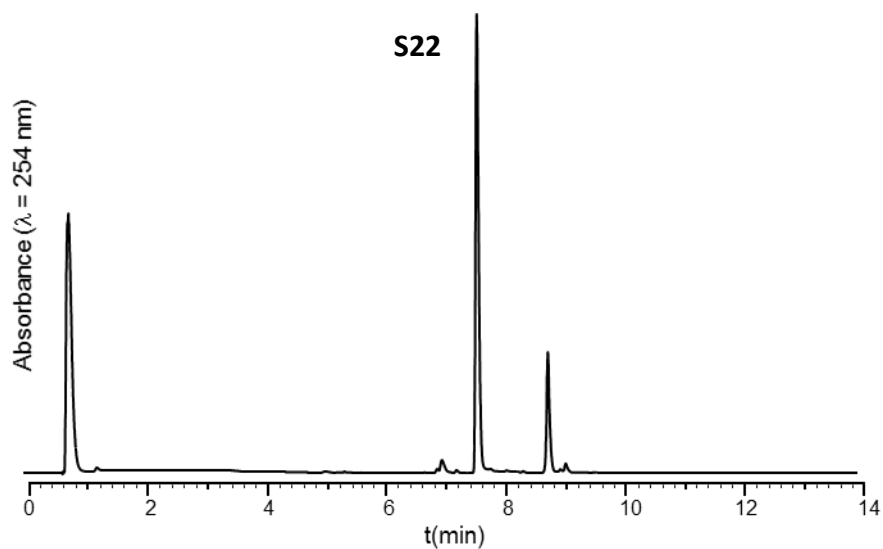
To an ice-cold solution of benzylamine (1.02 mL, 9.33 mmol) in NMP (0.05M) was added 5-formylsalicylic acid (3.1 g, 18.7 mmol, 2 equiv.), HATU (6.03 g, 15.9 mmol, 1.7 equiv.) and DIEA (3.25 mL, 18.7 mmol, 2 equiv.). The resulting solution was stirred for 24 h at 25 °C. Then the mixture was diluted with EtOAc (266 mL) and the organic layer was washed successively with 1 M HCl (x 7) and saturated NaCl aqueous solutions, dried over MgSO_4 , and concentrated under reduced pressure. The resulting crude white solid (2.01 g) was directly engaged in the next step without further purification. Analytical scale purification by flash chromatography (eluent: 9:1 DCM/ EtOAc) furnished pure **S22** for NMR characterization.

^1H NMR (600 MHz, $\text{DMSO}-d_6$): δ 13.36 (s, 1H, H_{11}), 9.64 (m, 1H, H_7), 8.48 (d, $J = 2.1$ Hz, 1H, H_8), 7.90 (dd, $J = 8.5, 2.1$ Hz, 1H, H_{10}), 7.35 – 7.28 (m, 5H, $\text{H}_{1+2+3+4+5}$), 7.05 (d, $J = 8.5$ Hz, 1H, H_9), 4.51 (d, $J = 5.9$ Hz, 2H, H_6).

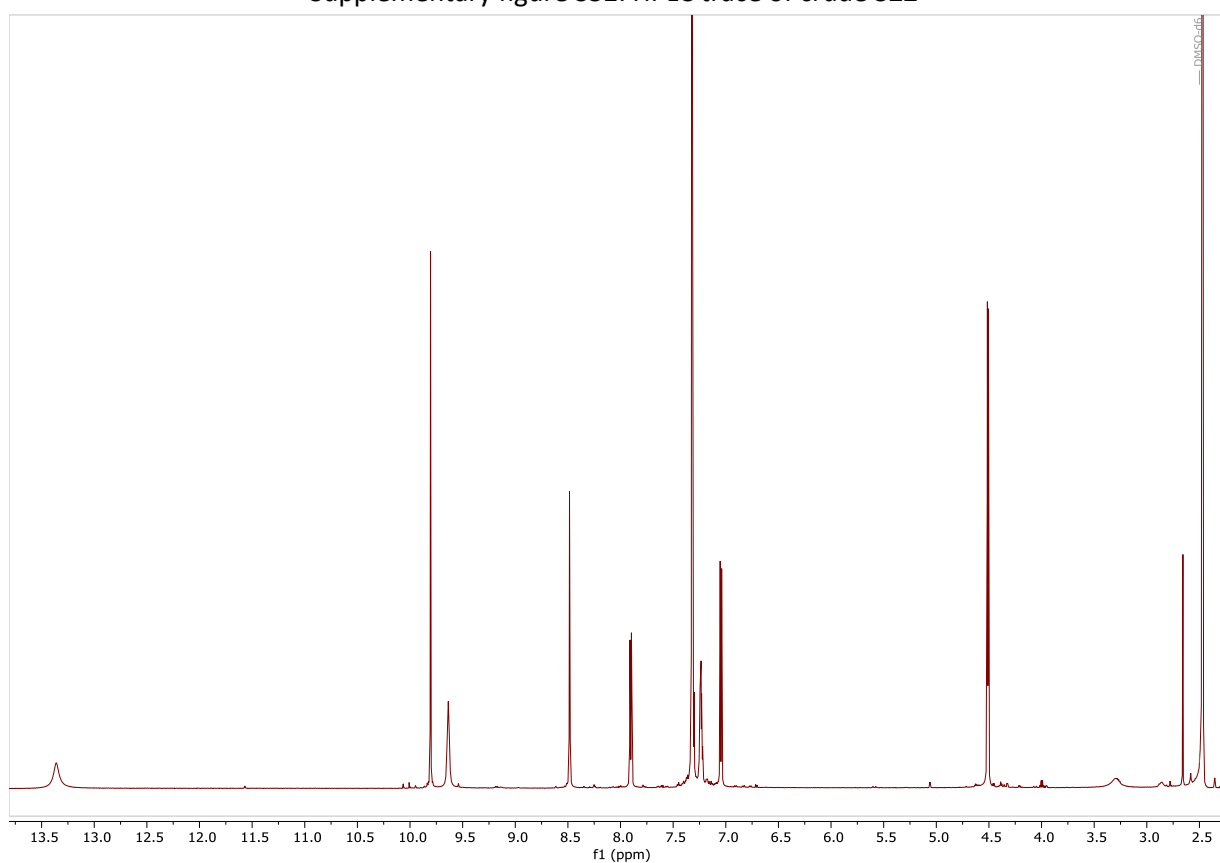
^{13}C NMR (151 MHz, $\text{DMSO}-d_6$): δ 190.7, 167.9, 165.4, 138.7, 134.3, 131.3, 128.4, 127.7, 127.4, 127.0, 118.5, 116.0, 42.5.

ESI-MS (m/z): [M] calcd. for $\text{C}_{15}\text{H}_{13}\text{NO}_3$: 255.1 found: 255.0

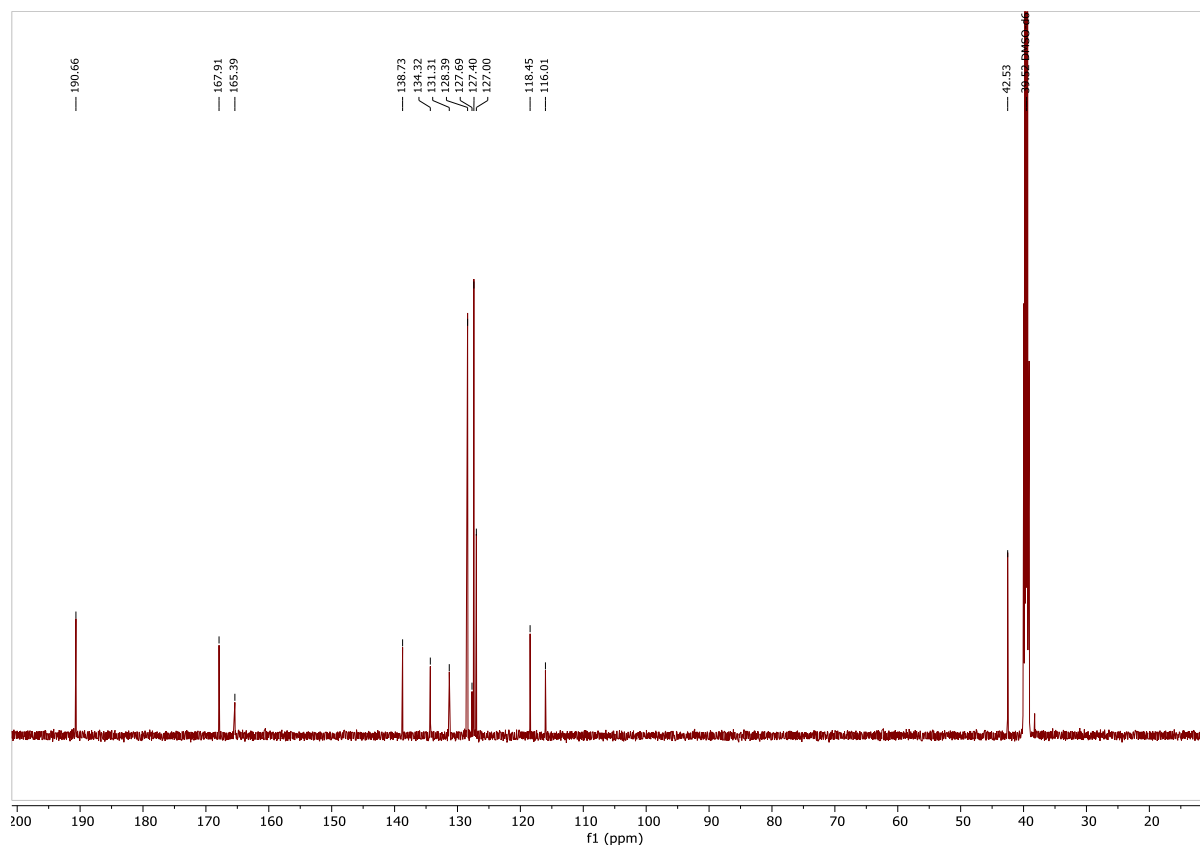
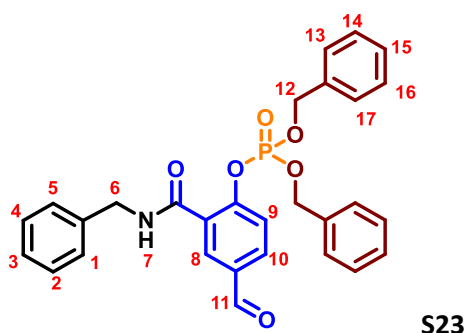
HPLC analysis: $t_R = 7.51$ min (Aeris Widepore XB-C18 2, gradient: 3-90% B' over 15 min, 60 °C)



Supplementary figure S32: HPLC trace of crude **S22**



Supplementary figure S33: ¹H-NMR spectra of **S22** (600 MHz) in DMSO-*d*₆

Supplementary figure S34: ^{13}C -NMR spectrum of **S22** (151 MHz) in $\text{DMSO}-d_6$ 

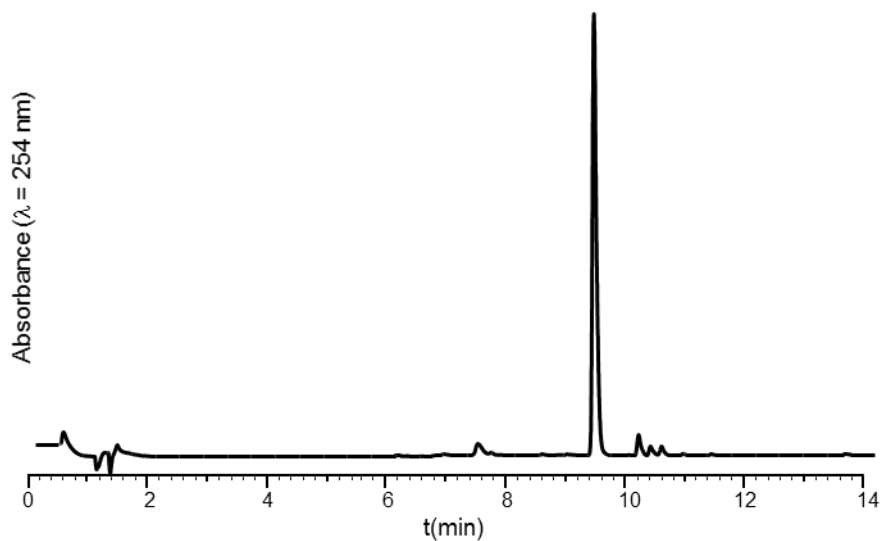
S22 (1 g, 3.92 mmol if it was pure) was dissolved in anhydrous MeCN (0.05M) by refluxing the mixture for few minutes. The resulting solution is then cooled in an ice bath and CCl_4 (1.88 mL, 19.6 mmol, 5 equiv.), DIEA (1.42 mL, 8.82 mmol, 2,1 equiv.) and DMAP (48 mg, 1.9 mmol, 0,1 equiv.) were added. One minute later, dibenzylphosphite (1.24 mL, 5.7 mmol, 1.6 equiv.) was added dropwise and the reaction was stirred for 30 min at 0 °C. Then the mixture was diluted with EtOAc (266 ml) and washed successively with 1 M HCl (x 3), saturated NaHCO_3 (x 3), and saturated NaCl aqueous solutions, dried over MgSO_4 , and concentrated under reduced pressure. The resulting crude white solid (1.77 g) was directly engaged in the next step without further purification.

^1H NMR (600 MHz, CDCl_3): δ 8.43 (m, 1H, H_7), 7.87 (dd, J = 8.5, 2.2 Hz, 1H, H_8), 7.46 (dd, J = 8.6, 2.3 Hz, 1H, H_{10}), 7.47 – 7.18 (m, 11H, $\text{H}_9+13+14+15+16+17$), 5.01 – 4.95 (m, 4H, H_{12}), 4.56 (d, J = 5.7 Hz, 2H, H_6).

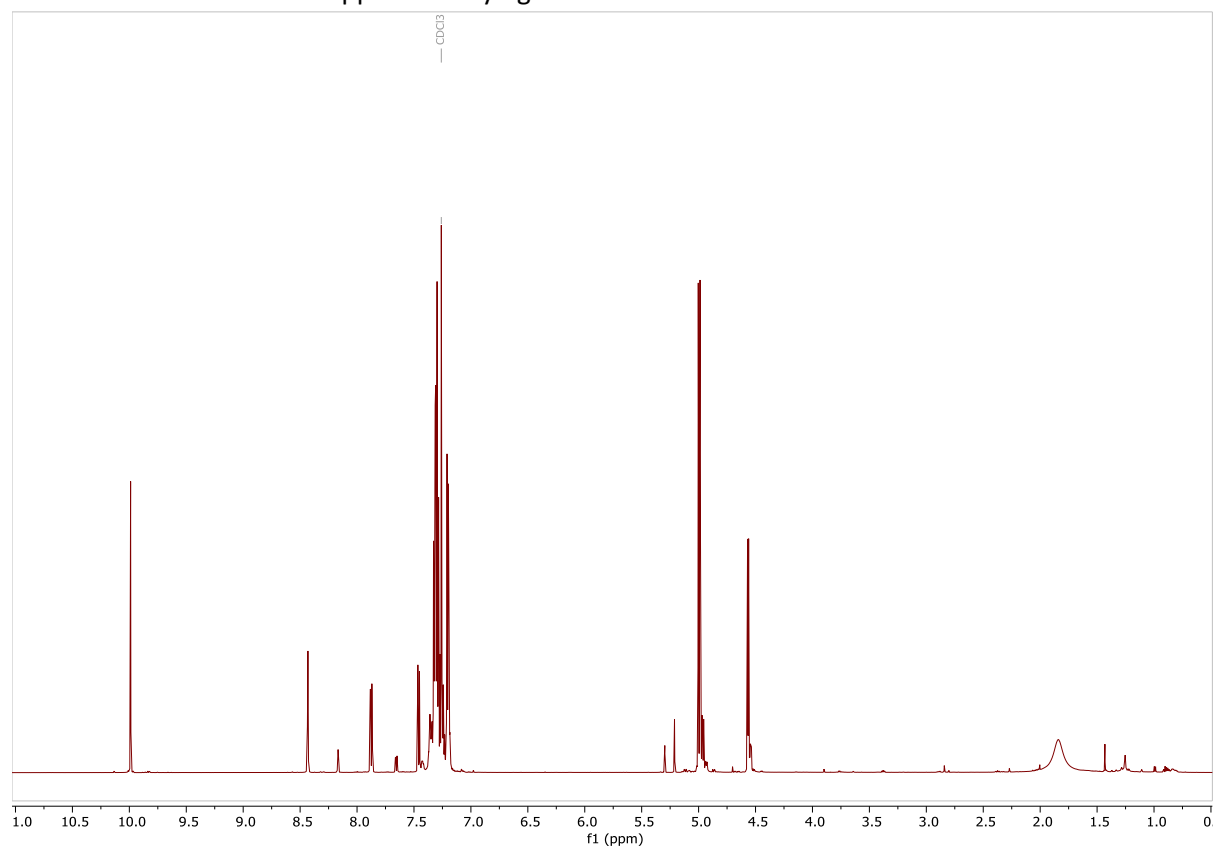
^{13}C NMR (151 MHz, CDCl_3 -*d*): δ 190.6, 163.8, 151.9, 138.1, 135.1, 132.1, 129.5, 129.3, 129.1, 129.08, 129.06, 128.6, 128.4, 128.0, 121.9, 120.5, 71.2, 44.6.

ESI-MS (m/z): [M] calcd. for $\text{C}_{29}\text{H}_{26}\text{NO}_6\text{P}$: 515.2 found: 515.2

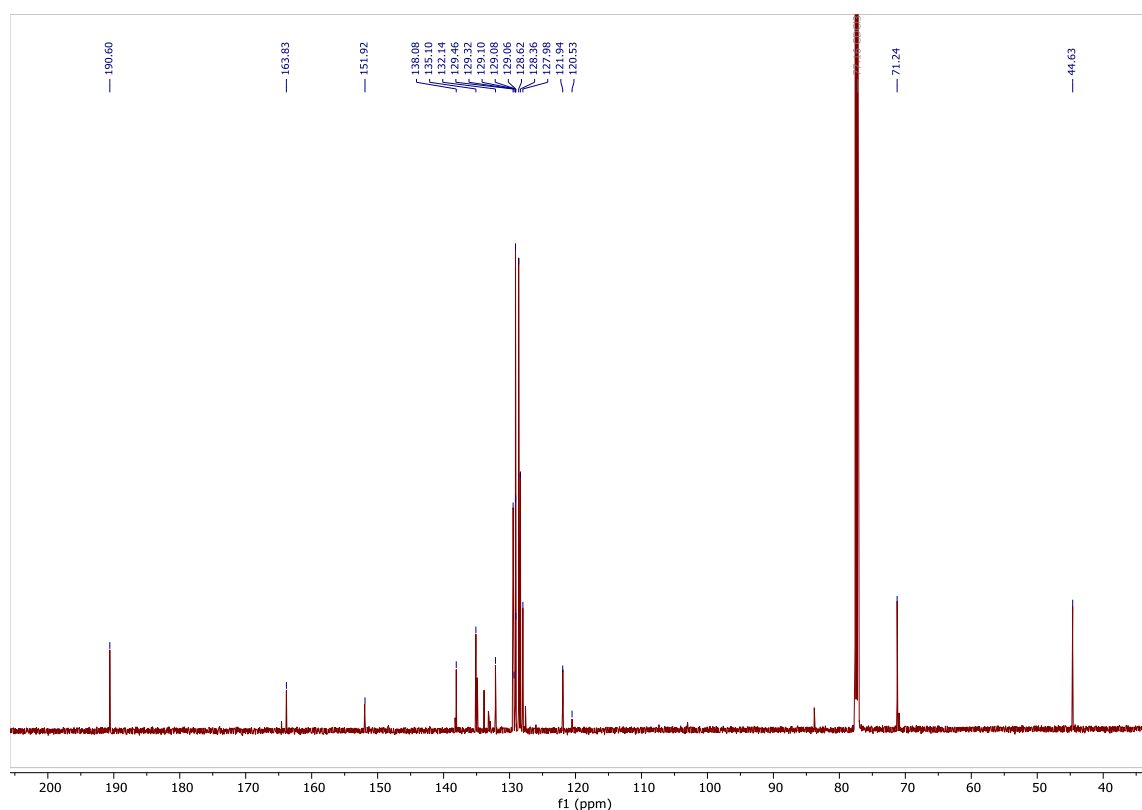
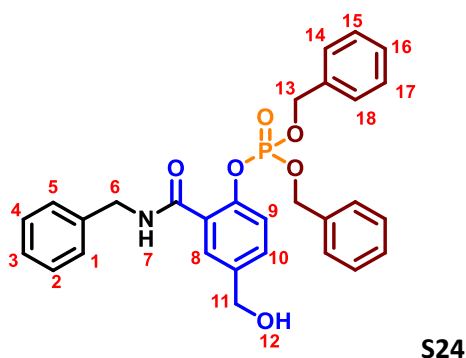
HPLC analysis: t_R = 9.53 min (Aeris Widespore XB-C18 2, gradient: 3-90% B' over 15 min, 60 °C)



Supplementary figure S35: HPLC trace of crude **S23**



Supplementary figure S36: ^1H -NMR spectra of crude **S23** (600 MHz) in CDCl_3

Supplementary figure S37: ^{13}C -NMR spectrum of crude **S23** (151 MHz) in CDCl_3 

A stirred solution of **S23** (700 mg, 1.36 mmol if it was pure) in anhydrous THF (13.6 mL, 0.2 M aldehyde) was cooled to 0° C and NaBH_4 (51 mg, 1.36 mmol, 1 equiv.) was added. The resulting reaction mixture was stirred at 0° C for 15 min. Then the mixture was diluted with EtOAc and washed successively with 1 M HCl (x3), saturated NaHCO_3 (x3), and saturated NaCl aqueous solutions, dried over MgSO_4 , and concentrated under reduced pressure. The product was purified by chromatography (eluent: 8:2 DCM/EtOAc) affording an amorphous white solid (695 mg, 37% over three steps).

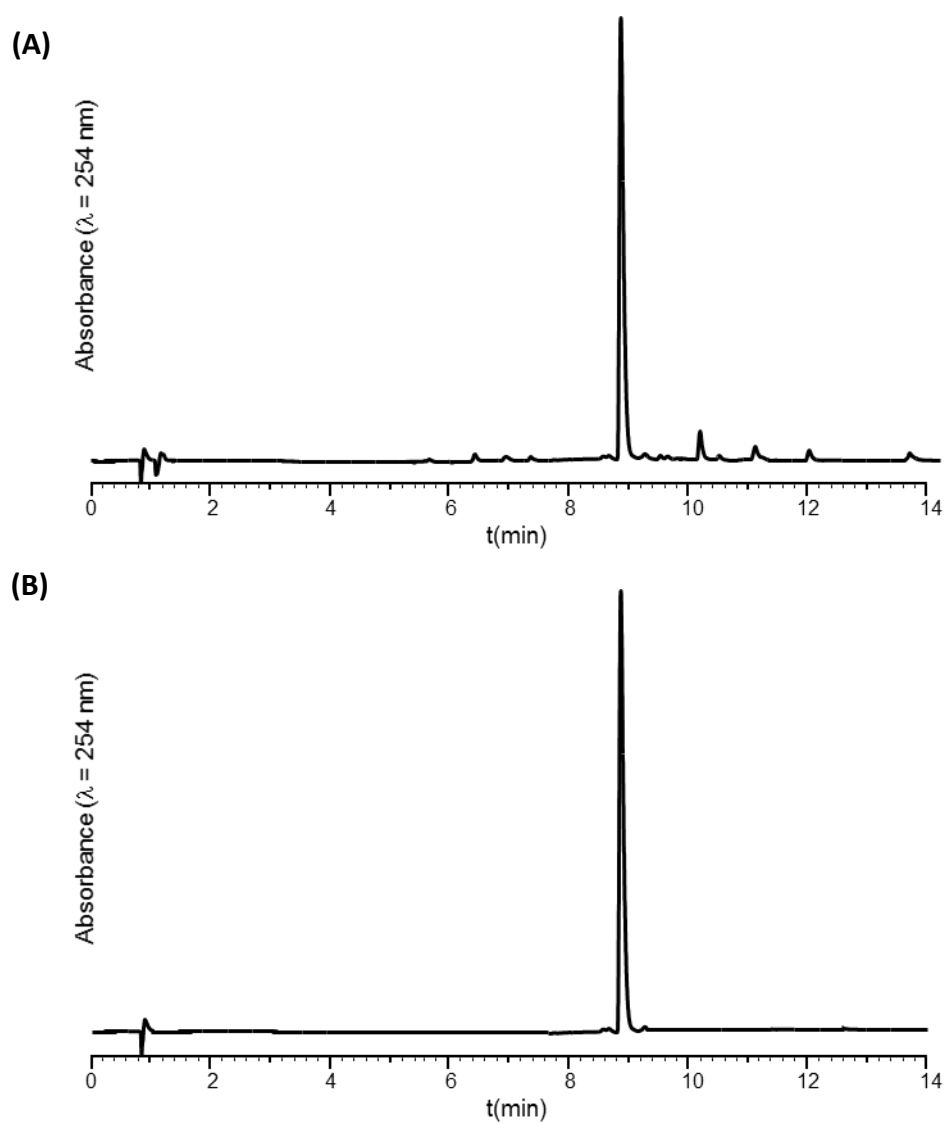
^1H NMR (600 MHz, CDCl_3): δ 7.45 (t, J = 5.8 Hz, 1H, H_7), 7.37 (dd, J = 8.6, 2.3 Hz, 1H, H_8), 7.35 – 7.16 (m, 12H, $\text{H}_{9+10+14+15+16+17+18}$), 4.93 – 4.85 (m, 4H, H_{13}), 4.59 (bs, 2H, H_{11}), 4.47 (d, J = 5.7 Hz, 2H, H_6).

^{13}C NMR (151 MHz, CDCl_3): δ 165.1, 146.8, 138.3, 135.2, 130.8, 130.1, 129.2, 129.0, 128.9, 128.8, 128.5, 128.3, 127.8, 121.3, 120.4, 70.8, 64.4, 44.5.

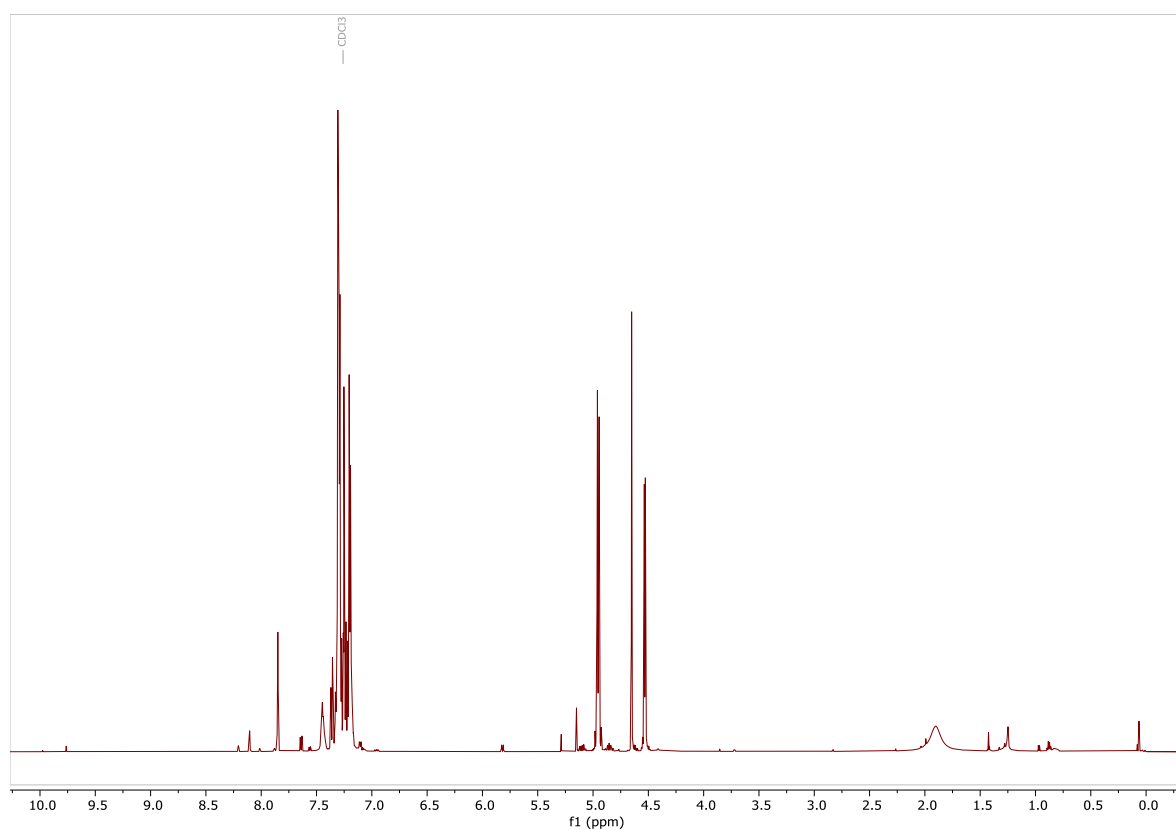
Molar extinction coefficient: $\epsilon^{280} = 2718 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$ in 5:5:0.01 $\text{H}_2\text{O}/\text{MeCN}/\text{TFA}$

ESI-MS (m/z): [M] calcd. for $\text{C}_{29}\text{H}_{28}\text{NO}_6\text{P}$: 517.2 found: 517.2

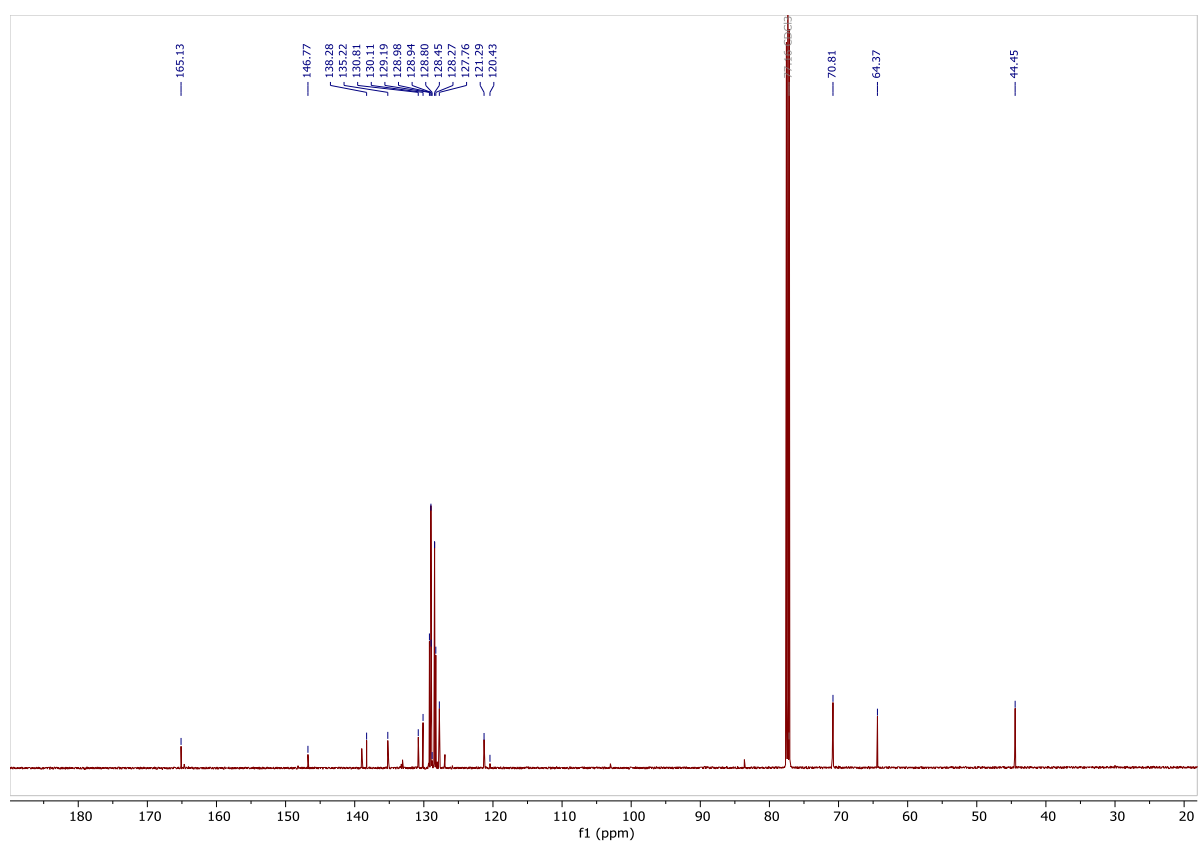
HPLC analysis: t_R = 9.01 min (Aeris Widespore XB-C18 2, gradient: 3-90% B' over 15 min, 60 °C)



Supplementary figure S38: HPLC traces of crude (A) and purified (B) **S24**



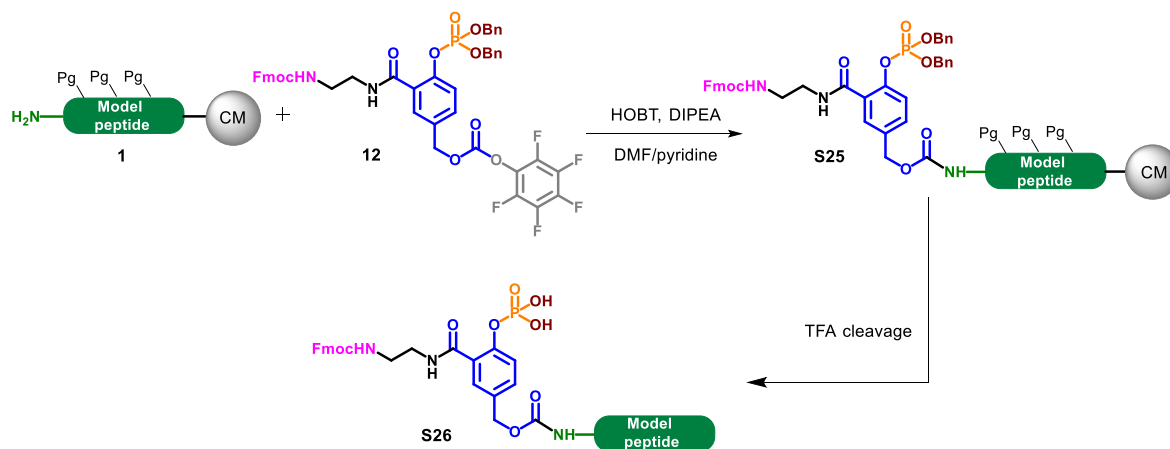
Supplementary figure S39: ^1H -NMR spectra of **S24** (600 MHz) in CDCl_3



Supplementary figure S40: ^{13}C -NMR spectrum of **S24** (151 MHz) in CDCl_3

6-3- Linker introduction on model peptide

The same model peptide **1** was used for all the optimization work with the second generation linker

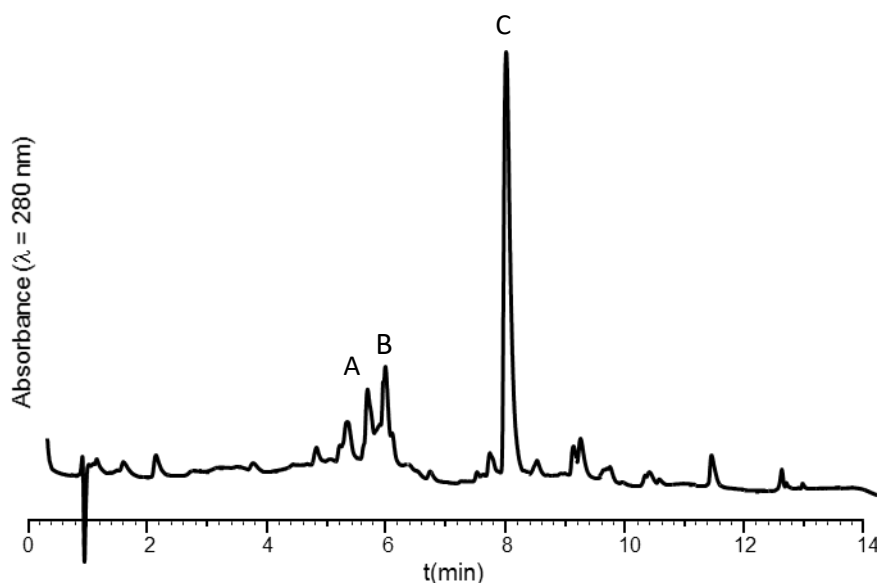


Peptidyl resin **1** (25 μ mol) was washed three times by 0.1% TFA in DCM and 1% DIEA in DCM. To this resin was added a solution of second generation linker precursor **12** (34 mg, 1.5 equiv.) in 1:1 DMF/pyridine (10.5 mM solid-supported peptide concentration), DIEA (9.2 μ L 1.8 equiv.) and hydroxybenzotriazole hydrate (HOBT) (3.8 mg, 1.5 equiv.). The mixture was stirred at room temperature for 5 h and the same coupling procedure was repeated one more time. Then the resin was washed with DCM/NMP. Capping of potential unreacted amine groups was achieved by treatment with acetic anhydride as described in protocol PS1. The reaction was monitored by Kaiser Test [2] and analytical scale TFA cleavage (protocol PS3).

- Compound **S26**

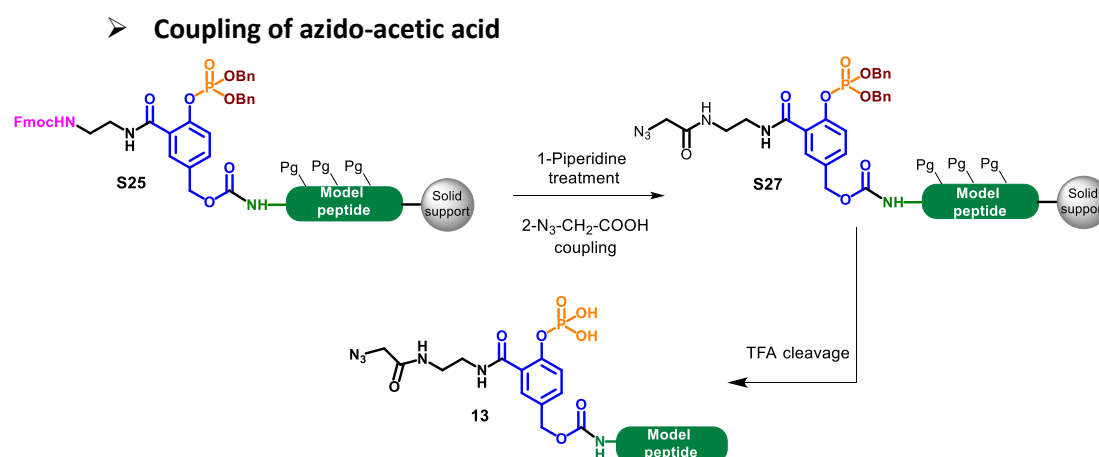
ESI-MS (m/z): [M] calcd. for $C_{197}H_{286}N_{55}O_{64}P$: 4479.8, found: 4479.0 (average mass, deconvoluted)

HPLC analysis: t_R = 8.12 min (Aeris Widepore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C)



Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (5.64)	3941.2	3940.7	1
B (5.98)	3983.3	3982.6	Ac-(Muc1) ₂ -W-NH ₂
C (8.12)	4479.8	4479.0	S26

Supplementary figure S41: HPLC trace and MS analysis of crude **S26** (average masses, deconvoluted)



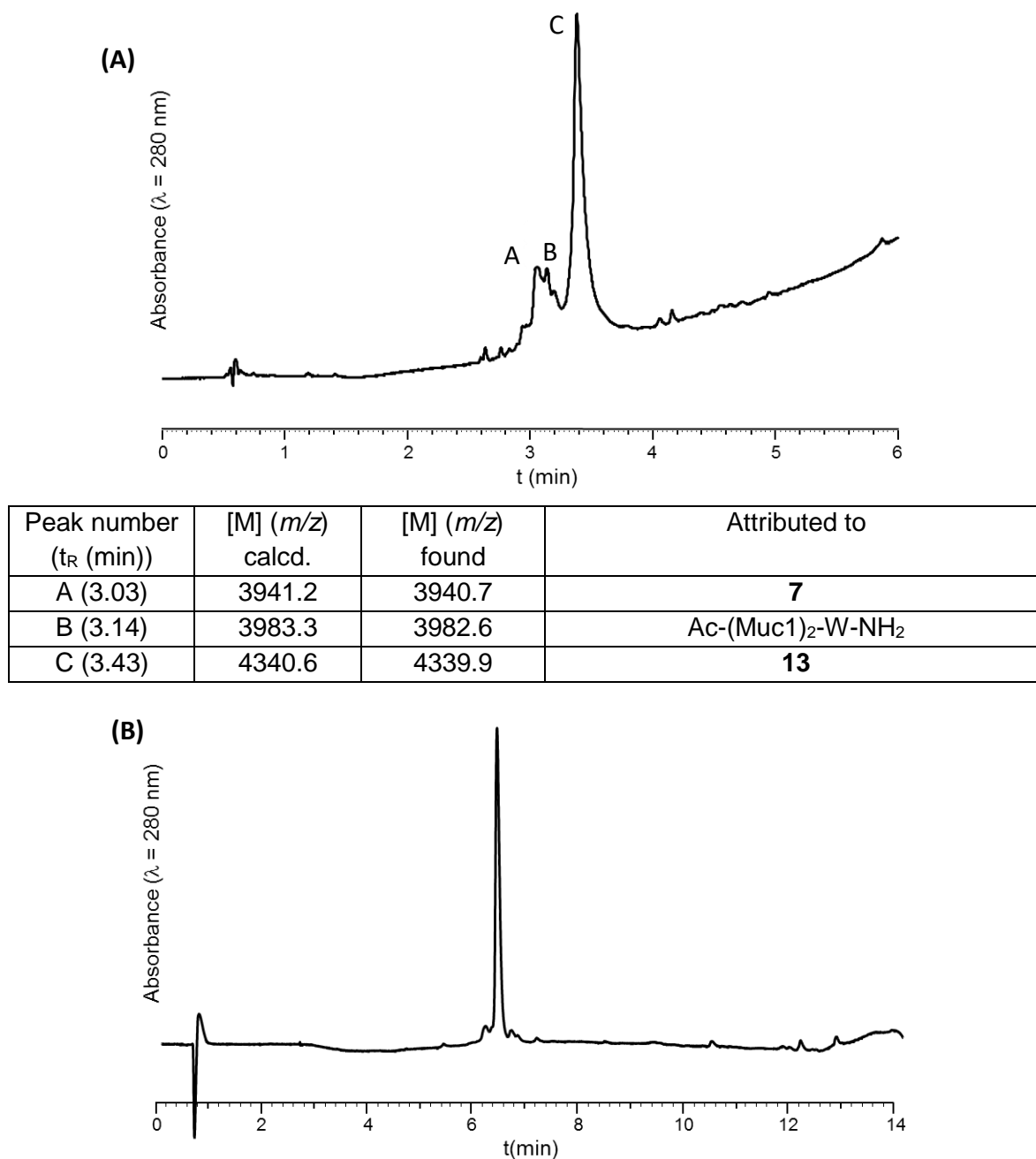
Fmoc group was deprotected by three successive treatments with 20% piperidine in NMP for 3 min. Peptidyl resin **S25** (25 μ mol) was washed with NMP (5 mL, 1 min). N_3-CH_2-COOH (19 μ L, 10 equiv.) in NMP (1.5 mL), DIC (39 μ L, 10 equiv.) and Oxyma (36 mg, 10 equiv.) in NMP (2 mL,) were added and the reactor was stirred for 6 h followed by extensive washing. Next, the same coupling procedure was repeated one more time. Capping of potential unreacted amine groups was achieved by treatment

with acetic anhydride as described in protocol PS1. The reaction was monitored by Kaiser Test [2] and analytical scale TFA cleavage (protocol PS3).

ESI-MS (m/z): [M] calcd. for $C_{184}H_{278}N_{58}O_{63}P$: 4340.6, found: 4339.9 (average mass, deconvoluted)

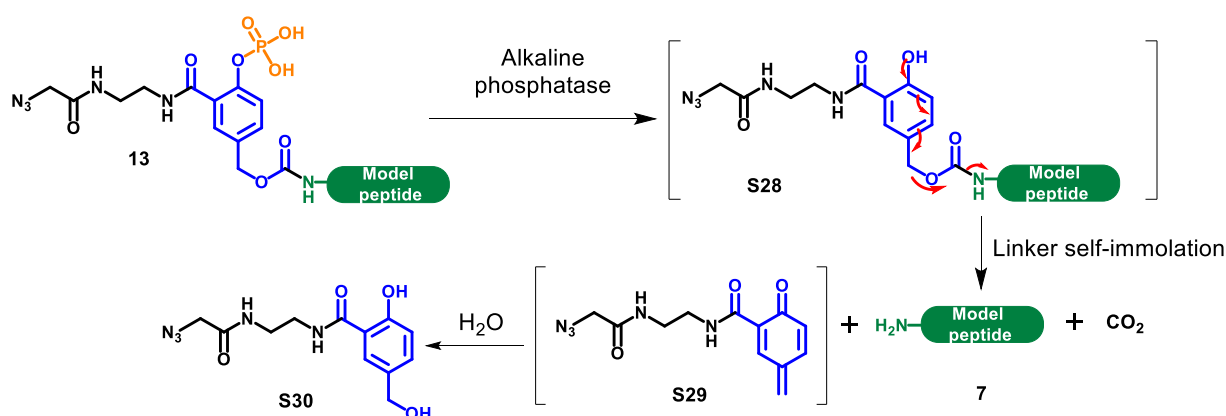
HPLC analysis: t_R = 6.54 min (Aeris Widepore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C)

HPLC purification: Nucleosil C18, gradient: 20-31% B over 20 min, 23%



Supplementary figure S42: HPLC traces and MS analysis (average masses, deconvoluted) of crude (A) and purified (B) compound **13** (Chromolith, gradient: 5-50% B over 5 min and Aeris Widepore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C, respectively)

6-4- Solution phase enzymatic assays



Enzymatic linker cleavage was carried out with commercial alkaline phosphatases.* The azidolinker-functionalized peptide **13** was incubated at 37 °C (except for Lambda PP where $T = 30$ °C according to manufacturer recommendations) with the enzyme (0.4 enzyme unit/10nmol of peptide) in the buffer** (100 μM peptide concentration). Linker cleavage was monitored by analytical HPLC. Note that intermediates **S28** and **S29** were not observed in the LC-MS monitoring of the reaction, indicating fast linker self-immolation through 1,6-elimination and water addition on the quinone methide.

Relative rates were determined by HPLC peak integration at $\lambda = 280$ nm, taking into account the predicted molar absorption coefficient of Trp and the 2nd generation linker (see p S41-S47) at 280 nm: $\epsilon_{\text{Trp}} = 5500 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$ [3]; $\epsilon_{\text{linker}} = 2718 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$.

Enzymes and buffers used

❖ Alkaline Phosphatase Calf Intestinal (CIP)

*- Calf Intestinal Alkaline Phosphatase (CIP) enzyme: 1,000 U/mL, suspension in 50% glycerol, 10 mM Tris-HCl buffer salt, 1 mM MgCl_2 , 50 mM KCl, 0.1 mM ZnCl_2 , pH 8.2

** - Alkaline Phosphatase commercial buffer: 50 mM Tris-HCl, 0.1 mM EDTA, pH 8.5

❖ Antarctic Phosphatase

*-Antarctic Phosphatase: 5,000 U/mL, suspension in 50% glycerol, 10 mM Tris-HCl, 1 mM MgCl_2 , 0.01 mM ZnCl_2 , pH 7.4

** - Antarctic Phosphatase commercial buffer: 50 mM Bis-Tris-Propane HCl, 1 mM MgCl_2 , 0.1 mM ZnCl_2 , pH 6

❖ Shrimp Alkaline Phosphatase (rSAP)

*-Shrimp Alkaline Phosphatase (rSAP) enzyme: 1,000 U/mL, suspension in 50% glycerol, 25 mM Tris-HCl, 1 mM MgCl_2 , pH 7.5

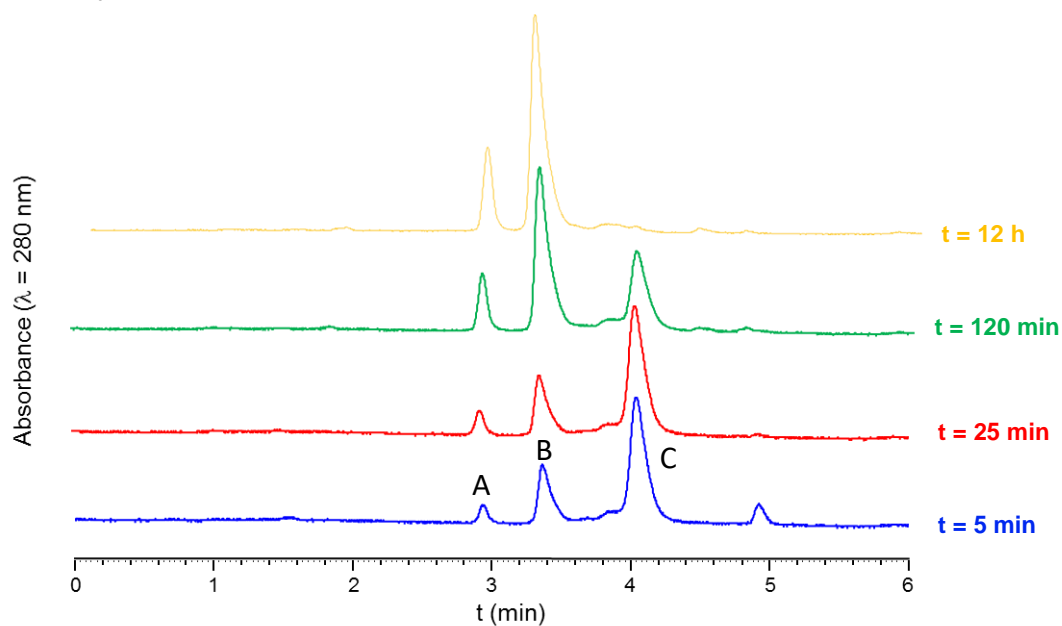
** -Shrimp Phosphatase commercial buffer: 50 mM Tris-HCl, 10 mM MgCl_2 , pH 9

❖ Lambda Protein Phosphatase

*-Lambda Protein Phosphatase (Lambda PP) : 400,000 U/mL, suspension in 50% glycerol, 50 mM HEPES, 100 mM NaCl, 2 mM DTT, 0.01% Brij 35, 0.1 mM EGTA, 0.1 mM MnCl_2 , pH 7.5

** - Protein MetalloPhosphatases buffer (PMP): 50 mM HEPES, 100 mM NaCl, 2 mM DTT, 0.01% Brij 35, pH 7.5 (supplemented with 1 mM MnCl_2 just prior to use)

** - CutSmart buffer: 50 mM potassium acetate, 20 mM Tris-Acetate, 10 mM magnesium acetate, 100 $\mu\text{g/ml}$ BSA, pH 7.9



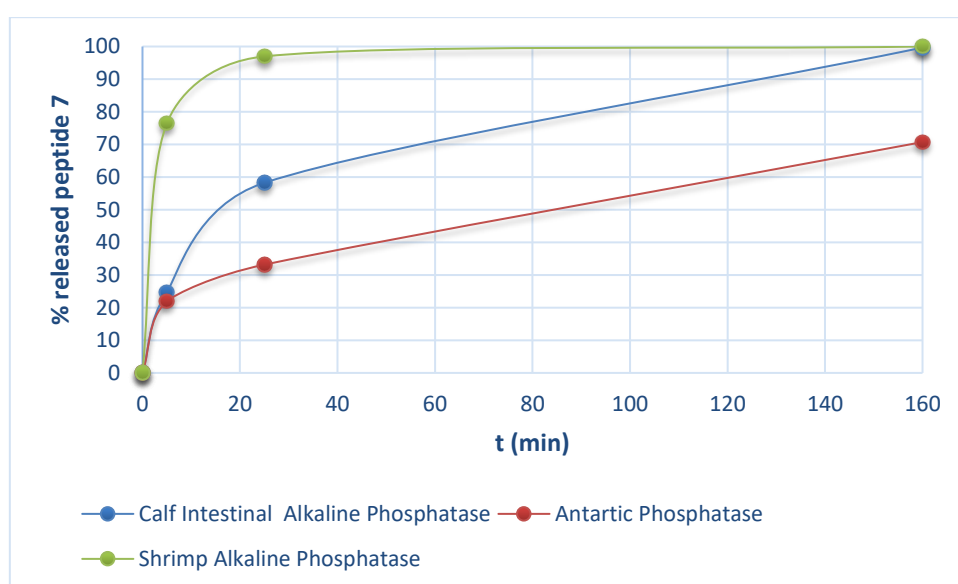
Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (2.98)	293.1	293.1	S30
B (3.37)	3941.2	3940.5	7
C (4.03)	4340.6	4339.9	13

Supplementary figure S43: HPLC traces (Chromolith, gradient: 5-50% B over 5 min) and MS analysis (average masses, deconvoluted, for peaks B and C) of an example of enzymatic linker cleavage (enzyme = antartic phosphatase)

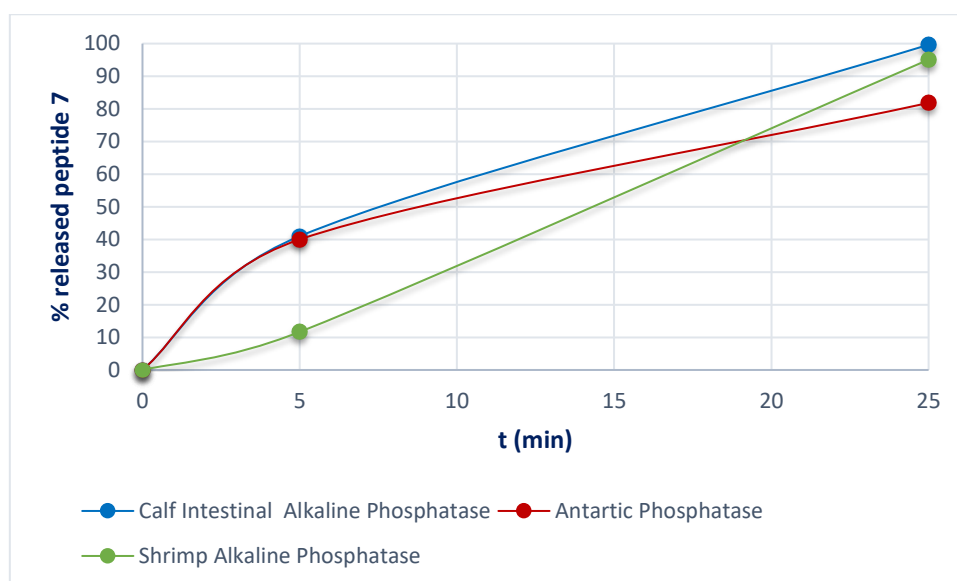
- Compound **7**

ESI-MS (m/z): [M] calcd. for $C_{171}H_{263}N_{53}O_{55}$: 3941.2, found: 3940.7 (average mass, deconvoluted)

HPLC analysis: t_R = 3.37 min (Chromolith, gradient: 5-50% B' over 5 min)

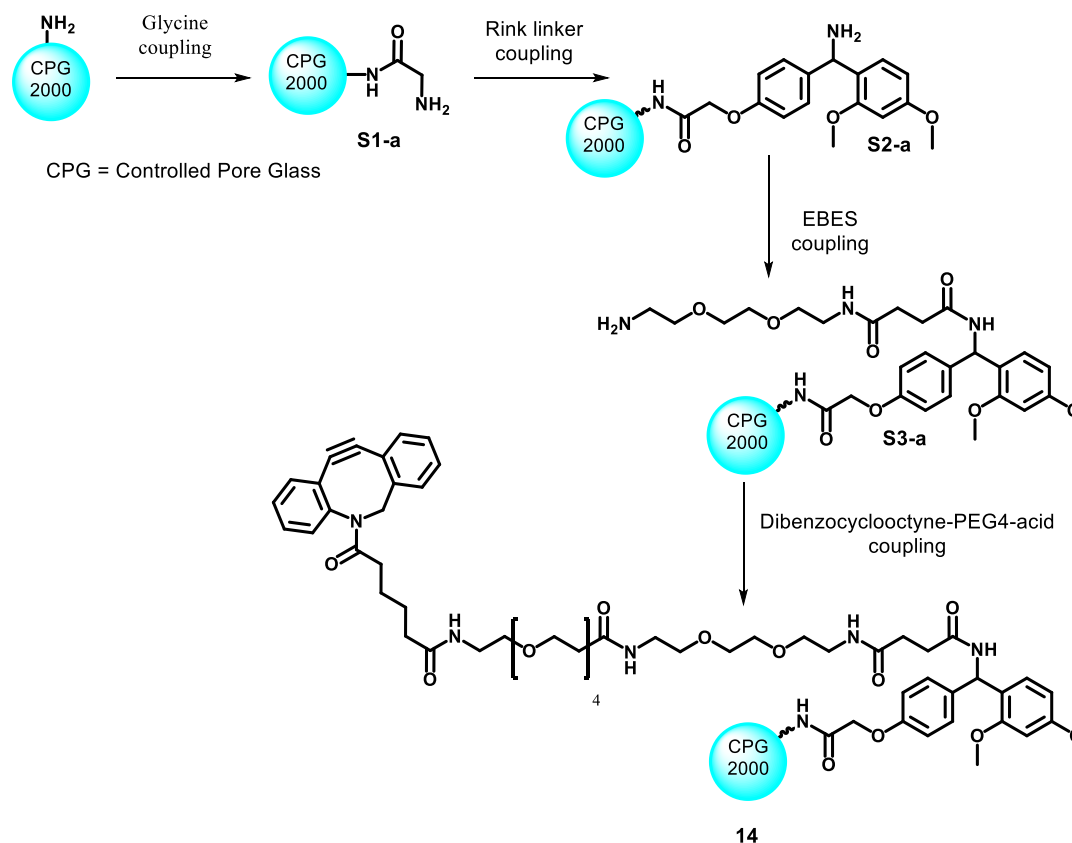


Supplementary figure S44: Solution phase enzymatic release of the model peptide **7** from **13** using different phosphatases, with the commercial buffer for each enzyme



Supplementary figure S45: Graphical representation of the in solution enzymatic release of the model peptide **7** from **13** using different phosphatases with the CutSmart buffer

6-5- Solid supported enzymatic cleavage

➤ Synthesis of DBCO-EBES-NH₂ on CPG 200 resin

Aminopropyl CPG2000 (25 μmol , 39 $\mu\text{mol/g}$) was washed with NMP (5 mL, 1 min). Fmoc-Gly-OH (112 mg, 15 equiv.) in NMP (1.5 mL), PyAOP (261 mg, 20 equiv.) in NMP/DCM 65:35 (2 mL, final NMP/DCM ratio 8:2) and DIEA (131 μL , 30 equiv.) were added and the reactor was stirred for 6 h followed by extensive washing. Capping of potential unreacted amine groups was achieved through treatment with acetic anhydride as described in protocol PS1. Fmoc group was deprotected by three successive treatments with 20% piperidine in NMP for 3 min. The same procedure was repeated for the coupling of Fmoc-Rink-OH (202 mg, 15 equiv.) and Fmoc-EBES-OH (176 mg, 15 equiv.).

Dibenzocyclooctyne-PEG4-acid (DBCO) (218 mg, 15 equiv.) in NMP (1.5 mL), PyAOP (261 mg, 20 equiv.) in NMP/DCM 65:35 (2 mL, final NMP/DCM ratio 8:2) and DIEA (131 μL , 30 equiv.) were added and the reactor was stirred for 12 h followed by extensive washing.

To quantify DBCO loading, a small amount of the solid support was treated with Fmoc-Lys(N₃)-OH (25 equiv.) dissolved in H₂O/MeCN/AcOH 75:20:5 and the reaction was stirred for 24 h at 37 °C followed by extensive washing and UV titration of the fluorenylmethyl-piperidine adduct after Fmoc deprotection.

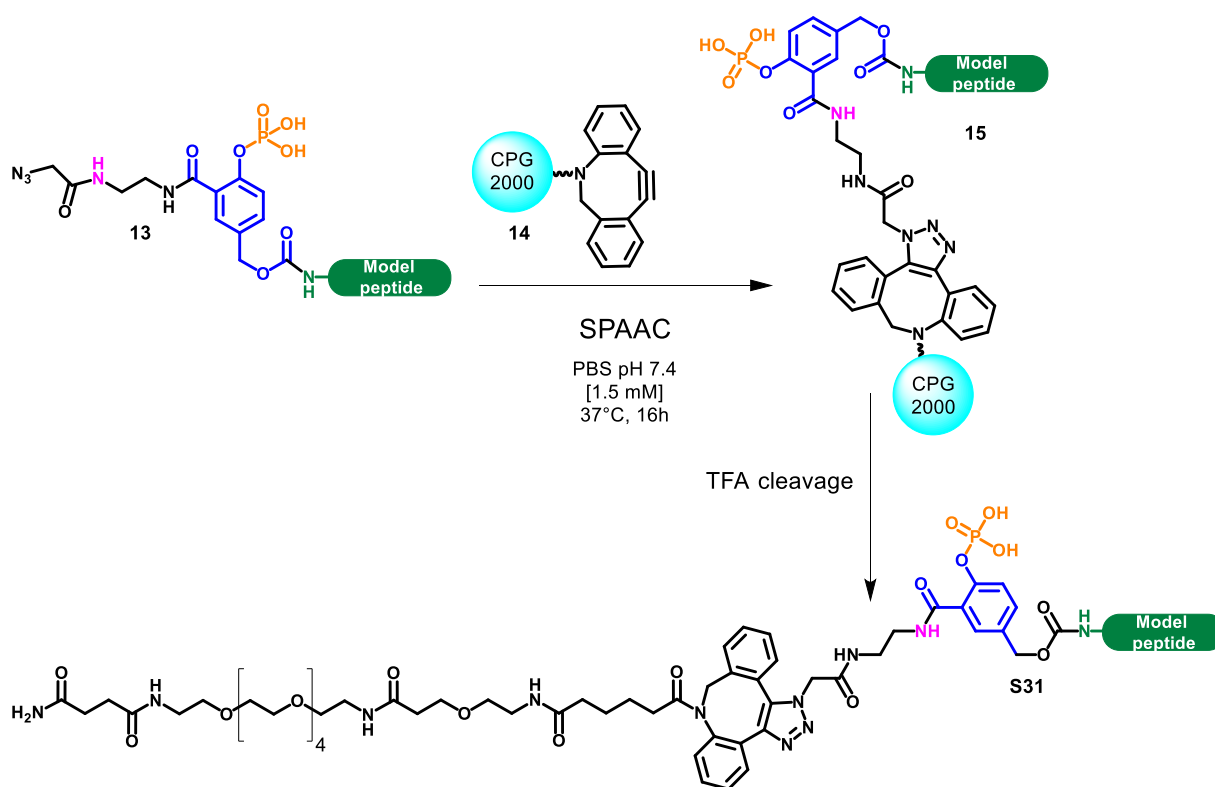
Note that DBCO undergoes a rearrangement upon TFA treatment used for analytical Rink linker cleavage, leading to a product with the same mass but a distinctive UV spectrum [4].

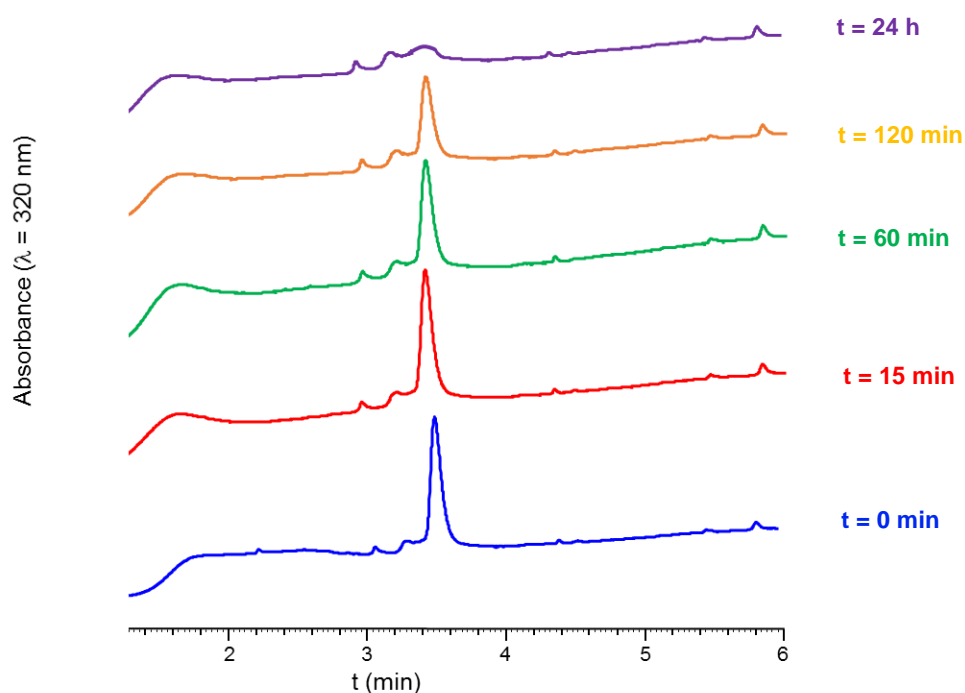
We also noticed that the DBCO solid support **14** was not stable for long term storage even when conserved at -80 °C, therefore, **14** should be engaged in the SPAAC reaction directly after functionalization.

➤ **Strain-promoted azide/alkyne cycloaddition to graft the azidolinker-derivatized peptide (**13**) on solid support (**14**)**

Azidolinker-derivatized peptide **13** (1.3 mg, 0.5 µmol) was dissolved in 333 µL of PBS (1.5 mM peptide concentration), pH 7.4 and then added to DBCO-functionalized solid support **14** (1.5 µmol, 3 equiv.) and the reaction was stirred for 16 h at 37 °C. Capping of the excess of DBCO was achieved by treatment with 15 equiv. of azidoacetic acid for 5 h under the same conditions.

Note that the reaction did not work when performed in H₂O/MeCN/AcOH 75:20:5 or in pure water. No disappearance of **13** was noticed in those cases. Also, 3 equiv. of the DBCO solid support were needed to drive the reaction to completion.





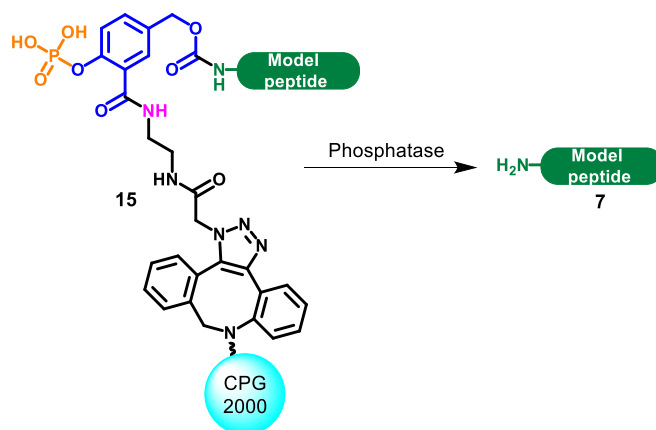
Supplementary figure S46: HPLC traces (Chromolith, gradient: 5-50% B over 5 min) of SPAAC ligation monitoring showing the disappearance of the starting azide **13**

- Compound **S31**

ESI-MS (m/z): [M] calcd. for $C_{215}H_{315}N_{62}O_{69}P$: 4903.2, found: 4902.5 (average mass, deconvoluted)

HPLC analysis: t_R = 4.72 min (Chromolith, gradient: 51-50% B over 5 min)

➤ **Enzymatic cleavage of the linker**

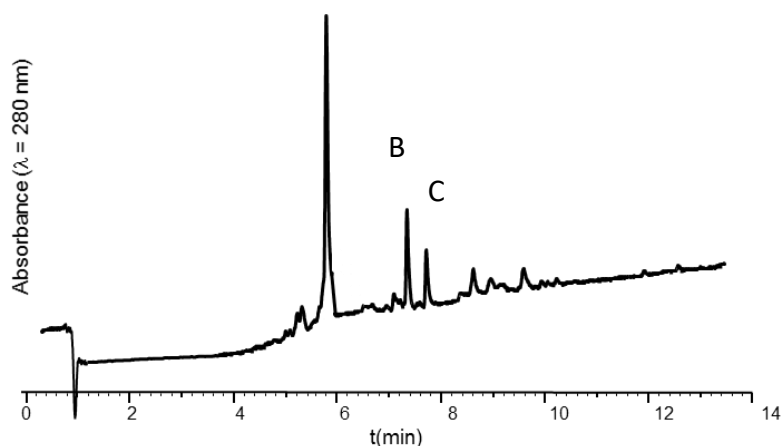


Enzymatic cleavage of the linker was carried out with commercial phosphatases*. The CGP2000-supported peptide **15** was incubated with the enzyme (4 enzyme unit/10nmol of peptide) in the appropriate buffer** (100 μ M peptide concentration). Release in solution of peptide **7** was monitored by analytical HPLC.

*Shrimp alkaline phosphatase ** CutSmart buffer at 37 °C

*Lambda phosphatase ** PMP buffer at 30 °C

A



Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (5.90)	3941.2	3940.7	7
B (7.39)	-	747.2	Non peptidic compound*
C (7.71)	-	883.3	Non peptidic compound*

*: Note that peaks B and C are also observed when treating the starting alkyne solid support **14** under the same conditions, and thus does not correspond to a peptide.

Supplementary figure S47: Typical HPLC trace (Aeris Widepore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C) and MS analysis (average masses, deconvoluted, for peak A) of compound **7** released through lambda phosphatase-mediated linker cleavage from **15**

- Compound **7**

ESI-MS (m/z): calcd. for $C_{171}H_{263}N_{53}O_{55}$: 3941.3, found: 3940.7 (average mass, deconvoluted)

HPLC analysis: t_R = 5.90 min Aeris Widepore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C)

7- Kinetics studies

The goal of this study was to obtain a rough evaluation of kinetic constants in order to compare the cleavage rate on support and in solution for each enzyme, and not a much more demanding complete enzyme kinetics studies, out of the scope of this work. Therefore, enzymatic release mediated reactions were considered as a pseudo first-order reactions, with a k_{obs} linearly dependant to enzyme concentration, as generally assumed for cases where $K_m \gg [S]$.

Observed rate constants (k_{obs}) were calculated according to the classical method for first-order reactions rate determination: A plot was drawn with $\ln [\text{substrate}]$ (mol/L) vs. time (10^3 s) and the slope of the linear fit gave the rate constant in 10^{-3} s^{-1} .

All enzymatic assays were performed at 100 μM initial linker-peptide concentration. As a consequence of the much slower kinetics observed in solution vs solid-phase reactions, different amounts of enzymes were used in order to monitor the reaction during a few hours: a 50-fold increase in the case of b-galactosidase, and a 10 fold for phosphatases.

For a direct comparison of reaction rates in solution and on solid support, a normalized rate constant k' was therefore calculated by dividing k_{obs} by the relative concentration of enzymes, expressed in enzyme unit (U) per nmol of linker-peptide.

The buffer used for each enzyme was chosen following our different optimization works:

- For β -galactosidase: incubation at 37 °C with the enzyme in a 20 mM phosphate buffer, 0.1 mM Mg^{2+} , pH 7.3.
- For Shrimp alkaline phosphatase: incubation at 37 °C with the enzyme in a solution of 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 μ g/ml BSA PEG, pH 7.9 (CutSmart buffer).
- For Lambda phosphatase: incubation at 30 °C with the enzyme in 50 mM HEPES, 100 mM NaCl, 2 mM DTT, 0.01% Brij 35, 0.1 mM EGTA, 0.1 mM $MnCl_2$, pH 7.5 (PMP buffer).

Table S4: Comparison of $k'_{\text{solution}}/k'_{\text{support}}$ for the different solid supports explored for the β -galactosidase-cleavable linker

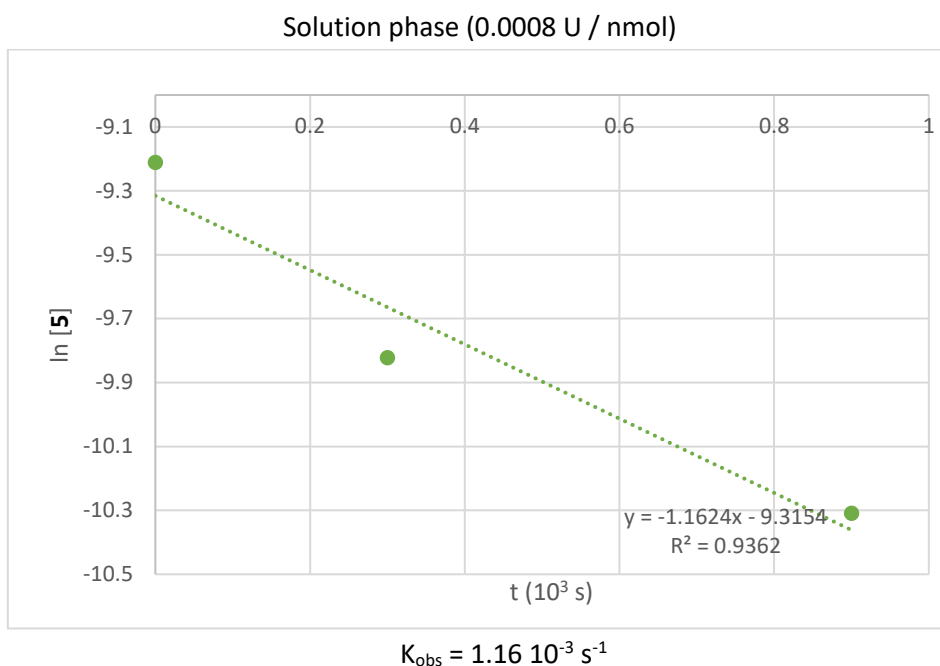
Enzyme	Solution-phase cleavage				Cleavage on solid support					$k'_{\text{solution}}/k'_{\text{support}}$
	substrate	$U_{\text{enzyme}}/\text{nmol}$	$k_{\text{obs}} (10^{-3} \text{ s}^{-1})$	$k' (10^{-3} \text{ s}^{-1})$	substrate	$U_{\text{enzyme}}/\text{nmol}$	solid support	$k_{\text{obs}} (10^{-3} \text{ s}^{-1})$	$k' (10^{-3} \text{ s}^{-1})$	
β -galactosidase	5	0.0008	1.16	1450	9a	0.4	CPG2000	0.0094	0.0235	61 702
							CPG1000	0.0007	0.0018	805 555
							PEGA ¹⁹⁰⁰	Cannot be fitted to a first order kinetic curve		

Table S5: Comparison of $k'_{\text{solution}}/k'_{\text{support}}$ for the different phosphatases used in this work

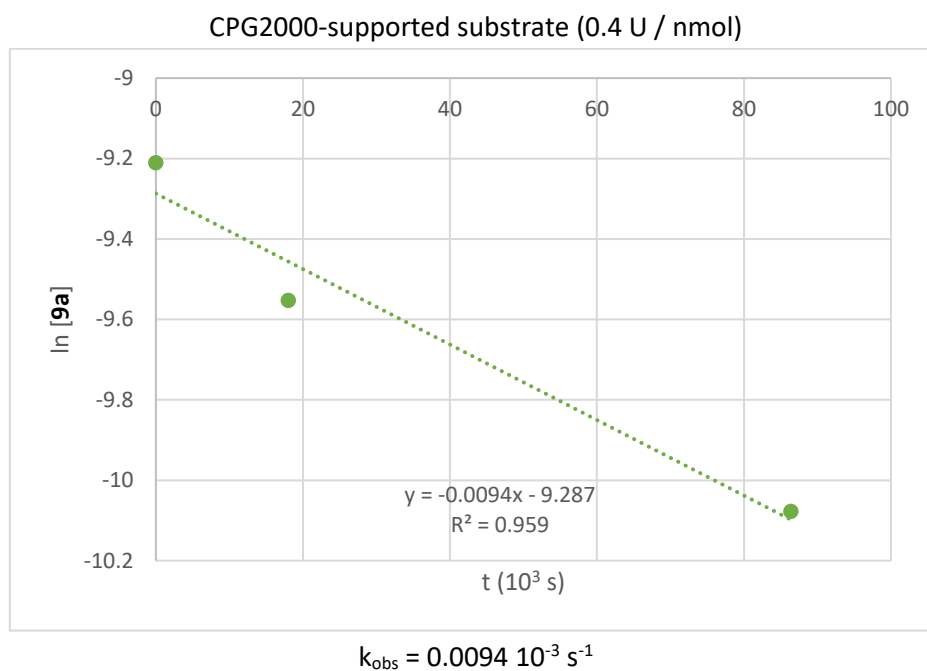
Enzyme	Solution-phase cleavage				Cleavage on solid support				$k'_{\text{solution}}/k'_{\text{support}}$
	substrate	$U_{\text{enzyme}}/\text{nmol}$	$k_{\text{obs}} (10^{-3} \text{ s}^{-1})$	$k' (10^{-3} \text{ s}^{-1})$	substrate	$U_{\text{enzyme}}/\text{nmol}$	$k_{\text{obs}} (10^{-3} \text{ s}^{-1})$	$k' (10^{-3} \text{ s}^{-1})$	
β -galactosidase	5	0.0008	1.16	1450	9a	0.4	0.0094	0.0235	61 702
Shrimp phosphatase	13	0.04	2.16	54	15	0.4	0.0026	0.0065	8 308
Lambda phosphatase	13	0.04	0.147	3.68	15	0.4	0.0072	0.0180	204

Note that the kinetic constant cannot be compared from one enzyme to another, as the definition of one enzyme unit (U) varies. However, the normalized $k'_{\text{solution}}/k'_{\text{support}}$ can be directly compared and is representative of the relative efficiency of enzyme cleavage on solid support and in solution.

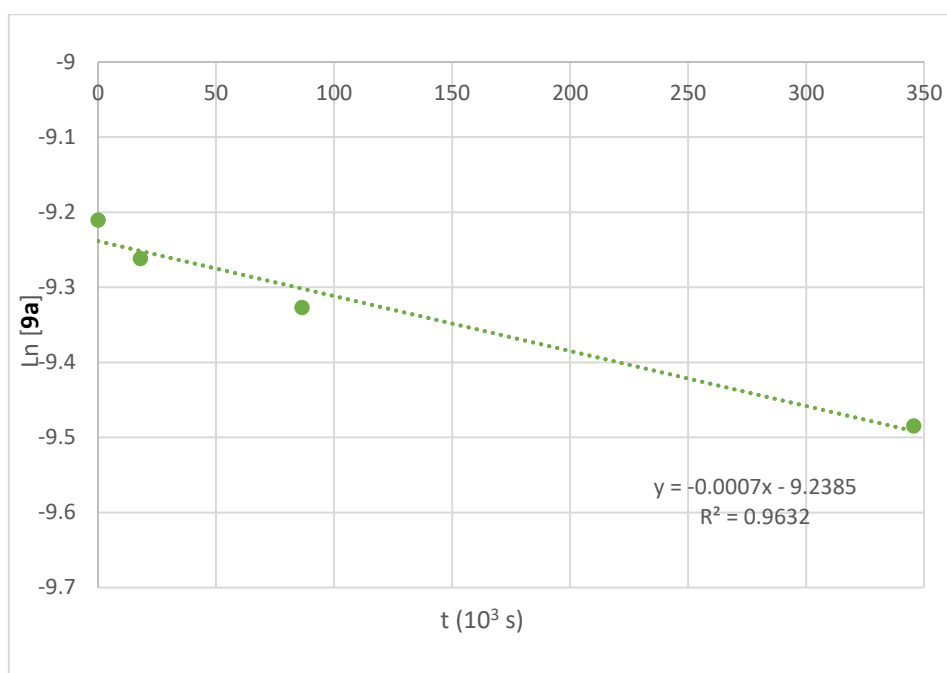
➤ **β-galactosidase**



Note that the reaction kinetics were very fast with β-galactosidase in solution, so an amount of U_{enzyme} 50 times lower compared to other enzymes was used.

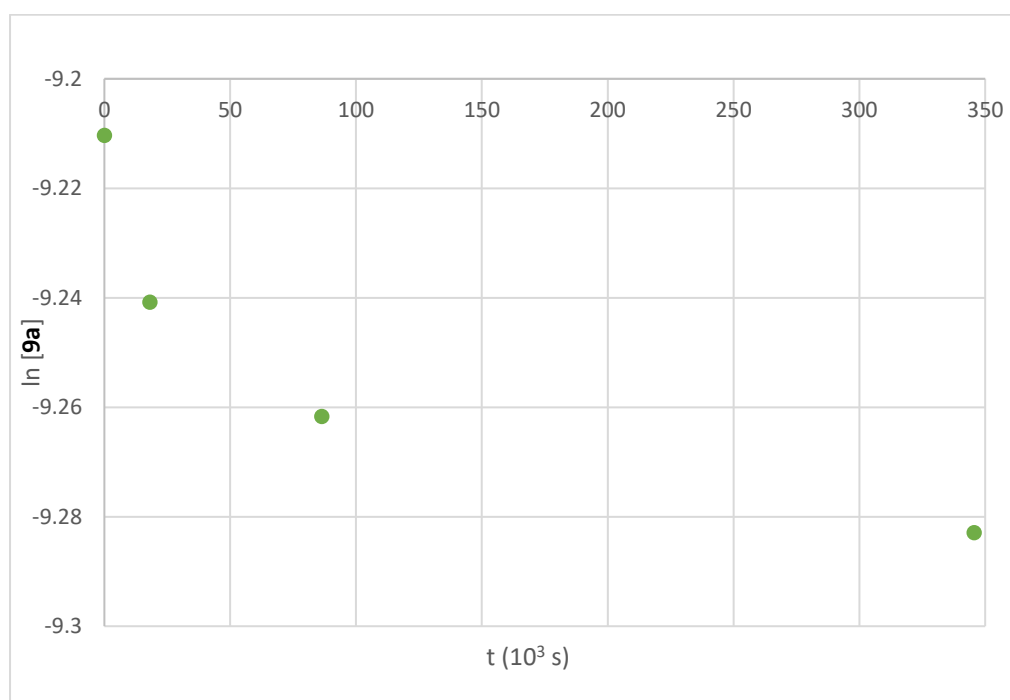


CPG1000-supported substrate (0.4 U / nmol)



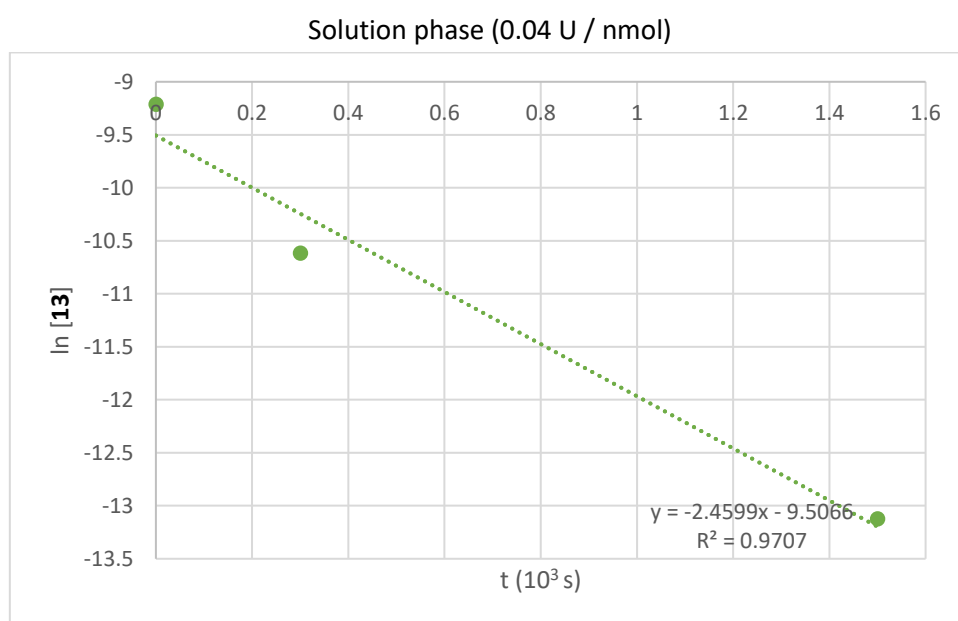
$$k_{\text{obs}} = 0.0007 \cdot 10^{-3} \text{ s}^{-1}$$

PEGA1900-supported substrate (0.4 U / nmol)

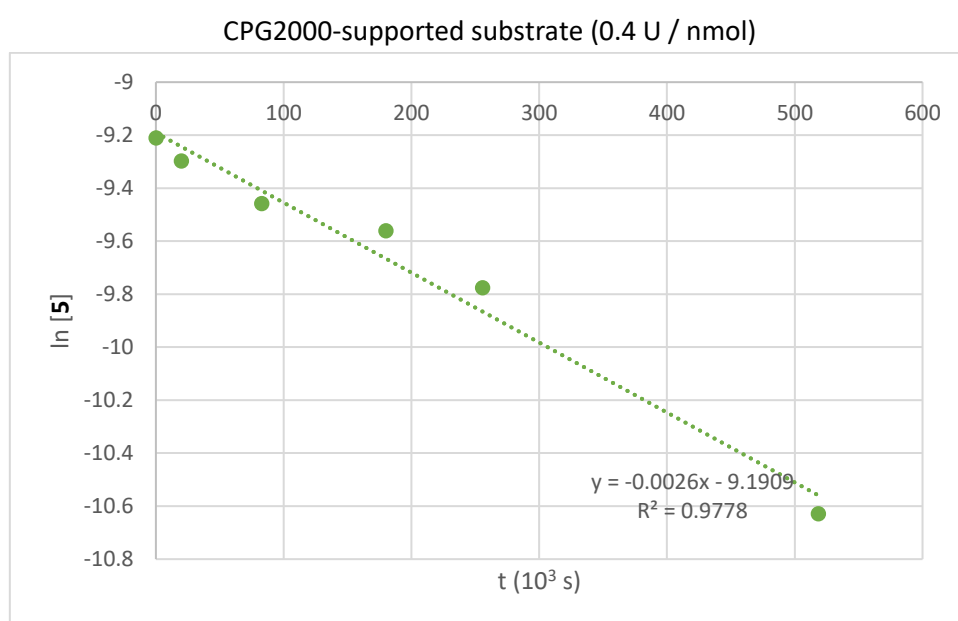


The values cannot be fitted to a first-order kinetics. The k_{obs} decreases upon time, representative of the plateau observed in the green curve of supplementary figure S8, pS15.

➤ **Shrimp phosphatase**

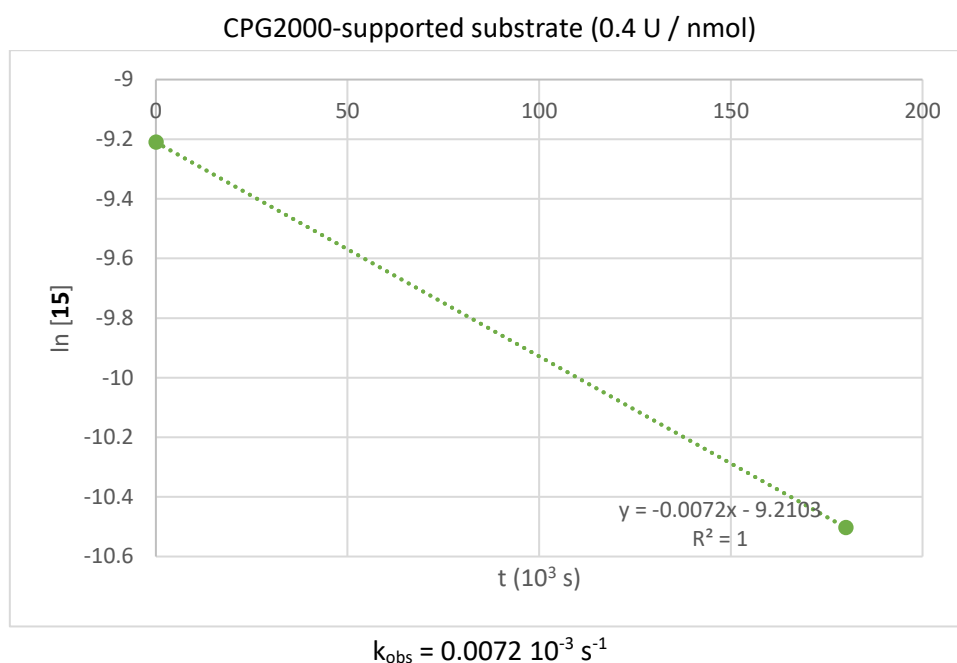
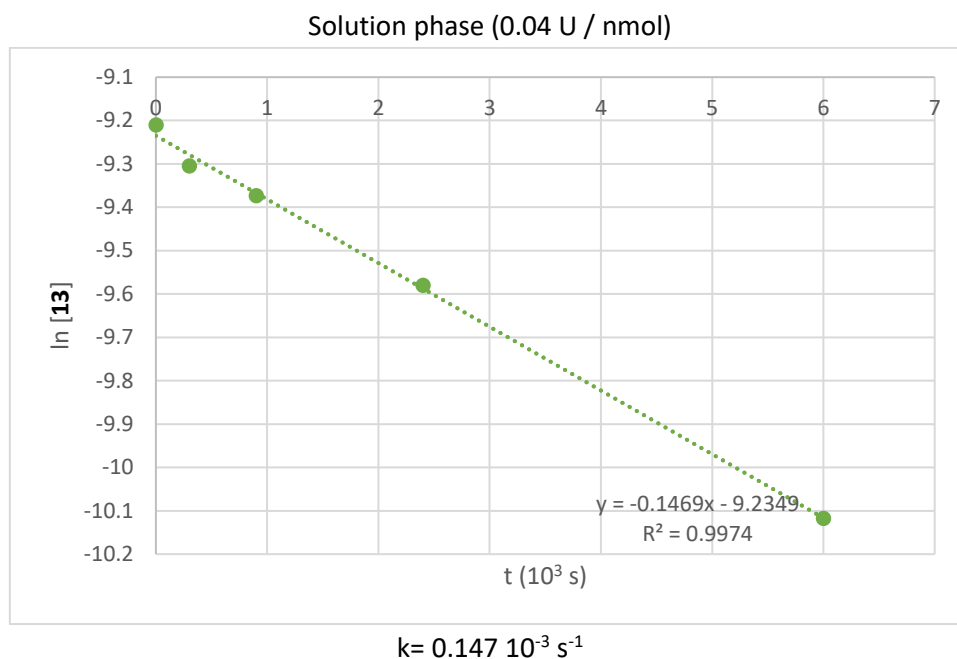


$$k_{\text{obs}} = 2.16 \cdot 10^{-3} \text{ s}^{-1}$$



$$k_{\text{obs}} = 0.0026 \cdot 10^{-3} \text{ s}^{-1}$$

SI-p S63

➤ **Lambda phosphatase**

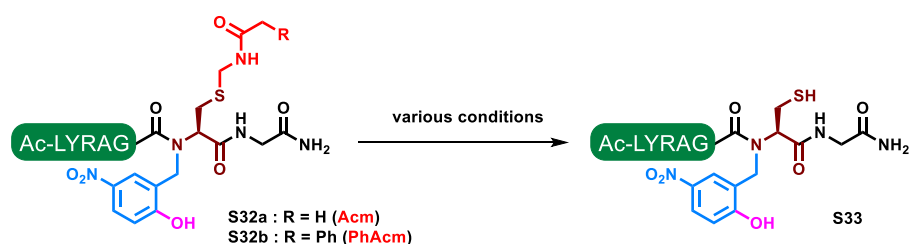
Note that in this case, only one value has been used to estimate the kinetics constant: 72.6% linker cleavage after 50 h reaction (average value from two distinct experiments: 73.3% and 71.8%, respectively). This is however sufficient to estimate the observed kinetic constant. A $\pm 10\%$ error in the measured value would give $k_{\text{obs}} = 0.0055$ and $0.0097 \cdot 10^{-3} \text{ s}^{-1}$, respectively, corresponding to $k_{\text{solution}}/k_{\text{support}} = 267$ and 152 , instead of 204 , thus quite negligible as compared as the orders of magnitude of difference observed with shrimp phosphatase and β -galactosidase ($k_{\text{solution}}/k_{\text{support}} = 61\,702$ and $8\,308$, respectively). A massive $\pm 20\%$ error would correspond to 359 and 102 ratios.

8- Optimization of a strategy to temporarily mask the reactivity of *N*-Hnb-Cys crypto-thioesters in order to enable successive N-to-C ligations.

In order to enable the SPCL elongation through N-to-C successive NCLs, the thioester reactivity of internal segments should be kept in a dormant state during a first NCL involving their N-terminal cysteine residue. To achieve this goal, we chose to protect the cysteine of an *N*-Hnb-Cys device with a NCL-resistant protective group that could be easily deprotected in order to switch on the crypto-thioester properties on demand.

Two types of protecting groups have been employed in related works using other *N*-S-shift-based crypto-thioesters: the transition metal-labile acetamidomethyl (Acm) group [5], and the photolabile 2-nitrobenzyl and 6-nitroveratryl groups [6]. Considering that photocleavage is not ideal for solid-phase reactions due to absorption of the support, we selected Acm as a worthy candidate. In addition, we explored the enzyme-labile phenylacetamidomethyl group (PhAcm) that can be cleaved by penicillin-G acylase (PGA).

In order to identify the best suited protecting group for our needs, we synthesized two model masked crypto-thioesters (**S32a** and **S32b**) and systematically screened different conditions reported in the literature for their cleavage.



8-1- Synthesis of model crypto-thioesters **S32a** and **S32b**.

Model peptides **S32a** and **S32b** was conducted using protocol PS1 starting from Rink Tentagel R (0.21 mmol/g).

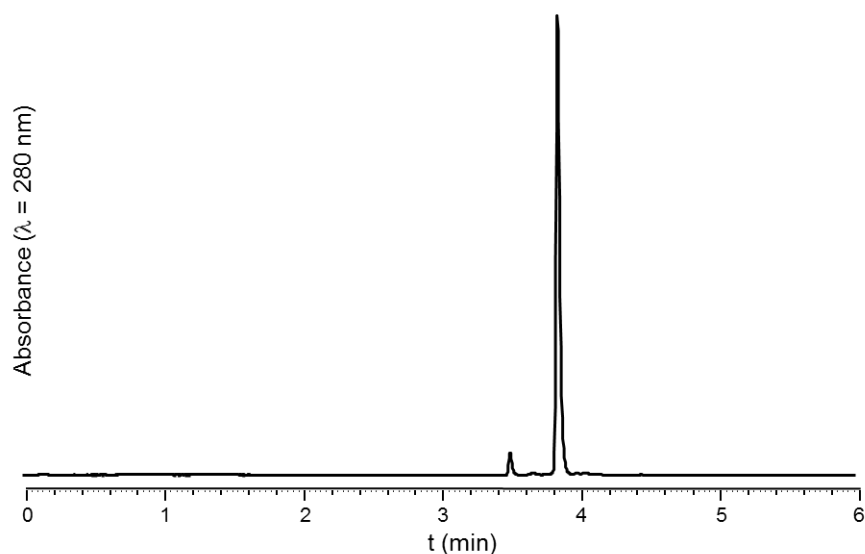
Hnb was introduced using the previously described [7] non optimized protocol: 25 μ mol of H-Cys(Acm)-Gly-Rink-Tentagel R or H-Cys(PhAcm)-Gly-Rink-Tentagel R was washed with 1:1 DMF/MeOH (4 x 3 mL, 30 s) then treated with a 9:9:2 DMF/MeOH/AcOH mixture (3 mL, 5 min). The resin was washed with 1:1 DMF/MeOH (3 x 3 mL, 30 s), 2-hydroxy-5-nitrobenzaldehyde (42 mg, 10 equiv.) in 2 mL 1:1 DMF/MeOH (125 mM aldehyde concentration) was then added, and the reactor was stirred for 1 h. The reactor was drained and the resin was washed with 1:1 DMF/MeOH (3 x 3 mL, 5 s). Without delay, a fresh solution of sodium cyanoborohydride (32 mg, 20 equiv.) in 3 mL 9:9:2 DMF/MeOH/AcOH (250 mM, NaBH₃CN concentration) was added and the reactor stirred for 1 h. The reactor was drained and the resin was washed with 1:1 DMF/MeOH (4 x 3 mL, 30 s), NMP (3 x 3 mL, 30 s), 20% piperidine in NMP (3 x 3 mL, 3 min), NMP (3 x 3 mL, 30 s), dichloromethane (3 x 5 mL, 30 s) and NMP (2 x 3 mL, 30 s).

Cleavage was performed following protocol PS4, and resulting crude peptides **S32a** and **S32b** were purified by semi-preparative HPLC.

- Ac-LYRAA-(Hnb)C(Acm)-G-NH₂ **S32a**

ESI-MS (*m/z*): [M] calcd. for C₄₄H₆₅N₁₃O₁₃S: 1015.5, found: 1015.6

HPLC analysis: *t_R* = 3.98 min (Chromolith, gradient: 5-50% B over 5 min)

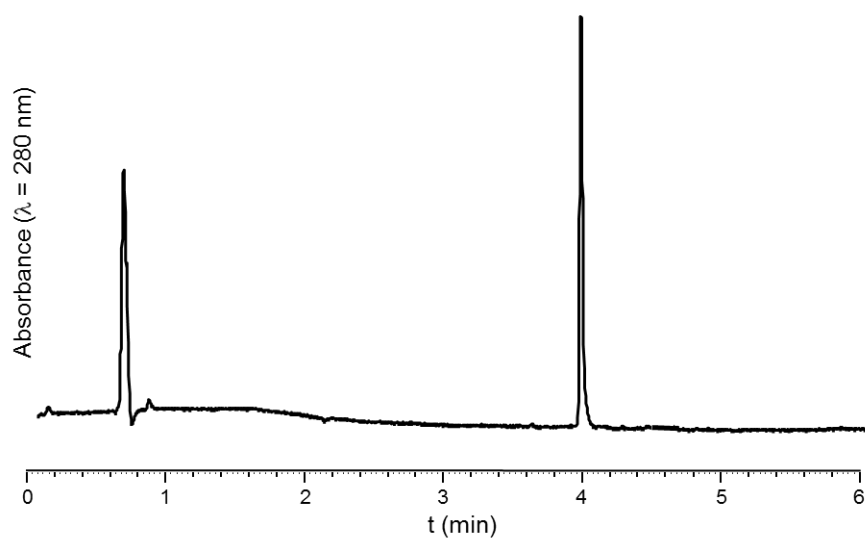


Supplementary figure S48: HPLC trace of purified **S32a**

- Ac-LYRAA-(Hnb)C(PhAcm)-G-NH₂ **S32b**

ESI-MS (*m/z*): [M] calcd. for C₅₀H₆₉N₁₃O₁₃S: 1091.5, found: 1091.4

HPLC analysis: *t_R* = 4.03 min (Chromolith, gradient: 5-50% B over 5 min)



Supplementary figure S49: HPLC trace of purified **S32b**

8-2- Screening of different conditions for the deprotection of the Acm group in S32a.

We systematically screened conditions described in the literature for the deprotection of Acm groups in peptides, using either transition metals salts (Ag^I , Hg^{II} , Pd^{II}) or 2,2'-dithiodipyridine (DTP). Note we applied different protocols directly adapted from the literature rather than trying to homogenize solvents, additives, and number of equivalents of the reagents.

Reactions were monitored by analytical HPLC. Results are summarized in table S6.

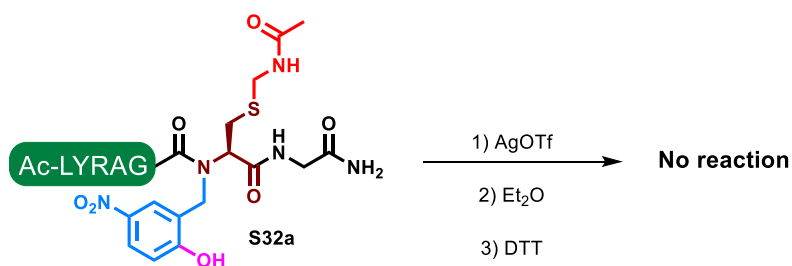
Table S6: Conditions used for the removal of Acm protecting groups in masked crypto-thioester **S32a**.

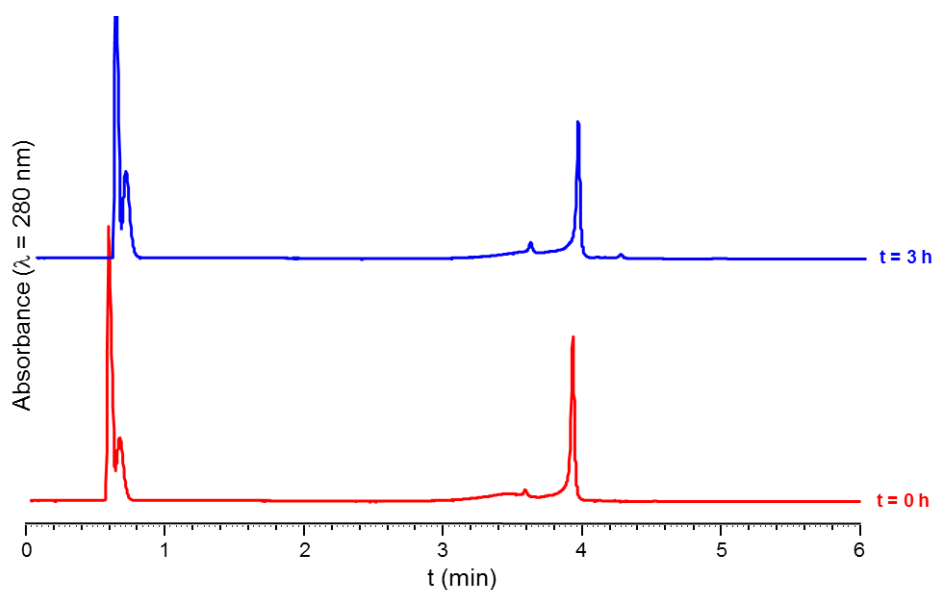
Entry	Conditions used	Results
1	10 equiv. AgOTf, 10 equiv. anisole, TFA, 3 h, 0 °C then 50 equiv. DTT	No reaction (< 0.5%)
2	40 equiv. AgBF ₄ , 98:2 TFA/anisole, 12 h, then 80 equiv. DTT	Incomplete reaction (~5%)
3	10 equiv. Hg(OAc) ₂ , 20 equiv. β-mercaptoethanol, 5 M Gu.HCl, 9:1 H ₂ O/AcOH	Nearly quantitative conversion of the starting material but numerous unidentified side product
4	30 equiv. 2,2'-dithiodipyridine, 98:2 TFA/thioanisole, then 50 equiv. TCEP	Incomplete reaction (~25%)
5	15 equiv. PdCl ₂ , 6M Gu.HCl, 0.2 M phosphate buffer pH 7,	Complete and clean reaction

➤ Use of silver trifluoromethanesulfonate (AgOTf)

The use of AgOTf to remove Acm cysteine protecting group following a protocol adapted from reference [8] was unsuccessful: only trace amounts (< 0.5%) of the expected product **S33** were observed after 3 h.

Conditions: Peptide **S32a** (0.25 mM final concentration) was dissolved in TFA then AgOTf (40 equiv.) and anisole (10 equiv.) were added and the reaction was stirred for 1 h at 0 °C. Then 10 volumes Et₂O were added, the resulting precipitate was recovered by centrifugation, washed twice with Et₂O then treated with DTT (80 equiv.) in 50% AcOH for 2 h at room temperature.

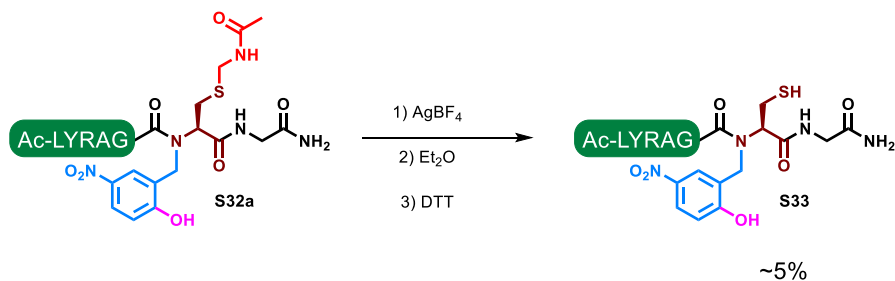




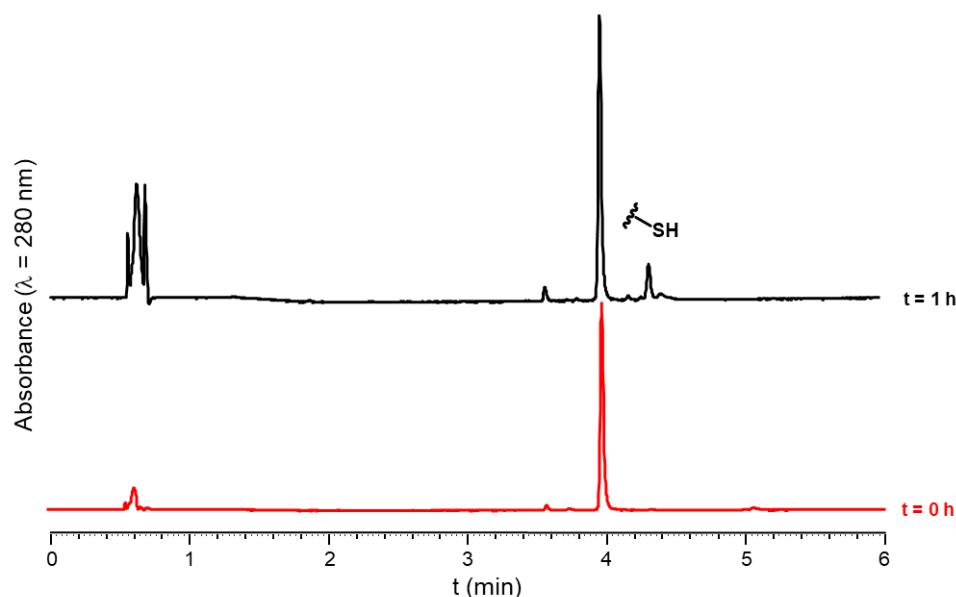
Supplementary figure S50: HPLC traces (Chromolith, gradient: 5-50% B over 5 min) corresponding to the reaction of **S32a** with AgOTf (table S6, entry 1)

➤ Use of silver tetrafluoroborate (AgBF_4)

The use of AgBF_4 to remove AcM cysteine protecting group following a protocol adapted from reference [8] gave slightly better results than AgOTf however the reaction was incomplete and only small amounts ($\sim 5\%$) of the expected product **S33** were observed after 12 h.



Conditions: Peptide **S32a** (0.25 mM final concentration) was dissolved in 98:2 TFA/anisole (10 equiv. anisole) then AgBF_4 (10 equiv.) was added and the reaction stirred for 1 h at 0 °C. Then 10 volumes Et_2O were added, the resulting precipitate was recovered by centrifugation, washed twice with Et_2O then treated with DTT (80 equiv.) in 50% AcOH for 2 h at room temperature.

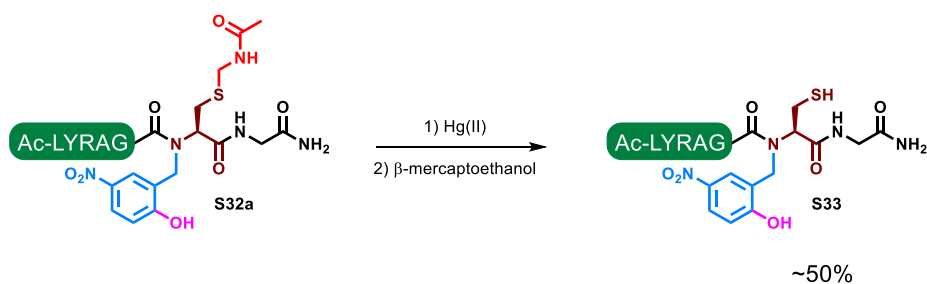


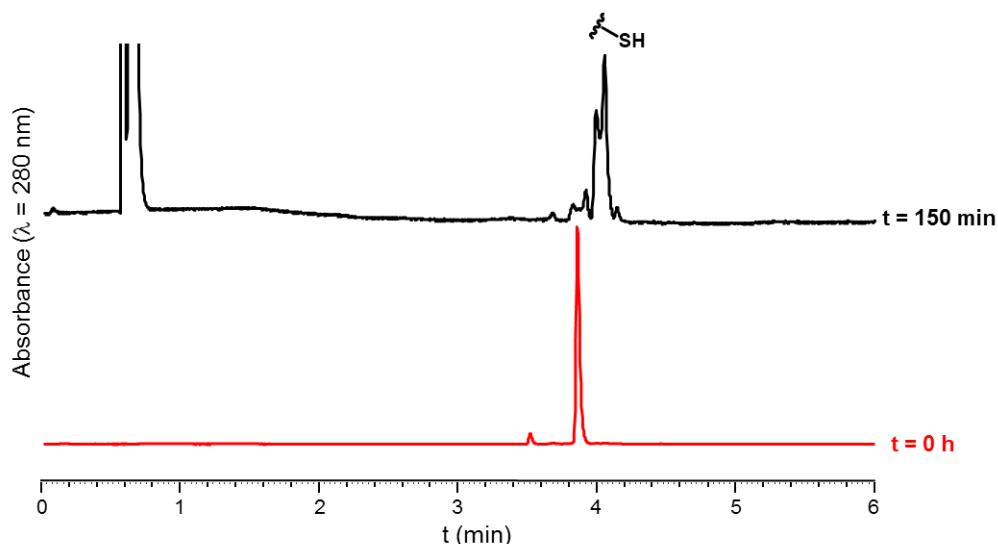
Supplementary figure S51: HPLC traces (Chromolith, gradient: 5-50% B over 5 min) corresponding to the reaction of **S32a** with AgBF_4 (table S6, entry 2)

➤ Use of mercury (II) acetate ($\text{Hg}(\text{OAc})_2$)

Acm was successfully removed using $\text{Hg}(\text{II})$ acetate following a protocol adapted from reference [9] nevertheless, the reaction was not clean and numerous unidentified side products were observed. Area of the peak corresponding to the expected product **S33** only accounted for ~50% of the total area of the different peaks observed at $\lambda = 280 \text{ nm}$.

Conditions: Peptide **S32a** (0.25 mM final concentration) was dissolved in 9:1 6 M Gu.HCl then mercury acetate (10 equiv.) in AcOH (10% vol AcOH final concentration) were added and the reaction stirred for 90 min at room temperature. Next, β -mercaptoethanol (20 equiv.) was added and the reaction stirred for one additional hour at 50°C .



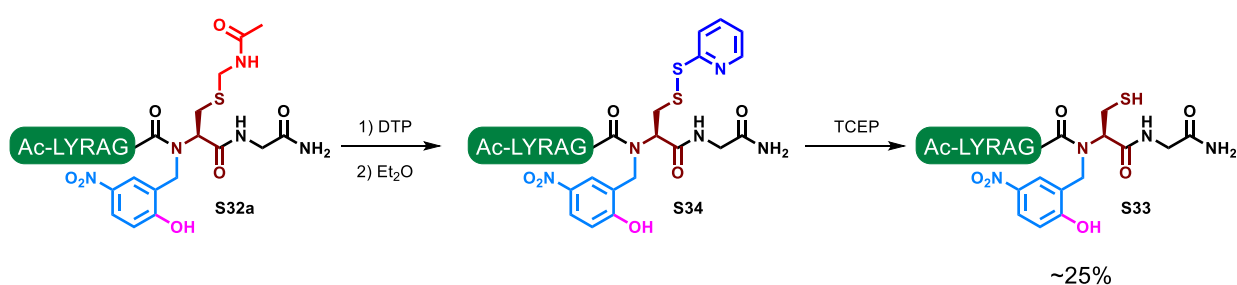


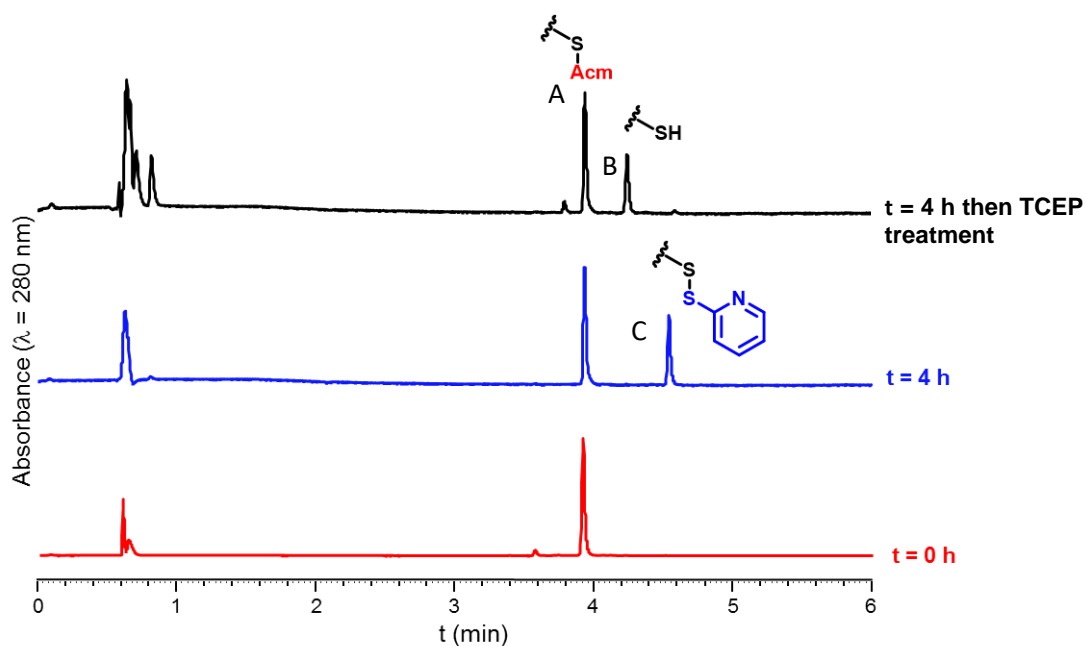
Supplementary figure S52: HPLC traces (Chromolith, gradient: 5-50% B over 5 min) corresponding to the reaction of **S32a** with Hg(II) (table S6, entry 3)

➤ Use of 2,2'-dithiodipyridine (DTP)

The use of DTP following a protocol adapted from reference [10] gave better results: the AcM group was cleanly transformed into a 2-thiopyridyl disulfide that was further reduced into the free thiol by a TCEP treatment. However, the reaction was not complete.

Conditions: Peptide **S32a** (0.25 mM final concentration) was dissolved in 98:2 TFA/thioanisole then DTP (30 equiv.) was added and the reaction stirred for 4 h at room temperature. Next, the peptide was precipitated by dilution into 10 volumes of an ice-cold diethyl ether/petroleum ether 1:1 mixture, recovered by centrifugation and washed twice with diethyl ether. Then the crude product was treated with TCEP (50 equiv.) in 200 mM phosphate buffer, pH 7.





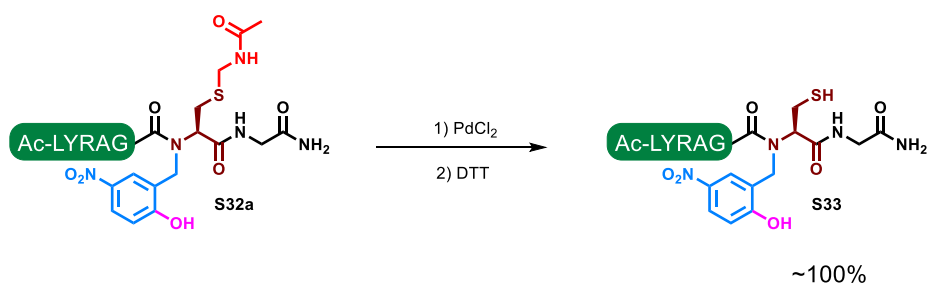
Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (3.92)	1015.5	1015.6	S32a
B (4.23)	944.4	944.4	S33
C (4.57)	1053.4	1053.4	S34

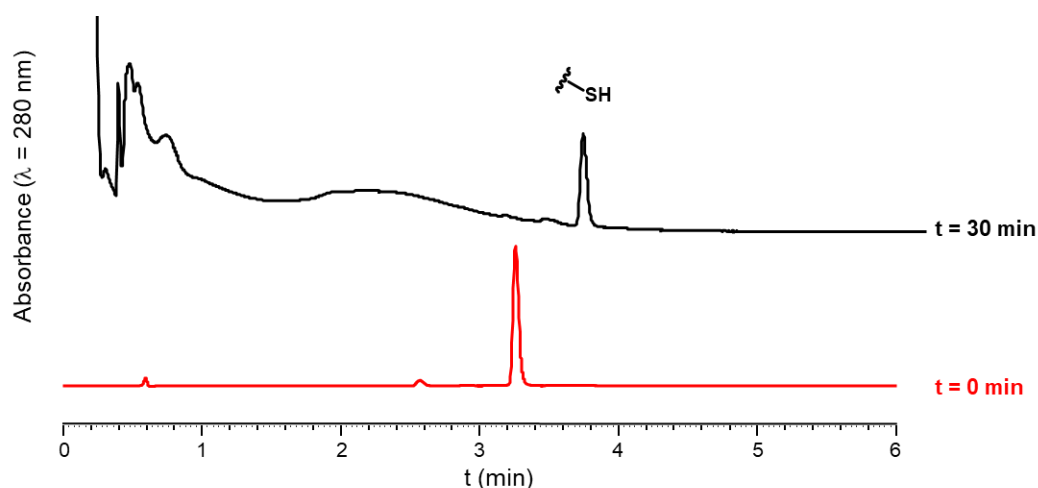
Supplementary figure S53: HPLC traces (Chromolith, gradient: 5-50% B over 5 min) and MS analysis corresponding to the reaction of **S32a** with DTP (table S6, entry 4)

➤ Use of palladium(II) chloride (PdCl_2)

When PdCl_2 was used following a protocol developed by Brik and co-workers [11], a quantitative deprotection was obtained, and no significant by-products were observed.

Conditions: Peptide **S32a** was dissolved in 6 M Gu.HCl (0.25 mM peptide concentration). Next, PdCl_2 (15 equiv.) was added and the reaction stirred for 30 min. Then DTT (50 equiv.) was added, leading to the precipitation of the palladium salts. The supernatant was analyzed by HPLC.



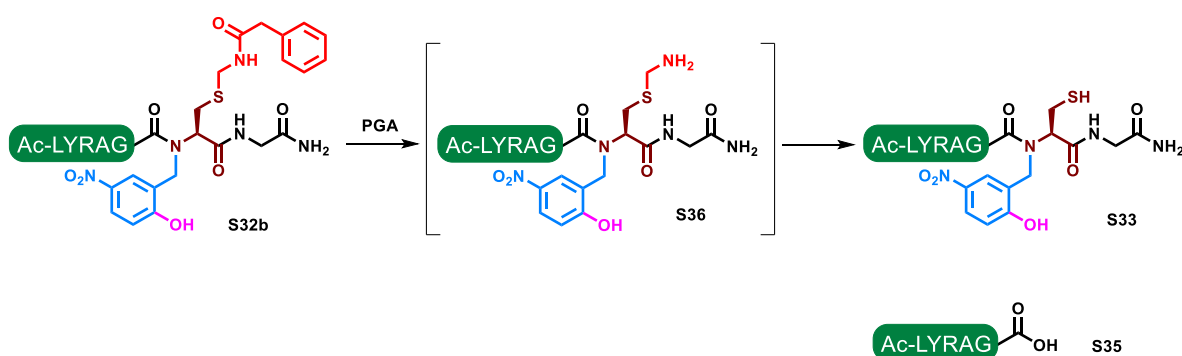


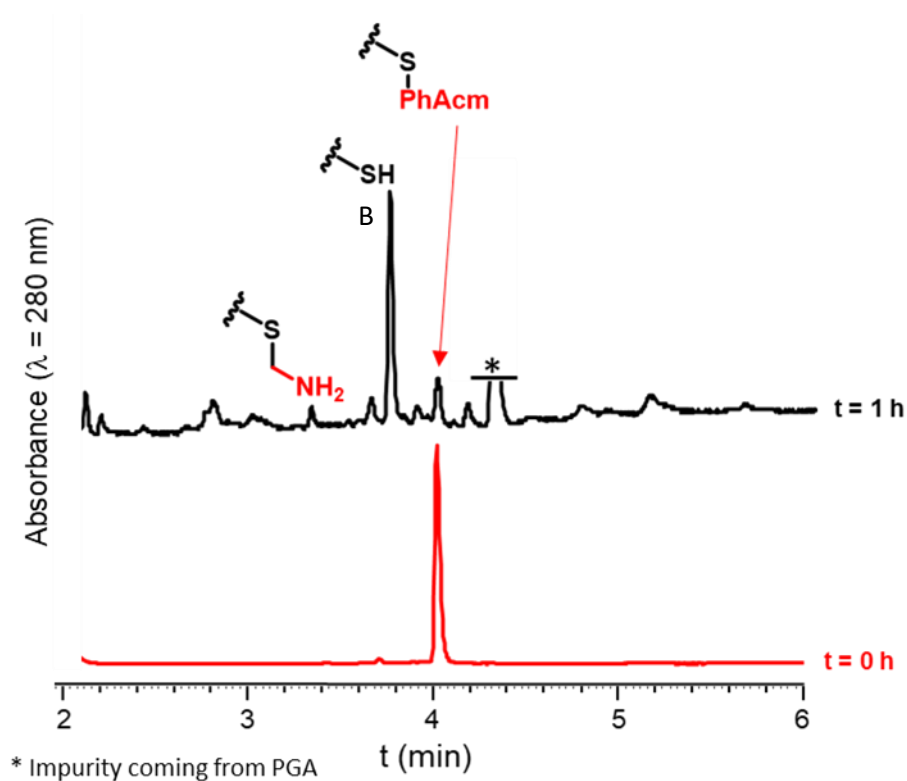
Supplementary figure S54: HPLC traces (Chromolith, gradient: 15-50% B over 5 min) corresponding to the reaction of **S32a** with PdCl_2 (table S6, entry 4). Note that the baseline perturbation is caused by remains of palladium salts in the supernatant

8-3- PGA-mediated deprotection of the PhAcm group in **S32b**.

The use of penicillin G amidase to remove PhAcm cysteine protecting group was rather effective, as the reaction was nearly complete after 1 h, but was not very clean. Moreover, extended reaction times led to the degradation of the desired product and formation of the hydrolysed crypto-thioester (Ac-LYRAAG-OH, **S35**).

Conditions: Peptide (0.25 mM final concentration) **S32b** was incubated with PGA (5 U/1 μmol peptide) in 20 mM Na_2HPO_4 - NaH_2PO_4 buffer, 50 mM TCEP, pH 7.8 at 37 °C for 1 h [12].





Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (4.03)	1091.5	1091.4	S32b
B (3.80)	944.4	944.4	S33
C (3.38)	973.4	973.4	S36
D (2.81)	634.7	634.3	S35

Supplementary figure S55: HPLC traces and MS analysis (Chromolith, gradient: 5-50% B over 5 min) of the treatment of **S32b** with PGA

9- Synthesis of the different segments needed for SPCL

9-1- N-terminal segment 16

Sequence:

H-¹AP^{DT}RPAGSTAP^{PA}HGVTSAP^{DT}RPAGSTAP^{PA}HGVTSAP^{DT}RPAGSTAP^{PA}HGVTS-⁶⁰S-(Hnb)C(StBu)-G-NH₂

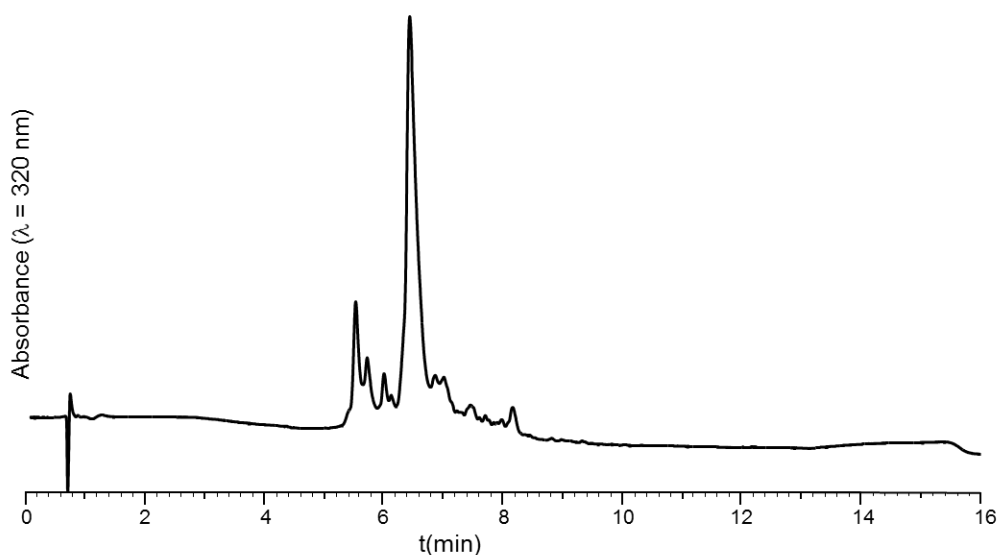
The peptide was synthesized using protocol PS1 with double couplings for consecutive proline residues (highlighted in blue in the above sequence). Pseudoproline dipeptides were used at ^{DT} positions in order to prevent aspartimide formation during SPPS elongation: Fmoc-Asp(OtBu)-Thr($\psi^{\text{Me,Me}}$ pro)-OH building block was incorporated using slightly modified coupling conditions (3.5 equiv. dipeptide, 3.4 equiv. HATU, 10 equiv. DIEA in NMP for 2 h). Hnb was introduced following protocol PS2. Analytical scale cleavage was performed following protocol PS4.

- Compound **S37**

Elongation yield: 75%. Determined by the ratio between the quantity of fluorenylpiperidine released during final Fmoc deprotection (Protocol PS6) and the quantity released during the Fmoc deprotection of the C-terminal Gly residue (UV titration at 301 nm, $\epsilon = 7800 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$).

ESI-MS (m/z): [M] calcd. for C₂₅₆H₃₉₉N₇₉O₈₆S₂: 6023.5, found: 6022.6 (average mass, deconvoluted)

HPLC analysis: $t_R = 6.56 \text{ min}$ (Aeris Widespore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C)



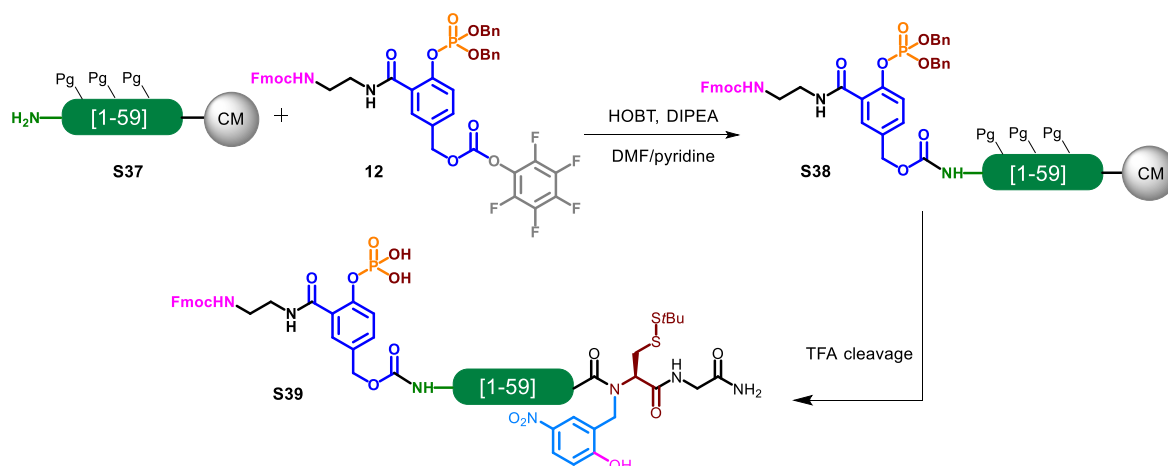
Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (5.6)	5994.5	5994.2	Ac-[2-60]-(Hnb)C(StBu)-G-NH ₂
B (5.6)	5986.4	5985.8	β -elimination on Cys(StBu) + piperidine addition on resulting dehydroalanine (see p S76)
C (6.58)	6023.5	6022.6	S37

Supplementary figure S56: HPLC trace and MS analysis of crude **S37** (average masses, deconvoluted)

SI-p S74

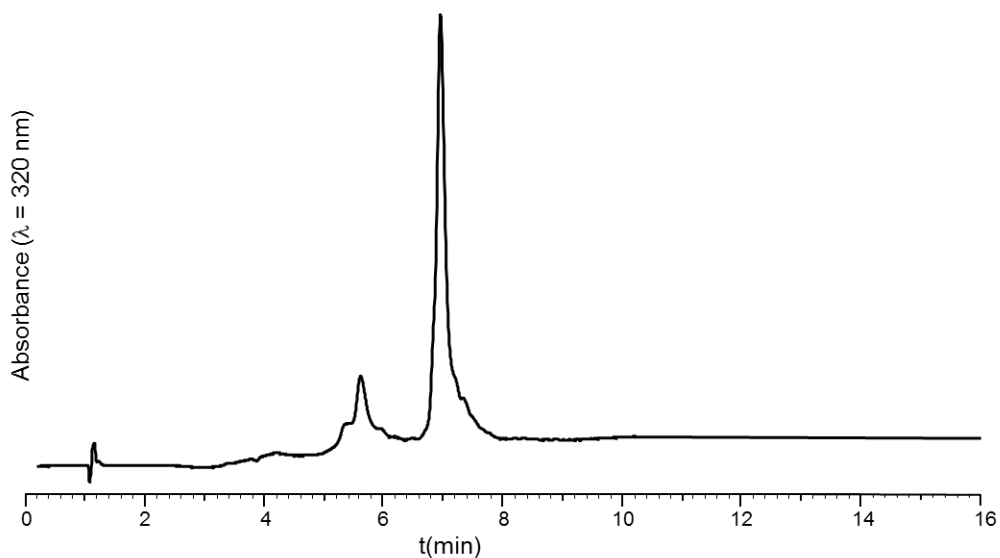
➤ Introduction of linker on the first segment

The same protocol as for the linker introduction on model peptide was used (see p S48-S49).

• Compound **S39**

ESI-MS (*m/z*): [M] calcd. for C₂₆₇H₄₁₂N₈₁O₉₃PS₂: 6339.6, found: 6338.8 (average mass, deconvoluted)

HPLC analysis: *t_R* = 6.98 min (Aeris Widepore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C)

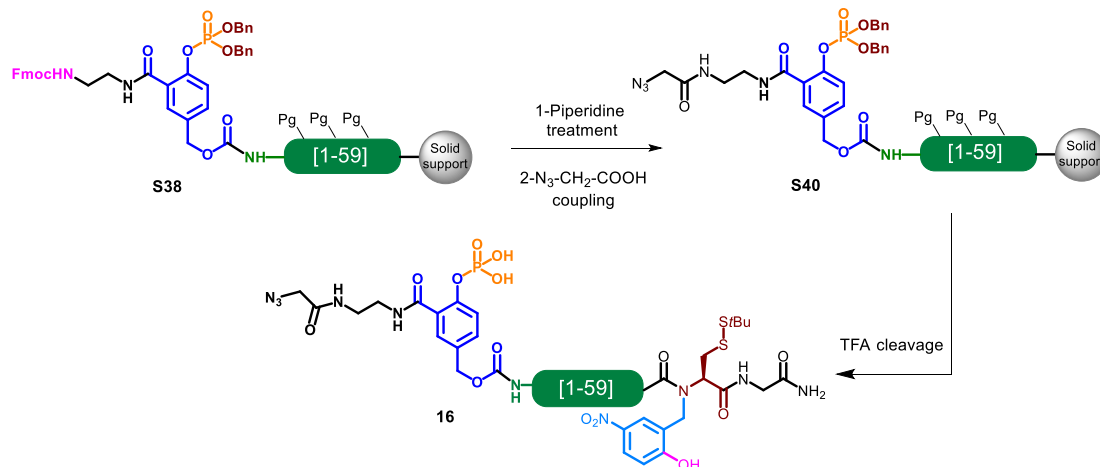


Peak number (<i>t_R</i> (min))	[M] (<i>m/z</i>) calcd.	[M] (<i>m/z</i>) found	Attributed to
A (5.63)	5994.5	5994.2	Ac-[2-60]-(Hnb)C(StBu)-G-NH ₂
B (6.98)	6339.6	6338.8	S39

Supplementary figure S57: HPLC trace and MS analysis of crude **S39** (average masses, deconvoluted)

➤ **Coupling of azido-acetic acid**

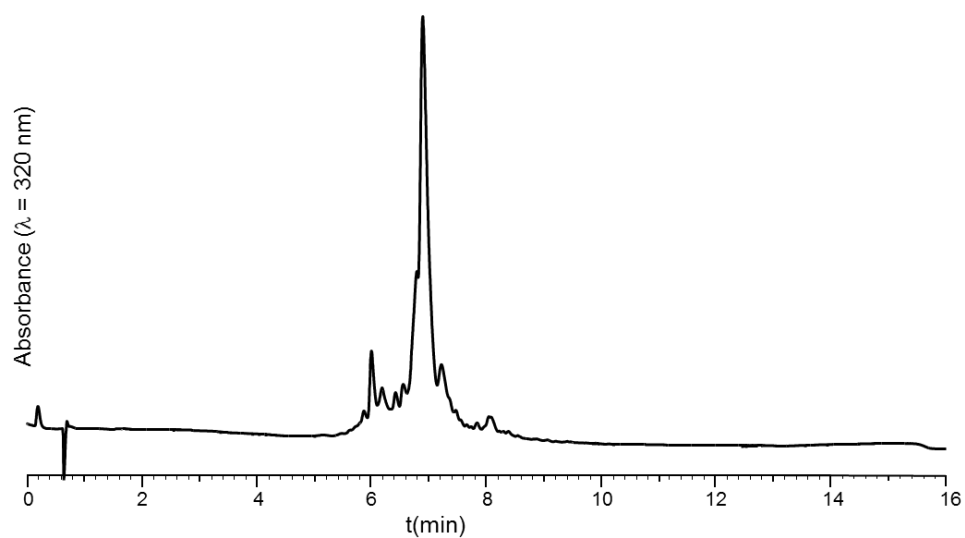
The same protocol as for coupling of azido-acetic acid on the model peptide bearing the linker was used (see p S49-S50).

• **Compound 16**

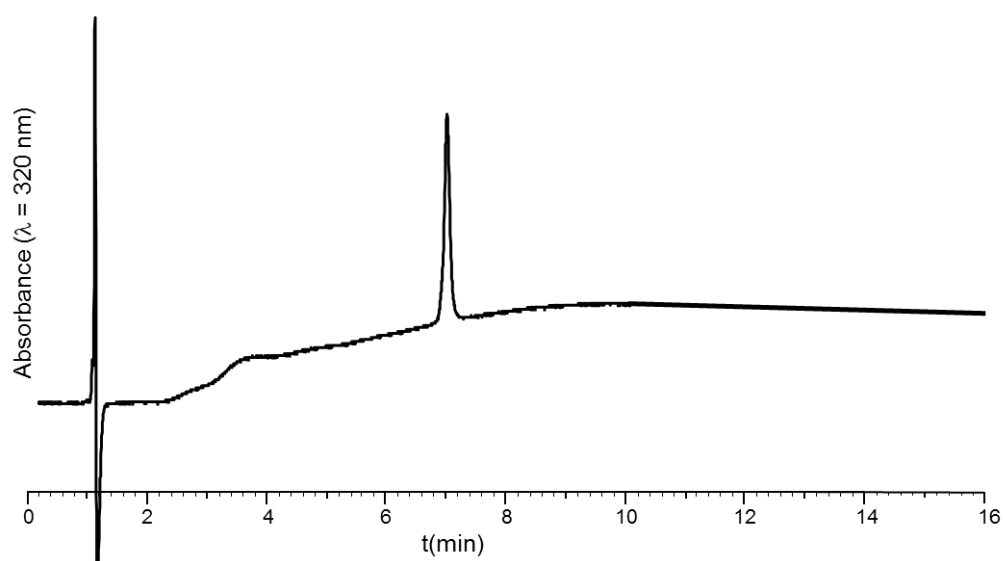
ESI-MS (m/z): [M] calcd. for $\text{C}_{269}\text{H}_{414}\text{N}_{84}\text{O}_{94}\text{PS}_2$: 6423.8, found: 6422.0 (average mass, deconvoluted)

HPLC analysis: t_R = 3.54 min (Chromolith, gradient: 5-50% B over 5 min)

HPLC purification: [5 mg/mL] Nucleosil C18, gradient: 18-30% B over 20 min affording a white solid after lyophilisation (22% yield).



Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (6.02)	-	6079.5	not attributed
B (6.94)	6519.8	6519.0	TFA ester on a Ser or Thr residue
C (7.04)	6423.8	6422.0	16
D (7.20)	-	6105.7	not attributed



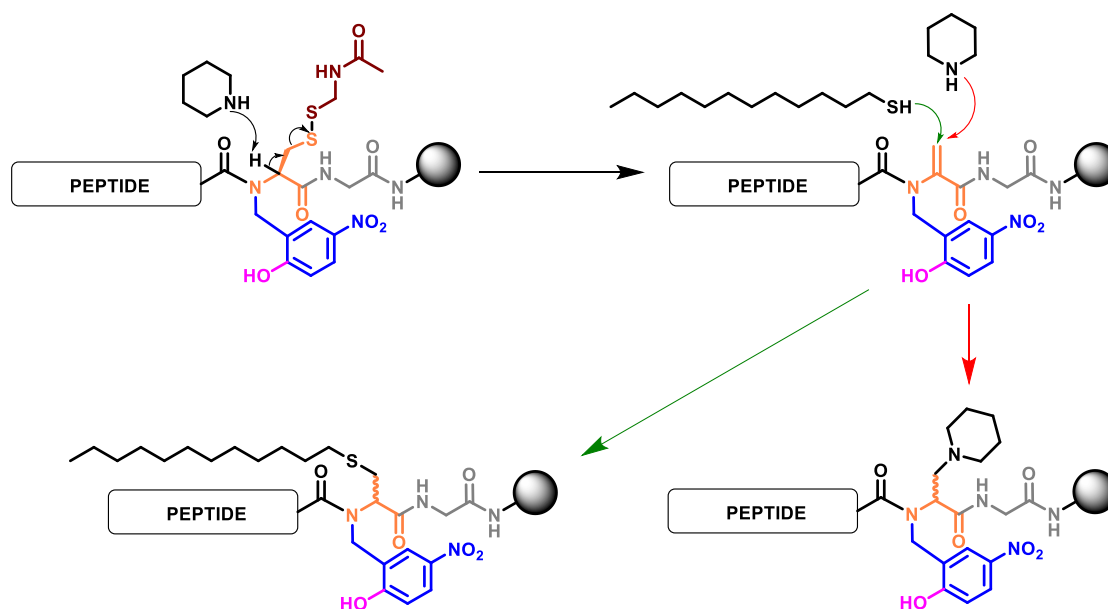
Supplementary figure S58: HPLC traces and MS analysis of crude (A) and purified (B) compound **16** (average masses, deconvoluted)

9-2- Compound 18: Median segment

Sequence: H-⁶¹CPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGV¹⁰⁰S-(Hnb)C(Acm)-G-NH₂

The peptide was synthesized using protocol PS1 with double couplings for consecutive proline residues (highlighted in blue in the above sequence). Pseudoproline dipeptides were used at DT positions in order to prevent aspartimide formation during SPPS elongation: Fmoc-Asp(OtBu)-Thr($\psi^{\text{Me,Me}}$ pro)-OH building block was incorporated using slightly modified coupling conditions (3.5 equiv. dipeptide, 3.4 equiv. HATU, 10 equiv. DIEA in NMP for 2 h). Hnb was introduced following protocol PS2. Cleavage was performed following protocol PS4.

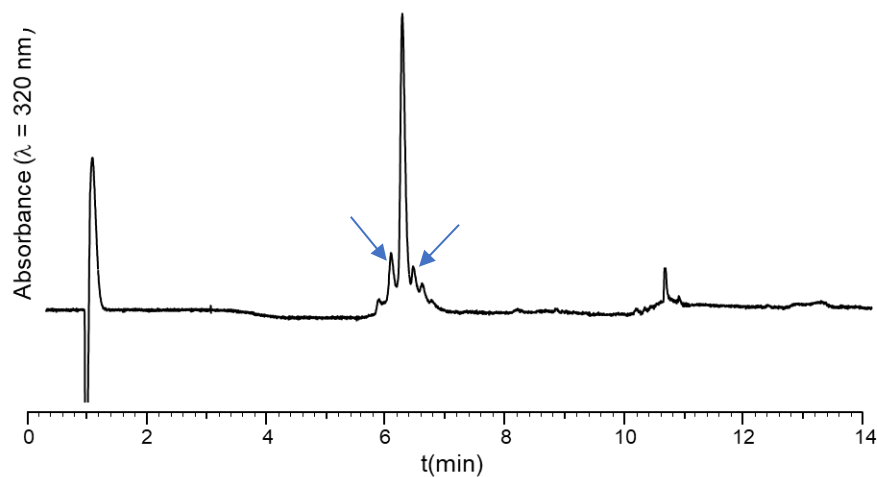
In first attempt using classical coupling and deprotection conditions, we noticed the formation of a co-product (~10%) which mass is consistent with the formation of a 3-(1-piperidinyl)alanine at the Cys(Acm) position, which hampered the purification of the target segment. The formation of this side-product likely occurs through base-induced β -elimination followed by Michael-type addition of piperidine during the repeated Fmoc cleavage steps. Note that we already observed this side reaction, which is sequence- and temperature-dependant (see reference [13] and synthesis of **S37**, p S72), in the case of *N*-Hnb-Cys(*StBu*) cryptothioesters. Acm-protected cysteines are known to be generally more sensitive than Trt- and *StBu*-protected ones towards this side-reaction [14]. In order to limit the amount of the piperidine-alanyl side-product, dodecanethiol (5% v/v) was added to the piperidine treatment cocktail to induce a competitive nucleophilic attack leading to a hydrophobic dodecanesulfanyl derivative easier to separate from the desired product (fig. S59).



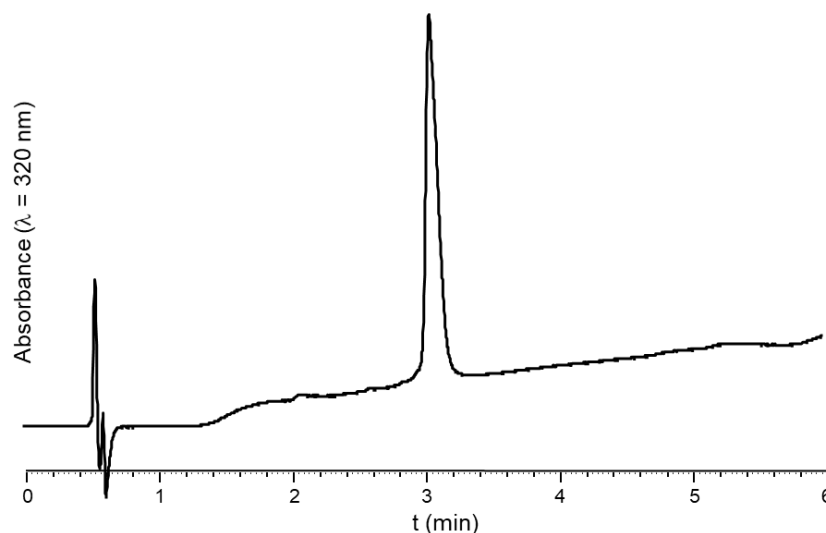
Supplementary figure S59: putative mechanism of β -elimination followed by nucleophilic additions
Elongation yield: 81%. Determined by the ratio between the quantity of fluorenylpiperidine released during final Fmoc deprotection (Protocol PS6 - Procedure for selective Hnb ester cleavage to allow UV titration of Fmoc deprotection) and the quantity released during the Fmoc deprotection of the C-terminal Gly residue (UV titration at 301 nm, $\epsilon = 7800 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$).

ESI-MS (m/z): [M] calcd. for $C_{175}H_{271}N_{55}O_{60}S_2$: 4169.5, found: 4169.1 (average mass, deconvoluted)

HPLC purification: [5 mg/mL] Nucleosil C18, gradient: 10-15% B over 20 min, affording a white solid after lyophilisation (16% yield).



Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (6.11)	-	4418.8	not attributed
B (6.23)	4169.5	4169.1	18
C (6.41)	4148.9	4149.1	β -elimination on Cys(Acm) + piperidine addition on resulting dehydroalanine
D (10.65)	4266.5	4266.0	β -elimination on Cys(Acm) + dodecanethiol addition on resulting dehydroalanine



Supplementary figure S60: HPLC traces and MS analysis (average masses, deconvoluted) of crude (A) and purified (B) compound **18** (Aeris Widepore XB-C18 2, gradient: 3-50% B' over 15 min, 60°C and Chromolith, gradient: 5-50% B over 5 min, respectively)

9-3- Compound 21: C-terminal segment

Sequence:

[illegible]

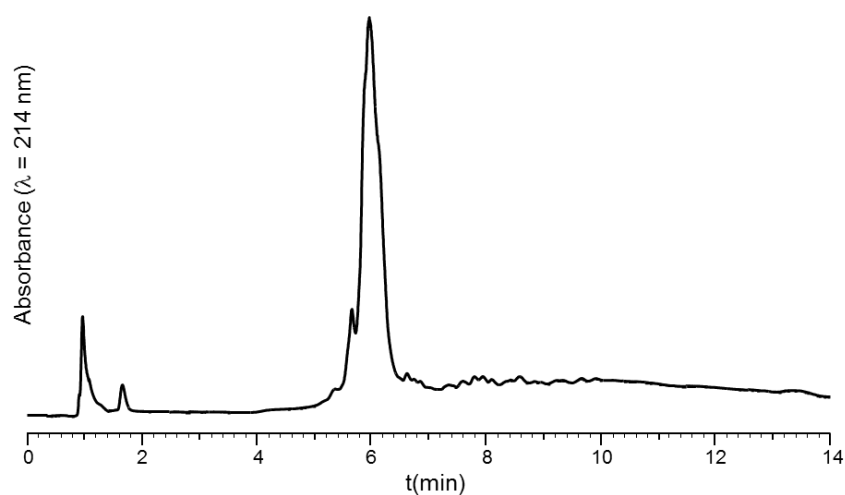
The peptide was synthesized using protocol PS1 with double couplings for consecutive proline residues (in blue in the sequence). Pseudoproline dipeptides were used at **DT** positions in order to prevent aspartimide formation during SPPS elongation: Fmoc-Asp(OtBu)-Thr($\psi^{\text{Me,Me}}$ pro)-OH building block was incorporated (3.5 equiv. dipeptide, 3.4 equiv. HATU, 10 equiv. DIEA in NMP for 2 h). Cleavage was performed following protocol PS4.

Elongation yield: 93%. Determined by the ratio between the quantity of fluorenylpiperidine released during final Fmoc deprotection and the quantity released during the Fmoc deprotection of the C-terminal Ser residue (UV titration at 301 nm, $\epsilon = 7800 \text{ L.mol}^{-1}.\text{cm}^{-1}$).

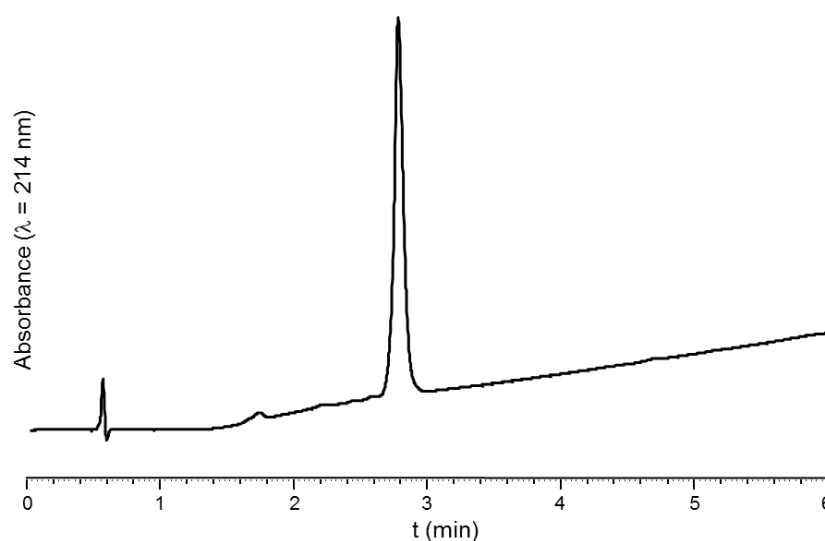
ESI-MS (*m/z*): [M] calcd. for C₂₄₀H₃₇₇N₇₅O₈₂S₁: 5657.1, found: 5656.2 (average mass, deconvoluted)

HPLC analysis: t_R = 2.78 min Chromolith, gradient: 20-50% B over 5 min, respectively

HPLC purification: [5 mg/mL] Nucleosil C18, gradient: 16-21% B over 20 min, affording a white solid after lyophilisation (29% yield).



Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (5.88)	3514.8	3514.1	Ac-[124-160]
B (6.01)	5753.1	5753.2	TFA ester on a Ser or Thr residue
C (6.01)	5657.1	5656.3	21



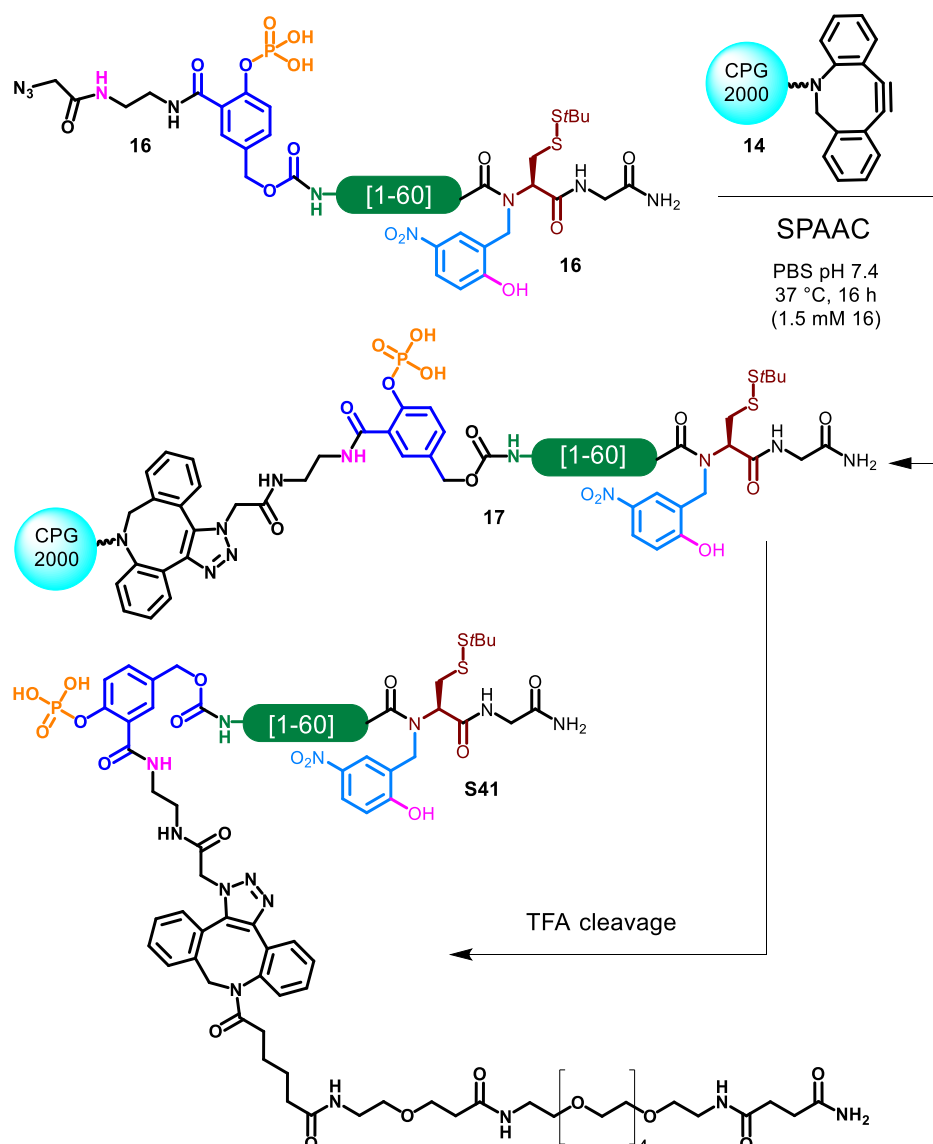
Supplementary figure S61: HPLC traces and MS analysis (average masses, deconvoluted) of crude (A) and purified (B) compound **21** (Aeris Widepore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C, and Chromolith, gradient: 20-50% B over 5 min, respectively)

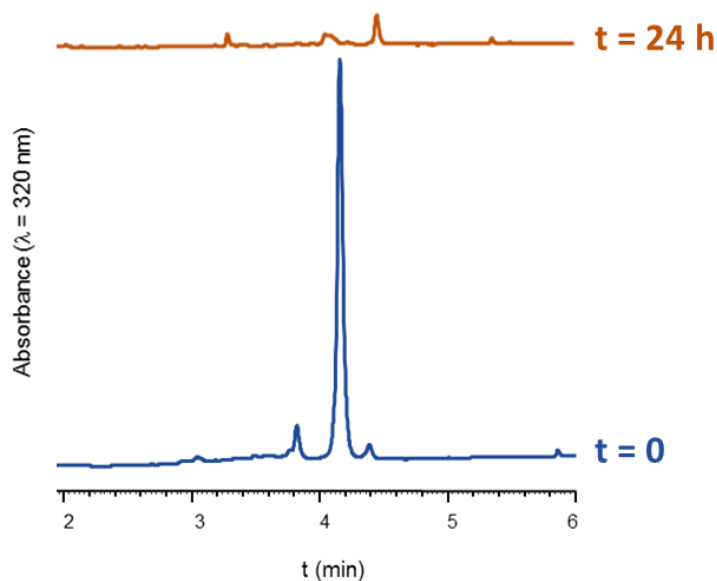
10- Synthesis of polypeptide 23 through two successive solid supported NCL

We observed that storage of the intermediates during a few days or weeks led to a reduced yield upon cleavage (either TFA- or enzyme-mediated), and attributed this finding to the aggregation of the peptide on solid support, as already suspected in other examples of SPCL. [15] For this reason, and with the aim to obtain a good overall yield, the five assembly steps were carried out in a single continuous process, immediately between steps.

10-1- Grafting of the first segment 16 on solid support *via* SPAAC reaction to obtain 19

The same protocol as for the grafting of the model peptide 13 on the solid support was used. (see p S53-S54)





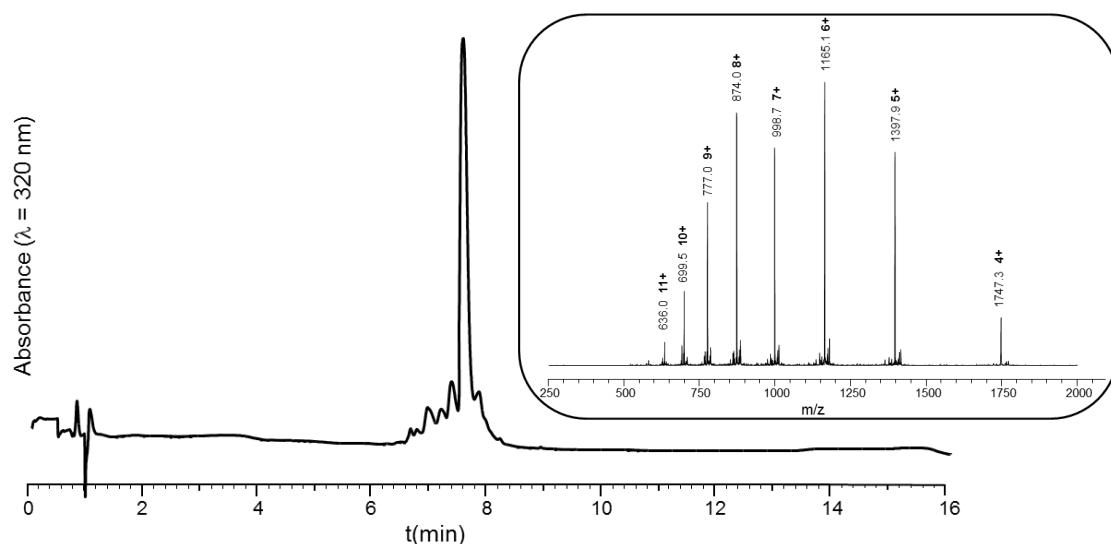
Supplementary figure S62: HPLC monitoring of the SPAAC reaction showing the disappearance of the starting azide **16** (HPLC conditions: Chromolith, gradient: 5-50% B over 5 min)

For analytical purpose, an aliquot of **17** was cleaved (TFA-mediated cleavage of the Rink linker, protocol: PS3) to give compound **S41**.

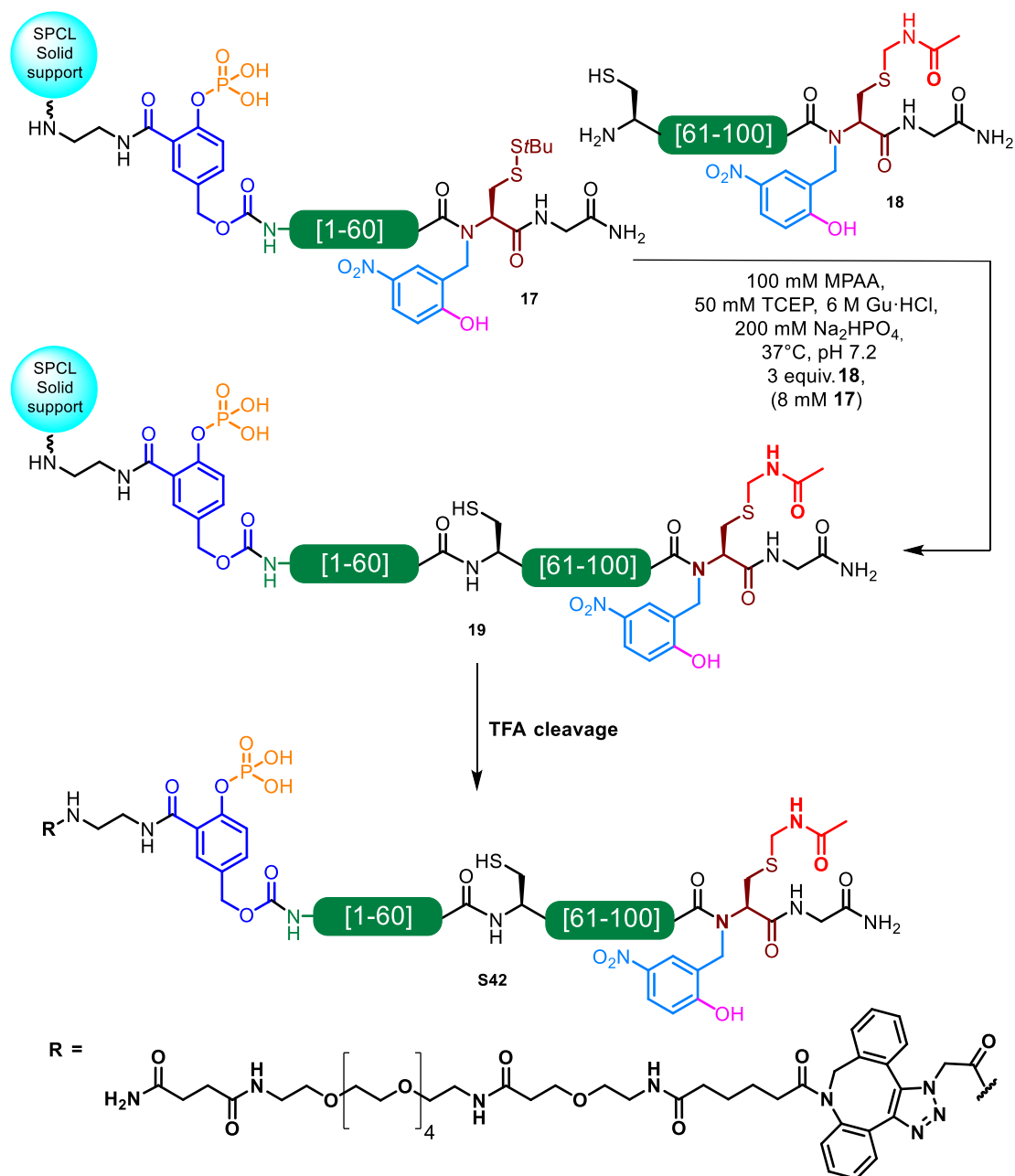
- Compound **S41**

ESI-MS (m/z): [M] calcd. for $C_{300}H_{451}N_{88}O_{100}PS_2$: 6986.4, found: 6984.4 (average mass, deconvoluted)

HPLC analysis: $t_R = 7.96$ min (Aeris Widepore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C)



Supplementary figure S63: HPLC trace of released compound **S41**. Insert: MS spectrum corresponding to the major peak.

10-2- NCL between solid-supported crypto-thioester **17** and cysteinyl peptide **18**

Under argon, solid supported crypto-thioester **17** (300 nmol, 8 mM final concentration) was washed (3 x 30 min) with the NCL buffer (see protocol PS6), then the buffer was carefully removed leaving the solid support as dry as possible. Next cysteinyl peptide **18** (4.3 mg, 900 nmol, 3 equiv., 24 mM final concentration) was dissolved in the NCL buffer (38 μ L), added to the solid support and the resulting mixture was gently stirred under argon at 37 °C for 18 h. Note that it is crucial for the reaction completion to wash the crypto-thioester solid support with the NCL buffer before starting the NCL reaction.

For analytical purpose, an aliquot of support **19** was cleaved (TFA-mediated cleavage of the Rink linker, protocol: PS3) to give compound **S42**.

After each analytical cleavage after ligations and prior to HPLC and MS characterization, the crude released product was treated with a solution of 50 mM TCEP solution in 0.2 M phosphate buffer pH 7 to reduce possible disulfides that could be formed.

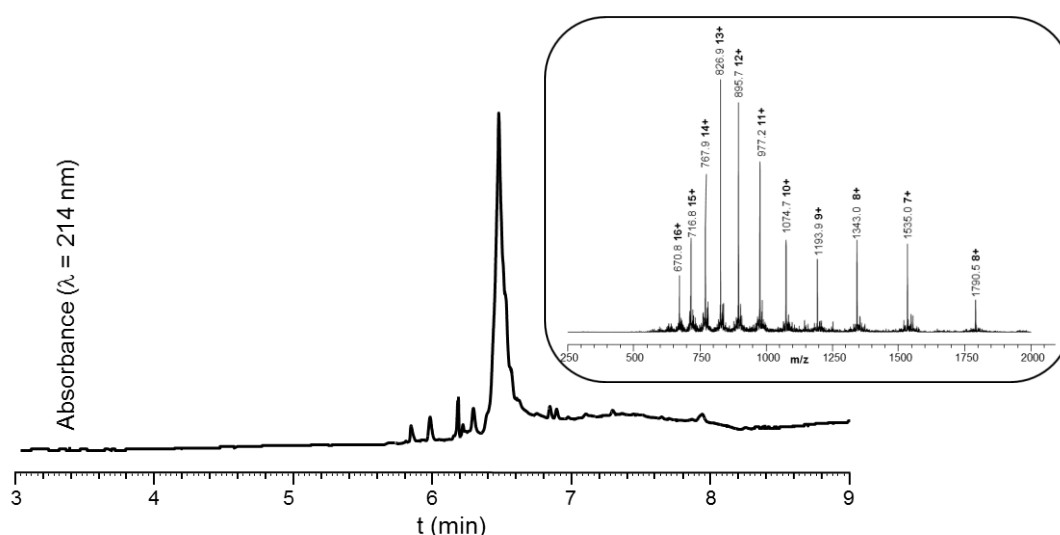
Disulfide reduction protocol:

To 100 μ L of a solution of the crude released product in water, 100 μ L of a 50 mM TCEP solution pH 7 was added. Next, the mixture was left at room temperature for 20 min, then quenched through addition of 25 μ L of 10% TFA in water.

- Compound **S42**

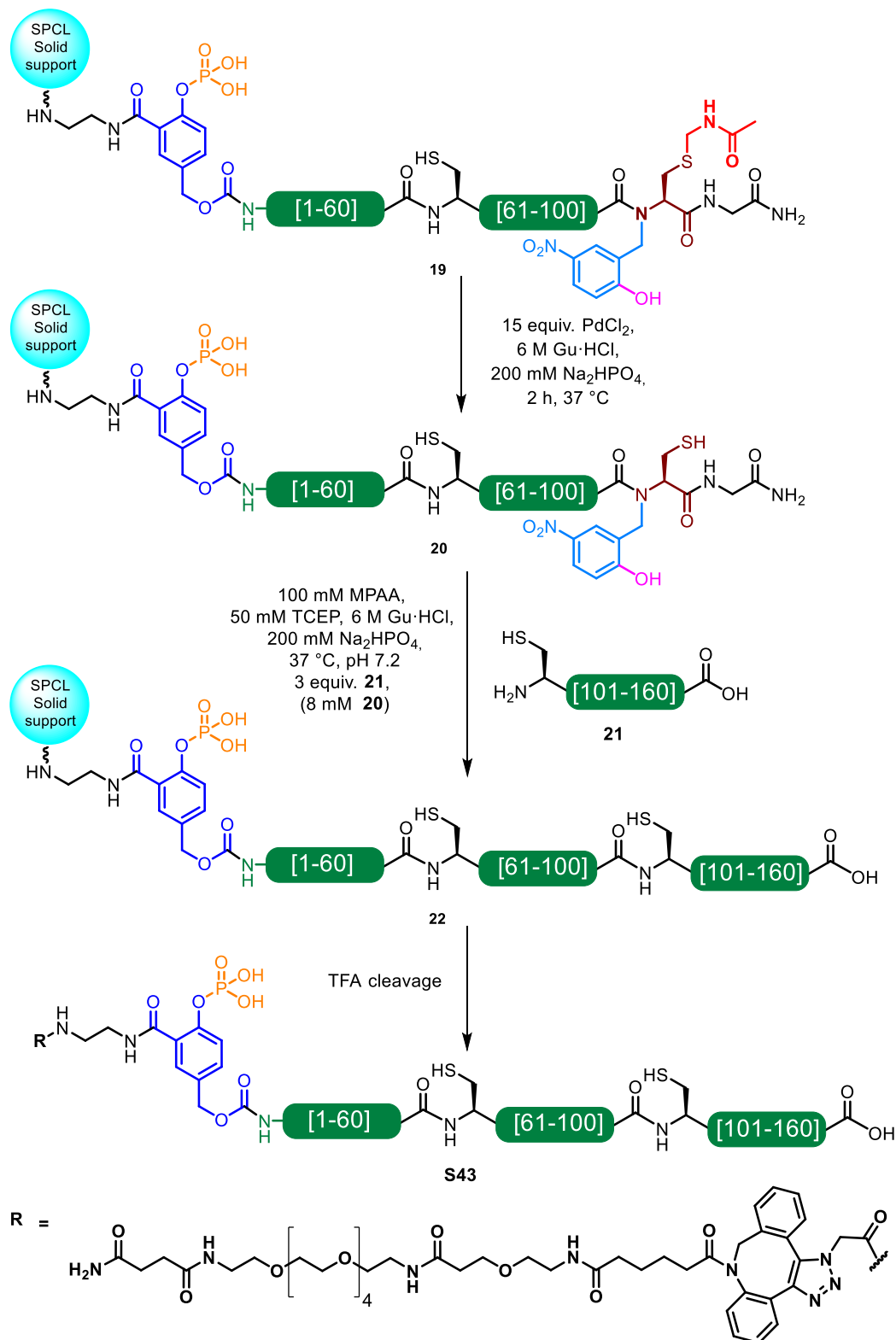
ESI-MS (m/z): [M] calcd. for $C_{459}H_{698}N_{139}O_{155}PS_2$: 10738.4, found: 10737.1 (average mass, deconvoluted)

HPLC analysis: t_R = 6.51 min (Aeris Widespore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C)



Supplementary figure S64: HPLC trace of released compound **S42** obtained after a first solid-supported NCL. Insert: MS spectrum corresponding to the major peak.

10-3- Removal of Acm protecting group and second NCL between solid supported cryptothioester 20 and cysteinyl peptide 21



PdCl₂ (15 equiv.) was dissolved in 6 M guanidinium chloride, 200 mM Na₂HPO₄ buffer and stirred at 37 °C for 30 min under an argon atmosphere. Then the solution was added to the solid supported Acm-

protected crypto-thioester **19** (1 mM supported peptide concentration) and the reaction stirred for 2 h under argon. Next the solid support was extensively washed similarly to post-ligations washings (see protocol S6)

Under argon, crypto-thioester solid support **20** (300 nmol, 8 mM final concentration) was washed (3 x 45 min) with the NCL buffer, the complete removal of the palladium salts from the solid support could be monitored according to the intensity of the solid support colour (usually 3 washing cycles are enough to restore the initial solid support color)

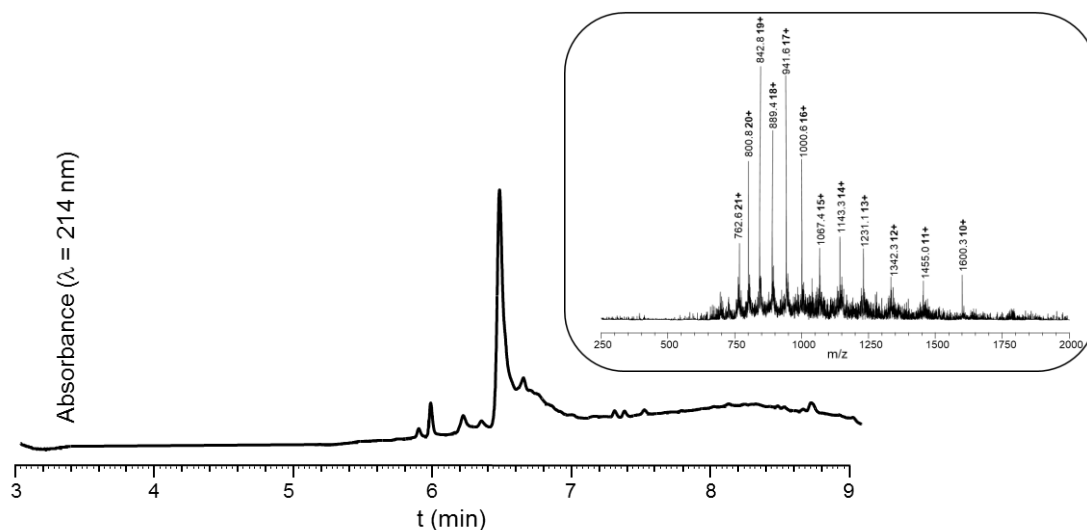
Then, the buffer was carefully removed leaving the solid support as dry as possible. Next cysteinyl peptide **21** (5.8 mg, 900 nmol, 3 equiv., 24 mM final concentration) was dissolved in NCL buffer (38 μ L), added to the solid support and the resulting mixture was gently stirred under argon at 37 °C for 18 h.

For analytical purpose, an aliquot of support **22** was cleaved (TFA-mediated cleavage of the Rink linker, protocol: PS3) to give compound **S43**. The crude released product was treated with a solution of 50 mM TCEP solution in 0.2 M phosphate buffer pH 7 to reduce possible disulfide that could be formed as described for **S42** (protocol PS6).

➤ Compound **S43**

ESI-MS (m/z): [M] calcd. for $C_{684}H_{1054}N_{209}O_{231}PS_2$: 15996.1, found: 15993.7 (average mass, deconvoluted)

HPLC analysis: t_R = 6.57 min (Aeris Widepore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C)



Supplementary figure S65: HPLC trace of released compound **S43** obtained after the second solid-supported NCL. Insert: MS spectrum corresponding to the major peak.

10-4- Enzymatic cleavage and HPLC purification

Enzymatic linker cleavage was carried out with lambda phosphatase. Solid-supported polypeptide **22** was incubated at 30 °C with the enzyme (4 enzyme unit/10nmol of peptide) in the PMP buffer (100 µM solid-supported peptide concentration) for 48 h to afford ~20 % release yield of **23**. The enzymatic treatment was repeated 4 more times for 24 h to afford an overall 60% release yield.

The quantity of ligation product **23** released in solution was evaluated by integration of the HPLC peak at $\lambda = 214$ nm in the crude release mixture, using tryptophan as an external calibrant. 214 nm molar epsilon coefficients were calculated according to Kuipers and and Gruppen [16]: $\epsilon_{23} = 299655 \text{ L.mol}^{-1}.\text{cm}^{-1}$ and $\epsilon_{\text{Trp}} = 29050 \text{ L.mol}^{-1}.\text{cm}^{-1}$ in $\text{H}_2\text{O}/\text{MeCN}/\text{formic acid } 80:20:0.1$. The release yield was calculated by dividing the quantity of released **23** by the initial amount of the N_3 -linker-segment1 **16** immobilized by SPAAC, evaluated by UV titration at 280 nm ($\epsilon_{16} = \epsilon_{\text{2nd generation linker}} = 2718 \text{ L.mol}^{-1}.\text{cm}^{-1}$).

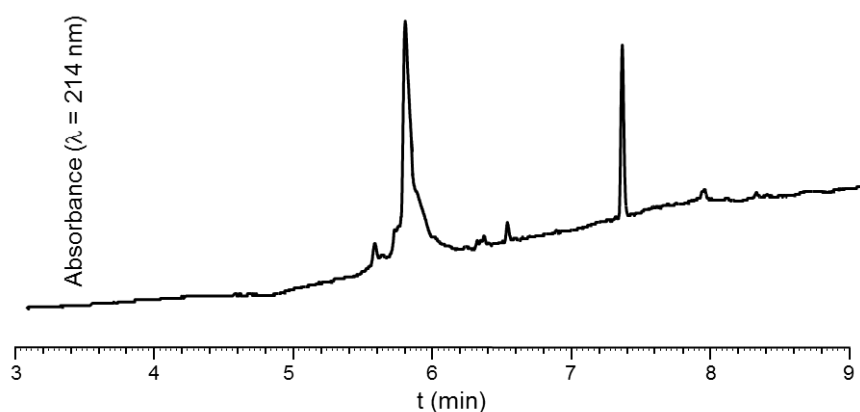
- Compound **23**

ESI-MS (m/z): [M] calcd. for $\text{C}_{640}\text{H}_{1002}\text{N}_{200}\text{O}_{217}\text{S}_2$: 15034.1, found: 15032.4 (average mass, deconvoluted)

ESI-HRMS (m/z): [M] calcd. for $\text{C}_{640}\text{H}_{1002}\text{N}_{200}\text{O}_{217}\text{S}_2$: 15025.2961, found: 15025.3414 (deconvoluted)

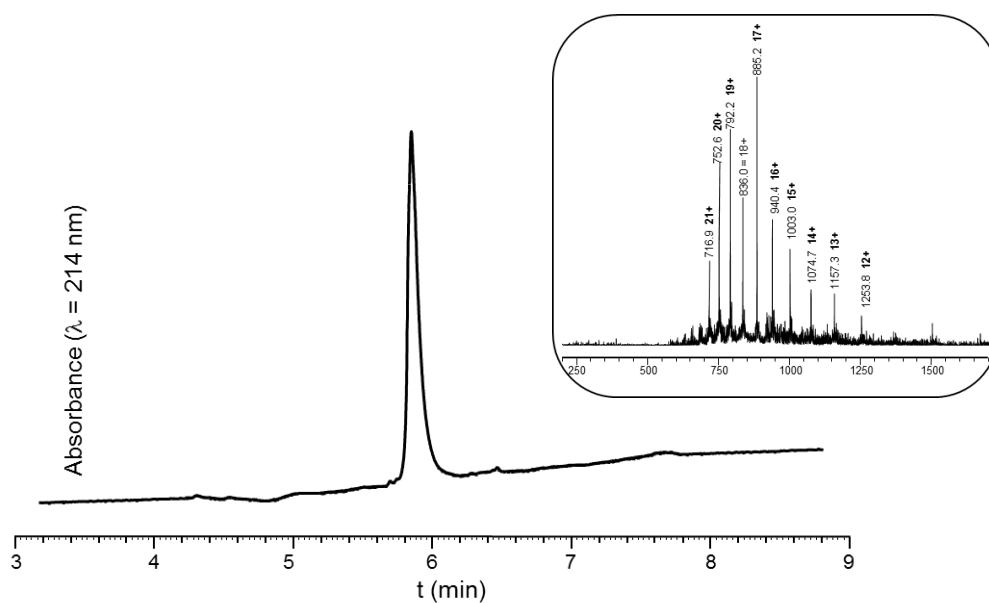
HPLC analysis: $t_R = 5.84$ min (Aeris Widedpore XB-C18 2, gradient: 3-50% B' over 15 min, 60 °C)

HPLC purification: [0.25 mg/mL] Jupiter C4, gradient: 11-17% B over 20 min, affording a white solid after lyophilisation (13% overall yield, determined relatively to the initial amount of immobilized **16**).

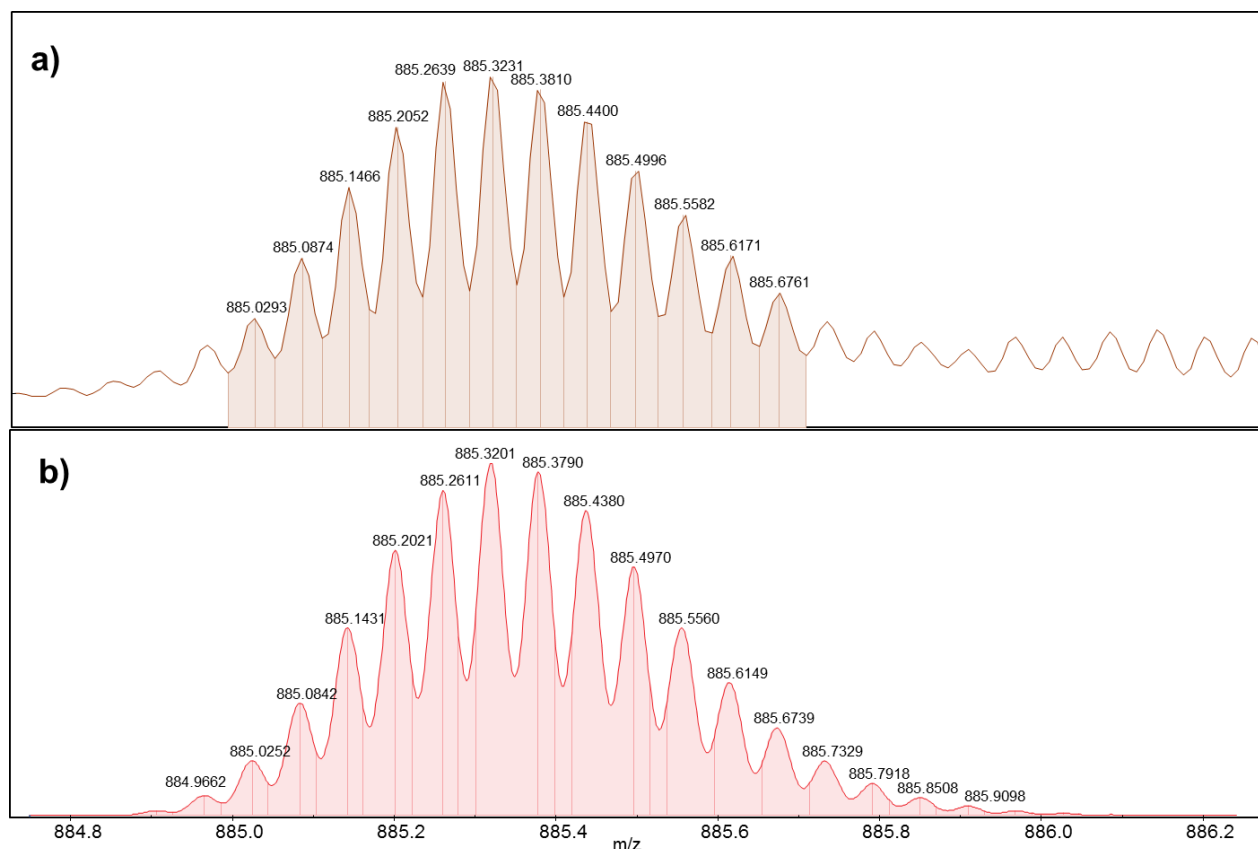


Peak number (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (5.84)	15034.1	15032.4	23
B (7.43)	-	1120.7	Non peptidic compound*

*: Note that peaks B is also observed when treating the starting alkyne solid support **14** under the same conditions, and thus does not correspond to a peptide.



Supplementary figure S66: HPLC traces and MS analysis of crude (A) and purified (B) compound **23**.
Insert: mass spectrum of purified compound.



Supplementary figure S67: a) zoom on the $[M+17H]^{17+}$ peak of the high-resolution mass spectrum; b) theoretical isotopic distribution of the $[M+17H]^{17+}$ peak

11- References

- Legigan, T., Clarhaut, J., Tranoy-Opalinski, I., Monvoisin, A., Renoux, B., Thomas, M., Le Pape, A., Lerondel, S., and Papot, S. (2012) The First Generation of β -Galactosidase-Responsive Prodrugs Designed for the Selective Treatment of Solid Tumors in Prodrug Monotherapy. *Angew. Chem. Int. Ed.*, **51** (46), 11606–11610.
- Kaiser, E., Colescott, R.L., Bossinger, C.D., and Cook, P.I. (1970) Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal Biochem*, **34** (2), 595–598.
- Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.*, **4** (11), 2411–2423.
- Shi, W., Tang, F., Ao, J., Yu, Q., Liu, J., Tang, Y., Jiang, B., Ren, X., Huang, H., Yang, W., and Huang, W. (2020) Manipulating the Click Reactivity of Dibenzazacyclooctynes: From Azide Click Component to Caged Acylation Reagent by Silver Catalysis. *Angew. Chem. Int. Ed.*, **59** (45), 19940–19944.
- Sato, K., Shigenaga, A., Kitakaze, K., Sakamoto, K., Tsuji, D., Itoh, K., and Otaka, A. (2013) Chemical Synthesis of Biologically Active Monoglycosylated GM2-Activator Protein Analogue Using N -Sulfanylethylanilide Peptide. *Angew. Chem. Int. Ed.*, **52** (30), 7855–7859.
- a) Erlich, L.A., Kumar, K.S.A., Haj-Yahya, M., Dawson, P.E., and Brik, A. (2010) N-methylcysteine-mediated total chemical synthesis of ubiquitin thioester. *Org. Biomol. Chem.*, **8** (10), 2392–2396. b) Aihara, K., Yamaoka, K., Naruse, N., Inokuma, T., Shigenaga, A., and Otaka, A. (2016) One-

Pot/Sequential Native Chemical Ligation Using Photocaged Crypto-thioester. *Org. Lett.*, **18** (3), 596–599.

7. Terrier, V.P., Adihou, H., Arnould, M., Delmas, A.F., and Aucagne, V. (2016) A straightforward method for automated Fmoc-based synthesis of bio-inspired peptide crypto-thioesters. *Chem. Sci.*, **7** (1), 339–345.

8. Sato, K., Kitakaze, K., Nakamura, T., Naruse, N., Aihara, K., Shigenaga, A., Inokuma, T., Tsuji, D., Itoh, K., and Otaka, A. (2015) The total chemical synthesis of the monoglycosylated GM2 ganglioside activator using a novel cysteine surrogate. *Chem. Commun.*, **51** (49), 9946–9948.

9. Nishimura, O., Kitada, C., and Fujino, M. (1978) New method for removing the S-p-methoxybenzyl and S-t-butyl groups of cysteine residues with mercuric trifluoroacetate. *Chem. Pharm. Bull.*, **26** (5), 1576–1585.

10. Harris, K.M., Flemer, S., and Hondal, R.J. (2007) Studies on deprotection of cysteine and selenocysteine side-chain protecting groups. *J. Pept. Sci.*, **13** (2), 81–93.

11. Maity, S.K., Jbara, M., Laps, S., and Brik, A. (2016) Efficient Palladium-Assisted One-Pot Deprotection of (Acetamidomethyl)Cysteine Following Native Chemical Ligation and/or Desulfurization To Expedite Chemical Protein Synthesis. *Angew. Chem. Int. Ed.*, **55** (28), 8108–8112.

12. Royo, M., Alsina, J., Giralt, E., Slomczynska, U., and Albericio, F. (1995) S-Phenylacetamidomethyl (Phacm): an orthogonal cysteine protecting group for Boc and Fmoc solid-phase peptide synthesis strategies. *J. Chem. Soc., Perkin Trans. 1*, (9), 1095.

13. Abboud, S.A., and Aucagne, V. (2020) An optimized protocol for the synthesis of N-2-hydroxybenzyl-cysteine peptide crypto-thioesters. *Org. Biomol. Chem.*, **18**, 8199–8208.

14. Lukszo, J., Patterson, D., Albericio, F., and Kates, S.A. (1996) 3-(1-Piperidinyl)alanine formation during the preparation of C-terminal cysteine peptides with the Fmoc/t-Bu strategy. *Lett Pept Sci*, **3** (3), 157–166.

15. a) Brik, A., Keinan, E., and Dawson, P.E. (2000) Protein Synthesis by Solid-Phase Chemical Ligation Using a Safety Catch Linker. *J. Org. Chem.*, **65** (12), 3829–3835. b) Yu, R.R., Mahto, S.K., Justus, K., Alexander, M.M., Howard, C.J., and Ottesen, J.J. (2016) Hybrid phase ligation for efficient synthesis of histone proteins. *Org. Biomol. Chem.*, **14** (9), 2603–2607.

16. Kuipers, B.J.H., and Gruppen, H. (2007) Prediction of Molar Extinction Coefficients of Proteins and Peptides Using UV Absorption of the Constituent Amino Acids at 214 nm To Enable Quantitative Reverse Phase High-Performance Liquid Chromatography–Mass Spectrometry Analysis. *J. Agric. Food Chem.*, **55** (14), 5445–5451.

Chapter 3: An improved method for the solid-phase synthesis of *N*-Hnb-Cys crypto thioester peptides

Table of contents

I. Bibliographic study	204
1) Peptide thioesters and NCL.....	204
2) Fmoc-based synthesis of peptide thioesters	206
A. Overview of the most popular approaches.....	207
a) Thioesterification of protected peptide acids.....	207
b) The use of sulfonamide safety catch linkers	208
c) Benzimidazolones as thioester precursors	209
d) Other imides as thioester precursors.....	210
e) Hydrazides as thioester precursors.....	210
B. Intramolecular acyl shift-based devices	211
a) <i>O</i> -S shift-based devices	212
b) <i>N</i> -S shift-based devices	213
i- <i>N</i> -S shift systems requiring acidic conditions for the rearrangement	213
ii- <i>N</i> -S shift-based crypto-thioesters	216
C. <i>N</i> -(2-hydroxybenzyl)cysteine (<i>N</i> -Hnb-Cys) thioesterification device	222
a) Limitation of other <i>N</i> -S shift based crypto-thioesters	222
b) Design of the <i>N</i> -Hnb-Cys thioesterification device	222
II. Aim of this work: an optimized protocol for the synthesis of <i>N</i>-Hnb-Cys peptide crypto-thioesters	226
III. References for the bibliographic study part	228
IV. Publication 3 (Published in Organic & Biomolecular Chemistry)	236
V. Supporting Information	247

I. Bibliographic study

1) Peptide thioesters and NCL

As mentioned in the [first chapter](#) of this manuscript, one of the most practical and robust methods for chemical protein synthesis is the native chemical ligation (NCL) reaction. This method reported by Kent and co-workers in 1994 [1] has become an often used and powerful method for the connection of unprotected peptides segments *via* a native amide bond. It was inspired from a reaction observed by Wieland and co-workers in 1953. [2] This pioneering work consisted in coupling a thioester derivative of valine with cysteine in water in a chemo- and regioselective way to obtain the corresponding dipeptide, *N*-valyl cysteine. This coupling was found to be particularly efficient with cysteine compared to other amino acids like alanine or glycine due to the formation of the thioester intermediate. (fig. 1).

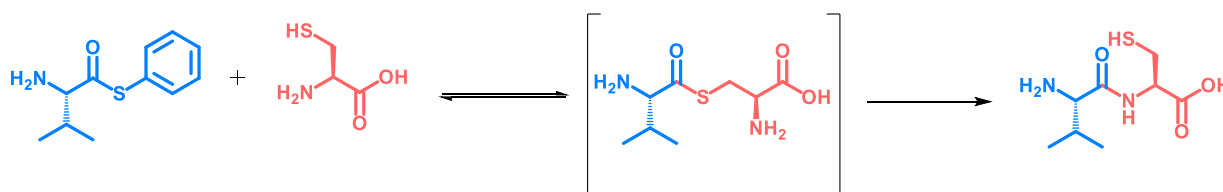


Figure 1: Formation of *N*-valyl cysteine through a trans-thioesterification / *S*-*N* acyl shift mechanism

The seminal work of Kent was also inspired by the strategy reported by Kemp [3,4] and based on the use of a “thiol capture auxiliary” (C-terminal 6-hydroxy-4-mercaptodibenzofuran ester) for the formation of an amide bond through an intramolecular *O*-*N* acyl transfer (fig. 2).

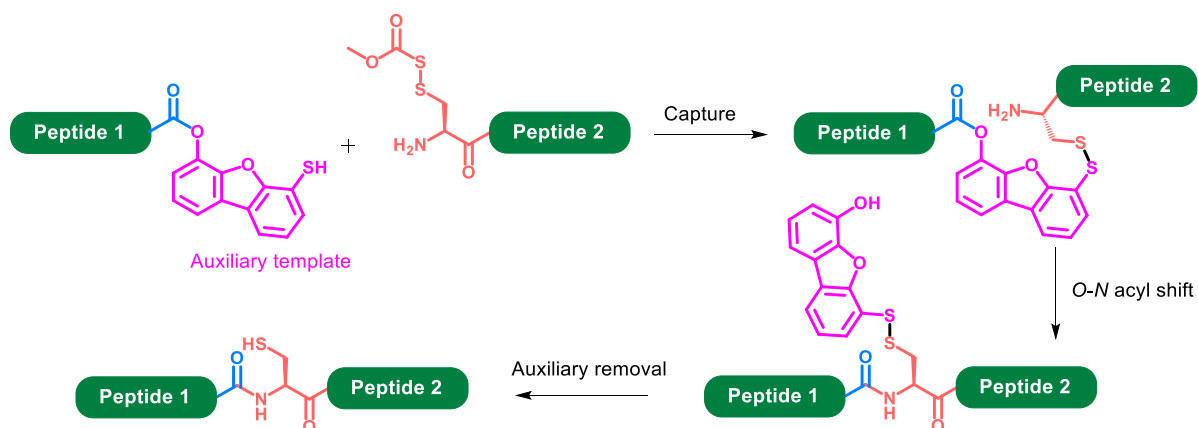


Figure 2: Chemoselective amide bond formation promoted by the 6-hydroxy-4-mercaptodibenzofuran auxiliary

In NCL, a peptide having a C-terminal thioester reacts with the N-terminal cysteine residue of another peptide in aqueous solution, at neutral pH and at mild temperatures to form an

intermediate thioester with the cysteine thiol through a trans-thioesterification reaction. This intermediate then spontaneously rearranges by an *S-N* acyl transfer to form a native amide bond between the peptides (fig. 3), following a mechanism exactly similar as in Wieland's synthesis of *N*-valyl cysteine.

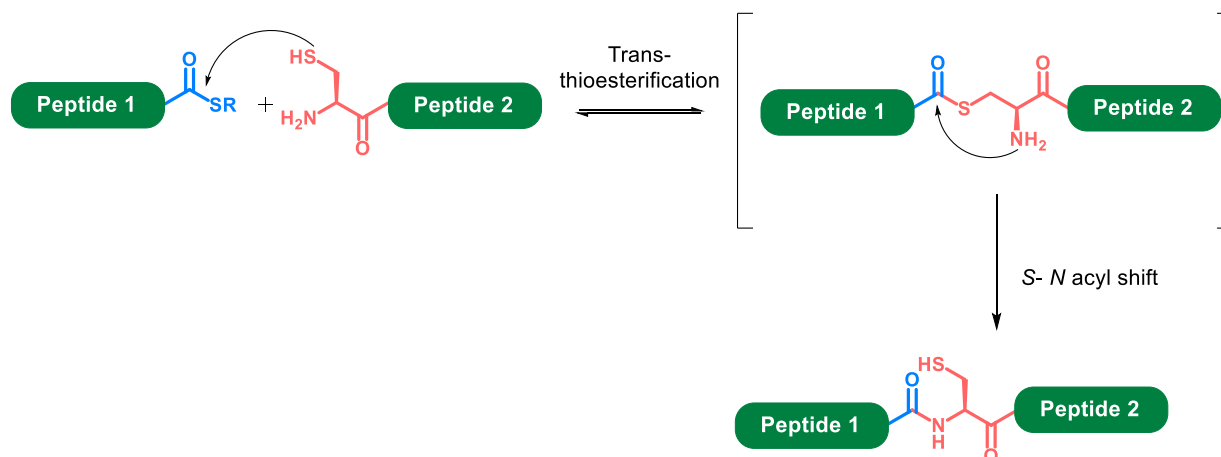


Figure 3: Native chemical ligation mechanism

Since the first example of the use of NCL for protein synthesis mentioned above, countless proteins have been synthesized with this new method. As starting materials for NCL, peptide C- α -thioesters play a key role in protein synthesis strategies, it was therefore crucial to develop straightforward methods for their preparation.

Peptide thioesters are easily prepared when using the Boc-SPPS strategy, using thioester-containing linkers (fig. 4) like the 3-mercaptopropionamide linker (MPAL) [5,6] or the 4-mercaptophenylacetamide linker. [7]

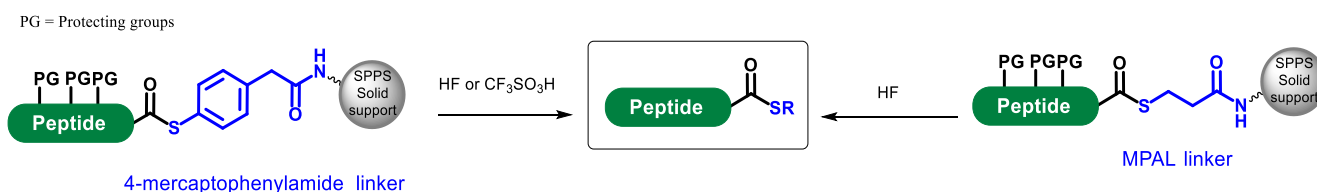


Figure 4: Synthesis of peptide thioesters by Boc-SPPS

However, this strategy is less and less used due to many disadvantages compared to the Fmoc strategy. The repeated TFA-mediated *N*- α -Boc deprotection during long synthesis may lead to modification and/or degradation of sensitive peptides and the main drawback remains the use of dangerous and harsh anhydrous HF final cleavage. However, it must be said that Boc-SPPS remains superior to Fmoc-SPPS in some cases of peptides with so-called “difficult sequences”, where the SPPS is problematic due to aggregation of the protected peptide on the resin: repeated TFA treatments used to deprotect the Boc group tend to promote disaggregation.

2) Fmoc-based synthesis of peptide thioesters

The other option is the use of the Fmoc-SPPS strategy, which is nowadays much more widely used than the historical Boc-SPPS. Nonetheless, this strategy is challenging for peptide thioesters synthesis due their incompatibility with the piperidine treatments repeatedly used to deprotect Fmoc groups. Indeed, nucleophilic piperidine readily reacts with the electrophilic carbonyl moiety of the thioester (fig. 5).

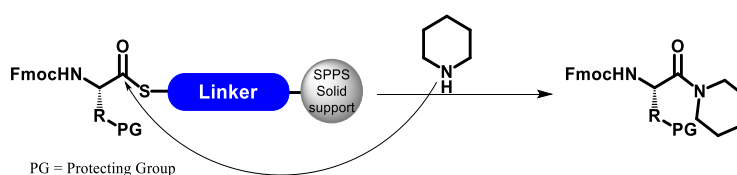


Figure 5: Side reaction with piperidine nucleophilic attack

1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU), a non-nucleophilic base [8] was used in some reports along with HOBt as a replacement to piperidine in order to prevent this side reaction, however this strategy did not prevent completely the degradation of thioesters during the coupling the first amino acids. Indeed, it appears that the peptide thioester is most vulnerable to aminolysis-derived C-terminal deletion sequence formation during the first amino acids addition and that its resistance to nucleophilic attacks is increased as the peptide chain grows thanks to steric hindrance.[9,10]

In a similar work, Aimoto and coworkers [11] examined the synthesis of peptide thioesters using a mixture of 1-methylpyrrolidine, hexamethylenimine, and HOBt for the removal of the Fmoc groups. HOBt was used to reduce the basicity of the Fmoc deprotecting mixture and thus thioester aminolysis and hydrolysis. However, it has been showed later that this protocol induces the partial epimerization of the C-terminal residue. [12]

Thus the success of native chemical ligation and related strategies has sparked intensive research effort devoted to the development of alternative methods. To overcome thioesters sensitivity towards repetitive piperidine treatments used for Fmoc-SPPS, thioesters moieties can be generated after the elongation of the peptide. The great advantage of this strategy is that exposure of the thioester group to all the reagents required to assemble the polypeptide chain is avoided. Many ingenious developments made over the years on the Fmoc-based synthesis of peptide thioesters have led to a variety of methods which are based on different various mechanisms. However, despite huge efforts, no universal Fmoc-SPPS based strategy has yet emerged. The aim of this bibliographic study is to provide a broad overview of the most representative methods without being exhaustive.

A. Overview of the most popular approaches

a) Thioesterification of protected peptide acids

An early used approach relies on the generation of the thioester moiety through thioesterification of a protected peptide carboxylic acid in solution (fig. 6-a). [13,14]

This late-stage thioesterification approach can also be performed on solid support to avoid solubility problems encountered with protected peptides in solution. One possibility is the grafting of the peptide on the solid support through a side chain function (fig. 6-c). [15–18] In this case, the C-terminal carboxylic moiety is protected as an Pd(0)-labile allyl ester during SPPS. After elongation, the allyl group is removed and a thioester function is introduced by coupling the resulting carboxylic acid either with a thiol or an α -amino thioester derivative of an amino acid. Another similar approach is based on the peptide grafting on solid support via a backbone amide linker (BAL) (fig. 6-b). [19,20] An advantage of the BAL approach towards side-chain anchoring is that it is not dependent on the nature of C-terminal amino acids. Moreover, *N*-alkylation of the C-terminal residue prevents epimerization *via* oxazolone formation during the coupling to introduce the thioester function.

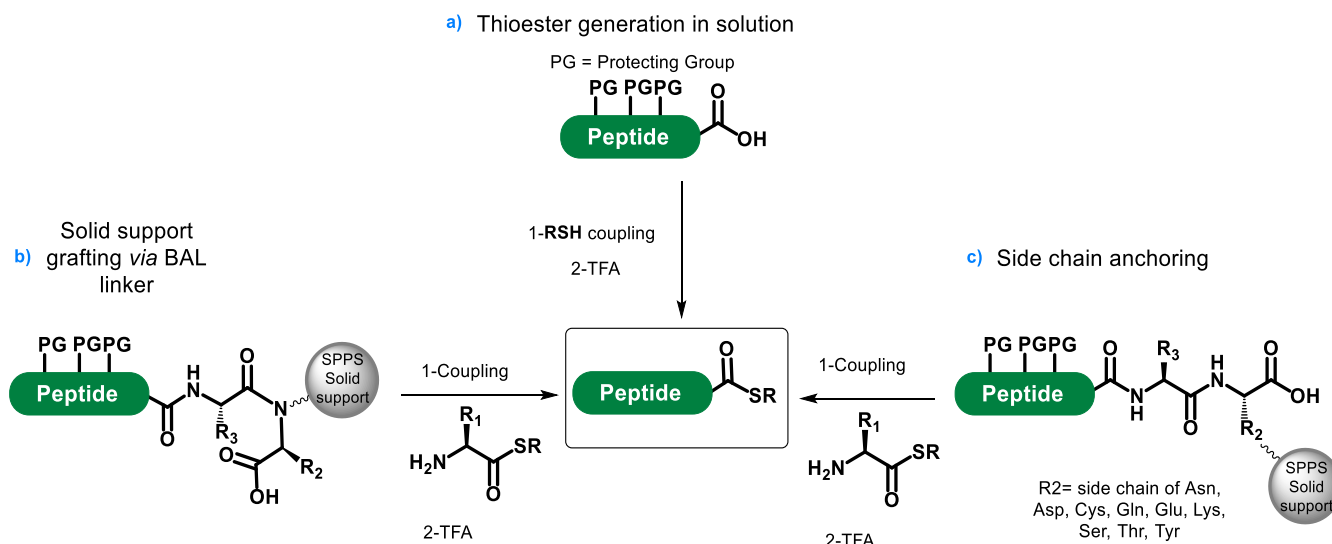


Figure 6: Solution- or solid-phase introduction of thioester moiety on protected peptide acids

b) The use of sulfonamide safety catch linkers

Another widely used method in the Fmoc-SPPS-based synthesis of peptide thioesters involves the use of safety-catch linkers inspired from Kenner's sulfonamide linker developed for peptide C-terminal diversification (fig. 7-a). [21] After peptide elongation on an alkylsulfonamide resin, the peptidyl-sulfonamide bond is *N*-alkylated with iodoacetonitrile [22,23] or trimethylsilyldiazomethane [23] making it vulnerable to nucleophilic attack. Thiolysis releases into solution fully protected peptide thioesters, which are then treated with TFA for final removal of the protecting groups. In order to avoid solubility problems associated with fully-protected peptides in solution, Melnyk and co-workers extended the method to an intramolecular version. Alkylation of the sulfonamide linker with *S*-triisopropylsilyl β -mercapto-ethanol using the Mitsunobu reaction, followed by fluoride-mediated thiol deprotection allows intramolecular thiolysis on solid support after (fig. 7-b). [24] In a complementary approach, Burlina and co-workers demonstrated the possibility of performing *in situ* thiolysis under NCL conditions of an *N*-methyl *N*-peptidyl sulfonamide linker, since the latter resists to TFA treatment (fig. 7-c). [25]

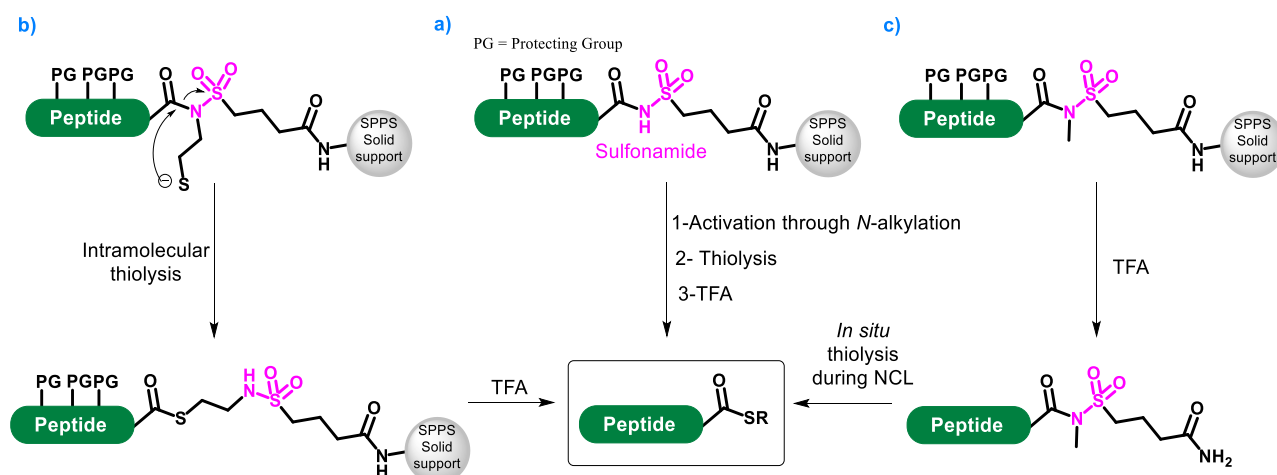


Figure 7: Use of sulfonamide safety catch linkers for thioesters generation

c) Benzimidazolones as thioester precursors

Within the same concept of safety catch linkers, Dawson and co-workers [26] reported a new safety catch principle based on the activation of a 3,4-diaminobenzoyl (Dbz) linker with *p*-nitrophenylchloroformate followed by a spontaneous ring closure leading to the formation of an *N*-acyl benzimidazolone (Nbz) (fig. 8). These Nbz peptides can be used for *in situ* generation of peptide thioesters during NCL due to the ease of Nbz thiolysis at neutral pH. The main drawback of this approach was the acylation of the free aniline moiety during Fmoc-SPPS and therefore several next generation versions of the Dbz linker were designed where the aniline was monomethylated, [27] or protected with an allyloxycarbonyl [28], or a 2-chlorobenzyloxycarbonyl group. [29]

The usefulness of the approach has been demonstrated by many works, [26,27,30,31] and remains up to day one of the most widely used method to generate thioesters, thanks to the commercial availability of the Dbz linker. However, Nbz peptides are highly reactive, thus quite sensitive to hydrolysis during purification and NCL.

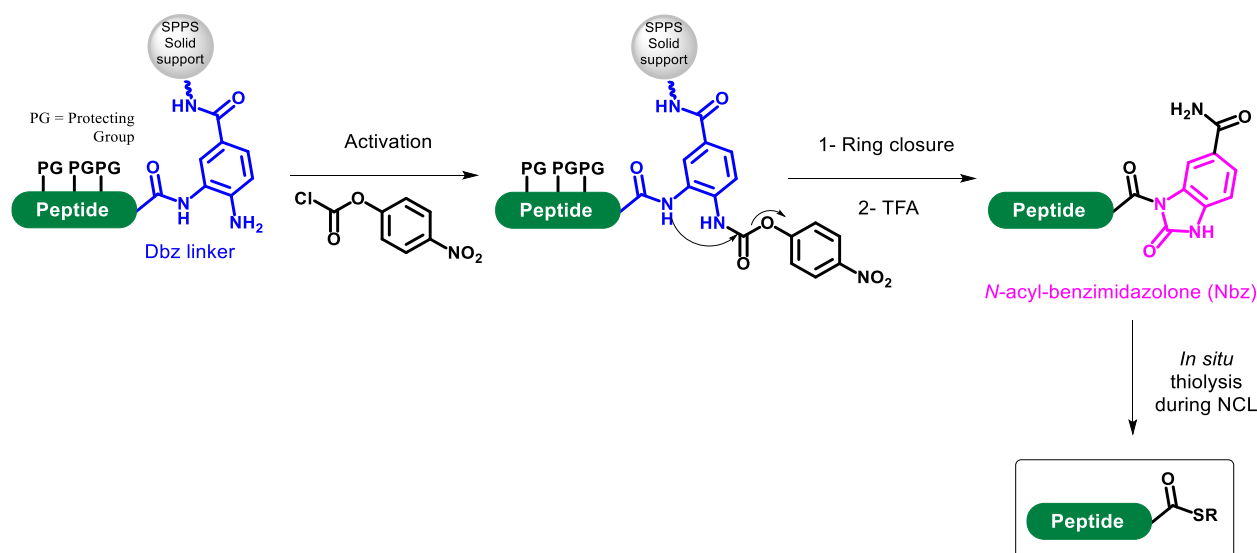


Figure 8: Synthesis of peptide benzimidazolones as highly reactive precursors for peptide thioesters

d) Other imides as thioester precursors

Similar approaches to Dawson's method have also been recently reported, based on pyroglutamyl imides [32] and cyclic *N*-acyl-urethanes [33] as precursors for peptide α -thioesters, obtained from the cyclization of glutamic acid and serine residues, respectively, (fig. 9). However, high temperatures and long reaction times are needed to obtain the peptide thioesters in these cases.

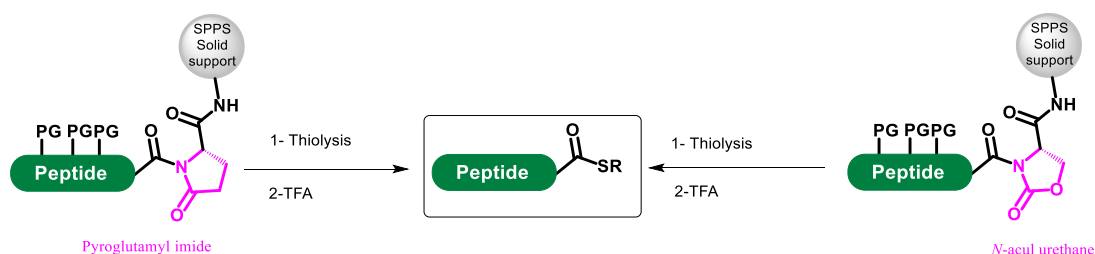


Figure 9: Thioester functionality generation from an imide precursor

e) Hydrazides as thioester precursors

Another safety-catch linker reported for the synthesis of C-terminal thioesters is the hydrazide linker. After SPPS elongation on an *N*-aryl hydrazide [34] solid support, *N*-bromosuccinimide (NBS) is used to oxidize hydrazide into the electrophilic acyl diazene followed by aminolysis with thioesters derivative of amino acid to release the protected thioester into solution. Further TFA treatment led to the formation of unprotected peptide thioesters (fig. 10-a). [35]

A very popular method that evolved from this work was reported by Lei Liu and co-workers, [36] building on the simplicity to synthesize C-terminal hydrazide peptides using

hydrazinecarboxylate loaded on a Wang resin, or hydrazine loaded on a 2-chlorotrityl resin as starting SPPS solid supports. The great advance of Liu's strategy was the possible selective activation of the hydrazide in solution by the action of sodium nitrite on the deprotected peptide under acidic conditions to form an acyl azide, followed by thiolysis to form a thioester, thus allowing its one-pot-three-steps utilization in NCL (fig. 10-b). Despite some difficulties that have been reported during the use of peptide hydrazides featuring C-terminal Asn, Gln [37] and Lys residues, [38] in NCL, this method represented a major advance in the field of protein chemical synthesis and was widely used since then. [39–42] Today this method is indeed the most used strategy, [43] and, besides its simplicity, offers the significant advantage to afford a stable, unreactive thioester precursors that can be activated on demand, a very appealing feature to perform N-to-C successive assembly of multiple segments.

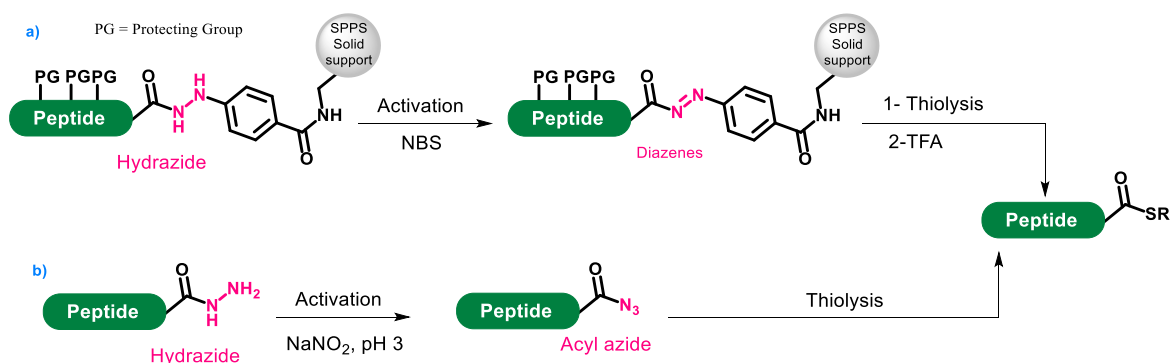


Figure 10: Thioester functionality generation from hydrazide oxidation using both solution- and solid-phase methods

B. Intramolecular acyl shift-based devices

As stated above, the most widely used thioester precursors are nowadays Liu's hydrazides and Dawson's Nbz. However, if these two methodologies are easy to implement and significantly expanded the applicability of NCL-based protein chemical synthesis, they suffer from some limitations and there was still a need for alternative synthetic strategies. In particular, both methods imply a post-elongation activation step that can be tricky to implement, and is not automatable on a peptide synthesizer.

A new strategy that have been increasingly used in the last decade is based on the spontaneous formation of a thioester from either a β -mercapto amide or a β -mercapto ester "intramolecular thioesterification devices" *via* a reversible intramolecular *N*-S (fig. 11-a) [44,45] or *O*-S acyl shift, respectively (fig. 11-b). [46,47] In both cases, a large excess of external

thiol intercepts the *S*-acyl products and shift the equilibrium in favor of the peptide thioester through trans-thioesterification.

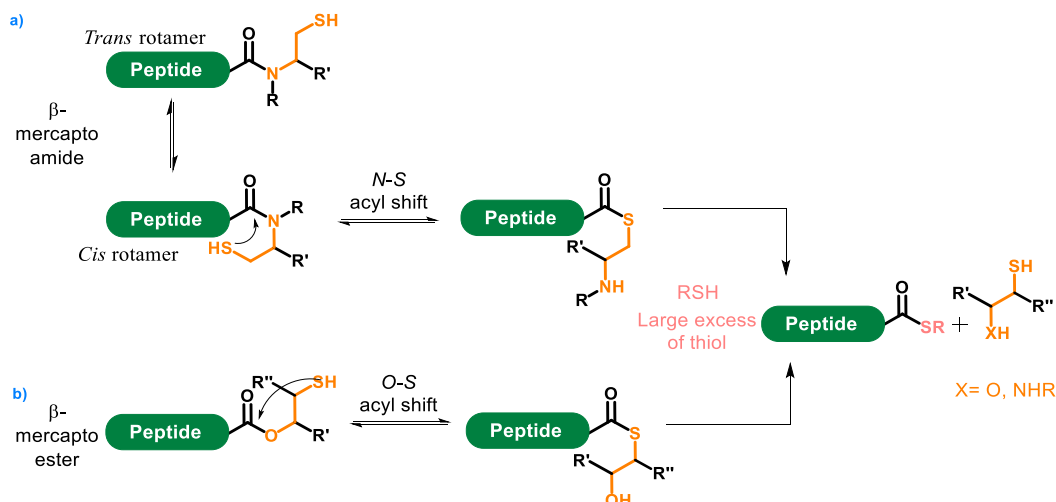
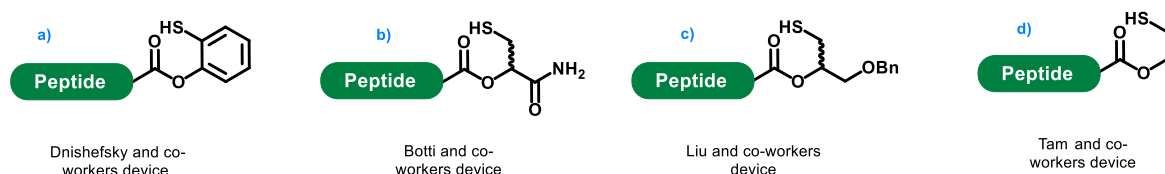


Figure 11: Principle of *N*-S and *O*-S acyl shift-based thioester precursors

a) *O*-S shift-based devices

The ability of peptidyl- β -mercapto-esters derivatives to rearrange into thioesters was exploited by Danishefsky [48] in 2004 when he introduced an *o*-mercapto-phenol ester-based intramolecular thioesterification device (fig. 12-a). In the same year, Botti and coworkers independently described a system based on a 2-hydroxy-3-mercaptopropionamide ester (Hmp) (fig. 12-b) β -mercapto-ester. [49] In both cases, the esters were able to rearrange spontaneously into thioesters under NCL conditions, but the reaction was accompanied by concomitant hydrolysis to form the resulting carboxylic acid. Moreover, the Hmp unit was found to be labile during standard Fmoc-SPPS. Few years later, Lei Liu and co-workers [50] described a device bearing an electron-releasing group $-\text{CH}_2\text{OBn}$ instead of the electron withdrawing amide of Botti's device (fig. 12-c). Liu's device was less prompt to hydrolysis during NCL than the previous ones, but still showed a significant amount of this side-reaction ($\sim 10\%$). A minimal device based on a β -mercaptoethanol ester was also reported by Tam and Eom [51] to prepare peptide thioesters *via* acid-catalyzed *O*-S shift (fig. 12-d). More recently, Tietze [52] revisited the Hmp device to introduce a hydrophilic tag and exploited it to facilitate the synthesis of a transmembrane protein segment (see chapter 4, bibliographic study focused on related solubility tags).

Figure 12: Intramolecular thioesterification devices based on *O*-*S* acyl shift

b) *N*-*S* shift-based devices

According to the protein chemical synthesis database (PCS) maintained by Melnyk and co-workers, [43] the use of *O*-*S* acyl shift-based systems declined after 2013 due to use of more convenient *N*-*S* acyl shift-based devices. The combination of advantageous aspects such as their relative ease of synthesis using Fmoc-SPPS, their high stability during SPPS, and lower susceptibility to hydrolysis and their ability to act as efficient acyl donors in both acidic and neutral conditions makes *N*-*S* acyl shift-based systems an appealing substitute to pre-formed thioesters, highly complementary to Dawson's Nbz and Liu's hydrazides peptide thioester precursors.

i- *N*-*S* shift systems requiring acidic conditions for the rearrangement

In 2005, Aimoto and his co-workers [53] developed the first method for an acid-induced intramolecular *N*-*S* acyl shift of a primary amide (i.e. not a secondary amide where two different acyl groups are branched on the same nitrogen such as electronically-activated acyl donors imides or *N*-acyl sulfonamides). This acyl shift is mediated by an auxiliary: The *N*-4,5-dimethoxy-2-mercapto-benzyl (Dmmb) group (fig. 13) and can be performed either in solution or on-resin. Note that this example is the only report involving a γ -mercapto ester device (and not β -). These *N*-*S* acyl shift properties were found as a side reaction that were further exploited in the above cited article, but never found further application to protein chemical synthesis. A major limitation of this device concerns the difficult coupling on the sterically-hindered *N*-Dmmb secondary amine required for the elongation of the peptide. Moreover, the requirement for acidic conditions to mediate the *N*-*S* acyl shift is another severe limitation: protonation of the resulting amine is needed to prevent the reverse *S*-*N* acyl shift, but acidic pH slows down trans-thioesterification, and is incompatible with one-pot NCL, as this reaction requires a near neutral pH.

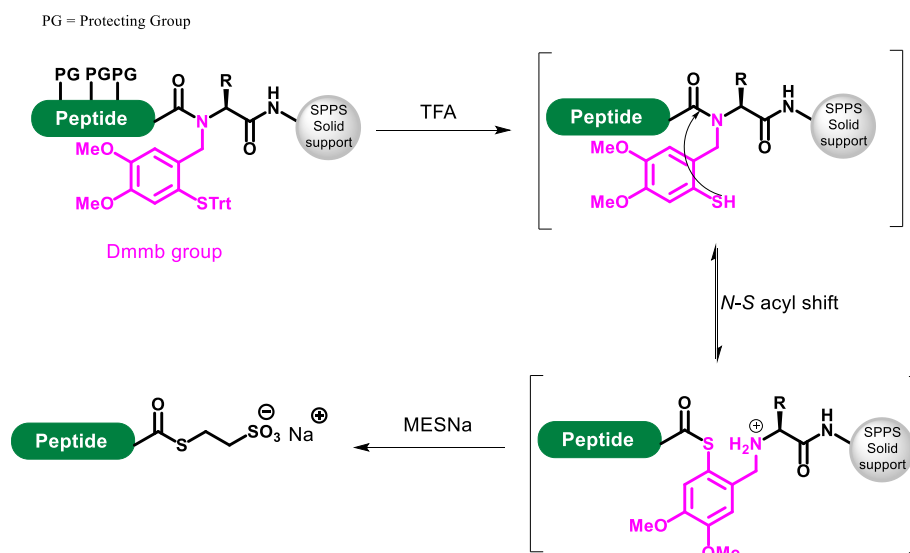


Figure 13: The Dmmb device

A year later, Nakahara [54] reported the synthesis of a peptide carrying a 5-mercaptomethyl proline derivative at its C-terminus. This peptide was easily converted to a 3-mercaptopropionic acid (MPA) thioester by treatment of the side-chain deprotected peptidyl resin with aqueous MPA under acidic conditions and microwave irradiation (fig. 14).

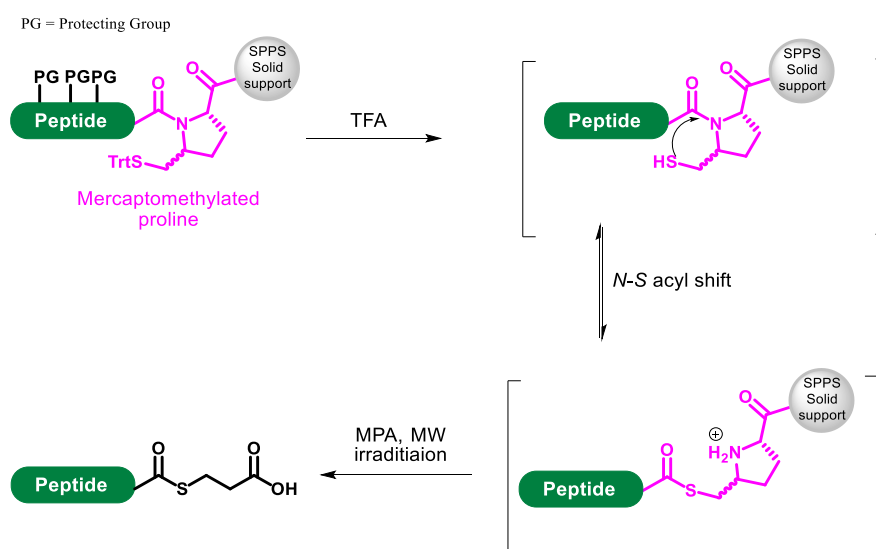


Figure 14: The mercaptomethylproline device

Another acidic pH-mediated *N*-S-acyl shift device was reported by Hojo, [55] who prepared peptides having an *N*-alkyl cysteine derivatives (fig. 15-a) as the C-terminus residue. These peptides were readily converted into peptide thioester in aqueous MPA.

Another illustration is the thioethylalkylamido (TEA) thioester surrogate (fig. 15-c) developed by Tam. [56] This *N*-S acyl shift device is cost-effective, conveniently prepared in one-step from commercially available starting materials.

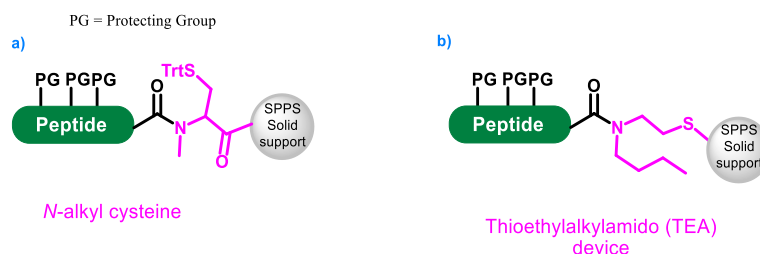


Figure 15: Other *N*-S shift-based devices operating under acidic conditions

It also important to mention the work of Otaka and his-co-workers [57] who reported *N*-S acyl-transfer-mediated synthesis of peptide thioesters using *N*-aminoacyl-*N*-sulfanylethyl-anilides (SEALide) (fig. 16). The *N*-aryl group acts as a better leaving group as compared to the above cited *N*-alkyl derivatives, but acylation of the *N*-alkyl aniline to introduce the first amino acid of the segment is notably difficult. The main limitation of this work was the epimerization of the C-terminal amino acid of the segments attached to this SEALide device during the acid-mediated *N*-S shift.

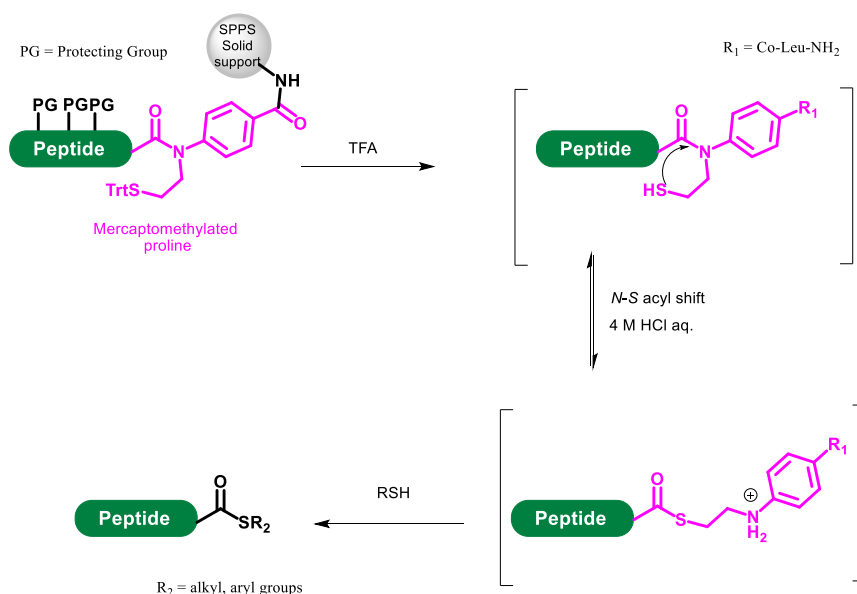


Figure 16: SEALides peptide thioester precursors

In a very elegant approach, Lei Liu and his-co-workers [58] reported the use of an enamide-based device (fig. 17-a). The final TFA cleavage step affords the targeted thioester functionality through an intramolecular *N*-S acyl shift that liberates an enamine, which

isomerize into an imine that is further hydrolyzed into a ketone, thus making the thioesterification process irreversible. Nevertheless, this method requires long exposures to TFA making it unsuitable for sensitive targets. Therefore, Zheng and his co-workers [59] reported a very similar enamide containing device where no long TFA treatment was needed (fig. 17-b).

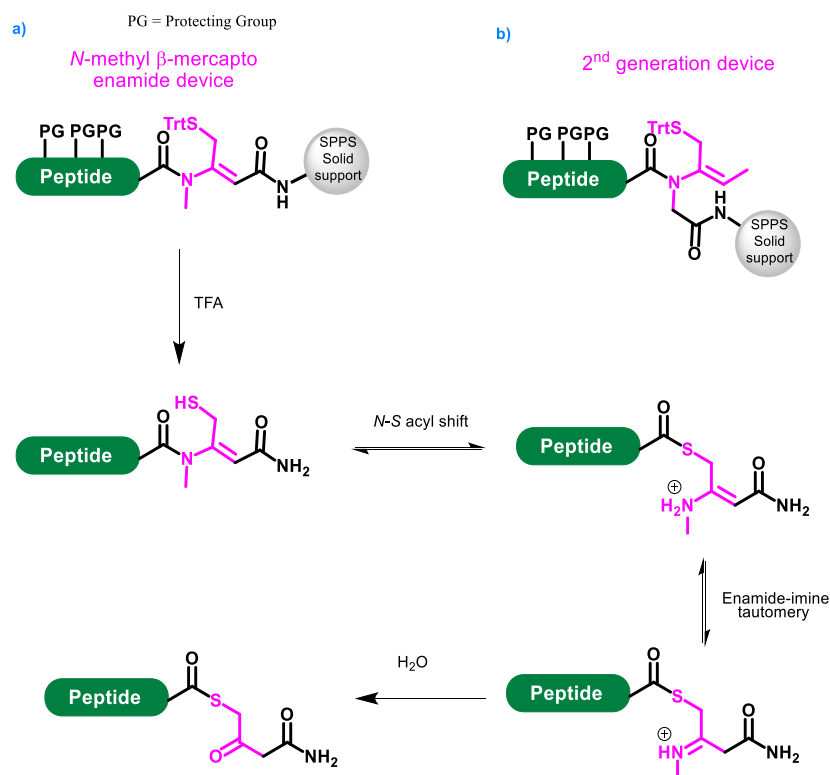


Figure 17: Enamide based devices

ii- *N*-S shift-based crypto-thioesters

Some methods relying on C-terminal β -mercapto amide devices that can operate under neutral conditions have been described, thus making possible one pot reactions (formation of the thioester function and NCL). Such segments have been called crypto-thioester peptides. This approach is particularly interesting because it minimizes the post-SPPS steps required before performing NCL.

In 2009, Kawakami and Aimoto reported peptides containing a cysteinyl prolyl ester (CPE) [60,61] moiety at the C-terminus (CPE peptides). These CPE peptides are spontaneously transformed into a diketopiperazine thioester *via* a reversible intramolecular *N*-S acyl shift followed by an irreversible *N*-O shift that constitutes the driving force of the reaction. This

diketopiperazine is readily transformed into a more reactive peptide thioester by trans-thioesterification with sodium mercaptoethanesulfonate (MESNa) (fig. 18).

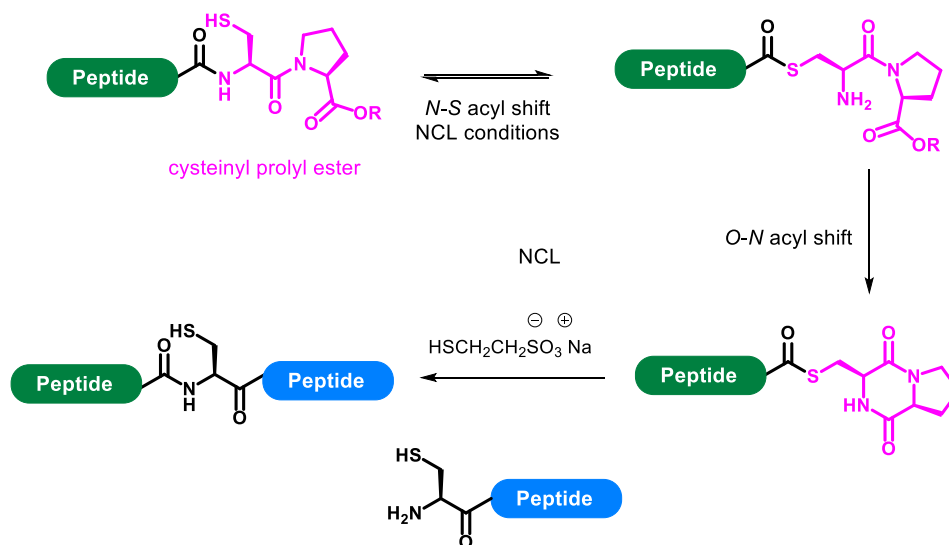


Figure 18: Thioester formation and one-pot ligation of CPE peptides

One year later, the Melnyk [62] (Fig. 19-a) and Chuan-Fa Liu groups [63] (Fig. 19-b) independently reported the design of a latent thioester surrogate based on the bis(2-sulfanylethyl) amino group (SEA^{on}). The SEA^{on} group is a tertiary amide bearing two thiol groups in β of the amide nitrogen. It spontaneously rearranges into a thioester conditions NCLat neutral to mildly acidic pH, followed by trans-thioesterification with mercaptophenylacetic acid (MPAA). The symmetry of the device makes that both amide bond rotamers are available for the nucleophilic attack of one thiol on the carbonyl, and the second thiol has been predicted to act as an intramolecular proton donor for the amine leaving group in a concerted mechanism [64] thus explaining the fast acyl shift kinetics under neutral conditions.

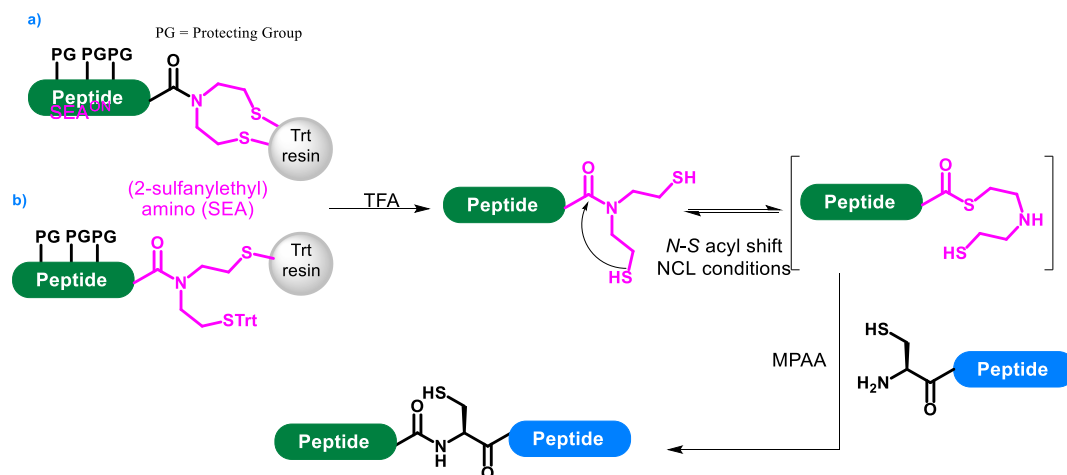
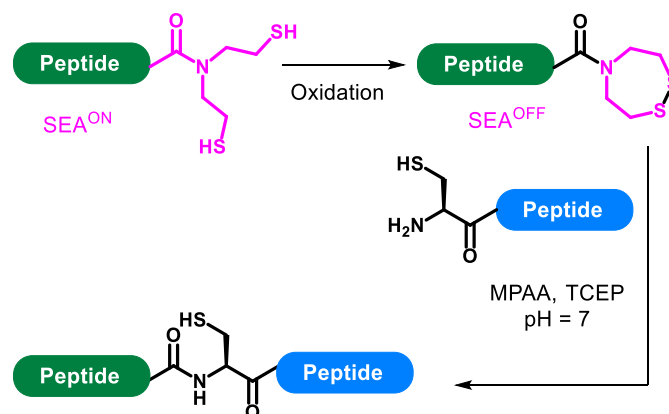


Figure 19: SEA crypto-thioesters

Advantageously, the crypto-thioester properties of SEA^{ON} can be masked as a stable SEA^{OFF}, through oxidation to form an intramolecular disulfide bond. SEA^{OFF} is converted on demand into an active thioester by treatment with a combination of the reducing agent *tris*-carboxyethylphosphine (TCEP) and MPAA as the exogenous thiol (Fig. 20).

Figure 20: SEA^{ON} and SEA^{OFF} devices

Note that Melnyk and his co-workers [65] have used the ability to mask the SEA device crypto-thioester properties for a one-pot three segments ligation strategy. The first ligation between a pre-formed thioester segment and a cysteinyl-peptide having a C-terminal SEA^{OFF} motif is carried out under strictly deoxygenated conditions and in absence of strong reductants like TCEP. Then, the second ligation is carried out under classical NCL condition, in the presence of TCEP, between the peptide obtained by the first ligation and a cysteinyl-peptide (fig. 21).

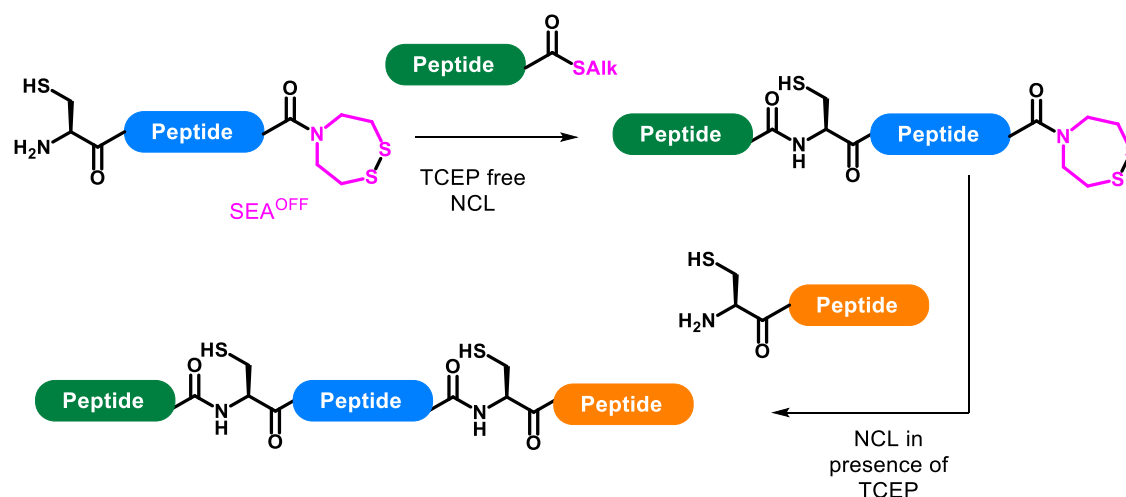


Figure 21: The use of the SEA device for three segments one-pot ligations

Building on their first work on the SEALide device, [57] Otaka and his co-workers [66] examined the utility of SEALide peptides as a crypto-thioesters and found out that the SEALide moiety could work as a crypto-thioester under neutral NCL conditions to afford the corresponding NCL product. Therefore, they solved the epimerization problem encountered with the device when thioesterification was performed under acidic conditions.

It should be pointed out that the SEALide system does not bear an internal acidic function capable of assisting thioester formation, therefore, the rearrangement needs phosphate salts to proceed, through a probable proton jump mechanism (fig. 22). These interesting features enable one-pot three segments “kinetically controlled ligations”, if performing the first NCL in absence of phosphate salts, and the second upon addition of a phosphate buffer.

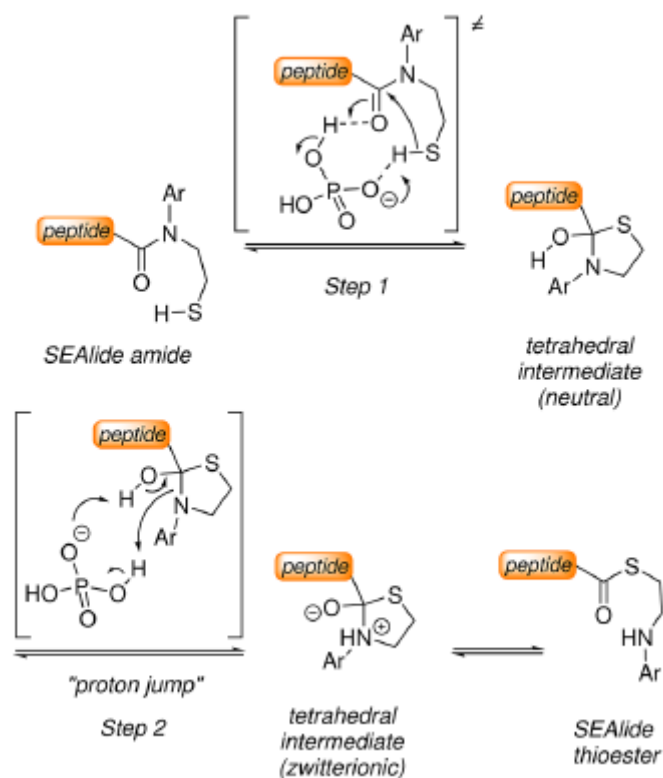


Figure 22: Proposed mechanism for the rearrangement of SEALide by phosphate salts (figure taken from ref. [67])

Chuan-Fa Liu [68] reported the use of peptides carrying a C-terminal 2-pyrrolidinemethanethiol (PMT) (fig. 23) as crypto-thioesters that were shown to ligate efficiently with cysteinyl-peptides at slightly acidic pH (5-6). However, this ligation requires very high peptide concentrations (15 to 35 mM) to proceed to completion.

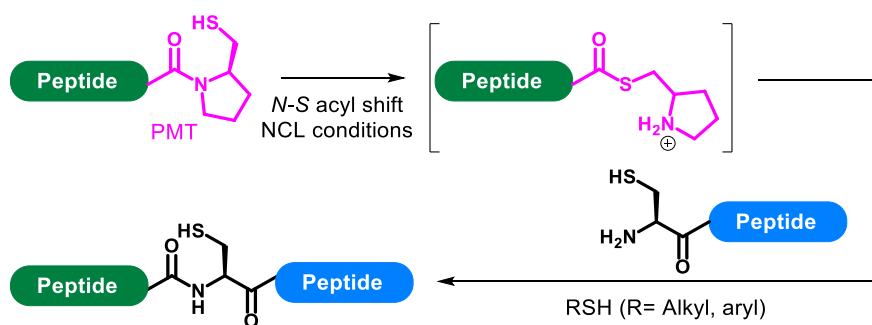


Figure 23: PMT cryptothioester device

In another example of crypto-thioesters, Offer and his co-workers [69] described the use of α -methylcysteines for the *in situ* thioester formation (fig. 24). Note that the driving force of the formation of thioester with this device could be explained by the formation of a 5-membered

ring intermediate favored by the presence of gem-disubstituents, probably through the Thorpe Ingold effect. Nevertheless, ligation requires high MPAA concentration and reaction kinetics are slow.

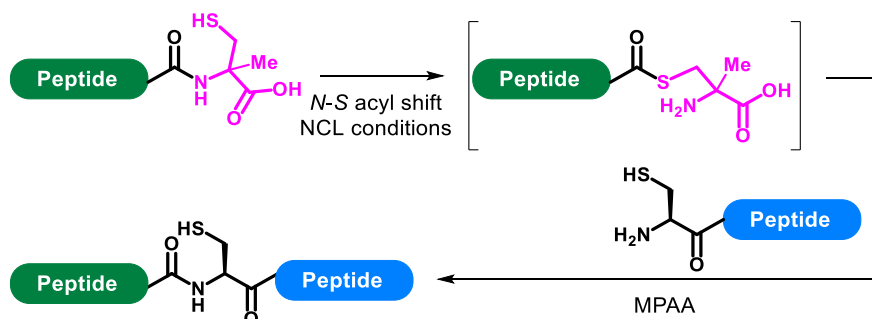


Figure 24: Methylcysteines for the *in situ* thioester formation

Another illustration is the work of Hojo, [70] who used peptides having the C-terminal *N*-alkylcysteine (NAC) with a free carboxy group as crypto-thioesters (fig. 25), for NCL reactions under mild acidic condition (pH ~5). It was shown that C-terminal acid cysteine leads to faster NCLs than the amide counterpart, probably due to intramolecular acid-base catalysis. However again, reaction kinetics are very slow.

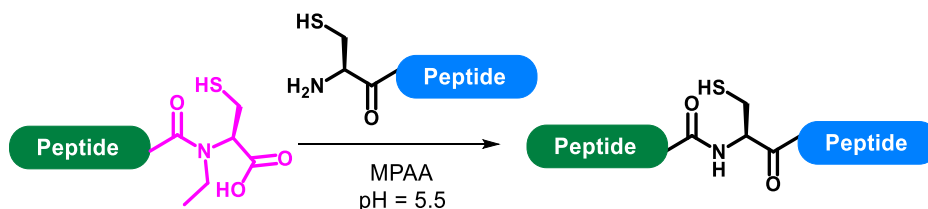
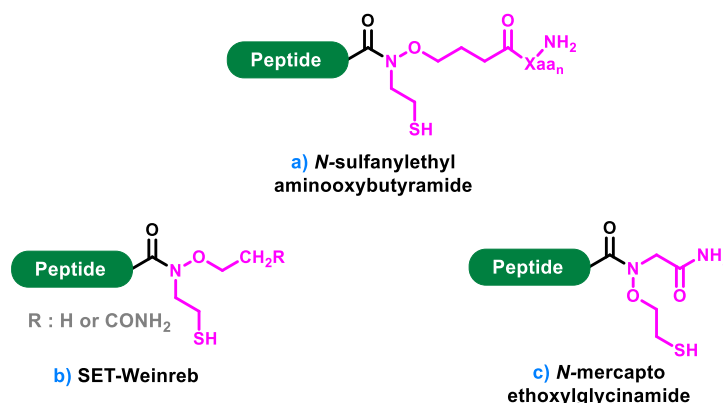


Figure 25: NAC device

Besides standard amides, *N*-alkoxy derivatives have also been reported in the past few years taking advantage of the fact that hydroxylamine ethers are better leaving group than amines. For example, we can mention the use of *N*-sulfanylethylaminooxybutyramides (SEAoxy) by Yoshiya (fig. 26-a), [71] the development of very similar “sulfanylethyl Weinreb amide derivatives” (fig. 26-b), by Chuan-Fa Liu, [72] and the more recent introduction by Chatterjee [73] of *N*-mercaptoethoxyglycinamide (MEGA) (fig. 26-c).

Figure 26: *N*-alkoxy amide-based crypto-thioesters

C. *N*-(2-hydroxybenzyl)cysteine (*N*-Hnb-Cys) thioesterification device

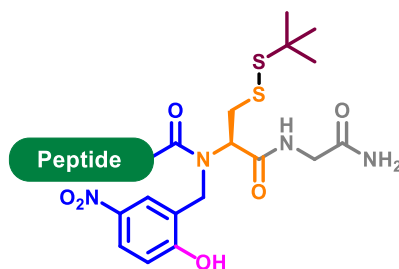
a) Limitation of other *N*-S shift based crypto-thioesters

As mentioned before, numerous methods for the synthesis of peptide thioesters by Fmoc-SPPS has been investigated, but none of these methods is yet used universally due to several drawbacks associated with the different strategies. The strategy based on thioester formation through *in situ* *N*-S acyl shift under NCL conditions seems promising, nevertheless, the existing devices are often limited by:

- The complex synthesis of the thioesterification device
- The difficult coupling of the C-terminal amino acid on the hindered secondary amine of many of these devices
- The instability of the peptides carrying these devices in aqueous medium (e. g. during purification and storage), due to an early *N*-S acyl transfer followed by hydrolysis
- The slow kinetics of these devices in NCL as compared to a preformed-thioester, Liu's hydrazides or Dawson's Nbz.

b) Design of the *N*-Hnb-Cys thioesterification device

Therefore, our group recently introduced a straightforward, inexpensive and fully automated strategy for the Fmoc-SPPS synthesis of peptide crypto-thioesters [74,75] based on an *N*-(2-hydroxybenzyl)cysteine (*N*-Hnb-Cys) structure (fig. 27).

Figure 27: *N*-Hnb-Cys thioesterification device

The design of this thioesterification device was based on *N*-alkyl cysteines that however show extremely slow kinetics under NCL conditions. [55] To accelerate *N*-S acyl shift at neutral pH, our group took inspiration from the mechanism of *in vivo* protein splicing promoted by class 1 inteins. [76] The first step in protein splicing indeed consists of a self-catalyzed intramolecular *N*-S acyl transfer. The mechanism involves stabilization of the oxyanion of the tetrahedral intermediate through hydrogen bonding (fig. 28, green residues), as well as intramolecular *N*-protonation by a highly conserved histidine residue to induce N-C bond cleavage and form the thioester (fig. 28, red).

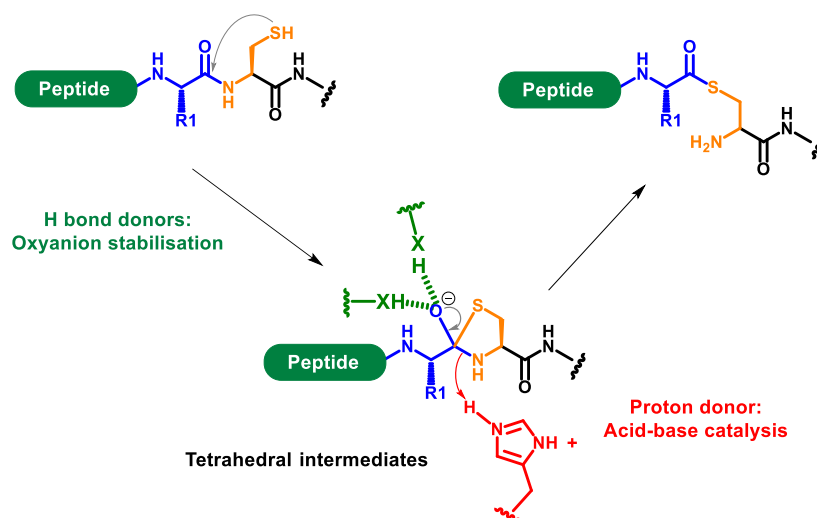


Figure 28: Inteins mechanism

Therefore, the *N*-alkyl group of the cysteine was functionalized by a proton and hydrogen-bond donor group in order to try to mimic both the histidine side chain and oxyanion-stabilizing residues. The key to this design is a phenol group with a pKa close to neutral, as this would result in a low energy cost for an intramolecular proton transfer at pH 7. Phenol groups are also excellent hydrogen bond donors as compared to alcohols or amides. [77]

Gratifyingly, the presence of the phenol group was found to be critical for achieving fast ligation rates, since an *O*-methylated version [74] shows a massive decrease in NCL kinetics (~ 60 fold), that becomes comparable to *N*-alkyl cysteine and other previously reported *N*-S shift-based crypto-thioesters.

According to recent density-functional theory (DFT) computational studies, [78,79] the para-nitrophenol group was found to play three different roles: it activates the amide by hydrogen bond donation to the carbonyl oxygen, followed by concerted thiolate attack and oxyanion protonation by the phenol, and stabilizes of the tetrahedral intermediate by a phenoxide/hydroxyl hydrogen bond (fig. 29).

If the phenol group thus apparently indeed acts as both hydrogen bond and proton donor, the underlying mechanism is thus quite different from intein self-catalysis that served as inspiration.

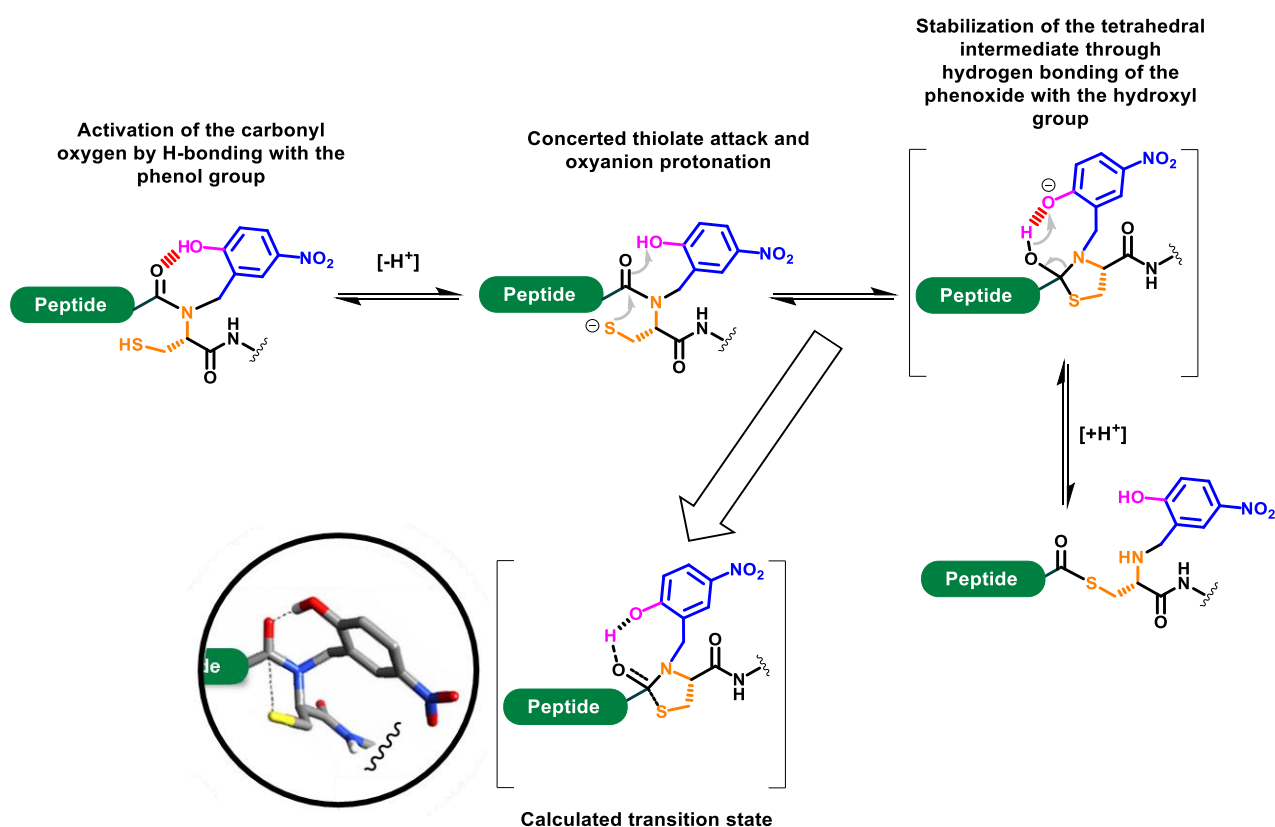


Figure 29: Mechanism of *N*-S acyl transfer of *N*-(2-hydroxybenzyl) cysteine derivatives according to DFT calculations in water

The hydroxyl group also plays another crucial role, during the SPPS: it behaves as an acyl transfer auxiliary which greatly facilitates the *N*-acylation of the sterically-hindered secondary amine *via* *O*-acylation followed by an intramolecular *O*-*N* shift (fig. 30). [80] This mechanism is reminiscent from the work of Sheppard and co-workers who introduced the 2-hydroxy-4-methoxybenzyl (Hmb) backbone amide protecting group [81] which usefulness has been demonstrated by the Fmoc syntheses of numerous difficult peptides, and shows considerably facilitated *N*-acylation as compared to the parent 2,4-dimethoxybenzyl group (Dmb), for which the preparation of a dipeptide building block in solution is preferred (Fmoc Xaa1-(Dmb)Xaa2-OH).

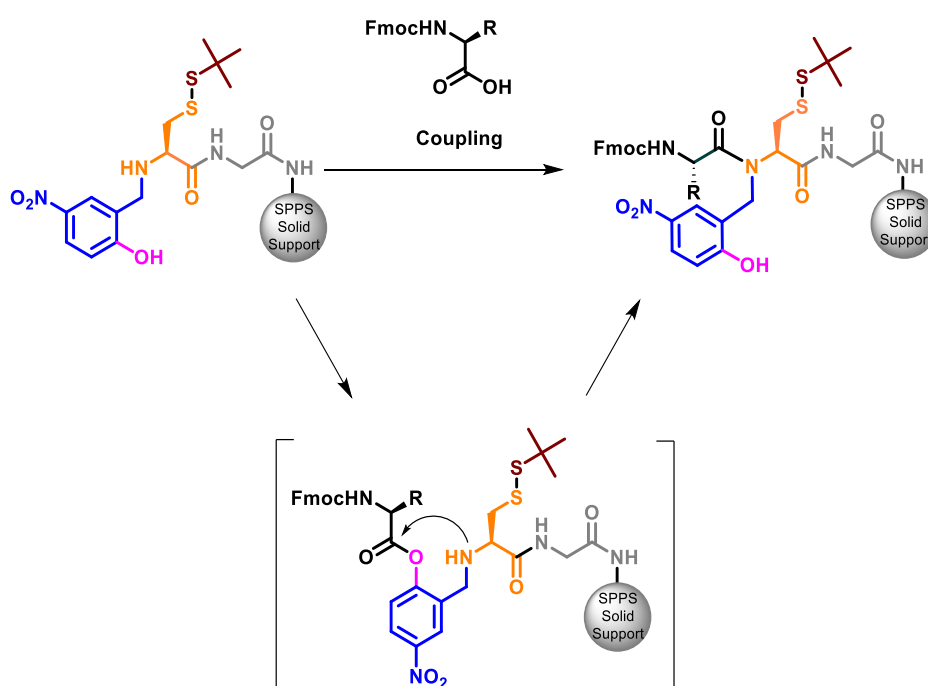


Figure 30: Assistance of the hydroxyl group for secondary amine acylation

Another key feature of the device is the protection of the cysteine thiol group with a *S*tBu disulfide stable to Fmoc-SPPS, in order to prevent premature spontaneous *N*-*S* acyl shift after cleavage, during purification and handling. The *S*-*S*tBu group is readily removed under NCL conditions, which typically include MPAA and TCEP, thus unleashing the crypto-thioesters properties.

A glycine spacer was added between the cysteine and the solid support (fig. 31) to prevent a variety of side reactions such as Cys amide hydrolysis during TFA cleavage when a Rink amide

linker was used, [82,83] as well as Cys epimerization and β -elimination during SPPS in the case of Cys esters. [84]

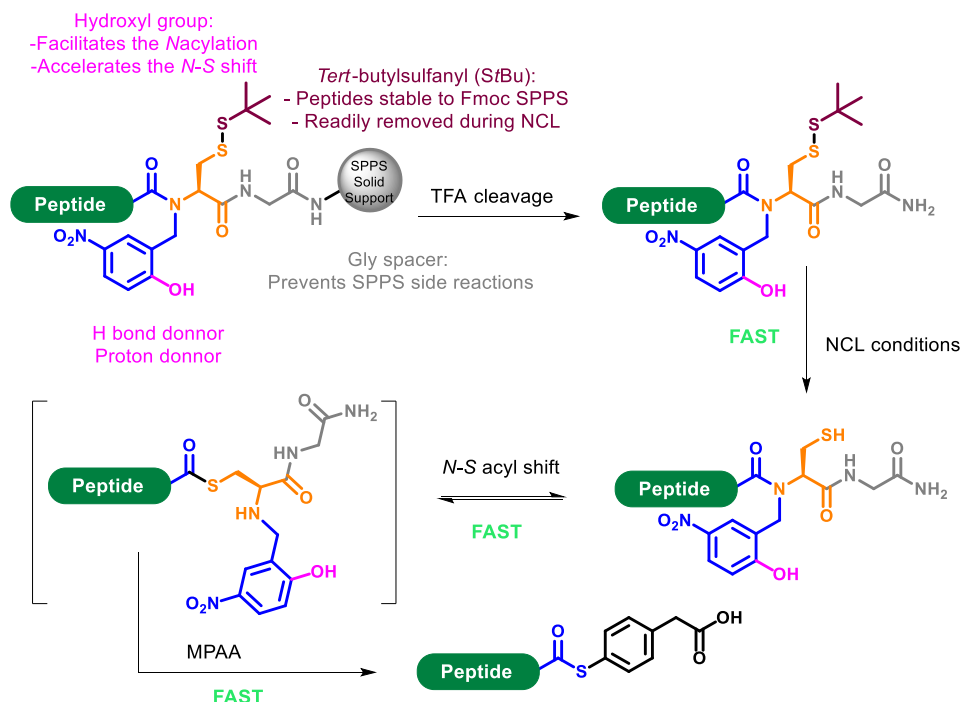


Figure 31: C-terminal *N*-(2-hydroxy-4-nitrobenzyl) cysteine thioesterification device

II. Aim of this work: an optimized protocol for the synthesis of *N*-Hnb-Cys peptide crypto-thioesters

The *N*-Hnb-Cys(StBu) device was used in our team [73,74,79,80] and others [81] for the synthesis of numerous proteins without any issue. However, when it was used in the course of this PhD thesis, we observed the formation of a large proportion (up to 20%) of a co-product showing the same molecular weight as the desired compound. The main difference with previous reports was the use of ChemMatrix instead of Tentagel as solid support.

Therefore, we undertook a systematic study to characterize this co-product and identify the parameters favoring its formation. We discovered that epimerization from L- to D-cysteine arise during the solid-supported reductive amination step from a tautomeric equilibrium between the aldimine and a rearranged achiral ketimine forms [88] (Fig. 32) prior to reduction into amine.

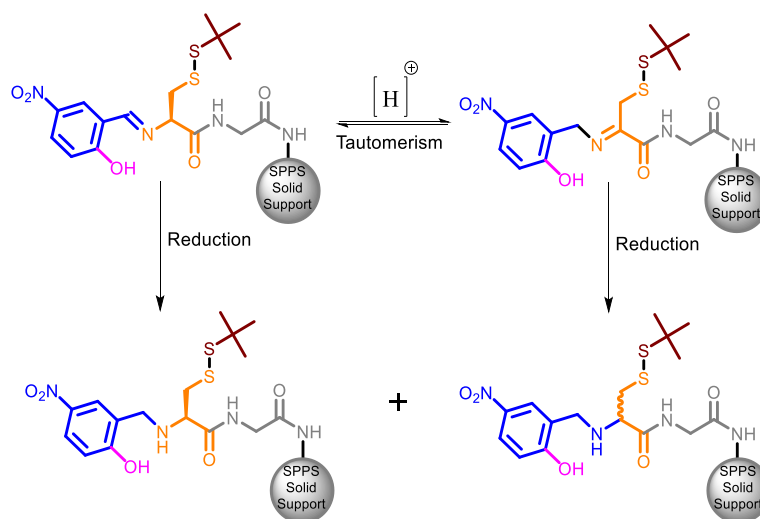


Figure 32: Putative epimerization mechanism: Imine tautomerism

We also found out that this side-reaction was very sensitive to temperature, dilution, reaction time and nature of the solid support. Meanwhile during our optimization process, we noticed the formation of other side products which mass corresponds to $[M-2]$ of the desired product. A thorough study allowed us to identify these side products as imidazolidinones formed from non-reduced imines. [89,90]

Gratifyingly, the understanding of the mechanism leading to these side reactions allowed us to optimize a protocol that abolished the formation of all side products. These results are presented in the following article published in *Organic and Biomolecular Chemistry*.

III. References for the bibliographic study part

1. Dawson, P., Muir, T., Clark-Lewis, I., and Kent, S. (1994) Synthesis of proteins by native chemical ligation. *Science*, **266** (5186), 776–779.
2. Wieland, T., Bokelmann, E., Bauer, L., Lang, H.U., and Lau, H. (1953) Über Peptidsynthesen. 8. Mitteilung Bildung von S-haltigen Peptiden durch intramolekulare Wanderung von Aminoacylresten. *Justus Liebigs Ann. Chem.*, **583** (1), 129–149.
3. Kemp, D.S., and Galakatos, N.G. (1986) Peptide synthesis by prior thiol capture. 1. A convenient synthesis of 4-hydroxy-6-mercaptodibenzofuran and novel solid-phase synthesis of peptide-derived 4-(acyloxy)-6-mercaptodibenzofurans. *J. Org. Chem.*, **51** (10), 1821–1829.
4. Kemp, D.S., Galakatos, N.G., Bowen, B., and Tan, K. (1986) Peptide synthesis by prior thiol capture. 2. Design of templates for intramolecular *O,N*-acyl transfer. 4,6-Disubstituted dibenzofurans as optimal spacing elements. *J. Org. Chem.*, **51** (10), 1829–1838.
5. Hojo, H., and Aimoto, S. (1991) Polypeptide Synthesis Using the *S*-Alkyl Thioester of a Partially Protected Peptide Segment. Synthesis of the DNA-Binding Domain of *c*-Myb Protein (142–193)–NH₂. *BCSJ*, **64** (1), 111–117.
6. Raz, R., Burlina, F., Ismail, M., Downward, J., Li, J., Smerdon, S.J., Quibell, M., White, P.D., and Offer, J. (2016) HF-Free Boc Synthesis of Peptide Thioesters for Ligation and Cyclization. *Angew. Chem. Int. Ed.*, **55** (42), 13174–13179.
7. Bang, D., Pentelute, B.L., Gates, Z.P., and Kent, S.B. (2006) Direct On-Resin Synthesis of Peptide- α Thiophenylesters for Use in Native Chemical Ligation. *Org. Lett.*, **8** (6), 1049–1052.
8. Ghosh, N. (2004) DBU (1,8-diazabicyclo[5.4.0]undec-7-ene) - A Nucleophilic Base. *Synlett*, (3), 574–575.
9. Bu, X., Xie, G., Law, C.W., and Guo, Z. (2002) An improved deblocking agent for direct Fmoc solid-phase synthesis of peptide thioesters. *Tetrahedron Lett.*, **43** (13), 2419–2422.
10. Clippingdale, A.B., Barrow, C.J., and Wade, J.D. (2000) Peptide thioester preparation by Fmoc solid phase peptide synthesis for use in native chemical ligation. *J. Pept. Sci.*, **6** (5), 225–234.
11. Li, X., Kawakami, T., and Aimoto, S. (1998) Direct preparation of peptide thioesters using an Fmoc solid-phase method. *Tetrahedron Lett.*, **39** (47), 8669–8672.

12. Hasegawa, K., Sha, Y.L., Bang, J.K., Kawakami, T., Akaji, K., and Aimoto, S. (2001) Preparation of phosphopeptide thioesters by Fmoc-and Fmoc(2-F)-solid phase synthesis. *Lett Pept Sci*, **8** (3–5), 277–284.
13. von Eggelkraut-Gottanka, R., Klose, A., Beck-Sickinger, A.G., and Beyermann, M. (2003) Peptide α thioester formation using standard Fmoc-chemistry. *Tetrahedron Lett.*, **44** (17), 3551–3554.
14. Kitagawa, K., Adachi, H., Sekigawa, Y., Yagami, T., Futaki, S., Gu, Y.J., and Inoue, K. (2004) Total chemical synthesis of large CCK isoforms using a thioester segment condensation approach. *Tetrahedron*, **60** (4), 907–918.
15. Nakamura, K., Hanai, N., Kanno, M., Kobayashi, A., Ohnishi, Y., Ito, Y., and Nakahara, Y. (1999) Design and synthesis of silyl ether-based linker for solid-phase synthesis of glycopeptides. *Tetrahedron Letters*, **40** (3), 515–518.
16. Li, L., and Wang, P. (2007) Transfer allyl esters to thioesters in solid phase condition: synthesis of peptide thioesters by Fmoc chemistry. *Tetrahedron Letters*, **48** (1), 29–32.
17. Ficht, S., Payne, R.J., Guy, R.T., and Wong, C.-H. (2008) Solid-Phase Synthesis of Peptide and Glycopeptide Thioesters through Side-Chain-Anchoring Strategies. *Chem. Eur. J.*, **14** (12), 3620–3629.
18. Lelièvre, D., Barta, P., Aucagne, V., and Delmas, A.F. (2008) Preparation of peptide thioesters using Fmoc strategy through hydroxyl side chain anchoring. *Tetrahedron Letters*, **49** (25), 4016–4019.
19. Alsina, J., Yokum, T.S., Albericio, F., and Barany, G. (1999) Backbone Amide Linker (BAL) Strategy for *N* ^{α} -9-Fluorenylmethoxycarbonyl (Fmoc) Solid-Phase Synthesis of Unprotected Peptide *p*-Nitroanilides and Thioesters. *J. Org. Chem.*, **64** (24), 8761–8769.
20. Gross, C.M., Lelievre, D., Woodward, C.K., and Barany, G. (2005) Preparation of protected peptidyl thioester intermediates for native chemical ligation by N α -9-fluorenylmethoxycarbonyl (Fmoc) chemistry: considerations of side-chain and backbone anchoring strategies, and compatible protection for N-terminal cysteine. *J Pept Res*, **65** (3), 395–410.
21. Kenner, G.W., McDermott, J.R., and Sheppard, R.C. (1971) The safety catch principle in solid phase peptide synthesis. *J. Chem. Soc. D*, (12), 636.
22. Shin, Y., Winans, K.A., Backes, B.J., Kent, S.B.H., Ellman, J.A., and Bertozzi, C.R. (1999) Fmoc-Based Synthesis of Peptide- α Thioesters: Application to the Total Chemical

- Synthesis of a Glycoprotein by Native Chemical Ligation. *J. Am. Chem. Soc.*, **121** (50), 11684–11689.
23. Ingenito, R., Bianchi, E., Fattori, D., and Pessi, A. (1999) Solid Phase Synthesis of Peptide C-Terminal Thioesters by Fmoc/ *t*-Bu Chemistry. *J. Am. Chem. Soc.*, **121** (49), 11369–11374.
24. Ollivier, N., Behr, J.-B., El-Mahdi, O., Blanpain, A., and Melnyk, O. (2005) Fmoc Solid-Phase Synthesis of Peptide Thioesters Using an Intramolecular *N*, *S*-Acyl Shift. *Org. Lett.*, **7** (13), 2647–2650.
25. Burlina, F., Morris, C., Behrendt, R., White, P., and Offer, J. (2012) Simplifying native chemical ligation with an *N*-acylsulfonamide linker. *Chem. Commun.*, **48** (20), 2579.
26. Blanco-Canosa, J.B., and Dawson, P.E. (2008) An Efficient Fmoc-SPPS Approach for the Generation of Thioester Peptide Precursors for Use in Native Chemical Ligation. *Angew. Chem. Int. Ed.*, **47** (36), 6851–6855.
27. Blanco-Canosa, J.B., Nardone, B., Albericio, F., and Dawson, P.E. (2015) Chemical Protein Synthesis Using a Second-Generation *N*-Acylurea Linker for the Preparation of Peptide-Thioester Precursors. *J. Am. Chem. Soc.*, **137** (22), 7197–7209.
28. Mahto, S.K., Howard, C.J., Shimko, J.C., and Ottesen, J.J. (2011) A reversible protection strategy to improve Fmoc-SPPS of peptide thioesters by the *N*-Acylurea approach. *Chembiochem*, **12** (16), 2488–2494.
29. Weidmann, J., Dimitrijević, E., Hoheisel, J.D., and Dawson, P.E. (2016) Boc-SPPS: Compatible Linker for the Synthesis of Peptide *o*-Aminoanilides. *Org. Lett.*, **18** (2), 164–167.
30. Pascal, R., Chauvey, D., and Sola, R. (1994) Carboxyl-protecting groups convertible into activating groups. Carbamates of *o*-aminoanilides are precursors of reactive *N*-acylureas. *Tetrahedron Lett.*, **35** (34), 6291–6294.
31. Palà-Pujadas, J., Albericio, F., and Blanco-Canosa, J.B. (2018) Peptide Ligations by Using Aryloxycarbonyl- *o*-methylaminoanilides: Chemical Synthesis of Palmitoylated Sonic Hedgehog. *Angew. Chem. Int. Ed.*, **57** (49), 16120–16125.
32. Tofteng, A.P., Sørensen, K.K., Conde-Frieboes, K.W., Hoeg-Jensen, T., and Jensen, K.J. (2009) Fmoc Solid-Phase Synthesis of C-Terminal Peptide Thioesters by Formation of a Backbone Pyroglutamyl Imide Moiety. *Angew. Chem. Int. Ed.*, **48** (40), 7411–7414.

33. Elashal, H.E., Sim, Y.E., and Raj, M. (2017) Serine promoted synthesis of peptide thioester-precursor on solid support for native chemical ligation. *Chem. Sci.*, **8** (1), 117–123.
34. Wolman, Y., Gallop, P.M., and Patchornik, A. (1961) Peptide synthesis via oxidation of hydrazines. *J. Am. Chem. Soc.*, **83** (5), 1263–1264.
35. Camarero, J.A., Hackel, B.J., de Yoreo, J.J., and Mitchell, A.R. (2004) Fmoc-Based Synthesis of Peptide α -Thioesters Using an Aryl Hydrazine Support. *J. Org. Chem.*, **69** (12), 4145–4151.
36. Fang, G.-M., Li, Y.-M., Shen, F., Huang, Y.-C., Li, J.-B., Lin, Y., Cui, H.-K., and Liu, L. (2011) Protein Chemical Synthesis by Ligation of Peptide Hydrazides. *Angew. Chem. Int. Ed.*, **50** (33), 7645–7649.
37. Tian, X., Li, J., and Huang, W. (2016) Optimal peptide hydrazide ligation with C-terminus Asp, Asn, and Gln hydrazides. *Tetrahedron Lett.*, **57** (38), 4264–4267.
38. Siman, P., Karthikeyan, S.V., Nikolov, M., Fischle, W., and Brik, A. (2013) Convergent Chemical Synthesis of Histone H2B Protein for the Site-Specific Ubiquitination at Lys34. *Angew. Chem. Int. Ed.*, **52** (31), 8059–8063.
39. Thompson, R.E., Liu, X., Alonso-García, N., Pereira, P.J.B., Jolliffe, K.A., and Payne, R.J. (2014) Trifluoroethanethiol: An Additive for Efficient One-Pot Peptide Ligation–Desulfurization Chemistry. *J. Am. Chem. Soc.*, **136** (23), 8161–8164.
40. Huang, Y.-C., Chen, C.-C., Li, S.-J., Gao, S., Shi, J., and Li, Y.-M. (2014) Facile synthesis of C-terminal peptide hydrazide and thioester of NY-ESO-1 (A39-A68) from an Fmoc-hydrazine 2-chlorotrityl chloride resin. *Tetrahedron*, **70** (18), 2951–2955.
41. Zheng, J.-S., Tang, S., Qi, Y.-K., Wang, Z.-P., and Liu, L. (2013) Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat Protoc*, **8** (12), 2483–2495.
42. Zheng, J.-S., Tang, S., Guo, Y., Chang, H.-N., and Liu, L. (2012) Synthesis of Cyclic Peptides and Cyclic Proteins via Ligation of Peptide Hydrazides. *ChemBioChem*, **13** (4), 542–546.
43. Agouridas, V., El Mahdi, O., Cargoët, M., and Melnyk, O. (2017) A statistical view of protein chemical synthesis using NCL and extended methodologies. *Bioorg. Med. Chem.*, **25** (18), 4938–4945.
44. Kawakami, T. (2014) Peptide Thioester Formation via an Intramolecular N to S Acyl Shift for Peptide Ligation, in *Protein Ligation and Total Synthesis I*, vol. 362, Springer International Publishing, Cham, pp. 107–135.

45. Kang, J., and Macmillan, D. (2010) Peptide and protein thioester synthesis via N→S acyl transfer. *Org. Biomol. Chem.*, **8** (9), 1993.
46. Wang, C., and Guo, Q.-X. (2012) Theoretical study on formation of thioesters via O-to-S acyl transfer. *Sci. China Chem.*, **55** (10), 2075–2080.
47. Kim, B. (2016) The C-Terminal O-S Acyl Shift Pathway under Acidic Condition to Propose Peptide-Thioesters. *Molecules*, **21** (11), 1559.
48. Warren, J.D., Miller, J.S., Keding, S.J., and Danishefsky, S.J. (2004) Toward Fully Synthetic Glycoproteins by Ultimately Convergent Routes: A Solution to a Long-Standing Problem. *J. Am. Chem. Soc.*, **126** (21), 6576–6578.
49. Botti, P., Villain, M., Manganiello, S., and Gaertner, H. (2004) Native Chemical Ligation through in Situ O to S Acyl Shift. *Org. Lett.*, **6** (26), 4861–4864.
50. Zheng, J.-S., Cui, H.-K., Fang, G.-M., Xi, W.-X., and Liu, L. (2010) Chemical Protein Synthesis by Kinetically Controlled Ligation of Peptide O-Esters. *ChemBioChem*, **11** (4), 511–515.
51. Eom, K.D., and Tam, J.P. (2011) Acid-Catalyzed Tandem Thiol Switch for Preparing Peptide Thioesters from Mercaptoethyl Esters. *Org. Lett.*, **13** (10), 2610–2613.
52. Baumruck, A.C., Tietze, D., Steinacker, L.K., and Tietze, A.A. (2018) Chemical synthesis of membrane proteins: a model study on the influenza virus B proton channel. *Chem. Sci.*, **9** (8), 2365–2375.
53. Kawakami, T., Sumida, M., Nakamura, K., Vorherr, T., and Aimoto, S. (2005) Peptide thioester preparation based on an N-S acyl shift reaction mediated by a thiol ligation auxiliary. *Tetrahedron Letters*, **46** (50), 8805–8807.
54. Nagaike, F., Onuma, Y., Kanazawa, C., Hojo, H., Ueki, A., Nakahara, Y., and Nakahara, Y. (2006) Efficient Microwave-Assisted Tandem *N*- to S -Acyl Transfer and Thioester Exchange for the Preparation of a Glycosylated Peptide Thioester. *Org. Lett.*, **8** (20), 4465–4468.
55. Hojo, H., Onuma, Y., Akimoto, Y., Nakahara, Y., and Nakahara, Y. (2007) *N*-Alkyl cysteine-assisted thioesterification of peptides. *Tetrahedron Lett.*, **48** (1), 25–28.
56. Taichi, M., Hemu, X., Qiu, Y., and Tam, J.P. (2013) A Thioethylalkylamido (TEA) Thioester Surrogate in the Synthesis of a Cyclic Peptide via a Tandem Acyl Shift. *Org. Lett.*, **15** (11), 2620–2623.

57. Tsuda, S., Shigenaga, A., Bando, K., and Otaka, A. (2009) *N* → *S* Acyl-Transfer-Mediated Synthesis of Peptide Thioesters Using Anilide Derivatives. *Org. Lett.*, **11** (4), 823–826.
58. Zheng, J.-S., Chang, H.-N., Wang, F.-L., and Liu, L. (2011) Fmoc Synthesis of Peptide Thioesters without Post-Chain-Assembly Manipulation. *J. Am. Chem. Soc.*, **133** (29), 11080–11083.
59. Zheng, J.-S., Chen, X., Tang, S., Chang, H.-N., Wang, F.-L., and Zuo, C. (2014) A New Method for Synthesis of Peptide Thioesters via Irreversible *N*-to-*S* Acyl Transfer. *Org. Lett.*, **16** (18), 4908–4911.
60. Kawakami, T., and Aimoto, S. (2009) The use of a cysteinyl prolyl ester (CPE) autoactivating unit in peptide ligation reactions. *Tetrahedron*, **65** (19), 3871–3877.
61. Kawakami, T., Shimizu, S., and Aimoto, S. (2010) Peptide Thioester Formation by an *N* to *S* Acyl Shift Reaction at the Cysteinyl Prolyl Cysteine Position. *BCSJ*, **83** (5), 570–574.
62. Ollivier, N., Dheur, J., Mhidia, R., Blanpain, A., and Melnyk, O. (2010) Bis(2-sulfanylethyl)amino Native Peptide Ligation. *Org. Lett.*, **12** (22), 5238–5241.
63. Hou, W., Zhang, X., Li, F., and Liu, C.-F. (2011) Peptidyl *N*, *N*-Bis(2-mercaptoethyl)-amides as Thioester Precursors for Native Chemical Ligation. *Org. Lett.*, **13** (3), 386–389.
64. Raibaut, L., Cargoët, M., Ollivier, N., Chang, Y.M., Drobecq, H., Boll, E., Desmet, R., Monbaliu, J.-C.M., and Melnyk, O. (2016) Accelerating chemoselective peptide bond formation using bis(2-selenylethyl)amido peptide selenoester surrogates. *Chem. Sci.*, **7** (4), 2657–2665.
65. Ollivier, N., Vicogne, J., Vallin, A., Drobecq, H., Desmet, R., El Mahdi, O., Leclercq, B., Goormachtigh, G., Fafeur, V., and Melnyk, O. (2012) A One-Pot Three-Segment Ligation Strategy for Protein Chemical Synthesis. *Angew. Chem. Int. Ed.*, **51** (1), 209–213.
66. Sato, K., Shigenaga, A., Tsuji, K., Tsuda, S., Sumikawa, Y., Sakamoto, K., and Otaka, A. (2011) *N*-Sulfanylethylanilide Peptide as a Crypto-Thioester Peptide. *ChemBioChem*, **12** (12), 1840–1844.
67. Agouridas, V., Diemer, V., and Melnyk, O. (2020) Strategies and open questions in solid-phase protein chemical synthesis. *Curr Opin Chem Biol*, **58**, 1–9.
68. Yang, R., Qi, L., Liu, Y., Ding, Y., Kwek, M.S.Y., and Liu, C.-F. (2013) Chemical synthesis of *N*-peptidyl 2-pyrrolidinemethanethiol for peptide ligation. *Tetrahedron Letters*, **54** (29), 3777–3780.

69. Burlina, F., Papageorgiou, G., Morris, C., White, P.D., and Offer, J. (2014) Insitu thioester formation for protein ligation using α -methylcysteine. *Chem. Sci.*, **5** (2), 766–770.
70. Asahina, Y., Nabeshima, K., and Hojo, H. (2015) Peptidyl *N*-alkylcysteine as a peptide thioester surrogate in the native chemical ligation. *Tetrahedron Lett.*, **56** (11), 1370–1373.
71. Tsuda, S., Mochizuki, M., Sakamoto, K., Denda, M., Nishio, H., Otaka, A., and Yoshiya, T. (2016) *N*-Sulfanylethylaminooxybutyramide (SEAoxy): A Crypto-Thioester Compatible with Fmoc Solid-Phase Peptide Synthesis. *Org. Lett.*, **18** (22), 5940–5943.
72. Rao, C., and Liu, C.-F. (2017) Peptide Weinreb amide derivatives as thioester precursors for native chemical ligation. *Org. Biomol. Chem.*, **15** (12), 2491–2496.
73. Shelton, P.M.M., Weller, C.E., and Chatterjee, C. (2017) A Facile *N*-Mercaptoethoxyglycinamide (MEGA) Linker Approach to Peptide Thioesterification and Cyclization. *J. Am. Chem. Soc.*, **139** (11), 3946–3949.
74. Terrier, V.P., Adihou, H., Arnould, M., Delmas, A.F., and Aucagne, V. (2016) A straightforward method for automated Fmoc-based synthesis of bio-inspired peptide crypto-thioesters. *Chem. Sci.*, **7** (1), 339–345.
75. Lelièvre, D., Terrier, V.P., Delmas, A.F., and Aucagne, V. (2016) Native Chemical Ligation Strategy to Overcome Side Reactions during Fmoc-Based Synthesis of C-Terminal Cysteine-Containing Peptides. *Org. Lett.*, **18** (5), 920–923.
76. Binschik, J., and Mootz, H.D. (2013) Chemical Bypass of Intein-Catalyzed *N*-S Acyl Shift in Protein Splicing. *Angew. Chem. Int. Ed.*, **52** (15), 4260–4264.
77. Hunter, C.A. (2004) Quantifying Intermolecular Interactions: Guidelines for the Molecular Recognition Toolbox. *Angew. Chem. Int. Ed.*, **43** (40), 5310–5324.
78. Gautier, A., Nauton, L., and Aucagne, V. DFT computational studies of the *N*-Hnb-Cys thioesterification device. (Unpublished results).
79. Bi, S., Liu, P., Ling, B., Yuan, X., and Jiang, Y. (2018) Mechanism of *N*-to-S acyl transfer of *N*-(2-hydroxybenzyl) cysteine derivatives and origin of phenol acceleration effect. *Chin. Chem. Lett.*, **29** (8), 1264–1268.
80. Miranda, L.P., Meutermans, W.D., Smythe, M.L., and Alewood, P.F. (2000) An activated O \rightarrow N acyl transfer auxiliary: efficient amide-backbone substitution of hindered 'difficult' peptides. *J. Org. Chem.*, **65** (18), 5460–5468.

81. Johnson, T., Quibell, M., Owen, D., and Sheppard, R.C. (1993) A reversible protecting group for the amide bond in peptides. Use in the synthesis of 'difficult sequences'. *J. Chem. Soc., Chem. Commun.*, (4), 369–372.
82. Creighton, C.J., Romoff, T.T., Bu, J.H., and Goodman, M. (1999) Mechanistic Studies of an Unusual Amide Bond Scission. *J. Am. Chem. Soc.*, **121** (29), 6786–6791.
83. Teixidó, M., Albericio, F., and Giralt, E. (2008) Solid-phase synthesis and characterization of *N*-methyl-rich peptides: *N*-methyl-rich peptides. *The Journal of Peptide Research*, **65** (2), 153–166.
84. Lukszo, J., Patterson, D., Albericio, F., and Kates, S.A. (1996) 3-(1-Piperidiny)alanine formation during the preparation of C-terminal cysteine peptides with the Fmoc/*t*-Bu strategy. *Lett Pept Sci*, **3** (3), 157–166.
85. Terrier, V.P., Delmas, A.F., and Aucagne, V. (2017) Efficient synthesis of cysteine-rich cyclic peptides through intramolecular native chemical ligation of *N*-Hnb-Cys peptide crypto-thioesters. *Org. Biomol. Chem.*, **15** (2), 316–319.
86. Martinez, G., Hograindleur, J.-P., Voisin, S., Abi Nahed, R., Abd El Aziz, T.M., Escoffier, J., Bessonnat, J., Fovet, C.-M., De Waard, M., Hennebicq, S., Aucagne, V., Ray, P.F., Schmitt, E., Bulet, P., and Arnoult, C. (2016) Spermaurin, an La1-like peptide from the venom of the scorpion *Scorpio maurus palmatus*, improves sperm motility and fertilization in different mammalian species. *Mol. Hum. Reprod.*, molehr;gaw075v2.
87. De Rosa, L., Di Stasi, R., and D'Andrea, L.D. (2019) Total chemical synthesis by native chemical ligation of the all-D immunoglobulin-like domain 2 of Axl. *Tetrahedron*, **75** (7), 894–905.
88. Huisgen, R. (1955) Structure and mechanism in organic chemistry, von C. K. Ingold. *Angew. Chem.*, **67** (13), 360–360.
89. Adebomi, V., Cohen, R.D., Wills, R., Chavers, H.A.H., Martin, G.E., and Raj, M. (2019) CyClick Chemistry for the Synthesis of Cyclic Peptides. *Angew. Chem. Int. Ed.*, **58** (52), 19073–19080.
90. Koo, B., Dolan, N.S., Wucherer, K., Munch, H.K., and Francis, M.B. (2019) Site-Selective Protein Immobilization on Polymeric Supports through N-Terminal Imidazolidinone Formation. *Biomacromolecules*, **20** (10), 3933–3939.

IV. Publication 3 (Published in Organic & Biomolecular Chemistry)

These results have been submitted (22 Aug 2020) and accepted (18 Sep 2020) for publication in Organic & Biomolecular Chemistry. In this thesis manuscript, we have chosen to reproduce the original paper.

Abboud S.A., and Aucagne V. (2020) An optimized protocol for the synthesis of *N*-2-hydroxybenzyl-cysteine peptide crypto-thioesters. *Org. Biomol. Chem.*, **18**, 8199–8208. DOI : 10.1039/D0OB01737J

An optimized protocol for the synthesis of *N*-2-hydroxybenzyl-cysteine peptide crypto-thioesters.

Skander A. Abboud^a and Vincent Aucagne^{a,*}

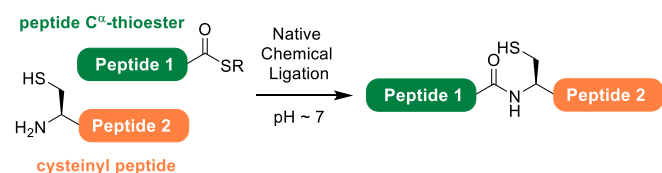
We herein report a robust upgraded synthetic protocol for the synthesis of *N*-Hnb-Cys crypto-thioester peptides, useful building blocks for segment-based chemical protein synthesis through native chemical ligation. We recently observed the formation of an isomeric co-product when using a different solid support than the originally-reported one, thus hampering the general applicability of the methodology. We undertook a systematic study to characterize this compound and identify the parameters favouring its formation. We show here that epimerization from L- to D-cysteine occurred during the key solid-supported reductive amination step. We also observed the formation of imidazolidinones by-products arising from incomplete reduction of the imine. Structural characterization combined with the deciphering of underlying reaction mechanisms allowed us to optimize conditions that abolished the formation of all these side-products.

Received 22nd August 2020,
Accepted 18th September 2020

DOI: 10.1039/d0ob01737j

Introduction

The advent of the native chemical ligation (NCL) reaction¹ (scheme 1) opened the realm of proteins to synthetic organic chemists, leading to a new paradigm in protein engineering. NCL substantially pushed away the limits of solid phase peptide synthesis (SPPS, typically limited to a few dozen amino-acid residues),² by providing a simple solution for the assembly of SPPS-synthesized segments. More than 25 years after its discovery, NCL is still, by far, the most widely used approach for chemical protein synthesis³ and has led to tremendous applications to functional proteins of more than 300 residues.⁴



Scheme 1: The native chemical ligation (NCL) reaction

NCL is based on the chemoselective coupling of unprotected peptides bearing a C-terminal thioester and an N-terminal cysteine residue, respectively, to form a native amide bond at a neutral pH, which is optimal for the reaction. Peptide C α -thioesters segments are key building blocks. However, the thioester group is not stable towards piperidine treatments used repeatedly in Fmoc-based SPPS for the deprotection of the *N*-Fmoc group. Given that SPPS is conducted from the C- to N-

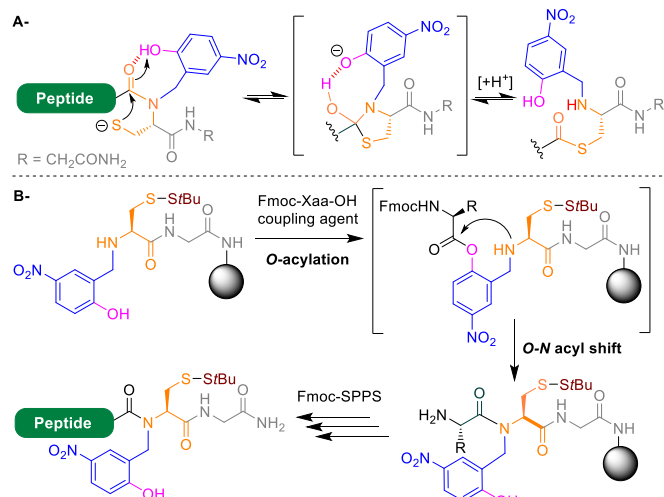
terminus, Fmoc-SPPS is essentially incompatible to the direct synthesis of C-terminal thioesters. A vast number of synthetic methodologies have been explored to circumvent this issue⁵ but none is yet generally applicable. Most strategies rely on the introduction of the thioester function after the Fmoc-SPPS elongation of the peptide segment. The most widely used thioester precursors are Liu's hydrazides⁶ and Dawson's *N*-acylureas (Nbz/Me-Nbz).⁷ These two C-terminal appendages conveniently allow for the *in situ* generation of a thioester (either before or during NCL, respectively), classically through thiolysis using a large excess of the arylthiol mercaptophenylacetic acid (MPAA).⁸ If these two methodologies are easy to implement and significantly expanded the applicability of NCL-based protein chemical synthesis, they suffer from some limitations⁹ and there is still a need for alternative synthetic strategies.¹⁰

Another approach has been increasingly used during the past decade. It exploits the reversible rearrangement through *N*-to-*S* acyl shift of a β - or γ -mercapto amide intramolecular thioesterification device, leading a β - or γ -amino thioester, respectively.¹¹ In most cases, this process necessitates acidic conditions to shift the *N*-*S* shift equilibrium by protonation of the amine of the rearranged product. *Trans*-thioesterification with an exogenous thiol and subsequent isolation of the resulting thioester is thus needed prior to NCL. A few β -mercapto amides, later coined as "crypto-thioesters",^{11k} are able to rearrange and undergo *trans*-thioesterification in tandem with NCL, thus considerably simplifying the use of *N*→*S* shift devices for chemical protein synthesis. Varieties of elegant

^a Centre de Biophysique Moléculaire, CNRS UPR 4301, Rue Charles Sadron 45071 Orléans cedex 2, France. E-mail: vincent.aucagne@cnrs-orleans.fr.

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

designs of such crypto-thioesters have been proposed. For example, cysteine-prolyl esters (CPE) developed by Aimoto^{11f} and later extended to cysteine-prolyl imides (CPI) by Hayashi and Okamoto^{11r} are based on an intramolecular trapping of the amine group. *Bis*-sulfanylethylamides (SEA) introduced by Melnyk¹¹ⁱ and Liu^{11j} incorporate a second thiol group that is thought to accelerate the *N*-S shift equilibrium through intramolecular protonation, and Hojo's *N*-alkyl cysteines (NAC)^{11o} presumably proceed following a similar mechanism, involving in this case the C-terminal Cys carboxylic acid. Otaka's *N*-sulfanylethylanilides (SEAlide)^{11k} exploit the destabilization of the amide by an *N*-aryl substituent and Offer's C^α-methyl cysteinamides,¹¹ⁿ an α -gem-dialkyl substitution. Despite remarkable applications, all these crypto-thioesters show dramatically slower NCL kinetics than preformed alkyl thioesters. An additional drawback of most of these methodologies lies in the difficult coupling of the C-terminal residue of the peptide segment, requiring *N*-acylation of a sterically-hindered secondary amine. For example, the delicate use of moisture-sensitive coupling agent PyBroP¹¹ⁱ and Fmoc-amino acid fluorides,¹¹ⁱ or long couplings at high temperature using HATU^{11e,o} are typically required for acylation of SEA or NAC devices, respectively. Synthetically-demanding preformation of a building block by *N*-acylation in solution has been shown to be a preferable solution for SEAlides^{11d,k,12} and NAC.¹³ To tackle these limitations, our group reported a few years ago a simple methodology based on an *N*-(2-hydroxybenzyl) cysteine device (*N*-Hnb-Cys).^{11p} *N*-Hnb-Cys shows superior *N*→*S*-shift/*trans*-thioesterification/NCL overall kinetics, owing to an internal catalysis by the phenol group acting as both an H-bond donor¹⁴ binding to the amide oxygen atom and proton donor that stabilizes the *N*-S shift tetrahedral intermediate (scheme 2A).¹⁵ The phenol has an additional effect: it assists the *N*-acylation of the *N*-Hnb-Cys device through an *O*-acylation/*O*→*N*-acyl shift mechanism,¹⁶ therefore greatly facilitating the introduction of the C-terminal residue of peptide segments (scheme 2B).



Scheme 2: A) Thioester formation mechanism through *N*→*S* acyl shift of *N*-Hnb-Cys crypto-thioesters; B) Mechanism of *N*-acylation of the Hnb-Cys device.

N-Hnb-Cys crypto-thioesters synthesis is straightforward, and can easily be automated on a standard peptide synthesizer

using inexpensive building blocks. These crypto-thioesters have been extensively used by us¹⁷ and others¹⁸ for the synthesis of a wide range of targets, including cyclic peptides,^{17c} C-terminal-cysteine-containing peptides^{17a} and other disulfide rich miniproteins,^{17b,d} as well as a tyrosine kinase receptor domain,¹⁸ illustrating the wide applicability of the method. However, we recently noticed side-reactions when using the initially disclosed synthetic procedure, and decided to take the opportunity for an in-depth study of a key reduction amination step to provide a robust synthetic protocol.

Results and discussion

N-Hnb-Cys device synthesis starts with the Fmoc-SPPS coupling of a glycine, then a cysteine bearing an *S*-tert-butylsulfanyl (StBu) protecting group, on a SPPS resin equipped with a Rink amide linker.¹⁹ The Gly residue is introduced to prevent SPPS side reactions like β -elimination of Cys esters and amide hydrolysis during TFA cleavage of Cys directly linked to a Rink linker.^{11p} The StBu protecting group is preferred to the classical trityl to afford the crypto-thioester in a dormant state, unable to rearrange into a hydrolyzable thioester during handling, purification and storage. StBu is cleaved within a few minutes under classical NCL conditions (including TCEP and MPAA), leading to the active *N*-Hnb-Cys crypto-thioester species. The *N*-Hnb group is incorporated after Gly and Cys couplings through a two-steps reductive amination reaction to afford the thioesterification device (see fig. 1). The best protocol we identified in our initial report consists in (1) treatment with a dilute acetic acid solution in order to protonate the amine, followed by washings to eliminate excess acid, (2) imine formation through incubation with 10 equiv. of 2-hydroxy-4-nitro benzaldehyde for one hour in 1:1 DMF/MeOH, followed by washings to eliminate excess aldehyde and (3) reduction with a large excess of sodium cyanoborohydride in 9:9:2 DMF/MeOH/AcOH. These conditions led to quantitative Hnb introduction.²⁰

In all our reported applications,¹⁷ *N*-Hnb-Cys crypto-thioesters were synthesized on an amino-Tentagel R resin, a PEG-grafted polystyrene. Recently, during attempts to use ChemMatrix, a PEG-based polymer known for its superior properties for the synthesis of long or "difficult" peptide sequences,²¹ we observed a satellite to the major peak: cleavage of the resin after the Fmoc group deprotection of the peptide segment first residue, an *O*-tBu-serine in this case, similarly led to a mixture of the expected H-Ser-(Hnb)Cys(StBu)-Gly-NH₂ (**4**) together with a significant side-product (**5**) (8-15%) showing the same *m/z* in MS analysis (fig. 1B), indicating an isomerization during either the *N*-Hnb-Cys device synthesis or the Ser coupling.

This result was somehow surprising, as we previously reported the synthesis of **4** on Tentagel without observing any significant impurities.²² This prompted us to investigate the influence of the polymer support on the generation of its isomer **5**, since resins do not behave as inert carriers but rather act as co-solvents that can affect various parameters of a chemical

reaction such as yield, purity, rate and parasite reactions.²³ A resin screening study (table 1) showed that the amount of side-product is indeed highly dependent on the nature of the polymer support: the poly-lysine-based resin Spheritide²⁴ and ChemMatrix gave the worst results, followed by the PEG-grafted polystyrene resin Novagel²⁵ and polystyrene (see supporting information, scheme p S4) for polymer structures).

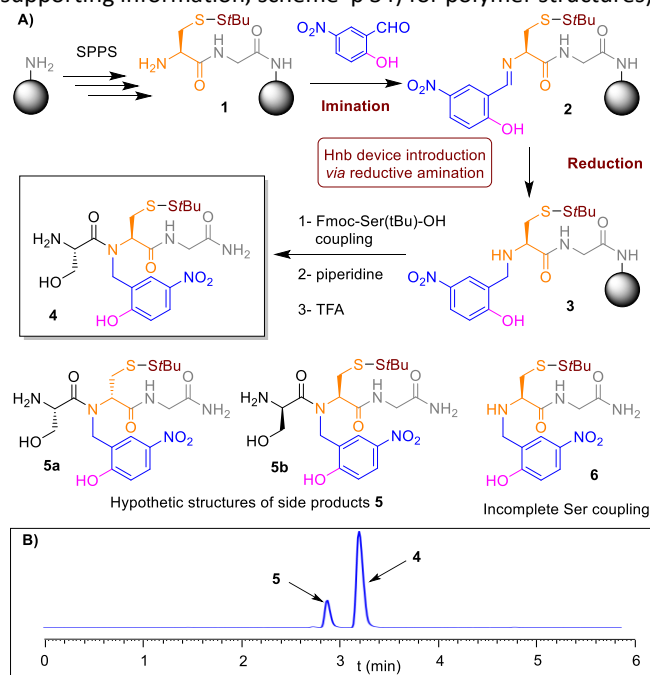


Figure 1: A) Synthesis of the *N*-Hnb-Cys thioesterification device through solid phase reductive amination (conditions: see text). B) HPLC chromatogram ($\lambda = 320$ nm) obtained when using a ChemMatrix resin, showing a satellite to the major peak after serine coupling and TFA cleavage.

We also observed the formation of **5** –albeit in a lower extent– when using Tentagel, in contradiction with our initial report. Several repetitions of the latter experiment led to similar results, even if the amount of **5** slightly varied upon batches (5–8%).

Table 1 Resin effect on side product **5** formation during the synthesis of **4**.^[a]

Entry	Resin	Loading (mmol/g)	4 (%) ^[b]	5 (%) ^[b]
1	ChemMatrix	0.54	89	11
2	Tentagel R	0.16	93	7
3	Polystyrene	0.89	91	9
4	Spheritide	0.15	86	14
5	Novagel	0.77	90	10

[a] Reaction conditions¹¹⁰: imination with 10 equiv. aldehyde, 1 equiv. AcOH in 1:1 DMF/MeOH, 12.5 mM peptidyl resin, RT; reduction with 20 equiv. NaBH₃CN in 9:9:2 DMF/MeOH/AcOH, 12.5 mM peptidyl resin, RT; [b] Relative rates determined by integration of HPLC peaks at $\lambda = 320$ nm (λ_{max} Hnb).

We hypothesized that one of the chiral amino acid residues epimerized during the synthesis, either Cys (leading to compound **5a**) or Ser (compound **5b**), and we decided to carry out a systematic study to understand and minimize this side

reaction, focusing on the ChemMatrix resin that gave particularly high amounts of **5**.

At first, we assumed that the serine residue could have likely epimerized during its coupling. Indeed, Ser is known to be one of the proteogenic amino acids most prone to epimerization upon coupling,²⁶ after Cys²⁷ and His.²⁸ We thought that the *O*→*N* shift mechanism assumed for Hnb-Cys *N*-acylation could potentially have enhanced this propensity.

Therefore, we tried to optimize Ser coupling, focusing on reaction conditions having been described to reduce epimerization in specific cases (table 2). We did not observe any significant improvement in the **4** / **5** ratios (9–11% **5**), and even a slight increase when using *N*-methyl-morpholine (NMM) as the base (entry 5). We also monitored the completion of the coupling: the reaction went to completion in most cases except when using NMM, or 1:1 DCM/DMF as solvent (entries 5 and 6, respectively, observation of the non-acylated compound **6**).

Table 2 Effectiveness of different coupling conditions for Ser coupling on **3**.^[a]

Entry	Coupling agent	Base	Solvent	4 + 5 (%) ^[b]	4 : 5 ratio ^[b]	6 (%) ^[b]
1	HCTU	DIEA	DMF	100	89:11	-
2	HATU	DIEA	DMF	100	90:10	-
3	DIC/Oxyma	-	DMF	100	90:10	-
4	DIC/HOBt	-	DMF	100	90:10	-
5	HCTU	NMM	DMF	93	85:15	7
6	HCTU	DIEA	DCM/DMF	78	91:9	22

[a] 18 h coupling, 10 mM peptidyl resin, 10 equiv. Fmoc-Ser(tBu)-OH, 9.5 equiv. coupling agent, 20 equiv. base, RT; [b] Relative rates determined by integration of HPLC peaks at $\lambda = 320$ nm.

Puzzled by these results, the next hypothesis we explored was Cys(*StBu*) epimerization, either as the sole source of isomer formation, or in addition to Ser epimerization. For this purpose we coupled an extra Val residue after Ser, in order to add a third chiral centre and obtain diastereoisomers discernible by classical achiral RP-HPLC, instead of the co-eluted enantiomers **5a** and **5b**. This led to a mixture of four HPLC peaks showing the same *m/z*, in a 89.2 : 9 : 1.6 : 0.2 ratio (fig. 2B). These peaks were supposed to correspond to all four diastereoisomers **7a-d** arising from Cys and/or Ser epimerization, the major peak being obviously the non-epimerized product **7a**. In order to unambiguously attribute the three remaining peaks to putative structures **7b-d**, we generated three HPLC standards corresponding to Cys epimerization (**7b**), Ser epimerization (**7c**) and both Ser and Cys epimerization (**7d**). In the case of **7b** and **7d**, we synthesized their enantiomeric counterpart **7e** and **7f**, respectively, as the latter compounds were less synthetically demanding while being equivalent in achiral HPLC (fig. 2A).

As expected, this led in each case to a major peak accompanied with three minor ones. Co-injection of the four crude mixtures

led to four well-separated HPLC peaks (SI, fig. S8). This allowed us to identify the peaks in the initial chromatogram: as anticipated, the most important epimerization corresponds to D-cysteine (**7b** + **7d**: 9.4%), even if a small amount of D-serine (**7c** + **7d**: 1.8%) was also detected. (fig. 2B).

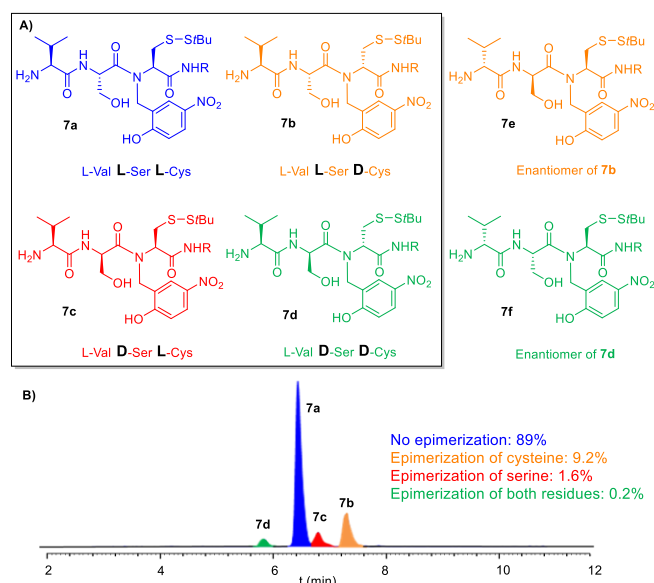
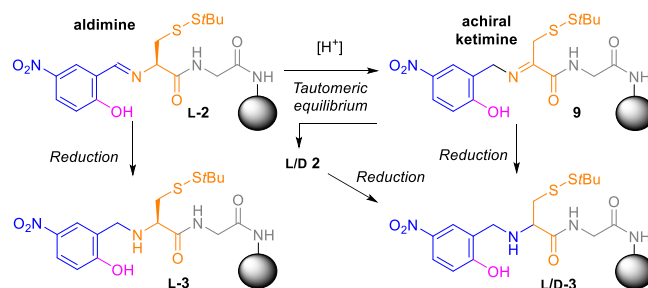


Figure 2: A) Structures of the four diastereoisomer products **7a-d** and the HPLC standards **7e** and **7f**. R = CH₂CONH₂. B) HPLC Chromatogram (λ = 320nm) of crude mixture **7**. % determined by HPLC peak integration.

Having clearly established that it is the cysteine residue that mostly epimerizes, we were quite relieved about the efficiency of our previously-described protocol for applications to chemical protein synthesis: indeed, this meant that the epimerized residue will not be present in ligation products as a consequence of the expulsion of the *N*-Hnb-D/L-Cys thioesterification device during NCL.²⁹ However, D-Cys formation complicates the purification of the crypto-thioester segments and is deleterious in terms of isolated yields and product purity. We thus decided to pursue a systematic investigation to determine the origin of Cys epimerization. We started by checking the enantiomeric purity of our commercial cysteine building block, through formation of diastereoisomers by coupling with D- and L-versions of a chiral amine. No D-Cys isomer was detected (supporting information, p S15). Next, in order to assess if epimerization happens during the coupling of the cysteine residue, we synthesized the tripeptide H-Ser-Cys(SiBu)-Gly-NH₂ (**8**) devoid of Hnb group, together with an HPLC standard corresponding to the epimerized product (supporting information, p S16). We did not detect any D-Cys, which means that the epimerization occurs during the introduction of the Hnb group.

We hypothesized that Cys epimerization could likely arise from a tautomeric equilibrium between aldimine **2** and a rearranged achiral ketimine **9** (scheme 3) prior to reduction into amine **3**. Such imine tautomerism has been widely investigated in synthetic organic chemistry,³⁰ and can be catalysed by strong bases³¹ acids³² metals³³ or high temperatures.³⁴ This rearrangement has been exploited in protein and peptide

chemistry for the bio-inspired conversion of N-terminal amines to ketones in aqueous media, through a ketimine intermediate which is further hydrolyzed.³⁵ Careful re-examination of the literature in the solid phase synthesis field told us that this side reaction has also been observed during the grafting of amino acid derivatives on a backbone amide linker-functionalized resin through reductive amination, and was even intentionally used for the complete racemization of chiral amino acids.³²



Scheme 3: Putative epimerization mechanism through amine tautomerism

To probe the imine tautomerism hypothesis, we varied the imination time prior to treatment with sodium cyanoborohydride, followed by coupling of Fmoc-Ser(*t*Bu)-OH. Consistently, prolongation from one to four hours dramatically enhanced the formation of the side-product **5**, (table 3, entries 1 and 2, respectively). Reduction of the imination time to 5 min only led to a moderate decrease of the formation of **5** (table 3, entry 3): this could be explained by a slow reduction of the imine by NaBH₃CN, leaving enough time for the deleterious tautomeric equilibrium to proceed during the reduction step. An overnight treatment of the reduced compound (**3**) under the same reduction conditions did not change the **4** : **5** ratio, confirming that epimerization does not happen after reduction of the imine.

Table 3 Effect of the variation of imination reaction time on epimerization.

Entry	Reducing agent	Imination ^[a] time	Reduction ^[b] time	4 (%) ^[c]	5 (%) ^[c]
1 ^[d]	NaBH ₃ CN	1 h	1 h	89	11
2	NaBH ₃ CN	4 h	1 h	80	20
3	NaBH ₃ CN	5 min	1 h	92	8

[a] Imination conditions : 10 equiv. aldehyde, 1 equiv. AcOH, 1:1 DMF/MeOH, 12.5 mM peptidyl resin, RT; [b] Reduction conditions : 9:9:2 DMF/MeOH/AcOH, 12.5 mM peptidyl resin, RT; [c] Relative rates determined by HPLC peak integration at λ = 320 nm. [d]: initial conditions from reference.^{11p}

If things started to be clearer on the rationale behind the formation of byproduct **5**, we remained intrigued by the low reproducibility of our results (present and past experiments^{11p}). We questioned if the relative notion of “room” temperature, or other biases upcoming from the conduction of the reaction on a solid support (efficiency of the washing steps, actual final concentrations, etc.) could be responsible for this lack of reproducibility. High temperatures favor the formation of **5**: up

to 32% when the reaction is performed at 60 °C, and only 8% at 4 °C (supporting information, table S2). Epimerization depends on concentration: the more the reaction is diluted the more the rate of the epimer increases (supporting information, table S3). Moreover, the amount of acetic acid used during the imine formation step also modulates the formation of **5** (table 4). Use of a large excess (10% v/v, 140 equiv., entry 4) dramatically increased epimerization, while a lower concentration (1%, entry 3) slightly reduced it, as compared with our initial conditions where no extra AcOH was added in addition to the 1 equiv. acetate counter anion arising from prior amine protonation (entry 2).

Table 4 Variation of acetic acid amount during imination reaction.^[a]

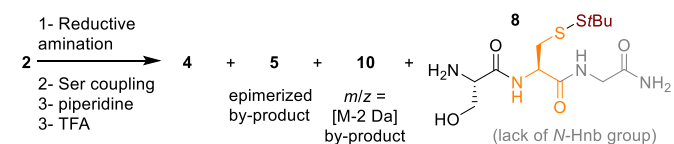
Entry	Additives	4 (%)	5 (%) ^[b]
1	No additive	89	11
2	1 equiv. AcOH	89	11
3	14 equiv. AcOH (1% v/v)	91	9
4	140 equiv. AcOH (10% v/v)	68	32

[a] Reaction conditions: Imination in 1:1 DMF/MeOH for 1 h at RT, 12.5 mM peptidyl resin; reduction in DMF/MeOH/AcOH (9:9:2) for 1 h at RT. [b] Relative rates determined by integration of HPLC peaks at $\lambda = 320$ nm.

We believe that this high overall sensitivity to reaction conditions is likely responsible for the divergence between our initial report and this work. This prompted us to start a re-optimization of the reductive amination, willing to establish a highly reproducible protocol.

Having demonstrated that imine **2** does undergo epimerization, we focused on reducing its lifetime during the reductive amination process. We thus re-investigated one-pot imination/reduction, followed by Ser coupling, with the goal to minimize the formation of **5** as well as of non- or *bis*-alkylated products observed in our initial study when trying different conditions to optimize the reductive amination step. We explored the use of sodium cyanoborohydride, in addition to pyridine borane and sodium triacetoxyborohydride, two other classical reagents for solution-phase one-pot reductive amination³⁶ (table 5).³⁷ The one pot process greatly minimized epimer **5** formation, down to a ~ 2% rate very close to epimerization accounting from the Ser residue (1.8%, fig. 2). However, significant amounts (2–17%) of non-alkylated compound **8** were also produced, in line with our initial study. Surprisingly, use of sodium tri-acetoxyborohydride led to the formation of four different peaks (**10a–d**, entry 8) which masses correspond to [M–2 Da] of the desired product, and no traces of either **4** or **5**. After re-investigation of the LC-MS chromatograms of all crude products, small amounts of the same new byproducts **10** were also observed in most other one-pot reductive aminations (entries 1–7), while being absent when using our initial protocol. Use of pyridine borane in DMF/MeOH

gave satisfactory results in terms of epimer **5** and [M–2 Da] **10** byproducts formation, albeit accompanied with the generation of non-alkylated compound (entry 4), that can be minimized using prolonged reaction time (entry 7). If the latter conditions proved to be highly reproducible and relatively well suited for manual introduction of the Hnb group, it was difficult to adapt to an automated version, due to the observed instability of the reducing agent upon storage in solution for a few hours.



Scheme 4: Mixture of products obtained after one pot reductive amination

Table 5 One pot reductive amination.^[a]

Entry	Reducing agent	Solvent	4 ^[b] (%)	5 ^[b] (%)	10 ^[b] (%)	8 ^[c] (%)
1 ^[d]	NaBH ₃ CN	DMF/MeOH	84.3	2.2	4.2	9.3
2	NaBH ₃ CN	MeOH	94.6	1.9	0.6	2.9
3	NaBH ₃ CN	THF	84	3.8	8.7	3.5
4 ^[d]	Pyridine·BH ₃	DMF/MeOH	86.4	1.8	< 0.1	11.8
5 ^[d,e]	Pyridine·BH ₃	DMF/MeOH	95.5	2	< 0.1	2.5
6	Pyridine·BH ₃	MeOH	80.5	2.1	0.3	17.1
7	Pyridine·BH ₃	THF	75.4	2.4	17.7	4.5
8 ^[f]	NaBH(OAc) ₃	THF/MeOH	0	0	100	0

[a] 5 min reactions using 10 equiv. aldehyde, 20 equiv. reducing agent, 1% AcOH as additive, RT, 12.5 mM peptidyl resin; [b] Relative rates determined by HPLC peak integration at $\lambda = 320$ nm; [c] Relative rate determined by coupling of an extra Trp residue and integration of HPLC peaks at $\lambda = 280$ nm (see supporting information p S45); [d] 1:1 DMF/MeOH ratio; [e] 20 min reaction; [f] 1:1 THF/MeOH ratio.

Prior to further optimization of a synthetic protocol compatible with routine automation on a peptide synthesizer, we wanted to decipher the formation of the [M–2 Da] byproducts **10**. Considering that this mass shift is consistent with compounds that would have not been reduced, we tried imine formation followed by overnight Ser coupling, without treatment by any reducing agent. This led to a chromatogram essentially identical to the one obtained with sodium tri-acetoxyborohydride, demonstrating that **10** indeed arose from incomplete reduction of the imine. We next wanted to characterize the four products **10a–d** by NMR in order to elucidate their chemical structure. Separation of the four peaks proved to be very challenging, but we succeeded in isolating small quantities of the three compounds **10a–c**, as well as **10d** in mixture with **10c**. Disappointingly, extensive NMR characterization was not feasible due to the presence of two or more different sets of NMR peaks in each purified compound, probably arising from conformers in slow exchange, which considerably complicated the spectra.³⁸ However, combination of 1D and 2D ¹H, ¹³C and ¹⁵N experiments (reported in detail in the supporting

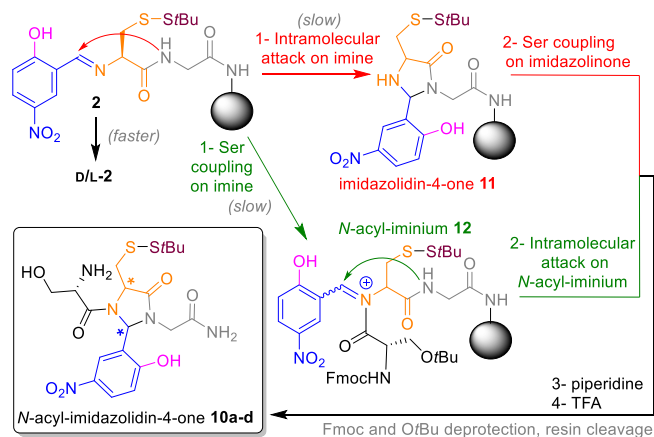
information, p S27-S41) allowed us to elaborate a solid hypothesis for the structures of **10a-d**. In a few words, we observed for the four products the disappearance of the benzylic CH₂ signals observed in both **4** and **5**, accompanied with the apparition of new CH singlets at 6.0-6.7 ppm. This suggested the formation of compounds C-disubstituted in α of the nitrophenol ring.³⁹ Moreover, ¹H-¹⁵N HSQC spectra of **10a-d** showed the disappearance of the NH amide signal of the Cys-Gly peptide bond, while glycine C-terminal CONH₂ and serine NH₃⁺ (TFA salt) remained essentially unchanged: this suggested an *N*-substitution of the Cys-Gly amide nitrogen. We consequently deduced the probable *N*-acyl imidazolidinone structures shown in scheme 5. This hypothesis was further strengthened by comparison with literature NMR data for related compounds: ¹H and ¹³C chemical shifts of the newly formed CH signals in **10a-d** are consistent with those reported for structurally-related *N*-acyl-imidazolidinones (see supporting information, p S42).⁴⁰

From a mechanistic point of view, these five-membered rings could be obtained through two different known pathways (scheme 5). In the first one, a highly reactive *N*-acyliminium would be formed from the coupling of serine with the non-reduced imine, followed by intramolecular trapping with the Cys-Gly amide nitrogen acting as a nucleophile (green pathway). Very similar reactions were reported in the literature using T3P as a strong carboxylic acid activator.⁴¹ In our case, overnight reaction using HCTU could be sufficient to promote coupling with the poorly nucleophilic imine, and assistance of the Hnb phenol group for imine *N*-acylation, similar to the *O*-acylation/*O* → *N*-acyl shift mechanism during amine *N*-acylation of the reduced compound **3**, can also be advocated. In a second plausible hypothesis, nucleophilic attack of the Cys-Gly amide nitrogen on the non-reduced imine would generate imidazolidinone **11**, followed by serine coupling (red pathway). Related reactions were recently exploited by Francis and Raj groups for protein immobilization on polymeric supports and for the synthesis of conformationally rigid macrocycles through chemo- and diastereoselective cyclization ("CyClick chemistry"), respectively.⁴²

Note that, in both hypotheses, imine **2** is expected to have a long lifespan during the overnight Ser coupling, thus being consistent with a massive epimerization of Cys through aldimine-ketimine tautomerism prior to imidazolidinone formation. Moreover, addition of the Cys-Gly amide nitrogen on the *N*-acyliminium or imine is likely to proceed in a poorly-diastereoselective fashion, affording epimeric products at the benzylic position. These two considerations are fully consistent with the formation of comparable amounts of each of the imidazolidones **10a-d**⁴³ that thus likely correspond to the four possible diastereoisomeric structures depicted in scheme 5.⁴⁴

Having elucidated the structure of all byproducts (**5a-b**, **10a-d**) together with drawing up plausible hypotheses that explain their formation, we turned back to our initial goal of protocol optimization. Our findings taught us that the lifespan of imine **2** should be kept as short as possible. In consequence, we designed a simple protocol based on a short imination

time, combined with very fast reduction of the imine. Considering that we observed slightly higher formation of **5**



Scheme 5: Mechanistic hypotheses for imidazolidinones **10a-d** formation.

when using either no acid additive or a large excess (see above, table 4), we conducted the imination through a 5 min treatment with a tenfold excess of 5-hydroxy-2-nitrobenzaldehyde in a 1:1DMF/MeOH solvent mixture supplemented with 1% AcOH. After prompt resin washes, we treated the imine resin **2** with a large excess of sodium borohydride (NaBH₄) in DMF. This reducing agent is stronger than sodium cyanoborohydride and in contrast with the latter, NaBH₄ is known to rapidly reduce imines without requiring iminium formation through protonation. It is also known to react with acids and methanol, reason why we used neat DMF. Very gratifyingly, these conditions resulted in (1) no observation of **10a-d** upcoming from incomplete reduction, (2) no formation of non- or bis-alkylated products and (3) only 2% epimerization, thus very close to the 1.8% D-Ser formation during coupling (fig. 2).⁴⁵ On top of that, we obtained equivalent results after keeping the NaBH₄ solution in DMF for at least one day at room temperature, an ideal feature for automated *N*-Hnb-Cys cryptothioesters preparation.

Table 6 Two-steps NaBH₄ reductive amination.^[a]

Entry	Imination time	Reduction time	4 ^[b] (%)	5 ^[b] (%)	10 ^[b] (%)	8 ^[c] (%)
1	5 min	20 min	98	2	<0.1	<0.1

[a] Reaction conditions : Imination with 20 equiv. aldehyde in DMF/MeOH 1:1 + 1% AcOH, RT, 12.5 mM peptidyl resin; reduction with 20 equiv. NaBH₄ in DMF, RT, 12.5 mM peptidyl resin. [b] Relative rates determined by HPLC peak integration at λ = 320 nm. [c] Relative rate determined by coupling of an extra Trp residue and integration of HPLC peaks at λ = 280 nm (see supporting information p S45).

With such a gold standard protocol for reductive amination in hands, we wanted to assess the actual Cys vs Ser epimerization under these conditions, and thus coupled an extra Val for diastereoisomer formation. We were pleased to observe that Cys(StBu) epimerization was reduced to as low as 0.3% (supporting information, supplementary figure S9b). The remaining 1.7% corresponds to Ser epimerization (product **7b**) during its coupling. We finally re-investigated the coupling

conditions of the Ser residue on the *N*-Hnb-Cys(*StBu*) device (**3**), DIC/Oxyma giving significantly improved results (0.6 % Ser epimerization, see supplementary information figure S10 and table S7). Gratifyingly, these conditions also gave excellent results for the coupling of Cys(*Trt*) and His(*Trt*) (0.8% and 0.2% epimerization, respectively, see supporting information tables S5 and S6).

The new reductive amination protocol was finally challenged with the automated synthesis of two different peptides. The model peptide crypto-thioester H-Leu-Tyr-Arg-Ala-Ala-(Hnb)Cys(*StBu*)-Gly-NH₂ (**S23**) was synthesized on Tentagel, ChemMatrix and Rink-MBHA-polystyrene resins (supporting information fig. S42) giving clean products devoid of any detectable above-mentioned side-products arising from epimerization or incomplete reduction. Synthesis of a very long 62 amino acids peptide crypto-thioester, the sequence of which derives from the human mucin MUC1⁴⁶ (supporting information, fig. S47, compound **S24**) also gave excellent results, fully validating our optimized protocol.

Conclusions

In conclusion, an in-depth reinvestigation of previously described protocols for *N*-Hnb-Cys peptide crypto-thioesters synthesis led to the identification of a variety of deleterious by-products. The major one corresponds to the epimerization of the Cys residue during solid-supported reductive amination. Fine understanding of the structures of the by-products combined with the deciphering of their mechanism of formation allowed to reduce the total amount of epimerization from 11 to less than 1% and to optimize a highly-reproducible synthetic protocol leading to clean peptide crypto-thioester segments. In addition to providing a gold standard automatable protocol for the preparation of *N*-Hnb-Cys peptide crypto-thioesters for applications to the chemical synthesis of proteins, the findings herein reported could be significant for the synthesis of *N*-alkylated chiral amino acids or peptide through reductive amination, either in solution or on solid phase.⁴⁷

Experimental Section

All experimental protocols and compounds characterization are detailed in the supporting information file.

Recommended protocol for the synthesis of *N*-Hnb-Cys crypto-thioesters.

Reductive amination: 25 μ mol of a H-Cys(*StBu*)Gly-Rink-resin synthesized through standard Fmoc-SPPS was washed with a 1:1 DMF/MeOH mixture (4 \times 3 mL for 30 s). 2-Hydroxy-5-nitrobenzaldehyde (42 mg, 0.25 mmol, 10 equiv.) in 2 mL 44.5:44.5:1 DMF/MeOH/AcOH (125 mM aldehyde concentration) was then added, and the reaction vessel was stirred for 5 min at room temperature. Solvents were drained and the resin was washed with 1:1 DMF/MeOH (3 \times 3 mL for 5 s) then DMF (3 \times 3 mL for 5 s). Without delay, a solution of sodium borohydride (19 mg, 0.5 mmol, 20 equiv.) in 2 mL DMF (250 mM borohydride concentration) was added and the

reaction vessel was stirred for 20 min at room temperature. Solvents were drained and the resin was washed with DMF (4 \times 3 mL for 30 s), 20% v/v piperidine in NMP (3 \times 3 mL for 3 min), NMP (3 \times 3 mL for 30 s), dichloromethane (3 \times 5 mL for 30 s) and NMP (3 \times 3 mL for 30 s).

Important notes (1): it is critical to limit to 5 min the incubation with the aldehyde, to keep subsequent washes as short as possible and to add without delay the NaBH₄ solution after washes. The Prelude synthesizer that we use operates its six reaction vessels (RV) in a successive fashion rather than in parallel. We therefore prefer to program the reductive amination protocol one RV after the other.

Important notes (2): "fresh" solutions of NaBH₄ can be prepared up to at least 24 h in advance before use without any loss of efficiency. Use peptide-grade DMF for all three reductive amination steps.

Coupling of the C-terminal amino acid of the peptide crypto-thioester segment sequence: as disclosed in our previous works, the *N*-acylation kinetics highly depends on the C-terminal residue. If non-hindered amino-acids can be easily introduced through standard Fmoc-SPPS protocols, beta-substituted ones (Ile, Val, Thr) and Pro require prolonged coupling (slow *O*-to-*N* acyl shift^{17a}) and epimerization-prone residue (Cys/His) needs particular attention.

For Ala / Arg / Asn / Gln / Gly / Leu / Lys / Met / Phe / Ser / Trp / Tyr: protected amino acids (0.25mmol, 10 equiv.) were coupled using HCTU (98 mg, 0.238 mmol, 9.5 equiv.) and DIEA (87 μ L, 0.5 mmol, 20 equiv.) in NMP (2.5 mL) for 3 \times 3 h. A single 30 min coupling is sufficient for Gly and Ala,^{11p} but we prefer to routinely use this long triple coupling protocol for synthesizer programming simplicity.

For Ile / Pro / Thr / Val: protected amino acids (0.25mmol, 10 equiv.) were coupled using HCTU (98 mg, 0.238 mmol, 9.5 equiv.) and DIEA (87 μ L, 0.5 mmol, 20 equiv.) in NMP (2.5 mL) for 2 \times 18 h.

For epimerization-prone amino acids Cys / His: protected amino acids (0.25mmol, 10 equiv.) were coupled using DIC (39 μ L, 0.25 mmol, 10 equiv.) and oxyma (45 mg, 0.075 mmol, 10 equiv.) in NMP (2.5 mL) for 12 h.

Note that unprotected carboxylic acid side chains of Asp and Glu are known to react intramolecularly with the C α thioester function leading to C-terminal anhydrides and thus the obtention of *iso*Asp and *iso*Glu-containing ligation product, respectively,^{9e,48} and we do not recommend the use of Asp- and Glu-*N*-Hnb-Cys crypto-thioesters. We observed this deleterious side-reaction for Asp, and did not check it for Glu.

In all cases, we routinely perform a capping step of eventual unreacted amine groups by four consecutive treatments with acetic anhydride (143 μ L, 1.51 mmol, 60 equiv), *i*Pr₂NEt (68 μ L, 0.39 mmol, 15.5 equiv.) and HOBt (6 mg, 0.044 mmol, 1.8 equiv.) in NMP (3 mL) for 7 min.

Crypto-thioester peptide elongation: classical Fmoc-SPPS elongation using uronium coupling agents (HBTU/HOBt, HCTU, HATU) should be preferred. In our hands, this gave similar results in terms of qualitative HPLC chromatogram as for peptide segments not incorporating a C-terminal *N*-Hnb-Cys device.

Important note (3): phosphonium coupling reagents should be avoided, as we observed side-reaction corresponding to *O*-*P* bond formation on the Hnb phenol group when using PyAOP or PyBOP.

Important note (4): microwave- and other heat-accelerated syntheses of *N*-Hnb-Cys crypto-thioesters probably remains to be optimized, as by-products arising from β -elimination of the Cys(StBu) followed by Michael-type addition of piperidine were observed.⁴⁹ Note that we detected small amounts of this by-product at room temperature for a very long peptide (compound **S24**, see supporting information, p S47-S48), but this side-reaction is probably sequence-dependant: indeed, a study on a short model peptide did not show any trace of a by-product even after a very long piperidine treatment (supporting information, p S48-S49).

Important note (5): variable amount (5-90 %) of *O*-acylation of the Hnb phenol are observed after each coupling. This ester is cleaved within seconds upon piperidine treatment during Fmoc deprotection. As we make use of 10 equiv. of Fmoc-Xaa-OH for each coupling, that is not a problem in terms of elongation yields. As a consequence, a **final piperidine treatment** (standard 3 x 3 min, 20% in DMF or NMP) is thus necessary even if the N-terminal residue is not Fmoc-protected (Boc, acetyl, biotin, etc...).¹⁸ Moreover, UV titration of the fluorenylmethyl-piperidine adduct after Fmoc deprotection is useless unless using a prior treatment with hydroxylamine/imidazole⁵⁰ for selective ester cleavage before piperidine treatment (supporting information, p S3).

Conflict of interest

There are no conflicts to declare.

Acknowledgements

We thank Dr Guillaume Gabant, Dr Cyril Colas and the mass spectrometry platforms of CBM and FR2708 federation for the MS and LC/HRMS analyses, as well as Hervé Meudal and the CBM nuclear magnetic resonance platform for assistance in the recording of NMR spectra and their analyses. We also thank Jean-Baptiste Madinier for his assistance in automated peptide synthesis. University of Orléans and doctoral school SSBCV (ED549) are gratefully acknowledged for a PhD fellowship for SAA, as well as French national funding agency ANR (EasyMiniprot project, ANR-15-CE07-0022) for financial support.

Notes and references

- a) P. Dawson, T. Muir, I. Clark-Lewis, S. Kent, *Science*, 1994, **266**, 776–779; b) V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoët, J.-C. M. Monbaliu, O. Melnyk, *Chem. Rev.*, 2019, **119**, 7328–7443.
- Note that Pentelute and his co-workers recently reported a flow-based approach for the synthesis of polypeptides of more than 200 aa See for example: N. Hartrampf, A. Saebi, M. Poskus, Z. P. Gates, A. J. Callahan, A. E. Cowfer, S. Hanna, S. Antilla, C. K. Schissel, A. J. Quartararo, X. Ye, A. J. Mijalis, M. D. Simon, A. Loas, S. Liu, C. Jessen, T. E. Nielsen, B. L. Pentelute, *Science*, 2020, **368**, 980–987.
- For a database dedicated to protein chemical synthesis, see: V. Agouridas, O. El Mahdi, M. Cargoët, O. Melnyk, *Bioorg. Med. Chem.*, 2017, **25**, 4938–4945.
- See for example: a) M. T. Weinstock, M. T. Jacobsen, M. S. Kay, *Proc. Natl. Acad. Sci. USA*, 2014, **111**, 11679–11684; b) A. Pech, J. Achenbach, M. Jahnz, S. Schülzchen, F. Jarosch, F. Bordusa, S. Klussmann, *Nucleic Acids Res.*, 2017, **45**, 3997–4005; c) J. Weidmann, M. Schnölzer, P. E. Dawson, J. D. Hoheisel, *Cell Chem. Biol.*, 2019, **26**, 645–651; d) M. Wang, W. Jiang, X. Liu, J. Wang, B. Zhang, C. Fan, L. Liu, G. Peña-Alcántara, J.-J. Ling, J. Chen, T. F. Zhu, *Chem.*, 2019, **5**, 848–857; e) H. Sun, A. Brik, *Acc. Chem. Res.*, 2019, **52**, 12, 3361–3371.
- a) F. Mende, O. Seitz, *Angew. Chem. Int. Ed.*, 2011, **50**, 1232–1240; b) H. Li, S. Dong, *Sci. China Chem.*, 2017, **60**, 201–213.
- G.-M. Fang, Y.-M. Li, F. Shen, Y.-C. Huang, J.-B. Li, Y. Lin, H.-K. Cui, L. Liu, *Angew. Chem. Int. Ed.*, 2011, **50**, 7645–7649.
- a) J. B. Blanco-Canosa, P. E. Dawson, *Angew. Chem. Int. Ed.*, 2008, **47**, 6851–6855; b) J. B. Blanco-Canosa, B. Nardone, F. Albericio, P. E. Dawson, *J. Am. Chem. Soc.*, 2015, **137**, 7197–7209.
- E. C. B. Johnson, S. B. H. Kent, *J. Am. Chem. Soc.*, 2006, **128**, 6640–6646.
- For peptide hydrazides, see for example: a) P. Siman, S. V. Karthikeyan, M. Nikolov, W. Fischle, A. Brik, *Angew. Chem. Int. Ed.*, 2013, **52**, 8059–8063; b) S.K. Mong, A.A. Vinogradov, M.D. Simon, B.L. Pentelute, *ChemBioChem*, 2014, **15**, 721–733; c) X. Tian, J. Li, W. Huang, *Tetrahedron Lett.*, 2016, **57**, 4264–4267; For Nbz, see for example: d) S. K. Mahto, C. J. Howard, J. C. Shimko, J. J. Ottesen, *Chembiochem*, 2011, **12**, 2488–2494; e) J. Mannuthodikayil, S. Singh, A. Biswas, A. Kar, W. Tabassum, P. Vydyam, M. K. Bhattacharyya, K. Mandal, *Org. Lett.*, 2019, **21**, 9040–9044.
- For recent examples, see: a) H. E. Elashal, Y. E. Sim, M. Raj, *Chem. Sci.*, 2017, **8**, 117–123; b) D. T. Flood, J. C. J. Hintzen, M. J. Bird, P. A. Cistrone, J. S. Chen, P. E. Dawson, *Angew. Chem. Int. Ed.*, 2018, **57**, 11634–11639.
- See for examples: a) T. Kawakami, M. Sumida, K. Nakamura, T. Vorherr, S. Aimoto, *Tetrahedron Lett.*, 2005, **46**, 8805–8807; b) N. Ollivier, J. B. Behr, O. El-Mahdi, A. Blanpain, O. Melnyk, *Org. Lett.*, 2005, **7**, 2647–2650; c) Nagaike, F., Onuma, Y., Kanazawa, C., Hojo, H., Ueki, A., Nakahara, Y., Nakahara, Y. *Org. Lett.*, 2006, **8**, 4465–4468; d) Y. Ohta, S. Itoh, A. Shigenaga, S. Shintaku, N. Fujii, A. Otaka, *Org. Lett.*, 2006, **8**, 467–470; e) H. Hojo, Y. Onuma, Y. Akimoto, Y. Nakahara, Y. Nakahara, *Tetrahedron Lett.*, 2007, **48**, 25–28; f) Kawakami, T., Aimoto, S. *Chem. Lett.*, 2007, **36**, 76–77; g) J. Kang, J. P. Richardson, D. Macmillan, *Chem. Commun.*, 2009, 407–409; h) S. Tsuda, A. Shigenaga, K. Bando, A. Otaka, *Org. Lett.*, 2009, **11**, 823–826; i) N. Ollivier, J. Dheur, R. Mhidia, A. Blanpain, O. Melnyk, *Org. Lett.*, 2010, **12**, 5238–5241; j) W. Hou, X. Zhang, F. Li, C.-F. Liu, *Org. Lett.*, 2011, **13**, 386–389; k) K. Sato, A. Shigenaga, K. Tsuji, S. Tsuda, Y. Sumikawa, K. Sakamoto, A. Otaka, *ChemBioChem*, 2011, **12**, 1840–1844; l) R. Sharma, J. P. Tam, *Org. Lett.*, 2011, **13**, 5176–5179; m) J. S. Zheng, H. N. Chang, F. L. Wang, L. Liu, *J. Am. Chem. Soc.*, 2011, **133**, 11080–11083; n) F. Burlina, G. Papageorgiou, C. Morris, P. D. White, J. Offer, *Chem. Sci.*, 2014, **5**, 766–770; o) Y. Asahina, K. Nabeshima, H. Hojo, *Tetrahedron Lett.*, 2015, **56**, 1370–1373; p) V. P. Terrier, H. Adihou, M. Arnould, A. F. Delmas, V. Aucagne, *Chem. Sci.*, 2016, **7**, 339–345; q) P. M. M. Shelton, C. E. Weller, C. Chatterjee, *J. Am. Chem. Soc.*, 2017, **139**, 3946–3949; r) M. Yanase, K. Nakatsu, C. J. Cardos, Y. Konda, G. Hayashi, A. Okamoto, *Chem. Sci.*, 2019, **10**, 5967–5975; s) K. Nakatsu, M. Yanase, G. Hayashi, A. Okamoto, *Org. Lett.*, 2020, **22**, 4670–4674.

- K. Sakamoto, K. Sato, A. Shigenaga, K. Tsuji, S. Tsuda, H. Hibino, Y. Nishiuchi, A. Otaka, *J. Org. Chem.*, 2012, **77**, 6948–6958.
- Y. Asahina, H. Hojo, *J. Org. Chem.*, 2020, **85**, 1458–1465.
- C. A. Hunter, *Angew. Chem. Int. Ed.*, 2004, **43**, 5310–5324.
- Note that DFT calculations on *N*-Hnb-Cys mechanism were undertaken by others, see: S. Bi, P. Liu, B. Ling, X. Yuan, Y. Jiang, *Chin. Chem. Lett.*, 2018, **29**, 1264–1268.
- a) T. Johnson, M. Quibell, D. Owen, R.C. Sheppard, *J. Chem. Soc., Chem. Commun.*, 1993, 369–372. b) L. P. Miranda, W. D. Meutermans, M. L. Smythe, P. F. Alewood, *J. Org. Chem.*, 2000, **65**, 5460–5468.
- a) D. Lelièvre, V. P. Terrier, A. F. Delmas, V. Aucagne, *Org. Lett.*, 2016, **18**, 920–923; b) G. Martinez, J.-P. Hograindleur, S. Voisin, R. Abi Nahed, T. M. Abd El Aziz, J. Escoffier, J. Bessonnat, C.-M. Fovet, M. De Waard, S. Hennebicq, V. Aucagne, P. F. Ray, E. Schmitt, P. Bulet, C. Arnoult, *Mol. Hum. Reprod.*, 2016, **10**, 116–131; c) V. P. Terrier, A. F. Delmas, V. Aucagne, *Org. Biomol. Chem.*, 2017, **15**, 316–319; d) K. Loth, A. Vergnes, C. Barreto, S. N. Voisin, H. Meudal, J. Da Silva, A. Bressan, N. Belmadi, E. Bachère, V. Aucagne, C. Cazevielle, H. Marchandin, R. D. Rosa, P. Bulet, L. Touqui, A. F. Delmas, D. Destoumieux-Garzón, *mBio*, 2019, **10**, e01821–19.
- L. De Rosa, R. Di Stasi, L. D. D’Andrea, *Tetrahedron*, 2019, **75**, 894–905.
- H. Rink, *Tetrahedron Lett.*, 1987, **28**, 3787–3790.
- Note that we tried different other conditions found in the literature for solid-supported reductive aminations, involving either a one pot process or two consecutive steps, but with inferior results, resulting in either uncomplete reaction, bis-alkylation or to other unidentified byproducts. See reference.^{11p}
- F. García-Martín, M. Quintanar-Audelo, Y. García-Ramos, L. J. Cruz, C. Gravel, R. Furic, S. Côté, J. Tulla-Puche, F. Albericio, *J. Comb. Chem.*, 2006, **8**, 213–220.
- In our seminal study, a notable side-product was only observed when introducing a Cys(Trt) as the first amino-acid of the peptide segment, which we attributed to epimerization considering the know propensity of Cys to epimerize during coupling (see reference 27). We also observed an expected co-product after coupling of Asn, due to incomplete deprotection of the *N*-Trt group (see: M. Friede, S. Denery, J. Neimark, S. Kieffer, H. Gausepohl, J. P. Briand, *Pept. Res.*, 1992, **5**, 145–147).
- a) A. Czarnik, *Biotechnol. Bioeng.*, 1998, **61**, 77–79. b) Y. Feng, K. Burgess, *Biotechnol. Bioeng.*, 2000, **71**, 3–8.
- <https://spheritech.com/spheritide-Ks/>.
- J.H. Adams, R.M. Cook, D. Hudson, V. Jammalamadaka, M.H. Lyttle, M.F. Songster, *J. Org. Chem.*, 1998, **63**, 3706–3716.
- See for example: a) N. Robertson, L. Jiang, R. Ramage, *Tetrahedron*, 1999, **55**, 2713–2720. b) W. Van Den Nest, S. Yuval, F. Albericio, *J. Pept. Sci.*, 2001, **7**, 115–120. c) Y.E. Jad, S.N. Khattab, B.G. de la Torre, T. Govender, H.G. Kruger, A. El-Faham, F. Albericio, *Org. Biomol. Chem.*, 2014, **12**, 8379–8385.
- See for example: a) T. Kaiser, G. J. Nicholson, H. J. Kohlbau, W. Voelter, *Tetrahedron Lett.*, 1996, **37**, 1187–1190; b) Y. Han, F. Albericio, G. Barany, *J. Org. Chem.*, 1997, **62**, 4307–4312.
- a) H. Hibino, Y. Nishiuchi, *Tetrahedron Lett.*, 2011, **52**, 4947–4949.
- Note that in our previously reported work, a model peptide generated using this device was proven to be epimerization-free after NCL (see reference 11p).
- R. Huisgen, *Angew. Chem.* 1955, **67**, 360–360.
- a) D. A. Jaeger, D. J. Cram, *J. Am. Chem. Soc.*, 1971, **93**, 5153–5161; b) T. F. Buckley, H. Rapoport, *J. Am. Chem. Soc.*, 1982, **104**, 4446–4450; c) G. Cainelli, D. Giacomini, A. Trerè, P. P. Boyl, *J. Org. Chem.*, 1996, **61**, 5134–5139.
- C. G. Booramra, K. M. Burrow, L. A. Thompson, J. A. Ellman, *J. Org. Chem.*, 1997, **62**, 1240–1256.
- M. Petrisko, J. Krupka, *Res. Chem. Intermed.*, 2005, **31**, 769–778.
- A. Lawson, J. O. Stevens, *J. Chem. Soc.*, 1968, 1514–1515.
- See for example: a) H. B. Dixon, V. Moret, *Biochem J.* 1965, **94**, 463–469; b) H. B. Dixon, R. Fields, *Methods Enzymol.*, 1972, **25**, 409–419; c) T. Hara, A. Tainosho, K. Nakamura, T. Sato, T. Kawakami, S. Aimoto, *J. Pept. Sci.*, 2009, **15**, 369–376.
- E. W. Baxter, A. B. Reitz, in *Organic Reactions*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2002, **59**, pp. 1–714.
- Note that we also tested the use of Lewis acids in combination with NaBH₃CN, and tried to conduct the reaction in THF/DMF, but this led either to a greater amount of epimerization, or the formation of unidentified byproducts. See supporting information, table S1, entries 1, 6 and 8.
- Recording of the NMR spectra at higher temperatures did not afford a single peak set, but rather a massive loss of resolution, consistent with slow conformational exchange in the NMR timescale.
- In the case of the **10c/10d** mixture, available in larger quantity than pure **10a-c**, we were able to record a ¹H-¹³C HMBC spectrum which confirmed that the newly appeared CH signals were coupled to carbon signals of the phenol ring, see supporting information p S41.
- a) P. Gomes, M. J. Araújo, M. Rodrigues, N. Vale, Z. Azevedo, J. Iley, P. Chambel, J. Morais, R. Moreira, *Tetrahedron*, 2004, **60**, 5551–5562; b) A. L. Satz, J. Cai, Y. Chen, R. Goodnow, F. Gruber, A. Kowalczyk, A. Petersen, G. Naderi-Oboodi, L. Orzechowski, Q. Strebel, *Bioconjugate Chem.*, 2015, **26**, 1623–1632; c) Y. Wang, A. Ø. Madsen, F. Diness, M. Meldal, *Chem. Eur. J.*, 2017, **23**, 13869–13874.
- W. P. Unsworth, G. Coulthard, C. Kitsiou, R. J. K. Taylor, *J. Org. Chem.*, 2014, **79**, 1368–1376.
- a) V. Adebomi, R. D. Cohen, R. Wills, H. A. H. Chavers, G. E. Martin, M. Raj, *Angew. Chem. Int. Ed.*, 2019, **58**, 19073–19080. b) B. Koo, N. S. Dolan, K. Wucherer, H. K. Munch, M. B. Francis, *Biomacromolecules*, 2019, **20**, 3933–3939.
- Relative amounts of **10a-d** depend on the reaction conditions (solvent, additive, reducing agent etc.), probably because of variable amounts of Cys epimerization and low diastereoselectivity of the cyclization.
- ¹H NOESY and ROESY analyses were performed to try to determine distereoisomers configurations, however signals overlap did not allow stereochemical attribution.
- Note that we also tested the use of Lewis acids in combination with NaBH₄, and tried alternative reducing agents, but this led either to a greater amount of epimerization, or the formation of other byproducts. See supporting information, table S1.
- See for example: a) V. Aucagne, I. E. Valverde, P. Marceau, M. Galibert, N. Dendane, A. F. Delmas, *Angew. Chem. Int. Ed.* 2012, **51**, 11320–11324; b) I. E. Decostaire, D. Lelièvre, V. Aucagne, A. F. Delmas, *Org. Biomol. Chem.*, 2014, **12**, 5536–5543; c) M. Galibert, V. Piller, F. Piller, V. Aucagne, A. F. Delmas, *Chem. Sci.*, 2015, **6**, 3617–3623.
- For recent examples of solid-supported aminations of peptides, see: a) A.-B. M. Abdel-Aal, G. Papageorgiou, M. Quibell, J. Offer, *Chem. Commun.*, 2014, **50**, 8316–8319; b) K. Pels, T. Kodadek, *ACS Comb. Sci.*, 2015, **17**, 152–155; c) A.-B. M. Abdel-Aal, G. Papageorgiou, R. Raz, M. Quibell, F. Burlina, J. Offer, *J. Pept. Sci.*, 2016, **22**, 360–367; d) S. F. Loibl, A. Dallmann, K. Hennig, C. Judds, O. Seitz, *Chem. Eur. J.*, 2018, **24**, 3623–3633. e) F. Burlina, A.-B. M. Abdel-Aal, R. Raz, I. Pinzuti, G. Papageorgiou, J. Li, R. Antobus, S. R. Martin, S. Kunzelmann, B. Stieglitz, J. Offer, *Commun. Chem.*, 2019, **2**, 111.
- See for example: a) M. Jbara, E. Eid, A. Brik, *Org. Biomol. Chem.*, 2018, **16**, 4061–4064; b) N. Kamo, G. Hayashi, A. Okamoto, *Chem. Commun.*, 2018, **54**, 4337–4340.
- Dr Maria Volkova, personal communication.

J. J. Díaz-Mochón, L. Bialy, M. Bradley, *Org. Lett.*, 2004, **6**, 1127–1129.

V. Supporting Information

These results have been submitted (22 Aug 2020) and accepted (18 Sep 2020) for publication in Organic & Biomolecular Chemistry. Here, we reproduced the corresponding supporting information file as published.

An optimized protocol for the synthesis of *N*-2-hydroxybenzyl-cysteine peptide crypto-thioesters.

Skander A. Abboud and Vincent Aucagne*

Centre de Biophysique Moléculaire, CNRS UPR 4301, Rue Charles Sadron 45071 Orléans cedex 2,
France. Email : aucagne@cnrs-orleans.fr

Supporting Information

1-General information.....	249
2-General procedures for manual and automatic SPPS.....	250
3-Optimization of the reductive amination and coupling of first amino acid	251
3-1- Synthesis of cysteinyl peptide resins 1a-e	251
3-2- Solid-supported reductive amination	252
3-3 Characterization of compounds 4 and 5	255
4-Synthesis of HPLC standards 7a , 7c , 7d and 7f	256
5-Identification of the origin of the Cys epimerization	262
6-Effect of temperature and concentration on Cys epimerization during reductive amination	264
7-NMR characterisation of the [M-2] imidazolidinone byproducts 10a-d	264
7-1- Synthesis and isolation of products 4 , 5 and 10a-d	264
7-2 NMR analysis of 4 , 5 and 10a-d	270
7-3 Structure elucidation of imidazolinones 10a-d	280
7-4- Comparison with ¹ H and ¹³ C spectra of previously reported <i>N</i> ¹ -acyl- <i>N</i> ³ -alkyl-2-aryl-imidazolidin-4-ones	289
8-Optimization of the coupling of the first amino acid.....	290
8-1- Cysteine	290
8-2- Histidine	291
8-3 Serine	292
9-Quantification of the reductive amination yield.....	292
10- Examples of crypto-thioesters synthesized with the optimized method	293
11-Investigation of the formation of piperidinylalanine byproduct	295
12-References.....	297

1- General information

All reagents and solvents were used without further purification. Fmoc-Rink MBHA polystyrene resin, Fmoc-Rink linker, aminomethyl NovaGel resin and Oxyma were purchased from Merck Biosciences (Nottingham, UK). Protected amino acids were purchased from Gyros Protein Technology (Uppsala, Sweden). Aminomethyl TentaGel R resin was purchased from Rapp polymers (Tuebingen, Germany). HCTU and HATU were purchased from IRIS Biotech GmbH (Marktredwitz, Germany). DIEA was purchased from Carlo ERBA (Val-de-Reuil, France). Rink amide ChemMatrix resin was purchased from Biotage (Uppsala, Sweden). Spheritide resin was a gift from CEM (Orsay, France). Peptide synthesis grade DMF was obtained from VWR (Fontenay-sous-Bois, France). All other chemicals were from Sigma Aldrich (St-Quentin-Fallavier, France) and solvents from SDS-Carlo Erba (Val de Reuil, France). Ultrapure water was obtained using a Milli-Q water system from Millipore (Molsheim, France). Polypropylene syringes fitted with polypropylene frits were obtained from Torviq (Niles, MI, USA) and were equipped with PTFE stopcock bought from Biotage. ^1H and ^{13}C NMR spectra were recorded on a Bruker AVANCE III 600 instrument, at a constant temperature of 25°C. Chemical shifts are reported in parts per million from low to high field and referenced to tetramethylsilane (TMS). Coupling constants (*J*) are reported in hertz (Hz). Standard abbreviations indicating multiplicity were used as follows: s = singlet, d = doublet, dd = doublet of doublets, m = multiplet, b = broad signal. HPLC analyses were carried out on a Hitachi Chromaster system equipped with a 5160 pump, a 5430 diode array detector and a 5260 auto sampler and semi-preparative purifications were carried out on a Hitachi LaChromElite system equipped with a L-2130 pump, a L-2455 diode array detector and a L-2200 auto sampler. Chromolith High Resolution RP-18e (150 Å, 10 × 4.6 mm, 3 mL/min flow rate) columns were used for analysis, Nucleosil C18 (300 Å, 5 µm, 250 × 10 mm, 3 mL/min flow rate) and Jupiter C4 (300 Å, 5 µm, 250 × 10 mm, 3 mL/min flow rate) for purification. Solvents A and B are 0.1 % TFA in H₂O and 0.1 % TFA in MeCN, respectively.) Each gradient was followed by a washing step to elute any compound not eluted during the gradient (up to 95% B/A over 0.5 min, then isocratic 95% B/A over 0.5 min for Chromolith). LC-ESI-MS analyses were carried out on an Agilent 1260 Infinity HPLC system, coupled with a 6120 mass spectrometer, and fitted with an Aeris Widepore XB-C18 2 (3.6 µm, 150 × 2.1 mm, 0.5 mL/min flow rate, 60°C) column. The reported *m/z* values correspond to the monoisotopic ions if not specified otherwise. Solvents A' and B' were 0.1 % formic acid in H₂O and 0.1 % formic acid in MeCN, respectively. Gradient: 3% B'/A' for 1 min, then 3 to 50% B'/A' over 15 min. Low resolution MS of pure compounds were obtained using this system. High-resolution ESI-MS analyses were performed on a maXisTM ultra-high-resolution Q-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany), using the positive mode. Relative rates of Hnb-containing compounds were determined by HPLC peak integration at $\lambda = 320$ nm which is a wavelength of maximum absorption (λ_{max}) for the Hnb group.

2-General procedures for manual and automatic SPPS

Protocol PS1 – peptide elongation: Manual couplings were performed on polypropylene syringes fitted with polypropylene frits using rotation stirring. Fmoc-based solid phase peptide syntheses (SPPS) were carried out on a Prelude synthesizer from Protein technologies. The side-chain protecting groups used were Arg(Pbf), Asp(OtBu), Cys(Acm), Cys(Trt), Cys(StBu), Glu(OtBu), His(Trt), Ser(tBu), Thr(tBu), Trp(Boc) and Tyr(tBu). Syntheses were performed on a 0.025 mmol-per-reactor scale. Protected amino acids (0.25mmol, 10 equiv.) were coupled using HCTU (98 mg, 0.238 mmol, 9.5 equiv.) and *i*Pr₂NEt (87 μ L, 0.5 mmol, 20 equiv.) in NMP (3 mL) for 30 min. Coupling of Fmoc-Ser(tBu)-OH *N*-Hnb-cysteine secondary amine were performed for 18 h. For recommended coupling protocols of the different C-terminal amino acids of peptide segments, see the experimental part of the article. Capping of eventual unreacted amine groups was achieved by treatment with acetic anhydride (143 μ L, 1.51 mmol, 60 equiv.), *i*Pr₂NEt (68 μ L, 0.39 mmol, 15.5 equiv.) and HOBT (6 mg, 0.044 mmol, 1.8 equiv.) in NMP (3 mL) for 7 min (4 x 7 min in the case of *N*-Hnb-cysteine secondary amine). Fmoc group was deprotected by three successive treatments with 20% piperidine in NMP (3 mL) for 3 min. As stated in the experimental section of the article, variable amount (5-90 %) of *O*-acylation of the Hnb phenol are observed after each coupling. This ester is cleaved within seconds upon piperidine treatment during Fmoc deprotection. As a consequence, a final piperidine treatment (standard 3 x 3 min, 20% in DMF or NMP) is thus necessary even if the N-terminal residue is not Fmoc-protected.

Protocol PS2 – Cleavage: The crude peptide was deprotected and cleaved from the resin through a treatment with TFA/H₂O/*i*Pr₃SiH/phenol, 88:5:2:5 for 2 h, and the peptide was precipitated by dilution into an ice-cold diethyl ether/petroleum ether 1:1 mixture, recovered by centrifugation and washed twice with diethyl ether.

Protocol PS3 – Small peptides cleavage (less than 4 residues): they were deprotected and cleaved from the resin through a treatment with TFA/H₂O/*i*Pr₃SiH, 93:5:2 for 30 min following by concentration under vacuum.

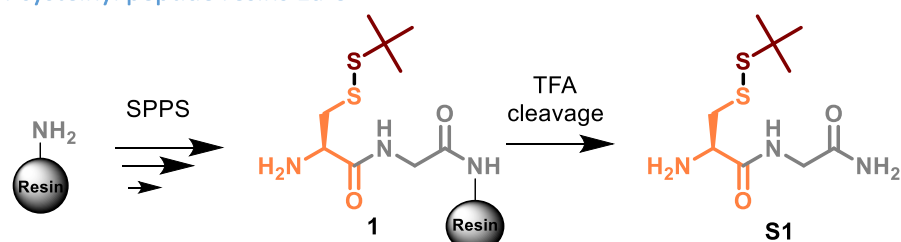
Protocol PS4 – Procedure for selective Hnb ester cleavage to allow UV titration of Fmoc deprotection: As a consequence of the formation of variable amount (5-90 %) of *O*-acylated Hnb during each coupling, this ester being cleaved upon piperidine treatment during Fmoc deprotection, standard UV titration of the fluorenylmethyl-piperidine adduct after Fmoc deprotection is useless unless using a prior treatment for selective ester cleavage before piperidine treatment.

Ester cleavage mixture was prepared as follows: 1.25 g (1.80 mmol) of NH₂OH·HCl and 0.918 g (1.35 mmol) of imidazole were suspended in 5 mL of NMP, the mixture was sonicated until complete dissolution. This solution can be stored for few months at -20°C. 5 volumes of this solution is diluted with 1 volume of DCM prior to utilization, and the resin is treated for 3x20 min for quantitative ester cleavage.

The fluorenylmethyl-piperidine adduct is quantified by UV spectroscopy at $\lambda = 301$ nm ($\epsilon = 7800$ mol⁻¹ cm⁻¹) in order to evaluate the Fmoc SPPS elongation yield of crypto-thioester peptides.

3-Optimization of the reductive amination and coupling of first amino acid

3-1- Synthesis of cysteinyl peptide resins 1a-e



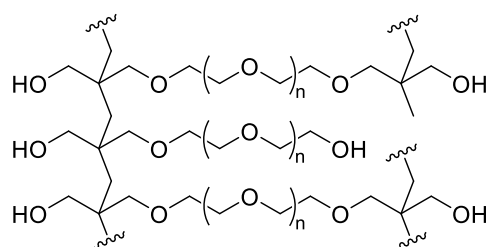
1a-Resin = Rink ChemMatrix

1b-Resin = Rink Polystyrene

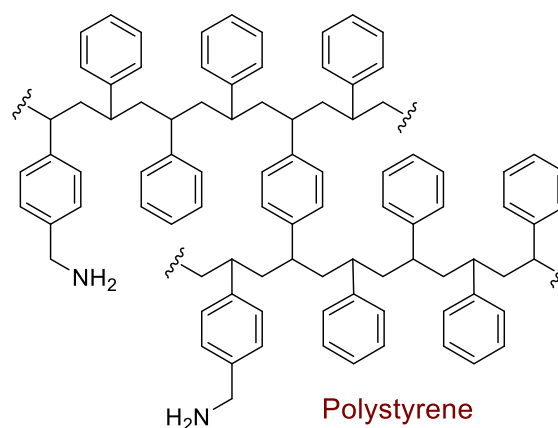
1c-Resin = Rink Tentagel

1d-Resin = Rink Novagel

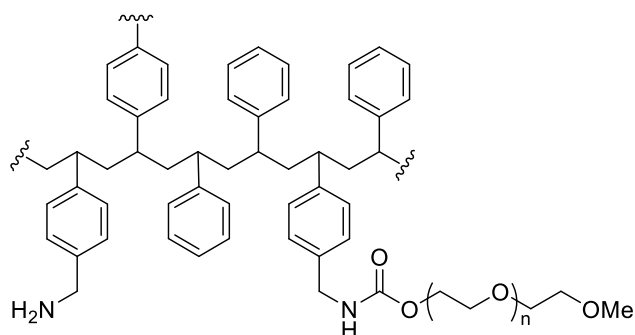
1e-Resin = Rink Spheritide



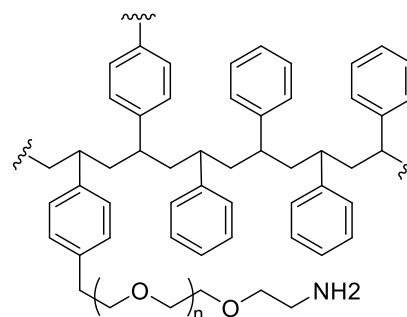
ChemMatrix



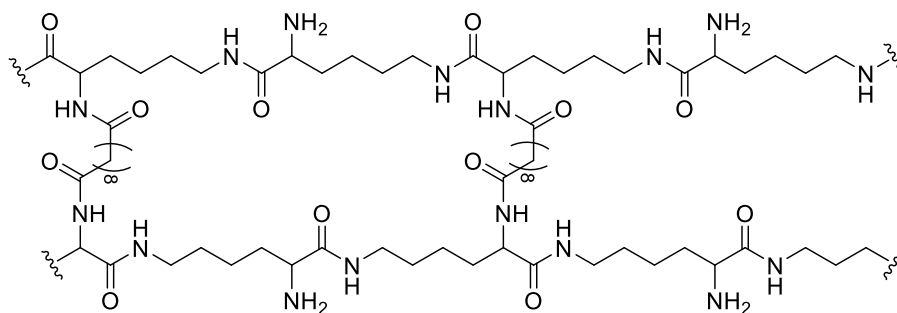
Polystyrene



Novagel



Tentagel



Spheritide

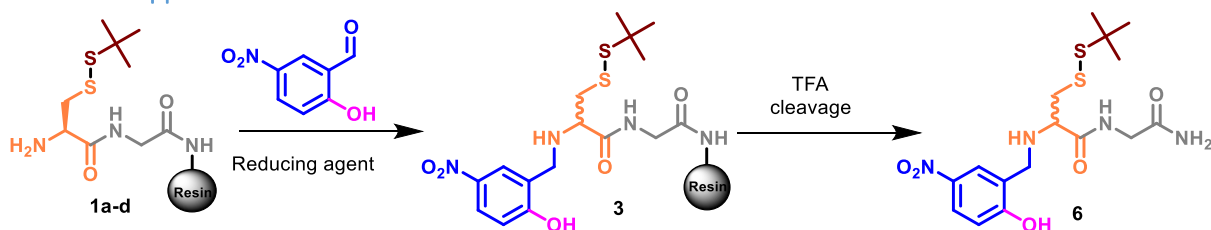
Peptidyl resins **1a-e** were obtained through manual SPPS (protocol PS1) starting from different resins. Resins loading: Rink ChemMatrix: 0.54 mmol/g, Rink-MBHA polystyrene: 0.89 mmol/g, Tentagel R: 0.164 mmol/g, Novagel: 0.77 mmol/g, Rink Sphertide: 0.15 mmol/g. In the cases of Tentagel and Novagel, Fmoc-Rink linker was coupled manually. Small amounts of peptide-resins were cleaved for analytical purpose (protocol PS3 - small peptides cleavage) to give peptide **S1**.

- H-Cys(StBu)-Gly-NH₂ (**S1**)

ESI-MS (m/z): [MH]⁺ calcd. for C₉H₂₀N₃O₂S₂: 266.1, found: 266.1

HPLC analysis: t_R = 2.14 min (Chromolith, gradient: 1-50% B/A over 15 min)

3-2- Solid-supported reductive amination



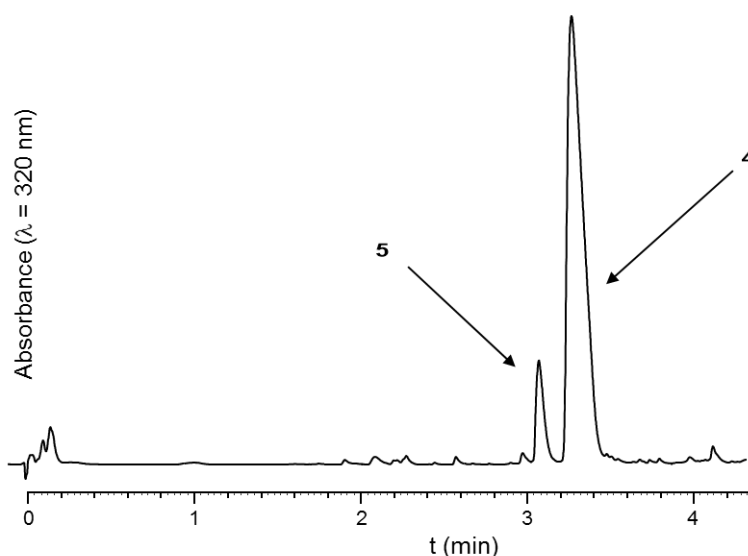
- H-(Hnb)Cys(StBu)-Gly-NH₂ (**6**)

ESI-MS (m/z): [MH]⁺ calcd. for C₁₆H₂₅N₄O₅S₂: 417.1, found: 417.0

HPLC analysis: t_R = 6.8 min (Aeris Widespore XB-C18 2, gradient: 3-50% B'/A' over 15 min)

Protocol PS5-1: Previously reported reductive amination protocol with NaBH₃CN [1] (see main text, Table 1, entry 1)

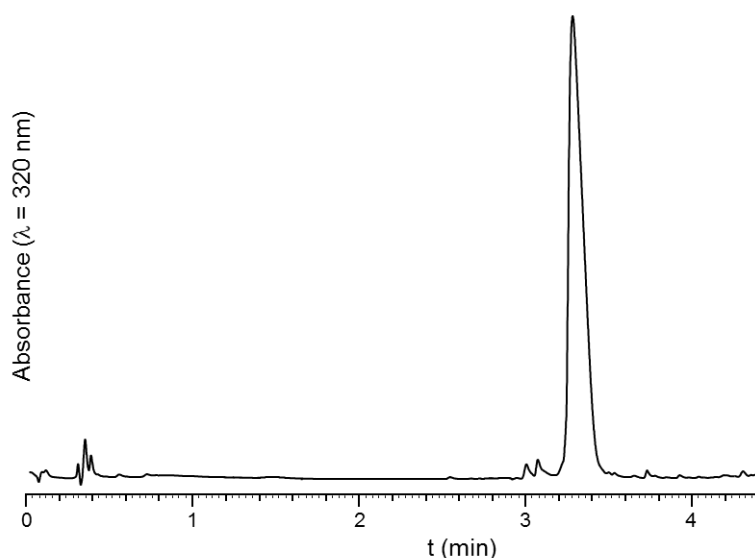
25 μ mol of H-Cys(StBu)-Gly-Rink-resin (**1a-d**) was washed with 1:1 DMF/MeOH (4 x 3 mL, 30 s) then treated with 9:9:2 DMF/MeOH/AcOH mixture (3 mL, 5 min). The resin was washed with 1:1 DMF/MeOH (3 x 3 mL, 30 s) and 2-hydroxy-5-nitrobenzaldehyde (42 mg, 10 equiv.) in 2 mL 1:1 DMF/MeOH (125 mM aldehyde concentration) was then added, and the reactor was stirred for 1 h. The reactor was drained and the resin was washed with 1:1 DMF/MeOH (3 x 3 mL, 5 s). Without delay, a fresh solution of sodium cyanoborohydride (32 mg, 20 equiv.) in 3 mL 9:9:2 DMF/MeOH/AcOH (250 mM, NaBH₃CN concentration) was added and the reactor stirred for 1 h. The reactor was drained and the resin was washed with 1:1 DMF/MeOH (4 x 3 mL, 30 s), NMP (3 x 3 mL, 30 s), 20% piperidine in NMP (3 x 3 mL, 3 min), NMP (3 x 3 mL, 30 s), dichloromethane (3 x 5 mL, 30 s) and NMP (2 x 3 mL, 30 s). Peptidyl resins were cleaved using protocol PS2 - small peptides cleavage.



Supplementary figure S1: HPLC trace (Chromolith, gradient: 5-50% B/A over 5 min) of **4** (resin: ChemMatrix) obtained using previously reported two-steps reductive amination with NaBH_3CN protocol PS2-1 corresponding to table 1, entry 1)

Protocol PS5-2 : Optimized reductive amination protocol with NaBH_4 (see main text, table 6, entry 1)

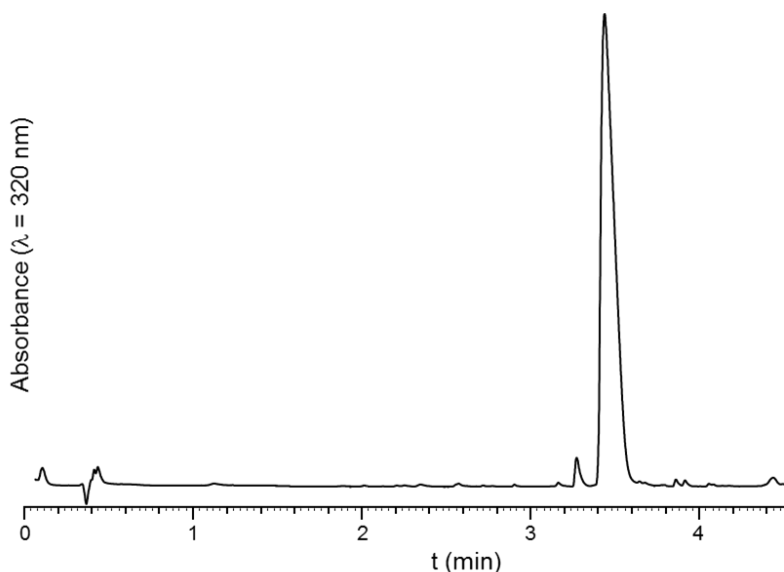
25 μmol of H-Cys(StBu)-Gly-Rink-resin (**1a-c**) was washed with 1:1 DMF/MeOH (4 x 3 mL, 30 s). 2-Hydroxy-5-nitrobenzaldehyde (42 mg, 10 equiv.) in 2 mL 44.5:44.5:1 DMF/MeOH/AcOH (125 mM aldehyde concentration) was then added, and the reactor was stirred for 5 min. The reactor was drained and the resin was washed with 1:1 DMF/MeOH (3 x 3 mL, 5 s) then DMF (3 x 3 mL, 5 s). Without delay, a fresh solution of sodium borohydride (19 mg, 20 equiv.) in 2 mL DMF (250 mM borohydride concentration) was added and the reactor was stirred for 20 min. The reactor was drained and the resin was washed with DMF (4 x 3 mL, 30 s), 20% v/v piperidine in NMP (3 x 3 mL, 3 min), NMP (3 x 3 mL, 30 s), dichloromethane (3 x 5 mL, 30 s) and NMP (3 x 3 mL, 30 s). Peptidyl resins were cleaved using protocol PS2 - small peptides cleavage.



Supplementary figure S2: HPLC trace (Chromolith, gradient: 5-50% B/A over 5 min) of **4** obtained using two-steps reductive amination with NaBH₄

Protocol PS5-3 : Optimized one pot reductive amination protocol with borane pyridine complex (see main text, table 5, entry 5)

25 μmol of H-Cys(StBu)-Gly-Rink-ChemMatrix **1a** was washed with 3 mL of a 1:1 DMF/MeOH mixture (4 x 3 mL, 30 s). A solution of 2-hydroxy-5-nitrobenzaldehyde (42 mg, 10 equiv.) in 1 mL 1:1 DMF/MeOH (250 mM, aldehyde concentration) was prepared, next acetic acid (20 μL, 14 equiv., 1% v/v final) was added and finally a solution of reducing agent (46.5 mg, 20 equiv.) in 1 mL DMF/MeOH (250 mM final borane concentration) was added. The mixture was added to the resin without delay, and the reactor was stirred for 20 min. The reactor was drained and the resin was washed with 1:1 DMF/MeOH (4 x 3 mL, 30 s), NMP (3 x 3 mL, 30 s), 20% piperidine in NMP (3 x 3 mL, 3 min), NMP (3 x 3 mL, 30 s), dichloromethane (3 x 5 mL, 30 s) and NMP (3 x 3 mL, 30 s).



Supplementary figure S3: HPLC trace (Chromolith, gradient: 5-50% B/A over 5 min) of **4** obtained through one-pot reductive amination with borane pyridine complex

- Other reductive amination protocols tried not detailed in the manuscript

❖ Table S1: Different other reductive amination conditions tested following literature protocols^[a]

Entry	Ref.	One pot / two steps	Reducing agent	Additive (equiv.)	Desired product 4 (%) ^[b]	Epimer 5 (%) ^[b]	Imidazolidinones 10 (%) ^[b]	Other side products (%) ^[b] [<i>m/z</i>] ^[c]
1 ^[d]	[2]	One pot	NaBH ₃ CN	SnCl ₂ (20)	91	6	-	3 [447.0]
2 ^[e]	[3]	Two steps	BH ₃ ·Me ₂ NH	-	66	4.5	25	4.5 [534.1]
3 ^[f]	[4]	Two steps	BH ₃ ·Me ₂ S	-	90	9	0.5	-
4 ^[g]	[5]	Two steps	NaBH ₄	ZnCl ₂ (15)	75	8	-	16 [534.1]
5 ^[g]	[6]	Two steps	NaBH ₄	MgCl ₂ (10)	95	2.5	2	1.6 [534.1]
6 ^[h]	[7]	One pot	NaBH(OAc) ₃	-	<2	-	-	<98 [581.9]
7 ^[i]	[8]	Two steps	NaBH ₄	TiCl ₄ (4)	51	5	-	27 [534.1] 17 [409.1]
8 ^[j]	[9]	One pot	NaBH ₃ CN	Ti(<i>i</i> PrO) ₄ (2)	14	0.7	-	70.1 [353.0] 15.2 [385.1]

[a] All reactions were conducted at room temperature at a 12.5 mM peptidyl resin concentration using 20 equiv. of the reducing agent; [b] Relative rates determined by HPLC peak integration at $\lambda = 320$ nm; [c] *m/z* not attributed to putative structures except for the [M-2Da] byproducts that likely correspond to imidazolidinones **10a-d**; [d] 5 min reaction in MeOH; [e] Imination in DMF/MeOH 1:1 for 5 min, reduction in DMF/MeOH 1:1 for 5 min; [f] Imination in DMF/MeOH 1:1 for 5 min, reduction in THF for 5 min; [g] Imination in DMF for 5 min, reduction in DMF for 5 min; [h] 1 h reaction in THF/DMF 1:1; [i] Imination in THF for 5 min, reduction in THF for 5 min; [j] 5 min reaction in THF.

The different conditions described in table S1 were inspired from the following reports. Note that in some cases, reaction conditions were slightly modified and adapted to the current work.

3-3 Characterization of compounds 4 and 5.

- H-Ser-(Hnb)Cys(StBu)-Gly-NH₂ (**4**)

ESI-MS (*m/z*): [MH]⁺ calcd. for C₁₉H₃₀N₅O₇S₂: 504.2, found: 504.1

HPLC analysis: *t_R* = 7.30 min (Aeris Widespore XB-C18 2, gradient: 3-50% B'/A' over 15 min)

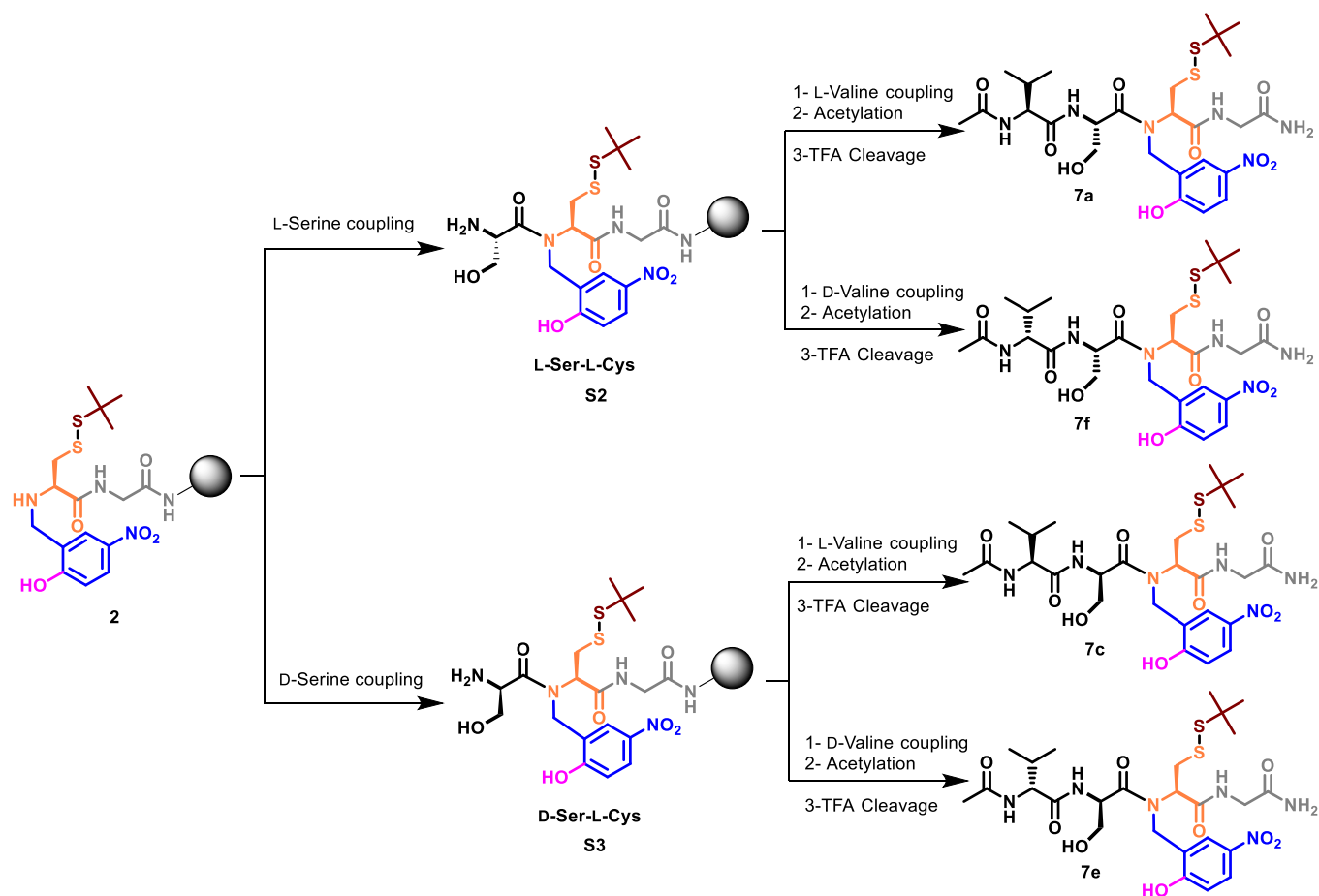
- H-Ser-(Hnb)D-Cys(StBu)-Gly-NH₂ (**5a**) and H-D-Ser-(Hnb)Cys(StBu)-Gly-NH₂ (**5b**)

ESI-MS (*m/z*): [MH]⁺ calcd. for C₁₉H₃₀N₅O₇S₂: 504.2, found: 504.1

HPLC analysis: *t_R* = 6.5 min (Aeris Widespore XB-C18 2, gradient: 3-50% B'/A' over 15 min)

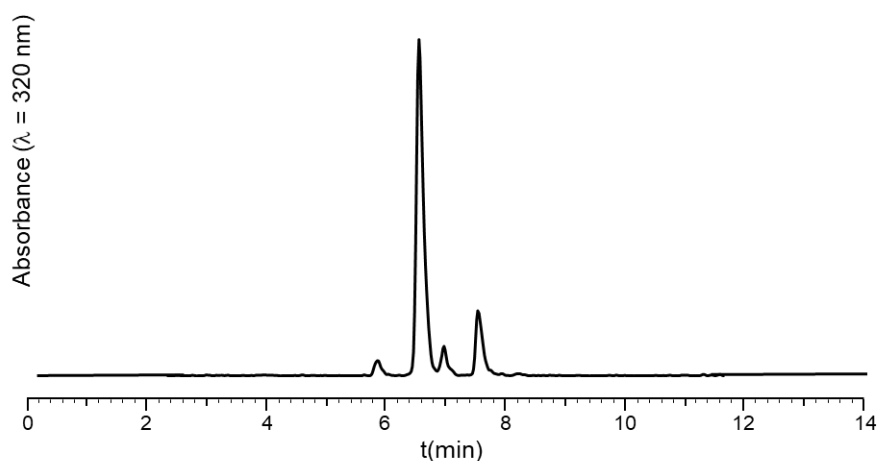
4-Synthesis of HPLC standards 7a, 7c, 7d and 7f

In the case of 7b and 7d, we synthesized their enantiomeric counterpart 7e and 7f, respectively, as the later compounds were less synthetically demanding while being equivalent in achiral HPLC (fig. 2A).



The four HPLC standards (**7a**, **7c**, **7f** and **7e**) were synthesized from **2** according to general coupling protocol PS1. *N*-acetylation was performed with 20% anhydride acetic in DCM for 20 min followed by piperidine treatment. Peptidyl resins were then cleaved following protocol PS1-small peptides cleavage.

- Ac-Val-Ser-(Hnb)Cys(StBu)-Gly-NH₂ (**7a**)

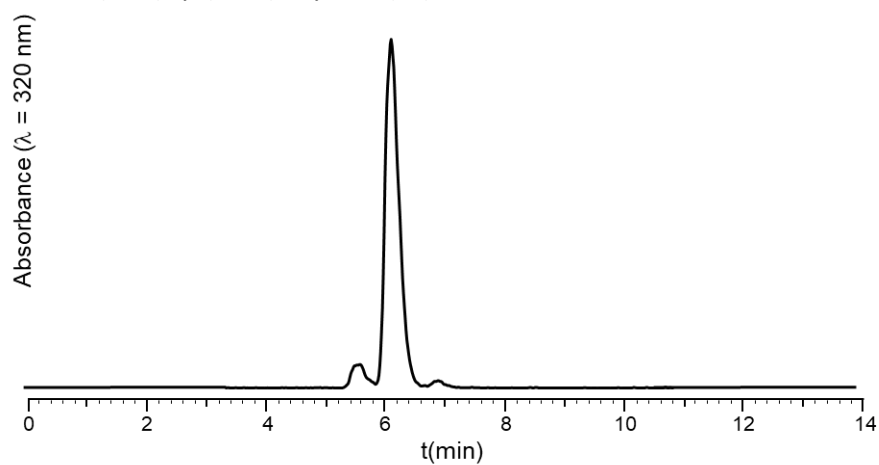


Supplementary figure S4: HPLC trace of crude **7a**

ESI-MS (*m/z*): [MH]⁺ calcd. for C₂₆H₄₁N₆O₉S₂: 645.2, found: 645.2

HPLC analysis: *t*_R = 6.77 min (Aeris Widepore XB-C18 2, gradient: 23-33% B'/A' over 15 min)

- Ac-Val-D-Ser-(Hnb)Cys(StBu)-Gly-NH₂ (**7c**)



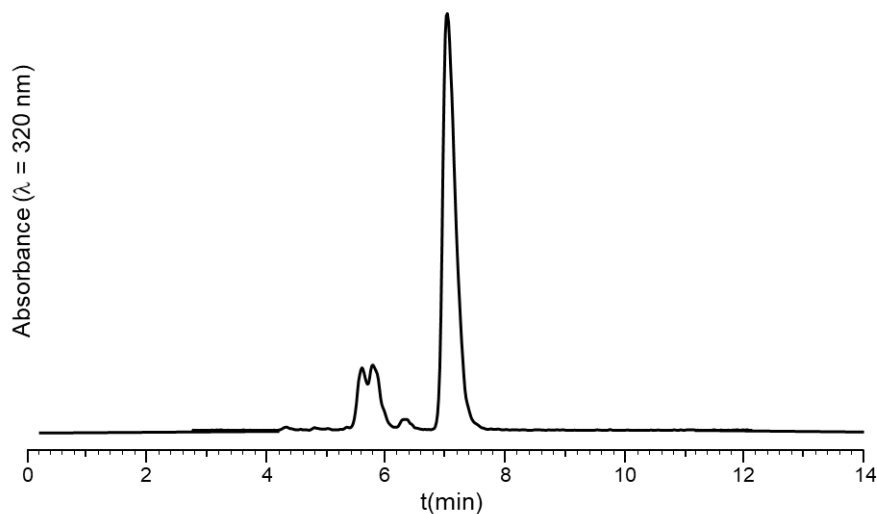
Supplementary figure S5: HPLC trace of crude **7c**

ESI-MS (*m/z*): [MH]⁺ calcd. for C₂₆H₄₁N₆O₉S₂: 645.2, found: 645.2

HPLC analysis: *t*_R = 6.23 min (Aeris Widepore XB-C18 2, gradient: 23-33% B'/A' over 15 min)

- Ac-D-Val-D-Ser-(Hnb)Cys(StBu)-Gly-NH₂ (**7e**)

For synthesis simplicity reasons, **7e** has been synthesized to be used as an HPLC standard for its enantiomer Ac-Val-Ser-(Hnb)D-Cys(StBu)-Gly-NH₂ (**7b**).



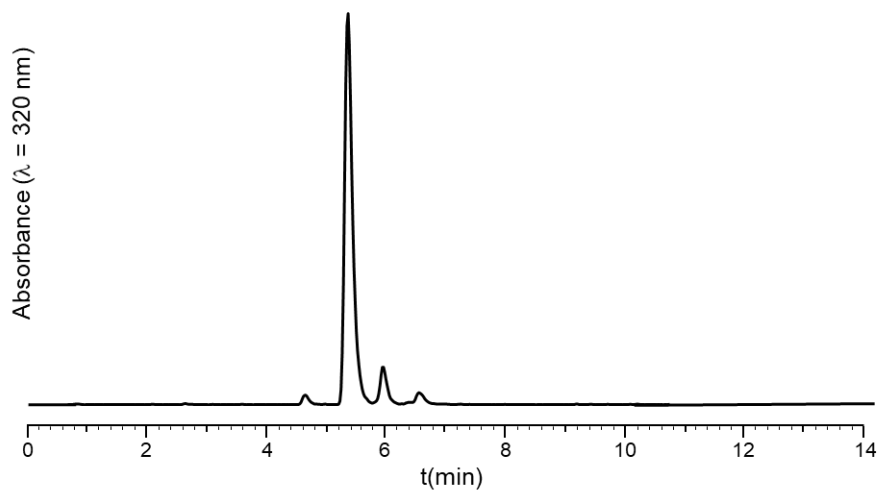
Supplementary figure S6: HPLC trace of crude **7e**

ESI-MS (m/z): [MH]⁺ calcd. for C₂₆H₄₁N₆O₉S₂: 645.2, found: 645.1

HPLC analysis: t_R = 7.01 min (Aeris Widedore XB-C18 2, gradient: 23-33% B'/A' over 15 min)

- Ac-D-Val-Ser-(Hnb)Cys(StBu)-Gly-NH₂ (**7f**)

For synthesis simplicity reasons, **7f** has been synthesized to be used as an HPLC standard for its enantiomer Ac-Val-D-Ser-(Hnb)D-Cys(StBu)-Gly-NH₂ (**7d**).

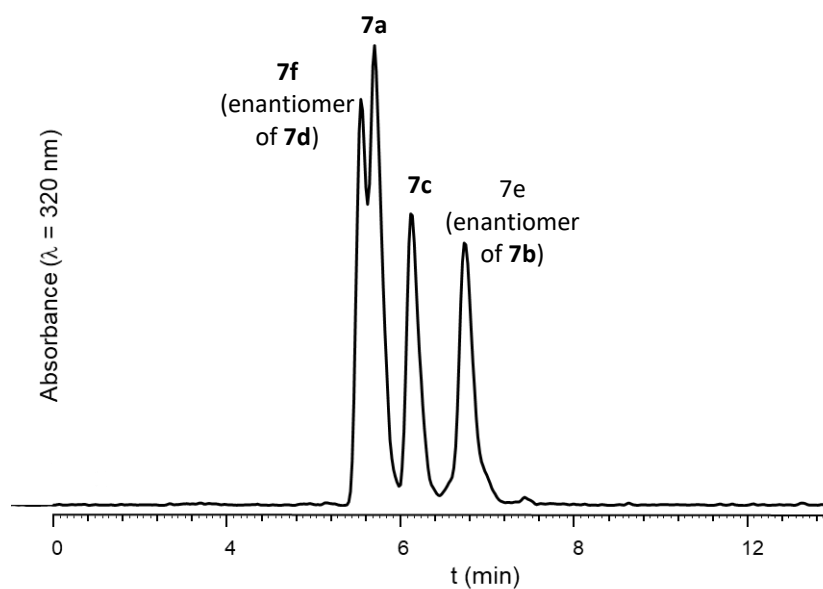


Supplementary figure S7: HPLC trace of crude **7f**

ESI-MS (m/z): [MH]⁺ calcd. for C₂₆H₄₁N₆O₉S₂: 645.2, found: 645.2

HPLC analysis: t_R = 5.56 min (Aeris Widedore XB-C18 2, gradient: 23-33% B'/A' over 15 min)

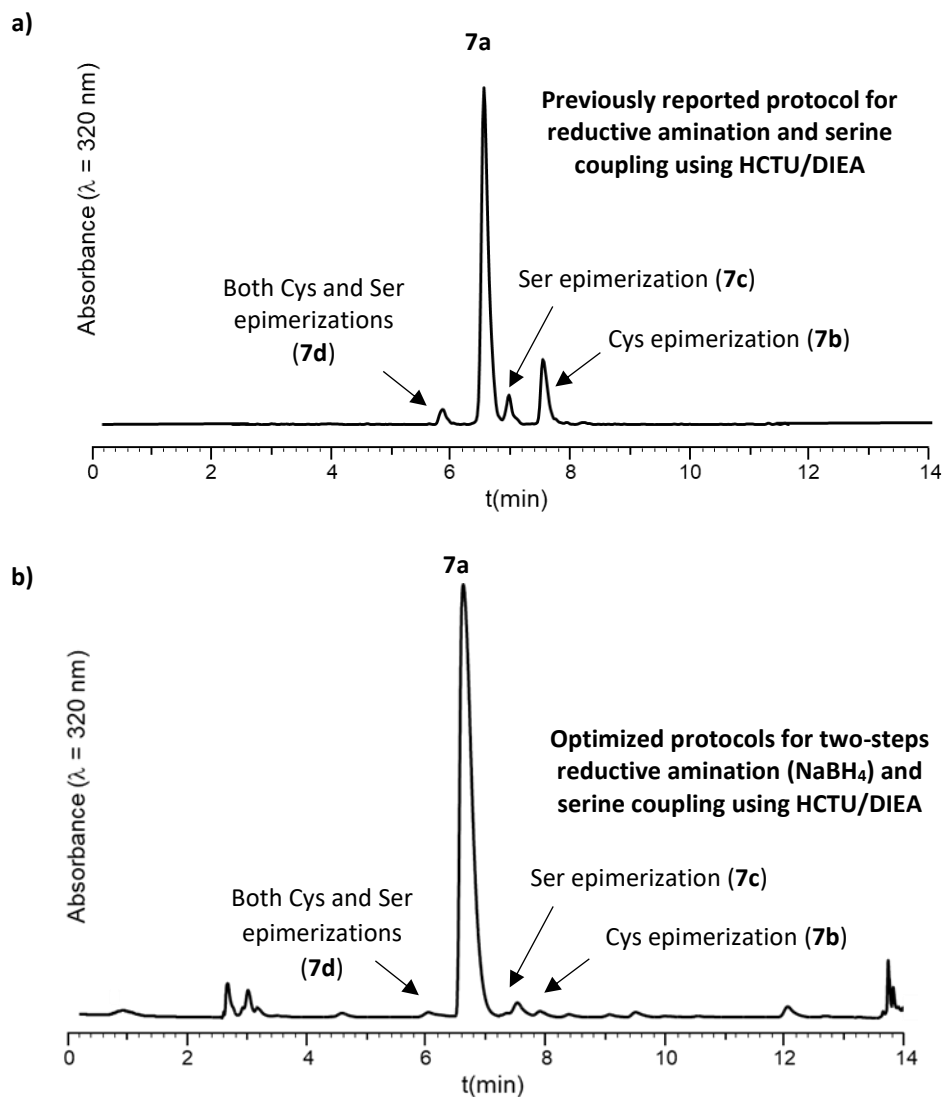
- Co-injection of the four HPLC standards **7a**, **7b**, **7e** and **7f**



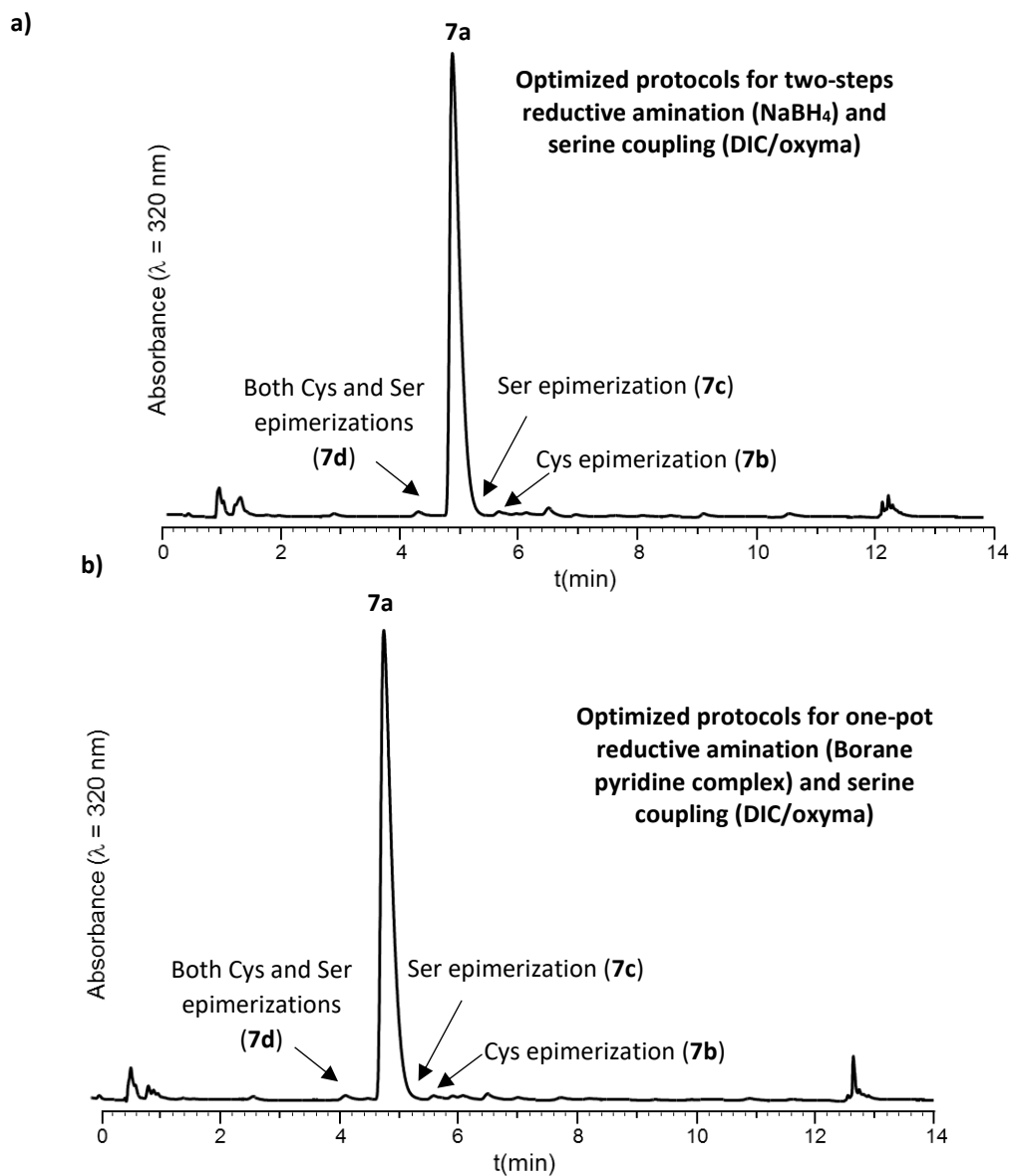
Supplementary figure S8: HPLC trace (Aeris Widepore XB-C18 2, gradient: 23-33% B'/A' over 15 min) of co-injection of the 4 HPLC standards

- Use of the HPLC standards for quantification of Ser and Cys epimerization

The mixture of the 4 HPLC standards was co-injected with the three crude mixtures obtained from the initial and optimized reductive amination protocols in order to unambiguously identify compounds arising from Cys, Ser and both Cys and Ser epimerization.



Supplementary figure S9: HPLC trace (Aeris Widespore XB- C18 2, gradient: 23-33% B'/A' over 15 min) of crude **7** obtained using previously reported reductive amination protocol (resin: ChemMatrix) (S9a) and the optimized protocol for two-steps reductive amination (NaBH_4) (resin: ChemMatrix) (S9b). Serine was coupled using the HCTU/DIEA standard protocol in both cases

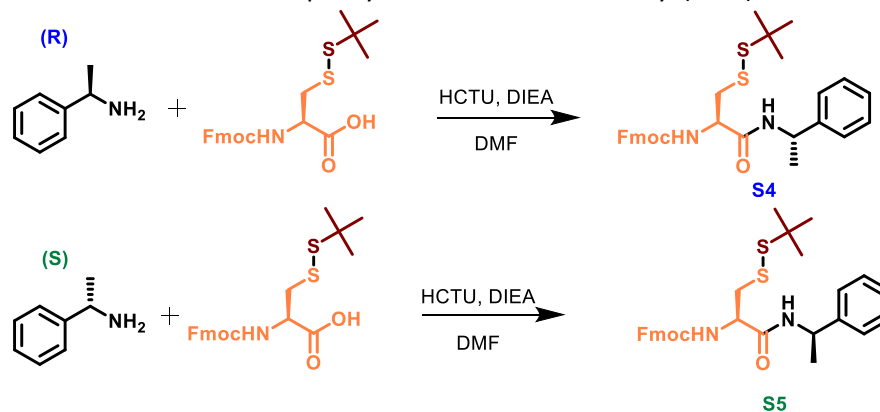


Supplementary figure S10: HPLC trace trace (Aeris Widepore XB- C18 2, gradient: 3-90 % B'/A' over 15 min) of crude **7** obtained using optimized two-steps reductive amination (NaBH_4) protocol (resin: ChemMatrix) (S10a) and the optimized one-pot reductive amination (borane pyridine complex) (resin: ChemMatrix) protocol (S10b). Serine was coupled using the optimized DIC/Oxyma protocol in both cases

5-Identification of the origin of the Cys epimerization

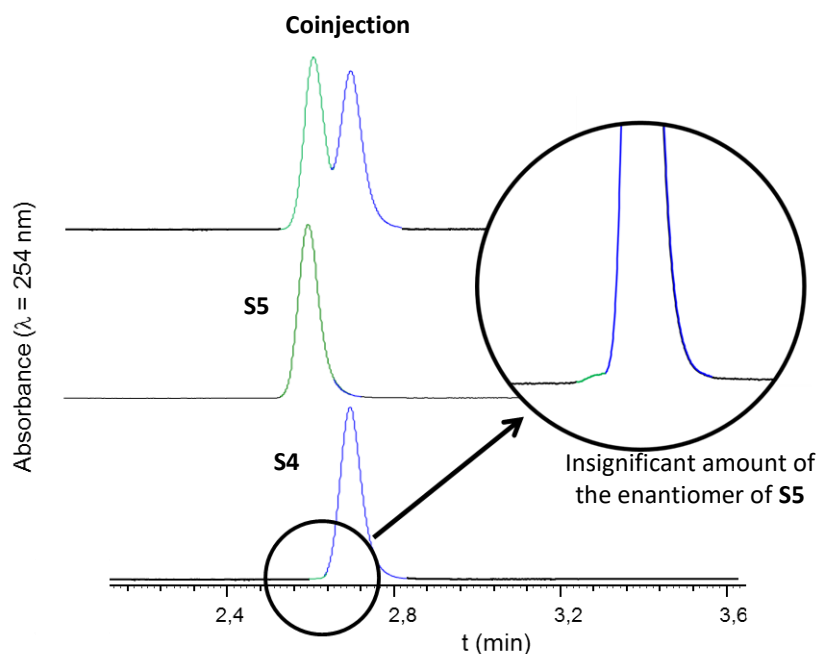
As described in the article main text, the following experiments allowed us to unambiguously conclude that Cys epimerization occurred during reductive amination.

- Verification of the enantiomeric purity of commercial Fmoc-Cys(StBu)-OH



(*R*)- and (*S*)-methylbenzylamine (89 μ l, 0.7 mmol, 1 equiv.) were coupled to Fmoc-Cys(StBu)-OH (302 mg, 0.7 mmol) using HCTU (350 mg, 0.83 mmol, 1.2 equiv.) and DIEA (122 μ l, 0.14 mmol, 2 equiv.) in DMF (5 ml) for 30 min giving **S4** and **S5**, respectively. No further purification was performed, crude products were directly analysed by HPLC.

S5 was used as a HPLC standard for its enantiomer H-D-Cys(StBu)-(*R*)-methylbenzylamine that would be present in **S4** if Fmoc-Cys(StBu)-OH was not optically pure.



Supplementary figure S11: HPLC trace of L-Cys(StBu)-(*R*)-methylbenzylamine (**S4**), L-Cys(StBu)-(*S*)-methylbenzylamine (**S5**) and co-injection of both

- **H-L-Cys(StBu)-(R)-methylbenzylamine (S4)**

ESI-MS (m/z): $[MH]^+$ calcd. for $C_{30}H_{34}N_2O_3S_2$: 535.2, found: 535.7

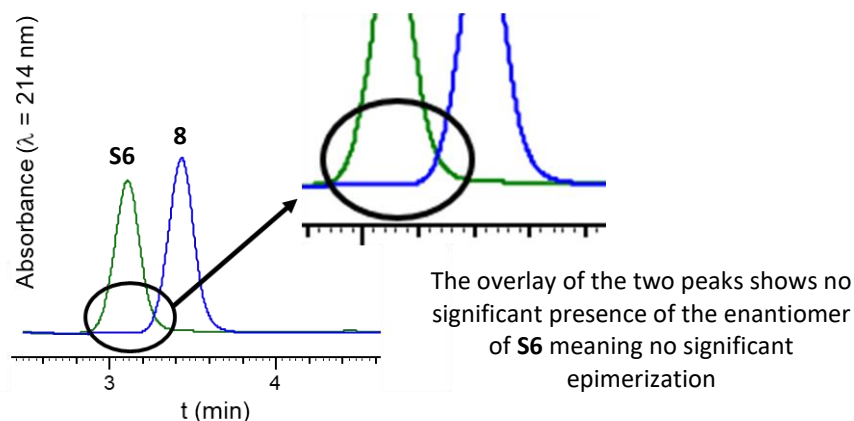
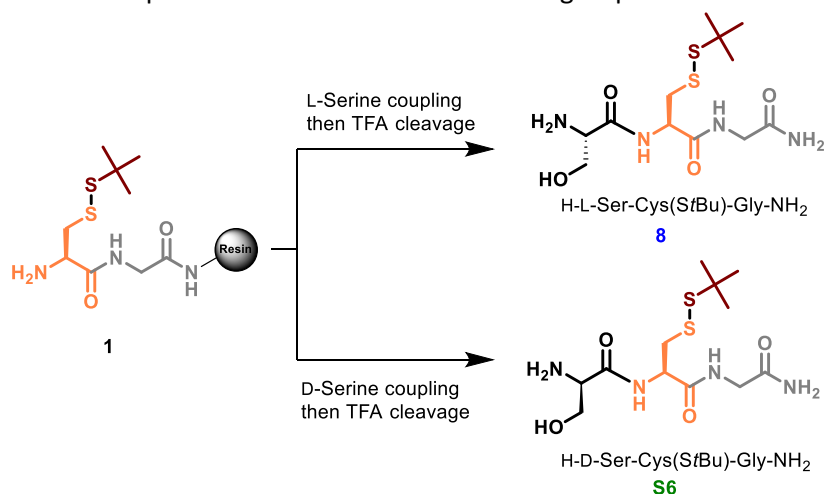
HPLC analysis: $t_R = 2.73$ min, (Chromolith, gradient: 10-50% B/A over 5 min)

- **H-L-Cys(StBu)-(S)-methylbenzylamine (S5)**

ESI-MS (m/z): $[MH]^+$ calcd. for $C_{30}H_{34}N_2O_3S_2$: 535.2, found: 535.7

HPLC analysis: $t_R = 2.63$ min (Chromolith, gradient: 10-50% B/A over 5 min)

- Verification of the epimerization rate without the Hnb group



Supplementary figure 12: HPLC trace of the overlay of H-L-Ser-Cys(StBu)-Gly-NH₂ (**8**) and H-D-Ser-Cys(StBu)-Gly-NH₂ (**S6**) peaks

S6 was used as a HPLC standard for its enantiomer H-L-Ser-D-Cys(StBu)-Gly-NH₂ that would be formed if L-Cys was epimerized during peptide elongation.

- **H-L-Ser-L-Cys(StBu)-Gly-NH₂ (8)**

ESI-MS (*m/z*): [MH]⁺ calcd. for C₁₂H₂₅N₄O₄S₂: 353.1, found: 353.5

HPLC analysis: *t_R* = 5.44 min (Chromolith, gradient: 20-30% B/A over 5 min)

- **H-D-Ser-L-Cys(StBu)-Gly-NH₂ (S6)**

ESI-MS (*m/z*): [MH]⁺ calcd. for C₁₂H₂₅N₄O₄S₂: 353.1, found: 353.5

HPLC analysis: *t_R* = 5.32 min, (Chromolith, gradient: 20-30% B/A over 5 min)

6-Effect of temperature and concentration on Cys epimerization during reductive amination

❖ Table S2: Effect of temperature on side product **5** formation during the synthesis of **4**.^[a]

Entry	Temperature (°C)	4 (%) ^[b]	5 (%) ^[b]
1	5	92	8
2	25	89	11
3	60	68	32

[a] Reductive amination conditions : imination with 10 equiv. aldehyde, 1 equiv. AcOH in 1:1 DMF/MeOH, 12.5 mM peptidyl resin, RT; reduction with 20 equiv. NaBH₃CN in 9:9:2 DMF/MeOH/AcOH, 12.5 mM peptidyl resin [b] Relative rates determined by HPLC peak integration at λ = 320 nm.

❖ Table S3: Effect of peptidyl resin concentration on side product **5** formation during the synthesis of **4**.^[a]

Entry	[Peptidyl resin] (mM)	equiv. aldehyde	equiv. NaBH ₃ CN	4 (%) ^[b]	5 (%) ^[b]
1	12.5	10	20	89	11
2	3	10	20	84	16
3	25	10	20	89	11
4	12.5	2	4	77	23
5	12.5	50	100	89	11
6	12.5	2	20	78	22

[a] Reductive amination conditions : imination with 10 equiv. aldehyde, 1 equiv. AcOH in 1:1 DMF/MeOH, 12.5 mM peptidyl resin, RT; reduction with 20 equiv. NaBH₃CN in 9:9:2 DMF/MeOH/AcOH, 12.5 mM peptidyl resin [b] Relative rates determined by HPLC peak integration at λ = 320 nm.

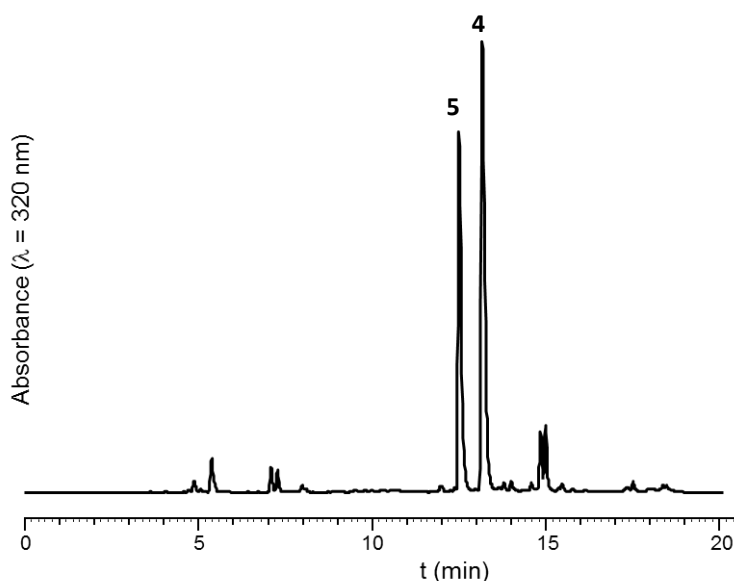
7-NMR characterisation of the [M-2] imidazolidinone byproducts 10a-d

7-1- Synthesis and isolation of products 4, 5 and 10a-d

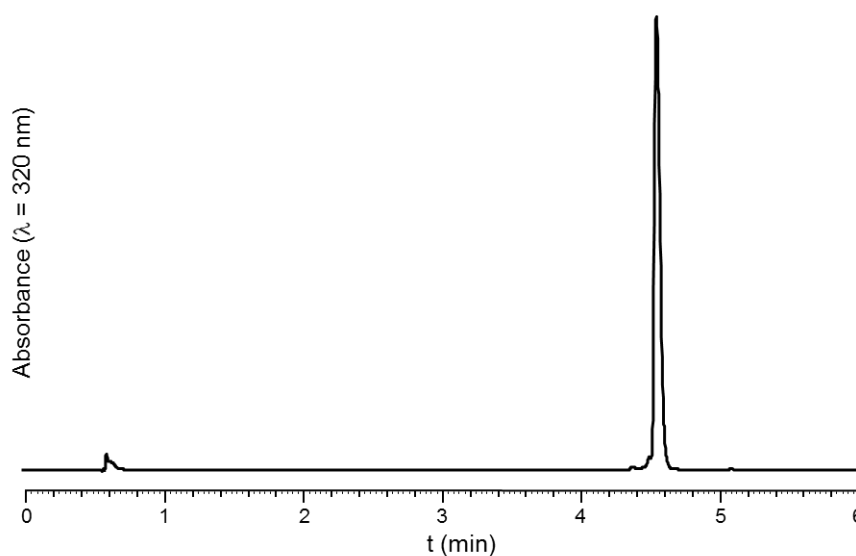
- Synthesis of **4** and **5** *via* reductive amination with NaBH₃CN using a protocol designed to promote the formation of the epimerized product.

25 μmol of H-Cys(StBu)-Gly-Rink-ChemMatrix **1a** was washed with 1:1 DMF/MeOH mixture (4 x 3 mL, 30 s) then swollen in 3 mL of a 9:9:2 DMF/MeOH/AcOH mixture for 5 min. The resin was washed with 1:1 DMF/MeOH mixture (3 x 3 mL, 30 s), then 2-hydroxy-5-nitrobenzaldehyde (42mg, 10 equiv.) in 8 mL of 9:9:2 DMF/MeOH/AcOH was added and the reactor was stirred for 4 h at 40°C. The reactor was drained and the resin was washed with 1:1 DMF/MeOH mixture (3 x 3 mL, 5 s). A fresh solution of

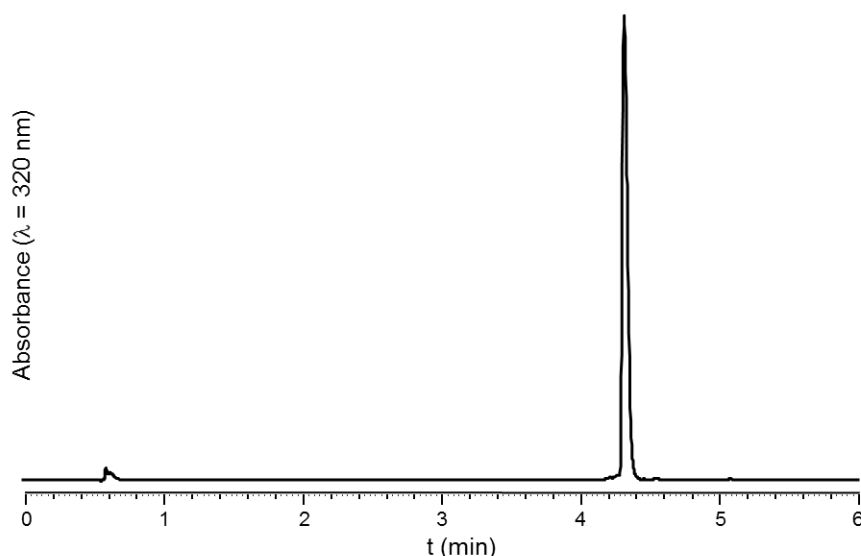
sodium cyanoborohydride (31.5 mg, 20 equiv.) in 3 mL of 9:9:2 DMF/MeOH/AcOH (250 mM, sodium cyano-borohydride concentration) was added and the reactor stirred for 1h. The reactor was drained and the resin was washed with 1:1 DMF/MeOH (4 × 3 mL, 30 s), NMP (3 × 3 mL, 30 s), 20% piperidine in NMP (3 × 3 mL, 3 min), NMP (3 × 3 mL, 30 s), dichloromethane (3 × 5 mL, 30 s) and NMP (2 × 3 mL, 30 s). Fmoc-Ser(*t*Bu)-OH was then coupled for 18 h followed by piperidine-mediated Fmoc deprotection, following protocol PS1. The peptidyl-resin was cleaved (protocol PS3 small peptides cleavage) to give a 42:58 mixture of **4** and **5**. The mixture was purified by semi preparative HPLC (Nucleosil C18, gradient: 15-50% B/A over 20 min) to give pure **4** and **5** for NMR characterization (see p S7 for MS characterization).



Supplementary figure S13: HPLC trace (Nucleosil C18, gradient: 5-50% B/A over 20 min) of the **4/5** mixture obtained using a protocol designed to promote epimerization



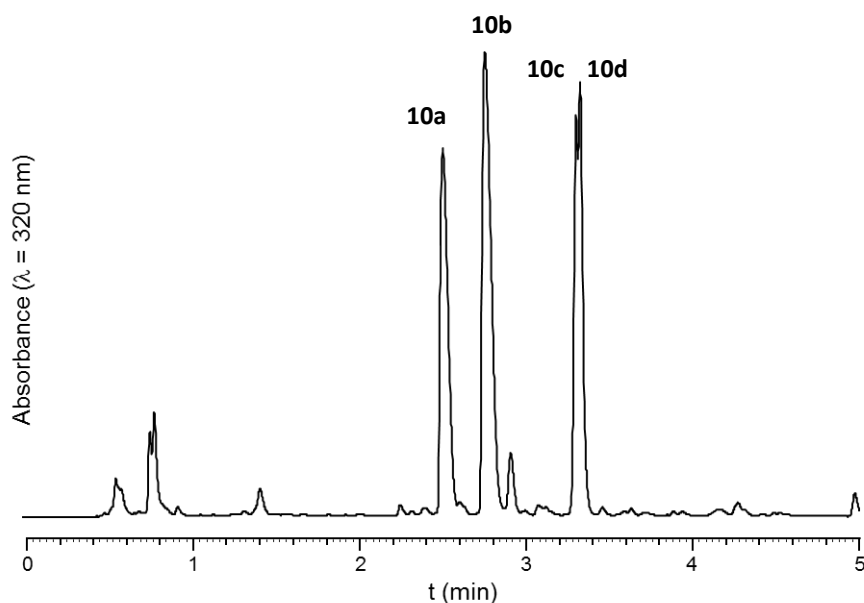
Supplementary figure S14: HPLC trace (Chromolith, gradient: 5-50% B/A over 5 min) of purified **4**



Supplementary figure S15: HPLC trace (Chromolith, gradient: 5-50% B/A over 5 min) of purified **5**

- Synthesis and purification of imidazolidinones **10a-d**

25 μ mol of H-Cys(*StBu*)-Gly-Rink-ChemMatrix **1a** was washed with 1:1 DMF/MeOH mixture (4 x 3 mL, 30 s) then swollen in 3 mL of a 9:9:2 DMF/MeOH/AcOH mixture for 5 min. The resin was washed with 1:1 DMF/MeOH mixture (3 x 3 mL, 30 s), then 2-hydroxy-5-nitrobenzaldehyde (42 mg, 10 equiv.) in 2 mL 1:1 DMF/MeOH (125 mM aldehyde concentration) was added, and the reactor was stirred for 2 h. The reactor was drained and the resin was washed with 1:1 DMF/MeOH (4 x 3 mL, 30 s), NMP (3 x 3 mL, 30 s), 20% piperidine in NMP (3 mL, 3 min, x 3), NMP (3 x 3 mL, 30 s), dichloromethane (3 x 5 mL, 30 s) and NMP (2 x 3 mL, 30 s). Fmoc-Ser(*tBu*)-OH was then coupled for 18 h followed by piperidine-mediated Fmoc deprotection, following general protocol PS1. The peptidyl-resin was cleaved (protocol PS3 small peptides cleavage) to give a mixture of **10a-d**. The mixture was purified by semi preparative HPLC (Nucleosil C18, gradient: 15-50% B/A over 20 min).



Supplementary figure S16: HPLC trace of imidazolidinones **10a-d**

- Imidazolidinone **10a**

ESI-HRMS (m/z): $[MH]^+$ calcd. for $C_{19}H_{28}N_5O_7S_2$: 502.1430, found: 502.1429

HPLC analysis: t_R = 2.47 min (Chromolith, gradient: 5-50% B/A over 5 min)

- Imidazolidinone **10b**

ESI-HRMS (m/z): $[MH]^+$ calcd. for $C_{19}H_{28}N_5O_7S_2$: 502.1430, found: 502.1426

HPLC analysis: t_R = 2.78 min (Chromolith, gradient: 5-50% B/A over 5 min)

- Imidazolidinone **10c**

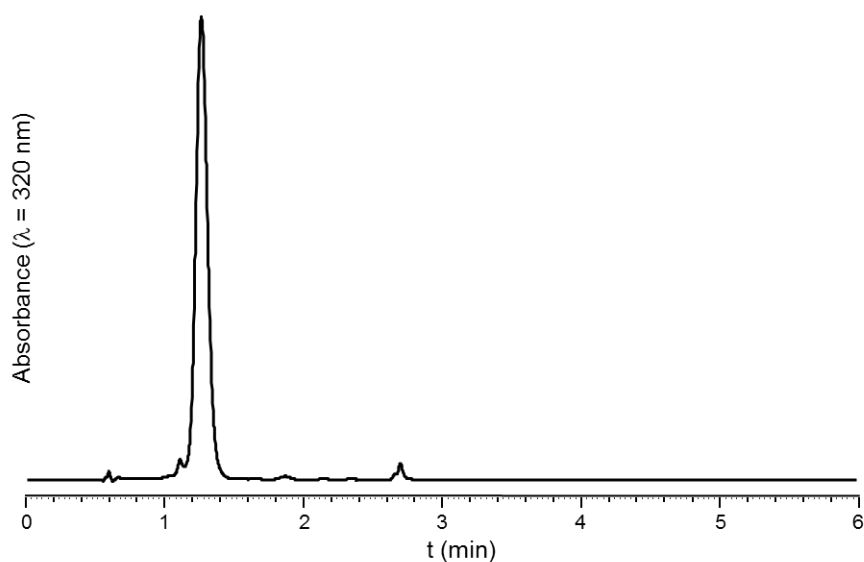
ESI-HRMS (m/z): $[MH]^+$ calcd. for $C_{19}H_{28}N_5O_7S_2$: 502.1430, found: 502.1428

HPLC analysis: t_R = 3.31 min (Chromolith, gradient: 5-50% B/A over 5 min)

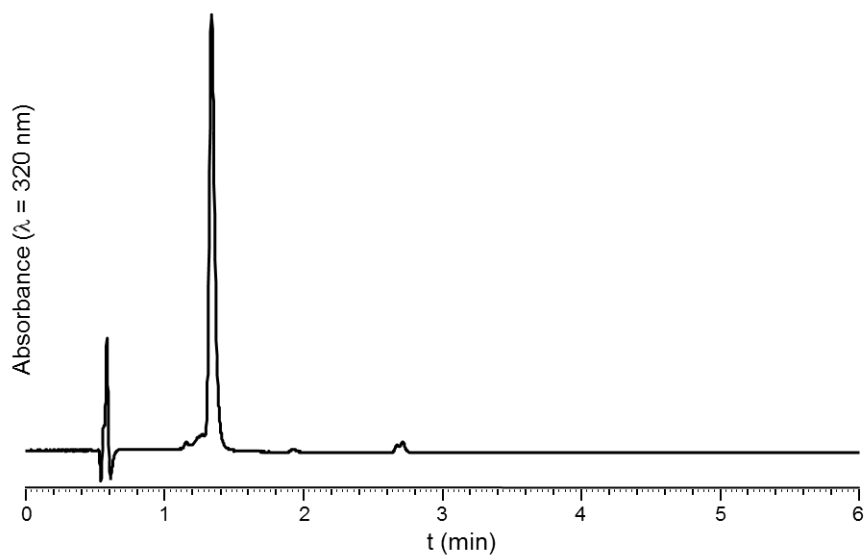
- Imidazolidinone **10d**

ESI-HRMS (m/z): $[MH]^+$ calcd. for $C_{19}H_{28}N_5O_7S_2$: 502.1430, found: 502.1423

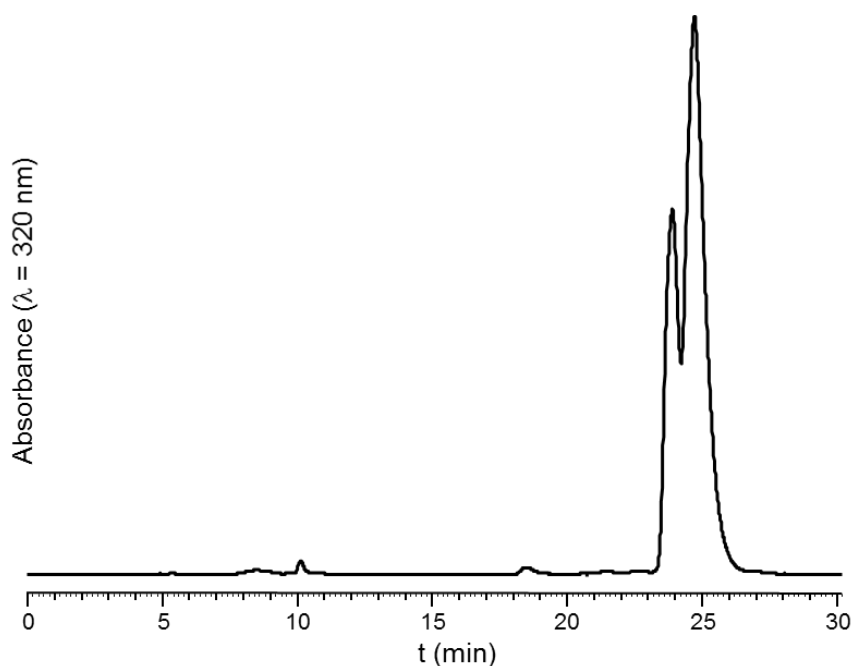
HPLC analysis: t_R = 3.33 min (Chromolith, gradient: 5-50% B/A over 5 min)



Supplementary figure S17: HPLC trace (Chromolith, gradient: 20-50% B/A over 5 min) of purified imidazolidinone **10a**

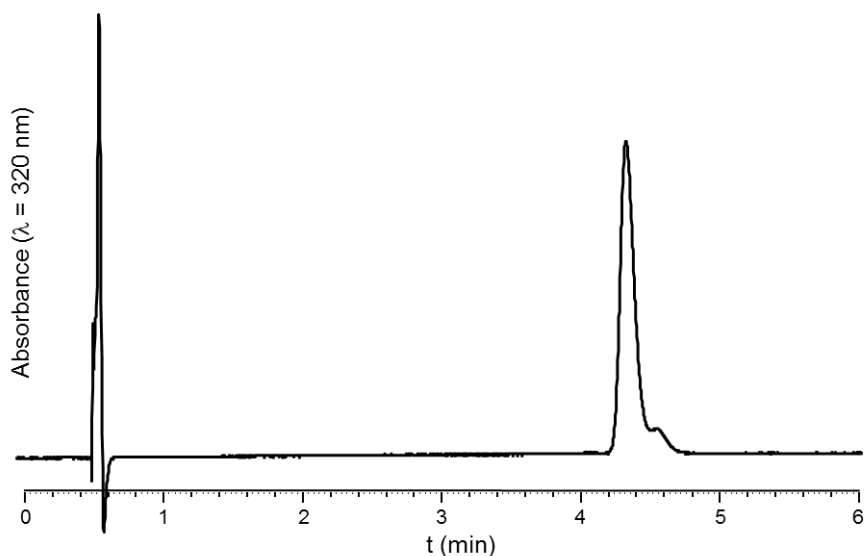


Supplementary figure S18: HPLC trace (Chromolith, gradient: 20-50% B/A over 5 min) of purified imidazolidinone **10b**

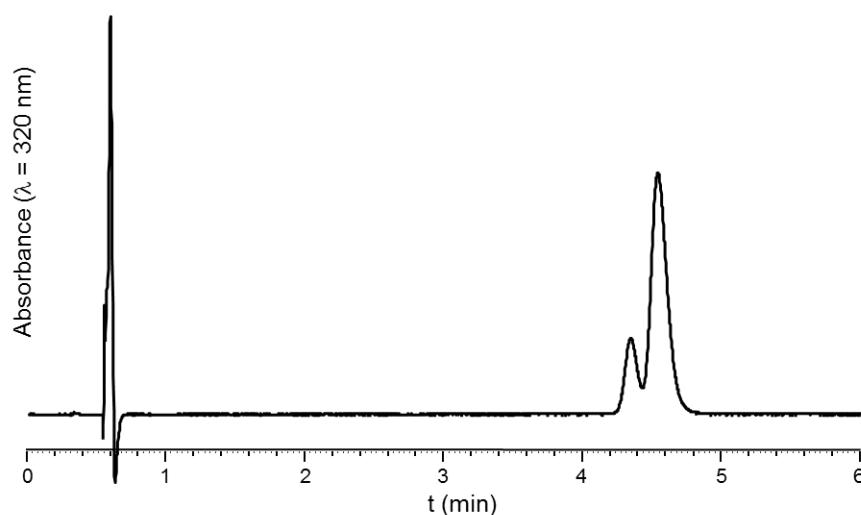


Supplementary figure S19: HPLC trace (Nucleosil C4, gradient: Isocratic 23% B/A over 20 min) of purified imidazolidinones **10c** and **10d** mixture (33:67 ratio)

Then, the 33:67 mixture of **10c** and **10d** were subjected to a second round of semi-preparative HPLC in order to try to separate them. HPLC purification: Nucleosil C4, gradient: Isocratic 23 % B/A over 20 min at 40°C. This leads to 91:9 and 19:81 **10c/10d** mixtures, respectively, but in very small amounts, which complicated their NMR characterization due to low signal/noise ratios. As a consequence, we mainly based the NMR characterization of **10c** and **10d** on the mixture.



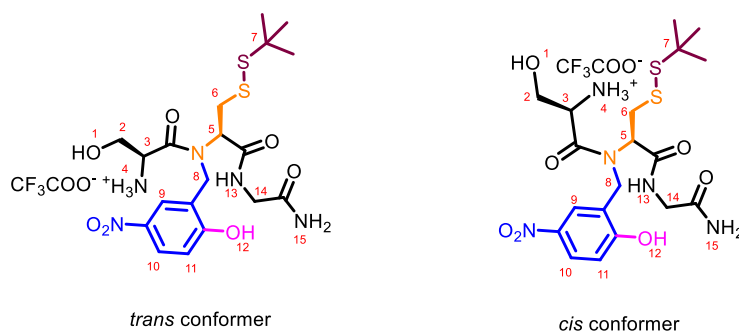
Supplementary figure S20: HPLC trace (Chromolith, gradient: Isocratic 20% B/A over 5 min) of purified imidazolidinone **10c**



Supplementary figure S21: HPLC trace (Chromolith, gradient: Isocratic 20% B/A over 5 min) of purified imidazolidinone **10d**

7-2 NMR analysis of 4, 5 and 10a-d

- NMR analysis of **4**



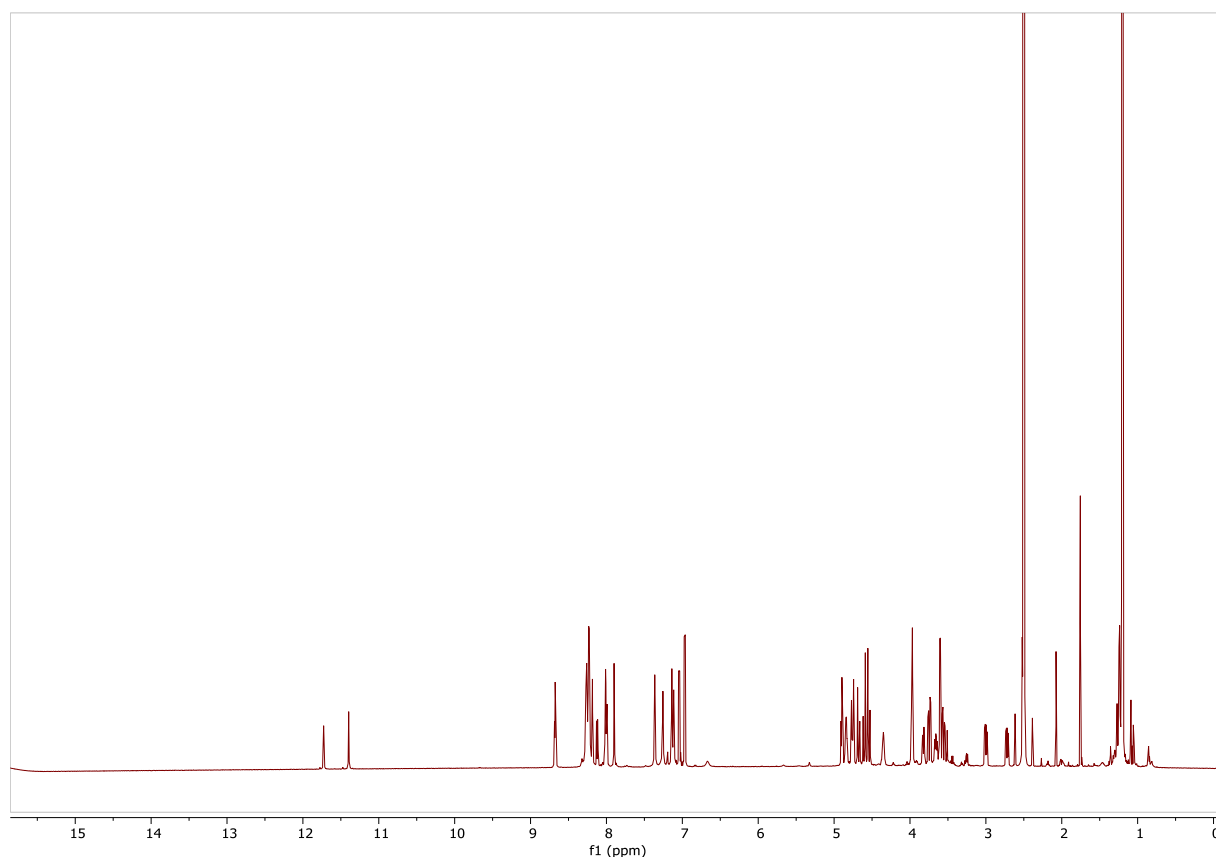
NMR analysis of pure **4** shows two sets of peaks corresponding to two amide bond rotamers (57:43 ratio), as expected from the *N,N*-disubstituted nature of the Ser-Cys bond. Identification of *cis* and *trans* conformers was not possible with NOESY and ROESY analysis due to the overlap between signals.

The following color code is used in the NMR description: **Major conformer**, **minor conformer**, **both**.

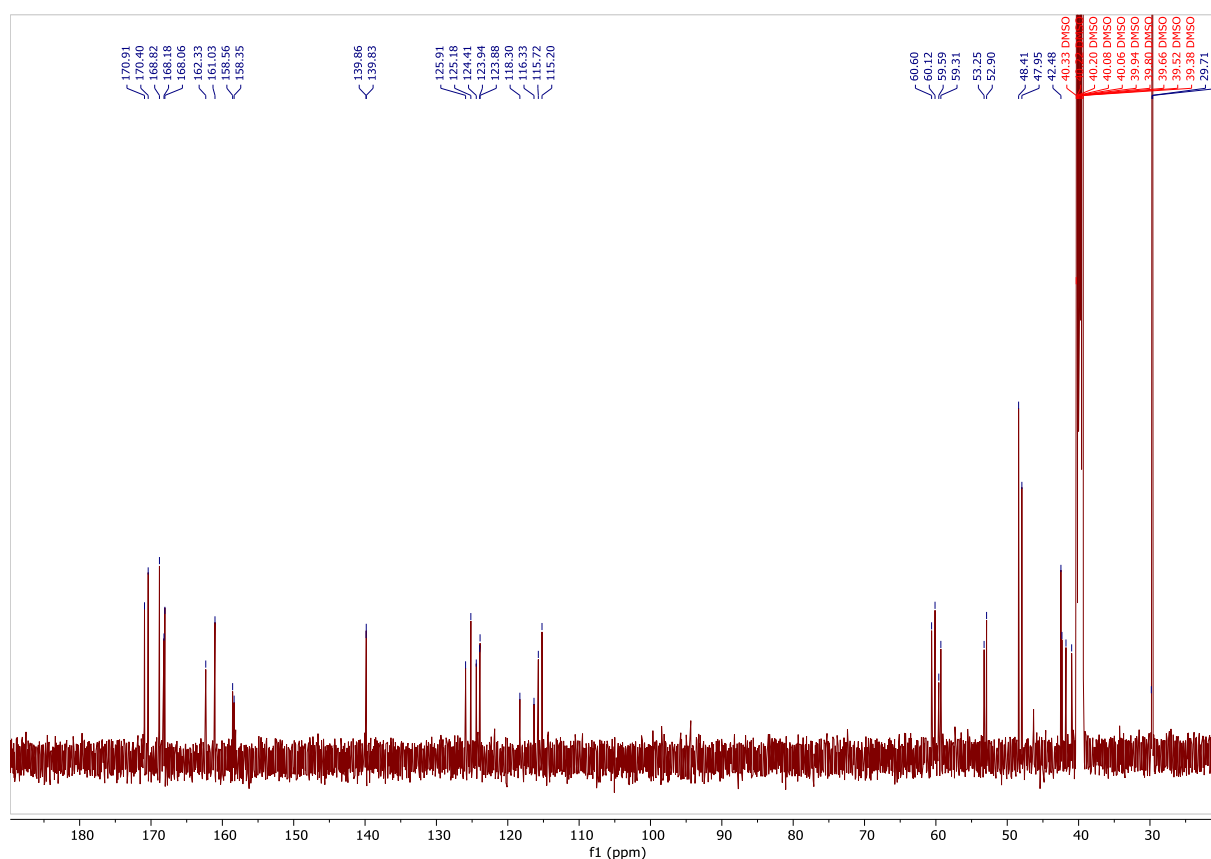
^1H NMR (600 MHz, DMSO- d_6) δ 11.73 (s, 1H, H_{12}), 11.40 (s, 1H, H_{12}), 8.68 (t, $J = 5.5$ Hz, 1H, H_{13}), 8.26 (d, $J = 5.3$ Hz, 3H, H_4), 8.23 (d, $J = 5.1$ Hz, 3H, H_4), 8.19 (d, $J = 2.8$ Hz, 1H, H_9), 8.12 (dd, $J = 9, 2.8$ Hz, 1H, H_{10}), 8.03 – 7.98 (m, 2x1H, $\text{H}_{10} + \text{H}_{13}$), 7.90 (d, $J = 2.8$ Hz, 1H, H_9), 7.36 (bs, 1H, H_{15b}), 7.26 (bs, 1H, H_{15a}), 7.14 (s, 1H, H_{15a}), 7.12 (s, 1H, H_{15a}), 7.04 (d, $J = 9.0$ Hz, 1H, H_{11}), 6.97 (d, $J = 9.0$ Hz, 1H, H_{11}), 4.90 (bs, 1H, H_5), 4.86–4.81 (m, 1H, H_3), 4.79 – 4.73 (m, 2xH, H_5 , H_{8b}), 4.68 (d, $J = 17.1$ Hz, 1H, H_{8a}), 4.60 (d, $J = 17.0$ Hz, 1H, H_{8b}), 4.54 (d, $J = 17.0$ Hz, 1H, H_{8a}), 4.39 – 4.32 (m, 1H, H_3), 3.97 (d, $J = 4.7$ Hz, 2H, H_2), 3.82 (dd, $J = 11.72, 4.01$ Hz, 1H, H_{2b}), 3.75 (dd, $J = 16.9, 6.0$ Hz, 1H, H_{14b}), 3.66 (dd, $J = 11.7, 7.1$ Hz, 1H, H_{2a}), 3.61–3.58 (m, 2H, H_{14}), 3.56 (dd, $J = 16.9, 4.9$ Hz, 1H, H_{14a}), 3.33 (dd, $J = 13.5, 7.6$ Hz, 1H, H_{6a}), 3.26 (dd, $J = 13.1, 7.7$ Hz, 1H, H_{6a}), 3.00 (dd, $J = 13.4, 6.7$ Hz, 1H, H_{6b}), 2.72 (dd, $J = 12.3, 6.2$ Hz, 1H, H_{6b}), 1.21 (s, 9H, H_7), 1.20 (s, 9H, H_7).

^{13}C NMR (190 MHz, DMSO-*d*₆) δ 170.9 C_{CO}, 170.4 C_{CO}, 168.8 2C_{CO}, 168.2 C_{CO}, 168.1 C_{CO}, 162.3 C_{OH}, 161.0 C_{OH}, 139.9 2C_{NO2}, 125.9 C_{Ar} + C₉, 125.2 C_{Ar}, 124.4 C₁₀, 123.94 C₁₀, 123.88 C₉, 115.7 C₁₁, 115.2 C₁₁, 60.6 C₂, 60.1 C₂, 59.6 C₅, 59.31 C₅, 53.25 C₅, 52.9 C₅, 48.4 C₈, 47.95 C₈, 42.5 C₁₄, 42.3 C₁₄, 41.8 C₆, 41.0 C₆, 29.7 C₇, 29.6 C₇ (CMe₃ and CF₃COO⁻ have not been identified due to low signal-to-noise ratio)

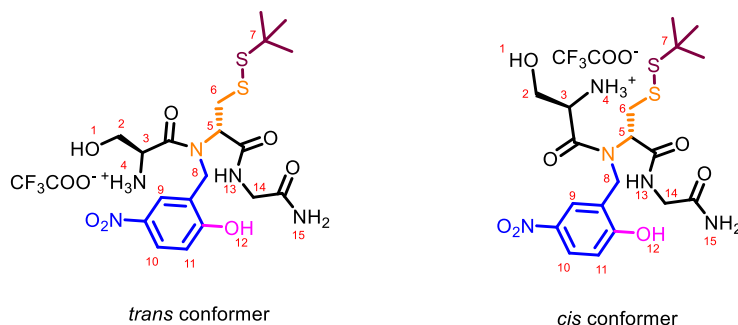
^{15}N NMR (60 MHz, DMSO-*d*₆) δ 111.9 CONH, 107.9 CONH, 105.1 CONH₂, 104.7 CONH₂, 92.4 2 x NH₃⁺ (chemical shifts obtained from 2D ^1H - ^{15}N HSQC)



Supplementary figure S22: ^1H -NMR spectrum of **4**

Supplementary figure S23: ^{13}C -NMR spectrum of **4**

- NMR analysis of **5**



NMR analysis of pure **5** shows two sets of peaks corresponding to two amide bond rotamers (71:29 ratio), as expected from the *N,N*-disubstituted nature of the Ser-Cys bond. Identification of *cis* and *trans* conformers was not possible with NOESY and ROESY analysis due to the overlap between signals.

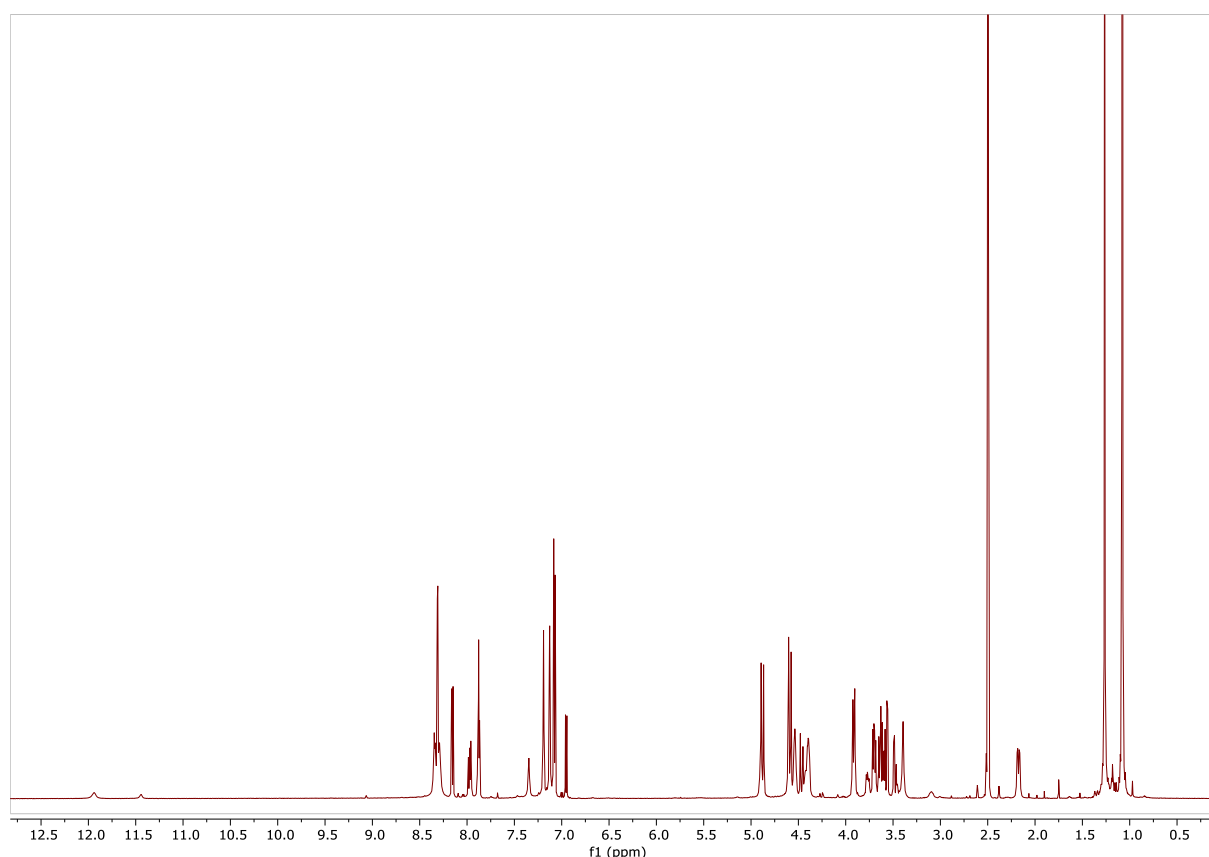
The following color code is used in the NMR description: **Major conformer**, **minor conformer**, **both**. (Assignment based on the combined analysis of 2D ^1H COSY, ^1H - ^{13}C and ^1H - ^{15}N HSQC spectra)

^1H NMR (600 MHz, $\text{DMSO}-d_6$) δ 11.94 (s, 1H, H_{12}), 11.44 (s, 1H, H_{12}), 8.37 – 8.32 (m, 3H + 1H, $\text{H}_4 + \text{H}_{13}$), 8.31 (d, $J = 2.8$ Hz, 1H, H_9), 8.30 – 8.27 (m, 3H, H_4), 8.15 (dd, $J = 9.0, 2.8$ Hz, 1H, H_{10}), 7.98 (dd, $J = 8.8, 2.8$ Hz, 1H, H_{10}), 7.96 (d, $J = 2.9$ Hz, 1H, H_9), 7.90–7.85 (m, 3H, H_{13}), 7.19 (bs, 2x1H, $\text{H}_{15b} + \text{H}_{15b}$), 7.13 (bs, 2x1H, $\text{H}_{15a} + \text{H}_{15a}$), 7.08 (d, $J = 8.9$ Hz, 1H, H_{11}), 6.95 (d, $J = 8.8$ Hz, 1H, H_{11}), 4.93–4.83 (m, 2x1H, $\text{H}_{5b} \text{H}_{8a}$), 4.59 (d, $J = 15.8$ Hz, 2x1H, $\text{H}_{8b} + \text{H}_{8a}$), 4.56–4.50 (m, 1H, H_3), 4.47 (d, $J = 16.4$ Hz, 1H, H_{8b}), 4.44 – 4.40 (m,

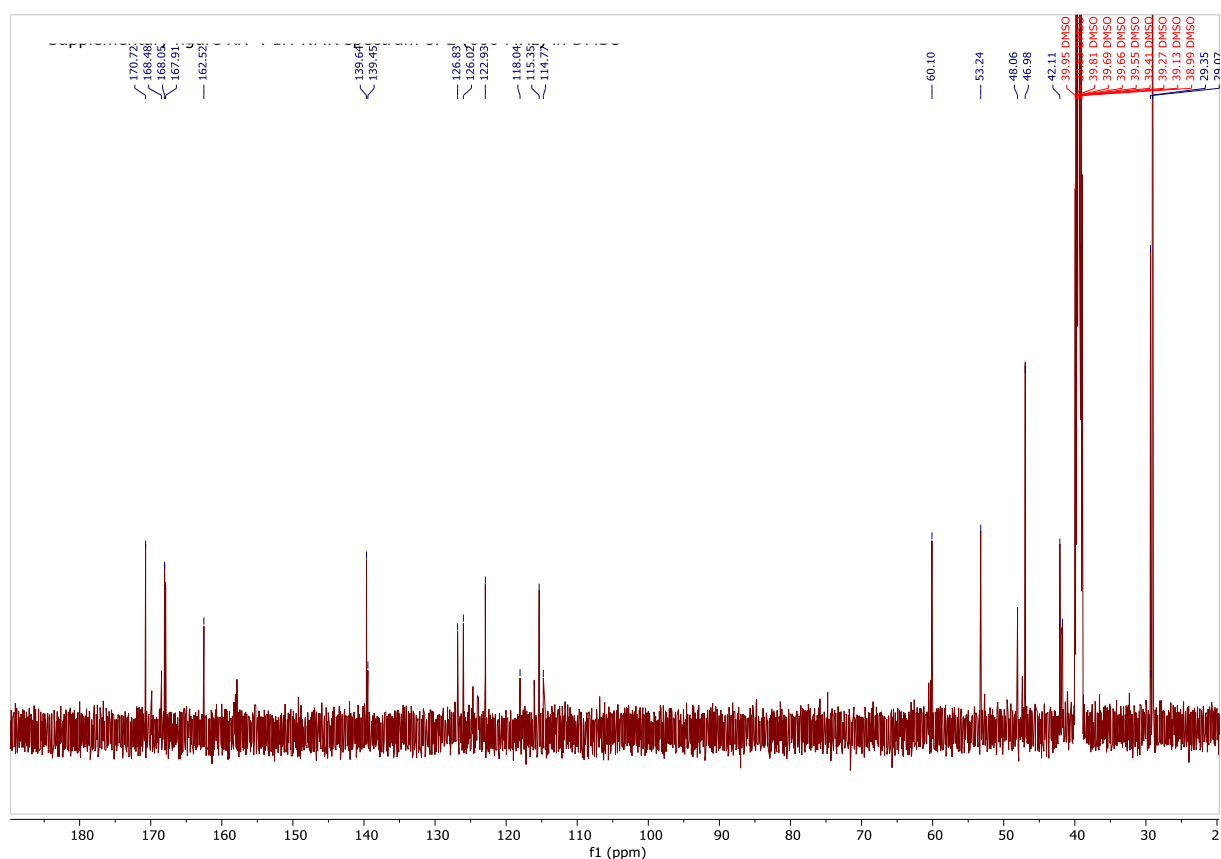
2x1H, $H_3 + H_5$), 3.92 (m, 2x1H, $H_{2a} H_{2a}$), 3.77 (dd, $J = 11.3, 6.7$ Hz, 1H H_{2b}), 3.70 (dd, $J = 11.9, 7.0$ Hz, 1H, H_{2b}), 3.66 – 3.62 (m, 2x1H, $H_{6b} H_{14b}$), 3.62 – 3.58 (m, 2x1H, $H_{14b} H_{14b}$), 3.49 (d, $J = 5.6$ Hz, 1H, H_{14a}), 3.46 (d, $J = 5.6$ Hz, 1H, H_{14a}), 3.39 (s, 1H, H_{6a}), 3.1–3.05 (m, 1H, H_{6b}), 2.18 (dd, $J = 13.5, 4.7$ Hz, 1H, H_{6a}), 1.26 (s, 9H, H_7), 1.08 (s, 9H, H_7).

^{13}C NMR (190 MHz, DMSO- d_6) δ 170.7 C_{CO} , 169.9 C_{CO} , 168.5 2C_{CO} , 168.1 C_{CO} , 167.9 C_{CO} , 162.5 2C_{OH} , 139.6 C_{NO_2} , 139.4 C_{NO_2} , 126.8 C_9 , 126.0 C_{10} , 124.7 C_{10} , 124.0 C_9 , 123.9 C_{Ar} , 122.9 C_A , 115.3 C_{15} , 114.8 C_{15} , 60.3 C_2 , 60.13 C_2 , 60.10 C_5 , 59.2 C_5 , 53.2 C_3 , 52.5 C_3 , 47.0 C_8 , 42.1 2C_{14} , 41.8 C_8 , 41.7 C_6 , 40.9 C_6 , 29.3 C_7 , 29.1 C_7 . (CMe_3 and CF_3COO^- have not been identified due to low signal-to-noise ratio)

^{15}N NMR (60 MHz, DMSO- d_6) δ 111.1 CONH , 105.0 CONH , 104.9 CONH_2 , 104.8 CONH_2 , 91.5 NH_3^+ , 92.3 NH_3^+ (chemical shifts obtained from 2D ^1H - ^{15}N HSQC)



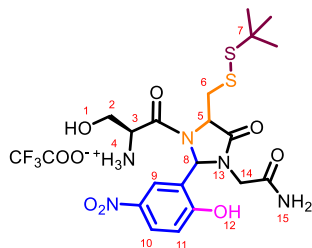
Supplementary figure S24: ^1H -NMR spectrum of **5**

Supplementary figure 25: ^{13}C -NMR spectrum of **5**

- NMR analysis of imidazolidinones **10a-d**

Note that ^1H NMR spectra of pure **10a-d** were extremely complex, showing two or three different sets of peak contributing to more than 5 % in each spectrum, and probably other minor species. Based on the HPLC purity of the purified compounds, we attribute this spectral complexity to the presence of a mixture of conformers due to the two expected *cis-trans* amide rotamers, combined with different conformations of the 5-membered heterocycle. Identification of *cis* and *trans* conformers was not possible with NOESY and ROESY analysis due to the overlap between signals. For similar reasons, assignment of the relative C5/C8 stereochemistry was also not possible.

- NMR analysis of imidazolidinone **10a**



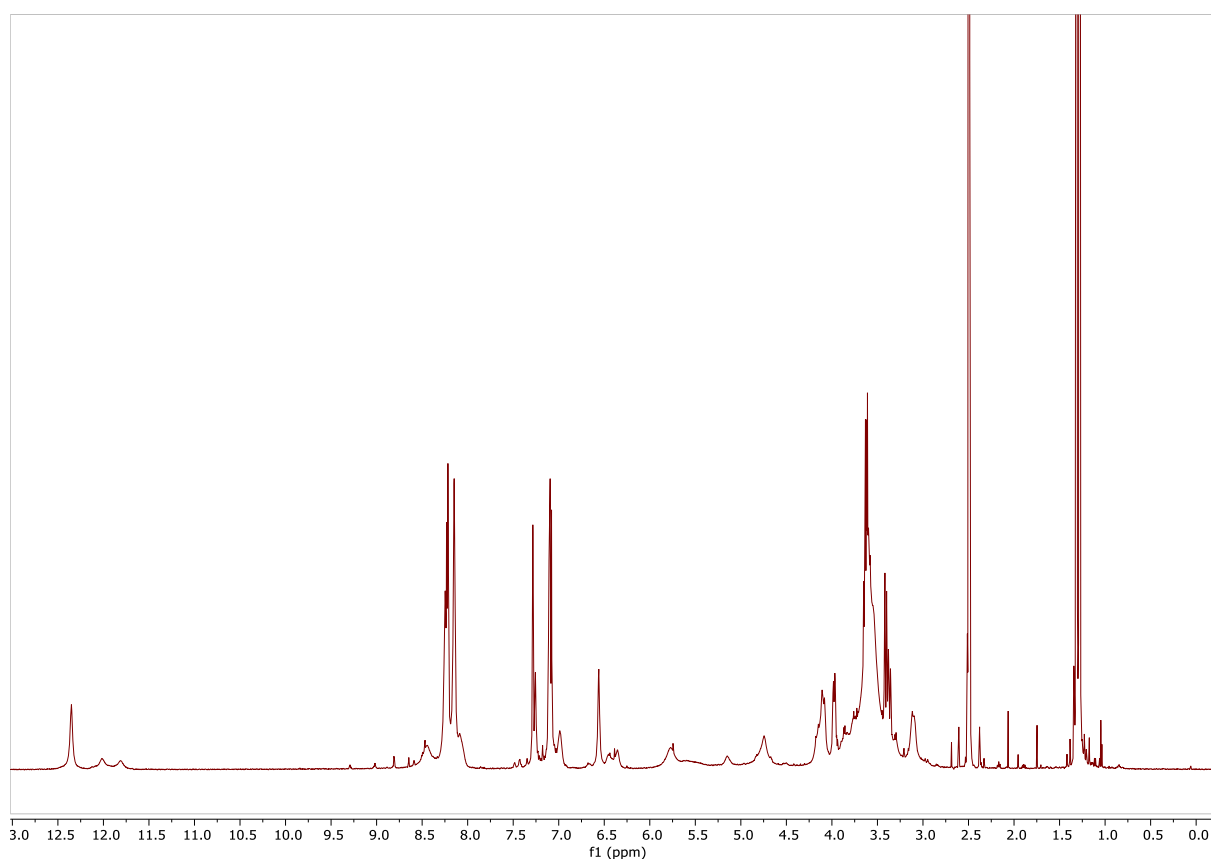
Three major conformers were observed in a 65:21:14 ratio. A red/blue/orange colour code is accordingly used to describe peaks in the following $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ characterization. A black colour refers to a signal or a multiplet not clearly attributed to only one of the conformers. Only unambiguously

identified signals are reported (assignment based on the combined analysis of 2D ^1H COSY, ^1H - ^{13}C and ^1H - ^{15}N HSQC spectra).

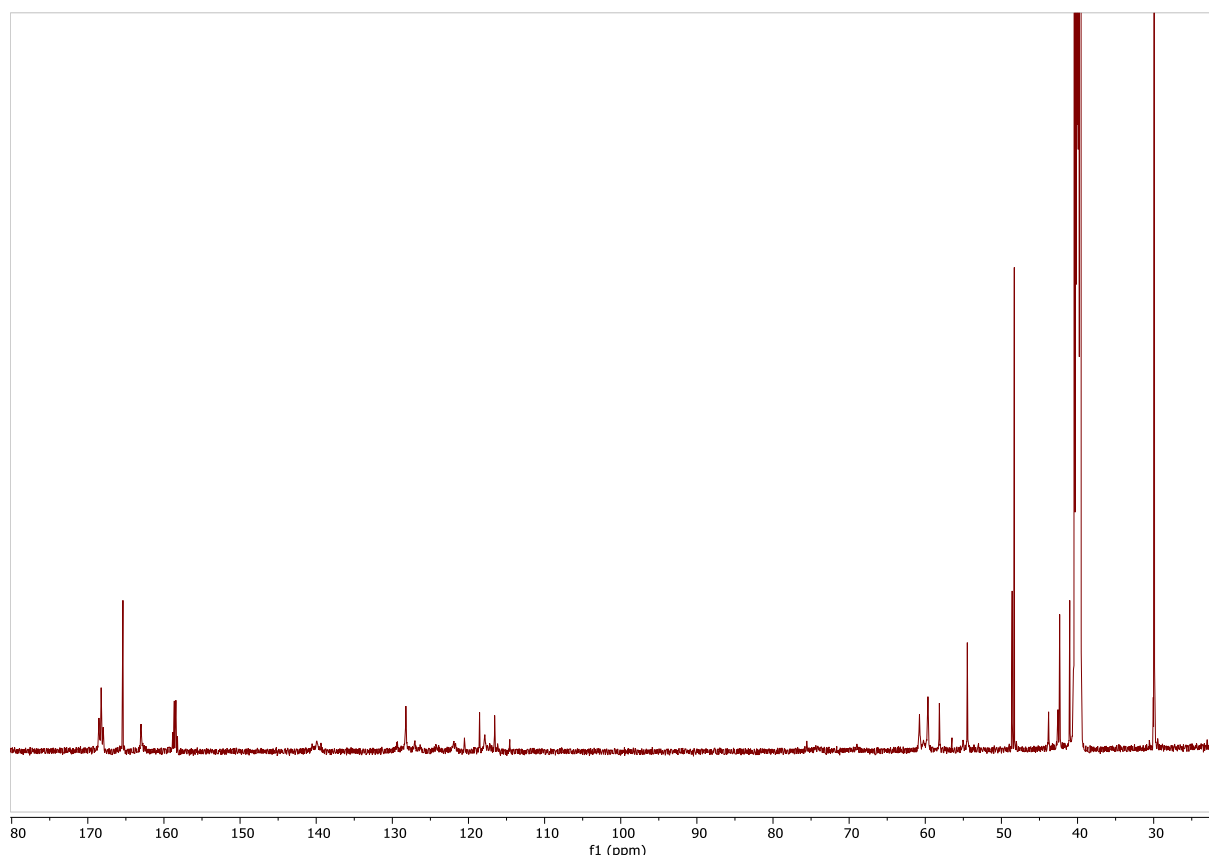
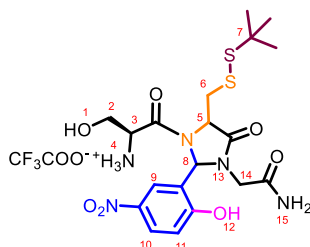
^1H NMR (600 MHz, DMSO- d_6) δ 12.35 (bs, 1H, H_{12}), 12.01 (bs, 1H, H_{12}), 11.81 (bs, 1H, H_{12}) 8.26 – 8.20 (m, 4H + 2x1H + 2x1H, H_4 + H_{10} + H_9), 8.16 – 8.13 (m, 3H, H_4), 7.29– 7.24 (m, 2x2H, H_{15}), 7.12 – 6.99 (m, 2x1H, H_{11}), 6.56 (s, 1H, H_8), 6.45 (bs, 1H, H_8), 6.36 (bs, 1H, H_8), 5.20 (bs, 1H, H_5), 4.75 (bs 1H, H_5), 4.15 – 4.09 (m, 1H + 2H, H_3 + H_{14}), 3.97 (d, J = 10.7 Hz, 2H, H_2), 3.67 – 3.56 (m, 2H + 2H + 1H, H_6 + H_2 + H_3), 3.44 – 3.36 (m, 2H, H_6), 3.11 (d, J = 16.0 Hz, 2H, H_{14}), 1.31 (s, 9H, H_7), 1.28 (s, 9H, H_7).

^{13}C NMR (151 MHz, DMSO- d_6) δ 168.5 C_{CO} , 168.3 C_{CO} , 168.0 C_{CO} 165.4 C_{CO} , 163.0 2 C_{OH} , 128.2 H_{10} , 118.5 H_{11} , 116.6 H_{11} , 76.0 C_8 , 74.1 C_8 , 68.9 C_8 , 60.8 C_2 , 59.6 C_2 , 58.1 C_5 , 56.5 C_5 , 55.0 C_3 , 54.5 C_3 , 42.6 C_{14} , 42.3 C_{14} , 41.0 C_6 , 40.5 C_6 , 30.0 C_7 , 29.9 C_7 .

^{15}N NMR (60 MHz, DMSO- d_6) δ 105.8 CONH_2 , 90.8 NH_3^+



Supplementary figure S26: ^1H -NMR spectrum of imidazolidinone **10a**

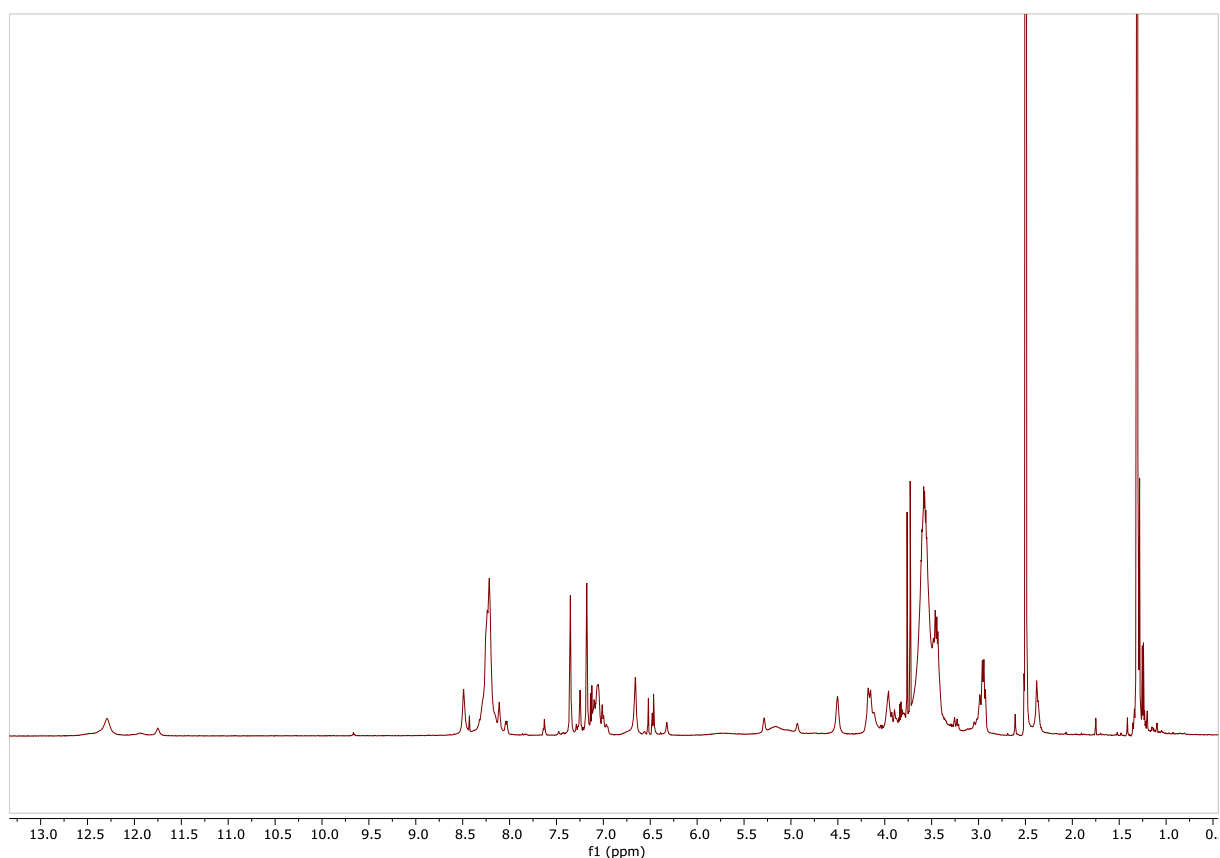
Supplementary figure 27: ^{13}C -NMR spectrum of imidazolidinone **10a** in DMSO➤ NMR analysis of imidazolidinone **10b**

Three major conformers were observed in a 57:23:20 ratio. A red/blue/orange colour code is accordingly used to describe peaks in the following $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ characterization. A black colour refers to a signal or a multiplet not clearly attributed to only one of the conformers. Only unambiguously identified signals are reported (assignment based on the combined analysis of 2D ^1H COSY, ^1H - ^{13}C and ^1H - ^{15}N HSQC spectra).

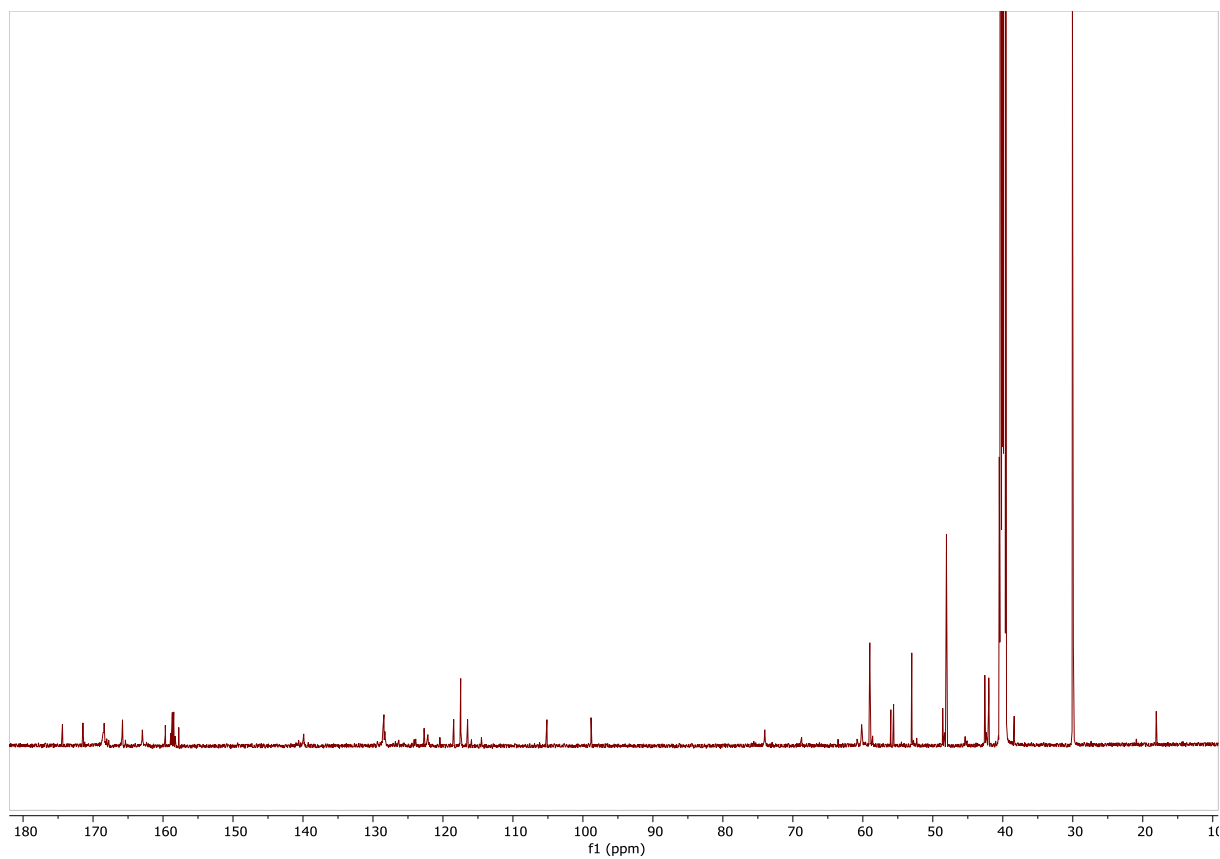
^1H NMR (600 MHz, DMSO- d_6) δ 12.29 (s, 1H, H_{12}), 11.75 (s, 1H, H_{12}), 8.49 (bs, 1H, H_{10}), 8.27 – 8.20 (m, 1H + 1H + 3H, H_9 + H_{10} + H_4), 8.11 (bs, 3H, H_4), 8.03 (d, J = 8.8 Hz, 1H, H_9), 7.35 (s, 1H, H_{15}), 7.18 (s, 1H, H_{15}), 7.09 – 7.03 (m, 2x1H, H_{11}), 6.66 (s, 1H, H_8), 6.56 (s, 1H, H_8), 6.32 (s, 1H, H_8), 5.29 (bs, 1H, H_5), 4.93 (bs, 1H, H_5), 4.50 (bs, 1H, H_5), 4.20 – 4.09 (m, 1H + 2H, H_3 + H_{14}), 3.96 (d, J = 10.7 Hz, 2H, H_2), 3.70 – 3.35 (m, 2H + 2H + 1H, H_6 + H_2 + H_3), 3.01–2.91 (m, 1H, H_6), 2.40–2.34 (m, 2H, H_{14}), 1.31 (s, 9H, H_7), 1.28 (s, 9H, H_7).

^{13}C NMR (151 MHz, DMSO- d_6) δ 168.5 C_{CO} , 168.4 C_{CO} , 165.8 C_{CO} , 165.3 C_{CO} , 162.9 2 C_{OH} , 128.4 H_{10} , 126.79 H_9 , 126.3 H_9 , 118.5 H_{11} , 117.5 H_{11} , 75.4 C_8 , 74.2 C_8 , 69.0 C_8 , 63.5 C_2 , 60.8 C_2 , 60.1 C_5 , 59.0 C_5 , 58.6 C_5 , 53.0 C_3 , 52.3 C_3 , 42.6 C_{14} , 42.5 C_{14} , 42.0 C_6 , 38.3 C_6 , 29.9 C_7 , 29.9 C_7 .

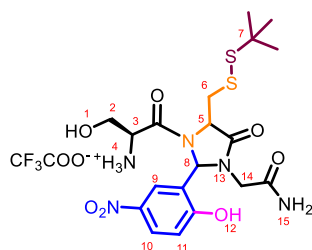
^{15}N NMR (60 MHz, DMSO- d_6) δ 106.1 CONH_2 , 92.8 NH_3^+



Supplementary figure S28: ^1H -NMR spectrum of **10b**



Supplementary figure S29: ^{13}C -NMR spectrum of **10b**

➤ NMR analysis of imidazolidinones **10c** and **10d** 67:33 mixture

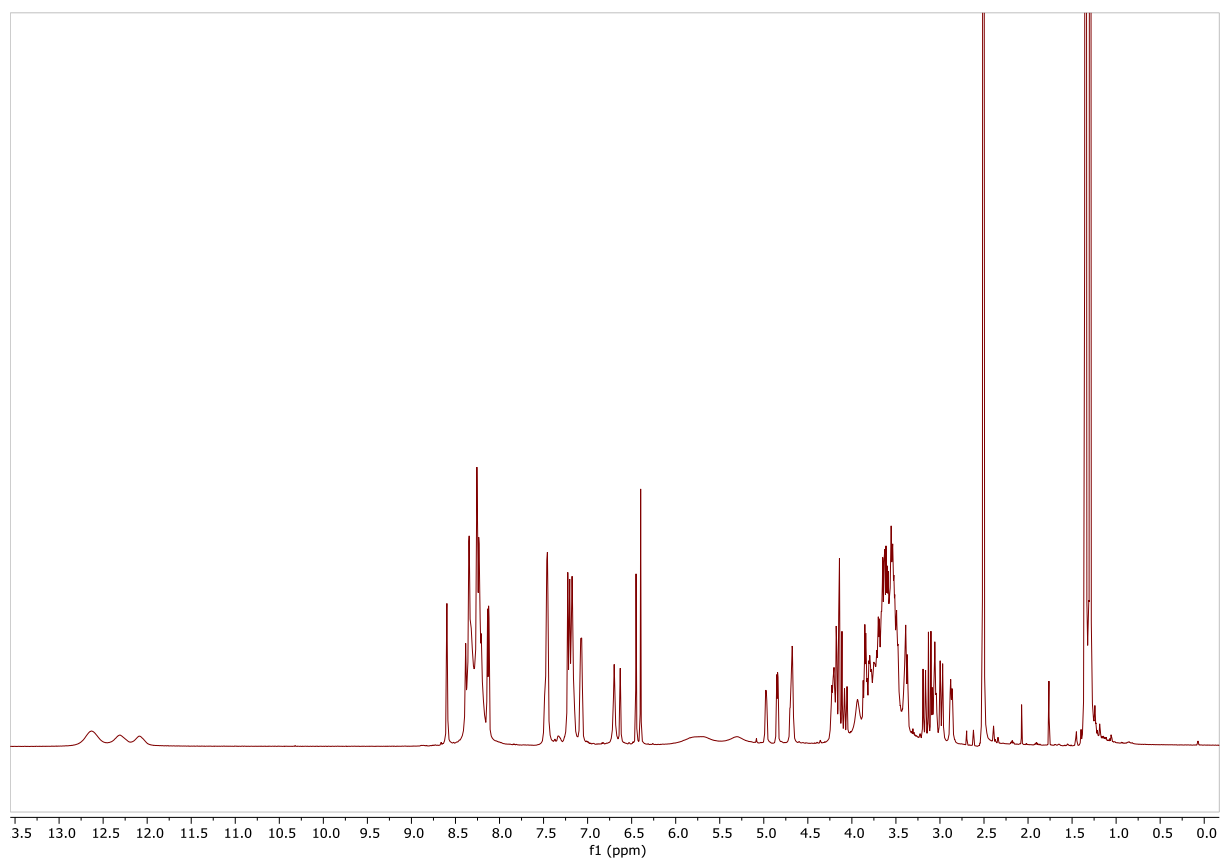
Two major conformers of both **10c** and **10d** were observed. A red/blue code colour was used to respectively describe the **major** and the **minor** conformer of **10c** in the following $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ characterization. Similarly, a purple/green code colour was used to respectively describe the **major** and the **minor** conformer of **10d** in the following $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ characterization. A black colour refers to a signal or a multiplet not clearly attributed to only one of the compounds **10c** and **10d**. Only unambiguously identified signals are reported (assignment based on the combined analysis of 2D ^1H COSY, ^1H - ^{13}C , ^1H - ^{15}N HSQC as well as ^1H of **10c** and **10d*** spectra).

* Small amounts of impure isolated **10c** and **10d** (see supplementary figures 25 and 26) were analysed as well.

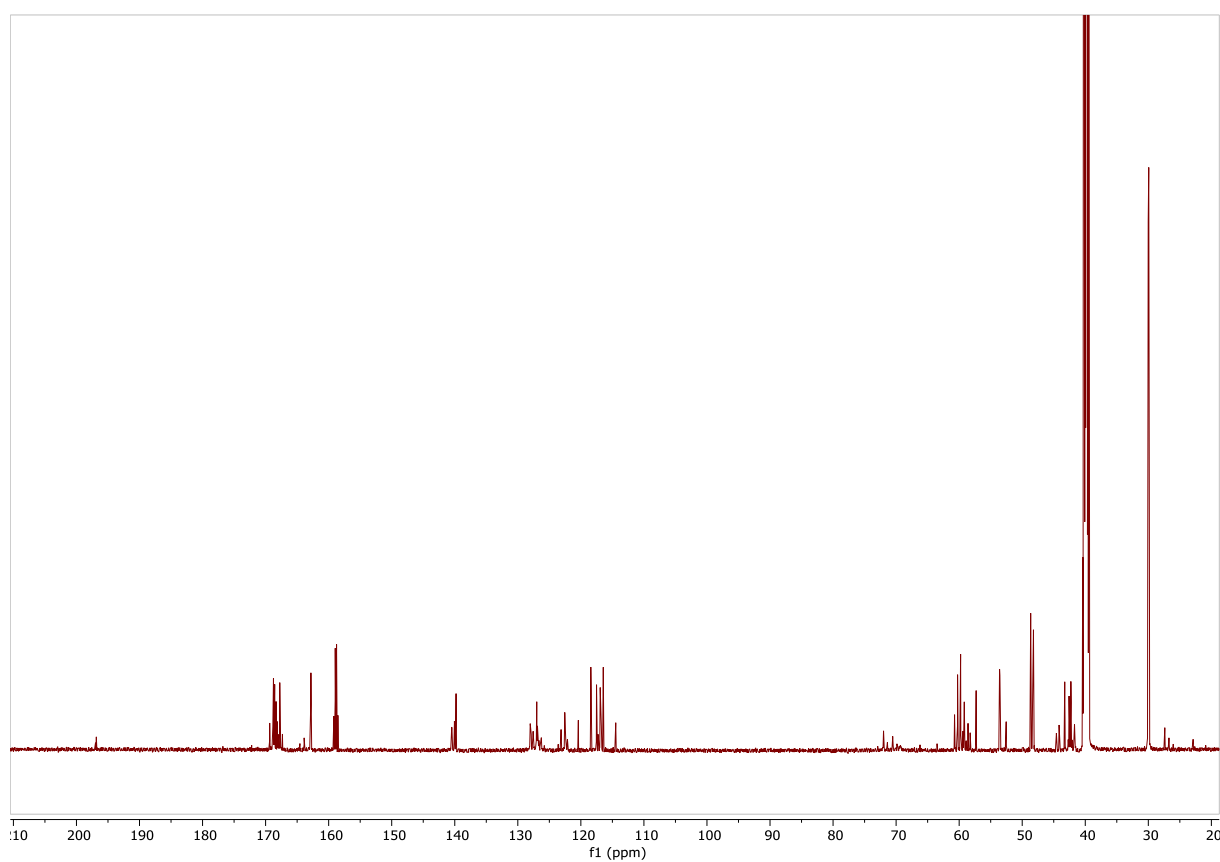
^1H NMR (600 MHz, DMSO- d_6) δ 12.62 (m, 2x1H H_{12-10d} , H_{12-10c}), 12.29 (s, 1H, H_{12-10c}), 12.07 (s, 1H H_{12-10d}) 6.70 (s, 1H, H_{8-10d}), 6.63 (s, 1H, H_{8-10c}), 6.45 (s, 1H, H_{8-10d}), 6.40 (s, 1H, H_{8-10c}), 4.98 (dd, $J = 7.3, 3.2$ Hz, 1H, H_{5-10d}), 4.85 (dd, $J = 8.1, 3.4$ Hz, 1H, H_{5-10c}), 4.70 – 4.64 (m, 2x1H, H_{5-10d} , H_{5-10c})

^{13}C NMR (151 MHz, DMSO- d_6) δ 71.6 C_{8-10d} , 70.8 C_{8-10c} , 69.5 C_{8-10d} , 69.4 C_{8-10c} , 59.2 C_{5-10c} , 59.0 C_{5-10d} , 58.8 C_{5-10d} , 57.3 C_{5-10c}

^{15}N NMR (60 MHz, DMSO- d_6) δ 105.7 CONH₂, 104.8 CONH₂, 104.5 CONH₂, 104.4 CONH₂, 91.9 2xNH₃⁺, 91.7 2xNH₃⁺



Supplementary figure S30: ^1H -NMR spectrum of imidazolidinones **10c/10d** 67:33 mixture



Supplementary figure S31: ^{13}C -NMR spectrum of imidazolidinones **10c/10d** 67:33 mixture

7-3 Structure elucidation of imidazolinones 10a-d

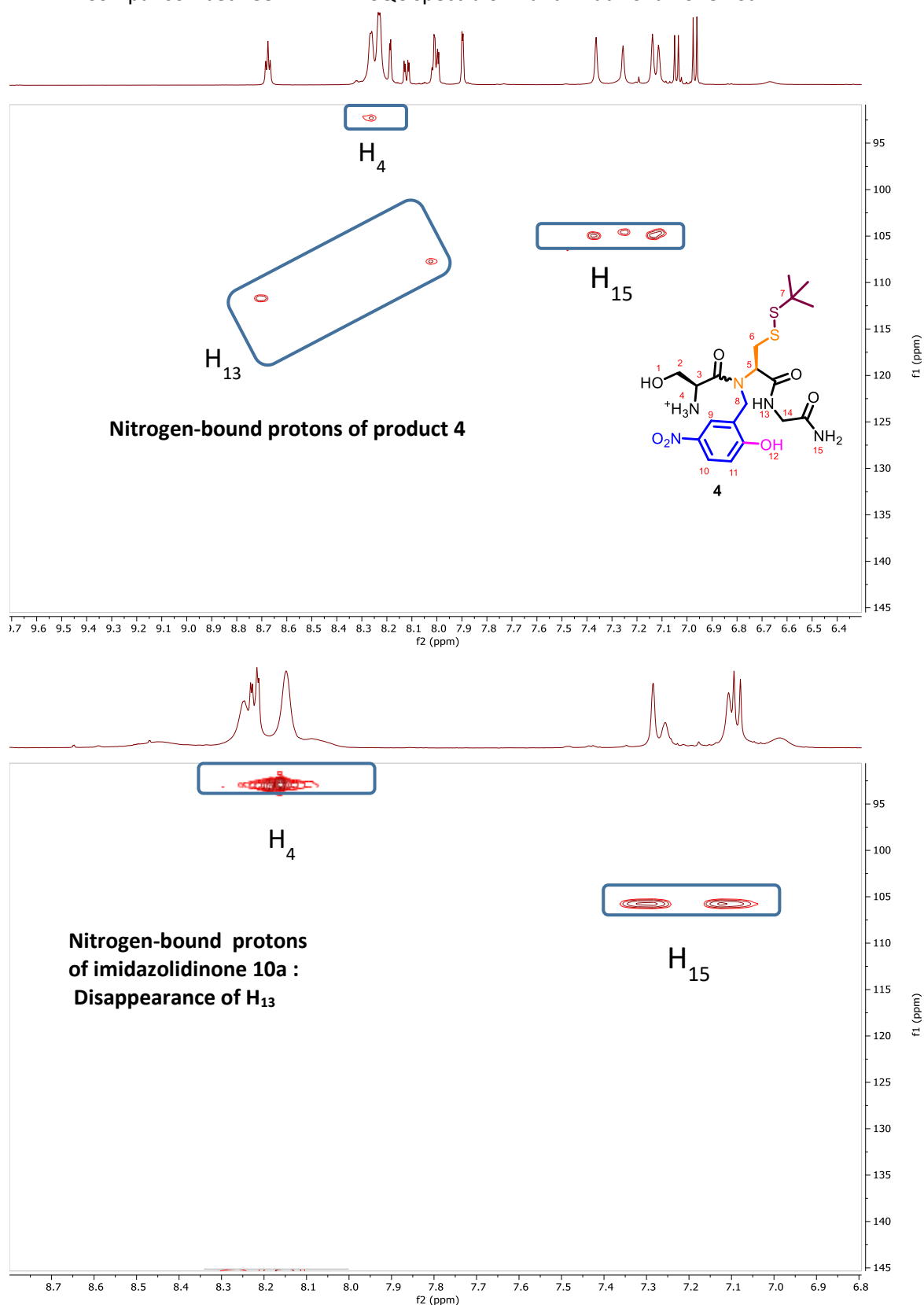
Combination of 1D and 2D ^1H , ^{13}C and ^{15}N experiments allowed us to conclude on the *N*-acyl-imidazolidin-4-ones structure of **10a-d**.

1- We observed for the four products the disappearance of the benzylic CH_2 signals observed in both **4** and **5**, accompanied with the apparition of new CH singlets at 6.0-6.7 ppm. This suggested the formation of compounds C-disubstituted in α of the nitrophenol ring.

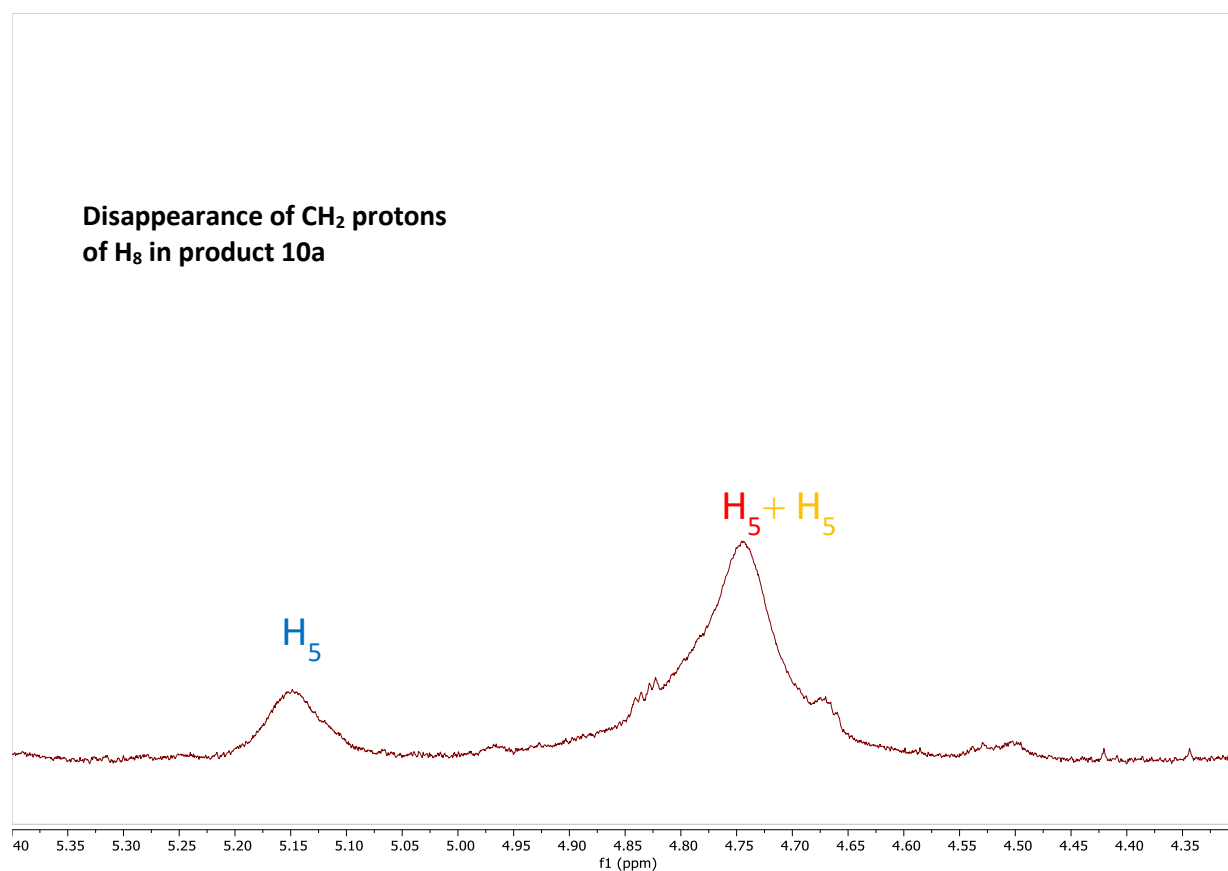
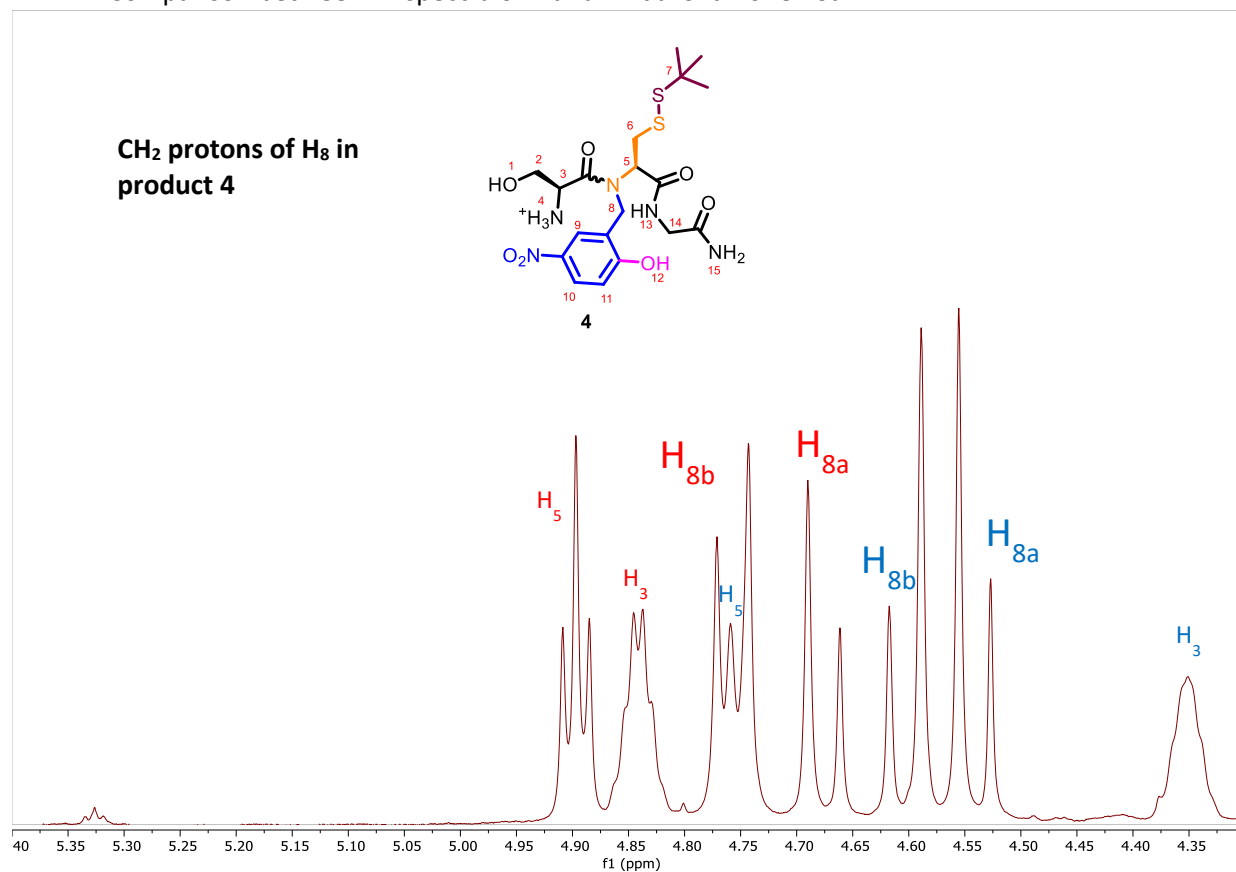
2- ^1H - ^{15}N HSQC spectra of **10a-d** showed the disappearance of the NH amide signal of the Cys-Gly peptide bond, while glycine C-terminal CONH_2 and serine NH_3^+ (TFA salt) remained essentially unchanged: this suggested an *N*-substitution of the Cys-Gly amide nitrogen.

We consequently deducted the probable *N*-acyl imidazolidinone structures shown in scheme 5 (see main article).

- Compound **10a**
- Comparison between ^1H - ^{15}N HSQC spectra of **4** and imidazolidinone **10a**

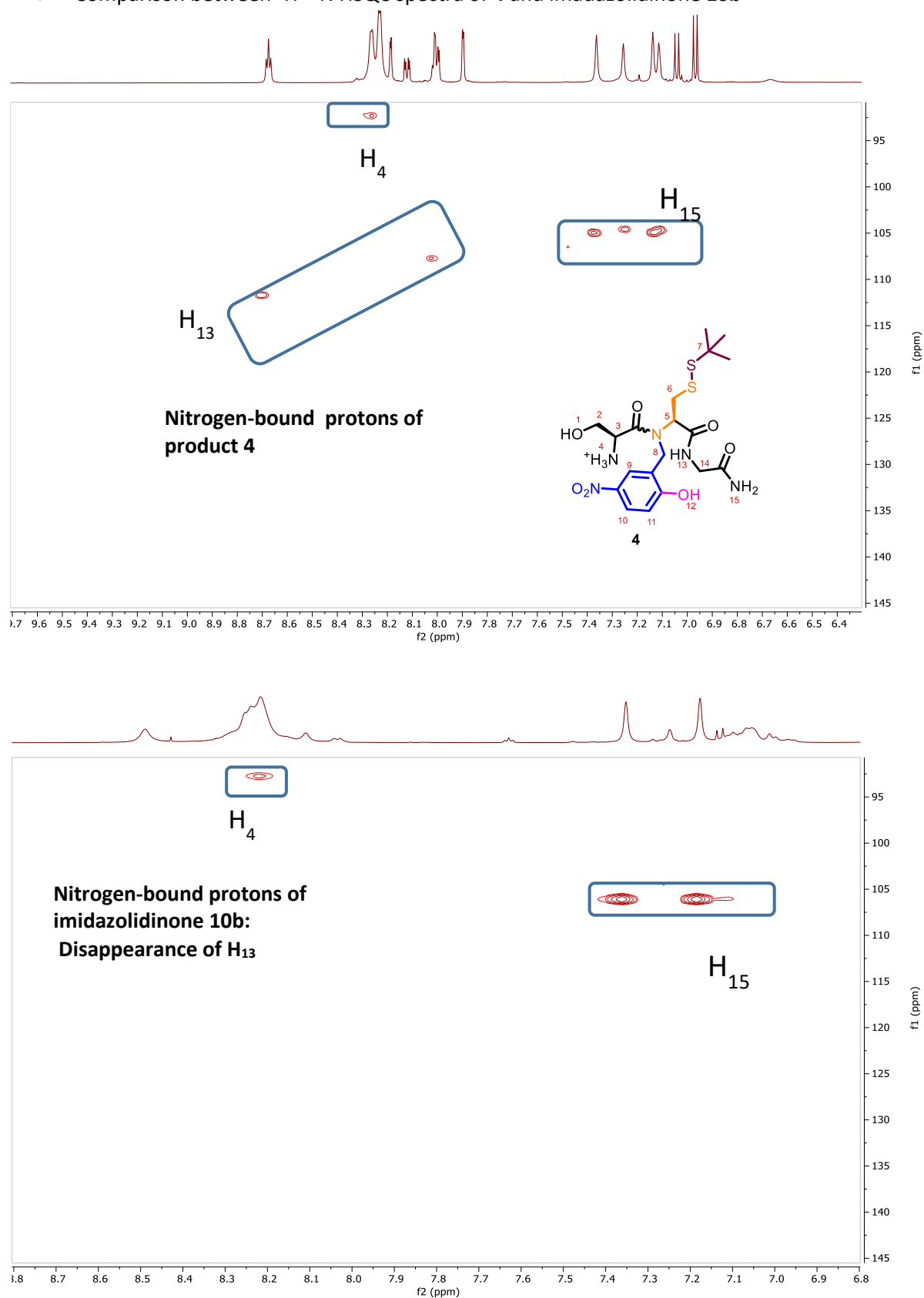


Supplementary figure S32: Comparison of ^1H - ^{15}N HSQC NMR spectra of **4** and **10a** in $\text{DMSO}-d_6$

➤ Comparison between ^1H spectra of **4** and imidazolidinone **10a**Supplementary figure S33: zoom on ^1H -NMR spectra of **4** and **10a** in $\text{DMSO-}d_6$

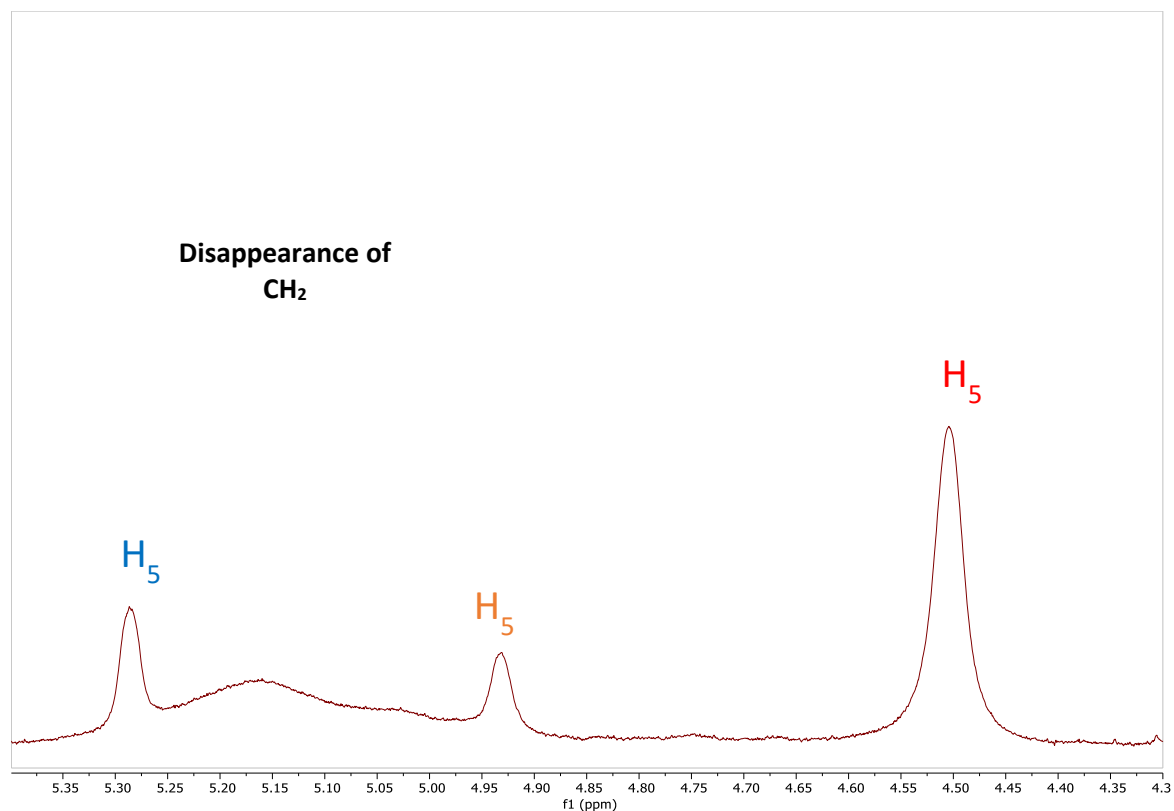
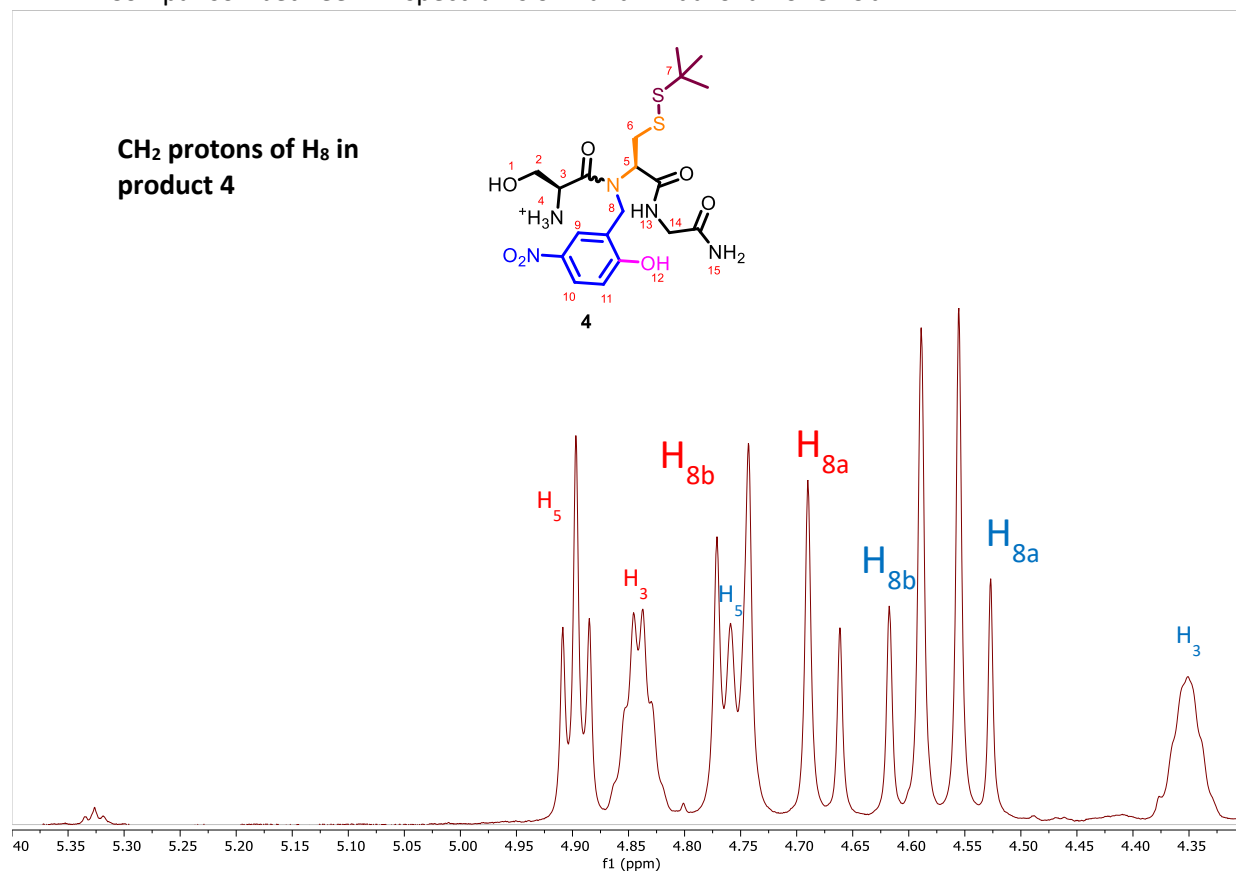
- **Compound 10b**

➤ Comparison between ^1H - ^{15}N HSQC spectra of **4** and imadazolidinone **10b**



Supplementary figure S34: comparison of ^1H - ^{15}N HSQC spectra of **4** and **10b** in $\text{DMSO}-d_6$

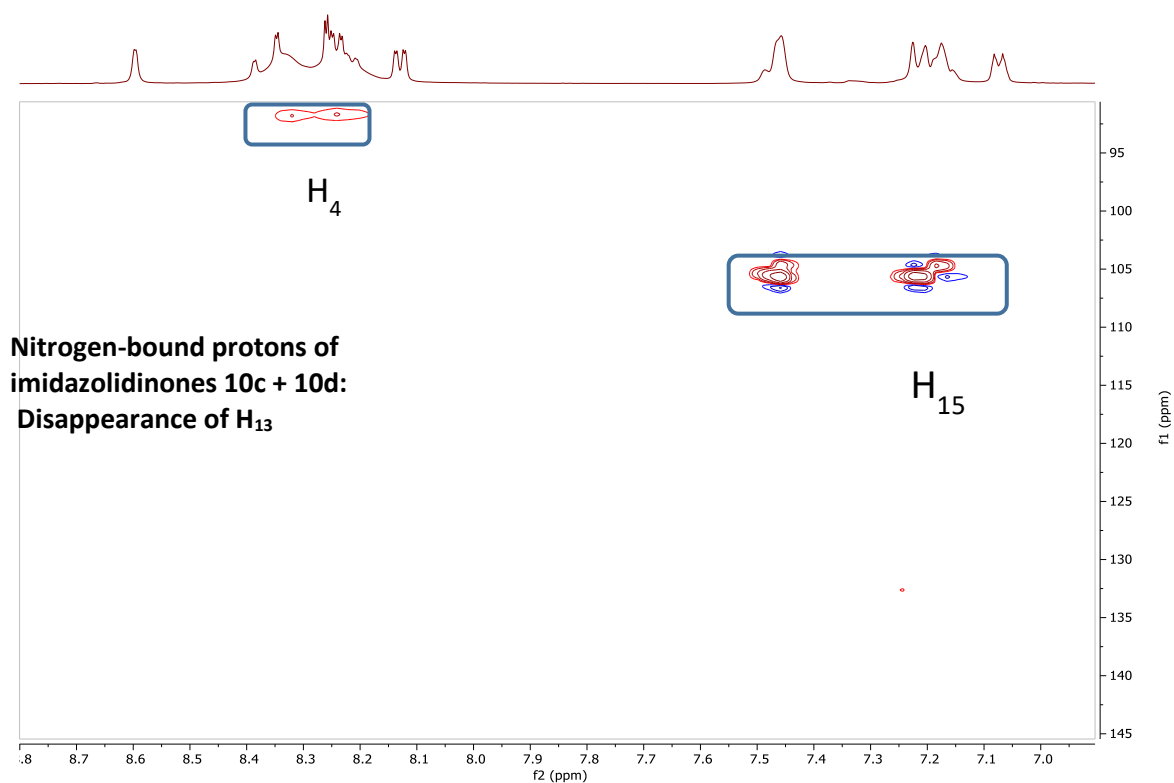
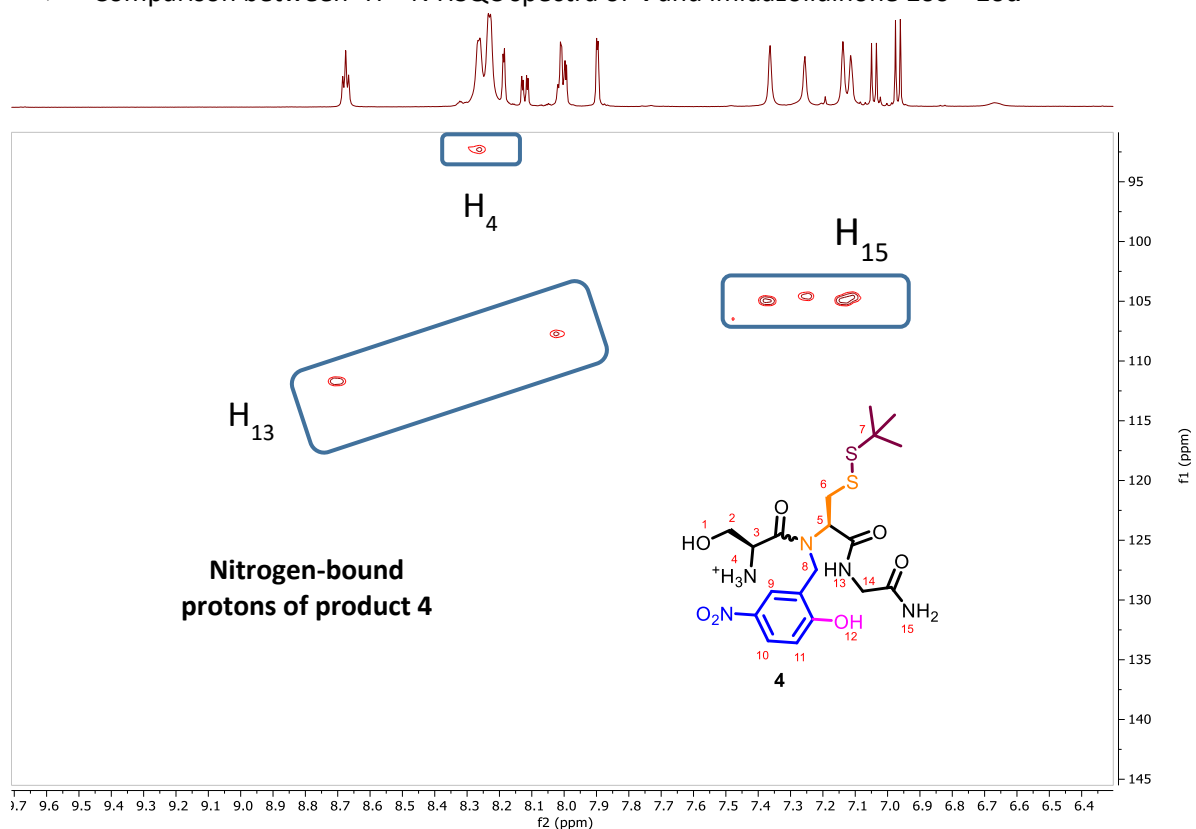
➤ Comparison between ^1H spectra of **4** and imidazolidinone **10b**



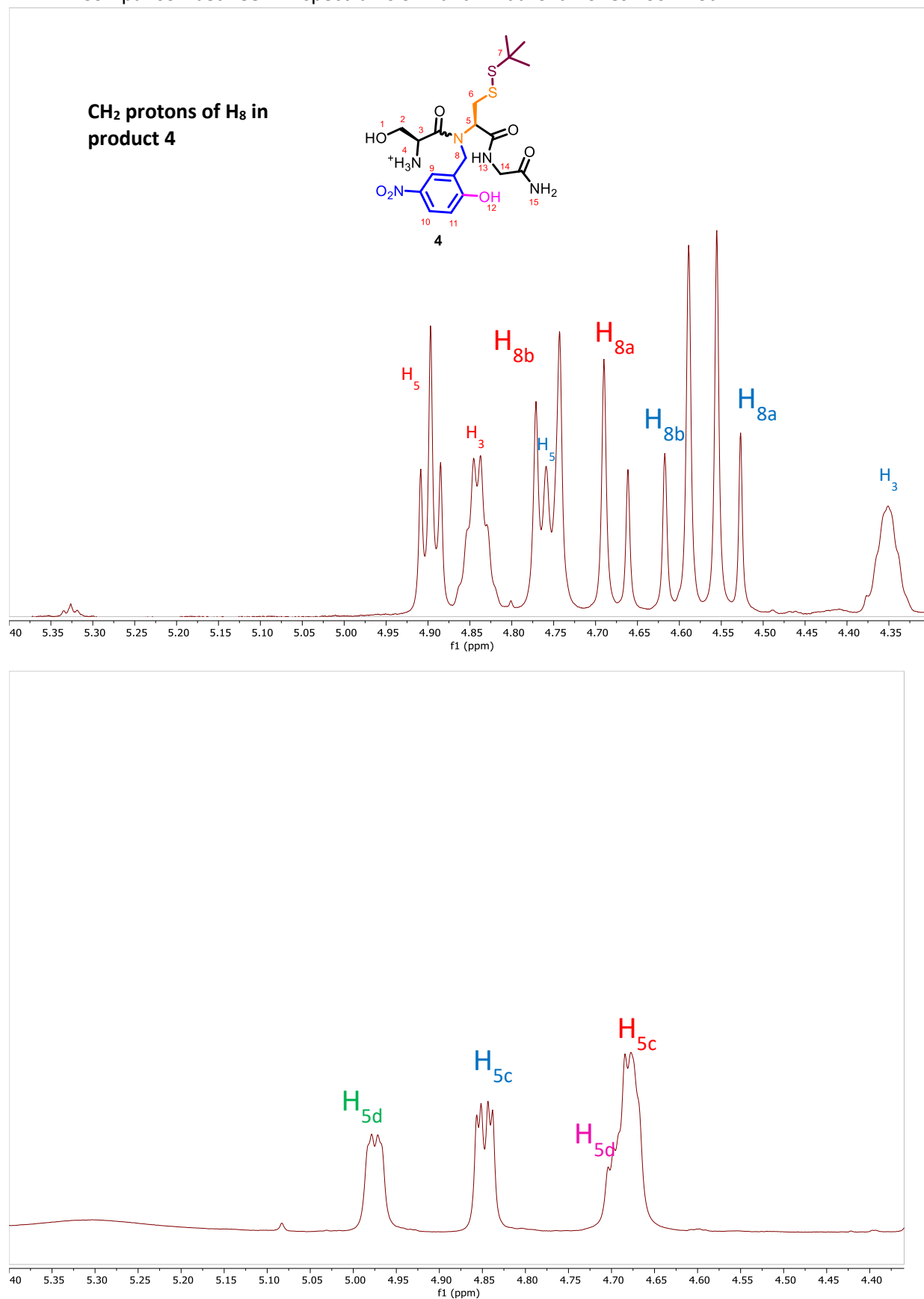
Supplementary figure S35: zoom on ^1H -NMR spectrum of product **4** and **10b** in $\text{DMSO}-d_6$

- **Compounds 10c/10d 77:33 mixture**

- Comparison between ^1H - ^{15}N HSQC spectra of **4** and imidazolidinone **10c + 10d**



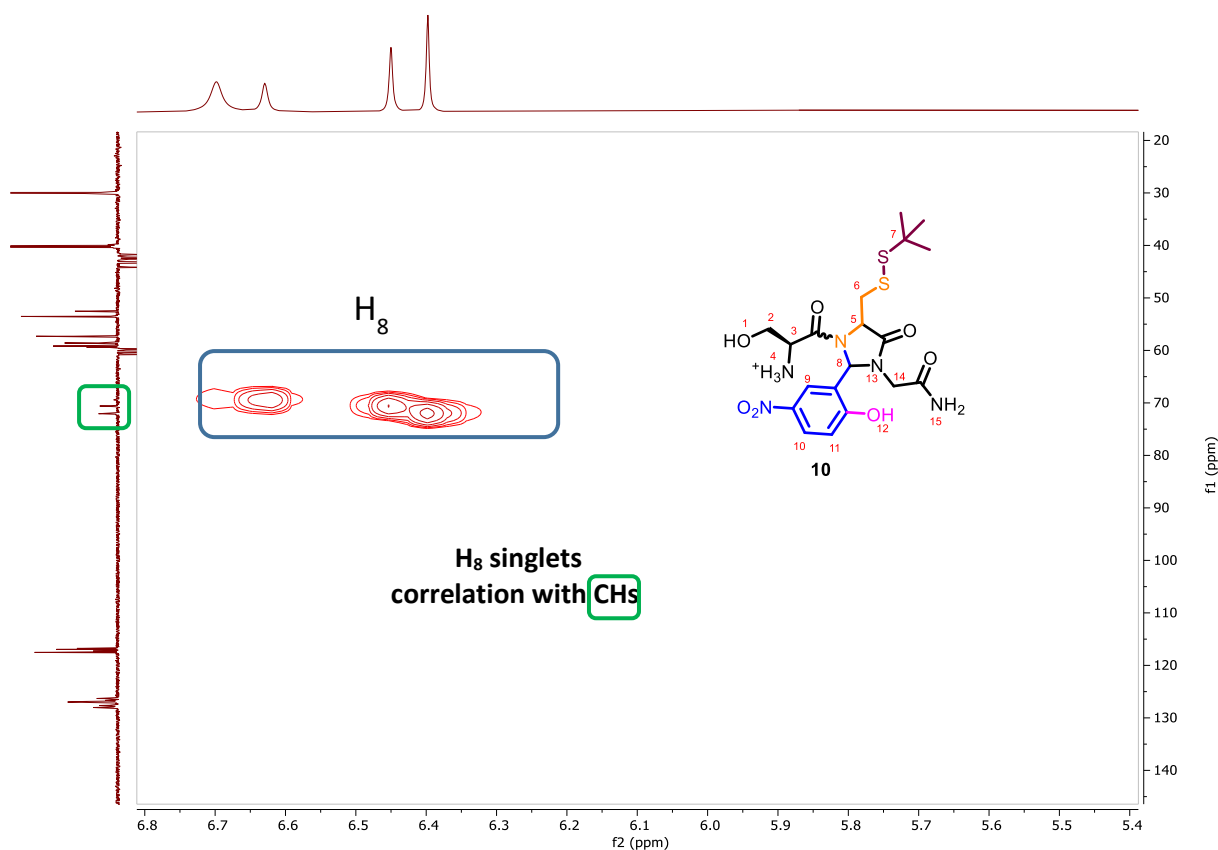
Supplementary figure S36: comparison of ^1H - ^{15}N HSQC spectra of **4** and **10c/10d** in DMSO-*d*₆

➤ Comparison between ^1H spectra of **4** and imidazolidinones **10c** + **10d**Supplementary figure S37: zoom on ^1H -NMR spectra of **4** and **10c** + **10d** in $\text{DMSO}-d_6$

The fact that we collected the mixture of **10c** and **10d** for NMR analysis yielded in bigger amount thus we could run additional analysis like DEPT 135 or HMBC. As explained above, in the case of **10c** and **10d**, only two sets of peaks were obtained (presumably *cis-trans* amide bond rotamers), yielding much simpler spectra than for **10a** and **10b**, where at least three sets of peaks corresponding to different conformers were identified.

➤ ^1H - ^{13}C HSQC and DEPT-135 spectra of imidazolidinones **10c/10d** 67:33 mixture

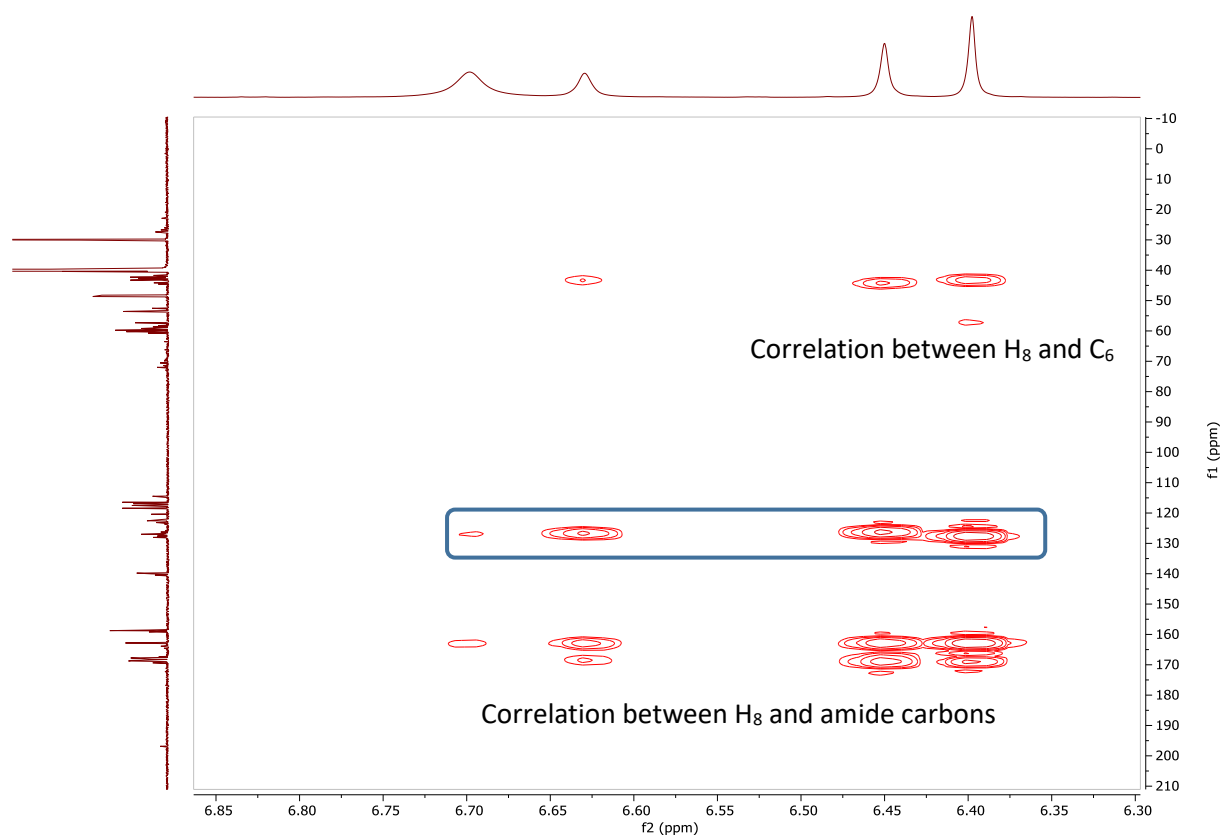
In the ^1H - ^{13}C HSQC and DEPT-135 NMR spectra we can clearly see the correlation between the newly formed H8 singlets at 6.4-6.7 ppm and CH carbons at 70-75 ppm (fig S44).



Supplementary figure S38: ^1H - ^{13}C HSQC and DEPT-135 NMR spectra of imidazolidinones **10c** + **10d** in DMSO

➤ ^1H - ^{13}C HMBC spectrum of imidazolidinones **10c/10d** 67:33 mixture

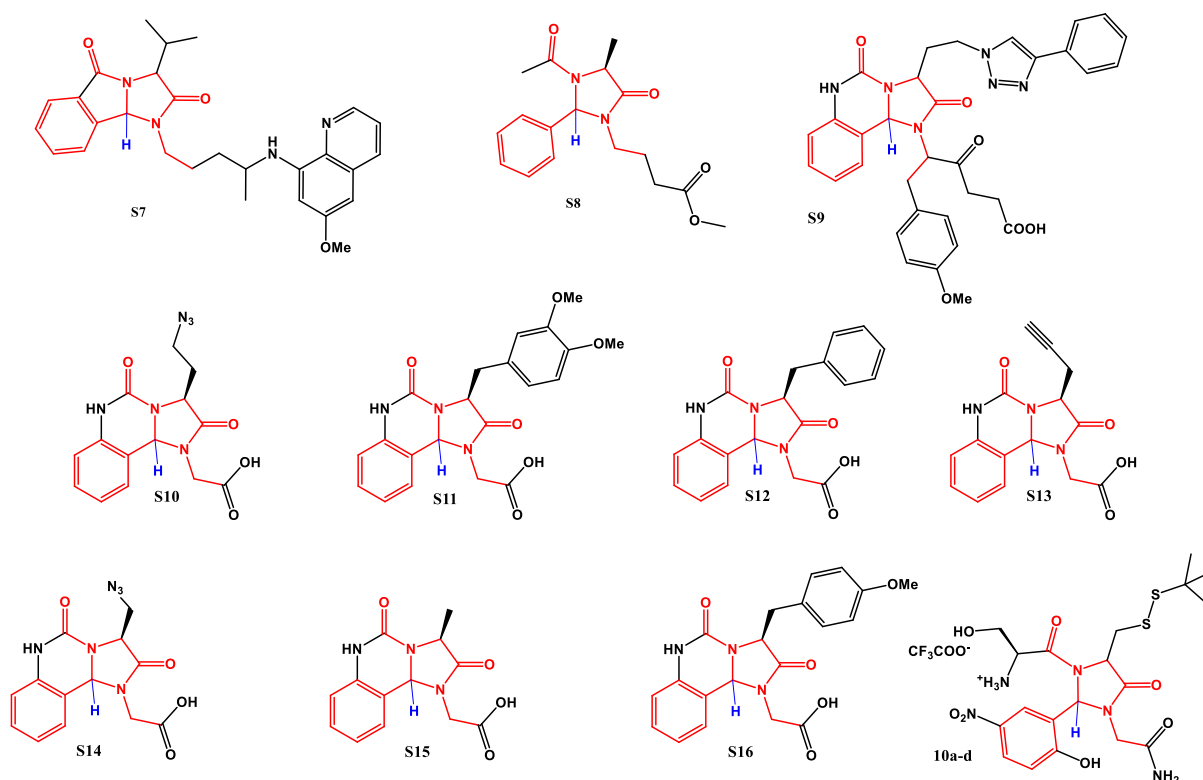
The HMBC spectra validates our attribution of these newly formed singlets as H8 protons since it clearly shows Correlation between H8 and respectively: C5, C6, aromatic carbons and amide carbons.



Supplementary figure S39: ^1H - ^{13}C HMBC NMR spectrum of imidazolidinones **10c + 10d** in $\text{DMSO-}d_6$

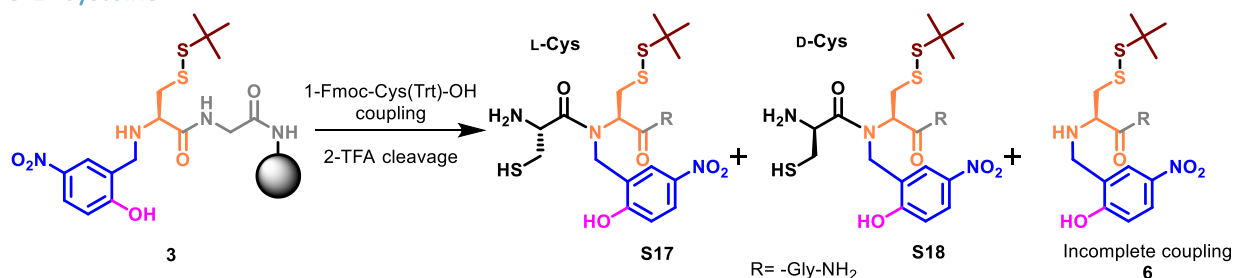
7-4- Comparison with ^1H and ^{13}C spectra of previously reported *N*¹-acyl-*N*³-alkyl-2-aryl-imidazolidin-4-ones❖ Table S4: Chemical shift of H2 and C2 of *N*¹-acyl-*N*³-alkyl-2-aryl-imidazolidin-4-ones previously described in the literature

Compound	Reference	^1H ppm	^{13}C ppm
S7	[10]	5.76 and 5.59	74.1 and 73.9
S8	[11]	5.98-6.23	74.1 and 73.9
S9	[12]	6.15	71.0
S10	[12]	6.27	70.1
S11	[12]	6.00	70.7
S12	[12]	6.03	70.3
S13	[12]	6.36	70.8
S14	[12]	6.34	71.0
S15	[12]	6.41	69.3
S16	[12]	5.97	70.3
10a	This work	6.56, 6.45 and 6.36	76.0, 74.1 and 69.6
10b	This work	6.66, 6.56 and 6.32	75.4, 74.2 and 69.0
10c	This work	6.70 and 6.63	69.5 and 69.4
10d	This work	6.45 and 6.40	71.6 and 70.8

Supplementary figure S40: Structures of *N*¹-acyl-*N*³-alkyl-2-aryl-imidazolidin-4-ones previously reported in the literature (H8 protons highlighted in blue)

8-Optimization of the coupling of the first amino acid

8-1- Cysteine



- H-Cys-(Hnb)Cys(StBu)-Gly-NH₂ (**S17**)

ESI-MS (*m/z*): [MH]⁺ calcd. for C₁₉H₃₀N₅O₆S₃: 520.1, found: 520.1

HPLC analysis: t_R = 8.2 min (Aeris Widepore XB-C18 2, gradient: 3-50% B'/A' over 15 min)

- H-Cys-(Hnb)Cys(StBu)-Gly-NH₂ (**S18**)

ESI-MS (*m/z*): [MH]⁺ calcd. for C₁₉H₃₀N₅O₆S₃: 520.1, found: 520.1

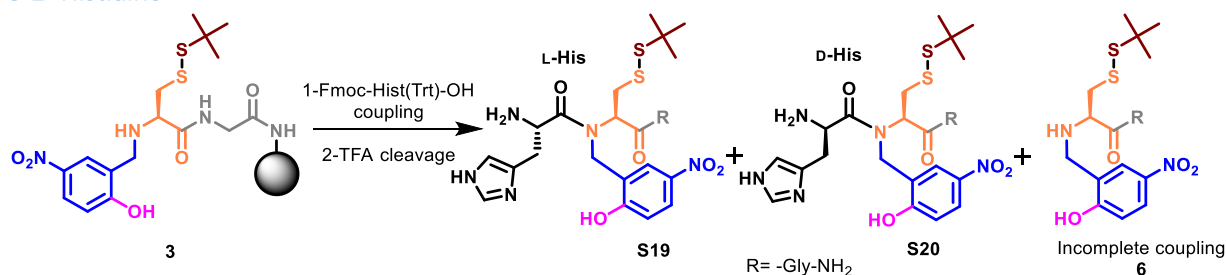
HPLC analysis: t_R = 7.4 min (Aeris Widepore XB-C18 2, gradient: 3-50% B'/A' over 15 min)

❖ Table S5: Optimization of the coupling of cysteine

Entry	AA coupled	Coupling conditions	S17 + S18 (%) ^[b]	S17:S18 ratio (%) ^{[b], [c]}	6 (%) ^[b]
1 ^[a]	Cys	HCTU/DIEA	91.4	96.5 : 3.5	9.6
2 ^[a]	Cys	HATU/DIEA	97.9	96.7 : 3.3	2.1
3 ^[a]	Cys	DIC/Oxyma	98.3	99.2 : 0.8	1.7
4 ^[a]	Cys	DIC/HOBT	93.2	98.4 : 1.6	6.8

[a] Coupling conditions : 18 h coupling in NMP, 10 mM peptidyl resin, RT, 10 equiv. Fmoc-Cys(Trt)-OH, 9.5 coupling agent, 20 equiv. DIEA and 10 equiv. Oxyma or HOBT; [b] Relative rates determined by HPLC peak integration at λ = 320 nm; [c] **S18** is coeluted with its enantiomer arising from Cys(StBu) epimerization during reductive amination under optimized conditions, determined to be 0.3%. Therefore, this amount was deducted from the **S18** rate relative to **S17** showed in this table, in order to show only Cys(Trt) epimerization.

8-2- Histidine



- H-His-(Hnb)Cys(StBu)-Gly-NH₂ (**S19**)

ESI-MS (m/z): [MH]⁺ calcd. for C₂₂H₃₂N₇O₇S₂: 554.2, found: 554.1

HPLC analysis: t_R = 5.8 min (Aeris Widespore XB-C18 2, gradient: 3-50% B'/A' over 15 min)

- H-His-(Hnb)Cys(StBu)-Gly-NH₂ (**S20**)

ESI-MS (m/z): [MH]⁺ calcd. for C₂₂H₃₂N₇O₇S₂: 554.2, found: 554.1

HPLC analysis: t_R = 5.7 min (Aeris Widespore XB-C18 2, gradient: 3-50% B'/A' over 15 min)

❖ Table S6: Optimization of the coupling of histidine

Entry	AA coupled	Coupling conditions	S19 + S20 (%) ^[b]	S19/S20 ratio (%) ^[b] [c]	6 (%) ^[b]
1 ^[a]	His	HCTU/DIEA	99.3	98.7 : 1.3	0.7
2 ^[a]	His	DIC/Oxyma	99.6	99.8 : 0.2	0.4

[a] Coupling conditions : 18 h coupling in NMP, 10 mM peptidyl resin, RT, 10 equiv. Fmoc-Cys(Trt)-OH, 9.5 coupling agent, 20 equiv. DIEA and 10 equiv. Oxyma

[b] Relative rates determined by HPLC peak integration at λ = 320 nm. [c] Relative rate determined by coupling of an extra Trp residue and integration of HPLC peaks at λ = 280 nm; [c] **S20** is coeluted with its enantiomer arising from Cys(StBu) epimerization during reductive amination under optimized conditions, determined to be 0.3%. Therefore, this amount was deducted from the **S20** rate relative to **S19** showed in this table, in order to show only Cys(Trt) epimerization.

8-3 Serine

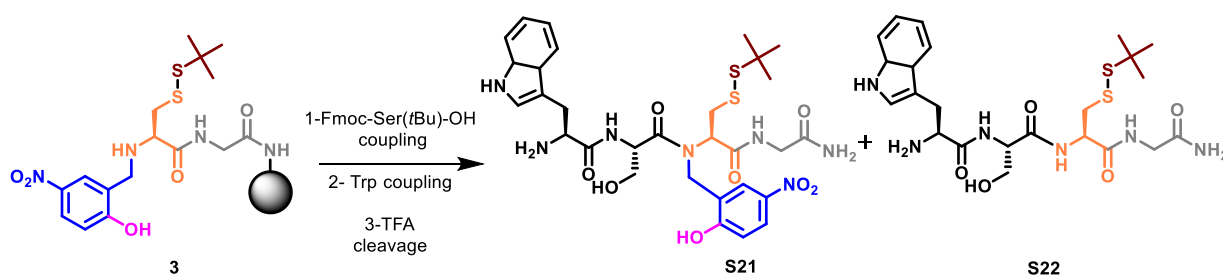
Table S7: Optimization of the coupling of serine

Entry	AA coupled	Coupling conditions	4 + 5 (%) ^[b]	4/5a ratio (%) ^{[b] [c]}	6 (%) ^[b]
1 ^{[a][d]}	Ser	DIC/Oxyma	99.5	99.1 : 0.9	0.5

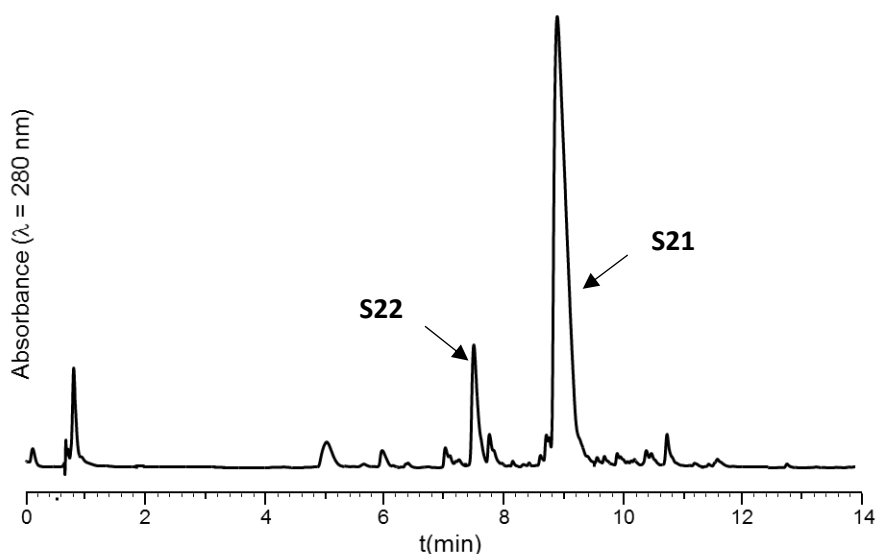
[a] Coupling conditions : 18 h coupling in NMP, 10 mM peptidyl resin, RT, 10 equiv. Fmoc-Cys(Trt)-OH, 9.5 DIC, 10 equiv. Oxyma; [b] Relative rates determined by HPLC peak integration at $\lambda = 320$ nm; [c] **5a** is coeluted with its enantiomer **5a** arising from Cys(StBu) epimerization during reductive amination under optimized conditions, determined to be 0.3%. Therefore, this amount was deducted from the **5a** rate relative to **4** showed in this table, in order to show only Cys(Trt) epimerization; [d] Reaction performed on three different resins: tentagel, ChemMatrix and polystyrene with similar results;

9-Quantification of the reductive amination yield

The coupling of a Trp residue allows the rate quantification of the non-Hnb bearing compound **S22** by integration of HPLC peaks at $\lambda = 280$ nm taking into account the molar absorption coefficient of Trp and Hnb at 280 nm: $\epsilon_{\text{Trp}} = 5500 \text{ L.mol}^{-1}.\text{cm}^{-1}$ $\epsilon_{\text{Hnb}} = 2395 \text{ L.mol}^{-1}.\text{cm}^{-1}$



Fmoc-Trp(Boc)-OH (10 equiv.) was coupled for 1 h on H-Ser(tBu)-(Hnb)Cys(StBu)-Gly-Rink ChemMatrix followed by piperidine treatment and TFA cleavage using general protocol PS1.



Supplementary figure S41: HPLC trace (Aeris Widepore XB-C18 2, gradient: 3-50% B'/A' over 15 min) of H-Trp-Ser-(Hnb)Cys(StBu)-Gly-NH₂ (**S21**) and H-Trp-Ser-Cys(StBu)-Gly-NH₂ (**S22**)

- H-Trp-Ser-(Hnb)Cys(StBu)Gly-NH₂ (**S21**)

ESI-MS (m/z): [MH]⁺ calcd. for C₃₀H₄₀N₇O₈S₂: 690.2, found: 690.2

HPLC analysis: t_R = 8.6 min (Aeris Widepore XB-C18 2, gradient: 3-50% B'/A' over 15 min)

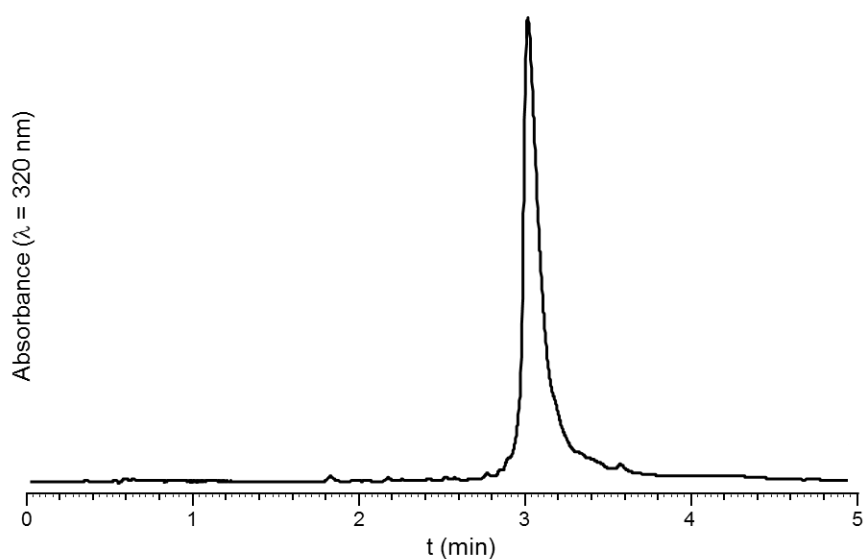
- H-Trp-Ser-Cys(StBu)Gly-NH₂ (**S22**)

ESI-MS (m/z): [MH]⁺ calcd. for C₃₀H₄₀N₇O₈S₂: 690.2, found: 690.2

HPLC analysis: t_R = 7.2 min (Aeris Widepore XB-C18 2, gradient: 3-50% B'/A' over 15 min)

10- Examples of crypto-thioesters synthesized with the optimized method

- LYRAA-(Hnb)Cys(StBu)-Gly-NH₂ (**S23**)

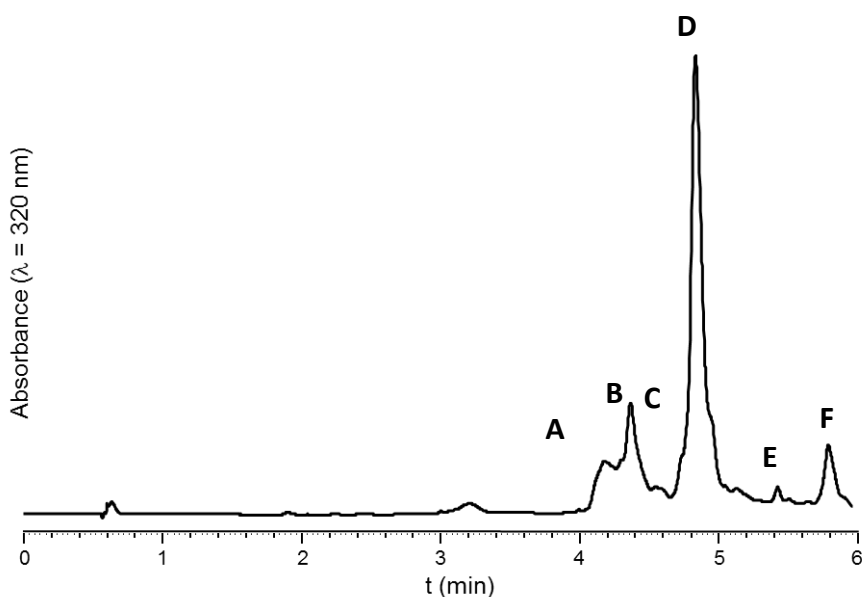


Supplementary figure S42: HPLC trace (Chromolith, gradient: 20-50% B/A over 5 min) of crude **S23**

Peak (t_R (min))	[MH] ⁺ (m/z) calcd.	[MH] ⁺ (m/z) found	Attributed to
A (3.29)	991.4	991.4	S23

- (MUC1)₃-(Hnb)Cys(StBu)-Gly-NH₂ (**S24**)

(Muc1)₃ = APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS

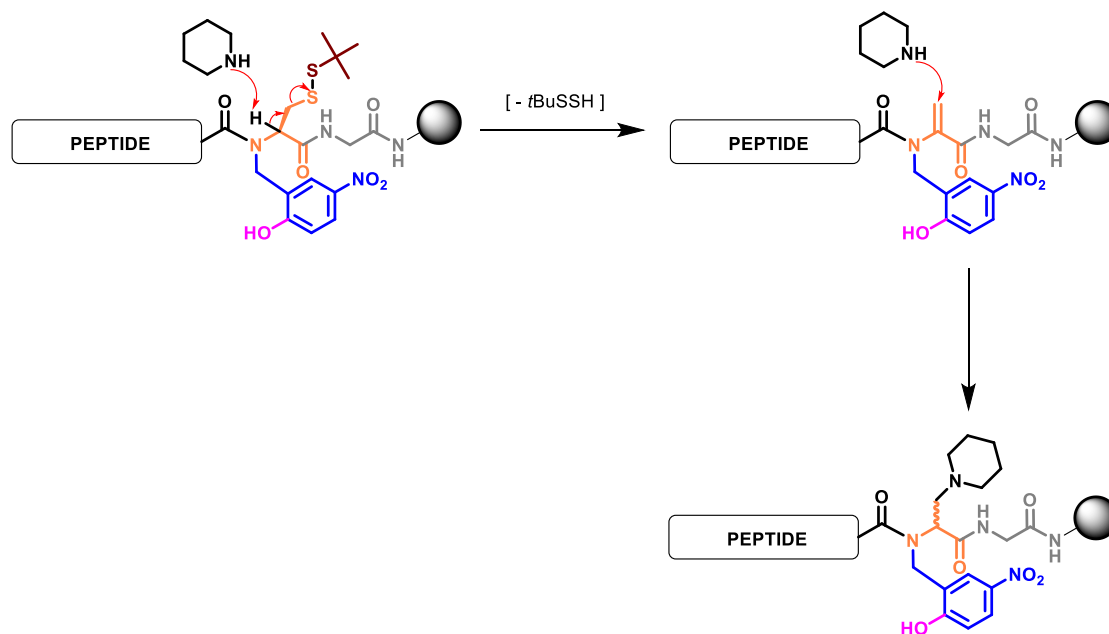


Supplementary figure S43: HPLC trace (Chromolith, gradient: 20-50% B/A over 5 min) of crude **S24**

Peak (t _R (min))	[MH] ⁺ (m/z) calcd. ^[a]	[MH] ⁺ (m/z) Found ^{[a] [b]}	Attributed to
A (4.22)	5681.2	5680.5	Ac-RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS-(Hnb)Cys(StBu)-Gly-NH ₂
B (4.42)	5986.4	5985.8	β-elimination + piperidine addition
C (4.61)	5923.5	5923.2	Ac-RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS-(Hnb)Cys(StBu)-Gly-NH ₂ + Trt
D (4.91)	6023.5	6022.8	S24
E (5.5)	2977.2	2976.7	Ac-PAHGVTSAPDTRPAPGSTAPPAHGVTS-(Hnb)Cys(StBu)-Gly-NH ₂
F (5.89)	1618.8	1617.7	Ac-PGSTAPPAHGVTS-(Hnb)Cys(StBu)-Gly-NH ₂

[a]: Average masses; [b] The multiply-charged envelope was deconvoluted using the charge deconvolution tool in Agilent OpenLab CDS ChemStation software.

As mentioned in the experimental part of the article, small amount of byproducts whose mass is consistent with β -elimination of the Cys(*StBu*) followed by Michael-type addition of piperidine is sometimes observed (figure S50). The amount was particularly high in the case of (MUC1)₃-(Hnb)Cys(*StBu*)-Gly-NH₂ **S33** (peak B, ~ 6 % based on the integration of extracted ion chromatogram for the most abundant multicharged species $[M+7H]^{7+}$).



Supplementary figure S44: putative mechanism of β -elimination followed by piperidine addition

11-Investigation of the formation of piperidinylalanine byproduct

In many cases of peptides as long as, or longer than **S24**, no piperidylalanine byproduct could be detected, which led us to question if this reaction was sequence-dependant, and not caused the presence of the *N*-Hnb group.

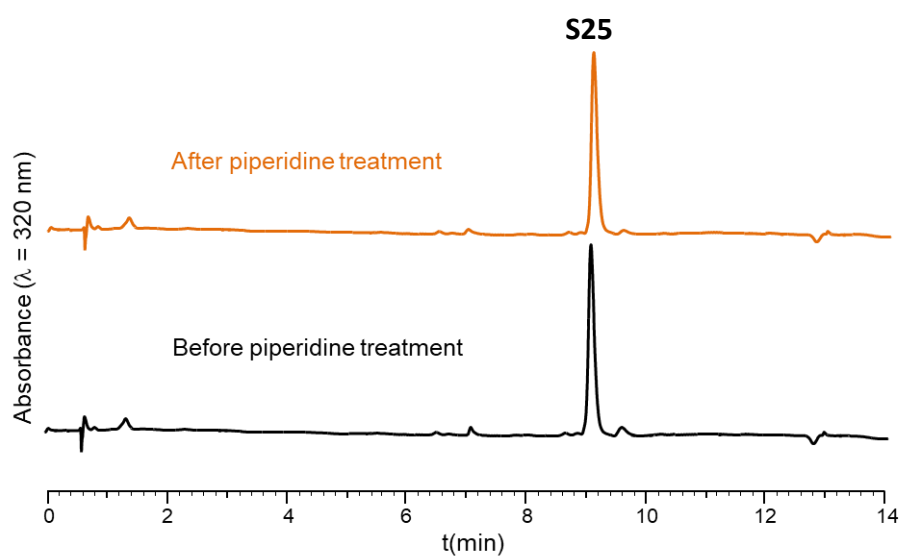
Peptidyl resin Ac-Ser(*tBu*)-(Hnb)Cys(*StBu*)-Gly-Rink ChemMatrix (**S25**) was treated with 20% piperidine in NMP for 42 hours at room temperature (equivalent to 280 deprotection cycles of 3 x 3 min). TFA cleavage using protocol PS1-small peptide cleavage gave tripeptide Ac-Ser-(Hnb)Cys(*StBu*)-Gly-NH₂ (**S26**). No trace of a β -elimination/piperidine addition by-product could be detected by LC-MS analysis. Moreover, a sample not treated with piperidine gave an essentially identical chromatogram.

This means that the formation of the β -elimination byproduct is very probably sequence-dependant, and not a general side reaction associated with the *N*-Hnb-Cys device.

- Ac-Ser-(Hnb)Cys(*StBu*)-Gly-NH₂ (**S25**)

ESI-MS (m/z): $[MH]^+$ calcd. for C₂₁H₃₂N₅O₈S₂: 546.16, found: 546.1

HPLC analysis: t_R = 9.03 min (Aeris Widepore XB-C18 2, gradient: 3-50% B'/A' over 15 min)



Supplementary figure S45: HPLC traces (Chromolith, gradient: 5-50% B/A over 5 min) of **S25** before and after piperidine treatment

12-References

- [1] V. P. Terrier, H. Adihou, M. Arnould, A. F. Delmas, V. Aucagne, *Chem. Sci.* **2016**, 7, 339–345
- [2] E.R. Koft, *Tetrahedron* **1987**, 43, 5775–5780.
- [3] J.H. Billman, J.W. MacDowell, *J. Org. Chem.* **1961**, 26, 1437–1440.
- [4] A. Johansson, E.-L. Lindstedt, T. Olsson, J. Huuskonen, K. Rissanen, W. Shi, S. Styring, C. Tommos, K. Warncke, B.R. Wood, *Acta Chem. Scand.* **1997**, 51, 351–335.
- [5] a) S. Kim, C.H. Oh, J.S. Ko, K.H. Ahn, Y.J. Kim, *J. Org. Chem.* **1985**, 50, 1927– 1932. b) S. Bhattacharyya, A. Chatterjee, S.K. Duttachowdhury, *J. Chem. Soc., Perkin Trans. 1* 1994, 1–2. c) X. Li, A.K. Yudin, *J. Am. Chem. Soc.* **2007**, 129, 14152– 14153.
- [6] J. Brussee, R.A.T.M. van Benthem, C.G. Kruse, A. van der Gen, *Tetrahedron Asymmetry* **1990**, 1, 163–166.
- [7] A.F. Abdel-Magid, K.G. Carson, B.D. Harris, C.A. Maryanoff, R.D. Shah, *J. Org. Chem.* **1996**, 61, 3849– 3862.
- [8] E. Podyacheva, O.I. Afanasyev, A.A. Tsygankov, M. Makarova, D. Chusov, *Synthesis* **2019**, 51, 2667– 2677.
- [9] a) K.A. Neidigh, M.A. Avery, J.S. Williamson, S. Bhattacharyya, *J. Chem. Soc., Perkin Trans. 1* **1998**, 2527– 2532. b) J.C. DiCesare, C.E. White, W.E. Rasmussen, B.M. White, C.B. McComas, L.E. Craft, *Synth. Commun* **2005**, 35, 663– 668
- [10] P. Gomes, M. J. Araújo, M. Rodrigues, N. Vale, Z. Azevedo, J. Iley, P. Chambel, J. Morais, R. Moreira, *Tetrahedron* **2004**, 60, 5551–5562
- [11] A. L. Satz, J. Cai, Y. Chen, R. Goodnow, F. Gruber, A. Kowalczyk, A. Petersen, G. Naderi-Oboodi, L. Orzechowski, Q. Strebel, *Bioconjugate Chem.* **2015**, 26, 1623–1632.
- [S12] Y. Wang, A.Ø. Madsen, F. Diness, M. Meldal, *Chem. Eur. J.* **2017**, 23, 13869–13874.

Chapter 4: A straightforward methodology to overcome solubility challenges for N-terminal cysteinyl peptide segments used for native chemical ligation

I. Bibliographic study	299
1) Poor solubility of peptide segments: a frequent problem in NCL-based chemical protein synthesis	299
2) Use of solubility-enhancing additives during NCL	300
3) Backbone modification	301
A. Backbone N-alkylation	301
B. Ser/Thr <i>O</i> -acyl isopeptides	302
4) Solubilizing tags	304
A. Permanent solubilizing tags	305
B. Solubilizing tags cleaved in a second step after NCL	305
a) C-terminal solubilizing tags	305
b) Solubilizing tags incorporated on amino acids side chains	306
i. Solubilizing tags on cysteine side-chain	307
ii. Solubilizing tags on glutamine, glutamic acid, arginine and aspartic acid side-chains	308
iii. Solubilizing tags on Lysine side-chains	309
c) Solubilizing tags incorporated on amides backbone	310
C. Solubilizing tags cleaved during NCL	299
a) NCL-labile solubilizing tags for C-terminal thioester segments	300
b) Towards a NCL-labile solubilizing tags for N-terminal cysteinyl segments	313
II. Aim of this work	315
III. References for the bibliographic study part	316
IV. Publication 4 (Published in Chemical Science)	310
V. Supporting Information	331

I. Bibliographic study

1) Poor solubility of peptide segments: a frequent problem in NCL-based chemical protein synthesis

The poor solubility of peptides and proteins is a common and troubling issue in peptide and protein chemical synthesis. Besides “standard” solubility, which is usually an easily measurable parameter for small molecules in a given solvent, peptides and proteins can also undergo aggregation over time, and slowly form precipitates from an initially clear solution. The term aggregation is used to describe a number of different processes during which peptide molecules associate into larger species through non-covalent interactions. Depending on the peptides or proteins involved, these species can be disordered or highly organized, such as in the cases of the β -amyloid peptides, amylin and the prion protein, involved in Alzheimer’s disease, type 2 diabetes and Creutzfeldt-Jakob disease, respectively. [1]

Low solubility and aggregation are also a common problem when handling protected peptides, which combine a highly hydrophilic amide-based peptide backbone with highly lipophilic side chain protective groups, resulting in low solubility in both polar (including water) and apolar solvents.

This problem tends to become more pregnant as the size of the peptide increases, and is one of the main reasons for the development of chemoselective ligation strategies for chemical protein synthesis through the assembly of unprotected peptides, as a more general alternative to the coupling of protected segments.

Poor solvation is also a frequent problem encountered during solid phase peptide synthesis. If the solid support acts as a co-solvent and disfavor interactions between distinct peptide chains owing to the “pseudo-dilution” effect, [2] so-called “difficult sequences” [3] tends to form solid-supported aggregates that prevent efficient peptide elongations. A wide number of methodologies have been developed to solve this issue. [4–6] If an extensive description of these methodologies is clearly out of the scope of the present bibliographic study, a few of it have been successfully extended to the solubilization of unprotected peptide segments used in NCL and will be very briefly described in the course of this chapter.

Even though unprotected peptides are generally much more soluble (in this case, in aqueous media) than protected ones, the poor solubility or tendency to aggregate of some of the peptide segments is recognized as one of the major current limitations of NCL-based protein synthesis. This problem can be encountered either during the NCL reaction itself, or during the purification and characterization of a segment. If such solubility issues are clearly anticipated when synthesizing a very hydrophobic target such as a transmembrane protein, segments from soluble hydrophilic proteins also frequently prove to be problematic. Segments containing hydrophobic clusters (several contiguous residues with lipophilic side-chains), other “aggregation hotspots” or ones including beta sheets in the folded protein 3D structure tend to be generally more prone to solubility and aggregation issues. Solubility also tends to decrease when the pH is getting closer to the isoelectric point of the segment (pI is the pH at which a molecule carries no net electrical charge). However, there is no such general rules easy to define, and the solubility of a given segment is often extremely hard to predict: the most common way is usually to synthesize the segment and check experimentally its solubility.

Many methodologies have been developed to circumvent these segment solubility issues, and the present bibliographic study aims at briefly presenting the different strategies. Note that the recent reviews elaborated by the group of Guo [7], or the one from the group of Tietze focusing on highly hydrophobic peptides [8] constitute useful guides for those who are getting started in this field.

2) Use of solubility-enhancing additives during NCL

A particularly efficient approach to improve the solubility of peptide segments during the NCL reaction relies on the use of additives. The use of denaturants like the chaotropic salt guanidine hydrochloride (Gu.HCl) [9] often considerably improves the solubility of the peptide segments and the ligation product, and 6 M Gu.HCl is classically used even for segment devoid of any solubility issues. We can also mention urea as another classical protein denaturant used to improve solubility during NCL, [10,11] however, cyanate produced from urea can react with amines and thiols thus limiting its utilization. [12]

The addition of organic co-solvents such as dimethylformamide (DMF) [13,14] or dimethyl sulfoxide (DMSO) [13] have been proven to exhibit good ability to dissolve

difficult peptide. Moreover, the effect of alcohols on the solubility of proteins and peptides has been extensively studied, especially fluorinated alcohols like trifluoroethanol (TFE) [15] or hexafluoroisopropanol (HFIP). [16]

All these solvents have been widely used to solubilize peptide segments taking advantage of their ability to disrupt hydrogen bonds and hydrophobic interactions. However the addition of organic solvents have been shown to promote in some cases the oxidization of segments containing Met and Cys residues, as well as the hydrolysis of thioesters. [17]

The addition of detergents can also be a great help for hydrophobic segment solubilization. For instance, *n*-octyl- β -D-glucoside, a non-ionic detergent has attracted much attention as an additive in NCL. [11,18,19,19] In addition, *n*-dodecylphosphocholine [20] and sodium dodecyl sulfate [21,22] were also proved to be useful for solubility issues during NCL reactions. Nevertheless, the addition of the two latter detergents can hamper HPLC purification [23] and contaminate final ligation products. [10] Approaches which mimic the native environment of transmembrane domains such as ligation reactions in lipid bilayers [24,25] or lipidic cubic phase [26] were also explored.

The use of solubility-enhancing additive is widely used for the NCL reaction itself, but does not solve the solubility problems encountered during the purification, characterization and handling of peptide segment before NCL. A wide number of methods based on the covalent modification of the peptide segments have been developed, and tend to be more and more frequently used in recent reports of NCL-based protein synthesis, and especially for very long ones. [27]

3) Backbone modification

Besides the addition of external solubility-enhancing additives, backbone modification is another way to address low solubility issues.

A. Backbone *N*-alkylation

Backbone amide nitrogen protection through *N*-alkylation or introduction of proline residues is frequently used to minimize aggregation by disfavoring secondary structure formation and preventing intermolecular beta-sheets formation. [28,29]

In the context of NCL-based protein synthesis, Kent and Johnson [30] demonstrated that the introduction of *N*-methylated residues (fig. 1) as well as prolines in a transmembrane peptide segment greatly increased the solubility in aqueous solvents and allowed the ligation to be performed. However, these modifications will permanently remain in the peptide sequence.

Hmb (2-hydroxy-4-methoxybenzyl) was introduced in 1993 as a TFA-labile backbone-amide protecting group for Fmoc-SPPS in order to inhibit interchain association during peptide elongation. [31]

Danishefsky [32] demonstrated the utility of *N*-Hmb backbone protection (fig. 1) for unprotected segments used in NCL. For this purpose, he *O*-acetylated the phenol group on solid phase, in order to make the resulting *O*-Ac-Hmb resistant to TFA cleavage owing to the electron-withdrawing ability of the ester group that destabilize the corresponding carbocation. Deacetylation in solution prior to NCL regenerated the Hmb groups which were cleaved by TFA treatment after NCL-based assembly of the target proteins.

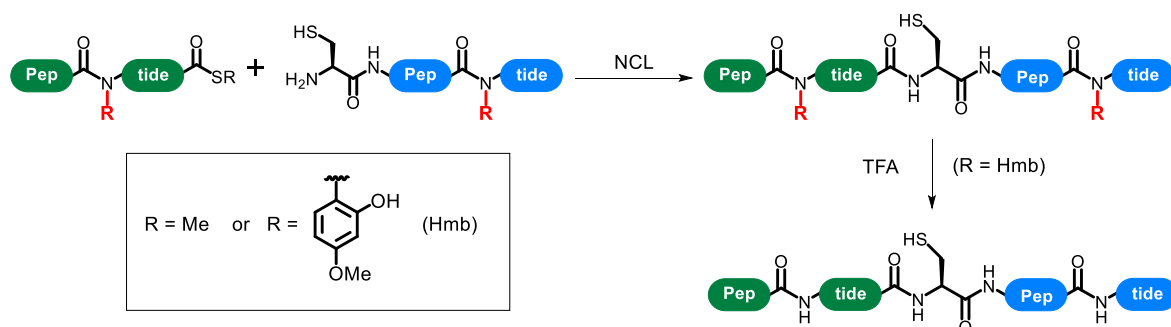


Figure 1: *N*-methylation and Hmb backbone modifications to improve solubility of peptide segments used for NCL

B. Ser/Thr *O*-acyl isopeptides

O-acyl isopeptides were initially developed by Kiso and co-workers [33] to reduce the aggregation during the Fmoc-SPPS of difficult peptides, using *N*-Boc dipeptide building blocks (fig. 2). This isopeptide structure is stable to TFA cleavage that deprotects the *N*-Boc group of the Ser or Thr residue, is also stable under classical HPLC purification conditions ($\text{H}_2\text{O}/\text{MeCN}/0.1\% \text{ TFA}$, $\text{pH} \sim 2$) but undergoes a fast spontaneous intramolecular *O*-*N* acyl migration under neutral conditions, generating the native amide bond. These pH-triggered properties make this strategy very well adapted to the

synthesis of peptides susceptible to aggregation when deprotected and in solution, in addition to making more efficient their solid phase syntheses.

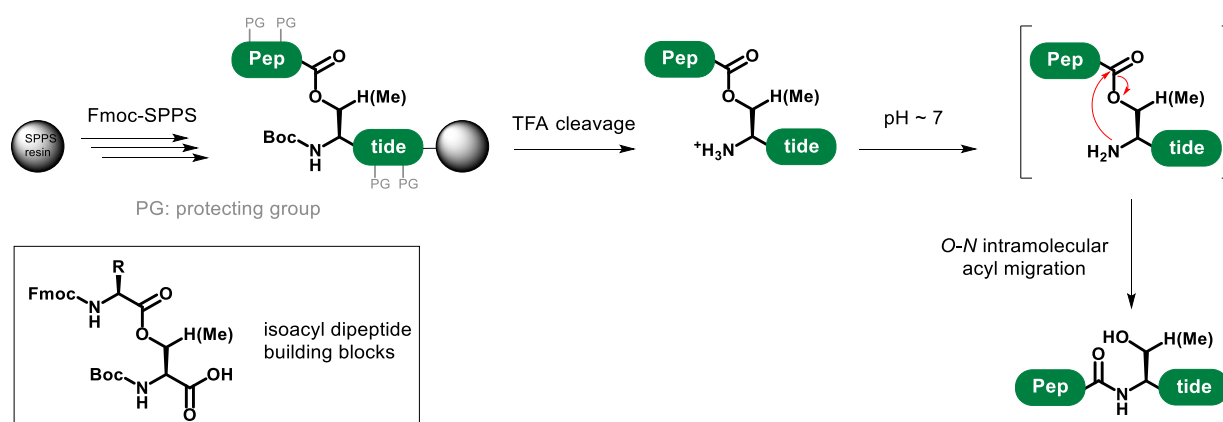


Figure 2: *O*-acyl isopeptide generation by Fmoc-SPPS

These properties also make it very suitable as a traceless approach in NCL, as it has been shown that the rearrangement readily happens under standard NCL conditions without unwanted by-products (fig. 3-a). For instance, the *O*-acyl isopeptide-containing segment approach was used as a traceless modification that was rapidly removed under NCL conditions for the synthesis of very long protein targets by Lei Liu and co-workers. [34] [35]

In some cases where the ligation product is still prone to aggregation, it is advantageous to keep the *O*-acyl isopeptide unit during the NCL-based protein assembly and further purification, and trigger the rearrangement in a further synthetic step (fig. 3-b). In order to make compatible both approaches, several strategies have been described:

Since very slow *O*-*N* acyl shift was noticed in organic solvents, Kiso and his co-workers [36] exploited this feature by conducting NCL in DMF in the presence of an organic base (triethylamine). [13].

Within the same context, Melnyk [37] took advantage of the fact that when using SEA crypto-thioester, NCL could be performed at pH = 3 thus preserving the *O*-acyl moiety during the course of NCL.

Another approach is the use of a TFA- and NCL-stable *N*-protective group for the *O*-acyl isodipeptide building block, instead of the Boc group. The group is deprotected in a second step after NCL, followed by *O*-*N* acyl migration under neutral conditions. For example, Akaji and his co-workers [38] used the Pd(0)-labile allyloxycarbonyl (Aloc)

amine protecting group, Nakahara and his co-workers [39] used azides as amine precursors. In the latter case, NCL was conducted without TCEP to avoid reduction of the azide. After protein assembly, TCEP was employed to trigger the rearrangement of the isopeptide by reducing the azido group thus enabling the *O-N* acyl shift.

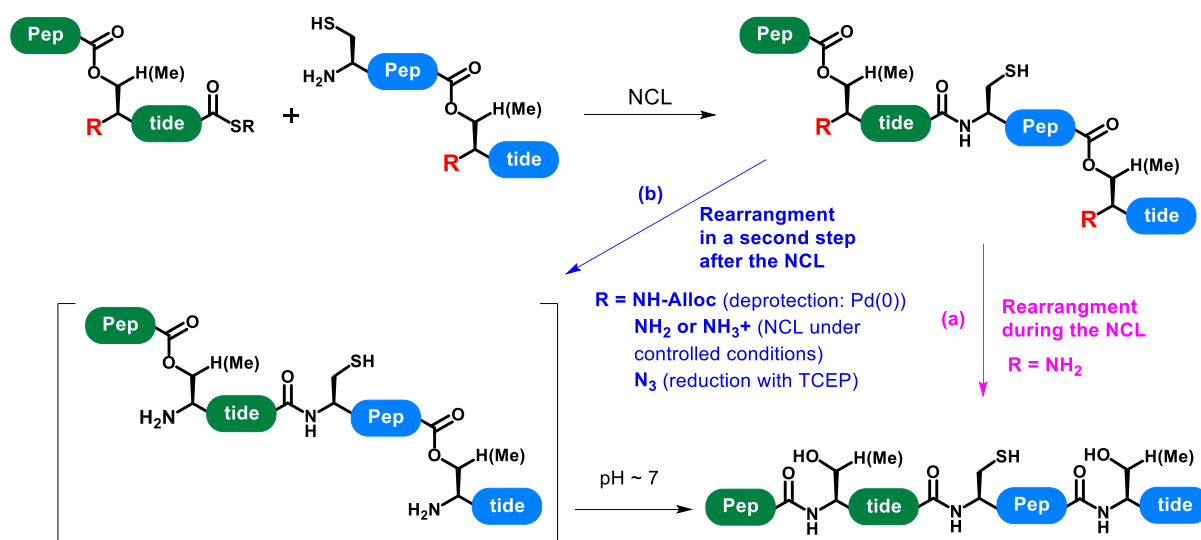


Figure 3: The two approaches used for the use of *O*-acyl isopeptides-containing SH segments in NCL-based chemical protein synthesis

4) Solubilizing tags

Another chemical modification to overcome solubility problem is the introduction of solubilizing tags linked to the segment, either as a permanent modification or through a cleavable linker. These linkers could be either stable to NCL conditions, thus necessitating an extra step after the NCL to remove the tag, or labile under NCL conditions, thus generating the native sequence concomitantly with the ligation reaction. They can be incorporated either at the N- or C-terminus of the segment, linked to a side chain functional group, or as an *N*-substituent of a backbone amide. The solubilizing tags are generally composed of homo-oligomerized amino acids such as lysines or arginines bearing cationic charge side chain and having good water solubility, but single Arg residues, [40] oligoethyleneglycol [41] and tertiary amines [42] have also been successfully used.

A. Permanent solubilizing tags

Pessi and his co-workers reported the addition of three lysing residues at the N-terminus of a hydrophobic segment in order to generate a transmembrane protein segment (fig. 4).

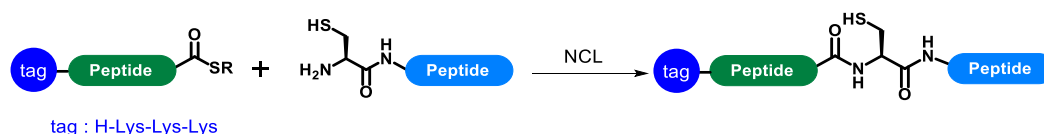


Figure 4: Permanent tag strategy reported by Pessi and his co-workers

Only few illustrations of the use of permanent tags in NCL are reported in the literature: indeed, these tags can affect the structure and function of the synthesized protein which considerably limits their scope.

B. Solubilizing tags cleaved in a second step after NCL

Numerous approaches were developed where the solubilizing tag is removed after the protein assembly.

a) C-terminal solubilizing tags

Installation of a solubilizing tag to the C-terminus of a peptide can be achieved by temporary modification of the C $^{\alpha}$ -carboxylic acid through a cleavable linker (fig. 5). The advantage of this approach is that most of the linkers used are commercially available, and that the tagged segment can be synthesized using conventional SPPS approaches.

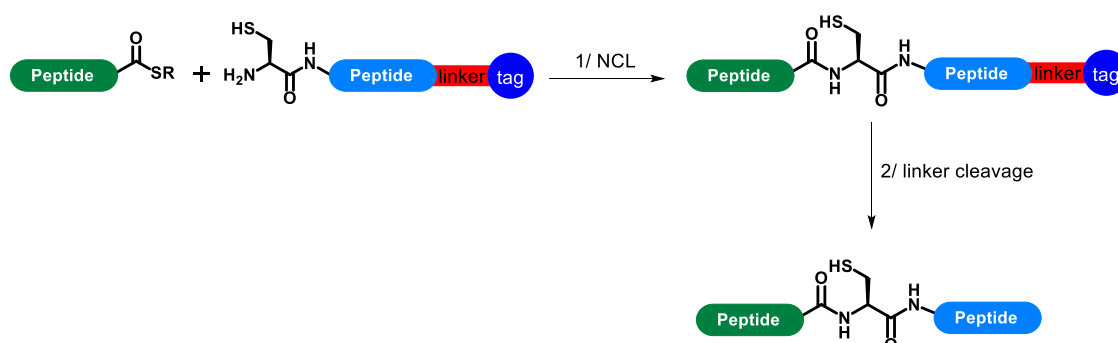


Figure 5: Attachment of solubilizing tag to C-terminus of peptide

Harris and Brimble reported the use of an ester-linked solubilizing tags based on the 4-hydroxy-methylbenzoic acid (HMBA) linker stable under both Boc- and Fmoc-SPPS conditions (fig. 6). [43,44] The tag can be removed by saponification.

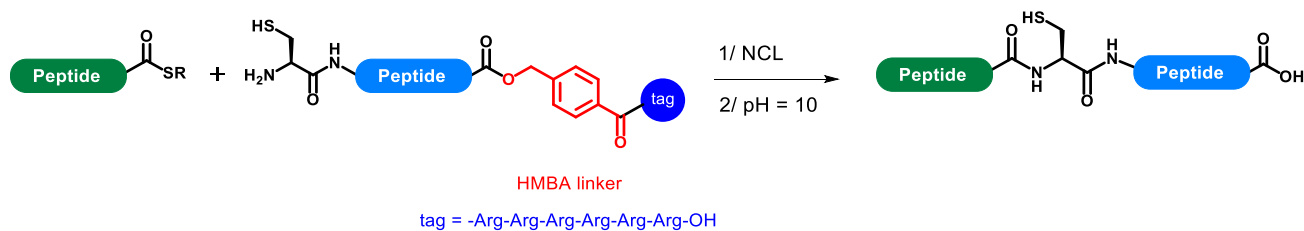


Figure 6: HMBA linker for the introduction of a C-terminal solubilizing tag

Within the same context, Brik [45] reported the attachment of an oligoArg tag to the C-terminus of a hydrophobic segment *via* a 3,4-diaminobenzoic acid linker (Dbz), removed upon treatment with NaNO_2 to generate a benzotriazole amide which undergoes spontaneous hydrolysis (fig. 7).

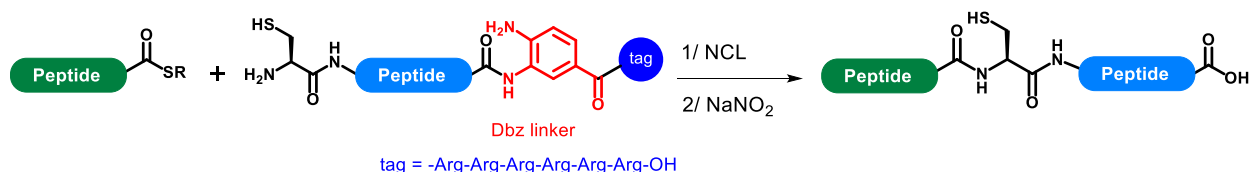


Figure 7: Dbz linker for the introduction of a C-terminal solubilizing tag

Another option is the use of a proteolytic cleavage site between the segment and the hydrophilic tag (fig. 8), a strategy introduced by Kent [46] after he failed to synthesize the HIV-1 protease (HIV-1 PR1) due to the hydrophobicity of the C-terminal cysteinyl peptide segment. In order to overcome this issue, he elegantly took advantage of the auto-proteolytic cleavage during natural HIV-1 protease maturation, by adding 10 C-terminal residues from the reverse transcriptase (RT) protein adjacent to the HIV-1 PR1 segment in the polyprotein precursor, followed by 10 arginine residues as a solubilizing tag. Folding of the protein efficiently led to the removal of the tag and peptide linker through autoproteolysis.

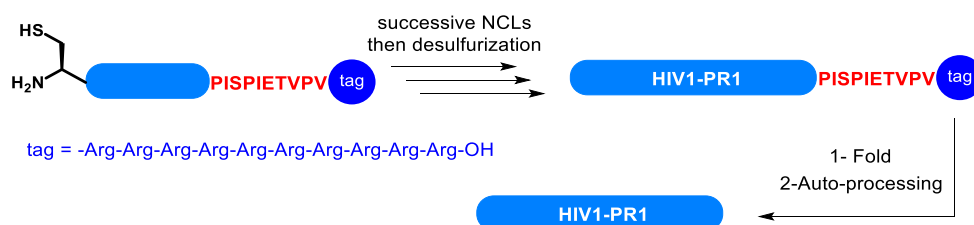


Figure 8: Synthesis of the HIV-1PR by Kent

b) Solubilizing tags incorporated on amino acids side chains

Solubilizing tags can also be inserted in the middle of a segment taking advantage of the several reactive functional groups in the amino acids side chains (fig. 9).

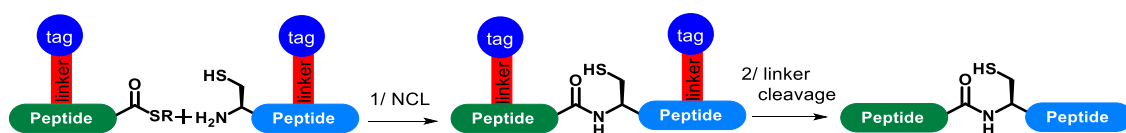
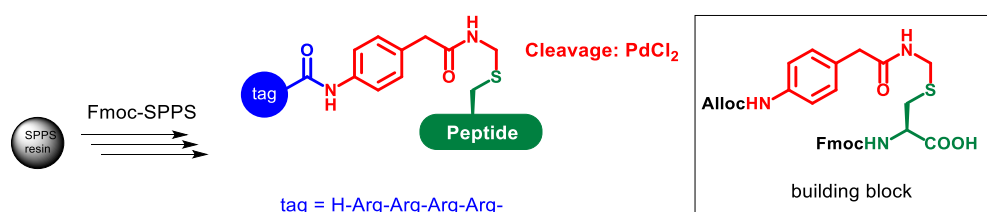


Figure 9: Attachment of solubilizing tag on amino acids side chains

i- Solubilizing tags on cysteine side-chain

Cysteine side chain was the most frequently used to introduce solubilizing tags. As an example, we can cite the solubilizing tag developed by Brik [47] based on a phenylacetamidomethyl (Phacm)-derived linker. In this case, Pd(0)-mediated Alloc deprotection after Fmoc-SPPS elongation of the segment generates a free amine that is used to introduce a tetra-arginine solubilizing sequence through SPPS elongation. This tag was easily removed by treatment with Pd (II) salts after NCL (fig. 10).

Figure 10: The attachment of solubilizing tag *via* Alloc-Phacm Linker and its palladium assisted removal upon protein synthesis

Another Cys side chain solubilizing tag was developed by Danishefsky [40] who used as a linker a derivative of the acetamidomethyl (Acm) group, (fig. 12) often used as a cysteine protecting group for the chemical synthesis and semi synthesis of peptide and proteins. This tag was efficiently removed along with the other Acm protecting group of cysteine residues in the protein using AgOAc.

Yoshiya [48] investigated the introduction of a solubilizing Trt-K₁₀ tag (fig. 11) to synthesize the hepatitis B virus capsid protein known to be very difficult to handle. The tag can be chemoselectively attached to cysteine side chain thiol groups in unprotected peptide segments by simply adding a trityl alcohol-based introducing reagent under acidic conditions that generate the corresponding trityl carbocation. The Trt-K₁₀ solubilizing tag can be removed by a TFA treatment.

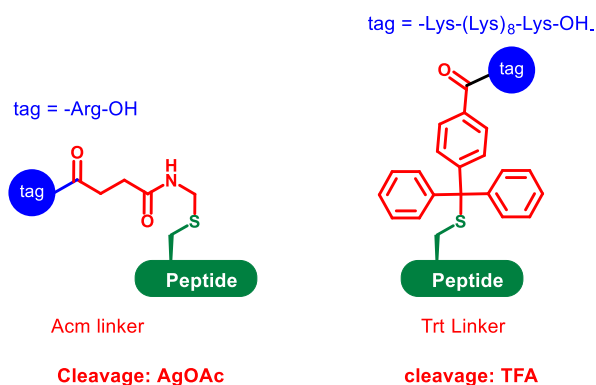


Figure 11: Cysteine side chain based linkage for the introduction of solubilizing tags

ii- Solubilizing tags on glutamine, glutamic acid, asparagine and aspartic acid side-chains

The Trt linker was combined with a (Dbz) linker by Yoshiya and his group to develop a dual linker Lys₆-based solubilizing tag system for Asp/Asn/Glu/Gln-containing peptides (fig. 12). To remove the solubilizing tag, the Dbz moiety is activated using NaNO₂, and the native side chain is generated *via* hydrolysis or ammonolysis. [49]

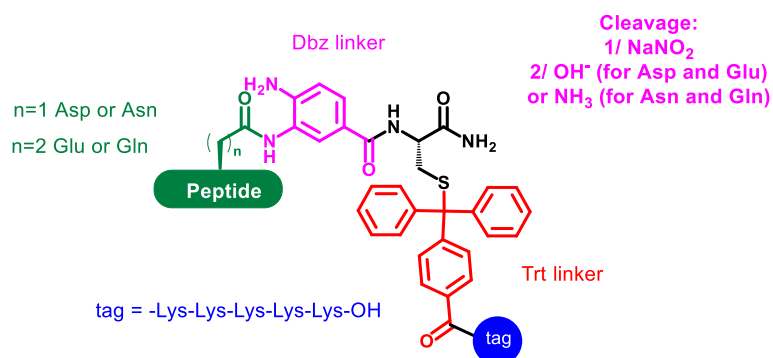


Figure 12: Dual linker tag for Asx/Glx containing peptides

A photocleavable Gln side chain solubilizing tag was introduced by Liu [50] using a nitrobenzene-derived *N*-Alloc-protected building block (fig. 13). Cleavage by UV irradiation at 365 nm cleanly generates the free glutamine side-chain.

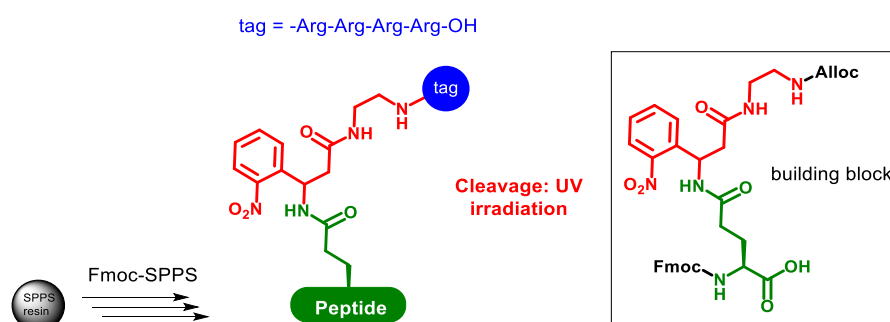


Figure 13: Nitrobenzene Gln side chain linker tag

iii- Solubilizing tags on Lysine side-chains

Lys side chain was also used by Aucagne and Kay [51] to introduce an hexalysine tag through a Ddae (1-(4,4-dimethyl-2,6-dioxocyclo-hexylidene)-3-[2-(2-aminoethoxy)ethoxy]-propyl) linker derived from the Dmab protecting group (fig. 14). [52] The Lys side chain is protected by a Dde group removed after SPPS elongation of the segment by a hydrazine treatment, followed by simple reaction with of a Fmoc-protected enol-based building block then Fmoc-SPPS elongation of the solubilizing peptide sequence. This linker is cleaved by nitrogen-based nucleophiles such as hydrazine or hydroxylamine, but is stable to thiolates present during NCL ligations. Building on this work, Kay [53] recently reported a second generation linker 2-(7-amino-1-hydroxyheptylidene)-5,5-dimethylcyclohexane-1,3-dione (Ddap) (fig. 14), that is more stable in aqueous solvents and easier to handle compared to the Ddae linker.

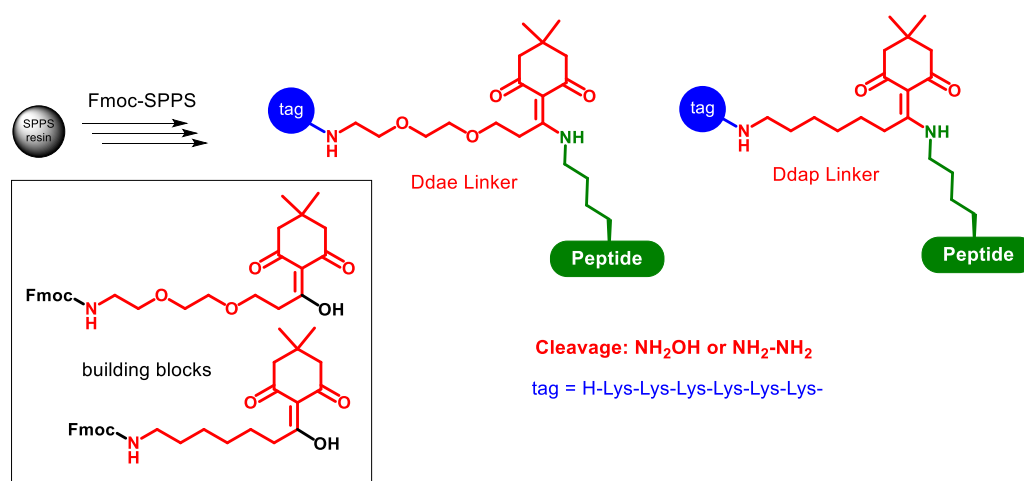


Figure 14: Solubilizing strategies developed by Kay's group

Canaline (Can) is a non-proteinogenic amino acid containing an aminooxy group in its side chain. Yoshiya [54] took advantage of the stability of canalinamides under both neutral NCL conditions and acidic HPLC purification conditions to develop a new linker to introduce a solubilizing tag (fig. 15). The latter is spontaneously removed after NCL and purification under controlled pH conditions (pH 4-5) that promote the attack of the aminooxy group on the amide, resulting in the liberation of the Lys side chain amine. A limitation of this approach is that the Can aminooxy group reacts with NaNO₂, making it unsuitable to prepare a Can-containing peptide thioester by the hydrazide method.

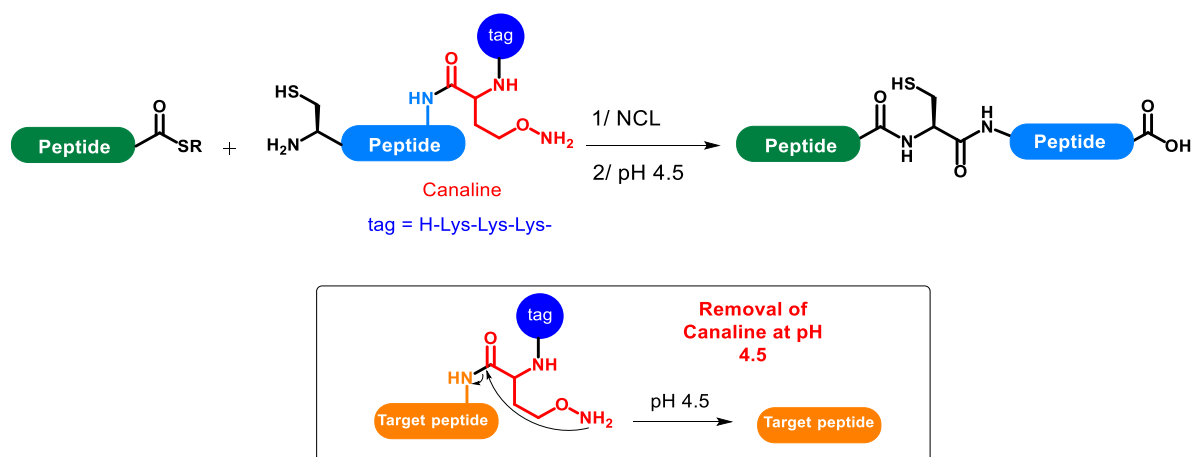


Figure 15: Solubilizing tag strategy using Can-Containing peptides

c) Solubilizing tags incorporated on amides backbone

Solubilizing tags incorporated on amino acids side chains are dependent on a particular amino acid thus limiting their scope. A more general approach for the introduction of these tags in the middle of the segment consists in installing them on amides backbone (fig.16).

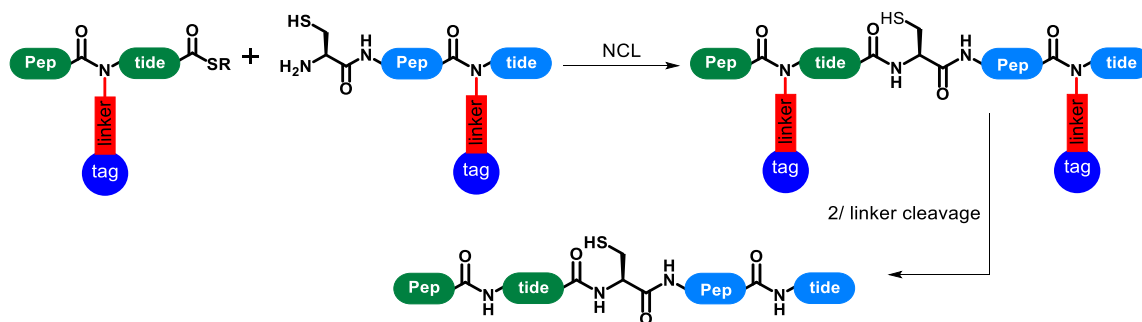


Figure 16: Attachment of solubilizing tag on backbone amides

Lei Liu [55] developed a TFA-labile backbone linker derived from the Hmb group to introduce a solubilizing tag, attached to the α -amino group of Gly residues (fig. 17). Activation of the TFA-lability properties is induced by an ingenious mechanism: a *N*-methyl-*N*-methylaminoethylcarbonyl protection of the phenol group of the linker is removed under the neutral conditions used for the NCL through a spontaneous cyclisation.

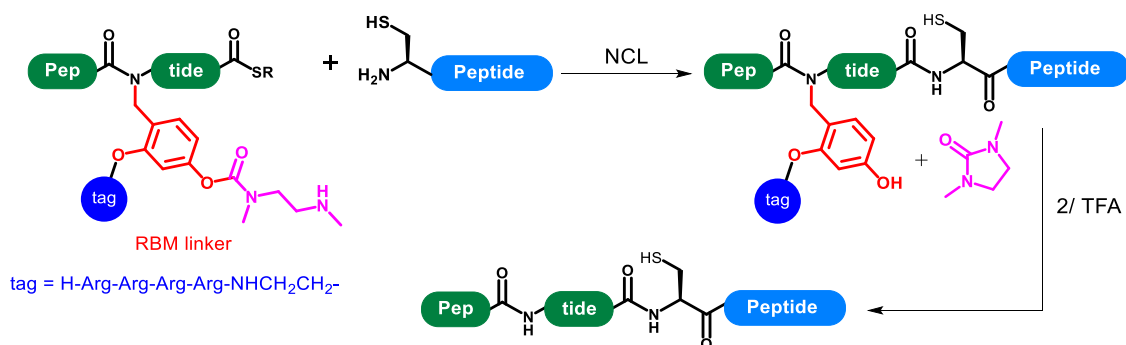
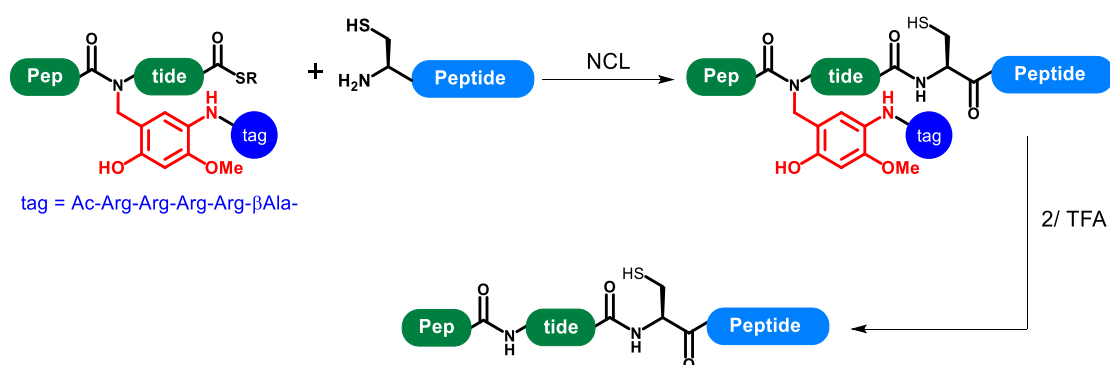


Figure 17: Removable backbone modification (RBM) solubilizing tag developed by Liu

A second generation linker was developed, [56] that is more generally applicable because the tag can be introduced on any amino acid and is no longer limited to Gly residues (fig. 18).

Figure 18: 2nd generation removable backbone modification reported by Liu

Although being very sufficient to solubilize hydrophobic segments and to inhibit aggregation, these solubilizing tags have several significant disadvantages. The linker generally needs to be prepared through multistep synthesis in solution, adds tedious and low-yield steps for the general synthesis and, most importantly, necessitates an additional cleavage step after the NCL, sometimes employing harsh conditions [57] which limits the application of the solubilizing tags.

C. Solubilizing tags cleaved during NCL

An advantageous alternative is the introduction of a solubilizing tag that can be removed *in situ* during the ligation process without the need of any additional cleavage step. This strategy is well suited when the segment has limited solubility in HPLC buffers, precluding its characterization and purification, but is soluble under NCL

conditions. This is often the case thanks to guanidinium chloride and other additives that are typically used for NCL reactions.

a) NCL-labile solubilizing tags for C-terminal thioester segments

Introducing a temporary solubilizing tag at the C-terminus of the peptide *via* a thioester linkage (fig. 19) is the most common approach to generate NCL-labile solubilizing tags.

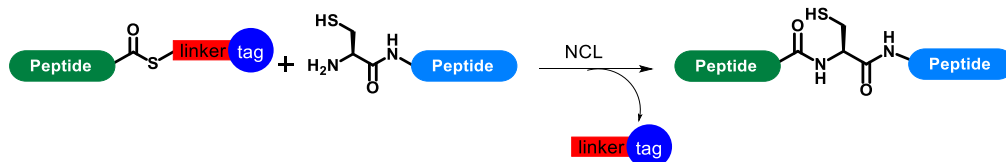


Figure 19: C-terminal thioester NCL-labile solubilizing tags

Such C-terminal tags are generally introduced by Boc-SPPS using a MPAL linker for the incorporation of the thioester moiety, installed after the SPPS elongation of a solubilizing peptide sequence (fig. 20). Many successful examples for this method were reported. For instance, Aimoto [22] used a penta-Arg tag for the NCL-based synthesis of the C-terminal region of the opioid receptor like 1, ORL1 (288-370). Kent used a similar oligoArg tag strategy for the synthesis of a 121 aa integral membrane protein, [58] IGF-1 [59] and ribonuclease A. [60] Lu and his-coworkers also used an identical strategy to synthesize a 100 residues protein. [61] We can also cite the synthesis of a masked analogue of the fish antifreeze potentiating protein (AFPP) by Brimble using an oligoLys tag. [62]

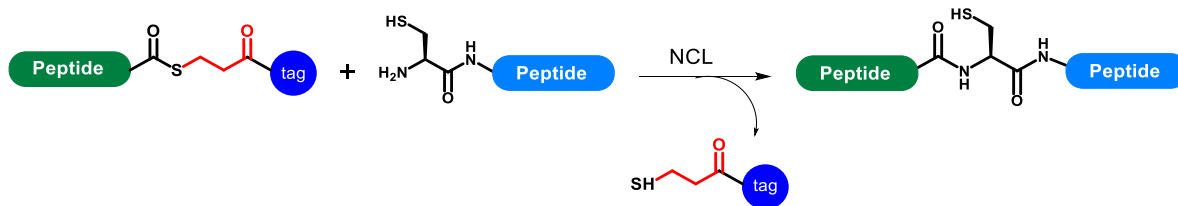


Figure 20: Introduction of solubilizing tag on the thioester leaving group

However, the reported methods require Boc SPPS to introduce the linker or to generate the tagged thioester segment, which means using hazardous HF cleavage. A useful alternative is the 2-hydroxy-3-mercaptopropionic acid (Hmp) linker which is compatible with Fmoc-based SPPS (see chapter 3). This *O*-*S*-shift crypto-thioesters-based approach was introduced by Tietze and her team, [41] but suffers from the limitations of the Hmp linker: up to 25% of hydrolysis was observed during NCL (fig. 21).

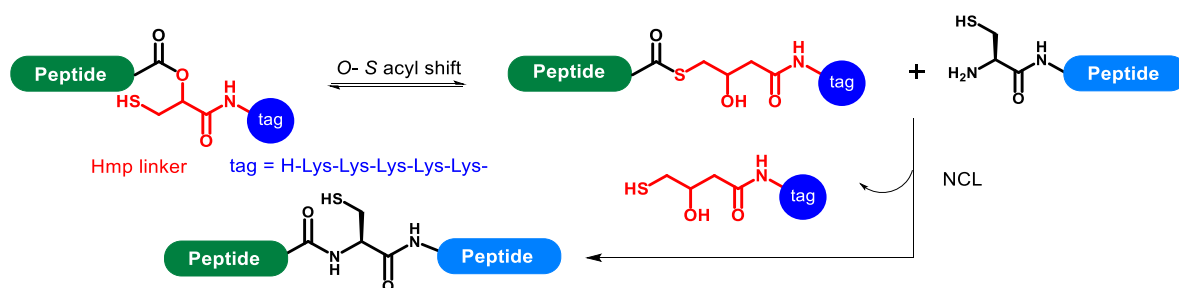


Figure 21: The use of ADO-K₅ tag for the NCL based synthesis a membrane protein reported by Tietze and her team

b) Towards a NCL-labile solubilizing tags for N-terminal cysteinyl segments

In the only example found in the literature, Valiyaveetil developed a solubilizing tag that can be introduced at the N-terminus of a hydrophobic peptide segment by forming an intermolecular disulfide bond with the N-terminal Cys side chain (fig. 22). [63] This tag was successfully used to synthesize NaK, a bacterial nonselective cation channel, through expressed protein ligation. However, the strategy used by the authors is associated with severe limitations. The solubilizing tag is introduced in solution on the hydrophobic cysteinyl segment, which thus needs to be soluble for this purpose, and the method used for its introduction relies on the air-mediated formation of an intermolecular disulfide with a small hydrophilic cysteine-containing peptide (H-Arg-Arg-Arg-Cys-NH₂), resulting in very low yield introduction of the tag due to non-controlled formation of homo- and heterodisulfides. This incomplete leads in turn to very low purification yields for the segment (~ 2%). Moreover, in their work, the authors did not demonstrate the possibility to use directly the tagged segment in the NCL reaction, and preferred to remove the tag through reduction in a first step.

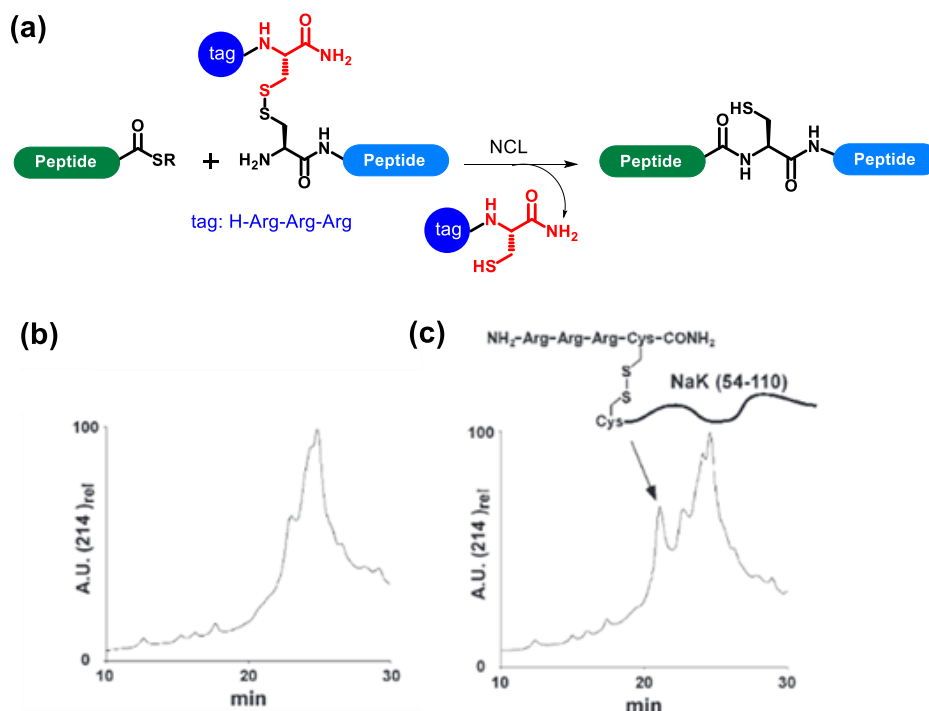


Figure 22: a) strategy developed by Valiyaveetil for the synthesis of NaK; b) HPLC trace of the crude cysteinyl segment without the tag; c) crude cysteinyl segment after introduction of the tag (figure b and c adapted from ref. [63])

II. Aim of this work

We thought that developing a straightforward and high-yielding method for the introduction of such a disulfide-linked solubilizing tag on N-terminal cysteinyl segments, and extending the concept to the concomitant NCL/tag cleavage could be extremely valuable and generally applicable in chemical protein synthesis.

This new method should ideally meet the following requirements:

- It must be compatible with Fmoc solid phase peptide synthesis
- It should be easily introduced on the peptide segment without requiring non-commercial reagents or building blocks, or additional solution phase steps
- Its removal should be traceless and fast under NCL conditions.

The aim of this work of my PhD work was the development and optimization of such a method. The strategy we developed makes use of a key solid-supported disulfide formation to introduce a 2-amino-2,2-dimethylsulfanyl (Ades) moiety that can be further used as an NCL labile handle to introduce a solubilizing tag (here an hexalysine) through further standard Fmoc-SPPS (fig. 23).

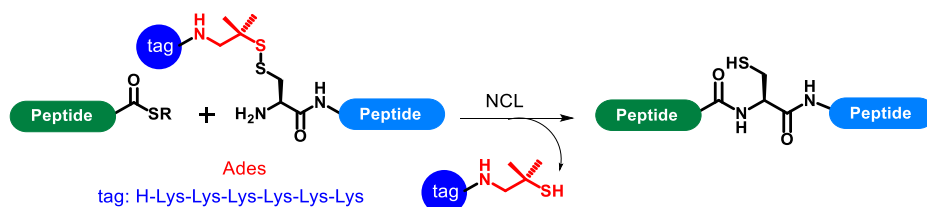


Figure 23: New solubilizing tag developed in this work

This tag overcomes major handling and purification problems encountered with hydrophobic peptides and is highly complementary to reminiscent methodologies aimed at solubilizing the C-terminal thioester segments. We exemplified the broad utility of this method through the synthesis, purification and ligation of a model polypeptide protein (MUC1) and also of a well-known difficult small protein target, SUMO-2. The two targets were obtained in high yields and purities.

III. References for the bibliographic study part

1. Fink, A.L. (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Folding and Design*, **3** (1), R9–R23.
2. Jayalekshmy, P., and Mazur, S. (1976) Pseudodilution, the solid-phase immobilization of benzyne. *J. Am. Chem. Soc.*, **98** (21), 6710–6711.
3. Kent, S.B.H. (1988) Chemical Synthesis of Peptides and Proteins. *Annu. Rev. Biochem.*, **57** (1), 957–989.
4. Schnölzer, M., Alewood, P., Jones, A., Alewood, D., and Kent, S.B.H. (2007) In Situ Neutralization in Boc-chemistry Solid Phase Peptide Synthesis: Rapid, High Yield Assembly of Difficult Sequences. *Int J Pept Res Ther*, **13** (1–2), 31–44.
5. Mutter, M., Nefzi, A., Sato, T., Sun, X., Wahl, F., and Wöhr, T. (1995) Pseudo-prolines (psi Pro) for accessing ‘inaccessible’ peptides. *Pept Res*, **8** (3), 145–153.
6. Simmonds, R.G. (2009) Use of the Hmb backbone-protecting group in the synthesis of difficult sequences. (*Int. J. Pept. Protein Res.*, **47** (1–2), 36–41.
7. Zhao, D.-D., Fan, X.-W., Hao, H., Zhang, H.-L., and Guo, Y. (2019) Temporary Solubilizing Tags Method for the Chemical Synthesis of Hydrophobic Proteins. *COC*, **23** (1), 2–13.
8. Mueller, L.K., Baumruck, A.C., Zhdanova, H., and Tietze, A.A. (2020) Challenges and Perspectives in Chemical Synthesis of Highly Hydrophobic Peptides. *Front. Bioeng. Biotechnol.*, **8**, 162.
9. Palmer, D.C. (2001) Guanidine, in *Encyclopedia of Reagents for Organic Synthesis* (eds. John Wiley & Sons, Ltd), John Wiley & Sons, Ltd, Chichester, UK, pp. rg013.
10. Kwon, B., Tietze, D., White, P.B., Liao, S.Y., and Hong, M. (2015) Chemical ligation of the influenza M2 protein for solid-state NMR characterization of the cytoplasmic domain: Chemical Ligation of Influenza M2 for SSNMR. *Protein Science*, **24** (7), 1087–1099.
11. Lahiri, S., Brehs, M., Olschewski, D., and Becker, C.F.W. (2011) Total Chemical Synthesis of an Integral Membrane Enzyme: Diacylglycerol Kinase from *Escherichia coli*. *Angew. Chem. Int. Ed.*, **50** (17), 3988–3992.
12. Stark, G.R. (1964) On the Reversible Reaction of Cyanate with Sulfhydryl Groups and the Determination of NH₂-Terminal Cysteine and Cystine in Proteins. *J. Biol. Chem.*, **239**, 1411–1414.

13. Dittmann, M., Sauermann, J., Seidel, R., Zimmermann, W., and Engelhard, M. (2010) Native chemical ligation of hydrophobic peptides in organic solvents. *J. Pept. Sci.*, **16** (10), 558–562.
14. Dittmann, M., Sadek, M., Seidel, R., and Engelhard, M. (2012) Native chemical ligation in dimethylformamide can be performed chemoselectively without racemization: NCL in organic solvents. *J. Pept. Sci.*, **18** (5), 312–316.
15. Kochendoerfer, G.G., Salom, D., Lear, J.D., Wilk-Orescan, R., Kent, S.B.H., and DeGrado, W.F. (1999) Total Chemical Synthesis of the Integral Membrane Protein Influenza A Virus M2: Role of Its C-Terminal Domain in Tetramer Assembly [†]. *Biochemistry*, **38** (37), 11905–11913.
16. Shen, F., Tang, S., and Liu, L. (2011) Hexafluoro-2-propanol as a potent cosolvent for chemical ligation of membrane proteins. *Sci. China Chem.*, **54** (1), 110–116.
17. Zuo, C., Tang, S., and Zheng, J.-S. (2015) Chemical synthesis and biophysical applications of membrane proteins: Chemical synthesis of membrane proteins. *J. Pept. Sci.*, **21** (7), 540–549.
18. Bianchi, E., Ingenito, R., Simon, R.J., and Pessi, A. (1999) Engineering and Chemical Synthesis of a Transmembrane Protein: The HCV Protease Cofactor Protein NS4A. *J. Am. Chem. Soc.*, **121** (33), 7698–7699.
19. Bouchenna, J., Sénéchal, M., Drobecq, H., Vicogne, J., and Melnyk, O. (2019) Total Chemical Synthesis of All SUMO-2/3 Dimer Combinations. *Bioconjugate Chem.*, **30** (11), 2967–2973.
20. Chu, N.K., and Becker, C.F.W. (2009) Chapter 9 Semisynthesis of Membrane-Attached Prion Proteins, in *Methods in Enzymology*, vol. 462, Elsevier, pp. 177–193.
21. Valiyaveetil, F.I., MacKinnon, R., and Muir, T.W. (2002) Semisynthesis and Folding of the Potassium Channel KcsA. *J. Am. Chem. Soc.*, **124** (31), 9113–9120.
22. Sato, T., Saito, Y., and Aimoto, S. (2005) Synthesis of the C-terminal region of opioid receptor like 1 in an SDS micelle by the native chemical ligation: effect of thiol additive and SDS concentration on ligation efficiency. *J. Peptide Sci.*, **11** (7), 410–416.
23. Loo, R.R.O., Dales, N., and Andrews, P.C. (1994) Surfactant effects on protein structure examined by electrospray ionization mass spectrometry. *Protein Sci.*, **3** (11), 1975–1983.

24. Otaka, A., Ueda, S., Tomita, K., Yano, Y., Tamamura, H., Matsuzaki, K., and Fujii, N. (2004) Facile synthesis of membrane-embedded peptides utilizing lipid bilayer-assisted chemical ligation. *Chem. Commun.*, (15), 1722.
25. Ingale, S., Buskas, T., and Boons, G.-J. (2006) Synthesis of Glyco(lipo)peptides by Liposome-Mediated Native Chemical Ligation. *Org. Lett.*, **8** (25), 5785–5788.
26. Hunter, C.L., and Kochendoerfer, G.G. (2004) Native Chemical Ligation of Hydrophobic Peptides in Lipid Bilayer Systems. *Bioconjugate Chem.*, **15** (3), 437–440.
27. Zhang, B., Li, Y., Shi, W., Wang, T., Zhang, F., and Liu, L. (2020) Chemical Synthesis of Proteins Containing 300 Amino Acids. *Chem. Res. Chin. Univ.*, **36** (5), 733–747.
28. Md. Abdur Rauf, S., Arvidsson, P.I., Albericio, F., Govender, T., Maguire, G.E.M., Kruger, H.G., and Honarparvar, B. (2015) The effect of N-methylation of amino acids (Ac-X-OMe) on solubility and conformation: a DFT study. *Org. Biomol. Chem.*, **13** (39), 9993–10006.
29. Boehringer, R., Kieffer, B., and Torbeev, V. (2018) Total chemical synthesis and biophysical properties of a designed soluble 24 kDa amyloid analogue. *Chem. Sci.*, **9** (25), 5594–5599.
30. Johnson, E.C.B., and Kent, S.B.H. (2006) Studies on the Insolubility of a Transmembrane Peptide from Signal Peptide Peptidase. *J. Am. Chem. Soc.*, **128** (22), 7140–7141.
31. Johnson, T., Quibell, M., Owen, D., and Sheppard, R.C. (1993) A reversible protecting group for the amide bond in peptides. Use in the synthesis of ‘difficult sequences’. *J. Chem. Soc., Chem. Commun.*, (4), 369–372.
32. Levinson, A.M., McGee, J.H., Roberts, A.G., Creech, G.S., Wang, T., Peterson, M.T., Hendrickson, R.C., Verdine, G.L., and Danishefsky, S.J. (2017) Total Chemical Synthesis and Folding of All- L and All- D Variants of Oncogenic KRas(G12V). *J. Am. Chem. Soc.*, **139** (22), 7632–7639.
33. Sohma, Y., Sasaki, M., Hayashi, Y., Kimura, T., and Kiso, Y. (2004) Novel and efficient synthesis of difficult sequence-containing peptides through O–N intramolecular acyl migration reaction of O-acyl isopeptides. *Chem. Commun.*, (1), 124–125.
34. Wang, Z., Xu, W., Liu, L., and Zhu, T.F. (2016) A synthetic molecular system capable of mirror-image genetic replication and transcription. *Nature Chem*, **8** (7), 698–704.
35. Xu, W., Jiang, W., Wang, J., Yu, L., Chen, J., Liu, X., Liu, L., and Zhu, T.F. (2017) Total chemical synthesis of a thermostable enzyme capable of polymerase chain reaction. *Cell Discov*, **3** (1), 17008.

36. Sohma, Y., Kitamura, H., Kawashima, H., Hojo, H., Yamashita, M., Akaji, K., and Kiso, Y. (2011) Synthesis of an O-acyl isopeptide by using native chemical ligation to efficiently construct a hydrophobic polypeptide. *Tetrahedron Letters*, **52** (52), 7146–7148.
37. Desmet, R., Pauzuolis, M., Boll, E., Drobecq, H., Raibaut, L., and Melnyk, O. (2015) Synthesis of Unprotected Linear or Cyclic O -Acyl Isopeptides in Water Using Bis(2-sulfanylethyl)amido Peptide Ligation. *Org. Lett.*, **17** (13), 3354–3357.
38. Kawashima, H., Kuruma, T., Yamashita, M., Sohma, Y., and Akaji, K. (2014) Synthesis of an O -acyl isopeptide by using native chemical ligation in an aqueous solvent system. *J. Pept. Sci.*, **20** (5), 361–365.
39. Hojo, H., Katayama, H., Tano, C., Nakahara, Y., Yoneshige, A., Matsuda, J., Sohma, Y., Kiso, Y., and Nakahara, Y. (2011) Synthesis of the sphingolipid activator protein, saposin C, using an azido-protected O-acyl isopeptide as an aggregation-disrupting element. *Tetrahedron Letters*, **52** (5), 635–639.
40. Brailsford, J.A., Stockdill, J.L., Axelrod, A.J., Peterson, M.T., Vadola, P.A., Johnston, E.V., and Danishefsky, S.J. (2018) Total chemical synthesis of human thyroid-stimulating hormone (hTSH) β -subunit: Application of arginine-tagged acetamidomethyl (AcmR) protecting groups. *Tetrahedron*, **74** (15), 1951–1956.
41. Baumruck, A.C., Tietze, D., Steinacker, L.K., and Tietze, A.A. (2018) Chemical synthesis of membrane proteins: a model study on the influenza virus B proton channel. *Chem. Sci.*, **9** (8), 2365–2375.
42. Eid, E., Boross, G.N., Sun, H., Msallam, M., Singh, S.K., and Brik, A. (2020) Total Chemical Synthesis of ISGylated-Ubiquitin Hybrid Chain Assisted by Acetamidomethyl Derivatives with Dual Functions. *Bioconjugate Chem.*, **31** (3), 889–894.
43. Harris, P., and Brimble, M. (2009) Synthesis of an Arginine Tagged [Cys155-Arg180] Fragment of NY-ESO-1: Elimination of an Undesired By-product Using 'In House' Resins. *Synthesis*, **2009** (20), 3460–3466.
44. Harris, P.W.R., and Brimble, M.A. (2010) Toward the total chemical synthesis of the cancer protein NY-ESO-1. *Biopolymers*, **94** (4), 542–550.
45. Bondalapati, S., Eid, E., Mali, S.M., Wolberger, C., and Brik, A. (2017) Total chemical synthesis of SUMO-2-Lys63-linked diubiquitin hybrid chains assisted by removable solubilizing tags. *Chem. Sci.*, **8** (5), 4027–4034.

46. Johnson, E.C.B., Malito, E., Shen, Y., Rich, D., Tang, W.-J., and Kent, S.B.H. (2007) Modular Total Chemical Synthesis of a Human Immunodeficiency Virus Type 1 Protease. *J. Am. Chem. Soc.*, **129** (37), 11480–11490.
47. Maity, S.K., Mann, G., Jbara, M., Laps, S., Kamnesky, G., and Brik, A. (2016) Palladium-Assisted Removal of a Solubilizing Tag from a Cys Side Chain To Facilitate Peptide and Protein Synthesis. *Org. Lett.*, **18** (12), 3026–3029.
48. Tsuda, S., Mochizuki, M., Ishiba, H., Yoshizawa-Kumagaye, K., Nishio, H., Oishi, S., and Yoshiya, T. (2018) Easy-to-Attach/Detach Solubilizing-Tag-Aided Chemical Synthesis of an Aggregative Capsid Protein. *Angew. Chem. Int. Ed.*, **57** (8), 2105–2109.
49. Tsuda, S., Masuda, S., and Yoshiya, T. (2019) Solubilizing Trityl-Type Tag To Synthesize Asx/Glx-Containing Peptides. *ChemBioChem*, **20** (16), 2063–2069.
50. Huang, Y.-C., Li, Y.-M., Chen, Y., Pan, M., Li, Y.-T., Yu, L., Guo, Q.-X., and Liu, L. (2013) Synthesis of Autophagosomal Marker Protein LC3-II under Detergent-Free Conditions. *Angew. Chem. Int. Ed.*, **52** (18), 4858–4862.
51. Jacobsen, M.T., Petersen, M.E., Ye, X., Galibert, M., Lorimer, G.H., Aucagne, V., and Kay, M.S. (2016) A Helping Hand to Overcome Solubility Challenges in Chemical Protein Synthesis. *J. Am. Chem. Soc.*, **138** (36), 11775–11782.
52. Bycroft, B.W., Chan, W.C., Chhabra, S.R., and Hone, N.D. (1993) A novel lysine-protecting procedure for continuous flow solid phase synthesis of branched peptides. *J. Chem. Soc., Chem. Commun.*, (9), 778.
53. Fulcher, J., Petersen, M., giesler, riley, Cruz, Z., eckert, debra, Francis, J.N., kawamoto, eric, Jacobsen, M., and Kay, M. (2019) Chemical Synthesis of Shiga Toxin Subunit B Using a Next-Generation Traceless “Helping Hand” Solubilizing Tag.
54. Tsuda, S., Nishio, H., and Yoshiya, T. (2018) Peptide self-cleavage at a canaline residue: application to a solubilizing tag system for native chemical ligation. *Chem. Commun.*, **54** (64), 8861–8864.
55. Zheng, J.-S., Yu, M., Qi, Y.-K., Tang, S., Shen, F., Wang, Z.-P., Xiao, L., Zhang, L., Tian, C.-L., and Liu, L. (2014) Expedient Total Synthesis of Small to Medium-Sized Membrane Proteins via Fmoc Chemistry. *J. Am. Chem. Soc.*, **136** (9), 3695–3704.
56. Zheng, J.-S., He, Y., Zuo, C., Cai, X.-Y., Tang, S., Wang, Z.A., Zhang, L.-H., Tian, C.-L., and Liu, L. (2016) Robust Chemical Synthesis of Membrane Proteins through a General Method of Removable Backbone Modification. *J. Am. Chem. Soc.*, **138** (10), 3553–3561.

57. Zuo, C., Tang, S., Si, Y.-Y., Wang, Z.A., Tian, C.-L., and Zheng, J.-S. (2016) Efficient synthesis of longer A β peptides via removable backbone modification. *Org. Biomol. Chem.*, **14** (22), 5012–5018.
58. Johnson, E.C.B., and Kent, S.B.H. (2007) Towards the total chemical synthesis of integral membrane proteins: a general method for the synthesis of hydrophobic peptide- α thioester building blocks. *Tetrahedron Lett.*, **48** (10), 1795–1799.
59. Sohma, Y., Pentelute, B.L., Whittaker, J., Hua, Q., Whittaker, L.J., Weiss, M.A., and Kent, S.B.H. (2008) Comparative Properties of Insulin-like Growth Factor 1 (IGF-1) and [Gly7D-Ala]IGF-1 Prepared by Total Chemical Synthesis. *Angew. Chem. Int. Ed.*, **47** (6), 1102–1106.
60. Boerema, D.J., Tereshko, V.A., and Kent, S.B.H. (2008) Total synthesis by modern chemical ligation methods and high resolution (1.1 Å) X-ray structure of ribonuclease A. *Biopolymers*, **90** (3), 278–286.
61. Zhan, C., Zhao, L., Chen, X., Lu, W.-Y., and Lu, W. (2013) Total chemical synthesis of dengue 2 virus capsid protein via native chemical ligation: Role of the conserved salt-bridge. *Bioorganic & Medicinal Chemistry*, **21** (12), 3443–3449.
62. Yang, S.-H., Wojnar, J.M., Harris, P.W.R., DeVries, A.L., Evans, C.W., and Brimble, M.A. (2013) Chemical synthesis of a masked analogue of the fish antifreeze potentiating protein (AFPP). *Org. Biomol. Chem.*, **11** (30), 4935.
63. Linn, K.M., Derebe, M.G., Jiang, Y., and Valiyaveetil, F.I. (2010) Semisynthesis of NaK, a Na⁺ and K⁺ Conducting Ion Channel. *Biochemistry*, **49** (21), 4450–4456.

IV. Publication 4 (Published in Chemical Science)

These results have been submitted (1 Nov 2020) and accepted (10 Jan 2021) for publication in Chemical Science. In this thesis manuscript, we have chosen to reproduce the original paper.

Abboud S.A., Cisse E., Doudeau M., Bénédicti H., Aucagne V. (2021) A straightforward methodology to overcome solubility challenges for N-terminal cysteinyl peptide segments used in native chemical ligation. *Chem. Sci.*, 2021, 12, 3194-3201. DOI: 10.1039/D0SC06001A

EDGE ARTICLE

A large variety of linkers used to attach the tag to the peptide segment and designed to be cleaved in an additional step after the NCL (scheme 1A) have been reported. Cleavage conditions include treatments with acids,^{11,14c-d} bases,^{12b} pH 4.5 buffer,^{17c} sodium nitrite^{12c,13} and nucleophiles,^{17b,17d} transition metal catalysis,^{12c,14a-b,15,17a} UV irradiation¹⁶ and autoprotoleolysis.^{12a}

An advantageous alternative is the introduction of the solubilizing tag on the C-terminal thioester moiety (scheme 1B).^{12b,18-20} In this case, the tag is cleaved in the course of the NCL reaction without needing an additional synthetic step, after playing its solubilizing role during the purification, characterization and handling of the problematic segment. This approach was pioneered by Aimoto¹⁸ and Kent¹⁹ for Boc-SPPS-based thioester synthesis, and later extended by Tietze²⁰ to Fmoc-SPPS using a β -mercaptoester precursor converted *in situ* into a thioester during NCL through an $O \rightarrow S$ acyl shift²¹ (a so-called peptide crypto-thioester²²). If this one-pot NCL/tag cleavage approach is clearly not suited to extreme situations where the ligation product remains insoluble or prone to aggregation, it is expected to be widely applicable in the many cases when the low-solubility/aggregation behavior is associated with a single isolated segment of the protein. Indeed, the additional segment ligated to the problematic one can further play the role of solubility tag, strikingly demonstrated by the synthesis of fragments of transmembrane proteins using these techniques.^{18,20} However, the strategy is inherently limited to thioester segments and not suited for N-terminal cysteinyl counterparts.²³

In a closely related context, Valiyaveetil²⁴ proposed the idea to introduce a solubility tag linked through a disulfide to the N-terminal Cys of a cysteinyl segment synthesized by Boc-SPPS, the disulfide bond being cleaved in a first additional reduction step prior to NCL. In this case, the tag (Arg-Arg-Arg-Cys-NH₂) was introduced in solution through air oxidation-mediated formation of a mixed disulfide with the crude segment. This resulted in a very low yield in the tagged segment due to both insolubility of the non-tagged segment and the non-directed formation of the disulfide leading to complex mixtures. Nevertheless, the authors succeeded in the highly demanding semi-synthesis of ion channels through expressed protein ligation²⁵ (EPL) with a recombinant thioester, demonstrating the feasibility of the approach.

We thought that devising a straightforward methodology for the introduction of such a disulfide-linked solubilizing tag on N-terminal cysteinyl segments, and extending the concept of concomitant NCL/tag cleavage could be extremely valuable and generally applicable in chemical protein synthesis (scheme 1C). Indeed, disulfides are readily cleaved under standard NCL conditions generally including reducing agents (e. g. TCEP) and a large excess of aromatic thiols like 4-mercaptophenylacetic acid (MPAA).²⁶

We herein describe a straightforward methodology for the introduction of such a tag through Fmoc-SPPS, which can be easily automated on a standard peptide synthesizer, and

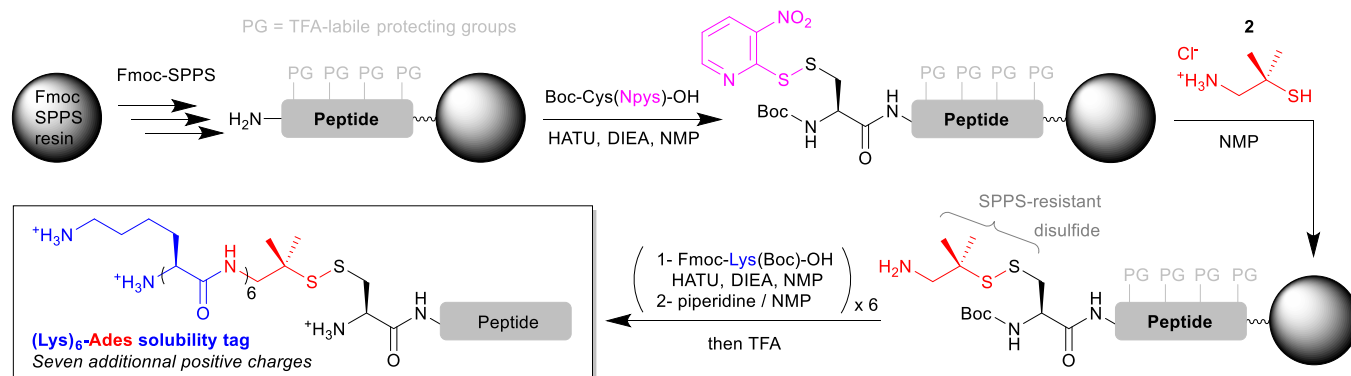
exemplified the utility of this method through the synthesis of SUMO-2, a previously reported difficult target.^{12c}

Results and discussion

The disulfide-based linker bridging the N-terminal cysteinyl segment and the solubilizing tag is the cornerstone of the strategy. Ideally, this linker should be (1) introduced through an automation-friendly solid phase step using commercially available materials, (2) bear a primary aliphatic amine or a suitable precursor for further Fmoc-SPPS elongation of a hydrophilic peptide sequence and (3) be stable to the Fmoc-SPPS conditions, including elongation and TFA-based cleavage. Considering these requirements, we reasoned that incorporation of the N-terminal cysteinyl residue through the coupling of Boc-Cys(Npys)-OH would be ideal. The Npys group²⁷ (S-3-nitro-2-pyridinesulfonyl) is classically used for the directed formation of a mixed disulfide on an N-terminal cysteine either on solid support²⁸ or in solution, through reaction with a thiol. The simplest solution for automated synthesis would be a reaction on solid support with an amino-thiol. We selected two commercially available candidates: cysteamine (2-amino-ethane-1-thiol, **1**), and 2-amino-1,1-dimethyl-ethane-1-thiol (**2**). Preliminary experiments with a model tripeptide (see ESI p S5-S11) showed that disulfide formation was quantitative through simple incubation in NMP for 1 h with an excess (10 equiv.) of **2** as its hydrochloride salt. We also demonstrated that the disulfide was stable to a long piperidine treatment mimicking the repeated Fmoc deprotection conditions needed for the SPPS elongation of a hydrophilic peptide tag. Contrastingly, reaction with **1** was much less clean, and the resulting disulfide was not stable to the piperidine treatment. These results are in accordance with the known higher stability of tertiary thiol disulfides derivatives of cysteine like S-StBu towards Fmoc-SPPS conditions as compared to simple non-hindered primary thiol disulfides.²⁹

Having in hand a robust method for the solid phase introduction of the linker (referred hereafter as Ades, 2-amino-1,1-dimethylethyl-1-sulfanyl), we applied it to two different long peptides, followed by the automated introduction of a (Lys)₆ hydrophilic tag through repeated couplings of Fmoc-Lys(Boc)-OH under standard conditions (scheme 2). We started with a 41 amino acids (aa) model sequence devoid of any solubility problem, derived from the human mucin MUC1³⁰ variable number tandem repeat (VNTR) region made of a duplicated 20 aa sequence (**7**, scheme 3). Gratifyingly, **7** was obtained in excellent yield and purity without needing of any further optimization of the synthetic protocol.

To test the concomitant NCL/tag cleavage, we synthesized the 41 aa crypto-thioester segment counterpart **3a**, also derived from MUC1. This segment is equipped with an *N*-(2-hydroxy-4-nitrobenzyl)cysteine-based device (*N*-Hnb-Cys) capable of forming thioester *in situ* under NCL conditions (scheme 3).³¹ *N*-Hnb-Cys crypto-thioesters are straightforward to synthesize.



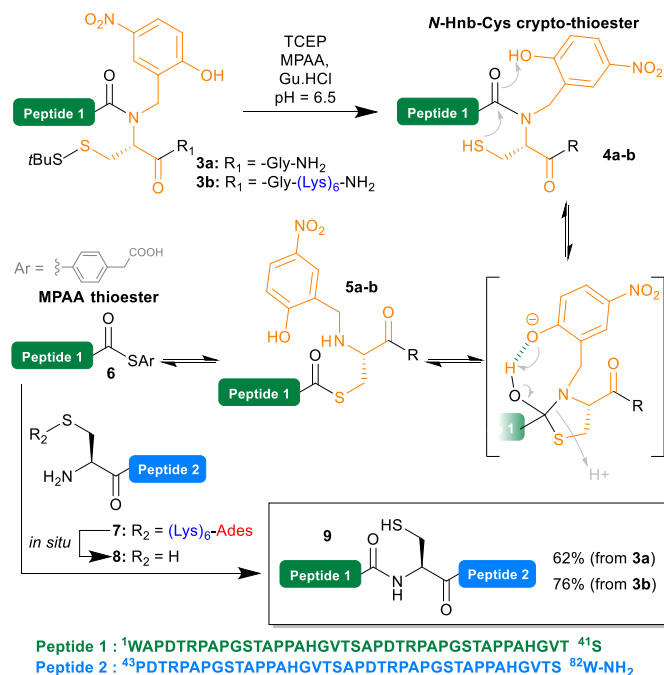
Scheme 2: General synthetic strategies for the introduction of (Lys)₆-Ades solubility tags on N-terminal cysteinyl segments using automated Fmoc-SPPS.

through automated Fmoc-SPPS and show fast ligation kinetics owing to internal catalysis by a judiciously placed phenol group (scheme 3).^{31a}

To our delight, the ligation of **3a** with **7** proceeded very cleanly, the (Lys)₆-Ades tag being cleaved within seconds under the NCL conditions, giving the target 82 aa polypeptide **9** in a 62% isolated yield (ESI p S18-S19).

Additionally, we took the opportunity of this work to demonstrate the applicability of the C-terminal thioester solubilization strategy to *N*-Hnb-Cys crypto-thioesters. Quite expectedly, introduction of a (Lys)₆ tag was straightforward, and gave segment **3b** in good yields without needing any optimization. As for **3a**, NCL with **7** proceeded cleanly, and with comparable kinetics (ESI, fig. S15), giving **9** in an excellent 76% isolated yield (scheme 3).

The cysteine residue in **9** was further desulfurized under classical conditions³² to give an alanine (ESI, p S20-S21) such as in the native MUC1 sequence.



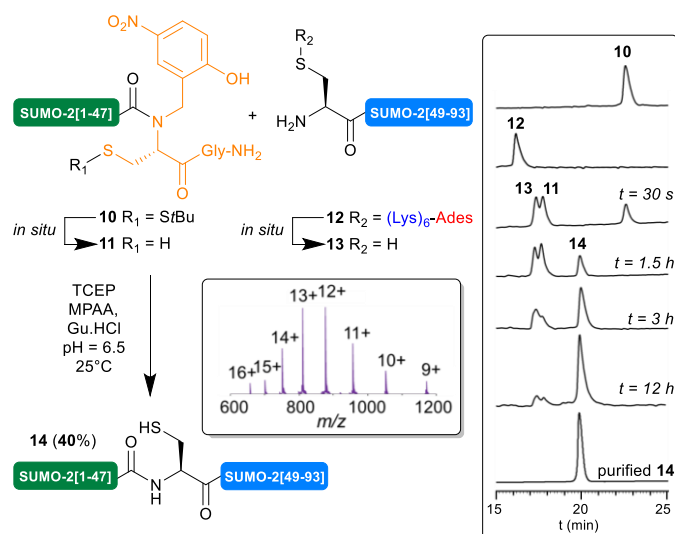
Scheme 3: Native chemical ligation reaction showing the putative (Hnb)Cys self-catalyzed *N*→*S* acyl shift mechanism.

Encouraged by these results, we then implemented this approach for the synthesis of the 93 aa SUMO-2. Small ubiquitin-related modifiers³³ (SUMO) were first discovered in mammals in 1996.³⁴ To date, five SUMO isoforms have been identified in humans, SUMO-1, 2, 3, 4 and 5. SUMOylation is a post-translational modification (PTM) consisting in the covalent attachment of the C-terminus of SUMO proteins *via* an isopeptide bond to specific lysine residues in target proteins. An enzymatic cascade controls the attachment, involving activator (E1), conjugating (E2), and sometimes ligase (E3) enzymes. This PTM is reversible, through deSUMOylation by sentrin/SUMO-specific proteases (SENPs).³⁵ In contrast with ubiquitin,^{36,37} only few examples of the chemical synthesis of SUMO proteins, their dimers and conjugates, have been reported.^{5d,12c,38,39} One illustration is the synthesis by Brik^{12c} of SUMO-2-diubiquitin hybrid chains. In this work, the authors reported the low solubility, tendency to aggregation and unusual HPLC behavior of an N-terminal cysteinyl SUMO-2[46-93]⁴⁰ segment. To circumvent this issue, they developed an elegant C-terminal solubilizing tag in which a 3,4-diaminobenzoic acid (Dbz) linker⁴¹ was employed to attach a poly-Arg tag to the C-terminus of this segment. We thought that it could be interesting to challenge our (Lys)₆-Ades methodology with this benchmark target.

We applied our synthetic protocol to (Lys)₆-Ades-SUMO-2[48-93] (**12**),⁴² which gave a clean and soluble crude product that was purified by standard HPLC. As anticipated from Brik's report, in sharp contrast with **12**, the non-tagged version exhibited anomalous HPLC behavior (ESI, S11, fig. S11) and the crude peptide was barely soluble in HPLC solvents: 0.05 mg/mL in 8:2:0.01 H₂O/MeCN/TFA, thus further validating the choice of this target for this work. Gratifyingly, NCL of **12** with SUMO-2[1-47] *N*-Hnb-Cys crypto thioester **10** proceeded very cleanly as shown in the analytical HPLC monitoring⁴³ of the reaction (scheme 4), giving the expected SUMO-2[1-93] **14** in a good 40% isolated yield. As observed with the MUC1-derived model segments, the HPLC chromatograms show nearly instantaneous conversion of **12** into non-tagged **13**, concomitantly with the slightly slower conversion of *S*-StBu-protected dormant *N*-Hnb-Cys crypto-thioester **10** into an active form **11**.⁴⁴

Note that in accordance with Melnyk's findings,⁴⁵ 15% of a side product showing a molecular mass of [M-18] *m/z* relative to the desired product **14** was observed when performing the reaction

at 37°C for a prolonged time and was attributed to aspartimide formation at one of the Asp-Gly sites. Lowering the temperature to 25°C nearly abolished this side reaction (ESI p S22-S25).



SUMO-2[1-93]: ¹ADEKPKGKGVKTENNNDHINLVKVGQDGSVVQFKIKRHTPLSKLXKAY CE
RQGLSXQRIRFRFDGQPINETDTPAQLXEDEDTIDVFQQQTG⁹³G (X = norleucine)

Scheme 4: NCL-based SUMO-2 synthesis using the (Lys)₆-Ades methodology.

Finally, in order to further characterize the synthesized SUMO-2 **14** from a biochemical point of view, it was folded by simple solubilization into a neutral buffer as described.^{39b,39f,46} Its three-dimensional structure was evaluated by circular dichroism, showing a spectrum essentially identical with the previously reported ones⁴⁶ for recombinant and synthetic SUMO-2 (ESI fig. S21).

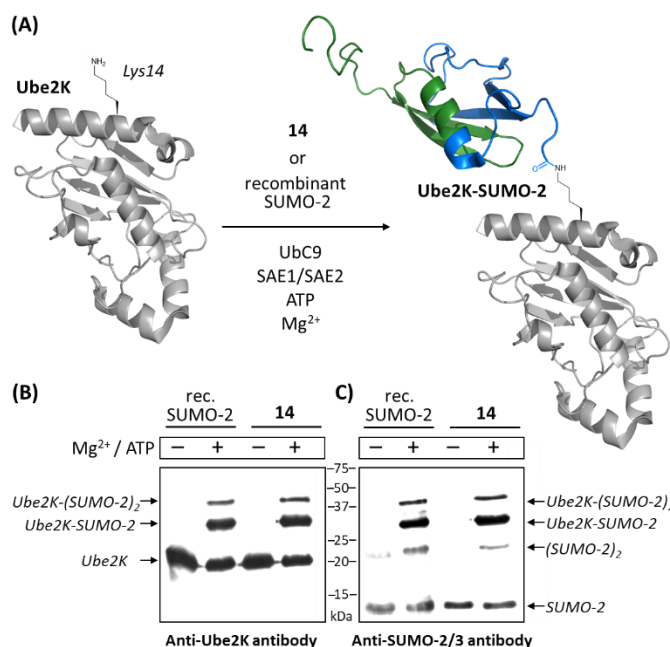


Figure 1: *In vitro* SUMOylation of Ube2K with **14** and a recombinant SUMO-2 used as a control. (A) schematic representation of the reaction (cartoons based on NMR and X-ray structures of SUMO-2 and Ube2K, PDB ID: 2BEP and 2N1W, respectively); (B) western-blot analysis using an anti-Ube2K antibody; (C) western-blot analysis using an anti-SUMO-2/3 antibody.

We then compared **14** to a commercially available recombinant version of SUMO-2 for its ability to act as a substrate of the SUMO E2 conjugating enzyme Ubc9⁴⁷ and a SUMO E1 activating enzyme (namely a heterodimeric complex of SAE1 and SAE2 proteins)⁴⁸ for conjugation to the ubiquitin-conjugating enzyme Ube2K (also known as E2-25K and Huntington Interacting Protein 2). This target is reported to be among the best *in vitro* substrates for Ubc9-dependent SUMOylation known thus far, although its SUMOylated lysine residue (Lys14) is not surrounded by a consensus SUMOylation motif.⁴⁹ Using anti-SUMO-2/3 and anti-Ube2K antibodies, western-blot analyses first showed that synthetic SUMO-2 is recognized as efficiently as the recombinant one. The appearance of an intense band recognized by both antibodies and migrating at the expected molecular mass of a SUMO-modified Ube2K demonstrated the successful SUMOylation of the acceptor protein in both cases. Weaker bands likely corresponding to di-SUMOylated Ube2K and di-SUMO-2 were also observed, consistent with the presence of a SUMOylable lysine in SUMO-2 (Lys11). As expected, no reactions occurred in the absence of ATP and Mg²⁺, required cofactors for the E1 enzyme. Altogether, these experiments demonstrate the fully functional nature of synthesized **14** (Fig. 1 and ESI p S27-S29).

Conclusions

Poor solubility and aggregation of peptide segments are main bottlenecks for the chemical synthesis of proteins using native chemical ligation. Numerous strategies for the solubilization of problematic segments through temporary modification have been developed such as the introduction of solubilizing tags, but often require complex synthetic strategies, in-house synthesis of building blocks or extra steps to generate the native protein sequence after the NCL-based assembly. In this work, we have introduced a straightforward methodology for the temporary solubilization of N-terminal cysteinyl segments, based on the introduction of an oligolysine tag through a disulfide linkage with the N-terminal cysteine residue. This (Lys)₆-Ades tag is easily incorporated in the target segment through automated solid-phase synthesis using commercially available building blocks, is stable during handling, purification and storage of the segment, while being cleaved within seconds under NCL conditions⁵⁰ to generate *in situ* the reactive free cysteine. We exemplified the broad potential of this method through the NCL-based synthesis of a model polypeptide derived from the human mucin MUC1, in addition to a well-known difficult small protein target, SUMO-2. Due to its overall simplicity and efficiency, we believe that this strategy will advantageously complement existing methodologies in the synthesis of other challenging proteins.

Conflict of interests

There are no conflicts to declare.

Acknowledgements

We warmly thank Jean-Baptiste Madinier for assistance in peptide synthesis, Dr Cyril Colas for HRMS analyses and Dr Franck Costein for his help for the acquisition of CD spectra. The University of Orléans and doctoral school SSBCV (ED549) are gratefully acknowledged for PhD fellowships for SAA and EC, and ANR (EasyMiniprot, ANR-15-CE07-0022) for financial support.

Notes and references

- 1 P. Dawson, T. Muir, I. Clark-Lewis and S. Kent, Synthesis of proteins by native chemical ligation, *Science*, 1994, **266**, 776–779.
- 2 V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoët, J.-C. M. Monbaliu and O. Melnyk, Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations, *Chem. Rev.*, 2019, **119**, 7328–7443.
- 3 (a) M. T. Weinstock, M. T. Jacobsen and M. S. Kay, Synthesis and folding of a mirror-image enzyme reveals ambidextrous chaperone activity, *Proc. Natl. Acad. Sci. USA*, 2014, **111**, 11679–11684; (b) A. Pech, J. Achenbach, M. Jahnz, S. Schülzchen, F. Jarosch, F. Bordusa and S. Klussmann, A thermostable D-polymerase for mirror-image PCR, *Nucleic Acids Res.*, 2017, **45**, 3997–4005; (c) J. Weidmann, M. Schnölzer, P. E. Dawson and J. D. Hoheisel, Copying Life: Synthesis of an Enzymatically Active Mirror-Image DNA-Ligase Made of D-Amino Acids, *Cell Chem. Biol.*, 2019, **26**, 645–651; (d) Z. Wang, W. Xu, L. Liu and T. F. Zhu, A synthetic molecular system capable of mirror-image genetic replication and transcription, *Nat. Chem.*, 2016, **8**, 698–704; (e) W. Xu, W. Jiang, J. Wang, L. Yu, J. Chen, X. Liu, L. Liu and T. F. Zhu, Total chemical synthesis of a thermostable enzyme capable of polymerase chain reaction, *Cell Discov.*, 2017, **3**, 17008; (f) S. Tang, L.-J. Liang, Y.-Y. Si, S. Gao, J.-X. Wang, J. Liang, Z. Mei, J.-S. Zheng and L. Liu, Practical Chemical Synthesis of Atypical Ubiquitin Chains by Using an Isopeptide-Linked Ub Isomer, *Angew. Chem. Int. Ed.*, 2017, **56**, 13333–13337; (g) H. Sun and A. Brik, The Journey for the Total Chemical Synthesis of a 53 kDa Protein, *Acc. Chem. Res.*, 2019, **52**, 3361–3371.
- 4 (a) G. G. Kochendoerfer, D. Salom, J. D. Lear, R. Wilk-Orescan, S. B. H. Kent and W. F. DeGrado, Total Chemical Synthesis of the Integral Membrane Protein Influenza A Virus M2: Role of Its C-Terminal Domain in Tetramer Assembly, *Biochemistry*, 1999, **38**, 11905–11913; (b) M. Dittmann, J. Sauermann, R. Seidel, W. Zimmermann and M. Engelhard, Native chemical ligation of hydrophobic peptides in organic solvents, *J. Pept. Sci.*, 2010, **16**, 558–562; (c) F. Shen, S. Tang and L. Liu, Hexafluoro-2-propanol as a potent cosolvent for chemical ligation of membrane proteins, *Sci. China Chem.*, 2011, **54**, 110–116; (d) M. Dittmann, M. Sadek, R. Seidel and M. Engelhard, Native chemical ligation in dimethylformamide can be performed chemoselectively without racemization: NCL in organic solvents, *J. Pept. Sci.*, 2012, **18**, 312–316.
- 5 (a) F. I. Valiyaveetil, R. MacKinnon and T. W. Muir, Semisynthesis and Folding of the Potassium Channel KcsA, *J. Am. Chem. Soc.*, 2002, **124**, 9113–9120; (b) S. Ingale, T. Buskas and G.-J. Boons, Synthesis of Glyco(lipo)peptides by Liposome-Mediated Native Chemical Ligation, *Org. Lett.*, 2006, **8**, 5785–5788; (c) S. Lahiri, M. Brehms, D. Olschewski and C. F. W. Becker, Total Chemical Synthesis of an Integral Membrane Enzyme: Diacylglycerol Kinase from *Escherichia coli*, *Angew. Chem. Int. Ed.*, 2011, **50**, 3988–3992; (d) J. Bouchenna, M. Sénéchal, H. Drobecq, J. Vicogne and O. Melnyk, Total Chemical Synthesis of All SUMO-2/3 Dimer Combinations, *Bioconjugate Chem.*, 2019, **30**, 2967–2973.
- 6 E. C. B. Johnson and S. B. H. Kent, Studies on the Insolubility of a Transmembrane Peptide from Signal Peptide Peptidase, *J. Am. Chem. Soc.*, 2006, **128**, 7140–7141.
- 7 E. Bianchi, R. Ingenito, R. J. Simon and A. Pessi, Engineering and Chemical Synthesis of a Transmembrane Protein: The HCV Protease Cofactor Protein NS4A, *J. Am. Chem. Soc.*, 1999, **121**, 7698–7699.
- 8 A. M. Levinson, J. H. McGee, A. G. Roberts, G. S. Creech, T. Wang, M. T. Peterson, R. C. Hendrickson, G. L. Verdine and S. J. Danishefsky, Total Chemical Synthesis and Folding of All- α and All- β Variants of Oncogenic KRas(G12V), *J. Am. Chem. Soc.*, 2017, **139**, 7632–7639.
- 9 (a) H. Hojo, H. Katayama, C. Tano, Y. Nakahara, A. Yoneshige, J. Matsuda, Y. Sohma, Y. Kiso and Y. Nakahara, Synthesis of the sphingolipid activator protein, saposin C, using an azido-protected O-acyl isopeptide as an aggregation-disrupting element, *Tetrahedron Lett.*, 2011, **52**, 635–639 (b) Y. Sohma, H. Kitamura, H. Kawashima, H. Hojo, M. Yamashita, K. Akaji and Y. Kiso, Synthesis of an O-acyl isopeptide by using native chemical ligation to efficiently construct a hydrophobic polypeptide, *Tetrahedron Lett.*, 2011, **52**, 7146–7148; (c) H. Kawashima, T. Kuruma, M. Yamashita, Y. Sohma and K. Akaji, Synthesis of an O-acyl isopeptide by using native chemical ligation in an aqueous solvent system, *J. Pept. Sci.*, 2014, **20**, 361–365; (d) R. Desmet, M. Pauzuolis, E. Boll, H. Drobecq, L. Raibaut and O. Melnyk, Synthesis of Unprotected Linear or Cyclic O-Acyl Isopeptides in Water Using Bis(2-sulfanylethyl)amido Peptide Ligation, *Org. Lett.*, 2015, **17**, 3354–3357.
- 10 D.-D. Zhao, X.-W. Fan, H. Hao, H.-L. Zhang and Y. Guo, Temporary Solubilizing Tags Method for the Chemical Synthesis of Hydrophobic Proteins, *Curr. Org. Chem.*, 2019, **23**, 2–13.
- 11 (a) J.-S. Zheng, M. Yu, Y.-K. Qi, S. Tang, F. Shen, Z.-P. Wang, L. Xiao, L. Zhang, C.-L. Tian and L. Liu, Expedient Total Synthesis of Small to Medium-Sized Membrane Proteins via Fmoc Chemistry, *J. Am. Chem. Soc.*, 2014, **136**, 3695–3704; (b) J.-S. Zheng, Y. He, C. Zuo, X.-Y. Cai, S. Tang, Z. A. Wang, L.-H. Zhang, C.-L. Tian and L. Liu, Robust Chemical Synthesis of Membrane Proteins through a General Method of Removable Backbone Modification, *J. Am. Chem. Soc.*, 2016, **138**, 3553–3561; (c) S. Tang, C. Zuo, D.-L. Huang, X.-L. Cai, L.-H. Zhang, C.-L. Tian, J.-S. Zheng and L. Liu, Chemical synthesis of membrane proteins by the removable backbone modification method, *Nat. Protoc.*, 2017, **12**, 2554–2569.
- 12 (a) E. C. B. Johnson, E. Malito, Y. Shen, D. Rich, W.-J. Tang and S. B. H. Kent, Modular Total Chemical Synthesis of a Human Immunodeficiency Virus Type 1 Protease, *J. Am. Chem. Soc.*, 2007, **129**, 11480–11490; (b) P. W. R. Harris and M. A. Brimble, Toward the total chemical synthesis of the cancer protein NY-ESO-1,

- Biopolymers*, 2010, **94**, 542–550; (c) S. Bondalapati, E. Eid, S. M. Mali, C. Wolberger and A. Brik, Total chemical synthesis of SUMO-2-Lys63-linked diubiquitin hybrid chains assisted by removable solubilizing tags, *Chem. Sci.*, 2017, **8**, 4027–4034.
- 13 S. Tsuda, S. Masuda and T. Yoshiya, Solubilizing Trityl-Type Tag To Synthesize Asx/Glx-Containing Peptides, *ChemBioChem*, 2019, **20**, 2063–2069.
- 14 (a) S. K. Maity, G. Mann, M. Jbara, S. Laps, G. Kamnesky and A. Brik, Palladium-Assisted Removal of a Solubilizing Tag from a Cys Side Chain To Facilitate Peptide and Protein Synthesis, *Org. Lett.*, 2016, **18**, 3026–3029; (b) J. A. Brailsford, J. L. Stockdill, A. J. Axelrod, M. T. Peterson, P. A. Vadola, E. V. Johnston and S. J. Danishefsky, Total chemical synthesis of human thyroid-stimulating hormone (hTSH) β -subunit: Application of arginine-tagged acetamidomethyl (AcmR) protecting groups, *Tetrahedron*, 2018, **74**, 1951–1956; (c) S. Tsuda, M. Mochizuki, H. Ishiba, K. Yoshizawa-Kumagaye, H. Nishio, S. Oishi and T. Yoshiya, Easy-to-Attach/Detach Solubilizing-Tag-Aided Chemical Synthesis of an Aggregative Capsid Protein, *Angew. Chem. Int. Ed.*, 2018, **57**, 2105–2109; (d) S. Tsuda, S. Masuda and T. Yoshiya, The versatile use of solubilizing trityl tags for difficult peptide/protein synthesis, *Org. Biomol. Chem.*, 2019, **17**, 1202–1205.
- 15 For a recent example of the incorporation of a tertiary amine as solubilizing tag, see: E. Eid, G. N. Boross, H. Sun, M. Msallam, S. K. Singh and A. Brik, Total Chemical Synthesis of ISGylated-Ubiquitin Hybrid Chain Assisted by Acetamidomethyl Derivatives with Dual Functions, *Bioconjugate Chem.*, 2020, **31**, 889–894.
- 16 Y.-C. Huang, Y.-M. Li, Y. Chen, M. Pan, Y.-T. Li, L. Yu, Q.-X. Guo and L. Liu, Synthesis of Autophagosomal Marker Protein LC3-II under Detergent-Free Conditions, *Angew. Chem. Int. Ed.*, 2013, **52**, 4858–4862.
- 17 (a) Z. Tan, S. Shang and S. J. Danishefsky, Rational development of a strategy for modifying the aggregability of proteins, *Proc. Natl. Acad. Sci. USA*, 2011, **108**, 4297–4302; (b) M. T. Jacobsen, M. E. Petersen, X. Ye, M. Galibert, G. H. Lorimer, V. Aucagne and M. S. Kay, A Helping Hand to Overcome Solubility Challenges in Chemical Protein Synthesis, *J. Am. Chem. Soc.*, 2016, **138**, 11775–11782; (c) S. Tsuda, H. Nishio and T. Yoshiya, Peptide self-cleavage at a canaline residue: application to a solubilizing tag system for native chemical ligation, *Chem. Commun.*, 2018, **54**, 8861–8864; (d) J. M. Fulcher, M. E. Petersen, R. J. Giesler, Z. S. Cruz, D. M. Eckert, J. N. Francis, E. M. Kawamoto, M. T. Jacobsen and M. J. Kay, Chemical Synthesis of Shiga Toxin Subunit B Using a Next-Generation Traceless “Helping Hand” Solubilizing Tag, *Org. Biomol. Chem.*, 2019, **17**, 10237–10244.
- 18 T. Sato, Y. Saito and S. Aimoto, Synthesis of the C-terminal region of opioid receptor like 1 in an SDS micelle by the native chemical ligation: effect of thiol additive and SDS concentration on ligation efficiency, *J. Peptide Sci.*, 2005, **11**, 410–416.
- 19 E. C. B. Johnson and S. B. H. Kent, Towards the total chemical synthesis of integral membrane proteins: a general method for the synthesis of hydrophobic peptide-athioester building blocks, *Tetrahedron Lett.*, 2007, **48**, 1795–1799.
- 20 A. C. Baumruck, D. Tietze, L. K. Steinacker and A. A. Tietze, Chemical synthesis of membrane proteins: a model study on the influenza virus B proton channel, *Chem. Sci.*, 2018, **9**, 2365–2375.
- 21 P. Botti, M. Villain, S. Manganiello and H. Gaertner, Native Chemical Ligation through in Situ O to S Acyl Shift, *Org. Lett.*, 2004, **6**, 4861–4864.
- 22 K. Sato, A. Shigenaga, K. Tsuji, S. Tsuda, Y. Sumikawa, K. Sakamoto, A. Otaka, N-sulfanylethylanilide peptide as a cryptothioester peptide, *ChemBioChem*, 2011, **12**, 1840–1844.
- 23 Note that, besides cationic solubilization tags, concomitant rearrangement of Ser/Thr O-acyl isopeptides during NCL has been exploited, taking advantage of the aggregation-inhibiting properties of the depsipeptide linkage during HPLC purification of segments: see for example references 3d, 3e and 9a.
- 24 K. M. Linn, M. G. Derebe, Y. Jiang and F. I. Valiyaveetil, Semisynthesis of NaK, a Na⁺ and K⁺ Conducting Ion Channel, *Biochemistry*, 2010, **49**, 4450–4456.
- 25 T. W. Muir, D. Sondhi and P. A. Cole, Expressed protein ligation: A general method for protein engineering, *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 6705–6710.
- 26 E. C. B. Johnson and S. B. H. Kent, Insights into the Mechanism and Catalysis of the Native Chemical Ligation Reaction, *J. Am. Chem. Soc.*, 2006, **128**, 6640–6646.
- 27 R. Matsueda, S. Higashida, R. J. Ridge and G. R. Matsueda, Activation of conventional S-protecting groups of cysteine by conversion into the 3-nitro-2-pyridinesulfonyl (Npys) group, *Chem. Lett.*, 1982, **11**, 921–924.
- 28 W. Tegge, W. Bautsch and R. Frank, Synthesis of cyclic peptides and peptide libraries on a new disulfide linker, *J. Pept. Sci.*, 2007, **13**, 693–699.
- 29 (a) C. Seidel, E. Klauschen, K. D. Kaufmann and H. Kunzek, Selektiv S-geschützte Cysteinpeptide. Zur Synthese von Cysteinpeptiden unter Verwendung der S-Äthylmercapto-Schutzgruppe, *J. Prakt. Chem.*, 1976, **318**, 229–247; (b) D. Andreu, F. Albericio, N. A. Solé, M. C. Munson, M. Ferrer and G. Barany, Formation of Disulfide Bonds in Synthetic Peptides and Proteins, 1994, In: Pennington M.W., Dunn B.M. (eds) *Peptide Synthesis Protocols. Methods in Molecular Biology*, vol 35. Humana Press, Totowa, NJ.
- 30 For other examples of use of MUC1 VNTRs as models in protein chemical synthesis, see: (a) V. Aucagne, I. E. Valverde, P. Marceau, M. Galibert, N. Dendane and A. F. Delmas, Towards the Simplification of Protein Synthesis: Iterative Solid-Supported Ligations with Concomitant Purifications, *Angew. Chem. Int. Ed.*, 2012, **51**, 11320–11324; (b) I. E. Decostaire, D. Lelièvre, V. Aucagne and A. F. Delmas, Solid phase oxime ligations for the iterative synthesis of polypeptide conjugates, *Org. Biomol. Chem.*, 2014, **12**, 5536–5543; (c) M. Galibert, V. Piller, F. Piller, V. Aucagne and A. F. Delmas, Combining triazole ligation and enzymatic glycosylation on solid phase simplifies the synthesis of very long glycoprotein analogues, *Chem. Sci.*, 2015, **6**, 3617–3623; (d) S. F. Loibl, Z. Harpaz,

R. Zitterbart and O. Seitz, Total chemical synthesis of proteins without HPLC purification, *Chem. Sci.*, 2016, **7**, 6753–6759.

31 (a) V. P. Terrier, H. Adihou, M. Arnould, A. F. Delmas and V. Aucagne, A straightforward method for automated Fmoc-based synthesis of bio-inspired peptide crypto-thioesters, *Chem. Sci.*, 2016, **7**, 339–345; (b) D. Lelièvre, V. P. Terrier, A. F. Delmas and V. Aucagne, Native Chemical Ligation Strategy to Overcome Side Reactions during Fmoc-Based Synthesis of C-Terminal Cysteine-Containing Peptides, *Org. Lett.*, 2016, **18**, 920–923; (c) G. Martinez, J.-P. Hograindleur, S. Voisin, R. Abi Nahed, T. M. Abd El Aziz, J. Escoffier, J. Bessonnat, C.-M. Fovet, M. De Waard, S. Hennebicq, V. Aucagne, P. F. Ray, E. Schmitt, P. Bulet and C. Arnould, Spermaurin, an La1-like peptide from the venom of the scorpion *Scorpio maurus palmatus*, improves sperm motility and fertilization in different mammalian species, *Mol. Hum. Reprod.*, 2016, **10**, 116–131; (d) V. P. Terrier, A. F. Delmas and V. Aucagne, Efficient synthesis of cysteine-rich cyclic peptides through intramolecular native chemical ligation of N-Hnb-Cys peptide crypto-thioesters, *Org. Biomol. Chem.*, 2017, **15**, 316–319; (e) K. Loth, A. Vergnes, C. Barreto, S. N. Voisin, H. Meudal, J. Da Silva, A. Bressan, N. Belmadi, E. Bachère, V. Aucagne, C. Cazevielle, H. Marchandin, R. D. Rosa, P. Bulet, L. Touqui, A. F. Delmas, D. Destoumieux-Garzón, The Ancestral N-Terminal Domain of Big Defensins Drives Bacterially Triggered Assembly into Antimicrobial Nanonets, *mBio*, 2019, **10**, e01821–19; (f) L. De Rosa, R. Di Stasi, L. D. D'Andrea, *Tetrahedron*, 2019, **75**, 894–905; (g) S. A. Abboud and V. Aucagne, An optimized protocol for the synthesis of N-2-hydroxybenzyl-cysteine peptide crypto-thioesters, *Org. Biomol. Chem.*, 2020, **18**, 8199–8208.

32 Q. Wan and S. J. Danishefsky, Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides, *Angew. Chem. Int. Ed.*, 2007, **46**, 9248–9252.

33 Z.-J. Han, Y.-H. Feng, B.-H. Gu, Y.-M. Li and H. Chen, The post-translational modification, SUMOylation, and cancer, *Int. J. Oncol.*, 2018, **52**, 1081–1094.

34 M. J. Matunis, E. Coutavas and G. Blobel, A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex., *J. Cell Biol.*, 1996, **135**, 1457–1470.

35 C. Guo and J. M. Henley, Wrestling with stress: Roles of protein SUMOylation and deSUMOylation in cell stress response: SUMOylation in Cell Stress, *IUBMB Life*, 2014, **66**, 71–77.

36 G. Goldstein, M. Scheid, U. Hammerling, D. H. Schlesinger, H. D. Niall and E. A. Boyse, Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells., *Proc. Natl. Acad. Sci. USA*, 1975, **72**, 11–15.

37 For recent reviews on the chemical synthesis of ubiquitin derivatives, see for example: (a) M. P. Mulder, K. F. Witting and H. Ovaa, Cracking the Ubiquitin Code: The Ubiquitin Toolbox, *Curr. Issues Mol. Biol.*, 2020, **37**, 1–20; (b) R. Meledin, S. Mali and A. Brik, Pushing the Boundaries of Chemical Protein Synthesis: The Case of Ubiquitin Chains and Polyubiquitinated Peptides and Proteins, *Chem. Rec.*, 2016, **16**, 509–519.

38 For a review on the chemical synthesis of SUMO derivatives, see: O. Melnyk and J. Vicogne, Total chemical synthesis of SUMO proteins, *Tetrahedron*, 2016, **57**, 4319–4324.

39 (a) E. Boll, H. Drobecq, N. Ollivier, L. Raibaut, R. Desmet, J. Vicogne and O. Melnyk, A novel PEG-based solid support enables the synthesis of >50 amino-acid peptide thioesters and the total synthesis of a functional SUMO-1 peptide conjugate, *Chem. Sci.*, 2014, **5**, 2017–2022; (b) T. G. Wucherpennig, V. R. Pattabiraman, F. R. P. Limberg, J. Ruiz-Rodríguez and J. W. Bode, Traceless Preparation of C-Terminal α -Ketoacids for Chemical Protein Synthesis by α -Ketoacid-Hydroxylamine Ligation: Synthesis of SUMO2/3, *Angew. Chem. Int. Ed.*, 2014, **53**, 12248–12252; (c) E. Boll, H. Drobecq, N. Ollivier, A. Blanpain, L. Raibaut, R. Desmet, J. Vicogne and O. Melnyk, One-pot chemical synthesis of small ubiquitin-like modifier protein–peptide conjugates using bis (2-sulfanylethyl)amido peptide latent thioester surrogates, *Nat. Protoc.*, 2015, **10**, 269–292; (d) H. Drobecq, E. Boll, M. Sénéchal, R. Desmet, J.-M. Saliou, J.-J. Lacapère, A. Mougél, J. Vicogne and O. Melnyk, A Central Cysteine Residue Is Essential for the Thermal Stability and Function of SUMO-1 Protein and SUMO-1 Peptide–Protein Conjugates, *Bioconjugate Chem.*, 2016, **27**, 1540–1546; (e) M. P. C. Mulder, R. Merckx, K. F. Witting, D. S. Hameed, D. El Atmioui, L. Lelieveld, F. Liebelt, J. Neefjes, I. Berlin, A. C. O. Vertegaal and H. Ovaa, Total Chemical Synthesis of SUMO and SUMO-Based Probes for Profiling the Activity of SUMO-Specific Proteases, *Angew. Chem. Int. Ed.*, 2018, **57**, 8958–8962; (f) J. Bouchenna, M. Sénéchal, H. Drobecq, N. Stankovic-Valentin, J. Vicogne and O. Melnyk, The Role of the Conserved SUMO-2/3 Cysteine Residue on Domain Structure Investigated Using Protein Chemical Synthesis, *Bioconjugate Chem.*, 2019, **30**, 2684–2696; (g) S. Baldauf, A. O. Ogunkoya, G. N. Boross and J. W. Bode, Aspartic Acid Forming α -Ketoacid–Hydroxylamine (KAHA) Ligations with (S)-4,4-Difluoro-5-oxaproline, *J. Org. Chem.*, 2020, **85**, 1352–1364.

40 Note that for practical reasons, in this case the native cysteine Cys49 was mutated to a serine, and that ligation was performed at an Ala48->Cys site which was desulfurized after the NCL to restore the native alanine residue.

41 J. B. Blanco-Canosa and P. E. Dawson, An Efficient Fmoc-SPPS Approach for the Generation of Thioester Peptide Precursors for Use in Native Chemical Ligation, *Angew. Chem.*, 2008, **120**, 6957–6961.

42 SUMO-2 methionines residues (Met1, Met44, Met55 and Met78) were substituted by isosteric norleucines in **10** and **12**, following the common use in chemical protein synthesis to avoid air-mediated oxidation of the side chain sulfides into sulfoxides.

43 Note that the HPLC monitoring of the NCL reaction was delicate, due to the *in situ* formation of non-tagged SUMO-2[48-93] **S11** that required extensive washes of the column between analyses.

44 S-tert-butylsulfanyl (S-StBu) is used to afford the crypto-thioester in a dormant state, unable to rearrange into a reactive thioester during handling, purification and storage. Under classical NCL conditions (including TCEP and MPAA), StBu is readily cleaved within a few minutes affording active N-Hnb-Cys crypto-thioester species.

45 J. Bouchenna, M. Sénéchal, H. Drobecq, J. Vicogne and O. Melnyk, The Problem of Aspartimide Formation During Protein Chemical Synthesis Using SEA-Mediated Ligation, 2019, In: Iranzo O., Roque A. (eds) Peptide and Protein Engineering. Springer Protocols Handbooks. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-0720-6_2

46 M. Sénéchal, J. Bouchenna, J. Vicogne and O. Melnyk, Comment on "N-terminal Protein Tail Acts as Aggregation Protective Entropic Bristles: The SUMO Case", *Biomacromolecules*, 2020, **21**, 3480–3482.

47 L. M. Lois and C. D. Lima, Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1, *EMBO J.*, 2005, **24**, 439–451.

48 B. A. Schulman and J. W. Harper, Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signaling pathways, *Nat. Rev. Mol. Cell Biol.*, 2009, **10**, 319–331.

49 A. Pichler, P. Knipscheer, E. Oberhofer, W. J. van Dijk, R. Körner, J. V. Olsen, S. Jentsch, F. Melchior and T. K. Sixma, SUMO modification of the ubiquitin-conjugating enzyme E2-25K, *Nat. Struct. Mol. Biol.*, 2005, **12**, 264–269.

50 Note that we only tested the (Lys)₆-Ades solubilizing tag under the most classically used conditions for NCL, including a large excess of TCEP and MPAA. In rare cases, NCL reactions are performed under non-reductive conditions, either making use of pre-formed arylthioesters (such as MPAA thioesters), or using azaheterocycles as catalysts instead of an external arylthiol: the strategy herein described would expectedly not be applicable.

V. Supporting Information

These results were accepted for publication in Chemical Science. Here we reproduced the corresponding supporting information file.

A straightforward methodology to overcome solubility challenges for N-terminal cysteinyl segments used for native chemical ligation.

Skander A. Abboud, El hadji Cisse, Michel Doudeau, Hélène Bénédicti, Vincent Aucagne*

Centre de Biophysique Moléculaire, CNRS UPR 4301, Rue Charles Sadron 45071 Orléans cedex 2,
France. E-mail: aucagne@cnrs-orleans.fr

Supporting Information

1- General information.....	333
2- General procedures for solid phase peptide synthesis	333
3- Optimization of the disulfide linker on a model tripeptide	336
3-1- Synthesis of Boc-Cys(Npys)-peptidyl resin S3	336
3-2- Reaction of cysteamine hydrochloride 1.HCl with S3 to give S4	337
3-3- Reaction of 2-amino-1,1-dimethyl-ethane-1-thiol hydrochloride with S2 to give S5	338
3-4- Stability of disulfide-containing peptidyl resins S4 and S5 towards piperidine treatment.....	339
3-5- Optimization of the reaction of 2 with S3	341
4- Synthesis and purification of MUC1 derived peptide segments	342
4-1- H-Trp-(MUC1) ₂ -(Hnb)Cys(StBu)-Gly-NH ₂ crypto-thioester (3a)	342
4-2 H-Cys(K ₆ -Ades)-(MUC1) ₂ -Trp-NH ₂ cysteinyl peptide (7)	343
4-3 H-Trp-(MUC1) ₂ -(Hnb)Cys(StBu)-Gly-(Lys) ₆ -NH ₂ crypto-thioester (3b)	344
5- Synthesis and purification of SUMO-2 derived peptide segments.....	345
5-1 SUMO-2[1-47]-(Hnb)Cys(StBu)-Gly-NH ₂ crypto-thioester (10)	345
5-2- SUMO-2[48-93] cysteinyl peptide (S11).....	346
5-3- Cys(K ₆ -Ades)-SUMO-2[49-93] cysteinyl peptide (12).....	347
6- Native chemical ligations (NCL)	349
6-1 NCL between 3a and 7 to give 9	349
6-2- NCL between 3b and 7 to give 9	350
6-3- Desulfurization of 9	351
6-4 NCL between 10 and 12 to give SUMO-2[1-93] (14).....	353
7- Functional and structural characterization of SUMO-2[1-93] (14).....	357
7-1- Circular dichroism	357
7-2- Enzymatic conjugation	358

1- General information

All reagents and solvents were used without further purification. Protected amino acids were purchased from Gyros Protein Technology (Uppsala, Sweden). 1-Amino-2-methyl-2-propanethiol hydrochloride (**2**.HCl) was purchased from Key organics (Camelford, UK). DIEA was purchased from Carlo Erba (Val-de-Reuil, France). Rink amide ChemMatrix resin was purchased from Biotage (Uppsala, Sweden). Peptide synthesis grade DMF was obtained from VWR (Fontenay-sous-Bois, France). Fmoc-Cys(NPys)-OH was purchased from BACHEM (Bubendorf, Switzerland). All other chemicals were from Sigma Aldrich (St-Quentin-Fallavier, France) and solvents from Carlo Erba. Ultrapure water was obtained using a Milli-Q water system from Millipore (Molsheim, France). Polypropylene syringes fitted with polypropylene frits were obtained from Torviq (Niles, MI, USA) and were equipped with PTFE stopcocks bought from Biotage. HPLC analyses were carried out on a Chromaster system equipped with a 5160 pump, a 5430 diode array detector and a 5260 auto sampler and semi-preparative purifications were carried out on a LaChromElite system equipped with a Hitachi L-2130 pump, a Hitachi L-2455 diode array detector and a Hitachi L-2200 auto sampler. Chromolith High Resolution RP-18e (150 Å, 10 × 4.6 mm, 3 mL/min flow rate) columns were used for analysis, Nucleosil C18 (300 Å, 5µm, 250 × 10mm, 3 mL/min flow rate) and Jupiter C4 (300 Å, 5µm, 250 × 10mm, 3 mL/min flow rate) for purification. Solvents A and B are 0.1 % TFA in H₂O and 0.1 % TFA in MeCN, respectively. Each gradient was followed by a washing step to elute any compound not eluted during the gradient (up to 95% B/A over 0.5 min, then isocratic 95% B/A for 0.5 min for the HR Chromolith). LC-MS analyses were carried out on an Agilent 1260 Infinity HPLC system, coupled with an Agilent 6120 mass spectrometer (ESI + mode), and fitted with an Aeris Widepore XB-C18 2 (3.6 µm, 150 × 2.1 mm, 0.5 mL/min flow rate, 60°C) column. The reported *m/z* values correspond to the monoisotopic ions if not specified otherwise. Solvents A' and B' were 0.1 % formic acid in H₂O and 0.1 % formic acid in MeCN, respectively. Gradient: 3% B'/A' for 0.6 min, then 3 to 50% B'/A' over 10.8 min. Low resolution MS of pure compounds were obtained using this system. When needed, the multiply-charged envelope was deconvoluted using the charge deconvolution tool in Agilent OpenLab CDS ChemStation software to obtain the average [M] value. High resolution ESI-MS analyses were performed on a maXisTM ultra-high-resolution Q-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany), using the positive mode. The multiply-charged envelope was deconvoluted using the charge deconvolution algorithm in Bruker Data Analysis 4.1 software to obtain the monoisotopic [M] value.

For yield calculations purposes, the quantities of purified peptides were determined by weight, taking into account a molecular weight including trifluoroacetate counter-ions (one per Arg, His, Lys and N-terminal amine of the peptide sequence) but not water content.

2- General procedures for solid phase peptide synthesis

Protocol PS1- Peptide elongation: Automated Fmoc-based solid phase peptide syntheses (SPPS) were carried out on a Prelude synthesizer from Protein technologies. Manual couplings were performed on polypropylene syringes fitted with polypropylene frits using rotation stirring. The side-chain protecting groups used were Arg(Pbf), Asp(OtBu), Cys(Npys), Cys(Trt), Cys(StBu), Glu(OtBu), Gln(Trt), His(Trt), Lys(Boc), Ser(tBu), Thr(tBu), and Tyr(tBu). Syntheses were performed on a 0.025 mmol-per-reactor scale. Protected amino acids (0.25mmol, 10 equiv.) were coupled using HATU (90 mg, 0.238 mmol, 9.5 equiv.) and DIEA (87 µL, 0.5 mmol, 20 equiv.) in NMP (3 mL) for 30 min. Coupling on *N*-Hnb-cysteine secondary amine were performed through three successive couplings for 2 h. Capping of possible unreacted amine groups was achieved by treatment with acetic anhydride (143 µL, 1.51 mmol, 60 equiv.), DIEA (68 µL, 0.39 mmol, 15.5 equiv.) and HOBt (6 mg, 0.044 mmol, 1.8 equiv.) in NMP (3 mL)

for 7 min (4 x 7 min in the case of *N*-Hnb-cysteine secondary amine). Fmoc group was deprotected by three successive treatments with 20% piperidine in NMP (3 mL) for 3 min.

Protocol PS2-Reductive amination: introduction of the hydroxyl nitrobenzyl group (Hnb): 25 μ mol of H-Cys(StBu)-Gly-Rink-resin (1a-c) was washed with 1:1 DMF/MeOH (4 x 3 mL, 30 s). 2-Hydroxy-5-nitrobenzaldehyde (42 mg, 10 equiv.) in 2 mL 44.5:44.5:1 DMF/MeOH/AcOH (125 mM aldehyde concentration) was then added, and the reactor was stirred for 5 min. The reactor was drained and the resin was washed with 1:1 DMF/MeOH (3 x 3 mL, 5 s) then DMF (3 x 3 mL, 5 s). Without delay, a fresh solution of sodium borohydride (19 mg, 20 equiv.) in 2 mL DMF (250 mM borohydride concentration) was added and the reactor was stirred for 20 min. The reactor was drained and the resin was washed with DMF (4 x 3 mL, 30 s), 20% v/v piperidine in NMP (3 x 3 mL, 3 min), NMP (3 x 3 mL, 30 s), dichloromethane (3 x 5 mL, 30 s) and NMP (3 x 3 mL, 30 s).

Protocol PS3-Boc-Cys(Ades) introduction: 25 μ mol of peptidyl resin was washed with 3 mL of NMP. A solution of Boc-Cys(Npys)-OH (93.5 mg, 10 equiv.), HATU (90.3 mg, 9.5equiv.) and DIEA (87 μ L, 20 equiv.) in 2.5 mL of NMP (100 mM Boc-Cys(Npys)-OH) was added to the resin and the reactor was stirred for 4 hours. The reactor was drained and the resin was washed with DCM (3 x 3 mL, 30 s) and NMP (3 x 3 mL, 30 s). A solution of 2-amino-1,1-dimethyl-propane-1-thiol hydrochloride **2** (35.5 mg, 10 equiv.) in 2.5 mL of NMP (100 mM thiol) was then added to the resin and the reactor was stirred for 5 min. The reactor was drained off and the resin was washed with DCM (3 x 3 mL, 30 s) and NMP (3 x 3 mL, 30 s).

Protocol PS4-Peptide cleavage: The crude peptide was deprotected and cleaved from the resin through a treatment with TFA/H₂O/*i*Pr₃SiH/phenol, 88:5:2:5 for 2 h, and the peptide was precipitated by dilution into an ice-cold diethyl ether/petroleum ether 1:1 mixture, recovered by centrifugation and washed twice with diethyl ether.

Protocol PS5-Small peptides cleavage (less than 4 residues): Small peptides were cleaved from the resin through a treatment with TFA/H₂O/*i*Pr₃SiH, 93:5:2 for 30 min following by concentration under vacuum.

Protocol PS6 - Procedure for selective Hnb ester cleavage to allow UV titration of Fmoc deprotection: As a consequence of the formation of variable amount (5-90 %) of *O*-acylated Hnb during each coupling, this ester being cleaved upon piperidine treatment during Fmoc deprotection, standard UV titration of the fluorenylmethyl-piperidine adduct after Fmoc deprotection is useless unless using a prior treatment for selective ester cleavage before piperidine treatment. Ester cleavage mixture was prepared as follows: 1.25 g (1.80 mmol) of NH₂OH·HCl and 0.92 g (1.35 mmol) of imidazole were suspended in 5 mL of NMP and the mixture was sonicated until complete dissolution. This solution can be stored for few months at -20 °C. 5 volumes of this solution is diluted with 1 volume of DCM prior to utilization, and the resin is treated with the resulting mixture for 3 x 20 min for quantitative ester cleavage. The fluorenylmethyl-piperidine adduct is quantified by UV spectroscopy at $\lambda = 301$ nm ($\epsilon = 7800$ L mol⁻¹ cm⁻¹) in order to evaluate the Fmoc SPPS elongation yield of crypto-thioester peptides.

Protocol PS7-General procedure for native chemical ligation:

Preparation of the NCL buffer (100 mM MPAA, 50 mM TCEP, 6 M guanidinium chloride, 200 mM sodium phosphate, pH = 6.5) :

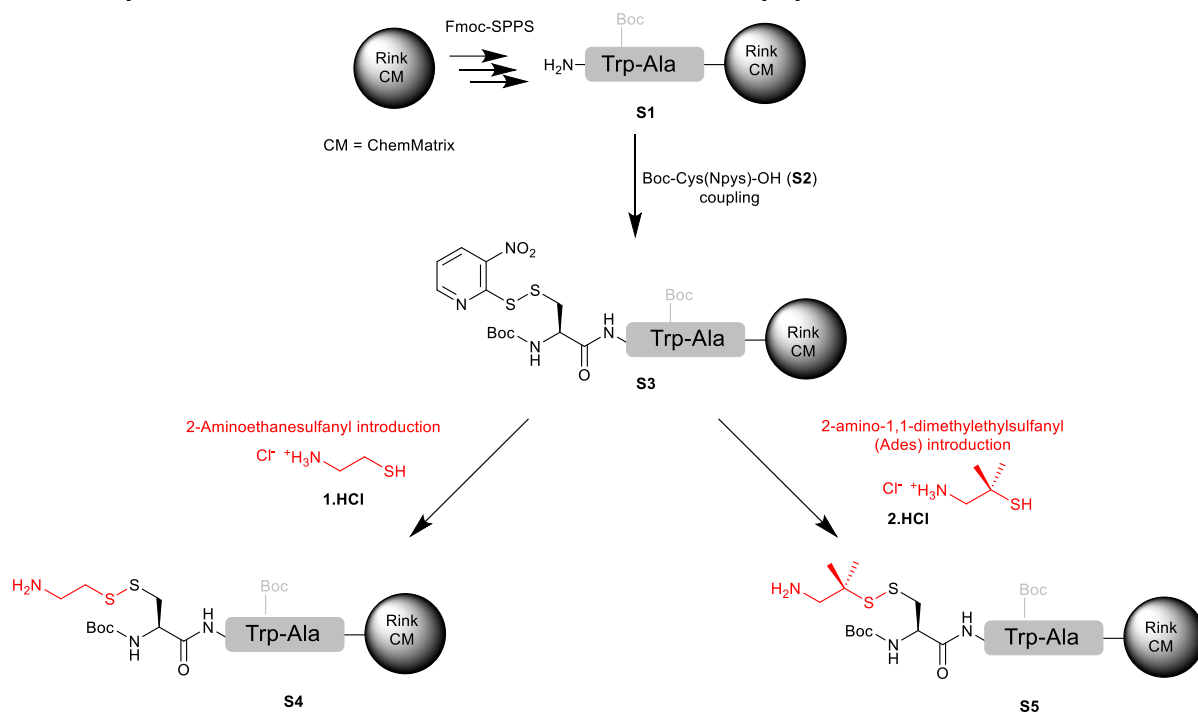
16.8 mg 4-mercaptophenylacetic acid (MPAA, 0.18 mmol), 14.3 mg tris-carboxyethylphosphine hydrochloride (TECP, 0.057 mmol) and 10.3 mg of dry NaOH powder (0.26 mmol) were weighted into a vial, which was sealed with a septum and purged with an argon flow.

A 200 mM disodium hydrogen phosphate, 6 M guanidinium chloride aqueous solution was prepared by dissolving 356 mg disodium hydrogen phosphate dihydrate (2 mmol) and 5.73 g guanidinium chloride in water (10 mL final volume). This solution can be kept at 4°C for several months. Just before use, this solution was thoroughly deoxygenated through successive vacuum/argon cycles.

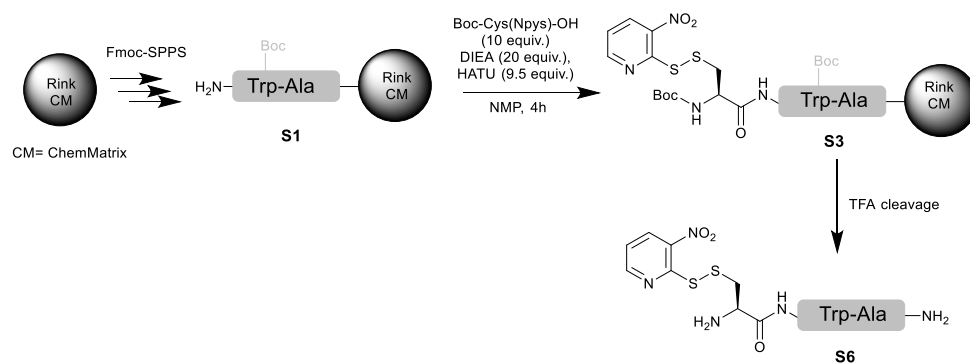
Under argon, 1 mL of this solution was added to the MPAA/TCEP/NaOH vial, followed by sonication upon complete dissolution to give the NCL buffer which was immediately used.

NCL reaction: The cysteinyl peptide and crypto-thioester were weighted in a centrifuge tube, which was sealed with a septum and purged with an argon flow. The volume of the NCL buffer appropriate to reach the desired final peptide segments concentration was added under argon, the reaction vessel was sealed with Parafilm, and the resulting yellow solution was allowed to stir at 25 °C or 37 °C. The reaction was quenched by dilution into 20 volumes of an 80:20:0.1 water/MeCN/TFA mixture.

3- Optimization of the disulfide linker on a model tripeptide

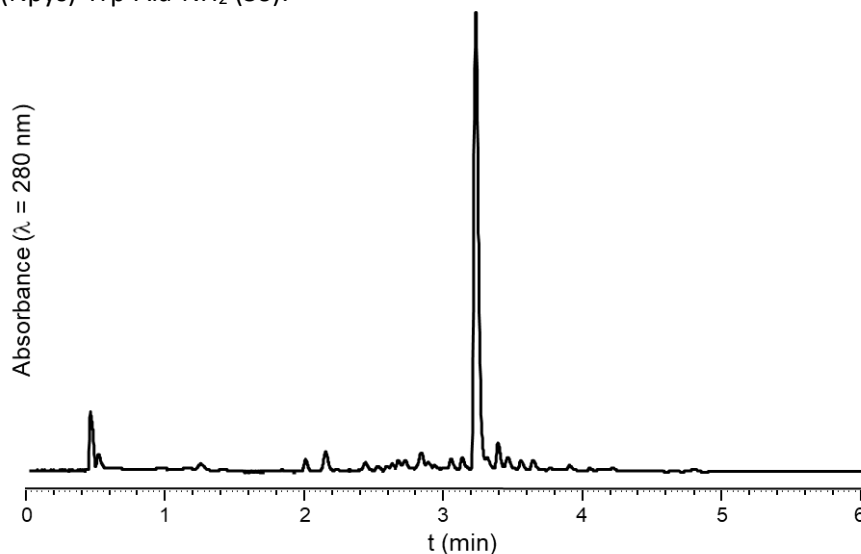


3-1- Synthesis of Boc-Cys(Npys)-peptidyl resin **S3**



Peptidyl resin **S3** was obtained through manual SPPS (protocol PS1) starting from Rink-ChemMatrix resin (0.4 mmol/g). The coupling of Boc-Cys(Npys)-OH was performed as described in protocol PS3. A small amount of the peptidyl resin was cleaved for analytical purpose to afford tripeptide **S6** (protocol PS5 small peptides cleavage).

- H-Cys(Npys)-Trp-Ala-NH₂ (**S6**):

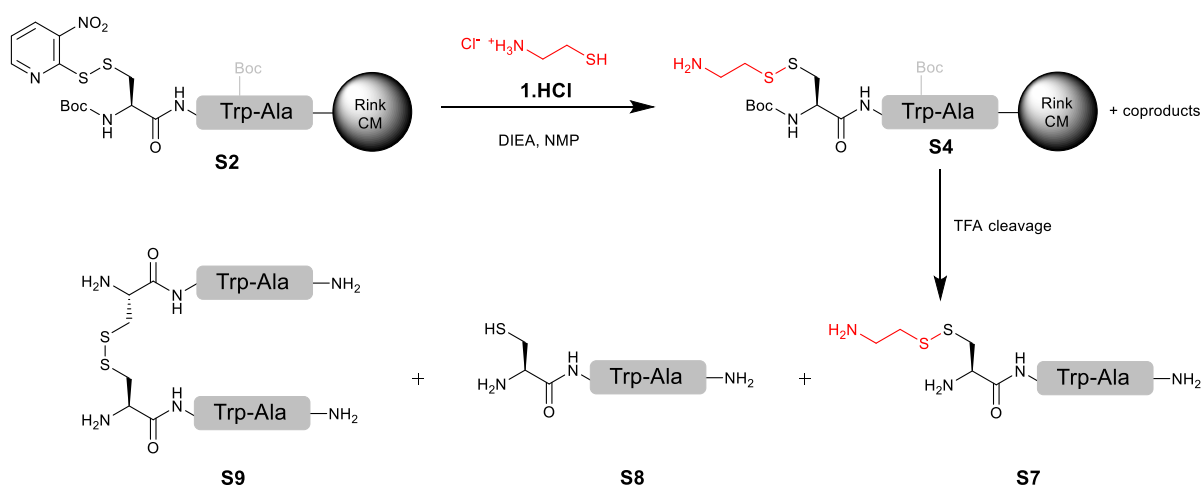


Supplementary figure S1: HPLC trace of crude **S6**

ESI-MS (m/z): [M] calcd. for C₂₂H₂₅N₇O₅S₂: 531.1, found: 531.1

HPLC analysis: t_R = 3.22 min (Chromolith, gradient: 5-50% B/A over 5 min)

3-2- Reaction of cysteamine hydrochloride 1.HCl with **S3** to give **S4**

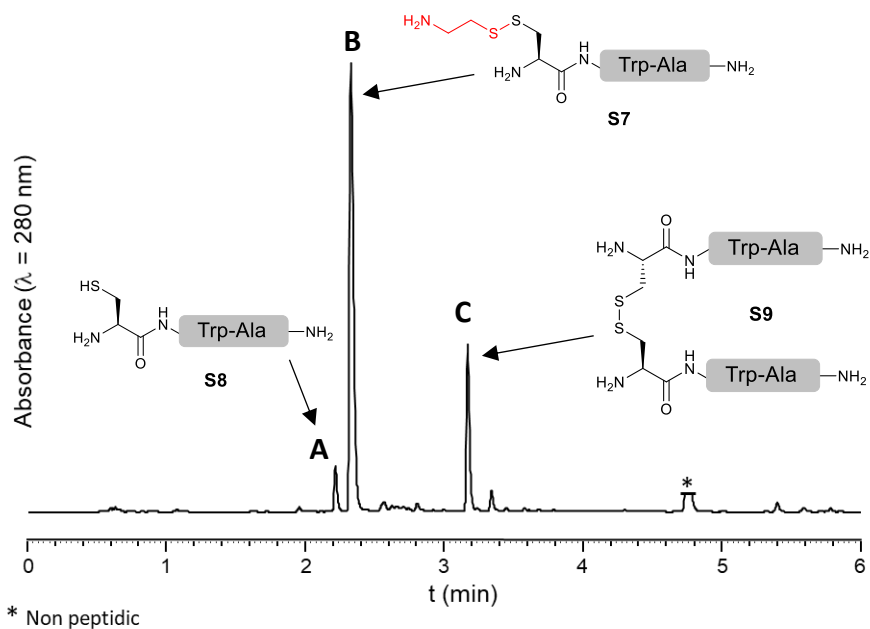


10 μ mol of peptidyl resin **S2** was washed with 3 mL of NMP. A solution of 2-aminoethanethiol hydrochloride (8 mg, 10 equiv.) in 1 mL of NMP (100 mM thiol concentration) was added to the resin. Then DIEA (37.8 μ L, 20 equiv.) was added and the reactor was stirred for 5 min. The reactor was drained and the resin was washed with DCM (3 \times 3 mL, 30 s) and NMP (3 \times 3 mL, 30 s). A small amount of the peptidyl resin was cleaved for analytical purpose affording crude **S7** (protocol PS5 small peptides cleavage).

- H-Cys(SCH₂CH₂NH₂)-Trp-Ala-NH₂ (**S7**):

ESI-MS (*m/z*): [M] calcd. for C₁₉H₂₈N₆O₃S₂: 452.2, found: 452.1

HPLC analysis: *t_R* = 2.22 min (Chromolith, gradient: 5-50% B/A over 5 min)

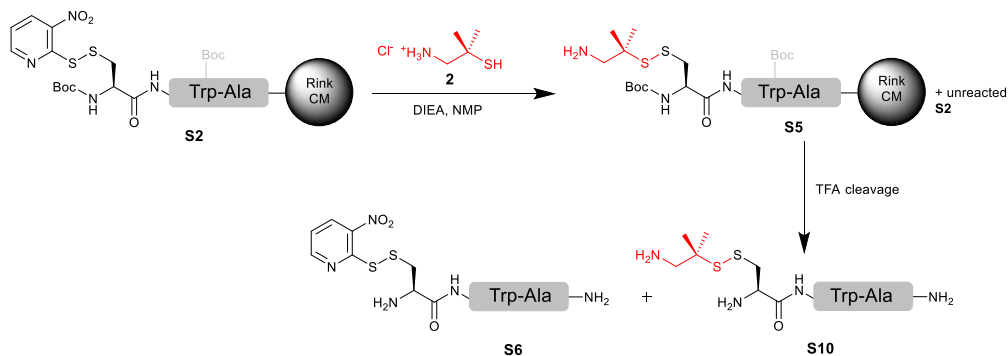


Peak (<i>t_R</i> (min))	[M] (<i>m/z</i>) calcd.	[M] (<i>m/z</i>) found	Attributed to
A (2.36)	377.1	377.2	S8
B (2.22)	452.2	452.1	S7
C (3.20)	752.3	752.2	S9
* (5.00)	No MS signal		Non peptidic compound

Supplementary figure S2: HPLC trace and MS characterisation of crude **S7**

3-3- Reaction of 2-amino-1,1-dimethyl-ethane-1-thiol hydrochloride with **S2** to give **S5**

(Non-optimized conditions)

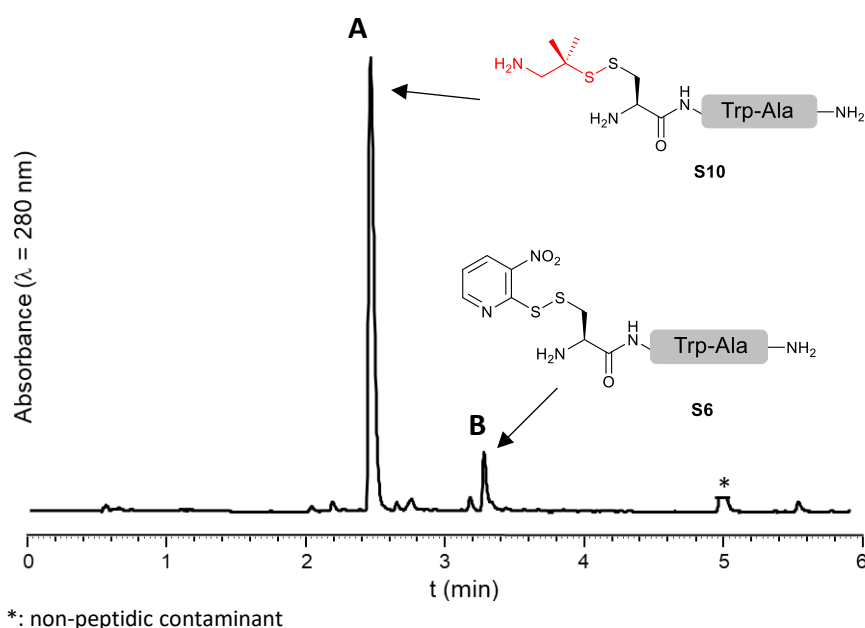


10 μmol of peptidyl resin **S3** was washed with 3 mL of NMP. A solution of 2-amino-1,1-dimethylpropane-1-thiol hydrochloride **2** (14 mg, 10 equiv.) in 1 mL of NMP (100 mM thiol concentration) was added to the resin. Then DIEA (35 μL , 20 equiv.) was added and the reactor was stirred for 5 min. The reactor was drained off and the resin was washed with DCM (3 \times 3 mL, 30 s) and NMP (3 \times 3 mL, 30 s). A small amount of the peptidyl resin was cleaved for analytical purpose affording **S10** (protocol PS5 small peptides cleavage).

- H-Cys(H-Ades)-Trp-Ala-NH₂ (**S10**):

ESI-MS (m/z): [M] calcd. for C₂₁H₃₂N₆O₃S₂: 480.2, found: 480.2

HPLC analysis: t_R = 2.49 min (Chromolith, gradient: 5-50% B/A over 5 min)



Peak (t _R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (2.49)	480.2	480.2	S10
B (3.31)	377.1	377.2	S6
* (5.00)	No MS signal		Non peptidic compound

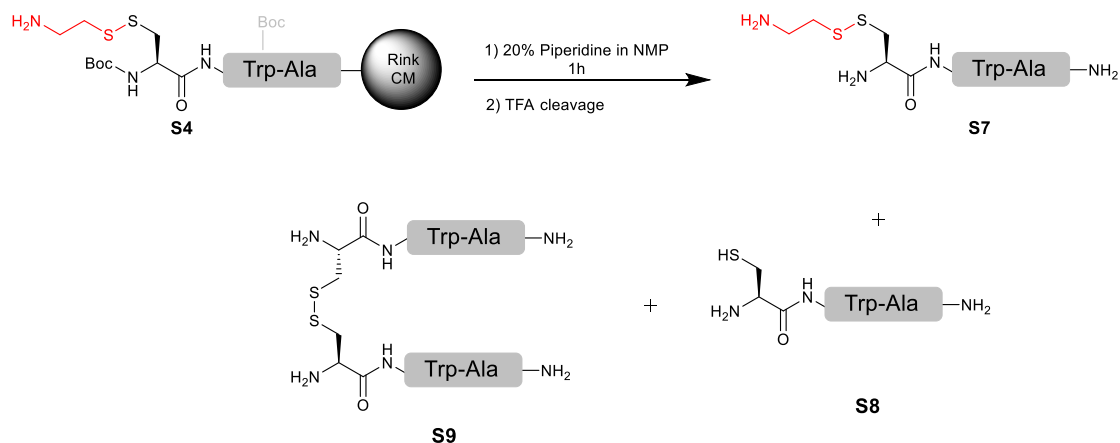
Supplementary figure S3: HPLC trace and MS characterisation of crude **S10** obtained under non-optimized conditions

3-4- Stability of disulfide-containing peptidyl resins **S4** and **S5** towards piperidine treatment

Resins were washed with 3 mL of NMP. A solution of 20 % v/v piperidine in NMP was added and the reactor was stirred for 1 h. The reactor was drained off and the resin was washed with DCM (3 \times 3 mL, 30 s) and NMP (3 \times 3 mL, 30 s). A small amount of the peptidyl resin was cleaved for analytical purpose (protocol PS5 small peptides cleavage).

- **Peptidyl resin S4**

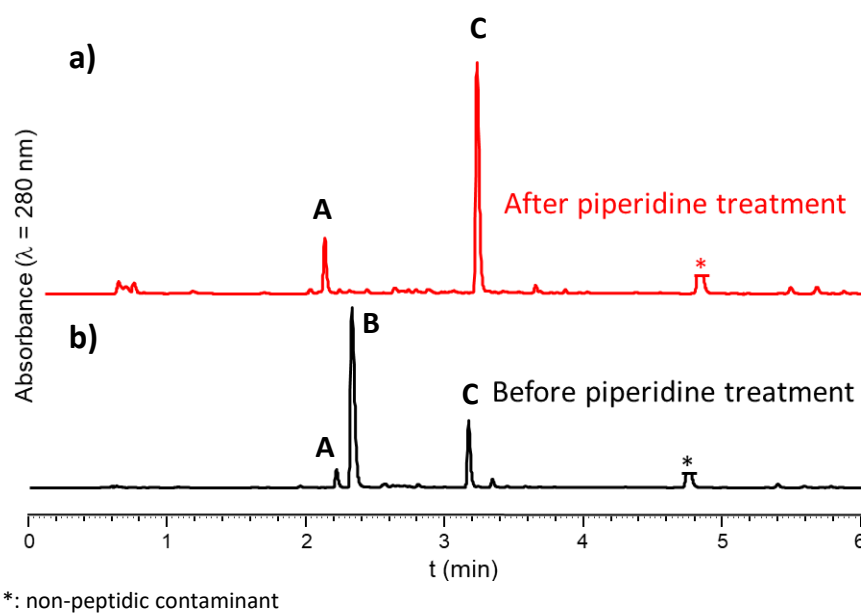
Complete degradation of **S4** was noticed upon treatment with piperidine, as assessed by TFA cleavage followed by analytical HPLC.



- Supplementary table S1: Rates of the products formed after piperidine treatment of peptidyl resin **S4** followed by TFA cleavage.

Entry	Piperidine treatment	S7 (%) ^a	S8 (%) ^a	S9 (%) ^a
1	No	79	10	11
2	Yes	-	58	42

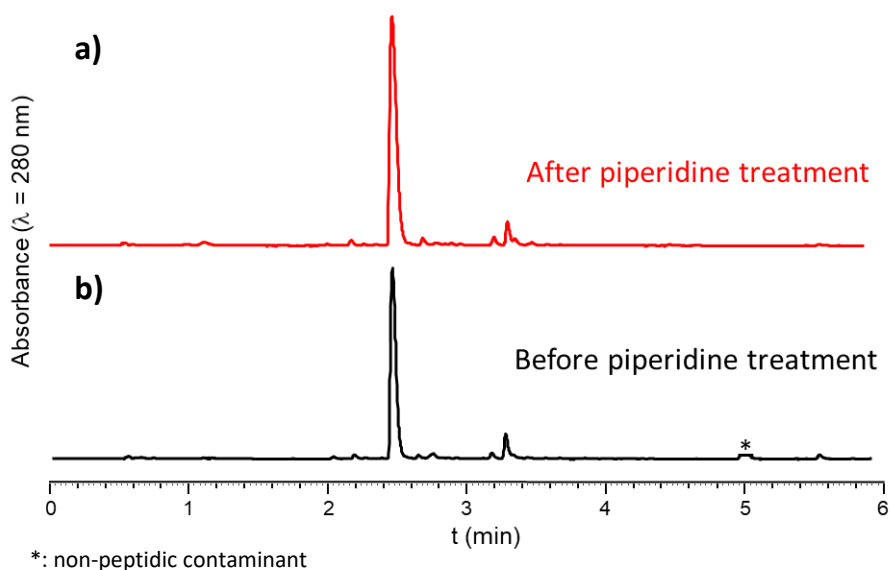
^a: Relative rates determined by HPLC peak integration at $\lambda = 280$ nm, taking into account the predicted molar absorption coefficient of Trp and disulfide at 280 nm: $\epsilon_{\text{Trp}} = 5500 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$; $\epsilon_{\text{SS}} = 125 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$. ¹



Supplementary figure S4: HPLC trace (Chromolith, gradient: 5-50% B/A over 5 min) of TFA-cleaved peptidyl resin **S4** with (a) and without (b) prior piperidine treatment

- **Peptidyl resin S5**

No degradation of **S5** was noticed upon treatment with piperidine, as assessed by TFA cleavage followed by analytical HPLC, showing no evolution of the major peak, corresponding to H-Cys(H-Ades)-Trp-Ala-NH₂ (**S10**).



Supplementary figure S5: HPLC trace (Chromolith, gradient: 5-50% B/A over 5 min) of TFA-cleaved peptidyl resin **S5** with (a) and without (b) prior piperidine treatment

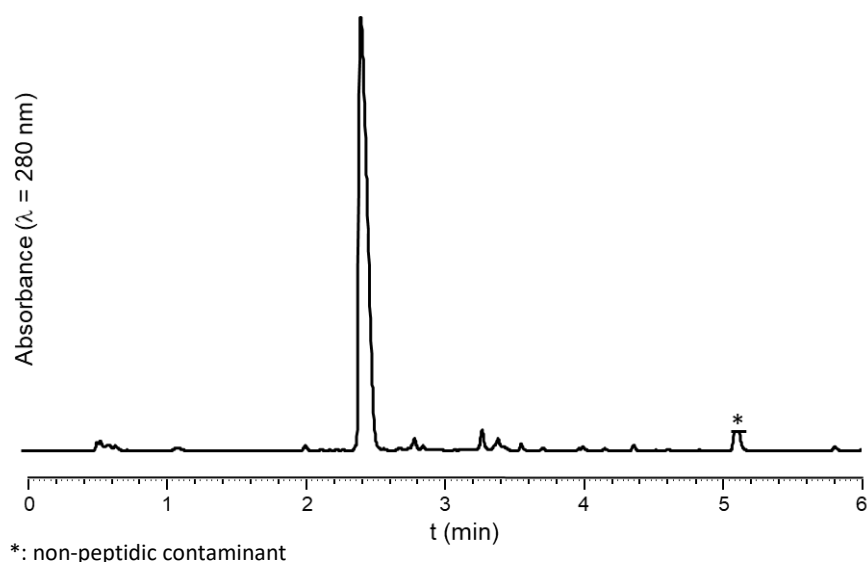
3-5- Optimization of the reaction of **2** with **S3**

- Table S2: Different conditions screened for the introduction of Ades

Entry	Equiv. 1	Base (20 equiv.)	Time (min)	S6 (%) ^a	S10 (%) ^a
1	10	DIEA	5	4	96
2	10	NMM	5	11	89
3	10	-	5	20	80
4	10	-	60	1	99
5	3	-	60	25	75
6	3 x 1.5	-	3 x 20	7	93

^a: Relative rates determined by HPLC peak integration at $\lambda = 280$ nm taking into account the predicted molar absorption coefficient of Trp, disulfide ¹ and S-Npys at 280 nm: $\epsilon_{\text{Trp}} = 5500 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$; $\epsilon_{\text{SS}} = 125 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$; $\epsilon_{\text{Npys}} = 1884 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$ (ϵ_{Npys} at 280 nm was determined from the molar absorption coefficient of 2,2'-dithiobis(5-nitropyridine) at 280 nm measured in 8:2:0.01 water/MeCN/TFA).

The use of 10 equiv. of 2-amino-1,1-dimethylethane-1-thiol hydrochloride **2** for 1 h without base (table S2, entry 4) gave satisfying results in terms of conversion and impurities rates.



Supplementary figure S6: HPLC trace (Chromolith, gradient: 5-50% B/A over 5 min) of crude **S10** obtained using optimized conditions (table S2, entry 4)

4- Synthesis and purification of MUC1 derived peptide segments

4-1- H-Trp-(MUC1)₂-(Hnb)Cys(StBu)-Gly-NH₂ crypto-thioester (**3a**)

Sequence: H-WAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS-(Hnb)C(StBu)-G-NH₂

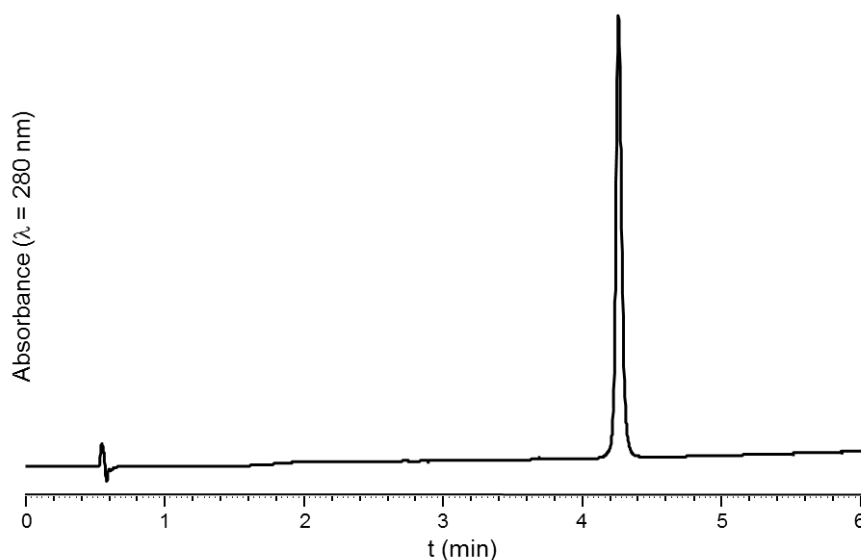
The peptide was synthesized following protocol PS1. Hnb was introduced following protocol PS2. Cleavage was performed following protocol PS4.

Elongation yield: 87%. Determined by the ratio between the quantity of fluorenylpiperidine released during final Fmoc deprotection (Protocol PS6) and the quantity released during the Fmoc deprotection of the C-terminal Gly residue (UV titration at 301 nm, $\epsilon = 7800 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$).

ESI-MS (*m/z*): [M] calcd. for C₁₈₇H₂₈₄N₅₆O₆₀S₂: 4340.7, found: 4340.2 (average mass, deconvoluted)

HPLC purification: [5mg/mL] Nucleosil C18, gradient: 22-32% B/A over 20 min affording a white solid after lyophilisation (21% yield).

HPLC analysis: $t_R = 4.32 \text{ min}$ (Chromolith, gradient: 5-50% B/A over 5 min)



Supplementary figure S7: HPLC trace of purified compound **3a**

4-2 H-Cys(K₆-Ades)-(MUC1)₂-Trp-NH₂ cysteinyl peptide (**7**)

Sequence: H-C(H-KKKKKK-Ades)PDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSW-NH₂

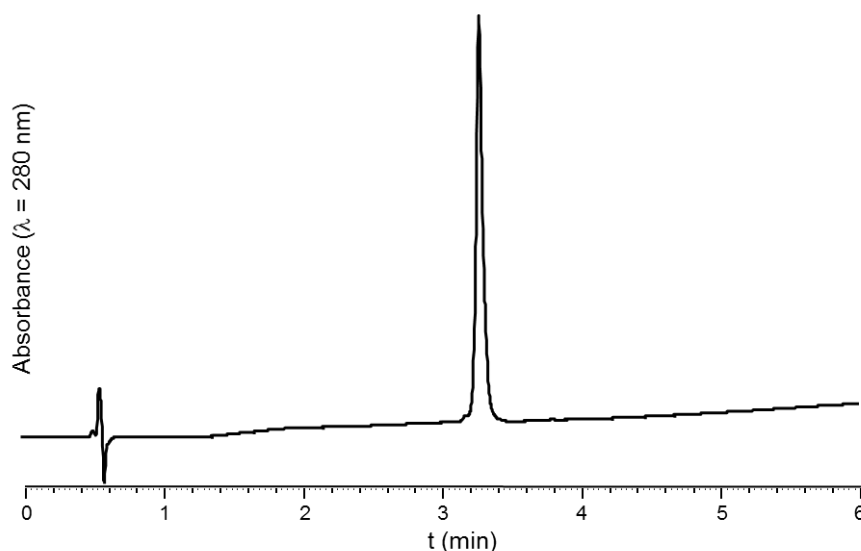
The peptide was synthesized following protocol PS1. Ades was introduced following protocol PS3. Cleavage was performed following protocol PS4.

Elongation yield: 86%. Determined by the ratio between the quantity of fluorenylpiperidine released during final Fmoc deprotection and the quantity released during the Fmoc deprotection of the C-terminal Trp residue (UV titration at 301 nm, $\epsilon = 7800 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$).

ESI-MS (m/z): [M] calcd. for C₂₁₁H₃₄₅N₆₆O₆₁S₂: 4846.5, found: 4845.5 (average mass, deconvoluted)

HPLC purification: [5mg/mL] Nucleosil C18, gradient: 12-22% B/A over 20 min, affording a white solid after lyophilisation (22% yield).

HPLC analysis: $t_R = 3.35 \text{ min}$ (Chromolith, gradient: 5-50% B/A over 5 min)



Supplementary figure S8: HPLC trace of purified compound **7**

4-3 H-Trp-(MUC1)₂-(Hnb)Cys(StBu)-Gly-(Lys)₆-NH₂ crypto-thioester (**3b**)

Sequence: H-WAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS-(Hnb)C(StBu)-GKKKKKK-NH₂

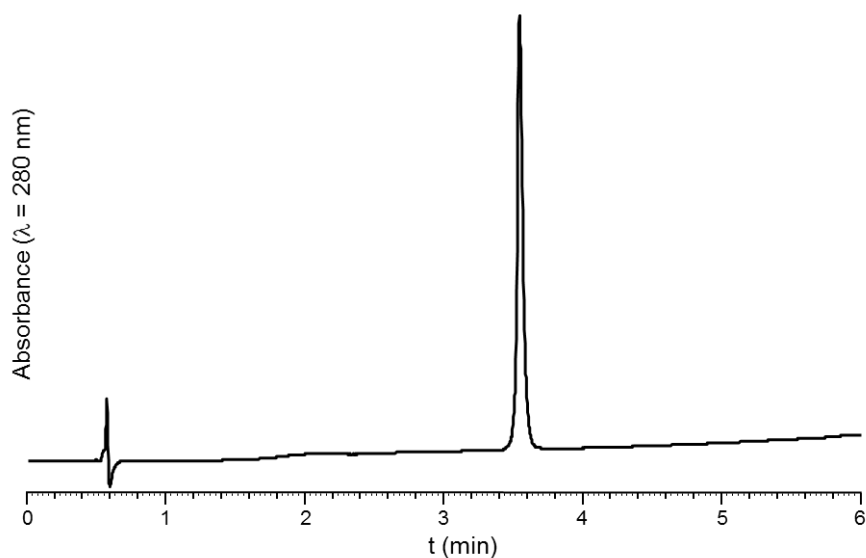
The peptide was synthesized using protocol PS1. Hnb was introduced using protocol PS2. Cleavage was performed following protocol PS4.

Elongation yield: 90%. Determined by the ratio between the quantity of fluorenylpiperidine released during final Fmoc deprotection (Protocol PS6) and the quantity released during the Fmoc deprotection of the C-terminal Lys residue (UV titration at 301 nm, $\epsilon = 7800 \text{ L.mol}^{-1}\text{.cm}^{-1}$).

ESI-MS (m/z): [M] calcd. for C₂₂₃H₃₅₇N₇₁O₆₆S₂: 5110.8, found: 5110.3 (average mass, deconvoluted)

HPLC purification: [5mg/mL] Nucleosil C18, gradient: 22-32% B/A over 20 min affording a white solid after lyophilisation (22% yield).

HPLC analysis: $t_R = 3.54 \text{ min}$ (Chromolith, gradient: 5-50% B/A over 5 min)



Supplementary figure S9: HPLC trace of purified compound **3b**

5- Synthesis and purification of SUMO-2 derived peptide segments

5-1 SUMO-2[1-47]-(Hnb)Cys(StBu)-Gly-NH₂ crypto-thioester (**10**)

Sequence: H-XADEKPKEGVKTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLXKAY-(Hnb)C(StBu)-G-NH₂
(X= Norleucine)

The peptide was synthesized using protocol PS1. Hnb was introduced using protocol PS2. D26 and G27 were introduced through the coupling of dipeptide Fmoc-Asp(OtBu)-(Dmb)Gly-OH to prevent aspartimide formation during SPPS (3.5 equiv. dipeptide, 3.4 equiv. HATU, 10 equiv. DIEA in NMP for 2 h). Cleavage was performed following protocol PS4.

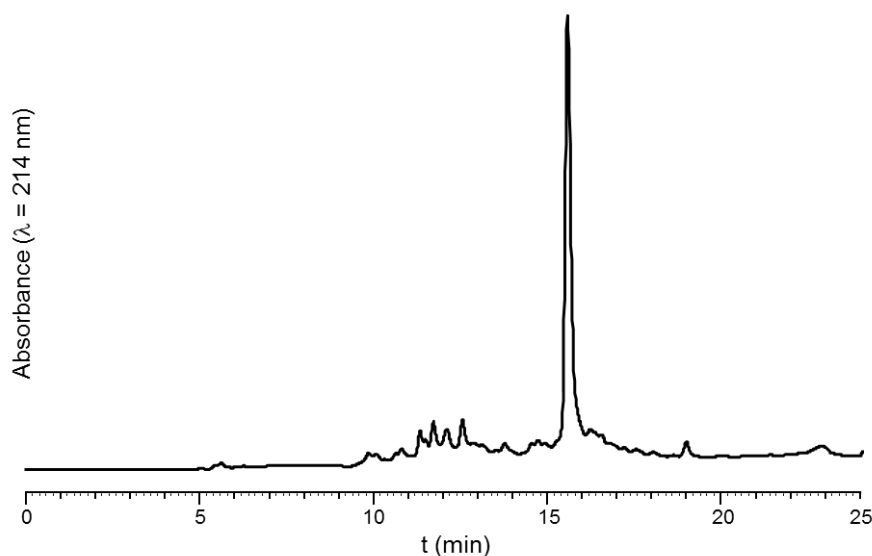
Elongation yield: 90%. Determined by the ratio between the quantity of fluorenylpiperidine released during final Fmoc deprotection (Protocol PS6) and the quantity released during the Fmoc deprotection of the C-terminal Gly residue (UV titration at 301 nm, $\epsilon = 7800 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$).

ESI-MS (m/z): [M] calcd. for C₂₅₀H₄₀₉N₇₁O₇₄S₂: 5657.5, found: 5657.6 (average mass, deconvoluted)

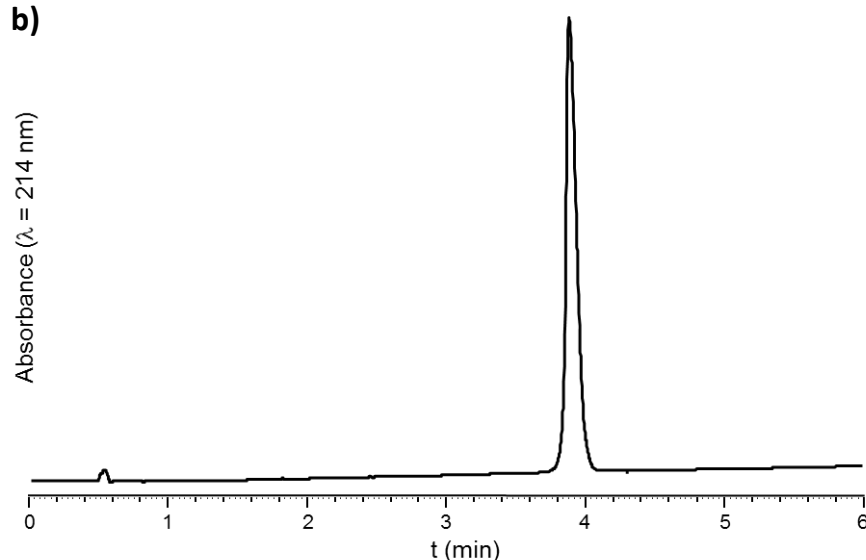
HPLC purification: [5mg/mL] Nucleosil C18, gradient: 25-40% B/A over 20 min affording a white solid after lyophilisation (37% yield).

HPLC analysis: $t_R = 3.88 \text{ min}$ (Chromolith, gradient: 5-50% B/A over 5 min)

a)



b)



Supplementary figure S10: HPLC traces of crude (a) and purified (b) compound **10** (HPLC conditions: Nucleosil C18, gradient: 5-50% B/A over 40 min, and Chromolith, gradient 5-50% B/A over 5 min, respectively)

5-2- SUMO-2[48-93] cysteinyl peptide (S11)

Sequence: H-CERQGLSXRQIRFRFDGQPINETDTPAQLEXEDEDTIDVFQQQTGG-OH (X= Norleucine)

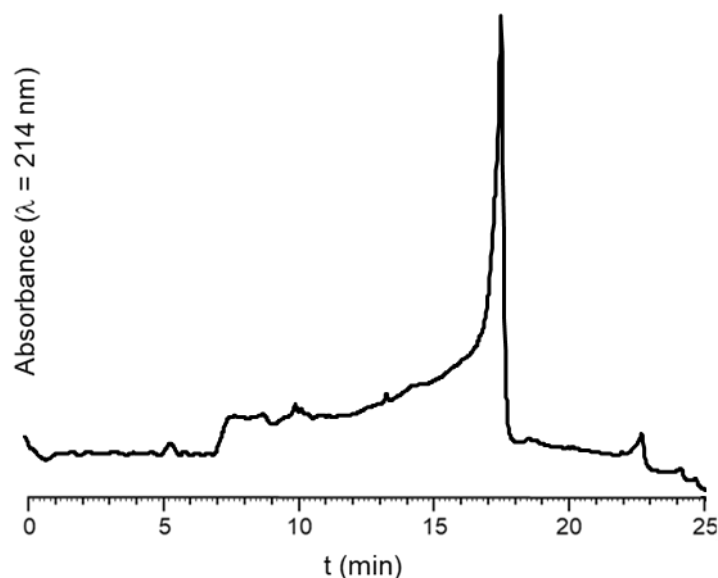
The peptide was synthesized following protocol PS1. D63 and G64 were introduced through the coupling of dipeptide Fmoc-Asp(OtBu)-(Dmb)Gly-OH to prevent aspartimide formation during SPPS (3.5 equiv. dipeptide, 3.4 equiv. HATU, 10 equiv. DIEA in NMP for 2 h). Q88 and Q89 were subjected to double couplings. Cleavage was performed following protocol PS4.

Elongation yield: 79%. Determined by the ratio between the quantity of fluorenylpiperidine released during final Fmoc deprotection and the quantity released during the Fmoc deprotection of the C-terminal Gly residue (UV titration at 301 nm, $\epsilon = 7800 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$).

Solubility: **S11** was not soluble at 5 mg/mL and 1 mg/mL in 8:2:0.01 water/MeCN/TFA, which are the conditions used to purify the other segments. We were able to solubilize crude peptide at very low concentrations (0.056 mg/mL and 0.1 mg/mL in 8:2:0.01 and 6:4:0.01 water/MeCN/TFA, respectively). However, we observed a non-conventional HPLC behaviour, and dragging of the product leading to an overlap of the impurities and the desired product, in accordance with the finding of Brik and co-workers.²

ESI-MS (*m/z*): [M] calcd. for C₂₂₅H₃₅₄N₆₆O₈₀S₁: 5295.7, found: 5295.9 (average mass, deconvoluted)

HPLC Analysis: t_R = 17.2 min (Jupiter C4, gradient: 10-90% B/A over 20 min)



Supplementary figure S11: HPLC trace of crude peptide **S11**

5-3- Cys(K₆-Ades)-SUMO-2[49-93] cysteinyl peptide (12)

Sequence: H-C(H-KKKKKK-Ades)ERQGLSXRQIRFRFDGQPINETDTPAQLEXEDEDITDVFQQQTGG-OH

(X= Norleucine)

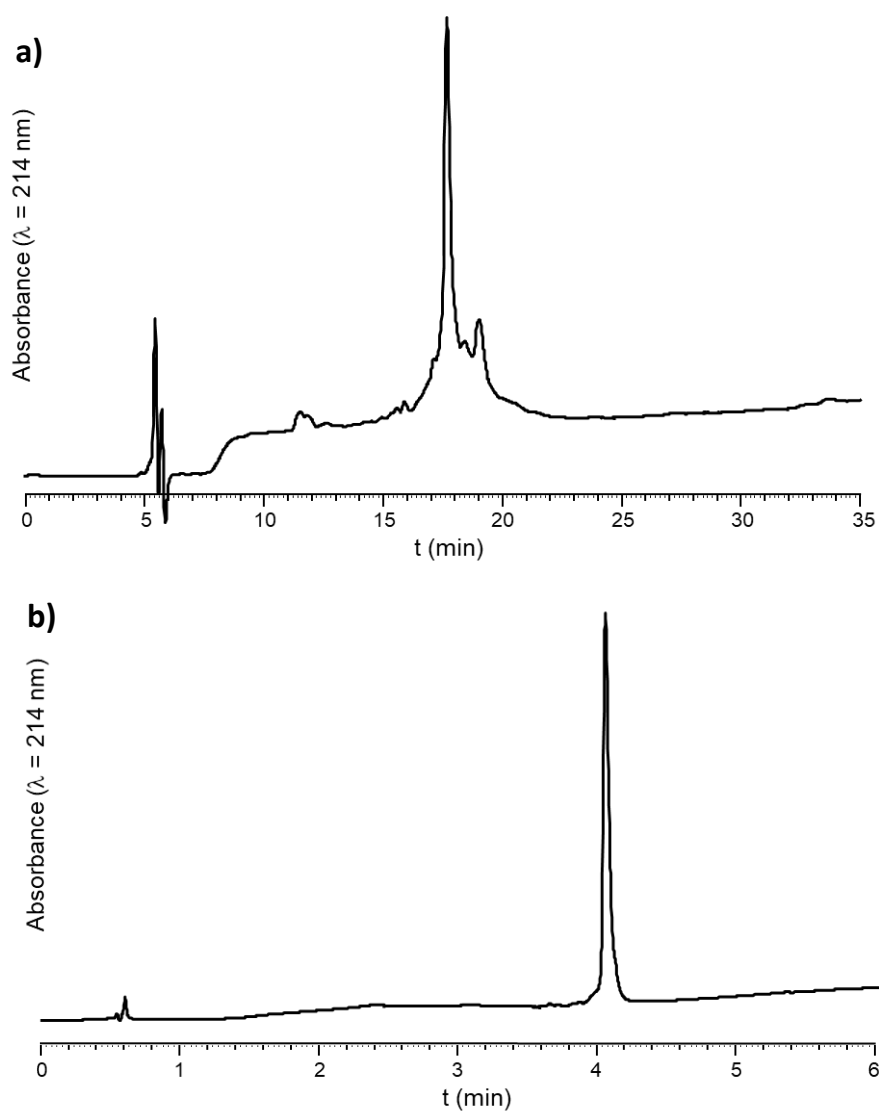
The peptide was synthesized using protocol PS1- Peptide elongation. Ades was introduced following protocol PS3. D63 and G64 were introduced through the coupling of dipeptide Fmoc-Asp(OtBu)-(Dmb)Gly-OH to prevent aspartimide formation during SPPS (3.5 equiv. dipeptide, 3.4 equiv. HATU, 10 equiv. DIEA in NMP for 2 h). Q88 and Q89 were subjected to double couplings. Cleavage was performed following protocol PS4. Boc-Cys(Ades) introduction was performed following the protocol PS3.

Elongation yield: 82%. Determined by the ratio between the quantity of fluorenylpiperidine released during final Fmoc deprotection and the quantity released during the Fmoc deprotection of the C-terminal Gly residue (UV titration at 301 nm, ε = 7800 L.mol⁻¹.cm⁻¹).

ESI-MS (*m/z*): [M] calcd. for C₂₆₅H₄₃₅N₇₉O₈₆S₂: 6167.9, found: 6166.8 (average mass, deconvoluted)

HPLC purification: [1 mg/mL] Jupiter C4, gradient: 7-42% B/A over 20 min affording a white solid after lyophilisation (14% yield).

HPLC analysis: t_R = 4.06 min (Chromolith, gradient: 5-50% B/A over 5 min)



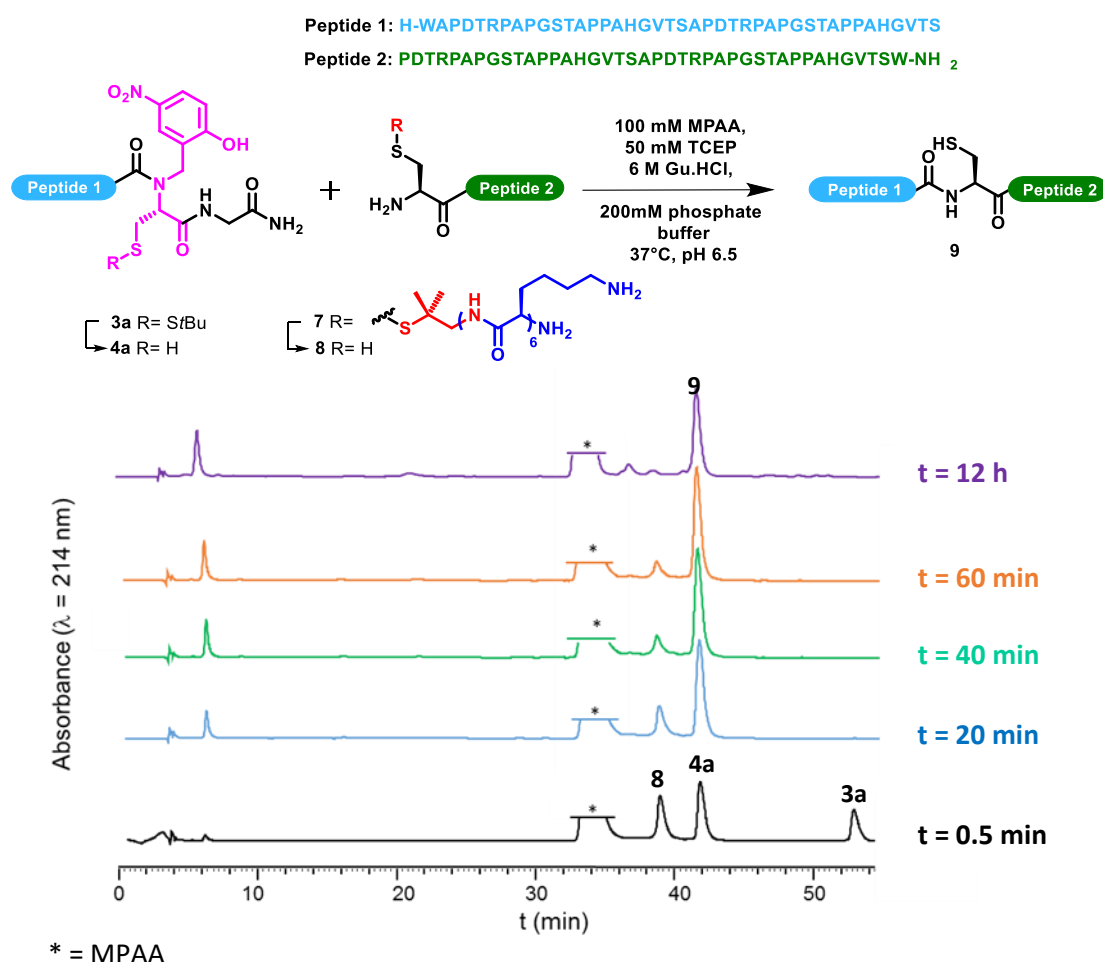
Supplementary figure S12: HPLC traces of crude (a) and purified (b) compound **12** (HPLC conditions: Nucleosil C18, gradient: 5-50% B/A over 40 min, and Chromolith, gradient 5-50% B/A over 5 min, respectively)

6- Native chemical ligations (NCL)

6-1 NCL between 3a and 7 to give 9

Crypto-thioester **3a** (3.0 mg, 0.6 μ mol, 2 mM final concentration) and cysteinyl peptide **7** (5.4 mg, 0.9 μ mol, 1.5 equiv., 3 mM final concentration) were dissolved in NCL buffer (300 μ L) and the resulting yellow solution was gently stirred under inert atmosphere at 37°C for 12 h.

Note that ligation product **9** coeluted with *S*-deprotected cryptothioester **4a**. LC-MS analysis indicated total consumption of **4a** after 12 h reaction.



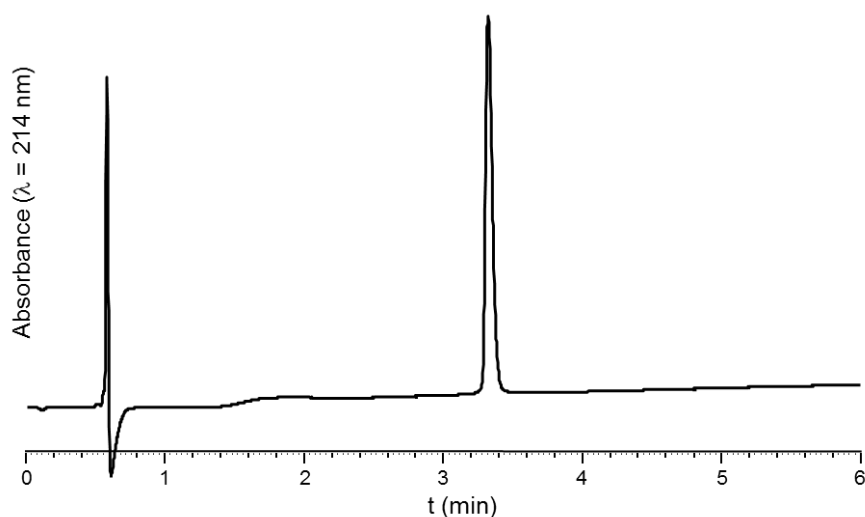
Supplementary figure S13: HPLC (Nucleosil C18, gradient: 20-50% B/A over 60 min) monitoring of ligation between compounds **3a** and **7**

- Ligation product (**9**)

HPLC purification: [5mg/mL] Jupiter C4, gradient: 10-24% B/A over 20 min affording a white solid after lyophilisation (62% yield)

ESI-MS (*m/z*): [*M*] calcd. for C₃₄₂H₅₂₃N₁₀₅O₁₁₀S₁: 7897.5, found: 7896.7 (average mass, deconvoluted)

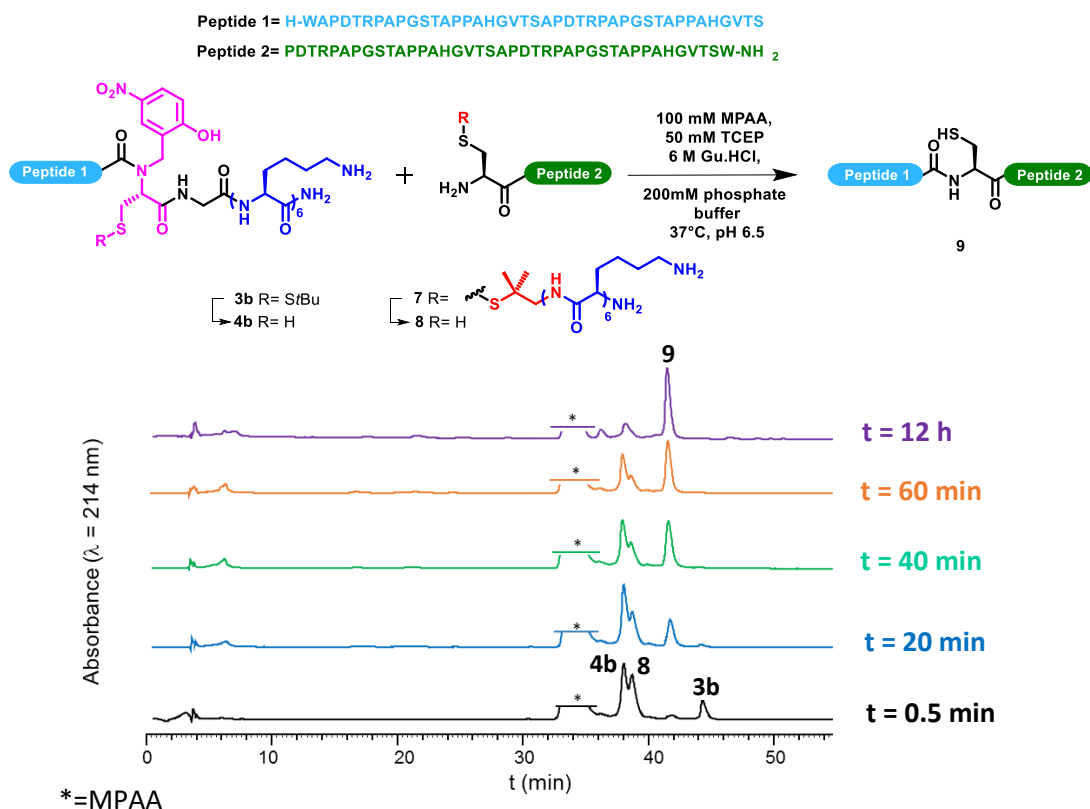
HPLC analysis: *t_R* = 3.38 min (Chromolith, gradient: 5-50% B/A over 5 min)



Supplementary figure S14: HPLC trace of purified ligation product **9**

6-2- NCL between **3b** and **7** to give **9**

Crypto-thioester **3b** (3.0 mg, 0.4 μ mol, 2 mM final concentration) and cysteinyl peptide **7** (3.6 mg, 0.6 μ mol, 1.5 equiv., 3 mM final concentration) were dissolved in NCL buffer (200 μ L) and the resulting yellow solution was gently stirred under inert atmosphere at 37°C for 12 h.



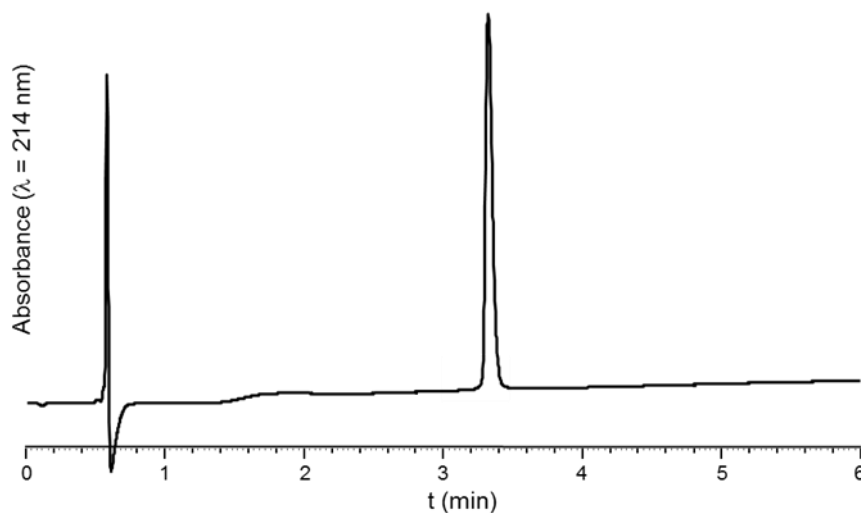
Supplementary figure S15: HPLC (Nucleosil C18, gradient: 20-50% B/A over 60 min) monitoring of ligation between compounds **3b** and **7**

- Ligation product (**9**)

HPLC purification: [5mg/mL] Jupiter C4, gradient: 10-24% B/A over 20 min affording a white solid after lyophilisation (76% yield)

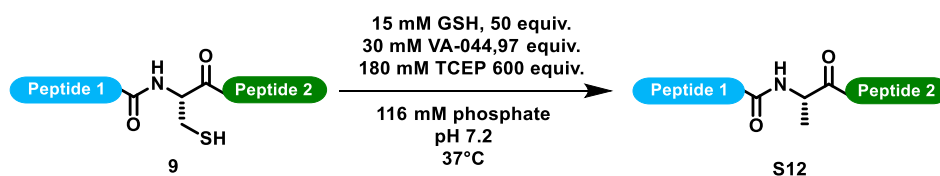
ESI-MS (*m/z*): [M] calcd. for C₃₄₂H₅₂₃N₁₀₅O₁₁₀S₁: 7897.5, found: 7896.7 (average mass, deconvoluted)

HPLC analysis: *t_R* = 3.38 min (Chromolith, gradient: 5-50% B/A over 5 min)



Supplementary figure S16: HPLC trace of ligation product **9** between compounds **3b** and **7**

6-3- Desulfurization of **9**

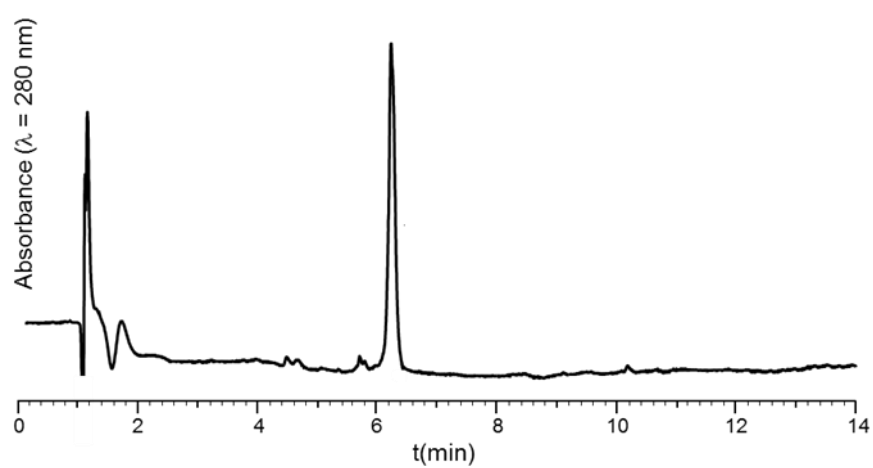


Cysteine-containing peptide **9** (0.55 mg, 61 nmol) was desulfurized by incubation for 12 h at 37°C in a 116 mM phosphate buffer (203 µL, 0.3 mM peptide) containing 15 mM reduced glutathione (GSH), 30 mM 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044) and 180 mM TCEP, pH = 7.2. LC-MS monitoring showed completion of the reaction. The crude product **S12** was not purified.

- Desulfurized product (**S12**)

ESI-MS (*m/z*): [M] calcd. for C₃₄₂H₅₂₃N₁₀₅O₁₁₀: 7865.5 average, found: 7865.7 (average mass, deconvoluted)

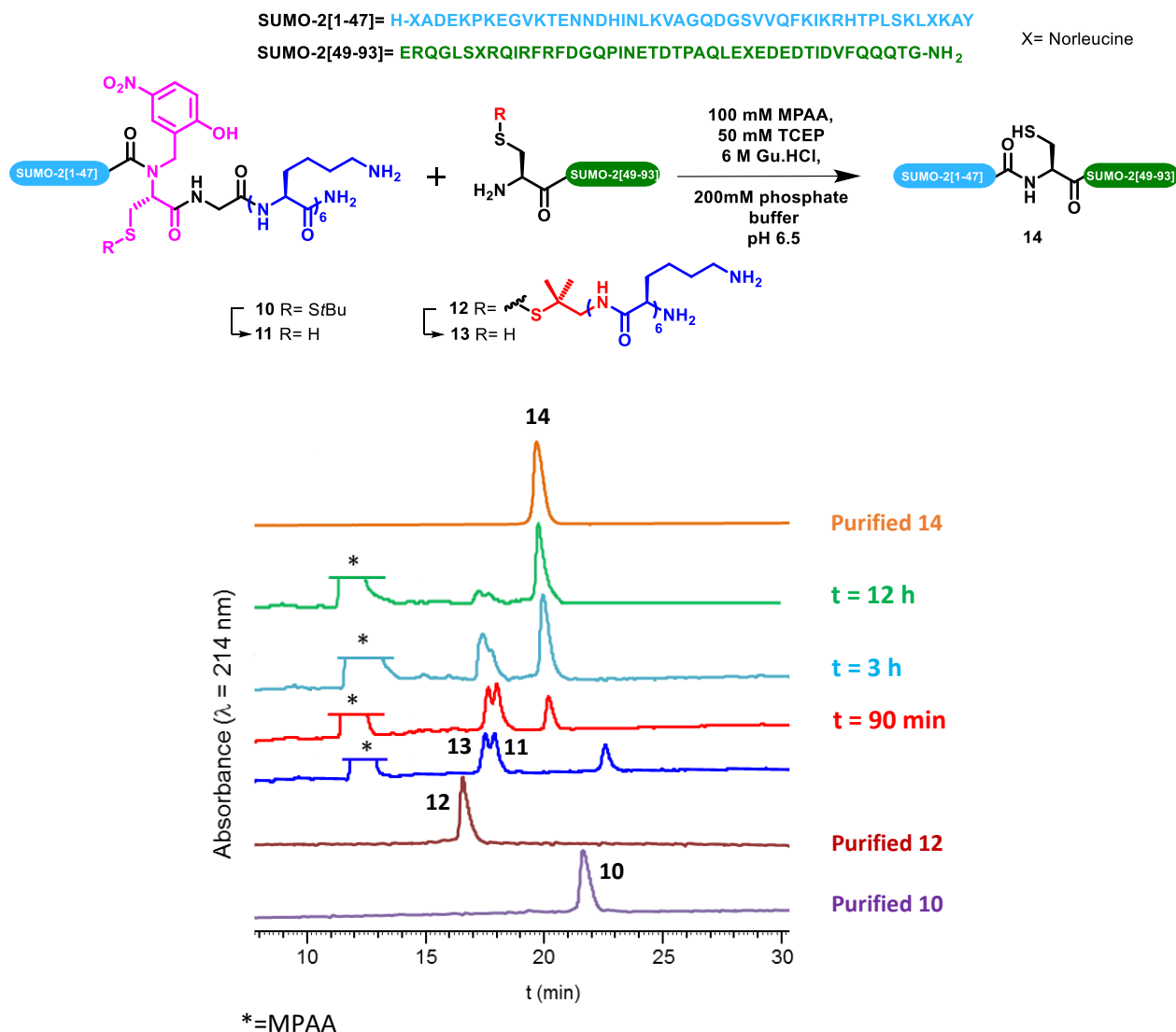
HPLC analysis: *t_R* = 6.14min (Aeris Widepore XB-C18 2, gradient: 3-90 % B'/A' over 15 min)



Supplementary figure S17: HPLC trace of crude desulfurized product **S12**

6-4 NCL between **10** and **12** to give SUMO-2[1-93] (**14**)

Crypto-thioester **10** (3.5 mg, 0.44 μ mol, 2 mM final concentration) and cysteinyl peptide **12** (4.97 mg, 0.66 μ mol, 1.5 equiv., 3 mM final concentration) were dissolved in NCL buffer (220 μ L) and the resulting yellow solution was gently stirred under inert atmosphere at 25°C for 12 h.



Supplementary figure S18: HPLC monitoring (Nucleosil C18, gradient: 20-50% B/A over 50 min) of ligation between compounds **10** and **12**

In accordance with the finding of Melnyk and co-workers³ during the NCL-based synthesis of a SUMO-2 derivative using *bis*(2-sulfenylethyl)amido (SEA) crypto-thioesters,⁴ a by-product **S13** which mass is consistent with aspartimide formation at one of the two Asp-Gly sites (-18 Da in MS analysis) was formed when conducting the reaction at 37°C for 12 h. The by-product **S13** co-eluted with SUMO-2 **14** in HPLC and LC-MS, and the relative rate (18% aspartimide) was quantified by integration of the extracted ion chromatograms corresponding to the expected product and aspartimide ($[M+12H]^{12+}$ peak, the most intense one of the multicharged envelopes) of the crude mixture. Similarly to Melnyk's findings, when the reaction was performed at 25°C for 12 h, no significant amount of the aspartimide by-product was formed (<5%).

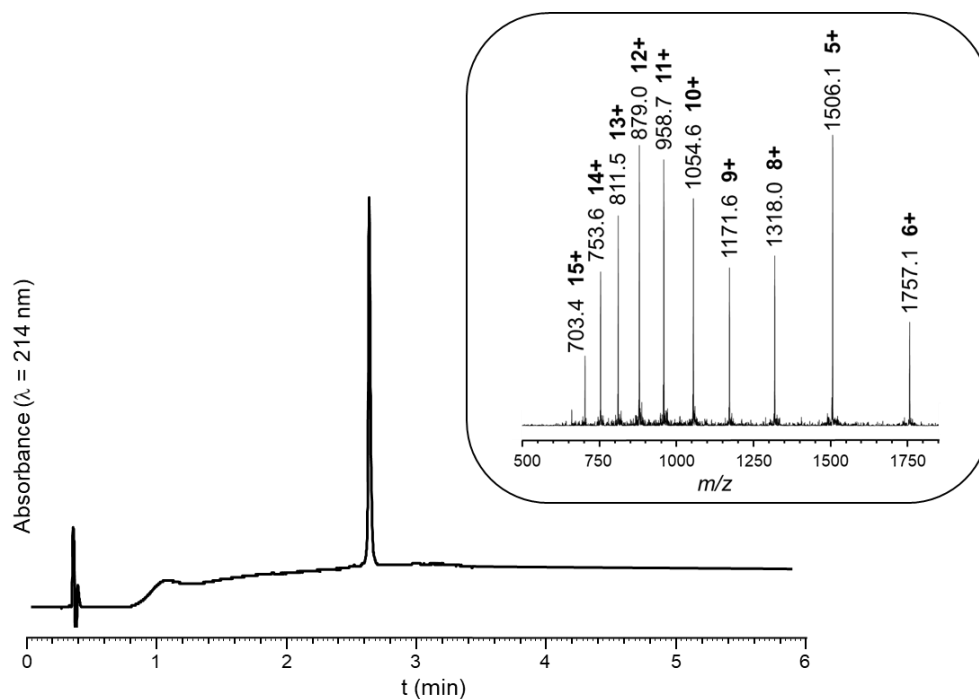
- Ligation product SUMO-2[1-93] (**14**)

HPLC purification: [2mg/mL] Jupiter C4, gradient: 20-90% B/A over 40 min affording a white solid after lyophilisation (40% yield)

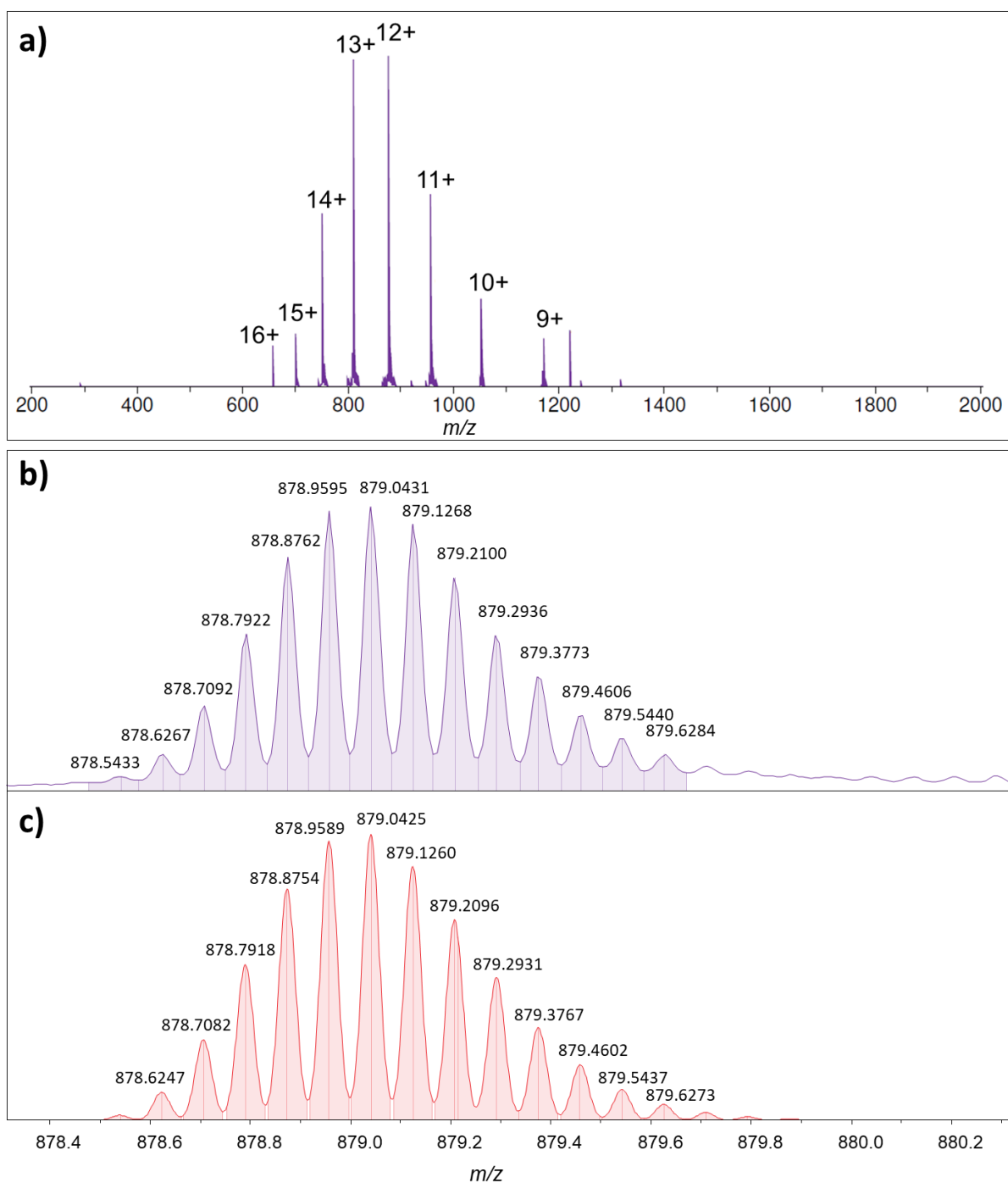
ESI-MS (m/z): [M] calcd. for $C_{459}H_{739}N_{133}O_{149}S$: 10536.6, found: 10535.4 (average mass, deconvoluted)

ESI-HRMS (m/z): [M] calcd. for $C_{459}H_{739}N_{133}O_{149}S$: 10530.4059, found: 10530.4060 (deconvoluted)

HPLC analysis: t_R = 2.61 min (Chromolith, gradient: 20-50% B/A over 5 min)



Supplementary figure S19: HPLC trace and ESI mass spectrum of purified SUMO-2[1-93] **14**



Supplementary figure S20: a) High resolution mass spectrum of purified SUMO-2[1-93] **14**; b) zoom on the $[M+12H]^{12+}$ peak; c) theoretical isotopic distribution of the $[M+12H]^{12+}$ peak

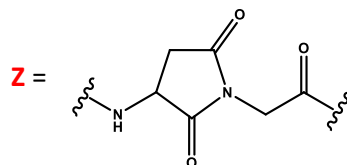
- Aspartimide by-product **S13**

Two possible sequences (since there are two aspartimide prone DG sites):

H-XADEKPKEGVKTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLXKAYCERQGLSXRQIRFRF**Z**QPINETDTPAQL
XEDEDTIDVFQQQTGG-OH

H-XADEKPKEGVKTENNDHINLKVAGQ**Z**SVVQFKIKRHTPLSKLXKAYCERQGLSXRQIRFRFDGQPINETDTPAQL
XEDEDTIDVFQQQTGG-OH

X = norleucine



Note that we did not try to confirm the aspartimide structure nor to identify which of the two DG sequence could be more prompt for aspartimide formation during NCL.

ESI-MS (*m/z*): [M] calcd. for C₄₅₉H₇₃₇N₁₃₃O₁₄₈S: 10518.6, found: 10516.7 (average mass, deconvoluted)

HPLC analysis: t_R = 2.61 min (Chromolith, gradient: 20-50% B/A over 5 min)

7- Functional and structural characterization of SUMO-2[1-93] (**14**)

7-1- Circular dichroism

A 11.9 μM solution of SUMO-2 (**14**) in 10 mM sodium phosphate buffer (pH 7.2) containing 50 mM sodium fluoride was prepared and incubated at 25 °C for 12 h. Peptide concentration was determined by UV spectrophotometry at $\lambda = 280 \text{ nm}$.¹

The circular dichroism spectra were recorded on a Jasco 810 spectropolarimeter (J-810) at 25 °C over the range 190-260 nm using 0.1 cm path-length cell and by averaging 20 scans.

A 2 nm bandwidth, 1 nm data pitch were used for spectral acquisition and 2 s integration time.

The results are presented as $[\theta]_{\text{MRW}}$ against the wavelength between 190 and 260 nm.⁵

$$[\theta]_{\text{MRW}} = \frac{M \times \theta_{\text{obs}}}{(N-1) \times 10 \times d \times c}$$

with:

$[\theta]_{\text{MRW}}$ = mean residue ellipticity ($\text{deg.cm}^2.\text{dmol}^{-1}$)

M = molecular weight (g.mol^{-1})

θ_{obs} = observed ellipticity (mdeg)

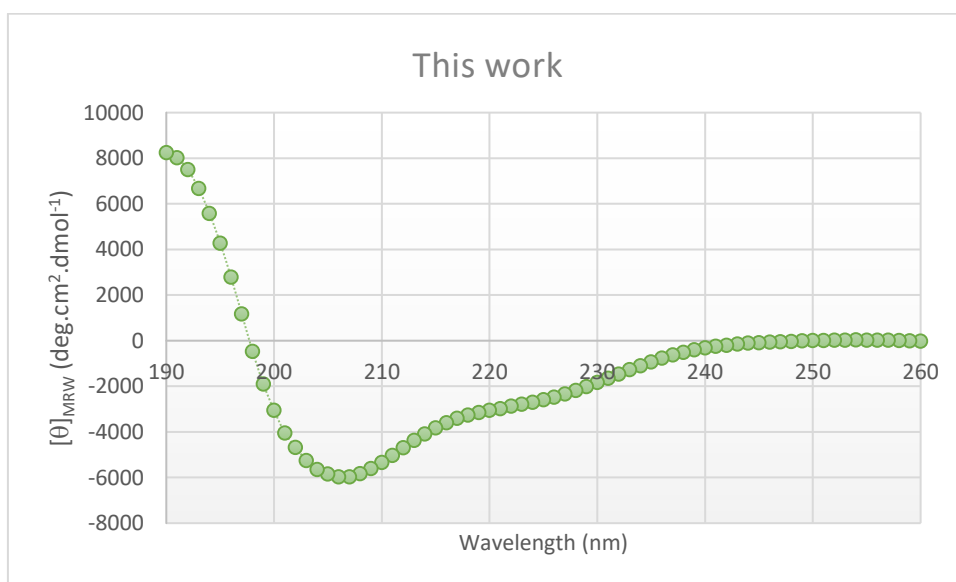
N = number of amino acids

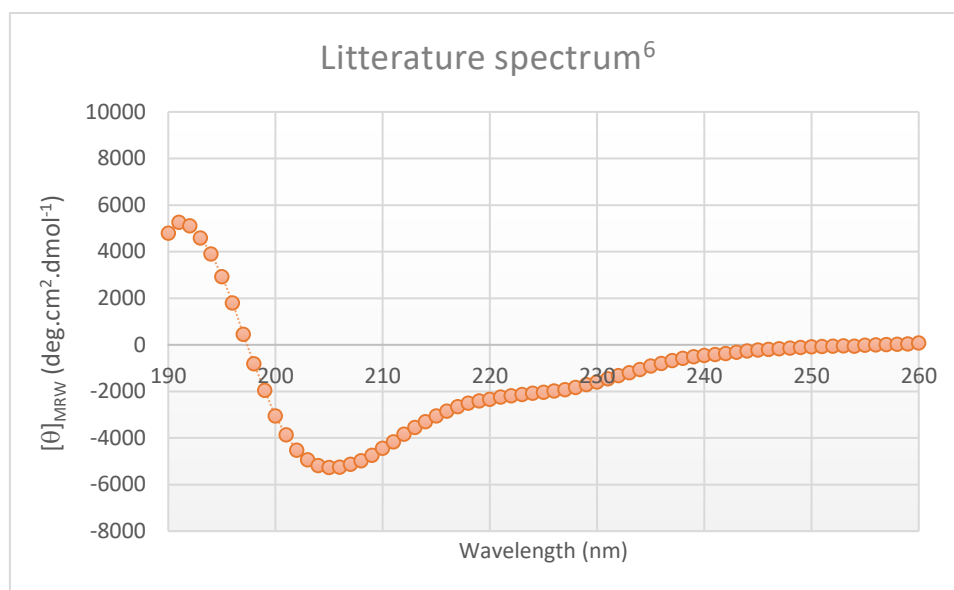
d = pathlength (cm)

c = concentration (g.L^{-1})

The far-UV CD spectrum of **14** was compared to the one reported by Melnyk and his co-workers,⁶ who analyzed [2-93]SUMO-2 at a $\sim 10\text{-}22 \mu\text{M}$ concentration.

Spectra were almost identical in the 200-260 nm range.





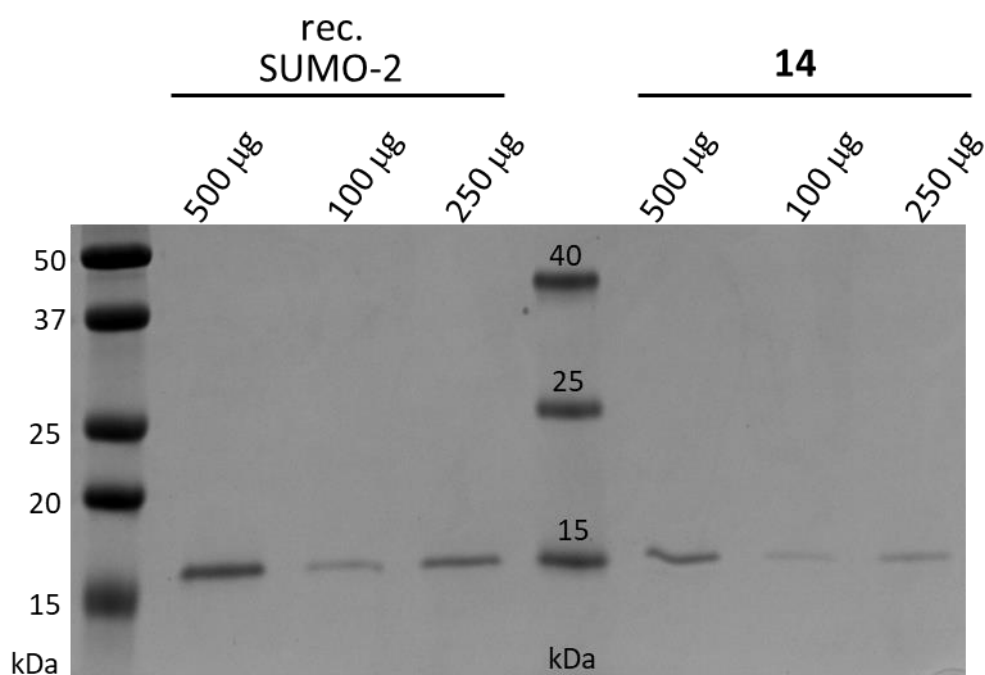
Supplementary figure S21: CD spectrum of **14** compared to a literature spectrum

7-2- Enzymatic conjugation

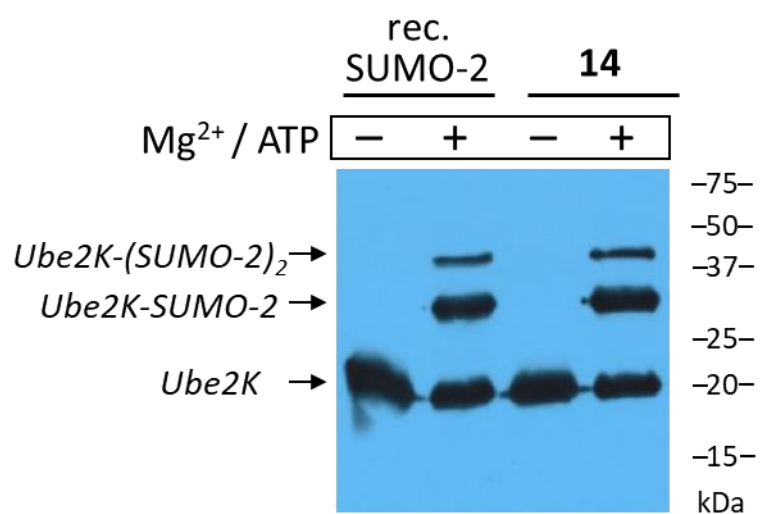
To perform enzymatic conjugation of synthetic SUMO-2, we used a SUMO-2 conjugation kit (Ref # K-715) developed by BostonBiochem containing SUMO E1 enzyme (SAE1/SAE2), SUMO E2 (Ubc9), recombinant SUMO-2, reaction buffer and a Mg^{2+} /ATP mix and we purchased the Ube2K (E2-25K) SUMOylation target from R & D System (Ref SP-200).

A solution of synthetic SUMO-2 (**14**) was prepared at the same concentration of recombinant SUMO-2 (250 μ M) and in the same buffer (50 mM HEPES, 150 mM NaCl, 1 mM DTT, pH 7.5). SUMO conjugation reactions in the presence of synthetic or recombinant SUMO-2 were then performed according to the kit recommendations for 60 min. Mg^{2+} /ATP solution was omitted to obtain negative controls. Reaction was stopped by the addition of reducing SDS-PAGE sample buffer.

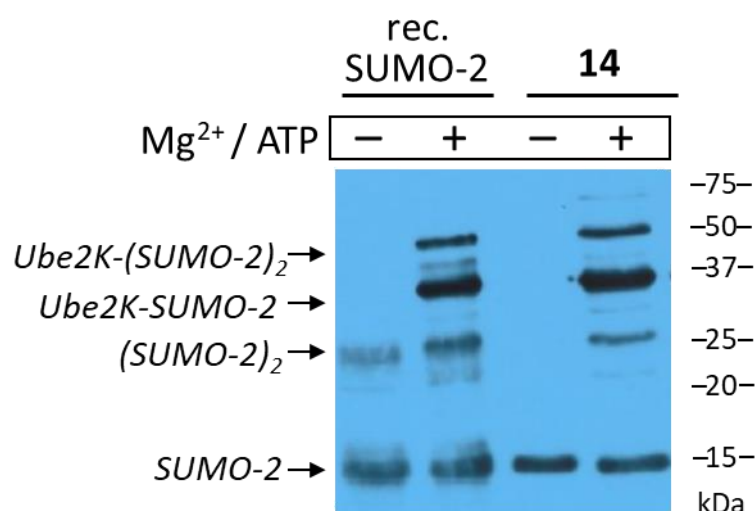
SUMO-2-conjugation of Ube2K was then analysed by western-blot (SDS-PAGE 15%) after heating the sample for 5 min at 90 °C. For this purpose, primary antibodies directed against SUMO-2/3 (Abcam, Ab1371) and Ube2K (Novus Biologicals, MAB6609) and secondary anti-mouse antibodies coupled to HRP (Invitrogen, Ref 61-6520) were used. The chemiluminescent HRP substrate Super signal west dura (Thermo Scientific, Ref 34075) allowed to reveal the corresponding proteins on a CL-Xposure film (Thermo Scientific, Ref 34090).



Supplementary figure S22: SDS-PAGE analysis of **14** and a recombinant SUMO-2 (coomassie blue staining)

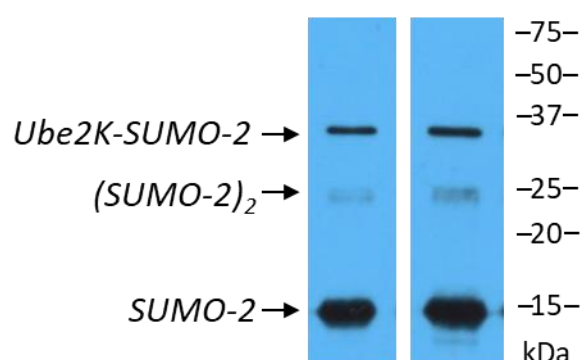


Supplementary figure S23: Anti-Ube2K western-blot of enzymatic SUMOylation of Ube2K by **14** and a recombinant SUMO-2



Supplementary figure S24: Anti-SUMO-2/3 western-blot of enzymatic SUMOylation of Ube2K by **14** and a recombinant SUMO-2

Another sample of synthetic SUMO-2 **14** was also denaturated in 6 M guanidium chloride and refolded⁷ by diluting it ten times in reaction buffer (50 mM HEPES, 150 mM NaCl, 1 mM DTT, pH 7.5) such to obtain a final concentration of 250 μ M before being tested as previously described in the enzymatic conjugation reaction. No difference was observed in the conjugation test between the sample simply diluted into buffer and the one subjected to this refolding protocol. Note that a different batch of conjugating kit was used for this experiment, explaining the lower Ube2K SUMOylation rates as compared to figure S24.



Supplementary figure S25: Anti-SUMO-2/3 western-blot of enzymatic SUMOylation of Ube2K by **14** either simply dissolved in buffer (left) or subjected to guanidinium chloride-mediated denaturation followed by refolding (right)

References:

- 1 C. N. Pace, F. Vajdos, L. Fee, G. Grimsley and T. Gray, How to measure and predict the molar absorption coefficient of a protein, *Protein Sci.*, 1995, **4**, 2411–2423.
- 2 S. Bondalapati, E. Eid, S. M. Mali, C. Wolberger and A. Brik, Total chemical synthesis of SUMO-2-Lys63-linked diubiquitin hybrid chains assisted by removable solubilizing tags, *Chem. Sci.*, 2017, **8**, 4027–4034.
- 3 J. Bouchenna, M. Sénéchal, H. Drobecq, J. Vicogne and O. Melnyk, The Problem of Aspartimide Formation During Protein Chemical Synthesis Using SEA-Mediated Ligation, 2019, In:
Iranzo O., Roque A. (eds) Peptide and Protein Engineering. Springer Protocols Handbooks. Humana, New York, NY.
- 4 N. Ollivier, J. Dheur, R. Mhidia, A. Blanpain and O. Melnyk, Bis(2-sulfanylethyl)amino Native Peptide Ligation, *Org. Lett.*, 2010, **12**, 5238–5241.
- 5 N. Greenfield and G. D. Fasman, Computed circular dichroism spectra for the evaluation of protein conformation, *Biochemistry*, 1969, **8**, 4108–4116.
- 6 M. Sénéchal, J. Bouchenna, J. Vicogne and O. Melnyk, Comment on “N-terminal Protein Tail Acts as Aggregation Protective Entropic Bristles: The SUMO Case”, *Biomacromolecules*, 2020, **21**, 3480–3482.
- 7 For a recent study on SUMO-2 refolding after denaturation, see: T. Nandi, A. Yadav, S. R. Koti Ainavarapu, Experimental comparison of energy landscape features of ubiquitin family proteins, *Proteins*, 2019, **88**, 449–461.

Chapter 5: Synthesis of a SUMO proteases-resistant SUMOylated peptide through solid-supported orthogonal chemical ligations

I- Bibliographic study	363
1) Triazolo peptides as peptidomimetics	363
A. The peptide bond	363
B. Amide bond mimics	364
C. 1,4-disubstituted 1,2,3-triazoles as <i>trans</i> amide bond mimics	364
D. Peptidomimetic triazole ligation (PTL)	366
E. Triazoles as proteases-resistant amide bond mimics	366
2) Synthesis and semi-synthesis of SUMOylated peptides and proteins	371
A. SPPS based synthesis of SUMO proteins	373
B. NCL-based synthesis of SUMOylated peptides and proteins	373
C. KAHA ligation-based synthesis of SUMO proteins and SUMOylated peptides	374
D. Semi-synthesis of SUMOylated substrates	376
II- Aim of this work: towards chemical biology tools to decipher neurofibromin SUMOylation	377
1) Neurofibromatosis and SUMOylation	377
2) Synthetic strategy	378
III- Results and discussion	379
1) Synthesis of the three peptide segments	379
2) SPCL-mediated target assembly	382
IV- Experimental part	387
1) General information	387
2) General procedures	387
3) Synthesis of the different segments needed for SPCL	389
A. SUMO-2 [1-46] crypto-thioester (1)	389
B. SUMO-2 [47-93]-Propargylamide (S6)	391
C. K ₆ Ades-SUMO-2 [47-93]-propargylamide (7)	392
D. Model azido-tetrapeptide (9)	394
E. PML derived peptide (S7)	395
4) SPCL-based synthesis of (15)	397
V- References	402

I- Bibliographic study

Note that this bibliographic study introduction is separated in two essentially independent parts, the first one concerns the use of 1,4-disubstituted-1,2,3-triazoles as amide bond mimics in peptides and proteins, while the second one is focused on the synthesis and semi-synthesis of SUMO proteins and derivatives.

1) Triazolopeptides as peptidomimetics

A. The peptide bond

Peptides and proteins are linear polymers formed by linking the α -carboxyl group of one amino acid to the α -amino group of another through an amide bond, also referred to as peptide bond. The amide bond has a partially double-bond character resulting from resonance between mesomeric forms, keeping its four atoms planar (fig. 1). Two rotamers, *cis* or *trans* can coexist in equilibrium. However *N*-monosubstituted amide bond in non-constrained peptides usually exhibit > 99.9% *trans* configuration due to a steric clash between side chains in the *cis* form, an exception being the *N,N*-disubstituted Xaa-Pro bond, where both rotamers are present. Another consequence of resonance is the high polarity of the amide bond: it exhibits a strong dipolar moment, the nitrogen atom is electropositive, and the oxygen atom is electronegative. This also leads to very strong hydrogen bonding properties: acceptor for the carbonyl oxygen, and donor for the NH proton.

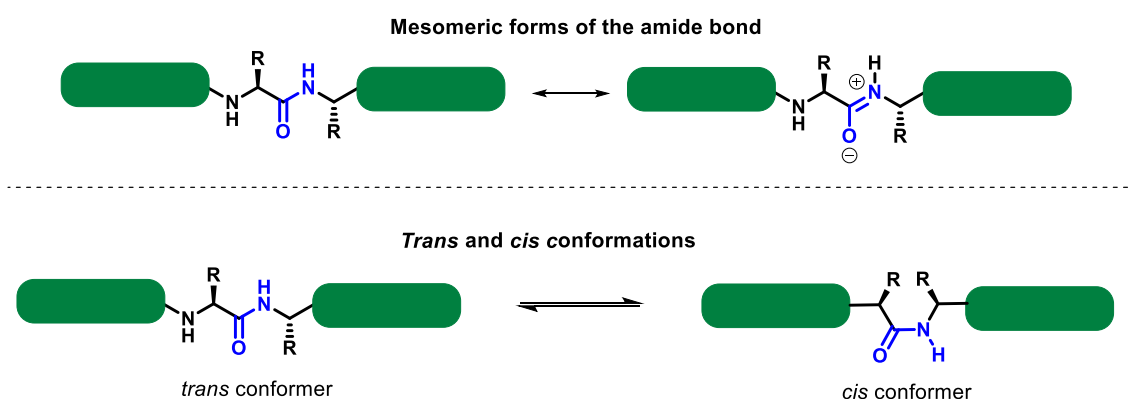


Figure 1: The amide bond in peptides and proteins

B. Amide bond mimics

Although extremely stable towards chemical hydrolysis, peptide bonds can be very unstable *in vivo* due to enzymatic hydrolysis by proteases.[1] This vulnerability has motivated chemists to develop amide bond surrogates, to generate peptidomimetic compounds able to retain the conformational and electrostatic properties of the native peptide and to interact with its biological receptor. Therefore, many synthetic amide bond mimics capable of reproducing some of its characteristics have been reported (fig. 2), [2–13] including 1,4-disubstituted 1,2,3-triazoles on which the first part of this bibliographic study will concentrate.

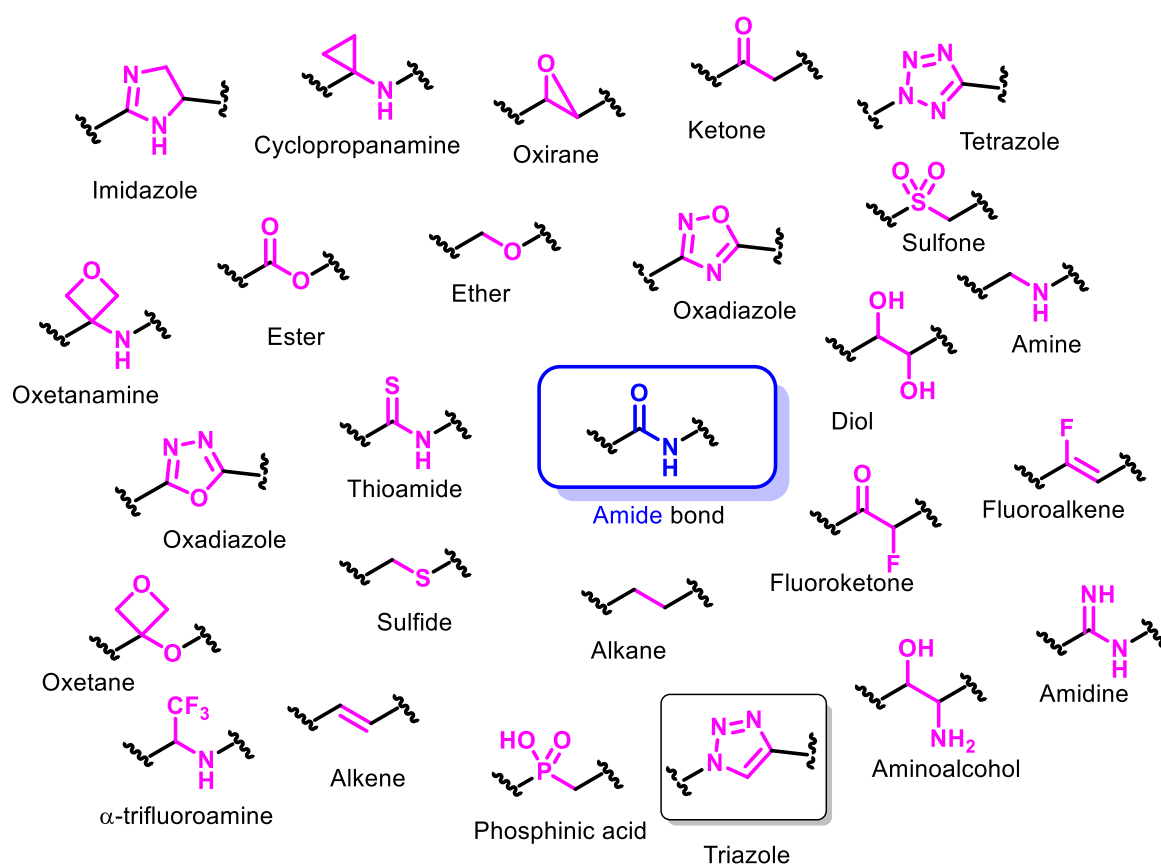


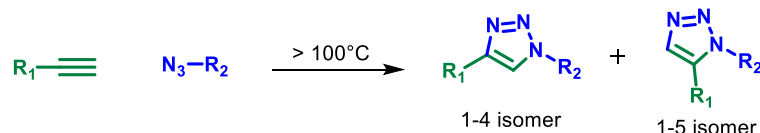
Figure 2: Examples of amide bond mimics

C. 1,4-disubstituted 1,2,3-triazoles as *trans* amide bond mimics

1,4-Disubstituted-1,2,3-triazoles can be conveniently obtained *via* the copper (I)-catalyzed azide/alkyne cycloaddition (CuAAC) reaction independently introduced by Sharpless [14] and Meldal. [15] Contrasting with the non-catalyzed Huisgen 1,3-dipolar cycloaddition, [16] CuAAC is completely regioselective in the formation of the 1,4-disubstituted triazoles towards the 1,5 isomer, occurs with fast kinetics under mild conditions, and is extremely

chemoselective (fig. 3). Moreover, azide and alkyne functional groups are largely chemically inert in the absence of catalysts. These properties have made CuAAC a prominent reaction in many fields of research, including material, polymer chemistry, drug discovery and chemical biology.

a) Huisgen's 1,3 dipolar cycloaddition



b) Copper-catalyzed 1,3 dipolar cycloaddition

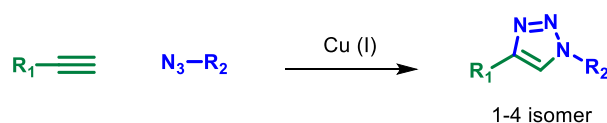


Figure 3: Synthesis of 1,2,3-triazoles *via* 1,3-dipolar cycloadditions

The 1,4-disubstituted-1,2,3-triazole heterocycle shares numerous conformational and electrostatic similarities with the *trans*-amide: it has a planar structure, shows similar polarisation and other electronic properties, and features hydrogen bond donors and acceptors sites. [17–19] (fig. 4) Probably with a bit of exaggeration, triazoles has been qualified as “universal peptidomimetics” that could potentially accommodate any peptide secondary structure. [20] However the triazole has an additional hydrogen bond acceptor site, the overall dipole moment of the triazole (about 5 D) is higher than that of the amide bond (about 3.5 D) and, most importantly, the two substituents are separated by two atoms in the amide bond (3.9 Å) versus three in the 1,4-triazole (5.0 Å). This divergences with amide can lead to poor mimics in some cases.

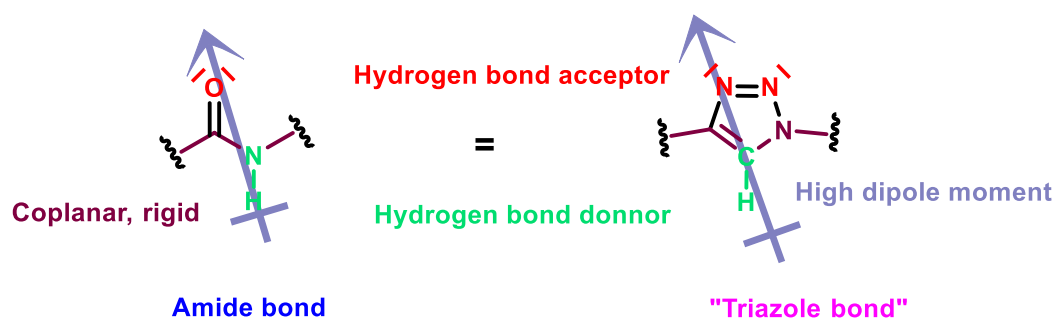


Figure 4: 1,4-Disubstituted-1,2,3-triazoles as amide bond mimics [11]

D. Peptidomimetic triazole ligation (PTL)

Capitalizing on these amide-mimicking properties as well as on the chemoselectivity of the reaction, our group [11] first reported in 2011 the application of CuAAC as a “pseudo-native” ligation for the assembly of unprotected peptide segments into a bioactive triazole-containing protein (fig. 5). This method was called Peptidomimetic Triazole Ligation (PTL) and its utility was demonstrated with the synthesis of a triazole-containing analogue of the 97 amino acid protein cystatin A which retained full bioactivity.

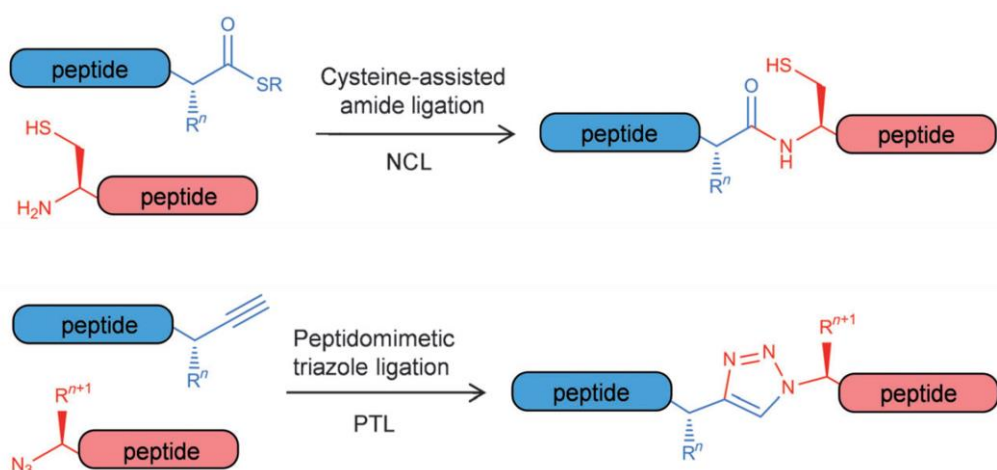


Figure 5: Use of CuAAC for chemical protein synthesis: peptidomimetic triazole ligation (fig. reproduced from ref [11])

E. Triazoles as proteases-resistant amide bond mimics

Besides having a similar structure, triazole peptidomimetics have been proven to be useful tools to enhance the bioavailability of the parent peptide, due to their absence of susceptibility towards proteolytic degradation. Some relevant examples are presented in the following section.

a) Amide-to-triazole mutation to enhance the *in vivo* half-life of peptides

A few examples of the amide-to-triazole mutations of proteolysis-sensitive peptides used for diagnostic or therapeutic purposes have been reported.

Bombesin (BBN) is a 14-amino acid peptide that shows high affinity for the gastrin-releasing peptide receptor (GRPr) that is overexpressed in numerous tumours. Thus, BBN can be used as targeting unit for applications in cancer imaging, however it is rapidly degraded by proteases upon injection. In this context, Mindt and co-workers focused on a minimal binding

sequence, BBN [7-14] and investigated the effect of the systematic replacement of each peptide bond with a 1,4-disubstituted 1,2,3-triazole, synthesized through on-resin CuAAC. [21] This “triazole scan” study allowed to identify triazolo-peptides with an enhanced stability and improved tumour-targeting capability. [22–25]

Kisspeptins [26] are neuropeptides responsible for the central control of the reproductive system in mammals and other vertebrates, acting by triggering hormone release cascades upon binding to their receptor GPR54. However poor pharmacodynamics of this peptide limits its utilization for therapeutic and veterinary applications. Our group reported the design of a derivative of kisspeptin 10 (KP10, 10 aa) where an amide-to-triazole mutation was introduced at one site known to be very sensitive to proteolytic degradation. This led to a considerably improved stability in blood serum, [27] basis for the development of a highly stable and potent agonist for *in vivo* applications. [28]

b) Triazolo-analogues of protease substrates as protease inhibitors

Cathepsins are proteases involved in a wide range of physiological and pathological processes. Cathepsin K and S are relevant targets for the treatment of different diseases, and despite great improvement, current pharmacological inhibitors are still limited in terms of safety and side effects. Within this context, our group reported the synthesis of analogues of peptides developed as selective substrates for Cat S and Cat K, where the protease-labile bond was replaced by a triazole. [29] This substrate-to-inhibitor translation strategy led however to moderate inhibitors (μM range), whereas similar peptidomimetics based on the introduction of an aza-amino acid residue [30] were much more potent (nM range).

c) Triazole-based deubiquitinases-resistant ubiquitin analogues

Protein post-translational modifications (PTMs) refers to the covalent and generally enzymatic modification of proteins following their ribosomal biosynthesis. This modification can occur on the amino acid side chains or at the C- or N-terminus. [31] It includes phosphorylation, glycosylation, amidation, acetylation, methylation, ubiquitination, SUMOylation and many others.

Ubiquitination is the attachment of the small protein ubiquitin (Ub, 76 aa) [32] (fig. 6) through the formation of an isopeptide bond between the C-terminal Gly residue of ubiquitin and the

amine of a Lys side chain of the substrate protein. The name of ubiquitin came from the fact that it was believed that this protein was present ubiquitously in both prokaryotes and eukaryotes. However, further studies proved that it only exists in eukaryotes. [33]

Human ubiquitin: ¹M¹⁰Q¹⁰I¹⁰F¹⁰V¹⁰K¹⁰T¹⁰L¹⁰T¹⁰G¹⁰K¹⁰T¹⁰I¹⁰T¹⁰L¹⁰E¹⁰V¹⁰P¹⁰S¹⁰D¹⁰T¹⁰I¹⁰E¹⁰N¹⁰V¹⁰K¹⁰A¹⁰K¹⁰I¹⁰Q¹⁰D¹⁰K¹⁰E¹⁰G¹⁰I¹⁰P¹⁰P¹⁰D¹⁰Q¹⁰
⁵⁰Q⁵⁰R⁵⁰L⁵⁰I⁵⁰F⁵⁰A⁵⁰G⁵⁰K⁵⁰Q⁵⁰L⁵⁰E⁵⁰D⁵⁰G⁵⁰R⁵⁰T⁵⁰L⁵⁰S⁵⁰D⁵⁰Y⁵⁰N⁵⁰I⁵⁰Q⁵⁰K⁵⁰E⁵⁰S⁵⁰T⁵⁰L⁵⁰H⁵⁰L⁵⁰V⁵⁰L⁵⁰R⁵⁰L⁵⁰R⁵⁰G⁵⁰G⁵⁰

Figure 6: Ubiquitin (Ub) structure with ubiquitinable residues highlighted in blue

In addition to protein mono-ubiquitination, oligo-Ub chains can be formed (fig. 7) through the attachment of one Ub to the N-terminus or any of the seven Lys residues of another (highlighted in fig. 6). The lysine residues used for Ub-Ub conjugation is assumed to determine the biological function of the poly-Ub chain. [34] This PTM is reversed by a class of proteases called deubiquitinases (DUBs) which hydrolyze the isopeptide bond. [35–37]

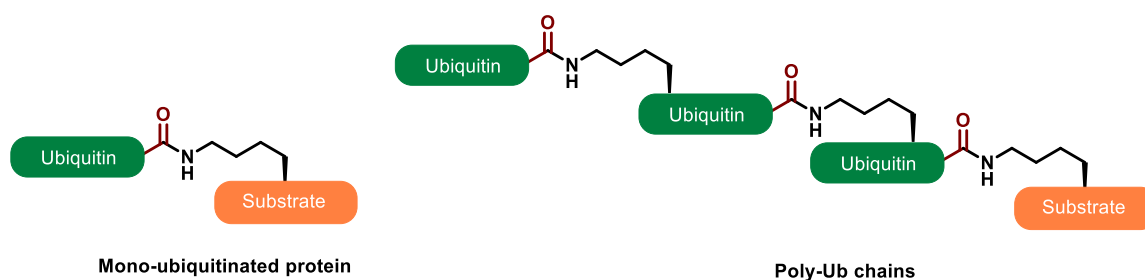


Figure 7: Ubiquitinated proteins

The difficulties in obtaining well-defined Ub- and poly-Ub-modified proteins by enzymatic preparations opened the way for the development of their chemical syntheses and semi-syntheses: many native and non-native ubiquitination methods have been developed.

The production of DUBs-stable analogues has proven useful for applications in cell biology and proteome research, otherwise precluded by the lability of native isopeptide bonds in living environments or cell lysates. In this context, triazoles have been used as amide mimics to precisely replace the isopeptide bond. For example, Ovaa and co-workers synthesised di-Ub probes through peptidomimetic triazole ligation of two chemically-synthesized Ub, [38,39]

one incorporating either an azido-norvaline ($n = 1$) or azido-norleucine ($n=2$) mutation at any of the seven lysine positions, and the other a C terminal propargylamide instead of the Cter Gly residue (fig. 8).

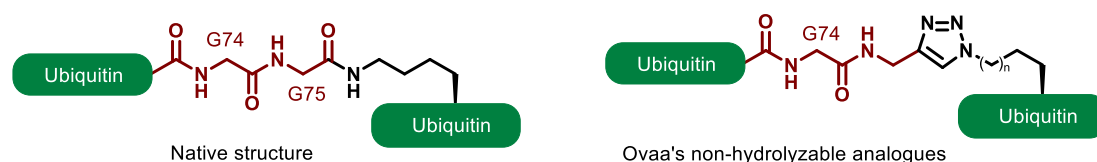


Figure 8: Triazole as an isopeptide bond mimetics

Note that in addition to this precise modification exploiting the amide-mimicking properties of the triazole in order to obtain a compound with a very similar structure, CuAAC has been widely used as a bioconjugation method to link two Ubs or an Ub with an acceptor protein, using recombinant protein expression or hybrid chemical synthesis/recombinant methods. A variety of linkage structures are presented in fig. 9.

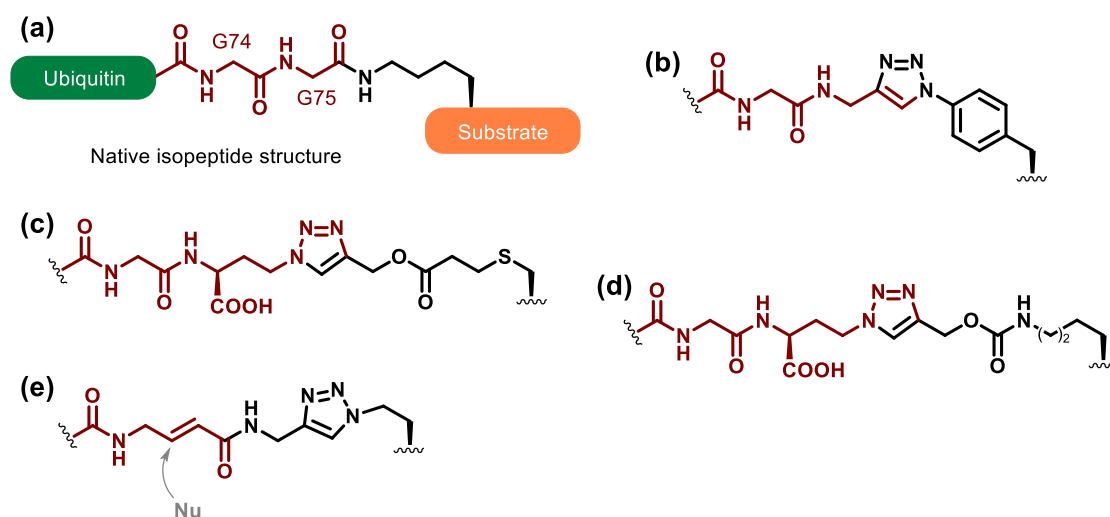


Figure 9: Triazole-containing linkages for non-enzymatic ubiquitination

For example, the azidophenylalanine-based linkage (fig. 9-b) developed by Mootz [40] has been demonstrated to be an excellent mimics using hybrid quantum mechanical and molecular mechanical (QM/MM) methods. Longer linkages more divergent from the native structure have also been introduced by Marx (fig. 9-c) [41] and (fig. 9-d) [42–49] shown to be resistant to DUBs [41] and to maintain both structural and dynamic properties of a natural isopeptide-linked Ub dimer. [48] The CuAAC-mediated ubiquitination approach has also been used by Kessler and co-workers to introduce an electrophilic site, to act as a trap for the catalytic cysteine residue of DUBs, and thus lead to the formation of covalent adducts (fig. 9-e).

2) Synthesis and semi-synthesis of SUMOylated peptides and proteins

SUMOylation is a PTM that plays a very important role in the normal functions of cells and thereby has become one of research hotspots in recent years. Similarly to ubiquitination, it involves the conjugation of a member of the small ubiquitin-like modifier (SUMO) family of proteins to the side chain of a lysine residue in target proteins to form an isopeptidic bond. A large variety of proteins have been proposed as SUMO targets based on the presence of consensus SUMOylation sequences, the most common being the tetrapeptides Ψ KXD and Ψ KXE where Ψ is a hydrophobic amino acid and X any aa.

SUMOylation can mask part of the surface of the protein and modify its conformation, affecting its interactions with other macromolecules. It can also modify its subcellular localization, inhibit or facilitate its degradation, affect enzyme activity, and is associated with genomic stability and DNA repair. [50]

SUMO proteins were first identified in mammals in 1996 [51], where they were found to be covalently linked to the GTPase activating protein RanGAP1. In humans, four SUMO proteins have been identified, SUMO-1, SUMO-2, SUMO-3 and SUMO-4 [50,52,53] (97, 93, 92 and 94 amino acid residues respectively (fig. 10)).

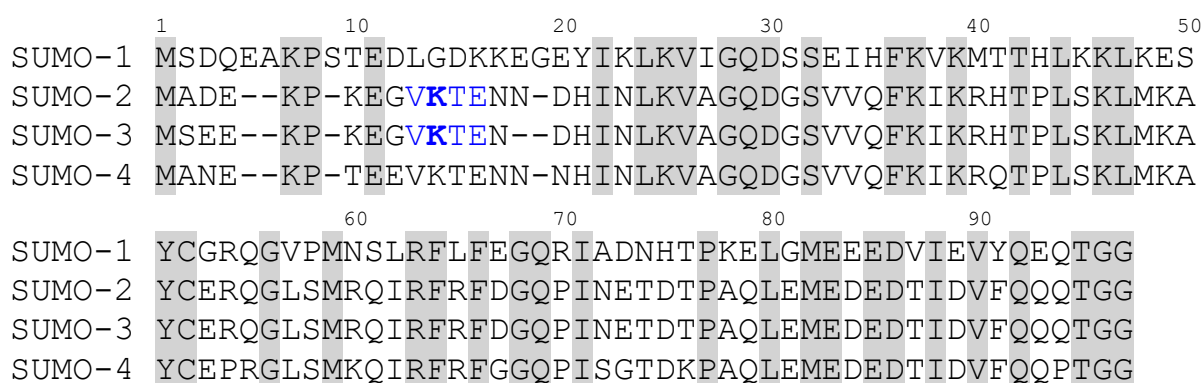


Figure 10: Sequences of human SUMO1-2-3-4 isoforms with conserved residues highlighted in grey, SUMOylable lysine in SUMO-2 and -3 in bold and consensus sequence in blue (sequence alignment taken from ref. [54])

SUMO-2 and SUMO-3 differ from each other by only three N-terminal residues and have yet to be functionally distinguished. They show ~50% sequence identity with SUMO-1. [55] SUMO-1 is usually conjugated to proteins as a monomer, while SUMO-2 and -3 form poly-SUMO chains on proteins, due to the presence of a consensus SUMOylation site in their N-

terminal regions (highlighted in blue in fig. 10). The fourth variant, SUMO-4 was discovered more recently and is probably not conjugated to proteins under physiological conditions. [56] It shares 86% amino acid sequence identity with SUMO-2. In contrast with SUMO-1,-2 and -3 which are expressed across many tissues, SUMO-4 was detected mainly in the kidney. [57]

Despite the low similarity in their primary amino acid sequence, three-dimensional structure of the folded core of SUMO and ubiquitin proteins are nearly superimposable (fig. 11).

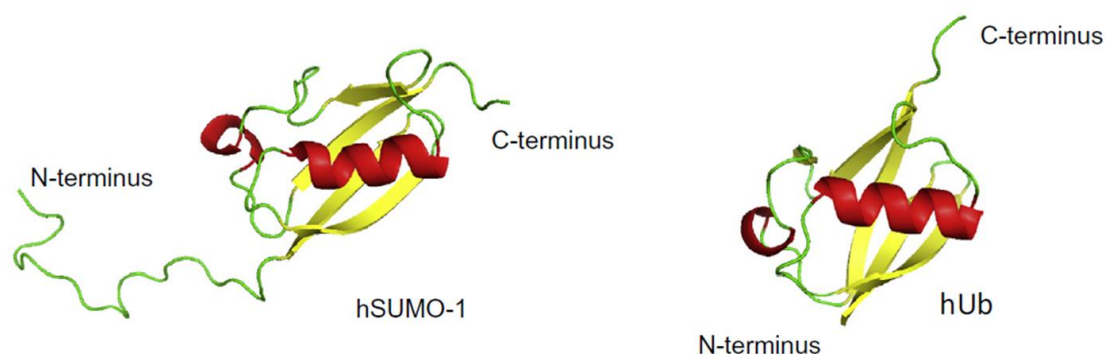


Figure 11: Ub and SUMO proteins showing similar three-dimensional structure (this figure was partially reproduced from the review of Melnyk [58])

An enzymatic pathway that is similar to the ubiquitination pathway controls the attachment of a SUMO to a protein. [52] It involves an enzymatic cascade with three types of biocatalysts: activator (E1), conjugating (E2), and ligase (E3) enzymes. Contrary to ubiquitin, the E3 enzyme is not always mandatory. It should be stressed that, like for ubiquitination, the reaction is reversible; SUMO can be deSUMOylated by sentrin/SUMO-specific proteases (SENPs). [59]

While the semi- or total chemical synthesis of ubiquitinated proteins or ubiquitin oligomers has drawn a lot of attention, the chemical synthesis of SUMO proteins remains relatively unexplored, with only few examples reported. An overview is given in the next pages.

A. SPPS based synthesis of SUMO proteins

Due to its large size (92-97aa), most reports of SUMO syntheses used a ligation-based approach from two or more segments (*vide infra*). However, In 2018, Ovaa's group [60] reported a solid-phase peptide synthesis approach of the SUMO isoforms 1 2 and 3 by incorporating aggregation breakers such as pseudoprolines [61] and *N*-dimethoxybenzyl groups [62] at strategic positions (fig. 20). This direct synthetic approach permitted N- and C-terminal labelling to incorporate dyes and reactive warhead with generation of activity-based probes (ABPs) for SUMO-1, SUMO-2, and SUMO-3-specific proteases as a final aim. This fully synthetic linear approach opened the way for the total synthesis of SUMO proteins, however the use of expensive building blocks and limitations inherent to the synthesis of long peptides with SPPS led other groups to consider the use of native chemical ligation for the synthesis of SUMO proteins.

B. NCL-based synthesis of SUMOylated peptides and proteins

Due to the presence of a cysteine residue in the middle of SUMO sequences, NCL is a method a choice for the chemical synthesis of these proteins. The first total synthesis of a SUMOylated peptide was reported in 2014 when Melnyk and co-workers [63,64] described the assembly of SUMO-peptide conjugates in a one-pot, three-segments ligation approach (fig. 12). Note that to facilitate the synthesis, the C-terminal Gly residue was mutated into a cysteine.

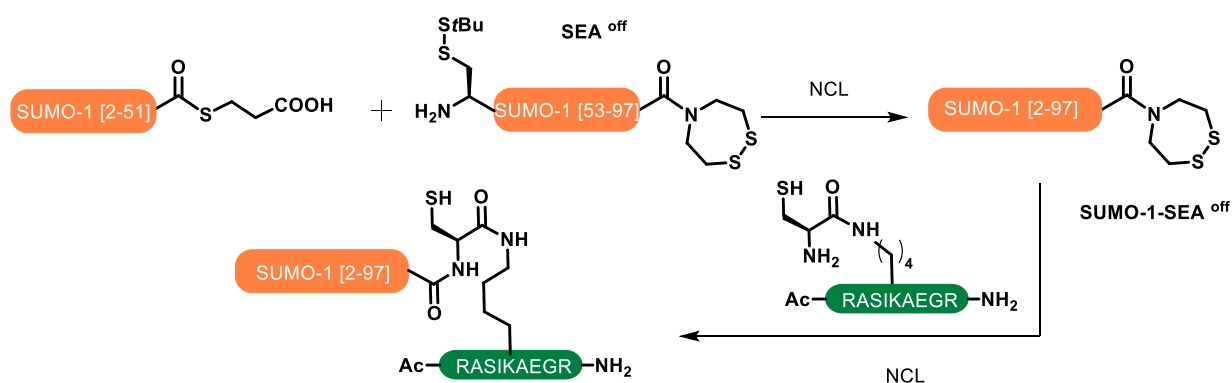


Figure 12: First report of the synthesis of a SUMOylated peptide by NCL

In 2016, the method was improved through the use of a δ -mercaptolysine that can serve as surrogate for the NCL, then be desulfurized to give the native Lys residue. [65] This approach known as isopeptide chemical ligation (ICL) is frequently used for preparing ubiquitin

conjugates. [66] More recently, the same group described the chemical synthesis of SUMO-2, SUMO-3, and SUMO-2/3 dimers [67]

In another illustration of the synthesis of SUMO conjugates by ligation, we can mention the work of Brik who used NCL for the assembly of four different SUMO-2-di-ubiquitin hybrid chains (fig. 13). [68]

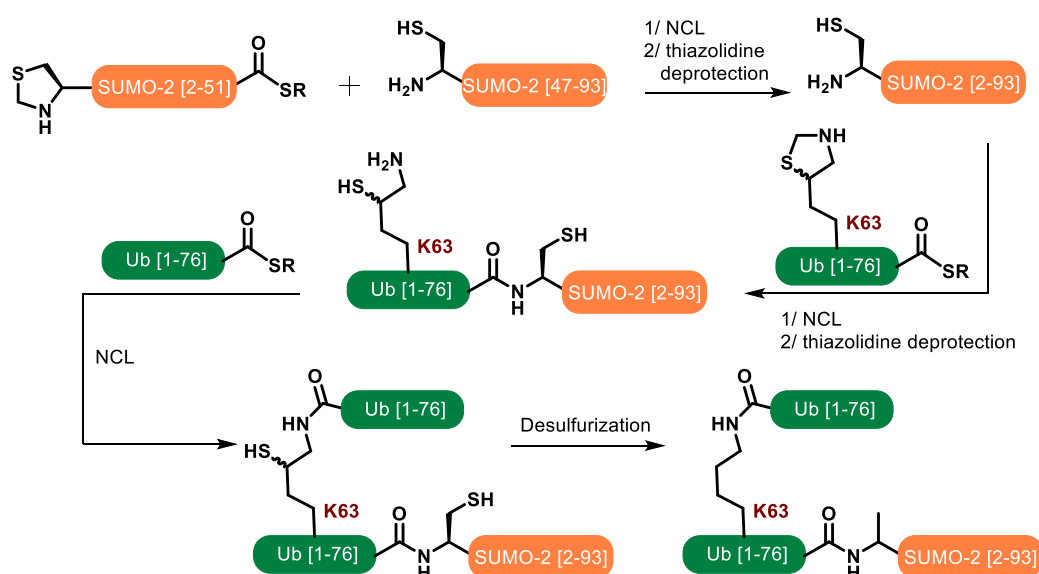


Figure 13: NCL based synthesis of SUMO-2-di-ubiquitin hybrid chains

C. KAHA ligation-based synthesis of SUMO proteins and SUMOylated peptides

KAHA ligation, based on a chemoselective reaction between C-terminal α -ketoacids and N-terminal hydroxylamines peptide segments to form amides under aqueous conditions and introduced by Bode's team in 2006 [69] was used for the synthesis of SUMO proteins. [70] SUMO-2 and SUMO-3 were assembled from respectively four and three segments affording synthetic proteins with close structures to the native ones (fig. 14). Note that a key partner in this ligation is the 5-membered cyclic hydroxylamine (S)-5-oxaproline, which forms a homoserine as the ligation product.

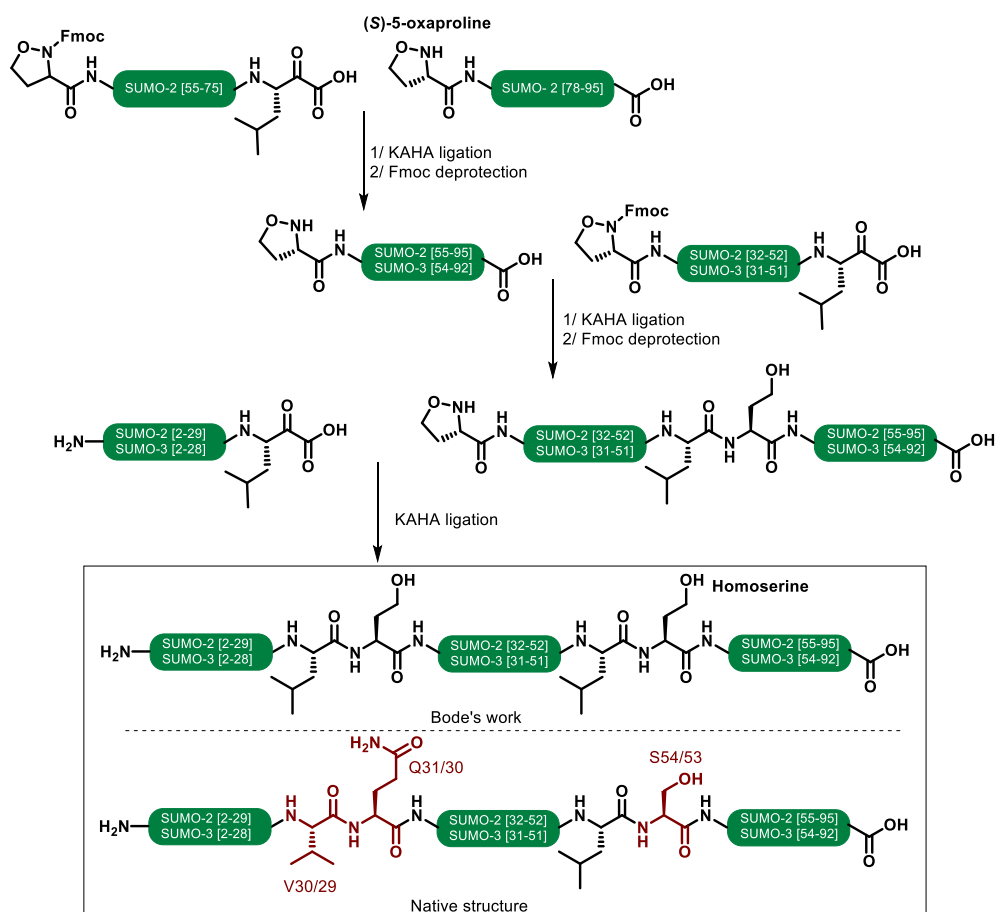


Figure 14: Synthesis of SUMO-2 and SUMO-3 proteins by successive KAHA ligations

More recently, the same team designed a 4,4-difluoro version allowing the formation of a native aspartic acid at the ligation site. As a proof of concept, this new version of KAHA ligation was used for the synthesis of SUMO-3 (fig. 15). [71]

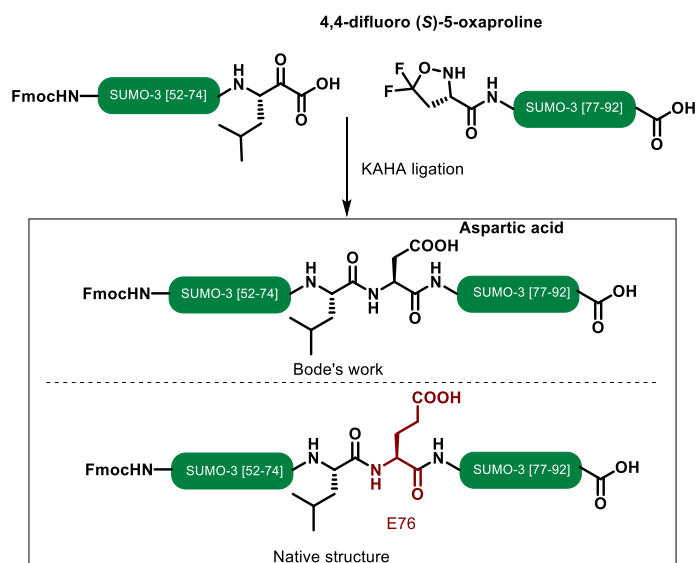


Figure 15: Chemical synthesis of a SUMO-3 segment prepared by KAHA ligation

D. Semi-synthesis of SUMOylated substrates

The first report of the semi-synthesis of SUMO proteins was the work of Mootz and his team [72] who developed a chemical conjugation approach to site-specifically introduce stable SUMO modifications into fully recombinant substrate proteins. To this end, a native lysine residue in the substrate protein was mutated into a cysteine and an azide functionality was introduced following a chemoselective reaction with iodoacetamidoethyl azide. The SUMO-2 protein was expressed in fusion with a mutant intein to generate a C-terminal thioester that was reacted with propargylamine to afford the terminal alkyne group (fig. 16).

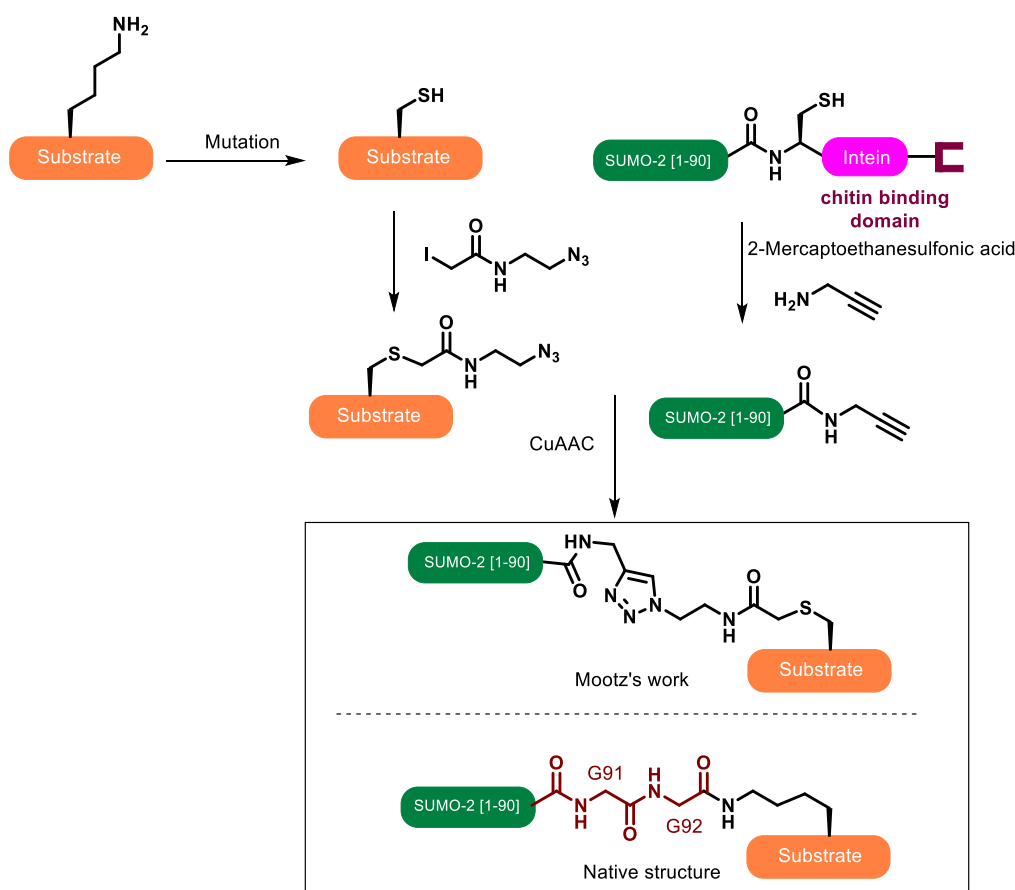


Figure 16: Strategy for the generation of synthetic SUMO-2 conjugates [72]

In a very close approach, and similarly to his work on ubiquitination (fig. 9b), Mootz also reported the conjugation of a para-azidophenylalanine (AzF)-containing recombinant proteins with a SUMO-2 alkyne. [73] AzF was introduced by unnatural amino acid mutagenesis [74] at the position of the acceptor lysine. The two approaches described above were later applied

by the same group to RanGAP1-SUMO-1 conjugates [75] They found that introducing the azide at the desired position *via* cysteine alkylation was very effective compared to the use of genetically-encoded Azf.

Liu and his team [76] reported the cysteine-aminoethylation assisted chemical ubiquitination (CAACU) strategy to produce SUMO-modified histone analogues. The CAACU key step is the site-specific installation of an *N*-alkylated 2-bromoethylamine derivative onto a Cys residue of recombinant histones which is followed by chemical ligation with a recombinant SUMO thioester. Note that this ligation is assisted by the removable auxiliary 2-mercapto-2-phenethyl group developed by Seitz and his team. [77]

Metanis and his group developed selenium-containing analogues of modified Lys residues in order to afford traceless isopeptidic bonds upon deselenization. To demonstrate the utility of this Se-ICL method, it was used for the semi-synthesis of a SUMOylated peptide using a recombinant SUMO thioester. [78]

II- Aim of this work: towards chemical biology tools to decipher neurofibromin SUMOylation

The imbalance of SUMOylation and deSUMOylation has been associated with the occurrence and progression of various diseases such as cancer, bacterial and viral infections, neurodegenerative and cardiac diseases. [56] Understanding the role of SUMOylation at the molecular level is fundamental to the discovery of new therapeutic targets and also for the development of new treatments for these diseases.

1) Neurofibromatosis and SUMOylation

Recently, the Bénédicti group at CBM demonstrated that SUMOylation plays a curtail role in neurofibromatosis type 1, [79] a common genetic disease which affects the nervous system caused by constitutional mutations in the NF1 gene. This gene is approximately 350 kb in size, contains 61 exons and is responsible for the expression of the 2818 amino acid protein neurofibromin (Nf1), which acts as a tumour suppressor. The mutation rate in the NF1 gene is one of the highest known for human genes. [80]

They identified lysine 1731 as a major SUMOylation site in the SecPH domain of Nf1 (fig. 17). Several patient mutants localized in the SecPH domain have an altered SUMOylation profile.

Mutants with altered SUMOylation profile appear to be more unstable, meaning that disruption of SUMOylation of Nf1 may contribute to the disease.



Figure 17: Nf1 domains with SUMOylated SecPH domain highlighted

As a general statement, major difficulties in studying protein SUMOylation are due to the low amount of detectable SUMOylated proteins as well as the very high sensitivity towards SUMO proteases of the isopeptide bond between SUMO and the substrate.

The long-term goal of the work described in this chapter is to explore the development of innovative chemical biology tools to overcome this difficulty in order to facilitate the study of Nf1 SUMOylation. This chapter focuses on the development of a synthetic methodology to afford a SUMOylated peptide wherein the isopeptide bond is replaced by a triazole, non-hydrolyzable bond amide mimic.

2) Synthetic strategy

Our idea was to assemble SUMOylated peptides by using two orthogonal chemical ligation reactions: NCL and PTL. In this way, there is no need for masking/unmasking of a reactive group to perform the two successive ligations (fig. 18).

The two ligations have been implemented on solid phase, using the enzyme-cleavable linker-based SPCL methodology described in [chapter 2](#), with the initial idea to extend the strategy to the synthesis of SUMO protease-stable di-, tri- or tetra-SUMOylated peptides using the same building blocks. A massive advantage of the orthogonal PTL / NCL methodology is that the total number of steps would be halved as compared with a traditional approach requiring masking/unmasking steps for successive NCLs.

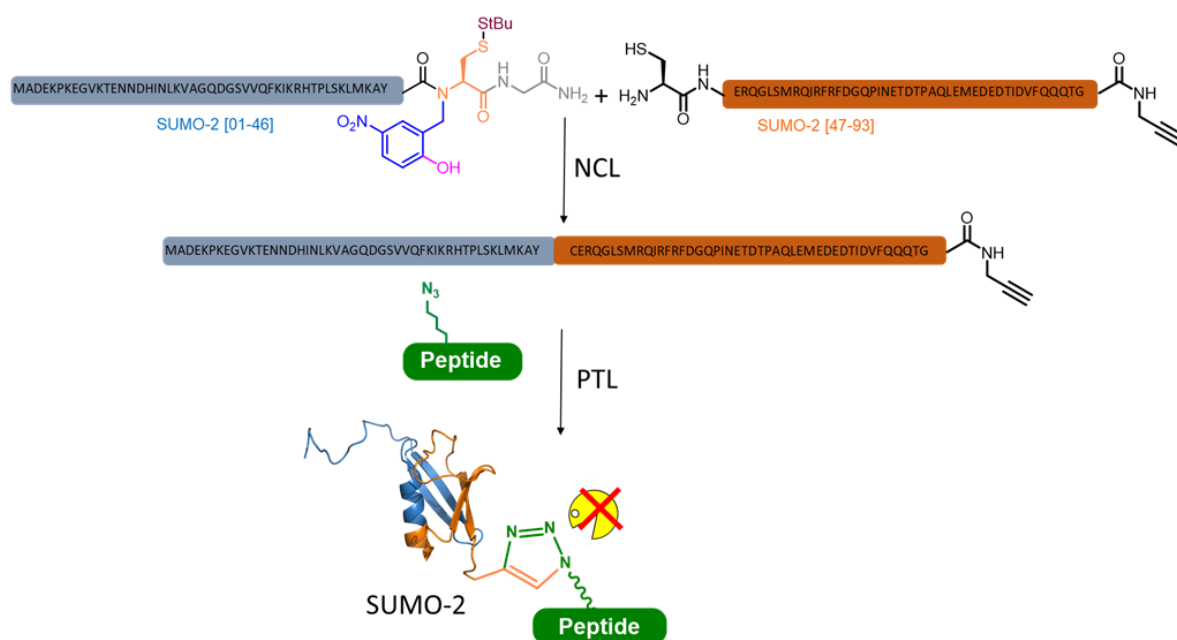


Figure 18: The two orthogonal chemical ligations used for the assembly of the SUMOylated peptide

III- Results and discussion

1) Synthesis of the three peptide segments

A. N-terminal segment (5)

The N-terminal segment **5** was equipped with both our second-generation linker functionalized by an azide group, and an *N*-Hnb-Cys device where thiol is protected by an *S*-StBu disulfide, labile under reductive NCL conditions.

SUMO-2 [1-46] crypto-thioester peptidyl resin **1** was synthesized as already described in [chapter 4](#). The second generation enzyme cleavable linker precursor **2** was introduced through the nucleophilic amine attack of **1** on the activated carbonate. Then, a piperidine treatment removed the Fmoc protecting group, and the coupling of azido acetic acid followed by TFA cleavage and HPLC purification gave **5** in a 15% yield (fig. 19).

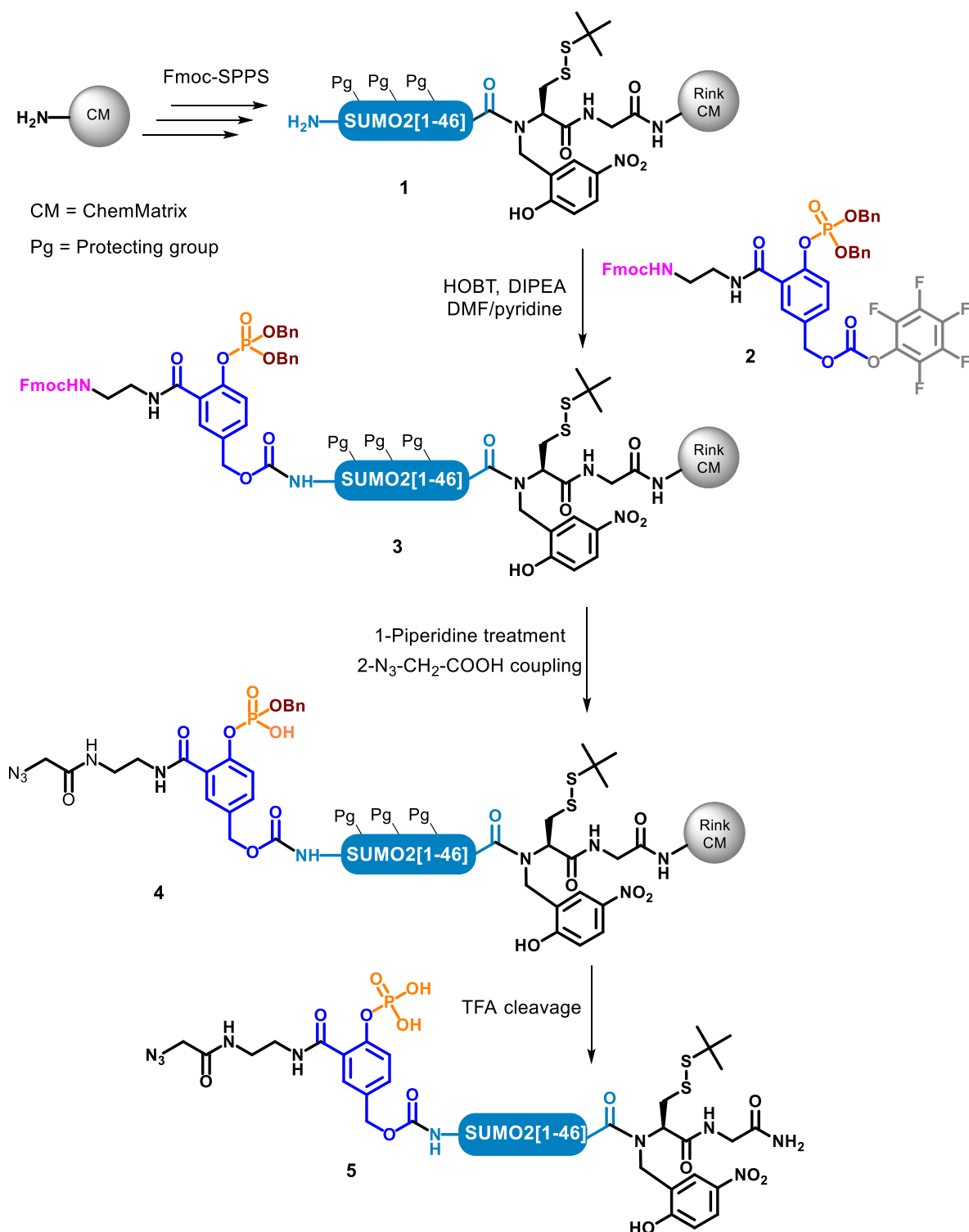
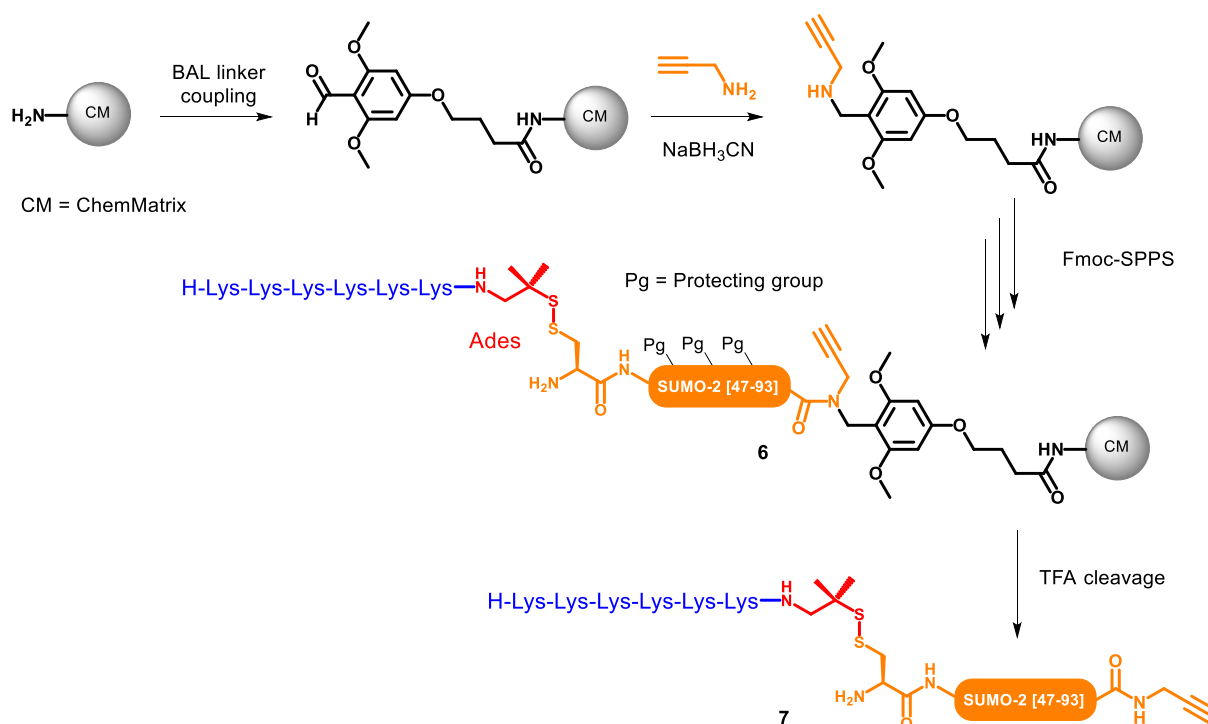


Figure 19: Synthesis of compound 5

B. Cysteiny / alkyne segment 7

The second segment was equipped with both an N-terminal cysteine for NCL and a propargylamide at its C-terminus for subsequent PTL. Our SPPS strategy was similar to the one used for the segment containing a C-terminal carboxylic acid and described in [chapter 4](#)

(compound **12**, chap. 4 numbering), except that the starting resin was here equipped with a backbone amide linker (BAL) loaded with propargylamine following a previously described reductive amination protocol. [11] D16 and G17 were introduced through the coupling of dipeptide Fmoc-Asp(OtBu)-(Dmb)Gly-OH to prevent aspartimide formation during SPPS, and Q88 and Q89 were subjected to double couplings to ensure the completion of the reaction. A hexa-lysine-Ades solubilizing tag was introduced following the protocol reported in [chapter 4](#) of this manuscript. Subsequent cleavage from the solid support and HPLC purification gave **7** in a 35% yield (fig. 20). Note that the counterpart **S6** devoid of a solubilizing tag was also synthesized, and showed similar low solubility and anomalous HPLC behaviour observed for [chapter 4](#)'s peptide **S11**.

Figure 20: Synthesis of compound **7**

C. Azido-tetrapeptide **9**

A tetrapeptide containing the consensus SUMOylation motif Ψ KXE, and whose lysine residue was replaced by an azido-norleucine (aka Lys(N₃)) was synthesized through manual Fmoc-SPPS. Subsequent cleavage from the solid support and HPLC purification gave **9** in a 65% yield (fig. 21).

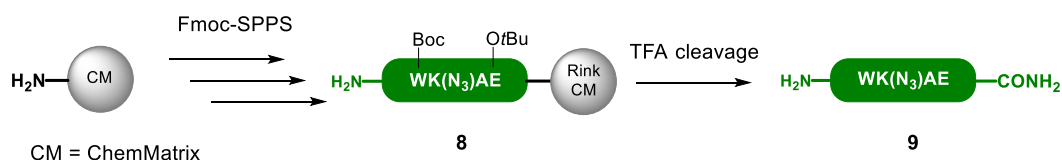
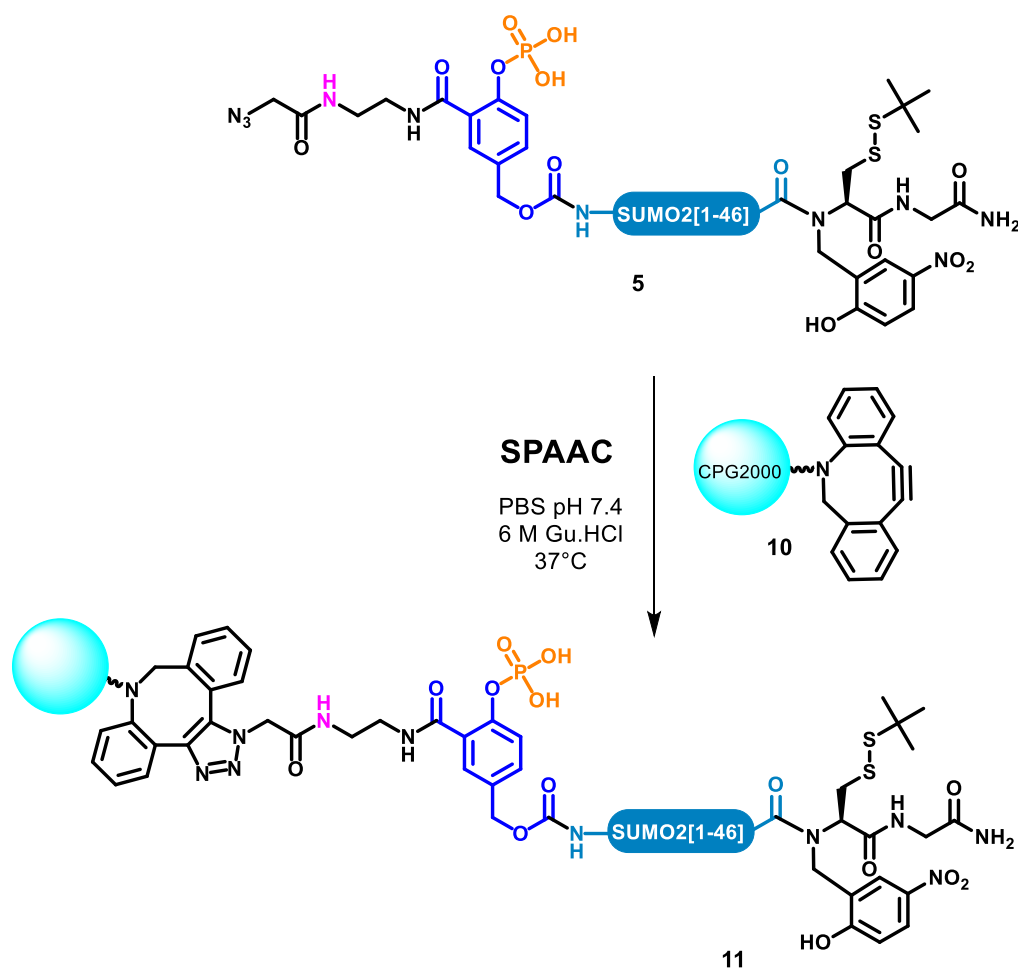


Figure 21: Synthesis of compound 9

2) SPCL-mediated target assembly

Azide-functionalized N-terminal segment **5** was immobilized on a cyclooctyne-CPG2000 solid support (**10**) through strain-promoted azide/alkyne cycloaddition (fig. 22), following a similar protocol as the one developed in [chapter 2](#) except that 6 M Gu.HCl had to be added to the mixture in order to dissolve the segment. Disappearance of **5** was monitored by analytical HPLC. Once the reaction was completed, excess cyclooctyne was capped *via* SPAAC using a large excess of azido acetic acid.

Figure 22: SPAAC-mediated immobilization of **5**

Native chemical ligation of peptidyl resin **11**, and K₆Ades-SUMO-2[47-93]-propargylamide (**7**) was first attempted using 2 equiv. of **7** (fig. 23). However, the reaction never went to completion even after a prolonged reaction time (up to 64 h), as judged by analytical cleavage (fig. 24A).

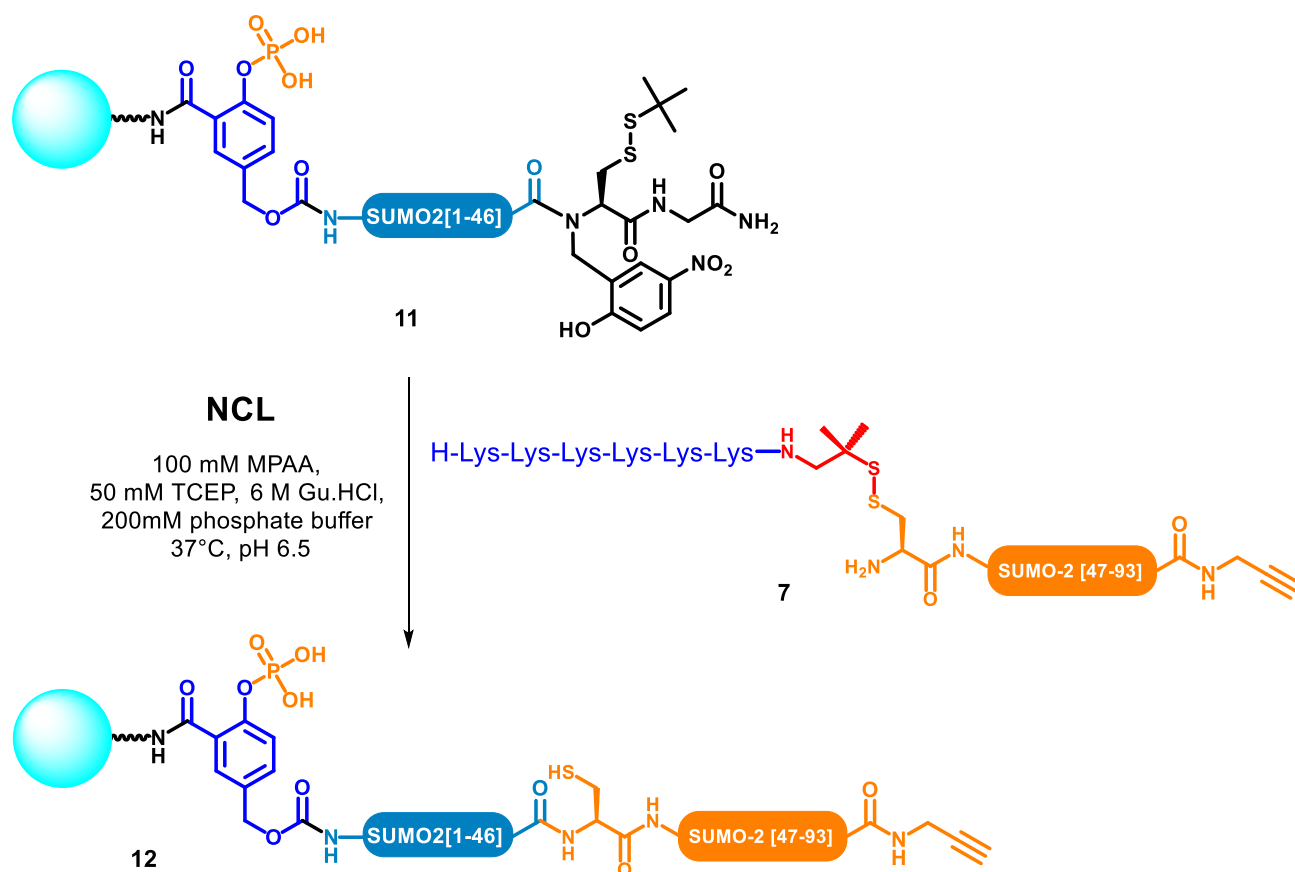


Figure 23: First solid-supported ligation to generate **12**

Several conditions were tested to optimize the reaction. Repeating the reaction under the same conditions or increasing the concentration of **11** from 5 to 9 mM did not improve significantly the conversion rate. Varying the pH from 6.5 to 8.5 or 7.5 resulted in lower conversion rates in concordance with the previous publication of our team showing that the optimum pH for NCL using the Hnb thioesterification device is 6.5. [81] Next, knowing that the solubilizing tag is nearly instantaneously removed under NCL conditions, we reasoned that low conversion could be the result of solubilisation issues of **7**. Melnyk and his co-workers reported the use of *N*-octyl glucoside as an additive to ensure the solubilization of starting materials and products throughout the synthesis process when assembling SUMO proteins. [67,82] However, the use of this detergent did not improve the conversion rate in our case.

Use of 10 equivalents of **7** for 64 h led to a nearly complete reaction, nonetheless, we thought that the use of a large excess of the segment limited the scope of this approach.

Given that we noticed that in incomplete reactions, segment **11** just lost the *StBu* protecting group without formation of the MPAA thioester, we thought that the problem could come from a lower concentration of the reagents than expected, due to dilution with the water present inside the CPG beads. Satisfyingly, the treatment of **11** with the NCL buffer (3 x 20 min) prior to the ligation resolved the problem and **12** was obtained with nearly 100 % conversion (fig. 24B).

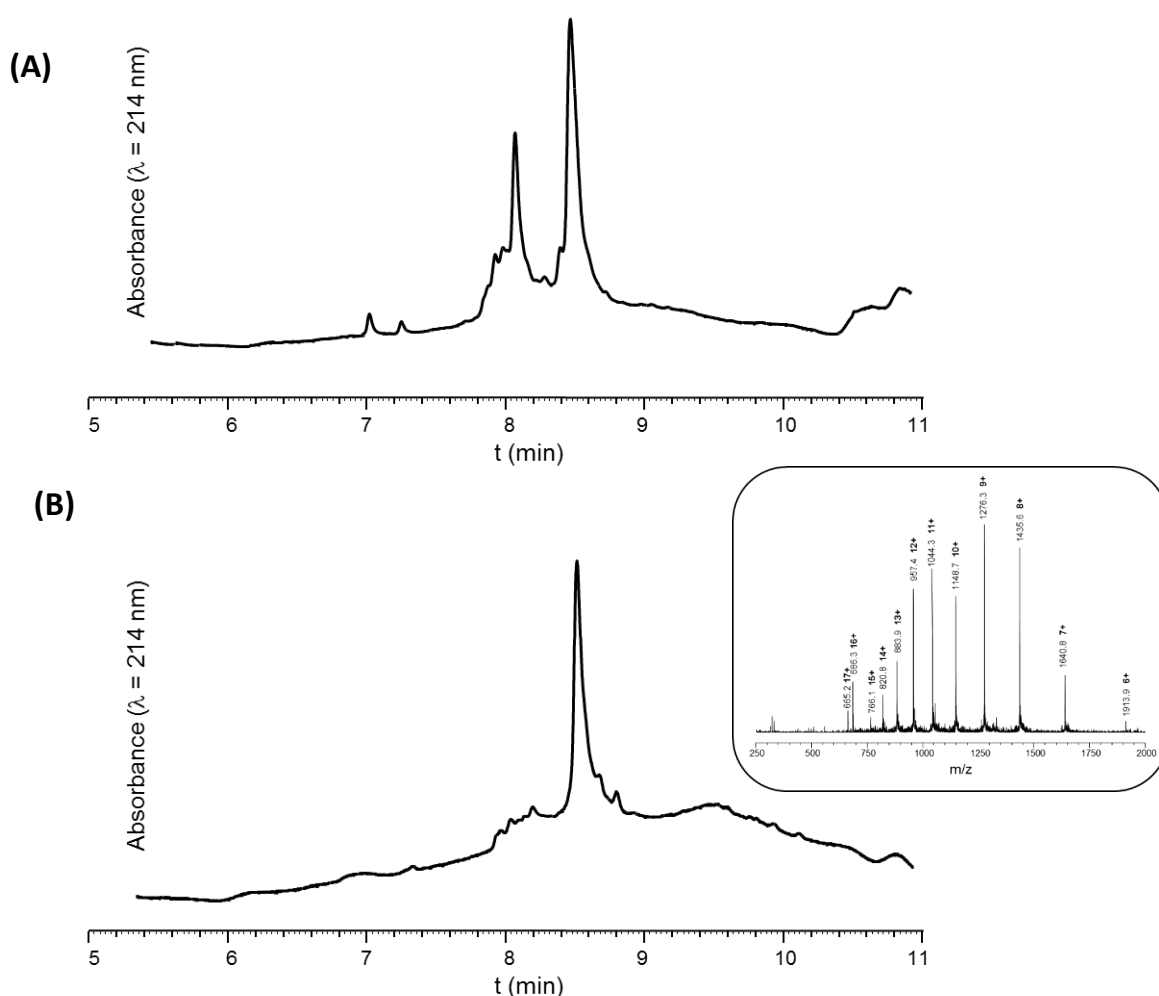


Figure 24: HPLC traces of incomplete (A) and complete (B) NCL reaction between **11** and **7**.
Insert: MS spectrum corresponding to the major peak.

Having in hands solid supported alkyne **12**, we next reacted it with tetrapeptide **9** (fig. 25). The reaction was carried out using a large excess of **9** and of the copper catalyst ($\text{CuBr} \cdot \text{Me}_2\text{S}$). Reactions were conducted under the conditions previously optimized in our team for solid

phase PTL [2]: in a mixture of NMP and 200 mM HEPES buffer at pH 7.5, using the copper(I) ligand *tris*-(hydroxypropyl-triazolylmethyl)amine (THPTA), [83,84] in the presence of *tert*-butanol and aminoguanidine as reactive oxygen species (ROS) scavengers. [84]

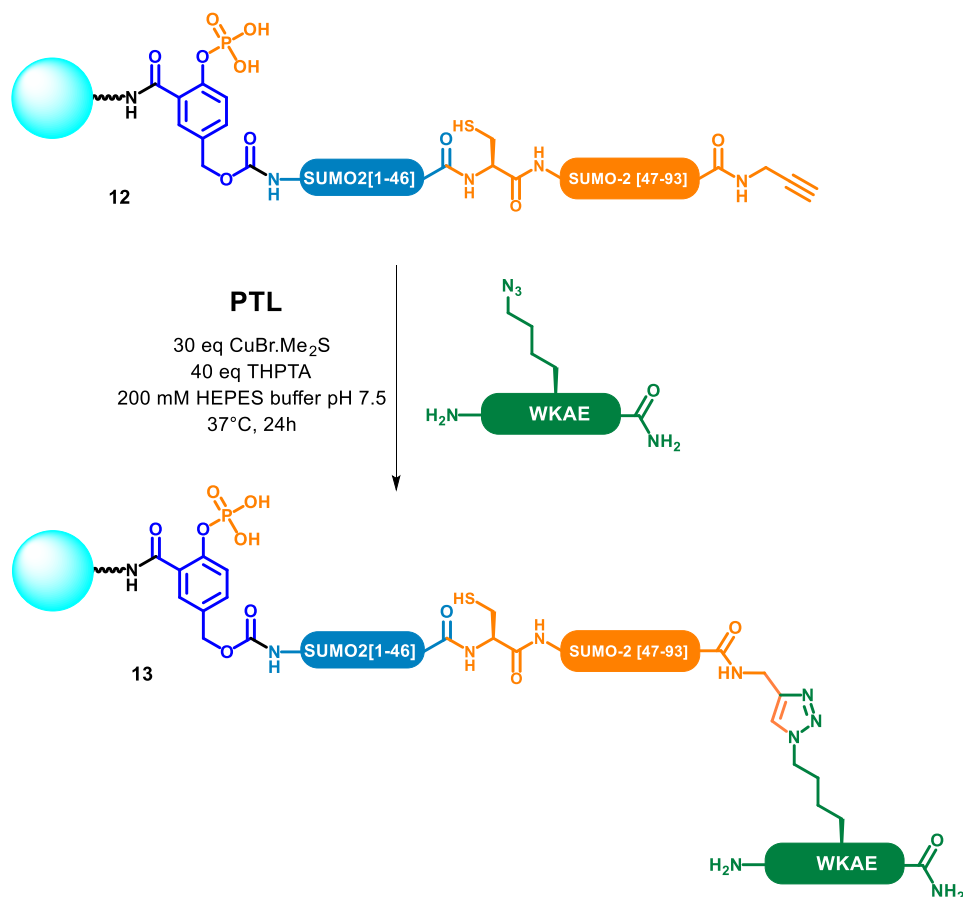


Figure 25: Second solid-supported ligation to generate **13**

Note that, the reaction was at first attempted with another model peptide **14**, biotin-labelled and containing a consensus SUMOylation sequence (fig. 26). The sequence of this peptide is derived from the promyelocytic leukemia protein (PML), a well-known SUMOylated protein (⁴⁸⁵PRKVIKMESEE⁴⁹⁵). It also worth mentioning that methionine was replaced by norleucine and an extra alanine residue was added. However, despite many attempts, the reaction with **14** did not work at all. As **14** showed atypical physicochemical properties (slow formation of a gel in HPLC solvents), we thought that formation of micelles, soluble aggregates or other supramolecular structures could likely explain this total absence of reactivity.

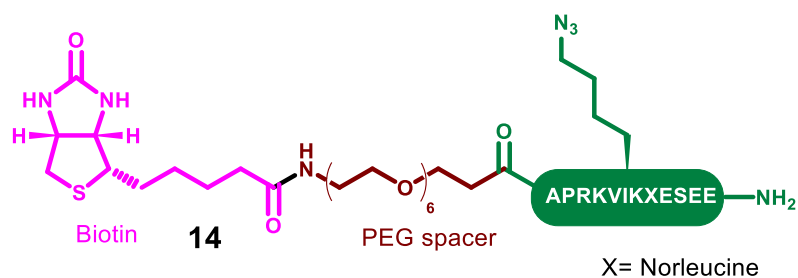


Figure 26: Structure of the PML derived peptide **14**

Finally, enzymatic cleavage of the linker was carried out with lambda phosphatase under the conditions optimized in [chapter 2](#) to afford the expected triazole-modified SUMOylated peptide **15** which was not purified (fig. 27).

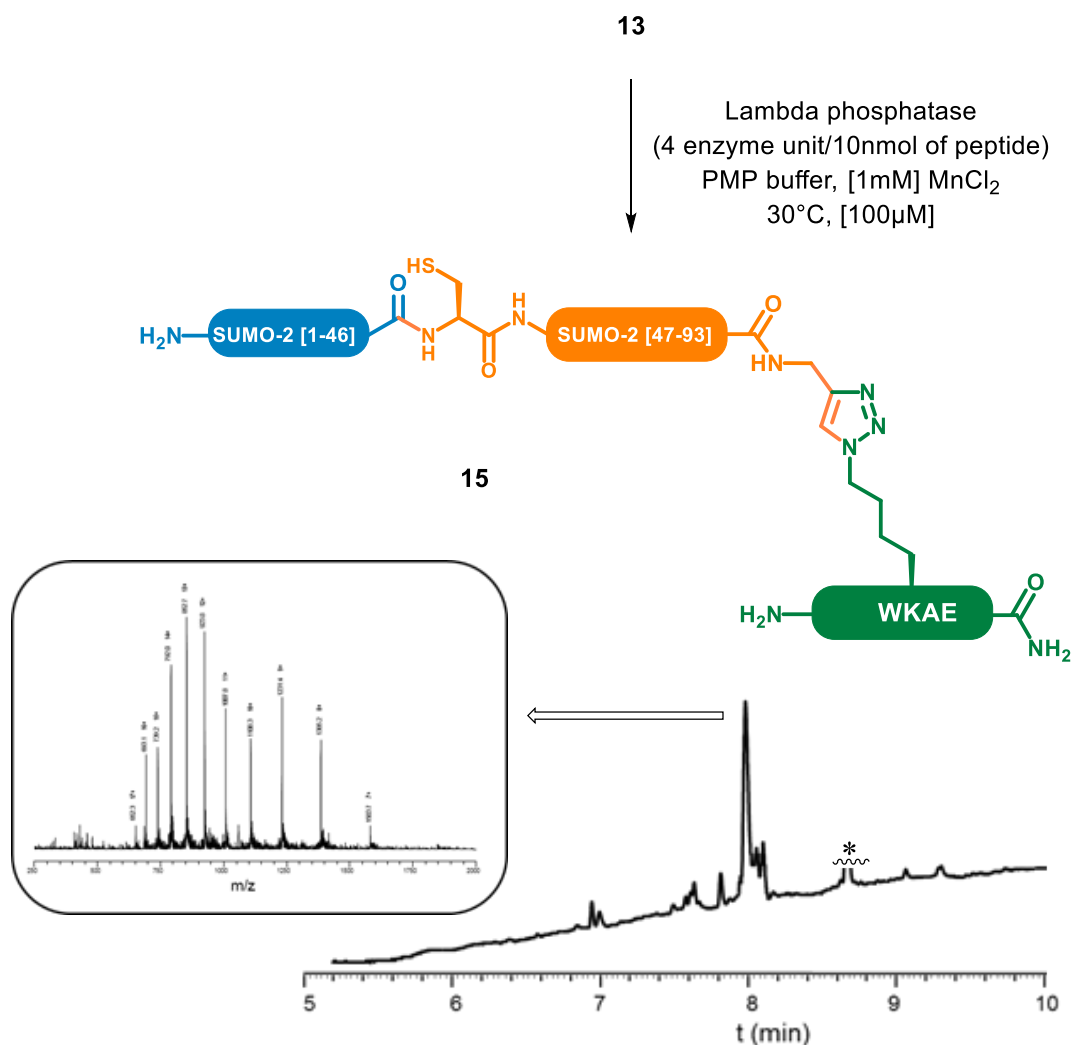


Figure 27: Enzyme-mediated release of **15**. *: non peptidic compound

IV- Experimental part

1) General information

All reagents and solvents were used without further purification. Protected amino acids were purchased from Gyros Protein Technology (Uppsala, Sweden). 1-Amino-2-methyl-2-propanethiol hydrochloride was purchased from Key organics (Camelford, UK). DIEA was purchased from Carlo ERBA (Val-de-Reuil, France). Rink ChemMatrix resin was purchased from Biotage (Uppsala, Sweden). Peptide synthesis grade DMF was obtained from VWR (Fontenay-sous-Bois, France). All other chemicals were purchased from Sigma Aldrich (St-Quentin-Fallavier, France) and solvents from SDS-Carlo Erba (Val de Reuil, France). Ultrapure water was obtained using a Milli-Q water system from Millipore (Molsheim, France). Lambda Protein Phosphatase (Lambda PP) and CutSmart buffer were purchased from New England Biolabs (Évry, France). Polypropylene syringes fitted with polypropylene frits were obtained from Torviq (Niles, MI, USA) and were equipped with PTFE stopcock bought from Biotage. HPLC analyses were carried out on a Chromaster system equipped with a 5160 pump, a 5430 diode array detector and a 5260 auto sampler and semi-preparative purifications were carried out on a LaChromElite system both equipped with a Hitachi L-2130 pump, a Hitachi L-2455 diode array detector and a Hitachi L-2200 auto sampler. Chromolith High Resolution RP-18e (150 Å, 10 × 4.6 mm, 3 mL/min flow rate) columns were used for analysis, Nucleosil C18 (300 Å, 5 µm, 250 × 10 mm, 3 mL/min flow rate) and Jupiter C4 (300 Å, 5 µm, 250 × 10 mm, 3 mL/min flow rate) for purification. Solvents A and B are 0.1 % TFA in H₂O and 0.1 % TFA in MeCN, respectively. Each gradient was followed by a washing step to elute any compound not eluted during the gradient (up to 95% B/A over 0.5 min, then isocratic 95% B/A over 0.5 min for the HR Chromolith). ESI-MS analyses were carried out on an Agilent 1260 Infinity HPLC system, coupled with an Agilent 6120 mass spectrometer (ESI + mode), and fitted with an Aeris Widepore XB-C18 2 (3.6 µm, 150 × 2.1 mm, 0.5 mL/min flow rate, 60°C) column. The reported *m/z* values correspond to the monoisotopic ions if not specified otherwise. Solvents A' and B' were 0.1 % formic acid in H₂O and 0.1 % formic acid in MeCN, respectively. Gradient: 3% B'/A' for 1 min, then 3 to 50% B'/A' over 15min. Low resolution MS of pure compounds were obtained using this system. When needed, the multiply-charged envelope was deconvoluted using the charge deconvolution tool in Agilent OpenLab CDS ChemStation software to obtain the average [M] value. For yield calculations purposes, the quantities of purified peptides were determined by weight, taking into account a molecular weight including trifluoroacetate counter-ions (one per Arg, His, Lys and N-terminal amine of the peptide sequence) but not water content.

2) General procedures

Protocol PS1- Peptide elongation: Manual couplings were performed on polypropylene syringes fitted with polypropylene frits using rotation stirring. Fmoc-based solid phase peptide syntheses (SPPS) were carried out on a Prelude synthesizer from Protein technologies. The side-chain protecting groups used were Arg(Pbf), Asp(*t*Bu), Cys(Npys), Cys(Trt), Cys(*St*Bu), Glu(*t*Bu), Gln(Trt), His(Trt), Lys(Boc), Ser(*t*Bu), Thr(*t*Bu), and Tyr(*t*Bu). Syntheses were performed on a 0.025 mmol-per-reactor scale. Protected amino acids (0.25mmol, 10 equiv.) were coupled using HATU (90.49 mg, 0.238 mmol, 9.5 equiv.) and *i*Pr₂NEt (87 µL, 0.5 mmol, 20 equiv.) in NMP (2.5 mL) for 30 min. Coupling on *N*-Hnb-cysteine secondary amine were performed for 18 h. Capping of eventual unreacted amine groups was achieved by treatment with acetic anhydride (143 µL, 1.51 mmol, 60 equiv.), *i*Pr₂NEt (68 µL, 0.39 mmol, 15.5 equiv.) and HOBt (6 mg, 0.044 mmol, 1.8 equiv.) in NMP (3 mL) for 7 min (4 x 7 min in the case of *N*-Hnb-cysteine secondary amine). Fmoc group was deprotected by three successive treatments with 20% piperidine in NMP (3 mL) for 3 min.

Protocol PS2 - Reductive amination: introduction of the Hnb group: 25 μ mol of H-Cys(StBu)-Gly-Rink-resin (1a-c) was washed with 1:1 DMF/MeOH (4 x 3 mL, 30 s). 2-Hydroxy-5-nitrobenzaldehyde (42 mg, 10 equiv.) in 2 mL 44.5:44.5:1 DMF/MeOH/AcOH (125 mM aldehyde concentration) was then added, and the reactor was stirred for 5 min. The reactor was drained and the resin was washed with 1:1 DMF/MeOH (3 x 3 mL, 5 s) then DMF (3 x 3 mL, 5 s). Without delay, a fresh solution of sodium borohydride (19 mg, 20 equiv.) in 2 mL DMF (250 mM borohydride concentration) was added and the reactor was stirred for 20 min. The reactor was drained and the resin was washed with DMF (4 x 3 mL, 30 s), 20% v/v piperidine in NMP (3 x 3 mL, 3 min), NMP (3 x 3 mL, 30 s), dichloromethane (3 x 5 mL, 30 s) and NMP (3 x 3 mL, 30 s).

Protocol PS3 - Boc-Cys(Ades) introduction: 25 μ mol of peptidyl resin was washed with 3 mL of NMP. A solution of Boc-Cys(Npys)-OH (93.5 mg, 10 equiv.), HATU (90.3 mg, 9.5 equiv.) and DIEA (87 μ L, 20 equiv.) in 2.5 mL of NMP (100 mM Boc-Cys(Npys)-OH) was added to the resin and the reactor was stirred for 4 hours. The reactor was drained and the resin was washed with DCM (3 x 3 mL, 30 s) and NMP (3 x 3 mL, 30 s). A solution of 2-amino-1,1-dimethyl-propane-1-thiol hydrochloride **2** (35.5 mg, 10 equiv.) in 2.5 mL of NMP (100 mM thiol) was then added to the resin and the reactor was stirred for 5 min. The reactor was drained off and the resin was washed with DCM (3 x 3 mL, 30 s) and NMP (3 x 3 mL, 30 s).

Protocol PS4 - Peptide cleavage: The crude peptide was deprotected and cleaved from the resin through a treatment with TFA/H₂O/*i*Pr₃SiH/phenol, 88:5:2:5 for 2 h, and the peptide was precipitated by dilution into an ice-cold diethyl ether/petroleum ether 1:1 mixture, recovered by centrifugation and washed twice with diethyl ether.

Protocol PS5 - Native chemical ligation:

Preparation of the NCL buffer (100 mM MPAA, 50 mM TCEP, 6 M guanidinium chloride, 200 mM sodium phosphate, pH = 6.5):

16.8 mg 4-mercaptophenylacetic acid (MPAA, 0.18 mmol), 14.3 mg tris-carboxyethylphosphine hydrochloride (TCEP, 0.057 mmol) and 10.3 mg of dry NaOH powder (0.26 mmol) were weighted into a vial, which was sealed with a septum and purged with an argon flow.

A 200 mM disodium hydrogen phosphate, 6 M guanidinium chloride aqueous solution was prepared by dissolving 356 mg disodium hydrogen phosphate dihydrate and 5.73 g guanidinium chloride in water (10 mL final volume). This solution can be kept at 4°C for several months. Just before use, this solution was thoroughly deoxygenated through successive vacuum/argon cycles.

Under argon, 1 mL of this solution was added to the MPAA/TCEP/NaOH vial, followed by sonication upon complete dissolution to give the NCL buffer which was immediately used.

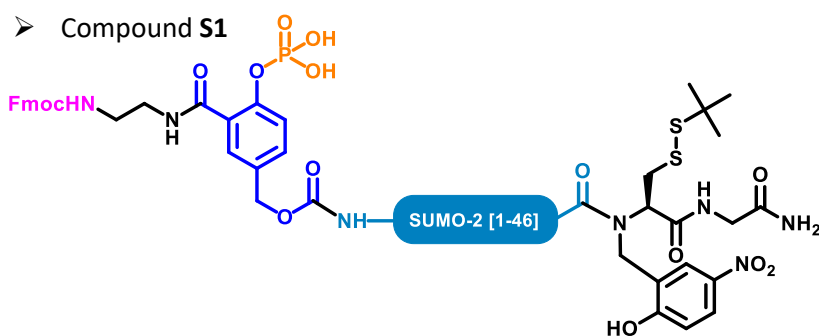
NCL reaction: The cysteinyl peptide and crypto-thioester peptidyl resin were weighted in two separate centrifuge tubes, which were sealed with a septum and purged with an argon flow. The volume of the NCL buffer appropriate to reach the desired final peptide segments concentration was added to the cysteinyl peptide under argon and then transferred to the tube containing the crypto-thioester peptidyl resin. The reaction vessel was sealed with Parafilm, and the resulting yellow solution was allowed to stir at 37 °C for 18 h. The resin was washed with the 200 mM disodium hydrogen phosphate, 6 M guanidinium chloride deoxygenated aqueous solution (3 x 1 mL, 30 s), 200 mM deoxygenated disodium hydrogen phosphate solution (3 x 1 mL, 30 s) and 6 M guanidinium chloride deoxygenated aqueous solution (pH = 3 ~4 to quench the reaction) (3 x 1 mL, 30 s) then drained and finally extensively washed with deoxygenated water (pH = 3 ~4).

3) Synthesis of the different segments needed for SPCL

A. SUMO-2 [1-46] crypto-thioester (**1**)

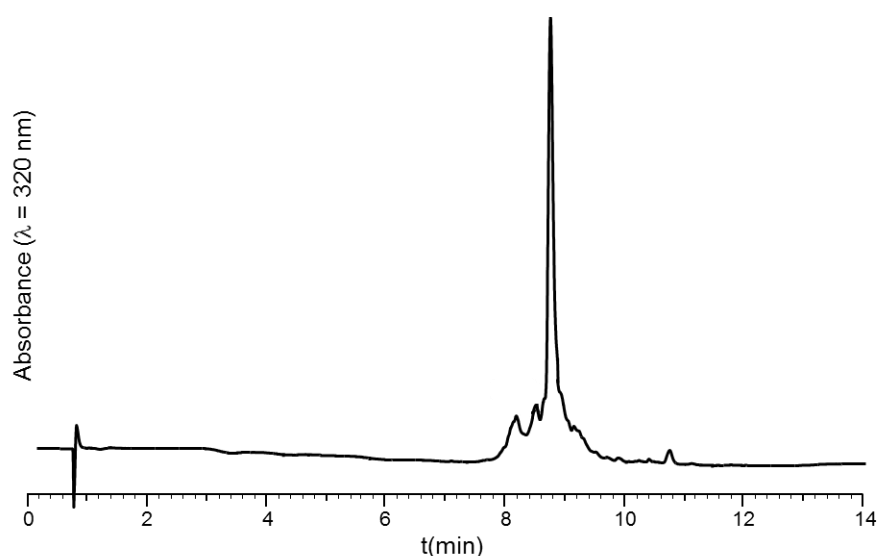
Sequence: XADEKPKEGVKTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLXKAY-(Hnb)C(StBu)-G-NH₂
(X= Norleucine)

Peptidyl resin **1** was synthesized as described in chapter 4 (see p S14-15 of the supporting information part). Briefly, SPPS elongation was conducted following protocol PS1 on Rink amide ChemMatrix (0.45 mmol/g). Hnb was introduced using protocol PS2. D26 and G27 were incorporated through the coupling of the dipeptide Fmoc-Asp(OtBu)-(Dmb)Gly-OH to prevent aspartimide formation during SPPS. To this resin was added a solution of second generation linker precursor **2** (40 mg, 1.5 equiv.) in 1:1 DMF/pyridine (10.5 mM solid-supported peptide concentration), DIPEA (11 μ L 1.8 equiv.) and hydroxybenzotriazole (HOBt) (6 mg, 1.5 equiv.). The mixture was stirred at room temperature for 5 h and the same coupling procedure was repeated one more time. Capping of eventual unreacted amine groups was achieved by treatment with acetic anhydride 60 equiv.), DIEA (15.5 equiv.) and HOBt (1.8 equiv.) in NMP for 7 min. The reaction was monitored by Kaiser Test [85] and analytical scale TFA cleavage (a few beads of resin) to give compound **S1**.



ESI-MS (m/z): [M] calcd. For C₂₆₁H₄₂₂N₇₃O₈₁S₂P₁: 5973.5, found: 5972.5 (average mass, deconvoluted)

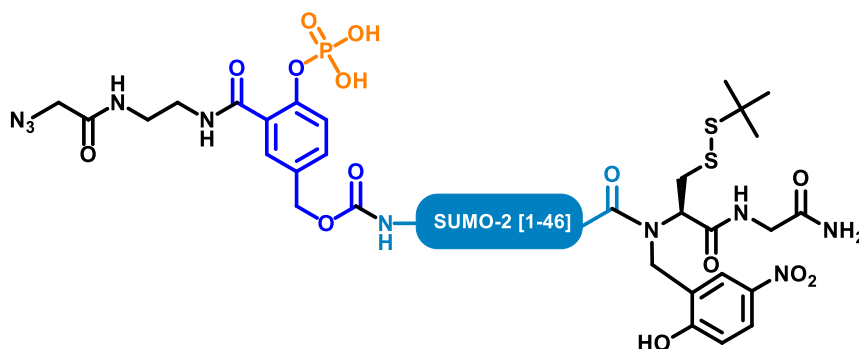
HPLC analysis: t_R = 8.94 min (Aeris Widepore XB-C18 2, gradient: 03-50% B'/A' over 15 min, 60°C)



Supplementary figure S2: HPLC trace of crude **S1**

The Fmoc group of peptidyl resin **3** was deprotected by three successive treatments with 20% piperidine in NMP for 3 min, then $\text{N}_3\text{-CH}_2\text{-COOH}$ (19 μL , 10 equiv.) in NMP (1.5 mL), DIC (39 μL , 10 equiv.) and Oxyma (36 mg, 10 equiv.) in NMP (2 mL) were added and the reactor was stirred for 6 h followed by extensive washing. The same procedure was repeated one more time. Capping of eventual unreacted amine groups was achieved by treatment with acetic anhydride 60 equiv.), DIEA (15.5 equiv.) and HOBt (1.8 equiv.) in NMP for 7 min. The reaction was monitored by Kaiser Test [85] and analytical scale TFA cleavage (a few beads of resin). Cleavage was performed following protocol PS4, and the crude product was purified by semi-preparative HPLC.

➤ Compound **5**

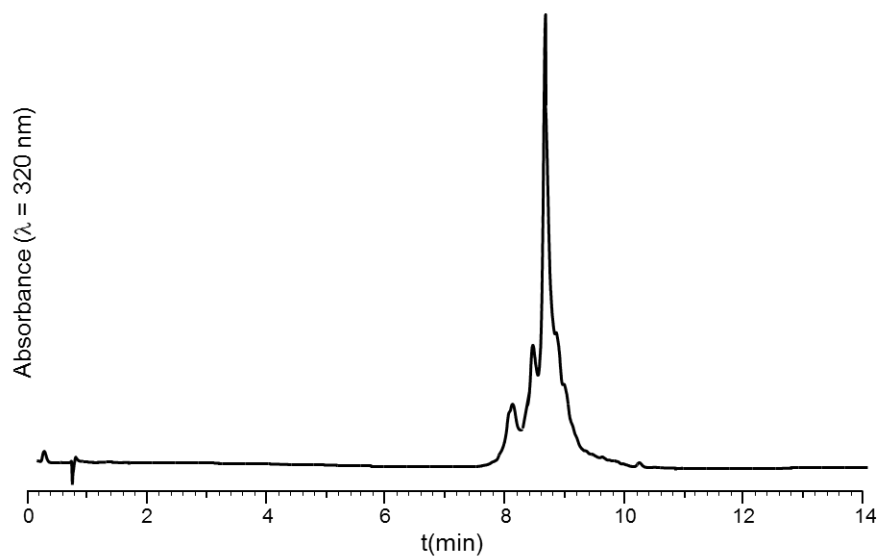


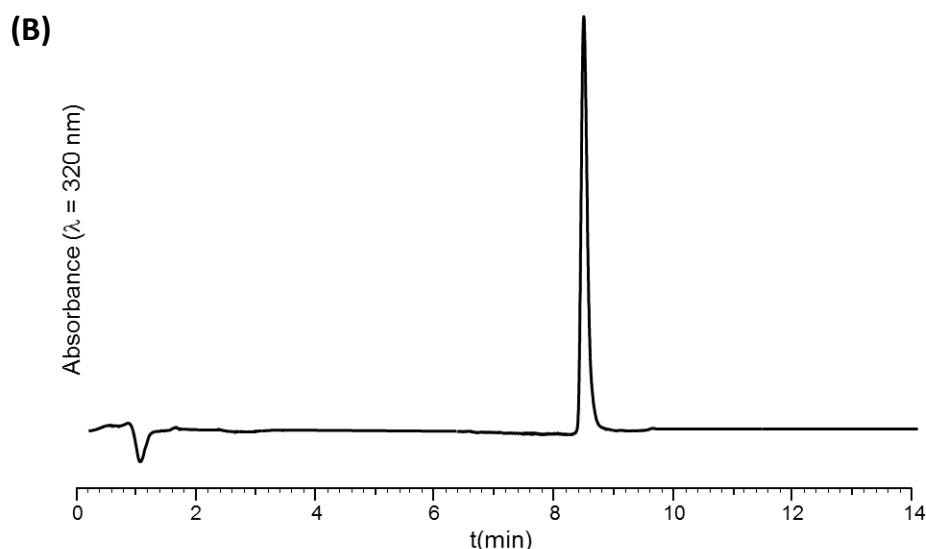
ESI-MS (m/z): [M] calcd. For $\text{C}_{263}\text{H}_{424}\text{N}_{76}\text{O}_{82}\text{S}_2\text{P}_1$: 6057.7, found: 6055.8 (average mass, deconvoluted)

HPLC analysis: t_R = 8.62 min (Aeris Widepore XB-C18 2, gradient: 03-50% B'/A' over 15 min, 60°C)

HPLC purification: [5mg/mL] Nucleosil C18, gradient: 31-35% B/A over 20 min, affording a white solid after lyophilisation (15% yield).

(A)





Supplementary figure S3: HPLC traces of crude (A) and purified (B) compound 5

B. SUMO-2 [47-93]-Propargylamide (S6)

Sequence: H-CERQGLS^XRQIRFRF^DGQPINETDTPAQL^EEEDEDTIDVF^QQQTG-NH-CH₂-C≡CH

^X= Norleucine,

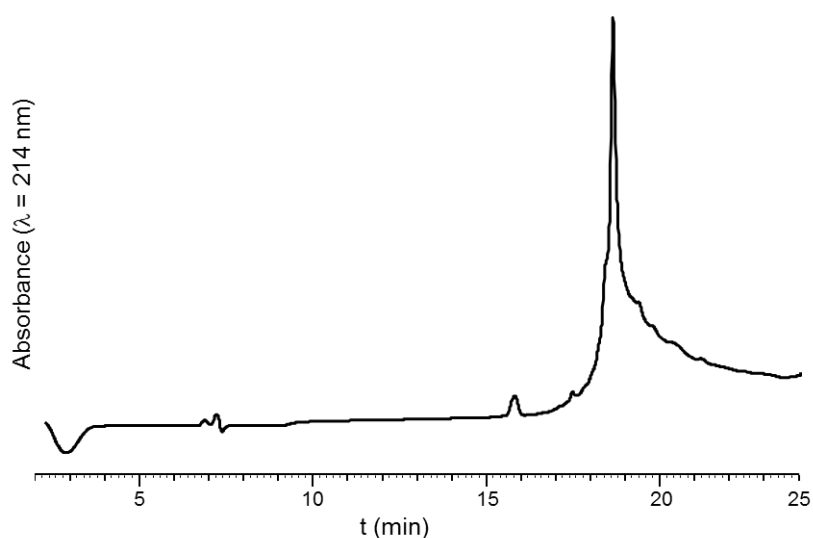
BAL linker was loaded on aminomethyl ChemMatrix® resin (1.25 equiv. BAL linker, 1.25 equiv. HATU, 5 equiv. DIEA in NMP for 2 h). Then the resin was treated with 9:9:2 DMF/MeOH/AcOH mixture (3 mL, 5 min). Next, the resin was washed and propargylamine (5 equiv.) and NaBH₃CN (5 equiv.) were added and the mixture was stirred for 12 h at 50° C. Then, ³⁵Gly was coupled manually (tenfold excess) and the rest of the peptide was synthesized following protocol PS1. ^{D16} and ^{G17} were introduced through the coupling of dipeptide Fmoc-Asp(OtBu)-(Dmb)Gly-OH to prevent aspartimide formation during SPPS (3.5 equiv. dipeptide, 3.4 equiv. HATU, 10 equiv. DIEA in NMP for 2 h). ^{Q88} and ^{Q89} were subjected to double coupling to ensure the completion of the reaction. Cleavage was performed following protocol PS4.

Elongation yield: 83%. Determined by the ratio between the quantity of fluorenylpiperidine released during final Fmoc deprotection and the quantity released during the Fmoc deprotection of the C-terminal Gly residue (UV titration at 301 nm, ε = 7800 L.mol⁻¹.cm⁻¹).

Solubility: **S6** was not soluble at 5 mg/mL and 1 mg/mL in 8:2:0.01 water/MeCN/TFA, which are the conditions used to purify the other segments. We were able to solubilize crude peptide at very low concentrations (0.1 mg/mL and 0.17 mg/mL in 8:2:0.01 and 6:4:0.01 water/MeCN/TFA, respectively). However, we observed a non-conventional HPLC behaviour, and dragging of the product leading to an overlap of the impurities and the desired product.

ESI-MS (*m/z*): [M] calcd. For C₂₂₆H₃₅₄N₆₆O₇₈S₁ 5275.7, found: 5274.9 (average mass, deconvoluted)

HPLC Analysis: t_R = 18.8 min (Jupiter C4, gradient: 10-90% B/A over 20 min)



Supplementary figure S4: HPLC trace of crude peptide **S6**

C. K₆Ades-SUMO-2 [47-93]-propargylamide (**7**)

Sequence:

H-C(H-KKKKKK-Ades)ERQGLSXRQIRFRFDGQPINETDTPAQLXEDEDTIDVFQQQTG-NH-CH₂-C≡CH

X= Norleucine

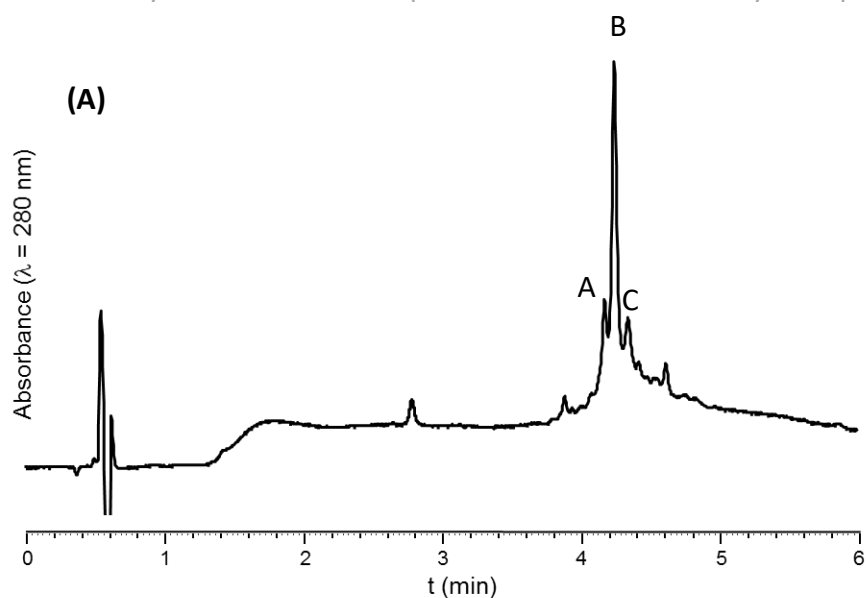
Peptide elongation was performed as described for **S6** up to Glu2. Ades was introduced following protocol PS3. Lys₆ elongation was done following protocol PS1. Cleavage was performed following protocol PS4, and the crude product was purified by semi-preparative HPLC.

Elongation yield: 85%. Determined by the ratio between the quantity of fluorenylpiperidine released during final Fmoc deprotection and the quantity released during the Fmoc deprotection of the C-terminal Gly residue (UV titration at 301 nm, $\epsilon = 7800 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$).

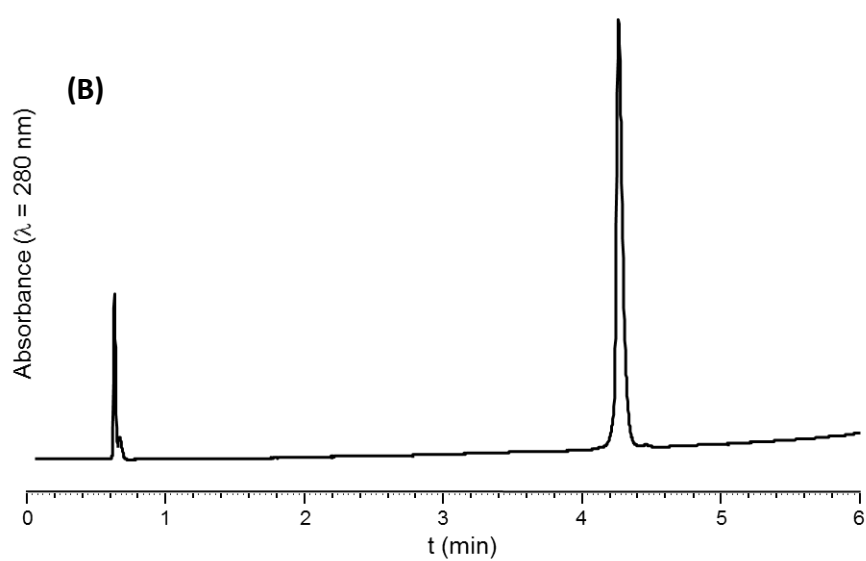
ESI-MS (*m/z*): [M] calcd. For C₂₆₆H₄₃₆N₇₉O₈₄S₂: 6148.9, found: 6147.2 (average mass, deconvoluted)

HPLC analysis: $t_R = 4.36 \text{ min}$ (Chromolith, gradient: 20-50% B/A over 5 min)

HPLC purification: [5mg/mL] Jupiter C4, gradient: 7-42% B/A over 20 min, 35%



Peak (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (4.19)	6035.8	6033.3	Deletion (missing a Lys residue)
B (4.28)	6148.9	6147.2	7
C (4.37)	6110.9	6109.0	Peptide without propargylamide (C-terminal acid)



Supplementary figure S5: HPLC traces and MS analysis of crude (A) and purified (B) compound **7**

D. Model azido-tetrapeptide (9)

Sequence: H-WBAE-NH₂ B= azido-norleucine

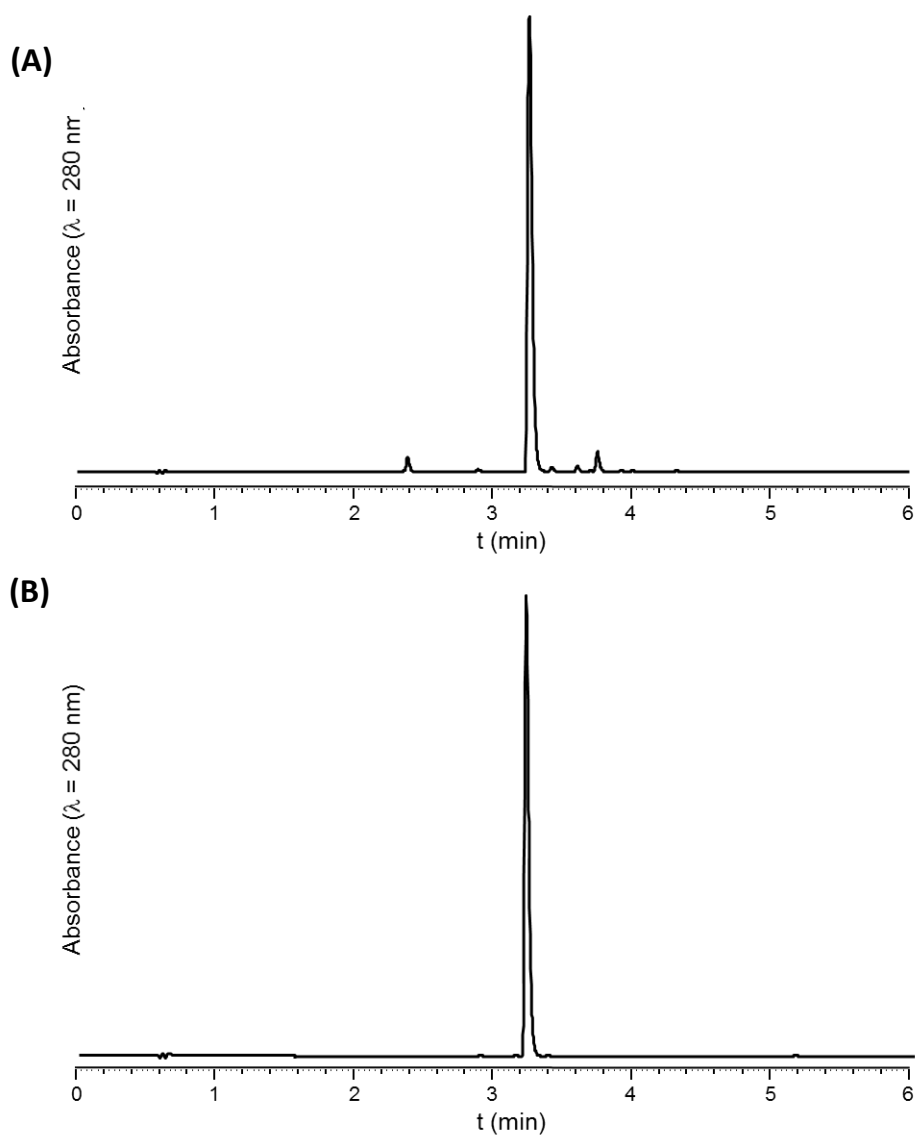
The peptide was synthesized on Rink amide ChemMatrix (0.45 mmol/g) following protocol PS1. Cleavage was performed following protocol PS4, and the crude product was purified by semi-preparative HPLC.

Elongation yield: 83%. Determined by the ratio between the quantity of fluorenylpiperidine released during final Fmoc deprotection and the quantity released during the Fmoc deprotection of the C-terminal Glu residue (UV titration at 301 nm, $\epsilon = 7800 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$).

ESI-MS (m/z): [M] calcd. For C₂₅H₃₅N₉O₆: 557.3, found: 557.2

HPLC analysis: $t_R = 3.31 \text{ min}$ (Chromolith, gradient: 05-50% B/A over 5 min)

HPLC purification: [5mg/mL] Jupiter C4, gradient: 10-20% B/A over 20 min, 65%



Supplementary figure S6: HPLC traces of crude (A) and purified (B) **9**

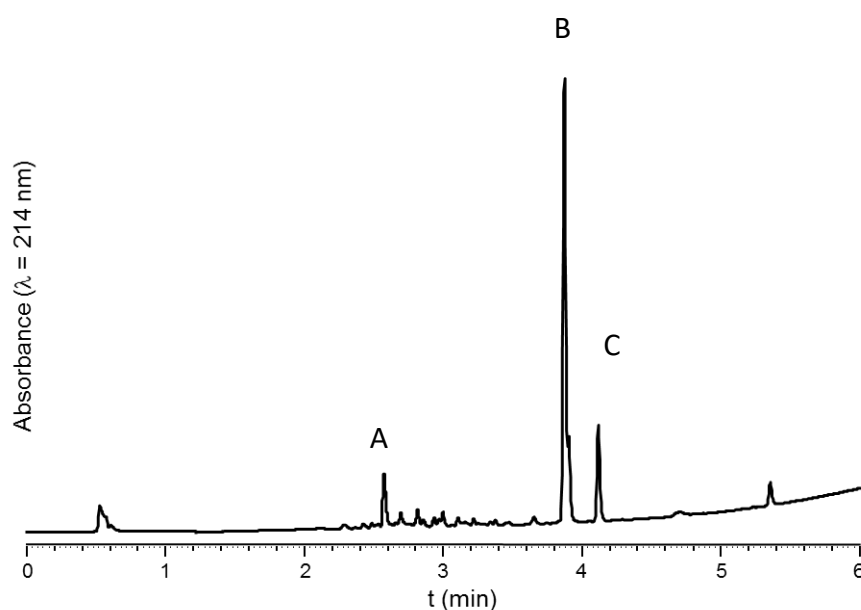
E. PML derived peptide (S7)

Sequence: H- APRKVI**B**XESEE-NH₂ **B**= Lys(N₃) **X**= Norleucine

Elongation yield: 50%. Determined by the ratio between the quantity of fluorenylpiperidine released during final Fmoc deprotection and the quantity released during the Fmoc deprotection of the C-terminal Gln residue (UV titration at 301 nm, $\epsilon = 7800 \text{ L.mol}^{-1}.\text{cm}^{-1}$).

ESI-MS (m/z): [M] calcd. For C₆₁H₁₀₆N₂₀O₁₉: 1424.6, found: 1424.2

HPLC analysis: $t_R = 3.89 \text{ min}$ (Chromolith, gradient: 5-90% B/A over 5 min)



Peak (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (2.57)	1325.1	1324,7	Deletion of valine
B (3.89)	1424.7	1423.2	S7
C (4.15)	1142.9	1142,3	Ac-[4-12]

Supplementary figure S7: HPLC trace and MS analysis of crude **S7**

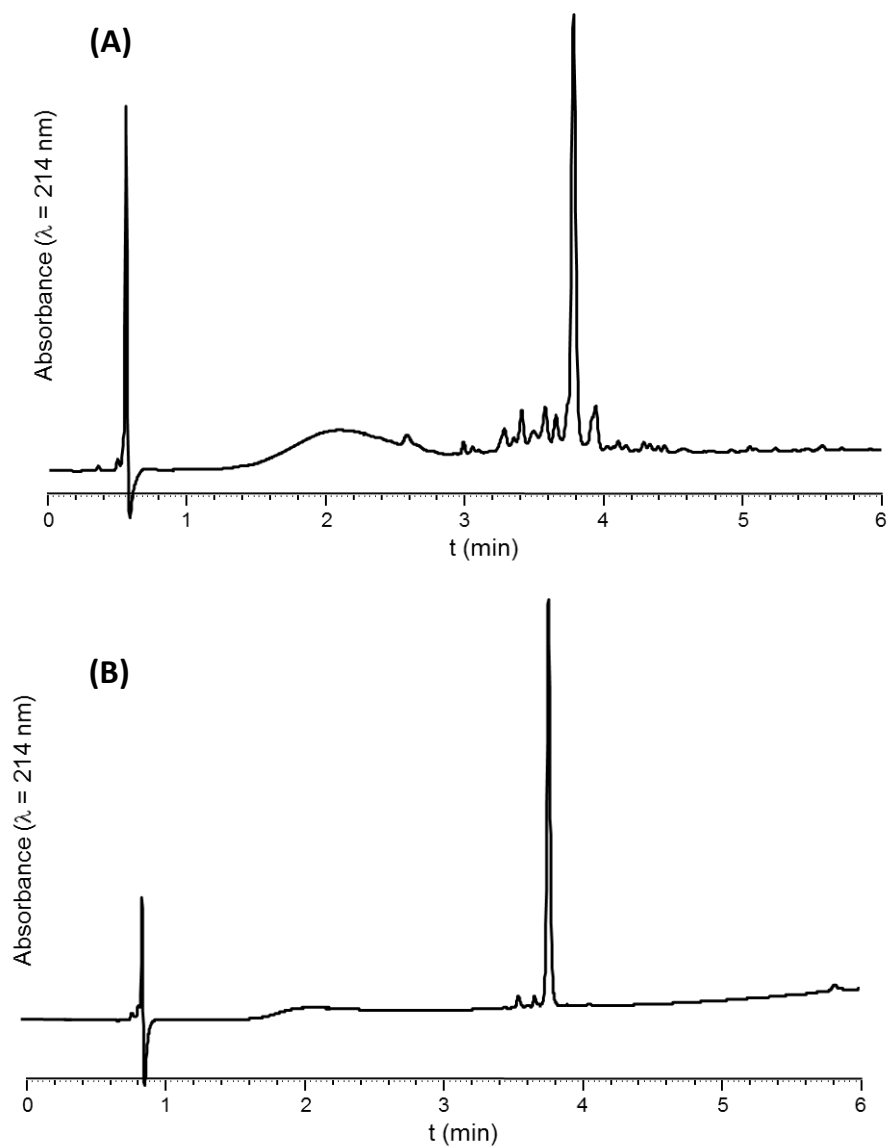
➤ **Coupling of PEG6-biotin to afford product (14)**

(11.7 μmol) of resin was washed with NMP (2 mL, 1 min). Biotin-PEG₆ acid (13.6 mg, 2 equiv.) in NMP (585 μL), HATU (7.8 mg, 1.75 equiv.) in NMP (585 μL), and DIEA (8.1 μL mg, 4 equiv.) were added and the reactor was stirred for 12 h followed by extensive washing. Cleavage was performed following protocol PS4, and the crude product was purified by semi-preparative HPLC.

ESI-MS (m/z): [M] calcd. For C₈₆H₁₄₉N₂₃O₂₈S₁: 1986.3, found: 1984.7

HPLC analysis: $t_R = 3.78$ min (Chromolith, gradient: 05-90% B/A over 5 min)

HPLC purification: [5mg/mL] Jupiter C4, gradient: isocratic 24 B/A over 20 min, to afford a white powder (37%).



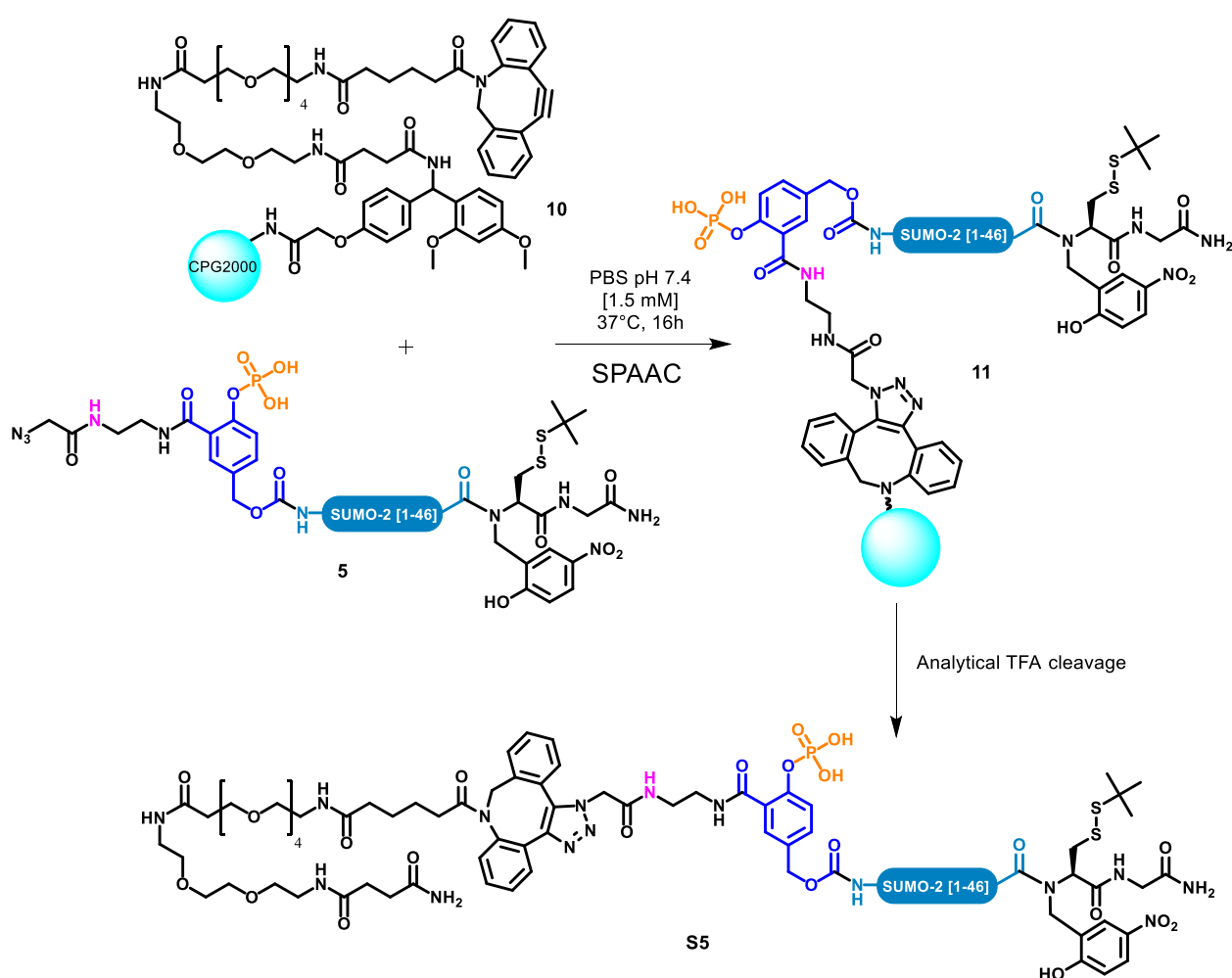
Supplementary figure S8: HPLC traces of crude (A) and purified (B) **14**

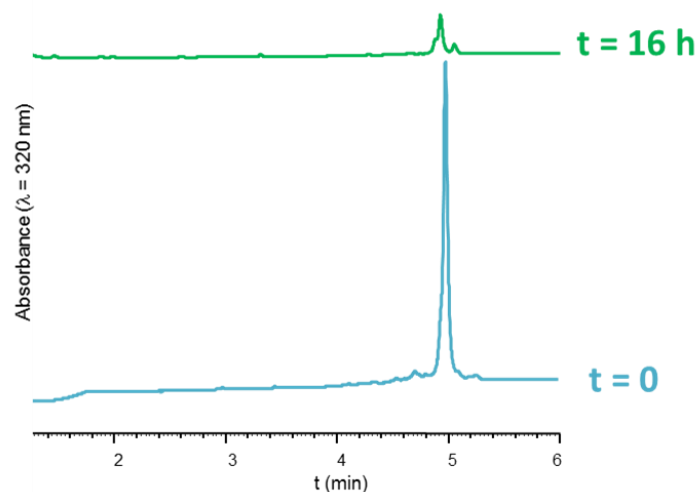
4) SPCL-based synthesis of (15)

➤ Strain-promoted azide/alkyne cycloaddition to graft the azidolinker-derivatized peptide 5 on solid support 10

Cyclooctyne-functionalized CPG2000 support **10** was synthesized as described in [chapter 2](#) (compound **14** in [chapter 2](#) numbering, synthesis described pS54-55 of the supporting information part). Azidolinker-derivatized peptide **5** (0.5 μ mol, 1 mg) was dissolved in 333 μ L of PBS pH 7.4 containing 6 M Gu.HCl (1.5mM peptide concentration), and added to **10** (1.5 μ mol, 3 equiv.). The reaction was stirred for 16 h at 37 °C, and disappearance of **5** was monitored by HPLC. The excess of DBCO was capped by addition of 15 equiv. azidoacetic acid followed by stirring for additional 5 h.

For analytical purpose, an aliquot of support **11** was cleaved (TFA-mediated cleavage of the Rink linker, protocol: PS4) to give compound **S5**.



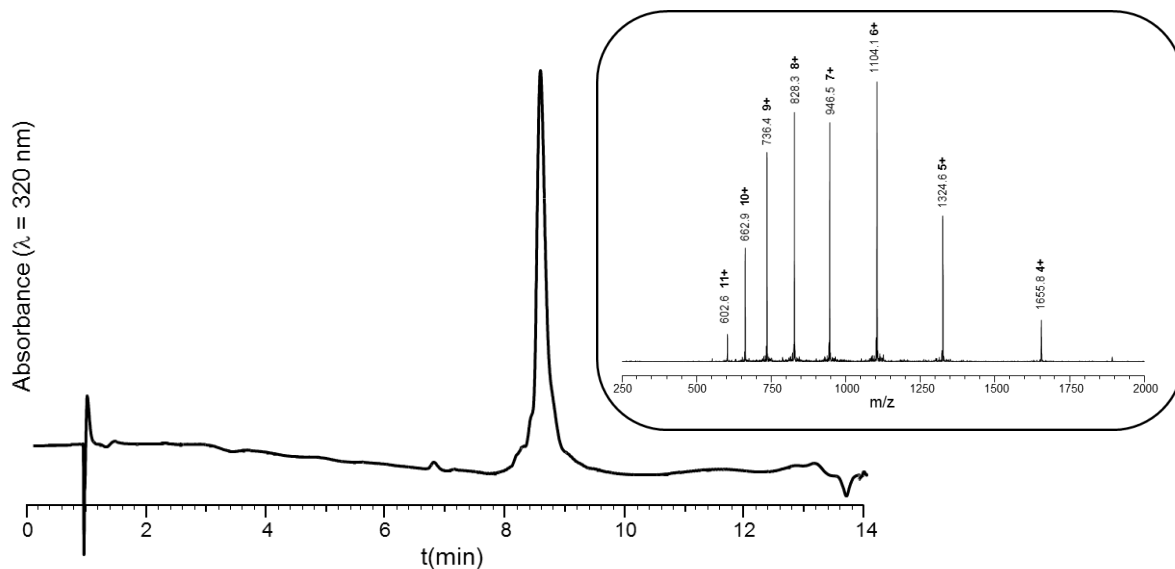


Supplementary figure S9: HPLC monitoring of the SPAAC reaction showing the disappearance of the starting azide **11** (HPLC conditions: Chromolith, gradient: 5-50% B/A over 5 min)

➤ Compound **S5**

LC-MS (m/z): [M] calcd. For $C_{294}H_{461}N_{80}O_{88}S_2P_1$: 6619.4, found: 6618.6

HPLC analysis: t_R = 8.87 min (Aeris Widepore XB-C18 2, gradient: 03-50% B'/A' over 15 min, 60°C)



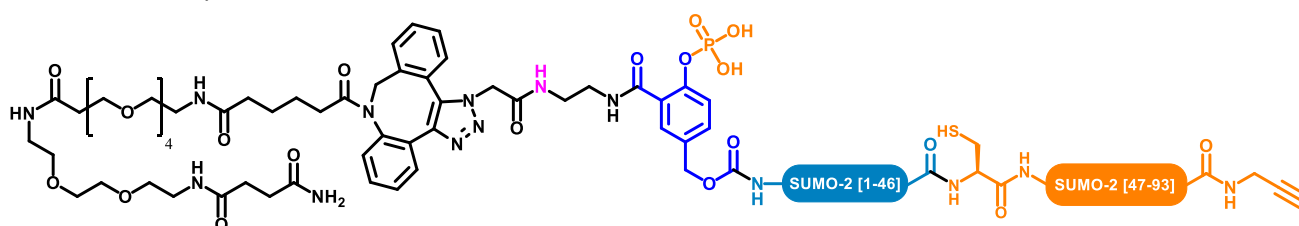
Supplementary figure S10: HPLC trace of released compound **S5** Insert: MS spectrum corresponding to the major peak.

A. First ligation (NCL)

Solid supported crypto-thioester **11** (200 nmol, 5 mM final concentration) was washed (3 x 30 min) with the NCL buffer, and then the buffer was carefully removed leaving the solid support as dry as possible (see protocol PS7). Next cysteinyl peptide **7** (2.5 mg, 400 nmol, 2 equiv., 1 mM final concentration) was dissolved in NCL buffer (40 μ L), added to the solid support and the resulting mixture was gently stirred under inert atmosphere at 37°C for 18 h. Note that it is crucial for the reaction completion to wash the crypto-thioester solid support with the NCL buffer (3 x 20 min) before starting the NCL reaction (see the optimization work below).

For analytical purpose, an aliquot of support **12** was cleaved (TFA-mediated cleavage of the Rink linker, protocol: PS4) to give compound **S6**.

➤ Compound **S6**



ESI -MS (m/z): [M] calcd. For $C_{504}H_{791}N_{142}O_{161}S_1P_1$: 11478.6, found: 11477.5 (average mass, deconvoluted)

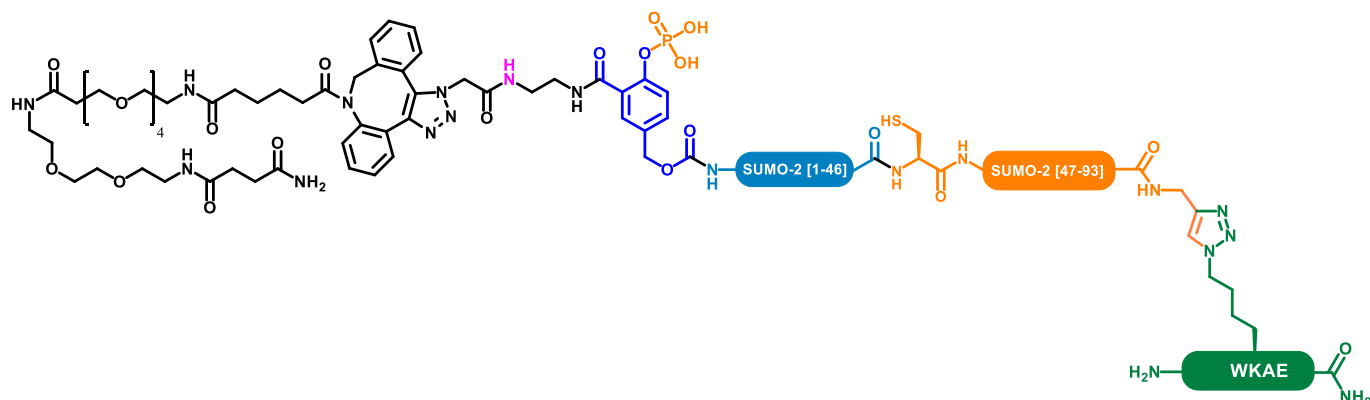
HPLC analysis: t_R = 8.77 min (Aeris Widepore XB-C18 2, gradient: 03-50% B'/A' over 15 min, 60°C)

B. Second ligation (PTL)

A solution of the azide-containing peptide **9** (1 mg, 6 equiv.) dissolved in 160 μ L of a 1:1 mixture of NMP and 200 mM HEPES buffer pH 7.5 was added to a 2 mL microcentrifuge tube containing the azide functionalized solid support **12** (200 nmol). To the suspension were subsequently added a 1 M aqueous aminoguanidine solution (6 μ L, 30 equiv.), a 1 M aqueous tert-butanol solution (6 μ L, 30 equiv.). The tube was sealed with a rubber septum and the resulting mixture was further deoxygenated through several successive vacuum (15 mbar) / argon cycles. Then, a mixture containing copper(I) bromide-dimethylsulfide complex (30 equiv.) and THPTA (40 equiv.) dissolved under argon in 20 μ L of deoxygenated NMP was added, the resulting suspension was further deoxygenated and was stirred for 24 h at 37 °C. The resin was then extensively washed under argon with NMP, de-ionized water and then repeatedly treated (3 \times 1 mL, 30 s) with a pH 7 aqueous buffer containing 6 M guanidinium chloride, 0.1 M EDTA and 0.1 M sodium dihydrogenophosphate. Next the reaction was washed with 6 M guanidinium chloride deoxygenated aqueous solution (pH = 3~4) (3 \times 1 mL, 30 s) then drained and finally extensively washed with deoxygenated water.

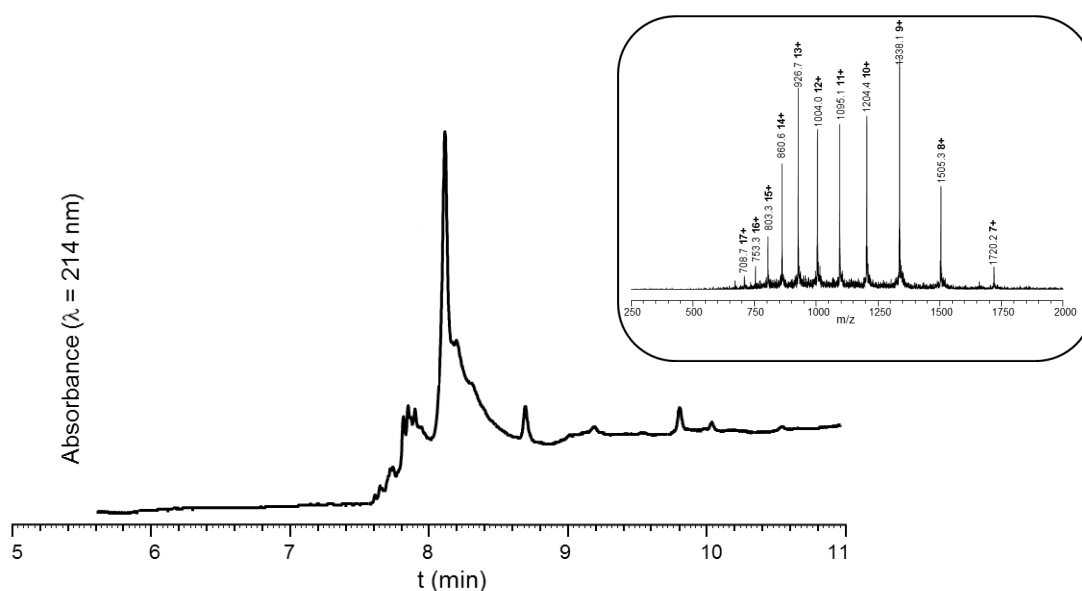
For analytical purpose, an aliquot of support **13** was cleaved (TFA-mediated cleavage of the Rink linker, protocol: PS4) to give compound **S7**.

➤ Compound **S7**



ESI -MS (m/z): [M] calcd. For $C_{529}H_{826}N_{151}O_{167}S_1P_1$: 12036.2, found: 12035.4 (average mass, deconvoluted)

HPLC analysis: t_R = 8.31 min (Aeris Widespore XB-C18 2, gradient: 03-50% B'/A' over 15 min, 60°C)



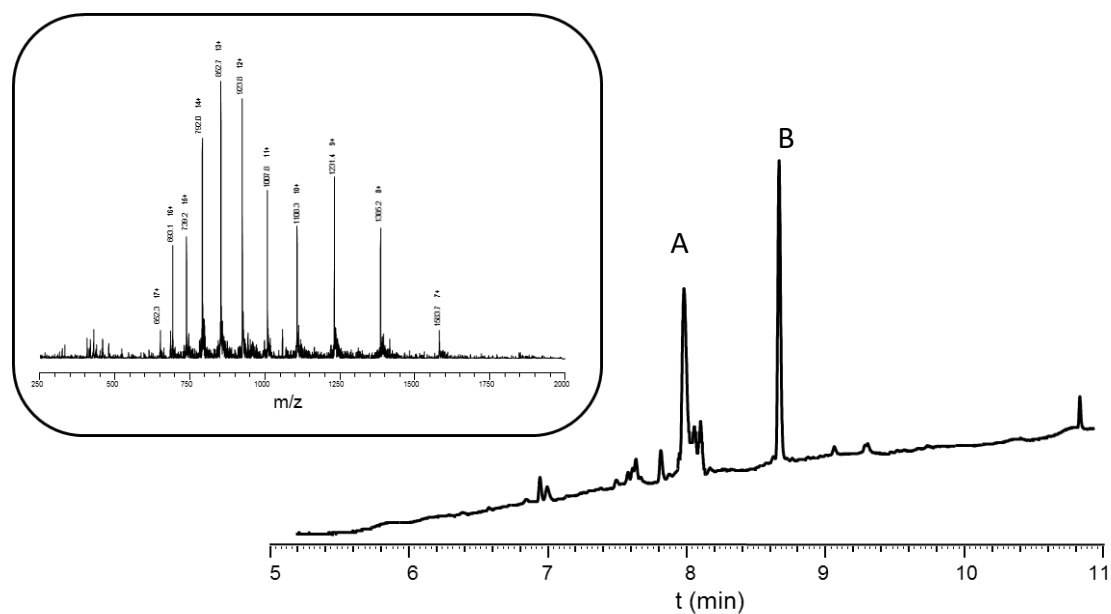
Supplementary figure S11: HPLC trace released compound **S7**. Insert: MS spectrum corresponding to the major peak.

C. Enzymatic cleavage of the SPCL linker to afford (**15**)

Enzymatic hydrolysis was carried out with lambda phosphatase. The targeting device grafted on CGP2000 solid support **13** was incubated at 30°C with the enzyme (4 enzyme unit/10nmol of peptide) in the PMP buffer ([100 μ M] of the peptide) for 24 h.

ESI -MS (m/z): [M] calcd. For $C_{485}H_{774}N_{142}O_{153}S_1$: 11074.4, found: 11073.2 (average mass, deconvoluted)

HPLC analysis: t_R = 7.98 min (Aeris Widespore XB-C18 2, gradient: 03-50% B'/A' over 15 min, 60°C)



Peak (t_R (min))	[M] (m/z) calcd.	[M] (m/z) found	Attributed to
A (7.98)	11074.4	11073.2	15
B (8.77)	-	1121.7	Non peptidic compound*

*: Note that peaks B is also observed when treating the starting alkyne solid support **10** under the same conditions, and thus does not correspond to a peptide.

Supplementary figure S12: HPLC trace of enzymatically-released **15**. Insert: MS spectrum corresponding to the major peak.

V- References

1. Hruby, V.J. (2003) Peptide Science: Exploring the Use of Chemical Principles and Interdisciplinary Collaboration for Understanding Life Processes. *J. Med. Chem.*, **46** (20), 4215–4231.
2. Galibert, M., Piller, V., Piller, F., Aucagne, V., and Delmas, A.F. (2015) Combining triazole ligation and enzymatic glycosylation on solid phase simplifies the synthesis of very long glycoprotein analogues. *Chem. Sci.*, **6** (6), 3617–3623.
3. Inokuchi, E., Yamada, A., Hozumi, K., Tomita, K., Oishi, S., Ohno, H., Nomizu, M., and Fujii, N. (2011) Design and synthesis of amidine-type peptide bond isosteres: application of nitrile oxide derivatives as active ester equivalents in peptide and peptidomimetics synthesis. *Org. Biomol. Chem.*, **9** (9), 3421.
4. Jones, R.C.F., and Ward, G.J. (1988) Amide bond isosteres: Imidazolines in pseudopeptide chemistry. *Tetrahedron Lett.*, **29** (31), 3853–3856.
5. Abraham, R.J., Ellison, S.L.R., Schonholzer, P., and Thomas, W.A. (1986) A theoretical and crystallographic study of the geometries and conformations of fluoro-olefins as peptide analogues. *Tetrahedron*, **42** (7), 2101–2110.
6. Christos, T.E., Arvanitis, A., Cain, G.A., Johnson, A.L., Pottorf, R.S., Tam, S.W., and Schmidt, W.K. (1993) Stable isosteres of neurotensin c-terminal pentapeptides derived by modification of the amide function. *Bioorg. Med. Chem. Lett.*, **3** (6), 1035–1040.
7. Howard, J.A.K., Hoy, V.J., O'Hagan, D., and Smith, G.T. (1996) How good is fluorine as a hydrogen bond acceptor? *Tetrahedron*, **52** (38), 12613–12622.
8. Lin, J., Toscano, P.J., and Welch, J.T. (1998) Inhibition of dipeptidyl peptidase IV by fluoroolefin-containing N-peptidyl-O-hydroxylamine peptidomimetics. *Proc. Natl. Acad. Sci. USA*, **95** (24), 14020–14024.
9. Choudhary, A., and Raines, R.T. (2011) An Evaluation of Peptide-Bond Isosteres. *ChemBioChem*, **12** (12), 1801–1807.
10. Bagiński, M., Piela, L., and Skolnick, J. (1993) The ethylene group as a peptide bond mimicking unit: A theoretical conformational analysis. *J. Comput. Chem.*, **14** (4), 471–477.
11. Valverde, I.E., Lecaille, F., Lalmanach, G., Aucagne, V., and Delmas, A.F. (2012) Synthesis of a Biologically Active Triazole-Containing Analogue of Cystatin A Through Successive Peptidomimetic Alkyne-Azide Ligations. *Angew. Chem. Int. Ed.*, **51** (3), 718–722.

12. Davis, M.R., Singh, E.K., Wahyudi, H., Alexander, L.D., Kunicki, J.B., Nazarova, L.A., Fairweather, K.A., Giltrap, A.M., Jolliffe, K.A., and McAlpine, S.R. (2012) Synthesis of sansalvamide A peptidomimetics: triazole, oxazole, thiazole, and pseudoproline containing compounds. *Tetrahedron*, **68** (4), 1029–1051.
13. Perez, J.J. (2018) Designing Peptidomimetics. *Curr. Top. Med. Chem.*, **18** (7), 566–590.
14. Rostovtsev, V.V., Green, L.G., Fokin, V.V., and Sharpless, K.B. (2002) A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective ‘ligation’ of azides and terminal alkynes. *Angew. Chem. Int. Ed. Engl.*, **41** (14), 2596–2599.
15. Tornøe, C.W., Christensen, C., and Meldal, M. (2002) Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.*, **67** (9), 3057–3064.
16. Huisgen, R. (1963) 1,3-Dipolar Cycloadditions. Past and Future. *Angew. Chem. Int. Ed. Engl.*, **2** (10), 565–598.
17. Brik, A., Alexandratos, J., Lin, Y.-C., Elder, J.H., Olson, A.J., Wlodawer, A., Goodsell, D.S., and Wong, C.-H. (2005) 1,2,3-Triazole as a Peptide Surrogate in the Rapid Synthesis of HIV-1 Protease Inhibitors. *ChemBioChem*, **6** (7), 1167–1169.
18. Hou, J., Liu, X., Shen, J., Zhao, G., and Wang, P.G. (2012) The impact of click chemistry in medicinal chemistry. *Expert Opin. Drug Discov.*, **7** (6), 489–501.
19. Jacobsen, Ø., Maekawa, H., Ge, N.-H., Görbitz, C.H., Rongved, P., Ottersen, O.P., Amiry-Moghaddam, M., and Klaveness, J. (2011) Stapling of a 310-Helix with Click Chemistry. *J. Org. Chem.*, **76** (5), 1228–1238.
20. Ko, E., Liu, J., Perez, L.M., Lu, G., Schaefer, A., and Burgess, K. (2011) Universal Peptidomimetics. *J. Am. Chem. Soc.*, **133** (3), 462.
21. Valverde, I.E., Bauman, A., Kluba, C.A., Vomstein, S., Walter, M.A., and Mindt, T.L. (2013) 1,2,3-Triazoles as Amide Bond Mimics: Triazole Scan Yields Protease-Resistant Peptidomimetics for Tumor Targeting. *Angew. Chem. Int. Ed.*, **52** (34), 8957–8960.
22. Valverde, I.E., Vomstein, S., and Mindt, T.L. (2016) Toward the Optimization of Bombesin-Based Radiotracers for Tumor Targeting. *J. Med. Chem.*, **59** (8), 3867–3877.
23. Mascarin, A., Valverde, I.E., Vomstein, S., and Mindt, T.L. (2015) 1,2,3-Triazole Stabilized Neurotensin-Based Radiopeptidomimetics for Improved Tumor Targeting. *Bioconjugate Chem.*, **26** (10), 2143–2152.

24. Valverde, I.E., Vomstein, S., Fischer, C.A., Mascarin, A., and Mindt, T.L. (2015) Probing the Backbone Function of Tumor Targeting Peptides by an Amide-to-Triazole Substitution Strategy. *J. Med. Chem.*, **58** (18), 7475–7484.
25. Valverde, I.E., Huxol, E., and Mindt, T.L. (2014) Radiolabeled antagonistic bombesin peptidomimetics for tumor targeting: Radiopeptidomimetics. *J. Label Compd. Radiopharm*, **57** (4), 275–278.
26. Lee, J.-H., Miele, M.E., Hicks, D.J., Phillips, K.K., Trent, J.M., Weissman, B.E., and Welch, D.R. (1996) KiSS-1, a Novel Human Malignant Melanoma Metastasis-Suppressor Gene. *J. Natl. Cancer Inst*, **88** (23), 1731–1737.
27. Beltramo, M., Robert, V., Galibert, M., Madinier, J.-B., Marceau, P., Dardente, H., Decourt, C., De Roux, N., Lomet, D., Delmas, A.F., Caraty, A., and Aucagne, V. (2015) Rational Design of Triazololipopeptides Analogs of Kisspeptin Inducing a Long-Lasting Increase of Gonadotropins. *J. Med. Chem.*, **58** (8), 3459–3470.
28. Decourt, C., Robert, V., Anger, K., Galibert, M., Madinier, J.-B., Liu, X., Dardente, H., Lomet, D., Delmas, A.F., Caraty, A., Herbison, A.E., Anderson, G.M., Aucagne, V., and Beltramo, M. (2016) A synthetic kisspeptin analog that triggers ovulation and advances puberty. *Sci Rep*, **6** (1), 26908.
29. Galibert, M., Wartenberg, M., Lecaille, F., Saidi, A., Mavel, S., Joulin-Giet, A., Korkmaz, B., Brömme, D., Aucagne, V., Delmas, A.F., and Lalmanach, G. (2018) Substrate-derived triazolo- and azapeptides as inhibitors of cathepsins K and S. *Eur. J. Med. Chem*, **144**, 201–210.
30. Epinette, C., Croix, C., Jaquillard, L., Marchand-Adam, S., Kellenberger, C., Lalmanach, G., Cadene, M., Viaud-Massuard, M.-C., Gauthier, F., and Korkmaz, B. (2012) A selective reversible azapeptide inhibitor of human neutrophil proteinase 3 derived from a high affinity FRET substrate. *Biochemical Pharmacology*, **83** (6), 788–796.
31. Voet, D., Voet, J.G., and Pratt, C.W. (2006) *Fundamentals of biochemistry: life at the molecular level*, Wiley, Hoboken, N.J.
32. Goldstein, G., Scheid, M., Hammerling, U., Schlesinger, D.H., Niall, H.D., and Boyse, E.A. (1975) Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. *Proc. Natl. Acad. Sci. USA*, **72** (1), 11–15.

33. Ciechanover, A. (2005) Intracellular Protein Degradation: From a Vague Idea, through the Lysosome and the Ubiquitin-Proteasome System, and onto Human Diseases and Drug Targeting (Nobel Lecture). *Angew. Chem. Int. Ed.*, **44** (37), 5944–5967.
34. Li, W., and Ye, Y. (2008) Polyubiquitin chains: functions, structures, and mechanisms. *Cell. Mol. Life Sci.*, **65** (15), 2397–2406.
35. Love, K.R., Catic, A., Schlieker, C., and Ploegh, H.L. (2007) Mechanisms, biology and inhibitors of deubiquitinating enzymes. *Nat Chem Biol*, **3** (11), 697–705.
36. Reyes-Turcu, F.E., and Wilkinson, K.D. (2009) Polyubiquitin Binding and Disassembly By Deubiquitinating Enzymes. *Chem. Rev.*, **109** (4), 1495–1508.
37. Komander, D., Clague, M.J., and Urbé, S. (2009) Breaking the chains: structure and function of the deubiquitinases. *Nat Rev Mol Cell Biol*, **10** (8), 550–563.
38. Flierman, D., van der Heden van Noort, G.J., Ekkebus, R., Geurink, P.P., Mevissen, T.E.T., Hospenthal, M.K., Komander, D., and Ovaa, H. (2016) Non-hydrolyzable Diubiquitin Probes Reveal Linkage-Specific Reactivity of Deubiquitylating Enzymes Mediated by S2 Pockets. *Cell Chem. Biol*, **23** (4), 472–482.
39. Zhang, X., Smits, A.H., van Tilburg, G.B.A., Jansen, P.W.T.C., Makowski, M.M., Ovaa, H., and Vermeulen, M. (2017) An Interaction Landscape of Ubiquitin Signaling. *Molecular Cell*, **65** (5), 941-955.e8.
40. Weikart, N.D., Sommer, S., and Mootz, H.D. (2012) Click synthesis of ubiquitin dimer analogs to interrogate linkage-specific UBA domain binding. *Chem. Commun.*, **48** (2), 296–298.
41. Zhao, X., Scheffner, M., and Marx, A. (2019) Assembly of branched ubiquitin oligomers by click chemistry. *Chem. Commun.*, **55** (87), 13093–13095.
42. Eger, S., Scheffner, M., Marx, A., and Rubini, M. (2010) Synthesis of Defined Ubiquitin Dimers. *J. Am. Chem. Soc.*, **132** (46), 16337–16339.
43. Zhao, X., Lutz, J., Höllmüller, E., Scheffner, M., Marx, A., and Stengel, F. (2017) Identification of Proteins Interacting with Ubiquitin Chains. *Angew. Chem. Int. Ed.*, **56** (49), 15764–15768.
44. Schneider, D., Schneider, T., Rösner, D., Scheffner, M., and Marx, A. (2013) Improving bioorthogonal protein ubiquitylation by click reaction. *Bioorganic & Medicinal Chemistry*, **21** (12), 3430–3435.

45. Eger, S., Castrec, B., Hübscher, U., Scheffner, M., Rubini, M., and Marx, A. (2011) Generation of a Mono-ubiquitinated PCNA Mimic by Click Chemistry. *ChemBioChem*, **12** (18), 2807–2812.
46. Schneider, D., Schneider, T., Aschenbrenner, J., Mortensen, F., Scheffner, M., and Marx, A. (2016) Anionic surfactants enhance click reaction-mediated protein conjugation with ubiquitin. *Bioorg. Med. Chem.*, **24** (5), 995–1001.
47. Schneider, T., Schneider, D., Rösner, D., Malhotra, S., Mortensen, F., Mayer, T.U., Scheffner, M., and Marx, A. (2014) Dissecting Ubiquitin Signaling with Linkage-Defined and Protease Resistant Ubiquitin Chains. *Angew. Chem. Int. Ed.*, **53** (47), 12925–12929.
48. Zhao, X., Mißun, M., Schneider, T., Müller, F., Lutz, J., Scheffner, M., Marx, A., and Kovermann, M. (2019) Structural and Dynamic Characterization of Artificially Linked Ubiquitin Dimers by NMR spectroscopy. *ChemBioChem*, cbic.201900146.
49. Rösner, D., Schneider, T., Schneider, D., Scheffner, M., and Marx, A. (2015) Click chemistry for targeted protein ubiquitylation and ubiquitin chain formation. *Nat Protoc*, **10** (10), 1594–1611.
50. Hickey, C.M., Wilson, N.R., and Hochstrasser, M. (2012) Function and regulation of SUMO proteases. *Nat Rev Mol Cell Biol*, **13** (12), 755–766.
51. Matunis, M.J., Coutavas, E., and Blobel, G. (1996) A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J. Cell Biol.*, **135** (6), 1457–1470.
52. Johnson, E.S. (2004) Protein Modification by SUMO. *Annu. Rev. Biochem.*, **73** (1), 355–382.
53. Princz, A., and Tavernarakis, N. (2017) The role of SUMOylation in ageing and senescent decline. *Mech Ageing Dev*, **162**, 85–90.
54. Da Silva-Ferrada, E., Lopitz-Otsoa, F., Lang, V., Rodríguez, M.S., and Matthiesen, R. (2012) Strategies to Identify Recognition Signals and Targets of SUMOylation. *Biochemistry Research International*, **2012**, 1–16.
55. Mukhopadhyay, D., and Dasso, M. (2007) Modification in reverse: the SUMO proteases. *Trends Biochem. Sci.*, **32** (6), 286–295.
56. Yang, Y., He, Y., Wang, X., liang, Z., He, G., Zhang, P., Zhu, H., Xu, N., and Liang, S. (2017) Protein SUMOylation modification and its associations with disease. *Open Biol.*, **7** (10), 170167.

57. Bohren, K.M., Nadkarni, V., Song, J.H., Gabbay, K.H., and Owerbach, D. (2004) A M55V Polymorphism in a Novel *SUMO* Gene (*SUMO-4*) Differentially Activates Heat Shock Transcription Factors and Is Associated with Susceptibility to Type I Diabetes Mellitus. *J. Biol. Chem.*, **279** (26), 27233–27238.
58. Melnyk, O., and Vicogne, J. (2016) Total chemical synthesis of SUMO proteins. *Tetrahedron Letters*, **57** (39), 4319–4324.
59. Guo, C., and Henley, J.M. (2014) Wrestling with stress: Roles of protein SUMOylation and deSUMOylation in cell stress response: SUMOylation in Cell Stress. *IUBMB Life*, **66** (2), 71–77.
60. Mulder, M.P.C., Merks, R., Witting, K.F., Hameed, D.S., El Atmioui, D., Lelieveld, L., Liebelt, F., Neefjes, J., Berlin, I., Vertegaal, A.C.O., and Ova, H. (2018) Total Chemical Synthesis of SUMO and SUMO-Based Probes for Profiling the Activity of SUMO-Specific Proteases. *Angew. Chem. Int. Ed.*, **57** (29), 8958–8962.
61. Haack, T., and Mutter, M. (1992) Serine derived oxazolidines as secondary structure disrupting, solubilizing building blocks in peptide synthesis. *Tetrahedron Lett.*, **33** (12), 1589–1592.
62. Blaakmeer, J., Tjisse-Klasen, T., and Tesser, G.I. (2009) Enhancement of solubility by temporary dimethoxybenzyl-substitution of peptide bonds: Towards the synthesis of defined oligomers of alanine and of lysylglutamyl-glycine. *Int. J. Pept. Protein Res.*, **37** (6), 556–564.
63. Boll, E., Drobecq, H., Ollivier, N., Raibaut, L., Desmet, R., Vicogne, J., and Melnyk, O. (2014) A novel PEG-based solid support enables the synthesis of >50 amino-acid peptide thioesters and the total synthesis of a functional SUMO-1 peptide conjugate. *Chem. Sci.*, **5** (5), 2017–2022.
64. Boll, E., Drobecq, H., Ollivier, N., Blanpain, A., Raibaut, L., Desmet, R., Vicogne, J., and Melnyk, O. (2015) One-pot chemical synthesis of small ubiquitin-like modifier protein–peptide conjugates using bis (2-sulfanylethyl)amido peptide latent thioester surrogates. *Nat Protoc*, **10** (2), 269–292.
65. Drobecq, H., Boll, E., Sénéchal, M., Desmet, R., Saliou, J.-M., Lacapère, J.-J., Mougél, A., Vicogne, J., and Melnyk, O. (2016) A Central Cysteine Residue Is Essential for the Thermal Stability and Function of SUMO-1 Protein and SUMO-1 Peptide–Protein Conjugates. *Bioconjugate Chem.*, **27** (6), 1540–1546.

66. Ajish Kumar, K.S., Haj-Yahya, M., Olschewski, D., Lashuel, H.A., and Brik, A. (2009) Highly Efficient and Chemoselective Peptide Ubiquitylation. *Angew. Chem. Int. Ed.*, **48** (43), 8090–8094.
67. Bouchenna, J., Sénéchal, M., Drobecq, H., Vicogne, J., and Melnyk, O. (2019) Total Chemical Synthesis of All SUMO-2/3 Dimer Combinations. *Bioconjugate Chem.*, **30** (11), 2967–2973.
68. Bondalapati, S., Eid, E., Mali, S.M., Wolberger, C., and Brik, A. (2017) Total chemical synthesis of SUMO-2-Lys63-linked diubiquitin hybrid chains assisted by removable solubilizing tags. *Chem. Sci.*, **8** (5), 4027–4034.
69. Bode, J.W., Fox, R.M., and Baucom, K.D. (2006) Chemoselective Amide Ligations by Decarboxylative Condensations of *N*-Alkylhydroxylamines and α -Ketoacids. *Angew. Chem. Int. Ed.*, **45** (8), 1248–1252.
70. Wucherpfennig, T.G., Pattabiraman, V.R., Limberg, F.R.P., Ruiz-Rodríguez, J., and Bode, J.W. (2014) Traceless Preparation of C-Terminal α -Ketoacids for Chemical Protein Synthesis by α -Ketoacid-Hydroxylamine Ligation: Synthesis of SUMO2/3. *Angew. Chem. Int. Ed.*, **53** (45), 12248–12252.
71. Baldauf, S., Ogunkoya, A.O., Boross, G.N., and Bode, J.W. (2020) Aspartic Acid Forming α -Ketoacid–Hydroxylamine (KAHA) Ligations with (*S*)-4,4-Difluoro-5-oxaproline. *J. Org. Chem.*, **85** (3), 1352–1364.
72. Weikart, N.D., and Mootz, H.D. (2010) Generation of Site-Specific and Enzymatically Stable Conjugates of Recombinant Proteins with Ubiquitin-Like Modifiers by the Cul-Catalyzed Azide-Alkyne Cycloaddition. *ChemBioChem*, **11** (6), 774–777.
73. Sommer, S., Weikart, N.D., Brockmeyer, A., Janning, P., and Mootz, H.D. (2011) Expanded Click Conjugation of Recombinant Proteins with Ubiquitin-Like Modifiers Reveals Altered Substrate Preference of SUMO2-Modified Ubc9. *Angew. Chem. Int. Ed.*, **50** (42), 9888–9892.
74. Chin, J.W., Santoro, S.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002) Addition of *p*-Azido- L -phenylalanine to the Genetic Code of *Escherichia coli*. *J. Am. Chem. Soc.*, **124** (31), 9026–9027.
75. van Treel, N.D., and Mootz, H.D. (2014) SUMOylated RanGAP1 prepared by click chemistry: CLICK SUMOylation. *J. Pept. Sci.*, **20** (2), 121–127.

76. Chu, G.-C., Pan, M., Li, J., Liu, S., Zuo, C., Tong, Z.-B., Bai, J.-S., Gong, Q., Ai, H., Fan, J., Meng, X., Huang, Y.-C., Shi, J., Deng, H., Tian, C., Li, Y.-M., and Liu, L. (2019) Cysteine-Aminoethylation-Assisted Chemical Ubiquitination of Recombinant Histones. *J. Am. Chem. Soc.*, **141** (8), 3654–3663.
77. Loibl, S.F., Harpaz, Z., and Seitz, O. (2015) A Type of Auxiliary for Native Chemical Peptide Ligation beyond Cysteine and Glycine Junctions. *Angew. Chem. Int. Ed.*, **54** (50), 15055–15059.
78. Dardashti, R.N., Kumar, S., Sternisha, S.M., Reddy, P.S., Miller, B.G., and Metanis, N. (2020) Selenolysine: A New Tool for Traceless Isopeptide Bond Formation. *Chem. Eur. J.*, **26** (22), 4952–4957.
79. Bergoug, M. (2020) Etude de la SUMOylation de la neurofibromine, la protéine responsable de la neurofibromatose de type 1.
80. Messiaen, L.M., Callens, T., Mortier, G., Beysen, D., Vandenbroucke, I., Van Roy, N., Speleman, F., and Paepe, A.D. (2000) Exhaustive mutation analysis of the NF1 gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects. *Hum. Mutat.*, **15** (6), 541–555.
81. Terrier, V.P., Adihou, H., Arnould, M., Delmas, A.F., and Aucagne, V. (2016) A straightforward method for automated Fmoc-based synthesis of bio-inspired peptide crypto-thioesters. *Chem. Sci.*, **7** (1), 339–345.
82. Bouchenna, J., Sénéchal, M., Drobecq, H., Vicogne, J., and Melnyk, O. (2019) The Problem of Aspartimide Formation During Protein Chemical Synthesis Using SEA-Mediated Ligation.
83. Presolski, S.I., Hong, V.P., and Finn, M.G. (2011) Copper-Catalyzed Azide–Alkyne Click Chemistry for Bioconjugation. *Curr Protoc Chem Biol*, **3** (4), 153–162.
84. Hong, V., Presolski, S.I., Ma, C., and Finn, M.G. (2009) Analysis and Optimization of Copper-Catalyzed Azide–Alkyne Cycloaddition for Bioconjugation. *Angew. Chem. Int. Ed.*, **48** (52), 9879–9883.
85. Kaiser, E., Colecott, R.L., Bossinger, C.D., and Cook, P.I. (1970) Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal Biochem*, **34** (2), 595–598.

Conclusion and perspectives

The total synthesis of long proteins requires the assembly of multiple segments through successive ligations. The need of numerous intermediate purifications is a strong limitation, as this leads generally to very low overall yield. An early-recognized solution to this problem is the assembly of the protein on a solid support. Although extremely appealing, solid phase chemical ligation (SPCL) has been so far essentially limited to proof-of-concepts, one of the main reasons being the difficulty to immobilize a first peptide segment on a water-compatible solid support through a dedicated SPCL linker. A wide range of linkers have been developed but their cleavage requires chemical conditions that proved to be incompatible with sensitive protein targets. In the present manuscript, research was focused on the development of new chemo-enzymatic methodologies for the total synthesis of proteins through solid-supported chemical ligation.

When synthesizing proteins on solid support, several major technological bottlenecks must always be taken into consideration: (1) the choice of an adequate solid support that allows the protein to be assembled with good yields. (2) The choice of the linker that will be used for the specific immobilization of a first pure and unprotected peptide segment on a suitable solid support. (3) The synthesis of appropriately functionalized peptide segments. (4) The handling of poorly soluble or aggregation-prone segments. (5) The assembly of segments by successive ligations, which is confronted with the need for continuous development of methods for temporary ligation and masking of reactive groups. (6) Control the conditions necessary for the cleavage of these linkers under very mild conditions, in order to release from the solid support the protein assembled by successive ligations.

During this work, we tried to provide answers to the bottlenecks mentioned above:

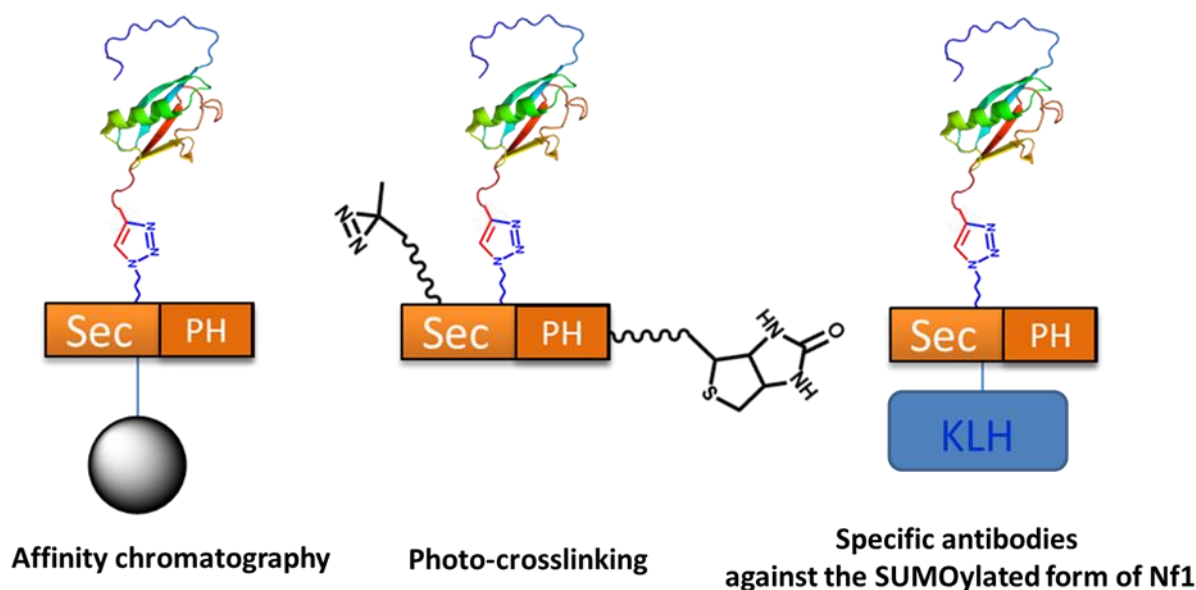
We reported a new SPCL strategy based on an enzymatic release of a polypeptide from a CPG2000 solid support. A huge amount of work was devoted to obtain a suitable SPCL linker and to find the optimal conditions for its mild and selective cleavage. We also had to cautiously choose an adequate solid support compatible with linker cleavage condition given that the nature of the solid support can affect the cleavage rate as demonstrated in this work. As mentioned above, a challenging step of SPCL strategy is the design of masked reactive groups

in order to avoid side reactions. To meet this need, we improved the previous method reported by our group for the synthesis of *N*-2-hydroxybenzyl-cysteine peptide crypto-thioesters to obtain a highly reproducible low racemization solid-phase automatable method for the synthesis of clean crypto-thioesters. In addition, and in order to answer the very frequently encountered problem of low soluble or aggregation-prone segments, we developed a straightforward methodology to overcome solubility challenges for N-terminal cysteinyl peptide segments.

We also reported an original SPCL synthetic route which combines two types of orthogonal chemical ligations, NCL and PTL. In this way, there is no need for masking/unmasking of reactive groups to perform successive ligations, therefore limiting the number of steps and intermediate purifications. This strategy was used to synthesize a 97 residues polypeptide where an isopeptide bond is replaced by a triazole, a non-hydrolyzable bond amide mimic. The resistance of the synthesized molecules to SENP SUMO-proteases will open new research avenues potentially applicable to targets other than Nf1.

The long-term objective of this work, which will be carried out in our group under a thesis project entitled: “**Development of tools for the study of neurofibromin SUMOylation**” is to use this original methodology for the synthesis of mono-, di- or tri-SUMOylated Nf1-derived peptides, where the isopeptide linkages are replaced by non-hydrolysable triazoles. Capitalizing on the recent arrival in our team of Mélanie Chenon as a permanent member (Ingénieur d'études) at CBM, specialized in the recombinant production of proteins, besides SEC-PH-derived peptides, this futur thesis project will also intend to extend our purely chemical approach to the semi-synthesis of a triazolo-SUMOylated version of the full length SEC-PH, through recombinant expression of a protein in which an azidonorleucine residue will be site-specifically introduced through unnatural aminoacid mutagenesis.

As a proof on concept, these triazole containing SUMOylated Nf1-derived peptides will be used to demonstrate their functionality and their resistance to SUMO-proteases. The next step would be to use these peptides to develop probes in order to isolate specific partners of the SUMOylated form of Nf1. This could be done by affinity chromatography or by photo-cross-linking with the aim of identifying proteins that specifically interact with the SUMOylated form of Nf1. The production of specific antibodies against the SUMOylated form of Nf1 could also be considered in order to generate specific detection tools (see figure below).



We expect that the strategies put forwards in this thesis will greatly help this following thesis project. Besides this exciting challenge, this thesis work draws out several varieties of other perspectives:

- We firmly believe that the future extension of our new SPCL strategy to the synthesis of longer targets would open a new pathway to facilitate the preparation of biologically relevant proteins with important pharmacological properties. This calls for the improvement of the AcM protected cryptho-thioester segment synthesis. As demonstrated in this work, this segment is prone to base-induced β -elimination followed by Michael-type addition of piperidine during the subsequent Fmoc cleavage steps affording a side product that cannot be separated from the desired product. The addition of dodecanthiol to induce a competitive nucleophilic attack generating a hydrophobic dodecanesulfanyl derivative easier to separate from the desired product was a convenient approach that allowed us to obtain a pure 40 aa segment in a good yield. However, this method was not applicable to larger segments thus limiting the size of our final target. A deep study of this side reaction and the identification of the parameters favoring its formation would help in developing a method that abolishes the formation of the side product. This might require the use of another method for Fmoc deprotection (a less nucleophilic base such as DBU?) or the use of another protecting group which should meet the following requirements: It absolutely must be completely stable to basic and acidic treatments as well as to ligation conditions; *O*-

Nitrobenzyl (oNB), a cysteine protecting group removed by photolysis ($\lambda = 300\text{-}400\text{ nm}$), may be a valid candidate, for example.

- The deciphering of imidazolidinones structure and the assumptions made on their formation mechanism reported in **chapter 3** could be significant for the synthesis of *N*-alkylated chiral amino acids or peptide through reductive amination, either in solution or on solid phase.
- We strongly believe that our solubilizing strategy will advantageously complement existing methodologies in the synthesis of challenging proteins. We also think that in addition to a controlled release from the solid support, the linkers presented in this work could be used for *à la carte* introduction of solubilizing groups. In case of very difficult segments, we could consider combining the two strategies where the solubilizing tag reported in **chapter 4** would be introduced on the N-terminal cysteine side chain while the linker could be installed at any other residues side chain followed by the introduction of a poly Lys or Arg tag, thus installing two solubilizing tags at two distinct positions and which cleavages are triggered by different conditions.

From the numerous examples cited in this work, it is clear that solid phase peptide synthesis combined to the use of chemical ligation is a promising strategy for the chemical synthesis of challenging proteins. However, in an honest statement, even though conceptually innovative, the SPCL linkers developed in this work are far to address all the problems related to this extremely appealing approach, and optimization of the solid-phase ligations proved to be extremely difficult as compared to solution phase ligation approaches. The future of SPCL is in the synthesis of large targets that, by the traditional in solution synthesis approach, would be extremely laborious. One could debate on the advantages of one method over the other but an opportune middle ground solution would be the use of a mixed solid phase/solution assembly strategy or the use of soluble polymers as supports in order to try to avoid the support deleterious effects in reaction kinetics. All of this is with the objective of enabling the chemical synthesis of tailor-modified proteins as chemical biology tools that at the moment are only fascinating targets.

Développement de méthodologies chimio-enzymatiques pour la synthèse de protéines par ligation chimique sur support solide

La synthèse totale de protéines requiert l'assemblage de segments peptidiques déprotégés par des réactions de ligation, telle la *native chemical ligation* (NCL). Lorsque de nombreuses ligations successives sont mises en œuvre, la nécessité de purifier les intermédiaires réactionnels conduit souvent à de faibles rendements globaux. Une solution pour s'affranchir de ces étapes de purification est d'assembler les protéines par ligation sur support solide (SPCL). Cette approche a cependant été jusqu'à présent essentiellement limitée à des preuves de concept, l'une des raisons principales étant la difficulté d'immobiliser le premier segment peptidique sur un support adapté *via* un bras qui peut être coupé une fois les ligations effectuées. De nombreux bras ont été décrits, mais leur coupure nécessite des conditions incompatibles avec certaines protéines cibles. L'objectif principal de cette thèse était de développer une nouvelle génération de bras en explorant une approche enzymatique plutôt que purement chimique, afin d'induire une coupure douce et sélective. Un premier bras coupé par la β -galactosidase a été étudié, mais les cinétiques de coupures sur support solide se sont avérées très lentes. Nous avons émis l'hypothèse que la grande taille de l'enzyme (540 kDa) ralentissait sa diffusion à l'intérieur du support. Par conséquent, un second bras conçu pour être coupé par des enzymes plus petites a été synthétisé. Dans le cas de la lambda-phosphatase (25 kDa), le rapport entre les vitesses de coupure en solution et sur support a été multiplié par un facteur 300, comparé au premier bras. Ces travaux ont aussi traité d'autres limites de la synthèse de protéines par ligation telle que la manipulation de segments peu solubles. Nous avons ainsi développé un *tag* solubilisant qui peut être introduit de façon automatisée sur une cystéine N-terminale, par l'intermédiaire d'un pont disulfure. Le segment ainsi modifié peut être facilement purifié, et le pont disulfure est ensuite coupé instantanément dans les conditions de NCL. Le développement de techniques plus efficaces de synthèse des segments a également été abordé, dans le cadre d'une étude approfondie d'un protocole décrit précédemment pour la synthèse de peptides crypto-thioesters *N*-Hnb-Cys, capables de former des thioesters *in situ* lors de la NCL. Ce travail a conduit à l'identification d'une variété de co-produits et à l'optimisation d'un protocole de synthèse facile à mettre en œuvre, automatisable et reproductible. L'utilité de toutes ces méthodes a été démontrée par la synthèse d'un polypeptide de 160 acides aminés, par ligation de trois segments sur support solide.

Mots clés : Peptides ; Protéines ; SPSS ; SPCL ; Bras enzymo-labile ; Ligation chimique

Development of chemo-enzymatic methodologies for the total synthesis of proteins through solid-supported chemical ligation

The total synthesis of proteins requires the assembly of unprotected peptide segments through ligation reactions, such as native chemical ligation (NCL). The need of intermediate purifications is a strong limitation, as this generally leads to low overall yield when numerous successive ligations are implemented. An early-recognized solution to this problem is the assembly of the protein on a solid support. Although extremely appealing, solid phase chemical ligation (SPCL) has been so far essentially limited to proof-of-concepts, one of the main reasons being the difficulty to immobilize a first peptide segment on a water-compatible solid support through a dedicated SPCL linker which could be cleaved after assembly completion. A wide range of linkers have been described, but their cleavage requires chemical conditions that proved to be incompatible with some sensitive targets. The main aim of this work was the development of a next generation of linkers by exploring an enzymatic rather than a purely chemical approach, in order to induce a cleavage under extremely mild and selective conditions. A first β -galactosidase cleavable linker was explored however, cleavage kinetics on solid support were slow. We assumed that the large size of the enzyme (540 kDa) hampered its diffusion inside the solid support. Hence, a second linker designed to be cleaved by smaller enzymes was synthesized. In the case of lambda-phosphatase (25 kDa), the relative cleavage rate, solid vs solution phase, was increased by a factor of 300. Besides linkers, this work focused on addressing other bottlenecks associated to NCL-based chemical protein synthesis such as the handling of poorly soluble segments. Therefore, we developed a method that makes use of a key solid-supported disulfide formation to introduce a 2-amino-2,2-dimethylsulfanyl (Ades) moiety that can be further used as an NCL labile handle to introduce a solubilizing tag on N-terminal cysteinyl peptides through standard Fmoc-SPSS. We also worked on the development of more efficient synthesis methods by the in-depth reinvestigation of a previously described protocol for *N*-Hnb-Cys peptide crypto-thioesters synthesis, which are capable of forming thioesters *in situ* under NCL conditions. This optimization work led to the identification of a variety of deleterious by-products and to the optimization of an automatable and highly reproducible synthetic protocol leading to clean peptide crypto-thioester segments. The utility of all these methods was demonstrated by the solid supported synthesis of a 160 residues polypeptide assembled from three segments.

Keywords: Peptides ; Proteins ; SPSS ; SPCL ; Enzyme-cleavable linkers ; Chemical ligation