Université de Lille

École Doctorale Biologie – Santé de Lille (EDBSL)

THESIS

To pursue the title of

PhD at the University of Lille

Specialty: chemistry and physical chemistry - MED

Presented by

Olga FIRSTOVA

Iterative peptide assembly in one-pot using *N*-selenoethyl cysteine chemistry with potential application for antifreeze protein synthesis

Defended the 4th of July, 2023

Jury members:

Reporters:	Pr. Vladimir TORBEEV *(University of Strasbourg)*
	Pr. Philippe KAROYAN *(Sorbonne University, Paris)*
Examiner:	Dr. Sabine Chierici *(Grenoble Alpes University)*
Jury president:	Pr. Philippe COTELLE *(University of Lille)*
Thesis director:	Dr. Oleg MELNYK *(Institut de Biologie de Lille)*
Mentor:	Dr. Vincent DIEMER *(Institut de Biologie de Lille)*

Institut de Biologie de Lille, CNRS UMR 9017

Institut Pasteur de Lille

1 Rue du Professeur Calmette 59019 Lille Cedex France

Acknowledgments

Content

Abbreviations

Ac	Acetyl
ACN	Acetonitrile
Alk	Alkyl
Ar	Aryl
Bn	Benzyl
Boc	*tert*-butyloxycarbonyl
Bz	Benzoyl
CMC	Critical micellar concentration
DCE	Dichloroethane
DCM	Dichloromethane
DIC	*N,N*'-Diisopropylcarbodiimide
DIEA	*N,N*-diisopropylethylamine
DMF	*N,N*-dimethylformamide
DMSO	Dimethylsulfoxyde
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
Et	Ethyl
Et_2O	Diethyl ether
Fmoc	9-fluorenylmethoxycarbonyl
Gdn·HCl	Guanidinium chloride
HATU	*N*-[(dimethylamino)-*1H*-1,2,3-triazolo[4,5-*b*]pyridino-1-ylmethylene]-*N*-methylmethanaminium hexafluorophosphate *N*-oxide
HBTU	*N*-[(1*H*-benzotriazol-1-yl)(dimethylamino)methylene]-*N*-methylmethanaminium hexafluorophosphate *N*-oxide
HCTU	2-(6-Chloro-1-H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate
HOBt	1-Hydroxybenzotriazole
HPLC	High-performance liquid chromatography
KAHA	Ketoacid-hydroxylamine ligation
KCL	Kinetically controlled ligation
LC-MS	Liquid chromatography–mass spectrometry
LG	Leaving Group
MALDI-TOF	Matrix-assisted laser desorption ionisation-time of flight
Me	Methyl
MESNa	sodium 2-mercaptoethanesulfonate
MPA	3-mercaptopropionic acid
MPAA	4-mercaptophenylacetic acid
MPSNa	Sodium 3-mercaptopropiosulfonate
MS	Mass spectrometry
MTG	Methyl thioglycolate
NCL	Native chemical ligation
NMP	*N*-methyl-2-pyrrolidinone

NMR	Nuclear magnetic resonance spectroscopy
PG	Protecting group
Ph	Phenyl group
PyBOP	Benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate
SEA	*N*,*N*-*Bis*(2-sulfanylethyl)amide
SeEA	*N*,*N*-*Bis*(2-selenylethyl)amide
SEAlide	*N*-sulfanylethylanilide
Sec	Selenocysteine
SetCys	*N*-selenoethyl cysteine
SPCL	Solid phase chemical ligation
SPPS	Solid phase peptide synthesis
STL	Serine/threonine ligation
***t*Bu**	*tert*-butyl
TCEP·HCl	*tris*(2-carboxyethyl)phosphine hydrochloride
TFA	Trifluoroacetic acid
THF	Tetrahydrofurane
Thz	Cysteine thiazolidine
TIS	Triisopropylsilane
TFET	Trifluoroethanethiol
TPPTS	Triphenylphosphine-3,3',3''-trisulfonic acid trisodium salt
TTFMPP	Bis(3-sulfonatophenyl)(3,5-di-trifluoroMethylphenyl)phosphine, disodium salt
TMPP	Tris(4,6-dimethyl-3-sulfanatophenyl)phosphine trisodium salt hydrate
Trt	Trityl
TS	Transition state
UV	Ultra-violet

Abstract

Proteins are macromolecules naturally produced by all living organisms and constitute essential components of organic life. Their functions and structure are the result of spatial organization of a polypeptide chain made of the concatenation of 21 different amino acid (AA) residues. To access polypeptides by chemical synthesis, Merrifield in 1963 introduced the solid phase peptide synthesis (SPPS), a chemical method that is appropriate for the preparation of peptides and small proteins usually with a size less than 50 AAs. This approach guarantees high product purity and allows to easily modify a protein sequence or to functionalize it on demand. To facilitate the access to longer peptide sequences with a size greater than 50 AAs, a complementary strategy was developed based on the concatenation of smaller peptide segments. To make this second approach work, chemists have a few chemical ligation reactions to covalently and chemoselectively connect unprotected peptide segments together.

This thesis manuscript consists of two separated parts. The first part is dedicated to the development of an innovative method of polypeptide assembly in one-pot based on native chemical ligation (NCL) reaction and *N*-selenoethyl cysteine (SetCys), the artificial amino acid where amine and thiol groups are bonded *via* selenoethyl bridge forming a cyclic structure. The key feature of SetCys is its ability to lose its selenoethyl appendage *via* C-N bond breaking under mild conditions. I will describe the application of NCL in non-standard conditions, the chemical properties of SetCys and the challenge to combine these chemistries for designing an elegant and chemoselective two-step method enabling peptide segment concatenation.

The second part describes the definition and properties of ice binding proteins (IBPs) and their sub-group called antifreeze proteins (AFPs), which are produced by cold coping organisms in nature. I will focus on a particular protein produced by *Tenebrio molitor* (*Tm*) beetle, aiming to access its bioinspired analogue by chemical protein synthesis for the first time. In perspective, the innovative method described in the first part of the manuscript would be applied for the synthesis of *Tm*AFP.

Résumé

Les protéines sont les macromolécules naturellement produites par tous les organismes vivants et elles constituent des composants essentiels du monde vivant. Leur fonctions et structure sont le résultat de l'organisation dans l'espace d'une chaîne polypeptidique construite par l'assemblage de 21 acides aminés (AA) différents. Afin d'accéder aux polypeptides par voie chimique, en 1963, Merrifield développa le synthèse peptidique sur support solide (solid phase peptide synthesis, SPPS), une méthode chimique adaptée à l'obtention de peptides et de protéines de petite taille, plutot de moins de 50 AAs. Cette approche assure la haute pureté des produits obtenus et permet de modifier facilement les séquences peptidiques ou de les fonctionnaliser. Afin de faciliter l'accès aux peptides composés d'un nombre d'acide aminés supérieur à 50, une stratégie complémentaire a été développée qui est basée sur assemblage de segments peptidiques non protégés plus courts. Pour rendre cette deuxième approche efficace, les chimistes ont développé quelques réactions de ligation chimique qui permettent de connecter deux peptides non-protégés de manière covalente et chimioselective.

Ce manuscrit de thèse se compose de deux parties distinctes. La première partie est dédiée au développement d'une méthode innovante d'assemblage de polypeptides basée sur la réaction de ligation chimique native (NCL) et la *N*-sélénoéthyl cystéine (SetCys), un acide aminé artificiel dérivé de la cystéine où les groupes amine et thiol sont liés *via* un pont sélénoéthyle pour former une structure cyclique. Une propriété importante de SetCys est sa capacité à perdre spontanément son bras sélénoéthyle *via* la rupture de la liaison C-N dans des conditions douces. Je décrirai l'application de la réaction de NCL dans des conditions non standard, les propriétés chimiques de SetCys et le défi de combiner ces chimies pour développer une méthode élégante et chimiosélective en deux étapes permettant la concaténation de segments peptidiques.

La deuxième partie définit et décrit les propriétés des protéines capables de s'attacher à la surface de la glace (ice-binding proteins, IBPs) et d'une catégorie particulière appelée protéines antigel (antifreeze proteins, AFPs), qui sont produites par des organismes adaptés au froid dans la nature. Je me concentrerai sur une protéine particulière produite par le coléoptère *Tenebrio molitor (Tm)*, et je décrirai les première tentatives visant à accéder à un analogue par synthèse chimique. En perspective, la méthode innovante décrite dans la première partie du manuscrit pourrait être appliquée àla synthèse de *Tm*AFP.

Tools to access protein molecules – state of art

1 Protein synthesis

1.1 Introduction

Proteins are macromolecules naturally produced by all living organisms and they are key building blocks of huge part of organic life. Proteins possess various special conformations what gives them specific properties and functions. Their structure is a result of organization, or folding, of linear polypeptides of different length. Polypeptides consist of individual amino acids, minimal units for protein construction, that are attached one to another *via* amide bonds.

Protein structure includes four levels (**Figure 1**). The sequence of attached amino acids determines the so-called primary protein structure, linear peptide (**Figure 1, a**). A peptide chain organised for example in α-helix, β-strand or β-sheet due to hydrogen bond formation represents the secondary protein structure (**Figure 1, b**). The next level is called tertiary structure, a single protein molecule, a monomer where a polypeptide is folded in a 3-dimentional globular structure driven by hydrophobic interactions, salt bridges, hydrogen bonds or disulfide bonds. The quaternary structure is a more complex special organisation of two or more tertiary structured monomers. This multimer is stabilised by monomers intercalations, salt or disulfide bridges and hydrogen bonding.

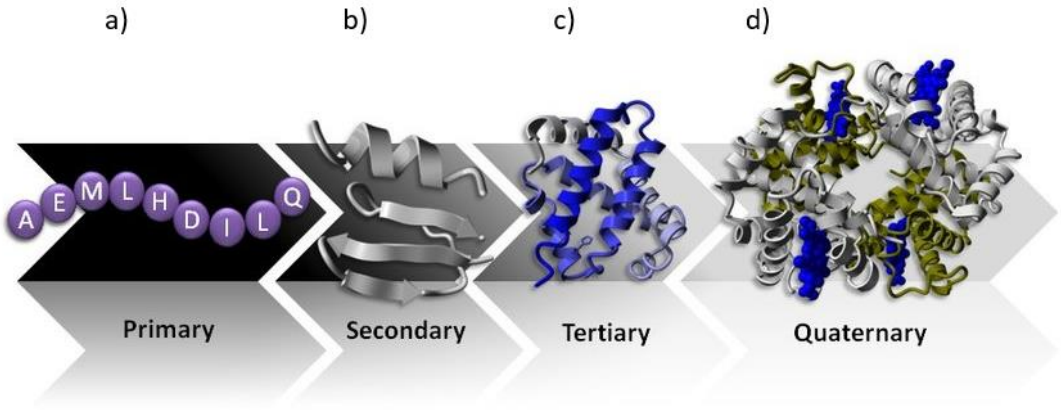


Figure 1. *Four levels of protein organization: primary, secondary, tertiary and quaternary (adapted from Joosten R. et al*[1]*).*

Protein structures have tremendous number of different size, organisation, symmetry and properties. Those natural macromolecules had billions years of evolution and differentiation. Simple proteins building blocks are present not only on our Earth. For example, the traces of amides were found on comet 67P/Churyumov-Gerasimenko by Rosetta's Philae lander (detected by mass spectrometry)[2]. Glycine was detected on several meteorites[3], as well as glycine precursors[4].

Proteins are associated with specific functions in nature. Tissues of all living organisms are structured by proteins alongside with other biomolecules. Cellular membranes, enzymes, hormones, antibodies involve protein molecules. Venoms and pheromones are often

small proteins or peptides[5]. Such vast diversity of natural proteins inspired discoveries of new application fields for proteins and the design of new bioinspired structures.

The application of peptides and proteins grows in various fields between academy and industry. For example, in the medical field, protein scaffolds endowed with biocidal properties are in the curse of development. Proteins are recruited for materials functionalization, in the synthesis of dendrimers, of polymers and the preparation of various hybrid materials. Protein applications were appreciated and described in a dedicated review[6].

To produce proteins, biological and chemical pathways were developed. In biological methods, also known as recombinant methods, target proteins are produced by bacteria or cells. The genome of bacteria or cells is modified by genetic recombination so that target proteins can be expressed by the molecular machinery of the host. Although recombinant approaches offer the possibility of producing large proteins, they are not very flexible in terms of AA sequence modifications when non-coded AAs have to be considered. Furthermore, residual biological contaminants can be present, even after final purification. An alternative to biological methods is to use chemical synthesis. In this case, the amide bonds that connect the AAs of the target proteins are formed *via* chemical reactions. This approach guarantees high product purity and allows to easily equip a protein with the modifier of interest, to insert post-translational modifications or to modify the primary sequence of proteins by introducing different amino acids, including D-amino acids[7].

Successful for the preparation of peptides and small proteins (< 50 AAs), the connection of single AAs by chemical reactions turns out to be less and less efficient as the length of the peptide sequence gradually increases. Although the elongation process is based on highly reliable AA coupling, the increasing number of iterations induces inevitably results in an increase of the amount of by-products formed. To circumvent this problem and to facilitate the access to longer AA sequences, a complementary strategy was developed where the full length of the target peptide sequences is produced by concatenation of smaller peptide segments (< 50 AAs each). To make this approach work, chemists have designed a few chemical ligation reactions to covalently connect chemoselectively two unprotected peptide segments together. These reactions can be chained one after another leading to larger and more complex peptide structures, as illustrated with the synthesis of functional L-enzymes[8] and D-enzymes[9] or 450 AA polyubiquitin[10].

In this chapter, I will describe chemical ligation reactions that are classically used in modern chemical protein synthesis to assemble peptide segments and I will emphasize the native chemical ligation (NCL), the most widely used reaction in the field up to now. I will focus on the scope and the mechanism of this reaction before illustrating the NCL-based strategies that are usually exploited to concatenate several peptide segments. In this context, I will highlight how the innovative chemistry developed by the CBF team and based on redox-controlled chemical devices fits with the major challenges of the field. At the end, this state of

the art will be used as starting point for my PhD works aiming at developing novel one-pot, redox-controlled, iterative assemblies of peptide segments.

1.2 Chemical ligation: a milestone for the modern chemical protein synthesis

Chemical ligation reactions are the methods of choice in modern chemical protein synthesis to access quickly large sequences of AAs. The term chemical ligation has been introduced by Schnölzer and Kent in 1992 and refers to a chemo- and regioselective reaction between two peptide segments that allows the formation of longer polypeptides[11]. Chemical ligation reactions are usually performed in water and under mild conditions starting from unprotected peptide segments.

Ligations leading to the formation of hydrazone, thioester, oxime and triazole junctions have been developed to connect two peptide segments (**Figure 2**)[11–15]. Relying on reactive groups located at the N- and C- terminal ends of the peptide segments (X and Y in **Figure 2**), these reactions occur selectively in spite of the diversity of functional groups (carboxylic acid, amine, thiol…) present around the reactive site. However, hydrazone, thioester, oxime or triazole ligations have the disadvantage of leading to non-native junctions, restricting their scope to the preparation of conjugates.

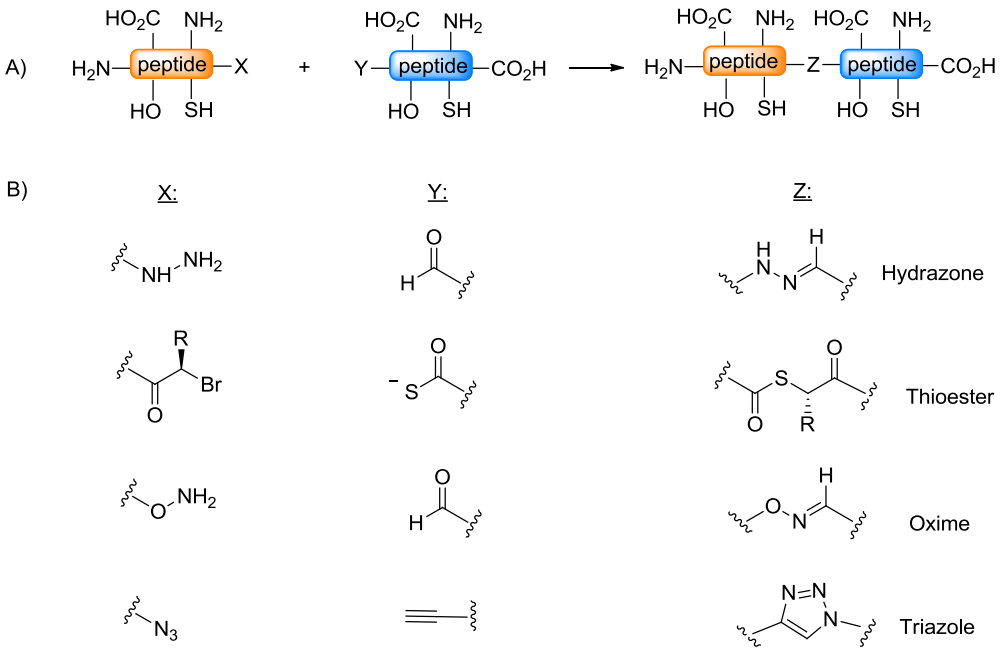


***Figure 2**. Examples of non-native ligation (adapted from Agouridas et al[16]).*

Assembling peptide segments *via* native amide bonds is of great interest and has stimulated the development of a series of conceptually related ligations (**Figure 3,** adapted from the review[16]). In all these reactions, the coupling partners are connected through a highly chemo

selective capture step and the subsequent rearrangement of the covalently-linked intermediate leads to the formation of the final amide bond.

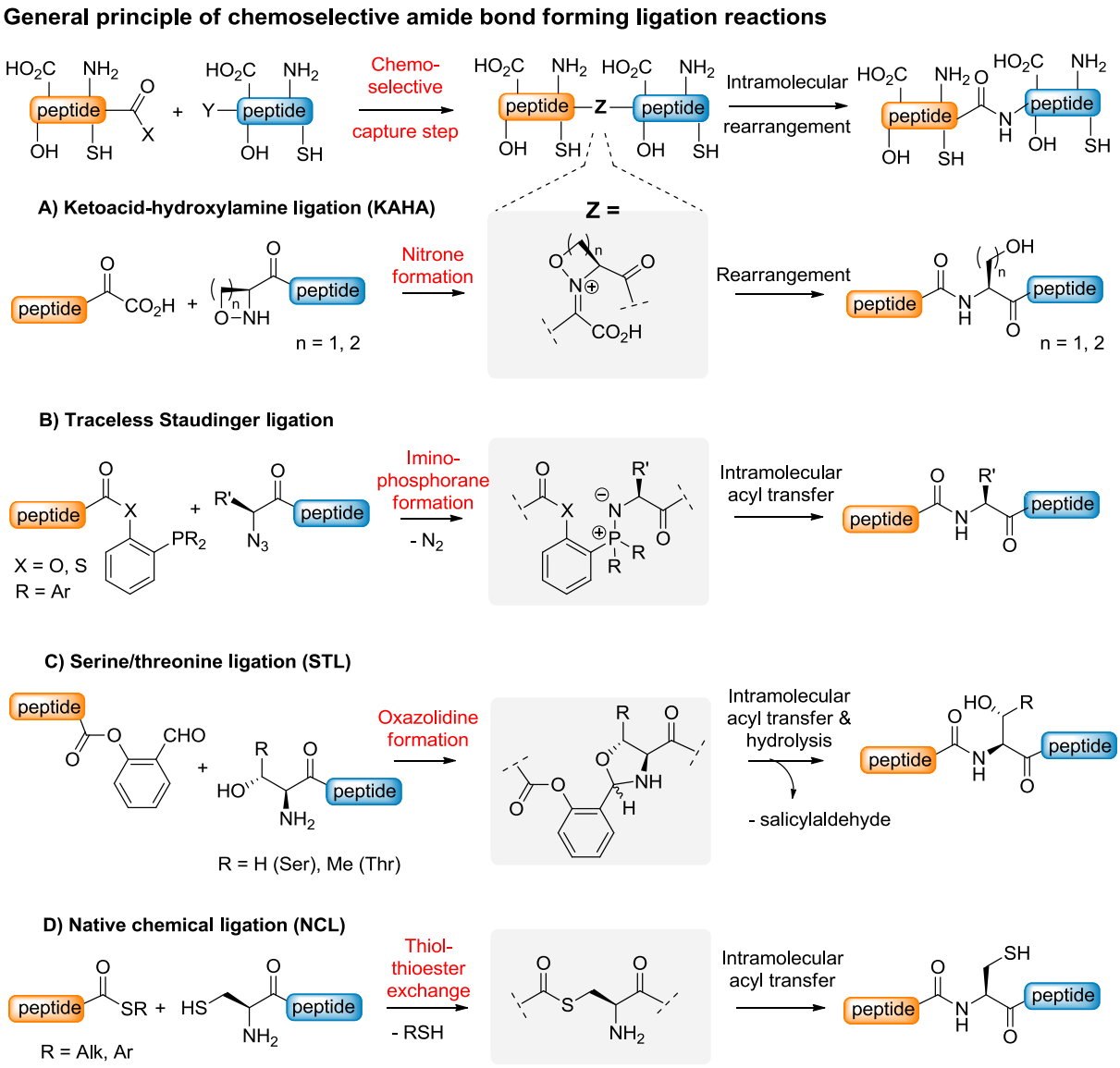


***Figure 3**. Types of ligation reactions with the formation of native amide bond: a) ketoacid-hydroxylamine (KAHA) ligation[17], b) Traceless Staudinger ligation[18,19], c) Serine/threonine ligation (STL)[20], d) Native chemical ligation (NCL)[21].*

The main ligation leading to native peptides include the α-ketoacid-hydroxylamine (KAHA) ligation, the Staudinger traceless ligation, the serine/threonine ligation (STL) and the native chemical ligation (NCL).

α-ketoacid-hydroxylamine (KAHA) ligation

Reported by Bode in 2006, this reaction promotes the formation of a native peptide bond to serin, threonine or homoserine starting from two peptides functionalized respectively with a hydroxylamine at the N-terminus and an α-keto acid at the C-terminus (**Figure 3, a**) [17].

The benefit of this reaction that proceeds through the formation of an intermediate nitrone, is that no additional reactants is required and that the only by-products are water and carbon dioxide. In the initial version, KAHA was carried out in an organic solvent such as DMF, DMSO or MeOH and involved unsubstituted hydroxylamines. Later, the conditions were adapted for aqueous media by using *O*-substituted hydroxylamine derivatives, what allowed also the acceleration of the ligation process[22].

Staudinger traceless ligation

Under its classical form, the Staudinger reaction is a covalent capture of an azide derivative by a phosphine in aqueous media with formation of intermediate iminophosphorane that hydrolyses resulting into an amine[23]. This reaction was successfully applied for chemoselective peptide ligation independently by two scientific groups in 2000, headed by C. Bertozzi[18] and R. Raines[24]. In the adapted ligation method, a peptide bearing a C-ter phosphinothioester function reacts with a peptide equipped an N-terminal azide. An aza-ylide intermediate forms and, by intramolecular acyl transfer from sulfur to nitrogen, transforms to an amidophosphonium salt. The latter spontaneously hydrolyzes to form a native peptide bond (**Figure 3, b**).

Serine/threonine ligation (STL)

In 1990s Tam et al. developed a pseudoproline ligation method[19,25,26] which was a basis for the conception of serine/threonine ligation (STL) introduced in 2013[20]. The reaction between the aldehyde and β-amino alcohol functionalities occurs *via* intermediate oxazolidine formation followed by intramolecular acyl transfer. Hydrolysis of the rearranged product is performed in a separate step and results in the formation of a native peptide bond (**Figure 3, c**).

Although those reactions found synthetic utility, the most widely used ligation reaction to form a peptide bond at the junction of two peptide segments is the **Native Chemical Ligation (NCL)**[27]. This reaction based on thiol/thioester exchange chemistry is the subject of the next section (**Figure 3, d**).

2 The Native Chemical Ligation (NCL)

The native chemical ligation (NCL) was reported for the first time by Kent in 1994 and this discovery is reminiscent from the works of Wieland in the 1950s[28,29]. Inspired by the role of acyl donor played by *S*-acetyl coenzyme A in natural cellular processes, Wieland showed that dipeptides could be produced in aqueous media under mild conditions by reacting valine thiophenyl ester with AAs such as glycine, alanine or cysteine. The process proved to be particularly efficient in the case of the cysteine residue, suggesting the key role of the cysteine thiol group on the outcome of the reaction. As depicted in **Figure 4**, Wieland hypothesised that

the coupling proceeded through the formation of a thioester intermediate produced from the reactants by thiol/thioester exchange. A subsequent acyl migration from sulfur to nitrogen was proposed to explain the formation of the final dipeptide.

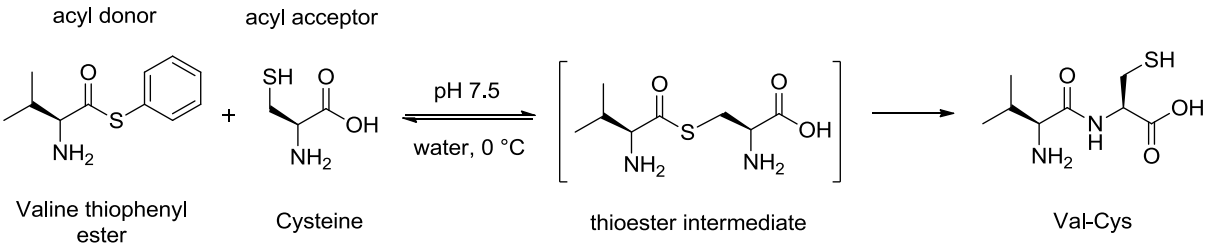


***Figure 4**. The principle of thiol-thioester exchange followed by native amide bond formation, described by Wieland[28].*

Later, this principle was applied by Kent to the concatenation of unprotected peptide segments and has led to the emergence of the NCL. Under its classical form, the NCL involves a peptide armed with a N-terminal cysteine and another one possessing C-terminal thioester moiety (**Figure 5**). These compounds react chemo selectively by thiol/thioester exchange to give an intermediate thioester. The latter rearranges *via S,N*-acyl shift leading to the formation of the native amide bond and to the elongation of the peptide sequence.

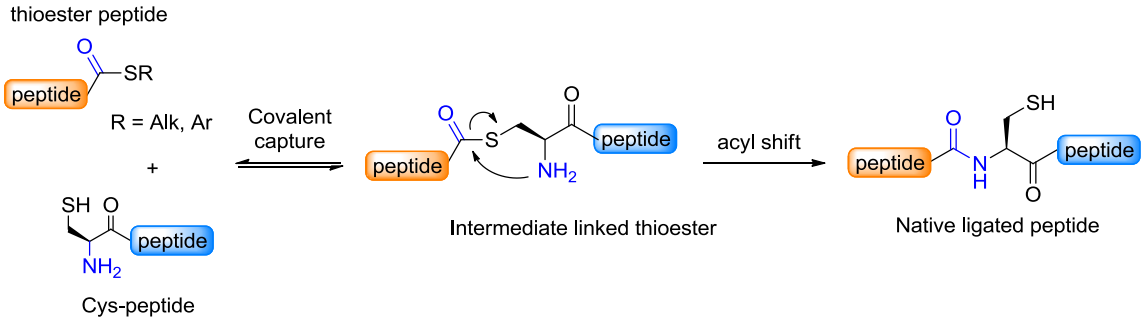


***Figure 5.** Mechanism of the native chemical reaction.*

Since its discovery, NCL has been the subject of many studies and developments and has found wide synthetic applications in chemical biology, medicinal chemistry, and material science[16]. Lots of proteins were synthesised by NCL and methods derived thereof[16], including D-proteins, cyclic or branched architectures and peptide sequences with post-translational modifications (ubiquitylation, glycosylation, phosphorylation, acetylation…)[27,30]. NCL was also applied to produce conjugates, enabling the labelling of biomolecules or the synthesis of complex lipids and oligonucleotides. Finally, NCL was extensively used to synthesize peptide materials such as high molecular weight peptide-based biopolymers, hydrogels or dendrimers and to design peptide-functionalized surfaces. Thus, it is important to have a deep look on NCL mechanism and reactivity of the reacting partners.

2.1 Reactivity of the NCL coupling partners

Experimentally, the NCL reaction between of cysteinyl peptide and a peptide thioester is often carried out at 25-37 °C and around neutral pH in a buffered solution. Denaturants such as guanidinium chloride are usually added to limit aggregation phenomena and to improve the solubility of the NCL coupling partners.

2.1.1 Cysteine reactivity

As previously described, NCL mechanism involves the thiol group of an *N*-terminal Cys peptide. The cysteine AA is essential for living cells for its properties such as metals chelation and redox activity. In contrast, NCL mechanism takes advantage of the nucleophilic and acido basic properties of the cysteine AA.[31] The pK_a of the thiol group for N-terminal cysteine is 7.0 and 8.0[32]. Under the NCL conditions, Cys thiol is partially deprotonated and exists under thiolate form which is the active nucleophilic form acting during the NCL capture step.

To increase the rate of the NCL, Cys can be replaced by selenocysteine (Sec)[33–35]. Nucleophilic properties of the selenium element towards carbonyl group is higher than the one of sulfur element[36]. Due to lower pK_a of selenols comparing with thiols (~ 3 pK_a units less), it is easier to transform selenol into selenoate at a neutral or mildly acidic working pH. This makes acyl donor capture step more efficient when Sec-peptide is used instead of thiol-based during NCL.

As NCL reaction considers thiol-thioester exchange, the conclusion might be that this reaction is restricted to the formation of XX-Cys junctions. This is indeed the limitation of the NCL as cysteine is not an abundant amino acids in natural proteins[37,38]. To increase NCL scope, methods that combine NCL with a post-ligation desulfurization step were developed[39]. The conversion of Cys into Ala by desulfurization allows the formation of XX-Ala junction in a target peptide (**Figure 6,** R = H). Following the same strategy, the use of beta-mercapto-amino acids enables to access potentially all types of junctions (**Figure 6,** R ≠ H). However, as for other ligation methods, there is the risk of thioester epimerization at long reaction time[40].

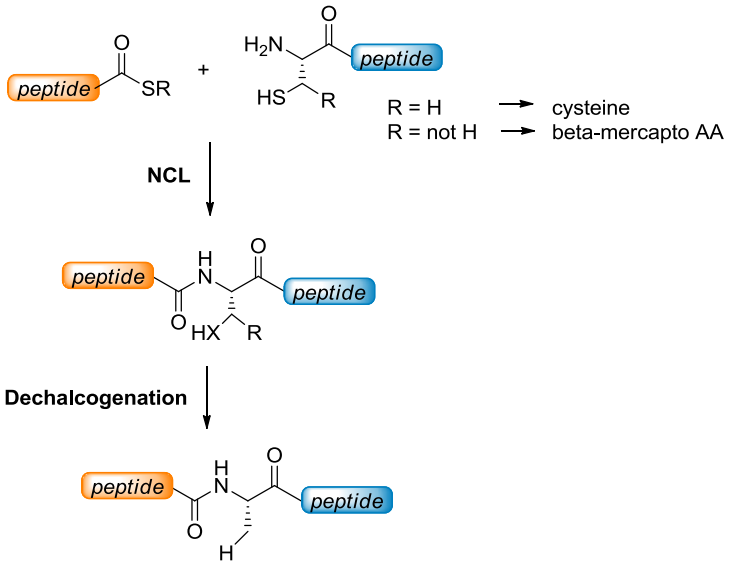


***Figure 6**. Formation of XX-Ala junction when alanine presents in a target peptide with use of a beta-mercapto-amino acid.*

To perform the NCL, not only cysteine reactivity matters, but also of a thioester, the second reacting partner during the ligation.

2.1.2 Thioester reactivity

Thioesters are the core reactants for many bioinspired reactions such as various types of bioconjugation, and it is important to have a deeper look at their structure and reactivity[41–43]. To discuss the role of thioesters in NCL, it is interesting to compare them with oxoesters and amides. Thioesters, oxoesters and amides are structurally related groups, where carbonyl group is linked with a sulfur, oxygen or nitrogen heteroatom.

Electrons delocalization towards carbonyl group is different for every heteroatom, what explains different reactivity of those esters. To explain the variation in reactivity, C(O)-X bonds lengths have been measured (**Figure 7**)[44,45]. C(O)-NH and C(O)-O are very similar whereas C(O)-S bond is significantly longer. These data are related to the strength of resonance between the heteroatom and the carbonyl group: the resonance structure II indicates the structure stabilization through resonance orbital interactions of the X n_s and n_p lone pairs with the antibonding σ* and π* molecular orbitals of the carbonyl group, and this interaction is significantly weaker for thioesters. It means that thiols are better leaving groups during the NCL activation step and have higher reactivity towards nucleophiles.

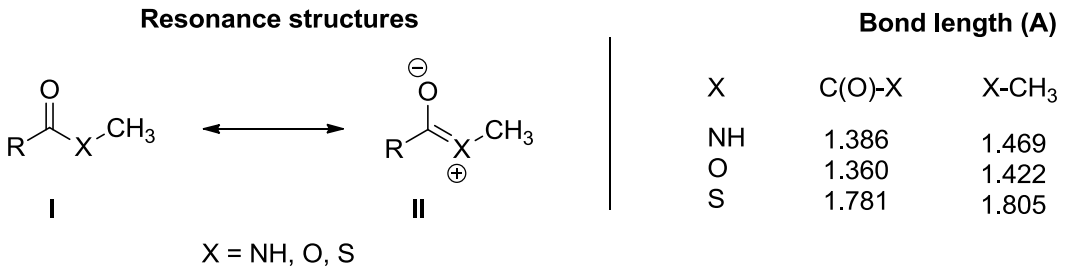


***Figure 7**. Resonance structures for amides (X = NH), oxoesters (X = O), and thioesters (X = S); and bonds characteristics.*[44,45]

In particular, the higher reactivity of the thioester towards thiol nucleophiles enables to perform thiol/thioester exchange reaction under mild conditions and to produce native peptides by NCL. However, this reactivity is highly modulated by the nature of leaving thiol group. The more stable the thiolate released by the nucleophilic attack of the thioester is, the more reactive the thioester is. For example, aryl thioesters are much more reactive than alkyl thioesters in NCL reaction.

The nature of C-terminal AA of the peptide thioester also changes the NCL rate by several orders of magnitude. The main contribution comes from the steric hindrance of the AA side chain. For example, β-branched amino acid residues threonine, valine and isoleucine are considered among the worst-case scenarios for NCL kinetics. In contrast, peptides equipped with C-ter alanine or glycine are easily coupled by NCL due to the small size of these AAs. Note that the very slow NCL kinetics observed for peptidyl prolyl thioesters is attributed to an intramolecular interaction that disfavours the nucleophilic attack on the thioester carbonyl[46]. Note also that peptide prolyl thioesters are prone to peptide deletion by-product formation, which occurs through a diketopiperazine intermediate[47].

It is important to mention that thioesters are susceptible to hydrolysis in aqueous media, and the hydrolysis rate varies depending on thioester nature, the media and the time thioester is incubated in the media[48,49]. Applying that to NCL, while thiol-thioester exchange is a reversible process, thioester hydrolysis is irreversible and may become a problematic competitive reaction if its rate is higher that the ligation rate (**Figure 8**).

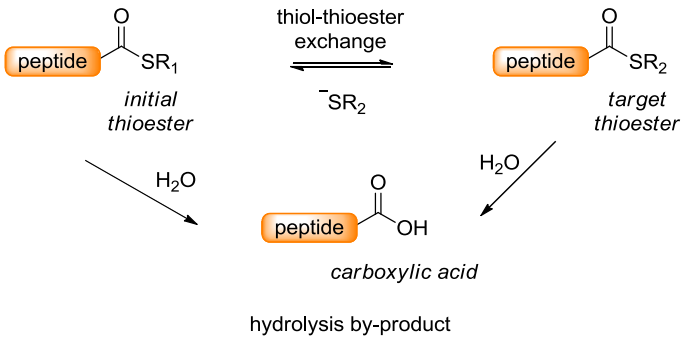


***Figure 8**. Irreversible hydrolysis of thioesters may occur in aqueous media as a competitive reaction to thiol-thioester exchange.*

2.2 Synthesis of the NCL coupling partners

Both cysteinyl and thioester peptides can be produced chemically on solid support with some differences in procedures.

2.2.1 Solid support peptide synthesis (SPPS)

Peptides synthesis by sequential coupling of amino acids on solid support (SPPS) has been reported for the first time by Merrifield in 1963[50]. Sequential assembly of amino acids on solid support is based on cycles of coupling and deprotection steps leading to the elongation of the peptide sequence in the C-to-N direction (**steps 1-3, Figure 9**). Practically, each AA is coupled to the N-terminus of the growing peptide chain by activation of its carboxylic acid function. Once the amide bond has been formed, the protecting group of the alpha amine of the coupled AA is removed. At the end of the elongation process, the assembled peptide is cleaved from the solid support and the protecting group that masked the AA side chains are removed simultaneously, providing the desired peptide segment (**step 5, Figure 9**). Note that the alpha amines of the growing peptide chains that have not reacted with the activated amino acid are systematically acetylated during a capping step (**step 4, Figure 9**), avoiding the formation of deletion product that are difficult to separate during the final purification step of the peptide by HPLC.

Originally, Boc-protected AAs were used as building-blocks for the synthesis of peptide segments by SPPS. In this first strategy, the Boc group is removed at each deprotection step by treatment with TFA and the cleavage of the assembled peptide from the solid support demand highly corrosive and toxic anhydrous HF media. Glassware is not resistant to hydrofluoric acid, so reactions with HF must be held in non-transparent flasks made of polytetrafluoroethylene or some metal alloys. To avoid the use of anhydrous HF, a second strategy was later developed based on Fmoc chemistry. Since the Fmoc groups that protect the alpha amine of the AAs are removed in basic conditions, the cleavage from the solid support and the full deprotection of the peptide is performed in TFA, facilitating the synthesis of peptide segments by SPPS. SPPS concept is particularly adapted to automation[51]. Moreover, various solid support, linkers, side chain protective groups and non-canonical AAs have been developed to enrich the peptides prepared by SPPS.

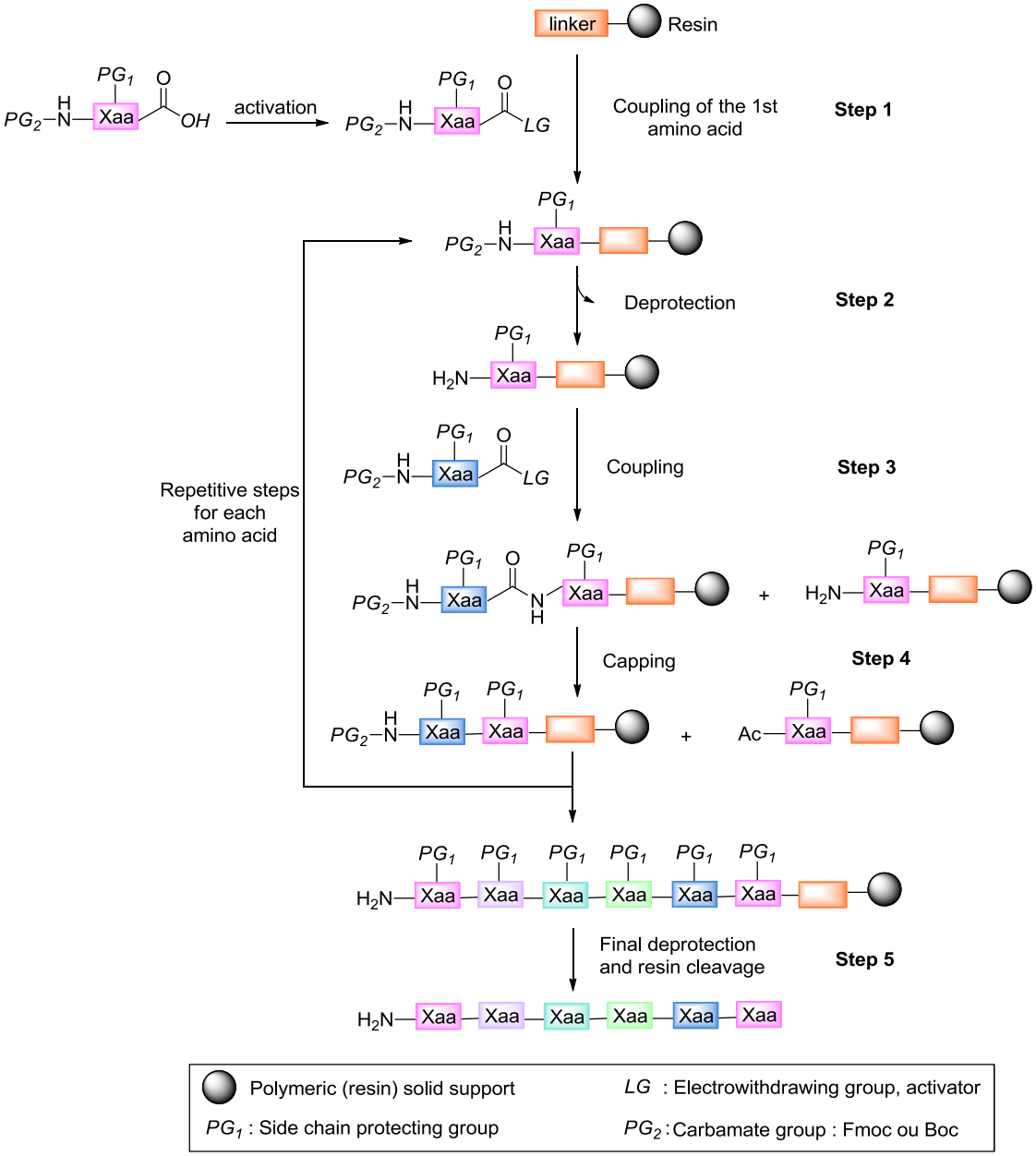


***Figure 9**. Principle of the solid phase peptide synthesis(SPPS).*

SPPS is the method of choice for synthesizing cysteinyl peptides used as coupling partner in NCL reactions since the cysteine residue is introduced, as the others AAs, during the elongation process.

2.2.2 Synthesis of cysteinyl peptides

For example, N-terminal cysteinyl peptides can be obtained starting from commercially available Rink amine solid support, as illustrated in **Figure 10**. In contrast, the introduction of the thioester function at the C-terminal extremity of the coupling partner is more

complex and needs the design of specific strategies. The next section describes the main strategies developed in the field up to now.

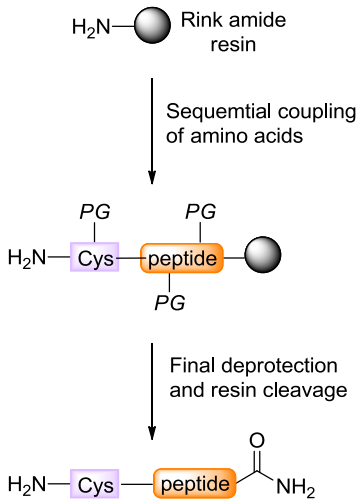


***Figure 10**. Synthesis of cysteinyl peptides by SPPS from a Rink amine solid support.*

2.2.3 Synthesis of peptide thioesters

As the use of peptidyl thioesters has become popular in ligation chemistry, the design of specific resin linkers enabling the production of thioester functionality at the *C*-terminal extremity of peptides was necessary. Two strategies, with their advantages and limitations, were developed depending on whether the thioester group is introduced before or after the elongation of the peptide sequence by SPPS.

The first strategy is to apply SPPS protocol using linkers that contain in their structure the thioester function. In this case, the elongation process and the cleavage from the solid support provide directly the peptide thioester, as illustrated in **Figure 11** with the 3-mercaptopropioamide linker (MPAL)[52]. The main drawback of this strategy is that thioester groups are known to be unstable under the basic conditions, preventing the use of Fmoc-SPPS. This is why Boc-SPPS strategy has to be used, which requires HF media as discussed previously. Note that Boc-SPPS production of peptide thioesters has been adapted to avoid the use of anhydrous HF[52].

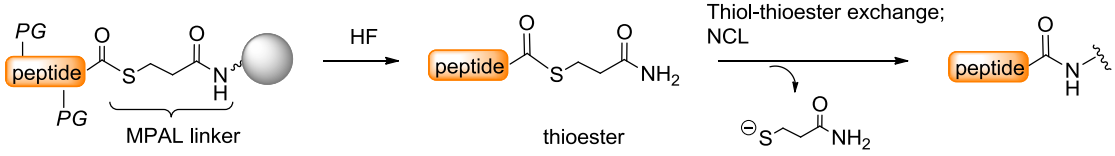


***Figure 11**. Peptide with MPAL that remains as a thioester after resin cleavage; such a peptide thioester can be directly used for NCL.*

A second strategy to obtain a peptide thioester is the design of linkers enabling the formation of the thioester functionality after elongation of the peptide sequence on solid support. As depicted **Figure 12**, the first option is to start from peptide acid and to use carboxylic acid activating procedures relying on DIC/HOBt or PyBOP/DIEA reagents. Thiolysis of the activated thioester with thiols such as ethyl-3-mercaptopropionate[53] or *p*-acetamidothiophenol[54] provide the expected peptide thioester which is then fully deprotected under acidic conditions. The main drawback of this method is the low solubility of a protected peptide in the reaction mixture and the risk of epimerization.

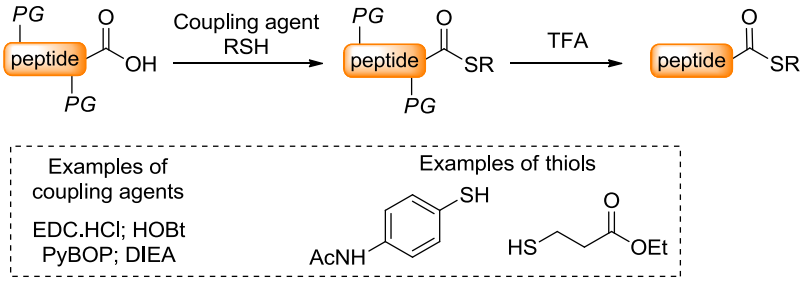


***Figure 12**. Introduction of thioester functionality in solution via thiolysis of carboxylic acid in the presence of an activating agent, base and a thiol followed by peptide deprotection in TFA[53,54].*

Note that methods were designed to activate directly on solid support the carboxylic acid functionality of the peptide assembled by SPPS. For example, the BAL linker (backbone amide linker) enables to anchor the backbone of the elongated peptide to the solid support, the C-terminal acid carboxylic functionality remaining free to be coupled with a thioester AA (**Figure 13**)[55].

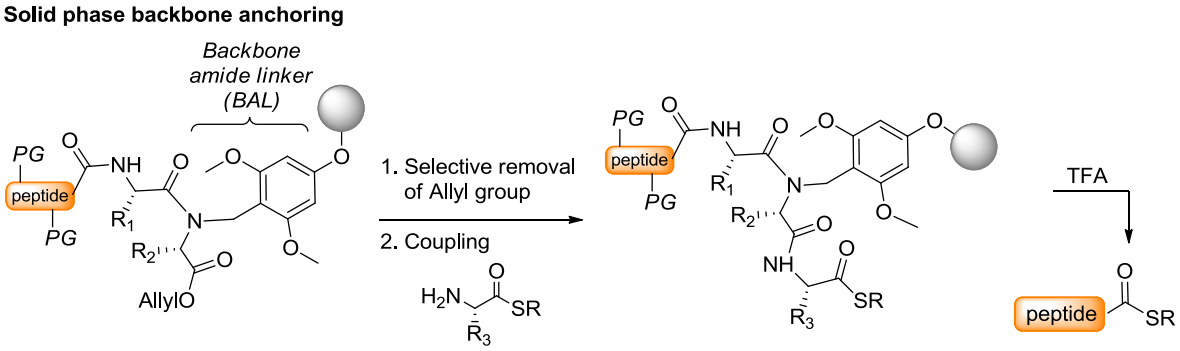


***Figure 13**. Synthesis of peptide thioester solid phase backbone anchoring, where the thiolysis occurs prior to resin cleavage with TFA[55].*

Apart from carboxylic group activation, alkylation, acylation, oxidation or condensation reactions were also used to trigger the formation of the thioester functionality

after peptide elongation by SPPS (**Figure 14**). As illustrated **Figure 14, a**, alkylation was employed to modulate the reactivity of the Kenner safety-catch sulfonamide linker [56,57]. Stable during the elongation by Fmoc-SPPS, the *N*-acyl sulfonamide group of the linker is activated by N-alkylation with TMS-CH_2N_2 or iodoacetonitrile, increasing its electrophilicity towards sulfur nucleophiles. Thiolysis leads therefore to the release of the expected peptide thioester in solution whereas subsequent TFA treatment enables the full deprotection of the peptide segment. Note that *N*-alkyl *N*-acyl sulfonamide linkage is stable when treated with TFA. As a consequence, *N*-alkyl *N*-peptidyl sulfonamides can be directly used as a thioester surrogates in NCL reactions.[58]

Acylation is another approach to obtain electrophilic imides and to produce peptide thioesters by thiolysis. A frequently used system is the 3,4-diaminobenzoic linker (Dbz) (**Figure 14, b**)[59–61]. After elongation of the peptide sequence by Fmoc-SPPS, acylation of the Dbz linker followed by TFA treatment provides unprotected *N*-peptidyl benzimidazolinones, which can be either converted into peptide thioester by thiolysis or engaged directly in NCL reaction as acylating agent. Note that a second generation of Dbz linker has been reported[62].

Prepared by Fmoc-SPPS starting from hydrazine solid support, peptide hydrazides are extremely popular precursors for the synthesis of peptide thioesters. A first method is to oxidise the hydrazide group with $NaNO_2$ (**Figure 14, c**). Thiolysis of the resulting *N*-acyl azide intermediate provides the target thioester. The oxidation step has to be handled at -10 °C to avoid undesired Curtius rearrangement[63–65]. An alternative method to access peptide thioester from peptide hydrazides is reminiscent of Knorr pyrazole synthesis. This method includes the condensation between hydrazide functionality and acetyl acetone with the formation of intermediate *N*-acyl pyrazole, which is then converted into the final thioester by thiolysis (**Figure 14, d**).

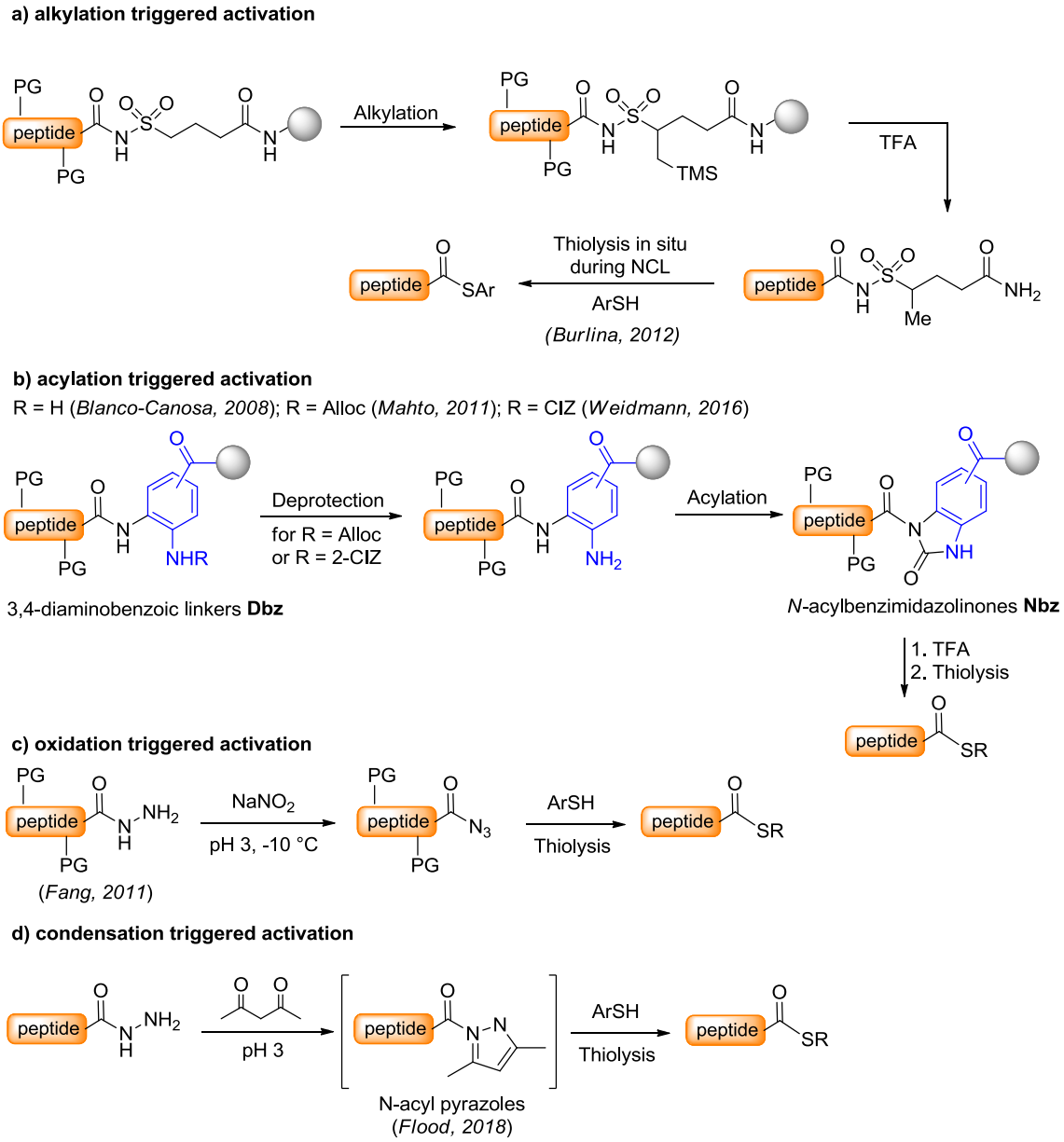


***Figure 14**. Examples of thiolysis starting from pre-activated C-ter amide; a) via alkylation[66], b) via acylation, c) via oxidation[67], d) via condensation[68].*

Finally, the last method classically used to produce peptide thioester is based on *N*,*S*-acyl shift systems. *N*,*S*-acyl shift systems are amide bonds *N*-functionalized with an alkylthiol appendage that rearrange through an intramolecular process to give a transient thioester (**Figure 15**). Present at the C-terminal extremity of the peptide, *N*,*S*-acyl shift systems provides the desired peptide thioester by thiol/thioester exchange reaction with an excess of exogenous thiol. The reaction is usually performed in acidic conditions since *N*,*S*-acyl shift is favoured at this working pH, probably due to the protonation of the amine in the rearranged form. Note also that the process is often carried out in the presence of TCEP to prevent thiols from oxidation.

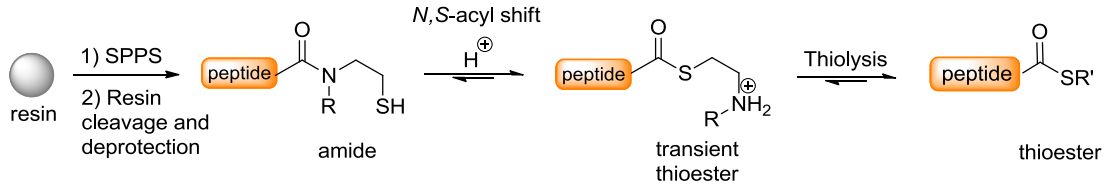


***Figure 15**. General scheme for N,S-acyl shift systems where the formation of transient thioester is driven by pH decrease. A target thioester is obtained by thiolysis of a transient thioester, becoming a leaving group.*

A variety of *N,S*-acyl shift systems has been developed up to now (**Figure 16**). Some of them including mercaptomethylated proline *(Nagaike)*[69] and thiazolidine *(Sharma)*[70] derivatives as well as *N*-2-(thioethyl)glycine residue (*Wierzbicka*)[71] were anchored to a solid support[72]. In these cases, thiol-thioester exchange leads not only to the formation of the target thioester but also to the cleavage of the peptide from a solid support. The peptide thioesters are simply recovered from the reaction mixture by filtration. In-solution systems are also popular due to facile set-up and monitoring of the reaction. Among in-solution *N,S*-acyl shift systems, *N*-alkylcysteine (*Hojo*)[73] bis(2-sulfanylethyl)amido (SEA)[74] and SEAlide thioester precursors are the most widely used. In case of SEAlide *(Tsuda)*[75], the rearrangement of the *N,S*-acyl shift system is induced at neutral pH by the addition of phosphate, in contrast to other cases.

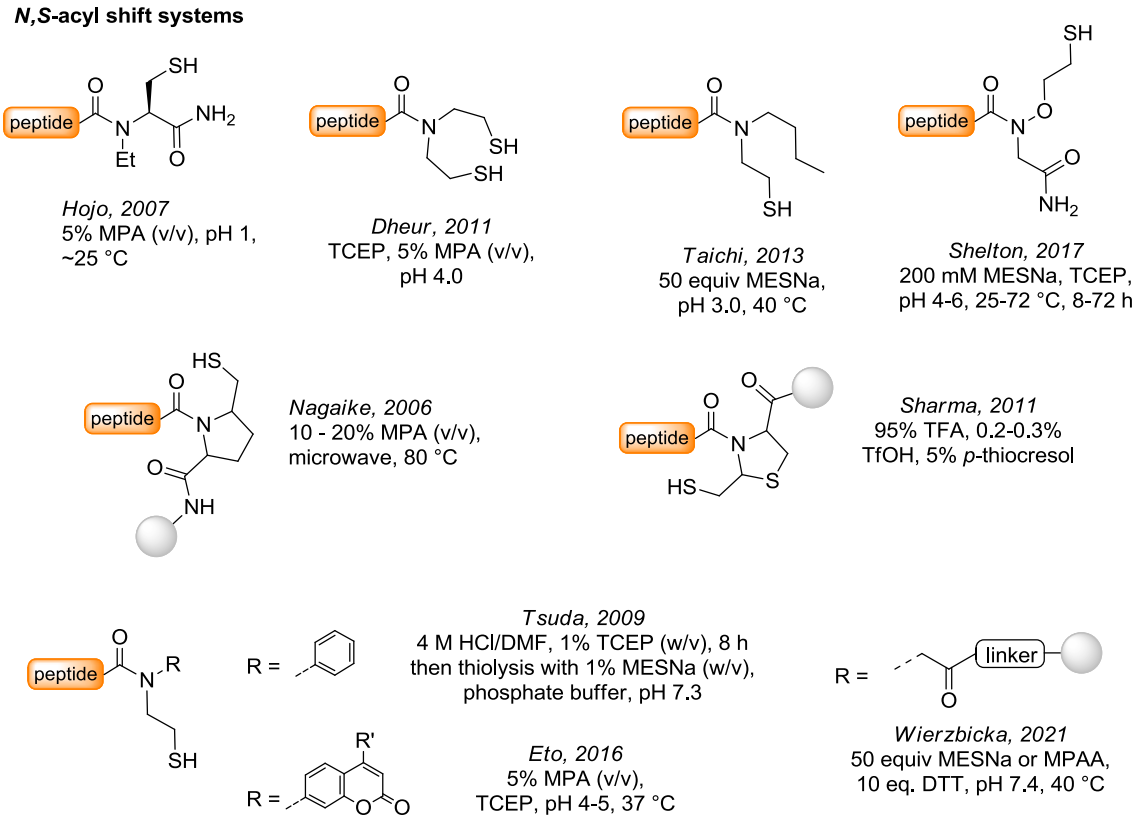


***Figure 16**. N,S-shift acyl donor systems that can be converted into thioesters via N,S-shift and thiol-thioester exchange.*

The synthesis of peptide alkyl thioesters by Fmoc-SPPS using the SEA chemistry was reported by our team 10 years ago[74,76] and a solid support functionalized with a SEA linker was designed for this purpose (**Figure 17**). Structurally, the two sulfanylethyl moieties of the SEA group are connected to a trityl-functionalized solid support through thioether linkages, making the SEA linker cleavable under acidic conditions. The free secondary amine of the SEA group is used as an anchoring point for the assembly of the peptide sequence by Fmoc-SPPS. At the end of the elongation process, the SEA peptide is released in solution and fully deprotected by reaction with TFA. The SEA group is then converted into MPA thioester by *N*,*S*-acyl shift and subsequent thiol/thioester exchange reaction with MPA. Practically, the procedure is carried out in a glovebox, in the presence of TCEP to keep SEA thiols reduced, and with an excess of MPA, in buffered solution at mildly acidic pH.

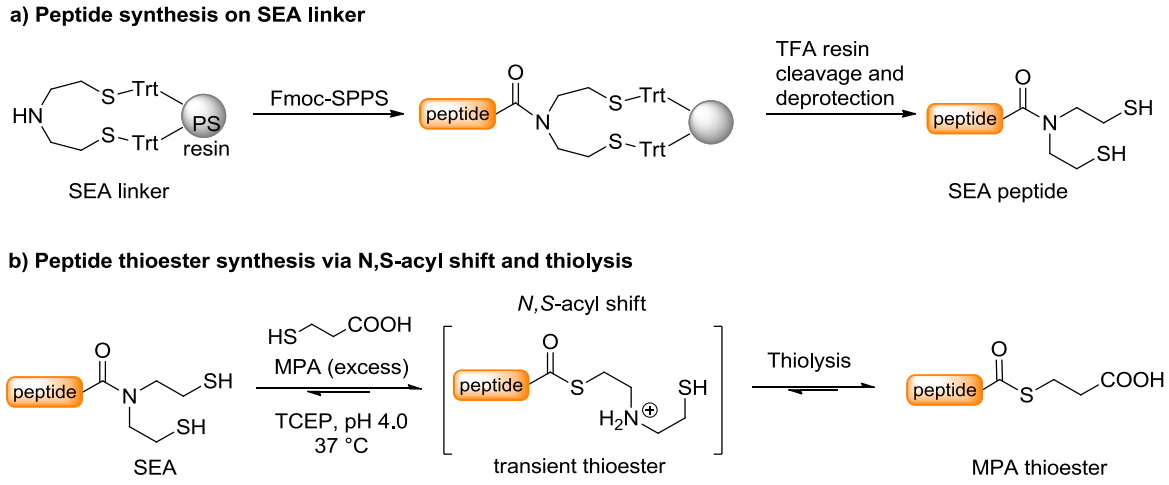


***Figure 17**. Synthesis of peptide MPA thioester using the SEA chemistry.*

In practice, alkyl peptide thioester derived from MPA, MTG or MESNa are often used as starting material for NCL due to their synthetical accessibility and relative stability to hydrolysis, compared to more reactive thioesters such as arylthioesters. To boost the reactivity of peptide alkyl thioesters, the addition of nucleophilic catalysts is crucial.

2.3 Nucleophilic catalysis of NCL

This section dedicated to nucleophilic catalysis is presented as a resume of the review "**Pedal to the metal: the catalysis of the native chemical ligation reaction**" which describes the general principles and usefulness of nucleophilic catalysis for accelerating NCL. This review was recently published by our team with my co-authorship[77]. The full document is available as **annex A**.

As previously described, the mechanism of NCL involves two steps: the capture of the cysteinyl peptide by the peptide thioester and the rearrangement of the thioester-linked intermediate by *N*,*S*-acyl shift leading to the final native peptide (**Figure 18**). At millimolar concentrations of peptide, the capture is the rate limiting step of the process and this process is very slow starting from low reactive peptide alkyl thioesters such as those derived from MPA or MESNa thiols. To compensate this low reactivity, nucleophilic catalysts are often added to enhance the NCL rate by *in situ* formation of a more efficient acyl donor than the starting peptidyl thioester.

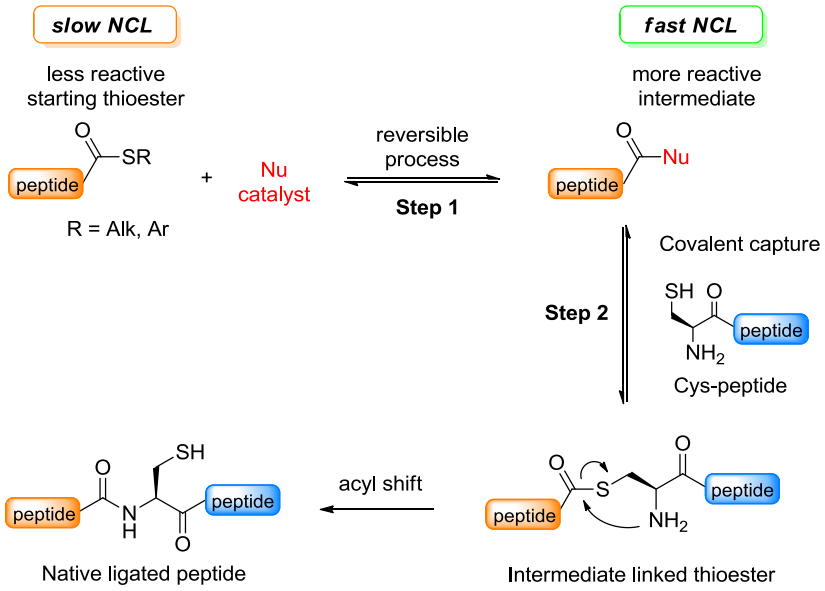


***Figure 18**. General scheme of NCL catalyzed by a nucleophilic additive.*

There are four general groups of nucleophilic catalysts developed to boost NCL ligation rate: thiolates, selenoates, phosphines and *N*sp^2 species such as imidazole derivatives (**Figure 19**). Although selenium and phosphorus-based compounds were reported in the literature to accelerate NCLs, I will focus here on the most employed additives, i.e. thiols, and on *N*sp^2-based catalysts. The latter are imidazole-derived, non-reducing alternative to thiols and are therefore of great interest for my PhD project dealing with the assembly of peptide segments under redox-controlled conditions.

***Figure 19**. Four main groups of nucleophilic NCL catalysts: thiolates, imidazole-derived, selenolates and phosphines.*

Thiol catalysts

Some thiols are known to be efficient catalysts of the NCL reaction. Catalysis proceeds through the *in situ* formation of an activated intermediate thioester by thiol/thioester exchange reaction (**Figure 20).**

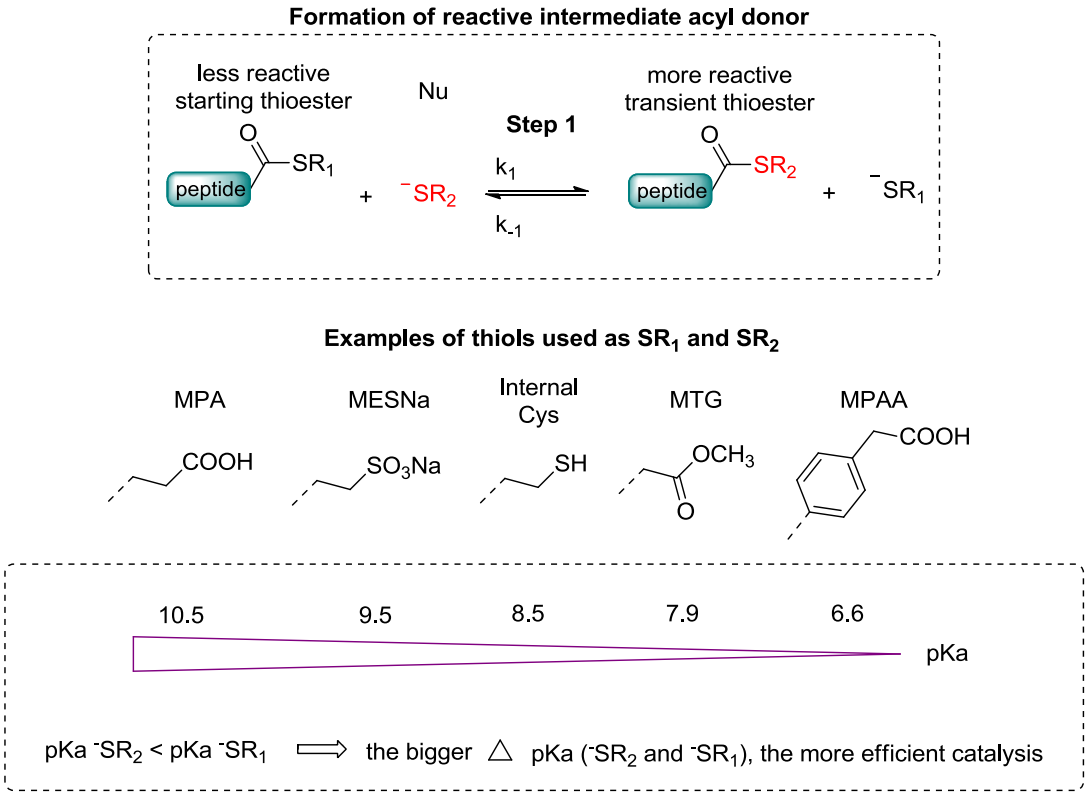


Figure 20*. Focus on Nu catalysis of the NCL-like step where the more reactive transient thioester is formed from the starting, less reactive thioester. The thiols examples are commonly used for NCL as thioesters and catalysts.*

According to Jenck's works, the equilibrium between starting and activated thioester is dictated by the dissociation constants (pK_as) of the thiol catalyst (HS-R_2) and the leaving thiol (HS-R_1) in the thiol/thioester exchange process. This process leads to the formation of a more reactive acylating species and therefore to NCL catalysis if the pK_a of R_2-SH is lower than the pK_a of R_1-SH. In this case, R_2S^- is more stabilized than R_1S^- at the working pH of the reaction and is therefore a better leaving group. However, the formation of the activated thioester in these conditions is thermodynamically disfavoured and a large excess of catalyst R_2-SH is usually added in the reaction mixture to force the formation of the activated thioester.

Thiols such as 3-mercaptopropionic acid (MPA), sodium 2-mercaptoethyl sulfonate (MESNa), methyl thioglycolate (MTG) or 3-mercaptophenyl acetic acid (MPAA) were successfully used as additives in NCL. However, only MTG or MPAA can provide a significant rate acceleration owing to their pK_a. It is important to note that not only added thiols can catalyse the reaction, but also thiols of internal Cys residues, due to their mid-range pK_a (**Figure 20**).

Among all the thiol catalysts developed up to now, MPAA is the gold standard *catalyst* for accelerating NCL, and for several reasons. First, MPAA is a catalyst of wide scope due to its pK_a of 6.6 which is below all the pK_a of the alkyl thiols classically used for peptide thioester synthesis. Furthermore, MPAA is easily soluble around neutral pH and can be used at high concentration in the ligation mixture. Finally, MPAA is relatively odourless and commercially available. However, its use has some limitations. The disadvantage of MPAA is HPLC complication as in some cases the analysed proteins co-elute with MPAA at close or idem retention time. It is therefore recommended to remove it from the reaction mixture prior to HPLC analysis and purification by acidifying the reaction mixture and extracting MPAA with diethyl ether.

A general limitation for thiol use is that in large excess they are not compatible with one –pot NCL/metal-free desulfurisation processes (Cys to Ala conversion at the ligation junction), as thiols and especially aryl thiols[78] quench this radical process. To overcome the pointed limitations, a thiol catalyst may be removed from the reaction mixture prior to desulfurisation. For example, trifluoroethanethiol (TFET) has a low boiling point what gives the possibility to evaporate the thiol upon NCL completion. Otherwise, the use of non-thiol catalysts for NCL is an alternative solution.

Azole catalysts

N-acyl azoles are efficient acyl donors toward S-nucleophiles. To improve NCL kinetics, *in situ* formation of *N*-acyl azole intermediates can be achieved by reaction between an azole catalyst and a peptide thioester. As for thiol catalysts, the intermediate formation process is thermodynamically unfavoured (shifted toward the formation of the thioester). Thus, high concentrations (2-5 M) of an azole catalyst are required to observe significant rate enhancements. Three types of azole catalysts were reported up to now: imidazole, 1,2,4-triazole and 2-methylimidazole[79,80] (**Figure 21**).

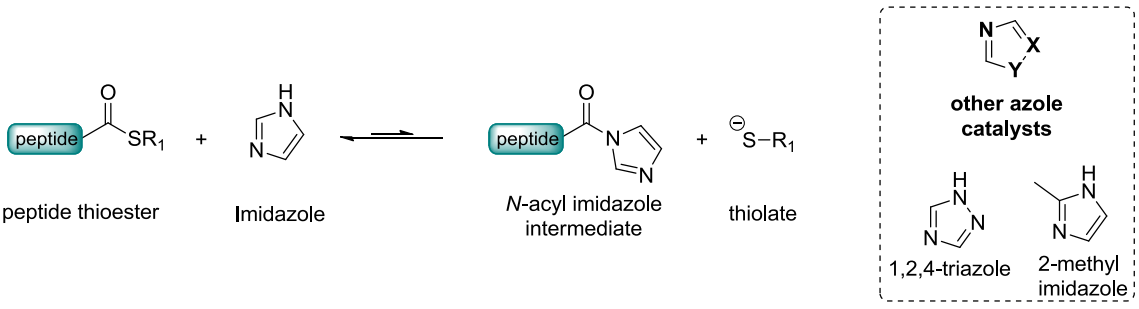


***Figure 21**. Activation of peptide thioester by formation of N-acyl imidazole intermediate.*

It is important to note that the ability of imidazole or related compounds to catalyze thioester hydrolysis has been clearly established. Catalysis proceeds through the formation of an acyl-imidazole intermediate that then reacts with water (**Figure 22**)[81].

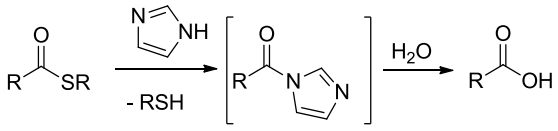


***Figure 22**. Thioester hydrolysis catalyzed by imidazole or histidine.*

To summarise, since NCL was reported in 1994 by Kent, it has been developed a lot. First, synthetic pathways were developed to synthesize the thioester and cysteinyl peptides used as coupling partner in the reaction. Second, the scope of the NCL was also widened with tools such as beta-mercapto amino acid junctions at the ligation site. Several groups of catalysts appeared to accelerate the NCL. Despite the NCL is a powerful tool in chemical protein synthesis, its scope in its classical form is restricted to the connection of two peptide segments. Is it possible to chain several NCL reactions and to produce this way longer and more complex polypeptides? Principle and recent advances in the field are described in the next section.

3 Assembly of polypeptides using two NLC reactions

Using several NCL ligations to assemble peptide segments produced by SPPS is an appealing strategy to access complex polypeptides. This strategy is even more interesting when the assembly process is performed in one-pot, i.e., without any intermediate purification. Indeed, intermediate purifications by reverse-phase liquid chromatography[82] are known to significantly decrease the global yield of the assembly process.

To orient the assembly process and to obtain a homogeneous polypeptide in terms of AA sequence after several NCLs, it is necessary at some point to control the reactivity of the species that are involved in each ligation, i.e., the thioester group or the Cys residue. There are two different approaches:

- the assembly in N-to-C direction targeting thioester group reactivity;
- the assembly in C-to-N direction targeting Cys residue reactivity.

Both concepts are classically used in chemical protein synthesis and will be illustrated in this section through the assembly of peptide segments *via* two successive NCLs. I will describe one-pot processes and their advantages in chemical protein synthesis.

3.1 Assembly in the N-to-C direction

Playing with thioester reactivity is an efficient strategy to assemble a protein in the N-to-C direction, for example by joining three peptides *via* two successive ligations (**Figure 23**). In this approach, the assembly starts with the ligation between a peptide thioester A and a bifunctional peptide B, armed respectively at its N- and C terminal extremities with a cysteine residue and another thioester function. Because of the difference of reactivity between the two

thioester functions present in the reaction mixture, the cyclization of the bifunctional B by NCL is disfavoured. The first ligation only provides the elongated peptide AB. Once the reaction is over, a cysteinyl peptide C is added to reaction mixture leading to the formation of the peptide ABC through a second NCL reaction.

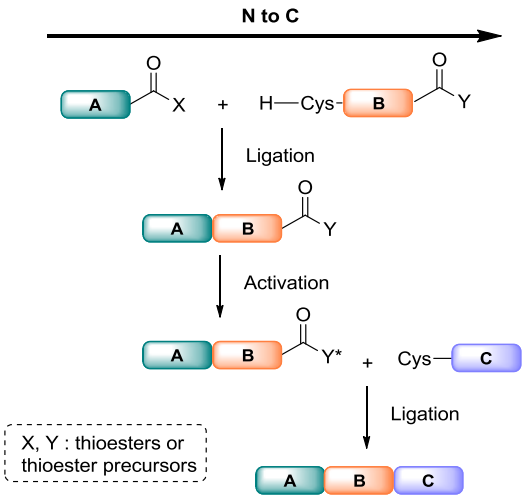


***Figure 23**. General scheme of three peptides assembly in one pot in N-to-C direction.*

To make this assembly strategy work, it is necessary to discriminate two thioester functions in term of reactivity. One possibility is to use two thioester functions that have very different acylating properties. This approach is known as Kinetically controlled ligation[83] (KCL). Another possibility is to shut down temporally the reactivity of one thioester function using a latent thioester. Both approaches will be discussed in the next section.

3.1.1 Kinetically controlled ligation (KCL)

The kinetically controlled ligation (KCL) concept proposed by Kent is based on competitive ligation reactions, playing on the difference of reactivity between aryl and alkyl thioesters[83]. The idea is to perform the NCL between a peptide armed with a C-ter aryl thioester and a bifunctional peptide equipped with a N-ter cysteine and a C-ter alkyl thioester (**Figure 24**). In the absence of a thiol catalyst such as MPAA, the ligation involving the aryl thioester is several orders of magnitude faster than the one involving the alkyl thioester. As a consequence, the ligation between the two coupling partners is kinetically favoured compared to the oligomerization ($k_1 >> k_2$) or the cyclization ($k_1 >> k_3$) of the bifunctional segment. Once the first ligation is completed, a cysteinyl peptide as well as a thiol catalyst are added in the reaction mixture. The presence of an added thiol catalyst accelerates the second ligation between the elongated alkyl thioester produced during the first ligation and the added cysteinyl peptide.

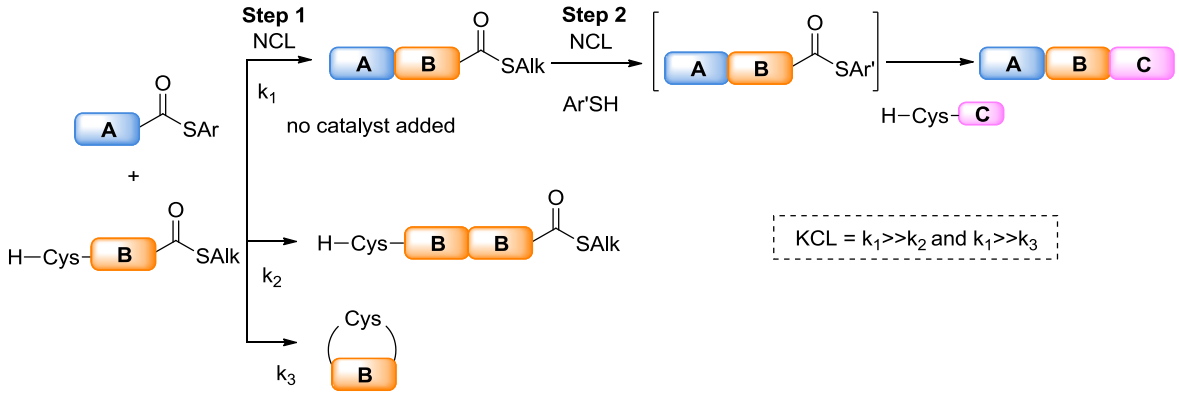


***Figure 24**. Kinetically controlled ligation, where $k_1>>k_2$ and $k_1>>k_3$.*

An example of successful KCL application was demonstrated by Kent with synthesis of small amphiphilic protein crambin, naturally produced by a cabbage. The assembly of six fragments was performed *via* two KCLs of three short peptides with further purification and terminal ligation[83].

Note that during KCLs all the thioesters present in the reaction mixture are reactive, but with very different rates. This not the case of peptide assembly strategy based on latent thioesters.

3.1.2 Peptide assembly based on latent thioesters

Latent thioesters by definition are silent under classical NCL conditions and must be activated by addition of reagents or modification of physicochemical parameters of the reaction mixture (pH, temperature, redox power, etc.), leading to the formation of a reactive thioester. Various latent thioesters have been designed so far. They include peptide hydrazides or *N*,*S*-acyl shift systems such SEA group[84].

Peptide hydrazides

As previously discussed, peptide hydrazides are activated by treatment with $NaNO_2$ and further thiolysis. It makes them popular for sequential peptide assembly as hydrazides are unreactive under classical NCL conditions[63]. However, the $NaNO_2$ activation step can't be performed in the presence of thiol additive, especially the one used in large excess to catalyse the NCL reaction. It is therefore impossible to perform two successive NCL reactions in one-pot: the elongated peptide hydrazide needs to be first isolated from the thiol catalyst prior to the next activation step (**Figure 25**).

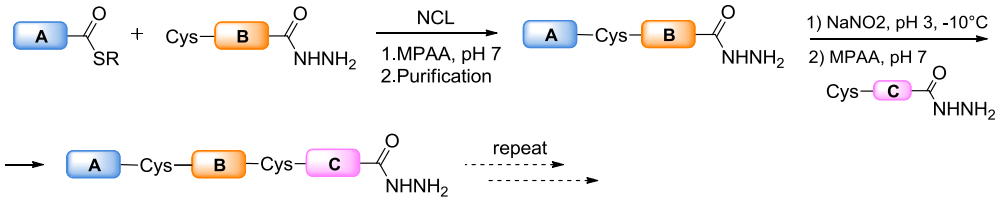


***Figure 25**. Assembly of three peptides in N-to-C direction when hydrazide is the thioester precursor.*

To make hydrazides compatible with sequential NCLs in one pot, Dawson performed the activation of the peptide hydrazide with acetylacetone instead of $NaNO_2$, through intermediate acyl pyrazole formation, which is compatible with sequential NCLs (**Figure 26**)[68].

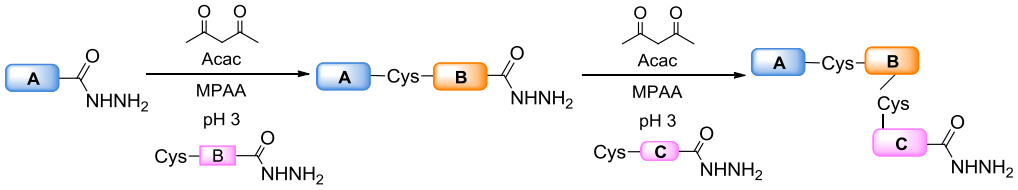


***Figure 26**. Hydrazides activation with Acac applied to one pot NCLs.*

This hydrazide activation technique was applied to the semisynthesis of serotonylated histone H3 by combining peptide desulfurisation and NCL[85]. Currently, the use of hydrazide precursors for NCL is one of the most practical and popular alongside with *N,S*-acyl shift systems.

Latent thioester surrogate based on *N,S*-acyl shift system

Some *N,S*-acyl shift systems are considered as latent thioesters as an amide functionality can be switched to thioester. With such systems, the intermediate thioester produced by intramolecular rearrangement can be used in situ as acyl donor precursor in NCL-like reactions. It can be operated as two-state switch system, where the amide form is unreactive while the thioester form is reactive.

Among the latent thioester surrogates presented in the sub-section 2.2.3, I will focus on the example of the tool bis(2-sulfanylethyl)amide (SEA) that has been developed by our team [76].

SEA functionality can exist in reduced and oxidized states (**Figure 27**). The peptide obtained by SPPS contains SEA in reduced form, indicated as SEA^{on}. SEA^{on} can be oxidized, for example using iodine, with the formation of a disulfide bond, turning to SEA^{off} state. Reversely, SEA group can be reduced in strong reducing media, for example by phosphines.

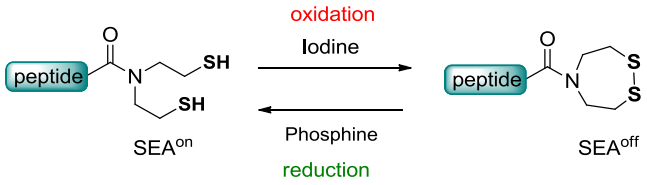


***Figure 27**. Transition between reduced and oxidized state of SEA.*

To sum up, the SEA group is a two-state, redox-controlled chemical device and such a property was exploited by the CBF team to design a latent thioester that enables to place the assembly of peptide segments under redox control. SEA^{off} group is unreactive during classical NCL reaction catalysed by MPAA since MPAA is not reductive enough to reduce the disulfide bond of the SEA group (**Figure 28**, X = S**)**. In contrast, the presence of strong reductants such as DTT or TCEP in the reaction mixture induces the ring-opening of SEA group, triggering its rearrangement into SEA thioester. The latter can react *in situ* with a cysteinyl peptide to give a ligation product. Note that the same principle works with the selenium analogue of the SEA group, i.e., the bis(2-selenylethyl)amide (SeEA) system. [86,87] (**Figure 28**, X = Se). Besides providing extended redox-control during complex peptide assemblies, SeEA group is also more reactive than its sulfur analogue and gives higher ligation rates.

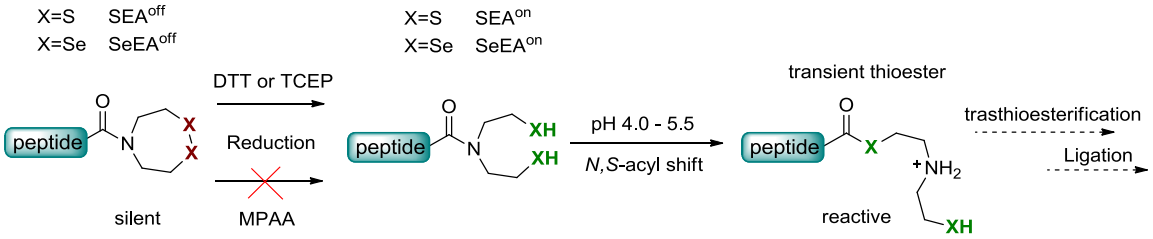


***Figure 28**. SEA main properties: reduction by strong reductant and transient thioester formation.*

Another modified version of the SEA group is ^{oxo}SEA which provides fast *N*,*S*-acyl shift and highly efficient ligation even at nanomolar scale due to enhanced electrophilicity of SEA amide (**Figure 29**)[88].

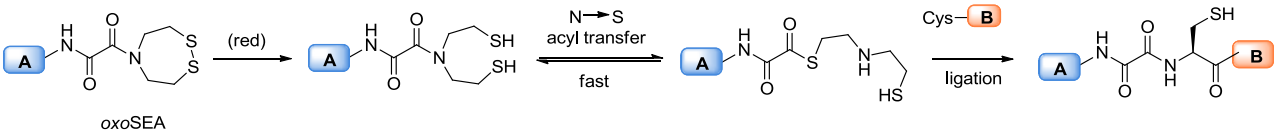


***Figure 29**. SEA-derived device oxoSEA applied to induce the NCL.*

The latency of the SEA group was applied to synthesize several proteins. For example, the K1 domain of the hepatocyte growth factor (HGF) starting from three peptide segments was produced in one-pot using SEA chemistry (**Figure 30**)[84]. The first ligation is a classical NCL

between peptides A and B, catalysed by MPAA and performed at pH 7.2. During this step, SEA group of the peptide B remains unreactive under its oxidized form. Once the reaction is over, TCEP was added in the reaction mixture to induce SEA reduction and pH was lowered down to 5.5 to favour SEA rearrangement into a reactive thioester. This thioester generated *in situ* was coupled with cysteinyl peptide C. The target product was isolated with the yield of 40 % and then folded.

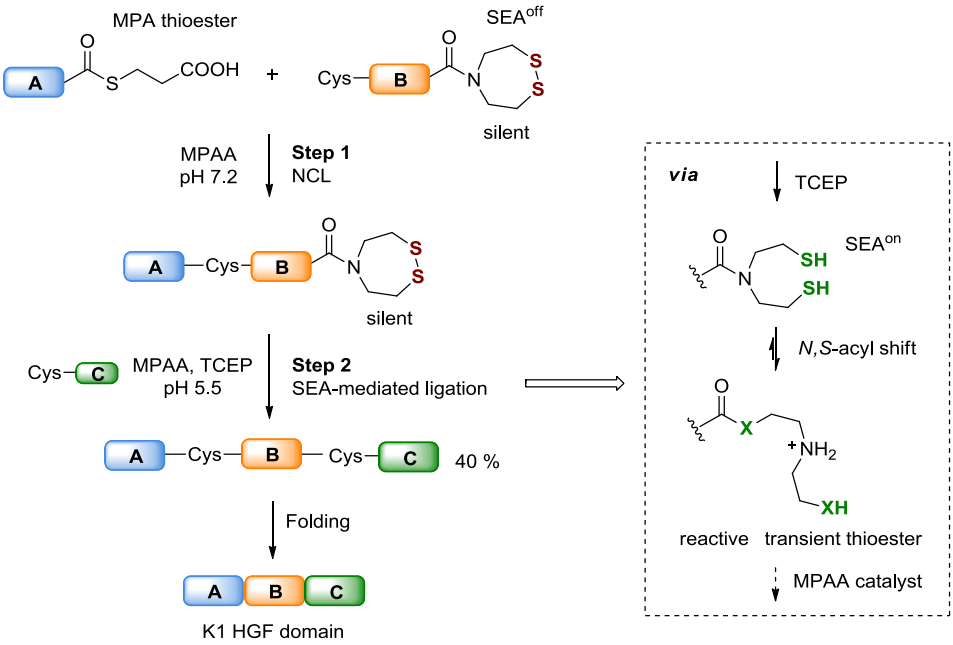


***Figure 30**. K1 HGF domain synthesis by one pot assembly of three peptide fragments using redox-control of the SEA group.*

The property of surrogate thio- or selenoesters to remain silent under NCL conditions has great practical interest in performing more than one ligation in one pot. This concept will be more elaborated in section 4.

3.2 Assembly in the C-to-N direction

NCL is a highly chemoselective reaction. However, to discriminate the reactivity of Cys residues during complex peptide assemblies, the use of Cys protecting groups is compulsory.

In the following scheme (**Figure 31**), assembly of 3 peptide segments by 2 successive ligations, the Cys-residue of the middle peptide segment B is temporally masked by a protecting group. The thioester present at the other extremity of the peptide B is available to react with the Cys-peptide C during the first NCL reaction. Upon NCL completion, the protecting group is removed and the resulting Cys-peptide BC can be coupled to the peptide thioester A through the second ligation. The masking of cysteine is necessary to avoid phenomena of oligomerization or cyclization.

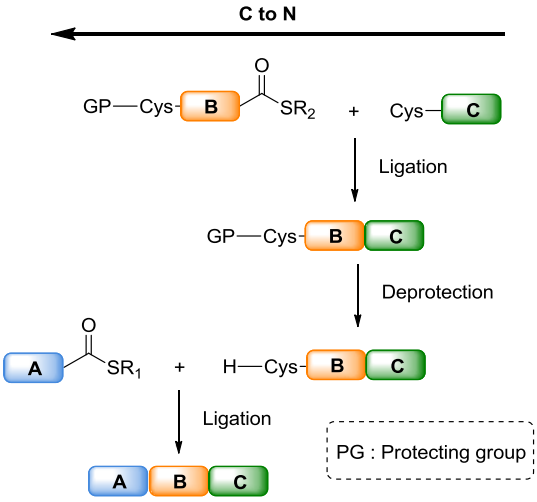


***Figure 31**. Peptide assembling strategy in C-to-N direction.*

It is then interesting to discuss the existing protective groups for Cys-residue and their compatibility with concatenation of three or more peptides.

3.2.1 Cys protecting groups

The protecting group (PG) of cysteine used to assemble peptide segments by NCL must be stable under the ligation conditions and be eliminated under mild conditions (**Figure 32**). They can be attached to the alpha amine function of the cysteine residue (Msc)[89], to the its thiol group (Acm[90,91], Mapoc[92]) or to both of them (Thz)[91,93] and are removed under basic conditions, using light irradiation or in the presence of reagents such as transition metals or *O*-methyl hydroxylamine.

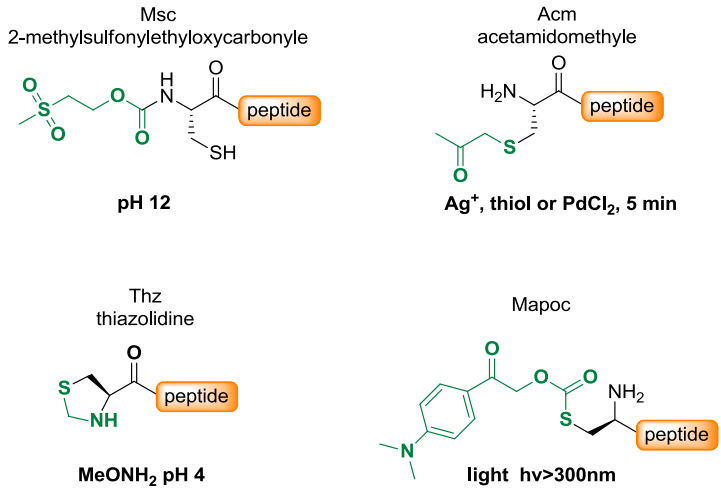


***Figure 32**. Protecting groups for C-to-N assembly of peptides.*

Note that some Cys protecting groups are compatible with one-pot NCL/deprotection or NCL/deprotection/NCL sequences. These methods enable to minimize intermediate purification steps and to improve the global yield of the target protein synthesis. In 2004, Bang

used for example this strategy to synthesize the crambin protein starting from three peptide segments without performing intermediate purifications[91]. A Thz protecting group was used to mask transiently one of the key cysteines present in the target peptide sequence and to orient the peptide assembly.

More recently, synthesis of snake toxin mambalgin, acting as painkiller, was performed in one-pot by Zuo et *al.* with the overall yield of 35 % (**Figure 33)**. In this approach, an azide group was used to temporarily mask the amine of an *N*-ter cysteine. The latter was converted into the reactive amine by addition of TCEP, a strong reductant, in aqueous media.

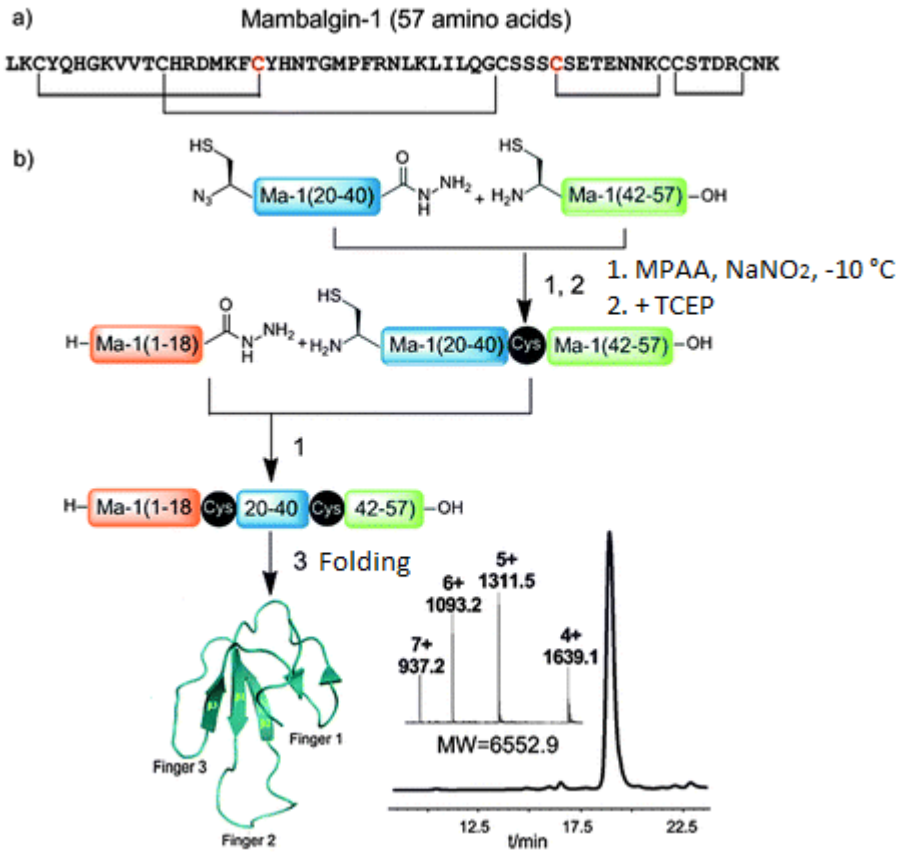


Figure 33*. Synthesis of snake toxin mambalgin in one pot by three peptide fragments assembly in C-to-N direction (adapted from Pan et al[94]).*

This last example suggests that using the reducing power of the reaction mixture for controlling cysteine reactivity during peptide assemblies appears as an attractive and promising alternative to pH change or light irradiation. This idea is bioinspired as changing of cysteine residue redox state is a strategy used by nature to control biological processes[31]. In organic synthesis, controlling the redox properties of thiols is often achieved *via* formation of disulfide bonds. As illustrated in **Figure 34**, several cysteine PGs forming disulfide bonds are available to protect the thiol group of cysteine[95]. But, such acyclic disulfides are labile in the reducing conditions imposed by the thiol catalyst of NCL and are therefore not adapted for redox-controlled peptide assembly.

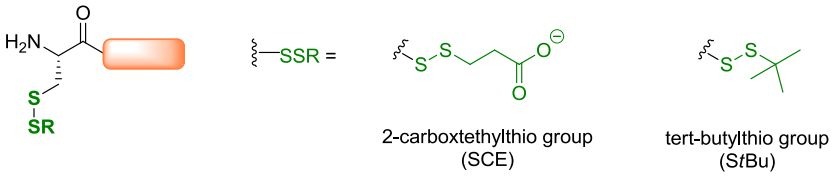


***Figure 34**. Examples of redox cleavable protecting groups for N-terminal cysteine.*

Cyclic dichalcogenides are less prone to be reduced than acyclic derivatives and such a property opens the way towards new strategies of assembly. In this field, the CBF developed recently the selenoethyl cysteine residue (SetCys)[96] and this work is the heart of my PhD works.

3.2.2 The SetCys residue as redox controlled cysteine surrogate

SetCys (*N*-selenoethyl cysteine) is the artificial amino acid where amine and thiol groups are linked *via* a selenoethyl bridge forming a cyclic structure. After reduction of the selenosulfide bond, this residue has the capacity to lose spontaneously its selenoethyl arm, leading in a traceless manner leading to native a cysteine (**Figure 35**). The cleavage of selenoethyl arm occurs *via* carbon-nitrogen bond break under mild conditions in aqueous media. To compare, classically carbon-nitrogen bond cleavage is difficult and demands harsh conditions, metal catalysis or radical reactions. Note that the presence of selenium atom is needed to make the *N*-alkyl substituent cleavable. Indeed, sulfur analogue of SetCys is not cleavable under the same conditions.

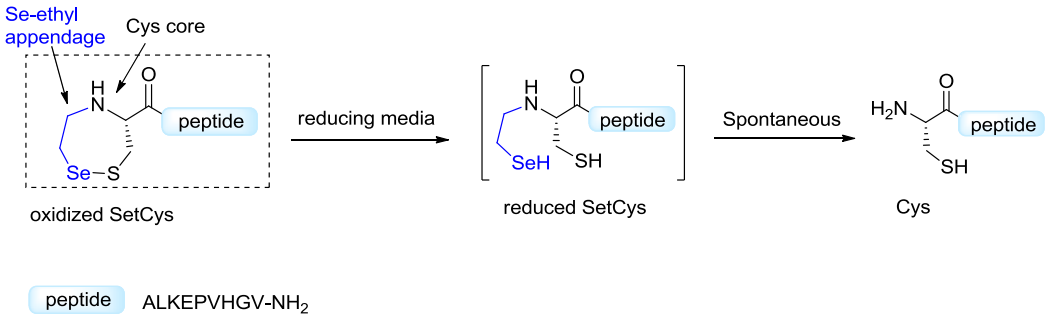


***Figure 35**. SetCys structure and the key property of latent cysteine.*

From a mechanistic point of view, it is unlikely that the selenoethyl limb cleavage involves radical intermediates since the reaction media contained large excess of MPAA and sodium ascorbate, which are known to be powerful radical quencher [78]. Note that the presence of sodium ascorbate was necessary to avoid deselenization caused by TCEP present in excess. The SetCys decomposition occurs rather *via* an ionic mechanism where an intramolecular substitution of the ammonium group by the selenoate ion leads to the release of the native cysteine and to the formation of an episelenide (**Figure 36, a**)[39,97]. In the presence of TCEP, the latter decomposes into ethylene and TCEP=Se.

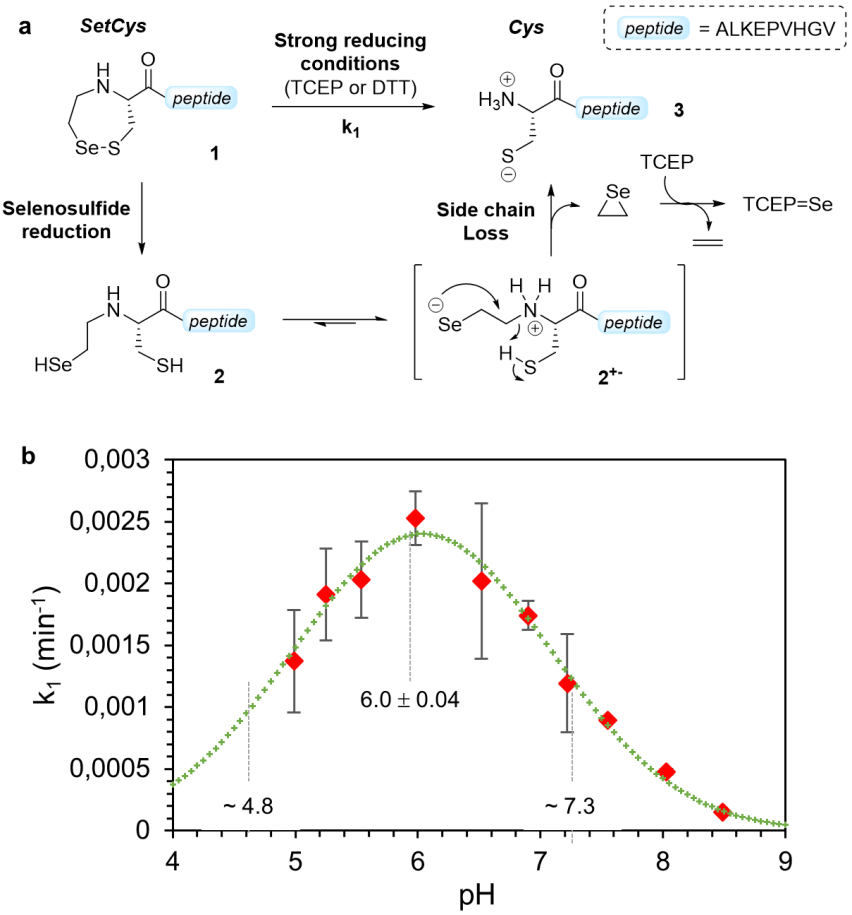


***Figure 36**. Mechanistic studies of SetCys reactivity a) the proposed mechanism of SetCys decomposition; b) Rate of the SetCys decomposition as a function of pH. (adapted from paper[96]).*

Some experimental data are support this anionic mechanism. For example, it was established that the rate of the SetCys to Cys decomposition is maximal between pH 4.8 and and 7.3 (**Figure 36, b**). Since this values are close to the pK_as of the selenol and ammonium groups[98], this pH range likely corresponds to the predominance zone of the selenoate/ammonium species $\mathbf{2}^{+-}$ (**Figure 36, a**). This intermediate contains in its structure the best nucleophile and nucleofuge for the intramolecular substitution and is believed to be responsible for the high rates of SetCys decomposition.

To prospect the capacity of SetCys to act as cysteine surrogate in NCL reaction, a model SetCys peptide was dissolved in various ligation media (**Figure 37**). In the absence of any coupling partner, the SetCys residue turned out to be stable when was incubated with MPAA: no loss of the selenoethyl arm is observed under classical NCL conditions. In the presence of the peptide thioester, the formation of ligated peptide containing the SetCys residue in the middle of its sequence was formed, suggesting a transient ring-opening of SetCys during the process and the capture of the intermediate opened form by the thioester. However, this reaction is kinetically disfavoured when a competitive cysteinyl peptide is introduced in the

reaction mixture. In other words, SetCys is silent during classical NCL reaction and does not interfere with the outcome of the reaction.

The situation is clearly different when a strong reductant such as TCEP is added in the reaction mixture. Complete reduction of Se-S bond triggers the decomposition of SetCys into Cys, and this process in the presence of a peptide thioester leads to the formation of a native ligated peptide.

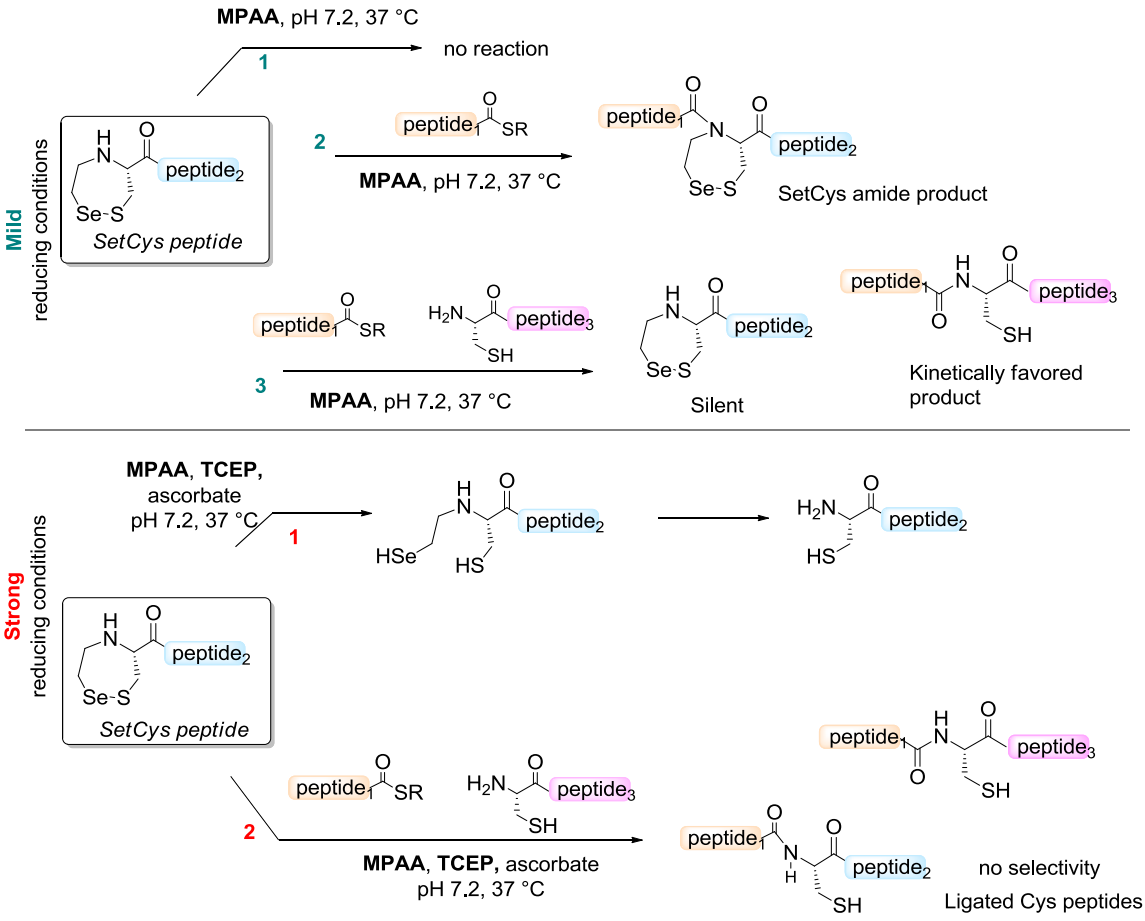


***Figure 37**. SetCys peptide properties in ligation media. Under mild reducing conditions, SetCys remains silent in the presence of NCL reacting partners. Under strong reducing conditions, SetCys is reduced what leads to its activation and loss of selectivity. (adapted from paper[96]).*

SetCys chemistry was applied by the CBF team to design new strategies that give access to linear and cyclic analogues of the HGF K1 subdomain. A detailed protocol, recently reported by Firstova *et al.*, showed how a mixture of three peptides was concatenated in one pot to produce selectively the linear sequence (**Figure 38**).

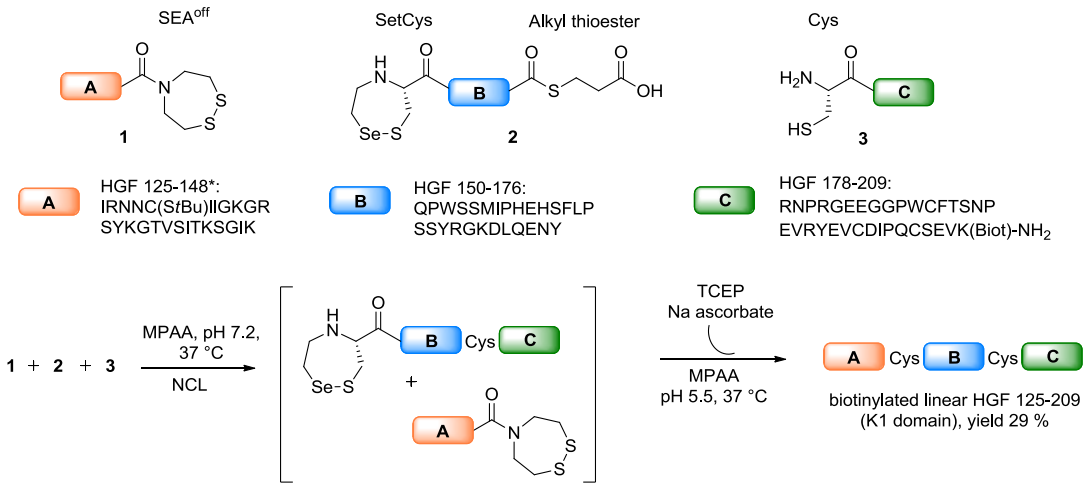


***Figure 38**. One-pot strategy to access the linear K1 domain with use of SetCys chemistry. (adapted from paper[99]).*

Experimentally, a bifunctional peptide equipped with a N-ter SetCys and C-ter alkyl thioester (peptide B) was mixed with two monofunctional SEA^{off} and Cys peptides (peptides A and C). The first ligation performed in the presence of MPAA led to the formation of the BC segment, the SEA group and the SetCys residue remaining silent under the reductive conditions of the reaction. Then, the addition of TCEP induced the activation of the SEA latent thioester and the conversion of SetCys into Cys. These two processes worked synergistically during the second ligation step to provide the final ABC. Note that the pH was decreased from 7.2 to 5.5 during the second ligation since SetCys to Cys conversion as well as the rearrangement of the SEA group are faster at more acidic pH.

To summarise, N-to-C and C-to-N elongation strategies relying on the use of latent thioester and Cys protecting groups were developed to concatenate several peptide segments. These strategies or a combination thereof were used to access complex proteins *via* linear or convergent peptide assembly (**Figure 39**).

Convergent strategy relies on the assembly of intermediate peptide blocks that have been already produced by concatenation of smaller peptide segments. For example, the described synthesis of the protein Crambin is related to convergent strategy[91], as well as "covalent dimer" HIV-1 protease synthesis described by Torbeev and co-workers[100].

This approach is more secure and efficient in terms yields than linear strategy where the peptide segments are introduced one by one in the peptide sequence though successive ligations. However, such a linear process can be improved thanks to the development of one-pot iterative process. This is the subject of the next section.

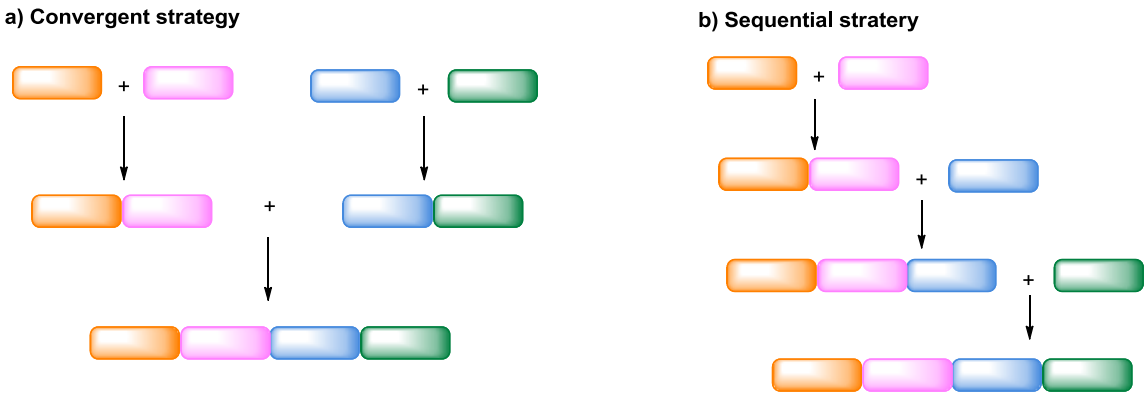


***Figure 39**. Strategies of multiple peptide fragments concatenation. a) Convergent strategy; b) Sequential strategy.*

4 Iterative assembly of peptide segments

Iterative or sequential process is the repetition of the same coupling process to insert one by one the peptide segments that compose the final polypeptide. This strategy is similar to SPPS, but instead of single amino acids there is an assembly of peptides one after another. Researchers are making efforts to design a process with no intermediate purifications between elongation cycles to improve the yield of the assembly process.

One solution is to perform the ligation on solid support. As for SPPS, the washing of the solid phase after each reaction enables to remove reagents from previous reactions and to repeat multiple elongation cycles without any perturbations. As illustrated **Figure 40**, the K1 subdomain of HGF, possessing biological activity and composed of 90 amino acids was synthesised by our team in the N-to-C direction on water-compatible resin using a acetoacetyloxime (AcA) linker that could be cleaved in a traceless manner *via* transoximation.[101] Each elongation cycle involves the coupling of a cysteinyl peptide to the solid support by NCL and the activation of the latent SEA thioester that functionalizes the other extremity of the peptide segment by a strong reductant. The SEA^{on} group is then converted into a MPA thioester which is ready to react in another elongation cycle.

However, assembly on solid support brings some difficulties[102]. It is not practical to follow the process analytically as the resin must be cleaved for every sample prior to analysis. Another problem that might appear is the modification of reactivity of the peptide segment due to the presence of the solid support. Diffusion and local concentrations of reacting molecules are not the same than in solution.

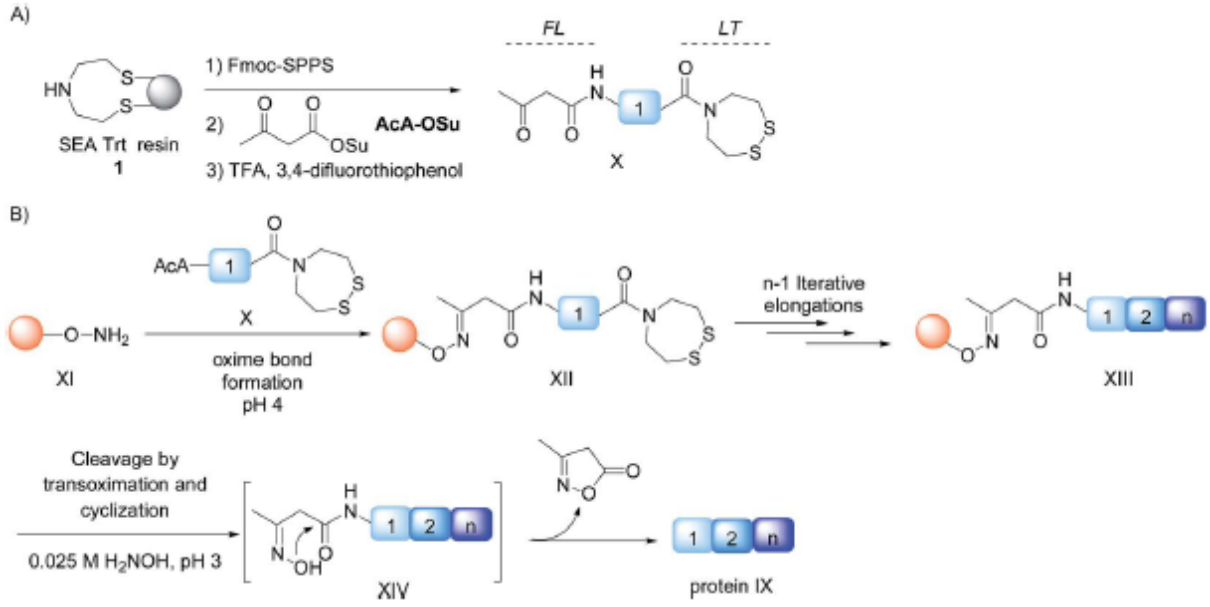


***Figure 40**. Iterative polypeptide assembly on solid support which is cleavable in traceless manner (adapted from paper[101]).*

Another possibility to perform iterative peptide assemblies is to use in-solution process with no intermediate purifications. Although analytical monitoring of the process is not problematic in this and that it is easier to control the mixing and stirring of a batch, accumulation of reagents after each elongation cycle in the reaction mixture remains the main problem of the approach. Ideal iterative process relies:

- on an efficient coupling process of peptide segments, to avoid the side-reactions that will increase the complexity of the reaction mixture composition at each elongation cycle;
- on a coupling process that requires no drastic changes of the reaction conditions and minimum intervention.

To perform one-pot, iterative peptide assemblies, one possibility is to use light to unmask the reactivity of the key Cys residue or thioester needed to orient the peptide assembly at each elongation activation. However, this approach requires a special set-up. Note that some methods using the pH or the temperature to control the peptide assembly were also reported. In my thesis, I want to use the SetCys chemistry and the ability of this cysteine surrogate to lose its selenoethyl appendage under well-defined redox conditions to develop a novel one-pot, iterative assembly process using temperature as an external mean of activation.

Combining native chemical ligation and SetCys chemistries – results and discussion

5 Main goals

Rendering chemical protein synthesis more popular can be done primarily by process simplification, which can include many aspects such as reducing the number of chemical steps, isolation steps, limiting costs and waste produced. Considering the importance of accessing long polypeptides by chemical synthesis, our main goal was to simplify polypeptide elongation. In particular, we wanted to develop an NCL based one-pot elongation cycle to access polypeptides by iterative coupling of peptide segments, with the important constraint of adding all reagents needed for the elongation cycle at the beginning of the process (**Figure 41**). To precise, the elongation should be efficient, devoid of significant by-products formation and not require intermediate purifications. In such a process, the elongation proceeds from C-to-N. Upon the end of the first ligation step (step 1), the N-ter part is activated (blocked cysteine, step 2) to react in the next iteration. Because all the reagents needed for the elongation cycle are introduced at the beginning of the process, a critical part of my work was to find a way to achieve this activation step *externally*. Once the two-step process is completed, the next bifunctional peptide is added to continue the elongation.

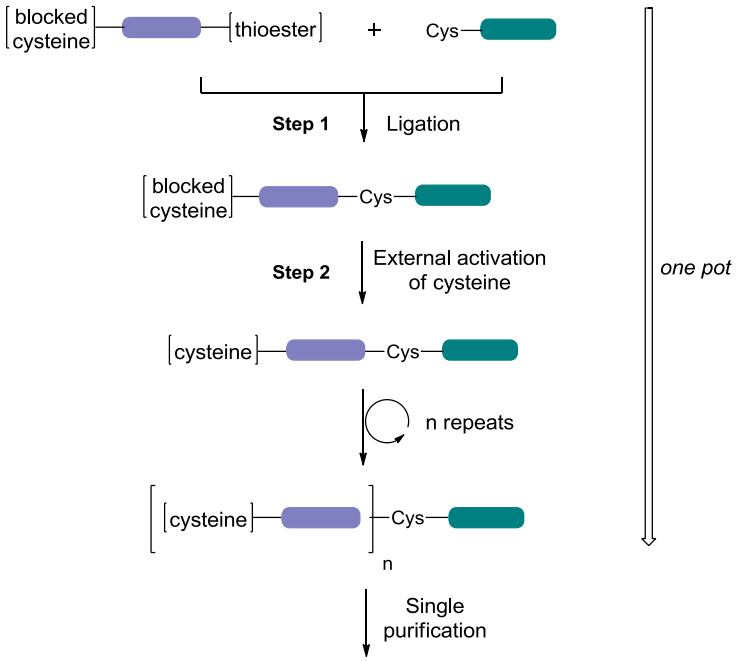


***Figure 41**. General concept of novel one-pot peptide assembly in C-to-N direction.*

To develop such a one-pot process, we propose to exploit SetCys chemistry. For that, we plan to develop NCL-like ligation using bifunctional peptide segments functionalized with SetCys and a thioester group at the N- and C-termini respectively. Thioester plays the role of acyl donor for the NCL ligation, SetCys represents the precursor of cysteine needed for the next NCL. The key point is to keep SetCys residue as silent as possible during NCL. Knowing that

NCL reaction proceeds efficiently in reducing conditions, and that the reactivity of SetCys strongly depends on the reducing power of the mixture, I had to find a way to control each chemical process with minimal interference by finding an appropriate redox system, keeping in mind that all the reagents are added at the beginning of the process. This control is important in order to avoid side reactions that would interfere with the peptide elongation process.

Controlling the reducing power of the media can be managed by adding a phosphine and/or thiols at the beginning of the elongation process. In a previous work[96], a two redox state strategy relying on the SetCys chemistry was used for assembling peptide segments. It was established that in mild reductive conditions imposed by aromatic thiol MPAA, SetCys remains mostly in its cyclic selenosulfide form, and therefore is silent during the NCL reaction. Under such conditions, the selenoethyl arm loss is not observed. In contrast, under strong reductive conditions such as in presence of TCEP, SetCys is opened reductively and efficiently converted into Cys *via* a spontaneous loss of the selenoethyl arm (**Figure 42**).

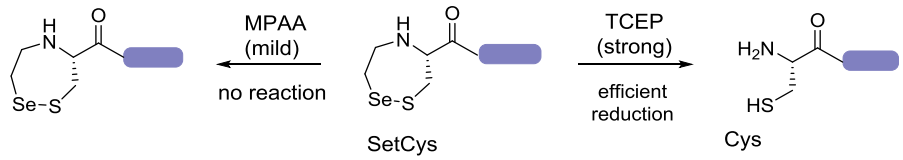


Figure 42. *Properties of SetCys in the presence of mild and strong reducing agents.*

Thus, running the NCL step in the presence of MPAA only followed by TCEP addition at the second step guaranties good selectivity. However, this method has two main limitations. First, the purification of the reaction mixture is required to eliminate the TCEP before the next ligation step. Otherwise, premature ring opening of the SetCys and its conversion to Cys leads to selectivity loss. Second, the ligation mixture must be manipulated between NCL and SetCys-to-Cys conversion (addition of TCEP, pH adjustment, dilution), and this is typically what we want to avoid.

For the development of a novel strategy based on SetCys chemistry, we need to delay SetCys opening and Se-ethyl arm loss during the first NCL step (**Figure 43**), and find a parameter to externally trigger the SetCys-to-Cys conversion in the second step.

To achieve this goal, it is necessary to work at constant reducing power. Thus, the first aim of the project is to find appropriate redox conditions by studying the parameters and understanding the mechanisms implicated in the ring opening of SetCys by different S and P nucleophiles, and its consequences on the SetCys into Cys conversion. The second objective is to use the temperature as an external trigger of SetCys-to-Cys conversion. As we had to change significantly the mixture composition to control SetCys reactivity, we had to design the conditions well adapted for the NCL step.

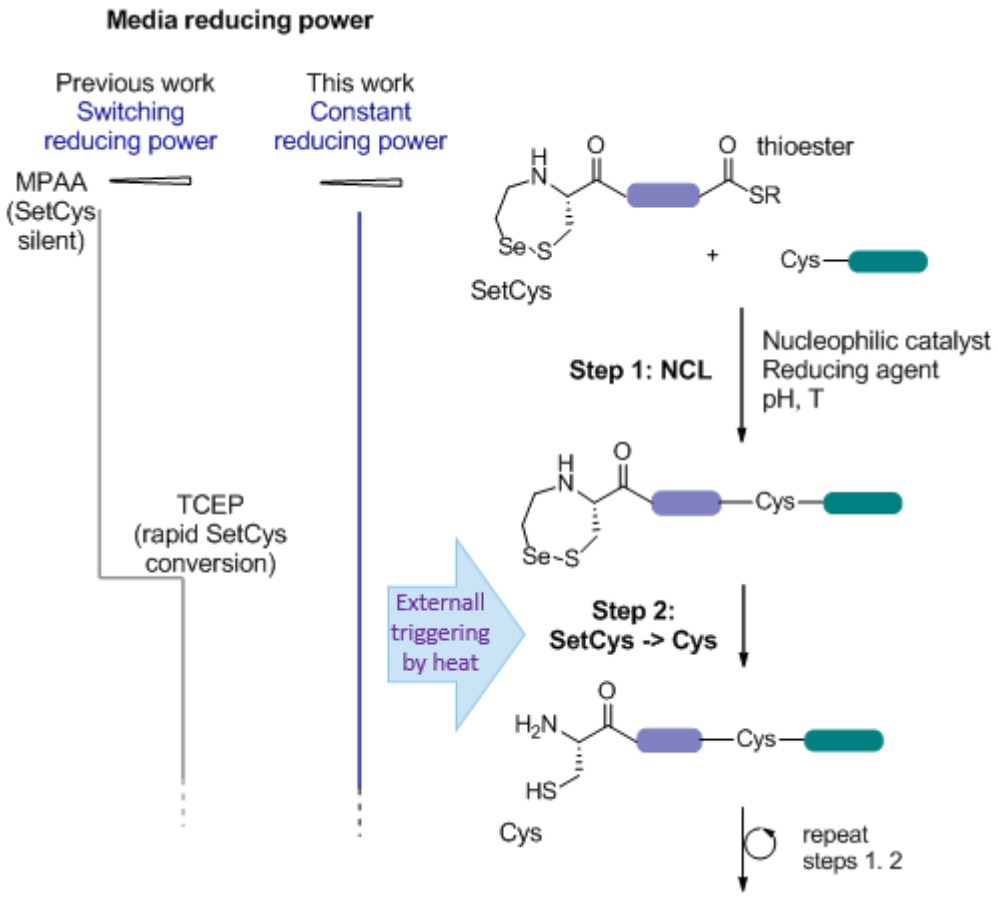


***Figure 43**. The scheme of novel peptides concatenation with use of SetCys reactivity, externally triggered by heat.*

Once the novel ligation strategy will be confirmed with simple model peptides, we plan to apply this ligation strategy for *Tm*AFP analogues chemical synthesis. *Tm*AFPs are natural anti-freeze proteins which have a practical interest and will be discussed in following chapters.

6 Two-step process set up

In the strategy presented in **Figure 43**, masked *N*-ter Cys must be silent during the NCL reaction but activated subsequently without changing the composition of the mixture. This is highly challenging and one potential solution to this problem is to play with the temperature, which can be varied externally. Thus, once initial peptides are fully converted, the temperature is raised to accelerate the second step, the conversion of SetCys to Cys. Next, the reaction mixture is cooled down to initial temperature and the next bifunctional peptide is added to perform another NCL reaction. The main advantage of this approach is that no additives must be added between the steps to demask Cys, as soon as all reagents are present in the reaction mixture from the beginning. This includes the control of the pH which is carried out before running the first NCL iteration.

Prior to the process set-up, I first had to obtain peptides with functionalities necessary for mechanistic and kinetic studies.

6.1 Synthesis of SetCys

As it was discussed in chapter 3.2.2 of bibliography part, the synthetic access to SetCys was reported by Diemer et al[96], and later detailed extensively in the protocol article published by Firstova et al[99] (the paper is available as **Annex C**). Briefly, its synthesis includes selenoethyl appendage construction, its attachment to cysteine core and Fmoc-protected cyclic structure formation (**Figure 44**).

For my thesis I reproduced the published[96] Fmoc-SetCys-OH synthesis in 2-2.5 gram scale of starting materials and with reasonable product quantity of 500 mg at the end. This quantity was sufficient for many sets of peptidic synthesis.

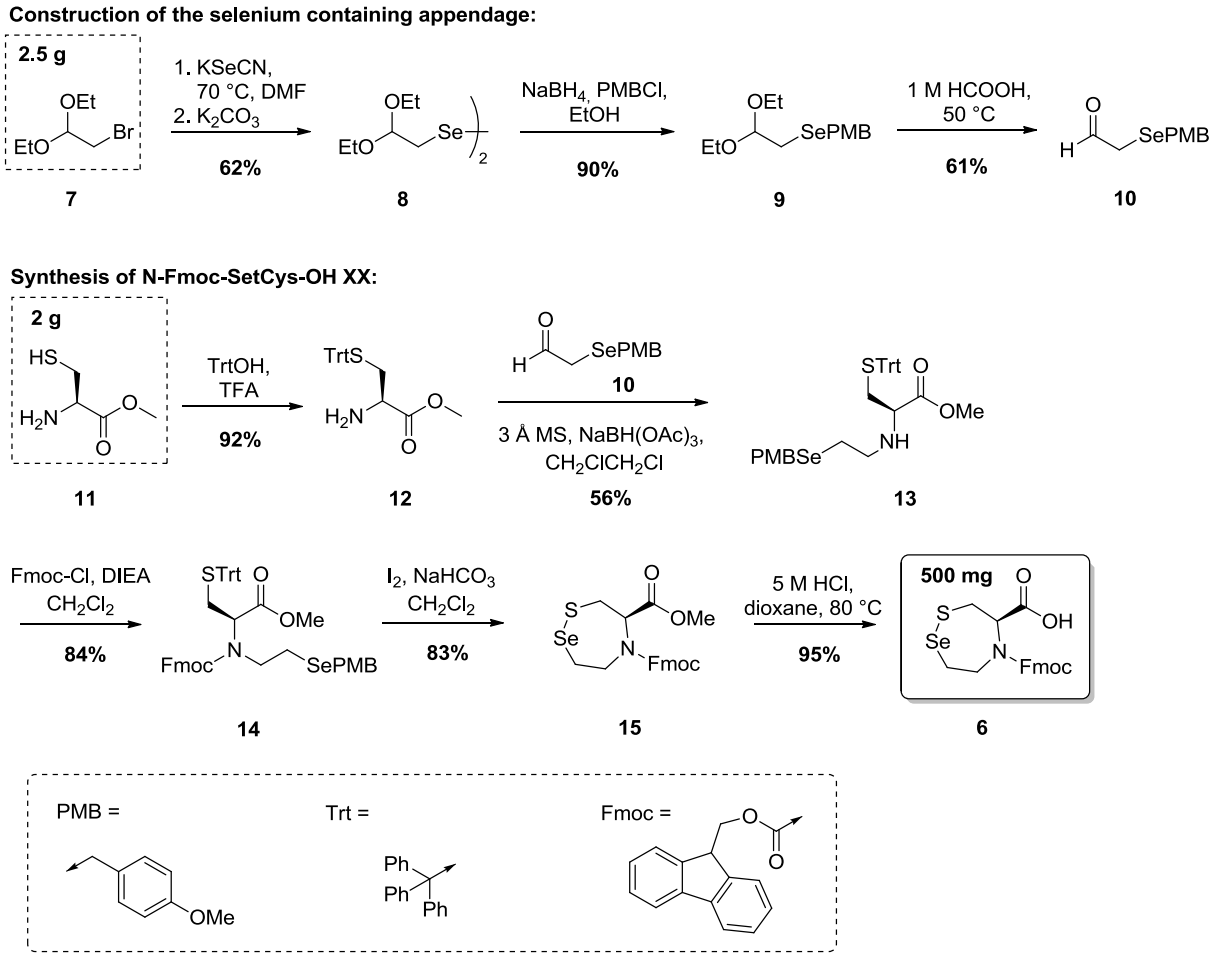


***Figure 44**. Synthesis of Fmoc-SetCys-OH to be used during SPPS. Adapted from Firstova et al.*[99]

The SetCys amino acid was designed to be attached to a peptide by SPPS. Once a peptide sequence is obtained by automated synthesis, SetCys is grafted manually. While a common amino acid coupling takes about 2 minutes, SetCys is incubated for 2 hours in the presence of the peptide and activators (**Figure 45**).

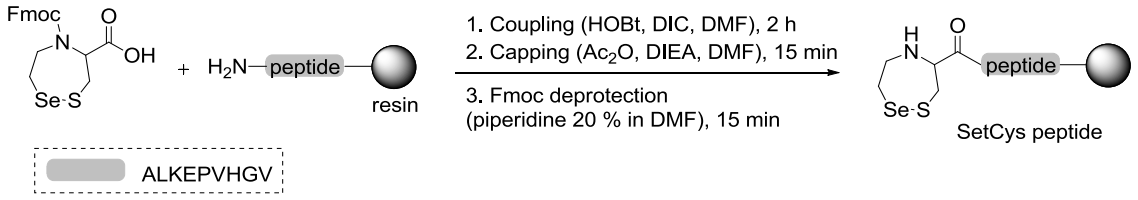


***Figure 45**. The final step of SetCys peptide synthesis on solid support: Fmoc-SetCys-OH attachment to a peptide followed by classical steps of capping and Fmoc deprotection.*

In the next section I will describe the synthesis of peptides that I used to set-up the innovative process.

6.2 Chemical synthesis of reacting peptides

First of all, the reacting peptides have been synthesized on solid support using Fmoc-strategy followed by TFA-induced peptide deprotection and cleavage from the solid support. The *N*-ter cysteinyl peptide is obtained on Rink amide resin (**Figure 46, a**). The bifunctional peptide possessing SetCys on its N-extremity and MPA thioester on C-extremity, was synthesized on SEA resin (**Figure 46, b**), and Fmoc-protected SetCys was grafted to the peptide at the last step of SPPS as it was shown in **Figure 45**. The SEA peptide was deprotected and cleaved from the solid support using TFA with appropriate cleavage cocktail. Then, the thioester functionality was obtained by transthioesterification of the SEA peptide in solution with an excess of mercaptopropionic acid (MPA). MPA thioesters are classically used as acyl donors during NCL in the presence of MPAA used for the catalytic step (discussed in chapter 2.3 of bibliography).

For all experiments, I used short model peptides of 9 or 10 amino acids devoid of internal cysteines in their sequences. The sequences of the used peptides were inspired from the sequence of POL HIV DNA polymerase[103]. Being easily soluble in water, POL derivatives are routinely used in the laboratory for proof-of-concept studies and optimization experiments.

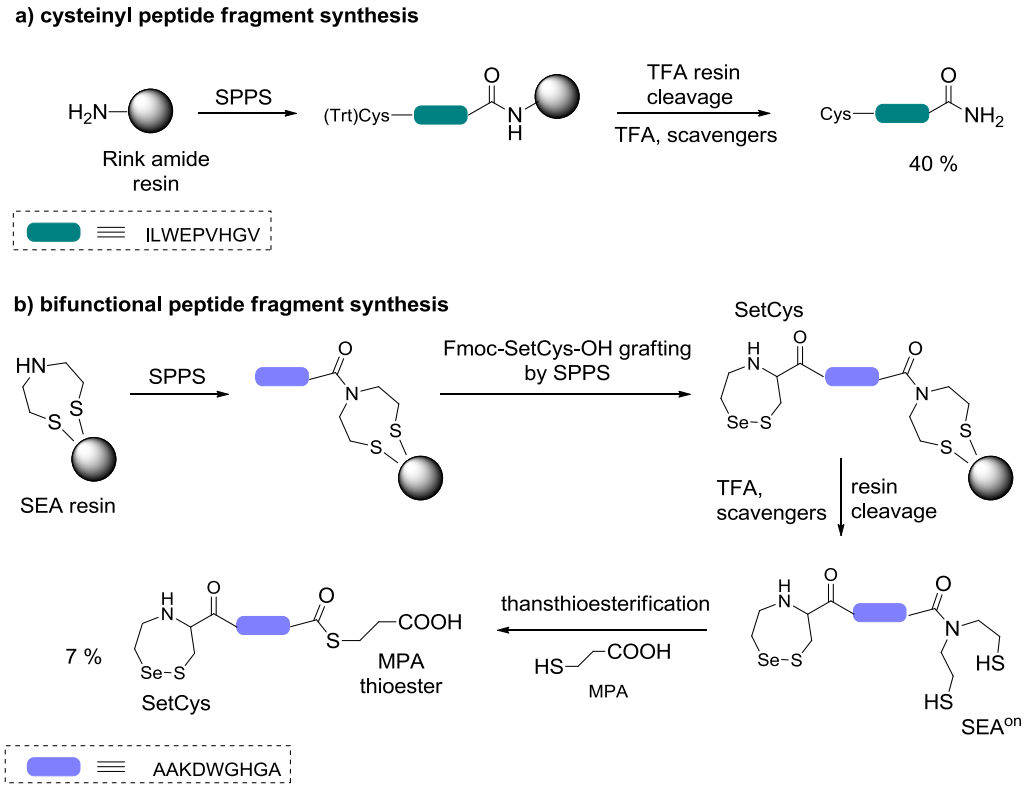


***Figure 46**. Chemical synthesis of model peptides on SPPS. a) Cys-peptide on Rink amide resin; b) bifunctional peptide with SetCys appended by SPPS obtained on SEA resin after resin cleavage and transthioesterification steps.*

Once the peptides were obtained and purified, I could start exploring SetCys chemistry in the ligation media.

6.3 Study of SetCys-to-Cys conversion

The first attempt to assemble the NCL and SetCys chemistries was the ligation between the bifunctional peptide and the Cys-peptide in the presence of MPAA and TCEP, as both are classical reagents for NCL under common ligation conditions: pH 7.2, 37 °C (**Figure 47**). The pH was controlled by phosphate buffer, guanidinium chloride (Gdn·HCl) presented as denaturant and solubilizing agent, sodium ascorbate was added to avoid the deselenization by TCEP. As all following experiments, the reaction was carried out in the glovebox and was monitored by HPLC.

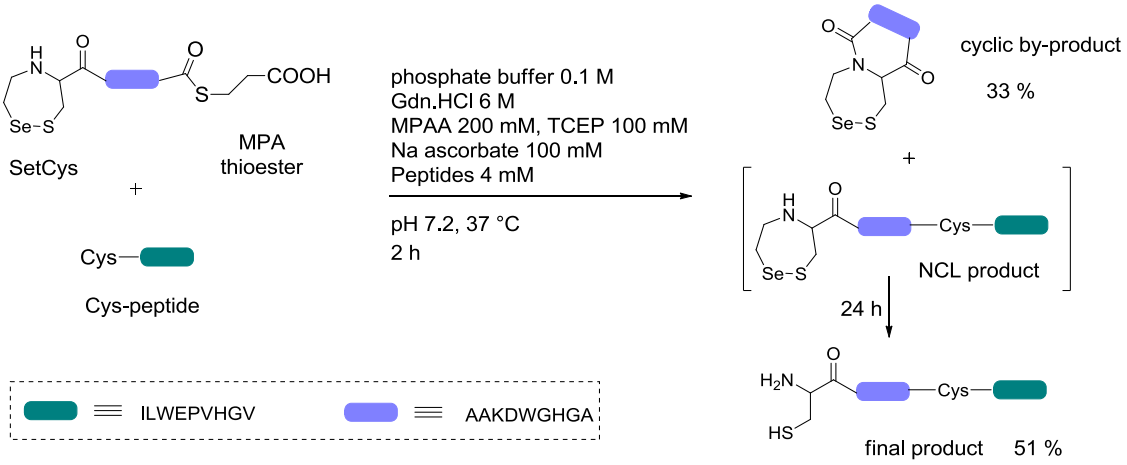


***Figure 47**. NCL performed between the bifunctional SetCys peptide and Cys-peptide in the presence of strong reductant TCEP, which induced the cyclic by-product formation due to SetCys premature activation.*

However, as it was expected, the process was not selective. Due to Se-S bridge fast opening the SetCys amide was prematurely activated and participated in the formation of a cyclic by-product. Interestingly, the loss of selenoethyl arm by the cyclic product was not observed. The target NCL product was formed within 1.5 hours, and then was losing Se-ethyl limb during 8 days. The last reaction corresponds to demasking of N-ter cysteine for another potential ligation.

As the peptides were reacting in strong reducing media, we could expect to detect peptides with reduced SetCys during mass analysis. However, all molecular weights corresponded to SetCys oxidized form, surely due to samples exposure to air oxygen during MPAA extraction with ether and transferring to HPLC vial.

This experiment confirmed the theoretical data and the incompatibility of TCEP use for my project. Thus, I had to select a phosphine with lower reducing power.

6.3.1 Studies on SetCys reduction by thiols and phosphines

In the bibliography chapter 3.2.2 the cysteine surrogate SetCys was presented, where I described its main properties applied to peptides ligation. During the set-up of the innovative two-step process it was important to deeper study the reactivity of SetCys in the presence of thiols and phosphines needed for ligation.

This part of the work is presented as a summary of the publication "**Thiol catalysis of selenosulfide bond cleavage by a triarylphosphine**", where I am the first author[104]. The full version of the paper is presented as **Annex B**.

Chalcogens are of high importance in organic synthesis and peptide chemistry. The reduction of dichalcogenides by phosphines is being the subject of several studies. For example, alkyl and aryl disulfides can be reduced by phosphines such as *tris*(2-carboxyethyl)phosphine hydrochloride (TCEP) or 3,3',3''-phosphanetriyltris(benzenesulfonic acid) trisodium salt (TPPTS), a mild reductant (**Figure 48, b**). Meantime, alkylselenosulfides can be effectively reduced only by an alkylphosphine with strong reducing potential such as TCEP, taking into account the limitation of its use due to causing undesired deselenization. In all these reduction methods, the participation of water has been clearly established (**Figure 48, a**).

In this work, our interest was focused on redox properties of SetCys, developed in our team as Cys surrogate. SetCys is a modified cysteine with a cyclic structure containing Se-S bridge. We had to find optimal conditions for SetCys conversion into Cys by losing its selenoethyl appendage (**Figure 48, c**) while avoiding deselinization.

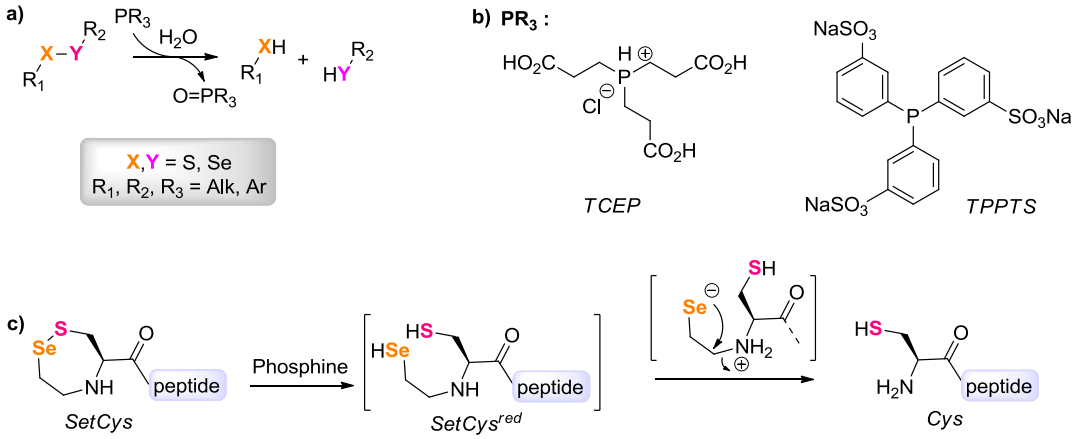


***Figure 48**. a) Reduction of mixed dichalcogenides with phosphines in the presence of water; b) TCEP and TPPTS have been used in related studies and in this work; c) The reaction of our interest, SetCys to Cys conversion.*

We could establish, that 4-mercaptophenylacetic acid (MPAA), a thiophenol derivative used for catalyzing NCL, is a powerful catalyst of SetCys selenosulfide bond

reduction by the water-soluble triarylphosphine TPPTS. Being a milder reductant compared to TCEP, TPPTS has the advantage of not inducing deselenization reaction.

Apart from MPAA, methyl thioglycolate (MTG) and sodium 2-mercaptopropylsulfonate (MPSNa) were also tested to accelerate SetCys reduction, but these thiol additives demonstrated significantly weaker catalytic effect than MPAA. Experimentally, MPAA enables to accelerate the conversion more than 3-fold (t½ 75.4 h vs 22.3 h) at concentration of 2 mM only. By applying higher concentrations of MPAA, we could observe more rapid SetCys to Cys transformation. We could conclude that MPAA attack on the SetCys Se-S bond (**Figure 49, path b**) is rate limiting and depends on MPAA concentration when the TPPTS is at 100 mM. Note that TCEP remains efficient in the direct Se-S bond reduction even in the absence of MPAA (**Figure 49, path a**).

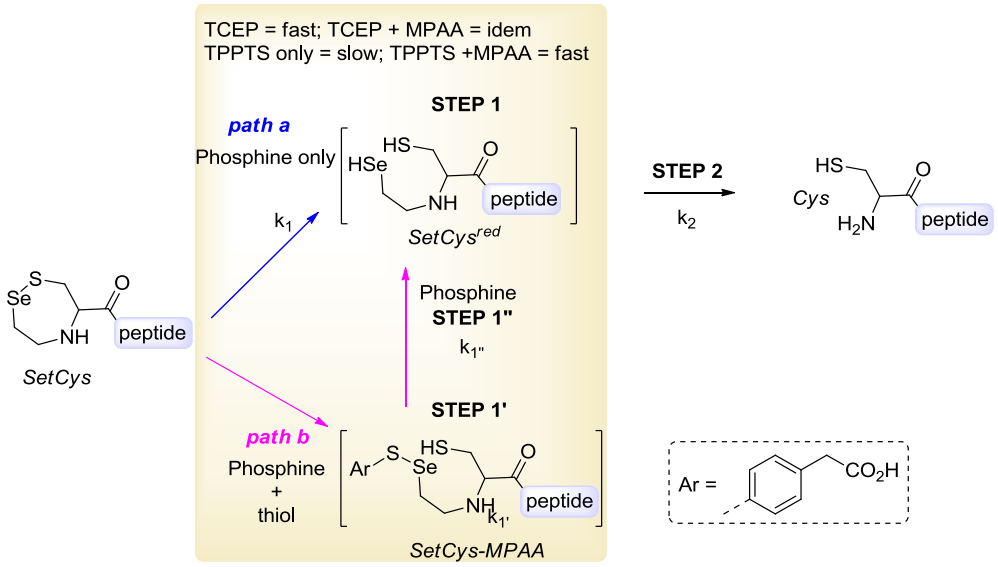


***Figure 49**. Proposed mechanisms of SetCys to Cys conversion. Path a: in the presence of a phosphine only the Se-S bridge is reduced with spontaneous Se-ethyl loss. b) Se-S reduction is promoted by MPAA-thiolate attacking Se, when a modest phosphine reductant such as TPPTS is used.*

To mention, apart from TPPTS, two other aromatic phosphines were interesting candidates (**Figure 50**). They are bis(3-sulfonatophenyl) (3,5-di-trifluoromethylphenyl) phosphine (TTFMPP) and tris(4,6-dimethyl-3-sulfanatophenyl)phosphine (TMPP). However, both phosphines were abandoned as they have low solubility in aqueous media and were precipitating when Gdn·HCl concentration was ≥ 2 M, while I kept Gdn·HCl 6 M for all experiments.

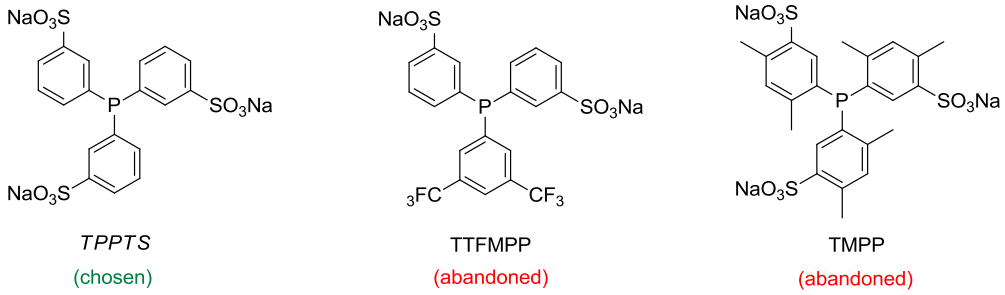


***Figure 50**. Aromatic phosphines with moderate reducing power that have been tested. TPPTS is chosen, TTFMPP and TMPP are abandoned.*

From the obtained results we could understand how to control kinetics of SetCys to Cys conversion *via* reduction of dichalcogenides. Practically, SetCys is the masked cysteine that can be liberated in kinetically controlled manner. This knowledge was applied to development of novel method of peptides concatenation, where at least two ligations are performed in one-pot. Therefore, peptides ligation must be managed in the presence of TPPTS with no thiol additives, to keep SetCys intact. This brings two conclusions.

First, as SetCys cannot be activated by TPPTS alone, the external activation is required. As it was proposed in main goals part, SetCys can be efficiently activated by heat because it requires a high activation energy. Heat enables also to accelerate the reduction of the SetCys selenosulfide bond, a step that is mandatory for observing the formation of Cys residue from SetCys.

Second, a thiol cannot be used as catalyst for NCL. This is why, relying on theoretical data described in chapter 2.3 of bibliography, I plan to use imidazole as perspective non-thiol nucleophilic catalyst.

These two ideas will be developed in following chapters.

6.3.2 Temperature effect on SetCys-to-Cys conversion

The classical temperature for performing NCL is 37 °C, referred to physiological conditions. However, a higher temperature is needed to accelerate SetCys to Cys conversion, and the goal was to investigate the range of effective heat.

First of all, a monofunctional SetCys model peptide was synthesized by SPPS on Rink amide resin to study SetCys response to heat and other parameters (**Figure 51**).

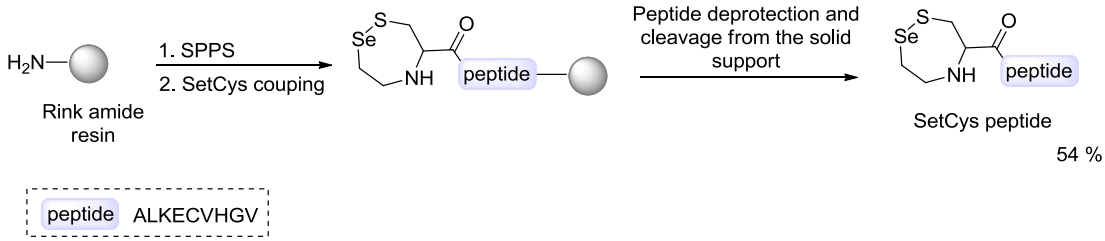
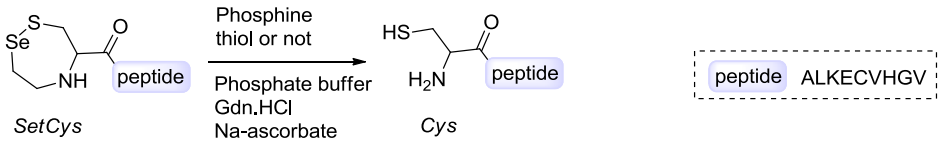


***Figure 51**. Synthesis of monofunctional SetCys peptide on solid support.*

For all kinetic studies of the heat effect, the media contained Gdn·HCl and sodium ascorbate as for reference essays with phosphines. Already having in hand the kinetics of TCEP-induced SetCys conversion, it was useful to use the same conditions (TCEP 100 mM, MPAA 200 mM, pH 6.0) and to vary only the heat (**Figure 52, a**; presented in the paper – **Annex B**). The heat was progressively increased with step of 10 °C for experiments between 30 °C and 60 °C, including also the reference of 37 °C.



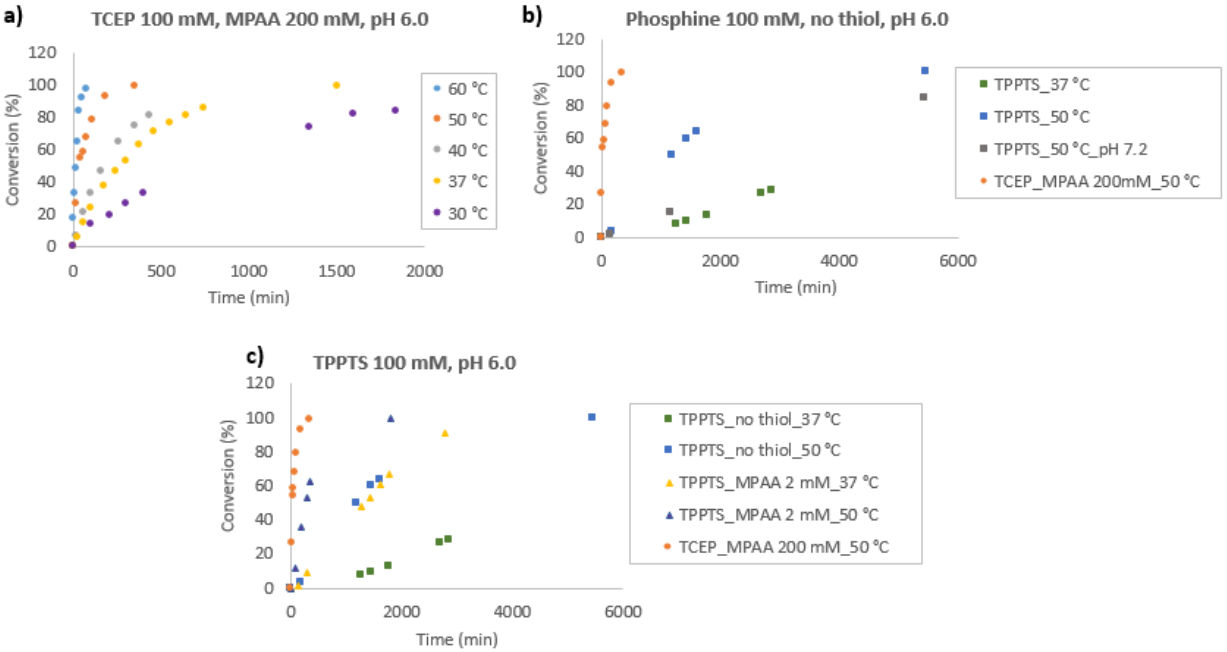


***Figure 52**. a) TCEP-induced conversion in the presence of MPAA at pH 6.0 under variated heat (from **Annex B**); b) TPPTS-induced conversion with no thiol added, at 37 and 50 °C in the useful pH range (TCEP reference included as orange dots); c) TPPTS-induced conversion at pH 6, at selected temperatures and in the presence of a thiol (TCEP reference - orange dots, TPPTS with no thiol – green and blue squares).*

We discovered that adding the heat of 10 degrees, the rate of selenoethyl arm cleavage increases about 3.5 times. For example, after heating at 30 °C for over 33 hours, the conversion of only 80 % was reached. The reference heat at 37 °C lead to full conversion in 25 hours.

Despite the highest rate was reached while heating at 60 °C, this temperature might be damaging for peptidic molecules during long exposure. We therefore chose to activate SetCys conversion at 50 °C as more convenient.

The next step was to incubate SetCys-peptide with milder phosphine, TPPTS, taking into account the condition of the absence of added thiols for the desired two-step innovative process (**Figure 52, b**). The incubation was carried out at reference 37 °C and selected 50 °C. At 37 °C the conversion was very slow and reached only 40 % in 4 days, which is promising result as SetCys must remain unreactive under NCL conditions. NCL is often carried out at pH 7.2, so the heat at 50 °C was performed for pH 7.2 and 6.0, to anticipate the influence of pH on conversion rate. The combination of pH 7.2 and 50 °C resulted the kinetics almost as slow as at pH 6 and 37 °C, what confirms pH dependence of the selenoethyl limb cleavage rate, i.e. increasing pH slowers the breakdown of selenoethyl arm. The combination of pH 6.0 and 50 °C resulted more rapid conversion with the half-life time 20 h and total time of 91 hours, meaning that applying heat is the effective way to promote Cys demasking when this process is induced by a mild reductant such as TPPTS.

It is interesting to compare the kinetic curves for different phosphines, TCEP and TPPTS. With TCEP the beginning of kinetic curve is stip and straight, going through a plateau by the end. In contrast, the presence of TPPTS results the conversion delay for 2-3 hours, so the curve is smooth on the corresponding area. Then the curve becomes stiper and ends on a plateau.

The curve beginning was more straight when MPAA 2 mM were added to TPPTS at pH 6.0, heating at 37 °C and 50 °C (**Figure 52, c**). We observed that in the presence of MPAA 2 mM the kinetic curve remains slightly delayed, but the half-life time and total conversion time were reduced. For example, the presence of MPAA 2 mM at 37 °C recovers kinetics for TPPTS at 50 °C in the absence of the thiol.

The obtained results demonstrated that SetCys surrogate can be transformed into Cys by heat in the presence of TPPTS and in the absence of thiol. SetCys conversion into Cys under such conditions proceeds faster at pH 6. During the NCL step, the temperature must be no higher than 37 °C, and raised up to 50 °C to induce the cysteine demasking by selenoethyl arm cleavage. Applying this to two-step elongation process, pH still must be compromised between SetCys optimal conditions for demasking (pH 6) and optimal pH for thioester reactivity in NCL, which is around 7. In ideal, the NCL step completion should take no more than 3 hours while SetCys remains intact. In the next sub-chapter, I will describe the optimization of NCL step where thiol catalyst is replaced with azole one, and how the new conditions can be combined with redox properties of SetCys.

6.4 Optimization of the NCL

6.4.1 NCL catalysis by imidazole

As a reminder, some azole derivatives such as imidazole are good nucleophilic agents toward thioesters and lead to the in situ formation of activated acyl azole intermediates (**Figure 53**). A large excess of the catalyst is required to accelerate the process.

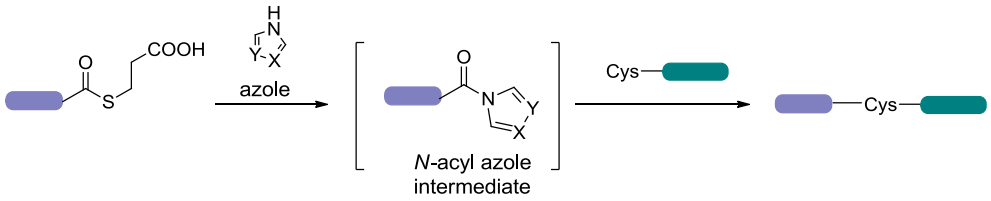


***Figure 53**. General scheme of azole-catalyzed NCL.*

Imidazole-catalyzed NCL (Im-NCL) was demonstrated by Sakamoto et al.[79] NCL kinetics was increased in the presence of large imidazole excess (**Figure 54**). In addition, imidazole plays the role of the buffer for the pH range 6.2 – 7.8.[105]

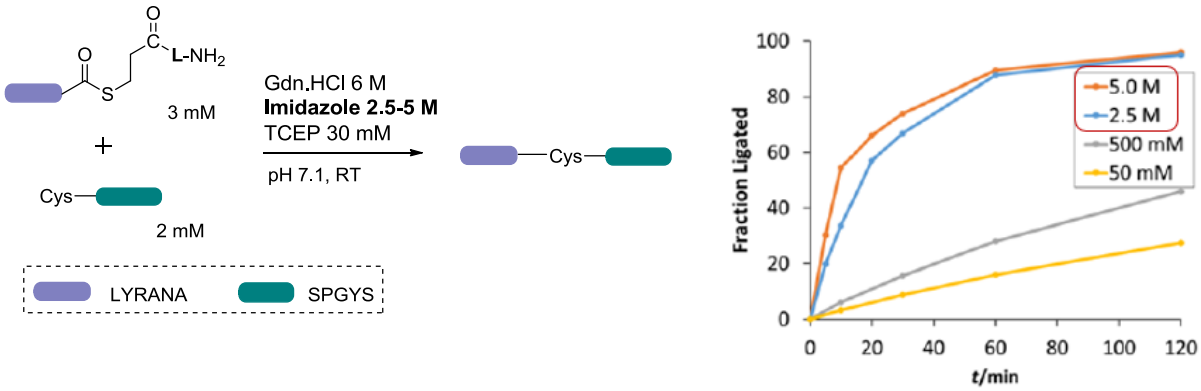


***Figure 54**. Imidazole catalyzed ligation with effective catalyst range 2.5-5 M. Kinetic data and figure on the left are taken from Sakamoto's publication[79].*

The authors demonstrated that imidazole was efficient in the concentration range between 2.5 and 5 M. Note that Gdn·HCl and imidazole are both taken in large excess, so they play the role of co-solvents. We noticed that to avoid any precipitations by working with 5 M of imidazole, the Gdn·HCl concentration had to be decreased to 3 M.

Relying on these data, I performed the ligation between bifunctional and cysteinyl peptides in the adapted conditions (Gdn·HCl 6 M, TPPTS 10 mM, peptides 3 mM, pH 7.4, 20 °C) and variated imidazole concentration (**Figure 55**).

It is then important to explain those adjustments. The Gdn·HCl concentration remains as high as 6 M to avoid components precipitation. Sodium ascorbate was not added as TPPTS does not induce deselenization, unlikely to TCEP (see the paper in **Annex B**). The concentration of TPPTS was decreased to 10 mM to decrease the risk of precipitation of the phosphine. As imidazole is known to catalyze thioesters hydrolysis, I switched from classical 37 °C towards lower temperature, 20 °C, in order to limit MPA thioester hydrolysis. As this would decrease the Im-NCL rate as well, the pH was slightly increased up to 7.4.

To select optimal imidazole concentration, I ran the ligation with 2.5 M, 4.2 M and 5 M of the catalyst. The lowest rate was at 2.5 M of imidazole, 5 M addition led to precipitation during the ligation despite the highest ligation rate, and 4.2 M was optimal as the rate was nearly as high as for 5 M but with no precipitations. In all cases, the kinetic curve shape corresponded to the one of reference publication. However, the yield of the ligation product didn't reach 100 % in this work due to formation of side products that will be discussed below in details. In the reference paper the starting peptides were consumed in 2 hours after the ligation start, while in this work the conversion took between 3 and 5 hours.

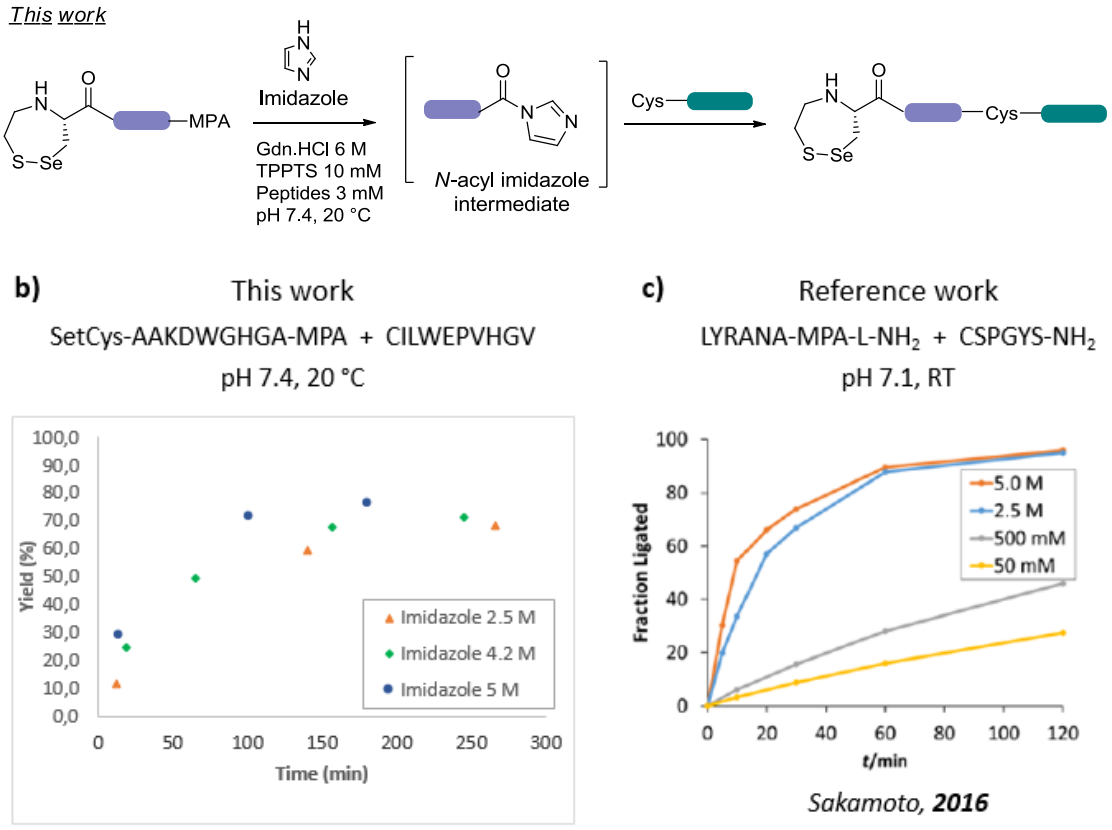


***Figure 55**. a) Im-NCL between bifunctional and cysteinyl peptides in adjusted conditions; b) The kinetics for this work; c) the kinetics for the reference paper[79].*

Comparing the reference and actual works, differences can be found in peptides sequences and nature of N-ter thioesters. The indicated room temperature from the reference paper could be variated, as we don't know the termostating conditions. In this work, to

understand the limitations of the reaction, we decided to explore the conditions by changing the additives, pH and temperature prior to modifying the reacting peptides nature.

The starting point for further optimizations was the use of imidazole concentration of 4.2 M as the most optimal. Imidazole 4.2 M and guanidinium chloride 6 M are considered as co-solvents, and water takes only 15 % of solution by weight. This low percentage of water might also less promote the thioester hydrolysis.

The full view on the ligation and all formed products is presented in **Figure 56**. Apart from NCL product (71 %), thioester hydrolysis (13 %), bifunctional peptide cyclization (5 %) and oligomerization (2 %) products were detected.

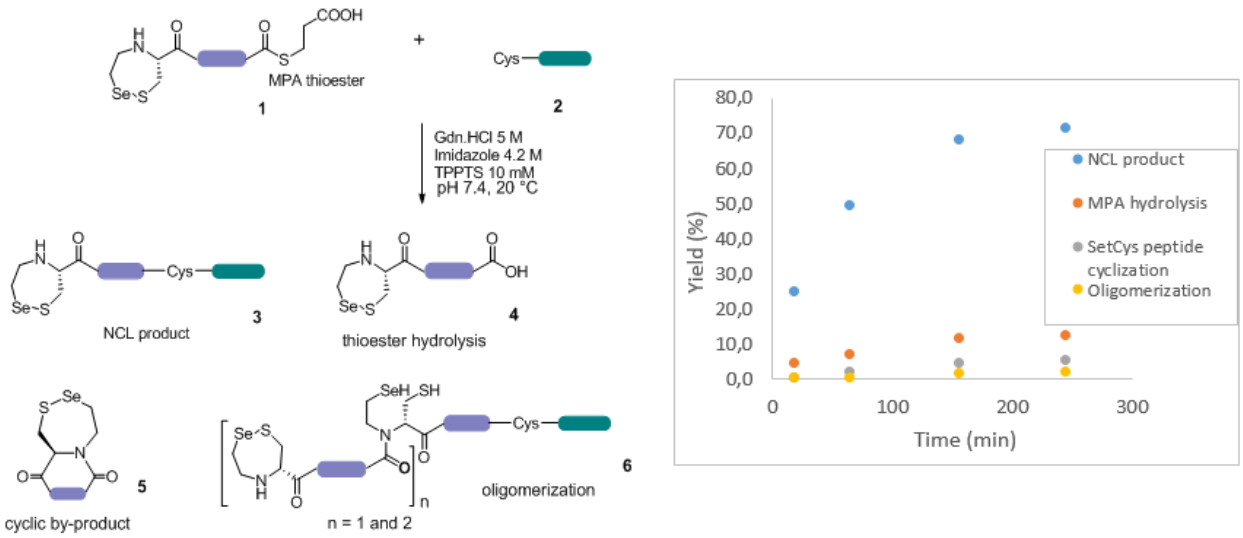


***Figure 56**. Im-NCL between bifunctional and Cys-peptides with formation of target NCL product alongside with thioester hydrolysis, cyclization and oligomerization by-products.*

Unfortunately, the use of imidazole as NCL catalyst promoted peptide thioester hydrolysis at the rate more than expected. Hydrolysis of the thioester segment leads to the accumulation of non-reacted Cys-component in the mixture which will complicate the process during following NCL iterations. I varied several parameters to minimize its formation.

6.4.2 Temperature adjustment for NCL

After several trials we found that decreasing the temperature decreases the extent of hydrolysis. We found that working at 10 °C, while keeping the pH at 7.4, was a good compromise for minimize hydrolysis while keeping the rate of NCL at a practical level (**Figure 57**).

With optimized temperature, the target product yield slightly decreased down to 64 %, but the conversion took 25 hours instead of 5 hours when heated at 20 °C. The thioester hydrolysis product forms at 10 °C in much lower quantities (2-4 %), and this is what we expected. However, the cyclization of bifunctional peptide significantly increased and reached

18 %. As the NCL rate decreased, we suggested that the bifunctional peptide had more time for cyclization. Following the reaction, we could still detect the polymerization products with 1 or 2 additional SetCys peptides grafted in chain on opened SetCys of the desired ligation product.

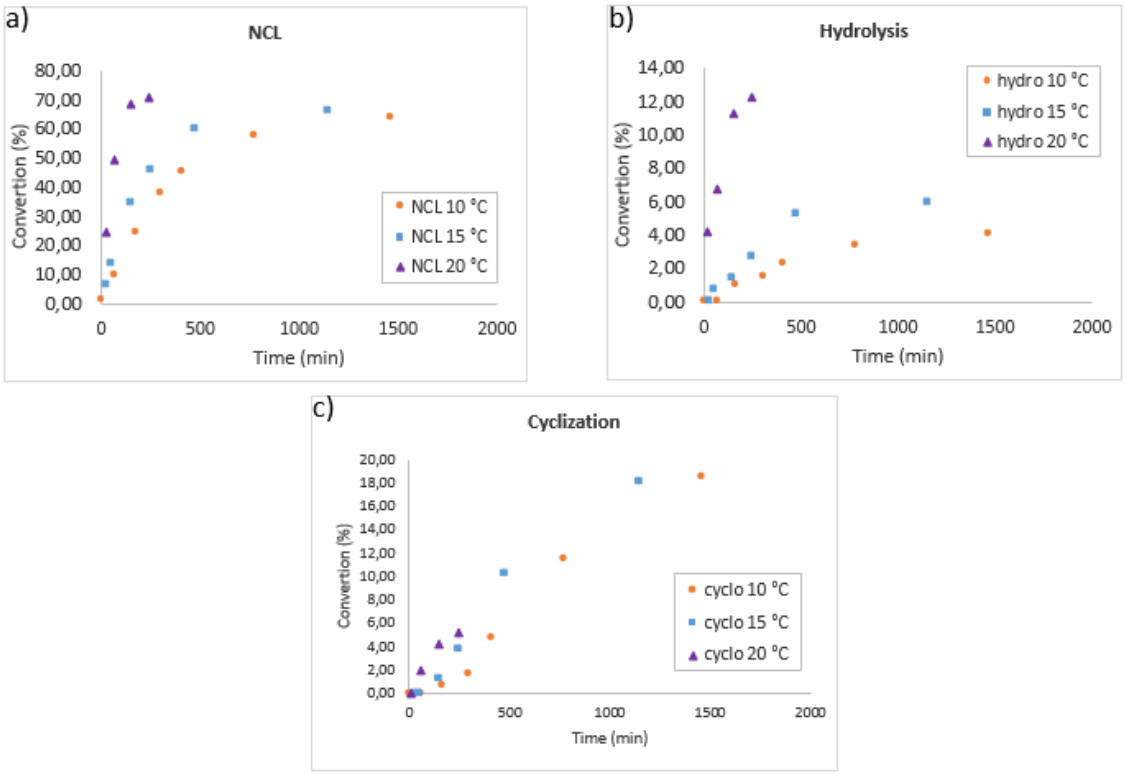


***Figure 57**. Kinetic curves for temperature adjustment experiments. a) Im-NCL; b) MPA hydrolysis; c) bifunctional peptide cyclization.*

On the presented HPLC monitoring of the process we can clearly see that slow kinetics of NCL is critical as by-products actively form in 7 hours after the starting point, and cumulate for 24 hours in total (**Figure 58**).

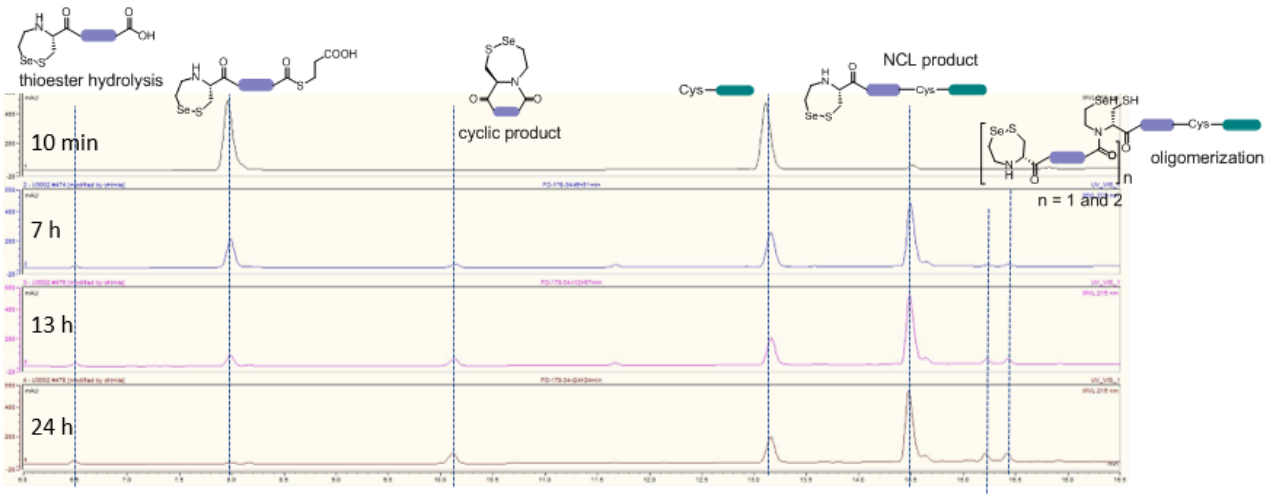


***Figure 58**. HPLC monitoring of Im-NCL with excessive formation of by-products due to low ligation rate.*

6.4.3 Peptides concentration adjustment

One of the easiest ways to accelerate the ligation can be the increase of reacting peptides concentration. I have used the concentration of 3 mM for both reacting peptides, and I varied the peptides concentration to see its impact on the extent of by-product formation.

We found that the increase of peptides concentration to 5 mM leads to not only faster NCL kinetics, but also favors the oligomerization of the bifunctional peptide (takes 2-6 % of yield). A lower concentration of 1-2 mM favours side-reactions such as bifunctional peptide cyclization and MPA thioester hydrolysis.

The next question we had is whether the rate-limiting step is the formation of transient acyl-imidazole or the capture step with Cys-peptide.

6.4.4 Imidazole catalyst activation by addition of a base

To compensate the NCL rate, we tried to co-catalyse the thioester activation with a base (DABCO, for example[106]) by increasing its nucleophilicity (**Figure 59**). However, this idea did not work since the rate of undesired cyclization was increased. Thus, we had to accelerate the target reaction sufficiently by other means.

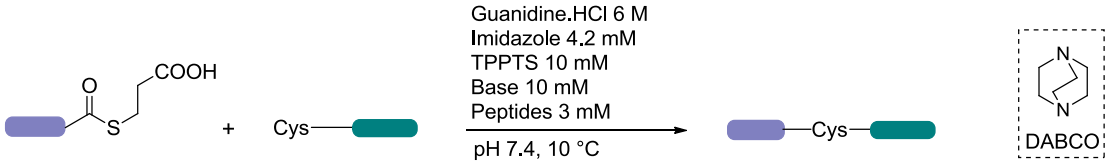


***Figure 59**. Attempt to increase the ligation rate by adding a base (DABCO) as co-catalyst.*

6.4.5 Improving acyl acceptor capability of peptide nucleophile during NCL capture step – replacing Cys with Sec

I tried to increase the ligation rate from the side of Cys-peptide, the acyl acceptor. Cys-peptide was replaced by Sec-peptide[107,108], to favor intermolecular thioester capture over cyclization. Due to lower pK_a of Sec, this idea could also allow working at lower pH that would be more beneficial during the step of SetCys to Cys conversion. However, this option was abandoned due to the slow Sec-dimer reduction into reactive monomers, causing additional by-products formation (**Figure 60**).

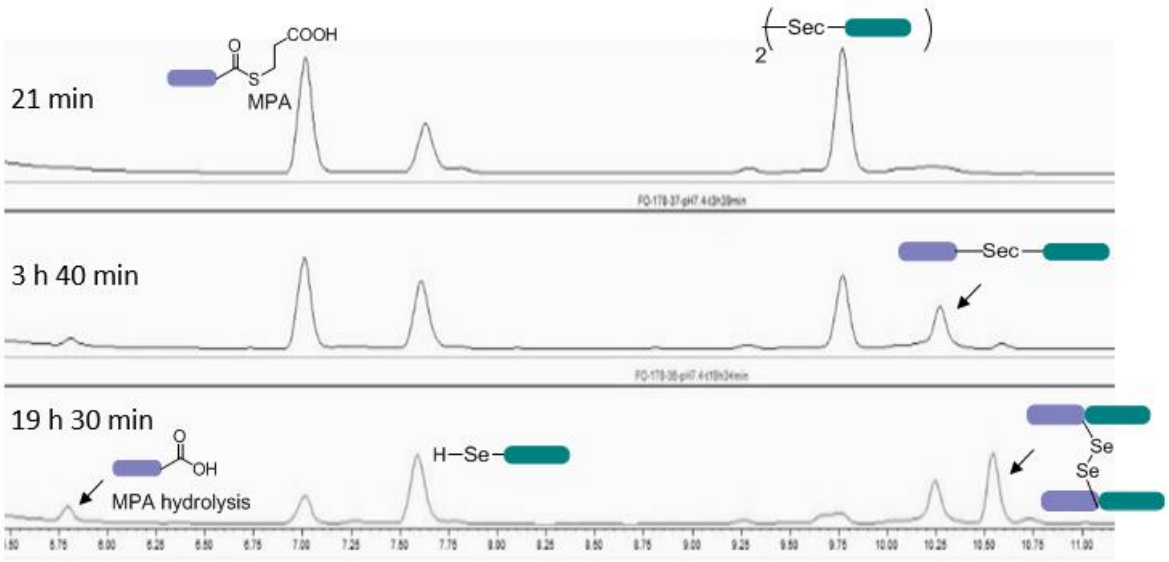


***Figure 60**. HPLC following of Im-NCL with Sec-peptide dimer instead of Cys-peptide as the second reacting partner.*

6.4.6 Thioester reactivity adjustment

The last, but not least modification was varying thioester reactivity. It means the replacement of MPA thiol leaving group with another one more active, methyl thioglycolate (MTG) with lower pK_a, what makes MTG potentially better leaving group (see the publication in **Annex A**).

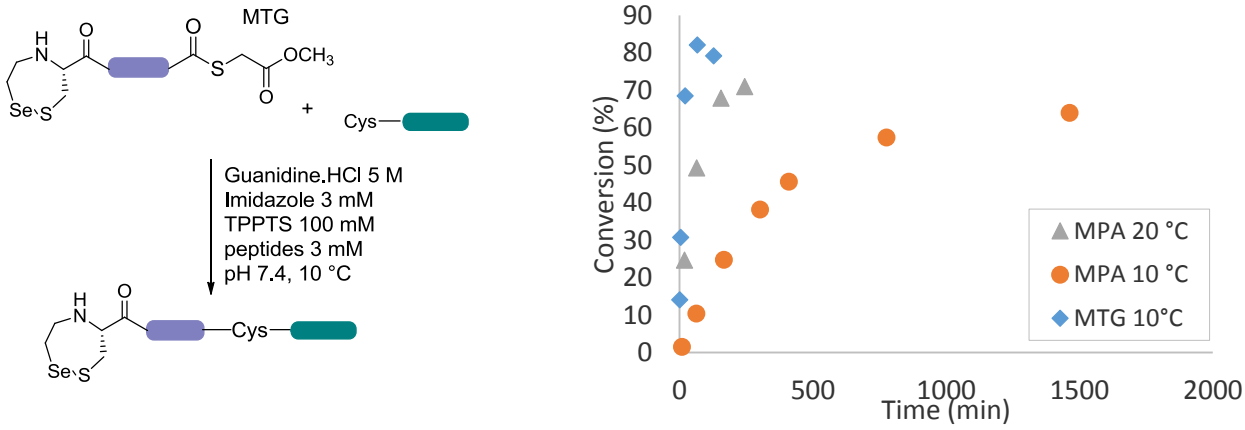


***Figure 61**. Im-NCL with more active MTG thioester and comparison of kinetic curves MPA vs MTG.*

As I plan to use the innovative process for multiple ligations in one pot, TPPTS must not be limiting too early. This is why it was decided to use 100 mM concentration of the phosphine. Intermediate trials demonstrated that the increase of TPPTS concentration does not affect the NCL step. Next, to solubilize the increased quantity of the phosphine, I adapted the concentrations of co-solvents: 5 M of guanidinium chloride and 3 M of imidazole. Despite the

decreased concentration of the catalyst, the NCL rate and yield were surprisingly high (**Figure 61**). At 10 °C, MTG thioester increased NCL rate 10-times, what allowed to decrease the formation of side products. Being indeed a better leaving group than MPA, MTG is faster activated by imidazole. As a result, the NCL speed increased nearly 10 times, and the reaction takes 2,5 h instead of 20 h with the yield of NCL product of 85 %. To compare, this yield is even higher than for MPA thioester at 20 °C.

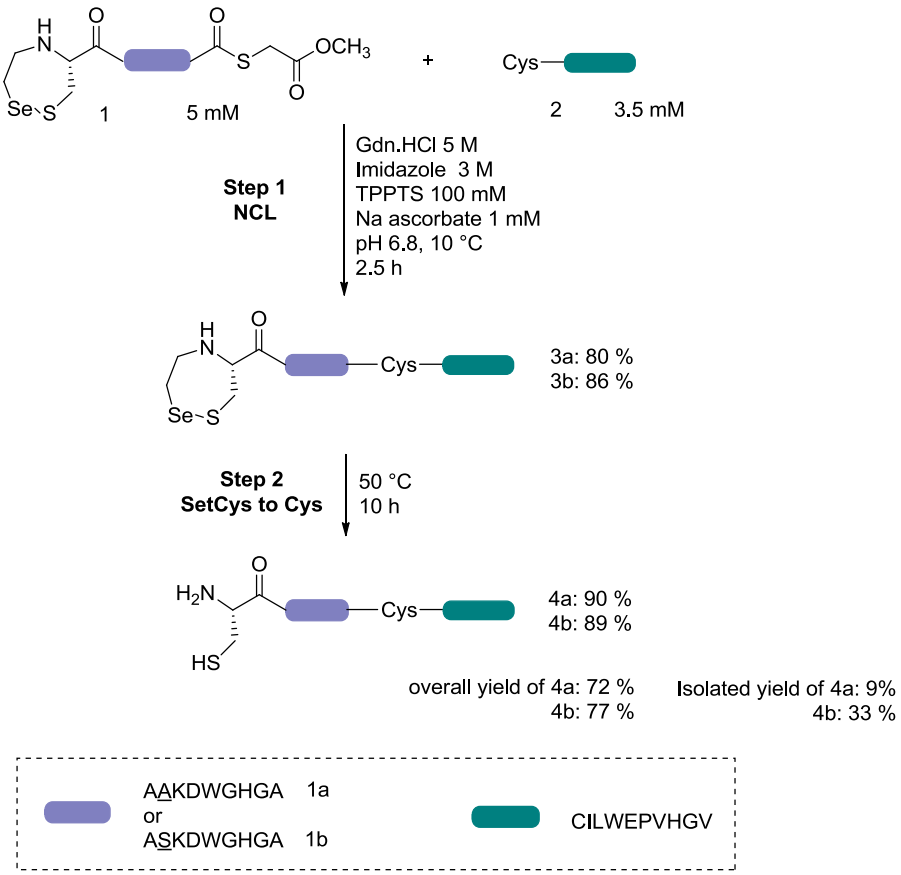


***Figure 62**. Two-step innovative process controlled by temperature switch; analytical and isolated yields are presented for the final products.*

Nevertheless, the process can be optimized even more. Earlier we have studied the NCL step using Cys-peptide and SetCys-bifunctional peptides in 1:1 ratio. Since we cannot avoid the partial unmasking of SetCys during NCL due to liberation of 3 mM of MTG thiolate, and hence bifunctional peptide loss in by-product formation, we decided to use a slight excess of the bifunctional peptide to compensate for this loss. It is important that the Cys-peptide is consumed entirely during NCL to avoid its participation in the next NCL iterations. In contrast, the accumulation of cyclic by-product is less problematic for being poorly reactive during NCL. Another adjustment was done for the pH that had to be compromised between SetCys and

thioester reactivities. The pH value of 6.8 was chosen. Prior to the two-step process performance under given conditions, the preliminary SetCys to Cys conversion was carried out with the monofunctional SetCys peptide (the same used for phosphines reactivity studies) which is described in experimental part (**Annex C**).

Under final selected conditions, I performed the two-step process for two similar bifunctional model peptides reacting with the Cys-peptide (**Figure 62**). The peptide **1a** was used for all the optimization experiments, while the peptide **1b** was synthesized to confirm the reactivity using another sequence. The difference is in one alanine (**1a**) replaced with serine (**1b**). Analytical yields for the first step are 80 % and 86 % for ligation products **3a** and **3b** respectively. Analytical yields for the second step are 90 % and 89 % for the SetCys conversion products **4a** and **4b** respectively. The overall analytical yields of the products **4a** and **4b** are 72 % and 77 % respectively, isolated yields are 9 % (purification issue) and 33%.

The HPLC profiles for both reactions are presented in **Figure 63**.

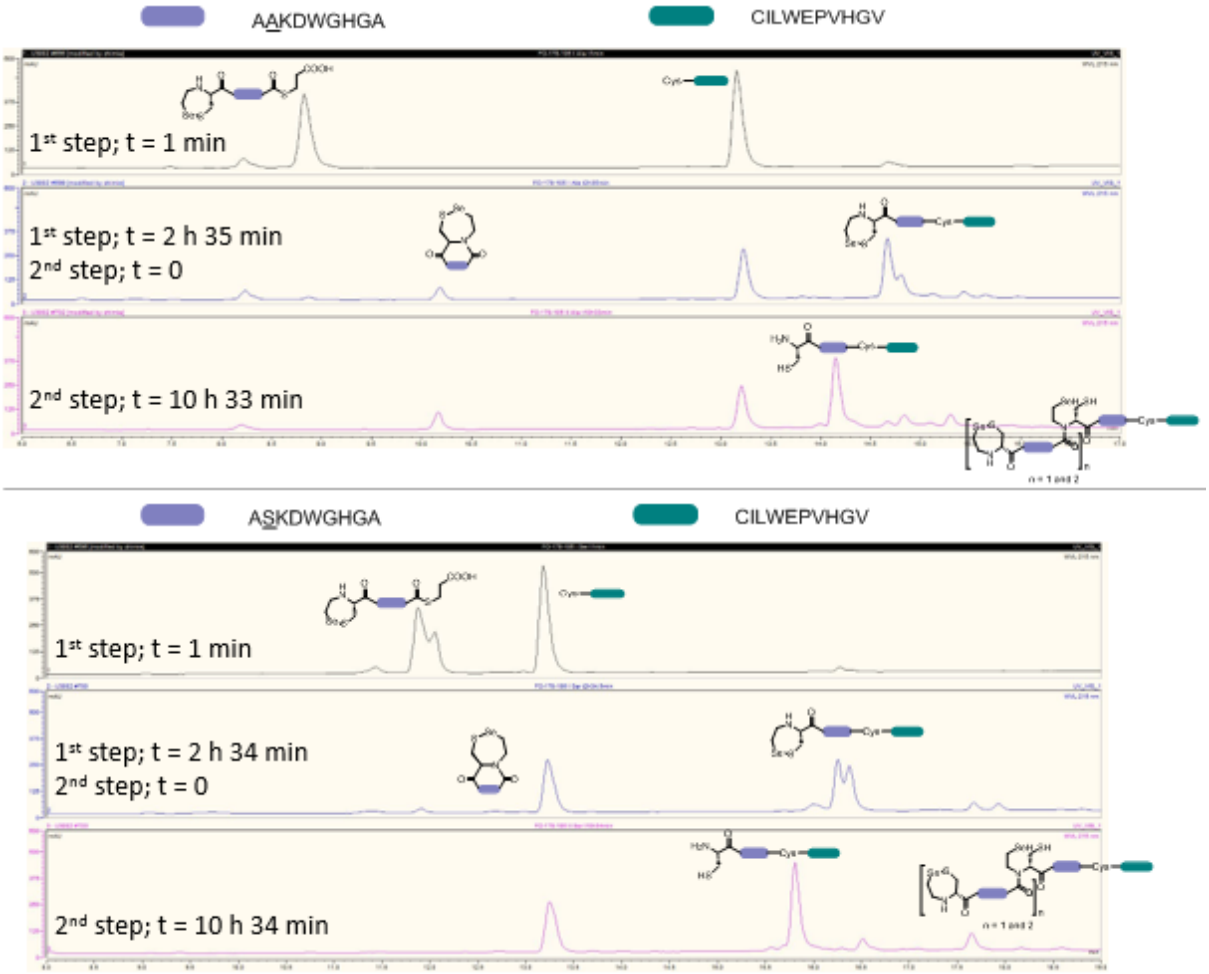


***Figure 63**. HPLC monitoring of the two-step process for two similar peptide sequences*

For both batches, the Im-NCL step took 2.5 hours at 10 °C, then I switched the temperature to 50 °C to induce the SetCys to Cys conversion step that took 10.5 hours. The HPLC profile was cleaner for the reaction with serine-modified bifunctional peptide. Both batches were prepared at the same time and the same solution was used to solubilize the peptides. Perhaps, the Ala to Ser replacement could affect the reactivity of the bifunctional peptide. It is also important to note that the HPLC profile of the Ser-modified peptide appears as double-peak, as well as the SetCys containing NCL-product. This effect depends on exposure time for an aliquoted sample prior to injection into chromatograph column. The longer the exposure, the bigger is reduced SetCys peak. Oxidized and reduced forms of SetCys appear in the presence of thiols (Cys-peptide and MTG thiolate) while only the oxidized form is detected in the absence of thiols.

6.4.7 Other optimization trials

Clearly, the novel two-step process was significantly optimized as all the conditions have been changed and compromised. At its current state, the process is selective, but the discussed above by-products still cumulate.

Trying to accelerate the NCL step, I used 2-methyl imidazole as nucleophilic catalyst, but no improvements were achieved in term of yields.

Next, sodium ascorbate at 1 mM can be added to the reaction mixture. As this is a powerful antioxidant, its presence will be beneficial for further iterations. Experimentally I could check that the presence of Na ascorbate at low concentration of 1 mM did not change the reaction profile and did not increase the thioester hydrolysis.

Once the reaction time is explored and well known, an automation of the process can be applied. For example, we were curious to run two steps process in PCR machine for automation of temperature switch and excellent temperature control. However, the reaction profile was messy due to exposure to air oxygen, as I had to open the Eppendorf lead to mix thermostated peptides solutions.

7 Conclusion and perspectives

7.1 Innovative SetCys-based strategy development

SetCys has been applied as a tool to access long peptides by one-pot sequential ligations. The innovative proof-of-concept method for polypeptide elongation demanded in depth study of SetCys reactivity under conditions appropriate for NCL, as well as significant optimization of the NCL step. The main challenge of the development was to compromise the elaborated conditions at every stage of this work, to put together NCL and SetCys chemistries.

I managed to find sufficiently mild reducing conditions for SetCys to Cys conversion with delay, enabling selective NCL step. At the same time, I optimized the NCL step by changing thioester reactivity and catalyst, avoiding excessive formation of by-products.

Further improvements could be applied to the step of NCL. First, another imidazole derivative can be investigated as nucleophilic catalyst. Next, it would be interesting to work with longer model peptides, in order to potentially decrease the rate of cyclization of bifunctional peptide due to longer distance between SetCys and thioester functionalities. Finally, the second two-step iteration should be performed by adding the next bifunctional peptide. In this case the side-product accumulation will be more critical, thus their formation must be as low as possible.

7.2 The innovative strategy application

As it was mentioned in the main goals section, we planned to apply this ligation strategy for the chemical synthesis of natural anti-freeze proteins (*Tm*AFP) analogues, which have a practical interest and will be discussed in following chapters (**Figure 64**). To shed some light on this subject, I suggest that the novel synthesis strategy will not be compatible with *Tm*AFP analogues synthesis, because this is a cysteine-rich protein. Each peptide fragment contains 4 cysteines, meaning that three peptides assembly results 12 cysteines which can catalyse SetCys opening, what would be harmful for the process selectivity.

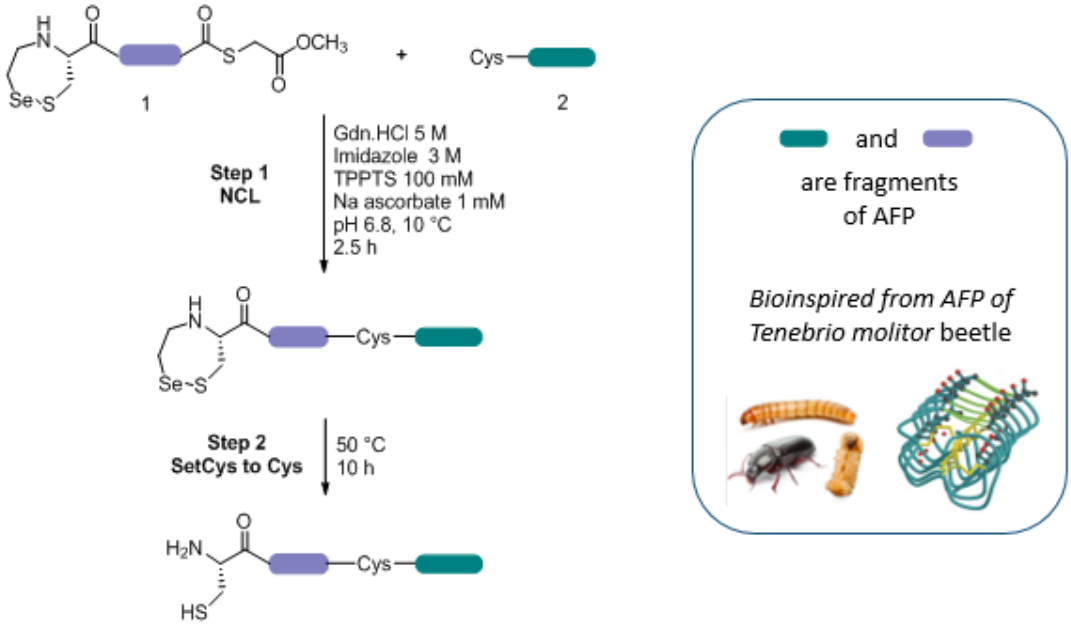


***Figure 64**. The innovative SetCys-based strategy as the tool to access a bioinspired antifreeze protein.*

Antifreeze proteins, their place in nature and application

8 Ice-binding proteins: types, properties, applications

Ice-binding proteins are macromolecules that possess the capacity to bind ice crystals at the stage of nucleation. Naturally present in plants and living organisms, ice-binding proteins protect their tissues from lethal damage due to ice crystal growth. The capacity of those proteins to protect organisms from cryo-injury has attracted scientist's interest and brought a wide field of discoveries. In past few decades, certain progress has been achieved in the development of anti-icing coatings, storage of food products and biological tissues.

As an introduction of my works on ice binding proteins, a bibliography part will be first dedicated to the different types of ice binding proteins found in nature, to their properties and to their actual and potential application in academia and industry. I will then focus on a specific and efficient antifreeze protein produced by *Tenebrio molitor* (*Tm*) beetle, an insect that resists arctic winter. This natural molecule has bioinspired my research aiming to design and synthesize a functional protein for improving the cryopreservation of *Schistosoma mansoni* (*Sm*), a parasite studied by the CBF team. Reproducing the complex life cycle of *Sm* at the laboratory is difficult and time-consuming. Therefore, cryopreserving some stages of the parasite instead of maintaining continuously its life cycle is of great interest. The synthetic efforts towards the designed antifreeze proteins capable of increasing the viability of *Sm* stages after freeze-storage will conclude this chapter dedicated to ice-binding proteins.

8.1 What are ice-biding proteins?

By definition, ice-binding proteins (IBPs) are molecules that adsorb on ice surface and control the growth of ice crystal. They are produced by cold-coping organisms to protect them from cryoinjuries (**Figure 65**).

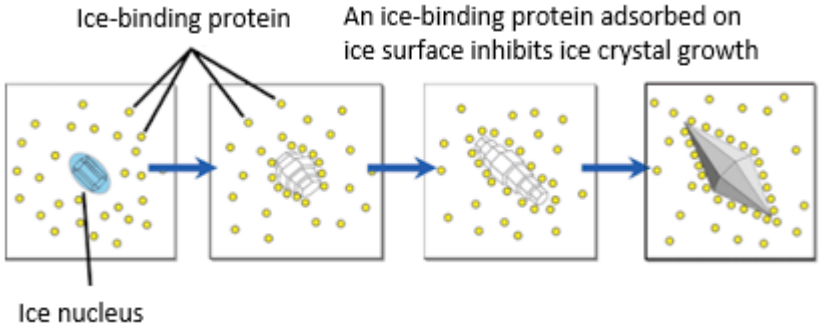


***Figure 65**. IBPs action via absorption on the surface of a forming ice crystal.*

A tremendous number of studies shades light on the large presence of ice-binding proteins in nature, on their diversity of structures and on their properties [109–111].

IBPs are found in fish[109], plants, bacteria, fungi[112], microalgae[113], insects[114,115], diatoms[113] and even yeasts[116]. They are expressed mainly in extracellular liquids, blood and

shell. Ice-binding activity was reported for the first time in 1964 by Ramsay[117] who demonstrated that the peripheral liquids of *Tenebrio molitor* beetle was frozen 8 °C below its theoretical freezing point. This discovery was followed by characterization of IBPs in Antarctic marine fish[109,118].

IBPs include sub-groups such as ice-nucleating proteins (INPs), antifreeze proteins (AFPs) and glycoproteins (AFGPs)[119]. Ice-nucleating proteins, found mostly in cold-tolerant organisms, initiate a seasonal freezing of body liquids. Some other organisms, to survive in cold environment, express AFPs or AFGPs that inhibit damages due to ice nucleation and growth. The term "antifreeze" point on the particular property of such proteins to decrease the freezing point of liquids what helps to resist the lethal freezing.

It is worth to note that the climate conditions are remarkably different for various freeze-resisting animals. Arctic winter means stable and dry cold conditions, while continental winter may bring repetitive temperature variations, which may lead to ice thawing-freezing cycles. Also, marine environment is opposite to dry arctic media. This fact forces various organisms to adapt differently and to produce IBPs having different properties. Those properties can be described and measured as thermal hysteresis (TH) and ice recrystallization inhibition (IRI).

8.2 Main properties of antifreeze proteins

Thermal hysteresis (TH) refers to the difference between melting and freezing points of a solution (**Figure 66**). In AFP-containing solution, the temperature gap results from the irreversible binding of the protein to ice surface. Such an interaction inhibits the growth of nucleated ice crystals until the temperature decreases to the non-equilibrium freeze point. Below this temperature, the burst of ice crystal can be observed[120].

Thermal hysteresis depends on the AFP concentration and is traditionally measured by nanoliter osmometry. Briefly, a small drop of AFP solution suspended in oil is snap-frozen and the resulting multicrystalline ice is warmed slowly and melted until only one small ice crystal remains. This first step of the process allows to determine the melting point of the AFP solution. In the second step of the process, the temperature is lowered slowly and continuously until the burst of ice crystal occurs. The difference between the melting and the depressed freezing points provides the thermal hysteresis of the AFP solution. Recently, microfluidic systems were used to facilitate the sample handling and crystals formation[121,122].

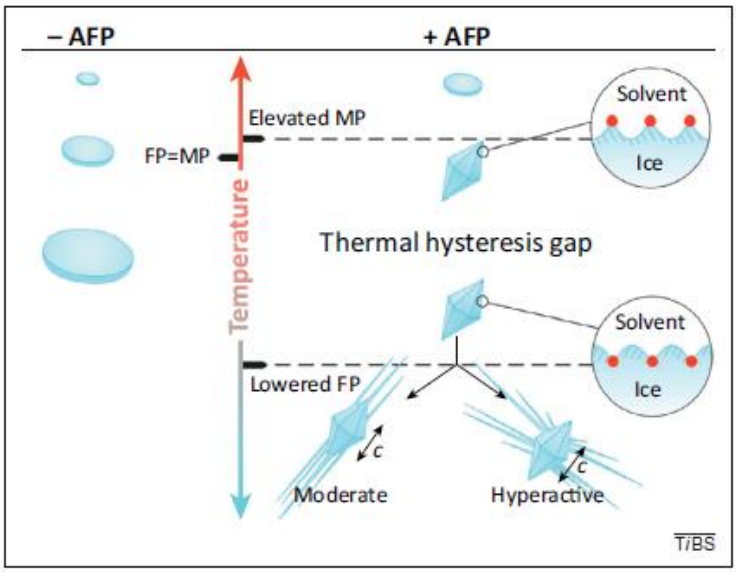


***Figure 66**. Thermal hysteresis induced by the adsorption of AFP (red dots) on ice surface.*

AFPs are classically classified in several categories depending on their TH activity and their structure (**Figure 67**)[121,123,124].

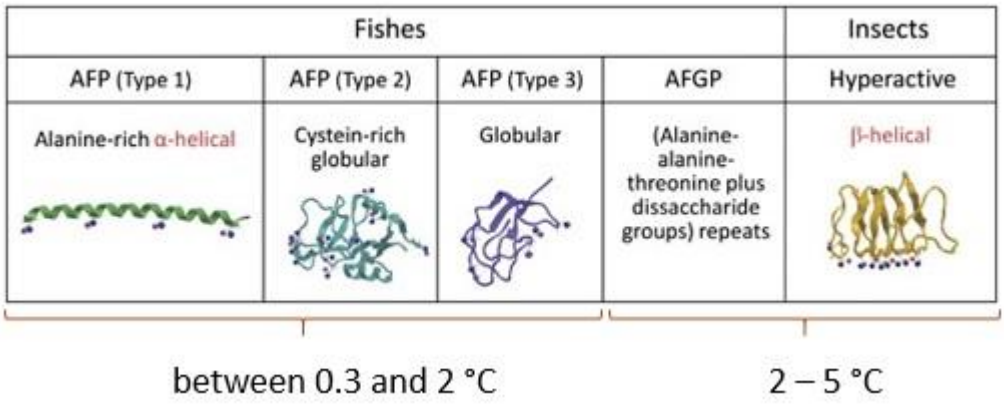


***Figure 67**. Main groups of antifreeze proteins produced by fishes or insects and their TH.*

I-III types AFPs, also called moderate AFPs, are mainly expressed by marine animals surviving during the winter in oceans or in ice-ladden lakes at near to zero or sub-zero temperatures[120]: fishes, like Antarctic notothenioids[109,111], ocean pout[110,125] and winter flounder[126,127]. According to different studies, moderate AFPs are able to modify TH by 0.5 – 2 °C at concentrations 10-40 mg/mL[114,119,128].

- Type I AFPs are 3.3–4.5 kDa Ala-rich repetitive sequence with the periodicity of about 11 residues and α-helical structure[129,130].

- Types II AFPs are globular, cysteine-rich lectin-like proteins of 11–24 kDa. Their folding in some cases is stabilized by disulfide bridges between Cys residues[131].
- Type III AFPs adopt a non-repetitive globular structure, devoid of cysteines, of about 6–7 kDa[132].
- Type IV AFP, a 12.2 kDa Gln-rich lipoprotein-like antifreeze protein, has been found only in blood plasma of the longhorn sculpin at a very low concentration. It has a faint antifreeze effect[133].

AFGPs contain from 4 to 50 tripeptide repeats of Ala–Ala–Thr with a disaccharide galactose-*N*-acetylgalactosamine attached to each Thr residue. Their mass range is 2-33 kDa. Their antifreeze activity is in some cases higher than for AFPs types I-IV and comparable with the one of hyperactive AFPs, however the handling and characterization in the laboratory is more laborious due to more complex composition.

Hyperactive AFPs, unlikely moderate AFP, act in non-colligative manner, at micromolar concentrations. It means that a hyperactive AFP may provide the same activity as a moderate AFP, but in a concentration 300 or 500-fold lower. Their low concentration minimizes their effect on osmotic pressure[134]. Hyperactive AFPs are often highly regular and shaped proteins such as β-solenoid type structures, they are found in insects and microorganisms undergoing extreme cold, which can be as low as -70 °C. Hyperactive AFPs decrease the freezing point by 3-6 °C in average[114].

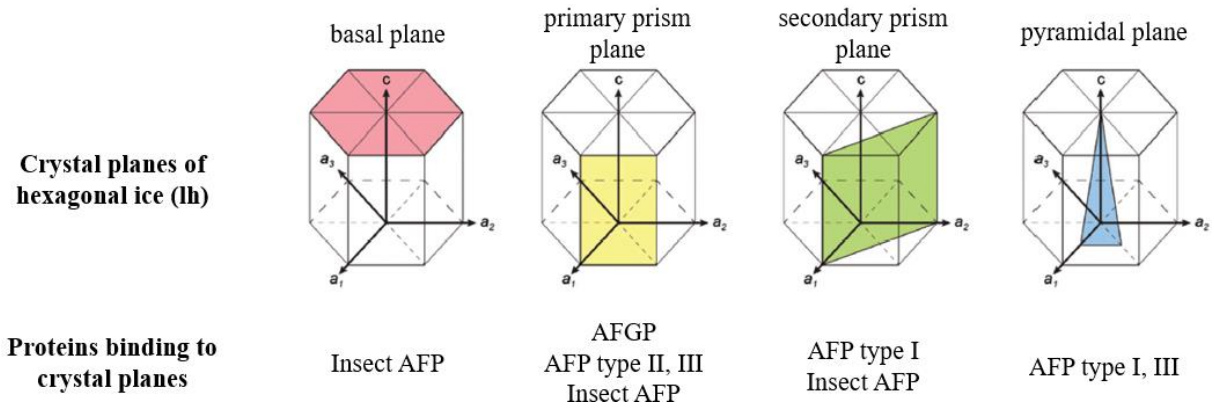


***Figure 68**. Crystal planes of hexagonal ice that are bound by different AFPs.*

It was established that the binding of AFPs occurs specifically at one or several faces of ice crystals depending on the type of AFPs considered (**Figure 68**, adapted from [120,135]). For instance, AFPs with moderate TH are known to mainly bind to primary, secondary or pyramidal planes[136,137]. In contrast, it has been experimentally demonstrated that the higher TH activity of hyperactive AFPs is due to absorption to ice crystals on its basal and prism faces[121]. This different affinity of AFPs to ice crystal planes was also explained by modelling of protein–ice interactions[138]. Besides modulating the TH activity, the absorption of AFPs to specific faces of

the ice/water interface also modifies both the rate of ice crystal's growth and the morphology of ice crystal.

Ice recrystallization inhibition is another essential property of AFPs. As it was mentioned, in some natural habitat temperature fluctuations take place, leading to thawing and recrystallization of ice in tissues of some organisms. Upon the recrystallization process, small crystals can growth and form larger ones, damaging cells (**Figure 69**). Some AFPs have remarkable capacity to inhibit the recrystallization by absorbing on ice crystal surface[139].

***Figure 69**. Cells damage from freeze-thaw cycle.*

To estimate IRI[140–142], an IBP containing sample is flash-frozen from room temperature to -40 °C to form a multicrystalline sample state. The ice is then annealed at - 6 °C over a certain period of time, allowing ice recrystallization to occur. The resulting IRI activity of the IBP corresponds to the lowest IBP-concentration where ice recrystallization cannot be observed.

It is interesting to discuss that both AFP cryoprotective activities, HT and IRI, are equally essential. It means that an AFP, to be functional, must possess both properties simultaneously. However, it was observed that AFPs with moderate TH activity have usually a higher IRI potential. In contrast, hyperactive AFPs with high TH activity have nearly no IRI power[143]. It can be explained by the natural habitat of the cold-adapted organisms. Mentioned earlier temperature fluctuations during the winter force those organisms to produce AFPs with high IRI capacity, as the ice in body liquids may thaw and freeze rapidly many times. At the same time, arctic winter demands a stable antifreeze protection against constant and extreme cold by producing AFPs with high TH activity. The latter case means slow cooling prior to winter and slow progressive thawing upon the winter ends. This observation is practically important for the design of bioinspired AFP mimics. Depending on the temperature range and other specificity of using them, TH or IRI activity should be in priority.

8.3 AFP application

Antifreeze proteins have found application in many fields of industry and academic research. Depending on specific needs, thermal hysteresis capacity or ice recrystallization inhibition can be the priority property.

8.3.1 Industrial applications

Some ice-binding proteins are currently tested on their capacity to control atmospheric glaciation processes[144] and climate[145] by modifying atmospheric precipitations time and power.

AFPs may be applied in a broad range of water-based materials or influence ice shaping during the production of porous composite materials[146].

AFPs were found to be efficient "green inhibitors" of the crystallization of clathrate hydrates, crystallohydrate-like solid agglomerations of gas molecules and frozen water. Clathrates are formed during the storage or transport of some organic gases or liquids. In contact with air moisture, under certain atmospheric conditions (temperature, pressure), hydrate crystals may form and perturb the matter flow. In addition, during the studies, AFPs demonstrated hydration inhibitory capacity which is competitive to the one of the commercial hydrate inhibitors polyvinylpyrolidone[147] or its substituent methanol[148] that inhibits hydrate crystals formation in gas pipelines. According to another study[149], the hydration inhibitory capacity of methanol was improved by adding a *Rm*AFP1, a beetle AFP, what represents a hybrid approach that may demand less quantities of needed AFP once mixed with methanol.

Anti-icing materials are of high interest especially to protect metallic surfaces such as aluminium which is a common material in aeronautics and cooling equipment fabrication among other applications[150,151]. Classically, a hydrophobic ZrO_2 coating is used to protect Al layer. However, this method was shown the way less efficient comparing to protein coating. For the assay, an AFP was linked to an Al-binding protein. As a result, this coating keeps providing efficient ice-shaping in a temperature range between -5 and -26 °C. In another application, polymer films[152] were successfully modified by AFPs through conjugation. This modification conferred to a polymer films the additional freeze resistance.

Agriculture can also benefit from cold hardiness boost by introducing into plants genes responsible for AFPs expression in different parts of a plant (leaves, core, etc.)[153].

Last but not least, food technology constantly challenges food freezing and chilling problems. Freeze-thaw cycles allow ice crystals growth, what leads to loss of food quality. One of primary solutions is the addition of sugar or polysaccharides derivatives as they are humectants and structuring agents. However, sugar is not tolerated by insulin-deficient people, thus, other options must have been explored. One of them is AFPs addition. AFP types I-III and their analogues are more demanded in this field, as IRI capacity is crucial here[154]. The successful use of those proteins was demonstrated for storage and re-freezing preservation of meat, fish, dough and prepared breads[155], ice-cream based products[156]. The latter has been the most successful attempt of freeze-thawing resistance to recrystallization, probably due to synergic antifreeze action of AFPs, sugar and fat molecules contained within the product.

8.3.2 Biomedical application

Cryosurgery is being studied as an alternative to classical surgery approach that destroys targeted part of tissue and that demands only local anesthesia in some cases. Apart from demanding less anesthesia, it might facilitate the chirurgical manipulation as soft tissues are "immobilized", what increases a surgery precision and decreases tissues degradation. This surgery method requires a short-term tissue freezing that would be not lethal for cells[157]. In a study with liver tissue, a natural antifreeze proteins of type I from the winter flounder was used[158]. The study revealed that after the first freeze – thaw cycle, the AFP I was still active and could protect tissue from cryoinjury, while after the second freeze – thaw cycle the cell membranes within the tissue were much weaker. Cell leakage allowed water penetrating the cells and forming intracellular ice crystals, a phenomenon that resulted in lethal cells cryodamage. Thus, AFPs may be a solution for cryosurgery, if freeze – thaw cycle was managed only once.

8.3.3 Cryopreservation

Cryopreservation is an approach to a preservation of a biological material from metabolic injuries and aging at low temperatures. Such a preservation can be extended in time and must be reversible with temperature increase. Different kind of biological material are currently being cryopreserved: cell lines[159], tissues, organs[160] or embryos[161]. The effectiveness of freezing for any samples is evaluated by looking at metabolism, cells recovery and size (shrinkage effect) compared to control unfreezed sample. As cryopreserved samples often demand water-containing media, the main focus of cryopreservation research is the design of the solution media itself[162,163], which controls the osmotic effects causing tissues dehydration and ice formation in extra- and intracellular liquids. During freezing, the solutes are excluded from the growing extracellular ice crystals, thus, they become more concentrated with time. Those effects generally lead to cell damage or death[164].

The first additive used to reduce the tissues cryodamage is glycerol[165]. This additive was utilized to develop the high glycerol/slow freeze and low glycerol/rapid freeze techniques [166,167]. Later, extracellular cryoprotectants such as hydroxyethyl starch (HES), dextran, polyvinyl pyrrolidone and serum albumin have been studied as well for the same purpose. These macromolecules were found to minimize osmotic stress due to the addition and removal of glycerol, and facilitate the procedure in general. Nevertheless, the risk in thawing remains another huge issue because of ice recrystallization.

DMSO and glycerol are commonly employed cell penetrating agents, which are effective but not without disadvantages: cell damage is decreased but not fully prevented due to solvent toxicity and the necessity to remove them after thawing prior to further manipulations[160].

A few works examined the interest of AFPs for cryopreservation. Some improvements have been achieved by using the AFP of an arctic yeast *Leucosporidium*[168], as the main

cryoprotective component mixed with glycerol in PBS. The *Le*AFP concentration range was varied between 0-1.5 mg/mL, and the rate of cooling and thawing was also adjusted. Interestingly, the additive was efficient against ice recrystallization at any rate of cooling/thawing, resulting in almost quantitative cells recovery. This experiment clearly shows the minimal active AFP concentration (0.4 mg/mL) and the damaging osmotic effect above the limit of 0.8 mg/mL. The absence of this protein in the cooling mixture resulted in only 32 % recovered red blood cells at maximum[169].

Clearly, it is still challenging to store living matter under hypothermal conditions. A lot of improvements must be implemented into existing techniques to achieve quantitative cryopreservation. Meanwhile, antifreeze proteins have already demonstrated a potential efficacy, thus, the studies should be pursued.

Futuristic application in cryopreservation. It is not a secret that many human inventions are inspired from Nature. A long evolutionary path allowed to some terrestrial and marine organisms to cryopreserve and inhibit themselves during cold periods. Knowing the mechanisms of this cryo-protective systems, humans may think about cryogenic systems for future interstellar travelling. Perhaps, it may be managed nearly as it is shown in science fiction novels, *via* immersion into an antifreeze solution and injection in blood or other body liquids. The idea might be developed starting from an example of a cold-adapted beetle *Rhagium inquisitor* that preserves itself from freezing by cumulating antifreeze proteins in intracellular and intestine media, hemolymph[170], or as discussed earlier through studies on organisms recovery after being overcooled.

Nowadays, the most powerful space agencies are making the first steps in designing cryogenic cameras for “hibernation” that could maintain hypothermal conditions, using the existing knowledge from cryopreservation studies[171].

8.3.4 Cryopreservation and Schistosomiasis

Although highly optimized, classical cryopreservation techniques can induce significant (cryo)injuries that impair survival and developmental potential of cryopreserved living organisms. For example, no effective cryopreservation techniques are available for *Schistosoma* species, a dangerous parasite[172]. This fact seriously impacts its study. Several laboratories in the world, including the Chemical Biology of Flatworms team in Lille (CBF)[173], concentrates its research on Schistosomiasis, a tropical neglected disease that affects 260 million people in the world. The infestation occurs through the contact with fresh water, contaminated with *Schistosoma* parasites. *Schistosoma mansoni (Sm)*, one of the most frequent *Schistosoma* species infecting human, is actively studied with the aim to design novel therapeutics against this parasite.

The *Sm* life cycle in nature is summarized in **Figure 70**. When eggs are in contact with water, they hatch to release miracidia that swim to find mollusc to infest and transform into sporocyts.

***Figure 70**. Schistosoma mansoni life cycle.*

By asexual reproduction, a sporocyte multiply to thousands cercariae. Once in water, cercariae are searching a final host, usually a rodent or a human, cross its skin, transform to schistosomula and penetrate into the vascular system. Then, they migrate towards the mesenteric vessels, finish their transformation to *Schistosoma mansoni* male or female, form permanent couples and start massive eggs production close to the intestine capillaries. Eggs cross the vessels and are then rejected with feces. Once they are in contact of fresh water, the cycle starts again and reproduces.

Female can produce up to 300 eggs per day and since 50 to 80% of eggs are flushed out by the blood circulation, a big part of eggs stuck into the liver. They promote the formation of granulomas and of a strong inflammatory response. Over the time, they induce a liver fibrosis, associated with portal hypertension and increase the risk of liver and colorectal cancer. This complex problem brought to scientists the challenge to find efficient treatment against *Sm*. There is one existing efficient drug against *Sm*, Praziquantel[172], which has several drawbacks such as high required dosage, efficiency decrease in case of re-infestation and above all, lack of efficacy on *Sm* larvae. Therefore, it is still necessary to keep looking for new solutions and thus to maintain the parasite in the laboratories for drug screening.

Studying the biology of *Sm* and performing screening campaigns requires maintaining its complex life-cycle[174] involving a fresh water snail as intermediate host and a definitive mammal host (hamsters in the lab) (**Figure 71**).

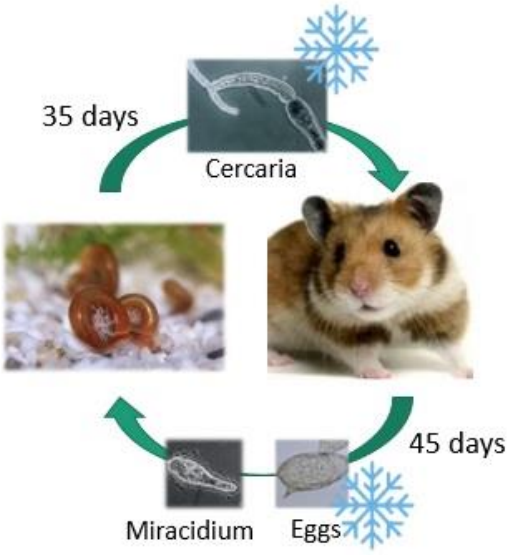


***Figure 71**. Schistosoma mansoni cycle in the lab might benefit from the cryopreservation of parasite at stages of cercaria and eggs.*

In absence of in vitro culture models, of effective cryopreservation techniques and considering the duration needed to produce the different stages, this life-cycle cannot be stopped. Here is the practical interest of efficient cryopreservation of *Sm*, in particular their cercariae and egg stages. First, having in hand the working method, we would be able to freeze cercariae until we need to obtain some more adult worms. This way, it would be possible to stop constant use and sacrificing hamsters. Similarly, being able to cryopreserve the eggs obtained from livers of infected animals properly, we could launch a new parasite cycle at any moment. In addition, a common benefit of cryopreservation of eggs is the ease of transportation of such biological materials between research centres. At present time, the parasites are transported inside alive infested molluscs. This is done just after having infested the snails and before cercariae can develop to avoid any possible contamination.

The cryopreservation of *Sm* stages was attempted by many different research teams at the end of the last century. In fact, the nature of additives in the media, the final temperature of freeze-storage and the cooling rate must be considered. The first successful *Sm* recovery after being frozen was reported in 1977 for the stage schistosomula, as both cercariae and schistosomula lifespan at atmospheric conditions is no more than a few days[175]. Methanol (17.5 %) was used as cryopreserving agent, being not highly toxic at short exposure time. A two-step cooling procedure was designed. The first step is a slow cooling between room temperature and -38 °C at the rate of 0.56 °C/min, allowing homogeneous ice nucleation inside and outside cell membranes. The second step is rapid freezing *via* sample immersion in liquid nitrogen (-196 °C), where the rate is estimated as 10’000 °C/min. It is important to note that the step of slow cooling was limited at intermediate sub-zero temperature in order to avoid osmotic cells dehydration. After 30 min at -196 °C, the schistosomula sample was incubated in a 37 °C bath

for 10 min while largely diluted in collagen based media (to decrease methanol concentration). This method resulted in a viability of only 0.44% for schistosomula, calculated as maturated schistosoma inside their host compared to a non-frozen schistosomula sample. Later, the next step forward the successful cryopreservation was done by replacement of methanol with ethanediol (35 %), increasing the viability to up to 5 % of control test[176]. From another hand, keeping methanol as a cryo-conservative agent with single addition at 0 °C but modifying the cooling rate by starting the rapid cooling from 0 °C, the authors increased the parasite viability up to 6 %[177]. The consequent improvements of cooling rate and media nature resulted in 8 % of viable *Sm* stage[178]. Other trials of cryopreservation in the presence of DMSO or under different cooling conditions resulted in no viable *Sm* after temperature recovery. Also, all those efforts are valid for very short overcooling time not exceeding 1 hour.

Regarding promising cryopreservation capacity of AFPs in non-colligative manner and low cytotoxicity of those macromolecules in comparison with more toxic small molecules such as methanol, DMSO or ethanediol, a potentially attractive approach for cryopreserving invertebrates such as *Sm* would be to use AFPs. Due to one of the highest measured TH, the AFP expressed by *Tenebrio molitor* beetle would be an interesting candidate for cryo-protection experiments with *Schistosoma mansoni*.

8.4 *Tenebrio molitor* AFP (*Tm*AFP)

Our interest is focused on an antifreeze protein expressed by a beetle *Tenebrio molitor* (*Tm*AFP), in its larvae stage. Its thermal hysteresis reaches 3-4 °C depending on impurities presence, at concentration about 1 mg/mL[179]. TH of *Tm*AFP is 10-fold higher than for AFP III[180]. In contrast, IRI capacity is 4-fold lower than for a moderate AFP III at concentration 0.5 mg/mL. From structural point of view, AFPs isolated from *Tm* are small β-helical proteins (above 10 kDa) stabilized by regularly spaced disulfide bridges This hyperactive protein consist of several repeats of 12 amino acids residues with a consensus sequence CTxSxxCxxAxT[181] (**Figure 72**). The most common *Tm*AFP isoform consists of 8 repeats, however 7 – 10 repeat isoforms were also isolated[182,183].

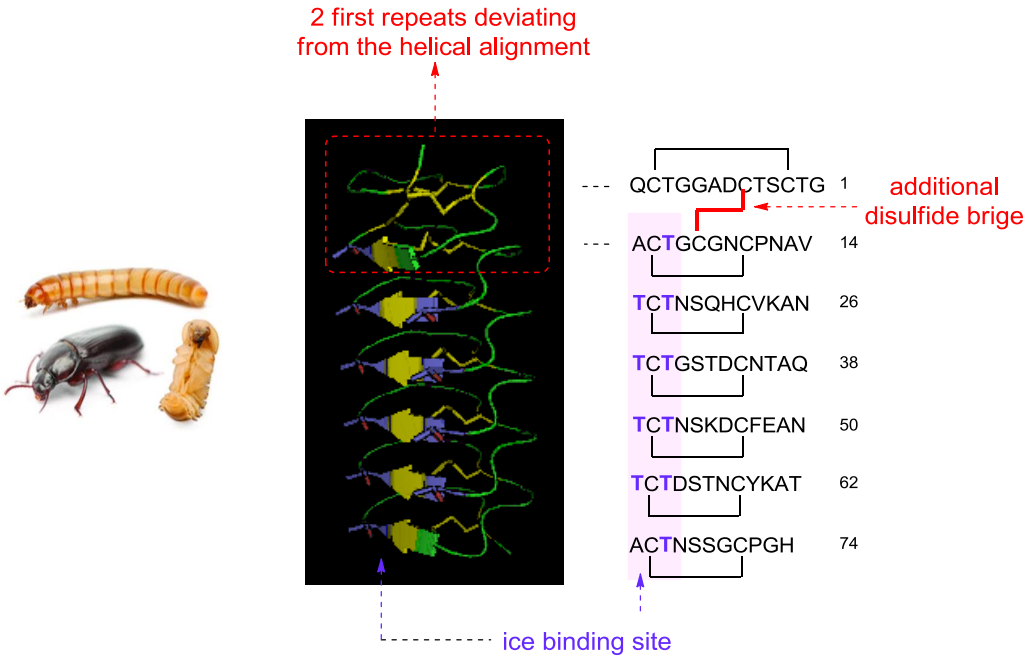


***Figure 72**. 2D and X-Ray structures of the natural 7 loop protein (PDB: 1EZG) from which the TmAFP mimics are bio-inspired. In yellow and purple are respectively shown the disulfide bonds and the threonines involved in the ice-binding site of TmAFP.*

The β-helical structure of *Tm*AFPs was established using X-Ray analysis[123,184]. As shown in **Figure 72**, the first two repeats are distorted due to an inter-coils disulfide bridge. The five following coils are highly similar and form regular loops that are stabilized by a unique intra-loop disulfide bridge. In this highly ordered folded secondary structure, the threonine residues of the TCT motif are aligned on one face of the protein and are known to form the ice-binding site.

To confirm the involvement of this array of Thr residues on the ability of *Tm*AFP to bind ice, a series of mutants were produced where Thr residues of the protein were substituted by other amino acids, such as Tyr, Leu or Lys residues[185]. Mutations led to a massive decrease of the TH activity when the Thr residues belonged to the putative ice-binding site (**Figure 72**, T26Y, T62Y, T38Y, T40Y, T40L, T40 K). In contrast, the replacement of Thr located on another face of *Tm*AFP helix (**Figure 72**, T43Y) didn't change significantly the anti-freeze properties of the protein.

Furthermore, Marshall and co-workers[179] showed several years ago that the antifreeze activity of *Tm*AFPs highly depends on the length of the β-helical structure, i.e., on its number of loops (**Figure 73**). To investigate such a relationship, several analogues in terms of length were produced and purified by modifying the wild-type *Tm*AFP 4-9 made of 7 coils. TH measurement revealed that the deletion of one coil compared to the natural protein led to a drastic decrease of TH (from 3 °C to 0.5 °C at 60 µM). In contrast, the insertion of additional loops increased the anti-freeze activity of the protein, reaching a maximum TH value of 6 °C

for the analogue made of 9 loops. According to Marshall's study, further increasing of the length of the protein tends to impair the anti-freeze properties of the proteins.

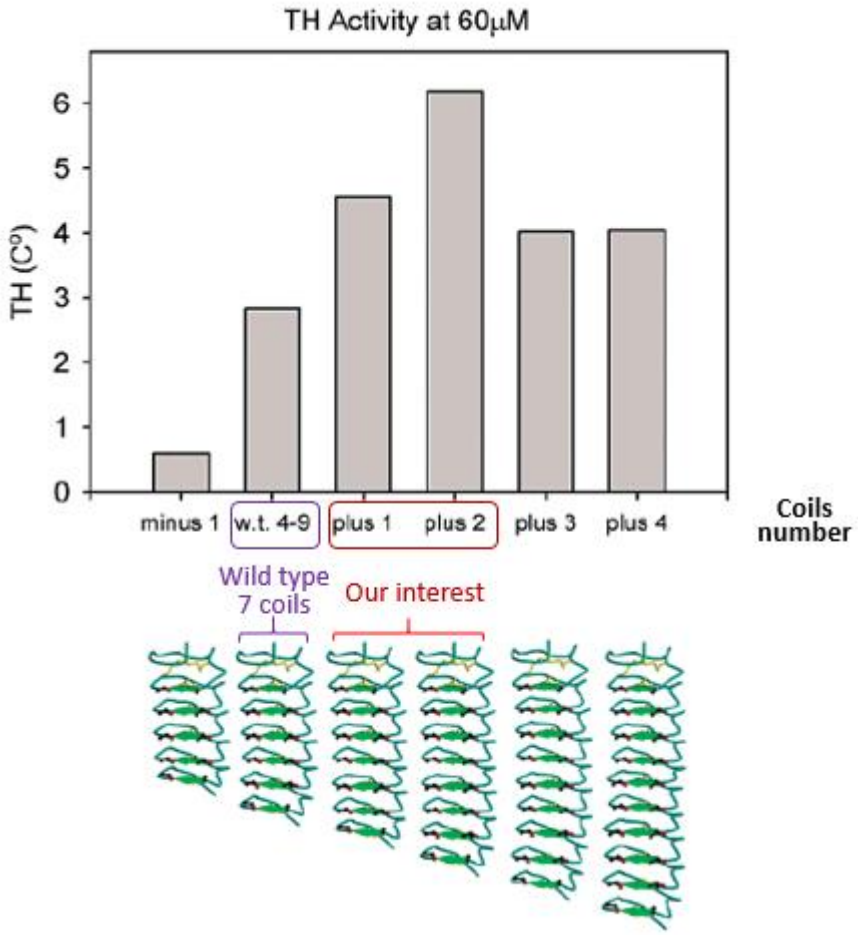


***Figure 73**. TH activity of TmAFP made of increasing number of loops (adapted from Marshall et al., 2004 [71]).*

In conclusion, *Tm*AFPs are small β-helical proteins that have high TH activity. It has been shown that their ice-binding site is well identified and composed of an array of Thr located on one face of the highly regular helix structure. Furthermore, the TH activity of this family of proteins can be easily tuned by modifying the length of the helix, i.e., the number of loops the protein is made of. As consequence, *Tm*AFP is an attractive model to design novel bio-inspired proteins having potent TH activity. Due their properties, these *Tm*AFP mimics could have interesting applications in various fields, and especially for the cryopreservation of *Schistosoma mansoni*. Maintaining the life cycle of this parasite at the laboratory is complex, time consuming and expensive and requires animal sacrifice. To improve the process, the cryopreservation of some parasite stages can be a solution. Existing cryopreservation techniques entered biomedical routine for single cells or cell cultures conservation, but not for such a complex organism as *Sm* larvae. Several attempts to freeze and thaw them resulted in only about 5 % of theoretical survival of parasites. This is why we wanted to test *Tm*AFP mimics as additive for improving

our capacity to cryopreserve *Sm* larval stages. The design of such structures as well as my results towards their synthesis are described in the next section.

Interest to an antifreeze protein and synthesis of its bioinspired analogue – results and discussion

9 Results and discussion

The aim of this work which has started prior to the beginning of my thesis is to produce *Tm*AFPs by chemical synthesis.

Up to now, natural *Tm*AFPs as well as their analogues have been produced by recombinant method[184,186] or isolated from *Tenebrio Molitor* larvae[187]. In the latter case the protein extraction demands animal sacrificing, and for both cases the isolation of the target compounds requires extended and laborious purifications, leading to poor yields.[179,188] Moreover, overexpressed in bacteria proteins often requires further refolding. Another drawback of recombinant protein production is the presence of contaminants from the expression system that might be immunogenic and therefore induce immunological reactions during cryopreservation experiments with *Sm* stages.

To solve these problems, we plan to use chemical synthesis to produce the targeted *Tm*AFPs. Synthetic approach allows chemists to obtain non-immunogenic D-proteins and to easily insert modifications in a polypeptide sequence (for one or more AAs). It provides less complex mixtures and homogeneous peptide in terms of peptide length, facilitating the subsequent purification step. In spite all the advantages provided by chemical synthesis, no chemical synthesis of *Tm*AFPs has been reported so far in the literature. In this section, I will describe the synthetic strategy that I used to assemble *Tm*AFPs by chemical ligations, highlighting challenges that I faced during this synthetic work. To simplify the block chemical synthesis of the target compounds, we designed analogues bio-inspired from the regular structure of the wild -type protein.

9.1.1 Design of *Tm*AFP analogue

As previously discussed, the repetitive structure of *Tm*AFPs is stabilized by a series of disulfide bridges, positioning two rows of Thr residue that form the ice-binding site on one face of the helix. Even if the loops are very similar in terms of structure, their sequence are not strictly identical. As a consequence, synthesizing natural *Tm*AFPs requires the preparation of a large number of peptide segments and their assembly by chemical ligations.

To simplify the assembly of *Tm*AFPs by chemical ligations and to potentially optimize their ice-binding properties, we designed analogues that are inspired from the natural proteins and composed of a strictly identical repeats (**Figure 74**). In these *Tm*AFP mimics, the repeated module (CTGSKDCFEATTCTGSTNCYKATT) connected by sequential chemical ligations is made of 24 residues and correspond to two loops of the final helix. By design, the cysteine residues needed for the stabilization of the helicoidal structure by formation of disulfide bridges as well as the threonine residues involved in the ice-binding site are remaining in the sequence of the module. The other amino acids correspond to a consensus sequence of *Tm*AFP variants reported in the literature.

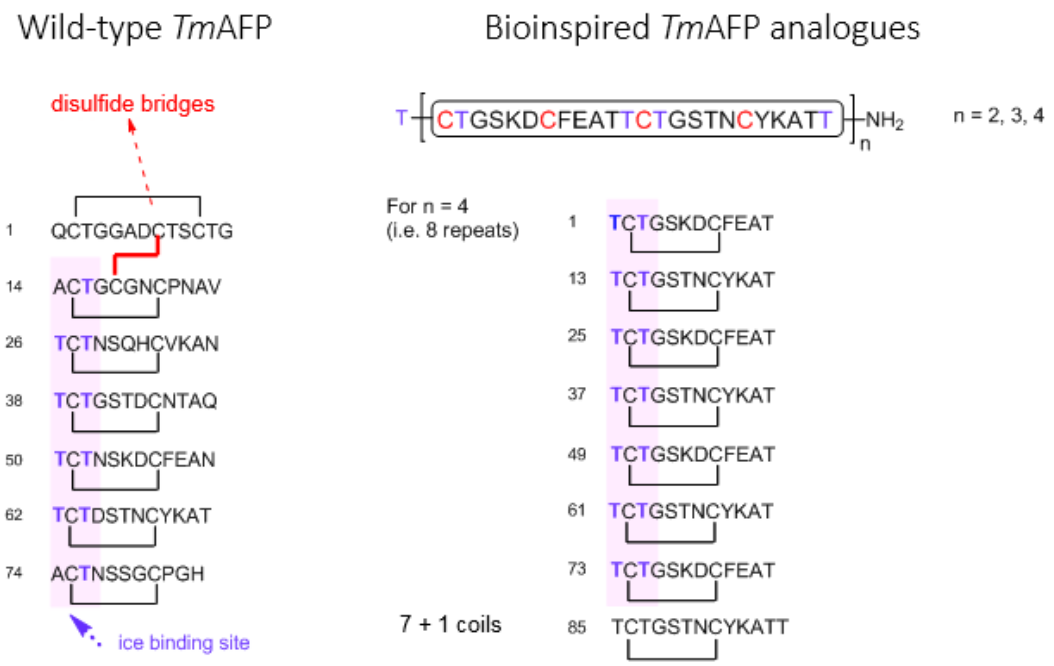


***Figure 74**. The WT and bioinspired TmAFP structures.*

As their natural parent protein, *Tm*AFP mimics are predicted to fold in a highly regular structure and lack the extra disulfide bond found in the natural protein for the first two loops, which could complicate the subsequent folding step. Note also that the designed AFPs share some common features with natural *Tm*AFP isoforms only identified at transcriptomic level (see for example: UniProt-Q6DLX5), since they are composed of regular repeats and lack the first two distorted loops. Since the number of reactions defines the size of the final polypeptide, it is possible to obtain polypeptides of a needed length. Another possibility to investigate structural features by using chemical synthesis is the facile insertion of modifications in the initial module of the *Tm*AFP sequence by replacing some amino acids of the consensus structure by others. In this way, we plan to optimize the ice-binding surface as well as the formation of the disulfide bridge pattern of this cysteine-rich family of proteins.

9.1.2 Synthesis of the designed *Tm*AFP analogue

The sequence (CTGSKDCFEATTCTGSTNCYKATT) obtained by conventional solid phase peptide synthesis is the starting material to assemble the repetitive structure of the targeted AFP by sequential chemical ligation. To master the chemistry of this cysteine-rich sequence, we planned to assemble an increasing numbers of modules (2, 3 and 4) (**Figure 75**).

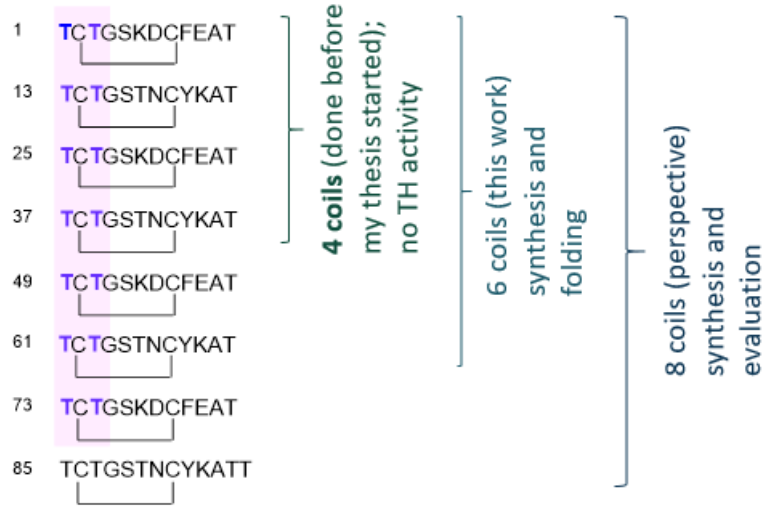


***Figure 75**. Progressive mastering of the chemistry leading to the assembly of TmAFP analogues by chemical ligations.*

Assembling two modules have been done prior to my thesis as a part of a preliminary study (**Figure 76**). We chose to ligate the precursors by applying classical NCL ligation. The resulting polypeptide with four cysteines was successfully folded in a mixture of oxidized and reduced glutathione within one week, leading to a helix with 4 loops stabilized by 4 disulfide bonds. The prepared *Tm*AFP mimic was assayed for anti-freeze properties but no TH was detected. This result is in accordance with Marshall's work about activity – loop number relationship (as shown in **Figure 75**): only β-helical proteins composed at least of 6 loops exhibit TH activity.

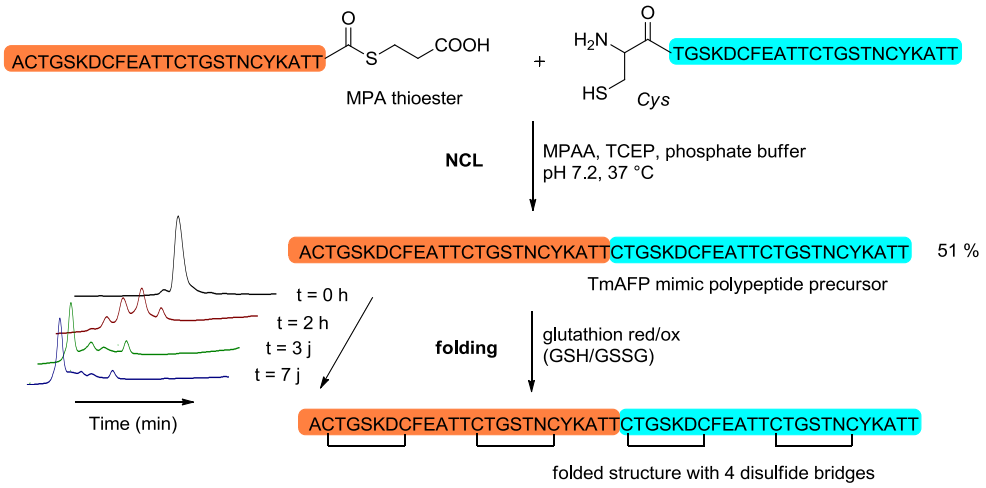


***Figure 76**. Preliminary results: assembly of the TmAFP analogue made of 4 loops.*

As a first step towards longer and more active *Tm*AFP mimics, I tested the concatenation of three peptide modules using a one-pot, two-step ligation process designed in our team some years ago and based on the SEA chemistry[189] (**Figure 77**).

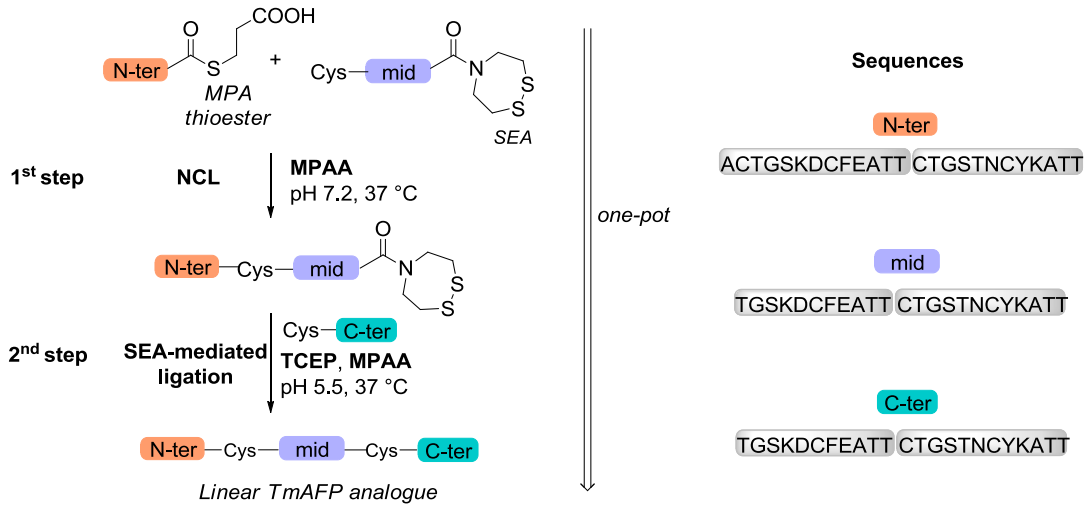


***Figure 77**. Three-peptide segment assembly using NCL and SEA chemistries.*

In the first step, the N-ter segment is connected to the middle segment by classical NCL reaction. The SEA group that functionalizes the C-terminal extremity of the middle segment remains silent as a cyclic disulfide under the weak reducing conditions imposed by the large excess of MPAA catalyst. The second ligation is then triggered by addition of the strong reductant TCEP. As depicted in **Figure 78**, the latter efficiently reduces the disulfide bond of the SEA group ($SEA^{off} \rightarrow SEA^{on}$), allowing its rearrangement into a transient thioester and the ligation step with the *C*-terminal cysteinyl peptide segment to proceed.

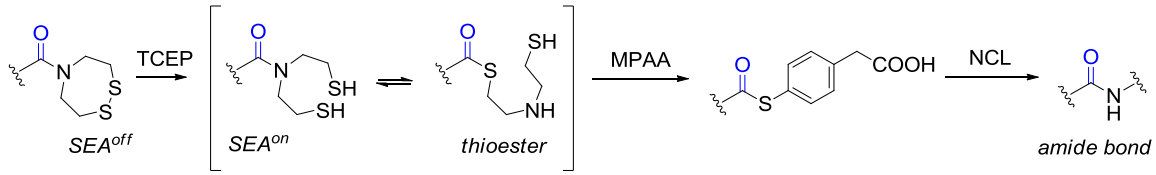


***Figure 78**. The reminder about the mechanism of SEA-mediated NCL.*

The two-step process is designed to be concise and elegant, however the nature of the peptide brought us several challenges to overcome during peptide segments synthesis and their assembly by chemical ligation.

9.1.3 Synthesis mastering and problem solving

The chemical synthesis of the *Tm*AFP analogue made of 6 loops started with the synthesis of the N-ter, middle and C-ter peptide segments by Fmoc-SPPS. We chose to introduce manually the [Ala-Thr] unit in the middle of the peptide sequence as pseudo-proline (ψ-Pro, or Fmoc-Ala-Thr($\psi^{Me,Me}$ pro)-OH) (**Figure 79**). By masking some NHs in the backbone

of the growing peptide, the aim is to minimize the aggregation that could interfere with the elongation process.[190,191] Stable during the elongation of the peptide sequence, the acetal function that masks the Thr residue is easily removed during TFA resin cleavage.

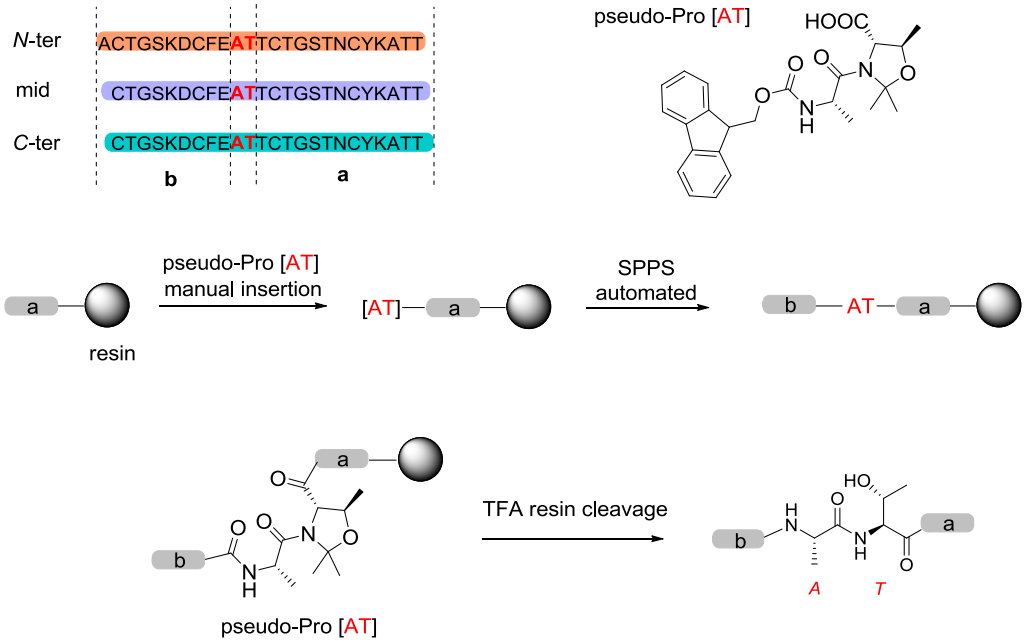


***Figure 79**. Pseudo-proline-mediated synthesis of the peptide segments needed for the assembly of the TmAFP analogue.*

9.1.3.1 Synthesis of the peptide segments

A SEA solid support was used to produce by Fmoc-SPPS the *N*-terminal fragment functionalized with a MPA thioester (**Figure 80, a**). After elongation of the peptide sequence, cleavage from the solid support led to the release of an intermediate SEA^{on} peptide. A transthioesterification step was then performed at pH 4 in the presence of MPA (mercaptopropionic acid) to convert the crude peptide into the desired peptide MPA thioester. After the HPLC purification, the compound was isolated with an overall yield of 18%.

C-terminal peptide fragment is obtained by SPPS on Rink amide resin (**Figure 80, b**). TFA-mediated cleavage from the solid support followed by HPLC purification provided the target peptide with 14% overall yield.

For both peptides, cysteine side chain was protected with trityl (Trt) protected group, which was removed during the resin cleavage with TFA, resulting in unprotected cysteine thiols.

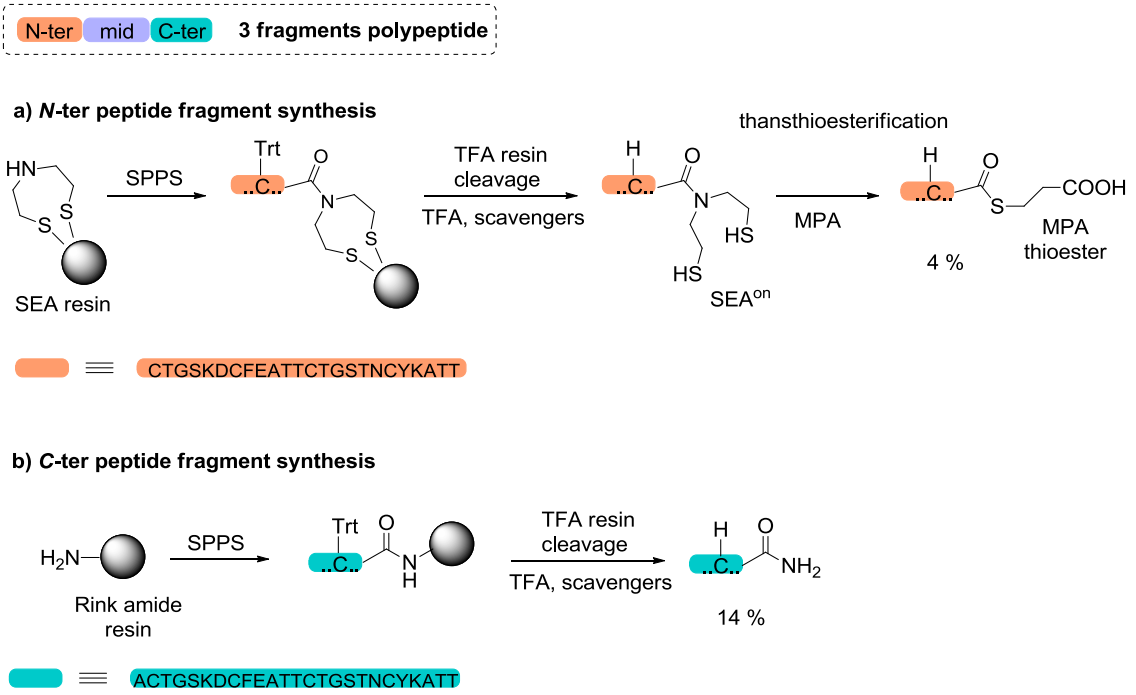


***Figure 80**. Synthesis of N- and C-terminal fragments of the designed TmAFP analogue.*

Synthesis of the middle segment is more complex and was performed starting from a SEA solid support. Once the middle sequence is obtained by Fmoc-SPPS and liberated from the solid support, the SEA^{on} group must be closed by disulfide bond formation (**Figure 81, a**). Classically, SEA^{on} is converted to SEA^{off} by oxidation in solution with iodine prior to the HPLC purification step. Meantime, internal cysteines of the sequence must be protected during SEA oxidation. One option is to use S*t*Bu protected cysteines during the SPPS step. Since the S*t*Bu protecting group is stable under the acidic conditions used to cleave the elongated peptide from SEA resin, the SEA^{on} group released in solution can be selectively oxidized with iodine. The S*t*Bu protective groups are then removed under the conditions of the NCL or the SEA ligation.

The problem I faced during the first attempts to produce the middle segment according to this classical strategy is that the peptide fragment armed with four S*t*Bu- protected cysteines becomes hardly soluble in aqueous media. This is why we switched to classical *Trt* protecting group for Cys, which is TFA labile. To overcome the problem of unprotected cysteines oxidation during SEA inactivation in the presence of iodine, we search for a solution enabling self-protection of the cysteines by forming internal disulfide bonds under redox-controlled conditions (**Figure 81, b**). Classically, glutathione redox pair is used to initiate disulfide bridge formation, however this is an expensive reagent for a scale of a crude peptide over 100 mg. Thus, the use of cystine was considered to be a good and cheap alternative to promote SEA oxidation and selective intramolecular disulfide bond formation (**Figure 81, b**).

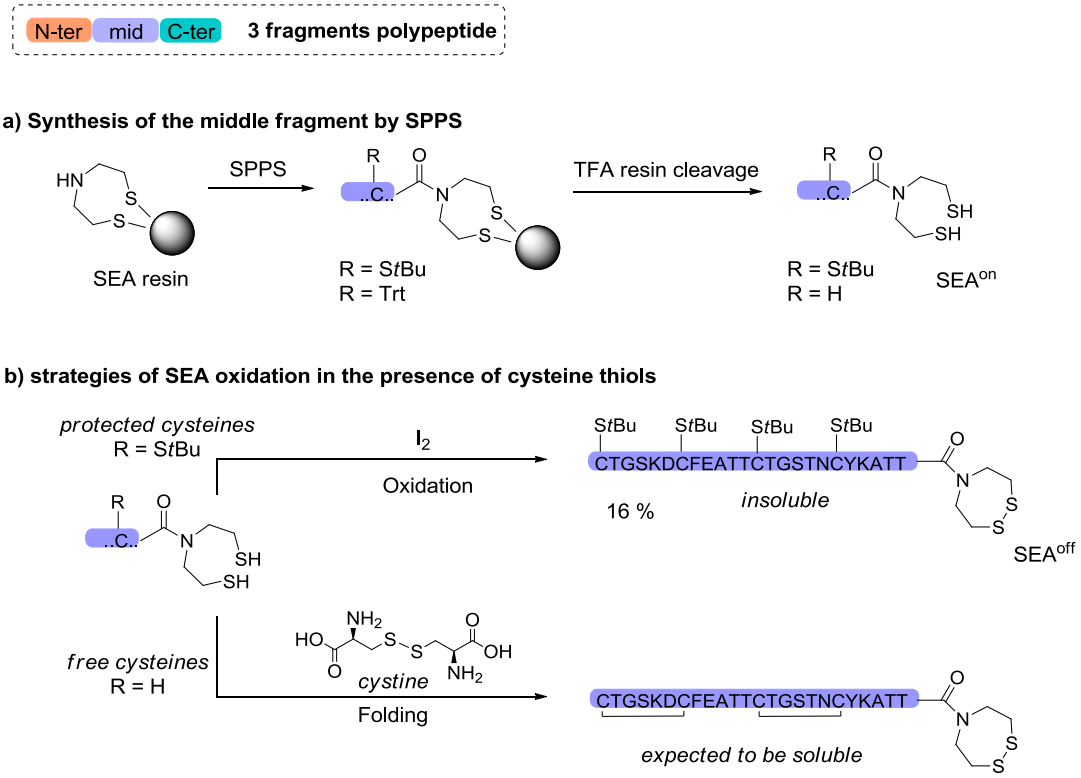


***Figure 81**. a) The middle peptide fragment synthesis on SEA solid support and TFA resin cleavage; b) Overcoming the low solubility of StBu-protected cysteines by adopting a self-protection strategy with the formation of internal disulfide bonds.*

Prior to synthesizing the new middle fragment with *Trt*-protected cysteines, the optimization of cystine-driven oxidation was first tested on already prepared *C*-terminal fragment. Cystine is used in its oxidized state, forming the redox pair with cysteines in the peptide sequence (**Figure 82**).

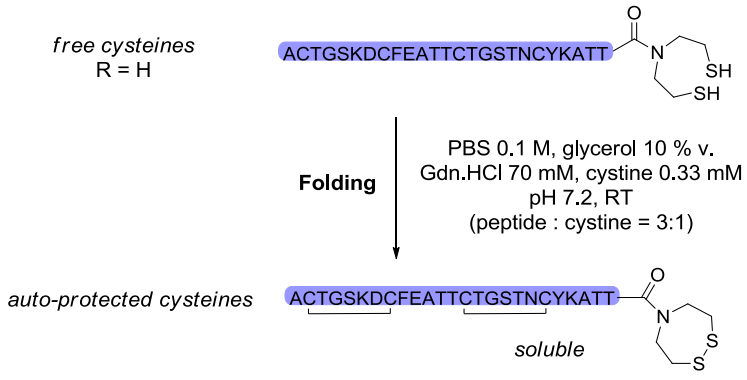


***Figure 82**. Cystine-mediated oxidation of the SEA group and Cys residues: optimization of the reaction conditions on a model compound*

The best ratio between peptide unprotected cysteines and cystine was established experimentally and was cysteine/cystine 3:1. The oxidation, initiated by cystine, occurs in the PBS media with 10 % vol. of glycerol, with minimal necessary addition of Gdn·HCl (70 mM)

to solubilize the peptide (1 mM). The self-protection occurs properly within 17 hours even for the crude starting material resulting in a unique structure with the two expected disulfide bridges. The position of two formed coils was proved by enzymatic digestion with Endoproteinase GluC, selectively cleaving at C-ter of glutamic acid (E) at pH 7.8, resulting in two separated peptide fragments containing each one disulfide bridge (**Figure 83**)[192].

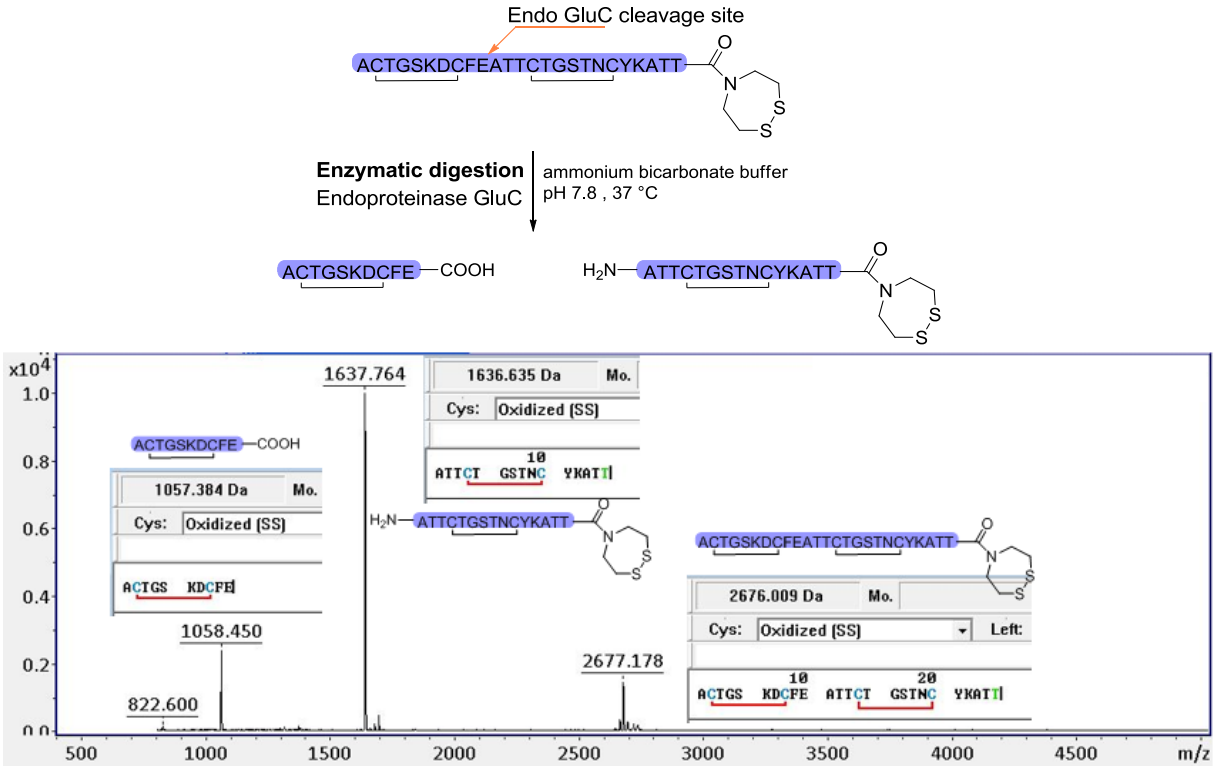


***Figure 83**. The middle peptide fragment after oxidation with cystine of the SEA^{on} and MALDI-TOF analysis of the fragments obtained after the peptide enzymatic digestion.*

The fact that only one isomer is produced in this experiment among the three possible combinations is intriguing. As disulfide formation occurs under thermodynamic conditions, obtaining only one isomer strongly suggest that stabilizing forces guide the folding toward the expected isomer, although the peptide segment contains only two loops.

Once the correct folding was confirmed on the C-terminal peptide fragment, the middle fragment was synthesized and folded under same conditions. As a result, the peptide with oxidized SEA and disulfide bridges was formed (**Figure 84**).

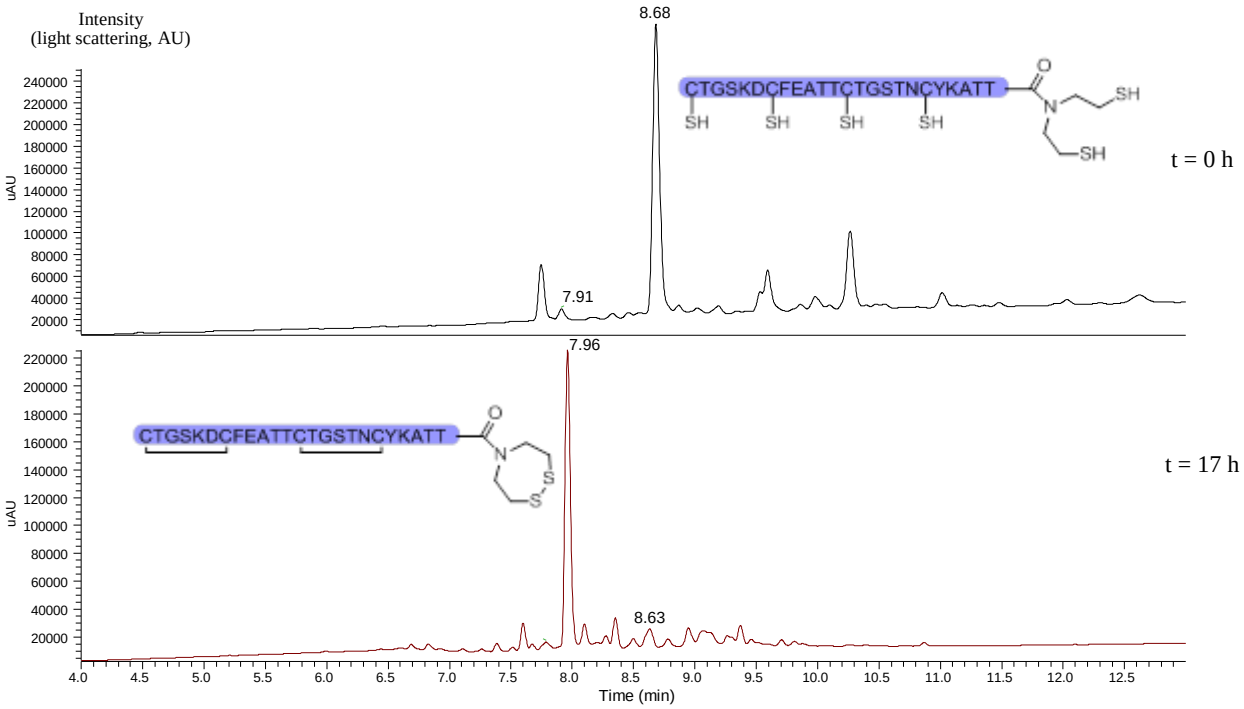


***Figure 84**. UPLC monitoring of the oxidation of the crude middle segment (t_{ret} 8.68) with the formation of two disulfide bonds and one SEA^{off} group (t_{ret} 7.96) within 17 hours.*

The chosen strategy provided the self-protected middle fragment with an overall yield of 27 %. The product is easily soluble in aqueous media compared to its S*t*Bu-protected analogue (**Figure 85**).

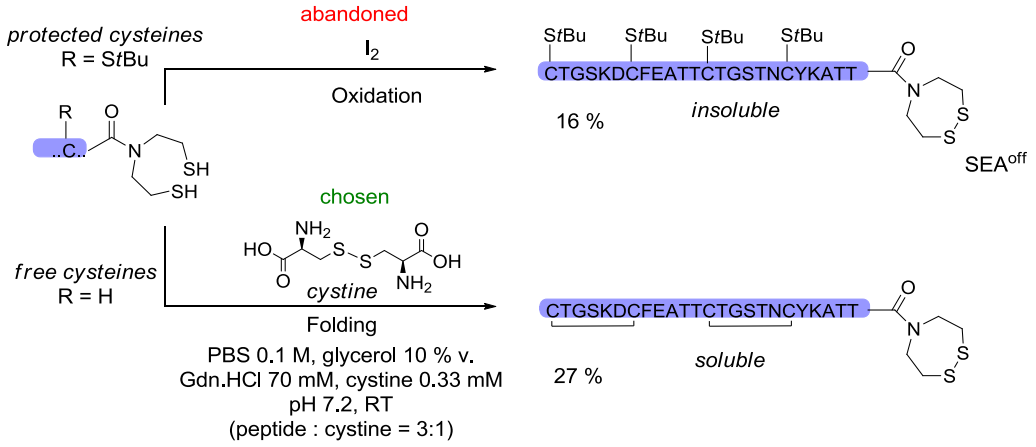


***Figure 85**. The synthesis of the middle peptide fragment by iodine oxidation (abandoned) and cystine oxidation (chosen).*

The self-protection strategy for the middle segment preparation allowed me to resolve the solubility problem previously observed with S*t*Bu protecting group. The N-ter, C-ter and middle peptide segments in hand, I tested the strategy leading to the 6 loop *Tm*AFP mimic and based on the NCL and SEA chemistries.

9.1.3.2 Assembly of the 6 loop *Tm*AFP mimic using SEA chemistry

As previously mentioned, I plan to use a one-pot, two-step process to produce the 6 loop *Tm*AFP mimic. During the first step, the *N*-ter and middle peptide fragments are connected by NCL (**Figure 86, a**). The reaction was carried out under classical NCL conditions: phosphate buffer to maintain pH 7.2, guanidinium chloride as a denaturant, *n*-octylglucoside as an additional solubilizing agent and an excess of MPAA (200 mM). MPAA has a dual role in the NCL reaction. Firstly, MPAA catalyses the NCL reaction by forming a more reactive arylthioester through thiol/thioester exchange. Secondly, MPAA is expected to reduce selectively the disulfide bridges of the middle fragments (the SEA^{off} remaining silent in these conditions) and to keep under reduced form all the cysteines present in the peptide segments during the ligation process, especially the one involved in the formation of the amide bond. Experimentally, the expected ligation product L1 containing only free cysteines was mainly detected in the reaction mixture by UPLC-MS analysis.

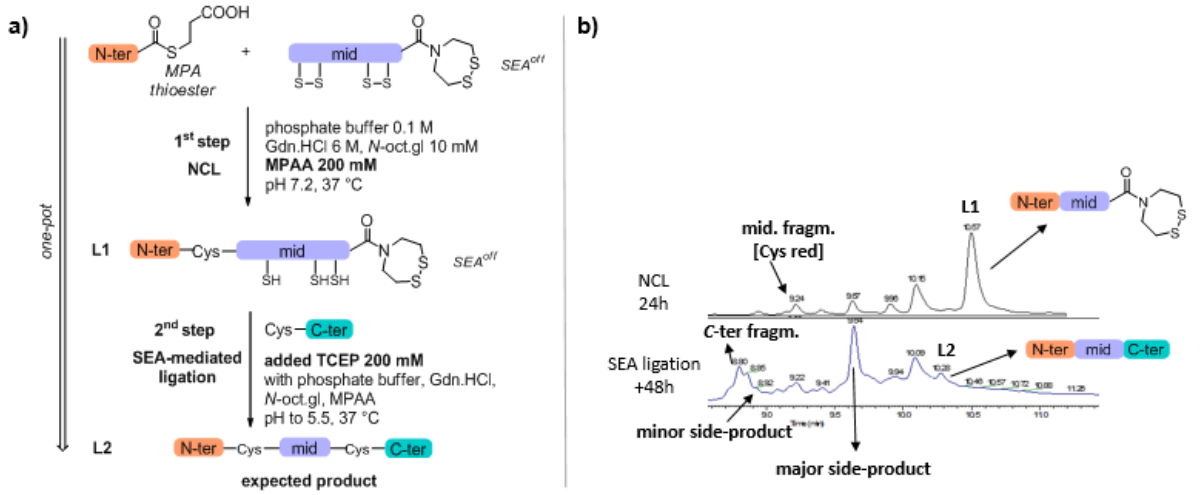


***Figure 86**. First attempt to assemble the 6 loop TmAFP mimic using SEA chemistry. a) Reaction conditions for the two-step process. b) UPLC-MS monitoring of the process.*

As the desired ligation product L1 was formed during the NCL step, it was possible to run the second step, i.e., the SEA-mediated ligation between the intermediate peptide L1 and C-ter fragment. To promote the ligation, the SEA group was activated by adding TCEP, a more powerful reducing agent than MPAA. Practically, TCEP and the C-ter peptide fragment were dissolved in the media identical to what was used for the step 1. Once added, pH is adjusted to 5.5. Note that the addition of reagents dilutes twice the reaction media, what decreases peptide concentration and decreases the ligation rate. By monitoring the progress of the second step, we observed the formation of a by-product going with the target product decomposition (**Figure 86, b**).

This main by-product was isolated by HPLC and its mass (2474.2 Da) corresponds to a cyclic peptide having the same sequence than the middle segment. To confirm the cyclic structure of the isolated compound, the latter was digested with pepsin, an enzyme that cleaves the peptide sequence after aromatic residues (Phe, Tyr, Trp). As the enzymatic cleavage at the Tyr residue gave a single product with a mass of M + 16 Da, we concluded that the peptide had a cyclic structure prior to digestion. (**Figure 87**).

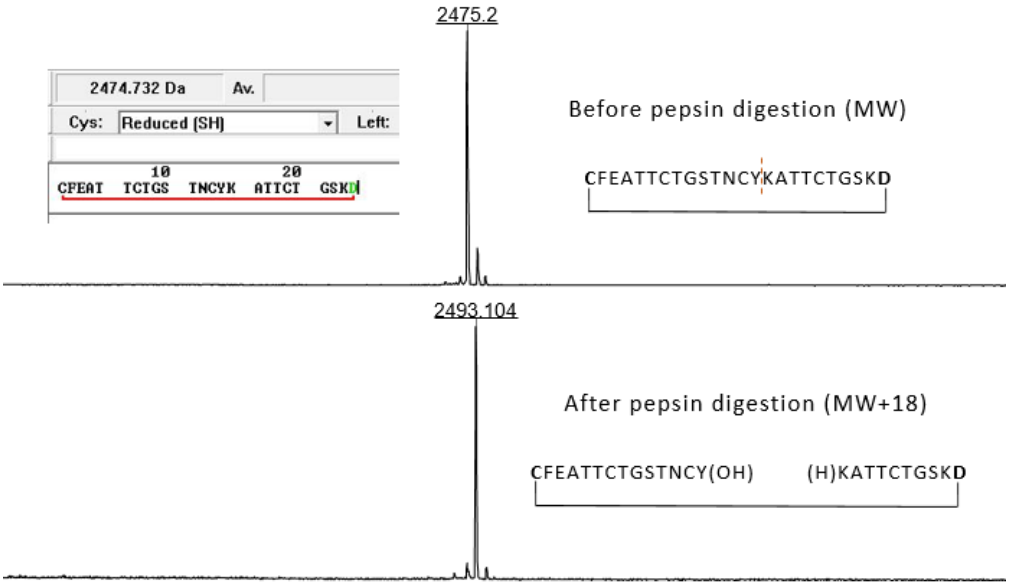


***Figure 87**. MALDI-TOF analysis of the cyclic by-product before and after pepsin digestion.*

Note that the MS/MS analysis was not informative to characterize this main by-product. This is due to the absence of well-defined N- or C-terminal extremities in the cyclic structure, what prevents the assignation of ions produced by fragmentation of the peptide chain.

The cyclic by-product detected and isolated at the end of the concatenation process couldn't result from the direct cyclization of the middle peptide segment by intramolecular SEA

***Figure 88**. Unexpected cleavage of the peptide sequence at the Asp-Cys junction explaining the formation of linear and cyclic by-products during the assembly of the 6 loop TmAFP mimic.*

ligation. Indeed, the SEA group is latent under the reductive conditions of the NCL reaction. To explain the formation of the cyclic by-product, we hypothesize that a cleavage occurs between Asp and Cys residues during the assembly of the target sequence by NCL and SEA ligations. This side reaction leads to polypeptide fragmentation at multiple sites and cyclization of the sequence (**Figure 88**).

Another side-product was detected by mass analysis in the reaction mixture. Its molecular weight corresponds to the C-terminal peptide segment truncated of six amino acids at its N-terminus (**Figure 89**). Its formation also supported the hypothesis of the peptide sequence cleavage to Asp-Cys junction (**Figure 90**).

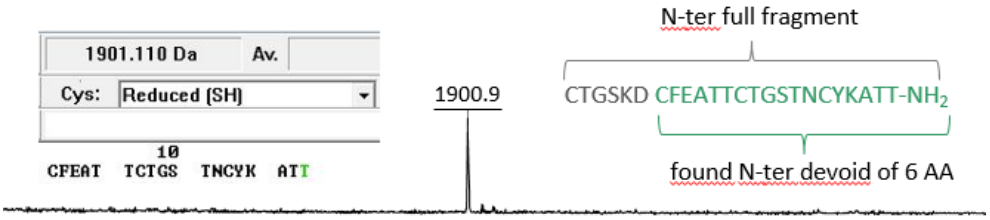


***Figure 89**. MALDI-TOF analysis of the linear by-product.*

In support of this hypothesis, the cleavage of Asp-Cys sequence assisted by intramolecular acid catalysis was described recently by Pane et al. and by Shimamoto et al. in another context [193,194]. In the field of chemical protein synthesis, this side reaction has not been described before. According to our procedure, the first ligation occurs at pH 7.2, between C-terminal MPA peptide thioester and middle (auto-protected) segment featuring a SEA^{off} on its C-terminus. At this stage, we noticed that the extent of target product formation due to Asp-Cys cleavage is limited (**Figure 86, b**). The second ligation step between SEA^{on} functionality and Cys-peptide occurs at pH 5.5 because SEA-mediated ligation proceeds optimally at mildly acidic pH. We imagine that these slightly acidic conditions favour the cysteine liberation and the cleavage of the Asp-Cys bond, as proposed in **Figure 90**.

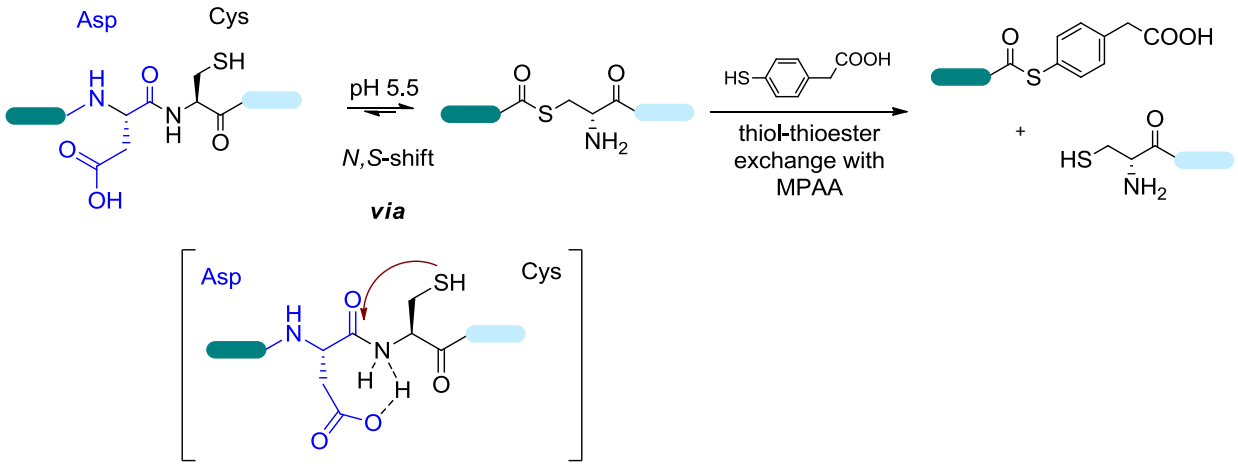


***Figure 90**. The proposed mechanism of Asp-Cys cleavage.*

The thioester formed upon the acid catalysed *N*,S-acyl shift undergoes a trans thiol-thioester exchange process with MPAA, leading to the cleavage of the peptide sequence. We suppose that the side chain of Asp plays an important role in this process, because its carboxylic acid is in close vicinity to Asp-Cys peptidic bond that can promote the amide bond weakening by internal protonation. Two cleavages must proceed to liberate a fragment having an *N*-terminal cysteine and a C-terminal thioester functionality, before it can cyclize by classical NCL.

In search for a solution to this problem we replaced Asp (D) by Glu (E) at the problematic positions. Indeed, although Glu is the homolog of Asp, the presence of one additional methylene group in its side-chain makes backbone amide protonation and thus activation by the pendent carboxylic acid less favourable. To test this hypothesis, we synthesized the appropriate N-terminal, middle and C-terminal peptide segments containing the EC instead of the DC dyad in their sequence (**Figure 91**). They were synthesized as previously described for their parent molecules by Fmoc-SPPS starting from Rink amine or SEA solid support. Note that the method we developed to protect cysteines by formation of internal disulfide bridges was successfully extended to the production of the SEA^{off} middle peptide fragment with 16% global yield.

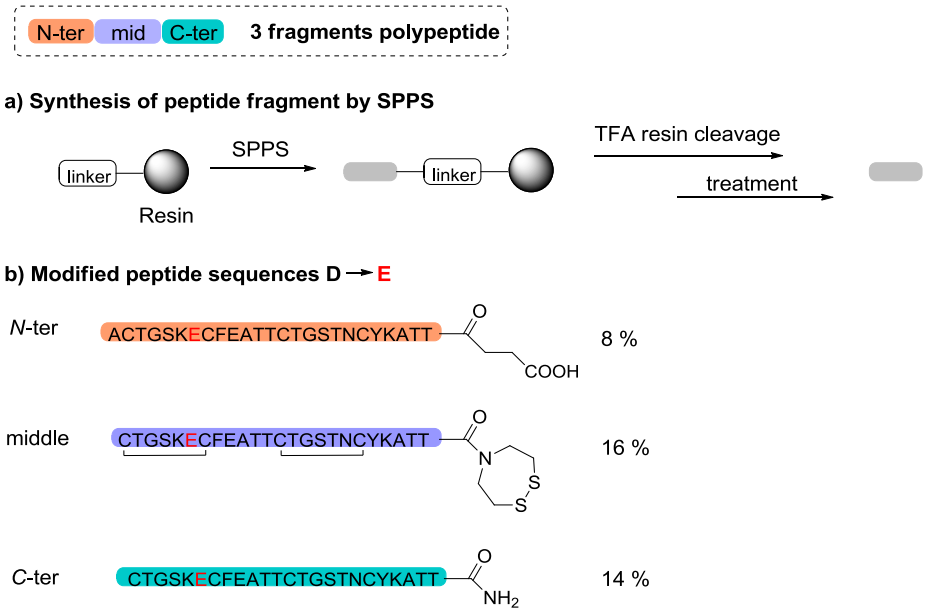


***Figure 91**. Second-generation of peptide segments designed for the assembly of the 6 loop TmAFP mimic. a) The general scheme of synthetic strategy; b) Sequences of the N-ter, middle, and C-ter peptide fragments, where DC dyads of amino acids were replaced by EC dyads of amino acids.*

This second generation of peptide fragments were then assembled following the one-pot, two-step process described in **Figure 86**. Experimentally, no undesired *N*,S-shift was observed during the peptide assembly and the modified sequence was found to be stable during the ligation. Unfortunately, once the forming peptide could accumulate in the reaction mixture

without being decomposed, we faced another problem which is the peptide precipitation during the second ligation step, despite the presence of the solubilising agent *n*-octylglucoside in the mixture (**Figure 92, a**). To remind, the concatenation of three peptide segments to obtain the six coils polypeptide is an intermediate goal towards longer polypeptides synthesis (**Figure 75**). As the precipitation occurred at the stage of six coils peptide synthesis, we can conclude that the next stage of eight coils peptide synthesis would be more challenging. It was thus necessary to anticipate the increasing tendency for precipitation.

To improve the solubility of the final polypeptide L2, we armed *N*-terminal peptide fragment with poly-Arg solubilizing tag. The latter was synthesized as described in figure 27. Transthioesterification of the SEA^{on} peptide produced by Fmoc-SPPS from a SEA solid support provided the desired peptide with 4% yield. Luckily, the insertion of poly-Arg tag in the peptide sequence prevented precipitation during the one-pot, two-step process and allowed the assembly of the target polypeptide L2 with 51% HPLC yield (**Figure 92, b**). To recover this compound as pure as possible, it was necessary to optimize HPLC conditions.

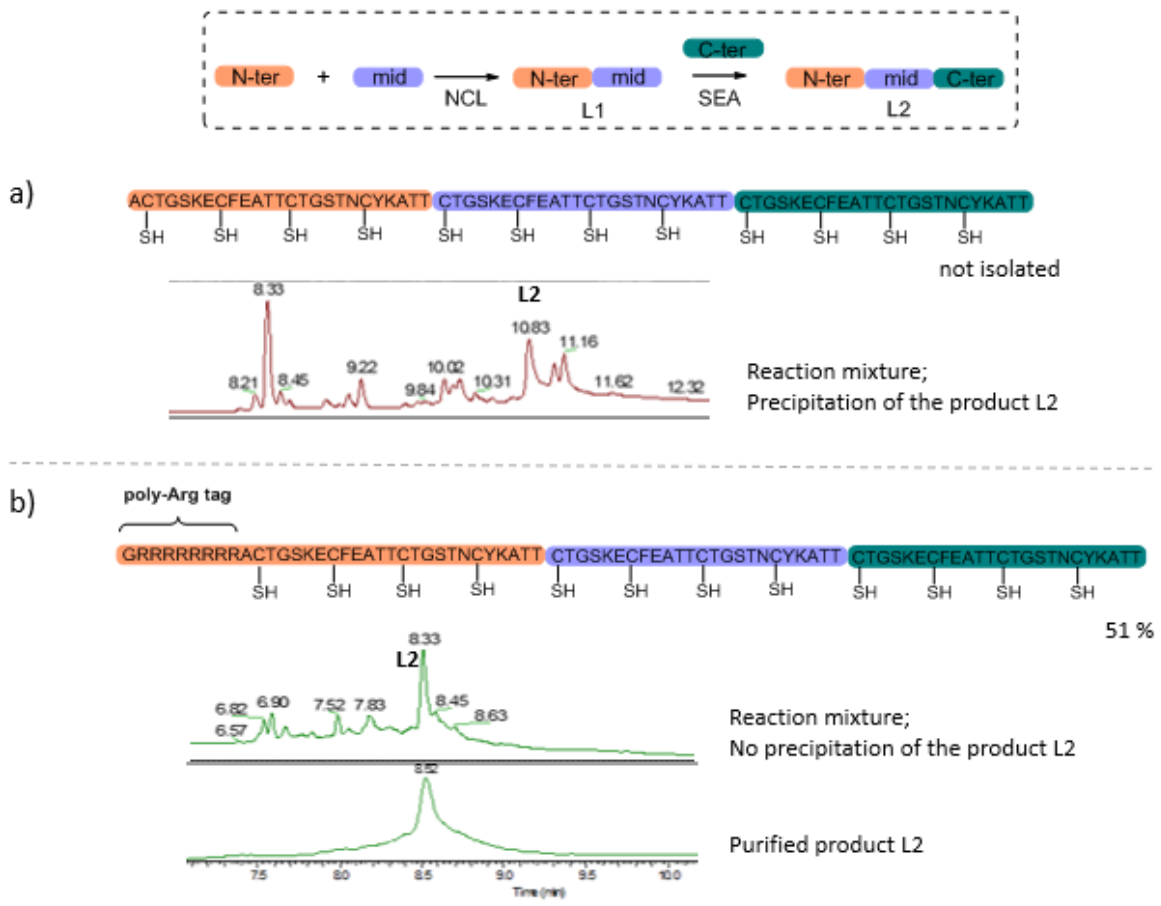


***Figure 92**. a) The reaction profile when the product precipitates; b) the profile of successfully formed product with poly-Arg solubility tag and its purification.*

9.1.4 Purification

Performing purification with classical eluent system (water / ACN with 0.1% TFA) didn't allow to obtain the polypeptide L2 with the required purity, what can be seen in **Figure 92, b**, as a peak with very wide foot. Trying to improve the HPLC purification step, we varied multiple parameters of analytical HPLC to find the best conditions. Relying on a study[195], we have varied the TFA percentage in the eluent in the range 0.1 / 0.2 / 0.3 / 0.4 percent, used more polar acid HCl as a counter-ion, replaced acetonitrile by isopropanol (IPA) or used mixed eluent composed of IPA and water in ratio 4:6. The latter solvent was the best choice to reveal some co-eluted impurities.

9.1.5 Folding

Folding is the process of formation the most thermodynamically stable structure among all possible ones. In case of Cys-rich proteins like *Tm*AFP, the number of possible combinations grows with cysteines number increase, what makes the folding process longer and more complex. This is why, to avoid additional system complications coming from impurities, it is very important to purify the polypeptide well before the folding (**Figure 93**). Unlikely the peptide with 4 formed disulphide bridges, this longer polypeptide has not been folded properly yet. Following the folding of the product L2 by HPLC, polypeptide with impurities gives no signals with good resolution. In addition, the folding will result 6 coils formed by 12 cysteines. It is a complex system with multiple transition folding states, and impurities must not be present to not interfere the interpretation of the folding process.

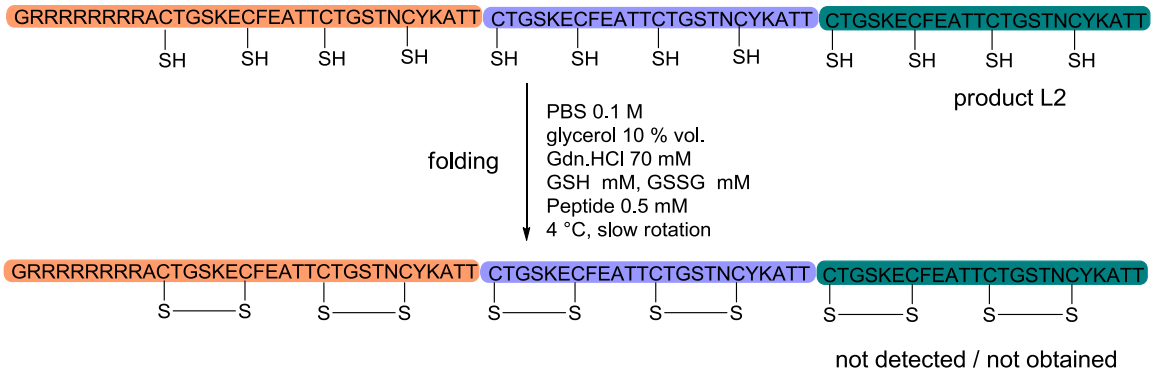


***Figure 93**. Expected product of folding with six formed disulfide bridges that was not detected by UPLC-MS monitoring.*

As it was discussed in bibliography part, every amino acid in a sequence brings its contribution to peptide properties due to its side chain bulkiness, electronic properties, etc. Thus, the antifreeze properties of the designed protein are in certain dependence on the polypeptide sequence. In its turn, every new modification of the sequence demands new peptide

synthesis, what is time consuming in total. Unfortunately, due to the limitation of time given for the thesis development, it was necessary to define a priority line of the project, what was the development of SetCys chemistry. This is why the study of *Tm*AFP analogue will be delegated to further researchers, and I present here the perspective of future work.

9.2 Conclusion and perspectives

During my experimental part of the thesis I managed to make few steps towards mastering of a *Tm*AFP analogue chemistry. I obtained the polypeptide non-folded precursor corresponding to the six coils structure. It became possible after the insertion of several modifications. First, I improved the solubility of the middle short SEA containing peptide fragment by replacing S*t*Bu protecting groups on Cys by Trt, while keeping one-pot ligation selective due to redox control. Second, the unexpected instability of the target sequence was removed by insertion of a modification in the sequence. The aspartic acids preceding cysteine residues caused undesired cleavage and peptide cyclization. Replacing the problematic aspartic acid residues by glutamic ones solved the problem, probably due to the greater distance between the side-chain carboxylic acid and the scissile peptide bond. Finally, increasing the solubility of the target sequence was compulsory. For that, the AFP polypeptide was extended on its N-terminal side with a poly-arginine solubilizing tag.

The next stage is the folding, which still must be performed with success. Apart from the purity issue, we plan to improve the folding procedure by variation of temperature, pH and peptide concentration. Classically, low temperatures (0 – 10 °C) are applied to minimize polypeptide precipitation. Thus, the temperature might be kept in this range to reach better folding yields. Eventually, the folding might be conducted with some ice which could serve as a template for facilitating the correct formation of the coiled structure. The peptide concentration is also very important as effective dilution favours intramolecular disulfide bond formation, while too concentrated peptide might form intermolecular disulfide bridges.

As it was discussed in main goals section, the terminal perspective of *Tm*AFP analogue to design and to synthesis an active protein that would improve the viability of *Schistosoma mansoni* parasite stages after being frozen. The synthesised and correctly folded molecules will be first tested for their TH and IRI capacity in the laboratory facilities, and the active samples will be incubated with *Tm* parasite stages.

In my opinion, the cryopreservation studies may be performed in non-linear mode. It means that the designed final *Tm*AFP analogues structures, obtained by recombinant methods, can be ordered elsewhere and tested with *Sm* in parallel with chemical protein synthesis. Having in hand such a preliminary data, it would be, from one hand, more purposeful to keep

synthesizing a molecule with confirmed activity, and from another hand, it is interesting to compare the activity of the same molecule obtained by chemical and biological pathways. The latter reason is related to the possible TH modification due to biological impurities potentially present in an overexpressed protein.

References

(1) Joosten, R. X-Ray Structure Re-Refinement. Combining Old Data with New Methods for Better Structural Bioinformatics. *Electron. Notes Theor. Comput. Sci. - ENTCS* **2010**.

(2) Goesmann, F.; Rosenbauer, H.; Bredehöft, J. H.; Cabane, M.; Ehrenfreund, P.; Gautier, T.; Giri, C.; Krüger, H.; Le Roy, L.; MacDermott, A. J.; McKenna-Lawlor, S.; Meierhenrich, U. J.; Caro, G. M. M.; Raulin, F.; Roll, R.; Steele, A.; Steininger, H.; Sternberg, R.; Szopa, C.; Thiemann, W.; Ulamec, S. Organic Compounds on Comet 67P/Churyumov-Gerasimenko Revealed by COSAC Mass Spectrometry. *Science* **2015**, *349* (6247), aab0689. https://doi.org/10.1126/science.aab0689.

(3) Herbst, E.; van Dishoeck, E. F. Complex Organic Interstellar Molecules. *Annu. Rev. Astron. Astrophys.* **2009**, *47* (1), 427–480. https://doi.org/10.1146/annurev-astro-082708-101654.

(4) Jiménez-Serra, I.; Martín-Pintado, J.; Rivilla, V. M.; Rodríguez-Almeida, L.; Alonso Alonso, E. R.; Zeng, S.; Cocinero, E. J.; Martín, S.; Requena-Torres, M.; Martín-Domenech, R.; Testi, L. Toward the RNA-World in the Interstellar Medium—Detection of Urea and Search of 2-Amino-Oxazole and Simple Sugars. *Astrobiology* **2020**, *20* (9), 1048–1066. https://doi.org/10.1089/ast.2019.2125.

(5) Holford, M.; Daly, M.; King, G. F.; Norton, R. S. Venoms to the Rescue. *Science* **2018**, *361* (6405), 842–844. https://doi.org/10.1126/science.aau7761.

(6) Muttenthaler, M.; King, G. F.; Adams, D. J.; Alewood, P. F. Trends in Peptide Drug Discovery. *Nat. Rev. Drug Discov.* **2021**, *20* (4), 309–325. https://doi.org/10.1038/s41573-020-00135-8.

(7) Koniev, O.; Wagner, A. Developments and Recent Advancements in the Field of Endogenous Amino Acid Selective Bond Forming Reactions for Bioconjugation. *Chem. Soc. Rev.* **2015**, *44* (15), 5495–5551. https://doi.org/10.1039/C5CS00048C.

(8) Canne, L. E.; Botti, P.; Simon, R. J.; Chen, Y.; Dennis, E. A.; Kent, S. B. H. Chemical Protein Synthesis by Solid Phase Ligation of Unprotected Peptide Segments. *J. Am. Chem. Soc.* **1999**, *121* (38), 8720–8727. https://doi.org/10.1021/ja9836287.

(9) Xu, W.; Jiang, W.; Wang, J.; Yu, L.; Chen, J.; Liu, X.; Liu, L.; Zhu, T. F. Total Chemical Synthesis of a Thermostable Enzyme Capable of Polymerase Chain Reaction. *Cell Discov.* **2017**, *3* (1), 1–10. https://doi.org/10.1038/celldisc.2017.8.

(10) Xu, L.; Huang, J.-F.; Chen, C.-C.; Qu, Q.; Shi, J.; Pan, M.; Li, Y.-M. Chemical Synthesis of Natural Polyubiquitin Chains through Auxiliary-Mediated Ligation of an Expressed Ubiquitin Isomer. *Org. Lett.* **2018**, *20* (2), 329–332. https://doi.org/10.1021/acs.orglett.7b03515.

(11) Schnölzer, M.; Kent, S. B. Constructing Proteins by Dovetailing Unprotected Synthetic Peptides: Backbone-Engineered HIV Protease. *Science* **1992**, *256* (5054), 221–225. https://doi.org/10.1126/science.1566069.

(12) Edelmira Cabezas, Meng Wang, Paul W. H. I. Parren, Robyn L. Stanfield, ‖ and; Arnold C. Satterthwait. *A Structure-Based Approach to a Synthetic Vaccine for HIV-1†*. ACS Publications. https://doi.org/10.1021/bi0003691.

(13) Shao, J.; Tam, J. P. Unprotected Peptides as Building Blocks for the Synthesis of Peptide Dendrimers with Oxime, Hydrazone, and Thiazolidine Linkages. *J. Am. Chem. Soc.* **1995**, *117* (14), 3893–3899. https://doi.org/10.1021/ja00119a001.

(14) Dirksen, A.; Dirksen, S.; Hackeng, T. M.; Dawson, P. E. Nucleophilic Catalysis of Hydrazone Formation and Transimination: Implications for Dynamic Covalent Chemistry. *J. Am. Chem. Soc.* **2006**, *128* (49), 15602–15603. https://doi.org/10.1021/ja067189k.

(15) Valverde, I.; Lecaille, F.; Lalmanach, G.; Aucagne, V.; Delmas, A. Synthesis of a Biologically Active Triazole-Containing Analogue of Cystatin A Through Successive Peptidomimetic Alkyne-Azide Ligations. *Angew. Chem. Int. Ed Engl.* **2012**, *51*, 718–722. https://doi.org/10.1002/anie.201107222.

(16) Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoët, M.; Monbaliu, J.-C. M.; Melnyk, O. Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* **2019**, *119* (12), 7328–7443. https://doi.org/10.1021/acs.chemrev.8b00712.

(17) Bode, J. W.; Fox, R. M.; Baucom, K. D. Chemoselective Amide Ligations by Decarboxylative Condensations of N-Alkylhydroxylamines and α-Ketoacids. *Angew. Chem. Int. Ed.* **2006**, *45* (8), 1248–1252. https://doi.org/10.1002/anie.200503991.

(18) Saxon, E.; Armstrong, J. I.; Bertozzi, C. R. A "Traceless" Staudinger Ligation for the Chemoselective Synthesis of Amide Bonds. *Org. Lett.* **2000**, *2* (14), 2141–2143. https://doi.org/10.1021/ol006054v.
(19) Liu, C. F.; Tam, J. P. Peptide Segment Ligation Strategy without Use of Protecting Groups. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (14), 6584–6588. https://doi.org/10.1073/pnas.91.14.6584.
(20) Zhang, Y.; Xu, C.; Lam, H. Y.; Lee, C. L.; Li, X. Protein Chemical Synthesis by Serine and Threonine Ligation. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (17), 6657–6662. https://doi.org/10.1073/pnas.1221012110.
(21) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266* (5186), 776–779. https://doi.org/10.1126/science.7973629.
(22) Pusterla, I.; Bode, J. W. The Mechanism of the α-Ketoacid-Hydroxylamine Amide-Forming Ligation. *Angew. Chem. Int. Ed Engl.* **2012**, *51* (2), 513–516. https://doi.org/10.1002/anie.201107198.
(23) Staudinger, H.; Meyer, J. Über neue organische Phosphorverbindungen III. Phosphinmethylenderivate und Phosphinimine. *Helv. Chim. Acta* **1919**, *2* (1), 635–646. https://doi.org/10.1002/hlca.19190020164.
(24) Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. Staudinger Ligation: A Peptide from a Thioester and Azide. *Org. Lett.* **2000**, *2* (13), 1939–1941. https://doi.org/10.1021/ol0060174.
(25) Tam, J. P.; Miao, Z. Stereospecific Pseudoproline Ligation of N-Terminal Serine, Threonine, or Cysteine-Containing Unprotected Peptides. *J. Am. Chem. Soc.* **1999**, *121* (39), 9013–9022. https://doi.org/10.1021/ja991153t.
(26) Tam, J. P.; Rao, C.; Liu, C. F.; Shao, J. Specificity and Formation of Unusual Amino Acids of an Amide Ligation Strategy for Unprotected Peptides. *Int. J. Pept. Protein Res.* **1995**, *45* (3), 209–216. https://doi.org/10.1111/j.1399-3011.1995.tb01482.x.
(27) Agouridas, V.; El Mahdi, O.; Cargoët, M.; Melnyk, O. A Statistical View of Protein Chemical Synthesis Using NCL and Extended Methodologies. *Bioorg. Med. Chem.* **2017**, *25* (18), 4938–4945. https://doi.org/10.1016/j.bmc.2017.05.050.
(28) Wieland, T.; Bokelmann, E.; Bauer, L.; Lang, H. U.; Lau, H. Über Peptidsynthesen. 8. Mitteilung Bildung von S-Haltigen Peptiden Durch Intramolekulare Wanderung von Aminoacylresten. *Justus Liebigs Ann. Chem.* **1953**, *583* (1), 129–149. https://doi.org/10.1002/jlac.19535830110.
(29) Brenner, M.; Zimmermann, J. P.; Wehrmüller, J.; Quitt, P.; Photaki, I. Eine neue Umlagerungsreaktion und ein neues Prinzip zum Aufbau von Peptidketten. *Experientia* **1955**, *11* (10), 397–399. https://doi.org/10.1007/BF02158504.
(30) *The Protein Chemical Synthesis Database - An interactive and educational tool for looking at chemically synthesized proteins (Find us on twitter @PcsDb)*. The Protein Chemical Synthesis Database. http://pcs-db.fr/ (accessed 2023-03-26).
(31) Poole, L. B. The Basics of Thiols and Cysteines in Redox Biology and Chemistry. *Free Radic. Biol. Med.* **2015**, *80*, 148–157. https://doi.org/10.1016/j.freeradbiomed.2014.11.013.
(32) Maguire, O. R.; Zhu, J.; Brittain, W. D. G.; Hudson, A. S.; Cobb, S. L.; O'Donoghue, A. C. N-Terminal Speciation for Native Chemical Ligation. *Chem. Commun.* **2020**, *56* (45), 6114–6117. https://doi.org/10.1039/D0CC01604G.
(33) Gieselman, M. D.; Xie, L.; van Der Donk, W. A. Synthesis of a Selenocysteine-Containing Peptide by Native Chemical Ligation. *Org. Lett.* **2001**, *3* (9), 1331–1334. https://doi.org/10.1021/ol015712o.
(34) Quaderer, R.; Sewing, A.; Hilvert, D. Selenocysteine-Mediated Native Chemical Ligation. *Helv. Chim. Acta* **2001**, *84* (5), 1197–1206. https://doi.org/10.1002/1522-2675(20010516).
(35) *Selenocysteine in Native Chemical Ligation and Expressed Protein Ligation | Journal of the American Chemical Society*. https://pubs.acs.org/doi/10.1021/ja005885t (accessed 2023-03-26).
(36) Reich, H. J.; Hondal, R. J. Why Nature Chose Selenium. *ACS Chem. Biol.* **2016**, *11* (4), 821–841. https://doi.org/10.1021/acschembio.6b00031.
(37) Moura, A.; Savageau, M. A.; Alves, R. Relative Amino Acid Composition Signatures of Organisms and Environments. *PLoS ONE* **2013**, *8* (10), e77319. https://doi.org/10.1371/journal.pone.0077319.
(38) Miseta, A.; Csutora, P. Relationship Between the Occurrence of Cysteine in Proteins and the Complexity of Organisms. *Mol. Biol. Evol.* **2000**, *17* (8), 1232–1239. https://doi.org/10.1093/oxfordjournals.molbev.a026406.

(39) Wan, Q.; Danishefsky, S. J. Free-Radical-Based, Specific Desulfurization of Cysteine: A Powerful Advance in the Synthesis of Polypeptides and Glycopolypeptides. *Angew. Chem. Int. Ed Engl.* **2007**, *46* (48), 9248–9252. https://doi.org/10.1002/anie.200704195.
(40) Haase, C.; Rohde, H.; Seitz, O. Native Chemical Ligation at Valine. *Angew. Chem. Int. Ed.* **2008**, *47* (36), 6807–6810. https://doi.org/10.1002/anie.200801590.
(41) Cappadocia, L.; Lima, C. D. Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. *Chem. Rev.* **2018**, *118* (3), 889–918. https://doi.org/10.1021/acs.chemrev.6b00737.
(42) Lynen, F.; Reichert, E. Zur Chemischen Struktur Der „aktivierten Essigsäure". *Angew. Chem.* **1951**, *63* (2), 47–48. https://doi.org/10.1002/ange.19510630207.
(43) Mills, K. V.; Johnson, M. A.; Perler, F. B. Protein Splicing: How Inteins Escape from Precursor Proteins. *J. Biol. Chem.* **2014**, *289* (21), 14498–14505. https://doi.org/10.1074/jbc.R113.540310.
(44) Della Védova, C. O.; Romano, R. M.; Oberhammer, H. Gas Electron Diffraction Analysis on S-Methyl Thioacetate, CH3C(O)SCH3. *J. Org. Chem.* **2004**, *69* (16), 5395–5398. https://doi.org/10.1021/jo0493828.
(45) Kitano, M.; Kuchitsu, K. Molecular Structure of *N* -Methylformamide as Studied by Gas Electron Diffraction. *Bull. Chem. Soc. Jpn.* **1974**, *47* (3), 631–634. https://doi.org/10.1246/bcsj.47.631.
(46) Pollock, S. B.; Kent, S. B. H. An Investigation into the Origin of the Dramatically Reduced Reactivity of Peptide-Prolyl-Thioesters in Native Chemical Ligation. *Chem. Commun.* **2011**, *47* (8), 2342–2344. https://doi.org/10.1039/C0CC04120C.
(47) Raibaut, L.; Seeberger, P.; Melnyk, O. Bis(2-Sulfanylethyl)Amido Peptides Enable Native Chemical Ligation at Proline and Minimize Deletion Side-Product Formation. *Org. Lett.* **2013**, *15* (21), 5516–5519. https://doi.org/10.1021/ol402678a.
(48) Jencks, W. P.; Cordes, S.; Carriuolo, J. The Free Energy of Thiol Ester Hydrolysis. *J. Biol. Chem.* **1960**, *235* (12), 3608–3614. https://doi.org/10.1016/S0021-9258(18)64517-X.
(49) Bracher, P. J.; Snyder, P. W.; Bohall, B. R.; Whitesides, G. M. The Relative Rates of Thiol-Thioester Exchange and Hydrolysis for Alkyl and Aryl Thioalkanoates in Water. *Orig. Life Evol. Biosphere J. Int. Soc. Study Orig. Life* **2011**, *41* (5), 399–412. https://doi.org/10.1007/s11084-011-9243-4.
(50) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154. https://doi.org/10.1021/ja00897a025.
(51) Merrifield, R. B. Automated Synthesis of Peptides. *Science* **1965**, *150* (3693), 178–185. https://doi.org/10.1126/science.150.3693.178.
(52) Hojo, H.; Aimoto, S. 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. *Bull. Chem. Soc. Jpn.* **1991**, *64* (1), 111–117. https://doi.org/10.1246/bcsj.64.111.
(53) Kitagawa, K.; Adachi, H.; Sekigawa, Y.; Yagami, T.; Futaki, S.; Gu, Y. J.; Inoue, K. Total Chemical Synthesis of Large CCK Isoforms Using a Thioester Segment Condensation Approach. *Tetrahedron* **2004**, *60* (4), 907–918. https://doi.org/10.1016/j.tet.2003.11.037.
(54) von Eggelkraut-Gottanka, R.; Klose, A.; Beck-Sickinger, A. G.; Beyermann, M. Peptide Athioester Formation Using Standard Fmoc-Chemistry. *Tetrahedron Lett.* **2003**, *44* (17), 3551–3554. https://doi.org/10.1016/S0040-4039(03)00582-3.
(55) Alsina, J.; Kates, S. A.; Barany, G.; Albericio, F. Backbone Amide Linker Strategies for the Solid-Phase Synthesis of C-Terminal Modified Peptides. *Methods Mol. Biol. Clifton NJ* **2005**, *298*, 195–208. https://doi.org/10.1385/1-59259-877-3:195.
(56) Kenner, G. W.; McDermott, J. R.; Sheppard, R. C. The Safety Catch Principle in Solid Phase Peptide Synthesis. *J. Chem. Soc. Chem. Commun.* **1971**, No. 12, 636–637. https://doi.org/10.1039/C29710000636.
(57) Heidler, P.; Link, A. N-Acyl-N-Alkyl-Sulfonamide Anchors Derived from Kenner's Safety-Catch Linker: Powerful Tools in Bioorganic and Medicinal Chemistry. *Bioorg. Med. Chem.* **2005**, *13* (3), 585–599. https://doi.org/10.1016/j.bmc.2004.10.045.
(58) *Simplifying native chemical ligation with an N-acylsulfonamide linker - Chemical Communications (RSC Publishing)*. https://pubs.rsc.org/en/content/articlelanding/2012/cc/c2cc15911b (accessed 2023-04-08).

(59) Blanco-Canosa, J. B.; Dawson, P. E. An Efficient Fmoc-SPPS Approach for the Generation of Thioester Peptide Precursors for Use in Native Chemical Ligation. *Angew. Chem. Int. Ed Engl.* **2008**, *47* (36), 6851–6855. https://doi.org/10.1002/anie.200705471.
(60) *A Reversible Protection Strategy To Improve Fmoc-SPPS of Peptide Thioesters by the N-Acylurea Approach - Mahto - 2011 - ChemBioChem - Wiley Online Library*. https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/cbic.201100472 (accessed 2023-04-09).
(61) *Chemical Protein Synthesis Using a Second-Generation N-Acylurea Linker for the Preparation of Peptide-Thioester Precursors | Journal of the American Chemical Society*. https://pubs.acs.org/doi/10.1021/jacs.5b03504 (accessed 2023-04-09).
(62) Blanco-Canosa, J. B.; Nardone, B.; Albericio, F.; Dawson, P. E. Chemical Protein Synthesis Using a Second-Generation N-Acylurea Linker for the Preparation of Peptide-Thioester Precursors. *J. Am. Chem. Soc.* **2015**, *137* (22), 7197–7209. https://doi.org/10.1021/jacs.5b03504.
(63) *Protein chemical synthesis by ligation of peptide hydrazides - PubMed*. https://pubmed.ncbi.nlm.nih.gov/21648030/ (accessed 2023-04-08).
(64) Camarero, J. A.; Hackel, B. J.; de Yoreo, J. J.; Mitchell, A. R. Fmoc-Based Synthesis of Peptide Alpha-Thioesters Using an Aryl Hydrazine Support. *J. Org. Chem.* **2004**, *69* (12), 4145–4151. https://doi.org/10.1021/jo040140h.
(65) Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L. Chemical Synthesis of Proteins Using Peptide Hydrazides as Thioester Surrogates. *Nat. Protoc.* **2013**, *8* (12), 2483–2495. https://doi.org/10.1038/nprot.2013.152.
(66) Burlina, F.; Morris, C.; Behrendt, R.; White, P.; Offer, J. Simplifying Native Chemical Ligation with an N-Acylsulfonamide Linker. *Chem. Commun.* **2012**, *48* (20), 2579–2581. https://doi.org/10.1039/C2CC15911B.
(67) Fang, G.-M.; Li, Y.-M.; Shen, F.; Huang, Y.-C.; Li, J.-B.; Lin, Y.; Cui, H.-K.; Liu, L. Protein Chemical Synthesis by Ligation of Peptide Hydrazides. *Angew. Chem. Int. Ed Engl.* **2011**, *50* (33), 7645–7649. https://doi.org/10.1002/anie.201100996.
(68) Flood, D. T.; Hintzen, J. C. J.; Bird, M. J.; Cistrone, P. A.; Chen, J. S.; Dawson, P. E. Leveraging the Knorr Pyrazole Synthesis for the Facile Generation of Thioester Surrogates for Use in Native Chemical Ligation. *Angew. Chem. Int. Ed.* **2018**, *57* (36), 11634–11639. https://doi.org/10.1002/anie.201805191.
(69) Nagaike, F.; Onuma, Y.; Kanazawa, C.; Hojo, H.; Ueki, A.; Nakahara, Y.; Nakahara, Y. Efficient Microwave-Assisted Tandem N- to S-Acyl Transfer and Thioester Exchange for the Preparation of a Glycosylated Peptide Thioester. *Org. Lett.* **2006**, *8* (20), 4465–4468. https://doi.org/10.1021/ol0616034.
(70) Sharma, R. K.; Tam, J. P. Tandem Thiol Switch Synthesis of Peptide Thioesters via N-S Acyl Shift on Thiazolidine. *Org. Lett.* **2011**, *13* (19), 5176–5179. https://doi.org/10.1021/ol202047q.
(71) Wierzbicka, M.; Waliczek, M.; Dziadecka, A.; Stefanowicz, P. One-Pot Cyclization and Cleavage of Peptides with N-Terminal Cysteine via the N,S-Acyl Shift of the N-2-[Thioethyl]Glycine Residue. *J. Org. Chem.* **2021**, *86* (17), 12292–12299. https://doi.org/10.1021/acs.joc.1c01045.
(72) Ollivier, N.; Behr, J.-B.; El-Mahdi, O.; Blanpain, A.; Melnyk, O. Fmoc Solid-Phase Synthesis of Peptide Thioesters Using an Intramolecularn,S-Acyl Shift. *Org. Lett.* **2005**, *7* (13), 2647–2650. https://doi.org/10.1021/ol050776a.
(73) Hojo, H.; Onuma, Y.; Akimoto, Y.; Nakahara, Y.; Nakahara, Y. N-Alkyl Cysteine-Assisted Thioesterification of Peptides. *Tetrahedron Lett.* **2007**, *48* (1), 25–28. https://doi.org/10.1016/j.tetlet.2006.11.034.
(74) Dheur, J.; Ollivier, N.; Vallin, A.; Melnyk, O. Synthesis of Peptide Alkylthioesters Using the Intramolecular N,S-Acyl Shift Properties of Bis(2-Sulfanylethyl)Amido Peptides. *J. Org. Chem.* **2011**, *76* (9), 3194–3202. https://doi.org/10.1021/jo200029e.
(75) Tsuda, S.; Shigenaga, A.; Bando, K.; Otaka, A. N-->S Acyl-Transfer-Mediated Synthesis of Peptide Thioesters Using Anilide Derivatives. *Org. Lett.* **2009**, *11* (4), 823–826. https://doi.org/10.1021/ol8028093.
(76) Ollivier, N.; Dheur, J.; Mhidia, R.; Blanpain, A.; Melnyk, O. Bis(2-Sulfanylethyl)Amino Native Peptide Ligation. *Org. Lett.* **2010**, *12* (22), 5238–5241. https://doi.org/10.1021/ol102273u.

(77) Diemer, V.; Firstova, O.; Agouridas, V.; Melnyk, O. Pedal to the Metal: The Homogeneous Catalysis of the Native Chemical Ligation Reaction. *Chem. – Eur. J.* **2022**, *28* (16), e202104229. https://doi.org/10.1002/chem.202104229.
(78) Rohde, H.; Schmalisch, J.; Harpaz, Z.; Diezmann, F.; Seitz, O. Ascorbate as an Alternative to Thiol Additives in Native Chemical Ligation. *ChemBioChem* **2011**, *12* (9), 1396–1400. https://doi.org/10.1002/cbic.201100179.
(79) Sakamoto, K.; Tsuda, S.; Mochizuki, M.; Nohara, Y.; Nishio, H.; Yoshiya, T. Imidazole-Aided Native Chemical Ligation: Imidazole as a One-Pot Desulfurization-Amenable Non-Thiol-Type Alternative to 4-Mercaptophenylacetic Acid. *Chem. - Eur. J.* **2016**, *22* (50), 17940–17944. https://doi.org/10.1002/chem.201604320.
(80) Fife, T., H.; DeMark, B., R. *General-base-catalyzed intramolecular aminolysis of thiol esters. Cyclization of S-n-propyl o-(2-imidazolyl)thiolbenzoate. Relationship of the uncatalyzed and base-catalyzed nucleophilic reactions | Journal of the American Chemical Society*. https://pubs.acs.org/doi/10.1021/ja00518a041 (accessed 2023-04-05).
(81) Bruice, T. C. *Imidazole Catalysis. VI.1 The Intramolecular Nucleophilic Catalysis of the Hydrolysis of an Acyl Thiol. The Hydrolysis of n-Propyl γ-(4-Imidazolyl)-thiolbutyrate*. ACS Publications. https://doi.org/10.1021/ja01529a048.
(82) Loibl, S. F.; Harpaz, Z.; Zitterbart, R.; Seitz, O. Total Chemical Synthesis of Proteins without HPLC Purification. Electronic Supplementary Information (ESI) Available. See DOI: 10.1039/C6sc01883a Click Here for Additional Data File. *Chem. Sci.* **2016**, *7* (11), 6753–6759. https://doi.org/10.1039/c6sc01883a.
(83) Bang, D.; Pentelute, B. L.; Kent, S. B. H. Kinetically Controlled Ligation for the Convergent Chemical Synthesis of Proteins. *Angew. Chem. Int. Ed.* **2006**, *45* (24), 3985–3988. https://doi.org/10.1002/anie.200600702.
(84) Ollivier, N.; Vicogne, J.; Vallin, A.; Drobecq, H.; Desmet, R.; El Mahdi, O.; Leclercq, B.; Goormachtigh, G.; Fafeur, V.; Melnyk, O. A One-Pot Three-Segment Ligation Strategy for Protein Chemical Synthesis. *Angew. Chem. Int. Ed.* **2012**, *51* (1), 209–213. https://doi.org/10.1002/anie.201105837.
(85) Sun, Z.: Wei, T.; Cao, Y.; Li, X. *Protocol for semisynthesis of serotonylated histone H3 by rapid protein desulfurization in tandem with native chemical ligation | Elsevier Enhanced Reader*. https://doi.org/10.1016/j.xpro.2022.102042.
(86) Raibaut, L.; Drobecq, H.; Melnyk, O. Selectively Activatable Latent Thiol and Selenolesters Simplify the Access to Cyclic or Branched Peptide Scaffolds. *Org. Lett.* **2015**, *17* (14), 3636–3639. https://doi.org/10.1021/acs.orglett.5b01817.
(87) Raibaut, L.; Cargoët, M.; Ollivier, N.; Chang, Y. M.; Drobecq, H.; Boll, E.; Desmet, R.; Monbaliu, J.-C. M.; Melnyk, O. Accelerating Chemoselective Peptide Bond Formation Using Bis(2-Selenylethyl)Amido Peptide Selenoester Surrogates. *Chem. Sci.* **2016**, *7* (4), 2657–2665. https://doi.org/10.1039/C5SC03459K.
(88) Snella, B.; Grain, B.; Vicogne, J.; Capet, F.; Wiltschi, B.; Melnyk, O.; Agouridas, V. Fast Protein Modification in the Nanomolar Concentration Range Using an Oxalyl Amide as Latent Thioester**. *Angew. Chem. Int. Ed.* **2022**, *61* (29). https://doi.org/10.1002/anie.202204992.
(89) Tesser, G. I.; Balvert-Geers, I. C. The Methylsulfonylethyloxycarbonyl Group, a New and Versatile Amino Protective Function. *Int. J. Pept. Protein Res.* **1975**, *7* (4), 295–305. https://doi.org/10.1111/j.1399-3011.1975.tb02444.x.
(90) Veber, D., F.; Milkowski, J., D.; Larga, S., L.; Denkewalter, R., G.; Hirschmann, L. *Acetamidomethyl. A novel thiol protecting group for cysteine - PubMed*. https://pubmed.ncbi.nlm.nih.gov/5040849/ (accessed 2023-04-11).
(91) Duhee Bang; Neeraj Chopra; Stephen B. H. Kent. *Total Chemical Synthesis of Crambin*. ACS Publications. https://doi.org/10.1021/ja0385078.
(92) Ueda, S.; Fujita, M.; Tamamura, H.; Fujii, N.; Otaka, A. Photolabile Protection for One-Pot Sequential Native Chemical Ligation. *ChemBioChem* **2005**, *6* (11), 1983–1986. https://doi.org/10.1002/cbic.200500272.
(93) Wintermann, F.; Engelbrecht, S. Reconstitution of the Catalytic Core of F-ATPase (Aβ)3γ from Escherichia Coli Using Chemically Synthesized Subunit γ. *Angew. Chem. Int. Ed.* **2013**, *52* (4), 1309–1313. https://doi.org/10.1002/anie.201206744.

(94) Pan, M.; He, Y.; Wen, M.; Wu, F.; Sun, D.; Li, S.; Zhang, L.; Li, Y.; Tian, C. One-Pot Hydrazide-Based Native Chemical Ligation for Efficient Chemical Synthesis and Structure Determination of Toxin Mambalgin-1. *Chem Commun* **2014**, *50* (44), 5837–5839. https://doi.org/10.1039/C4CC00779D.
(95) J. Spears, R.; McMahon, C.; Chudasama, V. Cysteine Protecting Groups: Applications in Peptide and Protein Science. *Chem. Soc. Rev.* **2021**, *50* (19), 11098–11155. https://doi.org/10.1039/D1CS00271F.
(96) Diemer, V.; Ollivier, N.; Leclercq, B.; Drobecq, H.; Vicogne, J.; Agouridas, V.; Melnyk, O. A Cysteine Selenosulfide Redox Switch for Protein Chemical Synthesis. *Nat. Commun.* **2020**, *11* (1), 2558. https://doi.org/10.1038/s41467-020-16359-6.
(97) Metanis, N.; Keinan, E.; Dawson, P. E. Traceless Ligation of Cysteine Peptides Using Selective Deselenization. *Angew. Chem. Int. Ed Engl.* **2010**, *49* (39), 7049–7053. https://doi.org/10.1002/anie.201001900.
(98) *Syntheses of Related Compounds of Selenocysteamine and Their Complex Formation with Metal Ions*. https://www.jstage.jst.go.jp/article/cpb1958/19/6/19_6_1089/_article/-char/ja/ (accessed 2023-04-11).
(99) Firstova, O.; Agouridas, V.; Diemer, V.; Melnyk, O. A Selenium-Based Cysteine Surrogate for Protein Chemical Synthesis. *Methods Mol. Biol. Clifton NJ* **2022**, *2530*, 213–239. https://doi.org/10.1007/978-1-0716-2489-0_15.
(100) Torbeev, V. Yu.; Kent, S. B. H. Convergent Chemical Synthesis and Crystal Structure of a 203 Amino Acid "Covalent Dimer" HIV-1 Protease Enzyme Molecule. *Angew. Chem. Int. Ed.* **2007**, *46* (10), 1667–1670. https://doi.org/10.1002/anie.200604087.
(101) Ollivier, N.; Desmet, R.; Drobecq, H.; Blanpain, A.; Boll, E.; Leclercq, B.; Mougel, A.; Vicogne, J.; Melnyk, O. A Simple and Traceless Solid Phase Method Simplifies the Assembly of Large Peptides and the Access to Challenging Proteins. *Chem. Sci.* **2017**, *8* (8), 5362–5370. https://doi.org/10.1039/C7SC01912B.
(102) Agouridas, V.; Diemer, V.; Melnyk, O. Strategies and Open Questions in Solid-Phase Protein Chemical Synthesis. *Curr. Opin. Chem. Biol.* **2020**, *58*, 1–9. https://doi.org/10.1016/j.cbpa.2020.02.007.
(103) *MeSH Browser*. https://meshb.nlm.nih.gov/record/ui?name=Gene+Products%2C+pol (accessed 2023-04-11).
(104) Firstova, O.; Melnyk, O.; Diemer, V. Thiol Catalysis of Selenosulfide Bond Cleavage by a Triarylphosphine. *J. Org. Chem.* **2022**, *87* (14), 9426–9430. https://doi.org/10.1021/acs.joc.2c00934.
(105) *Imidazole buffer Solution BioUltra, 1M water 288-32-4*. http://www.sigmaaldrich.com/ (accessed 2023-04-20).
(106) Tallon, S.; Lawlor, A. C.; Connon, S. J. Urea-Catalyzed Transthioesterification: Towards a New Kinetic Resolution Methodology. *Arkivoc* **2011**, *2011* (4), 115–126. https://doi.org/10.3998/ark.5550190.0012.410.
(107) Hondal, R. J.; Nilsson, B. L.; Raines, R. T. Selenocysteine in Native Chemical Ligation and Expressed Protein Ligation. *J. Am. Chem. Soc.* **2001**, *123* (21), 5140–5141. https://doi.org/10.1021/ja005885t.
(108) Quaderer, R.; Sewing, A.; Hilvert, D. Selenocysteine-Mediated Native Chemical Ligation. *Helv. Chim. Acta* **2001**, *84* (5), 1197–1206. https://doi.org/10.1002/1522-2675(20010516)84.
(109) DeVries, A. L.; Wohlschlag, D. E. Freezing Resistance in Some Antarctic Fishes. *Science* **1969**, *163* (3871), 1073–1075. https://doi.org/10.1126/science.163.3871.1073.
(110) Howard, E. I.; Blakeley, M. P.; Haertlein, M.; Petit-Haertlein, I.; Mitschler, A.; Fisher, S. J.; Cousido-Siah, A.; Salvay, A. G.; Popov, A.; Muller-Dieckmann, C.; Petrova, T.; Podjarny, A. Neutron Structure of Type-III Antifreeze Protein Allows the Reconstruction of AFP-Ice Interface. *J. Mol. Recognit. JMR* **2011**, *24* (4), 724–732. https://doi.org/10.1002/jmr.1130.
(111) DeVries, A. L.; Komatsu, S. K.; Feeney, R. E. Chemical and Physical Properties of Freezing Point-Depressing Glycoproteins from Antarctic Fishes. *J. Biol. Chem.* **1970**, *245* (11), 2901–2908.
(112) Duman, J. G.; Olsen, T. M. Thermal Hysteresis Protein Activity in Bacteria, Fungi, and Phylogenetically Diverse Plants. *Cryobiology* **1993**, *30* (3), 322–328. https://doi.org/10.1006/cryo.1993.1031.

(113) Raymond, J. A.; Sullivan, C. W.; DeVries, A. L. Release of an Ice-Active Substance by Antarctic Sea Ice Diatoms. *Polar Biol.* **1994**, *14* (1), 71–75. https://doi.org/10.1007/BF00240276.
(114) Duman, J. G. Antifreeze and Ice Nucleator Proteins in Terrestrial Arthropods. *Annu. Rev. Physiol.* **2001**, *63* (1), 327–357. https://doi.org/10.1146/annurev.physiol.63.1.327.
(115) Liou, Y.-C.; Tocilj, A.; Davies, P. L.; Jia, Z. Mimicry of Ice Structure by Surface Hydroxyls and Water of a B-Helix Antifreeze Protein. **2000**, *406*, 3.
(116) Buzzini, P.; Margezin, R. *Cold-Adapted Yeasts*. Springer, Berlin, Heidelberg, 2014, 3-22. https://doi.org/10.1007/978-3-642-39681-61.
(117) Ramsay, J., A.; *The Rectal Complex of the Mealworm Tenebrio Molitor L (Coleoptera, Tenebrionidae); Further Studies on the Rectal Complex. (Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, Vols 248, 253, Nos 748, 788) by Ramsay, J A and A V Grimstone:: Very Good Pamphlet (1964) | Plurabelle Books Ltd*. https://www.abebooks.com/Rectal-Complex-Mealworm-Tenebrio-Molitor-Coleoptera/30952759644/bd (accessed 2023-02-15).
(118) DeVries, A. L. Glycoproteins as Biological Antifreeze Agents in Antarctic Fishes. *Science* **1971**, *172* (3988), 1152–1155. https://doi.org/10.1126/science.172.3988.1152.
(119) Duman, J. G. Animal Ice-Binding (Antifreeze) Proteins and Glycolipids: An Overview with Emphasis on Physiological Function. *J. Exp. Biol.* **2015**, *218* (12), 1846–1855. https://doi.org/10.1242/jeb.116905.
(120) Voets, I. K. From Ice-Binding Proteins to Bio-Inspired Antifreeze Materials. *Soft Matter* **2017**, 16.
(121) Scotter, A. J.; Marshall, C. B.; Graham, L. A.; Gilbert, J. A.; Garnham, C. P.; Davies, P. L. The Basis for Hyperactivity of Antifreeze Proteins. *Cryobiology* **2006**, *53* (2), 229–239. https://doi.org/10.1016/j.cryobiol.2006.06.006.
(122) Fairley, K.; Westman, B. J.; Pham, L. H.; Haymet, A. D. J.; Harding, M. M.; Mackay, J. P. Type I Shorthorn Sculpin Antifreeze Protein: RECOMBINANT SYNTHESIS, SOLUTION CONFORMATION, AND ICE GROWTH INHIBITION STUDIES*. *J. Biol. Chem.* **2002**, *277* (27), 24073–24080. https://doi.org/10.1074/jbc.M200307200.
(123) Graether, S. P.; Kuiper, M. J.; Gagné, S. M.; Walker, V. K.; Jia, Z.; Sykes, B. D.; Davies, P. L. Beta-Helix Structure and Ice-Binding Properties of a Hyperactive Antifreeze Protein from an Insect. *Nature* **2000**, *406* (6793), 325–328. https://doi.org/10.1038/35018610.
(124) Raymond, J. A.; Wilson, P.; DeVries, A. L. Inhibition of Growth of Nonbasal Planes in Ice by Fish Antifreezes. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86* (3), 881–885. https://doi.org/10.1073/pnas.86.3.881.
(125) Garnham, C. P.; Natarajan, A.; Middleton, A. J.; Kuiper, M. J.; Braslavsky, I.; Davies, P. L. Compound Ice-Binding Site of an Antifreeze Protein Revealed by Mutagenesis and Fluorescent Tagging. *Biochemistry* **2010**, *49* (42), 9063–9071. https://doi.org/10.1021/bi100516e.
(126) Haymet, A. D. J.; Ward, L. G.; Harding, M. M. Winter Flounder "Antifreeze" Proteins: Synthesis and Ice Growth Inhibition of Analogues That Probe the Relative Importance of Hydrophobic and Hydrogen-Bonding Interactions. *J. Am. Chem. Soc.* **1999**, *121* (5), 941–948. https://doi.org/10.1021/ja9801341.
(127) Gronwald, W.; Chao, H.; Reddy, D. V.; Davies, P. L.; Sykes, B. D.; Sönnichsen, F. D. NMR Characterization of Side Chain Flexibility and Backbone Structure in the Type I Antifreeze Protein at near Freezing Temperatures. *Biochemistry* **1996**, *35* (51), 16698–16704. https://doi.org/10.1021/bi961934w.
(128) Graham, L. A.; Marshall, C. B.; Lin, F.-H.; Campbell, R. L.; Davies, P. L. Hyperactive Antifreeze Protein from Fish Contains Multiple Ice-Binding Sites. *Biochemistry* **2008**, *47* (7), 2051–2063. https://doi.org/10.1021/bi7020316.
(129) Sicheri, F.; Yang, D. S. C. Ice-Binding Structure and Mechanism of an Antifreeze Protein from Winter Flounder. *Nature* **1995**, *375* (6530), 427–431. https://doi.org/10.1038/375427a0.
(130) Sun, T.; Lin, F.-H.; Campbell, R. L.; Allingham, J. S.; Davies, P. L. An Antifreeze Protein Folds with an Interior Network of More than 400 Semi-Clathrate Waters. *Science* **2014**, *343* (6172), 795–798. https://doi.org/10.1126/science.1247407.

(131) Liu, Y.; Li, Z.; Lin, Q.; Kosinski, J.; Seetharaman, J.; Bujnicki, J. M.; Sivaraman, J.; Hew, C.-L. Structure and Evolutionary Origin of Ca2+-Dependent Herring Type II Antifreeze Protein. *PLOS ONE* **2007**, *2* (6), e548. https://doi.org/10.1371/journal.pone.0000548.
(132) Jia, Z.; DeLuca, C. I.; Chao, H.; Davies, P. L. Structural Basis for the Binding of a Globular Antifreeze Protein to Ice. *Nature* **1996**, *384* (6606), 285–288. https://doi.org/10.1038/384285a0.
(133) Gauthier, S. Y.; Scotter, A. J.; Lin, F.-H.; Baardsnes, J.; Fletcher, G. L.; Davies, P. L. A Re-Evaluation of the Role of Type IV Antifreeze Protein. *Cryobiology* **2008**, *57* (3), 292–296. https://doi.org/10.1016/j.cryobiol.2008.10.122.
(134) Bar-Dolev, M.; Celik, Y.; Wettlaufer, J. S.; Davies, P. L.; Braslavsky, I. New Insights into Ice Growth and Melting Modifications by Antifreeze Proteins. *J. R. Soc. Interface* **2012**, *9* (77), 3249–3259. https://doi.org/10.1098/rsif.2012.0388.
(135) Oude Vrielink, A. S.; Aloi, A.; Olijve, L. L. C.; Voets, I. K. Interaction of Ice Binding Proteins with Ice, Water and Ions. *Biointerphases* **2016**, *11* (1), 018906. https://doi.org/10.1116/1.4939462.
(136) Knight, C. A.; Cheng, C. C.; DeVries, A. L. Adsorption of Alpha-Helical Antifreeze Peptides on Specific Ice Crystal Surface Planes. 10.
(137) Pertaya, N.; Marshall, C. B.; Celik, Y.; Davies, P. L.; Braslavsky, I. Direct Visualization of Spruce Budworm Antifreeze Protein Interacting with Ice Crystals: Basal Plane Affinity Confers Hyperactivity. *Biophys. J.* **2008**, *95* (1), 333–341. https://doi.org/10.1529/biophysj.107.125328.
(138) Strom, C. S.; Liu, X. Y.; Jia, Z. Why Does Insect Antifreeze Protein from Tenebrio Molitor Produce Pyramidal Ice Crystallites? *Biophys. J.* **2005**, *89* (4), 2618–2627. https://doi.org/10.1529/biophysj.104.056770.
(139) Olijve, L. L. C.; Oude Vrielink, A. S.; Voets, I. K. A Simple and Quantitative Method to Evaluate Ice Recrystallization Kinetics Using the Circle Hough Transform Algorithm. *Cryst. Growth Des.* **2016**, *16* (8), 4190–4195. https://doi.org/10.1021/acs.cgd.5b01637.
(140) Sidebottom, C.; Buckley, S.; Pudney, P.; Twigg, S.; Jarman, C.; Holt, C.; Telford, J.; McArthur, A.; Worrall, D.; Hubbard, R.; Lillford, P. Heat-Stable Antifreeze Protein from Grass. *Nature* **2000**, *406* (6793), 256–256. https://doi.org/10.1038/35018639.
(141) Raymond, J. A.; Fritsen, C. H. Semipurification and Ice Recrystallization Inhibition Activity of Ice-Active Substances Associated with Antarctic Photosynthetic Organisms. *Cryobiology* **2001**, *43* (1), 63–70. https://doi.org/10.1006/cryo.2001.2341.
(142) Rahman, A. T.; Arai, T.; Yamauchi, A.; Miura, A.; Kondo, H.; Ohyama, Y.; Tsuda, S. Ice Recrystallization Is Strongly Inhibited When Antifreeze Proteins Bind to Multiple Ice Planes. *Sci. Rep.* **2019**, *9* (1), 2212. https://doi.org/10.1038/s41598-018-36546-2.
(143) Olijve, L. L. C.; Meister, K.; DeVries, A. L.; Duman, J. G.; Guo, S.; Bakker, H. J.; Voets, I. K. Blocking Rapid Ice Crystal Growth through Nonbasal Plane Adsorption of Antifreeze Proteins. *Proc. Natl. Acad. Sci.* **2016**, *113* (14), 3740–3745. https://doi.org/10.1073/pnas.1524109113.
(144) Pandey, R.; Usui, K.; Livingstone, R., A.; Fischer, S., A.; Pfaendtner, J.; Backus, E., G.; Weidner, T. *Ice-nucleating bacteria control the order and dynamics of interfacial water | Science Advances*. https://www.science.org/doi/10.1126/sciadv.1501630 (accessed 2022-09-08).
(145) Pummer, B. G.; Budke, C.; Augustin-Bauditz, S.; Niedermeier, D.; Felgitsch, L.; Kampf, C. J.; Huber, R. G.; Liedl, K. R.; Loerting, T.; Moschen, T.; Schauperl, M.; Tollinger, M.; Morris, C. E.; Wex, H.; Grothe, H.; Pöschl, U.; Koop, T.; Fröhlich-Nowoisky, J. Ice Nucleation by Water-Soluble Macromolecules. *Atmos Chem Phys* **2015**, 15.
(146) Deville, S. Ice-Templating, Freeze Casting: Beyond Materials Processing. *J. Mater. Res.* **2013**, *28* (17), 2202–2219. https://doi.org/10.1557/jmr.2013.105.
(147) Walker, V. K.; Zeng, H.; Ohno, H.; Daraboina, N.; Sharifi, H.; Bagherzadeh, S. A.; Alavi, S.; Englezos, P. Antifreeze Proteins as Gas Hydrate Inhibitors. *Can. J. Chem.* **2015**, *93* (8), 839–849. https://doi.org/10.1139/cjc-2014-0538.
(148) Gordienko, R.; Ohno, H.; Singh, V. K.; Jia, Z.; Ripmeester, J. A. Towards a Green Hydrate Inhibitor: Imaging Antifreeze Proteins on Clathrates. *PLoS ONE* **2010**, *5* (2), 7.
(149) Perfeldt, C. M.; Chua, P. C.; Daraboina, N.; Friis, D.; Kristiansen, E.; Ramløv, H.; Woodley, J. M.; Kelland, M. A.; von Solms, N. Inhibition of Gas Hydrate Nucleation and Growth: Efficacy of an Antifreeze Protein from the Longhorn Beetle Rhagium Mordax. *Energy Fuels* **2014**, *28* (6), 3666–3672. https://doi.org/10.1021/ef500349w.

(150) Jeong, Y.; Jeong, S.; Nam, Y. K.; Kang, S. M. Development of Freeze-Resistant Aluminum Surfaces by Tannic Acid Coating and Subsequent Immobilization of Antifreeze Proteins. *Bull. Korean Chem. Soc.* **2018**, *39* (4), 559–562. https://doi.org/10.1002/bkcs.11406.
(151) Gwak, Y. Creating Anti-Icing Surfaces via the Direct Immobilization of Antifreeze Proteins on Aluminum. *Sci. Rep.* 9.
(152) Esser-Kahn, A. P.; Trang, V.; Francis, M. B. Incorporation of Antifreeze Proteins into Polymer Coatings Using Site-Selective Bioconjugation. 6.
(153) Gupta, R.; Deswal, R. Antifreeze Proteins Enable Plants to Survive in Freezing Conditions. *J. Biosci.* **2014**, *39* (5), 931–944. https://doi.org/10.1007/s12038-014-9468-2.
(154) Boonsupthip, W.; Lee, T.-C. Application of Antifreeze Protein for Food Preservation: Effect of Type III Antifreeze Protein for Preservation of Gel-Forming of Frozen and Chilled Actomyosin. *J. Food Sci.* **2003**, *68* (5), 1804–1809. https://doi.org/10.1111/j.1365-2621.2003.tb12333.x.
(155) Ribotta, P. D.; León, A. E.; Añón, M. C. Effect of Freezing and Frozen Storage of Doughs on Bread Quality. *J. Agric. Food Chem.* **2001**, *49* (2), 913–918. https://doi.org/10.1021/jf000905w.
(156) Feeney, R. E.; Yeh, Y. Antifreeze Proteins: Current Status and Possible Food Uses. *Trends Food Sci. Technol.* **1998**, *9* (3), 102–106. https://doi.org/10.1016/S0924-2244(98)00025-9.
(157) Koushafar, H.; Pham, L.; Lee, C.; Rubinsky, B. Chemical Adjuvant Cryosurgery with Antifreeze Proteins. *J. Surg. Oncol.* **1997**, *66* (2), 114–121. https://doi.org/10.1002/(SICI)1096-9098(199710)66:2<114::AID-JSO8>3.0.CO;2-G.
(158) Coger, R.; Rubinsky, B.; Fletcher, G. Microscopic Pattern of Ice Crystal Growth in the Presence of Thermal Hysteresis Proteins. *J. Offshore Mech. Arct. Eng.* **1994**, *116* (3), 173–179. https://doi.org/10.1115/1.2920147.
(159) Brockbank, K. G. M.; Campbell, L. H.; Greene, E. D.; Brockbank, M. C. G.; Duman, J. G. Lessons from Nature for Preservation of Mammalian Cells, Tissues, and Organs. *In Vitro Cell. Dev. Biol. Anim.* **2011**, *47* (3), 210–217. https://doi.org/10.1007/s11626-010-9383-2.
(160) Bakhach, J. The Cryopreservation of Composite Tissues: Principles and Recent Advancement on Cryopreservation of Different Type of Tissues. *Organogenesis* **2009**, *5* (3), 119–126. https://doi.org/10.4161/org.5.3.9583.
(161) Makarevich, A. V.; Kubovicová, E.; Popelková, M.; Fabian, D.; Cikos, S.; Pivko, J.; Chrenek, P. Several Aspects of Animal Embryo Cryopreservation: Anti-Freeze Protein (AFP) as a Potential Cryoprotectant. *Zygote Camb. Engl.* **2010**, *18* (2), 145–153. https://doi.org/10.1017/S0967199409990141.
(162) Farrant, J.; Meares, P.; Echlin, P.; Richards, R. E.; Franks, F. Water Transport and Cell Survival in Cryobiological Procedures. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1977**, *278* (959), 191–205. https://doi.org/10.1098/rstb.1977.0037.
(163) Vaz, C. Cryopreservation and Freeze-Drying Protocols.
(164) Pegg, D. E. Mechanisms of Freezing Damage. *Symp. Soc. Exp. Biol.* **1987**, *41*, 363–378.
(165) Smith, A. U. PREVENTION OF HÆMOLYSIS DURING FREEZING AND THAWING OF RED BLOOD-CELLS. *The Lancet* **1950**, *256* (6644), 910–911. https://doi.org/10.1016/S0140-6736(50)91861-7.
(166) Meryman, H. T.; Hornblower, M. A Method for Freezing and Washing Red Blood Cells Using a High Glycerol Concentration. *Transfusion (Paris)* **1972**, *12* (3), 145–156. https://doi.org/10.1111/j.1537-2995.1972.tb00001.x.
(167) Rowe, A. W.; Eyster, E.; Kellner, A. Liquid Nitrogen Preservation of Red Blood Cells for Transfusion: A Low Glycerol — Rapid Freeze Procedure. *Cryobiology* **1968**, *5* (2), 119–128. https://doi.org/10.1016/S0011-2240(68)80154-3.
(168) Lee, J. K.; Park, K. S.; Park, S.; Park, H.; Song, Y. H.; Kang, S.-H.; Kim, H. J. An Extracellular Ice-Binding Glycoprotein from an Arctic Psychrophilic Yeast. *Cryobiology* **2010**, *60* (2), 222–228. https://doi.org/10.1016/j.cryobiol.2010.01.002.
(169) Lee, S. G.; Koh, H. Y.; Lee, J. H.; Kang, S.-H.; Kim, H. J. Cryopreservative Effects of the Recombinant Ice-Binding Protein from the Arctic Yeast Leucosporidium Sp. on Red Blood Cells. *Appl. Biochem. Biotechnol.* **2012**, *167* (4), 824–834. https://doi.org/10.1007/s12010-012-9739-z.
(170) Kristiansen, E.; Pedersen, S.; Ramløv, H.; Zachariassen, K. E. Antifreeze Activity in the Cerambycid Beetle Rhagium Inquisitor. *J. Comp. Physiol. B* **1999**, *169* (1), 55–60. https://doi.org/10.1007/s003600050193.

(171) Jang, T. H.; Park, S. C.; Yang, J. H.; Kim, J. Y.; Seok, J. H.; Park, U. S.; Choi, C. W.; Lee, S. R.; Han, J. Cryopreservation and Its Clinical Applications. *Integr. Med. Res.* **2017**, *6* (1), 12–18. https://doi.org/10.1016/j.imr.2016.12.001.
(172) McManus, D. P.; Dunne, D. W.; Sacko, M.; Utzinger, J.; Vennervald, B. J.; Zhou, X.-N. Schistosomiasis. *Nat. Rev. Dis. Primer* **2018**, *4* (1), 1–19. https://doi.org/10.1038/s41572-018-0013-8.
(173) *Biology and chemistry of Platyhelminthes*. Institut Pasteur de Lille. https://pasteur-lille.fr/en/center-of-research/research-units/center-of-infection-and-immunity-of-lille/biology-and-chemistry-of-platyhelminthes/ (accessed 2022-09-08).
(174) *Chemical Biology of Flatworms- Schistosoma mansoni cycle of life*. http://chemicalbiologyflatworms.org/spip.php?article11 (accessed 2022-09-08).
(175) James, E. R.; Farrant, J. Recovery of Infective Schistosoma Mansoni Schistosomula from Liquid Nitrogen: A Step towards Storage of a Live Schistosomiasis Vaccine. *Trans. R. Soc. Trop. Med. Hyg.* **1977**, *71* (6), 498–500. https://doi.org/10.1016/0035-9203(77)90143-2.
(176) Stirewalt, M.; Lewis, F. A.; Cousin, C. E.; Leef, J. L. Cryopreservation of Schistosomules of Schistosoma Mansoni in Quantity. *Am. J. Trop. Med. Hyg.* **1984**, *33* (1), 116–124. https://doi.org/10.4269/ajtmh.1984.33.116.
(177) James, E. R. Cryo Preservation of Schistosoma Mansoni Schistosomula Using 40 Percent Volume Volume Methanol and Rapid Cooling. *Cryo Letters* **1980**, *1* (15), 535–546.
(178) James, E. R. Schistosoma Mansoni: Cryopreservation of Schistosomula by Two-Step Addition of Ethanediol and Rapid Cooling. *Exp. Parasitol.* **1981**, *52* (1), 105–116. https://doi.org/10.1016/0014-4894(81)90066-7.
(179) Marshall, C. B.; Daley, M. E.; Sykes, B. D.; Davies, P. L. Enhancing the Activity of a β-Helical Antifreeze Protein by the Engineered Addition of Coils. *Biochemistry* **2004**, *43* (37), 11637–11646. https://doi.org/10.1021/bi0488909.
(180) Yu, S. O.; Brown, A.; Middleton, A. J.; Tomczak, M. M.; Walker, V. K.; Davies, P. L. Ice Restructuring Inhibition Activities in Antifreeze Proteins with Distinct Differences in Thermal Hysteresis. *Cryobiology* **2010**, *61* (3), 327–334. https://doi.org/10.1016/j.cryobiol.2010.10.158.
(181) *Plant antifreeze proteins and their expression regulatory mechanism | SpringerLink*. https://link.springer.com/article/10.1007/s11632-005-0057-1 (accessed 2022-09-13).
(182) Graham, L. A.; Liou, Y.-C.; Walker, V. K.; Davies, P. L. Hyperactive Antifreeze Protein from Beetles. *Nature* **1997**, *388* (6644), 727–728. https://doi.org/10.1038/41908.
(183) Liou, Y. C.; Thibault, P.; Walker, V. K.; Davies, P. L.; Graham, L. A. A Complex Family of Highly Heterogeneous and Internally Repetitive Hyperactive Antifreeze Proteins from the Beetle Tenebrio Molitor. *Biochemistry* **1999**, *38* (35), 11415–11424. https://doi.org/10.1021/bi990613s.
(184) Liou, Y.-C.; Davies, P. L.; Jia, Z. Crystallization and Preliminary X-ray Analysis of Insect Antifreeze Protein from the Beetle Tenebrio Molitor. **2000**, 3.
(185) Marshall, C. B.; Daley, M. E.; Graham, L. A.; Sykes, B. D.; Davies, P. L. Identification of the Ice-Binding Face of Antifreeze Protein from *Tenebrio Molitor*. *FEBS Lett.* **2002**, *529* (2–3), 261–267. https://doi.org/10.1016/S0014-5793(02)03355-0.
(186) Bar, M.; Bar-Ziv, R.; Scherf, T.; Fass, D. Efficient Production of a Folded and Functional, Highly Disulfide-Bonded Beta-Helix Antifreeze Protein in Bacteria. *Protein Expr. Purif.* **2006**, *48* (2), 243–252. https://doi.org/10.1016/j.pep.2006.01.025.
(187) Tomalty, H. E.; Graham, L. A.; Eves, R.; Gruneberg, A. K.; Davies, P. L. Laboratory-Scale Isolation of Insect Antifreeze Protein for Cryobiology. *Biomolecules* **2019**, *9* (5), 180. https://doi.org/10.3390/biom9050180.
(188) Liou, Y.-C.; Daley, M. E.; Graham, L. A.; Kay, C. M.; Walker, V. K.; Sykes, B. D.; Davies, P. L. Folding and Structural Characterization of Highly Disulfide-Bonded Beetle Antifreeze Protein Produced in Bacteria. *Protein Expr. Purif.* **2000**, *19* (1), 148–157. https://doi.org/10.1006/prep.2000.1219.
(189) Ollivier, N.; Dheur, J.; Mhidia, R.; Blanpain, A.; Melnyk, O. Bis(2-Sulfanylethyl)Amino Native Peptide Ligation. *Org. Lett.* **2010**, *12* (22), 5238–5241. https://doi.org/10.1021/ol102273u.
(190) *"Peptides: Structure and Function. [Proceedings of the 9th American Peptide Symposium Held in Toronto, Ontario, Can., June 23-28, 1985]" Reference Detail | CAS SciFinder*^{n}. https://scifinder-n-cas-org.ressources-electroniques.univ-

lille.fr/searchDetail/reference/640726e13136c60ac942999c/referenceDetails (accessed 2023-03-07).
(191) Wöhr, T.; Wahl, F.; Nefzi, A.; Rohwedder, B.; Sato, T.; Sun, X.; Mutter, M. Pseudo-Prolines as a Solubilizing, Structure-Disrupting Protection Technique in Peptide Synthesis. *J. Am. Chem. Soc.* **1996**, *118* (39), 9218–9227. https://doi.org/10.1021/ja961509q.
(192) Endoproteinase Glu-C.
(193) Pane, K.; Verrillo, M.; Avitabile, A.; Pizzo, E.; Varcamonti, M.; Zanfardino, A.; Di Maro, A.; Rega, C.; Amoresano, A.; Izzo, V.; Di Donato, A.; Cafaro, V.; Notomista, E. Chemical Cleavage of an Asp-Cys Sequence Allows Efficient Production of Recombinant Peptides with an N-Terminal Cysteine Residue. *Bioconjug. Chem.* **2018**, *29* (4), 1373–1383. https://doi.org/10.1021/acs.bioconjchem.8b00083.
(194) Shimamoto, S.; Mitsuoka, N.; Takahashi, S.; Kawakami, T.; Hidaka, Y. Chemical Digestion of the -Asp-Cys- Sequence for Preparation of Post-Translationally Modified Proteins. *Protein J.* **2020**, *39* (6), 711–716. https://doi.org/10.1007/s10930-020-09940-x.
(195) Chen, Y.; Mehok, A. R.; Mant, C. T.; Hodges, R. S. Optimum Concentration of Trifluoroacetic Acid for Reversed-Phase Liquid Chromatography of Peptides Revisited. *J. Chromatogr. A* **2004**, *1043* (1), 9–18. https://doi.org/10.1016/j.chroma.2004.03.070.

General conclusion

Proteins are macromolecules naturally produced by all living organisms and constitute essential components of organic life. Proteins and peptides have found application in various fields, for example, therapeutics or materials development. Therefore, it is hard to overestimate the importance of producing those biomolecules. One well developed approach for accessing proteins is recombinant methods that exploit the natural machinery of bacterial, or cell cultures. Meantime, chemical synthesis significantly extended peptides and proteins fabrication, allowing to obtain molecules with particular properties (antibacterial, protease resistant, etc.) and to modulate easily peptidic sequence and whole composition.

To access polypeptides by chemical synthesis, Merrifield in 1963 introduced the solid phase peptide synthesis (SPPS), a chemical method that is appropriate for the preparation of peptides and small proteins with a size usually less than 50 AAs. To facilitate the access to longer peptide sequences with a size greater than 50 AAs, a complementary strategy was developed based on the concatenation of smaller peptide segments. To make this second approach work, chemists exploit a few chemical ligation reactions to covalently and chemoselectively connect unprotected peptide segments together.

During my PhD fellowship I have worked on a novel method of peptide ligation, allowing to assemble several peptide fragments in concise and sequential manner. To do so, I faced the challenge to master Native chemical ligation reaction and combine this chemistry with a novel chemical tool, SetCys cysteine surrogate, which can be activated upon reduction of its internal selenosulfide bond. The main complexity was to compromise both type of reactivities and to find conditions to allow polypeptide synthesis with relatively high yield and low by-products formation. This part of the work was published and presented during a conference.

Regarding the protein target, my main interest was in antifreeze proteins (AFPs), which are produced by cold coping organisms in nature. In particular, I was interested by a protein produced by *Tenebrio molitor* (*Tm*) beetle to protect itself from cryodamage. During my PhD, I aimed to produce *Tm*AFP analogues with antifreeze activity by chemical synthesis. Due to the lack of time, I could only make the first steps toward synthesis of this challenging protein, mastering its chemistry.

The application field of antifreeze proteins is impressive, as they can improve existing methods of cryopreservation of nutrition products, cells, tissues and more complex systems such as organs or living organisms.

Annexes

- **Annex A. Pedal to the Metal: The Homogeneous Catalysis of the Native Chemical Ligation Reaction.** Diemer, V.; Firstova, O.; Agouridas, V.; Melnyk, O. *Chem. – Eur. J.* **2022**, *28*; e202104229;
- **Annex B. Thiol Catalysis of Selenosulfide Bond Cleavage by a Triarylphosphine** (main paper and Supporting Information). Firstova, O.; Diemer, V.; Melnyk, *O. J. Org. Chem.* **2022**, *87*, 9426-9430;
- **Annex C. Experimental part of the manuscript;**
- **Annex D. A Selenium-Based Cysteine Surrogate for Protein Chemical Synthesis.** Firstova, O.; Agouridas, V.; Diemer, V.; Melnyk, O. *Methods Mol. Biol.* **2022**, *2530*, 213–239.

Annex A

Pedal to the Metal: The Homogeneous Catalysis of the Native Chemical Ligation Reaction

Diemer, V.; Firstova, O.; Agouridas, V.; Melnyk, O.

Chem. – Eur. J. **2022**, *28* (16)

The pre-printed version is presented here.

The peer-reviewed version is available by

DOI 10.1002/chem.202104229

Pedal to the metal: the homogeneous catalysis of the native chemical ligation reaction

Vincent Diemer,[a] Olga Firstova,[a] and Vangelis Agouridas,*[b] Oleg Melnyk*[a]

a] Dr. V. Diemer, O. Firstova, Dr. O. Melnyk
Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 9017 - CIIL - Center for Infection and Immunity of Lille, F-59000 Lille, France
E-mail: oleg.melnyk@ibl.cnrs.fr

[b] Dr. V. Agouridas
Centrale Lille, F-59000 Lille, France
E-mail: vangelis.agouridas@ibl.cnrs.fr

Abstract: The native chemical ligation reaction of peptide thioesters with cysteinyl peptides is a pivotal chemical process in the production of native or modified peptides and proteins, and well beyond in the preparation of various biomolecule analogs and materials. To benefit from this reaction at its fullest and to access all the possible applications, the experimentalist needs to know the factors affecting its rate and how to control it. This concept article presents the fundamental principles underlying the rate of the native chemical ligation and its homogeneous catalysis by nucleophiles. It has been prepared to serve as a quick guide in the search for an appropriate catalyst.

Introduction

By enabling the concatenation of unprotected peptide segments in water under mild conditions, the Native Chemical Ligation (NCL) reaction has become one of the most valuable transformation for accessing synthetic or semisynthetic proteins (Figure 1A).[1-4] The core principle of this process consists in opposing a peptide thioester and a cysteinyl (Cys) peptide whose chemoselective reaction results in the formation of a native peptide bond between the two segments and thus, in the formation of a larger and well-defined polypeptide. Since the advent of NCL in 1994, its synthetic scope has been significantly enriched, providing protein chemists with an eclectic repertoire of practical methods for accessing a tremendous diversity of protein targets.[5][6] This gain in operational synthetic capacity is reflected in the advances made in numerous domains where NCL is used as a routine tool, including the fields of chemical biology and biomedical research (Figure 1B).[7]

Considered as a pivotal transformation for the production of proteins, NCL has seen its mechanism come under scrutiny very shortly after its discovery.[8] The apparent straightforward equation represented in Figure 1A in reality conceals a two-step process, which starts with a thiol-thioester exchange leading to a capture intermediate (Figure 2A, Step 1). Then, the latter rearranges through an intramolecular *S*-to-*N* acyl group transfer to yield the product with a native peptide bond to cysteine (Figure 2A, Step 2). In standard conditions, experimental evidence suggests that the reaction is kinetically governed by step 1 which proceeds through tetrahedral intermediate **1** (Figure 2A). Logically, the rate of step 1 is under the dependence of the reactivity of the thioester and Cys thiol groups. While the reactivity of the Cys thiol is mainly determined by the fraction of Cys thiolate species present at the working pH,[9] the reactivity of the thioester partner largely depends on two factors: the nature of the C-terminal residue and the type of thiol leaving group installed (Figure 2B). Thanks to their straightforward synthetic accessibility and good stability, peptide alkyl thioesters remain the most appreciated acylating source used for protein semi- or total synthesis.[2, 6] However, practically useful ligation rates cannot usually be attained with such thioesters at millimolar concentration due to their tempered reactivity. As a consequence, the NCL with peptide alkyl thioesters needs to be promoted most of the time by addition of catalysts. The aim of this concept article is to examine the fundamental principles underlying the homogeneous catalysis of the NCL reaction and to guide the reader in the search for an appropriate catalyst.

The guiding principles of NCL catalysis

Among the several modes of catalytic activation, covalent catalysis by nucleophiles [10] provides the most significant rate enhancements for NCL. This approach consists in using a nucleophile additive (Nu) intended to react with the thioester partner for generating an activated acyl-Nu intermediate *in situ* (Figure 3A). During the reaction, the starting thioester and the acyl-Nu intermediate coexist in solution. Although the latter is usually formed in tiny amounts, its enhanced acylating capability makes it kinetically the main acylating species in the process. To achieve this goal, the nucleophile must be chosen carefully according to the nature of the starting thioester so that the energy barrier going from reactants to tetrahedral intermediate **3** is lower than that leading to tetrahedral intermediate **1** (Figure 3B). In practice, nucleophilic catalysis of NCL is usually carried out by the addition of various sulfur-, selenium-, nitrogen- or phosphorus-derived nucleophiles, which all present some benefits and drawbacks reviewed hereafter (Figure 3C).

Thiolates, a type of sulfur-based nucleophiles, have rapidly emerged as successful additives for the activation of peptide thioesters [11] and are still the most widely deployed solution for enhancing NCL's rate today (Figure 4A). Their efficiency as catalysts can be rationalized according to the pioneering work of Jencks and coworkers, which allows to qualitatively appreciate the extent of activation of a thioester upon thiol-thioester exchanges.[11] Indeed, the acylating power of the activated thioester, and thus the rate of NCL, largely depends on the acidity of the thiol catalyst R_2SH relative to that of the departing thiol R_1SH (Figure 4A). The use of a thiol additive R_2SH having a lower acid dissociation constant (pK_a) than that of R_1SH results in efficient catalysis because the thiolate R_2S^- is a better leaving group. However, in such a case, the formation of the activated thioester is a kinetically disfavored situation unless a very significant excess of catalyst is introduced in the reaction mixture as compensation.

A selection of thiols used either for thioester synthesis or NCL catalysis are listed by increasing acidity in Figure 4B (3-mercaptopropionic acid, MPA; sodium 2-mercaptoethyl sulfonate, MESNa[12]; methyl thioglycolate, MTG [13]; trifluoroethanethiol, TFET [14]; 4-hydroxythiophenol [15]; 3-mercaptophenyl acetic acid, MPAA [15]; thiophenol [8]). Despite the large diversity of commercially available thiols, just a handful of them are used as catalysts, one stringent selection criterion being their good to excellent water-solubility. Among these, MPAA has a pK_a well-below those of alkyl thiols

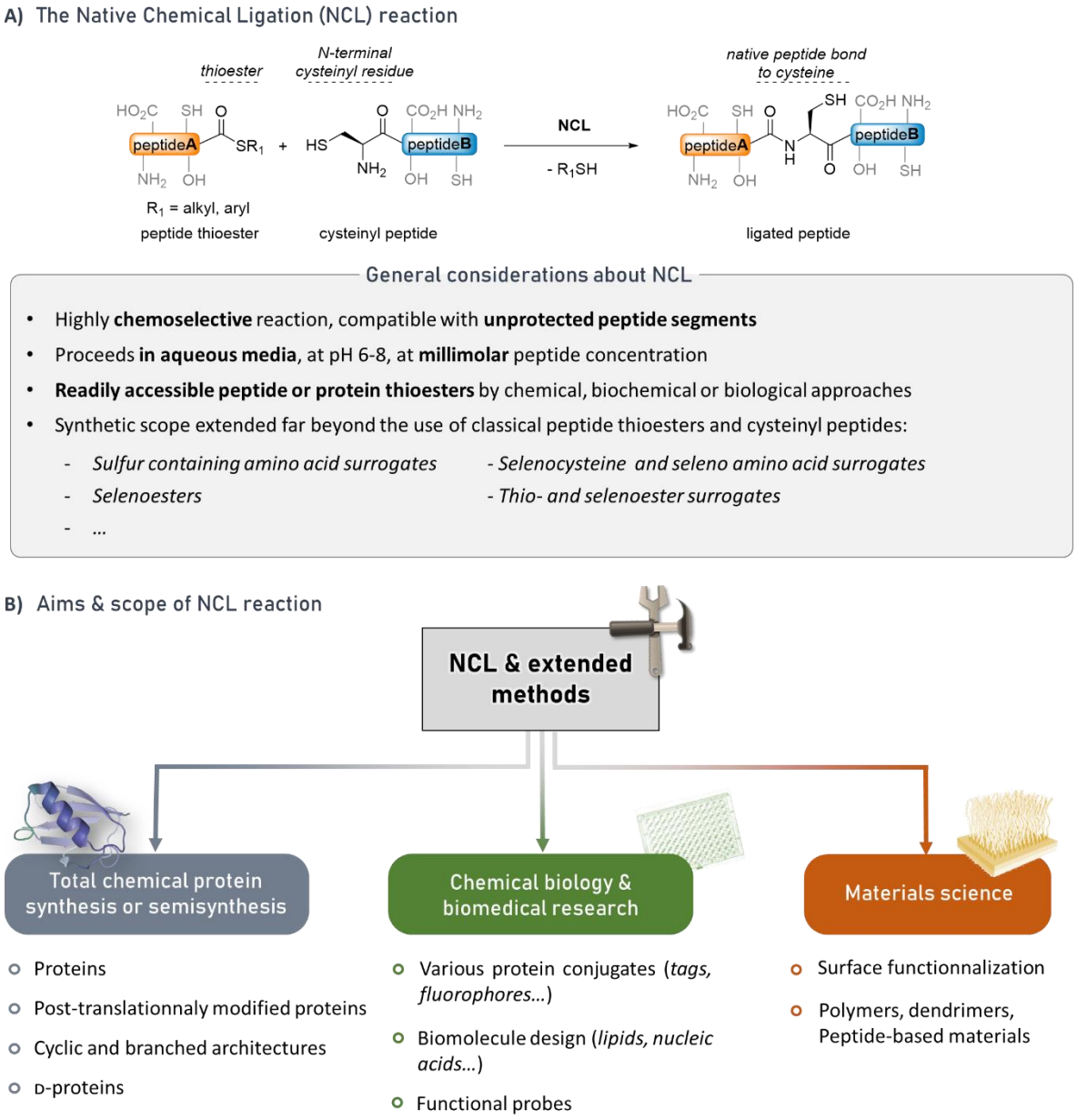


Figure 1. General considerations about the Native Chemical Ligation (NCL) reaction. A) Principle of the native chemical ligation reaction (functional groups on amino acid side chains are represented in gray on each peptide segment. For the sake of clarity, they will not be shown in the following figures). B) Most common applications of the NCL reaction. Various examples can be found in refs [3, 4, 7].

classically used for thioester synthesis, is odorless and highly water soluble at the working pH of NCL (≈ 7) due to its carboxylic acid moiety. Therefore, MPAA is by far the most widely applied catalyst in the field. In addition, it is important to note that thiol moieties harbored by the sidechain of internal Cys residues can in some cases offer a significant catalytic contribution owing to their mid-range pK_a.[16]

The use of thiols, and more particularly of aryl thiols such as MPAA, can however prove problematic in some specific situations. This is especially the case when the NCL reaction needs to be combined with a subsequent Metal-Free Desulfurization (MFD) step in a one-pot process (Figure 5A).[17] This popular approach was developed to extend the scope of NCL by converting the Cys junction residue to an alanine (Ala) residue, a significantly more represented amino acid in protein sequences. The mechanism of MFD is critically dependent on the formation of a Cys-derived thiyl radical intermediate, which must combine with the added phosphine to make the desulfurization process successful. Aryl thiols are powerful thiyl radical scavengers and therefore kill the MFD reaction especially when used in large excess as discussed above. Consequently, their use is not compatible with one-pot NCL/MFD approaches unless the aryl thiol can be removed prior to adding the reagents for the MFD step. For example, the MPAA-derivative equipped with a hydrazide handle and designed by Brik and coll. was removed by using an aldehyde scavenger solid support (Figure 5B).[18] Another option consists in using TFET as a catalyst, which can be removed by evaporation after ligation.[14] Although being less powerful NCL catalysts than MPAA, alkyl thiols such as MTG [13] are practicable MFD-compatible thiol additives. Imidazole and N-heterocyclic analogues thereof,[19-21] *i.e.* sp^2 nitrogen-derived nucleophiles, provide an additional option in this regard. Indeed, they do not interfere with MFD reaction conditions and can recapitulate the catalytic properties of thiol additives through the formation of acyl imidazole type intermediates. They must however be introduced at much higher loadings to this end

because the equilibrated process depicted in Figure 5C favors the thioester species.

The last decade has witnessed the development of various thioester surrogates, which broaden the scope of classical thioesters by permitting the design of novel peptide assembly strategies.[22] The class of thioester precursors referred to as

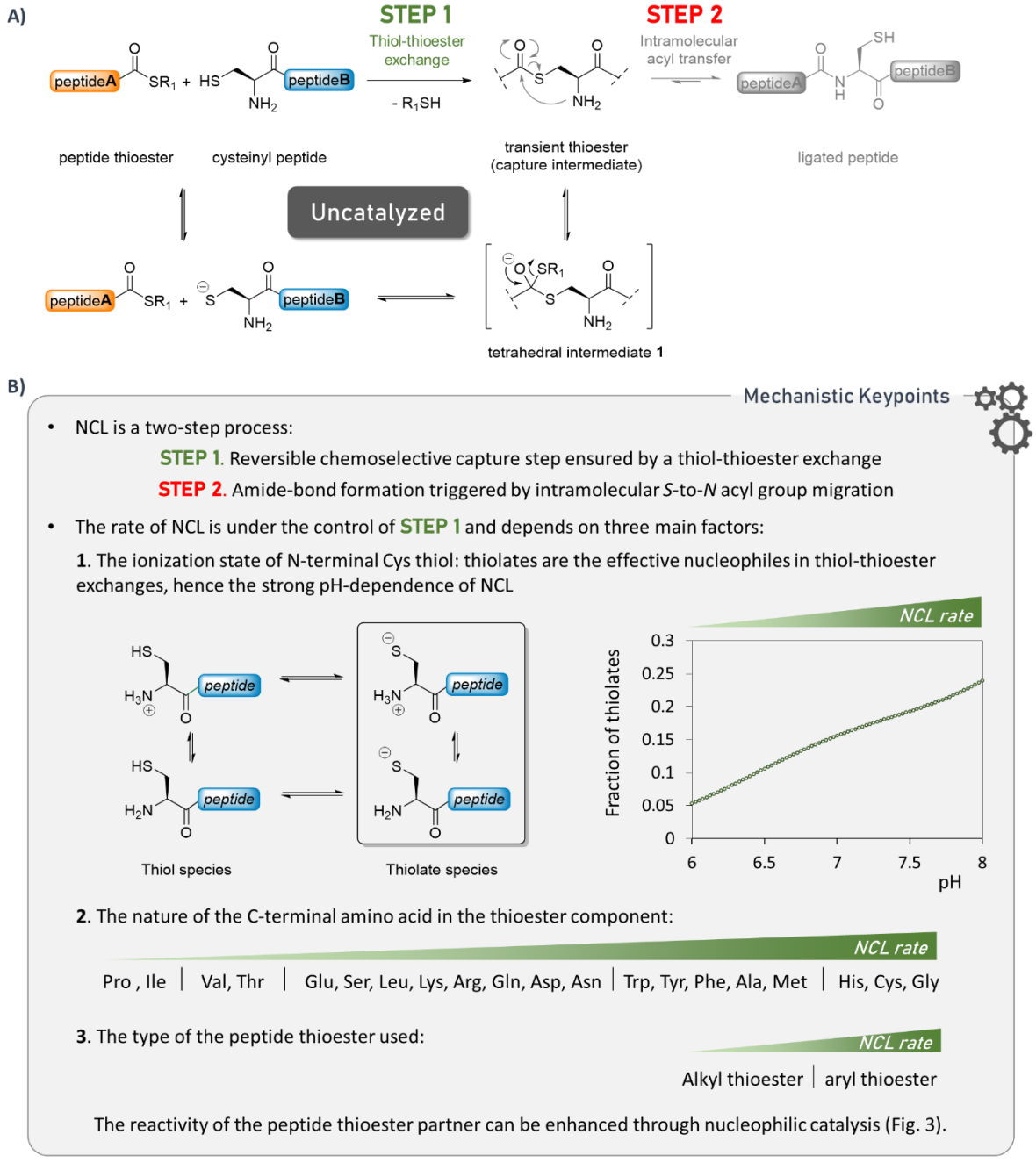


Figure 2. Insights into the mechanism of the uncatalyzed NCL reaction. A) NCL is a two-step process involving a thiol-thioester exchange for the capture step. B) NCL is usually under the kinetic control of the thiol-thioester exchange step and depends on three main factors: the ionization state of the N-terminal cysteine residue (the fraction of thiolate was calculated based on the model and the data in ref. [9]) and the environment as well as the nature of the thioester partner.

N,S-acyl shift systems is particularly popular (Figure 6).[23, 24] Including the *bis*(2-sulfanylethyl)amido (SEA[25]) and the *N*-sulfanylethylanilide (SEAlide[26]) devices, they consist of special amides that are able to rearrange *in situ* into thioesters through an *N,S*-acyl shift (Figure 6, STEP 1). Similar to classical thioesters, these transient thioesters can then be involved in an NCL-like process to provide ligated peptides (Figure 6A, STEPS 2 and 3; Figure 6B, STEP 2). Classical aryl thiol additives can promote the NCL-like part of the process but not the *N,S*-acyl shift, which is usually rate limiting and catalyzed by other mechanisms. This specific feature of the *N,S*-acyl shift systems has prompted the development of dedicated catalysts. For example, the SEA group has benefited from the design of tailored selenol-based additives (Figure 6A).[27, 28] Thanks to the low pK_a of the selenol group (pKa ~ 5), the catalyst is still present in its nucleophilic selenolate form in the working acidic conditions that are favorable for the *N,S*-acyl shift rearrangement step (Figure 6A, STEP 1). Its reaction with the SEA thioester intermediate results in the *in situ* formation of a selenoester (Figure 6A, STEP 2), a type of acyl donor that proved particularly powerful under classical NCL conditions.[29]

In another example, the rearrangement of the SEAlide group to generate a thioester intermediate depends on the presence of phosphate.[26] Otaka and coworkers have developed a dual catalyst featuring a phosphonic acid able to catalyze the rearrangement of the SEAlide device (Figure 6B, STEP 1) and an aryl thiol moiety for promoting the NCL-like part of the process (Figure 6B, STEP 2).[30]

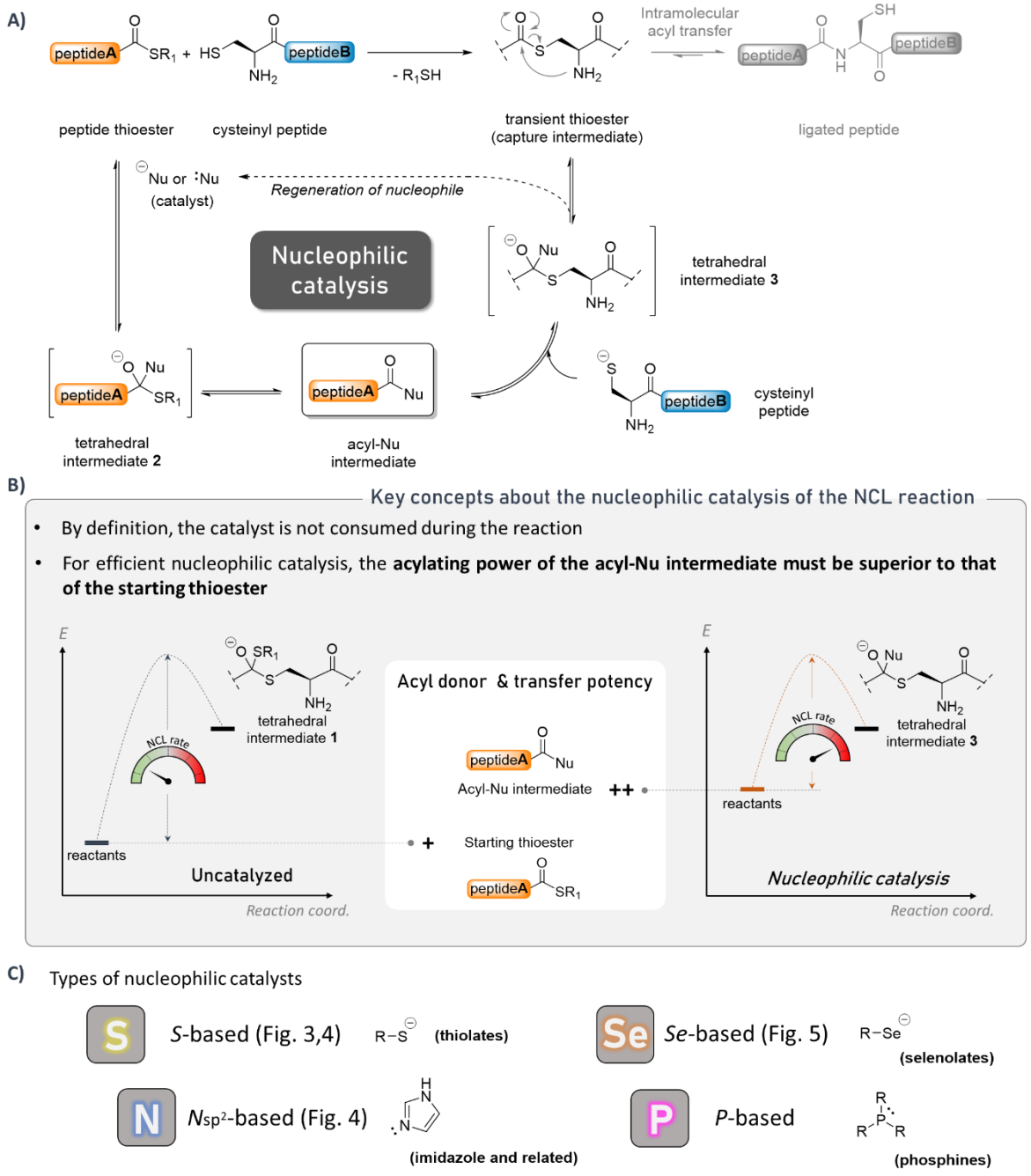


Figure 3. Activation of the thioester partner through nucleophilic acyl substitution is key to NCL rate enhancement by *in situ* generation of more efficient acyl donors. A) The mechanism of the NCL reaction catalyzed by nucleophile additives involves the formation of an acyl-nucleophile intermediate. B) Nucleophilic catalysis operates on the condition that the acyl-Nu intermediate is a better acyl transfer reactant than the starting thioester. C) Types of nucleophilic catalysts.

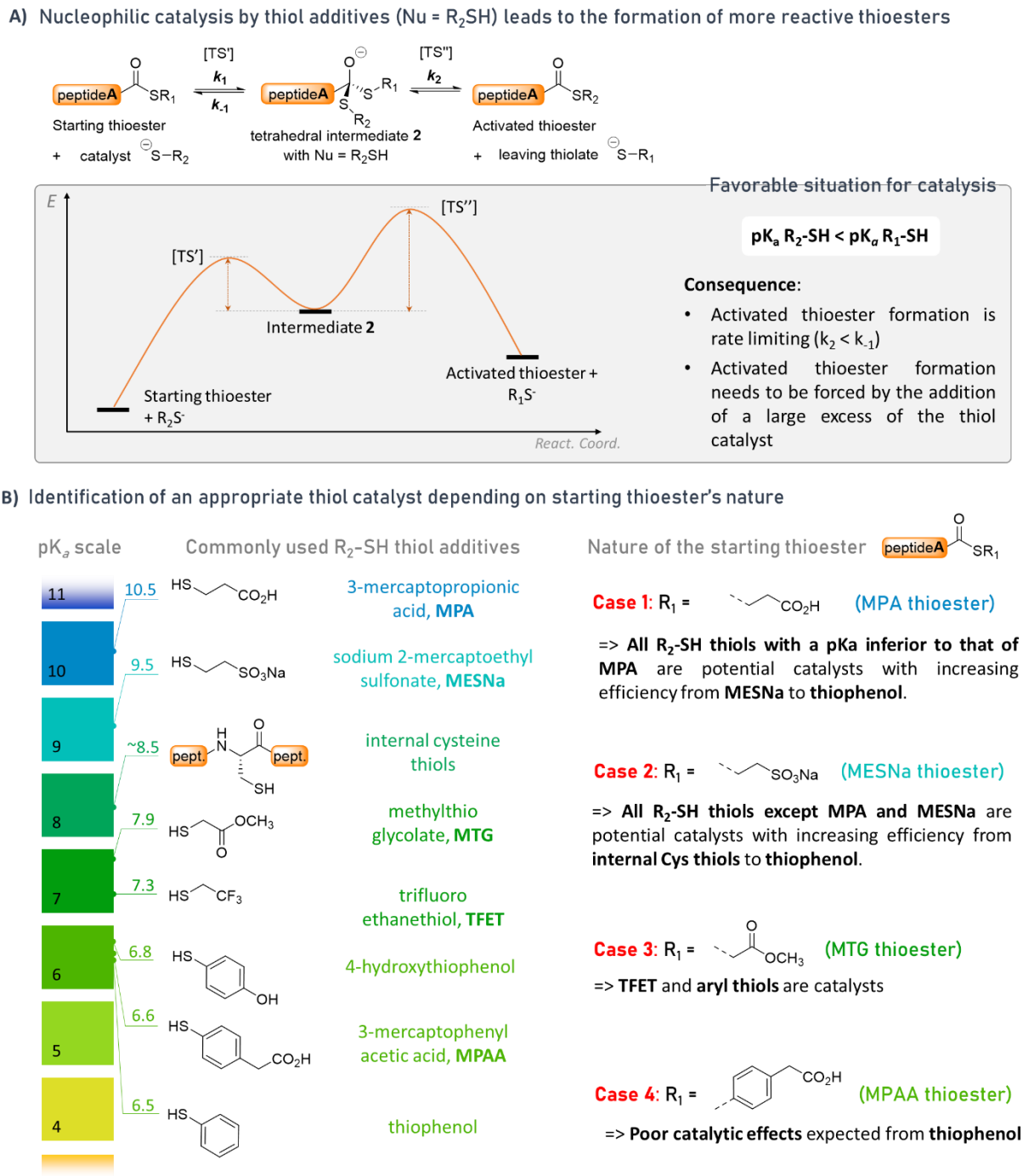
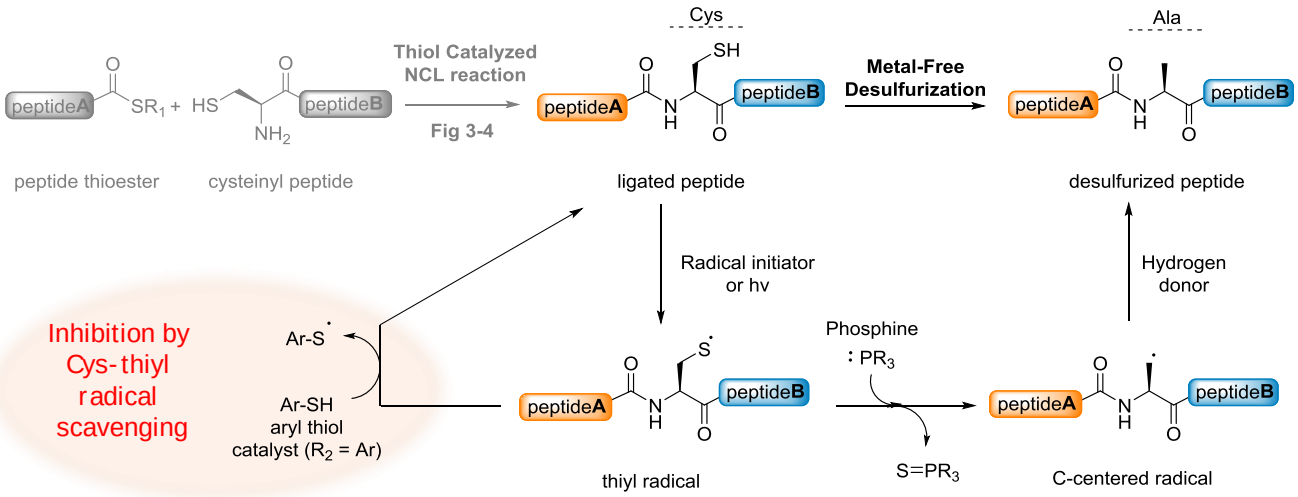


Figure 4. NCL catalysis by thiol additives. A) Principle of NCL catalysis by thiols. B) Identification of an appropriate thiol catalyst depending on starting thioester's nature: as a rule of thumb, thiol catalysts being more acidic (from top to bottom on the pK_a scale) than the departing thiol have a catalytic effect. Examples are provided at the left of the figure.

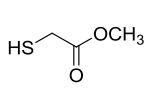
A) Compatibility of thiol additives with metal-free desulfurization (MFD)



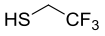
General considerations about metal-free desulfurization (MFD)

- MFD considerably extends the scope of NCL (Ala is 5.7 fold more abundant than Cys in proteins)
- NCL/MFD one-pot processes save time and yield by limiting work-up & purification steps
- Arylthiol catalysts (e.g. MPAA) inhibit MFD by neutralizing the intermediate Cys-thiyl radical

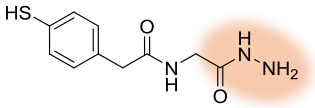
B) Which thiol catalyst to use for a one-pot NCL/MFD sequence?



MTG: compatible

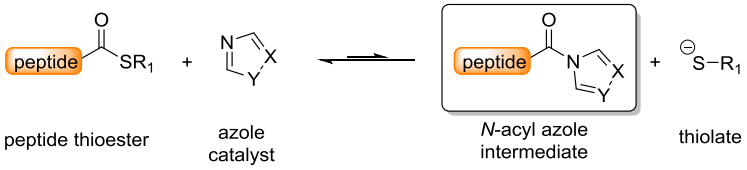


TFET easily eliminated by evaporation (bp 35 °C) before MFD

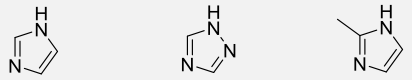


MPAA hydrazide removed prior to MFD by capture on an aldehyde resin and filtration

C) Nsp^2 based-catalysts: a powerful alternative to thiol catalysts for one-pot NCL/MFD approaches



Azole catalysts



- *N*-acyl azoles are efficient acyl donors toward *S*-nucleophiles
- Process thermodynamically unfavored (shifted toward the formation of thioesters)

  => High concentrations of the azole catalyst are required to observe significant rate enhancements (> 2-5 M)

Figure 5. Compatibility of the NCL catalyst with one-pot post-ligation desulfurization methods. A) General mechanism of metal-free desulfurization (MFD) and its inhibition by aryl thiols. B) Some thiols used for one-pot NCL/MFD protocols. C). Nucleophilic catalysis by Nsp^2-nucleophiles is compatible with one-pot NCL/MFD protocols.

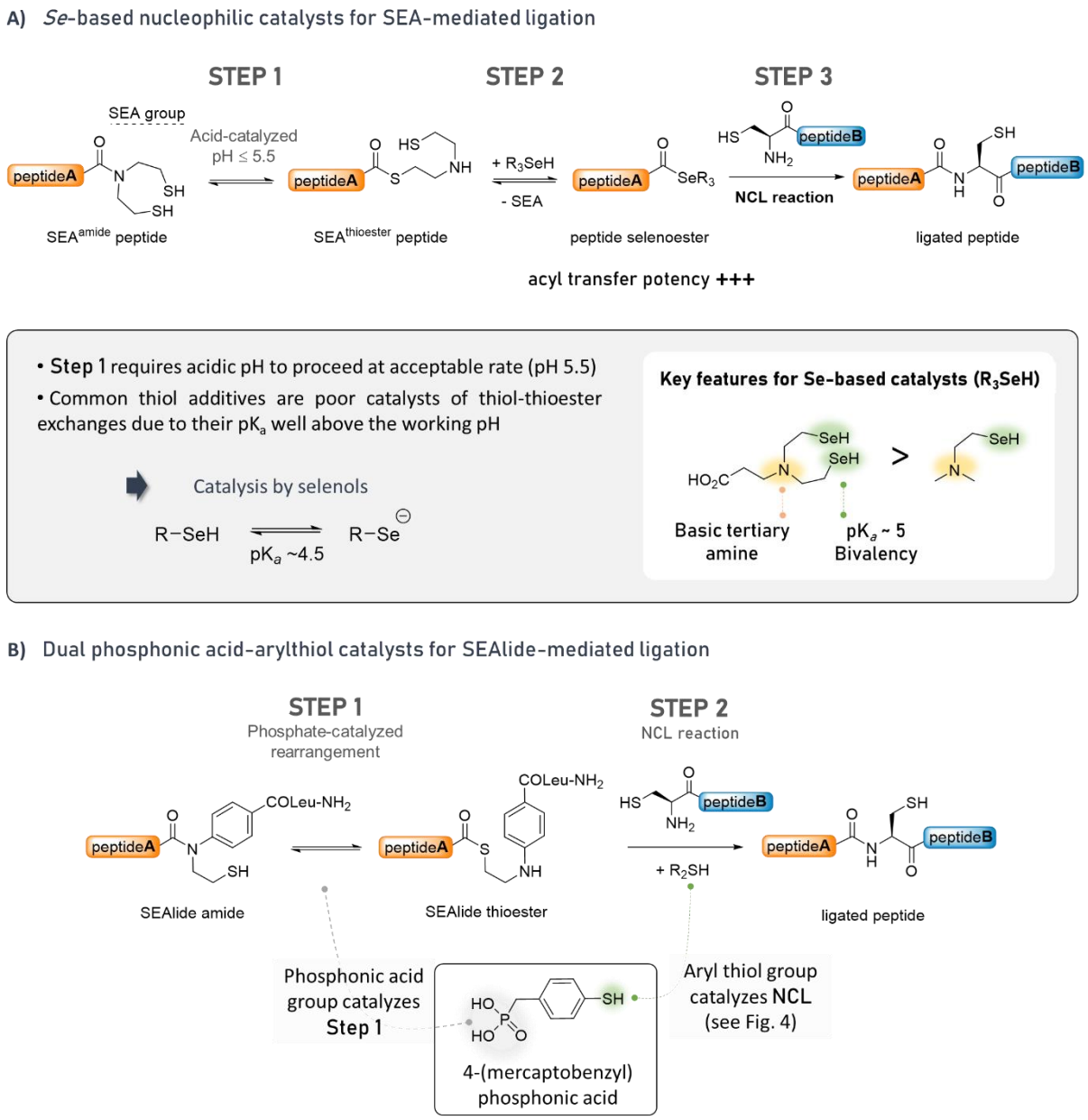


Figure 6. Catalysis of *N,S*-acyl shift-mediated ligations. A) Catalysis of SEA ligation by selenol compounds. B). Catalysis of SEAlide-mediated ligation by a dual phosphonic acid-aryl thiol catalyst.

Summary and outlook

In conclusion, the rapid concatenation of peptide segments in water using NCL is central to protein chemical synthesis and many other applications of this chemoselective reaction. The catalysis of NCL and related reactions enables to save time, to force the formation of difficult junctions and to minimize the impact of protein aggregation and potential side reactions. Throughout the years, this quest for rate enhancement has been sustained by an in-depth understanding of reaction mechanisms based on experimental and theoretical approaches.

A limitation of existing catalysts is the large excesses used to achieve useful synthetic rates. Therefore, future research might see the emergence of novel catalysis approaches enabling the acceleration of ligation reactions while significantly reducing the heavy excesses of nucleophiles. A potential strategy can be to increase the effective concentration of the nucleophilic catalyst at the thioester site by incorporating it in the thiol handle. By becoming intramolecular, the exchange process leading to the formation of the activated acyl-Nu species might not be limiting anymore. Such an approach using a Cys thiol as the internal nucleophile has been examined by Schmalisch and Seitz, who reported significant rate enhancements.[31]

Non-covalent strategies for increasing the effective concentration of the nucleophilic catalyst at the thioester site remains to be developed. The concepts underpinning template-assisted NCL reactions might serve this purpose, albeit to the expense of increased synthetic work and reduced atom economy.[2, 32, 33] Templated NCL has been developed for enabling NCL at very low reactant concentration such as in the labelling of proteins exposed to the cell surface.[34] The method relies on the high affinity between pairs of nucleic acids or peptides for increasing the effective concentration of reactants and thus the ligation yields. Likewise, we can imagine that the biomolecular systems used so far for promoting reactant

encounter could be simplified in order to extend their application to catalysis.

Acknowledgements

This work was supported by the Centre National de la Recherche Scientifique (CNRS), the University of Lille Nord de France, Institut Pasteur de Lille.

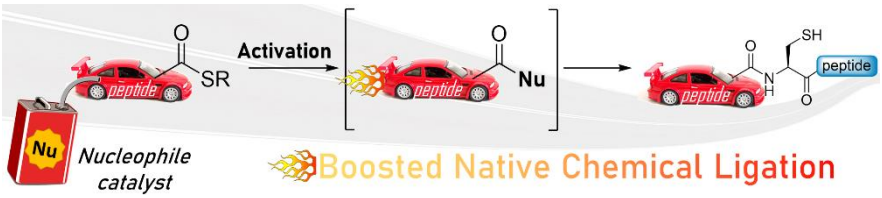
Conflict of interest

The authors declare no conflict of interest.

Keywords: Native Chemical Ligation • homogeneous catalysis • nucleophilic catalysis • thiol-thioester exchange • desulfurization • chemoselectivity

[1] Dawson, P. E., Muir, T. W., Clark-Lewis, I., Kent, S. B. H., *Science* **1994,** *266*, 776-779.
[2] Agouridas, V., El Mahdi, O., Diemer, V., Cargoet, M., Monbaliu, J.-C. M., Melnyk, O., *Chem. Rev.* **2019,** *12*, 7328-7443.
[3] Conibear, A. C., Watson, E. E., Payne, R. J., Becker, C. F. W., *Chem. Soc. Rev.* **2018,** *47*, 9046-9068.
[4] Brik, A., Dawson, P. E., Liu, L. (Eds), Total Chemical Synthesis of Proteins, Wiley-VCH, Weinheim, **2021**.
[5] Protein Chemical Synthesis database (PCS-db). http://pcs-db.fr, 2017 (accessed 20 July 2021)
[6] Agouridas, V., El Mahdi, O., Cargoët, M., Melnyk, O., *Bioorg. Med. Chem.* **2017,** *25*, 4938-4945.
[7] Agouridas, V., El Mahdi, O., Melnyk, O., *J. Med. Chem.* **2020,** *63*, 15140-15152.
[8] Dawson, P. E., Churchill, M. J., Ghadiri, M. R., Kent, S. B. H., *J. Am. Chem. Soc.* **1997,** *119*, 4325-4329.
[9] Maguire, O. R., Zhu, J., Brittain, W. D. G., Hudson, A. S., Cobb, S. L., O'Donoghue, A. C., *Chem. Commun.* **2020,** *56*, 6114-6117.
[10] van der Helm, M. P., Klemm, B., Eelkema, R., *Nat. Rev. Chem.* **2019,** *3*, 491-508.
[11] Hupe, D. J., Jencks, W. P., *J. Am. Chem. Soc.* **1977,** *99*, 451-464.
[12] Evans, T. C., Jr., Benner, J., Xu, M. Q., *Protein Sci.* **1998,** *7*, 2256-64.
[13] Huang, Y.-C., Chen , C.-C., Gao, S., Wang, Y.-H., Xiao, H., Wang, F., Tian, C.-L., Li, Y.-M., *Chem. Eur. J.* **2016,** *22*, 7623-7628.
[14] Thompson, R. E., Liu, X., Alonso-Garcia, N., Pereira, P. J., Jolliffe, K. A., Payne, R. J., *J. Am. Chem. Soc.* **2014,** *136*, 8161-4.
[15] Johnson, E. C., Kent, S. B. H., *J. Am. Chem. Soc.* **2006,** *128*, 6640-6.
[16] Tsuda, S., Yoshiya, T., Mochizuki, M., Nishiuchi, Y., *Org. Lett.* **2015,** *17*, 1806-1809.
[17] Wan, Q., Danishefsky, S. J., *Angew. Chem. Int. Ed.* **2007,** *46*, 9248-52.
[18] Moyal, T., Hemantha, H. P., Siman, P., Refua, M., Brik, A., *Chem. Sci.* **2013,** *4*, 2496-2501.
[19] Sakamoto, K., Tsuda, S., Nishio, H., Yoshiya, T., *Chem. Commun.* **2017,** *53*, 12236-12239.
[20] Sakamoto, K., Tsuda, S., Mochizuki, M., Nohara, Y., Nishio, H., Yoshiya, T., *Chem. Eur. J.* **2016,** *22*, 17940-17944.
[21] Chisholm, T. S., Clayton, D., Dowman, L. J., Sayers, J., Payne, R. J., *J. Am. Chem. Soc.* **2018,** *29*, 9020–9024.
[22] Raibaut, L., Ollivier, N., Melnyk, O., *Chem. Soc. Rev.* **2012,** *41*, 7001-7015.
[23] Diemer, V., Bouchenna, J., Kerdraon, F., Agouridas, V., Melnyk, O., N,S- and N,Se-Acyl Transfer Devices in Protein Synthesis. In *Total Chemical Synthesis of Proteins*, Brik, A., Liu, L., Dawson, P., Eds. Wiley: 2021; pp 59-85.
[24] Melnyk, O., Agouridas, V., *Curr. Opin. Chem. Biol.* **2014,** *22*, 137-145.
[25] Ollivier, N., Dheur, J., Mhidia, R., Blanpain, A., Melnyk, O., *Org. Lett.* **2010,** *12*, 5238-5241.
[26] Sato, K., Shigenaga, A., Tsuji, K., Tsuda, S., Sumikawa, Y., Sakamoto, K., Otaka, A., *ChemBioChem* **2011,** *12*, 1840-4.
[27] Kerdraon, F., Bogard, G., Snella, B., Drobecq, H., Pichavant, M., Agouridas, V., Melnyk, O., *Molecules* **2021,** *26*.
[28] Cargoët, M., Diemer, V., Snella, B., Desmet, R., Blanpain, A., Drobecq, H., Agouridas, V., Melnyk, O., *J. Org. Chem.* **2018,** *83*, 12584–12594.
[29] Durek, T., Alewood, P. F., *Angew. Chem. Int. Ed.* **2011,** *50*, 12042-5.
[30] Eto, M., Naruse, N., Morimoto, K., Yamaoka, K., Sato, K., Tsuji, K., Inokuma, T., Shigenaga, A., Otaka, A., *Org. Lett.* **2016,** *18*, 4416-9.
[31] Schmalisch, J., Seitz, O., *Chem. Commun.* **2015,** *51*, 7554-7557.
[32] Seitz, O., *J. Pept. Sci.* **2019,** *25*, e3198.
[33] Giesler, R. J., Erickson, P. W., Kay, M. S., *Curr. Opin. Chem. Biol.* **2020,** *58*, 37-44.
[34] Reinhardt, U., Lotze, J., Zernia, S., Morl, K., Beck-Sickinger, A. G., Seitz, O., *Angew. Chem. Int. Ed.* **2014,** *53*, 10237-41.

Entry for the Table of Contents



Nucleophilic catalysis is key to the acceleration of the native chemical ligation (NCL) reaction and thus, to its application for protein synthesis, functionalization and utilization in chemical biology. The present concept article discusses the fundamental principles of NCL rate and catalysis with emphasis on thiol nucleophiles.

Institute and/or researcher Twitter usernames: @Melnyk_Group; @PcsDb

Annex B

Thiol Catalysis of Selenosulfide Bond Cleavage by a Triarylphosphine

Firstova, O.; Diemer, V.; Melnyk, O.

The ChemRxiv version is presented here

10.26434/chemrxiv-2022-9qmlh

The peer-reviewed version is available by

J. Org. Chem. 2022, 87, 9426-9430

DOI 10.1021/acs.joc.2c00934

Thiol catalysis of selenosulfide bond cleavage by a triarylphosphine

Olga Firstova[a], Oleg Melnyk[a]*, Vincent Diemer[a]*

[a] Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 9017 - CIIL - Center for Infection and Immunity of Lille, F-59000 Lille, France

* Corresponding authors

Abstract

The arylthiol 4-mercaptophenylacetic acid (MPAA) is a powerful catalyst of selenosulfide bond reduction by the triarylphosphine 3,3',3''-phosphanetriyltris(benzenesulfonic acid) trisodium salt (TPPTS). Both reagents are water-soluble at neutral pH and are particularly adapted for working with unprotected peptidic substrates. Contrary to trialkylphosphines such as *tris*(2-carboxyethyl)phosphine hydrochloride (TCEP), TPPTS has the advantage of not inducing deselenization reactions. We believe that the work reported here will be of value for those manipulating selenosulfide bonds in peptidic or protein molecules.

Main

The reduction of dichalcogenides by phosphines is a long-standing reaction that has been the subject of several mechanistic studies. The reductive cleavage of disulfide bonds into thiols by phosphines has concentrated most efforts primarily due to the importance of thiols in synthetic organic chemistry and peptide/protein chemistry (X, Y = S; Fig. 1a).[1-4] In-depth studies from several authors scrutinized the reaction of various disulfides with triphenylphosphine (R_3 = Ph, Fig. 1a).[5-9] These investigations pointed out the critical role played by water in the reduction process,[5-7,9] and the higher propensity of aryl disulfides or mixed alkyl aryl disulfides over alkyl disulfides to be cleaved by triphenylphosphine.

Comparatively, alkyl selenosulfides (R_1, R_2 = Alk, X = S, Y = Se, Fig. 1a) are reduced with difficulty by triphenylphosphine.[10] In practice, their cleavage require more reducing phosphorus compounds such as trialkylphosphines, among which *tris*(2-carboxyethyl)phosphine hydrochloride (TCEP, Fig. 1b) is particularly popular for reactions conducted in aqueous media. However, a limitation of using trialkylphosphines for the reduction of alkyl selenosulfides is their propensity to deselenize the alkyl selenol component.[11-14]

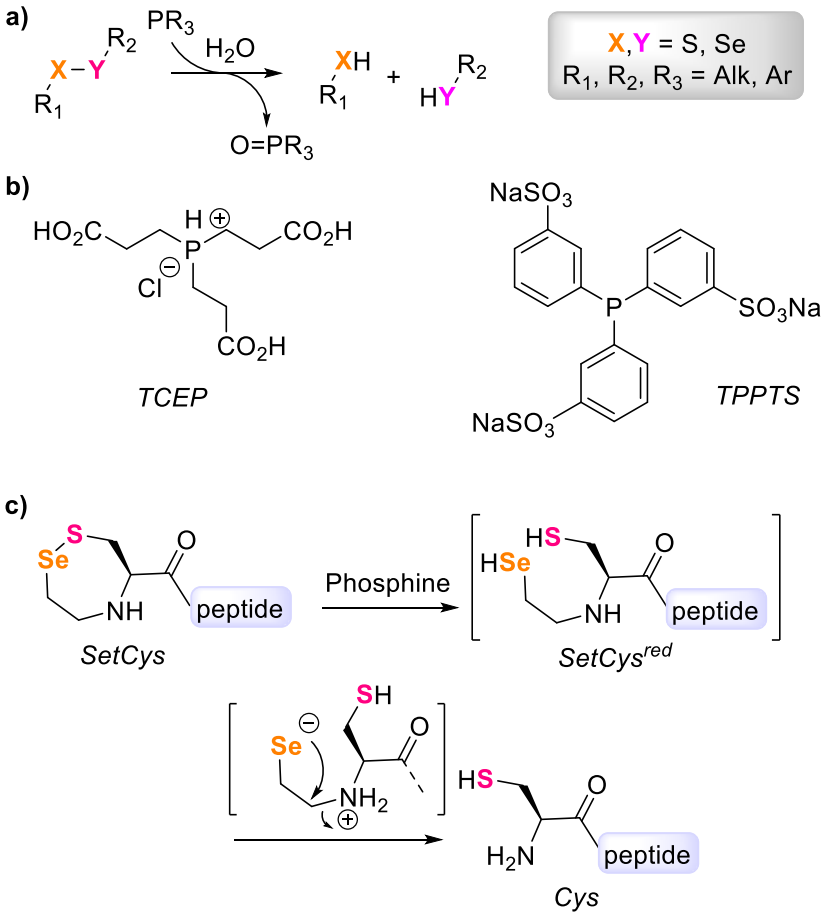


Figure 1. a) Reduction of dichalcogenides by phosphines. b) Water-soluble phosphines studied in this work. c) Conversion of SetCys into Cys upon reduction by a phosphine.

Our interest in the chemistry of selenosulfides comes from the recent discovery that *N*-(2-selenoethyl)cysteine, in the form of its cyclic selenosulfide called SetCys, can be converted into cysteine residue (Cys) in the presence of TCEP (Fig. 1c).[15] The loss of the 2-selenoethyl arm proceeds through the reductive opening of the cyclic selenosulfide bond by the phosphine, which is followed by the spontaneous cleavage of the carbon-nitrogen bond through an intramolecular nucleophilic substitution mechanism. Cysteine surrogates such as SetCys are useful tools for protein chemical synthesis because such chemical devices enable to control the concatenation of peptide segments using the native chemical ligation (NCL[16]). NCL is a widely used method in the field of protein chemical synthesis to connect unprotected peptide segments in water by reacting a C-terminal peptide thioester with a N-terminal cysteinyl peptide (Fig. 2a).[2,17] To achieve synthetic useful rates, NCL is often catalyzed by thiol additives that activate the thioester component of the reaction through thiol-thioester exchange.[18,19] The gold standard catalyst for promoting NCL is 4-mercaptophenyl acetic acid (MPAA),[20] but other thiol additives such as methyl thioglycolate (MTG),[21] or sodium 2-mercaptopropylsulfonate (MPSNa),[22] are also described in some works (Fig. 2b).

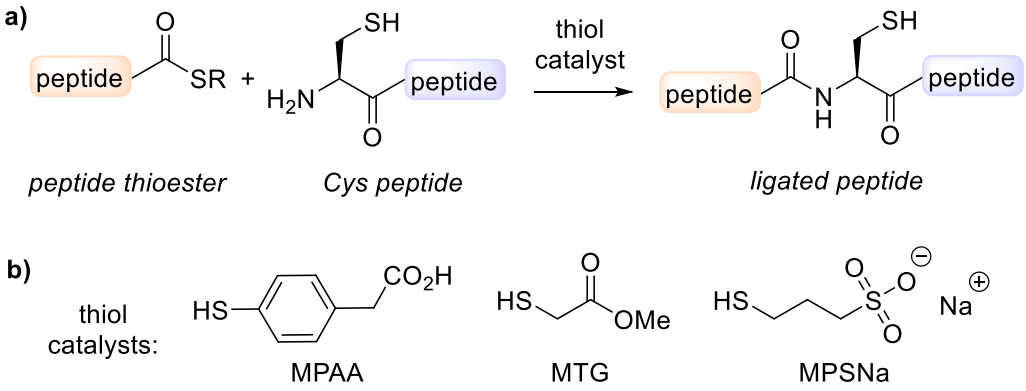


Figure 2. a) NCL reaction. b) Some thiols used as additives during NCL.

Since the conversion of SetCys into Cys can be controlled by the cleavage of the Se-S bond during elaborated peptide assemblies by NCL, it is important to have a full understanding of the SetCys reactivity regarding different types of reductants that might be present in the reaction mixture. We therefore investigated the propensity of SetCys to be converted into Cys in the presence of TCEP or TPPTS (3,3',3''-phosphanetriyltris(benzenesulfonic acid) trisodium salt), two water soluble phosphines that differ markedly by their reductive power (Fig. 1b). Considering the large excesses of thiol additive usually needed for catalyzing NCL, we also examined if their presence has any role in the SetCys conversion into Cys. The studied thiols, MPPA, MTG and MPSNa, cover by their pK_a the range of thiols used for NCL catalysis.

SetCys peptide **1** was used as a model compound for studying the kinetics of SetCys conversion into Cys (Fig. 3a). The results as function of the nature and the concentration of phosphines and thiol additives used are gathered in Fig. 3b-f. All the experiments were performed at pH 6.0 since previous investigations showed that the rate of the intramolecular substitution process leading to the loss of the 2-selenoethyl arm is at its highest at this pH value.[15] Furthermore, sodium ascorbate (100 mM) was systematically added in all reaction mixtures. Sodium ascorbate is a powerful radical scavenger that is mandatory when TCEP is used as the reductant to inhibit the deselenization of SetCys residue,[15] even in the presence of large excesses of MPAA.[23] Interestingly, we found that sodium ascorbate is no longer required when TPPTS is used as the reductant (see Supplementary Information). Nevertheless, sodium ascorbate was used in all experiments to enable comparison of kinetic data. Note also that some lots of TPPTS used in this study were found to contain some traces of formaldehyde. Since TPPTS was used in large excess in our experiments, its contamination by formaldehyde resulted in the formation of a thiazolidine byproduct by reaction with Cys peptide **2** (~ 5-10%). The quality of TPPTS relative to its contamination by formaldehyde can be easily checked as described in the experimental section (see Fig. 6) and Supplementary Information.

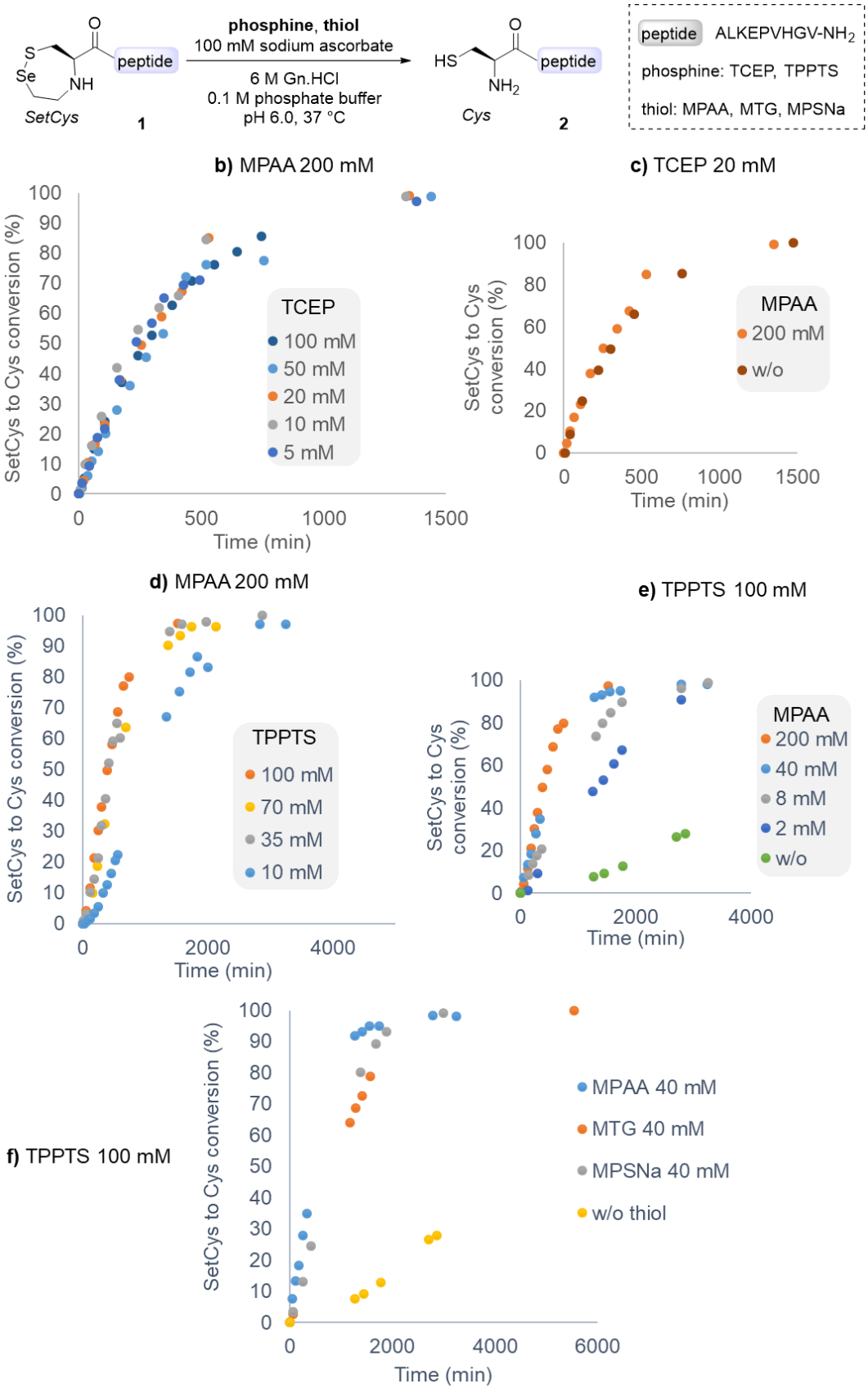


Figure 3. Rate of SetCys conversion into Cys in the presence of varying concentrations of phosphine (TCEP, TPPTS) and thiol (MPAA, MTG, MPSNa). The reactions were monitored at 215 nm by HPLC and/or UPLC-MS.

In the presence of 200 mM MPAA, the rate of SetCys to Cys conversion is independent of TCEP concentration (Fig. 3b). Conversely, at a fixed TCEP concentration of 20 mM, no difference in the rate of SetCys conversion into Cys is observed whether MPAA is present or not (Fig. 3c). In addition, in the presence of TCEP/MPAA, the rate of SetCys conversion into Cys is independent of peptide concentration (see Supplementary Information). Taken together, these data strongly support the mechanism proposed in Fig. 4 path a. According to this mechanism, the reduction of the cyclic selenosulfide by TCEP (STEP 1) is a fast process and the rate of SetCys to Cys conversion dictated by the intramolecular substitution process leading to C-N bond cleavage (STEP 2).

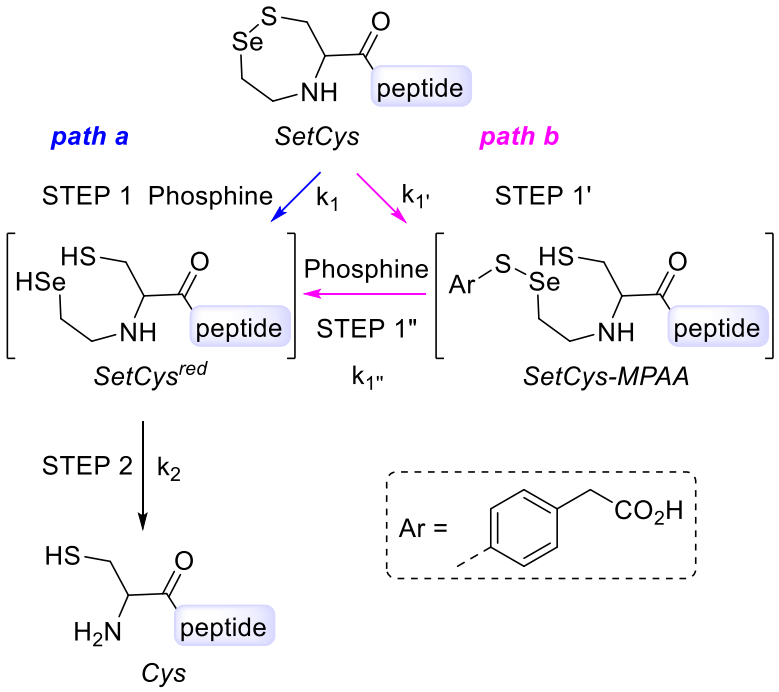


Figure 4. Proposed mechanisms for the conversion of SetCys into Cys upon reaction with a phosphine.

We next examined the influence of TPPTS and MPAA concentration on the rate of SetCys to Cys conversion. In the presence of 200 mM MPAA, the rate of SetCys to Cys conversion increased significantly when TPPTS concentration was increased from 10 to 35 mM and then plateaued (Fig. 3d). This result shows that the reduction process involving TPPTS is no longer rate limiting at concentrations greater than 35 mM. If the reaction proceeded as depicted in Fig. 4 path a, the intramolecular substitution process (STEP 2) would dictate the observed rate. However, the half-reaction time for the SetCys to Cys conversion using 100 mM TPPTS and 200 mM MPAA was below that expected on the basis of the TCEP experiments described above (t½ 6.5 h *vs* t½ 4.6 h). This observation led us to examine the role of MPAA in the reaction. Using 100 mM of TPPTS, we found that the rate of the SetCys to Cys conversion is dramatically dependent on MPAA concentration (Fig. 3e). In absence of MPAA (green dots in Fig. 3e), the t½ is 75.4 h, while it is more than 10-fold higher when MPAA is present at 200 mM (orange dots in Fig. 3e, t½ 6.5 h). Consequently, another step is rate limiting in the presence of TPPTS that is neither the reduction of the SetCys selenosulfide bond by the phosphine nor the intramolecular substitution process, but that involves MPAA.

To sum-up at this stage, the data accumulated so far show that TPPTS can cleave on its own the SetCys selenosulfide bond and trigger the SetCys to Cys conversion, that is through a mechanism corresponding to path a in Fig. 4 as proposed above for TCEP. However and contrary to TCEP, such a direct reduction of the selenosulfide bridge by TPPTS is very slow. The results point out that MPAA is a powerful catalyst of SetCys reductive opening by the phosphine. Although MTG and MPSNa are also able to accelerate SetCys reduction, these thiol additives proved to be significantly less efficient than MPAA (Fig. 3f). Experimentally, adding as low as 2 mM of MPAA enables to accelerate the rate of the process more than 3-fold (green dots in Fig. 3e t½ 75.4 h *vs* dark blue dots t½ 22.3 h). Apparently, a step involving MPAA but not the phosphine is rate limiting at 100 mM TPPTS.

One plausible explanation for the above results is to consider the nucleophilic cleavage of the SetCys cyclic selenosulfide bond by MPAA thiol leading to an acyclic selenosulfide intermediate, i.e. SetCys-MPAA, as being rate limiting (Fig. 4, path b, STEP 1'). The formed SetCys-MPAA is then reduced by the phosphine TPPTS (STEP 1"), an event that unmask the selenol group and triggers the intramolecular substitution process leading to the production of the Cys peptide (STEP 2). Attempts to detect SetCys-MPAA intermediate formation by LC-MS was unsuccessful. However, previous experiments argue for its formation.[15] Typically, incubating for days a SetCys peptide with 200 mM MPAA in the absence of phosphine results in no apparent reaction. However, evidence for the transient formation of SetCys-MPAA intermediate was obtained when a SetCys peptide was treated with MPAA and a peptide thioester, an experiment that enabled the isolation of a ligation product featuring a SetCys residue at the ligation junction (Fig. 5). The formation of such a compound shows that the cyclic structure of SetCys is opened by reduction with MPAA. The resulting SetCys-MPAA intermediate is then trapped *in situ* by the peptide thioester through an NCL-like reaction before SetCys ring closure. A few experimental and computational works examined the nucleophilic attack of thiolates on dialkyl selenosulfides.[24-26] These studies revealed that reaction rates of thiolates at selenium are several orders of magnitude higher than those at sulfur. Such results suggest that the MPAA-induced ring-opening of SetCys would rather lead to an alkyl aryl selenosulfide intermediate, as depicted in Fig. 4 and 5.

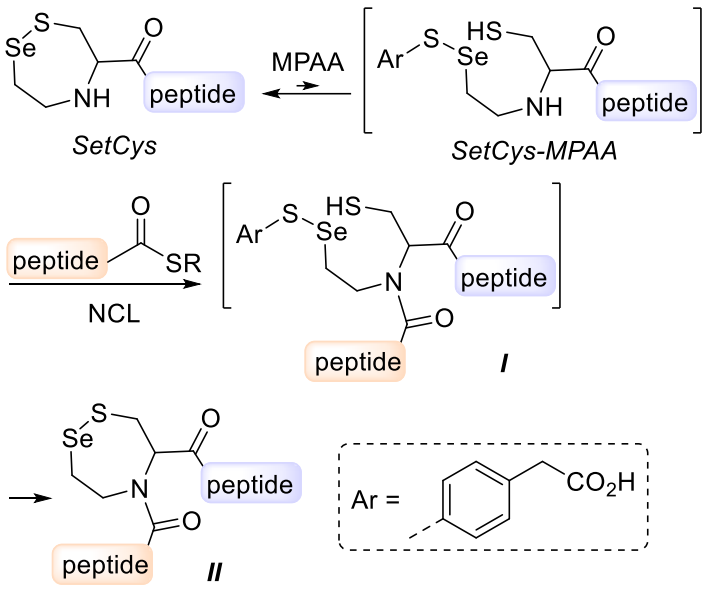


Figure 5. The reaction of a SetCys peptide with MPAA in the presence of a peptide thioester yields ligation product *II*.

In conclusion, the thiophenol derivative MPAA is a powerful catalyst of SetCys selenosulfide bond reduction by the water-soluble triarylphosphine TPPTS. TPPTS is a mild reductant compared to TCEP that has the advantage of not inducing deselenization reactions when combined with MPAA. We believe that the work reported here will be of value for those manipulating selenosulfide bonds in peptidic or protein molecules.

EXPERIMENTAL SECTION

*Synthesis of model SetCys peptide **1***. Peptide **1** was synthesized as previously reported in the literature.[15]

*Conversion of SetCys peptide **1** into Cys peptide **2**. General procedure illustrated with 200 mM MPAA and 100 mM TPPTS.* The reaction is performed under nitrogen atmosphere. To a solution of Gn·HCl (287 mg) in 0.1 M, pH 7.0 sodium phosphate buffer (300 µL) were added sodium ascorbate (9.90 mg, 50.0 µmol, 100 mM), TPPTS (28.4 mg, 50.0 µmol, 100 mM) and MPAA (16.8 mg, 100 µmol, 200 mM) and the pH of the mixture was adjusted to 6.0 by adding 6 M NaOH. SetCys peptide **1** (0.48 mg, 0.32 µmol, 1.00 mM) was then dissolved in the above solution (320 µL) and the reaction mixture was heated at 37 °C. The conversion of SetCys into Cys was monitored by HPLC at 215 nm. For each analysis, an 8 µL aliquot was taken from the reaction mixture and quenched by adding 10% AcOH in water (100 µL). The sample was then extracted with Et_2O to remove MPAA prior to analysis by reversed-phase HPLC or UPLC-MS.

TPPTS quality control

A straitghforward method to check for the presence of formaldehyde in TPPTS (or any other reagent used for converting SetCys into Cys) is to incubate Cys peptide **3** with TPPTS (70 mM) at pH 6.0 and 37 °C and quantify the amount of thiazolidine byproduct **4** formed overnight by UPLC-MS (Fig. 6, see also Supplementary Information).[27] No formaldehyde could be detected in Gn·HCl, MPAA and sodium ascorbate.

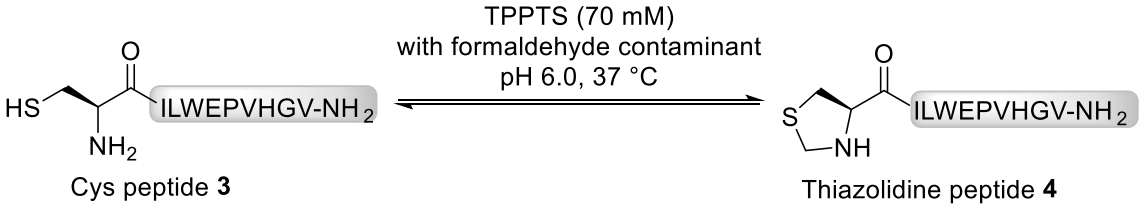


Figure 6. The presence of formaldehyde in TPPTS can be assayed through the propensity of Cys peptides to form stable thiazolidine products upon reaction with aldehydes. The thiazolidine product **4** is detected by UPLC-MS analysis (peptide **3** $[M+H]^+$ calcd av. 1152.4, peptide **4** $[M+H]^+$ calcd av. 1164.4).

ASSOCIATED CONTENT

Supplementary procedures and analysis data (pdf).

AUTHOR INFORMATION

Corresponding Authors

Oleg Melnyk – CNRS

Center of Infection and Immunity of Lille, Lille, France.

orcid.org/0000-0002-3863-561;

Email: Oleg.melnyk@ibl.cnrs.fr

Vincent Diemer – CNRS

Center of Infection and Immunity of Lille, Lille, France.

orcid.org/0000-0002-5735-4630;

Vincent.diemer@ibl.cnrs.fr

Authors

Olga Firstova – CNRS

Center of Infection and Immunity of Lille, Lille, France.

NOTES

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This project has received financial support from the CNRS through the 80|Prime program (grant to O. F.)

REFERENCES

(1) Parker, A. J.; Kharasch, N. The Scission of the Sulfur-Sulfur Bond. *Chem. Rev.* **1959**, *59*, 583-628.

(2) Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoet, M.; Monbaliu, J.-C. M.; Melnyk, O. Native Chemical Ligation and Extended Methods. Mechanisms, Catalysis, Scope and Limitations. *Chem. Rev.* **2019**, *12*, 7328-7443.

(3) Krall, N.; da Cruz, F. P.; Boutureira, O.; Bernardes, G. J. L. Site-Selective Protein-Modification Chemistry for Basic Biology and Drug Development. *Nat. Chem.* **2016**, *8*, 103-113.

(4) Boutureira, O.; Bernardes, G. J. L. Advances in Chemical Protein Modification. *Chem. Rev.* **2015**, *115*, 2174-2195.
(5) Overman, L. E.; O'Connor, E. M. Nucleophilic Cleavage of the Sulfur-Sulfur Bond by Phosphorus Nucleophiles. Iv. Kinetic Study of the Reduction of Alkyl Disulfides with Triphenylphosphine and Water. *J. Am. Chem. Soc.* **1976**, *98*, 771-775.
(6) Overman, L. E.; Petty, S. T. Nucleophilic Cleavage of the Sulfur-Sulfur Bond by Phosphorus Nucleophiles. III. Kinetic Study of the Reduction of a Series of Ethyl Aryl Disulfides with Triphenylphosphine and Water. *J. Org. Chem.* **1975**, *40*, 2779-2782.
(7) Overman, L. E.; Matzinger, D.; O'Connor, E. M.; Overman, J. D. Nucleophilic Cleavage of the Sulfur-Sulfur Bond by Phosphorus Nucleophiles. Kinetic Study of the Reduction of Aryl Disulfides with Triphenylphosphine and Water. *J. Am. Chem. Soc.* **1974**, *96*, 6081-6089.
(8) Humphrey, R. E.; McCrary, A. L.; Webb, R. M. Reduction of Alkyl Disulphides with Triphenylphosphine. *Talanta* **1965**, *12*, 727-732.
(9) Moore, C. G.; Trego, B. R. The Reaction of Triphenylphosphine with Organic Disulphides and Polysulphides. *Tetrahedron* **1962**, *18*, 205-218.
(10) Crich, D.; Krishnamurthy, V.; Hutton, T. K. Allylic Selenosulfide Rearrangement: A Method for Chemical Ligation to Cysteine and Other Thiols. *J. Am. Chem. Soc.* **2006**, *128*, 2544-2545.
(11) Cross, R. J.; Millington, D. Selenium Abstraction from Diethyl Diselenide by Tertiary Phosphines. *J. Chem. Soc. Chem. Commun.* **1975**, 455-456.
(12) Chu, J. Y. C.; Marsh, D. G. Photochemistry of Organochalcogen Compounds. 2. Photochemical Deselenation of Benzyl Diselenide by Triphenylphosphine. *J. Org. Chem.* **1976**, *41*, 3204-3205.
(13) Metanis, N.; Keinan, E.; Dawson, P. E. Traceless Ligation of Cysteine Peptides Using Selective Deselenization. *Angew. Chem. Int. Ed.* **2010**, *49*, 7049-7053.
(14) Dery, S.; Reddy, P. S.; Dery, L.; Mousa, R.; Dardashti, R. N.; Metanis, N. Insights into the Deselenization of Selenocysteine into Alanine and Serine. *Chem. Sci.* **2015**, *6*, 6207-6212.
(15) Diemer, V.; Ollivier, N.; Leclercq, B.; Drobecq, H.; Vicogne, J.; Agouridas, V.; Melnyk, O. A Cysteine Selenosulfide Redox Switch for Protein Chemical Synthesis. *Nat Commun* **2020**, *11*, 2558.
(16) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266*, 776-779.
(17) *Total Chemical Synthesis of Proteins*; Wiley-VCH: Weinheim, 2021.
(18) Diemer, V.; Firstova, O.; Agouridas, V.; Melnyk, O. Pedal to the Metal: The Homogeneous Catalysis of the Native Chemical Ligation Reaction. *Chemistry* **2022**, e202104229.
(19) Dawson, P. E.; Churchill, M. J.; Ghadiri, M. R.; Kent, S. B. H. Modulation of Reactivity in Native Chemical Ligation through the Use of Thiol Additives. *J. Am. Chem. Soc.* **1997**, *119*, 4325-4329.
(20) Johnson, E. C.; Kent, S. B. H. Insights into the Mechanism and Catalysis of the Native Chemical Ligation Reaction. *J. Am. Chem. Soc.* **2006**, *128*, 6640-6646.
(21) Huang, Y.-C.; Chen , C.-C.; Gao, S.; Wang, Y.-H.; Xiao, H.; Wang, F.; Tian, C.-L.; Li, Y.-M. Synthesis of L- and D-Ubiquitin by One-Pot Ligation and Metal-Free Desulfurization. *Chem. Eur. J.* **2016**, *22*, 7623-7628.
(22) Ayers, B.; Blaschke, U. K.; Camarero, J. A.; Cotton, G. J.; Holford, M.; Muir, T. W. Introduction of Unnatural Amino Acids into Proteins Using Expressed Protein Ligation. *Biopolymers* **1999**, *51*, 343-354.
(23) Rohde, H.; Schmalisch, J.; Harpaz, Z.; Diezmann, F.; Seitz, O. Ascorbate as an Alternative to Thiol Additives in Native Chemical Ligation. *ChemBioChem* **2011**, *12*, 1396-1400.
(24) Steinmann, D.; Nauser, T.; Koppenol, W. H. Selenium and Sulfur in Exchange Reactions: A Comparative Study. *J. Org. Chem.* **2010**, *75*, 6696-6699.
(25) Lothrop, A. P.; Snider, G. W.; Ruggles, E. L.; Patel, A. S.; Lees, W. J.; Hondal, R. J. Selenium as an Electron Acceptor During the Catalytic Mechanism of Thioredoxin Reductase. *Biochemistry* **2014**, *53*, 654-663.

(26) Bachrach, S. M.; Demoin, D. W.; Luk, M.; Miller, J. V. Nucleophilic Attack at Selenium in Diselenides and Selenosulfides. A Computational Study. *J. Phys. Chem. A* **2004**, *108*, 4040-4046.
(27) Liu, C. F.; Tam, J. P. Peptide Segment Ligation Strategy without Use of Protecting Groups. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 6584-6588.

Supporting Information for

Thiol catalysis of selenosulfide bond cleavage by a triarylphosphine

Olga Firstova, Oleg Melnyk*, Vincent Diemer*

Univ. Lille, CNRS, INSERM, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 9017 - CIIL - Center for Infection and Immunity of Lille, F-59000 Lille, France

Corresponding authors:

Dr Oleg Melnyk, E-mail: oleg.melnyk@ibl.cnrs.fr

Dr Vincent Diemer, E-mail : vincent.diemer@ibl.cnrs.fr

Table of contents

1. Instrumentation

Analytical HPLC as well as micro preparative HPLC were performed with a Thermofisher system on a reverse phase column XBridge BEH300 C18 (3.5 µm, 300 Å, 4.6 × 150 mm) using a linear gradient of increasing concentration of eluent B in eluent A (eluent A: 0.1% by vol. of TFA in water; eluent B: 0.1% vol. of TFA in acetonitrile or in acetonitrile/water 4/1 v/v). The column eluant was monitored by UV at 215 nm.

Purified products as well as reaction mixtures were characterized by analytical LC-MS (Waters 2695 LC/ZQ 2000 quadripole) on a reverse phase column XBridge BEH300 C18 (3.5 µm, 300 Å, 4.6 × 150 mm) using a linear gradient of increasing concentration of eluent B in eluent A. The column eluate was monitored by UV at 215 nm, evaporative light scattering (ELS, waters 2424) and electrospray ionization mass spectrometry (ESI-MS). Using similar eluents, analytical UPLC-MS analyses were performed on System Ultimate 3000 UPLC (Thermofisher) equipped with a column Acquity peptide BEH300 C18 (1.7 µm, 2.1 × 100 mm), a diode array detector, a charged aerosol detector (CAD) and a mass spectrometer (Ion trap LCQfleet).

MALDI-TOF mass spectra were recorded with a Bruker Autoflex Speed using alpha cyano 4-hydroxycinnaminic acid, sinapinic acid or 2,5-dihydroxybenzoic acid (DHB) as matrix.

2. Characterization of peptides **1** and **2**

Peptide **1** was synthesized as previously reported in the literature [1] and was characterized by LC-MS and MALDI-TOF MS analyses (Figure S 1 and Figure S 2).

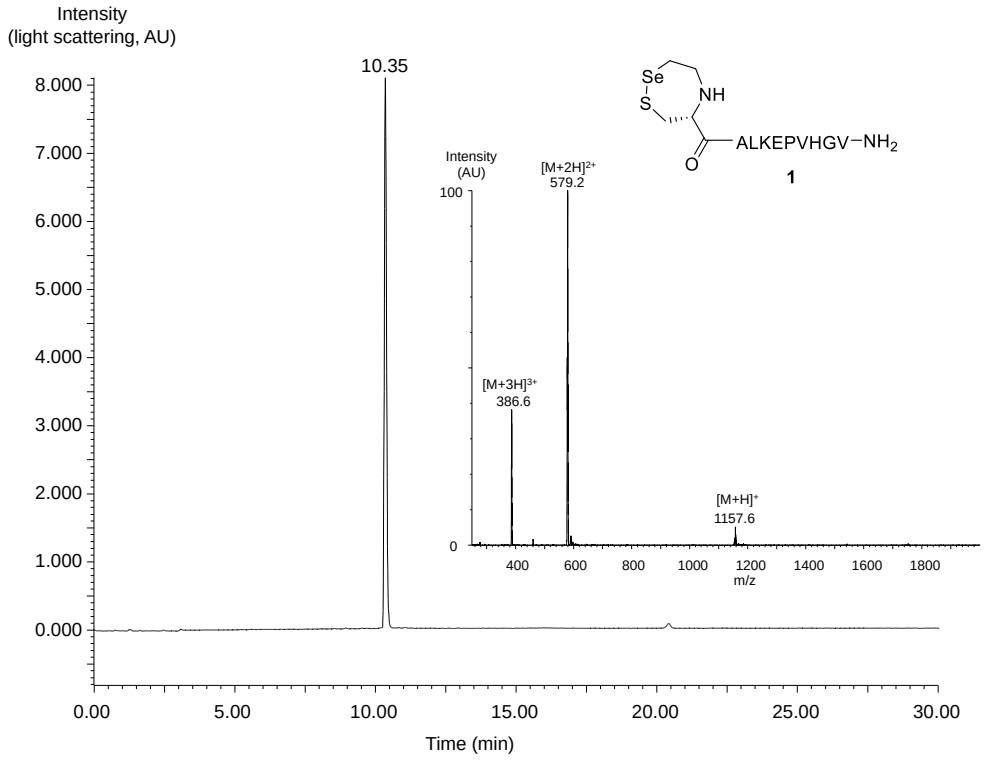


*Figure S 1. LC-MS analysis of SetCys-ALKEPVHGV-NH_2 peptide **1**. LC trace: eluent A 0.10% TFA in water, eluent B 0.10% TFA in CH_3CN/water 4/1 v/v. C18 column, gradient 0-100% B in 30 min, 30 °C, 1 mL/min, light scattering detection. MS trace: m/z calcd. for $[M+H]^+$ (monoisotopic mass): 1157.5, found: 1157.6.*

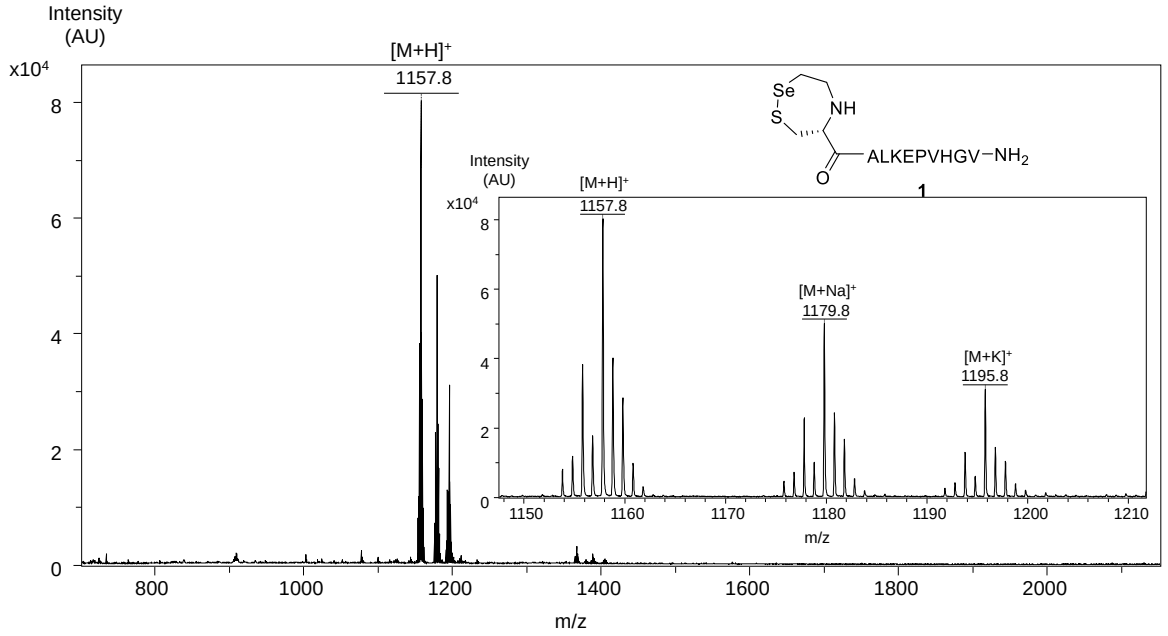


*Figure S 2. MALDI-TOF MS analysis of SetCys-ALKEPVHGV-NH_2 peptide **1**. Matrix α-cyano-4-hydroxycinnamic acid, positive detection mode, calcd. for $[M+H]^+$ (monoisotopic): 1157.5, found: 1157.8.*

The Cys peptide **2** [1] obtained by decomposition of peptide **1** was isolated by micro preparative HPLC (C18 column, eluent A 0.10% TFA in water, eluent B 0.10% TFA in acetonitrile, gradient 0-50% B in 15 min, 1 mL/min, 50 °C) and was characterized by UPLC-MS analysis (Figure S 3).

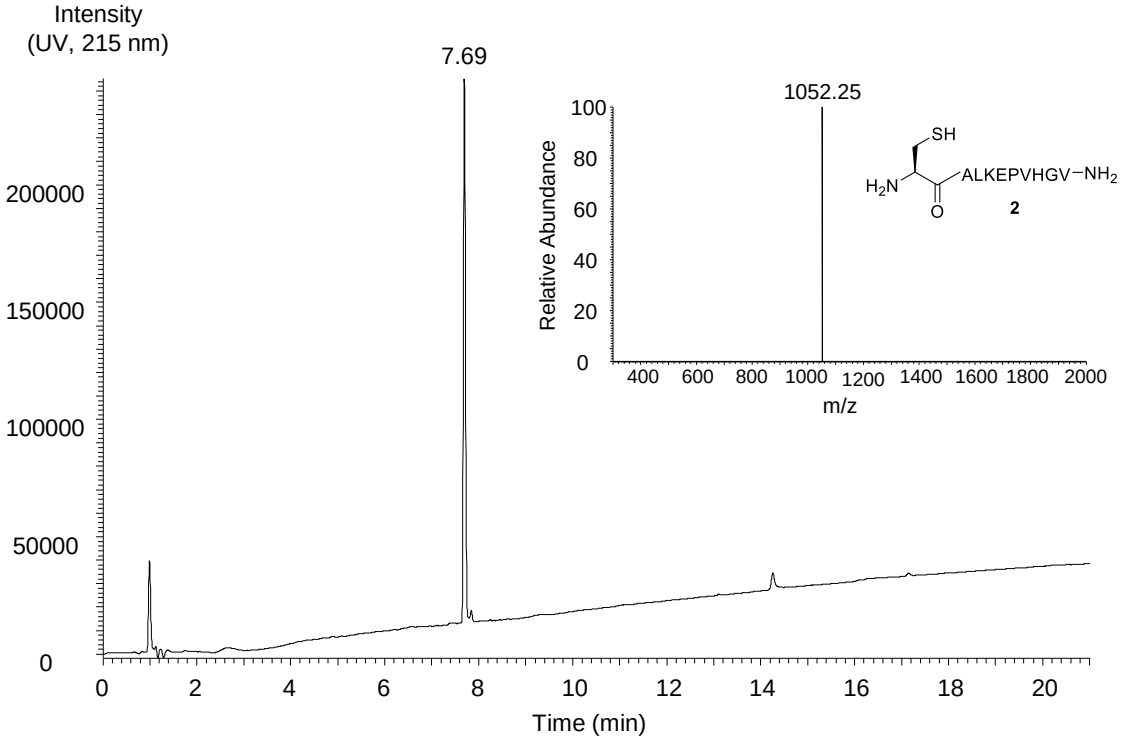


*Figure S 3. UPLC-MS analysis of Cys-ALKEPVHGV-NH_2 peptide **2**. LC trace: eluent A 0.10% TFA in water, eluent B 0.10% TFA in CH_3CN, C18 column, gradient 0-70% B in 20 min, 50 °C, 0.4 mL/min, detection at 215 nm. MS trace: m/z calcd. for $[M+H]^+$ (average mass): 1152.28, found: 1152.25.*

3. Effect of sodium ascorbate on the TPPTS-induced conversion of SetCys into Cys

TPPTS-induced conversion of SetCys into Cys in the presence of sodium ascorbate

As depicted in Figure S 4, two concentrations of TPPTS (35 mM and 70 mM) were tested to study the SetCys to Cys conversion in the presence of sodium ascorbate.

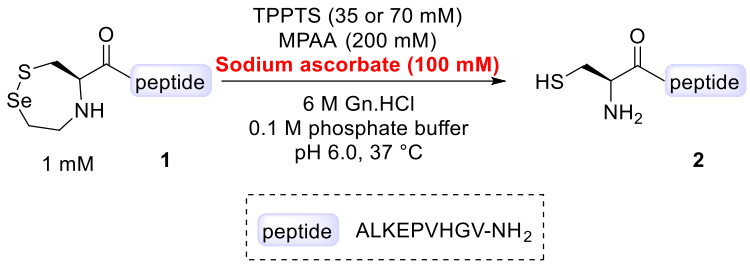
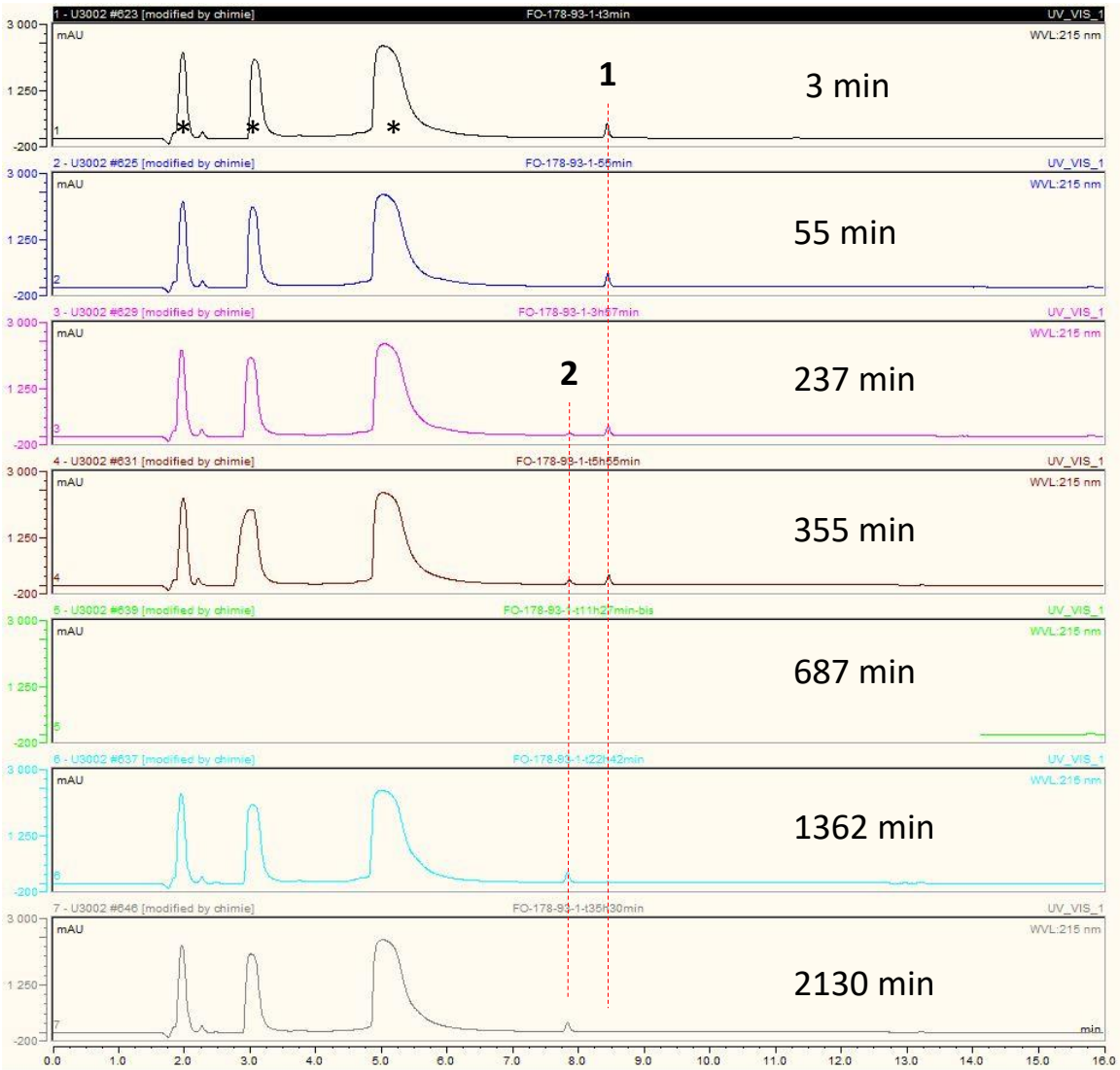


Figure S 4. Reactions studied.

The two reaction were performed according to the general procedure described in the experimental section. Figure S 5 provides the HPLC monitoring for a 70 mM concentration of TPPTS.

A)



B)

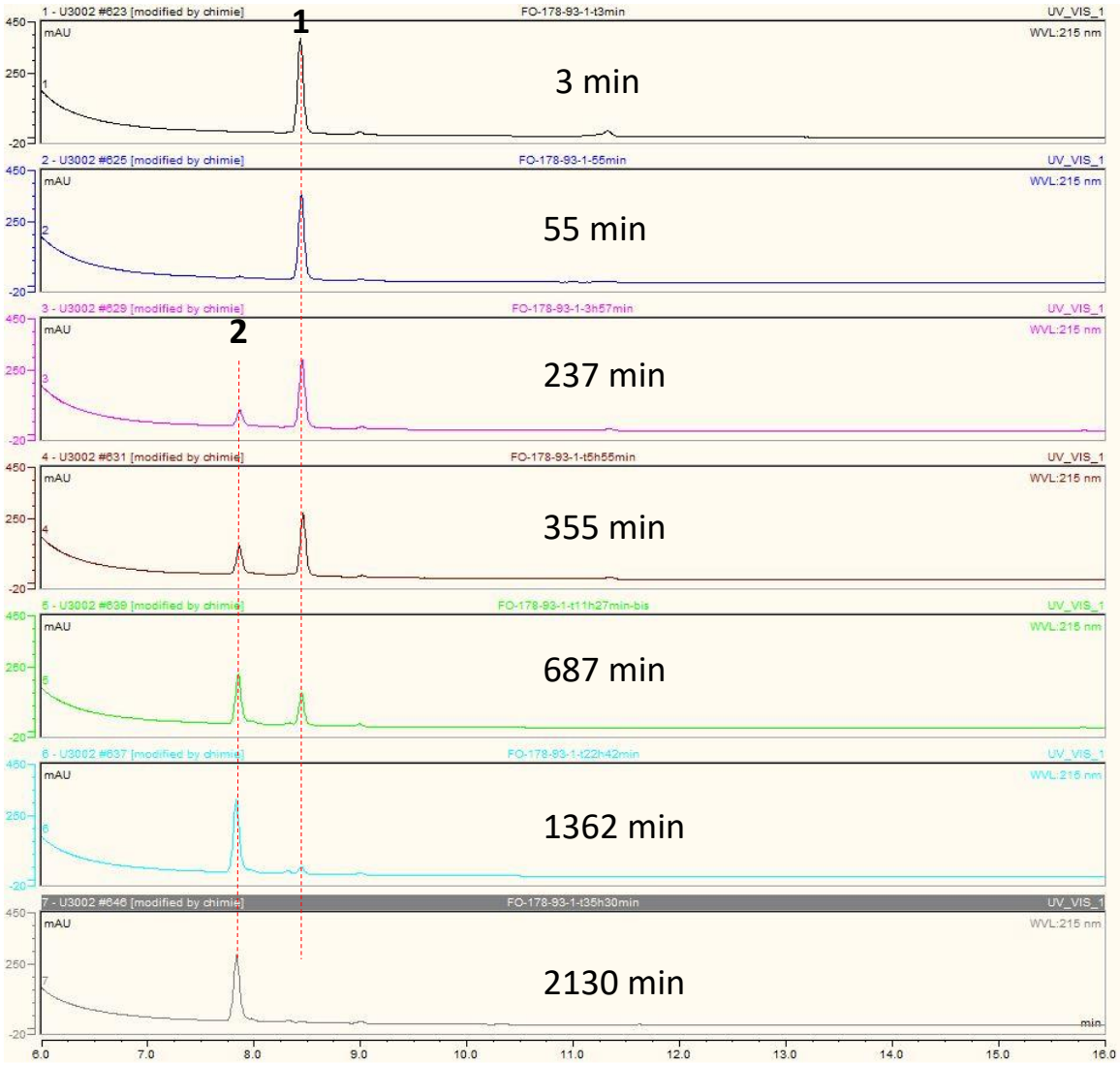


*Figure S 5. Conversion of SetCys peptide **1** into Cys peptide **2** in the presence of TPPTS (70 mM), MPAA (200 mM) and sodium ascorbate (100 mM) in sodium phosphate buffer (0.1 M), 6 M Gn·HCl, pH 6.0, 37 °C. A) HPLC chromatograms, *: non peptidic, B) Zoom of the peptidic signals.*

TPPTS-induced conversion of SetCys into Cys in the absence of sodium ascorbate

The following conditions (Figure S 6) were tested to study the SetCys to Cys conversion in the absence of sodium ascorbate.

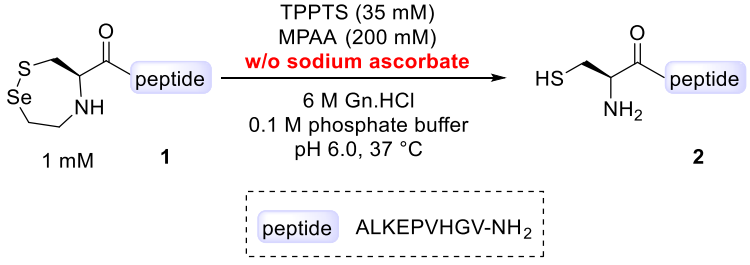
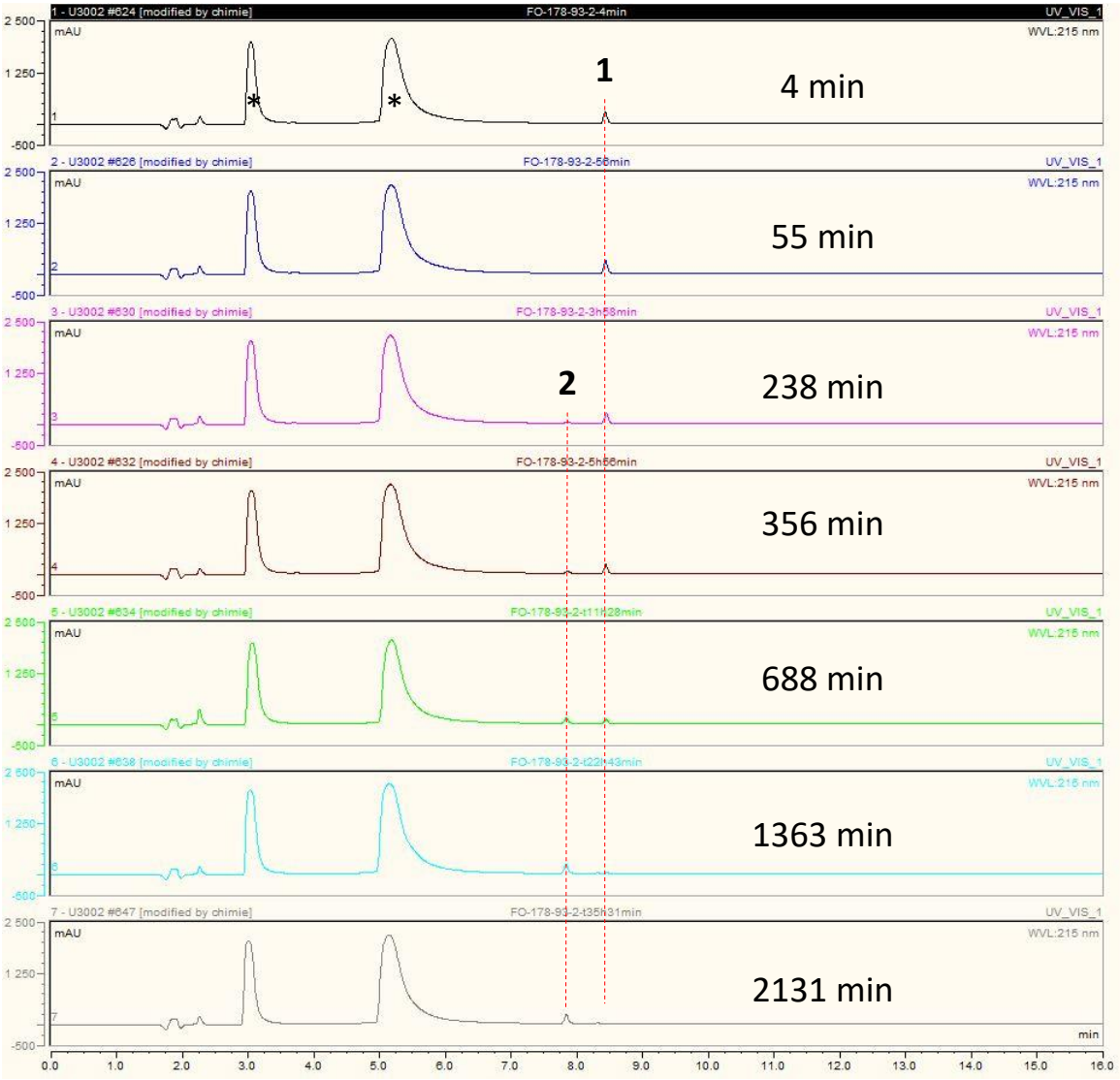


Figure S 6. Reaction studied.

The reaction was performed under nitrogen atmosphere. To a solution of Gn·HCl (287 mg, 6 M) in 0.1 M, pH 7.0 sodium phosphate buffer (300 µL) were added TPPTS (9.94 mg, 17.5 µmol, 35 mM) and MPAA (16.8 mg, 100 µmol, 200 mM) and the pH of the mixture was adjusted to 6.0 by adding 6 M NaOH. SetCys peptide **1** (0.48 mg, 0.32 µmol, 1.00 mM) was then dissolved in the above solution (320 µL) and the reaction mixture was heated at 37 °C using a hotplate stirrer equipped with a heating block. The conversion of SetCys into Cys was monitored by HPLC at 215 nm. For each analysis, a 4 µL aliquot was taken from the reaction mixture and quenched by adding 10% AcOH in water (100 µL). The sample was then extracted with Et_2O to remove MPAA prior to HPLC analysis.

Figure S 7 provides the HPLC monitoring of the reaction.

A)



B)

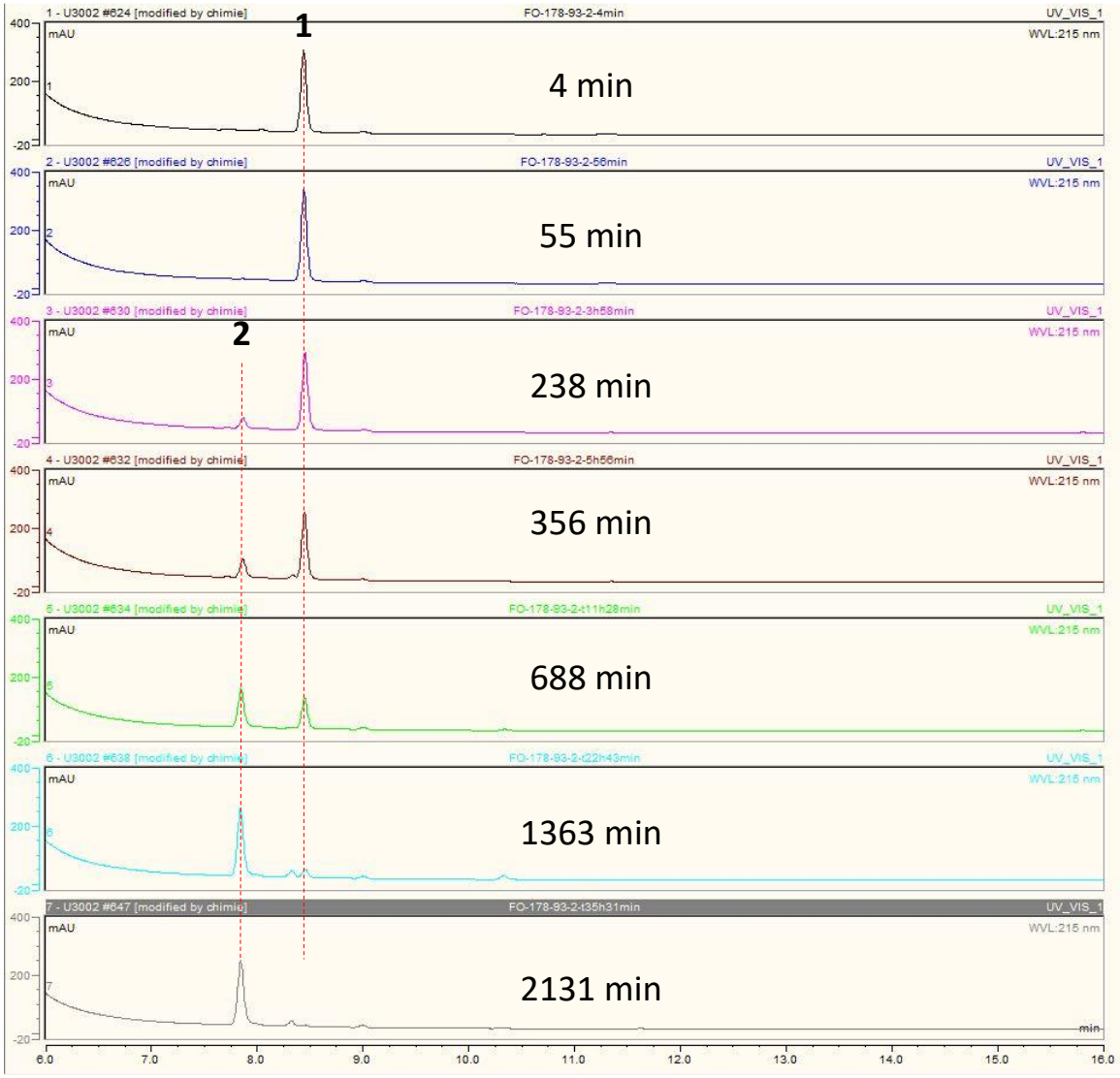


*Figure S 7. Conversion of SetCys peptide **1** into Cys peptide **2** in the presence of TPPTS (35 mM) and MPAA (200 mM) in sodium phosphate buffer (0.1 M), 6 M Gn·HCl, pH 6.0, 37 °C. A) HPLC chromatograms, *: non peptidic, B) Zoom of the peptidic signals.*

When TPPTS was used as reductant, no significant deselenization of the SetCys residue (leading to the *N*-ethyl cysteine) was detected in the reaction mixture, even in the absence of sodium ascorbate.

Note also that the addition of sodium ascorbate in the reaction mixture has no significant effect on the rate of the SetCys to Cys conversion (Figure S 8).

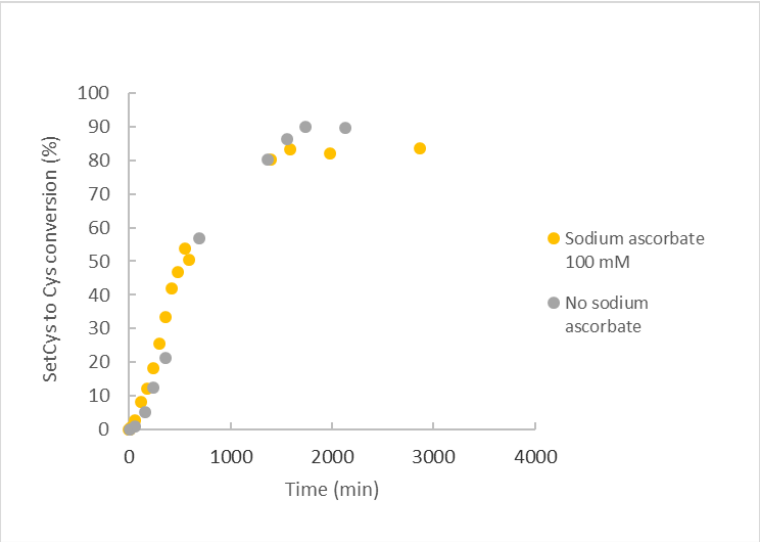


Figure S 8. TPPTS-induced conversion of SetCys into Cys in the absence and in the presence of sodium ascorbate. Reaction conditions: [MPAA] = 200 mM, [TPPTS] = 35 mM, pH 6.0, 37 °C.

4. Effect of the SetCys peptide concentration on the TCEP-induced conversion of SetCys into Cys

The conditions depicted in Figure S 9 were tested to study the effect of the SetCys peptide concentration on the TCEP-induced conversion of SetCys into Cys.

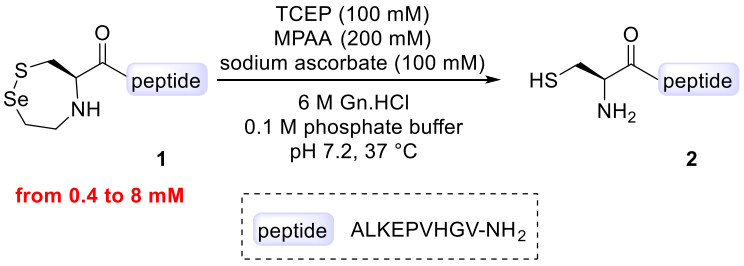


Figure S 9. Reactions studied.

*General procedure (peptide **1** at 0.4 mM).* The reaction is performed under nitrogen atmosphere. To a solution of Gn·HCl (1.15 g) in 0.1 M, pH 7.0 sodium phosphate buffer (1.20 mL) were added sodium ascorbate (39.6 mg, 200 µmol, 100 mM), TCEP·HCl (57.2 mg, 200 µmol, 100 mM) and MPAA (67.2 mg, 400 µmol, 200 mM) and the pH of the mixture was adjusted to 7.2 by adding 6 M NaOH. SetCys peptide **1** (0.95 mg, 0.64 µmol, 0.40 mM) was then dissolved in the above solution (1.58 mL) and the reaction mixture was heated at 37 °C using a hotplate stirrer equipped with a heating block. The conversion of SetCys into Cys was monitored by HPLC at 215 nm. For each analysis, a 20 µL aliquot was taken from the reaction mixture and quenched by adding 10% AcOH in water (100 µL). The sample was then extracted with Et_2O to remove MPAA prior to HPLC analysis.

Kinetics of SetCys to Cys conversion at 0.4, 1 and 8 mM concentration of SetCys peptide **1** are given in Figure S 10.

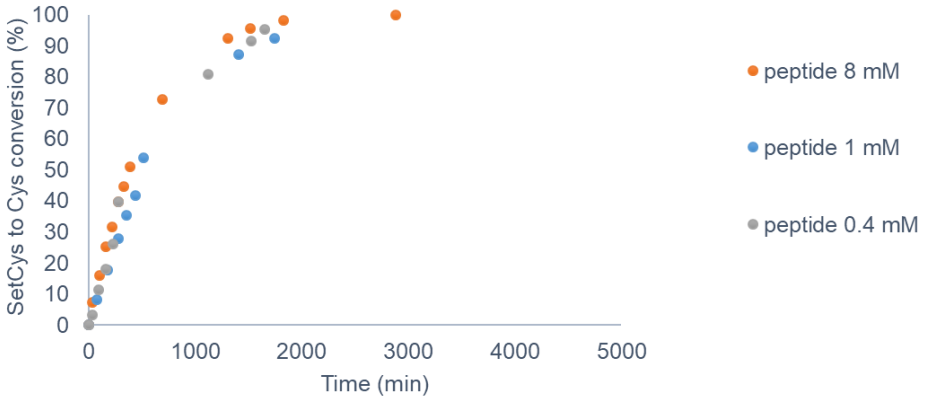


Figure S 10. Effect of the SetCys peptide concentration on the TCEP-induced conversion of SetCys into Cys. Reaction conditions: [MPAA] = 200 mM, [TCEP] = 100 mM, [ascorbate] = 100 mM, pH 7.2, 37 °C.

5. Effect of the pH on the TPPTS-induced conversion of SetCys into Cys

The conditions depicted in Figure S 11 were tested to study the effect of the pH on the TPPTS-induced conversion of SetCys into Cys.

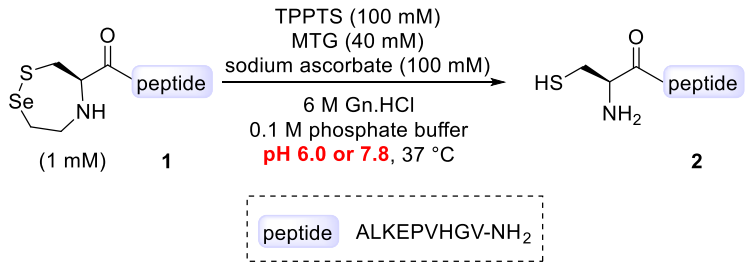


Figure S 11. Reactions studied.

Procedure for pH 7.8. The reaction was performed under nitrogen atmosphere. To a solution of Gn·HCl (115 mg) in 0.1 M, pH 7.0 sodium phosphate buffer (120 µL) were added sodium ascorbate (3.96 mg, 20.0 µmol, 100 mM), TPPTS (11.4 mg, 20.0 µmol, 100 mM) and MTG (0.72 µL, 8.0 µmol, 40 mM) and the pH of the mixture was adjusted to 7.8 by adding 6 M NaOH. SetCys peptide **1** (0.253 mg, 0.169 µmol, 1.00 mM) was then dissolved in the above solution (169 µL) and the reaction mixture was heated at 37 °C using a hotplate stirrer equipped with a heating block. The conversion of SetCys into Cys was monitored by HPLC at 215 nm. For each analysis, a 4 µL aliquot was taken from the reaction mixture and quenched by adding 10% AcOH in water (100 µL) prior to HPLC analysis.

Kinetics of SetCys to Cys conversion at pH 6.0 and 7.8 are given in Figure S 12.

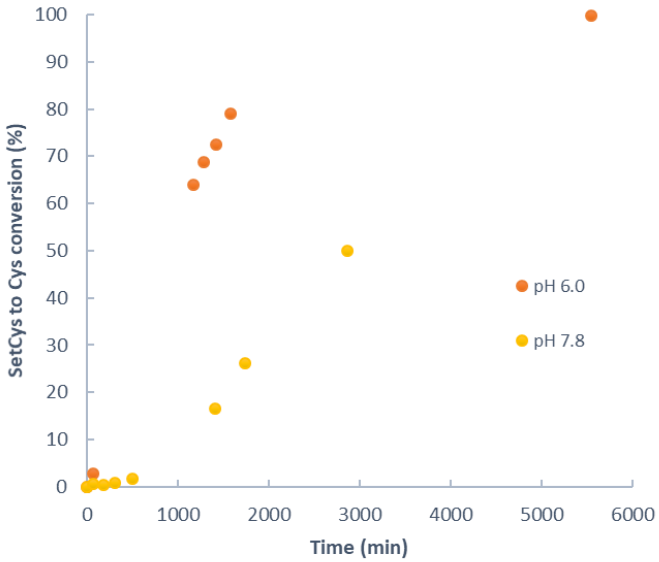


Figure S 12. Effect of the pH on the TPPTS-induced conversion of SetCys into Cys. Reaction conditions: [MTG] = 40 mM, [TPPTS] = 100 mM, [ascorbate] = 100 mM, 37 °C.

6. TPPTS quality control

As depicted in Figure S 13, the presence of formaldehyde in TPPTS can be assayed through the propensity of Cys peptide **3** to form stable thiazolidine products upon reaction with aldehydes. The thiazolidine product **4** is detected by UPLC-MS analysis (peptide **3** $[M+H]^+$ calcd av. 1152.4, peptide **4** $[M+H]^+$ calcd av. 1164.4).

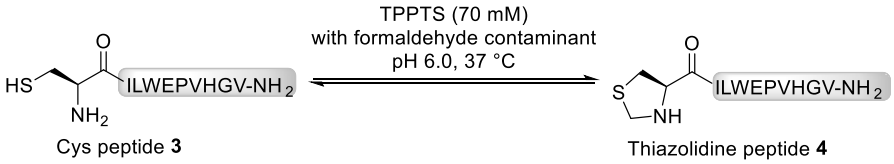
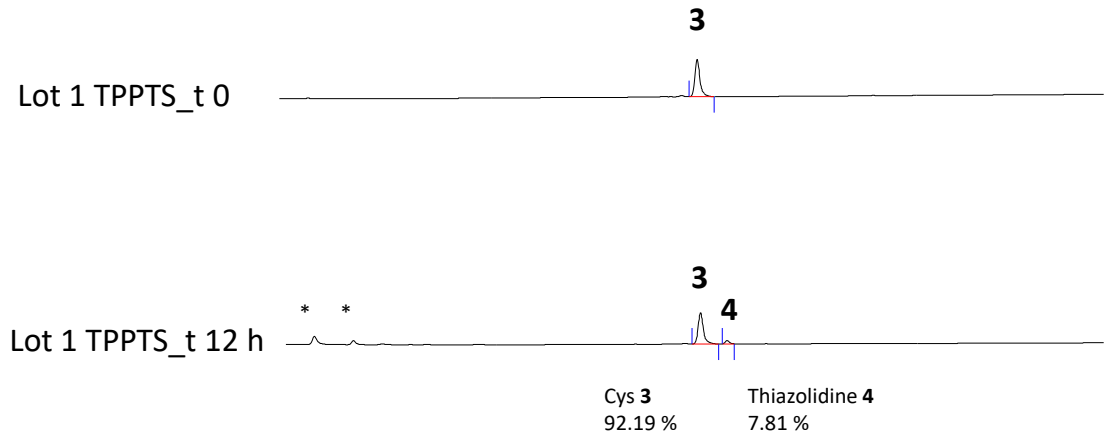


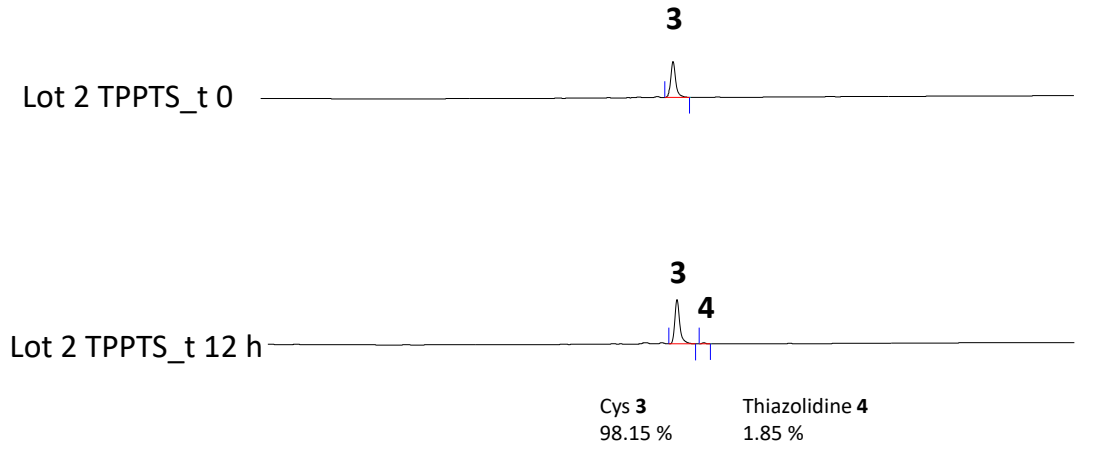
Figure S 13. Reaction used to detect formaldehyde contaminant contained in TPPTS.

General procedure. Cys-ILWEPVHGV peptide **3** (1 mM final concentration) was dissolved in sodium phosphate buffer (0.1 M, pH 6.0) containing TPPTS (70 mM final concentration). The solution was incubated at 37 °C and analyzed by UPLC-MS (Figure S 14). For each analysis, a 4 µL aliquot was taken from the reaction mixture and quenched by adding 10% AcOH in water (100 µL). The samples were injected in UPLC-MS system (27 µL).

A)



B)



C)

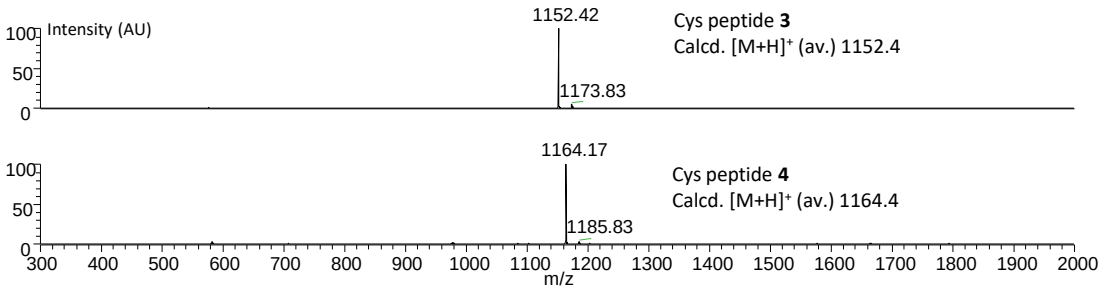


*Figure S 14. Analysis of thiazolidine by-product formation after 12 h. A) HPLC chromatogram obtained with TPPTS lot 1 (Sigma Aldrich). B) HPLC chromatogram obtained with TPPTS lot 2 (TCI). C) MS traces for **3** and **4**. *: non peptidic.*

7. References

[1] Diemer, V.; Ollivier, N.; Leclercq, B.; Drobecq, H.; Vicogne, J.; Agouridas, V.; Melnyk, O. A Cysteine Selenosulfide Redox Switch for Protein Chemical Synthesis. *Nat Commun* **2020**, *11*, 2558.

Annex C

Experimental part for peptides ligation, SetCys application and antifreeze protein synthesis

Content

1 General information

1.1 Equipment

Conventional laboratory glassware.

Microfuge tubes (0.5, 1.5 mL and 5 mL safe-lock tubes), plastic tubes for centrifugation (15 mL and 50 mL).

Glove box equipped with a block heater (37 °C), a set of adjustable pipettes (0.5–10, 2–20, 20–200, 100–1,000 µL), corresponding pipette tips, a pH-meter and a vortex shaker.

Argon gas (O_2 < 0.01 ppm, H_2O < 0.02 ppm).

Automated peptide synthesizer without microwaves.

Rotatory evaporator.

TLC UV cabinet (254 nm).

TLC plates (Silica gel 60 F254 on aluminium sheets).

1.2 Reagents and solvents

Deionized water (≥ 18 MΩ.cm).

O-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HCTU) and *N*-Fmoc protected amino acids were obtained from Iris Biotech GmbH. Side-chain protecting groups used for the amino acids were Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(O*t*Bu)-OH, Fmoc-Glu(O*t*Bu)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Pro-OH, Fmoc-Phe-OH, Fmoc-Ser(*t*Bu)-OH, Fmoc-Thr(*t*Bu)-OH, Fmoc-Tyr(*t*Bu)-OH, Fmoc-Val-OH, Fmoc-Cys(S*t*Bu)-OH, Fmoc-Cys(Trt)-OH, Fmoc-Ala-Thr[psi(Me,Me)Pro]-OH. Resins Rink amide (NovaSyn TGT), SEA-PS are purchased from Novabioche and X'ProChem respectively.

4-mercaptophenylacetic acid (MPAA), 3-mercaptopropionic acid (MPA), methyl thioglycolate (MTG), *tris*(2-carboxyethyl)phosphine hydrochloride (TCEP), imidazole, guanidine hydrochloride (Gdn·HCl) were purchased from Sigma-Aldrich. 3,3′,3″-Phosphanetriyltris (benzenesulfonic acid) trisodium salt (TPPTS) was purchased from TCI Chemicals. Starting materials for the synthesis of SetCys amino acid, and other reagents were purchased from Acros Organics or Merck and were of the purest grade available.

High purity grade solvents for classical organic synthesis and peptide chemistry: acetonitrile (ACN), chloroform ($CHCl_3$), cyclohexane (Cy), 1,2-dichloroethane (DCE), dichloromethane (DCM), dimethyl sulfoxide (DMSO), diethyl ether (Et_2O), *N*,*N*-dimethylformamide (DMF) for peptide synthesis, dioxane, ethanol (EtOH), ethyl acetate (AcOEt), *n*-heptanes, *n*-hexanes, methanol (MeOH), 1-methyl-2-pyrrolidinone (NMP).

High purity grade reagents for classical organic synthesis and peptide chemistry: sodium bicarbonate ($NaHCO_3$), sodium chloride (NaCl), potassium carbonate (K_2CO_3), hydrochloric acid (HCl, 37%), sodium hydroxide pellets (NaOH), anhydrous magnesium sulfate ($MgSO_4$).

Silica gel (60-200 µm, 60 Å).

2 Health and safety recommendations

All organic solvents and reagents used in this protocol should be handled inside a chemical fume hood with appropriate personal protective equipment (lab coat, gloves and protective glasses). Solvents were removed using a rotary evaporator placed inside a fume hood. Particular care must be taken during the manipulation of selenium containing reagents and synthetic intermediates due to their potential toxicity. Trifluoroacetic acid is strongly corrosive and toxic. Therefore, it should be handled with the greatest attention.

3 Analysis and purifications

Unless otherwise stated, the eluents for all analyses and purifications of peptides and proteins by liquid chromatography are: eluent A = 0.1% TFA in H_2O; eluent B = 0.1% TFA in ACN.

3.1 UPLC-MS analyses

UPLC–MS analyses were performed on a LC system equipped with a C18 reverse-phase column (2.1 × 100 mm, pore size: 300 Å, particle size: 1.7 µm) or C3 column (2.1 × 100 mm, pore size: 300 Å, particle size: 1.8 µm), a diode array detector and a mass spectrometer (Ion trap LCQfleet). The mass detection was performed at Ionization mode: ES^+, m/z range 350–2040, capillary voltage 3 kV, cone voltage 30 V, extractor voltage 3 V, RF lens 0.2 V, source temperature 120 °C, dessolvation temperature 350 °C.

3.2 Analytical HPLC analyses

The reactions were monitored by analytical HPLC (Dionex) on a reverse phase column XBridge BEH300 C18 (3.5 µm, 300 Å, 2.1 × 150 mm) or reverse phase column ZORBAX 300 BEH300-C3 (3.5 µm, 300 Å, 2.1 × 150 mm). The column eluate was monitored by UV at 215 and 280 nm.

3.3 Preparative HPLC

Purification of crude peptides were performed with a preparative HPLC Gilson or Buchi system. For the loading scale within 30 mg, a reverse phase column XBridge BEH300 Prep C18 (5 µm, 300 Å, 10 × 250 mm) or ZORBAX 3005B-C3 (5 µm, 300 Å, 9.4 × 250 mm) and appropriate linear gradient of increasing concentration of eluent B in eluent A (flow rate of 6 mL/min) are used. For the loading scale between 30 and 100 mg, a reverse phase column XBridge Prep C18 (5 µm, 19 × 150 mm) and appropriate linear gradient of increasing concentration of eluent B in eluent A (flow rate of 20 mL/min) are used. Selected peptide fractions were then combined and lyophilized.

3.4 MALDI-TOF analyses

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

3.5 NMR analysis

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

4 Buffer preparation

0.1 M phosphate buffer for ligations

0.1 M, pH 7.4 phosphate buffer was prepared by dilution of a 0.2 M stock solution.[1] The latter was obtained by mixing a 0.2 M solution of $NaH_2PO_4{\cdot}2H_2O$ in water (19 mL) and a 0.2 M solution of $Na_2HPO_4{\cdot}2H_2O$ (81 mL).
0.2 M solution of $NaH_2PO_4{\cdot}2H_2O$ in water: 3.12 g of $NaH_2PO_4{\cdot}2H_2O$ in 100 mL of water.
0.2 M solution of $Na_2HPO_4{\cdot}2H_2O$ in water: 7.12 g of $Na_2HPO_4{\cdot}2H_2O$ in 200 mL of water.
It is recommended to stock at 4 °C for 6 months.

Phosphate buffered saline 0.1 M

Phosphate buffered saline was prepared from commercially available powdered salts by dissolving in pure water. One powder standard must be dissolved in 1 L of water giving the buffer at pH 7.4. The pH was corrected if needed.

Hydrochloric acid (HCl) 6 M

6 M solution of HCl in water was prepared by diluting 37% HCl (10 mL) with deionised water (final volume: 20 mL).

Sodium hydroxide (NaOH) 6 M preparation

6 M solution of NaOH in water was prepared by dissolving NaOH (4.8 g) in deionised water (final volume: 20 mL).

5 Synthesis of Fmoc-protected SetCys amino acid

The synthetic strategy proposed for the preparation of the Fmoc-protected SetCys amino acid **6** involves a convergent approach with a total of five steps from the affordable and commercially available cysteine methyl ester hydrochloride **11**. After the protection of Cys thiol side chain, intermediate **12** undergoes a reductive alkylation to install the selenated appendage on the α-amino group. Amine protection by a Fmoc group is followed by a tandem oxidative deprotection/cyclization process to afford the cyclic selenosulfide **15**. Final liberation of the acid moiety is achieved by HCl-mediated hydrolysis to provide Fmoc-SetCys-OH **6** with an overall yield of 34% (**Figure 1**).

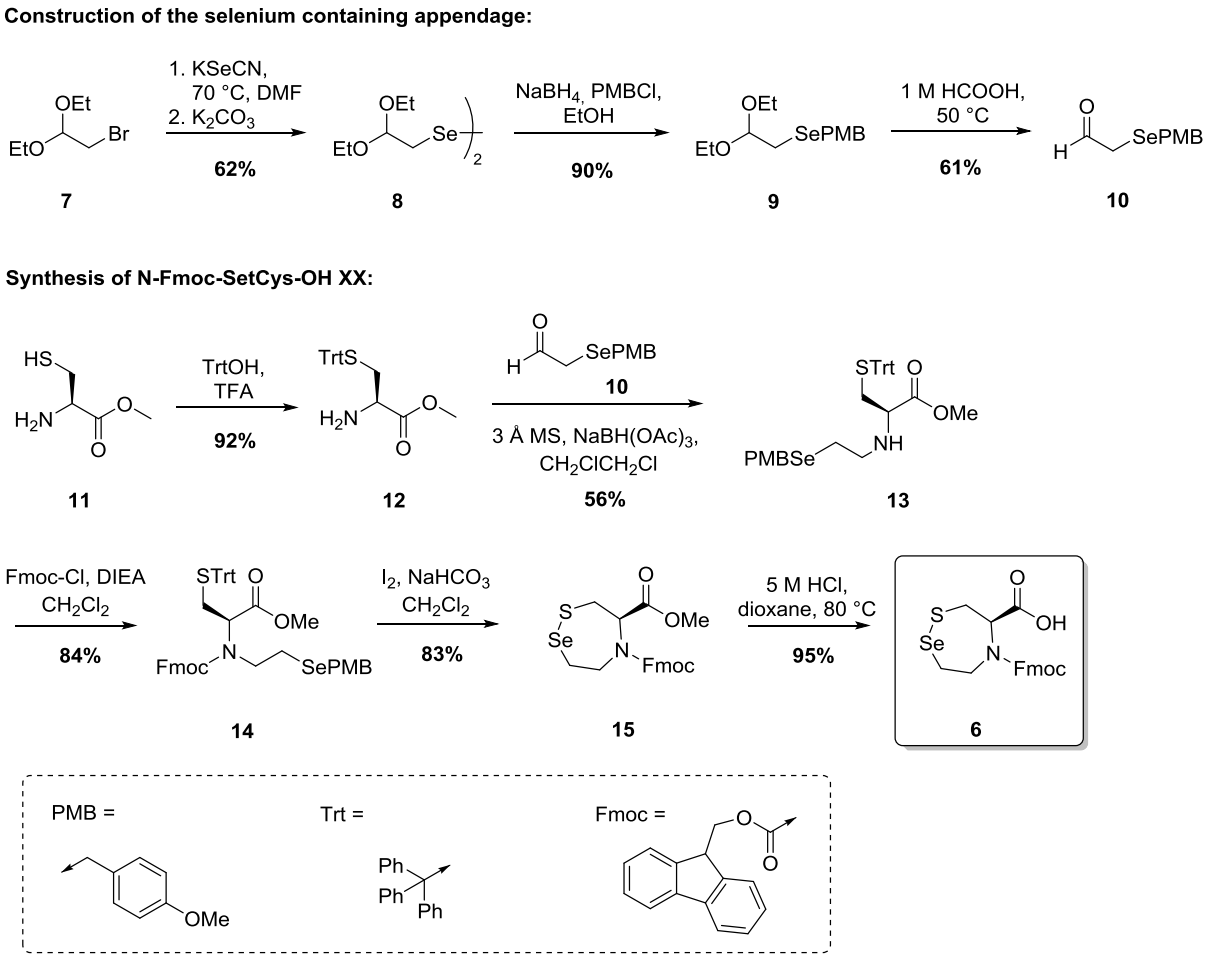
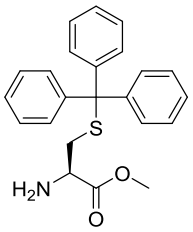


Figure 1. *Synthesis of the Fmoc-protected SetCys amino acid* ***6.***

Synthesis of H-Cys(Trt)-OMe 12



12

Chemical Formula: $C_{23}H_{23}NO_2S$
Molecular Weight: 377,5020

To a solution of cysteine methyl ester hydrochloride 11 (856 mg, 4.99 mmol) in TFA (5 mL) was added triphenylmethanol (1.43 g, 5.5 mmol) and the mixture was stirred at RT for 5 h. After evaporation of the TFA, the residue was dissolved in MeOH (27 mL) and the mixture was stirred at RT until the yellow color disappeared. The methanol was then evaporated under reduced pressure and the residue was suspended in 0.25 M K_2CO_3 (30 mL)**.

Workup. The obtained aqueous layer was transfered in a separatory funnel and extracted with Et_2O (2 × 20 mL) and the combined organic layers were dried over $MgSO_4$ and filtered out. After evaporation of the solvent, purification of the crude by column chromatography (DCM/MeOH 98:2) provided the S-trityl protected amino acid 12 (1.73 g) as a viscous oil with 92% yield.

*The addition of K_2CO_3 solution or concentrated HCl to the mixture can induce CO_2 release and bubbling.

*The pH of the aqueous solution of K_2CO_3 must be at least pH 9 after crude dissolution to best achieve free amine extraction by the organic solvent. If the pH is too low, add more 0.25 M K_2CO_3 aqueous solution.

NMR data are in agreement with the literature (**Figure 2, Figure 3**).[2]

^{1}H NMR (300 MHz, $CDCl_3$) δ 7.41-7.44 (m, 6H), 7.18-7.31 (m, 9H), 3.65 (s, 3H), 3.20 (dd, *J* = 4.8 and 7.8 Hz, 1H), 2.59 (dd, *J* = 4.8 and 12.4 Hz, 1H), 2.46 (d, *J* = 7.8 and 12.4 Hz, 1H) ppm. **^{13}C NMR** (75 MHz, $CDCl_3$) δ 174.3 (C), 144.6 (3 × CH), 129.7 (6 × CH), 128.1 (6 × CH), 126.9 (3 × CH), 66.9 (C), 53.9 (CH), 52.3 (CH_3), 37.2 (CH_2) ppm.

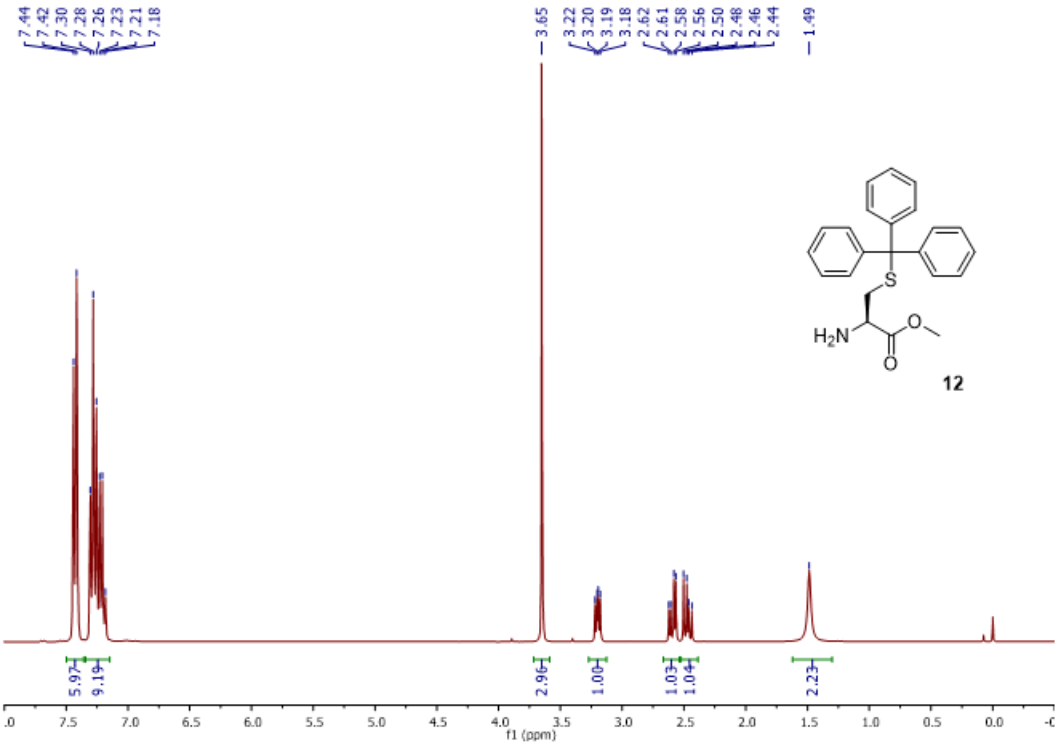


Figure 2. *^{1}H NMR (300 MHz) spectrum of H-Cys(Trt)-OMe* ***12*** *($CDCl_3$, 293 K).*

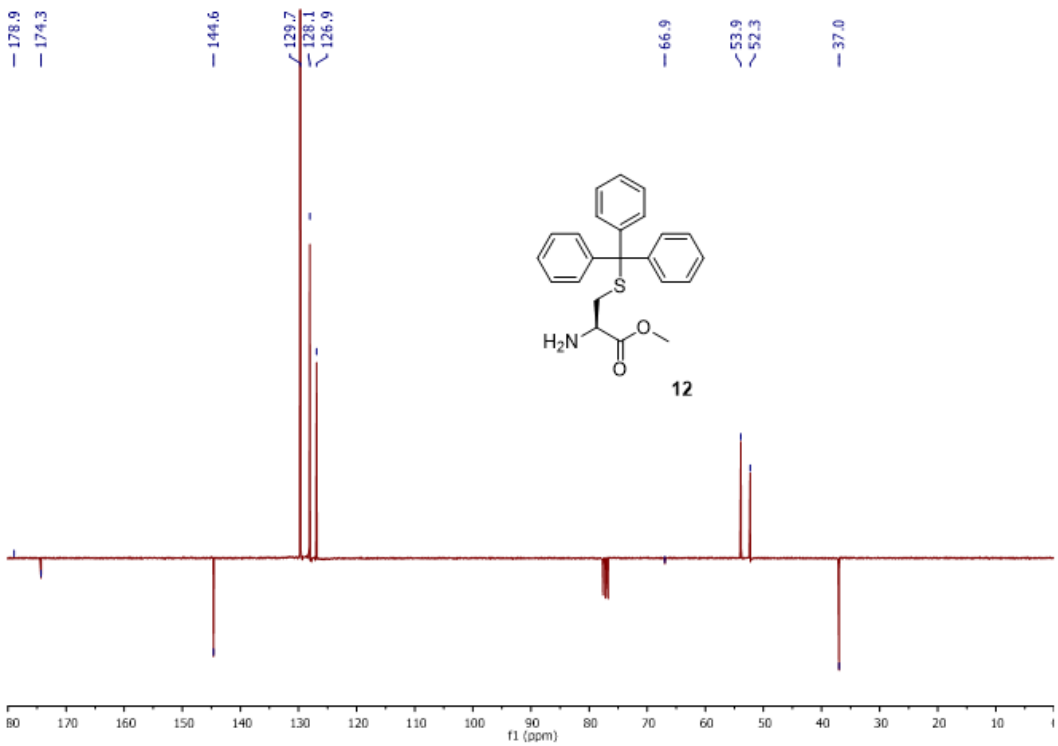


Figure 3. *^{13}C JMOD NMR (75 MHz) spectrum of H-Cys(Trt)-OMe* ***12*** *($CDCl^3$, 293 K).*

Synthesis of *bis*(2,2-diethoxyethyl)diselenide 8

The following procedure adapted from Krief *et al*[2] was used to prepare diselenide **8**.

In a two-neck flask equipped with a septum and a reflux condenser, potassium selenocyanate (576 mg, 4.00 mmol) was dissolved in DMF (10 mL) under argon atmosphere. Then the solution was heated at 75 °C (oil bath). The resulting solution turned to a clear yellowish color. A solution of 2-bromo-1,1-diethoxyethane **7** (4.00 mmol, 601 µL) in DMF (4 mL) was introduced in the flask dropwise. The reaction mixture was stirred at 75 °C for 3 h. The solution became brownish. A solution of K_2CO_3 (552 mg, 4.00 mmol) in deionized water (1.6 mL) was added dropwise. The reaction mixture has been stirring for 3 h at 75 °C.

Workup. The oil bath was removed, and 20 mL of deionized water was added at RT. The mixture was transfered to a separatory funnel. The aqueous layer was extracted with Et_2O (2 × 20 mL)*, the combined organic layers were washed with water (2 × 20 mL) and dried over anhydrous $MgSO_4$. The solvent was filtered and evaporated to dryness under reduced pressure in a rotatory evaporator. The crude product was purified by column chromatography on silica gel (hexane/EtOAc 95:5) to obtain diselenide **8** as a yellow oil (485 mg, 62%).

*After the extraction with Et_2O, the aqueous layer has to be neutralized with bleach since the reaction with KSeCN leads to the formation of one equivalent of cyanide anion.

Check the identity and purity of the synthesized compound by NMR. **^{1}H NMR** (300 MHz, $CDCl_3$, 298 K) δ 4.70 (t, *J* = 5.7 Hz, 2H), 3.63–3.73 (m, 4H), 3.52–3.61 (m 4H), 3.22 (d, *J* = 5.7 Hz, 4H), 1.22 (t, *J* = 7.1 Hz, 12H).

Synthesis of (2,2-diethoxyethyl)(4-methoxybenzyl)selenide 9

In round-bottom flask maintained under an argon atmosphere, diselenide **8** (900 mg, 2.29 mmol) was dissolved in absolute ethanol (14 mL). The flask was placed in an ice bath, then $NaBH_4$ * (175 mg, 4.62 mmol) was added portionwise under a stream of argon gas. The ice bath was removed and the mixture was allowed to stir at RT for 30 min. A white precipitate initially formed and redissolved resulting in a homogeneous solution. Then the flask was placed again in an ice bath, and 4-methoxybenzyl chloride (561 µL, 4.14 mmol) was added dropwise. We observed the formation of white precipitate. The ice bath was removed again and the reaction mixture was allowed to stir for 3 h at RT.

Workup. The crude suspension was filtered through a celite pad in order to remove the white precipitate. The clear yellowish filtrate was concentrated under reduced pressure in a rotatory evaporator and 30 mL of water was added. The aqueous mixture was transfered to a separatory funnel and extracted it with Et_2O (2 × 25 mL). The combined organic layers were dried over $MgSO_4$, the solvent filtered and evaporated under reduced pressure in a rotatory

evaporator. The crude product was purified by silica gel column chromatography (hexanes/EtOAc 95:5) to obtain selenide **9** as a yellow oil (1.19 mg, 90%).

*The addition of $NaBH_4$ leads to active liberation of molecular hydrogen. The excess of gas should be evacuated through an outlet.

Check the identity and purity of the synthesized compound by ^{1}H NMR. **^{1}H NMR** (300 MHz, $CDCl_3$, 298 K) δ 7.23 (d, *J* = 8.6 Hz, 2H), 6.82 (d, *J* = 8.6 Hz, 2H), 4.59 (t, *J* = 5.6 Hz, 1H), 3.82 (s, 2H), 3.79 (s, 3H), 3.59–3.70 (m, 2H), 3.46–3.56 (m, 2H), 2.62 (d, J = 5.6 Hz, 2H), 1.22 (t, *J* = 7.1 Hz, 6H).

Synthesis of 2-((4-methoxybenzyl)selenyl)acetaldehyde 10

Aldehyde **10** was prepared by a procedure similar to that reported by Abbas *et al.*[3]

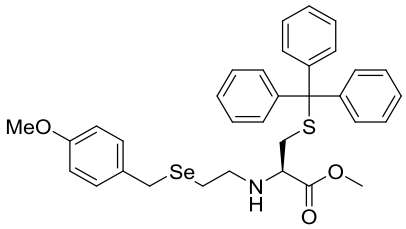
In a round-bottom flask equipped with a reflux condenser, acetal **9** (1.78 g, 5.61 mmol) was suspended in 1 M aqueous formic acid solution (24 mL). The flask was placed in an oil bath and heat the mixture overnight at 50 °C (oil bath temperature).

Workup. The oil bath was removed and water (55 mL) was added at RT. The aqueous mixture was transfered to a separatory funnel and extracted with Et_2O (2 × 55 mL). The combined organic layers were washed with water (40 mL) and the organic solution was dried over $MgSO_4$. The solvent was filtered and evaporated under reduced pressure in a rotatory evaporator. Aldehyde **10** (838 mg, 61%) is directly used in the next step without further purifications*.

*If aldehyde **10** cannot be used for the next step on the same day, we advise to store it at -20 °C under an argon atmosphere overnight.

Check the identity and purity of the synthesized compound by ^{1}H NMR (300 MHz, $CDCl_3$, 298 K) δ 9.36 (t, *J* = 4.2 Hz, 1H), 7.22 (d, *J* = 8.7 Hz, 2H), 6.84 (d, *J* = 8.7 Hz, 2H), 3.79 (s, 3H), 3.64 (s, 2H), 3.12 (d, *J* = 4.1 Hz, 2H).

Synthesis of *N*-[PMBSe-$(CH_2)_2$]-Cys(Trt)-OMe 13



13

Chemical Formula: $C_{33}H_{35}NO_3SSe$
Molecular Weight: 604,6670

After evaporation of the solvent under reduced pressure, aldehyde **10** was dissolved in anhydrous 1,2-dichloroethane (50 mL) under argon. Activated powdered 3 Å molecular sieves* (4 g), sodium triacetoxyborohydride (1.02 g, 4.81 mmol) and a solution of H-Cys(Trt)-OMe **12** (1.30 g, 3.44 mmol) in DCE (35 mL) were successively added to the solution of aldehyde.

Workup. After 18 h of stirring at RT, the reaction mixture was filtered on a Büchner funnel and the solid was washed with additional portions of DCM. The filtrate was evaporated under reduced pressure. Then 0.5 M K_2CO_3 (50 mL) was added to the residue and the aqueous layer was extracted with DCM (3 × 50 mL). The combined organic layers were dried over $MgSO_4$, filtered, and the solvent was evaporated under reduced pressure. Purification of the crude mixture by column chromatography (cyclohexane/ethyl acetate 80:20) afforded the *N*-alkylated amino acid **13** as a clear yellow oil (1.17 g, 56%).

*The powdered 3 Å molecular sieve was activated under high vacuum (< 1 mbar), by heating the flask with a heatgun for ~15 min. The flask containing the molecular sieve was filled with argon during cooling to RT.

^{1}H NMR (300 MHz, $CDCl_3$) δ 7.41 (d, *J* = 7.4 Hz, 6H), 7.19-7.32 (m, 9H), 7.17 (d, *J* = 8.6 Hz, 2H), 6.79 (d, *J* = 8.6 Hz, 2H), 3.77 (s, 3H), 3.71 (s, 2H), 3.66 (s, 3H), 2.99 (t, *J* = 6.5 Hz, 1H), 2.37-2.74 (m, 6H), 1.44-1.91 (m, 1H) ppm (*Figure 4*). **^{13}C NMR** (75 MHz, $CDCl_3$) δ 173.5 (C), 158.4 (C), 144.6 (3 × C), 131.1 (C), 129.9 (2 × CH), 129.6 (6 × CH), 127.9 (6 × CH), 126.7 (3 × CH), 113.9 (2 × CH), 66.8 (C), 60.2 (CH), 55.2 (CH_3), 51.9 (CH_3), 47.3 (CH_2), 34.6 (CH_2), 26.3 (CH_2), 23.9 (CH_2) ppm (*Figure 4*, **Figure 5**). **MALDI-TOF** (**Figure 6**).

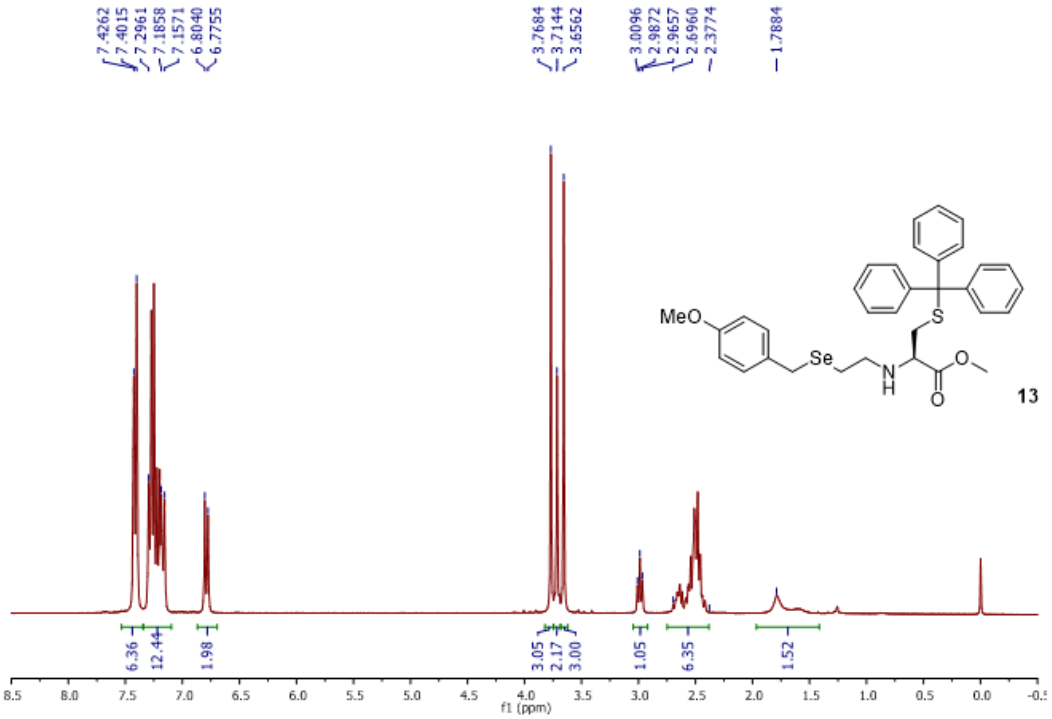


Figure 4. *^{1}H NMR (300 MHz) spectrum of compound **13** ($CDCl_3$, 293 K).*

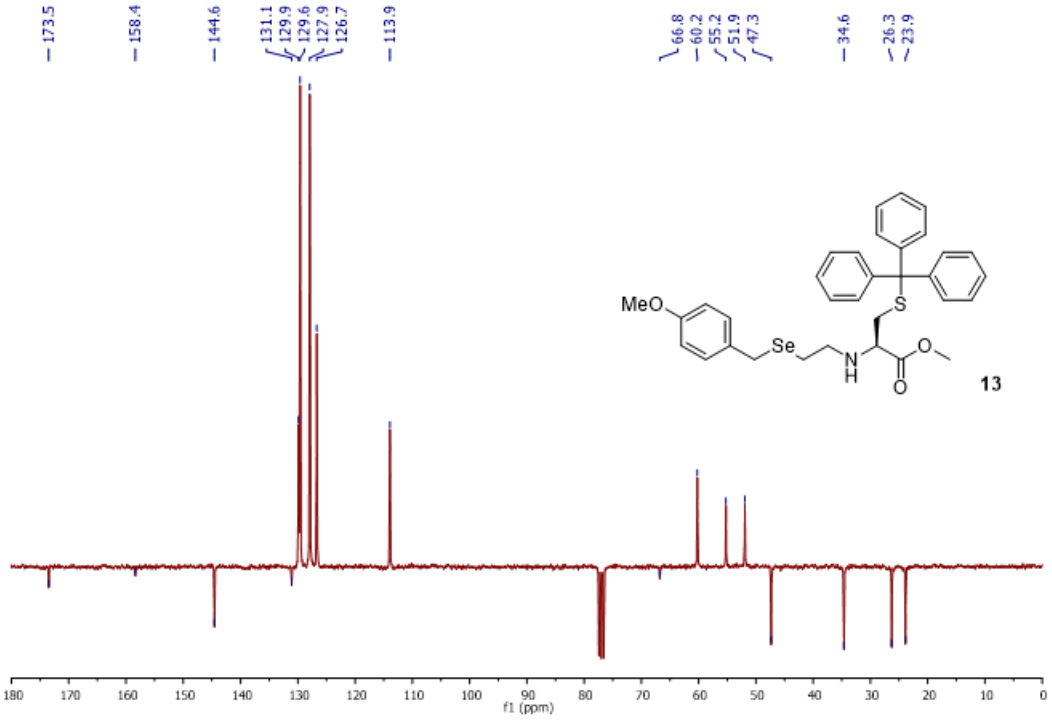


Figure 5. *^{13}C JMOD NMR (75 MHz) spectrum of compound **13** ($CDCl_3$, 293 K).*

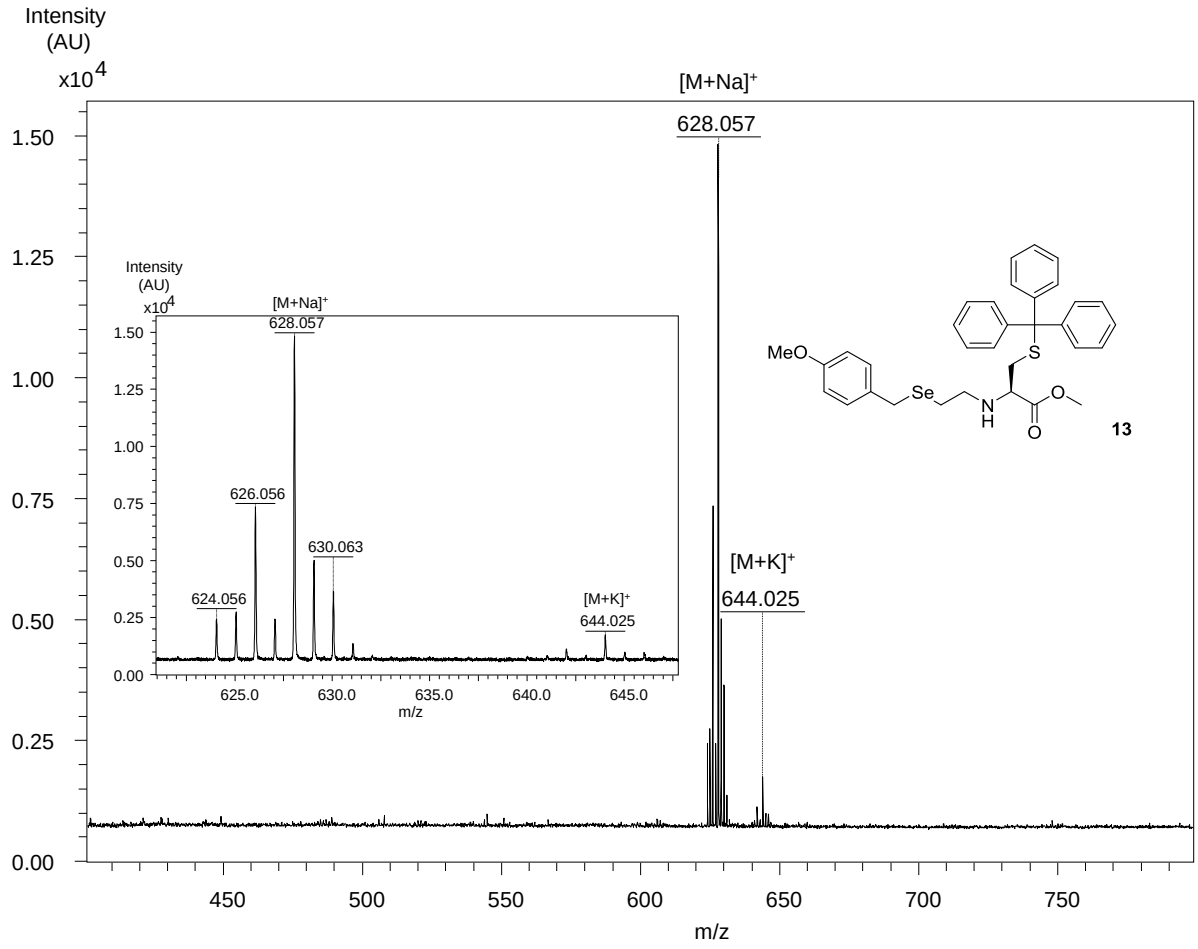
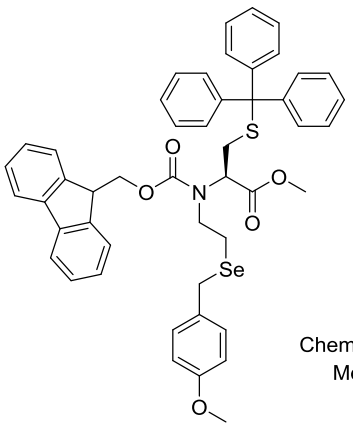


Figure 6. *MALDI-TOF analysis of N-[PMBSe-(CH₂)₂]-Cys(Trt)-OMe* ***13****. Matrix 2,5-dihydroxybenzoic acid (DHB), positive detection mode, [M+Na]⁺ (monoisotopic) 628.14, found 628.06.*

Synthesis of *N*-Fmoc-*N*-[PMBSe-(CH₂)₂]-Cys(Trt)-OMe 14



14

Chemical Formula: $C_{48}H_{45}NO_5SSe$
Molecular Weight: 826,9100

To a solution of *N*-[PMBSe(CH_2)$_2$]-Cys(Trt)-OMe **13** (455 mg, 0.75 mmol) in DCM (3.75 mL) were successively added DIEA (0.137 mL, 0.79 mmol) and FmocCl (204 mg, 0.79 mmol). The reaction mixture was stirred at RT for 24 h and was then diluted with DCM (20 mM).

Workup. The obtained organic layer was washed with water (2 × 20 mL) and was dried over $MgSO_4$. After evaporation of the solvent under reduced pressure, purification of the crude by column chromatography (cyclohexane/EtOAc 7:3) provided the expected compound **14** (522 g) as a white solid with 84% yield.

The titled compound was characterized by ^{1}H and ^{13}C NMR as a mixture of two conformers*. A/B = 37:63. The ratio A/B was calculated from the CH signals of the Fmoc protecting group. **

* There are strong conformational restrictions in compounds **14**, **15** and **6**. As a consequence, NMR spectra of these compounds at 298 K are complex due to the overlapping of the signals given by all the conformers present in solution. Because no clear description of the NMR spectra can be provided, FID files are given to facilitate the identification of the synthesized compounds by direct comparison.

The presence of the Fmoc group in compound **14 induces conformational restrictions in the molecule. However, variable temperature NMR experiments performed in deutered DMSO show that such conformational barriers can be overcome by heating the sample, leading to the coalescence of NMR signals above 70 °C. Once the selenosulfide bond is formed by iodine oxidation, the conformational restrictions are so strong in the resulting cyclic species (compounds **15** and **6**) that several conformers are detected by NMR in solution, whatever the temperature.

^{1}H NMR (300 MHz, $CDCl_3$) δ 7.80 (d, *J* = 9.0 Hz, 0.64H, 2H_A), 7.74 (d, *J* = 7.6 Hz, 2H_B), 7.49-7.59 (m, 2H_A + 2H_B), 7.12-7.49 (m, 21H_A + 19H_B + residual $CDCl_3$), 7.08 (d, *J* = 8.3 Hz, 2H_B), 6.89 (d, *J* = 8.3 Hz, 2H_A), 6.71 (d, *J* = 8.4 Hz, 2H_B), 4.34-4.60 (m, 2H_A + 2H_B), 4.20 (t, *J* = 5.9 Hz, H_B), 4.13 (t, *J* = 5.2 Hz, H_A), 3.75 (s, 3H_A), 3.73 (s, 2H_A), 3.68 (s, 3H_B) 3.40-3.63 (m, H_A + 5H_B), 3.02-3.40 (m, 4H_A + 2H_B), 2.45-3.02 (m, 3H_A + 3H_B), 2.17-2.45 (m, 2H_A + 2H_B) ppm (*Figure 7*). **^{13}C NMR** (75 MHz, $CDCl_3$) δ 170.0 (C, B), 169.7 (C, A), 158.4 (C, A +B), 155.2 (C, A + B), 144.5 (C, A + B), 143.7 (C, A + B), 141.4 (C, A), 141.3 (C, B), 131.3 (C, A), 130.9 (C, B), 129.8 (CH, A + B), 129.6 (CH, A + B), 128.0 (CH, A + B), 127.7 (CH, A + B), 127.1 (CH, A + B), 126.8 (CH, A + B), 124.3-125.3 (m, CH, A + B), 120.0 (CH, A + B), 113.9 (A + B), 67.2 (CH_2, A), 67.0 (CH_2, B), 60.6 (CH, B), 59.8 (CH, A), 55.1 (CH_3, A+ B), 52.3 (CH_3, B), 52.2 (CH_3, A), 49.6 (CH_2, A), 49.2 (CH_2, B), 47.2 (CH, B), 47.1 (CH, A), 31.5 (CH_2, A), 31.2 (CH_2, B), 26.5 (CH_2, A + B), 21.3 (CH_2, B), 20.8 (CH_2, A) ppm (**Figure 8**). **HSQC** (**Figure 9**), **MALDI-TOF** (**Figure 10**).

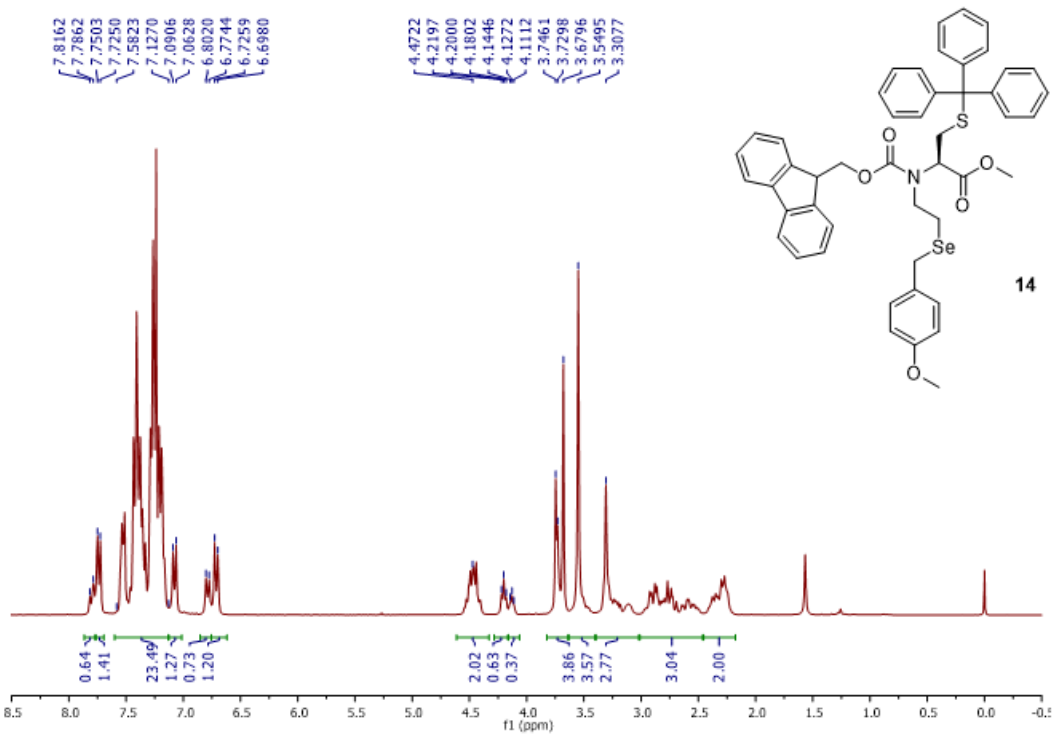


Figure 7. *^{1}H NMR (300 MHz) spectrum of N-Fmoc-N-[PMBSe-$(CH_2)_2$]-Cys(Trt)-OMe **14** ($CDCl_3$, 293 K).*

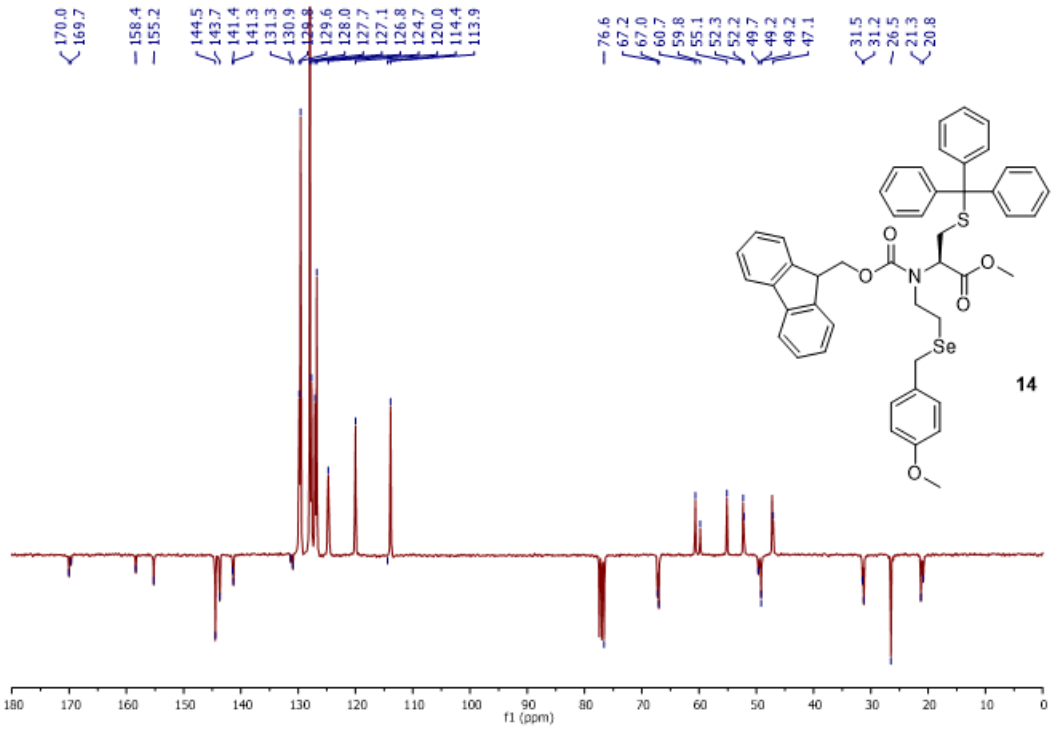


Figure 8. *^{13}C JMOD NMR (75 MHz) spectrum of N-Fmoc-N-[PMBSe-$(CH_2)_2$]-Cys(Trt)-OMe **14** ($CDCl_3$, 293 K).*

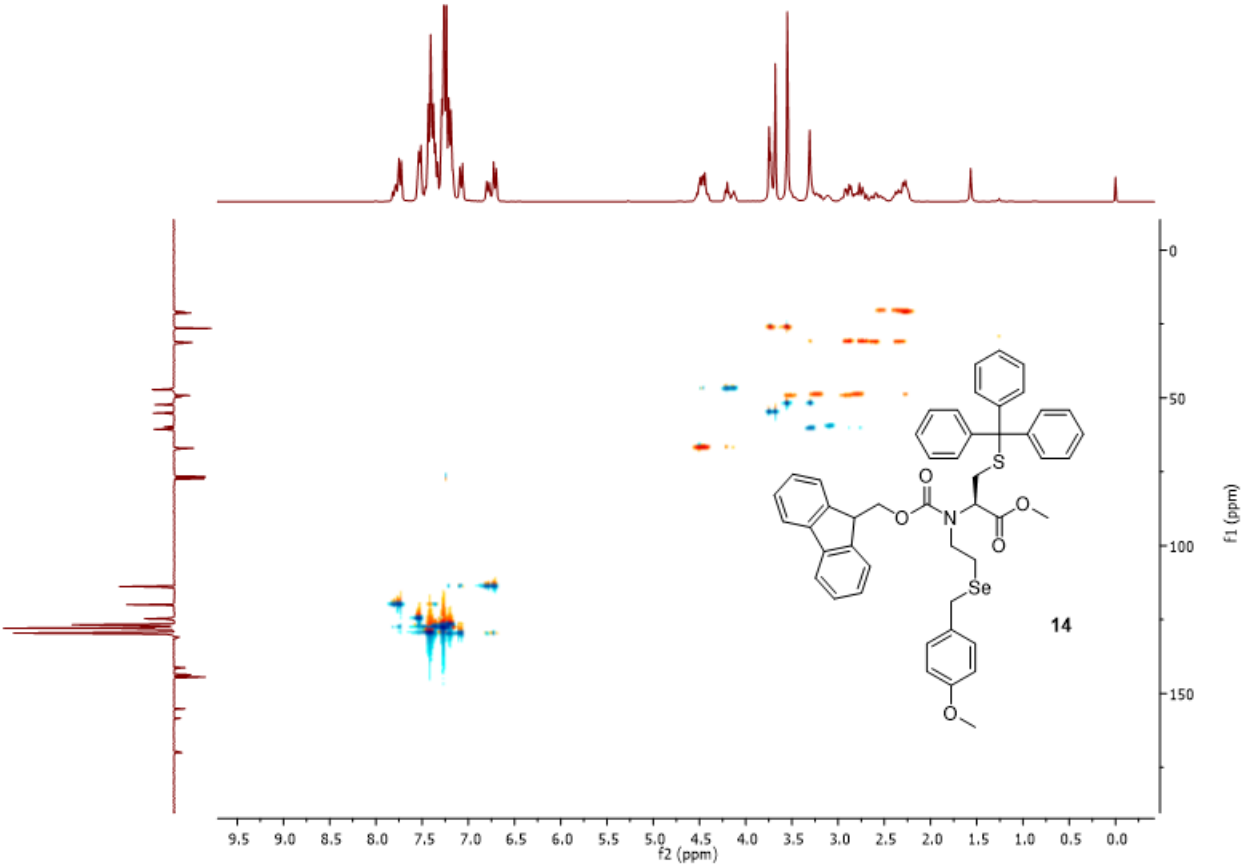


Figure 9. *^{1}H-^{13}C HSQC spectrum of N-Fmoc-N-[PMBSe-(CH2)2]-Cys(Trt)-OMe* ***14*** *($CDCl_3$, 293 K).*

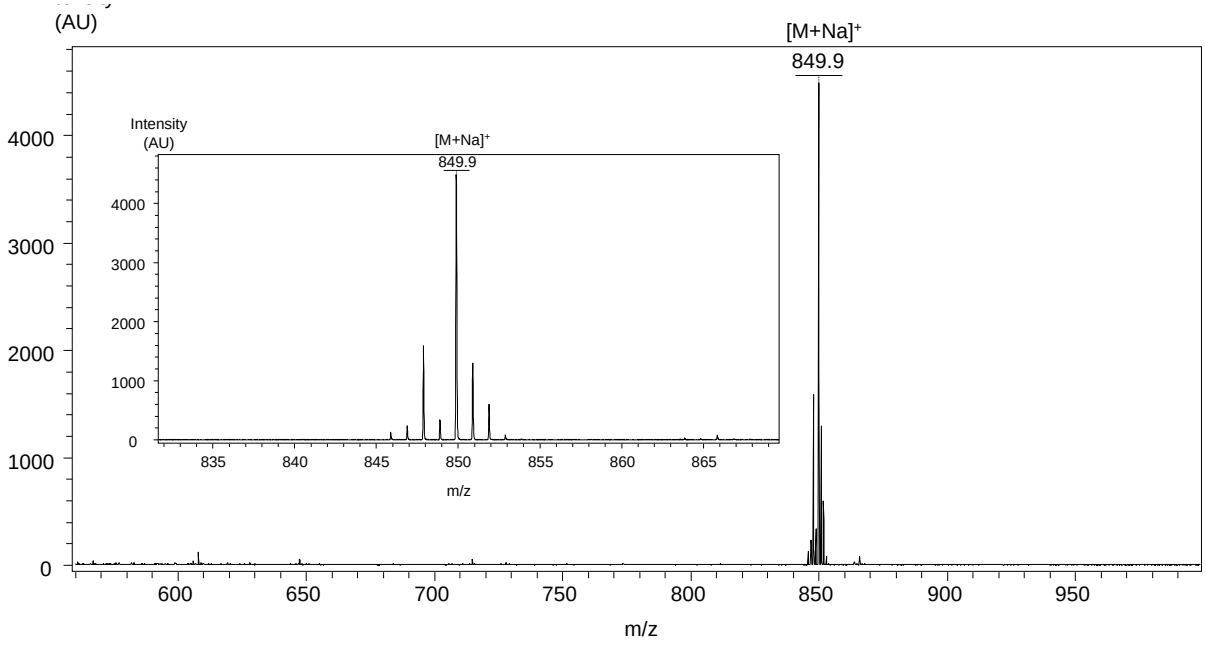
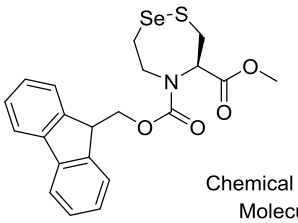


Figure 10*. MALDI-TOF analysis of N-Fmoc-N-[PMBSe-$(CH_2)_2$]-Cys(Trt)-OMe* ***14****. Matrix 2,5-dihydroxybenzoic acid (DHB), positive detection mode, $[M+Na]^+$ (monoisotopic) 850.22, found 849.91.*

Synthesis of Fmoc-SetCys-OMe 15



15

Chemical Formula: $C_{21}H_{21}NO_4SSe$
Molecular Weight: 462,4220

To a solution of *N*-Fmoc-*N*-[PMBSe-$(CH_2)_2$]-Cys(Trt)-OMe 14 (1.67 g, 2.02) in DCM (50 mL) was added $NaHCO_3$ (508 mg, 6.06 mmol). The mixture was cooled to 0 °C and I2 (1.54 g, 6.06 mmol) dissolved in DCM (50 mL) was added dropwise. The reaction mixture was stirred at RT for 30 min*.

Workup. 1 M $Na_2S_2O_3$ (50 mL) was added to remove the excess of I2. The organic layer was separated and the aqueous layer was extracted with DCM (2 × 50 mL). Note that the separation between the aqueous and organic layers was improved by addition of brine (50 mL). All the organic layers were then combined, washed with brine (50 mL) and dried over $MgSO_4$. After evaporation of the solvent*, purification of the crude by column chromatography (cyclohexane/EtOAc 8:2) provided the expected compound 15 (780 mg) as a glass with 83% yield.

* For TLC analysis, one drop of the crude mixture is poured in a biphasic mixture containing ~500 µL of a 1 M $Na_2S_2O_3$ solution and ~200 µL of AcOEt. The tube is vigorously shaken and the upper layer is spotted on the TLC plate against the reference for the starting material.

* We noticed that evaporation of the crude mixture to dryness before purification systematically resulted in lower isolated yields. Therefore, we recommend to concentrate the mixture to a final volume of 5-10 mL. The silica gel chromatography column is loaded with the solution (final volume of 5-10 mL) obtained after concentrating the crude mixture under reduced pressure.

The titled compound was characterized by ^1H and ^{13}C NMR as a mixture of two main conformers. A/B = 45:55. The ratio A/B was calculated from the aromatic CH signals of the Fmoc protecting group. Due to overlapping signals, no clear description of the ^1H NMR spectrum can be provided (**Figure 11**).

^{13}C NMR (75 MHz, $CDCl_3$) δ 170.0 (C), 169.7 (C), 155.6 (C), 155.2 (C), 143.90 (C), 143.88 (C), 143.84 (C), 143.6 (C), 141.6 (C), 141.53 (C), 141.51 (C), 127.91 (CH), 127.88 (CH), 127.86 (CH), 127.7 (CH), 127.34 (CH), 127.30 (CH), 127.25 (CH), 127.1 (CH),

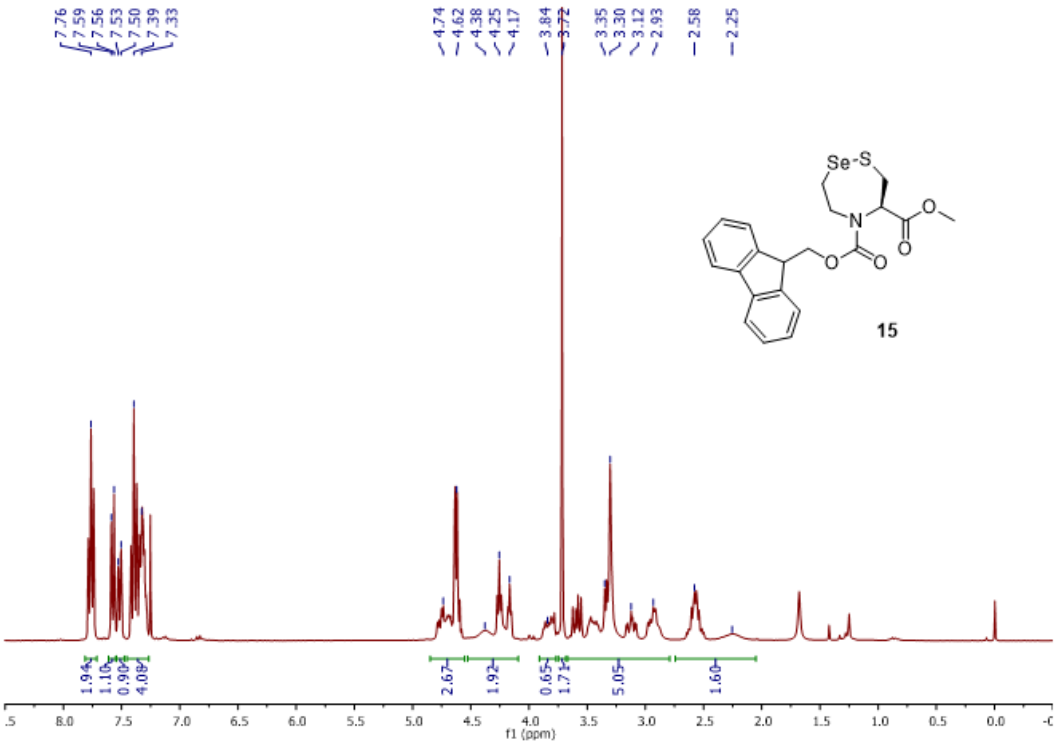


Figure 11. *^{1}H NMR (300 MHz) spectrum of Fmoc-SetCys-OMe **15** ($CDCl_3$, 293 K).*

124.8 (CH), 124.4 (CH), 124.3 (CH), 67.1 (CH_2), 66.8 (CH_2), 62.7 (CH), 52.7 (CH_3), 52.4 (CH_3), 49.8 (CH_2), 47.4 (CH), 47.2 (CH), 38.6 (CH_2), 29.3 (CH_2), 29.1 (CH_2) ppm. **IR** (ATR, cm^{-1}) 2948, 1740, 1697, 1474, 1449, 1417, 1278, 738 (**Figure 12**). **HRMS** (ES^+) Calcd. for $C_{21}H_{21}NO_4NaSSe$: 486.0254, found: 486.0246. **$[\alpha]_D^{20}$** (c 1.0, $CHCl_3$): -29 °.

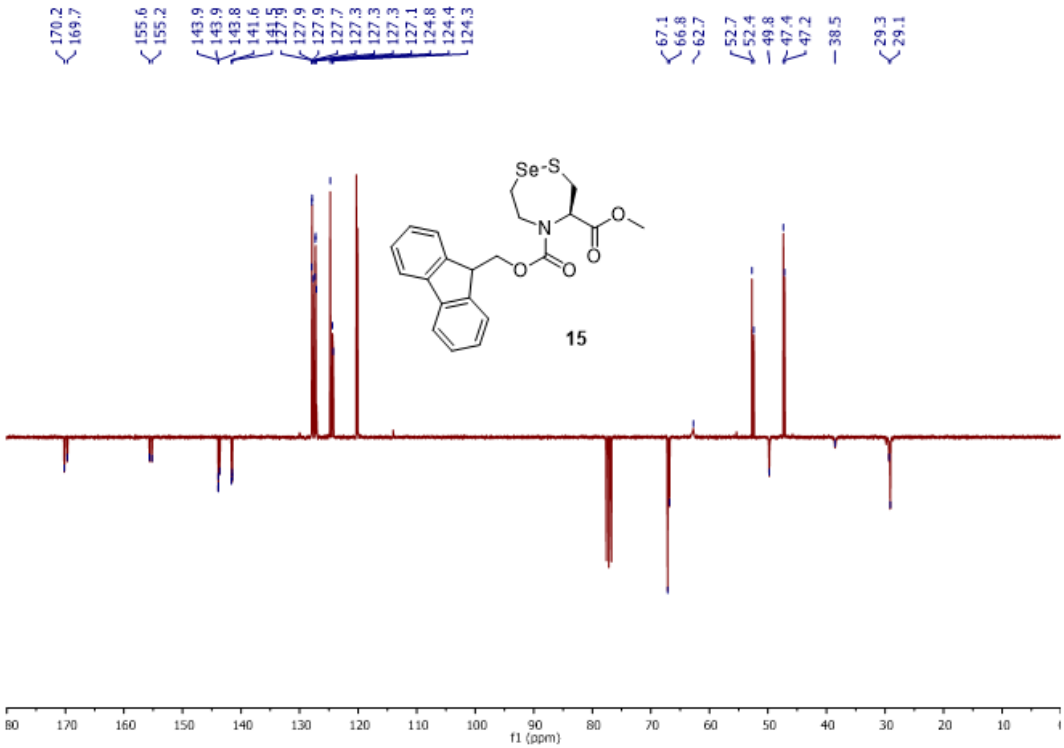


Figure 12. *^{13}C JMOD NMR (75 MHz) spectrum of Fmoc-SetCys-OMe **15** ($CDCl_3$, 293 K).*

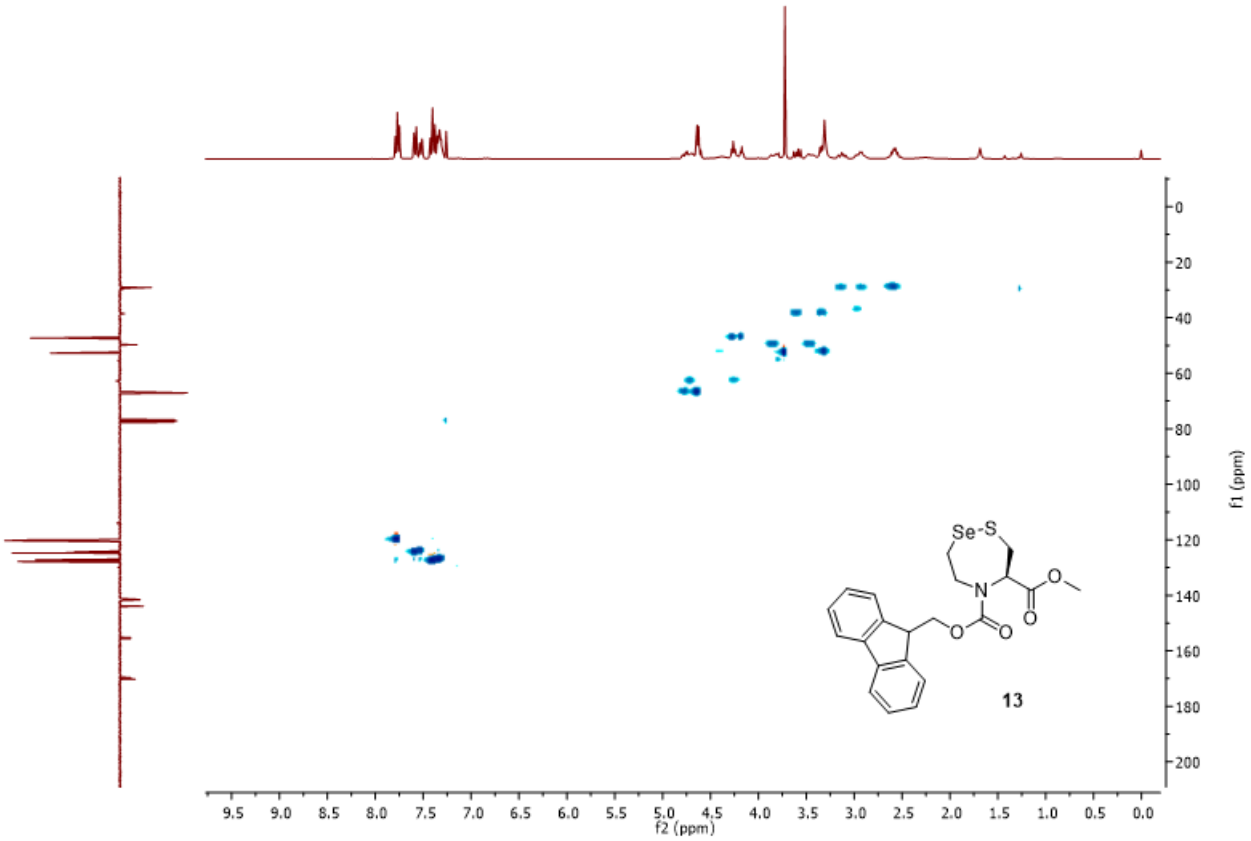


Figure 13. *^{1}H-^{13}C HSQC spectrum of Fmoc-SetCys-OMe* ***15*** *($CDCl_3$, 293 K).*

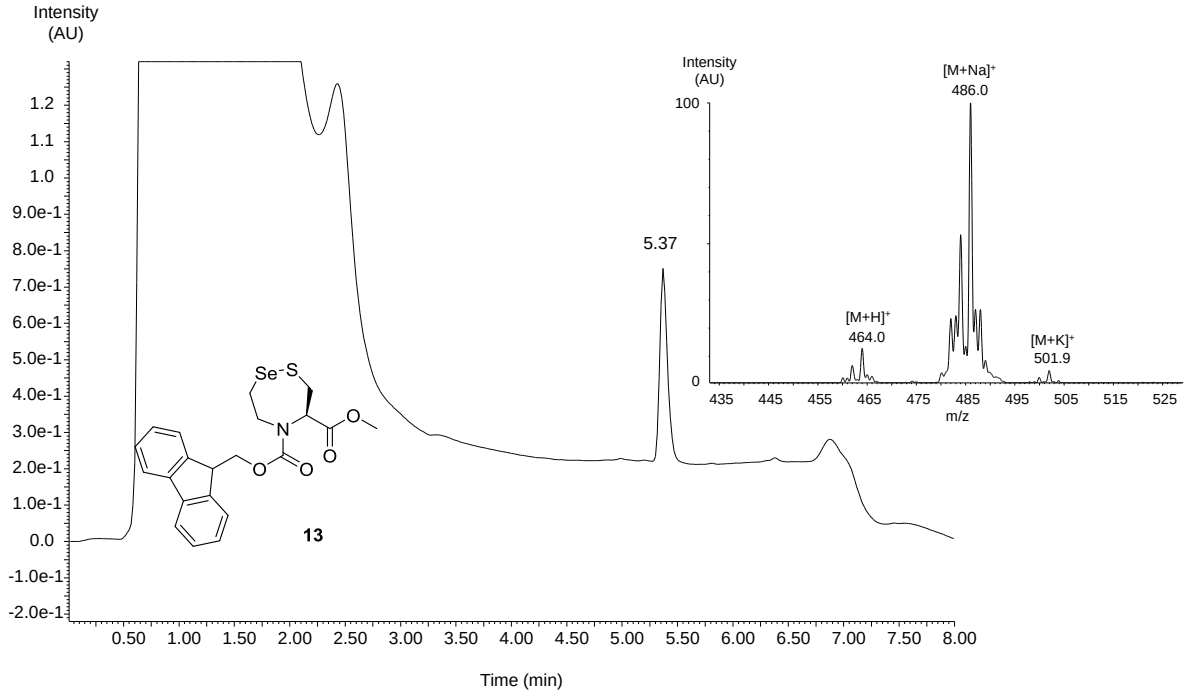


Figure 14*. LC-MS analysis of Fmoc-SetCys-OMe* ***15****. LC trace: eluent C 0.10% formic acid in water, eluent D 0.10% formic acid in CH_3CN/water: 4/1 by vol. C18 column, gradient 0-100% D in 4 min, 50 °C, 0.5 mL/min, UV detection. MS trace: $[M+H]^+$ m/z (monoisotopic mass) 464.04, found 464.00; $[M+Na]^+$ m/z calcd. (monoisotopic mass) 486.04, found 486.00.*

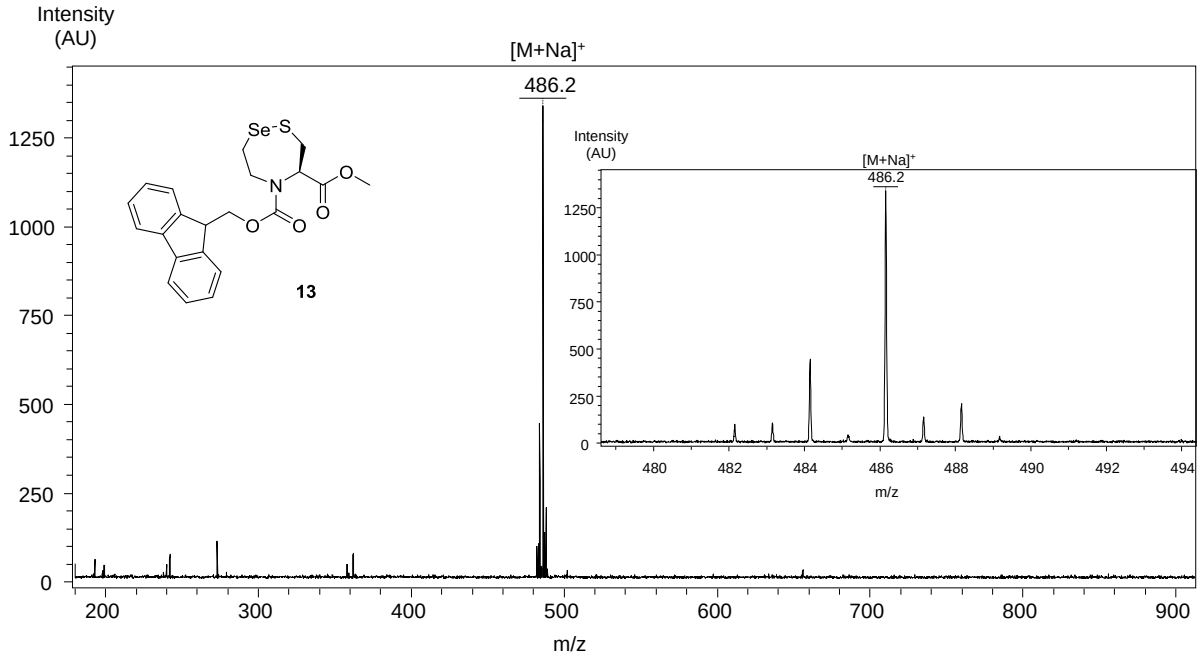
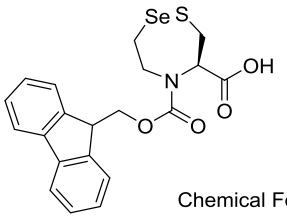


Figure 15. *MALDI-TOF analysis of Fmoc-SetCys-OMe* ***15****. Matrix 2,5-dihydroxybenzoic acid (DHB), positive detection mode,* $[M+Na]^+$ *(monoisotopic) 486.04, found 486.2.*

Synthesis of Fmoc-SetCys-OH 6



6

Chemical Formula: $C_{20}H_{19}NO_4SSe$
Molecular Weight: 448,3950

To a solution of **Fmoc-SetCys-OMe 15** (532 mg) in dioxane (22 mL) was added 5 M HCl (11 mL) and the mixture was refluxed for 12 h at 80 °C (oil bath temperature). The reaction mixture was cooled to RT and 5% K_2CO_3 * (200 mL) was added to raise the pH above 9.

Workup. The aqueous layer was then washed with E_2O (2 × 100 mL), acidified with concentrated HCl* until pH ≈ 2-3 and extracted with DCM (3 × 100 mL). The organic layers (DCM) were combined and dried over $MgSO_4$. Evaporation of the solvent under reduced provided the expected amino acid **6** (488 mg, 95%) as a white solid which was used in the next step without further purification.

* The addition of K_2CO_3 solution or concentrated HCl to the mixture can induce CO_2 release and bubbling.

The titled compound was characterized by ^{1}H and ^{13}C NMR as a mixture of two main conformers (**Figure 16, Figure 17**). Due to overlapping signals, no clear description of the ^{1}H NMR spectrum can be provided. **IR** (ATR, cm^{-1}) 3000 (broad signal), 1699 (broad), 1476, 1450, 1419, 1286, 1190, 740. **HRMS** (ES^{+}): Calcd. for $C_{20}H_{19}NO_4NaSSe$: 472.0098, found: 472.0116. **$[\alpha]_D^{20}$** (c 1.0, $CHCl_3$): -32 ° (**Figure 18, Figure 19**).

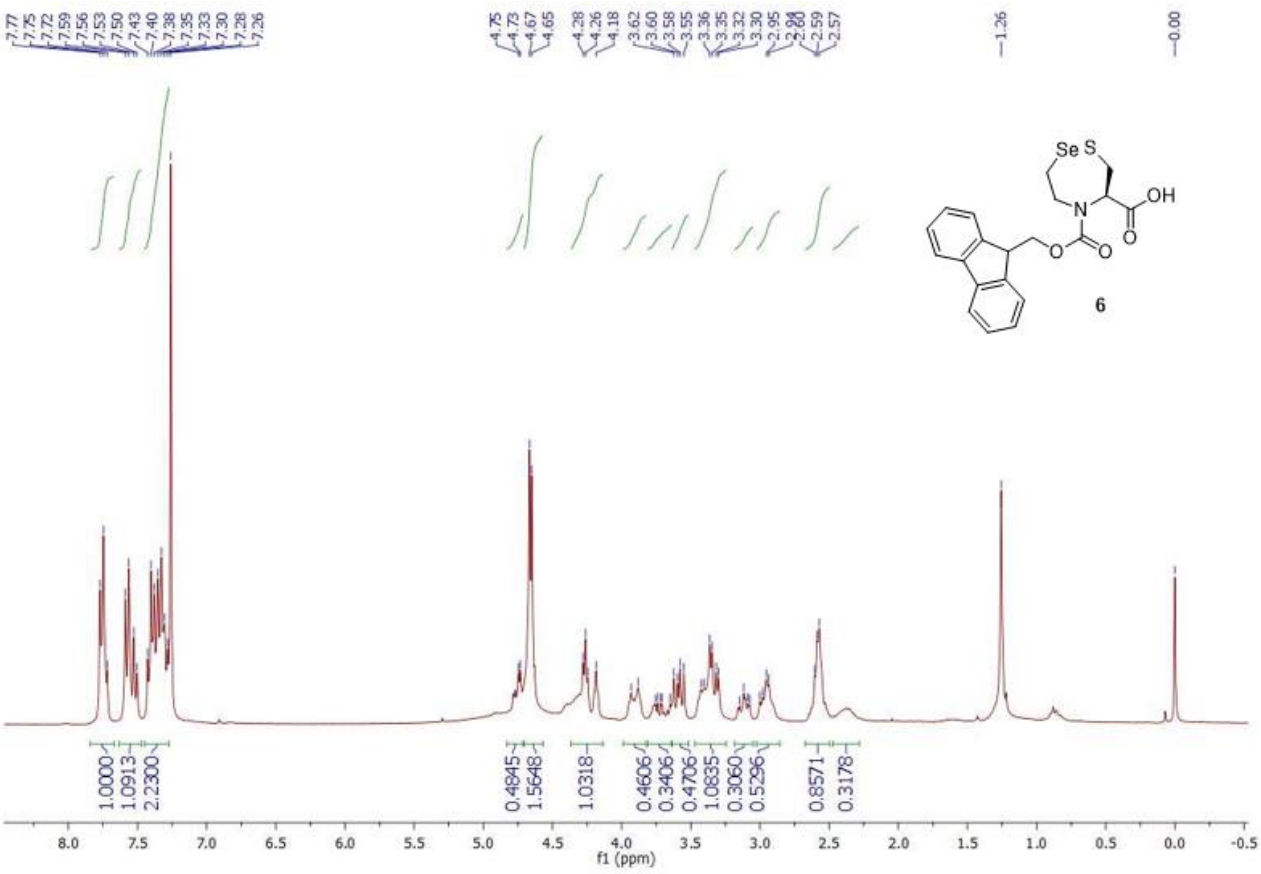


Figure 16. *^{1}H NMR (300 MHz) spectrum of Fmoc-SetCys-OH **6** ($CDCl_3$, 293 K).*

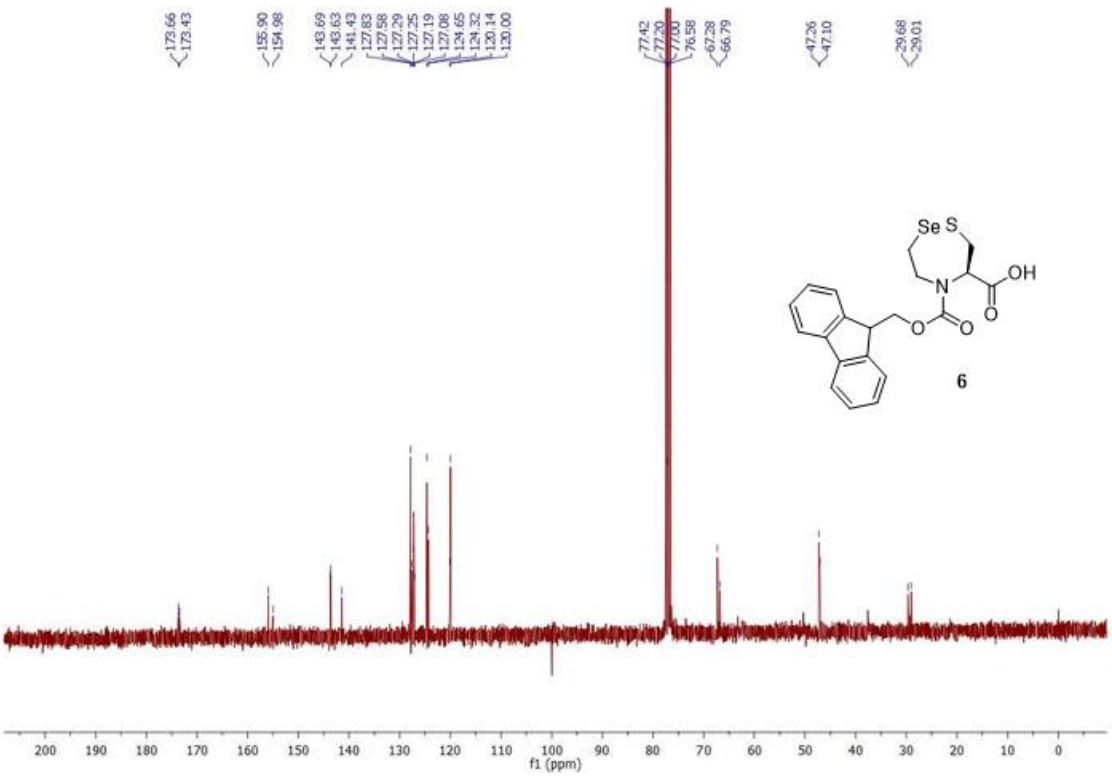


Figure 17. *^{13}C NMR (75 MHz) spectrum of Fmoc-SetCys-OH* ***6*** *($CDCl_3$, 293 K).*

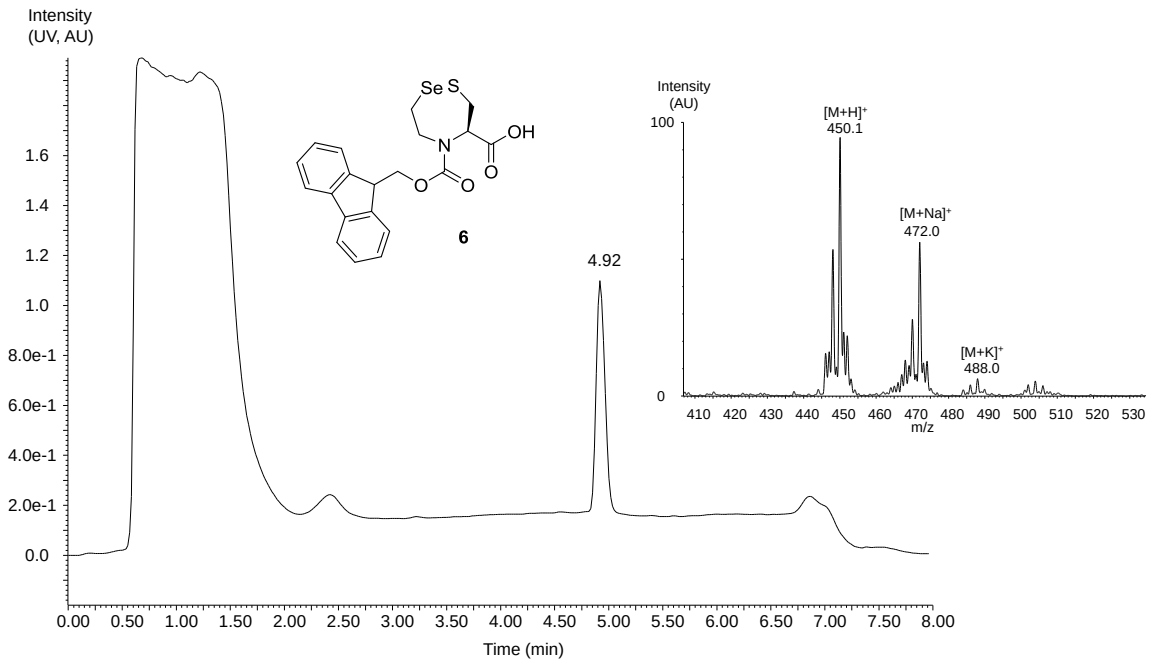


Figure 18*. LC-MS analysis of Fmoc-SetCys-OH* ***6****. LC trace: eluent C 0.10% formic acid in water, eluent D 0.10% formic acid in CH_3CN/water: 4/1 by vol. C18 column, gradient 0-100% D in 4 min, 50 °C, 0.5 mL/min, UV detection. MS trace: $[M+H]^+$ m/z (monoisotopic mass) 450.02, found 450.13; $[M+Na]^+$ m/z (monoisotopic mass) 472.02, found 472.00.*

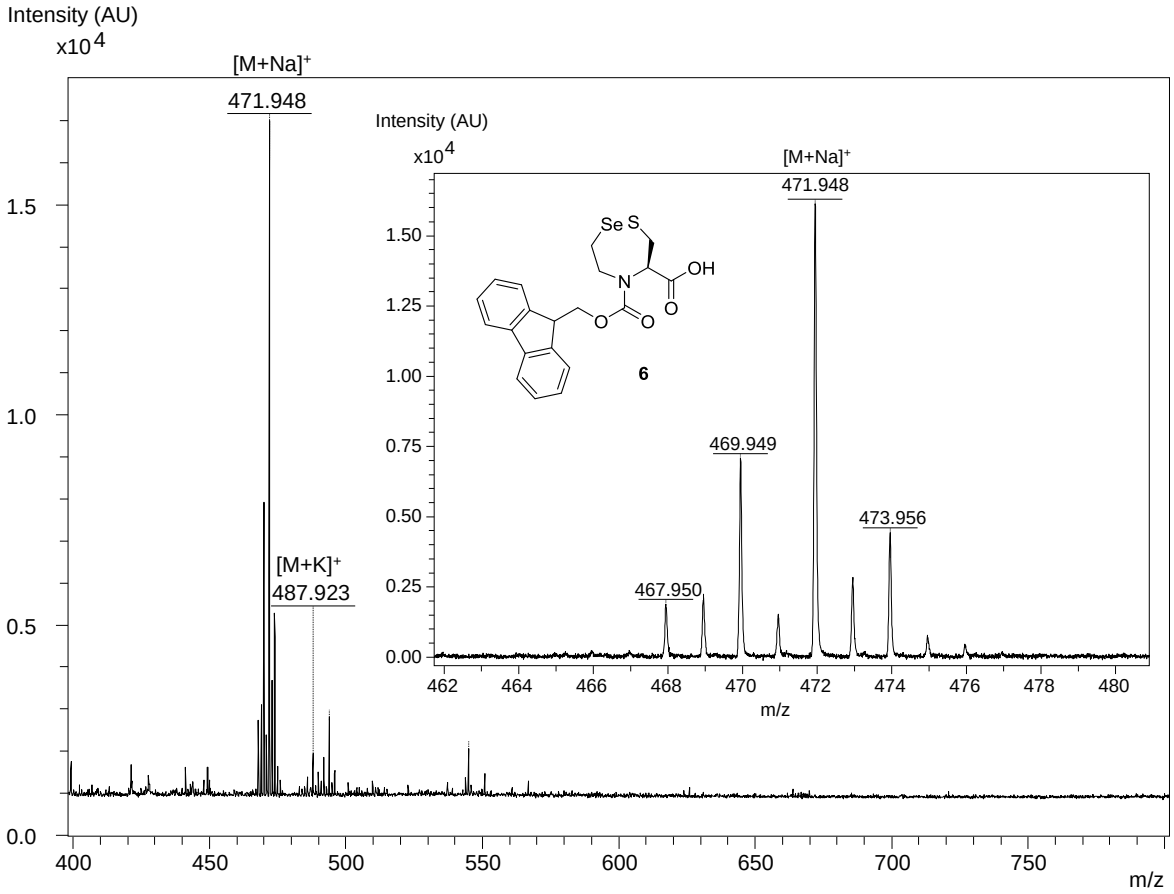


***Figure 19**. MALDI-TOF analysis of Fmoc-SetCys-OH **6**. Matrix 2,5-dihydroxybenzoic acid (DHB), positive detection mode, $[M+Na]^+$ (monoisotopic) 472.01, found 471.95.*

6 Synthesis of peptide segments

6.1 General procedure

Sequences of the target peptide segments were assembled using standard Fmoc chemistry on an automated peptide synthesizer. SEA PS or Rink amide resins were used as solid support (**Figure 20**). Crude amide peptides synthesized from Rink amid resin were directly purified by preparive HPLC. Note that the first amino acid coupled to the SEA PS resin (**Figure 20, B**), Fmoc-SetCys-OH and Fmoc-(ΨPro)-OH were manually introduced in the peptide sequence. At the end of the elongation process, the peptides were cleaved in TFA using a cocktail of scavengers. The peptides were then precipitated by dilution of the TFA solution in a 1:1 (v/v) mixture of Et_2O and heptane (300 mL for 15 mL of solution) and isolated by centrifugation (400 rpm for 10 min). Unless otherwise stated, the crude peptide was dissolved in eluent A and lyophilized.

SEA^{on} peptide obtained after cleavage from SEA PS resin were converted into SEA^{off} peptide by oxidation or into peptidyl thioester by transthioesterification. Detailed protocols are provided for each compounds.

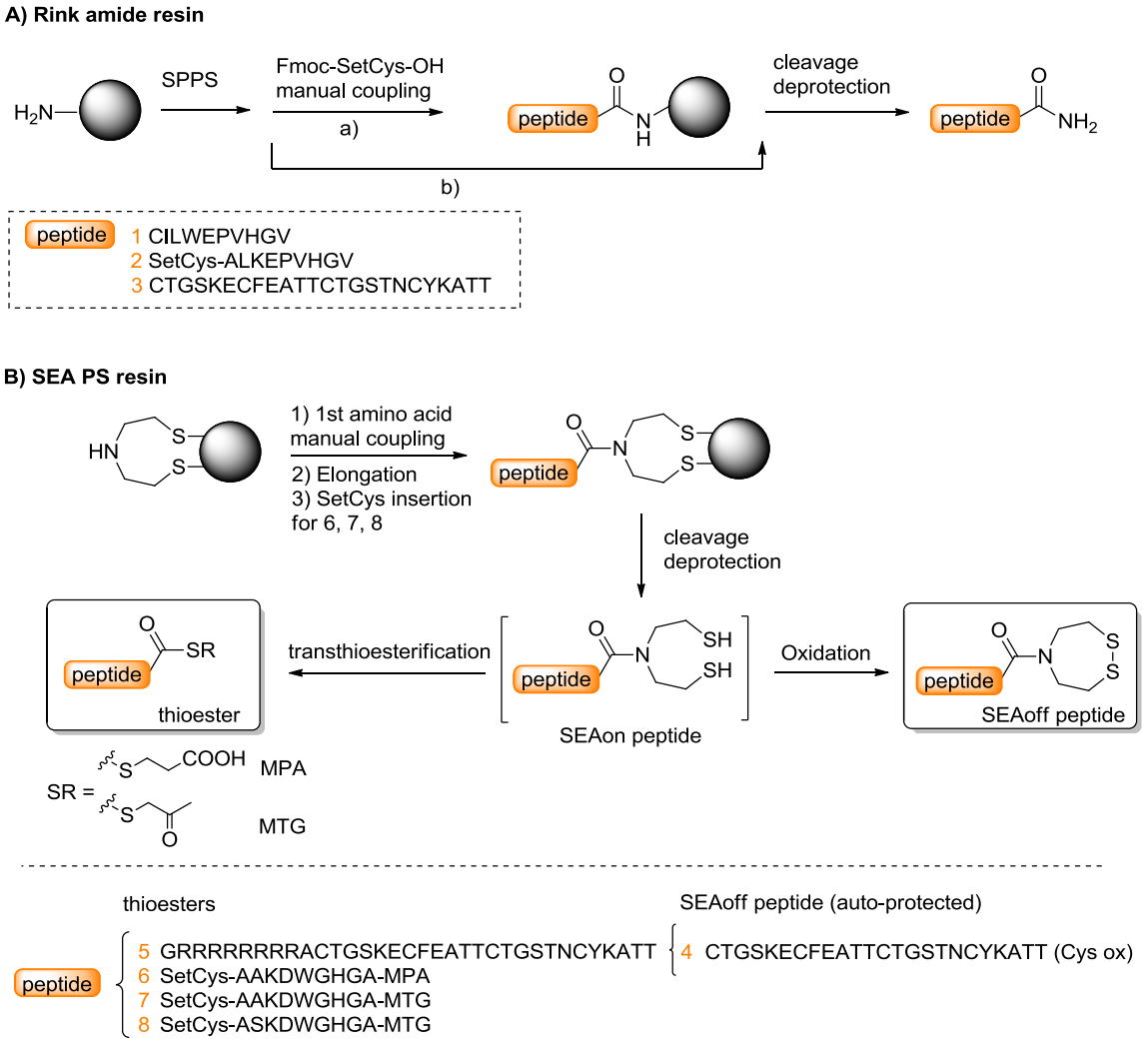


***Figure 20**. General approach for peptides synthesis.*

Automated elongation of the peptide sequences on solid support

Couplings were performed using 5-fold molar excess of each Fmoc-L- amino acid, 4.5-fold molar excess of HBTU, and 10-fold molar excess of DIEA. A capping step was performed after each coupling with DMF/Ac_2O/DIEA 85/10/5, followed by Fmoc deprotection with piperidine/DMF 80/20 (v/v). At the end of the synthesis, the resin was washed with CH_2Cl_2, diethylether (3 × 2 min) and dried in vacuo.

Manual coupling of the Fmoc-SetCys-OH amino acid

To a solution of Fmoc-SetCys-OH (44.8 mg, 0.10 mmol, 2 eq) in DMF (2.5 mL) were successively added HOBt (15.3 mg, 0.10 mmol, 2 eq) and DIC (15.6 µL, 0.10 mmol, 2 eq) and the mixture was stirred at RT for 2 min. The preactivated amino acid was then added to the resin (0.10 mmol, 1 eq) swelled in a minimal volume of DMF. The beads were agitated at RT for 3 h and washed with DMF (3 × 2 min). The resin was then treated with a 80:20 mixture of DMF and piperidine (2 × 10 min) to remove the Fmoc protecting group from SetCys-OH amino acid. Finally, the solid support was washed with DMF (3 × 2 min).

Manual coupling of the first amino acid on SEA PS resin

The first amino acid (10 eq) was coupled to SEA PS resin (1 eq) using HATU (10 eq)/DIEA (20 eq) activation in the minimal volume of DMF. The amino acid was preactivated for 2 min and was then added to the resin swelled in the minimal volume of DMF. The beads were agitated at RT for 1.5 h and the resin was then washed with DMF (3 × 2 min) and drained. The absence of unreacted secondary amino groups was checked using chloranyl colorimetric assay. A capping step was then performed using Ac_2O/DIEA/DMF 10:5:85 (2 × 5 min) and the resin was washed with DMF (3 × 2 min). The resin was then treated with a 80:20 mixture of DMF and piperidine (2 × 10 min) to remove the Fmoc protecting group of the SetCys amino acid. Finally, the solid support was washed with DMF (3 × 2 min) and returned to automated synthesiser to complete the sequence assembling.

Manual coupling of Fmoc-(ΨPro)-OH

To a solution of Fmoc-(ΨPro)-OH (67.88 mg, 0.15 mmol, 1.5 eq) in DMF (2.5 mL) were successively added HATU (54.15 mg, 0.15 mmol, 1.5 eq) and DIEA (34.88 µL, 0.20 mmol, 2 eq) and the mixture was shaken at RT for 1 min. The preactivated amino acid was then added to the resin fragment TCTGSTNCYKATT-(SEA resin) (0.10 mmol, 1 eq) swelled in a minimal volume of DMF. The beads were agitated at RT for 1.5 h and washed with DMF (3 × 2 min). The resin was then treated with 80:20 mixture of DMF and piperidine (2 × 10 min) to remove the Fmoc protecting group of the SetCys amino acid. Finally, the solid support was washed with DMF (3 × 2 min) and returned to automated synthesiser to complete the sequence assembling.

For the part of SetCys based ligation, 10 AA sequences were used, derived from the HIV-1 HXB2 POL region of the viral genome[4]. They are good models as they are facile water soluble and devoid of internal cysteines.

6.2 Synthesis of peptide amide segments

This sub-chapter explains in details the peptide synthesis on Rink resin (**Figure 20, a**).

6.2.1 Synthesis of CILKEPVHGV-NH_2 (1)

Peptide 1 was synthesized on 0.1 mmol scale as described in the general procedure. TFA/H_2O/ TIS / EDT (89/1/5/5) was used as cleavage cocktail. Purification of the crude by HPLC (column C18, 15-40 % in 40 min, % "B" = 25.5 %) provided the titled peptide (59 mg, yield 40 %) as a white powder.

Peptide 1 was characterized by UPLC-MS and MALDI-TOF analysis (**Figure 21**, **Figure 22**).

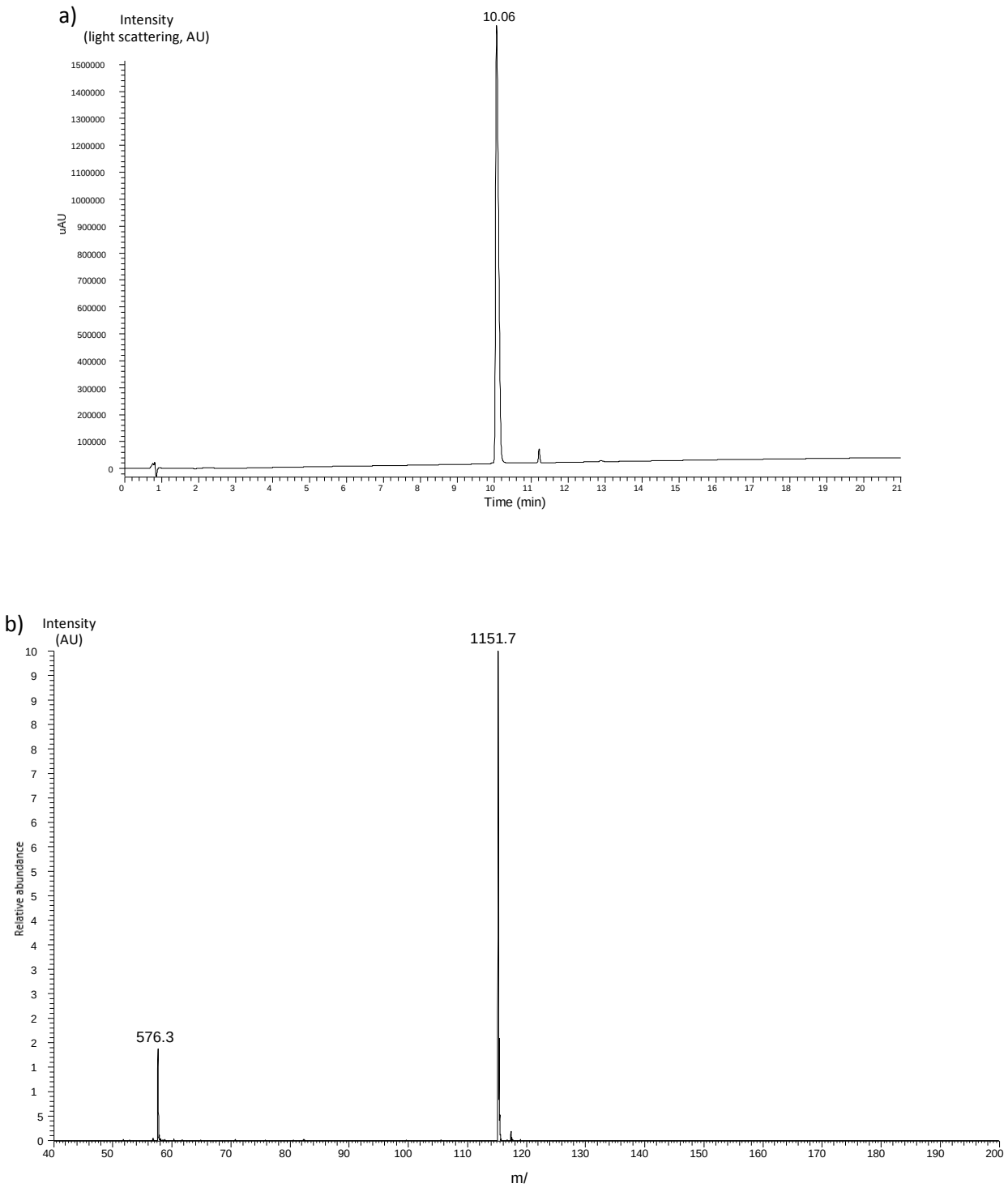


Figure 21. *Peptide **1**. Analysis by UPLC-MS. a) UPLC on C18 column, 50 °C, gradient 0-70 % "B" in 20 min, 0.4 mL min^{-1}, light scattering detection; b) MS trace: m/z for $[M+H]^+$ (monoisotopic mass) 1151.75, $[M+2H]^{2+}$ 576.33, found for M 1150.71. Calcd. for 1151.07.*

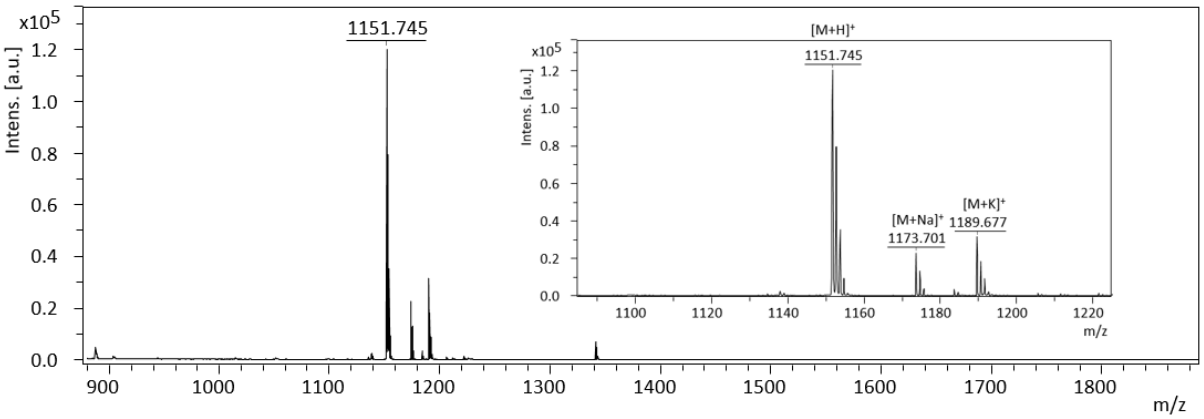
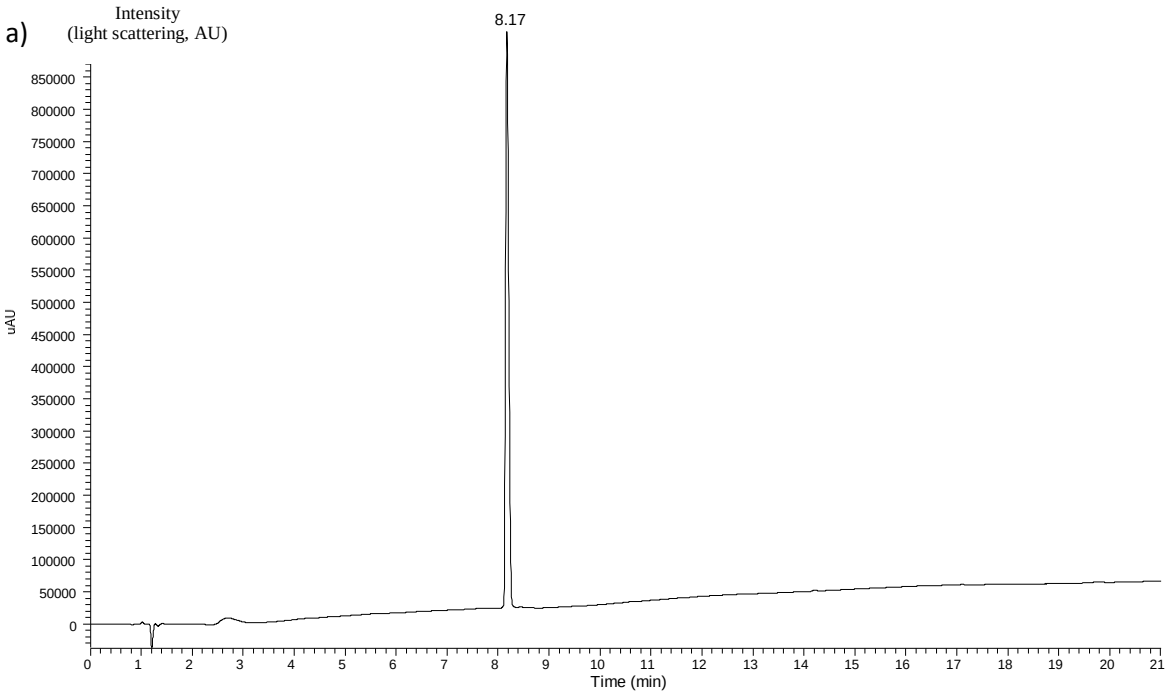


Figure 22. *MALDI-TOF analysis of **1**. Matrix α-cyano-4-hydroxycinnamic acid, positive detection mode, calcd. for [M] : 1151.38, observed mass for [M+H]$^+$: 1151.75 (monoisotopic).*

6.2.2 Synthesis of peptide SetCys-ALKEPVHGV-NH$_2$ (2)

Peptide 2 was synthetized on 0.1 mmol scale as described in the general procedure. TFA/H_2O/ TIS / thioanisol / thiophenol (90.5/2.5/3/2/2) was used as cleavage cocktail. Purification of the crude by HPLC (column C18, 10-30 % "B" in 30 min; % "B" = 21.2) provided the titled peptide (46 mg, yield 54 %) as a white powder.

Peptide 2 was characterized by UPLC-MS and MALDI-TOF analysis (**Figure 23**, **Figure 24**).



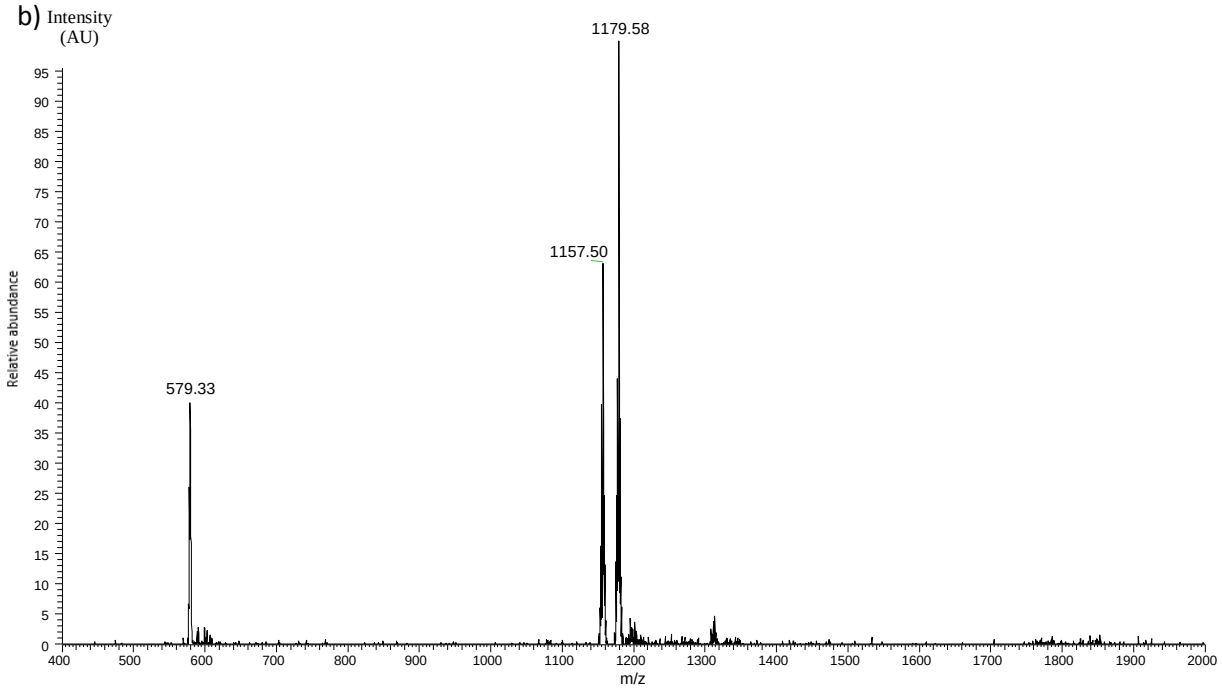


Figure 23. *Peptide **2**. Analysis by UPLC-MS. a) UPLC on C18 column, 50 °C, gradient 0-70 % "B" in 20 min, 0.4 mL min^{-1}, light scattering detection; b) MS trace: m/z for [M+H]$^{+}$ (monoisotopic mass) 1157.50, [M+2H]$^{2+}$ 579.33, found for M 1156.58. Calcd. for M 1156.12.*

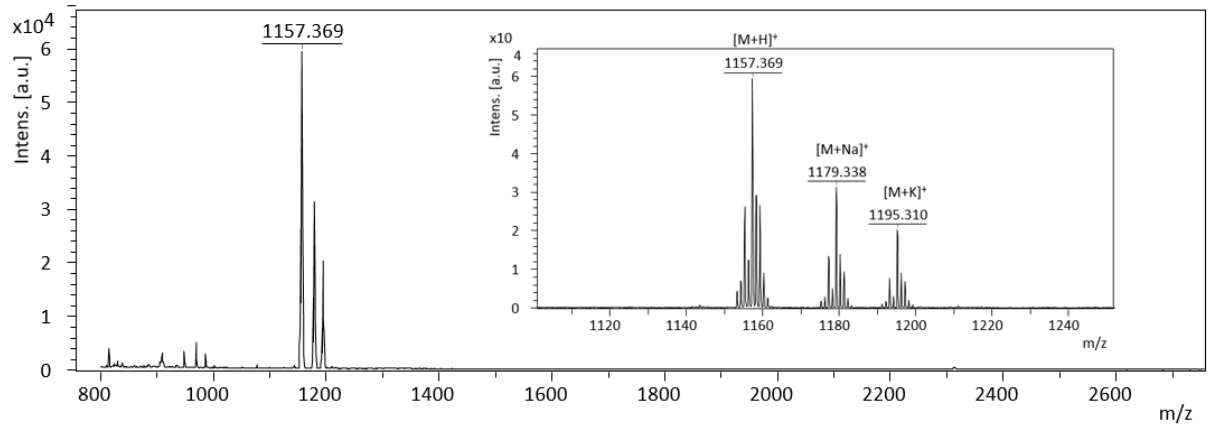


Figure 24. *MALDI-TOF analysis of **2**. Matrix α-cyano-4-hydroxycinnamic acid, positive detection mode, calcd. for [M] : 1156.12, observed mass for [M+H]$^{+}$: 1157.33 (monoisotopic).*

6.2.3 Synthesis of CTGSKECFEATTCTGSTNCYKATT-NH_2 (3)

Peptide 3 was syntheized on 0.1 mmol scale as described in the general procedure. To manually introduce Fmoc-(ΨPro)-OH (AT), the fragment TCTGSTNCYKATT-NH_2 was removed from automated synthesizer and transferred into a reactor for manual coupling. Once Fmoc-(ΨPro)-OH is introduced, the automated synthesis can be pursued. TFA/H_2O/ TIS / EDT (89/1/5/5) was used as cleavage cocktail. Purification of the crude by HPLC (column

C18, 0-30 “B” in 60 min; % “B” = 23.30) provided the titled peptide (112 mg, yield 35%) as a white powder.

Peptide 3 was characterized by UPLC-MS and MALDI-TOF analysis (**Figure 25**, **Figure 26**).

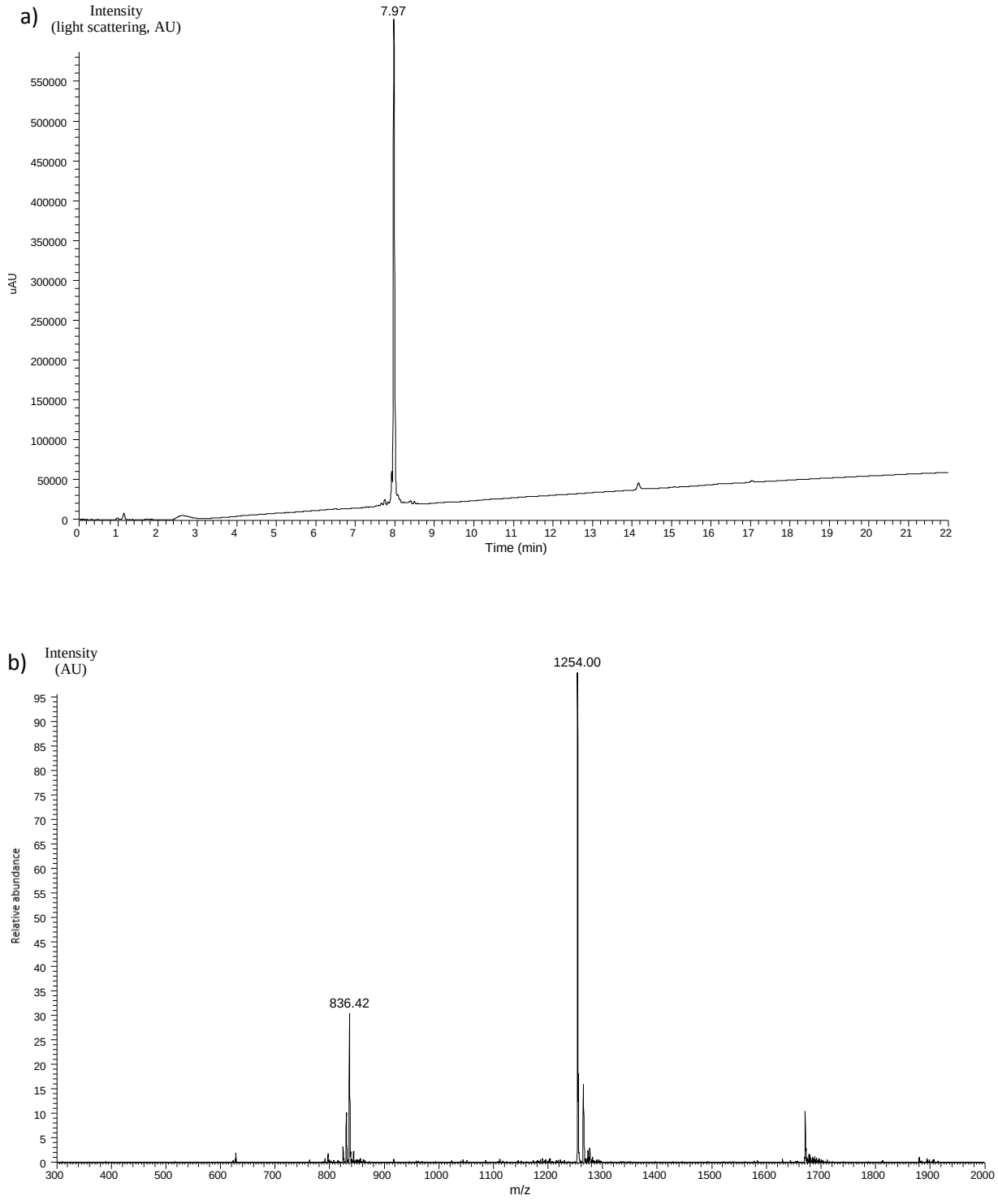


Figure 25. *Peptide **3**. Analysis by UPLC-MS. a) UPLC on C18 column, 50 °C, gradient 0-70 % “B” in 20 min, 0.4 mL min^{-1}, light scattering detection; b) MS trace: m/z for $[M+2H]^{2+}$ 1254.00, $[M+3H]^{3+}$ 836.42, found for M (average) 2506.13. Calcd. for M 2505.03.*

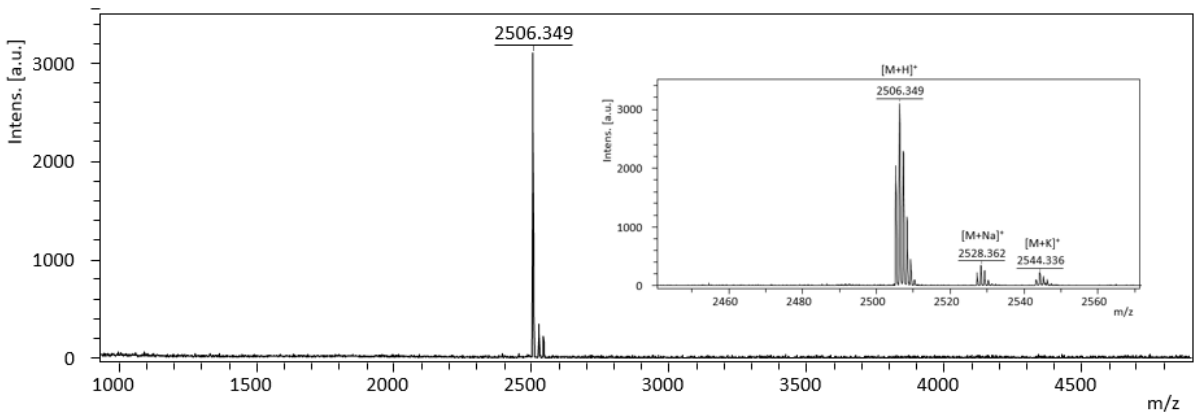


Figure 26. *MALDI-TOF analysis of **3**. Matrix α-cyano-4-hydroxycinnamic acid, positive detection mode, calcd. for [M] : 2505.03, observed mass for $[M+H]^+$: 2506.35 (monoisotopic).*

6.3 Synthesis of SEA^{off} peptide segments

Typical procedures for the synthesis of SEA^{off} peptide segments (**Figure 20, B**) were described in previous papers[5,6]. For a detailed protocol see the protocol article[7] (**Annex D**).

After the resin cleavage of the sequence CTGSKECFEATTCTGSTNCYKATT-SEA^{on}, to convert SEA^{on}→SEA^{off}, iodine solution is classically used. It demands to use S-*t*Bu protecting group for internal cysteines to avoid random disulfide bridges formation. However, this PG makes peptide segments poorly soluble in water. To overcome the problem, cystine was used to convert SEA^{on}→SEA^{off} and to protect internal cysteines in folding-like manner. Unlikely classical method with glutathione redox use, here the peptide concentration is higher and only cystine$_{ox}$ is used, as concentration of internal cysteines becomes significant and takes the role of cysteine$_{red}$. Note that 1 mM of the peptide (**Figure 27**) contains 4 internal cysteines and 2×SH of SEA → 6 mM of SH in total.

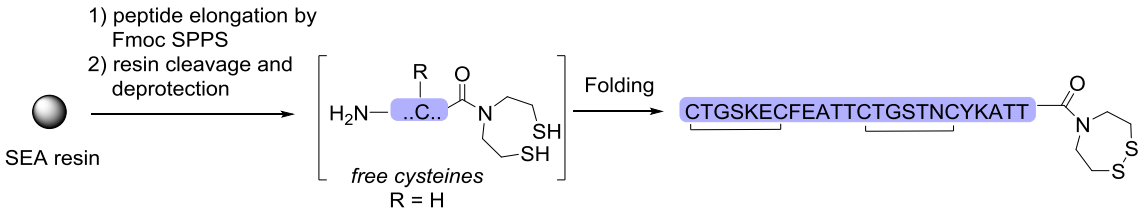


Figure 27*. Self-protection of thiols via folding process.*

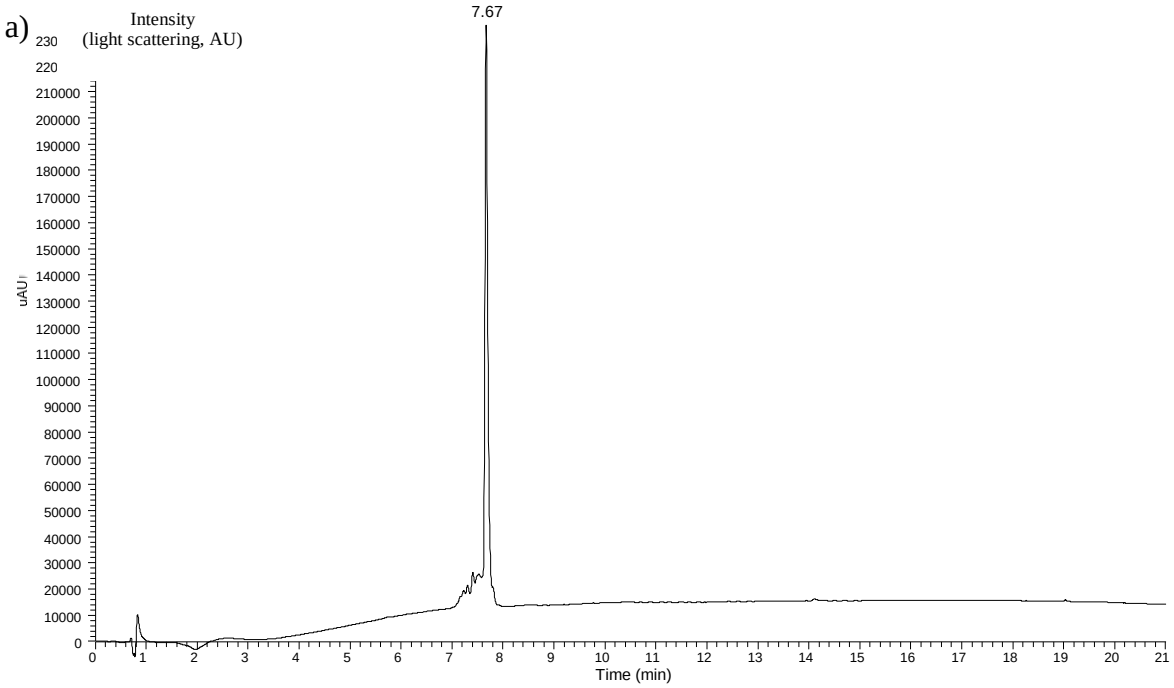
6.3.1 Synthesis of CTGSKECFEATTCTGSTNCYKATT-SEA^{off} (4)

Peptide 4 was synthesized on 0.1 mmol scale as described in the general procedure. TFA/H_2O/ TIS / EDT (89/1/5/5) was used as cleavage cocktail.

Powder of cystine (0.5 mg/mL) was added in the solution of PBS 0.1 M containing 10% of glycerol by volume. The mixture was acidified to pH 2-3 with HCl 6 M stirred and heated at 50 °C with aid of oil bath for 2 h. When cystine is solubilized completely, the

solution was cooled down to RT and the pH was increased at 7.2 with NaOH 6 M. Then the peptide 1 mM (20 mg) final concentration was dissolved in Gdn·HCl 6 M (87 μL) and mixed with 7.4 mL of cystine solution. This mixture was incubated at 25°C for 17h (cystine: cysteine ≡ 1: 3). Within this time all disulfide bridges were correctly formed. The reaction was followed by HPLC-MS.

Upon the reaction completion, the reaction mixture was centrifuged to separate misfolded precipitates. The folded peptide containing in the liquid phase was then purified by HPLC (column C18, gradient 10-30% "B" in 35 min, % "B" = 18.2 %), resulting the titled peptide (13.6 mg, yield 27%) as a white powder. Peptide 4 was characterized by UPLC-MS and MALDI-TOF analysis (**Figure 28**).



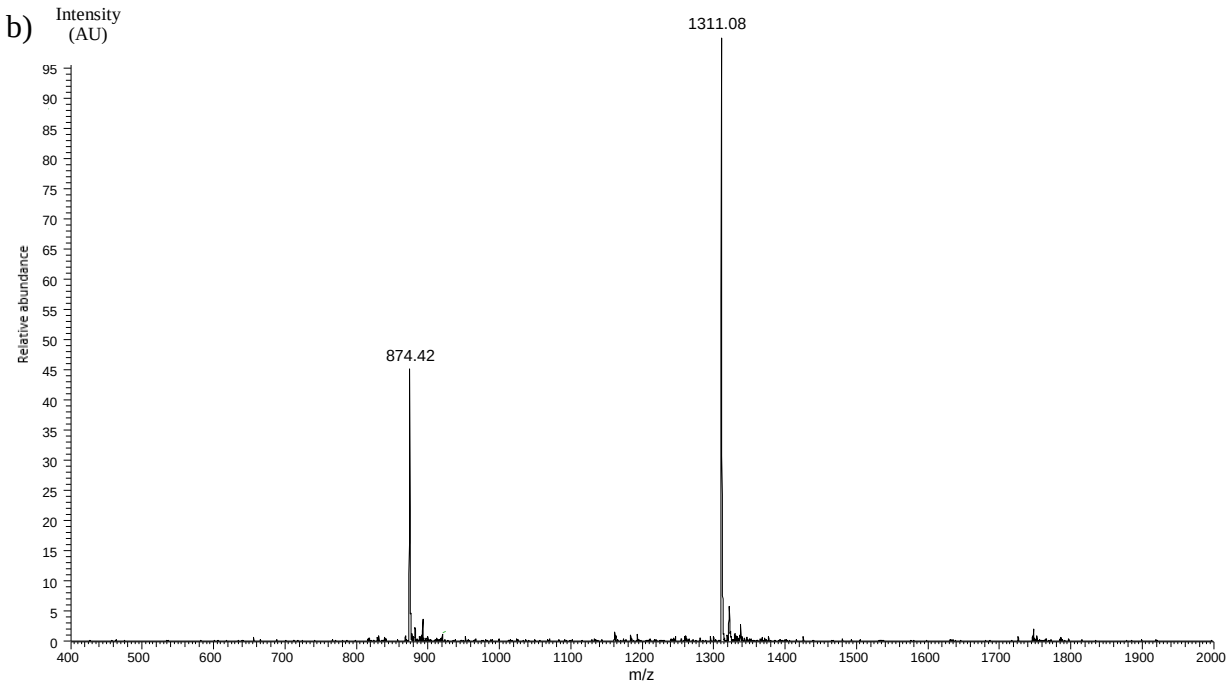


Figure 28. *Peptide **4**. Analysis by UPLC-MS. a) UPLC on C18 column, 50 °C, gradient 0-70 % "B" in 20 min, 0.4 mL min^{-1}, light scattering detection; b) MS trace: m/z for $[M+2H]^{2+}$ 1311.08, $[M+3H]^{3+}$ 874.42, found for M (average) 2620.21. Calcd. for M 2620.97.*

6.4 Synthesis of peptidyl thioester that doesn't include SetCys residue in their sequence

The SEA^{on} peptide recovered after precipitation was converted into the corresponding MPA thioester by reaction with MPA or MTG (5% in volume) at pH 4.0.

MPA (0.75 mL) or MTG (0.75 mL), Gdn·HCl 6 M aqueous (14.25 mL), TCEP 100 mM (487.3 mg), the cleaved peptide 0.1 mmol were introduced into a glovebox in separate vessels. After letting the reagents to degase inside a glovebox for 20 min, MPA or MTG and TCEP were mixed together. Then, the pH was set as 4.0 with help of NaOH 6 M. Then the obtained solution was poured into the peptide to solubilize it. The reaction was allowed to proceed at 37 °C. The transthioesterification progress was monitored by UPLC-MS by collecting 5 µL aliquots and quenching them in 100 µL of AcOH 10 %. Once the process is over (monitored by HPLC), concentrated AcOH (1.5 mL) was added. The reaction mixture was removed from glovebox and extracted with Et_2O (4 ×10 mL) to remove MPA / MTG under fuming hood, to avoid the mercaptan smell. The crude solution was filtered through a disposable syringe filter prior to the purification.

6.4.1 Synthesis of GRRRRRRRRACTGSKECFEATTCTGSTNCYKATT-MPA (5)

Peptide 5 was synthetized on 0.1 mmol scale as described in the general procedure. TFA/H_2O/ TIS / thioanisol / thiophenol (87/3/5/2.5/2.5) was used as cleavage cocktail.

The transthioesterification was performed following the general procedure 6.4. MPA (0.75 mL) was used as added thiol. The peptide 5 was purified by HPLC (column C3, 70 °C, gradient 0-30% "B" in 40 min, % "B" = 18.2 %) resulting the titled peptide (181 mg, 51 %) as

a white powder. Peptide 5 was characterized by UPLC-MS and MALDI-TOF analysis (**Figure 29**).

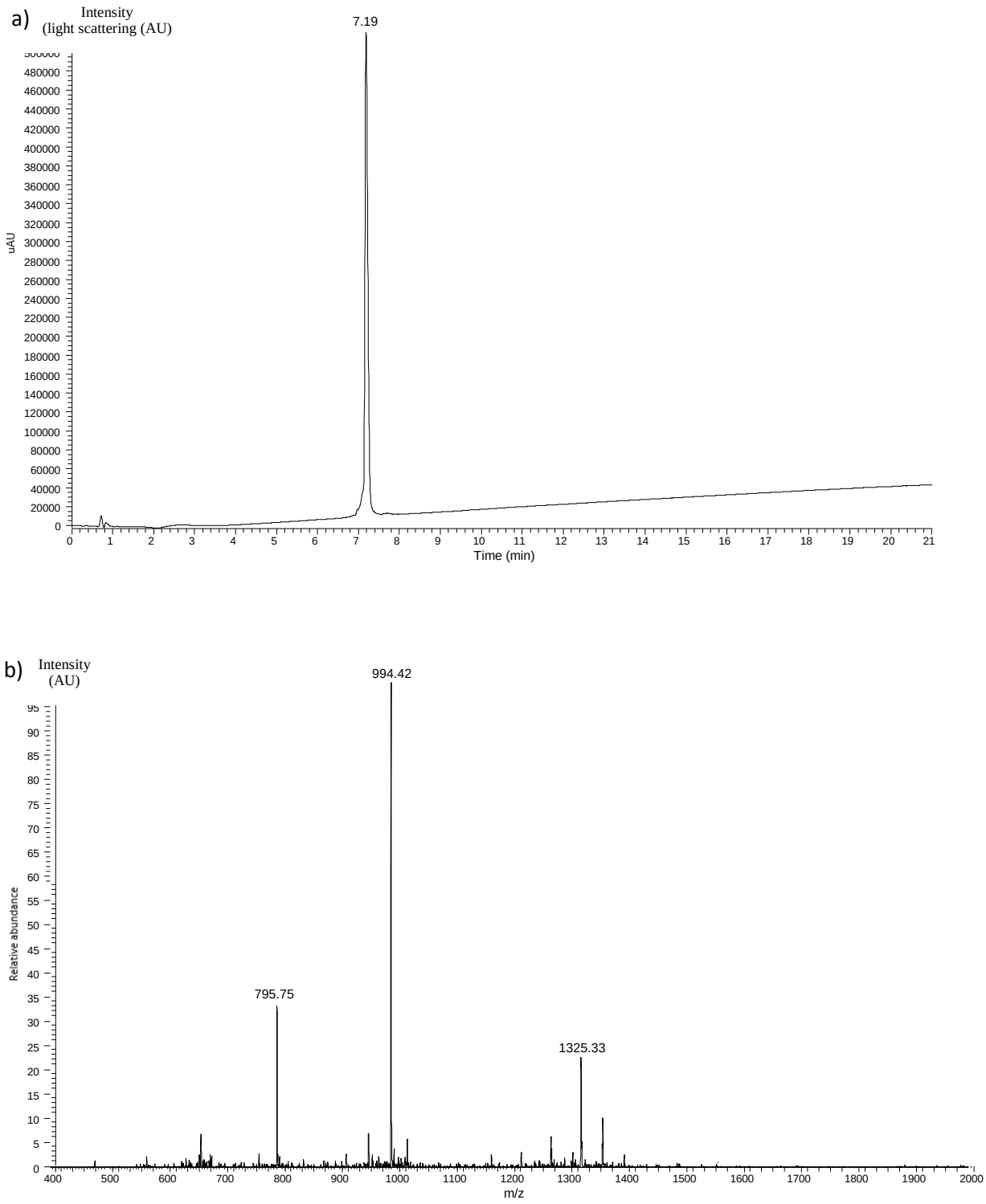


Figure 29. *Peptide **5**. Analysis by UPLC-MS. a) UPLC on C3 column, 70 °C, gradient 0-70 % "B" in 20 min, 0.4 mL min^{-1}, light scattering detection; b) MS trace: m/z for $[M+3H]^{3+}$ 1325.33, $[M+4H]^{4+}$ 994.42, $[M+5H]^{5+}$ 795.75, found for M (average) 3973.34. Calcd. for M 3973.51.*

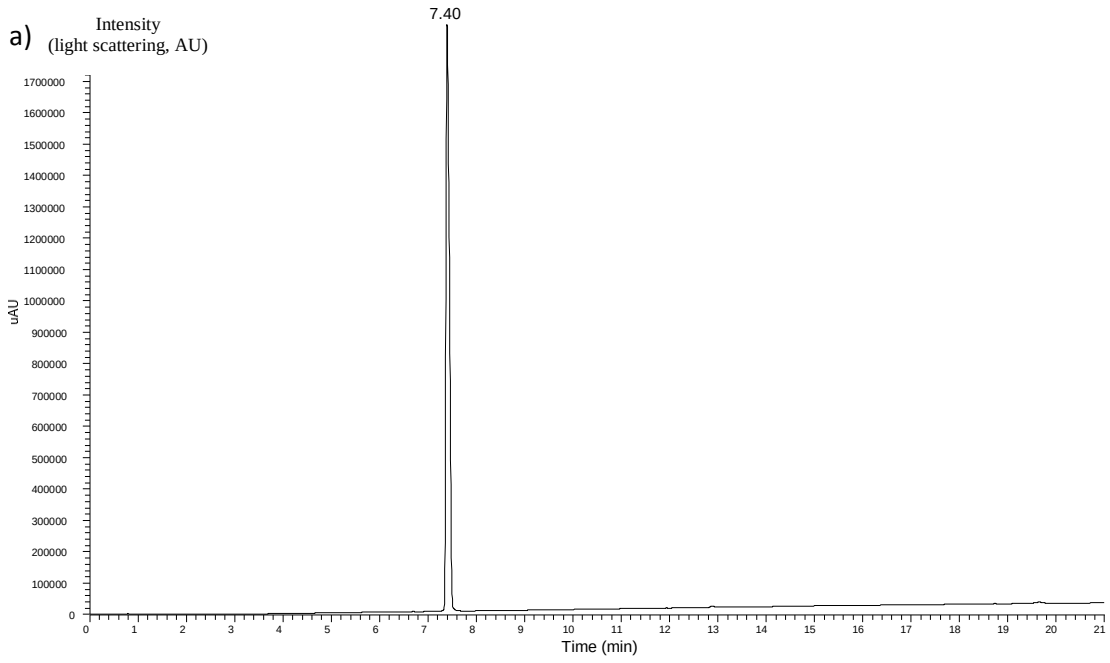
6.5 Synthesis of peptidyl thioesters that include SetCys residue in their sequence

The SEA^{on} peptide recovered after precipitation was converted into the corresponding MPA thioester by reaction with MPA or MTG (5% in volume) at pH 4.0. MPA (0.75 mL) or MTG (0.75 mL), Gdn·HCl 6 M aqueous (14.25 mL), the cleaved peptide 0.1 mmol were introduced into a glovebox in separate vessels. After letting the reagents to degase inside a glovebox for 20 min, MPA or MTG and TCEP were mixed together. Then, the pH was set as 4.0 with help of NaOH 6 M. Then the obtained solution was poured into the peptide to solubilize it. The reaction was allowed to proceed at 37 °C. The trasthioesterification progress was monitored by UPLC-MS by collecting 5 µL aliquots and quenching them in 100 µL of AcOH 10 %. Once the process is over (monitored by HPLC), concentrated AcOH (1.5 mL) was added. The reaction mixture was removed from glovebox and extracted with Et_2O (4 ×10 mL) to remove MPA / MTG under fuming hood, to avoid the mercaptane smell. The crude solution was filtered through a disposable syringe filter prior to the purification.

6.5.1 Synthesis of SetCys-AAKDWGHGA-MPA (6)

Peptide 6 was synthesized on 0.1 mmol scale as described in the general procedure 6.5. TFA/H_2O/ TIS / thioanisol / thiophenol (89.5/2.5/3/2.5/2.5) was used as cleavage cocktail.

The transthioesterification was performed following the general procedure 6.5. MPA (0.75 mL) was used as added thiol. The peptide 6 was purified by HPLC (column C18, gradient 5-30% “B” in 40 min, % “B” = 17.5 %) resulting the titled peptide as white powder (15.23 mg, yield 13 %) as a white powder. Peptide 6 was characterized by UPLC-MS analysis (**Figure 30**).



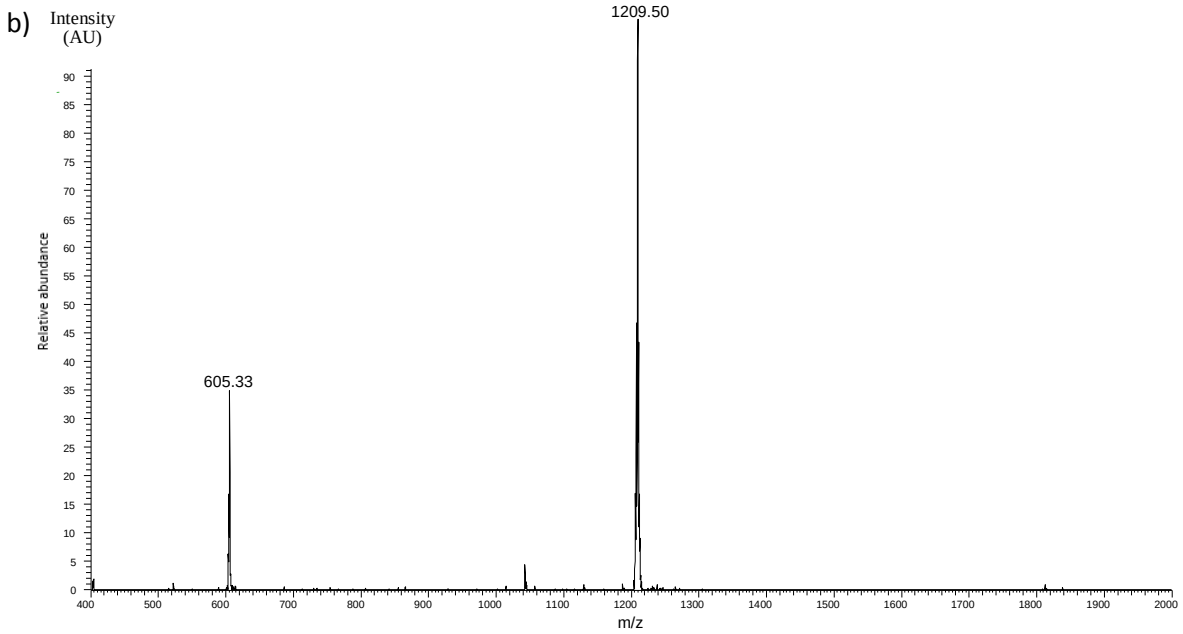


Figure 30. *Peptide **6**. Analysis by UPLC-MS. a) UPLC on C18 column, 50 °C, gradient 0-70 % "B" in 20 min, 0.4 mL min^{-1}, light scattering detection; b) MS trace: m/z for [M+H]$^{+}$ (monoisotopic mass) 1209.50, [M+2H]$^{2+}$ 605.33, found for M 1208.58. Calcd. for M 1208.09.*

6.5.2 Synthesis of SetCys-AAKDWGHGA-MTG (7)

Peptide 7 was synthesized on 0.1 mmol scale as described in the general procedure 6.5. TFA/H_2O/ TIS / thioanisol / thiophenol (89.5/2.5/3/2.5/2.5) was used as cleavage cocktail.

The transthioesterification was performed following the general procedure 6.5. MTG (0.75 mL) was used as added thiol. The peptide 7 was purified by HPLC (column C18, gradient 10-30% "B" in 40 min, % "B" = 19.1 %) resulting 13 % yield, 14.5 mg as a white powder. Peptide 7 was characterized by UPLC-MS (**Figure 31**).

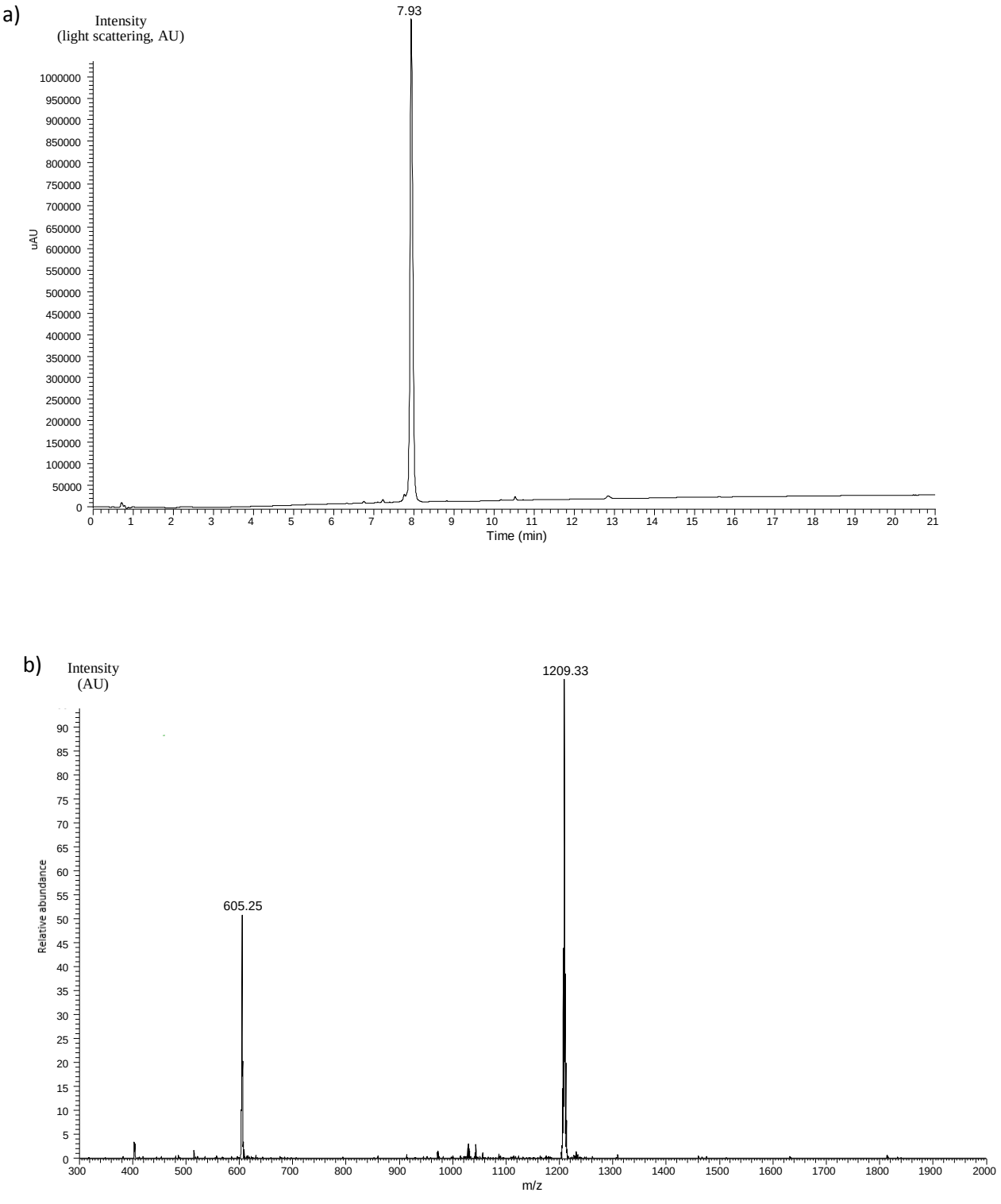


Figure 31. *Peptide **7**. Analysis by UPLC-MS. a) UPLC on C18 column, 50 °C, gradient 0-70 % "B" in 20 min, 0.4 mL min^{-1}, light scattering detection; b) MS trace: m/z. for $[M+H]^+$ 1209.33 (monoisotopic), $[M+2H]^{2+}$ 605.25, found for M 1208.42. Calcd. for M 1208.09.*

6.5.3 Synthesis of SetCys-ASKDWGHGA-MTG (8)

Peptide 8 was synthesized on 0.1 mmol scale as described in the general procedure 6.5. TFA/H_2O/ TIS / thioanisol / thiophenol (89.5/2.5/3/2.5/2.5) was used as cleavage cocktail.

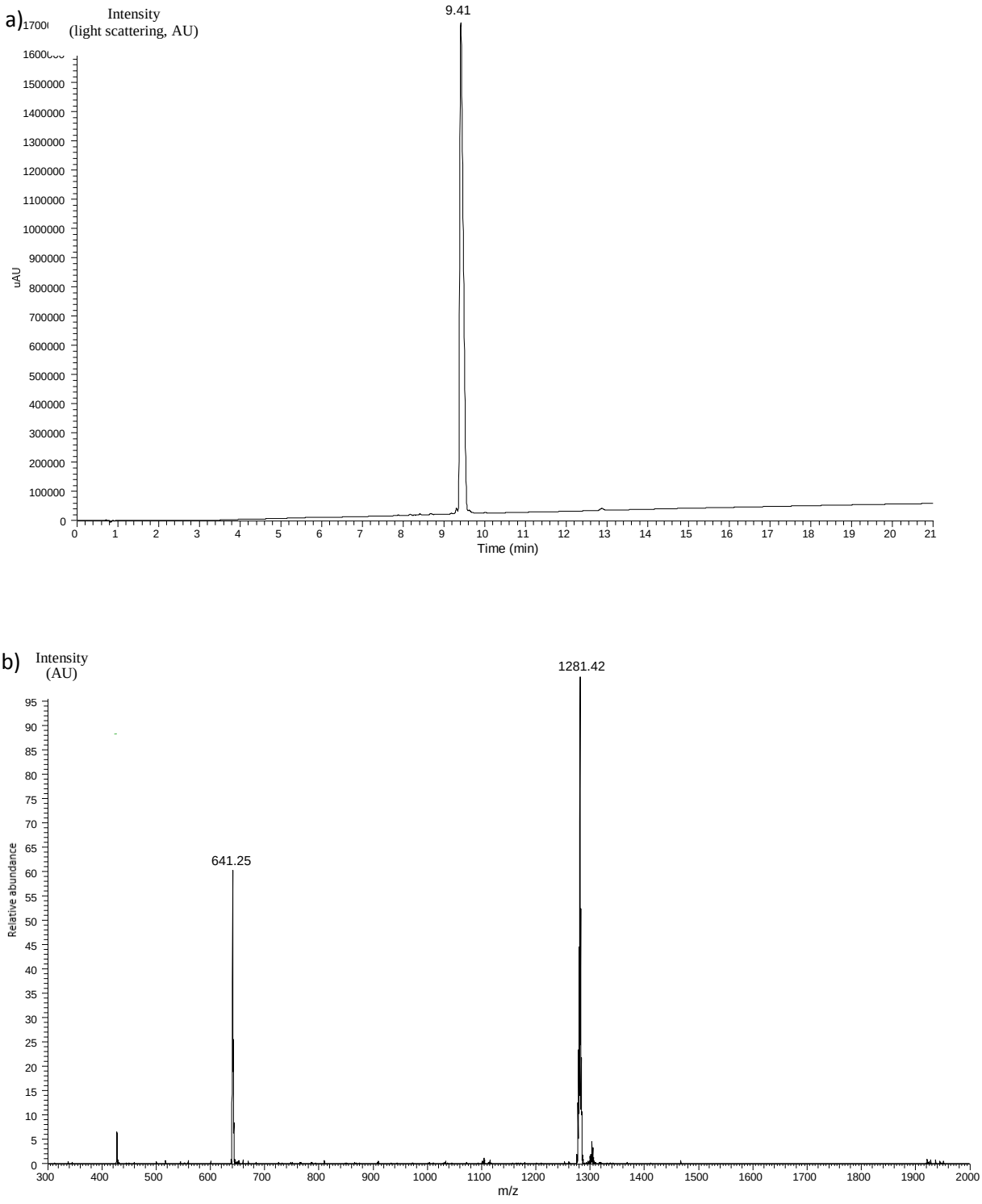


Figure 32. *Peptide **8**. Analysis by UPLC-MS. a) UPLC on C18 column, 50 °C, gradient 0-70 % "B" in 20 min, 0.4 mL min^{-1}, light scattering detection; b). MS trace: m/z found for $[M+H]^+$ (monoisotopic mass) 1281.42, $[M+2H]^{2+}$ 641.25, found for M 1280.46. Calcd. for M 1280.54.*

The transthioesterification was performed following the general procedure 6.5. MTG (0.75 mL) was used as added thiol. The peptide 8 was purified by HPLC (column C18, gradient 15-35% "B" in 40 min, % "B" = 24.5 %) resulting 14 % yield, 16 mg as a white powder. Peptide 8 was characterized by UPLC-MS analysis (**Figure 32**).

7 SetCys reactivity assays

All preliminary studies of SetCys chemical properties under desired conditions were tested on SetCys amided peptide (**Figure 33**). The reactions were performed under nitrogen atmosphere. All reagents were placed into a glovebox prior to manipulation.

The reactions were carried out using a hotplate stirrer equipped with a heating block. The conversion of SetCys into Cys was monitored by HPLC at 215 nm (both C18 and C3 columns may be used; 50 °C, gradient 0-40 % "B"). For each analysis, a 4 µL aliquot was taken from the reaction mixture and quenched by adding 10% AcOH in water (100 µL). The sample was then extracted with Et_2O to remove MPAA prior to HPLC analysis. In case when MPAA is not added, no extraction with Et_2O needed.

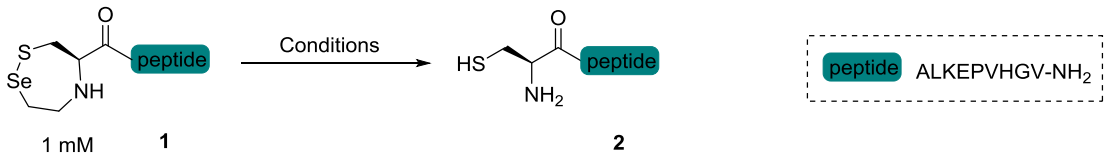


Figure 33. *Assays on SetCys to Cys conversion under various conditions.*

7.1 The choice of appropriate phosphine

In this experiment, the influence of phosphines on SetCys conversion was studied. TCEP, TPPTS, and other aromatic phosphines were tested. TTFMPP (Bis(3-sulfonatophenyl)(3,5-di-trifluoroMethylphenyl)phosphine disodium salt), TMPP (Tris(4,6-dimethyl-3-sulfanatophenyl)phosphine trisodium salt hydrate) were not soluble at required concentration. The method hereby is given only for TPPTS (**Figure 34**).

Gdn·HCl (574 mg, 6 M) were dissolved in sodium phosphate buffer 0.1 M, pH 7.0 (600 µL), resulting a total volume of 1 mL. In the obtained solution were progressively added: sodium ascorbate (18.6 mg, 100 mM), TPPTS (5.68 mg, 10 mM) and MPAA (33.6 mg, 200 mM). The pH of the mixture was adjusted to 6.5 by adding 6 M NaOH. SetCys peptide 1 (0.48 mg, 1 mM) was then dissolved in 320 µL of the above solution and the reaction mixture was heated at 37 °C.

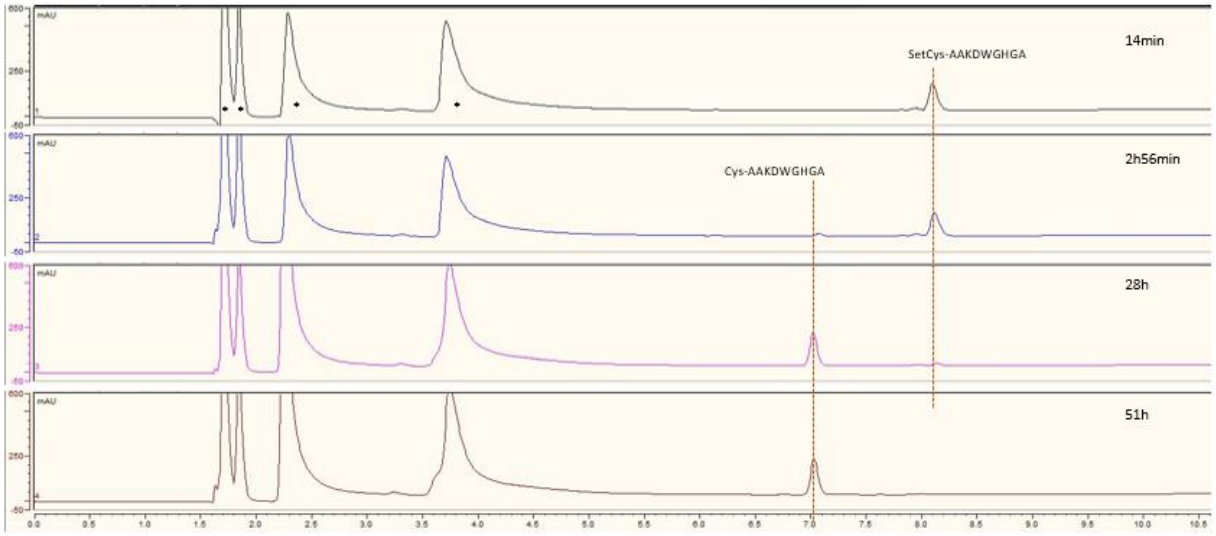


Figure 34. *HPLC profile of SetCys to Cys conversion.*

7.2 Phosphine concentration assays

The essay was performed first for TCEP and then for TPPTS. The method hereby is given only for TPPTS.

Gdn·HCl (287 mg, 6 M) were dissolved in sodium phosphate buffer 0.1 M, pH 7.0 (300 µL), resulting a total volume of 500 µL. In the obtained solution were progressively added: sodium ascorbate (9.9 mg, 100 mM), TPPTS (9.94 mg, 35 mM / 19.88 mg, 70 mM) and MPAA (16.8 mg, 200 mM). The pH of the mixture was adjusted to 6.0 by adding 6 M NaOH. SetCys peptide 1 (0.48 mg, 1 mM) was then dissolved in 320 µL of the above solution and the reaction mixture was heated at 37 °C.

7.3 Influence of temperature on selenoethyl arm loss

Gdn·HCl (287 mg, 6 M) were dissolved in sodium phosphate buffer 0.1 M, pH 7.0 (300 µL), resulting a total volume of 500 µL. In the obtained solution were progressively added: sodium ascorbate (9.9 mg, 100 mM), TPPTS (9.94 mg, 35 mM / 19.88 mg, 70 mM) and MPAA (16.8 mg, 200 mM). The pH of the mixture was adjusted to 6.0 by adding 6 M NaOH. SetCys peptide 1 (0.48 mg, 1 mM) was then dissolved in 320 µL of the above solution and the reaction mixture was heated at 37 /50 / 60 °C.

7.4 TPPTS quality control

TPPTS contamination with formaldehyde was observed upon long storage of TPPTS in commercial flask made of a polymer. This contamination was detected by thiazolidine formation (

Figure *35*, see in annex B, supplementary information). To check the phosphine quality, the following procedure was applied.

Cys-ILWEPVHGV peptide **1** (1 mM final concentration) was dissolved in sodium phosphate buffer (0.1 M, pH 6.0) containing TPPTS (70 mM final concentration). The solution was incubated at 37 °C and analyzed by UPLC-MS (

Figure 36). For each analysis, a 4 µL aliquot was taken from the reaction mixture and quenched by adding 10% AcOH in water (100 µL). The samples were injected in UPLC-MS system (27 µL).

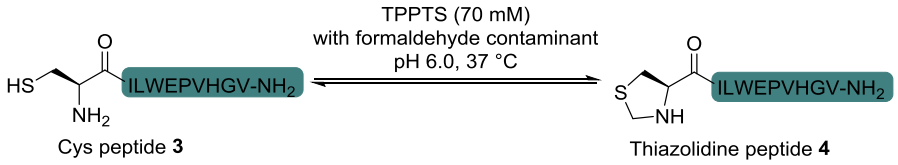


Figure 35. *The presence of formaldehyde in TPPTS can be assayed through the propensity of Cys peptides to form stable thiazolidine products upon reaction with aldehydes.*

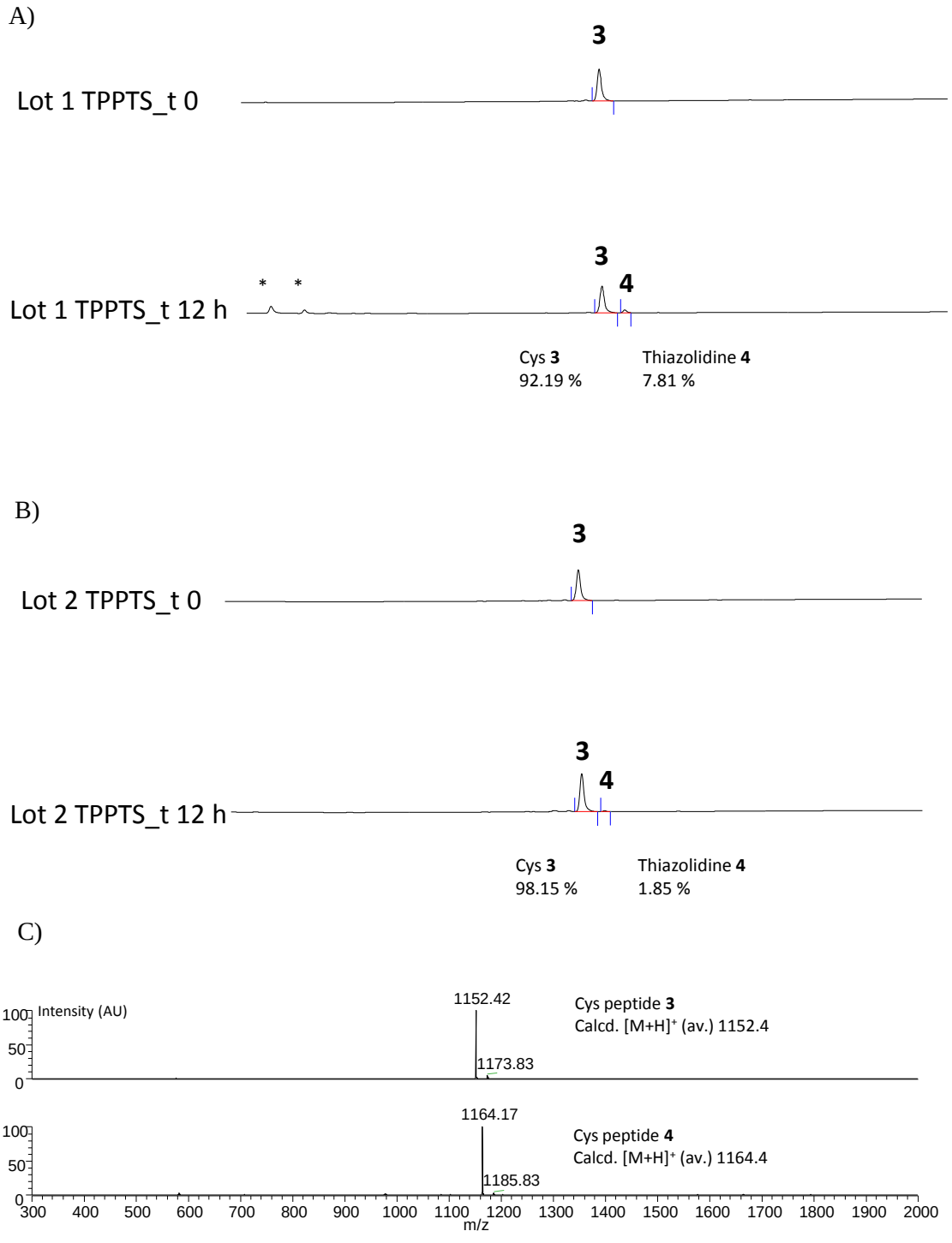


Figure 36. *Analysis of thiazolidine by-product formation after 12 h. A) HPLC chromatogram obtained with TPPTS lot 1 (Sigma Aldrich). B) HPLC chromatogram obtained with TPPTS lot 2 (TCI). C) MS traces for 3 and 4. *: non-peptidic.*

8 Adaptation of Native chemical ligation

As the reaction conditions were adapted for SetCys conversion, the same was done for NCL step. To find the compromise conditions, many parameters were tested with simple thioester peptides reacting with Cys-peptide. Next, the compatibility of the most promising conditions with the bifunctional peptide was verified: presence of TPPTS, specific temperature and pH.

8.1 Experiments with simple MPA-peptide devoid of SetCys under classical reducing conditions

To improve the NCL kinetics and reaction selectivity, following parameters were variated: concentration of peptides, MPAA, Na-ascorbate, different pH, presence or absence of a phosphine. The sequence RLKEPVHGA-MPA was used because it was included in preliminary studies of this project (**Figure 37**).

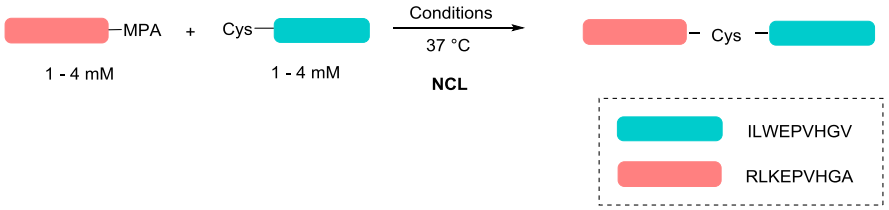


Figure 37. *General scheme of conditions optimization for NCL between simple MPA thioester peptide and Cys-peptide.*

The following procedure is described for pH optimization, in the absence of phosphine and Na-ascorbate. Peptides concentration for most of essays was chosen 1 mM as max dilution and the worst-case scenario for NCL kinetics.

Gdn·HCl (574 mg, 6 M) were dissolved in sodium phosphate buffer 0.1 M, pH 7.0 (600 µL), resulting a total volume of 1 mL. This solution was inserted into MPAA (33.6 mg, 200 mM) and vigorously vortexed to facilitate MPAA solubilisation. The pH of the mixture was adjusted to 7.0 by adding 6 M NaOH. The obtained solution (550 µL) was inserted into Cys-peptide (759 µg, 1 mM) and then into MPA peptide (781 µg, 1 mM). The reaction mixture was incubated at 37 °C. The reaction was followed by HPLC (column C18, 50 °C, gradient 12-38 % "B" in 15 min, 1 mL/min) at wavelength 215 nm (**Figure 38**).

For each analysis, a 4 µL aliquot was taken from the reaction mixture and quenched by adding 10% AcOH in water (100 µL). The samples were injected in UPLC-MS system (15 µL).

Note that when Gdn·HCl, MPAA, Na-ascorbate and a phosphine are all required for the reaction, they must be mixed in following order: Gdn·HCl, Na-ascorbate, phosphine, MPAA. it is due to different solubility and roles of additives.

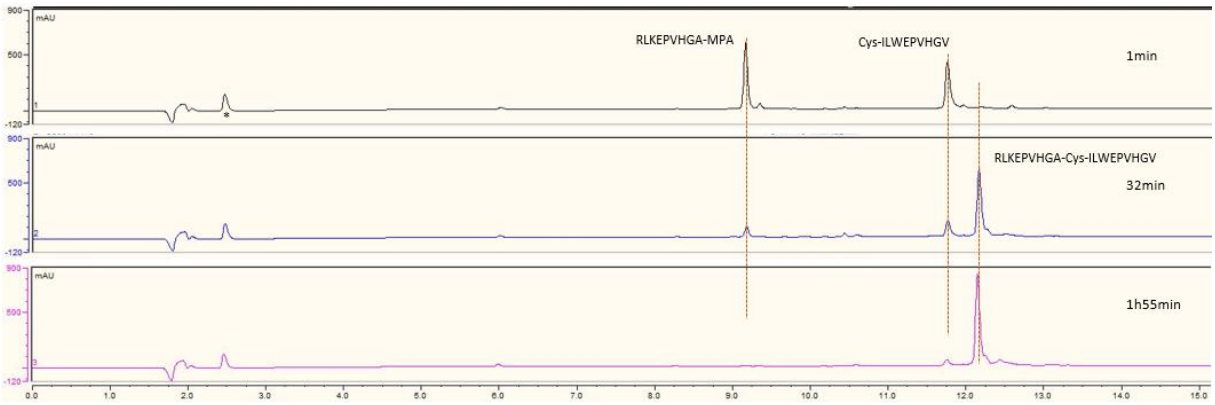


Figure 38. *HPLC profile of HPLC in reducing conditions.*

8.2 Buffer nature studies on NCL with bifunctional peptide

It was demonstrated that phosphate buffer may co-catalyze thiol-thoester exchange (**Figure 39**). In some cases, it can increase the thioester hydrolysis level. This is why the following procedure was carried out with phosphate buffer 0.1 M and MES buffer 0.1 M, to compare their effects. No difference between both buffers was detected.

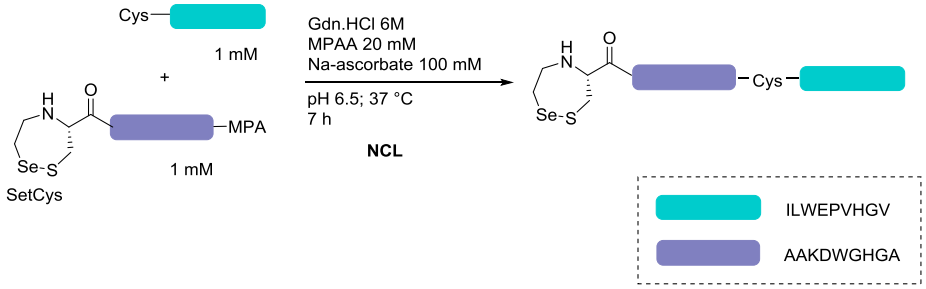


Figure 39. *NCL with bifunctional peptide carried out in different buffers.*

Gdn·HCl (574 mg, 6 M) were dissolved in MES 0.1 M, pH 7.0 (600 µL), resulting a total volume of 1 mL. Na-ascorbate (18.6 mg, 100 mM) was dissolved in given solution. This solution was inserted into MPAA (3.36 mg, 20 mM) and vigorously vortexed to facilitate MPAA solubilisation. The pH of the mixture was adjusted to 6.5 by adding 6 M NaOH. The obtained solution (330 µL) was inserted into Cys-peptide 1 (441.5 µg, 1 mM) and then into MPA peptide 6 (496.3 µg, 1 mM). The reaction mixture was incubated at 37 °C for 7 h. The reaction was followed by HPLC (column C18, 50 °C, gradient 12-38 % "B" in 15 min, 1 mL/min) at wavelengths 215 nm and 280 nm (**Figure 40**).

For each analysis, a 4 µL aliquot was taken from the reaction mixture and quenched by adding 10% AcOH in water (100 µL). The samples were injected in UPLC-MS system (15 µL).

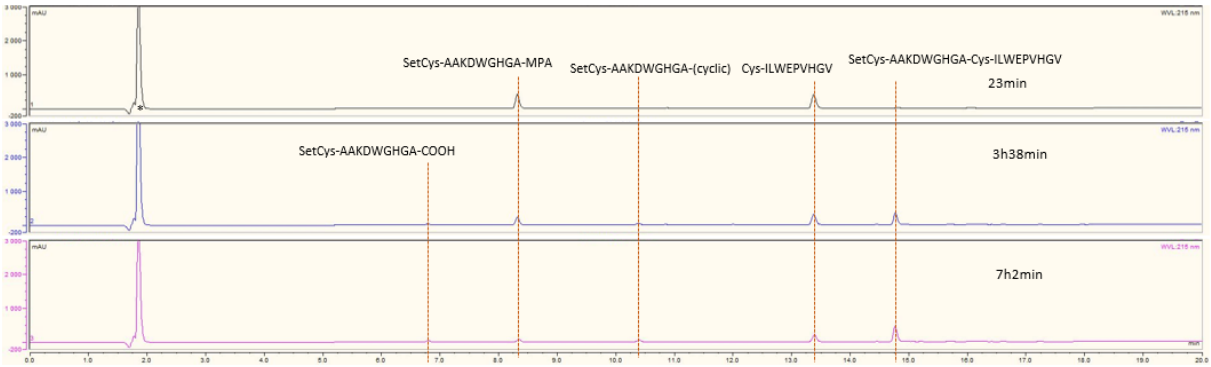


Figure 40. *HPLC profile of NCL in reducing media with bifunctional thioester peptide.*

8.3 Experiments with simple MPA-peptides devoid of SetCys under non-reducing imidazole media

Imidazole plays the role of nucleophilic catalyst and has its buffering capacity. Thus, no other buffer needed. To adapt non-reducing conditions, the pH, reaction temperature and additives concentration (imidazole, Na ascorbate, TPPTS) were adjusted (**Figure 41**).

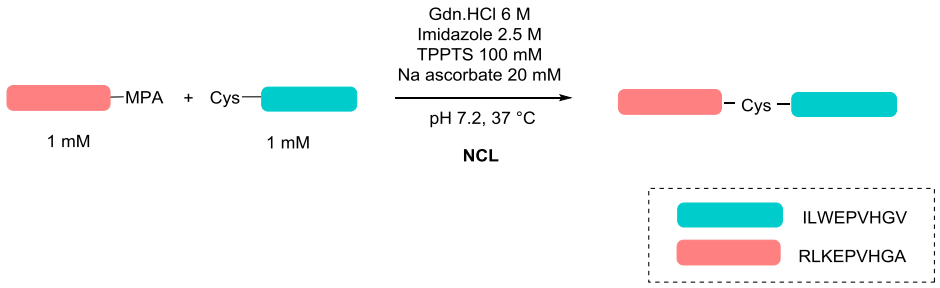


Figure 41. *Optimization of NCL between simple MPA peptide and Cys-peptide catalysed with imidazole.*

To obtain 1 mL of the solution, Gdn·HCl 6 M (573 mg), imidazole 2.5 M (238.25 mg), TPPTS 100 mM (56.8 mg) and sodium ascorbate 20 mM (3.72 mg) were placed into a volumetric flask of 1 mL. To solubilize the mixture of powders, distilled water (106 µL) and HCl 6 M (250 µL) were added progressively into the same flask. The right water/HCl ratio was set experimentally, by pH control during the solubilization. The solubilization was facilitated by vortexing and sonication. The final solution of pH is 7.2 was transferred into a recipient convenient for storage.

The obtained solution (330 µL) was inserted into Cys-peptide 1 (455.2 µg, 1 mM) and then into MPA peptide (473.98 µg, 1 mM). The reaction mixture was incubated at 37 °C with shaking 300 rpm for 20 h.

The process was monitored by HPLC (column C3, gradient 12-38 % “B” in 15 min) at wavelength 215 nm (**Figure 42**), and controlled by UPLC-MS (column C18, gradient 0-70 % “B” in 20 min). For each analysis, an aliquot of 2 µL was quenched in 100 µL of AcOH 10 %.

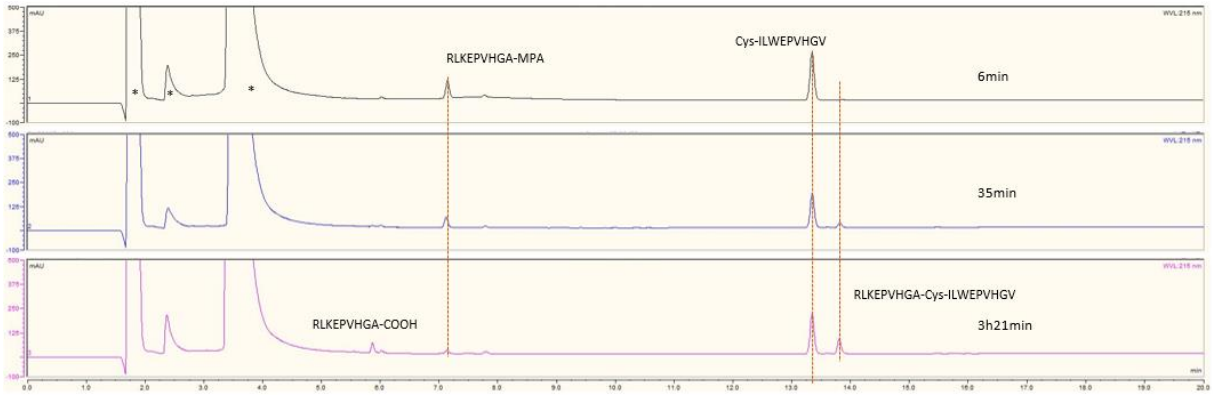


Figure 42. *HPLC profile of NCL in non-reducing media, with high MPA hydrolysis rate.*

The below protocol describes the most optimal conditions applied to attempt of using Sec-peptide instead of Cys-peptide, as an attempt to improve the NCL kinetics (**Figure 43**).

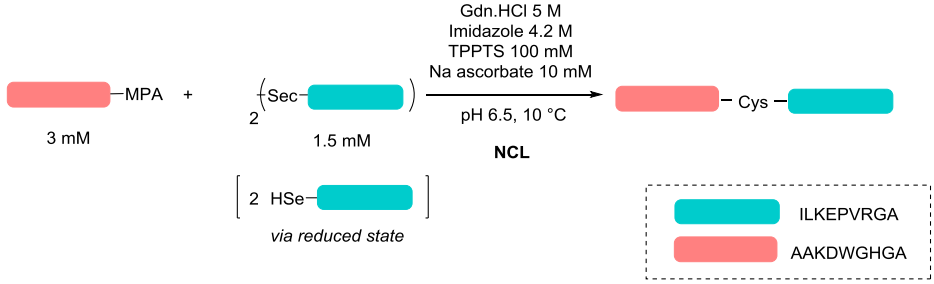


Figure 43*. Optimization of NCL between simple MPA peptide and Sec-peptide catalysed with imidazole.*

To obtain 1 mL of the solution, Gdn·HCl 5 M (477.5 mg), imidazole 3 M (285.9 mg), TPPTS 100 mM (56.8 mg) and sodium ascorbate 10 mM (1.86 mg) were placed into a volumetric flask of 1 mL. To solubilize the mixture of powders, distilled water (90 µL) and HCl 6 M (240 µL) were added progressively into the same flask. The right water/HCl ratio was set experimentally, by pH control during the solubilization. The solubilization was facilitated by vortexing and sonication. The final solution of pH is 6.50 was transferred into a recipient convenient for storage.

The prepared solution (350 µL) was added into Sec-peptide (obtained elsewhere) (1.5 mM, 1.426 mg) and the dissolved peptide was placed at 10 °C for 30 min to reduce diselenide with presenting TPPTS. This solution was transferred to MPA peptide (obtained elsewhere) (3 mM, 1.508 mg). The reaction mixture was incubated at 10 °C with shaking 300 rpm for 20 h.

The process was monitored by HPLC (column C3, gradient 12-38 % "B" in 15 min) at wavelength 215 nm (**Figure 44**), and controlled by UPLC-MS (column C18, gradient 0-70 % "B" in 20 min). For each analysis, an aliquot of 2 µL was quenched in 100 µL of AcOH 10 %.

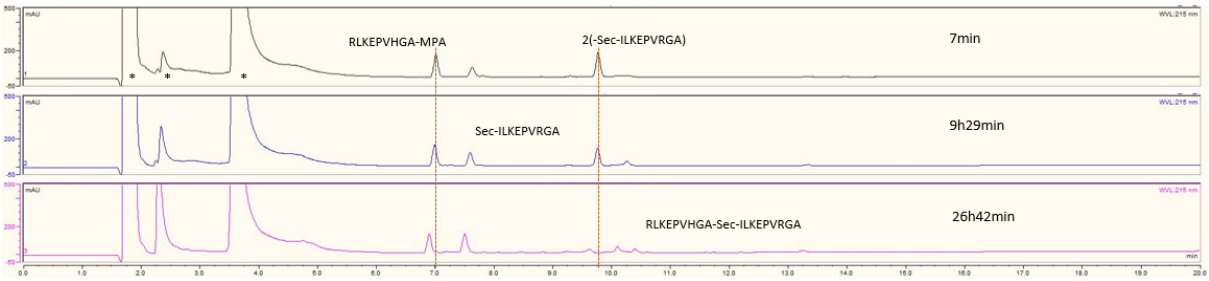


Figure 44. *HPLC profile of NCL with Sec-peptide instead of Cys-peptide. The chromatogram t point 26h42 min is shifted due to buffer system changing*

8.4 Co-catalysing NCL with base addition

To increase the NCL rate by increasing the nucleophilicity of the catalyst, a base (DABCO) was added. The experiment was performed with no base as negative control, 10 mM and 40 mM of base to find the optimal concentration.* The given below protocol is described for 40 mM DABCO (**Figure 45**).

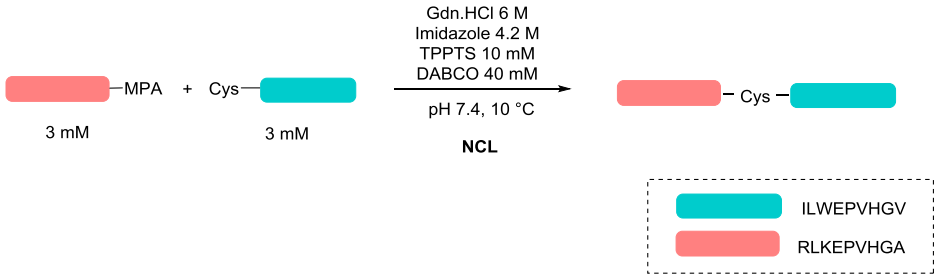


Figure 45. *Essay of NCL between simple MPA peptide and Sec-peptide catalysed with imidazole and co-catalysed with DABCO.*

To obtain 1 mL of the solution, Gdn·HCl 6 M (573 mg), imidazole 4.2 M (285.9 mg), TPPTS 10 m (5.68 mg) and DABCO 40 mM (4.49 mg) were placed into a volumetric flask of 1 mL. To solubilize the mixture of powders, distilled water (70 µL) and HCl 6 M (310 µL) were added progressively into the same flask. The right water/HCl ratio was set experimentally, by pH control during the solubilization. The solubilization was facilitated by vortexing and sonication. The final solution of pH is 7.40 was transferred into a recipient convenient for storage.

The prepared solution (350 µL) was added into Cys-peptide 1 (3 mM, 1.448 mg) and then transferred to MPA peptide 2 (3 mM, 1.508 mg). The reaction mixture was incubated at 10 °C with shaking 300 rpm for 20 h.

The process was monitored by HPLC (column C3, gradient 12-38 % “B” in 15 min) (**Figure 46**) and controlled by UPLC-MS (column C18, gradient 0-70 % “B” in 20 min). For each analysis, an aliquot of 2 µL was quenched in 100 µL of AcOH 10 %.

*For the negative control without DABCO, no base was added during the buffer preparation. For the 10 mM DABCO, solutions with no base and 40 mM base were mixed in 3:1 ratio.

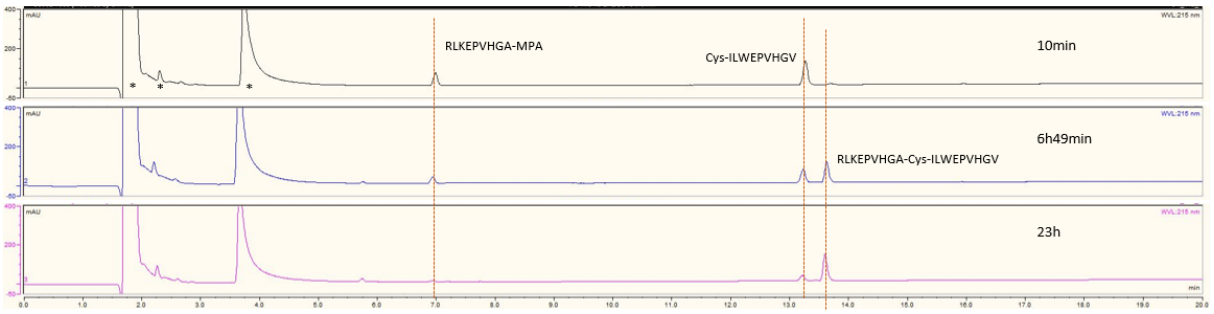


Figure 46. *HPLC profile of NCL in non-reducing media with added base.*

8.5 Non-classical imidazole catalysed NCL with simple MTG thioester devoid of SetCys

To increase the rate of NCL, MPA thioester was replaced with MTG. his strategy allowed to reduce the reaction time from 20 h to 2 h (10 times). The pH, concentration of phosphine and peptides, were re-optimized (**Figure 47**). Moreover, relying on published data, 2-methyl imidazole was also tested to catalyse the NCL, but it gave no kinetics improvement.

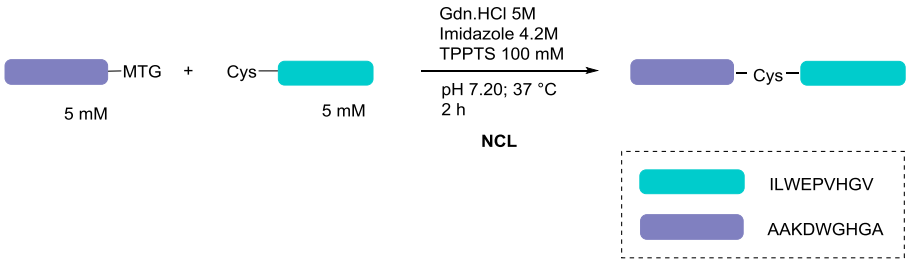


Figure 47. *Optimisation of NCL between simple MTG peptide and Cys-peptide catalysed with imidazole.*

To obtain 1 mL of the solution, Gdn·HCl 6 M (573 mg), imidazole 4.2 M (285.9 mg), TPPTS 100 mM (56.8 mg) were placed into a volumetric flask of 1 mL. To solubilize the mixture of powders, distilled water (73 µL) and HCl 6 M (320 µL) were added progressively into the same flask. The right water/HCl ratio was set experimentally, by pH control during the solubilization. The solubilization was facilitated by vortexing and sonication. The final solution of pH is 7.40 was transferred into a recipient convenient for storage.

The prepared solution (377.7 µL) was added into Cys-peptide 1 (5 mM, 2.6 mg) and then transferred to MPA peptide 2 (5 mM, 2.53 mg). The reaction mixture was incubated at 10 °C with shaking 300 rpm for 2 h.

The process was monitored by HPLC (column C3, gradient 12-38 % “B” in 15 min) at wavelengths 215 nm and 280 nm, and controlled by UPLC-MS (column C18, gradient 0-70 % “B” in 20 min). For each analysis, an aliquot of 2 µL was quenched in 100 µL of AcOH 10 %.

9 Innovative imidazole-based protocol

The non-reducing imidazole-based media combined with MTG thioester use was chosen to be applied for the innovative two-step one-pot process. Tested with bifunctional peptide, the protocol described in chapter 8.6 was the most beneficial in term of NCL rate and selectivity. In addition, highly concentrated final solution was improved by reasonable decrease of imidazole concentrations in order to avoid any precipitations during the reaction. The described below protocol was tested for NCL with bifunctional peptide prior to start the two-step process.

The solution containing all reagents except peptides is prepared in advance and may be stored for a while at -80 °C. As TPPTS is stable in aqueous solutions for many hours, no glovebox is needed during the buffer preparation.

The reactions were carried out using a hotplate stirrer equipped with a heating block. The reactions were monitored by HPLC at 215 nm, column C3. For each analysis, a 2 µL aliquot was taken from the reaction mixture and quenched by adding 10% AcOH in water (100 µL).

9.1 Preparation of stock solutions

Stock solution **A**. To obtain 2 mL of the solution, Gdn·HCl 5 M (955 mg), imidazole 3 M (408.45 mg) and TPPTS 100 mM (113.60 mg) were placed into a volumetric flask of 2 mL. To solubilize the mixture of powders, distilled water (150 µL) and HCl 6 M (640 µL) were added progressively into the same flask. The right water/HCl ratio was set experimentally, by pH control during the solubilization. The solubilization was facilitated by vortexing and sonication. The final solution of pH is 6.80 was transferred into a recipient convenient for storage.

Stock solution **B**. 200 µL of solution A was transferred to another recipient containing 198 µg of Na-ascorbate (5 mM).

Stock solution **C**. The final solution to be used for our chemistry contains Na-ascorbate 1 mM. Thus, 4 parts of solution A was mixed with 1 part of solution B (4 : 1), for example 400 µL A and 100 µL B. The pH is checked to remain 6.80. If it changed, it was corrected by adding NaOH or HCl 6 M. Prior to be used, the solution must be inserted in a glovebox to degaze for 20-30 min with opened Eppendorf flask lead.

9.2 Innovative two-step one pot process

Cys-peptide 1 (4.5 mM, 1.05 mg) was dissolved in 170.2 µL of the solution C which was incubated at 10 °C inside a glovebox for 20 min (**Figure 48**). Then, the mixture with Cys-peptide 1 was transferred into another Eppendorf vessel to dissolve the bifunctional peptide (for SetCys-AAKDWGHGA-MTG: 5 mM, 1.04 mg). The NCL time starts upon the bifunctional peptide is completely dissolved. After the addition of both peptides, the pH remains nearly constant (Δ 0.4) and doesn't have to be corrected. The reaction mixture was

incubated at 10 °C with shaking 300 rpm for 2.5 h. Then the shaking was stopped and the temperature rosen up to 55 °C for 10 h, to proceed SetCys→Cys conversion.

The two-steps process was monitored by HPLC (column C3, gradient 12-38 % "B" in 15 min) at wavelengths 215 nm and 280 nm (**Figure 49**), and controlled by UPLC-MS (column C18, gradient 0-70 % "B" in 20 min). For each analysis, an aliquot of 2 µL was quenched in 100 µL of AcOH 10 %. The HPLC profile with Ser-modified bifunctional peptide is presented in **Figure 50**.

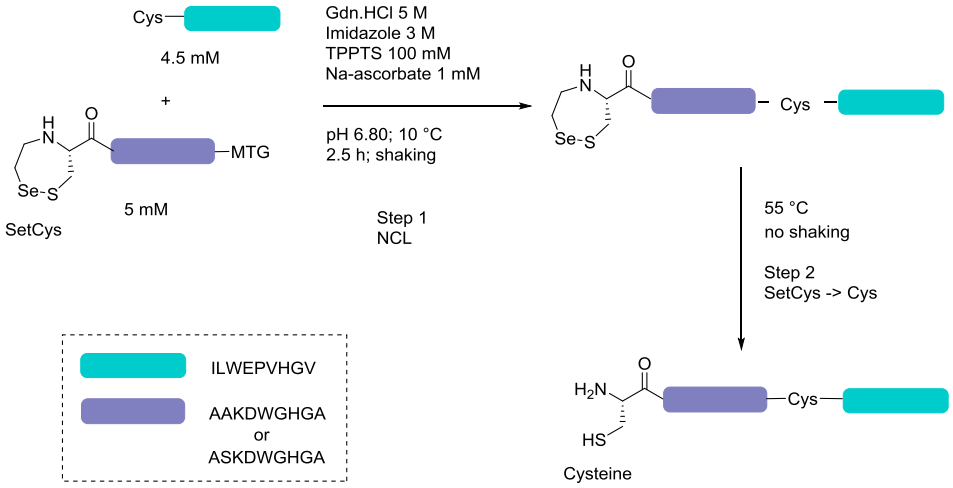


Figure 48. *Two step one-pot process describing NCL followed by cysteine deprotection.*

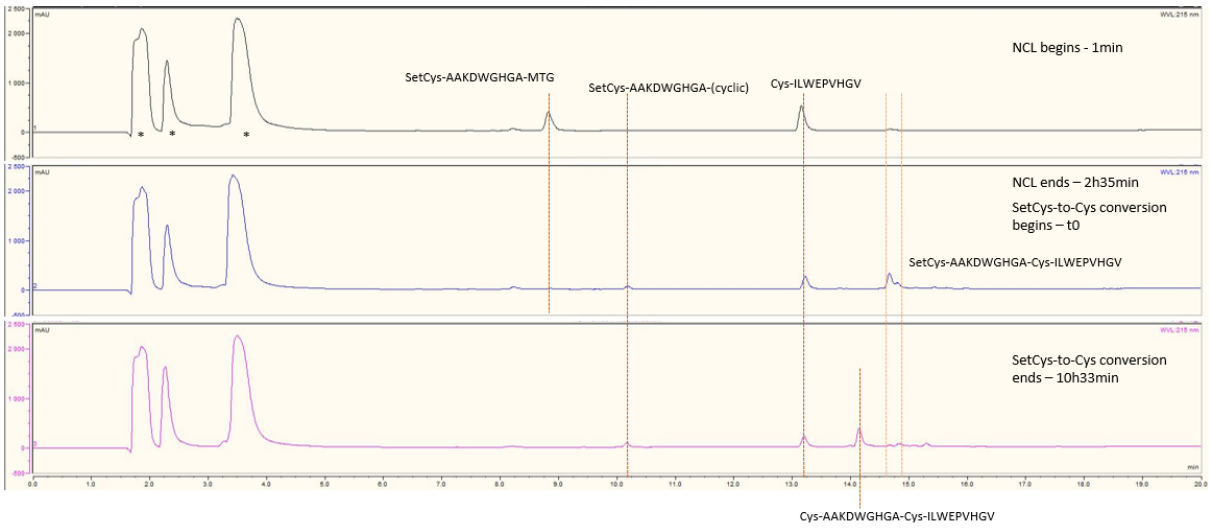


Figure 49. *HPLC profile of 2-step one pot process under optimized conditions, where Ala mutation was on bifunctional peptide.*

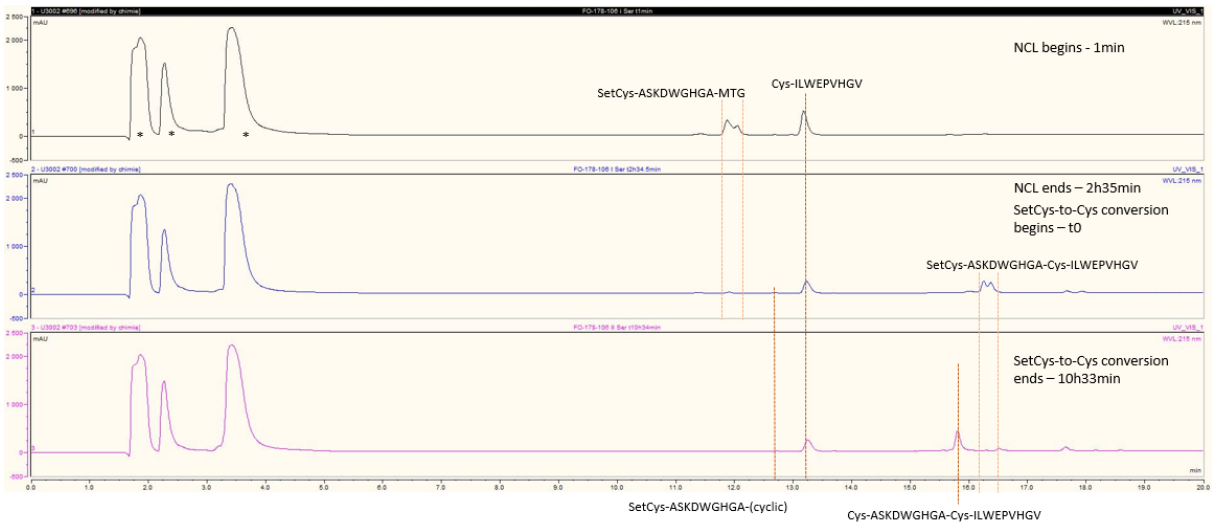


Figure 50. *HPLC profile of 2-step one pot process under optimized conditions, where Ser mutation was on bifunctional peptide.*

Isolated products

To prove the concept, the product was isolated for the 1st iteration that includes one pot process: ligation followed by SetCys to Cys conversion.

The final product 9a (Ala) was isolated by HPLC (column C18, gradient 22 – 38 % "B" in 70 min, the "B" = 30.17 %) resulting the titled peptide as white powder (130.5 mg, yield 9 %), described in **Figure 51**.

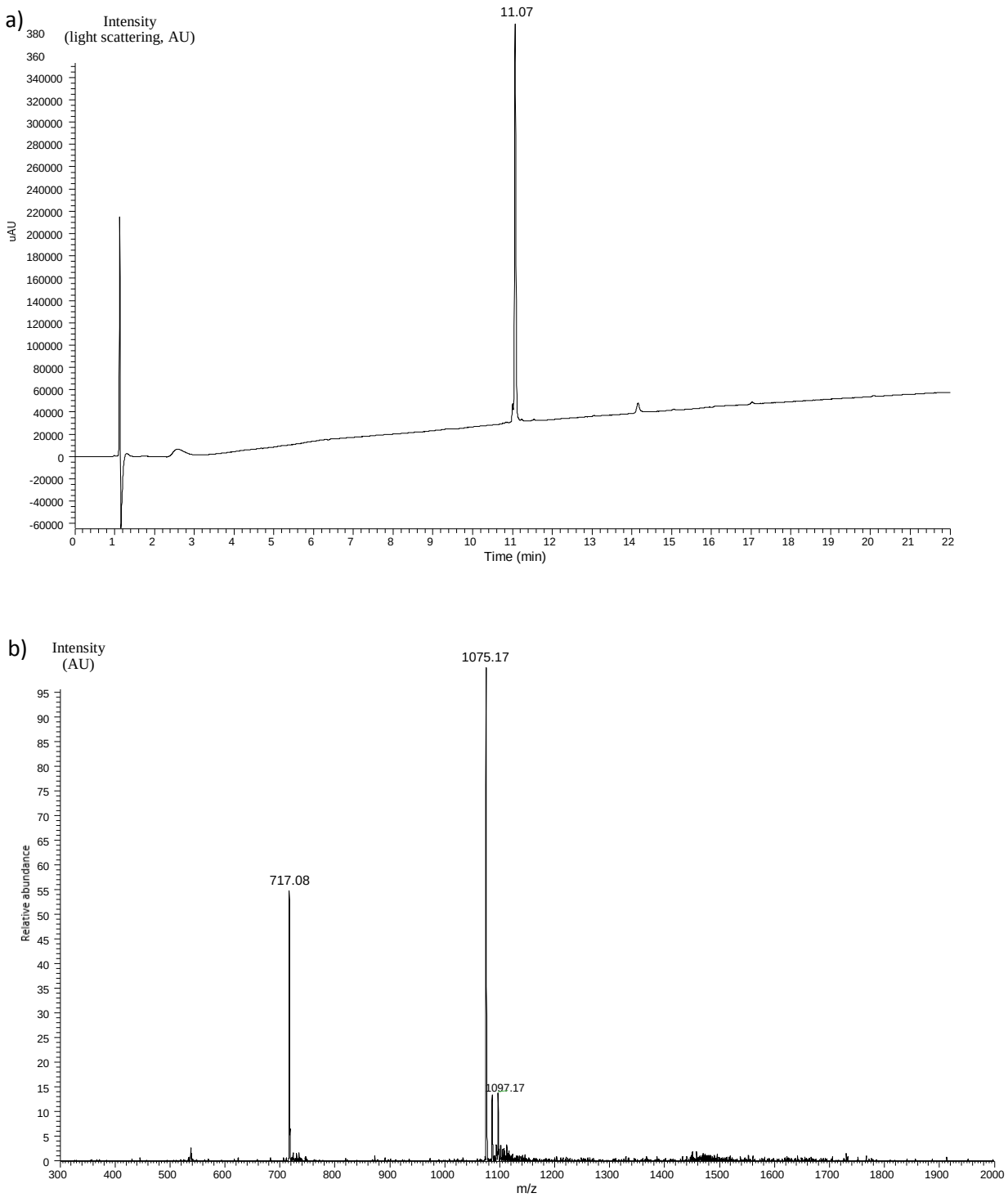


***Figure 51**. peptide **9a**. Analysis by UPLC-MS. a) UPLC on C18 column, 50 °C, gradient 0-70 % "B" in 20 min, 0.4 mL min^{-1}, light scattering detection; b) MS trace: m/z for $[M+2H]^{2+}$ 1075.17, $[M+3H]^{3+}$ 717.08, found for M (average) 2148.29. Calcd. for M 2147.78.*

The final product 9b (Ser) was obtained from 2 mg of each starting material and isolated by HPLC (column C18, gradient 18 – 38 % "B" in 70 min, the "B" = 32.9 %) resulting the titled peptide as white powder (1.04 mg, yield 33 %), described in **Figure 52**.

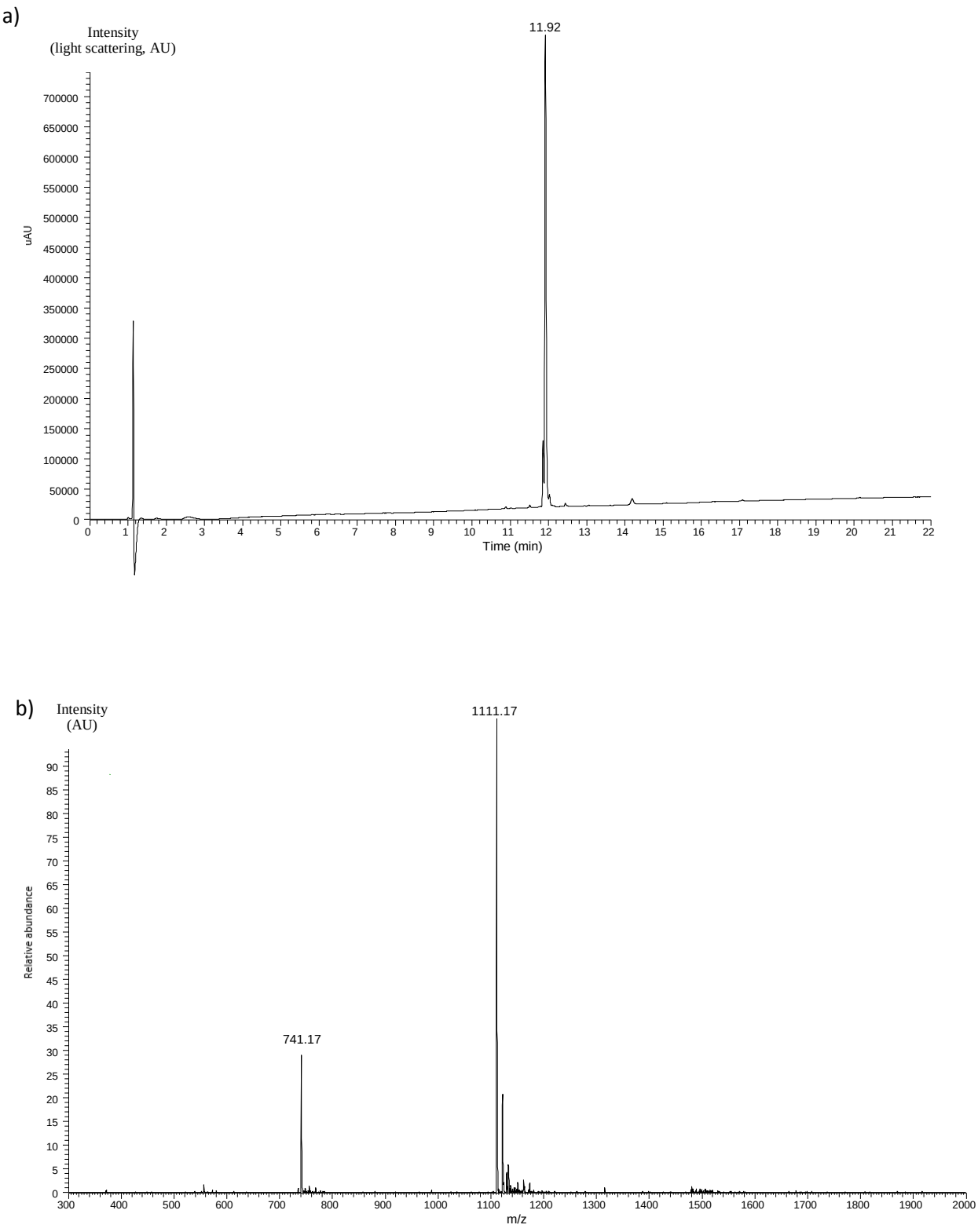


Figure 52. *Peptide* ***9b****. Analysis by UPLC-MS. a)UPLC on C18 column, 50 °C, gradient 0-70 % "B" in 20 min, 0.4 mL min^{-1}, light scattering detection; b) MS trace: m/z for [M+2H]$^{2+}$ 1111.17, [M+3H]$^{3+}$ 741.17, found for M (average) 2220.43. Calcd. for M 2217.54.*

9.3 Innovative two-step one pot process in PCR machine

A partial automation of the process was attempted in a PCR machine, under non-inert atmosphere. The idea was to benefit from the excellent heat control that such a system can achieve, and to perform two steps without interventions. The solution containing all reagents was degassed inside a glovebox, then removed and placed into PCR machine to thermostate at 10 °C for 20 min. Next, both reacting peptides were subsequently dissolved in the solution manually (with air contact) and placed back in PCR machine. The temperature 10 °C was hold for 2 h to complete NCL and then was increased automatically up to 55 °C to trigger SetCys conversion during 10 h. In total time 12 h the sample was analysed by UPLC-MS. It was demonstrated that the present oxygen made the reaction profile messy, so the protocol still demands improvements.

10 Antifreeze protein chemistry

10.1 One pot AFP peptide concatenation

The described below protocol was applied to all peptide mutants throughout the sequence optimization. The optimization steps consist of amino acid replacement (D→E) in all three peptide fragments and insertion of polyArg-tag (solubilizing) to N-terminus of N-terminal peptide fragment. Here the last peptides modification is demonstrated (**Figure 53**).

6 M Gdn·HCl (573 mg), *N*-octyl glucoside 20 mM (5.83 mg) were dissolved 0.1 M pH 7.2 sodium phosphate buffer resulting 1 mL of total volume. The solution was then degassed during 30 min. MPAA 200 mM (33.6 mg) was dissolved in this solution. NaOH (6 M) was then added to adjust the pH to 7.2. The MPA peptide 1 (5.07 mg) and the auto-protected SEA^{off} peptide 2 (2.95 mg) were dissolved in the above solution (346 µL) and stirred at 37 °C for 20h under nitrogen atmosphere.

After 20 h, the exchange of the SEA group by MPA was started by adding 346 µL of solution containing 6 M Gdn·HCl (573 mg), *N*-octyl glucoside 20 mM (5.83 mg), MPAA 200 mM (33.6 mg) and TCEP 200 mM (19.6 mg). The final TCEP concentration was 100 mM. The obtained solution was transferred into amided Cys-peptide 3 (2.644 mg). After 30 h, the reaction mixture was acidified with acetic acid (69 µL), extracted with Et_2O (4 x 1 mL) to remove MPAA and purified on preparative HPLC (detection at 215 nm, 6 mL/min, 0 to 12 % eluent B in 2 min, then 12 to 40 % eluent B in 30 min, C3 XBridge column, 70 °C) to give 4.25 mg of thioester peptide 4 (48 %).

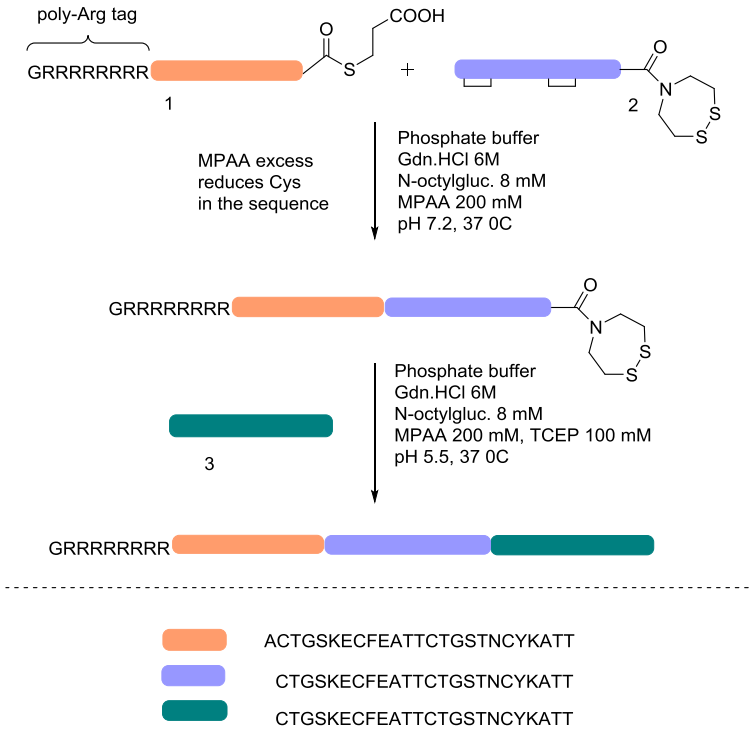
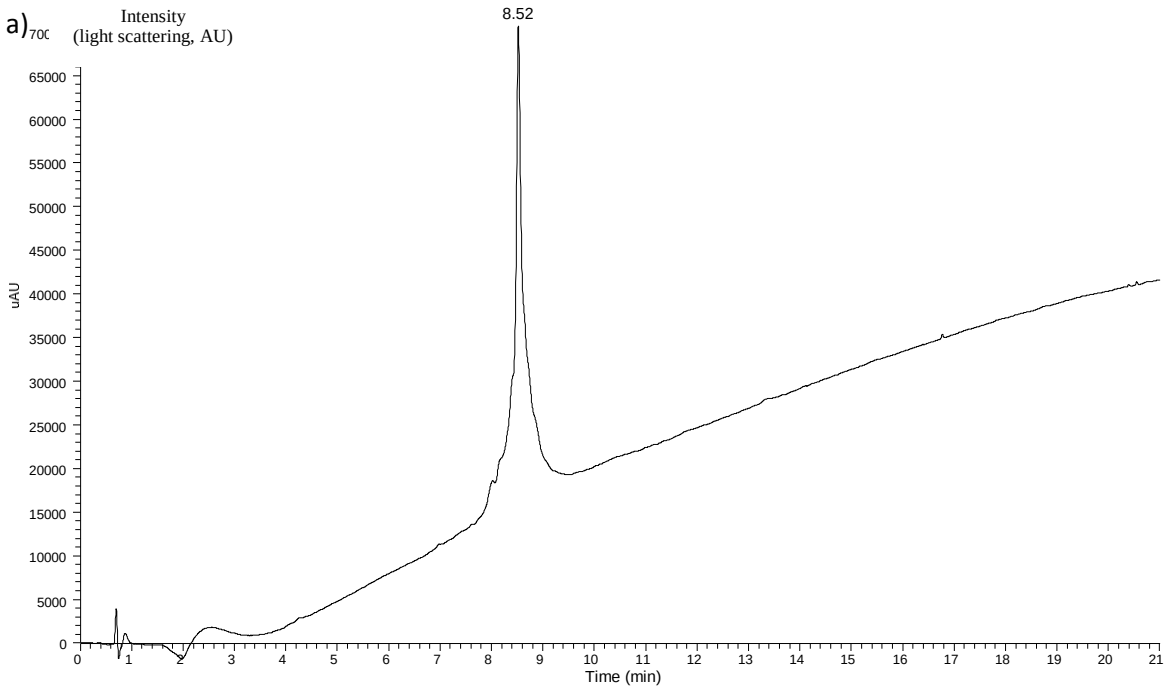


Figure 53. *One pot SEA ligation to obtain AFP polypeptide.*

The final product 10 was obtained and isolated by HPLC (column C3, gradient 12 – 40 % "B" in 30 min, the "B" = 20.3 %) resulting the titled peptide as white powder (4.25 mg, yield 48 %), described in **Figure 54**.



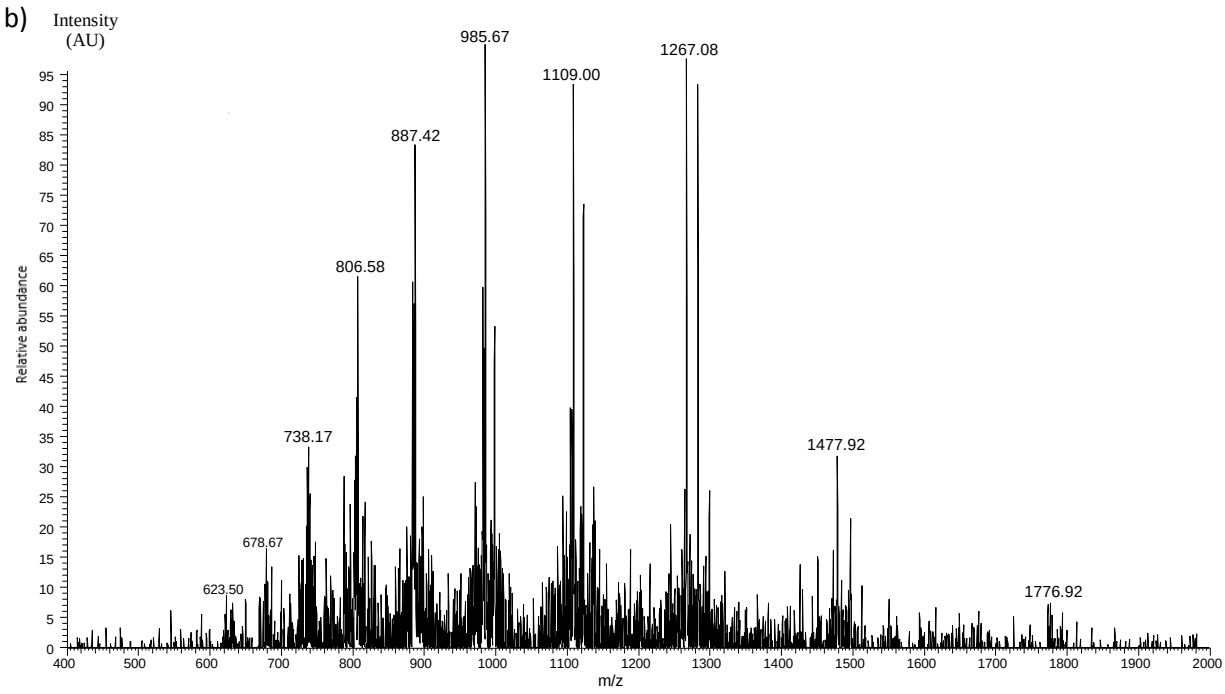


***Figure 54**. Peptide **10**. Analysis by UPLC-MS. a) UPLC on C3 column, 70 °C, gradient 0-70 % "B" in 20 min, 0.4 mL min-1, light scattering detection; b) MS trace: m/z for $[M+5H]^{5+}$ 1776.92, $[M+6H]^{6+}$ 1477.92, $[M+7H]^{7+}$ 1267.08, $[M+8H]^{8+}$ 1109.00, $[M+9H]^{9+}$ 985.67, $[M+10H]^{10+}$ 887.42, $[M+11H]^{11+}$ 738.17, $[M+12H]^{12+}$ 678.67, $[M+13H]^{13+}$ 623.50; found for M (average) 8863.88. Calcd. for M 8863.28.*

10.2 Folding of the AFP polypeptides

Stock solution of the obtained polypeptide 10 was prepared. For that, 1 mL of Gdn·HCl 6 M (573.18 mg) was prepared. 1.041 mg of the obtained polypeptide 4 (1.041 mg) was dissolved in 30 µL of Gdn·HCl 6 M. The final solution may be stored at -80 °C for a while.

10.3 Folding of the obtained polypeptide

To obtain a functional protein, the polypeptide 10 (**Figure 55**) undergone several trials of folding via disulfide bonds formation. The process did not function, perhaps due to remaining impurities after the purification step.

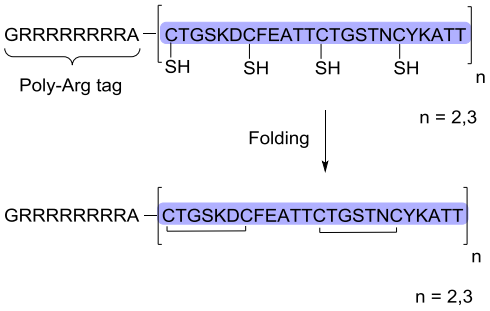


Figure 55. *Folding via disulfide bridges formation.*

For the folding solution, a peptide must be of concentration 1 mg/mL. Gdn·HCl is required at concentration of 0.2 M maximum. To fit this, for one folding experiment with 250 µg of the peptide, we take 7.9 µL of the stock solution.

The buffer (PBS 0.1 M containing 10 % by vol. of glycerol, named PBS/glycerol) was used to prepare the following solutions *. Stock solution A: GSH 100 mM (6 mg/200 µL); stock solution B: GSSG 3.265 mM (2 mg/mL). Those solutions, as well as polypeptide stock solutions, were thermostated at cold room (4 C°). To the PBS/glycerol (228.6 µL) in an Eppendorf recipient, the solution A (2.5 µL) and then solution B (15 µL) were added. The pH was checked to be 7.00. If necessary, it was corrected by adding NaOH 2 M. Then, the 7.9 µL of peptide stock solution were added and vortexed gently. The mixture in an Eppendorf was placed in a tube rotator for days at 4 C°.

*GSH and GSSG stock solutions must be freshly prepared and used straightaway.

References

(1) Sodium Phosphate Buffer (0.2 M). *Cold Spring Harb. Protoc.* **2010**, *2010* (1), pdb.rec12105. https://doi.org/10.1101/pdb.rec12105.

(2) Krief, A.; Dumont, W.; Delmotte, C. Reaction of Organic Selenocyanates with Hydroxides: The One-Pot Synthesis of Dialkyl Diselenides from Alkyl Bromides. *Angew. Chem. Int. Ed Engl.* **2000**, *39* (9), 1669–1672. https://doi.org/10.1002/(sici)1521-3773(20000502)39:9<1669::aid-anie1669>3.0.co;2-6.

(3) Abbas, M.; Bethke, J.; Wessjohann, L. A. One Pot Synthesis of Selenocysteine Containing Peptoid Libraries by Ugi Multicomponent Reactions in Water. *Chem. Commun.* **2006**, No. 5, 541–543. https://doi.org/10.1039/B514597J.

(4) Zawahir, Z.; Neamati, N. Inhibition of HIV-1 Integrase Activity by Synthetic Peptides Derived from the HIV-1 HXB2 Pol Region of the Viral Genome. *Bioorg. Med. Chem. Lett.* **2006**, *16* (19), 5199–5202. https://doi.org/10.1016/j.bmcl.2006.07.022.

(5) Ollivier, N.; Dheur, J.; Mhidia, R.; Blanpain, A.; Melnyk, O. Bis(2-Sulfanylethyl)Amino Native Peptide Ligation. *Org. Lett.* **2010**, *12* (22), 5238–5241. https://doi.org/10.1021/ol102273u.

(6) Ollivier, N.; Vicogne, J.; Vallin, A.; Drobecq, H.; Desmet, R.; El Mahdi, O.; Leclercq, B.; Goormachtigh, G.; Fafeur, V.; Melnyk, O. A One-Pot Three-Segment Ligation Strategy for Protein Chemical Synthesis. *Angew. Chem. Int. Ed.* **2012**, *51* (1), 209–213. https://doi.org/10.1002/anie.201105837.

(7) Ollivier, N.; Raibaut, L.; Blanpain, A.; Desmet, R.; Dheur, J.; Mhidia, R.; Boll, E.; Drobecq, H.; Pira, S. L.; Melnyk, O. Tidbits for the Synthesis of Bis(2-Sulfanylethyl)Amido (SEA) Polystyrene Resin, SEA Peptides and Peptide Thioesters. *J. Pept. Sci.* **2014**, *20* (2), 92–97. https://doi.org/10.1002/psc.2580.

Annex D

A Selenium-Based Cysteine Surrogate for Protein Chemical Synthesis

Firstova, O.; Agouridas, V.; Diemer, V.; Melnyk, O.

The ChemRxiv version is presented here,

10.26434/chemrxiv-2021-5nn16

The peer-reviewed version is available

Methods Mol. Biol. Clifton NJ 2022, 2530, 213–239

DOI 10.1007/978-1-0716-2489-0_15

Olga Firstova

Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 9017 - CIIL - Center for Infection and Immunity of Lille, F-59000 Lille, France**.**

Vangelis Agouridas

Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 9017 - CIIL - Center for Infection and Immunity of Lille, F-59000 Lille, France.

Centrale Lille, F-59000 Lille, France.

Vincent Diemer

Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 9017 - CIIL - Center for Infection and Immunity of Lille, F-59000 Lille, France.

Oleg Melnyk

Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 9017 - CIIL - Center for Infection and Immunity of Lille, F-59000 Lille, France.

Corresponding authors:

Oleg Melnyk

Email address: oleg.melnyk@ibl.cnrs.fr

Vincent Diemer

Email address: vincent.diemer@ibl.cnrs.fr

A selenium-based cysteine surrogate for protein chemical synthesis

Olga Firstova, Vangelis Agouridas, Vincent Diemer*, Oleg Melnyk*

We provide in this protocol detailed procedures for the synthesis of SetCys cysteine surrogate and its use for chemical synthesis of proteins through the redox-controlled assembly of three peptide segments in one-pot.

Keywords

N-(2-selenoethyl)cysteine, cysteine surrogate, native chemical ligation (NCL), *bis*(2-sulfanylethyl)amido (SEA)-mediated ligation, chemical protein synthesis, sulfur, selenium, one-pot.

1. Introduction

Modern chemical protein synthesis is usually achieved by the concatenation of unprotected peptide segments through the application of chemoselective peptide bond-forming reactions. Considered as a figurehead in the field, the native chemical ligation (NCL) reaction involves the coupling of a peptide thioester with a cysteinyl (Cys) peptide in aqueous media under neutral conditions. [1,2,3,4] Many target proteins exceed 100 amino acids and are thus too long to be accessed by the coupling of just two peptide segments, if the latter have to be produced by solid phase peptide synthesis (SPPS). [5] The access to challenging protein targets usually requires the resort to two or more NCL reactions, [6] and thus potentially to set up a large number of chemical and isolation steps. In this context and logically, the field of chemical protein synthesis has encountered a growing interest for economy-oriented synthetic approaches [7], which are sustainable as well as time and cost effective by increasing the yield of isolated material through the reduction of intermediate purification steps. To achieve such an objective, solid-phase protein assembly techniques were developed in which elongated polypeptides are simply separated at each step of the process from the reagents and by-products by washing of the solid phase [6,8,9]. An attractive alternative to access large proteins is to use one-pot multi-segment concatenation strategies [6,10]. However, the application of such strategies assumes to have temporarily silent and activatable thioester and/or cysteine surrogates at disposal, so as to trigger ligations in a specific order and to ensure that peptide segments get connected in the proper way.

The *bis*(2-sulfanylethyl)amido (SEA)-mediated ligation is an NCL-related process involving a peptide equipped with a latent thioester moiety in the form of a cyclic disulfide (SEA^{off} peptide

in Figure 1A,B) [11,12].While the SEA^{off} group is unaffected under weak reducing conditions (Figure 1A), the addition of a strong reductant such as *tris*(2-carboxyethyl)phosphine (TCEP) in the *milieu* triggers its ring-opening by reduction of the disulfide bond and its reaction with Cys peptides to yield a ligated product (Figure 1B). Recently, we have developed a cysteine surrogate called SetCys whose reactivity is also dependent on the reducing power of the medium (Figure 1C,D) [13]. The SetCys amino acid is a cyclic selenosulfide analogue of cysteine, which under mild reducing conditions does not interfere with a classical NCL reaction (Figure 1C). In contrast, subjecting SetCys to strong reducing conditions results in the spontaneous loss of its *N*-selenoethyl appendage and thus to its conversion into an N-terminal Cys residue (Figure 1D). If a peptide thioester segment is present in the reacting mixture, the conversion of SetCys to Cys followed by a classical NCL leads to the production of a native polypeptide of larger size (Figure 1D).

[Insert Figure 1 near here]

Figure 1 clearly shows the parallel between SEA^{off} and SetCys regarding their behavior toward the reducing power of the medium. This protocol article illustrates how the redox-controllable SEA^{off} and SetCys groups can be exploited hand-in-hand for the synthesis of proteins through a one-pot three-segment assembly process (Figure 2). The synthesis involves SEA^{off} peptide **1**, the bifunctional peptide **2** featuring a SetCys residue at its N-terminus and a C-terminal alkyl thioester, and finally a Cys peptide **3**. By virtue of the properties presented for each group in Figure 1, the elongation of the polypeptide chain takes place in the C-to-N direction. Indeed, since weakly reducing conditions are first applied, the SEA^{off} and the SetCys groups remain silent. Therefore, partners **2** and **3** first react by classical NCL to form the elongated intermediate **4**. Addition of TCEP and sodium ascorbate (see later) in the reaction mixture and a slight pH-adjustment provide the conditions to activate the SEA^{off} and SetCys groups and thus, an event that triggers the coupling of intermediate polypeptide **4** to SEA peptide **1** with the concomitant conversion of SetCys into Cys. The simultaneous presence of three peptide segments in the reaction mixture from the beginning of the one-pot process is an hallmark of this approach, which is illustrated by the chemical synthesis of a biotinylated analogue of hepatocyte growth factor (HGF) K1 domain (Figure 2).

[Insert Figure 2 near here]

In addition to the assembly of the K1 domain, the detailed synthesis of a conveniently Fmoc-protected SetCys amino acid **6** for solid phase peptide synthesis as well as its incorporation into bifunctional peptides are described hereafter.

2. Materials

2.1 Health and safety recommendations

All organic solvents and reagents used in this protocol should be handled inside a chemical fume hood with appropriate personal protective equipment (lab coat, gloves and protective glasses). Solvents were removed using a rotary evaporator placed inside a fume hood. Particular care must be taken for the manipulation of selenium containing reagents and synthetic intermediates due to their potential toxicity. Trifluoroacetic acid is strongly corrosive and toxic. Therefore, it should be handled with the greatest attention.

2.2 General laboratory equipments, solvents and reagents

2.2.1 General materials for organic synthesis and peptide chemistry

1. Conventional laboratory glassware.
2. Microfuge tubes (0.5, 1.5 mL and 5 mL safe-lock tubes; Eppendorf), plastic tubes (15 mL and 50 mL) for centrifugation.
3. Glove box (Jacomex) equipped with a block heater (37°C), a set of adjustable pipettes (0.5–10, 2–20, 10–50, 10–100, 100–1,000 µL, Eppendorf), corresponding pipette tips, a pH meter (Eutech instrument) and a vortex shaker.
4. Argon gas (O_2 < 0.01 ppm, H_2O < 0.02 ppm, Air Liquide).
5. Rotatory evaporator.
6. TLC UV Cabinet (254 nm).
7. TLC plates (Silica gel 60 F_{254} on aluminium sheets).
8. Deionized water (≥ 18 MΩ.cm).
9. High purity grade solvents for classical organic synthesis and peptide chemistry: acetonitrile, chloroform ($CHCl_3$), cyclohexane (Cy), 1,2-dichloroethane (DCE), dichloromethane (DCM), diethyl ether (Et_2O), *N*,*N*-dimethylformamide (DMF) for peptide synthesis, dioxane, ethanol (EtOH), ethyl acetate (EtOAc), *n*-heptanes, *n*-hexanes, methanol (MeOH).
10. *N*,*N*-diisopropylethylamine for peptide synthesis (DIEA).
11. Guanidine hydrochloride (Gn·HCl).

12. High purity grade reagents for classical organic synthesis and peptide chemistry: sodium bicarbonate ($NaHCO_3$), sodium chloride (NaCl), potassium carbonate (K_2CO_3), hydrochloric acid (HCl, 37%), sodium hydroxide (NaOH), anhydrous magnesium sulfate ($MgSO_4$).
13. Silica gel (60-200 µm, 60 Å).

2.2.2 HPLC analysis & purification

2.2.2.1. Analytical UPLC-MS analysis

1. Analytical UPLC–MS analyses were performed on Dionex Ultimate 3000 UHPLC system (Thermofisher) equipped with an Acquity UPLC Peptide BEH300 C18 reverse phase-column (2.1 × 100 mm, pore size: 300 Å, particle size: 1.7 µm), a diode array detector, a charged aerosol detector (CAD) and a mass spectrometer (Ion trap LCQfleet).
2. MS-grade trifluoroacetic acid.
3. LC-MS grade acetonitrile (ACN).
4. Eluant A for UPLC-MS analysis: TFA 0.1% (vol/vol) in deionized water. Add 2.5 mL of TFA in 2.5 L of deionized water.
5. Eluant B for UPLC-MS analysis: TFA 0.1% (vol/vol) in LC-MS grade ACN. Add 2.5 mL of TFA in 2.5 L of ACN.
6. Standard method for analyzing peptides by UPLC-MS. Gradient: 0-70% eluent B in 20 min, flow rate: 0.4 mL/min, temperature: 50 °C, UV detection at 215 nm.

2.2.2.2. Preparative HPLC

1. Preparative HPLC were performed on a PLC2020 Gilson system equipped with a C18 reverse-phase column.
   - for small quantities of crude peptide (< 20 mg): Waters XBridge BEH300 C18 reverse-phase column (10 × 250 mm; pore size: 300 Å; particle size: 5 µm).
   - for large quantities of crude peptide (> 20 mg): Waters XBridge Prep C18 reverse-phase column (19 × 150 mm; pore size: 130 Å, particle size: 5 µm).
2. UV-grade trifluoroacetic acid.
3. HPLC-grade ACN
4. TFA 10% (vol/vol) in deionized water. Add 100 mL of TFA carefully to 900 mL of deionized water.

5. Eluent A for semi-preparative HPLC. TFA 0.1% (vol/vol) in deionized water. Add 10 mL of TFA 10% (vol/vol) in water to 990 mL of deionized water.
6. Eluent B for semi-preparative HPLC. TFA 0.1% (vol/vol) in ACN. Add 10 mL of 10% aqueous TFA (vol/vol) to 990 mL of ACN.

2.2.3 Lyophilization

1. Lyophilizer (Christ Gamma 2-20) equipped with a manifold.
2. Lyophilizer flasks (VWR, 300 mL).
3. Liquid nitrogen.
4. Dewar.

2.2.4 MALDI-TOF analysis

1. MALDI-TOF mass spectrometer (Autoflex Speed, Bruker).
2. α-Cyano-4-hydroxycinnamic acid.
3. 2,5-Dihydroxybenzoic acid.
4. A solution of α-cyano-4-hydroxycinnamic acid (10 mg/mL) in 50% aqueous ACN containing 0.1% TFA.
5. A solution of 2,5-dihydroxybenzoic acid (16 mg/mL) in 50% aqueous ACN containing 0.1% TFA.
6. To perform MALDI-TOF analysis, mix 1 µL of α-cyano-4-hydroxycinnamic acid solution or 2,5-dihydroxybenzoic acid solution with 1 µL of the sample on the MALDI plate. Let dry at room temperature in air prior to analysis.

2.2.5 NMR analysis

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

2.2.5 Polarimetric analysis

1. Optical rotations were recorded at 20 °C on a Perkin-Elmer 343 digital polarimeter at 589 nm with a cell path length of 10 cm.

2.2.6. HRMS analysis

1. For HRMS analyses, this work has benefited from the facility and expertise of "Service HPLC-MASSE" (Institut de Chimie des Substances Naturelles, Centre de Recherche de Gif sur Yvette, CNRS, 91198, Gif sur Yvette, France).

2.3 Synthesis of the Fmoc-SetCys-OH amino acid 6

2.3.1 Synthesis of *bis*(2,2-diethoxyethyl)diselenide 8

1. Potassium selenocyanate (KSeCN).
2. 2-bromo-1,1-diethoxyethane (**7** in Figure 3).

2.3.2 Synthesis of (2,2-diethoxyethyl)(4-methoxybenzyl)selenide 9

1. Sodium borohydride ($NaBH_4$).
2. 4-methoxybenzyl chloride (PMBCl).

2.3.3 Synthesis of 2-((4-methoxybenzyl)selenyl)acetaldehyde 10

1. Formic acid. A 1 M aqueous solution is prepared by adding formic acid (0.37 mL, 10 mmol) in a 10 mL volumetric flask and completing with deionized water. Mix the flask thoroughly before use.

2.3.4 Synthesis of H-Cys(Trt)-OMe 12

1. Cysteine methyl ester hydrochloride (**11** in Figure 3).
2. Trifluoroacetic acid (TFA).
3. Triphenylmethanol.
4. The 0.25 M aqueous solution of K_2CO_3 is prepared by dissolving K_2CO_3 (3.45 g, 25.0 mmol) in deionized water (100 mL).

2.3.5 Synthesis of *N*-[PMBSe-$(CH_2)_2$]-Cys(Trt)-OMe 13

1. Powdered 3 Å molecular sieves.
2. Sodium triacetoxyborohydride ($NaBH(OAc)_3$).
3. The 0.5 M aqueous solution of K_2CO_3 is prepared by dissolving K_2CO_3 (6.90 g, 50.0 mmol) in water (100 mL).

2.3.6 Synthesis of *N*-Fmoc-*N*-[PMBSe-$(CH_2)_2$]-Cys(Trt)-OMe 14

1. 9-Fluorenylmethoxycarbonyl chloride (Fmoc-Cl).

2.3.7 Synthesis of Fmoc-SetCys-OMe 15

1. Iodine (I_2).
2. Sodium sulfate pentahydrate ($Na_2S_2O_3 \cdot 5\ H_2O$). An aqueous solution of $Na_2S_2O_3$ is prepared by dissolving $Na_2S_2O_3 \cdot 5\ H_2O$ (24.8 g, 100 mmol) in water (100 mL).

2.3.8 Synthesis of Fmoc-SetCys-OH 6

1. ~5 M HCl. The solution is prepared by adding HCl (37%, 8.3 mL) to deionized water (11.7 mL) in a graduated cylinder.

2.4 Synthesis of the bifunctional peptide 2 (SetCys-HGF 150-176-MPA)

1. SEA-PS (polystyrene) solid support, loading: 0.16 mmol/g (prepared according to ref [11]).
2. 1-Hydroxybenzotriazole hydrate (HOBt).
3. *N*,*N*'-Diisopropylcarbodiimide (DIC).
4. Acetic anhydride (Ac_2O).
5. Piperidine.
6. Capping mixture (SPPS): 10 % Ac_2O, 5 % DIEA in DMF (vol/vol/vol). Add 2 mL of Ac_2O and 1 mL of DIEA to 17 mL of DMF.
7. Piperidine solution for Fmoc removal (SPPS): 20% piperidine in DMF (vol/vol). Add 4 mL of piperidine in 16 mL of DMF.
8. Triisopropylsilane (TIS).
9. Thioanisole.
10. Thiophenol.
11. 3-Mercaptopropionic acid (MPA).

2.5 One-pot C-to-N three peptide segment assembly

1. 4-Mercaptophenylacetic acid (MPAA).
2. Disodium hydrogen phosphate dodecahydrate ($Na_2HPO_4 \cdot 12H_2O$).
3. Sodium dihydrogen phosphate dihydrate ($NaH_2PO_4 \cdot 2H_2O$).
4. 0.1 M sodium phosphate buffer pH 7.2. Prepare 36 mL of a 0.2 M solution of disodium hydrogen phosphate in deionized water. Prepare 10 mL of a 0.2 M solution of sodium

dihydrogen phosphate in deionized water. Mix the two solutions and verify the pH. Dilute the prepared sodium phosphate buffer two-fold prior use.

5. *Tris*(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl).
6. Sodium ascorbate.
7. Glacial acetic acid (AcOH).

3 Methods

2.1 Synthesis of the Fmoc-SetCys-OH amino acid 6

The synthetic strategy proposed for the preparation of the Fmoc-protected SetCys amino acid **6** involves a convergent approach with a total of five steps from the affordable and commercially available cysteine methyl ester hydrochloride **11** (Figure 3). After the protection of Cys thiol side chain, intermediate **12** undergoes a reductive alkylation to install the selenated appendage on the α-amino group. Amine protection by a Fmoc group is followed by a tandem oxidative deprotection/cyclization process to afford the cyclic selenosulfide **15**. Final liberation of the acid moiety is achieved by HCl-mediated hydrolysis to provide Fmoc-SetCys-OH **6** with an overall yield of 34% (*see* **Notes §1** and **2**).

[Insert Figure 3 near here]

3.1.1 Synthesis of *bis*(2,2-diethoxyethyl)diselenide 8

Diselenide **8** was prepared by a procedure similar to that reported by Krief *et al.* [14]

1. In a two-neck flask equipped with a septum and a reflux condenser and maintained under argon atmosphere, dissolve potassium selenocyanate (576 mg, 4.00 mmol) in DMF (10 mL). The resulting solution turns to a clear yellowish color.
2. Place the flask in an oil bath and heat the mixture at 75 °C (bath temperature). Introduce dropwise by syringe a solution of 2-bromo-1,1-diethoxyethane **7** (4.00 mmol, 601 µL) in DMF (4 mL). Stir the reaction mixture at 75 °C for 3 h. The solution becomes brownish.
3. Add dropwise by syringe a solution of K_2CO_3 (552 mg, 4.00 mmol) in deionized water (1.6 mL). Keep stirring the reaction mixture for 3 h at 75 °C.
4. *Workup*. Remove the oil bath, add at RT 20 mL of deionized water and transfer the mixture to a separatory funnel. Extract the aqueous layer with Et_2O (2 × 20 mL) (*see* **Note §3**), wash the combined organic layers with water (2 × 20 mL) and dry over

anhydrous $MgSO_4$. Filter and evaporate the solvent to dryness under reduced pressure in a rotatory evaporator. Purify the crude product by column chromatography on silica gel (hexane/EtOAc 95:5) to obtain diselenide **8** as a yellow oil (485 mg, 62%).

5. Check the identity and purity of the synthesized compound by NMR. **^{1}H NMR** (300 MHz, $CDCl_3$, 298 K) δ 4.70 (t, *J* = 5.7 Hz, 2H), 3.63–3.73 (m, 4H), 3.52–3.61 (m 4H), 3.22 (d, *J* = 5.7 Hz, 4H), 1.22 (t, *J* = 7.1 Hz, 12H).

3.1.2 Synthesis of (2,2-diethoxyethyl)(4-methoxybenzyl)selenide 9

Selenide **9** was prepared by a procedure similar to that reported by Abbas *et al.* [15]

1. In round-bottom flask maintained under an argon atmosphere, dissolve diselenide **8** (900 mg, 2.29 mmol) in absolute ethanol (14 mL).
2. Place the flask in an ice bath and add $NaBH_4$ (175 mg, 4.62 mmol) portionwise under a stream of argon gas (*see* **Note §4**). Remove the ice bath and allow the mixture to stir at RT for 30 min. A white precipitate initially forms and redissolves resulting in a homogeneous solution.
3. Place again the flask in an ice bath and add 4-methoxybenzyl chloride (561 µL, 4.14 mmol) dropwise. A white precipitate forms. Remove the ice bath and stir the reaction mixture for 3 h at RT.
4. *Workup.* Filter the crude suspension through a celite pad in order to remove the white precipitate. Concentrate the clear yellowish filtrate under reduced pressure in a rotatory evaporator and add 30 mL of water. Transfer the aqueous mixture to a separatory funnel and extract it with Et_2O (2 × 25 mL). Dry the combined organic layers over $MgSO_4$, filter and evaporate the solvent under reduced pressure in a rotatory evaporator. Purify the crude product by silica gel column chromatography (hexanes/EtOAc 95:5) to obtain selenide **9** as a yellow oil (1.19 mg, 90%).
5. Check the identity and purity of the synthesized compound by ^{1}H NMR. **^{1}H NMR** (300 MHz, $CDCl_3$, 298 K) δ 7.23 (d, *J* = 8.6 Hz, 2H), 6.82 (d, *J* = 8.6 Hz, 2H), 4.59 (t, *J* = 5.6 Hz, 1H), 3.82 (s, 2H), 3.79 (s, 3H), 3.59–3.70 (m, 2H), 3.46–3.56 (m, 2H), 2.62 (d, J = 5.6 Hz, 2H), 1.22 (t, *J* = 7.1 Hz, 6H).

3.1.3 Synthesis of 2-((4-methoxybenzyl)selenyl)acetaldehyde 10

Aldehyde **10** was prepared by a procedure similar to that reported by Abbas *et al.* [15]

1. In a round-bottom flask equipped with a reflux condenser, suspend acetal **9** (1.78 g, 5.61 mmol) in 1 M aqueous formic acid solution (24 mL).

2. Place the flask in an oil bath and heat the mixture overnight at 50 °C (oil bath temperature).
3. *Workup*. Remove the oil bath and add water (55 mL) at RT. Transfer the aqueous mixture to a separatory funnel and extract it with Et_2O (2 × 55 mL). Wash the combined organic layers with water (40 mL) and dry the organic solution over $MgSO_4$. Filter and evaporate the solvent under reduced pressure in a rotatory evaporator. Aldehyde **10** (838 mg, 61%) is directly used in the next step without further purifications (*see* **Note §5**).
4. Check the identity and purity of the synthesized compound by ^{1}H NMR. **^{1}H NMR** (300 MHz, $CDCl_3$, 298 K) δ 9.36 (t, *J* = 4.2 Hz, 1H), 7.22 (d, *J* = 8.7 Hz, 2H), 6.84 (d, *J* = 8.7 Hz, 2H), 3.79 (s, 3H), 3.64 (s, 2H), 3.12 (d, *J* = 4.1 Hz, 2H).

3.1.4 Synthesis of H-Cys(Trt)-OMe 12

1. In a round-bottom flask, dissolve under argon cysteine methyl ester hydrochloride **11** (856 mg, 4.99 mmol) in TFA (5 mL).
2. Add triphenylmethanol (1.43 g, 5.49 mmol) and stir the mixture at RT for 5 h. The solution turns brown upon addition of triphenylmethanol.
3. Evaporate the TFA under reduced pressure in a rotatory evaporator and take up the residue in MeOH (27 mL). Keep on stirring at RT until the yellow color disappears.
4. Evaporate MeOH under reduced pressure in a rotatory evaporator and suspend the residue in a 0.25 M K_2CO_3 aqueous solution (30 mL) (*see* **Notes §6** and **7**).
5. *Workup*. Transfer the aqueous mixture in a separatory funnel and extract it with Et_2O (2 × 20 mL). Dry the combined organic layers over $MgSO_4$, filter and evaporate the solvent under reduced pressure in a rotatory evaporator. Purify the crude product by silica gel column chromatography (DCM/MeOH 98:2) to obtain the trityl-protected cysteine **12** (1.73 g, 92%) as a viscous oil.
6. Check the identity and purity of the synthesized compound by NMR. **^{1}H NMR** (300 MHz, $CDCl_3$, 298 K) δ 7.41-7.44 (m, 6H), 7.18-7.31 (m, 9H), 3.65 (s, 3H), 3.20 (dd, *J* = 4.8 and 7.8 Hz, 1H), 2.59 (dd, *J* = 4.8 and 12.4 Hz, 1H), 2.46 (d, *J* = 7.8 and 12.4 Hz, 1H) ppm. **^{13}C NMR** (75 MHz, $CDCl_3$, 298 K) δ 174.3 (C), 144.6 (3 × CH), 129.7 (6 × CH), 128.1 (6 × CH), 126.9 (3 × CH), 66.9 (C), 53.9 (CH), 52.3 (CH_3), 37.2 (CH_2) ppm.

3.1.5 Synthesis of *N*-[PMBSe-(CH_2)$_2$]-Cys(Trt)-OMe 13

4. In a round-bottom flask, dissolve under argon aldehyde **10** (838 mg, 3.45 mmol) in DCE (50 mL).

5. To this solution, successively add ≈ 4 g of activated powdered 3 Å molecular sieve (*see* **Note §8**), sodium triacetoxyborohydride (1.02 g, 4.81 mmol) and a solution of trityl-protected cysteine **12** (1.30 g, 3.44 mmol) in DCE (35 mL).
6. After 18 h of stirring at RT, check the completion of the reaction by TLC.
7. *Workup*. Filter the reaction mixture on a Büchner funnel and wash the solid with additional portions of DCM. Concentrate the filtrate under reduced pressure in a rotatory evaporator and take up the residue in a 0.5 M aqueous K_2CO_3 solution (50 mL). Transfer the aqueous mixture to a separatory funnel and extract it with DCM (3 × 50 mL). Dry the combined organic layers over $MgSO_4$, filter and evaporate the solvent under reduced pressure in a rotatory evaporator. Purify the crude by silica gel column chromatography (Cy/EtOAc 80:20) to obtain *N*-alkylated cysteine **13** as a clear yellow oil (1.17 g, 56%).
8. Check the identity and purity of the synthesized compound. **^{1}H NMR** (300 MHz, $CDCl_3$, 298 K) δ 7.41 (d, *J* = 7.4 Hz, 6H), 7.19-7.32 (m, 9H), 7.17 (d, *J* = 8.6 Hz, 2H), 6.79 (d, *J* = 8.6 Hz, 2H), 3.77 (s, 3H), 3.71 (s, 2H), 3.66 (s, 3H), 2.99 (t, *J* = 6.5 Hz, 1H), 2.37-2.74 (m, 6H), 1.44-1.91 (m, 1H) ppm. **^{13}C NMR** (75 MHz, $CDCl_3$, 298 K) δ 173.5 (C), 158.4 (C), 144.6 (3 × C), 131.1 (C), 129.9 (2 × CH), 129.6 (6 × CH), 127.9 (6 × CH), 126.7 (3 × CH), 113.9 (2 × CH), 66.8 (C), 60.2 (CH), 55.2 (CH_3), 51.9 (CH_3), 47.3 (CH_2), 34.6 (CH_2), 26.3 (CH_2), 23.9 (CH_2) ppm. **HRMS** (ES^+): Calcd. for $C_{33}H_{35}NO_3SNaSe$: 628.1401, found: 628.1378. **$[\alpha]_D^{20}$** -19 ° (*c* 1.0, chloroform).

3.1.6 Synthesis of *N*-Fmoc-*N*-[PMBSe-(CH_2)$_2$]-Cys(Trt)-OMe 14

1. In a round-bottom flask maintained under argon, dissolve compound **13** (455 mg, 0.752 mmol) in DCM (3.75 mL)
2. Successively add DIEA (0.137 mL, 0.786 mmol) and Fmoc-Cl (204 mg, 0.788 mmol).
3. Stir the reaction mixture at RT for 24 h.
4. *Workup*. Dilute the reaction mixture with DCM (20 mL). Transfer the DCM solution to a separatory funnel and wash it with water (2 × 20 mL). Dry the organic layer over $MgSO_4$, filter and evaporate the solvent under reduced pressure in a rotatory evaporator. Purify the crude mixture by silica gel column chromatography (Cy/EtOAc 7:3) to obtain Fmoc-protected compound **14** as a white solid (522 mg, 84%).
5. Check the identity and purity of the synthesized compound. **^{1}H NMR** (300 MHz, $CDCl_3$, 298 K) and **^{13}C NMR** (75 MHz, $CDCl_3$, 298 K). Use the provided NMR FID files for compound **14** to make comparisons (*see* **Notes §9** and **10**). **HRMS** (ES^+): Calcd. for $C_{48}H_{45}NO_5SNaSe$: 850.2081, found: 850.2089. **$[\alpha]_D^{20}$** (c 1.0, $CHCl_3$): -38 °.

3.1.7 Synthesis of Fmoc-SetCys-OMe 15

1. In a round-bottom flask maintained under argon, dissolve compound **14** (1.67 g, 2.02 mmol) in DCM (50 mL) and add $NaHCO_3$ (508 mg, 6.06 mmol).
2. Place the flask in an ice bath and add dropwise by syringe a solution of I_2 (1.54 g, 6.06 mmol) in DCM (50 mL).
3. Remove the ice bath and stir the reaction mixture at RT for 30 min. Check the complete consumption of the starting material by TLC (*see* **Note §11**).
4. *Workup*. Add 1 M $Na_2S_2O_3$ (50 mL) to remove the excess of I_2. Separate the organic and aqueous phases in a separatory funnel (*see* **Note 12**) and extract the aqueous layer with DCM (2 × 50 mL). Wash the combined DCM layers layer with brine (50 mL) and dry over $MgSO_4$. Filter and concentrate the solvent under reduced pressure in a rotatory evaporator (*see* **Note §13**). Purify the crude by silica gel column chromatography (Cy/EtOAc 8:2) (*see* **Note §14**) to obtain Fmoc-SetCys-OMe **15** as a colorless glass (780 mg, 83%).
5. Check the identity and purity of the synthesized compound. **^{1}H NMR** (300 MHz, $CDCl_3$, 298 K) and **^{13}C NMR** (75 MHz, $CDCl_3$, 298 K). Use the provided NMR FID files for compound **15** to make comparisons (*see* **Notes §9** and **10**). **HRMS** (ES^+) Calcd. for $C_{21}H_{21}NO_4NaSSe$: 486.0254, found: 486.0246. **$[\alpha]_D^{20}$** (c 1.0, $CHCl_3$): -29 °.

3.1.8 Synthesis of Fmoc-SetCys-OH 6

1. In a round-bottom flask equipped with reflux condenser and maintained under argon, dissolve compound **15** (532 mg, 1.15 mmol) in dioxane (22 mL).
2. Add 5 M HCl (11 mL).
3. Place the flask in an oil bath and heat the mixture for 12 h (overnight) at 80 °C (oil bath temperature).
4. Remove the oil bath and add at RT a 5% aqueous K_2CO_3 solution (200 mL) in order to raise the pH above 9 (*see* **Note §6**).
5. *Workup*. Transfer the mixture to a separatory funnel and wash it with Et_2O (2 × 100 mL). Acidify the aqueous layer with 12 N HCl until pH ≈ 2-3 (*see* **Note §6**) and extract it with DCM (3 × 100 mL). Dry the combined DCM layers over $MgSO_4$, filter and evaporate the solvent under reduced pressure in a rotatory evaporator. Fmoc-SetCys-OH **6** is isolated as a light yellow solid (488 mg, 95%) and is pure enough to be used in Fmoc-SPPS without further purification.

6. Check the structure and purity of the synthesized compound. **^{1}H NMR** (300 MHz, $CDCl_3$, 298 K) and **^{13}C NMR** (75 MHz, $CDCl_3$, 298 K). Use the provided NMR FID files for compound **6** to make comparisons (*see* **Notes §9** and **10**). **HRMS** (ES^+): Calcd. for $C_{20}H_{19}NO_4NaSSe$: 472.0098, found: 472.0116. **$[\alpha]_D^{20}$** (c 1.0, $CHCl_3$): -32 °.

3.2 Synthesis of the bifunctional peptide 2 (SetCys-HGF 150-176-MPA)

The incorporation of Fmoc-SetCys-OH residue **6** at the N-terminus of a bifunctional peptide 2 is achieved as described in Figure 4. The peptide is first elongated on a SEA-functionalized solid support using an automated peptide synthesizer. Fmoc-SetCys-OH **6** is then manually coupled to the elongated peptide using DIC/HOBt activation. After Fmoc removal, the peptide is cleaved from the solid support, liberating a bifunctional N-terminal SetCys and a C-terminal SEA^{on} group. The latter is immediately subjected to a SEA / 3-mercaptopropionic (MPA) exchange reaction in the absence of TCEP to avoid any opening of the SetCys group by reduction of the selenosulfide bond. The whole detailed procedure is described below.

[Insert Figure 4 near here]

Elongation of the peptide on SEA-functionalized solid support:

1. Elongate the HGF 150-176 peptide segment by Fmoc-SPPS (0.05 mol scale) on a SEA-functionalized solid support (loading 0.16 mmol/g) using the protocol described in ref [16]. At the end of the automated process, the peptidyl solid support is in suspension in DCM.
2. Transfer the peptidyl solid support obtained in step 1 into a glass reactor for manual solid phase synthesis equipped with a fritted glass and a stopcock at its bottom. Drain the DCM.
3. Wash the beads with DMF (3 × 10 mL) and drain.
4. Swell the peptidyl solid support in the minimal volume of DMF.
5. In a 5-mL microfuge tube, dissolve the SetCys amino acid **6** (44.8 mg, 0.100 mmol) in DMF (2.5 mL). Successively add HOBt (15.3 mg, 0.100 mmol) and DIC (15.6 µL, 0.100 mmol) and stir the mixture for 2 min.
6. Add the pre-activated amino acid (from step 5) to the reactor containing the solid support. Shake the mixture on a wrist-action shaker for 3 h.
7. Wash the beads with DMF (3 × 10 mL) and drain (*see* **Note §15**).

8. Add 10 mL of capping mixture, gently agitate on the wrist-action shaker for 15 min and drain.
9. Wash the beads with DMF (3 × 10 mL) and drain.
10. Add the piperidine solution for Fmoc removal (10 mL), shake for 10 min and drain the solid support.
11. Repeat step 10 a second time.
12. Wash the beads with DMF (3 × 10 mL), DCM (3 × 10 mL) and Et_2O (3 × 10 mL) and drain.

Cleavage of the peptide from the solid support:

13. Dry the solid support inside the reactor by connecting a vacuum pump to the lower stopcock.
14. Add a 10 mL cocktail of TFA/H_2O/TIS/thioanisole/thiophenol 87.5/2.5/5/2.5/2.5 (v/v/v/v/v) to the dry solid support. Shake the mixture on a wrist-action shaker for 2 h (*see* **Note §16**).
15. Precipitate the SEA^{on} peptide by adding dropwise the TFA solution from step 14 to an ice-cooled and stirred solution (200 mL) of Et_2O and *n*-heptane (1:1 v/v).
16. Transfer the suspension containing the precipitated SEA^{on} peptide to 50 mL plastic tubes (*see* **Note §17**), centrifuge at 4500 rpm for 5 min and discard carefully the supernatant.
17. Resuspend the pellet in Et_2O (100 mL), vortex the suspension, centrifuge for 5 min and discard carefully the supernatant.
18. Let the peptide dry by placing the open tubes for about 5 min in the fume hood (*see* **Note §18**).

SEA^{on} / MPA exchange reaction:

19. Fill a 1.5-mL microfuge tube with MPA (0.8 mL).
20. Fill a 50 mL plastic tube with deionized water (15 mL)
21. Fill a 1.5-mL microfuge tube with 6 M NaOH (1 mL).
22. Weigh guanidine hydrochloride (10 g) in a 50 mL plastic tube.
23. Introduce into a glove box the cleaved peptide (step 18) as well as all the reagents needed for the SEA^{on} / MPA exchange reaction (steps 19-22) (*see* **Notes §19** and **20**).
24. Into the glovebox, mix MPA (0.625 mL) and water (11.9 mL) in an empty 50 mL plastic tube. Adjust the pH of the solution to 4.0 by adding 6 M NaOH (≈ 400 µL) with a micropipette. Add this solution to the peptide.

25. Add guanidine hydrochloride (≈ 8.6 g) until complete solubilization of the peptide (*see* **Note §21**) and readjust the pH to 4.0 by addition of 6 M NaOH (≈ 250 µL).
26. Let the reaction proceed at 37°C.
27. Monitor the progress of the SEA^{on} / MPA exchange reaction by UPLC-MS analysis (*see* **Note §22**). Collect 5 µL aliquots of the reaction mixture and dilute these samples with 10% AcOH in water (100 µL) prior to UPLC-MS analysis.
28. Once the SEA^{on} / MPA exchange reaction is over (16 h), add AcOH (1.5 mL). Remove the reaction mixture from glovebox and extract the reaction mixture with Et_2O (4 ×10 mL) to remove MPA. Filter the crude solution through a disposable syringe filter.
29. Purify the peptide by preparative HPLC. Use the following elution conditions. Gradient: 0-10% eluent B in 5 min then 10-35% in 30 min, flow rate: 20 mL/min, RT. Analyze the fractions by MALDI-TOF mass spectrometry.
30. Pool the pure fractions in a 50 mL plastic tube, freeze the solution with liquid nitrogen and lyophilize. The bifunctional peptide **2** (25.3 mg, 20% overall from SEA PS solid support) was isolated as a white powder (*see* **Note §23**).
31. Check the identity and purity of the synthesized peptide by UPLC-MS analysis. R_t = 11.24 min. *m/z* = 1766.25 ($[M+2H]^{2+}$), 1177.75 ($[M+3H]^{3+}$), 883.50 ($[M+4H]^{4+}$); calcd. for M (average): 3530.25, found: 3530.82.

SEA^{off} peptide **1** and cysteinyl peptide **3** can be prepared according to the published procedures [17].

3.3 One-pot C-to-N three peptide segment assembly of K1 domain

1. Prepare two 1.5-mL microfuge tubes, each containing 287 mg of Gn·HCl (*see* **Note §24**).
2. Prepare two 1.5-mL microfuge tubes, each containing 16.8 mg of MPAA (*see* **Note §25**).
3. Weigh sodium ascorbate (19.8 mg) in a 1.5-mL microfuge tube (*see* **Note §26**).
4. Weigh TCEP·HCl (28.6 mg) in a 1.5-mL microfuge tube.
5. Weigh SEA^{off} peptide **1** (2.18 mg, 0.608 µmol, 1 equiv.) in a 0.5-mL microfuge tube.
6. Weigh SetCys peptide **2** (2.50 mg, 0.610 µmol, 1 equiv.) in a 0.5-mL microfuge tube.
7. Weigh cysteinyl peptide **3** (2.78 mg, 0.609 µmol, 1 equiv.) in a 0.5-mL microfuge tube.
8. Transfer 1 mL of 0.1 M sodium phosphate buffer (pH 7.2) in a 1.5-mL microfuge tube.

9. Prepare two 1.5-mL microfuge tubes filled respectively with 1 mL of NaOH 6 N and 1 mL of HCl 6 M as well as several 1.5-mL microfuge tubes containing 100 µL of 10% AcOH in deionized water (vol/vol).
10. Cap all the tubes (steps 1-9) and transfer them to the glove box under nitrogen atmosphere (*see* **Notes §19** and **20**).

First ligation: Assembly of the elongated SetCys peptide **4**.

11. Transfer 300 µL of 0.1 M, pH 7.2 phosphate buffer (step 8) to one of the microfuge tubes containing Gn·HCl (see step 1). Vortex until dissolution of the guanidinium salt and transfer this solution to one of the microfuge tubes containing MPAA (see step 2). Adjust the pH of the solution to 7.3 by addition of 6 M NaOH (*see* **Note §27**). The final concentration of MPAA is 200 mM.
12. Transfer 152 µL of the MPAA solution prepared in step 11 to the microfuge tube containing peptide **1** (step 5). Vortex until dissolution of peptide **1** and transfer the solution to the microfuge tube containing peptide **3** (step 7). Vortex until dissolution of peptide **3** and transfer the solution to the microfuge tube containing peptide **2** (step 6). Vortex until dissolution of peptide **2** (*see* **Note §28**). The final concentration of each peptide is 4 mM. Note that the dissolution of peptides **1**, **2** and **3** leads to a decrease of the pH to 6.8.
13. Place the reaction tube in a heating block kept at 37 °C and allow the mixture to react for 4.5 h.
14. Check by UPLC-MS analysis that peptide segments **2** and **3** have been consumed to produce the elongated SetCys peptide **4** and that SEA peptide segment **1** has not reacted. Transfer 3 µL of the reaction mixture to a microfuge tube containing 100 µL of AcOH 10% (vol/vol) in deionized water. Vortex and remove the sample from the glove box. Extract the sample with diethyl ether (4 × 0.5 mL) to remove MPAA (*see* **Note §29**) and add TCEP (3 mg) (*see* **Note §30**). Inject 10 µL of the reaction sample into the UPLC-MS system

Second ligation: Assembly of the biotinylated HGF 125-209 polypeptide **5**.

15. Transfer 300 µL of 0.1 M, pH 7.2 phosphate buffer (see step 8) to the second microfuge tube containing Gn·HCl (see step 1). Vortex until dissolution of the guanidinium salt and transfer the solution to the microfuge tube containing sodium ascorbate (step 3). Vortex until dissolution of sodium ascorbate and transfer the solution to the microfuge

tube containing TCEP (step 4). Vortex until dissolution of TCEP and transfer the solution to the second microfuge tubes containing MPAA (step 2). Adjust the pH of the solution to 5.4 by addition of 6 M NaOH (≈ 60 µL) (*see* **Note §27**). The final concentration of MPAA, TCEP and sodium ascorbate is 200 mM.

16. Transfer 152 µL of the solution prepared in step 15 to the centrifuge tube containing the first ligation mixture (step 13). Adjust the pH of the solution to 5.5 by addition of 6 M HCl (≈ 2.5 µL) (*see* **Note §31**). The final concentrations of MPAA, TCEP and sodium ascorbate are respectively 200 mM, 100 mM and 100 mM.
17. Place the reaction tube in a heating block kept at 37 °C.
18. Monitor the ligation reaction leading to polypeptide **5** by UPLC-MS analysis. Transfer 4 µL of the reaction mixture to a microfuge tube containing 100 µl of AcOH 10% (v/v) in deionized water. Vortex and remove the sample from the glove box. Extract the sample with diethyl ether (4 × 0.5 mL) to remove MPAA. Inject 10 µL of the reaction sample into the UPLC-MS system to confirm that peptide segments **1** and **4** have been consumed to produce the polypeptide **5**.
19. After completion of the reaction (40 h), remove the tube containing the reaction mixture from the glove box.
20. Transfer the content of the reaction tube to a 15-mL plastic tube and add 8 mL of a 7.5% AcOH solution in deionized water (v/v). Extract the mixture with diethyl ether (4 × 5 mL) to remove the MPAA. Filter the crude solution through a disposable syringe filter.
21. Purify the target polypeptide **5** by preparative HPLC. Use the following elution conditions. Gradient: 0-10% eluent B in 5 min then 10-35% in 30 min, 6 mL/min, RT. Analyze the fractions by MALDI-TOF mass spectrometry.
22. Pool the purified fractions in a 50 mL plastic tube, freeze the solution with liquid nitrogen and lyophilize it for 2 days. The reaction yielded the polypeptide **5** (2.05 mg, 29%) as a white powder.
23. Check the purity and identity of the synthesized peptide by UPLC-MS and MALDI-TOF mass spectrometry. **LC-MS analysis**. R_t = 11.40 min. *m/z* = 1666.17 ($[M+6H]^{6+}$), 1428.42 ($[M+7H]^{7+}$), 1250.08 ($[M+8H]^{8+}$), 1111.25 ($[M+9H]^{9+}$), 1000.25 ($[M+10H]^{10+}$), 909.42 ($[M+11H]^{11+}$), 833.75 ($[M+12H]^{12+}$), 769.58 ($[M+13H]^{13+}$), 714.67 ($[M+14H]^{14+}$), 667.17 ($[M+15H]^{15+}$). Calcd. for M (average): 9993.40, found: 9992.14. **MALDI-TOF analysis** (matrix: 2,5-dihydroxybenzoic acid). Calcd. for $[M+H]^+$ (average): 9994.4, found: 9993.1.

4 Notes

Note §1. The synthesis of the SetCys amino acid **6** was repeated several times on gram scale by different experimenters. The yields were quite reproducible.

Note §2. The synthesis of the SetCys amino acid **6** by an experienced chemist requires 3 weeks. Once synthesized, the SetCys residue is stable and can be stored for months in the fridge (4 °C).

Note §3. After extraction, the aqueous layer has to be neutralized with bleach since the reaction with KSeCN leads to the formation of one equivalent of cyanide anion.

Note §4. The addition of $NaBH_4$ leads to active liberation of molecular hydrogen. The excess of gas should be evacuated through an outlet.

Note §5. If aldehyde **10** cannot be used for the next step on the same day, we advise to store it at -20 °C under an argon atmosphere overnight.

Note §6. The addition of K_2CO_3 solution or concentrated HCl to the mixture can induce CO_2 release and bubbling.

Note §7. The pH of the aqueous solution of K_2CO_3 must be at least pH 9 after crude dissolution to best achieve free amine extraction by the organic solvent. If the pH is too low, add more 0.25 M K_2CO_3 aqueous solution.

Note §8. The powdered 3 Å molecular sieve was activated under high vacuum (< 1 mbar), by heating the flask with a heatgun for ~15 min. The flask containing the molecular sieve was filled with argon during cooling to RT.

Note §9. There are strong conformational restrictions in compounds **14**, **15** and **6**. As a consequence, NMR spectra of these compounds at 298 K are complex due to the overlapping of the signals given by all the conformers present in solution. Because no clear description of the NMR spectra can be provided, FID files are given to facilitate the identification of the synthesized compounds by direct comparison. These files can be opened using for example TopSpin® software from Bruker.

Note §10.The presence of the Fmoc group in compound **14** induces conformational restrictions in the molecule. However, variable temperature NMR experiments performed in deutered DMSO show that such conformational barriers can be overcome by heating the sample, leading to the coalescence of NMR signals above 70 °C. Once the selenosulfide bond is formed by iodine oxidation, the conformational restrictions are so strong in the resulting cyclic species (compounds **15** and **6**) that several conformers are detected by NMR in solution, whatever the temperature.

Note §11. For TLC analysis, one drop of the crude mixture is poured in a biphasic mixture containing ~500 µL of a 1 M $Na_2S_2O_3$ solution and ~200 µL of EtOAc. The tube is vigorously

shaken and the upper layer is spotted on the TLC plate against the reference for the starting material.

Note §12. If needed, add brine to improve the separation between aqueous and organic layers.

Note §13. We noticed that evaporation of the crude mixture to dryness before purification systematically resulted in lower isolated yields. Therefore, we recommend to concentrate the mixture to a final volume of 5-10 mL.

Note §14. The silica gel chromatography column is loaded with the solution (final volume of 5-10 mL) obtained after concentrating the crude mixture under reduced pressure.

Note §15. At this stage, we highly recommend to check the completion of the coupling step. Take few polymer beads in order to perform a microcleavage and mass analysis. If coupling is not complete, repeat steps 3-7. Perform microcleavage as described in steps 12-16 using 1 mL of TFA cocktail.

Note §16. The shaking power should be set appropriately to guarantee the mixing, but not too vigorous to avoid the continuous projection of TFA on the reactor stopper. Plastic stoppers are weakly resistant to TFA. Their degradation products may contaminate the peptide-containing cleavage solution.

Note §17. For safety reasons, the centrifuge tubes should be correctly balanced by mass.

Note §18. If the tranthioesterification step is not performed straightaway, dissolve the dried peptide in water with 0.1 % TFA, freeze the solution with liquid nitrogen and lyophilize it.

Note §19. The glove box allows performing transthioesterification as well as ligation experiments in a nitrogen atmosphere with oxygen levels below 10 ppm. Very low oxygen levels are recommended since molecular oxygen oxidizes thiol or selenol species (SEA^{on} thiols, Cys thiol, MPAA …) into unproductive dichalcogenides. This side-reaction cannot be reversed when reactions are performed in the absence of TCEP as for some chemical transformations described in this protocol.

Note §20. The tubes that contain peptide segments, reagents and buffers must be tightly capped during transfer into the glove box, which requires purging the airlock by pumping and filling with nitrogen.

Note §21. Guanidine hydrochloride is a strong peptide denaturant. It was added at the beginning of the SEA^{on} / MPA exchange reaction to improve the low solubility of the SEA^{on} peptide in the MPA solution at pH 4.

Note §22. We noticed that bifunctional peptides such as compound **2** can overreact, leading to the formation of cyclic SetCys peptides by intramolecular ligation. To minimize this side

reaction, the transthioesterification step has to be carefully monitored and stopped immediately after completion.

Note §23. The yield is calculated, based on the quantity of solid support used to elongate the bifunctional peptide **2**.

Note §24. Due to its denaturant properties, guanidine hydrochloride is used as classical additive during ligation. At a 6 M concentration, it avoids any partial folding of the peptide segments that could mask reactive ends and affect ligation efficiency.

Note §25. MPAA catalyzes NCL and SEA ligation processes through the *in situ* formation of an arylthioester peptide segment by a thiol-thioester exchange mechanism. Usually, the aryl thioester peptide derived from the alkyl thioester or SEA peptide reacts quickly with the cysteinyl component so that it is usually not observed in the HPLC traces of the crude ligation mixture.

Note §26. Sodium ascorbate is a strong radical quencher. It avoids the TCEP-induced deselenization side-reaction that converts the SetCys residue into unproductive *N*-ethyl cysteine.

Note §27. Dissolution of high concentrations of TCEP and/or MPAA in the phosphate buffer results in a significant decrease of the pH well below pH 5.5. MPAA is poorly soluble in these conditions. MPAA is highly soluble in water at and above pH 5.5 due to the ionization of its carboxylic acid group. MPAA will therefore solubilize gradually upon addition of NaOH and vortexing. At this stage, a precise control of the pH is not mandatory. However, the addition of NaOH must be done carefully to avoid a significant increase of the pH, that would require the addition of large volumes of HCl in return.

Note §28. Peptides **1**, **2** and **3** must be added in the MPAA solution in the order described in the protocol (**1** -> **3** -> **2**) to avoid the formation of a SetCys cyclic by-product at the beginning of the reaction. In presence of MPAA, the SetCys residue and the thioester group of bifunctional peptide **2** can react, leading to the cyclisation of the peptidic backbone by intramolecular ligation. However, this side-reaction is kinetically disfavored compared to classical NCL between peptides **1** and **2**. The SetCys reactivity is not an issue as long as a competitive cysteine is present in the reaction mixture.

Note §29. The large excess of MPAA used for catalyzing the ligation must be extracted with diethyl ether before running the analytical or the semi-preparative HPLC systems. This arylthiol absorbs strongly at the wavelengths used for the UV detection by the HPLC systems and can co-elute with the target peptide. The efficient extraction of MPAA requires the acidification of

the reaction mixture well below the p*K*a of the aromatic thiol. This is usually achieved by adding 10% AcOH (v/v) to the reaction mixture.

Note §30. By enabling the reduction of all the dichalcogenides that are contained in the sample after the MPAA extraction, the addition of TCEP facilitates the analysis of the ligation mixture. Following this approach, HPLC chromatograms show a single set of peaks corresponding to the reactants and the products of the ligation under their reduced forms.

Note §31. The pH of the reaction was decreased to 5.5 in order to accelerate the SEA ligation between peptide segments **1** and **4**. In the SEA ligation, the peptide thioester which reacts with the cyteinyl peptide to give the ligated product is produced *in situ* from the SEA^{on} peptide by *N*,*S*-acyl shift (Figure 1). Such a rearrangement is pH-dependent and is favored under mild acidic conditions, typically pH 5.5. Decreasing the pH to 5.5 also accelerate the spontaneous decomposition of the SetCys residue into cysteine. Previous studies [13] showed that such a reaction proceeds through an intramolecular substitution of the ammonium group by the selenoate ion (Figure 1). The rate of SetCys decomposition is optimal around pH 5.5-6.0 since the ammonium group and the selenoate ion are the predominant species at this working pH and are respectively the best leaving group and nucleophile for the substitution reaction. We also noticed that the departure of the 2-selenoethyl limb is greatly facilitated by the nearby SetCys thiol, see ref [13].

Acknowledgements

This study was supported by the Centre National de la Recherche Scientifique (CNRS), the University of Lille Nord de France, Institut Pasteur de Lille.

Figures

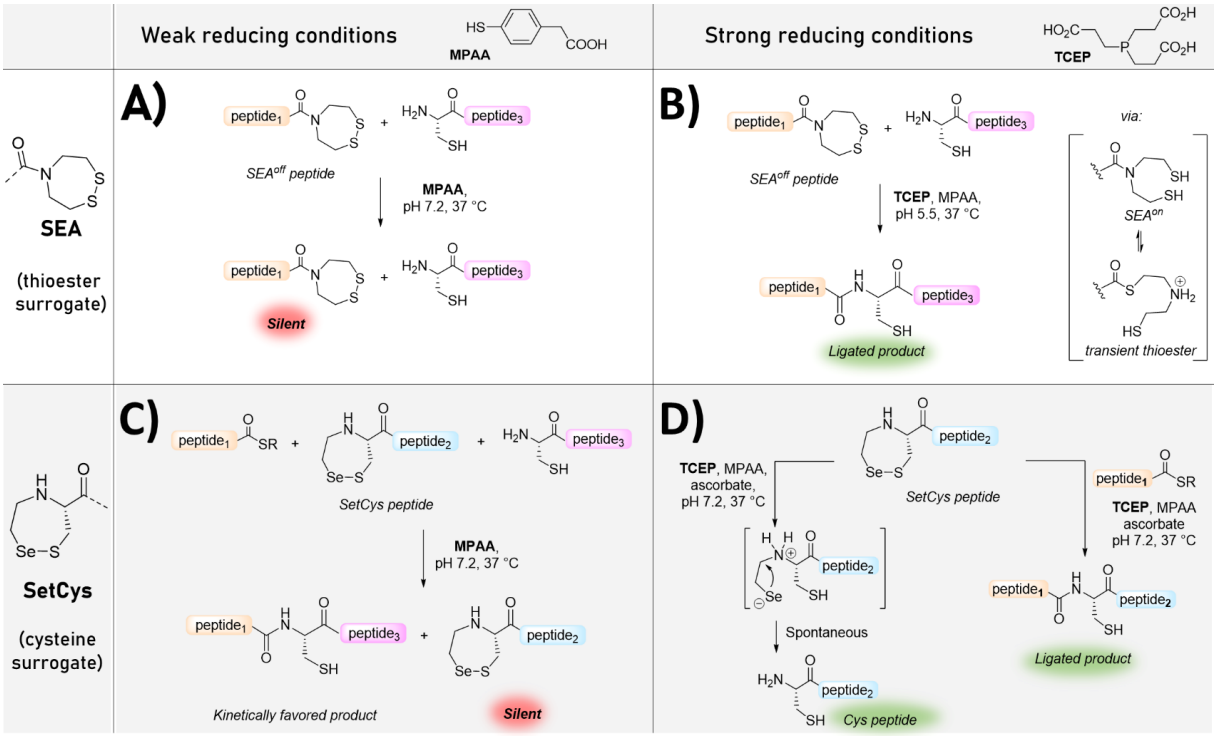


Figure 1. The reactivity of *N*-(2-selanylethyl)cysteine (SetCys) and SEA^{off} peptides is controlled by the nature of the reducing agent present in solution.

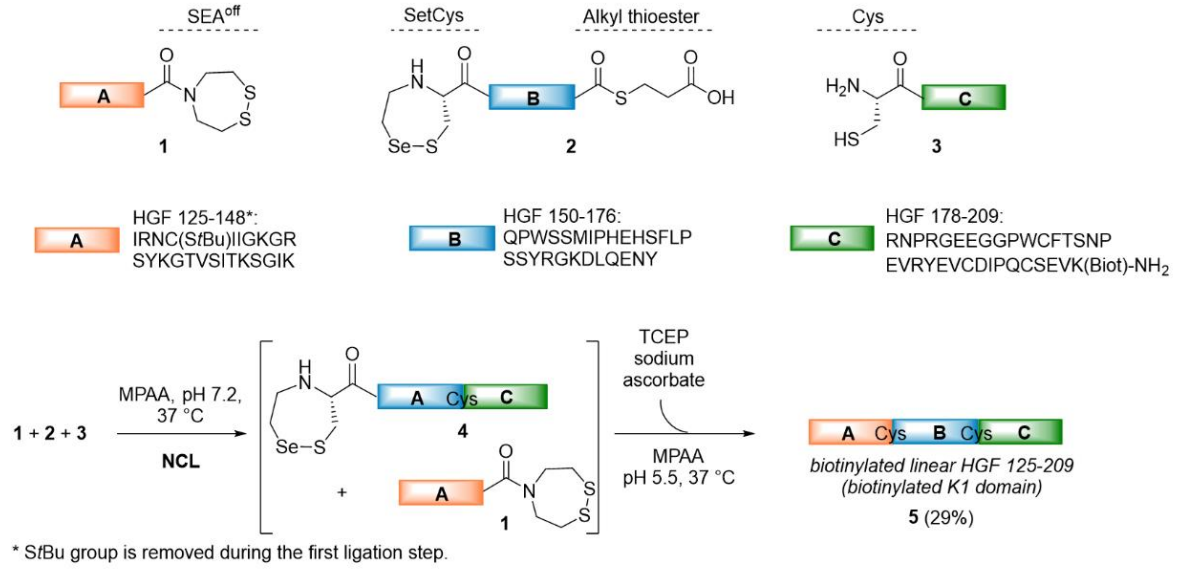


Figure 2. Application of SetCys to the one-pot three peptide segment assembly of proteins is illustrated in this protocol by the total chemical synthesis of biotinylated HGF 125-209 polypeptide, which corresponds to the kringle 1 (K1) domain of HGF.

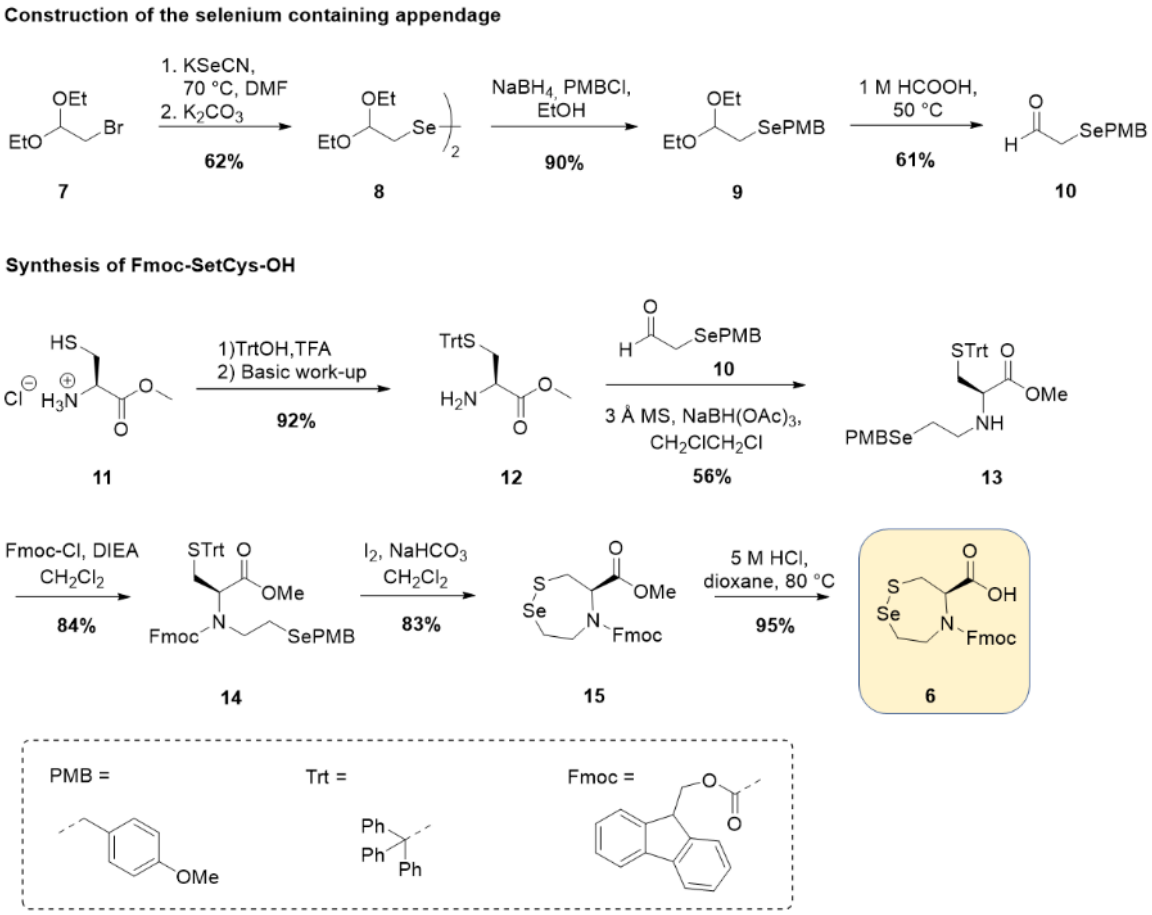


Figure 3. Synthesis of the Fmoc-protected SetCys amino acid **6**.

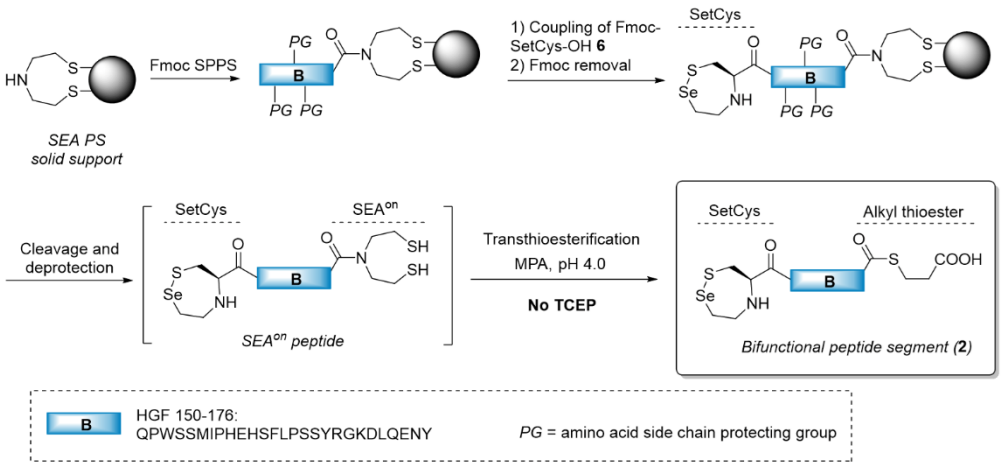


Figure 4. General approach for the preparation of *N*-terminal SetCys peptide thioesters.

[1]Dawson P E, Muir T W, Clark-Lewis I, Kent S B H (1994) Synthesis of proteins by native chemical ligation. Science 266:776-779

[2] Agouridas V, El Mahdi O, Diemer V, Cargoët M, Monbaliu J-C M, Melnyk O (2019) Native chemical ligation and extended methods. mechanisms, catalysis, scope and limitations. Chem Rev 119:7328−7443

[3] Conibear A C, Watson E E, Payne R J, Becker C F W (2018) Native chemical ligation in protein synthesis and semi-synthesis. Chem Soc Rev 47:9046−9068

[4] Brik A, Dawson P E, Liu L (eds) (2021) Total chemical synthesis of proteins. Wiley-VCH, Weinheim

[5] Agouridas V, El Mahdi O, Cargoët M, Melnyk O (2017) A statistical view of protein chemical synthesis using NCL and extended methodologies. Bioorg Med Chem 25:4938-4945

[6] Raibaut L, Ollivier N, Melnyk O (2012) Sequential native peptide ligation strategies for total chemical protein synthesis. Chem Soc Rev 41:7001-7015

[7] Hayashi Y (2016) Pot economy and one-pot synthesis. Chem Sci 7:866-880

[8] Agouridas V, Diemer V, Melnyk O (2020) Strategies and open questions in solid-phase protein chemical synthesis. Curr Opin Chem Biol 58:1-9

[9] Loibl S F, Harpaz Z, Zitterbart R, Seitz O (2016) Total chemical synthesis of proteins without HPLC purification. Chem Sci 7:6753-6759

[10] Zuo C, Zhang B, Yan B, Zheng J-S (2019) One-pot multi-segment condensation strategies for chemical protein synthesis. Org Biomol Chem 17: 727-744

[11] Ollivier N, Dheur J, Mhidia R, Blanpain A, Melnyk O (2010) *Bis*(2-sulfanylethyl)amino native peptide ligation. Org Lett 12:5238-5241

[12] Ollivier N, Vicogne J, Vallin A, Drobecq H, Desmet R, El-Mahdi O, Leclercq B, Goormachtigh G, Fafeur V, Melnyk O (2012) A one-pot three-segment ligation strategy for protein chemical synthesis. Angew Chem Int Ed 51:209-213

[13] Diemer V, Ollivier N, Leclercq B, Drobecq H, Vicogne J, Agouridas V, Melnyk O (2020) A cysteine selenosulfide redox switch for protein chemical synthesis. Nat Commun 11:2558

[14] Krief A, Dumont W, Delmotte C (2000) Reaction of organic selenocyanates with hydroxides: the one-pot synthesis of dialkyl diselenides from alkyl bromides. Angew Chem Int Ed 39:1669-1672

[15] Abbas M, Bethke J, Wessjohann L A (2006) One pot synthesis of selenocysteine containing peptoid libraries by Ugi multicomponent reactions in water. Chem Commun 541–543

[16] Bouchenna J, Sénéchal M, Drobecq H, Vicogne J, Melnyk O (2020) The problem of aspartimide formation during protein chemical synthesis using SEA-mediated ligation. In: Iranzo O, Roque A C (eds) Peptide and protein engineering: From concepts to Biotechnological applications, Springer Protocol Handbooks, pp 13-28

[17] Boll E, Drobecq H, Ollivier N, Blanpain A, Raibaut L, Desmet R, Vicogne J, 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 10:269–292