



**UNIVERSITÉ DE STRASBOURG**

**EDSC**  
École Doctorale des  
Sciences Chimiques

**ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES – ED222**  
**Institut de Science et d'Ingénierie Supramoléculaires – UMR 7006**

**THÈSE** présentée par :

**Elise NAUDIN**

soutenue le : **17 juillet 2020**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Chimie

**Développement de protéines  
synthétiques *de novo* catalysant le  
transfert de groupements acyles**

**THÈSE dirigée par :**

**M. TORBEEV Vladimir**

Maître de conférences, HDR, université de Strasbourg

**RAPPORTEURS :**

**M. GUICHARD Gilles**

**M. MELNYK Oleg**

Directeur de recherches, CNRS, université de Bordeaux

Directeur de recherches, CNRS, université de Lille

---

**AUTRES MEMBRES DU JURY :**

**M. WAGNER Alain**

**M. TORBEEV Vladimir**

Directeur de recherches, CNRS, université de Strasbourg

Maître de conférences, HDR, université de Strasbourg



# Remerciements, acknowledgements

J'aimerais remercier les membres du jury, le Dr. Gilles Guichard, directeur de recherches à l'université de Bordeaux, le Dr. Oleg Melnyk, directeur de recherches à l'université de Lille et le Dr. Alain Wagner, directeur de recherches à l'université de Strasbourg de m'avoir fait l'honneur de juger ces travaux de thèse. Je remercie aussi le Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation de m'avoir attribué une bourse de thèse ainsi que le programme ERC (European Research Council) d'avoir financé ces travaux.

I would like to warmly thank my supervisor, Dr. Vladimir Torbeev for the great opportunity to work in his group. I have learnt so much from our discussion, your knowledge and your passion for protein science that you have been able to share with me. I have really appreciated your support as well as the freedom and the trust you have given me. Thank you for giving me the possibility to work in such good conditions, to learn so many techniques, to go beyond my comfort zone, to travel and to meet the top world scientists. I will never forget the wonderful chance you gave me to visit the laboratory of Prof. DeGrado at San Francisco, I could not have dreamt better opportunity as a PhD student. Thank you for everything!

I would like to express my eternal gratitude to Bill DeGrado for his precious advises and his priceless guidance in the course of my PhD studies. I would like also to thank you for your warm welcome as well as for your expertise and your knowledge on protein design you shared with me. Thank you for the scientific (and less scientific) moments we shared together during my visit in your lab, and thank you again for the wonderful day we spent with Susan around a baseball game and an amazing sunset at the Golden Gate Bridge. I hope to meet you up again someday! I would like to thank all your lab members for their welcome, kindness, positive mood and valuable help. Big thanks to Nick, Nathan, Sophia and Marco that have particularly helped me for my project.

I would like to thank all the present and past members of the lab for the time we spent together in such a great atmosphere. I would like to specifically thank the post-docs, Ana, Robrecht, Boris, Jérémy and Abhishek for our scientific and non-scientific discussions, for your strong support, help, and advices. You have been always here for me and it has meant a lot to me. Thank you to all the students, Adeline, Claudie, Robertinah, Joris and Louis that I supervised during my PhD and who assisted me in my research. It was a great experience to teach you all the techniques necessary for your project.

Special thanks to my PhD colleagues and friends, Régis, Aromal and Valentin. We have spent so many good and funny moments together, I will never forget. You have been a

true support for me, and the words fail me to express all the gratitude I have for you. You have been always present to help me and our discussions have helped me to move forward so many times. Without you, my thesis would have been quite sad. Thank you for everything!

Je tiens aussi à remercier toutes les personnes du staff de l'ISIS qui ont participé au bon fonctionnement de ma thèse. Un grand merci à Annia pour ton soutien et d'avoir toujours été là pour moi. Ta porte était toujours ouverte quand j'en avais besoin et nos discussions interminables ont beaucoup compté pour moi. Un merci tout particulier à Muriel, malgré ton emploi du temps de ministre, tu as toujours trouvé le temps de m'aider dans mes projets. Tes conseils et ton expérience ont été précieux ! Merci à Jean-Louis qui a toujours été là dans les galères techniques, pour ton aide et tes conseils, j'ai beaucoup appris à tes côtés. Merci à Philippe, Thierry, Fabien, Cyril, Fabienne, Marie-Claude, Arthur, Raphaël et merci à tous les membres de l'ISIS que j'ai pu fréquenter au cours de ces années pour votre gentillesse et votre sympathie. J'aimerais remercier le Centre de Biologie Intégrative de l'IGBMC et plus particulièrement Alastair, Pierre et Catherine pour m'avoir toujours bien accueillie et pour avoir partagé avec moi votre expertise avec passion.

Je tiens à remercier plus généralement toutes les personnes qui m'ont aidée lors de mes études, ma thèse et mon année de monitorat. Votre pédagogie, votre motivation pour la recherche et l'enseignement et vos conseils ont beaucoup compté pour moi et continueront de m'accompagner.

J'aimerais remercier mes amis qui m'ont toujours soutenue et qui m'ont changé les idées quand j'en avais le plus besoin. Une immense pensée pour les copines Pauline, Marie et Marie qui ont toujours été là pour moi et qui je sais, le seront pour la vie ; sans oublier mon petit Maxence, Valentin, Alex, Hadrien, Pierre, Thibault et la petite dernière Caroline, vous êtes ma seconde famille. Une pensée très particulière à mes amis de la fac et tout spécialement au Grignard crew : Marion, Marina, Céline, Xavier, Jonathan, Vincent, Elyse et Marco (et vos chéri(e)s !), merci d'avoir été à mes côtés tout au long de mes études supérieures. Merci aussi à Virgile, Joachim, Simon et Soumia pour la force que vous m'avez véhiculée.

Je remercie du fond du cœur toute ma famille, Marc pour tes précieux conseils, et plus particulièrement mes parents, Clément et Margot qui ont été d'un soutien sans faille. Un grand merci pour votre aide, vos encouragements et votre confiance que vous m'avez apporté tout au long de mes études.

J'aimerais remercier tout spécialement Valentin qui a été d'un immense soutien moral. Tu as su me réconforter et me rassurer dans les moments de doute. Merci pour ta patience, tes encouragements et merci d'avoir toujours cru en moi. Tu es mon pilier, mon moteur et ma force.

## Table of contents

<b>Abbreviations .....</b>	<b>5</b>
<b>Résumé de thèse.....</b>	<b>9</b>
<b>Chapter 1. General introduction.....</b>	<b>35</b>
<b>1.1. Functional <i>de novo</i> design: is it possible to efficiently install a catalytic function on a small structured scaffold?.....</b>	<b>36</b>
1.1.1. Natural enzymes.....	36
1.1.2. Installing a catalytic activity into a large and natural protein scaffold .....	40
1.1.3. Installing a catalytic activity into minimalistic <i>de novo</i> protein scaffold .....	50
1.1.4. Incorporation of a catalytic function on small-size structured and non-proteinogenic scaffolds.....	65
1.1.5. Conclusions.....	69
<b>1.2. Amide bond formation, a common but challenging reaction .....</b>	<b>71</b>
1.2.1. Natural and biological synthesis of amide bond .....	72
1.2.2. State-of-the-art of chemical approaches to form amide bond.....	74
1.2.3. Total synthesis of proteins .....	77
1.2.4. Conclusions.....	89
<b>1.3. Objectives and thesis outline .....</b>	<b>90</b>
<b>1.4. References .....</b>	<b>93</b>
<b>Chapter 2. Design, synthesis and characterization of <i>de novo</i> catalytic proteins for acyl transfer reactions .....</b>	<b>111</b>
<b>2.1. Introduction.....</b>	<b>112</b>
<b>2.2. Protein scaffold selection and protein design.....</b>	<b>115</b>
<b>2.3. Total chemical synthesis of the analogues .....</b>	<b>119</b>
2.3.1. Sequential Fmoc-SPPS .....	119
2.3.2. Fmoc-SPPS and two-segment ligation.....	120
2.3.3. One-pot three-segment ligation of peptides synthesized by Boc-SPPS .	123
2.3.4. Sequential Boc-SPPS.....	126

<b>2.4.</b>	<b>Biophysical characterization .....</b>	<b>128</b>
<b>2.5.</b>	<b>Crystal structures .....</b>	<b>134</b>
<b>2.6.</b>	<b>Optimization of catalytic assays for amide bond formation .....</b>	<b>139</b>
2.6.1.	Fluorescent catalytic assay based on FRET with a quencher .....	140
2.6.2.	Fluorescent catalytic assay based on FRET with two fluorophores.....	143
2.6.3.	Catalytic assays followed by HPLC – 1 <sup>st</sup> generation assay .....	146
2.6.4.	Catalytic assays followed by LC/MS – 2 <sup>nd</sup> generation assay .....	151
<b>2.7.</b>	<b>Characterization of the acyl transferase catalytic activity .....</b>	<b>154</b>
<b>2.8.</b>	<b>Scope of reactivity for acyl acceptor molecules and catalytic promiscuity .....</b>	<b>162</b>
<b>2.9.</b>	<b>Conclusions .....</b>	<b>165</b>
<b>2.10.</b>	<b>References .....</b>	<b>168</b>
<b>Chapter 3. Libraries of analogues and substrates: towards better understanding of the catalytic activity of <i>de novo</i> protein.....</b>		
<b>3.1.</b>	<b>Introduction.....</b>	<b>174</b>
<b>3.2.</b>	<b>Library of the second generation of DSD analogues .....</b>	<b>176</b>
3.2.1.	Design of the library of analogues.....	176
3.2.2.	Design of the synthetic route for 'negative' strand analogues .....	181
3.2.3.	Synthesis of the library of analogues .....	183
3.2.4.	Catalytic assays of the libraries of analogues .....	194
<b>3.3.</b>	<b>Library of substrates .....</b>	<b>200</b>
3.3.1.	Chemical synthesis of a library of peptide- <sup>α</sup> thioester substrates.....	201
3.3.2.	Chemical synthesis of a library of acyl acceptor substrates .....	204
<b>3.4.</b>	<b>Conclusions .....</b>	<b>206</b>
<b>3.5.</b>	<b>References .....</b>	<b>209</b>
<b>Chapter 4. Design, synthesis and characterization of elongated DSD analogues.....</b>		
<b>4.1.</b>	<b>Introduction.....</b>	<b>212</b>
<b>4.2.</b>	<b>Design of elongated analogues .....</b>	<b>215</b>
<b>4.3.</b>	<b>Total chemical synthesis of the elongated analogues.....</b>	<b>220</b>
<b>4.4.</b>	<b>Crystal structure.....</b>	<b>223</b>

4.5.	<b>Catalytic characterization</b> .....	<b>226</b>
4.6.	<b>Biophysical characterization: towards a better stabilization of the structure?</b> .....	<b>229</b>
4.7.	<b>Conclusions</b> .....	<b>230</b>
4.8.	<b>References</b> .....	<b>232</b>
<b>Chapter 5. An unexpected proteolytic activity?</b> .....		<b>233</b>
5.1.	<b>Introduction</b> .....	<b>234</b>
5.2.	<b>Hydrolytic activity of the Hom-N2 analogue</b> .....	<b>235</b>
5.2.1.	Hydrolytic activity with pNPA as substrate .....	235
5.2.2.	Unexpected proteolytic activity: a discovery by serendipity .....	236
5.3.	<b>Conclusions and perspectives</b> .....	<b>241</b>
5.4.	<b>References</b> .....	<b>243</b>
<b>Chapter 6. General conclusions, significance and perspectives</b> .....		<b>245</b>
<b>Chapter 7. Experimental part</b> .....		<b>251</b>
7.1.	<b>Materials, methods and general procedures</b> .....	<b>252</b>
7.1.1.	General procedures for solid-phase peptide synthesis.....	255
7.1.2.	General procedure for peptide thioesterification.....	259
7.1.3.	General procedure for concentration determination .....	260
7.1.4.	General procedures for catalytic assays .....	260
7.2.	<b>Chapter 2</b> .....	<b>263</b>
7.3.	<b>Chapter 3</b> .....	<b>280</b>
7.4.	<b>Chapter 4</b> .....	<b>302</b>
7.5.	<b>Chapter 5</b> .....	<b>311</b>
7.6.	<b>References</b> .....	<b>313</b>





# Abbreviations

**2-CT:** 2-chlorotrityl

**aaRS:** Aminoacyl-tRNA synthetase

**Å:** Angström

**Acm:** Acetamidomethyl

**ACTR:** Activation domain from the p160 transcriptional co-activator for thyroid hormone and retinoid receptors

**Ala (A):** Alanine

**Alloc:** Allyloxycarbonyl

**Asn (N):** Asparagine

**Ar:** Argon

**Arg (R):** Arginine

**Asp (D):** Aspartate

**a. u.:** Arbitrary unit

**Boc:** *tert*-butyloxycarbonyl

**Bom:** benzyloxymethyl

**Bzl:** Benzyl

**c:** Concentration

**°C:** Degree Celsius

**CD:** Circular Dichroism

**CSD:** Cambridge structural database

**Cys (C):** Cysteine

**δ:** Chemical shift

**d:** Doublet

**DCC:** *N,N'*-Dicyclohexylcarbodiimide

**DCM:** Dichloromethane

**DF:** DueFerri

**DIC:** *N,N'*-Diisopropylcarbodiimide

**DIEA:** Diisopropylethylamine

**DMF:** *N,N*-Dimethylformamide

**DMSO:** Dimethylsulfoxide

**DNA:** Deoxyribonucleic acid

**DSD:** Domain-Swapped Dimer

**DTT:** DL-Dithiothreitol

**ESI:** Electrospray Ionization

**EtOH:** Ethanol

**Eq:** Equivalent

**Fmoc:** Fluorenylmethyloxycarbonyl

**Gln (Q):** Glutamine

**Glu (E):** Glutamate

**Gly (G):** Glycine

**Gn-HCl:** Guanidine hydrochloride

**h:** hour

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

**HBTU:** 3-[Bis(dimethylamino)methylumyl]-3H-benzotriazol-1-oxide hexafluorophosphate

**HEPES:** 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid

**His (H):** Histidine

**HOBt:** 1-Hydroxy-1H-benzotriazole

**HPLC:** High Performance Liquid Chromatography

**Hz:** Hertz

**Ile (I):** Isoleucine

**$k_{cat}$ :** Turnover number

**$K_d$ :** Dissociation constant

**$K_m$ :** Michaelis-Menten constant

**$k_{obs}$ :** Observed rate constant

**LC-MS:** Liquid Chromatography-Mass Spectrometry

**Leu (L):** Leucine

**Lys (K):** Lysine

**m:** Multiplet

**M:** Mole per litre or molar

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

**MALDI-TOF:** Matrix Assisted Laser Desorption Ionization - Time of Flight

**MD:** Molecular Dynamics

**Me:** Methyl

**MeCN:** Acetonitrile

**MeDbz:** *o*-Amino(methyl)aniline

**MeNbz:** *N*-acyl-*N'*-methyl-benzimidazolinone or *N*-acyl-*N'*-methylurea

**MES:** 2-(*N*-morpholino)ethanesulfonic acid

**Mes:** 2-Mercaptoethanesulfonate

**MesNa:** Sodium 2-mercaptoethanesulfonate

**Met (M):** Methionine

**MOPS:** 3-(*N*-Morpholino)propanesulfonic acid

**MPA:** 3-Merceptopropionic acid

**MPAA:** 4-Mercaptophenylacetic acid

**MRE:** Mean residue ellipticity  
**mRNA:** Messenger ribonucleic acid  
**MS:** Mass Spectrometry  
**NCL:** Native chemical ligation  
**NMM:** *N*-methylmorpholine  
**NMP:** *N*-methylpyrrolidinone  
**OcHex:** Cyclohexyl ester  
**p:** *para*  
**PAL:** Peptide amide linker  
**Pbf:** 2,2,4,6,7-pentamethylidhydrobenzofuran-5-sulfonyl  
**PDB:** Protein data bank  
**Pd/C:** Palladium on carbon  
**PEG:** Polyethylene glycol  
**PEG 550 MME:** Polyethylene glycol monomethyl ether 550  
**pH:** Potential of hydrogen  
**Phe (F):** Phenylalanine  
**PLP:** pyridoxal 5'-phosphate  
**pNP:** *para*-nitrophenol  
**pNPA:** *para*-nitrophenylacetate  
**ppm:** Parts per million  
**Pro (P):** Proline  
**RMSD:** Root-Mean-Square Deviation  
**RNA:** Ribonucleic acid  
**RPM:** Revolutions per minute  
**r.t.:** Room temperature  
**s:** Singlet  
**SEC:** Size-exclusion chromatography  
**Ser (S):** Serine  
**SPE:** Solid-phase extraction  
**SPPS:** Solid-phase peptide synthesis  
**t:** triplet  
**tBu:** *tert*-butyl  
**TCEP:** Tris(2-carboxyethyl)phosphine  
**TFA:** Trifluoroacetic acid  
**Thr (T):** Threonine  
**Thz:** Thiazolidine-4-carboxylic acid  
**TIPS:** Triisopropylsilane  
**Tos:** Tosyl

**Tris:** Tris(hydroxymethyl)aminomethane

**tRNA:** tranfert ribonucleic acid

**Trp:** Tryptophan

**Trt:** Trityl

**TI:** Tetrahedral intermediate

**TS:** Transition state

**Tyr (Y):** Tyrosine

**UHPLC:** Ultra High Performance Liquid Chromatography

**UV:** Ultraviolet

**v/v :** volume per volume percentage concentration

**Val (V):** Valine

**Vis :** Visible

**w/v:** weigh per volume percentage concentration

**Xan:** Xanthyl

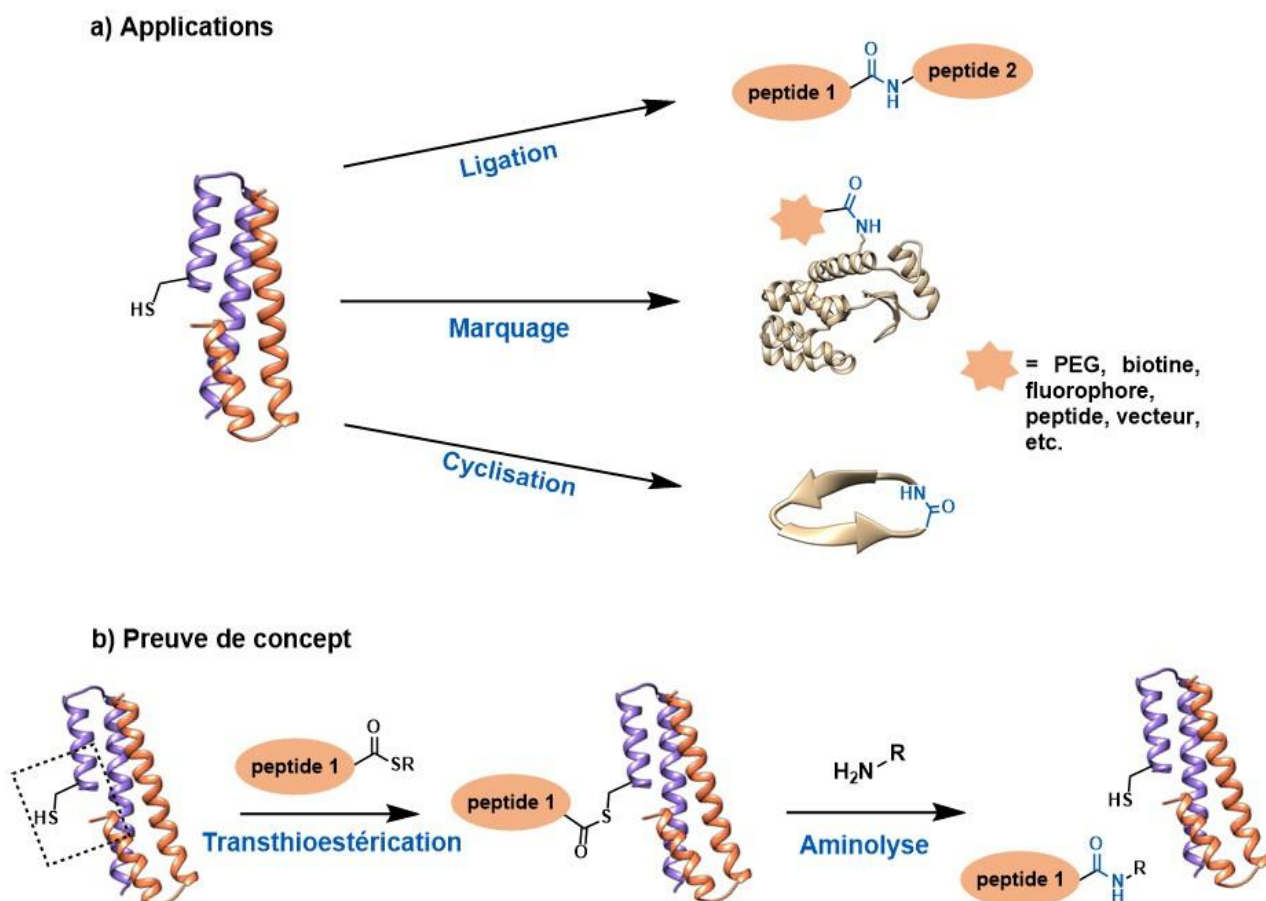
# Résumé de thèse

## 1. Introduction

La conception et les applications de nouvelles protéines non-naturelles, dites *de novo*, avec des structures réfléchies et maîtrisées n'ont de limites que notre imagination. Outre les nombreuses applications pratiques via l'incorporation de fonctions innovantes, l'étude de ces protéines nous permet aussi de mieux comprendre les principes fondamentaux qui régissent les structures et les fonctions de protéines et d'enzymes naturelles.<sup>1</sup>

Pendant ces travaux de thèse, nous avons introduit une activité catalytique innovante à partir de principes rationnels sur un châssis protéique *de novo*, c'est-à-dire que la structure protéique et la fonction catalytique ont entièrement été élaborées et construites au laboratoire. L'avantage de l'utilisation de protéines *de novo* est notamment de pouvoir facilement contrôler l'effet de chaque acide aminé sur le repliement, la stabilité et la fonction. De nombreux exemples ont démontré la robustesse et la modularité de cette approche ce qui a permis la création de nouveaux catalyseurs protéiques entre autres pour des réactions d'oxydoréduction et d'hydrolyse.<sup>2,3</sup>

Dans notre cas, nous nous sommes intéressés à la catalyse de la ligation entre peptides et plus largement de la formation de liaisons amide qui se fait via des réactions d'addition/élimination sur des thioesters. Malgré l'ubiquité de la fonction amide dans les médicaments, les polymères synthétiques et les biomolécules telles que les protéines, la formation de liaisons amide demeure un défi synthétique, et des progrès en termes de catalyse, de réduction de déchets et de chimiosélectivité peuvent effectivement être accomplis.<sup>4</sup> Dans ce contexte, le développement d'une protéine *de novo* catalysant les liaisons amide rassemblerait tous les prérequis mentionnés ci-dessus. Pour cela, nous avons introduit un résidu catalytique, une cystéine, à la surface d'une protéine *de novo* sans incorporer de groupes pour l'interaction avec les substrats dans le but de favoriser, dans un premier temps, une catalyse modérée sans restriction de substrats. La conception d'une telle protéine catalytique pour la formation de liaisons amide représenterait ainsi un nouvel outil universel pour la synthèse chimique de protéines (ligation et cyclisation de peptides) mais aussi pour le marquage de protéines (**Figure 1a**).



**Figure 1. Potentielles applications et principes de notre protéine catalytique.** a) La catalyse de la formation de liaisons amide trouve de multiples applications telles que la ligation et la cyclisation de peptides et le marquage de protéines. b) Nous avons conçu une protéine *de novo* qui porte une fonction de catalyse prédéfinie pour les réactions de transfert de groupements acyles. Une première étape de transthioestérification entre un donneur d'acyle et la cystéine catalytique forme un adduit covalent qui pourra ensuite réagir avec un second substrat, l'accepteur d'acyle. L'encadré noir en pointillé marque la position du site actif.

Le but de cette thèse était donc de confirmer la faisabilité de ce concept en démontrant que notre protéine *de novo* puisse induire une accélération de la vitesse de réaction entre un peptide- $\alpha$ thioester et un accepteur d'acyle, principalement des amines primaires (**Figure 1b**). Un modèle de protéine catalytique a d'abord été réfléchi et conçu de manière rationnelle et minimaliste afin de vérifier qu'un minimum d'éléments fonctionnels pouvait être suffisant pour installer une activité catalytique. Nous avons ensuite développé plusieurs méthodes de synthèse pour cette protéine. L'optimisation de la synthèse chimique a fait partie intégrante du processus de développement de la protéine *de novo*. Après avoir caractérisé les propriétés biophysiques et structurales de la protéine, nous avons élaboré un test catalytique qui a permis de mieux comprendre l'activité de la protéine *de novo*. Nous nous sommes ensuite intéressés à la synthèse de librairie d'analogues afin de vérifier l'influence des résidus voisins à la cystéine catalytique sur les paramètres cinétiques. En se

reposant sur de l'intuition et de la modélisation sur ordinateur, une librairie de vingt analogues a été imaginée et synthétisée via une approche chimique et convergente. En parallèle, la synthèse de séries de substrats sous forme de donneurs d'acyle ou d'accepteurs d'acyle a été conduite afin de caractériser dans le futur l'activité catalytique et la polyvalence de substrats de manière plus détaillée. Finalement, nous avons commencé à explorer de nouvelles pistes pour faire évoluer notre protéine catalytique qui restent à approfondir dans le futur. D'un côté, un analogue allongé a été conçu grâce à de multiples outils de modélisation sur ordinateur afin d'améliorer la stabilité et l'activité de ce dernier. D'un autre côté, nous avons émis l'hypothèse de l'émergence d'une potentielle nouvelle activité protéolytique grâce à l'oligomérisation d'un analogue.

Dans cette thèse, nous avons reporté l'achèvement de la première étape de ce long projet en construisant une protéine *de novo* avec une activité catalytique prédéterminée de transfert de groupements acyles. Cela représente un point de départ prometteur pour le développement et l'amélioration de l'activité catalytique via un processus de cycles itératifs entre étape de modélisation, de synthèse et de caractérisation des analogues protéiques.

## 2. Résultats et discussions

**Mécanisme et conception de la protéine catalytique.** La **Figure 2** représente les principales étapes du mécanisme supposé pour la catalyse de transfert de groupements acyles (ici, la ligation de peptides) par notre protéine *de novo*. Une réaction de transthioestérification aurait d'abord lieu entre un peptide- $\alpha$ thioester, le substrat donneur d'acyle, et la cystéine catalytique. Après être passé par un premier intermédiaire tétraédrique, le thioester ramifié peptide-protéine résultant réagirait avec un H<sub>2</sub>N-peptide, le substrat accepteur d'acyle. La dissociation du second intermédiaire tétraédrique libérerait le produit de la ligation peptidique et la protéine catalytique libre. Les deux intermédiaires chargés négativement dans ce mécanisme à double déplacement, aussi appelé ping-pong, peuvent être stabilisés par des groupes N-H des liaisons amide, stratégie similaire aux trous oxyanion trouvés dans les protéases à sérine et à cystéine.<sup>5</sup>

Précédemment, il a été déduit que le résidu phosphosérine (pSer) pouvait stabiliser les hélices  $\alpha$  lorsque celui-ci est introduit au niveau de l'extrémité N-terminale.<sup>6</sup> Cette stabilisation a pour origine la formation de liaisons hydrogène entre les NHs non appariés des liaisons peptidiques et le groupe phosphate qui est chargé négativement. En fait, les intermédiaires tétraédriques des étapes d'addition-élimination du mécanisme catalytique représenté **Figure 2** sont aussi chargés négativement et sont structurellement semblables

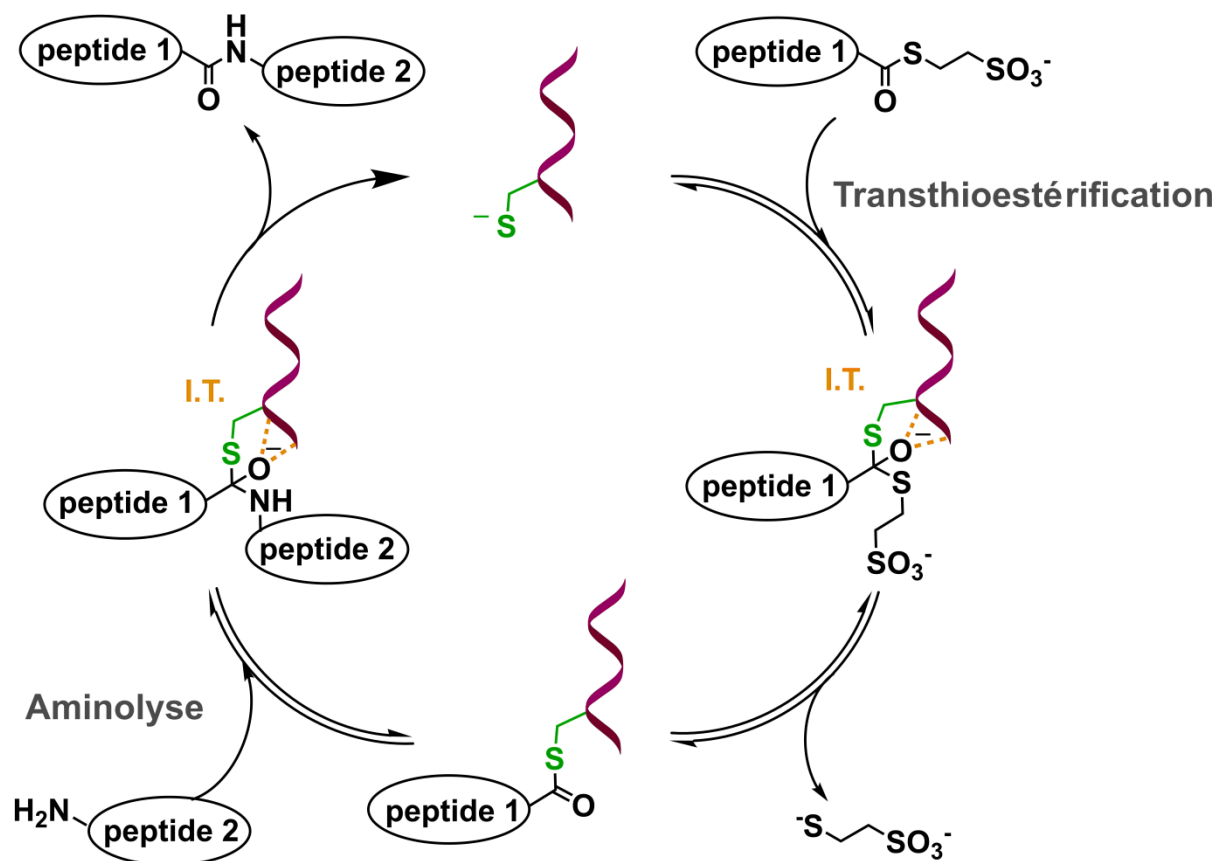
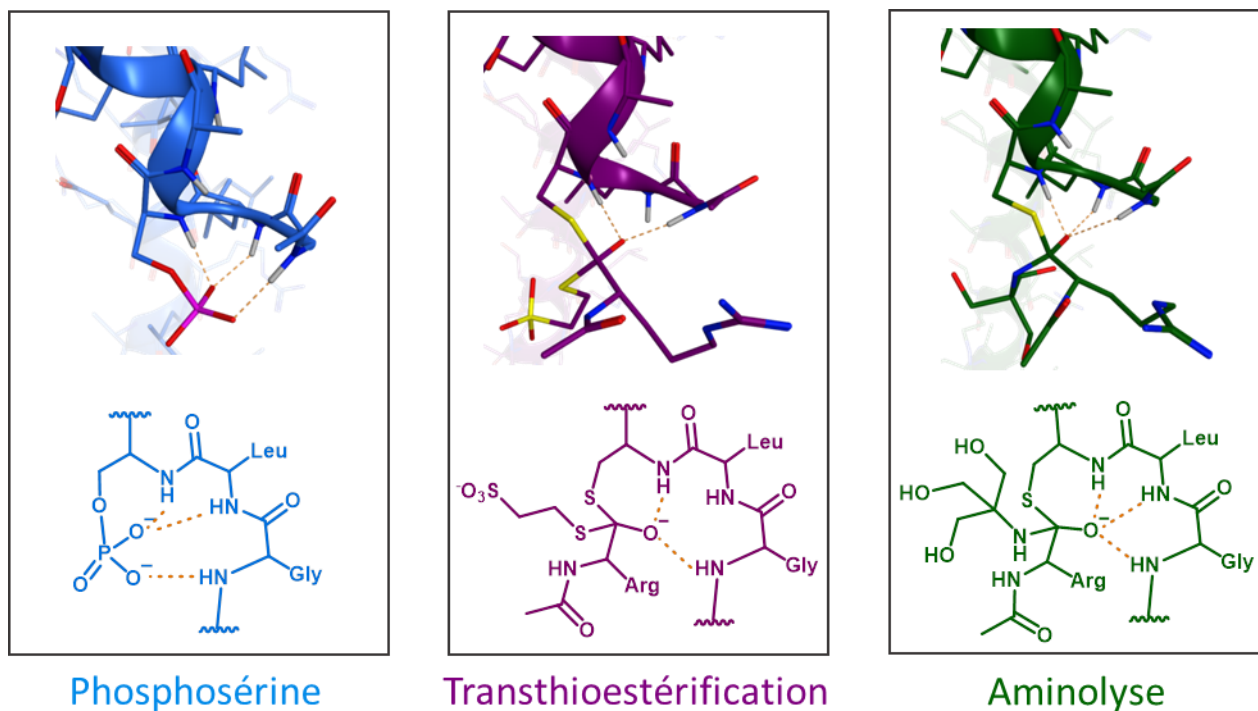


Figure 2. Mécanisme putatif pour la catalyse de transfert de groupements acyles (ici, la ligation de peptides) en présence de la protéine catalytique *de novo*. I.T. fait référence aux intermédiaires tétraédriques où l'oxyanion chargé négativement est stabilisé grâce à un réseau de liaisons hydrogène (pointillés oranges).

au résidu pSer (voir les modélisations moléculaires sur la **Figure 3**). Ainsi, l'extrémité N-terminale d'une hélice  $\alpha$  représente un motif structural approprié pour la stabilisation des intermédiaires tétraédriques.

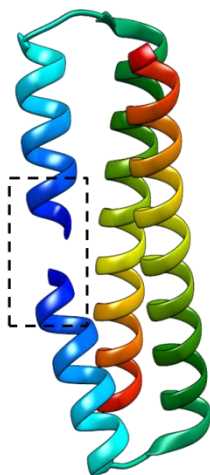
C'est pourquoi nous avons sélectionné la protéine « Domain-Swapped Dimer » (DSD) précédemment conçue par le laboratoire de DeGrado et qui est composée de deux séquences identiques dimérisant spontanément pour former un faisceau de trois hélices  $\alpha$  (**Figure 4**).<sup>7</sup> Dans ce châssis protéique, nous avons exploré l'extrémité N-terminale de l'hélice  $\alpha$  contenant des groupes NH donateurs de liaisons hydrogène non appariés comme motif structural pouvant stabiliser les intermédiaires tétraédriques chargés négativement. De plus cette protéine contient ces deux extrémités N-terminales l'une en face de l'autre créant ainsi une cavité appropriée pour l'installation d'une activité catalytique. En se basant sur l'étude analysant la phosphosérine dans les hélices  $\alpha$ , nous avons privilégié la position N2 (deuxième résidu à l'extrémité N-terminale dans la conformation hélice  $\alpha$ ) pour placer notre cystéine catalytique.<sup>6</sup> A partir de la structure DSD de départ, les lysines ont été remplacées par des résidus arginines (chargés positivement mais non nucléophiles) et un homodimère





**Figure 3. Modélisation moléculaire de la phosphosérine et des intermédiaires tétraédriques à la position N2 de la protéine DSD.** Les liaisons hydrogène stabilisantes sont montrées en pointillés oranges et ont été détectées pour chaque oxyanion et les groupes NHs non appariés de l'extrémité N-terminale. L'étape d'aminolyse a été modélisée avec du tris(hydroxyméthyl)aminométhane (Tris). Le logiciel MOE a été utilisé pour modéliser ces trois structures et les liaisons H ont été générées pendant la minimisation utilisant Amber 10 comme champ de force.

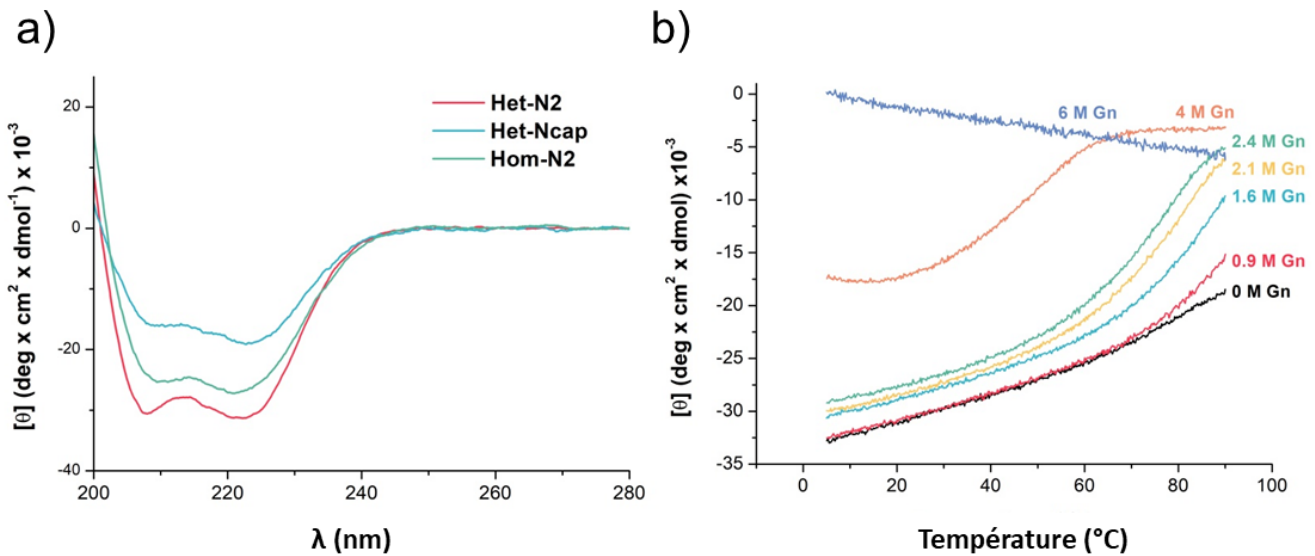
**Hom-N2** a été conçu avec deux résidus cystéine incorporés à la position N2 de chaque monomère. Simultanément, un hétérodimère a été généré : les deux brins ont été différenciés, l'un chargé négativement et l'autre positivement pour favoriser l'hétérodimérisation. La cystéine catalytique a été introduite en position N2 du brin négatif donnant l'analogue **Het-N2**. En outre, deux principales structures servant de contrôles ont été conçues : *i) Het-Ncap* correspond au même hétérodimère mais la cystéine se trouve en position Ncap (résidu précédant le premier acide aminé qui adopte une conformation en hélice  $\alpha$  : c'est aussi le premier acide aminé de la séquence), et nous permettra de vérifier l'importance de la position de l'acide aminé catalytique dans la structure; *ii) tronq-Cys-N2* fait référence à l'analogue tronqué de la protéine **Het-N2** correspondant aux douze premiers résidus N-terminaux du brin contenant la cystéine, et nous permettra d'évaluer la pertinence de la structure entière pour l'activité catalytique.



**Figure 4. Structure de la protéine DSD.** L'homodimère est représenté avec une coloration arc-en-ciel allant du bleu au rouge dans la direction N-terminale vers C-terminale. Le rectangle en pointillé noir illustre la localisation du site actif au niveau de la cavité formée par les deux extrémités N-terminales.

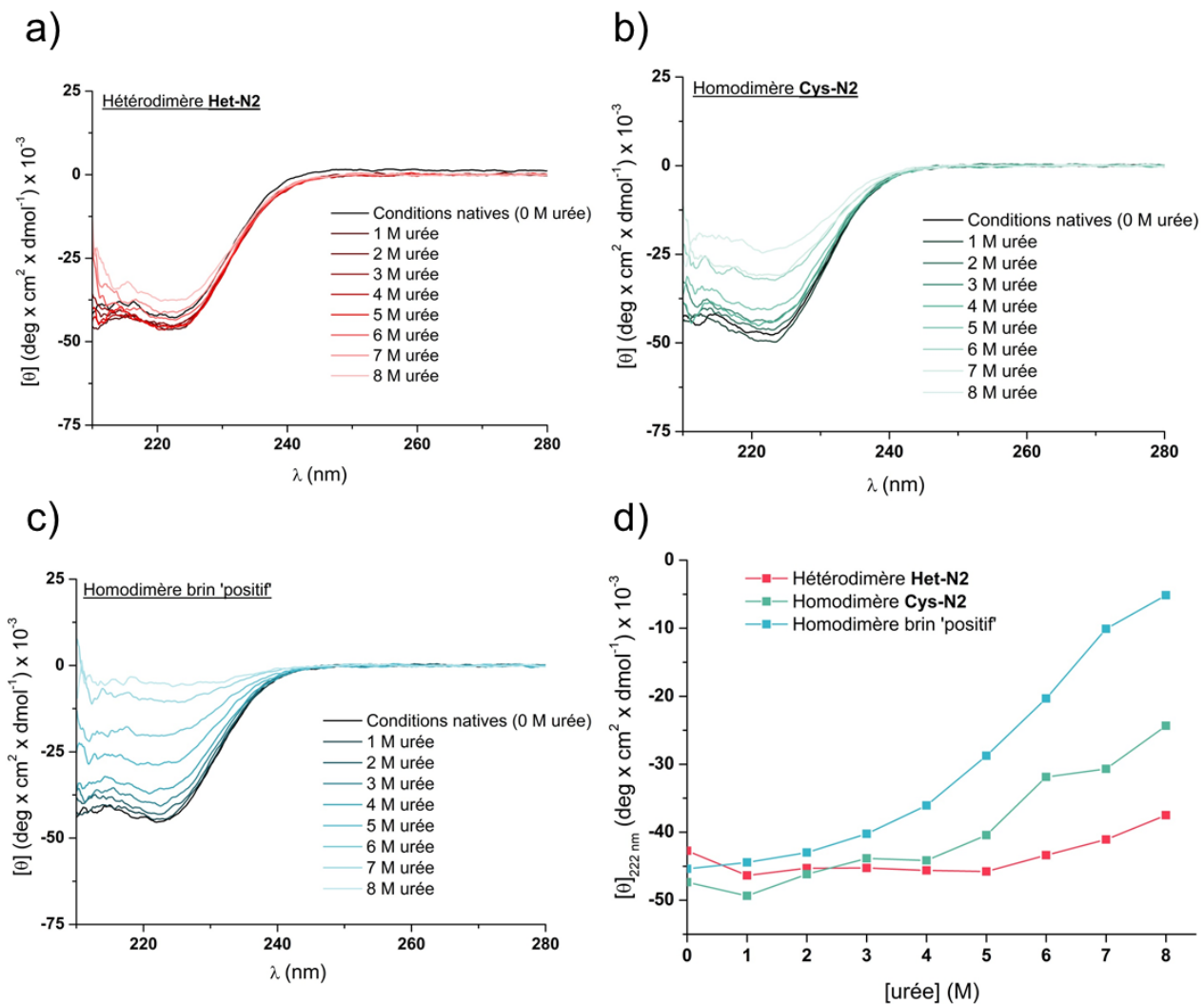
**Synthèse chimique.** Différentes approches synthétiques ont été élaborées pour la production des monomères des analogues DSD, jouant alors sur des avantages distincts. La première méthode est basée sur la Fmoc-SPPS (synthèse peptidique sur phase solide) et bénéficie d'un accès aux fragments peptidiques qui composent un monomère via l'utilisation d'un synthétiseur automatique (qui fonctionne uniquement par Fmoc-SPPS). Les fragments sont ensuite assemblés par ligation chimique native (LCN). La deuxième méthode basée sur la synthèse du monomère entier par assemblage manuel en Boc-SPPS est la plus utile pour produire efficacement des dizaines de milligrammes en utilisant une seule étape de purification CLHP. La troisième méthode s'appuie sur une stratégie de ligation à trois segments peptidiques par ajout séquentiel des segments (de l'extrémité C-terminale à N-terminale) dans le même milieu réactionnel (c'est-à-dire sans purification intermédiaire). Cette dernière est la plus appropriée pour introduire de la diversité dans la séquence des analogues grâce à une approche combinatoire par ligation peptidique.

**Caractérisations biophysiques et structurales.** L'assemblage et le repliement en hélices  $\alpha$  de l'analogue **Het-N2** ainsi que l'homodimère **Hom-N2** et le contrôle **Het-Ncap** dans un tampon phosphate et à pH neutre ont pu être confirmés par spectroscopie DC (dichroïsme circulaire, spectres **Figure 5a**). Il est à noter que l'intensité de l'ellipticité moyenne par résidu observée sur les spectres ne peut pas être prise en compte car les spectres n'ont pas été enregistrés sur la même machine et que seule la concentration de la protéine **Het-N2** a pu être calculée de manière précise, ce qui permet en général une interprétation plus exacte des mesures. Les autres analyses biophysiques, plus poussées, ont été majoritairement réalisées avec la protéine **Het-N2** qui est supposée posséder l'activité catalytique la plus efficace. En conditions natives, la structure de la protéine **Het-N2** reste très stable entre 5°C et 90°C d'après la mesure de l'ellipticité à 222 nm, indicateur



**Figure 5. Dichroïsme circulaire des différents analogues et courbes de dénaturation de Het-N2.** a) Les trois analogues montrent des spectres typiques des structures hélicoïdales. b) Courbes de dénaturation de **Het-N2** mesurées par DC à 222 nm en présence de multiples concentrations en chlorure de guanidinium (Gn-HCl).

d'une structure en hélice  $\alpha$  (Figure 5b). Jusqu'à 2 M de dénaturant (chlorure de guanidinium) dans le tampon, cette protéine retient un fort caractère hélicoïdal à température ambiante, confirmant ainsi la stabilité intrinsèque du châssis protéique qui pourra par la suite être optimisée. Grâce à l'ultracentrifugation analytique, il a été démontré que la protéine **Het-N2** était sous forme de dimère en solution lorsque les deux brins la constituant étaient mélangés. Pour confirmer que le dimère détecté dans l'échantillon de **Het-N2** lors de l'expérience de vitesse de sédimentation correspond bien à l'hétérodimère attendu et non à un mélange d'homodimères (les brins s'associant séparément), nous avons réalisé des expériences de dénaturation en présence d'urée. Ici, l'urée a été préférée à la guanidine car cette dernière déstabiliserait trop les interactions électrostatiques essentielles à la formation de l'hétérodimère à cause de son fort pouvoir ionique. En mesurant les courbes de dénaturation par DC de l'hétérodimère **Het-N2** ainsi que des homodimères correspondant à chaque monomère (**Cys-N2** pour le brin contenant la cystéine et de l'autre côté, le brin 'positif'), nous avons prouvé que l'hétérodimère était approximativement 10 fois plus stable que l'homodimère **Cys-N2** et 100 fois plus stable que l'homodimère du brin 'positif' (Figure 6). De plus, d'après les données de spectrométrie de masse en conditions natives, l'espèce hétérodimérique a été majoritairement observée avec très légère proportion d'homodimères (< 5 %).

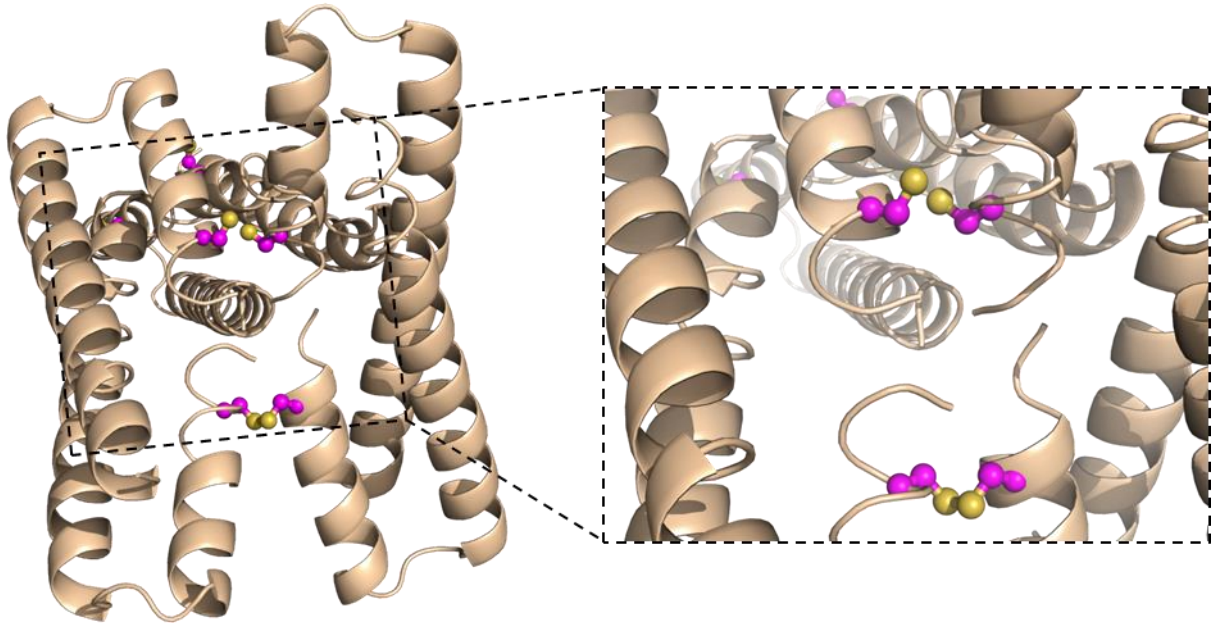


**Figure 6. Expérience de dénaturation par l'urée. a-c)** Les spectres de dichroïsme circulaire montre une diminution de la structure en hélice  $\alpha$  avec l'augmentation de la concentration en urée. **d)** Comparaison des courbes de dénaturation à 222 nm engendrées par la hausse en urée dans la solution.

**Caractérisations structurales.** En collaboration avec le Centre de Biologie Intégrative de l'Institut de génétique et de biologie moléculaire et cellulaire (IGBMC), les trois protéines **Het-N2** et **Het-Ncap** ainsi que l'homodimère **Hom-N2** ont été cristallisées. Par diffraction des rayons X, les structures ont été déterminées avec de bonnes résolutions, et le problème de phase a été résolu par remplacement moléculaire avec la structure de la protéine originale DSD (code PDB : 1G6U).<sup>7</sup> Les trois analogues synthétiques sont bien repliés et structurés formant le dimère désiré (homodimère pour **Hom-N2** et hétérodimères pour **Het-N2** et **Het-Ncap**) avec une structure globale formant un faisceau de trois hélices  $\alpha$ .

Globalement, la structure de l'homodimère **Hom-N2** se superpose bien avec la structure cristallographique de la protéine originale DSD avec seulement quelques variations

au niveau du squelette protéique résultant à un RMSD calculé par rapport aux C $\alpha$  de 1.280Å. De manière intéressante, deux copies de la protéine sont covalamment liées par un pont disulfure à cause de l'oxydation entre deux cystéines d'homodimères distincts. Par conséquence, quatre extrémités N-terminales se retrouvent dans un environnement proche formant alors une cavité avec deux cystéines non oxydées l'une en face de l'autre (**Figure 7**).

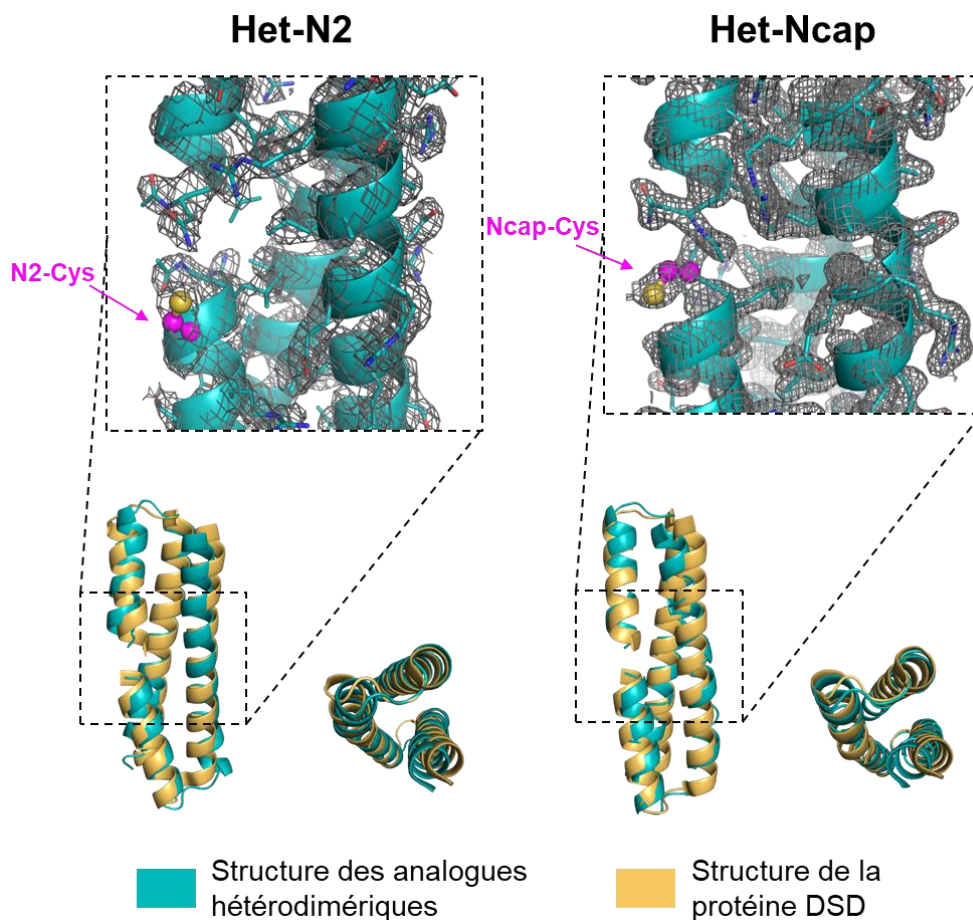


**Figure 7. Structure cristallographique de la protéine Hom-N2.** L'unité asymétrique est composée de trois copies de la protéine dont deux sont covalamment liées par un pont disulfure. Par conséquence, quatre extrémités N-terminales se retrouvent dans un environnement proche comme le montre l'image de droite (correspondant à un zoom au niveau de l'encadré en pointillé sur l'unité asymétrique à gauche). Les cystéines catalytiques sont représentées en magenta et les atomes de soufre en jaune.

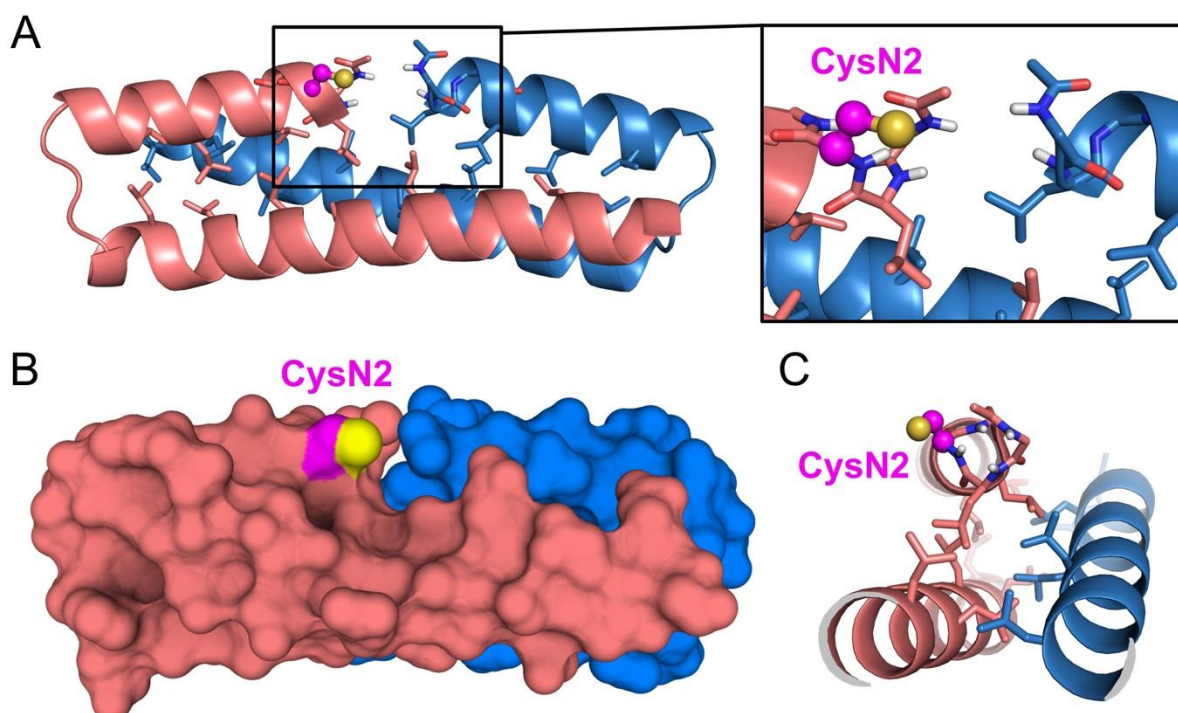
D'autre part, en comparaison avec la structure de départ de la protéine DSD, seuls des petits changements dans le squelette protéique ont été observés pour chaque hétérodimères **Het-N2** et **Het-Ncap** (en moyenne, RMSD par rapport aux carbones  $\alpha$  de 1.839Å pour **Het-N2** et de 1.793Å pour **Het-Ncap**, comme supporté par la **Figure 8**). Toutes ces informations structurales sur les analogues synthétisés soutiennent que le châssis protéique de la DSD est modulaire et accepte de multiples modifications pour y insérer une activité catalytique.

Surtout, en étudiant de plus près la structure cristallographique de l'analogue **Het-N2**, il a été clairement observé que la cystéine catalytique était positionnée à proximité de la crevasse créée par les deux extrémités N-terminales (**Figure 9**). Les NHs non appariés des groupes amide sont bien localisés par rapport au résidu Cys pour agir en tant que trou oxyanion afin de stabiliser les états de transition (c'est-à-dire les intermédiaires

tétraédriques). Cette structure est en cohérence avec notre modèle et incarne un bon point de départ pour de futures optimisations de la séquence aux alentours du site catalytique et de cette cavité.



**Figure 8. Structures cristallographiques de Het-N2 (gauche) et de Het-Ncap (droite).** Un zoom au niveau des sites catalytiques révèle que les résidus cystéine (en magenta, représentation boules-bâtonnets) sont exposés au solvant. La carte de densité  $2F_o-F_c$  est montrée au niveau  $1.0\sigma$ . Les structures des hétérodimères **Het-N2** et **Het-Ncap** (en cyan) se superposent bien à la structure de la protéine DSD d'origine (en jaune).



**Figure 9. Analyse du site catalytique de l'analogue Het-N2.** Le site catalytique formé par la cystéine en position N2 (CysN2) est localisé à l'interface des deux extrémités N-terminales. **A)** Vue horizontale de l'interface hélice-hélice avec les groupes NHs non appariés pointant vers la cavité. **B)** Représentation de la surface. **C)** Vue le long de l'axe superhélicoïdal.

**Caractérisation préliminaire de l'activité catalytique.** Notre but est d'accélérer les réactions de transfert de groupements acyles sur des peptide- $\alpha$ thioesters avec divers groupements amines. Pour se faire, nous avons développé des tests catalytiques permettant de suivre l'évolution et l'accélération de ces réactions. Le mécanisme réactionnel est composé des deux étapes principales : la transthioestérification avec la formation d'un produit ramifié suivi par l'aminolyse formant la liaison amide désirée qui est en compétition avec l'hydrolyse donnant l'acide carboxylique correspondant en tant que produit secondaire (**Figure 10**). Nous avons essayé différentes possibilités pour surveiller nos réactions catalytiques (en parallèle de la réaction contrôle sans catalyseur) pour trouver le meilleur compromis entre la collecte de données quantitatives et la possibilité de suivre plusieurs conditions en même temps.

De premiers tests catalytiques ont été réalisés par suivi HPLC et ont permis d'apercevoir les tendances générales sur l'activité catalytiques de nos trois variants **Het-N2**, **Het-Ncap** et **Hom-N2**. Un substrat de type peptide- $\alpha$ thioester (0.5 mM) a été mis en présence d'une quantité équimolaire d'une des protéines catalytiques. La formation d'un adduit covalent résultant d'une réaction de transthioestérification entre le substrat et la cystéine catalytique d'un des analogues a été instantanément observée (environ 40 % formé

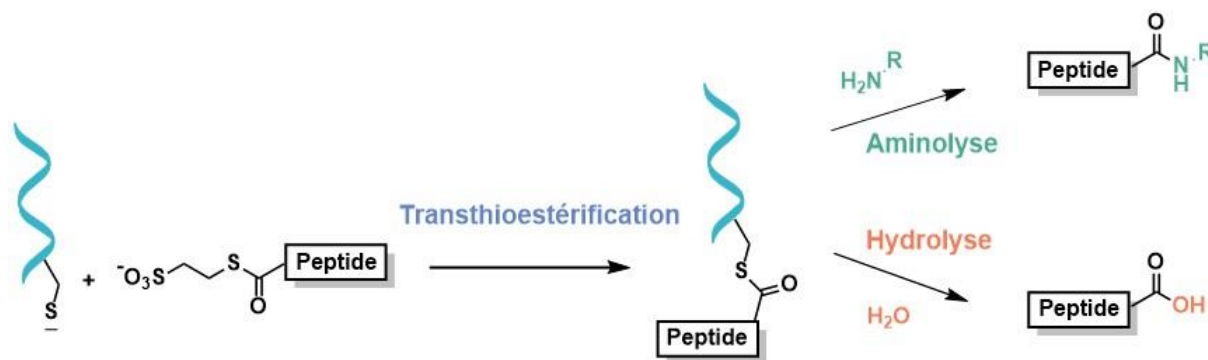


Figure 10. Voie catalytique pour la réaction de transfert de groupements acyles sur des substrats de type peptide- $\alpha$ thioester. Après une première étape de transthioestérification, l'aminolyse de l'adduit covalent sur la protéine catalytique est en compétition avec la réaction secondaire d'hydrolyse.

directement). Cet intermédiaire covalent de type Cys-thioester est généralement remarqué tout au long de la réaction, ce qui est évocateur d'un mécanisme ping-pong comme anticipé dans notre hypothèse (Figure 2). Cet adduit ramifié a été soumis au même moment à une réaction d'aminolyse ou d'hydrolyse, les deux réactions étant en compétitions. Ici, du Tris (tris(hydroxyméthyl)aminométhane) a été ajouté en large excès (250 mM) et a été adopté comme accepteur d'acyle standard car son  $pK_a$  de 8,1 est similaire à celui de l'amine N-terminale dans les peptides.<sup>8</sup> Lors de ce test préliminaire, le temps de demi-vie du substrat thioester correspondant au temps de réaction nécessaire pour consommer la moitié de la quantité initiale de peptide- $\alpha$ thioester, soit par aminolyse soit par hydrolyse, a été estimé. Ainsi, les analogues ont pu être approximativement comparés entre eux, de même qu'avec une condition contrôle sans protéine catalytique présente dans le milieu réactionnel. Les résultats de ces tests préliminaires sont récapitulés dans le Tableau 1. Sans surprise, la protéine **Het-N2** accélère très nettement la réaction et de façon plus efficace que les autres analogues **Het-Ncap** et **Hom-N2** qui présentent néanmoins de meilleurs paramètres cinétiques que la réaction sans catalyseur. La localisation du site catalytique au niveau des extrémités N-terminales semble donc être appropriée pour accélérer les transferts de groupements acyles. Cela a aussi permis de souligner l'importance de la position de la cystéine catalytique dans la séquence. Nous supposons que la plus faible activité catalytique de l'homodimère **Hom-N2** par rapport à la protéine **Het-N2** pourrait venir de l'encombrement stérique généré par la proximité des deux cystéines catalytiques, ce qui diminuerait l'accès à l'accepteur d'acyle (le Tris). L'analogue **Hom-N2** a donc été mis de côté lors des prochaines caractérisations catalytiques.



**Tableau 1. Paramètres cinétiques pour la réaction aminolyse en compétition avec l'hydrolyse sur un peptide- $\alpha$ thioester.** Conditions : 0.5 mM peptide- $\alpha$ thioester, 250 mM Tris, avec ou sans 0.5 mM de protéine catalytique

Conditions / protéine catalytique	$\tau_{1/2}$ (heures)	$k_{obs}$ ( $M^{-1} \cdot s^{-1}$ )
Réaction sans catalyseur	77.5	0.007
<b>Het-N2</b>	0.75	0.74
<b>Het-Ncap</b>	10	0.056
<b>Hom-N2</b>	8	0.069

Cependant, les résultats de ces tests préliminaires sont approximatifs et ne permettent pas de connaître le ratio d'aminolyse par rapport à l'hydrolyse. C'est pourquoi un suivi par LC/MS a été optimisé ce qui a permis d'avoir accès aux changements relatifs de concentrations entre le substrat peptide- $\alpha$ thioester et les produits d'aminolyse et d'hydrolyse dans le temps. De plus, les conditions précédentes n'étaient pas compatibles avec le criblage de nombreuses conditions et de nombreux substrats car la quantité nécessaire de protéines catalytiques était très importante et que le temps d'analyse était trop important. Il est à noter que ces conditions (notamment à cause des plus fortes concentrations en réactifs comparées aux prochaines conditions) entraînent des cinétiques plus rapides et ne sont donc pas comparables avec les résultats présentés ci-dessous. En effet, nous avons privilégié des conditions permettant de cribler un nombre important de tests catalytiques en réduisant la quantité nécessaire des réactifs.

**Activité d'acyle transférase de la protéine Het-N2.** Suite à l'optimisation d'un nouveau test catalytique, l'évaluation de l'activité catalytique de l'analogue **Het-N2** (100 $\mu$ M) et la comparaison aux contrôles (**Het-Ncap** et **tronq-Cys-N2**) ont été réalisées plus précisément avec un substrat thioester de type peptide- $\alpha$ Mes (200  $\mu$ M) (Mes = 2-mercaptoéthanesulfonate), comme donneur d'acyle, et en présence d'un large excès d'un accepteur d'acyle (200 mM) dans un tampon de phosphate de sodium à 50 mM, à pH 7,5. A noter que le Tris a été adopté comme accepteur d'acyle standard pour son  $pK_a$  similaire à celui de l'amine N-terminale dans les peptides. Les changements relatifs dans le temps des différentes concentrations ont été suivis via l'acquisition de spectres de chromatographie en phase liquide couplée à la spectrométrie de masse et en intégrant les ions détectés correspondants. La réaction contrôle (sans catalyseur) a aussi été enregistrée pour chaque condition. Ce nouveau test optimisé nous a permis de déduire des paramètres cinétiques

pour comparer la réactivité entre les diverses conditions comprenant le ratio aminolyse sur hydrolyse (A/H) et la constante de vitesse observée ( $k_{\text{obs}}$ ).

En présence de protéines catalytiques, l'adduit ramifié est instantanément observé et résulte de la réaction de transthioestérification entre le substrat thioester et la cystéine catalytique de l'analogue **Het-N2**, **Het-Ncap** ou **tronq-Cys-N2** avec libération du 2-mercaptoéthanesulfonate (Mes). Cet intermédiaire covalent de type Cys-thioester est généralement remarqué tout au long de la réaction, ce qui est évocateur d'un mécanisme ping-pong comme anticipé dans notre hypothèse (Figure 2). L'hétérodimère **Het-N2** avec la cystéine en position N2 est effectivement l'analogue qui entraîne la plus forte accélération de la réaction. Les contrôles (**Het-Ncap** et **tronq-Cys-N2**) n'influencent quasiment pas la vitesse de la réaction avec une accélération aux alentours de 1,5 fois comparé à la réaction sans protéine catalytique (Figure 11). Ceci a permis de démontrer l'importance des points suivants : *i*) la position de la cystéine influence les cinétiques (**Het-Ncap**), les intermédiaires tétraédriques étant certainement plus ou moins bien stabilisés ; *ii*) la structure complète de la protéine est nécessaire pour l'efficacité catalytique (**tronq-Cys-N2**).

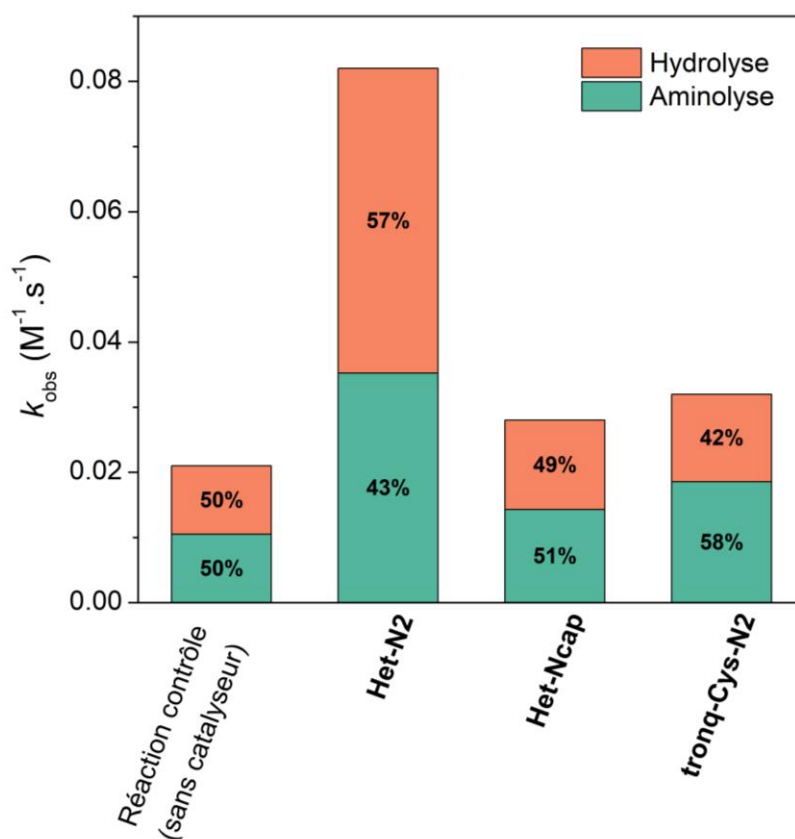
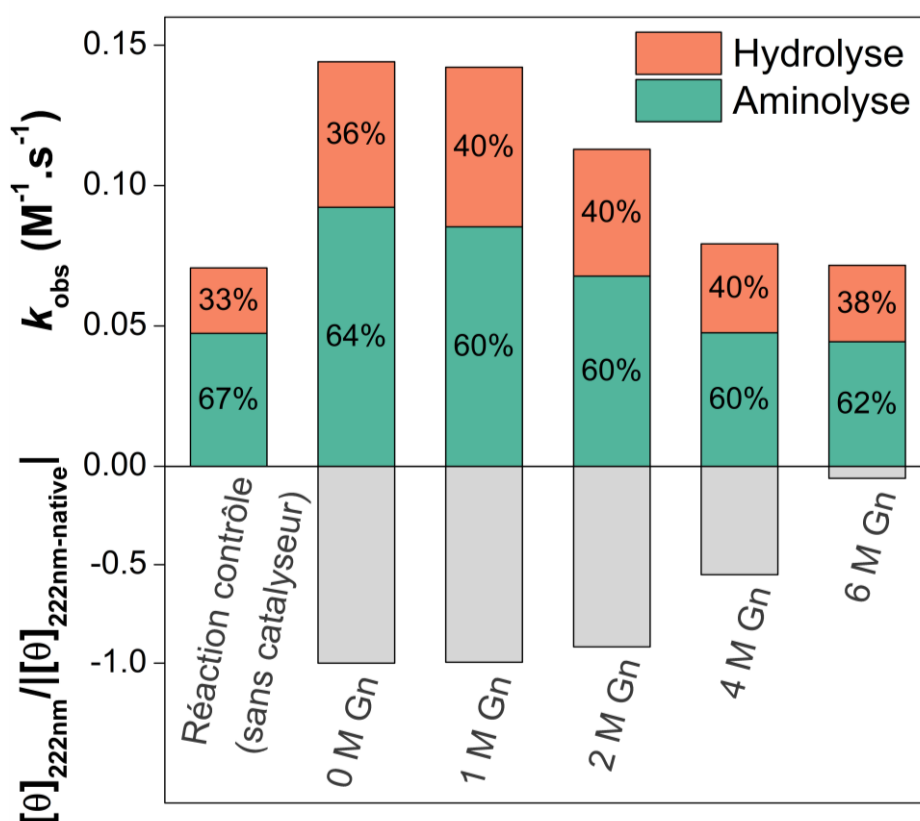


Figure 11. Constante de vitesse observée ( $k_{\text{obs}}$ , calculée à partir de la diminution du peptide-Phe-<sup>35</sup>thioester) et pourcentages d'aminolyse (avec du Tris) et d'hydrolyse du substrat thioester en absence ou en présence de protéines catalytiques. La protéine **Het-N2** catalyse le plus efficacement la réaction de transfert d'acyles.

Afin de prouver de manière plus claire que la structure et le repliement global de **Het-N2** sont nécessaires à son activité catalytique, un même test catalytique a été réalisé avec de multiples concentrations de dénaturant (chlorure de guanidinium) dans le milieu (**Figure 12**). A faible concentration de guanidine, où la protéine catalytique **Het-N2** est toujours bien structurée d'après les mesures de DC, les vitesses de réaction sont similaires aux conditions natives (sans dénaturant). Cependant, à de fortes concentrations en guanidine (6M) qui entraînent une déstructuration complète de la protéine, la vitesse de réaction diminue jusqu'à atteindre la vitesse de la réaction contrôle (sans catalyseur), prouvant que la structure de **Het-N2** est essentielle pour la catalyse. Ce résultat a permis de valider notre hypothèse : le châssis protéique basé sur la protéine DSD est primordial pour la catalyse en rassemblant les éléments cruciaux au niveau de la cavité formée par les deux extrémités N-terminales et qui favorisent ainsi la catalyse.



**Figure 12.** Constante de vitesse observée ( $k_{obs}$ , calculée à partir de la diminution du peptide-Arg<sup>6</sup>thioester) et pourcentages d'aminolyse (avec du Tris) et d'hydrolyse du substrat thioester en présence de **Het-N2**. La même réaction a été réalisée à de multiples concentrations de dénaturant (chlorure de guanidinium, Gn-HCl) et donc à différents degrés de repliement de **Het-N2** selon les mesures de DC à 222nm. A forte concentration de guanidine, la protéine **Het-N2** se déstructure progressivement ce qui entraîne une baisse de la vitesse de réaction ( $k_{obs}$ ) jusqu'à atteindre la vitesse de la réaction contrôle sans catalyseur.

**Champ de réactivité de la protéine Het-N2.** Nous avons finalement pu observer que la protéine catalytique **Het-N2** était capable d'accélérer de nombreuses réactions de transfert de groupes acyles. Même si les paramètres cinétiques sont propres à chaque réaction, la protéine **Het-N2** accélère à chaque fois la vitesse de la réaction indépendamment de la nature du peptide- $\alpha$ thioester et de l'accepteur de groupe acyle (nucléophile) (voir le **Tableau 2** pour un récapitulatif des réactions testées). Sans modifier le ratio d'aminolyse sur hydrolyse par rapport à la réaction sans catalyseur, la protéine catalytique **Het-N2** accélère la vitesse de réaction entre différents thioesters et du Tris. Des changements au niveau des paramètres cinétiques ont été observés en fonction de la nature du résidu à l'extrémité C-terminale du thioester. De même, la catalyse de la réaction de transferts d'acyles en présence des nucléophiles distincts (plus particulièrement des amines) sur un même thioester a été constatée avec l'hétérodimère **Het-N2**. Le résultat le plus remarquable est probablement l'accélération de la formation de la liaison peptidique entre un peptide- $\alpha$ thioester et un tripeptide Gly-Gly-Gly en tant que nucléophile même si une majorité d'hydrolyse a été observée

**Tableau 2. Paramètres cinétiques pour diverses réactions de transfert de groupements acyles catalysées par Het-N2.** \* Accélération par rapport à la réaction contrôle (sans catalyseur) correspondante.

Donneur d'acyle en présence de Tris	$k_{\text{obs}}$ ( $\text{M}^{-1} \cdot \text{s}^{-1}$ )	Ratio A/H	Accélération*
peptide1-Arg	0.143	1.89	2
peptide2-Arg	0.143	2.03	1.8
peptide-Phe	0.082	0.76	3.9
peptide-Gly	0.173	1.66	2.4
Accepteur d'acyle en présence de peptide1-Arg- $\alpha$ thioester	$k_{\text{obs}}$ ( $\text{M}^{-1} \cdot \text{s}^{-1}$ )	Ratio A/H	Accélération*
Tris	0.143	1.89	2
MeONH <sub>2</sub>	0.412	2.33	3.2
MeNH <sub>2</sub>	0.187	0.37	3.1
Gly	0.108	0.09	2.6
Gly-Gly	0.162	0.15	2.9
Gly-Gly-Gly	0.177	0.29	2.7
Histamine	0.693	0.48	1.2

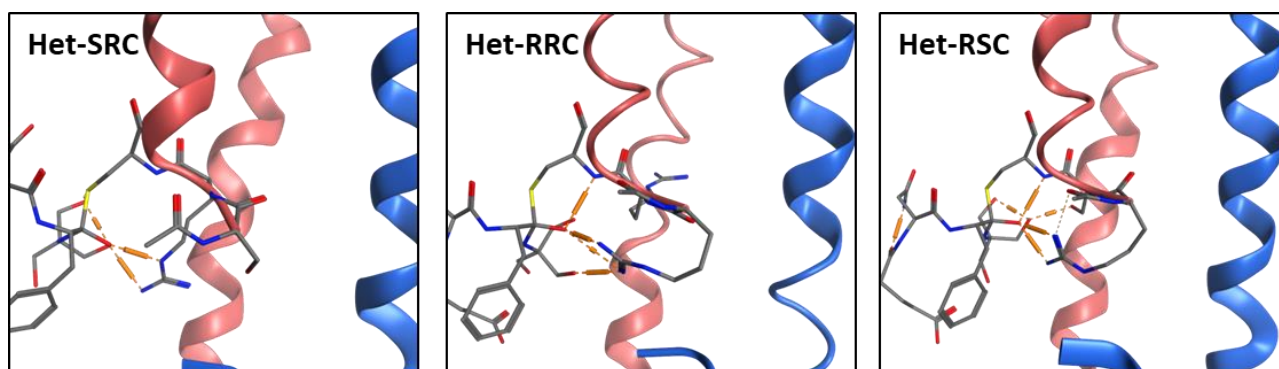
**Conception et synthèse d'une librairie de protéines catalytiques.** Pour mieux comprendre les facteurs essentiels qui influencent voire améliorent l'activité catalytique du châssis protéique DSD, une librairie d'analogues a été synthétisée. Une vingtaine de nouveaux analogues basés sur la séquence de l'hétérodimère **Het-N2**, où les acides aminés Ncap et N1 à côté de la cystéine catalytique diffèrent, ont été produits (*i.e.* librairie de brins 'négatifs'). Ils ont été nommés selon le code **Het-AA(Ncap)AA(N1)C** (avec AA(Ncap) et AA(N1) donnés par le code à une lettre de l'acide aminé correspondant) et leur séquence commune est la suivante :

### Analogues de Het-N2 :

Brin 'positif': Ac-NLAALRSELQALRREGFSPERLAALESRLQALERRLAALRSRLQALRG-NH<sub>2</sub>

Brin 'négatif' Ac-**Ncap-N1-C**ALRSELQALRREGFSPPEELAALESELQALERELAALRSELQALRG-NH<sub>2</sub>

Les acides aminés en position Ncap et N1 ont été sélectionnés pour leur potentiel à apporter des liaisons hydrogène stabilisantes complémentaires (exemple pour les analogues portant un résidu arginine en Ncap ou en N1), à mieux positionner la cystéine catalytiques (variation du résidu hydrophobique en position N1), à influencer le pK<sub>a</sub> de la cystéine ou à apporter une base ou un acide général(e) pouvant être intégré(e) dans le mécanisme réactionnel (substitution par des résidus histidine, tyrosine ou glutamate). La modélisation sur ordinateur a permis, par exemple, de visualiser les possibles interactions (liaisons H) stabilisantes entre l'oxyanion d'un intermédiaire tétraédrique et un résidu Arg placé en position Ncap ou N1 (voir **Figure 13**).



**Figure 13. Modélisation moléculaire de l'intermédiaire tétraédrique pour l'étape d'aminolyse sur les nouveaux analogues Het-SRC, Het-RRC et Het-RSC.** Les liaisons hydrogène sont représentées en orange et sont clairement visibles après modélisation entre les résidus arginines placés en position Ncap ou en N1 et l'oxygène chargé négativement. Le brin 'positif' est montré en bleu et le brin 'négatif' en rose saumon.

Pour produire la librairie de brins 'négatifs', nous avons optimisé une nouvelle procédure pour effectuer de multiples synthèses en parallèle via une approche convergente. (Figure 14). Le monomère relatif au brin 'négatif' a été divisé en deux fragments peptidiques entre le résidu N1 (c'est-à-dire le second acide aminé de la séquence) et la cystéine catalytique en position N2 (troisième acide aminé dans la séquence). Le fragment C-terminal contenant la cystéine reste inchangé pour tous les analogues et a été synthétisé en utilisant un processus d'assemblage séquentiel des acides aminés le constituant par Boc-SPPS. D'autre part, le segment peptidique N-terminal est composé de deux acides aminés (Ncap et N1) ce qui risque de poser des problèmes de solubilité lors des étapes de thioestérification et de ligation et/ou de rétention sur les colonnes en phase inverse (C18) lors de l'analyse de réactions et lors des purifications. C'est pourquoi nous avons décidé de synthétiser les différents fragments N-terminaux par le biais d'un connecteur (« linker ») de type *N*-acylurée permettant d'y coupler une séquence additionnelle (ici, LDA) pour améliorer le caractère amphiphile du produit final. Ce connecteur positionné à l'extrémité C-terminale du dipeptide désiré est aussi un bon groupe partant. Il permet ainsi de générer *in situ* un thioester en présence d'un dérivé thiophénol lors de l'étape de ligation avec le fragment commun C-terminal. La cystéine nécessaire à la ligation correspond à la cystéine catalytique en position N2. Suivant cette stratégie de synthèse, 20 analogues différant de leurs résidus Ncap et N1 ont été produits de manière efficace.

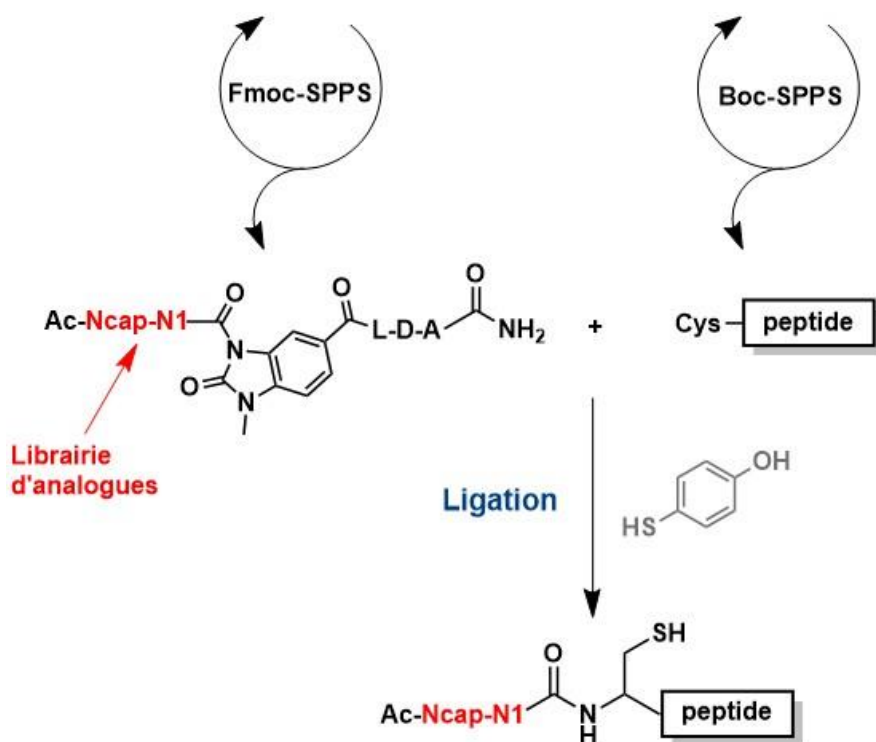


Figure 14. Stratégie de synthèse pour produire en parallèle la librairie de brins 'négatifs'.

Des tests catalytiques par suivi LC/MS ont révélé de légers changements en termes de réactivité par rapport à la protéine **Het-N2** d'origine (voir **Figure 15**). De manière surprenante, l'activité catalytique de notre standard **Het-N2** reste une des meilleures. En modifiant le résidu hydrophobique en position N1 (barres orange, vertes et rouges sur la **Figure 15**), nous avons remarqué que l'analogue **Het-GAC** avec une chaîne latérale moins hydrophobique présente une activité catalytique plus faible, alors que les autres résidus hydrophobiques entraînent des cinétiques semblables à **Het-N2**. Les acides aminés tyrosine et tryptophane portant un cycle hétéroaromatique induisent remarquablement un plus fort taux d'aminolyse qui devient alors le produit majoritaire. Au contraire, si une histidine est présente en position N1 (barres bleues claires), celle-ci favorise encore plus l'hydrolyse que la protéine **Het-N2**. De plus, les résidus arginines en position N1 ou Ncap généralement conduisent à de plus faibles vitesses de réaction (barres jaunes). Dans ce cas, nous avons simultanément constaté que la diminution en concentration de l'adduit covalent (via l'aminolyse ou l'hydrolyse) sur les analogues arginines est plus lente, suggérant ainsi une plus grande stabilisation de cet adduit ramifié ou que le centre électrophile du carbonyl et/ou le site catalytique est moins accessible pour la seconde réaction.

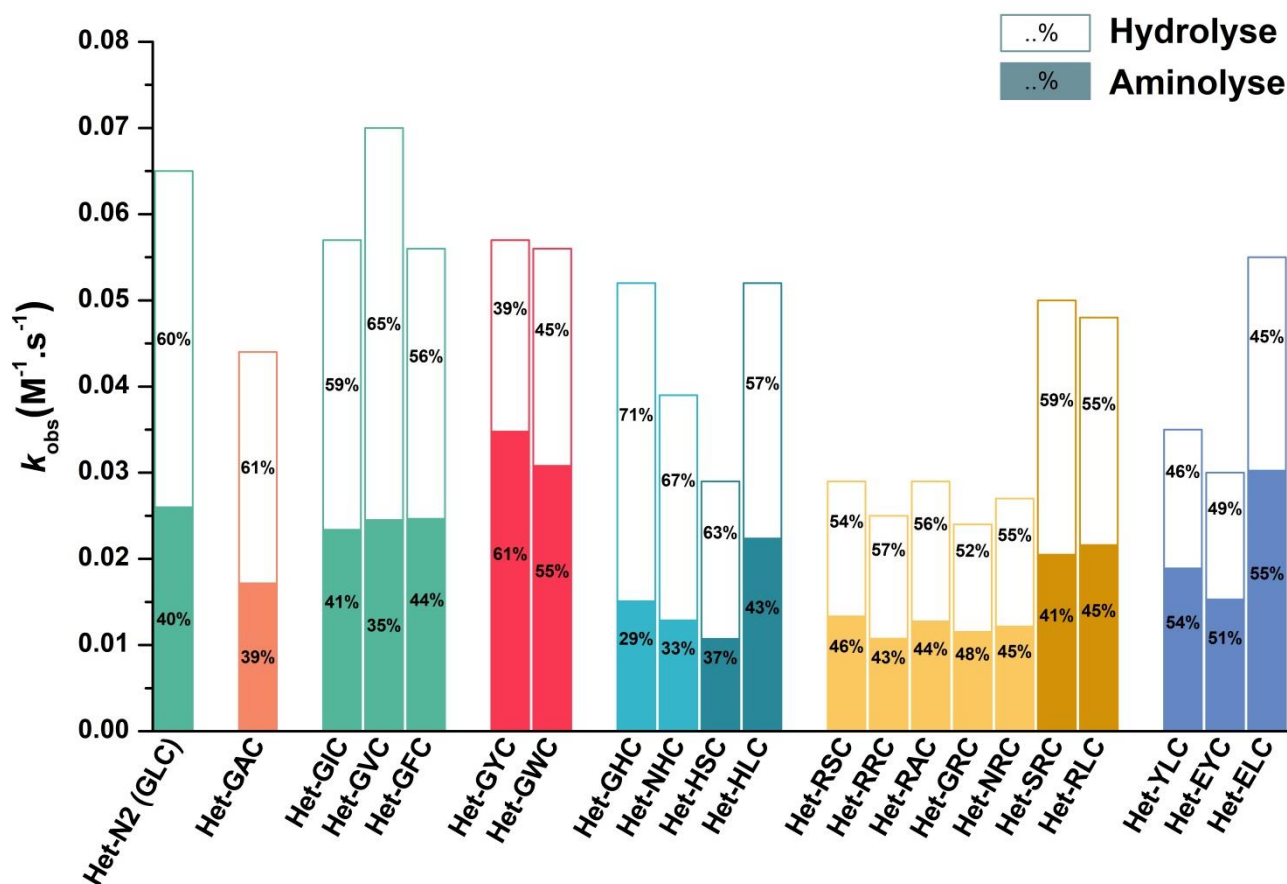
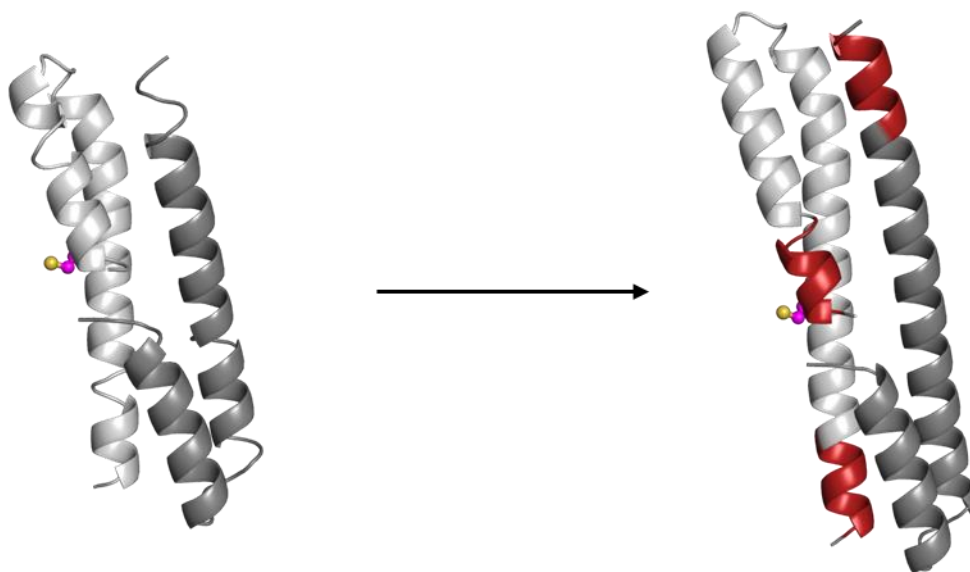


Figure 15. Constantes cinétiques et pourcentages d'aminolyse et d'hydrolyse induits par les différents analogues de la librairie d'hétérodimère contenant différentq brins 'négatif'.

**Synthèse d'une librairie de substrats.** L'impact de la nature du résidu C-terminal du substrat thioester et de la nature de l'accepteur d'acyle sur les paramètres cinétiques a déjà été démontré. Pour gagner encore plus en compréhension sur l'unique réactivité des analogues DSD, des librairies de substrats ont été synthétisées. La première librairie est composée de substrats thioesters dont le résidu C-terminal varie (20 peptides avec les différents acides aminés naturels à la dernière position). Une seconde librairie composée de peptides accepteurs d'acyles avec l'amine N-terminale libre a aussi été produite (les 20 résidus canoniques ont été placés sur la position N-terminale). Le but sera d'utiliser les différents substrats de ces deux librairies dans le test catalytique afin de déduire leur influence sur l'activité catalytique puis de retransmettre ces informations sur le châssis protéique DSD.

**Conception d'analogues allongés.** Nous avons utilisé la modélisation sur ordinateur avec notamment le logiciel Rosetta pour construire des analogues potentiellement plus stables. Le but était de pouvoir exploiter ces nouveaux analogues dans des conditions plus dénaturantes tout en gardant leur structure tridimensionnelle. Pour cela, les hélices  $\alpha$  ont été allongées (augmentation de la surface de contact) et le cœur hydrophobique a été optimisé pour y favoriser les interactions (voir **Figure 16**, les parties rouges correspondent aux heptades supplémentaires créant ainsi un châssis asymétrique avec un brin plus long que l'autre). En allongeant l'hélice contenant la cystéine catalytique, nous avons aussi avancé l'hypothèse d'une potentielle influence sur l'activité catalytique grâce à l'effet du macrodipôle de l'hélice. Il a été supposé que cet effet pouvait affecter la catalyse enzymatique en



**Figure 16. Principe de la conception de l'analogue allongé.** En partant de la structure cristallographique de **Hom-N2**, nous avons gardé uniquement une cystéine en position N2 conduisant à un hétérodimère. Les hélices ont été allongées d'un heptade chacune (représenté en rouge). La cystéine est illustrée en magenta et l'atome de soufre en jaune.



facilitant la déprotonation du résidu nucléophile en position N-terminale de l'hélice, et en stabilisant la charge négative de l'intermédiaire tétraédrique via des interactions favorables avec le champ électrique généré à l'extrémité N-terminale par le macrodipôle de l'hélice.<sup>9</sup> Le processus de conception sur ordinateur de ces nouveaux analogues a compris plusieurs étapes clés : l'obtention des paramètres de Crick définissant la superhélice de **Hom-N2** (seule structure cristallographique disponible à ce moment-là), la production d'une superhélice plus longue suivant les paramètres de la structure de **Hom-N2**, superposition de la structure cristallographique et du modèle généré, fusion des deux superhélices pour former l'analogue désiré. La séquence a ensuite été optimisée (surtout au niveau du cœur hydrophobique) avec le logiciel Rosetta et plus particulièrement avec le script « fixbb » qui permet de modifier les résidus voulus par une série d'acides aminés préalablement définie tout en gardant le squelette protéique intact. Les structures générées sont ensuite notées et les séquences apportant les meilleures stabilisations de la structure 3D sont sélectionnées. Deux séquences ont été choisies en maintenant, dans les deux cas, la cystéine catalytique en position N2 sur le brin le plus long.

Les deux protéines correspondantes ont été synthétisées en assemblant les deux monomères par SPPS et ligation chimique native. Cependant, des résultats catalytiques préliminaires pour des réactions de transfert de groupes acyles ont montré une plus faible activité des analogues conçus et allongés comparé à notre protéine standard **Het-N2**. De même, les mesures de courbes de fusion par DC en présence de fortes concentrations en dénaturant (jusqu'à 6M de guanidine) suggèrent que ces analogues allongés soient moins stables contre la dénaturation chimique en comparaison à la protéine **Het-N2** (**Figure 17**). L'acquisition de structures cristallographiques à haute résolution des protéines allongées n'a pas fourni d'informations particulières soulignant cette déstabilisation (voir **Figure 18**). En effet, la structure obtenue se superpose correctement avec la structure modèle générée par ordinateur (RMSD entre les C $\alpha$  des deux structures est de 1.77Å), ce qui montre que nous avons produit la structure voulue et modélisée par ordinateur. Pour mieux comprendre l'origine de la baisse d'activité et de stabilité, de nouveaux analogues seront à synthétiser et à analyser. Nous envisageons notamment d'implémenter les multiples changements un après l'autre : uniquement la synthèse de la structure allongée avec la séquence originale puis avec l'optimisation de la séquence. Chaque étape du processus de conception sera reprise avec des synthèses intermédiaires pour comprendre l'impact de chaque modification. Il sera par la suite possible d'orienter la modélisation vers une structure plus stable.

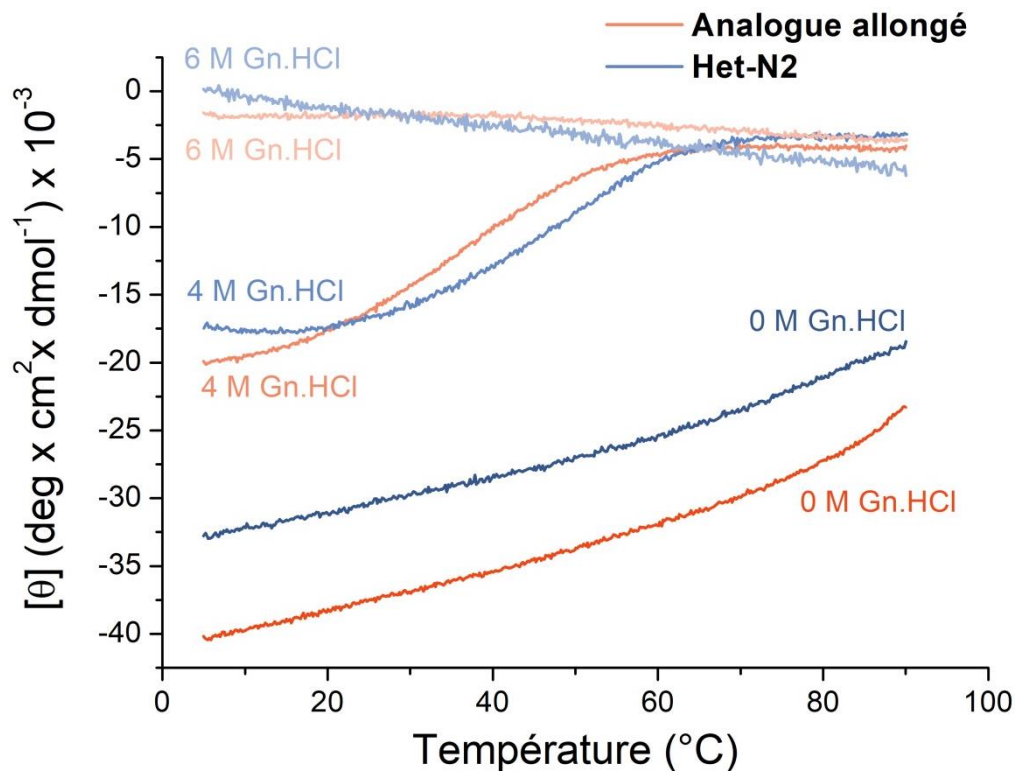


Figure 17. Courbes de dénaturation mesurées par dichroïsme circulaire à 222 nm pour l'analogue allongé (orange) et la protéine standard Het-N2 (violet) à différentes concentrations en chlorure de guanidinium (Gn-HCl).

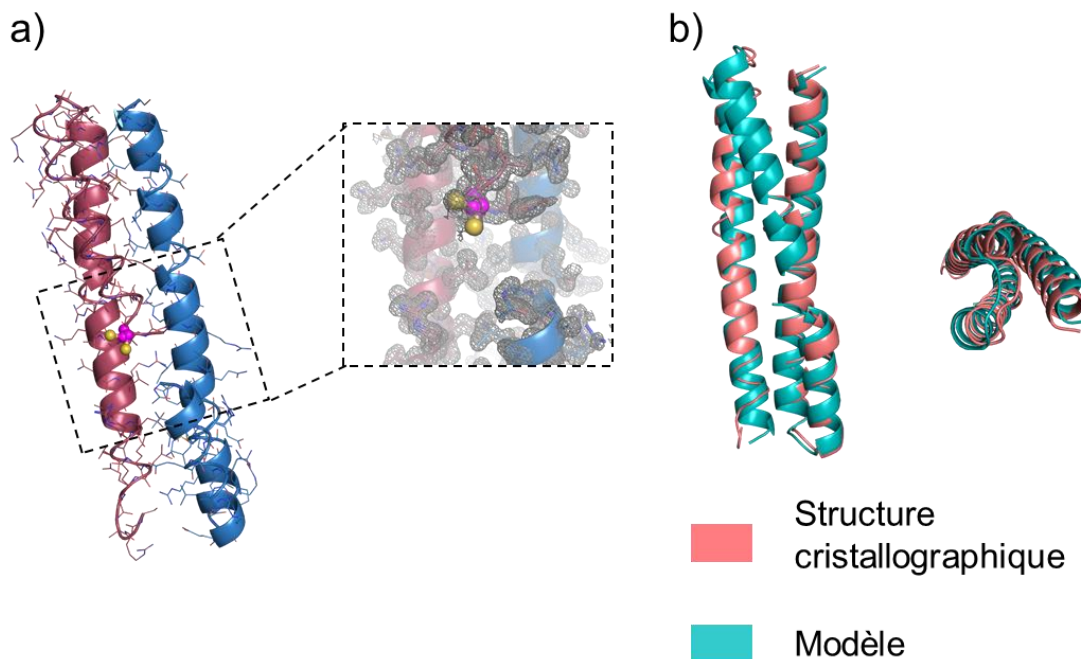


Figure 18. Structure cristallographique d'un des analogues allongés. a) Le brin le plus long contenant la cystéine catalytique est représenté en rouge, et l'autre brin en bleu. Le zoom au niveau du site catalytique (encadré en pointillé) révèle bien la cystéine (magenta en représentation boules et bâtonnets) exposée au solvant. La carte de densité  $2F_o - F_c$  est montrée au niveau  $1.0\sigma$ . b) Superposition de la structure cristallographique (en rose saumon) avec le modèle conçu sur ordinateur (cyan) avec une vue de côté sur la gauche et une vue de dessus sur la droite.

**Vers une activité protéolytique.** Grâce à la modularité démontrée de notre châssis protéique, il nous est permis d'imaginer de nouveaux arrangements afin de faire émerger de nouvelles activités catalytiques. Pour cela, la structure cristallographique de l'analogue **Hom-N2 (Figure 7)** formant un assemblage d'ordre supérieur par le biais d'un pont disulfure nous a inspiré afin d'étendre notre gamme d'applications. Il a été suggéré qu'une activité catalytique primitive pouvait se manifester au cours de l'évolution à partir de réarrangements structuraux de simples peptides formant ainsi des superstructures. En suivant ce plausible processus, nous envisageons d'atteindre de nouvelles fonctions et plus spécifiquement une activité protéolytique. Via l'oligomérisation de notre homodimère, nous espérons la formation de sites catalytiques distincts avec notamment deux cystéines catalytiques en face à face comme dans l'arrangement cristallographique. Cette exploration fera l'objet de futures expériences qui seront ensuite suivies par de la modélisation avancée sur ordinateur en partant de la structure cristallographique déjà obtenues. Si notre hypothèse s'avère valide, nous considérerons la construction de cet assemble d'ordre supérieur grâce à des liaisons plus stables (par exemple, via l'alkylation de cystéines) en comparaison aux ponts disulfures.

### 3. Conclusion générale

Pour conclure, nous avons démontré qu'une structure protéique complètement artificielle pouvait catalyser des réactions de transferts de groupements acyles sur des thioester en passant par la formation d'un lien covalent de type Cys-thioester sur cette protéine. La synthèse chimique nous a permis de rapidement obtenir différents analogues et de pouvoir facilement incorporer des modifications dans la séquence. Par conception rationnelle, nous avons introduit une unique cystéine en position N2 d'une hélice  $\alpha$  (**Het-N2**) sur un analogue de la protéine DSD. La structure cristallographique a confirmé que l'assemblage hétérodimérique de la protéine **Het-N2** avait une structure globale similaire au châssis protéique originale de la protéine DSD. Nous avons clairement établi l'importance de la position de la cystéine catalytique sur l'efficacité de l'activité. De plus, le repliement 3D de la protéine *de novo* est nécessaire à l'activité catalytique. La protéine **Het-N2** ainsi développée accélère modestement une variété de réactions de transferts de groupements d'acyles (hydrolyse et/ou aminolyse de thioester) acceptant un large spectre de substrats. Ainsi, la protéine catalytique **Het-N2** peut être comparée aux enzymes primitives (dites primordiales) possédant une efficacité catalytique modérée et acceptant plusieurs substrats.<sup>10</sup> L'optimisation d'une approche convergente pour la synthèse d'une librairie d'analogues a ensuite permis de mieux comprendre comment certains acides aminés

influençaient la réactivité de la protéine catalytique. Une caractérisation biophysique, structurale et catalytique plus détaillée est planifiée pour étudier les nouveaux analogues importants, et notamment ceux qui favorisent plus l'aminolyse. Cela nous permettra de mieux comprendre les principes fondamentaux de la catalyse enzymatique pour la formation de liaisons amide. Les bibliothèques de substrats nous permettront d'avoir une meilleure compréhension du système et nous pourrons ainsi accéder à la conception d'un analogue optimisé pour les réactions de transferts d'acyles.

Finalement, ces travaux de thèse représentent l'achèvement de la première étape de ce conséquent projet et auront permis de valider notre nouveau concept : une protéine non-naturelle, dite *de novo*, peut être utilisée pour l'accélération de la formation de liaison amide. Cette approche nous permet de mieux comprendre, étape par étape, les paramètres fondamentaux à cette catalyse. L'architecture modulaire de la protéine **Het-N2**, la modélisation sur ordinateur et la synthèse de nouvelles bibliothèques combinatoires permettront d'atteindre de plus fortes activités catalytiques en faisant varier la composition et la structure du site actif. Une approche itérative sera mise en place afin d'améliorer étape par étape et point par point notre système catalytique. Les travaux futurs seront ainsi dirigés vers l'optimisation de la séquence pour l'aminolyse de thioester afin de développer un nouvel outil pratique pour la ligation et la cyclisation de peptides ainsi que la modification de protéines.

#### 4. Références

- (1) Nanda, V.; Koder, R. L. Designing Artificial Enzymes by Intuition and Computation. *Nat Chem* **2010**, 2 (1), 15–24.
- (2) Watkins, D. W.; Jenkins, J. M. X.; Grayson, K. J.; Wood, N.; Steventon, J. W.; Le Vay, K. K.; Goodwin, M. I.; Mullen, A. S.; Bailey, H. J.; Crump, M. P.; MacMillan, F.; Mulholland, A. J.; Cameron, G.; Sessions, R. B.; Mann, S.; Anderson, J. L. R. Construction and in Vivo Assembly of a Catalytically Proficient and Hyperthermostable de Novo Enzyme. *Nat Commun* **2017**, 8 (1), 358.
- (3) Burton, A. J.; Thomson, A. R.; Dawson, W. M.; Brady, R. L.; Woolfson, D. N. Installing Hydrolytic Activity into a Completely de Novo Protein Framework. *Nat Chem* **2016**, 8 (9), 837–844.
- (4) Pattabiraman, V. R.; Bode, J. W. Rethinking Amide Bond Synthesis. *Nature* **2011**, 480 (7378), 471–479.
- (5) Simón, L.; Goodman, J. M. Enzyme Catalysis by Hydrogen Bonds: The Balance between Transition State Binding and Substrate Binding in Oxyanion Holes. *J. Org. Chem.* **2010**, 75 (6), 1831–1840.

- (6) Andrew, C. D.; Warwicker, J.; Jones, G. R.; Doig, A. J. Effect of Phosphorylation on  $\alpha$ -Helix Stability as a Function of Position. *Biochemistry* **2002**, *41* (6), 1897–1905. <https://doi.org/10.1021/bi0113216>.
- (7) Ogihara, N. L.; Ghirlanda, G.; Bryson, J. W.; Gingery, M.; DeGrado, W. F.; Eisenberg, D. Design of Three-Dimensional Domain-Swapped Dimers and Fibrous Oligomers. *Proc Natl Acad Sci USA* **2001**, *98* (4), 1404–1409.
- (8) Sievers, A.; Beringer, M.; Rodnina, M. V.; Wolfenden, R. The Ribosome as an Entropy Trap. *Proc Natl Acad Sci USA* **2004**, *101* (21), 7897–7901.
- (9) Hol, W. G. J.; van Duijnen, P. T.; Berendsen, H. J. C. The  $\alpha$ -Helix Dipole and the Properties of Proteins. *Nature* **1978**, *273* (5662), 443–446..
- (10) Khersonsky, O.; Roodveldt, C.; Tawfik, D. S. Enzyme Promiscuity: Evolutionary and Mechanistic Aspects. *Curr Op Chem Biol* **2006**, *10* (5), 498–508.



# **Chapter 1.    General introduction**

## 1.1. Functional *de novo* design: is it possible to efficiently install a enzymatic-like function on a small structured scaffold?

### 1.1.1. *Natural enzymes*

Enzymes are protein machines functioning as biocatalysts that accelerate the rates of chemical reactions in living organisms. These protein catalysts have fascinated scientists since their discovery because of their superb capabilities to accelerate a myriad of reactions by a factor of up to  $10^{19}$  under mild conditions and with high control of selectivity.<sup>1</sup> Enzymes are playing multiple important roles in cellular processes such as metabolic pathways.<sup>2</sup> They are classified based on the type of chemical reaction they catalyze corresponding to the Enzyme Commission (EC) number. Seven higher level classes exist nowadays: oxidoreductases (EC 1, oxidation/reduction reactions), transferases (EC 2, transfer of a functional group from a molecule to another), hydrolases (EC 3, hydrolysis of various bonds), lyases (EC 4, removal of groups from the substrates, other than hydrolysis), isomerases (EC 5, isomerization within a molecule), ligases (EC 6, joining together of two molecules with simultaneous breakdown of a diphosphate bond in a nucleoside triphosphate) and translocases (EC 7, movement of ions or molecules across membranes).<sup>3</sup> While catalysis generally takes place in a precise localization within enzyme named the active site, some enzymes need the presence of cofactors (metal ions or small organic molecules) to participate to the catalytic mechanism. Binuclear metallohydrolases are, for example, a group of enzyme using two metal ions to catalyze the hydrolysis of amides, esters, and phosphoric esters.<sup>4</sup> This group is very diverse, contains a wide variety of structures and bears different mechanisms of catalysis with various contributions of the metal ions. On the other hand, organic cofactors directly participate in the reaction mechanism and generally transfer electrons (like nicotinamide adenine dinucleotide or flavin adenine dinucleotide) or chemical groups such as acetyl group by the acetyl coenzyme A cofactor.<sup>5-7</sup>

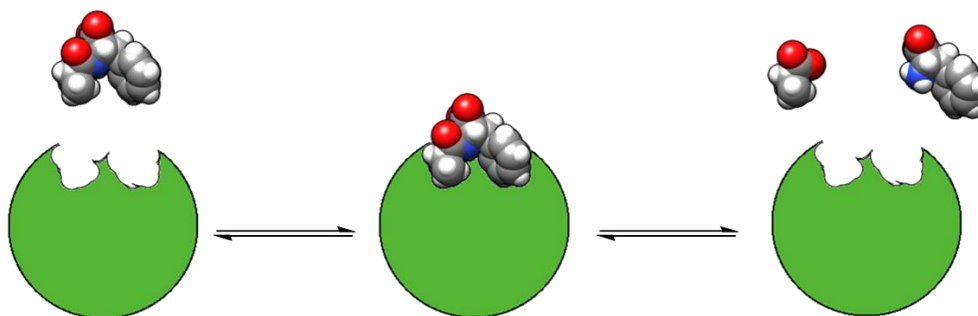
In addition to rate acceleration, enzymes also control substrate specificity, regiochemistry and stereochemistry of the reactions. These features operate thanks to the structural and conformational complementarity between the enzyme and its substrate. Historically, this concept was largely explained using the lock-and-key hypothesis first suggested by Emil Fischer: the substrate, the key, specifically binds and fills the active site of the enzyme, the lock (**Figure 1.1a**).<sup>8</sup> This principle was then extended by the induced-fit theory proposed by Daniel Koshland, where the specific substrate causes the right changes in the three-dimensional structure of the enzyme upon binding, which allow catalytic groups for proper alignment leading to catalysis (**Figure 1.1b**).<sup>9,10</sup> Actually, enzymes have to find the



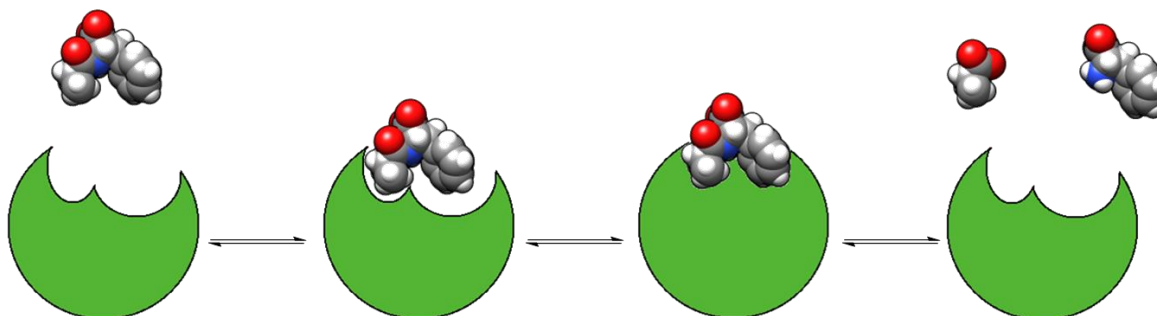
good balance between substrate binding, transition state geometry and product dissociation that often results in flexibility of the overall protein fold and/or local motions around the catalytic site.<sup>11</sup> Compared to small and rigid molecular catalysts, enzymes couple reactivity to protein motions in order to fulfill numerous roles such as precise arrangement of functional groups (e.g. general acid/base), enhanced binding of substrate, removal of water from catalytic site and stabilization of transition states (TS) notably through electrostatic interactions.

A rigorous understanding of the origin(s) of the catalytic power of enzymes remains a challenge. Despite much effort and multiple breakthroughs already attained, experts in the field are continuing to probe, adjust and expand our understanding of enzyme mechanisms. It is first important to mention that enzymes do not alter the equilibrium of reactions. But, the enzymes allow the reaction equilibria to be reached more rapidly. To do so, enzymes lower the energy barrier of activation that separates the substrates and the products when compared to the uncatalyzed reaction (**Figure 1.2a**). In the presence of a protein catalyst, the formation of the transition state (TS) is energetically more favorable (by lowering the activation energy) than the corresponding TS in solution and reaction rate will therefore be increased.<sup>12</sup> TS is defined as the point of highest energy relative to substrates on the reaction coordinate profile. Among other factors that favor catalysis within enzymes, the transition

a) Lock-and-key model

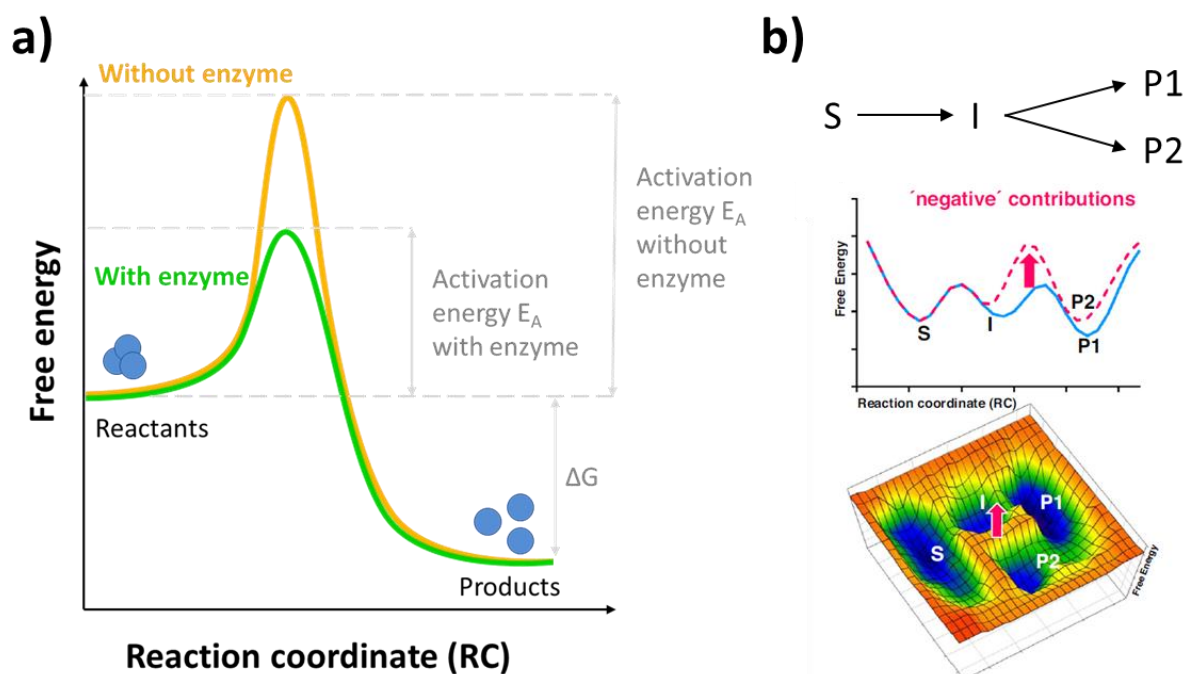


b) Induced-fit model



**Figure 1.1. Traditional models for substrate binding and enzyme catalysis.** **a)** The lock-and-key model proposes that shapes of the substrate and the catalytic site of the enzyme are complementary. **b)** In the induced-fit theory, the enzyme undergoes conformational changes upon binding the substrate that allow catalytic groups for proper alignment necessary for catalysis.

state stabilization theory is dominating and consists in the stabilization of the reaction intermediates (or transition states) through extra interactions provided by the enzyme as hypothesized by Linus Pauling.<sup>13,14</sup> It was also demonstrated that enzymes improve their packing efficiency during the binding with substrates or products, which is largely enhanced during TS binding through positively cooperative binding resulting in a higher catalytic rate.<sup>15</sup> In the TS step, this process involves a benefit in enthalpy (thanks to stronger noncovalent interactions) and a cost in entropy (due to reduced dynamics in the global structure). Nevertheless, due to the preorganization of the substrate and the enzyme, the reaction can be therefore considered as an intramolecular process and no more as an intermolecular one.<sup>16</sup> Hydrogen/deuterium (H/D) exchange of backbone amide NHs measured by mass spectrometry can be used to probe the dynamics and conformational properties of an enzyme and so the positively cooperative binding phenomenon. A reduction in exchange of backbone amide NHs should be observed when the enzyme is better packed. H/D exchange experiments in hypoxanthine-guanine phosphoribosyltransferase indicate, for example, a decrease of 21 NHs exchanges in the equilibrium substrate/product complex and a decrease of 34 NHs exchanges with the transition-state analogue complex compared to the free enzyme state.<sup>17</sup> These results emphasize that the enzyme structure is getting better packed upon binding of the substrate or the product, and the packing efficiency is even greater through the binding of the TS analogue. On top of that, the TS stabilization theory is



**Figure 1.2. Free energy changes of a catalyzed reaction by enzyme.** a) Two dimensional free energy plot showing that enzymes lower the energy barrier of activation separating the reactants and the products compared to uncatalyzed reaction in solution. b) Two dimensional and three dimensional free energy landscapes showing 'negative' contributions that destabilize the unwanted TS by increasing its energy of activation, which leads to a favored formation of the desired product P1. S corresponds to the substrates, I to intermediates and P2 to the side product. *Figure b) is adapted with permission from Figure 1(a) and (e) of Vögeli, et al. (2018) Curr. Opin. Chem. Biol. 47, 94-100. Copyright (2020) Elsevier.*

considered as an important factor for catalysis and thus, finds applications notably by the conception of TS analogues. Actually, transition state analogues convert the short-lived transition state to a stable thermodynamic state and induce tighter binding compared to substrates leading to efficient inhibitors for key enzymes like the DADMe-Immucillin-H, a TS mimic of the human purine nucleoside phosphorylase, for the gout treatment.<sup>18</sup> Beside the stabilization of favorable transition state, some enzymes appear to prevent alternative reactions by increasing the energy of activation of competing TSs, which explains the specificity of the catalyzed reaction and the formation of one unique product among others.<sup>19</sup> In this way, enzyme catalysis may be represented as a three-dimensional free energy landscape showing the possible reaction outcomes passing through different TSs (**Figure 1.2b**).

Other elements were also proposed to be crucial for high catalytic activity of some enzymes. Thanks to the precise geometry of the active site, enzymes are able to bring the substrate and the reactive moieties together in proximity and proper orientation to favor the reaction (*i.e.* intramolecular reaction). In this case, the high concentration of reactants in the form of near attack conformation (close to the transition state structure) is responsible for higher rates.<sup>20</sup> A theory introduced by Haldane known as ground-state destabilization considers the catalyzed reaction prior to the assumption of the TS stabilization: enzymes do not bind the substrate perfectly and therefore can apply a strain or a distortion on the substrate resulting in an additional energy of activation.<sup>14,21</sup> Moreover, catalytic sites are mostly characterized as a deep and buried pocket inside enzyme structure, which creates specific microenvironments promoting catalysis. By studying the first elucidated crystal structure of an enzyme, the lysozyme, Max Perutz described for the first time the interior of this enzyme as a medium of low dielectric constant in which the strength of the electrostatic interactions can be comparable to those in organic solvents rather than in pure water.<sup>22</sup> This medium effect fluctuates depending on the stabilizing interactions needed for the TS state.<sup>23</sup> Another well-known principle that enzyme may utilize to accelerate reactions is based on the covalent catalysis, also called nucleophilic catalysis, which involves the formation of a transitory covalent bond between the substrate and the enzyme or a cofactor. This strategy notably allows subdividing demanding reactions into several more feasible steps through the formation of more reactive intermediates (*e.g.* the catalytic serine of protease is both a better nucleophile than water and a better leaving group than the amide of peptide bond).<sup>24,25</sup> More recently, new hypotheses arose revealing that a complete understanding of the origin of catalytic activity of enzymes is still missing. One important contribution to catalytic mechanism of an enzyme can be quantum mechanical tunneling in transfer of hydrogen ( $H^+$ ,  $H^-$  or  $H^\bullet$ ). This model involves protein fluctuation to generate a reactive configuration along the heavy-atom coordinate from which the hydrogen transfers via nuclear tunneling.<sup>26,27</sup>

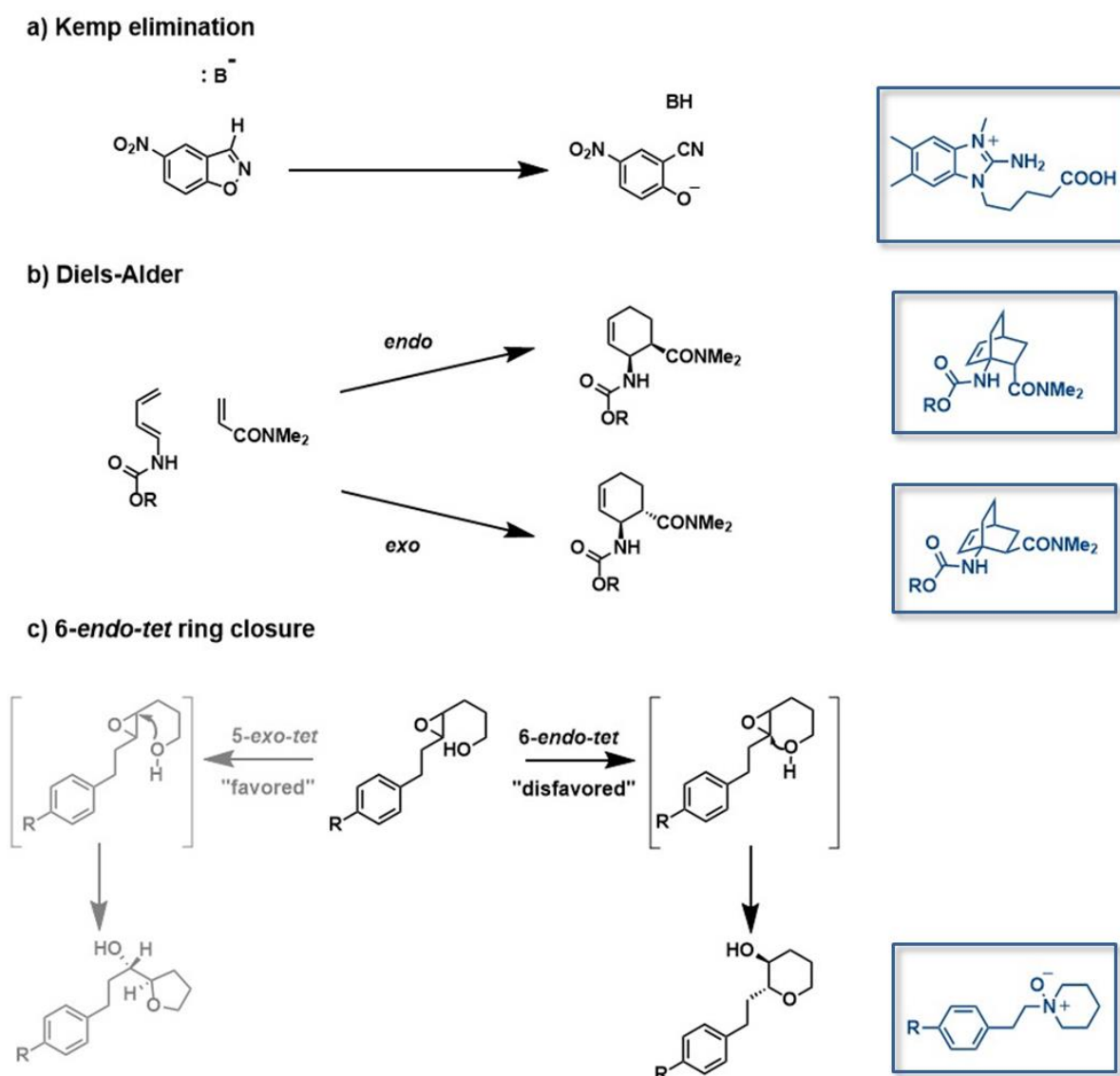
Hydrogen tunneling is directly linked to the importance of protein dynamics for catalysis (by modulation of the reactive coordinates and so of the tunneling barrier); rapid and localized vibration seems to play a role too. The latter was also found to be involved in transition state formation.<sup>28</sup> In a larger extent, coupled networks (small and long range that is to say second- and third-shell interactions) providing a framework for the maintenance of catalytic activity also appear to be critical.<sup>29</sup>

This subsection has drawn an overall picture of factors influencing enzyme catalysis revealing an extreme complexity that may be at the origin of the high specificity and high efficiency of enzymes. Although we have a clearer understanding of the different strategies used by enzymes to accelerate reaction, some of these strategies may be nonessential depending of the studied enzyme. Thus, having a complete understanding of enzyme functioning is still behind a lot of effort and research in this field. To challenge our understanding, one approach consists of designing a new enzyme from scratch or to design a new catalytic function different to those found in nature. Moreover, being able to reproduce selectivity, rate-enhancements and product specificity of natural enzymes in aqueous conditions is an outstanding task that would allow developing powerful designed technologies.

### **1.1.2. *Installing a catalytic activity into a large and natural protein scaffold***

Functional design of both already existing and abiological enzymatic activity is a challenging assignment that is seen as the Holy Grail of protein designer.<sup>30</sup> Developing such protein catalysts provides significant benefits in terms of fundamental understanding and constitutes new tools for industrial processes, medicine and chemical biology studies.<sup>31</sup> Designing catalytic functions using natural protein scaffold enables to take advantage of the intrinsic stability of protein folding. Compared to completely *de novo* catalytic protein, in which scaffold and function are designed from scratch, the use of natural protein scaffold avoids having to predict stability of novel scaffold created in the laboratory and thus, seems to be a more tractable task.

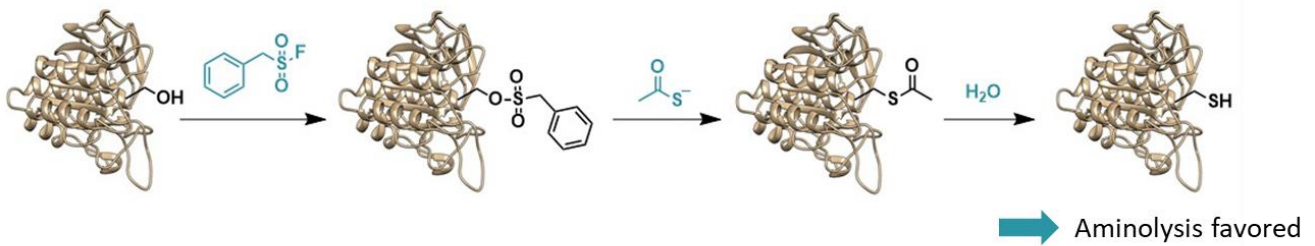
**Catalytic antibodies.** Catalytic antibodies are one historical approach to develop protein catalyst that led to the acceleration of a wide range of chemical reactions. Catalytic antibody principle exploits the diversity and the specificity of mammalian immune system and is based on the TS stabilization hypothesis from Pauling, in which the enzyme provides extra and complementary interaction to stabilize the TS resulting in a rate enhancement. In response to a hapten, here a stable TS analogue, antibodies carrying tailored binding pockets for the TS mimic are produced.<sup>32,33</sup> The production of catalytic antibodies takes advantage of the rapid and adaptive immune responses. The resulting catalytic antibodies are then highly specific to the TS promoting the catalysis of a number of reactions.<sup>34</sup> The key step consists in defining with a high precision the best hapten for the desired reaction, and then synthesizing it. Several strategies to develop haptens were established such as the widespread design of TS analogues by intuition, docking simulations or quantum mechanical models, and the more recent 'bait-and-switch' strategy dealing with haptens that bear charges to induce the installation of complementary charges in the binding site of the antibody.<sup>35</sup> Such developed catalytic antibodies were prone to catalyzing a large number of chemical reactions, notably non-natural reactions (e.g. Kemp elimination, a ring-opening reaction initiated by deprotonation of the substrate, see **Scheme 1.1a**),<sup>36</sup> stereoselectivity-controlled reactions (e.g. Diels-Alder with selection of either *endo* or *exo* attack of the dienophile on the diene, see **Scheme 1.1b**)<sup>37</sup> and disfavored processes (e.g. formation of the disfavored anti-Baldwin<sup>38</sup> 6-*endo-tet* ring closure of an epoxyalcohol, see **Scheme 1.1c**).<sup>39</sup> Thanks to these features, catalytic antibodies find multiple synthetic and therapeutic applications including the control of an enantioselective key step in the total synthesis of (-)- $\alpha$ -multistriatin,<sup>40</sup> antibody-catalyzed cocaine hydrolysis for treatment to addiction<sup>41</sup> and antibody-directed prodrug activation.<sup>42</sup> Despite promising catalytic efficiencies of some antibodies providing a platform for the study of biocatalytic processes, most of them did not fulfil the initial expectations.<sup>33</sup> Limited catalytic rate and efficiency were generally attributed to imperfect TS analogues, inhibition by product, a lack of electrostatic environment for chemical reaction (*i.e.* selection towards binding and not catalysis), differences in time scales of evolution between natural enzymes and catalytic antibodies and inadequacies of the immunoglobulin fold with a lack of conformational flexibility.<sup>43,44</sup>



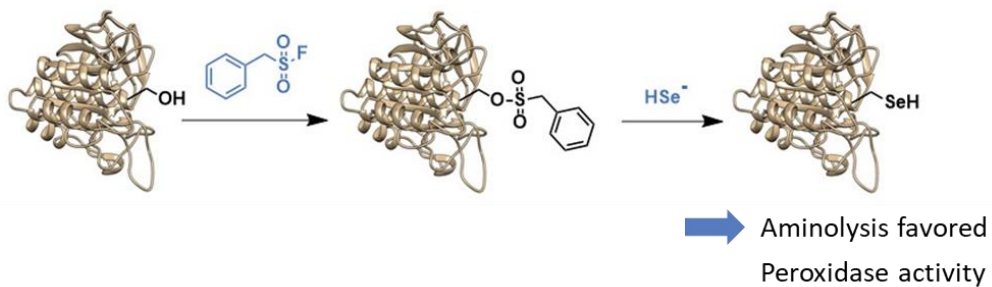
**Scheme 1.1.** Examples of chemical reactions accelerated by catalytic antibodies. **a)** The abiological Kemp elimination through deprotonation of the substrate. **b)** Diels-Alder reaction: with two different haptens, different catalytic antibodies were selected to either catalyze the *endo* or the *exo* cycloaddition. **c)** The disfavored 6-*endo-tet* ring closure was exclusively catalyzed. Designed haptens that were selected for each indicated reaction are shown in blue in box.

**Protein redesign.** The emergence of natural protein redesign (or reengineering) as a prompt route for installing new catalytic activity was mainly facilitated by the continuous development of tools for structural characterization of biomolecules and for directed evolution. Without extensive work needed, the major goals are here to create new catalytic activity, modulate the existing function, increase the stability of the protein, better tolerate specific environment and so on.<sup>45–47</sup> Before the era of recombinant DNA, chemical mutation was intelligently applied to post-translationally modify an amino acid in or near the active site resulting in a semisynthetic enzyme carrying new catalytic functionalities.<sup>48</sup> For example, Bender and Koshland simultaneously transformed the catalytic serine of the subtilisin, a protease, to a cysteine in a three-step process (Figure 1.3a).<sup>49,50</sup> The resulting thiolsubtilisin

## a) Thiolsubtilisin

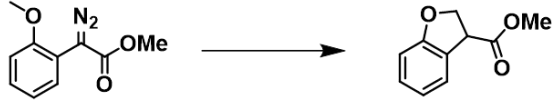
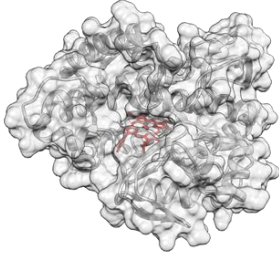
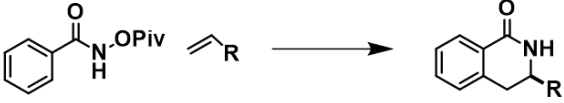
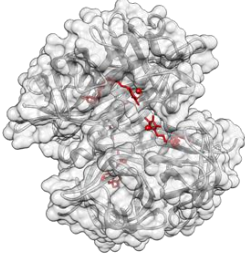


## b) Selenosubtilisin



**Figure 1.3. Redesign of the subtilisin, a serine protease.** a) Transformation of the subtilisin to thiolsubtilisin that favored more aminolysis than hydrolysis. b) Chemical reactions for the conversion of the subtilisin to selenosubtilisin, which favored even more aminolysis than hydrolysis and also carried some peroxidase activity. *Subtilisin structure from PDB: 1SUD*

surprisingly carried a very poor protease activity. More than ten years later, Kaiser deduced that this semisynthetic thiolsubtilisin largely favored aminolysis of active ester (*p*-chlorophenyl ester) and could be used for the catalysis of peptide ligation (more will be discussed later in section 1.2.3).<sup>51</sup> As demonstrated by the latest example, only few changes can be sufficient to effectively alter a function and/or to obtain a new one. A more recent example showed that a single mutation of the catalytic Tyr to an Ala was able to convert a pyridoxal 5'-phosphate (PLP)-dependent alanine racemase into an aldolase by playing with the intrinsic catalytic promiscuity of the PLP cofactor.<sup>52</sup> In general, once a sufficient level of activity is attained, directed evolution can be used to enhance the activity, improve the stability, increase the tolerance to non-natural medium and better control the stereoselectivity.<sup>53,54</sup> Directed evolution is also employed to modify original activity of enzyme such as modification of substrate specificity towards a non-natural one. In this way, systematic reengineering of aminoacyl-tRNA synthetases (aaRSs) was investigated to allow the introduction of non-canonical amino acids to the genetic code. Thanks to aaRSs crystal structures, library of mutants can be produced by modifying residues in the amino acid-binding site. Then, the aaRS mutants capable to catalyze the aminoacylation of cognate suppressor tRNA with the desired unnatural amino acid are selected.<sup>55</sup> In this way, numerous aaRSs were reengineered to recognize more than 70 unnatural amino acids. Despite the established efficiency of taking advantage of original characteristics of active sites, such

Catalyzed reaction	Protein structure	Ref.
a) 		58
b) 		59

**Figure 1.4. Protein reengineering with synthetic organometallic cofactors.** a) Insertion of an iridium porphyrin in a P450 variant led to a highly efficient catalyst with enzyme-like kinetic parameters for various abiological insertions of carbenes into C-H bonds. b) Non-native transition metal complex of rhodium(III) with an attached biotin moiety was docked into an engineered streptavidin and was shown to catalyze the asymmetric C-H bond activation. Cofactors are depicted in red. Structures from PDB: 1I07 and 5CSE

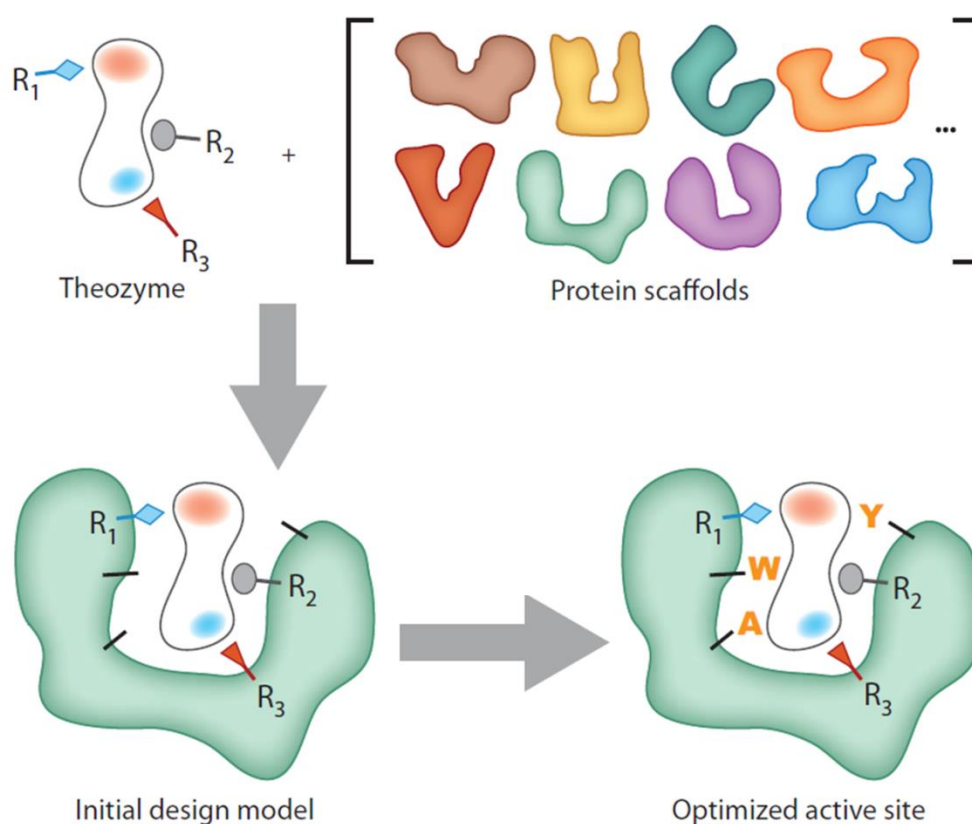
reengineered activities and applications are generally similar (at least in a mechanistic point of view) to the initial function of the protein. To go further and reach more distinct functionalities, some investigations were, for instance, focused on the introduction of amino acids that bear completely different chemical characteristics in order to expand the range of catalyzed reactions. The subtilisin, which was discussed hereinabove, was transformed to selenosubtilisin by a two-step chemical conversion of the catalytic serine to a selenocysteine (Sec; **Figure 1.3b**).<sup>45</sup> The resulting semisynthetic enzyme showed the highest ability for aminolysis of acyl-enzyme intermediate compared to the wild type subtilisin and the thiolsubtilisin (aminolysis over hydrolysis selectivity for selenosubtilisin was 14000 times and 20 times higher, respectively).<sup>56</sup> But more surprisingly, the incorporation of selenocysteine into this protease led to a completely novel activity of peroxidase.<sup>57</sup>

Finally, the use of organic and organometallic cofactors was also investigated in order to introduce completely novel catalytic activities thanks to their intrinsic chemical properties. One can simply replace the natural counterpart into a protein scaffold (e.g. insertion of an iridium porphyrin in a P450 variant in combination with directed evolution led to a highly efficient catalyst with enzyme-like kinetic parameters for various abiological insertions of carbenes into C-H bonds – **Figure 1.4a**)<sup>58</sup> or can install the catalytic moiety into an inert scaffold (e.g. a non-native transition metal complex of rhodium(III) with an attached biotin



moiety was docked into an engineered streptavidin and was shown to catalyze asymmetric C-H bond activation – **Figure 1.4b**).<sup>59</sup>

**Computational enzyme design.** The major goal of computer-aided design of enzyme is to find the best protein scaffold able to accept the hypothesized active site for diversified reactions including non-biological ones. Computational enzyme design has to face two main challenges: *i*) sufficient sampling of candidate protein scaffolds to identify optimal locations for the introduction of active-site residues; *ii*) modelling the active site with adequate precision to enable the acceleration of the target reaction.<sup>30</sup> To facilitate this process, powerful computer programs such as the widespread Rosetta suite (especially RosettaMatch and RosettaDesign),<sup>60,61</sup> ORBIT<sup>62,63</sup> and SABER<sup>64</sup> were developed. Computational enzyme design generally followed same main steps (**Figure 1.5**). First, an appropriate catalytic mechanism corresponding to the target reaction is defined without forgetting the necessary functional groups. Second, quantum mechanical calculations are carried out to generate the structure of rate-limiting transition state(s) (TS(s)) and the three-dimensional arrangements of amino acid side chains that are optimal to stabilize the TS(s). The resulting model corresponds to the idealized requisite composition of a minimal active



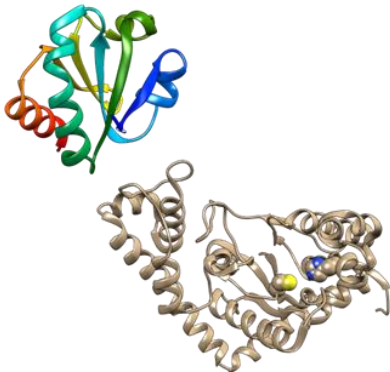
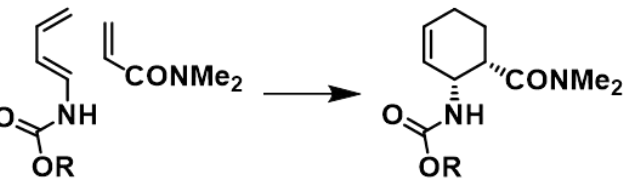
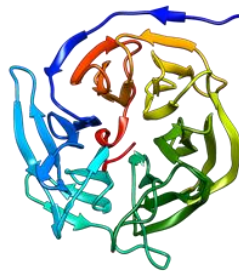
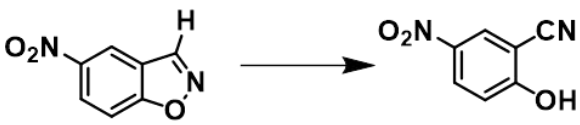
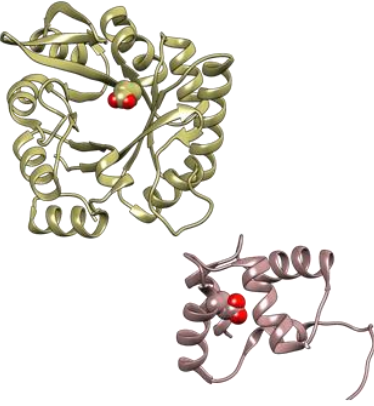
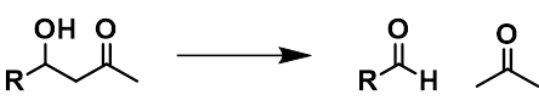
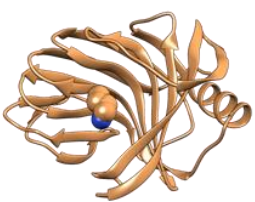

**Figure 1.5 Computational enzyme design concept.** First, a theozyme is modelled by quantum mechanics to define the idealized active site that would stabilize the transition state(s). The theozyme is then matched to complementary pockets of a set of proteins provided by the Protein Data Bank (PDB). Finally, residues in the surrounding of the selected sites are optimized for good packing and binding of the TS(s). The figure is reprinted with permission from Figure 3 of D. Hilvert (2013) *Annu. Rev. Biochem.* 82, 447-470. Copyright (2020) Annual Reviews.

site, named a theozyme.<sup>65</sup> This theozyme is then grafted *in silico* into structurally characterized proteins from the Protein Data Bank (PDB) by sequentially attaching each side chain of the theozyme to the backbone of the protein scaffold. Residues in the surrounding of the selected pocket are then mutated and optimized for good packing and fold stability that would complement the geometric and electronic features of the modeled TS. Algorithms iteratively look for the conformational space of the TS and the side chains to minimize. Finally, the designed structures are filtered, ranked and evaluated on the basis of empirical criteria like global energies, the catalytic site geometry, the TS binding energy, the packing and shape complementarity between the designed pocket and the TS. Promising designs are synthesized, purified and tested for their ability to catalyze the desired reaction.<sup>33</sup>

In early work, Stephen Mayo identified and optimized possible active sites for nucleophilic hydrolysis of *para*-nitrophenyl acetate (*p*NPA) mediated by a His residue on an inert protein scaffold of thioredoxin from *Escherichia coli* (*E. coli*) (**Figure 1.6a**).<sup>63</sup> The designed proteins demonstrated significant catalytic activity proving the feasibility of applying a computer-aided approach to insert a catalytic function on inert protein scaffold. This concept for enzyme design was thereafter implemented to accelerate a variety of reactions utilizing different strategies of catalysis. Designed protein catalyst for the abiological Diels-Alder cycloaddition used, for example, the principle of proximity effect to preorganize the two substrates in a shape-complementary pocket in order to accelerate this bimolecular reaction (**Figure 1.6b**).<sup>66</sup> The cycloaddition between a dienophile and a diene is a reaction of high synthetic value by creating two C-C bonds and up to four stereocenters. As the natural counterpart does not exist, designed Diels-Alderases are of great interest. In this study, two of the fifty designed enzymes presented substantial catalytic activity that revealed to be substrate specific with the intended stereoselectivity but were less active than previous catalytic antibody.<sup>37</sup>

Moreover, Kemp elimination (especially of 5-nitrobenzoxazole) is a model reaction for proton transfer and allows to test our ability to properly position acid and base groups and to tune their reactivity through their installation in networks of residues. By using the aforementioned protocol, David Baker and collaborators produced eight designed active enzymes that catalyzed the base-directed opening of 5-nitrobenzoxazole with rate enhancement comparable to catalytic antibodies (**Figure 1.6c**).<sup>36,67</sup> Crystal structure of the resulting enzyme showed nearly atomic accuracy compared to the designs.<sup>67</sup>

Then, covalent catalysis, also called nucleophilic catalysis, is a classic strategy for enzymes that was also exploited in mechanism of several computationally-designed catalytic proteins. Nucleophilic residue is always accompanied with a complex constellation of amino acids in order to perturb its  $pK_a$ , to add acid-base catalysis (by amino acid side chains or water molecules), to stabilize the TSs, which represents a more challenging task to design.<sup>68</sup>

Reaction type	Protein structure	Ref.
<p>a) Ester hydrolysis</p> $\text{R}-\overset{\text{O}}{\parallel}{\text{C}}-\text{O-Aryl} \longrightarrow \text{R}-\overset{\text{O}}{\parallel}{\text{C}}-\text{OH} + \text{HO-Aryl}$		<p>63</p> <p>70</p>
<p>b) Diels-Alder</p> 		<p>66</p>
<p>c) Kemp elimination</p> 		<p>67</p> <p>75</p>
<p>d) Retro-aldol</p> 		<p>69</p>
<p>e) Superoxide dismutase</p> $2 \overset{\ominus}{\text{O}}_2 + 2 \text{H}^+ \longrightarrow \text{O}_2 + \text{H}_2\text{O}_2$		<p>73</p>

**Figure 1.6. Computational enzyme design on natural protein scaffold.** Multiple reactions were catalyzed by grafting optimized catalytic site onto natural protein scaffold. **a)** Ester hydrolysis by either placing a His residue on a thioredoxin protein or designing a catalytic dyad Cys-His on another scaffold. **b)** Diels-Alderase activity designed on a six-bladed  $\beta$ -propeller scaffold. **c)** To catalyze Kemp elimination, a Glu residue as general base was positioned either by computational design or by more minimalist design. **d)** Retro-aldol activity was designed in a jelly-roll fold. **e)** An iron-binding site was designed on the thioredoxin protein for superoxide dismutase activity. Structures from PDB: 2TRX, 3ULV, 3I1C, 2RKX, 2KZ2, 3B5L.

Computational design successfully managed, for example, to generate more than 30 active enzymes for the multistep retro-aldol cleavage (model substrate: 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone) in which a C-C bond is broken through the formation of a Schiff base with a catalytic Lys residue (**Figure 1.6d**).<sup>69</sup> Most of the successful designs contained an ordered water molecule to mediate proton shuffling, whereas more complex networks of amino acids to activate the Lys or to assist the C-C bond cleavage were less effective, suggesting difficulties in modelling larger network of functional groups. Nucleophilic catalysis strategy was also exploited to design esterase activity by placing a catalytic dyad Cys-His and an oxyanion hole into protein scaffolds (**Figure 1.6a**).<sup>70</sup> In this two-step acylation/deacylation mechanism, the cysteine acylation showed rapid kinetics (up to 4000-fold acceleration over background reaction), but the turnover was limited by the slow deacylation. X-ray structures exhibited that either the histidine or the nucleophilic cysteine were misplaced compared to design models, reflecting again difficulties in correctly positioning multiple and mutually dependent residues. Finally, artificial metalloenzymes were designed to benefit from the intrinsic properties of cofactors such as their better ease for redox chemistry and radical reactions. Automated programs like DEZYMER<sup>71</sup> and Metal Search<sup>72</sup> were developed to identify sites in protein scaffold that would allow placement of residues with the proper geometry for metal binding. A superoxide dismutase activity was, for example, observed by incorporating a high-affinity iron-binding site into the thioredoxin protein from *E. coli* (**Figure 1.6e**).<sup>73</sup> However, designing catalytically-active metal binding site is not trivial (proper geometry for binding, negative design to prevent undesired interactions, poor parametrization of metal ions, second and third coordination shell to consider), limiting the efficiency of designed metalloenzymes.<sup>68</sup>

In general, computational enzyme design gave only low catalytic efficiency compared to natural enzymes. But this method was still found to be necessary to generate active designed catalyst at reasonable frequency. Control experiments with randomized libraries that approximate the amino acid composition without fixing the positions of the catalytic residues (in contrary to the theozyme approach) were much less efficient than computational design and gave far fewer successful hits.<sup>74</sup> Nevertheless, rational and minimalist design of protein catalyst coupled with simpler computational procedures is still on par with more extensive computational design such as the theozyme procedure. The regulatory protein calmodulin (CaM) was designed to form an allosterically controlled Kemp eliminase through intuition and screening of the possible positions of the catalytic residue (Asp or Glu) in a hydrophobic pocket followed by docking of the substrate.<sup>75</sup> Upon binding of Ca(II) ion, a CaM protein containing a single mutation opened the hydrophobic pocket resulting in substantial catalytic activity for Kemp elimination that was not far from previously characterized Kemp eliminase.<sup>36,67</sup>

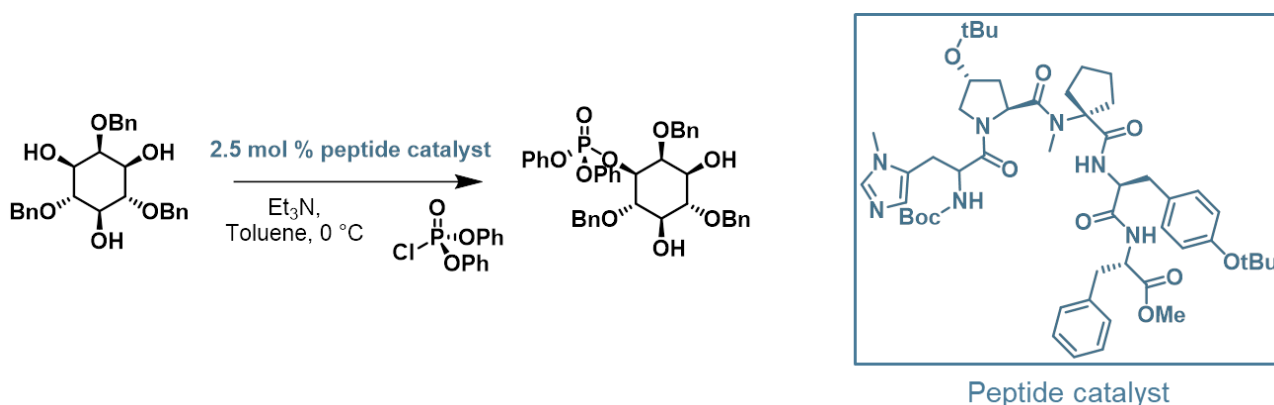
Installing effective catalytic functionalities comparable to natural enzymes is a demanding task and generally did not reach the expectations after the first round of design. The reasons of this low success rate were principally assigned to the oversimplification of the exact placement of essential residues, of the backbone behavior and of conformational dynamics as well as the underestimation of the role of long range interactions (*i.e.* with second- and third-shell residues).<sup>76</sup> In an attempt to improve catalytic efficiency of the first generation of designed catalysts, iterative processes including precise structural characterization, more rational redesign steps and directed laboratory evolution are involved. After careful analysis of the crystal structure of a poorly active Diels-Alderase,<sup>66</sup> the active site turned out to be actually open on one side letting the substrates to be solvent exposed. Players from Foldit, a crowdsourcing folding game, were challenged to remodel active site loops to increase interactions with substrates.<sup>77</sup> Several iterations of design and structural characterizations led to the generation of an extra helix-turn-helix motif with an increase of more than 18-fold of catalytic activity. Another example enabled the production a Kemp eliminase with one of the highest kinetic parameters from a completely inactive variant through multiple iterative cycles of detailed biophysical and structural analysis combined with molecular dynamics (MD) simulations.<sup>78</sup> But the most widespread solution to achieve higher rates is to apply high-throughput methods such as directed evolution to optimize the sequences of initially designed scaffolds yielding man-made enzymes with higher catalytic efficiencies.<sup>79,80</sup> Through an iterative process of random mutations and selection of the active analogues, directed evolution typically provides access to a fine-tuned reshaping of the catalytic site and of the positions of the catalytic residues.<sup>81</sup> Directed evolution also demonstrated its effectiveness for correcting the function of the higher-shell residues that participate to protein conformational dynamics for a better enzyme packing around the TS(s) as well as for an efficient substrate recognition and product release.<sup>82,83</sup> Surprisingly, this process of optimization can also induce serendipitous discovery such as the creation of a new active site pocket with the appropriate placement of the catalytic residues.<sup>82</sup> Nevertheless, as directed evolution works better with protein catalysts already displaying some level of the desired activity, computational enzyme design still plays a key role in installing a catalytic activity on inert protein scaffolds from scratch even if the initial efficiency is modest.<sup>84</sup>

The process of computational design focuses on generating active site to stabilize TS as principal strategy to catalyze the intended reaction, but it is clearly not sufficient according to the quite low success rate (the number of successful designs is low compared to the number of tested designs). As discussed in the previous subsection, enzymes use much more intricate approaches favoring high catalytic rate accelerations, which are then more or less accessible after numerous cycles of directed evolution. Still, some features may be

difficult to approach through directed evolution alone without some degree of initial design; this includes large-scale motion such as loop and domain motion and flexibility or allosteric control to regulate the activity of the catalyst.<sup>30</sup> Thus, efforts on improving computational tools and our knowledge for more precise chemical intuition would be necessary. Several lines of research to enhance computational enzyme design based on our understanding of natural enzymes are conceivable: *i*) more complex theozyme taking into account longer-range interactions; *ii*) taking into account all parts of the catalytic cycle from substrate binding to product release; *iii*) accommodation of backbone flexibility; *iv*) second- and third-shell electrostatic interactions; *v*) explicit interactions with bulk water; *vi*) protein dynamics; *vii*) applying computational design for more complex reactions.<sup>68</sup> However, incorporating all these details in design will largely complicate computational calculations that would be even more time consuming and finding a protein scaffold that accepts all these considerations will be more laborious. Although this constitutes a true challenge, it will benefit from the continuous development of design software.

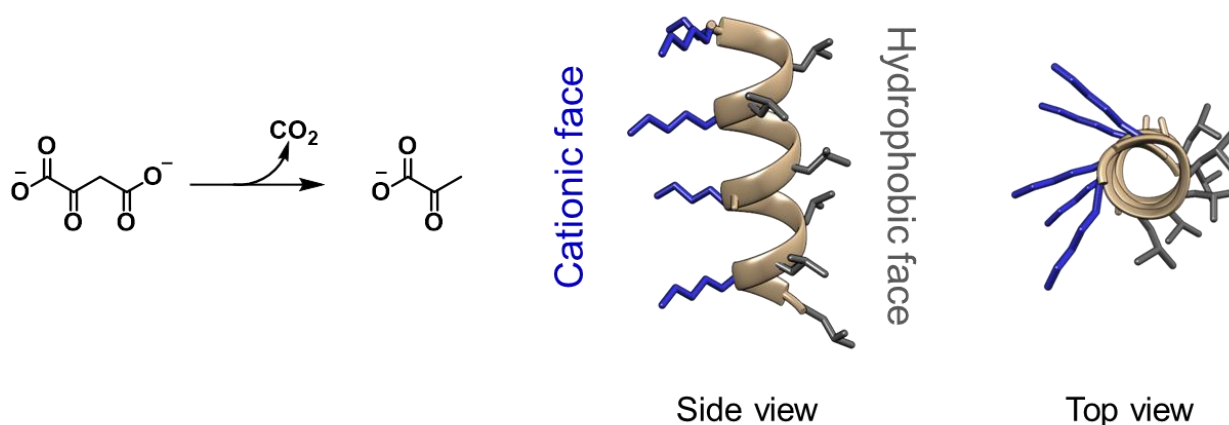
### 1.1.3. Installing a catalytic activity into minimalistic de novo protein scaffold

Short polypeptides have proven to be efficient catalysts for a wide range of chemical reactions especially for various asymmetric transformations suggesting that small protein scaffolds can also be good candidates to serve as enzyme mimics.<sup>85</sup> Polypeptides composed of few amino acids are indeed suitable frameworks for enabling chiral catalysis since they may already adopt a secondary structure (e.g.  $\beta$ -turn)<sup>86</sup> convenient for transfer of chirality (in **Scheme 1.2**, an example of catalytic asymmetric phosphorylation of an inositol derivative by a  $\beta$ -turn peptide).<sup>87</sup> Moreover, the modularity of such peptides (accessible through split-and-mix libraries for example)<sup>88</sup> promotes simpler structural modification and fine-tuning of



**Scheme 1.2. Asymmetric catalysis by a polypeptide.** Here, an asymmetric phosphorylation of an inositol derivative was catalyzed by an artificial polypeptide with a  $\beta$ -turn like conformation.

reactivity and selectivity. Highly structured polypeptides allow for a higher level of preorganization that may be beneficial for some reactions. The  $\alpha$ -helix provides an attractive structural scaffold particularly due to its amphipathic characteristics. For example, oxaldie peptide, an  $\alpha$ -helical peptide of 15-residue length and mostly composed of Leu and Lys, was found to catalyze decarboxylation of oxaloacetate to pyruvate through imine formation (**Figure 1.7**).<sup>89</sup> Two strategies were explored for lowering  $pK_a$  of possible amine donors, either through helix macrodipole interaction in the case of the N-terminal amine, or via coulombic interactions of the ammonium groups of several neighboring lysine side chains. Both strategies yielded substantial catalytic activity, demonstrating that an explicit binding pocket is not necessary in contrary to a catalytic antibody. Further improvement of the scaffold by controlling the monomeric packing with the addition of a polyproline chain (*i.e.* to avoid oligomerization to helical bundle through hydrophobic interactions) and by stabilizing the folding slightly enhanced the kinetic parameters of oxaldie-type peptides.<sup>90,91</sup> Although obtaining stable three-dimensional structures from scratch requires more sophisticated design process, it seems to provide sufficient control over substrate binding and reaction mechanism that is on par with designed enzymes using large and natural protein scaffold. So one can ask whether an entire protein scaffold is always mandatory for enzyme-like catalysis or small structured proteins can also enable sufficiently high catalytic activities.

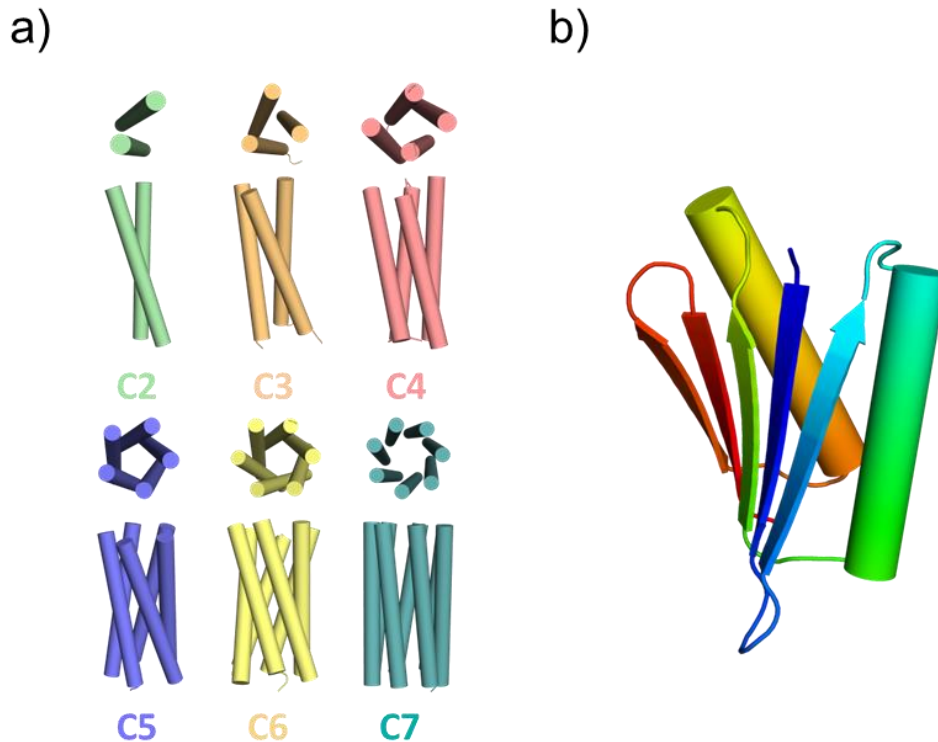


**Figure 1.7. Decarboxylation of oxaloacetate to pyruvate by oxaldie peptide.** The  $\alpha$ -helix organization is enabled by its repeating pattern, one face was grafted with Leu residues creating a hydrophobic side and the other was grafted by Lys residues creating a cationic side. Spatial clustering of amine groups of lysine side chains lowers  $pK_a$  through coulombic interactions allowing imine formation, a necessary step for catalyzing the reaction. Peptide has been modelled with Chimera software following the 1-letter sequence of oxaldie: LAKLLKALAKLLKK. Leucine residues are depicted in dark grey, lysine in dark blue and alanine in tan.

**Use of natural versus de novo protein scaffolds.** As mentioned before, the use of natural protein scaffold eliminates the need to predict stability of a new scaffold created from scratch and thus, seems to be a more tractable task. However, designing a precise folded protein structure has recently become easier because of better understanding of fundamental rules necessary to build three-dimensional structures and development of

modern computational tools. For instance, an assembly of helical bundles is now well understood and controlled, and highly stable coiled-coils with different oligomeric states are now easily accessible (**Figure 1.8a**).<sup>92</sup> This may notably explain the widespread use of coiled-coil structures to design *de novo* catalytic protein (as mentioned below via the large number of successful functional designs of helical bundles). Naturally occurring proteins are actually restricted in terms of folding diversity, which is estimated between 1000 and 10000 distinct types of natural folds.<sup>93</sup> Evolution has only explored a tiny amount of possible protein structures as it proceeds by sequential mutation and selection which limits the diffusion across the entire protein structural space.<sup>94</sup> Completely *de novo* proteins have the advantage of being unrelated to any protein family, which allows to explore the “dark matter” of protein space and so to reach protein structures that do not exist in nature.<sup>93</sup> Pentameric, hexameric and heptameric helical bundles, which are not commonly found in nature, were successfully designed and let the possibility of further functionalization thanks to the clearly defined central channel (**Figure 1.8a**).<sup>95</sup> Baker lab demonstrated the feasibility of designing and production of a completely *de novo* protein, the Top7 protein, carrying a globular fold with a novel topology (**Figure 1.8b**).<sup>96</sup> With the possibility to access new topologies and new folds for a small protein scaffold, the dependency on natural scaffold can be suppressed, which would allow for a better control of functional design and may be more appropriate for reaching catalysis of novel reactions. In fact, working with a small *de novo* protein reduces the complexity of the scaffold and allows to better understand the effects of each modification and the contribution of each side chain to structure, stability and function.<sup>97</sup> In this way, *de novo* design of small catalytic protein is a good strategy to establish the minimum structure required for function. Other advantages of using small *de novo* scaffold for installing catalytic activity are the ease to work with notably due to a greater control of stability, the possibility of modification, the more straightforward synthesis and as a result lower price of production, which may fit better with industrial applications. However, compared to natural scaffolds, some features look harder to introduce such as how to reproduce protein dynamics and bulk solvent removal in the active site, which would represent the future challenges of *de novo* design.





**Figure 1.8. Fully *de novo* protein scaffolds.** a) Coiled-coil assembly with different and controlled oligomeric states. Five-, six- and seven-helix bundles are not commonly found in nature. b) The Top 7 protein presents a globular fold with a completely novel topology. Structures from PDB: 4DZM, 4DZL, 3R4A, 4PN8, 4PN9, 4PNA, 1QYS.

**Design of binding function as a first step towards catalysis.** Since the early days of *de novo* functional protein design, metal ions and organic cofactors have been appreciated for providing catalytic activity due to their intrinsic and diverse catalytic properties. That is why initial work first concentrated to master binding of metal ions and organic cofactors (e.g., heme) before studying their catalytic properties. Moreover, around one third of naturally occurring proteins binds at least one metal ion, which plays essential structural and functional roles.<sup>98</sup> As a result of finding only a limited set of metal ions in biological processes, the same metal can provide different functions such as oxygen binding, electron transfer and catalysis.<sup>99</sup> A lot of effort was therefore invested to better understand the influence of protein matrix in tuning metal ion properties. In this way, *de novo* protein scaffolds are highly suitable for an identification of the important features for a desired function because of their relative simplicity.

Diiron class of protein was, for example, heavily studied through *de novo* design to elucidate how the environment of the metal-binding site can influence the functional properties of the protein.<sup>100</sup> By careful retro-structural analysis of natural carboxylated-bridge diiron proteins and computational modelling, a set of four-helix bundles were designed to contain a diiron center with four Glu and two His residues in the hydrophobic core as

coordination site. The resulting DueFerri (DF) family is composed of different designs like dimeric helix-loop-helix motif,<sup>101</sup> heterotetramer<sup>102</sup> or monomer of the four-helix bundle.<sup>103</sup> Upon binding of several types of metal ions, DF protein fold is stabilized and presents highest affinity for metal ions that classically bind in octahedral or tetrahedral geometries (Zn, Co, Mn, Fe).<sup>104</sup> This family was extensively studied by X-ray crystallography to determine adjustments and flexibility of the coordination site with different ligands.<sup>100</sup> Iron-sulfur (FeS) clusters were also studied in particular for their reducing potential and their capacity for long-range electron transfer. Here again, multiple *de novo* proteins were designed taking advantage of a stable platform that allows to examine influence of the environment on the FeS-cluster properties. If designs took mainly inspiration from naturally containing-FeS proteins and reused native sequences,<sup>105–107</sup> *de novo* design also enabled placement of a four-iron four-sulfur cluster in a completely novel environment compared to natural FeS proteins such as the hydrophobic core of a four-helix bundle.<sup>108</sup> This opens new possibilities in terms of practical applications. Through computational design, metal ions, here zinc ions, were also used to mediate the formation of homodimer with orientation preferences.<sup>109</sup> Zinc ions largely increased the affinity between the two monomers, thereby increasing the potency of applications for this tight dimeric scaffold.

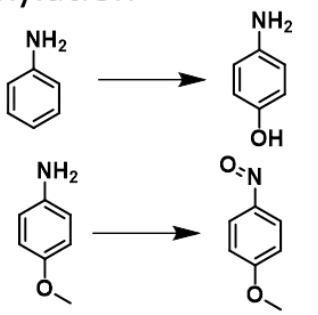
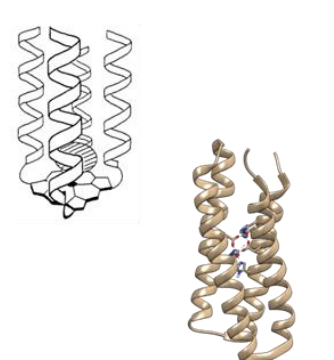
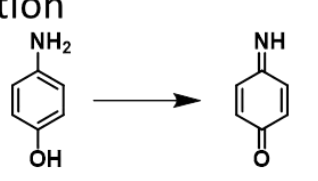
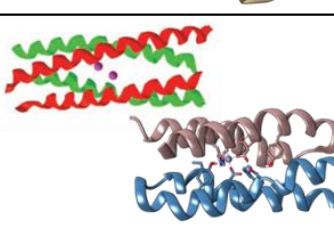
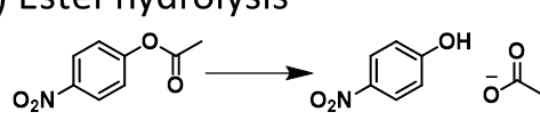
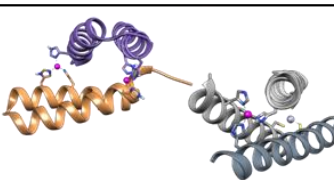
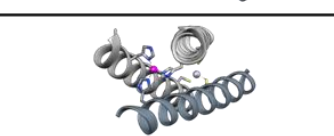
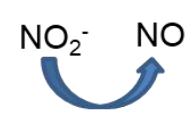
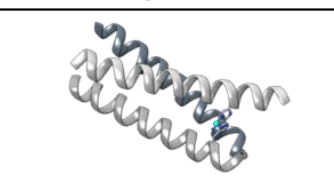
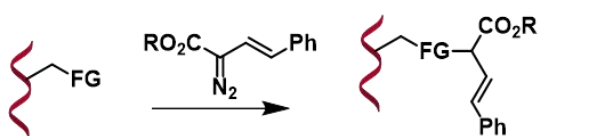
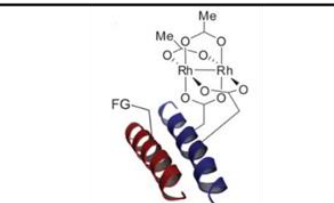
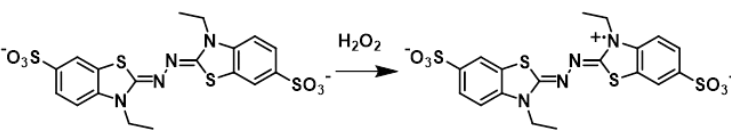
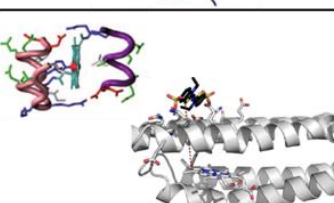
Heme proteins are another category of metalloproteins that have attracted a lot of interest due to their natural abundance in a variety of protein scaffolds enabling diverse functions like electron transfer, oxidation, storage and transport.<sup>110</sup> In early work, peptide-based hemoprotein mimetics were designed by taking inspiration from natural heme-binding proteins, using native sequence of interest for binding and covalently linking the heme cofactor to the polypeptides.<sup>110,111</sup> Mimochrome protein family is one this class of hemoprotein mimics, which was basically designed with simple molecular modeling studies based on the  $\beta$ -chain F helix of hemoglobin.<sup>112</sup> Two helices were covalently bridged by a deuteroporphyrin molecule thanks to lysine side chains resulting in a helix-heme-helix sandwich conformation convenient for spectroscopy analysis. The mimochrome protein was further designed to better understand the implication of peptidic sequence composition and length in controlling heme properties.<sup>113</sup> Later, a maquette protein was developed from first principles to form a four-helix bundle with His-guided binding of heme molecules in the hydrophobic core.<sup>114</sup> After several iterative cycles of design, oxygen was stably bound to heme thanks to hydrophobic core packing and loop design to reduce helical-interface flexibility leading to an efficient O<sub>2</sub>-transport protein that discriminated CO binding. Computational methods were also included in the process of design to produce helical bundle that specifically interacts in the core with non-natural cofactors of interest.<sup>115–117</sup> With improvements of computational design methodologies and a combination of fully extensive core packing in concert with ligand-binding site, this recently led to the design of a novel four-

helical protein that binds a specified abiological porphyrin with sub-Å accuracy compared to the model and precision necessary for catalysis.<sup>118</sup> All these studies bring closer the possibility to create from scratch efficient and catalytic active protein with unrivaled properties.

**Design of *de novo* minimalistic metalloenzymes.** As binding of metals ions and organic cofactors into *de novo* small structured proteins was mastered, a number of applications for catalysis have followed. One of the earliest *de novo* designed catalytic protein was proposed by Kaiser with the helichrome protein.<sup>119</sup> The helichrome was composed of four  $\alpha$ -helices covalently linked at their N-termini to an iron porphyrin. The resulting hydrophobic pocket above the porphyrin ring was a suitable binding and catalytic site for hydroxylation of aniline molecules (**Figure 1.9a**).

Coming back to the DF (DueFerri) protein family, DeGrado lab further modified the diiron heterotetramer variant to optimize a binding pocket to receive the substrate, 4-aminophenol.<sup>120</sup> The resulting designed protein provided an efficient artificial phenol oxidase activity dependent on O<sub>2</sub> molecules (**Figure 1.9b**). Interestingly, the similar activity of oxidase with a broader range of substrates was also demonstrated by designing another analogue of the DF family, the helix-loop-helix dimer, in order to optimize active site cleft and to better stabilize the scaffold with loop adjustment (**Figure 1.9b**).<sup>121</sup> Using the single-chain DF analogue, the oxidation activity was subsequently altered towards the catalysis of arylamine *N*-hydroxylation by remodeling the substrate cavity and modifying the coordination sphere of the diiron center, especially with the addition of a third His ligand (**Figure 1.9a**).<sup>122</sup> These examples highlight the great capacity of *de novo* scaffolds to tune catalytic activity through rational and straightforward modifications of the close and more distant environment of the catalytic site.

The deep biophysical and structural characterization of metal ions in *de novo* scaffold can furnish substantial clues for serendipitous discovery of catalytic activity. Indeed, crystal structure of the zinc-mediated homodimer (discussed before) revealed a cleft nearby the coordination site that would be suitable for substrate binding.<sup>123</sup> Thanks to the inherent power of zinc, the interface of this dimer showed primitive activity for *para*-nitrophenyl acetate (*p*NPA) and *p*-nitrophenyl phosphate hydrolysis (**Figure 1.9c**). Further design to get a monomeric assembly allowed performing several rounds of directed evolution, which led to a highly proficient Zn-containing ester hydrolase with catalytic efficiency surpassing the hydrolytic activities of the naturally occurring metalloenzymes.<sup>124</sup> This may illustrate a plausible pathway of the evolution, where metal ions would have templated the assembly of short peptides and creation of a cavity essential to install catalytic activity that would be later enhanced with consecutive mutations. Analogous coordination site for zinc (ZnHis<sub>3</sub>O) was

Reaction type	Protein structure	Ref.
<b>a) Hydroxylation</b> 		119  122
<b>b) Oxidation</b> 		120  121
<b>c) Ester hydrolysis</b> 		123  125
<b>d) CO<sub>2</sub> hydration</b> $\text{CO}_2 + \text{H}_2\text{O} \longrightarrow \text{H}_2\text{CO}_3 \longrightarrow \text{H}^+ + \text{HCO}_3^-$		125
<b>e) Nitrite reductase</b> 		127
<b>f) Functional group (FG) acylation</b> 		128- 130
<b>g) Peroxidase</b> 		132  133

**Figure 1.9. Examples of minimalistic metalloenzyme.** **a)** Hydroxylation and N-hydroxylation of arylamine have been catalyzed by helichrome and DueFerri (DF) proteins, respectively. **b)** Oxidation of phenol derivatives has been accelerated by both the heterotetramer and the helix-loop-helix dimer of the DF family in an oxygen-dependent manner. **c)** Hydrolysis of ester has been demonstrated with two different scaffolds coordinating a catalytic zinc ion. **d)** CO<sub>2</sub> hydration has been catalyzed by a coiled-coil trimer containing a zinc active site and a mercury stabilizing site. **e)** An analogous scaffold as in d) binds copper ion and has carried nitrite reductase activity. **f)** Through coiled-coil assembly, a rhodium site has catalyzed the acylation of diverse side chains (FG for functional group of amino acids). **g)** Peroxidase activity has been designed in a mimochrome variant and in a maquette protein bearing a c-type cytochrome. Structures from figures of the cited papers or PDB: 2LFD, 2KIK, 3V1C, 3PBJ.

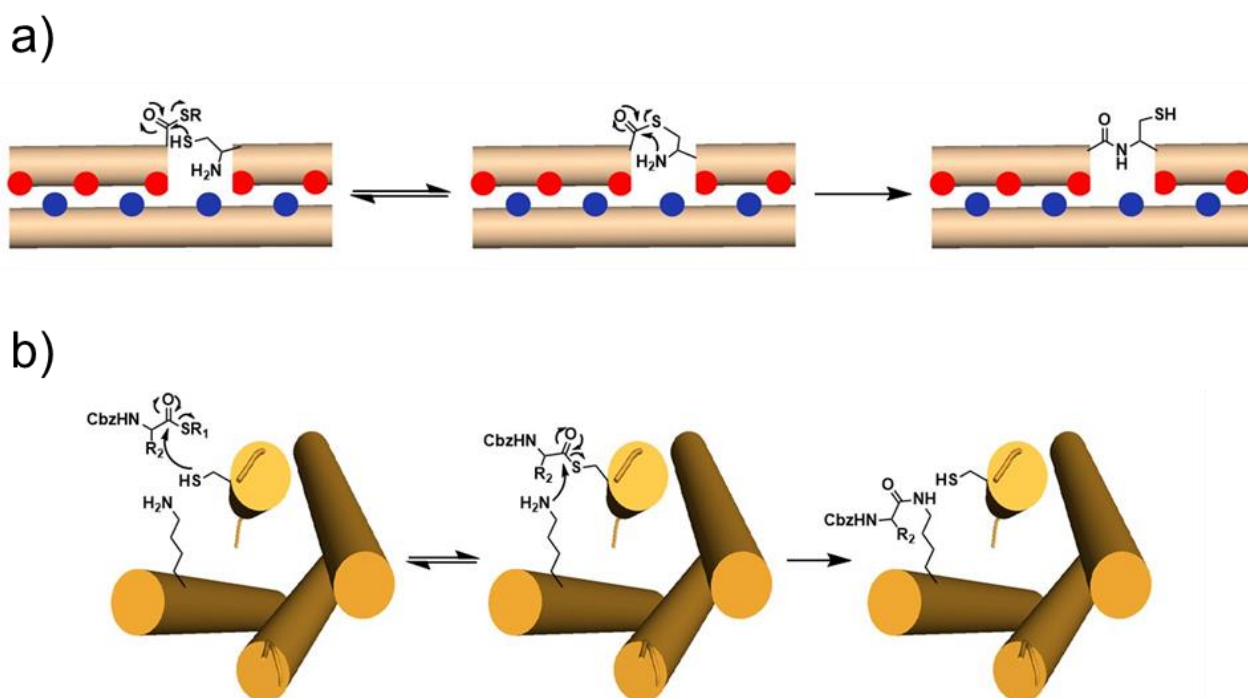
introduced in a three-stranded coiled coils stabilized by a trigonal coordination site for mercury ( $\text{HgCys}_3$ ) leading to efficient *p*NPA hydrolysis too, and also valuable  $\text{CO}_2$  hydration activity (**Figure 1.9c-d**).<sup>125</sup> Although the hydrolytic activity of *p*NPA was a bit less efficient compared to the previous example<sup>123</sup> probably due to a center less accessible to substrate, this *de novo* protein demonstrated the possibility to design two spatially distant coordination sites that bind different ligands with different functions (*i.e.* catalysis and stabilization) into the same scaffold. An additional study on the location of the coordination sites in the coiled-coil scaffold established no significant change in catalytic activity while binding affinity of metal and substrate access were slightly affected.<sup>126</sup> A related peptide with mercury binding site removed was designed to bind copper instead of zinc in a  $\text{CuHis}_3$  site and resulted in a redox system with nitrite reductase activity (**Figure 1.9e**).<sup>127</sup> Moreover, *de novo* metalloprotein design facilitates the use of abiological metal ions to access to a wider range of chemical reaction. A dirhodium site was, for example, introduced in a helix for site-specific and proximity-driven covalent modifications of amino acid (*e.g.* aromatic, Asn, Gln) side chains mediated by specific coiled-coil recognition between the catalyst and the peptide substrate (**Figure 1.9f**).<sup>128–130</sup>

Finally, heme proteins have also demonstrated to be capable of efficient catalysis. Previously described mimochrome proteins were further designed for better accessibility of the heme site and to create a cavity for substrate accommodation. Immobilized to a coated gold electrode, the resulting mimochrome showed substantial electron transfer and in solution the protein displayed efficient peroxidase activity comparable to natural enzyme with several substrates in different media (*e.g.* water, buffer, buffer with 50 % trifluoroethanol – **Figure 1.9g**).<sup>131,132</sup> The maquette approach was also further extended to catalyze peroxidase activity.<sup>133</sup> Backbone of the protein was covalently grafted with a *c*-type cytochrome facilitating the fully assembling of the functional protein *in vivo*. The resulting four-helix maquette, named C45, retains oxidation and oxidative dehalogenation activities coupled with  $\text{H}_2\text{O}_2$  for a diverse set of substrates and is as proficient as certain natural peroxidases, but with an enhanced chemical and thermal stability (**Figure 1.9g**).

Based on precise knowledge for the installation of diverse cofactors in *de novo* protein scaffold and their highly efficient intrinsic catalytic properties, a number of successful protein catalysts were produced for a range of chemical reactions. However, this activity is completely dependent on the binding and the nature of the cofactor, thus limiting to some extent the scope of possible catalyzed reactions.

**Design of *de novo* catalytic proteins lacking cofactors.** Although designing *de novo* catalytic proteins without cofactors may be a more challenging task, strategies used by enzymes such as proximity effect and  $\text{p}K_a$  depression can be envisaged and employed in a

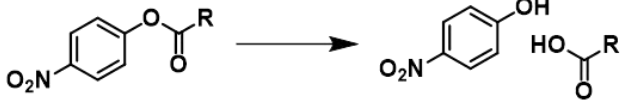
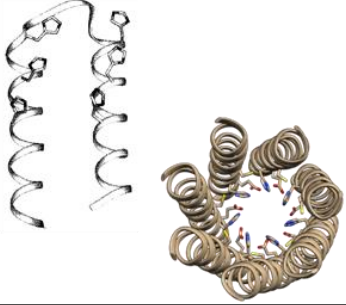
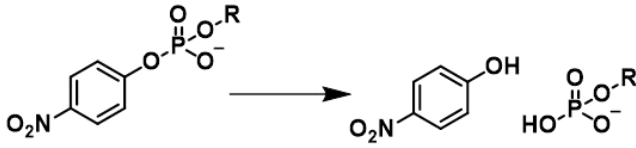
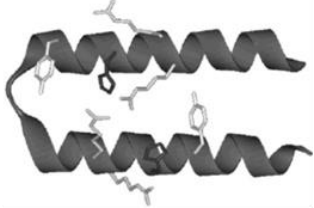
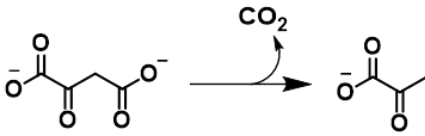
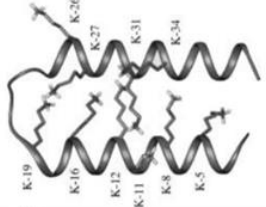
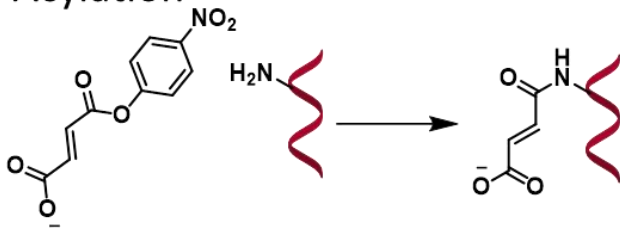
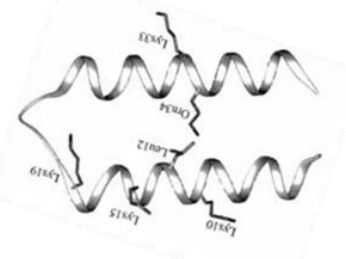
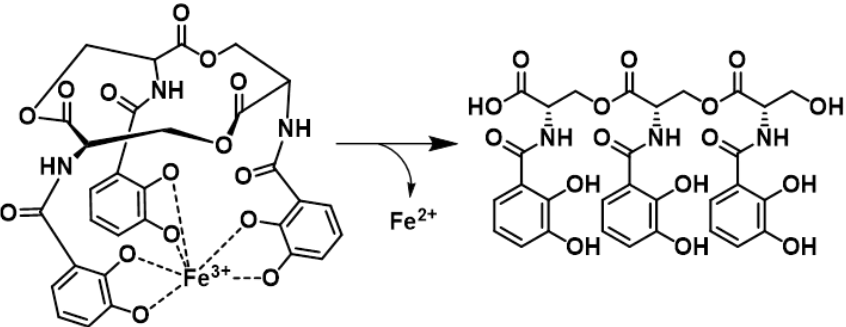
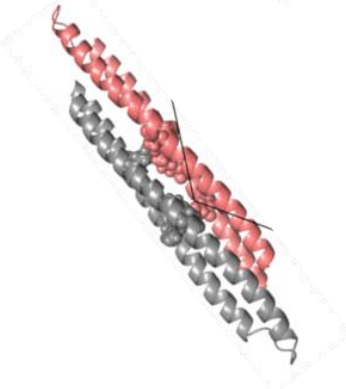
*de novo* protein scaffold. Previously, the design of a synthetic peptide was reported to catalyze condensation of peptide fragments through Native Chemical Ligation (NCL) mechanism with high sequence- and diastereoselectivity due to preorganization of substrates on the catalyst surface (Figure 1.10a).<sup>134</sup> Rate acceleration of 4100 times over uncatalyzed reaction was observed without the need of catalytic residue but only as a result of strong hydrophobic interactions and charge complementarity forming a specific coiled-coil assembly between the catalyst and specific substrates.<sup>135</sup> The bimolecular reaction is mainly driven by the high effective concentration and by the entropy gain, both coming from the proximity effect through coiled-coil assembly. Further investigations developed this concept towards self-replicating systems possessing dynamic error-correcting properties,<sup>136</sup> chiroselective amplification ability for the production of homochiral peptides from racemic mixtures<sup>137</sup> and conditional selectivity allowing amplification of one or more products depending on the reaction conditions (*i.e.* pH and salt concentration of the medium).<sup>138</sup> Interestingly, these peptide-based systems describe self-replicating and self-organized chemical networks with characteristics essential for evolution and may support their putative role in the origin of life. This supramolecular approach based on self-assembling designed peptides was also exploited to accelerate aminoacyl transfer, a process used by ribosomes and nonribosomal peptides synthetases (Figure 1.10b).<sup>139</sup> In the reaction mechanism, different features of natural enzymes were mimicked such as covalent catalysis with aminoacyl- $\alpha$ -thioester



**Figure 1.10. Catalysis due to coiled-coil assembly.** **a)** Self-replication system, where the Native Chemical Ligation has been catalyzed via proximity effect mediated by hydrophobic interactions (not represented) and charge complementarity during coiled-coil assembly. Charges are depicted in blue (positive) and in red (negative). **b)** Aminoacylation of lysine residues driven by proximity effect (with the anchored cysteine via coiled-coil formation),  $pK_a$  modulation and general acid/base contributions (not presented).

substrate loaded on a catalytic cysteine, precise positioning in close proximity of the aminoacyl donor (*i.e.* aminoacyl substrate anchored to the cysteine) and the acceptor (*i.e.* side chain of a lysine),  $pK_a$  modulation by placing the lysine in the hydrophobic core of a four-helix bundle, general acid/base contributions of the surrounding residues.<sup>140</sup>

Manipulating the reactivity of an individual residue notably through  $pK_a$  depression is a common strategy employed in helical bundles to enhance catalytic efficiency. In early work, a reactive site was engineered on the surface of a four-helix bundle (antiparallel dimer of a helix-loop-helix peptide) and was composed of six histidines. Detailed analyses showed rate enhancements for the hydrolysis and transesterification with trifluoroethanol (used as acyl acceptor substrate and to enhance helical propensity) of different *p*-nitrophenyl esters thanks to a cooperativity of unprotonated and protonated histidine residues allowing nucleophilic catalysis and transition state stabilization, respectively (**Figure 1.11a**).<sup>141</sup> Additional studies were undertaken to better understand and control this cooperativity between  $pK_a$  depression, TS conformation and general acid/base of a specific network of amino acids around the HisH<sup>+</sup>-His catalytic motif.<sup>142,143</sup> In the same *de novo* scaffold, another intricate constellation of catalytic amino acids consisting of two histidines flanked by four arginines and two tyrosines was designed to accelerate the hydrolysis of activated alkyl aryl phosphate diesters and unactivated dialkyl phosphate diesters (**Figure 1.11b**).<sup>144</sup> Supplementary examples using this dimer of helix-loop-helix motif demonstrated that  $pK_a$  modulation is now well controlled in order to, for instance: *i*) catalyze oxaloacetate decarboxylation through imine formation on a lysine residue positioned in the hydrophobic core and flanked by an arginine amino acid (**Figure 1.11c**);<sup>145</sup> or *ii*) accelerate the rate of lysine and ornithine acylation by increasing the nucleophilicity of these residues and by  $pK_a$  depression (**Figure 1.11d**).<sup>146</sup> More recently, a functional catalytic Cys-His-Glu triad was installed for the first time into a *de novo* protein scaffold in the laboratory of D. Woolfson.<sup>147</sup> A designed coiled-coil heptamer showed to be a robust framework as it resisted to the incorporation of one triad on each monomer, meaning a total of 21 mutations in the hydrophobic core of the barrel, by retaining substantial stability and it presented an accessible channel convenient for substrate accessibility to the active site. Resulting hydrolytic activity of *p*NPA relying on a covalent catalysis of cysteine residues was on par with redesigned enzymes based on natural scaffolds (**Figure 1.11a**).<sup>63,70</sup>

Reaction type	Protein structure	Ref.
<p>a) Ester hydrolysis</p> 		<p>141</p> <p>147</p>
<p>b) Phosphodiester hydrolysis</p> 		<p>144</p>
<p>c) Oxaloacetate decarboxylation</p> 		<p>145</p>
<p>d) Acylation</p> 		<p>146</p>
<p>e) Ferric enterobactin hydrolysis</p> 		<p>148</p>

**Figure 1.11. Examples of *de novo* catalytic proteins working without cofactors.** **a)** Hydrolytic activity has been installed into a helix-loop-helix dimer (only one monomer shown) and into the lumen of a coiled-coil heptamer. **b)** Hydrolysis of various phosphodiester has been catalyzed by a *de novo* four-helix bundle (only one monomer shown in the model). **c)** Oxaloacetate decarboxylation has been catalyzed via imine formation on a catalytic lysine of a helix-loop-helix dimer (only one monomer depicted). **d)** Acylation of lysine and ornithine has been accelerated in strategic positions within a helix-loop-helix dimer (one monomer shown in the model). **e)** Catalysis of ferric enterobactin hydrolysis has been observed *in vivo* rescuing a life-sustaining function for a modified strain of *E. coli* unable to grow in iron-limited medium. Structures from figures of the cited papers or PDB: 5EZC.



Besides rational approaches to design catalytic functions, screening of large library of *de novo* sequences was also explored to generate non-natural scaffolds carrying a desired catalytic activity. For example, Hecht laboratory found one construct within a large library of *de novo* helical bundles that hydrolyzed the siderophore ferric enterobactin rescuing a life-sustaining function for a modified strain of *E. coli* unable to grow in iron-limited medium (see **Figure 1.11e**).<sup>148</sup> Although the *de novo* catalytic protein was selected via screening of a large library of variants and not by rational design, this is a rare example of an unnatural protein that is efficient enough for maintaining living cells.

Miniaturized *de novo* protein scaffolds demonstrated great and promising catalytic activities because of increasing fundamental knowledge of natural enzyme capabilities. *De novo* frameworks are generally very robust and modular allowing to control every single modification. It is now possible to construct precise networks of amino acids to tune the reactivity of a catalytic residue, which is one of the essential features of natural enzymes.



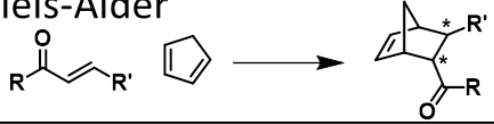
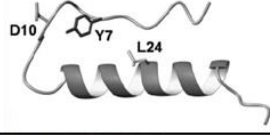
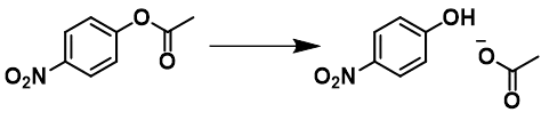
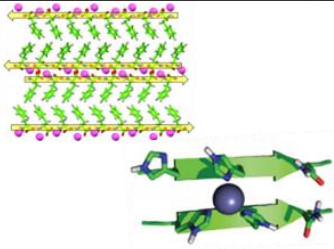
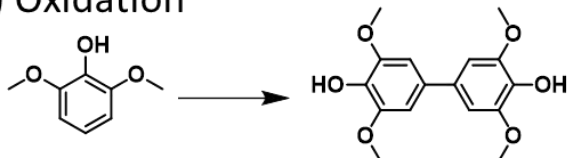
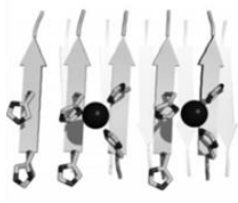
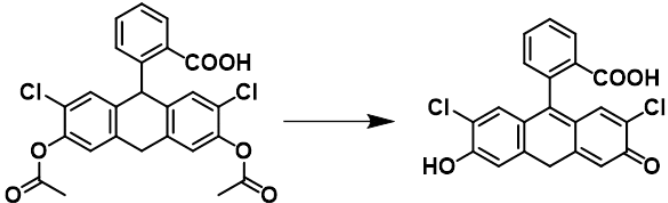
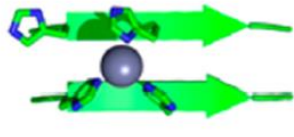
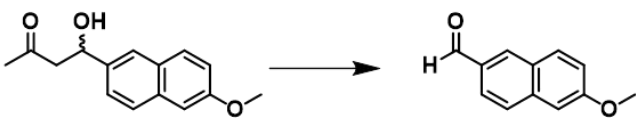
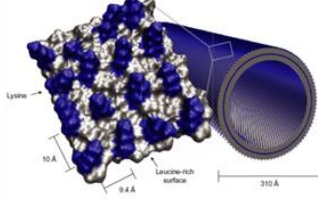
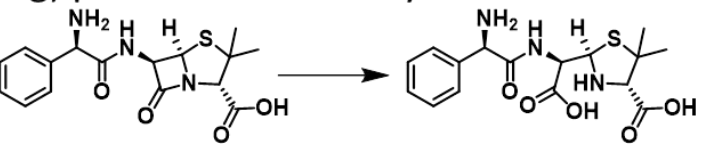
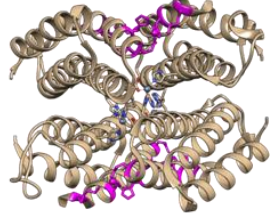
**Going further with small-size *de novo* proteins.** The robustness and the modularity of *de novo* protein scaffolds allow for using multiple strategies to install and enhance catalytic activity. As for enzyme design based on natural protein framework, directed laboratory evolution can be employed to improve the initial activity. Generally, this iterative process would require a monomeric protein for a well-defined and controlled folding and assembly *in vivo*, whereas the *de novo* scaffolds are often formed by oligomerization of several smaller sub-units such as the widely employed helical bundle arrangement. As discussed before, directed evolution was successfully applied to an artificial four-helix monomeric protein and demonstrated the possibility to use this technique starting from a *de novo* miniaturized scaffold to obtain highly proficient catalytic proteins for hydrolysis of ester even surpassing activities of natural enzymes.<sup>124</sup>

One advantage of small-size *de novo* proteins is their accessibility by total chemical protein synthesis allowing an easier and controlled incorporation of non-canonical amino acids within the sequence in order to expand the scope of reactivity for example. A cofactor-type amino acid was introduced in a small designed protein with a  $\beta\beta\alpha$  motif. The pyridoxamine cofactor was utilized as an amino acid chimera notably to have a good control of the incorporation and the position of the cofactor in the synthetic protein.<sup>149</sup> Influence of the surrounding amino acids on the activity was evaluated with the synthesis of various variants and this resulted in an artificial protein carrying transamination activity for the conversion of pyruvic acid to alanine (see **Figure 1.12a**). The addition of copper ion, a known chelator of transamination intermediates, provoked further acceleration of the reaction with the change in the stereoselectivity. In addition, non-natural amino acids capable to chelate copper ions (*i.e.* L-3-pyridylalanine and L-4-pyridylalanine) in order to graft an active site

were introduced on a small natural scaffold of bovine pancreatic polypeptide composed of a polyproline helix backfolded on a  $\alpha$ -helix.<sup>150</sup> The resulting metalloenzyme catalyzed Diels-Alder and Michael addition with good control of enantioselectivity (**Figure 1.12b**).

Large supramolecular peptide self-assemblies like amyloids is another tool to produce efficient catalytic material with substantial activity. Simple peptides can self-interact to form structure with diverse physical (e.g. enhanced thermal stability) and structural characteristics and thus, generate a high catalytic activity sometimes exceeding those of natural enzymes by weight.<sup>151</sup> The self-assembled motif can be actually as simple as a single amino acid coordinated to metal ions and still exhibit catalytic activity. Phenylalanine molecules showed coordination with zinc ions in a [2:1] ratio forming then a supramolecular amyloid-like cross- $\beta$ -sheet superstructure.<sup>152</sup> The surface of the architecture presented a dense ordered array of zinc ions accessible by substrates resulting in remarkable hydrolase activity of *p*NPA (**Figure 1.12c**) and CO<sub>2</sub> hydration activity. Thanks to its catalytic abilities such as lowering  $pK_a$  of bound water and stabilizing and positioning hydroxide of the substrates for reactions, zinc ions were also utilized with a  $\beta$ -sheet-forming heptapeptide comprising alternated hydrophobic and hydrophilic residues.<sup>153</sup> Such amphiphilic pattern is known to form extended amyloid  $\beta$ -structures via hydrophobic interactions all oriented on the same side. On the other hand, hydrophilic residues were selected for their capabilities to coordinate zinc ions leading to efficient catalysis of acyl ester hydrolysis with *p*NPA as referring substrate (**Figure 1.12c**). By binding copper ions on similar superassembly of heptapeptides, the range of reactivities was expanded to a substrate oxidation via oxygen activation (**Figure 1.12d**), phosphoester hydrolysis and tandem reaction of hydrolysis and oxidation (**Figure 1.12e**).<sup>154,155</sup> Interestingly, the supramolecular self-assembling peptide architecture can carry out a set of activities even with similar structure due to the fact that small changes propagate along amyloid structure and generate larger modifications in the macromolecular assembly.<sup>151</sup> By exploiting nanotube-like superstructure and cross- $\beta$  grooves on the surface, adjustments of individual amino acids displayed the possibility to catalyze a variety of chemical reactions including enantioselective transformation such as condensation and retro-aldol cleavage (**Figure 1.12f**).<sup>156</sup> Finally, supramolecular protein assembly is not only reserved to  $\beta$ -sheets but can also be applied to  $\alpha$ -helices. For instance, monomeric four- $\alpha$ -helix bundle assembled to form a zinc-mediated higher-ordered tetramer that bore metallo- $\beta$ -lactamase activity thanks to catalytic zinc sites at the interface of the assembly in addition to the structural zinc sites (**Figure 1.12g**).<sup>157</sup> This supramolecular assembly showed *in vivo*  $\beta$ -lactamase activity, which was useful for functional screening and optimization through directed evolution and which conferred ampicillin resistance to *E. coli*.

*De novo* protein scaffolds show again their remarkable modularity and tunability of their catalytic properties. Non-proteinogenic amino acid incorporation enables larger scope of

Reaction type	Protein structure	Ref.
a) Transamination 		149
b) Diels-Alder 		150
c) Ester hydrolysis 		152 153
d) Oxidation 		154
e) Hydrolysis and oxidation tandem 		155
f) Retro-Aldol 		156
g) $\beta$ -lactamase activity 		157

**Figure 1.12. Extra catalytic activity brought by non-natural amino acid incorporation and peptide self-assembly.** **a)** A pyridoxamine cofactor has been introduced in a peptide sequence leading to the transformation of pyruvate to alanine. **b)** Diels-Alder reactions have been catalyzed by a peptide thanks to the incorporation of copper-binding amino acid like L-3-pyridylalanine and L-4-pyridylalanine. **c)** Rate of pNPA hydrolysis has been accelerated by a phenylalanine/zinc  $\beta$ -sheet superstructure or by heptapeptide prone to self-assemble and to bind zinc ion. **d)** Oxidation of dimethoxyphenol has been catalyzed by superassembly of heptapeptides. **e)** Similar superstructure of heptapeptide carried catalytic activity for tandem reaction of hydrolysis and oxidation. **f)** Retro-Aldol reaction has been catalyzed by peptide-forming nanotubes. **g)** A zinc-mediated self-assembly of four-helix bundle unit resulted in efficient  $\beta$ -lactamase activity *in vivo*. Catalytic zinc sites are depicted in magenta. Structures from figures of the cited papers or PDB: 4U9D.

reactivity. Supramolecular assembly enables rate enhancements and atom economy to generate catalytic active sites when compared to the molecular weight of a typical enzyme; however lack of high-resolution structural characterization of active fibrils slows down development towards efficiency enhancement. Additionally, as for enzyme design based on natural scaffolds, directed evolution has proven to be an indispensable technique for further catalytic efficiency improvement.

**Future challenges.** Despite the number of successful examples of small *de novo* catalytic proteins carrying rate enhancements for a given chemical reaction, the catalytic efficiency is generally quite far from the natural enzyme counterparts. Employed strategies are different compared to computational enzyme design based on natural protein scaffolds and TS stabilization, and usually rely on either controlled positioning of cofactors, creation of a non-specific hydrophobic pocket for substrate accessibility to the active site, proximity effects driven by specific coiled-coil assembly,  $pK_a$  modulation, nucleophilic catalysis or general acid/base contributions. However, sub-Å structural characterization is often missing, which may complicate further design and slow down the iterative process of catalytic improvement. Additionally, important catalytic strategies employed by enzymes are still challenging to design and to incorporate into *de novo* protein framework. Dynamics are mainly missing and precise amino acids constellation can still be further improved to consider more electrostatics and to fine-tune vibrational modes helping to reach the transition state and for hydrogen-tunneling notably. To do so, multiple states enzyme design will need to be better elaborated for such purposes.<sup>158,159</sup> Multistate design takes into consideration multiple conformational states or includes positive and negative design principles but still needs to be implemented to the process of fully *de novo* design of a catalytic protein. Nevertheless, dynamic multistate design is under progress and recently, a protein was successfully engineered to switch between two novel targeted conformations on a millisecond timescale.<sup>160</sup> Finally, *de novo* catalytic proteins were designed to accelerate only a limited set of chemical reactions and were essentially focused on a reduced number of reaction models such as oxidation, hydroxylation, *p*-nitrophenyl ester and phosphodiester hydrolysis, CO<sub>2</sub> hydration, oxaloacetate decarboxylation or Diels-Alder cycloaddition. Therefore, future *de novo* functional design would have to aim at more challenging multistep and tandem reactions.

#### 1.1.4. ***Incorporation of a catalytic function on small-size structured and non-proteinogenic scaffolds***

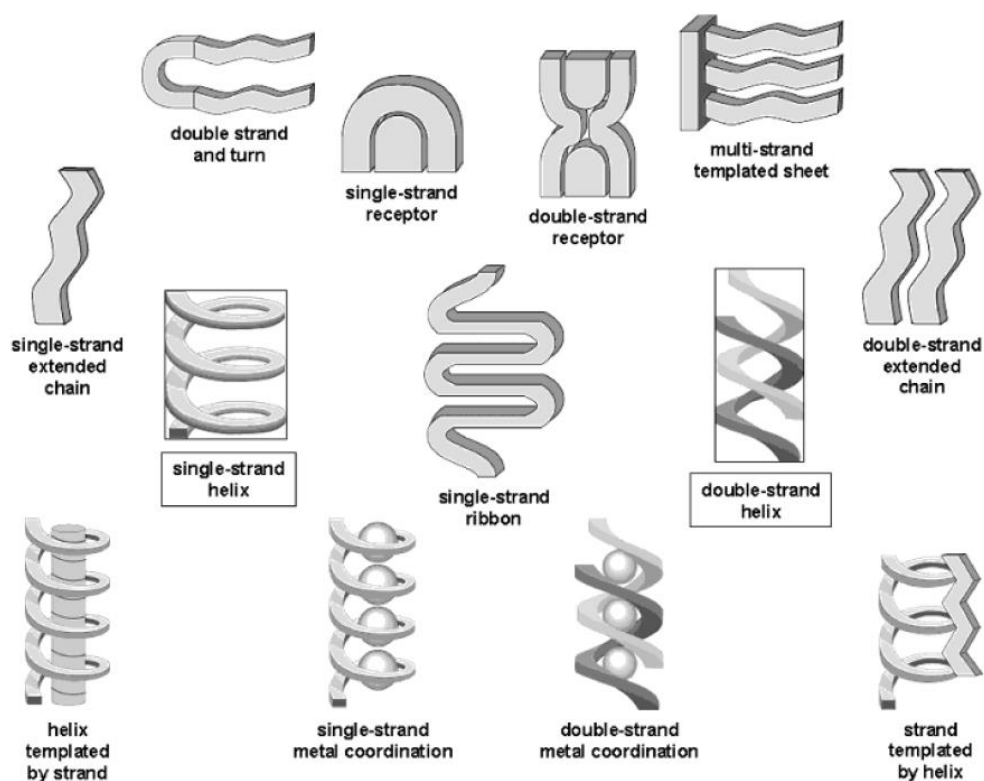
**Ribozymes.** Since the discovery of the catalytic properties of RNA molecules (Nobel Prize of Chemistry in 1989), their important catalytic role in cellular functions was explicitly demonstrated.<sup>161</sup> Indeed, RNA molecules have the ability to fold into complex structures forming binding sites for metal ions and substrates and thus forming potential active sites. Natural RNA catalysis encompasses protein biosynthesis by the ribosome, RNA strand scission and RNA splicing of introns involving excision of the intron sequence followed by ligation of the flanking exons.<sup>162</sup> To do so, RNA molecules utilize various catalytic strategies including substrate orientation to achieve TS, electrostatic stabilization of TS, general acid/base contribution and metal ions assistance. Due to the pervasive roles of RNA in modern biology, the primordial theory of “RNA world” has emerged in which the RNA molecules would have performed dual function: storing the genetic information in the sequence of RNA and the catalytic properties defining the derived phenotype.<sup>163</sup> Hypothetically, RNA was able to catalyze its own replication and primitive metabolic function, though this theory is still under debate.<sup>164</sup>

Interestingly, artificial ribozymes and deoxyribozymes were produced to catalyze numerous of chemical reactions through *in vitro* evolution and selection from combinatorial libraries of nucleic acid sequences.<sup>165</sup> For example, a selected ribozyme that presents two hairpin loops and one internal loop catalyzed C-C bond formation between two organic molecules via Diels-Alder cycloaddition.<sup>166</sup> This ribozyme performed the reaction with multiple turnovers and with good enantioselectivity like protein catalyst. The enantioselectivity was reversed by employing a synthetic “mirror-image” of the ribozyme. A hammerhead-based ribozyme, which is a common structure found in catalytic RNA, was designed to cleave a synthetic RNA substrate under allosteric control.<sup>167</sup> Here, the authors took the advantage of RNA that can specifically bind DNA or RNA with high accuracy through sequence complementary and designed a highly specific DNA effector to trigger the formation of the active state.

RNA can form convenient artificial scaffold to catalyze numerous chemical reactions and supports enzyme features such as enantioselectivity and allosteric control. RNA catalysts can also be subjected to directed evolution to increase the catalytic efficiency. However, the five natural nucleotides (adenine, cytosine, guanine, thymine and uracil) limit the structural and chemical diversity. High-resolution structural characterization is also harder to obtain for RNA or DNA-based molecules, thus a deep understanding of the relationship between structure and function is still missing compared to protein catalysts.

**Foldamers.** Foldamers are a novel class of molecules consisting of any oligomer that folds into a well-defined three-dimensional structure in solution.<sup>168,169</sup> Foldamers are based on a broad pallet of building blocks increasing the range of potential conformations (as shown in **Figure 1.13**). Foldamers are generally related to an element of secondary structure and present two main classes of conformation: single-stranded foldamers such as peptidomimetics and multiple-stranded foldamers that fold and need to be associated to another partner like nucleotidomimetics. Foldamers is a large family of molecules and support a wide diversity of sizes, shapes and arrangements, which offers the possibility to create new synthetic catalysts and could expand the range of applications over protein catalysts.

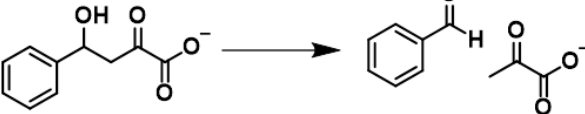
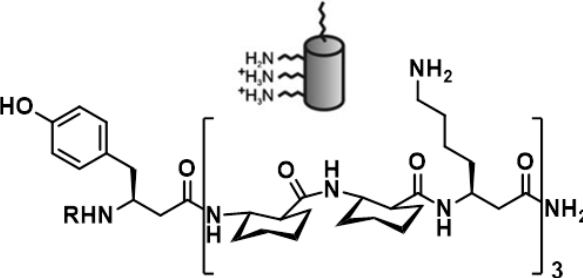
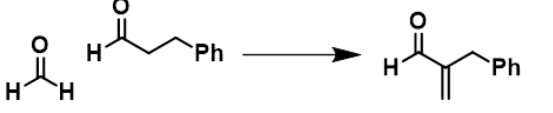
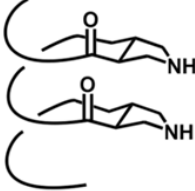
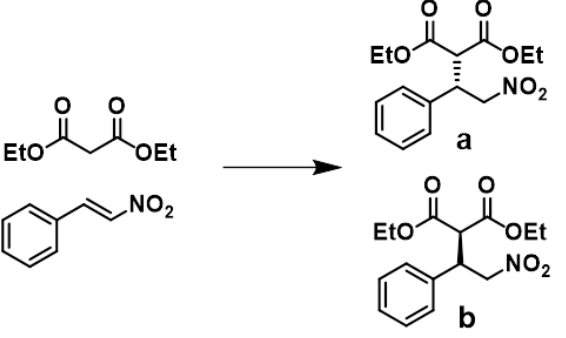
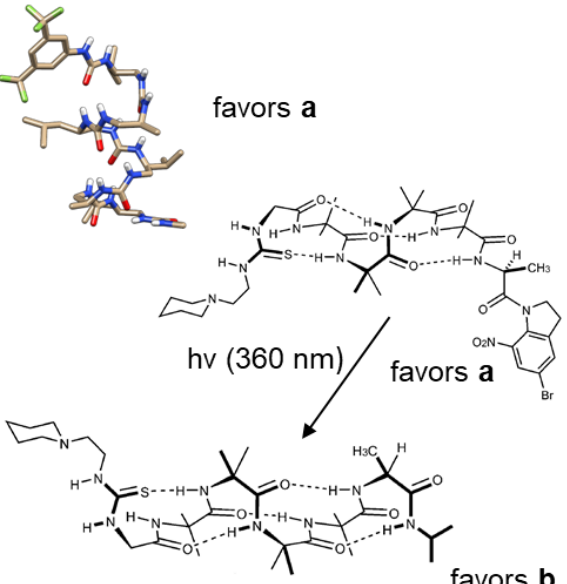
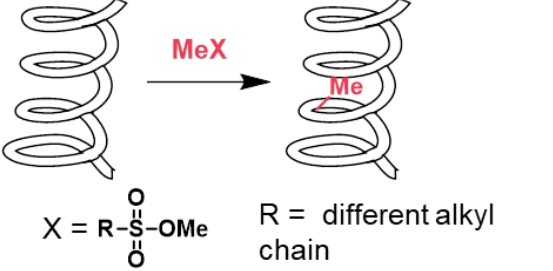
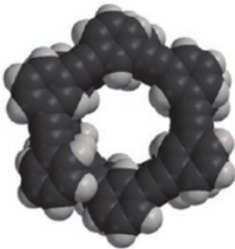
Peptidomimetics resemble peptide structure, therefore they control the spatial disposition of functional groups and are considered as aliphatic foldamers. Peptidomimetics have generally a modified backbone compared to native peptides such as  $\beta$ -peptide foldamers in which the amino group is linked to the  $\beta$ -carbon and not to the  $\alpha$ -carbon as in natural amino acid. A  $\beta$ -peptide composed of cyclically constrained *trans*-2-aminocyclohexanecarboxylic acid favoring helical conformation was a good framework to align  $\beta^3$ -homolysine along one side of the helix.<sup>170</sup> The close proximity of the amine groups of  $\beta^3$ -homolysine side chains was exploited to accelerate a retro-aldol reaction via  $pK_a$



**Figure 1.13.** Different possibilities of secondary structure reached by foldamers. There can be single-stranded foldamers as well as multiple-stranded foldamers. *The figure is reprinted with permission from Figure 1 of Hill, et al. (2001) Chem. Rev., 101, 3893-4011. Copyright (2020) American Chemical Society.*

depression catalysis and imine formation (**Figure 1.14a**). Helices were also obtained with oligomers composed of  $\alpha$ - and  $\beta$ -amino acids and more particularly of cyclically constrained pyrrolidine-derived  $\beta$ -amino acids for tuning the secondary structure. With a good placement and orientation of the pyrrolidine units along the helix, the resulting reactive dyad showed to work in tandem to catalyze crossed aldol reaction (see **Figure 1.14b**).<sup>171</sup> A similar catalytic foldamer was explored for macrocycle ring closure.<sup>172</sup> Moreover, helical oligourea foldamers were reported by Guichard and Palomo groups to be efficient and enantioselective catalysts for Michael addition of carbonyl pronucleophiles to nitroalkenes through synergic activation with an achiral base and the chiral H-bond donor foldamer (**Figure 1.14c**).<sup>173</sup> High selectivity strongly correlated with helix folding and modularity of the system were clearly demonstrated, which proved that foldamers are robust framework for chiral catalysis in the same way as peptide-based catalysts. Clayden and coworkers went a step further by proposing a switchable foldamer catalyst via photochemical deprotection causing an insertion of a single H-bond, which enabled interconverting the screw sense of the oligomer resulting in a change of the enantioselectivity of the catalyzed Michael addition reaction (**Figure 1.14c**).<sup>174</sup> Although the latter was a quite poor catalyst with an important catalyst loading necessary compared to the previous oligourea (10 mol % and 0.1 mol %, respectively), it demonstrated the possibility to insert switch and so to finely control reaction. Another promising area offered by foldamers is their capacity to form helical cavity that would serve to isolate the substrate from the solvent or as reactive sieves favoring catalysis with the ideally sized substrates over smaller or larger substrates.<sup>175</sup> Moore and collaborators challenged this concept with *m*-phenyleneethynylene oligomers forming a helical conformation and demonstrated rate enhancement depending of the size and the shape of the substrate even if the system needs to be further developed to have a true substrate discrimination and to match with natural enzymes (**Figure 1.14d**).<sup>176</sup>

Foldamers represent a wide field of research including small and structured molecules. Thanks to the large range of building blocks, foldamers can reach a variety of secondary structures. Design of catalytic foldamers is still in its infancy but promises to supplement the scope of reactions and possibility of *de novo* catalytic peptides.

Reaction type	Foldamer structure	Ref.
<p>a) Retro-Aldol</p> 		170
<p>b) Crossed aldol</p> 		171
<p>c) Michael addition</p> 		173  174
<p>d) Reactive sieves</p> 		175

**Figure 1.14. Catalysis by various type of foldamers.** **a)** A reaction of retro-aldol has been catalyzed by a helical  $\beta$ -peptide, in which  $\beta^3$ -homolysine have been aligned resulting in  $pK_a$  depression. **b)** Pyrrolidine-derived  $\beta$ -amino acids have been well positioned in a helical structure allowing working in tandem for crossed aldol catalysis. **c)** Michael addition has been enantioselectively catalyzed by helical oligoureia foldamers or by switchable foldamers causing inversion of the enantioselectivity. **d)** *m*-Phenyleneethynylene oligomers have worked as reactive sieves, in which methylation was more or less enhanced depending of the size and the shape of the reagent. Structures or structural models from figures of the cited papers or CSD: DEKLUD.



### 1.1.5. Conclusions

Since decades, a large number of proteins has been reengineered or even designed from scratch to carry tailor-made catalytic activity. But the robust and routine design of artificial enzymes with activity comparable with natural enzymes is an unresolved problem for the moment. Still, design of new catalytic activity has provided some new insights towards the understanding of enzyme function. In this section, I have compared the design of enzymes based on natural protein scaffolds and the entirely *de novo* catalytic proteins built from artificial framework.

On one hand, computational enzyme design on naturally occurring proteins as well as catalytic antibodies have initially focused their efforts on stabilizing the transition state(s). The resulting catalytic efficiencies were often found to be far from natural enzymes, thus, this strategy alone is clearly not robust enough. General deficiencies in the designed enzymes are attributed to a lack of precise placement of catalytic residues and a lack of implementation in the design process of backbone behavior, long range interactions and protein conformation. Several iterative cycles for further optimization, in particular by directed evolution, are able to improve catalytic activities. Nevertheless, this approach is quite time consuming and includes advanced computational calculations and several rounds of laboratory evolution. Additionally, complexity of such larger natural protein scaffolds has not been helpful for precise understanding and to use it as a practical tool.

On the other hand, fully *de novo* catalytic protein, where scaffold and function are designed from scratch, provides a valuable alternative for a better comprehension of enzyme catalysis and for practical applications. Designing from scratch a simplified three-dimensional structure especially in the case of helical barrels is now quite well understood, subsequently, the functionalization of such scaffolds has been widely demonstrated. Importantly, the stability of *de novo* scaffolds is entirely controlled and the contribution of each side chains for folding, stability and catalytic function can be easily evaluated. The ability to perform catalysis as efficient as some natural enzymes (e.g. the maquette protein C45 with peroxidase activity comparable to some natural counterparts)<sup>133</sup> with extremely simplified framework suggests that a defined structure allowing precise preorganization of catalytic residues may be more critical than the complexity of the protein topologies. This was also demonstrated by the extensive use of helical structures and coiled-coil assemblies that allow for precise and controlled placement of each amino acid. Moreover, the stepwise modifications of such miniaturized scaffolds may provide appropriate models for the emergence of primitive enzymes (e.g. zinc-mediated dimerization leading to the formation of a cleft pertinent for the installation of catalytic activity and further optimized by directed evolution).<sup>123,124</sup> Generally, *de novo* simplified catalytic proteins lack specific and buried

substrate binding site, which may reduce the catalytic efficiency due to high  $K_m$  values but may also yield to promiscuous activity attractive to develop general and practical tools. Finally, foldamer molecules have supported the proposition that a small and structured scaffold can accommodate efficient catalytic properties. For example, a chiral scaffold with a defined twisted orientation is sufficient to transfer the chirality to small molecules for enantioselective reactions. The modularity and ease of modification of foldamers and *de novo* simplified proteins have also resulted in a fine-tuning of such activity.

Of course, many challenges need still to be resolved such as the incorporation of dynamics into *de novo* protein scaffold or also the possibility to catalyze difficult multistep reactions. But, the utilization of such miniaturized protein remains a suitable tool for fundamental research as well as applied research for the development of new technologies.

## 1.2. Amide bond formation, a common but challenging reaction

Amide bond is widespread in many synthetic and natural molecules. Although amide linkage is prevalent in biomolecules and more particularly in proteins in which peptide bond links two amino acids, this moiety is also broadly employed in drugs such as paracetamol and penicillin,<sup>177,178</sup> and in synthetic polymers like nylon.<sup>179</sup> Nature and synthetic chemists have indeed privileged amide due to its high stability towards various conditions such as high temperature and multiple chemicals.<sup>180</sup> Radzicka and Wolfenden have estimated that, for instance, the peptide bond of the dipeptide glycylglycine is cleaved by hydrolysis with half-life of more than 300 years in neutral buffer at room temperature (25 °C).<sup>181</sup> This is mainly attributed to the property of amide bond to form resonating structures by delocalizing the lone pair of the nitrogen into the carbonyl group, thus providing a partial double bond and planar character of amide between nitrogen and carbon (as shown in **Figure 1.15**).<sup>180,182</sup> Furthermore, amide bond possesses a large dipole moment due to the great electronegativity of oxygen and to a less extent of nitrogen too. This confers interesting properties to the amide bond, which can act as H-bond donor through the oxygen as well as H-bond acceptor via NH. These properties are notably exploited by proteins to form secondary structures like  $\alpha$ -helix and  $\beta$ -sheet but also serve for drug molecules to interact with the target for example. However, despite the ubiquity of amide bond, its formation persists to be a synthetic challenge.<sup>183</sup> Amide bond synthesis is basically a condensation between a carboxylic acid and an amine. Nevertheless, mixing these two entities together leads to an acid-base reaction with the production of a stable salt.<sup>184</sup> To overcome the thermodynamic barrier, nature has evolved complex machines as introduced hereinafter. This section will then present the development of chemical techniques that mainly focus on carboxylic acid activation. We will finish by introducing methods for total synthesis of proteins involving peptide bond formation.

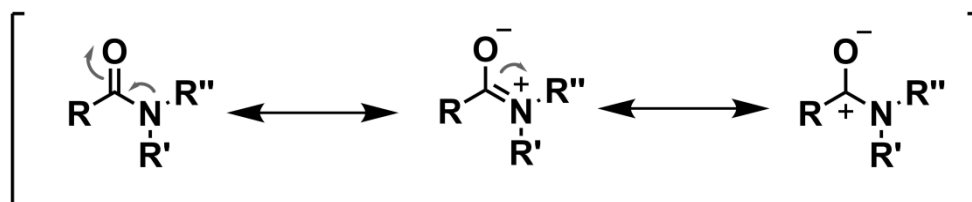


Figure 1.15. Resonance structures of an amide bond.

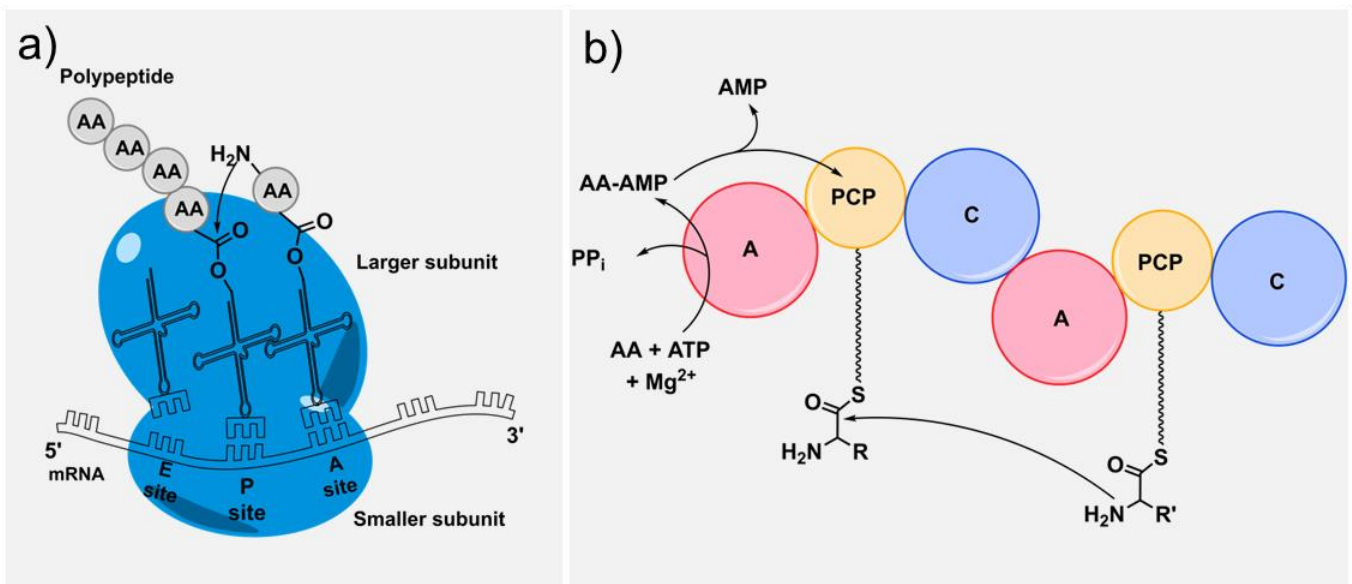
### 1.2.1. *Natural and biological synthesis of amide bond*

To efficiently accelerate the formation of amide, living organisms involve complex macromolecular machines including, for instance, the ribosome for protein synthesis and peptide bond formation. Ribosomes are large macromolecular complexes composed of multiple RNAs and associated proteins and divided into two subunits, a smaller subunit binding the messenger RNA (mRNA) and a larger subunit binding the aminoacylated transfer RNAs (tRNAs) (see **Figure 1.16a**).<sup>185</sup> A particular protein is encoded by a gene sequence that is first transcribed to a specific mRNA. The ribosome produces proteins by translating the proper mRNA used as template and more precisely by decoding the mRNA codon per codon in a 5'-3' direction in order to determine the sequence of amino acids. A codon is formed of three nucleotides and is related to a specific amino acid. Due to the degeneracy of the genetic code (64 different codons for 20 amino acids and stop), different codons can code for the same amino acid. A codon is recognized by a tRNA through its anticodon part, which delivers the appropriate amino acid to the ribosome. Each tRNA is specifically loaded by the correct amino acid via an ester bond thanks to the aminoacyl tRNA synthetase. Moreover, the ribosome contains three sites for tRNA binding (**Figure 1.16a**).<sup>186</sup> First, the charged aminoacyl-tRNA recognizes and binds the corresponding codons of the mRNA in the A site (aminoacyl site) of the ribosome. The peptide bond is then formed between the amino acid of the tRNA of the A site and the growing peptide charged in the tRNA of the P site (peptidyl site). The growing polypeptide chain is so transferred to the tRNA of the A site. Finally, translocation occurs to move the free tRNA from the P site to the E site (exit site) and the tRNA with the polypeptide attached to the P site. Thus, this complex machinery enables to join the different amino acids via the formation of amide bond to assemble the entire polypeptide chain in an N-to-C-terminal direction. At the end of the synthesis, the target protein is released and upon correct folding serves its predefined biological function.

Amide formation for the assembly of polypeptides takes also place in the non-ribosomal peptide synthesis. The non-ribosomal synthesis is generally found in microorganisms like bacteria and fungi and is independent of mRNA allowing the production of a wide range of peptide metabolites that bear cyclic and/or branched structure and can contain non-proteinogenic amino acids (e.g. D-amino acids) and unusual backbone modifications (e.g. N-methylation).<sup>187,188</sup> Consequently, such synthesized peptide metabolites carry a variety of biological functions like antibiotics with gramicidin S or immunosuppressive activity with the cyclosporine A. Nonribosomal peptide synthetases are actually composed of a particular arrangement of modules. The identity and order of the modules specify the sequence of the assimilated units, the modification and chemistry occurring at each

incorporation. Therefore, each module is responsible for the incorporation of a building block in the growing chain and is defined by several domains that may be capable of other functions such as racemization or formylation. Three domains are commonly present for peptide elongation (like in **Figure 1.16b**). These domains are involved in the recognition and activation of the amino acid (adenylation (A) domain), the propagation of the growing chain between the catalytic centers (thiolation and peptidyl carrier protein (PCP) domain) and in forming amide bond (condensation (C) domain).<sup>189</sup> Interestingly, the growing peptide is covalently attached to the domains by a thioester linkage and amide bond formation with new amino acids proceeds through reaction of the amino group of the latter on thioester. It is noteworthy that thioester is a common moiety used to activate carboxylic acid as this group benefits from an enhanced reactivity towards most of the nucleophiles due to the poorer orbital interactions between sulfur atom and carbonyl moiety.<sup>190</sup> Finally, product release is catalyzed by the thioesterase (TE) domain by either hydrolysis or macrocyclization.

Activation in the form of acyl-thioester to catalyze amide bond is a common strategy found in other biological processes. For instance, ubiquitination of lysine residues by a dedicated enzyme cascade proceeds via ubiquitin-thioester activated species.<sup>191</sup> In other words, the ubiquitin E3 ligase catalyzes the formation of an isopeptide/amide bond between a lysine residue and the C-terminus of the ubiquitin protein covalently attached and activated through a thioester bond on the enzyme.<sup>192</sup> Sortase A, a transpeptidase from *Staphylococcus aureus*, is another enzyme where transiently formed thioesters play a key role. Sortase A catalyzes the covalent anchoring of surface proteins bearing an LPXTG recognition motif to



**Figure 1.16. Amide bond formation for polypeptide synthesis.** **a)** Simplified cartoon depiction of ribosomal protein synthesis. Proteins are assembled by following the mRNA template. The A site serves as an entry point for the specific charged aminoacyl-tRNA and elongation of the chain occurs in the P site by aminolysis of the next amino acid (AA). **b)** Schematic representation of the non-ribosomal peptide synthetase working by modules. Classically, modules are composed of at least three domains: the A domain for activation of the specific amino acid (AA), the PCP domain where the growing chain propagates through the different catalytic centers and the C domain catalyzing the amide bond formation by aminolysis of a thioester-activated carbonyl. AA stands for amino acid.

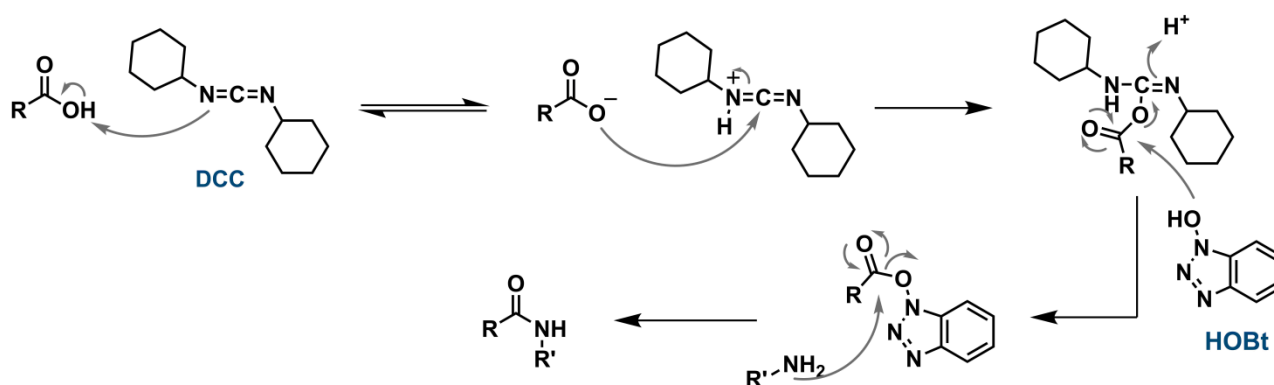
the cell wall.<sup>193</sup> After cleavage of the motif on the threonyl carboxylate by a catalytic cysteine, a covalently branched thioester adduct is formed between the cleaved protein and the enzyme allowing for the catalysis of an amide bond formation with the amino group of the pentaglycine peptide attached to a peptidoglycan. Butelase 1, a naturally occurring ligase, belongs to the family of asparginyl endopeptidase and carries the fastest peptide ligation activity found in nature.<sup>194</sup> Butelase 1 is mainly involved in the catalysis of cyclization of plant peptides and especially of cyclotides through amide bond formation. During biocatalysis, an HV dipeptide from the N/D-HV recognition motif is cleaved via the attack of the catalytic cysteine forming, here again, a thioester in the form of an acyl-Cys-enzyme intermediate, which favors the nucleophilic attack of the N-terminal amino acid leading to the native amide linkage.

Nature has evolved complex macromolecular machineries to catalyze amide bond formation. These machines are very specific and extremely proficient with high turnover. The complexity of such macromolecular assemblies like ribosomes reveals how challenging can be to catalyze amide bond in a remarkably efficient manner. These systems employ diverse sophisticated strategies to reach their effectiveness like activation of carboxylic acid via formation of activated ester or transient thioester, bringing in close proximity the activated acyl-moiety with the amino group, remove water from catalytic site, and so on. Chemistry researcher in the laboratory does not have the possibility to handle such elaborate molecular systems and needs to develop completely different techniques to form amide bond.

### 1.2.2. *State-of-the-art of chemical approaches to form amide bond*

As mentioned before, mixing a carboxylic acid with an amine leads to an acid-base reaction with the production of a stable salt. Organic chemistry has historically dealt with this thermodynamic barrier by completing direct condensation at high temperature.<sup>195–197</sup> Due to broad incompatibility with other functionalities, modern methods prefer acting through selective activation of the acid. Activation of the carboxy component can occur with a variety of reagents. Then, reaction with amine can proceed either after formation and isolation of the intermediate acylating agent, or after formation of the activated carboxy group with the immediate treatment with amine, or with activation of the acid generated *in situ* in presence of the amine. Development of new methodologies was mostly oriented towards the field of peptide synthesis, but organic molecules such as drugs also benefit from these improvements leading to a large choice of activation techniques that may be useful for challenging synthetic issues. Beside acyl azides and anhydrides, one of the most widespread

approaches is the generation of acyl chloride from carboxylic acid followed by the coupling reaction with the required amine.<sup>184</sup> However, the use of acyl chloride is rather limited for practical reasons, particularly in peptide synthesis, considering hydrolysis and racemization. Indeed, reporting of new coupling reagents has not ceased since the synthesis and usage of the first carbodiimide, the dicyclohexylcarbodiimide (DCC, in 1955), as activator of carboxylic acid.<sup>198</sup> To reduce epimerization over the formation of O-acylurea intermediate, the use of additives like 1-hydroxy-1H-benzotriazole (HOBT) has demonstrated to be effective in peptide bond formation (**Scheme 1.3**).<sup>199</sup> The activation of carboxylic acid through an ester complements carbodiimides approach by providing the active esters with good leaving groups prone to react with amine under mild conditions and normally with reduced racemization. Formed esters with pentafluorophenol,<sup>200,201</sup> or *N*-hydroxysuccinimide<sup>202,203</sup> are efficiently activated derivatives that can be used in various applications for amide bond formation. Indeed, NHS-activated ester derivatives are widely used as a bioconjugation tool for amino group modification through stable amide bond formation (*i.e.* N-terminus and  $\epsilon$ -amino group of lysine) and find multiple applications in fluorescent labelling and cross-linking of biomolecules (for elucidation of different kinds of interactions, for instance).<sup>204</sup> Nowadays, many alternative reagents generating the OBt ester *in situ* are used such as phosphonium salts and uronium/guanidinium salts.<sup>205</sup> Among the uronium/guanidinium salts, HBTU ((2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) activator is now well-established especially in solid-phase peptide synthesis (SPPS) with elaborated protocols to increase yield and reduce racemization (including precautions to add equimolar quantity compared to the carboxylic acid to avoid formation of guanidinium by-product by reaction with amine).<sup>206,207</sup> Basically, the mechanism of acid activation includes the attack of carboxylate anions on HBTU to form an O-acyl(tetramethyl)isouronium salt intermediate (**Scheme 1.4**). The released OBt ions react on the intermediate affording the activated OBt ester. Addition of amine leads to the formation of an amide bond with the acylating reagent. The HATU ((1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide

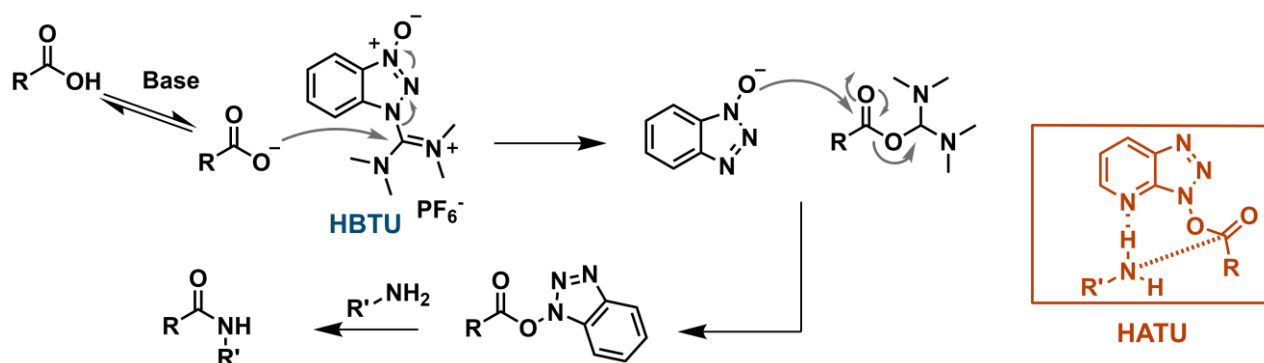


**Scheme 1.3.** Amide bond formation through DCC and HOBT activation of carboxylic acid. Addition of HOBT was demonstrated to reduce racemization.

hexafluorophosphate) derivative containing an azabenzotriazole moiety results in more rapid and complete couplings with even less epimerization compared to the counterpart HBTU and has proven to be more effective in difficult couplings.<sup>208</sup> It is postulated that this rate enhancement arises from stabilization of the amine through neighboring group effect of the extra nitrogen group of the azabenzotriazole cycle (**Scheme 1.4**).

However, these methods to activate carboxylic acid in order to generate amide are considered expensive and nonelegant in terms of poor atom economy due to stoichiometric activators and reagents, waste quantity and the general use of environmentally unfriendly solvent (*i.e.* *N,N*-dimethylformamide (DMF) and *N*-methylpyrrolidinone (NMP)). Consequently, the ACS Green Chemistry Institute has classified the formation of amide bond as a top challenge in organic chemistry and has encouraged researches for the development of catalytic or sustainable, meaning direct, amide formation in more desirable solvents.<sup>209</sup> Although numerous strategies have been attempted like boron-based catalysts,<sup>210,211</sup> metal catalysts,<sup>212</sup> biomimetic catalyst,<sup>213</sup> catalytic and oxidative amidation from alcohols or aldehydes,<sup>214,215</sup> they are still in their infancy and need to demonstrate their scalability. Beside organic and organometallic approaches, engineering of biocatalysts has proven to be highly potent for amide bond formation catalysis with high specificity, regioselectivity (no protecting group needed) and enantioselectivity.<sup>216</sup> But due to the very restricted scope of reactivity of such developed enzymes, applications are rather limited to few industrial reactions.

Despite the wide range of methods for amide bond formation, this reaction persists to be a synthetic challenge, and progress in terms of catalysis, waste reduction, chemoselectivity and scalability of new techniques can still be accomplished.<sup>183</sup> Nevertheless, available and common methodologies to form amide bond through carboxylic acid activation are robust and enable access of challenging reactions. Development in this area mainly benefits from effort made for peptide synthesis.



**Scheme 1.4. Amide bond formation via activation of carboxylic acid with HBTU.** The red box depicts the neighboring effect of HATU through extra interaction between nitrogen of the pyridine ring and the amine, which provides acceleration of the rate reaction and less epimerization compared to HBTU.



### 1.2.3. Total synthesis of proteins

The most important interest in controlling amide bond formation is probably the access to proteins. Total synthesis of peptides and proteins is of great interest for detailed studies of proteins of high biological significance in terms of folding, structure-activity relationship, as well as for synthesis of biomedically important peptides and proteins.<sup>217</sup> Thus, tools for protein production are essential for fundamental and practical research and will be generally presented in this subsection.

**Protein production via biotechnological processes.** The principle of protein production by biotechnology is based on the overexpression of a recombinant gene by a host cell machinery and was first completely demonstrated by Cohen, *et al.* in 1973.<sup>218-220</sup> Typically, the gene for the target protein is introduced into a host cell via a vector. The vector can be either a plasmid, viral or an artificial chromosome. The most commonly used vector, especially for bacterial expression systems, is the plasmid in which the recombinant DNA is inserted in a cloning site containing restriction sites for specific DNA cleavage followed by ligation of the target gene. The cloning vectors are then introduced to the host cell by a process of artificial induction of cell competence (e.g. salt, heat shock treatments or electroporation) followed by transformation with the genetic material. Plasmids generally also contain a selectable marker such as an antibiotic resistance gene allowing the transformed cells to grow in a selective media in contrary to non-transformed cells. The recombinant DNA is then expressed upon activation of a promoter and the target protein is produced in large quantity by the internal system of the host cell. It is now possible to choose between different cell systems depending on the expressed genes and upcoming applications. Bacterial systems, which are the most widespread hosts notably with *E. coli*, are suitable for production of a large quantity of proteins usually required for structural characterization thanks to the ease of genetic manipulation and their rapid growth. Yeast systems are preferred when posttranslational modifications are prescribed and insect or mammal cells when splicing of mRNA and correct folding with the help of chaperones are necessary.<sup>221,222</sup>

Production of proteins by recombinant approaches is a powerful tool in particular for the synthesis of large libraries of mutants through mutagenesis techniques. Alanine scanning, in which key positions are mutated one by one to alanine residue, allows for instance for a better understanding of the contribution of mutated amino acids in folding, inter-molecular interactions or function.<sup>223</sup> Directed evolution also requires large libraries of recombinant proteins via various techniques of mutagenesis (e.g. random mutagenesis, insertion and deletion) in order to engineer improved protein stability, enhanced catalytic

activity (for example, for computationally *de novo* designed enzyme, see section 1.1.2.) or to engineer altered substrate affinity.<sup>53,79</sup> Furthermore, recombinant DNA technologies enable the extensive production of protein-based drugs such as the widely used insulin.

However, recombinant techniques are extremely dependent of internal host system machineries and are limited in respect to flexibility, diversity and fine-tuning of modifications. Development of standard and routine procedures for the incorporation of non-canonical amino acids still demands efforts for improvement. Although engineered techniques for aminoacyl-tRNA synthetases modification allowing recognition and incorporation of unnatural amino acids and the development of cell free expression have been successful to extend the scope of synthesis,<sup>55,224</sup> they remain laborious and not commonly and largely used.

For these reasons, total chemical synthesis of proteins complements well biotechnological tools by providing access to native peptide sequences as well as modified analogues containing non-canonical amino acids. Chemical synthesis or semi-synthesis of protein holds several advantages in particular for the incorporation of non-natural and unusual building blocks,<sup>225,226</sup> the precise introduction of post-translational modifications and cross-linkers,<sup>227–229</sup> the production of mirror-image protein analogue<sup>230,231</sup> and proteins with non-linear backbone topology.<sup>232,233</sup>

**Solution-phase peptide synthesis.** Historically, the first peptides were assembled in solution through diverse methods that had been improved over the time following the development of new protecting groups and activation strategies. With only a little understanding about composition and structure of proteins, Theodor Curtius and Emil Fischer successfully pioneered the synthesis of the first peptides.<sup>234</sup> First, Theodor Curtius reported the synthesis in 1882 of a *N*<sup>ε</sup>-protected glycylglycine dipeptide by treatment of a salt of glycine with benzoylchloride.<sup>235</sup> He then developed the azide coupling method and managed to assemble polyglycine peptides of defined lengths.<sup>236</sup> In parallel, Emil Fischer synthesized the first completely unprotected dipeptide (glycylglycine) from diketopiperazine opening in 1901.<sup>237</sup> Moreover, he developed another method for peptide coupling with acyl chloride derivatives and handled the synthesis of an octadecapeptide composed of leucine and glycine residues.<sup>238,239</sup> During his course to the remarkable synthesis of polypeptides, Emil Fischer postulated while writing to his former supervisor, Adolf Baeyer, in 1905: “My entire yearning is directed toward the first synthetic enzyme. If its preparation falls into my lap with the synthesis of a natural protein material, I will consider my mission fulfilled”. Emil Fischer was a visionary, though he did not achieve his admirable goal at that time, mostly due to a lack of appropriate synthetic methods such as reversible protecting groups. It

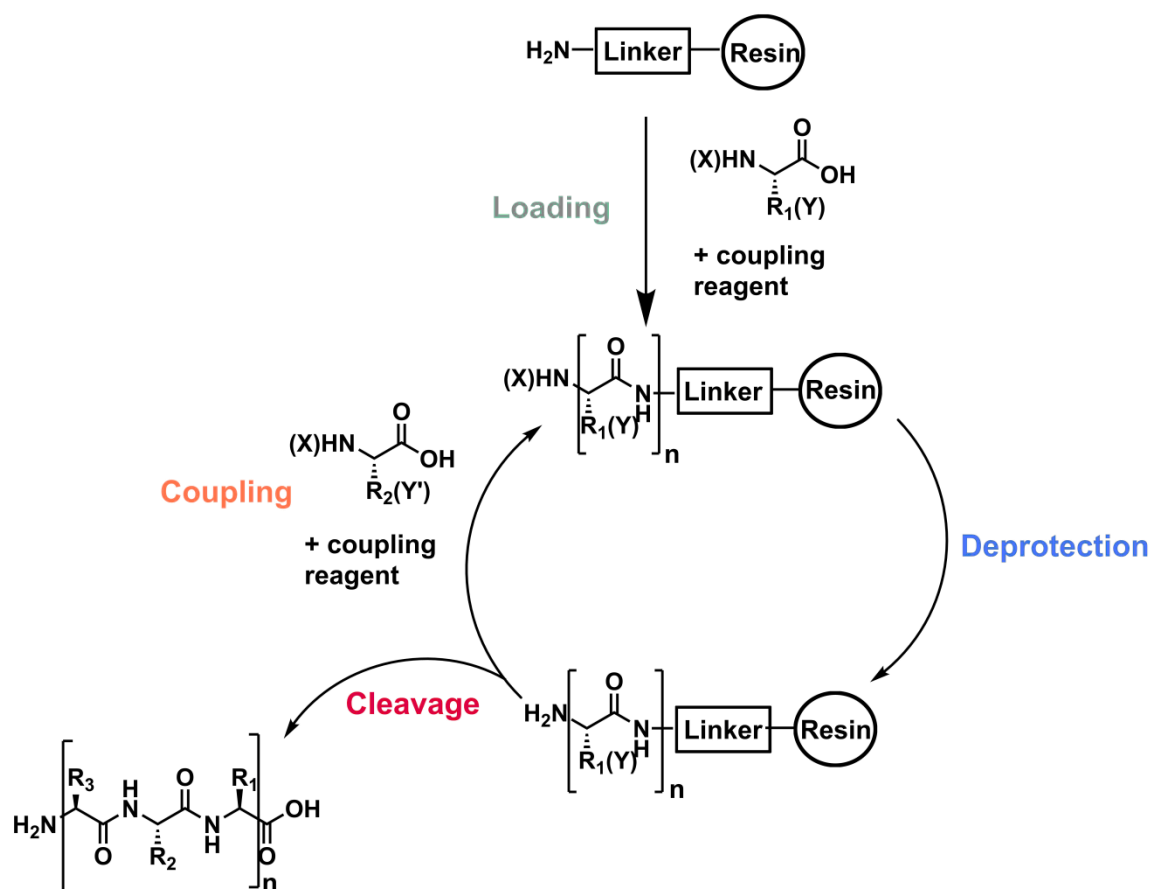
actually took several decades to develop robust synthetic methodologies in order to produce biologically relevant proteins.

Several advancements like the introduction of the benzyloxycarbonyl-protecting group (Cbz) by Bergmann and Zervas<sup>240</sup> or of the activating agent tetraethyl pyrophosphite by Anderson<sup>241</sup> led to the synthesis of oxytocin, a nine-residue long peptide hormone with biological activity. For this work, Vincent du Vigneaud was awarded the Nobel Prize in 1955.<sup>242</sup> Finally, in the late sixties, Hirschmann and coworkers accomplished the first complete synthesis by segment condensation in solution of an enzyme, the ribonuclease S, with catalytic activity recovery.<sup>243,244</sup> This demonstrated the chemical nature of proteins and enzymes as Emil Fischer had postulated more than 50 years before.

Despite the rapid development of practical tools such as robust and orthogonal protecting groups or activator reagents and the number of successful syntheses in solution phase, this approach suffers from solubility issues of fully protected peptide fragments, too low concentration in solvent slowing down the rate of reaction, lack of chiral integrity and difficult purification. All the steps necessary to assemble the amino acids and the segments with intermediate purifications lead to laborious syntheses with typically low yields. Nowadays, solution-phase peptide synthesis is usually employed to synthesize very short peptide fragments, but is not appropriate anymore for the total synthesis of larger proteins.

**Solid-phase peptide synthesis (SPPS).** While benefiting from progress in terms of orthogonal protecting groups and activation agents, the next major breakthrough in protein synthesis was introduced by Robert Bruce Merrifield in 1963 with the development of a method simplifying the sequential steps of chemical protein synthesis, and which is called the solid-phase peptide synthesis (SPPS).<sup>245</sup> Basically, the growing polypeptide chain is attached to an insoluble polymer at the C-terminus, thus allowing for a large excess of reagents that are simply removed by filtration and washing, while the desired peptide stays linked to the polymer support. The use of excess of reagents leads to faster and more efficient reaction yielding to peptide with higher purity compared to the solution-phase approach. Robert Bruce Merrifield demonstrated the potential and the robustness of his technique with the synthesis of the fully assembled ribonuclease A (124-residue long) by SPPS bearing native catalytic activity, in 1969 exactly at the same moment as Hirschmann.<sup>246,247</sup> Robert Bruce Merrifield finally won the Nobel Prize in 1984 for the development of chemical synthesis on a solid matrix and the large impact of the SPPS has been established on multiple fields such as biochemistry and medicine.

Typically, an insoluble cross-linked polymeric support (e.g. cross-linked polystyrene) is used in which a linker is grafted for peptide attachment and peptide cleavage in specific conditions during the final step. This technique is based on the stepwise addition of protected amino acids in a C-to-N terminal direction (see **Figure 1.17**). A first  $N^t$ -protected amino acid is covalently loaded to the polymer via activation of the carboxylic acid. Reactive groups on side chain of amino acids are also protected to avoid side reactions and would be generally deprotected only during the cleavage step. To proceed with high yields of amino acid coupling, excess of activated building blocks is used in comparison with the reactive group on the polymer. The polymer with the attached growing peptide is easily isolated from by-products by several rounds of washing and filtration. Orthogonal deprotection of  $N^t$ -amino group is subsequently performed with excess of reagent that is then removed by filtration and washing. The resulting free amine reacts with the next activated amino acid added in excess and repeated cycles of N-terminus deprotection and coupling are performed until polypeptide chain assembly is completed. In the final step named the cleavage step, the expected peptide is detached from the solid support and side chain protecting groups are removed to afford the peptide crude product. Thanks to the repetitive steps of this relatively simple approach for peptide synthesis and the use of a single reaction vessel containing the



**Figure 1.17. Principle of solid-phase peptide synthesis (SPPS).** The SPPS is based on the stepwise addition of protected amino acids in a C-to-N terminal direction. Repeated cycles of coupling and N-terminus deprotection are performed until polypeptide chain assembly is completed. The cleavage releases the expected unprotected peptide.

insoluble beads of resin, automatization of the procedure is possible and was actually directly proven by Robert Bruce Merrifield in 1966.<sup>248</sup>

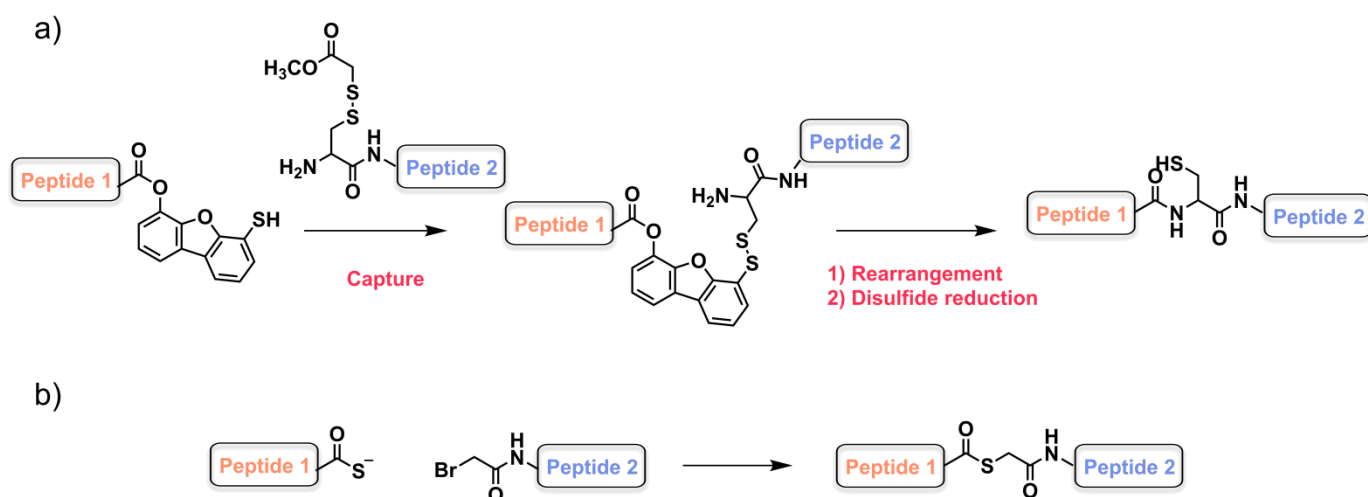
Two principal protecting group systems, the Boc/benzyl and the Fmoc/*t*Bu, are employed in SPPS. The historical approach has utilized the Boc/benzyl protecting group scheme and relies on a *N*<sup>ε</sup>-amino group protected with a Boc (*tert*-butyloxycarbonyl) moiety. This approach is not completely orthogonal as both *N*<sup>ε</sup>-protecting group and side chain protecting group are removed in acidic conditions. While the Boc group is removed with trifluoroacetic acid (TFA), cleavage and side chain deprotection need a stronger acid, which is liquid hydrogen fluoride (HF). Boc-SPPS can be advantageous for 'difficult' sequences to synthesize as TFA used for each *N*<sup>ε</sup> deprotection has the potency to temporarily destroy potential peptide-resin aggregates that are responsible for incomplete couplings and/or deprotection.<sup>249</sup> However, the handling of hazardous chemicals during the cleavage step (*i.e.* HF) has motivated researchers to find alternatives. Procedures with a less hazardous acid during cleavage have been elaborated, notably with trifluoromethanesulfonic acid (TFMSA).<sup>250</sup> Otherwise, another system using Fmoc/*t*Bu protecting groups for SPPS has been developed avoiding harsh and highly acidic conditions. This approach is completely orthogonal meaning that the *N*<sup>ε</sup>-Fmoc (9-Fluorenylmethoxycarbonyl) group is removed in basic conditions (commonly 20 % of piperidine in DMF) and side chains are deprotected during cleavage in acidic mixture (usually 95 % of TFA). These conditions are easier to set up, more prone to automatization thanks to peptide synthesizer and allow for the incorporation of building blocks sensitive to highly acidic medium such as phospho-residues.

Peptides of approximately 50-residues can be readily prepared by SPPS (difficulties of synthesis are highly dependent on the peptide sequence). However, sequences beyond this length are very hard to synthesize and purify to homogeneity due to the accumulation of by-products. The solution to this problem is a convergent approach for chemical synthesis, where several shorter peptides are efficiently synthesized by SPPS and condensed after purification to form a larger product.

**Chemical ligation of peptides.** Techniques of peptide ligation can allow synthesis of full-length protein up to ~ 300 residues by assembling small peptide fragments efficiently synthesized by SPPS and purified by preparative reverse phase high performance liquid chromatography (HPLC). Thus, ligation methods are based on the use of unprotected peptides that have the advantage to be highly soluble notably in buffers containing chaotropic agent (*i.e.* 6 M guanidinium hydrochloride or 8 M urea) in contrary to segment condensation methods working with protected peptides hardly soluble even in organic solvents. In this way, a ligation reaction has to be highly chemoselective to allow only reaction at the junction site without generating side reactions that may involve side chain

functionalities that bear distinct reactivities (*i.e.* phenol,  $\epsilon$ -amino group, alcohol, carboxylic acid, thiol, thioethers or heteroaromatic ring).

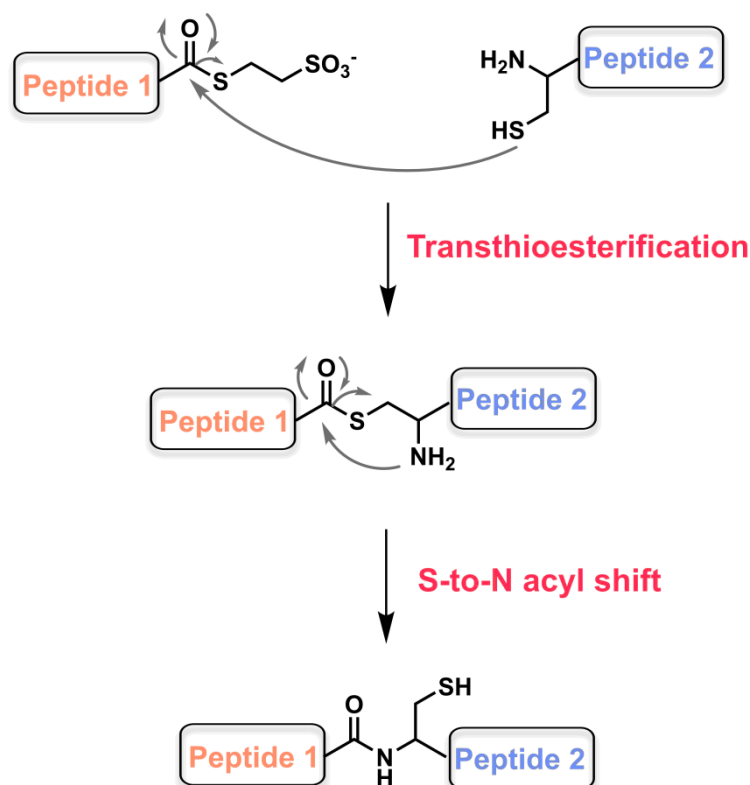
Early attempts of chemoselective ligation with unprotected synthetic peptide involved the prior thiol capture strategy developed by Daniel Kemp<sup>251</sup> and the original chemical ligation producing a non-native bond developed by Stephen Kent.<sup>252</sup> Although the first technique still needed protection of cysteine residues, rapid and chemoselective conversion to the ligated product was observed. Prior thiol capture consists in the synthesis of the N-terminal fragment with the  $C^{\alpha}$  carboxy group mildly activated by an auxiliary template, the 6-hydroxy-4-mercaptodibenzofuran (**Scheme 1.5a**). The capture of the C-terminal peptide fragment is driven by a disulfide exchange between the cysteine residue at the N-terminus of the latter peptide and the free thiol of the template. Therefore, the template places the two peptide fragments in proximity and an O-to-N acyl shift leads to the native peptide bond at the junction site. Finally, the reductive cleavage of the cysteine-template disulfide releases the ligated peptide product. Although, the prior thiol capture strategy allowed the synthesis of a 39-residue peptide, it did not turn into a routine synthetic method. On the other hand, the chemical ligation developed by S. Kent in 1992 employed for the first time completely unprotected peptides. This ligation strategy described the formation of a non-native bond, a thioester bond, at the junction site and involved the chemoselective and nucleophilic reaction of a C-terminal thioacid on an N-terminal bromoacetylated peptide (**Scheme 1.5b**). The resulting thioester linkage at a glycine-glycine junction possesses structural similarities with native amide bond. This technique contributed to the synthesis of a monomer of the human immunodeficiency virus-1 (HIV-1) protease composed of 99 amino acids.<sup>252</sup>



**Scheme 1.5. Early ligation of peptides.** a) Prior thiol capture strategy involves peptide 1 with the  $C^{\alpha}$  carboxy group mildly activated by an auxiliary template and a Cys-containing peptide 2. b) Chemical ligation with non-native thioester bond formation that is nearly isosteric to amide bond.

**Native chemical ligation of peptides.** One of the major milestones in total chemical synthesis of proteins and in peptide ligation came with the development of the native chemical ligation (NCL) technique published by S. Kent and coworkers in 1994.<sup>253</sup> This approach is now the most commonly used for convergent synthesis of peptides and yields to a native amide bond at the junction site through a chemoselective reaction between two unprotected peptides. The NCL relies on the distinct reactivity of thiols towards thioesters in aqueous buffer. The NCL starts with a reversible transthioesterification between the N-terminal peptide- $\alpha$ -thioester and the thiol moiety of a Cys-peptide fragment (**Scheme 1.6**). This step is followed by an irreversible S-to-N acyl shift that provides a native and stable amide bond at the ligation junction. Although the rate of the NCL is dependent of the nature of the amino acids at the junction site and the nature of the thioester, the reaction is often catalyzed by the addition of aryl-thiols. Generally, large concentration of aromatic thiols is necessary for thiol-thioester exchange to occur leading to highly reactive aryl-thioester.<sup>254</sup>

At the time when S. Kent developed the NCL, he demonstrated the potency and robustness of his ligation methodology with the synthesis of the 72-amino acid long human interleukin 8 (IL-8).<sup>253</sup> Nowadays, the NCL is widely used and synthesis of proteins of more than 200 amino acids can be attained.<sup>255</sup> A recent work of Ashraf Brik has even reported the impressive synthesis of a 53 kDa tetraubiquitinated  $\alpha$ -globin composed of 472 amino acids mediated through sequential NCL and isopeptide NCL on lysine.<sup>256</sup>

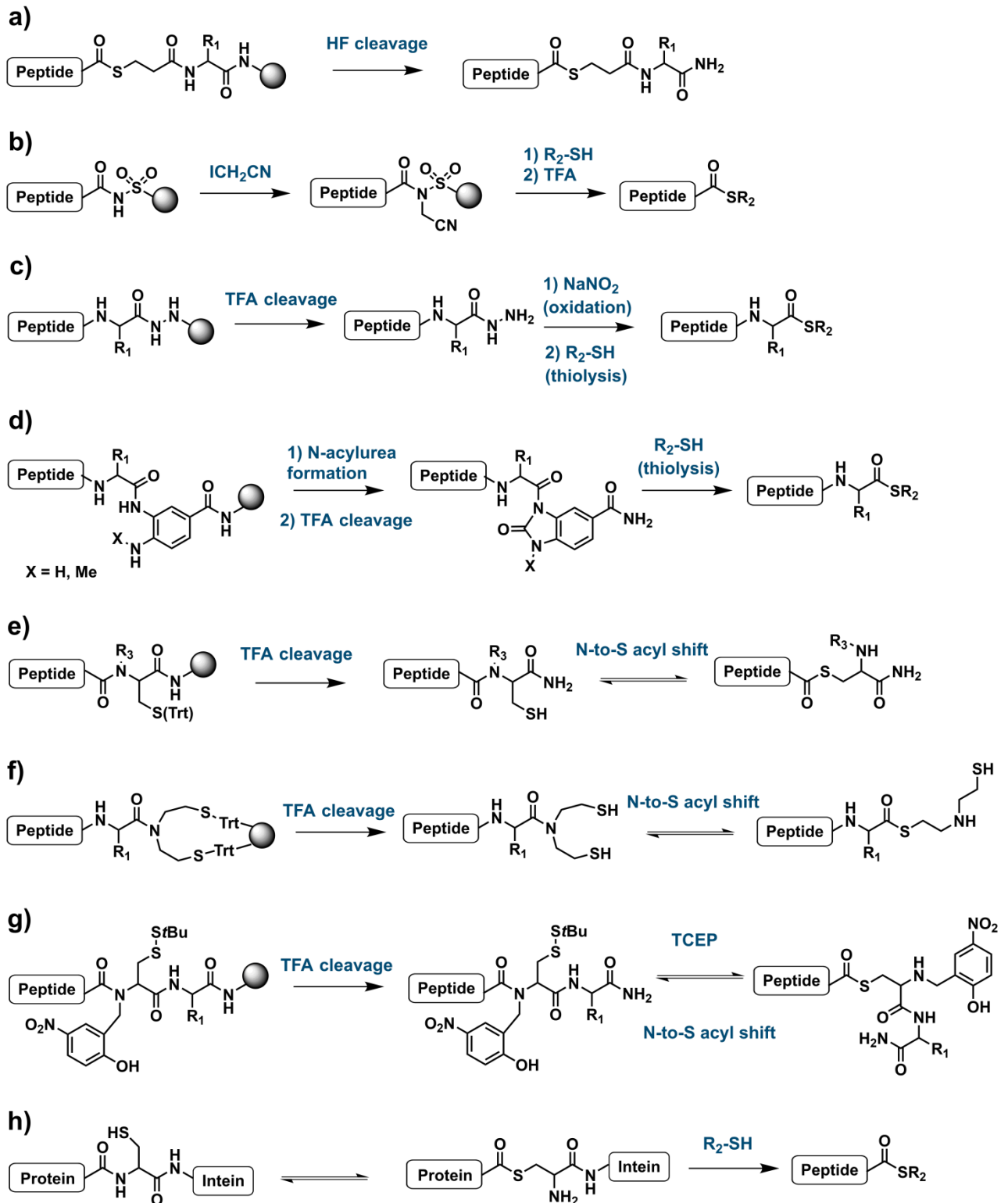


**Scheme 1.6. The Native chemical ligation (NCL).** The NCL involves a transthioesterification followed by an S-to-N acyl shift that irreversibly gives a stable amide bond.

One of the preliminary and essential steps before performing native chemical ligation is the preparation of peptide- $\alpha$ thioester. As we discussed before, thioester are usually employed in living organism to activate carboxylic acid to mainly assist in the formation of amide bond, however the production of peptide- $\alpha$ thioester by chemical methods can be a critical step. The preparation of thioesters can readily be performed by Boc/benzyl-SPPS by coupling an amino acid on a thiol moiety such as 3-mercaptopropionic acid (MPA) thus forming a thioester linkage on solid phase (see **Scheme 1.7a**). The peptide- $\alpha$ thioester is directly obtained during the synthesis and recovered after cleavage from polymer support. However, the Boc-SPPS is not widely available in laboratories and pre-existing thioester moiety on resin is not compatible with Fmoc chemistry because of its lability towards secondary amine (*i.e.* piperidine) used during the different Fmoc-deprotection steps. Among the different strategies developed to circumvent this problem, methods to form the thioester during the final step of cleavage as well as methods to form crypto-thioesters that would be converted afterwards into thioesters were established. To prepare peptide- $\alpha$ thioesters after completion of the Fmoc-SPPS and during the cleavage step, the Kenner's acylsulfonamide safety-catch linker was adapted (**Scheme 1.7b**).<sup>257,258</sup> Basically, this specific linker can be activated by alkylation to provide an *N*-alkyl acylsulfonamide that is susceptible to thiol attack yielding to the cleavage of the peptide with formation of a C-terminal peptide thioester. However, this safety-catch method is not so straightforward and alkylation of the sulfonamide prior to nucleophilic cleavage can also lead to reactions with other residues of the peptide. For this reason, efforts were dedicated towards the development of crypto-thioesters in order to recuperate the desired thioester by using specific and controlled conditions. Few key examples are presented hereinafter.

One widely used method to form thioester was developed by L. Liu.<sup>259,260</sup> Here, peptides- $\alpha$ hydrazide were used as thioester surrogates and can be first easily obtained by Fmoc-SPPS using a resin containing a hydrazide linker (**Scheme 1.7c**). Peptides- $\alpha$ hydrazide are subsequently oxidized by treatment with sodium nitrite to form an acyl azide derivative, which directly undergoes thiolysis upon the addition of an external thiol (for example, 4-mercaptophenylacetic acid (MPAA) or sodium 2-mercaptoethanesulfonate (MesNa)). Another frequent method utilized in total chemical synthesis of protein to form thioester was established by P. Dawson via peptide synthesis on a diaminobenzoyl linker (Dbz) (as shown in **Scheme 1.7d**).<sup>261</sup> After completion of peptide assembly on the linker, treatment with *p*-nitrophenylchloroformate and followed by TFA leads to the *N*-acyl urea (Nbz) peptide. Thiolytic cleavage of the Nbz peptide can easily be performed. Improvement of the linker such as monomethylation has avoided formation of by-products coming from Dbz diacylation.<sup>262</sup> Multiple strategies to form thioester peptides have also involved post-SPPS N-to-S acyl shift systems with, for instance, the introduction of an *N*-alkylcysteine developed by H. Hojo





**Scheme 1.7. Chemical methods to form peptide- $\alpha$ thioester.** **a)** Boc-SPPS. **b)** Safety-catch linker. **c)** Hydrazide method. **d)** Formation of *N*-acyl urea (Nbz) peptide from diaminobenzoyl linker. **e)** *N*-alkyl cysteine for *N*-to-*S* acyl shift. **f)** Bis(2-sulfanylethyl)amido (SEA) system. **g)** *N*-(2-hydroxybenzyl)cysteine building block allowing *N*-to-*S* acyl shift for thioester formation. **h)** Thioester formation via expressed protein ligation, where a thioester is formed through intein rearrangement and transthioesterification.

(**Scheme 1.7e**),<sup>263</sup> the bis(2-sulfanylethyl)amido (SEA) system by O. Melnyk (**Scheme 1.7f**)<sup>264,265</sup> or the introduction of *N*-(2-hydroxybenzyl)cysteine building block by V. Aucagne (**Scheme 1.7g**).<sup>266</sup>

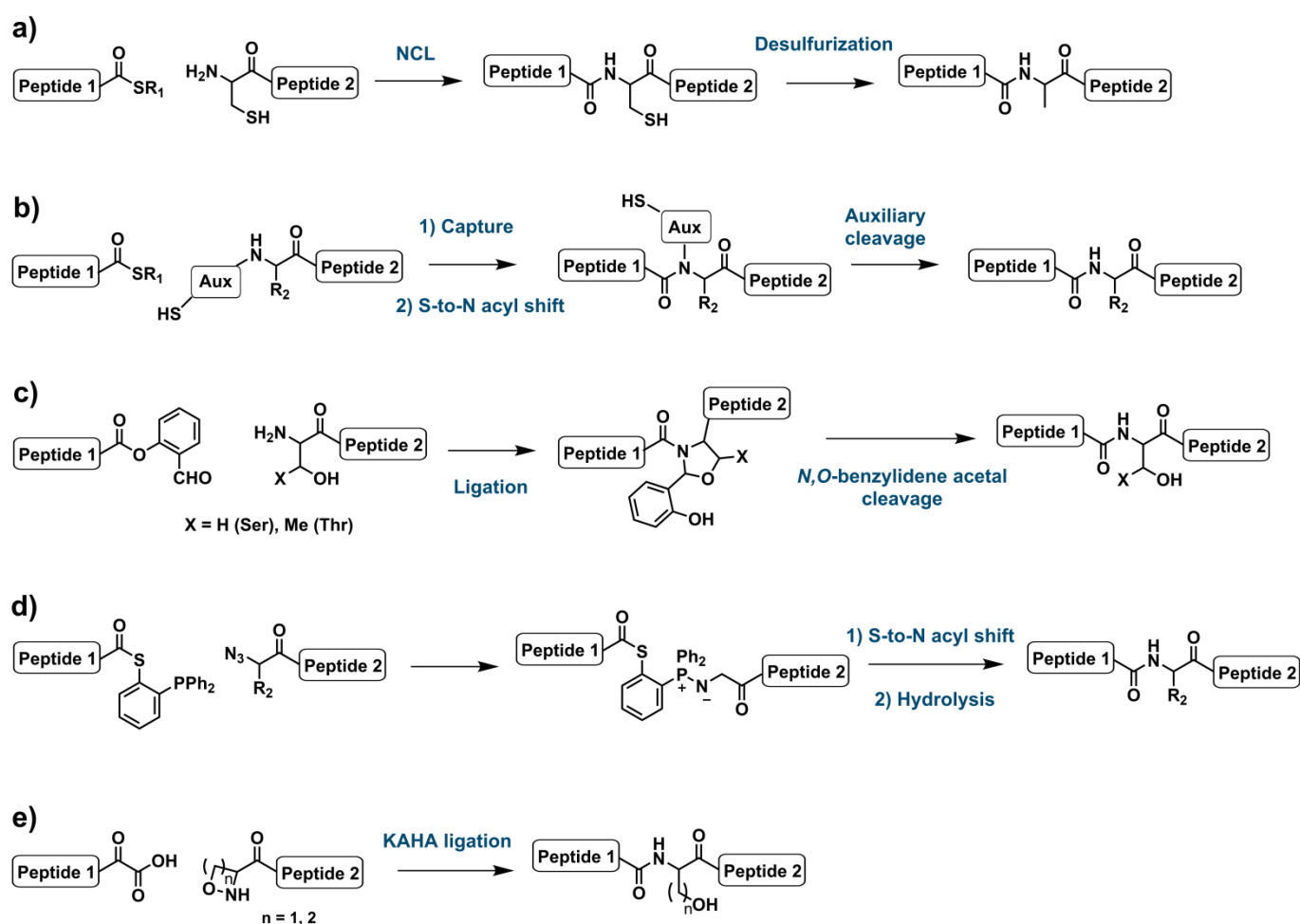
Besides chemical methods to produce thioester, a more biological approach takes inspiration from intein system to generate an intrinsic thioester on recombinant protein, which is then transthioesterified with an alkyl- or aryl-thiol (**Scheme 1.7h**).<sup>267</sup> This method, called expressed protein ligation (EPL), enables semisynthesis of large protein by assembling together a synthetic peptide and a recombinant protein. Peptide chemists can now play with a large toolkit for thioester formation allowing for the access to more and more synthetically challenging proteins.

However, the major and general drawback of the native chemical ligation approach is the requirement for a cysteine residue in the sequence, which is one of the rarely occurring amino acids in native proteins. That is why several complementary NCL-techniques have been conceived to expand the scope of chemical protein synthesis.

**Complementary and alternative methods to native chemical ligation.** To enlarge the scope of possible synthesized proteins, complementary methods to NCL were developed. Probably, the most useful and widespread technique was introduced by Dawson and Yan and consists in a post-ligation desulfurization of the thiol of the involved cysteine leading to a more abundant alanine residue (**Scheme 1.8a**).<sup>268</sup> Later, a milder and non-metal based reduction of the cysteine was elaborated to allow broader range of applications.<sup>269</sup> This concept was extended to many other residues with the elaboration of their  $\beta$ -,  $\gamma$ - and  $\delta$ -thiol derivatives enabling NCL followed by desulfurization to recover the native amino acid.<sup>270</sup> Another approach for ligation site containing no Cys amino acid resides in the use of thiol-containing auxiliary that is branched on the N-terminal amino group and mimics the presence of a cysteine (see **Scheme 1.8b**).<sup>271,272</sup> After ligation reaction, this auxiliary is selectively removed from the resulting amide bond. However, auxiliaries are very sensitive to the amino acids found at the junction site, thus this method is less used for practical reasons.

Other ligation methodologies have involved thiol-independent reactivities to form native amide bond in contrary to the native chemical ligation. The Ser/Thr ligation employs, for instance, a salicylaldehyde (SAL) ester on a peptide C-terminus and a serine or threonine residue at the N-terminus of another unprotected peptide (**Scheme 1.8c**).<sup>273</sup> This ligation is actually based on the chemoselective reaction between the SAL ester and the 1,2-hydroxylamine moiety of Ser or Thr forming an *N,O*-benzylidene acetal intermediate, which is subsequently subjected to acidolysis affording a native peptide bond at a Xaa-Ser/Thr junction. Furthermore, the Staudinger reaction was adapted for ligation of peptides in order to exploit the chemoselectivity of phosphine towards azide.<sup>274,275</sup> The resulting traceless

Staudinger ligation is based on the reaction between a phosphinothioester peptide and an N-terminal azidopeptide to form an amide bond free from any residual components (**Scheme 1.8d**). A peptide- $\alpha$ thioester is first transthioesterified with a phosphinothiol, which then reacts with the azido group of a second peptide generating an iminophosphorane intermediate. An S-to-N acyl transfer yields an amidophosphonium salt, which upon hydrolysis gives the native amide bond with release of phosphine oxide. Another alternative to the NCL providing native amide bond from two unprotected peptides is the  $\alpha$ -ketoacid-hydroxylamine (KAHA) decarboxylative ligation reported by J. Bode (**Scheme 1.8e**).<sup>276</sup> Through a rather complex mechanism, the C-terminal ketoacid and an N-terminal hydroxylamine derivative rearrange spontaneously in acidic conditions to form an amide bond. Limitations in terms of solubility in aqueous media necessitated improvements of the technique. The 5-oxaproline building block was found to be a stable and practical substitute to hydroxylamines for robust and easy way to execute ligation, but provided a homoserine, a non-canonical residue, at the ligation site.<sup>277</sup> Later, the synthesis of a four-membered ring alkoxyamine derivative, the oxazetidine,



**Scheme 1.8. Complementary and alternative methods to native chemical ligation.** **a)** Post-NCL desulfurization affording an alanine residue. **b)** Auxiliary-assisted ligation. In the Scheme, Aux stands for auxiliary. **c)** Ser/Thr ligation. **d)** Traceless Staudinger ligation between a phosphinothioester peptide and an azidopeptide. **e)** KAHA ligation between a C-terminal ketoacid and an N-terminal hydroxylamine derivative.

induced rapid serine-forming ligations for synthesis of proteins with fully native sequence.<sup>278</sup>

**Enzyme-catalyzed ligation.** Besides chemical methods, enzyme-mediated ligation of peptide fragments is highly promising thanks to their inherent chemoselectivity, their high chiral integrity and the potential use at low catalytic amount. These enzymes are also potentially powerful tools for peptide cyclization.<sup>279</sup> Enzymes that entail peptide ligation bear similar catalytic mechanisms. Basically, an acyl donor substrate (e.g. peptide- $\alpha$ -thioester, activated ester or a recognized sequence by the enzyme) is attacked by the catalytic and nucleophile residue to constitute a branched adduct in the form of a covalently acyl-enzyme intermediate. This branched adduct undergoes then aminolysis with an  $\alpha$ -amino group (coming either from the same peptide substrate for cyclization or from a second peptide substrate for peptide ligation) to form the expected peptide bond.

Historically, because of the rare existence of ligases in nature, the first ligase was engineered from a protease (the most outstanding example is the subtilisin, see subsection 1.1.2.) that catalyzed the opposite reaction (*i.e.* cleavage of peptides). Like all biocatalysts, proteases are capable to catalyze the reverse reaction that is to say the peptide ligation under specific conditions. A double mutation in the catalytic site of the subtilisin provided a great enhancement of the ligation reaction in comparison to hydrolysis.<sup>280</sup> The resulting subtiligase carried a mutation of the catalytic serine to a cysteine improving the aminolysis over hydrolysis ratio and a second mutation of a proline to an alanine reducing the steric crowding of the active site and so, facilitating the entry of the second substrate. Among the wide range of applications of the engineered subtiligase, this work has resulted in the impressive synthesis of a 124-residue ribonuclease A by assembling six peptide fragments.<sup>281,282</sup> Furthermore, newer and more efficient variants of proteolytic proteins with ligation catalytic activity allowing for larger scope of substrates were recently developed, proving the robustness and the promising possibilities of this method.<sup>283,284</sup>

Moreover, sortase A and butelase 1, two naturally occurring ligases that we have previously presented in subsection 1.2.1., have also been utilized as a synthetic tool to catalyze peptide ligation. On one hand, the sortase A catalyzes the transpeptidation of an LPXTG peptide acceptor and the pentaglycine N-terminal extension of lipid II in bacterial cell wall. Consequently, by using a substrate including the LPXTG sorting motif at the C-terminus and a second substrate with at least one Gly residue at the N-terminus, the sortase A enzyme has been extensively applied for site-specific protein modification, also referred to as “sortagging”, as well as for peptide and protein ligation and cyclization.<sup>285</sup> However, the absolute need of the sorting motif and the relatively poor catalytic efficiency (between 0.1 and 1.0 molar ratio of required enzyme) have limited the application of this enzyme. On the other hand, butelase 1, an Asp/Asn-specific peptide ligase, was found to be the fastest

peptide ligase known with the catalytic efficiency as high as  $542\ 000\ \text{M}^{-1}\cdot\text{s}^{-1}$ .<sup>194</sup> Indeed, butelase 1 recognizes the NHV motif and cleaves after asparagine residue forming a transient acyl-Cys-enzyme intermediate, which is intercepted by N-terminal  $\alpha$ -amino group of a second peptide substrate with creation of an amide bond. Applications for peptide ligation have been demonstrated as well as bioconjugation of protein and peptide macrocyclization.<sup>286,287</sup>

However, all these enzymatic approaches share a common disadvantage. Engineered and natural ligases possess restrictions to amino acids sequences that can bind to their substrate pockets. Actually, the catalytic efficiency is coupled to substrate specificity in these enzymes meaning that substrates with low affinity would not result in efficient catalysis. Moreover, the overall complexity of these large and evolved structures complicates the improvement and further tuning of their enzymatic properties.

Finally, available tools for total synthesis of proteins are varied to expand the range of possible synthesized proteins. Biotechnological and chemical methods allow reaching various proteins for diverse applications. Larger-sized proteins are more easily produced via recombinant techniques, whereas peptides and proteins with controlled and precise insertion of labels, post-translational modifications and of unnatural amino acids are more conveniently attained through chemical synthesis.

#### **1.2.4. Conclusions**

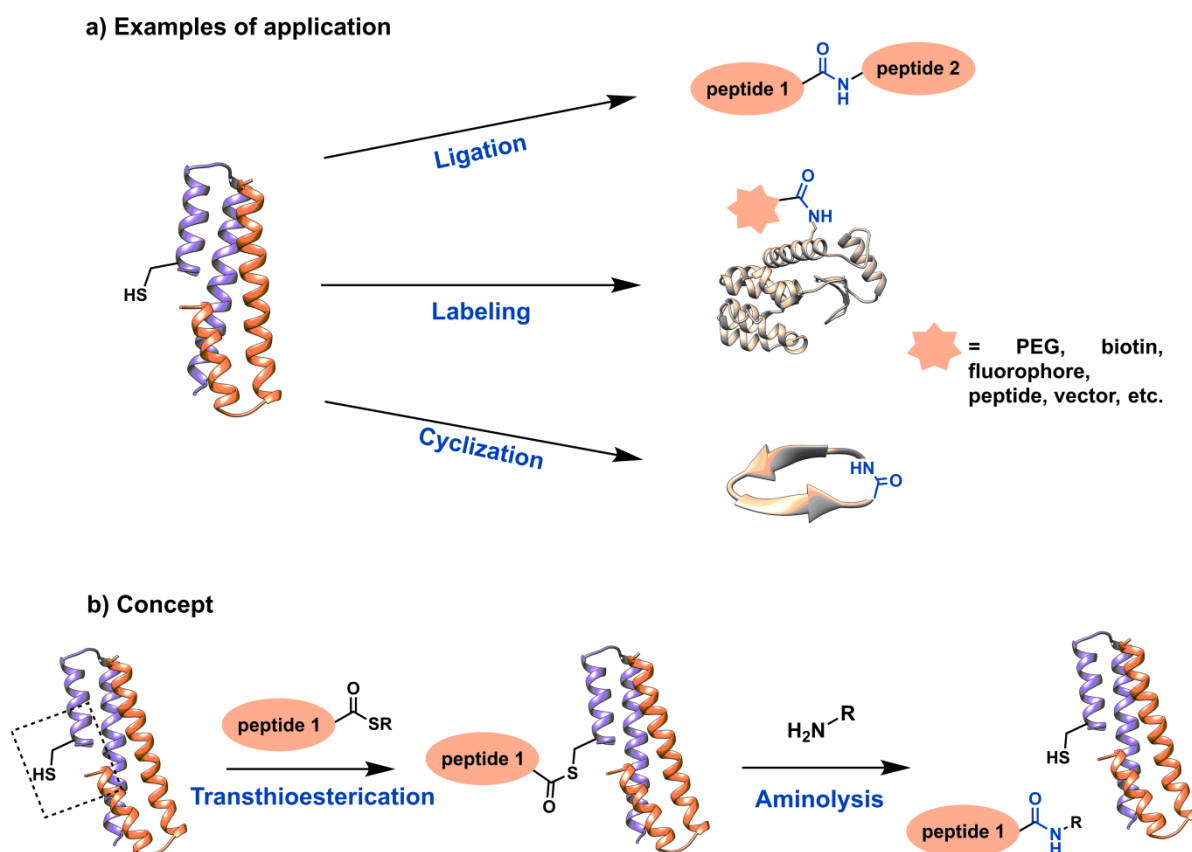
Understanding amide bond formation is of great interest to develop new synthetic tools as well as to produce valuable molecules like synthetic drugs, polymers and proteins. In the last century, a lot of effort has been directed to the development of novel methodologies for amide bond formation, which has been beneficial in the field of peptide and protein synthesis for therapeutic and fundamental needs.

Biomolecular and chemical tools are complementary for the production of proteins. On one hand, biotechnologies enable the production of large proteins and protein complexes but standard and routine procedures for incorporation of non-canonical amino acids and precise placement of post-translational modifications are currently missing. On the other hand, total chemical synthesis of proteins overcomes these limitations as each step of the synthesis is well controlled and enables tailor-made synthesis of complex peptides. Expansion of the different chemical methodologies such as the SPPS or the NCL has enabled the synthesis of larger and larger peptides and proteins. Possibilities for synthesis of

novel topologies or for incorporation of unusual modifications are rather unlimited thanks to chemistry. However, current synthetic approaches are highly demanding in terms of atom economy and global efficiencies are generally sequence dependent. For instance, activation of carboxylic acids for amide bond formation usually needs equimolar quantity of the coupling agent and coupling and deprotection steps in SPPS utilize excess of reagents and environmentally unfriendly solvents. Although NCL is performed on unprotected peptide fragments in aqueous buffer, high concentration of aryl thiols is generally used to efficiently catalyze the reaction. Development and improvement of chemical methodologies have to be continued in order to meet these challenges in terms of chemical efficiency in a more environmentally friendly manner.

### 1.3. Objectives and thesis outline

Designing *de novo* proteins with a desired and innovative catalytic function finds practical applications by addressing contemporary challenges and provides with the possibility to better understand fundamental principles of protein function.<sup>30,97</sup> Developing such protein catalysts would constitute new tools for industrial processes, medicine and chemical biology studies. In this thesis, I applied rational minimalist design to install a novel function into an already existing *de novo* scaffold. In the laboratory, we are involved in the development of new approaches to facilitate peptide bond formation in a high-throughput manner for application in total chemical synthesis of proteins. On my side, I have been particularly interested in the catalysis of peptide ligation and more broadly of amide bond formation through addition/elimination (A/E) reactions on thioesters. In the previous section, we have seen that progresses in relation to catalysis, waste reduction and chemoselectivity can still be ameliorated for amide bond formation. From this perspective, the design of a *de novo* catalytic protein that is rather small and modular compared to a natural enzyme would gather all the above-mentioned prerequisites. In this way, the final aim of this project would be to develop a novel general tool that can be used as a peptide ligase and/or for peptide/protein labeling (**Figure 1.18a**). By installing a catalytic residue in a specific environment on a protein scaffold, the main goal of my thesis was to demonstrate the feasibility of this concept to accelerate the reaction between a peptide- $\alpha$ thioester and an acyl acceptor, principally an amine (**Figure 1.18b**).



**Figure 1.18. Possible applications and schematic principle of the designed catalytic protein.** a) Catalyzing amide bond formation finds multiple applications such as peptide ligation, peptide and protein labeling and peptide cyclization. b) Our designed *de novo* protein is expected to accelerate the reaction between an acyl donor (a peptide-<sup>α</sup>thioester) and an acyl acceptor (an amine) through a first step of transthioesterification to form a branched adduct and a second step of aminolysis. The square dashed box represents the position of the active site.

The second chapter of my thesis explores the feasibility of the concept: is it possible to install a challenging function such as the catalysis of amide bond formation in a *de novo* protein by simple rational design based on few assumptions and chemical intuition? In this chapter, the concept and the protein design principles will be largely explained. Optimization of the chemical synthesis of the catalytic proteins and of the catalytic assays will be presented as an inherent part of the development of the *de novo* catalytic protein. Biophysical and structural characterization as well as general catalytic description of the first generation of our *de novo* proteins have enabled us to better understand our new system. The third chapter focuses on the synthesis of a library of new analogues. Based on intuition and simple modeling, a library of twenty analogues has been designed and a chemical and convergent approach has been applied for its synthesis. I demonstrate that the environment in proximity of the catalytic residue can moderately modulate kinetic parameters. In parallel, the synthesis of libraries of peptide substrates for acyl donors and acyl acceptors is reported for future catalytic characterizations. In the fourth chapter, I will describe the computational design of an elongated analogue that should possess better stability. Finally, the fifth chapter

is dedicated to the concomitant hydrolytic and potential proteolytic activity of one specific analogue that was discovered by serendipity.

I report in this thesis the successful completion of the first step in this long-term project by building a completely *de novo* protein with predefined acyl transfer catalytic activity. Different paths have been explored to better understand our system and some has not reached our expectations. Nevertheless, this thesis represents a promising starting point for development of efficient catalysts and several more iterative cycles will be necessary to reach high levels of catalytic activity.



## 1.4. References

- (1) Wolfenden, R.; Snider, M. J. The Depth of Chemical Time and the Power of Enzymes as Catalysts. *Acc. Chem. Res.* **2001**, *34* (12), 938–945.
- (2) Caspi, R.; Foerster, H.; Fulcher, C. A.; Hopkinson, R.; Ingraham, J.; Kaipa, P.; Krummenacker, M.; Paley, S.; Pick, J.; Rhee, S. Y.; Tissier, C.; Zhang, P.; Karp, P. D. MetaCyc: A Multiorganism Database of Metabolic Pathways and Enzymes. *Nucleic Acids Res.* **2006**, *34* (suppl\_1), D511–D516.
- (3) Tipton, K.; Boyce, S. History of the Enzyme Nomenclature System. *Bioinformatics* **2000**, *16* (1), 34–40.
- (4) Mitić, N.; Smith, S. J.; Neves, A.; Guddat, L. W.; Gahan, L. R.; Schenk, G. The Catalytic Mechanisms of Binuclear Metallohydrolases. *Chem. Rev.* **2006**, *106* (8), 3338–3363.
- (5) Klinman, J. P. The Mechanism of Enzyme-Catalyzed Reduced Nicotinamide Adenine Dinucleotide-Dependent Reductions. Substituent and Isotope Effects in the Yeast Alcohol Dehydrogenase Reaction. *J. Biol. Chem.* **1972**, *247* (24), 7977–7987.
- (6) Dourado, D. F. A. R.; Swart, M.; Carvalho, A. T. P. Why the Flavin Adenine Dinucleotide (FAD) Cofactor Needs To Be Covalently Linked to Complex II of the Electron-Transport Chain for the Conversion of FADH<sub>2</sub> into FAD. *Chem. Eur. J.* **2018**, *24* (20), 5246–5252.
- (7) Yan, Y.; Harper, S.; Speicher, D. W.; Marmorstein, R. The Catalytic Mechanism of the ESA1 Histone Acetyltransferase Involves a Self-Acetylated Intermediate. *Nat. Struct. Biol.* **2002**.
- (8) Fischer, E. Einfluss Der Configuration Auf Die Wirkung Der Enzyme. *Ber. Dtsch. Chem. Ges.* **1894**, *27* (3), 2985–2993.
- (9) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44* (2), 98–104.
- (10) Koshland, D. E. The Key–Lock Theory and the Induced Fit Theory. *Angew. Chem., Int. Ed. Engl.* **1995**, *33* (23–24), 2375–2378.
- (11) Gutteridge, A.; Thornton, J. Conformational Change in Substrate Binding, Catalysis and Product Release: An Open and Shut Case? *FEBS Lett.* **2004**, *567* (1), 67–73.
- (12) Robinson, P. K. Enzymes: Principles and Biotechnological Applications. *Essays Biochem.* **2015**, *59*, 1–41.
- (13) Pauling, L. Molecular Architecture and Biological Reactions. *Chem. Eng. News* **1946**, *24* (10), 1375–1377.
- (14) Kraut, J. How Do Enzymes Work? *Science* **1988**, *242* (4878), 533–540.
- (15) Williams, D. H. Enzyme Catalysis from Improved Packing in Their Transition-State Structures. *Curr. Opin. Chem. Biol.* **2010**, *14* (5), 666–670.
- (16) Williams, D. H.; Stephens, E.; O'Brien, D. P.; Zhou, M. Understanding Noncovalent Interactions: Ligand Binding Energy and Catalytic Efficiency from Ligand-Induced Reductions in Motion within Receptors and Enzymes. *Angew. Chem. Int. Ed.* **2004**, *43* (48), 6596–6616.

- (17) Wang, F.; Shi, W.; Nieves, E.; Angeletti, R. H.; Schramm, V. L.; Grubmeyer, C. A Transition-State Analogue Reduces Protein Dynamics in Hypoxanthine-Guanine Phosphoribosyltransferase. *Biochemistry* **2001**, *40* (27), 8043–8054.
- (18) Schramm, V. L. Transition States, Analogues, and Drug Development. *ACS Chem. Biol.* **2013**, *8* (1), 71–81.
- (19) Vögeli, B.; Erb, T. J. ‘Negative’ and ‘Positive Catalysis’: Complementary Principles That Shape the Catalytic Landscape of Enzymes. *Curr. Opin. Chem. Biol.* **2018**, *47*, 94–100.
- (20) Bruice, T. C.; Lightstone, F. C. Ground State and Transition State Contributions to the Rates of Intramolecular and Enzymatic Reactions. *Acc. Chem. Res.* **1999**, *32* (2), 127–136.
- (21) Cannon, W. R.; Singleton, S. F.; Benkovic, S. J. A Perspective on Biological Catalysis. *Nat. Struct. Biol.* **1996**, *3* (10), 821–833.
- (22) Perutz, M. F. Concluding Remarks. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **1967**.
- (23) Hollfelder, F.; Kirby, A. J.; Tawfik, D. S. On the Magnitude and Specificity of Medium Effects in Enzyme-like Catalysts for Proton Transfer. *J. Org. Chem.* **2001**, *66* (17), 5866–5874.
- (24) Lienhard, G. E. Enzymatic Catalysis and Transition-State Theory. *Science* **1973**, *180* (4082), 149–154..
- (25) Bell, R. M.; Koshland, D. E. Covalent Enzyme-Substrate Intermediates. *Science* **1971**, *172* (3989), 1253–1256.
- (26) Knapp, M. J.; Klinman, J. P. Environmentally Coupled Hydrogen Tunneling. *Eur. J. Biochem.* **2002**, *269* (13), 3113–3121.
- (27) Nagel, Z. D.; Klinman, J. P. Tunneling and Dynamics in Enzymatic Hydride Transfer. *Chem. Rev.* **2006**, *106* (8), 3095–3118.
- (28) Schramm, V. L.; Schwartz, S. D. Promoting Vibrations and the Function of Enzymes. Emerging Theoretical and Experimental Convergence. *Biochemistry* **2018**, *57* (24), 3299–3308.
- (29) Benkovic, S. J.; Hammes-Schiffer, S. A Perspective on Enzyme Catalysis. *Science* **2003**, *301* (5637), 1196–1202.
- (30) Nanda, V.; Koder, R. L. Designing Artificial Enzymes by Intuition and Computation. *Nat. Chem.* **2010**, *2* (1), 15–24.
- (31) Khoury, G. A.; Smadbeck, J.; Kieslich, C. A.; Floudas, C. A. Protein Folding and de Novo Protein Design for Biotechnological Applications. *Trends Biotechnol.* **2014**, *32* (2), 99–109.
- (32) Wentworth, P.; Janda, K. D. Catalytic Antibodies. *Curr. Op. Chem. Biol.* **1998**, *2* (1), 138–144.
- (33) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational Enzyme Design. *Angew. Chem. Int. Ed.* **2013**, *52* (22), 5700–5725.
- (34) Schultz, P. G.; Lerner, R. A. From Molecular Diversity to Catalysis: Lessons from the Immune System. *Science* **1995**, *269* (5232), 1835–1842.
- (35) Xu, Y.; Yamamoto, N.; Janda, K. D. Catalytic Antibodies: Hapten Design Strategies and Screening Methods. *Bioorg. Med. Chem.* **2004**, *12* (20), 5247–5268.
- (36) Thorn, S. N.; Daniels, R. G.; Auditor, M.-T. M.; Hilvert, D. Large Rate Accelerations in Antibody Catalysis by Strategic Use of Haptenic Charge. *Nature* **1995**, *373* (6511), 228–230.

- (37) Gouverneur, V. E.; Houk, K. N.; Pascual-Teresa, B. de; Beno, B.; Janda, K. D.; Lerner, R. A. Control of the Exo and Endo Pathways of the Diels-Alder Reaction by Antibody Catalysis. *Science* **1993**, 262 (5131), 204–208.
- (38) Baldwin, J. E. Rules for Ring Closure. *J. Chem. Soc., Chem. Commun.* **1976**, No. 18, 734–736.
- (39) Janda, K. D.; Shevlin, C. G.; Lerner, R. A. Antibody Catalysis of a Disfavored Chemical Transformation. *Science* **1993**, 259 (5094), 490–493.
- (40) Sinha, S. C.; Keinan, E. Catalytic Antibodies in Organic Synthesis. Asymmetric Synthesis of (-)- $\alpha$ -Multistriatin. *J. Am. Chem. Soc.* **1995**, 117 (12), 3653–3654.
- (41) Mets, B.; Winger, G.; Cabrera, C.; Seo, S.; Jamdar, S.; Yang, G.; Zhao, K.; Briscoe, R. J.; Almonte, R.; Woods, J. H.; Landry, D. W. A Catalytic Antibody against Cocaine Prevents Cocaine's Reinforcing and Toxic Effects in Rats. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, 95 (17), 10176–10181.
- (42) Jones, L. H.; Wentworth, P. The Therapeutic Potential for Catalytic Antibodies From a Concept to a Promise. *Mini-Rev. Med. Chem.* **2001**, 1 (2), 125-132.
- (43) Hilvert, D. Critical Analysis of Antibody Catalysis. *Annu. Rev. Biochem* **2000**, 69 (1), 751-793.
- (44) Stewart, J. D.; Benkovic, S. J. Catalytic Antibodies: Mechanistic and Practical Considerations. *Chem. Soc. Rev.* **1993**, 22 (4), 213–219.
- (45) Toscano, M. D.; Woycechowsky, K. J.; Hilvert, D. Minimalist Active-Site Redesign: Teaching Old Enzymes New Tricks. *Angew. Chem. Int. Ed.* **2007**, 46 (18), 3212–3236.
- (46) Reetz, M. T.; Soni, P.; Acevedo, J. P.; Sanchis, J. Creation of an Amino Acid Network of Structurally Coupled Residues in the Directed Evolution of a Thermostable Enzyme. *Angew. Chem. Int. Ed.* **2009**, 121 (44), 8418–8422.
- (47) Zumárraga, M.; Bulter, T.; Shleev, S.; Polaina, J.; Martínez-Arias, A.; Plou, F. J.; Ballesteros, A.; Alcalde, M. In Vitro Evolution of a Fungal Laccase in High Concentrations of Organic Cosolvents. *Chem. Biol.* **2007**, 14 (9), 1052–1064.
- (48) Kaiser, E. T.; Lawrence, D. S. Chemical Mutation of Enzyme Active Sites. *Science* **1984**, 226 (4674), 505–511.
- (49) Polgar, L.; Bender, M. L. A New Enzyme Containing a Synthetically Formed Active Site. Thiol-Subtilisin1. *J. Am. Chem. Soc.* **1966**, 88 (13), 3153–3154..
- (50) Neet, K. E.; Koshland, D. E. The Conversion of Serine at the Active Site of Subtilisin to Cysteine: A "Chemical Mutation." *Proc. Natl. Acad. Sci. U.S.A.* **1966**, 56 (5), 1606–1611.
- (51) Nakatsuka, T.; Sasaki, T.; Kaiser, E. T. Peptide Segment Synthesis Catalyzed by the Semisynthetic Enzyme Thiolsubtilisin. *J. Am. Chem. Soc.* **1987**, 109 (12), 3808–3810.
- (52) Seebeck, F. P.; Hilvert, D. Conversion of a PLP-Dependent Racemase into an Aldolase by a Single Active Site Mutation. *J. Am. Chem. Soc.* **2003**, 125 (34), 10158–10159.
- (53) Arnold, F. H. Design by Directed Evolution. *Acc. Chem. Res.* **1998**, 31 (3), 125–131.
- (54) Farinas, E. T.; Bulter, T.; Arnold, F. H. Directed Enzyme Evolution. *Curr. Opin. Biotechnol.* **2001**, 12 (6), 545–551.
- (55) Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, 79 (1), 413-444.

- (56) Wu, Z. P.; Hilvert, D. Conversion of a Protease into an Acyl Transferase: Selenosubtilisin. *J. Am. Chem. Soc.* **1989**, *111* (12), 4513–4514.
- (57) Wu, Z. P.; Hilvert, D. Selenosubtilisin as a Glutathione Peroxidase Mimic. *J. Am. Chem. Soc.* **1990**, *112* (14), 5647–5648.
- (58) Dydio, P.; Key, H. M.; Nazarenko, A.; Rha, J. Y.-E.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F. An Artificial Metalloenzyme with the Kinetics of Native Enzymes. *Science* **2016**, *354* (6308), 102–106.
- (59) Hyster, T. K.; Knörr, L.; Ward, T. R.; Rovis, T. Biotinylated Rh(III) Complexes in Engineered Streptavidin for Accelerated Asymmetric C–H Activation. *Science* **2012**, *338* (6106), 500–503.
- (60) Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Röthlisberger, D.; Baker, D. New Algorithms and an in Silico Benchmark for Computational Enzyme Design. *Protein Sci.* **2006**, *15* (12), 2785–2794.
- (61) Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D. De Novo Enzyme Design Using Rosetta3. *PLoS One* **2011**, *6* (5), e19230.
- (62) Dahiyat, B. I.; Mayo, S. L. De Novo Protein Design: Fully Automated Sequence Selection. *Science* **1997**, *278* (5335), 82–87.
- (63) Bolon, D. N.; Mayo, S. L. Enzyme-like Proteins by Computational Design. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (25), 14274–14279.
- (64) Nosrati, G. R.; Houk, K. N. SABER: A Computational Method for Identifying Active Sites for New Reactions. *Protein Sci.* **2012**, *21* (5), 697–706.
- (65) Tantillo, D. J.; Jiangang, C.; Houk, K. N. Theozymes and Compuzymes: Theoretical Models for Biological Catalysis. *Curr. Opin. Chem. Biol.* **1998**, *2* (6), 743–750.
- (66) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St.Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **2010**, *329* (5989), 309–313.
- (67) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* **2008**, *453* (7192), 190–195.
- (68) Hilvert, D. Design of Protein Catalysts. *Annu. Rev. Biochem.* **2013**, *82* (1), 447–470.
- (69) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008**, *319* (5868), 1387–1391.
- (70) Richter, F.; Blomberg, R.; Khare, S. D.; Kiss, G.; Kuzin, A. P.; Smith, A. J. T.; Gallaher, J.; Pianowski, Z.; Helgeson, R. C.; Grjasnow, A.; Xiao, R.; Seetharaman, J.; Su, M.; Vorobiev, S.; Lew, S.; Forouhar, F.; Kornhaber, G. J.; Hunt, J. F.; Montelione, G. T.; Tong, L.; Houk, K. N.; Hilvert, D.; Baker, D. Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis. *J. Am. Chem. Soc.* **2012**, *134* (39), 16197–16206.

- (71) Hellinga, H. W.; Richards, F. M. Construction of New Ligand Binding Sites in Proteins of Known Structure: I. Computer-Aided Modeling of Sites with Pre-Defined Geometry. *J. Mol. Biol.* **1991**, *222* (3), 763–785.
- (72) Clarke, N. D.; Yuan, S.-M. Metal Search: A Computer Program That Helps Design Tetrahedral Metal-Binding Sites. *Proteins.* **1995**, *23* (2), 256–263.
- (73) Pinto, A. L.; Hellinga, H. W.; Caradonna, J. P. Construction of a Catalytically Active Iron Superoxide Dismutase by Rational Protein Design. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (11), 5562–5567.
- (74) Kipnis, Y.; Baker, D. Comparison of Designed and Randomly Generated Catalysts for Simple Chemical Reactions. *Protein Sci.* **2012**, *21* (9), 1388–1395.
- (75) Korendovych, I. V.; Kulp, D. W.; Wu, Y.; Cheng, H.; Roder, H.; DeGrado, W. F. Design of a Switchable Eliminase. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (17), 6823–6827.
- (76) Baker, D. An Exciting but Challenging Road Ahead for Computational Enzyme Design. *Protein Sci.* **2010**, *19* (10), 1817–1819.
- (77) Eiben, C. B.; Siegel, J. B.; Bale, J. B.; Cooper, S.; Khatib, F.; Shen, B. W.; Players, F.; Stoddard, B. L.; Popovic, Z.; Baker, D. Increased Diels-Alderase Activity through Backbone Remodeling Guided by Foldit Players. *Nat. Biotechnol.* **2012**, *30* (2), 190–192.
- (78) Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L. Iterative Approach to Computational Enzyme Design. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (10), 3790–3795.
- (79) Zeymer, C.; Hilvert, D. Directed Evolution of Protein Catalysts. *Annu. Rev. Biochem.* **2018**, *87* (1), 131–157.
- (80) Bunzel, H. A.; Garrabou, X.; Pott, M.; Hilvert, D. Speeding up Enzyme Discovery and Engineering with Ultrahigh-Throughput Methods. *Curr. Opin. Struct. Biol.* **2018**, *48*, 149–156.
- (81) Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Grütter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D. Precision Is Essential for Efficient Catalysis in an Evolved Kemp Eliminase. *Nature* **2013**, *503* (7476), 418–421.
- (82) Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D. Evolution of a Designed Retro-Aldolase Leads to Complete Active Site Remodeling. *Nat. Chem. Biol.* **2013**, *9* (8), 494–498.
- (83) Preiswerk, N.; Beck, T.; Schulz, J. D.; Milovnik, P.; Mayer, C.; Siegel, J. B.; Baker, D.; Hilvert, D. Impact of Scaffold Rigidity on the Design and Evolution of an Artificial Diels-Alderase. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (22), 8013–8018.
- (84) Bloom, J. D.; Meyer, M. M.; Meinhold, P.; Otey, C. R.; MacMillan, D.; Arnold, F. H. Evolving Strategies for Enzyme Engineering. *Curr. Opin. Struct. Biol.* **2005**, *15* (4), 447–452.
- (85) Davie, E. A. C.; Mennen, S. M.; Xu, Y.; Miller, S. J. Asymmetric Catalysis Mediated by Synthetic Peptides. *Chem. Rev.* **2007**, *107* (12), 5759–5812.
- (86) Sculimbrene, B. R.; Morgan, A. J.; Miller, S. J. Nonenzymatic Peptide-Based Catalytic Asymmetric Phosphorylation of Inositol Derivatives. *Chem. Commun.* **2003**, No. 15, 1781–1785.

- (87) Sculimbrene, B. R.; Morgan, A. J.; Miller, S. J. Enantiodivergence in Small-Molecule Catalysis of Asymmetric Phosphorylation: Concise Total Syntheses of the Enantiomeric d-Myo-Inositol-1-Phosphate and d-Myo-Inositol-3-Phosphate. *J. Am. Chem. Soc.* **2002**, *124* (39), 11653–11656.
- (88) Krattiger, P.; Kovasy, R.; Revell, J. D.; Ivan, S.; Wennemers, H. Increased Structural Complexity Leads to Higher Activity: Peptides as Efficient and Versatile Catalysts for Asymmetric Aldol Reactions. *Org. Lett.* **2005**, *7* (6), 1101–1103.
- (89) Johnsson, K.; Allemann, R. K.; Widmer, H.; Benner, S. A. Synthesis, Structure and Activity of Artificial, Rationally Designed Catalytic Polypeptides. *Nature* **1993**, *365* (6446), 530–532.
- (90) Taylor, S. E.; Rutherford, T. J.; Allemann, R. K. Design, Synthesis and Characterisation of a Peptide with Oxaloacetate Decarboxylase Activity. *Bioorg. Med. Chem. Lett.* **2001**, *11* (19), 2631–2635.
- (91) Taylor, S. E.; Rutherford, T. J.; Allemann, R. K. Design of a Folded, Conformationally Stable Oxaloacetate Decarboxylase. *J. Chem. Soc., Perkin Trans. 2* **2002**, No. 4, 751–755.
- (92) Woolfson, D. N. Coiled-Coil Design: Updated and Upgraded. *Fibrous Proteins: Structures and Mechanisms* **2017**, 35–61.
- (93) Woolfson, D. N.; Bartlett, G. J.; Burton, A. J.; Heal, J. W.; Niitsu, A.; Thomson, A. R.; Wood, C. W. De Novo Protein Design: How Do We Expand into the Universe of Possible Protein Structures? *Cur. Opin. Struct. Biol.* **2015**, *33*, 16–26.
- (94) Huang, P.-S.; Boyken, S. E.; Baker, D. The Coming of Age of de Novo Protein Design. *Nature* **2016**, *537* (7620), 320–327.
- (95) Thomson, A. R.; Wood, C. W.; Burton, A. J.; Bartlett, G. J.; Sessions, R. B.; Brady, R. L.; Woolfson, D. N. Computational Design of Water-Soluble  $\alpha$ -Helical Barrels. *Science* **2014**, *346* (6208), 485–488.
- (96) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **2003**, *302* (5649), 1364–1368.
- (97) Dawson, W. M.; Rhys, G. G.; Woolfson, D. N. Towards Functional de Novo Designed Proteins. *Curr. Opin. Chem. Biol.* **2019**, *52*, 102–111.
- (98) Peacock, A. F. Incorporating Metals into de Novo Proteins. *Curr. Opin. Chem. Biol.* **2013**, *17* (6), 934–939.
- (99) DeGrado, W. F.; Summa, C. M.; Pavone, V.; Natri, F.; Lombardi, A. De Novo Design and Structural Characterization of Proteins and Metalloproteins. *Annu. Rev. Biochem.* **1999**, *68* (1), 779–819.
- (100) Calhoun, J. R.; Natri, F.; Maglio, O.; Pavone, V.; Lombardi, A.; DeGrado, W. F. Artificial Diiron Proteins: From Structure to Function. *Pept. Sci.* **2005**, *80* (2–3), 264–278.
- (101) Lombardi, A.; Summa, C. M.; Geremia, S.; Randaccio, L.; Pavone, V.; DeGrado, W. F. Retrostructural Analysis of Metalloproteins: Application to the Design of a Minimal Model for Diiron Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (12), 6298–6305.
- (102) Computational de Novo Design, and Characterization of an A2B2 Diiron Protein. *J. Mol. Biol.* **2002**, *321* (5), 923–938.

- (103) Calhoun, J. R.; Kono, H.; Lahr, S.; Wang, W.; DeGrado, W. F.; Saven, J. G. Computational Design and Characterization of a Monomeric Helical Dinuclear Metalloprotein. *J. Mol. Biol.* **2003**, *334* (5), 1101–1115.
- (104) Torres Martin de Rosales, R.; Faiella, M.; Farquhar, E.; Que, L.; Andreato, C.; Pavone, V.; Maglio, O.; Natri, F.; Lombardi, A. Spectroscopic and Metal-Binding Properties of DF3: An Artificial Protein Able to Accommodate Different Metal Ions. *J. Biol. Inorg. Chem.* **2010**, *15* (5), 717–728.
- (105) Nanda, V.; Rosenblatt, M. M.; Osyczka, A.; Kono, H.; Getahun, Z.; Dutton, P. L.; Saven, J. G.; DeGrado, W. F. De Novo Design of a Redox-Active Minimal Rubredoxin Mimic. *J. Am. Chem. Soc.* **2005**, *127* (16), 5804–5805.
- (106) Gibney, B. R.; Mulholland, S. E.; Rabanal, F.; Dutton, P. L. Ferredoxin and Ferredoxin–Heme Maquettes. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93* (26), 15041–15046.
- (107) Scott, M. P.; Biggins, J. Introduction of a  $[4\text{Fe-4S (S-Cys)}_4]^{+1,+2}$  Iron-Sulfur Center into a Four- $\alpha$  Helix Protein Using Design Parameters from the Domain of the Fx Cluster in the Photosystem I Reaction Center. *Protein Sci.* **1997**, *6* (2), 340–346.
- (108) Grzyb, J.; Xu, F.; Weiner, L.; Reijerse, E. J.; Lubitz, W.; Nanda, V.; Noy, D. De Novo Design of a Non-Natural Fold for an Iron–Sulfur Protein: Alpha-Helical Coiled-Coil with a Four-Iron Four-Sulfur Cluster Binding Site in Its Central Core. *Biochim. Biophys. Acta, Bioenergetics* **2010**, *1797* (3), 406–413.
- (109) Der, B. S.; Machius, M.; Miley, M. J.; Mills, J. L.; Szyperski, T.; Kuhlman, B. Metal-Mediated Affinity and Orientation Specificity in a Computationally Designed Protein Homodimer. *J. Am. Chem. Soc.* **2012**, *134* (1), 375–385.
- (110) Reedy, C. J.; Gibney, B. R. Heme Protein Assemblies. *Chem. Rev.* **2004**, *104* (2), 617–650.
- (111) Lombardi, A.; Natri, F.; Pavone, V. Peptide-Based Heme–Protein Models. *Chem. Rev.* **2001**, *101* (10), 3165–3190.
- (112) Natri, F.; Lombardi, A.; Morelli, G.; Maglio, O.; D’Auria, G.; Pedone, C.; Pavone, V. Hemoprotein Models Based on a Covalent Helix–Heme–Helix Sandwich: 1. Design, Synthesis, and Characterization. *Chem. Eur. J.* **1997**, *3* (3), 340–349.
- (113) Lombardi, A.; Natri, F.; Sanseverino, M.; Maglio, O.; Pedone, C.; Pavone, V. Miniaturized Hemoproteins: Design, Synthesis and Characterization of Mimochrome II. *Inorg. Chim. Acta* **1998**, *275–276*, 301–313.
- (114) Koder, R. L.; Anderson, J. L. R.; Solomon, L. A.; Reddy, K. S.; Moser, C. C.; Dutton, P. L. Design and Engineering of an O<sub>2</sub> Transport Protein. *Nature* **2009**, *458* (7236), 305–309.
- (115) Cochran, F. V.; Wu, S. P.; Wang, W.; Nanda, V.; Saven, J. G.; Therien, M. J.; DeGrado, W. F. Computational De Novo Design and Characterization of a Four-Helix Bundle Protein That Selectively Binds a Nonbiological Cofactor. *J. Am. Chem. Soc.* **2005**, *127* (5), 1346–1347.
- (116) Fry, H. C.; Lehmann, A.; Saven, J. G.; DeGrado, W. F.; Therien, M. J. Computational Design and Elaboration of a de Novo Heterotetrameric  $\alpha$ -Helical Protein That Selectively Binds an Emissive Abiological (Porphinato)Zinc Chromophore. *J. Am. Chem. Soc.* **2010**, *132* (11), 3997–4005.
- (117) Fry, H. C.; Lehmann, A.; Sinks, L. E.; Asselberghs, I.; Tronin, A.; Krishnan, V.; Blasie, J. K.; Clays, K.; DeGrado, W. F.; Saven, J. G.; Therien, M. J. Computational de Novo Design and

Characterization of a Protein That Selectively Binds a Highly Hyperpolarizable Abiological Chromophore. *J. Am. Chem. Soc.* **2013**, *135* (37), 13914–13926.

(118) Polizzi, N. F.; Wu, Y.; Lemmin, T.; Maxwell, A. M.; Zhang, S.-Q.; Rawson, J.; Beratan, D. N.; Therien, M. J.; DeGrado, W. F. De Novo Design of a Hyperstable Non-Natural Protein–Ligand Complex with Sub-Å Accuracy. *Nat. Chem.* **2017**, *9* (12), 1157–1164.

(119) Sasaki, T.; Kaiser, E. T. Helichrome: Synthesis and Enzymic Activity of a Designed Hemeprotein. *J. Am. Chem. Soc.* **1989**, *111* (1), 380–381.

(120) Kaplan, J.; DeGrado, W. F. De Novo Design of Catalytic Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (32), 11566–11570.

(121) Faiella, M.; Andreozzi, C.; de Rosales, R. T. M.; Pavone, V.; Maglio, O.; Nistri, F.; DeGrado, W. F.; Lombardi, A. An Artificial Di-Iron Oxo-Protein with Phenol Oxidase Activity. *Nat. Chem. Biol.* **2009**, *5* (12), 882–884.

(122) Reig, A. J.; Pires, M. M.; Snyder, R. A.; Wu, Y.; Jo, H.; Kulp, D. W.; Butch, S. E.; Calhoun, J. R.; Szyperki, T.; Solomon, E. I.; DeGrado, W. F. Alteration of the Oxygen-Dependent Reactivity of de Novo De Fe<sub>2</sub> Proteins. *Nat. Chem.* **2012**, *4* (11), 900–906.

(123) Der, B. S.; Edwards, D. R.; Kuhlman, B. Catalysis by a De Novo Zinc-Mediated Protein Interface: Implications for Natural Enzyme Evolution and Rational Enzyme Engineering. *Biochemistry* **2012**, *51* (18), 3933–3940.

(124) Studer, S.; Hansen, D. A.; Pianowski, Z. L.; Mittl, P. R. E.; Debon, A.; Guffy, S. L.; Der, B. S.; Kuhlman, B.; Hilvert, D. Evolution of a Highly Active and Enantiospecific Metalloenzyme from Short Peptides. *Science* **2018**, *362* (6420), 1285–1288.

(125) Zastrow, M. L.; Peacock, A. F. A.; Stuckey, J. A.; Pecoraro, V. L. Hydrolytic Catalysis and Structural Stabilization in a Designed Metalloprotein. *Nat. Chem.* **2012**, *4* (2), 118–123.

(126) Zastrow, M. L.; Pecoraro, V. L. Influence of Active Site Location on Catalytic Activity in de Novo-Designed Zinc Metalloenzymes. *J. Am. Chem. Soc.* **2013**, *135* (15), 5895–5903.

(127) Tegoni, M.; Yu, F.; Bersellini, M.; Penner-Hahn, J. E.; Pecoraro, V. L. Designing a Functional Type 2 Copper Center That Has Nitrite Reductase Activity within  $\alpha$ -Helical Coiled Coils. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (52), 21234–21239.

(128) Popp, B. V.; Ball, Z. T. Structure-Selective Modification of Aromatic Side Chains with Dirhodium Metallopeptide Catalysts. *J. Am. Chem. Soc.* **2010**, *132* (19), 6660–6662.

(129) V. Popp, B.; T. Ball, Z. Proximity-Driven Metallopeptide Catalysis: Remarkable Side-Chain Scope Enables Modification of the Fos BZip Domain. *Chem.Sci.* **2011**, *2* (4), 690–695.

(130) Chen, Z.; Vohidov, F.; Coughlin, J. M.; Stagg, L. J.; Arold, S. T.; Ladbury, J. E.; Ball, Z. T. Catalytic Protein Modification with Dirhodium Metallopeptides: Specificity in Designed and Natural Systems. *J. Am. Chem. Soc.* **2012**, *134* (24), 10138–10145.

(131) Ranieri, A.; Monari, S.; Sola, M.; Borsari, M.; Battistuzzi, G.; Ringhieri, P.; Nistri, F.; Pavone, V.; Lombardi, A. Redox and Electrocatalytic Properties of Mimochrome VI, a Synthetic Heme Peptide Adsorbed on Gold. *Langmuir* **2010**, *26* (23), 17831–17835.



- (132) Nastri, F.; Lista, L.; Ringhieri, P.; Vitale, R.; Faiella, M.; Andreozzi, C.; Travascio, P.; Maglio, O.; Lombardi, A.; Pavone, V. A Heme–Peptide Metalloenzyme Mimetic with Natural Peroxidase-Like Activity. *Chem. Eur. J.* **2011**, *17* (16), 4444–4453.
- (133) Watkins, D. W.; Jenkins, J. M. X.; Grayson, K. J.; Wood, N.; Steventon, J. W.; Le Vay, K. K.; Goodwin, M. I.; Mullen, A. S.; Bailey, H. J.; Crump, M. P.; MacMillan, F.; Mulholland, A. J.; Cameron, G.; Sessions, R. B.; Mann, S.; Anderson, J. L. R. Construction and in Vivo Assembly of a Catalytically Proficient and Hyperthermostable de Novo Enzyme. *Nat. Commun.* **2017**, *8* (1), 358.
- (134) Severin, K.; Lee, D. H.; Kennan, A. J.; Ghadiri, M. R. A Synthetic Peptide Ligase. *Nature* **1997**, *389* (6652), 706–709.
- (135) Kennan, A. J.; Haridas, V.; Severin, K.; Lee, D. H.; Ghadiri, M. R. A de Novo Designed Peptide Ligase: A Mechanistic Investigation. *J. Am. Chem. Soc.* **2001**, *123* (9), 1797–1803.
- (136) Severin, K.; Lee, D. H.; Martinez, J. A.; Vieth, M.; Ghadiri, M. R. Dynamic Error Correction in Autocatalytic Peptide Networks. *Angew. Chem. Int. Ed.* **1998**, *37* (1–2), 126–128.
- (137) Saghatelian, A.; Yokobayashi, Y.; Soltani, K.; Ghadiri, M. R. A Chiroselective Peptide Replicator. *Nature* **2001**, *409* (6822), 797–801.
- (138) Yao, S.; Ghosh, I.; Zutshi, R.; Chmielewski, J. Selective Amplification by Auto- and Cross-Catalysis in a Replicating Peptide System. *Nature* **1998**, *396* (6710), 447–450.
- (139) Wilcoxon, K. M.; Leman, L. J.; Weinberger, D. A.; Huang, Z.-Z.; Ghadiri, M. R. Biomimetic Catalysis of Intermodular Aminoacyl Transfer. *J. Am. Chem. Soc.* **2007**, *129* (4), 748–749.
- (140) Leman, L. J.; Weinberger, D. A.; Huang, Z.-Z.; Wilcoxon, K. M.; Ghadiri, M. R. Functional and Mechanistic Analyses of Biomimetic Aminoacyl Transfer Reactions in de Novo Designed Coiled Coil Peptides via Rational Active Site Engineering. *J. Am. Chem. Soc.* **2007**, *129* (10), 2959–2966.
- (141) Broo, K. S.; Brive, L.; Ahlberg, P.; Baltzer, L. Catalysis of Hydrolysis and Transesterification Reactions of p-Nitrophenyl Esters by a Designed Helix–Loop–Helix Dimer. *J. Am. Chem. Soc.* **1997**, *119* (47), 11362–11372.
- (142) Baltzer, L.; Broo, K. S.; Nilsson, H.; Nilsson, J. Designed Four-Helix Bundle Catalysts—the Engineering of Reactive Sites for Hydrolysis and Transesterification Reactions of p-Nitrophenyl Esters. *Bioorg. Med. Chem.* **1999**, *7* (1), 83–91.
- (143) Nilsson, J.; Baltzer, L. Reactive-Site Design in Folded-Polypeptide Catalysts—The Leaving Group pKa of Reactive Esters Sets the Stage for Cooperativity in Nucleophilic and General-Acid Catalysis. *Chem. Eur. J.* **2000**, *6* (12), 2214–2220.
- (144) Razkin, J.; Lindgren, J.; Nilsson, H.; Baltzer, L. Enhanced Complexity and Catalytic Efficiency in the Hydrolysis of Phosphate Diesters by Rationally Designed Helix-Loop-Helix Motifs. *ChemBiochem* **2008**, *9* (12), 1975–1984.
- (145) Allert, M.; Baltzer, L. Setting the Stage for New Catalytic Functions in Designed Proteins—Exploring the Imine Pathway in the Efficient Decarboxylation of Oxaloacetate by an Arg–Lys Site in a Four-Helix Bundle Protein Scaffold. *Chem. Eur. J.* **2002**, *8* (11), 2549–2560.
- (146) Andersson, L. K.; Caspersson, M.; Baltzer, L. Control of Lysine Reactivity in Four-Helix Bundle Proteins by Site-Selective pKa Depression: Expanding the Versatility of Proteins by Postsynthetic Functionalisation. *Chem. Eur. J.* **2002**, *8* (16), 3687–3697.

- (147) Burton, A. J.; Thomson, A. R.; Dawson, W. M.; Brady, R. L.; Woolfson, D. N. Installing Hydrolytic Activity into a Completely de Novo Protein Framework. *Nat. Chem.* **2016**, *8* (9), 837–844.
- (148) Donnelly, A. E.; Murphy, G. S.; Digianantonio, K. M.; Hecht, M. H. A de Novo Enzyme Catalyzes a Life-Sustaining Reaction in Escherichia Coli. *Nat. Chem. Biol.* **2018**, *14* (3), 253–255.
- (149) Shogren-Knaak, M. A.; Imperiali, B. Modulating Pyridoxamine-Mediated Transamination through a B $\beta$  Motif Peptide Scaffold. *Bioorg. Mei. Chem.* **1999**, *7* (9), 1993–2002.
- (150) Coqui re, D.; Bos, J.; Beld, J.; Roelfes, G. Enantioselective Artificial Metalloenzymes Based on a Bovine Pancreatic Polypeptide Scaffold. *Angew. Chem. Int. Ed.* **2009**, *48* (28), 5159–5162.
- (151) Zozulia, O.; Dolan, M. A.; Korendovych, I. V. Catalytic Peptide Assemblies. *Chem. Soc. Rev.* **2018**, *47* (10), 3621–3639.
- (152) Makam, P.; Yamijala, S. S. R. K. C.; Tao, K.; Shimon, L. J. W.; Eisenberg, D. S.; Sawaya, M. R.; Wong, B. M.; Gazit, E. Non-Proteinaceous Hydrolase Comprised of a Phenylalanine Metallo-Supramolecular Amyloid-like Structure. *Nat. Catal.* **2019**, *2* (11), 977–985.
- (153) Rufo, C. M.; Moroz, Y. S.; Moroz, O. V.; St hr, J.; Smith, T. A.; Hu, X.; DeGrado, W. F.; Korendovych, I. V. Short Peptides Self-Assemble to Produce Catalytic Amyloids. *Nat. Chem.* **2014**, *6* (4), 303–309.
- (154) Makhlynets, O. V.; Gosavi, P. M.; Korendovych, I. V. Short Self-Assembling Peptides Are Able to Bind to Copper and Activate Oxygen. *Angew. Chem. Int. Ed.* **2016**, *55* (31), 9017–9020.
- (155) Lengyel, Z.; Rufo, C. M.; Moroz, Y. S.; Makhlynets, O. V.; Korendovych, I. V. Copper-Containing Catalytic Amyloids Promote Phosphoester Hydrolysis and Tandem Reactions. *ACS Catal.* **2018**, *8* (1), 59–62.
- (156) Omosun, T. O.; Hsieh, M.-C.; Childers, W. S.; Das, D.; Mehta, A. K.; Anthony, N. R.; Pan, T.; Grover, M. A.; Berland, K. M.; Lynn, D. G. Catalytic Diversity in Self-Propagating Peptide Assemblies. *Nat. Chem.* **2017**, *9* (8), 805–809.
- (157) Song, W. J.; Tezcan, F. A. A Designed Supramolecular Protein Assembly with in Vivo Enzymatic Activity. *Science* **2014**, *346* (6216), 1525–1528.
- (158) Leaver-Fay, A.; Jacak, R.; Stranges, P. B.; Kuhlman, B. A Generic Program for Multistate Protein Design. *PLoS One* **2011**, *6* (7), e20937.
- (159) L ffler, P.; Schmitz, S.; Hupfeld, E.; Sterner, R.; Merkl, R. Rosetta:MSF: A Modular Framework for Multi-State Computational Protein Design. *PLoS Comput. Biol.* **2017**, *13* (6), e1005600.
- (160) Davey, J. A.; Damry, A. M.; Goto, N. K.; Chica, R. A. Rational Design of Proteins That Exchange on Functional Timescales. *Nat. Chem. Biol.* **2017**, *13* (12), 1280–1285.
- (161) Doudna, J. A.; Lorsch, J. R. Ribozyme Catalysis: Not Different, Just Worse. *Nat. Struct. Mol. Biol.* **2005**, *12* (5), 395–402.
- (162) Doudna, J. A.; Cech, T. R. The Chemical Repertoire of Natural Ribozymes. *Nature* **2002**, *418* (6894), 222–228.
- (163) Joyce, G. F. The Antiquity of RNA-Based Evolution. *Nature* **2002**, *418* (6894), 214–221.
- (164) Robertson, M. P.; Joyce, G. F. The Origins of the RNA World. *Cold Spring Harbor Perspect. Biol.* **2012**, *4* (5), a003608.

- (165) Jäschke, A. Artificial Ribozymes and Deoxyribozymes. *Curr. Opin. Struct. Biol.* **2001**, *11* (3), 321–326.
- (166) Seelig, B.; Keiper, S.; Stuhlmann, F.; Jäschke, A. Enantioselective Ribozyme Catalysis of a Bimolecular Cycloaddition Reaction. *Angew. Chem. Int. Ed.* **2000**, *39* (24), 4576–4579.
- (167) Porta, H.; Lizardi, P. M. An Allosteric Hammerhead Ribozyme. *Nat. Biotechnol.* **1995**, *13* (2), 161–164.
- (168) Hill, D. J.; Mio, M. J.; Prince, R. B.; Hughes, T. S.; Moore, J. S. A Field Guide to Foldamers. *Chem. Rev.* **2001**, *101* (12), 3893–4012.
- (169) Goodman, C. M.; Choi, S.; Shandler, S.; DeGrado, W. F. Foldamers as Versatile Frameworks for the Design and Evolution of Function. *Nat. Chem. Biol.* **2007**, *3* (5), 252–262.
- (170) Müller, M. M.; Windsor, M. A.; Pomerantz, W. C.; Gellman, S. H.; Hilvert, D. A Rationally Designed Aldolase Foldamer. *Angew. Chem. Int. Ed.* **2009**, *48* (5), 922–925.
- (171) Girvin, Z. C.; Gellman, S. H. Exploration of Diverse Reactive Diad Geometries for Bifunctional Catalysis via Foldamer Backbone Variation. *J. Am. Chem. Soc.* **2018**, *140* (39), 12476–12483.
- (172) Girvin, Z. C.; Andrews, M. K.; Liu, X.; Gellman, S. H. Foldamer-Templated Catalysis of Macrocyclic Formation. *Science* **2019**, *366* (6472), 1528–1531.
- (173) Bécart, D.; Diemer, V.; Salaün, A.; Oiarbide, M.; Nelli, Y. R.; Kauffmann, B.; Fischer, L.; Palomo, C.; Guichard, G. Helical Oligoureia Foldamers as Powerful Hydrogen Bonding Catalysts for Enantioselective C–C Bond-Forming Reactions. *J. Am. Chem. Soc.* **2017**, *139* (36), 12524–12532.
- (174) Le Bailly, B. A. F.; Byrne, L.; Clayden, J. Refoldable Foldamers: Global Conformational Switching by Deletion or Insertion of a Single Hydrogen Bond. *Angew. Chem. Int. Ed.* **2016**, *55* (6), 2132–2136.
- (175) Smaldone, R. A.; Moore, J. S. Reactive Sieving with Foldamers: Inspiration from Nature and Directions for the Future. *Chem. Eur. J.* **2008**, *14* (9), 2650–2657.
- (176) Smaldone, R. A.; Moore, J. S. Foldamers as Reactive Sieves: Reactivity as a Probe of Conformational Flexibility. *J. Am. Chem. Soc.* **2007**, *129* (17), 5444–5450.
- (177) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *J. Comb. Chem.* **1999**, *1* (1), 55–68.
- (178) Roughley, S. D.; Jordan, A. M. The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates. *J. Med. Chem.* **2011**, *54* (10), 3451–3479.
- (179) Marchildon, K. Polyamides – Still Strong After Seventy Years. *Macromol. React. Eng.* **2011**, *5* (1), 22–54.
- (180) Mahesh, S.; Tang, K.-C.; Raj, M. Amide Bond Activation of Biological Molecules. *Molecules* **2018**, *23* (10), 2615.
- (181) Radzicka, A.; Wolfenden, R. Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases. *J. Am. Chem. Soc.* **1996**, *118* (26), 6105–6109.
- (182) Kemnitz, C. R.; Loewen, M. J. "Amide Resonance" Correlates with a Breadth of C–N Rotation Barriers. *J. Am. Chem. Soc.* **2007**, *129* (9), 2521–2528.

- (183) Pattabiraman, V. R.; Bode, J. W. Rethinking Amide Bond Synthesis. *Nature* **2011**, *480* (7378), 471–479.
- (184) Montalbetti, C. A. G. N.; Falque, V. Amide Bond Formation and Peptide Coupling. *Tetrahedron* **2005**, *61* (46), 10827–10852.
- (185) Ben-Shem, A.; Loubresse, N. G. de; Melnikov, S.; Jenner, L.; Yusupova, G.; Yusupov, M. The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution. *Science* **2011**, *334* (6062), 1524–1529.
- (186) Wilson, D. N.; Cate, J. H. D. The Structure and Function of the Eukaryotic Ribosome. *Cold Spring Harbor Perspect. Biol.* **2012**, *4* (5), a011536.
- (187) Schwarzer, D.; Finking, R.; Marahiel, M. A. Nonribosomal Peptides: From Genes to Products. *Nat. Prod. Rep.* **2003**, *20* (3), 275.
- (188) Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev.* **1997**, *97* (7), 2651–2674.
- (189) Strieker, M.; Tanović, A.; Marahiel, M. A. Nonribosomal Peptide Synthetases: Structures and Dynamics. *Curr. Opin. Struct. Biol.* **2010**, *20* (2), 234–240.
- (190) Yang, W.; Drueckhammer, D. G. Understanding the Relative Acyl-Transfer Reactivity of Oxoesters and Thioesters: Computational Analysis of Transition State Delocalization Effects. *J. Am. Chem. Soc.* **2001**, *123* (44), 11004–11009.
- (191) Pickart, C. M. Mechanisms Underlying Ubiquitination. *Annu. Rev. Biochem.* **2001**, *70* (1), 503–533.
- (192) Berndsen, C. E.; Wolberger, C. New Insights into Ubiquitin E3 Ligase Mechanism. *Nat. Struct. Mol. Biol.* **2014**, *21* (4), 301–307.
- (193) Ton-That, H.; Liu, G.; Mazmanian, S. K.; Faull, K. F.; Schneewind, O. Purification and Characterization of Sortase, the Transpeptidase That Cleaves Surface Proteins of *Staphylococcus Aureus* at the LPXTG Motif. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (22), 12424–12429.
- (194) Nguyen, G. K. T.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J. P. Butelase 1 Is an Asx-Specific Ligase Enabling Peptide Macrocyclization and Synthesis. *Nat. Chem. Biol.* **2014**, *10* (9), 732–738.
- (195) Hofmann, A. W. Ueber Die Darstellung Der Amide Einbasischer Säuren Der Aliphatischen Reihe. *Ber. Dtsch. Chem. Ges.* **1882**, *15* (1), 977–984.
- (196) Mitchell, J. A.; Reid, E. Emmet. The Preparation of Aliphatic Amides. *J. Am. Chem. Soc.* **1931**, *53* (5), 1879–1883.
- (197) Jursic, B. S.; Zdravkovski, Z. A Simple Preparation of Amides from Acids and Amines by Heating of Their Mixture. *Synth. Commun.* **1993**, *23* (19), 2761–2770.
- (198) Sheehan, J. C.; Hess, G. P. A New Method of Forming Peptide Bonds. *J. Am. Chem. Soc.* **1955**, *77* (4), 1067–1068.
- (199) König, W.; Geiger, R. Eine Neue Methode Zur Synthese von Peptiden: Aktivierung Der Carboxylgruppe Mit Dicyclohexylcarbodiimid Unter Zusatz von 1-Hydroxy-Benzotriazolen. *Chem. Ber.* **1970**, *103* (3), 788–798.
- (200) Kisfaludy, L.; Schön, I.; Szirtes, T.; Nyéki, O.; Lőw, M. A Novel and Rapid Peptide Synthesis. *Tetrahedron Lett.* **1974**, *15* (19), 1785–1786.

- (201) Kisfaludy, L.; Schön, I. Preparation and Applications of Pentafluorophenyl Esters of 9-Fluorenylmethoxycarbonyl Amino Acids for Peptide Synthesis. *Synthesis* **1983**, 1983 (4), 325–327.
- (202) Anderson, G. W.; Zimmerman, J. E.; Callahan, F. M. N-Hydroxysuccinimide Esters in Peptide Synthesis. *J. Am. Chem. Soc.* **1963**, 85 (19), 3039–3039.
- (203) Anderson, G. W.; Zimmerman, J. E.; Callahan, F. M. The Use of Esters of N-Hydroxysuccinimide in Peptide Synthesis. *J. Am. Chem. Soc.* **1964**, 86 (9), 1839–1842.
- (204) 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.
- (205) Valeur, E.; Bradley, M. Amide Bond Formation: Beyond the Myth of Coupling Reagents. *Chem. Soc. Rev.* **2009**, 38 (2), 606–631.
- (206) Dourtoglou, V.; Ziegler, J.-C.; Gross, B. L'hexafluorophosphate de O-benzotriazolyl-N,N-tetramethyluronium: Un reactif de couplage peptidique nouveau et efficace. *Tetrahedron Lett.* **1978**, 19 (15), 1269–1272.
- (207) Coin, I.; Beyermann, M.; Bienert, M. Solid-Phase Peptide Synthesis: From Standard Procedures to the Synthesis of Difficult Sequences. *Nat. Protoc.* **2007**, 2 (12), 3247–3256.
- (208) Carpino, L. A. 1-Hydroxy-7-Azabenzotriazole. An Efficient Peptide Coupling Additive. *J. Am. Chem. Soc.* **1993**, 115 (10), 4397–4398.
- (209) Bryan, M. C.; Dunn, P. J.; Entwistle, D.; Gallou, F.; Koenig, S. G.; Hayler, J. D.; Hickey, M. R.; Hughes, S.; Kopach, M. E.; Moine, G.; Richardson, P.; Roschangar, F.; Steven, A.; Weiberth, F. J. Key Green Chemistry Research Areas from a Pharmaceutical Manufacturers' Perspective Revisited. *Green Chem.* **2018**, 20 (22), 5082–5103.
- (210) Ishihara, K.; Ohara, S.; Yamamoto, H. 3,4,5-Trifluorobenzeneboronic Acid as an Extremely Active Amidation Catalyst. *J. Org. Chem.* **1996**, 61 (13), 4196–4197.
- (211) Al-Zoubi, R. M.; Marion, O.; Hall, D. G. Direct and Waste-Free Amidations and Cycloadditions by Organocatalytic Activation of Carboxylic Acids at Room Temperature. *Angew. Chem. Int. Ed.* **2008**, 47 (15), 2876–2879.
- (212) Lundberg, H.; Tinnis, F.; Selander, N.; Adolfsson, H. Catalytic Amide Formation from Non-Activated Carboxylic Acids and Amines. *Chem. Soc. Rev.* **2014**, 43 (8), 2714–2742.
- (213) Wu, H.; Handoko; Raj, M.; Arora, P. S. Iterative Design of a Biomimetic Catalyst for Amino Acid Thioester Condensation. *Org. Lett.* **2017**, 19 (19), 5122–5125.
- (214) Tamaru, Y.; Yamada, Y.; Yoshida, Z. Direct Oxidative Transformation of Aldehydes to Amides by Palladium Catalysis. *Synthesis* **1983**, 1983 (6), 474–476.
- (215) Allen, C. L.; Williams, J. M. J. Metal-Catalysed Approaches to Amide Bond Formation. *Chem. Soc. Rev.* **2011**, 40 (7), 3405–3415.
- (216) Dorr, B. M.; Fuerst, D. E. Enzymatic Amidation for Industrial Applications. *Curr. Opin. Chem. Biol.* **2018**, 43, 127–133.
- (217) Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D. The Future of Peptide-Based Drugs. *Chem/ Biol. Drug Des.* **2013**, 81 (1), 136–147.

- (218) Nallamsetty, S.; Waugh, D. S. A Generic Protocol for the Expression and Purification of Recombinant Proteins in Escherichia Coli Using a Combinatorial His 6 -Maltose Binding Protein Fusion Tag. *Nat. Protoc.* **2007**, 2 (2), 383–391.
- (219) Rosano, G. L.; Ceccarelli, E. A. Recombinant Protein Expression in Escherichia Coli: Advances and Challenges. *Front. Microbiol.* **2014**, 5.
- (220) Cohen, S. N.; Chang, A. C. Y.; Boyer, H. W.; Helling, R. B. Construction of Biologically Functional Bacterial Plasmids In Vitro. *Proc. Natl. Acad. Sci. U.S.A* **1973**, 70 (11), 3240–3244.
- (221) Çelik, E.; Çalık, P. Production of Recombinant Proteins by Yeast Cells. *Biotechnol. Adv.* **2012**, 30 (5), 1108–1118.
- (222) Hunter, M.; Yuan, P.; Vavilala, D.; Fox, M. Optimization of Protein Expression in Mammalian Cells. *Curr. Protoc. in Protein Sci.* **2019**, 95 (1), e77.
- (223) Morrison, K. L.; Weiss, G. A. Combinatorial Alanine-Scanning. *Curr. Opin. Chem. Biol.* **2001**, 5 (3), 302–307.
- (224) Takatsujii, R.; Shinbara, K.; Katoh, T.; Goto, Y.; Passioura, T.; Yajima, R.; Komatsu, Y.; Suga, H. Ribosomal Synthesis of Backbone-Cyclic Peptides Compatible with In Vitro Display. *J. Am. Chem. Soc.* **2019**, 141 (6), 2279–2287.
- (225) Kent, S. B. H. Novel Protein Science Enabled by Total Chemical Synthesis. *Protein Sci.* **2019**, 28 (2), 313–328.
- (226) Bondalapati, S.; Jbara, M.; Brik, A. Expanding the Chemical Toolbox for the Synthesis of Large and Uniquely Modified Proteins. *Nat. Chem.* **2016**, 8 (5), 407–418.
- (227) Wang, P.; Dong, S.; Shieh, J.-H.; Peguero, E.; Hendrickson, R.; Moore, M. A. S.; Danishefsky, S. J. Erythropoietin Derived by Chemical Synthesis. *Science* **2013**, 342 (6164), 1357–1360.
- (228) Samarasimhareddy, M.; Mayer, D.; Metanis, N.; Veprintsev, D.; Hurevich, M.; Friedler, A. A Targeted Approach for the Synthesis of Multi-Phosphorylated Peptides: A Tool for Studying the Role of Phosphorylation Patterns in Proteins. *Org. Biomol. Chem.* **2019**, 17 (42), 9284–9290.
- (229) Dang, B.; Wu, H.; Mulligan, V. K.; Mravic, M.; Wu, Y.; Lemmin, T.; Ford, A.; Silva, D.-A.; Baker, D.; DeGrado, W. F. De Novo Design of Covalently Constrained Mesosize Protein Scaffolds with Unique Tertiary Structures. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, 114 (41), 10852–10857.
- (230) Mandal, K.; Uppalapati, M.; Ault-Riché, D.; Kenney, J.; Lowitz, J.; Sidhu, S. S.; Kent, S. B. H. Chemical Synthesis and X-Ray Structure of a Heterochiral {D-Protein Antagonist plus Vascular Endothelial Growth Factor} Protein Complex by Racemic Crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, 109 (37), 14779–14784.
- (231) Weinstock, M. T.; Jacobsen, M. T.; Kay, M. S. Synthesis and Folding of a Mirror-Image Enzyme Reveals Ambidextrous Chaperone Activity. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, 111 (32), 11679–11684.
- (232) Mandal, K.; Pentelute, B. L.; Bang, D.; Gates, Z. P.; Torbeev, V. Y.; Kent, S. B. H. Design, Total Chemical Synthesis, and X-Ray Structure of a Protein Having a Novel Linear-Loop Polypeptide Chain Topology. *Angew. Chem. Int. Ed.* **2012**, 51 (6), 1481–1486.
- (233) Boehringer, R.; Kieffer, B.; Torbeev, V. Total Chemical Synthesis and Biophysical Properties of a Designed Soluble 24 KDa Amyloid Analogue. *Chem. Sci.* **2018**, 9 (25), 5594–5599.

- (234) Kimmerlin, T.; Seebach, D. '100 Years of Peptide Synthesis': Ligation Methods for Peptide and Protein Synthesis with Applications to  $\beta$ -Peptide Assemblies. *J. Pept. Res.* **2005**, *65* (2), 229–260.
- (235) Curtius, T. Ueber Einige Neue Der Hippursäure Analog Constituirte, Synthetisch Dargestellte Amidosäuren. *J. Prakt. Chem.* **1882**, *26* (1), 145–208.
- (236) Curtius, T. Verkettung von Amidosäuren I. Abhandlung. *J. Prakt. Chem.* **1904**, *70* (1), 57–72.
- (237) Fischer, E.; Fourneau, E. Ueber Einige Derivate Des Glykocolls. *Ber. Dtsch. Chem. Ges.* **1901**, *34* (2), 2868–2877.
- (238) Fischer, E. Synthese von Polypeptiden. IX. Chloride Der Aminosäuren Und Ihrer Acylderivate. *Ber. Dtsch. Chem. Ges.* **1905**, *38* (1), 605–619.
- (239) Fischer, E. Synthese von Polypeptiden. XVII. *Ber. Dtsch. Chem. Ges.* **1907**, *40* (2), 1754–1767.
- (240) Bergmann, M.; Zervas, L. Über Ein Allgemeines Verfahren Der Peptid-Synthese. *Ber. Dtsch. Chem. Ges.* **1932**, *65* (7), 1192–1201.
- (241) Anderson, G. W.; Blodinger, J.; Welcher, A. D. Tetraethyl Pyrophosphite as a Reagent for Peptide Syntheses1. *J. Am. Chem. Soc.* **1952**, *74* (21), 5309–5312.
- (242) du Vigneaud, V.; Ressler, C.; Swan, J. M.; Roberts, C. W.; Katsoyannis, P. G. The Synthesis of Oxytocin1. *J. Am. Chem. Soc.* **1954**, *76* (12), 3115–3121.
- (243) Denkwalter, R. G.; Veber, D. F.; Holly, F. W.; Hirschmann, R. Total Synthesis of an Enzyme. I. Objective and Strategy. *J. Am. Chem. Soc.* **1969**, *91* (2), 502–503.
- (244) Hirschmann, R.; Nutt, R. F.; Veber, D. F.; Vitali, R. A.; Varga, S. L.; Jacob, T. A.; Holly, F. W.; Denkwalter, R. G. Total Synthesis of an Enzyme. V. Preparation of Enzymatically Active Material. *J. Am. Chem. Soc.* **1969**, *91* (2), 507–508.
- (245) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154.
- (246) Gutte, B.; Merrifield, R. B. Total Synthesis of an Enzyme with Ribonuclease A Activity. *J. Am. Chem. Soc.* **1969**, *91* (2), 501–502.
- (247) Gutte, B.; Merrifield, R. B. The Synthesis of Ribonuclease A. *J. Biol. Chem.* **1971**, *246* (6), 1922–1941.
- (248) Merrifield, R. B.; Stewart, J. Morrow.; Jernberg, Nils. Instrument for Automated Synthesis of Peptides. *Anal. Chem.* **1966**, *38* (13), 1905–1914.
- (249) Muttenthaler, M.; Albericio, F.; Dawson, P. E. Methods, Setup and Safe Handling for Anhydrous Hydrogen Fluoride Cleavage in Boc Solid-Phase Peptide Synthesis. *Nat. Protoc.* **2015**, *10* (7), 1067–1083.
- (250) Gates, Z. P.; Dhayalan, B.; Kent, S. B. H. Obviation of Hydrogen Fluoride in Boc Chemistry Solid Phase Peptide Synthesis of Peptide-Athioesters. *Chem. Commun.* **2016**, *52* (97), 13979–13982.
- (251) Kemp, D. S.; Galakatos, N. G.; Bowen, B.; Tan, K. Peptide Synthesis by Prior Thiol Capture. 2. Design of Templates for Intramolecular O,N-Acyl Transfer. 4,6-Disubstituted Dibenzofurans as Optimal Spacing Elements. *J. Org. Chem.* **1986**, *51* (10), 1829–1838.
- (252) Schnolzer, M.; Kent, S. B. Constructing Proteins by Dovetailing Unprotected Synthetic Peptides: Backbone-Engineered HIV Protease. *Science* **1992**, *256* (5054), 221–225.

- (253) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, 266 (5186), 776–779.
- (254) Johnson, E. C. B.; Kent, S. B. H. Insights into the Mechanism and Catalysis of the Native Chemical Ligation Reaction. *J. Am. Chem. Soc.* **2006**, 128 (20), 6640–6646.
- (255) Torbeev, V. Y.; 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.
- (256) Sun, H.; Brik, A. The Journey for the Total Chemical Synthesis of a 53 KDa Protein. *Acc. Chem. Res.* **2019**, 52 (12), 3361–3371.
- (257) Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. Fmoc-Based Synthesis of Peptide-AThioesters: Application to the Total Chemical Synthesis of a Glycoprotein by Native Chemical Ligation. *J. Am. Chem. Soc.* **1999**, 121 (50), 11684–11689.
- (258) Ingenito, R.; Bianchi, E.; Fattori, D.; Pessi, A. Solid Phase Synthesis of Peptide C-Terminal Thioesters by Fmoc/t-Bu Chemistry. *J. Am. Chem. Soc.* **1999**, 121 (49), 11369–11374.
- (259) 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.* **2011**, 50 (33), 7645–7649.
- (260) 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.
- (261) 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.* **2008**, 120 (36), 6957–6961.
- (262) 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.
- (263) Hojo, H.; Onuma, Y.; Akimoto, Y.; Nakahara, Y.; Nakahara, Y. N-Alkyl Cysteine-Assisted Thioesterification of Peptides. *Tetrahedron Lett.* **2007**, 48 (1), 25–28.
- (264) Ollivier, N.; Dheur, J.; Mhidia, R.; Blanpain, A.; Melnyk, O. Bis(2-Sulfanylethyl)Amino Native Peptide Ligation. *Org. Lett.* **2010**, 12 (22), 5238–5241.
- (265) 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.
- (266) Terrier, V. P.; Adihou, H.; Arnould, M.; Delmas, A. F.; Aucagne, V. A Straightforward Method for Automated Fmoc-Based Synthesis of Bio-Inspired Peptide Crypto-Thioesters. *Chem. Sci.* **2015**, 7 (1), 339–345.
- (267) Muir, T. W.; Sondhi, D.; Cole, P. A. Expressed Protein Ligation: A General Method for Protein Engineering. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, 95 (12), 6705–6710.
- (268) Yan, L. Z.; Dawson, P. E. Synthesis of Peptides and Proteins without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization. *J. Am. Chem. Soc.* **2001**, 123 (4), 526–533.



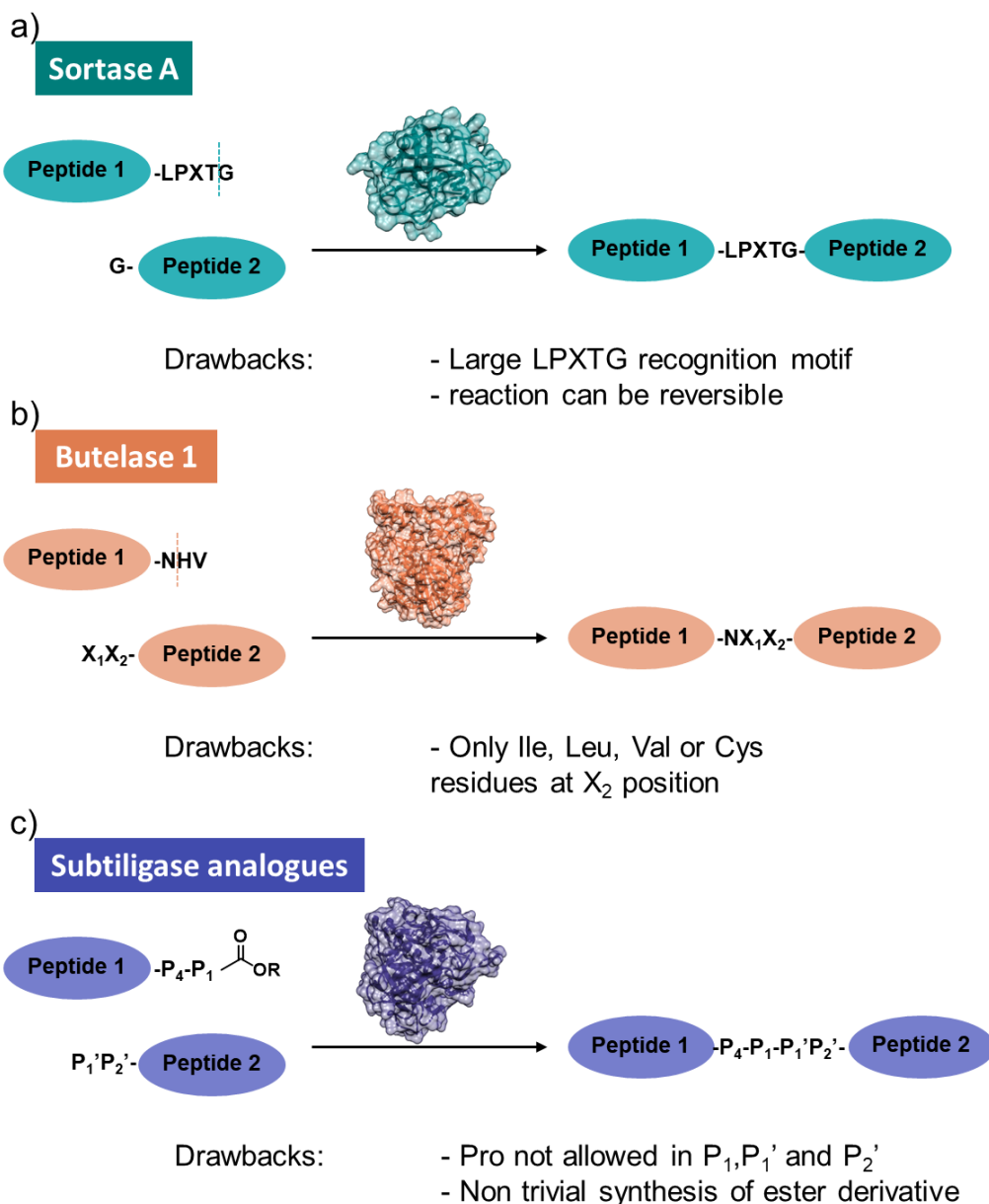
- (269) Wan, Q.; Danishefsky, S. J. Free-Radical-Based, Specific Desulfurization of Cysteine: A Powerful Advance in the Synthesis of Polypeptides and Glycopolypeptides. *Angew. Chem. Int. Ed.* **2007**, *46* (48), 9248–9252.
- (270) Malins, L. R.; Payne, R. J. Recent Extensions to Native Chemical Ligation for the Chemical Synthesis of Peptides and Proteins. *Curr. Opin. Chem. Biol.* **2014**, *22*, 70–78.
- (271) Loibl, S. F.; Harpaz, Z.; Seitz, O. A Type of Auxiliary for Native Chemical Peptide Ligation beyond Cysteine and Glycine Junctions. *Angew. Chem. Int. Ed.* **2015**, *54* (50), 15055–15059.
- (272) Nadler, C.; Nadler, A.; Hansen, C.; Diederichsen, U. A Photocleavable Auxiliary for Extended Native Chemical Ligation. *Eur. J. Org. Chem.* **2015**, *2015* (14), 3095–3102.
- (273) 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.
- (274) 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.
- (275) Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. High-Yielding Staudinger Ligation of a Phosphinothioester and Azide To Form a Peptide. *Org. Lett.* **2001**, *3* (1), 9–12.
- (276) Bode, J. W.; Fox, R. M.; Baucom, K. D. Chemoselective Amide Ligations by Decarboxylative Condensations of N-Alkylhydroxylamines and  $\alpha$ -Ketoacids. *Angew. Chem. Int. Ed.* **2006**, *45* (8), 1248–1252.
- (277) Pattabiraman, V. R.; Ogunkoya, A. O.; Bode, J. W. Chemical Protein Synthesis by Chemoselective  $\alpha$ -Ketoacid–Hydroxylamine (KAHA) Ligations with 5-Oxaproline. *Angew. Chem. Int. Ed.* **2012**, *51* (21), 5114–5118.
- (278) Pusterla, I.; Bode, J. W. An Oxazetidine Amino Acid for Chemical Protein Synthesis by Rapid, Serine-Forming Ligations. *Nat. Chem.* **2015**, *7* (8), 668–672.
- (279) Schmidt, M.; Toplak, A.; Quaedflieg, P. J. L. M.; van Maarseveen, J. H.; Nuijens, T. Enzyme-Catalyzed Peptide Cyclization. *Drug Discov. Today Technol.* **2017**, *26*, 11–16.
- (280) Abrahmsen, L.; Tom, J.; Burnier, J.; Butcher, K. A.; Kossiakoff, A.; Wells, J. A. Engineering Subtilisin and Its Substrates for Efficient Ligation of Peptide Bonds in Aqueous Solution. *Biochemistry* **1991**, *30* (17), 4151–4159.
- (281) Weeks, A. M.; Wells, J. A. Subtiligase-Catalyzed Peptide Ligation. *Chem. Rev.* **2020**, *120* (6), 3127–3160.
- (282) Jackson, D. Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J. A. A Designed Peptide Ligase for Total Synthesis of Ribonuclease A with Unnatural Catalytic Residues. *Science* **1994**, *266* (5183), 243–247.
- (283) Schmidt, M.; Toplak, A.; Quaedflieg, P. J. L. M.; Ippel, H.; Richelle, G. J. J.; Hackeng, T. M.; van Maarseveen, J. H.; Nuijens, T. Omniligase-1: A Powerful Tool for Peptide Head-to-Tail Cyclization. *Adv. Synth. Catal.* **2017**, *359* (12), 2050–2055.
- (284) Toplak, A.; Nuijens, T.; Quaedflieg, P. J. L. M.; Wu, B.; Janssen, D. B. Peptiligase, an Enzyme for Efficient Chemoenzymatic Peptide Synthesis and Cyclization in Water. *Adv. Synth. Catal.* **2016**, *358* (13), 2140–2147.

- (285) Antos, J. M.; Truttmann, M. C.; Ploegh, H. L. Recent Advances in Sortase-Catalyzed Ligation Methodology. *Curr. Opin. Struct. Biol.* **2016**, *38*, 111–118.
- (286) Nguyen, G. K. T.; Qiu, Y.; Cao, Y.; Hemu, X.; Liu, C.-F.; Tam, J. P. Butelase-Mediated Cyclization and Ligation of Peptides and Proteins. *Nat. Protoc.* **2016**, *11* (10), 1977–1988.
- (287) Nguyen, G. K. T.; Cao, Y.; Wang, W.; Liu, C. F.; Tam, J. P. Site-Specific N-Terminal Labeling of Peptides and Proteins Using Butelase 1 and Thiodepsipeptide. *Angew. Chem. Int. Ed.* **2015**, *54* (52), 15694–15698.

**Chapter 2. Design, synthesis and  
characterization of *de novo* catalytic  
proteins for acyl transfer reactions**

## 2.1. Introduction

In this chapter, I apply rational minimalist design to install a function of catalysis into an already existing *de novo* protein scaffold for miscellaneous acyl transfer reactions and more particularly for amide bond formation. Amide linkages are encountered in a variety of synthetic and natural molecules such as drugs, synthetic polymers and proteins.<sup>1,2</sup> Despite impressive progress of various methods to particularly activate carboxylic acid, amide bond formation remains to be a synthetic challenge and further improvement related to catalysis, waste reduction and chemoselectivity can still be accomplished.<sup>3</sup> In the field of total chemical synthesis of proteins, these issues are especially important for amino acid activation and the need of high excesses of reagents during coupling in solid-phase peptide synthesis (SPPS). Additionally, for the peptide ligation, the most widespread method, the native chemical ligation (NCL), requires for instance a cysteine residue in the sequence, which one of the less common amino acids found in native protein. We are more interested in developing alternatives for the latter. Besides different chemical approaches (such as post-ligation desulfurization of Cys to Ala, traceless Staudinger ligation, auxiliary-mediated ligations, etc.),<sup>4-6</sup> enzyme-catalyzed ligation strategies were also attempted. Naturally occurring ligases such sortase A<sup>7,8</sup> or butelase 1<sup>9,10</sup> have been utilized for peptide ligation or cyclization, which has illustrated the feasibility of high catalytic efficiencies for condensation of acyl donor and acceptor peptides substrates and their potential for practical applications.<sup>11</sup> In addition, reengineered proteases have been demonstrated to perform condensation of peptides.<sup>12</sup> For instance, previously reengineered variant of subtilisin was successfully used to assemble a 124-residue ribonuclease A.<sup>13</sup> Furthermore, newer and more efficient variants of proteolytic proteins with ligation catalytic activity were recently developed.<sup>14,15</sup> However, both natural ligases and reengineered proteases possess restrictions to amino acids sequences that can bind to their substrate pockets (**Figure 2.1**); in other words, the catalytic efficiency is coupled to substrate specificity in these enzymes. Moreover, the overall complexity of these large and evolved structures complicates the improvement and further tuning of their enzymatic properties.



**Figure 2.1. Enzyme-catalyzed peptide bond formation.** Overview of the necessary prerequisites at the junction site for the enzymes generally used to catalyze peptide ligation: a) the sortase A, b) the butelase 1, c) subtiligase analogues such as peptiligase or omniligase. Structures from PDB: 3RCC, 6DHI, 5OX2.

To have a more general tool that is not limited by substrate scope and can be compatible with a variety of reaction conditions, we introduced a catalytic ligation activity from first principles into a *de novo* protein scaffold. *De novo* proteins are characterized by their miniaturized framework, in which folding and stability are well understood. Their robustness, modularity and designability allow for introducing functional residues in a rather flexible manner. Moreover, designing an explicit binding site is not necessarily a prerequisite since the goal is mainly to obtain some substantial acceleration and to develop general tool via promiscuous activity compatible with multiple-substrate binding. On the other hand, by precise placement of the catalytic residues and by adjustment of the environment near the

resulting catalytic site, we can also get better insights into the fundamental principles necessary for enabling such functions.

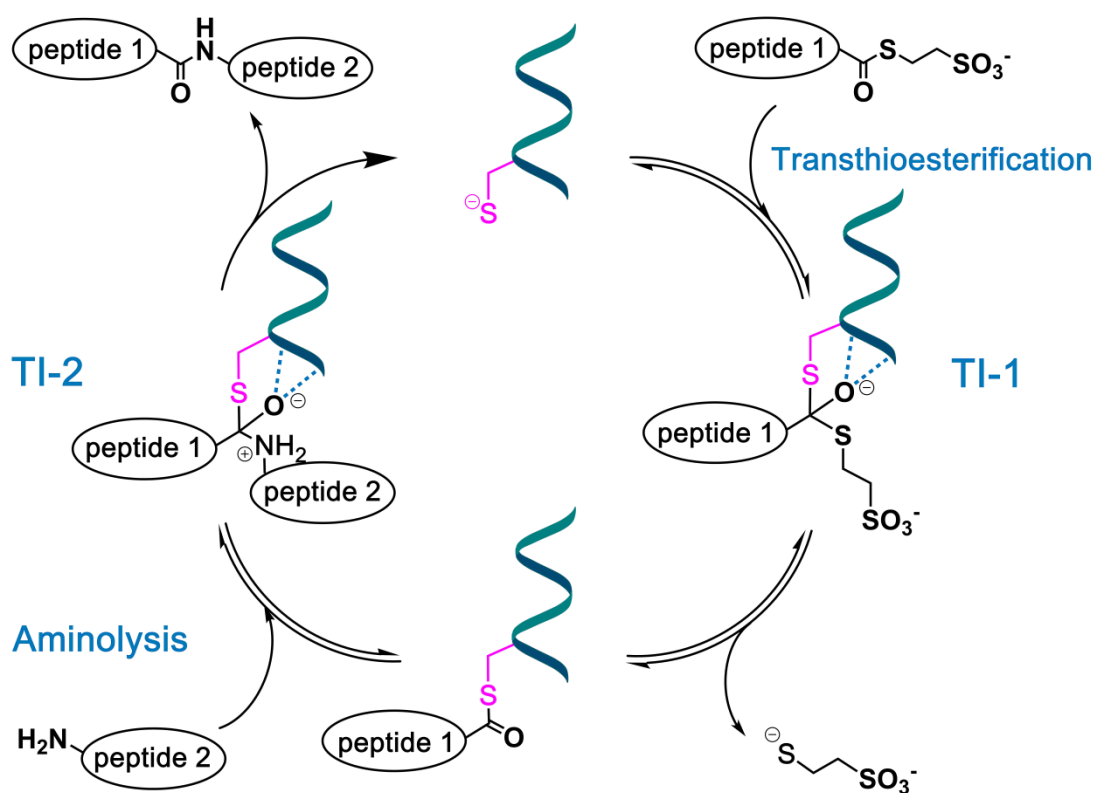
To design the active site, we used a minimalistic approach, which is based on chemical logic and intuition and implies the incorporation of minimal number of elements (such as amino acid residues or cofactors) into a *de novo* protein scaffold that is initially devoid of catalytic properties in order to convert it into catalytically active variant. Two recent examples of *de novo* catalytic proteins developed by a minimalistic and rational method include a thermostable hydrolase based on a 7-helix bundle and containing Cys-His-Glu triads as catalytic apparatus<sup>16</sup> and a promiscuous peroxidase that utilizes heme as a cofactor bound to a 4-helix bundle scaffold.<sup>17</sup> In our case, the minimal functional elements required for the ligation reaction are a principal nucleophilic residue (such as Cys or Ser) and an “oxyanion hole” arrangement of hydrogen-bond donors to stabilize the negative charge of the tetrahedral intermediates in this reaction.<sup>18</sup> These functional motifs are present in both natural ligases and reengineered proteases. We selected Cys as a key catalytic residue and not Ser for the following reasons: *i*) the thiol group of Cys has lower  $pK_a$  (8.5 in unstructured peptide)<sup>19</sup> which can potentially be used as a single residue instead of catalytic dyads or triads (*e.g.* classical Ser-His-Asp that promotes the deprotonation of hydroxyl in Ser), because the correct placement of two or three residues can already complicate the design; *ii*) the activation of peptide acyl donor in the form of peptide anchored to protein scaffold via Cys-thioester bond can be advantageous over the related Ser-oxoester (see **Scheme 2.1** for the putative mechanism), because the resulting thioester group benefits from an enhanced reactivity towards most of the nucleophiles due to the poorer orbital interactions between sulfur atom and carbonyl moiety.<sup>20</sup> Another advantage of Cys as a catalytic residue is the possibility of using synthetic peptide- $\alpha$ thioesters as acyl donor substrates, which can undergo a facile transthioesterification with Cys, thus forming the covalently attached peptide-protein species. Peptide- $\alpha$ thioesters are accessible by both Boc/benzyl and Fmoc/*t*Bu SPPS<sup>21</sup> and are commonly used in enzyme-free total chemical synthesis of proteins, especially via the NCL.<sup>22</sup>

The major aim of my thesis was to confirm the proof-of-concept for this approach; that is to engineer a *de novo* protein catalyzing the formation of an amide bond. In the second chapter of my thesis, the *de novo* protein scaffold selection and protein design will be first explained. The development of an efficient synthetic path was then explored followed by the biophysical and structural characterization of the designed analogues. Finally, a catalytic assay for the acyl transfer reaction was optimized to allow for the systematic catalytic characterization and the evaluation of reactivity scope.

## 2.2. Protein scaffold selection and protein design

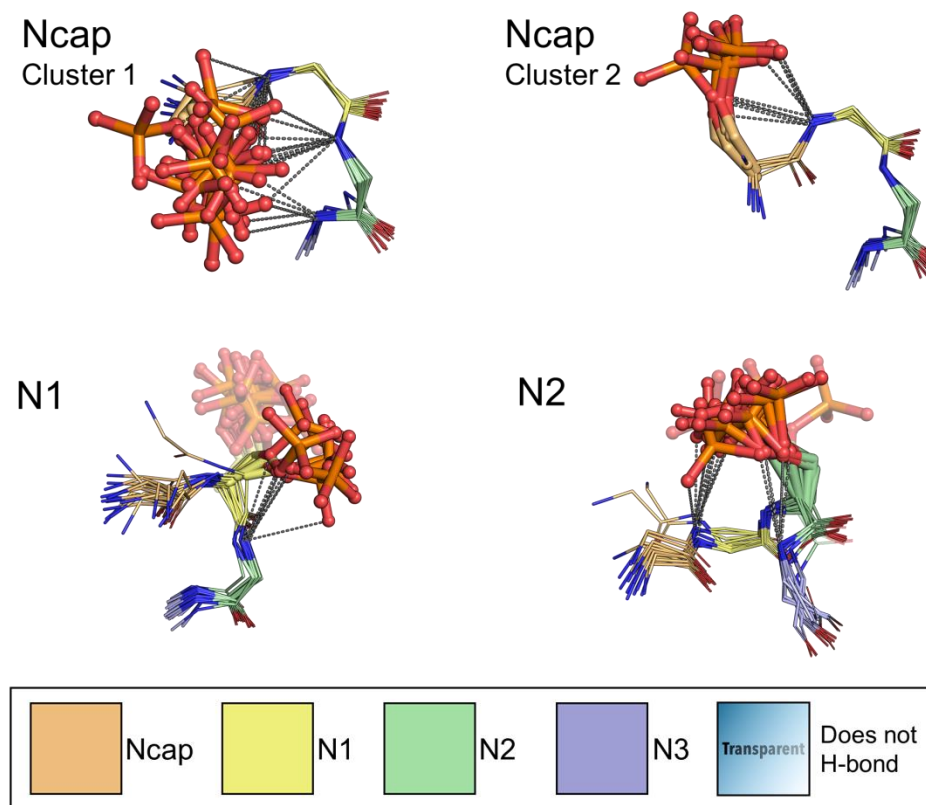
**Scheme 2.1** outlines the principal steps in the putative mechanism of catalysis of peptide ligation by a *de novo* protein. A peptide- $\alpha$ thioester would first react with the catalytic Cys and would undergo transthioesterification. After passing through the first tetrahedral intermediate (TI-1), the resulting branched peptide-protein thioester would be well oriented to react with H<sub>2</sub>N-peptide, the acyl acceptor substrate. The collapse of the second tetrahedral intermediate (TI-2) would lead to the release of the ligated product and free catalytic protein. For this double-displacement ping-pong mechanism to operate efficiently, both negatively charged intermediates of the two addition-elimination steps need to be stabilized, for example, through H-bonding with amides of protein scaffold serving as an “oxyanion hole”. Therefore, protein framework in a *de novo* protein must include nucleophilic Cys placed in such an environment.

Previously, it was demonstrated that phospho-serine (pSer) when introduced at N-terminus of an  $\alpha$ -helix can lead to its stabilization presumably due to the productive H-bonding of unpaired amides of  $\alpha$ -helical N-terminus to negatively charged phosphate group.<sup>23</sup>



**Scheme 2.1.** Putative mechanism for the catalysis of acyl transfer reaction (e.g. peptide ligation) by *de novo* catalytic protein (depicted schematically in cyan). TI-1 and TI-2 refer to the key tetrahedral intermediates of transthioesterification and aminolysis steps, respectively, where negatively charged oxyanions are stabilized by H-bonding network (shown in blue).

This stabilizing effect was most prominent when pSer was placed at N2 (second residue at N-terminus of  $\alpha$ -helix) position. By analyzing the structural data from PDB in collaboration with W. F. DeGrado and coworkers, we provided support to these experimental results: clustering and alignment of crystal structures containing pSer at Ncap (residue preceding the first amino acid that adopts  $\alpha$ -helical conformation), N1 and N2 (first and second positions at N-terminus of  $\alpha$ -helix) illustrate that backbone amides at  $\alpha$ -helical N-termini indeed can contribute 1-to-2 H-bonds to one of the distal oxygens of the phospho-group (**Figure 2.2**). The tetrahedral intermediates of addition-elimination steps of acyl-transfer catalytic mechanism shown in **Scheme 2.1** are negatively charged and bear similarities in structure to pSer as demonstrated by molecular modeling shown in **Figure 2.3**, therefore, the N-terminus of  $\alpha$ -helix represents a suitable structural motif for stabilization of the respective tetrahedral intermediates.



**Figure 2.2. Structural alignment of pSer-containing fragments from structures deposited in PDB (accessed in February 2020) with pSer located at Ncap, N1 and N2 positions of the corresponding  $\alpha$ -helical fragments.** pSer at Ncap positions form 2 clusters, due to a bimodal distribution of its psi dihedrals. In both clusters, all of the pSers form H-bonds with downstream amides. However, only 7 out of 23 pSer's in the N1 position hydrogen bond to the helix backbone. The geometry of N2 pSer's appears as stabilizing as Ncap pSer's; 13 out of 16 fragments participate in H-bonds with its preceding and/or succeeding amide. pSer at the Ncap and N2 positions are observed contributing bivalent H-bonds, whereas only 1 H-bond is observed if pSer is in the N1 position. Example hydrogen bonds are represented by dashed lines, and phosphate groups that do not H-bond with helix backbone are transparent.



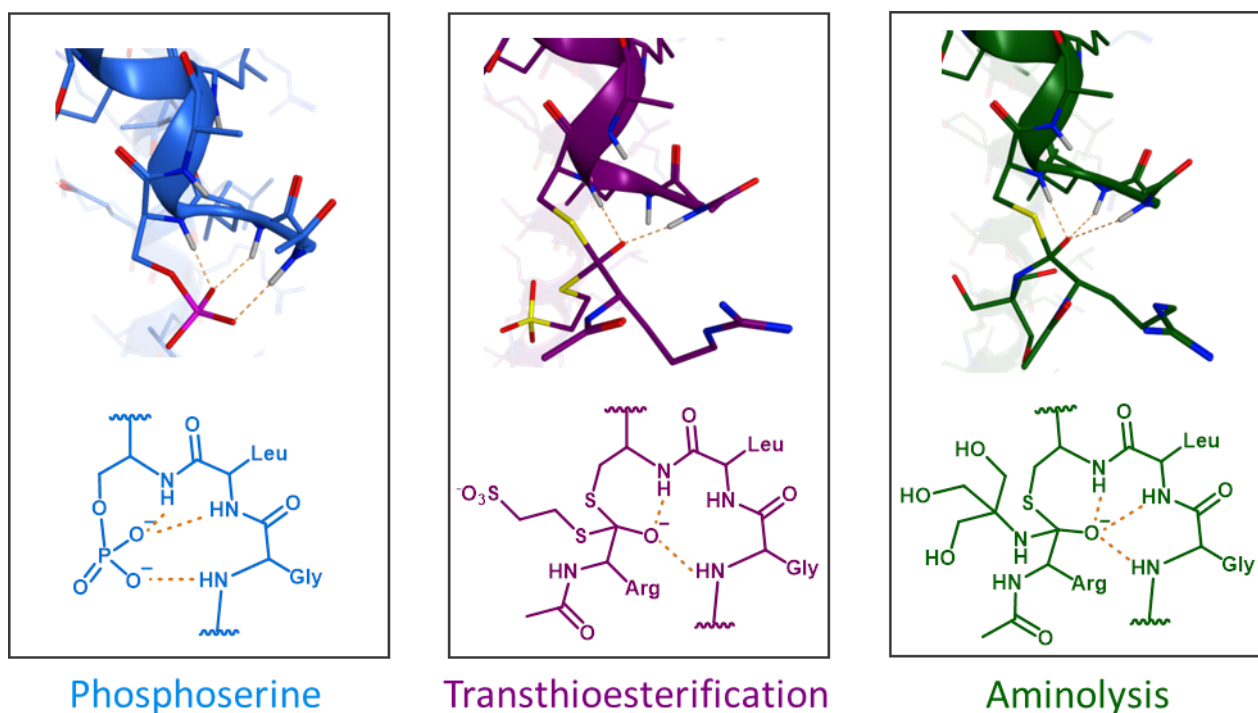


Figure 2.3. Computational modeling of phosphoserine and tetrahedral intermediates of both transthoesterification and aminolysis steps with Cys introduced at N2 position of the DSD protein (PDB id: 1g6u). Stabilizing hydrogen bonds (orange dotted lines) between the three oxyanions and unpaired N-H groups of amide at the N-terminus were detected. Aminolysis was modelled using tris(hydroxymethyl)aminomethane (Tris) as standard acyl acceptor. MOE software was used to construct the phosphoserine and the two tetrahedral intermediates and minimization using Amber 10 force field was performed.

To test the feasibility of such catalytic strategy we have chosen the *de novo* designed domain-swapped dimer (DSD) protein that is composed of two identical sequences that spontaneously dimerize to form a three- $\alpha$ -helical bundle.<sup>24</sup> In this scaffold, the two adjacent  $\alpha$ -helical N-termini are facing each other creating a cavity that represents a proper location to incorporate an active site (see Figure 2.4). The environment at the juxtaposition of two N-termini is in the proximity of the hydrophobic interior of the helical bundle and, at the same time, solvent exposed, therefore, this part of the protein can permit robust access of various peptide substrates and allow for catalysis without introducing any steric restrictions in substrate structures. We also supposed that catalytic cysteine may benefit from helix macrodipole effect at N-terminus positions in order to depress the  $pK_a$  of the nucleophile residue.

The Arg residues were used to substitute all Lys in the original DSD scaffold to prevent unwanted acylation of Lys side-chains and Gly or Asn were generally employed at the N-cap position instead of Ser in the DSD protein. A homodimer **Hom-N2**

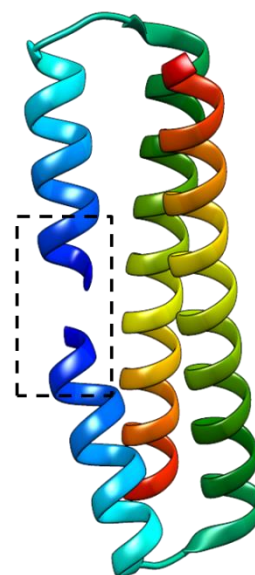
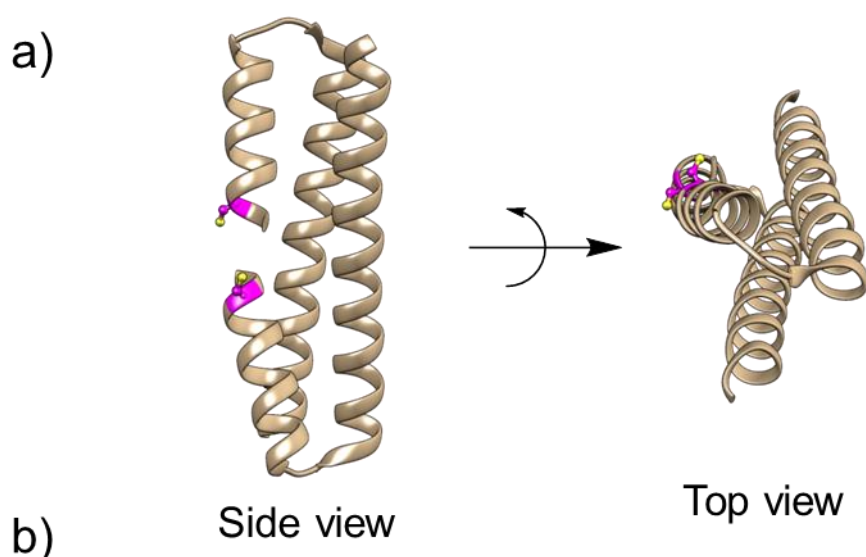


Figure 2.4. DSD structure. The homodimer is represented with rainbow color scheme from blue to red in the N-terminal-to-C-terminal direction. The black dashed box depicts the localization of the potential active site in the cavity created by the two N-termini.

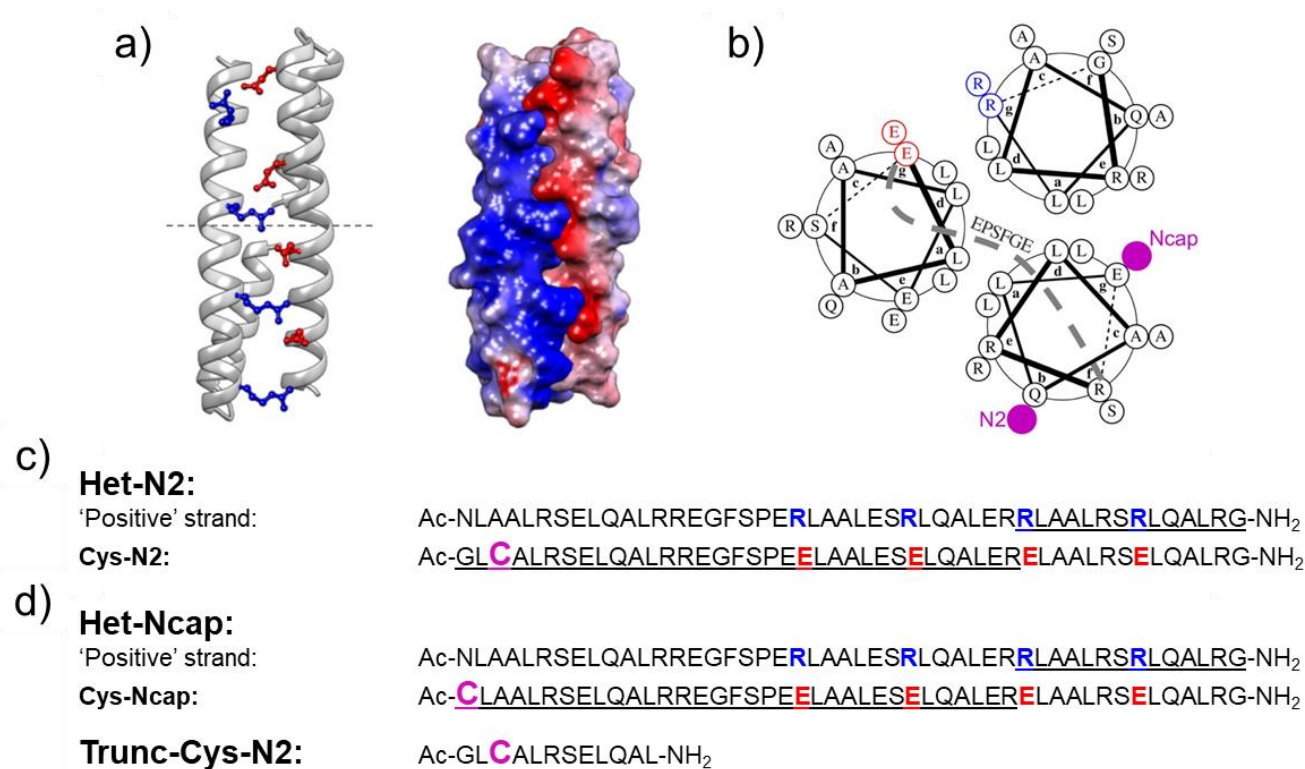
was first designed with two cysteine residues incorporated at the N2 positions of the two monomers, based on the most efficient stabilizing effects with pSer-containing peptides (for computational model and sequence see **Figure 2.5**).<sup>23</sup> Charge patterning of the two segments of DSD scaffold was used to redesign homodimer into heterodimer in order to distinguish the two N-termini for their further functionalization. Thus, one strand was patterned with negatively charged residues (Glu) ('negative' strand), while another one with positively charged (Arg) ('positive' strand) (**Figure 2.6a-b**). Subsequently, catalytic Cys residue was engineered into the 'negative' strand (containing excess of Glu) at N2 position resulting in the **Cys-N2** peptide fragment. The sequences of the corresponding peptide fragments are depicted in **Figure 2.6c**. The heterodimer resulting from complexation of **Cys-N2** with 'positive' strand is abbreviated as **Het-N2**. Concurrently, two control scaffolds were designed in order to test the importance of the N-terminal position of the catalytic cysteine and to evaluate the relevance of the entire structure for the catalytic activity. First, a second heterodimer (named as **Het-Ncap**) was engineered where the 'negative' strand (abbreviated as **Cys-Ncap**) contained the Cys residue at Ncap position and the 'positive' strand remained the same as for **Het-N2** (sequences are shown in **Figure 2.6d**). Second, a truncated variant of **Cys-N2** strand was considered corresponding to the first twelve N-terminal residues and is abbreviated as **trunc-Cys-N2** (see **Figure 2.6d**).



### Hom-N2:

Ac-GL**C**ALRSELQALRREGFSPEELAALESELQALERRLAALRSRLQALRG-NH<sub>2</sub>

**Figure 2.5. Homodimer design.** a) Computational modeling of the two catalytic Cys residues (in magenta) at N2 position of each monomer. b) Sequence of the monomer allowing spontaneous dimerization.



**Figure 2.6. Design of heterodimers.** a) Left: charged residue patterning by grafting negative charges (Glu in red) on the strand containing the Cys catalytic residue and positive charges (Arg in blue) onto the other helix. Right: resulting cumbic surface is shown. b) Helical wheel diagram of the heterodimeric protein that corresponds to the half-structure depicted above the dashed line in a). Leucine residues constitute the hydrophobic core (a and d positions). Glu and Arg residues at e and g positions promote heterodimerization. The DrawCoil 1.0 software (<https://grigoryanlab.org/drawcoil/>) was used to arrange the wheel diagram using the same color code as in a). c) Sequences of the two monomers of **Het-N2**. d) Sequences of the two control constructs with Cys at Ncap position in **Het-Ncap** and a 12-residue N-terminal fragment **trunc-Cys-N2**. The sequence fragments of **Het-N2** and **Het-Ncap** that are depicted on the wheel diagram in b) are underlined.

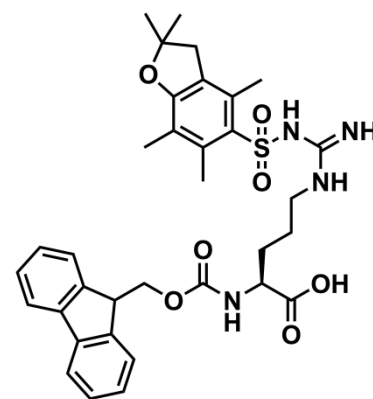
## 2.3. Total chemical synthesis of the analogues

### 2.3.1. Sequential Fmoc-SPPS

Previously, the original DSD protein was synthesized by the sequential chain assembly of 48 residues via machine-assisted Fmoc/*t*Bu-SPPS<sup>25</sup> on PAL (peptide amide linker) resin using either pentafluorophenyl activated ester or free acids activated *in situ* by HBTU/HOBt/DIEA as coupling methods.<sup>24,26</sup> In our hands, using different protocols (Rink amide resin and HATU coupling reagent), the sequential assembly of the 48-residue DSD protein using machine-assisted Fmoc/*t*Bu-SPPS resulted in too many by-products in the crude material. Although the desired peptide was detected, this strategy was abandoned due to anticipated poor yields after purification. Later, with the arrival of a microwave synthesizer

in the laboratory, the full strand Fmoc-SPPS was tried again with the sequential assembly of the 48 residues that composed the 'positive' strand of heterodimers. Even with microwave assistance, the synthesis failed due to the accumulation of too many by-products in the final crude mixture. This time, all the original lysine residues of the original DSD protein were modified by arginine amino acids. Arginines are generally harder to couple due to the bulkiness of the side chain and of the corresponding protecting group (see **Figure 2.7** for the Fmoc-L-Arg(Pbf)-OH building block structure). Thus, we mainly attributed problems of full-length strand synthesis by Fmoc-SPPS to the recurrent presence of arginine residues in the heptad repeat within the helix. Playing with time of coupling and the number of couplings (double or even triple coupling for key residues) did not improve much the global efficiency of the Fmoc-SPPS. Sequential addition of the full-length 48 residues by Fmoc-SPPS was thus abandoned to favor convergent approaches and Boc-SPPS.

Nevertheless, automated microwave Fmoc-SPPS was employed to synthesize the 12-residue **trunc-Cys-N2** control peptide in a good yield of 32 %.

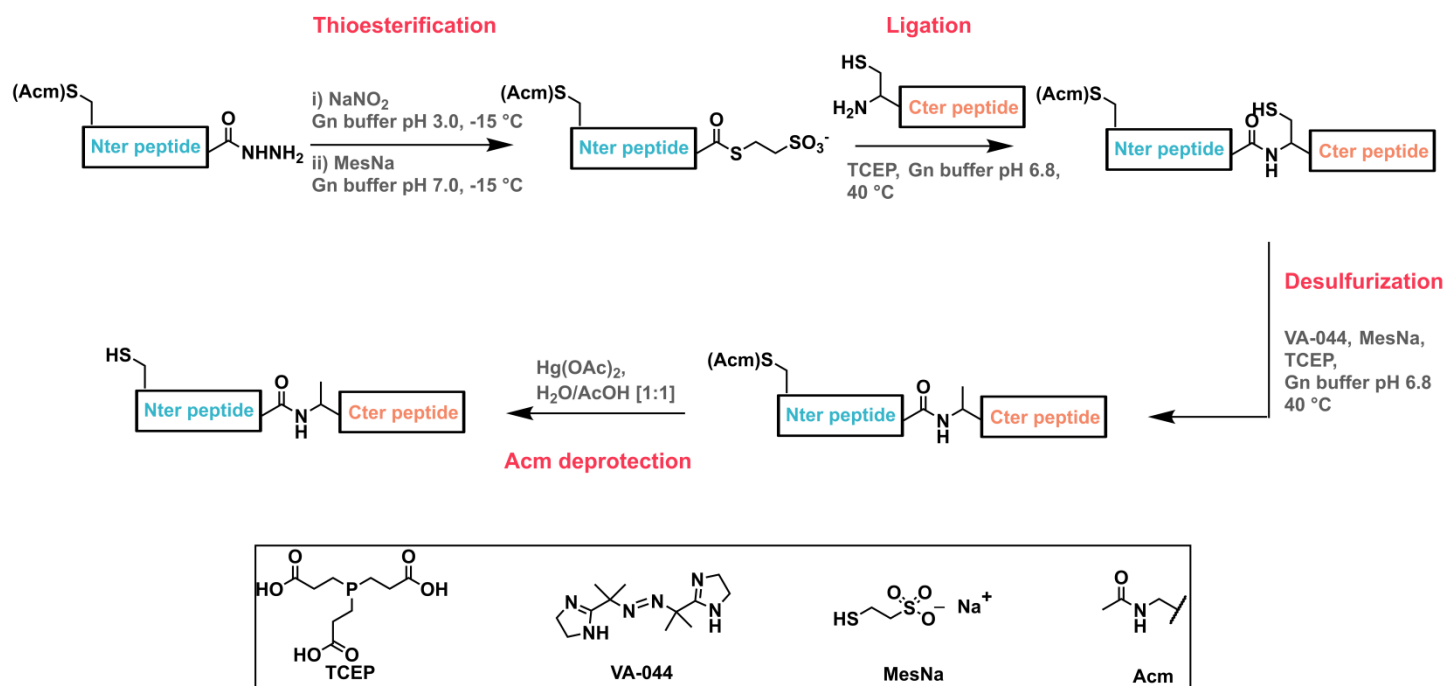


**Figure 2.7. Structure of the Fmoc-L-Arg(Pbf)-OH building block.** Pbf stands for 2,2,4,6,7-pentamethyl-2H-benzofuran-5-sulfonyl.

### 2.3.2. Fmoc-SPPS and two-segment ligation

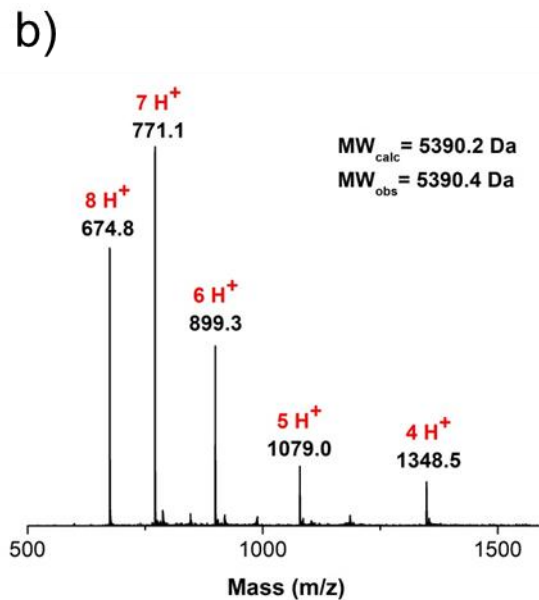
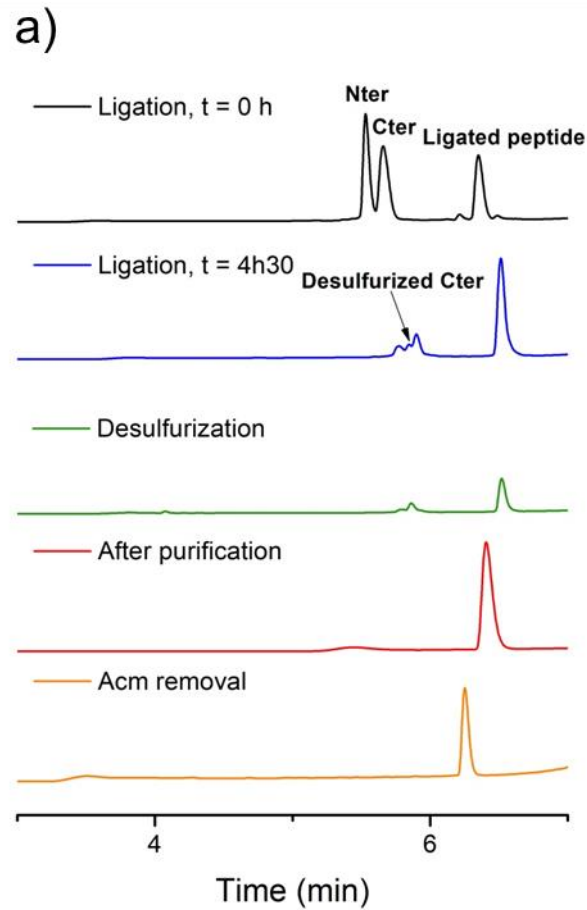
A synthetic strategy was then elaborated for the production of the homodimer **Hom-N2** analogue based on the native chemical ligation (NCL)<sup>22</sup> of two peptide fragments followed by desulfurization and in which the peptides were prepared by Fmoc/*t*Bu-based SPPS (**Scheme 2.2**). We decided to split the **Hom-N2** sequence between the 23<sup>rd</sup> and 24<sup>th</sup> residues. This Ala-Ala junction was convenient because: *i*) it splits the sequence into two parts of almost equal length; *ii*) peptide-Ala- $\alpha$ thioester is relatively easy to obtain and it is reactive in NCL;<sup>27</sup> *iii*) the other Ala can be provided via desulfurization of the cysteine at the ligation junction.<sup>28</sup> The N-terminal peptide segment contains the catalytic Cys residue that needs to remain intact at the end of the synthetic process. This particular cysteine was protected by an acetamidomethyl (Acm) group that is stable under the cleavage conditions (95 % of TFA) and desulfurization.<sup>4,29</sup>

The N-terminal peptide- $\alpha$ hydrazide was prepared with high purity by Fmoc-SPPS on 2-chloro-trityl resin prefunctionalized with a hydrazine group (2-Cl-Trt-NHNH<sub>2</sub>) and was



**Scheme 2.2. Two-segment ligation.** The N-terminal peptide- $\alpha$ hydrazide was first thioesterified to then undergo condensation with the C-terminal peptide via NCL. Desulfurization of the cysteine was performed directly after ligation in a one-pot fashion. After isolation of the ligated product, the catalytic cysteine was deprotected.

directly introduced in thioesterification reaction without any purification. To do so, the peptide- $\alpha$ hydrazide was first activated through oxidation to the corresponding peptide- $\alpha$ azide in the presence of  $\text{NaNO}_2$  in guanidine hydrochloride (Gn-HCl) buffer at pH 3.0. Upon completion of the reaction, sodium-mercaptoethanesulfonate (MesNa), solubilized in a neutral pH buffer, was directly added to the reaction mixture. The thiolysis took a few minutes judging by HPLC (shift of the retention time) and LC/MS and gave a satisfactory isolated yield (22 % from crude). The C-terminal Cys-peptide fragment was synthesized by machine-assisted Fmoc-SPPS with the isolated yield after HPLC purification of 6 %. Ligation between purified peptide- $\alpha$ thioester and peptide C-terminal was then performed (**Scheme 2.2**). The progress of the reaction was monitored by HPLC and LC/MS (**Figure 2.8**). After one hour, desulfurization of both starting cysteine-containing peptides and ligated products was observed. The reducing agent tris(2-carboxyethyl)phosphine (TCEP), which is added to the reaction mixture to prevent oxidation, may initiate a radical reaction that cannot be quenched in a deoxygenated buffer.<sup>30</sup> The desulfurization of the starting material slowed down the ligation because the cysteine-containing starting material got depleted. After five hours, only an insignificant decrease of the reactants was observed by HPLC. Therefore, desulfurization was subsequently performed in the presence of a radical initiator (VA-044) in a reducing environment (TCEP at high concentration, 200 mM) to obtain the original Ala. In one hour, the reaction was completed according to the LC/MS with a quite good yield of 44 % after purification. As was reported previously,<sup>4</sup> the desulfurization step did not affect Acm-



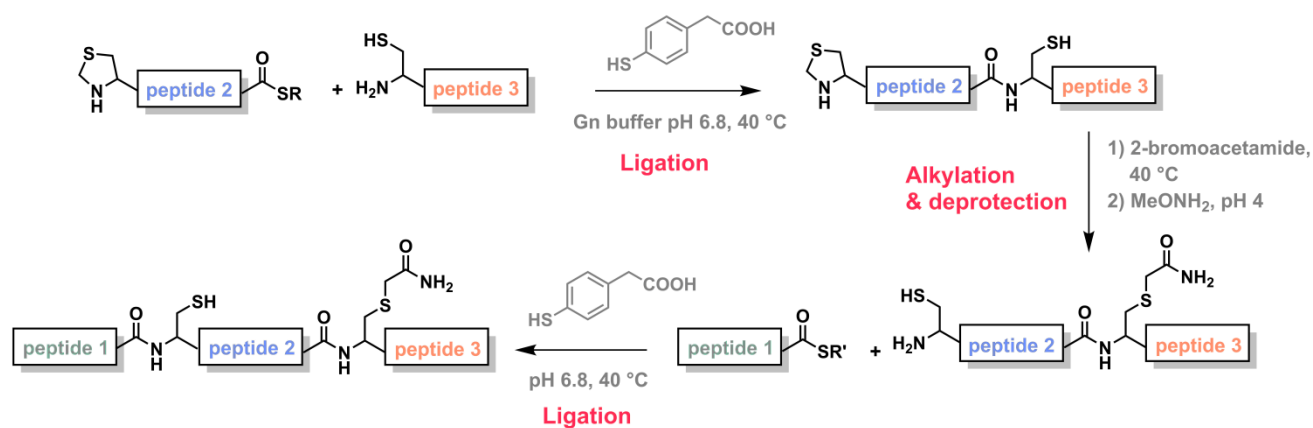
**Figure 2.8. Two-segment ligation of the protein Hom-N2.** a) HPLC monitoring of the ligation between the N-terminal and C-terminal peptide fragments and the subsequent Cys desulfurization to Ala. After purification, the acetamidomethyl (Acm) protecting group was removed. b) LC/MS spectrum of the purified **Hom-N2**.

protected cysteine. Deprotection of the Ac<sub>m</sub>-group was performed in the presence of a Hg(OAc)<sub>2</sub> reagent.<sup>31</sup> During work-up, the reaction mixture was treated with dithiothreitol (DTT), which was added as a chelating and reducing agent, and the pH was adjusted from ~2 to 8.5 to deprotonate thiol groups of DTT and to activate it to coordinate to Hg<sup>2+</sup>. The purification by HPLC furnished the desired full-length **Hom-N2** with a yield of 25 %. Surprisingly, this quite low yield was not necessarily expected due to the straightforward reaction according to HPLC and LC/MS analysis. We presumed that the loss of product may come from a reduced binding to the preparative column due to the presence of acetic acid at high concentration (solvent of reaction was half water, half acetic acid). Alternative solution may have been to evaporate the high amount of acetic acid before purification.

Nevertheless, this synthetic approach based on Fmoc-SPPS of peptide segment followed by ligation/desulfurization and Ac<sub>m</sub> deprotection was not retained. The yield of Fmoc-SPPS even by splitting in two the length of produced peptide was still low, especially for the C-terminal Cys-peptide fragment. Some Ac<sub>m</sub> deprotection was also observed after cleavage of the N-terminal peptide. In addition, the multiple steps of purification decreased the global yield of the synthesis. Other strategies were therefore exploited with the major aims to lower the number of purification steps and to optimize time of syntheses.

### 2.3.3. *One-pot three-segment ligation of peptides synthesized by Boc-SPPS*

This method consists in a one-pot three-segment ligation, which means that sequential NCL were performed without intermediate purification (see **Scheme 2.3** for the synthetic pathway). The three peptide fragments were prepared by '*in situ* neutralization' Boc/benzyl-SPPS.<sup>32</sup> Boc-SPPS can be advantageous for 'difficult' sequences to synthesize as trifluoroacetic acid (TFA) used to deprotect the N<sup>α</sup>H<sub>2</sub> after each coupling has the potency to temporarily disintegrate potential peptide-resin aggregates that are responsible for incomplete couplings and/or deprotection.<sup>33</sup> Time of couplings and deprotections are also reduced compared to Fmoc-SPPS (10 to 15 min and 2 min against 30 to 40 min and 20 min, respectively) but Boc-SPPS is not accessible by an automated machine in our laboratory. Nevertheless, manual peptide synthesis allows a fine-tuning of each step and especially of each coupling by following the completeness of reactions with colorimetric test such as the Kaiser test, also called the ninhydrin test.<sup>34</sup> Moreover, peptide-<sup>α</sup>thioesters are directly and easily available by Boc-SPPS as thioester linkages are resistant to the synthetic conditions (*i.e.* acidic N<sup>α</sup>H<sub>2</sub> deprotection and acidic cleavage), whereas it is not the case during Fmoc-



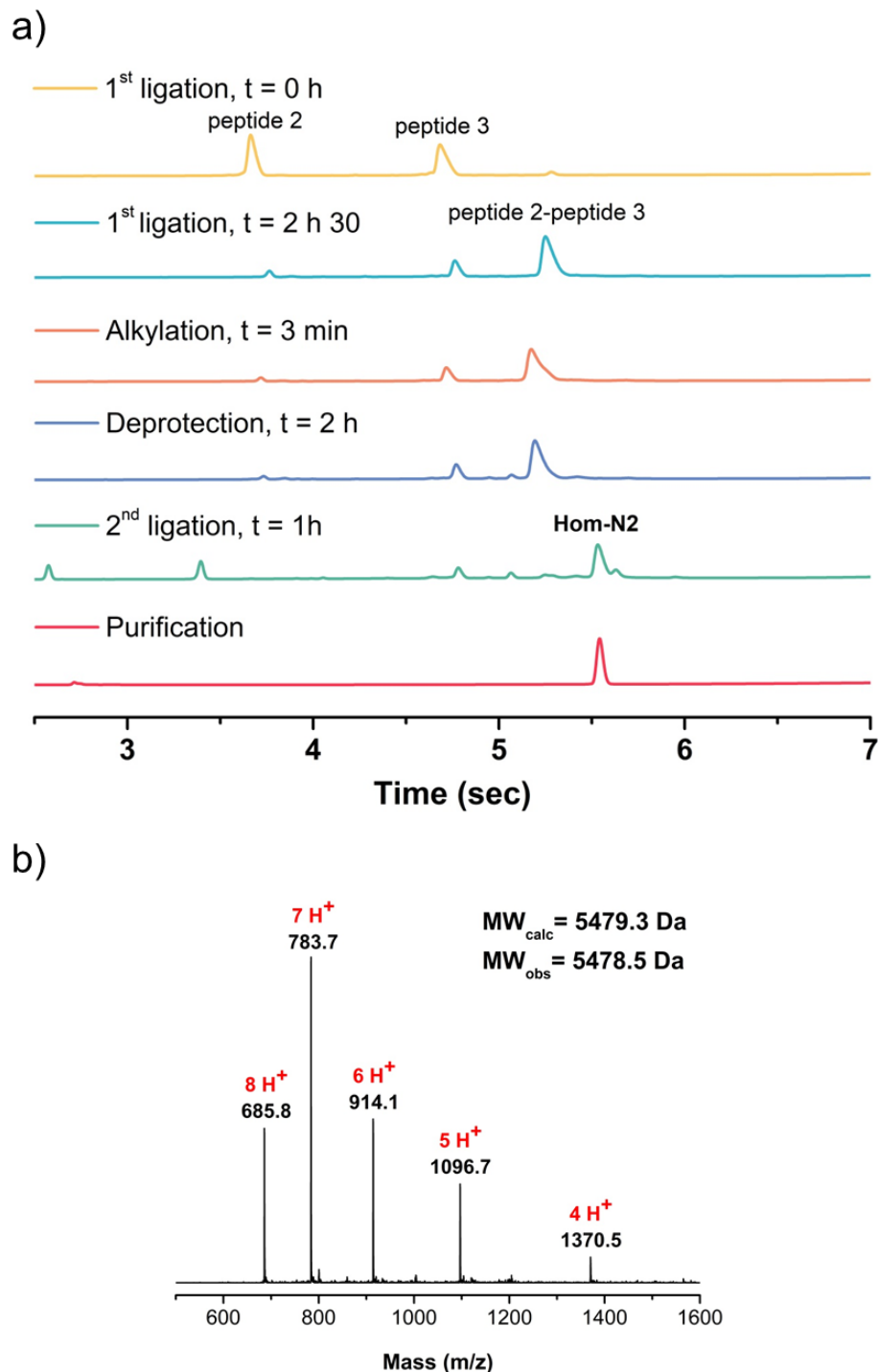
**Scheme 2.3. One-pot three-segment ligation.** Native Chemical Ligation (NCL) was performed between peptide 2 and 3 in presence of MPAA as catalyst. The remaining cysteine was capped by alkylation. Subsequently, the thiazolidine protecting group was removed and the resulting free cysteine participated to the second ligation with peptide 1 giving the final **Hom-N2** protein.

SPPS notably due to the amine deprotection at basic conditions in which an extra thioesterification step is subsequently needed.

This synthetic approach was validated with the synthesis of the **Hom-N2** protein. The strand was split in three peptide fragments. The junction site between the middle and C-terminal fragments (peptide 2 and peptide 3 in the **Scheme 2.3**) was located between the 23<sup>rd</sup> and 24<sup>th</sup> residues exactly like in the previously described method. The second junction site of the N-terminal peptide and the middle one (peptide 1 and peptide 2 in the **Scheme 2.3**) was established between the 2<sup>nd</sup> and 3<sup>rd</sup> residues meaning just before the catalytic cysteine, thus this residue participates to the second ligation. The C-terminal peptide (peptide 3) is identical to previous method, however Boc-SPPS provided better yield than Fmoc-SPPS (14 % against 6 %, respectively). The peptide 2-<sup>o</sup>thioester was synthesized with a final yield after purification of 16 % by first coupling a linker composed of 3-mercaptopropionic acid (MPA) and two arginine residues that correspond to the thioester moiety. This middle fragment contained also an N-terminal thiazolidine protecting group (corresponding to the protected catalytic cysteine). The progress of the 3-segment ligation reaction was monitored by HPLC and LC/MS (**Figure 2.9**). The middle peptide segment was first ligated with the C-terminal fragment using NCL in presence of 4-mercaptophenylacetic acid (MPAA) to catalyze the reaction via transthioesterification with peptide 2-<sup>o</sup>thioester resulting in a more reactive aryl thioester. The resulting cysteine was capped by alkylation with 2-bromoacetamide. The excess of alkylating reagent was then quenched with a large excess of MPAA followed by the deprotection of thiazolidine group to give the catalytic free cysteine. This resulting cysteine residue allowed the second ligation to attach Ncap and N1 residues corresponding to the peptide 1 previously synthesized by Boc-SPPS to directly obtain a thioester with a linker (*i.e.* composed of Ala, Arg and Tyr residues) for better



solubility and purification. The global yield of this one-pot three-segment ligation was 30 % after purification.



**Figure 2.9. Three-segment one-pot ligation of the protein Hom-N2.** a) HPLC monitoring of the ligation between the peptide 2<sup>-α</sup>thioester and peptide 3 followed by alkylation of the remaining Cys residue. The Thz protecting group was removed and the resulting Cys was utilized for the second ligation with peptide 1<sup>-α</sup>thioester. Ligated Hom-N2 was isolated by HPLC. b) LC/MS spectrum of the purified Hom-N2

This method to synthesize full-length strand is very convenient. Syntheses of the peptide fragments were performed by Boc-SPPS in large scale and were then assembled in a one-pot fashion to avoid multiple steps of purifications. Although this method was first used to synthesize only the **Hom-N2** protein, it would also allow achieving combinatorial synthesis of multiple analogues via a convergent approach to ligate different types of the three peptides.

#### 2.3.4. Sequential Boc-SPPS

Due to the success of Boc-SPPS to synthesize peptide fragments that constitute an entire 48-residue strand, this technique was used to sequentially assemble the 48-residue monomers of the **Het-N2** and **Het-Ncap** proteins, meaning the 'positive', **Cys-N2** and **Cys-Ncap** strands. Manual syntheses was performed on up to 0.4 mmol scale and single-step purification led to the desired peptide with good purity and global yield around 2 %. Analysis of pure peptides is shown in **Figure 2.10**. Even if the yields were rather low, this method allows synthesizing large quantity of pure peptides in a minimum of time and effort.

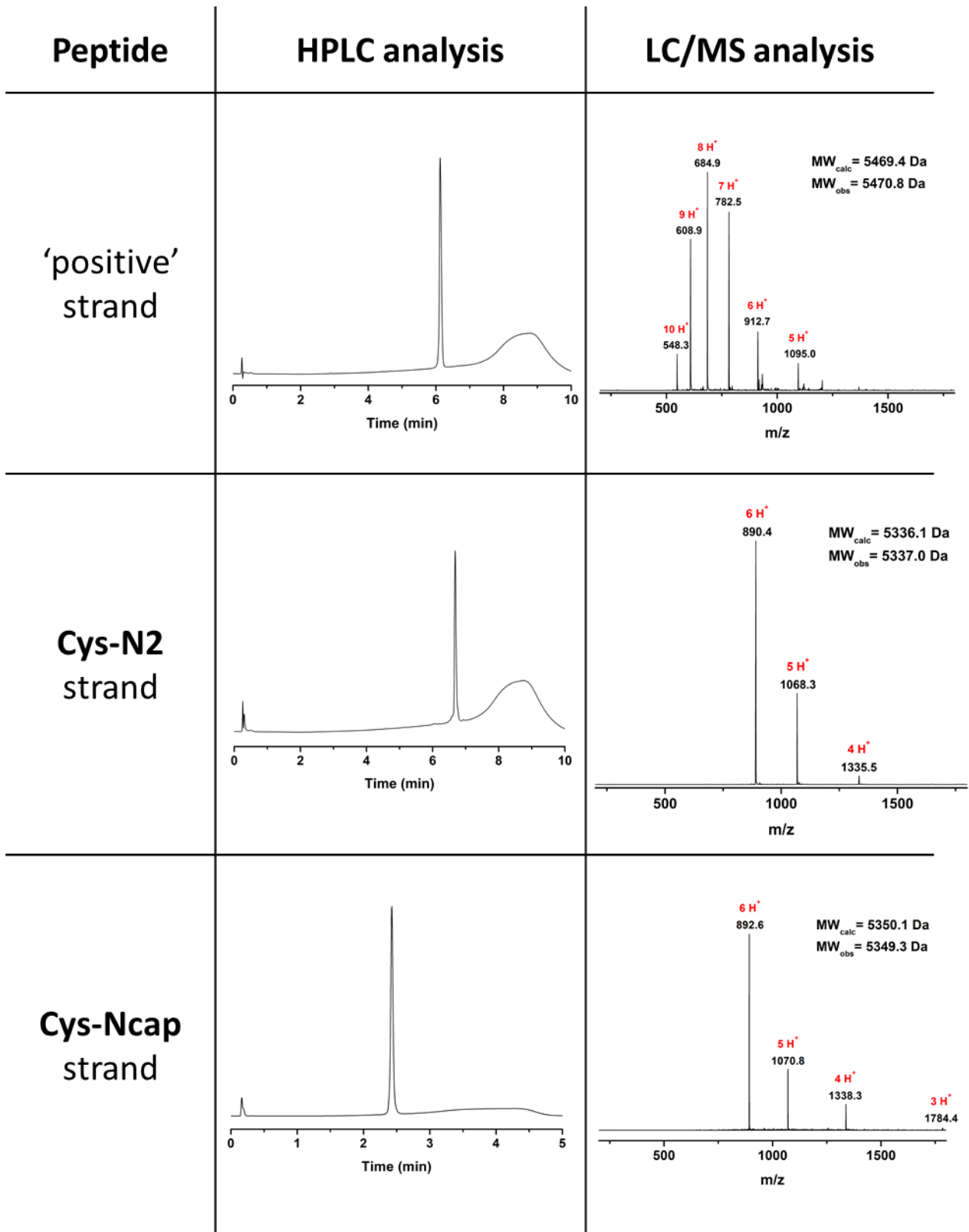
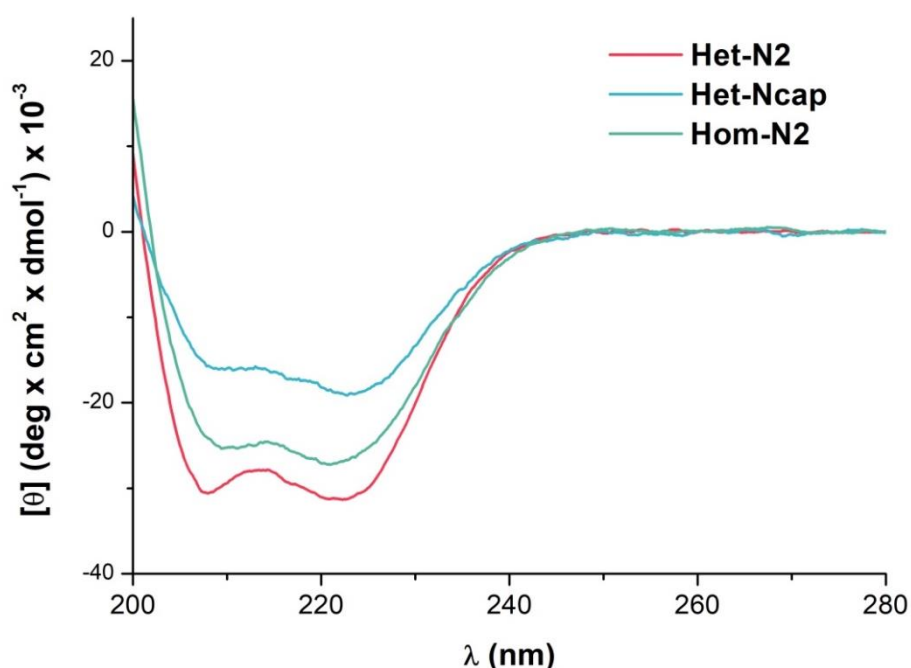


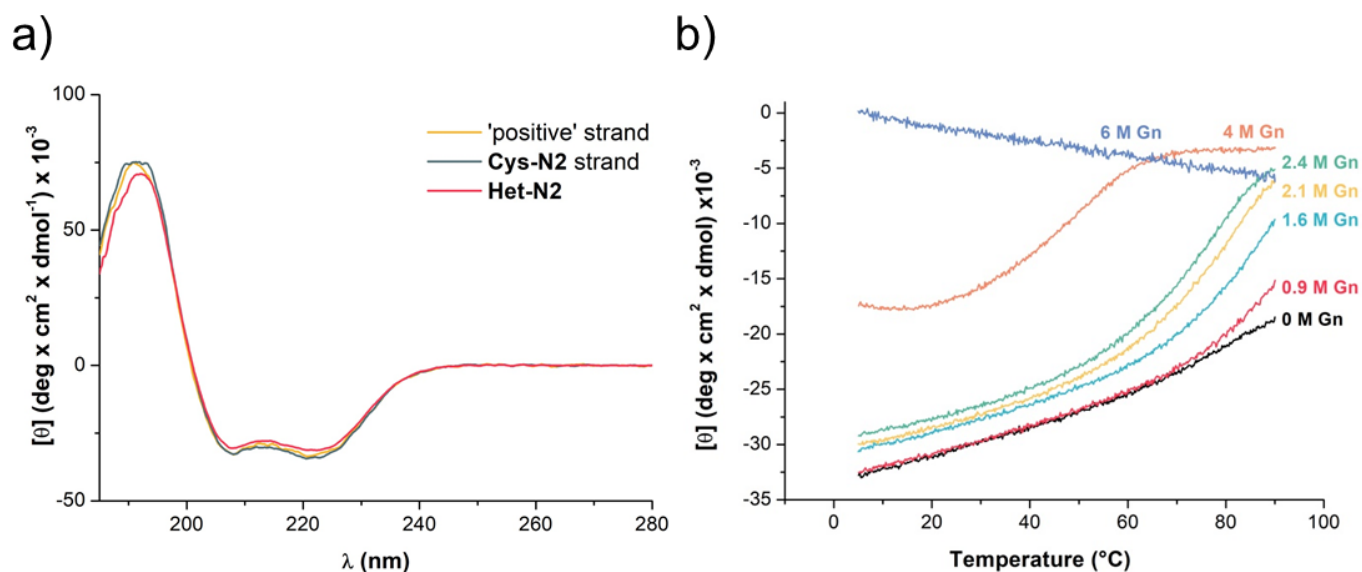
Figure 2.10. HPLC and LC/MS analysis of the different peptides that constitute the Het-N2 and Het-Ncap heterodimers.

## 2.4. Biophysical characterization

The assembly and folding of constructs **Het-N2**, **Het-Ncap** and **Hom-N2** proceeded spontaneously by dissolving the corresponding ‘positive’ and ‘negative’ strands in a 1:1 ratio for the heterodimers and only the monomer for the homodimer in phosphate buffer at near neutral pH. Anticipated  $\alpha$ -helical structure was confirmed by circular dichroism (CD) spectroscopy for the three proteins (see **Figure 2.11**). However, it is important to say that intensity of the ellipticity cannot be taken into account as the exact concentration was only determined for the **Het-N2** protein leading to a more precise interpretation of the measurements. Further biophysical analyses were mainly performed on **Het-N2** protein, which is the analogue we expect to carry the highest catalytic activity. Interestingly, the two strands of **Het-N2** taken separately were also bearing strong  $\alpha$ -helix signal as shown in **Figure 2.12a**. The stability of folding of heterodimer **Het-N2** against thermal and chemical (increased concentration of guanidine hydrochloride, Gn·HCl) denaturation was then evaluated by measuring intensity of ellipticity at 222 nm [ $\theta_{222 \text{ nm}}$ ], a strong indicator of  $\alpha$ -helicity. In native conditions, **Het-N2** stayed substantially folded from 5 °C to 90 °C (**Figure 2.12b**). Up to 2 M Gn·HCl in the buffer, **Het-N2** retained a high helical content at room temperature. Increasing the concentration of denaturant led to lowering of an apparent melting temperature and at 6 M Gn·HCl the **Het-N2** was found to be unfolded.

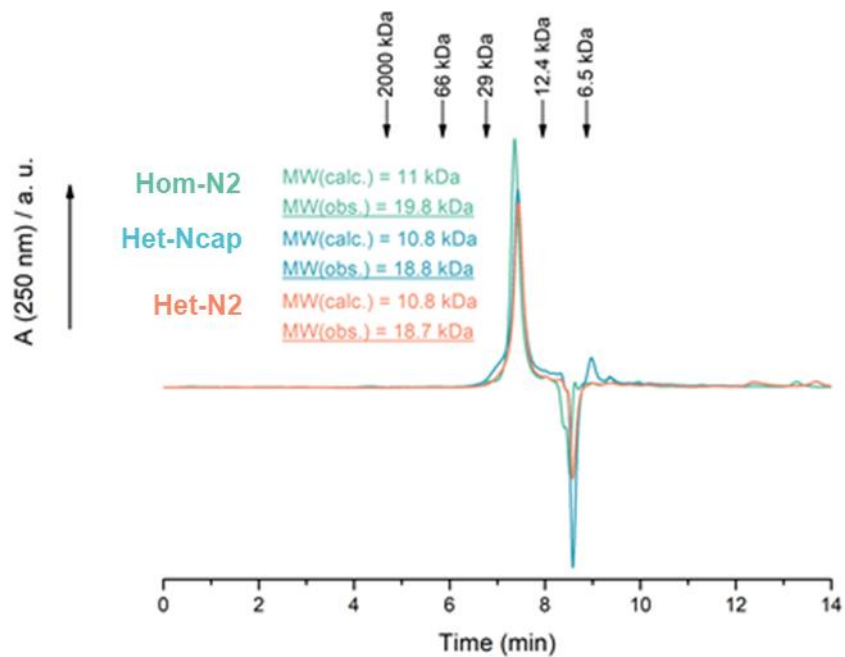


**Figure 2.11.** CD spectra of the different DSD-analogues. **Het-N2**, **Het-Ncap** and **Hom-N2** showed typical CD spectra of  $\alpha$ -helix.

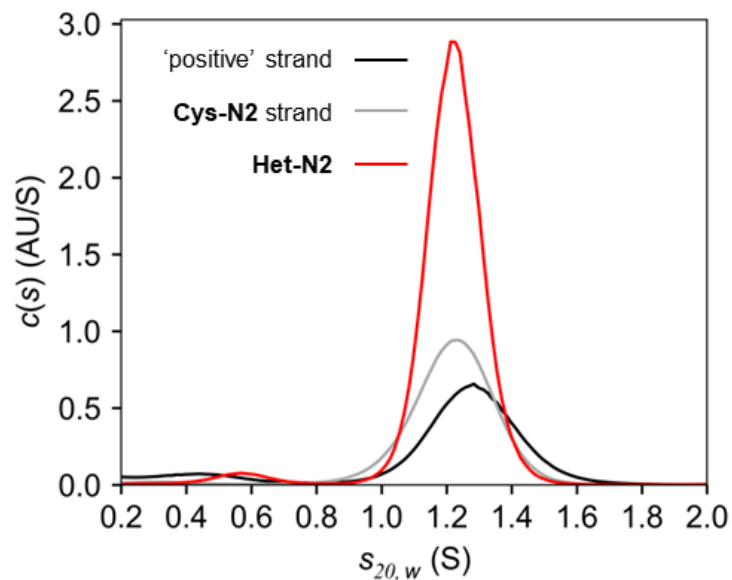


**Figure 2.12. CD measurements of Het-N2 protein.** a) CD spectra for the catalytic protein **Het-N2** and the respective isolated 'positive' and **Cys-N2** strands. All showed high  $\alpha$ -helicity content. b) Melting curves measured by CD at 222 nm for the catalytic protein **Het-N2** at different concentrations of guanidine hydrochloride (Gn-HCl).

The next step was to estimate the oligomerization state of the proteins and to verify if the dimer is well formed in solution. Moreover, single monomeric strand also demonstrated high helicity content (Figure 2.12a), which may be due to self-assembly to form homodimer that was not expected. Therefore, we also needed to validate that the heterodimerization was favored compared to the homodimerization of the two strands taken separately. First attempt to estimate the oligomerization state of **Het-N2**, **Het-Ncap** and **Hom-N2** analogues was performed by size-exclusion chromatography (SEC). According to the elution profiles, the dimers were characterized as monodisperse (sharp elution peak, see Figure 2.13). The difference between observed molecular weights (18.7-19.8 kDa) and calculated ones (10.8-11 kDa) may come from a higher oligomerization state in solution or may correspond to the dimeric species with an overestimation of the weight due to the elongated shape of the *de novo* proteins compared to more globular topologies of the references. The dimeric composition of **Het-N2** was better confirmed by analytical ultracentrifugation (AUC), as shown in Figure 2.14. However, it was not clear if the dimer detected for **Het-N2** sample by sedimentation velocity experiment corresponded to heterodimer or a mixture of heterodimer and the corresponding homodimers, because the individual 'positive' and **Cys-N2** strands gave similar sedimentation profiles at 200  $\mu$ M. Ideally, measurements of AUC at lower concentrations of proteins would have been necessary to assess dissociation constant of heterodimer and the potential homodimers.



**Figure 2.13. Size-exclusion chromatography of Het-N2, Het-Ncap and Hom-N2 in phosphate buffer, pH 7.5.** The references are composed of Blue Dextran (for the determination of the void volume, 2000 kDa), aprotinin from bovine lung (6.5 kDa), cytochrome c from equine heart (12.4 kDa), carbonic anhydrase from bovine erythrocytes (29 kDa) and albumin from bovine serum (66 kDa).



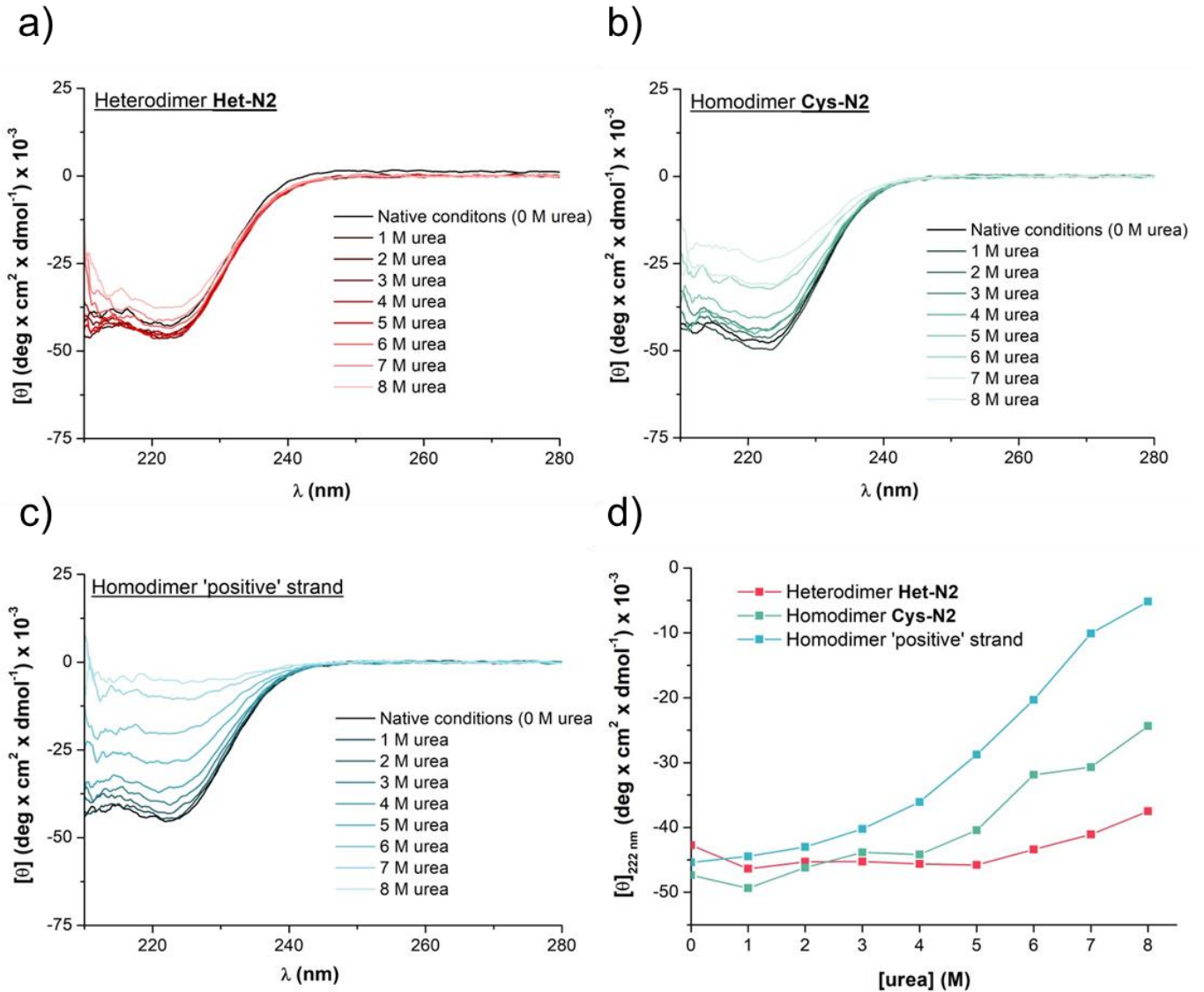
**Figure 2.14. Sedimentation profile of Het-N2 heterodimer protein and isolated 'positive' and Cys-N2 strands (each at c 200  $\mu$ M in phosphate buffer).** Assuming dimerization of 'positive' strand and Cys-N2 peptide, the concentration of respective homodimers corresponds to 100  $\mu$ M, which is 50% of Het-N2 heterodimer concentration. This explains weaker amplitudes for 'positive' and Cys-N2 samples.  $s$ ('positive' strand) =  $1.28 \pm 0.13$  S;  $s$ (Cys-N2 strand) =  $1.22 \pm 0.12$  S;  $s$ (Het-N2) =  $1.22 \pm 0.08$  S.

However, a minimum of 200  $\mu\text{M}$  of protein was required to detect the protein signal at 240 nm. In fact, the protein has no aromatic residues strongly absorbing at 280 nm like tryptophan and the buffer necessitate containing TCEP (tris(2-carboxyethyl)phosphine) as reducing agent to avoid artefacts due to cysteine oxidation. The wavelength of detection at 240 nm was selected as TCEP gave only a lower background signal at this wavelength and the protein is still absorbing, but it demands higher concentration.

To demonstrate that the dimer detected for the **Het-N2** sample by sedimentation velocity experiment corresponded to a heterodimer rather than a mixture of the corresponding homodimers we turned to urea denaturation (**Figure 2.15**). Urea was preferred as a chemical denaturant, because the electrostatic interactions between solvent-exposed sidechains were used to stabilize the heterodimer, and the favorability of these interactions would decrease significantly at the very high ionic strength required to denature the protein with Gn-HCl. Urea denaturation by monitoring  $\alpha$ -helical content by CD spectroscopy was therefore performed for **Het-N2** as well as for **Cys-N2** and the 'positive' strand homodimers to compare their stabilities (total concentration of each dimer was 10.5  $\mu\text{M}$ , **Figure 2.15**). The **Het-N2** assembly proved to be very resistant against increased concentrations of urea since only slight change in the intensity of the ellipticity, especially at 222 nm, was observed (**Figure 2.15a**). In contrary, denaturation was more pronounced for both homodimers and in particular for the 'positive' strand homodimer, which showed almost complete denaturation at 8 M urea (**Figure 2.15b-c**). Making a few assumptions one can estimate the  $K_d$  at 8 M urea where all the peptides are at least partially denaturated (**Figure 2.15d**). We assume that the monomer is disordered at 8 M urea (0 ellipticity) and the three-dimensional structures are completely folded into dimer at 0 to 1 M urea corresponding to  $[\theta]_{222} \sim -45 \times 10^3 \text{ deg.cm}^2.\text{dmol}^{-1}$ . Thus, the concentrations of folded dimeric structures at 8 M urea are: 8.8  $\mu\text{M}$  for **Het-N2**, 5.7  $\mu\text{M}$  for **Cys-N2** dimer and 1.2  $\mu\text{M}$  for dimer of 'positive' strand, which corresponds to  $K_d$  values at 8 M urea: 0.33  $\mu\text{M}$ , 4.0  $\mu\text{M}$  and 72  $\mu\text{M}$ , respectively. In Gibbs free energy terms (at  $T = 25 \text{ }^\circ\text{C}$  or 298.15 K), it will correspond to  $\sim 1.5 \text{ kcal/mol}$  and  $\sim 3.2 \text{ kcal/mol}$  destabilization of **Cys-N2** and 'positive' strand homodimers in comparison to heterodimer **Het-N2**. Consequently, the corresponding homodimers are approximately 10- and 100-times less stable than heterodimer, respectively.

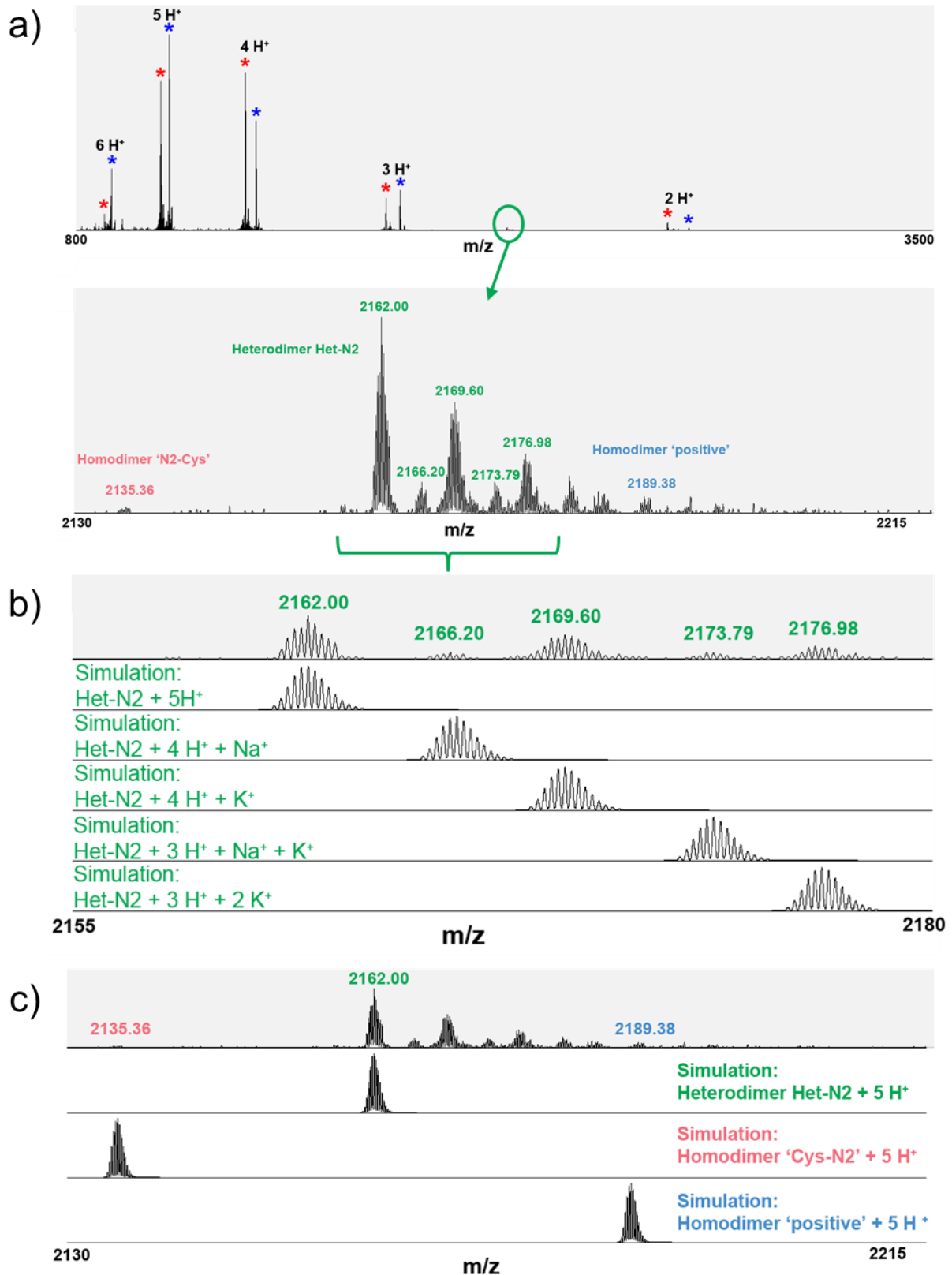
Furthermore, 'native' electrospray mass-spectrometry was attempted to verify the presence in 'solution' of heterodimer **Het-N2** but also of homodimers of **Cys-N2** and 'positive' strand. Although masses for the monomeric strands were mostly detected probably due to dissociation during ionization, ions corresponding to dimeric species were also identified (**Figure 2.16**). Predominantly heterodimeric species were observed with minor quantities of homodimers ( $< 5\%$  with our conditions of ionization), which correlates well with previous urea denaturation experiments.

All these biophysical characterizations strongly support the presence in majority of an  $\alpha$ -helical heterodimer **Het-N2** with probably minor quantities of the corresponding homodimers in solution. The designed heterodimer **Het-N2** is also very robust and stable in solution as demonstrated via thermal and chemical denaturation experiments.



**Figure 2.15. Urea denaturation experiments.** a-c) Circular dichroism spectra showing decrease of  $\alpha$ -helical structure upon increase of the concentration of urea in the buffer (50 mM sodium phosphate, 5 mM TCEP-HCl, pH 7.5) at  $c$  10.5  $\mu\text{M}$  of heterodimer **Het-N2** (a) and homodimers formed from **Cys-N2** (b) or 'positive' strands (c). d) A comparison of denaturation curves for heterodimer **Het-N2** and two homodimers.





**Figure 2.16.** ‘Native’ mass-spectrometry of Het-N2 indicates mainly heterodimer and minute amounts of Cys-N2 and ‘positive’ strand homodimers detected in the mixture. **a)** In the top panel, ions for Cys-N2 strand are labeled with red stars, for ‘positive’ strand with blue stars and for dimeric species in green. Bottom panel is a zoom onto the area with dimeric species. **b)** Experimental data (top panel) versus mass-spectra simulations (below) for various ion adducts of Het-N2. Formula for Het-N2 is  $C_{464}H_{795}N_{153}O_{141}S$ . **c)** Experimental data (top panel) versus simulations (below) for homodimer species. Formula for homodimer of Cys-N2 strand is  $C_{458}H_{772}N_{140}O_{148}S_2$  and formula for homodimer of ‘positive’ strand is  $C_{470}H_{818}N_{166}O_{134}$ .

## 2.5. Crystal structures

To better characterize synthesized DSD analogues, we performed X-ray crystallographic experiments in collaboration with the Structural Biology and Genomics Technology Platform of Institut de génétique et de biologie moléculaire et cellulaire (IGBMC). Crystals were obtained for the three constructs **Het-N2**, **Het-Ncap** and **Hom-N2** by high-throughput screening. Samples of proteins were prepared in 100 mM MES buffer at pH 6.5 containing 100 mM NaCl and 10 mM TCEP at protein concentration between 10 and 13 mg/mL. Designed **Het-N2** as well as **Het-Ncap** and **Hom-N2** were crystallized each in different conditions. The X-ray data were collected at 2.33 Å (from anisotropic data, 4.43 Å in the worst direction), 1.45 Å and 2 Å resolution for the three proteins, respectively, and structures were solved by molecular replacement using the known structure of the original DSD protein (PDB ID: 1G6U).<sup>24</sup> All the protein analogues were well folded forming the intended dimer (homodimer for **Hom-N2** and heterodimers for **Het-N2** and **Het-Ncap**) with a three- $\alpha$ -helical bundle topology (**Figure 2.17-2.20**).

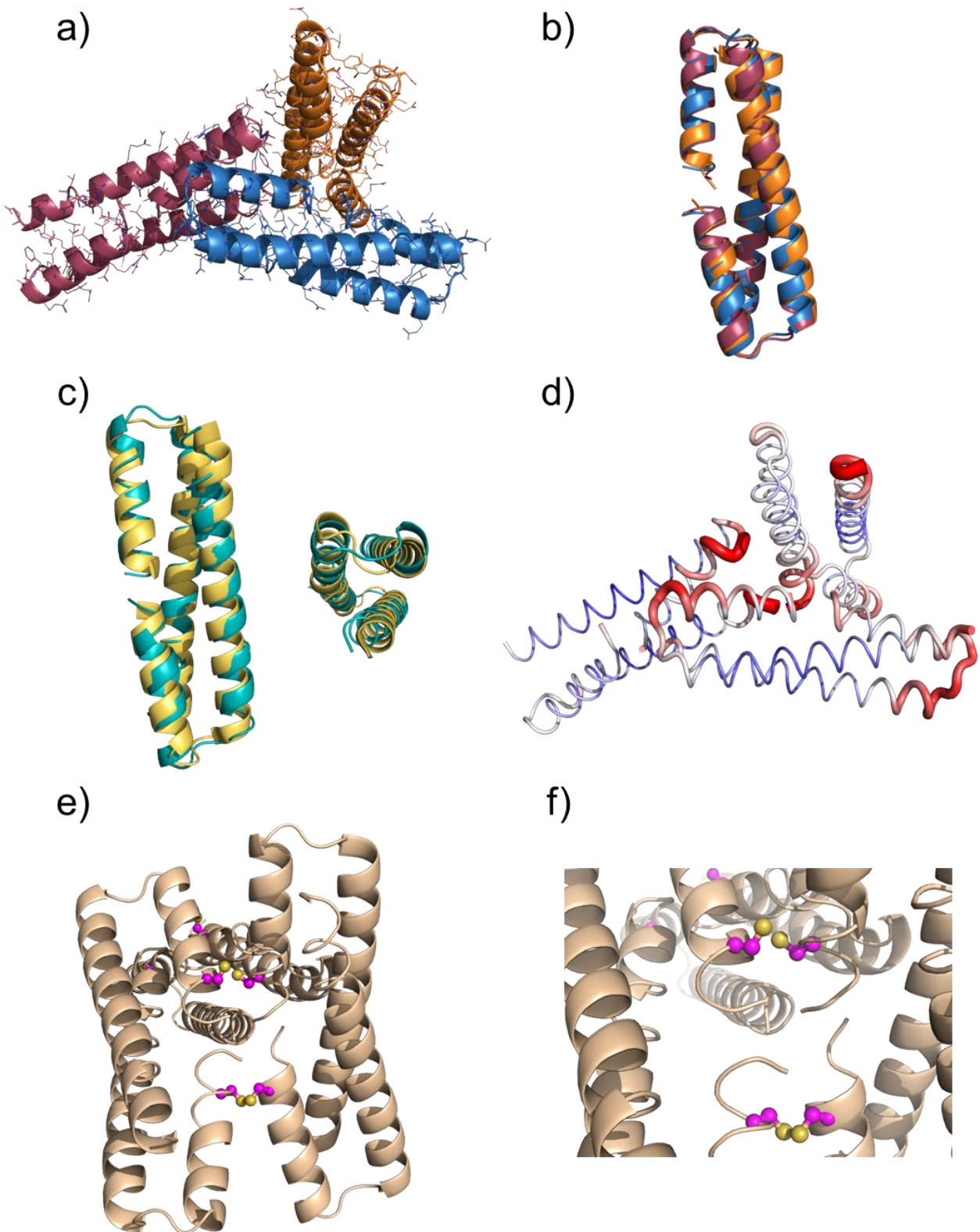
In the crystal structure of the homodimer **Hom-N2**, the asymmetric unit is composed of three homodimers (**Figure 2.17a**). The three copies of the protein overlay well confirming that no alternative conformation of the backbone is present (RMSD on C $\alpha$  varies from 0.643 Å to 0.996 Å, **Figure 2.17b**). Globally, the scaffold of **Hom-N2** construct superimposes well with the crystal structure of the original DSD protein with the observation of only small shifts in the backbone resulting in a RMSD on C $\alpha$  of 1.280 Å (**Figure 2.17c**). The B factor, also called the Debye-Waller factor, in the overall structure of the analogue **Hom-N2** is generally enhanced in the extremities and the loop indicating higher thermal motions in these regions (**Figure 2.17d**). Remarkably, two copies of the **Hom-N2** are covalently joined by a single disulfide bond formed due to oxidation of two corresponding cysteines (**Figure 2.17e-f**). Four N-termini are consequently in close proximity to each other forming a cavity with two other cysteine residues facing to each other. This resulting tetramer is contacted by the third homodimer by one of its helical C-termini and one of its loops. This interesting crystal structure will be discussed further in Chapter 5.

The asymmetric unit of crystal structure of both heterodimers **Het-N2** and **Het-Ncap** are composed of four copies but belong to different space groups (P1 and C2, respectively) as shown in **Figure 2.18a**. For each heterodimers, the corresponding backbones of the four copies thoroughly overlay validating the robustness of the design giving no alternative conformations (mean RMSD on C $\alpha$  for the overlay of the four copies of 0.332 Å for **Het-N2** and 0.522 Å for **Het-Ncap**, **Figure 2.18b**). Here again, B factors are higher in extremities and loop regions as represented in **Figure 2.18c**, which are in general more flexible areas. In

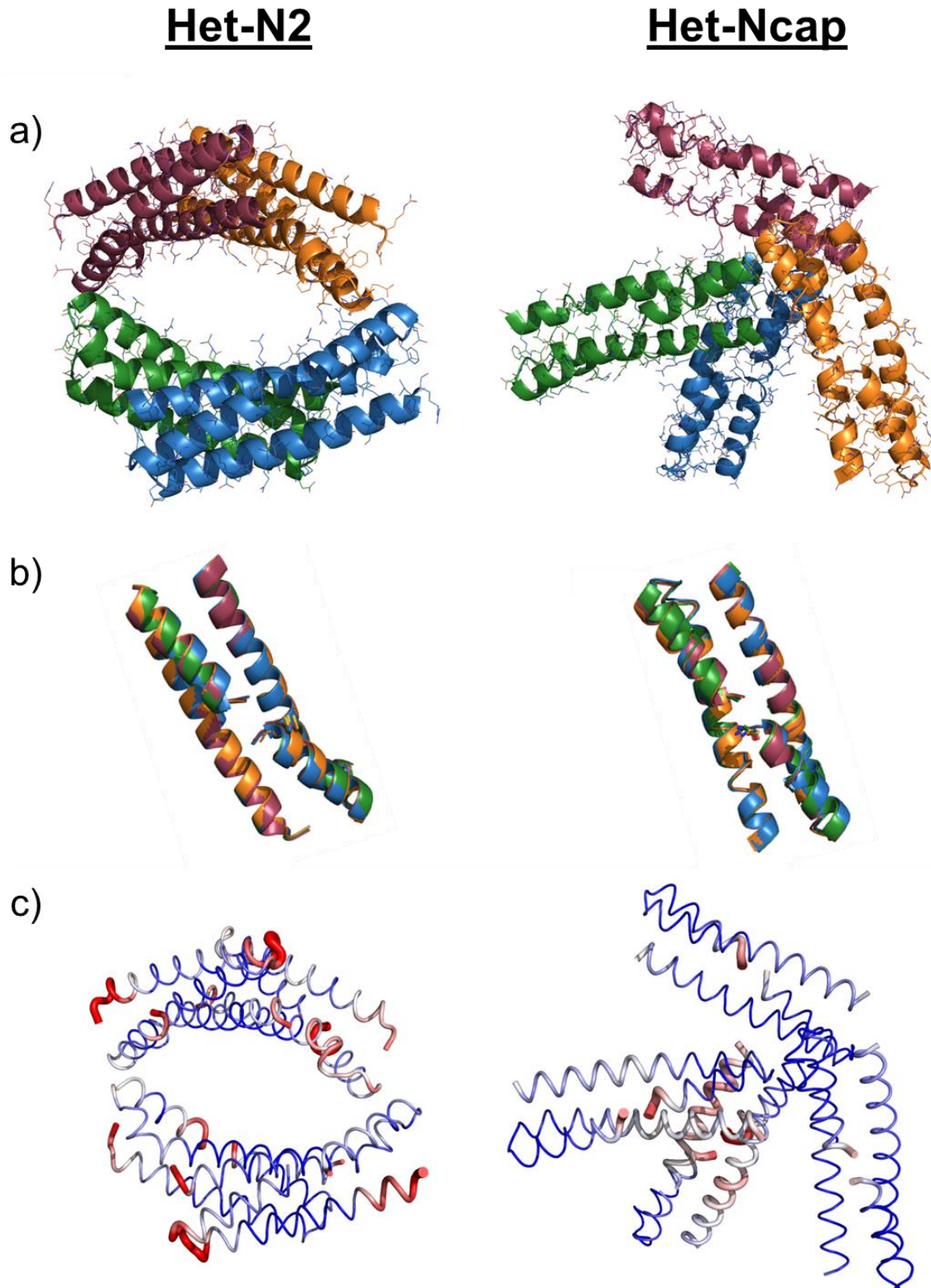
comparison with the original crystal structure of the DSD protein, small shifts of the backbone were observed (on average, 1.839 Å of RMSD on C $\alpha$  for **Het-N2** and 1.793 Å for **Het-Ncap**), resulting in changes of the coiled-coil pitch from 217.7 Å for DSD protein to 199.2 Å for **Het-N2** and 177 Å for **Het-Ncap** (Figure 2.19).<sup>35</sup> Substitution of Arg residues for Lys resulted in rotamer adjustments to globally maintain salt bridges between the two monomers. In both structures, the Cys residues were found to face the solvent (see zoomed area in Figure 2.19), thus being accessible to reactants.

Importantly, by looking in more details to the crystal structure of **Het-N2**, the Cys residue lies adjacent to a thin crevasse formed by the abutting N-terminal ends of the two helices (Figure 2.20). The Cys residue of **Het-N2** is positioned to interact with amide NH groups at the N-terminus of the helix, where its thiol group is within hydrogen-bonding distance of its own amide. The remaining amide NH groups are also well positioned to act as an oxyanion hole to stabilize charge in the transition states and tetrahedral intermediates. The N-terminal end of the second helix provides additional possibilities for stabilization of the anionic tetrahedral intermediate. Thus, the structure of **Het-N2** validates the design and sets the groundwork for future optimization of the sequence surrounding the binding site.

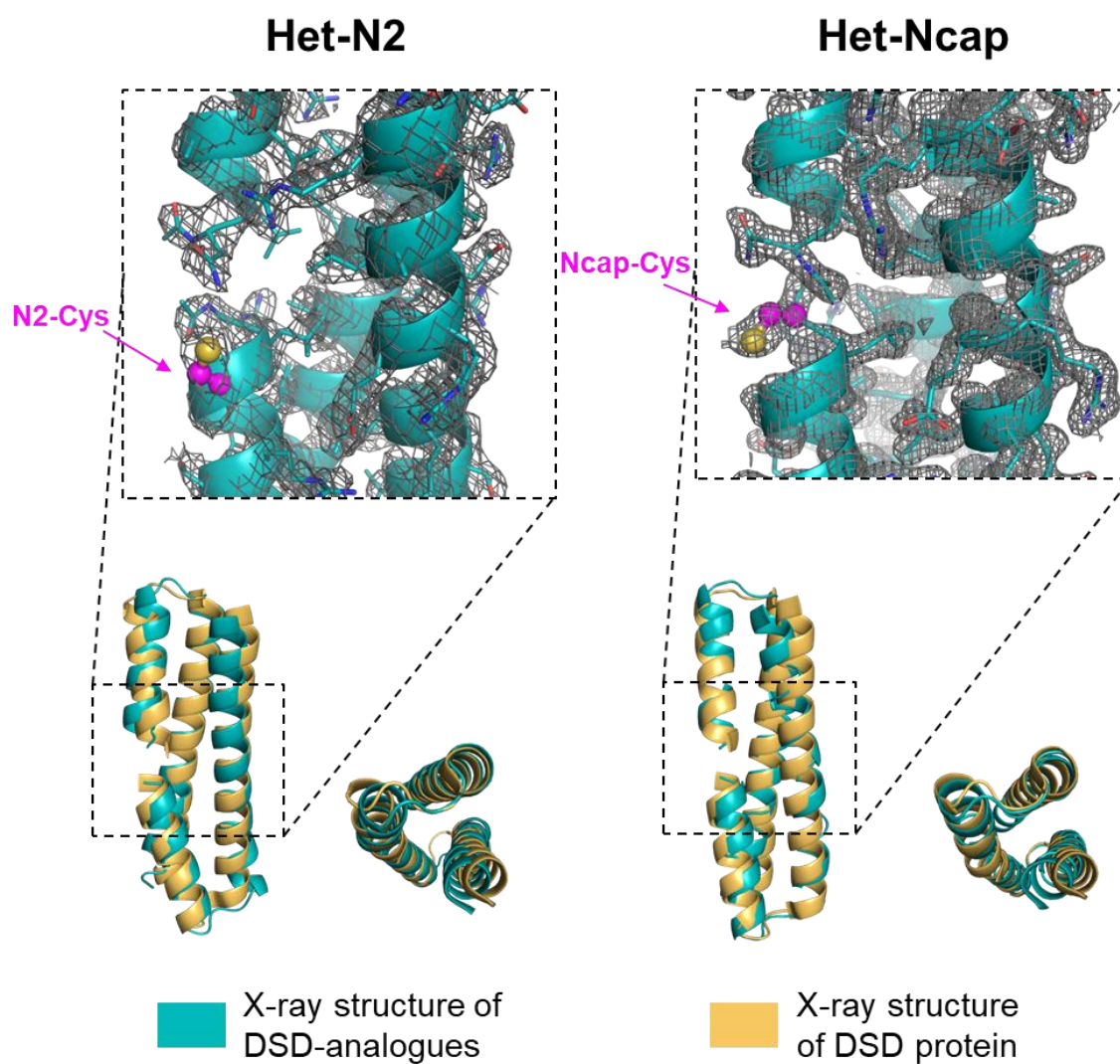
Crystal structures provided important information about the three-dimensional geometry of the different designed analogues. Obtained X-ray structures overlaid well with the original DSD protein, which confirmed the robustness and the modularity of this scaffold by allowing multiple modifications (*i.e.* Lys to Arg, charge patterning for heterodimers, Cys introduction) without perturbing the global structure and folding. However, these crystal structures do not report on the mechanism of catalysis and more derivatives that mimic tetrahedral intermediate need to be produced. To obtain such a mimic, we plan to protect the catalytic cysteine by oxidative sulfitolysis in presence of sodium sulfite (Na<sub>2</sub>SO<sub>3</sub>) and sodium tetrathionate (Na<sub>2</sub>S<sub>4</sub>O<sub>6</sub>) in aqueous buffer, which would lead to a negatively charged adduct on cysteine (protein)-S-SO<sub>3</sub><sup>-</sup>.<sup>36</sup> Moreover, this protecting group will be also advantageous to avoid cysteine oxidation and formation of disulfide bridges between different copies of the protein during crystallization process as it was the case with the **Hom-N2** protein.



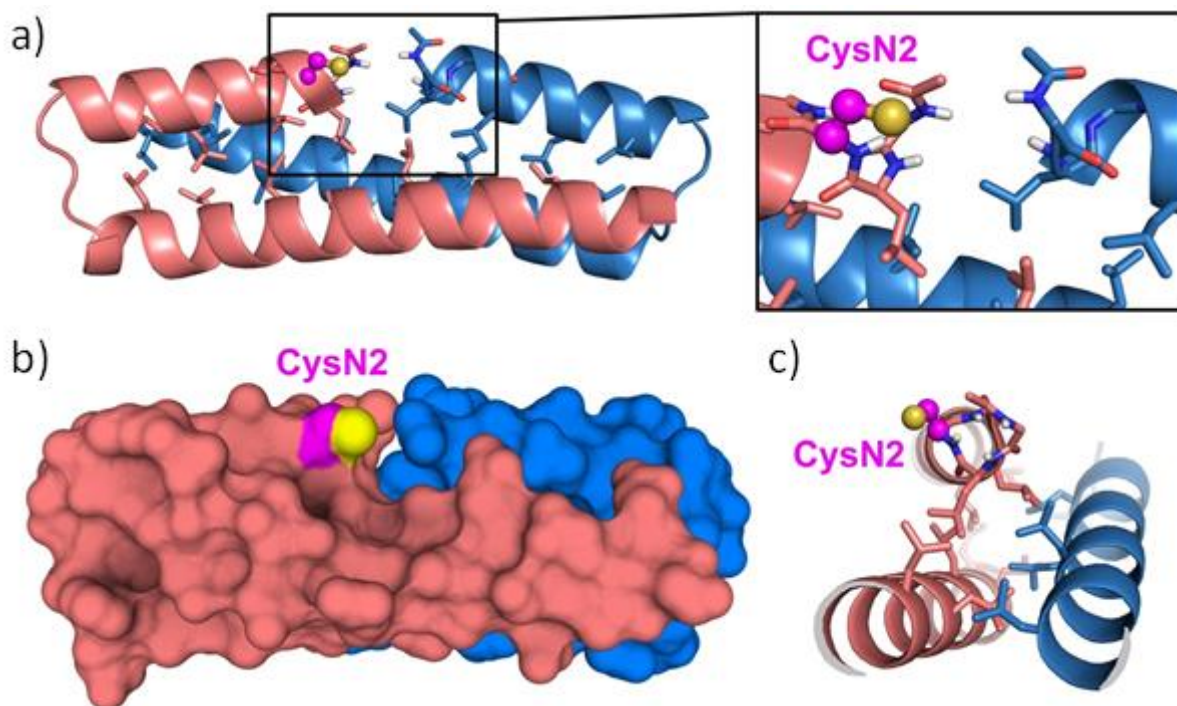
**Figure 2.17. Crystal structure of the homodimer Hom-N2.** **a)** Asymmetric unit of the crystal structure included in the monoclinic space group C2. **b)** Overlay of the three copies of the structure comprised in the asymmetric unit. **c)** The **Hom-N2** structure (depicted in cyan) superimposes with the crystal structure of the original DSD protein (in yellow), side view on left and top view on right of the alignment are represented. **d)** Cartoon putty B-factors representation in Pymol highlights regions with higher B factors (in red) and with lower B factors depicted in blue. **e)** Representation of the crystal structure with catalytic cysteines depicted in magenta with sulfur atom in yellow. Two copies of the homodimer are covalently joined by a single disulfide bond that emerged due to oxidation of the two corresponding cysteines. **f)** Zoom on the four N-termini that are in close proximity to each other, due to a disulfide bridge, forming a deep cavity with two other cysteine residues facing each other.



**Figure 2.18. Crystal structures of the two heterodimer proteins Het-N2 (left) and Het-Ncap (right).** a) Four copies of heterodimer were observed in each asymmetric unit of the corresponding X-ray structures. b) Overlay of the four heterodimers for each X-ray structure. For **Het-N2**, all-atom RMSD between the four copies vary between 1.166 Å and 1.496 Å. For **Het-Ncap**, all-atom RMSD between the four copies are between 1.110 Å and 1.486 Å. c) Cartoon putty B-factors representation in Pymol highlights regions with higher B factors (in red) and with lower B factors depicted in blue.



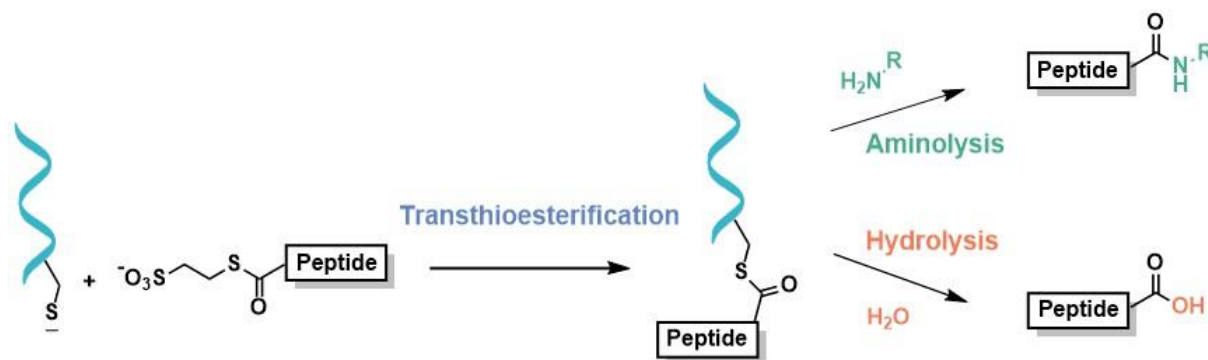
**Figure 2.19.** X-ray structures of the synthesized DSD-analogues **Het-N2** (left) and **Het-Ncap** (right). Catalytic sites are shown revealing solvent-exposed cysteine residues depicted in magenta with sulfur atom in yellow in ball-and-sticks representation. The simulated annealing  $2F_o-F_c$  electron density map is shown at  $1.0\sigma$  level. The X-ray structures of **Het-N2** and **Het-Ncap** (depicted in cyan) are superimposed onto the original DSD structure (in yellow).



**Figure 2.20.** The active site is located at the interface of two N-termini of  $\alpha$ -helices with nucleophilic CysN2 positioned next to a small cavity. a) Horizontal view illustrating helix-helix interface with the unpaired NH groups pointing towards the cavity. b) Surface representation. c) View along the helical axis.

## 2.6. Optimization of catalytic assays for amide bond formation

We aim at accelerating acyl transfer reactions on peptide- $\alpha$ thioester with peptides bearing a free N-terminal amine or miscellaneous amines. To do so, we needed to develop assays to follow the evolution of such reactions. The reaction mechanism is composed of two main steps: the transthioesterification with the formation of a branched adduct of the catalytic protein followed by the aminolysis resulting in the formation of the desired amide bond in competition with hydrolysis leading to the formation of the carboxylic acid side product (**Figure 2.21**). In this way, we tried different possibilities to follow the catalyzed reaction (in parallel with uncatalyzed reaction) to find a compromise between collection of quantitative data of the formed products and intermediates and convenience of running various assays at the same time.



**Figure 2.21. Catalytic pathways of acyl transfer reactions of peptide- $\alpha$ -thioester.** After the first step of transthioesterification, aminolysis of the resulting branched adduct is in competition with the side reaction of hydrolysis.

### 2.6.1. Fluorescent catalytic assay based on FRET with a quencher

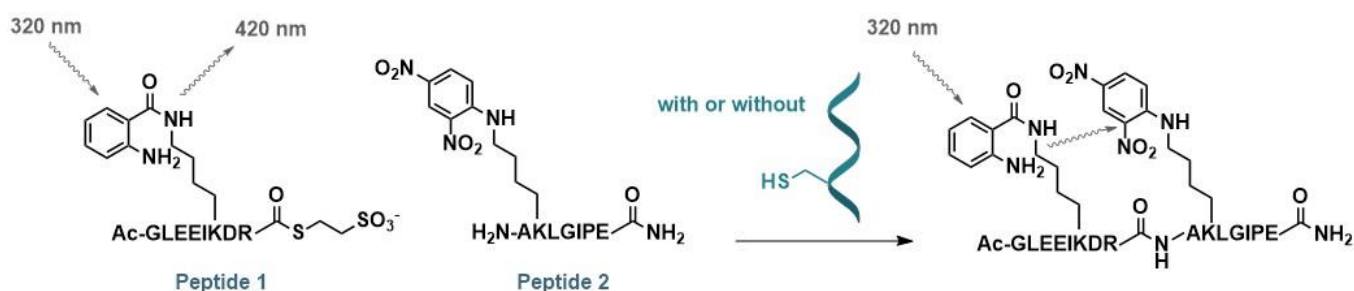
Continuous assays are the most precise ways to get kinetic curves to extract the rate of reaction, whereas fixed-time (or discontinuous) assays entails to take aliquots of an enzymatic reaction at fixed interval of time allowing for the monitoring of multiple reactions in parallel like when using a microplate reader. Fluorometric assays provide an access to both continuous and discontinuous assays and necessitate fluorescent substrates. Fluorescence is the capacity of a molecule to absorb light at a specific wavelength resulting in the emission of light at a different wavelength, in other words fluorescence occurs when an orbital electron goes into its excited singlet state followed by relaxation to its ground state by emitting a photon. Difference in the intensity of fluorescence of a substrate enables to measure enzymatic reaction. Although this method is very sensitive and quantitative, it can suffer from intrinsic artefacts such as photobleaching of the fluorophore. Additionally, FRET (Förster resonance energy transfer) assays are indeed appropriate to follow a bimolecular reaction that assembles two substrates together like in our case. FRET consists in a nonradiative energy transfer through dipole-dipole coupling between two molecules (a donor and an acceptor) and requires that the emission spectrum of the donor covers, at least partially, the absorption spectrum of the acceptor molecule. The efficiency of the energy transfer is directly linked to the distance between the donor and the acceptor. In this way, FRET experiments can, for example, directly provide information about assembly of two peptides or report on peptide proteolysis. The donor fluorophore can be coupled to a quencher that would absorb the emitted light or to another fluorophore that would absorb the emitted light and then reemit at another wavelength. In our case, by placing the donor and the acceptor on different



peptide substrates (one on the peptide- $\alpha$ thioester and the other on peptides bearing a free N-terminal amine), we would precisely monitor the ligation of the two peptides and be able to determine the rate acceleration in the presence of our catalytic protein.

We first decided to inspect the possibility to follow the catalysis of peptide assembly by fluorescence with a fluorophore conjugated to a first peptide fragment and a quencher on the second. Basically, the decrease of fluorescence meaning that ligation of the two peptide substrates occurs, would allow us to extract kinetic parameters of peptide aminolysis for our putative catalytic variants. The sequences of these two peptide substrates were designed as follows (**Scheme 2.4**): *i*) sequences are based on intrinsically disordered peptide model to have accessible extremities for the reaction; *ii*) peptide1- $\alpha$ thioester is acetylated to avoid intramolecular aminolysis, and first prepared as acyl hydrazide; *iii*) the peptide2 substrate bears an N-terminal free amine in order to participate in a direct aminolysis; *iv*) 2-aminobenzoyl (Abz)<sup>37</sup> is used as the fluorophore and anchored to a Lys residue of the peptide1- $\alpha$ thioester on resin; *v*) 2,4-dinitrophenyl (Dnp)<sup>37</sup> is used as the quencher and is directly introduced to the peptide2 via the commercially available building block Fmoc-L-Lys(Dnp)-OH.

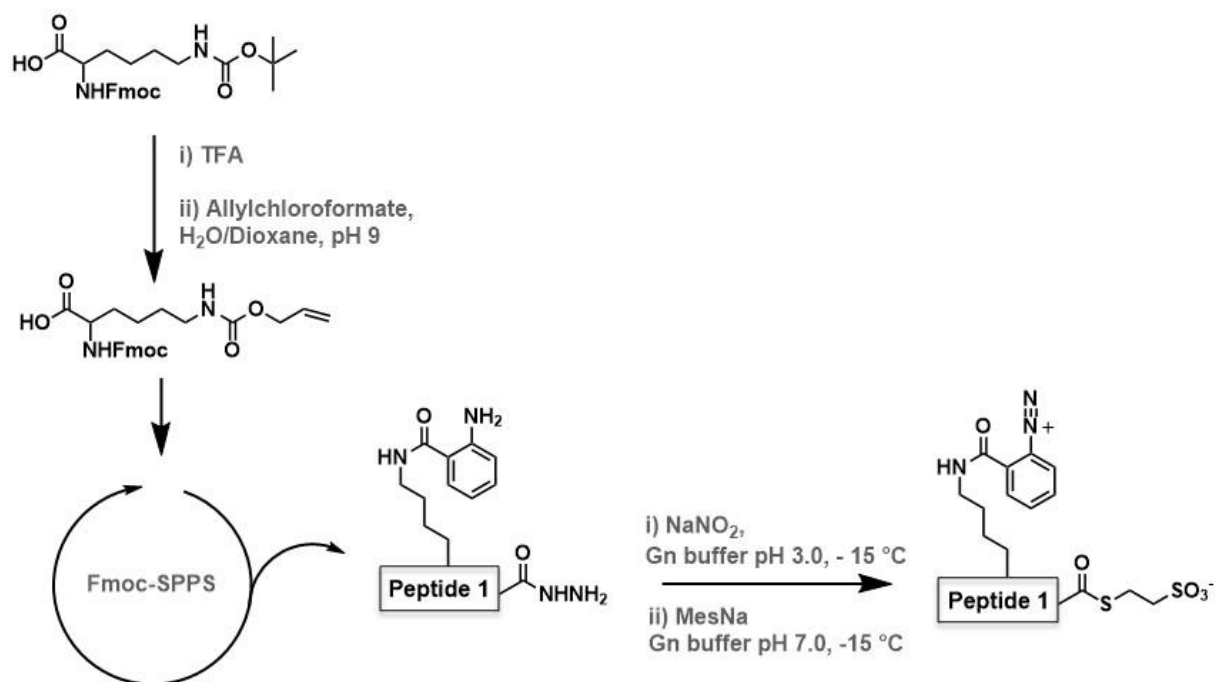
We first focused on the synthesis of peptide1 which is the N-terminal peptide fragment containing the Abz fluorophore. Fmoc-Lys(Alloc)-OH, whose side chain can be orthogonally deprotected with a catalyst (palladium-tetrakis) and a hydrogen donor (phenylsilane),<sup>38</sup> was first synthesized from the commercially available Fmoc-Lys(Boc)-OH in a two-step transformation (**Scheme 2.5**). First, quantitative deprotection of Boc-group was carried out in an acidic medium (100% of trifluoroacetic acid). After removing the excess of reagent under vacuum, the resulting Fmoc-Lys-OH was dissolved in an alkaline mixture (pH 9) in order to enable free nucleophilic amine of the lateral chain to react with the allylchloroformate added in slight excess (1.2-fold). After extraction and purification by



**Scheme 2.4. Principle of the catalytic assay based on FRET with quencher.** A peptide1- $\alpha$ thioester bearing an Abz fluorophore can react with a peptide2 containing a free N-terminal amine and a Dnp quencher moiety. As the two separated peptides are too far in solution, fluorescence of peptide1 is observed. Once the two peptides are ligated (with or without acceleration of the reaction in presence of our catalytic protein), the acceptor quenches the emission of the donor and fluorescence is not observed anymore.

reverse phase flash chromatography, the expected Fmoc-Lys(Alloc)-OH was obtained with high purity compatible for further SPPS and a yield of 51 %. Then, Fmoc-SPPS on 2-chloro-trityl resin prefunctionalized with a hydrazine group (2-Cl-Trt-NHNH<sub>2</sub>) was performed to assemble the peptide including the Lys(Alloc) residue. After final deprotection and acetylation of the N-terminal amine, the Alloc protecting group was selectively removed with tetrakis(triphenylphosphine)palladium(0) and phenylsilane under argon atmosphere.<sup>38</sup> The fluorophore building block (Boc-Abz-OH) was coupled on the resulting free amine following standard procedure for solid support coupling. After cleavage, the peptide was retrieved in a good purity allowing thioesterification to be performed on crude material. Following the standard two-step procedure of oxidation of acyl hydrazide to acyl azide and then thioesterification, a side reaction was also observed on the aryl amine of the Abz fluorophore. Actually, in the presence of sodium nitrite at acidic pH, diazotization reaction on aryl amine occurred following similar mechanism as for the transformation of peptide- $\alpha$ hydrazide to peptide- $\alpha$ azide.

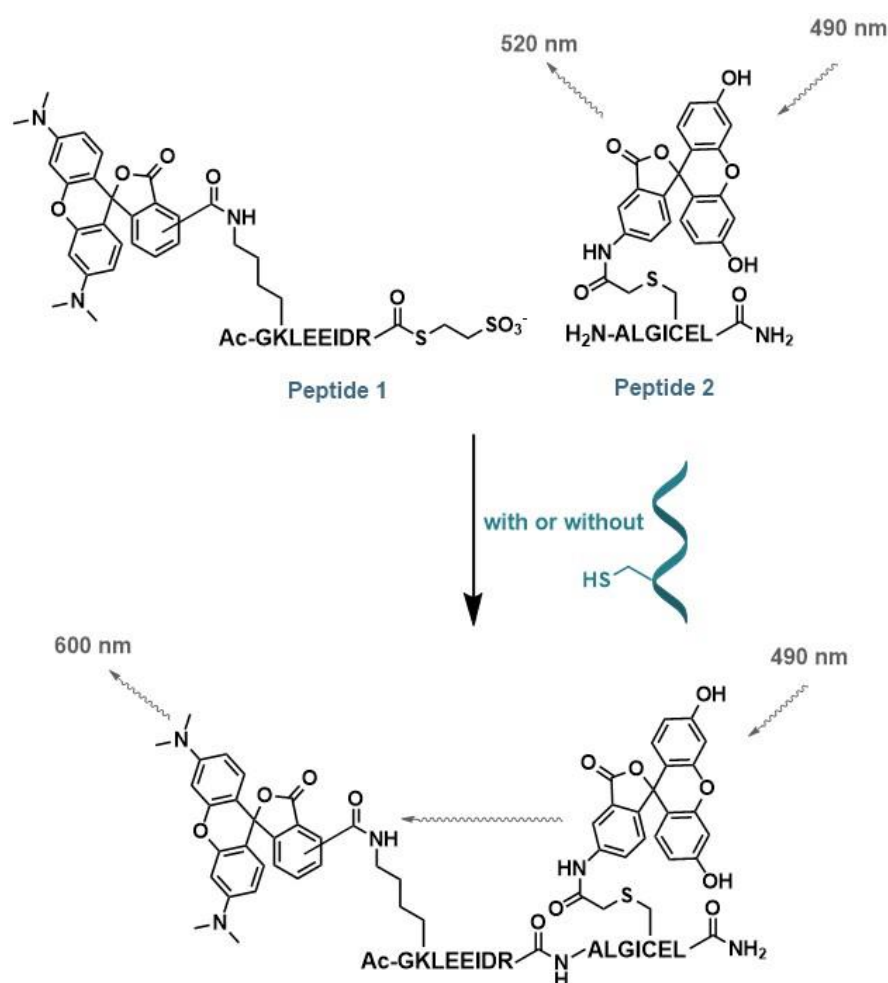
Although synthetic alternatives would have been possible to attempt such as protection of the aryl amine or inversion of the fluorophore and the quencher on the respective peptide, this strategy for monitoring catalytic assay was abandoned. FRET with quencher is actually more prone to artefacts than FRET in the presence of a second fluorophore as acceptor, so meanwhile we also tried to elaborate another method for catalytic assays with two fluorophores for FRET experiments.



**Scheme 2.5. Synthetic pathway of peptide1- $\alpha$ thioester carrying Abz fluorophore.** Fmoc-Lys(Alloc)-OH building block was first synthesized from Fmoc-Lys(Boc)-OH in two steps. Peptide1 was assembled by Fmoc-SPPS on resin. Resulting peptide was thioesterified in two steps, however, a side reaction of diazotization on the Abz fluorophore was observed.

### 2.6.2. Fluorescent catalytic assay based on FRET with two fluorophores

As many assays had to be planned and as we wanted also to get a first proof-of-concept that FRET can be conveniently used for our assays, we decided to select non expensive dyes. We chose the pair fluorescein/rhodamine as donor and acceptor for FRET experiments, respectively.<sup>39,40</sup> As before, by conjugating one fluorophore on the first peptide fragment and the second fluorophore on the second peptide fragment, we expected to extract kinetic parameters of peptide aminolysis for our catalytic analogues via modification of the light emission (FRET only observed for ligated peptides) while exciting at the same wavelength. The sequences of these new peptide substrates were designed as follows (**Scheme 2.6**): *i*) peptide sequences are similar as for previous FRET experiments with quencher; *ii*) peptide1-<sup>α</sup>thioester is acetylated and the peptide2 substrate bears a N-terminal free amine in order to catalyze a direct aminolysis between the two peptides; *iii*) fluorescein is used as the donor fluorophore and anchored to a cysteine residue of peptide2 thanks to a

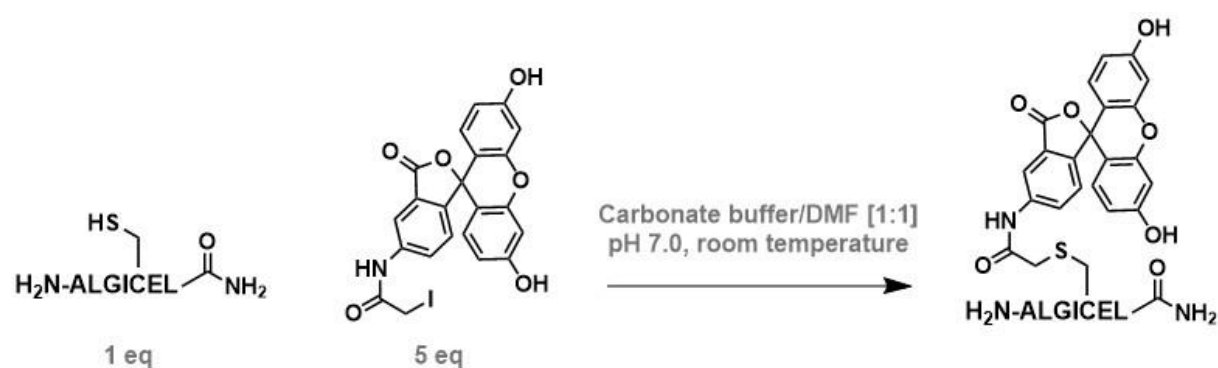


**Scheme 2.6.** Principle of the catalytic assay based on FRET with two fluorophores. Upon ligation between peptide1-<sup>α</sup>thioester and peptide2, FRET can be measured while exciting the fluorescein moiety at 490 nm and detecting emission of rhodamine derivative at around 600 nm. Tetramethylrhodamine coupled via an activated NHS ester on a lysine residue is represented here.

iodoacetamido moiety; *iv*) rhodamine B or tetramethylrhodamine (shown in **Scheme 2.6**) are used as the acceptor fluorophore and are conjugated to a Lys residue of peptide1 (no other free amine present in this peptide) thanks to either a isothiocyanate moiety or a NHS (N-succinimidyl) ester, respectively.

We first focused on the production of peptide2 carrying the fluorescein (**Scheme 2.7**). Synthesis of peptide2 was performed by Fmoc/tBu-SPPS in a very good yield of 74 % after purification. The free cysteine was then alkylated by the 5-(iodoacetamido)fluorescein in solution of carbonate buffer/DMF [1:1]. The reaction was spontaneous and excess of alkylating reagent was quenched by DTT before HPLC purification.

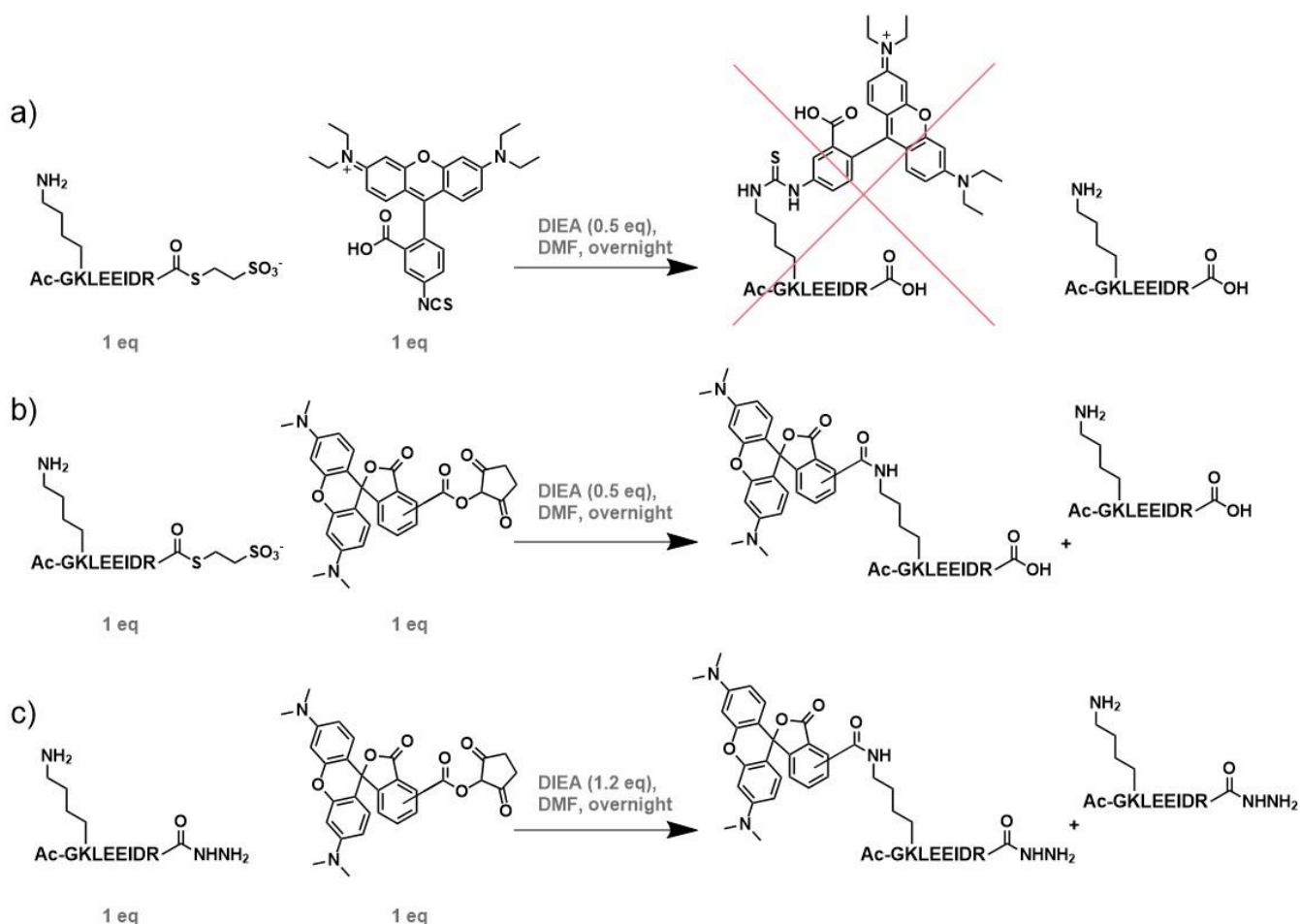
The conjugation of rhodamine derivatives on peptide1 was more challenging. Fmoc-SPPS on 2-Cl-Trt-NHNH<sub>2</sub> resin was performed to assemble the peptide1-<sup>α</sup>hydrazide. The resulting crude peptide was thioesterified following the standard two-step procedure (oxidation with NaNO<sub>2</sub> followed by thioesterification with MesNa).<sup>27</sup> Attempts to label the lysine residue of peptide1-<sup>α</sup>thioester with NHS-ester derivative of tetramethylrhodamine and isothiocyanate derivative of rhodamine B were unsuccessful even under anhydrous conditions, because DIEA (for deprotonation of the amine moiety on lysine) resulted in immediate hydrolysis of thioester moiety (**Scheme 2.8 a-b**). Both labels were present in the form of isomer mixtures that complicated HPLC reaction monitoring. Eventually, purification of both NHS ester and isothiocyanate rhodamines needs to be done prior to labelling. Nevertheless, substantial labelling was observed with these conditions in the presence of NHS-ester derivative of tetramethylrhodamine, whereas no labelling was detected in presence of the isothiocyanate moiety even with 5-fold excess of dye and higher concentration of base (**Scheme 2.8 a-b**). Thus, we decided to keep the NHS-ester derivative and changed our synthetic strategy to first label the precursor peptide1-<sup>α</sup>hydrazide and then perform thioesterification (no primary aryl amines are present). Mixture of peptide1-<sup>α</sup>hydrazide and NHS-5(6)carboxyrhodamine in a 1:1 ratio in presence of a slight excess of DIEA (1.2 eq) converted the peptide in a sufficient amount enabling HPLC purification to isolate the conjugated peptide (**Scheme 2.8 c**). However, due to very low amount recovered,



**Scheme 2.7.** Labelling of the N-terminal peptide2 fragment by alkylation of the cysteine residue. Peptide2 was mixed with an excess of 5-(iodoacetamido)fluorescein in a solution of carbonate buffer and DMF. Reaction was complete instantaneously and excess of alkylating reagent was quenched with DTT.

yield could not be calculated and further thioesterification was not tested.

Finally, this strategy was also abandoned due to too low amount of obtained products. Although synthesis of labelled peptide2 by alkylation was very efficient, coupling rhodamine derivative through NHS ester was more laborious. FRET approaches to follow our catalytic assays are actually not convenient as our protein analogues (**Het-N2** especially) do not bear enzyme-like kinetic parameters (see hereinafter) and so need larger amount of substrates and catalytic protein to observe rate acceleration. Producing labelled substrates in large quantity in such way seems not appropriate for the moment. Additionally, FRET experiments would provide only information about ligation of the two peptide substrates and not about formation of the intermediate branched adduct and the product of hydrolysis, which are also important parameters for our kinetics. For these reasons, we decided to move to less precise methods but more informative in terms of complete composition of the catalytic mixture, such as analytical HPLC and LC/MS.

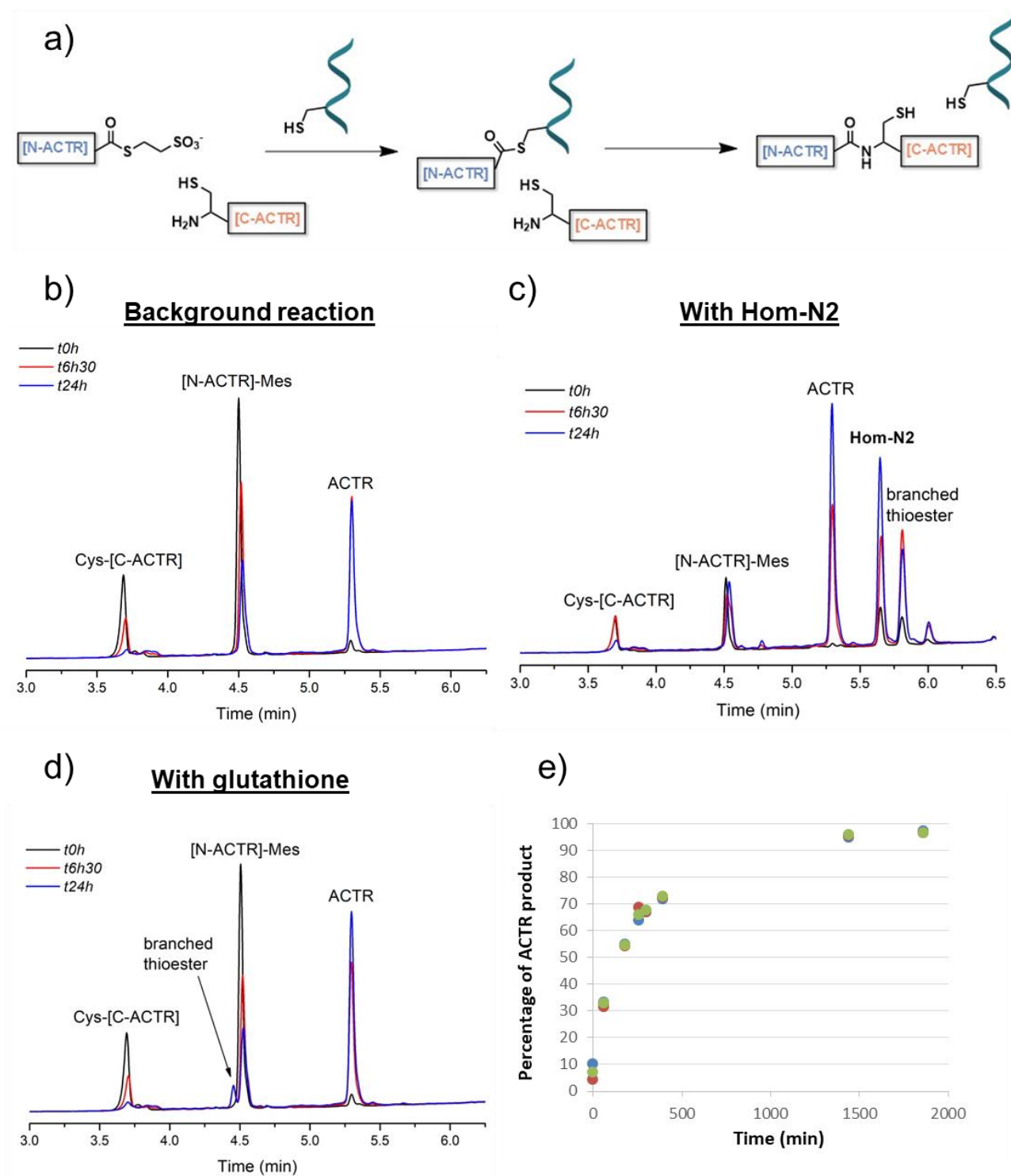


**Scheme 2.8. Labelling of peptide1 with rhodamine derivatives.** a) Attempt to label the peptide1- $\alpha$ thioester with rhodamine B isothiocyanate. Hydrolysis of thioester was observed with no labelling of the amino group of the lysine residue. b) Labelling of peptide1- $\alpha$ thioester with 5(6)-carboxytetramethylrhodamine N-succinimidyl ester (NHS). Although thioester was completely hydrolyzed, partial labelling of the peptide was detected. c) Labelling of peptide1- $\alpha$ hydrazide with 5(6)-carboxytetramethylrhodamine N-succinimidyl ester (NHS). Sufficient amount of labelled peptide was obtained to isolate it by preparative HPLC.

### 2.6.3. Catalytic assays followed by HPLC – 1<sup>st</sup> generation assay

This discontinuous assay consists of mixing the different non-labelled substrates in presence or in absence of a catalytic protein in phosphate buffer with a reducing agent (here, TCEP, tris(2-carboxyethyl)phosphine) to avoid oxidation of the catalytic cysteine. The reaction mixture was let under stirring and aliquots at different time points were taken off for HPLC and LC/MS analysis. Areas of the different peaks on HPLC spectra are easily accessible and allow to estimate evolution of concentration of the products and of the branched adduct intermediate. At this stage, peptide substrates were used at quite significant concentration (around 0.5 mM) and equimolar quantity of catalytic protein. The advantage to follow the catalytic assay by HPLC is that only a good separation of the peptides is needed and can be adjusted via the method of analysis.

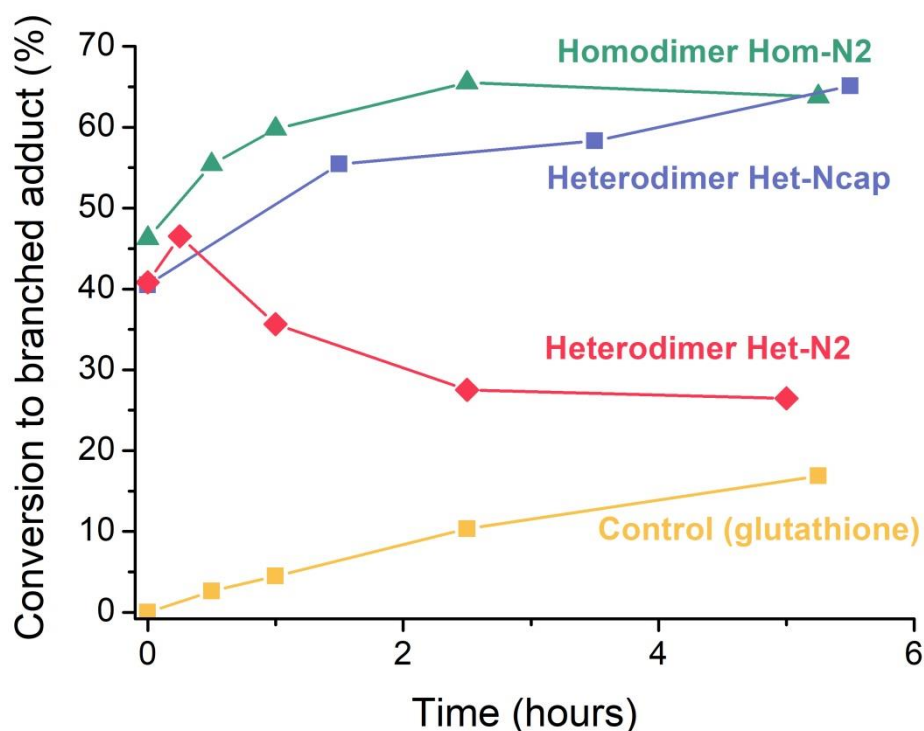
Since our goal is to develop new catalysts for chemical protein synthesis, our first attempt with such discontinuous assay was to test the effect of the homodimer protein **Hom-N2** on the NCL reaction of two peptide fragments. For the first experiments, two already synthesized peptides of the laboratory, a 30-residue peptide- $\alpha$ thioester fragment and a 17-residue Cys-peptide fragment used in the synthesis of intrinsically-disordered activation domain from p160 transcriptional co-activator (ACTR), were chosen (**Figure 2.22a**).<sup>41</sup> As these peptide fragments are coming from an intrinsically disordered protein, they should not bear any folded secondary structure in phosphate buffer and so we expect the extremities to be accessible for the reaction. Transthioesterification is the first step in NCL, where thiol group of the Cys-peptide attacks thioester group of N-terminal peptide- $\alpha$ thioester and is followed by intramolecular S-to-N acyl transfer to form native peptide bond. Typically, exogenous aryl-thiols (e.g. mercaptophenol or 4-mercaptophenylacetic acid) are added in excess to convert peptide- $\alpha$ alkylthioesters into more reactive peptide- $\alpha$ arylthioesters *in situ*.<sup>42</sup> In our experiments, instead of aryl-thiols we added our **Hom-N2** homodimer in stoichiometric amounts with respect to ligating peptides (**Figure 2.22a**). As control reactions, we tested the same reaction without any catalytic protein (*i.e.* background reaction) as well as in the presence of glutathione, a natural tripeptide that also contains an internally placed Cys residue. Ligation of N-terminal peptide fragment of ACTR synthesized as Mes thioester (Mes = 2-mercaptoethanesulphonate) with C-terminal Cys-peptide leads to the expected 47-residue ACTR product, where reaction is complete after 24 hours (concentrations of [N-ACTR]- $\alpha$ Mes and [C-ACTR] fragments are 0.7 and 0.5 mM, respectively) (**Figure 2.22b** and **e**). In the presence of **Hom-N2** protein, we observed instantaneous formation of branched thioester that is the product of transthioesterification of [N-ACTR]- $\alpha$ Mes fragment with **Hom-N2** (see **Figure 2.22a, c** and **e**), which then reacted with Cys-[C-ACTR] fragment to form



**Figure 2.22. Native Chemical Ligation catalytic assay.** **a)** Scheme of the NCL between the [N-ACTR]-<sup>q</sup>Mes and the Cys-[C-ACTR] fragments in presence of the catalytic protein **Hom-N2**. A first step of transthioesterification between [N-ACTR]-Mes and the catalytic cysteine would lead to the formation of a branched thioester that would then react with the Cys-[C-ACTR] fragment. **b)** HPLC monitoring of the background reaction without catalyst. **c)** HPLC monitoring of the NCL between [N-ACTR]-Mes and the Cys-[C-ACTR] in presence of **Hom-N2** homodimer. Formation of branched thioester was clearly observed. **d)** HPLC monitoring of the NCL reaction in presence of glutathione. A lower and slower formation of the branched thioester was detected. **e)** Formation of the ligated ACTR product calculated from areas of the peak on HPLC. Blue dots correspond to the background reaction, red dots to the reaction with **Hom-N2** and green dots to the reaction with glutathione. No changes are observed with additives.

ACTR product. The HPLC and LC/MS monitoring showed rates of product formation similar to the conditions without any added catalyst (Figure 2.22e). In the third experiment, the same reaction was performed in the presence of glutathione, where conversion rate was again similar to the previous cases; however, formation of branched adduct of glutathione was at substantially lower level (Figure 2.22d and e). The results indicate that **Hom-N2** homodimer indeed interferes with the NCL of peptide fragments by participating in a very fast transthioesterification and suggests a unique reactivity of Cys residue at N-terminus of DSD construct. However, in the subsequent step the resulting branched adduct is not much more reactive than [N-ACTR]-Mes thioester and the net effect on catalysis of the native chemical ligation is subtle.

While performing NCL in the presence of **Hom-N2** catalyst, we noticed a substantial amount of thioester hydrolysis side product. Subsequently, we performed experiments to quantify the rate of hydrolysis that were also used to follow more carefully the evolution of branched adduct. [N-ACTR]-<sup>α</sup>Mes fragment alone was subjected to reaction with the different designed analogues **Het-N2**, **Het-Ncap** and **Hom-N2** (0.5 mM of thioester and of catalytic proteins). In all cases, we instantaneously observed the formation of branched adduct on the catalytic protein, which was then entirely hydrolyzed within 50 hours, whereas the formation of branched adduct with glutathione was much slower. Interestingly, by looking closely at the six first hours of the reaction, substantial differences in the evolution of the branched adduct were detected between the analogues (Figure 2.23). The three protein catalysts **Het-N2**,



**Figure 2.23. Branched adduct evolution.** Changes in concentration of [N-ACTR]-protein branched thioester resulting from the transthioesterification between [N-ACTR]-<sup>α</sup>Mes thioester and **Het-N2**, **Het-Ncap**, **Hom-N2** or glutathione.



**Het-Ncap** and **Hom-N2** promptly reacted with the [N-ACTR]-<sup>α</sup>thioester to directly form more than 40 % of branched adduct compared to the remaining free catalyst. **Het-Ncap** and **Hom-N2** analogues reached 65 % of conversion to branched adduct after 5 hours of reaction before being consumed by hydrolysis, whereas the branched adduct formed on the **Het-N2** protein was directly hydrolyzed after less than one hour of reaction. Half-lives ( $\tau_{1/2}$ ) were approximately estimated based on LC/MS monitoring of the reaction and confirmed a greater acceleration of thioester hydrolysis in presence of **Het-N2**:  $\tau_{1/2}$  of thioester substrate [N-ACTR]-Mes was equal to 1 h in the presence of **Het-N2**, while half-lives were equal to 20 h for **Het-Ncap** and 24 h for **Hom-N2**. The corresponding experiments for glutathione and for background reaction resulted in  $\tau_{1/2} = 40$  h and 96 h, respectively. Assuming second order reaction kinetics, the apparent observed kinetic constants were calculated from half-lives and are summarized in **Table 2.1**.

In addition, we performed direct aminolysis reaction using large excess of Tris ( $c = 250$  mM, tris(hydroxymethyl)aminomethane) as model amine because the  $pK_a$  of amino group in Tris is 8.1 (at 25 °C), which is similar to the value for N-terminal peptide amino group.<sup>43</sup> The resulting branched adduct formed through transthioesterification of [N-ACTR]-<sup>α</sup>Mes ( $c = 0.5$  mM) fragment and the different designed analogues **Het-N2**, **Het-Ncap** and **Hom-N2** ( $c = 0.5$  mM) showed substantial reactivity in aminolysis reaction (see **Table 2.2**). Faster kinetic of aminolysis in comparison with hydrolysis is in agreement with a general preference of thioesters to react faster with amines than with oxygen-containing nucleophiles (e.g.  $\text{OH}^-$ ).<sup>20</sup> Here again, the **Het-N2** protein clearly presented the fastest kinetics in comparison to **Het-Ncap** and **Hom-N2**, which also accelerated the reaction compared to background reaction. However, the side reaction of hydrolysis was also observed and precise quantification of the ratio of aminolysis versus hydrolysis was not achievable with this method of assay monitoring.

**Table 2.1. Kinetic parameters for hydrolysis of [N-ACTR]-<sup>α</sup>Mes with or without protein catalysts.** Comparison of the kinetic parameters between the analogues **Het-N2**, **Het-Ncap** and **Hom-N2**.

Conditions / catalytic protein	$\tau_{1/2}$ (hours)	$k_{\text{obs}}$ ( $\text{M}^{-1} \cdot \text{s}^{-1}$ )
Background reaction (without catalyst)	96	0.006
Glutathione (negative control)	40	0.014
<b>Het-N2</b>	1	0.56
<b>Het-Ncap</b>	20	0.028
<b>Hom-N2</b>	24	0.023

Thanks to this catalytic assay, several parameters were confirmed to not affect kinetics of the reaction. Catalytic proteins coming from different batches of synthesis and even from different method of synthesis carried same catalytic profiles. For instance, the protein **Hom-N2** that was provided either by the 2-segment ligation strategy or by the 3-segment ligation strategy (with the difference of an alanine residue or an alkylated cysteine in position 24, respectively) resulted in similar acceleration of thioester hydrolysis and half-life of the [N-ACTR]-<sup>o</sup>Mes of around 24 h for both cases. Moreover, for practical reasons, we decided to move to a shorter fragment of the [N-ACTR]-<sup>o</sup>Mes peptide that is easier to produce in large quantity (Ac-GRLEEIDR-<sup>o</sup>Mes, called later peptide1-Arg-<sup>o</sup>Mes, corresponding to the eight C-terminal residues of [N-ACTR]). The catalytic hydrolysis and aminolysis of this shorter substrate in the presence of **Het-N2** occurred at similar enhanced rate with half-life of thioester around 1 hour, suggesting therefore that length of the thioester substrate does not influence the kinetics of the catalytic protein **Het-N2**.

Monitoring catalytic properties by HPLC and LC/MS allowed to get a first estimation of the catalytic efficiencies of the different analogues. The catalytic protein **Het-N2** undoubtedly furnishes the highest acceleration of hydrolysis and aminolysis of thioester substrate. Compared to the analogue **Het-Ncap**, this suggests that position of the catalytic cysteine is essential for hydrolysis presumably by providing better stabilization for the tetrahedral intermediates. The protein **Hom-N2** presents lower kinetic rate in comparison to **Het-N2** that might indicate some steric hindrance due to the two catalytic cysteines in close proximity. Nevertheless, this catalytic assay by HPLC monitoring was not convenient for our purposes mostly because we did not have access to absolute ratio of aminolysis over hydrolysis and the set up needed a lot of peptide material (for the substrates and the catalytic proteins), which is not compatible with the screening of conditions necessary to better characterize the analogue **Het-N2**. A next generation of catalytic assay has to be compatible with multiple time points and at the same time with a more precise monitoring of product formation.

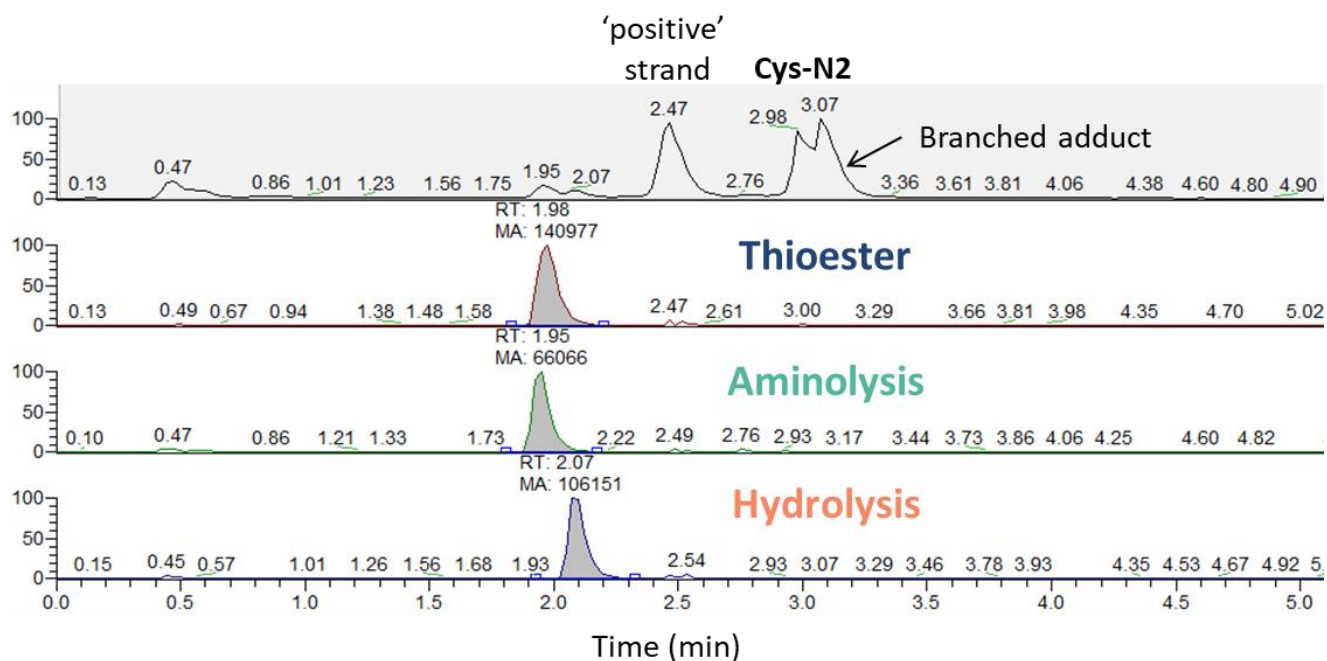
**Table 2.2. Kinetic parameters for aminolysis of [N-ACTR]-<sup>o</sup>Mes with Tris and concomitant hydrolysis in presence or in absence of protein catalysts.** Comparison of the kinetic parameters between the analogues **Het-N2**, **Het-Ncap** and **Hom-N2**.

Conditions / catalytic protein	$\tau_{1/2}$ (hours)	$k_{\text{obs}}$ ( $\text{M}^{-1} \cdot \text{s}^{-1}$ )
Background reaction (without catalyst)	77.5	0.007
<b>Het-N2</b>	0.75	0.74
<b>Het-Ncap</b>	10	0.056
<b>Hom-N2</b>	8	0.069

### 2.6.4. Catalytic assays followed by LC/MS – 2<sup>nd</sup> generation assay

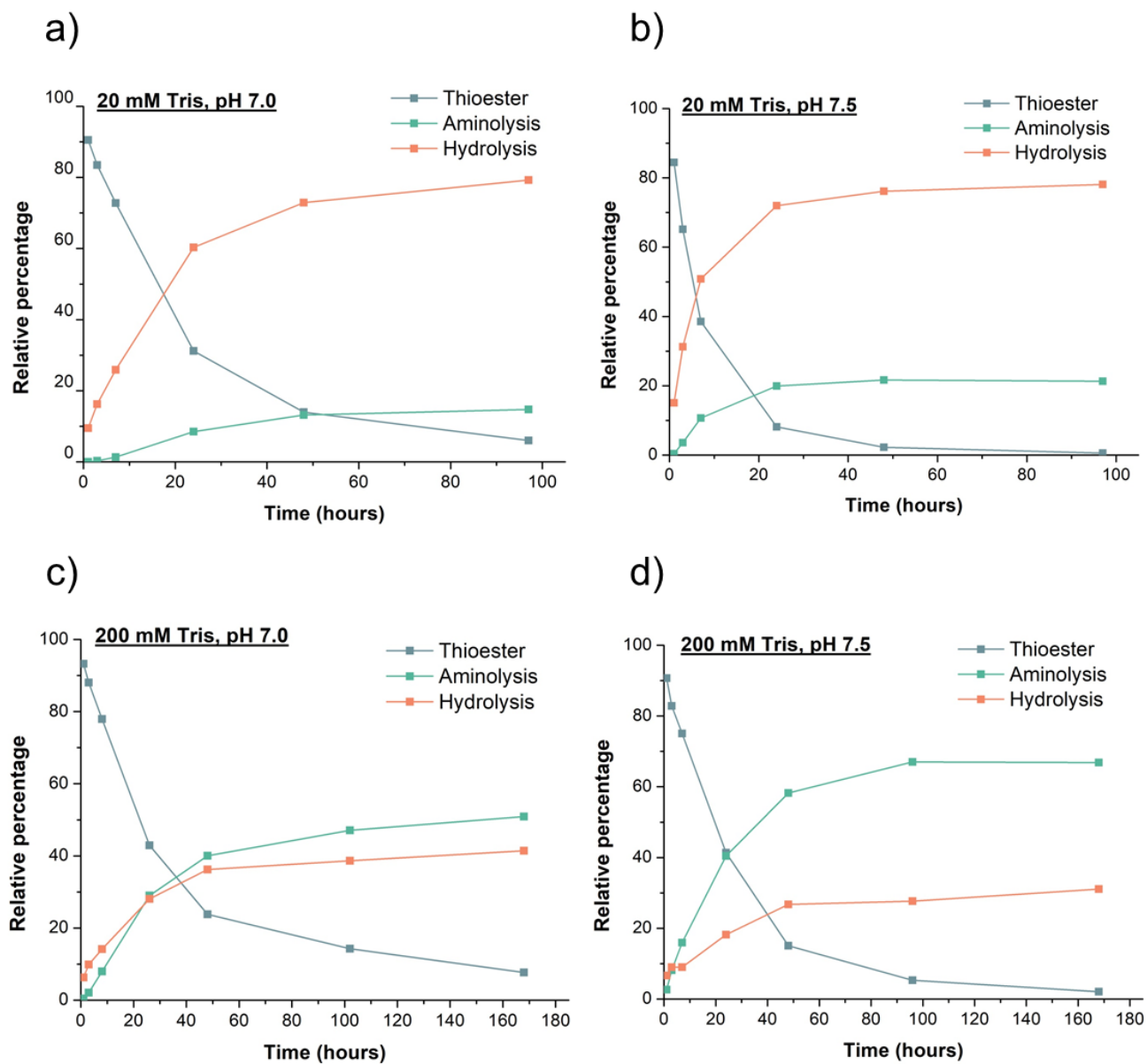
Relative changes of concentrations in time can actually be monitored by acquiring liquid chromatography electrospray ionization mass-spectra (LC/ESI-MS) and integrating the corresponding detected ions. LC/MS is a very sensitive technique that enables to generally work at lower concentrations compared to analytical HPLC. With optimized LC method, compounds of the reaction mixture can be separated allowing for a better separation of the detected ions facilitating the integration and the interpretation. The **Figure 2.24** illustrates the analysis of an aliquot of a catalytic assay with this new method, in which ions for the peptide-<sup>α</sup>thioester, the product of aminolysis and the product of hydrolysis were detected and integrated (ions for the ‘positive’ and **Cys-N2** strands of **Het-N2** as well as the branched adduct are also detected). Although quantitative results can be obtained via calibration curves of the detected compounds, we decided to extract relative percentage of detected and integrated substrate and product derivatives, which already provides substantial information and is valuable to evaluate different conditions over the time.

With this method of analysis, catalytic assays to evaluate the activity of **Het-N2** were performed using 100  $\mu$ M of peptide catalyst or control and peptide-<sup>α</sup>Mes thioester substrates (c 200  $\mu$ M) as an acyl donor in 50 mM sodium phosphate buffer with TCEP. Concentrations were reduced compared to assay of the 1<sup>st</sup> generation in order to maximize the number of assays. To allow screening of multiple conditions in parallel, reactions were let at room



**Figure 2.24.** LC/MS monitoring of an aliquot of a catalytic assay in the presence of **Het-N2**. Ions for the thioester substrate, product of aminolysis and product of hydrolysis can be detected and integrated. Relative concentrations are then deduced. Ions corresponding to **Het-N2** and the resulting branched adduct are also observed.

temperature in a non-stirred vial. To have substantial amount of aminolysis compared to hydrolysis of the thioester substrate, concentration of acyl acceptor, here Tris as a model amine, and pH of the buffer were optimized. Catalytic assays with 20 mM of Tris as acyl acceptor at pH 7.0 or 7.5 were first tested in the presence of peptide<sup>1</sup>-Arg-<sup>α</sup>thioester (see sequence in **Table 2.4**) and the **Het-N2** catalytic protein. In these conditions, hydrolysis was much more favored (**Figure 2.25a-b**). At pH 7.5, rate of thioester consumption was globally faster and aminolysis was a bit higher than at pH 7.0 (21 % of aminolysis against 15 %, respectively). To enhance aminolysis, the concentration of acyl acceptor was increased to 200 mM. Again, the catalytic assay was compared at pH 7.0 and 7.5. At such concentration of Tris, aminolysis was detected in majority compared to the product of hydrolysis (see **Figure 2.25c-d**). Additionally, pH 7.5 provided higher rate of thioester consumption and larger amount of aminolysis: at pH 7.0,  $k_{\text{obs}} = 0.062 \text{ M}^{-1} \cdot \text{s}^{-1}$  and ratio aminolysis/hydrolysis = 1.23, whereas at pH 7.5,  $k_{\text{obs}} = 0.143 \text{ M}^{-1} \cdot \text{s}^{-1}$  and ratio aminolysis/hydrolysis = 1.89. Therefore, we established that the best conditions for our catalytic assays to follow by LC/MS and to be used to further characterize **Het-N2** are: 200  $\mu\text{M}$  of peptide-<sup>α</sup>Mes, 200 mM of acyl acceptor and 100  $\mu\text{M}$  of **Het-N2** or other protein catalysts in phosphate buffer at pH 7.5.

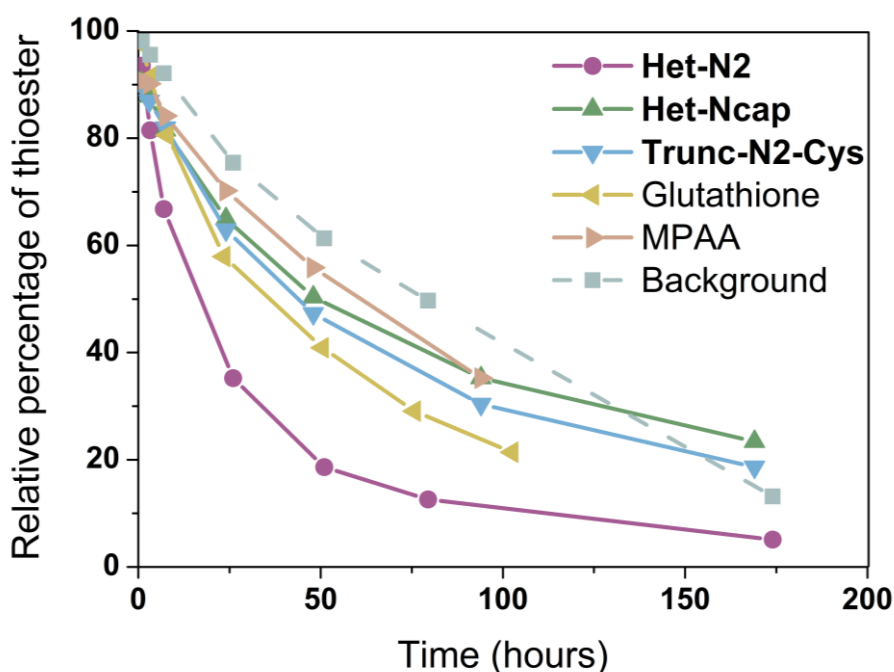


**Figure 2.25. Optimization of the conditions for the catalytic assay.** Concentration of acyl acceptor, here Tris, and pH were tested. Conditions: a) 20 mM Tris, pH 7.0; b) 20 mM Tris, pH 7.5; c) 200 mM Tris, pH 7.0; d) 200 mM Tris, pH 7.5. Last condition was selected for further catalytic characterization.

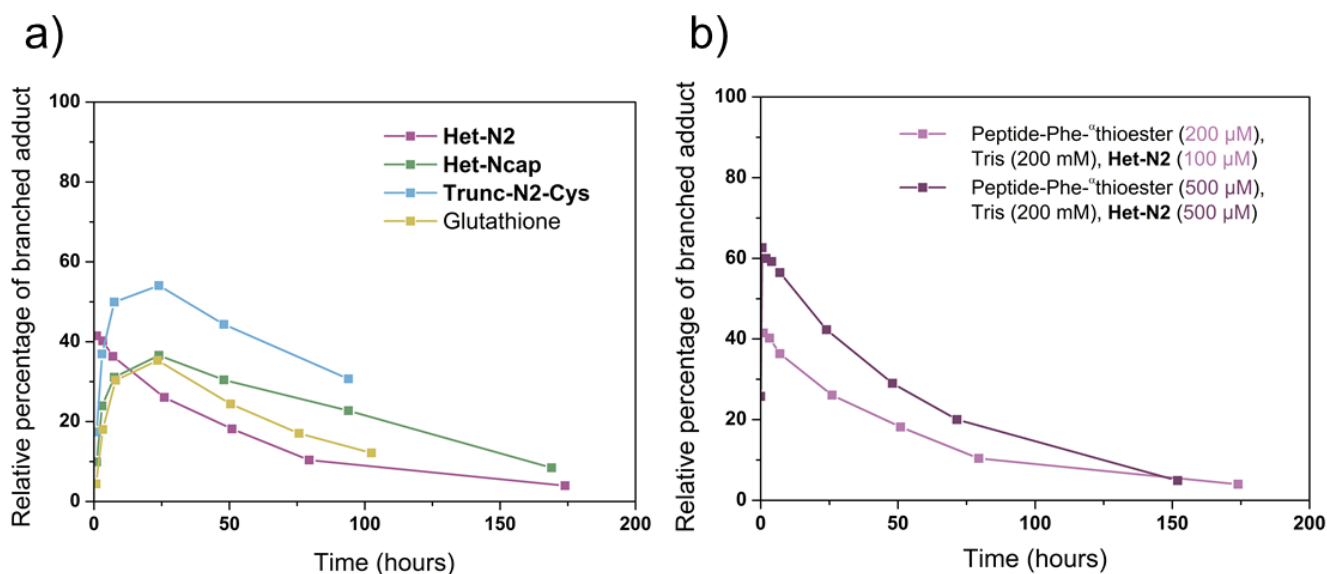
## 2.7. Characterization of the acyl transferase catalytic activity

To evaluate the acyl-transferase catalytic activity of the most active variant, **Het-N2**, conditions of the 2<sup>nd</sup> generation assay were employed. Controls were selected to further validate the essential parameters influencing the catalysis of **Het-N2** such as the importance of catalytic cysteine localization with **Het-Ncap**, the importance of the full-length scaffold with the truncated analogue **trunc-Cys-N2** as well as with other thiols like glutathione and MPAA (4-mercaptophenylacetic acid). We focused our study on heterodimer analogues that are more interesting for further functionalization and optimization (*i.e.* the ‘positive’ strand can also serve for installing extra catalytic or binding residues or general acid/base) and that possesses more efficient catalytic activity (*i.e.* **Hom-N2** carried lower catalytic activity according to assays of the 1<sup>st</sup> generation that may be due to steric hindrance of the two catalytic cysteines in close proximity). Catalytic assays were acquired via the 2<sup>nd</sup> generation assay (*i.e.* 200  $\mu$ M of peptide- $\alpha$ Mes, 200 mM of acyl acceptor with Tris as the standard acyl acceptor and 100  $\mu$ M of **Het-N2** or other protein catalysts in phosphate buffer at pH 7.5) and relative changes of concentrations in time were monitored by acquiring liquid chromatography electrospray ionization mass-spectra (LC/ESI-MS) and integrating the corresponding detected ions. Background reaction (without protein catalyst) was also recorded for most of the conditions.

As an example, **Figure 2.26** shows kinetics of consumption of peptide-Phe- $\alpha$ thioester



**Figure 2.26.** Peptide-Phe- $\alpha$ thioester consumption in the presence of **Het-N2**, **Het-Ncap**, **trunc-N2-Cys**, glutathione and MPAA (4-mercaptophenylacetic acid) in comparison to background reaction (without catalyst).



**Figure 2.27. Branched adduct evolution.** **a)** Changes in concentration of peptide-protein branched thioester resulting from the transthoesterification between peptide-Phe- $\alpha$ -thioester and **Het-N2**, **Het-Ncap**, **trunc-Cys-N2** or glutathione. **b)** Changes in concentration of peptide-protein thioester at different initial concentrations of catalyst and thioester. At higher concentrations of **Het-N2** and peptide-Phe- $\alpha$ -thioester, the branched adduct is formed in larger amounts as expected for bimolecular reaction.

substrate in the presence of different catalysts (c 100  $\mu\text{M}$ ), and **Figure 2.27a** illustrates the evolution of branched adduct resulting from the transthoesterification between the substrate-thioester and the catalytic cysteine. In the presence of **Het-N2**, peptide-Phe- $\alpha$ -thioester transthoesterified most rapidly forming peptide-catalytic protein branched thioester adduct accompanied by the release of Mes (41 % of branched adduct formed after 1 hour of reaction in the presence of **Het-N2**, whereas 10 % and 17 % were observed for **Het-Ncap** and **trunc-Cys-N2**, respectively). The percentage of branched adduct were higher upon increasing the concentration of **Het-N2** (from 200  $\mu\text{M}$  to 500  $\mu\text{M}$ , see **Figure 2.27b**) as expected for this bimolecular step. Accumulation of the covalently bound Cys-thioester intermediate supports a ping-pong mechanism as expected in the hypothesized catalytic mechanism outlined in **Scheme 2.1**. For comparison, the same reaction performed with **Het-Ncap**, **trunc-Cys-N2** and Cys-containing glutathione proceeded much slower (**Figure 2.25a**), and when MPAA was added at c 100  $\mu\text{M}$  the product of thioester exchange was not observed. It should be mentioned that MPAA is a common aryl-thiol catalyst for native chemical ligation (NCL), however, used at much higher concentrations (10-200 mM).<sup>42</sup> The subsequent rate-limiting step involved aminolysis by Tris that occurred in competition with hydrolysis. Thus, the ratio of aminolysis over hydrolysis (A/H) and the observed rate constant ( $k_{\text{obs}}$ ) calculated from the depletion of initial peptide- $\alpha$ -thioester substrate concentration were used as quantifiers to characterize and compare reactivity of the catalysts in different conditions (**Table 2.3**).

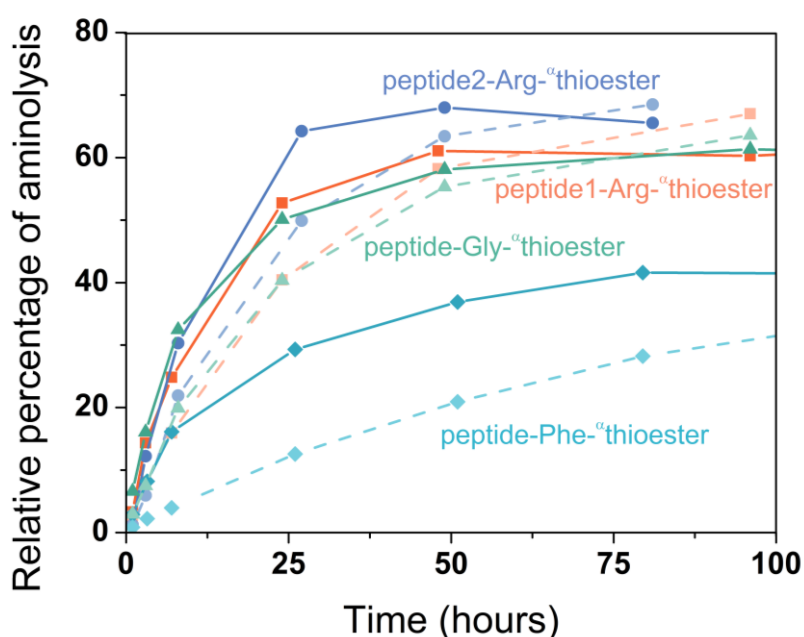
For **Het-N2** construct showing appreciable catalytic effects in **Figure 2.26**, the  $k_{\text{obs}}$  is 0.082  $\text{M}^{-1}\cdot\text{s}^{-1}$ , which is four times higher than the background reaction without catalyst ( $k_{\text{obs}}$

**Table 2.3.** Comparison of catalytic activity between the DSD-analogue Het-N2 and controls (Het-Ncap, trunc-Cys-N2, glutathione and MPAA). Conditions: 100  $\mu\text{M}$  catalyst, 200  $\mu\text{M}$  peptide-Phe- $\alpha$ Mes, 200 mM Tris, pH 7.5.

Catalyst	$k_{\text{obs}}$ ( $\text{M}^{-1}\cdot\text{s}^{-1}$ )	Aminolysis/Hydrolysis ratio	Acceleration compared to background reaction
Het-N2	0.082	0.76	3.9
Het-Ncap	0.028	1.03	1.4
trunc-Cys-N2	0.032	1.40	1.5
Glutathione	0.039	0.66	1.5
MPAA	0.023	0.97	0.9

0.021  $\text{M}^{-1}\cdot\text{s}^{-1}$ ). Although aminolysis was slightly more favored with other catalysts, they influenced the rate of the reaction less significantly with a maximum acceleration of 1.5 times over background including the **Het-Ncap** protein (Figure 2.26 and Table 2.3), highlighting the importance of the N2 position of the catalytic Cys for rate acceleration.

In addition to peptide-Phe- $\alpha$ thioester, other thioester substrates with different C-terminal amino acids (Arg, Gly) and sequences were assayed (see Table 2.4). Both Arg and Gly thioesters were found to undergo faster reactions in the presence and absence of **Het-N2** than peptide-Phe- $\alpha$ thioester (Table 2.4 and Figure 2.28), which is most likely caused by intramolecular catalytic effect of Arg side chain and favorable steric effect due to the absence of side chain in Gly. However, in comparison to peptide-Phe- $\alpha$ thioester, the observed rate constant  $k_{\text{obs}}$  for Arg and Gly thioesters was only two times faster compared to the respective background reaction (see Table 2.4). Another difference is the A/H ratio, which for peptide-Phe- $\alpha$ thioester was found to be 0.76 indicating hydrolysis to be a major outcome, whereas for Arg and Gly thioester substrates A/H was in the range of 1.7-2, which corresponds to a more pronounced aminolysis. Two distinct sequences that were studied for Arg thioester showed



**Figure 2.28.** Comparison of aminolysis product formation by reaction of Tris on different thioester substrates in presence (solid lines) or in absence (dashed lines) of Het-N2.

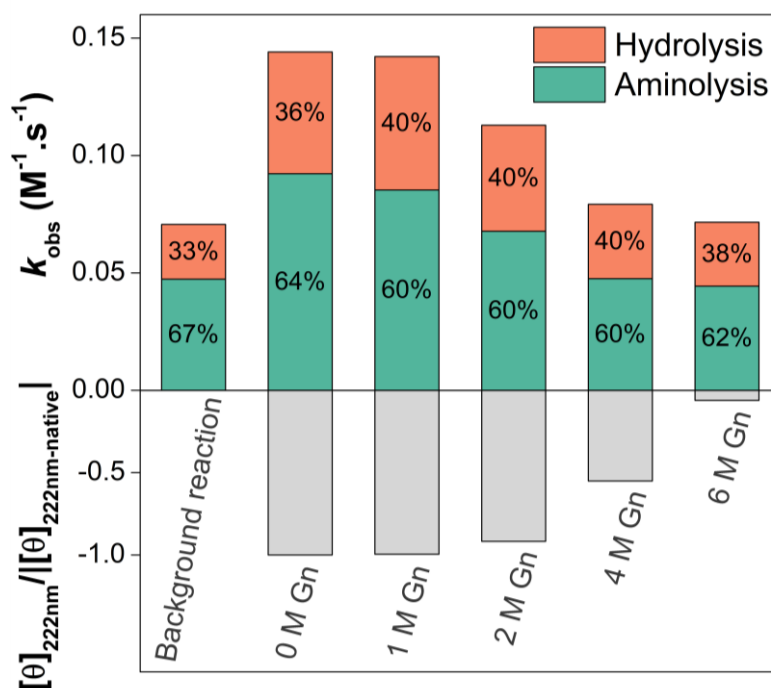


very similar results (Table 2.4). Thus, we can conclude that it is the C-terminal amino acid on the acyl donor that plays a decisive role in influencing the catalytic reaction and its outcome with **Het-N2** catalyst.

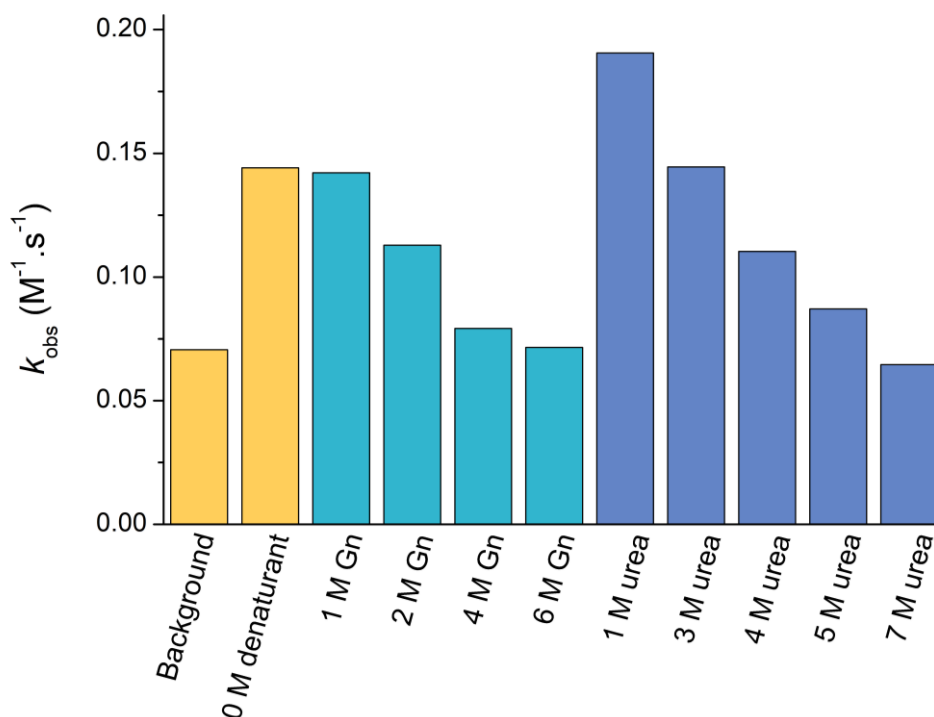
**Table 2.4. Kinetic parameters of acyl transfer reaction with various thioester substrates in the presence of Het-N2.** Conditions: 100  $\mu\text{M}$  **Het-N2**, 200  $\mu\text{M}$  of peptide- $\alpha$ thioester, 200 mM Tris, pH 7.5.

Thioester	Sequence	$k_{\text{obs}}$ ( $\text{M}^{-1} \cdot \text{s}^{-1}$ )	Aminolysis/Hydrolysis ratio	Acceleration compared to background reaction
peptide1-Arg	Ac-GRLEEIDR- $\alpha$ Mes	0.143	1.89	2
peptide2-Arg	Ac-VALENR- $\alpha$ Mes	0.143	2.03	1.8
peptide-Phe	Ac-VALENF- $\alpha$ Mes	0.082	0.76	3.9
peptide-Gly	Ac-LYRAG- $\alpha$ Mes	0.173	1.66	2.4

The fact that the 12-residue **trunc-Cys-N2** is less efficient than **Het-N2** (Figure 2.26 and Table 2.3) suggests that the overall stable structure of DSD is required for stabilization of oxyanion at N-terminus of  $\alpha$ -helix. Using peptide1-Arg- $\alpha$ thioester and Tris as acyl acceptor we further confirmed the significance of the folded structure of **Het-N2** for the catalytic activity by increasing the concentration of a denaturant in the reaction mixture (Figure 2.29) while keeping an apparent pH of 7.5. The  $k_{\text{obs}}$  was found to correlate to the degree of folding based



**Figure 2.29. Observed rate constant and outcome of the reaction between peptide1-Arg- $\alpha$ thioester substrate and Tris in presence of Het-N2 as a function of the concentration of denaturant (Gn-HCl) and the degree of folding of Het-N2 based on circular dichroism measurements.** At high concentrations of Gn-HCl, Het-N2 was progressively more unfolded corresponding to a decrease of the reaction rate down to a background level.



**Figure 2.30.** Observed rate constant  $k_{\text{obs}}$  depending on the concentration of denaturant (guanidine or urea). Conditions: 100  $\mu\text{M}$  of **Het-N2**, 200  $\mu\text{M}$  of peptide1-Arg- $\alpha$ thioester and 200 mM of Tris, pH 7.5.

on circular dichroism (CD) data at different guanidine hydrochloride (Gn-HCl) concentrations: when the protein **Het-N2** was highly  $\alpha$ -helical the catalytic activity was at maximum, whereas at high concentration (6 M) of Gn-HCl provoking unfolding of the protein the rate decreased to a background level. Same effect on the catalytic activity of **Het-N2** was also observed upon increasing the concentration of urea (**Figure 2.30**). The clear loss of catalytic activity upon protein unfolding demonstrated the importance of the fully folded DSD scaffold that may provide a distinct environment for tetrahedral intermediates stabilization.

The pH of the buffer also modulated the rate of reaction and A/H ratio. Upon increasing the pH, the rate of peptide1-Arg- $\alpha$ thioester consumption in the presence of **Het-N2** increased with an inflection point at 8.6, which was approximately similar for the reaction without protein catalyst (**Figure 2.31**). Maximum of aminolysis product was observed at pH 8 with a relative final product percentage of 66%. Below pH 5 (pH 6 without catalyst), thioester stayed very stable during time and above pH 10 hydrolysis was predominant with reaction almost as fast as background. Additionally, to rule out any effect of the buffer on catalysis, we performed assays with different concentrations of phosphate (*i.e.* 50, 200, 500 mM) and in two other buffer systems (50 mM MES or 50 mM MOPS) at pH 7.5. The catalytic activity of **Het-N2** remained similar in different buffer conditions with a slight increase of hydrolysis at higher concentration of phosphate (500 mM) as shown in **Figure 2.32**.

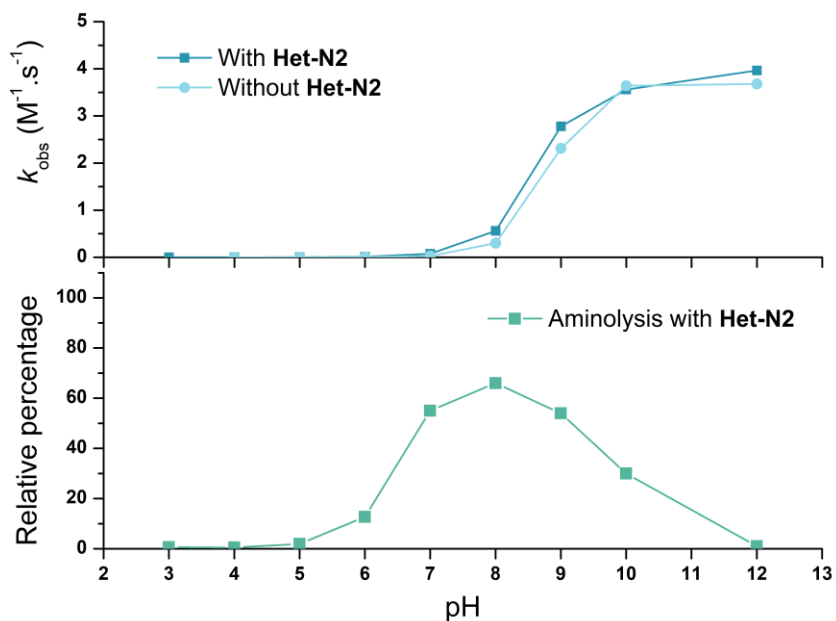


Figure 2.31. Observed rate constant  $k_{\text{obs}}$  and relative percentage of aminolysis at different pHs. Conditions: peptide1-Arg-<sup>o</sup>thioester (200  $\mu\text{M}$ ) and Tris (200 mM) in the presence or absence of Het-N2 (100  $\mu\text{M}$ ).

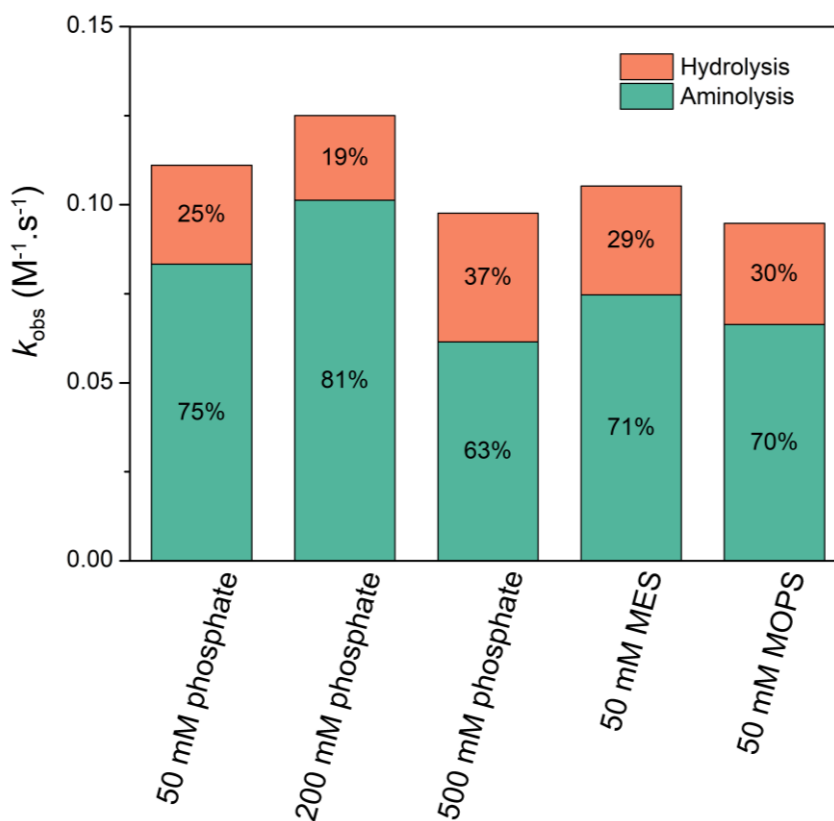
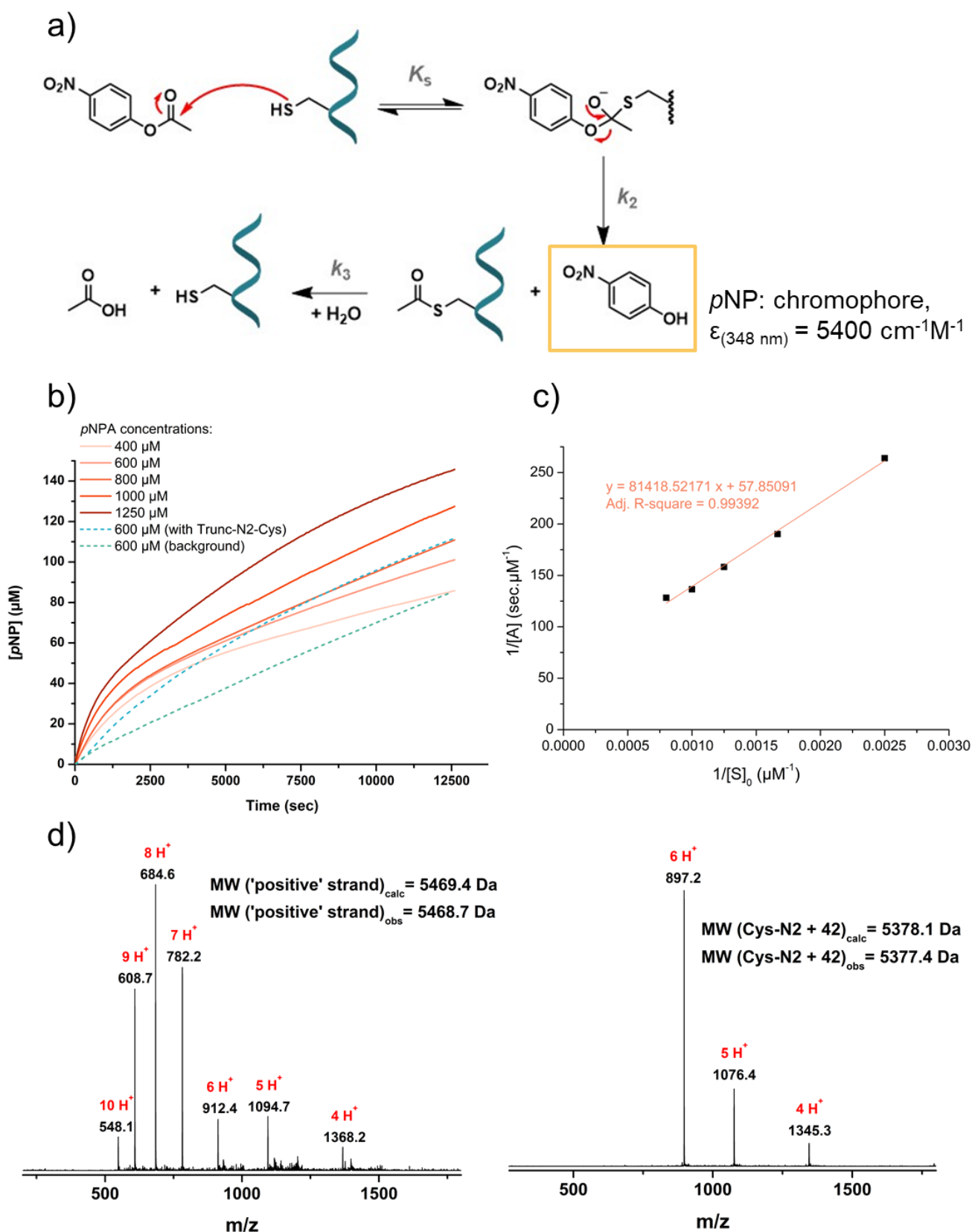


Figure 2.32. Catalytic activity of Het-N2 for the reaction between peptide1-Arg-<sup>o</sup>Mes and Tris in different buffers (50 mM, 200 mM or 500 mM sodium phosphate; 50 mM MES; 50 mM MOPS). Catalytic activity of Het-N2 remained similar, meaning that the buffer did not influence the kinetics.

Finally, in order to complete the general catalytic characterization of **Het-N2** for acyl transfer reactions and since the hydrolysis of thioester substrates was prominent, its hydrolytic activity was evaluated in more detail using a continuous assay with *para*-nitrophenyl acetate (*p*NPA) as substrate (**Figure 2.33a**).<sup>44</sup> Observed kinetic curves showed a burst phase and a linear phase (**Figure 2.33b**), and when the reaction mixture was analyzed by LC-MS, the mass of the **Cys-N2** peptide with a mass shift of +42 Da corresponding to acetyl thioester intermediate was detected (**Figure 2.33d**). These findings further support the ping-pong mechanism of the catalyzed hydrolysis of *p*NPA by **Het-N2** protein through the formation of a branched adduct.<sup>44</sup> Interestingly, the truncated variant **trunc-N2-Cys** did not present such a pronounced burst phase suggesting a less efficient hydrolytic activity (as shown in **Figure 2.33b**). By analyzing the kinetic equations for **Het-N2** at different [*p*NPA]<sub>0</sub> (from 400 μM to 1250 μM), the double-reciprocal plot was deduced (**Figure 2.33c**) allowing to determine  $k_{\text{cat}}$  of  $3.46 \times 10^{-4} \text{ sec}^{-1}$  and  $K_{\text{m}}$  of 1.4 mM, corresponding to catalytic efficiency ( $k_{\text{cat}}/K_{\text{m}}$ ) of  $0.25 \text{ M}^{-1}\text{sec}^{-1}$ . Compared to previously reported *de novo* hydrolases, the catalytic efficiency of **Het-N2** is nearly 10 times lower, however, no additional catalytic residues (e.g. histidine) are present in **Het-N2**, thus the catalytic effect emerges mainly from the interactions with N-terminus of  $\alpha$ -helix.<sup>16,45</sup>

Thanks to rational and minimalist design, a catalytic protein was produced. The resulting **Het-N2** protein shows substantial catalytic activity for acyl transfer reaction and is notably capable of accelerating aminolysis and hydrolysis of various peptide- $\alpha$ thioesters as well as hydrolysis of activated ester such as *p*NPA. General catalytic characterization emphasized the importance of the catalytic cysteine position and the full-length and folded scaffold for the efficiency of the function. The next subsection will demonstrate the catalytic promiscuity of **Het-N2**.



## 2.8. Scope of reactivity for acyl acceptor molecules and catalytic promiscuity

To test the ability of **Het-N2** to catalyze acyl transfer reactions other than aminolysis with Tris, we carried out the catalytic assays with several other acyl acceptors. Reactions of miscellaneous amines and few alcohols with peptide1-Arg- $\alpha$ -thioester were studied (**Table 2.5**). Amines with lower  $pK_a$  (i.e. methoxyamine with  $pK_a$  of 4.7) provided with the fastest kinetics ( $k_{obs}$   $0.412 \text{ M}^{-1}\cdot\text{s}^{-1}$ ) that represents more than three times acceleration over background, and aminolysis was slightly more favored ( $A/H = 2.33$ ) in comparison to the catalyzed reaction with Tris. In this particular case, the rate enhancement is likely resulting from  $\alpha$ -effect well-known for methoxyamine, where nucleophilicity of amine is increased due to the adjacent oxygen atom containing lone electron pairs.<sup>46</sup> In contrast, substrates with higher  $pK_a$  like methylamine (10.6) resulted in the excess of hydrolyzed product ( $A/H = 0.37$ ) even if the rate of thioester consumption ( $k_{obs}$   $0.187 \text{ M}^{-1}\cdot\text{s}^{-1}$ ) was similar to the reaction with Tris. We also investigated catalysis of peptide bond formation by **Het-N2** using glycine ( $pK_a = 9.6$ ), a dipeptide Gly-Gly ( $pK_a = 8.2$ ) and a tripeptide Gly-Gly-Gly ( $pK_a = 7.8$ )<sup>47</sup> as acyl

**Table 2.5. Kinetics parameters for various acyl transfer reactions on peptide1-Arg- $\alpha$ -thioester (200  $\mu\text{M}$ ) catalyzed by Het-N2 (100  $\mu\text{M}$ ).**

Aminolysis vs Hydrolysis (A vs H)				
Acyl acceptor (200mM)	$pK_a$	$k_{obs} (\text{M}^{-1}\cdot\text{s}^{-1})$	A/H ratio	Acceleration over background reaction
Tris	8.1	0.143	1.89	2
MeONH <sub>2</sub>	4.7	0.412	2.33	3.2
MeNH <sub>2</sub>	10.6	0.187	0.37	3.1
Gly	9.6	0.108	0.09	2.6
Gly-Gly	8.2	0.162	0.15	2.9
Gly-Gly-Gly	7.8	0.177	0.29	2.7
Histamine	9.8	0.693	0.48	1.2
Esterification vs hydrolysis (E vs H)				
Acyl acceptor (200 mM)	$pK_a$	$k_{obs} (\text{M}^{-1}\cdot\text{s}^{-1})$	E/H ratio	Acceleration over background reaction
Trifluoroethanol	12.4	0.388	Max: 1.83	4.4
MeOH	15.2	0.157	0.2	3
EtOH	15.9	No reaction	-	-

acceptors. Glycine was less reactive and more prone to hydrolysis than the dipeptide ( $A/H = 0.09$  and  $0.15$ , respectively) and the tripeptide ( $A/H = 0.29$ ), illustrating an influence of the  $pK_a$  of acyl acceptor on the reactivity (**Table 2.5**). Recurrently, the catalytic protein **Het-N2** affords up to three-fold acceleration for acyl transfer reaction on peptide- $\alpha$ thioester, independently of the different acyl acceptors and without perturbing the ratio aminolysis/hydrolysis compared to the respective background reactions.

A peculiar change of reactivity was noticed for histamine ( $pK_a$  (amine) =  $9.8$  and  $pK_a$  (imidazole) =  $6.04$ ). A rather high observed rate constant was detected ( $k_{obs}$  of  $0.693 \text{ M}^{-1}\cdot\text{s}^{-1}$ ), however, in the background reaction without **Het-N2** the consumption of thioester was almost as fast and aminolysis was more pronounced compared to the reaction with **Het-N2** ( $A/H$  for uncatalyzed reaction is  $0.85$  versus  $0.48$  for catalyzed). This may indicate an additional catalytic effect of the imidazole ring on the initial thioester that is more accessible in solution than the Cys-thioester intermediate covalently bound to **Het-N2**. Influence of imidazole used as separate additive (*i.e.* in synergy with the catalytic protein) will be presented hereinafter.

Alcohols as acyl acceptors demonstrated moderate reactivity (**Table 2.5**). Esterification with trifluoroethanol, an alcohol with a reduced  $pK_a$  of  $12.4$ , was observed in the presence of catalytic protein **Het-N2** with  $k_{obs} = 0.388 \text{ M}^{-1}\cdot\text{s}^{-1}$  and up to  $61\%$  of detected ester formation compared to  $k_{obs} = 0.087 \text{ M}^{-1}\cdot\text{s}^{-1}$  and  $39\%$  of ester for background reaction, but the resulting ester product was not stable and decomposed through hydrolysis after reaching a maximum of ester production. Methanol ( $pK_a = 15.2$ ) reached a conversion of  $17\%$  of ester product with a three-fold faster rate compared to the uncatalyzed reaction. Interestingly, ethanol ( $pK_a = 15.9$ ) did not react in the acyl transfer with peptide1-Arg- $\alpha$ thioester.

Furthermore, we demonstrated that addition of azide and cyanide was able to modify kinetics of the acyl transfer reaction between peptide1-Arg- $\alpha$ thioester and Tris. The increasing concentrations of these additives with and without catalytic proteins led to higher observed rates of reactions and higher percentage of hydrolysis (see **Table 2.6** and **Figure 2.34**). The most contrasting effect was discovered in the presence of azide, where the high concentration of  $\text{NaN}_3$  ( $154 \text{ mM}$ ) in synergy with **Het-N2** contributed to sixteen-fold acceleration compared to background reaction (without additives and catalyst) and an eight-fold acceleration compared to conditions without additives and with catalyst. However, the  $A/H$  ratio dropped from  $1.89$  to  $0.1$ . Similar tendency was observed with  $\text{KCN}$  in the reaction mixture (**Figure 2.34**). Presumably, transient acyl azides and acyl cyanides are the reactive intermediates that facilitate hydrolysis,<sup>48,49</sup> although, these intermediates were not directly detected by LC-MS.

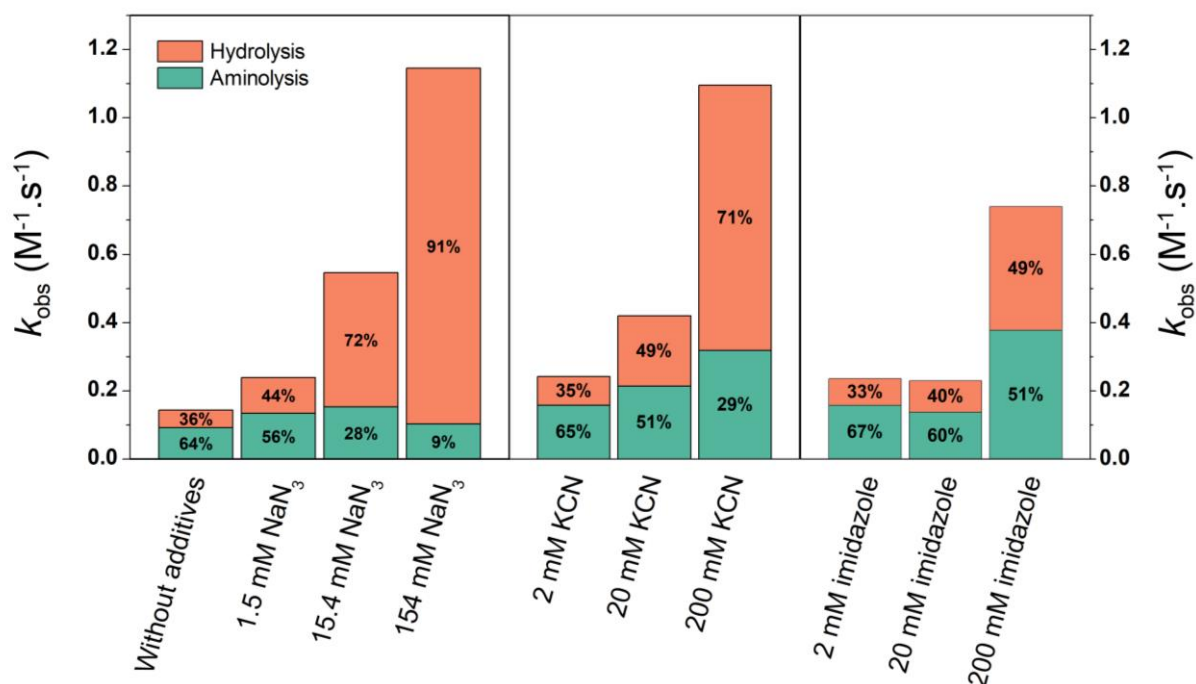
After observing the unusual reactivity of histamine as acyl acceptor, we evaluated the effect of imidazole as an additive. When imidazole was at low concentrations ( $2$  and  $20 \text{ mM}$ )

no considerable reactivity difference was observed, whereas at higher concentration (200 mM), the observed rate constant increased 5-fold to  $0.740 \text{ M}^{-1}\cdot\text{s}^{-1}$  with aminolysis being more favored in comparison to other additives with a final A/H ratio of 1.04 (see **Table 2.6** and **Figure 2.34**). Previously, imidazole added at high concentrations (2.5 M) was reported to catalyze native chemical ligation.<sup>50</sup>

In conclusion, the designed **Het-N2** protein reasonably accelerates various acyl transfer reactions with different peptide- $\alpha$ thioesters as well as different acyl acceptors. A lot of effort is still needed to suppress the side reaction of hydrolysis; nevertheless **Het-N2** does not affect the aminolysis over hydrolysis ratio compared to the background reaction without catalyst. The catalyst **Het-N2** influences only the rate of all tested reactions with acceleration between 2-fold and 4-fold as determined by the 2<sup>nd</sup> generation assay.

**Table 2.6. Further acceleration in presence of additives and Het-N2.** Additives were used for reaction between Tris (200 mM) and peptide1-Arg- $\alpha$ thioester (200  $\mu\text{M}$ ) in the presence of **Het-N2** (100  $\mu\text{M}$ ). Listed accelerations correspond to the comparison with the reaction without additives and with catalyst, and in brackets with the reaction without additives and catalyst.

Additives	$k_{\text{obs}} (\text{M}^{-1}\cdot\text{s}^{-1})$	A/H ratio	Acceleration
$\text{NaN}_3$ (154 mM)	1.146	0.1	8.0 (16.2)
KCN (200 mM)	1.095	0.41	7.7 (15.5)
Imidazole (200 mM)	0.74	1.04	5.2 (10.5)



**Figure 2.34. Modulation of the kinetics of  $\text{NaN}_3$ , KCN or imidazole in comparison to the kinetics without additives.** Condition: peptide1-Arg- $\alpha$ thioester (200  $\mu\text{M}$ ), Tris (200 mM) and **Het-N2** (100  $\mu\text{M}$ ).



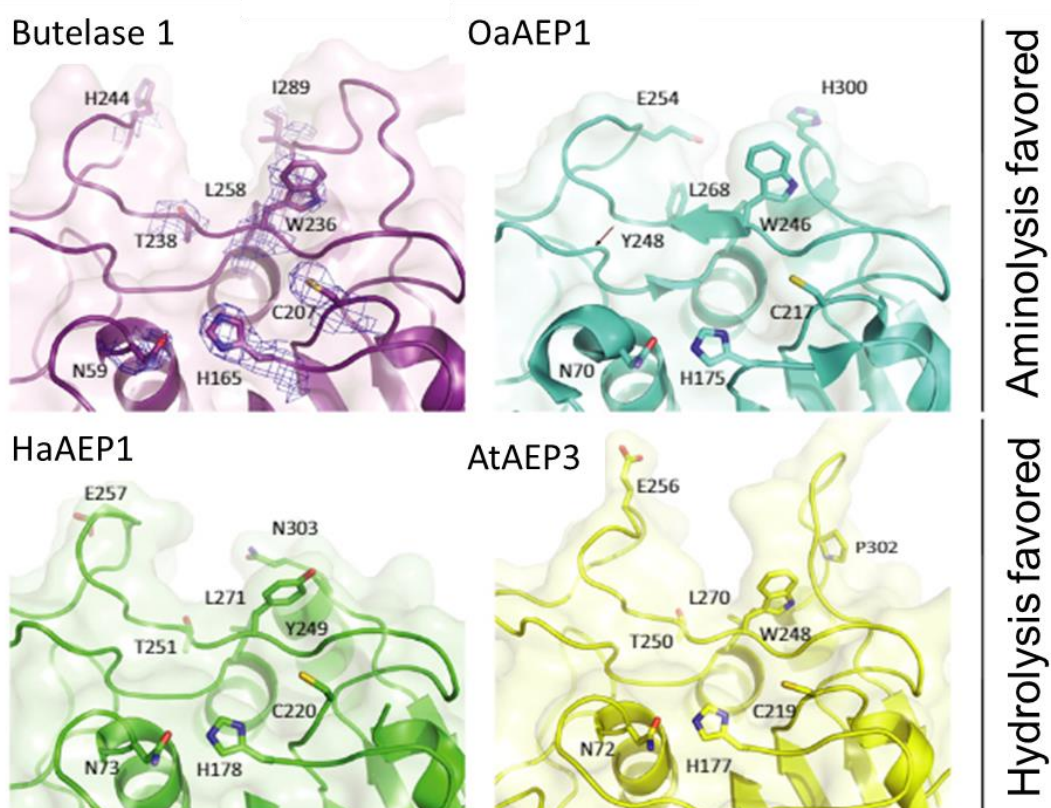
## 2.9. Conclusions

To conclude, we demonstrated that a completely artificial protein scaffold enables to enhance reactivity of peptide- $\alpha$ thioesters to form an amide bond through the formation of a covalently bound Cys-thioester intermediate. By rational design, we introduced a single cysteine residue at N-terminal positions in a *de novo* protein scaffold resulting in **Het-N2**, **Het-Ncap** and also **Hom-N2** analogues. X-ray structure unambiguously confirmed their assembly into heterodimeric structures for **Het-N2** and **Het-Ncap** and the homodimeric structure for **Hom-N2** that are all similar to the original DSD protein scaffold. We demonstrated the importance of the environment for catalytic Cys and the relationship between folding of the *de novo* structure and the catalytic activity. Indeed, the **Het-N2** construct displayed the highest catalytic efficiency, demonstrating the importance of the position of the catalytic cysteine (compared to **Het-Ncap**), the fully folded three-dimensional structure (compared to a truncated analogue, **trunc-Cys-N2**) and an accessible catalytic site (compared to **Hom-N2** that possesses two catalytic cysteines but a lower catalytic activity probably due to steric clashes between the two sites). In addition, the designed **Het-N2** was found to modestly accelerate a variety of acyl transfer reactions with a wide range of substrates. Thus, **Het-N2** can be compared to primitive (or primordial) enzymes possessing moderate catalytic efficiency and substrate promiscuity.<sup>51</sup> Such modest catalytic efficiency coupled to catalytic promiscuity is believed to be essential for evolution of enzymes with distinct specificities. The next step is to apply computational design and combinatorial library approaches taking advantage of the modular architecture of **Het-N2** scaffold to vary the composition at the active site and improve its catalytic properties.

Enabling selective catalysis of aminolysis versus hydrolysis by design is particularly challenging aspect. Recently, several X-ray structures were reported for natural proteolytic enzymes which also possess peptide macrocyclization or ligation activities and provide insights into how such selectivity can be achieved.<sup>52-57</sup> The principal catalytic residue is either Ser or Cys and catalysis is assisted by few other residues such as His, Asp/Glu, Tyr and backbone amides. The enhancement of ligation activity over proteolysis is believed to occur by exclusion of water from the acyl-enzyme covalent adduct via additional structural elements such as “capping” lid domain,<sup>52</sup> binding of follower peptide<sup>53</sup> or hydrophobicity of the active site.<sup>54-57</sup> Remarkably, structural comparison of the enzymes with ligation (aminolysis) activity with the homologous variants possessing enhanced proteolytic (hydrolysis) activity showed that the structures of the related ligases and proteases are very similar with only subtle changes observed in the catalytic pocket. Furthermore, minimal

changes in active site composition are sufficient to repurpose a ligase into a protease and vice versa (see **Figure 2.35**).

Thus, it is not surprising that the branched thioester adduct observed in our work where electrophilic thioester moiety is exposed to aqueous buffer undergoes fast hydrolysis in competition with aminolysis. The choice to use Cys and not Ser as the principle nucleophilic residue is not sufficient to bias the reactivity towards the reaction with amines, despite the reported higher relative reaction rates of thioesters towards nitrogen nucleophiles than oxygen nucleophiles in comparison to oxoesters.<sup>20,58</sup> On the other hand, our data indicate that the environment nearby catalytic Cys can slightly modify the aminolysis versus hydrolysis ratio. This result can be explained by the small alterations of the conformation of thioester moiety and the corresponding tetrahedral intermediates, resulting in changes of their stabilities and predisposition to react with either amines or water. Indeed, previous theoretical work indicates that the possible transition states for nucleophilic attack of thioesters by either nitrogen or oxygen nucleophiles may have distinct geometrical preferences.<sup>20</sup> Thus, catalytic proteins designed in this work may further be studied to explore the intricate features of thioesters' reactivity including the role of stereoelectronic effects thanks to the great advantage of *de novo* protein carrying modular scaffold. With the



**Figure 2.35. Catalytic site structures of different asparaginyl endopeptidases (AEPs).** High overall structural conservation is observed among the family of AEPs. Minimal changes in the catalytic site may be at the origin of differences in activity. The figure is reprinted with permission from Figure 1.c of James et al. (2019) *Plant J.*, 98, 988-999. Copyright (2020) John Wiley and Sons.

main goal to achieve higher catalytic activities, the future work will be directed towards understanding the structural requirements for aminolysis versus hydrolysis by optimizing the sequence for a particular reaction path and developing new general and practical tools for peptide ligation and protein modification.

## 2.10. References

- (1) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *J. Comb. Chem.* **1999**, *1* (1), 55–68.
- (2) Hasegawa, S.; Horike, S.; Matsuda, R.; Furukawa, S.; Mochizuki, K.; Kinoshita, Y.; Kitagawa, S. Three-Dimensional Porous Coordination Polymer Functionalized with Amide Groups Based on Tridentate Ligand: Selective Sorption and Catalysis. *J. Am. Chem. Soc.* **2007**, *129* (9), 2607–2614.
- (3) Pattabiraman, V. R.; Bode, J. W. Rethinking Amide Bond Synthesis. *Nature* **2011**, *480* (7378), 471–479.
- (4) 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.* **2007**, *46* (48), 9248–9252.
- (5) Hemantha, H. P.; Narendra, N.; Sureshbabu, V. V. Total Chemical Synthesis of Polypeptides and Proteins: Chemistry of Ligation Techniques and Beyond. *Tetrahedron* **2012**, *68* (47), 9491–9537.
- (6) Nadler, C.; Nadler, A.; Hansen, C.; Diederichsen, U. A Photocleavable Auxiliary for Extended Native Chemical Ligation. *Eur. J. Org. Chem.* **2015**, *2015* (14), 3095–3102.
- (7) Wu, Z.; Guo, X.; Guo, Z. Sortase A-Catalyzed Peptide Cyclization for the Synthesis of Macrocyclic Peptides and Glycopeptides. *Chem. Commun.* **2011**, *47* (32), 9218–9220.
- (8) Antos, J. M.; Truttmann, M. C.; Ploegh, H. L. Recent Advances in Sortase-Catalyzed Ligation Methodology. *Curr. Opin. Struct. Biol.* **2016**, *38*, 111–118.
- (9) Nguyen, G. K. T.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J. P. Butelase 1 Is an Asx-Specific Ligase Enabling Peptide Macrocyclization and Synthesis. *Nat. Chem. Biol.* **2014**, *10* (9), 732–738.
- (10) Nguyen, G. K. T.; Qiu, Y.; Cao, Y.; Hemu, X.; Liu, C.-F.; Tam, J. P. Butelase-Mediated Cyclization and Ligation of Peptides and Proteins. *Nat. Protoc.* **2016**, *11* (10), 1977–1988.
- (11) Schmidt, M.; Toplak, A.; Quaedflieg, P. J. L. M.; van Maarseveen, J. H.; Nuijens, T. Enzyme-Catalyzed Peptide Cyclization. *Drug Discov. Today Technol.* **2017**, *26*, 11–16.
- (12) Weeks, A. M.; Wells, J. A. Subtiligase-Catalyzed Peptide Ligation. *Chem. Rev.* **2019**.
- (13) Jackson, D. Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J. A. A Designed Peptide Ligase for Total Synthesis of Ribonuclease A with Unnatural Catalytic Residues. *Science* **1994**, *266* (5183), 243–247.
- (14) Toplak, A.; Nuijens, T.; Quaedflieg, P. J. L. M.; Wu, B.; Janssen, D. B. Peptiligase, an Enzyme for Efficient Chemoenzymatic Peptide Synthesis and Cyclization in Water. *Adv. Synth. Catal.* **2016**, *358* (13), 2140–2147.
- (15) Schmidt, M.; Toplak, A.; Quaedflieg, P. J. L. M.; Ippel, H.; Richelle, G. J. J.; Hackeng, T. M.; van Maarseveen, J. H.; Nuijens, T. Omniligase-1: A Powerful Tool for Peptide Head-to-Tail Cyclization. *Adv. Synth. Catal.* **2017**, *359* (12), 2050–2055.

- (16) Burton, A. J.; Thomson, A. R.; Dawson, W. M.; Brady, R. L.; Woolfson, D. N. Installing Hydrolytic Activity into a Completely de Novo Protein Framework. *Nat. Chem.* **2016**, *8* (9), 837–844.
- (17) Watkins, D. W.; Jenkins, J. M. X.; Grayson, K. J.; Wood, N.; Steventon, J. W.; Le Vay, K. K.; Goodwin, M. I.; Mullen, A. S.; Bailey, H. J.; Crump, M. P.; MacMillan, F.; Mulholland, A. J.; Cameron, G.; Sessions, R. B.; Mann, S.; Anderson, J. L. R. Construction and in Vivo Assembly of a Catalytically Proficient and Hyperthermostable de Novo Enzyme. *Nat. Commun.* **2017**, *8* (1), 358.
- (18) Simón, L.; Goodman, J. M. Enzyme Catalysis by Hydrogen Bonds: The Balance between Transition State Binding and Substrate Binding in Oxyanion Holes. *J. Org. Chem.* **2010**, *75* (6), 1831–1840.
- (19) Poole, L. B. The Basics of Thiols and Cysteines in Redox Biology and Chemistry. *Free Radic. Biol. Med.* **2015**, *80*, 148–157.
- (20) Yang, W.; Drueckhammer, D. G. Understanding the Relative Acyl-Transfer Reactivity of Oxoesters and Thioesters: Computational Analysis of Transition State Delocalization Effects. *J. Am. Chem. Soc.* **2001**, *123* (44), 11004–11009.
- (21) 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.
- (22) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266* (5186), 776–779.
- (23) Andrew, C. D.; Warwicker, J.; Jones, G. R.; Doig, A. J. Effect of Phosphorylation on  $\alpha$ -Helix Stability as a Function of Position. *Biochemistry* **2002**, *41* (6), 1897–1905.
- (24) Ogihara, N. L.; Ghirlanda, G.; Bryson, J. W.; Gingery, M.; DeGrado, W. F.; Eisenberg, D. Design of Three-Dimensional Domain-Swapped Dimers and Fibrous Oligomers. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (4), 1404–1409.
- (25) Behrendt, R.; White, P.; Offer, J. Advances in Fmoc Solid-Phase Peptide Synthesis. *J. Pept. Sci.* **2016**, *22* (1), 4–27.
- (26) Ghirlanda, G.; Lear, J. D.; Lombardi, A.; DeGrado, W. F. From Synthetic Coiled Coils to Functional Proteins: Automated Design of a Receptor for the Calmodulin-Binding Domain of Calcineurin. *J. Mol. Biol.* **1998**, *281* (2), 379–391.
- (27) 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.* **2011**, *50* (33), 7645–7649.
- (28) Yan, L. Z.; Dawson, P. E. Synthesis of Peptides and Proteins without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization. *J. Am. Chem. Soc.* **2001**, *123* (4), 526–533.
- (29) Pentelute, B. L.; Kent, S. B. H. Selective Desulfurization of Cysteine in the Presence of Cys(Acm) in Polypeptides Obtained by Native Chemical Ligation. *Org. Lett.* **2007**, *9* (4), 687–690.
- (30) 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.
- (31) Kochendoerfer, G. G.; Chen, S.-Y.; Mao, F.; Cressman, S.; Traviglia, S.; Shao, H.; Hunter, C. L.; Low, D. W.; Cagle, E. N.; Carnevali, M.; Gueriguian, V.; Keogh, P. J.; Porter, H.; Stratton, S. M.;

- Wiedeke, M. C.; Wilken, J.; Tang, J.; Levy, J. J.; Miranda, L. P.; Crnogorac, M. M.; Kalbag, S.; Botti, P.; Schindler-Horvat, J.; Savatski, L.; Adamson, J. W.; Kung, A.; Kent, S. B. H.; Bradburne, J. A. Design and Chemical Synthesis of a Homogeneous Polymer-Modified Erythropoiesis Protein. *Science* **2003**, 299 (5608), 884–887.
- (32) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. In Situ Neutralization in Boc-Chemistry Solid Phase Peptide Synthesis. *Int. J. Pept. Res. Ther.* **2007**, 13 (1), 31–44.
- (33) Muttenthaler, M.; Albericio, F.; Dawson, P. E. Methods, Setup and Safe Handling for Anhydrous Hydrogen Fluoride Cleavage in Boc Solid-Phase Peptide Synthesis. *Nat. Protoc.* **2015**, 10 (7), 1067–1083.
- (34) Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I. Color Test for Detection of Free Terminal Amino Groups in the Solid-Phase Synthesis of Peptides. *Anal. Biochem.* **1970**, 34 (2), 595–598.
- (35) Grigoryan, G.; DeGrado, W. F. Probing Designability via a Generalized Model of Helical Bundle Geometry. *J. Mol. Biol.* **2011**, 405 (4), 1079–1100.
- (36) Weil-Ktorza, O.; Rege, N.; Lansky, S.; Shalev, D. E.; Shoham, G.; Weiss, M. A.; Metanis, N. Substitution of an Internal Disulfide Bridge with a Diselenide Enhances Both Foldability and Stability of Human Insulin. *Chem. Eur. J.* **2019**, 25 (36), 8513–8521.
- (37) Cezari, M. H. S.; Puzer, L.; Juliano, M. A.; Carmona, A. K.; Juliano, L. Cathepsin B Carboxydipeptidase Specificity Analysis Using Internally Quenched Fluorescent Peptides. *Biochem. J.* **2002**, 368 (1), 365–369.
- (38) Grieco, P.; Gitu, P. M.; Hruby, V. J. Preparation of ‘Side-Chain-to-Side-Chain’ Cyclic Peptides by Allyl and Alloc Strategy: Potential for Library Synthesis. *J. Pept. Res.* **2001**, 57 (3), 250–256.
- (39) Widmer, S.; J. Reber, M.; Müller, P.; E. Housecroft, C.; C. Constable, E.; M. Rossi, R.; Brühwiler, D.; J. Scherer, L.; F. Boesel, L. Incorporation of a FRET Dye Pair into Mesoporous Materials: A Comparison of Fluorescence Spectra, FRET Activity and Dye Accessibility. *Analyst* **2015**, 140 (15), 5324–5334.
- (40) Sapsford, K. E.; Berti, L.; Medintz, I. L. Materials for Fluorescence Resonance Energy Transfer Analysis: Beyond Traditional Donor–Acceptor Combinations. *Angew. Chem. Int. Ed.* **2006**, 45 (28), 4562–4589.
- (41) Schmidtgal, B.; Chaloin, O.; Bauer, V.; Sumyk, M.; Birck, C.; Torbeev, V. Dissecting Mechanism of Coupled Folding and Binding of an Intrinsically Disordered Protein by Chemical Synthesis of Conformationally Constrained Analogues. *Chem. Commun.* **2017**, 53 (53), 7369–7372.
- (42) Johnson, E. C. B.; Kent, S. B. H. Insights into the Mechanism and Catalysis of the Native Chemical Ligation Reaction. *J. Am. Chem. Soc.* **2006**, 128 (20), 6640–6646.
- (43) Sievers, A.; Beringer, M.; Rodnina, M. V.; Wolfenden, R. The Ribosome as an Entropy Trap. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, 101 (21), 7897–7901.
- (44) Bender, M. L.; Kezdy, F. J.; Wedler, F. C. Alpha-Chymotrypsin: Enzyme Concentration and Kinetics. *J. Chem. Educ.* **1967**, 44 (2), 84.
- (45) Marshall, L. R.; Zozulia, O.; Lengyel-Zhand, Z.; Korendovych, I. V. Minimalist de Novo Design of Protein Catalysts. *ACS Catal.* **2019**, 9 (10), 9265–9275.

- (46) Jencks, W. P.; Carriuolo, J. Reactivity of Nucleophilic Reagents toward Esters. *J. Am. Chem. Soc.* **1960**, *82* (7), 1778–1786.
- (47) Daragan, V. A.; Mayo, K. H. Peptide Dynamics in Triglycine: Coupling of Internal Bond Rotations and Overall Molecular Tumbling. *J. Phys. Chem.* **1994**, *98* (42), 10949–10956.
- (48) Hünig, S.; Schaller, R. The Chemistry of Acyl Cyanides. *Angew. Chem. Int. Ed. Engl.* **1982**, *21* (1), 36–49.
- (49) Eggerer, H.; Stadtman, E. R.; Poston, J. M. On the Role of Acetyl Cyanide in the Cyanide-Induced Acetylation of Amino Acids by Enzymes of *Clostridium Kluyveri*. *Arch. Biochem. Biophys.* **1962**, *98* (3), 432–443.
- (50) 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.
- (51) Khersonsky, O.; Roodveldt, C.; Tawfik, D. S. Enzyme Promiscuity: Evolutionary and Mechanistic Aspects. *Curr. Opin. Chem. Biol.* **2006**, *10* (5), 498–508.
- (52) Koehnke, J.; Bent, A.; Houssen, W. E.; Zollman, D.; Morawitz, F.; Shirran, S.; Vendome, J.; Nneoyiegbe, A. F.; Trembleau, L.; Botting, C. H.; Smith, M. C. M.; Jaspars, M.; Naismith, J. H. The Mechanism of Patellamide Macrocyclization Revealed by the Characterization of the PatG Macrocyclase Domain. *Nat. Struct. Mol. Biol.* **2012**, *19* (8), 767–772.
- (53) Chekan, J. R.; Estrada, P.; Covello, P. S.; Nair, S. K. Characterization of the Macrocyclase Involved in the Biosynthesis of RiPP Cyclic Peptides in Plants. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114* (25), 6551–6556.
- (54) Zauner, F. B.; Elsasser, B.; Dall, E.; Cabrele, C.; Brandstetter, H. Structural Analyses of *Arabidopsis Thaliana* Legumain  $\gamma$  Reveal the Differential Recognition and Processing of Proteolysis and Ligation Substrates. *J. Biol. Chem.* **2018**, *293* (23), 8934–8946.
- (55) Haywood, J.; Schmidberger, J. W.; James, A. M.; Nonis, S. G.; Sukhoverkov, K. V.; Elias, M.; Bond, C. S.; Mylne, J. S. Structural Basis of Ribosomal Peptide Macrocyclization in Plants. *eLife* **2018**, *7*, e32955.
- (56) Hemu, X.; Sahili, A. E.; Hu, S.; Wong, K.; Chen, Y.; Wong, Y. H.; Zhang, X.; Serra, A.; Goh, B. C.; Darwis, D. A.; Chen, M. W.; Sze, S. K.; Liu, C.-F.; Lescar, J.; Tam, J. P. Structural Determinants for Peptide-Bond Formation by Asparaginyl Ligases. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (24), 11737–11746.
- (57) James, A. M.; Haywood, J.; Leroux, J.; Ignasiak, K.; Elliott, A. G.; Schmidberger, J. W.; Fisher, M. F.; Nonis, S. G.; Fenske, R.; Bond, C. S.; Mylne, J. S. The Macrocyclizing Protease Butelase 1 Remains Autocatalytic and Reveals the Structural Basis for Ligase Activity. *Plant J.* **2019**, *98* (6), 988–999.
- (58) Connors, K. A.; Bender, M. L. The Kinetics of Alkaline Hydrolysis and N-Butylaminolysis of Ethyl p-Nitrobenzoate and Ethyl p-Nitrothiolbenzoate<sup>1a</sup>. *J. Org. Chem.* **1961**, *26* (7), 2498–2504.





**Chapter 3. Libraries of analogues and  
substrates: towards a better  
understanding of the catalytic activity  
of *de novo* protein**

### 3.1. Introduction

Enzymes are outstanding catalysts thanks to their high efficiency, chemoselectivity, regio- and stereospecificity as well as their biodegradability, which is appreciable for environmental concerns. However, natural enzymes only catalyze a restricted range of biological reactions and their effectiveness is indeed intricately linked to substrate specificity, which limits their applicability to a broader scope of chemical reactions.

*De novo* catalytic proteins with designed, small and defined structures can provide a good alternative to natural enzymes as they may combine advantages from natural catalysts and from organic catalysts. In other words, *de novo* enzymes can have the capacity to chemoselectively and stereospecifically carry out chemical transformations via transfer of chirality from the chiral scaffold to the substrate. Furthermore, such catalysts can possess higher catalytic promiscuity with a range of related substrates and be biodegradable. As an example, a *de novo* maquette protein has been designed to bind a c-type cytochrome and has in fact resulted in a proficient peroxidase activity for a diverse set of substrates.<sup>1</sup> The resulting catalytic protein has also exhibited enhanced chemical and thermal stability compared to natural enzymes. Although no concrete application has been so far demonstrated, such *de novo* designed proteins have the potential for further development to become general and practical tools for catalysis. Moreover, another advantage of miniaturized and designed frameworks in comparison to natural protein scaffolds is the possibility to deeper understand the contribution of individual residues for protein stability, folding or activity.<sup>2</sup>

One strategy to install a catalytic activity into a *de novo* inert scaffold consists in the exact placement of a constellation of amino acids to modulate the reactivity of the catalytic residues. The hydrolytic activity on ester substrates has been, for example, attained either with the incorporation of a network of histidine residues leading to  $pK_a$  depression in a four-helix bundle<sup>3</sup> or with the insertion of functional catalytic Cys-His-Glu triads into a designed coiled-coil heptamer.<sup>4</sup> Finding the combination of amino acid residues to facilitate amide bond formation represents a particular challenge and will allow for a better understanding of fundamental principles and prerequisites for peptide bond synthesis by enzymes.

In the previous chapter, I have presented data on catalytic activity for acyl transfer reactions using peptide- $\alpha$ thioesters as substrates and, furthermore, the acceleration of amide bond formation by installing a unique catalytic cysteine in a *de novo* protein scaffold. To go further, a network of residues has to be designed to improve this catalytic activity by contributing to the mechanism and modulating the activity of the cysteine. In addition, by using peptide- $\alpha$ thioesters as acyl donor substrates, we observed competition between

aminolysis and hydrolysis reaction pathways. We compared these effects with the natural proteases and ligases that belong to the same family of proteins, the legumains, which catalyze the respective reactions through substrate-enzyme thioester intermediates and bear similar three-dimensional structures with only subtle changes in the catalytic site.<sup>5</sup> Our designed catalytic proteins are therefore good candidates to further study the intricate features of thioesters' reactivity including the role of stereoelectronic effects thanks to the advantage of *de novo* protein carrying modular scaffold.

To improve catalytic activity and to better control the reaction outcome, one strategy is to apply an iterative approach based on several cycles of design, synthesis and characterization. Previously, iterative strategies have already demonstrated their potency by substantially increasing the success rate of computational design. Multiple iterative cycles of detailed biophysical and structural analysis combined with molecular dynamics (MD) simulations have impressively led to the production a Kemp eliminase with one of the highest kinetics parameters from a completely inactive variant.<sup>6</sup> Few other successful examples have involved directed evolution with high-throughput production of analogues resulting in stepwise enhancement of the catalytic activity. Thanks to directed evolution, a *de novo* Zn-containing catalytic protein was for instance further improved resulting in highly efficient activity on par with naturally occurring metalloenzymes.<sup>7</sup>

As protein chemists, we decided to apply a convergent approach to synthesize a library of new analogues as part of our iterative process. A set of different peptide fragments has been chemically synthesized and assembled to produce new protein variants. However, from the crystal structures of previously synthesized analogues **Het-N2**, **Het-Ncap** and **Hom-N2**, no clear insights concerning difference of reactivity or tetrahedral intermediates stabilization were deduced. Therefore, the design of the new sequences to assemble the library was mainly based on intuition and on few assumptions.

In parallel, libraries of substrates (*i.e.* one library for the acyl donor substrates and one for the acyl acceptor substrates) were synthesized to define in more details the sequence promiscuity of substrates already detected for **Het-N2** protein in Chapter 2. These libraries may help us to better understand the differences in aminolysis over hydrolysis depending on the nature of the substrates. The catalytic assays will be performed in the near future.

### 3.2. Library of the second generation of DSD analogues

To better understand the influence of the near environment of the cysteine on catalysis, a library of analogues was designed based on the sequence of the most efficient protein obtained so far, the **Het-N2** heterodimer. These different variants differ in their two first N-terminal residues, namely Ncap and N1 residues, which are in close proximity to the catalytic cysteine (at N2 position):

#### Het-N2 analogues:

'Positive' strand: Ac-NLAALRSELQALRREGFSPERLAALESRLQALERRLAALRSRLQALRG-NH<sub>2</sub>  
 'Negative' strand: Ac-**Ncap-N1-C**ALRSELQALRREGFSPEELAALESELQALERELAALRSELQALRG-NH<sub>2</sub>

Thus, direct impact on the catalysis and changes of kinetics constants may be observed through the incorporation of new surrounding amino acids. To build these new analogues of **Het-N2**, the 'positive' strand that remains unchanged compared to the original **Het-N2** protein is combined with the 'negative' strand containing a Cys residue at N2 position and various Ncap and N1 amino acids. To synthesize the different second strand bearing the catalytic N2 cysteine, we have taken advantage of the already present cysteine residue to apply a convergent approach via native chemical ligation in which a peptide including only Ncap and N1 residues is ligated to the Cys-containing C-terminal fragment that is invariant in all new analogues. The amino acids for incorporation at Ncap and N1 residues have been selected by inspection of literature and the naturally existing catalytic sites.

Besides the modifications on the 'negative' strand, few analogues have been also designed by modifying the sequence of the 'positive' strand and keeping the Cys-containing strand unchanged compared to **Het-N2**. Finally, kinetic parameters of the library of analogues have been investigated through optimized LC/MS assay (see subsection 2.6.4). To obtain an analogue with even greater catalytic activity, combining the best 'positive' and 'negative' strands will be later attempted.

#### 3.2.1. Design of the library of analogues

To examine the effect on catalytic activity by surrounding charged or hydrophobic residues, as well as the effect of residues introduced to stabilize tetrahedral intermediates, 21 sequences containing different amino acids at Ncap and N1 positions on the 'negative'

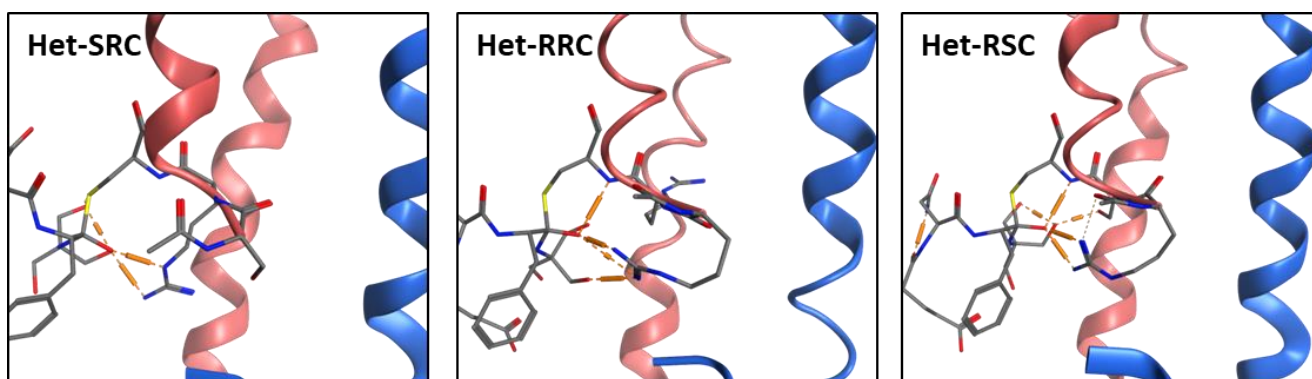
strand were designed (**Table 3.1**). The resulting analogues are abbreviated as follows: **AA(Ncap)AA(N1)C** strand (substitution of AA(Ncap) and AA(N1) by the 1-letter code of the corresponding amino acids) for the ‘negative’ strand alone and **Het-AA(Ncap)AA(N1)C** while considering the complexation of the ‘positive’ strand and the corresponding **AA(Ncap)AA(N1)Cys** strand forming the desired heterodimer.

**Table 3.1.** List of the ‘negative’ strand variants based on the corresponding sequence from Het-N2 construct.

<b>Ncap residue</b>	<b>N1 residue</b>	<b>Name of the resulting ‘negative’ stand</b>	<b>Reason for the modifications</b>
Ser	Arg	<b>SRC</b> strand	Phosphoserine stabilization
Arg	Arg	<b>RRC</b> strand	
Arg	Ser	<b>RSC</b> strand	
Gly	Arg	<b>GRC</b> strand	Extra stabilization via additional hydrogen bonds and electrostatic interactions with arginine
Arg	Leu	<b>RLC</b> strand	
Arg	Ala	<b>RAC</b> strand	
Asn	Arg	<b>NRC</b> strand	
Gly	Ala	<b>GAC</b> strand	Influence of the hydrophobic residue at N1 position
Gly	Ile	<b>GIC</b> strand	
Gly	Val	<b>GVC</b> strand	
Gly	Phe	<b>GFC</b> strand	
Gly	Tyr	<b>GYC</b> strand	
Gly	Trp	<b>GWC</b> strand	
Gly	His	<b>GHC</b> strand	His-Cys dyad
His	Leu	<b>HLC</b> strand	
Asn	His	<b>NHC</b> strand	
His	Ser	<b>HSC</b> strand	
Glu	Leu	<b>ELC</b> strand	Analogy with N-terminal acetyltransferases
Tyr	Leu	<b>YLC</b> strand	
Glu	Tyr	<b>EYC</b> strand	
Gly derivative	Leu	<b>G<sup>u</sup>LC</b> strand	Urea linkage providing an additional unpaired NH

The first set of analogues has been designed based on a study published by W. F. DeGrado and colleagues showing the stabilizing interactions of Ncap and N1 residues with phosphoserine at N2 position.<sup>8</sup> In this work, Ser and Arg residues at Ncap and N1 positions,

respectively, have shown to facilitate the transition from unfolded monomer to folded helical bundle upon phosphorylation of Ser residue at N2 due to the idealized geometry of Ser at Ncap and potential stabilizing hydrogen bonds and electrostatic interactions between Arg and phospho-Ser. Conversely, Arg at Ncap position would have destabilizing effect that could be nevertheless slightly counterbalanced thanks to formation of salt bridges with phospho-Ser. The modification of Arg to Ser at N1 position might provide less stabilizing effect on the phospho-Ser due to the smaller size and the neutral charge even if hydrogen bond are possible with phospho-Ser. In this way, sequences from this study have been selected resulting in three new analogues **Het-SRC**, **Het-RRC** and **Het-RSC**, which are expected to contribute to the extra stabilizing interactions with tetrahedral intermediates as shown by modeling in **Figure 3.1**. Following the same reasoning, the **Het-SRC** analogue should provide better stabilization of the oxyanion intermediates, therefore, leading to rate enhancement.



**Figure 3.1. Modeling of tetrahedral intermediate of aminolysis step on new analogues Het-SRC, Het-RRC and Het-RSC.** Hydrogen bonding is depicted in orange and is clearly visible by modeling between arginine at either Ncap or N1 position and the oxyanion. The 'negative' strand is shown in pink and the 'positive' strand in blue.

Other sequences with an arginine amino acid at either Ncap or N1 position were imagined due to its capability to donate hydrogen bonds for additional stabilization. By keeping the original residue of **Het-N2** at Ncap or N1 position, **Het-GRC** and **Het-RLC** have been conceived, respectively. Finally, two other analogues with arginine residue have also been designed, **Het-RAC** and **Het-NRC**.

A third batch of variants has been developed to evaluate the influence on catalysis of the hydrophobic residue at N1 position. Thus, the original leucine residue has been replaced by all other hydrophobic and/or aromatic residues (except Met residue) giving the following heterodimer proteins: **Het-GAC**, **Het-GIC**, **Het-GVC**, **Het-GFC**, **Het-GYC** and **Het-GWC**.

The natural cysteine proteases or ligases generally contain a catalytic triad or dyad that consists of a catalytic cysteine connected to at least a histidine residue.<sup>9,10</sup> This histidine typically mediates deprotonation of the cysteine to improve the nucleophilicity of the catalytic residue. The His-Cys dyad has already been utilized in complement to oxyanion hole as a

catalytic strategy to design ester hydrolases.<sup>11</sup> Moreover, histidine residues are often found in catalytic site as a proton shuttle and can be involved in general acid/base catalysis promoted by the imidazole ring.<sup>12</sup> In our case, histidine residue is interesting for being involved in cysteine deprotonation and possibly for being involved in the mechanism through deprotonation of the amine after the attack on the branched adduct. Therefore, four analogues containing a histidine residue have been conceived. Again, histidine has been accompanied with the original residues of **Het-N2** at either Ncap or N1 position resulting in **Het-GHC** and **Het-HLC** proteins. The two other histidine variants are: **Het-NHC** and **Het-HSC**.

By analyzing catalytic sites of an N-terminal acetyltransferase (e.g. NatA complex) that accelerates amide bond formation through acyl transfer from the acetyl CoA to N-terminal amine, it was noticed that Glu and Tyr amino acids may play an important role in the catalysis.<sup>13</sup> Thus, we decided to incorporate such residues in new analogues leading to **Het-ELC**, **Het-YLC** and **Het-EYC** proteins.

Additionally, we can take advantage provided by chemical synthesis to incorporate non-natural amino acids to expand our library. Our catalytic strategy mainly relies on the stabilization of tetrahedral intermediates (and presumably transition states) through hydrogen bonding with unpaired NHs of helix N-terminus. The team of Gilles Guichard has described binding of various anions to aliphatic oligoureases that has been mediated by the helical dipole and the interactions with the free NHs located at the extremity of the helix.<sup>14</sup> A building block has been incorporated at Ncap position to form a urea linkage with N1 amino acid (G<sup>u</sup> for the glycine derivative used to constitute a urea bond while coupling to previous amino acid) giving the analogue **Het-G<sup>u</sup>LC**, which would carry an additional unpaired NH at the N-terminus compared to the original **Het-N2** for oxyanion stabilization.

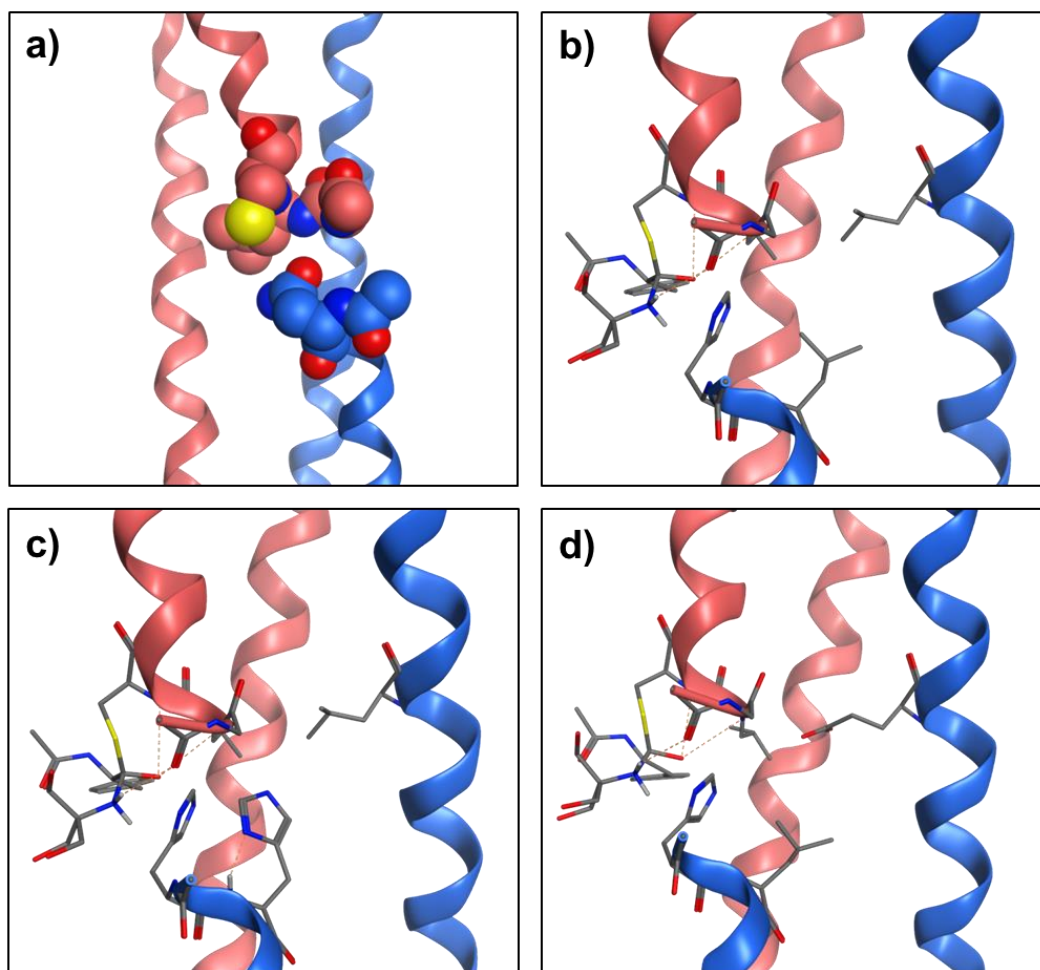
Finally, modifications on the 'positive' strand have been also investigated in order to take advantage of the heterodimeric structure and of the two N-termini facing each other creating a cavity. In this way, four analogues with sequences differing on the 'positive' strand were designed by inspecting the crystal structure of the **Het-N2** catalytic protein and by modeling the different modifications. These new designed sequences are summarized in **Table 3.2**. The crystal structure of the designed **Het-N2** protein has revealed that the asparagine at Ncap position of the 'positive' strand is oriented towards the oxyanion hole of the **Cys-N2** ('negative') strand (**Figure 3.2a**) instead of being oriented towards NHs of its own helix. Thus, we hypothesized that the catalytic activity of **Het-N2** may be slowed down by a lack of accessibility to the oxyanion hole due to steric hindrance caused by the asparagine of the 'positive' strand. To increase the access to the catalytic site and especially to the oxyanion hole, a first variant of the 'positive' strand has been conceived by modifying

the Asn to a Gly amino acid at Ncap position (first residue in the sequence) and has been named 'positive, G1' strand. The heterodimer resulting from complexation of the **Cys-N2** and the 'positive, G1' strands was named **Het-CysN2/'positive, G1'**. Then, three other analogues have been designed by adding a His residue in the 'positive' strand in order to assist in deprotonation steps, particularly in the rate-limiting step in which the nitrogen of the resulting amide bond needs to be deprotonated. A histidine residue has been therefore introduced at the Ncap position in close proximity to the nitrogen that has to be deprotonated to form the amide bond (according to modeling depicted in **Figure 3.2b**), resulting in 'positive, H1' strand. To depress the  $pK_a$  of this histidine residue that is expected to provide more efficient catalytic proteins, two other analogues have been envisaged. Thus, a second His amino acid was introduced adjacent to the first His leading to the 'positive, H1H2' strand (**Figure 3.2c**). In addition, a Glu residue has been introduced in the 36<sup>th</sup> position in the sequence, which is located in the hydrophobic core nearby the histidine at Ncap position, and providing the **Het-CysN2/'positive, H1E36'** heterodimer upon combining the strands **Cys-N2** and 'positive, H1E36' (**Figure 3.2d**).

**Table 3.2.** List of the designed 'positive' strand variants based on Het-N2 sequence. The modifications are indicated in bold and red.

Name of the new 'positive' strand	Sequence
'positive, G1' strand	Ac- <b>G</b> LAALRSELQALRREGFSPERLAALESRLQALERRLAALRSRLQALRG-NH <sub>2</sub>
'positive, H1' strand	Ac- <b>H</b> LAALRSELQALRREGFSPERLAALESRLQALERRLAALRSRLQALRG-NH <sub>2</sub>
'positive, H1H2' strand	Ac- <b>HH</b> AALRSELQALRREGFSPERLAALESRLQALERRLAALRSRLQALRG-NH <sub>2</sub>
'positive, H1E36' strand	Ac- <b>H</b> LAALRSELQALRREGFSPERLAALESRLQALERR <b>E</b> AALRSRLQALRG-NH <sub>2</sub>





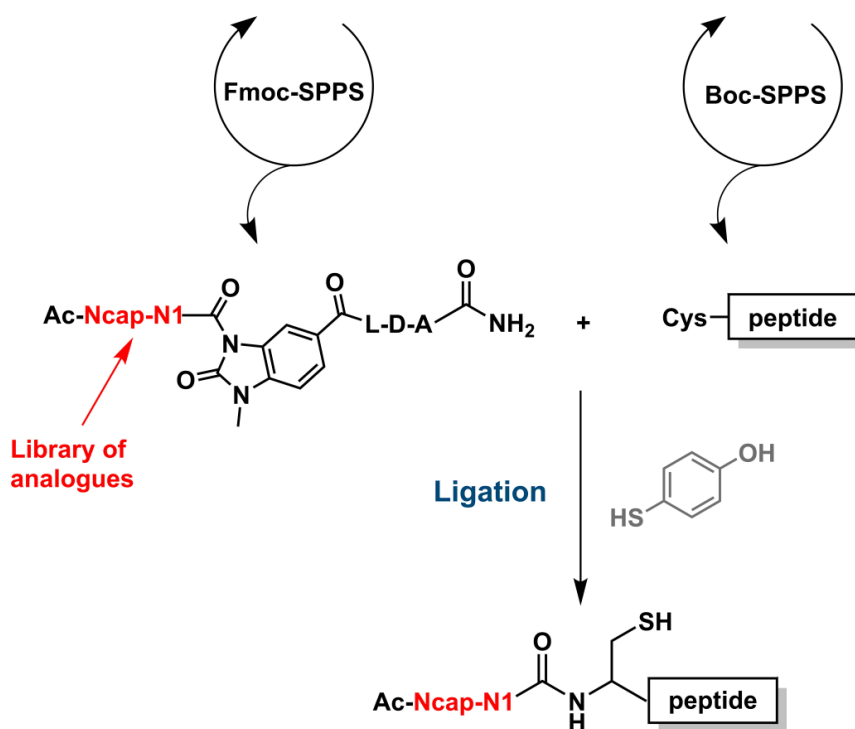
**Figure 3.2. Design of analogues of the 'positive' strand (depicted in blue).** **a)** Crystal structure of **Het-N2** showing the Asn of 'positive' strand oriented towards the oxyanion hole of the other strand. **b-d)** Model of the tetrahedral intermediate of the aminolysis step between a peptide- Phe- $\alpha$ thioester and Tris. **b)** Model of 'positive, H1' strand depicting the histidine in close proximity to the nitrogen of the amide bond in formation. **c)** Model of 'positive, H1H2' strand for potential  $pK_a$  depression. **d)** Model of the 'positive, H1E36' strand, the glutamate residue is located in the hydrophobic core nearby the histidine residue. H-bonds are depicted in orange dotted lines.

### 3.2.2. Design of the synthetic route for 'negative' strand analogues

The synthetic strategy to produce the library of 'negative' strand analogues is depicted in **Figure 3.3**. The monomer has been split into two peptide fragments between N1 residue (*i.e.* the second residue of the sequence) and the cysteine at N2 position (*i.e.* third residue of the sequence). The Cys-containing C-terminal fragment remains constant for all the analogues and can be synthesized in one large batch by Boc-SPPS. The N-terminal peptide segment is thus composed of only two amino acids (Ncap and N1 residues). However, we anticipated problem of solubility of some of those dipeptides especially during thioesterification and ligation steps and/or problem of retention on reverse phase HPLC columns during analysis, reaction monitoring and purification processes. Consequently, a

suitable balance of hydrophobic and hydrophilic residues are indispensable for peptide being soluble in aqueous buffer while having the possibility to analyze their purity and monitor progress of reactions. One approach would be to synthesize the different N-terminal fragments by Boc-SPPS in which a thioester moiety can be directly incorporated after coupling of several amino acids (at the C-terminal of the thioester linkage) that would enhance the solubility of the resulting peptide- $\alpha$ thioester. However, this approach was not retained for practical reasons, notably due to hazardous conditions during HF cleavage and the necessity of special equipment, which prevents from multiple cleavages in parallel. In fact, to be as efficient as possible in the synthesis of the N-terminal fragments, we have considered syntheses in parallel and Fmoc-SPPS is the most reliable to do so. By using Fmoc-SPPS, automatization of the procedure is actually available allowing multiple syntheses in parallel. Thus, for our dipeptides that have to be thioesterified for ligation, we have chosen the procedure introduced by P. Dawson and co-workers, which is based on *N*-acylurea linker.<sup>15</sup> This strategy allows for the incorporation of a spacer sequence before the linker coupling that is generally recommended because of difficulties to efficiently couple the linker directly on the resin (*i.e.* without spacer sequence). We have chosen a spacer sequence composed of LDA residues in order to *i)* enhance retention on the reverse phase columns for HPLC analysis and purification by providing additional hydrophobic residues, and *ii)* improve the final product solubility especially for hydrophobic dipeptides thanks to the charged Asp residue. The insertion of such a *N*-acyl-*N'*-methylurea (MeNbz) linker at the C-terminus of the desired dipeptides as a leaving group allows for *in situ* thioesterification with an aryl-thiol followed by ligation with the Cys-C-terminal fragment to generate the complete 'negative' strand. The cysteine residue required for the NCL corresponds to the catalytic cysteine at N2 position.

To facilitate the production of the library of analogues for evaluation of their catalytic properties, we planned to perform several ligations at the same time and on a small scale. Best variants would then be produced on a larger scale. However, HPLC purification is not appropriate for small scale ligations because of possible losses of too much product. For a first estimation of the catalytic activity, high purity of the protein analogue is actually not imperative and we have decided to separate ligation product from the unreacted peptide fragments, by-products, organic molecules and salts by solid-phase extraction (SPE), which is convenient for multiple small scale separations in parallel.

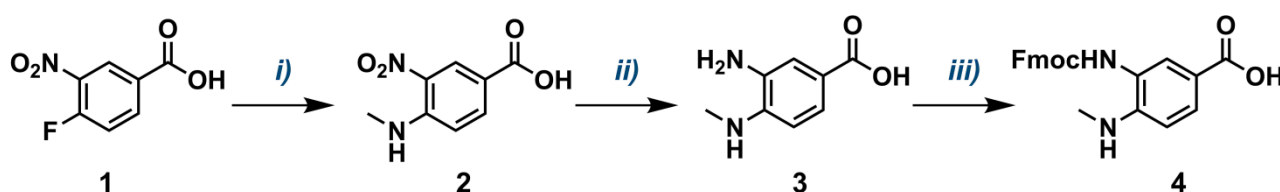


**Figure 3.3. Synthetic strategy to produce a library of ‘negative’ strand analogues in parallel.** The Cys-containing C-terminal fragment is common to all analogues and would be synthesized by Boc-SPPS. The N-terminal fragments containing different Ncap and N1 residues are to be synthesized by Fmoc-SPPS in the form of *N*-acyl-*N'*-methylurea (MeNbz) peptides. During ligation procedure, the N-terminal peptide is thioesterified *in situ* directly followed by the native chemical ligation with the C-terminal segment leading to the full-length ‘negative’ strand analogue.

### 3.2.3. Synthesis of the library of analogues

The synthesis of the library of ‘negative’ strand analogues has been divided into several steps. We have initially focused on the synthesis of the *N*-acyl-*N'*-methylurea (MeNbz) peptide fragment bearing different Ncap and N1 amino acids. First, a MeNbz linker surrogate, Fmoc-3-amino-4-(methylamino)benzoic acid (Fmoc-MeDbz), has been synthesized. Then, the N-terminal peptide fragments have been assembled on solid support by coupling a spacer sequence, followed by the MeDbz linker and the dipeptide Ncap-N1. Before cleavage, the MeDbz linker has been transformed to MeNbz resulting in a good leaving group for further thioesterification. Simultaneously, the C-terminal segment, common to all analogues, has been produced by Boc-SPPS. Finally, *in situ* thioesterification of the N-terminal peptide and native chemical ligation have proceeded to assemble the full-length analogue, which has been separated from the reaction mixture by solid phase extraction (SPE). In addition, the synthesis of the different ‘positive’ strands has been achieved by sequentially assembling the full-length sequence by Boc-SPPS.

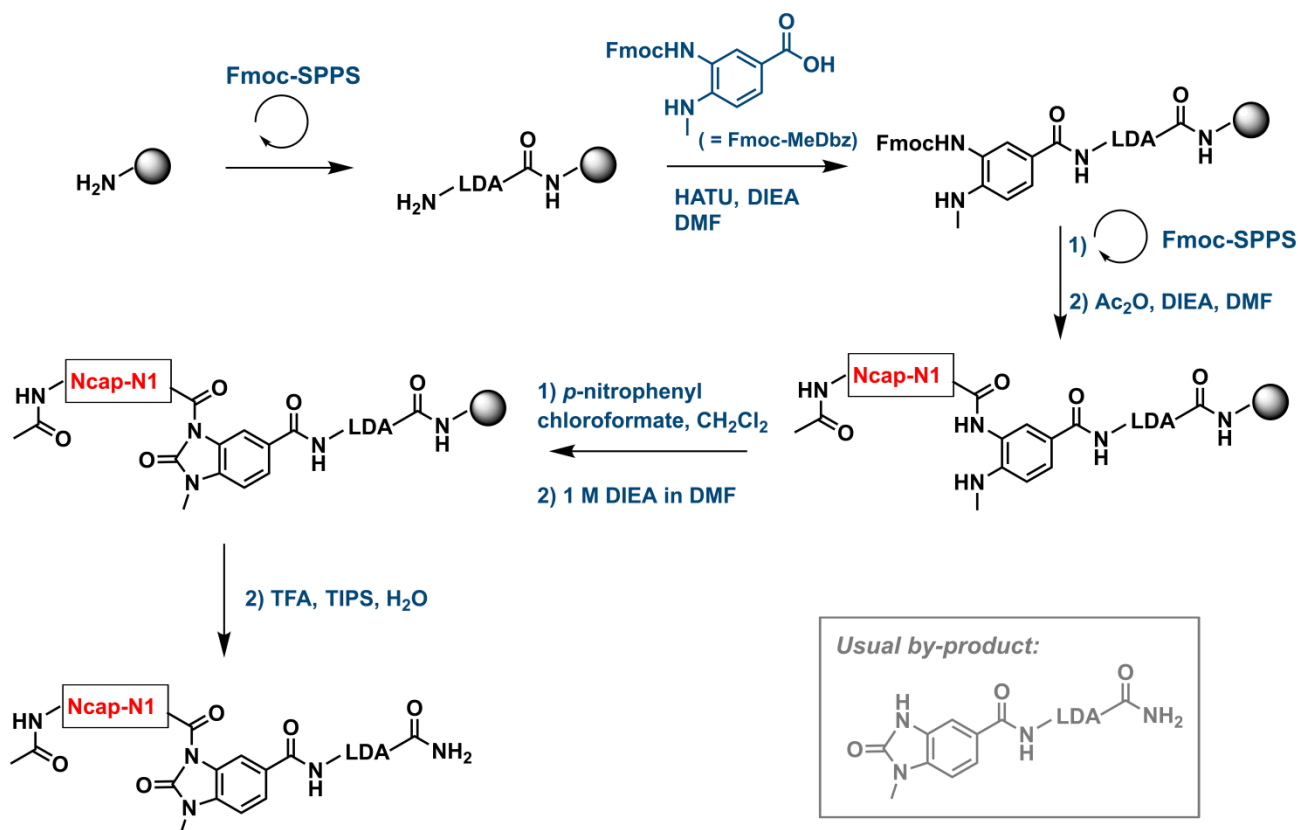
**Synthesis of the Fmoc-MeDbz building block.** Synthesis of the Fmoc-MeDbz building block was performed following the three-step procedure published by P. Dawson and coworkers on a gram scale (as shown in **Scheme 3.1**).<sup>15</sup> Nucleophilic aromatic substitution of methylamine on the commercially available 3-fluoro-4-nitrobenzoic acid **1** led to the 3-nitro-4-(methylamino)benzoic acid **2**. After complete consumption of the starting material, product was precipitated at acidic pH and was directly introduced in the second step of the synthesis, in which the nitro group was quantitatively reduced via catalytic hydrogenation with Pd/C. The resulting 3-amino-4-(methylamino)benzoic acid **3** was chemoselectively protected on the primary aryl amine by Fmoc-Cl providing the expected Fmoc-3-amino-4-(methylamino)benzoic acid (Fmoc-MeDbz) **4** with an overall yield of 62 %. According to HPLC, LC/MS and <sup>1</sup>H NMR, the final product was obtained in satisfactory purity and was introduced in SPPS without any further purification.



**Scheme 3.1. Synthetic route of the Fmoc-MeDbz linker.** Reaction conditions: *i*) MeNH<sub>2</sub>, MeOH, overnight, room temperature; *ii*) H<sub>2</sub>, Pd/C, MeOH, overnight, room temperature; *iii*) Fmoc-Cl, DIEA, H<sub>2</sub>O/MeCN, 2 h, room temperature.

**Fmoc-SPPS assembly of the N-terminal fragments as Ac-Ncap-N1-MeNbz-LDA peptide.** The synthetic method used to assemble the N-terminal peptide segments is described in **Scheme 3.2**. After incorporation of the spacer sequence by Fmoc-SPPS on Rink amide resin, the Fmoc-MeDbz linker was coupled following the same procedure as the amino acids, *i.e.* using HATU as a coupling agent in the presence of base (here, DIEA), but for a longer time to reach completion of the reaction. As the spacer sequence and the linker are common to all the peptide analogues, they were coupled in large batch and the resulting resin was then split to perform couplings of the dipeptides Ncap-N1 on a 0.05 mmol scale. After removal of the Fmoc protecting group of MeDbz, the target sequence of two residues was coupled. However, because of less efficient coupling engendered by the linker these Fmoc-amino acids were both double coupled. After final N-terminal deprotection and acetylation of the peptide on resin, the secondary amine of the linker was acylated with *p*-nitrophenyl chloroformate. Finally, the MeNbz entity was formed by an intramolecular cyclization in basic conditions (1 M DIEA in DMF). The different synthesized *N*-acylurea peptides were subjected to acidolytic cleavage conditions resulting in the expected fully deprotected peptides. The N-terminal fragment synthesis was mainly performed machine-assisted, but with the requirement of a manual handling for: *i*) the linker coupling owing to

solubility issue incompatible with automated synthesizer (that is to say the linker needs to be freshly solubilized before coupling); *ii*) acetylation due to incompatibility with the synthesizer conditions; *iii*) the acylation and cyclization steps because of the necessity of dry solvent and inert atmosphere.



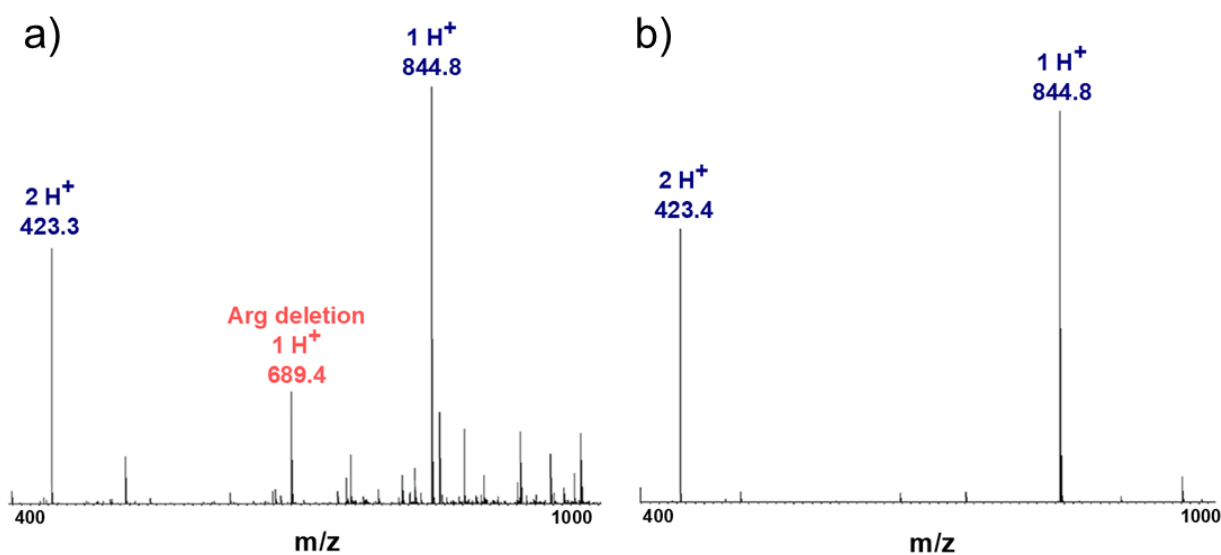
**Scheme 3.2.** Fmoc-SPPS assembly of the N-terminal fragments through MeDbz linker. Sequence spacer composed of LDA amino acids has been first coupled followed by the Fmoc-MeDbz linker. Then, the different N1 and Ncap residues have been coupled. Reaction with *p*-nitrophenyl chloroformate followed by basic conditions has allowed formation of the MeNbz moiety. The desired peptide has been cleaved from the resin in presence of TFA. The grey box represents a frequent by-product of the synthesis.

Syntheses of the 19 N-terminal peptide fragments are recapitulated in **Table 3.3**, and synthesis of the peptide containing the urea derivative will be discussed hereinafter in next subsection. The last entry of the table corresponds to the synthesis of the GA dipeptide derivative that was indeed previously performed by Boc-SPPS for synthesis presented in Chapter 4. Seven syntheses proceeded well with satisfying yield and did not require any purification before ligation. Although small amount of by-products was generally detected, the latter could be easily isolated from the final ligated product by SPE (*i.e.* elution conditions are highly distinct for by-products from short N-terminal fragments and full-length protein of 48 residues). Some sequences necessitated purification before ligation reaction due to either partial deletion of amino acids after the MeDbz linker or the considerable presence of by-product identified as the spacer sequence of LDA linked to the MeNbz moiety (represented in the grey box of **Scheme 3.2**). In the case of partial deletion, double coupling after the linker

did not prevent from incomplete reaction for few amino acids. This was mainly observed for amino acids that are known to be harder to couple such as Arg or the  $\beta$ -branched Ile. The deletion-containing by-products are not desired for further ligation step as they would also participate in the reaction and the resulting ligated by-product would be impossible to separate from the expected protein by only a solid phase extraction. This by-product was well removed by preparative HPLC purification as shown for the fragment Ac-RR-MeNbz-LDA in **Figure 3.4**. On the other hand, we postulated that the spacer sequence of LDA linked to the MeNbz moiety as by-product may arise from no coupling on the linker or from partial and spontaneous hydrolysis of the Ac-Ncap-N1-MeNbz-LDA peptide moiety. Despite

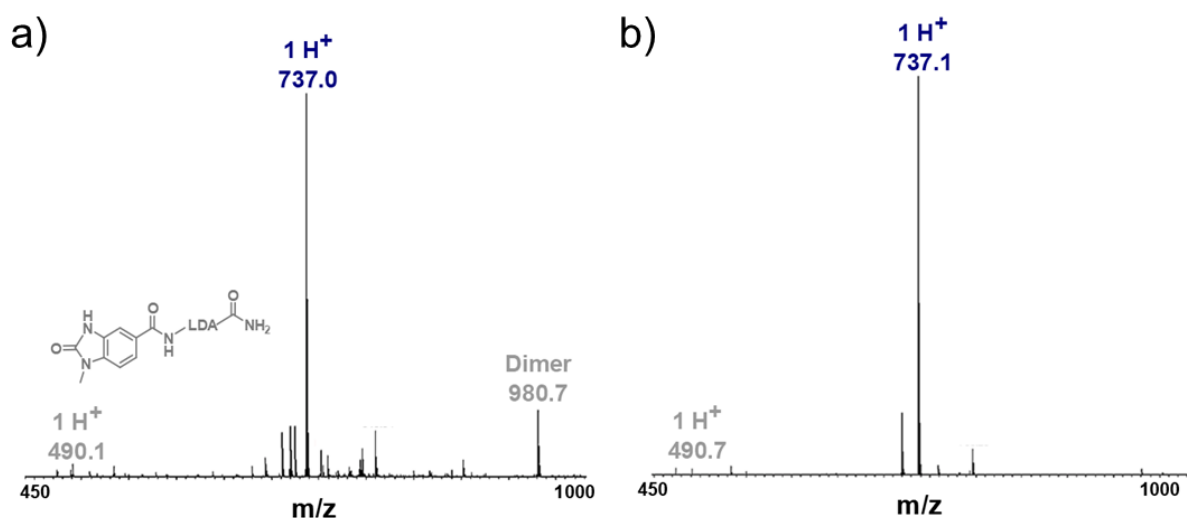
**Table 3.3. List of the N-terminal fragments synthesized by Fmoc-SPPS in the form of Ac-Ncap-N1-MeNbz-LDA peptide.** nd: not determined; no yield in % was calculated as the peptide was not isolated.

Peptide-MeNbz	Purification (Y/N)	MW calc. (Da)	MW obs. (Da)	Yield in % (Yield in mg)
Ac-SR-MeNbz-LDA	Y	775.8	775.5	45 (17.4)
Ac-RR-MeNbz-LDA	Y	844.9	844.3	23 (9.6)
Ac-RS-MeNbz-LDA	Y	775.8	775.5	36 (14)
Ac-GR-MeNbz-LDA	Y	745.8	745.4	27 (10.2)
Ac-RL-MeNbz-LDA	Y	801.9	801.5	21 (8.6)
Ac-RA-MeNbz-LDA	Y	759.8	759.4	34 (12.9)
Ac-NR-MeNbz-LDA	Y	802.8	802.4	28 (11.3)
Ac-GI-MeNbz-LDA	Y	702.8	702.0	17 (5.8)
Ac-GV-MeNbz-LDA	Y	688.7	688.1	22 (7.6)
Ac-GF-MeNbz-LDA	Y	736.8	736.1	30 (10.9)
Ac-GY-MeNbz-LDA	Y	752.8	752.1	17 (6.3)
Ac-GW-MeNbz-LDA	Y	775.8	775.2	9 (3.6)
Ac-GH-MeNbz-LDA	N	726.7	726.4	nd (32.9 mg of crude)
Ac-HL-MeNbz-LDA	N	782.9	782.3	nd (33.3 mg of crude)
Ac-NH-MPA-LDA	N (Boc-SPPS)	697.8	697.1	nd (117.3 mg of crude)
Ac-HS-MeNbz-LDA	N	756.8	756.4	nd (34.2 mg of crude)
Ac-EL-MeNbz-LDA	N	774.8	774.0	nd (28.8 mg of crude)
Ac-YL-MeNbz-LDA	N	808.9	808.0	nd (32.1 mg of crude)
Ac-EY-MeNbz-LDA	N	824.8	824.1	nd (32.8 mg of crude)
Ac-GA-MPA-LRY	Y (Boc-SPPS)	707.8	707.6	78 (113.8)



**Figure 3.4.** LC/MS spectra of the Ac-RR-MeNbz-LDA peptide. **a)** Crude peptide exhibiting an arginine deletion (in red). **b)** The peptide product after purification. The by-product from Arg deletion was removed.

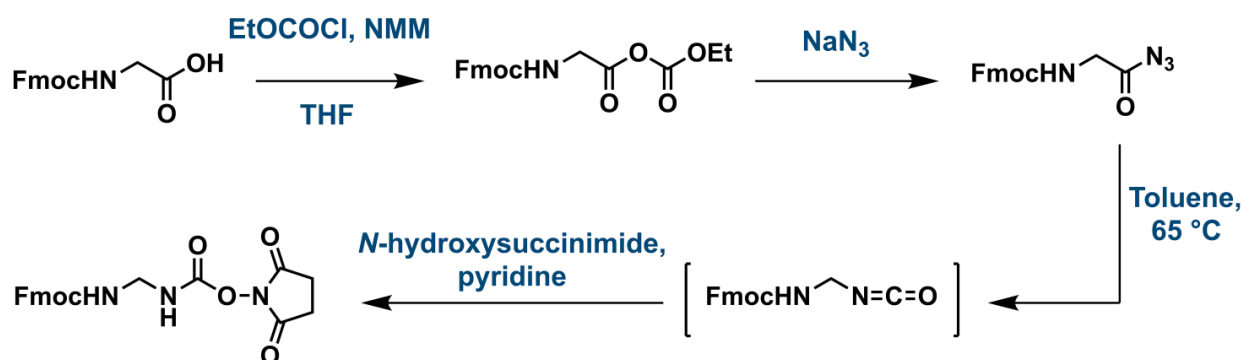
different reverse phase used for HPLC purification (*i.e.* C18, C8 and C4), the by-product was often co-eluting with the desired Ac-Ncap-N1-MeNbz-LDA peptide leading to partially pure product as shown in **Figure 3.5** with peptide Ac-GF-MeNbz-LDA. Nevertheless, this persistent impurity in such low amount would not be an obstacle for the following step because it would be unable to react in chemical ligation and is actually a regular by-product of the reaction as it would be released in the course of the reaction through thioesterification of the dipeptide. Then, the spacer sequence of LDA linked to the MeNbz moiety would be readily separated from ligation product as they possess both distinct properties and elution profiles in C18 reverse phase of the SPE.



**Figure 3.5.** LC/MS spectra of the Ac-GF-MeNbz-LDA peptide. **a)** Crude peptide exhibiting the spacer sequence of LDA linked to the MeNbz moiety (shown in grey) as by-product. **b)** The peptide product after purification. The by-product was still slightly present.

Interestingly, this frequent by-product consisting of spacer sequence of LDA linked to the MeNbz element was more present with peptides containing histidine residues. In the course of the purification, the amount of this by-product became even more significant resulting in the recovery of a very low amount of the desired peptide product. Thus, we associated that the formation of the by-product occurs via hydrolysis. To synthesize His-containing analogues, we opted for the following strategy: after cleavage and work up, the Ac-Ncap-N1-MeNbz-LDA peptide was solubilized in H<sub>2</sub>O/MeCN [1:1] + 0.1 % TFA and beads were immediately filtrated and peptide was directly flash frozen in liquid nitrogen for lyophilization. The peptide was not purified to avoid spontaneous hydrolysis and was directly utilized and solubilized for *in situ* thioesterification followed by ligation. Nevertheless, even with this protocol the Ac-NH-MeNbz-LDA peptide collapsed via hydrolysis in a rather large amount. That is why this analogue was synthesized by Boc-SPPS providing a peptide-<sup>α</sup>thioester in the form of Ac-NH-MPA-LDA (MPA stands for 3-mercaptopropionic acid).

**Synthesis of the glycine succinimidyl carbamate building block for urea linkage and peptide synthesis.** Synthesis of Fmoc protected glycine succinimidyl carbamate (Fmoc-Gly-CO-OSu) was performed following the four-step procedure reported by G. Guichard and coworkers (as depicted in [Scheme 3.3](#)).<sup>16</sup> Starting from Fmoc-Gly-OH, the carboxylic acid was first converted to a mixed anhydride by reaction with ethyl chloroformate in presence of *N*-methylmorpholine (NMM). The resulting mixed anhydride was then transformed to the relative acyl azide in presence of sodium azide (NaN<sub>3</sub>). Without any intermediate purification, the acyl azide underwent spontaneous Curtius rearrangement upon heating in toluene. In a one-pot fashion, the resulting isocyanate was treated with *N*-hydroxysuccinimide leading to the formation of the carbamate moiety of the resulting Fmoc-Gly-CO-OSu building block with 68 % of overall yield. The building block was used without any further purification due to suitable purity according to HPLC, <sup>1</sup>H and <sup>13</sup>C NMR analyses.



**Scheme 3.3. Synthetic route of the Fmoc-Gly-CO-OSu building block.** The Fmoc-Gly-OH was converted to a mixed anhydride followed by the formation of an acyl azide derivative. Through Curtius rearrangement, isocyanate was formed following by one-pot carbamate formation by reaction with *N*-hydroxysuccinimide.



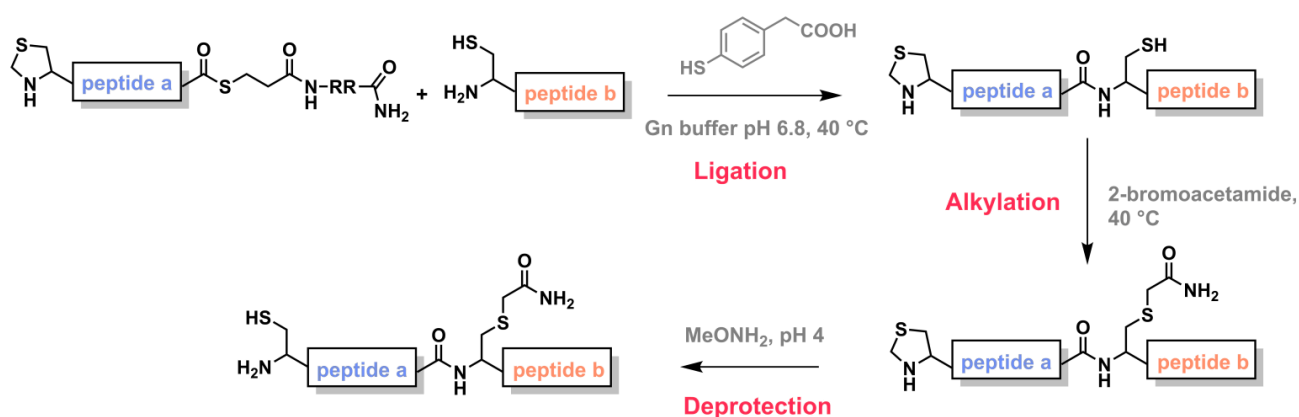
For solid phase Ac-**G<sup>14</sup>L**-MeNbz-LDA peptide synthesis, spacer sequence and Fmoc-MeDbz linker couplings were performed as described hereinabove. Two attempts were then carried out to couple the leucine and the glycine derivative forming a urea bond: *i*) triple coupling of each residue including an overnight coupling for the Fmoc-Gly-CO-OSu building block (coupling mixture for the Gly derivative was composed of 4 eq of Fmoc-Gly-CO-OSu dissolved in 1.4 mL DMF/DMSO [1:1] with 6 eq of DIEA) followed by N-terminal acetylation; *ii*) double coupling for both residue with time of reaction of 1 h for Fmoc-Leu-OH and 2 h for Fmoc-Gly-CO-OSu followed by N-terminal acetylation. However, both syntheses resulted in relatively noisy crude with only small amount of the desired product detected. Syntheses failed presumably due to the linker that decreases efficiency of the coupling and formation of the urea bond. Incorporation of the urea linkage on peptide derivative will be investigated again in the near future most probably by Boc-SPPS, which would be compatible with either Fmoc-Gly-CO-OSu or Boc-Gly-CO-OSu coupling.

#### **Synthesis of the C-terminal peptide segment common to all analogues.**

Synthesis of the C-terminal fragment for further ligation relies on the sequential assembly of the 46 residues by Boc-SPPS. However, the synthesis resulted in poor yield of less than 1 % and collected amount of protein (19.4 mg from a 0.4 mmol scale peptide synthesis) was not sufficient to constitute the entire library. A convergent approach was therefore applied consisting in splitting the full-length fragment in two at an Ala-Ala junction, synthesizing both peptide segments by Boc-SPPS and assembling them by native chemical ligation (NCL). This strategy resembles the one-pot three-segment ligation presented in Chapter 2, subsection 2.3.3 whereas here, second ligation to assemble Ncap and N1 residues (*i.e.* library of N-terminal fragments) is not directly performed in a one-pot fashion but after purification of the first ligation. Compared to the first synthetic approach employing sequential assembly by Boc-SPPS, an alanine residue (in the 24<sup>th</sup> position of the entire protein) is substituted by a pseudo homoglutamine (*i.e.* alkylated cysteine by 2-bromoacetamide). Catalytic assay with **Hom-N2** carrying either the original alanine or the alkylated cysteine showed no significant effect of this modification on catalysis (see subsection 2.6.3 of Chapter 2). Therefore, C-terminal fragment obtained by both approaches can be utilized for NCL with short N-terminal fragments to assemble the library of analogues.

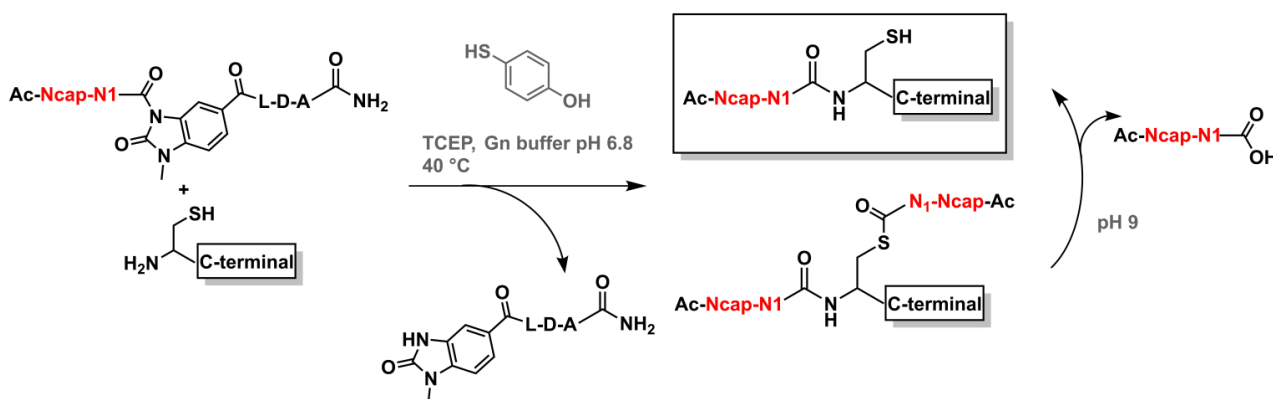
The second approach to assemble the C-terminal fragment is illustrated in **Scheme 3.4**. Both peptide **a**- $\alpha$ thioester and peptide **b** were synthesized using 'in situ neutralization' Boc/benzyl-SPPS at a 0.4 mmol scale. Peptide **a** was assembled using coupling of 3-mercaptopropionic acid (MPA) to form a peptide- $\alpha$ thioester on the resin and the N-terminal cysteine (*i.e.* the future N2 cysteine) was protected via a thiazolidine moiety. Syntheses and purifications of both peptide **a** and **b** proceeded properly resulting in large amount of pure

protein with isolated yield of 12 % (136.3 mg obtained) and 8 % (87.8 mg), respectively. Through NCL, peptide **a**-<sup>α</sup>thioester and Cys-peptide **b** were ligated together in presence of an aryl thiol, MPAA, as catalyst. The resulting cysteine at the junction site was capped by alkylation with 2-bromoacetamide. The excess of alkylating reagent was then quenched with a large excess of MPAA followed by the deprotection of thiazolidine group to give the free cysteine that would be used for further NCL to ligate the different N-terminal fragments composing the library. After HPLC purification of the reaction mixture, the C-terminal fragment was recovered with 32 % to 43 % of ligation yield (42 mg of isolated peptide in total).



**Scheme 3.4.** Synthesis of the C-terminal fragment by a ligation approach. Native Chemical Ligation (NCL) was performed between peptide **a**-<sup>α</sup>thioester and peptide **b** in presence of MPAA as catalyst. The remaining cysteine was capped by alkylation. Subsequently, the thiazolidine protecting group was removed and the resulting C-terminal fragment was isolated by HPLC purification.

**Ligation of the different N-terminal fragments with the C-terminal fragment to construct the library of analogues.** Native chemical ligations (NCL) were then performed to covalently join the diverse synthesized N-terminal fragments to the invariant Cys-C-terminal fragment. Reaction plan and reaction outcome are described in **Scheme 3.5**. For NCL, the N-terminal fragments in the form of Ac-**Ncap-N1**-MeNbz-LDA need to be converted *in situ* into dipeptide-<sup>α</sup>thioester before ligation of the two peptide segments. Generally, aryl thiols are used as a catalyst because aryl-thiolate is an efficient leaving group during the transthioesterification step of the NCL. Although 4-mercaptophenylacetic acid (MPAA) is commonly used as NCL catalyst, it is typically retained on C18 reverse phase and may elute in the same range as peptides. Thus, we selected a less hydrophobic aryl thiol, the 4-mercaptophenol (MPOH), to facilitate extraction of the ligated product on solid C18 phase. *N*-acylurea peptide (or thioester peptide for GA and NH sequences) was introduced in slight excess to consume the full quantity of C-terminal fragment that would elute too close to final product to be separated in SPE. *N*-acylurea peptide as well as thioester peptide underwent rapid thiolysis with MPOH leading to the release of the spacer sequence of LDA linked to the



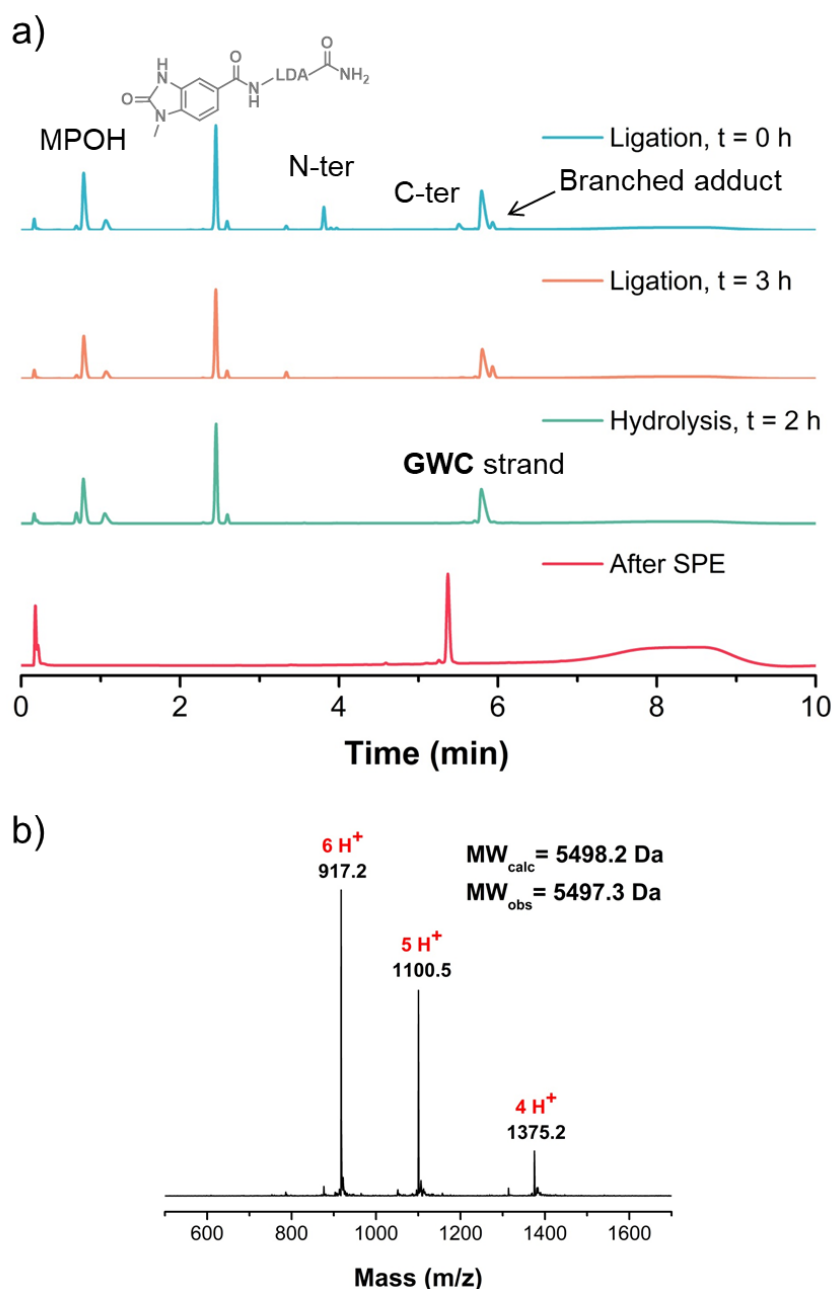
**Scheme 3.5.** Ligation reaction to assemble the library of 'negative' strand variants. Ac-Ncap-N1-MeNbz-LDA has been thioesterified *in situ* by MPOH leading to the release of the linker with the spacer sequence. Then, the resulting thioester has undergone NCL with the C-terminal peptide fragment giving the product represented in the black box. Excess of thioester generally reacts with the free cysteine through transthioesterification but the desired product can be recovered at pH 8.5.

MeNbz moiety or to the MPA, respectively. The resulting peptide- $\alpha$ thioester instantaneously reacted with the Cys-C-terminal peptide through NCL. The different ligation reactions were monitored by analytical LC/MS and HPLC (see an example in [Figure 3.6](#) with the synthesis of the **GWC** strand variant) and completion of the reaction were reached in less than one hour.

Interestingly, in the majority of ligation mixtures, the presence of a branched adduct emerged as a consequence of the transthioesterification between catalytic cysteine of the full length strand and excess of thioester peptide. As thioester groups are unstable under basic conditions, a pH adjustment to around 8.5-9 resulted in a rapid hydrolysis of such a branched adduct. Finally, solid-phase extraction (SPE) allowed for an efficient separation of expected ligated peptides from the reaction reagents (e.g. MPOH, TCEP) and by-products (e.g. linker with the spacer sequence) that eluted at CH<sub>3</sub>CN/H<sub>2</sub>O [3:7] whereas ligated product eluted at CH<sub>3</sub>CN/H<sub>2</sub>O [5:5]. Following this procedure, 20 analogues of full-length 'negative' strand were assembled (the list is summarized in [Table 3.4](#)). Some were synthesized using the C-terminal fragment with the original alanine residue and some with the alkylated cysteine, a modification that does not affect the ligation reaction and should not alter further catalytic assays.

**Synthesis of the library of 'positive' strand variants.** The four new analogues of 'positive' strand were synthesized by Boc-SPPS through sequential assembly of the full-length peptides. Starting from the same batch on 0.4 mmol scale, the resin was sequentially split to obtain the various peptides on 0.1 mmol scale following the process shown in [Figure 3.7](#). The shared sequence fragment 37-48 was synthesized from the 48<sup>th</sup> to the 37<sup>th</sup> residues in a common batch on 0.4 mmol scale. The resin was then divided in two to assemble on one side the full length 'positive, H1E36' strand on a 0.1 mmol scale and on the other side the

amino acids from the 36<sup>th</sup> to the 3<sup>rd</sup> position on a 0.3 mmol scale leading to a peptide fragment 3-48. The resin containing the latter was subsequently divided in three equal parts to finalize the synthesis of the different 'positive, G1', 'positive, H1' and 'positive, H1H2' strands by coupling the two last residues on a 0.1 mmol scale each. Thus, the four strands comprising the library of 'positive' strands were simultaneously obtained with yields around 2 % (between 7.8 and 11.4 mg of isolated peptide). Although purity of the analogues was not as good as achieved before for **Het-N2**, **Het-Ncap** and **Hom-N2** probably due to preparative column deterioration, it was judged adequate for further catalytic characterization as for the 'negative' strand library.



**Figure 3.6. Example of ligation with the GWC strand analogue.** a) HPLC monitoring of the ligation between Ac-GW-MeNbz-LDA peptide (N-ter) and the C-terminal fragment (C-ter) in the presence of MPOH for *in situ* thioesterification of the N-terminal peptide. b) LC/MS spectrum of the **GWC** strand after solid-phase extraction (SPE).

**Table 3.4. List of the synthesized ‘negative’ strand variants after ligation of the N-terminal and the C-terminal peptide fragments.** † Yields of ligation were estimated when the resulting lyophilized peptide (too low to be weighted) was solubilized in buffer to obtain stock solutions for catalytic assay. Concentration of the resulting solution was estimated with calibration curves on analytical HPLC (calculated from **Het-N2** strands) or with Nanodrop for analogues containing aromatic residues. Amount of peptides and yields were then deduced.

<b>Name of the synthesized ‘negative’ stand</b>	<b>C-terminal fragment containing an alanine or an alkylated cysteine residue</b>	<b>Ligation yield†</b>
<b>SRC</b> strand	alkylated Cys	19 %
<b>RRC</b> strand	Ala	16 %
<b>RSC</b> strand	Ala	22 %
<b>GRC</b> strand	Ala	21 %
<b>RLC</b> strand	Ala	21 %
<b>RAC</b> strand	Ala	21 %
<b>NRC</b> strand	Ala	24 %
<b>GAC</b> strand	alkylated Cys	25 %
<b>GIC</b> strand	Ala	14 %
<b>GVC</b> strand	alkylated Cys	10 %
<b>GFC</b> strand	Ala	20 %
<b>GYC</b> strand	alkylated Cys	20 %
<b>GWC</b> strand	alkylated Cys	20 %
<b>GHC</b> strand	alkylated Cys	27 %
<b>HLC</b> strand	alkylated Cys	23 %
<b>NHC</b> strand	alkylated Cys	29 %
<b>HSC</b> strand	alkylated Cys	27 %
<b>ELC</b> strand	alkylated Cys	21 %
<b>YLC</b> strand	alkylated Cys	23 %
<b>EYC</b> strand	alkylated Cys	23 %

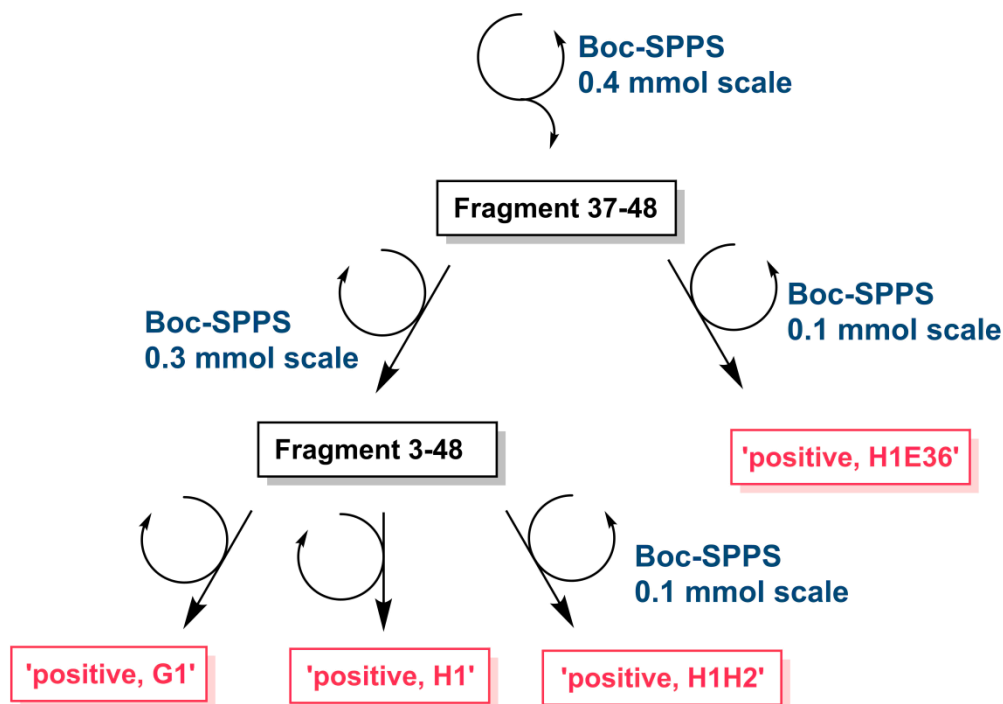


Figure 3.7. Synthetic strategy to obtain four different analogues of the 'positive' strand.

### 3.2.4. Catalytic assays of the libraries of analogues

The libraries of 'negative' and 'positive' strands were evaluated separately for their catalytic activity using the assay of the 2<sup>nd</sup> generation with LC/MS monitoring, which has been previously presented in subsection 2.6.4. To insure consistency of the acquired kinetic parameters for the different analogues, catalytic assays were all incubated at 25 °C in a PCR thermocycler using the incubation mode. For each analogue obtained by mixing the corresponding 'negative' and 'positive' strands at a concentration of 100 µM, reaction between peptide-Phe-<sup>α</sup>thioester (*c* 200 µM) and Tris (*c* 200 mM) was monitored by LC/MS and observed rate constant  $k_{\text{obs}}$  as well as the aminolysis/hydrolysis (A/H) were deduced.

**Catalytic evaluation of the library of 'negative' strands.** The different analogues of the library **Het-AA(Ncap)AA(N1)C** were obtained by mixing the corresponding **AA(Ncap)AA(N1)C** strand with the original 'positive' strand that composed also the **Het-N2** protein. Resulting  $k_{\text{obs}}$  and A/H for the 20 analogues are compared with kinetic parameters of **Het-N2** and are summarized in **Figure 3.8**.

Although no notable acceleration or substantial changes were observed while analyzing the library, some tendencies were emphasized. Interestingly, catalytic activity of

the standard **Het-N2** was one of the greatest. By modifying the hydrophobic residue at N1 position (orange, green and red bars in **Figure 3.8**), observed rate constants and A/H ratio were comparable to **Het-N2** protein except for the variant **Het-GAC** (orange bar) with a less hydrophobic side chain that provided a lower catalytic activity ( $k_{\text{obs}} = 0.044 \text{ M}^{-1}\cdot\text{s}^{-1}$  compared to  $0.065 \text{ M}^{-1}\cdot\text{s}^{-1}$  for **Het-N2**). Thus, we have supposed that a hydrophobic residue at N1 position may have an importance on catalysis by contributing in good packing that positions the catalytic cysteine. Remarkably, tyrosine and tryptophan amino acids bearing a heteroaromatic ring (red bars in **Figure 3.8**) favored aminolysis, which has become the main outcome of the reaction (61 % and 55 % of aminolysis, respectively).

However, when histidine residue was introduced at N1 position (blue bars in **Figure 3.8**), hydrolysis was more favored compared to the **Het-N2** protein (from 60 % with **Het-N2** to around 70 % for histidine at N1). Histidine residues are a common residue assisting the catalytic cysteine in either ligase or protease, and the reason why the catalytic site favored one pathway over the other is still not clearly understood. In our case, more experiments such as structural characterization as well as deeper catalytic studies would be necessary to better understand the influence of histidine residue on hydrolysis catalysis. Nevertheless, histidine incorporated at Ncap position (dark blue bars **Figure 3.8**) provided approximately similar A/H ratio in comparison to **Het-N2**. Interestingly, the analogue **Het-HLC** presented

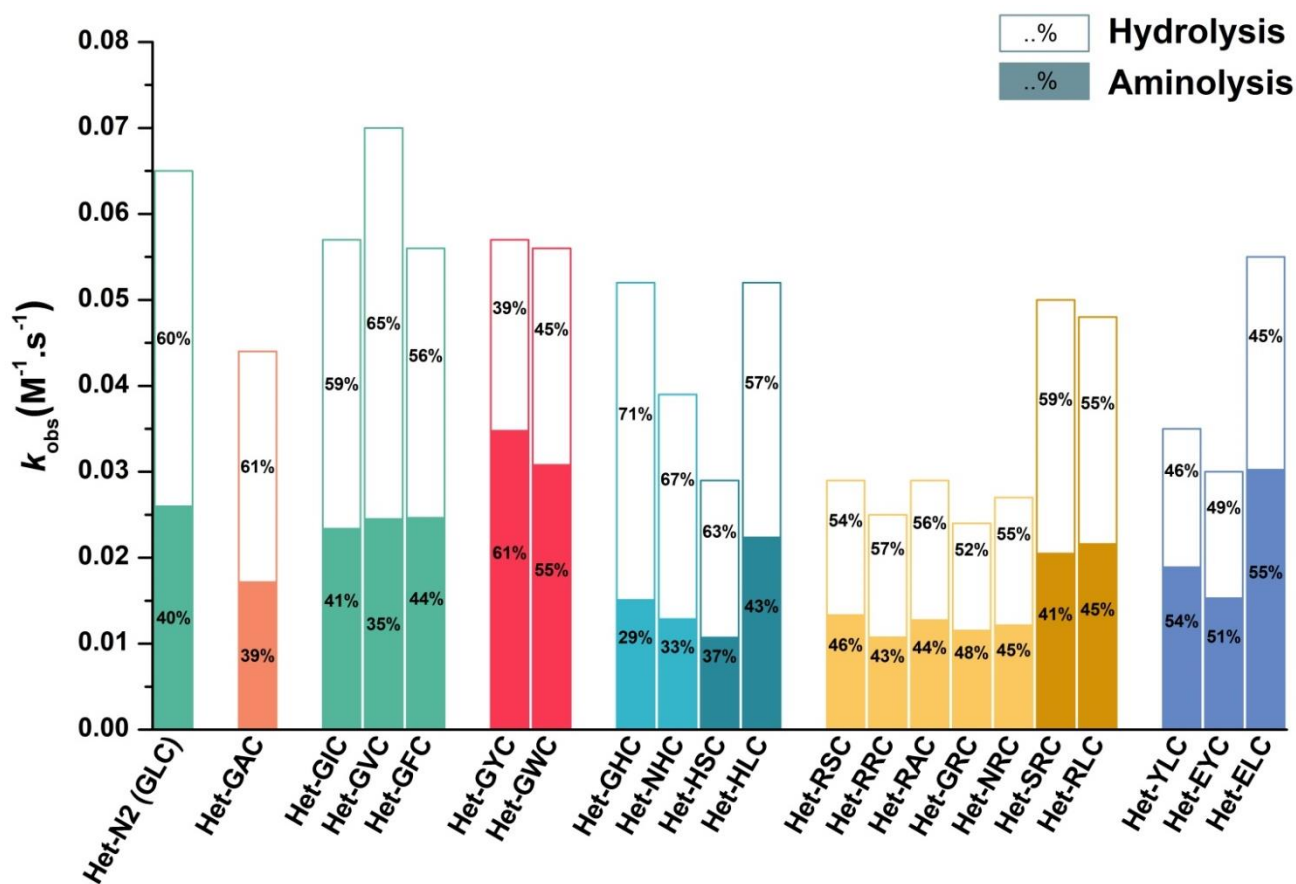


Figure 3.8. Observed rate constant and percentage of aminolysis and hydrolysis induced by the different analogues of the library containing heterodimers with diverse 'negative' strands.

higher rate of reaction than **Het-HSC** ( $k_{\text{obs}} = 0.052 \text{ M}^{-1}\cdot\text{s}^{-1}$  and  $0.029 \text{ M}^{-1}\cdot\text{s}^{-1}$ , respectively) suggesting again that leucine (and maybe also other hydrophobic residues) at N1 position may positively influence catalytic rate.

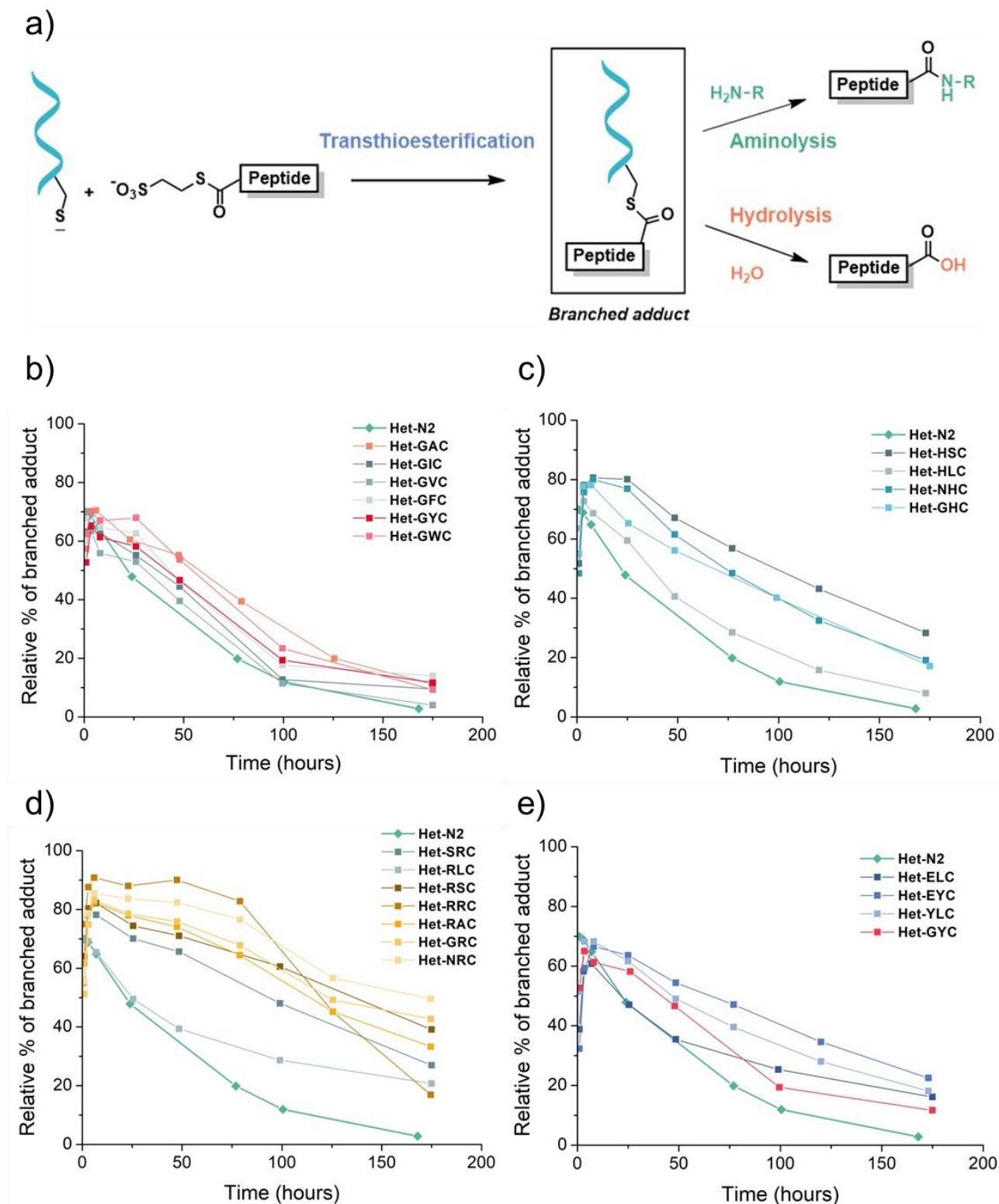
Surprisingly, arginine residue at either Ncap or N1 position (yellow bars **Figure 3.8**) mostly induced much lower rates compared to the original **Het-N2**, whereas **Het-SRC** and **Het-RLC** (brownish yellow bars **Figure 3.8**) provided almost similar observed rate constant  $k_{\text{obs}}$  compared to **Het-N2** ( $k_{\text{obs}} = 0.050 \text{ M}^{-1}\cdot\text{s}^{-1}$  and  $0.048 \text{ M}^{-1}\cdot\text{s}^{-1}$ , respectively, compared to  $0.065 \text{ M}^{-1}\cdot\text{s}^{-1}$  for **Het-N2**). According to the study published by W. F. DeGrado and coworkers on phosphoserine switch, Arg residue at Ncap or N1 position can supply stabilizing hydrogen bonding with phosphoserine at N2 position of  $\alpha$ -helix.<sup>8</sup> They demonstrated that serine at Ncap position and arginine at N1 position induced the most stabilizing effect as expected thanks to adapted geometries to such position and favorable interaction with the oxyanion. With our catalytic analogues, we also established that our variant **Het-SRC** induced fastest rate compared to the other analogues carrying an arginine residue at Ncap or N1 position, which may originate from better tetrahedral intermediates stabilization and compatible geometries with Ncap and N1 position. Here again, analogue containing a leucine residue at N1 position, **Het-RLC**, presented faster rate compared to the majority of other Arg analogues.

Finally, analogues presenting residues found in catalytic site of N-terminal acetyltransferases such as tyrosine and glutamate (purple bars **Figure 3.8**) slightly favored aminolysis (from 51 to 55 % of aminolysis observed), but possessed heterogeneous  $k_{\text{obs}}$ : the rates observed for **Het-YLC** ( $k_{\text{obs}} = 0.035 \text{ M}^{-1}\cdot\text{s}^{-1}$ ) and **Het-EYC** ( $k_{\text{obs}} = 0.030 \text{ M}^{-1}\cdot\text{s}^{-1}$ ) corresponded to around half of the one generated in the presence of **Het-N2** ( $k_{\text{obs}} = 0.065 \text{ M}^{-1}\cdot\text{s}^{-1}$ ), whereas **Het-ELC** ( $k_{\text{obs}} = 0.055 \text{ M}^{-1}\cdot\text{s}^{-1}$ ) showed almost similar observed rate constant compared to **Het-N2**.

From all these results, we have deduced that hydrophobic residue at N1 position is essential for providing the most efficient catalytic rate and some residues like tyrosine, tryptophan or glutamate can influence the A/H ratio in favor to aminolysis.

Interestingly, the rate of consumption of the branched adduct on catalytic protein (*i.e.* resulting from transthioesterification between peptide- $\alpha$ thioester substrate and the catalytic cysteine of the different variants) via aminolysis or hydrolysis correlated well with the global  $k_{\text{obs}}$  of the reaction (**Figure 3.9**). The original GLC-analogue (**Het-N2**) gave one of the highest global rate, and the consumption of branched adduct was the fastest observed. Moreover, evolution of branched adduct was very similar for all analogues containing a hydrophobic residue at N1 position (**Figure 3.9b**). **Het-N2** and **Het-GVC** analogues induced actually the highest  $k_{\text{obs}}$  and their consumptions of branched adduct were among the fastest.





**Figure 3.9. Evolution of the branched adduct for the different analogues of the library of ‘negative’ strands. a)** Catalytic pathway illustrating the formation of branched adduct through transthioesterification and consumption of this branched adduct via either aminolysis or hydrolysis. **b)** Branched adduct evolution for analogues containing a hydrophobic residue at N1 position. **c)** Branched adduct evolution for His-analogues in comparison to **Het-N2**. **d)** Branched adduct evolution for Arg-containing analogues in comparison to **Het-N2**. **e)** Branched adduct evolution for analogues containing a residue generally found in catalytic site of N-terminal acetyltransferases in comparison with **Het-N2**.

Generally, the maximum percentage of observed branched adduct was higher for His-analogues (around 80 %, whereas maximum of 70 % of branched adduct with **Het-N2**) as shown in **Figure 3.9c**. The observed rate constant for His-analogues was slightly lower than the original **Het-N2** analogue and consumption of the branched adduct was slightly slower too, which is even more noticeable for **Het-HSC** with reduced overall rate.

With arginine containing analogues, observed branched adduct attained higher maximum percentage (*i.e.* between 80 to 90 %, whereas maximum of 70 % of branched adduct with **Het-N2**) as depicted in **Figure 3.9d**. For Arg-analogues with lower overall  $k_{\text{obs}}$  (in brown and yellow in **Figure 3.9d**), consumption of the branched adduct was also slower suggesting a higher stabilization of the branched adduct or a less accessible electrophilic center and/or catalytic site.

Finally, evolution of branched adduct of analogues containing tyrosine or glutamate residues at Ncap or N1 position (*i.e.* residues typically found in catalytic site of N-terminal acetyltransferase) was globally similar and comparable with **Het-N2**. Here again, **Het-EYC** and **Het-YLC** proteins that presented lower overall  $k_{\text{obs}}$  showed a slightly slower consumption of the branched adduct.

Consumption of the branched adduct is intimately related to the global rate of the reaction as it corresponds to the rate-limiting step. Modification of the surrounding environment of the catalytic cysteine induces only slight changes in the overall kinetics. Some modifications generated lower rates coming from a slower consumption of the branched adduct. Some other modifications promoted one or another pathways of branched adduct consumption (*i.e.* aminolysis or hydrolysis). Further catalytic characterization as well as detailed structural characterization will be crucial for a better understanding on the influence of such modifications. Most probably, more drastic changes would be necessary to reaching much higher catalytic activity.

**Preliminary catalytic evaluation of the library of 'positive' strands.** To obtain the four other analogues differing from their 'positive' strand, the **Cys-N2** strand from **Het-N2** and the various designed 'positive' strand (see **Table 3.2**) were mixed together. However, the following preliminary results are coming from an incomplete set of data, meaning that the last aliquot was taken before completeness of the reaction. Thus, observed rate constants were deduced as more than the half of peptide- $\alpha$ thioester was consumed but the ratio of aminolysis over hydrolysis (A/H) was only estimated and calculated after 73 hours of reaction, which may be slightly different to the ratio at the end of the reaction. The measured kinetic parameters are gathered in **Table 3.5**.

**Table 3.5. Preliminary kinetics parameters for the library of heterodimer with different ‘positive’ strands in comparison to the original Het-N2.** † corresponds to A/H ratio calculated after 73 hours of reaction, before completeness of the reaction.

Name of the analogue	$k_{\text{obs}}$ ( $\text{M}^{-1}\cdot\text{s}^{-1}$ )	A/H ratio
<b>Het-N2</b>	0.065	0.67
<b>Het-CysN2/‘positive, G1’</b>	0.069	0.60 <sup>†</sup>
<b>Het-CysN2/‘positive, H1’</b>	0.063	0.54 <sup>†</sup>
<b>Het-CysN2/‘positive, H1H2’</b>	0.060	0.60 <sup>†</sup>
<b>Het-CysN2/‘positive, H1E36’</b>	0.063	0.82 <sup>†</sup>

The observed rate constants coming from the library of heterodimers with diverse ‘positive’ strands are very similar to the  $k_{\text{obs}}$  of the original **Het-N2**. Therefore, modifications on the second strand that is not containing the catalytic cysteine seem to not alter the rate of the reaction significantly, which may be due to rather large distance between modifications and the cysteine. By modifying the Asn residue to Gly at Ncap position (1<sup>st</sup> amino acid in the sequence), we desired to remove potential bulkiness of the asparagine that faced to oxyanion hole according to crystal structure. As virtually no change in the overall observed rate constant was detected for this modification, asparagine may indeed not obstruct the catalytic site in solution and may not influence kinetic parameters of the reaction. Although, here, the calculated A/H ratio does not correspond to the final one at the end of the reaction, it allows estimating the tendency of the reaction outcome. Typically, the A/H ratio in presence of the new synthesized analogues stayed approximately identical in comparison to the **Het-N2** protein. These results are rather in the range of error of the measurements, still, the following subtle propensities were observed: **Het-CysN2/‘positive, G1’**, **Het-CysN2/‘positive, H1’** and **Het-CysN2/‘positive, H1H2’** variants favored slightly more hydrolysis than the standard **Het-N2**, whereas **Het-CysN2/‘positive, H1E36’** induced a larger amount of aminolysis compared to **Het-N2**. Though the effect is indeed not substantial, the last analogue possesses a glutamate residue in the hydrophobic core in close proximity of a histidine that may assist in the  $pK_{\text{a}}$  depression of the His residue or in proton transfer to deprotonate the intermediate. The catalytic activity of the four analogues carrying different ‘positive’ strands will be evaluated again to confirm such catalytic tendency.

Finally, no substantial acceleration was observed for the diverse designed analogues containing either different Cys-‘negative’ strand or ‘positive’ strand. Nevertheless, modest influences on catalytic activity such as reduced rate or modulation of the A/H ratio were detected and are good starting point for further modifications. Thus, it would be interesting to combine certain modifications or even certain ‘negative’ and ‘positive’ strand variants

together. Further catalytic studies as well as structural characterization will be necessary to better understand such influences.

### 3.3. Library of substrates

Careful catalytic characterization of the first generation of analogues and more particularly of **Het-N2** has revealed intricate competition between aminolysis and hydrolysis, the two major outcomes of acyl transfer reaction on a peptide- $\alpha$ thioester. With our *de novo* catalytic protein, both pathways were always detected and aminolysis/hydrolysis (A/H) ratio was dependent on several external factors. In the same way, our designed protein demonstrated catalytic activity for a diverse set of acyl transfer reactions, but resulting rates of thioester consumption given by the observed rate constant  $k_{\text{obs}}$  were also contingent on reaction conditions. Among the elements influencing the reaction outcome, nature of the acyl donor (*i.e.* the peptide- $\alpha$ thioester) and of the acyl acceptor (*i.e.* the amine) play an important role and drastic changes in kinetic parameters were observed. For instance, in presence of the best DSD-variant, **Het-N2**, and a large excess of Tris, a higher rate of reaction and a higher A/H ratio were observed for peptide2-Arg- $\alpha$ thioester compared to peptide-Phe- $\alpha$ thioester ( $k_{\text{obs}} = 0.143 \text{ M}^{-1} \cdot \text{s}^{-1}$  versus  $0.082 \text{ M}^{-1} \cdot \text{s}^{-1}$ , respectively, and A/H = 2.03 versus 0.76, respectively). Interestingly, two distinct sequences that were studied for Arg thioester exhibited very similar results, therefore demonstrating the influence of the C-terminal residue of the peptide- $\alpha$ thioester substrate on kinetics with **Het-N2**.

To better understand key parameters that influence kinetics in our designed proteins, we have explored another strategy by synthesizing a library of peptide- $\alpha$ thioester substrates and a library of peptide substrates containing an N-terminal amino group. The first set of thioester substrates differs by their C-terminal residue, thus twenty peptides with the different canonical amino acids at last position have been synthesized. The second batch containing acyl acceptors has been produced in which the twenty natural amino acids were tested at the N-terminal position. The goal will be to examine the different substrates of the libraries in the optimized catalytic assay to extract the effectiveness of our catalytic protein depending on the nature of the substrates.

A common sequence for both libraries of peptide thioesters (acyl donors) and of acyl acceptors was first designed to adopt convenient properties, which were defined as follows: *i)* the sequence needs to contain hydrophobic residues to increase interactions with reverse

phase chromatography column, thus facilitating purifications and analyses; *ii*) polar and charged amino acids must also be incorporated to enhance solubility in aqueous buffer; *iii*) no amine is present in this sequence (except the N-terminal amine of acyl acceptors, and X = Lys) in order to not interfere with the aminolysis step of the assayed reactions. As a consensus sequence, we selected Ac-VALENX- $\alpha$ thioester, in which N-terminal amine was capped by acetylation and X corresponded to any natural amino acids for the acyl donor library (all sequences are shown in **Table 3.6**). We chose H<sub>2</sub>N-XVALEN- $\alpha$ amide for the acyl acceptor library (all sequences are shown in **Table 3.7**).

In this subsection, I will present the chemical synthesis of these libraries. Catalytic activity of our *de novo* designed proteins with the various substrates will be inspected in a near future.

### 3.3.1. Chemical synthesis of a library of peptide- $\alpha$ thioester substrates

**Solid-phase peptide synthesis.** The library of peptide- $\alpha$ thioesters was synthesized from their corresponding peptide- $\alpha$ hydrazides according to the method developed by L. Liu and coworkers.<sup>17</sup> Peptides were synthesized by Fmoc-SPPS using previously prepared 2-chlorotrityl-hydrazine (2-CT-NHNH<sub>2</sub>) resin. The first sequence (Ac-VALENR) was synthesized by manual SPPS on a 0.1 mmol scale, where control of coupling completion was done by chloranil test.<sup>18,19</sup> No double coupling was necessary according to the qualitative test and the synthesis reached 66 % (50 mg) of yield calculated from the crude material. The rest of the peptide library was prepared by machine-assisted SPPS enabling parallel syntheses and an optimized protocol consisting in double coupling for both first and last amino acids of each sequence was applied. However, yields of SPPS started to slowly decrease with time due to deterioration of resin. Hence, freshly prepared resin was utilized in the next syntheses, which improved the situation.

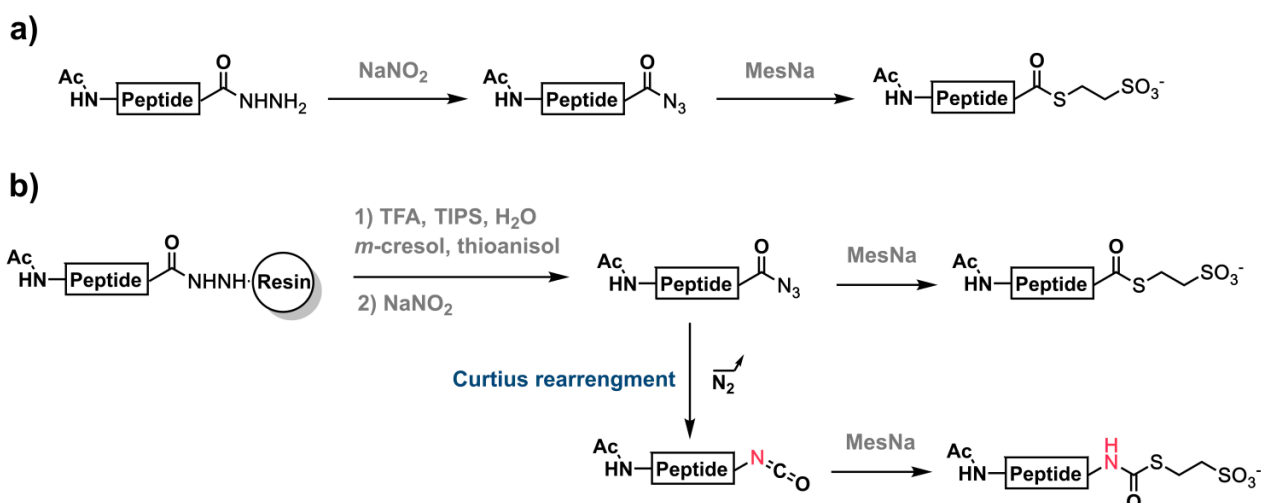
In addition, the methionine residue of Ac-VALENM peptide was almost totally oxidized to methionine sulfoxide and only traces of the desired peptide were observed by LC/MS analysis. Although simple reducing conditions (addition of dithiothreitol - DTT) and inert atmosphere during the cleavage step were employed, they were not sufficient owing to the high tendency of methionine to oxidize. The reaction is reversible but well-known reducing agents containing free thiols, like DTT, are not compatible with the enzymatic assays, unless a further step of purification is added, because thiols may compete with the catalytic cysteine present in the *de novo* catalytic protein. Thus, an alternative and selective reduction method

for methionine using  $\text{Me}_2\text{S}$  in  $\text{HCl}$  might be tested after thioesterification.<sup>20</sup> Otherwise, Met residue can simply be replaced by norleucine (Nle), which is nearly isosteric with methionine.

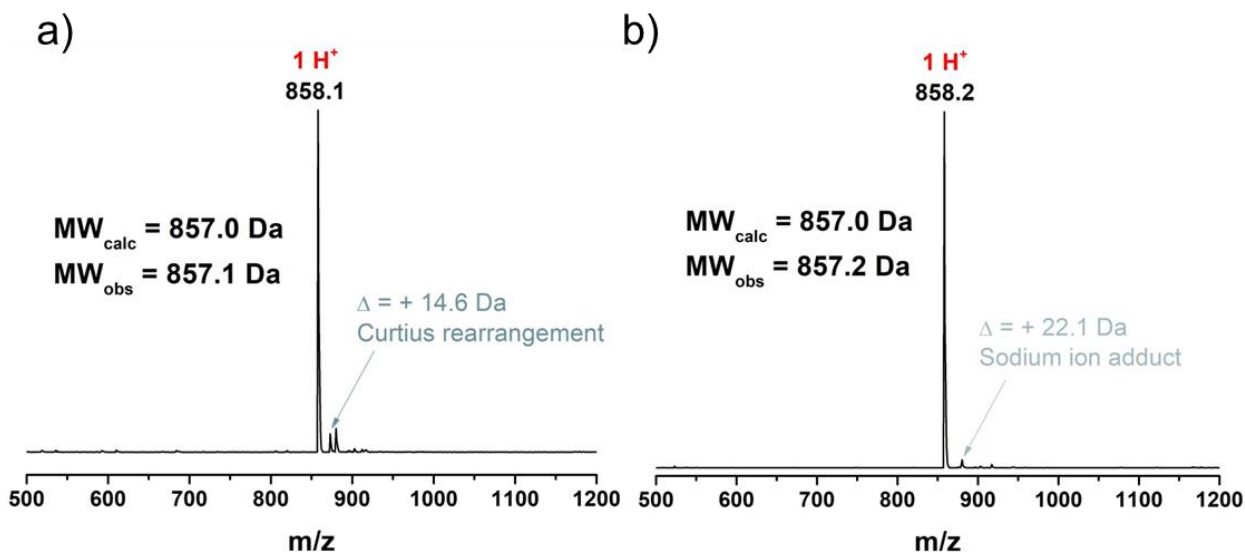
Furthermore, deprotection of tryptophan residue to remove the Boc group on the sequence Ac-VALENW was incomplete and a by-product with a mass of + 44 Da compared to the desired peptide was detected that corresponded to the remaining  $\text{CO}_2$  adduct. Completeness of the deprotection was finally achieved by incubating the peptide in  $\text{H}_2\text{O}$  + 0.1 % AcOH for 10 min and then lyophilizing it.

Otherwise, peptide- $\alpha$ hydrazides were generally obtained with acceptable purity, and thioesterification was directly performed on crude peptides.

**Thioesterification.** To form a thioester from the hydrazide surrogate, the method developed by L. Liu and coworkers starting from lyophilized crude or pure peptide via an acyl azide intermediate is the most commonly used in our laboratory.<sup>17</sup> More recently, the group of N. Mase adapted this method to directly produce acyl azide in the cleavage mixture (see **Scheme 3.6**).<sup>21</sup> We considered investigating this new procedure as it may reduce likelihood of hydrolysis and thioester can be obtained faster (*i.e.* one step of lyophilization is removed). VALENF was used as a model peptide to compare efficiency of both methods. Results showed that small amount of Curtius rearrangement took place with the second methodology and this side product was indeed not separable from the desired peptide by preparative HPLC (see **Figure 3.10a**); whereas the first procedure provided a peptide of higher purity without any notable side products (**Figure 3.10b**). For these reasons, the method of L. Liu was selected for the thioesterification of the rest of the library, which finally allowed



**Scheme 3.6. The two methods to obtain a peptide- $\alpha$ thioester from a peptide- $\alpha$ hydrazide.** **a)** The first procedure developed by L. Liu starts from lyophilized crude (or pure) peptide and goes through an acyl azide intermediate followed by direct thiolysis. **b)** The second methodology introduced by N. Mase consists in oxidation of the hydrazide surrogate in a one-pot fashion with cleavage. After precipitation of the acyl azide, the latter was solubilized in buffer in presence of a thiol allowing thioesterification of the peptide. However, this method favors Curtius rearrangement to a higher extent.



**Figure 3.10.** LC-MS analysis of the isolated peptide Ac-VALENF- $\alpha$ -thioester after HPLC purification. **a)** Following the procedure of N. Mase, Curtius rearrangement was observed and the desired peptide product was obtained with small amount of the resulting side product. **b)** Isolation of the peptide following L. Liu protocol.

performing all the reactions free from by-products resulting from Curtius rearrangement.

Following the first described method (**Scheme 3.6a**), transformation of peptide- $\alpha$ hydrazide to peptide- $\alpha$ thioester proceeded in straightforward fashion through oxidation and thioesterification steps. Efficiency and rate of the reaction were generally dependent on the C-terminal amino acid. After completion of the last step of thiolysis, pure peptides were isolated by preparative HPLC with variable yields of thioesterification (from 10 % to 49 %) as shown in **Table 3.6**. Nevertheless, hydrolysis of azide intermediate was very pronounced for the sequence Ac-VALEND when standard procedure was followed. Indeed, VALEND- $\alpha$ azide exhibited low stability and was completely hydrolyzed within five minutes presumably via intramolecular catalysis of the side chain carboxylate group of aspartic acid next to the acyl azide moiety. By decreasing the time between the addition of  $\text{NaNO}_2$  and MesNa, quantity of hydrolyzed peptide decreased while the quantity of desired thioester peptide increased. Around 50 % of each were obtained with one minute of the oxidation step directly followed by addition of the thiol MesNa. After HPLC purification, VALEND- $\alpha$ thioester was isolated with 16 % of thioesterification yield (see Entry 3 of **Table 3.6**). In this case, the alternative method of thioesterification developed by N. Mase would probably be helpful to overcome this problem by suppressing the formation of hydrolyzed peptide as in this method TFA is the main solvent during the oxidation step.<sup>21</sup>

### 3.3.2. Chemical synthesis of a library of acyl acceptor substrates

Synthesis of the library of peptides in the form of H<sub>2</sub>N-XVALEN-<sup>α</sup>amide (with X corresponding to the 20 different canonical amino acids) was carried out using standard Fmoc-SPPS procedure on Rink amide resin (0.1 mmol scale). The 20 peptide substrates were prepared by machine-assisted SPPS enabling multiple syntheses at the same time, and double coupling was programmed for the first residue (asparagine), valine residue and the last residue differing for each peptide. No noticeable issues were encountered. After cleavage all peptides presented acceptable purity that allowed for simple purification. Surprisingly, using the same condition of cleavage (*i.e.* 92.5 % TFA, 2.5 % DTT, 2.5 % TIPS, 2.5 % H<sub>2</sub>O) as for previously synthesized methionine-containing peptide-<sup>α</sup>hydrazide, the peptide MVALEN-<sup>α</sup>amide did not display any oxidation on the methionine residue. On the other hand, the peptide WVALEN-<sup>α</sup>amide still presented incomplete deprotection of Boc protecting group on tryptophan residue with recurrent CO<sub>2</sub> adduct even after longer cleavage reaction (*i.e.* 4 hours instead of generally 2 hours). Again, desired and completely deprotected peptide was retrieved by solubilizing it in a solution of water with 0.1 % of acetic acid followed by lyophilization. Peptides comprising the library of acyl acceptors were isolated with satisfactory purity for further catalytic assays and with yields comprising between 15 % and 84 %, which correspond to yields in a range between 11 mg and 57 mg of isolated peptides (as summarized in [Table 3.7](#)).



**Table 3.6.** List of the synthesized peptide-<sup>α</sup>thioesters constituting the library of acyl donors. Yields were calculated for the reaction of thioesterification starting with 10 mg of hydrazide crude. \* Methionine is oxidized.

Entry	Sequence of the peptide	MW calc. (Da)	MW obs. (Da)	Yield in % (in mg)
1	Ac-VALENA <sup>α</sup> Mes	781.9	781.0	46 % (5.5 mg)
2	Ac-VALENC <sup>α</sup> Mes	813.9	813.4	26 % (3.0 mg)
3	Ac-VALEND <sup>α</sup> Mes	825.9	824.9	16 % (1.8 mg)
4	Ac-VALENE <sup>α</sup> Mes	839.9	839.3	26 % (3.0 mg)
5	Ac-VALENF <sup>α</sup> Mes	858.0	857.2	21 % (2.5 mg)
6	Ac-VALENG <sup>α</sup> Mes	767.9	767.1	31 % (3.5 mg)
7	Ac-VALENH <sup>α</sup> Mes	847.9	847.3	42 % (4.8 mg)
8	Ac-VALENI <sup>α</sup> Mes	824.0	823.2	16 % (1.9 mg)
9	Ac-VALENK <sup>α</sup> Mes	839.0	838.4	49 % (5.7 mg)
10	Ac-VALENL <sup>α</sup> Mes	824.0	823.0	28 % (3.3 mg)
11	Ac-VALENM <sup>α</sup> Mes	842.0	857.4*	17% (2.0 mg)
12	Ac-VALENN <sup>α</sup> Mes	824.9	824.1	47 % (5.1 mg)
13	Ac-VALENP <sup>α</sup> Mes	807.9	807.3	27 % (3.0 mg)
14	Ac-VALENQ <sup>α</sup> Mes	838.9	838.2	13 % (1.6 mg)
15	Ac-VALENR <sup>α</sup> Mes	867.0	866.3	37 % (4.2 mg)
16	Ac-VALENS <sup>α</sup> Mes	797.9	797.3	31 % (3.8 mg)
17	Ac-VALENT <sup>α</sup> Mes	811.9	811.3	18 % (2.1 mg)
18	Ac-VALENV <sup>α</sup> Mes	809.9	809.2	48 % (5.4 mg)
19	Ac-VALENW <sup>α</sup> Mes	897.0	896.4	10 % (1.2 mg)
20	Ac-VALENY <sup>α</sup> Mes	874.0	873.2	47 % (5.4 mg)

Table 3.7. List of the synthesized peptides comprising the library of acyl acceptors.

Entry	Sequence of the peptide	MW calc. (Da)	MW obs. (Da)	Yield in % (in mg)
1	AVALEN- <sup>α</sup> amide	614.7	615.3	69 % (42 mg)
2	CVALEN- <sup>α</sup> amide	646.8	646.3	57 % (37 mg)
3	DVALEN- <sup>α</sup> amide	658.7	658.4	63 % (42 mg)
4	EVALEN- <sup>α</sup> amide	672.7	672.3	59 % (40 mg)
5	FVALEN- <sup>α</sup> amide	690.8	690.4	68 % (47 mg)
6	GVALEN- <sup>α</sup> amide	600.7	600.3	65 % (39 mg)
7	HVALEN- <sup>α</sup> amide	680.8	680.4	84 % (57 mg)
8	IVALEN- <sup>α</sup> amide	656.8	656.4	33 % (22 mg)
9	KVALEN- <sup>α</sup> amide	671.8	671.6	49 % (33 mg)
10	LVALEN- <sup>α</sup> amide	656.8	656.4	79 % (52 mg)
11	MVALEN- <sup>α</sup> amide	674.8	674.4	50 % (34 mg)
12	NVALEN- <sup>α</sup> amide	657.7	657.3	62 % (41 mg)
13	PVALEN- <sup>α</sup> amide	640.7	640.4	56 % (36 mg)
14	QVALEN- <sup>α</sup> amide	671.8	671.4	71 % (48 mg)
15	RVALEN- <sup>α</sup> amide	699.8	699.6	73 % (51 mg)
16	SVALEN- <sup>α</sup> amide	630.7	630.3	60 % (38 mg)
17	TVALEN- <sup>α</sup> amide	644.7	644.3	65 % (42 mg)
18	VVALEN- <sup>α</sup> amide	642.8	642.4	54 % (35 mg)
19	WVALEN- <sup>α</sup> amide	729.8	729.4	15 % (11 mg)
20	YVALEN- <sup>α</sup> amide	706.8	706.5	24 % (17 mg)

### 3.4. Conclusions

To improve the catalytic activity of our *de novo* catalytic proteins and to better understand the unique reactivity, an iterative approach was employed. Starting from crystal structure of the most efficient analogue obtained so far, the **Het-N2** protein, and from its catalytic characterization, a library of new analogues was designed. 21 analogues differing by their Ncap and N1 residues in the cysteine-containing strand were envisaged based on key examples of the literature, by modeling and based on few assumptions in order to: *i*) reach better stabilization of the tetrahedral intermediates through additional hydrogen bonding; *ii*) reach better placement of the cysteine (most probably affected by the

hydrophobic residue at N1 position); *iii*) or modulate reactivity of the cysteine via  $pK_a$  depression. Besides the library of 'negative' strands containing the catalytic cysteine, a library of four analogues of the 'positive' strands was also designed to take advantage of the asymmetry of the structure and of the cavity resulting from the two N-termini facing each other. Residues to reduce the bulkiness of the cavity or to provide additional general base/acid catalysis were introduced on this second strand.

The chemical approaches applied in this work resulted in the efficient synthesis of libraries of analogues. To produce the various 'negative' strands, a convergent strategy based on the ligation between the different N-terminal fragments and the common C-terminal peptide fragment was employed. As a result, 20 new variants of the 'negative' strand containing a cysteine at N2 position were produced with a minimum of synthetic steps. On the other hand, analogues of the 'positive' strand were sequentially assembled by Boc-SPPS starting from the same batch and then performing splitting of the resin.

Libraries of resulting 'negative' and 'positive' strands were tested separately for their ability to catalyze an acyl transfer reaction between a peptide-Phe- $\alpha$ thioester and Tris. Only subtle fluctuations of overall kinetic rate were noticed. For the 'negative' strand including the catalytic cysteine at N2 position and histidine residue at N1 position, more hydrolysis was observed, whereas tryptophan and tyrosine also at N1 induced more aminolysis, which became the major outcome of the reaction. Surprisingly, arginine residues that were expected to enhance the activity through additional hydrogen bonding with tetrahedral intermediates generated much lower overall rates compared to the parent **Het-N2**, which may result from better stabilization of the branched adduct or lesser accessibility to the electrophilic center in the reaction. The branched adduct consumption was indeed much slower with Arg residues in proximity to CysN2. Considering the other 'positive' strand, no significant changes were observed for overall kinetic parameters, which were similar to **Het-N2**. The fact that no substantial acceleration was observed with any of those new analogues suggests that more drastic modifications are needed. Our strategy based on few changes in the surrounding of the catalytic cysteine seems to be limited and further iterative cycles are clearly necessary in order to build more complex catalytic site and network of amino acids while maintaining substrate promiscuity. Nevertheless, in order to orient the next round of design and synthesis, further detailed studies on few synthesized analogues will be indispensable. Detailed catalytic characterization would be essential to better define these analogues in terms of substrate promiscuity and catalytic efficiency. Next step should also include structural characterization to inspect potential structural changes induced by the different amino acids in N-terminal positions. Crystal structure of such analogues in presence of a mimic of tetrahedral intermediates may be also very informative to evaluate the interaction between the protein and the oxyanion. Thus, we plan to protect the catalytic

cysteine by oxidative sulfitolysis leading to a charged adduct on the cysteine in the form of (protein)-S-SO<sub>3</sub><sup>-</sup>.<sup>22</sup> Furthermore, mixing different analogues from 'negative' strand library with others coming from the set of 'positive' strands (especially those inducing more aminolysis) may carry interesting catalytic activity that will be verified in the near future too.

Finally, to better define substrate promiscuity and also substrate dependency on catalytic activity of our designed proteins, libraries of short peptide substrates for both acyl donor (*i.e.* peptide-<sup>α</sup>thioester) and acyl acceptor (*i.e.* peptide bearing a free N-terminal amine) were synthesized. Both libraries will serve later to evaluate catalytic activity of our analogues. We expect that such results will guide us to further improve the design of our catalytic proteins.

### 3.5. References

- (1) Watkins, D. W.; Jenkins, J. M. X.; Grayson, K. J.; Wood, N.; Steventon, J. W.; Le Vay, K. K.; Goodwin, M. I.; Mullen, A. S.; Bailey, H. J.; Crump, M. P.; MacMillan, F.; Mulholland, A. J.; Cameron, G.; Sessions, R. B.; Mann, S.; Anderson, J. L. R. Construction and in Vivo Assembly of a Catalytically Proficient and Hyperthermostable de Novo Enzyme. *Nat. Commun.* **2017**, *8* (1), 358.
- (2) Dawson, W. M.; Rhys, G. G.; Woolfson, D. N. Towards Functional de Novo Designed Proteins. *Curr. Opin. Chem. Biol.* **2019**, *52*, 102–111.
- (3) Broo, K. S.; Brive, L.; Ahlberg, P.; Baltzer, L. Catalysis of Hydrolysis and Transesterification Reactions of P-Nitrophenyl Esters by a Designed Helix–Loop–Helix Dimer. *J. Am. Chem. Soc.* **1997**, *119* (47), 11362–11372.
- (4) Burton, A. J.; Thomson, A. R.; Dawson, W. M.; Brady, R. L.; Woolfson, D. N. Installing Hydrolytic Activity into a Completely de Novo Protein Framework. *Nat. Chem.* **2016**, *8* (9), 837–844.
- (5) James, A. M.; Haywood, J.; Leroux, J.; Ignasiak, K.; Elliott, A. G.; Schmidberger, J. W.; Fisher, M. F.; Nonis, S. G.; Fenske, R.; Bond, C. S.; Mylne, J. S. The Macrocyclizing Protease Butelase 1 Remains Autocatalytic and Reveals the Structural Basis for Ligase Activity. *Plant J.* **2019**, *98* (6), 988–999.
- (6) Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L. Iterative Approach to Computational Enzyme Design. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (10), 3790–3795.
- (7) Studer, S.; Hansen, D. A.; Pianowski, Z. L.; Mittl, P. R. E.; Debon, A.; Guffy, S. L.; Der, B. S.; Kuhlman, B.; Hilvert, D. Evolution of a Highly Active and Enantiospecific Metalloenzyme from Short Peptides. *Science* **2018**, *362* (6420), 1285–1288.
- (8) Signarvic, R. S.; DeGrado, W. F. De Novo Design of a Molecular Switch: Phosphorylation-Dependent Association of Designed Peptides. *J. Mol. Biol.* **2003**, *334* (1), 1–12.
- (9) Barrett, A. J.; Rawlings, N. D. Families and Clans of Cysteine Peptidases. *Perspect. Drug Discovery Des.* **1996**, *6* (1), 1–11.
- (10) Nguyen, G. K. T.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J. P. Butelase 1 Is an Asx-Specific Ligase Enabling Peptide Macrocyclization and Synthesis. *Nat. Chem. Biol.* **2014**, *10* (9), 732–738.
- (11) Richter, F.; Blomberg, R.; Khare, S. D.; Kiss, G.; Kuzin, A. P.; Smith, A. J. T.; Gallaher, J.; Pianowski, Z.; Helgeson, R. C.; Grjasnow, A.; Xiao, R.; Seetharaman, J.; Su, M.; Vorobiev, S.; Lew, S.; Forouhar, F.; Kornhaber, G. J.; Hunt, J. F.; Montelione, G. T.; Tong, L.; Houk, K. N.; Hilvert, D.; Baker, D. Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis. *J. Am. Chem. Soc.* **2012**, *134* (39), 16197–16206.
- (12) Fisher, Z.; Hernandez Prada, J. A.; Tu, C.; Duda, D.; Yoshioka, C.; An, H.; Govindasamy, L.; Silverman, D. N.; McKenna, R. Structural and Kinetic Characterization of Active-Site Histidine as a Proton Shuttle in Catalysis by Human Carbonic Anhydrase II. *Biochemistry* **2005**, *44* (4), 1097–1105.

- (13) Liszczak, G.; Goldberg, J. M.; Foy, H.; Petersson, E. J.; Arnesen, T.; Marmorstein, R. Molecular Basis for N-Terminal Acetylation by the Heterodimeric NatA Complex. *Nat. Struct. Mol. Biol.* **2013**, *20* (9), 1098–1105.
- (14) Diemer, V.; Fischer, L.; Kauffmann, B.; Guichard, G. Anion Recognition by Aliphatic Helical Oligoureas. *Chem. Eur. J.* **2016**, *22* (44), 15684–15692.
- (15) 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.
- (16) Fischer, L.; Semetey, V.; Lozano, J.-M.; Schaffner, A.-P.; Briand, J.-P.; Didierjean, C.; Guichard, G. Succinimidyl Carbamate Derivatives from N-Protected  $\alpha$ -Amino Acids and Dipeptides—Synthesis of Ureidopeptides and Oligourea/Peptide Hybrids. *Eur. J. Org. Chem.* **2007**, *2007* (15), 2511–2525.
- (17) 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>.
- (18) Christensen, T. A Qualitative Test for Monitoring Coupling Completeness in Solid Phase Peptide Synthesis Using Chloranil. *Acta Chem. Scand.* **1979**, *33b*, 763–766.
- (19) Vázquez, J.; Qushair, G.; Albericio, F. Qualitative Colorimetric Tests for Solid Phase Synthesis. In *Methods in Enzymology*; Combinatorial Chemistry, Part B; Academic Press, 2003; Vol. 369, pp 21–35.
- (20) Shechter, Y. Selective Oxidation and Reduction of Methionine Residues in Peptides and Proteins by Oxygen Exchange between Sulfoxide and Sulfide. *J. Biol. Chem.* **1986**, *261* (1), 66–70.
- (21) Sato, K.; Tanaka, S.; Yamamoto, K.; Tashiro, Y.; Narumi, T.; Mase, N. Direct Synthesis of N-Terminal Thiazolidine-Containing Peptide Thioesters from Peptide Hydrazides. *Chem. Commun.* **2018**, *54* (66), 9127–9130.
- (22) Weil-Ktorza, O.; Rege, N.; Lansky, S.; Shalev, D. E.; Shoham, G.; Weiss, M. A.; Metanis, N. Substitution of an Internal Disulfide Bridge with a Diselenide Enhances Both Foldability and Stability of Human Insulin. *Chem. Eur. J.* **2019**, *25* (36), 8513–8521.

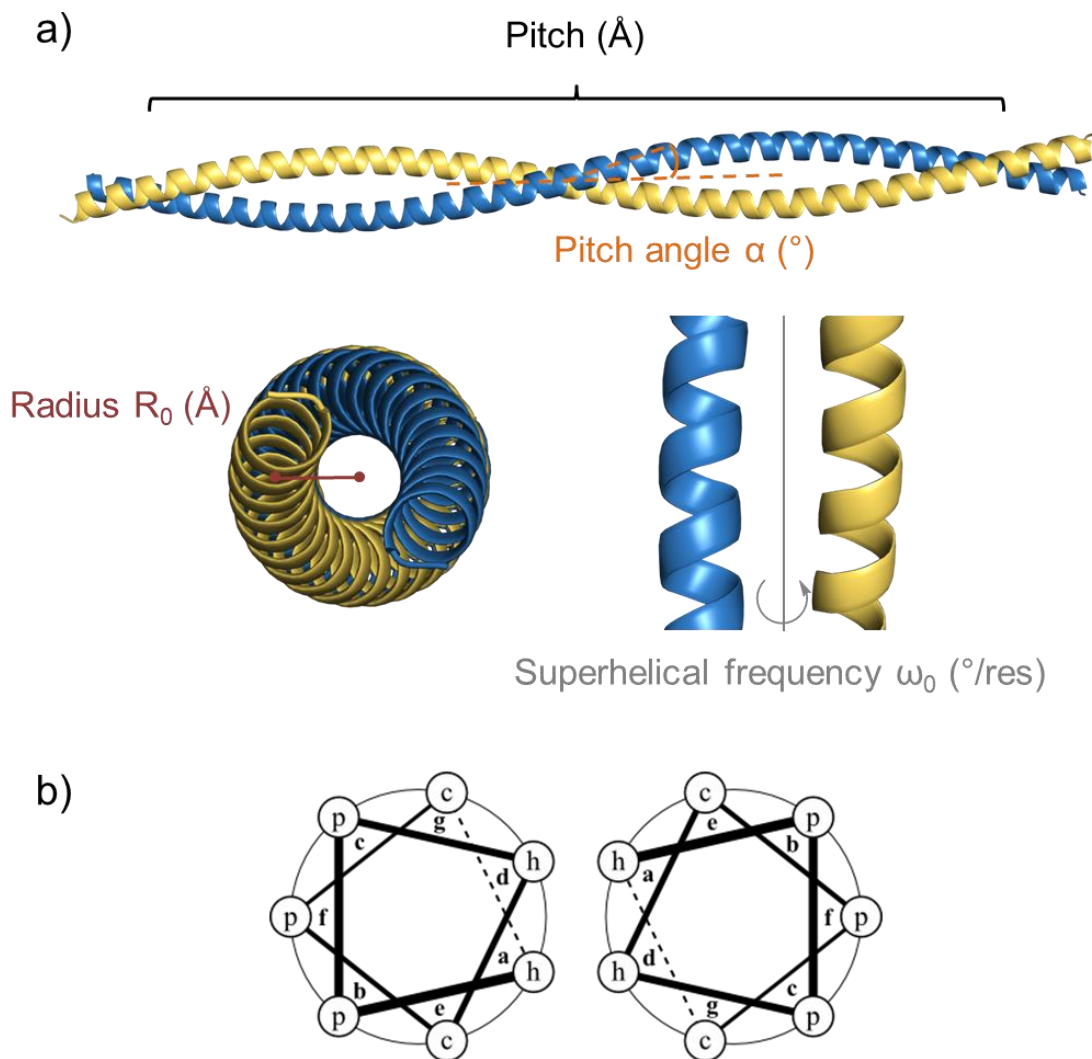
# **Chapter 4. Design, synthesis and characterization of elongated DSD analogues**

## 4.1. Introduction

By studying natural proteins and accumulating a large number of protein structures in the Protein Data Bank (PDB), we have expanded our knowledge for a better understanding of fundamental physical principles of protein folding. It has been then possible to deduce general rules, which furthermore have been incorporated into various algorithms for computational design methods in order to build protein structures that can be ranked using numerical score functions. Computational methodologies are constantly evolving to enable customizable design of protein scaffolds from scratch.<sup>1</sup> Unprecedented protein structures with three-dimensional folds previously not observed in Nature were produced such as the Top 7 protein designed by D. Baker presenting a completely novel globular topology.<sup>2</sup> Assessing such *de novo* protein frameworks is advantageous and is of great interest to diversify the functionalities of protein molecules.<sup>3</sup>

One approach for computational design is based on parametrization, which consists in a mathematical description of a desired and designable protein fold with a minimal number of defined parameters. This concept has been particularly applied for the design of coiled-coil assemblies taking into account the simplicity, the regularity and the designability of such a scaffold. In the fifties, F. Crick described, for the first time, parameters to define coiled-coil structures.<sup>4</sup> Coiled-coil proteins comprise  $\alpha$ -helices containing heptad repeats, abbreviated as *abcdefg* in the sequence, where *a* and *d* positions are generally filled by hydrophobic residues being mainly leucine, isoleucine and valine, and *e* and *g* correspond mostly to charged residues responsible of the selectivity of interactions. The hydrophobic face of the  $\alpha$ -helix drives the association with another  $\alpha$ -helix or several helices by coiling around and forming a knobs-into-holes hydrophobic packing.<sup>5</sup> This assembly forms a superhelical motif that is commonly left-handed and can be modeled by a few parameters with structural deviations between ideal and real structures contained within 1 Å.<sup>6</sup> Crick parametrization of coiled-coils includes among others the radius, the pitch (*i.e.* the height of one complete superhelix turn) of the superhelix and the superhelical frequency (*i.e.* the degree of twist between helices, meaning the angle for each residue of the superhelix around the coiled-coil axis) as represented in **Figure 4.1**. Thus, Crick's equations have been employed to design and produce *de novo* coiled-coil structures with, for example, the possibility to implement backbone motions.<sup>7</sup> Nowadays, dedicated and user-friendly software have been developed to assist in designing of coiled-coil scaffolds (CCCP and CC-builder, for example).<sup>6,8</sup> D. Woolfson and coworkers have successfully designed pentameric, hexameric and heptameric helical bundles via geometrical consideration and computational modeling providing a defined central channel for further functionalization.<sup>9</sup>



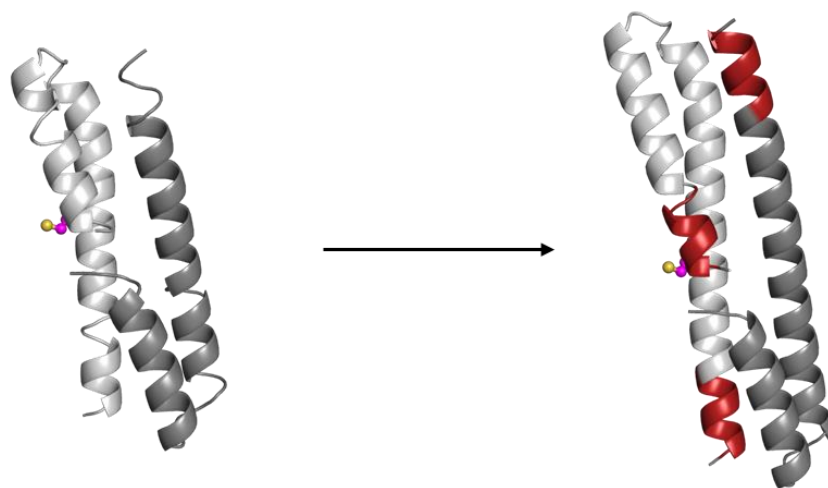


**Figure 4.1. Coiled-coil architecture.** **a)** The principal coiled-coil parameters are the pitch (periodicity of the pitch), the correlated pitch angle (angle between the helix and the coiled-coil axis), the radius of the superhelix and the superhelical frequency (angle for each residue of the superhelix around the coiled-coil axis). **b)** Helical wheel diagram representing the top view of a coiled-coil assembly. 'h' is for hydrophobic residues generally found in *a* and *d* positions of the heptad repeat, 'c' is for charged residues at both *e* and *g* position and 'p' is for polar residues.

Starting from one of our synthesized and characterized catalytic protein, **Hom-N2**, we utilized parametrization of our coiled coil and computational tools to design an elongated analogue (see **Figure 4.2**). By taking the crystal structure of the three-helical bundle, we broke the symmetry to form a heterodimer by elongating to one heptad repeat each helix, which implies that for the two helices facing to each other (*i.e.* both N-termini), only the helix containing the catalytic cysteine would be extended. In this way, the surface of the hydrophobic core and the number of hydrophobic interactions would increase leading to a plausible enhancement of stability of the heterodimeric assembly compared to the respective dissociated monomers. Indeed, a better stability especially against chemical denaturation is highly coveted to allow efficient catalysis in buffers with a high concentration of denaturant (*e.g.* 6 M Gn·HCl or 8 M urea) in contrary to **Het-N2** that demonstrated a loss of activity at

these denaturant concentrations (see subsection 2.7 and **Figures 2.29** and **2.30**). Native chemical ligation generally requires such conditions that result in unfolding of peptides for a better accessibility of the reactive groups at the N- and C-termini of peptides. Thus, having a stable protein conducting efficient catalysis in presence of a denaturant would be very attractive for further applications. Moreover, in this new design, we planned to elongate the helix containing the catalytic cysteine with seven additional residues that may result in an enhancement of the macrodipole effect. Helix macrodipole arises from the accumulation of dipole moments of each peptide bond, which are all approximately oriented and aligned along the helical axis from C-terminus to N-terminus. Possible effects of helix macrodipole have been proposed to play a role in anion binding at the N-terminus of helix (e.g. Cl<sup>-</sup> chloride channel, in which anion selectivity and stabilization are achieved via partial positive charges of N-termini and via chemical coordination with unpaired NH and some hydroxyl groups)<sup>10</sup> and rate enhancement of a reaction through  $pK_a$  depression for deprotonation of catalytic residues.<sup>11</sup> Interestingly, the subtilisin and the papain, two common proteases with unrelated three-dimensional structures, carry the nucleophilic group (Ser and Cys, respectively) at the N-terminus of a long helix. It has been hypothesized that helix macrodipole may affect catalysis by facilitating the deprotonation of the nucleophilic residue and by stabilizing the oxyanion of tetrahedral intermediates via favorable interactions with the electrical field of the helix (in addition to favorable H-bonding with unpaired NH).<sup>11</sup> Our catalytic protein is also expected to catalyze acyl transfer reactions but for the opposite reaction, *i.e.* peptide ligation or more generally amide bond formation. Therefore, the elongation of the helix may provide additional effect for the enhancement of catalytic efficiency.

In this chapter, I will present the design of elongated analogues with the subsequent chemical synthesis and structural characterization. Catalytic characterization and stability assessment will be also discussed.



**Figure 4.2. Principle of the elongated analogue design.** Starting from **Hom-N2** crystal structure, we kept only one cysteine at one N2 position to lead to heterodimer. Helices were elongated with one heptad repeat each, the additional heptads are represented in red. The cysteine at N2 position is depicted in magenta and the sulfur atom in yellow.

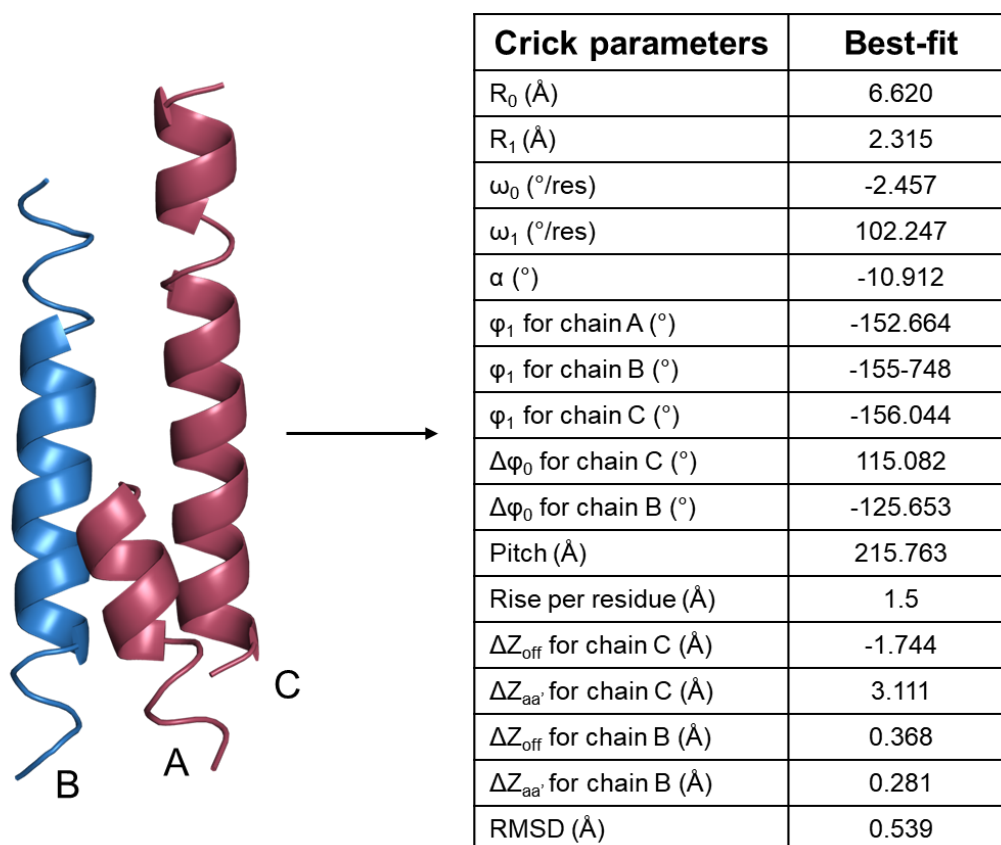
## 4.2. Design of elongated analogues

*Design described in this subsection was performed in the laboratory of W. F. DeGrado (UCSF, USA) with the help of Nicholas Polizzi, Nathan Schmidt and William F. DeGrado.*

To design elongated analogues, we first parametrized our three-helical bundle obtained from crystal structure. Thanks to the resulting Crick's parameters of our coiled-coil, we generated a longer bundle that was aligned to the X-ray structure. Both structures were combined to obtain an extended structure and the new residues were modified with respect to the heptad repeats. After relaxation of the protein model, sequence was optimized thanks to computational tools to identify sequences that would better stabilize the scaffold. All these aforementioned steps will be explained in more details in this subsection.

As starting scaffold, we worked with the crystal structure of **Hom-N2** because at that time it was the only structure we solved that contained our modifications (*i.e.* Lys of the original DSD to Arg residues and Cys incorporation). However, the asymmetric unit of this X-ray structure is composed of three copies of the homodimer and two of them are covalently linked by a single disulfide bridge (as shown in [Figure 2.17e-f](#) of subsection 2.5). Thus, we decided to work with the intact homodimer to abolish any influence on the structure of the disulfide bond. A truncated model of the bundle was generated to be then processed by the CCCP (Coiled-Coil Crick Parametrization) online program ([Figure 4.3](#)).<sup>6</sup> Our homodimer fitted well with Crick's parameters with a RMSD of 0.539 Å compared to an ideal coiled coil. The corresponding Crick's parameters for the catalytic protein **Hom-N2** are recapitulated in [Figure 4.3](#). Using the structure generator from CCCP software, a coiled-coil backbone was modeled following the Crick's parametrization of **Hom-N2** protein. A screenshot of the parameters used to model the backbone as well as the resulting coiled-coil protein (trimer constructed with a polyglycine sequence and with a longer chain length of 42 residues) are shown in [Figure 4.4](#).

The next step consisted in combining the crystal structure of **Hom-N2** and the modeled helical bundle composed of polyglycine to obtain an extended analogue while keeping the global shape of the framework (*e.g.* the loop). To do so, one monomer of the **Hom-N2** structure was first moved down to seven residues to enable insertion of the new heptad ([Figure 4.5a](#)). The resulting structure was aligned with the coiled-coil model, which was adjusted to contain only one additional heptad ([Figure 4.5a](#)). By defining the termini where each chain (*i.e.* chains A and B of **Hom-N2** and chains A, B and C for the modeled trimer) has to start and has to stop, a homemade python script (written by Nathan Schmidt) was employed to merge these ends and combine the PDBs to one PDB file ([Figure 4.5b](#)).



**Figure 4.3. Input structure for Crick's parametrization.** To be processed the **Hom-N2** structure was truncated. One strand is represented in red and the other in blue (chain B). Chain A corresponds to the N-terminal fragment of the strand and the corresponding C-terminal segment was named chain C. Resulting Crick parameters are summarized in the table. A definition of the different parameters can be found in Figure 4.1 and in ref. 6.

This procedure was performed separately for both strands of **Hom-N2** and the resulting PDB files for each chain were combined resulting in the desired extended scaffold (**Figure 4.5b**). Therefore, the strand containing the cysteine was elongated by 14 residues and was renamed 'strand +14, Ncap-N1C' and the other strand was elongated by 7 residues resulting in the 'strand +7'. By extending one strand more than the other, we anticipated a more favorable heterodimerization over homodimerization. A quick MASTER search was then performed on one portion of the designed protein bundle (delimited by the dashed black box in **Figure 4.5b**) to check the designability of the new construct, meaning if the scaffold can accept various sequences.<sup>12</sup> The MASTER program identified structural fragments from a PDB database (built by Marco Mravic from DeGrado laboratory) similar to the selected segment of our designed protein within a RMSD cutoff of 1.5 Å. Matches were detected with or without considering the loop (**Figure 4.5b**).

Polyglycine sequences of the resulting elongated analogue were then modified by repeating the sequence of one heptad and by placing the Cys at the N2 position of 'strand +14' (**Figure 4.5c**). To finish the elongation design process, relaxation of the structure with

Number of chains:

Chain length:

---

**Coiled-coil parameters**

Parameter (symbol, unit)	Star	End	# Samples	
Rise per residue ( $d$ , Å):	<input type="text" value="1.5"/>	<input type="text" value="1.5"/>	<input type="text" value="1"/>	<input checked="" type="checkbox"/> constrain*
Superhelical radius ( $R_0$ , Å):	<input type="text" value="6.620"/>	<input type="text" value="6.620"/>	<input type="text" value="1"/>	<input type="radio"/> adjust**
Superhelical frequency ( $\omega_0$ , °/aa): (negative means left-handed superhelix)	<input type="text" value="-2.457"/>	<input type="text" value="-2.457"/>	<input type="text" value="1"/>	<input type="radio"/> adjust**
Pitch angle ( $\alpha$ , °):	<input type="text" value="-10.912"/>	<input type="text" value="-10.912"/>	<input type="text" value="1"/>	<input type="radio"/> adjust**
$\alpha$ -helical radius ( $R_1$ , Å):	<input type="text" value="2.315"/>	<input type="text" value="2.315"/>	<input type="text" value="1"/>	
$\alpha$ -helical frequency ( $\omega_1$ , °/aa):	<input type="text" value="102.247"/>	<input type="text" value="102.247"/>	<input type="text" value="1"/>	

Symmetry (limits variable parameters):

$C_n$

$D_n$

Do not impose symmetry

Forcefield minimize final backbone (CA atoms held fixed)

Create a poly-alanine backbone, not a poly-glycine

---

**Chain-wise parameters**

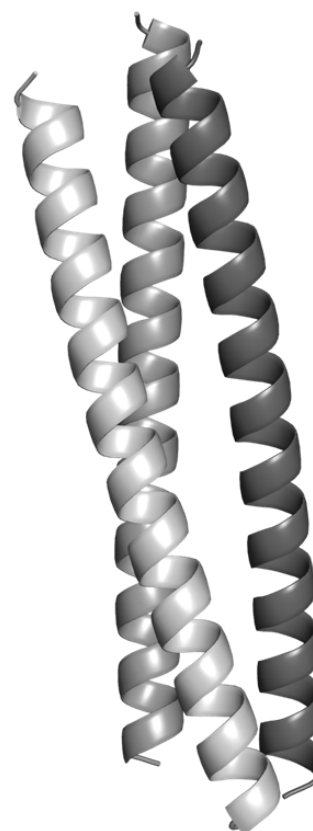
Parameter (symbol, unit)	Star	End	Steps
$\alpha$ -helical phase ( $\varphi_1$ , °):	<input type="text" value="-152.664; -152.664"/>	<input type="text" value="-152.664; -152.664"/>	<input type="text" value="1"/>
Superhelical phase offset ( $\Delta\varphi_0$ , °):	<input type="text" value="115.082; -125.65"/>	<input type="text" value="115.082; -125.65"/>	<input type="text" value="1"/>
Z offset ( $Z_{aa}$ , Å):	<input type="text" value="3.111; 0.281"/>	<input type="text" value="3.111; 0.281"/>	<input type="text" value="1"/>

Orientation versus first chain (**p** – parallel, **ap** – antiparallel):

Couple chain variations

---

Resulting number of structures (not to exceed 1,000):

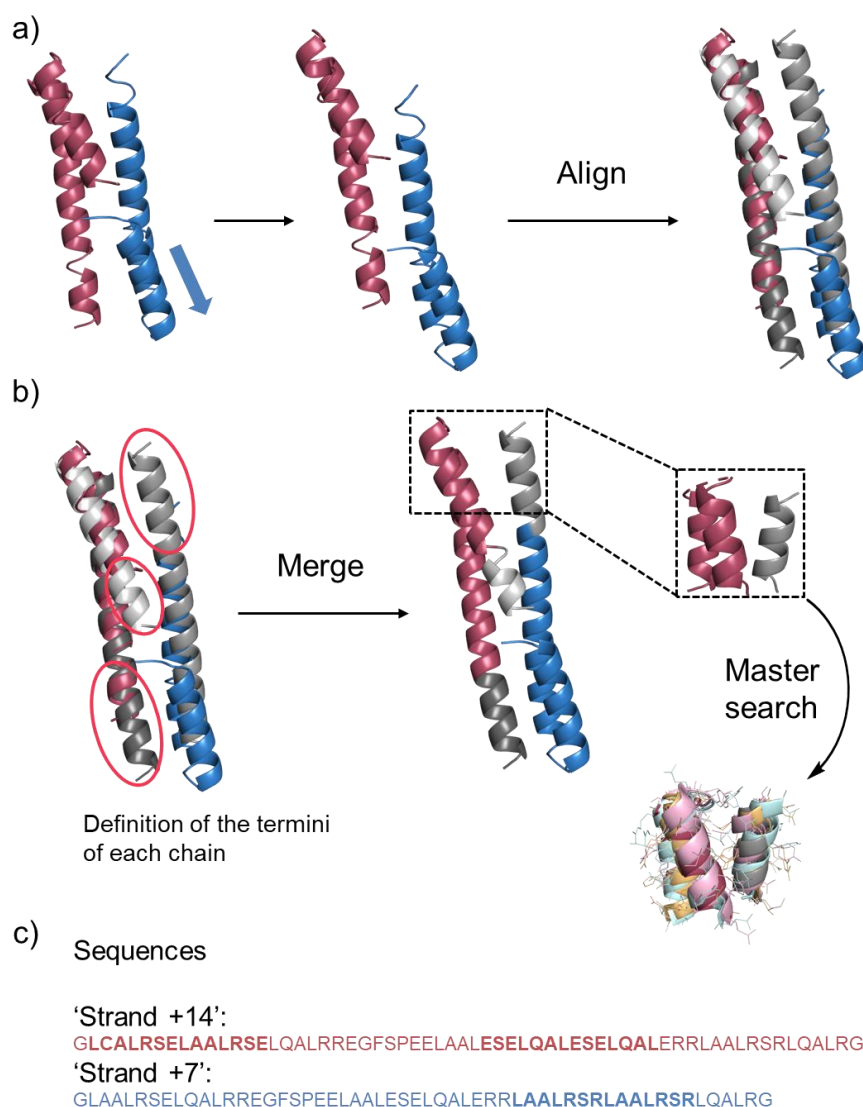


**Figure 4.4.** Employed input parameters (left side) used to design a coiled-coil model containing a polyglycine sequence (right side) based on Crick's parameters of Hom-N2.

Rosetta software was performed leading to a RMSD of 0.847 Å, which means that no significant clashes were generated during the design.

Finally, the sequence of the analogue was optimized especially for a potential better packing of the hydrophobic core (only composed of leucine residues for the moment), but also at  $f$  positions where the additional heptads were introduced (sequence in bold in **Figure 4.5c**) to possibly provide better stabilized rotamers or extra H-bonding through the helix. Thus, the fixbb (fix backbone) script of Rosetta software was used to allow modifications of specified positions of amino acids that were previously defined (in a resfile) meanwhile the backbone stayed fixed. Leucine residues of the hydrophobic core were allowed to be changed to other hydrophobic residues Tyr, Leu, Val, Ala, Trp, Phe and Ile and  $f$  positions of the heptad were allowed to be modified by any polar residues except cysteine and lysine. To favor fine-tuned packing, common rotamers were evaluated with a certain flexibility of the side chain torsion. Hundred structures were generated to probe a maximum of solutions that were then aligned to acquire the consensus sequence. Among the 40 positions allowed to be modified, 17 recurrent changes were noticed along the sequence of the extended analogue. The phenylalanine of the loop was changed to either valine or alanine. Nine modifications in the hydrophobic core have been calculated as favorable. Here, leucine residues were

substituted by either alanine or isoleucine. Surprisingly, the leucine at N1 position next to the cysteine was changed to alanine according to the modeling. However, it has been demonstrated that Ala amino acid with a side chain less hydrophobic than Leu induces a decrease of the catalytic activity (see subsection 3.2.4). Besides hydrophobic residues, 6 polar residues were modified by amino acids containing amide moiety in the side chain (*i.e.* Asn and Gln), hence providing additional hydrogen bonds between  $i$ ,  $i+4$  residues (see **Figure 4.6**). In parallel, the type of charge patterning was also modified by alternating positive and negative charges in a strand. As five charged residues are present along the helices, the 'strand +14, GAC' carried three negative charges and two positive charges, whereas the 'strand +7' carried two negative charges and three positive charges (see **Figure 4.7**).



**Figure 4.5. Design process to elongate the helices of the new analogue.** a) One strand of the crystal structure of **Hom-N2** was moved down to one heptad. The resulting scaffold was aligned with the coiled-coil model. b) The terminations of each chain were defined and merged to form a unique scaffold. A fragment of the resulting scaffold was analyzed to find similar other protein segments existing in a PDB database thanks to a MASTER search. c) Resulting sequence: the polyglycine sequence was modified according to surrounding heptad repeats (in bold).

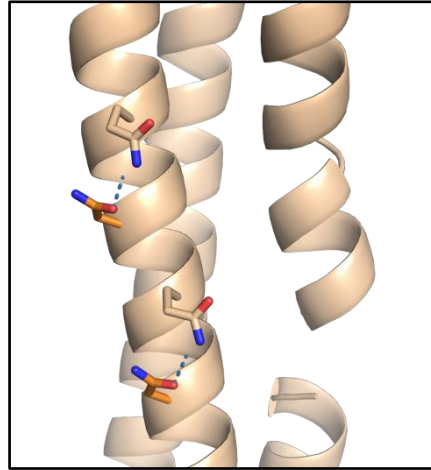


Figure 4.6. Additional hydrogen bonding between  $i, i+4$  residues thanks to modifications of polar residues to asparagine (in orange).

The design has finally resulted in the sequences listed below for the 'strand +14, GAC' and the 'strand 7' that form the analogue named **Het-elong, GAC**. As the alanine residue at N1 position may engender lower catalytic activity, we also decided to synthesize the 'strand +14, GLC' with the parent leucine at N1 position next to the catalytic cysteine to not disrupt the original catalytic environment resulting in the analogue **Het-elong, GLC**. Here, modifications concerning the hydrophobic core are depicted in blue, concerning the loop in purple and concerning polar residues in orange.

'Strand +14, GAC'

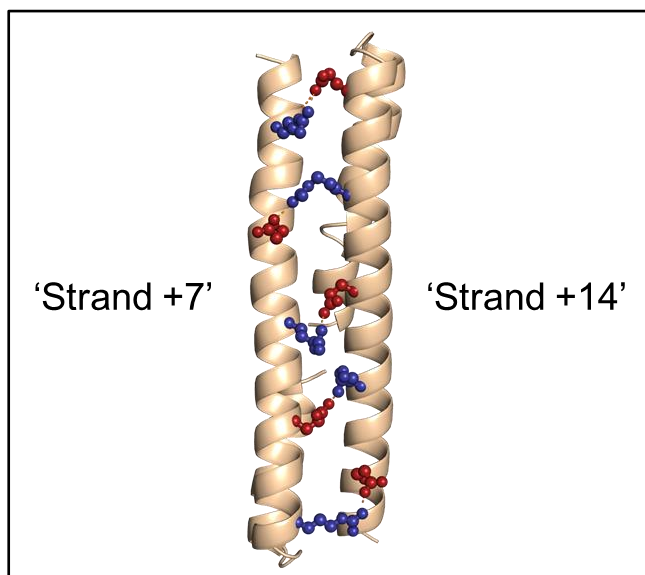
Ac-G**A**CALRSELAALR**Q**ELQA**A**RREG**V**SPEELAAL**E**NRLQA**E**NE**A**QA**I**ENRLAALRSELQA**A**RG

'Strand +14, GLC'

Ac-GL**C**ALRSELAALR**Q**ELQA**A**RREG**V**SPEELAAL**E**NRLQA**E**NE**A**QA**I**ENRLAALRSELQA**A**RG

'Strand +7'

Ac- G**A**AALRSE**I**QALRREG**A**SPERLAA**E**SELQA**I**EN**R**AAL**R**NE**L**AALRSRLQALRG



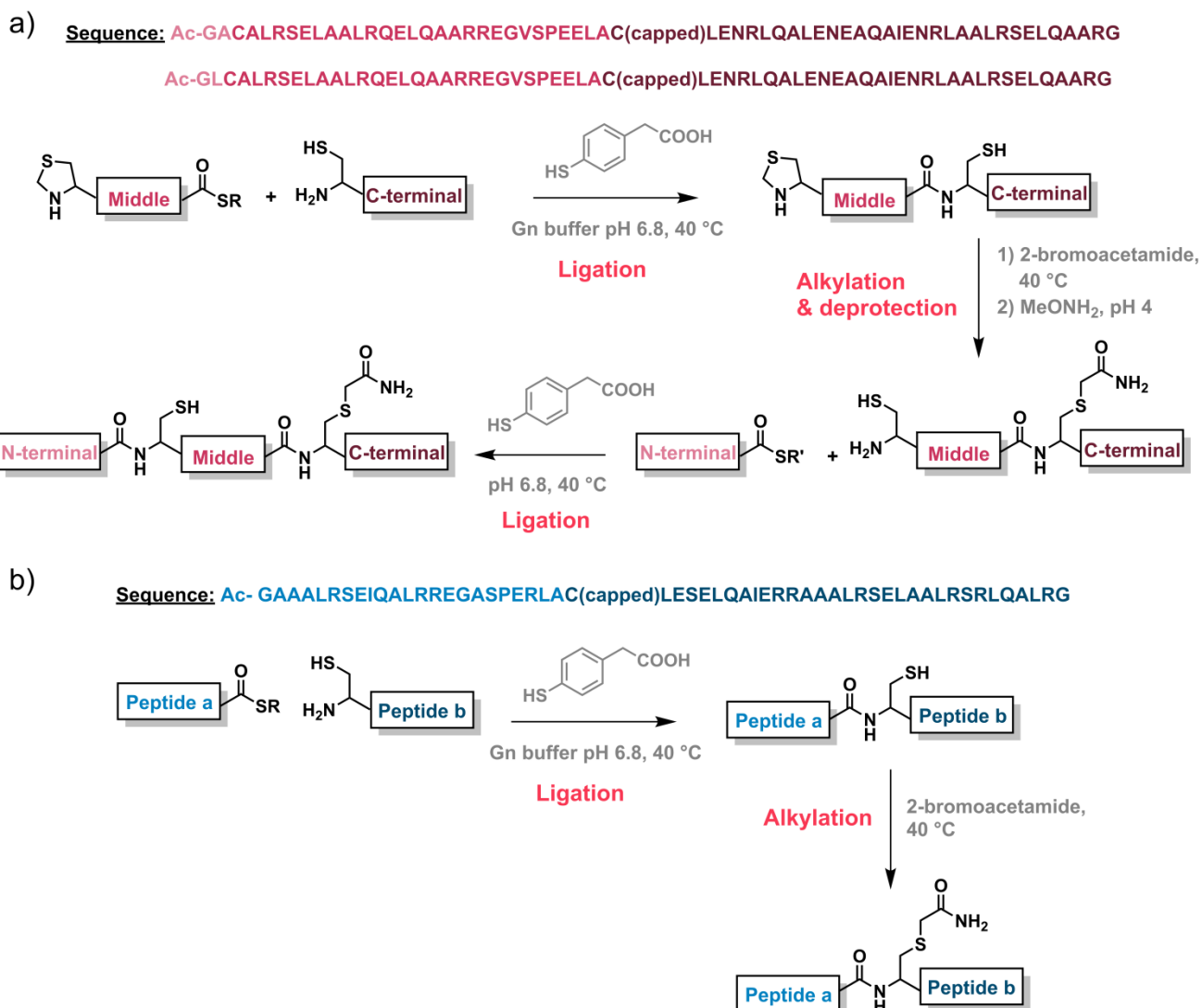
**Figure 4.7. Charge patterning of the elongated analogues.** The 'strand +14, GAC' carries three negative charges (Glu in red) and two positive charges (Arg, in blue), whereas the 'strand +7' carried two negative charges and three positive charges

### 4.3. Total chemical synthesis of the elongated analogues

Due to the rather low yields obtained during the 48-residue sequential assembly by Boc-SPPS of the strands composing **Het-N2** and **Het-Ncap**, we opted for a convergent approach through native chemical ligation to synthesize the 'strand +14, GAC', the 'strand +14, GLC' (62 residues each) and the 'strand +7' (composed of 55 residues). Both 'strands, +14' containing the catalytic cysteine at N2 position were divided into three fragments that were assembled via a one-pot three-segment ligation procedure as previously described for the synthesis of **Hom-N2** in subsection 2.3.3. The 'strand +7' was split into two peptide fragments, each synthesized by Boc-SPPS, which were joined by NCL followed by an alkylation of the remaining cysteine. Both strategies of ligation to assemble the different strands are recapitulated in **Scheme 4.1**.

For the synthesis of both 'strands +14' (as shown in **Scheme 4.1a**), the peptide was split at the junction between the second residue and the catalytic cysteine to take advantage of the latter for the second ligation and at an Ala-Ala junction right after the loop (between the 30<sup>th</sup> and the 31<sup>st</sup> amino acids in the sequence), where the second alanine would be therefore modified by a pseudo-homoglutamine (capped cysteine). Thus, only the N-terminal segment was varied as it was composed of the Ncap and N1 residue (at N1 position, Ala residue for the 'strand +14, GAC' and Leu residue for the 'strand +14, GLC'). Both N-terminal peptide fragments were synthesized by '*in situ* neutralization' Boc/benzyl-SPPS to directly obtain a thioester moiety by coupling MPA (3-mercaptopropionic acid). The fragment containing the leucine at N1 position was previously synthesized for the **Hom-N2** protein with an isolated





**Scheme 4.1. Reaction pathways for the synthesis of 'strand +14' and 'strand +7'.** a) Synthesis of both 'strand +14, GAC' and 'strand +14, GLC' using a one-pot three-segment ligation strategy. b) Synthesis of 'strand +7' common to both analogues by NCL. The shown sequence corresponds to the sequence in which Asn residues were replaced by either arginine or serine due to synthetic issues.

yield of 35 % and was used again for the synthesis of the 'strand +14, GLC'. The other fragment carrying an alanine residue at the second position was produced following the same protocol with a yield of 78 % after HPLC purification. The middle fragment was also assembled by Boc-SPPS as a peptide- $\alpha$ thioester with the cysteine protected in the form of a thiazolidine (yield after purification of 22 %, 154 mg of peptide obtained from a 0.2 mmol scale synthesis). The last peptide fragment corresponding to the C-terminus part of the strand was synthesized by Fmoc-SPPS using an automated microwave synthesizer. However, a significant accumulation of by-products generally observed with the Fmoc technique lowered down the yield to 4 % (corresponding to 12.6 mg of isolated peptide obtained from a 0.1 mmol scale synthesis). One-pot three-segment ligation in a C-to-N direction was then performed with the three pure peptide fragments (**Scheme 4.1a**). The

middle and the C-terminal peptides were first joined together by native chemical ligation using 4-mercaptophenylacetic acid (MPAA) as catalyst. The remaining cysteine was alkylated with 2-bromoacetamide and excess of alkylating agent was quenched upon addition of an excess of MPAA. Ring-opening of the thiazolidine was performed in presence of methoxyamine at pH 4.0 leading to the putative catalytic cysteine. The latter also served for the second ligation with either the N-terminal fragment containing the alanine or the one containing the leucine. The ligated 'strand +14, GAC' and 'strand +14, GLC' were isolated by preparative HPLC with 28 % and 30 % of yield, respectively.

On the other hand, the 'strand +7' common to both elongated analogue was split into two peptide fragments also at the Ala-Ala junction right after the loop (here, between the 23<sup>rd</sup> and 24<sup>th</sup> residues in the sequence), where the second alanine would be therefore modified by a pseudo-homoglutamine (capped cysteine). Both peptide segments were synthesized by Boc-SPPS to maximize yields. The peptide **a** was straightforwardly assembled as a peptide-<sup>α</sup>thioester thanks to the MPA moiety directly coupled on resin with a final yield after purification of 26 % (corresponding to 296 mg of isolated peptide obtained from a 0.4 mmol scale synthesis). However, full-length assembly of the peptide **b** failed while obtaining after cleavage a single peak on LC-MS. This main peak corresponded to the fragment of the fourteen first amino acids coupled that is acetylated at the N-terminus ( $MW_{\text{calc}} = 1594.9$  Da,  $MW_{\text{obs}} = 1595.1$  Da), suggesting that an acetyl group was coupled instead of the expected Asn residue, which stopped the synthesis. Actually, <sup>1</sup>H NMR of the employed Boc-L-Asn(Xan)-OH building block revealed the presence of acetic acid that most probably upon activation reacted with the growing peptide on resin instead of the Asn amino acid. To save time meanwhile optimizing the amount of peptides obtained in one solid-phase synthesis, we preferred to stick to the Boc-SPPS technique (due to too low yield by Fmoc-SPPS) and not necessarily wait for a new batch of Boc-L-Asn(Xan)-OH. Indeed, we judged that these Asn modifications would not impact much the stability of the protein folding and the catalytic activity compared to amino acids located in the hydrophobic core and in the close environment of the catalytic cysteine, respectively. Thus, we decided to work with the original amino acids present before the computational modeling (*i.e.* serine and arginine substituting the two asparagine residues present in the modeled sequence) leading to the sequence shown in dark teal in **Scheme 4.1b**. The new sequence of peptide **b** was finally successfully synthesized by Boc-SPPS on a 0.2 mmol scale with acceptable yield of 8 % after purification corresponding to 55.8 mg of isolated peptide. The NCL of the peptide fragments **a** and **b** followed by alkylation of the remaining cysteine proceeded smoothly resulting in the ligated 'strand +7' with 60 % of yield after purification (**Scheme 4.1b**).

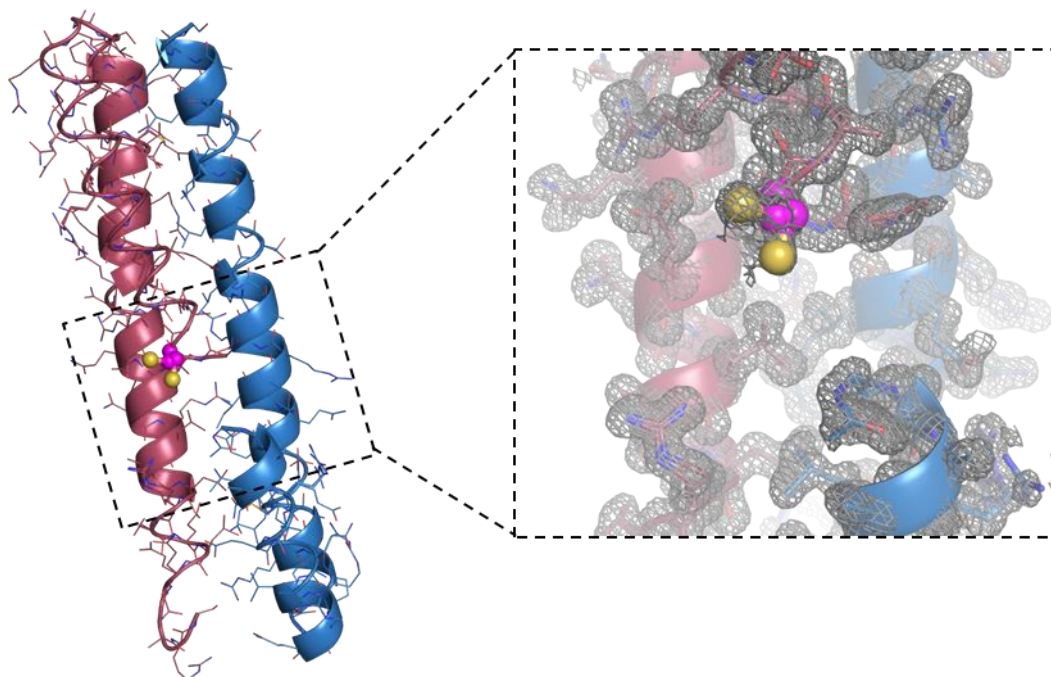
Formation of heterodimers **Het-elong, GAC** and **Het-elong, GLC** occurred spontaneously by mixing the corresponding 'strand +14' and 'strand +7' in a 1:1 ratio in buffer at neutral pH.

#### 4.4. Crystal structure

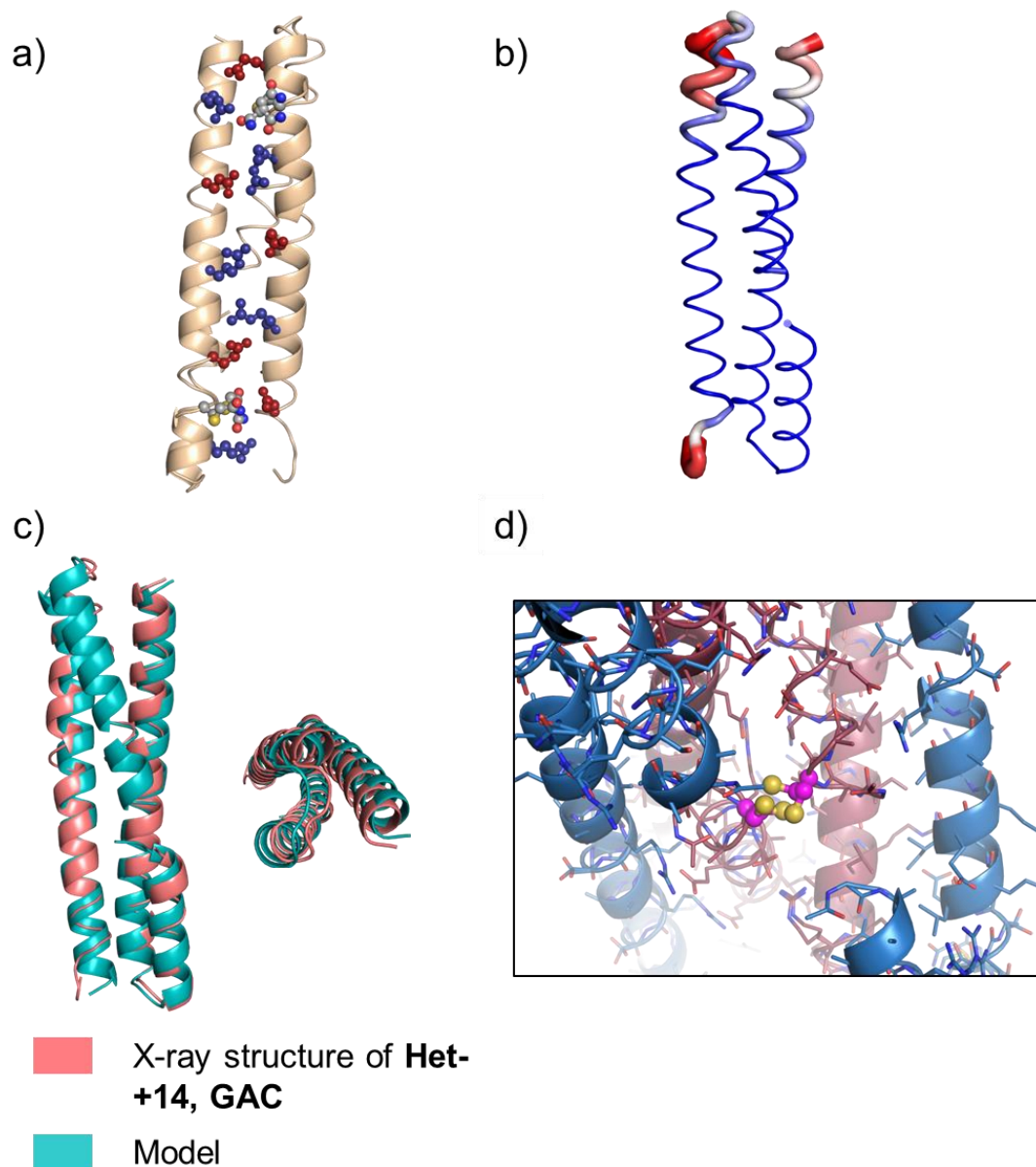
In collaboration with the Structural Biology and Genomics Technology Platform of Institut de génétique et de biologie moléculaire et cellulaire (IGBMC), we performed X-ray crystallographic experiments to verify the consistency of our design and to determine if the protein folds as designed and expected. By mixing samples of 'strand +14, GAC' and 'strand +7' in equimolar ratio (final protein concentration at 21.2 mg/mL) in 100 mM MES buffer at pH 6.5 containing 100 mM NaCl and 10 mM TCEP, crystals were obtained for the construct **Het-elong, GAC** by high-throughput screening followed by an optimization around the condition 20 % PEG 3350, 0.2 M ammonium sulfate and 0.1 M Bis Tris pH 5.5. X-ray data were collected at 1.15 Å of resolution and structure was solved by molecular replacement using the known structure of the original DSD protein (PDB ID: 1G6U).<sup>13</sup> The asymmetric unit is composed of one copy of the intended heterodimer that forms a three- $\alpha$ -helical bundle with the two strands of different lengths (**Figure 4.8**). Thanks to the high resolution of the structures, alternative conformations were noticed for several amino acids, such as few arginines, both capped cysteines and the catalytic cysteine as shown in the zoomed section in **Figure 4.8**. Charge patterning was globally respected by maintaining salt bridges between the two strands (**Figure 4.9a**). Interestingly, the capped cysteine that can be described as a pseudo homoglutamine (in grey in **Figure 4.9a**) may also participate in stabilizing intermolecular interactions via H-bonding with either arginine or glutamate residues (distances of 3.1-3.5 Å between residues). Moreover, the B or the Debye Waller factor indicates that the overall structure of **Het-elong, GAC** is rather stable with higher thermal motions in the loop and the C-termini regions (as shown in the cartoon putty depiction in **Figure 4.9b**). It is noteworthy that the C-terminus of the 'strand +14, GAC' with an enhanced B-factor is not adopting a helical fold in the crystal structure. The X-ray structure of our analogue **Het-elong, GAC** is globally comparable to our computationally designed model (all-atom RMSD of 2.5 Å and RMSD on C $\alpha$  of 1.77 Å). The major changes are essentially coming from the deviating C-terminus and a global shift of the backbone (see **Figure 4.9c**). The backbone shift may not be so relevant as it may be explained by the formation of a disulfide bridge between the two copies of the protein in different asymmetric units of the crystal, thus constraining the scaffold. Indeed, by generating the symmetric structures, one

copy of the analogue is found within the distance to form a disulfide bridge between the two proteins as illustrated in **Figure 4.9d**.

Here again, crystal structure provided important information about the potential three-dimensional topology of the new designed analogue. Obtained X-ray structure is in compliance with our model generated by different computational techniques. However, oxidation of the cysteine may have slightly perturbed the resulting structure. Protection of the cysteine by oxidative sulfitolysis forming a (protein)–S–SO<sub>3</sub><sup>−</sup> adduct seems again to be a good alternative meanwhile having the possibly to examine the interaction of an oxyanion with the catalytic environment.



**Figure 4.8.** Crystal structure of the elongated analogue **Het-elong, GAC**. The ‘strand +14, GAC’ is depicted in red and the ‘strand +7’ in blue. In the zoomed section (on the right) catalytic sites are shown revealing solvent-exposed cysteine residues depicted in magenta with sulfur atom in yellow in ball-and-sticks representation. The simulated annealing  $2F_o-F_c$  electron density map is shown at  $1.0\sigma$  level.



**Figure 4.9.** Analysis of the crystal structure of our synthesized **Het-elong, GAC**. **a)** Salt bridges are maintained as expected with the designed charge patterning. Arginine residues are depicted in blue and the glutamate in red. The capped cysteine shown in grey may also participate in intermolecular interactions. **b)** Cartoon putty B-factors representation in Pymol highlight regions with higher B factors (in red) and with lower B factors depicted in blue. **c)** The **Het-elong, GAC** structure (depicted in salmon pink) superimposes with the computationally designed model (in cyan), side view (on left) and top view (on right) of the alignment are represented. **d)** Two symmetric proteins of the crystal are depicted revealing nearby cysteine residues that may be oxidized.

## 4.5. Catalytic characterization

Catalytic activity of the two designed elongated analogues **Het-elong**, **GAC** and **Het-elong**, **GLC** was then evaluated. To do so, we followed acyl transfer reaction on a peptide- $\alpha$ thioester in presence or absence of an acyl acceptor (*i.e.* monitoring only hydrolysis or concomitant aminolysis and hydrolysis) by using catalytic assays of the 1<sup>st</sup> generation (see subsection 2.6.3). It is noteworthy that this catalytic assay was the only available at that time (the 2<sup>nd</sup> generation was developed later). The following conditions were employed: 0.5 mM of peptide- $\alpha$ Mes, with or without 250 mM Tris in presence of one of the elongated analogue in phosphate buffer (50 mM sodium phosphate, 0.1 M NaCl, pH 7.0). Aliquots at different time points were taken off to monitor the reaction by HPLC and LC/MS. Catalytic reactions were compared with the background reaction (*i.e.* without any catalyst) and with our most efficient variant **Het-N2**. Importantly, due to stock issues, thioester substrates were slightly different when we analyzed our designed analogues, *i.e.* a 30-residue thioester substrate named [N-ACTR]- $\alpha$ Mes was employed with **Het-elong**, **GAC**, whereas a shorter fragment of the [N-ACTR]- $\alpha$ Mes peptide called peptide1-Arg- $\alpha$ Mes was used for catalytic evaluation of **Het-elong**, **GLC**. Using **Het-N2**, it was demonstrated that length of the thioester substrate does not affect the kinetics and it is indeed the C-terminal amino acid of the acyl donor that plays a decisive role in influencing the catalytic reaction and its outcome. Here, both thioester substrates bear an Arg residue at the C-terminal position.

For each analogue, two conditions were assayed to test the catalytic activity of either the hydrolysis of thioester substrate alone or of the concomitant aminolysis and hydrolysis by adding some Tris in the reaction mixture. For both analogues, branched adduct resulting from transthioesterification between the peptide- $\alpha$ Mes and the catalytic cysteine was detected by LC/MS supporting a ping-pong mechanism as expected in the hypothesized catalytic mechanism (outlined in **Scheme 2.1**). However, with the assay of the 1<sup>st</sup> generation, evolution of such a branched adduct was estimated by measuring area of the corresponding peak detected by analytical HPLC. For **Het-elong**, **GAC**, peaks between the free analogue and the branched adduct were well separated and percentage of the branched adduct was deduced (**Figure 4.10**), whereas free **Het-elong**, **GLC** and the corresponding peptide-catalytic protein branched thioester adduct were co-eluting. Thus, only branched adduct evolution for **Het-elong**, **GAC** was compared to previously described analogues **Het-N2**, **Het-Ncap** and **Hom-N2**. Interestingly, the new elongated analogue showed a slower and less pronounced formation of the branched adduct as shown in **Figure 4.10**, suggesting a lower catalytic activity.

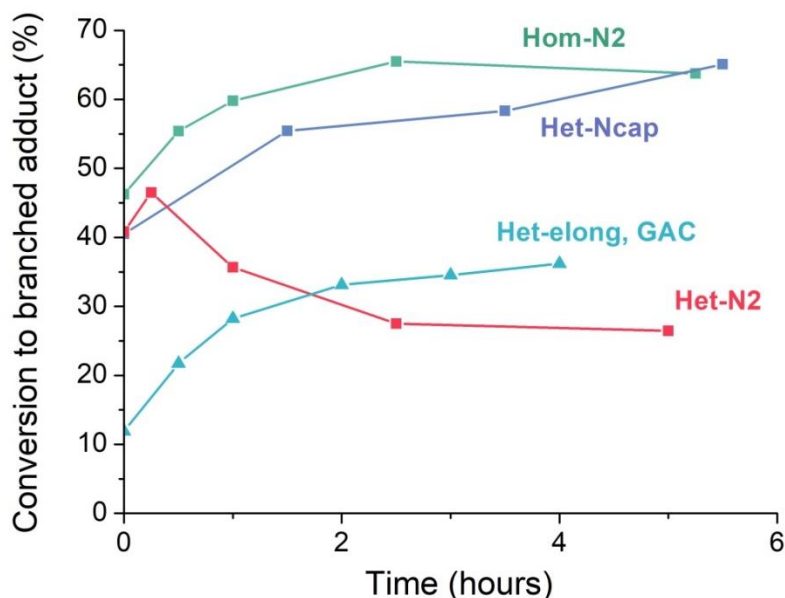


Figure 4.10. Evolution of the branched adducts resulting from transthioesterification between the peptide- $\alpha$ Mes and the catalytic cysteine. Het-elong, GAC analogue was compared with previously characterized Hom-N2, Het-Ncap and Het-N2 proteins.

When the two different peptide- $\alpha$ thioesters were introduced separately and in presence of the corresponding elongated analogue, catalysis of substrate hydrolysis was clearly noticed. Half-life of the thioester substrate [N-ACTR]- $\alpha$ Mes in presence of **Het-elong, GAC** was estimated at around 20 hours of reaction, whereas half-life of peptide1-Arg- $\alpha$ Mes in presence of **Het-elong, GLC** was around 7 hours (Table 4.1). Not surprisingly, the amino acid at N1 position, just next to the catalytic cysteine, affected the catalytic activity of the analogues and we demonstrated again the importance of the leucine residue (or at least larger hydrophobic residue) at this position in the sequence regardless the global structure of the catalytic variants. Nevertheless, both elongated analogues did not outcompete the **Het-N2** protein as they did not achieve superior rate acceleration as determined for **Het-N2** (see Table 4.1).

Table 4.1. Kinetic parameters for hydrolysis of thioester- $\alpha$ Mes for the elongated analogues in comparison to the reactions with Het-N2 and without catalyst.

Conditions / catalytic protein	$\tau_{1/2}$ (hours)	$k_{\text{obs}}$ ( $\text{M}^{-1} \cdot \text{s}^{-1}$ )
Background reaction (without catalyst)	96	0.006
<b>Het-N2</b>	1	0.56
<b>Het-elong, GAC</b>	20	0.028
<b>Het-elong, GLC</b>	7	0.079

Same tendency was observed in presence of Tris as acyl acceptor. The assay of the **Het-elong, GAC** protein resulted in an observed rate constant  $k_{\text{obs}}$  of  $0.023 \text{ M}^{-1}\cdot\text{s}^{-1}$ , while **Het-elong, GLC** better accelerated the rate of consumption of the thioester with a  $k_{\text{obs}}$  of  $0.079 \text{ M}^{-1}\cdot\text{s}^{-1}$  (Table 4.2). Therefore, the elongated analogues did not accelerate the rate of the consumption of thioester substrate in presence of Tris (the observed rate even lower for **Het-elong, GAC**) as it was noticed for previously described analogues in subsection 2.6.3.

The major drawback of the catalytic assay of 1<sup>st</sup> generation is the difficulty to assess the ratio of aminolysis over hydrolysis (A/H). At this stage, we cannot deduce if the elongated analogues favored more aminolysis over hydrolysis compared to former analogues. Therefore, it would be interesting, in the future, to test **Het-elong, GAC** and **Het-elong, GLC** with the catalytic assay of the 2<sup>nd</sup> generation, which enables to calculate a relative A/H ratio. Nevertheless, elongated analogues presented a less efficient catalytic activity compared to **Het-N2**, meaning that the design did not improve the activity and that a slightly larger helix macrodipole may not be an important factor to affect the catalysis. To understand the reasons of this lower efficiency, further design and production of resulting analogues would need to be performed.

Table 4.2. Kinetic parameters for concomitant aminolysis and hydrolysis of thioester-<sup>a</sup>Mes for the elongated analogues in comparison to the reactions with Het-N2 and without catalyst.

Conditions / catalytic protein	$\tau_{1/2}$ (hours)	$k_{\text{obs}}$ ( $\text{M}^{-1}\cdot\text{s}^{-1}$ )
Background reaction (without catalyst)	77.5	0.007
<b>Het-N2</b>	0.75	0.74
<b>Het-elong, GAC</b>	24	0.023
<b>Het-elong, GLC</b>	7	0.079



#### 4.6. Biophysical characterization: towards a better stabilization of the structure?

Stability of the designed elongated analogues was assessed in phosphate buffer at pH 6.9 (10 mM sodium phosphate, 500  $\mu$ M TCEP). Mean residual ellipticity was measured by CD spectroscopy for the most efficient extended analogue, **Het-elong, GLC**, at 222 nm, a strong indicator of  $\alpha$ -helicity. Stability of the folding of the heterodimer **Het-elong, GLC** was evaluated against thermal denaturation at increasing concentration of denaturant (*i.e.* 0 M Gn-HCl, 4 M Gn-HCl and 6 M Gn-HCl) and was compared to the stability of the **Het-N2** protein as a reference (see [Figure 4.11](#)). In native conditions, **Het-elong, GLC** stayed substantially folded from 5 °C to 90 °C and showed a larger mean residual ellipticity at 222 nm compared to **Het-N2**. However, at 4 M Gn-HCl complete denaturation was observed for both proteins with a melting temperature of 35 °C for **Het-elong, GLC** and of 46 °C for **Het-N2**, suggesting that the latter is more stable against chemical denaturation compared to the designed **Het-elong, GLC**. The contrary was indeed expected based on the extended hydrophobic interactions of **Het-elong, GLC** scaffold. At 6 M Gn-HCl, both proteins were completely unfolded.

CD measurements demonstrated that our designed extended protein did not present additional stabilization compared to the former **Het-N2**. **Het-elong, GLC** contains more electrostatic interactions in solvent-exposed faces that may be significantly affected at high ionic strength imposed by high concentrations of Gn-HCl. Therefore, it would be interesting to also verify this result in the presence of increased concentrations of urea.

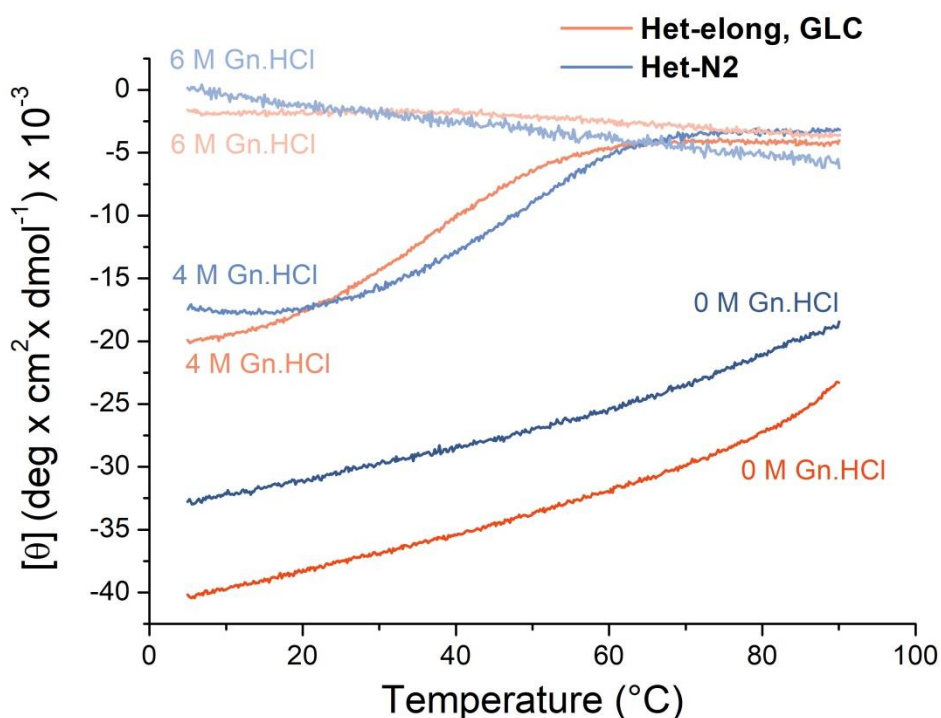


Figure 4.11. Melting curves measured by CD at 222 nm for the catalytic proteins Het-elong, GLC (orange curves) and Het-N2 (purple curves) at different concentrations of guanidine hydrochloride (Gn.HCl).

#### 4.7. Conclusions

In this chapter, the design of an extended analogue is described. Through different computational tools and software, each helix of the coiled-coil topology of the former protein **Hom-N2** was elongated and sequences were optimized. Consequently, we expected a better stabilization of the folding thanks to the extended hydrophobic core and an enhancement of the macrodipole effect on catalysis (*i.e.* via hypothesized higher  $pK_a$  depression of the catalytic cysteine and boosted interaction with the oxyanion of the tetrahedral intermediates). The resulting designed protein was indeed an asymmetric heterodimer, in which the strand containing the catalytic cysteine was extended by two heptads (one at the C-terminus and the other at the N-terminus) and the other strand was extended by one heptad at its N-terminus. The sequence was then optimized, and several positions including hydrophobic residues were modified. Two elongated heterodimers **Het-elong, GAC** and **Het-elong, GLC** generated by the design process were synthesized chemically using SPPS techniques followed by ligation methodologies. The formation of the intended heterodimer was confirmed with the analogue **Het-elong, GAC** by X-ray crystallography. The obtained crystal structure

overlays well with the designed model with a RMSD on C $\alpha$  of 1.77 Å. The small shift observed between the two backbones can be explained by the oxidation of the catalytic cysteine, which may constrain the protein scaffold. Nevertheless, Crick's parametrization of the crystal structure of coiled-coil **Het-elong, GAC** correlates well with the parametrization of the designed model and the former **Hom-N2**. Only slight fluctuations are noticed between the three scaffolds as shown in **Table 4.3**. Catalytic activity of **Het-elong, GLC** was finally higher than **Het-elong, GAC**, proving one more time the importance of the hydrophobic residue and especially leucine amino acid at N1 position next to the catalytic cysteine. However, **Het-elong, GLC** was less efficient than the initial **Het-N2**, suggesting that increasing the  $\alpha$ -helix of one heptad does not affect much the catalytic activity of the variants. Moreover, we showed that **Het-elong, GLC** was less stable against chemical denaturation than **Het-N2**.

In conclusion, we achieved the synthesis and the folding of our designed extended analogues. However, they did not meet our expectation in terms of catalysis and stability. A lot of changes were applied at the same time (*i.e.* elongation of the protein framework, hydrophobic core repacking, modification of the charge patterning and the loop sequence and few substitutions of polar residues at *f* position of heptads), therefore it complicates the interpretation of such results. To understand the origin of catalysis efficiency and stability reduction, we contemplate performing synthesis of other analogues by focusing on one modification after another. We would first synthesize an extended analogue by keeping original heptad sequences (thus keeping the original charge patterning and hydrophobic core) to first verify our hypothesis of the macrodipole effect and overall stabilization by increasing the hydrophobic interactions. Then, we would modify the hydrophobic packing on the elongated analogue and on the former **Het-N2** protein to check any possible impact on the folding stability. Finally, different charge patterning may be tested to identify the one providing the larger destabilization of homodimer formation.

**Table 4.3. Comparison of the key Crick parameters for the coiled-coil structure of the former Hom-N2, the designed model and the resulting Het-elong, GAC.**

<b>Crick parameters</b>	<b>Hom-N2</b>	<b>Designed model</b>	<b>X-ray structure of Het-elong, GAC</b>
R <sub>0</sub> (Å)	6.620	6.373	6.092
$\omega_0$ (°/res)	-2.457	-2.776	-2.841
Pitch angle $\alpha$ (°)	-10.912	-12.012	-11.454
Pitch (Å)	215.763	188.131	188.922
RMSD (Å)	0.539	0.775	0.720

## 4.8. References

- (1) Huang, P.-S.; Boyken, S. E.; Baker, D. The Coming of Age of de Novo Protein Design. *Nature* **2016**, *537* (7620), 320–327.
- (2) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **2003**, *302* (5649), 1364–1368.
- (3) Dawson, W. M.; Rhys, G. G.; Woolfson, D. N. Towards Functional de Novo Designed Proteins. *Curr. Opin. Chem. Biol.* **2019**, *52*, 102–111.
- (4) Crick, F. H. C. The Fourier Transform of a Coiled-Coil. *Acta Crystallogr.* **1953**, *6* (8–9), 685–689.
- (5) Crick, F. H. C. The Packing of  $\alpha$ -Helices: Simple Coiled-Coils. *Acta Crystallogr.* **1953**, *6* (8–9), 689–697.
- (6) Grigoryan, G.; DeGrado, W. F. Probing Designability via a Generalized Model of Helical Bundle Geometry. *J. Mol. Biol.* **2011**, *405* (4), 1079–1100.
- (7) Harbury, P. B.; Tidor, B.; Kim, P. S. Repacking Protein Cores with Backbone Freedom: Structure Prediction for Coiled Coils. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92* (18), 8408–8412.
- (8) Wood, C. W.; Bruning, M.; Ibarra, A. Á.; Bartlett, G. J.; Thomson, A. R.; Sessions, R. B.; Brady, R. L.; Woolfson, D. N. CCBUILDER: An Interactive Web-Based Tool for Building, Designing and Assessing Coiled-Coil Protein Assemblies. *Bioinformatics* **2014**, *30* (21), 3029–3035.
- (9) Thomson, A. R.; Wood, C. W.; Burton, A. J.; Bartlett, G. J.; Sessions, R. B.; Brady, R. L.; Woolfson, D. N. Computational Design of Water-Soluble  $\alpha$ -Helical Barrels. *Science* **2014**, *346* (6208), 485–488.
- (10) Dutzler, R.; Campbell, E. B.; Cadene, M.; Chait, B. T.; MacKinnon, R. X-Ray Structure of a CIC Chloride Channel at 3.0 Å Reveals the Molecular Basis of Anion Selectivity. *Nature* **2002**, *415* (6869), 287–294.
- (11) Hol, W. G. J.; van Duijnen, P. T.; Berendsen, H. J. C. The  $\alpha$ -Helix Dipole and the Properties of Proteins. *Nature* **1978**, *273* (5662), 443–446.
- (12) Zhou, J.; Grigoryan, G. Rapid Search for Tertiary Fragments Reveals Protein Sequence–Structure Relationships. *Protein Sci.* **2015**, *24* (4), 508–524.
- (13) Ogihara, N. L.; Ghirlanda, G.; Bryson, J. W.; Gingery, M.; DeGrado, W. F.; Eisenberg, D. Design of Three-Dimensional Domain-Swapped Dimers and Fibrous Oligomers. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (4), 1404–1409.

## **Chapter 5. An unexpected proteolytic activity?**

## 5.1. Introduction

*De novo* protein design represents a promising tool to develop efficient catalysts for biological and abiological reactions. Thanks to their modularity, ease of handling, enhanced stability as well as their catalytic efficiency in green solvents and notably in water and their biodegradability, *de novo* protein frameworks are of great interest for practical applications in academic research and in industrial processes.

Hydrolysis of various substrates such as ester and phosphodiester has been greatly explored by *de novo* designed protein catalysts.<sup>1</sup> Various catalytic strategies have been investigated using *de novo* scaffolds such as the incorporation of a coordination site for metal ions (e.g. zinc ions as in most metalloproteases),<sup>2,3</sup> the construction of amino acids networks predominantly for p*K*<sub>a</sub> depression<sup>4,5</sup> and introduction of catalytic triad into the lumen of helical barrels.<sup>6</sup> Many examples of *de novo* hydrolases found in literature could be explained by the presumed simplicity of the catalyzed reaction because one of the reagents corresponds to the solvent (*i.e.* water) surrounding the accessible catalytic site. Although pure design process of *de novo* catalytic proteins has rarely reached catalytic efficiencies of natural hydrolases, multiple rounds of laboratory evolution have demonstrated the possibility to attain such a effectiveness with the production of a *de novo* metalloenzyme carrying highly proficient esterase activity and stereospecificity.<sup>7</sup> Nevertheless, substantial catalytic activity for hydrolysis of unnatural substrates has been achieved and may find practical applications such as the degradation of pollutants and poisonings (e.g. *de novo* catalytic amyloids have catalyzed the hydrolysis of paraoxon, a largely used and toxic organophosphate pesticide).<sup>8</sup>

The more specialized class of hydrolases bearing proteolytic activity has never been designed from scratch with *de novo* protein scaffold. In fact, building proteolytic function from scratch represents a much more challenging and complex task because it involves the hydrolysis of peptide bonds, which display a better stability compared to ester linkage. Natural or engineered proteases find numerous applications, in particular for the degradation of biological materials. Notable examples of use of natural and generally optimized proteases have been demonstrated for the isolation of nucleic acids, tissue dissociation and digestion of proteins for proteomic studies, as well as for laundry detergents and enzyme-based therapeutics.<sup>9-11</sup>

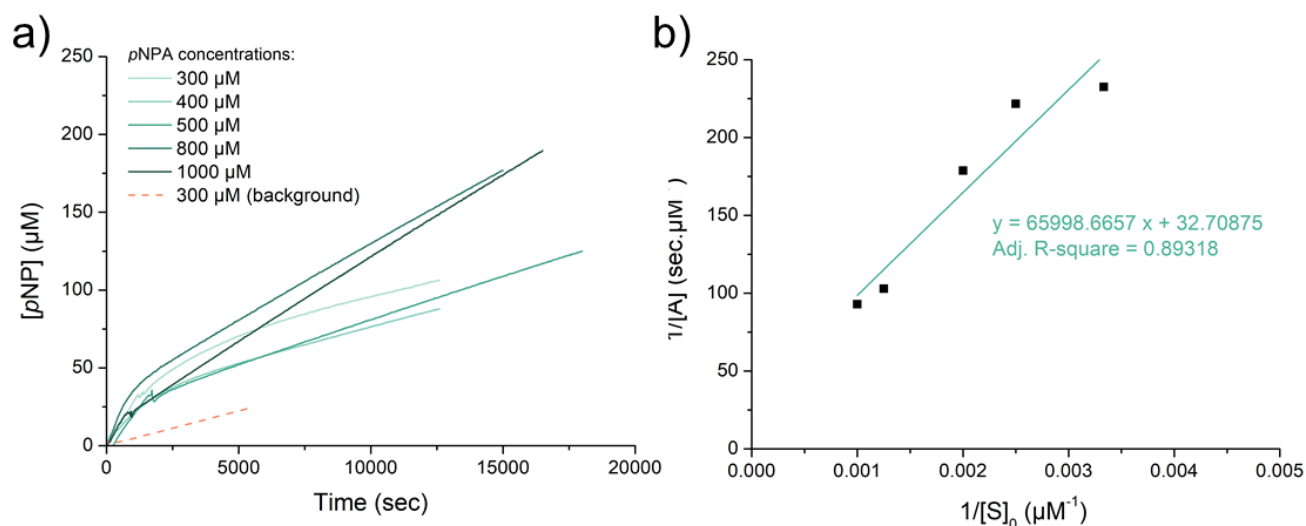
In our case, we have already demonstrated catalysis of hydrolysis on activated acyl group (*i.e.* peptide- $\alpha$ thioester and *para*-nitrophenyl acetate) with our *de novo* catalytic proteins and in particular with **Het-N2** heterodimer. Here, we will focus on the characterization of hydrolytic activity of the homodimer analogue **Hom-N2** and we will

present preliminary results that suggest a potential proteolytic activity. This chapter describes new finding for our system, which are put into perspective for further exploration.

## 5.2. Hydrolytic activity of the Hom-N2 analogue

### 5.2.1. Hydrolytic activity with pNPA as substrate

First, hydrolytic activity was evaluated with the well-established continuous *p*-nitrophenylacetate (*p*NPA) based assay. In the mechanism depicted in **Figure 2.33a**, the chromophore *p*-nitrophenol (*p*NP) is released via reaction of the cysteine residue of **Hom-N2** on the *p*NPA substrate, which permits monitoring the reaction by UV ( $\epsilon = 5400 \text{ L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$  at 348 nm). The last step in the mechanism corresponds to the hydrolysis of the acetyl-cysteine thioester adduct which upon reaction with water releases back the free protein catalyst. Each step passes through a tetrahedral intermediate bearing an oxyanion that is assumed to interact with unpaired NHs of **Hom-N2** via stabilizing hydrogen bonds. Observed kinetic curves showed a burst phase that corresponds to the rapid formation of the covalent adduct and a linear phase that is related to the rate-limiting step of the acetate release (**Figure 5.1a**) thus supporting the ping-pong mechanism of the catalyzed hydrolysis of *p*NPA by **Hom-N2** protein as previously described. For each tested condition (constant enzyme concentration at 50  $\mu\text{M}$  and a variable substrate concentration from 300  $\mu\text{M}$  to 1000  $\mu\text{M}$ ), kinetic curves were measured independently and kinetic parameters were calculated by fitting the equation  $[pNP] = At + B(1 - e^{-bt})$ . By analyzing the kinetic equations for **Hom-N2** at different  $[pPNA]_0$ , the double-reciprocal plot was deduced (**Figure 5.1b**) allowing to determine  $k_{\text{cat}}$  of  $6.11 \times 10^{-4} \text{ sec}^{-1}$  and  $K_m$  of 2.0 mM, corresponding to catalytic efficiency ( $k_{\text{cat}}/K_m$ ) of  $0.31 \text{ M}^{-1} \text{ sec}^{-1}$ . In comparison with **Het-N2**, preliminary kinetic parameters for the hydrolysis of *p*NPA are rather similar, whereas in acyl transfer reactions on peptide- $\alpha$ thioester substrate **Het-N2** provided greater catalytic efficiency than **Hom-N2**. This difference can be explained by the two catalytic cysteines in close proximity in **Hom-N2** protein, which may engender steric hindrance in the presence of larger substrates like peptide- $\alpha$ thioester. In contrast, the *p*NPA substrate and the resulting acetyl-cysteine thioester adduct may be not affected by the second catalytic cysteine present in the homodimer protein. At the same time, the large  $K_m$  value suggests the absence of proper binding pocket for the substrate. Compared to previously reported *de novo* hydrolases, the catalytic efficiencies for *p*NPA hydrolysis by **Hom-N2** and by **Het-N2** are nearly 10-times lower, however, no additional catalytic residues



**Figure 5.1. Characterization of the hydrolytic activity of Hom-N2 thanks to the continuous assay using pNPA as substrate.** **a)** Pre-steady-state kinetic curves of **Hom-N2** at different concentration of pNPA substrate (green curves). Burst phase and linear phase were well observed and fitted with the equation:  $[pNP] = At + B(1 - e^{-bt})$ . Background reaction is depicted in dashed orange line. **b)** Double-reciprocal plot for the hydrolytic activity of **Hom-N2** with pNPA as substrate.

(e.g. histidine) are present in our catalytic proteins, thus the catalytic effect emerges mainly from the interactions with N-terminus of  $\alpha$ -helix.<sup>1,6</sup>

The intrinsic hydrolytic activity on the standard ester substrate has been demonstrated for our homodimer protein **Hom-N2**. In the next subsection I will discuss the possible structural and compositional changes that result in a completely novel catalytic site. The obtained crystal structure of **Hom-N2** stimulated new hypotheses and a possible new design.

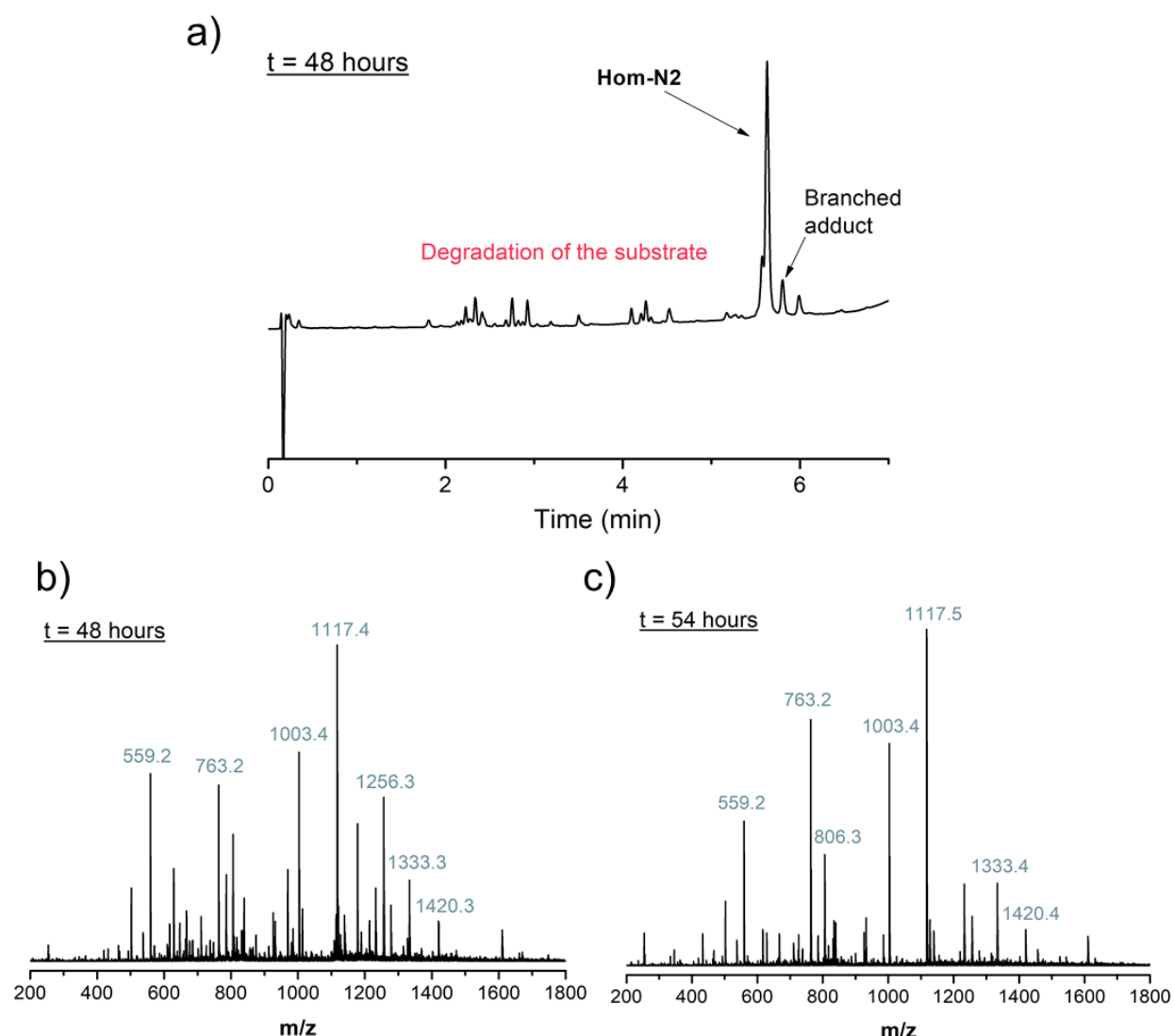
### 5.2.2. Unexpected proteolytic activity: a discovery by serendipity

While monitoring catalytic assays of the 1<sup>st</sup> generation in presence of the homodimer with N2 cysteine, **Hom-N2**, we observed degradation of the 30-residue thioester substrate [N-ACTR]-<sup>o</sup>Mes. Reaction conditions were set as follow: 0.5 mM [N-ACTR]-<sup>o</sup>Mes in presence of an equimolar amount of **Hom-N2** (0.5 mM) were mixed together in phosphate buffer (50 mM sodium phosphate, 100 mM NaCl) with 5 mM of TCEP at pH 7.0. It is noteworthy that the protein **Hom-N2** was synthesized with two different protocols: by a two-segment ligation or by a one-pot three-segment ligation leading to either the original alanine residue or an alkylated cysteine at the 24<sup>th</sup> position in sequence, respectively (see subsection 2.3.2 and 2.3.3). Both sequences of **Hom-N2** were assayed and resulted in identical acceleration of hydrolysis of the thioester substrate. Similarly, degradation of [N-ACTR]-<sup>o</sup>Mes was observed



after 48 hours or 72 hours of reaction with both sequences of **Hom-N2**, confirming that modification at position 24, which is rather far from the catalytic cysteines, does not affect the catalytic activity of the analogues.

Decomposition of the peptide [N-ACTR]-<sup>α</sup>Mes resulted in appearance of shorter peptide fragments of the substrate and was observed in three independent catalytic assays. One example of substrate degradation with HPLC and LC/MS profiles is shown in **Figure 5.2**. We were able to correlate detected masses to possible peptide fragments, which are recapitulated in **Table 5.1** for the three separate assays. To do so, we generated all masses for all possible fragments of different sizes coming from the [N-ACTR]-<sup>α</sup>Mes sequence thanks to a homemade python script and we selected only peptide segments with a difference of  $\pm 1$



**Figure 5.2.** Example of HPLC and LC/MS profile of the degradation of the [N-ACTR]-<sup>α</sup>Mes peptide. **a)** HPLC trace after 48 hours of reaction. The substrate is decomposed into smaller peaks. **b+c)** Mass profile corresponding to the different peptide fragments emerging from the degradation of [N-ACTR]-<sup>α</sup>Mes after **b)** 48 hours and **c)** 54 hours of reaction.

**Table 5.1. Detected peptide fragments in the three independent experiments.** Cutting sites are represented by a “|”

MW <sub>obs</sub> (Da)	MW <sub>calc</sub> (Da)	Corresponding fragment (in black)	Found in exp. 1	Found in exp. 2	Found in exp. 3
1610.5	1610.6 1610.7	EGQSDERALLDQLH   T D   QLHTLLSNTDATGLE   E	✓		
1419.4	1420.3	L   SNTDATGLEEIDR	✓		✓
1231.4	1231.2 1232.2	E   GQSDERALLDQ   L EGQSDERALLD   Q	✓	✓	✓
1218.3	1219.2	N   TDATGLEEIDR		✓	✓
1212.7	1212.3	D   QLHTLLSNTDA   T	✓		
1177.3	1177.1	S   NTDATGLEEID   R	✓		
1116.4	1117.1	EGQSDERALL   D	✓	✓	✓
1012.3	1013.0	Q   LHTLLSNTD   A	✓		
1002.4	1003.0	D   ATGLEEIDR	✓	✓	✓
951.3	952.1	A   LLDQLHTL   L			✓
931.4	931.9	A   TGLEEIDR	✓	✓	✓
906.0	906.8	L   SNTDATGLE   E			✓
846.2	846.9 846.9	T   DATGLEEI   D D   ATGLEEID   R			✓
838.2	839.0 839.0 839.0	A   LLDQLHT   L L   LDQLHTL   L L   DQLHTLL   S	✓		✓
833.1	833.8 833.8	H   TLLSNTDA   T T   LLSNTDAT   G	✓		
830.2	830.8	S   DERALLD   Q T   GLEEIDR		✓	✓
819.1	819.7 819.7	EGQSDER   A S   NTDATGLE   E			✓
762.2	762.7	H   TLLSNTD   A		✓	✓
737.2	737.9	A   LLDQLH   T	✓		✓
725.2	725.8	L   LDQLHT   L L   DQLHTL   L	✓	✓	✓
715.3	715.7 715.7	S   DERALL   D D   ERALLD   Q			✓
705.1	705.7	N   TDATGLE   E			✓
682.8	682.8	Q   LHTLLS   N		✓	✓
674.0	674.7	T   GLEEID   R			✓

Da compared to the observed molecular weight (error range of the ESI-MS). Some masses were commonly found in all three experiments, whereas for others it was detected only once or two times. These observations suggest that the scissile bonds are rather random without any clear pattern for the hydrolysis site.

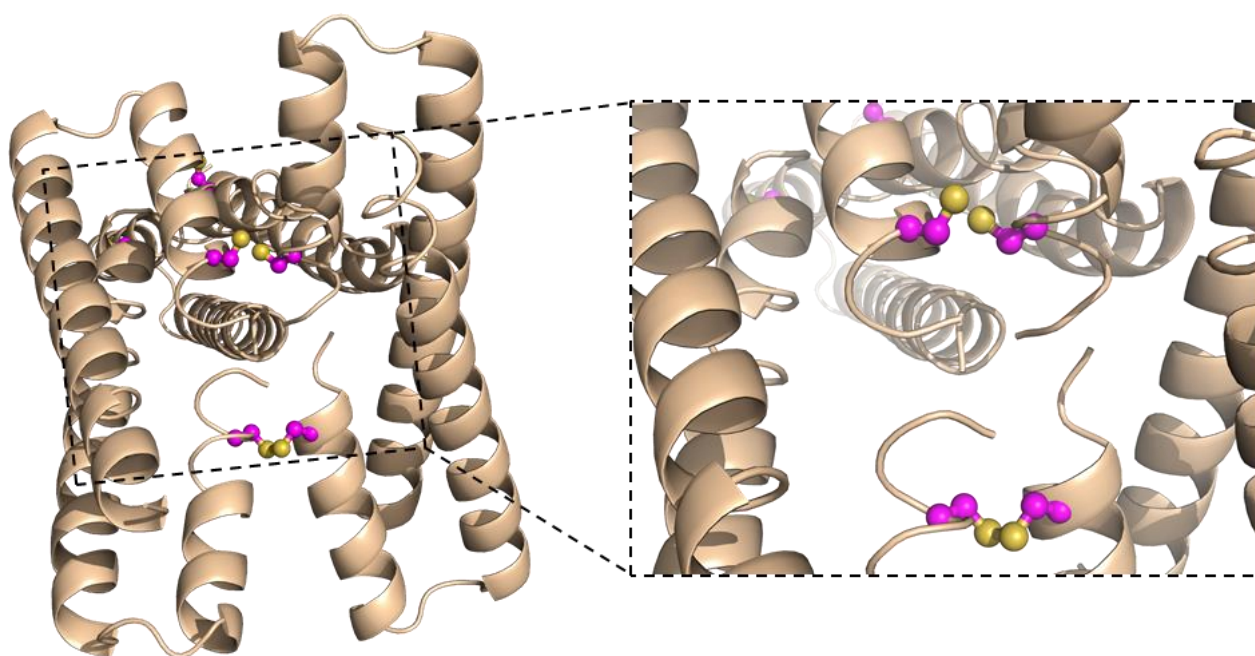
Overall, the **Hom-N2** catalytic protein stayed intact in the course of the reaction but additional peaks started also to appear slowly indicating a less intense decomposition compared to the substrate peptide. Importantly, the peptide [N-ACTR]-<sup>α</sup>Mes is coming from an intrinsically disordered protein that are known to be less resistant to protease degradation as they lack stable three-dimensional structure.

To better understand the origin of such protease activity and especially whether it is coming from bacterial contamination or from our **Hom-N2** protein, we envisaged to perform catalytic assay in 'sterile' conditions. In the laboratory, the easiest way was to add in the buffer 0.05 % (w/v) of sodium azide (NaN<sub>3</sub>) that is a well-known preservative and biocide usually found in stock solutions to avoid bacterial growth and then to filter the resulting buffer using a 0.22 μM (pore size) sterile membrane that typically retains bacteria. Although we did not noticed any degradation in this condition, we indeed discovered that NaN<sub>3</sub> was inducing acceleration of the catalytic reaction (half-life of [N-ACTR]-<sup>α</sup>Mes substrate of 3 h in presence of NaN<sub>3</sub> instead of 24 h without), which we finally confirmed during the detailed catalytic characterization of **Het-N2** (see subsection 2.7). As NaN<sub>3</sub> interfered in the reaction between the substrate and our **Hom-N2** protein, we could not surely certify at this stage that the degradation is derived from bacterial contamination. Further investigations are necessary and we can imagine sterile conditions without the addition of chemical compounds such as the use of autoclaved vessels, filtered buffer and manipulation in a sterilized zone (e.g. next to a flame and laminar flow hood).

Even if we are not certain that the proteolytic activity emerged from the homodimer protein **Hom-N2**, we actually noticed interesting structural topology in the crystal structure (**Figure 5.3**). Indeed, oxidation of two cysteine residues covalently links two different homodimers generating a specific arrangement in which the four N-termini are in close proximity to each other. The resulting tetramer is in contact with a third homodimer via one of its helical C-termini and one of its loops. This particular three-dimensional pattern creates a cavity with two other cysteine residues facing each other thus providing a completely different catalytic environment compared to the homodimer alone. We reasoned that these two cysteine residues in close proximity may affect their own reactivity through pK<sub>a</sub> depression for example, which would enable the attack on peptide bonds with subsequent cleavage via hydrolysis. One can also propose the analogy with aspartic proteases that catalyze peptide bond hydrolysis by activating water through deprotonation and by general

acid/base catalysis with the assistance of two aspartate/aspartic acid residues facing each other.<sup>12</sup> The intriguing decomposition of our substrate via peptide bond hydrolysis was detected after at least 48 hours of reaction that may correspond to the time for the protein **Hom-N2** to oligomerize through cysteine oxidation (the reducing agent, TCEP, is not very stable in phosphate buffer) and consequently forming such a peculiar catalytic site.

Proteolytic activity may certainly arise from bacterial contamination, it nevertheless challenges our way of thinking on our catalytic system. This crystal structure is finally a good starting point to design protease activity that would represent an unrivaled *de novo* catalytic protein.



**Figure 5.3. Crystal structure of the protein Hom-N2.** The asymmetric unit is composed of three copies of the homodimer. Two copies of the protein are covalently joined by a single disulfide bond that resulted due to oxidation of the two corresponding cysteines. Image on the right represents a zoomed region (dashed black box of left structure) of the four N-termini that are in close proximity to each other, due to the disulfide bridge, forming a deep cavity with two other cysteine residues facing to each other. The catalytic cysteines are depicted in magenta with sulfur atom in yellow.

### 5.3. Conclusions and perspectives

Finally, this chapter provides an alternative design strategy to develop *de novo* catalytic proteins bearing novel functionalities, while starting from the same protein scaffold. We focused on the potential capacity of our homodimer analogue, **Hom-N2**, to catalyze hydrolysis and more specifically proteolysis without any restriction, which would find notable applications. We demonstrated that the **Hom-N2** protein possesses intrinsic hydrolytic activity for *p*NPA substrate, which mainly emerges from the unpaired NHs of the two N-termini stabilizing the tetrahedral intermediates via H-bonding. On the other hand, using peptide-<sup>a</sup>Mes as substrate in catalytic assay of 1<sup>st</sup> generation in presence of **Hom-N2**, we noticed in distinct and independent experiments degradation of the peptide substrate in shorter fragments with time. The cuts in the sequence and resulting peptide segments were identified. Some cleavage sites were redundant and it was coming from both endo- and exo-protease activity without definite preferences. Although this peculiar protease activity could arise from bacterial contamination, such observations can be a good starting point to consider other applications for our system and more particularly to consider building protease activity with our **Hom-N2** protein as a starting framework. Indeed, crystal structure of the homodimer revealed an interesting spatial organization of three copies of the protein, which was driven by the oxidation of two cysteines residues from distinct **Hom-N2**. The resulting disulfide bridge brought together the four N-termini with two free cysteine residues facing each other and formed a cavity where a novel catalytic site can be installed. To specifically incorporate a proteolytic activity in such a three-dimensional arrangement, we can refer to natural proteases and especially papain, a cysteine protease, that also exploits the N-terminus of an  $\alpha$ -helix with the catalytic cysteine.<sup>13</sup>

With our system and our homodimer **Hom-N2**, we hypothesize the formation of a distinct catalytic site that differs from the original one and that may result in the emergence of novel catalytic activity through oligomerization driven by cysteine oxidation and leading to a unique globular shape. This actually constitutes a plausible enzyme evolutionary pathway in which primordial catalytic activity would emerge from the super-assembly of simple peptide sequences. One interesting features of such primitive enzymes is their speculated substrate and reactivity promiscuity. Thus, we can imagine a similar pathway in order to evolve our **Hom-N2** towards a promiscuous protease-like *de novo* catalytic protein.

To verify our hypothesis without further design process as a first attempt, we plan oligomerization experiments of the **Hom-N2** protein with subsequent catalytic characterization. In other words, we would let oxidize the homodimer and we would thus expect the spontaneous production of new oligomerization states via disulfide bridges

formation. The formation of higher-order structure would be monitored by either analytical size exclusion chromatography or native mass spectrometry. At different time points corresponding to various oligomerization states, aliquots would be taken off to verify the potential enhancement of the hydrolytic activity efficiency of *p*NPA substrate compared to the homodimer alone as previously shown. Simultaneously, we will verify the emergence of new catalytic activity such as  $\beta$ -lactamase activity (with nitrocefin as chromogenic substrate) and protease activity (instantaneous degradation of peptide substrate).

## 5.4. References

- (1) Marshall, L. R.; Zozulia, O.; Lengyel-Zhand, Z.; Korendovych, I. V. Minimalist de Novo Design of Protein Catalysts. *ACS Catal.* **2019**, *9* (10), 9265–9275.
- (2) Der, B. S.; Edwards, D. R.; Kuhlman, B. Catalysis by a De Novo Zinc-Mediated Protein Interface: Implications for Natural Enzyme Evolution and Rational Enzyme Engineering. *Biochemistry* **2012**, *51* (18), 3933–3940.
- (3) Zastrow, M. L.; Peacock, A. F. A.; Stuckey, J. A.; Pecoraro, V. L. Hydrolytic Catalysis and Structural Stabilization in a Designed Metalloprotein. *Nat. Chem.* **2012**, *4* (2), 118–123.
- (4) Broo, K. S.; Brive, L.; Ahlberg, P.; Baltzer, L. Catalysis of Hydrolysis and Transesterification Reactions of p-Nitrophenyl Esters by a Designed Helix–Loop–Helix Dimer. *J. Am. Chem. Soc.* **1997**, *119* (47), 11362–11372.
- (5) Razkin, J.; Lindgren, J.; Nilsson, H.; Baltzer, L. Enhanced Complexity and Catalytic Efficiency in the Hydrolysis of Phosphate Diesters by Rationally Designed Helix-Loop-Helix Motifs. *Chembiochem* **2008**, *9* (12), 1975–1984.
- (6) Burton, A. J.; Thomson, A. R.; Dawson, W. M.; Brady, R. L.; Woolfson, D. N. Installing Hydrolytic Activity into a Completely de Novo Protein Framework. *Nat. Chem.* **2016**, *8* (9), 837–844.
- (7) Studer, S.; Hansen, D. A.; Pianowski, Z. L.; Mittl, P. R. E.; Debon, A.; Guffy, S. L.; Der, B. S.; Kuhlman, B.; Hilvert, D. Evolution of a Highly Active and Enantiospecific Metalloenzyme from Short Peptides. *Science* **2018**, *362* (6420), 1285–1288.
- (8) Lengyel, Z.; Rufo, C. M.; Moroz, Y. S.; Makhlynets, O. V.; Korendovych, I. V. Copper-Containing Catalytic Amyloids Promote Phosphoester Hydrolysis and Tandem Reactions. *ACS Catal.* **2018**, *8* (1), 59–62.
- (9) Mótyán, J. A.; Tóth, F.; Tózsér, J. Research Applications of Proteolytic Enzymes in Molecular Biology. *Biomolecules* **2013**, *3* (4), 923–942.
- (10) Vojcic, L.; Pitzler, C.; Körfer, G.; Jakob, F.; Ronny Martinez; Maurer, K.-H.; Schwaneberg, U. Advances in Protease Engineering for Laundry Detergents. *New Biotechnol.* **2015**, *32* (6), 629–634.
- (11) Craik, C. S.; Page, M. J.; Madison, E. L. Proteases as Therapeutics. *Biochem. J.* **2011**, *435* (1), 1–16.
- (12) Brik, A.; Wong, C.-H. HIV-1 Protease: Mechanism and Drug Discovery. *Org. Biomol. Chem.* **2003**, *1* (1), 5–14.
- (13) Hol, W. G. J.; van Duijnen, P. T.; Berendsen, H. J. C. The  $\alpha$ -Helix Dipole and the Properties of Proteins. *Nature* **1978**, *273* (5662), 443–446.





## **Chapter 6. General conclusions, significance and perspectives**

Installing from scratch a catalytic activity on a protein scaffold is a complex task. Mimicking enzymes and equaling their great catalytic efficiency are challenging as enzyme activity has been finely tuned through evolutionary process. Nevertheless, by studying protein and enzyme properties, some fundamental and physical rules have been established. Our knowledge can be scrutinized by building from scratch novel protein structures and by incorporating catalytic activity for various reactions in natural and unnatural scaffolds. Designed *de novo* protein scaffolds generally present well-organized and miniaturized folds, which can have several advantages over larger and natural proteins including the ease to produce and to handle, the robustness, the modularity and the enhanced stability against chemical and thermal denaturation. By using a comprehensive design process, unique catalytic activities were successfully installed into *de novo* proteins resulting in hydrolases,<sup>1,2</sup> oxidoreductase,<sup>3</sup> retro-aldolase,<sup>4</sup> and others.

Our goal is to develop a *de novo* protein capable of catalyzing peptide ligation and more broadly amide bond formation through addition/elimination (A/E) reactions on thioesters without stringent restrictions at the junction site. The resulting protein would represent a highly useful chemical tool for a wide range of applications like peptide ligation and cyclization using any amino acids including non-canonical residues at ligation junction, and can also be used for protein labeling. This thesis demonstrated the feasibility of the main concept and summarizes the completion of the first steps in this long-termed project by building a completely *de novo* protein with predefined acyl transfer catalytic activity. We applied minimalist design, which implies only few assumptions and chemical intuition, by constructing an oxyanion hole within a small cavity containing an adjacent thiol nucleophile. To do so, we exploited the already designed domain-swapped dimer (DSD) protein and introduced a catalytic site in the N-terminal region in order to profit from the unpaired hydrogen-bond donors to stabilize negatively charged tetrahedral intermediates of A/E reactions at acyl group. We proved the robustness and the modularity of the DSD scaffold, which has accepted numerous modifications (*i.e.* transformation into heterodimer, Lys to Arg substitution, cysteine incorporation as catalytic residue) while preserving its three-dimensional shape as confirmed by crystal structures. By placing a unique catalytic cysteine residue at N2 position (second residue at N-terminus of  $\alpha$ -helix) of one strand leading to the heterodimer **Het-N2**, we clearly demonstrated acceleration of a diverse set of acyl transfer reactions (aminolysis, esterification concomitant with hydrolysis) on peptide- $\alpha$ thioesters. We also showed the importance of the full-length and folded structure for catalysis as well as the environment of the cysteine. This minimal set of functional elements is actually sufficient for the emergence of catalytic activity in a *de novo* protein, which can be therefore compared to primitive enzymes with modest activity and promiscuity of reactivity and substrates. Thanks to the modularity of the *de novo* protein scaffold, future steps will be focused on a better

understanding of the parameters that influence the reaction outcome and the catalytic efficiency. The major goal would be to perceive how to control the ratio of aminolysis over hydrolysis while enhancing catalytic rates. Importantly, as we would like to develop a general tool for various acyl transfer reactions, we will not necessarily aim at reaching catalytic efficiency of enzymes that is indeed strongly intertwined with reaction specificity. Thus, this optimal balance between promiscuity of substrates and catalytic efficiency will have to be considered.

By taking advantage of chemical synthesis and of computational modeling tools, we have started to explore different strategies to better understand our system and at the same time to attempt to enhance its catalytic capabilities. On one hand, libraries of analogues for both strands were produced by employing total chemical synthesis techniques via convergent or divergent approaches. For each analogue, a maximum of two modifications were introduced in close proximity to the catalytic N2-cysteine based on key examples of the literature, intuition and rational modeling in order to better stabilize the tetrahedral intermediates via additional H-bonding donors or to modulate the reactivity of the catalytic cysteine residue. Although no significant improvement in the catalyzed reaction rate was yet achieved (*i.e.* the **Het-N2** analogue staying one of the most efficient proteins studied in this work), subtle tendencies of the impact of modifications on kinetic parameters were noticed. Clearly, further iterative cycles are necessary in order to build more complex catalytic site and network of amino acids. To proceed with further rounds of optimization and to orient the design of the next generation of protein variants, detailed structural and functional characterizations of key analogues of the synthesized libraries will be performed. We will also evaluate the dependence of catalytic activity as a function of different substrates by characterizing the synthesized libraries of acyl donor and acyl acceptor peptide substrates. We expect that such results will guide us to further ideas how to improve the design and the effectiveness of our system. On the other hand, we also explored the possibility to modify our protein scaffold by extending the three- $\alpha$ -helix bundle and by simultaneously optimizing the sequence in order to provide a better stability against highly denaturing conditions and to enhance the reactivity of the designed analogue via helix macrodipole effect. We successfully obtained new analogues judged by high accuracy of superimposition of the resulting crystal structure and the designed model. However, stability and catalytic activity were not enhanced compared to the former **Het-N2**. Based on the existing experimental data, the decrease of stability and activity is not well understood and further investigations are required. We will notably consider the stepwise modification of the scaffold and the sequence to verify the influence of each single change. Such outcome will direct the next steps of computational design in order to adjust each parameter and to reach our expectations.

Furthermore, thanks to the chemical methods used to produce our catalytic proteins we are not restricted to natural and common amino acids and we are also able to incorporate other building blocks to take advantage of their unique properties. Additional attempts will be performed to introduce a urea linkage at the N-terminus providing an extra hydrogen-bond donor for enhancing the stabilization of the oxyanion intermediates. Effect on the substitution of the catalytic cysteine by a selenocysteine will be also explored. In fact, selenol groups are known to bear higher nucleophilicity and acidity in comparison to their thiol counterparts.<sup>5</sup> Such properties of selenol moiety and selenocysteine have already been exploited to increase the rate of native chemical ligation for example.<sup>6</sup> Moreover, catalytic serine of the protease subtilisin was converted to selenocysteine leading to the selenosubtilisin that highly promoted aminolysis compared to the wild type enzyme and the thiolsubtilisin analogue (where catalytic serine was converted to a cysteine residue).<sup>7</sup> Therefore, by incorporating this residue instead of the original cysteine, we predict similar influence on the catalysis and on the reaction outcome by favoring the aminolysis pathway to a larger extent. Nevertheless, resulting analogues may be more constraining to handle due to the higher potency of selenocysteine to oxidize (and also to deselenize in presence of phosphine).

Finally, we will also utilize the crystal structure of the homodimer analogue **Hom-N2** forming a higher-order assembly through cysteine oxidation to expand our range of applications. Following plausible enzyme evolutionary pathway in which primordial catalytic activity would emerge from the super-assembly of simple peptide sequences, we aim to attain novel catalytic activity and more specifically proteolytic activity. Through oligomerization, we expect the formation of new distinct catalytic sites with notably two cysteines facing each other resembling to the crystallographic arrangement. Oligomerization experiments will first be performed to verify if novel catalytic functionalities can randomly emerge. This will be followed by more advanced computational design starting from crystal structure. If our hypothesis turns out to be valid, we would then envisage constructing higher-order structure via more stable linkage (*i.e.* alkylation of a cysteine with a non-canonical amino acid like  $\beta$ -iodoalanine) compared to disulfide bridges that are sensitive to their environment and to redox reactions.

In conclusion, we demonstrated the proof-of-concept of our approach by building a *de novo* catalytic protein from scratch carrying preordained modest acyl transfer activity. Thanks to the established robustness and modularity of our protein scaffold, we can now envisage a wide range of strategies for optimization. This challenging project will demand a long-termed and iterative process to stepwise improve the catalytic efficiency and to better control the reaction outcome of our designed proteins. In the meantime, we expect to bring a better understanding of the fundamental principles and requirements of peptide bond formation catalysis by enzyme as well as the intricate features of thioesters' reactivity.

## References

- (1) Broo, K. S.; Brive, L.; Ahlberg, P.; Baltzer, L. Catalysis of Hydrolysis and Transesterification Reactions of P-Nitrophenyl Esters by a Designed Helix–Loop–Helix Dimer. *J. Am. Chem. Soc.* **1997**, *119* (47), 11362–11372.
- (2) Burton, A. J.; Thomson, A. R.; Dawson, W. M.; Brady, R. L.; Woolfson, D. N. Installing Hydrolytic Activity into a Completely de Novo Protein Framework. *Nat. Chem.* **2016**, *8* (9), 837–844.
- (3) Watkins, D. W.; Jenkins, J. M. X.; Grayson, K. J.; Wood, N.; Steventon, J. W.; Le Vay, K. K.; Goodwin, M. I.; Mullen, A. S.; Bailey, H. J.; Crump, M. P.; MacMillan, F.; Mulholland, A. J.; Cameron, G.; Sessions, R. B.; Mann, S.; Anderson, J. L. R. Construction and in Vivo Assembly of a Catalytically Proficient and Hyperthermostable de Novo Enzyme. *Nat. Commun.* **2017**, *8* (1), 358.
- (4) Omosun, T. O.; Hsieh, M.-C.; Childers, W. S.; Das, D.; Mehta, A. K.; Anthony, N. R.; Pan, T.; Grover, M. A.; Berland, K. M.; Lynn, D. G. Catalytic Diversity in Self-Propagating Peptide Assemblies. *Nat. Chem.* **2017**, *9* (8), 805–809.
- (5) 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.
- (6) 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.
- (7) Wu, Z. P.; Hilvert, D. Conversion of a Protease into an Acyl Transferase: Selenolsubtilisin. *J. Am. Chem. Soc.* **1989**, *111* (12), 4513–4514.



## **Chapter 7. Experimental part**

## 7.1. Materials, methods and general procedures

### **Reagents:**

All solvents, chemicals, and reagents were purchased from commercial sources and used without further purification. Coupling reagents (1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU), 3-[bis(dimethylamino)methylumyl]-3H-benzotriazol-1-oxide hexafluorophosphate (HBTU), *N,N'*-diisopropylcarbodiimide (DIC) and ethyl cyano(hydroxyimino)acetate (Oxyma Pure)), Fmoc- $\alpha$ -L-amino acids, resins for SPPS (solid-phase peptide synthesis) and DMF were purchased from Iris Biotech (Marktredwitz, Germany); Boc- $\alpha$ -L-amino acids were from Aapptec (Louisville, USA); diisopropylethylamine (DIEA), triisopropylsilane (TIPS), bromoacetic acid, 4-mercaptophenylacetic acid (MPAA), sodium 2-mercaptoethanesulfonate (MesNa), tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl), acetic anhydride and 4-methylpiperidine were from Sigma-Aldrich; trifluoroacetic acid (TFA, BioGrade) was from Halocarbon (Peachtree Corners, USA); hydrogen fluoride was from GHC Gerling Holz & Co Handels GmbH (Hanau, Germany). All other chemicals were purchased from Sigma-Aldrich or Roth. Reactions involving compounds that are sensitive to oxidation or hydrolysis were performed under an argon atmosphere (e.g. thioesterification and ligation).

### **LC-MS:**

Peptide masses were measured on a LC-MS instrument equipped with a Thermo Scientific Accela UHPLC (Hypers II GOLD column, 50x2.1 mm<sup>2</sup>, 1.9  $\mu$ m) integrated with a Thermo Scientific LCQ Fleet ion-trap. Deconvolution of data was performed in MagTran 1.03 (Amgen, Thousand Oaks, USA).

### **Analytical High Performance Liquid Chromatography (HPLC):**

Analytical reversed-phase HPLC was performed on a Thermo Scientific Dionex Ultimate 3000 UHPLC instrument. A Phenomenex Kinetex EVO-C18 column (50x2.1 mm<sup>2</sup>, 100 Å, 2.6  $\mu$ m) was used at 1 mL/min with a gradient of water with TFA (0.1 %, v/v) and acetonitrile with TFA (0.08 %, v/v). Analytical reversed-phase HPLC was also performed on a Nexera XR UHPLC instrument (Shimadzu, Kyoto, Japan) equipped with LC-20AD liquid chromatograph modules and an SPD-M20A Prominence diode array detector. A Phenomenex Kinetex XB-C18 column (50 x 2.1 mm, 100 Å, 2.6  $\mu$ m) was used at 1 mL/min with a gradient of water with TFA (0.1 %, v/v) and acetonitrile with TFA (0.08 %, v/v).



**Preparative HPLC:**

Purifications of peptides were performed on a preparative Shimadzu HPLC instrument equipped with two LC-20AP pumps and an SPD-20A Prominence UV/vis detector connected to an FRC-10A fraction collector. Phenomenex Kinetex XB-C18 column (250×21.2 mm<sup>2</sup>, 100 Å, 5 µm) was used at 10 mL/min with a gradient of water with TFA (0.1 %, v/v) and acetonitrile with TFA (0.08 %, v/v).

**Size-exclusion chromatography:**

SEC was performed on a Shimadzu Nexera XR UHPLC instrument. A Phenomenex Yarra 3µm SEC 2000 column was used (300×4.6 mm, 2.8 µm) with an isocratic mobile phase composed of NaH<sub>2</sub>PO<sub>4</sub> (10 mM), Na<sub>2</sub>HPO<sub>4</sub> (40 mM), NaCl (100 mM) and NaN<sub>3</sub> (0.05%, w/v), pH 7.5, at a flow of 0.5 mL/min, at 27 °C, detection at 250-280 nm. A gel filtration markers kit for protein molecular weights (6,500-66,000 Da) was used for calibration. The kit was composed of Blue Dextran (for the determination of the void volume), aprotinin from bovine lung (6.5 kDa), cytochrome c from equine heart (12.4 kDa), carbonic anhydrase from bovine erythrocytes (29 kDa) and albumin from bovine serum (66 kDa) and was purchased from Sigma-Aldrich.

**NMR spectroscopy:**

Liquid <sup>1</sup>H NMR analyses for modified amino acids and linkers were performed at room temperature on a 400 MHz Avance spectrometer (Bruker). The data were processed with TopSpin 2.0 (Bruker) and analyzed with MestReNova 10.0 (Mestrelab, Santiago de Compostela, Spain).

**Analytical Ultracentrifugation:**

Sedimentation velocity experiments were performed on a Beckman Coulter ProteomeLab XL-I ultracentrifuge at 20°C and 55,000 RPM using an An55-Ti rotor. 'Positive' strand and **Cys-N2** peptides were dialyzed against a buffer containing 50 mM sodium phosphate pH 7.5, 100 mM NaCl and 1 mM TCEP. Concentrations were determined by measuring absorbance with Thermo Scientific NanoDrop One instrument and volumes were adjusted to have three solutions of 'positive' strand, **Cys-N2** peptide and **Het-N2** heterodimer (a 1:1 mixture of 'positive' strand and **Cys-N2**) at 200 µM. 400 µL of each sample at 200 µM were loaded in 1.2 cm double sector cells with quartz windows. Absorbance scans at 240 nm were taken every 3 minutes for 17 h. Sedimentation data were analyzed using SEDFIT software<sup>1</sup> (version 16.1) and the continuous sedimentation coefficient distribution model c(s). Buffer density, buffer viscosity and peptides partial specific volumes were calculated using

SEDNTERP software.<sup>2</sup> GUSSE<sup>3</sup> was used to integrate the sedimentation peaks and to produce the graphs.

### **Circular Dichroism (CD):**

CD spectra were recorded in far-UV region (185-280 nm) using 20-25  $\mu\text{M}$  protein **Het-N2** concentrations in 10 mM sodium phosphate buffer, pH 6.9 at 25 °C. For urea denaturation experiments, CD spectra were recorded in far-UV region (210-280 nm) in buffer (50 mM sodium phosphate, 5 mM TCEP·HCl, pH 7.5) with increasing concentration of urea at c 10.5  $\mu\text{M}$  of heterodimer **Het-N2** and homodimers formed from **Cys-N2** or 'positive' strands. The Jasco J-810-1505 instrument was set as follows: Digital Integration Time (D.I.T.) 1 sec, band width 1 nm, data pitch 0.1 nm, scanning speed 100 nm/min. Every CD curve was obtained by averaging of 5 scans and subtracting the background signal of buffer. Melting curves in temperature range 5-90 °C were recorded at 222 nm using 50  $\mu\text{M}$  protein concentrations in 10 mM sodium phosphate buffer with 500  $\mu\text{M}$  TCEP, pH 6.9. Instrument parameters were set as follows: 1 °C/min, measurement every 0.2 °C, D.I.T. 2 sec, band width 1 nm.

For **Hom-N2** and **Het-Ncap** proteins, CD spectra were recorded in far-UV region (190-350 nm) using 20  $\mu\text{M}$  protein concentrations in 50 mM sodium phosphate, 0.1 M NaCl, pH 7 at room temperature (Jasco J-1500 instrument).

The obtained ellipticities were transformed to mean residue molar ellipticities using the following equation:

$$[\theta] = \frac{\theta \times 10^{-6}}{d \times c \times (n - 1)}$$

where:

$[\theta]$ : mean residue molar ellipticity [ $\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$ ]

$\theta$ : ellipticity [mdeg]

$d$ : pathlength of cuvette [mm] (here,  $d = 1$  mm)

$c$ : protein concentration [ $\mu\text{M}$ ]

$n$ : number of amino acids in the protein sequence

### **Native mass-spectrometry:**

Separated stock solutions of 'positive' and **Cys-N2** strands at 100  $\mu\text{M}$  were prepared in buffer of 100 mM  $\text{AcONH}_4$ , pH 7.5. The two strands were mixed in a 1:1 ratio and diluted with buffer to have the heterodimer **Het-N2** at a final concentration of 8.5  $\mu\text{M}$ . The sample was incubated for 5 minutes before analysis. Full MS spectra were acquired on a Thermo Scientific Exactive Plus EMR Orbitrap mass-spectrometer using direct infusion in positive ion mode at a 3.5 kV spray voltage and controlled with Exactive series (ver. 2.9 sp4) software.

Resolution of full MS and HCD scans was 70,000. Normalized collision energy for HCD spectra was 35 eV, in source CID 100 eV and data were acquired in profile mode and processed using Xcalibur 4.2 sp1.

### ***Molecular modelling:***

Molecular modeling and analyses were performed using MOE software (Molecular Operating Environment, Chemical Computing Group, Montréal, Canada). The Domain-Swaped Dimer (DSD) protein from PDB : 1G6U or crystal structures of our designed proteins were used as starting scaffolds. Energies of the resulting models were minimized by using the minimization force field AMBER10:EHT, with R-Field as a solvation mode. Molecular graphics and images were generated with MOE, UCSF Chimera (UCSF Resource for Biocomputing, Visualization, and Informatics) or PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC).

#### ***7.1.1. General procedures for solid-phase peptide synthesis***

##### ***General procedure I for the preparation of the 2-chlorotrityl hydrazine (2-CT-NHNH<sub>2</sub>) resin<sup>4</sup>***

The 2-CT-OH resin (1 g, 1.46 mmol/g active groups, mesh size 200-400) was washed several times with dry DCM and placed in a round bottom flask that was previously flushed several times with argon. Dry dichloromethane (10 mL) was added. The mixture was gently stirred allowing the resin to swell. Thionyl chloride (1.3 eq, 2.02 mmol, 147  $\mu$ L) was slowly added at 0 °C. The suspension was stirred under argon for 2 h allowing it to slowly warm up to room temperature. The solvent was removed and the resin was washed with DCM and then with DMF. The resulting 2-CT-Cl resin was swollen for 20 min in DMF (6 mL). A mixture of 64 wt. % hydrazine in water (3 eq, 4.7 mmol, 233  $\mu$ L) and DIEA (2.4 eq, 3.8 mmol, 667  $\mu$ L) in DMF (2 mL) was added to the resin at 0 °C. The suspension was stirred at room temperature for 90 min. The reaction was quenched by the addition of methanol (100  $\mu$ L). The resin was washed with DMF, water, DMF, methanol and diethyl ether and dried under reduced pressure. Generally, final mass of the resin is slightly above 1 g with a loading assumed to be 0.4 mmol/g based on the previously synthesized peptides in the laboratory.

##### ***General procedure II for manual Fmoc/tBu solid-phase peptide synthesis***

Peptides were prepared manually on usually a 0.1 mmol scale by Fmoc/tBu-SPPS following standard procedure.<sup>5</sup> Side chain protection of amino acids was as follows: Arg(Pbf), Asn(Trt), Asp(OtBu), Cys(Acm), Cys(Trt), Gln(Trt), Glu(OtBu), His(Trt),

Lys(Boc), Ser(*t*Bu), Thr(*t*Bu), Trp(Boc), Tyr(*t*Bu). The resin (Rink amide, 0.74 mmol/g loading, 100-200 mesh or 2-CT-NHNH<sub>2</sub> resin, 0.4 mmol/g loading, 200-400 mesh) was weighted to enable a synthesis on a 0.1 mmol scale. Resin was washed several times with DMF, where in the final wash the beads were entirely covered with the solvent. Then, resins were swollen for 20 min and stirred every 5 min. Finally, DMF was removed by vacuum filtration.

Coupling: The Fmoc-amino acid (4 eq, 0.4 mmol) was dissolved in 1 mL of a solution of HBTU dissolved at c 0.38 M in DMF (3.8 eq). For preactivation the 110  $\mu$ L of DIEA (6 eq) were added, then the reaction mixture was vortexed and let to react for 2 min. The resulting solution was poured onto the resin beads. The coupling reaction lasted 30 min with stirring every 5 min. The beads were rinsed and batch washed with DMF four times. The completion of the coupling was monitored by Kaiser test.<sup>6</sup> To do so, a few beads were transferred in a test tube and rinsed with EtOH. Then 50  $\mu$ L of a solution of ninhydrin (0.5 g in 10 mL of EtOH) and 50  $\mu$ L of KCN solution (0.4 mL of 1 mM KCN(aq) in 20 mL of pyridine) were added. The mixture was placed into the heating block at 115 °C for 3 min. 500  $\mu$ L of EtOH and 500  $\mu$ L of H<sub>2</sub>O were added to the tube and the color of the beads and the solution was inspected. A blue color of the beads and of the solution indicates an incomplete coupling. In the case of incomplete coupling a double coupling was performed.

Fmoc deprotection: The resin was rinsed twice with a solution of 20 % of 4-methylpiperidine in DMF. The beads were covered with the solution and let to react for 20 min with stirring every 5 min. The piperidine was removed and the beads were washed four times with DMF. The coupling and deprotection were repeated for each amino acid in the sequence of a peptide following from C-terminus to N-terminus.

Acetylation (0.1 mmol scale): After final deprotection, some peptides were acetylated. To do so, the N-termini were acetylated with a mixture of 700  $\mu$ L of DMF, 100  $\mu$ L of DIEA and 100  $\mu$ L of acetic anhydride. The reaction lasted for 6 min with stirring. The resin was washed with DMF and a Kaiser test was performed to validate the completeness of acetylation.

Cleavage: The resin was washed six times with DCM and dried by air flow using a vacuum filtration setup for 20 min. The resin was weighted: for 100 mg of resin, 1 mL of cleavage cocktail (95 % TFA, 2.5 % TIPS, 2.5 % H<sub>2</sub>O) was added to the beads. The resulting mixture was stirred for 2 hours at room temperature.

Work up: The cleavage mixture was transferred into several 50 mL centrifuge tubes and chilled diethyl ether (-20 °C) was added to reach a total volume of 40 mL. Afterwards, crude peptide product precipitated and was centrifuged at 4500 rpm for 3 min. The supernatant was discarded and the peptide was dissolved in 15 mL of a 1:1 mixture MeCN/H<sub>2</sub>O, 0.1 % TFA. The solution was filtered to remove the resin beads. An aliquot was taken for HPLC and

LC/MS analysis. The peptide solution was frozen in liquid nitrogen and then lyophilized. The peptide was then purified on a preparative HPLC or used as crude in the reactions.

### **General procedure III for automated Fmoc-tBu solid-phase peptide synthesis**

Machine-assisted peptide synthesis was performed using a Syro I (Biotage) synthesizer. Typically, the syntheses were performed on a 0.1 mmol scale with side-chain protection of the amino acids as follows: Arg(Pbf), Asn(Trt), Asp(OtBu), Cys(Acm), Cys(Trt), Gln(Trt), Glu(OtBu), His(Trt), Lys(Boc), Ser(tBu), Thr(tBu), Trp(Boc), Tyr(tBu). The resin (Rink amide, 0.74 mmol/g loading, 100-200 mesh or 2-CT-NH<sub>2</sub> resin, 0.4 mmol/g loading, 200-400 mesh) was weighted to enable a synthesis on a 0.1 mmol scale. Machine-assisted peptide syntheses on a Syro I synthesizer were programmed as follow: the resin was washed twice with DMF and then was swollen for 20 min, while vortexing every 3 min. Coupling reaction was performed with 5 eq of Fmoc-amino acid (c 0.5 M in DMF), 5 eq of HATU (c 0.5 M in DMF) and 10 eq of DIEA (c 2.5 M in NMP) for 30 min, while vortexing every 3 min. Fmoc-deprotection was performed in 2 stages: *i*) 40% 4-methylpiperidine in DMF for 3 min, and *ii*) 20% 4-methylpiperidine in DMF for 12 min, while vortexing every 3 min.

Acetylation (0.1 mmol scale): After final deprotection, some peptides were acetylated. To do so, the N-termini were acetylated with a mixture of 700  $\mu$ L of DMF, 100  $\mu$ L of DIEA and 100  $\mu$ L of acetic anhydride. The reaction lasted for 6 min with stirring. The resin was washed with DMF and a Kaiser test was performed to validate the completeness of acetylation.

Cleavage: The resin was washed six times with DCM and dried by air flow using a vacuum filtration setup for 20 min. The resin was weighted: for 100 mg of resin, 1 mL of cleavage cocktail (95 % TFA, 2.5 % TIPS, 2.5 % H<sub>2</sub>O) was added to the beads. The resulting mixture was stirred for 2 hours at room temperature.

Work up: The cleavage mixture was transferred into several 50 mL centrifuge tubes and chilled diethyl ether (-20 °C) was added to reach a total volume of 40 mL. Afterwards, crude peptide product precipitated and was centrifuged at 4500 rpm for 3 min. The supernatant was discarded and the peptide was dissolved in 15 mL of a 1:1 mixture MeCN/H<sub>2</sub>O, 0.1 % TFA. The solution was filtered to remove the resin beads. An aliquot was taken for HPLC and LC/MS analysis. The peptide solution was frozen in liquid nitrogen and then lyophilized. The peptide was then purified on a preparative HPLC or used as crude in the reactions.

### **General procedure IV for automated microwave Fmoc-tBu solid-phase peptide synthesis**

Automated microwave peptide synthesis was performed on a Liberty Blue (CEM) synthesizer. Typically, the syntheses were performed on a 0.1 mmol scale with side-chain protection of the amino acids as follows: Arg(Pbf), Asn(Trt), Asp(OtBu), Cys(Trt), Gln(Trt), Glu(OtBu), Ser(tBu), Tyr(tBu). The resin (Rink amide, 0.74 mmol/g loading, 100-200 mesh

or 2-CT-NH<sub>2</sub> resin, 0.4 mmol/g loading, 200-400 mesh) was weighted to enable a synthesis on a 0.1 mmol scale. The coupling reactions were performed by adding amino acids dissolved in DMF (2.5 mL, 0.2 M), the coupling reagent DIC in DMF (1.0 mL, 0.5 M) and Oxyma Pure in DMF (0.5 mL, 1 M) to the respective resin. Standard couplings were performed at 90 °C for 2 min (170 W for 15 s, 30 W for 110 s). Standard deprotections were performed using 20% (v/v) 4-methylpiperidine in DMF at 90 °C for 1 min (155 W 15 s, 32 W 50 s).

**Acetylation (0.1 mmol scale):** After final deprotection, some peptides were acetylated. To do so, the N-termini were acetylated with a mixture of 700 µL of DMF, 100 µL of DIEA and 100 µL of acetic anhydride. The reaction lasted for 6 min with stirring. The resin was washed with DMF and a Kaiser test was performed to validate the completeness of acetylation.

**Cleavage:** The resin was washed six times with DCM and dried by air flow using a vacuum filtration setup for 20 min. The resin was weighted: for 100 mg of resin, 1 mL of cleavage cocktail (95 % TFA, 2.5 % TIPS, 2.5 % H<sub>2</sub>O) was added to the beads. The resulting mixture was stirred for 2 hours at room temperature.

**Work up:** The cleavage mixture was transferred into several 50 mL centrifuge tubes and chilled diethyl ether (-20 °C) was added to reach a total volume of 40 mL. Afterwards, crude peptide product precipitated and was centrifuged at 4500 rpm for 3 min. The supernatant was discarded and the peptide was dissolved in 15 mL of a 1:1 mixture MeCN/H<sub>2</sub>O, 0.1 % TFA. The solution was filtered to remove the resin beads. An aliquot was taken for HPLC and LC/MS analysis. The peptide solution was frozen in liquid nitrogen and then lyophilized. The peptide was then purified on a preparative HPLC or used as crude in the reactions.

### **General procedure V for manual Boc/benzyl solid-phase peptide synthesis**

Each strand of the heterodimers and peptide fragments was prepared manually on 0.2, 0.3 or 0.4 mmol scale by 'in situ neutralization' Boc/Bzl solid-phase peptide synthesis (SPPS).<sup>7</sup> Side-chain protection of amino acids was as follows: Arg(Tos), Asn(Xan), Cys(4-MeBzl), Gln(Xan), Glu(OcHex), His(Bom), Ser(Bzl). The MPA(Trt) building block was also used to form thioester on resin. For 0.2 mmol scale synthesis, 167 mg of MBHA resin (1.2 mmol/g loading, 100-200 mesh) was used. Resin was washed several times with DMF followed by a solution of 5 % (v/v) DIEA in DMF, where in the final wash the beads were entirely covered with the latter. Then, resin was swollen for 20 min and stirred every 5 min. Finally, DMF with 5 % DIEA was removed by vacuum filtration.

**Coupling (0.2 mmol scale):** The Boc-amino acid (4 eq, 0.8 mmol) was dissolved in 2 mL of a solution of HBTU or HATU (c 0.38 M in DMF (3.8 eq)). For preactivation the 350 µL of DIEA (10 eq) were added, then the reaction mixture was vortexed and let to react for 2 min. The resulting solution was poured onto the resin beads. The coupling reaction lasted for 10 to 15

min with stirring every 5 min. The beads were rinsed and batch washed with DMF four times. The completion of the coupling was monitored by Kaiser test.<sup>6</sup> For Kaiser test protocol, see **General procedure II**. A blue color of the beads and of the solution indicates an incomplete coupling. In the case of incomplete coupling, a double coupling was performed.

Boc-deprotection: The resin was rinsed once with pure TFA. Then, the beads were covered with TFA and let to react for 2 min with continuous stirring. The TFA was removed and the beads were washed four times with DMF.

The coupling and deprotection were repeated for each amino acid in the sequence of a peptide following from C-terminus to N-terminus.

Trt-deprotection: In peptide- $\alpha$ thioester synthesis, after MPA(Trt) coupling as usual, Trt group was removed to then form a thioester linkage on the resin with the following amino acid. The resin was rinsed once with a mixture of 95 % TFA, 2.5 % TIPS and 2.5 % H<sub>2</sub>O. Then, the beads were covered with the same mixture and let to react for 2 min with continuous stirring. The TFA was removed and the beads were washed four times with DMF.

Acetylation (0.2 mmol scale): After final deprotection, some peptides were acetylated. To do so, the N-termini were acetylated with a mixture of 1.4 mL of DMF, 200  $\mu$ L of DIEA and 200  $\mu$ L of acetic anhydride. The reaction lasted for 6 min with stirring. The resin was washed with DMF and a Kaiser test was performed to validate the completeness of acetylation.

Cleavage: The resin was washed six times with DCM and dried by air flow using a vacuum filtration setup for 20 min. The cleavage mixture was composed of HF and a scavenger (either *p*-cresol or anisole) in approximately 9:1 ratio. HF was condensed to the reactor containing the resin and the scavenger using dry ice + ethanol bath. The resulting mixture was stirred for 1 hour at 0 °C.

Work up: After evaporation of the HF under reduced pressure, crude products were precipitated with chilled diethyl ether (-20 °C). Product precipitates were filtered on Celite® 545 and dissolved with a 1:1 mixture MeCN/H<sub>2</sub>O, 0.1 % TFA. An aliquot was taken for HPLC and LC-MS analysis. The peptide solution was frozen in liquid nitrogen and then lyophilized. The peptide was purified on a preparative HPLC.

### 7.1.2. General procedure for peptide thioesterification

#### **General procedure VI for peptide thioesterification starting from peptide- $\alpha$ hydrazide<sup>4</sup>**

Generally, due to acceptable purity, the crude peptide- $\alpha$ hydrazides were directly used in the following thioesterifications without any purification step.

Activation of peptide- $\alpha$ hydrazides: The peptide hydrazides were dissolved in acidic aqueous buffer (200 mM Na<sub>2</sub>HPO<sub>4</sub>, 6 M Gn·HCl, pH 3) at a 2.7 mM concentration. The temperature of

the solution was adjusted to  $-15\text{ }^{\circ}\text{C}$ , followed by a slow addition of  $\text{NaNO}_2$  (5 eq, dissolved in water at a c of 0.2 M). The solution was stirred at  $-15\text{ }^{\circ}\text{C}$  for 15 min. The formation of peptide- $\alpha$ azide could be followed by HPLC and LC-MS.

**Thioesterification:** To the resulting solution, a 0.2 M solution of sodium mercaptoethanesulfonate (MesNa) (50 eq) in neutral aqueous buffer (200 mM  $\text{Na}_2\text{HPO}_4$ , 6 M Gn-HCl, pH 7) was directly added in one pot fashion. The solution was stirred at  $-15\text{ }^{\circ}\text{C}$  for 15 min. The completion of the thioesterification was confirmed by HPLC and by LC-MS. The reaction mixture was warmed to room temperature and peptide- $\alpha$ thioester was directly isolated by preparative HPLC.

### 7.1.3. General procedure for concentration determination

As no aromatic residue (except a phenylalanine) that highly absorbs at 280 nm was present in our catalytic protein, concentration of protein samples was determined via calibration of HPLC setup when necessary (e.g. catalytic assay, crystallography, CD and so on.).

On a sample of freshly prepared and purified peptide, elemental analysis was performed to measure the content of C, H and N atoms in the powder. Measured percentage of nitrogen was compared to the expected percentage of nitrogen of the protein alone. The difference of the two values indicates the percentage in mass of protein present in the lyophilized powder. The same protein sample was then solubilized in phosphate buffer (10 mM sodium phosphate, 5 mM TCEP and pH 7.0) and serial dilution was performed to obtain a range of concentration between 6.25  $\mu\text{M}$  and 200  $\mu\text{M}$ . 10  $\mu\text{L}$  of each concentration was injected three times on analytical HPLC. Mean area of the three injections was determined for each concentration and calibration curve was deduced knowing the exact percentage of proteins in the sample. Thus, if the protein gets hydrated with time, we can still estimate concentration of proteins in different solutions using this calibration curve.

### 7.1.4. General procedures for catalytic assays

#### **General procedure VII for catalytic assays of 1<sup>st</sup> generation with HPLC monitoring**

Standard conditions for native chemical reaction: 0.5 mM of catalytic protein or controls, 0.7 mM of [N-ACTR]- $\alpha$ Mes, 0.5 mM Cys-[C-ACTR] dissolved in 50 mM sodium phosphate buffer, 100 mM NaCl, 5 mM of TCEP, pH 7.0. Total volume: 130  $\mu\text{L}$ .



Standard conditions for acyl transfer reaction: 0.5 mM of catalytic protein or controls, 0.5 mM of acyl donor (peptide- $\alpha$ thioester substrate), with or without 250 mM of Tris dissolved in 50 mM sodium phosphate buffer, 100 mM NaCl, 5 mM of TCEP, pH 7.0. Total volume: 200  $\mu$ L.

Procedure: Buffers were freshly prepared notably because of decomposition of TCEP in such a buffer. Peptide- $\alpha$ thioester was first solubilized in buffer (final volume of the reaction is used) at the final concentration (*i.e.* at either 0.5 or 0.7 mM). The resulting solution was then poured into the acyl acceptor if there is one (*i.e.* Cys-[ACTR], Tris or nothing if only hydrolysis is evaluated). The pH of the mixture was controlled and adjusted to 6.9-7.0. The reaction mixture was poured into the catalytic protein. If the protein is heterodimeric, the solution was first poured into 'positive' strand then into the 'negative' strand containing the catalytic cysteine. For conditions with glutathione, a stock solution at 33.8 mM of glutathione in buffer at pH 7.0 was prepared and then 3  $\mu$ L were added to previous reaction mixture to obtain a final concentration of glutathione of 0.5 mM. The pH of the resulting reaction mixture was adjusted to 6.9-7.0 giving the  $t = 0$  h. In general, the reactions were monitored for one week. Aliquots of 2  $\mu$ L of reaction mixture quenched in 50  $\mu$ L of H<sub>2</sub>O + 0.1 % TFA were taken to follow the reaction by HPLC and LC/MS at different time points. Branched adduct evolution was followed by integration of the areas of the corresponding HPLC peak normalized by the number of residues (absorbance at 220 nm of a protein depends on the number of amide bonds). Half-lives were estimated by inspecting LC/MS spectra.

### ***General procedure VIII for catalytic assays of 2<sup>nd</sup> generation with LC/MS monitoring***

Standard conditions: 100  $\mu$ M of catalytic protein, 200  $\mu$ M of acyl donor (peptide- $\alpha$ thioester substrate) and 200 mM of acyl acceptor dissolved in 50 mM sodium phosphate buffer, pH 7.5, 100 mM NaCl and 2 mM of TCEP. Total volume: 50  $\mu$ L.

Preparation of stock solutions: For a given heterodimer, the two monomers were first dissolved separately in phosphate buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 100 mM NaCl, pH 7.0) at a stock concentration of 750  $\mu$ M. 14 mM of TCEP were added to the solution of Cys-containing monomer to avoid oxidation during storage. pH of solutions was adjusted to 7.1. In parallel, a stock solution of acyl acceptor was prepared at concentration of 290 mM with 2.9 mM TCEP and pH was adjusted to 7.5-7.6. The stock solutions of peptide- $\alpha$ thioester substrates at 5 mM in anhydrous CH<sub>3</sub>CN were also prepared.

Preparation of reaction mixture: To obtain final expected concentrations in 50  $\mu$ L of reaction, 6.7  $\mu$ L of each stock solution of 'positive' and 'negative' strands were first mixed together. Then, 35  $\mu$ L of stock solution of acyl acceptor with TCEP were added followed by 2  $\mu$ L of peptide- $\alpha$ thioester solution (corresponding to time = 0 h). pH was controlled and

adjusted if necessary to 7.5. Resulting mixture was let to react for 1 week at room temperature (20-22 °C).

**Monitoring:** Aliquots were generally taken off after 1 h, 3 h, 7 h, 24 h, 48 h, 72 h, 96 h and 1 week of reaction. 2 µL of reaction mixture were diluted with 25 µL of water, 0.1 % TFA. 10 µL were then injected for LC-MS analysis. Kinetics were monitored by extracting percentage of detected ions between peptide- $\alpha$ thioester (starting material), peptide- $\alpha$ acyl acceptor and peptide- $\alpha$ OH (products).

Observed rate constants were deduced from half-life of peptide- $\alpha$ thioester (corresponding to 50 % of thioester in reaction mixture) using the rate law equation for second order reaction:

$$\tau_{1/2} = \frac{1}{k_{obs} \times [thioester]_0}$$

where:

$\tau_{1/2}$ : half-life of thioester (when reaching 50 % in reaction mixture),

$k_{obs}$ : observed rate constant,

$[thioester]_0$ : initial concentration of thioester.

### **General procedure IX for hydrolytic activity with para-nitrophenyl acetate (pNPA) as substrate**

Kinetic assays were performed in a 10 mm path-length quartz cuvette using a V-670 Jasco spectrophotometer at 348 nm (which corresponds to the pH-independent isobestic point for pNP,  $\epsilon_{pNP} = 5400 \text{ cm}^{-1} \cdot \text{M}^{-1}$ ) at 22 °C. Stock solutions of each strand of **Het-N2** were prepared separately at 105 µM in phosphate buffer (50 mM sodium phosphate, 100 mM NaCl, 500 µM TCEP, pH 7.0). ‘Positive’ and **Cys-N2** strands were mixed in a 1:1 ratio to obtain a final concentration of 50 µM in 315 µL of total volume (300 µL of peptide solution and 15 µL of pNPA solution in anhydrous CH<sub>3</sub>CN). 15 µL of a freshly prepared solution of pNPA (to obtain final concentration between 400 and 1250 µM) were added to the **Het-N2** solution. Measurements were continued by recording absorption values every 30 seconds for a total time of 3-4 hours. Data were background corrected (300 µL of buffer with TCEP + 15 µL of CH<sub>3</sub>CN).

All kinetic curves were fitted with the following equation containing an exponential (also called “burst” phase) and a linear part:  $[pNP] = At + B(1 - e^{-bt})$ .<sup>8</sup> By calculating A at different  $[S]_0$ , the double-reciprocal plot was then deduced and used to determine  $k_{cat}$  ( $k_{cat}[E]_0 = 1/y\text{-intercept}$ ) and  $K_m$  ( $K_m = \text{slope}/y\text{-intercept}$ ).

For catalytic assays measured in presence of **Hom-N2**, the cuvettes used allowed larger volume. Thus, the protein was dissolved in 1000 µL of phosphate buffer to have a final

concentration of 50  $\mu\text{M}$  in 1050  $\mu\text{L}$  (1000  $\mu\text{L}$  of peptide solution and 50  $\mu\text{L}$  of *p*NPA solution in anhydrous  $\text{CH}_3\text{CN}$ ). 50  $\mu\text{L}$  of a freshly prepared solution of *p*NPA (to obtain final concentration between 300 and 1000  $\mu\text{M}$ ) were added to the **Hom-N2** solution. Then, monitoring was performed as mentioned above.

## 7.2. Chapter 2

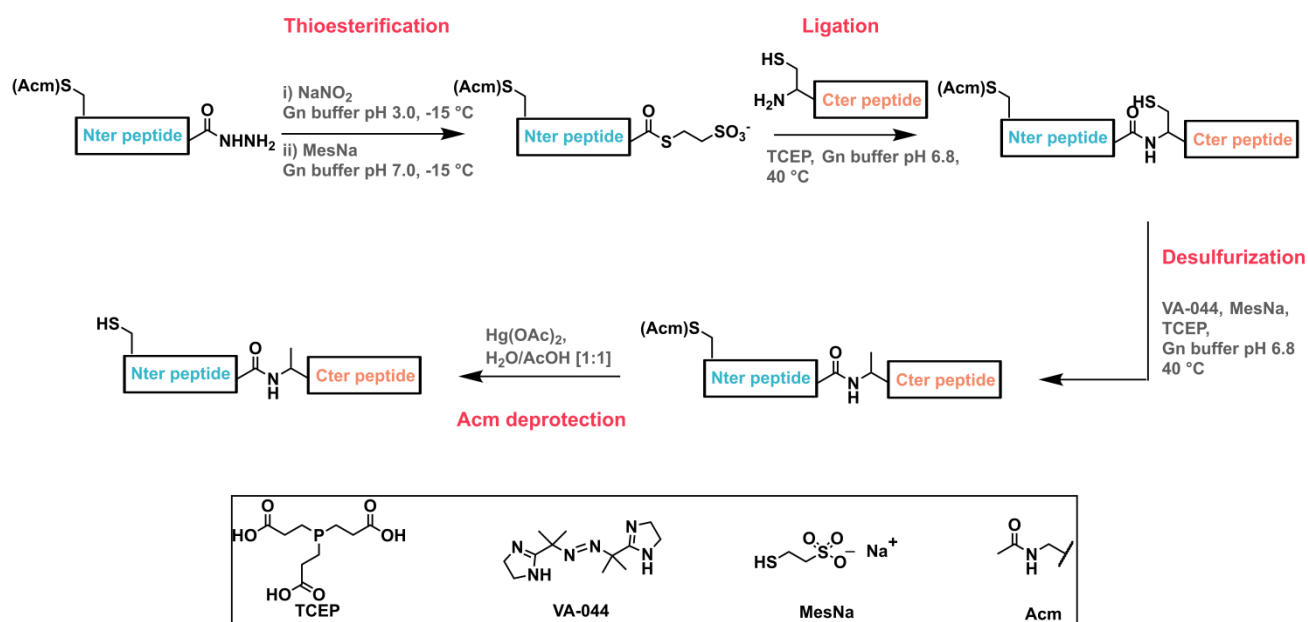
### **Synthesis of 12-residue trunc-Cys-N2 control protein**

Sequence: Ac-GLCALRSELQAL-amide

**Trunc-Cys-N2** was synthesized on a 0.1 mmol scale synthesis following **General procedure IV** on a Rink amide resin (0.74 mmol/g loading, 100-200 mesh); isolated yield was 41.8 mg (32 %) based on peptide synthesis scale. LC/MS (ESI):  $[\text{M} + 1 \text{H}^+] = 1315.0 \text{ m/z}$ ;  $[\text{M} + 2 \text{H}^+] = 657.9 \text{ m/z}$ ;  $\text{C}_{56}\text{H}_{99}\text{N}_{17}\text{O}_{17}\text{S}$ ; Average isotope calculated: 1314.6 Da  $[\text{M}]$ ; observed: 1313.9 Da.

**Synthesis of Hom-N2 by the 2-segment ligation method**

Synthetic route:

Sequence: Ac-GLCALRSELQALRREGFSPPEELAALESELQALERRLAALRSRLQALRG-<sup>α</sup>amide

Nter peptide-<sup>α</sup>Mes was synthesized on a 0.1 mmol scale according to the **General procedure III** as a hydrazide surrogate using a 2-CT-NHNH<sub>2</sub> resin (0.4 mmol/g loading, 200-400 mesh) followed by thioesterification according to the **General procedure VI**; isolated yield of thioesterification reaction was 16.2 mg (22 %). LC/MS (ESI): [M + 2 H<sup>+</sup>] = 1392.6 m/z; [M + 3 H<sup>+</sup>] = 928.6 m/z; C<sub>116</sub>H<sub>192</sub>N<sub>34</sub>O<sub>39</sub>S<sub>3</sub>; Average isotope calculated: 2783.2 Da [M]; observed: 2783.2 Da.

Cter peptide was synthesized on a 0.3 mmol scale according to the **General procedure III** on a Rink amide resin (0.74 mmol/g loading, 100-200 mesh); isolated yield was 50.1 mg (6 %) based on peptide synthesis scale. LC/MS (ESI): [M + 2 H<sup>+</sup>] = 1426.9 m/z; [M + 3 H<sup>+</sup>] = 951.7 m/z; [M + 4 H<sup>+</sup>] = 714.1 m/z; [M + 5 H<sup>+</sup>] = 571.6 m/z; [M + 6 H<sup>+</sup>] = 476.8 m/z; C<sub>120</sub>H<sub>215</sub>N<sub>43</sub>O<sub>35</sub>S; Average isotope calculated: 2852.4 Da [M]; observed: 2852.5 Da.

**Ligation of peptides.**

The C-terminal peptide (6.52 · 10<sup>-6</sup> mol, 1.7 eq, at c of 8.5 mM) was dissolved in buffer (200 mM Na<sub>2</sub>HPO<sub>4</sub>, 6 M Gn-HCl, 5 mM TCEP, pH 7). The resulting solution was added to the

N-terminal fragment ( $3.83 \cdot 10^{-6}$  mol, 1 eq, at c of 5 mM) and the pH was adjusted to 7.0 with a solution of 1 M NaOH. The resulting solution was degassed with argon and stirred under argon for 5 hours. The reaction was followed by HPLC and LC/MS. TCEP (dissolved in the same buffer, pH adjusted to 6.8-6.9) was added to the reaction mixture of the ligating peptides to a total concentration of TCEP of 200 mM. Subsequently, the radical initiator VA-044 ( $7.66 \cdot 10^{-5}$  mol, 20 eq) and sodium-mercaptoethanesulfonate ( $1.53 \cdot 10^{-4}$  mol, 40 eq) were added as solids. The resulting mixture was stirred for 1-2h at 40 °C under argon. The final product was subjected to HPLC purification.

Ligated product containing Cys(Acm) residue: isolated yield of the ligation was 9.2 mg (44 %). LC/MS (ESI):  $[M + 4 H^+] = 1365.9$  m/z;  $[M + 5 H^+] = 1092.9$  m/z;  $[M + 6 H^+] = 911.0$  m/z;  $[M + 7 H^+] = 781.0$  m/z;  $[M + 8 H^+] = 683.5$  m/z  $C_{234}H_{401}N_{77}O_{71}S$ ; Average isotope calculated: 5461.3 Da [M]; observed: 5460.1 Da.

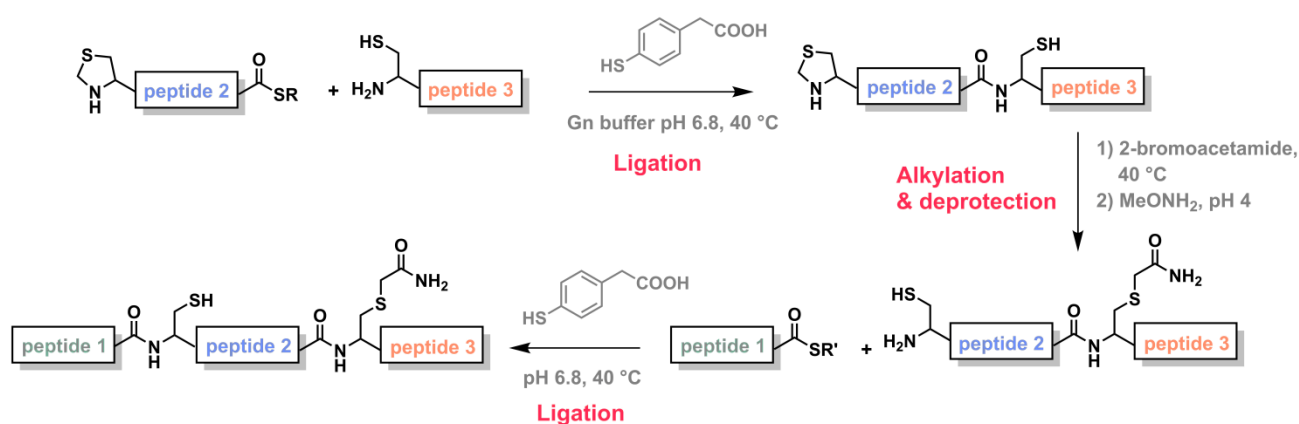
*Acm removal.*<sup>9,10</sup>

The resulting ligated product was dissolved in a 1 : 1 mixture of H<sub>2</sub>O/AcOH under argon to a total concentration of 0.1 mM. Then, Hg(OAc)<sub>2</sub> (7 eq) was added to the resulting mixture. After 2 h of treatment under argon, an equal volume of a quenching buffer (6 M Gn·HCl, 0.2 M DTT, pH 7.0) was added.

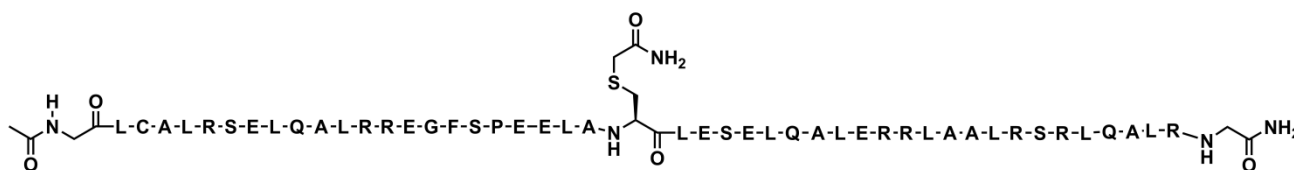
**Hom-N2** obtained after HPLC purification with an isolated yield of deprotection of 2.3 mg (25 %). LC/MS (ESI):  $[M + 4 H^+] = 1348.2$  m/z;  $[M + 5 H^+] = 1078.9$  m/z;  $[M + 6 H^+] = 899.2$  m/z;  $[M + 7 H^+] = 770.9$  m/z;  $[M + 8 H^+] = 674.7$  m/z;  $C_{231}H_{396}N_{76}O_{70}S$ ; Average isotope calculated: 5390.2 Da [M]; observed: 5389.6 Da.

### Synthesis of Hom-N2 by the three-segment ligation method

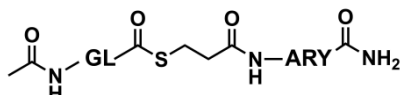
Synthetic route:



Sequence: Ac-GLCALRSELQALRREGFSPEELAC(capped)LESELQALERRLAALRSRLQALRG-<sup>α</sup>amide



Peptide 1-<sup>α</sup>thioester:



Peptide 1-<sup>α</sup>thioester was synthesized following the **General procedure V** on a 0.4 mmol scale using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated yield was 100 mg (35 %) based on peptide synthesis scale. LC/MS (ESI):  $[M + 1 H^+] = 708.3$  m/z;  $C_{31}H_{49}N_9O_8S$ ; Average isotope calculated: 707.8 Da  $[M]$ ; observed: 707.3 Da.

Peptide 2-<sup>α</sup>thioester:



Peptide 2-<sup>α</sup>thioester was synthesized following the **General procedure V** on a 0.2 mmol scale using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated yield was 89 mg (16 %) based on peptide synthesis scale. LC/MS (ESI):  $[M + 2 H^+] = 1394.1$  m/z;  $[M + 3 H^+] = 929.8$  m/z;  $[M + 4 H^+] = 697.7$  m/z;  $[M + 5 H^+] = 558.4$  m/z;  $C_{117}H_{196}N_{40}O_{35}S_2$ ; Average isotope calculated: 2787.2 Da  $[M]$ ; observed: 2786.9 Da.

Cys-peptide 3:



Cys-peptide 3 was synthesized following the **General procedure V** on a 0.4 mmol scale using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated yield was 155 mg (14 %) based on peptide synthesis scale. LC/MS (ESI):  $[M + 2 H^+] = 1426.7$  m/z;  $[M + 3 H^+] =$

= 951.5 m/z;  $[M + 4 H^+] = 713.9$  m/z;  $[M + 5 H^+] = 571.4$  m/z;  $[M + 6 H^+] = 476.6$  m/z;  
 $C_{120}H_{215}N_{43}O_{35}S$ ; Average isotope calculated: 2852.4 Da [M]; observed: 2851.9 Da.

#### *One-pot ligation.*

Peptide 3 (1.5 eq,  $8.76 \cdot 10^{-6}$  mol) was dissolved in ligation aqueous buffer (200 mM  $Na_2HPO_4$ , 6 M Gn·HCl, 20 mM TCEP, 15 mM MPAA, pH 7) to have a peptide concentration of 7.5 mM. The resulting solution was poured onto peptide 2 (1 eq,  $5.84 \cdot 10^{-6}$  mol, 5 mM). The pH was adjusted to 6.8. The reaction mixture was purged with argon and stirred at 40 °C, under Ar, during 3 hours. Completion of the reaction was followed by HPLC and LC/MS. Then, 2-bromoacetamide was introduced in 3-fold excess over total thiols ( $9.78 \cdot 10^{-5}$  mol) in solution at 1 M in neutral aqueous buffer (200 mM  $Na_2HPO_4$ , 6 M Gn·HCl, pH 7, purged with argon). The reaction mixture was stirred at 40 °C, under Ar, during 15 minutes. After verifying completion of the reaction, MPAA ( $2.6 \cdot 10^{-4}$  mol, ~ 200 mM) was added as solid into the reaction mixture to quench the excess of 2-bromoacetamide. The resulting solution was stirred at 40 °C, under Ar, during 30 minutes.  $MeONH_2 \cdot HCl$  ( $6.8 \cdot 10^{-4}$  mol) dissolved in 400  $\mu$ L of neutral aqueous buffer containing 20 mM TCEP was introduced in the reaction mixture to have a final and total concentration of 0.4 M. The pH decreased to ~ 4.0. The reaction solution was purged with argon and stirred at room temperature, under Ar, during 2.5 hours. Finally, peptide 1 (2 eq,  $1.17 \cdot 10^{-5}$  mol) was dissolved in 300  $\mu$ L of ligation aqueous buffer and added to reaction mixture. The pH was adjusted to 6.8 and the solution was purged with argon. Reaction mixture was stirred at 40 °C, under Ar, during 1 hour. After completion of the last ligation, the product was directly isolated by HPLC.

**Hom-N2:** isolated with 9.7 mg (30 %) ligation yield. LC-MS (ESI):  $[M + 4 H^+] = 1370.5$  m/z;  $[M + 5 H^+] = 1096.7$  m/z;  $[M + 6 H^+] = 914.1$  m/z;  $[M + 7 H^+] = 783.7$  m/z;  $[M + 8 H^+] = 685.8$  m/z;  
 $C_{233}H_{399}N_{77}O_{71}S_2$ ; Average isotope calculated: 5479.3 Da [M]; observed: 5478.5 Da.

### **Sequential Boc-SPPS to assemble the monomers composing Het-N2 and Het-Ncap**

#### *Synthesis of 'positive' strand*

Sequence: Ac-NLAALRSELQALRREGFSPERLAALESRLQALERRLAALRSRLQALRG-<sup>o</sup>amide

'Positive' strand: from a 0.4 mmol scale synthesis following the **General procedure V** using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated with 57 mg (3 %) yield based on peptide synthesis scale. LC-MS (ESI):  $[M + 5 H^+] = 1095.0$  m/z;  $[M + 6 H^+] = 912.7$

m/z; [M + 7 H<sup>+</sup>] = 782.5 m/z; [M + 8 H<sup>+</sup>] = 684.9 m/z; [M + 9 H<sup>+</sup>] = 608.9 m/z; [M + 10 H<sup>+</sup>] = 548.3 m/z; C<sub>235</sub>H<sub>409</sub>N<sub>83</sub>O<sub>67</sub>; Average isotope calculated: 5469.4 Da [M]; observed: 5470.8 Da.

### Synthesis of **Cys-N2**

Sequence: Ac-GLCALRSELQALRREGFSPEELAALESELQALERELAALRSELQALRG-<sup>α</sup>amide

**Cys-N2:** from a 0.3 mmol scale synthesis following the **General procedure V** using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated with 29.7 mg (2 %) yield based on peptide synthesis scale. LC-MS (ESI): [M + 4 H<sup>+</sup>] = 1335.5 m/z; [M + 5 H<sup>+</sup>] = 1068.3 m/z; [M + 6 H<sup>+</sup>] = 890.4 m/z; C<sub>229</sub>H<sub>386</sub>N<sub>70</sub>O<sub>74</sub>S; Average isotope calculated: 5336.1 Da [M]; observed: 5337.0 Da.

### Synthesis of **Cys-Ncap**

Sequence: Ac-CLAALRSELQALRREGFSPEELAALESELQALERELAALRSELQALRG-<sup>α</sup>amide

**Cys-Ncap:** from a 0.4 mmol scale synthesis following the **General procedure V** using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated with 34.8 mg (2 %) yield based on peptide synthesis scale. LC-MS (ESI): [M + 3 H<sup>+</sup>] = 1784.4 m/z; [M + 4 H<sup>+</sup>] = 1338.4 m/z; [M + 5 H<sup>+</sup>] = 1070.8 m/z; [M + 6 H<sup>+</sup>] = 892.6 m/z; C<sub>230</sub>H<sub>388</sub>N<sub>70</sub>O<sub>74</sub>S; Average isotope calculated: 5350.1 Da [M]; observed: 5349.3 Da.

## **Crystallization, data collection and structure refinement of homodimer Hom-N2**

The crystallization experiments were carried out by the sitting drop vapor diffusion method at 293 K using a Mosquito Crystal nanolitre dispensing robot (TTP Labtech) and screening against several commercially available screens: The Classics Suite (Qiagen), the JCSG-plus, Morpheus, and the BCS screen (Molecular dimensions).

For the **Hom-N2** homodimer, 0.15 μL of solution at 10.0 mg/mL in MES buffer (100 mM MES, 100 mM NaCl, 10 mM TCEP, pH 6.5) was mixed with 0.15 μL precipitant solution and equilibrated against 50 μL reservoir solution.



### **Crystallization, data collection and structure refinement of heterodimers Het-N2 and Het-Ncap**

The crystallization experiments were carried out by the sitting drop vapor diffusion method at 293 K using a Mosquito Crystal nanolitre dispensing robot (TTP Labtech) and screening against several commercially available screens: The Classics Suite (Qiagen), the JCSG-plus, Morpheus, and the BCS screen (Molecular dimensions).

For the **Het-N2** heterodimer, 0.15  $\mu\text{L}$  of solution at 10.1 mg/mL (1:1 stoichiometric ratio of 'positive' and **Cys-N2** strands) in MES buffer (100 mM MES, 100 mM NaCl, 10 mM TCEP, pH 6.5) was mixed with 0.15  $\mu\text{L}$  precipitant solution and equilibrated against 50  $\mu\text{L}$  reservoir solution. Crystals appeared after 3 months in condition 30% PEG 400, 0.1 M cadmium chloride, 0.1 M sodium acetate pH 4.6.

For the **Het-Ncap** heterodimer, 0.2  $\mu\text{L}$  of solution at 13 mg/mL (1:1 stoichiometric ratio of 'positive' and **Cys-Ncap** strands) in MES buffer (100 mM MES, 100 mM NaCl, 10 mM TCEP, pH 6.5) was mixed with 0.2  $\mu\text{L}$  precipitant solution and equilibrated against 50  $\mu\text{L}$  reservoir solution. A single crystal appeared overnight in condition 25% PEG 550 MME, 0.01 M zinc sulfate, 0.1 M HEPES pH 7.5.

Crystals were flash frozen by direct immersion in liquid nitrogen without further cryo-protection.

For the **Het-N2** heterodimer, data were collected at 100 K on an EIGER X 4M detector (Dectris) on the ID30A-3 beamline of the ESRF. 360° of data were collected using 0.1° rotation and 0.01 sec exposure per image. The data were indexed, integrated, and scaled using XDS,<sup>11</sup> before anisotropic truncation and correction were performed using the STARANISO server. The data were processed to 2.33 Å (ellipsoidal cutoff 2.28 Å, 2.49 Å, 3.48 Å) and belonged to the primitive triclinic space group P1, with unit cell dimensions  $a = 40.7$  Å,  $b = 49.9$  Å,  $c = 53.04$  Å,  $\alpha = 105.09^\circ$ ,  $\beta = 92.49^\circ$ ,  $\gamma = 110.15^\circ$ . The structure was solved by molecular replacement using PHASER<sup>12</sup> in the PHENIX<sup>13</sup> suite using the known DSD structure as a search model (PDB ID: 1G6U).<sup>14</sup> The asymmetric unit contains four copies of the heterodimeric DSD analogue, with a corresponding Matthews' coefficient<sup>15</sup> of 2.23 Å<sup>3</sup> Da<sup>-1</sup> and a solvent content of 44.92 %.

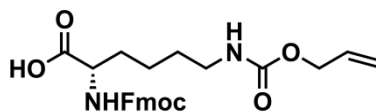
For the **Het-Ncap** heterodimer, data were collected at 100 K on a PILATUS 6M detector (Dectris) on the ID23-1 beamline of the ESRF. 180° of data were collected using 0.15° rotation and 0.037 sec exposure per image. The data were indexed, integrated, and scaled using XDS before merging the data in AIMLESS.<sup>16</sup> The data were processed to 1.45 Å and belonged to the C-centred monoclinic space group C2, with unit cell dimension  $a = 76.9$  Å,  $b = 73.1$  Å,  $c = 84.1$  Å,  $\beta = 105.02^\circ$ . The structure was solved as above. The

asymmetric unit contains four copies of the heterodimeric DSD analogue, with a corresponding Matthews' coefficient of  $2.63 \text{ \AA}^3 \text{ Da}^{-1}$  and a solvent content of 53.29 %.

The structures were refined using PHENIX and BUSTER with iterative model building performed in COOT.<sup>17</sup> The quality of the final refined models was assessed using MOLPROBITY<sup>18</sup> and PROCHECK.<sup>19</sup> Data collection and refinement statistics are given in **Table 7.1**. For definitions of the statistics given in the table see Table 1 of ref. 20.<sup>20</sup> The coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes 6Z0L (**Het-N2**) and 6Z0M (**Het-Ncap**).

**Table 7.1. Data collection and refinement statistics for Het-N2 and Het-Ncap.** The values given in parenthesis correspond to the highest resolution shells. For the Het-N2, the data were anisotropic, and the statistics for the anisotropic truncation performed by the STARANISO server are given.

	Het-N2	Het-Ncap
<b>PDB ID</b>	6Z0L	6Z0M
<b>Data Collection</b>		
Source	ESRF ID30A-3	ESRF ID23-1
Detector	EIGER X 4M	PILATUS 6M
Wavelength (Å)	0.968	0.97
Resolution range (Å)	21.76 - 2.33 (2.417 - 2.33)	52.09 - 1.45 (1.502 - 1.45)
Space group	P 1	C 1 2 1
Cell Dimensions: <i>a, b, c</i> (Å)	40.67, 49.95, 53.04	76.88, 73.10, 84.13
<i>α, β, γ</i> (°)	105.09, 92.49, 110.15	90, 105.021, 90
Total reflections	47719 (4014)	265700 (25795)
Unique reflections	14476 (139)	78982 (7859)
Multiplicity	3.3 (2.9)	3.4 (3.3)
Completeness (%)	54.82 (8.56)	99.15 (99.12)
Mean I/σ(I)	2.98 (0.26)	14.17 (1.49)
Wilson B-factor	14.56	20.91
R-merge	0.257 (3.230)	0.036 (0.697)
R-meas	0.307 (3.905)	0.043 (0.833)
R-pim	0.165 (2.166)	0.023 (0.452)
CC1/2	0.983 (0.151)	0.999 (0.811)
CC*	0.996 (0.526)	1 (0.946)
<b>Anisotropic truncation</b>		
Ellipsoidal resolution (Å) (direction)	3.48 (0.791 <i>a</i> * - 0.079 <i>b</i> * + 0.607 <i>c</i> *)	
	2.28 (0.069 <i>a</i> * + 0.730 <i>b</i> * - 0.680 <i>c</i> *)	
	2.49 (-0.516 <i>a</i> * + 0.615 <i>b</i> * + 0.597 <i>c</i> *)	
Best / worst diffraction limits (Å)	2.33 (2.53 – 2.33) / 4.43	
Total reflections (ellipsoidal)	29603 (1390)	
Unique reflections (ellipsoidal)	8746 (435)	
Mean I/σ(I) (ellipsoidal)	4.1 (1.8)	
R-merge (ellipsoidal)	0.246 (0.822)	
R-meas (ellipsoidal)	0.292 (0.983)	
R-pim (ellipsoidal)	0.157 (0.534)	
CC1/2 (ellipsoidal)	0.963 (0.503)	
Ellipsoidal completeness (%)	84.2 (67.0)	
<b>Refinement</b>		
Reflections used in refinement	8740 (139)	78982 (7850)
Reflections used for R-free	486 (6)	4028 (381)
R-work	0.223 (0.277)	0.152 (0.244)
R-free	0.283 (0.196)	0.207 (0.290)
CC(work)	0.772 (0.494)	0.968 (0.922)
CC(free)	0.823 (0.810)	0.955 (0.889)
Number of non-hydrogen atoms	3121	4040
macromolecules	2998	3380
ligands	38	38
solvent	85	622
Protein residues	400	400
RMS(bonds)	0.002	0.002
RMS(angles)	0.38	0.51
Ramachandran favored (%)	99.73	97.82
Ramachandran allowed (%)	0.27	2.18
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	3.39	1.93
Clashscore	7.91	5.9
Average B-factor	34.19	33.26
macromolecules	34.34	30.83
ligands	47.72	55.21
solvent	23.05	45.15

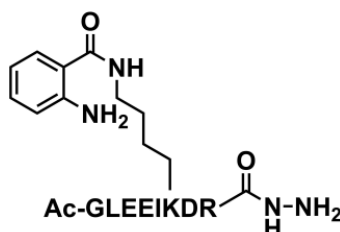
**Synthesis of the Fmoc-Lys(Alloc)-OH building block**

To 1 g of Fmoc-Lys(Boc)-OH ( $n = 2.1$  mmol) were added 3.2 mL of TFA. The reaction was stirred at room temperature for 20 min. The reaction was monitoring by TLC and HPLC. Then, the solution was evaporated under vacuum to remove TFA. Yellow oil was obtained and dissolved in 16 mL of distilled water. Saturated solution of  $\text{Na}_2\text{CO}_3$  (1.1 mL) was added until reaching pH 9. A precipitate was formed and 12 mL of 1,4-dioxane were added. The pH was carefully adjusted to 9 and the solution became clear. 273  $\mu\text{L}$  of allylchloroformate ( $n = 2.6$  mmol, 1.2 equivalents) were dissolved in 4 mL of 1,4-dioxane. The resulting solution was introduced to the reaction mixture and stirred at room temperature for 1 h and 10 min. Again, a TLC and HPLC monitoring was performed. Solvents and excess of reagent were then evaporated under vacuum and the retrieved product was solubilized in 50 mL of water. The pH was lowered to 2 with a solution of 6 M HCl. As a result, the precipitate was formed and then solubilized with the addition of ethyl acetate. Aqueous phase was extracted three times with ethyl acetate. The organic phases were collected together, washed with brine and dried with  $\text{Na}_2\text{SO}_4$ . The organic phase was evaporated under vacuum. The product that consisted of solid and oil was dissolved in acetonitrile and the same volume of water was added. The solution was then frozen in liquid nitrogen while stirring to homogenize the mixture and finally lyophilized. After extraction and lyophilization, 0.91 g of product was recovered with a yield 93.7 % before purification. Half of the Fmoc-Lys(Alloc)-OH obtained was then purified by reverse phase flash chromatography with a gradient of water + 0.1 % TFA and acetonitrile + 0.08 % TFA from 5 % to 80 %. The product was obtained in high purity as a white powder with isolated yield of 51 % (242 mg obtained from half of the crude).  $R_f$  ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  [9/1] + 0.5 % AcOH) = 0.79; LC/MS (ESI):  $\text{C}_{25}\text{H}_{28}\text{N}_2\text{O}_6$ ; Average isotope calculated: 452.5 g/mol [M]; observed: 452.0 g/mol.

**$^1\text{H}$  NMR (400 MHz,  $\text{DMSO}-d_6$ )  $\delta$ :** 7.89 ppm (d,  $J = 7.5$  Hz, 2H), 7.73 ppm (d,  $J = 7.5$  Hz, 2H), 7.61 ppm (d,  $J = 8.1$  Hz, 1H), 7.42 ppm (t,  $J = 7.4$  Hz, 2H), 7.33 ppm (t,  $J = 7.4$  Hz, 2H), 7.18 ppm (t,  $J = 5.8$  Hz, 1H), 5.90 ppm (m, 1H), 5.26 ppm (dd,  $J = 17.2$ ,  $J = 1.9$  Hz, 1H), 5.16 ppm (dd,  $J = 10.4$ ,  $J = 1.8$  Hz, 1H), 4.45 ppm (d,  $J = 5.3$  Hz, 1H), 4.32 ppm (m, 2H), 4.18 ppm (m, 1H), 2.98 ppm (q,  $J = 6.5$  Hz, 2H), 1.77 ppm (m, 1H), 1.54 ppm (m, 1H), 1.45 – 1.36 ppm (m, 2H), 1.32 ppm (m,  $J = 14.0$ , 5.9 Hz, 2H).

**Synthesis of peptide 1 conjugated with the 2-aminobenzoyl (Abz) fluorophore for FRET experiments using a quencher**

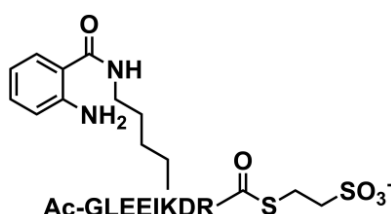
Sequence: Ac-GLEEIKDR- $\alpha$ hydrazide



To couple lysine, the previously synthesized Fmoc-Lys(alloc)-OH was used. Fmoc-SPPS was performed to assemble the peptide following **General procedure II** on a 0.1 mmol scale using a 2-CT-NHNH<sub>2</sub> resin (0.4 mmol/g loading, 200-400 mesh). After N-terminal acetylation, lysine deprotection by removing the alloc group was achieved.<sup>21</sup> The resin was washed with dichloromethane, and then swollen in dichloromethane previously purged with argon for 20 min. A solution of 296  $\mu$ L of phenylsilane in 1 mL of dichloromethane was degassed with argon and added to the beads. To initiate the deprotection, a second solution with 58 mg of Pd(Ph<sub>3</sub>)<sub>4</sub> dissolved in 2 mL of dichloromethane purged with argon was poured into the reaction vessel. The reaction lasted 30 min at room temperature while stirring manually. After being rinsed and batch washed the resin with dichloromethane and then DMF, a Kaiser test was realized to confirm the completeness of the deprotection. The same procedure as the one executed for the coupling of amino acid was applied to incorporate Boc-Abz-OH building block. The fluorophore compound (0.4 mmol) was pre-activated with 1 mL of HBTU (*c* 0.38 M in DMF) and 110  $\mu$ L of the base, DIEA. The mixture was vortexed and let to react for 2 min. The mixture was added to the resin and stirred every 5 min during 30 min. The beads were then rinsed twice and batch washed twice. Cleavage and workup were executed according to **General procedure II**.

39.1 mg of peptide crude were obtained and analysis showed expected mass: LC/MS (ESI): Average isotope calculated: 1134.3 Da [M]; observed: 1133.7 Da.

Sequence: Ac-GLEEIKDR- $\alpha$ Mes



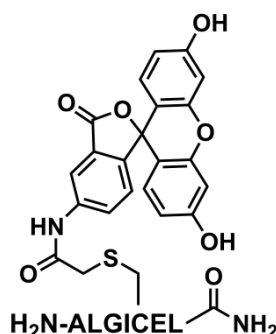
Following **General procedure VI**, thioesterification of peptide 1 was tried but side reaction consisting in diazotization reaction on aryl amine occurred. Synthesis was abandoned and actually another methodology to monitor catalytic assays was preferred.

### Synthesis of peptide 2 labelled with a fluorescein derivative for FRET experiments using two fluorophores

Sequence: ALGICEL- $\alpha$ amide

The peptide 2 was prepared on a 0.1 mmol scale synthesis following **General procedure II** on a Rink amide resin (0.74 mmol/g loading, 100-200 mesh); isolated with 53.4 mg (74 %) yield. LC-MS (ESI):  $[M + 1 H^+] = 717.2$  m/z;  $C_{31}H_{56}N_8O_9S$ ; Average isotope calculated: 716.9 Da  $[M]$ ; observed: 716.2 Da.

Sequence: ALGIC(fluorescein)EL- $\alpha$ amide



Reaction was performed on  $1 \times 10^{-7}$  mol of peptide. A solution of peptide 2 at 1 mM in carbonate buffer (0.1 M  $NaHCO_3$ , 5 mM TCEP, pH 7.5) was first prepared and pH was measured and adjusted if necessary to 7.5 (here, pH 7.7 was not adjusted). 5-(iodoacetamido)fluorescein was solubilized in anhydrous DMF at a concentration of 5 mM. Both solutions were mixed together in equal volume. The reaction was stirred at room temperature with protection against external light and was monitored by HPLC and LC/MS. After completion (after 30 minutes of reaction), the reaction was quenched with 10-time excess of DTT compared to the iodoacetamide fluorescein. Peptide was isolated by preparative HPLC but yield was not determined due to too small quantity obtained. LC-MS (ESI):  $[M + 1 H^+] = 1104.4$  m/z;  $[M + 2 H^+] = 552.7$  m/z;  $C_{53}H_{69}N_9O_{15}S$ ; Average isotope calculated: 1104.2 Da  $[M]$ ; observed: 1103.3 Da.

### Synthesis of peptide 1 labelled with a rhodamine derivative for FRET experiments using two fluorophores

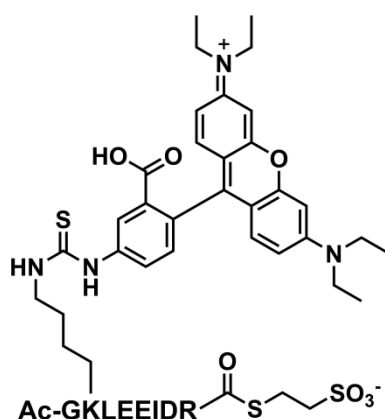
Sequence: Ac-GKLEEIDR- $\alpha$ hydrazide

The hydrazide peptide was obtained from a 0.1 mmol scale synthesis following manual Fmoc-SPPS of **General procedure II** using a 2-CT-NHNH<sub>2</sub> resin (0.4 mmol/g loading, 200-400 mesh). Peptide was not isolated due to good purity of the crude compatible with further thioesterification step. LC-MS (ESI): [M + 1 H<sup>+</sup>] = 1015.5 m/z; [M + 2 H<sup>+</sup>] = 508.4 m/z; C<sub>42</sub>H<sub>74</sub>N<sub>14</sub>O<sub>15</sub>; Average isotope calculated: 1015.1 Da [M]; observed: 1014.6 Da.

Sequence: Ac-GKLEEIDR- $\alpha$ Mes

Thioesterification of the peptide hydrazide surrogate was performed following **General procedure VI**; isolated yield of the thioesterification (on 2.46·10<sup>-5</sup> mol of peptide hydrazide) was 8.8 mg (32 %). LC/MS (ESI): [M + 1 H<sup>+</sup>] = 1125.7 m/z; [M + 2 H<sup>+</sup>] = 563.3 m/z; C<sub>44</sub>H<sub>76</sub>N<sub>12</sub>O<sub>18</sub>S<sub>2</sub>; Average isotope calculated: 1125.3 Da [M]; observed: 1124.7 Da.

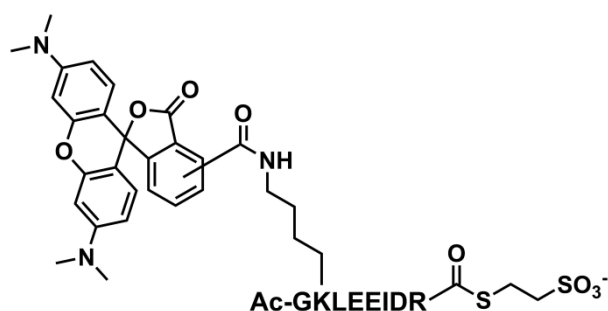
Sequence: Ac-GK(rhodamine B)LLEEIDR- $\alpha$ Mes (*not obtained*)



Solution of peptide 1- $\alpha$ Mes in anhydrous DMF at a concentration of 2 mM was prepared. In the same way, solution of rhodamine B isothiocyanate (in the form of mixed isomers) in anhydrous DMF was prepared to have a concentration of 2 mM. Stock solution of DIEA in anhydrous DMF at a concentration of 50 mM was also prepared. Same volume of both solutions of peptide and rhodamine B isothiocyanate was mixed together to obtain 1 mM of final concentration for each compound and 1  $\mu$ L of the stock solution of DIEA was added to have a final concentration of 0.5 mM. Reaction was stirred overnight under argon

and protected from light. After overnight reaction, no labelling was observed and 4  $\mu\text{L}$  of stock solution of DIEA was added to the reaction mixture. After five more hours of reaction, labelled product was still not monitored and only non labelled hydrolyzed thioester was detected.

Sequence: Ac-GK(tetramethylrhodamine)LEEIDR- $^{\alpha}\text{Mes}$  (*not obtained*)



Solution of peptide 1- $^{\alpha}\text{Mes}$  in anhydrous DMF at a concentration of 2 mM was prepared. In the same way, solution of 5(6)-carboxytetramethylrhodamine N-hydroxysuccinimide ester in anhydrous DMF was prepared to have a concentration of 2 mM. Stock solution of DIEA in anhydrous DMF at a concentration of 50 mM was also prepared. Same volume of both solutions of peptide and NHS-5(6)-carboxytetramethylrhodamine was mixed together to obtain 1 mM of final concentration for each compound and 1  $\mu\text{L}$  of the stock solution of DIEA was added to have a final concentration of 0.5 mM. Reaction was stirred overnight under argon at room temperature in the dark. After overnight reaction, peptide thioester was completely hydrolyzed but was partially labelled.

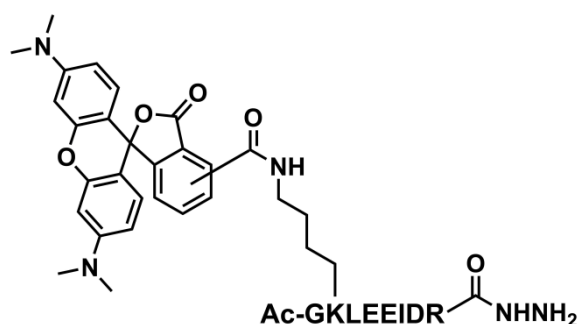
Peptide 1- $^{\alpha}\text{OH}$ : MS (MALDI-TOF,  $\alpha$ -cyano-4-hydroxycinnamic acid (CHCA) matrix):  $[\text{M} + 1 \text{H}^+] = 1001.7 \text{ m/z}$ .  $\text{C}_{42}\text{H}_{72}\text{N}_{12}\text{O}_{16}$ ; Average isotope calculated: 1001.1 Da [M]; observed: 1000.7 Da.

Labelled peptide 1- $^{\alpha}\text{OH}$  with tetramethylrhodamine: MS (MALDI-TOF, CHCA matrix):  $[\text{M} + 1 \text{H}^+] = 1414.9 \text{ m/z}$ .  $\text{C}_{67}\text{H}_{92}\text{N}_{14}\text{O}_{20}$ ; Average isotope calculated: 1413.6 Da [M]; observed: 1413.9 Da.

Due to complete hydrolysis, we decided to first label the precursor peptide 1- $^{\alpha}$ hydrazide and then perform thioesterification. We performed the reaction on the crude peptide to verify first if the reaction can work.



Sequence: Ac-GK(tetramethylrhodamine)LEEIDR- $\alpha$ hydrazide



10 mg of peptide 1- $\alpha$ hydrazide ( $9.9 \times 10^{-6}$  mol, 1 eq) were solubilized in 1250  $\mu$ L of anhydrous DMF. In parallel, 5.2 mg of NHS-5(6)-carboxytetramethylrhodamine ( $9.9 \times 10^{-6}$  mol, 1 eq) were dissolved in 500  $\mu$ L of anhydrous DMF. The two resulting solutions were mixed together and 2.1  $\mu$ L of DIEA ( $1.12 \times 10^{-5}$  mol, 1.2 eq) were added to the resulting mixture. The reaction was stirred under argon in the dark and monitored by HPLC and LC/MS. After 6 hours, evolution of the reaction was too slow and 2.1  $\mu$ L of DIEA (1.2 eq) was again added to the reaction mixture. The reaction was stirred overnight and was stopped when around 40 % of peptide was labelled. Labelled peptide with tetramethylrhodamine was then isolated by preparative HPLC. Yield was not determined due to too low amount recovered after purification. LC-MS (ESI):  $[M + 1 H^+] = 1427.8$  m/z;  $[M + 2 H^+] = 714.7$  m/z;  $[M + 3 H^+] = 476.6$  m/z;  $C_{67}H_{94}N_{16}O_{19}$ ; Average isotope calculated: 1427.6 Da [M]; observed: 1427.0 Da.

### **Synthesis of the N-terminal fragment of ACTR [N-ACTR]- $\alpha$ Mes**

Sequence: EGQSDERALLDQLHTLLSNTDATGLEEIDR- $\alpha$ Mes

[N-ACTR] - $\alpha$ Mes was previously synthesized by other members of the laboratory.<sup>22</sup> The synthesis was performed according to the **General procedure III** as a hydrazide surrogate using a 2-CT-NHNH<sub>2</sub> resin (0.4 mmol/g loading, 200-400 mesh) followed by thioesterification according to the **General procedure VI**. Average isotope calculated: 3463.7 Da [M]; observed: 3463.8 Da.

### **Synthesis of the C-terminal fragment of ACTR [C-ACTR]**

Sequence: CLGIPELVNQQQALEPK- $\alpha$ OH

[C-ACTR] was previously synthesized by other members of the laboratory.<sup>22</sup> The synthesis was performed according to the **General procedure III** using a Wang resin preloaded with a Fmoc-L-Lys(Boc). Average isotope calculated: 1809.1 Da [M]; observed: 1809.0 Da.

### **Synthesis of peptide1-Arg-<sup>α</sup>Mes**

Sequence: Ac-GRLEEIDR-<sup>α</sup>Mes

Peptide1-Arg-<sup>α</sup>Mes was synthesized by Fmoc-SPPS on a 0.1 mmol scale according to the **General procedure II** as a hydrazide surrogate using a 2-CT-NHNH<sub>2</sub> resin (0.4 mmol/g loading, 200-400 mesh) followed by thioesterification on crude according to the **General procedure VI**; isolated yield of the thioesterification (on 2.4·10<sup>-5</sup> mol of peptide hydrazide) was 16.0 mg (58 %). LC/MS (ESI): [M + 1 H<sup>+</sup>] = 1153.6 m/z; [M + 2 H<sup>+</sup>] = 577.4 m/z; C<sub>44</sub>H<sub>76</sub>N<sub>14</sub>O<sub>18</sub>S<sub>2</sub>; Average isotope calculated: 1153.3 Da [M]; observed: 1152.6 Da.

### **Synthesis of peptide-Phe-<sup>α</sup>Mes**

Sequence: Ac-VALENF-<sup>α</sup>Mes

Peptide-Phe-<sup>α</sup>Mes was synthesized by Fmoc-SPPS on a 0.1 mmol scale according to the **General procedure III** as a hydrazide surrogate using a 2-CT-NHNH<sub>2</sub> resin (0.4 mmol/g loading, 200-400 mesh) followed by thioesterification on crude according to the **General procedure VI**; isolated yield of the thioesterification (on 1.3·10<sup>-5</sup> mol of peptide hydrazide) was 2.5 mg (21 %). LC/MS (ESI): [M + 1 H<sup>+</sup>] = 858.2 m/z; C<sub>36</sub>H<sub>55</sub>N<sub>7</sub>O<sub>13</sub>S<sub>2</sub>; Average isotope calculated: 858.0 Da [M]; observed: 857.2 Da.

### **Synthesis of peptide-Gly-<sup>α</sup>Mes**

Sequence: Ac-LYRAG-<sup>α</sup>Mes

Peptide-Gly-<sup>α</sup>Mes was synthesized by Fmoc-SPPS on a 0.1 mmol scale according to the **General procedure IV** as a hydrazide surrogate using a 2-CT-NHNH<sub>2</sub> resin (0.4 mmol/g

loading, 200-400 mesh) followed by thioesterification on crude according to the **General procedure VI**; isolated yield of the thioesterification (on  $3.15 \cdot 10^{-5}$  mol of peptide hydrazide) was 7.3 mg (31 %). LC/MS (ESI):  $[M + 1 H^+] = 745.1$  m/z;  $C_{30}H_{48}N_8O_{10}S_2$ ; Average isotope calculated: 744.9 Da [M]; observed: 744.1 Da.

### **Synthesis of peptide2-Arg-<sup>α</sup>Mes**

Sequence: Ac-VALENR-<sup>α</sup>Mes

Peptide2-Arg-<sup>α</sup>Mes was synthesized by Fmoc-SPPS on a 0.1 mmol scale according to the **General procedure III** as a hydrazide surrogate using a 2-CT-NHNH<sub>2</sub> resin (0.4 mmol/g loading, 200-400 mesh) followed by thioesterification on crude according to the **General procedure VI**; isolated yield of the thioesterification (on  $1.32 \cdot 10^{-5}$  mol of peptide hydrazide) was 4.2 mg (37 %). LC/MS (ESI):  $[M + 1 H^+] = 867.3$  m/z;  $C_{33}H_{58}N_{10}O_{13}S_2$ ; Average isotope calculated: 867.0 Da [M]; observed: 866.3 Da.

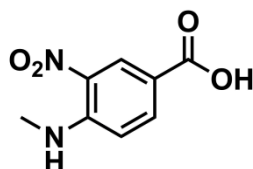
### **Hydrolytic activity of Het-N2 with p-NPA as substrate**

Hydrolytic activity of **Het-N2** was monitored with *para*-nitrophenyl acetate following the **General procedure IX**. Kinetics curves were monitored at different initial concentration of pNPA,  $[pNPA]_0$ . Kinetic parameters were extracted following the equation  $[pNP] = At + B(1 - e^{-bt})$  and are summarized in **Table 7.2**.<sup>8</sup>

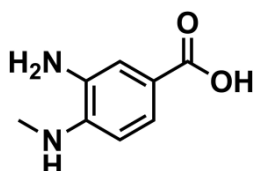
**Table 7.2. Kinetic parameters that were deduced from kinetic curves at each concentration of pNPA substrate by fitting the equation:  $[pNP] = At + B(1 - e^{-bt})$ . \*  $k_2/K_s$  is related to the initial burst phase efficiency**

$[pNPA]_0$	400 $\mu$ M	600 $\mu$ M	800 $\mu$ M	1000 $\mu$ M	1250 $\mu$ M
<b>A (<math>\mu</math>M sec<sup>-1</sup>)</b>	0.0038	0.0053	0.0063	0.0073	0.0078
<b>B (<math>\mu</math>M)</b>	38.4	35.7	31.9	36.8	52.0
<b>b (sec<sup>-1</sup>)</b>	0.00058	0.00076	0.00087	0.0011	0.00075
<b><math>k_2/K_s</math> (M<sup>-1</sup> sec<sup>-1</sup>)*</b>	1.45	1.26	1.10	1.11	1.88

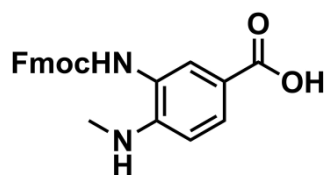
## 7.3. Chapter 3

**Synthesis of the Fmoc-MeDbz building block**<sup>23</sup>

2 g of fluoro-3-nitrobenzoic acid ( $10.8 \times 10^{-3}$  mol) were dissolved in 30 mL of MeOH. 8 mL of MeNH<sub>2</sub> in 40 % in MeOH was slowly added to the reaction mixture. Reaction was stirred overnight at room temperature. Completion of the reaction was monitored by taking 1  $\mu$ L of the reaction mixture and diluting it in 200  $\mu$ L of MeCN/ H<sub>2</sub>O [1:1] + 0.1 % TFA and injecting 2  $\mu$ L of the resulting solution on analytical HPLC. After overnight reaction, 30 mL of water was added to the reaction mixture and then acidified to pH 2.0 with HCl at 37 % until a yellow solid was completely precipitated. The solid was filtrated, washed once with filtrate and once with cold water. The yellow solid was dried under high vacuum. 2.05 g of 3-nitro-4-(methylamino)benzoic acid was obtained with a yield of 97 % (not purified). Rf (CH<sub>2</sub>Cl<sub>2</sub>/MeOH [9/1] + 1 % AcOH) = 0.68.



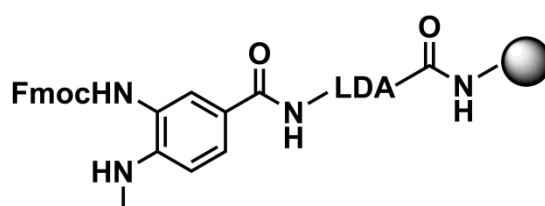
3-nitro-4-(methylamino)benzoic acid (2.05 g,  $10.45 \times 10^{-3}$  mol) was partially dissolved in 100 mL of MeOH. Solution was purged with argon before adding 100 mg of Pd/C (10 %). Reaction was again purged with argon followed by bubbling in the solution of one balloon of H<sub>2</sub>. Solution was stirred overnight, at room temperature under H<sub>2</sub>. The solution turned clear and greenish. Completion of the reaction was followed by TLC. At the end of the reaction, the palladium was removed by filtration on Celite®. MeOH was then removed under vacuum. 3-amino-4-(methylamino)benzoic acid was quantitatively obtained as solid (1.74 g, not purified). Rf (CH<sub>2</sub>Cl<sub>2</sub>/MeOH [9/1] + 1 % AcOH) = 0.5.



1.74 g of 3-amino-4-(methylamino)benzoic acid **3** ( $10.47 \times 10^{-3}$  mol, 1 eq) were dissolved in 20 mL of a solution of H<sub>2</sub>O/MeCN [1:1] containing 2 mL of DIEA ( $11.52 \times 10^{-3}$  mol, 1.1 eq). A solution of Fmoc-Cl (2.89 g in 12 mL MeCN,  $11.52 \times 10^{-3}$  mol, 1.1 eq) was poured dropwise. A precipitate (clear brown) was instantaneously observed. The reaction mixture was stirred 2 hours at room temperature. Completion of the reaction was verified by TLC. After completion of the reaction, the solvent was removed by rotary evaporator. Some water was added on the resulting solid. The solid was then filtrated and washed with the filtrate and cold MeCN until no more brown liquid flowed down. The resulting solid was then dried under high vacuum. 2.6 g of a light purple solid corresponding to the Fmoc-MeDbz building block was obtained with an overall yield of 62 % (some product was lost during MeCN washings). R<sub>f</sub> (Petroleum ether/AcOEt [4/6] + 1 % AcOH) = 0.32; LC/MS (ESI): C<sub>23</sub>H<sub>20</sub>N<sub>2</sub>O<sub>4</sub>; Average isotope calculated: 388.4 g/mol [M]; observed: 388.0 g/mol.

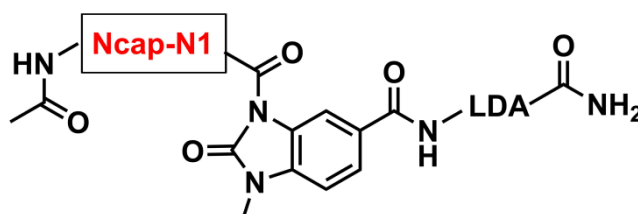
<sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ: 12.21 ppm (s, 1H), 8.72 ppm (s, 1H), 7.91 ppm (d, *J* = 7.6 Hz, 2H), 7.69 ppm (m, 4H), 7.43 ppm (t, *J* = 7.5 Hz, 2H), 7.34 ppm (m, 3H), 6.61 ppm (d, *J* = 8.6 Hz, 1H), 4.38 ppm (s, 2H), 4.29 ppm (s, 1H), 2.78 ppm (s, 3H).

### Assembly of the N-terminal fragments as Ac-Ncap-N1-MeNbz-LDA peptide<sup>23</sup>



The common part composed of the spacer sequence (LDA) and the linker (Fmoc-MeDbz) was first assembled on a common batch and the resulting resin was then split to perform couplings of the different sequences of Ncap and N1. Synthesis of the spacer sequence (LDA) was performed on a 0.4 mmol scale on a Rink amide resin (0.74 mmol/g loading, 100-200 mesh) following the **General procedure III**. After Fmoc deprotection of the leucine residue, the building block Fmoc-MeDbz was coupled manually. For a 0.4 mmol scale, 777 mg of synthesized Fmoc-MeDbz (2 mmol, 5eq) was dissolved in 5 mL of a solution of HATU (0.38 M in DMF, 4.75 eq). 488 μL of DIEA (7 eq) were then poured to the

resulting solution. The reaction mixture was vortexed and let for activation during 2 min. The resulting solution was then poured to the resin beads. The coupling reaction of the Fmoc-MeDbz lasted for 3 hours with constant stirring. The beads were rinsed and batch washed with DMF four times. DMF was well removed with vacuum. The resin was then split to perform the different syntheses of the N-terminal fragments on a 0.05 mmol scale.



*Fmoc-SPPS*: For each N-terminal fragment, the following protocol was applied. Automatic couplings and deprotections of the N1 and Ncap residues was performed following the **General procedure III** on a 0.05 mmol scale, the machine was programmed to do double coupling of both residues. After the final Fmoc deprotection, N-terminal acetylation of the peptides was performed by adding to the beads a mixture of DMF/DIEA/Ac<sub>2</sub>O (700  $\mu$ L/100  $\mu$ L/100  $\mu$ L). The reaction was let for 6 min while stirring. The beads were rinsed and batch washed with DMF four times.

*Cyclization of the linker to form the Nbz moiety*: The resin was washed with DCM and was then conditioned under argon. Resin was washed with dried DCM under argon. For a 0.05 mmol scale synthesis, 50.4 mg of *p*-nitrophenylchloroformate were dissolved in 1 mL of dried DCM. Resulting solution was added to the resin and reaction was let for 1 hour under argon with stirring. The beads were washed with DCM followed by DMF. One flow wash was performed with a solution of DIEA at 1 M in DMF and beads were then covered with the same solution. The reaction was let for 30 min under argon with stirring. Resin was washed with DMF followed with multiple washes of DCM and was dried with air flow.

*Cleavage*: The resin was weighted: for 100 mg of resin, 1 mL of cleavage cocktail (95 % TFA, 2.5 % TIPS, 2.5 % H<sub>2</sub>O) was added to the beads. The resulting mixture was stirred for 2 hours at room temperature.

*Work up*: The cleavage mixture was transferred into several 50 mL centrifuge tubes and chilled diethyl ether (-20 °C) was added to reach a total volume of 40 mL. Afterwards, crude peptide product precipitated and was centrifuged at 4500 rpm for 3 min. The supernatant was discarded and the peptide was dissolved in 15 mL of a 1:1 mixture MeCN/H<sub>2</sub>O, 0.1 % TFA. The solution was filtered to remove the resin beads. An aliquot was taken for HPLC and LC/MS analysis. The peptide solution was frozen in liquid nitrogen and

then lyophilized. For peptide containing histidine residues at either Ncap or N1 position, once the peptide was solubilized in MeCN/H<sub>2</sub>O, 0.1 % TFA, the last steps were rapidly performed to avoid spontaneous hydrolysis. The peptide was then purified on a preparative HPLC or used as crude in the ligation.

Ac-**SR**-MeNbz-LDA peptide was isolated with 17.4 mg (45 %) of yield based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 776.5 m/z; C<sub>33</sub>H<sub>49</sub>N<sub>11</sub>O<sub>11</sub>; Average isotope calculated: 775.8 Da [M]; observed: 775.5 Da.

Ac-**RR**-MeNbz-LDA peptide was isolated with 9.6 mg (23 %) of yield based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 844.9 m/z; [M + 2 H<sup>+</sup>] = 423.4 m/z; C<sub>36</sub>H<sub>56</sub>N<sub>14</sub>O<sub>10</sub>; Average isotope calculated: 844.9 Da [M]; observed: 844.3 Da.

Ac-**RS**-MeNbz-LDA peptide was isolated with 14 mg (36 %) of yield based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 776.5 m/z; C<sub>33</sub>H<sub>49</sub>N<sub>11</sub>O<sub>11</sub>; Average isotope calculated: 775.8 Da [M]; observed: 775.5 Da.

Ac-**GR**-MeNbz-LDA peptide was isolated with 10.2 mg (27 %) of yield based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 746.4 m/z; C<sub>32</sub>H<sub>47</sub>N<sub>11</sub>O<sub>10</sub>; Average isotope calculated: 745.8 Da [M]; observed: 745.4 Da.

Ac-**RL**-MeNbz-LDA peptide was isolated with 8.6 mg (21 %) of yield based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 802.5 m/z; C<sub>36</sub>H<sub>55</sub>N<sub>11</sub>O<sub>10</sub>; Average isotope calculated: 801.9 Da [M]; observed: 801.5 Da.

Ac-**RA**-MeNbz-LDA peptide was isolated with 12.9 mg (34 %) of yield based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 760.4 m/z; C<sub>33</sub>H<sub>49</sub>N<sub>11</sub>O<sub>10</sub>; Average isotope calculated: 759.8 Da [M]; observed: 759.4 Da.

Ac-**NR**-MeNbz-LDA peptide was isolated with 11.3 mg (28 %) of yield based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 803.4 m/z; C<sub>34</sub>H<sub>50</sub>N<sub>12</sub>O<sub>11</sub>; Average isotope calculated: 802.8 Da [M]; observed: 802.4 Da.

Ac-**GI**-MeNbz-LDA peptide was isolated with 5.8 mg (17 %) of yield based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 703.0 m/z; C<sub>32</sub>H<sub>46</sub>N<sub>8</sub>O<sub>10</sub>; Average isotope calculated: 702.8 Da [M]; observed: 702.0 Da.

Ac-**GV**-MeNbz-LDA peptide was isolated with 7.6 mg (22 %) of yield based on peptide synthesis scale. LC/MS (ESI):  $[M + 1 H^+] = 689.1$  m/z;  $C_{31}H_{44}N_8O_{10}$ ; Average isotope calculated: 688.7 Da [M]; observed: 688.1 Da.

Ac-**GF**-MeNbz-LDA peptide was isolated with 10.9 mg (30 %) of yield based on peptide synthesis scale. LC/MS (ESI):  $[M + 1 H^+] = 737.1$  m/z;  $C_{35}H_{44}N_8O_{10}$ ; Average isotope calculated: 736.8 Da [M]; observed: 736.1 Da.

Ac-**GY**-MeNbz-LDA peptide was isolated with 6.3 mg (17 %) of yield based on peptide synthesis scale. LC/MS (ESI):  $[M + 1 H^+] = 753.1$  m/z;  $C_{35}H_{44}N_8O_{11}$ ; Average isotope calculated: 752.8 Da [M]; observed: 752.1 Da.

Ac-**GW**-MeNbz-LDA peptide was isolated with 3.6 mg (9 %) of yield based on peptide synthesis scale. LC/MS (ESI):  $[M + 1 H^+] = 776.2$  m/z;  $C_{37}H_{45}N_9O_{10}$ ; Average isotope calculated: 775.8 Da [M]; observed: 775.2 Da.

Ac-**GH**-MeNbz-LDA peptide was not isolated due to its propensity to spontaneous hydrolysis. 32.9 mg of crude peptide were obtained. LC/MS (ESI):  $[M + 1 H^+] = 727.4$  m/z;  $C_{32}H_{42}N_{10}O_{10}$ ; Average isotope calculated: 726.7 Da [M]; observed: 726.4 Da.

Ac-**HL**-MeNbz-LDA peptide did not require purification step due to high quality of crude. 33.3 mg of crude peptide were obtained. LC/MS (ESI):  $[M + 1 H^+] = 783.3$  m/z;  $C_{36}H_{50}N_{10}O_{10}$ ; Average isotope calculated: 782.9 Da [M]; observed: 782.3 Da.

Ac-**HS**-MeNbz-LDA peptide was not isolated due to its propensity to spontaneous hydrolysis. 34.2 mg of crude peptide were obtained. LC/MS (ESI):  $[M + 1 H^+] = 757.4$  m/z;  $C_{33}H_{44}N_{10}O_{11}$ ; Average isotope calculated: 756.8 Da [M]; observed: 756.4 Da.

Ac-**EL**-MeNbz-LDA peptide did not require purification step due to high quality of crude. 28.8 mg of crude peptide were obtained. LC/MS (ESI):  $[M + 1 H^+] = 775.0$  m/z;  $C_{35}H_{50}N_8O_{12}$ ; Average isotope calculated: 774.8 Da [M]; observed: 774.0 Da.

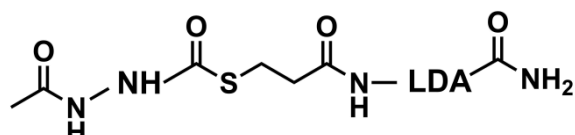
Ac-**YL**-MeNbz-LDA peptide did not require purification step due to high quality of crude. 32.1 mg of crude peptide were obtained. LC/MS (ESI):  $[M + 1 H^+] = 809.0$  m/z;  $C_{39}H_{52}N_8O_{11}$ ; Average isotope calculated: 808.9 Da [M]; observed: 808.0 Da.



Ac-**EY**-MeNbz-LDA peptide did not require purification step due to high quality of crude. 32.8 mg of crude peptide were obtained. LC/MS (ESI):  $[M + 1 H^+] = 825.1$  m/z;  $C_{38}H_{48}N_8O_{13}$ ; Average isotope calculated: 824.8 Da [M]; observed: 824.1 Da.

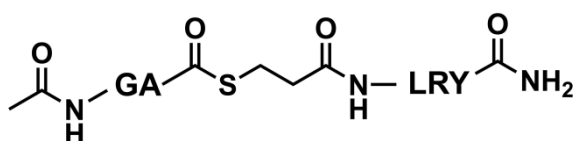
### **Boc-SPPS assembly of N-terminal fragments as thioester peptide**

Sequence: Ac-**NH**-MPA-LDA- $\alpha$ amide



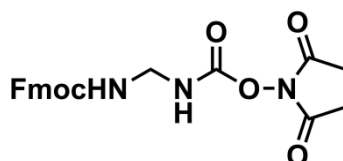
Ac-**NH**-MPA-LDA peptide was synthesized following the **General procedure V** on a 0.05 mmol scale using a MBHA resin (1.2 mmol/g loading, 100-200 mesh). Peptide did not require purification step due to high quality of crude. 117.3 mg of crude peptide were obtained. LC/MS (ESI):  $[M + 1 H^+] = 698.1$  m/z;  $C_{28}H_{43}N_9O_{10}S$ ; Average isotope calculated: 697.8 Da [M]; observed: 697.1 Da.

Sequence: Ac-**GA**-MPA-LRY- $\alpha$ amide



Ac-**GA**-MPA-LRY peptide was synthesized following the **General procedure V** on a 0.2 mmol scale using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated with 113.8 mg (78 %) of yield based on peptide synthesis scale. LC/MS (ESI):  $[M + 1 H^+] = 708.6$  m/z;  $C_{31}H_{49}N_9O_8S$ ; Average isotope calculated: 707.8 Da [M]; observed: 707.6 Da.

### **Synthesis of the glycine succinimidyl carbamate building block (Fmoc-Gly-CO-OSu)<sup>24</sup>**



The commercially available Fmoc-Gly-OH (1 g, 3.36 mmol, 1 eq) was conditioned with argon and was then dissolved in 10 mL of anhydrous THF. The solution was cooled

down to ~ -12 °C. Ethyl chloroformate (EtCOCl) (354 µL, 3.7 mmol, 1.1 eq) and *N*-methylmorpholine (NMM) (407 µL, 3.7 mmol, 1.1 eq) were poured to the solution and the resulting reaction mixture was stirred at -12 °C under argon for 20 min. The formation of a white suspension was observed. The reaction mixture was then warmed to -5 °C and aqueous solution of NaN<sub>3</sub> (546.1 mg in 1.7 mL of H<sub>2</sub>O, 8.4 mmol, 2.5 eq) was added. The reaction mixture was stirred under argon at -5 °C for 5 min. Then, EtOAc was poured into the reaction mixture and aqueous phase was extracted three times with EtOAc. Organic phases were collected together, washed with brine, dried with MgSO<sub>4</sub> and concentrated with rotary evaporator. To the resulting acyl azide, 10 mL of toluene were added under argon. The resulting solution was heated at 65 °C while stirring. Although the acyl azide derivative was not completely dissolved (white precipitate), gas release was observed and stopped after 15 min. After gas evolution had stopped, *N*-hydroxysuccinimide (387 mg, 3.36 mmol, 1 eq) and pyridine (272 µL, 3.36 mmol, 1eq) were added under argon. A more dense precipitate was instantaneously formed. After 5 min of stirring, mixture was cooled to room temperature. The residue was collected by filtration with 68 % (936 mg) of overall yield.

**<sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>) δ:** 8.94 ppm (t, *J* = 5.8 Hz, 1H), 8.23 ppm (t, *J* = 6.3 Hz, 1H), 7.89 ppm (d, *J* = 7.4 Hz, 2H), 7.72 ppm (d, *J* = 7.6 Hz, 2H), 7.42 ppm (t, *J* = 7.4 Hz, 2H), 7.34 ppm (t, *J* = 7.3 Hz, 2H), 4.40 ppm (t, *J* = 6.0 Hz, 2H), 4.32 ppm (m, 3H), 2.77 ppm (s, 4H).

**<sup>13</sup>C NMR (101 MHz, DMSO-*d*<sub>6</sub>) δ:** 171.28, 156.53, 152.28, 144.25, 141.16, 129.06, 128.67, 128.12, 127.57, 125.75, 120.56, 66.33, 48.53, 47.01, 25.77.

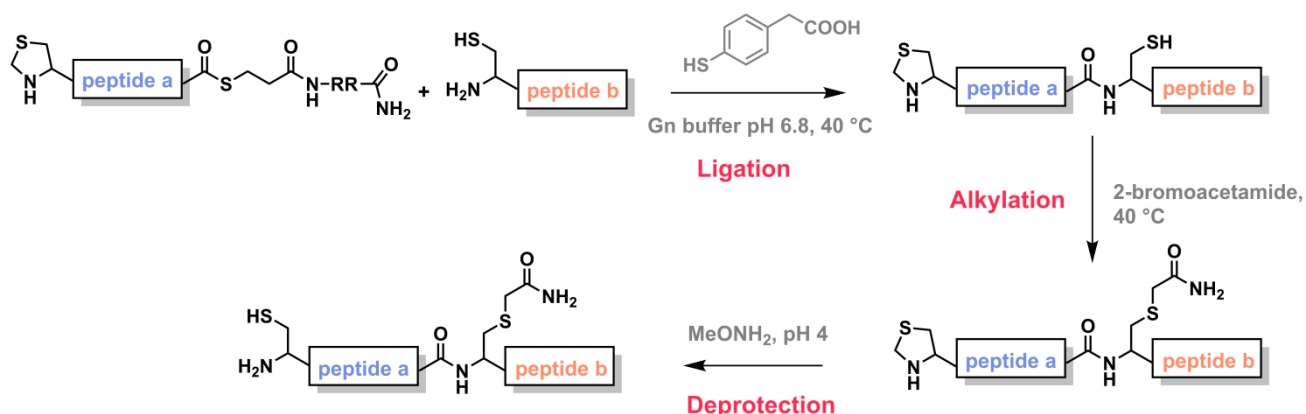
### **Sequential Boc-SPPS to assemble the common C-terminal fragment**

**Sequence:** CALRSELQALRREGFSPEELAALESELQALERELAALRSELQALRG-<sup>α</sup>amide

Sequential assembly of the 46-residue C-terminal peptide: from a 0.4 mmol scale synthesis following the **General procedure V** using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated with 19.4 mg (1 %) of yield based on peptide synthesis scale. LC-MS (ESI): [M + 3 H<sup>+</sup>] = 1708.7 m/z; [M + 4 H<sup>+</sup>] = 1281.9 m/z; [M + 5 H<sup>+</sup>] = 1025.8 m/z; [M + 6 H<sup>+</sup>] = 855.1 m/z; [M + 7 H<sup>+</sup>] = 732.8 m/z; C<sub>219</sub>H<sub>370</sub>N<sub>68</sub>O<sub>71</sub>S; Average isotope calculated: 5123.8 Da [M]; observed: 5123.9 Da.

**Ligation approach to assemble the common C-terminal fragment**

Synthetic route:



Sequence: CALRSELQALRREGFSPEELA<sup>C(capped)</sup>LESELQALERELAALRSELQALRG-<sup>α</sup>amide

Thz-peptide a- $\alpha$ -thioester

Thz-peptide a- $\alpha$ -thioester was synthesized following the **General procedure V** on a 0.4 mmol scale using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated yield based on peptide synthesis scale was 136.3 mg (12 %). LC/MS (ESI):  $[M + 2 H^+] = 1394.0$  m/z;  $[M + 3 H^+] = 930.1$  m/z;  $[M + 4 H^+] = 697.9$  m/z;  $[M + 5 H^+] = 558.7$  m/z; C<sub>117</sub>H<sub>196</sub>N<sub>40</sub>O<sub>35</sub>S<sub>2</sub>; Average isotope calculated: 2787.2 Da [M]; observed: 2787.4 Da.

Cys-peptide b



Cys-peptide b was synthesized following the **General procedure V** on a 0.4 mmol scale using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated yield based on peptide synthesis scale was 87.8 mg (8 %). LC/MS (ESI):  $[M + 2 H^+] = 1399.7$  m/z;  $[M + 3$

$H^+$ ] = 933.8 m/z;  $[M + 4 H^+] = 700.6$  m/z;  $C_{118}H_{205}N_{37}O_{39}S$ ; Average isotope calculated: 2798.2 Da  $[M]$ ; observed: 2798.3 Da.

### Ligation

Cys-peptide **b** (1.5 eq,  $8.94 \cdot 10^{-6}$  mol) was dissolved in ligation aqueous buffer (200 mM  $Na_2HPO_4$ , 6 M Gn-HCl, 20 mM TCEP, 15 mM MPAA, pH 7) to have a peptide concentration of 7.5 mM. The resulting solution was poured onto peptide **a**- $\alpha$ -thioester (1 eq,  $5.96 \cdot 10^{-6}$  mol, 5 mM). The pH was adjusted to 6.8. The reaction mixture was purged with argon and stirred at 40 °C, under Ar, during 4 hours. Completion of the reaction was followed by HPLC and LC/MS. Then, 2-bromoacetamide was introduced in 3-fold excess over total thiols ( $9.84 \cdot 10^{-5}$  mol) in solution at 1 M in neutral aqueous buffer (200 mM  $Na_2HPO_4$ , 6 M Gn-HCl, pH 7, purged with argon). The reaction mixture was stirred at 40 °C, under Ar, during 15 minutes. After verifying completion of the reaction, MPAA ( $2.6 \cdot 10^{-4}$  mol, ~ 200 mM) was added as solid into the reaction mixture to quench the excess of 2-bromoacetamide. The resulting solution was stirred at 40 °C, under Ar, during 30 minutes.  $MeONH_2 \cdot HCl$  ( $7.65 \cdot 10^{-4}$  mol) dissolved in 600  $\mu$ L of neutral aqueous buffer containing 20 mM TCEP was introduced in the reaction mixture to have a final and total concentration of 0.4 M. The pH decreased to ~ 4.0. The reaction solution was purged with argon and stirred at room temperature, under Ar, during 2.5 hours. After completion of the deprotection, the product was directly isolated by HPLC.

Alkylated C-terminal fragment: isolated with 32 % to 43 % of ligation yield (42 mg of isolated peptide in total). LC-MS (ESI):  $[M + 3 H^+] = 1738.5$  m/z;  $[M + 4 H^+] = 1304.3$  m/z;  $[M + 5 H^+] = 1043.7$  m/z;  $[M + 6 H^+] = 869.9$  m/z;  $C_{221}H_{373}N_{69}O_{72}S_2$ ; Average isotope calculated: 5212.9 Da  $[M]$ ; observed: 5213.4 Da.

### Ligation of the different N-terminal fragments with the C-terminal fragment to constitute the library of AA(Ncap)AA(N1)Cys strand

Sequence if Ala residue at the 24<sup>th</sup> position



Sequence if alkylated Cys residue at the 24<sup>th</sup> position

The C-terminal fragment ( $3.9 \cdot 10^{-7}$  mol, 5 mM, 1 eq) was dissolved in 78  $\mu$ L of ligation buffer (6 M Gn-HCl, 200 mM sodium phosphate, 25 mM MPOH, 100 mM TCEP, purged with argon, pH 6.9). The resulting solution was poured onto the N-terminal fragment Ac-Ncap-N1-MeNbz-LDA (or Ac-NH-MPA-LDA or Ac-**GA**-MPA-LRY) ( $4.68 \cdot 10^{-7}$  mol, 6 mM, 1.2 eq). The pH was adjusted to 6.9 and the reaction mixture was purged with argon. The reaction was stirred at 40 °C, for 4 hours, under argon. If branched adduct resulting from transthioesterification between the excess of thioester and the cysteine was observed, pH of the solution was increased to ~ 8.5. Reaction mixture was stirred at 40 °C under argon until completion of the branched thioester adduct hydrolysis.

When reaction was complete, solid-phase extraction (SPE) was performed using C18 cartridge of 6 mL. SPE cartridge was first conditioned with 7 volumes of MeCN + 0.1 % TFA followed by 10 volumes of water + 0.1 % TFA. The reaction mixture was diluted in a total volume of 4 mL of H<sub>2</sub>O + 0.1 % TFA. The resulting solution was loaded to the SPE cartridge. After elution of the loading volume, the cartridge was rinsed with 1 mL of H<sub>2</sub>O + 0.1 % TFA followed by 5 more mL of the same solvent. A first elution step was performed with 3 mL + 1 mL of solution of MeCN/H<sub>2</sub>O [1:9] + 0.1 % TFA. A second step of elution consisted in 5 mL + 5 mL + 2 mL of solution of MeCN/H<sub>2</sub>O [3:7] + 0.1 % TFA in which the MPOH and the spacer sequence with the MeNbz linker went out. The ligated peptide was recovered in a third step of elution with 5 mL + 5 mL + 1 mL of solution of MeCN/H<sub>2</sub>O [5:5] + 0.1 % TFA. The 4<sup>th</sup> step of elution consisted more of a cleaning step with 3 mL + 1 mL of solution of MeCN/H<sub>2</sub>O [9:1] + 0.1 % TFA. The SPE cartridge was finally washed and reconditioned for further utilization with 6 volumes of MeCN + 0.1 % TFA followed by 10 volumes of H<sub>2</sub>O + 0.1 % TFA. The SPE cartridge was stored at room temperature with some H<sub>2</sub>O + 0.1 % TFA to not getting dried. The recovered ligated peptide was finally analyzed by HPLC and LC/MS, flash frozen and lyophilized.

Yields were estimated when the resulting lyophilized peptide (too low to be weighted) was solubilized in buffer to obtain stock solutions for the catalytic assay. Concentration of the resulting solution was estimated with calibration curves on HPLC (calculated with **Het-N2** strands) or with Nanodrop for analogues containing aromatic residues. Amount of peptides and yields were then deduced.

**SRC** strand was isolated after ligation between the Ac-**SR**-MeNbz-LDA fragment and the C-terminal fragment bearing an alkylated cysteine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.4 mg (~ 19 %). LC-MS (ESI): [M + 3 H<sup>+</sup>] = 1833.7 m/z; [M + 4 H<sup>+</sup>] = 1375.3 m/z; [M + 5 H<sup>+</sup>] = 1100.5 m/z; [M + 6 H<sup>+</sup>] = 917.3 m/z; [M + 7 H<sup>+</sup>] = 786.2 m/z; C<sub>232</sub>H<sub>392</sub>N<sub>74</sub>O<sub>76</sub>S<sub>2</sub>; Average isotope calculated: 5498.3 Da [M]; observed: 5497.7 Da.

**RRC** strand was isolated after ligation between the Ac-**RR**-MeNbz-LDA fragment and the C-terminal fragment bearing an alanine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.3 mg (~ 16 %). LC-MS (ESI): [M + 4 H<sup>+</sup>] = 1370.5 m/z; [M + 5 H<sup>+</sup>] = 1096.9 m/z; [M + 6 H<sup>+</sup>] = 914.3 m/z; [M + 7 H<sup>+</sup>] = 783.6 m/z; C<sub>233</sub>H<sub>396</sub>N<sub>76</sub>O<sub>74</sub>S; Average isotope calculated: 5478.2 Da [M]; observed: 5479.4 Da.

**RSC** strand was isolated after ligation between the Ac-**RS**-MeNbz-LDA fragment and the C-terminal fragment bearing an alanine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.5 mg (~ 22 %). LC-MS (ESI): [M + 4 H<sup>+</sup>] = 1353.2 m/z; [M + 5 H<sup>+</sup>] = 1083.0 m/z; [M + 6 H<sup>+</sup>] = 902.6 m/z; [M + 7 H<sup>+</sup>] = 773.8 m/z; C<sub>230</sub>H<sub>389</sub>N<sub>73</sub>O<sub>75</sub>S; Average isotope calculated: 5409.1 Da [M]; observed: 5409.6 Da.

**GRC** strand was isolated after ligation between the Ac-**GR**-MeNbz-LDA fragment and the C-terminal fragment bearing an alanine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.4 mg (~ 21 %). LC-MS (ESI): [M + 4 H<sup>+</sup>] = 1345.4 m/z; [M + 5 H<sup>+</sup>] = 1076.5 m/z; [M + 6 H<sup>+</sup>] = 897.5 m/z; [M + 7 H<sup>+</sup>] = 769.3 m/z; C<sub>229</sub>H<sub>387</sub>N<sub>73</sub>O<sub>74</sub>S; Average isotope calculated: 5379.1 Da [M]; observed: 5378.4 Da.

**RLC** strand was isolated after ligation between the Ac-**RL**-MeNbz-LDA fragment and the C-terminal fragment bearing an alanine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.4 mg (~ 21 %). LC-MS (ESI): [M + 4 H<sup>+</sup>] = 1359.1 m/z; [M + 5 H<sup>+</sup>] = 1087.7 m/z; [M + 6 H<sup>+</sup>] = 906.7 m/z; [M + 7 H<sup>+</sup>] = 777.3 m/z; C<sub>233</sub>H<sub>395</sub>N<sub>73</sub>O<sub>74</sub>S; Average isotope calculated: 5435.2 Da [M]; observed: 5434.1 Da.

**RAC** strand was isolated after ligation between the Ac-**RA**-MeNbz-LDA fragment and the C-terminal fragment bearing an alanine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.4 mg (~ 21 %). LC-MS (ESI): [M + 4 H<sup>+</sup>] = 1348.7 m/z; [M + 5 H<sup>+</sup>] = 1079.4 m/z; [M + 6 H<sup>+</sup>] = 899.6 m/z; [M + 7 H<sup>+</sup>] = 771.3 m/z; C<sub>230</sub>H<sub>389</sub>N<sub>73</sub>O<sub>74</sub>S; Average isotope calculated: 5393.1 Da [M]; observed: 5391.9 Da.

**NRC** strand was isolated after ligation between the Ac-**NR**-MeNbz-LDA fragment and the C-terminal fragment bearing an alanine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.5 mg (~ 24 %). LC-MS (ESI):  $[M + 3 H^+] = 1812.3$  m/z;  $[M + 4 H^+] = 1359.7$  m/z;  $[M + 5 H^+] = 1088.2$  m/z;  $[M + 6 H^+] = 907.1$  m/z;  $[M + 7 H^+] = 777.6$  m/z;  $C_{231}H_{390}N_{74}O_{75}S$ ; Average isotope calculated: 5436.2 Da [M]; observed: 5436.2 Da.

**GAC** strand was isolated after ligation between the Ac-**GA**-MPA-LRY fragment and the C-terminal fragment bearing an alkylated cysteine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.5 mg (~ 25 %). LC-MS (ESI):  $[M + 4 H^+] = 1346.5$  m/z;  $[M + 5 H^+] = 1077.6$  m/z;  $[M + 6 H^+] = 898.2$  m/z;  $C_{228}H_{383}N_{71}O_{75}S_2$ ; Average isotope calculated: 5383.1 Da [M]; observed: 5382.8 Da.

**GIC** strand was isolated after ligation between the Ac-**GI**-MeNbz-LDA fragment and the C-terminal fragment bearing an alanine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.3 mg (~ 14 %). LC-MS (ESI):  $[M + 4 H^+] = 1335.1$  m/z;  $[M + 5 H^+] = 1068.5$  m/z;  $[M + 6 H^+] = 890.6$  m/z;  $C_{229}H_{386}N_{70}O_{74}S$ ; Average isotope calculated: 5336.1 Da [M]; observed: 5337.3 Da.

**GVC** strand was isolated after ligation between the Ac-**GV**-MeNbz-LDA fragment and the C-terminal fragment bearing an alkylated cysteine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.2 mg (~ 10 %). LC-MS (ESI):  $[M + 3 H^+] = 1804.6$  m/z;  $[M + 4 H^+] = 1353.9$  m/z;  $[M + 5 H^+] = 1083.3$  m/z;  $[M + 6 H^+] = 902.8$  m/z;  $C_{230}H_{387}N_{71}O_{75}S_2$ ; Average isotope calculated: 5411.2 Da [M]; observed: 5411.2 Da.

**GFC** strand was isolated after ligation between the Ac-**GF**-MeNbz-LDA fragment and the C-terminal fragment bearing an alanine at the 24<sup>th</sup> position with an estimated yield of ~ 0.4 mg (~ 20 %). LC-MS (ESI):  $[M + 4 H^+] = 1343.5$  m/z;  $[M + 5 H^+] = 1075.1$  m/z;  $[M + 6 H^+] = 896.2$  m/z;  $C_{232}H_{384}N_{70}O_{74}S$ ; Average isotope calculated: 5370.1 Da [M]; observed: 5370.7 Da.

**GYC** strand was isolated after ligation between the Ac-**GY**-MeNbz-LDA fragment and the C-terminal fragment bearing an alkylated cysteine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.4 mg (~ 20 %). LC-MS (ESI):  $[M + 4 H^+] = 1369.7$  m/z;  $[M + 5 H^+] = 1095.9$  m/z;  $[M + 6 H^+] = 913.4$  m/z;  $C_{234}H_{387}N_{71}O_{76}S_2$ ; Average isotope calculated: 5475.2 Da [M]; observed: 5474.4 Da.

**GWC** strand was isolated after ligation between the Ac-**GW**-MeNbz-LDA fragment and the C-terminal fragment bearing an alkylated cysteine at the 24<sup>th</sup> position with an estimated ligation

yield of ~ 0.4 mg (~ 20 %). LC-MS (ESI):  $[M + 4 H^+] = 1375.2$  m/z;  $[M + 5 H^+] = 1100.5$  m/z;  $[M + 6 H^+] = 917.2$  m/z;  $C_{236}H_{388}N_{72}O_{75}S_2$ ; Average isotope calculated: 5498.2 Da [M]; observed: 5497.3 Da.

**GHC** strand was isolated after ligation between the Ac-**GH**-MeNbz-LDA fragment and the C-terminal fragment bearing an alkylated cysteine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.6 mg (~ 27 %). LC-MS (ESI):  $[M + 3 H^+] = 1816.8$  m/z;  $[M + 4 H^+] = 1363.0$  m/z;  $[M + 5 H^+] = 1090.8$  m/z;  $[M + 6 H^+] = 909.2$  m/z;  $[M + 7 H^+] = 779.3$  m/z;  $C_{231}H_{385}N_{73}O_{75}S_2$ ; Average isotope calculated: 5449.2 Da [M]; observed: 5448.7 Da.

**HLC** strand was isolated after ligation between the Ac-**HL**-MeNbz-LDA fragment and the C-terminal fragment bearing an alkylated cysteine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.5 mg (~ 23 %). LC-MS (ESI):  $[M + 3 H^+] = 1836.4$  m/z;  $[M + 4 H^+] = 1377.2$  m/z;  $[M + 5 H^+] = 1102.2$  m/z;  $[M + 6 H^+] = 918.7$  m/z;  $[M + 7 H^+] = 787.4$  m/z;  $C_{235}H_{393}N_{73}O_{75}S_2$ ; Average isotope calculated: 5505.3 Da [M]; observed: 5505.8 Da.

**NHC** strand was isolated after ligation between the Ac-**NH**-MPA-LDA fragment and the C-terminal fragment bearing an alkylated cysteine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.6 mg (~ 29 %). LC-MS (ESI):  $[M + 3 H^+] = 1836.1$  m/z;  $[M + 4 H^+] = 1377.3$  m/z;  $[M + 5 H^+] = 1102.2$  m/z;  $[M + 6 H^+] = 918.7$  m/z;  $[M + 7 H^+] = 787.4$  m/z;  $C_{233}H_{388}N_{74}O_{76}S_2$ ; Average isotope calculated: 5506.2 Da [M]; observed: 5505.6 Da.

**HSC** strand was isolated after ligation between the Ac-**HS**-MeNbz-LDA fragment and the C-terminal fragment bearing an alkylated cysteine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.6 mg (~ 27 %). LC-MS (ESI):  $[M + 3 H^+] = 1826.9$  m/z;  $[M + 4 H^+] = 1370.5$  m/z;  $[M + 5 H^+] = 1096.8$  m/z;  $[M + 6 H^+] = 914.2$  m/z;  $[M + 7 H^+] = 783.5$  m/z;  $C_{232}H_{387}N_{73}O_{76}S_2$ ; Average isotope calculated: 5479.2 Da [M]; observed: 5478.8 Da.

**ELC** strand was isolated after ligation between the Ac-**EL**-MeNbz-LDA fragment and the C-terminal fragment bearing an alkylated cysteine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.4 mg (~ 21 %). LC-MS (ESI):  $[M + 3 H^+] = 1833.5$  m/z;  $[M + 4 H^+] = 1375.8$  m/z;  $[M + 5 H^+] = 1100.6$  m/z;  $[M + 6 H^+] = 917.4$  m/z;  $C_{234}H_{393}N_{71}O_{77}S_2$ ; Average isotope calculated: 5497.3 Da [M]; observed: 5498.1 Da.

**YLC** strand was isolated after ligation between the Ac-**YL**-MeNbz-LDA fragment and the C-terminal fragment bearing an alkylated cysteine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.5 mg (~ 23 %). LC-MS (ESI):  $[M + 3 H^+] = 1845.1$  m/z;  $[M + 4 H^+] = 1383.8$  m/z;



$[M + 5 H^+] = 1107.3 \text{ m/z}$ ;  $[M + 6 H^+] = 923.1 \text{ m/z}$ ;  $C_{238}H_{395}N_{71}O_{76}S_2$ ; Average isotope calculated: 5531.3 Da [M]; observed: 5531.9 Da.

**EYC** strand was isolated after ligation between the Ac-**EY**-MeNbz-LDA fragment and the C-terminal fragment bearing an alkylated cysteine at the 24<sup>th</sup> position with an estimated ligation yield of ~ 0.5 mg (~ 23 %). LC-MS (ESI):  $[M + 3 H^+] = 1849.8 \text{ m/z}$ ;  $[M + 4 H^+] = 1387.7 \text{ m/z}$ ;  $[M + 5 H^+] = 1110.6 \text{ m/z}$ ;  $[M + 6 H^+] = 925.8 \text{ m/z}$ ;  $C_{237}H_{391}N_{71}O_{78}S_2$ ; Average isotope calculated: 5547.3 Da [M]; observed: 5548.2 Da.

### **Synthesis of the library of 'positive' strands**

Starting from on a 0.4 mmol scale using a MBHA resin (1.2 mmol/g loading, 100-200 mesh), residues were assembled following the **General procedure V**. The resin was then split following the **Figure 3.6** to allow synthesis of each strand on a 0.1 mmol scale.

#### *'Positive, G1' strand*

**Sequence:** Ac-GLAALRSELQALRREGFSPERLAALESRLQALERRLAALRSRLQALRG-<sup>α</sup>amide

The 'positive, G1' strand was synthesized on a 0.1 mmol scale; isolated yield was 7.8 mg (2 %) based on peptide synthesis scale. LC/MS (ESI):  $[M + 5 H^+] = 1083.5 \text{ m/z}$ ;  $[M + 6 H^+] = 903.2 \text{ m/z}$ ;  $[M + 7 H^+] = 774.3 \text{ m/z}$ ;  $[M + 8 H^+] = 677.7 \text{ m/z}$ ;  $[M + 9 H^+] = 602.5 \text{ m/z}$ ;  $[M + 10 H^+] = 542.4 \text{ m/z}$ ;  $C_{233}H_{406}N_{82}O_{66}$ ; Average isotope calculated: 5412.3 Da [M]; observed: 5413.5 Da.

#### *'Positive, H1' strand*

**Sequence:** Ac-HLAALRSELQALRREGFSPERLAALESRLQALERRLAALRSRLQALRG-<sup>α</sup>amide

The 'positive, H1' strand was synthesized on a 0.1 mmol scale; isolated yield was 8.4 mg (2 %) based on peptide synthesis scale. LC/MS (ESI):  $[M + 6 H^+] = 916.3 \text{ m/z}$ ;  $[M + 7 H^+] = 785.7 \text{ m/z}$ ;  $[M + 8 H^+] = 687.6 \text{ m/z}$ ;  $[M + 9 H^+] = 611.3 \text{ m/z}$ ;  $[M + 10 H^+] = 550.3 \text{ m/z}$ ;  $C_{237}H_{410}N_{84}O_{66}$ ; Average isotope calculated: 5492.4 Da [M]; observed: 5493.0 Da.

#### *'Positive, H1H2' strand*

Sequence: Ac-HHAALRSELQALRREGFSPERLAALESRLQALERRLAALRSRLQALRG-<sup>α</sup>amide

The 'positive, H1H2' strand was synthesized on a 0.1 mmol scale; isolated yield was 8.9 mg (2 %) based on peptide synthesis scale. LC/MS (ESI): [M + 6 H<sup>+</sup>] = 920.3 m/z; [M + 7 H<sup>+</sup>] = 789.1 m/z; [M + 8 H<sup>+</sup>] = 690.7 m/z; [M + 9 H<sup>+</sup>] = 614.1 m/z; [M + 10 H<sup>+</sup>] = 552.7 m/z; [M + 11 H<sup>+</sup>] = 502.7 m/z; C<sub>237</sub>H<sub>406</sub>N<sub>86</sub>O<sub>66</sub>; Average isotope calculated: 5516.4 Da [M]; observed: 5517.3 Da.

*'Positive, H1E36' strand*

Sequence: Ac-HLAALRSELQALRREGFSPERLAALESRLQALERREAALRSRLQALRG-<sup>α</sup>amide

The 'positive, H1E36' strand was synthesized on a 0.1 mmol scale; isolated yield was 11.4 mg (2 %) based on peptide synthesis scale. LC/MS (ESI): [M + 4 H<sup>+</sup>] = 1377.8 m/z; [M + 5 H<sup>+</sup>] = 1102.5 m/z; [M + 6 H<sup>+</sup>] = 919.1 m/z; [M + 7 H<sup>+</sup>] = 788.0 m/z; [M + 8 H<sup>+</sup>] = 689.7 m/z; [M + 9 H<sup>+</sup>] = 613.2 m/z; [M + 10 H<sup>+</sup>] = 552.0 m/z; C<sub>236</sub>H<sub>406</sub>N<sub>84</sub>O<sub>68</sub>; Average isotope calculated: 5508.4 Da [M]; observed: 5509.2 Da.

### **Catalytic evaluation of the library of 'negative' strands**

Catalytic assays on the library of 'negative' strands were performed following the **General procedure VIII** and using as substrates the peptide-Phe-<sup>α</sup>thioester (Ac-VALENF-<sup>α</sup>Mes, 200 μM) and Tris (200 mM). Catalytic assays were incubated in a PCR thermocycler set at a constant temperature of 25 °C. Kinetic parameters were deduced as described in the procedure and are summarized in **Table 7.3**.

### **Catalytic evaluation of the library of 'positive' strands**

Catalytic assays on the library of 'negative' strands were performed following the **General procedure VIII** and using as substrates the peptide-Phe-<sup>α</sup>thioester (Ac-VALENF-<sup>α</sup>Mes, 200 μM) and Tris (200 mM). Catalytic assays were incubated in a PCR thermocycler set at a constant temperature of 25 °C. Kinetic parameters were deduced as described in the procedure and have been already summarized in **Table 3.5**.

**Table 7.3. Kinetic parameters of the different analogues that compose the library of ‘negative’ strands.** The reaction conditions are set as follow: 100  $\mu\text{M}$  of catalytic heterodimer, 100  $\mu\text{M}$  of peptide-Phe- $\alpha$ thioester and 200 mM of Tris in phosphate buffer pH 7.5 at 25  $^{\circ}\text{C}$ .

Analogues	$k_{\text{obs}}$ ( $\text{M}^{-1}\cdot\text{s}^{-1}$ )	A/H ratio	Acceleration over background
Het-N2 (reference)	0.065	0.67	4.3
Het-SRC	0.05	0.69	3.3
Het-RSC	0.029	0.85	1.9
Het-GHC	0.052	0.41	3.5
Het-RLC	0.048	0.82	3.2
Het-ELC	0.055	1.22	3.7
Het-GAC	0.044	0.63	2.9
Het-RRC	0.025	0.75	1.7
Het-RAC	0.029	0.8	1.9
Het-GRC	0.024	0.91	1.6
Het-NRC	0.027	0.81	1.8
Het-GIC	0.057	0.7	3.8
Het-GVC	0.07	0.55	4.7
Het-GFC	0.056	0.78	3.7
Het-GYC	0.057	1.6	3.8
Het-GWC	0.056	1.22	3.7
Het-HSC	0.029	0.58	1.9
Het-NHC	0.039	0.5	2.6
Het-HLC	0.052	0.74	3.5
Het-YLC	0.035	1.15	2.3
Het-EYC	0.03	1.04	2.0

**Solid-phase peptide synthesis of the library of peptide- $\alpha$ hydrazide substrates for further transformation to thioester**

Sequence: Ac-VALENX- $\alpha$ NHNH<sub>2</sub> (X is one of the 20 canonical amino acids)

The 2-chlorotrityl hydrazide (2-CT-NHNH<sub>2</sub>) resin was freshly prepared following the **General procedure I**. Solid-phase syntheses were performed on a 0.05 or a 0.1 mmol scale using the 2-CT-NHNH<sub>2</sub> resin and following the **General procedure II** for the peptide Ac-

VALENR-<sup>o</sup>NHNH<sub>2</sub> and the **General procedure III** for the other peptides. Due to acceptable purity according to HPLC and LC/MS analysis, thioesterification was directly performed on crude peptides.

The Ac-VALENM-<sup>o</sup>NHNH<sub>2</sub> peptide was cleaved using 92.5 % TFA, 2.5 % DTT, 2.5 % TIPS, 2.5 % H<sub>2</sub>O as cleavage cocktail and under argon.

As deprotection of Ac-VALENW-<sup>o</sup>NHNH<sub>2</sub> peptide was incomplete (*i.e.* by-product with a mass of + 44 Da corresponding to remaining CO<sub>2</sub> adduct), deprotection was finally achieved by incubating the peptide in H<sub>2</sub>O + 0.1 % AcOH during 10 min and then lyophilizing it. This step was repeated twice. The peptide was then used without any intermediate purification for the thioesterification step.

### **Thioesterification directly after cleavage**<sup>25</sup>

The effectiveness of this procedure was first tested with Ac-VALENF-<sup>o</sup>NHNH<sub>2</sub> as model peptide (on 50 mg of resin, around 0.016 mmol of peptide). Oxidation step was here performed directly in the cleavage mixture, thus special conditions were necessary for the cleavage to be compatible with next step. Here, cleavage cocktail was the following: 80 % TFA, 2.5 % TIPS, 2.5 % H<sub>2</sub>O, 5 % *m*-cresol, 10 % thioanisol (2 mL of solution for 50 mg of dried resin). After 2 h of reaction, the temperature was decreased to -15 °C and NaNO<sub>2</sub> was added in a one-pot manner (10 eq dissolved in 110 µL of water). The reaction mixture was stirred for 20 min. The crude peptide was precipitated with chilled Et<sub>2</sub>O (-20 °C) and centrifugated (3 min, 4500 RPM) two times. The supernatant was discarded and the peptide was then dried. MesNa (50 eq) was dissolved in the pH 7 neutral buffer (6 M Gn-HCl, 200 mM sodium phosphate). pH was adjusted to 7.8 and the buffer was added to the peptide pellet. pH was measured and adjusted if necessary to be at 7.5. The reaction was stirred under argon for 30 min at -15 °C. The thioester was directly purified by preparative HPLC and pure fractions were lyophilized.

Ac-VALENF-<sup>o</sup>Mes was isolated with 2.0 mg (15 %) of overall yield. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 858.2 m/z; C<sub>36</sub>H<sub>55</sub>N<sub>7</sub>O<sub>13</sub>S<sub>2</sub>; Average isotope calculated: 858.0 Da [M]; observed: 857.2 Da. However, even after purification, fractions that contain the desired peptide were still in presence of the side product of the Curtius rearrangement (+ 14.6 Da observed in the LC/MS spectrum). Complete separation was impossible to achieve due to co-elution.

**Thioesterification of peptide- $\alpha$ hydrazide crudes after cleavage and lyophilization to obtain the library of peptide- $\alpha$ Mes substrates**

Finally, thioesterification of all peptide surrogates was performed following the standard **General procedure VI** starting with 10 mg of crude peptide.

Ac-VALENA- $\alpha$ Mes peptide was isolated with 5.5 mg (46 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 782.0$  m/z;  $C_{30}H_{51}N_7O_{13}S_2$ ; Average isotope calculated: 781.9 Da [M]; observed: 781.0 Da.

Ac-VALENC- $\alpha$ Mes peptide was isolated with 3.0 mg (26 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 814.3$  m/z;  $C_{30}H_{51}N_7O_{13}S_3$ ; Average isotope calculated: 813.9 Da [M]; observed: 813.4 Da.

Ac-VALENE- $\alpha$ Mes peptide was isolated with 3.0 mg (26 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 840.3$  m/z;  $C_{32}H_{53}N_7O_{15}S_2$ ; Average isotope calculated: 839.9 Da [M]; observed: 839.3 Da.

Ac-VALENF- $\alpha$ Mes peptide was isolated with 2.5 mg (21 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 858.2$  m/z;  $C_{36}H_{55}N_7O_{13}S_2$ ; Average isotope calculated: 858.0 Da [M]; observed: 857.2 Da.

Ac-VALENG- $\alpha$ Mes peptide was isolated with 3.5 mg (31 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 768.1$  m/z;  $C_{29}H_{49}N_7O_{13}S_2$ ; Average isotope calculated: 767.9 Da [M]; observed: 767.1 Da.

Ac-VALENH- $\alpha$ Mes peptide was isolated with 4.8 mg (42 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 848.3$  m/z;  $C_{33}H_{53}N_9O_{13}S_2$ ; Average isotope calculated: 847.9 Da [M]; observed: 847.3 Da.

Ac-VALENI- $\alpha$ Mes peptide was isolated with 1.9 mg (16 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 824.2$  m/z;  $C_{33}H_{57}N_7O_{13}S_2$ ; Average isotope calculated: 824.0 Da [M]; observed: 823.2 Da.

Ac-VALENK- $\alpha$ Mes peptide was isolated with 5.7 mg (49 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 839.4$  m/z;  $C_{33}H_{58}N_8O_{13}S_2$ ; Average isotope calculated: 839.0 Da [M]; observed: 838.4 Da.

Ac-VALENL-<sup>o</sup>Mes peptide was isolated with 3.3 mg (28 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 824.0$  m/z;  $C_{33}H_{57}N_7O_{13}S_2$ ; Average isotope calculated: 824.0 Da [M]; observed: 823.0 Da.

Ac-VALENM(ox)-<sup>o</sup>Mes peptide was isolated in the oxidized form of the methionine with 2 mg (17 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 858.4$  m/z;  $C_{32}H_{55}N_7O_{14}S_3$ ; Average isotope calculated: 858.0 Da [M]; observed: 857.4 Da.

Ac-VALENN-<sup>o</sup>Mes peptide was isolated with 5.1 mg (47 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 825.1$  m/z;  $C_{31}H_{52}N_8O_{14}S_2$ ; Average isotope calculated: 824.9 Da [M]; observed: 824.1 Da.

Ac-VALENP-<sup>o</sup>Mes peptide was isolated with 3.0 mg (27 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 808.3$  m/z;  $C_{32}H_{53}N_7O_{13}S_2$ ; Average isotope calculated: 807.9 Da [M]; observed: 807.3 Da.

Ac-VALENQ-<sup>o</sup>Mes peptide was isolated with 1.6 mg (13 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 839.2$  m/z;  $C_{32}H_{54}N_8O_{14}S_2$ ; Average isotope calculated: 838.9 Da [M]; observed: 838.2 Da.

Ac-VALENR-<sup>o</sup>Mes peptide was isolated with 4.2 mg (37 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 867.3$  m/z;  $C_{33}H_{58}N_{10}O_{13}S_2$ ; Average isotope calculated: 867.0 Da [M]; observed: 866.3 Da.

Ac-VALENS-<sup>o</sup>Mes peptide was isolated with 3.8 mg (31 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 798.3$  m/z;  $C_{30}H_{51}N_7O_{14}S_2$ ; Average isotope calculated: 797.9 Da [M]; observed: 797.3 Da.

Ac-VALENT-<sup>o</sup>Mes peptide was isolated with 2.1 mg (18 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 812.3$  m/z;  $C_{31}H_{53}N_7O_{14}S_2$ ; Average isotope calculated: 811.9 Da [M]; observed: 811.3 Da.

Ac-VALENV-<sup>o</sup>Mes peptide was isolated with 5.4 mg (48 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 810.2$  m/z;  $C_{32}H_{55}N_7O_{13}S_2$ ; Average isotope calculated: 809.9 Da [M]; observed: 809.2 Da.

Ac-VALENW- $\alpha$ Mes peptide was isolated with 1.2 mg (10 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 897.4$  m/z;  $C_{38}H_{56}N_8O_{13}S_2$ ; Average isotope calculated: 897.0 Da [M]; observed: 896.4 Da.

Ac-VALENY- $\alpha$ Mes peptide was isolated with 5.4 mg (47 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 874.2$  m/z;  $C_{36}H_{55}N_7O_{14}S_2$ ; Average isotope calculated: 874.0 Da [M]; observed: 873.2 Da.

Ac-VALEND- $\alpha$ Mes. Special precautions were taken for the formation of the peptide Ac-VALEND- $\alpha$ Mes due to the propensity of the Ac-VALEND- $\alpha$ azide intermediate to hydrolyze. By decreasing the time of the oxidation step to 1 minute and therefore by directly adding the MesNa, only half of the acyl azide was hydrolyzed and the other half of desired Ac-VALEND- $\alpha$ Mes was then isolated by preparative HPLC.

Ac-VALEND- $\alpha$ Mes peptide was isolated with 1.8 mg (16 %) of thioesterification yield. LC/MS (ESI):  $[M + 1 H^+] = 825.9$  m/z;  $C_{31}H_{51}N_7O_{15}S_2$ ; Average isotope calculated: 825.9 Da [M]; observed: 824.9 Da.

### **Synthesis of the library of acyl acceptor substrates**

Sequence:  $H_2N$ -XVALEN- $\alpha$ amide (with X the 20 different canonical amino acids)

The different  $H_2N$ -XVALEN- $\alpha$ amide peptides were prepared on a 0.1 mmol scale synthesis following **General procedure III** on a Rink amide resin (0.74 mmol/g loading, 100-200 mesh). The crudes of peptide were purified by reverse-phase C18 HPLC.

$H_2N$ -AVALEN- $\alpha$ amide peptide was isolated with yield of 42 mg (69 %) based on peptide synthesis scale. LC/MS (ESI):  $[M + 1 H^+] = 616.3$  m/z;  $C_{26}H_{46}N_8O_9$ ; Average isotope calculated: 614.7 Da [M]; observed: 615.3 Da.

$H_2N$ -CVALEN- $\alpha$ amide peptide was isolated with yield of 37 mg (57 %) based on peptide synthesis scale. LC/MS (ESI):  $[M + 1 H^+] = 647.3$  m/z;  $C_{26}H_{46}N_8O_9S$ ; Average isotope calculated: 646.8 Da [M]; observed: 646.3 Da.

$H_2N$ -DVALEN- $\alpha$ amide peptide was isolated with yield of 42 mg (63 %) based on peptide synthesis scale. LC/MS (ESI):  $[M + 1 H^+] = 659.4$  m/z;  $C_{27}H_{46}N_8O_{11}$ ; Average isotope calculated: 658.7 Da [M]; observed: 658.4 Da.

H<sub>2</sub>N-EVALEN- $\alpha$ amide peptide was isolated with yield of 40 mg (59 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 673.3 m/z; C<sub>28</sub>H<sub>48</sub>N<sub>8</sub>O<sub>11</sub>; Average isotope calculated: 672.7 Da [M]; observed: 672.3 Da.

H<sub>2</sub>N-FVALEN- $\alpha$ amide peptide was isolated with yield of 47 mg (68 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 691.4 m/z; C<sub>32</sub>H<sub>50</sub>N<sub>8</sub>O<sub>9</sub>; Average isotope calculated: 690.8 Da [M]; observed: 690.4 Da.

H<sub>2</sub>N-GVALEN- $\alpha$ amide peptide was isolated with yield of 39 mg (65 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 601.3 m/z; C<sub>25</sub>H<sub>44</sub>N<sub>8</sub>O<sub>9</sub>; Average isotope calculated: 600.7 Da [M]; observed: 600.3 Da.

H<sub>2</sub>N-HVALEN- $\alpha$ amide peptide was isolated with yield of 57 mg (84 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 681.4 m/z; [M + 2 H<sup>+</sup>] = 341.3 m/z; C<sub>29</sub>H<sub>48</sub>N<sub>10</sub>O<sub>9</sub>; Average isotope calculated: 680.8 Da [M]; observed: 680.4 Da.

H<sub>2</sub>N-IVALEN- $\alpha$ amide peptide was isolated with yield of 22 mg (33 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 657.4 m/z; C<sub>29</sub>H<sub>52</sub>N<sub>8</sub>O<sub>9</sub>; Average isotope calculated: 656.8 Da [M]; observed: 656.4 Da.

H<sub>2</sub>N-KVALEN- $\alpha$ amide peptide was isolated with yield of 33 mg (49 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 672.5 m/z; [M + 2 H<sup>+</sup>] = 336.9 m/z; C<sub>29</sub>H<sub>53</sub>N<sub>9</sub>O<sub>9</sub>; Average isotope calculated: 671.5 Da [M]; observed: 671.6 Da.

H<sub>2</sub>N-LVALEN- $\alpha$ amide peptide was isolated with yield of 52 mg (79 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 657.4 m/z; C<sub>29</sub>H<sub>52</sub>N<sub>8</sub>O<sub>9</sub>; Average isotope calculated: 656.8 Da [M]; observed: 656.4 Da.

H<sub>2</sub>N-NVALEN- $\alpha$ amide peptide was isolated with yield of 41 mg (62 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 658.3 m/z; C<sub>27</sub>H<sub>47</sub>N<sub>9</sub>O<sub>10</sub>; Average isotope calculated: 657.7 Da [M]; observed: 657.3 Da.

H<sub>2</sub>N-PVALEN- $\alpha$ amide peptide was isolated with yield of 36 mg (56 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 641.4 m/z; C<sub>28</sub>H<sub>48</sub>N<sub>8</sub>O<sub>9</sub>; Average isotope calculated: 640.7 Da [M]; observed: 640.4 Da.



H<sub>2</sub>N-**Q**VALEN- $\alpha$ amide peptide was isolated with yield of 48 mg (71 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 672.4 m/z; C<sub>28</sub>H<sub>49</sub>N<sub>9</sub>O<sub>10</sub>; Average isotope calculated: 671.8 Da [M]; observed: 671.4 Da.

H<sub>2</sub>N-**R**VALEN- $\alpha$ amide peptide was isolated with yield of 51 mg (73 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 700.6 m/z; [M + 2 H<sup>+</sup>] = 350.7 m/z; C<sub>29</sub>H<sub>53</sub>N<sub>11</sub>O<sub>9</sub>; Average isotope calculated: 699.8 Da [M]; observed: 699.6 Da.

H<sub>2</sub>N-**S**VALEN- $\alpha$ amide peptide was isolated with yield of 38 mg (60 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 631.3 m/z; C<sub>26</sub>H<sub>46</sub>N<sub>8</sub>O<sub>10</sub>; Average isotope calculated: 630.7 Da [M]; observed: 630.3 Da.

H<sub>2</sub>N-**T**VALEN- $\alpha$ amide peptide was isolated with yield of 42 mg (65 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 645.3 m/z; C<sub>27</sub>H<sub>48</sub>N<sub>8</sub>O<sub>10</sub>; Average isotope calculated: 644.7 Da [M]; observed: 644.3 Da.

H<sub>2</sub>N-**V**VALEN- $\alpha$ amide peptide was isolated with yield of 35 mg (54 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 643.4 m/z; C<sub>28</sub>H<sub>50</sub>N<sub>8</sub>O<sub>9</sub>; Average isotope calculated: 642.8 Da [M]; observed: 642.4 Da.

H<sub>2</sub>N-**Y**VALEN- $\alpha$ amide peptide was isolated with yield of 17 mg (24 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 707.5 m/z; C<sub>32</sub>H<sub>50</sub>N<sub>8</sub>O<sub>10</sub>; Average isotope calculated: 706.8 Da [M]; observed: 706.5 Da.

The H<sub>2</sub>N-**M**VALEN- $\alpha$ amide peptide was cleaved under argon using 92.5 % TFA, 2.5 % DTT, 2.5 % TIPS, 2.5 % H<sub>2</sub>O as cleavage cocktail.

H<sub>2</sub>N-**M**VALEN- $\alpha$ amide peptide was isolated with yield of 34 mg (50 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 675.4 m/z; C<sub>28</sub>H<sub>50</sub>N<sub>8</sub>O<sub>9</sub>S; Average isotope calculated: 674.8 Da [M]; observed: 674.4 Da.

Cleavage of Ac-VALEN**W**- $\alpha$ NHNH<sub>2</sub> peptide was let for 4 hours. But deprotection of tryptophan residue was incomplete (*i.e.* by-product with a mass of + 44 Da corresponding to remaining CO<sub>2</sub> adduct), deprotection was finally achieved by incubating the peptide in H<sub>2</sub>O + 0.1 % AcOH during 10 min and then lyophilizing it. This step was repeated twice.

H<sub>2</sub>N-**W**VALEN- $\alpha$ amide peptide was isolated with yield of 11 mg (15 %) based on peptide synthesis scale. LC/MS (ESI): [M + 1 H<sup>+</sup>] = 730.4 m/z; C<sub>34</sub>H<sub>51</sub>N<sub>9</sub>O<sub>9</sub>; Average isotope calculated: 729.8 Da [M]; observed: 729.4 Da.

## 7.4. Chapter 4

### Merge two structures to form the extended scaffold

After moving down one monomer of the **Hom-N2** crystal structure, the resulting structure was aligned with the coiled-coil model generated by the CCCP software according to Crick parameters from **Hom-N2**. The terminations of each chain were defined (shown as 'start\_resnum' and 'end\_resnum' in the script options below) To merge both structures by the specified terminations and so to combine the two PDB files forming one unique PDB and structure, a python script 'pdb\_combine' (written by Nathan Schmidt from DeGrado lab) was employed.

To obtain the 'strand +14' the following options were used:

```
pdbs = ['polyG_fit_chA', 'xtal_struct', 'polyG_fit_chA2']
chains = ['A', 'A', 'B']
start_resnum = [21, 1, 31]
end_resnum = [27, 44, 41]
combined_pdb = 'combined_chA'
PC = pdb_combine(pdb, chains, start_resnum, end_resnum, combined_pdb)
```

And the resulting sequence was the following:

GGGGGGGGLCALRSELQALRREGFSPEELAALESELQALERRLAALRSRLQGGGGGGGGGGG

To obtain the 'strand +7' the following options were used:

```
pdbs = ['xtal_struct', 'polyG_fit2']
chains = ['B', 'C']
start_resnum = [1, 30]
end_resnum = [43, 41]
combined_pdb = 'combined_chB'
PC = pdb_combine(pdb, chains, start_resnum, end_resnum, combined_pdb)
```

And the resulting sequence was the following:

GLCALRSELQALRREGFSPEELAALESELQALERRLAALRSRLGGGGGGGGGGG

The PDB files of the 'strand +14' ('combined\_chA') and of the 'strand +7' ('combined\_chB') were merged manually to obtain the model of the extended heterodimer.

**MASTER search**

The software was downloaded from Grigoryan laboratory website.<sup>26</sup> The PDB file of the short fragment of our designed protein was first converted to PDS using the following command line:

```
for k in {1..1}; do ../createPDS --type query --pdb fragment0${k}.pdb --opdb tm${k}.pdb --pds
fragment0${k}.pds;done
```

Similarity of structure to our short peptide fragment was then searched within a PDB database collected by Marco Mravic from DeGrado laboratory. A RMSD cutoff of 1.5 Å was applied and the best 99 first matches were selected. The following command line was used:

```
for k in {1..1}; do ../master --query fragment0${k}.pds --targetList ../v2_162901_bc_30-
scPDB/list.txt --topN 99 --rmsdCut 1.5 --bbRMSD --matchOut fragment0${k}_match.m --
seqOut fragment0${k}_seq.s --structOut matches0${k};done
```

If the loop was not included in the search, 99 similar fragments were found in the following PDBs: 4HR1, 1QGI, 1HX1, 2R17, 3HTK, 3IEE, 3KMI, 3AGT, 4UY3, 1HX1, 2EHW, 4DYL, 4WPE, 4R0Y, 4EGW, 1GQE, 2O8P, 1GS9, 2YN7, 3C18, 4HR1, 1BZ4, 4HWI, 4RPE, 4U0R, 4HWC, 3A8Y, 3IEE, 4WAT, 3NVO, 1W07, 1NFN, 5C1F, 4HWI, 2EFK, 1V7M, 3CKC, 1NIG, 4WPE, 3M6Z, 4J2C, 2CSB, 1YAR, 2GSC, 3NFT, 2C5K, 3SWH, 3NZZ, 1JNR, 4PQZ, 4OO2, 3C64, 1BG1, 2RLD, 4XPX, 1UUR, 4DYL, 3QWE, 3JYS, 4I0U, 4GQ2, 2R25, 4XWJ, 4RGP, 3NFT, 3MCX, 2R17, 4Q25, 2YYK, 4DYL, 4W8J, 2XHE, 3N2O, 1BG1, 3JQ0, 2OKU, 3M7G, 1XZP, 2YYK, 4V3D, 4H5Y, 3QNK, 4ONR, 3SGH, 1U5P; 1V7M, 4W8J, 3KBT, 1ZRO, 3GN4, 3EZU, 1SUM, 3NFT, 1BG1, 4AKK, 2RBD, 1XIO, 1GQE, 4PAC.

If we took into account the presence of a loop, 5 similarities were found in the following PDB structures: 1NIG, 2GSC, 2RLD, 2FEF, 3KYG.

**Optimization of the sequence and the hydrophobic packing**

Starting from the following sequences, optimization of the residue composition were performed using the fixbb script of Rosetta software.

**Sequence to optimize:**

'Strand +14':

GLCALRSELAALRSELQALRREGFSPEELAALESELQALESELQALERRLAALRSRLQALRG

'Strand +7':

GLAALRSELQALRREGFSPEELAALESELQALERRLAALRSRLAALRSRLQALRG

The resulting structure was first relaxed with Rosetta to accommodate the new heptad repeats. The following command line was used:

```
../rosetta_bin_linux_2017.18.59451_bundle/main/source/bin/relax.static.linuxgccrelease -s
database ~/rosetta_bin_linux_2017.18.59451_bundle/main/database/ -s
extendHet+mutations.pdb -out:prefix relax_
```

The resulting relaxed structure superimposed with the input structure with a RMSD of 0.847 Å.

Specified positions of amino acids that were previously defined in a resfile were allowed to be changed (according to the instructions of the resfile) meanwhile the backbone stayed fixed. An extra sampling level was added to allow extra chi rotamers of the side chains with  $\pm 2$  standard deviation (in the command line: `-ex1:level 3 -ex2:level 3`). The following command line was used to generate 100 structures:

```
../rosetta_bin_linux_2017.18.59451_bundle/main/source/bin/fixbb.static.linuxgccrelease -s
database ../rosetta_bin_linux_2017.18.59451_bundle/main/database/ -s
relax_extendHet+mutations.pdb -nstruct 100 -resfile resfile.txt -ex1:level 3 -ex2:level 3
```

And the instructions of modifications were indicated as follows in a resfile:

```
USE_INPUT_SC
NATRO
START
```

```
2 A PIKAA YLVAWFI
5 A PIKAA YLVAWFI
7 A POLAR NOTAA KMC
9 A PIKAA YLVAWFI
12 A PIKAA YLVAWFI
14 A POLAR NOTAA KMC
16 A PIKAA YLVAWFI
19 A PIKAA YLVAWFI
24 A PIKAA YLVAWFI
29 A PIKAA YLVAWFI
32 A PIKAA YLVAWFI
34 A POLAR NOTAA KMC
36 A PIKAA YLVAWFI
39 A PIKAA YLVAWFI
41 A POLAR NOTAA KMC
43 A PIKAA YLVAWFI
46 A PIKAA YLVAWFI
```

48 A POLAR NOTAA KMC  
 50 A PIKAA YLVAWFI  
 53 A PIKAA YLVAWFI  
 57 A PIKAA YLVAWFI  
 60 A PIKAA YLVAWFI  
 3 B PIKAA YLVAWFI  
 6 B PIKAA YLVAWFI  
 10 B PIKAA YLVAWFI  
 13 B PIKAA YLVAWFI  
 18 B PIKAA YLVAWFI  
 23 B PIKAA YLVAWFI  
 26 B PIKAA YLVAWFI  
 30 B PIKAA YLVAWFI  
 33 B PIKAA YLVAWFI  
 35 B POLAR NOTAA KMC  
 37 B PIKAA YLVAWFI  
 40 B PIKAA YLVAWFI  
 42 B POLAR NOTAA KMC  
 44 B PIKAA YLVAWFI  
 47 B PIKAA YLVAWFI  
 49 B POLAR NOTAA KMC  
 51 B PIKAA YLVAWFI  
 54 B PIKAA YLVAWFI

Over the 100 generated structures, the consensus sequence was the following (modifications found in the hydrophobic core are depicted in blue, in the loop in purple and for polar residues in orange):

'Strand +14, GAC'

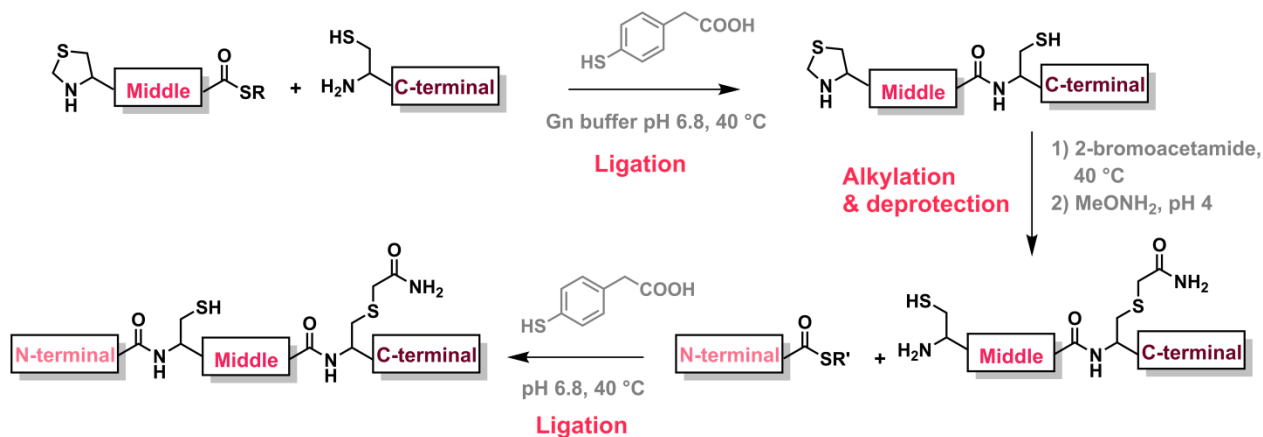
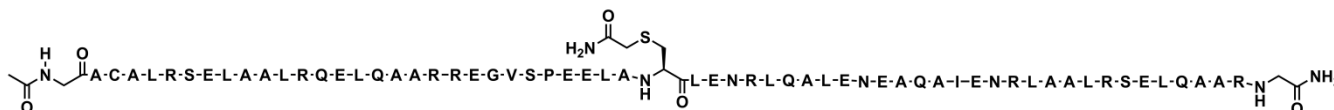
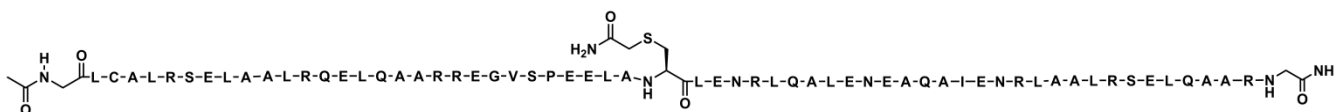
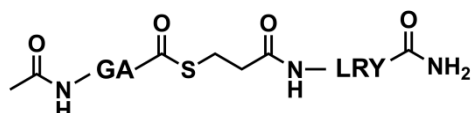
Ac-G**A**CALRSELAALR**Q**ELQA**A**RREG**V**SPEELAAL**E**NRLQALE**N**E**A**QAI**E**NRLAALRSELQA**A**RG

'Strand +7'

Ac- GA**A**AALRSE**I**QALRREG**A**SPERLAALESELQAI**E**NR**A**AALRN**E**LAALRSRLQALRG

**Synthesis of 'strand +14, GAC' and 'strand +14 GLC'**

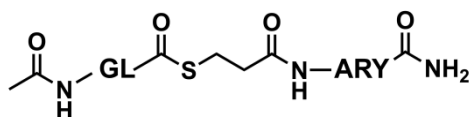
Synthetic route:

**Sequence of 'strand +14, GAC':**Ac-GACALRSELAALRQELQAARREGVSPEELAC(capped)LENRLQALENEAQAIENRLAALRSELQAARG-<sup>α</sup>amide**Sequence of 'strand +14, GLC':**Ac-GLCALRSELAALRQELQAARREGVSPEELAC(capped)LENRLQALENEAQAIENRLAALRSELQAARG-<sup>α</sup>amide*N-terminal peptide fragment for the 'strand +14, GAC,*

Ac-GA-MPA-LRY peptide was synthesized following the **General procedure V** on a 0.2 mmol scale using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated yield was

113.8 mg (78 %) based on peptide synthesis scale. LC/MS (ESI):  $[M + 1 H^+] = 708.6$  m/z;  $C_{31}H_{49}N_9O_8S$ ; Average isotope calculated: 707.8 Da [M]; observed: 707.6 Da.

*N-terminal peptide fragment for the 'strand +14, GLC'*



Ac-GL-MPA-ARY peptide was synthesized following the **General procedure V** on a 0.4 mmol scale using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated yield was 100 mg (35 %) based on peptide synthesis scale. LC/MS (ESI):  $[M + 1 H^+] = 708.3$  m/z;  $C_{31}H_{49}N_9O_8S$ ; Average isotope calculated: 707.8 Da [M]; observed: 707.3 Da.

*Middle peptide fragment*



The middle peptide fragment was synthesized following the **General procedure V** on a 0.2 mmol scale using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated yield was 154 mg (22 %) based on peptide synthesis scale. LC/MS (ESI):  $[M + 3 H^+] = 1160.9$  m/z;  $[M + 4 H^+] = 870.9$  m/z;  $[M + 5 H^+] = 697.0$  m/z;  $[M + 6 H^+] = 581.0$  m/z;  $C_{144}H_{249}N_{51}O_{45}S_2$ ; Average isotope calculated: 3479.0 Da [M]; observed: 3479.8 Da.

*C-terminal peptide fragment*



The C-terminal peptide fragment was synthesized following the **General procedure IV** (automated and microwave-assisted Fmoc-SPPS) on a 0.1 mmol scale using a Rink amide resin (0.74 mmol/g loading, 100-200 mesh); isolated yield was 12.6 mg (4 %) based on peptide synthesis scale. LC/MS (ESI):  $[M + 3 H^+] = 1184.9$  m/z;  $[M + 4 H^+] = 888.9$  m/z;  $[M + 5 H^+] = 711.2$  m/z;  $C_{147}H_{253}N_{51}O_{49}S$ ; Average isotope calculated: 3551.0 Da [M]; observed: 3551.6 Da.

*One-pot ligation*

The C-terminal peptide fragment ( $7.04 \cdot 10^{-6}$  mol, 1.5 eq) was dissolved in ligation aqueous buffer (200 mM  $\text{Na}_2\text{HPO}_4$ , 6 M  $\text{Gn}\cdot\text{HCl}$ , 20 mM TCEP, 15 mM MPAA, pH 7) to have a peptide concentration of 7.5 mM. The resulting solution was poured onto the middle fragment ( $4.69 \cdot 10^{-6}$  mol, 1 eq, 5 mM). The pH was adjusted to 6.8. The reaction mixture was purged with argon and stirred at 40 °C, under Ar, during 3 hours. Completion of the reaction was followed by HPLC and LC/MS. Then, 2-bromoacetamide ( $7.72 \cdot 10^{-5}$  mol) was introduced in 3-fold excess over total thiols in solution at 1 M in neutral aqueous buffer (200 mM  $\text{Na}_2\text{HPO}_4$ , 6 M  $\text{Gn}\cdot\text{HCl}$ , pH 7, purged with argon). The reaction mixture was stirred at 40 °C, under Ar, during 20 minutes. After verifying completion of the reaction, MPAA (final concentration ~ 200 mM) was added as solid into the reaction mixture to quench the excess of 2-bromoacetamide. The resulting solution was stirred at 40 °C, under Ar, during 30 minutes.  $\text{MeONH}_2\cdot\text{HCl}$  ( $5.68 \cdot 10^{-4}$  mol) dissolved in 400  $\mu\text{L}$  of neutral aqueous buffer containing 20 mM TCEP was introduced in the reaction mixture to have a final and total concentration of 0.4 M of methoxyamine. The pH decreased to ~ 4.0. The reaction solution was purged with argon and stirred at room temperature, under Ar, during 2.5 hours. Finally, one the N-terminal fragments ( $9.38 \cdot 10^{-6}$  mol, 2 eq) was dissolved in 300  $\mu\text{L}$  of ligation aqueous buffer and added to reaction mixture. The pH was adjusted to 6.8 and the solution was purged with argon. Reaction mixture was stirred at 40 °C, under Ar, during 1 hour. After completion of the last ligation, the product was directly isolated by HPLC.

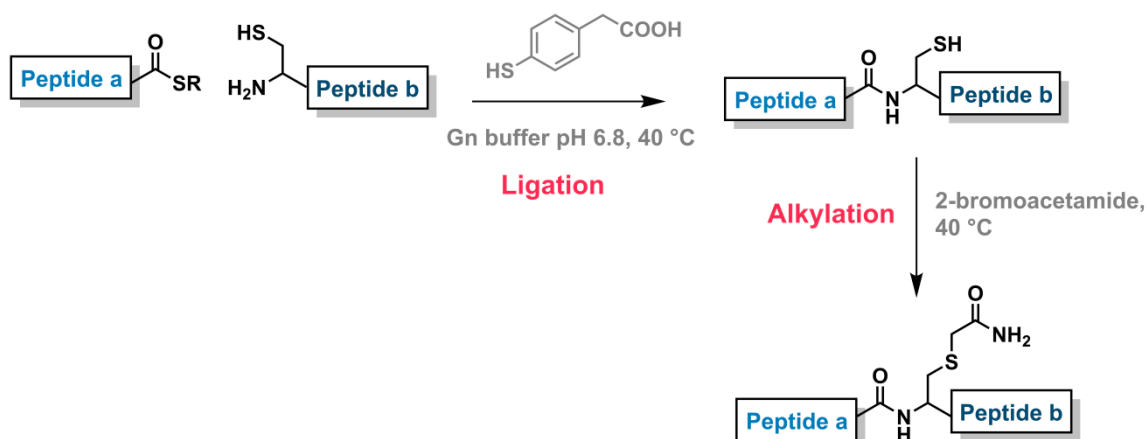
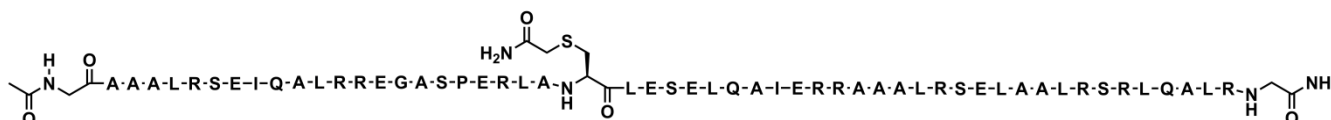
**‘Strand +14, GAC’:** isolated with 8.0 mg (28 %) of ligation yield. LC-MS (ESI):  $[\text{M} + 5 \text{H}^+] = 1366.9 \text{ m/z}$ ;  $[\text{M} + 6 \text{H}^+] = 1139.4 \text{ m/z}$ ;  $[\text{M} + 7 \text{H}^+] = 976.8 \text{ m/z}$ ;  $[\text{M} + 8 \text{H}^+] = 854.9 \text{ m/z}$ ;  $\text{C}_{284}\text{H}_{484}\text{N}_{96}\text{O}_{95}\text{S}_2$ ; Average isotope calculated: 6827.7 Da [M]; observed: 6830.4 Da.

**‘Strand +14, GLC’:** isolated with 9.7 mg (30 %) of ligation yield. LC-MS (ESI):  $[\text{M} + 4 \text{H}^+] = 1718.5 \text{ m/z}$ ;  $[\text{M} + 5 \text{H}^+] = 1374.6 \text{ m/z}$ ;  $[\text{M} + 6 \text{H}^+] = 1145.7 \text{ m/z}$ ;  $[\text{M} + 7 \text{H}^+] = 982.3 \text{ m/z}$ ;  $[\text{M} + 8 \text{H}^+] = 859.7 \text{ m/z}$ ;  $\text{C}_{287}\text{H}_{490}\text{N}_{96}\text{O}_{95}\text{S}_2$ ; Average isotope calculated: 6869.8 Da [M]; observed: 6869.6 Da.



**Synthesis of the 'strand +7'**

Synthetic route:

Sequence:Ac- GAAALRSEIQALRREGASPERLAC(capped)LESELQAIERRAAALRSELAALRSRLQALRG-<sup>d</sup>amide*Peptide a*

The peptide **a** was synthesized following the **General procedure V** on a 0.4 mmol scale using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated yield was 296 mg (26 %) based on peptide synthesis scale. LC/MS (ESI):  $[M + 3 H^+] = 955.9$  m/z;  $[M + 4 H^+] = 717.3$  m/z;  $[M + 5 H^+] = 574.1$  m/z;  $[M + 6 H^+] = 478.6$  m/z;  $C_{118}H_{207}N_{45}O_{36}S$ ; Average isotope calculated: 2864.3 Da  $[M]$ ; observed: 2865.3 Da.

*Peptide b*

The peptide **b** was synthesized following the **General procedure V** on a 0.2 mmol scale using a MBHA resin (1.2 mmol/g loading, 100-200 mesh); isolated yield was 55.8 mg

(8 %) based on peptide synthesis scale. LC/MS (ESI):  $[M + 3 H^+] = 1184.9$  m/z;  $[M + 4 H^+] = 889.1$  m/z;  $[M + 5 H^+] = 711.6$  m/z;  $[M + 6 H^+] = 593.3$  m/z;  $[M + 7 H^+] = 508.7$  m/z;  $C_{149}H_{265}N_{53}O_{45}S$ ; Average isotope calculated: 3551.1 Da [M]; observed: 3553.0 Da.

### Ligation

The peptide **b** (1.5 eq,  $7.04 \cdot 10^{-6}$  mol) was dissolved in ligation aqueous buffer (200 mM  $Na_2HPO_4$ , 6 M Gn·HCl, 20 mM TCEP, 15 mM MPAA, pH 7) to have a peptide concentration of 7.5 mM. The resulting solution was poured onto the middle fragment (1 eq,  $4.69 \cdot 10^{-6}$  mol, 5 mM). The pH was adjusted to 6.8. The reaction mixture was purged with argon and stirred at 40 °C, under Ar, during 3.5 hours. Completion of the reaction was followed by HPLC and LC/MS. Then, 2-bromoacetamide was introduced in 3-fold excess over total thiols ( $7.72 \cdot 10^{-5}$  mol) in solution at 1 M in neutral aqueous buffer (200 mM  $Na_2HPO_4$ , 6 M Gn·HCl, pH 7, purged with argon). The reaction mixture was stirred at 40 °C, under Ar, during 20 minutes. After verifying completion of the reaction, MPAA ( $2.04 \cdot 10^{-4}$  mol, final concentration ~ 200 mM) was added as solid into the reaction mixture to quench the excess of 2-bromoacetamide. The product was directly isolated by HPLC.

'**Strand +7**': isolated with 18.9 mg (60 %) of ligation yield. LC-MS (ESI):  $[M + 5 H^+] = 1212.3$  m/z;  $[M + 6 H^+] = 1010.5$  m/z;  $[M + 7 H^+] = 866.3$  m/z;  $[M + 8 H^+] = 758.1$  m/z;  $[M + 9 H^+] = 674.1$  m/z;  $[M + 10 H^+] = 606.5$  m/z;  $C_{254}H_{444}N_{90}O_{79}S$ ; Average isotope calculated: 6055.0 Da [M]; observed: 6057.2 Da.

### Crystal structure of Het-elong, GAC

The crystallization experiments were carried out by the sitting drop vapor diffusion method at 293 K using a Mosquito Crystal nanolitre dispensing robot (TTP Labtech) and screening against several commercially available screens: The Classics Suite (Qiagen), the JCSG-plus, Morpheus, and the BCS screen (Molecular dimensions).

For the **Het-elong, GAC** homodimer, 0.15  $\mu$ L of solution at 21.2 mg/mL (1:1 stoichiometric ratio of 'strand +14, GAC' and 'strand +7') in MES buffer (100 mM MES, 100 mM NaCl, 10 mM TCEP, pH 6.5) was mixed with 0.15  $\mu$ L precipitant solution and equilibrated against 50  $\mu$ L reservoir solution. Optimization was realized by seeding from previously obtained crystals (from the screening) around the condition 20 % PEG 3350, 0.2 M ammonium sulfate and 0.1 M Bis Tris pH 5.5. Different ratios of protein/precipitant were also tested.

## 7.5. Chapter 5

### Generation of peptide fragments

To generate all possible peptide fragments of all different lengths coming from a unique protein (here, [N-ACTR]-<sup>a</sup>Mes), a python script was written. The calculated peptide fragments were then identified in the mass profile allowing  $\pm 1$  Da of deviation.

The script was the following:

```

frag=[]
seq=['E','G','Q','S','D','E','R','A','L','L','D','Q','L','H','T','L','L','S','N','T','D','A','T','G','L','E','E','I','D','R']
print("la séquence est",seq)
n=int(input("longueur du fragment"))
b=len(seq)
masse = {'E': 129.12,'G': 57.05,'Q': 128.13, 'S': 87.08, 'D': 115.09,'R': 156.11,'A':71.08,'L':113.16,'H':137.16,'T':101.11,'N':114.01,'I':113.16,'P':97.05,'K':128.09,'V':99.06,'M':134.04,'X':167,'C':103,'Y':163.06,'F':147}
fragm=[]
for a in range(int(b)):
    frag.append("".join(seq[a:a+n]))
    a+=n

for a in range(int(b)):
    fragm=frag[a]
    z=0
    m=[]
    z=0
    a=0
    while z in range(len(fragm)):
        m.append(masse[fragm[a]])
        z+=1
        a+=1
    print("le fragment est: ",fragm, "de masse ",sum(m)+18)

```

**Hydrolytic activity of Hom-N2 with p-NPA as substrate**

Hydrolytic activity of **Hom-N2** was monitored with *para*-nitrophenyl acetate following the **General procedure IX**. Kinetics curves were monitored at different initial concentration of pNPA,  $[pNPA]_0$ . Kinetic parameters were extracted following the equation  $[pNP] = At + B(1 - e^{-bt})$  and are summarized in **Table 7.4**.<sup>8</sup>

**Table 7.4. Kinetic parameters that were deduced from kinetic curves at each concentration of pNPA substrate by fitting the equation:  $[pNP] = At + B(1 - e^{-bt})$ . \*  $k_2/K_s$  is related to the initial burst phase efficiency**

$[pNPA]_0$	300 $\mu\text{M}$	400 $\mu\text{M}$	500 $\mu\text{M}$	800 $\mu\text{M}$	1000 $\mu\text{M}$
<b>A (<math>\mu\text{M sec}^{-1}</math>)</b>	0.0043	0.00451	0.00559	0.00971	0.01075
<b>B (<math>\mu\text{M}</math>)</b>	58.9	31.3	24.8	32.4	13.3
<b>b (<math>\text{sec}^{-1}</math>)</b>	0.00053	0.00072	0.00085	0.0014	0.0019
<b><math>k_2/K_s</math> (<math>\text{M}^{-1} \text{sec}^{-1}</math>)*</b>	1.76	1.80	1.70	1.76	1.89

**Catalytic assay of 1<sup>st</sup> generation in 'sterile' conditions**

**General procedure VII** was adapted to work with bacteria-free conditions. To do so, a slightly different buffer was prepared to have 50 mM sodium phosphate, 100 mM NaCl, 5 mM 0.05 % (w/v)  $\text{NaN}_3$ , pH 7.0. Once the pH was adjusted, the buffer was filtrated using sterile filter with a membrane of 0.22  $\mu\text{m}$  (pore size). [N-ACTR]- $\alpha$ thioester was first solubilized in 288  $\mu\text{L}$  of fresh buffer at the final concentration of 0.5 mM. The pH of the mixture was controlled and adjusted to 6.9-7.0. The reaction mixture was poured into the catalytic protein **Hom-N2** (*c* 0.5 mM). The pH of the resulting reaction mixture was adjusted to 6.9-7.0 giving the  $t = 0$  h. The reaction was monitored during 5 days at room temperature. Aliquots of 2  $\mu\text{L}$  of reaction mixture quenched in 50  $\mu\text{L}$  of  $\text{H}_2\text{O} + 0.1$  % TFA were taken to follow the reaction by HPLC and LC/MS at different time points. Half-lives were estimated by inspecting LC/MS spectra.

For catalytic assays without sterile conditions (*i.e.* normal conditions in which decomposition of the thioester substrates was observed), the **General procedure VII** was strictly followed.

## 7.6. References

- (1) Schuck, P. Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys. J.* **2000**, *78* (3), 1606–1619.
- (2) Laue, T.; Shah, B. V.; Ridgeway, T. M.; Pelletier, S. L. Computer-aided interpretation of analytical sedimentation data for proteins <https://www.semanticscholar.org/paper/Computer-aided-interpretation-of-analytical-data-Laue-Shah/f53383e10a8969e4cbca338e57b9f07e13ff3d37> (accessed Apr 23, 2020).
- (3) Brautigam, C. A. Chapter Five - Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. In *Methods in Enzymology*; Cole, J. L., Ed.; Analytical Ultracentrifugation; Academic Press, 2015; Vol. 562, pp 109–133.
- (4) 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.
- (5) Coin, I.; Beyermann, M.; Bienert, M. Solid-Phase Peptide Synthesis: From Standard Procedures to the Synthesis of Difficult Sequences. *Nat. Protoc.* **2007**, *2* (12), 3247–3256.
- (6) Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I. Color Test for Detection of Free Terminal Amino Groups in the Solid-Phase Synthesis of Peptides. *Anal. Biochem.* **1970**, *34* (2), 595–598.
- (7) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. In Situ Neutralization in Boc-Chemistry Solid Phase Peptide Synthesis. *Int. J. Pept. Res. Ther.* **2007**, *13* (1), 31–44.
- (8) Bender, M. L.; Kezdy, F. J.; Wedler, F. C. Alpha-Chymotrypsin: Enzyme Concentration and Kinetics. *J. Chem. Educ.* **1967**, *44* (2), 84.
- (9) Kochendoerfer, G. G.; Chen, S.-Y.; Mao, F.; Cressman, S.; Traviglia, S.; Shao, H.; Hunter, C. L.; Low, D. W.; Cagle, E. N.; Carnevali, M.; Gueriguian, V.; Keogh, P. J.; Porter, H.; Stratton, S. M.; Wiedeke, M. C.; Wilken, J.; Tang, J.; Levy, J. J.; Miranda, L. P.; Crnogorac, M. M.; Kalbag, S.; Botti, P.; Schindler-Horvat, J.; Savatski, L.; Adamson, J. W.; Kung, A.; Kent, S. B. H.; Bradburne, J. A. Design and Chemical Synthesis of a Homogeneous Polymer-Modified Erythropoiesis Protein. *Science* **2003**, *299* (5608), 884–887.
- (10) Veber, D.; Milkowski, J.; Varga, S.; Denkwalter, R.; Hirschmann, R. Acetamidomethyl. A Novel Thiol Protecting Group for Cysteine. *J. Am. Chem. Soc.* **1972**, *94* (15), 5456–5461.
- (11) Kabsch, W. XDS. *Acta Crystallogr., Sect. D* **2010**, *66* (2), 125–132.
- (12) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (4), 658–674.
- (13) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkóczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L.-W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D. Macromolecular Structure Determination Using X-Rays, Neutrons and Electrons: Recent Developments in Phenix. *Acta Crystallogr., Sect. D* **2019**, *75* (10), 861–877.

- (14) Ogihara, N. L.; Ghirlanda, G.; Bryson, J. W.; Gingery, M.; DeGrado, W. F.; Eisenberg, D. Design of Three-Dimensional Domain-Swapped Dimers and Fibrous Oligomers. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (4), 1404–1409.
- (15) Matthews, B. W. Solvent Content of Protein Crystals. *J. Mol. Biol.* **1968**, *33* (2), 491–497.
- (16) Evans, P. R.; Murshudov, G. N. How Good Are my Data and What Is the Resolution? *Acta Crystallogr., Sect. D* **2013**, *69* (7), 1204–1214.
- (17) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr., Sect. D* **2010**, *66* (4), 486–501.
- (18) Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.; Deis, L. N.; Verma, V.; Keedy, D. A.; Hintze, B. J.; Chen, V. B.; Jain, S.; Lewis, S. M.; Arendall, W. B.; Snoeyink, J.; Adams, P. D.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. MolProbity: More and Better Reference Data for Improved All-Atom Structure Validation. *Protein Sci.* **2018**, *27* (1), 293–315.
- (19) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* **1993**, *26* (2), 283–291.
- (20) Karplus, P. A.; Diederichs, K. Assessing and Maximizing Data Quality in Macromolecular Crystallography. *Curr. Opin. Struct. Biol.* **2015**, *34*, 60–68.
- (21) Grieco, P.; Gitu, P. M.; Hruby, V. J. Preparation of ‘Side-Chain-to-Side-Chain’ Cyclic Peptides by Allyl and Alloc Strategy: Potential for Library Synthesis. *J. Pept. Res.* **2001**, *57* (3), 250–256.
- (22) Schmidtgal, B.; Chaloin, O.; Bauer, V.; Sumyk, M.; Birck, C.; Torbeev, V. Dissecting Mechanism of Coupled Folding and Binding of an Intrinsically Disordered Protein by Chemical Synthesis of Conformationally Constrained Analogues. *Chem. Commun.* **2017**, *53* (53), 7369–7372.
- (23) 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.
- (24) Fischer, L.; Semetey, V.; Lozano, J.-M.; Schaffner, A.-P.; Briand, J.-P.; Didierjean, C.; Guichard, G. Succinimidyl Carbamate Derivatives from N-Protected  $\alpha$ -Amino Acids and Dipeptides—Synthesis of Ureidopeptides and Oligourea/Peptide Hybrids. *Eur. J. Org. Chem.* **2007**, *2007* (15), 2511–2525.
- (25) Sato, K.; Tanaka, S.; Yamamoto, K.; Tashiro, Y.; Narumi, T.; Mase, N. Direct Synthesis of N-Terminal Thiazolidine-Containing Peptide Thioesters from Peptide Hydrazides. *Chem. Commun.* **2018**, *54* (66), 9127–9130.
- (26) Zhou, J.; Grigoryan, G. Rapid Search for Tertiary Fragments Reveals Protein Sequence–Structure Relationships. *Protein Sci.* **2015**, *24* (4), 508–524.



# Développement de protéines synthétiques *de novo* catalysant le transfert de groupements acyles

## Résumé

Le but de cette thèse est de développer une enzyme synthétique *de novo* capable de catalyser efficacement le transfert de groupes acyles, et notamment la formation de liaisons amide entre deux peptides en n'introduisant aucune restriction de séquence au niveau du site de jonction. En utilisant différents thioesters peptidiques comme substrats, nous avons démontré l'accélération des diverses réactions de transfert de groupements acyles. Des bibliothèques d'analogues ont été synthétisées et ont permis de prouver que la composition du site catalytique a une influence sur l'issue des réactions catalysées. Grâce à la modularité de notre structure protéique, nous avons aussi conçu des analogues de forme plus allongée et envisagé l'émergence de nouvelles activités catalytiques. Les résultats obtenus représentent un point de départ prometteur vers le développement de futurs catalyseurs protéiques pour la ligation et la cyclisation de peptides ainsi que pour le marquage des protéines.

**Mots clés :** conception de protéines *de novo*, synthèse totale de protéines, motif superhélice, catalyse enzymatique de liaisons amide.

## Résumé en anglais

The goal of this thesis is to develop a synthetic *de novo* enzyme for robust catalysis of acyl transfer reactions and more notably of amide bond formation between two peptides with no stringent restrictions to the amino acid composition at the ligation junction. Using peptide thioesters as acyl-donors, we demonstrated their catalyzed aminolysis concomitant with hydrolysis in various acyl transfer reactions. Libraries of analogues were chemically synthesized and proved that the environment at the catalytic site influences the reaction outcome. To further explore the modularity of our *de novo* protein scaffold, we designed elongated analogues and we also anticipated new emerging properties. The results obtained in this thesis represent a promising starting point for the development of efficient protein catalysts for protein labeling and peptide ligation and cyclization.

**Keywords:** *de novo* protein design, total synthesis of proteins, coiled-coil helices, enzymatic catalysis of amide bond formation