

Nature-inspired Circular-economy Recycling (NaCRe) for Proteins

Présentée le 24 septembre 2021

Faculté des sciences et techniques de l'ingénieur
Laboratoire des nanomatériaux supramoléculaires et interfaces - Chaire Constellium
Programme doctoral en science et génie des matériaux

pour l'obtention du grade de Docteur ès Sciences

par

Simone GIAVERI

Acceptée sur proposition du jury

Prof. D. Grundler, président du jury
Prof. F. Stellacci, Prof. S. Maerkl, directeurs de thèse
Prof. T. Weil, rapporteuse
Prof. F. G. Omenetto, rapporteur
Prof. M. W. Tibbitt, rapporteur

“Now we have a reason to live—
to learn, to discover, to be free!”

— Richard Bach

Jonathan Livingston Seagull

To my family.

Acknowledgements

During my doctoral studies I came across with people that left a strong positive mark on me. First of all, I'm grateful to Francesco, my advisor, for having supervised me during the last five years. He has been guiding, and motivating me all time, giving me always the freedom to find my path, to fly alone. With his creativity he widened my vision about Materials Science towards new shores. I highly appreciated his teachings, as well as the discussions together. Moreover, I'm grateful to him for supporting my ideas for my next studies.

I would like to thank Sebastian, my co-advisor, for having accepted me in his group, for having helped me to make this work possible experimentally, and especially for having stimulated my curiosity towards new disciplines, by exposing me to a highly interdisciplinary collaboration. Nothing would have been achieved without the help of Prof. Giuseppe Zerbi, Prof. Paolo Gronchi, and Prof. Daniele Dondi who strongly supported me from home, pushing me to think positive, when things were not easy.

I would like to express my gratitude to my collaborators Adeline M. Schmitt, Laura Roset Julià, Vincenzo Scamarcio, Dr. Anna Murello, Shiyu Cheng, Dr. Laure Menin, Dr. Daniel Ortiz, Dr. Luc Patiny, Dr. Sreenath Bolisetty, and Prof. Raffaele Mezzenga for their precious help, and to Prof. Bruno Correia, and Prof. Aleksandra Radenovic for the useful discussions.

Thanks to my colleagues in SuNMIL, and LBNC that shared with me most of the day in the lab. A special thank goes to Dr. Nadanai Laohakunakorn, Dr. Grégoire Michielin, Dr. Zoe Swank, Dr. Ivan Istomin, Dr. Francesca Volpetti, Dr. Anna Murello, and Matteo Gasbarri with whom I shared unique moments, and unforgettable experiences.

Finally, I would like to express all my gratitude to Andrea, Alessandro, and my family for being always with me from home.

Lausanne, June 9, 2021

S. G.

Abstract

Since a few decades our planet has been loaded with billion tons of synthetic polymer-based materials, commonly named plastics. The large scale of plastic production, associated with its limited recyclability, are the driving force for the accumulation of such materials into the environment. Garbage patches, *i.e.* islands of plastics in the ocean, are striking evidences of this issue. A sustainable handling, that is production and disposal, of synthetic polymer-based materials is one of the greatest challenges that humanity has to face.

Proteins and nucleic acids are natural polymers. Arguably, Nature produces an amount of such polymers that is higher with respect to man-made polymers. However, these materials do not accumulate into the environment, that is the approach used by Nature to handle these polymers is sustainable. The secret for its sustainability lies on the circularity in the materials' use. Indeed, proteins and nucleic acids are sequence-defined polymers that undergo depolymerization to monomers, and recycling into new materials by reassembling the so obtained monomers into arbitrarily different sequences. Organisms digest proteins into amino acids. These monomers are in turn polymerized to produce the protein of need, at the time of the protein synthesis.

In this thesis we show that this process is achievable outside living organisms. Indeed, we depolymerized structurally different short peptides, and peptides mixtures into their constitutive amino acids, and recycled such monomers into biotechnologically relevant proteins (green, and red fluorescent proteins), by using an amino acid-free cell-free transcription-translation system. We further applied our methodology to recycle proteins with high relevance in materials engineering such as β -lactoglobulin films, used for water filtration, or silk fibroin solutions into green fluorescent protein. We were also successful in recycling mixtures composed of

short peptides, and technologically relevant proteins into fluorescent proteins, as well as into bioactive enzymes (catechol 2,3-dioxygenase). Finally, we achieved multiple cycles of recycling (two cycles), and we demonstrated that the strategy can be expanded beyond the set of the twenty proteinogenic amino acids.

Presented herein is a nature-inspired approach to recycling of soft materials, where unknown mixtures of polymers are recycled into the polymer of need, by the local community, at the time of recycling. The materials are constantly transformed into different ones, without any external feed, and there is no way to distinguish a new polymer from a recycled one. The strength of this method lies in its compatibility with the principles of circular economy.

Keywords: recycling, sequence-defined polymers, protein-based materials, sustainability

Résumé

Depuis les dernières décennies, notre planète est surchargée de milliards de tonnes de matériaux polymériques synthétiques, communément appelés plastiques. La production de plastique à grande échelle, associée à la capacité limitée de son recyclage sont la force motrice de l'accumulation de ces matériaux dans l'environnement, comme représenté par l'exemple frappant des îles de plastique flottant dans l'océan. Élaborer un traitement durable, tant pour la production que pour l'évacuation de ces matériaux polymères synthétiques est l'un des plus grands enjeux que l'humanité doit affronter.

Les protéines et les acides nucléiques sont des polymères naturels. La nature produit de tels polymères à une quantité supérieure que celle des Hommes pour le plastique. Toutefois, ces matériaux ne s'accumulent pas dans l'environnement et ce, grâce à l'approche durable qu'utilise la nature pour traiter ces polymères. Le secret de cette durabilité repose sur la circularité de l'utilisation de ces matériaux. En effet, les protéines et acides nucléiques sont des polymères à séquences définies qui subissent leur dépolymérisation en monomères ainsi que leur recyclage en nouveaux matériaux en assemblant les monomères obtenus préalablement en nouvelles séquences arbitraires. Les organismes digèrent les protéines en acides aminés. Ces monomères sont ensuite polymérisés en retour pour produire la protéine nécessaire lors de la synthèse.

Dans cette thèse, nous montrons que ce procédé est accessible en dehors des organismes vivants. En effet, nous dépolymérisons de courts peptides de structures différentes ainsi que des mélanges de peptides, en leurs acides aminés constitutifs, puis nous recyclons ces monomères en protéines à intérêt biotechnologique (protéine fluorescente verte et rouge), en utilisant un système de transcription-translation acellulaire exempt d'acides aminés. Nous

avons ensuite appliqué notre méthode pour recycler des protéines possédant un haut intérêt en ingénierie de matériaux telles que les films de β -lactoglobuline utilisés pour filtrer l'eau, ou encore des solutions de fibroïne de soie, en protéine verte fluorescente. Nous avons réussi à recycler des mélanges composés de courts peptides et de protéines à intérêt technologique en protéines fluorescentes ainsi qu'en enzymes bioactives (catechol 2,3-dioxygénase). Enfin, nous avons accompli multiples cycles de recyclage (deux cycles) et avons démontré que la stratégie peut être étendue au-delà du jeu des vingt acides aminés protéinogènes.

Nous présentons ici une approche inspirée par la nature pour recycler des matériaux souples, où le mélange de polymères est recyclé en un polymère nécessaire, par la communauté locale, au temps de son recyclage. Les matériaux sont constamment transformés en éléments différents sans apport externe, et sans pouvoir faire la distinction entre un nouveau polymère et un polymère recyclé. La force de cette méthode repose en sa compatibilité avec les principes d'une économie circulaire.

Mots-clés : recyclage, polymères à séquence définie, matériaux à base de protéines, durabilité

Contents

Acknowledgements	i
Abstract (English/Français)	iii
List of Figures	xi
List of Tables	xxvii
Structure of the thesis	1
1 Introduction	3
1.1 Plastic pollution	3
1.2 Recycling of plastics	5
1.2.1 Mechanical recycling	5
1.2.2 Compatibilization	6
1.2.3 Chemical recycling	6
1.2.4 Repurposing or upcycling	10
1.2.5 Incineration	11
1.3 Biodegradable polymers	12
1.4 Recycling of natural polymers: a lesson from Nature	13
1.5 Sequence-defined polymers	14
1.5.1 Sequence-defined polymers through chemistry	15
1.5.2 Sequence-defined polymers through biology	17
	vii

Contents

I	Body	21
2	NaCRe: beyond recycling	23
3	Cell-free transcription-translation	25
3.1	Cell-free transcription-translation in PURE system	25
4	NaCRe: results and discussion	27
4.1	NaCRe: recycling short peptides	27
4.2	NaCRe: recycling technologically relevant proteins	34
4.3	NaCRe: beyond natural proteins	40
5	Non-natural proteins – synthetic SDPs	43
5.1	Non-natural proteins as synthetic SDPs: results	43
6	Conclusions and Outlook	47
II	Appendix	51
A	Materials	55
B	Methods	57
B.1	Calibrant expression	57
B.2	Film preparation	58
B.3	Selection of the model peptides	58
B.4	Depolymerization I (cleavage)	62
B.5	Depolymerization II (depolymerization)	63
B.6	Polymerase Chain Reaction (PCR)	64
B.7	CF protein TX-TL	65
B.8	Second NaCRe cycle	69
B.9	Peptide (and UAAs) analysis by mass spectrometry	71
B.10	Amino Acid Analysis by mass spectrometry (AAA)	71
B.11	Protein electrophoresis (SDS-PAGE)	73
B.12	Proteomic analysis	73

	Contents
B.13 Mass calibration	74
B.14 AFM imaging	75
C Additional Tables	77
D Additional Data	83
References	111
List of publications	123
Curriculum Vitae	125

List of Figures

1.1	Schematic illustration of microplastics accumulation in the deep sea. The figure has been reproduced from reference 6, with permission from AAAS.	4
1.2	Schematic illustration representing the plot of the Gibbs free energy as function of the reaction coordinate for the polymerization reaction of the monomer M at different temperatures ($T \ll T_C$, $T = T_C$, and $T \gg T_C$, where T_C is the equilibrium temperature at which $\Delta G_P = 0$). x is the free energy of the monomers that is function of the reaction temperature. The figure has been reproduced from reference 26, with permission from Springer Nature.	7
1.3	Schematic illustration of the ring opening polymerization of 3,4-T6GBL to obtain linear, and cyclic poly(3,4-T6GBL) (indicated as “Catalytic Polymer Synthesis”), and chemical recycling to monomer of poly(3,4-T6GBL) to achieve 3,4-T6GBL (indicated as “Catalytic Polymer Recycling”). 3,4-T6GBL is used to obtain virgin quality poly(3,4-T6GBL) in a potentially “infinite-recyclable” fashion. The figure has been reproduced from reference 35, with permission from Elsevier.	9
1.4	Schematic illustration of protein polymerization. The elongation step (1) through which one peptide bond is formed, the removal of the reacted tRNA (2), and the introduction of a new monomer (3) are sketched. The figure has been reproduced from reference 53, with permission from Springer Nature.	14
1.5	Schematic illustration of the synthesis of SD poly(alkoxyamine amide)s (a). Structure of the solid-phase, or soluble supports used (b). Chemical structure of a model SD oligomer achivable by using this method (c). The figure has been reproduced from reference 59, with permission from Springer Nature.	16

List of Figures

- 1.6 Schematic illustration of non-natural protein translation in PURE, by solely replacing the proteinogenic AAs (blue squares) with their correspondent UAA analogs (orange squares). The figure has been reproduced from reference 92, with permission from Elsevier. 19
- 2.1 Schematic illustration for the main concept of NaCRe. Multiple possible NaCRe cycles are shown. The illustrated examples are close to what is shown in this thesis. It should be clear that the overall concept of NaCRe goes beyond what is illustrated. The sketched process starts from three different short peptides (drawn as the ones used in this thesis, magainin II, glucagon, and somatostatin), and produces green fluorescent protein (GFP). In the second round of recycling, GFP, together with other arbitrary proteins, is used to produce red fluorescent protein (mScarlet-i). In the last recycling round mScarlet-i is recycled into something not specified, to stimulate the reader's imagination. Molecular graphics of the proteins 3D structures and of the AAs conformers were from PDB databank (protein 1(2LSA), protein 2(2MI1), protein 3(1GCN), protein 4(5B61), and protein 5(5LK4)) and PubChem (<https://pubchem.ncbi.nlm.nih.gov/compound/CID#section=3D-Conformer>, CID = 5950, 6322, 5960, 5961, 33032, 6274, 6306, 6106, 5962, 6137, 6140, 145742, 5951, 6305, 6057) respectively. All were edited in UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at UCSF, with support from NIH P41-GM103311. 24
- 3.1 Schematic illustration of the four main reactions (aminoacylation, transcription, translation, and energy regeneration) of cell-free protein synthesis in PURE. The figure has been reproduced from reference 73, with permission from Elsevier. . 26
- 4.1 Plots of the fluorescence signal resulting from the expression of mScarlet-i in our TX-TL system without the addition of any AA. 28

- 4.2 Recycling of magainin II, glucagon, and somatostatin 28 into mScarlet-i. Bar graphs showing the result of the amino acid analysis performed by using mass spectrometry on the result of depolymerization of magainin II (a), glucagon (b), and somatostatin 28 (c). The experimental results are represented with green bars to be compared with the gray bars that are the ideal reference concentrations of each AA calculated by assuming the complete conversion of the starting peptide into free AAs. The violet bars represent trace concentration of the AAs that theoretically should have not been observed, they are possibly the result of depolymerization of the digestion enzymes themselves. Such impurities are present for all the recovered AAs. The additive effect due to the impurities is by definition difficult to estimate, and probably contributes to slightly overestimate the green bars. This becomes more evident when the obtained depolymerization yield is close to 100%. (Note: cysteine is not detected by the amino acid analysis, hence the quantification of cysteine is n.a.). Plots of the fluorescence signal resulting from the expression of mScarlet-i in a TX-TL reaction (d). The green curves are data obtained performing NaCRe on magainin II, glucagon, and somatostatin 28 (Appendix B.7), the gray curves are obtained as the results of expression experiments with the TX-TL reactions supplemented with concentrations for each AA matching the gray bars shown in (a), (b), and (c). In the negative control expressions (violet curves), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the individual peptides, without adding the peptides initially. Bar-plots of the statistical mean of the results of the repeated injections (triplicates) of each sample are shown; error bars represent the standard deviation of the same data. The TX-TL reactions were all run in duplicates. The expression curves represent the statistical mean of the results at any acquisition time; the shadow represents the standard deviation of the same data. 30

- 4.3 Recycling of magainin II, glucagon, and somatostatin 28 into GFP. Plots of the fluorescence signal resulting from the expression of GFP in our TX-TL system. The green curve is obtained performing NaCRe on magainin II, glucagon, and somatostatin 28 (Appendix B.7). The gray curve (reference control) is obtained as the result of an expression experiment with the TX-TL system supplemented with concentrations of AAs matching the complete depolymerization of the initial materials (gray bars in 4.2a-c). In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the individual peptides, without adding the peptides initially. Details are explained in the caption of Figure 4.2. 31

- 4.4 Recycling of the mixture of magainin II, glucagon, and somatostatin 28 into GFP. Bar graphs showing the result of the amino acid analysis performed using mass spectrometry on the result of depolymerization of the mixture of magainin II, glucagon, and somatostatin 28 (a). Plot of the fluorescence signal resulting from the expression of GFP in a TX-TL reaction (b). The green curve is obtained performing NaCRe on the mixture of magainin II, glucagon, and somatostatin 28 (Appendix B.7). The gray curve (reference control) is obtained as the result of an expression experiment with the TX-TL system supplemented with concentrations of AAs matching the complete depolymerization of the initial materials (gray bars in (a)). In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the peptides mixture, without adding the peptides initially. Details are explained in the caption of Figure 4.2. 32

4.5	Recycling of β -lactoglobulin A into GFP. Bar graphs showing the result of the amino acid analysis performed using mass spectrometry on the result of depolymerization of β -lactoglobulin A (a). Plots of the fluorescence signal resulting from the expression of GFP (b) in a TX-TL reaction. The green curve is obtained performing NaCRe on β -lactoglobulin A (Appendix B.7). The gray curve (reference control) is obtained as the result of an expression experiment with the TX-TL system supplemented with concentrations of AAs matching the complete depolymerization of the initial material (gray bars in (a)). In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for β -lactoglobulin A, without adding β -lactoglobulin A initially. Details are explained in the caption of Figure 4.2.	35
4.6	Photograph of a film composed of β -lactoglobulin amyloid fibrils, deposited on a cellulose support.	35
4.7	AFM characterization of the β -lactoglobulin amyloids obtained from solubilizing the film powder, and deposited on cleaved mica surfaces, as prepared (a) and after depolymerization (b).	36
4.8	Recycling of a film composed of β -lactoglobulin amyloid fibrils into GFP. Bar graphs showing the result of the amino acid analysis performed using mass spectrometry on the result of depolymerization of the β -lactoglobulin film (a). Plots of the fluorescence signal resulting from the expression of GFP (b) in a TX-TL reaction. The green curve is obtained performing NaCRe on the β -lactoglobulin film (Appendix B.7). In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the film, without adding the film initially. The reference controls (gray bars in (a) and gray curve in (b)) are missing because the exact composition of the amyloids composing the film is unknown. Details are explained in the caption of Figure 4.2.	36

List of Figures

- 4.9 Recycling of a silk fibroin solution into GFP. Bar graphs showing the result of the amino acid analysis performed using mass spectrometry on the result of depolymerization of a silk fibroin solution (a). Plots of the fluorescence signal resulting from the expression of GFP (b) in a TX-TL reaction. The green curve is obtained performing NaCRe on the silk fibroin solution (Appendix B.7). The gray curve (reference control) is obtained as the result of an expression experiment with the TX-TL system supplemented with concentrations of AAs matching the complete depolymerization of the initial material (gray bars in (a)). In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the silk fibroin, without adding the silk fibroin initially. Details are explained in the caption of Figure 4.2. 37
- 4.10 Recycling of cysteine. In (a) the green curves are obtained performing NaCRe on magainin II, glucagon, and somatostatin 28 with and without spiking cysteine. In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the individual peptides, without adding the peptides initially. In (b) the green curve is obtained performing NaCRe on the mixture composed of glucagon, β -lactoglobulin A, and silk fibroin (Appendix B.7). The gray curve (reference control) is the result of an expression experiment with the TX-TL system supplemented with concentrations of AAs matching the complete depolymerization of the initial materials. In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the glucagon, β -lactoglobulin A, and silk fibroin mixture, without adding the three proteins initially. Details are explained in the caption of Figure 4.2. 38

- 4.11 Second NaCRe cycle. In (a) the green curve is obtained preforming a second cycle of NaCRe on the GFP produced by recycling the mixture composed of glucagon, β -lactoglobulin A, and silk fibroin (Appendix B.8). In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for GFP, without adding the protein initially. In (b) the green curve is obtained preforming NaCRe on the whole solution resulting from a first cycle of NaCRe in which glucagon, β -lactoglobulin A, and silk fibroin were recycled into GFP (Appendix B.8). In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the whole first cycle of NaCRe, without adding the whole first cycle of NaCRe initially. Details are explained in the caption of Figure 4.2. 39
- 4.12 Recycling of the mixture of glucagon, β -lactoglobulin A, and silk fibroin into the enzyme catechol 2,3-dioxygenase (CDO). Plots of the absorbance signal at 385 nm resulting from the conversion of catechol into 2-hydroxymuconate semialdehyde, catalyzed by the CDO enzyme expressed in our TX-TL system. The green curve is obtained preforming NaCRe on the mixture composed of glucagon, β -lactoglobulin A, and silk fibroin (Appendix B.7). In the negative control (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the glucagon, β -lactoglobulin A, and silk fibroin mixture, without adding the three proteins initially. The TX-TL reactions were all run in duplicates. The absorbance curves represent the statistical mean of the results at any acquisition time; the shadow represents the standard deviation of the same data. 40

List of Figures

- 5.1 Fluorescent images of SDS-PAGE protein gels used to identify proteins (GFP) containing single (a), or multiple (b) UAAs incorporations. In each gel, a green marker in lane 1 is used to indicate the protein bands corresponding to GFP obtained by using the 20 proteinogenic AAs (positive controls). Red markers in lanes Xa are used to indicate the protein bands corresponding to GFP with UAAs incorporations (samples). In (a), incorporations of L-canavanine, L-norleucine, L-3 hydroxy-norvaline, or β -t-butyl-L-alanine are shown in lanes 2a, 3a, 4a, and 5a respectively. In (b), incorporations of L-canavanine, L-canavanine + L-norleucine, L-canavanine + L-norleucine + L-3-hydroxy-norvaline, L-canavanine + L-norleucine + L-3-hydroxy-norvaline + 3-fluoro-L-valine, and L-canavanine + L-norleucine + L-3-hydroxy-norvaline + 3-fluoro-L-valine + S-(2-aminoethyl)-L-cysteine hydrochloride are shown in lanes 2a, 3a, 4a, 5a, and 6a respectively. Protein bands in lanes Xb should not be observed (negative controls). They are possibly due to the expression of GFP with trace concentration of natural AAs. The additional fluorescent bands visible at the bottom of each gel are due to tRNA-Lys-BODIPY-FL that is present in excess in each TX-TL reaction. 44
- 5.2 MS spectrum of the non-natural peptide H3(3+)For[L-norleucine] [3-fluoro-L-valine] [3-fluoro-L-valine] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-norleucine] [Serine] [Lysine]OH expressed at the N-terminus of GFP, and obtained by in gel digestion with Lys-C. Theoretical observed mass (m/z) = 538.639; experimental closest peak (m/z) = 538.640. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). . . 45
- 5.3 MS spectrum of the non-natural peptide H3(3+)For[L-norleucine] [3-fluoro-L-valine] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline][L-norleucine][3-fluoro-L-valine] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [Serine] [Lysine]OH expressed at the N-terminus of GFP, and obtained by in gel digestion with Lys-C. Theoretical observed mass (m/z) = 538.639; experimental closest peak (m/z) = 538.640. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 46

- D.1 Plots of the fluorescence signal resulting from the expression of mScarlet-i in our TX-TL system by using 3 different initial concentrations (25/75 v/v reference AAs mixture:nuclease-free water, 50/50 v/v reference AAs mixture:nuclease-free water, and 100/0 v/v reference AAs mixture:nuclease-free water) of the reference AAs mixture from magainin II, glucagon, and somatostatin 28 complete depolymerization. 84
- D.2 Plots of the fluorescence signal resulting from the expression of mScarlet-i in our TX-TL system (0 – 240 min) by using the reference AAs mixture from magainin II, glucagon, and somatostatin 28 complete depolymerization (grey curve), and substituting the reference AAs mixture with the negative controls (pink and purple curves); DNA(75 ng) was replaced by nuclease- free water (purple curve). Plots of the fluorescence signal resulting from the expression of mScarlet-i in our TX-TL system (270 – 845 min) by spiking the reference AAs mixture from magainin II, glucagon, and somatostatin 28 complete depolymerization and the negative control (with DNA) with a preheated (37 °C) stock water-AAs solution to get to a final expression reaction (0.39/2.5/25 v/v/v nuclease-free water:3mM AAs spike solution:expression (grey curve and pink curves), and the negative control (without DNA) with a preheated (37 °C) DNA(75 ng)-AAs solution to get to a final expression reaction (0.39/2.5/25 v/v/v DNA(75 ng):3mM AAs spike solution:expression (purple curve). (The plate reader sensitivity was exceptionally set to 80 % in this experiment). 85
- D.3 Plot of the mScarlet-i mass calibration curve in the plate reader (Appendix B.13). Error bars represent the variability of the expression by using different lots of PURE Frex™Solution II, and III, calculated as the standard deviation of the expression plateaus (RFU) in Figure D.4. 86
- D.4 Plots of the fluorescence signal resulting from the expression of mScarlet- i in our TX-TL system by using the reference AAs mixture from magainin II, glucagon, and somatostatin 28 complete depolymerization. Two lots of PURE-frex™Solution II, and III were used in order to quantify the variability of the expression plateau (RFU), as function of the PUREfrex™Solution II, and III lots. 86

List of Figures

- D.5 Plots of the fluorescence signal resulting from the expression in our TX- TL system of the GFP modified with the incorporation of L-norleucine, and L-canavanine. The green curve is obtained preforming NaCRe on the unnatural peptide (for recycling of L-norleucine, and L- canavanine), and supplementing the TX-TL system with the additional 18 proteinogenic AAs (Appendix B.7). The gray curve (reference control) is the result of an expression experiment with the TX-TL system supplemented with highly concentrated L-norleucine and L- canavanine, and with the additional 18 proteinogenic AAs. In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the unnatural peptide, without adding the unnatural peptide initially. In the additional negative control (light blue curve), the negative control depolymerization solution was substituted with nuclease-free water. 87
- D.6 MS spectrum of magainin II: H3(3+)HGlylleGlyLysPheLeuHisSerAlaLysLysPheGlyLysAlaPheValGlyGlulleMetAsnSerOH. Theoretical observed mass (m/z) = 822.782; experimental closest peak (m/z) = 822.784. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 88
- D.7 MS spectrum of glucagon: H3(3+)HHisSerGlnGlyThrPheThrSerAspTyrSerLysTyrLeuAspSerArgArgAlaGlnAspPheValGlnTrpLeuMetAsnThrOH. Theoretical observed mass (m/z) = 1161.213; experimental closest peak (m/z) = 1161.219. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 88
- D.8 MS spectrum of somatostatin 28: H4(4+)HSerAlaAsnSerAsnProAlaMetAlaProArgGluArgLysAlaGlyCys(H-)LysAsnPhePheTrpLysThrPheThrSerCys(H-)OH. Theoretical observed mass (m/z) = 787.623; experimental closest peak (m/z) = 787.625. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 89
- D.9 MS spectrum of the non-natural peptide: H2(2+)HMet(S-1CH2)Val(H-1F)Val(H-1F)Arg((CH2)-1O)Thr(CH2)Thr(CH2)Arg((CH2)-1O)Thr(CH2)Thr(CH2)Thr(CH2)Met(S-1CH2)SerLysOH. Theoretical observed mass (m/z) = 793.458; experimental closest peak (m/z) = 793.460. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 89

D.10 MS spectrum of magainin II cleaved fragment D1>5: H1(1+)HGlyIleGlyLysPheOH. Theoretical observed mass (m/z) = 521.308; experimental closest peak (m/z) = 521.309. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	90
D.11 MS spectrum of magainin II cleaved fragment D6>11: H1(1+)HLeuHisSerAlaLys LysOH. Theoretical observed mass (m/z) = 683.420; experimental closest peak (m/z) = 683.420. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	90
D.12 MS spectrum of magainin II cleaved fragment D12>15: H1(1+)HPheGlyLysAlaOH. Theoretical observed mass (m/z) = 422.240; experimental closest peak (m/z) = 422.240. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	91
D.13 MS spectrum of magainin II cleaved fragment D16>19: K1(1+)HPheValGlyGluOH. Theoretical observed mass (m/z) = 489.175; experimental closest peak (m/z) = 489.175. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	91
D.14 MS spectrum of magainin II cleaved fragment D20>23: K1(1+)HlleMetAsnSerOH. Theoretical observed mass (m/z) = 502.173; experimental closest peak (m/z) = 502.174. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	92
D.15 MS spectrum of glucagon cleaved fragment D1>5: K1(1+)HHisSerGlnGlyThrOH. Theoretical observed mass (m/z) = 567.192; experimental closest peak (m/z) = 567.193. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	92
D.16 MS spectrum of glucagon cleaved fragment D6>13: K1(1+)HPheThrSerAspTyrSer LysTyrOH. Theoretical observed mass (m/z) = 1048.402; experimental closest peak (m/z) = 1048.404. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	93

List of Figures

- D.17 MS spectrum of glucagon cleaved fragment D14>21: K1(1+)H1(1+)HLeuAspSer ArgArgAlaGlnAspOH. Theoretical observed mass (m/z) = 499.724; experimental closest peak (m/z) = 499.725. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 93
- D.18 MS spectrum of glucagon cleaved fragment D22>25: K1(1+)HPheValGlnTrpOH. Theoretical observed mass (m/z) = 617.248; experimental closest peak (m/z) = 617.249. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 94
- D.19 MS spectrum of glucagon cleaved fragment D26>29: H1(1+)HLeuMetAsnThrOH. Theoretical observed mass (m/z) = 478.233; experimental closest peak (m/z) = 478.233. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 94
- D.20 MS spectrum of somatostatin 28 cleaved fragment D1>7: H1(1+)HSerAlaAsnSer AsnProAlaOH. Theoretical observed mass (m/z) = 660.295; experimental closest peak (m/z) = 660.297. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 95
- D.21 MS spectrum of somatostatin 28 cleaved fragment D8>14: H3(3+)HMetAlaProArg GluArgLysOH. Theoretical observed mass (m/z) = 296.501; experimental closest peak (m/z) = 296.502. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 95
- D.22 MS spectrum of somatostatin 28 cleaved fragment D21>24: H2(2+)HPheTrpLys ThrOH. Theoretical observed mass (m/z) = 291.158; experimental closest peak (m/z) = 291.159. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 96
- D.23 MS spectrum of somatostatin 28 cleaved fragment D25>26: H1(1+)HPheThrOH. Theoretical observed mass (m/z) = 267.134; experimental closest peak (m/z) = 267.135. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 96

- D.24 MS spectrum of the non-natural peptide cleaved fragment D1>10: H₂(2+)HMet(S-1CH₂)Val(H-1F)Val(H-1F)Arg((CH₂)-1O)Thr(CH₂)Thr(CH₂)Arg((CH₂)-1O)Thr(CH₂)Thr(CH₂)Thr(CH₂)OH. Theoretical observed mass (m/z) = 629.352; experimental closest peak (m/z) = 629.354. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 97
- D.25 MS spectrum of the non-natural peptide cleaved fragment D1>6: H₂(2+)HMet(S-1CH₂)Val(H-1F)Val(H-1F)Arg((CH₂)-1O)Thr(CH₂)Thr(CH₂)OH. Theoretical observed mass (m/z) = 377.717; experimental closest peak (m/z) = 377.717. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). . . 97
- D.26 MS spectrum of the non-natural peptide cleaved fragment D2>5: H₁(1+)HVal(H-1F)Val(H-1F)Arg((CH₂)-1O)Thr(CH₂)OH. Theoretical observed mass (m/z) = 526.280; experimental closest peak (m/z) = 526.279. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 98
- D.27 MS spectrum of the non-natural peptide cleaved fragment D11>13: H₁(1+)HMet(S-1CH₂)SerLysOH. Theoretical observed mass (m/z) = 347.229; experimental closest peak (m/z) = 347.230. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 98
- D.28 MS spectrum of L-norleucine: H₁(1+)HMet(S-1CH₂)OH. Theoretical observed mass (m/z) = 132.102; experimental closest peak (m/z) = 132.102. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 99
- D.29 MS spectrum of L-canavanine: H₁(1+)HArg((CH₂)-1O)OH. Theoretical observed mass (m/z) = 177.098; experimental closest peak (m/z) = 177.098. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 99
- D.30 MS spectrum of DL-3-hydroxynorvaline: H₁(1+)HThr(CH₂)OH. Theoretical observed mass (m/z) = 134.081; experimental closest peak (m/z) = 134.081. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 100
- D.31 MS spectrum of 3-fluoro-DL-valine: H₁(1+)HVal(H-1F)OH. Theoretical observed mass (m/z) = 136.077; experimental closest peak (m/z) = 136.077. (Theoretical spectrum = red, experimental data = blue, and peak picking = green). 100

List of Figures

D.32 MS spectrum of L-lysine: H1(1+)HLysOH. Theoretical observed mass (m/z) = 147.113; experimental closest peak (m/z) = 147.113. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	101
D.33 MS spectrum of L-serine: H1(1+)HSerOH. Theoretical observed mass (m/z) = 106.050; experimental closest peak (m/z) = 106.049. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	101
D.34 MS spectrum of the (in-gel) cleaved peptide from modified GFP (reference control), incorporating L-norleucine = HMet(S-1CH ₂)OH: H2(2+)HLysSerAlaMet(S-1CH ₂)ProGluGlyTyrOH. Theoretical observed mass (m/z) = 432.727; experimental closest peak (m/z) = 432.726. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	102
D.35 MS spectrum of the (in-gel) cleaved peptide from modified GFP (sample), incorporating the recycled L-norleucine = HMet(S-1CH ₂)OH: H2(2+)HLysSerAlaMet(S-1CH ₂)ProGluGlyTyrOH. Theoretical observed mass (m/z) = 432.727; experimental closest peak (m/z) = 432.727. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	102
D.36 MS spectrum of the (in-gel) cleaved peptide from modified GFP (reference control), incorporating L-canavanine = HArg((CH ₂)-1O)OH: H2(2+)HValGlnGluArg((CH ₂)-1O)ThrIlePhePheOH. Theoretical observed mass (m/z) = 521.272; experimental closest peak (m/z) = 521.271. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	103
D.37 MS spectrum of the (in-gel) cleaved peptide from modified GFP (sample), incorporating the recycled L-canavanine = HArg((CH ₂)-1O)OH: H2(2+)HValGlnGluArg((CH ₂)-1O)ThrIlePhePheOH. Theoretical observed mass (m/z) = 521.272; experimental closest peak (m/z) = 521.272. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	103
D.38 MS spectrum of the b4 fragment of the isolated H2(2+)HLysSerAlaMet(S-1CH ₂)ProGluGlyTyrOH peptide from modified GFP (reference control), incorporating L-norleucine = HMet(S-1CH ₂)OH: HLysSerAlaMet(S-1CH ₂)(1+). Theoretical observed mass (m/z) = 400.255; experimental closest peak (m/z) = 400.255. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	104

D.39 MS spectrum of the b4 fragment of the isolated H2(2+)HLysSerAlaMet(S-1CH2)ProGluGlyTyrOH peptide from modified GFP (sample), incorporating the recycled L-norleucine = HMet(S-1CH2)OH: HLysSerAlaMet(S-1CH2)(1+). Theoretical observed mass (m/z) = 400.255; experimental closest peak (m/z) = 400.256. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	104
D.40 MS spectrum of the b6 fragment of the isolated H2(2+)HValGlnGluArg((CH2)-1O)ThrIlePhePheOH peptide from modified GFP (reference control), incorporating L-canavanine = HArg((CH2)-1O)OH: HValGlnGluArg((CH2)-1O)ThrIle(1+). Theoretical observed mass (m/z) = 729.389; experimental closest peak (m/z) = 729.389. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	105
D.41 MS spectrum of the b6 fragment of the isolated H2(2+)HValGlnGluArg((CH2)-1O)ThrIlePhePheOH peptide from modified GFP (sample), incorporating the recycled L-canavanine = HArg((CH2)-1O)OH: HValGlnGluArg((CH2)-1O)ThrIle(1+). Theoretical observed mass (m/z) = 729.389; experimental closest peak (m/z) = 729.389. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).	105
D.42 Fragmentation pattern of the isolated H2(2+)HLysSerAlaMet(S-1CH2)ProGluGlyTyrOH peptide from modified GFP (reference control), incorporating L-norleucine = HMet(S-1CH2)OH.	106
D.43 Fragmentation pattern of the isolated H2(2+)HLysSerAlaMet(S-1CH2)ProGluGlyTyrOH peptide from modified GFP (sample), incorporating the recycled L-norleucine = HMet(S-1CH2)OH.	106
D.44 Fragmentation pattern of the isolated H2(2+)HValGlnGluArg((CH2)-1O)ThrIlePhePheOH peptide from modified GFP (reference control), incorporating L-canavanine = HArg((CH2)-1O)OH.	107
D.45 Fragmentation pattern of the isolated H2(2+)HValGlnGluArg((CH2)-1O)ThrIlePhePheOH peptide from modified GFP (sample), incorporating the recycled L-canavanine = HArg((CH2)-1O)OH.	107
D.46 Electrophoresis gel (Agarose) of the PCR amplified mScarlet-i (a), and GFP templates (b).	108

List of Figures

- D.47 Coomassie stained (a), and fluorescence (b) images of the same SDS- PAGE protein gel used to prepare the expressed natural GFP (sample) (1), the modified GFP incorporating L-norleucine and L-canavanine (reference control) (2), and the modified GFP incorporating the recycled L-norleucine and L-canavanine (sample) (3) for the proteomic characterization. Modified GFP incorporating L-norleucine and L-canavanine (reference control), with Lys-BODIPY-FL inclusions (4), and modified GFP incorporating the recycled L-norleucine and L-canavanine (sample), with Lys-BODIPY-FL inclusions (5). Fluorescence image (c) of an additional SDS-PAGE protein gel, run in the same experimental conditions as the gel shown in (a-b), with a FluoroTect™Green_{Lys} tRNA in nuclease-free water solution (x). This is a control experiment to show that the fluorescent bands visible in (4), and (5) are not protein impurities but they are exclusively due to tRNA-Lys-BODIPY-FL. 108
- D.48 (a) Photograph of mScarlet-i, and GFP solutions obtained by NaCre from a mixture composed of glucagon, β -lactoglobulin A, and silk fibroin, purified by 6xHis tag at C-terminus, and inspected by using Invitrogen™E-Gel™Safe Imager™(emission max of the blue LED = 470 nm). The image has been taken in the lab by using an iPhone Xs. (b) Image of the Coomassie stained SDS-PAGE protein gel of the purified mScarlet-i (2-3), and GFP (4-5) shown in (a). For mScarlet-i, the calibrant expressed in *E. coli* cells has been added to the gel (1); red-fluorescent proteins are known to produce cleaved fragments when treated at high temperature with denaturants¹¹⁶. The GFP protein bands (4-5) are less intense with respect to mScarlet-i ones (2-3) since the GFP sample was buffer exchanged to prepare it for the second cycle of NaCre, as described in Appendix B.8. The gel has been used to prepare the samples prior to proteomic characterization. 109

List of Tables

4.1	Overview of the depolymerization and expression efficiencies for key experiments in this study. Minima are colored in blue (depolymerization) and red (expression). * Calculated as the ratio between the amount of each AA (nmol) and the number of its incorporations inside a single protein chain, see Table C.5. ** n.a. not assessable because E is present in the TX-TL system as Potassium glutamate (buffer), see Appendix B.7.	33
5.1	Frangmentation pattern of the isolated peptide H3(3+)For[L-norleucine] [3-fluoro-L-valine] [3-fluoro-L-valine] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-norleucine] [Serine] [Lysine]OH.	45
5.2	Frangmentation pattern of the isolated peptide H3(3+)For[L-norleucine] [3-fluoro-L-valine] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-norleucine] [3-fluoro-L-valine] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [Serine] [Lysine]OH.	46
C.1	Primary sequences of the depolymerized proteins.	78
C.2	Linear templates (gBlocks) for expressing mScarlet-i (a), GFP (b), and CDO (c) in PUREFrex™.	79
C.3	Forward (a) and reverse (b) primers for the PCR amplification of the gBlocks encoding mScarlet-i, and GFP. Forward (c) and reverse (d) primers for the PCR amplification of the gBlock encoding CDO.	80
C.4	Constructs for expressing mScarlet-i calibrant (a), and GFP calibrant (b) into BL21 (DE3) cells, after cloning them into the pET29b(+) vector.	80

List of Tables

C.5 Primary sequences of the expressed proteins.	81
--	----

Structure of the thesis

The thesis is divided in three parts: the introductory chapters (1-3), the chapters describing the results (4-5), and the conclusions (6). A description of the structure of the thesis is presented below.

In **Chapter 1** the problem of plastic pollution is presented, and the reader is guided step by step through the main methods that are currently used to recycle (or repurpose) synthetic polymer based materials. The techniques are described in detail by providing examples of the main materials they're applied to, as well as highlighting the advantages, and drawbacks of each approach. A short overview about biodegradable polymers follows. The chapter continues with a key discussion presenting the strategies used by Nature to recycling natural polymers. The chapter ends with an introduction to sequence-defined polymers, and a description of the main chemical, and biological methods that are currently used to synthesize them.

In **Chapter 2** the recycling approach defined in this thesis is introduced, and discussed by comparing it with current recycling techniques.

In **Chapter 3** a detailed description of transcription-translation in a purified cell-free system is presented.

In **Chapter 4** the main experimental results are discussed. This chapter is an extract of the manuscript entitled *Nature-inspired Circular-economy Recycling (NaCRe) for Proteins: Proof of Concept* that is available as a preprint (<https://www.biorxiv.org/content/10.1101/2020.09.23.309799v3>, doi: 10.1101/2020.09.23.309799 version 3).

In **Chapter 5** a set of preliminary experiments on the cell-free expression of proteins containing

List of Tables

several non-natural amino acids is discussed.

In **Chapter 6** the conclusions of the thesis are summarized, and the envisioned future developments are presented.

A detailed description of the materials, and methods, as well as a list of the additional data are reported in the **Appendix II**.

1 Introduction

Plastics is commonly used to indicate synthetic polymer-based materials. Polymers are chemical species composed of several structural units connected by covalent bonds¹. After a century of “polymer science”², plastics is used nowadays for fabricating materials for everyday needs (*e.g.* packaging), as well as for high-performance applications due to its competitive price, easy manufacturing, excellent durability, and high safeness³. Nevertheless, the end-of-life of plastics is still an open question⁴. This chapter begins with a brief overview on plastic pollution, followed by a detailed description of the strategies that are currently used to handle plastic waste streams. It continues with a comparison between the methods used nowadays to recycle man-made polymers, and the approaches used by Nature to recycle its natural polymers. The chapter ends with a detailed description of the state-of-the-art for producing a special class of polymers (*i.e.* sequence-defined polymers), by using chemical, or biological strategies respectively.

1.1 Plastic pollution

Plastics is resilient, and durable⁴. These properties are key when designing novel materials, although they may turn to be an issue if the end-of-life of such materials has not been initially planned. This is the case for synthetic polymer-based materials⁴. Plastics debris persist in the environment for centuries³. They accumulate in the oceans forming islands of plastics (*e.g.* the well-known Great Pacific Garbage Patch), driven by converging surface currents⁵.

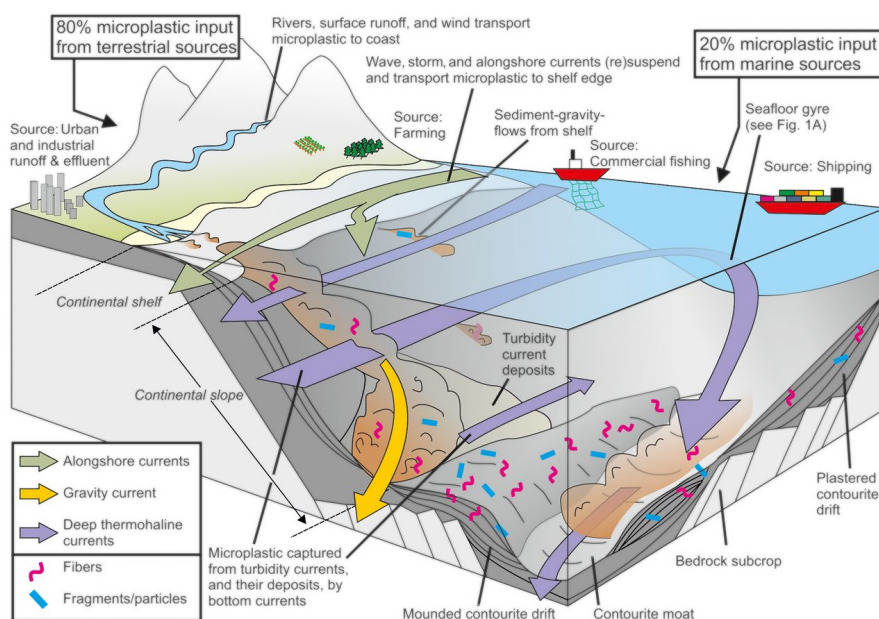


Figure 1.1 – Schematic illustration of microplastics accumulation in the deep sea. The figure has been reproduced from reference 6, with permission from AAAS.

Nevertheless, what we see is only a minimal fraction of the real amount of plastics that is present into the environment nowadays⁶. Plastic debris are fragmented by the action of light, and other weathering processes, into smaller particles such as beads, and fibers, commonly referred as “microplastics”^{7,8}. Such microparticles, originated from the degradation of plastic materials, have been found in soil, rivers, lakes, and oceans (Figure 1.1), and are dangerous for many organisms due to their size, and chemical composition^{6,8–10}.

This situation is not going to improve in the next years. The world population is growing¹¹, and the production of synthetic polymer-based materials is estimated to increase remarkably¹². In 2050 the weight amount of plastics present in the ocean is expected to become higher than fish¹². Together with food¹³ and energy supply¹⁴, the handling (*i.e.* production, use, and disposal) of such an enormous amount of synthetic polymer-based materials is one of the greatest sustainability challenges that humanity has to face^{3,15}.

Thus, the development of efficient recycling strategies able to address the end-of-life of synthetic polymeric materials is a priority for the planet, as well as an important economic opportunity for a new market¹⁶.

1.2 Recycling of plastics

The methods currently available to recycle plastics are severely limited¹⁶. In the followings a comprehensive overview of the techniques that are used nowadays is discussed.

1.2.1 Mechanical recycling

Mechanical recycling is the most used recycling technology to treat large scale plastic waste streams¹⁶. It is based on re-processing the polymers at high temperature by extrusion, and moulding¹⁷. Hence it is used to treat thermoplastic materials exclusively¹⁶.

If uncontaminated plastic wastes are recycled, the method is named “primary” recycling. This is the case when post-industrial waste, such as process scrap, or uncontaminated post-consumer waste is re-processed in a closed-loop fashion, to produce the same product as the original material¹⁸. In case of post-consumer waste, the materials are collected, sorted, ground, washed, and processed^{17,18}.

If the waste stream contains unknown post-consumer materials, the method is named “secondary” recycling. In this scenario, the materials are typically recycled into products of lower value with respect to the original materials (“downcycling”)¹⁸. The materials are collected from municipal solid wastes, identified and separated according to size, density, electrostatics, wettability, colour, and chemistry by using sieving, magnetic, triboelectric, flotation, and spectroscopic techniques; they are then ground, washed, and pelletized^{19,20}.

By mechanical recycling we recycle polyethylenes (PE), and polyethylene terephthalate (PET) that constitute 37%, and 9% of the annual plastic production¹⁶. However, these processes are costly, time and energy consuming, and often produce materials whose quality is lower than the original ones¹⁶.

Among the materials that are currently recycled by mechanical recycling²⁰, low density PE (widely used for packaging) can withstand several (~40) extrusions before showing lower mechanical properties, instead PET (highly used for bottles) suffers loss of ductility after a few (~3) cycles¹⁸. PET suffers significant molecular weight reduction due to re-processing at high temperature²¹, therefore bottles are commonly downcycled into fibers for textiles¹⁸.

Indeed processing at high temperatures (either in the case of “primary” or “secondary” recy-

cling) involves heating, and mechanical shearing of the polymer melt that promotes degradation pathways such as oxidation, and hemolytic scission of carbon-carbon backbone bonds. The so formed radicals generate chain scission, crosslinking, and branching²⁰. Finally impurities (of different types of plastics) during processing contribute to downgrade the properties of the output of mechanical recycling due to lack of miscibility between different polymers^{17,20}.

1.2.2 Compatibilization

As discussed above, the process of separating different plastic materials is complex, and expensive¹⁶. Mechanically recycling blends of polymers without the need of separating them is very convenient. This is true especially for polymers that are highly used for packaging, such as PE and isotactic polypropylene (*i*PP)²². However PE, and *i*PP are not miscible, *i.e.* they produce a brittle material if mechanically recycled together²². This is the case for several polymeric materials since the entropic contribution of mixing large macromolecules is almost negligible^{20,22}.

Compatibilization means supplementing the heterogeneous polymer blend with additives, commonly referred as compatibilizers. Compatibilizers act on the enthalpic contribution of mixing by lowering the interfacial tension between the different phases, improving adhesion, and stabilizing the resulting morphology during processing at high temperature²⁰. Examples are block, or graft co-polymers (*e.g.* PE-*i*PP copolymers), polymers containing polar groups (*e.g.* PMMA), and reactive polymers (*e.g.* PP grafted with maleic anhydride)²⁰. Recently multiblock PE-*i*PP copolymers with precise control on the length of the blocks have been synthesized. These polymers improve compatibilization in PE-*i*PP blends by co-crystallizing in the polymers' lamellae, or by bridging multiple phases²³.

1.2.3 Chemical recycling

Chemical recycling, also named "tertiary" recycling¹⁸, is based on the depolymerization of the polymeric material into its constituent monomers, under controlled conditions, in order to re-polymerize either the same virgin quality material²¹, or a different (co)polymer^{24,25}. As a result, the material is continuously kept in circulation, and the waste becomes a valuable resource for synthesizing new plastics⁴, reducing the burden on the planet³.

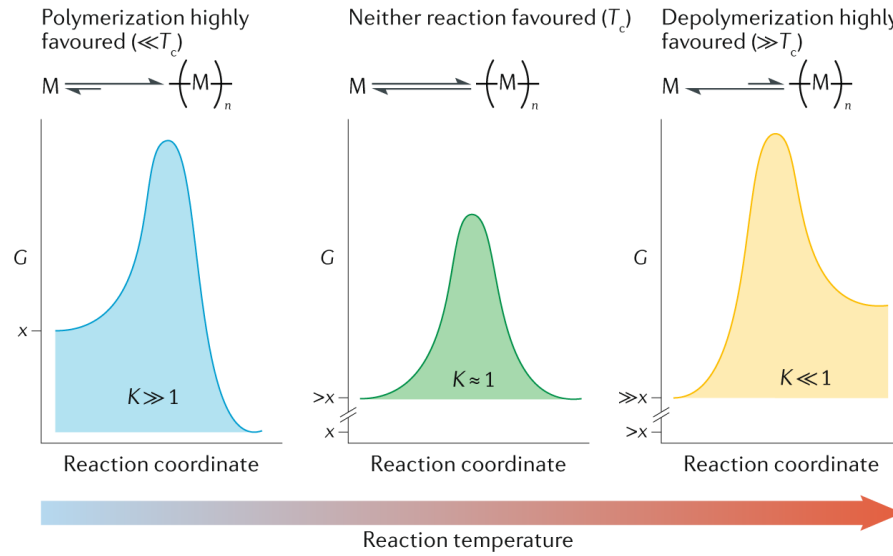


Figure 1.2 – Schematic illustration representing the plot of the Gibbs free energy as function of the reaction coordinate for the polymerization reaction of the monomer M at different temperatures ($T \ll T_c$, $T = T_c$, and $T \gg T_c$, where T_c is the equilibrium temperature at which $\Delta G_P = 0$). x is the free energy of the monomers that is function of the reaction temperature. The figure has been reproduced from reference 26, with permission from Springer Nature.

However, this approach is challenging due to thermodynamics that limits its applicability to a small set of polymers²⁶. In detail, the variation of the Gibbs free energy (ΔG_P) for a polymerization reaction can be written as:

$$\Delta G_P = \Delta H_P - T\Delta S_P \quad (1.1)$$

where (ΔH_P), and (ΔS_P) are the enthalpy, and the entropy variations for the polymerization reaction, and T is temperature. At equilibrium $\Delta G_P = 0$, and the polymerization system is at its critical temperature (T_c):

$$T_c = \frac{\Delta H_P}{\Delta S_P} \quad (1.2)$$

For the polymerization reaction to occur $\Delta G_P < 0$. Since for the majority of the polymerizations both ΔH_P and ΔS_P are negative²⁶, polymerization is favoured at $T < T_c$, de-polymerization is favoured at $T > T_c$, and T_c is the ceiling temperature (Figure 1.2).

In order to obtain complete depolymerization to monomer, the material needs to be heated at $T \gg T_c$. This is the reason why many polymers, such as polyethylene, cannot be recycled by chemical recycling. In fact, in the case of PE polymerization, the value of $\Delta G_P \ll 0$. Therefore,

Chapter 1. Introduction

the temperature required to depolymerize PE into ethylene is too high, and depolymerization produces mainly gases, waxes, and char²⁶. Finally, kinetics should be also considered, as depolymerization occurs if reactive chain ends are present²⁶.

Among the most common materials, whose depolymerization to monomer is thermodynamically favourable, nylon²⁷, PET^{28,29}, and some polyesters^{30–32} are listed below.

Nylon-6 is depolymerized into ϵ -caprolactam by heating nylon at 300 °C in ionic liquids, for a few hours, and collecting the monomers by distillation. The addition of *N,N*-dimethylaminopyridine catalyst increases the yield of depolymerization²¹. However, recycling of ϵ -caprolactam into virgin quality nylon is commercially not convenient because of the production scale of such material (few million tonnes per year), the difficulties in separating/recovering ϵ -caprolactam, and the volatility of market prices for ϵ -caprolactam²⁶.

Many strategies have been developed for recycling PET chemically³³, since cleaving the ester bond of PET is easier with respect to the carbon-carbon bond of polyolefins²⁶. One of the most established possibilities is depolymerizing PET by heating PET at 190 °C in excess of ethylene glycol, for a few hours, in presence of 1,5,7-triazabicyclo[4.4.0]dec-5-ene (TBD), to obtain bis(2-hydroxyethyl) terephthalate (BHET)²⁶. BHET is purified by recrystallization, and used to synthesize new virgin quality PET^{4,21}. Mixtures of TBD and methanesulfonic acid (MSA) as efficient protic ionic complexes have also been recently developed to catalyze the glycolysis of PET²⁹. Enzymatic depolymerization of PET is also possible⁴. Recently an highly efficient PET hydrolyse has been obtained by computer-aided enzyme engineering³⁴.

As discussed in section 1.2.1, PET suffers reprocessing at high temperature by mechanical recycling. Even if chain extenders can be used²⁶, the possibility to recycle PET by several chemical recycling strategies is encouraging. The commercial viability of such methods is still unclear⁴, however several efforts to develop competitive closed-loop recycling of PET have been put in place in the last decade²⁶.

Extensive research has been performed to develop polyesters that can be recycled multiple times by chemical recycling³⁵. As discussed above, polymers with low T_C can be quantitatively, and selectively depolymerized into monomers easily, by heating the material at relatively low temperatures²⁶. However, the production of such polymers requires industrially demanding

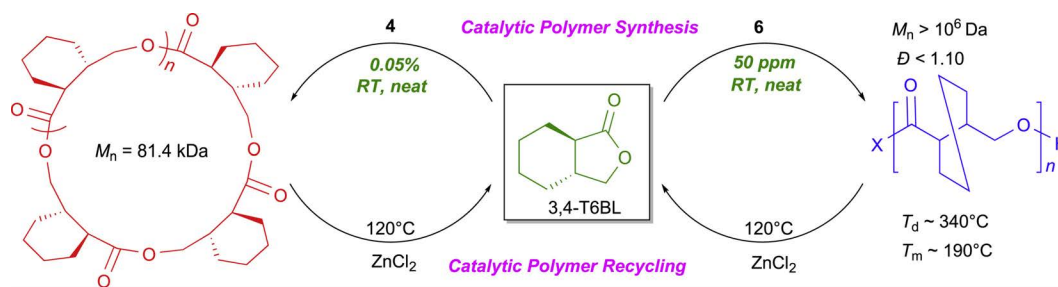


Figure 1.3 – Schematic illustration of the ring opening polymerization of 3,4-T6GBL to obtain linear, and cyclic poly(3,4-T6GBL) (indicated as “Catalytic Polymer Synthesis”), and chemical recycling to monomer of poly(3,4-T6GBL) to achieve 3,4-T6GBL (indicated as “Catalytic Polymer Recycling”). 3,4-T6GBL is used to obtain virgin quality poly(3,4-T6GBL) in a potentially “infinite-recyclable” fashion. The figure has been reproduced from reference 35, with permission from Elsevier.

conditions due to low synthesis temperatures, slow reaction rates, long reaction times, and moreover they exhibit limited thermostability, and crystallinity^{26,31}. This is the case for poly- γ -butyrolactone (PGBL). In fact, the low strain energy of the five-membered ring of γ -BL (GBL) leads to a small negative value of ΔH_p with respect to ΔS_p for the polymerization to occur³⁰. PGBL has been successfully polymerized recently by working at low temperatures, using kinetically strong catalysts, and modulating the reaction conditions so to shift the reaction towards propagation^{30,32}.

Starting from the five-membered ring structure of GBL, the 3,4-T6GBL monomer has been designed by trans-fusing a cyclohexyl ring at the α , and β positions of GBL³¹. This allows for preserving the chemical recyclability of PGBL, improving the thermodynamic polymerizability. Linear, and cyclic poly(3,4-T6GBL) has been synthesized by using Yttrium, Zinc, and Lanthanum complex catalysts, at room temperature, and achieving high polymerization yields, and low dispersity³¹. Both linear, and cyclic polymers are depolymerized back into 3,4-T6GBL by heating above 300 °C for a few hours (thermolysis), or at 120 °C by adding $ZnCl_2$ (chemolysis). 3,4-T6GBL is used to produce virgin quality poly(3,4-T6GBL) in a closed-loop fashion (Figure 1.3)³¹.

Recently, the synthesis of poly(4-carbomethoxyvalerolactone) that can be chemically recycled to monomer by a divergent depolymerization leading to 4-carbomethoxyvalerolactone, or 2-methyleneglutaric monomethyl ester has been shown²⁴. Polyesters (polyglycolic acids,

Chapter 1. Introduction

PGAs) that can be recycled into co-polymers (polyGAcoBL) by transesterification reactions in GBL have been reported²⁵.

So far the mechanical, or chemical recycling of thermoplastic polymers have been discussed solely. Thermosets are chemically, or thermally-resistant polymers that are chemically cross-linked in order to withstand harsh operating conditions. These materials are used in many high-tech applications, from automotive to electronics. However, they are difficult to recycle because they can't be reprocessed by melting, or solution processing³⁶.

To fill the need of recyclable thermoset polymers, several thermo-reversible materials based on Diels-Alder reactions have been developed³³. Moreover, chemically crosslinked polyurethane foams based on poly(β -methyl- δ -valerolactone) (PMVL) that can be depolymerized to achieve MVL upon heating have been studied³⁷.

However, since the majority of crosslinked polymers are designed to be temperature-stable, thermosets that can be depolymerized to monomers by lowering the pH have been developed³⁸. Such materials are obtained by low temperature ($\sim 50^\circ\text{C}$) polymerization of 4,4'-oxydianiline with paraformaldehyde, followed by cyclization to poly(hexahydrotriazine)s (PHTs) at $\sim 200^\circ\text{C}$. PHTs, and PHT-composites show outstanding mechanical properties, and solvent resistance at $\text{pH} > 3$. If pH is further lowered ($\text{pH} < 2$), PHTs are depolymerized into their monomers³⁸.

As further example of chemical recycling, bacteria such as *Ideonella sakaiensis* 201-F6 able to depolymerize PET into terephthalic acid, and ethylene glycol have been discovered³⁹.

1.2.4 Repurposing or upcycling

Repurposing, often named "upcycling" or valorization, is based on recycling a polymer into a different material that has a higher economic value. The properties, and the market of such material can be completely different from those of the initial polymer⁴.

As discussed in sections 1.2.1, and 1.2.3, PE can be successfully recycled by mechanical recycling, but not by chemical recycling. However, PE has been repurposed into liquid fuels

(diesel), and waxes by heating at 175 °C, for one day, in presence of light alkanes, and catalysts⁴⁰. The process is based on dehydrogenation of PE, and alkanes by pincer type Iridium complexes to form unsaturated olefins, Rhenium catalyzed cross metathesis of the so formed olefins, and final hydrogenation by Iridium catalyst⁴⁰. Recently, PE has been also upcycled into liquid long-chain alkylaromatics that are used as surfactants, lubricants, refrigeration fluids, and insulating oils⁴¹. The process runs at moderate temperature (280 °C), for one day, in presence of Pt/ γ -Al₂O₃ catalyst⁴¹.

Polycarbonate (PC) is a thermoplastic polymer mostly used for optical applications. The chemical recycling of PC is not efficient²⁶. Repurposing of PC into high-performance poly(aryl ether sulfone) (PSU) has been successfully achieved⁴². The applications of PSU materials range from producing reverse osmosis and water purification membranes, to fabricating medical equipments⁴². PC is treated at (190 °C), for 18 hours, in *N*-cyclohexyl-2-pyrrolidone, with K₂CO₃, to form reactive phenoxides that are in turn polycondensed with 4,4'-difluorodiphenylsulfone to produce PSU⁴².

These works are examples of manufacturing of value-added materials from plastic wastes⁴ *i.e.* generating new virgin quality materials without depleting petrochemical sources^{3,41}.

1.2.5 Incineration

According to data from 2015, only 14% of plastic wastes is collected for recycling²⁶. Most of plastics (40%) is landfilled, or leaked into the environment (32%)²⁶, *i.e.* the materials' value is lost²¹. Moreover, the majority of landfilled plastics is non-degradable, or degradable very slowly²¹.

Incineration, also named as "energy recovery"¹⁸, or "quaternary" recycling¹⁹ is based on producing value from the heat generated by burning plastic mixtures¹⁸. Approximately 14% of plastic wastes is incinerated²⁶. However, incineration suffers low efficiency in energy recovery²¹, and most importantly generation of green-house gases^{18,21} (most of plastics' end-of-life CO₂ emissions)⁴³, and toxins¹⁸.

When the thermal decomposition of polymers is run in oxygen-free conditions, in order to produce chemicals, the process is named "pyrolysis". The products of pyrolysis can be directly

used as fuels, or converted into monomers by cracking. In some cases (*e.g.* poly(methyl methacrylate), polystyrene), pyrolysis achieves the recovery of monomers⁴⁴.

1.3 Biodegradable polymers

A different approach to recycling of plastics is the use of biodegradable polymers *i.e.* polymers that, placed in a bioactive environment, degrade into CO₂, CH₄, water, biomass, and humic matter through the action of either microorganisms, or non-enzymatic processes⁴⁵. Examples of this polymers are polylactic acid (PLA), polycaprolactone (PCL), poly(butylene-succinate), poly(hydroxy-butyrates), polymers from starch, and cellulose⁴⁵. These polymers have been used for many applications ranging from packaging to hygiene products^{21,45}. However, bio-“degradation” basically means disassembling the material into components non-harmful for the environment, that is substantially different from depolymerization (*i.e.* recovery of the monomers), or recycling. In fact the materials’ value, that is the energy used to synthesize the resin, and to manufacture the material, is completely lost upon biodegradation^{4,21}. Furthermore, concerns have been raised on the slow degradation rates²¹, as well as on the intermediates of depolymerization⁴⁶. Finally, it is not obvious if the disposal of large quantities of biodegradable polymers into the environment is sustainable. In fact the excessive accumulation of depolymerization products, albeit non-harmful, may produce unseen environmental problems^{4,21}.

The majority of biodegradable polymers are also bio-based that is produced from renewable carbon resources⁴⁶. However, bio-based polymers are not necessarily bio-degradable. For example cellulose acetate (derived from polysaccharides), or bio-PE (derived from glucose fermentation) are not biodegradable⁴⁶. Furthermore bio-based polymers are quite resource-intensive because they require land for producing the raw materials⁴⁷, and they are energy-demanding for processing⁴.

1.4 Recycling of natural polymers: a lesson from Nature

Several strategies to recycle synthetic polymers have been discussed so far. During the last decade many steps forward have been achieved, mainly in the fields of chemical recycling (section 1.2.3), and repurposing (section 1.2.4). However, several limitations are still present. In order to limit plastic pollution one possibility could be replacing plastics with materials that are recycled more efficiently (*e.g.* glass). However, a glass bottle has a carbon footprint that is higher than a plastic one⁴⁸. Hence, replacing PET with glass would basically mean polluting the air more.

Another possibility is pausing to observe how Nature handles natural polymers. Natural polymers are sustainable because Nature produces polymer-based materials at a synthesis rate that is commensurate with their service life, and degradation. For example, wood is mainly composed of polymers such as cellulose, hemicellulose, and lignin⁴⁹. Wood takes long time to grow, remains in use for a long time, and biodegrades into the environment slowly (especially lignin)⁴⁹.

On the contrary, other natural polymers, such as proteins, and nucleic acids, are produced fast, but they are also recycled fast. This is possible because proteins (and nucleic acids, NAs) are seldom biodegraded (such as wood), but are mostly depolymerized into their monomers, *i.e.* the 20 proteinogenic amino acids (AAs). These monomers are in turn re-polymerized to produce a new polymer that can be very different from the parent ones.

This is the case because proteins (and NAs) are sequence-defined polymers (hereafter referred as SDPs)⁵⁰, that is their exceptional structural, and functional diversity depends on the sequence of their monomer building blocks, and not on their chemical diversity⁵¹.

Thus Nature is teaching us that the secret towards polymers' sustainability is the circularity in the materials' use, *i.e.* polymers don't undergo complete degradation into non-reusable compounds, but are depolymerized into building blocks that in turn are re-assembled to produce a different material.

For this approach to work, polymers need to be sequence-defined. Hence, a comprehensive overview of SDPs is presented in the following section.

1.5 Sequence-defined polymers

Sequence-defined polymers are macromolecules that have a “perfectly defined primary structure”⁵⁰, that is the control over the absolute position of each monomer building block in the polymer chain is achieved.

Nature is unrivalled in the ability of synthesizing SDPs. Cells are machines that produce proteins, and NAs continuously. To perform this task Nature takes advantage of “templated”, and “chain-walking” mechanisms⁵².

For protein synthesis, named “translation”, the genetic information templated in the messenger RNA (mRNA) is used by the ribosomes to polymerize the amino acid monomers, by progressing along the the mRNA chain. In detail, the process starts with the formation of an mRNA-ribosome complex. The complex is provided with AAs by the transfer RNAs (tRNAs). An AA-tRNA pair is formed by aminoacyl-tRNA synthetase (AARS) enzymes that acylate the AAs onto tRNAs. The AA-tRNAs bind the mRNA chain by Watson-Crick base-pairing between triads of nucleic acids (codon-anticodon pairing). The polymerization happens in a stepwise fashion in which pairs of AA-tRNAs react in the mRNA-ribosome complex to form one peptide bond in each step. After the amidation, the mRNA-ribosome complex moves forward along the mRNA guide, the reacted tRNA is removed, and a new AA-tRNA is introduced⁵³. The

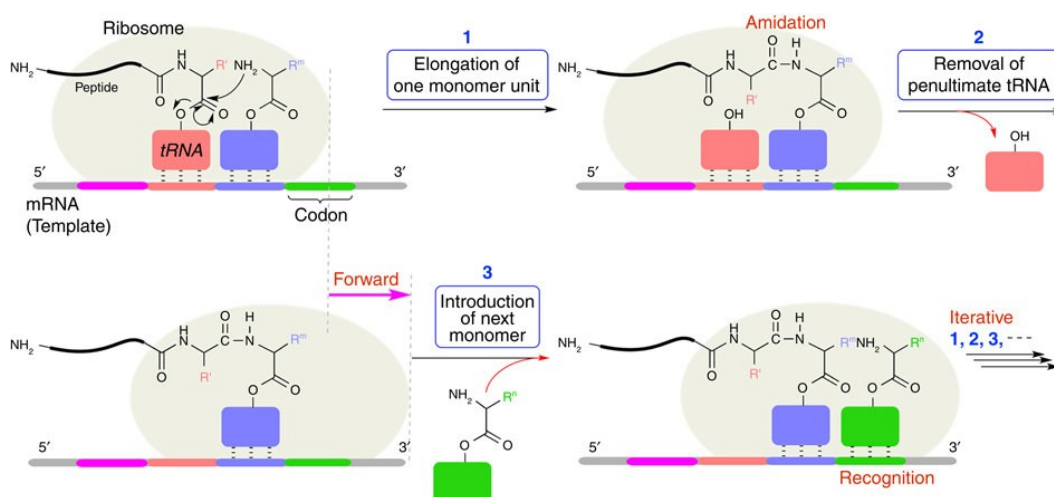


Figure 1.4 – Schematic illustration of protein polymerization. The elongation step (1) through which one peptide bond is formed, the removal of the reacted tRNA (2), and the introduction of a new monomer (3) are sketched. The figure has been reproduced from reference 53, with permission from Springer Nature.

process continues in iterative manner, as schematized in Figure 1.4.

Similarly, for the DNA synthesis, named “replication”, the information contained in the DNA chain is copied by polymerase enzymes that polymerize the nucleoside triphosphate monomers, by moving along the unwound DNA template.

Both processes work starting from a mixed pool of monomers that are provided to ribosomes, or polymerases by tRNAs, or simply by Watson-Crick base-pairing respectively.

Taking inspiration from Nature men has developed several methodologies to synthesize SDPs, either by chemistry, or by directly harvesting the biological machineries used to produce proteins, and NAs. In the followings, the most used strategies are summarized.

1.5.1 Sequence-defined polymers through chemistry

As discussed in section 1.5, the mechanism used by Nature to perform protein translation is based on iterative stepwise reactions. Chemists took inspiration from this method, and developed iterative polymerization techniques based on coupling self-reacting bifunctional monomers by means of protection-deprotection cycles⁵¹. The field was revolutionized by Merrifield that developed the so called Solid Phase Peptide Synthesis (SPPS) in 1963⁵⁴.

Starting from the first tetrapeptide synthesized by SPPS⁵⁴, more sophisticated techniques have been engineered, such as automated fast-flow peptide synthesis (AFPS). By using AFPS biologically active protein chains as long as 164 AAs has been recently synthesized, without the need of any ligation⁵⁵.

Taking advantage of the iterative coupling of monomers by means of protection-deprotection reactions, a variety of SDPs have been synthesized ranging from polypeptoids⁵⁶, to polyphosphates⁵⁷.

Another milestone has been set by Lutz that replaced the use of protection-deprotection cycles with chemoselective strategies, in combination with alternating monomers⁵⁸. An example of this methodology is the synthesis of SD poly(alkoxyamine amide)s by coupling a monomer containing an acyclic symmetric acid anhydride and alkyl bromides with a monomer containing a nitroxyl radical and a primary amine⁵⁹. The reaction is schematized in Figure 1.5. Other SDPs such as poly(alkoxyamine phosphodiester)s⁶⁰, and polyurethanes⁶¹ have been synthesized by using this approach.

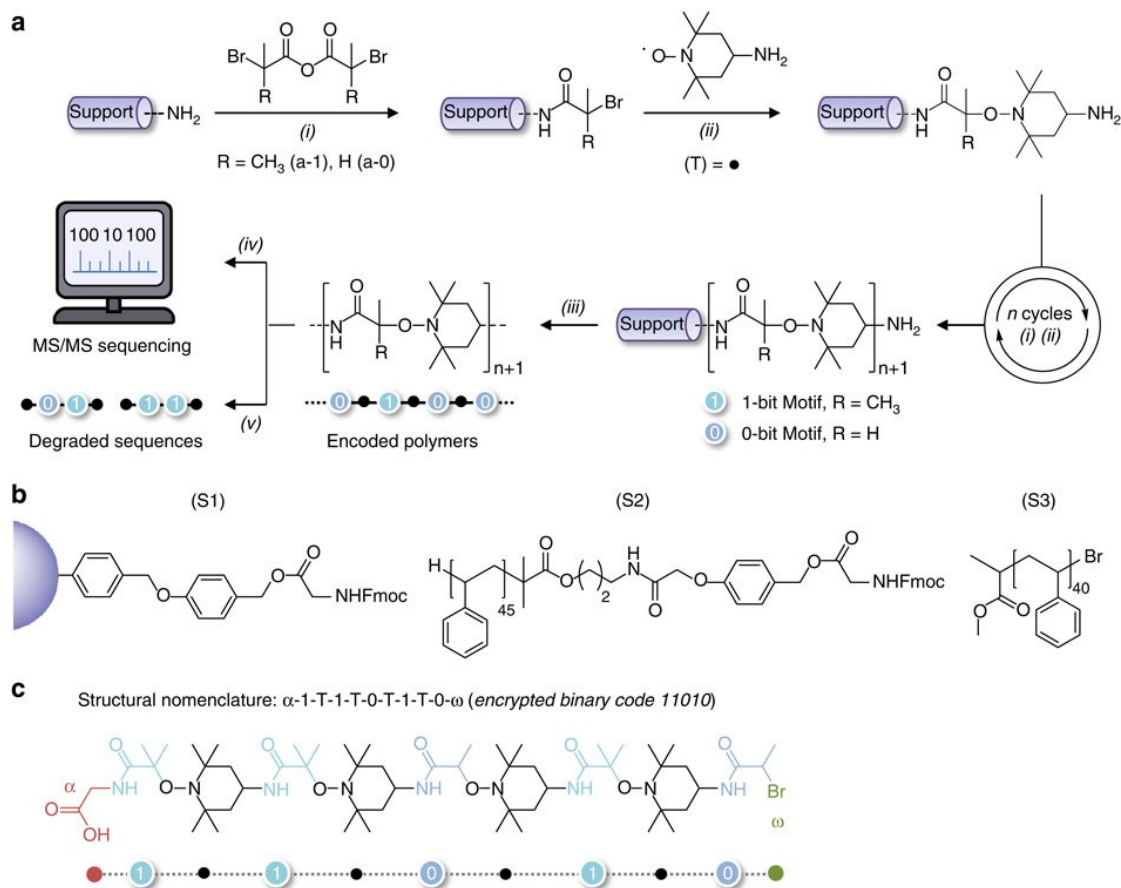


Figure 1.5 – Schematic illustration of the synthesis of SD poly(alkoxyamine amide)s (a). Structure of the solid-phase, or soluble supports used (b). Chemical structure of a model SD oligomer achievable by using this method (c). The figure has been reproduced from reference 59, with permission from Springer Nature.

Furthermore, techniques achieving high control over the polymer stereoconfiguration⁶², high yields and scalability⁶³, and photoligation⁶⁴ have been developed.

Moreover, elegant strategies to imitate proteins' polymerization machinery have been studied. In detail, a rotaxane-based small-molecule machine able to synthesize peptides by mimicking the chain walking mechanism of the ribosome along the mRNA template has been developed⁶⁵. A functionalized macrocycle (ribosome analog) polymerizes the AAs, weakly attached to a rigid thread (mRNA analog), by transacylating, and ligating them at the end of a functionalized arm (growing protein analog)⁶⁵. Another example of ribosome imitation has been achieved by using an inimer (initiator-monomer) that is able to polymerize vinyl oligomers⁵³.

Furthermore, high molecular weight SD polyethylene glycol (PEG), β -peptides, α -(D)-peptides, and copolymers of them have been obtained by emulating the tRNA-mRNA binding mechanism⁶⁶. This has been achieved by using a macrocycle (tRNA analog) composed of a peptide nucleic acid adapter (PNA), and a polymer building block (AA analog) separated by cleavable linkers. The macrocycle base-pairs with a designed DNA template (mRNA analog), the polymeric blocks are coupled by copper-catalyzed azide-alkyne cycloaddition (CuAAC), and the linkers are cleaved to release the full-length polymer⁶⁶.

1.5.2 Sequence-defined polymers through biology

An alternative strategy to produce SDPs is based on directly harvesting the biological machineries used to produce proteins, and nucleic acids.

This can be achieved by expressing “recombinant” proteins in “host” cells. Such cells, mostly from bacteria (*E. coli*), but also from yeast, fungi, or algae, are genetically modified organisms able to polymerize the protein of interest⁶⁷. The first animal protein successfully produced in *E. coli* was human insulin, obtained in 1978⁶⁸.

Recombinant protein expression is efficient, and very well established nowadays⁶⁷. However, this technique makes use of the whole cell machinery. Indeed, it is possible to remove the cell wall barriers, directly manipulate the reaction conditions, and avoid viability constraints by using cell-free protein synthesis (CFPS)⁶⁹. This methodology can be technologically performed in a crude extract, that is a cell lysate^{70,71}, or in a purified system composed solely of the essential elements for protein synthesis⁷². Lysates are not expensive, and produce proteins in high yields; purified systems are more flexible, and controllable⁶⁹. Among the purified systems, the protein synthesis by recombinant elements (PURE) has been developed by Ueda in 2001⁷². This system contains the *E. coli* elements necessary for transcription, and translation⁷³, is scalable^{74,75}, and can be adapted to express different types of proteins⁷⁶.

Nucleic acids (DNA) can also be “amplified” (*i.e.* polymerized) outside living organisms by using the polymerase chain reaction (PCR) invented by Mullis in 1988⁷⁷. This technique is based on the hybridization of two short oligonucleotides, named “primers”, on a target DNA template sequence by Watson-Crick base-pairing. A polymerase enzyme “extends” the primer sequence by polymerizing free nucleotides, while moving along the template strand. A precise

Chapter 1. Introduction

thermal cycle is needed to “denature” the DNA, “anneal” the primers, and “extend” the DNA sequence. The so polymerized DNA sequences are complementary to the primers, hence the process results in the exponential production of the targeted template⁷⁷.

The strategies described so far are based on the polymerization of proteins, or nucleic acids by using 20 proteinogenic, or 4 nucleoside monomers respectively. By using such a small set of building blocks, Nature is able to achieve a remarkable variety of functional SDPs⁷⁸.

However, the possibility to combine the advantages of using biological machineries with an expanded repertoire of monomers has been deeply investigated.

In detail, much work has been done to incorporate unnatural amino acids (UAAs) into protein sequences. Site-specific incorporation of a large set of UAAs has been achieved *in vitro* by Schultz who developed the so called site-specific mutagenesis in 1989⁷⁹. This approach is based on the degeneracy of three codons, named “stop codons”, in the mRNA chain. Such codons are triads of nucleic acids encoding the signal for terminating the protein synthesis. Since only one stop codon is necessary for such task, the remaining two stop codons can be used to program the incorporation of UAAs. Specifically, a mutation in the DNA is created so to produce a stop codon in the mRNA sequence. A tRNA bearing the anticodon complementary to the stop codon is chemo-enzymatically acylated with the UAA of interest to produce an UAA-tRNA pair⁸⁰. The so obtained UAA-tRNA hybridizes with the mRNA in the mRNA-ribosome complex, at the stop codon position, and the UAAs is incorporated into the growing protein chain⁷⁹.

Site-specific mutagenesis has been further developed by Chin who evolved an orthogonal synthetase-tRNA pair⁸¹, as well as an orthogonal mRNA-ribosome complex (ribo-X)⁸². The combination of such evolved systems allows for the efficient incorporation of a single UAA, at multiple positions, *in vivo*.

In order to expand the number of UAAs incorporations into the same protein sequence, techniques based on codon-anticodon pairs composed of quadruplets of nucleic acids have been introduced *in vitro*⁸³, and further evolved *in vivo*^{84,85}.

Futhermore, *in vitro* reprogramming of the genetic code has been achieved by using flexizymes (*i.e.* flexible tRNA acylation ribozymes) that facilitate the preparation of the UAA-tRNA pair⁸⁶. By using this approach more than 300 UAAs have been acylated⁸⁷, and a short SD polyester

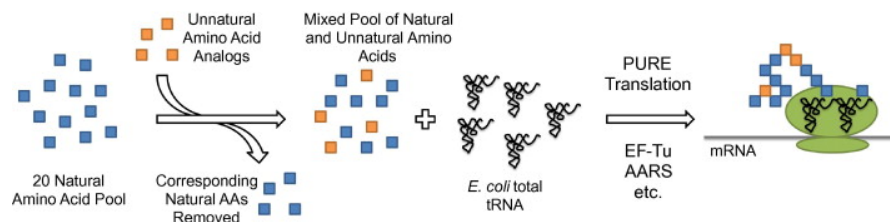


Figure 1.6 – Schematic illustration of non-natural protein translation in PURE, by solely replacing the proteinogenic AAs (blue squares) with their correspondent UAA analogs (orange squares). The figure has been reproduced from reference 92, with permission from Elsevier.

has been ribosomally polymerized⁸⁸.

Nature evolved aminoacyl-tRNA synthetases to prepare AA-tRNA pairs by discriminating amino acids among the small set of proteinogenic AAs⁸⁹. Indeed, Szostak discovered that a large set of UAAs can be enzymatically charged onto tRNAs by using the natural AARS⁸⁹. This is possible when UAAs are close “analogs” of their correspondent natural AAs⁸⁷. As sketched in Figure 1.6, by simply replacing the natural AAs with their correspondent analogs, ribosomal translation of peptides composed of several UAAs has been achieved in PURE system^{90–92}. Importantly, by using this approach, tRNAs are continuously re-acylated, during translation, as it occurs in natural protein expression⁸⁷.

Moreover, auxotrophic strains that is bacteria unable to biosynthesize some AAs have been used to produce proteins incorporating UAAs, when placed in media rich in the corresponding UAAs⁷⁹.

Finally, techniques for synthesizing nucleic acids composed of monomers beyond the limited set of natural nucleosides have been developed by using primer extension⁹³, or PCR⁹⁴. However, similarly to the strategy developed by Szostak, the PCR amplification of functionalized DNA is mostly limited to monomers whose structure is close to the correspondent natural counterparts⁹⁵.

Body

2 NaCRe: beyond recycling

In Chapter 1 several methods currently available for recycling synthetic polymer-based materials have been presented. Progress towards the development of “infinitely recyclable” polymers, as well as on “upcycling” polymers into a variety of higher-value products has been discussed deeply. Thus, recycling a material into itself is the current paradigm in recycling synthetic polymers nowadays.

On the contrary, Nature goes beyond this paradigm, as discussed in section 1.4. Indeed, Nature takes advantage of sequence-defined molecular structures (proteins, and NAs), reversibly cleavable backbones, and polymerization machineries able to work without separating the monomers, for recycling mixtures of n unknown SDPs into the $(n + 1)^{th}$ SDP of need, at the time of polymerization. The sequence, hence the properties, of the so obtained SDP can be completely different from the parent materials.

Much work has been done to synthesize SDPs, as described in section 1.5. However, most of the current research on SDPs obtained through chemical methods focuses mainly on the chemical properties of such materials^{51,96}. On the other side, current research on protein-^{78,97–105}, and DNA-based materials^{106–109} is centered on the nature of the material itself.

In this thesis we asked ourselves if we could establish a strategy for recycling polymers outside living organisms, inspired by the way Nature handles proteins, and NAs. We named this approach to recycling of soft materials nature-inspired circular-economy recycling (NaCRe). We present herein the feasibility of this vision by working with a variety of protein-based materials. A sketch of NaCRe is shown in Figure 2.1.

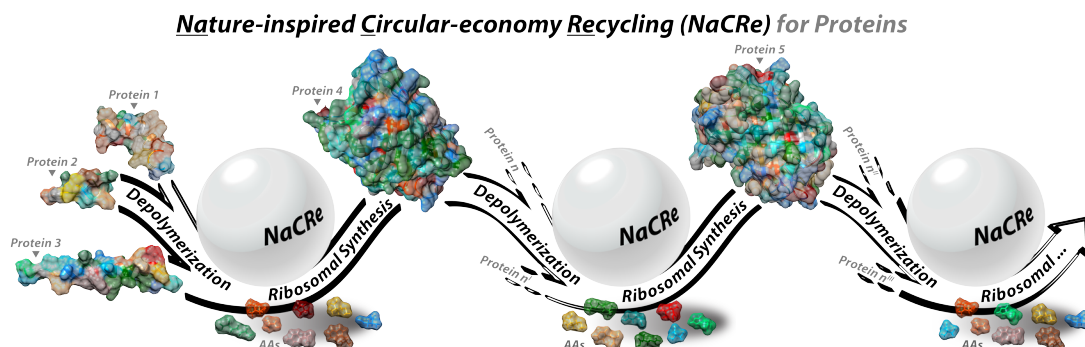


Figure 2.1 – Schematic illustration for the main concept of NaCRe. Multiple possible NaCRe cycles are shown. The illustrated examples are close to what is shown in this thesis. It should be clear that the overall concept of NaCRe goes beyond what is illustrated. The sketched process starts from three different short peptides (drawn as the ones used in this thesis, magainin II, glucagon, and somatostatin), and produces green fluorescent protein (GFP). In the second round of recycling, GFP, together with other arbitrary proteins, is used to produce red fluorescent protein (mScarlet-i). In the last recycling round mScarlet-i is recycled into something not specified, to stimulate the reader’s imagination. Molecular graphics of the proteins 3D structures and of the AAs conformers were from PDB databank (protein 1(2LSA), protein 2(2MI1), protein 3(1GCN), protein 4(5B61), and protein 5(5LK4)) and PubChem (<https://pubchem.ncbi.nlm.nih.gov/compound/CID#section=3D-Conformer>, CID = 5950, 6322, 5960, 5961, 33032, 6274, 6306, 6106, 5962, 6137, 6140, 145742, 5951, 6305, 6057) respectively. All were edited in UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at UCSF, with support from NIH P41-GM103311.

If this approach could be extended in the future for handling most of man-made polymers, there would be no way of distinguishing a new from a recycled material. Indeed mixtures of unknown SDPs would be transformed into the SDP of need at the time by the local community. This vision is based on re-engineering synthetic polymers in a way their building blocks could be kept in circulation. Indeed there would be no need of either downcycling materials (*i.e.* most of plastics recycling nowadays), or continuously generating raw materials from petrochemical, or biological sources^{3,110}.

Moreover the value of the materials would not be lost upon recycling. On the contrary, it would be possible to recycle low-value polymers into higher-value ones, hence creating value upon recycling¹⁶.

3 Cell-free transcription-translation

“Transcription”, and “translation” are precisely controlled polymerization reactions through which Nature performs the synthesis of mRNA, and protein sequences respectively^{51,52}.

A detailed description of DNA replication, and protein translation has been introduced in section 1.5. Similarly to DNA replication, in case of mRNA synthesis, the genetic information contained in the DNA chain is “transcribed” into an mRNA sequence by means of polymerase enzymes that polymerize the nucleoside triphosphate residues, by moving along the unwound DNA template.

In this thesis we made use of cell-free transcription-translation (TX-TL) in a purified system (PURE) in order to polymerize SDPs. This approach has been shortly discussed in section 1.5.2. However, given the fundamental role of such technique in this study, a detailed description of the TX-TL in PURE system, as well as of the main components of PURE is presented below.

3.1 Cell-free transcription-translation in PURE system

The PURE system is a mixture of enzymes, nucleic acids, ribosomes, energy sources, monomers, and buffers that are necessary for protein TX-TL⁷³. The protein components are obtained by recombinant expression in *E. coli*, and ribosomes are purified from *E. coli* cells⁷². The components of PURE are reconstituted to form a purified TX-TL system composed solely of molecules whose functionality is well known. This is a key difference with respect to crude extracts (discussed in section 1.5.2) where a minority of components participate in TX-TL, and

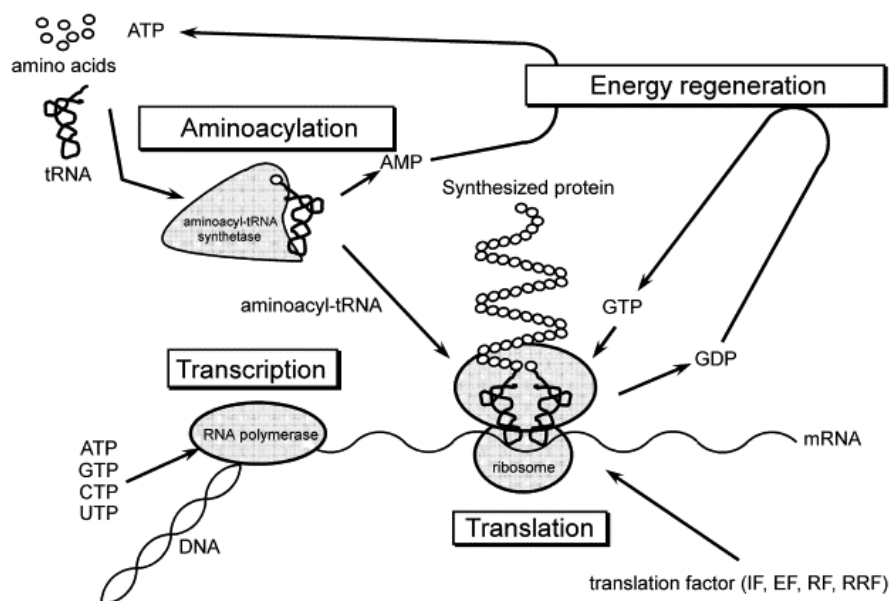


Figure 3.1 – Schematic illustration of the four main reactions (aminoacylation, transcription, translation, and energy regeneration) of cell-free protein synthesis in PURE. The figure has been reproduced from reference 73, with permission from Elsevier.

even inhibiting factors, such as proteases, and nucleases, are present⁷³.

In detail, protein polymerization in PURE system is achieved by means of 4 main reactions that are schematized in Figure 3.1. In aminoacylation, the AA-tRNA pair is formed by the 20 aminoacyl-tRNA synthetases that acylate the AAs onto tRNAs. In transcription, the genetic information contained in the DNA template is copied into an mRNA sequence by the T7 RNA polymerase that polymerizes ATP, GTP, CTP, and UTP nucleoside monomers. Pyrophosphatase catalyzes the conversion of pyrophosphates into phosphate ions. In translation, the protein of interest is synthesized by the ribosomes, using the information carried by the mRNA. The components involved in translation are the AA-tRNA pairs, the initiation factors (IF1, IF2, and IF3), the elongation factors (EF-G, EF-Tu, and EF-Ts), the termination factors (RF1, RF2, and RF3), and the ribosome recycling factor (RRF). The formylation of methionine to form N-formylmethionine (the starting AA) is performed by the methionyl-tRNA formyltransferase (MTF) enzyme, and 10-formyl-5,6,7,8-tetrahydrofolic acid donor. Energy regeneration is needed to power the whole system. The components involved in this process are creatine kinase, creatine phosphate, myokinase, nucleoside-diphosphate kinase, and nucleoside mono-, di-, and triphosphates⁷³.

4 NaCRe: results and discussion

This chapter is an extract of the manuscript entitled *Nature-inspired Circular-economy Recycling (NaCRe) for Proteins: Proof of Concept* that is available as a preprint (<https://www.biorxiv.org/content/10.1101/2020.09.23.309799v3>, doi: 10.1101/2020.09.23.309799 version 3). The authors of the manuscript are Simone Giaveri, Adeline M. Schmitt, Laura Roset Julià, Vincenzo Scamarcio, Anna Murello, Shiyu Cheng, Laure Menin, Daniel Ortiz, Luc Patiny, Sreenath Bolisetty, Raffaele Mezzenga, Sebastian J. Maerkl, and Francesco Stellacci. The author of the thesis is the first author of the manuscript, and is the main contributor for designing, and performing the experiments, characterizing the obtained materials, analyzing the data, and discussing them with the collaborators. All the work has been performed under the supervision of Prof. Francesco Stellacci, and Prof. Sebastian J. Maerkl.

4.1 NaCRe: recycling short peptides

The initial attempt to establish the feasibility of NaCRe was performed by enzymatically depolymerizing three peptides separately, and by recombining the AAs so achieved using the cell machinery to express a target protein.

The latter task was achieved in a standard method. We purchased a commonly-used cell-free transcription-translation (TX-TL) system (PURE, Protein synthesis Using Recombinant Elements, PUREfrex™, Kaneka Eurogentec SA, Appendix A) that is known to “transcribe” the information that we provided by feeding a specific DNA into a messenger RNA (mRNA), and then “translate” the mRNA code by “polymerizing” the target protein.

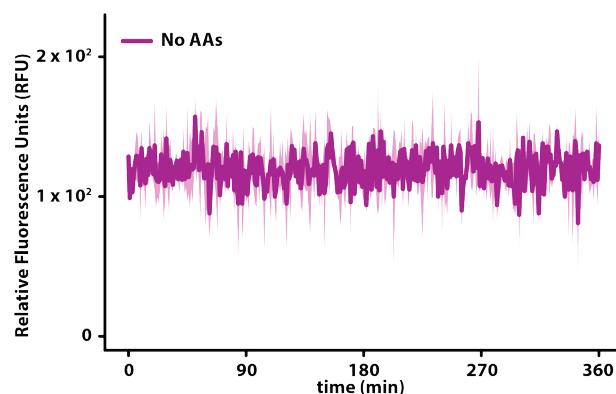


Figure 4.1 – Plots of the fluorescence signal resulting from the expression of mScarlet-i in our TX-TL system without the addition of any AA.

The main issue with commercial TX-TL systems is that they contain free AAs. We chose PUREfrexTM because it is composed of multiple separate solutions, with only one of them that contains free AAs, and it is relatively simple to replace such solution with a home-made one that is AAs-free. The home-made solution lacking the AAs was produced by using a protocol adapted from the original reference from Ueda and coworkers⁷². It should be noted that the PUREfrexTM system contains a single AA (glutamic acid) as a component of one of the other solutions. Hereafter, we will refer to this home-made AAs-free form of PUREfrexTM simply as TX-TL system.

To establish the absence of AAs in our TX-TL system, we performed control experiments that show the lack of any detectable protein expression (Figure 4.1). In order to have a simple way to detect protein expression in the TX-TL system, we decided to focus all the work presented here on expressing fluorescent proteins. As a first choice, we focused on mScarlet-i, a fluorescent protein whose sequence contains 19 of the 20 proteinogenic AAs with cysteine missing. For later work, we expressed green fluorescent protein (GFP) as it is the most commonly expressed fluorescent protein and it contains all 20 proteinogenic AAs.

We felt that it would be simpler to develop a robust depolymerization method starting with shorter molecules, thus our initial attempts were based on short peptides. We selected magainin II, and glucagon by reading the whole PDB databank searching for peptides composed of a short number n of residues ($20 \leq n \leq 30$), with no cysteine, and no unnatural/modified residues (see Appendix B.3).

From the hits, we selected commercially available peptides, presenting different secondary

structures, and different functions. Magainin II (Table C.1) is an antimicrobial peptide, and glucagon (Table C.1) is a peptide hormone. Somatostatin 28 (Table C.1), a peptide hormone, was selected *a posteriori* because it is rich in proline (missing in magainin II and glucagon), and structurally different from the other two peptides, *i.e.* disulfide cyclized. The three peptides together contain all 20 proteinogenic AAs (see Figure 4.2a-c for AAs contained in each peptide).

We depolymerized magainin II, glucagon, and somatostatin 28 by means of two consecutive enzymatic reactions, following the approach developed by Teixeira *et al*¹¹¹. We incubated the peptides first with thermolysin endoprotease (that cleaves at the N-terminus of Leu, Phe, Val, Ile, Ala, Met), then with leucine aminopeptidase (LAP), as described in Appendix B.4 and B.5. Mass spectrometric (MS) analysis of the materials before (Figures D.6-D.8), and after thermolysin treatment (Figures D.10-D.23) shows extensive cleavage at the N-terminus of the hydrophobic amino acids (see Appendix B.9). Cleaved fragments were incubated with LAP and depolymerized to their free AAs (Figure 4.2a-c).

For each AA we defined a depolymerization yield as the ratio between the amount of AAs produced by the depolymerization divided by the total amount of AAs present in the starting material (green and gray bars in Figure 4.2, respectively). Quantification was performed using MS (Appendix B.10). We achieved an average depolymerization yield of $\sim 66\% \pm 19\%$. The large standard deviation ($1\sigma=19\%$) is caused by the large variation between depolymerization yields of different AAs, with a maximum of $\sim 99\%$ for aspartic acid (for glucagon) and a minimum of $\sim 17\%$ for phenylalanine (averaged for all three peptides). We observed variations in yield also across peptides, for example alanine was efficiently recovered from the depolymerization of magainin II and glucagon, but not from somatostatin 28. We noticed that the aromatic AAs were consistently recovered in poor yields (for all three peptides), and that such yields were dependent on the number (type) of aromatic residues in the material to be depolymerized. Specifically, the recovery of the aromatics in glucagon was higher ($\sim 73\%$ for Trp, $\sim 52\%$ for Tyr, and $\sim 36\%$ for Phe) than in somatostatin 28 ($\sim 41\%$ for Trp, and $\sim 15\%$ for Phe), that was in turn higher than in magainin II (\sim null for Phe).

The free AAs achieved by depolymerizing separately the three peptides were combined, and added into the TX-TL system supplemented with an mScarlet-i DNA template (Table C.2,

Chapter 4. NaCRe: results and discussion

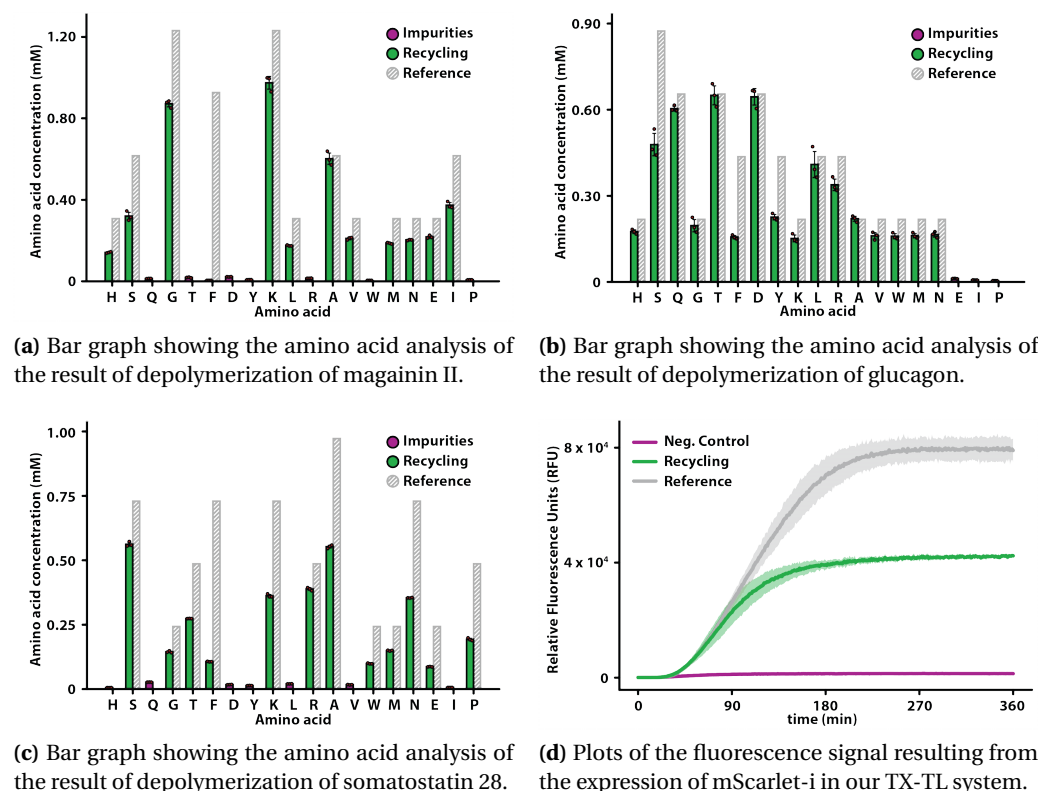


Figure 4.2 – Recycling of magainin II, glucagon, and somatostatin 28 into mScarlet-i. Bar graphs showing the result of the amino acid analysis performed by using mass spectrometry on the result of depolymerization of magainin II (a), glucagon (b), and somatostatin 28 (c). The experimental results are represented with green bars to be compared with the gray bars that are the ideal reference concentrations of each AA calculated by assuming the complete conversion of the starting peptide into free AAs. The violet bars represent trace concentration of the AAs that theoretically should have not been observed, they are possibly the result of depolymerization of the digestion enzymes themselves. Such impurities are present for all the recovered AAs. The additive effect due to the impurities is by definition difficult to estimate, and probably contributes to slightly overestimate the green bars. This becomes more evident when the obtained depolymerization yield is close to 100%. (Note: cysteine is not detected by the amino acid analysis, hence the quantification of cysteine is n.a.). Plots of the fluorescence signal resulting from the expression of mScarlet-i in a TX-TL reaction (d). The green curves are data obtained performing NaCRe on magainin II, glucagon, and somatostatin 28 (Appendix B.7), the gray curves are obtained as the results of expression experiments with the TX-TL reactions supplemented with concentrations for each AA matching the gray bars shown in (a), (b), and (c). In the negative control expressions (violet curves), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the individual peptides, without adding the peptides initially. Bar-plots of the statistical mean of the results of the repeated injections (triplicates) of each sample are shown; error bars represent the standard deviation of the same data. The TX-TL reactions were all run in duplicates. The expression curves represent the statistical mean of the results at any acquisition time; the shadow represents the standard deviation of the same data.

Appendix B.7). As shown in Figure 4.2d we successfully expressed mScarlet-i. As a reference control and yield reference, we ran a TX-TL reaction with a solution containing the concentration of each AA that would have been achieved had the depolymerization yield been 100% for each peptide (that ideal result of a complete depolymerization, Appendix B.7).

A first attempt to determine the efficiency of NaCRe was performed by comparing the fluorescence values of the expression plateau for the recycling curve with that for the reference control (the green and gray curves in Figure 4.2d respectively), leading to a yield of ~50%. We also used NaCRe to express GFP (see Table C.5). In this case we spiked cysteine into the free AAs solution obtained from the depolymerization of magainin II, glucagon, and somatostatin 28, the resulting yield for GFP was ~80% (Figure 4.3).

The results presented so far were achieved performing the depolymerization of each peptide separately, and by combining the obtained solutions at the end of the depolymerization process. In order to establish NaCRe as a recycling method that starts from mixtures of proteins and/or peptides, we also performed it starting with a mixture of the three peptides, depolymerizing them together, and expressing GFP. As shown in Figure 4.4, the process was

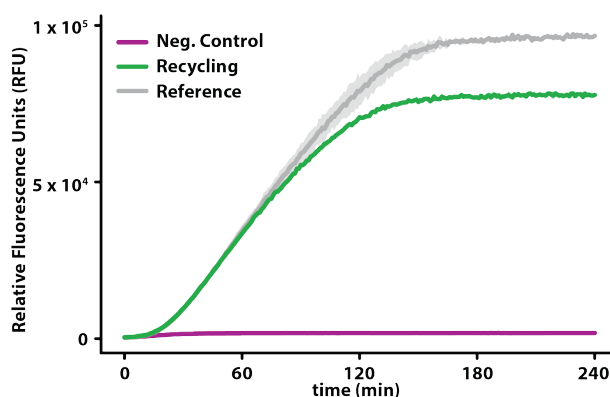


Figure 4.3 – Recycling of magainin II, glucagon, and somatostatin 28 into GFP. Plots of the fluorescence signal resulting from the expression of GFP in our TX-TL system. The green curve is obtained performing NaCRe on magainin II, glucagon, and somatostatin 28 (Appendix B.7). The gray curve (reference control) is obtained as the result of an expression experiment with the TX-TL system supplemented with concentrations of AAs matching the complete depolymerization of the initial materials (gray bars in 4.2a-c). In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the individual peptides, without adding the peptides initially. Details are explained in the caption of Figure 4.2.

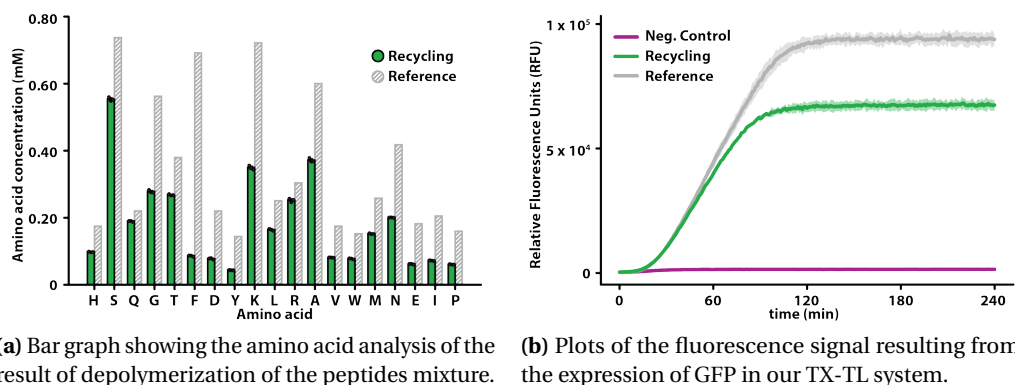


Figure 4.4 – Recycling of the mixture of magainin II, glucagon, and somatostatin 28 into GFP. Bar graphs showing the result of the amino acid analysis performed using mass spectrometry on the result of depolymerization of the mixture of magainin II, glucagon, and somatostatin 28 (a). Plot of the fluorescence signal resulting from the expression of GFP in a TX-TL reaction (b). The green curve is obtained performing NaCRe on the mixture of magainin II, glucagon, and somatostatin 28 (Appendix B.7). The gray curve (reference control) is obtained as the result of an expression experiment with the TX-TL system supplemented with concentrations of AAs matching the complete depolymerization of the initial materials (gray bars in (a)). In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the peptides mixture, without adding the peptides initially. Details are explained in the caption of Figure 4.2.

successful in depolymerization and expression, leading to a yield of ~70% that is approximately the same yield we obtained when expressing GFP starting from the product of the separate depolymerization of the peptides.

It would be obvious at this point to wonder about the difference in observed yields for the expression of mScarlet-i and GFP. First, the yields mentioned so far are relative yields (RY), defined as:

$$\frac{P_1}{P_2} \times 100 \quad (4.1)$$

where P_1 and P_2 are the fluorescence intensity signal for the NaCRe (P_1) and the reference (P_2) expressed proteins, averaged over the last 30 min of the experiment.

The evaluation of a yield for NaCRe is rather complex because of the sequence-defined nature of the product. In fact, when expressing a protein from a mixture of free AAs there will always be a limiting reactant. This limiting AA will be the one that determines the amount of protein expressed in the reference control.

By virtue of this definition, the limiting AA depends both on the proteins/peptides that were

4.1. NaCRe: recycling short peptides

depolymerized as well as on the specific sequence of the protein to be expressed. As shown in Table 4.1, when recycling the three peptides, the limiting AAs for expressing the reference mScarlet-i is either proline, tyrosine, or valine, while for GFP it is valine.

Note that the limiting AA does not necessarily need to be the AA with the lowest concentration in the reference reactant mixture, indeed in our case this was tyrosine. Also, the concentration of cysteine is irrelevant when expressing mScarlet-i because it lacks cysteine.

Therefore, the yield of NaCRe can be tailored by enriching the mixture of proteins to be depolymerized with proteins/protein-based materials that contain the residues that are highly used in the sequence of the protein to be expressed.

When determining the RY we make the implicit assumption that the limiting AA in NaCRe and in the reference control is the same. As shown in Table 4.1, this is not necessarily always the case.

Therefore, even though the RY is a simple measure of the efficiency of our process, it depends critically on the starting and final proteins, hence it is a powerful tool solely to compare and optimize the yield of NaCRe when starting and ending from and into the same proteins. The true efficiency of NaCRe should be its absolute yield (AY) defined as a mass-to-mass ratio of

Depolymerization (AAs) (nmol) of AAs in 10 μ l				Expression (protein chains)* (nmol) of protein chains in 25 μ l TX-TL system supplemented with 10 μ l of AAs from depolymerization				
AAs	Reference AAs	AAs		Reference (mScarlet-i)	NaCRe (mScarlet-i)	Reference (GFP)	NaCRe (GFP)	NaCRe (GFP)
(Type)	(Ideal)	from separated peptides	from peptides mix	(Ideal)	from separated peptides	(Ideal)	from separated peptides	from peptides mix
		(Experimental)	(Experimental)		(Experimental)		(Experimental)	(Experimental)
H	1.76	1.11 \pm 0.04	1.00 \pm 0.01	0.16	0.10 \pm 0.01	0.12	0.07 \pm 0.01	0.07 \pm 0.01
S	7.41	4.57 \pm 0.27	5.43 \pm 0.06	0.44	0.27 \pm 0.02	0.74	0.46 \pm 0.03	0.54 \pm 0.01
Q	2.19	2.17 \pm 0.04	1.90 \pm 0.01	0.27	0.27 \pm 0.01	0.27	0.27 \pm 0.01	0.24 \pm 0.01
G	5.67	4.05 \pm 0.18	2.80 \pm 0.01	0.19	0.14 \pm 0.01	0.26	0.18 \pm 0.01	0.13 \pm 0.01
T	3.81	3.14 \pm 0.14	2.70 \pm 0.01	0.22	0.19 \pm 0.01	0.24	0.20 \pm 0.01	0.17 \pm 0.01
F	6.98	0.88 \pm 0.02	0.87 \pm 0.06	0.63	0.08 \pm 0.01	0.58	0.07 \pm 0.01	0.07 \pm 0.01
D	2.19	2.29 \pm 0.13	0.83 \pm 0.06	0.14	0.14 \pm 0.01	0.12	0.13 \pm 0.01	0.05 \pm 0.01
Y	1.46	0.83 \pm 0.04	0.40 \pm 0.01	0.12	0.07 \pm 0.01	0.13	0.08 \pm 0.01	0.04 \pm 0.01
K	7.29	5.00 \pm 0.21	3.50 \pm 0.11	0.33	0.23 \pm 0.01	0.36	0.25 \pm 0.01	0.18 \pm 0.01
L	2.49	2.02 \pm 0.20	1.63 \pm 0.06	0.18	0.14 \pm 0.02	0.12	0.10 \pm 0.01	0.08 \pm 0.01
R	3.08	2.50 \pm 0.11	2.53 \pm 0.06	0.22	0.18 \pm 0.01	0.51	0.42 \pm 0.02	0.42 \pm 0.01
A	6.03	4.61 \pm 0.17	3.73 \pm 0.06	0.43	0.33 \pm 0.02	0.75	0.58 \pm 0.03	0.47 \pm 0.01
V	1.76	1.31 \pm 0.07	0.80 \pm 0.01	0.12	0.09 \pm 0.01	0.10	0.07 \pm 0.01	0.04 \pm 0.01
W	1.54	0.87 \pm 0.04	0.80 \pm 0.01	0.51	0.29 \pm 0.02	1.54	0.87 \pm 0.04	0.80 \pm 0.01
M	2.57	1.66 \pm 0.06	1.50 \pm 0.01	0.23	0.15 \pm 0.01	0.43	0.28 \pm 0.01	0.25 \pm 0.01
N	4.19	2.41 \pm 0.08	2.00 \pm 0.01	0.70	0.40 \pm 0.02	0.32	0.19 \pm 0.01	0.15 \pm 0.01
E	1.84	1.07 \pm 0.04	0.60 \pm 0.01	n.a.**	n.a.**	n.a.**	n.a.**	n.a.**
I	2.06	1.32 \pm 0.06	0.70 \pm 0.01	0.26	0.17 \pm 0.01	0.17	0.11 \pm 0.01	0.06 \pm 0.01
P	1.62	0.68 \pm 0.02	0.60 \pm 0.01	0.12	0.05 \pm 0.01	0.16	0.07 \pm 0.01	0.06 \pm 0.01

Table 4.1 – Overview of the depolymerization and expression efficiencies for key experiments in this study. Minima are colored in blue (depolymerization) and red (expression). * Calculated as the ratio between the amount of each AA (nmol) and the number of its incorporations inside a single protein chain, see Table C.5. ** n.a. not assessable because E is present in the TX-TL system as Potassium glutamate (buffer), see Appendix B.7.

the output divided by the input. When the limiting AA is the same for the NaCRe and reference control, the AY can be written as:

$$RY \times Y \quad (4.2)$$

where Y is the yield of expression of the TX-TL system. AY (mass-to-mass ratio) in the case of the expression of mScarlet-i is ~7% (see Appendix B.13). The present results show a mass-to-mass yield for NaCRe for the limiting AA of proline in the expression of mScarlet-i of ~15%. This is the most accurate measurement of the absolute yield of the process.

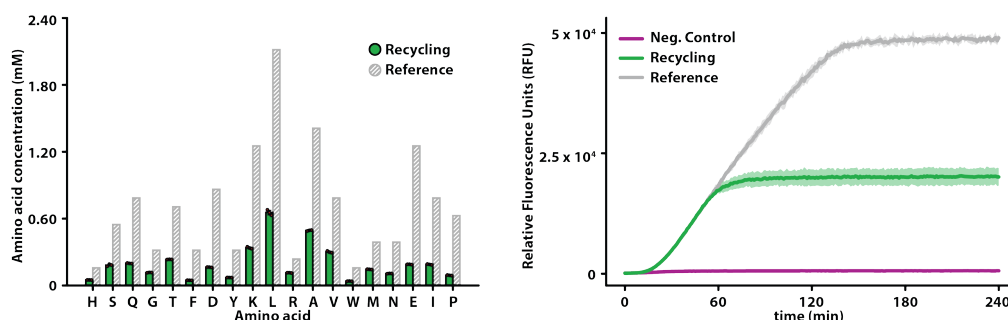
4.2 NaCRe: recycling technologically relevant proteins

To go beyond peptides, we performed NaCRe starting from larger proteins with defined tertiary structures. We started by recycling β -lactoglobulin A (~18 KDa, Table C.1), a protein that can be obtained in large quantities as a side product of bovine milk production. As shown in Figure 4.5a, β -lactoglobulin A was successfully depolymerized into its constitutive AAs with a yield comparable to the ones obtained for the peptides (see Appendix B.4 and B.5). These AAs were used to express GFP (Figure 4.5b, Appendix B.7). The RY for β -lactoglobulin A recycled into GFP was ~40%.

To better establish the potential of NaCRe we recycled technologically relevant materials. We first recycled a film composed of β -lactoglobulin amyloid fibrils, known to be able to adsorb a variety of different heavy metal ions with outstanding efficiency¹⁰³. Such amyloids are assemblies of peptides obtained from the hydrolysis of β -lactoglobulin chains (A and B) and their re-assembly into filamentous proteins with a typical cross- β secondary structure. Because amyloids have been postulated to be the ground state in the protein folding landscape¹¹², carrying out NaCRe starting from these systems ideally showcase the universality and the reach of the method.

A solution of amyloid fibrils was dried on a cellulose membrane, as shown in Figure 4.6 (see Appendix B.2). The dry film was removed from the support, the film powder was weighed, and first incubated with pepsin endoprotease (that cleaves at the C-terminus of Leu, Phe, Tyr, Trp), then with LAP (Appendix B.4 and B.5, respectively).

4.2. NaCre: recycling technologically relevant proteins



(a) Bar graph showing the amino acid analysis of the result of depolymerization of β -lactoglobulin A. (b) Plots of the fluorescence signal resulting from the expression of GFP in our TX-TL system.

Figure 4.5 – Recycling of β -lactoglobulin A into GFP. Bar graphs showing the result of the amino acid analysis performed using mass spectrometry on the result of depolymerization of β -lactoglobulin A (a). Plots of the fluorescence signal resulting from the expression of GFP (b) in a TX-TL reaction. The green curve is obtained performing NaCre on β -lactoglobulin A (Appendix B.7). The gray curve (reference control) is obtained as the result of an expression experiment with the TX-TL system supplemented with concentrations of AAs matching the complete depolymerization of the initial material (gray bars in (a)). In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for β -lactoglobulin A, without adding β -lactoglobulin A initially. Details are explained in the caption of Figure 4.2.

In order to support the mass spectrometry evaluation of the depolymerization process, we performed atomic force microscopy (AFM) analysis of the amyloid fibrils as prepared, and after full depolymerization. The images of the as prepared amyloids show an abundance of fibrils, that were absent after depolymerization (Figure 4.7).

The mass spectrometry result of the consecutive cleavage, and depolymerization is shown in Figure 4.8a. In this case we do not have a reference standard, as the exact amyloid composition is unknown due to the hydrolysis process. We note that methionine, and histidine were

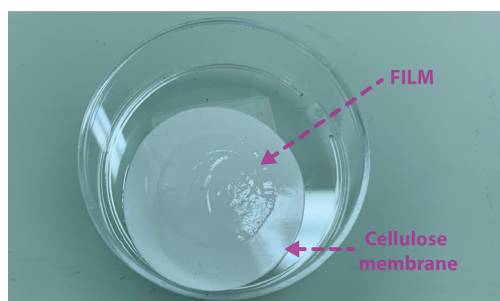


Figure 4.6 – Photograph of a film composed of β -lactoglobulin amyloid fibrils, deposited on a cellulose support.

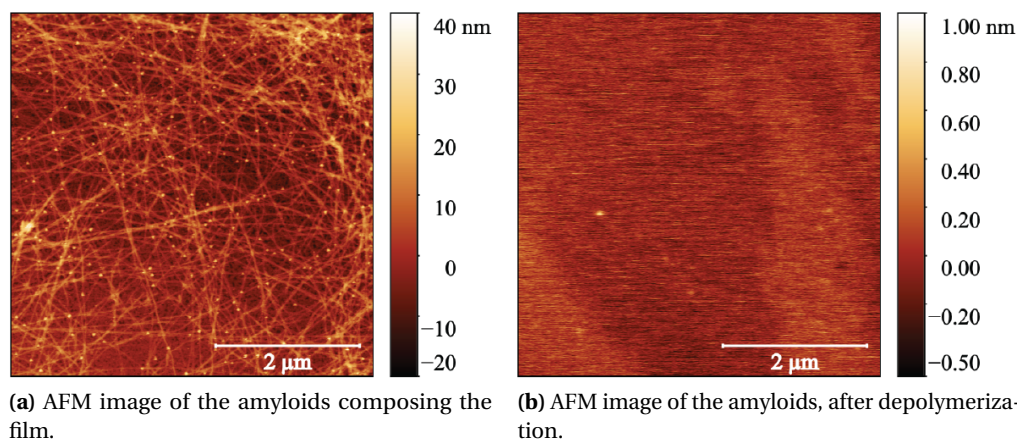


Figure 4.7 – AFM characterization of the β -lactoglobulin amyloids obtained from solubilizing the film powder, and deposited on cleaved mica surfaces, as prepared (a) and after depolymerization (b).

obtained only at low concentrations. As shown in Figure 4.8b, the free AAs obtained from the β -lactoglobulin film were recycled into GFP, by spiking cysteine, methionine, and histidine (see Appendix B.7).

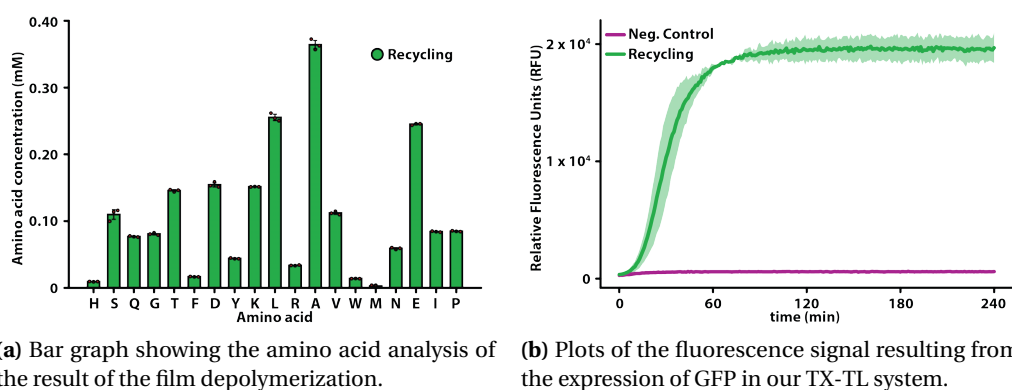


Figure 4.8 – Recycling of a film composed of β -lactoglobulin amyloid fibrils into GFP. Bar graphs showing the result of the amino acid analysis performed using mass spectrometry on the result of depolymerization of the β -lactoglobulin film (a). Plots of the fluorescence signal resulting from the expression of GFP (b) in a TX-TL reaction. The green curve is obtained performing NaCRE on the β -lactoglobulin film (Appendix B.7). In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the film, without adding the film initially. The reference controls (gray bars in (a) and gray curve in (b)) are missing because the exact composition of the amyloids composing the film is unknown. Details are explained in the caption of Figure 4.2.

4.2. NaCRe: recycling technologically relevant proteins

We then recycled a solution of silk fibroin (Table C.1), another technologically relevant protein used in many devices, ranging from biomedical⁹⁸ to electronic applications¹⁰⁰.

After incubating fibroin with thermolysin, and then LAP (see Appendix B.4 and B.5, respectively), we successfully recovered fibroin's free AAs (Figure 4.9a), and used them to express GFP in our TX-TL system (Figure 4.9b) spiked with cysteine, and methionine (see Appendix B.7). RY for silk fibroin recycling into GFP was ~95%.

Figures 4.8b and 4.9b demonstrate that NaCRe is capable of recycling high molecular weight polymeric structures, either composed of the supramolecular assembly of low molecular weight peptides or characterized by multiple high molecular weight chains.

As described above, we decided to spike cysteine every time we were expressing GFP because we could not detect cysteine, *i.e.* quantify it, in the AAs solutions from the depolymerizations. We then tried to assess if cysteine could be part of NaCRe by recycling magainin II, glucagon, and somatostatin 28 into GFP, without adding any cysteine (Appendix B.7). As shown in Figure 4.10a, spiking cysteine was not necessary, since the two recycling curves reach basically the same plateau, this means that cysteine from the disulfide cyclization of somatostatin 28 is

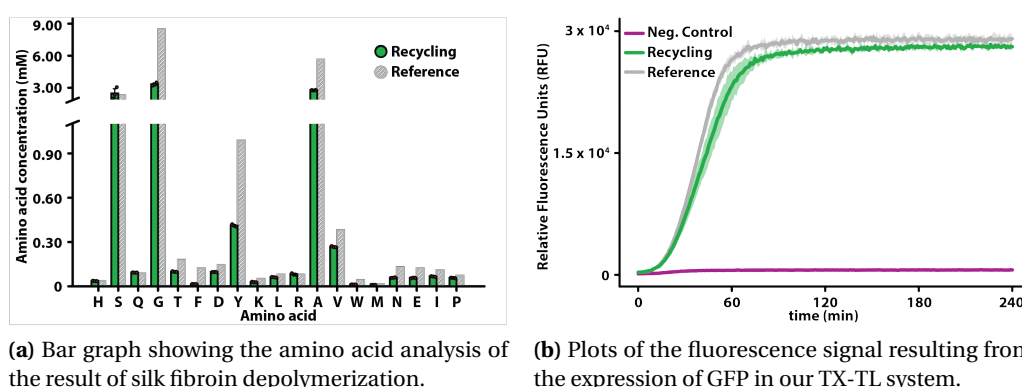


Figure 4.9 – Recycling of a silk fibroin solution into GFP. Bar graphs showing the result of the amino acid analysis performed using mass spectrometry on the result of depolymerization of a silk fibroin solution (a). Plots of the fluorescence signal resulting from the expression of GFP (b) in a TX-TL reaction. The green curve is obtained performing NaCRe on the silk fibroin solution (Appendix B.7). The gray curve (reference control) is obtained as the result of an expression experiment with the TX-TL system supplemented with concentrations of AAs matching the complete depolymerization of the initial material (gray bars in (a)). In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the silk fibroin, without adding the silk fibroin initially. Details are explained in the caption of Figure 4.2.

Chapter 4. NaCRe: results and discussion

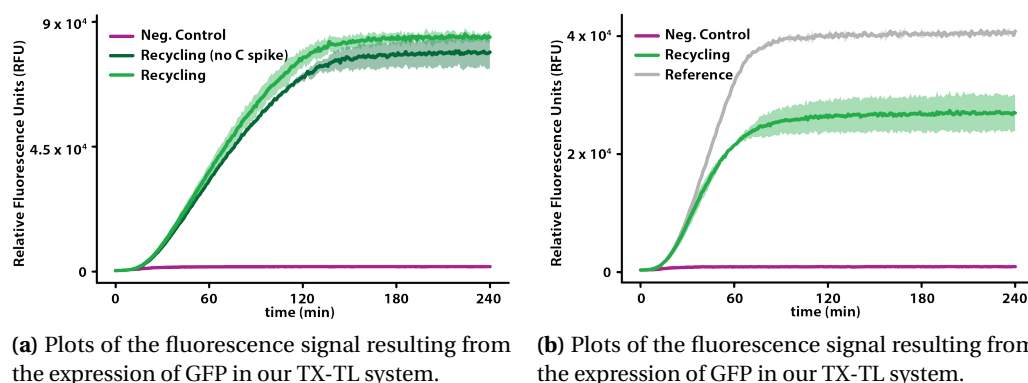


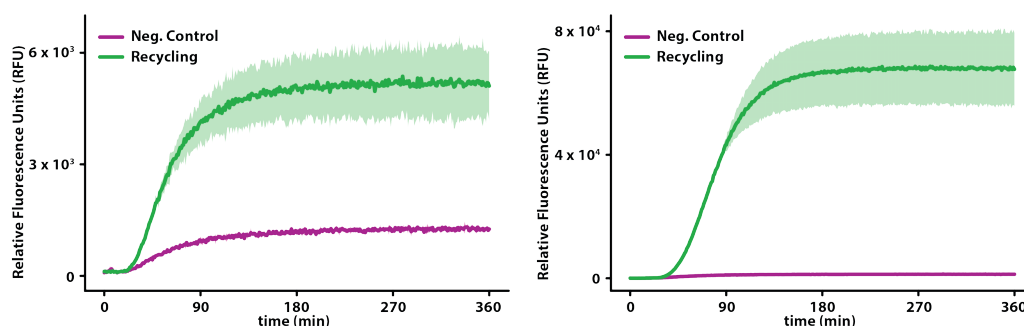
Figure 4.10 – Recycling of cysteine. In (a) the green curves are obtained preforming NaCRe on magainin II, glucagon, and somatostatin 28 with and without spiking cysteine. In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the individual peptides, without adding the peptides initially. In (b) the green curve is obtained preforming NaCRe on the mixture composed of glucagon, β -lactoglobulin A, and silk fibroin (Appendix B.7). The gray curve (reference control) is the result of an expression experiment with the TX-TL system supplemented with concentrations of AAs matching the complete depolymerization of the initial materials. In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the glucagon, β -lactoglobulin A, and silk fibroin mixture, without adding the three proteins initially. Details are explained in the caption of Figure 4.2.

recycled into GFP. This result strengthens the visionary idea of NaCRe, where materials are recycled into completely different ones, without the need of any external monomer feed, that is fulfilling the principles of a circular-economy model for polymers.

After proving that cysteine can be recycled by NaCRe (as well as the other AAs), we performed every experiment without the need of spiking any amino acid. We produced a mixture of low and high molecular weight proteins (glucagon, β -lactoglobulin A, and silk fibroin), and we successfully recycled it into GFP, as shown in Figure 4.10b. RY for recycling this mixture of proteins into GFP was $\sim 70\%$.

In order to show that NaCRe can undergo more than one complete cycle, we first scaled-up the NaCRe processes described just above to produce either GFP or mScarlet-i. We purified these proteins (see Figure D.48), and characterized them by proteomic analysis (see Appendix B.12). For GFP we identified 24 exclusive unique peptides (55 exclusive unique spectra), with 87% sequence coverage. For mScarlet-i we identified 21 exclusive unique peptides (48 exclusive

4.2. NaCRE: recycling technologically relevant proteins



(a) Plots of the fluorescence signal resulting from the expression of mScarlet-i in our TX-TL system. (b) Plots of the fluorescence signal resulting from the expression of mScarlet-i in our TX-TL system.

Figure 4.11 – Second NaCRE cycle. In (a) the green curve is obtained preforming a second cycle of NaCRE on the GFP produced by recycling the mixture composed of glucagon, β -lactoglobulin A, and silk fibroin (Appendix B.8). In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for GFP, without adding the protein initially. In (b) the green curve is obtained preforming NaCRE on the whole solution resulting from a first cycle of NaCRE in which glucagon, β -lactoglobulin A, and silk fibroin were recycled into GFP (Appendix B.8). In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the whole first cycle of NaCRE, without adding the whole first cycle of NaCRE initially. Details are explained in the caption of Figure 4.2.

unique spectra), with 77% sequence coverage. We then performed a second NaCRE cycle on the purified GFP (~0.1 mg) to produce mScarlet-i (Figure 4.11a), without the need of any spike AAs (see Appendix B.8).

After performing NaCRE starting from the mixture of low and high molecular weight proteins, we applied the same strategy to recycle a very complex mixture of proteins, that is our whole TX-TL system. As shown in Figure 4.11b, we successfully recycled into mScarlet-i the whole solution resulting from a first cycle of NaCRE in which glucagon, β -lactoglobulin A, and silk fibroin were recycled into GFP (see Appendix B.8). This experiment demonstrates the robustness of NaCRE that can perform multiple cycles of recycling for truly complex protein mixtures, in the presence of other polymers such as nucleic acids.

Starting from the same mixture of glucagon, β -lactoglobulin A, and silk fibroin, we have also performed NaCRE to obtain catechol 2,3-dioxygenase (CDO, see Table C.5), an enzyme which converts catechol into 2-hydroxymuconate semialdehyde¹¹³. Figure 4.12 shows that the product of NaCRE is indeed catalytically active.

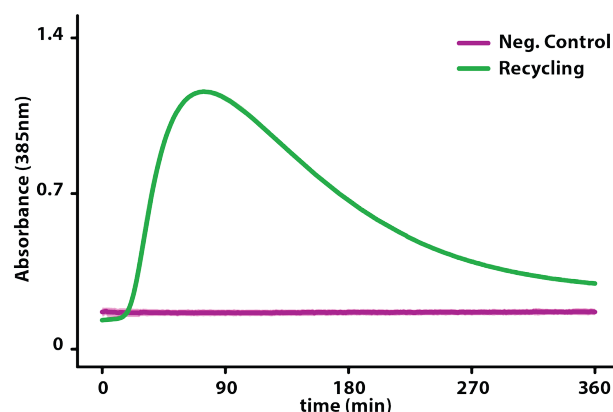


Figure 4.12 – Recycling of the mixture of glucagon, β -lactoglobulin A, and silk fibroin into the enzyme catechol 2,3-dioxygenase (CDO). Plots of the absorbance signal at 385 nm resulting from the conversion of catechol into 2-hydroxymuconate semialdehyde, catalyzed by the CDO enzyme expressed in our TX-TL system. The green curve is obtained performing NaCRe on the mixture composed of glucagon, β -lactoglobulin A, and silk fibroin (Appendix B.7). In the negative control (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the glucagon, β -lactoglobulin A, and silk fibroin mixture, without adding the three proteins initially. The TX-TL reactions were all run in duplicates. The absorbance curves represent the statistical mean of the results at any acquisition time; the shadow represents the standard deviation of the same data.

4.3 NaCRe: beyond natural proteins

After having shown that NaCRe is capable of recycling a variety of structurally different proteins, and protein-based materials, we demonstrated that NaCRe is not limited to the functionalities present in the 20 proteinogenic AAs. Thus, we recycled 2 unnatural amino acids (UAAs, L-norleucine, and L-canavanine) originating from a peptide containing several UAAs (see Table C.1), some present as DL-stereoisomers (3-Fluoro-DL-valine and DL-3-hydroxynorvaline).

The non-natural peptide was incubated first with thermolysin, then with LAP, as described in Appendix B.4 and B.5. MS analysis before (Figure D.9), and after thermolysin incubation (Figures D.24-D.27, and Appendix B.9) shows extensive cleavage. After depolymerization with LAP, we identified all the residues composing the non-natural peptide (Figures D.28-D.33, and Appendix B.9).

L-norleucine, and L-canavanine were successfully recycled into GFP (Figures D.34-D.45, and Appendix B.7 and B.12), following the protocol developed in references 89, and 90. The final

4.3. NaCRe: beyond natural proteins

product of this approach is a sequence-defined polymer composed of a set of monomers that goes beyond the 20 proteinogenic AAs. It should be noted that the GFP produced in this way is not fluorescent (Figure D.5). If one wanted to obtain from NaCRe proteins with their full set of biological properties then NaCRe should be based solely on the 20 proteinogenic AAs.

5 Non-natural proteins – synthetic SDPs

As discussed in section 1.5, proteins incorporating non-natural building blocks are synthetic SDPs composed of proteinogenic backbones, and modified lateral chains. Using the methodology developed by Szostak^{90–92} (section 1.5.2), we took advantage of the natural “translation” machinery to polymerize proteins containing several UAA residues by simply replacing the proteinogenic AAs with their corresponding “analogs”, in a PURE TX-TL system.

5.1 Non-natural proteins as synthetic SDPs: results

As a first attempt to express proteins containing non-natural amino acids, we decided to start with single UAA incorporations into GFP. Samples were obtained by performing TX-TL of GFP in PURE system supplemented with 19 proteinogenic AAs, and 1 non-natural residue. In the TX-TL of the negative controls, the non-natural monomer was omitted. We used SDS-PAGE protein electrophoresis for detecting the polymerization of the protein of interest. The protein bands corresponding to the samples were compared with the negative controls. We labelled the expressed proteins with BODIPY-FL, by supplementing the TX-TL system with commercially available tRNA-Lys-BODIPY-FL. This strategy allowed us to detect small amounts of the protein of interest in a fast, non-radioactive, and very sensitive way. Single incorporations into GFP of UAAs with a variety of non-natural lateral chains such as L-canavanine (arginine analog), L-norleucine (methionine analog), L-3-hydroxy-norvaline (threonine analog), 3-fluoro-L-valine (valine analog), S-(2-aminoethyl)-L-cysteine hydrochloride

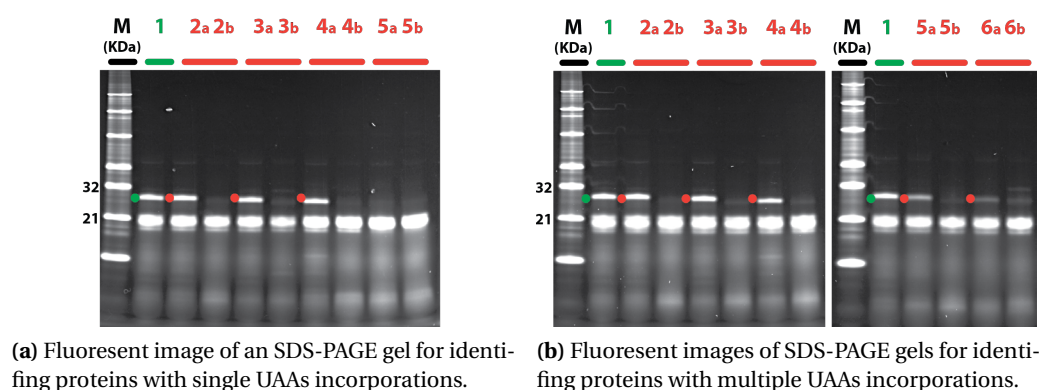


Figure 5.1 – Fluorescent images of SDS-PAGE protein gels used to identify proteins (GFP) containing single (a), or multiple (b) UAAs incorporations. In each gel, a green marker in lane 1 is used to indicate the protein bands corresponding to GFP obtained by using the 20 proteinogenic AAs (positive controls). Red markers in lanes Xa are used to indicate the protein bands corresponding to GFP with UAAs incorporations (samples). In (a), incorporations of L-canavanine, L-norleucine, L-3 hydroxy-norvaline, or β -t-butyl-L-alanine are shown in lanes 2a, 3a, 4a, and 5a respectively. In (b), incorporations of L-canavanine, L-canavanine + L-norleucine, L-canavanine + L-norleucine + L-3-hydroxy-norvaline, L-canavanine + L-norleucine + L-3-hydroxy-norvaline + 3-fluoro-L-valine, and L-canavanine + L-norleucine + L-3-hydroxy-norvaline + 3-fluoro-L-valine + S-(2-aminoethyl)-L-cysteine hydrochloride are shown in lanes 2a, 3a, 4a, 5a, and 6a respectively. Protein bands in lanes Xb should not be observed (negative controls). They are possibly due to the expression of GFP with trace concentration of natural AAs. The additional fluorescent bands visible at the bottom of each gel are due to tRNA-Lys-BODIPY-FL that is present in excess in each TX-TL reaction.

ride (lysine analog), 3-fluoro-L-tyrosine (tyrosine analog), 4-azido-L-homoalanine hydrochloride (methionine analog), quisqualic acid (glutamic acid analog), and L- α -(2-theinyl)glycine (isoleucine analog) were achieved successfully. Red markers in Figure 5.1a show the achieved single incorporations of L-canavanine, L-norleucine, or L-3-hydroxy-norvaline; on the contrary the incorporation of β -t-butyl-L-alanine (leucine analog) was not successful.

We then attempted to incorporate multiple UAAs into the same protein sequence. Figure 5.1b shows the successful incorporation of up to 5 UAAs into the GFP sequence. In this last case in fact the bands present in lane 6b are due to the competition between S-(2-aminoethyl)-L-cysteine hydrochloride (lysine analog), and Lys-BODIPY-FL. These experiments demonstrate that it is possible to polymerize ribosomally a protein sequence that is composed of ~30% of non-natural building blocks.

So far we achieved to polymerize GFP with several non-natural substitutions, that is we

5.1. Non-natural proteins as synthetic SDPs: results

obtained a sequence-defined polymer that is composed of natural, and non-natural residues. We were interested to see if we could polymerize a sequence composed of non-natural residues solely. Taking advantage of the multiple UAAs incorporations into GFP, we designed the DNA encoding an oligomer composed of 11 non-natural residues in a row, at the N-terminus of GFP. The DNA sequence was engineered such that the three ribosomal sites were all occupied by non-proteinogenic amino acids allowed by the ribosomal machinery. After TX-TL by supplementing the PURE system with 16 AAs, and 4 UAAs (L-norleucine, 3-fluoro-L-valine, L-canavanine, and L-3-hydroxy-norvaline), the non-natural peptide was obtained by cleaving it from GFP, and identified by MS as shown in Figure 5.2.

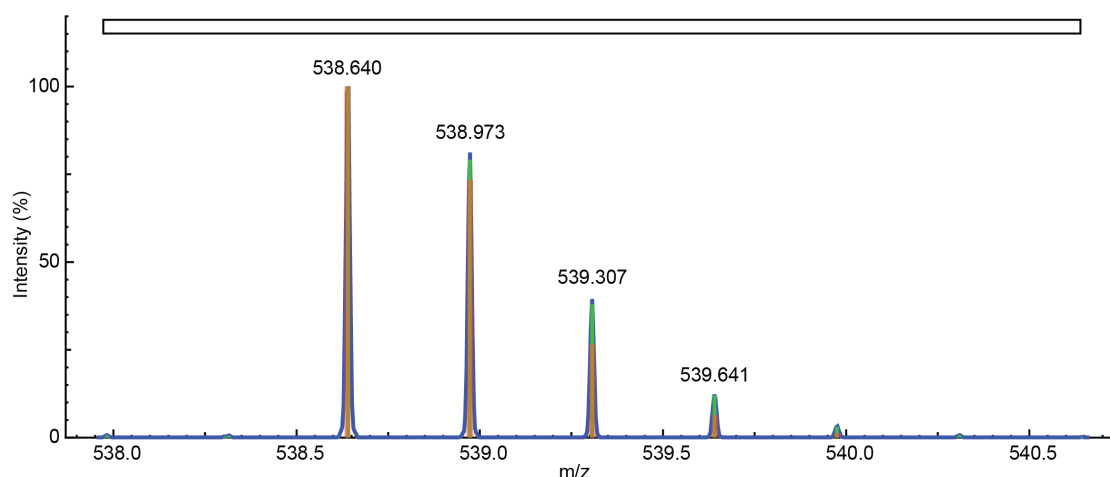


Figure 5.2 – MS spectrum of the non-natural peptide H3(3+)For[L-norleucine] [3-fluoro-L-valine] [3-fluoro-L-valine] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-norleucine] [Serine] [Lysine]OH expressed at the N-terminus of GFP, and obtained by in gel digestion with Lys-C. Theoretical observed mass (m/z) = 538.639; experimental closest peak (m/z) = 538.640. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

Identified fragments						
b						b₆(+1)
y	y ₁ (+1)	y ₂ (+1)	y ₃ (+1)	y ₄ (+1)	y ₅ (+1)	y ₆ (+1)

Table 5.1 – Frangmentation pattern of the isolated peptide H3(3+)For[L-norleucine] [3-fluoro-L-valine] [3-fluoro-L-valine] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-norleucine] [Serine] [Lysine]OH.

The non-natural peptide was further fragmented by high energy collision dissociation. The

proteomics analysis identified the primary structure of the peptide by using the fragments listed in Table 5.1.

Moreover, we designed the DNA encoding another non-natural peptide composed of the same but swapped residues. After TX-TL in PURE, and cleavage from GFP, the non-natural peptide was detected (see Figure 5.3), and its primary structure was identified (see Table 5.2).

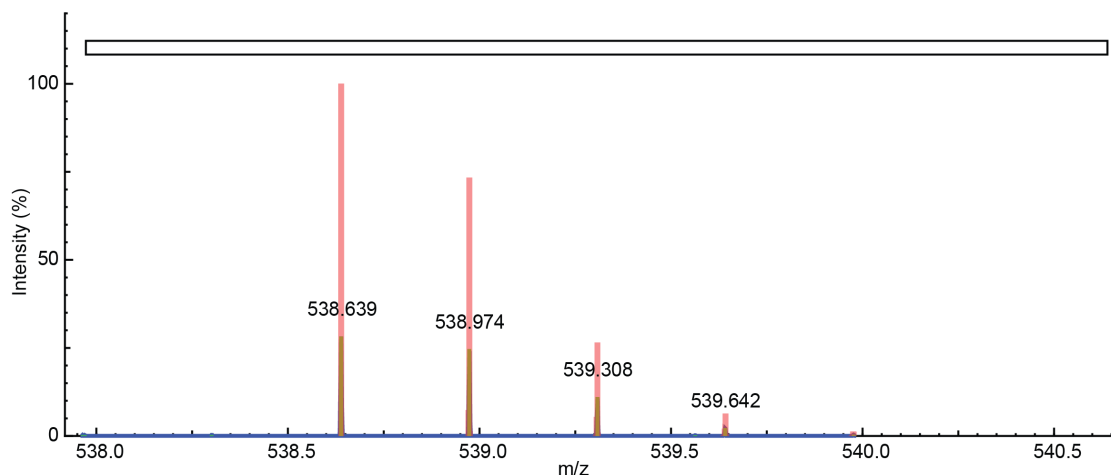


Figure 5.3 – MS spectrum of the non-natural peptide H3(3+)For[L-norleucine] [3-fluoro-L-valine] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline][L-norleucine][3-fluoro-L-valine] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [Serine] [Lysine]OH expressed at the N-terminus of GFP, and obtained by in gel digestion with Lys-C. Theoretical observed mass (m/z) = 538.639; experimental closest peak (m/z) = 538.639. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

Identified fragments									
b	b₄(+1)			b₇(+1)			b₈(+1)		
y	y ₂ (+1)	y ₄ (+1)	y ₅ (+2)	y ₆ (+1)	y ₇ (+1)	y ₈ (+1)	y ₉ (+1)	y ₁₀ (+1)	y ₁₁ (+3)

Table 5.2 – Frangmentation pattern of the isolated peptide H3(3+)For[L-norleucine] [3-fluoro-L-valine] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline][L-norleucine][3-fluoro-L-valine] [L-canavanine] [L-3-hydroxy-norvaline] [L-3-hydroxy-norvaline] [Serine] [Lysine]OH.

The results presented in this chapter are a small set of preliminary experiments intended to demonstrate the possibility of polymerizing synthetic sequence-defined polymers ribosomally, in a purified cell-free TX-TL system. However, much work is needed to scale-up this approach. For this reason, the non-natural peptide used in Chapter 4 (that has the same molecular structure of the first non-natural peptide described herein) has been obtained by SPPS.

6 Conclusions and Outlook

Presented in this thesis is a first step towards the development of a strategy for handling (*i.e.* producing, and disposing) man-made polymer-based materials in a sustainable fashion. NaCRe is inspired by the way Nature recycles proteins, and nucleic acids. It is based on sequence-defined polymers (not strictly necessary for the initial cycle) whose backbone can be reversibly depolymerized, and on polymerization reactions from unseparated mixtures of monomers.

This thesis is intended as proof-of-concept through which we demonstrate the overall feasibility of NaCRe by recycling peptides, proteins, protein-based materials, and mixtures of them into biotechnologically relevant proteins, outside living organisms.

Specifically, we have shown that short peptides, characterized by different secondary structures, hence functions, can be recycled by NaCRe into fluorescent proteins (that is polymers completely unrelated to the “parent” materials) through a series of enzymatic reactions carried out extracellularly. The process achieves comparable yields by either depolymerizing the initial materials separately, or starting from mixtures of them. Currently, the need of separating polymeric materials before recycling is complex, expensive, and time-consuming¹⁶. Indeed, the possibility to recycle mixtures of unknown polymers into the polymer of need at the time of recycling is visionary.

Moreover, we have shown that NaCRe is not limited to short peptides but can work with proteins characterized by well-defined tertiary structures, such as β -lactoglobulin. Furthermore, we have proven the potential of NaCRe by recycling engineering relevant proteins, and

Chapter 6. Conclusions and Outlook

protein-based materials such as silk fibroins, and β -lactoglobulin films, either composed of high molecular weight polymer chains, or by the supramolecular assembly of low molecular weight peptides.

We have then demonstrated that complex mixtures composed of peptides, and high molecular weight proteins (glucagon, β -lactoglobulin, and silk fibroins) can be recycled by NaCRe into fluorescent proteins, as well as into bioactive enzymes. Particularly noteworthy is that such process can be achieved without the need of any external monomer feed, that is we successfully generated by NaCRe high-value (virgin quality) polymers from unseparated mixtures of materials, without further depleting petrochemical sources³.

We have shown that NaCRe can be theoretically repeated “infinite times” by recycling glucagon, β -lactoglobulin, and silk fibroins into GFP that in turn was recycled into mScarlet-i. This result has been achieved by either purifying the GFP produced in the first cycle of NaCRe, or directly recycling the whole TX-TL system into mScarlet-i. This experiment is key since it demonstrates experimentally the robustness of NaCRe that is able to perform multiple cycles of recycling starting from dozens of proteins, even in presence of other polymers such as nucleic acids. Moreover, differently from the chemical recycling of “infinitely recyclable” polyesters, where linear polymers are recycled either into themselves, or into cyclic chains of the same monomer³¹, NaCRe is able to produce a different output material each time.

Finally, we have demonstrated that NaCRe is not limited to the set of the proteinogenic amino acids by recycling a few non-natural residues into a modified GFP chain, that is obtaining a sequence-defined polymer composed of natural, and non-natural monomers.

The results summarized above demonstrate that, were most of man-made polymers SDPs compatible with NaCRe, recycling an unknown mixture of polymers would mean producing the material of need, at the time of recycling, without being able to distinguish a new from a recycled material. Moreover, generating raw building blocks from petrochemical, or biological sources would not be needed because monomers would be kept in circulation, thus limiting the burden on the planet.

The work discussed in this thesis is a first step towards the development of a sustainable model for handling human-made polymers, inspired by Nature, and fulfilling the principles of

circular-economy, hence a variety of NaCRe developments can be envisioned.

Indeed, the experiments described so far have been focused on establishing NaCRe at the molecular level, that is achieving the depolymerization of proteins into their constitutive building blocks, and using such residues as the only source of monomers to perform the synthesis of an arbitrary different protein, which has nothing in common with the “parent” materials.

As summarized above, we have successfully recycled a technologically relevant film composed of β -lactoglobulin amyloids into GFP. However, an important step forward would be showing that the polymer produced by NaCRe could be used to build a realistic engineering device, and that such device could be in turn recycled into a different one, by performing consecutive NaCRe cycles.

For achieving this goal, a possibility could be starting from a protein-based composite material, composed of a matrix of β -lactoglobulin amyloids^{102,103}, and a filler. The matrix would be depolymerized to monomers by means of consecutive enzymatic reactions. The filler would be retrieved by filtration, or precipitation. The amino acids achieved by the digestion of the matrix would be recycled by NaCRe into silk fibroins. Such protein chains would be used to build a resistive switching memory device¹⁰¹. Such device would be in turn taken apart, and the fibroins would be depolymerized to amino acids that would be used as monomers for the synthesis of keratin. Keratin would be used to produce a photoresist material¹⁰⁴. The so obtained material (not exposed to light) could be recycled back into the initial composite material by depolymerizing keratin to monomers, and by polymerizing them into β -lactoglobulin. The composite material would be produced once more by feeding the amyloids, obtained from β -lactoglobulin, with the initial fillers.

The process discussed above is intended to show that NaCRe is envisioned to go beyond the molecular level, and become a powerful strategy to transform polymeric materials retrieved from realistic devices into different engineering relevant ones, that is building the device of need from a recycled one. Moreover, the possibility of recycling by NaCRe composite materials through the depolymerization of the matrix, and the separation of the fillers leads to a variety of processes, where both monomers, and fillers are kept in circulation.

In order to perform the process described above successfully, the optimization of time, and

Chapter 6. Conclusions and Outlook


temperatures for the enzymatic reactions, the improvement of the technological operations such as enzyme removal, as well as the scale up of the whole NaCRe process are needed. For this reason, a robotic platform able to automate the series of repetitive operations in NaCRe is currently being developed in our lab.

So far NaCRe has been established for proteins, and for sequences composed of proteino-genic backbones, and a few non-natural lateral chains. It would be interesting to change the backbone of the sequence-defined polymers involved in the process by performing NaCRe with nucleic acids such as DNA. In this case the ribosomal expression would be substituted with PCR, and the amino acids either with natural, or with functionalized nucleosides⁹⁵. A phosphorylation step would be needed between depolymerization, and synthesis.

After showing that NaCRe could work with either peptide, or phosphodiester backbones, the challenge would be attempting to perform NaCRe on sequence-defined polymers obtained through a fully synthetic approach. To achieve this goal, novel templated methods inspired by the PCR technique, which does not require any separation of the monomers, should be developed.

To conclude, this thesis envisions a truly circular approach to handling man-made polymers, inspired by Nature, by working with protein-based materials, outside living organisms. Were NaCRe to be successful in the future, the current paradigm of recycling a polymeric material into itself would be disrupted.

Appendix



The appendix is an extract of the *Supplementary Information* of the manuscript entitled *Nature-inspired Circular-economy Recycling (NaCRe) for Proteins: Proof of Concept* that is available as a preprint (<https://www.biorxiv.org/content/10.1101/2020.09.23.309799v3>, doi: 10.1101/2020.09.23.309799 version 3).

The authors of the manuscript are Simone Giaveri, Adeline M. Schmitt, Laura Roset Julià, Vincenzo Scamarcio, Anna Murello, Shiyu Cheng, Laure Menin, Daniel Ortiz, Luc Patiny, Sreenath Bolisetty, Raffaele Mezzenga, Sebastian J. Maerkl, and Francesco Stellacci.

The author of the thesis is the first author of the manuscript, and is the main contributor for designing, and performing the experiments, characterizing the obtained materials, analyzing the data, and discussing them with the collaborators. All the work has been performed under the supervision of Prof. Francesco Stellacci, and Prof. Sebastian J. Maerkl.

A Materials

Natural peptides and proteins. Magainin II, glucagon, somatostatin 28, β -lactoglobulin A from bovine milk, and silk fibroin solution were purchased from Sigma-Aldrich. β -lactoglobulin amyloids solution was kindly provided by Mezzenga's lab (ETH). *Non-natural peptide.* [L-norleucine][3-fluoro-DL-valine][3-fluoro-DL-valine][L-canavanine][DL-3-hydroxynorvaline][DL-3-hydroxynorvaline][L-canavanine][DL-3-hydroxynorvaline][DL-3-hydroxynorvaline][DL-3-hydroxynorvaline][L-norleucine][Ser][Lys] unnatural peptide was custom-synthesized by Sigma-Aldrich. *Expression of the calibrants.* pET29b(+) vector was purchased from Twist Bioscience. BL21 (DE3) cells were supplied by Lucigen. LB-Agar, Benzoinase, Imidazole, Magnesium acetate, Potassium glutamate were purchased from Sigma Aldrich. Kanamycin was supplied by MD Biomedical. Auto-induction TB medium was provided by Formedium. Protease inhibitor tablet was purchased from Roche. Glycerol, Sodium chloride, and HEPES were supplied by AppliChem. *PCR reagents.* gBlocks encoding GFP, mScarlet-i, and primers (fwd and rev) were purchased from IDT Integrated DNA Technologies. The gBlock encoding CDO was supplied by Twist Bioscience. 5x Phusion HF Buffer, dNTP Mix (10 mM), Phusion High-Fidelity DNA Polymerase ($2 \text{ U } \mu\text{L}^{-1}$), and DMSO were purchased from Thermo Fisher Scientific; nuclease-free water was supplied by Sigma-Aldrich. Q5 High-Fidelity 2x master mix was provided by New England Biolabs. 5x GelPilot DNA Loading Dye, and QIAquick PCR Purification Kit were purchased from Qiagen; GeneRuler 1 kb DNA Ladder (ready-to-use), and SYBR Safe DNA Gel Stain from Thermo Fisher Scientific. DNA Clean & ConcentratorTM was provided by Zymo Research. UltraPure Agarose was supplied by Invitrogen. 50x TAE buffer was pur-

Appendix A. Materials

chased from Jena Bioscience. *Cell-Free expression.* Magnesium acetate, Potassium glutamate, DL-Dithiothreitol (DTT), Creatine phosphate, Folinic acid, Spermidine, HEPES buffer, TCEP, catechol, Protector RNase Inhibitor, 20 proteinogenic AAs, L-canavanine, and L-norleucine were purchased from Sigma-Aldrich. ATP, GTP, CTP, and UTP were supplied by Thermo Fisher Scientific. tRNAs were purchased from Roche. PUREfrex™Solution II (enzymes), and PUREfrex™Solution III (ribosomes) were supplied by Kaneka Eurogentec^{SA}. FluoroTect™Green_{Lys} tRNA was provided by Promega. *Cleavage-depolymerization.* Thermolysin and pepsin were purchased from Promega. Leucine aminopeptidase (LAP) microsomal from porcine kidney (L9776, and L6007), TRIS hydrochloride, Calcium chloride, and Potassium hydroxide were supplied by Sigma-Aldrich. Fuming hydrochloric acid was purchased from ABCR Chemicals. *Mass Spectrometry.* Ammonium formate (LC/MS) was purchased from Thermo Fisher Scientific. Acetonitrile (ULC-MS) was supplied by Biosolve. Formic acid was purchased from Acros Organics. Trifluoroacetic acid, ethanol, Ammonium bicarbonate, and Iodoacetamide were supplied by Sigma-Aldrich. Dithioerythritol was purchased from Millipore. Chymotrypsin (sequencing grade), and trypsin (sequencing grade) were supplied by Promega. *Protein electrophoresis.* Precision Plus Protein™Unstained Protein Standards was purchased from Biorad. BenchMark™Fluorescent Protein Standard, NuPAGE™4-12% Bis-Tris mini protein gel, and 20x Novex™MES SDS Running Buffer were supplied by Thermo Fisher Scientific. 2x Laemmli buffer was provided by Sigma-Aldrich. InstantBlue stain was purchased from Lucerna-Chem. *Protein purification.* HisPur™Ni-NTA beads were provided by Thermo Fisher Scientific; MagneHis™protein purification system was supplied by Promega. *Filters-membranes-tools.* Amicon Ultra-0.5 mL Centrifugal Filters (3K, 10K, and 100K), 25 mm diameter, mixed cellulose esters (MCE) membranes, and C18 ZipTips were supplied by Millipore. 0.22 µm HPLC certified Nylon filter (PES) were purchased from Pall, and Protein LoBind Tubes from Eppendorf. Nunc™384-well optical bottom plates, HisPur™Ni-NTA beads, dialysis membranes, and 0.45 µm syringe filters, DynaMag™spin magnet, and Pierce™C18 StageTips were supplied by Thermo Fisher Scientific. SealPlate sealing film was purchased from Sigma-Aldrich. Polypropylene columns were provided by Bio-Rad.

All chemicals were used without any further purification.

B Methods

A detailed description of the methods used in this thesis are reported below.

B.1 Calibrant expression

Buffers preparation. Buffer A (NaCl (300 mM), HEPES (20 mM),), buffer B (NaCl (500 mM), HEPES (20 mM), imidazole (500 mM), pH 7.6), and storage buffer (HEPES (50 mM), Magnesium acetate (11.8 mM), Potassium glutamate (100 mM), pH 7.6) were prepared. *Expression.* The constructs were synthesized (as codon-optimized) for expression in *E. coli*, appended with a 6xHis tag at C-terminus, and cloned into pET29b(+) vector. mScarlet-i, and GFP constructs are reported in Table C.4. The plasmid was transformed into BL21 (DE3) cells by using the heat-shock method. Cells were plated onto LB-Agar plates containing kanamycin, and incubated overnight at 37 °C. A streak of colonies was picked, and grown in a LB broth (50 mL) containing kanamycin. The saturated overnight culture (40 mL) was inoculated into auto-induction TB medium (2 L) containing kanamycin, in a baffled flask (5 L). The culture was shaken at 37 °C for 3 h until the temperature was set to 20 °C for 18 h. The culture was harvested by centrifugation at 5000 rcf for 10 min in Thermo Fisher Scientific Lynx Sorvall. The pellet was resuspended in minimal volumes of buffer A, and frozen at –20 °C. *Purification.* The pellets were defrosted in 10/90 v/v glycerol:water, supplemented with Benzonase (5 µL), and 1 protease inhibitor tablet. The resuspended mixture was lysed by sonication, and spun down at 20000 rcf for 30 min in Thermo Fisher Scientific Lynx Sorvall. The soluble fraction was recovered, and

Appendix B. Methods

filtered by using a 0.45 μm syringe filter. The lysate was mixed with HisPur Ni-NTA beads (5 mL), and incubated at 4 °C for 60 min in rotator. The beads were transferred to a disposable polypropylene column, and washed with buffer A (20 column volumes). The proteins were purified by step-wise gradient purification by using 10/90 v/v buffer B:buffer A, 20/80 v/v buffer B:buffer A, 60/40 v/v buffer B:buffer A, and 100/0 v/v buffer B:buffer A (10 column volumes each). Fractions containing the desired proteins were pooled, dialyzed against storage buffer (3 L), concentrated in Amicon Ultra Centrifugal Filters (10K), and injected into GE Healthcare Superdex 200 26/600, pre-equilibrated in storage buffer. Peak fractions were pooled, and brought to (10 mgmL^{-1}) concentration approximately by using Amicon Ultra Centrifugal Filters (10K). *Quantification.* Proteins were quantified by 280 nm absorbance, and their predicted extinction coefficients (<https://web.expasy.org/protparam/>).

B.2 Film preparation

Filtration. The film was fabricated by vacuum filtration of the β -lactoglobulin amyloids solution (20 mgmL^{-1}) using a vacuum filtration assembly, and MCE membranes (pore size = 0.22 μm , diameter = 25 mm), following the protocol developed in Mezzenga's lab¹⁰². *Drying and film removal.* The film was left in the desiccator for 3 days to dry; the dry film was removed by a plastic spatula, and the powder was weighted.

B.3 Selection of the model peptides

The PDB database (updated on February 5, 2019) has been screened searching for 2 peptides composed of a number n of residues ($20 \leq n \leq 30$), cysteines = 0, and Unnatural/modified residues = 0, by using the script reported in the followings. The list of matches was further screened manually searching for commercially available peptides, presenting different secondary structures. Magainin II and glucagon were selected; together they contain all the proteinogenic amino acids except for cysteine and proline. Somatostatin 28 was selected *a posteriori*, as the source of the missing residues, looking for a highly structurally different material, i.e. disulfide cyclized. In detail, the Python3 script analyses all the PDB structure files contained in a given folder, filters the structures according to specified conditions, and creates

B.3. Selection of the model peptides

an output .txt file containing all the filtered chains. Bio, re, sys, os, joblib, multiprocessing, operator, and warnings are the Python Modules required. The script is reported below in B.1.

Listing B.1 – Python3 script for screening the PDB database

```
1
2  #Written by Anna Murello and Simone Giaveri (SuNMIL), EPFL
3  #January 2019
4
5  import Bio
6  from Bio.PDB import *
7  from Bio.Align import MultipleSeqAlignment
8  from distutils import spawn
9  import re
10 from Bio.SeqUtils.ProtParam import ProteinAnalysis
11 from Bio.PDB.PDBParser import PDBParser
12 from Bio.PDB.Polypeptide import three_to_one
13 from Bio.PDB.Polypeptide import is_aa
14 from Bio import Alphabet
15 from Bio.Data import IUPACData
16 from Bio.Data.SCOPData import protein_letters_3to1
17 import sys
18 from Bio import SeqIO
19 from Bio.PDB import PDBList
20 import os
21 import warnings
22 from Bio import BiopythonWarning
23 warnings.simplefilter('ignore', BiopythonWarning)
24 from joblib import Parallel, delayed
25 import multiprocessing
26 from operator import itemgetter
27 from Bio.SeqRecord import SeqRecord
28 from Bio.Seq import Seq
29
30 def Analysis (pdbfilename, l_max, l_min, aa1, aa2, count_aa1, count_aa2):
31  # this function reads a file (pdbfilename) and checks whether the filtering conditions
   are verified; if so it returns the sequence
32      Record = []
33
```

Appendix B. Methods

```
34     print(pdbfilename[:-4])
35
36     pdbfile = os.path.join(pdbdir, pdbfilename)
37
38     try:
39
40         parser = MMCIFParser()
41         structure = parser.get_structure('', pdbfile)
42
43         model = structure [0]  # the structure file may contain more than one model,
                                # the program analyses only the first one
44         if len(model)==1:
45
46             chain = model[ 'A' ]
47             seq = list()
48
49             for residue in chain: # checking the filtering conditions
50
51                 if is_aa(residue.get_resname(), standard=True):
52                     seq.append(three_to_one(residue.get_resname()))
53
54                 elif is_aa(residue.get_resname(), standard=False):
55                     seq.append('X')
56
57                 elif residue.get_resname()=='PYL' or residue.get_resname()=='XLE':
58                     seq.append('X')
59
60             myprot = str(''.join(seq))
61
62             length = len(myprot)
63
64
65             if length < l_max and length > l_min and myprot.count(aa1) < count_aa1 and
                myprot.count(aa2) != length \
66                 and myprot.count(aa2) < count_aa2:
67
68                 analysis = ProteinAnalysis(myprot)
69                 additional_features = {"length": length, "count_aa" : analysis.
```


B.3. Selection of the model peptides

```
count_amino_acids() , "count_X" : myprot.count('X')}  
70  
71 Record = [myprot, pdbfilename[: -4], additional_features]  
72  
73 except :  
74  
75     pass  
76  
77 return (Record)  
78  
79 # -----  
80  
81 with open("Output.txt", "w") as text_file :  
82     pdbdir = '/Users/simonegiaveri/Desktop/PythonPDB/Common_folder' # path for the  
        directoty containing the structure files  
83  
84     # in the following lines 'PYL' and 'XLE' amino acids are added to the dictionary of  
        the non-natural amino acids.  
85     protein_letters_3to1['PYL']='X'  
86     protein_letters_3to1['XLE']='X'  
87  
88     # in the following lines the filtering conditions are defined  
89     pdbfilenames = os.listdir(pdbdir)  
90     l_max = 50  
91     l_min = 5  
92     aa1 = 'C'  
93     aa2 = 'X'  
94     count_aa1 = 10  
95     count_aa2 = 1  
96  
97     results = Parallel(n_jobs=4)(delayed(Analysis)(pdbfilename, l_max, l_min, aa1, aa2,  
        count_aa1, count_aa2) for pdbfilename in pdbfilenames)  
98  
99     counter = 0  
100  
101 for a in results :  
102  
103     if a != []:
```

Appendix B. Methods

```
104         print(a[1],a[0],a[2], file=text_file)
105         print('\n', file=text_file)
106
107         counter = counter + 1
108
109         print('Number of filtered proteins = ', counter, file=text_file)
110
111         print('Total number of proteins = ',len(pdbfilenames), file=text_file)
112
113         print('Filtered as:', l_min, '< sequence length <', l_max, 'and number of ', aa1, '
114             <', count_aa1,'and number of ', aa2, '<', count_aa2, \
                ' and different from sequence length', file=text_file)
```

B.4 Depolymerization I (cleavage)

Enzymes preparation. Thermolysin was dissolved in buffer (Tris-HCl (50 mM), CaCl₂ (1 mM), KOH, pH 8) at 1 mgmL⁻¹ concentration; pepsin was reconstituted in water-HCl solution at 1.5 mgmL⁻¹ concentration. *Samples preparation (single cleavage reactions).* Magainin II and somatostatin 28 were prepared in (500 µL) buffer (Tris-HCl (50 mM), CaCl₂ (1 mM), KOH, pH 8), at 1 mgmL⁻¹ concentration. Glucagon was prepared in (500 µL) buffer (Tris-HCl (50 mM), CaCl₂ (1 mM), KOH, pH 9), at 1 mgmL⁻¹ concentration. β-lactoglobulin A was prepared in (500 µL) buffer (Tris-HCl (50 mM), CaCl₂ (1 mM), KOH, pH 8), at 2 mgmL⁻¹ concentration. Silk fibroin solution (50 mgmL⁻¹) was diluted in (500 µL) buffer (Tris-HCl (50 mM), CaCl₂ (1 mM), KOH, pH 8), at 2 mgmL⁻¹ concentration. β-lactoglobulin film powder was resuspended in (500 µL) water-HCl solution (pH 2.7), at 1.5 mgmL⁻¹ concentration. *Sample preparation (mixed cleavage reactions).* Magainin II, glucagon, and somatostatin 28 were prepared separately in (166.7 µL) buffer (Tris-HCl (50 mM), CaCl₂ (1 mM), KOH, pH 9), at 1 mgmL⁻¹ concentration; the solutions were then combined in equal volumes (500 µL). Glucagon, β-lactoglobulin A, and silk fibroin were prepared separately in (166.7 µL) buffer (Tris-HCl (50 mM), CaCl₂ (1 mM), KOH, pH 9), and combined in equal volumes (500 µL) to get a mixed protein solution at 1 mgmL⁻¹. *Sample preparation (Non-natural peptide).* [L-norleucine][3-fluoro-DL-valine][3-fluoro-DL-valine][L-canavanine][DL-3-hydroxynorvaline][DL-3-hydroxynorvaline][L-canavanine][DL-3-hydroxynorvaline][DL-3-

B.5. Depolymerization II (depolymerization)

hydroxynorvaline][DL-3-hydroxynorvaline][L-norleucine][Ser][Lys] was prepared in (500 μL) buffer (Tris-HCl (50 mM), CaCl_2 (1 mM), KOH, pH 8), at 2 mg mL^{-1} concentration. *Single cleavage reactions.* Magainin II, glucagon, somatostatin 28, and β -lactoglobulin A reactions (500 μL) were performed by 1/20 w/w thermolysin:protein. Reactions were run at 85°C , for 6 h into the Eppendorf Thermomixer C, at 300 rpm. Thermolysin was removed by cut-off filtration using Amicon Ultra-0.5 mL Centrifugal Filters (10K), previously washed with water, at 14000 rcf, 25°C in Eppendorf 5424R. Eluted solutions were frozen at -20°C for characterization, and further processing. The cleavage reaction (500 μL) for silk fibroin was incubated for additional 2 h. The cleavage of β -lactoglobulin film (500 μL) was performed by 1/20 w/w pepsin:protein. Reactions were run at 37°C , for 4 h into the Eppendorf Thermomixer C, at 450 rpm. The pH was adjusted to pH 8 using KOH and HCl before filtration. Pepsin was removed by cut-off filtration using Amicon Ultra-0.5 mL Centrifugal Filters (10K), previously washed with water, at 14000 rcf, 25°C in Eppendorf 5424R. Eluted solutions were frozen at -20°C for characterization, and further processing. *Mixed cleavage reactions.* The peptide mixture (500 μL) composed of magainin II, glucagon, and somatostatin 28, and the protein mixture (500 μL) composed of glucagon, β -lactoglobulin A, and silk fibroin were cleaved by using 1/20 w/w thermolysin:protein, following the same protocol described for single cleavage reactions. *Non-natural cleavage reactions.* The cleavage reaction (500 μL) of the unnatural peptide was performed by 1/20 w/w thermolysin:protein. Reactions were run at 85°C , for 8 h into the Eppendorf Thermomixer C, at 300 rpm. Thermolysin was removed by cut-off filtration using Amicon Ultra-0.5 mL Centrifugal Filters (10K), previously washed with water, at 14000 rcf, 25°C in Eppendorf 5424R. Eluted solutions were frozen at -20°C for characterization, and further processing.

B.5 Depolymerization II (depolymerization)

Enzymes preparation. Leucine aminopeptidase LAP (L9776 and L6007) were resuspended in nuclease-free water at 1 mg mL^{-1} concentration. *Samples preparation.* Cleaved samples were gently defrosted in ice. *Depolymerizations.* 80 μL of cleaved samples were supplemented with 20 μL of LAP solution. Reactions were run at 37°C , for 8 h into the Eppendorf Thermomixer C, at 300 rpm. LAP was removed by cut-off filtration using Amicon Ultra-0.5 mL

Centrifugal Filters (100K), previously washed with water, at 14000 rcf, 25 °C in Eppendorf 5424R. Eluted solutions were frozen at –20 °C for characterization, and further processing.

B.6 Polymerase Chain Reaction (PCR)

PCR batch (20 µL). The reaction was assembled by mixing 1 µL DNA linear gBlock template (1 ngµL⁻¹), 0.2 µL fwd. primer (50 µM), 0.2 µL rev. primer (50 µM), 4 µL 5x Phusion HF Buffer, 0.4 µL dNTP Mix (10 mM), 1 µL DMSO, 0.15 µL Phusion High-Fidelity DNA Polymerase (2 UµL⁻¹), and 13.05 µL nuclease-free water in a small PCR vial. *PCR batch (50 µL).* For amplifying the gBlock sequence encoding CDO, the reaction was assembled by mixing 1 µL DNA linear gBlock template (1 ngµL⁻¹), 2.5 µL fwd. primer (10 µM), 2.5 µL rev. primer (10 µM), 25 µL Q5 High-Fidelity 2x master mix, and 19 µL nuclease-free water in a small PCR vial. *PCR thermal cycle (20 µL batch).* Initialization was run at 98 °C for 2 min, denaturation at 98 °C for 20 s, annealing at 47 °C for 30 s, and extension at 72 °C for 45 s. Denaturation, annealing, and extension were repeated 35x. The reaction temperature was kept at 72 °C for additional 7 min, and decreased to 4 °C for storage. The whole thermal cycle was run into Thermo Fisher Scientific ProFlex™PCR System. *PCR thermal cycle (50 µL batch).* Initialization was run at 98 °C for 30 s, denaturation at 98 °C for 10 s, annealing at 70 °C for 30 s, and extension at 72 °C for 30 s. Denaturation, annealing, and extension were repeated 20x. The reaction temperature was kept at 72 °C for additional 5 min, and decreased to 4 °C for storage. *Casting of the gel.* The size of the amplified template was checked by running an agarose gel, prior to purification of the templated from the PCR batch. 1% Agarose gel was cast by mixing 0.4 g of Agarose into 40 mL of 1x TAE buffer; the suspension was heated in the microwave at 800 W for 90 s approximately, and added with 4 µL of SYBR Safe DNA Gel Stain. *Samples preparation.* 1 µL of PCR reaction was diluted adding 3 µL of nuclease-free water, and 1 µL of 5x GelPilot DNA Loading Dye; 5 µL of GeneRuler 1 kb DNA Ladder were used as reference. *Running conditions.* The gel was run at 60 V for 5 min followed by 120 V for 30 min in the Thermo Scientific EasyCast gel system. *Imaging.* The gel was imaged by using Thermo Fisher Scientific Benchtop 3UV transilluminator equipped with Kodak gel logic 100 imaging system, λ= 302 nm, 4s exposure. The gel is shown in Figure D.46. *Purification.* The PCR product was purified by combining multiple PCR batches, doubling the final volume by adding nuclease-free water,

and following the QIAquick PCR Purification Kit protocol. DNA was eluted by using 15 μ L of elution buffer per spin column. The 50 μ L batch was purified by using the DNA Clean & Concentrator protocol. *Quantification.* The final DNA concentration was measured using Witec NanoDrop 1000 spectrophotometer.

B.7 CF protein TX-TL

Energy solution preparation. The following solutions were prepared. SolutionA(-Salts - tRNAs - AAs) (2 mL): Creatine phosphate (147.06 mM), Folinic acid (0.15 mM), Spermidine (14.71 mM), DTT (7.4 mM), ATP (14.71 mM), GTP (14.71 mM), CTP (7.4 mM), UTP (7.4 mM), and HEPES (pH 7.6, 367.65 mM). Salts solution (2 mL): Magnesium acetate (184.38 mM), and Potassium glutamate (1.563 M). tRNAs solution (200 μ L): tRNAs (560 A_{260} mL⁻¹). tRNAs were quantified by using UV absorption A_{260} in Witec NanoDrop 1000 spectrophotometer. The three solutions were combined in a 25 μ L reaction, by mixing 3.4/1.6/2.5 v/v/v solutionA(-Salts - tRNAs - AAs):salts solution:tRNAs solution, in order to get the desired concentrations, adapted from Ueda and coworkers⁷³: Creatine phosphate (20 mM), Folinic acid (0.02 mM), Spermidine (2 mM), DTT (1 mM), ATP (2 mM), GTP (2 mM), CTP (1 mM), UTP (1 mM), HEPES (pH 7.6, 50 mM), Magnesium acetate (11.8 mM), Potassium glutamate (100 mM), and tRNAs (56 A_{260} mL⁻¹). For the CDO experiment the following premixed energy solution (2 mL) was prepared, substituting DTT with TCEP. Creatine phosphate (60 mM), Folinic acid (0.06 mM), Spermidine (6 mM), TCEP (3 mM), ATP (6 mM), GTP (6 mM), CTP (3 mM), UTP (3 mM), HEPES (pH 7.6, 150 mM), Magnesium acetate (35.4 mM), and Potassium glutamate (300 mM), and tRNAs solution (168 A_{260} mL⁻¹). *Cell-Free TX-TL reactions assembly (25 μ L).* 3.4 μ L of solutionA(-Salts - tRNAs - AAs), 1.6 μ L of salts solution, 2.5 μ L of tRNAs solution, 1.25 μ L PUREfrex™Solution II (enzymes), 1.25 μ L PUREfrex™Solution III (ribosomes), 0.5 μ L RNase inhibitor, 75 ng DNA, and 10 μ L of AAs were mixed in ice. Nuclease-free water was added to bring the reaction volume to 25 μ L. Cell-Free TX-TL reactions assembly (25 μ L) for CDO experiment. 8.33 μ L of premixed energy solution, 1.25 μ L PUREfrex™Solution II (enzymes), 1.25 μ L PUREfrex™Solution III (ribosomes), 0.5 μ L RNase inhibitor, 75 ng DNA, 10 μ L of AAs, and 2.5 μ L catechol in water solution (10 mM) were mixed in ice. Nuclease-free water was added to bring the reaction volume to 25 μ L. These volumes keep each reagent at the desired

Appendix B. Methods

concentration in the TX-TL reaction. *Cell-Free TX-TL reactions assembly (25 μ L) containing non-natural residues.* The reaction volume was supplemented with up to 12.5 μ L of AAs, and UAAs. *Magainin II, glucagon, and somatostatin 28 recycling into mScarlet-i. (Samples).* 10 μ L AAs solution was obtained by combining equal volumes (3.33 μ L) of magainin II, glucagon, and somatostatin 28 depolymerization solutions. *(Negative controls).* 10 μ L AAs solution was obtained by combining equal volumes (3.33 μ L) of magainin II, glucagon, and somatostatin 28 depolymerization solutions, prepared without adding the three peptides initially. *(Reference controls).* 10 μ L AAs solution was obtained by combining equal volumes (3.33 μ L) of three free AAs solutions, calculated from an ideal complete depolymerization of the initial peptides into free amino acids. *Magainin II, glucagon, and somatostatin 28 recycling into GFP. (Samples), (Negative controls), and (Reference controls)* as in *magainin II, glucagon, and somatostatin 28 recycling into mScarlet-i. (Spikes).* 0.5 μ L of L-cysteine hydrochloride in nuclease-free water solution (15 mM) was spiked in samples, negative controls, and reference controls TX-TL reactions. *Mixed magainin II, glucagon, and somatostatin recycling into GFP. (Samples).* 10 μ L AAs solution was obtained by the depolymerization of the magainin II, glucagon, and somatostatin 28 mixture. *(Negative controls).* 10 μ L AAs solution was obtained by preparing a depolymerization reaction of the magainin II, glucagon, and somatostatin 28 mixture, without adding the three peptides initially. *(Reference controls).* 10 μ L AAs solution was obtained by preparing a free AAs solution, calculated from an ideal complete depolymerization of the initial peptide mixture into free amino acids. *(Spikes).* 0.5 μ L of L-cysteine hydrochloride in nuclease-free water solution (15 mM) was spiked in samples, negative controls, and reference controls TX-TL reactions. *β -lactoglobulin A recycling into GFP. (Samples).* 10 μ L AAs solution was obtained by the depolymerization of β -lactoglobulin A. *(Negative controls).* 10 μ L AAs solution was obtained by preparing a depolymerization reaction of β -lactoglobulin A, without adding β -lactoglobulin A initially. *(Reference controls).* 10 μ L AAs solution was obtained by preparing a free AAs solution, calculated from an ideal complete depolymerization of β -lactoglobulin A into free amino acids. *(Spikes).* 0.5 μ L of L-cysteine hydrochloride in nuclease-free water solution (15 mM) was spiked in samples, negative controls, and reference controls TX-TL reactions. *Silk fibroin recycling into GFP. (Samples).* 10 μ L AAs solution was obtained by the depolymerization of silk fibroin solution. *(Negative controls).* 10 μ L AAs solution was obtained by preparing a depolymerization reaction of silk fibroin, without adding

silk fibroin initially. (*Reference controls*). 10 μ L AAs solution was obtained by preparing a free AAs solution, calculated from an ideal complete depolymerization of silk fibroin into free amino acids. (*Spikes*). 0.5 μ L of L-cysteine hydrochloride in nuclease-free water solution (15 mM), and 0.5 μ L of L-methionine in nuclease-free water solution (15 mM) were spiked in samples, negative controls, and reference controls TX-TL reactions. *β -lactoglobulin film recycling into GFP (Samples)*. 10 μ L AAs solution was obtained by the depolymerization of β -lactoglobulin film. (*Negative controls*). 10 μ L AAs solution was obtained by preparing a depolymerization reaction of β -lactoglobulin film, without adding β -lactoglobulin film powder initially. (*Spikes*). 0.5 μ L of L-cysteine hydrochloride in nuclease-free water solution (15 mM), 0.5 μ L of L-methionine in nuclease-free water solution (15 mM), and 0.5 μ L of L-histidine hydrochloride in nuclease-free water solution (15 mM) were spiked in samples, and negative controls TX-TL reactions. The free AAs solutions of all the reference controls were diluted (95/05 v/v reference control:endoprotease buffer, for magainin II, glucagon, somatostatin 28, and the peptides mix, and 90/10 v/v reference control:endoprotease buffer, for β -lactoglobulin A, and silk fibroin) and (80/20 v/v reference control:aminopeptidase buffer) consecutively, according to the cleavage, and depolymerization protocol that was undergone by the sample. *Non-natural residues recycling into GFP (Samples)*. 12.5 μ L AAs and UAAs solution was obtained by combining 9 μ L of the unnatural peptide depolymerization, 0.5 μ L of L-valine in nuclease-free water solution (15 mM), 0.5 μ L of L-threonine in nuclease-free water solution (15 mM), 0.5 μ L of L-leucine in nuclease-free water solution (15 mM), 0.5 μ L of L-lysine hydrochloride in nuclease-free water solution (15 mM), and 1.5 μ L of L-alanine, L-glycine, L-isoleucine, L-serine, L-proline, L-phenylalanine, L-tryptophan, L-tyrosine, L-aspartic acid, L-glutamic acid, L-histidine hydrochloride, L-glutamine, L-asparagine, and L-cysteine hydrochloride in nuclease-free water solution (5 mM each). An additional sample was prepared by spiking 1 μ L of FluoroTect™Green_{Lys} tRNA. (*Negative controls*). 9 μ L UAAs solution was obtained by preparing a depolymerization reaction of the unnatural peptide, without adding the unnatural peptide initially. An additional negative control was prepared by substituting 9 μ L of the negative control depolymerization with nuclease-free water. (*Reference controls*). 4.3 μ L AAs and UAAs solution was obtained by combining 0.4 μ L of L-canavanine in nuclease-free water solution (25 mM), 0.4 μ L of L-norleucine in nuclease-free water solution (25 mM), 0.5 μ L of L-valine in nuclease-free water solution (15 mM), 0.5 μ L of L-threonine in nuclease-free water

Appendix B. Methods

solution (15 mM), 0.5 μ L of L-serine in nuclease-free water solution (15 mM), 0.5 μ L of L-lysine hydrochloride in nuclease-free water solution (15 mM), and 1.5 μ L of L-alanine, L-glycine, L-isoleucine, L-leucine, L-proline, L-phenylalanine, L-tryptophan, L-tyrosine, L-aspartic acid, L-glutamic acid, L-histidine hydrochloride, L-glutamine, L-asparagine, and L-cysteine hydrochloride in nuclease-free water solution (5 mM each). An additional reference control was prepared by spiking 1 μ L of FluoroTect™Green_{Lys} tRNA. *Mixed glucagon, β -lactoglobulin A, and silk fibroin recycling into GFP. (Samples).* 10 μ L AAs solution was obtained by the depolymerization of the mixture composed of glucagon, β -lactoglobulin A, and silk fibroin. *(Negative controls).* 10 μ L AAs solution was obtained by preparing a depolymerization reaction of the glucagon, β -lactoglobulin A, and silk fibroin mixture, without adding the three materials initially. *(Reference controls).* 10 μ L AAs solution was obtained by preparing a free AAs solution, calculated from an ideal complete depolymerization of the initial protein mixture. *Mixed glucagon, β -lactoglobulin A, and silk fibroin recycling into CDO. (Samples).* 10 μ L AAs solution was obtained by the depolymerization of the mixture composed of glucagon, β -lactoglobulin A, and silk fibroin. *(Negative controls).* 10 μ L AAs solution was obtained by preparing a depolymerization reaction of the glucagon, β -lactoglobulin A, and silk fibroin mixture, without adding the three materials initially. *Cell-Free TX-TL reaction.* The reactions were gently mixed, transferred into a 384-well plate, sealed to avoid evaporation, spun down at 3000 rcf, 25 °C in Eppendorf 5810R, and incubated at 37 °C for 6 h (mScarlet-I, and CDO), and 4 h (GFP) in Thermo Fisher Scientific BioTek Synergy Mx plate reader. The plate reader parameters were the following: detection method = fluorescence, λ_{exc} = 569 nm (mScarlet-i), λ_{exc} = 488 nm (GFP), λ_{em} = 593 nm (mScarlet-i), λ_{em} = 507 nm (GFP), 1 min interval read, sensitivity = 90 % (mScarlet-i), sensitivity = 80 % (GFP), bottom optic position, fast continuous shaking. For the CDO experiment, the plate reader parameters were the following: detection method = absorbance, λ_{abs} = 385 nm (2-hydroxymuconate semialdehyde), 1 min interval read, fast continuous shaking. *Cell-Free TX-TL reaction containing non-natural residues.* The reactions were gently mixed, and incubated at 37 °C, for 6 h into the Eppendorf Thermomixer C, at 300 rpm. Additional expressions have been performed in the plate reader, as detailed above. *Data processing.* The TX-TL reactions were all run in duplicates. The expression curves represent the statistical mean of the results at any acquisition time; the shadow represents the standard deviation of the same data.

B.8 Second NaCRe cycle

Enzymes preparation. Thermolysin and LAP were prepared as described in Appendix B.4, and B.5. *Sample preparation.* 100 TX-TL reactions (25 μ L each) recycling a mixed solution (4 mL) of glucagon, β -lactoglobulin A, and silk fibroin into GFP were run in the plate reader, as described in Appendix B.7. In each reaction the volume of water was substituted by the same volume of AAs solution from recycling. The reactions were combined into three batches (750 μ L each), and the expressed GFP was purified by using MagneHis protein purification system. 30 μ L magnetic beads (15 min incubation, room temperature, rotating), 500 μ L binding/washing solution supplemented with 30 mg mL^{-1} Sodium Chloride (10 min incubation, 2 times repeat, room temperature, rotating), and 100 μ L elution buffer (500 mM imidazole, 15 min incubation, room temperature, rotating) were used for each batch. In each step beads were separated from the solution by using DynaMag spin magnet. The eluted batches were combined, diluted to 100 mM imidazole concentration in buffer (Tris-HCl (50 mM), CaCl_2 (1 mM), KOH, pH 8), buffer exchanged into the same buffer by using Amicon Ultra-0.5 mL Centrifugal Filters (3K), previously washed with water, at 14000 rcf, 25 °C in Eppendorf 5424R. The purified and buffer-exchanged GFP was recovered by reverse spinning at 1000 rcf, 25 °C, for 2 min, in Eppendorf 5424R. The fluorescence of the obtained GFP solution was inspected by using Invitrogen^{TME}-GelTM Safe ImagerTM (emission max of the blue LED = 470 nm), and the purity of the solution was checked by protein electrophoresis, as described in Appendix D.48. The whole purification process was performed a second time on the first supernatant solution. The obtained protein solutions were combined (90 μ L), and GFP was quantified by using Implen NanoPhotometer N60, and its predicted extinction coefficients (<https://web.expasy.org/protparam/>). *Sample preparation (without purification).* 8 TX-TL reactions (25 μ L each) recycling a mixed solution (4 mL) of glucagon, β -lactoglobulin A, and silk fibroin into GFP were run into the plate reader, as described in Appendix B.7. The reactions were combined, and filtrated by using Amicon Ultra-0.5 mL Centrifugal Filters (3K), previously washed with water, at 14000 rcf, 25 °C, in Eppendorf 5424R, in order to remove the unconsumed AAs during TX-TL of GFP. The retentate (50 μ L) was recovered by reverse spinning at 1000 rcf, 25 °C, for 2 min. *Cleavage reactions.* The cleavage reaction (150 μ L) of the purified GFP was performed by 1/10 w/w thermolysin:protein. Buffer (Tris-HCl (50 mM),

Appendix B. Methods

CaCl₂ (1 mM), KOH, pH 8) was added to bring the reaction volume to 150 μ L. The reaction was run at 85 °C, for 6 h into the Eppendorf Thermomixer C, at 300 rpm. Thermolysin was removed by cut-off filtration using Amicon Ultra-0.5 mL Centrifugal Filters (10K), previously washed with water, at 14000 rcf, 25 °C in Eppendorf 5424R. The eluted solution was frozen at –20 °C for further processing. The cleavage of the sample without purification (175 μ L) was performed by 50/100/25 v/v/v retentate:buffer (Tris-HCl (50 mM), CaCl₂ (1 mM), KOH, pH 8):thermolysin solution. *Depolymerizations*. Cleaved samples were gently defrosted in ice. 140 μ L of cleaved sample (from the cleavage of the purified GFP) was supplemented with 15 μ L of LAP solution. The depolymerization reaction was run at 37 °C, for 8 h into the Eppendorf Thermomixer C, at 300 rpm. LAP was removed by cut-off filtration using Amicon Ultra-0.5 mL Centrifugal Filters (100K), previously washed with water, at 14000 rcf, 25 °C in Eppendorf 5424R. The eluted solution was frozen at –20 °C for characterization, and further processing. The depolymerization of the cleaved sample (from the cleavage of the sample without purification) was performed as described in Appendix B.5. *Cell-Free TX-TL reactions assembly* (25 μ L). 3.4 μ L of solutionA(-Salts - tRNAs - AAs), 1.6 μ L of salts solution, 2.5 μ L of tRNAs solution, 1.25 μ L PUREfrex™Solution II (enzymes), 1.25 μ L PUREfrex™Solution III (ribosomes), 0.5 μ L RNase inhibitor, and 75 ng DNA were mixed in ice. The reactions were brought to volume by using the AAs solutions. *GFP (purified from the first NaCRe cycle) recycling into mScarlet-i. (Samples)*. The AAs solution was obtained by the depolymerization of the purified GFP, produced by the first NaCRe cycle. *(Negative controls)*. The AAs solution was obtained by preparing a depolymerization reaction of the purified GFP, produced by the first NaCRe cycle, without adding the purified GFP initially. *GFP (without purification) recycling into mScarlet-i. (Samples)*. The AAs solution was obtained by the depolymerization of GFP, produced by the first NaCRe cycle, together with the protein components of the TX-TL system. *(Negative controls)*. The AAs solution was obtained by preparing a depolymerization reaction of the mixture of GFP and protein components of the TX-TL system, without adding such mixture initially. *Cell-Free TX-TL reaction*. The reaction conditions were as described in Appendix B.7. *Data processing*. The data processing was performed as described in Appendix B.7.

B.9 Peptide (and UAAs) analysis by mass spectrometry

Sample preparation. Initial peptides, and samples of cleavage reactions were gently defrosted in ice and diluted to 0.1 mM concentration range by using the electrospray (ESI) solution (50/49.9/0.1 v/v/v acetonitrile:water:Formic acid). Glucagon peptide, [L-norleucine][3-fluoro-DL-valine][3-fluoro-DL-valine][L-canavanine][DL-3-hydroxynorvaline][DL-3-hydroxynorvaline][L-canavanine][DL-3-hydroxynorvaline][DL-3-hydroxynorvaline][DL-3-hydroxynorvaline][L-norleucine][Ser][Lys] unnatural peptide, and its cleaved fragments were desalted by Solid Phase Extraction using C18 ZipTips. Two steps elution was performed using first 60/39.9/0.1 v/v/v ACN:nuclease-free water:TFA then 80/19.9/0.1 v/v/v ACN:nuclease-free water:TFA. *Analysis.* Qualitative mass spectrometry analyses were performed on a Thermo Fisher Scientific LTQ Orbitrap Elite FTMS mass spectrometer operated in positive ionization mode, interfaced with a robotic chip-based nano-ESI source (TriVersa Nanomate, Advion Biosciences, Ithaca, NY, U.S.A.). A standard data acquisition and instrument control system was utilized (Thermo Fisher Scientific) whereas the ion source was controlled by Chipsoft 8.3.1 software (Advion BioScience). 20 µL of samples were loaded onto a 96-well plate (Eppendorf, Hamburg, Germany) within an injection volume of 5 µL. The experimental conditions for the ionization voltage were +1.4 kV and the gas pressure was set at 0.30 psi. The temperature of the ion transfer capillary was set to 300 °C. *Data processing.* Data were analyzed using XCalibur software (Thermo Fisher Scientific); compounds and fragments were identified by using apm2S software (https://ms.epfl.ch/applications/peptides_and_proteins/)^{114,115}.

B.10 Amino Acid Analysis by mass spectrometry (AAA)

Sample preparation. Samples of depolymerization reactions were gently defrosted in ice, and analyzed in triplicates without any further preparation. (For a more accurate quantification of serine, samples were additionally diluted 1/9 v/v sample:nuclease-free water). *Analysis.* Quantitative analyses were conducted on the 6530 Accurate-Mass Q-TOF LC/MS mass spectrometer coupled to the 1290 series UHPLC system (Agilent Technologies). 1.5 µL aliquots of the depolymerizations were injected onto a 2.1 x 100 mm, 2.7 µm Agilent Infinity-Lab Poroshell 120 HILIC-Z column heated at 25 °C. A binary gradient system consisted of A

Appendix B. Methods

(10/90 v/v 200 mM Ammonium Formate in Formic acid-water solution, pH 3:water), and B (10/90 v/v 200 mM Ammonium Formate in Formic acid-water solution, pH 3:acetonitrile). Sample separation was carried out at 0.5 mL min⁻¹ over a 16 min total run time. The initial condition was 0/100 v/v A:B. The proportion of the solvent B was linearly decreased from 0/100 v/v A:B to 30/70 v/v A:B, from 0 min to 10 min. From 10 min to 11 min the percentage of B was further increased linearly from 30/70 v/v A:B to 0/100 v/v A:B. The system was re-equilibrated in initial conditions for 3 min. Detection was operated in positive ionization mode using the Dual AJS Jet stream ESI Assembly. The instrument was operated in the 4 GHz high-resolution mode and calibrated in positive full scan mode using the ESI-L+ solution (Agilent Technologies). The nebulizer pressure was set at 45 psi, and the capillary voltage was set at 3.5 kV. AJS settings were as follows: drying gas flow, 7 L min⁻¹; drying gas temperature, 300 °C; nebulizer pressure, 45 psi; capillary voltage, 3500 V; fragmentor voltage, 75 V; skimmer voltage, 65 V; octopole 1 RF voltage, 750 V. *Data processing.* Data were analyzed by using MassHunter Qualitative Analysis (Agilent Technologies. Inc.) and quantification performed using MassHunter Quantitative Analysis (Agilent Technologies. Inc.). Standards for calibration curves were prepared at 3 mM, 1.5 mM, 0.3 mM, 0.03 mM, 0.015 mM, and 0.006 mM in the buffer (Tris-HCl (50 mM), CaCl₂ (1 mM), KOH, pH 8) to account for matrix effects. Standards were analyzed in duplicates, Extracted Ion Chromatograms (XIC) were generated using a MEW of ±50 ppm and peaks area obtained after automated integration. For calibration curves, a second order fitting *i.e.* $y = a + bx + cx^2$ was selected to better fit the experimental data. *Statistical analyses.* Bar-plots of the statistical mean of the results of the repeated injections (triplicates) of each sample are shown; error bars represent the standard deviation of the same data. *Calculation of the ideal amino acid concentrations in the complete depolymerization.* The ideal concentrations of each amino acid were estimated considering the consecutive dilutions for cleavage (95/05 v/v sample:endoprotease, for magainin II, glucagon, somatostatin 28, and the peptides mix, and 90/10 v/v sample:endoprotease, for β-lactoglobulin A, and silk fibroin) and depolymerization (80/20 v/v sample:aminopeptidase). For silk fibroin depolymerization, a 1:1 Fib-L:Fib-H was assumed; the signal peptides were removed from both Fib-L (<https://www.uniprot.org/uniprot/P21828>) and H (<https://www.uniprot.org/uniprot/P05790>).

B.11 Protein electrophoresis (SDS-PAGE)

Samples preparation. Cell-free expressions were gently defrosted in ice, and aliquoted (10 μ L). The aliquots were diluted 50/50 v/v sample:Laemmli buffer, and incubated at 98 °C, for 4 min into Thermo Fisher Scientific ProFlex™PCR System. Denaturated samples were loaded on 4-12% Bis-Tris mini protein gel. *Running conditions.* Gels were run at 100 V for 10 min followed by 150 V for 35 min in the Hoefer se260 mini-vertical gel electrophoresis unit. *Staining and washing.* Gels were washed in Milli-Q water for 1 h shaking prior to Coomassie staining for 1 h by using InstantBlue stain. Gels were destained in Milli-Q water for 1 h shaking. *Imaging.* Gels were imaged by using Vilber Lourmat Fusion Fx Imaging System, λ = AlexaFluor 488 nm, 3 s exposure, Biorad GelDoc Go Imaging System, white tray, auto-exposure, Image Lab 6.1, and by using iPhone Xs. Gels are shown in Figures D.47-D.48.

B.12 Proteomic analysis

Sample preparation. SDS-PAGE gel lanes were excised and washed twice in 50/50 v/v nuclease-free water:ethanol solution, containing 50 mM Ammonium bicarbonate for 20 min, and dried by vacuum centrifugation. Samples reduction was performed by using 10 mM Dithioerythritol for 1 h at 56 °C. A washing-drying step as above described was repeated before performing the alkylation step with 55 mM Iodoacetamide for 45 min at 37 °C in the dark. Samples were washed-dried again, and cleaved overnight at 37 °C by using chymotrypsin (non-natural GFP), or trypsin (GFP, mScarlet-i) at a concentration of 12.5 ng μ L⁻¹ in water-based solution containing 50 mM Ammonium bicarbonate and 10 mM CaCl₂. The resulting peptides were extracted by using 70/25/5 v/v/v Ethanol:water:Formic acid solution twice for 20 min with permanent shaking. Samples were further dried by vacuum centrifugation and stored at -20 °C. Peptides were desalted by Solid Phase Extraction using C18 StageTips. Two steps elution was performed using first 80/19.9/0.1 v/v/v ACN:water:TFA then 80/10/9.9/0.1 v/v/v/v ACN:TFE:water:TFA, and dried by vacuum centrifugation prior to LC-MS/MS injections. *Analysis.* Samples were resuspended in 97.9/2/0.1 v/v/v water:ACN:TFA solution and nano-flow separations were performed on a Dionex Ultimate 3000 RSLC nano UPLC system (Thermo Fisher Scientific) on-line connected with an Exploris 480 Orbitrap Mass Spectrometer (Thermo

Appendix B. Methods

Fischer Scientific). A capillary precolumn (Acclaim Pepmap C18, 3 μm -100 \AA , 2 cm x 75 μm ID) was used for 8 μL sample trapping and cleaning. A 50 cm long capillary column (75 μm ID; in-house packed using ReproSil-Pur C18-AQ 1.9 μm silica beads; Dr. Maisch) was then used for analytical separations at 250 nL min^{-1} over 90 min, biphasic gradients, by using A (97.9/2/0.1 v/v/v water:ACN:TFA), and B (90/9.9/0.1 v/v/v ACN:water:TFA). Acquisitions were performed through Top Speed Data-Dependent acquisition mode using a cycle time of 1 s. First MS scans were acquired with a resolution of 120000 (at 200 m/z) and the most intense parent ions were selected and fragmented by High energy Collision Dissociation (HCD) with a Normalized Collision Energy (NCE) of 30% using an isolation window of 2 m/z . Fragmented ions were acquired with a resolution 30000 (at 200 m/z) and selected ions were then excluded for the following 30 s. The experimental conditions for the ionization voltage were +1.6 kV; the temperature of the ion transfer capillary was set to 175 $^{\circ}\text{C}$. *Data processing*. Raw data were processed using SEQUEST in Proteome Discoverer v.2.4 against a concatenated database consisting of the Uniprot *E.coli* protein database (4391 entries), and GFP, or mScarlet-i sequence. Enzyme specificity was set to chymotrypsin, or trypsin, and a minimum of six amino acids was required for peptide identification. Up to two missed cleavages were allowed. A 1% FDR cut-off was applied both at peptide and protein identification level. For the database search, carbamidomethylation was set as a fixed modification, whereas oxidation (Met), acetylation (protein N-term), PyroGlu (N-term Q), and Phosphorylation (Ser,Thr,Tyr) were considered as variable modifications. Data were further processed and inspected in Scaffold 4.10 (Proteome Software, Portland, USA), and spectra of interest were manually validated. *Data processing (non-natural)*. Data were analyzed manually by focusing on a few peptides of interest by using XCalibur software (Thermo Fisher Scientific); peptides and fragments were identified by using apm2S software (https://ms.epfl.ch/applications/peptides_and_proteins/)^{114,115}.

B.13 Mass calibration

A calibration curve for mScarlet-i expression is reported in Figure D.3. *Sample preparation*. mScarlet-i calibrant dissolved in buffer (HEPES (50 mM), Magnesium acetate (11.8 mM), Potassium glutamate (100 mM), pH 7.6) at 9.6 mg mL^{-1} concentration was diluted 1/24 v/v calibrant solution:nuclease-free water. Five calibrators were prepared further diluting such

protein solution 0.5/24.5 v/v protein solution:(TX-TL reaction -0.5 μL of nuclease-free water), 1/24 v/v protein solution:(TX-TL reaction -1 μL of nuclease-free water), 1.5/23.5 v/v protein solution:(TX-TL reaction -1.5 μL of nuclease-free water), 2/23 v/v protein solution:(TX-TL reaction -2 μL of nuclease-free water), and 2.5/22.5 v/v protein solution:(TX-TL reaction -2.5 μL of nuclease-free water). TX-TL reactions (25 μL) is composed of 3.4 μL of solutionA(-Salts - tRNAs - AAs), 1.6 μL of salts solution, 2.5 μL tRNAs solution, 1.25 μL PUREfrex™Solution II (enzymes), 1.25 μL PUREfrex™Solution III (ribosomes), 0.5 μL RNase inhibitor, 75 ng DNA, 3.33 μL magainin II depolymerization solution, 3.33 μL glucagon depolymerization solution, 3.33 μL somatostatin 28 depolymerization solution, and 4.06 μL of nuclease-free water. *Data collection.* The solutions were gently mixed, transferred into a 384-well plate, sealed to avoid evaporation, spun down at 4000 rcf, 25 °C in Eppendorf 5810R, and incubated at 37 °C for 40 min in Thermo Fisher Scientific BioTek Synergy Mx plate reader. The plate reader parameters were the following: $\lambda_{\text{exc}} = 569 \text{ nm}$ (mScarlet-i), $\lambda_{\text{em}} = 593 \text{ nm}$ (mScarlet-i), 1 min interval read, sensitivity = 90% (mScarlet-i), bottom optic position, fast continuous shaking. *Data processing.* For each calibrator, the statistical mean of the data collected between 15 min and 30 min was calculated (mScarlet-i maturation time is approximately 40 min); a linear fit *i.e.* $y = a + bx$ was used to better fit the experimental data. *Statistical analysis.* Error bars represent the variability of the expression using different lots of PUREfrex™Solution II, and III, calculated as the standard deviation of the expression plateau for a magainin II, glucagon, and somatostatin 28 recycling into mScarlet-i (reference control experiment). Curves are shown in Figure D.4.

B.14 AFM imaging

Sample preparation. Solutions of as prepared fibrils and depolymerized fibrils ($\sim 0.2 \text{ mg mL}^{-1}$) have been drop-casted on freshly cleaved mica, dried overnight in ambient conditions and kept under vacuum in a desiccator for 1 h, to completely remove the residues of water. *Analysis.* AFM images were collected in ambient conditions in amplitude modulation mode on a Cypher S system (Asylum Research/Oxford Instrument) using a HQ:NSC18/AI BS cantilever from mikroMasch. The sensitivity and spring constant of the cantilever were calibrated by using the GetReal™ automated probe calibration method. AFM images are shown in Figure 4.7.

C Additional Tables

A list of tables reporting the primary structures of the proteins, and the DNA sequences, used in this study, is reported in the followings.

Appendix C. Additional Tables

a) HGlylleGlyLysPheLeuHisSerAlaLysLysPheGlyLysAlaPheValGlyGlulleMetAsnSerOH
b) HHISerGlnGlyThrPheThrSerAspTyrSerLysTyrLeuAspSerArgArgAlaGlnAspPheValGlnTrpLeuMetAsnThrOH
c) HSerAlaAsnSerAsnProAlaMetAlaProArgGluArgLysAlaGlyCysLysAsnPhePheTrpLysThrPheThrSerCysOH
d) HMet(S-1CH ₂)Val(H-1F)Val(H-1F)Arg((CH ₂)-1O)Thr(CH ₂)Thr(CH ₂)Arg((CH ₂)-1O)Thr(CH ₂)Thr(CH ₂)Thr(CH ₂)Met(S-1CH ₂)SerLysOH
e) HMetLysCysLeuLeuLeuAlaLeuAlaLeuThrCysGlyAlaGlnAlaLeulleValThrGlnThrMetLysGlyLeuAspIleGlnLysValAlaGlyThrTrpTyrSerLeuAlaMetAlaAlaSerAspIleSerLeuLeuAspAlaGlnSerAlaProLeuArgValTyrValGluGluLeuLysProThrProGluGlyAspLeuGlulleLeuLeuGlnLysTrpGluAsnAspGluCysAlaGlnLysLysIleAlaGluLysThrLysIleProAlaValPheLysIleAspAlaLeuAsnGluAsnLysValLeuValLeuAspThrAspTyrLysLysTyrLeuLeuPheCysMetGluAsnSerAlaGluProGluGlnSerLeuValCysGlnCysLeuValArgThrProGluValAspAspGluAlaLeuGluLysPheAspLysAlaLeuLysAlaLeuProMetHisIleArgLeuSerPheAsnProThrGlnLeuGluGluGlnCysHisIleOH
f) HSerValThrIleAsnGlnTyrSerAspAsnGlulleProArgAspIleAspAspGlyLysAlaSerSerValIleSerArgAlaTrpAspTyrValAspAspThrAspLysSerIleAlalleLeuAsnValGlnGlulleLeuLysAspMetAlaSerGlnGlyAspTyrAlaSerGlnAlaSerAlaValAlaGlnThrAlaGlyIleAlaHisLeuSerAlaGlyIleProGlyAspAlaCysAlaAlaAlaAsnValIleAsnSerTyrThrAspGlyValArgSerGlyAsnPheAlaGlyPheArgGlnSerLeuGlyProPhePheGlyHisValGlyGlnAsnLeuAsnLeulleAsnGlnLeuValIleAsnProGlyGlnLeuArgTyrSerValGlyProAlaLeuGlyCysAlaGlyGlyGlyArgIleTyrAspPheGluAlaAlaTrpAspAlalleLeuAlaSerSerAspSerSerPheLeuAsnGluGluTyrCysIleValLysArgLeuTyrAsnSerArgAsnSerGlnSerAsnAsnIleAlaAlaTyrIleThrAlaHisLeuLeuProProValAlaGlnValPheHisGlnSerAlaGlySerIleThrAspLeuLeuArgGlyValGlyAsnGlyAsnAspAlaThrGlyLeuValAlaAsnAlaGlnArgTyrIleAlaGlnAlaAlaSerGlnValHisValOH
g) HAsnIleAsnAspPheAspGluAspTyrPhe ... LysPheArgAlaLeuProCysValAsnCysOH

Table C.1 – Primary sequences of the depolymerized proteins: magainin II (a), glucagon (b), somatostatin 28 (c), non-natural peptide (d), β -lactoglobulin A (e), silk fibroin Light chain (f), and silk fibroin Heavy chain (g). The complete sequence of the silk fibroin Heavy chain is available at (<https://www.uniprot.org/uniprot/P05790>); the signal peptides were removed from both the fibroin chains.

<p>a) (5')GCACCATCAGCCAGAAAACCGAACCAGCCAGAAAACGACCTTTCTGTGGATCTTAAGGCTAGAGTACT AATACGACTCACTATAGGGAGACCACAACGGTTCCCTCTAGAAATAATTTGTTAACTTAAGAGGAGG AAAAAAAATGGTAAGTAAAGGTGAAGCGGTAATTAAAGAGTTTATGCGCTTTAAGGTACACATGGAAG GCTCTATGAATGGGCACGAATTCGAGATCGAAGGTGAGGGTGAAGGACGCCCTACGAGGGCACTCAGA CTGCAAAGTTAAAAGTGACGAAAGGTGGCCCTTACCGTTTAGCTGGGATATCCTGTGCGCCGAGTTTATG TATGGCAGTCGTGCGTTTCAAGCACCTGCTGACATCCCTGATTACTATAAGCAATCATTCCCCGAAGGC TTTAAGTGGGAGCGTGTATGAACCTTGAAGATGGCGGAGCTGTGACTGTTACACAAGACACGAGCTTGG AAGACGGAACCTGATCTACAAAGTGAATACGCGGTACGAACTTTCCGCCTGACGGTCCAGTGATGCA GAAAAAGACCATGGGATGGGAAGCTAGCACCGAACGTTTATATCCGGAGGACGGCGTGCTTAAAGGTGA CATTAAATGGCATTACGTTTGAAGGATGGTGGTCTTATTAGCAGATTTAAGACTACGTATAAAGCGA AGAAGCCTGTACAAATGCCGGGTGCTTATAACGTAGATCGTAACTTGATATTACCTCGCACAAATGAAGAC TATACAGTAGTAGAACAATACGAACGCTCCGAGGGACGCCACTCTACTGGGGGCATGGACGAATTATACA AATCTGGATTACGCAGCCGTGCCAGGCCAGTAATCCGCAGTGGACGGTACGGCTGGACCGGGGTCTAC AGGTAGTCGTACCACCATCATCACCCTAATAACGACTCAGGCTGCTACCTAGCATAACCCCTGGGGCCCT CTAAACGGGTCTTGAGGGGTTTTTGGCAGGAAAGAACATGTGAGCAAAAGG(3')</p>
<p>b) (5')GCACCATCAGCCAGAAAACCGAACCAGCCAGAAAACGACCTTTCTGTGGATCTTAAGGCTAGAGTACT AATACGACTCACTATAGGGAGACCACAACGGTTCCCTCTAGAAATAATTTGTTAACTTAAGAGGAGG AAAAAAAATGGTCTCTAAAGGTGAAGAATTATTACTGGTGTGTGCCAATTTTGGTTGAATTAGATGGT GATGTTAATGGTCACAAATTTCTGTCTCCGGTGAAGGTGAAGGTGATGCTACTTACGGTAAATTGACCTT AAAATTTATTTGTACTACTGGTAAATTGCCAGTTCATGGCCAACCTTAGTCACTACTTTAACTTATGGTGT TCAATGTTTTCTAGATACCCAGATCATATGAAACAACATGACTTTTTCAAGTCTGCCATGCCAGAAGGTTA TGTTCAAGAAAGAACTATTTTTTCAAAGATGACGGTAACCTACAAGACCAGAGCTGAAGTCAAGTTTGAA GGTGATACCTTAGTTAATAGAATCGAATTAAGGTATTGATTTTAAAGAAGATGGTAACATTTTAGGTCA CAAATTGGAATACAACTATAACTCTCAATGTTTACATCATGGCTGACAAACAAAGAATGGTATCAAAG TTAACTTCAAATTAGACACAACATTGAAGATGGTCTGTTCAATTAGCTGACCATTATCAACAAATACTC CAATTGGTGATGGTCCAGTCTTGTACCAGACAACCATTAATCTCACTCAATCTGCCTTATCCAAAGATC CAAACGAAAAGAGAGACCACATGGTCTTGTAGAATTTGTTACTGCTGCTGGTATTACCTTAGGTATGGAT GAATTGTACAAACACCACCATCATCACCCTAATAACGACTCAGGCTGCTACCTAGCATAACCCCTGGGG CCTCTAAACGGGTCTTGAGGGGTTTTTGGCAGGAAAGAACATGTGAGCAAAAGG(3')</p>
<p>c) (5')GATCTTAAGGCTAGAGTACTAATACGACTCACTATAGGGAGACCACAACGGTTCCCTCTAGAAATAAT TTTGTTTAACTTAAGAGGAGGAAAAAAAATGAACAAAGGTGAATGCGACCGGGCCATGTGCAGCTG CGTGATCGGACATGAGCAAGGCCCTGGAACACTACGTCGAGTTGCTGGGCCATCGAGATGGACCGTG ACGACCAGGGCCGTGTCTATCTGAAGGCTTGACCGAAGTGGATAAGTTTTCCCTGGTGCTACGCGAGGC TGACGAGCCGGGCATGGATTTTATGGGTTTCAAGGTTGTGGATGAGGATGCTCTCCGGCAACTGGAGCGG GATCTGATGGCATATGGCTGTGCCGTTGAGCAGCTACCCGAGGTGAACGAAGTTGTGGCCGCGCG TGCGCTTCCAGGCCCCCTCCGGGCATCACTTCGAGTTGTATGCAGACAAGGAATATACTGGAAAGTGGGG TTTGAATGACGTCAATCCGAGGCATGGCCGCGCATCTGAAAGGTATGGCGGCTGTGCGTTTCGACCAC GCCCTCATGTATGGCGACGAATTGCCGGCGACCTATGACCTGTTACCAAGGTGCTCGGTTTCTATCTGGCC GAACAGGTGCTGGACGAAAATGGCACGCGCGTCGCCAGTTTCTCAGTCTGTGACCAAGGCCACGAC GTGGCCTTCATTACCATCCGAAAAAGGCCGCTCCATCATGTGTCCTTCCACCTCGAAACCTGGGAAGA CTTGCTTCGCGCCGCGACCTGATCTCCATGACCGACACATCTATCGATATCGGCCAACCCGCCACGGCCT CACTCACGGCAAGACCATCTACTTCTCGACCCGTCCGGTAACCGCAACGAAGTGTCTGCGGGGGAGATT ACAACATCCCGGACCACAAACCGGTGACCTGGACCACCGACAGCTGGGCAAGGCGATCTTTACCACGA CCGATTCTCAACGAACGATTATGACCGTGCTGACCTGAGGATCTGAAGCTTGGGCCGAACAAAAAAT CATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCATCATTGAGTTTAAACGGTCTCCAG CTTGGCTGTTTTGGCGGATGAGAGAAGATTTTACGCTGATACAGATTAAATCAGAACGCAAGCGGTC TGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACTGACCCCATGCCAACTCAGAAAGT GAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAAGTCCAGGCATCAAT AAAACGAAAGGCTCAGTCGAAAGACTGGGCCCTTTCGTTTATCTGTTGTTGTCGGTGAAC(3')</p>

Table C.2 – Linear templates (gBlocks) for expressing mScarlet-i (a), GFP (b), and CDO (c) in PUREFrex™. T7 Promoter, Terminators (T7, and TrnB), Ribosome Binding Site (RBS), and Opening Reading Frame (ORF) are highlighted in blue, green, yellow, and red respectively.

Appendix C. Additional Tables

a) (5')GATCTTAAGGCTAGAGTAC(3')
b) (5')CAAAAAACCCCTCAAGAC(3')
c) (5')GATCTTAAGGCTAGAGTACTAATACGACTCACTATAGGGAGACC(3')
d) (5')AGTTCACCGACAAACAACAGATAAAACGAAAGGCC(3')

Table C.3 – Forward (a) and reverse (b) primers for the PCR amplification of the gBlocks encoding mScarlet-i, and GFP. Forward (c) and reverse (d) primers for the PCR amplification of the gBlock encoding CDO.

a) (5')ATGGTGTCTAAGGGTGAGGCCGTCATTAAGGAATTTATGCGTTTTAAGGTGCACATGGAAGGATCGATGAATGGTCATGAGTTCGAGATCGAGGGCGAGGGTGAGGGTCGTCCATATGAGGGCACTCAAACAGCCAA GTTAAAAGTTACCAAAGGTGGTCCGTTGCCTTTATGCTGGGATATTCTCTCCCACAATTCATGTACGGATC ACGCGCCTTCATTAAGCACCCCGCCGACATTCCCGACTACTACAAGCAAAGCTTTCAGAGGGTTTTAAGT GGGAGCGTGTTATGAATTCGAGGACGGCGGAGCGGTTACGGTAACGCAAGATACGAGTCTGGAAGACG GTACGTTAATCTATAAAGTGAAGCTGCGCGGCACAAATTTCCGCCGGACGGTCCGGTCATGCAGAAGAA GACTATGGGCTGGGAAGCATCGACGGAACGCTGTATCCAGAGGATGGTGTGTTGAAAGGGGACATCAA AATGGCACTGCGCTTAAAGGACGGCGGCCGCTACTTAGCAGACTTTAAGACAACGTATAAGGCCAAGAA ACCGGTGCAAATGCCTGGGGCCTATAATGTGGACCGTAAATTAGACATCACCTCACACAATGAGGACTACA CCGTGGTAGAGCAGTATGAACGCTCGGAAGGCCGTCCTACTGTTGGCATGGATGAAGTGTATAAGTC AGGCCTTCGTAGCCGCGCGCAGGCCAGCAATAGTGCAGTTGATGGTACGGCAGGCCCTGGAAGCACAGG CTCACGTCACCATCACCACCACCACTGA(3')
b) (5')ATGGTAAGCAAGGGAGAGGAGTTATTTACGGGTGTTGTGCCATCTTAGTAGAGTTAGACGGTGATGT CAATGGACATAAATTTTCAGTGAGCGGCGAGGGAGAAGGAGATGCGACCTACGGCAAGTTAACTTTAAA ATTTATTTGTACCACGGGCAAATGCCAGTCCCGTGGCCAACTTAGTAACCACCCTCACCTACGGTGTGC AGTGCTTTTCGCGCTACCCGGATCATATGAAACAACATGACTTCTTCAAAGCGCTATGCCGGAAGGCTAC GTACAGGAACGCACCATTTTCTTCAAGGACGATGGAATTTATAAAACCCGTGCAGAGGTGAAATTTGAAG GTGACACACTTGTAACCGCATTGAACTGAAAGGCATTGATTTTAAAGAGGACGGCAACATTCTGGGTCA TAACTGGAATACAATACTACAATTCGCACAACGTATACATTATGGCAGATAAACAGAAGAATGGCATTAAAG TCAATTTCAAATCCGCCACAACATCGAAGACGGCTCCGTGCAGCTCGCCGATCACTATCAGCAGAATACT CCGATTGGAGATGGCCCCGTTCTGTTACCGGATAACCACTATCTTAGTACCCAGAGCGCCCTCTCGAAAGA CCCTAACGAGAAACGCGACCATATGGTGCTGCTGGAATTCGTGACGGCAGCCGGAATTACCTTAGGCATG GACGAGTTATAACAACCATCATCACCACCATTA(3')

Table C.4 – Constructs for expressing mScarlet-i calibrant (a), and GFP calibrant (b) into BL21 (DE3) cells, after cloning them into the pET29b(+) vector.

<p>a) ForMetValSerLysGlyGluAlaValIleLysGluPheMetArgPheLysValHisMetGluGlySerMetAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyThrGlnThrAlaLysLeuLysValThrLysGlyGlyProLeuProPheSerTrpAspIleLeuSerProGlnPheMetTyrGlySerArgAlaPhelleLysHisProAlaAspIleProAspTyrTyrLysGlnSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyAlaValThrValThrGlnAspThrSerLeuGluAspGlyThrLeulleTyrLysValLysLeuArgGlyThrAsnPheProProAspGlyProValMetGlnLysLysThrMetGlyTrpGluAlaSerThrGluArgLeuTyrProGluAspGlyValLeuLysGlyAspIleLysMetAlaLeuArgLeuLysAspGlyGlyArgTyrLeuAlaAspPheLysThrThrTyrLysAlaLysLysProValGlnMetProGlyAlaTyrAsnValAspArgLysLeuAspIleThrSerHisAsnGluAspTyrThrValValGluGlnTyrGluArgSerGluGlyArgHisSerThrGlyGlyMetAspGluLeuTyrLysSerGlyLeuArgSerArgAlaGlnAlaSerAsnSerAlaValAspGlyThrAlaGlyProGlySerThrGlySerArgHisHisHisHisHisHisOH</p>
<p>b) ForMetValSerLysGlyGluGluLeuPheThrGlyValValProIleLeuValGluLeuAspGlyAspValAsnGlyHisLysPheSerValSerGlyGluGlyGluGlyAspAlaThrTyrGlyLysLeuThrLeuLysPhelleCysThrThrGlyLysLeuProValProTrpProThrLeuValThrThrLeuThrTyrGlyValGlnCysPheSerArgTyrProAspHisMetLysGlnHisAspPhePheLysSerAlaMetProGluGlyTyrValGlnGluArgThrIlePhePheLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeuValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHisLysLeuGluTyrAsnTyrAsnSerHisAsnValTyrIleMetAlaAspLysGlnLysAsnGlyIleLysValAsnPheLysIleArgHisAsnIleGluAspGlySerValGlnLeuAlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAspAsnHisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHisMetValLeuLeuGluPheValThrAlaAlaGlyIleThrLeuGlyMetAspGluLeuTyrLysHisHisHisHisHisHisOH</p>
<p>c) ForMetAsnLysGlyValMetArgProGlyHisValGlnLeuArgValLeuAspMetSerLysAlaLeuGluHisTyrValGluLeuLeuGlyLeulleGluMetAspArgAspAspGlnGlyArgValTyrLeuLysAlaTrpThrGluValAspLysPheSerLeuValLeuArgGluAlaAspGluProGlyMetAspPheMetGlyPheLysValValAspGluAspAlaLeuArgGlnLeuGluArgAspLeuMetAlaTyrGlyCysAlaValGluGlnLeuProAlaGlyGluLeuAsnSerCysGlyArgArgValArgPheGlnAlaProSerGlyHisHisPheGluLeuTyrAlaAspLysGluTyrThrGlyLysTrpGlyLeuAsnAspValAsnProGluAlaTrpProArgAspLeuLysGlyMetAlaAlaValArgPheAspHisAlaLeuMetTyrGlyAspGluLeuProAlaThrTyrAspLeuPheThrLysValLeuGlyPheTyrLeuAlaGluGlnValLeuAspGluAsnGlyThrArgValAlaGlnPheLeuSerLeuSerThrLysAlaHisAspValAlaPhelleHisHisProGluLysGlyArgLeuHisHisValSerPheHisLeuGluThrTrpGluAspLeuLeuArgAlaAlaAspLeulleSerMetThrAspThrSerIleAspIleGlyProThrArgHisGlyLeuThrHisGlyLysThrIleTyrPhePheAspProSerGlyAsnArgAsnGluValPheCysGlyGlyAspTyrAsnTyrProAspHisLysProValThrTrpThrThrAspGlnLeuGlyLysAlaIlePheTyrHisAspArgIleLeuAsnGluArgPheMetThrValLeuThrOH</p>

Table C.5 – Primary sequences of the expressed proteins: mScarlet-i (a), GFP (b), and CDO (c).

D Additional Data

A list of additional experiments (expressions, calibrations, further MS characterizations, DNA, and proteins gels) is reported in the followings.

Appendix D. Additional Data

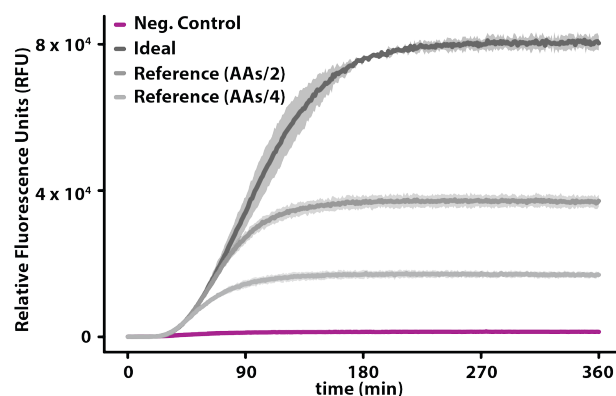


Figure D.1 – Plots of the fluorescence signal resulting from the expression of mScarlet-i in our TX-TL system by using 3 different initial concentrations (25/75 v/v reference AAs mixture:nuclease-free water, 50/50 v/v reference AAs mixture:nuclease-free water, and 100/0 v/v reference AAs mixture:nuclease-free water) of the reference AAs mixture from magainin II, glucagon, and somatostatin 28 complete depolymerization.

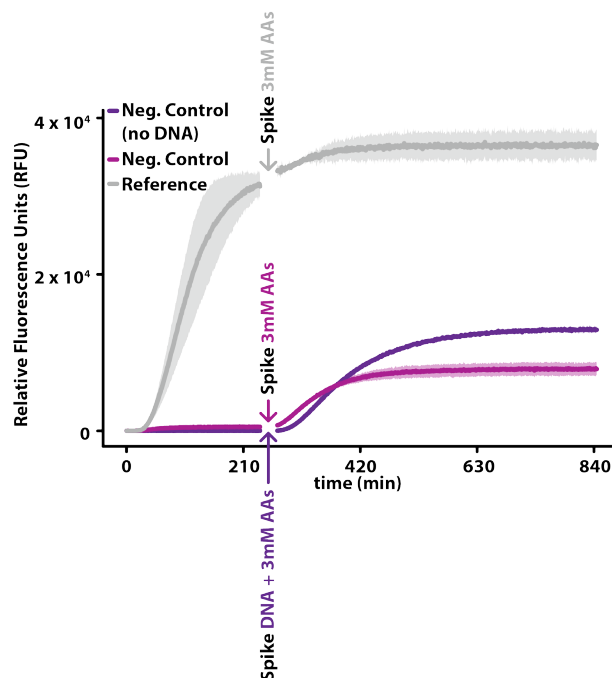


Figure D.2 – Plots of the fluorescence signal resulting from the expression of mScarlet-i in our TX-TL system (0 – 240 min) by using the reference AAs mixture from magainin II, glucagon, and somatostatin 28 complete depolymerization (grey curve), and substituting the reference AAs mixture with the negative controls (pink and purple curves); DNA(75 ng) was replaced by nuclease- free water (purple curve). Plots of the fluorescence signal resulting from the expression of mScarlet-i in our TX-TL system (270 – 845 min) by spiking the reference AAs mixture from magainin II, glucagon, and somatostatin 28 complete depolymerization and the negative control (with DNA) with a preheated (37 °C) stock water-AAs solution to get to a final expression reaction (0.39/2.5/25 v/v/v nuclease-free water:3mM AAs spike solution:expression (grey curve and pink curves), and the negative control (without DNA) with a preheated (37 °C) DNA(75 ng)-AAs solution to get to a final expression reaction (0.39/2.5/25 v/v/v DNA(75 ng):3mM AAs spike solution:expression (purple curve). (The plate reader sensitivity was exceptionally set to 80 % in this experiment).

Appendix D. Additional Data

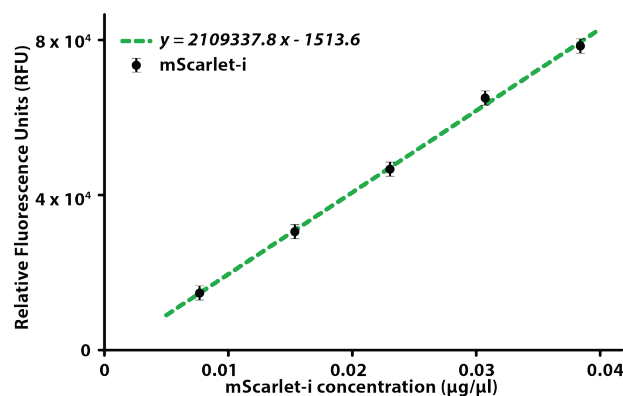


Figure D.3 – Plot of the mScarlet-i mass calibration curve in the plate reader (Appendix B.13). Error bars represent the variability of the expression by using different lots of PURE Frex™Solution II, and III, calculated as the standard deviation of the expression plateaus (RFU) in Figure D.4.

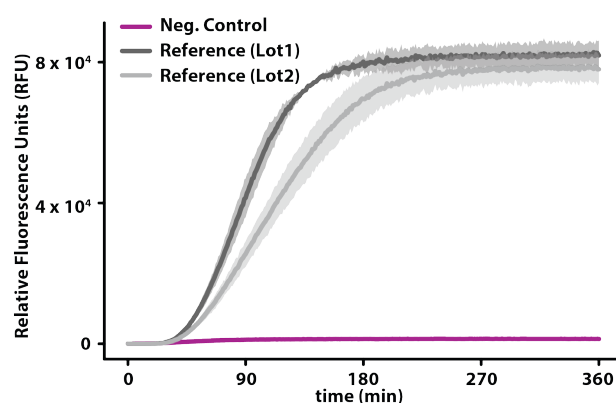


Figure D.4 – Plots of the fluorescence signal resulting from the expression of mScarlet- i in our TX-TL system by using the reference AAs mixture from magainin II, glucagon, and somatostatin 28 complete depolymerization. Two lots of PUREfrex™Solution II, and III were used in order to quantify the variability of the expression plateau (RFU), as function of the PUREfrex™Solution II, and III lots.

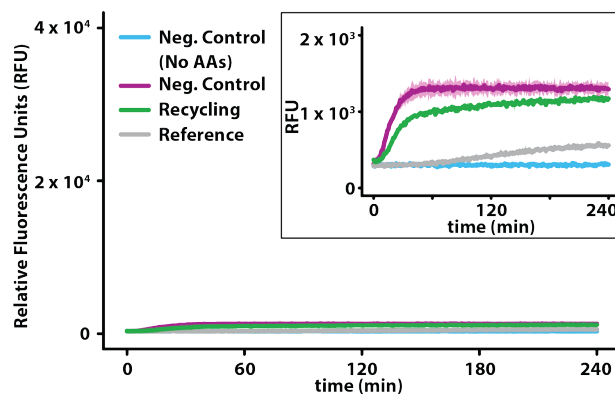


Figure D.5 – Plots of the fluorescence signal resulting from the expression in our TX- TL system of the GFP modified with the incorporation of L-norleucine, and L-canavanine. The green curve is obtained preforming NaCRe on the unnatural peptide (for recycling of L-norleucine, and L- canavanine), and supplementing the TX-TL system with the additional 18 proteinogenic AAs (Appendix B.7). The gray curve (reference control) is the result of an expression experiment with the TX-TL system supplemented with highly concentrated L-norleucine and L- canavanine, and with the additional 18 proteinogenic AAs. In the negative control expression (violet curve), the TX-TL system was supplemented with the solution resulting from the same depolymerization process used for the unnatural peptide, without adding the unnatural peptide initially. In the additional negative control (light blue curve), the negative control depolymerization solution was substituted with nuclease-free water.

Appendix D. Additional Data

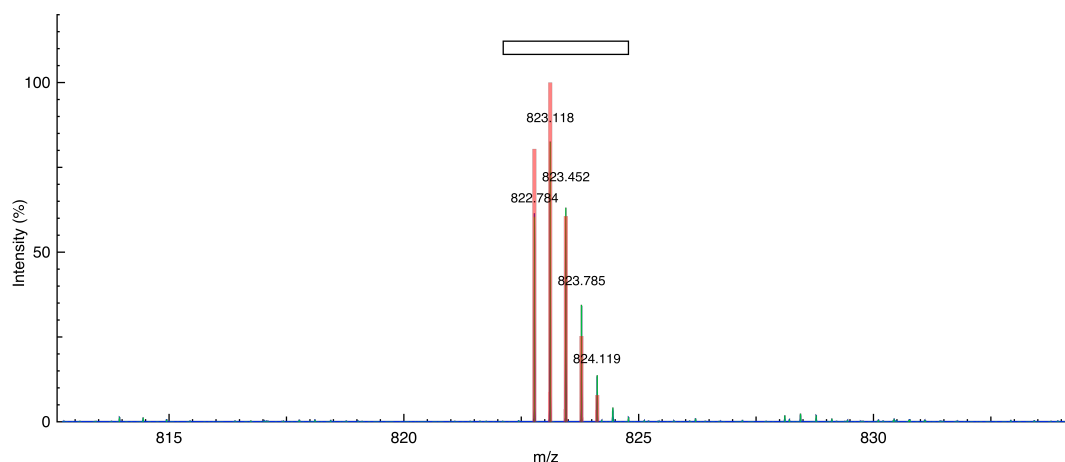


Figure D.6 – MS spectrum of magainin II: H3(3+)HGlylleGlyLysPheLeuHisSerAlaLysLysPheGlyLysAlaPheValGlyGlulleMetAsnSerOH. Theoretical observed mass (m/z) = 822.782; experimental closest peak (m/z) = 822.784. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

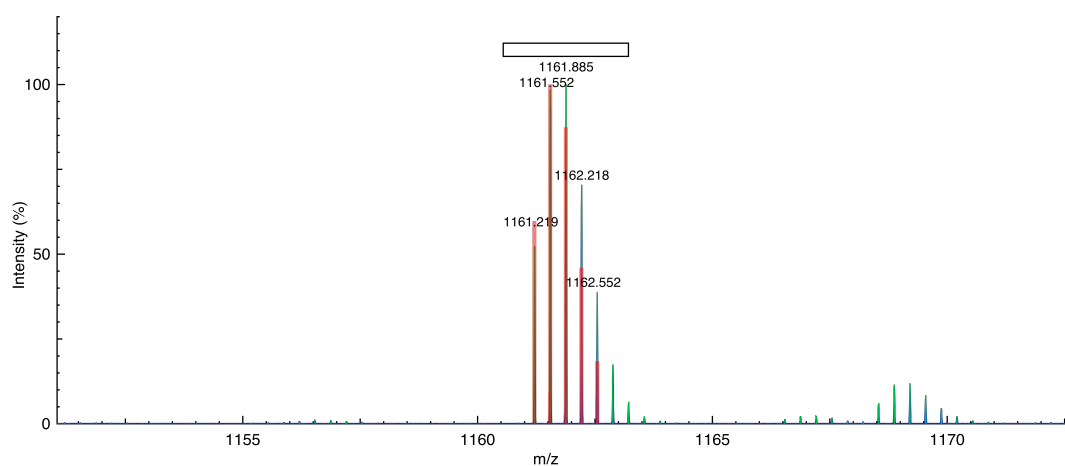


Figure D.7 – MS spectrum of glucagon: H3(3+)HHisSerGlnGlyThrPheThrSerAspTyrSerLysTyrLeuAspSerArgArgAlaGlnAspPheValGlnTrpLeuMetAsnThrOH. Theoretical observed mass (m/z) = 1161.213; experimental closest peak (m/z) = 1161.219. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

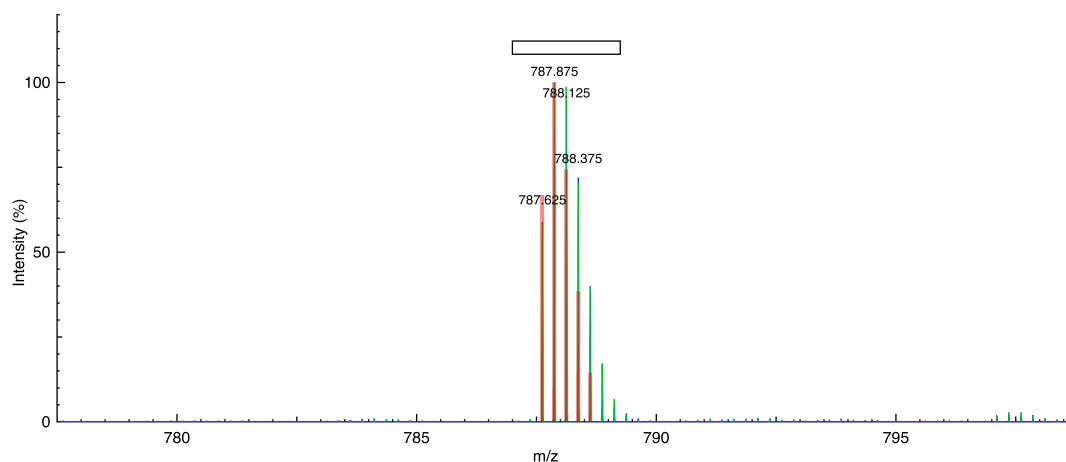


Figure D.8 – MS spectrum of somatostatin 28: H4(4+)HSerAlaAsnSerAsnProAlaMetAlaProArgGluArgLysAlaGlyCys(H-)LysAsnPhePheTrpLysThrPheThrSerCys(H-)OH. Theoretical observed mass (m/z) = 787.623; experimental closest peak (m/z) = 787.625. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

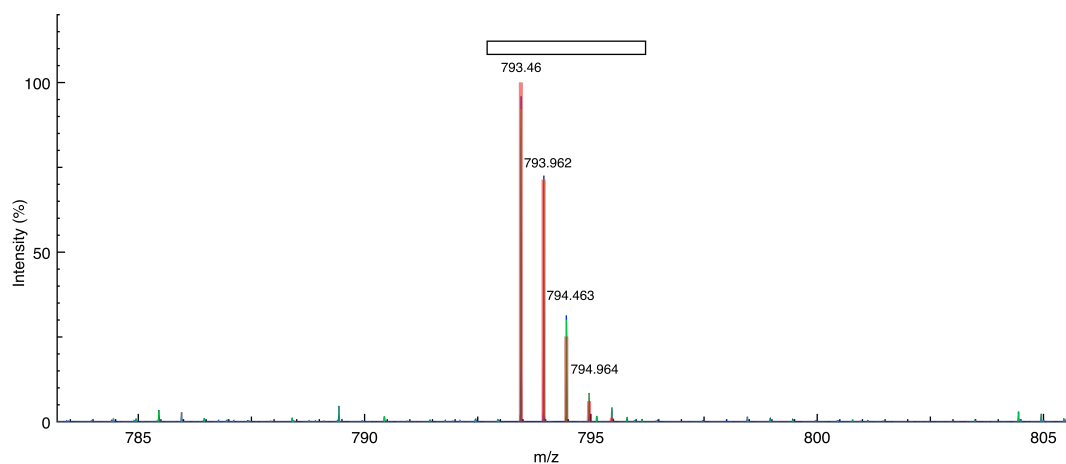


Figure D.9 – MS spectrum of the non-natural peptide: H2(2+)HMet(S-1CH2)Val(H-1F)Val(H-1F)Arg((CH2)-1O)Thr(CH2)Thr(CH2)Arg((CH2)-1O)Thr(CH2)Thr(CH2)Thr(CH2)Met(S-1CH2)SerLysOH. Theoretical observed mass (m/z) = 793.458; experimental closest peak (m/z) = 793.460. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

Appendix D. Additional Data

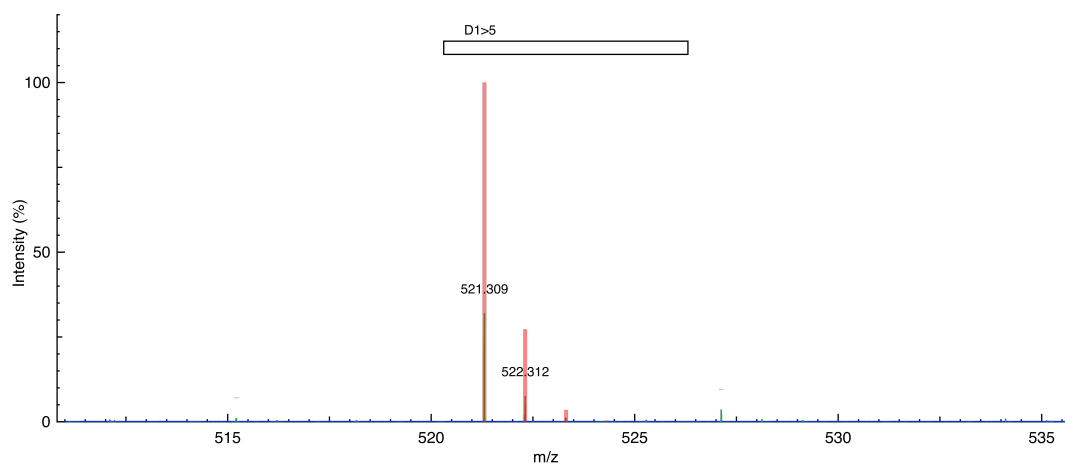


Figure D.10 – MS spectrum of magainin II cleaved fragment D1>5: H1(1+)HGlylleGlyLysPheOH. Theoretical observed mass (m/z) = 521.308; experimental closest peak (m/z) = 521.309. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

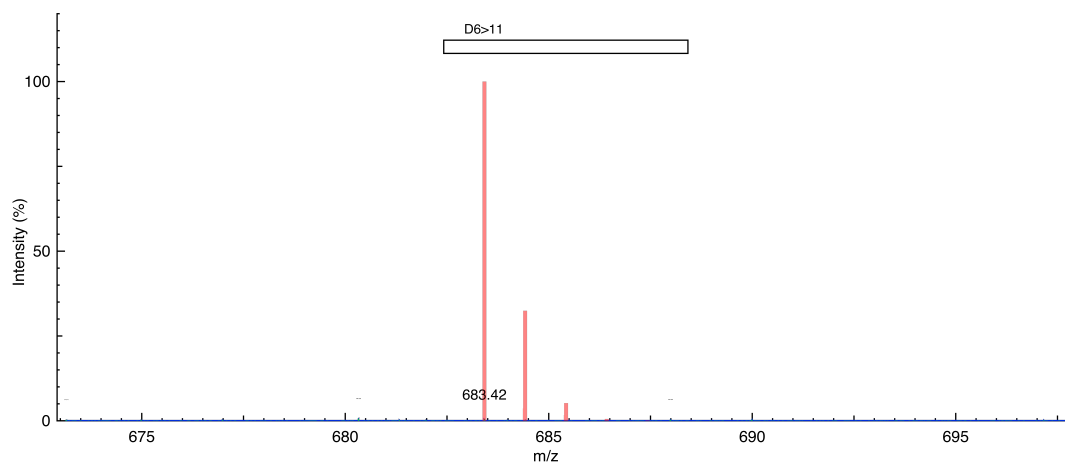


Figure D.11 – MS spectrum of magainin II cleaved fragment D6>11: H1(1+)HLeuHisSerAlaLysLysOH. Theoretical observed mass (m/z) = 683.420; experimental closest peak (m/z) = 683.420. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

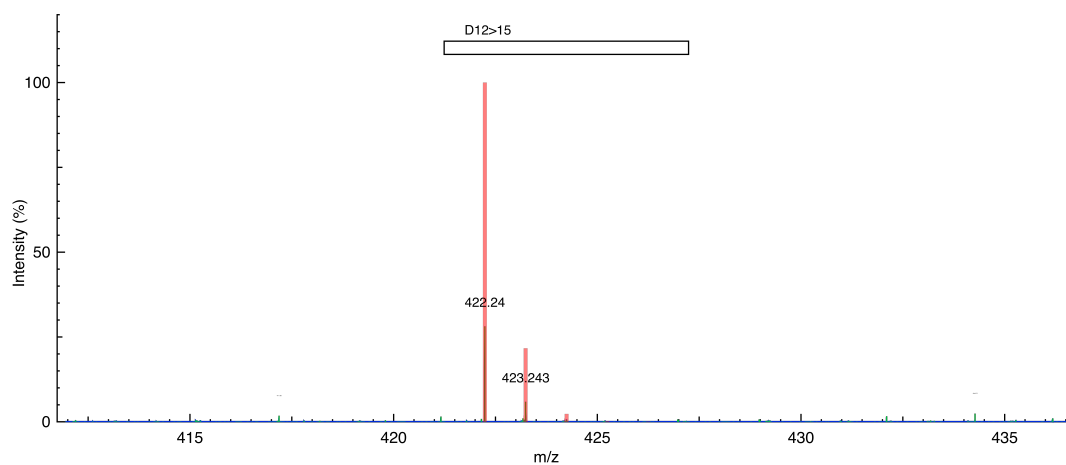


Figure D.12 – MS spectrum of magainin II cleaved fragment D12>15: H1(1+)HPheGlyLysAlaOH. Theoretical observed mass (m/z) = 422.240; experimental closest peak (m/z) = 422.240. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

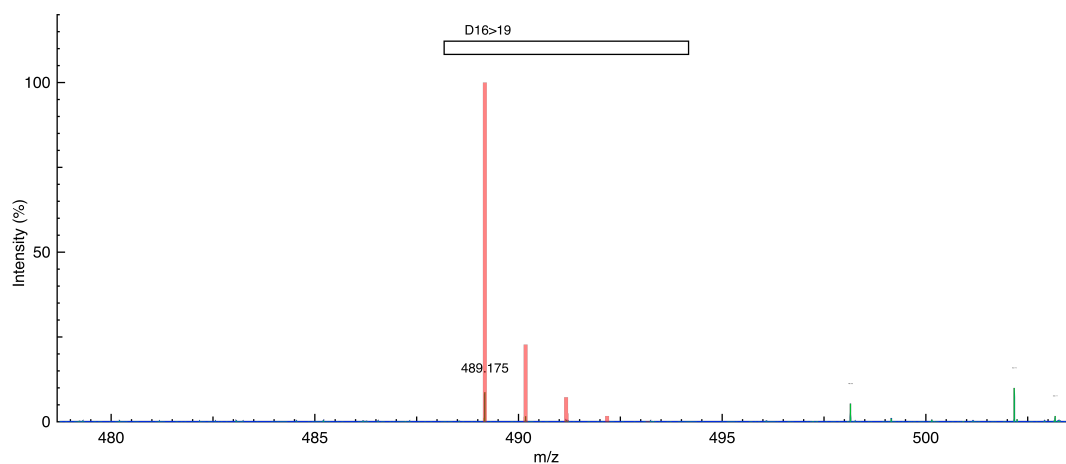


Figure D.13 – MS spectrum of magainin II cleaved fragment D16>19: K1(1+)HPheValGlyGluOH. Theoretical observed mass (m/z) = 489.175; experimental closest peak (m/z) = 489.175. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

Appendix D. Additional Data

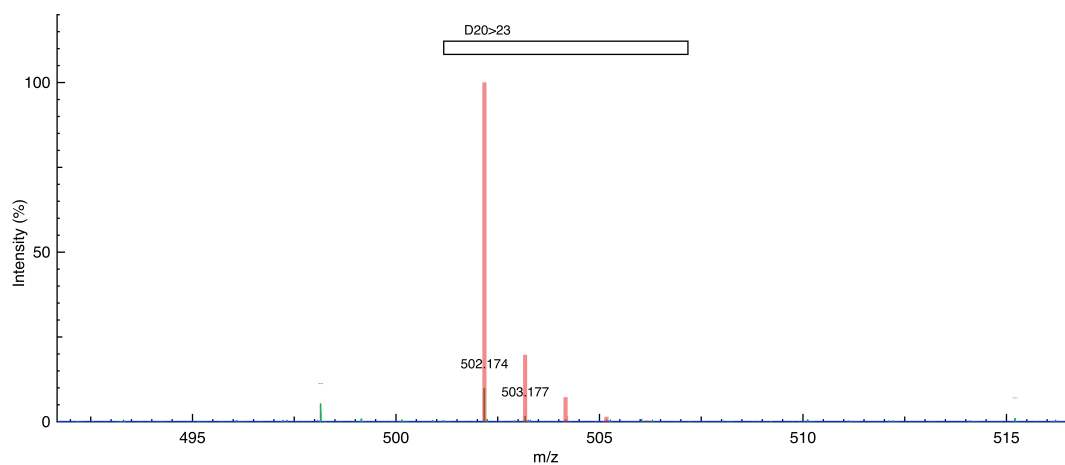


Figure D.14 – MS spectrum of magainin II cleaved fragment D20>23: K1(1+)HlleMetAsnSerOH. Theoretical observed mass (m/z) = 502.173; experimental closest peak (m/z) = 502.174. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

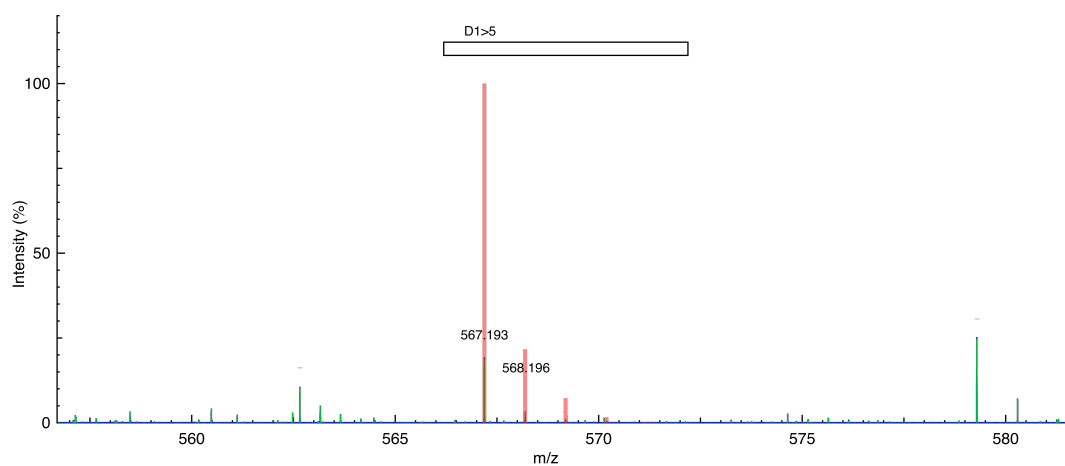


Figure D.15 – MS spectrum of glucagon cleaved fragment D1>5: K1(1+)HHisSerGlnGlyThrOH. Theoretical observed mass (m/z) = 567.192; experimental closest peak (m/z) = 567.193. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

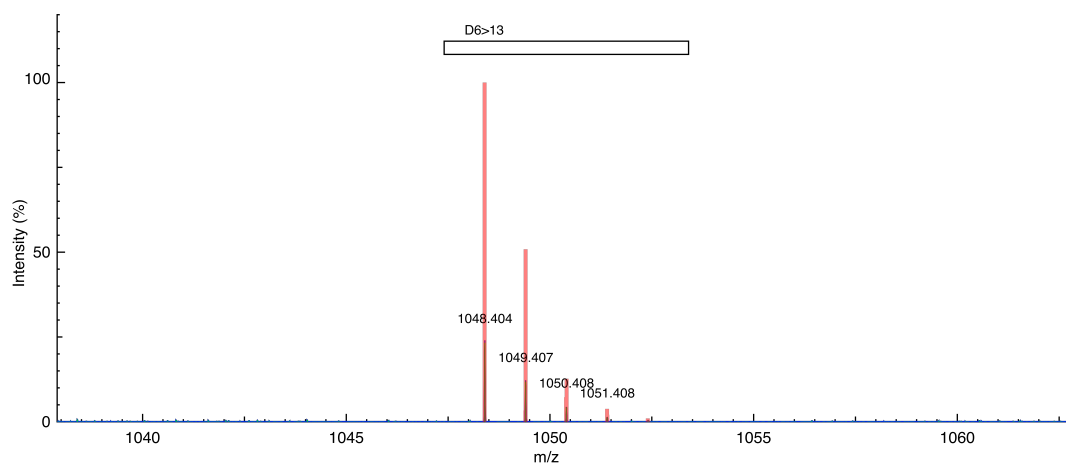


Figure D.16 – MS spectrum of glucagon cleaved fragment D6>13: K1(1+)HPheThrSerAsp TyrSerLysTyrOH. Theoretical observed mass (m/z) = 1048.402; experimental closest peak (m/z) = 1048.404. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

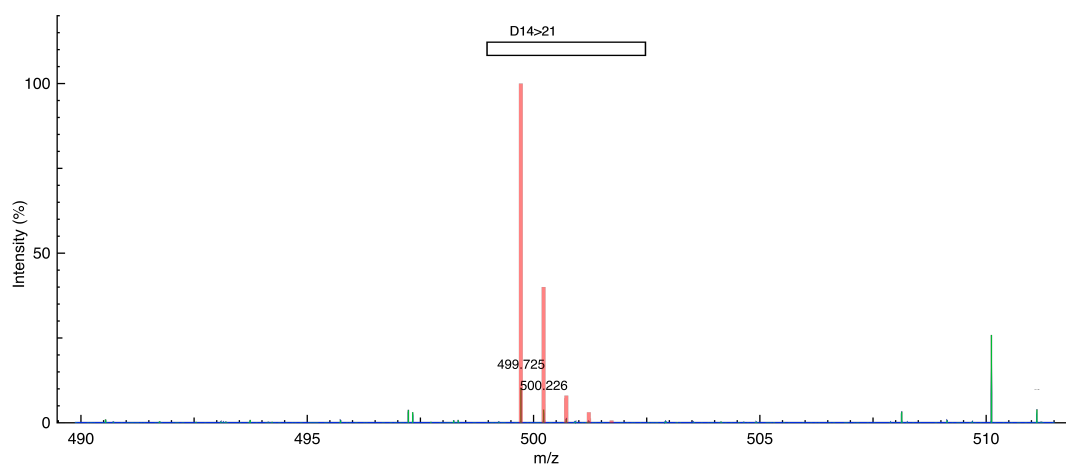


Figure D.17 – MS spectrum of glucagon cleaved fragment D14>21: K1(1+)H1(1+)HLeuAspSer ArgArgAlaGlnAspOH. Theoretical observed mass (m/z) = 499.724; experimental closest peak (m/z) = 499.725. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

Appendix D. Additional Data

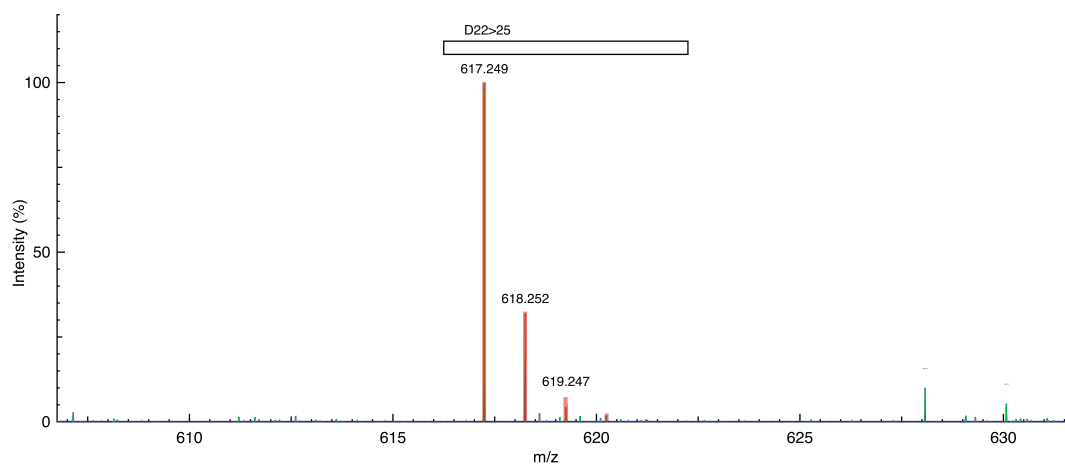


Figure D.18 – MS spectrum of glucagon cleaved fragment D22>25: K1(1+)HPheValGlnTrpOH. Theoretical observed mass (m/z) = 617.248; experimental closest peak (m/z) = 617.249. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

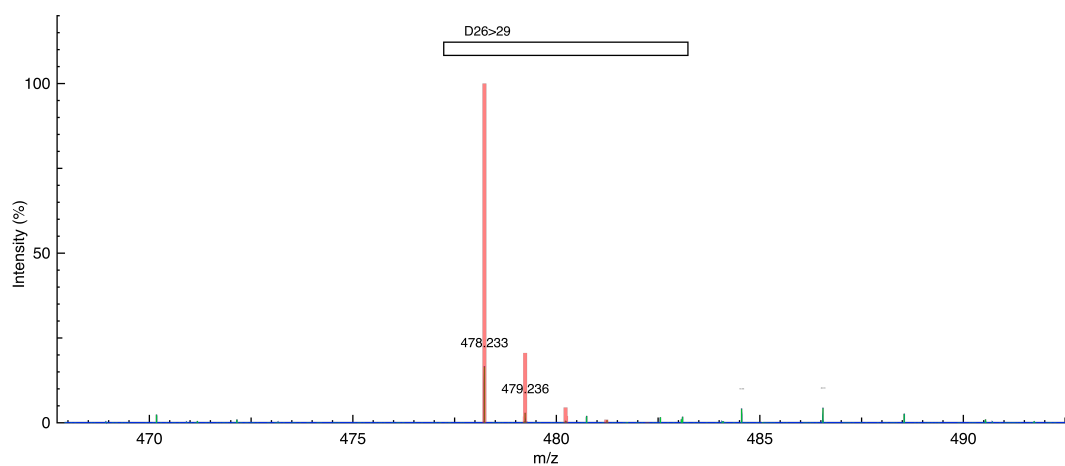


Figure D.19 – MS spectrum of glucagon cleaved fragment D26>29: H1(1+)HLeuMetAsnThrOH. Theoretical observed mass (m/z) = 478.233; experimental closest peak (m/z) = 478.233. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

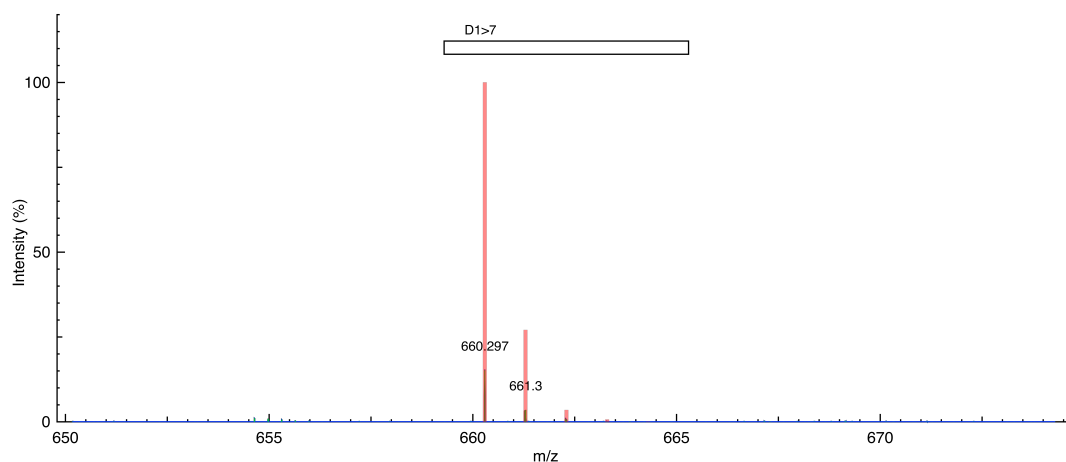


Figure D.20 – MS spectrum of somatostatin 28 cleaved fragment D1>7: H1(1+)HSerAla AsnSerAsnProAlaOH. Theoretical observed mass (m/z) = 660.295; experimental closest peak (m/z) = 660.297. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

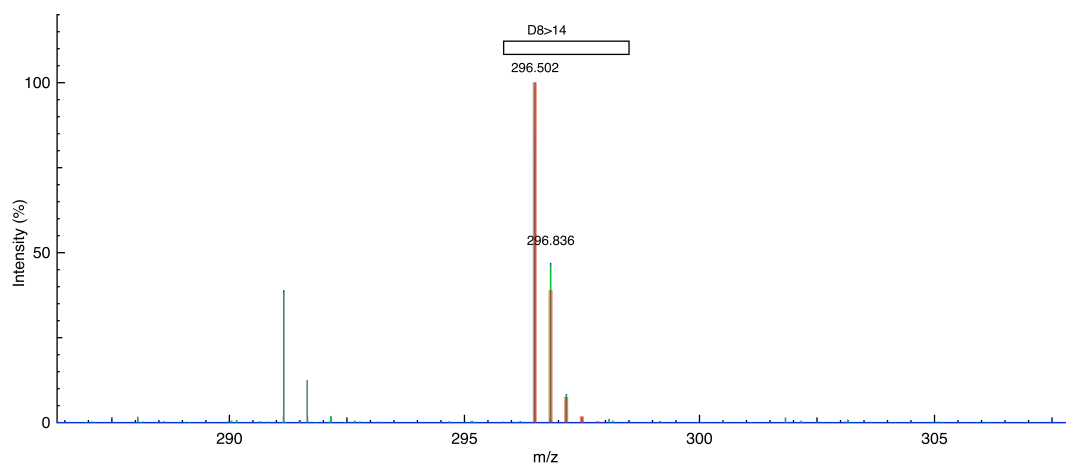


Figure D.21 – MS spectrum of somatostatin 28 cleaved fragment D8>14: H3(3+)HMet AlaProArgGluArgLysOH. Theoretical observed mass (m/z) = 296.501; experimental closest peak (m/z) = 296.502. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

Appendix D. Additional Data

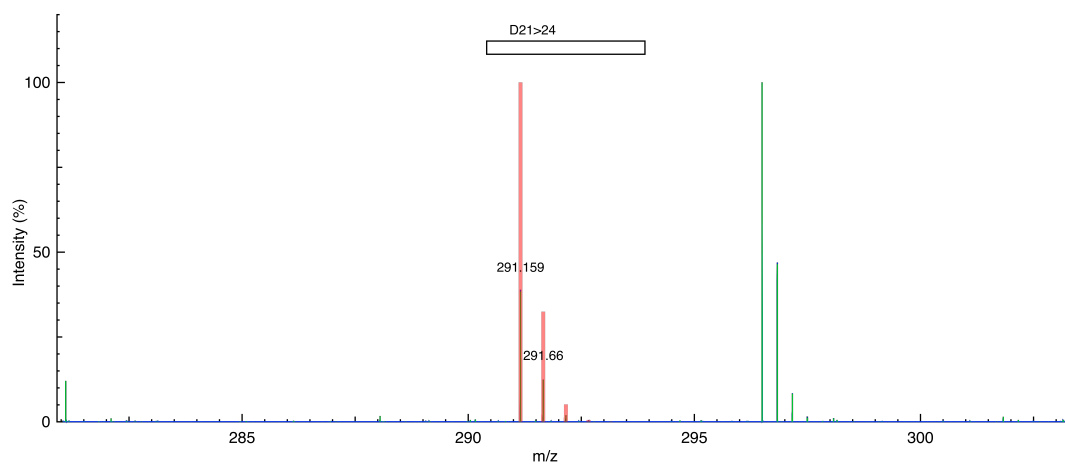


Figure D.22 – MS spectrum of somatostatin 28 cleaved fragment D21>24: H2(2+)HPheTrp LysThrOH. Theoretical observed mass (m/z) = 291.158; experimental closest peak (m/z) = 291.159. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

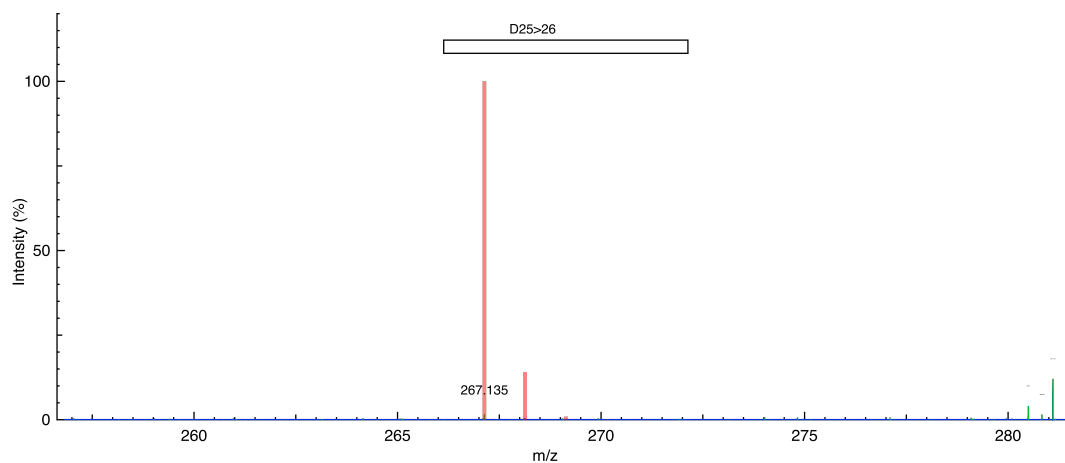


Figure D.23 – MS spectrum of somatostatin 28 cleaved fragment D25>26: H1(1+)HPheThrOH. Theoretical observed mass (m/z) = 267.134; experimental closest peak (m/z) = 267.135. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

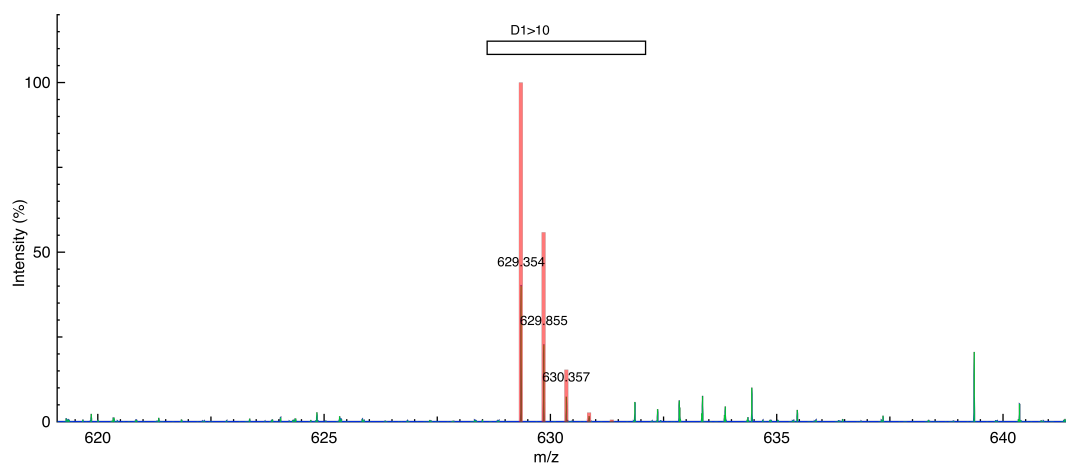


Figure D.24 – MS spectrum of the non-natural peptide cleaved fragment D1>10: H₂(2+)HMet(S-1CH₂)Val(H-1F)Val(H-1F)Arg((CH₂)-1O)Thr(CH₂)Thr(CH₂)Arg((CH₂)-1O)Thr(CH₂)Thr(CH₂)Thr(CH₂)OH. Theoretical observed mass (m/z) = 629.352; experimental closest peak (m/z) = 629.354. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

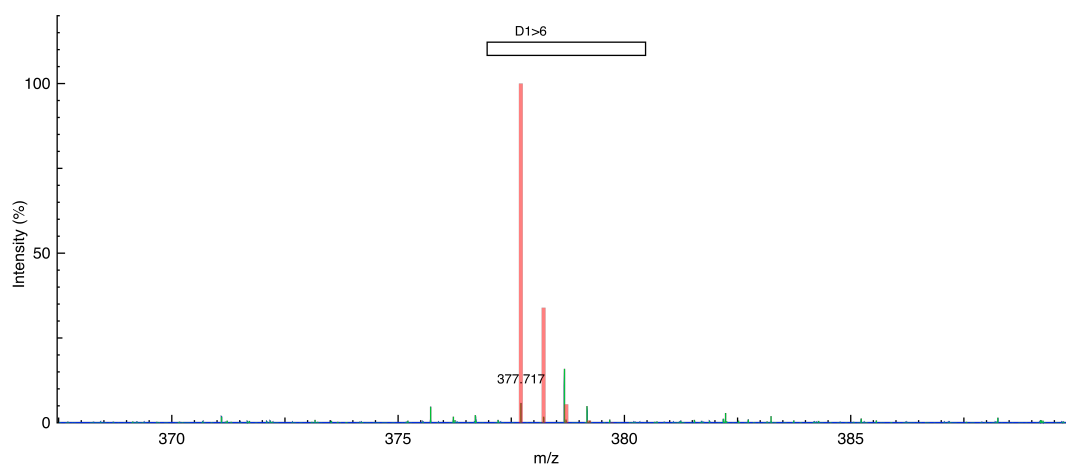


Figure D.25 – MS spectrum of the non-natural peptide cleaved fragment D1>6: H₂(2+)HMet(S-1CH₂)Val(H-1F)Val(H-1F)Arg((CH₂)-1O)Thr(CH₂)Thr(CH₂)OH. Theoretical observed mass (m/z) = 377.717; experimental closest peak (m/z) = 377.717. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

Appendix D. Additional Data

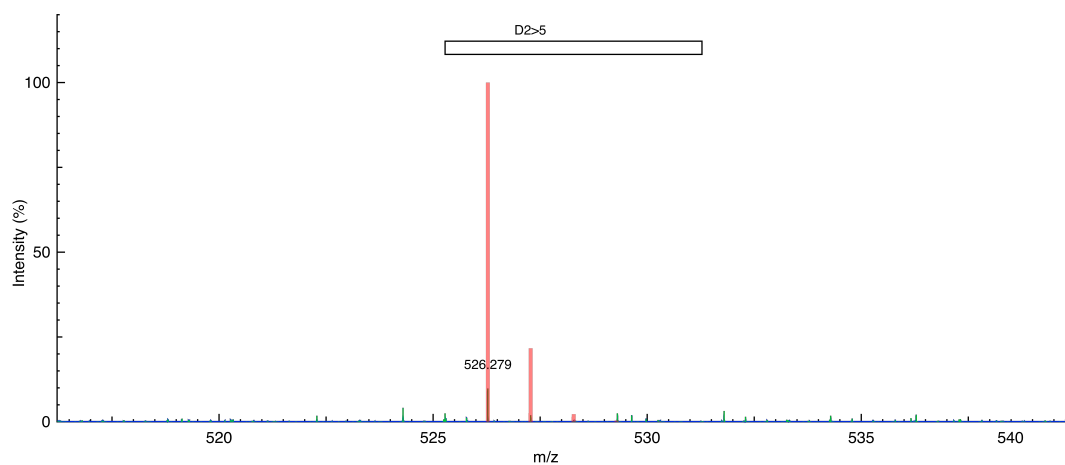


Figure D.26 – MS spectrum of the non-natural peptide cleaved fragment D2>5: H1(1+)HVal(H-1F)Val(H-1F)Arg((CH₂)-1O)Thr(CH₂)OH. Theoretical observed mass (m/z) = 526.280; experimental closest peak (m/z) = 526.279. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

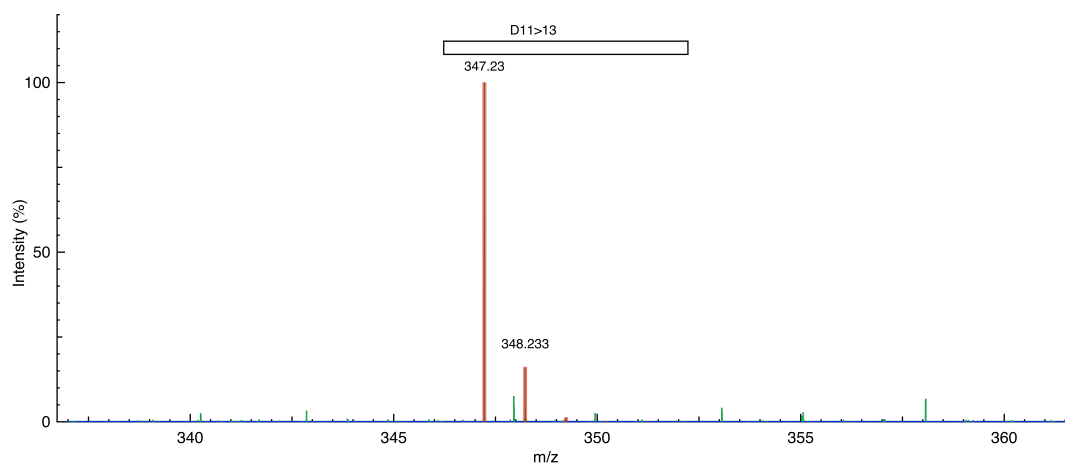


Figure D.27 – MS spectrum of the non-natural peptide cleaved fragment D11>13: H1(1+)HMet(S-1CH₂)SerLysOH. Theoretical observed mass (m/z) = 347.229; experimental closest peak (m/z) = 347.230. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

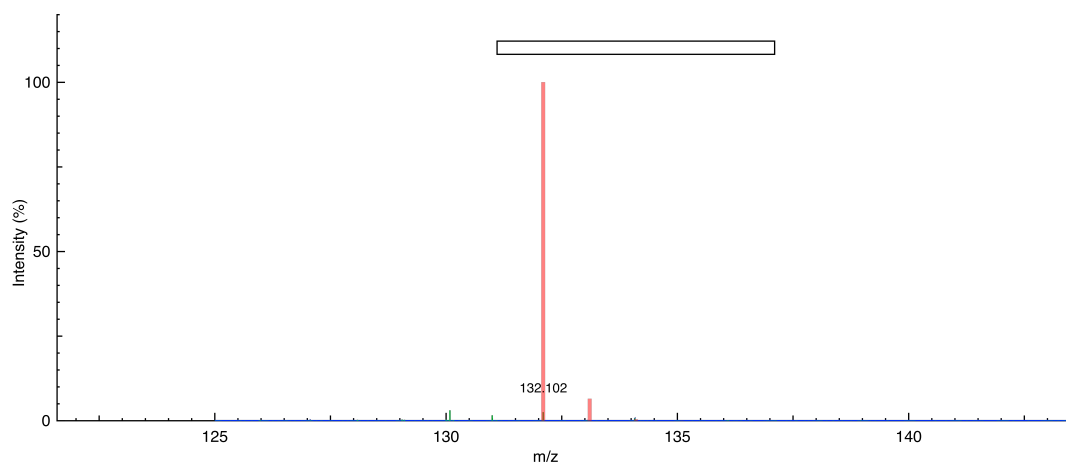


Figure D.28 – MS spectrum of L-norleucine: H1(1+)HMet(S-1CH₂)OH. Theoretical observed mass (m/z) = 132.102; experimental closest peak (m/z) = 132.102. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

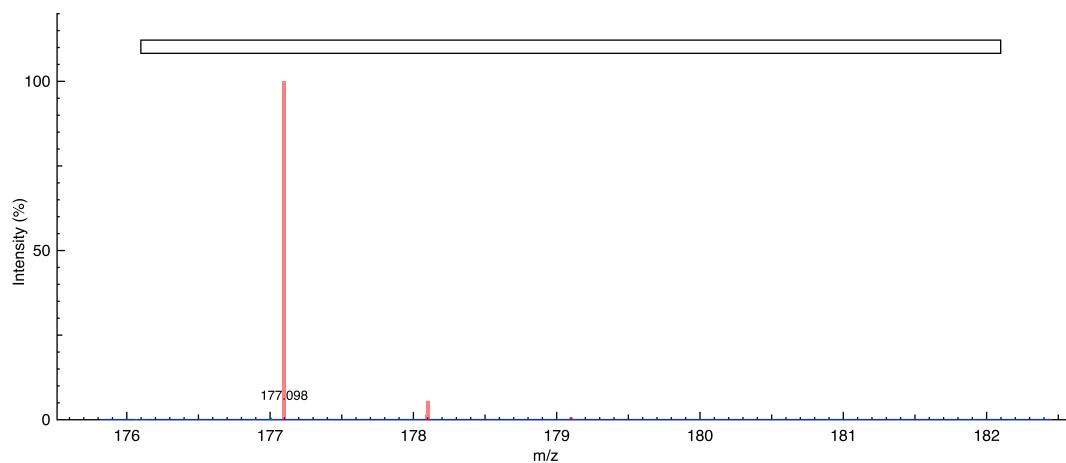


Figure D.29 – MS spectrum of L-canavanine: H1(1+)HArg((CH₂)-1O)OH. Theoretical observed mass (m/z) = 177.098; experimental closest peak (m/z) = 177.098. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

Appendix D. Additional Data

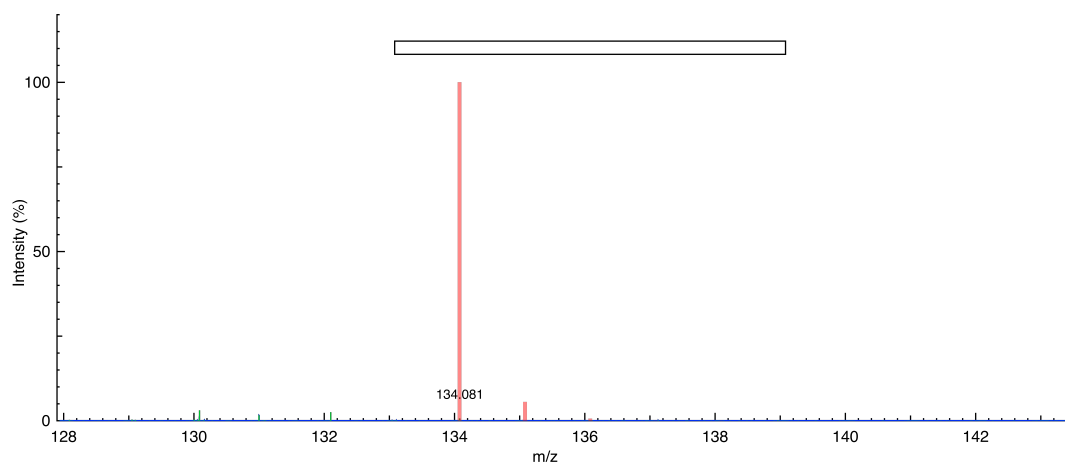


Figure D.30 – MS spectrum of DL-3-hydroxynorvaline: H1(1+)HThr(CH₂)OH. Theoretical observed mass (m/z) = 134.081; experimental closest peak (m/z) = 134.081. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

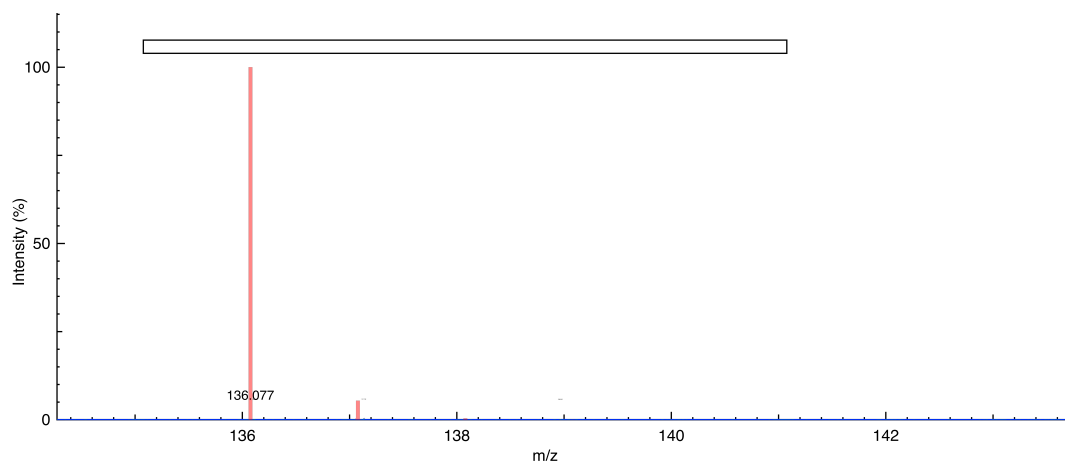


Figure D.31 – MS spectrum of 3-fluoro-DL-valine: H1(1+)HVal(H-1F)OH. Theoretical observed mass (m/z) = 136.077; experimental closest peak (m/z) = 136.077. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

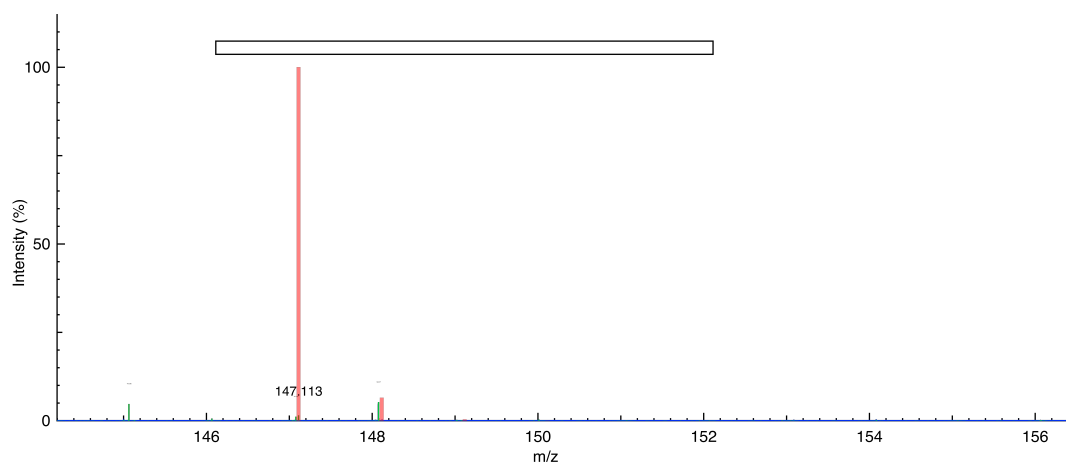


Figure D.32 – MS spectrum of L-lysine: H1(1+)HLysOH. Theoretical observed mass (m/z) = 147.113; experimental closest peak (m/z) = 147.113. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

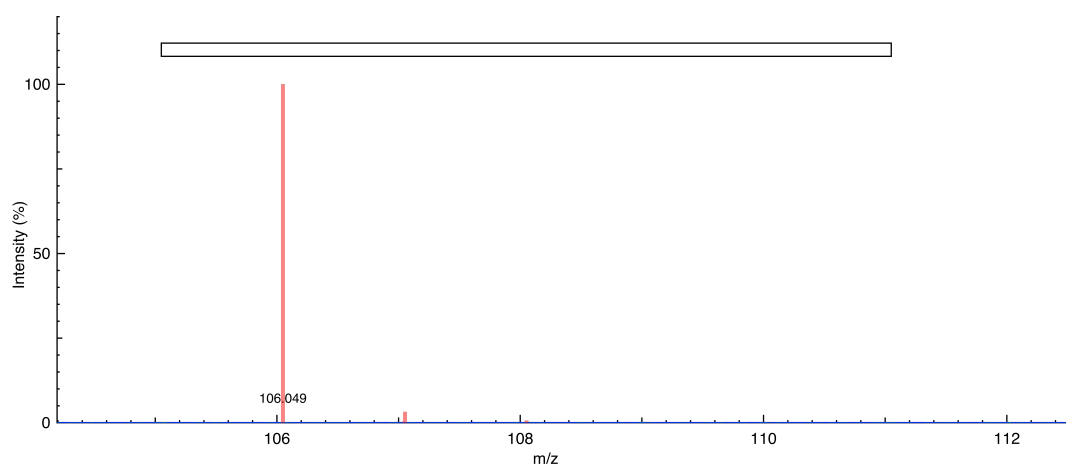


Figure D.33 – MS spectrum of L-serine: H1(1+)HSerOH. Theoretical observed mass (m/z) = 106.050; experimental closest peak (m/z) = 106.049. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

Appendix D. Additional Data

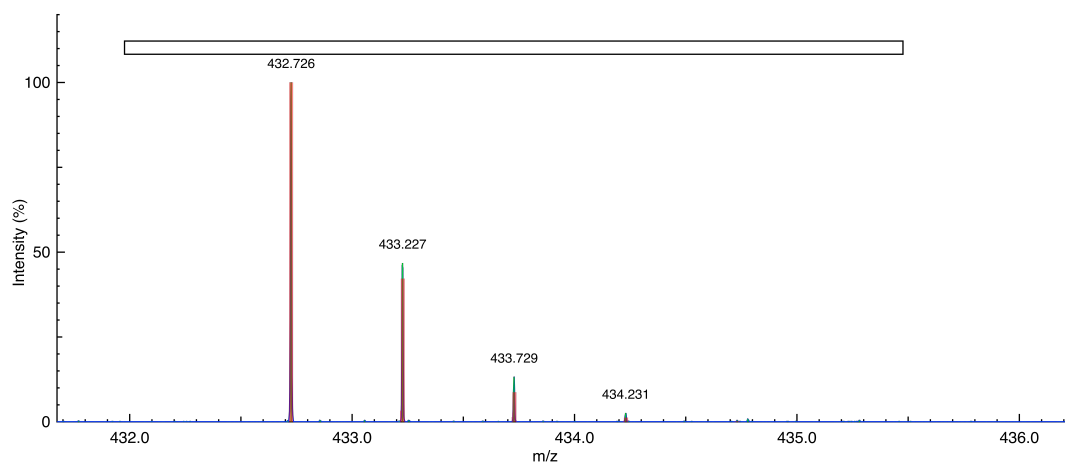


Figure D.34 – MS spectrum of the (in-gel) cleaved peptide from modified GFP (reference control), incorporating L-norleucine = HMet(S-1CH₂)OH: H₂(2+)HLysSerAlaMet(S-1CH₂)ProGluGlyTyrOH. Theoretical observed mass (m/z) = 432.727; experimental closest peak (m/z) = 432.726. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

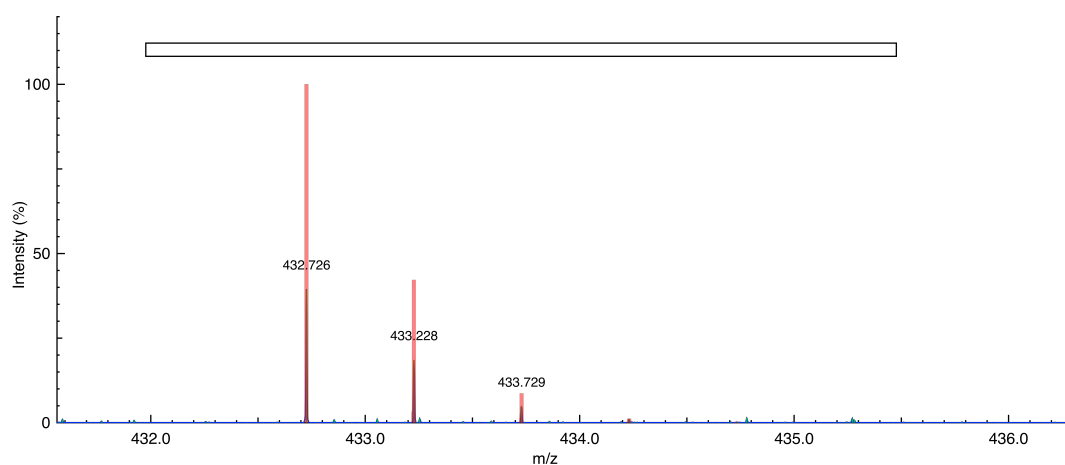


Figure D.35 – MS spectrum of the (in-gel) cleaved peptide from modified GFP (sample), incorporating the recycled L-norleucine = HMet(S-1CH₂)OH: H₂(2+)HLysSerAlaMet(S-1CH₂)ProGluGlyTyrOH. Theoretical observed mass (m/z) = 432.727; experimental closest peak (m/z) = 432.727. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

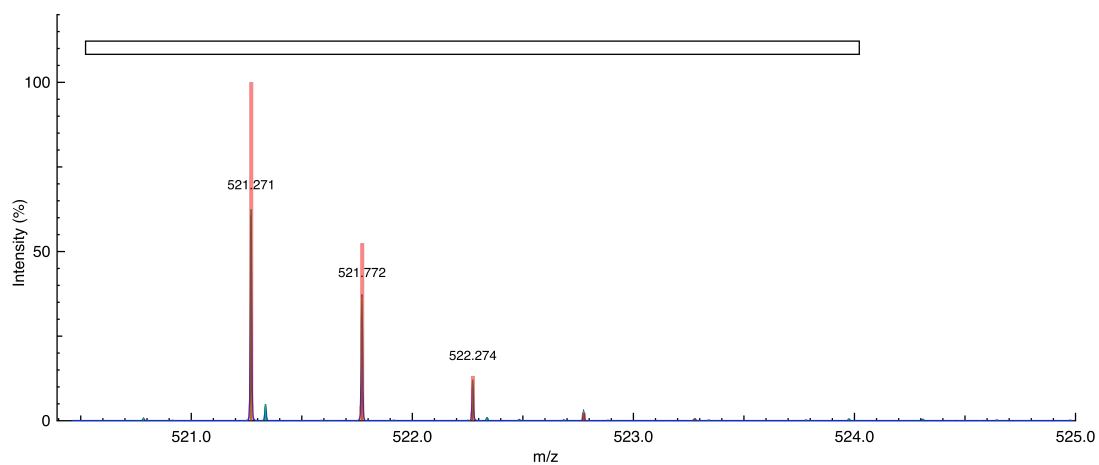


Figure D.36 – MS spectrum of the (in-gel) cleaved peptide from modified GFP (reference control), incorporating L-canavanine = HArg((CH₂)-1O)OH: H₂(2+)HValGlnGluArg((CH₂)-1O)ThrIlePhePheOH. Theoretical observed mass (m/z) = 521.272; experimental closest peak (m/z) = 521.271. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

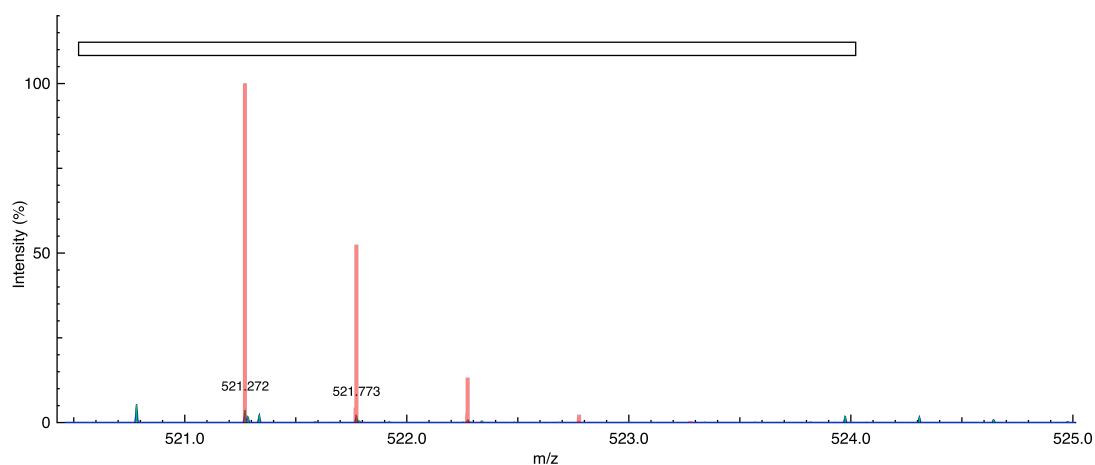


Figure D.37 – MS spectrum of the (in-gel) cleaved peptide from modified GFP (sample), incorporating the recycled L-canavanine = HArg((CH₂)-1O)OH: H₂(2+)HValGlnGluArg((CH₂)-1O)ThrIlePhePheOH. Theoretical observed mass (m/z) = 521.272; experimental closest peak (m/z) = 521.272. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

Appendix D. Additional Data

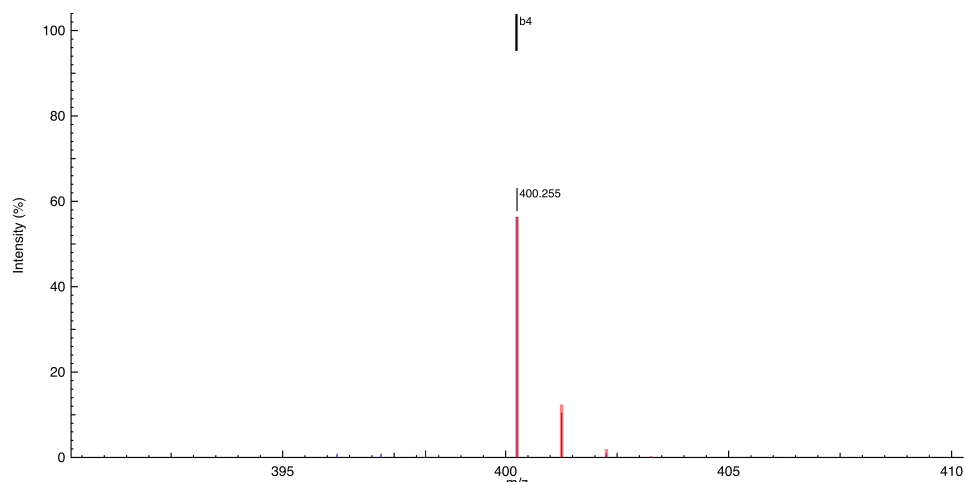


Figure D.38 – MS spectrum of the b4 fragment of the isolated H2(2+)HLysSerAlaMet(S-1CH2)ProGluGlyTyrOH peptide from modified GFP (reference control), incorporating L-norleucine = HMet(S-1CH2)OH: HLysSerAlaMet(S-1CH2)(1+). Theoretical observed mass (m/z) = 400.255; experimental closest peak (m/z) = 400.255. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

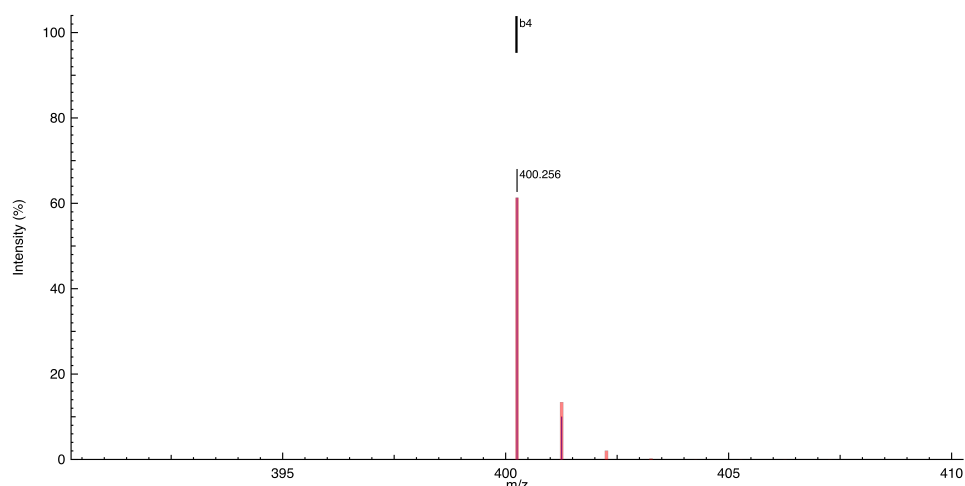


Figure D.39 – MS spectrum of the b4 fragment of the isolated H2(2+)HLysSerAlaMet(S-1CH2)ProGluGlyTyrOH peptide from modified GFP (sample), incorporating the recycled L-norleucine = HMet(S-1CH2)OH: HLysSerAlaMet(S-1CH2)(1+). Theoretical observed mass (m/z) = 400.255; experimental closest peak (m/z) = 400.256. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

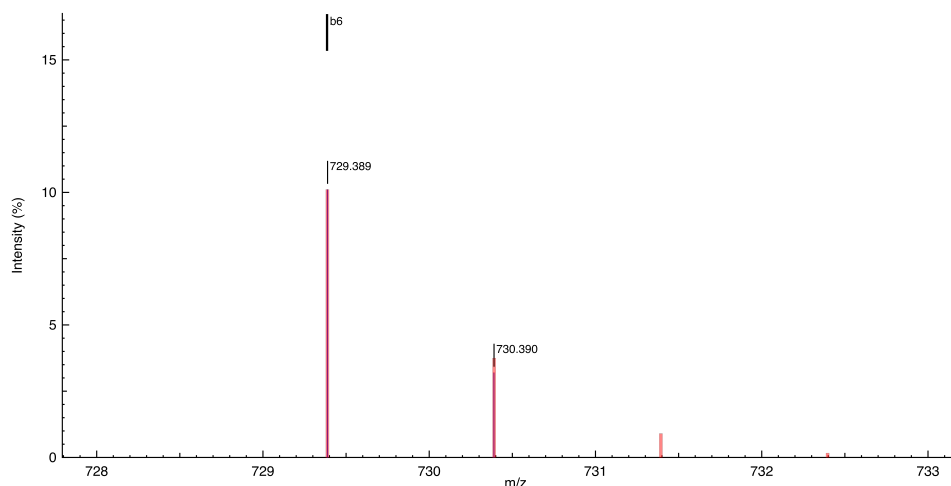


Figure D.40 – MS spectrum of the b6 fragment of the isolated H2(2+)HValGlnGluArg((CH2)-1O)ThrIlePhePheOH peptide from modified GFP (reference control), incorporating L-canavanine = HArg((CH2)-1O)OH: HValGlnGluArg((CH2)-1O)ThrIle(1+). Theoretical observed mass (m/z) = 729.389; experimental closest peak (m/z) = 729.389. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

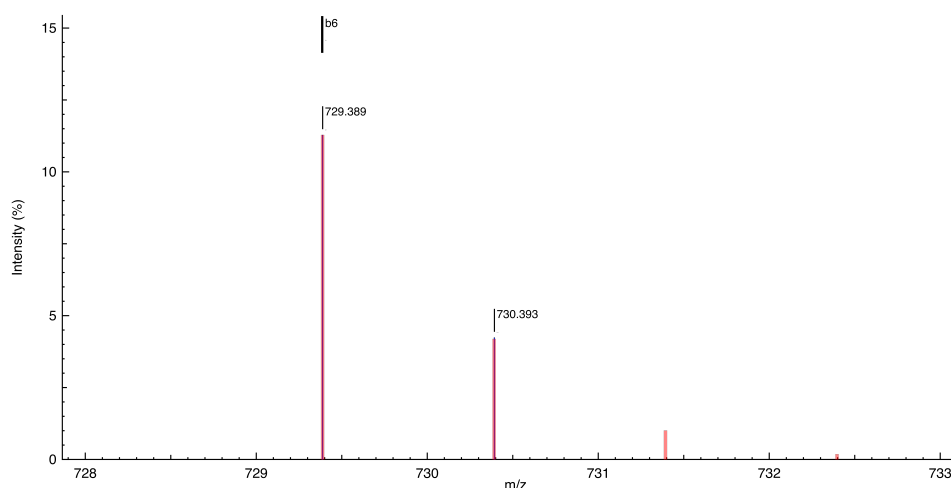
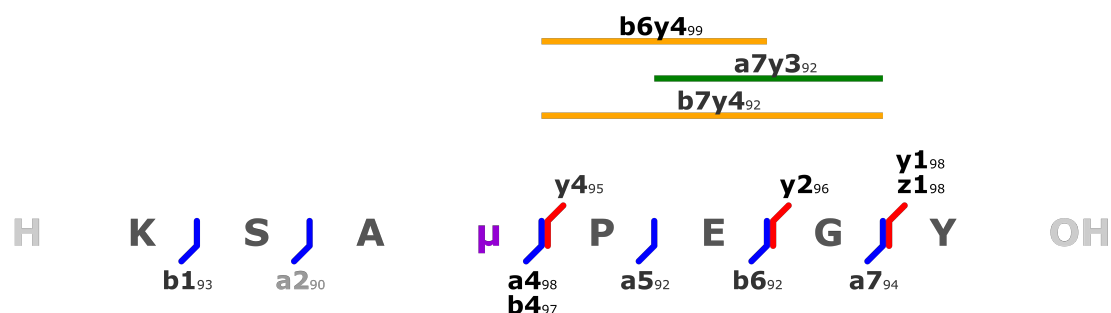
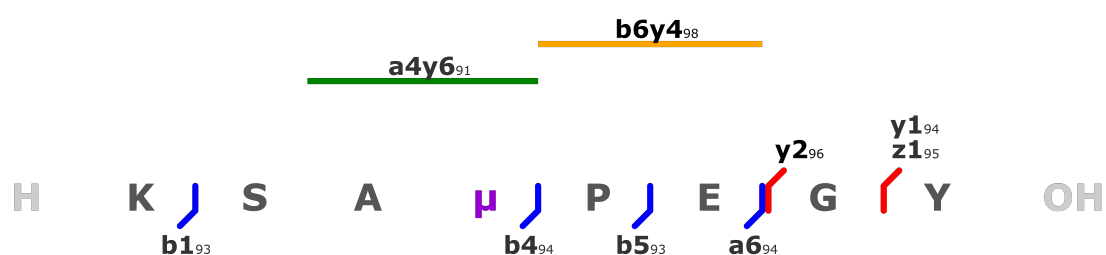


Figure D.41 – MS spectrum of the b6 fragment of the isolated H2(2+)HValGlnGluArg((CH2)-1O)ThrIlePhePheOH peptide from modified GFP (sample), incorporating the recycled L-canavanine = HArg((CH2)-1O)OH: HValGlnGluArg((CH2)-1O)ThrIle(1+). Theoretical observed mass (m/z) = 729.389; experimental closest peak (m/z) = 729.389. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).



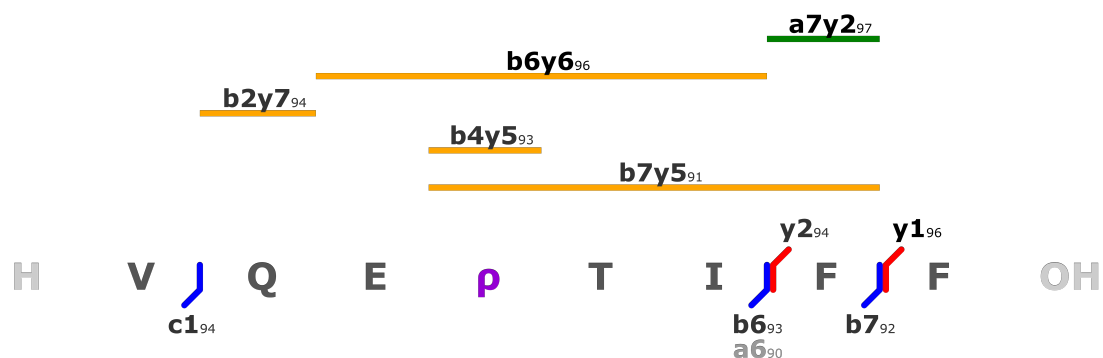
μ = Met(S-1CH2) = L-Norleucine

Figure D.42 – Fragmentation pattern of the isolated H2(2+)HLysSerAlaMet(S-1CH2)ProGluGly TyrOH peptide from modified GFP (reference control), incorporating L-norleucine = HMet(S-1CH2)OH.



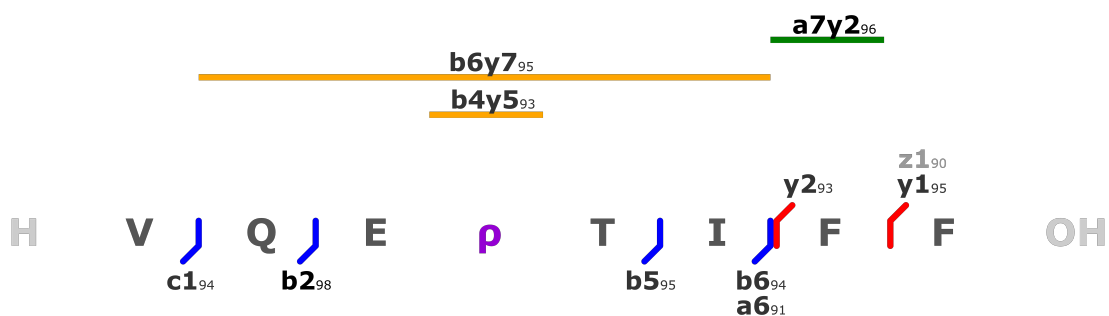
μ = Met(S-1CH2)= L-Norleucine

Figure D.43 – Fragmentation pattern of the isolated H2(2+)HLysSerAlaMet(S-1CH2)ProGluGly TyrOH peptide from modified GFP (sample), incorporating the recycled L-norleucine = HMet(S-1CH2)OH.



p = Arg((CH₂)-10)= L-Canavanine

Figure D.44 – Fragmentation pattern of the isolated H₂(2+)HValGlnGluArg((CH₂)-10)Thr IlePhePheOH peptide from modified GFP (reference control), incorporating L-canavanine = HArg((CH₂)-10)OH.



p = Arg((CH₂)-10)= L-Canavanine

Figure D.45 – Fragmentation pattern of the isolated H₂(2+)HValGlnGluArg((CH₂)-10)Thr IlePhePheOH peptide from modified GFP (sample), incorporating the recycled L-canavanine = HArg((CH₂)-10)OH.

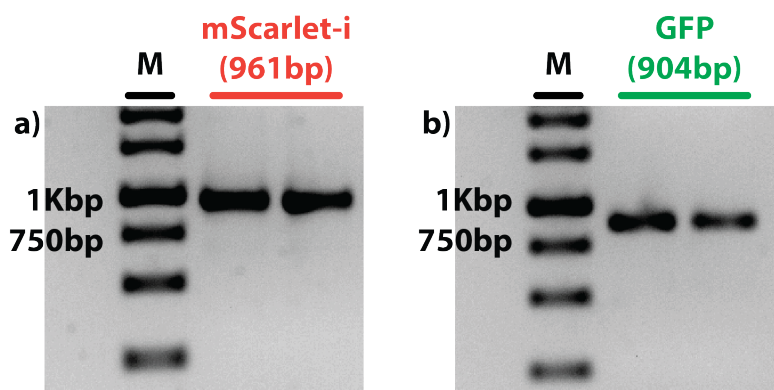


Figure D.46 – Electrophoresis gel (Agarose) of the PCR amplified mScarlet-i (a), and GFP templates (b).

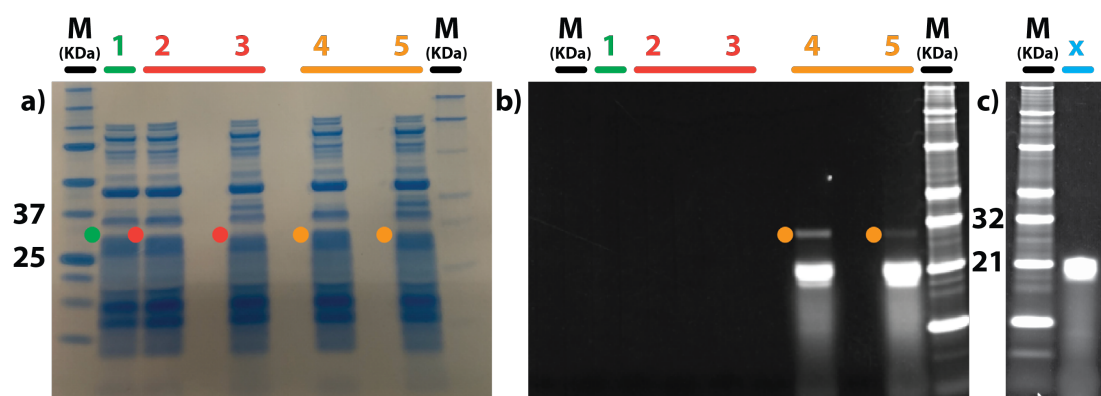


Figure D.47 – Coomassie stained (a), and fluorescence (b) images of the same SDS- PAGE protein gel used to prepare the expressed natural GFP (sample) (1), the modified GFP incorporating L-norleucine and L-canavanine (reference control) (2), and the modified GFP incorporating the recycled L-norleucine and L-canavanine (sample) (3) for the proteomic characterization. Modified GFP incorporating L-norleucine and L-canavanine (reference control), with Lys-BODIPY-FL inclusions (4), and modified GFP incorporating the recycled L-norleucine and L-canavanine (sample), with Lys-BODIPY-FL inclusions (5). Fluorescence image (c) of an additional SDS-PAGE protein gel, run in the same experimental conditions as the gel shown in (a-b), with a FluoroTect™GreenLys tRNA in nuclease-free water solution (x). This is a control experiment to show that the fluorescent bands visible in (4), and (5) are not protein impurities but they are exclusively due to tRNA-Lys-BODIPY-FL.

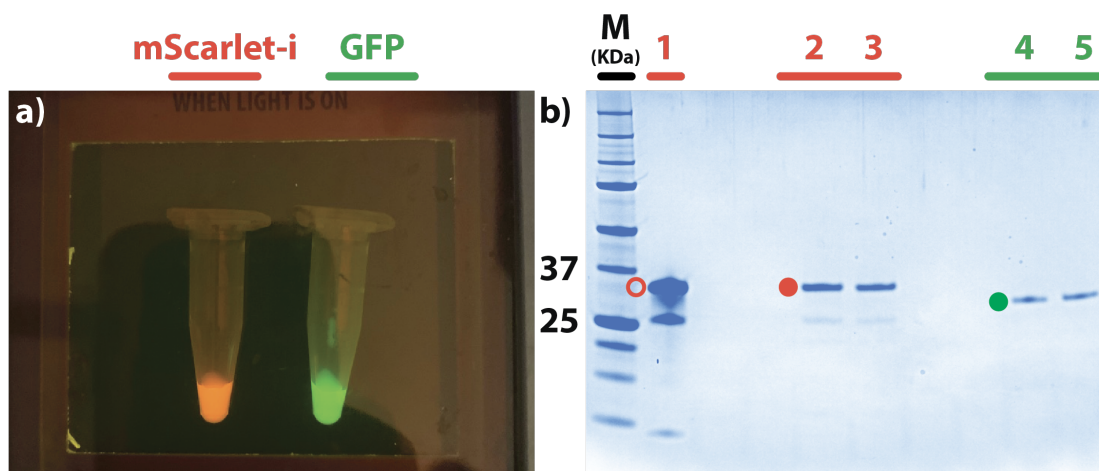


Figure D.48 – (a) Photograph of mScarlet-i, and GFP solutions obtained by NaCRe from a mixture composed of glucagon, β -lactoglobulin A, and silk fibroin, purified by 6xHis tag at C-terminus, and inspected by using InvitrogenTME-GelTMSafe ImagerTM (emission max of the blue LED = 470 nm). The image has been taken in the lab by using an iPhone Xs. (b) Image of the Coomassie stained SDS-PAGE protein gel of the purified mScarlet-i (2-3), and GFP (4-5) shown in (a). For mScarlet-i, the calibrant expressed in *E. coli* cells has been added to the gel (1); red-fluorescent proteins are known to produce cleaved fragments when treated at high temperature with denaturants¹¹⁶. The GFP protein bands (4-5) are less intense with respect to mScarlet-i ones (2-3) since the GFP sample was buffer exchanged to prepare it for the second cycle of NaCRe, as described in Appendix B.8. The gel has been used to prepare the samples prior to proteomic characterization.

References

- [1] Hermann Staudinger. Über polymerisation. *Berichte der deutschen chemischen Gesellschaft (A and B Series)*, 53(6):1073–1085, 1920.
- [2] Holger Frey and Tobias Johann. Celebrating 100 years of “polymer science”: Hermann staudinger’s 1920 manifesto. *Polymer Chemistry*, 11(1):8–14, 2020.
- [3] Haritz Sardon and Andrew P. Dove. Plastics recycling with a difference. *Science*, 360(6387):380–381, 2018.
- [4] Danielle E. Fagnani, Jessica L. Tami, Graeme Copley, Mackenzie N. Clemons, Yutan D. Y. L. Getzler, and Anne J. McNeil. 100th anniversary of macromolecular science viewpoint: Redefining sustainable polymers. *ACS Macro Letters*, 10:41–53, 2020.
- [5] Nikolai Maximenko, Jan Hafner, and Peter Niiler. Pathways of marine debris derived from trajectories of lagrangian drifters. *Marine pollution bulletin*, 65(1-3):51–62, 2012.
- [6] Ian A. Kane, Michael A. Clare, Elda Miramontes, Roy Wogelius, James J. Rothwell, Pierre Garreau, and Florian Pohl. Seafloor microplastic hotspots controlled by deep-sea circulation. *Science*, 368(6495):1140–1145, 2020.
- [7] Marcus Eriksen, Laurent C. M. Lebreton, Henry S. Carson, Martin Thiel, Charles J. Moore, Jose C. Borerro, Francois Galgani, Peter G. Ryan, and Julia Reisser. Plastic pollution in the world’s oceans: more than 5 trillion plastic pieces weighing over 250,000 tons afloat at sea. *PloS one*, 9(12):e111913, 2014.
- [8] Kara Lavender Law and Richard C. Thompson. Microplastics in the seas. *Science*, 345(6193):144–145, 2014.

References

- [9] Chelsea M. Rochman. Microplastics research—from sink to source. *Science*, 360(6384): 28–29, 2018.
- [10] Jenna R. Jambeck, Roland Geyer, Chris Wilcox, Theodore R. Siegler, Miriam Perryman, Anthony Andrady, Ramani Narayan, and Kara Lavender Law. Plastic waste inputs from land into the ocean. *Science*, 347(6223):768–771, 2015.
- [11] <https://ourworldindata.org/world-population-growth> (Last accessed on 2021-02-15).
- [12] <https://www.ellenmacarthurfoundation.org/assets/downloads/The-New-Plastics-Economy-Rethinking-the-Future-of-Plastics.pdf> (Last accessed on 2021-02-15).
- [13] David Tilman, Christian Balzer, Jason Hill, and Belinda L. Befort. Global food demand and the sustainable intensification of agriculture. *Proceedings of the national academy of sciences*, 108(50):20260–20264, 2011.
- [14] Dolf Gielen, Francisco Boshell, and Deger Saygin. Climate and energy challenges for materials science. *Nature materials*, 15(2):117–120, 2016.
- [15] https://read.oecd-ilibrary.org/environment/improving-plastics-management_c5f7c448-en (Last accessed on 2021-02-15).
- [16] Jeannette M. Garcia and Megan L. Robertson. The future of plastics recycling. *Science*, 358(6365):870–872, 2017.
- [17] Igor A. Ignatyev, Wim Thielemans, and Bob Vander Beke. Recycling of polymers: a review. *ChemSusChem*, 7(6):1579–1593, 2014.
- [18] AliReza Rahimi and Jeannette M. García. Chemical recycling of waste plastics for new materials production. *Nature Reviews Chemistry*, 1(6):1–11, 2017.
- [19] Narinder Singh, David Hui, Rupinder Singh, I. P. S. Ahuja, Luciano Feo, and Fernando Fraternali. Recycling of plastic solid waste: A state of art review and future applications. *Composites Part B: Engineering*, 115:409–422, 2017.
- [20] Kim Ragaert, Laurens Delva, and Kevin Van Geem. Mechanical and chemical recycling of solid plastic waste. *Waste Management*, 69:24–58, 2017.

-
- [21] Miao Hong and Eugene Y.-X. Chen. Chemically recyclable polymers: a circular economy approach to sustainability. *Green Chemistry*, 19(16):3692–3706, 2017.
- [22] Costantino Creton. Molecular stitches for enhanced recycling of packaging. *Science*, 355(6327):797–798, 2017.
- [23] James M. Eagan, Jun Xu, Rocco Di Girolamo, Christopher M. Thurber, Christopher W. Macosko, Anne M. LaPointe, Frank S. Bates, and Geoffrey W. Coates. Combining polyethylene and polypropylene: Enhanced performance with pe/ipp multiblock polymers. *Science*, 355(6327):814–816, 2017.
- [24] Grant W. Fahnhorst and Thomas R. Hoye. A carbomethoxylated polyvalerolactone from malic acid: synthesis and divergent chemical recycling. *ACS Macro Letters*, 7(2):143–147, 2018.
- [25] Xia Liu, Miao Hong, Laura Falivene, Luigi Cavallo, and Eugene Y.-X. Chen. Closed-loop polymer upcycling by installing property-enhancing comonomer sequences and recyclability. *Macromolecules*, 52(12):4570–4578, 2019.
- [26] Geoffrey W. Coates and Yutan D. Y. L. Getzler. Chemical recycling to monomer for an ideal, circular polymer economy. *Nature Reviews Materials*, 5(7):501–516, 2020.
- [27] Akio Kamimura and Shigehiro Yamamoto. An efficient method to depolymerize polyamide plastics: A new use of ionic liquids. *Organic Letters*, 9(13):2533–2535, 2007.
- [28] Kazuki Fukushima, Olivier Coulembier, Julien M. Lecuyer, Hamid A. Almegren, Abdullah M. Alabulrahman, Fares D. Alsewailam, Melanie A. Mcneil, Philippe Dubois, Robert M. Waymouth, Hans W. Horn, et al. Organocatalytic depolymerization of poly(ethylene terephthalate). *Journal of Polymer Science Part A: Polymer Chemistry*, 49(5):1273–1281, 2011.
- [29] Coralie Jehanno, Irma Flores, Andrew P. Dove, Alejandro J. Müller, Fernando Ruipérez, and Haritz Sardon. Organocatalysed depolymerisation of pet in a fully sustainable cycle using thermally stable protic ionic salt. *Green Chemistry*, 20(6):1205–1212, 2018.

References

- [30] Miao Hong and Eugene Y.-X. Chen. Completely recyclable biopolymers with linear and cyclic topologies via ring-opening polymerization of γ -butyrolactone. *Nature chemistry*, 8(1):42, 2016.
- [31] Jian-Bo Zhu, Eli M. Watson, Jing Tang, and Eugene Y.-X. Chen. A synthetic polymer system with repeatable chemical recyclability. *Science*, 360(6387):398–403, 2018.
- [32] Jian-Bo Zhu and Eugene Y.-X. Chen. Living coordination polymerization of a six-five bicyclic lactone to produce completely recyclable polyester. *Angewandte Chemie International Edition*, 57(38):12558–12562, 2018.
- [33] Haruo Nishida. Development of materials and technologies for control of polymer recycling. *Polymer journal*, 43(5):435–447, 2011.
- [34] V. Tournier, C. M. Topham, A. Gilles, B. David, C. Folgoas, E. Moya-Leclair, E. Kamionka, M.-L. Desrousseaux, H. Texier, S. Gavalda, et al. An engineered pet depolymerase to break down and recycle plastic bottles. *Nature*, 580(7802):216–219, 2020.
- [35] Xiaoyan Tang and Eugene Y.-X. Chen. Toward infinitely recyclable plastics derived from renewable cyclic esters. *Chem*, 5(2):284–312, 2019.
- [36] Timothy E. Long. Toward recyclable thermosets. *Science*, 344(6185):706–707, 2014.
- [37] Deborah K. Schneiderman, Marie E. Vanderlaan, Alexander M. Mannion, Tessie R. Panthani, Derek C. Batiste, Jay Z. Wang, Frank S. Bates, Christopher W. Macosko, and Marc A. Hillmyer. Chemically recyclable biobased polyurethanes. *ACS Macro Letters*, 5(4):515–518, 2016.
- [38] Jeannette M. García, Gavin O. Jones, Kumar Virwani, Bryan D. McCloskey, Dylan J. Boday, Gijs M. ter Huurne, Hans W. Horn, Daniel J. Coady, Abdulmalik M. Bintaleb, Abdullah M. S. Alabdulrahman, et al. Recyclable, strong thermosets and organogels via paraformaldehyde condensation with diamines. *Science*, 344(6185):732–735, 2014.
- [39] Shosuke Yoshida, Kazumi Hiraga, Toshihiko Takehana, Ikuo Taniguchi, Hironao Yamaji, Yasuhito Maeda, Kiyotsuna Toyohara, Kenji Miyamoto, Yoshiharu Kimura, and Kohei Oda. A bacterium that degrades and assimilates poly (ethylene terephthalate). *Science*, 351(6278):1196–1199, 2016.

-
- [40] Xiangqing Jia, Chuan Qin, Tobias Friedberger, Zhibin Guan, and Zheng Huang. Efficient and selective degradation of polyethylenes into liquid fuels and waxes under mild conditions. *Science advances*, 2(6):e1501591, 2016.
- [41] Fan Zhang, Manhao Zeng, Ryan D. Yappert, Jiakai Sun, Yu-Hsuan Lee, Anne M. LaPointe, Baron Peters, Mahdi M. Abu-Omar, and Susannah L. Scott. Polyethylene upcycling to long-chain alkylaromatics by tandem hydrogenolysis/aromatization. *Science*, 370(6515): 437–441, 2020.
- [42] Gavin O. Jones, Alexander Yuen, Rudy J. Wojtecki, James L. Hedrick, and Jeannette M. Garcia. Computational and experimental investigations of one-step conversion of poly (carbonate)s into value-added poly (aryl ether sulfone)s. *Proceedings of the National Academy of Sciences*, 113(28):7722–7726, 2016.
- [43] Jiajia Zheng and Sangwon Suh. Strategies to reduce the global carbon footprint of plastics. *Nature Climate Change*, 9(5):374–378, 2019.
- [44] Ina Vollmer, Michael J. F. Jenks, Mark C. P. Roelands, Robin J. White, Toon van Harmelen, Paul de Wild, Gerard P. van Der Laan, Florian Meirer, Jos T. F. Keurentjes, and Bert M. Weckhuysen. Beyond mechanical recycling: Giving new life to plastic waste. *Angewandte Chemie International Edition*, 59(36):15402–15423, 2020.
- [45] Richard A. Gross and Bhanu Kalra. Biodegradable polymers for the environment. *Science*, 297(5582):803–807, 2002.
- [46] Tadahisa Iwata. Biodegradable and bio-based polymers: future prospects of eco-friendly plastics. *Angewandte Chemie International Edition*, 54(11):3210–3215, 2015.
- [47] Mark Harvey and Sarah Pilgrim. The new competition for land: Food, energy, and climate change. *Food policy*, 36:S40–S51, 2011.
- [48] Jorgelina Pasqualino, Montse Meneses, and Francesc Castells. The carbon footprint and energy consumption of beverage packaging selection and disposal. *Journal of food Engineering*, 103(4):357–365, 2011.

References

- [49] Michael H. Ramage, Henry Burridge, Marta Busse-Wicher, George Fereday, Thomas Reynolds, Darshil U. Shah, Guanglu Wu, Li Yu, Patrick Fleming, Danielle Densley-Tingley, et al. The wood from the trees: The use of timber in construction. *Renewable and Sustainable Energy Reviews*, 68:333–359, 2017.
- [50] Jean-François Lutz. Aperiodic copolymers. *ACS Macro Letters*, 3(10):1020–1023, 2014.
- [51] Jean-François Lutz, Makoto Ouchi, David R. Liu, and Mitsuo Sawamoto. Sequence-controlled polymers. *Science*, 341(6146), 2013.
- [52] Nezha Badi and Jean-François Lutz. Sequence control in polymer synthesis. *Chemical Society Reviews*, 38(12):3383–3390, 2009.
- [53] Yusuke Hibi, Makoto Ouchi, and Mitsuo Sawamoto. A strategy for sequence control in vinyl polymers via iterative controlled radical cyclization. *Nature communications*, 7(1): 1–9, 2016.
- [54] Robert B. Merrifield. Solid phase peptide synthesis. i. the synthesis of a tetrapeptide. *Journal of the American Chemical Society*, 85(14):2149–2154, 1963.
- [55] Nina Hartrampf, Azin Saebi, Mackenzie Poskus, Zachary P. Gates, Alexander J Callahan, Amanda E. Cowfer, Stephanie Hanna, Sarah Antilla, Carly K. Schissel, Anthony J. Quarataro, et al. Synthesis of proteins by automated flow chemistry. *Science*, 368(6494): 980–987, 2020.
- [56] Jing Sun and Ronald N. Zuckermann. Peptoid polymers: a highly designable bioinspired material. *ACS nano*, 7(6):4715–4732, 2013.
- [57] Abdelaziz Al Ouahabi, Laurence Charles, and Jean-François Lutz. Synthesis of non-natural sequence-encoded polymers using phosphoramidite chemistry. *Journal of the American Chemical Society*, 137(16):5629–5635, 2015.
- [58] Sebastian Pfeifer, Zoya Zarafshani, Nezha Badi, and Jean-François Lutz. Liquid-phase synthesis of block copolymers containing sequence-ordered segments. *Journal of the American Chemical Society*, 131(26):9195–9197, 2009.

- [59] Raj Kumar Roy, Anna Meszynska, Chlo   Laure, Laurence Charles, Claire Verchin, and Jean-Fran  ois Lutz. Design and synthesis of digitally encoded polymers that can be decoded and erased. *Nature communications*, 6(1):1–8, 2015.
- [60] Gianni Cavallo, Abdelaziz Al Ouahabi, Laurence Oswald, Laurence Charles, and Jean-Fran  ois Lutz. Orthogonal synthesis of “easy-to-read” information-containing polymers using phosphoramidite and radical coupling steps. *Journal of the American Chemical Society*, 138(30):9417–9420, 2016.
- [61] Ufuk Saim Gunay, Benoit Eric Petit, Denise Karamessini, Abdelaziz Al Ouahabi, Jean-Arthur Amalian, Christophe Chendo, Michel Bouquey, Didier Gigmes, Laurence Charles, and Jean-Fran  ois Lutz. Chemoselective synthesis of uniform sequence-coded polyurethanes and their use as molecular tags. *Chem*, 1(1):114–126, 2016.
- [62] Jonathan C. Barnes, Deborah J. C. Ehrlich, Angela X. Gao, Frank A. Leibfarth, Yivan Jiang, Erica Zhou, Timothy F. Jamison, and Jeremiah A. Johnson. Iterative exponential growth of stereo-and sequence-controlled polymers. *Nature chemistry*, 7(10):810–815, 2015.
- [63] Susanne C. Solleder, Deniz Zengel, Katharina S. Wetzel, and Michael A. R. Meier. A scalable and high-yield strategy for the synthesis of sequence-defined macromolecules. *Angewandte Chemie International Edition*, 55(3):1204–1207, 2016.
- [64] Nicolas Zydziak, Waldemar Konrad, Florian Feist, Sergii Afonin, Steffen Weidner, and Christopher Barner-Kowollik. Coding and decoding libraries of sequence-defined functional copolymers synthesized via photoligation. *Nature communications*, 7(1):1–13, 2016.
- [65] Bartosz Lewandowski, Guillaume De Bo, John W. Ward, Marcus Papmeyer, Sonja Kuschel, Mar  a J. Aldegunde, Philipp M. E. Gramlich, Dominik Heckmann, Stephen M. Goldup, Daniel M. D’Souza, et al. Sequence-specific peptide synthesis by an artificial small-molecule machine. *Science*, 339(6116):189–193, 2013.
- [66] Jia Niu, Ryan Hili, and David R. Liu. Enzyme-free translation of dna into sequence-defined synthetic polymers structurally unrelated to nucleic acids. *Nature chemistry*, 5(4):282–292, 2013.

References

- [67] Germán L. Rosano and Eduardo A. Ceccarelli. Recombinant protein expression in *escherichia coli*: advances and challenges. *Frontiers in microbiology*, 5:172, 2014.
- [68] Michael R. Ladisch and Karen L. Kohlmann. Recombinant human insulin. *Biotechnology progress*, 8(6):469–478, 1992.
- [69] D. Calvin Harris and Michael C. Jewett. Cell-free biology: exploiting the interface between synthetic biology and synthetic chemistry. *Current opinion in biotechnology*, 23(5):672–678, 2012.
- [70] Erik D. Carlson, Rui Gan, C. Eric Hodgman, and Michael C. Jewett. Cell-free protein synthesis: applications come of age. *Biotechnology advances*, 30(5):1185–1194, 2012.
- [71] Yong-Chan Kwon and Michael C. Jewett. High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Scientific reports*, 5(1):1–8, 2015.
- [72] Yoshihiro Shimizu, Akio Inoue, Yukihide Tomari, Tsutomu Suzuki, Takashi Yokogawa, Kazuya Nishikawa, and Takuya Ueda. Cell-free translation reconstituted with purified components. *Nature biotechnology*, 19(8):751–755, 2001.
- [73] Yoshihiro Shimizu, Takashi Kanamori, and Takuya Ueda. Protein synthesis by pure translation systems. *Methods*, 36(3):299–304, 2005.
- [74] Kirsten Jackson, Takashi Kanamori, Takuya Ueda, and Z. Hugh Fan. Protein synthesis yield increased 72 times in the cell-free pure system. *Integrative Biology*, 6(8):781–788, 2014.
- [75] Barbora Lavickova and Sebastian J. Maerkl. A simple, robust, and low-cost method to produce the pure cell-free system. *ACS synthetic biology*, 8(2):455–462, 2019.
- [76] Yutetsu Kuruma and Takuya Ueda. The pure system for the cell-free synthesis of membrane proteins. *Nature protocols*, 10(9):1328, 2015.
- [77] Randall K. Saiki, David H. Gelfand, Susanne Stoffel, Stephen J. Scharf, Russell Higuchi, Glenn T. Horn, Kary B. Mullis, and Henry A. Erlich. Primer-directed enzymatic amplification of dna with a thermostable dna polymerase. *Science*, 239(4839):487–491, 1988.

-
- [78] Jan C. M. van Hest and David A. Tirrell. Protein-based materials, toward a new level of structural control. *Chemical communications*, (19):1897–1904, 2001.
- [79] Lei Wang and Peter G. Schultz. Expanding the genetic code. *Angewandte Chemie International Edition*, 44(1):34–66, 2005.
- [80] Christopher J. Noren, Spencer J. Anthony-Cahill, Michael C. Griffith, and Peter G. Schultz. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science*, 244(4901):182–188, 1989.
- [81] Jason W. Chin, Andrew B. Martin, David S. King, Lei Wang, and Peter G. Schultz. Addition of a photocrosslinking amino acid to the genetic code of escherichia coli. *Proceedings of the National Academy of Sciences*, 99(17):11020–11024, 2002.
- [82] Kaihang Wang, Heinz Neumann, Sew Y. Peak-Chew, and Jason W. Chin. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature biotechnology*, 25(7):770–777, 2007.
- [83] Takahiro Hohsaka, Yuki Ashizuka, Hikaru Taira, Hiroshi Murakami, and Masahiko Sisido. Incorporation of nonnatural amino acids into proteins by using various four-base codons in an escherichia coli in vitro translation system. *Biochemistry*, 40(37):11060–11064, 2001.
- [84] Heinz Neumann, Kaihang Wang, Lloyd Davis, Maria Garcia-Alai, and Jason W. Chin. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature*, 464(7287):441–444, 2010.
- [85] Jason W. Chin. Expanding and reprogramming the genetic code. *Nature*, 550(7674):53–60, 2017.
- [86] Yuki Goto, Takayuki Katoh, and Hiroaki Suga. Flexizymes for genetic code reprogramming. *Nature protocols*, 6(6):779–790, 2011.
- [87] Toby Passioura and Hiroaki Suga. Reprogramming the genetic code in vitro. *Trends in biochemical sciences*, 39(9):400–408, 2014.

References

- [88] Atsushi Ohta, Hiroshi Murakami, Eri Higashimura, and Hiroaki Suga. Synthesis of polyester by means of genetic code reprogramming. *Chemistry & biology*, 14(12):1315–1322, 2007.
- [89] Matthew C. T. Hartman, Kristopher Josephson, and Jack W. Szostak. Enzymatic aminoacylation of trna with unnatural amino acids. *Proceedings of the National Academy of Sciences*, 103(12):4356–4361, 2006.
- [90] Kristopher Josephson, Matthew C. T. Hartman, and Jack W. Szostak. Ribosomal synthesis of unnatural peptides. *Journal of the American Chemical Society*, 127(33):11727–11735, 2005.
- [91] Matthew C. T. Hartman, Kristopher Josephson, Chi-Wang Lin, and Jack W. Szostak. An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. *PloS one*, 2(10):e972, 2007.
- [92] E. Railey White, Timothy M. Reed, Zhong Ma, and Matthew C. T. Hartman. Replacing amino acids in translation: Expanding chemical diversity with non-natural variants. *Methods*, 60(1):70–74, 2013.
- [93] Oliver Thum, Stefan Jäger, and Michael Famulok. Functionalized dna: a new replicable biopolymer. *Angewandte Chemie International Edition*, 40(21):3990–3993, 2001.
- [94] Stefan Jäger, Goran Rasched, Hagit Kornreich-Leshem, Marianne Engeser, Oliver Thum, and Michael Famulok. A versatile toolbox for variable dna functionalization at high density. *Journal of the American Chemical Society*, 127(43):15071–15082, 2005.
- [95] Michal Hocek. Synthesis of base-modified 2'-deoxyribonucleoside triphosphates and their use in enzymatic synthesis of modified dna for applications in bioanalysis and chemical biology. *The Journal of organic chemistry*, 79(21):9914–9921, 2014.
- [96] Michael A. R. Meier and Christopher Barner-Kowollik. A new class of materials: Sequence-defined macromolecules and their emerging applications. *Advanced Materials*, 31(26):1806027, 2019.
- [97] Anika M. Jonker, Dennis W. P. M. Löwik, and Jan C. M. Van Hest. Peptide-and protein-based hydrogels. *Chemistry of Materials*, 24(5):759–773, 2012.

-
- [98] Fiorenzo G. Omenetto and David L. Kaplan. New opportunities for an ancient material. *Science*, 329(5991):528–531, 2010.
- [99] Dae-Hyeong Kim, Jonathan Viventi, Jason J. Amsden, Jianliang Xiao, Leif Vigeland, Yun-Soung Kim, Justin A. Blanco, Bruce Panilaitis, Eric S. Frechette, Diego Contreras, et al. Dissolvable films of silk fibroin for ultrathin conformal bio-integrated electronics. *Nature materials*, 9(6):511–517, 2010.
- [100] Hu Tao, David L. Kaplan, and Fiorenzo G. Omenetto. Silk materials—a road to sustainable high technology. *Advanced materials*, 24(21):2824–2837, 2012.
- [101] Hong Wang, Bowen Zhu, Hua Wang, Xiaohua Ma, Yue Hao, and Xiaodong Chen. Ultra-lightweight resistive switching memory devices based on silk fibroin. *Small*, 12(25):3360–3365, 2016.
- [102] Chaoxu Li, Jozef Adamcik, and Raffaele Mezzenga. Biodegradable nanocomposites of amyloid fibrils and graphene with shape-memory and enzyme-sensing properties. *Nature Nanotechnology*, 7(7):421–427, 2012.
- [103] Sreenath Bolisetty and Raffaele Mezzenga. Amyloid–carbon hybrid membranes for universal water purification. *Nature nanotechnology*, 11(4):365–371, 2016.
- [104] Shuihong Zhu, Wenbin Zeng, Zhaohui Meng, Wenhao Luo, Liyun Ma, Yanran Li, Changxu Lin, Qiaoling Huang, Youhui Lin, and Xiang Yang Liu. Using wool keratin as a basic resist material to fabricate precise protein patterns. *Advanced Materials*, 31(28):1900870, 2019.
- [105] Yao Wang, Priya Katyal, and Jin Kim Montclare. Protein-engineered functional materials. *Advanced healthcare materials*, 8(11):1801374, 2019.
- [106] Matthew R. Jones, Nadrian C. Seeman, and Chad A. Mirkin. Programmable materials and the nature of the dna bond. *Science*, 347(6224), 2015.
- [107] Nadrian C. Seeman and Hanadi F. Sleiman. Dna nanotechnology. *Nature Reviews Materials*, 3(1):1–23, 2017.

References

- [108] Pongphak Chidchob and Hanadi F. Sleiman. Recent advances in dna nanotechnology. *Current opinion in chemical biology*, 46:63–70, 2018.
- [109] Dong Wang, Jinhui Cui, Mingzhe Gan, Zhaohui Xue, Jing Wang, Peifeng Liu, Yue Hu, Yehudah Pardo, Shogo Hamada, Dayong Yang, et al. Transformation of biomass dna into biodegradable materials from gels to plastics for reducing petrochemical consumption. *Journal of the American Chemical Society*, 142(22):10114–10124, 2020.
- [110] Elsa A. Olivetti and Jonathan M. Cullen. Toward a sustainable materials system. *Science*, 360(6396):1396–1398, 2018.
- [111] Pedro F. Teixeira, Beata Kmiec, Rui M. M. Branca, Monika W. Murcha, Anna Byzia, Aneta Ivanova, James Whelan, Marcin Drag, Janne Lehtiö, and Elzbieta Glaser. A multi-step peptidolytic cascade for amino acid recovery in chloroplasts. *Nature chemical biology*, 13(1):15–17, 2017.
- [112] Pu Chun Ke, Ruhong Zhou, Louise C. Serpell, Roland Riek, Tuomas P. J. Knowles, Hilal A. Lashuel, Ehud Gazit, Ian W. Hamley, Thomas P. Davis, Marcus Fändrich, et al. Half a century of amyloids: past, present and future. *Chemical Society Reviews*, 49(15): 5473–5509, 2020.
- [113] Matthew Verosloff, James Chappell, Keith L. Perry, Jeremy R. Thompson, and Julius B. Lucks. Plant-dx: a molecular diagnostic for point-of-use detection of plant pathogens. *ACS synthetic biology*, 8(4):902–905, 2019.
- [114] Luc Patiny and Alain Borel. Chemcalc: a building block for tomorrow’s chemical infrastructure. *Journal of Chemical Information and Modeling*, 53(5):1223–1228, 2013.
- [115] Ronald F. S. Lee, Laure Menin, Luc Patiny, Daniel Ortiz, and Paul J. Dyson. Versatile tool for the analysis of metal–protein interactions reveals the promiscuity of metallodrug–protein interactions. *Analytical chemistry*, 89(22):11985–11989, 2017.
- [116] Robert E. Campbell, Oded Tour, Amy E. Palmer, Paul A. Steinbach, Geoffrey S. Baird, David A. Zacharias, and Roger Y. Tsien. A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences*, 99(12):7877–7882, 2002.

List of publications

During my doctoral studies I have been working on the following preprint:

- doi: 10.1101/2020.09.23.309799 version 3, <https://www.biorxiv.org/content/10.1101/2020.09.23.309799v3>

Simone Giaveri

Doctoral Assistant

🏠 Route du Grand-Mont, 63
CH-1052 Le Mont-sur-Lausanne, Switzerland
☎ +41 76 611 90 65 📧 simone_giaveri
✉ simone.giaveri@epfl.ch
🐦 @simone_giaveri
🌐 <https://people.epfl.ch/simone.giaveri>

Strengths

Passionate to start cutting-edge researches at the interface between multiple disciplines, leading them autonomously, looking for founding, and teaching how to further develop them.

Research

During my doctoral studies I developed a new vision about recycling of plastics, inspired by nature, under the supervision of Prof. Francesco Stellacci, and Prof. Sebastian J. Maerkl. This study demonstrates, as a proof-of-concept, that it is possible to envision a circular-economy model for polymers, where materials are transformed into new materials, not necessarily related to the parent ones. This vision is based on sequence-defined polymers whose backbone can be readily depolymerized, and polymerization reactions starting from random mixtures of monomers. The work has been developed at the interface between materials science, biochemistry, and synthetic biology.

Research Interests

Materials science, synthetic biology, metabolic engineering, synthetic metabolism, biochemistry, and combinations of them.

Education

- 2016 – 2021 🎓 **Ph.D.** in Materials Science and Engineering.
École Polytechnique Fédérale de Lausanne (EPFL).
- 2012 – 2015 🎓 **Master of Science Degree** in Materials Engineering and Nanotechnology (*cum laude*).
Politecnico di Milano, Milano, Italy.
- 2009 – 2012 🎓 **Bachelor of Science Degree** in Materials Engineering and Nanotechnology (109/110).
Politecnico di Milano, Milano, Italy.



Core Experience

- 2016 – 2021 ⚙️ **Doctoral studies:** *Nature-inspired Circular-economy Recycling (NaCRe) for Proteins: Proof of Concept* at SuNMIL-LBNC, under the supervision of Prof. Stellacci and Prof. Maerkl.
- 👥 **Main teaching assistantships:** *Thermodynamics for Materials Science* (2016 and 2017 - Prof. Stellacci, 2018 - Prof. Tileli), *MAN Maths1A* (2018 - Dr. Genoud).
- 👤 **Students' supervision:** V.L.P. De Mestral, and A.M. Schmitt (now PhD assistant at ETH Zurich).
- 📄 **Grant applications:** strong collaboration with Prof. Stellacci for writing the SNF Spark project CRSK-2_190167, and the ERC Advanced Grant (884114- NaCRe) based on my research, and its future development.
- 📖 **Main (attended) doctoral courses:** *3rd International Synthetic & Systems Biology Summer School SSBSS* (2016), *EuroTech Winter School - Molecular Engineering of Synthetic Biological Systems* (2017), *Mass spectrometry, principles and applications* EDCH-EPFL (2018), *Shaping the Future of Medicine SFM2019* (2019).
- 2014 – 2016 ⚙️ **Working experiences:** *High Temperature Organic Powder Coatings: characterization and innovation* (internship for master thesis at AkzoNobel, supervised by Prof. Gronchi), study of *Hydrophilic and superhydrophilic surfaces and materials for de-humidifying heat exchangers* (post-graduate fellowship, supervised by Prof. Gronchi).









Technical Experience

🧪 Polymerase chain reaction (PCR), cell-free transcription-translation (TX-TL), incorporation of unnatural amino acids (UAAs), protein electrophoresis, enzymatic cleavage and digestion of proteins, protein UV characterization, high-resolution mass spectrometry HRMS (sample preparation, and data analysis).



Publications

-  Nature-inspired Circular-economy Recycling (NaCRE) for Proteins: Proof of Concept. (doi: 10.1101/2020.09.23.309799 version 3, <https://www.biorxiv.org/content/10.1101/2020.09.23.309799v3>).
-  S. Giaveri, P. Gronchi, and A. Barzoni. IPN Polysiloxane-Epoxy Resin for High Temperature Coatings: Structure Effects on Layer Performance after 450 °C Treatment. *Coatings*, 7(12):213, 2017. (Cover Issue 12, 2017).


Attended Conferences

- 2021  Talk (in preparation): *IUPAC CCCE 2021*.
 Invited talk: *Seminar at Erb group, MPI Marburg*.
 Invited talk: *EDMX Research Day, EPFL*.
- 2017– 2020  Poster presentation: *NCCR Bio-Inspired Materials Annual Center Conference*.
- 2019  Poster presentation: *Sixth International Symposium Frontiers in Polymer Science*.
- 2018  Poster presentation: *Annual Meeting of the SCS Division of Polymers Colloids & Interfaces*.
- 2017  Invited talk: *NCCR Bio-Inspired Materials Annual Center Conference*.
 Conference attendance: *1st European Congress on Cell-Free Synthetic Biology ECCSB*.


Skills and Competences

-  **Softwares:** XcaliburTM (Thermo ScientificTM), MassHunter Workstation (Agilent), apm²S, UCSF Chimera, ChemDraw, SnapGene, Python (Bio, Matplotlib), SLIMS (Agilent), Illustrator, Affinity Designer, \LaTeX .
-  **Languages:** Italian (mother tongue), English (proficient), Spanish (good knowledge), French (beginner).


Personal Interests

-  Skiing, hiking, mountaneering, free climbing, fishing, playing ping-pong, dancing Swing.



Personal Informations

-  30 years old, Italian citizen, B permit in Switzerland, single.

Additional info

-  Member of the Italian Alpine Club (CAI) of my hometown since 2005 (lead climber/alpinist for many Alp tours).
-  Member of the best 100 Italian high school students for a stage at Scuola Superiore Sant' Anna (Pisa, Italy, June 2008).

References

- 2015 – today  Prof. Francesco Stellacci
Full Professor
Institute of Materials & Institute of Bioengineering
École Polytechnique Fédérale de Lausanne (EPFL)
Supramolecular Nano-Materials and Interfaces Laboratory, SuNMIL - Constellium Chair
<http://sunmil.epfl.ch/>
EPFL STI IMX SUNMIL MXG 030 (Bâtiment MXG) Station 12
CH-1015, Lausanne, VD, Switzerland
francesco.stellacci@epfl.ch
-  Prof. Sebastian J. Maerkl
Associate Professor
Institute of Bioengineering
École Polytechnique Fédérale de Lausanne (EPFL)
Laboratory of Biological Network Characterization, LBNC
<http://lbnc.epfl.ch>
EPFL STI IBI LBNC BM 5130 (Bâtiment BM) Station 17
CH-1015, Lausanne, VD, Switzerland
sebastian.maerkl@epfl.ch

