Nature-inspired Circular-economy Recycling (NaCRe) for Proteins: Proof of Concept

Simone Giaveri[1,2], Adeline M. Schmitt[1,3], Laura Roset Julià[1], Anna Murello[1], Laure Menin[4], Daniel Ortiz[4], Luc Patiny[4], Sreenath Bolisetty[5], Raffaele Mezzenga[5,6], Sebastian J. Maerkl[2], Francesco Stellacci[1,2*]*

Authors information

[1]Institute of Materials, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.
[2]Institute of Bioengineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.
[3]Faculty of Chemistry, University of Strasbourg, CNRS, Strasbourg, France.
[4]Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.
[5]Institute of Food, Nutrition and Health, ETH Zurich, Zürich, Switzerland.
[6]Department of Materials, ETH Zurich, Zürich, Switzerland.
[*]Corresponding authors (sebastian.maerkl@epfl.ch, francesco.stellacci@epfl.ch).

Abstract

In 2070, 10^{12} Kg of polymer-based materials (*i.e.* plastics) might be produced yearly, posing one of the greatest challenges that humanity has to face. Even though natural polymers, such as proteins and nucleic acids, are more abundant than synthetic ones, they are sustainable. The key property of such natural polymers is that they are sequence-defined. This allows for recycling to start with depolymerization into monomers, and end in the re-assembly of new polymers of arbitrarily different sequence. This process breaks a common recycling paradigm that a material is recycled only into itself. An organism digests proteins into amino acids. These are re-assembled into new proteins whose identity depends on the cell's needs at the time of protein synthesis. Here we show that the process described above is achievable extra-cellularly. Specifically, we depolymerized a mixture of different peptides and/or proteins into their amino acid constituents and used these amino acids to synthesize fluorescent proteins using an amino acid-free cell-free transcription-translation system. We were successful in recycling proteins with high relevance in materials engineering (β-lactoglobulin films, used for water filtration, or silk fibroin solutions) into a biotechnologically relevant protein (green fluorescent protein). The potential long-term impact of this approach to recycling lies in its compatibility with circular-economy models where raw materials remain in use as long as possible, thus reducing the burden on the planet.

Introduction

The world's projected population will be 11 billion by 2070[1]. One of the most daunting sustainability challenges linked to such a large population size will be the handling of all plastic products[2], *i.e.* the production and recycling of polymers[3]. Currently, there are only a few methods to recycle polymers and they are severely limited. Mechanical recycling inevitably leads to the original material but with lower quality[4]. A better possibility is chemical recycling[5,6], *e.g.* depolymerization of a polymer into its constituent monomers in order to re-polymerize the same virgin quality polymer[7], or, at best, co-polymers that are better than the starting materials (upcycling)[8], albeit with a certain degree of similarity between the recycled materials and the final products[7,9]. Recycling a material (and hence a polymer) into the same material is the current paradigm in recycling. To go beyond this paradigm, polymers are fully degraded into small molecules, burning the polymers and using the combustion products for further chemical production, ideally also extracting energy from the heat of combustion. Currently, this process is inefficient and too costly; in fact, while heat is recovered, most of the combustion products are released into the environment[5,6]. A better approach is the use of bio-sourced/-degradable polymers, *i.e.* materials that are derived from renewable sources and that can be degraded into environmentally benign substances[10,11]. This approach takes inspiration from the way nature handles some natural polymers such as lignin and cellulose. Yet, these natural materials grow slowly, remain in use a long time, and bio-degrade slowly. This balance is always present in nature's approach to recycling. Currently, man-made bio-sourced/-degradable polymers are produced for consumers products that often have very short lifetimes (days or less), but in the environment degrade over months or years. As a consequence, no matter how 'green' such materials will appear to be, there will be significant environmental concerns due to their accumulation into the environment. By 2070, ~10^{12} Kg of plastics are projected to be produced yearly[12]. Were all polymers to be bio-sourced and bio-degraded (*i.e.,* the best-case scenario), the sustainability problem would remain. Sourcing will generate issues in deforestation and in competition for land with food production[13]. The issues in disposing will be the accumulation into the environment. A major concern will be the deleterious effects of the intermediate degradation components on soils[4]. Moreover, the final degradation products will have negative ecological effects, as such large quantities will inevitably shift the equilibrium of local ecosystems[4].

It is clear that humans should move towards the use of models that rely on the principles of a circular economy[6,14] where materials, once produced, remain in use for the longest possible amount of time, taxing earth the minimum possible. The question is whether this is at all possible for polymers, that mostly lose quality upon recycling as opposed, for example, to metals. To address this question, one can be inspired by nature. It is undeniable that nature is sustainable: it takes most of its energy from the sun, food production is commensurate to population, and materials are used in a circular manner. While we have more than 1 billion tons of biological soft-matter produced on earth yearly, we do not have a sustainability concern with it. When pausing to observe nature's main polymers, *e.g.* proteins, each one characterized by its own specific sequence of monomers, the 20 proteinogenic amino acids (AAs), it is possible to admire the circularity in their use. A vast over-simplification of protein metabolism shows that, proteins can be depolymerized into AAs that, in turn, can be reassembled into a new protein by the ribosomal machinery of the cell. The newly formed protein can have a sequence that differs from any of the sequences of the original proteins. It is fair to say that this approach breaks the recycling paradigm, *i.e.* that materials are recycled into a lower version of themselves. In nature this is not the case, a protein can be of much higher complexity than its 'parent' proteins with which it has only the individual AA building blocks in common. Nature can achieve this impressive result because proteins are sequence-defined polymers (SDPs), *i.e.* their remarkable structural and property diversity derives from the sequence of the AAs that compose them, and not from their chemical diversity[15]. Furthermore, the backbone chemical bonds that link AAs are reversibly cleavable and there exists a machinery (the ribosome) capable of synthesizing proteins starting from a random mixture of

AAs. Nature is able of recycling a mixture of n SDPs into an arbitrary $(n+1)^{th}$ SDP whose sequence (and consequently property) can be completely different from any of the sequences of the n parent polymers. We call this approach to polymer recycling nature-inspired circular-economy recycling (NaCRe) (Fig. 1). Arguably, it would be a breakthrough, if, in future, a large quantity of human-made polymers could be NaCRe-compatible SDPs. Recycling any mixture of these polymers would mean producing a generic SDP: the one that is most needed at the moment by the local community. In this vision, there would be no way of distinguishing a new from a recycled polymer. In theory, a mixture of SDPs could be recycled into a SDP that had not yet been developed at the time its parent polymers were produced. Each of these facts would disrupt the economic model in polymer recycling. Currently, the multibillion dollar market in recycled plastics[16] is basically a cost to be paid by the environment. Using NaCRe it is theoretically possible to use low value polymers as starting material to produce high value polymers, thus creating value upon recycling[17].

Clearly, this vision will take decades (if not centuries) to be implemented, as the technological challenges are significant. This paper is intended as a proof-of-concept where we present the feasibility of the overall process, showing that it is possible to recycle mixtures of peptides and engineering-relevant proteins into fluorescent proteins (*i.e.* proteins with relevance in biotechnology).

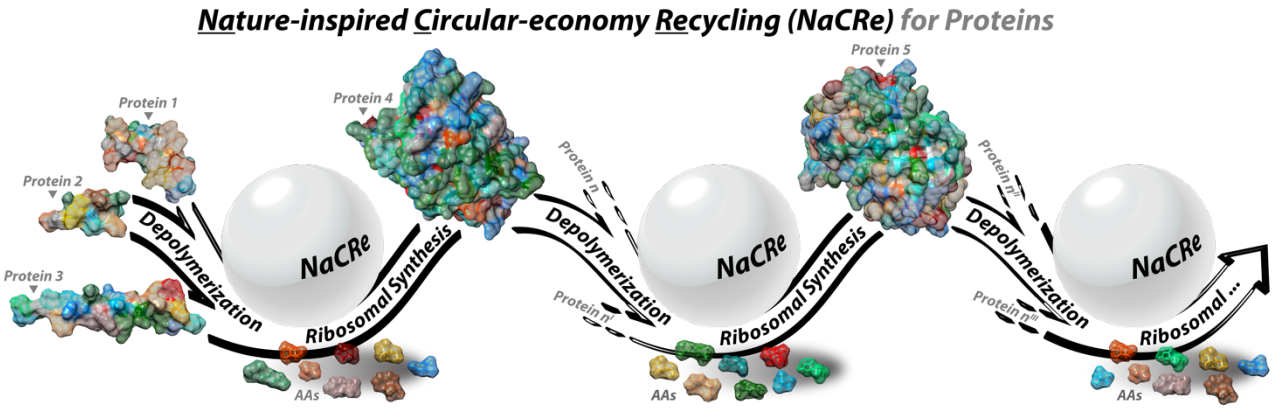


Figure 1 | Schematic illustration for the main concept of NaCRe. Multiple possible NaCRe cycles are shown. The illustrated examples are close to what is shown in this paper. It should be clear that the overall concept of NaCRe goes beyond what is illustrated. The sketched process starts from three different short peptides (drawn as the ones used in the paper, magainin II, glucagon, and somatostatin), and produces 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.

Results and Discussion

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, PURExfrex™, Kaneka Eurogentec SA, Supplementary Methods 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. The main issue with commercial TX-TL systems is that they contain free AAs. We chose PURExpress™ 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 coworkers[18]. It should be noted that the PURExpress™ 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 PURExpress™ 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 (Supplementary Fig. S1). 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 Supplementary Methods d). From the hits, we selected commercially available peptides, presenting different secondary structures, and different functions. Magainin II (Supplementary Table S1) is an antimicrobial peptide, and glucagon (Supplementary Table S1) is a peptide hormone. Somatostatin 28 (Supplementary Table S1), 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 Fig. 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.*[19] We incubated the peptides first with thermolysin endoprotease, then with leucine aminopeptidase (LAP), as described in Supplementary Methods e and f. Mass spectrometric (MS) analysis of the materials before (Supplementary Figs. S10-12), and after thermolysin treatment (Supplementary Figs. S14-27) shows extensive cleavage at the N-terminus of the hydrophobic amino acids (see Supplementary Methods i). Cleaved fragments were incubated with LAP and depolymerized to their free AAs (Fig. 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 Fig. 2, respectively). Quantification was performed using MS (Supplementary Methods j). We achieved an average depolymerization yield of ~66% ± 19%. The large standard deviation (1σ=19%) is caused by the large variation between depolymerization yields of different AAs, with a maximum of ~99% for aspartic acid (for glucagon) and a minimum of ~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. Phenylalanine was consistently recovered in poor yields, we believe that a possible cause for this behavior is the strong hydrophobic nature of this AA (and possibly of some fragments containing it) that lead to losses on the filters and throughout the process. 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 (Supplementary Table S3, Supplementary Methods h). As shown in Fig. 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, Supplementary Methods h).

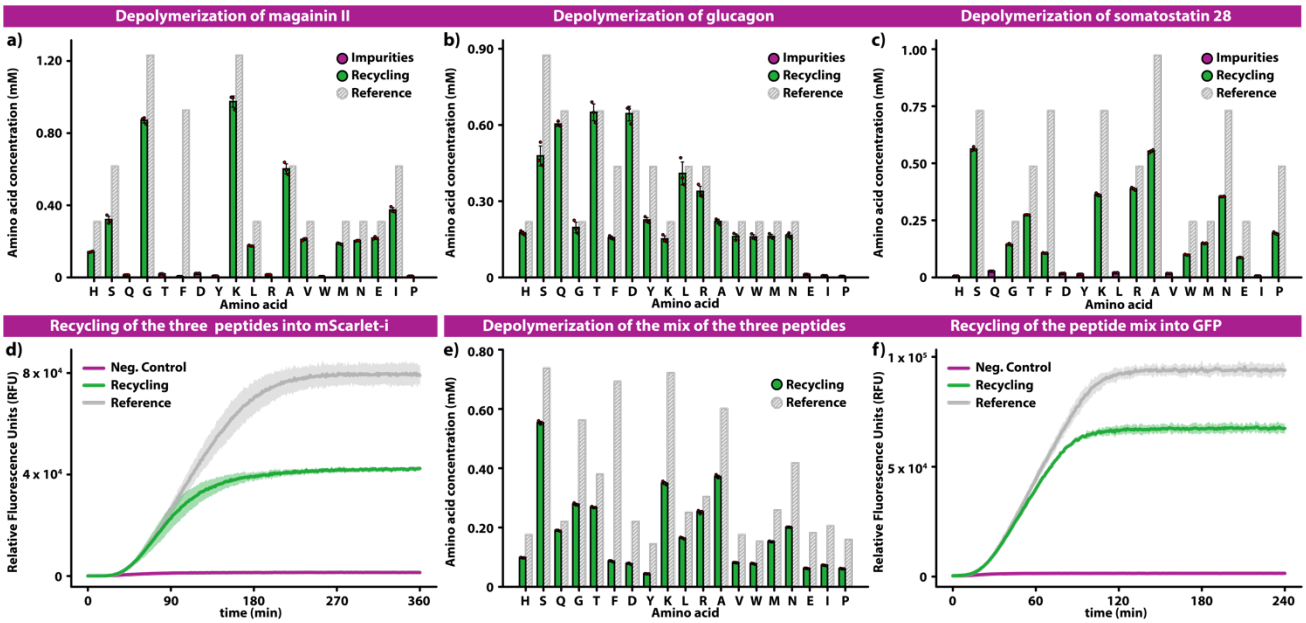


Figure 2 | Recycling of magainin II, glucagon, and somatostatin 28 into mScarlet-i. Bar graphs showing the result of the amino acid analysis performed using mass spectrometry of the result of the depolymerization of magainin II (**a**), glucagon (**b**), and somatostatin 28 (**c**), and their mixture (**e**). 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 AAs that theoretically should have not been observed, they are possibly the result of depolymerization of the digestion enzymes themselves. Plots of the fluorescence signal resulting from the expression of mScarlet-i (**d**) and GFP (**f**) in a TX-TL reaction. The green curves are data obtained preforming NaCRe, 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), (c) and (e). 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.

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 Fig. 2d respectively), leading to a yield of ~50%. We also used NaCRe to express GFP (see Supplementary Table S3). 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% (Supplementary Fig. S2). 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 and expressing GFP. As shown in Fig. 2e and f, the process was 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). They are the ratio between the fluorescent signal for the expression of the protein in the case of the complete NaCRe process to the signal achieved by a reference control. The Absolute Yield (AY) for NaCRe would be

the RY multiplied by the yield of expression in the TX-TL system. AY can be determined as a mass to mass ratio of the initial to the final product, and in the case of the expression of mScarlet-i is ~7%. 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 depolymerized as well as on the specific sequence of the protein to be expressed. As shown in Table 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. 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 1, this is not necessarily the case.

Table 1 | Overview of the depolymerization and expression efficiencies for key experiements in this study. Minima are colored in blue (depolymerization) and red (expression).

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

* Calculated as the ratio between the amount of each AA (nmol) and the number of its incorporations inside a single protein chain, see Supplementary Table S3.

** n.a. not assessable because E is present in the TX-TL system as Potassium glutamate (buffer), see Supplementary Methods h.

Even though the RY is a simple measure of the efficiency of our process, it is strongly affected by the nature of the true limiting AA in the NaCRe and in the reference control. In theory, the AY should be affected by the specific sequence of the product as the mass of the final product misses all of the free AAs left in solution because the true limiting AA has run out. In reality, it is known that expression in PURE systems stops due to a complexity of events that goes beyond AAs running out. The present results show a mass-to-mass yield for NaCRe for the limiting AA of proline in the expression of m-Scarlet-i of ~ 15%. This is the most accurate measurement of the absolute yield of the process.

To go beyond peptides, we performed NaCRe starting from larger proteins with defined tertiary structures. We started by recycling β-lactoglobulin A (~18 KDa, Supplementary Table S1), a protein that can be obtained in large quantities as a side product of bovine milk production. As shown in Fig. 3a, β-lactoglobulin A was successfully depolymerized into its constitutive AAs with a yield comparable to the ones obtained for the peptides (see Supplementary Methods e and f). These AAs were used to express GFP (Supplementary Fig. S3, Supplementary Methods h). The RY for β-lactoglobulin A recycled into GFP was ~40%.

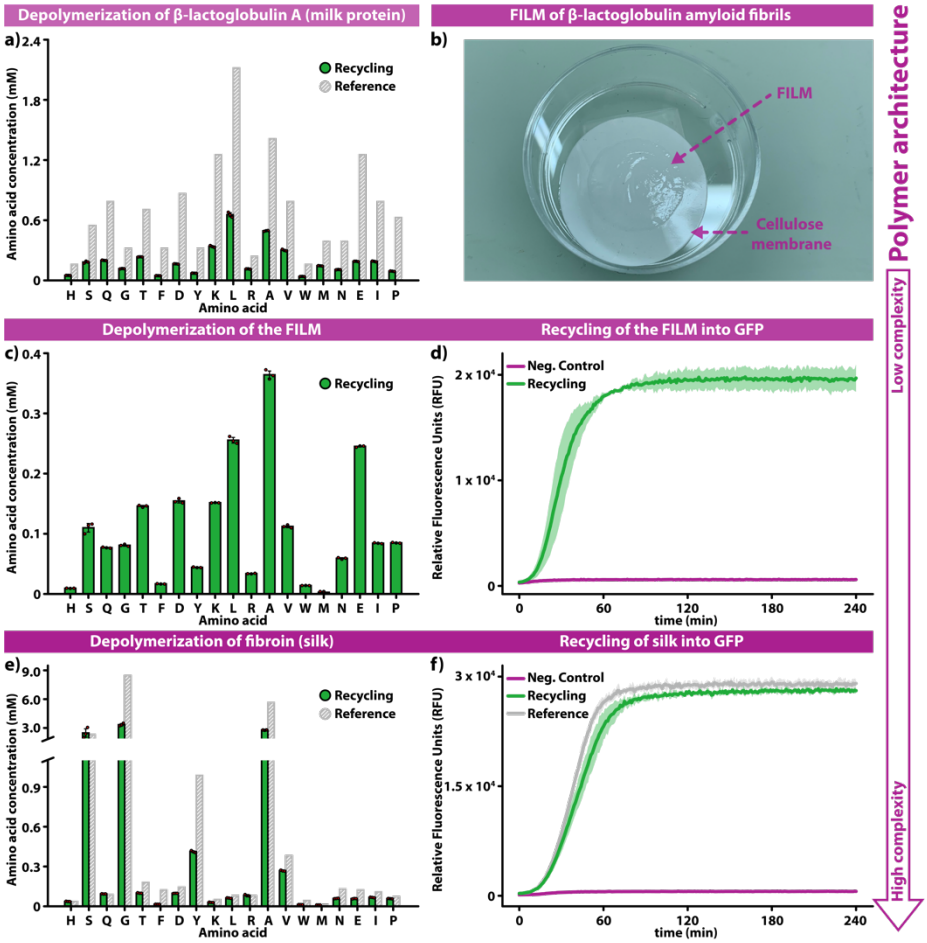


Figure 3 | Recycling of technologically relevant materials. Bar graphs of the results of the amino acid analysis for the depolymerization of β-lactoglobulin A (**a**), β-lactoglobulin amyloid film (**c**), and silk fibroin solution (**e**). The color scheme (and its meaning) is identical to the one used in Fig. 2, the ideal reference (gray bar) is missing from (**c**) because the exact composition of the amyloids composing the film is unknown. **b,** Photograph of a film composed of β-lactoglobulin amyloid fibrils. Real-time plots of the fluorescence signal resulting from the expression of GFP from the depolymerization of the amyloid film (**d**) and the silk fibroin solution (**f**); color scheme as in Fig. 2.

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 efficiency[20]. 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 landscape[21], 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 Fig. 3b (see Supplementary Methods c). The dry film was removed from the support, the film powder was weighed, and first incubated with pepsin endoprotease, then with LAP (Supplementary Methods e and f, respectively). 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 (Supplementary Fig. S50). The mass spectrometry result of the consecutive cleavage, and depolymerization is shown in Fig. 3c. 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 obtained only at low concentrations. As shown in Fig. 3d, the free AAs obtained from the β-lactoglobulin film were recycled into GFP, by spiking cysteine, methionine, and histidine (see Supplementary Methods h). We then recycled a solution of silk fibroin (Supplementary Table S1), another technologically relevant protein used in many devices, ranging from biomedical[22] to electronics applications[23]. After incubating fibroin with thermolysin, and then LAP (see Supplementary Methods e and f, respectively), we successfully recovered fibroin's free AAs (Fig. 3e), and used them to express GFP in our TX-TL system (Fig. 3f) spiked with cysteine, and methionine (see Supplementary Methods h). RY for silk fibroin recycling into GFP was ~95%. Fig. 3d and f 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 establish NaCRe in a way that did not depend on cysteine, by spiking cysteine every time we were expressing GFP. We then attempted to recycle magainin II, glucagon, and somatostatin 28 into GFP, without adding any cysteine (Supplementary Methods h). As shown in Supplementary Fig. S8, spiking cysteine was not necessary, since the two recycling curves reach approximately the same plateau, that is cysteine from the disulfide cyclization of somatostatin 28 is recycled into GFP. Finally we used suitable mixtures of low and high molecular weight proteins to produce GFP by using NaCRe, without the need of spiking any extra AA. We produced a mixture of glucagon, β-lactoglobulin A, and silk fibroin, and we successfully recycled it into GFP, as shown in Supplementary Fig. S9. RY for recycling this mixture of proteins into GFP was ~70%.

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 Supplementary Table S1), 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 Supplementary Methods e and f. MS analysis before (Supplementary Fig. S13), and after thermolysin incubation (Supplementary Figs. S28-31, and Supplementary Methods i) shows extensive cleavage. After depolymerization with LAP, we identified all the residues composing the non-natural peptide (Supplementary Figs. S32-37, and Supplementary Methods i). L-norleucine, and L-canavanine were successfully recycled into GFP (Supplementary Figs. S38-49, and Supplementary Methods h and l), following the protocol developed in references 24 and 25. The GFP produced in this way is most probably not fluorescent, that is the final product of this approach is a sequence-defined polymer composed of a set of monomers that goes beyond the 20 proteinogenic AAs. 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.

Conclusions

The results presented show that it is possible to envision a new way of recycling polymers were mixtures are transformed into a single targeted final polymer. We have shown this principle for peptides and proteins. The recycling of the β-lactoglobulin film, and of silk fibroin into GFP can be used as a proof of concept for the generality of NaCRe, showing its potential to recycle materials composed of complex molecular architectures. Particularly noteworthy is that such a process can be carried out also from protein templates with extremely robust secondary structures, as in the case of amyloids. NaCRe is a complex yet powerful way to think about recycling, were mixtures of (soft) materials are transformed into new (soft) materials, not (necessarily) related to the parent ones. NaCRe

has two key requirements, the materials must be sequence-defined macromolecules based on links that can be readily depolymerized, and the polymerization reaction must be based on an approach that can use a random mixtures of monomers as starting materials (this is what the ribosome does exceptionally well). Many efforts have recently focused on developing increasingly complex synthetic SDPs[26,27] for applications ranging from data storage[28–30] to catalysis [31,32]. It is true that these SDPs are all synthesized with approaches that are incompatible with NaCRe as they are based on step-growth synthesis and require the separation of the starting monomers. It should be noted though that significant efforts exist in creating a synthetic equivalent of the ribosome[33,34], that is a system capable of synthesizing SDPs starting from mixtures of monomers. With this work we hope to highlight an additional advantage of Sequence-Defined Polymers, their amazing ability to be recycled in ways that fulfill the vision of a circular economy.

Methods

Comprehensive information about the materials and procedures is detailed in the Supplementary Methods a and b-n sections of the Supplementary information respectively.

References

1. https://ourworldindata.org/world-population-growth.
2. https://ec.europa.eu/commission/news/eu-plastics-strategy-2018-nov-20_en.
3. https://www.ellenmacarthurfoundation.org/assets/downloads/The-New-Plastics-Economy-Rethinking-the-Future-of-Plastics.pdf.
4. Hong, M. & Chen, E. Y.-X. Chemically recyclable polymers: a circular economy approach to sustainability. *Green Chem.* **19**, 3692–3706 (2017).
5. Rahimi, A. & García, J. M. Chemical recycling of waste plastics for new materials production. *Nature Reviews Chemistry* **1**, 0046 (2017).
6. Coates, G. W. & Getzler, Y. D. Y. L. Chemical recycling to monomer for an ideal, circular polymer economy. *Nature Reviews Materials* **5**, 501–516 (2020).
7. Zhu, J.-B., Watson, E. M., Tang, J. & Chen, E. Y.-X. A synthetic polymer system with repeatable chemical recyclability. *Science* **360**, 398–403 (2018).
8. Tang, X. & Chen, E. Y.-X. Toward Infinitely Recyclable Plastics Derived from Renewable Cyclic Esters. *Chem* **5**, 284–312 (2019).
9. Jia, X., Qin, C., Friedberger, T., Guan, Z. & Huang, Z. Efficient and selective degradation of polyethylenes into liquid fuels and waxes under mild conditions. *Sci Adv* **2**, e1501591 (2016).
10. Gross, R. A. & Kalra, B. Biodegradable Polymers for the Environment. *Science* **297**, 803–807 (2002).
11. Iwata, T. Biodegradable and Bio-Based Polymers: Future Prospects of Eco-Friendly Plastics. *Angewandte Chemie International Edition* **54**, 3210–3215 (2015).
12. https://ourworldindata.org/plastic-pollution.
13. Harvey, M. & Pilgrim, S. The new competition for land: Food, energy, and climate change. *Food Policy* **36**, S40–S51 (2011).
14. Sardon, H. & Dove, A. P. Plastics recycling with a difference. *Science* **360**, 380–381 (2018).
15. Lutz, J.-F., Ouchi, M., Liu, D. R. & Sawamoto, M. Sequence-Controlled Polymers. *Science* **341**, 1238149 (2013).
16. OECD. Improving Plastics Management. (2018).
17. Garcia, J. M. & Robertson, M. L. The future of plastics recycling. *Science* **358**, 870–872 (2017).
18. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001).

19. Teixeira, P. F. *et al.* A multi-step peptidolytic cascade for amino acid recovery in chloroplasts. *Nature Chemical Biology* **13**, 15–17 (2016).
20. Bolisetty, S. & Mezzenga, R. Amyloid–carbon hybrid membranes for universal water purification. *Nature Nanotechnology* **11**, 365–371 (2016).
21. Ke, P. C. *et al.* Half a century of amyloids: past, present and future. *Chemical Society Reviews* **49**, 5473–5509 (2020).
22. Omenetto, F. G. & Kaplan, D. L. New Opportunities for an Ancient Material. *Science* **329**, 528–531 (2010).
23. Tao, H., Kaplan, D. L. & Omenetto, F. G. Silk Materials - A Road to Sustainable High Technology. *Advanced Materials* **24**, 2824–2837 (2012).
24. Josephson, K., Hartman, M. C. T. & Szostak, J. W. Ribosomal Synthesis of Unnatural Peptides. *Journal of the American Chemical Society* **127**, 11727–11735 (2005).
25. Hartman, M. C. T., Josephson, K. & Szostak, J. W. Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proceedings of the National Academy of Sciences* **103**, 4356–4361 (2006).
26. Lewandowski, B. *et al.* Sequence-Specific Peptide Synthesis by an Artificial Small-Molecule Machine. *Science* **339**, 189–193 (2013).
27. Niu, J., Hili, R. & Liu, D. R. Enzyme-free translation of DNA into sequence-defined synthetic polymers structurally unrelated to nucleic acids. *Nature Chemistry* **5**, 282–293 (2013).
28. Roy, R. K. *et al.* Design and synthesis of digitally encoded polymers that can be decoded and erased. *Nature Communications* **6**, 7237 (2015).
29. Dong, R. *et al.* Sequence-defined multifunctional polyethers via liquid-phase synthesis with molecular sieving. *Nature Chemistry* **11**, 136–145 (2019).
30. Zydziak, N. *et al.* Coding and decoding libraries of sequence-defined functional copolymers synthesized via photoligation. *Nature Communications* **7**, 13672 (2016).
31. Meier, M. A. R. & Barner-Kowollik, C. A New Class of Materials: Sequence-Defined Macromolecules and Their Emerging Applications. *Advanced Materials* **31**, 1806027 (2019).
32. Barnes, J. C. *et al.* Iterative exponential growth of stereo- and sequence-controlled polymers. *Nature Chemistry* **7**, 810–815 (2015).
33. Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature Biotechnology* **25**, 770–777 (2007).
34. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).

Acknowledgements

The research has been supported by the Swiss National Science Foundation (SNF) NCCR Bio-Inspired Materials, by the SNF Spark project CRSK-2_190167, and by the ERC Advanced Grant (884114-NaCRe). A.S received funding by the French National Research Agency (ANR) through the Programme d'Investissement d'Avenir under contract 17-EURE-0016.
We thank the Protein production and structure Core Facility (EPFL), and the Proteomics Core Facility (EPFL) for helping us to express the calibrants, and to perform the proteomic analyses respectively.
We thank Prof. Correia, Dr. Laohakunakorn, Dr. Swank, Dr. Olson, Dr. Istomin, Grégoire Michielin, and Matteo Gasbarri for useful discussions.

Author contributions

F.S. first conceived the work. S.G., F.S., S.J.M. designed the experiments, and discussed the results. S.G., A.M.S., and L.R.J. performed the experiments under the supervision of F.S. and S.J.M. S.G., D.O., and L.M. performed the MS characterizations, analyzed the data, and discussed the results. A.M. performed the AFM characterization. L.P. adapted the software for analyzing the MS data. A.M.,

and S.G. wrote the script for selecting the peptides. S.B., and S.G. prepared the amyloid film under the supervision of R.M. All the authors contributed to the writing of the paper.

Additional information

Supplementary information is available in the online version of the paper.

Competing financial interests

The authors declare no competing financial interests.

Supplementary Information

Nature-inspired Circular-economy Recycling (NaCRe) for Proteins: Proof of Concept

Simone Giaveri, Adeline M. Schmitt, Laura Roset Julià, Anna Murello, Laure Menin, Daniel Ortiz, Luc Patiny, Sreenath Bolisetty, Raffaele Mezzenga, Sebastian J. Maerkl, Francesco Stellacci

Table of Contents

Supplementary Methods (a. - n.)	Page 3
a. Materials	Page 3
b. Calibrant expression	Page 4
c. Film preparation	Page 5
d. Selection of the model peptides	Page 5
e. Depolymerization I (cleavage)	Page 8
f. Depolymerization II (depolymerization)	Page 9
g. Polymerase Chain Reaction (PCR)	Page 9
h. CF protein TX-TL	Page 10
i. Peptide (and UAAs) analysis by mass spectrometry	Page 12
j. Amino Acid Analysis by mass spectrometry (AAA)	Page 13
k. Protein electrophoresis (SDS-PAGE)	Page 14
l. Proteomic analysis	Page 14
m. Mass calibration	Page 15
n. AFM imaging	Page 16
Supplementary References	Page 16
Supplementary Table S1	Page 17
Supplementary Table S2	Page 18
Supplementary Table S3	Page 19
Supplementary Figure S1	Page 20
Supplementary Figure S2	Page 20
Supplementary Figure S3	Page 21
Supplementary Figure S4	Page 21
Supplementary Figure S5	Page 22
Supplementary Figure S6	Page 23
Supplementary Figure S7	Page 23
Supplementary Figure S8	Page 24
Supplementary Figure S9	Page 24
Supplementary Figure S10	Page 25
Supplementary Figure S11	Page 25
Supplementary Figure S12	Page 26
Supplementary Figure S13	Page 26
Supplementary Figure S14	Page 27
Supplementary Figure S15	Page 27
Supplementary Figure S16	Page 28

Supplementary Figure S17 Page 28
Supplementary Figure S18 Page 29
Supplementary Figure S19 Page 29
Supplementary Figure S20 Page 30
Supplementary Figure S21 Page 30
Supplementary Figure S22 Page 31
Supplementary Figure S23 Page 31
Supplementary Figure S24 Page 32
Supplementary Figure S25 Page 32
Supplementary Figure S26 Page 33
Supplementary Figure S27 Page 33
Supplementary Figure S28 Page 34
Supplementary Figure S29 Page 34
Supplementary Figure S30 Page 35
Supplementary Figure S31 Page 35
Supplementary Figure S32 Page 36
Supplementary Figure S33 Page 36
Supplementary Figure S34 Page 37
Supplementary Figure S35 Page 37
Supplementary Figure S36 Page 38
Supplementary Figure S37 Page 38
Supplementary Figure S38 Page 39
Supplementary Figure S39 Page 39
Supplementary Figure S40 Page 40
Supplementary Figure S41 Page 40
Supplementary Figure S42 Page 41
Supplementary Figure S43 Page 41
Supplementary Figure S44 Page 42
Supplementary Figure S45 Page 42
Supplementary Figure S46 Page 43
Supplementary Figure S47 Page 43
Supplementary Figure S48 Page 43
Supplementary Figure S49 Page 44
Supplementary Figure S50 Page 44
Supplementary Figure S51 Page 45
Supplementary Figure S52 Page 45

Supplementary Methods

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, Benzonase, 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. 5x Phusion HF Buffer, dNTP Mix (10 mM), Phusion High-Fidelity DNA Polymerase (2 U μl^{-1}), and DMSO were purchased from Thermo Fisher Scientific; nuclease-free water was supplied by Sigma-Aldrich. 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. UltraPure Agarose was supplied by Invitrogen. 50x TAE buffer was purchased from Jena Bioscience. *Cell-Free expression.* Magnesium acetate, Potassium glutamate, DL-Dithiothreitol (DTT), Creatine phosphate, Folinic acid, Spermidine, HEPES buffer, 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. PURExfrex™ Solution II (enzymes), and PURExfrex™ 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) was 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. *Filters-membranes-tools.* Amicon Ultra-0.5 ml Centrifugal Filters (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 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. 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 construct was synthesized (as codon-optimized) for expression in *E. coli*, appended with a 6xHis tag at C-terminus, and cloned into pET29b(+) vector. mScarlet-i construct is reported in Supplementary Table S2. 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 pellet was 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 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 protein was 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 protein 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 mg/ml) concentration approximately by using Amicon Ultra Centrifugal Filters (10K). *Quantification.* mScarlet-i was quantified by 280 nm absorbance, and its predicted extinction coefficients (https://web.expasy.org/protparam/).

c. Film preparation

Filtration. The film was fabricated by vacuum filtration of the β-lactoglobulin amyloids solution (20 mg ml^{-1}) using a vacuum filtration assembly, and MCE membranes (pore size = 0.22 μm, diameter = 25 mm), following the protocol developed in Mezzenga's lab[1]. *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.

d. 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 an output .txt file containing all the filtered chains. Bio, re, sys, os, joblib, multiprocessing, operator, and warnings are the Python Modules required.

```
#Written by Anna Murello and Simone Giaveri (SuNMIL), EPFL
#January 2019

import Bio
from Bio.PDB import *
from Bio.Align import MultipleSeqAlignment
from distutils import spawn
import re
from Bio.SeqUtils.ProtParam import ProteinAnalysis
from Bio.PDB.PDBParser import PDBParser
from Bio.PDB.Polypeptide import three_to_one
from Bio.PDB.Polypeptide import is_aa
from Bio import Alphabet
from Bio.Data import IUPACData
from Bio.Data.SCOPData import protein_letters_3to1
import sys
from Bio import SeqIO
from Bio.PDB import PDBList
import os
import warnings
from Bio import BiopythonWarning
warnings.simplefilter('ignore', BiopythonWarning)
from joblib import Parallel, delayed
```

```
import multiprocessing
from operator import itemgetter
from Bio.SeqRecord import SeqRecord
from Bio.Seq import Seq

def Analysis (pdbfilename, l_max, l_min, aa1, aa2, count_aa1, count_aa2):
# this function reads a file (pdbfilename) and checks whether the filtering conditions are verified; if so
#it returns the sequence
    Record = []

    print(pdbfilename[:-4])

    pdbfile = os.path.join(pdbdir, pdbfilename)

    try:

        parser = MMCIFParser()
        structure = parser.get_structure(",pdbfile)

        model = structure [0] # the structure file may contain more than one model, the program
#analyses only the first one
        if len(model)==1:

            chain = model['A']
            seq = list()

            for residue in chain: # checking the filtering conditions

                if is_aa(residue.get_resname(), standard=True):
                    seq.append(three_to_one(residue.get_resname()))

                elif is_aa(residue.get_resname(), standard=False):
                    seq.append('X')

                elif residue.get_resname()=='PYL' or residue.get_resname()=='XLE':
                    seq.append('X')

            myprot = str(".join(seq))

            length = len(myprot)


            if length < l_max and length > l_min and myprot.count(aa1) < count_aa1 and
myprot.count(aa2) != length \
                and myprot.count(aa2) < count_aa2:

                analysis = ProteinAnalysis(myprot)
                additional_features = {"length": length, "count_aa" : analysis.count_amino_acids(),
"count_X" : myprot.count('X')}

                Record = [myprot, pdbfilename[:-4], additional_features]

    except:

        pass
```

```
    return(Record)

#-----------------------------------------------------------------------------------------------

with open("Output.txt", "w") as text_file:
    pdbdir = '/Users/simonegiaveri/Desktop/PythonPDB/Common_folder' # path for the directory
#containing the structure files

    # in the following lines 'PYL' and 'XLE' amino acids are added to the dictionary of the non-natural
#amino acids.
    protein_letters_3to1['PYL']='X'
    protein_letters_3to1['XLE']='X'

    # in the following lines the filtering conditions are defined
    pdbfilenames = os.listdir(pdbdir)
    l_max = 50
    l_min = 5
    aa1 = 'C'
    aa2 = 'X'
    count_aa1 = 10
    count_aa2 = 1

    results = Parallel(n_jobs=4)(delayed(Analysis)(pdbfilename, l_max, l_min, aa1, aa2, count_aa1,
count_aa2) for pdbfilename in pdbfilenames)

    counter = 0

    for a in results:

        if a != []:
            print(a[1],a[0],a[2], file=text_file)
            print('\n', file=text_file)

            counter = counter + 1

    print ('Number of filtered proteins = ', counter, file=text_file)

    print('Total number of proteins = ',len(pdbfilenames), file=text_file)

    print('Filtered as:', l_min, '< sequence length <', l_max, 'and number of ', aa1, '<', count_aa1,'and
number of ', aa2, '<', count_aa2, \
        ' and different from sequence length', file=text_file)
```

e. Depolymerization I (cleavage)

Enzymes preparation. Thermolysin was dissolved in buffer (Tris-HCl (50 mM), CaCl2 (1 mM), KOH, pH 8) at 1 mg ml^{-1} concentration; pepsin was reconstituted in water-HCl solution at 1.5 mg ml^{-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 mg ml^{-1} concentration. Glucagon was prepared in (500 µl) buffer (Tris-HCl (50 mM), CaCl2 (1 mM), KOH, pH 9), at 1 mg ml^{-1} concentration. β-lactoglobulin A was prepared in (500 µl) buffer (Tris-HCl (50 mM), CaCl2 (1 mM), KOH, pH 8), at 2 mg ml^{-1} concentration. Silk fibroin solution (50 mg ml^{-1}) was diluted in (500 µl) buffer (Tris-HCl (50 mM), CaCl2 (1 mM), KOH, pH 8), at 2 mg ml^{-1} concentration. β-lactoglobulin film powder was resuspended in (500 µl) water-HCl solution (pH 2.7), at 1.5 mg ml^{-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 mg ml^{-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 mg ml^{-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-hydroxynorvaline][DL-3-hydroxynorvaline][L-norleucine][Ser][Lys] was prepared in (500 µl) buffer (Tris-HCl (50 mM), CaCl2 (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 6h 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 8h 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.

f. Depolymerization II (depolymerization)

Enzymes preparation. Leucine aminopeptidase LAP (L9776 and L6007) were resuspended in nuclease-free water at 1 mg/ml 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.

g. Polymerase Chain Reaction (PCR)

PCR batch (20 µl). The reaction was assembled mixing 1 µl DNA linear gBlock template (1 ng μl^{-1}), 0.2µl 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 U μl^{-1}), and 13.05 µl nuclease-free water in a small PCR vial. *PCR thermal cycle.* 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. *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, $\lambda = 302$ nm, 4s exposure. The gel is shown in Supplementary Fig. S51. *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 µl of elution buffer per spin column. Q*uantification.* The final DNA concentration was measured using Witec NanoDrop 1000 spectrophotometer.

h. 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 µl): 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 µl 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 coworkers[2]: 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}). *Cell-Free TX-TL reactions assembly (25 µl).* 3.4 µl of solutionA(-Salts - tRNAs - AAs), 1.6 µl of salts solution, 2.5 µl of tRNAs solution, 1.25 µl PURExfrex™ Solution II (enzymes), 1.25 µl PURExfrex™ Solution III (ribosomes), 0.5 µl RNAse inhibitor, 75 ng DNA, and 10 µl of AAs were mixed in ice. Nuclease-free water was added to bring the reaction volume to 25 µl. *Cell-Free TX-TL reactions assembly (25 µl) containing non-natural residues.* The reaction volume was supplemented with up to 12.5 µl of AAs, and UAAs. *Magainin II, glucagon, and somatostatin 28 recycling into mScarlet-i. (Samples).* 10 µl AAs solution was obtained by combining equal volumes (3.33 µl) of magainin II, glucagon, and somatostatin 28 depolymerization solutions. *(Negative controls).* 10 µl AAs solution was obtained by combining equal volumes (3.33 µl) of magainin II, glucagon, and somatostatin 28 depolymerization solutions, prepared without adding the three peptides initially. *(Reference controls).* 10 µl AAs solution was obtained by combining equal volumes (3.33 µl) 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 µl 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 µl AAs solution was obtained by the depolymerization of the magainin II, glucagon, and somatostatin 28 mixture. *(Negative controls).* 10 µl 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 µl 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 µl 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 µl AAs solution was obtained by the depolymerization of β-lactoglobulin A. *(Negative controls).* 10 µl AAs solution was obtained by preparing a depolymerization reaction of β-lactoglobulin A, without adding β-lactoglobulin A initially. *(Reference controls).* 10 µl 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 µl 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 µl AAs solution was obtained by the depolymerization of silk fibroin solution. *(Negative controls).* 10 µl AAs solution was obtained by preparing a depolymerization reaction of silk fibroin, without adding silk fibroin initially. *(Reference controls).* 10 µl 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 µl of L-cysteine hydrochloride in nuclease-free water solution (15 mM), and 0.5 µl 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 µl AAs solution was obtained by the depolymerization of β-lactoglobulin film. *(Negative controls).* 10 µl AAs solution was obtained by preparing a depolymerization reaction of β-lactoglobulin film, without adding β-lactoglobulin film powder initially. *(Spikes).* 0.5 µl of L-cysteine hydrochloride in nuclease-free water solution (15 mM), 0.5 µl of L-methionine in nuclease-free water solution (15 mM), and 0.5 µl 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 µl AAs and UAAs solution was obtained by combining 9 µl of the unnatural peptide depolymerization, 0.5 µl of L-valine in nuclease-free water solution (15 mM), 0.5 µl of L-threonine in nuclease-free water solution (15 mM), 0.5 µl of L-leucine in nuclease-free water solution (15 mM), 0.5 µl of L-lysine hydrochloride in nuclease-free water solution (15 mM), and 1.5 µl 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 µl of FluoroTect^{TM} Green$_{Lys}$ tRNA. *(Negative controls).* 9 µl of the unnatural peptide depolymerization were replaced by nuclease-free water. *(Reference controls).* 4.3 µl AAs and UAAs solution was obtained by combining 0.4 µl of L-canavanine in nuclease-free water solution (25 mM), 0.4 µl of L-norleucine in nuclease-free water solution (25 mM), 0.5 µl of L-valine in nuclease-free water solution (15 mM), 0.5 µl of L-threonine in nuclease-free water solution (15 mM), 0.5 µl of L-serine in nuclease-free water solution (15 mM), 0.5 µl of L-lysine hydrochloride in nuclease-free water solution (15 mM), and 1.5 µl 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 µl of FluoroTect^{TM} Green$_{Lys}$ tRNA. *Mixed glucagon, β-lactoglobulin A, and silk*

fibroin recycling into GFP. (Samples). 10 µl AAs solution was obtained by the depolymerization of the mixture composed of glucagon, β-lactoglobulin A, and silk fibroin. *(Negative controls).* 10 µl 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 µl AAs solution was obtained by preparing a free AAs solution, calculated from an ideal complete depolymerization of the initial protein mixture. *Cell-Free TX-TL reaction.* The reactions 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 6 h (mScarlet-i), and 4 h (GFP) in Thermo Fisher Scientific BioTek Synergy Mx plate reader. The plate reader parameters were the following: 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. *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. *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.

i. 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 5µl. The experimental conditions for the ionization voltage were +1.4kV 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 apm^{2}S software (https://ms.epfl.ch/applications/peptides_and_proteins/)[3,4].

j. 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 InfinityLab Poroshell 120 HILIC-Z column heated at 25° C. A binary gradient system consisted of A (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. Data from quantifications are shown in Supplementary Data x. *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).

k. Protein electrophoresis (SDS-PAGE). *Samples preparation.* Cell-free expressions were gently defrosted in ice, and aliquoted (10 µl). 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. Denaturized samples were loaded on 4-12% Bis-Tris mini protein gel. *Running conditions.* The gel was 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.* The gel was washed in Milli-Q water for 1 h shaking prior to Coomassie staining for 1 h by using InstantBlue stain. The gel was destained in Milli-Q water for 1 h shaking. *Imaging.* The gel was imaged by using Vilber Lourmat Fusion Fx Imaging System, λ = AlexaFluor 488 nm, 3 s exposure, and a photo of it was taken by using iPhone Xs. The gel is shown in Supplementary Fig. S52.

l. 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 at a concentration of 12.5 ng/µl 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:nuclease-free 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 ZipTips. Two steps elution was performed using first 80/19.9/0.1 v/v/v ACN:nuclease-free water:TFA then 80/10/9.9/0.1 v/v/v/v ACN:TFE:nuclease-free 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 nuclease-free water:ACN:TFA solution and nano-flow separations were performed on a Dionex Ultimate 3000 RSLC nano UPLC system (Thermo Fischer Scientific) on-line connected with an Exploris 480 Orbitrap Mass Spectrometer (Thermo Fischer Scientific). A capillary precolumn (Acclaim Pepmap C18, 3 µm-100Å, 2 cm x 75µm ID) was used for 8 µl sample trapping and cleaning. A 50cm 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 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 second. 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 200m/z) and selected ions

were then excluded for the following 30 s. The experimental conditions for the ionization voltage were +1.6kV; the temperature of the ion transfer capillary was set to 175 °C. *Data processing (natural GFP).* 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 sequence. Enzyme specificity was set to Chymotrypsin, 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. 62 exclusive unique peptides were identified, with 94 % sequence coverage. *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 apm^{2}S software (https://ms.epfl.ch/applications/peptides_and_proteins/)[3,4].

m. Mass calibration

A calibration curve for mScarlet-i expression is reported in Supplementary Fig. S7. *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 PURExfrex™ Solution II (enzymes), 1.25 µl PURExfrex™ 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: λ_{exc} = 569 nm (mScarlet-i), λ_{em} = 593 nm (mScarlet-i), λ_{em} = 507 nm (GFP), 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 PURExpress™ 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 Supplementary Fig. S6.

n. AFM imaging

Sample preparation. Solutions of as prepared fibrils and depolymerized fibrils (~ 0.2 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 Supplementary Fig. S50.

Supplementary References

1. Li, C., Adamcik, J. & Mezzenga, R. Biodegradable nanocomposites of amyloid fibrils and graphene with shape-memory and enzyme-sensing properties. *Nature Nanotechnology* **7**, 421–427 (2012).
2. Shimizu, Y., Kanamori, T. & Ueda, T. Protein synthesis by pure translation systems. *Methods* **36**, 299–304 (2005).
3. Patiny, L. & Borel, A. ChemCalc: A Building Block for Tomorrow's Chemical Infrastructure. *Journal of Chemical Information and Modeling* **53**, 1223–1228 (2013).
4. Lee, R. F. S., Menin, L., Patiny, L., Ortiz, D. & Dyson, P. J. Versatile Tool for the Analysis of Metal–Protein Interactions Reveals the Promiscuity of Metallodrug–Protein Interactions. *Analytical Chemistry* **89**, 11985–11989 (2017).

a)	**HGlyIleGlyLysPheLeuHisSerAlaLysLysPheGlyLysAlaPheValGlyGluIleMetAsnSerOH**
b)	**HHisSerGlnGlyThrPheThrSerAspTyrSerLysTyrLeuAspSerArgArgAlaGlnAspPheValGlnTrpLeuMet AsnThrOH**
c)	**HSerAlaAsnSerAsnProAlaMetAlaProArgGluArgLysAlaGlyCysLysAsnPhePheTrpLysThrPheThrSer CysOH**
d)	**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**
e)	**HMetLysCysLeuLeuLeuAlaLeuAlaLeuThrCysGlyAlaGlnAlaLeuIleValThrGlnThrMetLysGlyLeuAspIle GlnLysValAlaGlyThrTrpTyrSerLeuAlaMetAlaAlaSerAspIleSerLeuLeuAspAlaGlnSerAlaProLeuArg ValTyrValGluGluLeuLysProThrProGluGlyAspLeuGluIleLeuLeuGlnLysTrpGluAsnAspGluCysAlaGln LysLysIleIleAlaGluLysThrLysIleProAlaValPheLysIleAspAlaLeuAsnGluAsnLysValLeuValLeuAspThr AspTyrLysLysTyrLeuLeuPheCysMetGluAsnSerAlaGluProGluGlnSerLeuValCysGlnCysLeuValArgThr ProGluValAspAspGluAlaLeuGluLysPheAspLysAlaLeuLysAlaLeuProMetHisIleArgLeuSerPheAsnPro ThrGlnLeuGluGluGlnCysHisIleOH**
f)	**HSerValThrIleAsnGlnTyrSerAspAsnGluIleProArgAspIleAspAspGlyLysAlaSerSerValIleSerArgAla TrpAspTyrValAspAspThrAspLysSerIleAlaIleLeuAsnValGlnGluIleLeuLysAspMetAlaSerGlnGlyAsp TyrAlaSerGlnAlaSerAlaValAlaGlnThrAlaGlyIleIleAlaHisLeuSerAlaGlyIleProGlyAspAlaCysAlaAla AlaAsnValIleAsnSerTyrThrAspGlyValArgSerGlyAsnPheAlaGlyPheArgGlnSerLeuGlyProPhePheGly HisValGlyGlnAsnLeuAsnLeuIleAsnGlnLeuValIleAsnProGlyGlnLeuArgTyrSerValGlyProAlaLeuGly CysAlaGlyGlyGlyArgIleTyrAspPheGluAlaAlaTrpAspAlaIleLeuAlaSerSerAspSerSerPheLeuAsnGlu GluTyrCysIleValLysArgLeuTyrAsnSerArgAsnSerGlnSerAsnAsnIleAlaAlaTyrIleThrAlaHisLeuLeuPro ProValAlaGlnValPheHisGlnSerAlaGlySerIleThrAspLeuLeuArgGlyValGlyAsnGlyAsnAspAlaThrGly LeuValAlaAsnAlaGlnArgTyrIleAlaGlnAlaAlaSerGlnValHisValOH**
g)	**HAsnIleAsnAspPheAspGluAspTyrPhe ... LysPheArgAlaLeuProCysValAsnCysOH**

Supplementary Table S1. Primary sequences of the proteins we depolymerized: 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.

a)	(5′)GCACCATCAGCCAGAAAACCGAACCAGCCAGAAAACGACCTTTCTGTGGATCTTAAGGCTAGAGTACT AATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTAAGAAGGAGG AAAAAAAAATGGTAAGTAAAGGTGAAGCGGTAATTAAAGAGTTTATGCGCTTTAAGGTACACATGGAAG GCTCTATGAATGGGCACGAATTCGAGATCGAAGGTGAGGGTGAAGGACGCCCCTACGAGGGCACTCAGA CTGCAAAGTTAAAAGTGACGAAAGGTGGCCCCTTACCGTTTAGCTGGGATATCCTGTCGCCGCAGTTTATG TATGGCAGTCGTGCGTTCATCAAGCACCCTGCTGACATCCCTGATTACTATAAGCAATCATTCCCCGAAGGC TTTAAGTGGGAGCGTGTTATGAACTTTGAAGATGGCGGAGCTGTGACTGTTACACAAGACACGAGCTTGG AAGACGGAACCCTGATCTACAAAGTGAAATTACGCGGTACGAACTTTCCGCCTGACGGTCCAGTGATGCA GAAAAAGACCATGGGATGGGAAGCTAGCACCGAACGTTTATATCCGGAGGACGGCGTGCTTAAAGGTGA CATTAAAATGGCATTACGTTTGAAGGATGGTGGTCGTTATTTAGCAGATTTTAAGACTACGTATAAAGCGA AGAAGCCTGTACAAATGCCGGGTGCTTATAACGTAGATCGTAAACTTGATATTACCTCGCACAATGAAGAC TATACAGTAGTAGAACAATACGAACGCTCCGAGGGACGCCACTCTACTGGGGGCATGGACGAATTATACA AATCTGGATTACGCAGCCGTGCCCAGGCCAGTAATTCCGCAGTGGACGGTACGGCTGGACCGGGGTCTAC AGGTAGTCGTCACCACCATCATCACCACTAATAACGACTCAGGCTGCTACCTAGCATAACCCCTTGGGGCCT CTAAACGGGTCTTGAGGGGTTTTTTGGCAGGAAAGAACATGTGAGCAAAAGG(3′)
b)	(5′)GCACCATCAGCCAGAAAACCGAACCAGCCAGAAAACGACCTTTCTGTGGATCTTAAGGCTAGAGTACT AATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTAAGAAGGAGG AAAAAAAAATGGTCTCTAAAGGTGAAGAATTATTCACTGGTGTTGTCCCAATTTTGGTTGAATTAGATGGT GATGTTAATGGTCACAAATTTTCTGTCTCCGGTGAAGGTGAAGGTGATGCTACTTACGGTAAATTGACCTT AAAATTTATTTGTACTACTGGTAAATTGCCAGTTCCATGGCCAACCTTAGTCACTACTTTAACTTATGGTGT TCAATGTTTTTCTAGATACCCAGATCATATGAAACAACATGACTTTTTCAAGTCTGCCATGCCAGAAGGTTA TGTTCAAGAAAGAACTATTTTTTTCAAAGATGACGGTAACTACAAGACCAGAGCTGAAGTCAAGTTTGAA GGTGATACCTTAGTTAATAGAATCGAATTAAAAGGTATTGATTTTAAAGAAGATGGTAACATTTTAGGTCA CAAATTGGAATACAACTATAACTCTCACAATGTTTACATCATGGCTGACAAACAAAAGAATGGTATCAAAG TTAACTTCAAAATTAGACACAACATTGAAGATGGTTCTGTTCAATTAGCTGACCATTATCAACAAAATACTC CAATTGGTGATGGTCCAGTCTTGTTACCAGACAACCATTACTTATCCACTCAATCTGCCTTATCCAAAGATC CAAACGAAAAGAGAGACCACATGGTCTTGTTAGAATTTGTTACTGCTGCTGGTATTACCTTAGGTATGGAT GAATTGTACAAACACCACCATCATCACCACTAATAACGACTCAGGCTGCTACCTAGCATAACCCCTTGGGG CCTCTAAACGGGTCTTGAGGGGTTTTTTGGCAGGAAAGAACATGTGAGCAAAAGG(3′)
c)	(5′)GATCTTAAGGCTAGAGTAC(3′)
d)	(5′)CAAAAAACCCCTCAAGAC(3′)
e)	(5′)ATGGTGTCTAAGGGTGAGGCCGTCATTAAGGAATTTATGCGTTTTAAGGTGCACATGGAAGGATCGAT GAATGGTCATGAGTTCGAGATCGAGGGCGAGGGTGAGGGTCGTCCATATGAGGGCACTCAAACAGCCAA GTTAAAAGTTACCAAAGGTGGTCCGTTGCCTTTTAGCTGGGATATTCTCTCCCCACAATTCATGTACGGATC ACGCGCCTTCATTAAGCACCCCGCCGACATTCCCGACTACTACAAGCAAAGCTTTCCAGAGGGTTTTAAGT GGGAGCGTGTTATGAATTTCGAGGACGGCGGAGCGGTTACGGTAACGCAAGATACGAGTCTGGAAGACG GTACGTTAATCTATAAAGTGAAGCTGCGCGGCACAAATTTTCCGCCGGACGGTCCGGTCATGCAGAAGAA GACTATGGGCTGGGAAGCATCGACGGAACGCCTGTATCCAGAGGATGGTGTGTTGAAAGGGGACATCAA AATGGCACTGCGCTTAAAGGACGGCGGCCGCTACTTAGCAGACTTTAAGACAACGTATAAGGCCAAGAA ACCGGTGCAAATGCCTGGGGCCTATAATGTGGACCGTAAATTAGACATCACCTCACACAATGAGGACTACA CCGTGGTAGAGCAGTATGAACGCTCGGAAGGCCGTCACTCTACTGGTGGCATGGATGAACTGTATAAGTC AGGCCTTCGTAGCCGCGCGCAGGCCAGCAATAGTGCAGTTGATGGTACGGCAGGCCCTGGAAGCACAGG CTCACGTCACCATCACCACCACCACTGA(3′)

Supplementary Table S2. Linear templates (gBlocks) for expressing mScarlet-i, and GFP in PURExpressTM (a-b). T7 Promoter, T7 Terminator, Ribosome Binding Site (RBS), and Opening Reading Frame (ORF) are highlighted in blue, green, yellow, and red respectively. Primers for the PCR amplification (c-d). Construct for expressing mScarlet-i calibrant into BL21 (DE3) cells, after cloning it into the pET29b(+) vector (e).

a)	**ForMetValSerLysGlyGluAlaValIleLysGluPheMetArgPheLysValHisMetGluGlySerMetAsnGlyHisGlu PheGluIleGluGlyGluGlyGluGlyArgProTyrGluGlyThrGlnThrAlaLysLeuLysValThrLysGlyGlyProLeu ProPheSerTrpAspIleLeuSerProGlnPheMetTyrGlySerArgAlaPheIleLysHisProAlaAspIleProAspTyr TyrLysGlnSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGlyGlyAlaValThrValThrGlnAsp ThrSerLeuGluAspGlyThrLeuIleTyrLysValLysLeuArgGlyThrAsnPheProProAspGlyProValMetGlnLys LysThrMetGlyTrpGluAlaSerThrGluArgLeuTyrProGluAspGlyValLeuLysGlyAspIleLysMetAlaLeuArg LeuLysAspGlyGlyArgTyrLeuAlaAspPheLysThrThrTyrLysAlaLysLysProValGlnMetProGlyAlaTyrAsn ValAspArgLysLeuAspIleThrSerHisAsnGluAspTyrThrValValGluGlnTyrGluArgSerGluGlyArgHisSer ThrGlyGlyMetAspGluLeuTyrLysSerGlyLeuArgSerArgAlaGlnAlaSerAsnSerAlaValAspGlyThrAlaGly ProGlySerThrGlySerArgHisHisHisHisHisHisOH**
b)	**ForMetValSerLysGlyGluGluLeuPheThrGlyValValProIleLeuValGluLeuAspGlyAspValAsnGlyHisLys PheSerValSerGlyGluGlyGluGlyAspAlaThrTyrGlyLysLeuThrLeuLysPheIleCysThrThrGlyLysLeuPro ValProTrpProThrLeuValThrThrLeuThrTyrGlyValGlnCysPheSerArgTyrProAspHisMetLysGlnHisAsp PhePheLysSerAlaMetProGluGlyTyrValGlnGluArgThrIlePhePheLysAspAspGlyAsnTyrLysThrArgAla GluValLysPheGluGlyAspThrLeuValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspGlyAsnIleLeuGly HisLysLeuGluTyrAsnTyrAsnSerHisAsnValTyrIleMetAlaAspLysGlnLysAsnGlyIleLysValAsnPheLysIle ArgHisAsnIleGluAspGlySerValGlnLeuAlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeu ProAspAsnHisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHisMetValLeuLeuGlu PheValThrAlaAlaGlyIleThrLeuGlyMetAspGluLeuTyrLysHisHisHisHisHisHisOH**
c)	**ForMet(S-1CH2)ValSerLysGlyGluGluLeuPheThrGlyValValProIleLeuValGluLeuAspGlyAspValAsn GlyHisLysPheSerValSerGlyGluGlyGluGlyAspAlaThrTyrGlyLysLeuThrLeuLysPheIleCysThrThrGlyLys LeuProValProTrpProThrLeuValThrThrLeuThrTyrGlyValGlnCysPheSerArg((CH2)-1O)TyrProAspHis Met(S-1CH2)LysGlnHisAspPhePheLysSerAlaMet(S-1CH2)ProGluGlyTyrValGlnGluArg((CH2)-1O) ThrIlePhePheLysAspAspGlyAsnTyrLysThrArg((CH2)-1O)AlaGluValLysPheGluGlyAspThrLeuValAsn Arg((CH2)-1O)IleGluLeuLysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHisLysLeuGluTyrAsnTyrAsn SerHisAsnValTyrIleMet(S-1CH2)AlaAspLysGlnLysAsnGlyIleLysValAsnPheLysIleArg((CH2)-1O)His AsnIleGluAspGlySerValGlnLeuAlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAsp AsnHisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArg((CH2)-1O)AspHisMet(S-1CH2)Val LeuLeuGluPheValThrAlaAlaGlyIleThrLeuGlyMet(S-1CH2)AspGluLeuTyrLysHisHisHisHisHisHisOH**

Supplementary Table S3. Primary sequences of the proteins we expressed: mScarlet-i (a), GFP (b), and modified GFP (c).

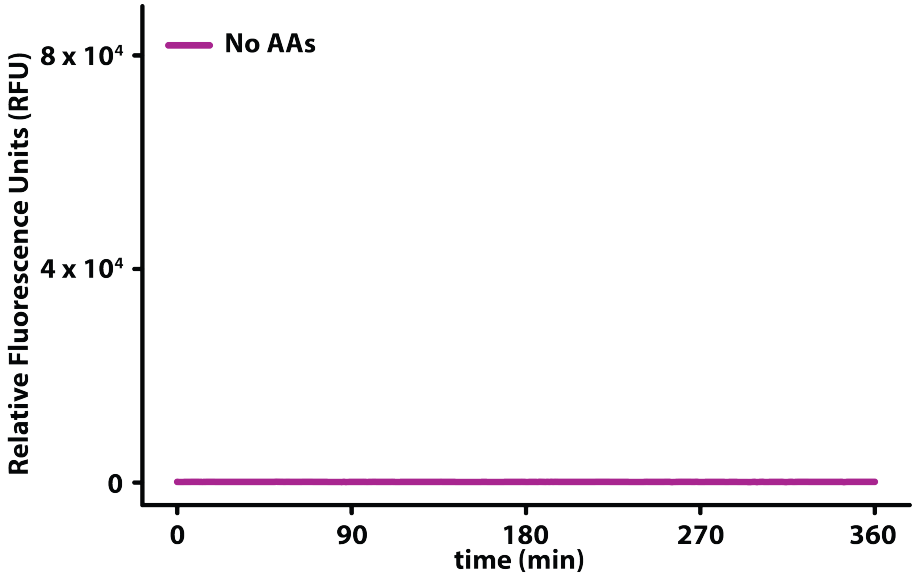


Figure S1. Plot of the fluorescence signal resulting from the expression of mScarlet-i in our TX-TL system without the addition of any AA. (The plate reader sensitivity was exceptionally set to 100 % in this experiment).

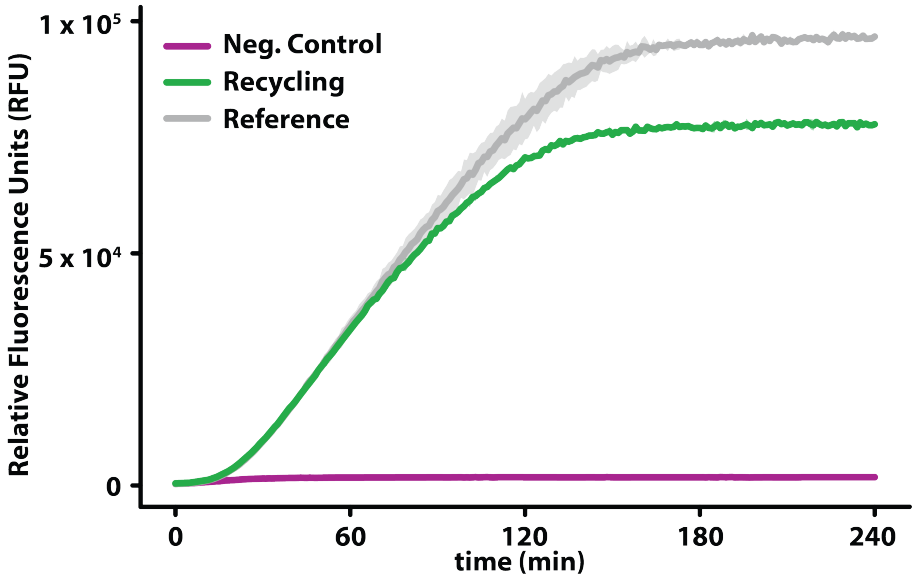


Figure S2. Plots of the fluorescence signal resulting from the expression of GFP in our TX-TL system. The green curve is obtained preforming NaCRe on magainin II, glucagon, and somatostatin 28 (Supplementary Methods h). The gray curve (reference control) is obtained as the result of an expression experiment with the TX-TL system supplemented with concentrations of AAs matching the complete depolymerization of the initial materials. 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.

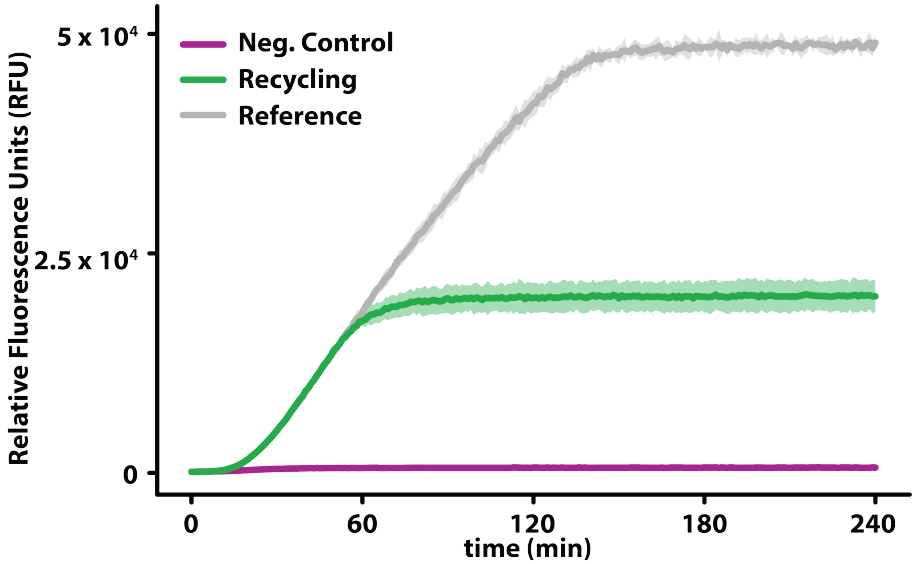


Figure S3. Plots of the fluorescence signal resulting from the expression of GFP in our TX-TL system. The green curve is obtained preforming NaCRe on β-lactoglobulin A (Supplementary Methods h). The gray curve (reference control) is obtained as the result of an expression experiment with our TX-TL system supplemented with concentrations of AAs matching the complete depolymerization of the initial material. 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.

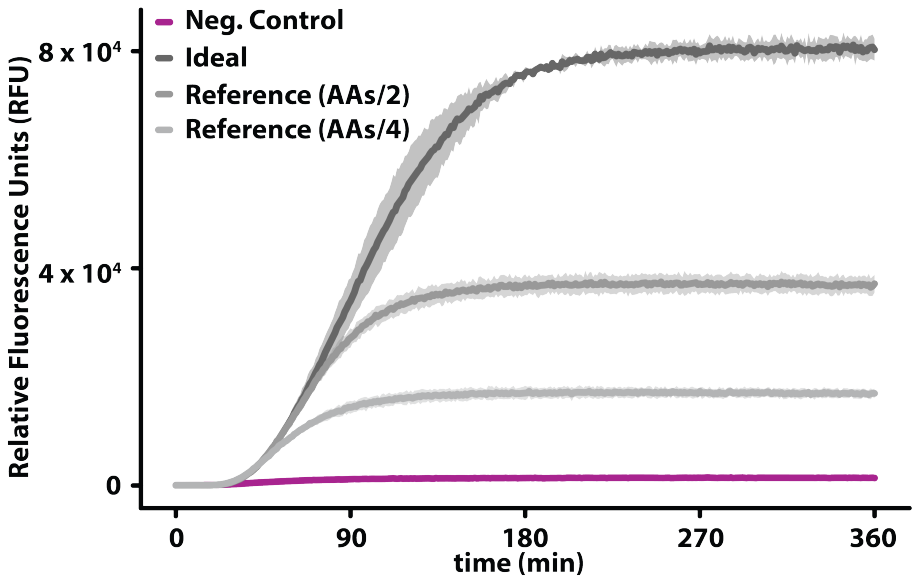


Figure S4. Plots of the fluorescence signal resulting from the expression of mScarlet-i in our TX-TL system by using 3 different initial concentrations (25/75 v/v reference AAs mixture:nuclease-free water, 50/50 v/v reference AAs mixture:nuclease-free water, and 100/0 v/v reference AAs mixture:nuclease-

free water) of the reference AAs mixture from magainin II, glucagon, and somatostatin 28 complete depolymerization.

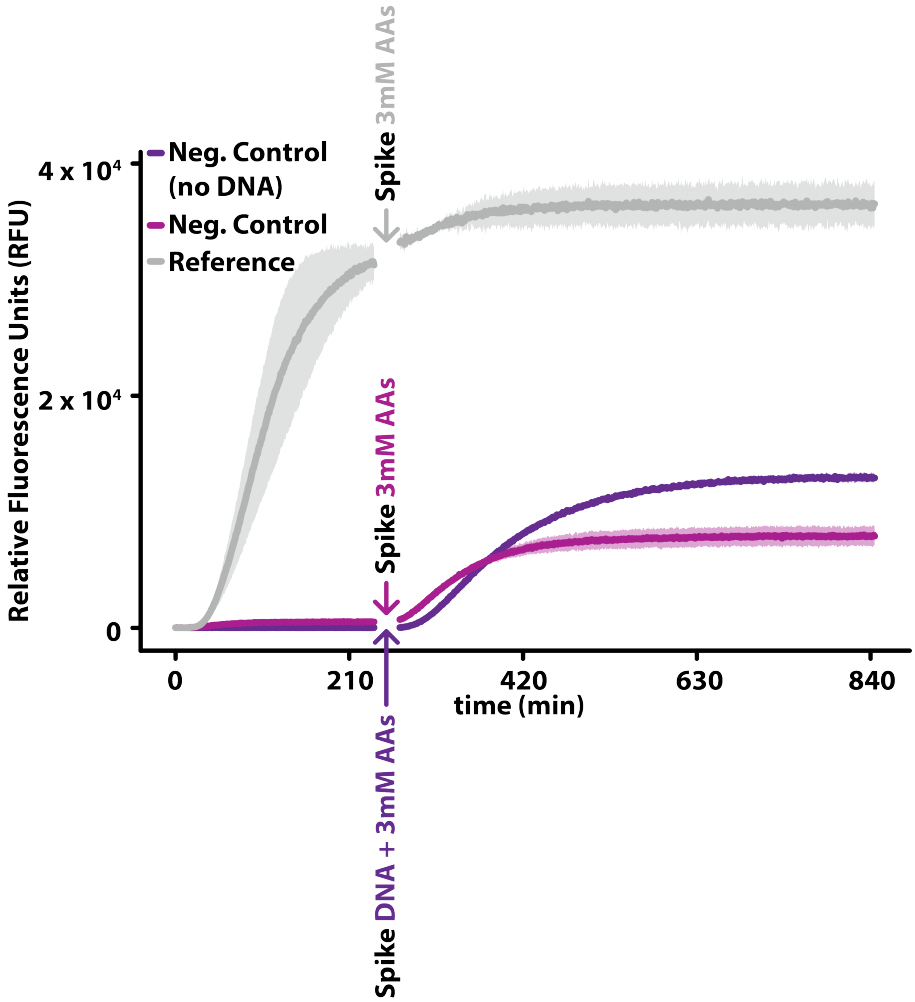


Figure S5. Plots of the fluorescence signal resulting from the expression of mScarlet-i in our TX-TL system (0 – 240 min) by using the reference AAs mixture from magainin II, glucagon, and somatostatin 28 complete depolymerization (grey curve), and substituting the reference AAs mixture with the negative 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).

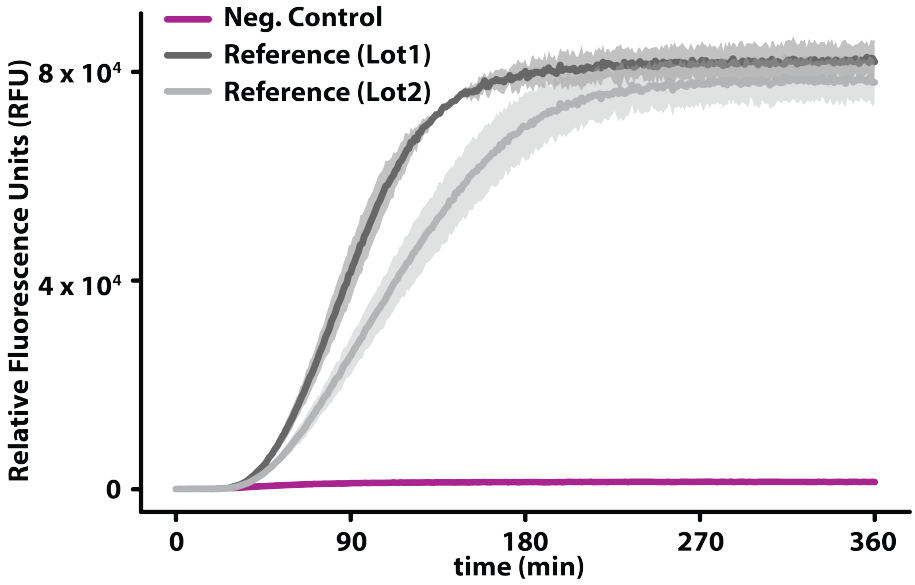


Figure S6. Plots of the fluorescence signal resulting from the expression of mScarlet-i in our TX-TL system by using the reference AAs mixture from magainin II, glucagon, and somatostatin 28 complete depolymerization. Two lots of PURExfrex™ Solution II, and III were used in order to quantify the variability of the expression plateau (RFU), as function of the PURExfrex™ Solution II, and III lots.

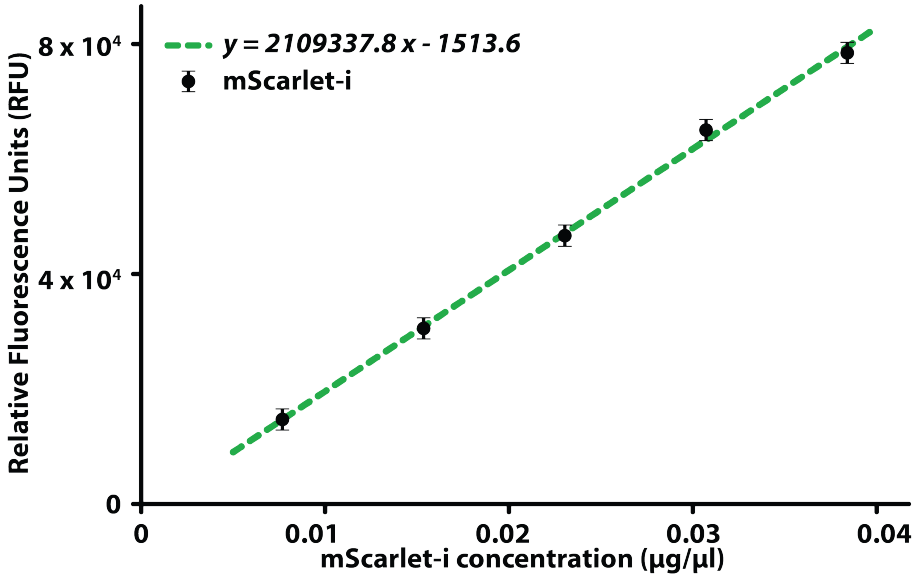


Figure S7. Plot of the mScarlet-i mass calibration curve in the plate reader (Supplementary Methods m). Error bars represent the variability of the expression by using different lots of PURE Frex™ Solution II, and III, calculated as the standard deviation of the expression plateaus (RFU) in Figure S8.

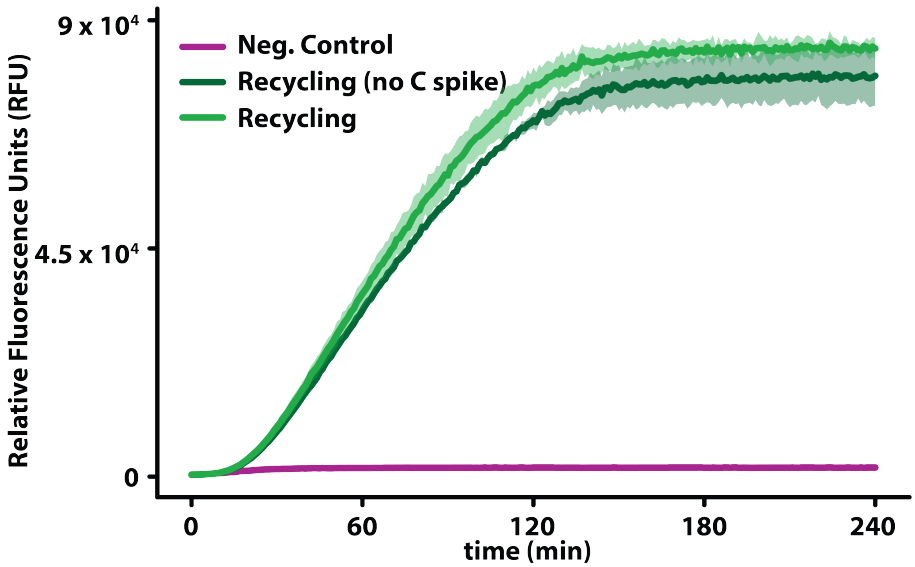


Figure S8. Plots of the fluorescence signal resulting from the expression of GFP in our TX-TL system. The green curves are data obtained preforming NaCRe on magainin II, glucagon, and somatostatin 28 with and without (0.5 µl spike of (15 mM) L-cysteine hydrochloride in nuclease-free water solution). 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.

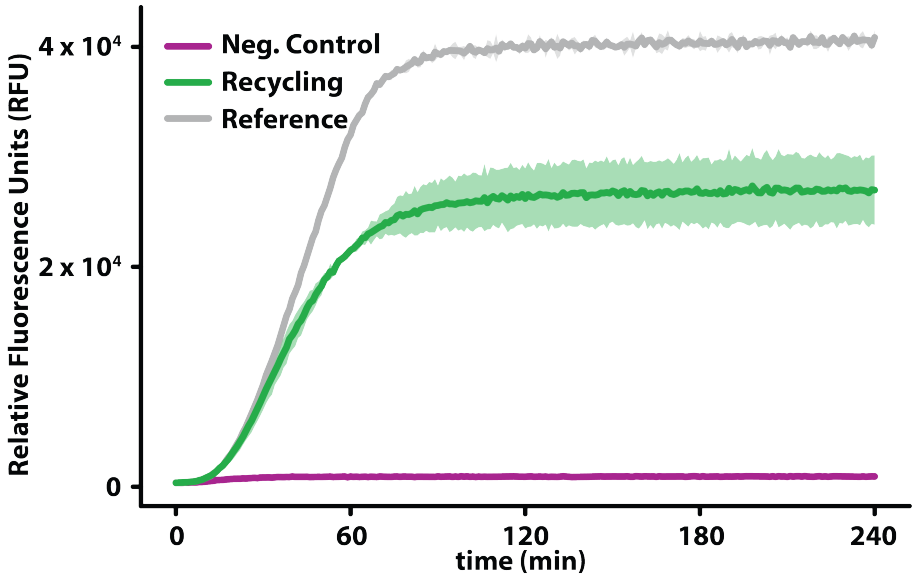


Figure S9. Plots of the fluorescence signal resulting from the expression of GFP in our TX-TL system. The green curve is obtained preforming NaCRe on the mixture composed of glucagon, β-lactoglobulin A, and silk fibroin (Supplementary Methods h). The gray curve (reference control) is the result of an 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.

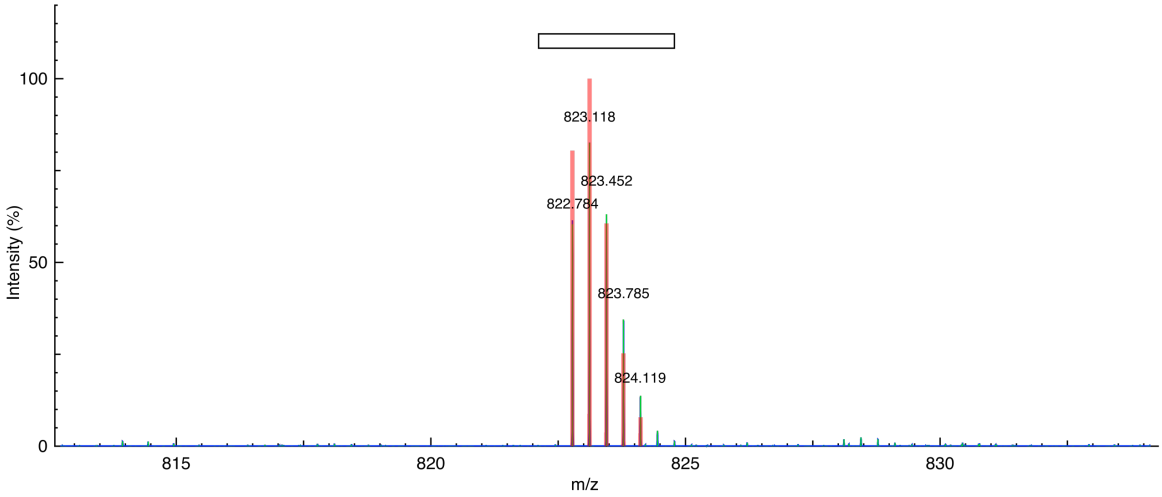


Figure S10. MS spectrum of magainin II: H3(3+)HGlyIleGlyLysPheLeuHisSerAlaLysLysPheGly LysAlaPheValGlyGluIleMetAsnSerOH. Theoretical observed mass (m/z) = 822.782; experimental closest peak (m/z) = 822.784. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

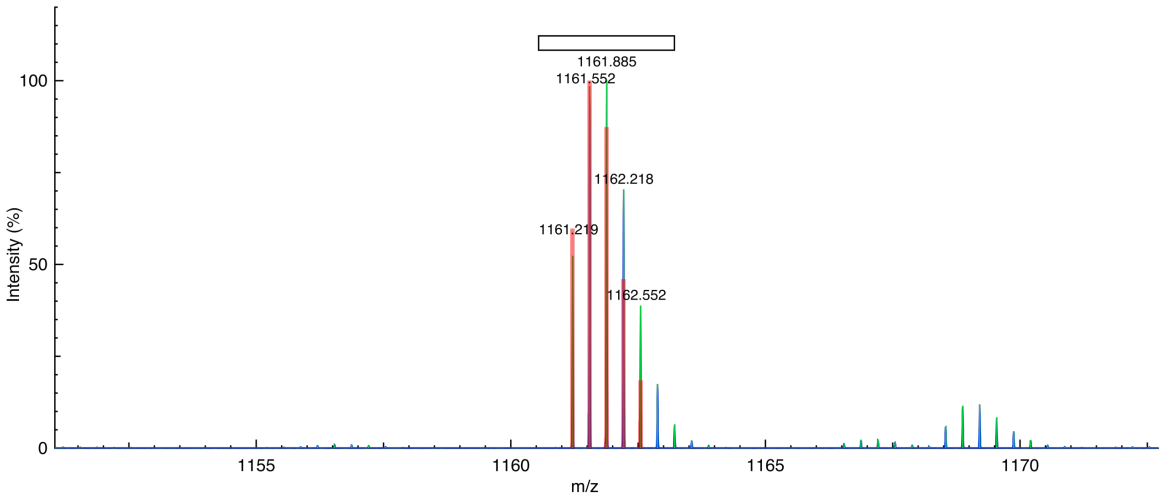


Figure S11. MS spectrum of glucagon: H3(3+)HHisSerGlnGlyThrPheThrSerAspTyrSerLysTyrLeu AspSerArgArgAlaGlnAspPheValGlnTrpLeuMetAsnThrOH. Theoretical observed mass (m/z) = 1161.213; experimental closest peak (m/z) = 1161.219. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

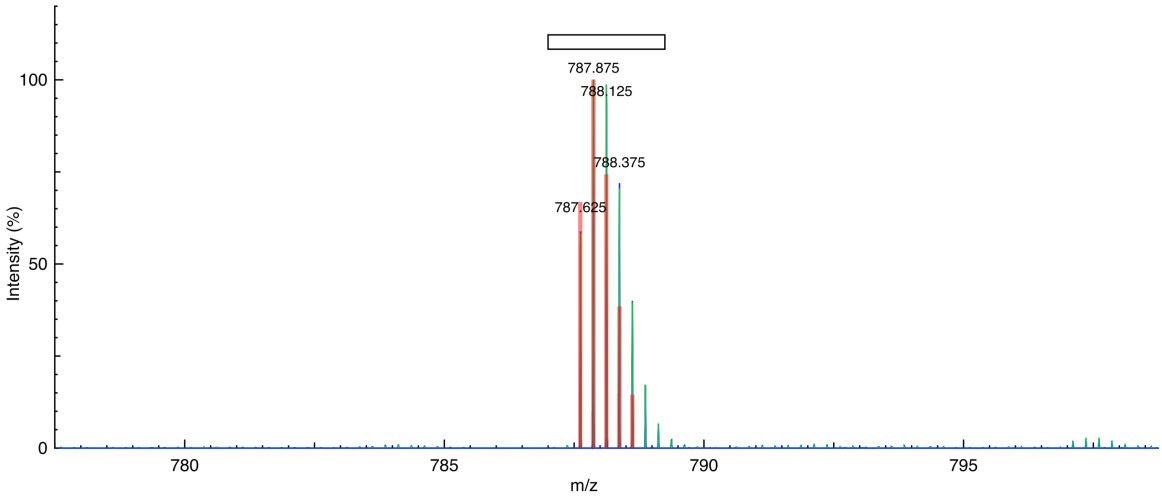


Figure S12. MS spectrum of somatostatin 28: H4(4+)HSerAlaAsnSerAsnProAlaMetAlaProArgGlu ArgLysAlaGlyCys(H-)LysAsnPhePheTrpLysThrPheThrSerCys(H-)OH. Theoretical observed mass (m/z) = 787.623; experimental closest peak (m/z) = 787.625. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

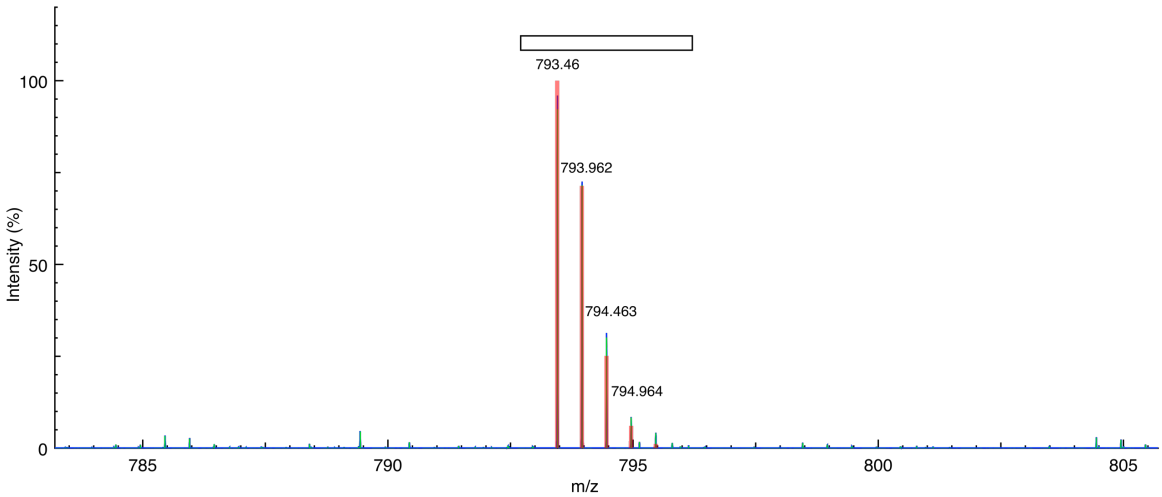


Figure S13. MS spectrum of the non-natural peptide: H2(2+)HMet(S-1CH2)Val(H-1F)Val(H-1F)Arg((CH2)-1O)Thr(CH2)Thr(CH2)Arg((CH2)-1O)Thr(CH2)Thr(CH2)Thr(CH2)Met(S-1CH2)SerLysOH. Theoretical observed mass (m/z) = 793.458; experimental closest peak (m/z) = 793.460. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

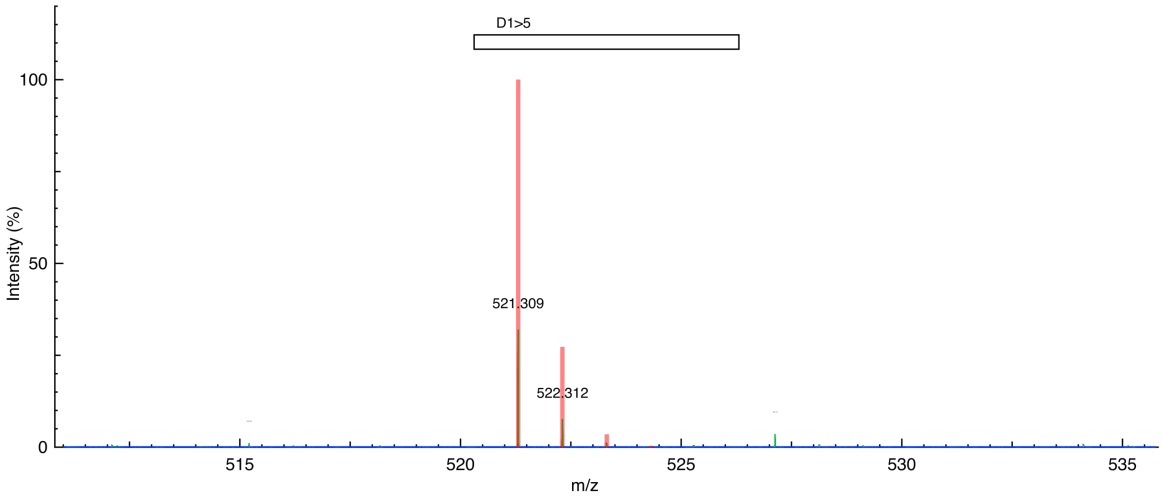


Figure S14. MS spectrum of magainin II cleaved fragment D1>5: H1(1+)HGlyIleGlyLysPheOH. Theoretical observed mass (m/z) = 521.308; experimental closest peak (m/z) = 521.309. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

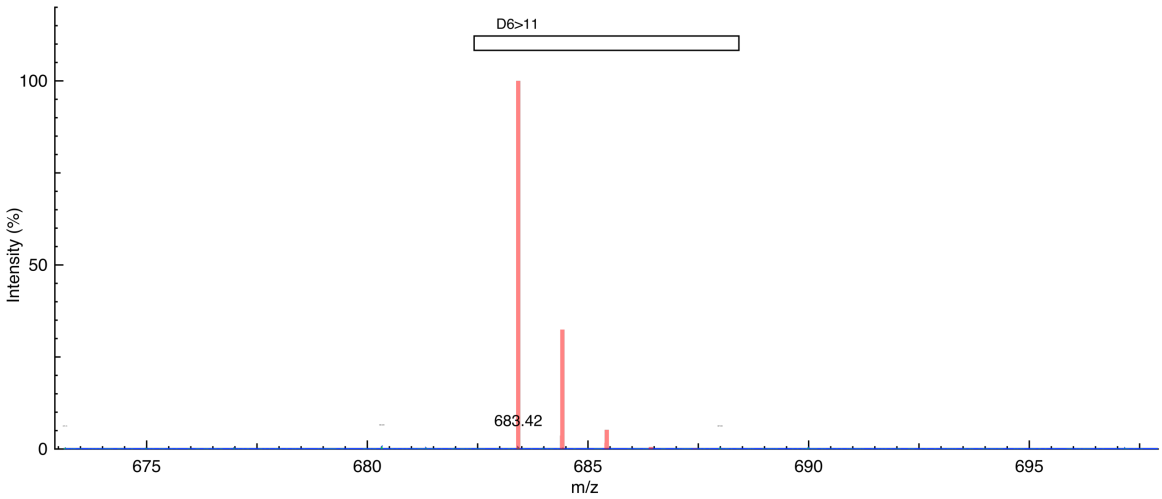


Figure S15. MS spectrum of magainin II cleaved fragment D6>11: H1(1+)HLeuHisSerAlaLysLysOH. Theoretical observed mass (m/z) = 683.420; experimental closest peak (m/z) = 683.420. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

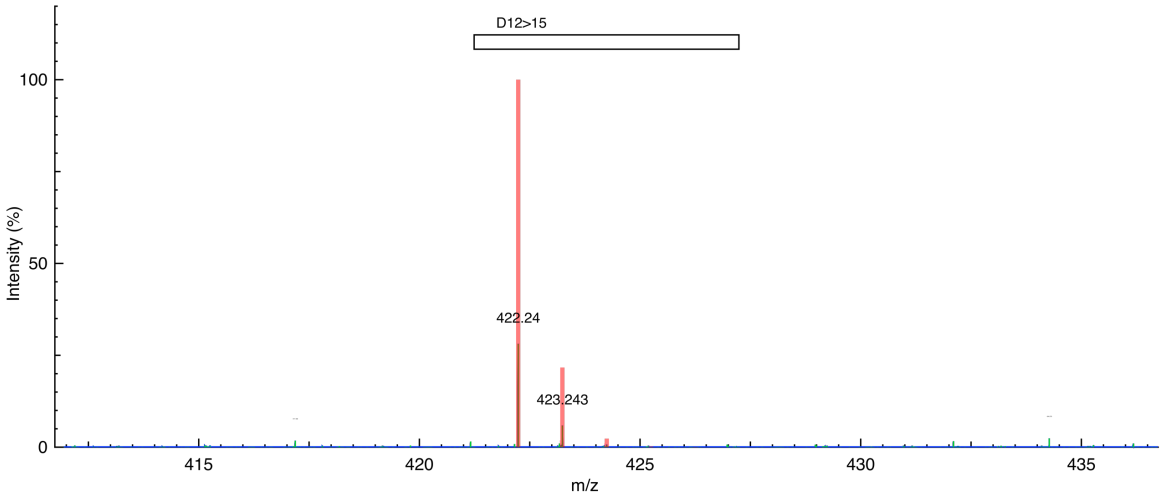


Figure S16. MS spectrum of magainin II cleaved fragment D12>15: H1(1+)HPheGlyLysAlaOH. Theoretical observed mass (m/z) = 422.240; experimental closest peak (m/z) = 422.240. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

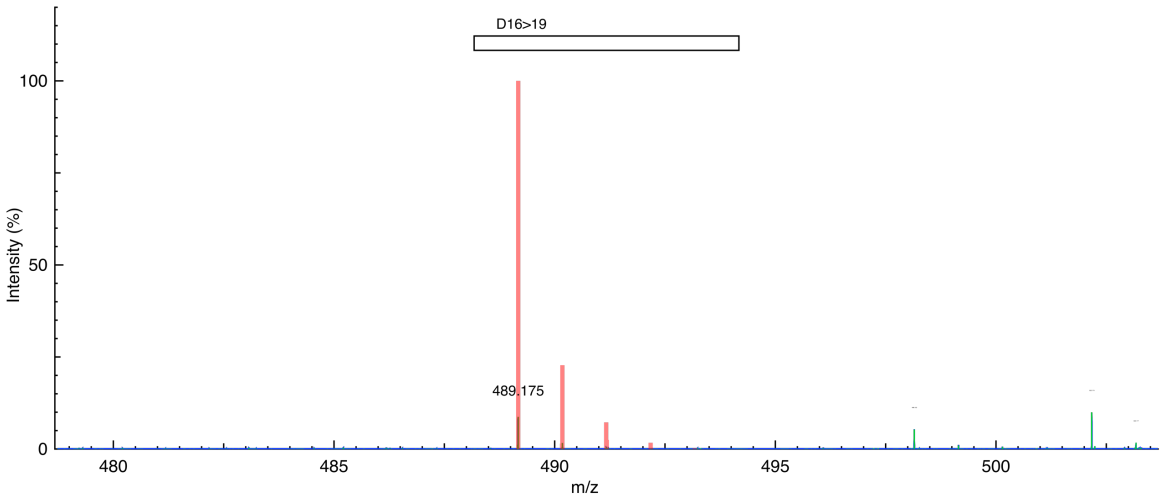


Figure S17. MS spectrum of magainin II cleaved fragment D16>19: K1(1+)HPheValGlyGluOH. Theoretical observed mass (m/z) = 489.175; experimental closest peak (m/z) = 489.175. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

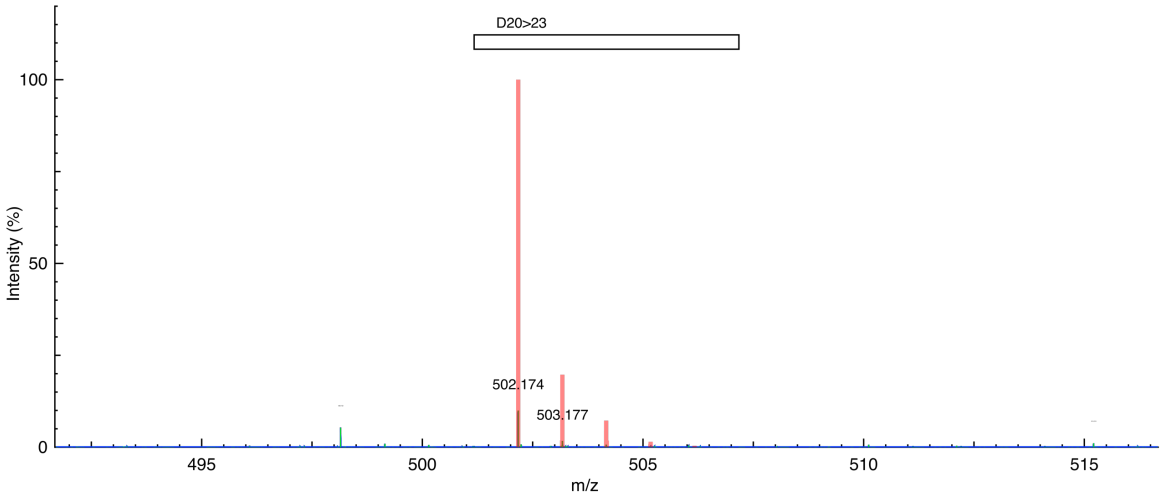


Figure S18. MS spectrum of magainin II cleaved fragment D20>23: K1(1+)HIleMetAsnSerOH. Theoretical observed mass (m/z) = 502.173; experimental closest peak (m/z) = 502.174. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

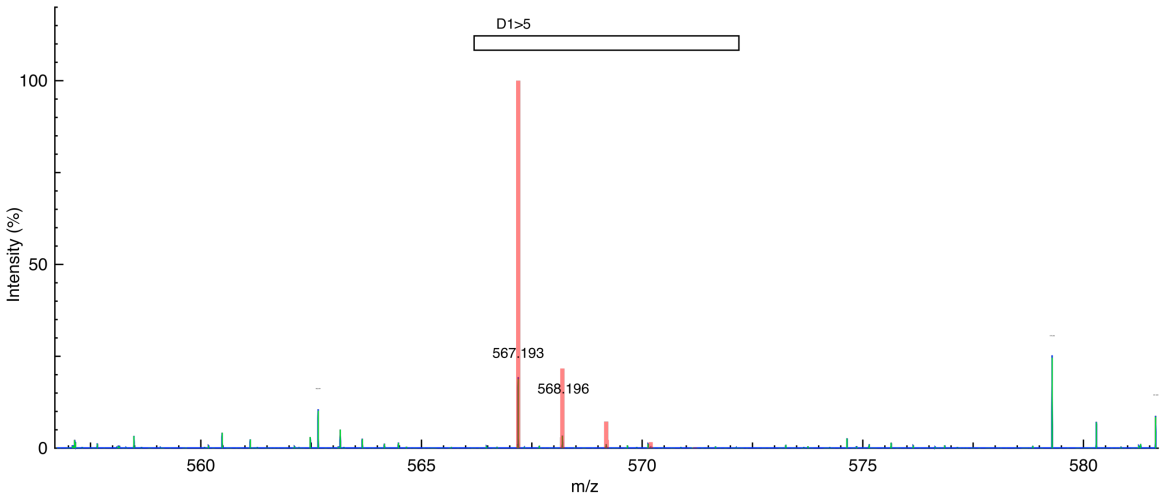


Figure S19. MS spectrum of glucagon cleaved fragment D1>5: K1(1+)HHisSerGlnGlyThrOH. Theoretical observed mass (m/z) = 567.192; experimental closest peak (m/z) = 567.193. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

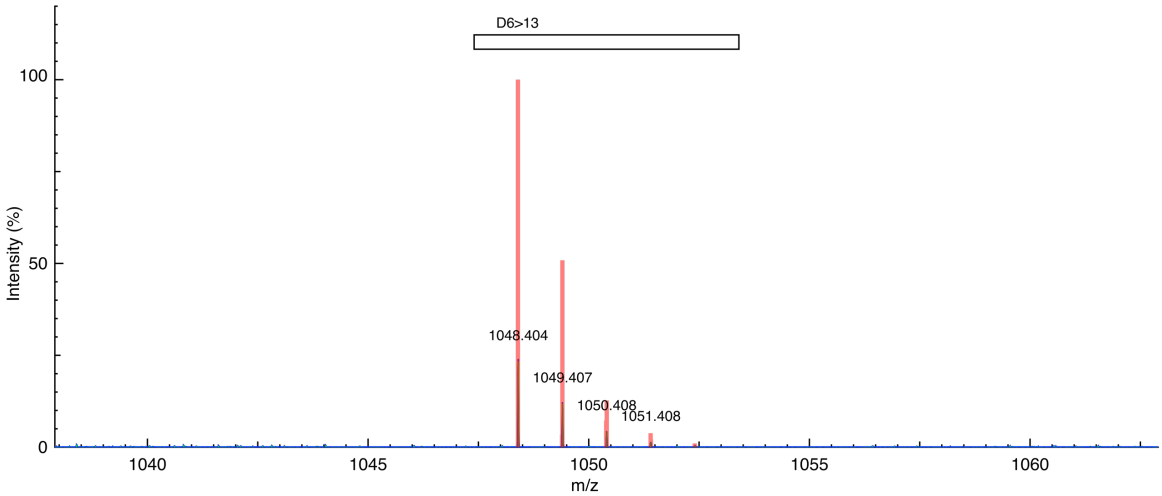


Figure S20. MS spectrum of glucagon cleaved fragment D6>13: K1(1+)HPheThrSerAspTyrSerLys TyrOH. Theoretical observed mass (m/z) = 1048.402; experimental closest peak (m/z) = 1048.404. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

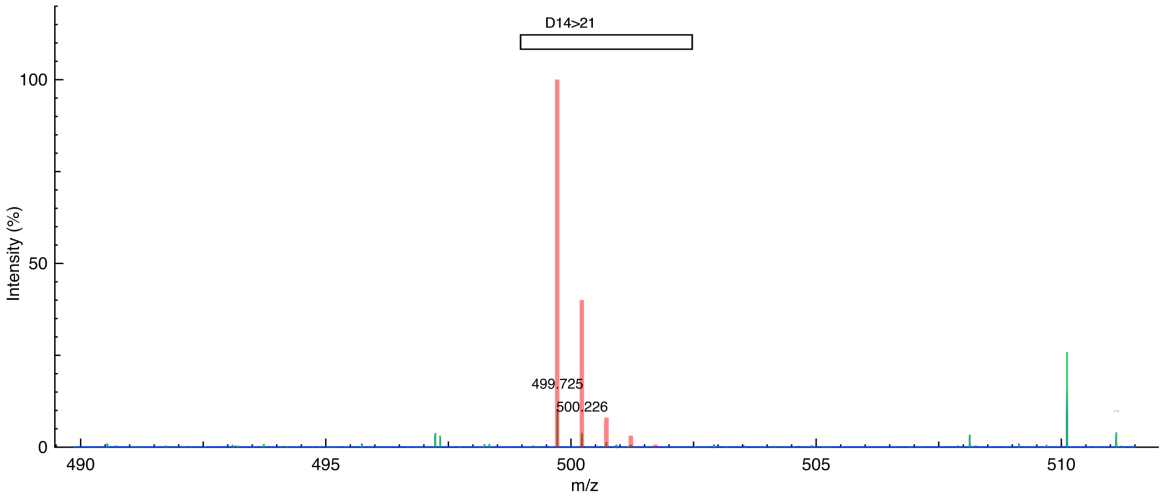


Figure S21. MS spectrum of glucagon cleaved fragment D14>21: K1(1+)H1(1+)HLeuAspSerArgArg AlaGlnAspOH. Theoretical observed mass (m/z) = 499.724; experimental closest peak (m/z) = 499.725. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

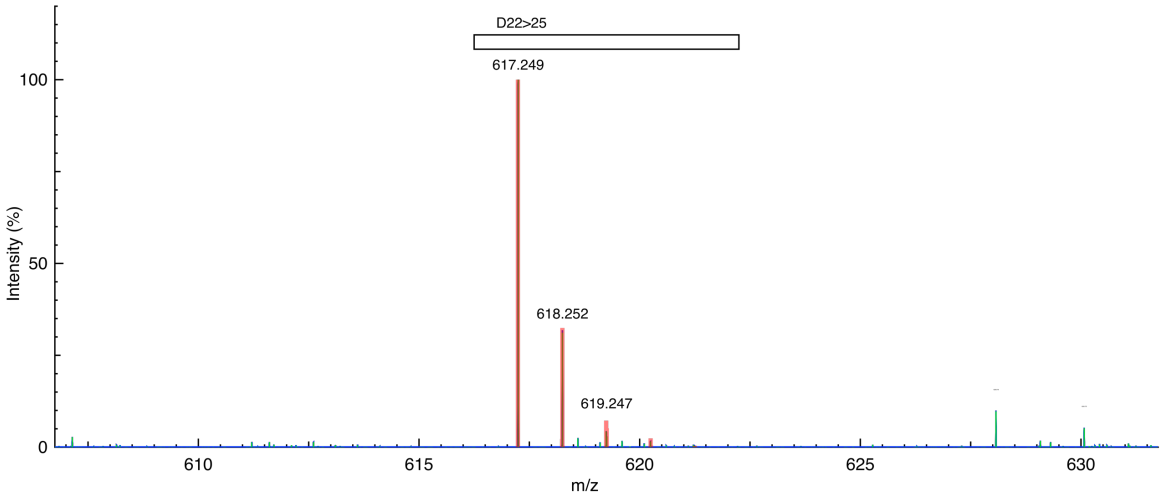


Figure S22. MS spectrum of glucagon cleaved fragment D22>25: K1(1+)HPheValGlnTrpOH. Theoretical observed mass (m/z) = 617.248; experimental closest peak (m/z) = 617.249. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

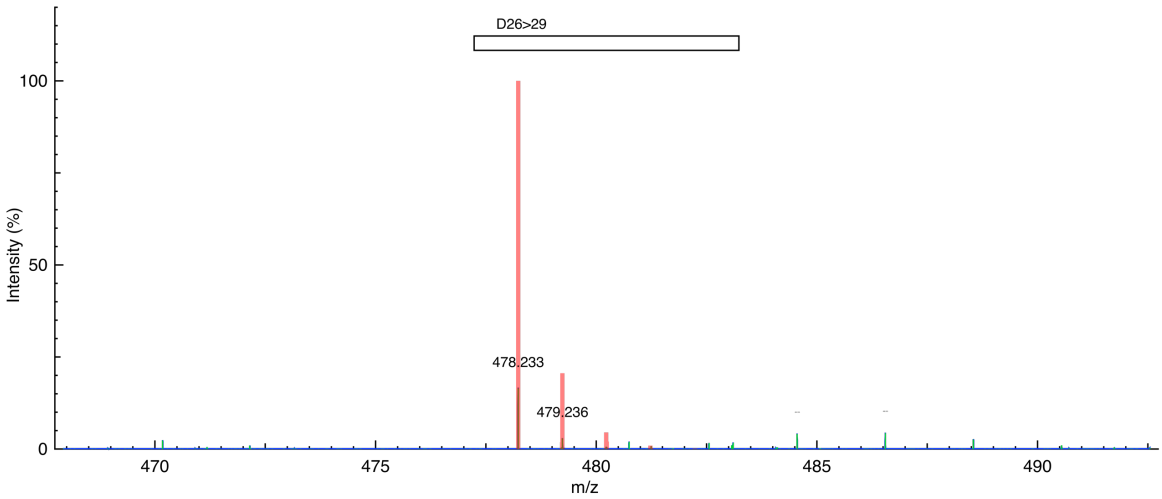


Figure S23. MS spectrum of glucagon cleaved fragment D26>29: H1(1+)HLeuMetAsnThrOH. Theoretical observed mass (m/z) = 478.233; experimental closest peak (m/z) = 478.233. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

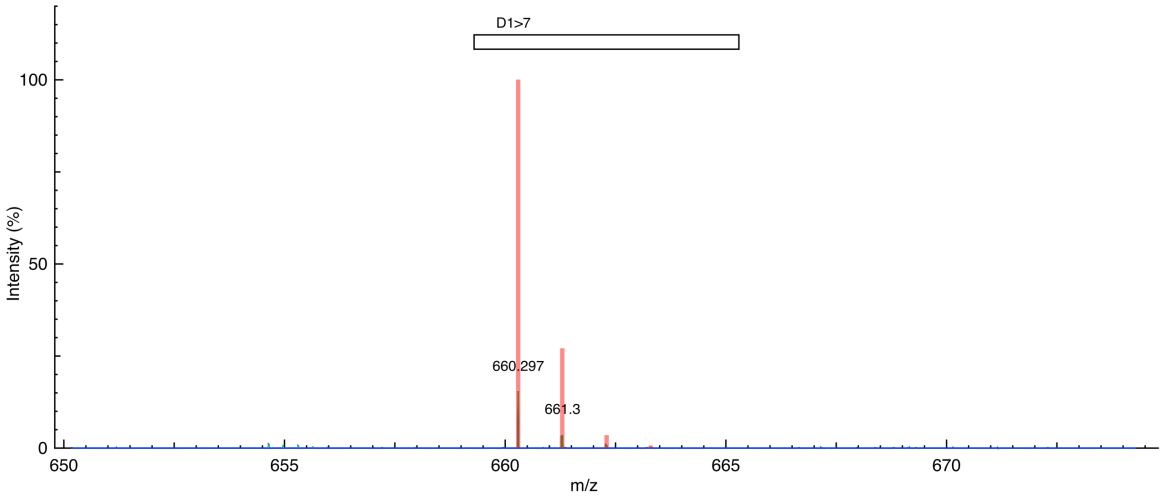


Figure S24. MS spectrum of somatostatin 28 cleaved fragment D1>7: H1(1+)HSerAlaAsnSerAsnPro AlaOH. Theoretical observed mass (m/z) = 660.295; experimental closest peak (m/z) = 660.297. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

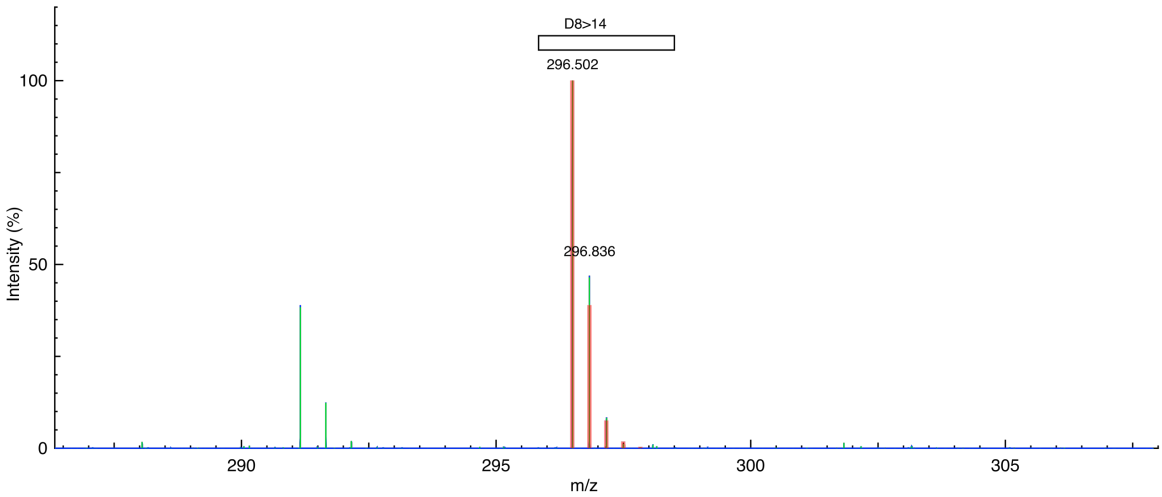


Figure S25. MS spectrum of somatostatin 28 cleaved fragment D8>14: H3(3+)HMetAlaProArgGlu ArgLysOH. Theoretical observed mass (m/z) = 296.501; experimental closest peak (m/z) = 296.502. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

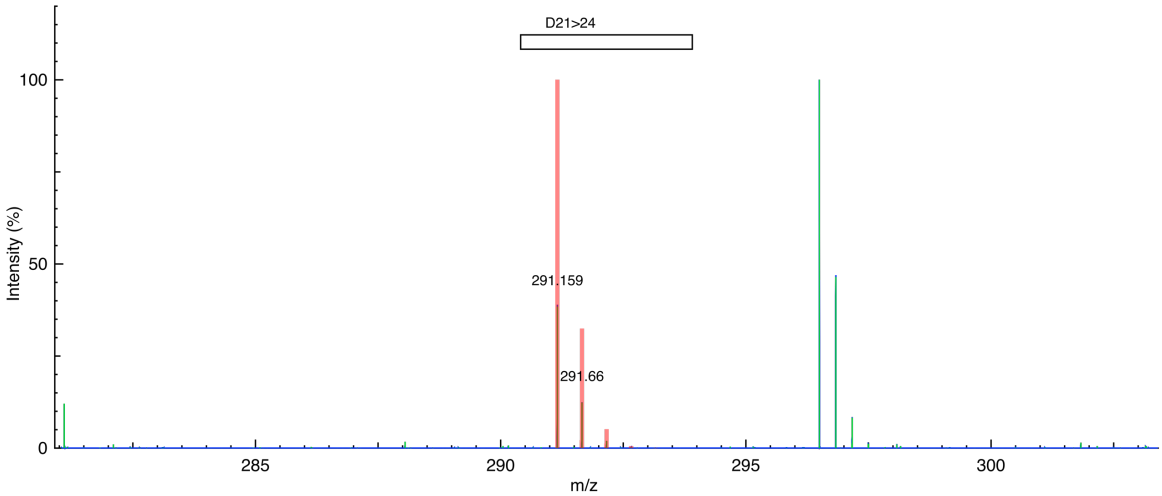


Figure S26. MS spectrum of somatostatin 28 cleaved fragment D21>24: H2(2+)HPheTrpLysThrOH. Theoretical observed mass (m/z) = 291.158; experimental closest peak (m/z) = 291.159. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

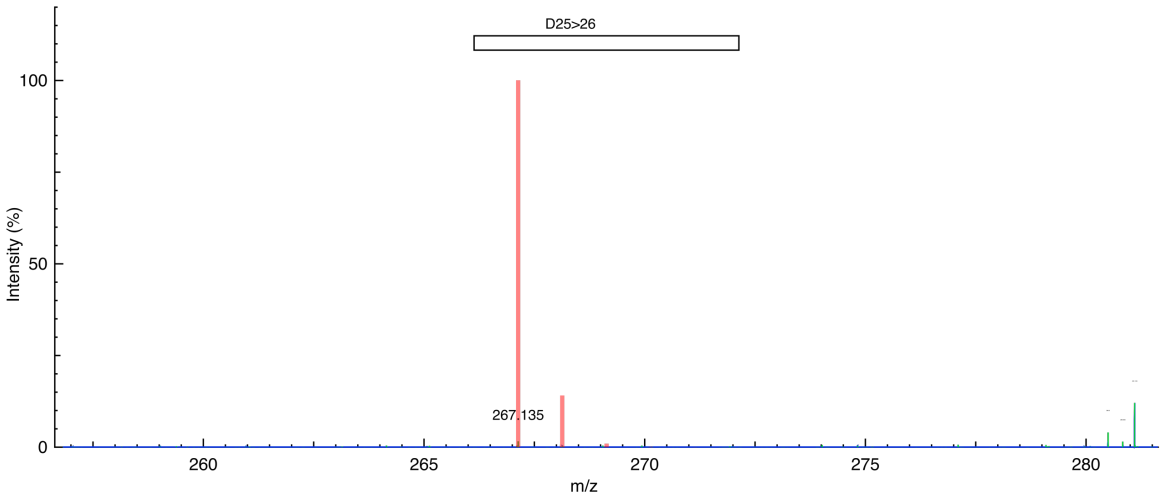


Figure S27. MS spectrum of somatostatin 28 cleaved fragment D25>26: H1(1+)HPheThrOH. Theoretical observed mass (m/z) = 267.134; experimental closest peak (m/z) = 267.135. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

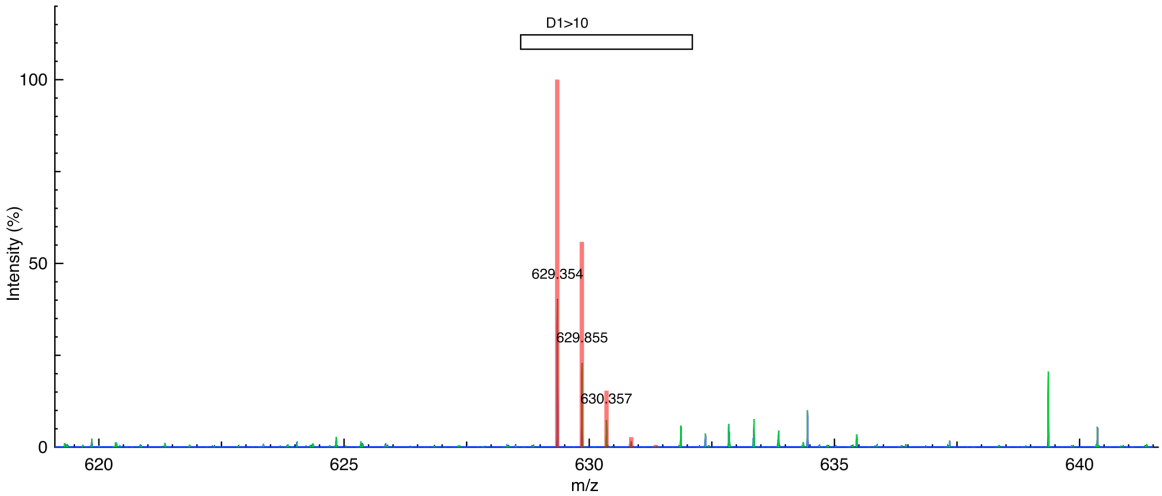


Figure S28. MS spectrum of the non-natural peptide cleaved fragment D1>10: H2(2+)HMet(S-1CH2)Val(H-1F)Val(H-1F)Arg((CH2)-1O)Thr(CH2)Thr(CH2)Arg((CH2)-1O)Thr(CH2)Thr(CH2) Thr(CH2)OH. Theoretical observed mass (m/z) = 629.352; experimental closest peak (m/z) = 629.354. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

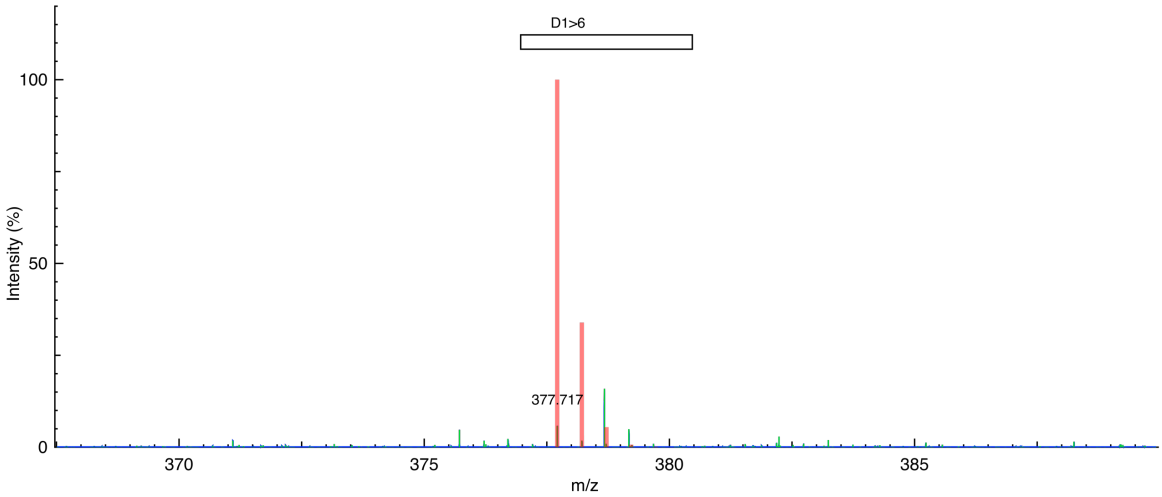


Figure S29. MS spectrum of the non-natural peptide cleaved fragment D1>6: H2(2+)HMet(S-1CH2)Val(H-1F)Val(H-1F)Arg((CH2)-1O)Thr(CH2)Thr(CH2)OH. Theoretical observed mass (m/z) = 377.717; experimental closest peak (m/z) = 377.717. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

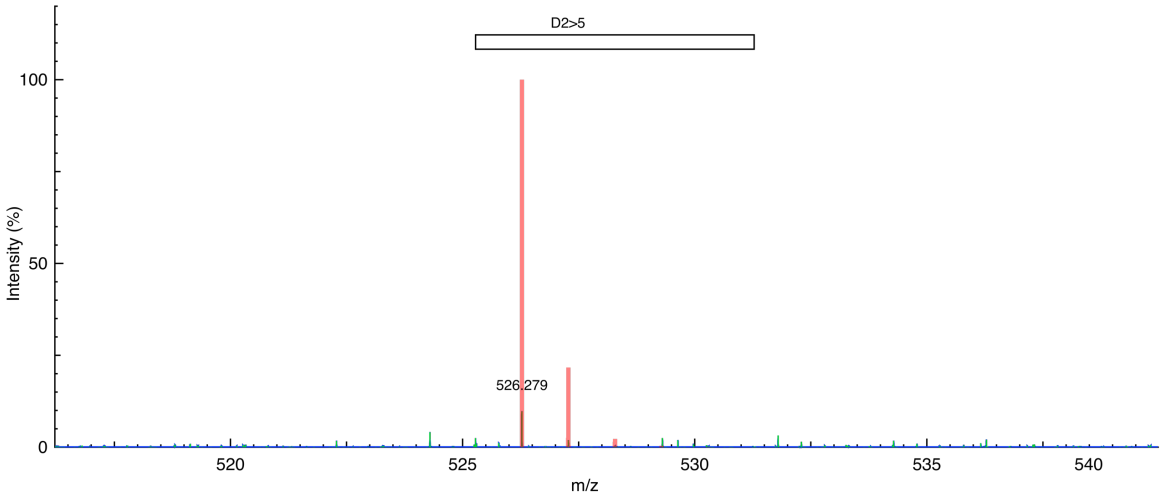


Figure S30. MS spectrum of the non-natural peptide cleaved fragment D2>5: H1(1+)HVal(H-1F)Val(H-1F)Arg((CH2)-1O)Thr(CH2)OH. Theoretical observed mass (m/z) = 526.280; experimental closest peak (m/z) = 526.279. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

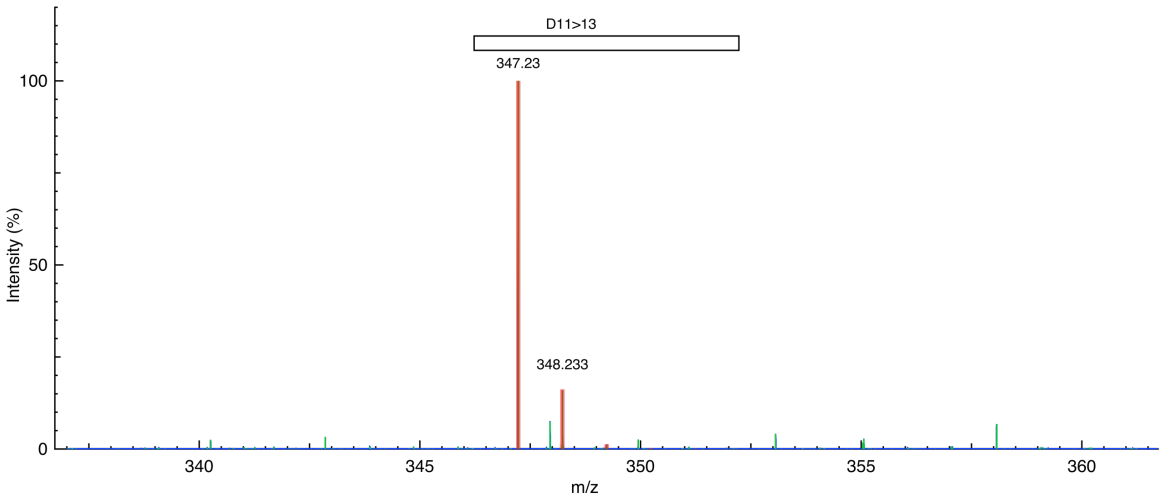


Figure S31. MS spectrum of the non-natural peptide cleaved fragment D11>13: H1(1+)HMet(S-1CH2)SerLysOH. Theoretical observed mass (m/z) = 347.229; experimental closest peak (m/z) = 347.230. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

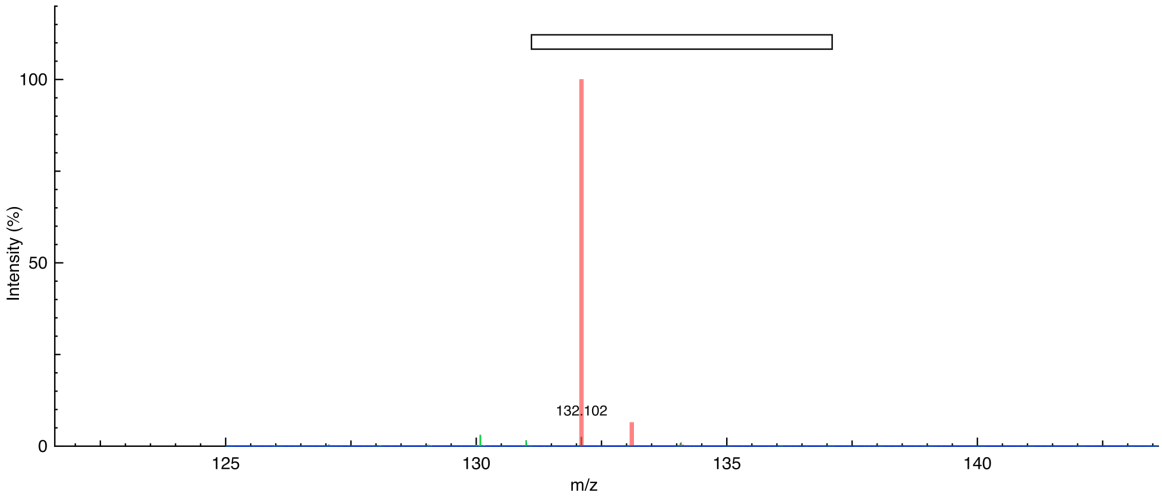


Figure S32. MS spectrum of L-norleucine: H1(1+)HMet(S-1CH2)OH. Theoretical observed mass (m/z) = 132.102; experimental closest peak (m/z) = 132.102. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

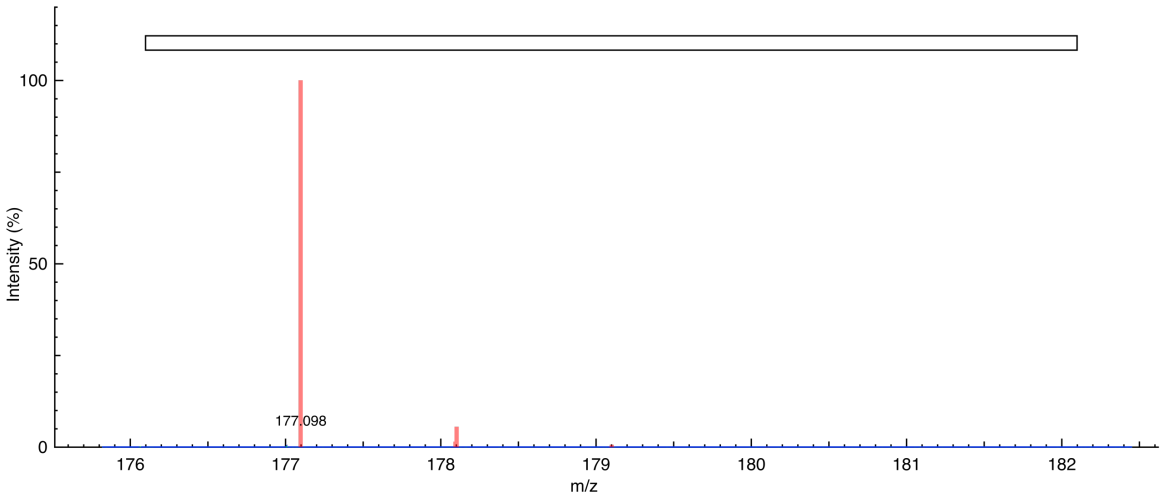


Figure S33. MS spectrum of L-canavanine: H1(1+)HArg((CH2)-1O)OH. Theoretical observed mass (m/z) = 177.098; experimental closest peak (m/z) = 177.098. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

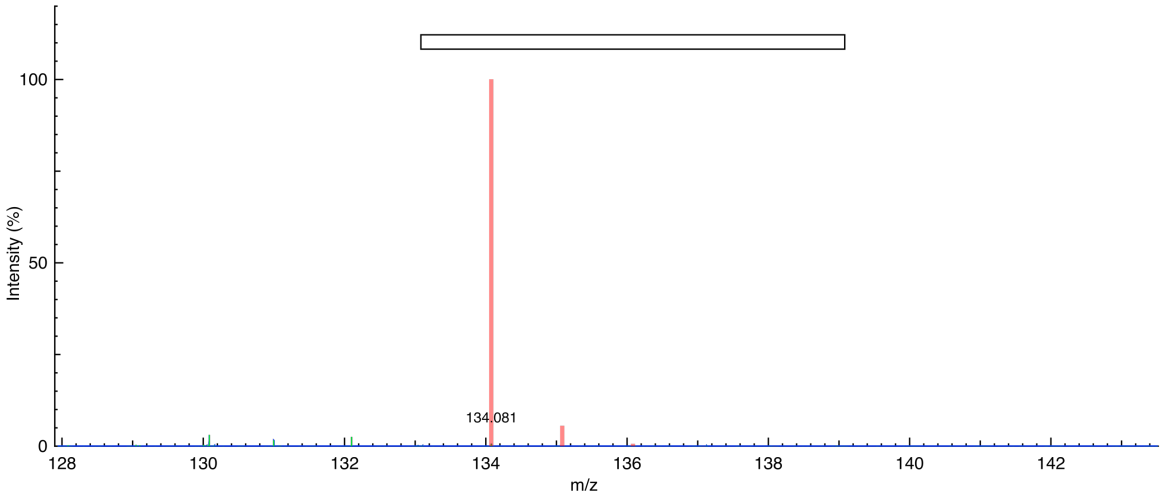


Figure S34. MS spectrum of DL-3-hydroxynorvaline: H1(1+)HThr(CH2)OH. Theoretical observed mass (m/z) = 134.081; experimental closest peak (m/z) = 134.081. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

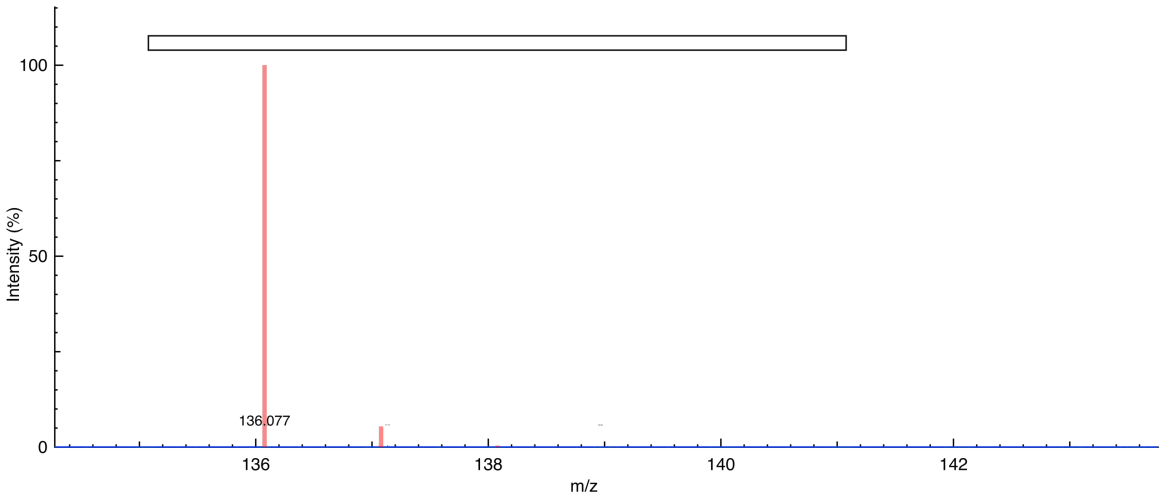


Figure S35. MS spectrum of 3-fluoro-DL-valine: H1(1+)HVal(H-1F)OH. Theoretical observed mass (m/z) = 136.077; experimental closest peak (m/z) = 136.077. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

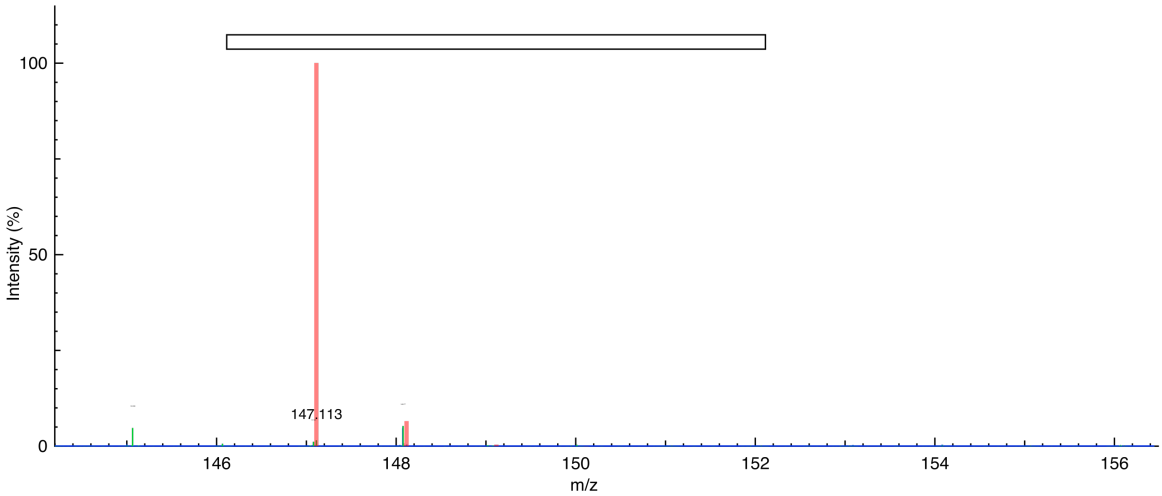


Figure S36. MS spectrum of L-lysine: H1(1+)HLysOH. Theoretical observed mass (m/z) = 147.113; experimental closest peak (m/z) = 147.113. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

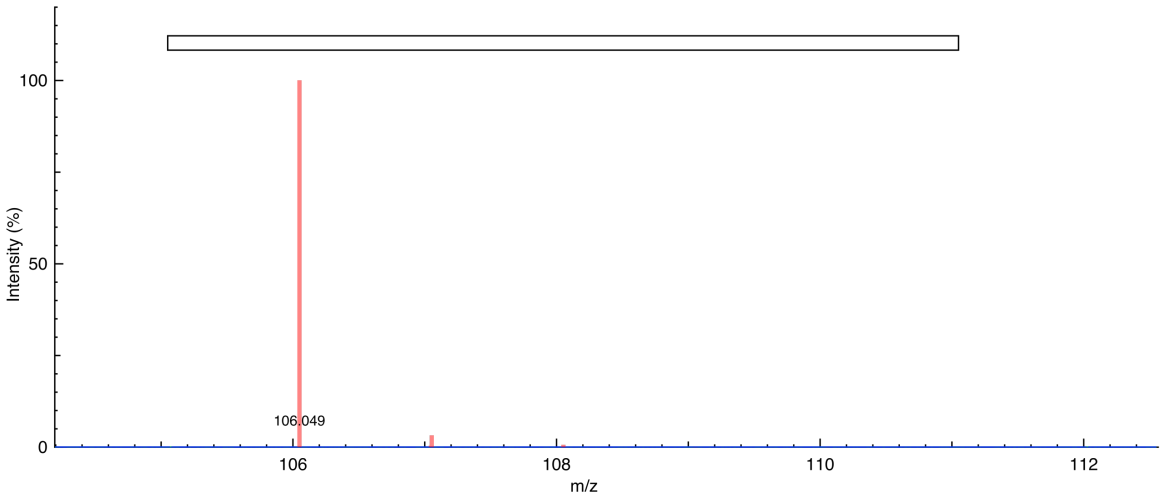


Figure S37. MS spectrum of L-serine: H1(1+)HSerOH. Theoretical observed mass (m/z) = 106.050; experimental closest peak (m/z) = 106.049. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

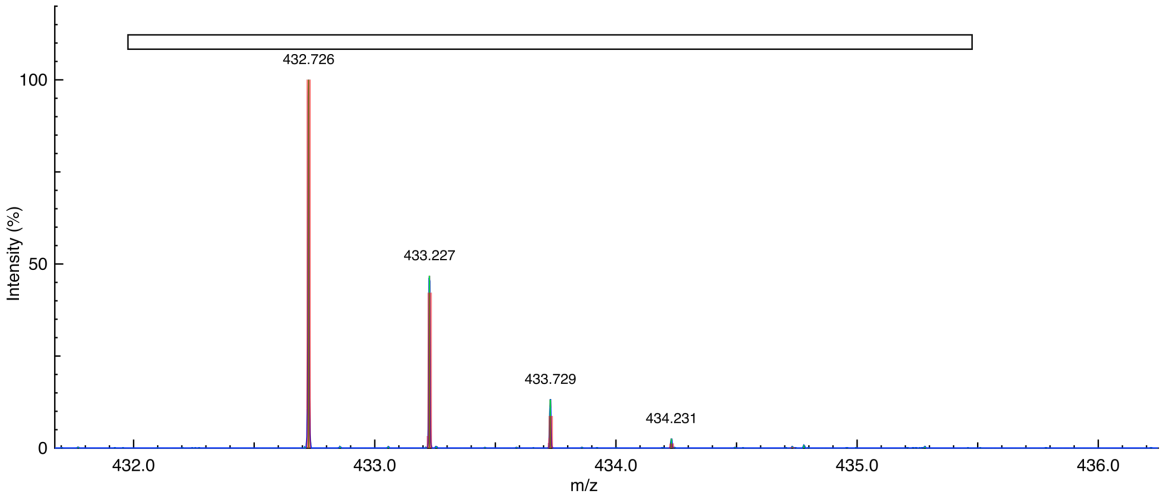


Figure S38. MS spectrum of the (in-gel) cleaved peptide from modified GFP (reference control), incorporating L-norleucine = HMet(S-1CH2)OH: H2(2+)HLysSerAlaMet(S-1CH2)ProGluGly TyrOH. Theoretical observed mass (m/z) = 432.727; experimental closest peak (m/z) = 432.726. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

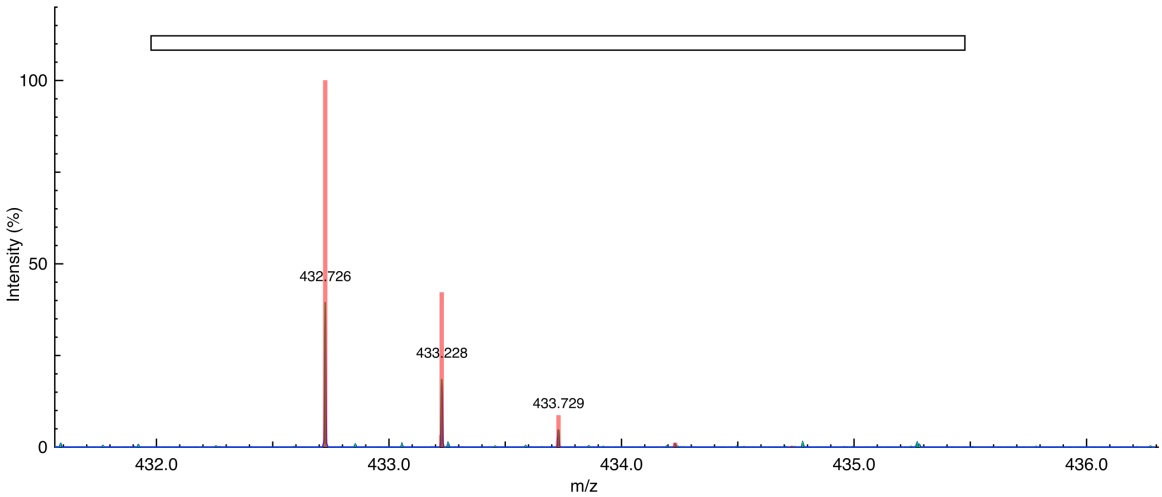


Figure S39. MS spectrum of the (in-gel) cleaved peptide from modified GFP (sample), incorporating the recycled L-norleucine = HMet(S-1CH2)OH: H2(2+)HLysSerAlaMet(S-1CH2)ProGluGlyTyrOH. Theoretical observed mass (m/z) = 432.727; experimental closest peak (m/z) = 432.727. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

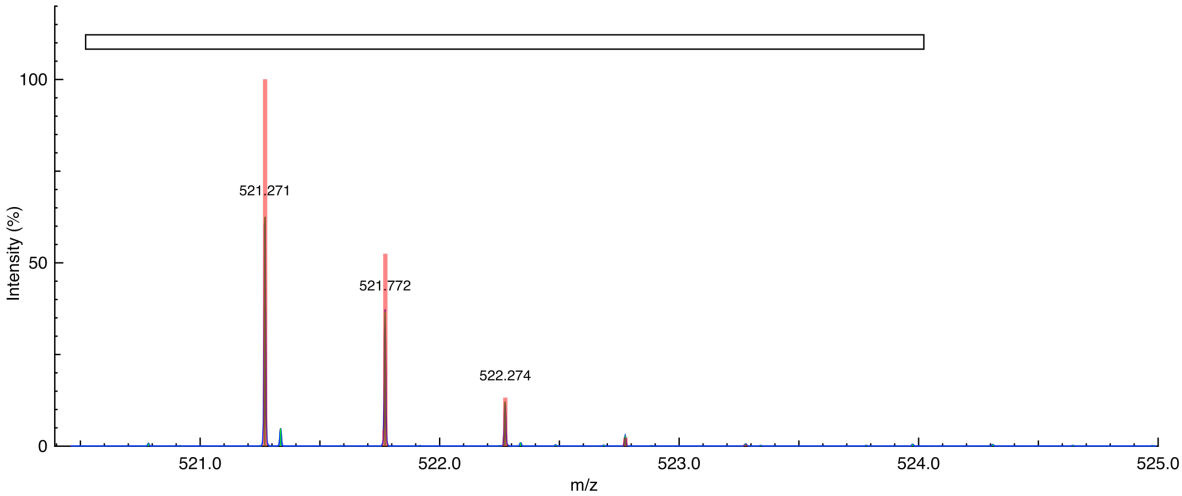


Figure S40. MS spectrum of the (in-gel) cleaved peptide from modified GFP (reference control), incorporating L-canavanine = HArg((CH2)-1O)OH: H2(2+)HValGlnGluArg((CH2)-1O)ThrIlePhe PheOH. Theoretical observed mass (m/z) = 521.272; experimental closest peak (m/z) = 521.271. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

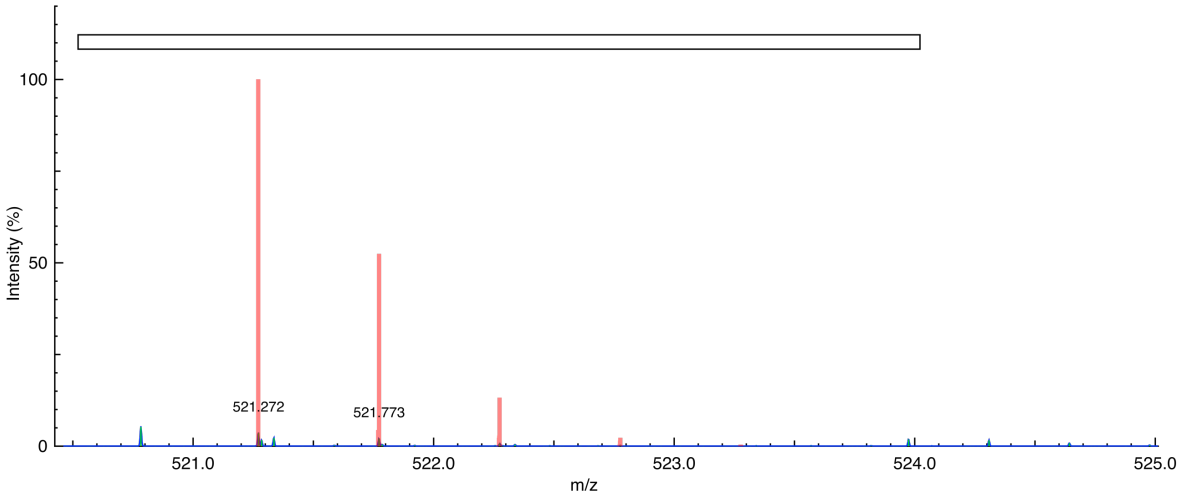


Figure S41. MS spectrum of the (in-gel) cleaved peptide from modified GFP (sample), incorporating the recycled L-canavanine = HArg((CH2)-1O)OH: H2(2+)HValGlnGluArg((CH2)-1O)ThrIlePhe PheOH. Theoretical observed mass (m/z) = 521.272; experimental closest peak (m/z) = 521.272. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

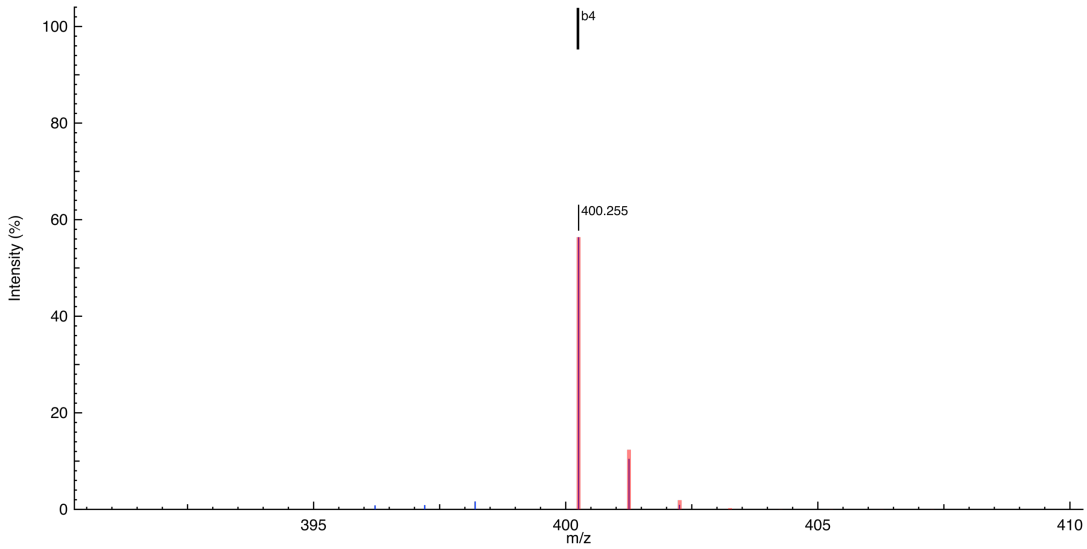


Figure S42. MS spectrum of the b4 fragment of the isolated H2(2+)HLysSerAlaMet(S-1CH2)ProGluGlyTyrOH peptide from modified GFP (reference control), incorporating L-norleucine = HMet(S-1CH2)OH: HLysSerAlaMet(S-1CH2)(1+). Theoretical observed mass (m/z) = 400.255; experimental closest peak (m/z) = 400.255. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

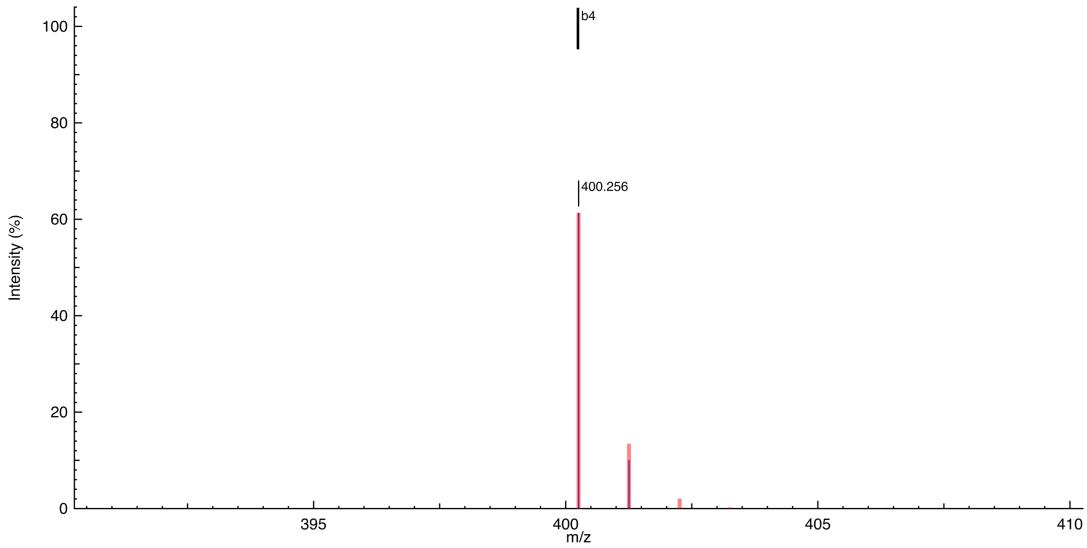


Figure S43. MS spectrum of the b4 fragment of the isolated H2(2+)HLysSerAlaMet(S-1CH2)ProGluGlyTyrOH peptide from modified GFP (sample), incorporating the recycled L-norleucine = HMet(S-1CH2)OH: HLysSerAlaMet(S-1CH2)(1+). Theoretical observed mass (m/z) = 400.255; experimental closest peak (m/z) = 400.256. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

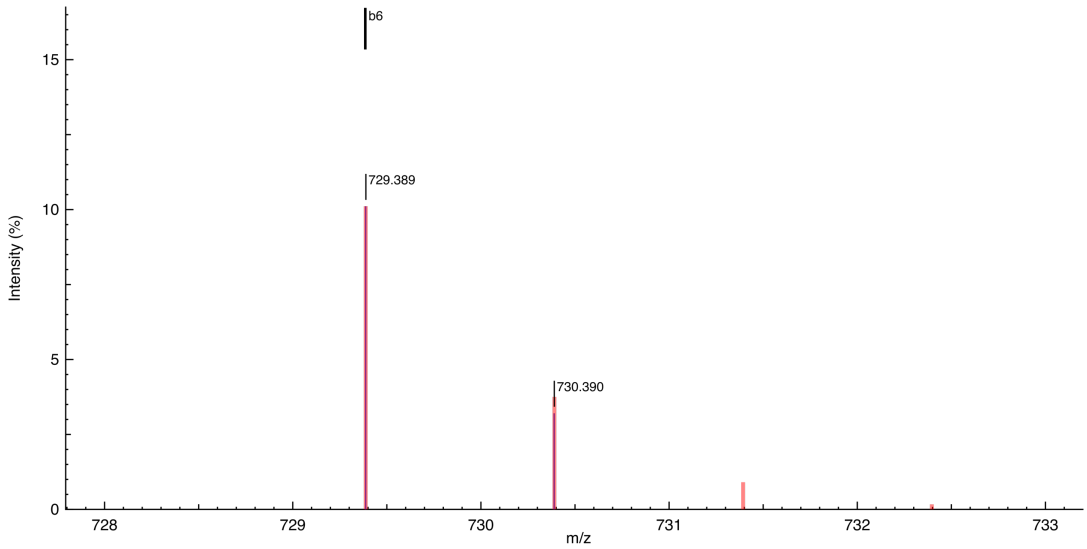


Figure S44. MS spectrum of the b6 fragment of the isolated H2(2+)HValGlnGluArg((CH2)-1O)ThrIlePhePheOH peptide from modified GFP (reference control), incorporating L-canavanine = HArg((CH2)-1O)OH: HValGlnGluArg((CH2)-1O)ThrIle(1+). Theoretical observed mass (m/z) = 729.389; experimental closest peak (m/z) = 729.389. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).

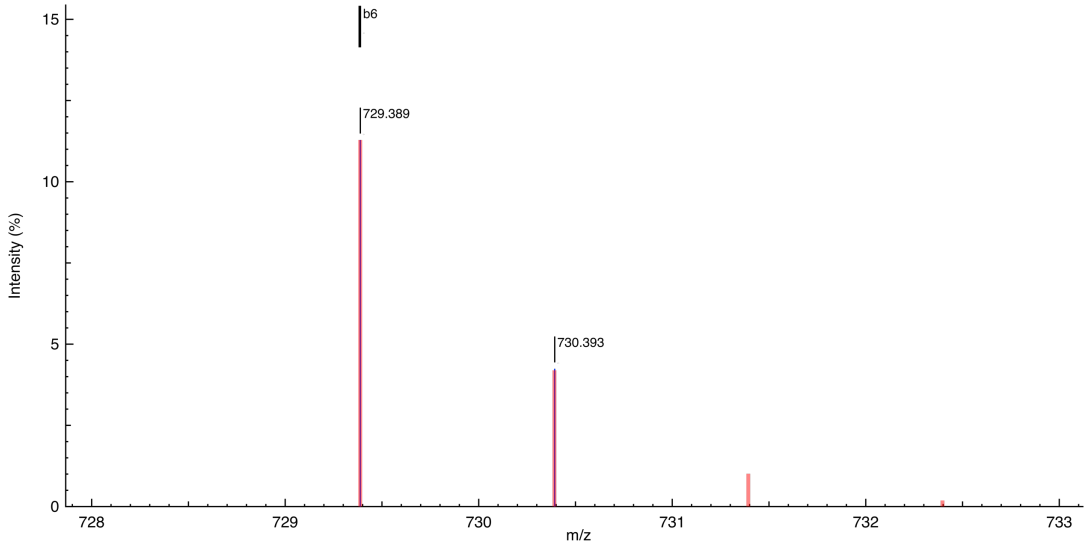
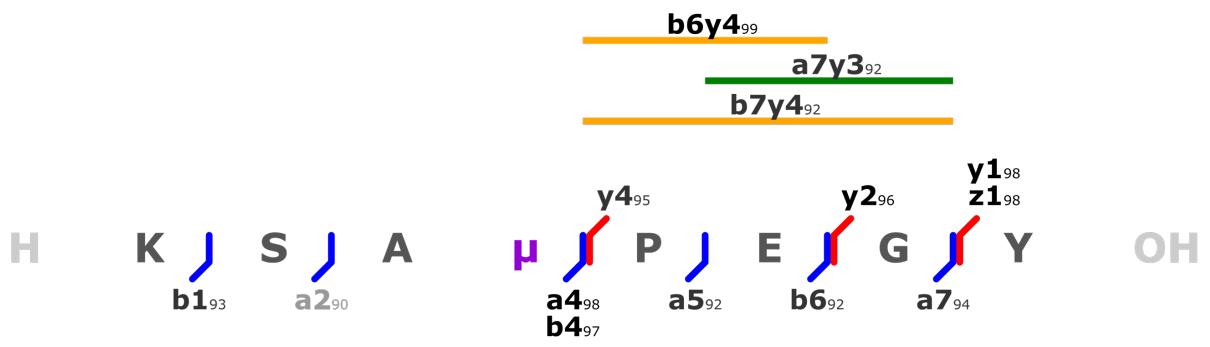
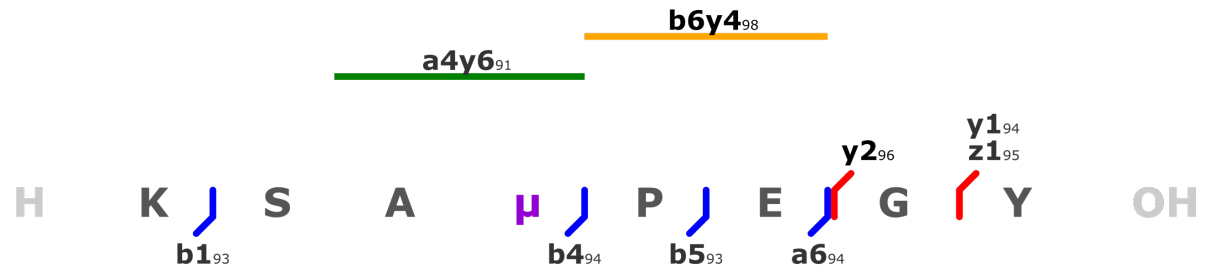


Figure S45. MS spectrum of the b6 fragment of the isolated H2(2+)HValGlnGluArg((CH2)-1O)ThrIlePhePheOH peptide from modified GFP (sample), incorporating the recycled L-canavanine = HArg((CH2)-1O)OH: HValGlnGluArg((CH2)-1O)ThrIle(1+). Theoretical observed mass (m/z) = 729.389; experimental closest peak (m/z) = 729.389. (Theoretical spectrum = red, experimental data = blue, and peak picking = green).



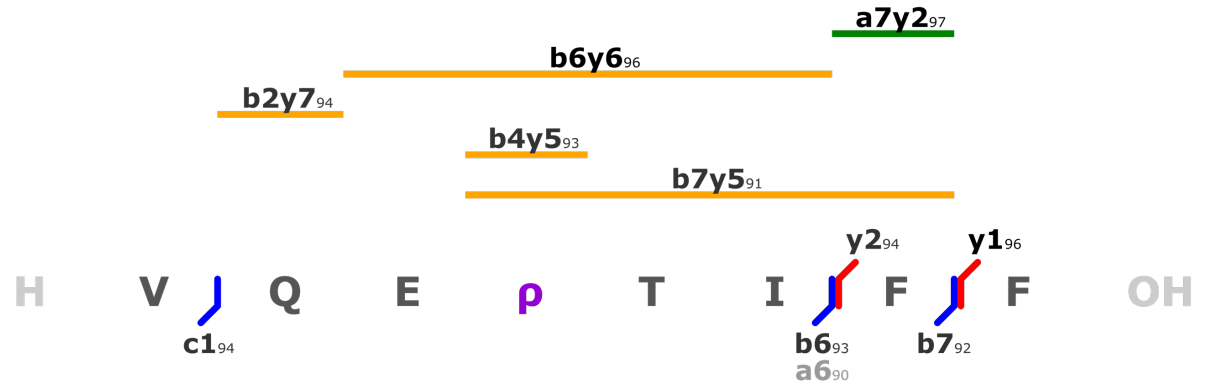
μ = Met(S-1CH2) = L-Norleucine

Figure S46. Fragmentation pattern of the isolated H2(2+)HLysSerAlaMet(S-1CH2)ProGluGlyTyrOH peptide from modified GFP (reference control), incorporating L-norleucine = HMet(S-1CH2)OH.



μ = Met(S-1CH2)= L-Norleucine

Figure S47. 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.



ρ = Arg((CH2)-1O)= L-Canavanine

Figure S48. Fragmentation pattern of the isolated H2(2+)HValGlnGluArg((CH2)-1O)ThrIlePhe PheOH peptide from modified GFP (reference control), incorporating L-canavanine = HArg((CH2)-1O)OH.

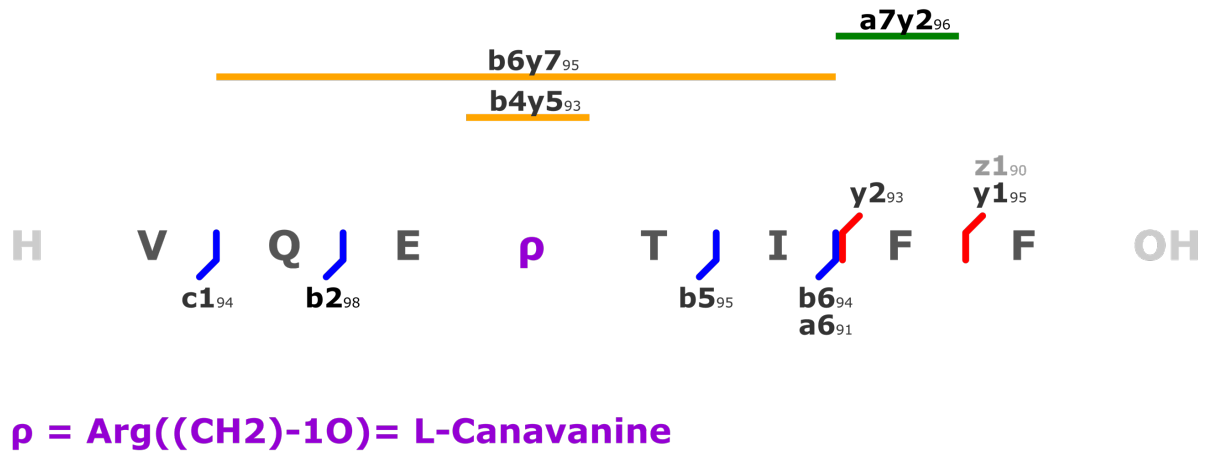


Figure S49. Fragmentation pattern of the isolated H2(2+)HValGlnGluArg((CH2)-1O)ThrIlePhe PheOH peptide from modified GFP (sample), incorporating the recycled L-canavanine = HArg((CH2)-1O)OH.

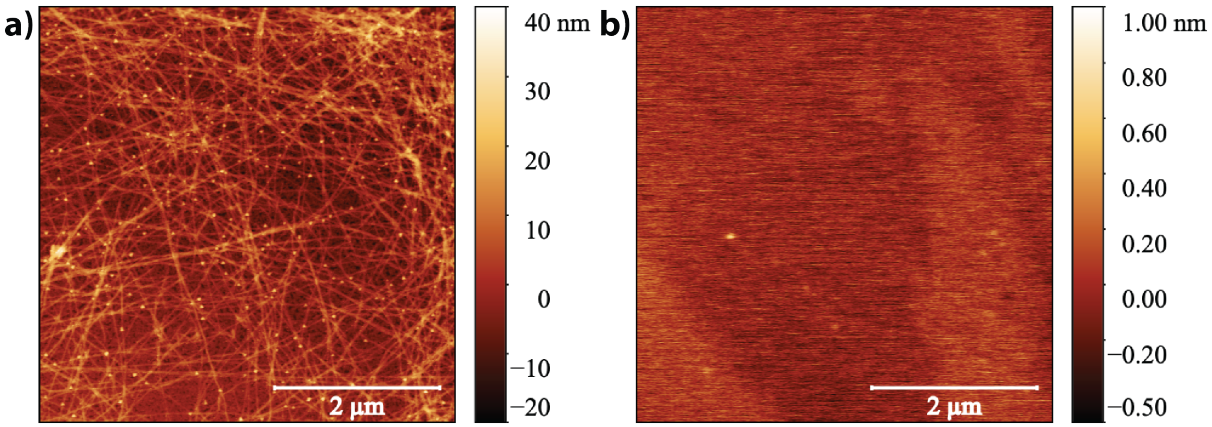


Figure S50. AFM characterization of the β-lactoglobulin amyloids obtained from solubilizing the film powder, and deposited on cleaved mica surfaces, as prepared (a) and after depolymerization (b).

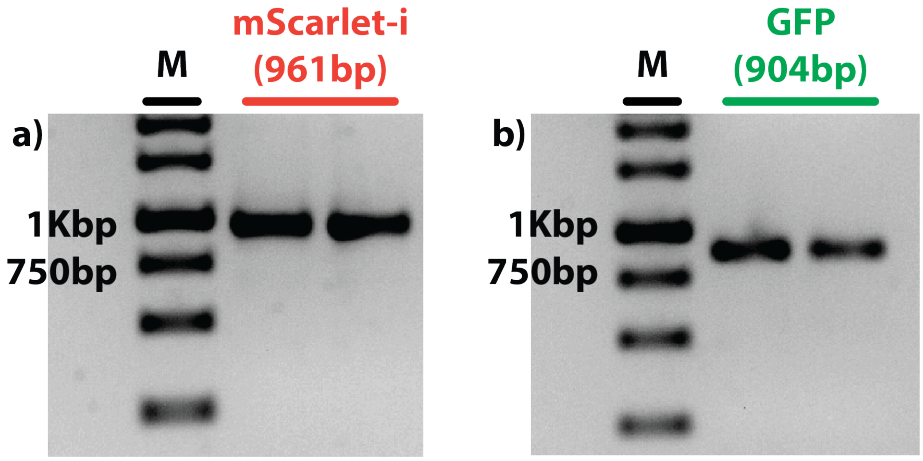


Figure S51. Electrophoresis gel (Agarose) of the PCR amplified mScarlet-i (a), and GFP templates (b).

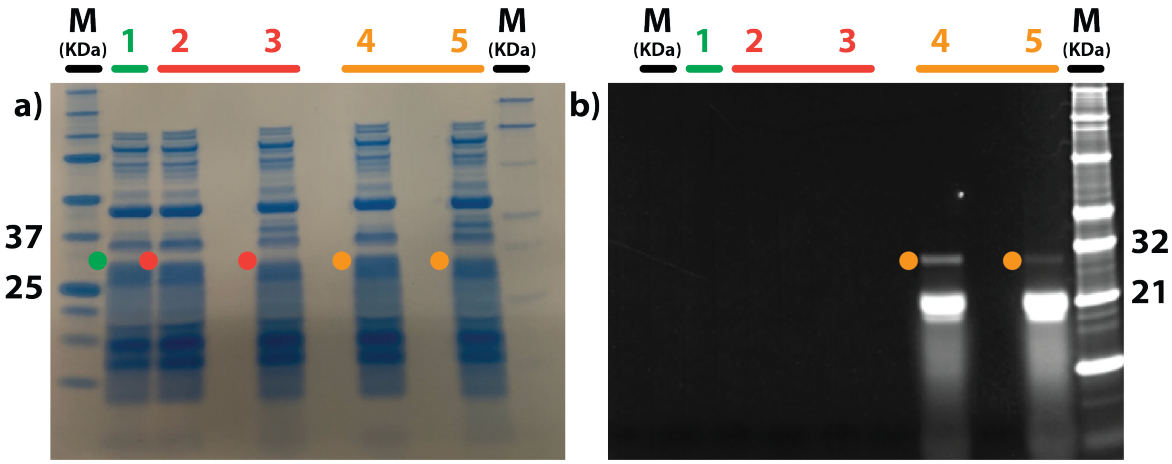


Figure S52. Coomassie stained (a), and fluorescence (b) images of the same SDS-PAGE protein gel used to prepare the expressed natural GFP (sample) (1), the modified GFP incorporating L-norleucine and L-canavanine (reference control) (2), and the modified GFP incorporating the recycled L-norleucine and L-canavanine (sample) (3) for 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).