Acyl transfer catalytic activity in *de novo* designed protein with N-terminus of α-helix as oxyanion-binding site

Elise A. Naudin,[1] Alastair G. McEwen,[2] Sophia K. Tan,[3] Pierre Poussin-Courmontagne,[2] Jean-Louis Schmitt,[1] Catherine Birck,[2] William F. DeGrado[3]* and Vladimir Torbeev[1]*

[1]Institut de Science et d'Ingénierie Supramoléculaires (ISIS), International Center for Frontier Research in Chemistry (icFRC), University of Strasbourg, CNRS (UMR 7006), Strasbourg, France

[2]Integrated Structural Biology Platform, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS (UMR 7104), INSERM (U1258), University of Strasbourg, Illkirch, France

[3]Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, USA

*Correspondence should be addressed to V. T. (torbeev@unistra.fr) or W. F. D. (william.degrado@ucsf.edu)

ABSTRACT: Design of catalytic proteins with functional sites capable of specific chemistry gains its momentum and a number of artificial enzymes have been recently reported including hydrolases, oxidoreductases, retro-aldolases and others. Our goal is to develop a peptide ligase for robust catalysis of amide bond formation and possessing no stringent restrictions to the amino acid composition at the ligation junction. Here we report the successful completion of the first step in this long-term project by building a completely *de novo* protein with predefined acyl transfer catalytic activity. We applied minimalist approach to rationally design an oxyanion hole within a small cavity and containing an adjacent thiol nucleophile. The N-terminus of α-helix with unpaired hydrogen-bond donors was exploited as a structural motif to stabilize negatively charged tetrahedral intermediates in nucleophilic addition-elimination reactions at acyl group. Cysteine acting as a principal catalytic residue was introduced at second residue position of α-helix N-terminus in a designed three-α-helix protein based on structural informatics prediction. We showed that this minimal set of functional elements is sufficient for the emergence of catalytic activity in a *de novo* protein. Using peptide-$^\alpha$thioesters as acyl-donors we demonstrated their catalyzed amidation concomitant with hydrolysis and proved that the environment at the catalytic site critically influences the reaction outcome. These results represent a promising starting point for the development of efficient catalysts for protein labeling, conjugation and peptide ligation.

INTRODUCTION

Designing proteins with the desired catalytic properties provides with a learning ground for understanding the fundamentals of enzyme catalysis and represents a major scientific challenge.[1-3] Moreover, the detailed knowledge of the enzyme mechanisms can be useful for deeper insights into chemical reactivity in general. In the recent years, there has been a significant progress in the design of enzymes using computational methods for a variety of reactions including retro-aldol reaction,[4] Kemp elimination,[5] Diels-Alder reaction[6] and ester hydrolysis.[7,8] The design strategy, proven to be general, consists of a combinatorial search of protein sequence, as well as optimization of backbone conformation and side chain rotamers in order to find the best arrangement of functional groups that stabilize transition-state of catalyzed reaction when it is docked into known protein scaffolds.[9] However, the catalytic efficiency of such computationally redesigned proteins turned out to be substantially lower in comparison to the observed catalytic rates for natural enzymes.[10] The currently available solution to achieve higher rates is to apply high-throughput methods such as directed evolution to optimize the sequences of initially designed scaffolds yielding man-made enzymes with higher catalytic efficiencies.[11,12] A recent highlight of the application of directed evolution strategy is the highly proficient Zn-containing hydrolase with catalytic efficiency surpassing the hydrolytic activities of the naturally occurring metalloenzymes.[13]

Installing a catalytic function increases the level of design complexity compared to a more tractable task of the design of three-dimensional protein structure.[14-16] The difficulty of recapitulating high catalytic efficiencies of natural enzymes by purely computational strategies is attributed to a lack of high enough precision of placement of critical catalytic residues and inability to control long-range inter-residue interactions including protein conformational dynamics. Several rounds of laboratory evolution are able to rectify these deficiencies by remodeling the structure of the active site,[17] as well as optimizing the surrounding second shell residues that contribute to an intricate balance between precise structural complementarity to transition state(s) of the catalyzed reaction and protein conformational flexibility mediating the efficient substrate binding and product release.[18,19]

At the present moment, methodologically more simple minimalist approach for *de novo* design of protein catalysts is still on par with computational methods.[20] Minimalist design is based on chemical logic and intuition and implies the

incorporation of a minimal number of elements (such as amino acid residues or cofactors) into a *de novo* protein scaffold that is initially devoid of catalytic properties in order to convert it into the catalytically active variant.[14,16,20] The recent examples of *de novo* catalytic proteins include a thermostable hydrolase based on a 7-helix bundle and containing multiple copies of Cys-His-Glu triad as catalytic apparatus,[21] a promiscuous peroxidase that utilizes heme as a cofactor bound to 4-helix bundle scaffold[22] and a hydrolase based on calmodulin, where the activity is regulated allosterically upon binding to Ca^{2+} ion.[23] In addition, it was shown that the catalytic proteins developed using minimalistic design can also be subjected to rounds of directed evolution to improve their properties.[24]

Our aim is to elaborate a protein catalyst that can be used to accelerate amide bond formation with applications for chemical or semi-synthesis of proteins, or for protein modification. Chemical approaches for production of proteins hold several advantages in comparison to molecular biology tools such as the possibility to incorporate a diverse set of non-canonical building blocks,[25,26] cross-links,[27] to produce mirror-image protein forms,[28,29] proteins with non-linear backbone topology,[30,31] or even constructs that are hybrids of polypeptides and other types of polymers.[32] These capabilities can enrich protein science with the concepts and tools from physical organic, supramolecular, polymer and other chemistries.

We are inspired by the naturally occurring ligases, such as butelase-1,[33] sortase A[34] or ubiquitin ligases[35] which illustrate the feasibility of high catalytic efficiencies for condensation of acyl donor and acceptor peptide substrates. In addition, reengineered proteases have been demonstrated to perform condensation of peptides.[36] For instance, previously reengineered variant of subtilisin was successfully used to assemble a 124-residue ribonuclease A.[37] Furthermore, newer and more efficient variants of proteolytic proteins with ligation catalytic activity were recently developed.[38,39] However, both natural ligases and reengineered proteases possess restrictions to amino acids sequences that can bind to their substrate pockets; in other words, the catalytic efficiency is coupled to substrate specificity in these enzymes. Moreover, the overall complexity of these large and evolved structures complicates the improvement and further tuning of their enzymatic properties. To have a more general tool that is not limited by substrate scope and can be compatible with a variety of reaction conditions, in this study we introduced a catalytic ligation activity from first principles into a *de novo* protein scaffold.

RESULTS AND DISCUSSION

Protein design. The minimal functional elements required for the ligation reaction are a principal nucleophilic residue (such as Cys or Ser) and an "oxyanion hole" arrangement[40] of hydrogen-bond donors to stabilize the negative charge of the tetrahedral intermediates in this reaction. These functional motifs are present in both natural ligases and reengineered proteases.[36] We selected Cys as a key catalytic residue in favor of Ser for the following reasons: (i) the thiol group of Cys has lower pK$_a$ (8.5 in unstructured peptides)[41] and potentially can be used as a single residue instead of catalytic dyads or triads (*e.g.* classical Ser-His-Asp) that promote deprotonation of hydroxyl in Ser, because the correct placement of two or three residues can already complicate the design; (ii) the activation of peptide acyl donor in the form of peptide anchored to protein

scaffold via Cys-thioester bond can be advantageous to the related Ser-oxoester, because the thioester group benefits from an enhanced reactivity towards most of the nucleophiles due to the poorer orbital interactions between sulfur atom and carbonyl moiety.[42]

It should be mentioned that in addition to ligases and proteases the activation in the form of acyl-thioesters occurs in non-ribosomal peptide or fatty acid synthases[43,44] and in protein splicing through self-catalyzed intein exclusion and ligation of two adjacent exteins[45] highlighting the utility of thioesters for acyl-transfer reactions. Another advantage of Cys as a catalytic residue is the possibility of using synthetic peptide-$^\alpha$thioesters as acyl donor substrates, which can undergo transthioesterication with Cys, thus forming covalently attached peptide-protein species. Peptide-$^\alpha$thioesters are accessible by both Boc/benzyl and Fmoc/tBu solid-phase peptide synthesis (SPPS)[46] and are commonly used in enzyme-free total chemical synthesis of proteins, either via silver-catalyzed condensation of partially protected peptide-$^\alpha$thioesters and amino-peptides[47] or with the help of native chemical ligation, which is a highly valuable chemoselective reaction of fully unprotected peptide-$^\alpha$thioesters and cysteinyl-peptides proceeding in aqueous solutions.[48]
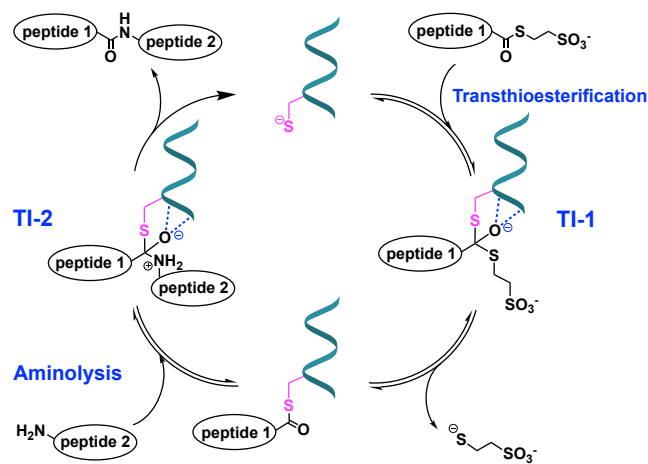
Scheme 1. Mechanism for the catalysis of acyl transfer reaction (*e.g.* peptide ligation) by *de novo* catalytic protein (depicted schematically in cyan). TI-1 and TI-2 refer to the key tetrahedral intermediates of transthioesterification and amidation steps, respectively, where negatively charged oxyanions are stabilized by H-bonding network (shown in blue).

Scheme 1 outlines the principal steps in the putative mechanism of catalysis of peptide ligation by a *de novo* protein. A peptide-$^\alpha$thioester reacts with the catalytic Cys and undergoes transthioesterification. After passing through the first tetrahedral intermediate (TI-1), the resulting branched peptide-protein thioester is posed to react with H$_2$N-peptide, the acyl acceptor substrate. The collapse of the second tetrahedral intermediate (TI-2) leads to the release of the ligated product and free catalytic protein. For this double-displacement ping-pong mechanism to operate efficiently, both negatively charged intermediates of the two addition-elimination steps need to be stabilized, for example, through H-bonding with amides of protein scaffold serving as an "oxyanion hole". Therefore,

protein framework in a *de novo* protein must include nucleophilic Cys placed in such an environment.

Previously, it was demonstrated that phospho-serine (pSer) when introduced at N-terminus of an α-helix can lead to its stabilization presumably due to productive H-bonding of unpaired amides of α-helical N-terminus to negatively charged phosphate group.[49] This stabilizing effect was most prominent when pSer was placed at N2 (second residue at N-terminus of α-helix) position.[49] By analyzing the structural data from PDB we provide support to these experimental results: clustering and alignment of crystal structures containing pSer at Ncap (residue preceding the first amino acid that adopts α-helical conformation), N1 and N2 (first and second positions at N-terminus of α-helix) illustrate that backbone amides at α-helical N-termini indeed can contribute 1-to-2 H-bonds to one of the distal oxygens of phospho-group (Figure 1). The tetrahedral intermediates of addition-elimination steps of acyl-transfer catalytic mechanism shown in Scheme 1 are negatively charged and bear similarities in structure to pSer, therefore, the N-terminus of α-helix represents a suitable structural motif for stabilization of the respective tetrahedral intermediates.
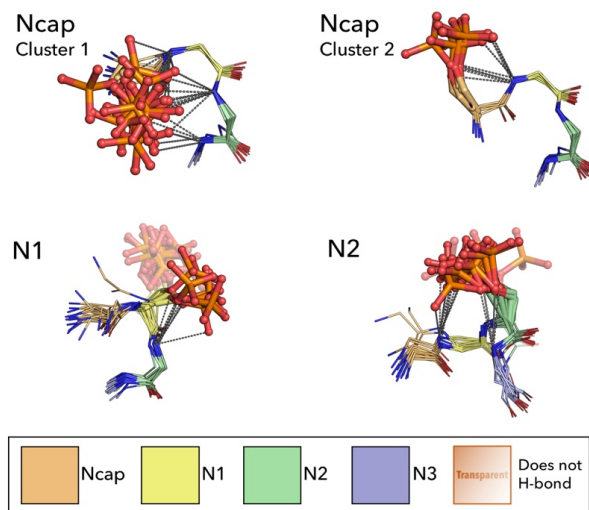

Figure 1. Structural alignment of pSer-containing fragments from structures in PDB (accessed in February 2020) with pSer located at Ncap, N1 and N2 positions of the corresponding α-helical fragments. The occurrences of pSer in non-redundant version of the PDB were determined using the database and methods previously described.[50] pSer at Ncap positions form 2 clusters, due to a bimodal distribution of its ψ-dihedrals. In both clusters, all of the pSers form H-bonds with downstream amides. However, only 7 out of 23 pSer's in the N1 position hydrogen bond to the helix backbone. The geometry of N2 pSer's appears as stabilizing as Ncap pSer's; 13 out of 16 fragments participate in H-bonds with its preceding and/or succeeding amide. pSer at the Ncap and N2 positions are observed contributing bivalent H-bonds, whereas only 1 H-bond is observed for pSer at N1 position. Example H-bonds are represented by dashed lines, and phosphate groups that do not H-bond with helix backbone are transparent. The PDB entries of the structures used in the alignment are listed in Table S1.

To test the feasibility of such catalytic strategy we have chosen the *de novo* designed domain-swapped dimer (DSD) protein that is composed of two identical sequences that spontaneously dimerize to form a three-α-helical bundle.[51] In this scaffold, the two adjacent α-helical N-termini are facing each other creating a cavity that represents a proper location to incorporate an active site. The environment at the juxtaposition of two N-termini is in the proximity of the hydrophobic interior of the helical bundle and, at the same time, solvent exposed, therefore, this part of the protein can permit access by various peptide substrates and allow for catalysis without introducing any steric restrictions in substrate structures.
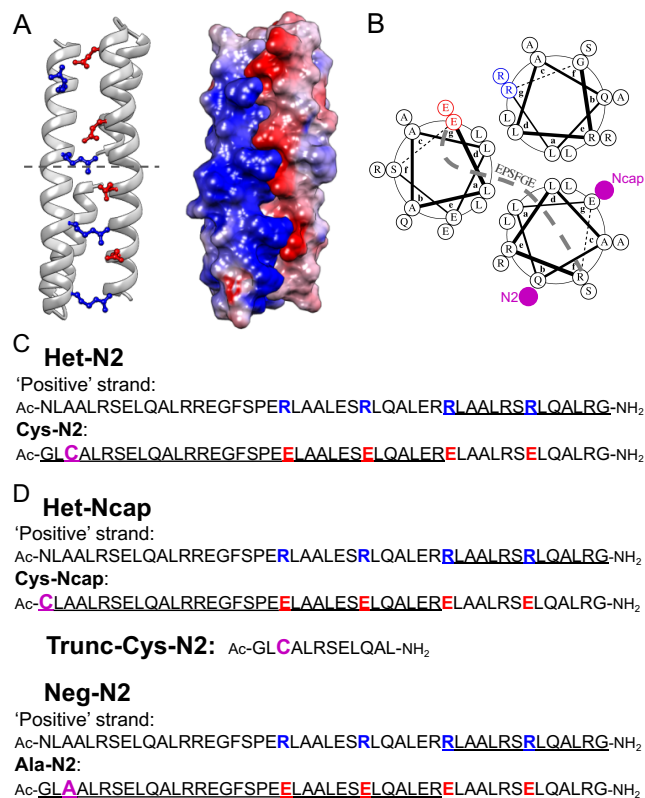

Figure 2. Design of heterodimers. (A) Left: charged residue patterning by grafting negative charges (Glu in red) on the strand containing the Cys catalytic residue and positive charges (Arg in blue) onto the other helix. Right: resulting columbic surface is shown. (B) Helical wheel diagram of the proteins that corresponds to the half-structure depicted above the dashed line in A. Leucine residues constitute the hydrophobic core (*a* and *d* positions). Glu and Arg residues at *e* and *g* positions promote heterodimerization. (C) Sequences of the two monomers of **Het-N2**. (D) Sequences of the three control constructs with Cys at Ncap position in **Het-Ncap**, a 12-residue N-terminal fragment **trunc-Cys-N2** and negative control **Neg-N2** with Ala at N2 position. The underlined sequence fragments correspond to that depicted on the wheel diagram in B.

Charge patterning of the two segments of DSD scaffold was used to redesign homodimer into heterodimer in order to distinguish the two N-termini for their further functionalization. Thus, one strand was patterned with negatively charged residues (Glu) ('negative' strand), while another one with positively charged (Arg) ('positive' strand) (Figure 2 A and B). The Arg residues were used to substitute all Lys in the original DSD scaffold to prevent unwanted acylation of Lys side-chains. Subsequently, catalytic Cys residue was engineered into the 'negative' strand (containing excess of Glu) at N2 position, based on the most efficient stabilizing effects with pSer-containing peptides,[49] resulting in the **Cys-N2** peptide fragment. The sequences of the corresponding peptide fragments are depicted in Figure 2C. The heterodimer resulting from complexation of **Cys-N2** with 'positive' strand is abbreviated as **Het-N2**. Concurrently, three control scaffolds were designed

in order to test the importance of the N-terminal position of the catalytic cysteine and to evaluate the relevance of the entire structure for the catalytic activity. First, a second heterodimer (named as **Het-Ncap**) was designed where the 'negative' strand (abbreviated as **Cys-Ncap**) contained the Cys residue at Ncap position and the 'positive' strand remained the same as for **Het-N2** (sequences are shown in Figure 2D). Second, a truncated variant of **Cys-N2** strand was considered corresponding to its twelve N-terminal residues and is abbreviated as **trunc-Cys-N2** (see Figure 2D). Finally, a negative control **Neg-N2** heterodimer consisting of **Ala-N2** strand with Ala replacing Cys in **Cys-N2** and the 'positive' strand was engineered.

Synthesis and folding. The chemical synthesis of the 48-residue 'positive' and 'negative' strands was realized by '*in situ* neutralization' Boc/benzyl-SPPS,[52] whereas shorter 12-residue **trunc-Cys-N2** was synthesized by Fmoc/tBu-SPPS[53] as described in Supporting Information (Table S2). The assembly and folding of **Het-N2** construct and the controls **Het-Ncap** and **Neg-N2** proceeded spontaneously by dissolving the corresponding 'positive' and 'negative' strands in phosphate buffer at near neutral pH and mixing the solutions in a 1:1 ratio. Anticipated α-helical structure was confirmed by circular dichroism (CD) spectroscopy for **Het-N2** protein and stability was accessed by thermal denaturation by monitoring ellipticity at 222 nm, a strong indicator of α-helicity (Figures S1 and S2). Based on these measurements, in the absence of denaturant **Het-N2** remained folded from 5 to 90 °C. Furthermore, the high helical content was observed with up to 2 M Gn·HCl (guanidine hydrochloride) in pH 6.9 buffer solution at temperature below 40 °C (Figure S2). Increasing the concentration of denaturant led to lowering of an apparent melting temperature and at 6 M Gn·HCl the **Het-N2** was found to be unfolded. The dimeric composition of **Het-N2** was confirmed by analytical ultracentrifugation (AUC) (Figure S3).

To demonstrate that the dimer detected for the **Het-N2** sample by sedimentation velocity experiment corresponded to a heterodimer rather than a mixture of the corresponding homodimers we turned to urea denaturation (Figure S4). Urea was preferred as a chemical denaturant, because the electrostatic interactions between solvent-exposed sidechains were used to stabilize the heterodimer, and the favorability of these interactions would decrease significantly at the very high ionic strength required to denature the protein with Gn·HCl. CD-monitored denaturation curves for **Het-N2** as well as for **Cys-N2** and the 'positive' strand homodimers were therefore used to determine their thermodynamic stabilities (Figure S4 D). **Het-N2** proved to be approximately 10-times more stable than **Cys-N2** homodimer and 100-times more stable than the 'positive' strand homodimer. Furthermore, predominantly heterodimeric species were observed by native electrospray mass-spectrometry with minor quantities of homodimers (< 5%) (Figure S5).

Crystallographic structures show the successful design of the oxyanion hole. Finally, the designed **Het-N2** as well as the control **Het-Ncap** were crystallized. The X-ray data were collected at 2.33 Å (from anisotropic data, 4.43 Å in the worst direction) and 1.45 Å resolution for **Het-N2** and **Het-Ncap**, respectively, and the structures were solved by molecular replacement using the known structure of the DSD protein (PDB ID: 1G6U).[51] In the crystal structures both **Het-N2** and **Het-Ncap** were found to form the intended heterodimers with a three-α-helical bundle arrangement (Figure 3 and Figures S6-

S7). In comparison with the original crystal structure of the DSD protein, small shifts of the backbone were observed (on average, 1.839 Å of RMSD on Cα for **Het-N2** and 1.793 Å for **Het-Ncap**), resulting in changes of the coiled-coil pitch[54] from 217.7 Å for DSD protein to 199.2 Å for **Het-N2** and 177 Å for **Het-Ncap**. Substitution of Arg residues for Lys resulted in rotamer adjustments to maintain salt bridges between the two monomers.
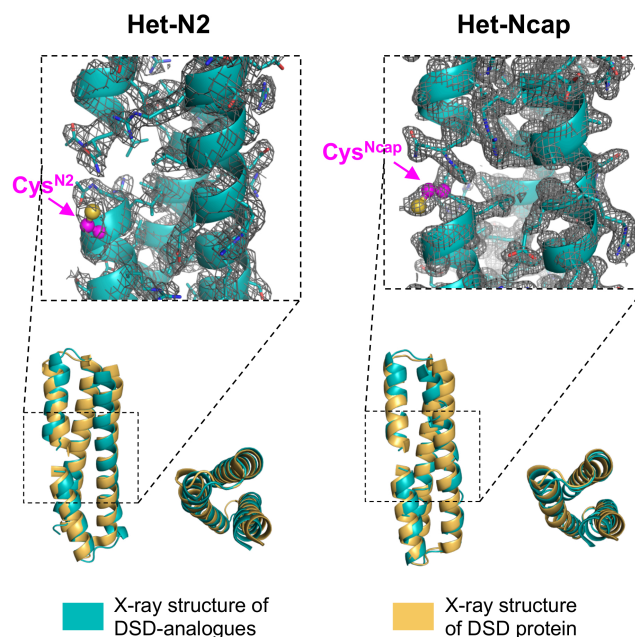

Figure 3. X-ray structures of the synthesized DSD-analogues **Het-N2** (left) and **Het-Ncap** (right). Catalytic sites are shown revealing solvent-exposed cysteine residues depicted in magenta with sulfur atom in yellow in ball-and-sticks representation. The simulated annealing $2F_o$-F_c electron density map is shown at 1.0σ level. The X-ray structures of **Het-N2** and **Het-Ncap** (depicted in cyan) are superimposed onto the original DSD structure (in yellow). PDB IDs: 6Z0L (**Het-N2**) and 6Z0M (**Het-Ncap**). Table S3 provides the refinement statistics. Figure S6 and Figure S7 show crystal packing and conformational variability of N-termini.
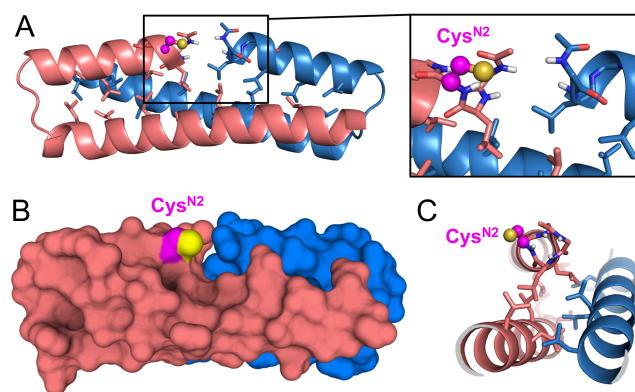

Figure 4. The active site in **Het-N2** is located at the interface of two N-termini of α-helices with nucleophilic Cys^{N2} positioned next to a small cavity. (A) Horizontal view illustrating helix-helix interface with the unpaired N-H groups pointing towards the cavity; (B) surface representation; (C) clipped section along the helical axis. Sulfur atom of Cys^{N2} is in yellow.
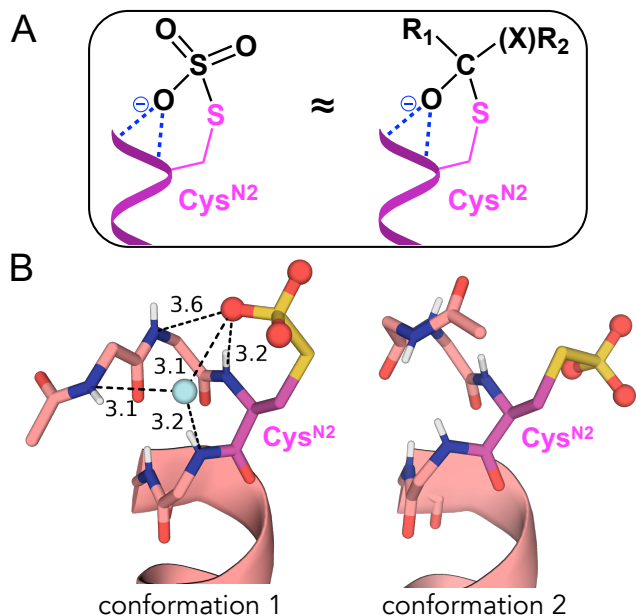
Figure 5. Cys-S-sulfonate mimicking tetrahedral intermediates. (A) Similarity of geometry and H-bonding interactions with catalytic protein for Cys-SO_3^- and tetrahedral intermediates for transthioesterification and amidation steps (X = S or NH_2^+). (B) Two distinct conformations of Cys-SO_3^- and helical N-termini observed in the crystal structure of **Het-N2-SO_3^-** (PDB ID 7BEY). Conformation 1 contains Cys-SO_3^- H-bonded to adjacent amides. In cyan is H_2O molecule mediating additional H-bonding interactions. H-bond donor-acceptor distances are shown in Å. More information is in Table S3, Figures S8 and S9.

Importantly, in **Het-N2** the Cys residue lies adjacent to a thin crevasse formed by the abutting N-terminal ends of the two helices (Figure 4). To validate the anticipated interactions of negatively charged tetrahedral intermediates with oxyanion hole, we solved an X-ray structure at 1.5 Å resolution for **Het-N2** construct where Cys^{N2} residue was converted to Cys-S-sulfonate to mimic the tetrahedral intermediates structures (Figure 5). Two **Het-N2** dimers were observed in an asymmetric unit differing in conformation of N-termini and orientation of sulfonate moiety. In the first conformation, one of the oxygens of sulfonate group is H-bonded to backbone amides, confirming the predictions shown in Figure 1. In the second conformation the sulfonate group points towards the solvent. Such conformational isomerism can be important for catalytic activity, providing key stabilizing interactions present in the first conformation and being available for reaction or facilitating product release illustrated by the second conformation. Thus, the **Het-N2-SO_3^-** structure validates the design and sets the groundwork for future optimization of the sequence surrounding the binding site.

Acyl-transferase activity. The evaluation of the acyl-transferase catalytic activity of **Het-N2** and its comparison to the controls (*i.e.* **Het-Ncap**, **trunc-Cys-N2**, **Neg-N2**, glutathione and also MPAA - 4-mercaptophenylacetic acid) were performed at fixed conditions (acyl donor, 200 µM; acyl acceptor, 200 mM; catalyst, 100 µM in 50 mM sodium phosphate buffer at pH 7.5) using liquid chromatography

electrospray ionization mass-spectrometry. Sequences and analytical data for acyl donor peptide-$^\alpha$thioester substrates are provided in Table S4. Tris (tris(hydroxymethyl)-aminomethane) was used as the standard acyl acceptor because of the pK_a 8.1 similar to N-terminal amine in peptides.[55] Background reaction (without protein catalyst) was also recorded. As an example, Figure 6 A shows kinetics of consumption of Ac-VALENF-$^\alpha$thioester substrate in the presence of different catalysts, and Figure 6 B illustrates the evolution of branched adduct resulting from the transthioesterification between the substrate-thioester and the catalytic cysteine. In the presence of **Het-N2**, peptide-thioester transthioesterified most rapidly forming peptide-catalytic protein branched thioester adduct accompanied by the release of Mes (2-mercaptoethanesulfonate).

The accumulation of this covalently bound Cys-thioester intermediate supports a ping-pong mechanism as expected in the hypothesized catalytic mechanism outlined in Scheme 1. For comparison, the same reaction performed with **Het-Ncap**, **trunc-Cys-N2** and Cys-containing glutathione proceeded much slower (Figure 6 A,B), did not occur with **Neg-N2**, and when MPAA was added at *c* 100 µM the product of thioester exchange was not observed. It should be mentioned that MPAA is a common aryl-thiol catalyst for native chemical ligation, however, used at much higher concentrations (10-200 mM).[56] The subsequent rate-limiting step involved amidation by Tris that occurred in competition with hydrolysis. Thus, the ratio of amidation over hydrolysis (A/H) and the observed pseudo first order rate constant (k_{obs}) calculated from the depletion of initial peptide-$^\alpha$thioester substrate concentration were used as quantifiers to characterize and compare reactivity of the catalysts at different conditions (Table 1).

For **Het-N2** construct showing appreciable catalytic effects in Figure 6 A, k_{obs} is $(1.1 \pm 0.1) \times 10^{-5}$ s^{-1} (at concentrations 200 µM for Ac-VALENF-$^\alpha$thioester and 100 µM for **Het-N2**), which is 5.5-times higher than k_{obs} $(2.0 \pm 0.3) \times 10^{-6}$ s^{-1} for the background reaction without catalyst. The other catalysts influenced the rate of the reaction less significantly with only 2-fold acceleration over background for **Het-Ncap** (Table S5), highlighting the importance of the N2 Cys position. Furthermore, Cys^{N2}-to-Ala substitution in negative control **Neg-N2** resulted in kinetics being similar to background level.

Next, for **Het-N2** the reaction was performed with different concentrations of Ac-VALENF-$^\alpha$thioester (from 0.2 to 2 mM) (examples of kinetic traces are displayed in Figure S10). The corrected initial rates of catalyzed peptide-$^\alpha$thioester consumption obtained by subtracting the observed initial rates of background reaction from catalyzed reaction were fitted to Michaelis-Menten equation (Figure 6 C). The estimated k_{cat} and K_m values are $(5.4 \pm 0.9) \times 10^{-5}$ s^{-1} and 0.3 ± 0.2 mM, respectively (standard error of the fitting is indicated), resulting in catalytic efficiency (k_{cat}/K_m) of ~ 0.2 M^{-1} s^{-1}. In addition, the catalytic turnover was confirmed by performing the experiment were fixed amounts of substrate (corresponding to starting concentration of 200 µM) were added two times sequentially (Figure S11). The **Het-N2** processed the two batches of substrate with similar efficiency.
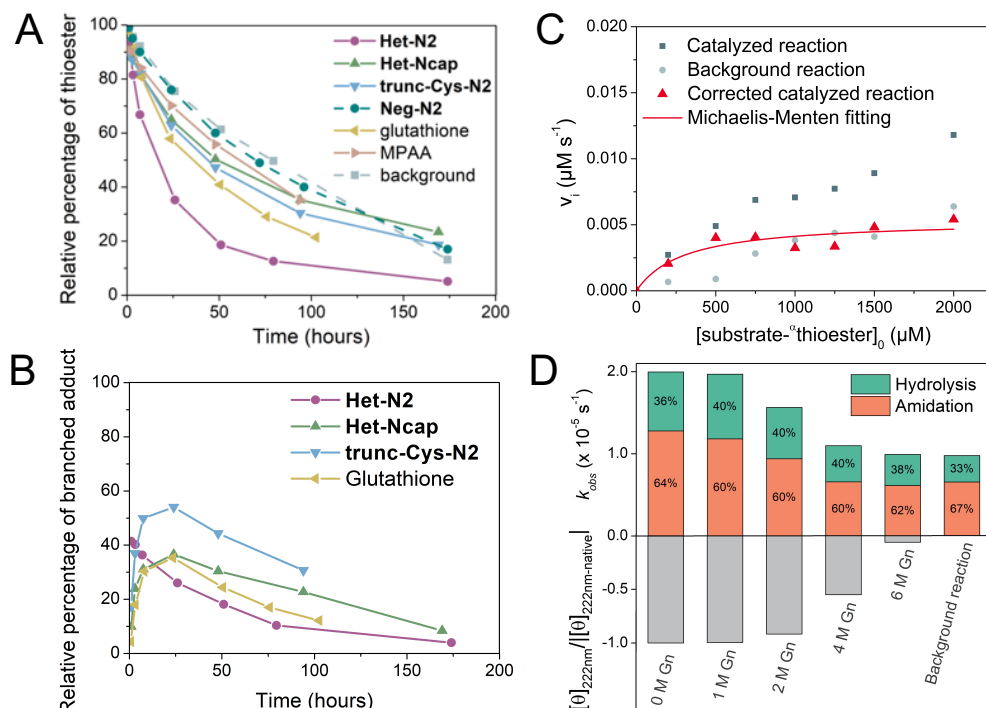
5

Figure 6. Characterization of catalytic activity. (A) Ac-VALENF-$^{\alpha}$thioester consumption in the presence of **Het-N2**, **Het-Ncap**, **trunc-N2-Cys**, glutathione and MPAA (4-mercaptophenylacetic acid) in comparison to background reaction and negative control **Neg-N2**. (B) Changes in concentration of peptide-protein branched thioester resulting from the transthioesterification between Ac-VALENF-$^{\alpha}$thioester and **Het-N2**, **Het-Ncap**, **trunc-Cys-N2** or glutathione. (C) Michaelis-Menten kinetics for the catalyzed reaction of Ac-VALENF-$^{\alpha}$thioester with **Het-N2**. (D) Observed rate constants and outcome of the reaction between Ac-GRLEEIDR-$^{\alpha}$thioester and Tris in presence of **Het-N2** as a function of the concentration of Gn·HCl and the degree of folding of **Het-N2** based on CD measurements.

The catalysis of peptide-$^{\alpha}$thioester amidation/hydrolysis by **Het-N2** can be compared with hydrolytic activity using a continuous assay with *para*-nitrophenyl acetate (*p*NPA) as a substrate,[57] which moreover was previously tested for other *de novo* designed protein catalysts.[20,21] The data obtained at pH 7 and 22 °C and presented in Figures S12-S14 and Table S6 allowed to obtain k_{cat} of 3.5×10^{-4} s^{-1} corresponding to rate enhancement k_{cat}/k_{uncat} of 138 and K_m of 1.4 mM resulting in catalytic efficiency of 0.25 M^{-1} s^{-1}, which is close to the value obtained with the thioester substrate. Compared to previously reported most efficient *de novo* hydrolases, the catalytic efficiency of **Het-N2** is nearly 10 times lower,[20,21] however, no additional catalytic residues (*e.g.* histidine) are present in **Het-N2**, thus the catalytic effect emerges mainly from the interactions with N-terminus of α-helix.

In addition to Ac-VALENF-$^{\alpha}$thioester, other thioester substrates with different sequences (Table S4) and C-terminal amino acids (Arg, Gly) were assayed. Both Arg and Gly thioesters were found to undergo faster reactions in the presence and absence of **Het-N2** than Ac-VALENF-$^{\alpha}$thioester (Table 1 and Figure S15), which is most likely caused by intramolecular catalytic effect of Arg side chain and favorable steric effect due to the absence of side chain in Gly. However, in comparison to Ac-VALENF-$^{\alpha}$thioester, the k_{obs} was only two times faster compared to the respective background reaction at standard reactant concentrations (Table 1). Another difference is the A/H ratio, which for Phe thioester was found to be 0.8 indicating hydrolysis to be a major outcome, whereas for Arg and Gly thioester A/H was in the range of 1.7-2.4, which corresponds to a more pronounced amidation. Two distinct sequences that were studied for Arg thioester showed similar results (Table 1)

suggesting that it is the C-terminal amino acid on the acyl donor that plays a decisive role in influencing the rate and its outcome with **Het-N2** catalyst.

The fact that the 12-residue **trunc-Cys-N2** is less efficient than **Het-N2** (Figure 6A) confirms that the overall stable structure of DSD is required for stabilization of oxyanion at N-terminus of α-helix. Using Ac-GRLEEIDR-$^{\alpha}$thioester and Tris as acyl acceptor we further proved the significance of the folded structure for the catalytic activity by increasing the concentration of a denaturant in the reaction mixture (Figure 6D). The k_{obs} was found to correlate to the degree of folding based on circular dichroism (CD) data at different Gn·HCl concentrations: when the protein **Het-N2** was highly α-helical the catalytic activity was at maximum, whereas at high concentration (6 M) of Gn·HCl causing unfolding of the protein the rate decreased to a background level. Same effect on the catalytic activity of **Het-N2** was also observed upon increasing the concentration of urea (Figure S16).

The pH of the buffer also modulated the rate of reaction and A/H ratio. Upon increasing the pH, the rate of Ac-GRLEEIDR-$^{\alpha}$thioester consumption in the presence of **Het-N2** increased with an inflection point at 8.6, which was similar for the reaction without protein catalyst (Figure S17). Maximum of amidation product was observed at pH 8 with a relative final product percentage of 66%. To rule out any effect of the buffer on catalysis, we performed assays with different concentrations of phosphate (*i.e.* 50, 200, 500 mM) and in two other buffer systems (50 mM MES or 50 mM MOPS) at pH 7.5. The catalytic activity of **Het-N2** remained similar in different buffer conditions with a slight increase of hydrolysis at higher concentration of phosphate (500 mM) as shown in Figure S18.

Table 1. Kinetic parameters for various acyl transfer reactions catalyzed by Het-N2 (100 µM). Additives were used for reaction between Tris (200 mM) and Ac-GRLEEIDR-$^\alpha$thioester (200 µM).

Amidation vs hydrolysis (A vs H)			
Acyl donor with Tris	k_{obs} (s^{-1})*	A/H ratio*	Acceleration**
Ac-GRLEEIDR-$^\alpha$SR	1.8×10^{-5}	2.4	2
Ac-VALENR-$^\alpha$SR	2.0×10^{-5}	2.0	1.8
Ac-VALENF-$^\alpha$SR	1.1×10^{-5}	0.8	5.5
Ac-LYRAG-$^\alpha$SR	2.4×10^{-5}	1.7	2.4
Acyl acceptor with Ac-GRLEEIDR-$^\alpha$SR	k_{obs} (s^{-1})*	A/H ratio*	Acceleration**
Tris	1.8×10^{-5}	2.4	2
MeONH$_2$	5.7×10^{-5}	2.3	3.2
MeNH$_2$	2.6×10^{-5}	0.4	3.1
Gly	1.5×10^{-5}	0.1	2.6
Gly-Gly	2.3×10^{-5}	0.15	2.9
Gly-Gly-Gly	2.5×10^{-5}	0.3	2.7
Histamine	9.6×10^{-5}	0.5	1.2
Esterification vs hydrolysis (E vs H)			
Acyl acceptor with Ac-GRLEEIDR-$^\alpha$SR	k_{obs} (s^{-1})	E/H ratio	Acceleration**
Trifluoroethanol	5.4×10^{-5}	Max.: 1.8	4.4
Methanol	2.2×10^{-5}	0.2	3
Ethanol	No reaction	-	-
Further acceleration			
Additives	k_{obs} (s^{-1})	A/H ratio	Acceleration†
NaN$_3$ (154 mM)	15.9×10^{-5}	0.1	8.8 (17.7)
KCN (200 mM)	15.2×10^{-5}	0.4	8.4 (16.9)
Imidazole (200 mM)	10.3×10^{-5}	1.0	5.7 (11.4)

*Estimated standard errors for k_{obs} and A/H are within 16% and 25% of their values, respectively

**Acceleration compared to background reaction without catalyst

†Acceleration compared to the reaction without additives and with catalyst, and in brackets to the reaction without additives and catalyst

Catalytic promiscuity. To test the ability of **Het-N2** to catalyze acyl transfer reactions other than amidation with Tris, we carried out the catalytic assays with several other acyl acceptors. Reactions of miscellaneous amines and few alcohols with Ac-GRLEEIDR-$^\alpha$thioester were studied (Table 1). Methoxyamine with lower pK$_a$ of 4.7 provided with the fastest kinetics (k_{obs} 5.7×10^{-5} s^{-1}) that represents more than three times acceleration over background and amidation was more favored (A/H = 2.33). In this particular case, the rate enhancement is likely resulting from α-effect well-known for methoxyamine, where nucleophilicity of amine is increased due to the adjacent oxygen atom containing lone electron pairs.[58] In contrast, substrates with higher pK$_a$ like methylamine (10.6) resulted in the excess of hydrolyzed product (A/H = 0.4) even if the rate of thioester consumption (k_{obs} 2.6×10^{-5} s^{-1}) was similar to the reaction with Tris. We also investigated catalysis of peptide bond formation by **Het-N2** using glycine (pK$_a$ = 9.6), a dipeptide Gly-Gly (pK$_a$ = 8.2) and a tripeptide Gly-Gly-Gly

(pK$_a$ = 7.8)[59] as acyl acceptors. The Ac-GRLEEIDR-$^\alpha$thioester was less reactive with glycine and more prone to hydrolysis than with dipeptide (A/H = 0.1 and 0.15, respectively) and tripeptide (A/H = 0.3), illustrating an influence of the pK$_a$ of acyl acceptor on the reactivity (Table 1). A peculiar change of reactivity was noticed for histamine (pK$_a$ (amine) = 9.8 and pK$_a$ (imidazole) = 6.04). A rather high observed rate constant was detected (k_{obs} 9.6×10^{-5} s^{-1}), however, in the background reaction without **Het-N2** the consumption of thioester was almost as fast and amidation was more pronounced compared to the reaction with **Het-N2** (A/H for uncatalyzed reaction is 0.9 versus 0.5 for catalyzed). This may indicate an additional catalytic effect of the imidazole ring on the initial thioester that is more accessible in solution than the Cys-thioester intermediate covalently bound to **Het-N2**. Influence of imidazole used as separate additive (*i.e.* in synergy with the catalytic protein) will be presented hereinafter.

Alcohols as acyl acceptors demonstrated moderate reactivity (Table 1). Esterification with trifluoroethanol, an alcohol with a reduced pK$_a$ of 12.4, was observed in the presence of catalytic protein **Het-N2** with k_{obs} = 5.4×10^{-5} s^{-1} and up to 61 % of detected ester formation compared to k_{obs} = 1.2×10^{-5} s^{-1} and 39 % of ester for background reaction, but the resulting ester product was not stable and decomposed through hydrolysis after reaching a maximum of ester production. Methanol (pK$_a$ = 15.2) reached a conversion of 17 % of ester product with a three-fold faster rate compared to the uncatalyzed reaction. Interestingly, ethanol (pK$_a$ = 15.9) did not react in the acyl transfer with Ac-GRLEEIDR-$^\alpha$thioester.

Furthermore, we demonstrated that addition of azide and cyanide was able to modify kinetics of the acyl transfer reaction between Ac-GRLEEIDR-$^\alpha$thioester and Tris. The increasing concentrations of these additives with and without catalytic proteins led to higher observed rates of reactions and higher percentage of hydrolysis (see Table 1 and Figure S19). The most contrasting effect was discovered in the presence of azide, where the high concentration of NaN$_3$ (154 mM) in synergy with **Het-N2** contributed to eighteen-fold acceleration compared to background reaction (without additives and catalyst) and a nine-fold acceleration compared to conditions without additives and with catalyst. However, the A/H ratio dropped from 2.4 to 0.1. Similar tendency was observed with KCN in the reaction mixture (Figure S18). Presumably, transient acyl azides and acyl cyanides are the reactive intermediates that facilitate hydrolysis,[60,61] although, these intermediates were not directly detected by LC-MS.

After observing the unusual reactivity of histamine as acyl acceptor, we evaluated the effect of imidazole as an additive. When imidazole was at low concentrations (2 and 20 mM) no considerable reactivity difference was observed, whereas at higher concentration (200 mM), observed rate constant increased nearly 6-fold to 10.3×10^{-5} s^{-1} with amidation being more favored compared to with the other additives with a final A/H ratio of 1.0. Previously, imidazole added at high concentrations (2.5 M) was reported to catalyze native chemical ligation.[62]

CONCLUSIONS

We describe the successful completion of the first step towards the design of a stable catalyst for thioester ligation. Our initial goals were to build a protein with a small active site cavity capable of activating a Cys thiol for attack of thioesters, and

stabilizing the anionic high-energy tetrahedral intermediates in the reaction mechanism. The designed **Het-N2** was found to modestly accelerate a variety of acyl transfer reactions with a wide range of substrates. Thus, **Het-N2** can be compared to primitive (or primordial) enzymes possessing moderate catalytic efficiency and substrate promiscuity.[63] Such modest catalytic efficiency coupled to catalytic promiscuity is believed to be essential for evolution of enzymes with distinct specificities.[63] The next step is to apply computational design and combinatorial library approaches taking advantage of the modular architecture of **Het-N2** scaffold to vary the composition at the active site and improve its catalytic properties.

The successful design of a catalyst also requires that the multiple chemical steps along the reaction path be understood, so that they can be optimized separately as needed. Enabling selective catalysis of amidation versus hydrolysis by design is particularly challenging aspect. Recently, several X-ray structures were reported for proteolytic enzymes which also possess peptide macrocyclization or ligation activities and provide insights into how such selectivity can be achieved.[64-69] The principal catalytic residue is either Ser[64,65] or Cys[66-69] and catalysis is assisted by few other residues such as His, Asp/Glu, Tyr and backbone amides. The enhancement of ligation activity over proteolysis is believed to occur by exclusion of water from the acyl-enzyme covalent adduct via additional structural elements such as "capping" lid domain,[64] binding of follower peptide[65] or hydrophobicity of the active site.[66-69] Remarkably, structural comparison of the enzymes with ligation (amidation) activity with the homologous variants possessing enhanced proteolytic (hydrolysis) activity showed that the structures of the related ligases and proteases are very similar with only subtle changes observed in the catalytic pocket.[65-69] Furthermore, minimal changes in active site composition are sufficient to repurpose a ligase into a protease and vice versa.[66-69]

Thus, it is not surprising that the branched thioester adduct observed in our work where electrophilic thioester moiety is exposed to aqueous buffer undergoes fast hydrolysis in competition with amidation. The choice to use Cys and not Ser as the principle nucleophilic residue is not sufficient to bias the reactivity towards the reaction with amines, despite the reported higher relative reaction rates of thioesters towards nitrogen nucleophiles than oxygen nucleophiles in comparison to oxoesters.[42,70] A possible way to further enhance selectivity towards amidation would be to incorporate selenocysteine as a principle catalytic residue since selenoesters are known to be ca. 100-times more reactive with amines than thioesters whereas hydrolysis rates are on the same order of magnitude,[71,72] however, it would require inert oxygen-free conditions due to the associated selenol's redox chemistry.[73]

Our data indicate that the environment nearby catalytic Cys can slightly modify the amidation versus hydrolysis ratio. Thus, with the main goal to achieve higher catalytic activities, our future work will be directed towards understanding the structural requirements for amidation versus hydrolysis by optimizing the sequence of the N-terminal residues of the two helices for a particular reaction path and developing new practical tools for peptide ligation and protein modification.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Materials and methods, details on chemical synthesis, crystallization, data collection and structure refinement, catalytic assays, supplementary tables and figures (PDF)

AUTHOR INFORMATION

Corresponding Authors

Vladimir Torbeev - Institut de Science et d'Ingénierie Supramoléculaires (ISIS), International Center for Frontier Research in Chemistry (icFRC), University of Strasbourg, CNRS (UMR 7006), Strasbourg, France; Email: torbeev@unistra.fr

William F. DeGrado - Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, USA; Email: william.degrado@ucsf.edu

Authors

Elise A. Naudin - Institut de Science et d'Ingénierie Supramoléculaires (ISIS), International Center for Frontier Research in Chemistry (icFRC), University of Strasbourg, CNRS (UMR 7006), Strasbourg, France

Alastair G. McEwen - Integrated Structural Biology Platform, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM (U1258), University of Strasbourg, CNRS (UMR 7104), Illkirch, France

Sophia K. Tan - Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, USA

Pierre Poussin-Courmontagne - Integrated Structural Biology Platform, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM (U1258), University of Strasbourg, CNRS (UMR 7104), Illkirch, France

Jean-Louis Schmitt - Institut de Science et d'Ingénierie Supramoléculaires (ISIS), International Center for Frontier Research in Chemistry (icFRC), University of Strasbourg, CNRS (UMR 7006), Strasbourg, France

Catherine Birck - Integrated Structural Biology Platform, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM (U1258), University of Strasbourg, CNRS (UMR 7104), Illkirch, France

Author Contributions

E. A. N., A. G. M., P. P.-C., J.-L. S., C. B., V. T. performed the experiments, E. A. N., A. G. M., S. K. T., C. B. analyzed the data, E. A. N., W. F. D., V. T. wrote the manuscript with the consultation from other authors.

Notes

The authors declare no competing financial interest.
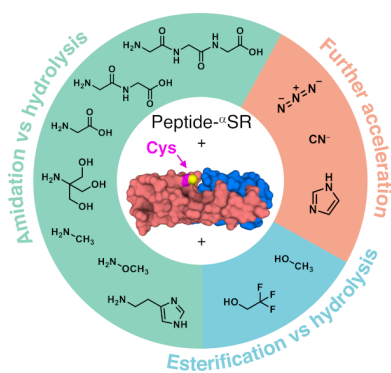
ACKNOWLEDGMENTS

REFERENCES

(1) Zanghellini, A. De Novo Computational Enzyme Design. *Curr. Opin. Biotechnol.* **2014**, *29*, 132-138.

(2) Kries, H.; Blomberg, R.; Hilvert, D. De Novo Enzymes by Computational Design. *Curr. Opin. Chem. Biol.* **2013**, *17*, 221-228.

(3) Hilvert, D. Design of Protein Catalysts. *Annu. Rev. Biochem.* **2013**, *82*, 447-470.

(4) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008**, *319*, 1387-1391.

(5) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* **2008**, *453*, 190-195.

(6) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St.Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **2010**, *329*, 309-313.

(7) Bolon, D. N.; Mayo, S. L. Enzyme-like Proteins by Computational Design. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14274-14279.

(8) Richter, F.; Blomberg, R.; Khare, S. D.; Kiss, G.; Kuzin, A. P.; Smith, A. J. T.; Gallaher, J.; Pianowski, Z.; Helgeson, R. C.; Grjasnow, A.; Xiao, R.; Seetharaman, J.; Su, M.; Vorobiev, S.; Lew, S.; Forouhar, F.; Kornhaber, G. J.; Hunt, J. F.; Montelione, G. T.; Tong, L.; Houk, K. N.; Hilvert, D.; Baker, D. Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis. *J. Am. Chem. Soc.* **2012**, *134*, 16197-16206.

(9) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational Enzyme Design. *Angew. Chem. Int. Ed.* **2013**, *52*, 5700-5725.

(10) Korendovych, I. V.; DeGrado, W. F. Catalytic Efficiency of Designed Catalytic Proteins. *Curr. Opin. Struct. Biol.* **2014**, *27*, 113-121.

(11) Zeymer, C.; Hilvert, D. Directed Evolution of Protein Catalysts. *Annu. Rev. Biochem.* **2018**, *87*, 131-157.

(12) Bunzel, H. A.; Garrabou, X.; Pott, M.; Hilvert, D. Speeding up Enzyme Discovery and Engineering with Ultrahigh-Throughput Methods. *Curr. Opin. Struct. Biol.* **2018**, *48*, 149-156.

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

(14) Nanda, V.; Koder, R. L. Designing Artificial Enzymes by Intuition and Computation. *Nat. Chem.* **2010**, *2*, 15-24.

(15) Huang, P.-S.; Boyken, S. E.; Baker, D. The Coming of Age of de Novo Protein Design. *Nature* **2016**, *537*, 320-327.

(16) Korendovych, I. V.; DeGrado, W. F. De Novo Protein Design, a Retrospective. *Q. Rev. Biophys.* **2020**, *53*, e3.

(17) Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Grütter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D. Precision Is Essential for Efficient Catalysis in an Evolved Kemp Eliminase. *Nature* **2013**, *503*, 418-421.

(18) Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D. Evolution of a Designed Retro-Aldolase Leads to Complete Active Site Remodeling. *Nat. Chem. Biol.* **2013**, *9*, 494-498.

(19) Preiswerk, N.; Beck, T.; Schulz, J. D.; Milovník, P.; Mayer, C.; Siegel, J. B.; Baker, D.; Hilvert, D. Impact of Scaffold Rigidity on the Design and Evolution of an Artificial Diels-Alderase. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 8013-8018.

(20) Marshall, L. R.; Zozulia, O.; Lengyel-Zhand, Z.; Korendovych, I. V. Minimalist de Novo Design of Protein Catalysts. *ACS Catal.* **2019**, *9*, 9265-9275.

(21) Burton, A. J.; Thomson, A. R.; Dawson, W. M.; Brady, R. L.; Woolfson, D. N. Installing Hydrolytic Activity into a Completely de Novo Protein Framework. *Nat. Chem.* **2016**, *8*, 837-844.

(22) Watkins, D. W.; Jenkins, J. M. X.; Grayson, K. J.; Wood, N.; Steventon, J. W.; Le Vay, K. K.; Goodwin, M. I.; Mullen, A. S.; Bailey, H. J.; Crump, M. P.; MacMillan, F.; Mulholland, A. J.; Cameron, G.; Sessions, R. B.; Mann, S.; Anderson, J. L. R. Construction and in Vivo Assembly of a Catalytically Proficient and Hyperthermostable de Novo Enzyme. *Nat. Commun.* **2017**, *8*, 358.

(23) Moroz, Y. S.; Dunston, T. T.; Makhlynets, O. V.; Moroz, O. V.; Wu, Y.; Yoon, J. H.; Olsen, A. B.; McLaughlin, J. M.; Mack, K. L.; Gosavi, P. M.; van Nuland, N. A. J.; Korendovych, I. V. New Tricks for Old Proteins: Single Mutations in a Nonenzymatic Protein Give Rise to Various Enzymatic Activities. *J. Am. Chem. Soc.* **2015**, *137*, 14905-14911.

(24) Moroz, O. V.; Moroz, Y. S.; Wu, Y.; Olsen, A. B.; Cheng, H.; Mack, K. L.; McLaughlin, J. M.; Raymond, E. A.; Zhezherya, K.; Roder, H.; Korendovych, I. V. A Single Mutation in a Regulatory Protein Produces Evolvable Allosterically Regulated Catalyst of Nonnatural Reaction. *Angew. Chem. Int. Ed.* **2013**, *52*, 6246-6249.

(25) Kent, S. B. H. Novel Protein Science Enabled by Total Chemical Synthesis. *Protein Sci.* **2019**, *28*, 313-328.

(26) Bondalapati, S.; Jbara, M.; Brik, A. Expanding the Chemical Toolbox for the Synthesis of Large and Uniquely Modified Proteins. *Nat. Chem.* **2016**, *8*, 407-418.

(27) Dang, B.; Wu, H.; Mulligan, V. K.; Mravic, M.; Wu, Y.; Lemmin, T.; Ford, A.; Silva, D.-A.; Baker, D.; DeGrado, W. F. De Novo Design of Covalently Constrained Mesosize Protein Scaffolds with Unique Tertiary Structures. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 10852-10857.

(28) Mandal, K.; Uppalapati, M.; Ault-Riché, D.; Kenney, J.; Lowitz, J.; Sidhu, S. S.; Kent, S. B. H. Chemical Synthesis and X-Ray Structure of a Heterochiral {D-Protein Antagonist plus Vascular Endothelial Growth Factor} Protein Complex by Racemic Crystallography. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14779-14784.

(29) Weinstock, M. T.; Jacobsen, M. T.; Kay, M. S. Synthesis and Folding of a Mirror-Image Enzyme Reveals Ambidextrous Chaperone Activity. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 11679-11684.

(30) Mandal, K.; Pentelute, B. L.; Bang, D.; Gates, Z. P.; Torbeev, V. Y.; Kent, S. B. H. Design, Total Chemical Synthesis, and X-Ray Structure of a Protein Having a Novel Linear-Loop Polypeptide Chain Topology. *Angew. Chem. Int. Ed.* **2012**, *51*, 1481-1486.

(31) Boehringer, R.; Kieffer, B.; Torbeev, V. Total Chemical Synthesis and Biophysical Properties of a Designed Soluble 24 KDa Amyloid Analogue. *Chem. Sci.* **2018**, *9*, 5594-5599.

(32) Kochendoerfer, G. G., Chen, S. Y.; Mao, F.; Cressman, S.; Traviglia, S.; Shao, H.; Hunter, C. L.; Low, D. W.; Cagle, E. N.; Carnevali, M.; Gueriguian, V.; Keogh, P. J.; Porter, H.; Stratton, S. M.; Wiedeke, M. C.; Wilken, J.; Tang, J.; Levy, J. J.; Miranda, L. P.; Crnogorac, M. M.; Kalbag, S.; Botti, P.; Schindler-Horvat, J.; Savatski, L.; Adamson, J. W.; Kung, A.; Kent, S. B.; Bradburne, J. A. Design and Chemical Synthesis of a Homogeneous Polymer-modified Erythropoiesis Protein. *Science* **2003**, *299*, 884-887.

(33) Nguyen, G. K. T.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J. P. Butelase 1 Is an Asx-Specific Ligase Enabling Peptide Macrocyclization and Synthesis. *Nat. Chem. Biol.* **2014**, *10*, 732-738.

(34) Piotukh, K.; Geltinger, B.; Heinrich, N.; Gerth, F.; Beyermann, M.; Freund, C.; Schwarzer, D. Directed Evolution of Sortase A Mutants with Altered Substrate Selectivity Profiles. *J. Am. Chem. Soc.* **2011**, *133*, 17536-17539.

(35) Zheng, N.; Shabek, N. Ubiquitin Ligases: Structure, Function, and Regulation. *Annu. Rev. Biochem.* **2017**, *86*, 129-157.

(36) Weeks, A. M.; Wells, J. A. Subtiligase-Catalyzed Peptide Ligation. *Chem. Rev.* **2019**, *120*, 3127-3160.

(37) Jackson, D. Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J. A. A Designed Peptide Ligase for Total Synthesis of Ribonuclease A with Unnatural Catalytic Residues. *Science* **1994**, *266*, 243-247.

(38) Toplak, A.; Nuijens, T.; Quaedflieg, P. J. L. M.; Wu, B.; Janssen, D. B. Peptiligase, an Enzyme for Efficient Chemoenzymatic Peptide Synthesis and Cyclization in Water. *Adv. Synth. Catal.* **2016**, *358*, 2140-2147.

(39) Schmidt, M.; Toplak, A.; Quaedflieg, P. J. L. M.; Ippel, H.; Richelle, G. J. J.; Hackeng, T. M.; van Maarseveen, J. H.; Nuijens, T. Omniligase-1: A Powerful Tool for Peptide Head-to-Tail Cyclization. *Adv. Synth. Catal.* **2017**, *359*, 2050-2055.

(40) Simón, L.; Goodman, J. M. Enzyme Catalysis by Hydrogen Bonds: The Balance between Transition State Binding and Substrate Binding in Oxyanion Holes. *J. Org. Chem.* **2010**, *75*, 1831-1840.

(41) Poole, L. B. The Basics of Thiols and Cysteines in Redox Biology and Chemistry. *Free Radic. Biol. Med.* **2015**, *80*, 148-157.

(42) Yang, W.; Drueckhammer, D. G. Understanding the Relative Acyl-Transfer Reactivity of Oxoesters and Thioesters: Computational Analysis of Transition State Delocalization Effects. *J. Am. Chem. Soc.* **2001**, *123*, 11004-11009.

(43) Finking, R.; Marahiel, M. A. Biosynthesis of Nonribosomal Peptides. *Annu. Rev. Microbiol.* **2004**, *58*, 453-488.

(44) Smith, S.; Witkowski, A.; Joshi, A. K. Structural and Functional Organization of the Animal Fatty Acid Synthase. *Prog. Lipid Res.* **2003**, *42*, 289-317.

(45) Perler, F. B. Protein Splicing Mechanisms and Applications. *IUBMB Life* **2005**, *57*, 469-476.

(46) Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoet, M.; Monbaliu, J.-C. M.; Melnyk, O. Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* **2019**, *119*, 7328-7442.

(47) Hojo, H.; Aimoto, S. Polypeptide Synthesis Using the S-Alkyl Thioester of a Partially Protected Peptide Segment. Synthesis of the DNA-Binding Domain of c-Myb Protein (142–193)-NH$_2$. *Bull. Chem. Soc. Jpn.* **1991**, *64*, 111-117.

(48) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266*, 776-779.

(49) Andrew, C. D.; Warwicker, J.; Jones, G. R.; Doig, A. J. Effect of Phosphorylation on α-Helix Stability as a Function of Position. *Biochemistry* **2002**, *41*, 1897-1905.

(50) Tan, S. K.; Fong, K. P.; Polizzi, N. F.; Sternisha, A.; Slusky, J. S. G.; Yoon, K.; DeGrado, W. F.; Bennett, J. S. Modulating Integrin αIIbβ3 Activity through Mutagenesis of Allosterically Regulated Intersubunit Contacts. *Biochemistry* **2019**, *58*, 3251-3259.

(51) Ogihara, N. L.; Ghirlanda, G.; Bryson, J. W.; Gingery, M.; DeGrado, W. F.; Eisenberg, D. Design of Three-Dimensional Domain-Swapped Dimers and Fibrous Oligomers. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 1404-1409.

(52) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. In Situ Neutralization in Boc-Chemistry Solid Phase Peptide Synthesis. *Int. J. Pept. Res. Ther.* **2007**, *13*, 31-44.

(53) Behrendt, R.; White, P.; Offer, J. Advances in Fmoc Solid-Phase Peptide Synthesis. *J. Pept. Sci.* **2016**, *22*, 4-27.

(54) Grigoryan, G.; DeGrado, W. F. Probing Designability via a Generalized Model of Helical Bundle Geometry. *J. Mol. Biol.* **2011**, *405*, 1079-1100.

(55) Sievers, A.; Beringer, M.; Rodnina, M. V.; Wolfenden, R. The Ribosome as an Entropy Trap. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7897-7901.

(56) Johnson, E. C. B.; Kent, S. B. H. Insights into the Mechanism and Catalysis of the Native Chemical Ligation Reaction. *J. Am. Chem. Soc.* **2006**, *128*, 6640-6646.

(57) Bender, M. L.; Kezdy, F. J.; Wedler, F. C. Alpha-Chymotrypsin: Enzyme Concentration and Kinetics. *J. Chem. Educ.* **1967**, *44*, 84-88.

(58) Jencks, W. P.; Carriuolo, J. Reactivity of Nucleophilic Reagents toward Esters. *J. Am. Chem. Soc.* **1960**, *82*, 1778-1786.

(59) Daragan, V. A.; Mayo, K. H. Peptide Dynamics in Triglycine: Coupling of Internal Bond Rotations and Overall Molecular Tumbling. *J. Phys. Chem.* **1994**, *98*, 10949-10956.

(60) Hünig, S.; Schaller, R. The Chemistry of Acyl Cyanides. *Angew. Chem. Int. Ed.* **1982**, *21*, 36-49.

(61) Eggerer, H.; Stadtman, E. R.; Poston, J. M. On the Role of Acetyl Cyanide in the Cyanide-Induced Acetylation of Amino Acids by Enzymes of Clostridium Kluyveri. *Arch. Biochem. Biophys.* **1962**, *98*, 432-443.

(62) Sakamoto, K.; Tsuda, S.; Mochizuki, M.; Nohara, Y.; Nishio, H.; Yoshiya, T. Imidazole-Aided Native Chemical Ligation: Imidazole as a One-Pot Desulfurization-Amenable Non-Thiol-Type Alternative to 4-Mercaptophenylacetic Acid. *Chem. Eur. J.* **2016**, *22*, 17940-17944.

(63) Khersonsky, O.; Roodveldt, C.; Tawfik, D. S. Enzyme Promiscuity: Evolutionary and Mechanistic Aspects. *Curr. Opin. Chem. Biol.* **2006**, *10*, 498-508.

(64) Koehnke, J.; Bent, A.; Houssen, W. E.; Zollman, D.; Morawitz, F.; Shirran, S.; Vendome, J.; Nneoyiegbe, A. F.; Trembleau, L.; Botting, C. H.; Smith, M. C. M.; Jaspars, M.; Naismith, J. H. The Mechanism of Patellamide Macrocyclization Revealed by the Characterization of the PatG Macrocyclase Domain. *Nat. Struct. Mol. Biol.* **2012**, *19*, 767-773.

(65) Chekan, J. R.; Estrada, P.; Covello, P. S.; Nair, S. K. Characterization of the Macrocyclase Involved in the Biosynthesis of RiPP Cyclic Peptides in Plants. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 6551-6556.

(66) Zauner, F. B.; Elsässer, B.; Dall, E.; Cabrele, C.; Brandstetter, H. Structural Analyses of Arabidopsis Thaliana Legumain G Reveal Differential Recognition and Processing of Proteolysis and Ligation Substrates. *J. Biol. Chem.* **2018**, *293*, 8934-8946.

(67) Haywood, J.; Schmidberger, J. W.; James, A. M.; Nonis, S. G.; Sukhoverkov, K. V.; Elias, M.; Bond, C. S.; Mylne, J. S. Structural Basis of Ribosomal Peptide Macrocyclization in Plants. *eLife* **2018**, *7*, e32955.

(68) Hemu, X.; El Sahili, A.; Hu, S.; Wong, K.; Chen, Y.; Wong, Y. H.; Zhang, X.; Serra, A.; Goh, B. C.; Darwis, D. A.; Chen, M. W.; Sze, S. K.; Liu, C. F.; Lescar, J.; Tam, J. P. Structural Determinants for Peptide-bond Formation by Asparaginyl Ligases. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 11737-11746.

(69) James, A. M.; Haywood, J.; Leroux, J.; Ignasiak, K.; Elliott, A. G.; Schmidberger, J. W.; Fisher, M. F.; Nonis, S. G.; Fenske, R.; Bond, C. S.; Mylne, J. S. The Macrocyclizing Protease Butelase 1 Remains Autocatalytic and Reveals the Structural Basis for Ligase Activity. *Plant J.* **2019**, *98*, 988-999.

(70) Connors, K. A.; Bender, M. L. The Kinetics of Alkaline Hydrolysis and n-Butylaminolysis of Ethyl p-Nitrobenzoate and Ethyl p-Nitrothiolbenzoate. *J. Org. Chem.* **1960**, *26*, 2498-2504.

(71) Chu, S.-H.; Mautner, H. G. Analogs of Neuroeffectors. V. Neighboring-Group Effects in the Reactions of Esters, Thiolesters, and Selenolesters. The Hydrolysis and Aminolysis of Benzoylcholine, Benzoylthiolcholine, Benzoylselenolcholine, and of their Dimethyla-mino Analogs. *J. Org. Chem.* **1966**, *31*, 308−312.

(72) Wu, Z.-P.; Hilvert, D. Conversion of a Protease into an Acyl Transferase: Selenolsubtilisin. *J. Am. Chem. Soc.* **1989**, *111*, 4513-4514.

(73) Wu, Z.-P.; Hilvert, D. Selenosubtilisin as a Glutathione Peroxidase Mimic. *J. Am. Chem. Soc.* **1990**, *112*, 5647-5648.

Table of contents artwork

Acyl transfer catalytic activity in *de novo* designed protein with N-terminus of α-helix as oxyanion-binding site

Elise A. Naudin,[1] Alastair G. McEwen,[2] Sophia K. Tan,[3] Pierre Poussin-Courmontagne,[2] Jean-Louis Schmitt,[1] Catherine Birck,[2] William F. DeGrado[3]* and Vladimir Torbeev[1]*

[1]Institut de Science et d'Ingénierie Supramoléculaires (ISIS), International Center for Frontier Research in Chemistry (icFRC), University of Strasbourg, CNRS (UMR 7006), Strasbourg, France

[2]Integrated Structural Biology Platform, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS (UMR 7104), INSERM (U1258), University of Strasbourg, Illkirch, France

[3]Department of Pharmaceutical Chemistry and the Cardiovascular Research Institute, University of California at San Francisco, San Francisco, California, United States

*Correspondence should be addressed to V. T. (torbeev@unistra.fr) or W. F. D. (william.degrado@ucsf.edu)

Table of contents:

Abbreviations:

2-CT: 2-chlorotrityl

4-MeBzl: 4-methylbenzyl

Boc: *tert*-butyloxycarbonyl

Bzl: Benzyl

c: concentration

CD: Circular Dichroism

DCM: Dichloromethane

DIEA: Diisopropylethylamine

DMF: *N,N*-Dimethylformamide

DSD: Domain-Swapped Dimer

ESI: Electrospray Ionization

EtOH: Ethanol

Eq: Equivalent

Fmoc: Fluorenylmethyloxycarbonyl

Gn: Guanidine

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

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

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HPLC: High Performance Liquid Chromatography

k_{cat}: turnover number

K_m: Michaelis-Menten constant

k_{obs}: Observed rate constant

LC-MS: Liquid Chromatography-Mass Spectrometry

MeCN: Acetonitrile

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

Mes: 2-Mercaptoethanesulfonate

MesNa: Sodium 2-mercaptoethanesulfonate

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

OcHex: Cyclohexyl ester

Pbf: 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl

PEG: Polyethylene glycol

PEG 550 MME: Polyethylene glycol monomethyl ether 550

*p*NP: *para*-nitrophenol

*p*NPA: *para*-nitrophenylacetate

RMSD: Root-Mean-Square Deviation

RPM: Revolutions per minute

SPPS: Solid Phase Peptide Synthesis

tBu: *tert*-butyl

TCEP: Tris(2-carboxyethyl)phosphine

TFA: Trifluoroacetic acid

TIPS: Triisopropylsilane

Tos: Tosyl

Tris: Tris(hydroxymethyl)aminomethane

Trt: Trityl

UHPLC: Ultra High Performance Liquid Chromatography

Xan: Xanthyl

1. Materials and methods

Reagents:

All solvents, chemicals, and reagents were purchased from commercial sources and used without further purification. Coupling reagents (1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate, HATU, and 3-[bis(dimethylamino)methyliumyl]-3H-benzotriazol-1-oxide hexafluorophosphate, HBTU), Fmoc-α-*L*-amino acids, resins for SPPS (solid-phase peptide synthesis) and DMF were purchased from Iris Biotech (Marktredwitz, Germany); Boc-α-*L*-amino acids were from Aapptec (Louisville, USA); diisopropylethylamine (DIEA) and 4-methylpiperidine were from Sigma-Aldrich; trifluoroacetic acid (TFA, BioGrade) was from Halocarbon (Peachtree Corners, USA); hydrogen fluoride was from GHC Gerling Holz & Co Handels GmbH (Hanau, Germany). All other chemicals were purchased from Sigma-Aldrich. Reactions involving compounds that are sensitive to oxidation or hydrolysis were performed under an argon atmosphere (*i.e.* thioesterification).

LC-MS:

Peptide masses were measured on a LC-MS instrument equipped with a Thermo Scientific Accela UHPLC (Hypers II GOLD C18 column, 50×2.1 mm^2, 1.9 µm) integrated with a Thermo Scientific LCQ Fleet ion-trap. Deconvolution of data was performed in MagTran 1.03 (Amgen, Thousand Oaks, USA).

Analytical High Performance Liquid Chromatography (HPLC):

Analytical reversed-phase HPLC was performed on a Thermo Scientific Dionex Ultimate 3000 UHPLC instrument. A Phenomenex Kinetex EVO C18 column (50×2.1 mm^2, 100 Å, 2.6 µm) was used at 1 mL/min with a gradient of water with TFA (0.1 %, *v/v*) and acetonitrile with TFA (0.08 %, *v/v*).

Semi-preparative HPLC:

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

Analytical Ultracentrifugation:

Sedimentation velocity experiments were performed on a Beckman Coulter ProteomeLab XL-I ultracentrifuge at 20°C and 55,000 RPM using an An55-Ti rotor. 'Positive' strand and **Cys-N2** peptides were dialyzed against a buffer containing 50 mM sodium phosphate pH 7.5, 100 mM NaCl and 1 mM TCEP. Concentrations were determined by measuring absorbance with Thermo Scientific NanoDrop One instrument and volumes were adjusted to have three solutions of 'positive' strand, **Cys-N2** peptide and **Het-N2** heterodimer (a 1:1 mixture of 'positive' strand and **Cys-N2**) at 200 µM. 400 µL of each sample at 200 µM concentration were loaded in 1.2 cm double sector cells with quartz windows. Absorbance scans at 240 nm were taken every 3 minutes for 17 h. Sedimentation data were analyzed using SEDFIT software[1] (version 16.1) and the continuous sedimentation coefficient distribution model c(s). Buffer density, buffer viscosity and peptides partial specific volumes were calculated using SEDNTERP software.[2] GUSSI[3] was used to integrate the sedimentation peaks and to produce the graphs.

Circular Dichroism (CD):

CD spectra were recorded in far-UV region (185-280 nm) using 20-25 µM protein concentrations in 10 mM sodium phosphate buffer, pH 6.9 at 25 °C. For urea denaturation experiments, CD spectra were recorded in far-UV region (210-280 nm) in buffer (50 mM sodium phosphate, 5 mM TCEP·HCl, pH 7.5) with increasing concentration of urea at *c* 10.5 µM of heterodimer **Het-N2** and homodimers formed from **Cys-N2** or 'positive' strands. The

Jasco J-810-1505 instrument was set as follows: Digital Integration Time (D.I.T.) 1 sec, band width 1 nm, data pitch 0.1 nm, scanning speed 100 nm/min. Every CD curve was obtained by averaging of 5 scans and subtracting the background signal of buffer. Melting curves in temperature range 5-90 °C were recorded at 222 nm using 50 µM protein concentrations in 10 mM sodium phosphate buffer with 500 µM TCEP, pH 6.9. Instrument parameters were set as follows: 1 °C/min, measurement every 0.2 °C, D.I.T. 2 sec, band width 1 nm.

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

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

where:
$[\theta]$: mean residue molar ellipticity [deg·cm^2·dmol^{-1}]
θ: ellipticity [mdeg]
d: pathlength of cuvette [mm] (here, d = 1 mm)
c: protein concentration [µM]
n: amount of amino acids in the protein sequence

Native mass-spectrometry:

Separated stock solutions of 'positive' and **Cys-N2** strands at 100 µM were prepared in buffer of 100 mM AcONH$_4$, pH 7.5. The two strands were mixed in a 1:1 ratio and diluted with buffer to have the heterodimer **Het-N2** at a final concentration of 8.5 µM. The sample was incubated for 5 minutes before analysis. Full MS spectra were acquired on a Thermo Scientific Exactive Plus EMR Orbitrap mass-spectrometer using direct infusion in positive ion mode at a 3.5 kV spray voltage and controlled with Exactive series (ver. 2.9 sp4) software. Resolution of full MS and HCD scans was 70,000. Normalized collision energy for HCD spectra was 35 eV, in source CID 100 eV and data were acquired in profile mode and processed using Xcalibur 4.2 sp1.

2. Chemical synthesis

2.1 General Procedure I for Boc/benzyl solid-phase peptide synthesis

*The procedure was used for the synthesis of 'positive' strand, **Cys-N2**, **Cys-Ncap** and **Ala-N2** peptides.*

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

Coupling (0.2 mmol scale): The Boc-amino acid (4 eq, 0.8 mmol) was dissolved in 2 mL of a solution of HBTU or HATU (*c* 0.38 M in DMF (3.8 eq)). For preactivation 350 µL of DIEA (10 eq) were added, then the reaction mixture was vortexed and let to react for 2 min. The resulting solution was poured onto the resin beads. The coupling reaction lasted for 10 to 15 min with stirring every 5 min. The beads were rinsed and batch washed with DMF four times. The completion of the coupling was monitored by Kaiser test.[5] To do so, a few beads were transferred to a test tube and rinsed twice with EtOH. Then 50 µL of a solution of ninhydrin (0.5 g in 10 mL of EtOH) and 50 µL of KCN solution (0.4 mL of 1 mM KCN$_{(aq)}$ in 20 mL of pyridine) were added. The mixture was placed into the heating block at 110 °C for 5 min. 500 µL of EtOH and 500 µL of H$_2$O were added to the tube and the color of the beads and the solution was inspected. A blue color of the beads and of the solution indicates an incomplete coupling. In the case of incomplete coupling a double coupling was performed.

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

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

Acetylation (0.2 mmol scale): For each monomer, the N-termini were acetylated with a mixture of 1.4 mL of DMF, 200 µL of DIEA and 200 µL of acetic anhydride. The reaction lasted for 6 min with stirring. The resin was washed with DMF and a Kaiser test was performed to validate the completeness of acetylation.

Cleavage: The resin was washed six times with DCM and dried by air flow using a vacuum filtration setup for 20 min. In the cleavage reaction for 100 mg of resin, 100 µL of *p*-cresol were added to the beads. Then, around 10 mL of HF were condensed to the reactor using dry ice + ethanol bath. The resulting mixture was stirred for 1 hour at 0 °C.

Work up: After evaporation of the HF under reduced pressure, crude products were precipitated with chilled diethyl ether (-20 °C). Product precipitates were filtered on Celite® 545 and dissolved with a 1 : 1 mixture MeCN/H_2O, 0.1 % TFA. An aliquot was taken for HPLC and LC-MS analysis. The peptide solution was frozen in liquid nitrogen and then lyophilized. The peptide was purified on a semi-preparative HPLC.

'Positive' strand: from a 0.4 mmol scale synthesis; isolated with 3 % yield. LC-MS (ESI): [M + 5 H$^+$] = 1095.0 m/z; [M + 6 H$^+$] = 912.7 m/z; [M + 7 H$^+$] = 782.5 m/z; [M + 8 H$^+$] = 684.9 m/z; [M + 9 H$^+$] = 608.9 m/z; [M + 10 H$^+$] = 548.3 m/z; $C_{235}H_{409}N_{83}O_{67}$; Average isotope calculated: 5469.4 Da [M]; observed: 5470.8 Da.

Cys-N2: from a 0.3 mmol scale synthesis; isolated with 2 % yield. LC-MS (ESI): [M + 4 H$^+$] = 1335.5 m/z; [M + 5H$^+$] = 1068.3 m/z; [M + 6H$^+$] = 890.4 m/z; $C_{229}H_{386}N_{70}O_{74}S$; Average isotope calculated: 5336.1 Da [M]; observed: 5337.0 Da.

Cys-Ncap: from a 0.4 mmol scale synthesis; isolated with 2 % yield. LC-MS (ESI): [M + 3 H$^+$] = 1784.4 m/z; [M + 4 H$^+$] = 1338.4 m/z; [M + 5 H$^+$] = 1070.8 m/z; [M + 6 H$^+$] = 892.6 m/z; $C_{230}H_{388}N_{70}O_{74}S$; Average isotope calculated: 5350.1 Da [M]; observed: 5349.3 Da.

Ala-N2: from a 0.1 mmol scale synthesis; isolated with 2 % yield. LC-MS (ESI): [M + 4 H$^+$] = 1326.8 m/z; [M + 5 H$^+$] = 1061.6 m/z; [M + 6 H$^+$] = 885.0 m/z; $C_{229}H_{386}N_{70}O_{74}$; Average isotope calculated: 5304.0 Da [M]; observed: 5303.7 Da.

Synthesis of Cys-S-sulfonate derivative of the Cys-N2 strand:[6,7] The **Cys-N2** strand (5.3 mg, 9.93×10^{-7} mol) was dissolved in 350 µL of solution containing 200 mM Na_2SO_3 and 20 mM $Na_2S_4O_4$ in 20 mM Tris·HCl, 1 mM EDTA, 8 M urea buffer, pH 7.3 to obtain a final concentration of peptide of 2.8 mM. The pH of the resulting reaction mixture was 7.1. The reaction mixture was stirred at room temperature for 4 h and the completion of the reaction was verified by HPLC and LC-MS. The Cys-S-sulfonate containing peptide was isolated by semi-preparative HPLC.

Cys-N2-SO$_3$H: isolated with 33 % yield (1.8 mg). LC-MS (ESI): [M + 3 H$^+$] = 1806.3 m/z; [M + 4 H$^+$] = 1355.0 m/z; [M + 5 H$^+$] = 1084.2 m/z; $C_{229}H_{386}N_{70}O_{77}S_2$; Average isotope calculated: 5416.2 Da [M]; observed: 5415.8 Da.

2.2. General Procedure II for Fmoc solid-phase peptide synthesis

*The procedure was used for the synthesis of **trunc-Cys-N2**.*

Automated microwave Fmoc/tBu solid-phase peptide synthesis:[8] automated microwave peptide synthesis was performed on a Liberty Blue (CEM) synthesizer. The synthesis was performed on a 0.1 mmol scale with side-chain protection of the amino acids as follows: Arg(Pbf), Cys(Trt), Gln(Trt), Glu(OtBu), Ser(tBu). The resin (Rink amide, 0.74 mmol/g loading, 100-200 mesh) was weighted to enable a synthesis on a 0.1 mmol scale. The coupling reactions were performed by adding amino acids dissolved in DMF (2.5 mL, 0.2 M), the coupling reagent DIC in DMF (1.0 mL, 0.5 M) and Oxyma Pure in DMF (0.5 mL,

1 M) to the respective resin. Standard couplings were performed at 90 °C for 2 min (170 W for 15 s, 30 W for 110 s). Standard deprotections were performed using 20% (v/v) 4-methylpiperidine in DMF at 90 °C for 1 min (155 W 15 s, 32 W 50 s).

The N-terminal acetylation, the cleavage and the work up were performed manually (for acetylation see the General Procedure I for Boc/benzyl solid-phase peptide synthesis). After final acetylation, the resin was washed six times with DCM and dried by air flow using a vacuum filtration setup for 20 min. In the cleavage reaction for 100 mg of resin, 1 mL of cleavage cocktail (95 % TFA, 2.5 % TIPS, 2.5 % H_2O) was added to the beads. The resulting mixture was stirred for 2 hours at room temperature. The cleavage mixture was transferred into several 50 mL centrifuge tubes and chilled diethyl ether (-20 °C) was added to reach a total volume of 40 mL. Afterwards, crude peptide product precipitated and was centrifuged at 4500 rpm for 3 min. The supernatant was discarded and the peptide was dissolved in 15 mL of a 1 : 1 mixture MeCN/H_2O, 0.1 % TFA. The solution was filtered to remove the resin beads. An aliquot was taken for HPLC and LC-MS analyses. The peptide solution was frozen in liquid nitrogen and then lyophilized. The peptide was purified on a semi-preparative HPLC.

Trunc-Cys-N2: from a 0.1 mmol scale synthesis; isolated yield 32 %. LC-MS (ESI): [M + 1H$^+$] = 1315.0 m/z; [M + 2H$^+$] = 657.9 m/z; $C_{56}H_{99}N_{17}O_{17}S$; Average isotope calculated: 1314.6 Da [M]; found 1313.9 Da.

2.3 General procedure III for the preparation of thioester substrates

The procedure was used for the synthesis of peptide-$^\alpha$thioesters.

Preparation of the 2-chlorotrityl hydrazide (2-CT-NHNH$_2$) resin:[9] The 2-CT-OH resin (1 g, 1.46 mmol/g active groups, mesh size 200-400) was washed several times with dry DCM and placed in a round bottom flask that was previously flushed several times with argon. Dry dichloromethane (10 mL) was added. The mixture was gently stirred allowing the resin to swell. Thionyl chloride (1.3 eq, 2.02 mmol, 147 μL) was slowly added at 0 °C. The suspension was stirred under argon for 2 h allowing it to slowly warm up to room temperature. The solvent was removed and the resin was washed with DCM and then with DMF. The resulting 2-CT-Cl resin was swollen for 20 min in DMF (6 mL). A mixture of 64 wt. % hydrazine in water (3 eq, 4.7 mmol, 233 μL) and DIEA (2.4 eq, 3.8 mmol, 667 μL) in DMF (2 mL) was added to the resin at 0 °C. The suspension was stirred at room temperature for 90 min. The reaction was quenched by the addition of methanol (100 μL). The resin was washed with DMF, water, DMF, methanol and diethyl ether and dried under reduced pressure. Final mass of the resin: 1.04 g with a loading assumed to be 0.4 mmol/g based on the previously synthesized peptides in the laboratory.

Fmoc/tBu solid-phase peptide synthesis: Ac-GRLEEIDR-$^\alpha$NHNH$_2$ was synthesized using standard manual SPPS procedure.[10] **Ac-LYRAG-$^\alpha$NHNH$_2$** was produced using automated microwave peptide synthesis (see General Procedure II). **Ac-VALENF-$^\alpha$NHNH$_2$** and **Ac-VALENR-$^\alpha$NHNH$_2$** were synthesized by machine-assisted peptide synthesis on a Syro I (Biotage) synthesizer. Typically, the syntheses were performed on a 0.1 mmol scale with side-chain protection of the amino acids as follows: Asn(Trt), Arg(Pbf), Asp(OtBu), Glu(OtBu), Tyr(tBu). The resin (2-CT-NHNH$_2$ resin, 0.4 mmol/g loading, 200-400 mesh) was weighted to enable a synthesis on a 0.1 mmol scale. Machine-assisted peptide syntheses on a Syro I synthesizer were performed as follows: the resin was washed twice with DMF and then was swollen for 20 min, while vortexing every 3 min. Coupling reaction was performed with 5 eq of Fmoc-amino acid (*c* 0.5 M in DMF), 5 eq of HATU (*c* 0.5 M in DMF) and 10 eq of DIEA (*c* 2.5 M in NMP) for 30 min, while vortexing every 3 min. Fmoc-deprotection was performed in 2 stages: *i)* 40% 4-methylpiperidine in DMF for 3 min, and *ii)* 20% 4-methylpiperidine in DMF for 12 min, while vortexing every 3 min.

For all peptides, N-terminal acetylation was then performed manually following General Procedure I. Then, resin drying, cleavage and work-up were carried out following the General Procedure II.

Due to acceptable purity, the crude peptide-$^\alpha$hydrazides were directly used in the subsequent thioesterifications without any purification step.

Activation of peptide-$^\alpha$hydrazides and thioesterification:[9] The peptide hydrazides were dissolved in acidic aqueous buffer (200 mM Na$_2$HPO$_4$, 6 M Gn·HCl, pH 3) at a 2.7 mM concentration. The temperature of the solution was adjusted to -15 °C, followed by a slow addition of NaNO$_2$ (5 eq, dissolved in water at c 0.2 M). The solution was stirred at -15 °C for 15 min. The formation of peptide-$^\alpha$azide could be followed by HPLC and LC-MS.

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

Ac-GRLEEIDR-$^\alpha$thioester: isolated with 58 % yield for thioesterification step based on the amount of crude hydrazide. LC-MS (ESI): [M + 1H$^+$] = 1153.6 m/z; C$_{44}$H$_{76}$N$_{14}$O$_{18}$S$_2$; Average isotope calculated: 1153.3 Da [M]; found: 1152.6 Da.

Ac-VALENF-$^\alpha$thioester: isolated with 21 % yield for thioesterification step based on the amount of crude hydrazide. LC-MS (ESI): [M + 1H$^+$] = 858.2 m/z; C$_{36}$H$_{55}$N$_7$O$_{13}$S$_2$; Average isotope calculated: 858.0 Da [M]; found: 857.2 Da.

Ac-LYRAG-$^\alpha$thioester: isolated with 31 % yield for thioesterification step based on the amount of crude hydrazide. LC-MS (ESI): [M + 1H$^+$] = 745.1 m/z; C$_{30}$H$_{48}$N$_8$O$_{10}$S$_2$; Average isotope calculated: 744.9 Da [M]; found: 744.1 Da.

Ac-VALENR-$^\alpha$thioester: isolated with 37 % yield for thioesterification step based the on amount of crude hydrazide. LC-MS (ESI): [M + 1H$^+$] = 867.3 m/z; C$_{33}$H$_{58}$N$_{10}$O$_{13}$S$_2$; Average isotope calculated: 867.0 Da [M]; found: 866.3 Da.

3.   Crystallization, data collection and structure refinement

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

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

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

For the **Het-N2-SO$_3$$^-$** heterodimer, 0.2 µL of solution at 12 mg/mL (1:1 stoichiometric ratio of 'positive' and **Cys-N2-SO$_3$H** strands) in MES buffer (100 mM MES, 100 mM NaCl, 10 mM TCEP, pH 6.5) was mixed with 0.2 µL precipitant solution and equilibrated against 50 µL reservoir solution. Crystals appeared after 2 weeks in 10% propan-2-ol, 0.2 M lithium sulfate, 0.1 M sodium phosphate citrate pH 4.2.

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

For the **Het-N2** heterodimer, data were collected at 100 K on an EIGER X 4M detector (Dectris) on the ID30A-3 beamline of the ESRF. 360° of data were collected using 0.1° rotation and 0.01 sec exposure per image. The data were indexed, integrated, and scaled using XDS,[11] before anisotropic truncation and correction were performed using the STARANISO server.[12] The data were processed to 2.33 Å (ellipsoidal cutoff 2.28 Å, 2.49 Å, 3.48 Å) and belonged to the primitive triclinic space group P 1, with unit cell dimensions a = 40.7 Å, b = 49.9 Å, c = 53.04 Å, $α$ = 105.09°, $β$ = 92.49°, $γ$ = 110.15°. The structure was

solved by molecular replacement using PHASER[13] in the PHENIX[14] suite using the known DSD structure as a search model (PDB ID: 1G6U).[15] The asymmetric unit contains four copies of the heterodimeric DSD analogue, with a corresponding Matthews' coefficient[16] of 2.23 $\text{Å}^3 \text{Da}^{-1}$ and a solvent content of 44.9 %.

For the **Het-Ncap** heterodimer, data were collected at 100 K on a PILATUS 6M detector (Dectris) on the ID23-1 beamline of the ESRF. 180° of data were collected using 0.15° rotation and 0.037 sec exposure per image. The data were indexed, integrated, and scaled using XDS before merging the data in AIMLESS.[17] The data were processed to 1.45 Å and belonged to the C-centered monoclinic space group $C\,2$, with unit cell dimension a = 76.9 Å, b = 73.1 Å, c = 84.1 Å, β = 105.02°. The structure was solved as above. The asymmetric unit contains four copies of the heterodimeric DSD analogue, with a corresponding Matthews' coefficient of 2.63 $\text{Å}^3 \text{Da}^{-1}$ and a solvent content of 53.3 %.

For the **Het-N2-SO$_3^-$** heterodimer, data were collected at 100 K on an EIGER X 16M detector (Dectris) on the Proxima 1 beamline of the Soleil Synchrotron. 360° of data were collected using 0.1° rotation and 0.02 sec exposure per image. The data were indexed, integrated, and scaled using XDS, before anisotropic truncation and correction were performed using the STARANISO server. The data were processed to 1.5 Å (ellipsoidal cutoff 1.41 Å, 1.56 Å, 2.49 Å) and belonged to the primitive monoclinic space group $P\,2_1$, with unit cell dimensions a = 26.5 Å, b = 68.7 Å, c = 47.3 Å, β = 98.23°. The structure was solved by molecular replacement using PHASER in the PHENIX suite using the **Het-N2** structure as a search model. The asymmetric unit contains four copies of the heterodimeric DSD analogue, with a corresponding Matthews' coefficient 1.95 $\text{Å}^3 \text{Da}^{-1}$ and a solvent content of 37.1 %.

The structures were refined using PHENIX and BUSTER[18] with iterative model building performed in COOT.[19] The quality of the final refined models was assessed using MOLPROBITY[20] and PROCHECK.[21] Data collection and refinement statistics are given in Table S3. For definitions of the statistics given in Table S3 see Table 1 of ref.[22] The coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes 6Z0L (**Het-N2**), 6Z0M (**Het-Ncap**) and 7BEY (**Het-N2-SO$_3^-$**).

4.   Catalytic assays

4.1   Acyl transfer reactions with peptide-$^\alpha$thioesters

Standard conditions for screening catalytic activity: 100 µM of catalytic protein, 200 µM of acyl donor (peptide-$^\alpha$thioester substrate), 200 mM of acyl acceptor in 50 mM sodium phosphate buffer, pH 7.5 containing 100 mM NaCl, 2 mM of TCEP and 4% v/v CH_3CN.

Demonstration of catalytic turnover: 100 µM of **Het-N2**, from 200 µM to 2 mM of acyl donor (Ac-VALENF-$^\alpha$thioester), 200 mM of acyl acceptor (Tris) in 50 mM sodium phosphate buffer, pH 7.5, 100 mM NaCl and 2 mM of TCEP (Figure 6C). Furthermore, we performed an experiment with sequential addition of acyl donor in which 100 µM of **Het-N2** were first mixed with 200 mM of acyl acceptor (Tris) and then 200 µM of acyl donor (Ac-VALENF-$^\alpha$thioester) were added two times (at time 0 h and at time 47 h) (Figure S11).

Preparation of stock solutions: For a given heterodimer, the two monomers were first dissolved separately in phosphate buffer (50 mM Na_2HPO_4, 100 mM NaCl, pH 7.0) at a stock concentration of 750 µM. 14 mM of TCEP were added to the solution of Cys-containing monomer. pH of solutions was adjusted to 7.1. In parallel, a stock solution of acyl acceptor was prepared at concentration of 290 mM with 2.9 mM TCEP and pH was adjusted to 7.5-7.6. The stock solutions of peptide-$^\alpha$thioester substrates at 5 mM in anhydrous CH_3CN were also prepared.

Preparation of reaction mixture: To obtain final expected concentrations in 50 µL of reaction, 6.7 µL of each stock solution of 'positive' and 'negative' strands were first mixed together at the bottom of 1.5 mL Eppendorf tubes. Then, 35 µL of stock solution of acyl acceptor with TCEP were added followed by 2 µL of peptide-$^\alpha$thioester solution

(corresponding to time = 0 h). pH was controlled and adjusted if necessary to 7.5 and the reaction mixture was transferred to glass HPLC vial or PCR microtube (0.2 mL). Resulting mixture was let to react for 1 week at room temperature (20-22 °C). The data displayed in Figure 6C and Figures S10-S11 were obtained at temperature of 25 °C precisely controlled by PCR Thermal Cycler C1000 from BioRad.

The pH in all cases was measured with an Inlab Ultra-Micro-ISM microelectrode (3 mm diameter) from Mettler TOLEDO (cat. N° 30244732) using pH meter SevenCompact from Mettler TOLEDO.

Monitoring: Aliquots were generally taken off after 1 h, 3 h, 7 h, 24 h, 48 h, 72 h, 96 h and 1 week of reaction. 2 µL of reaction mixture were diluted with 25 µL of water, 0.1 % TFA. 10 µL were then injected for LC-MS analysis. The relative percentages of peptide-$^\alpha$thioester (starting material), peptide-$^\alpha$acyl acceptor and peptide-$^\alpha$OH (products) were extracted from areas of the respective detected ions following the equation:

$$Relative\ percentage_{compound}$$
$$= \frac{Area_{compound}}{(Area_{peptide-thioester}\ +\ Area_{peptide-acyl\ acceptor}\ +\ Area_{peptide-OH})} \times 100$$

The relative percentage of branched adduct was calculated relative to peak areas of detected ions of free **Cys-N2** strand and branched adduct strand following the equation:

$$Relative\ percentage_{branched\ adduct} = \frac{Area_{branched\ adduct}}{(Area_{branched\ adduct}\ +\ Area_{free\ \textbf{Cys-N2}\ strand})} \times 100$$

Observed rate constants were deduced from half-life of peptide-$^\alpha$thioester (corresponding to 50 % of thioester in reaction mixture) using the pseudo-first order rate law equation given that the concentration of acyl-acceptor in the reaction mixture is several magnitudes higher than peptide-$^\alpha$thioester substrate concentration:

$\tau_{1/2} = \frac{ln2}{k_{obs}}$, where: $\tau_{1/2}$ - half-life of thioester (when reaching 50 %), k_{obs} - observed pseudo-first order rate constant.

The observed rate constants were found to be reproducible with standard deviation in 6-16% range based on duplicate or triplicate measurements. Furthermore, experiments were performed with different batches of synthesized **Het-N2** and the values were determined to be within the range for standard deviation.

4.2 Hydrolytic activity with *para*-nitrophenyl acetate (*p*NPA) as substrate

Kinetic assays were performed in a 10 mm path-length quartz cuvette using a V-670 Jasco spectrophotometer at 348 nm (ε_{pNP} = 5400 cm^{-1}·M^{-1}) at 22 °C. Stock solutions of each strand of **Het-N2** were prepared separately at 105 µM in phosphate buffer (50 mM sodium phosphate, 100 mM NaCl, 500 µM TCEP, pH 7.0). 'Positive' and **Cys-N2** strands were mixed in a 1:1 ratio to obtain a final concentration of 50 µM in 315 µL of total volume (300 µL of peptide solution and 15 µL of *p*NPA solution in anhydrous CH$_3$CN). 15 µL of a freshly prepared solution of *p*NPA (to obtain final concentration between 400 and 1250 µM) were added to the **Het-N2** solution. Measurements were continued by recording absorption values every 30 seconds for a total time of 3-4 hours. Data were background corrected (300 µL of buffer with TCEP + 15 µL of CH$_3$CN). Two controls were performed in presence of 50 µM of **trunc-N2-Cys** or without any peptide catalyst both at 600 µM of *p*NPA.
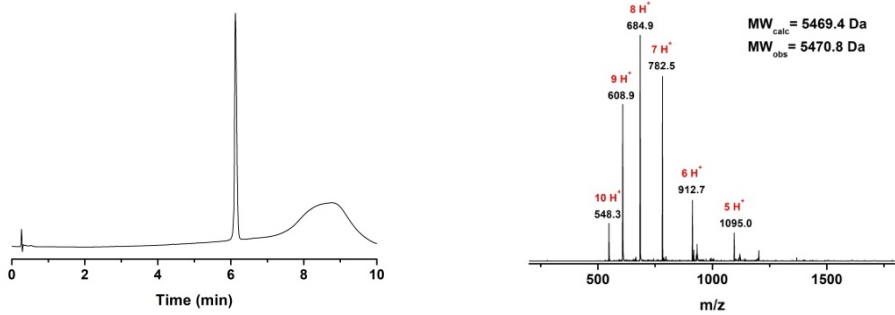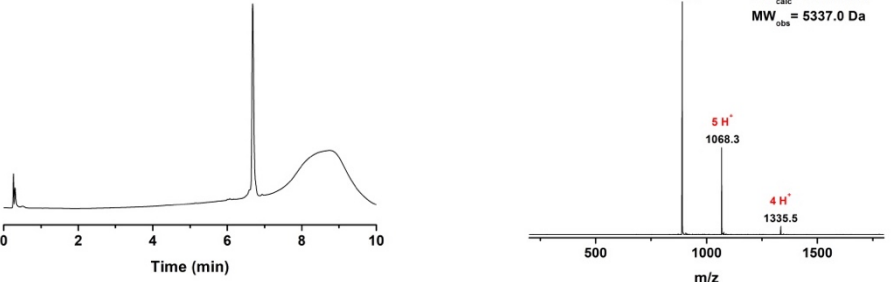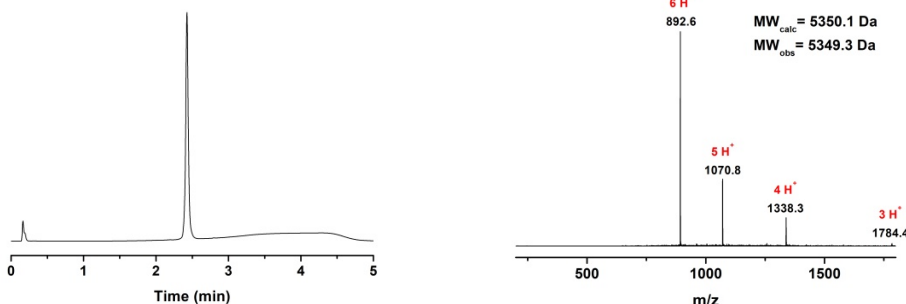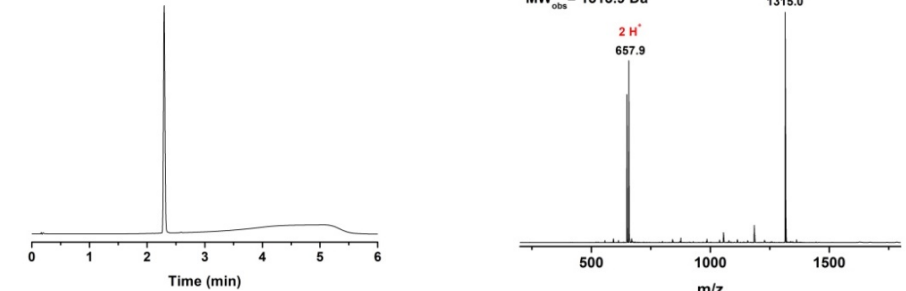
All kinetic curves were fitted with the following equation containing a linear part and an exponential (also called "burst" phase): $[pNP] = At + B(1 - e^{-bt})$.[23] By calculating A at different $[S]_0$, the double-reciprocal plot was then deduced and used to determine k_{cat} ($k_{cat}[E]_0$= 1/y-intercept) and K_m (K_m=slope/y-intercept).
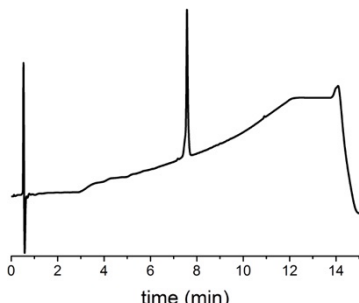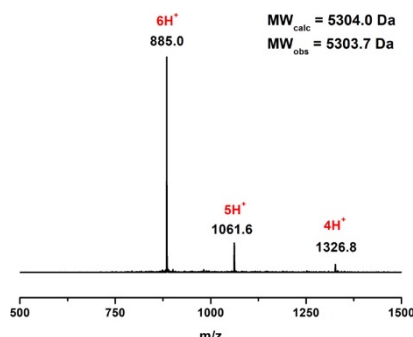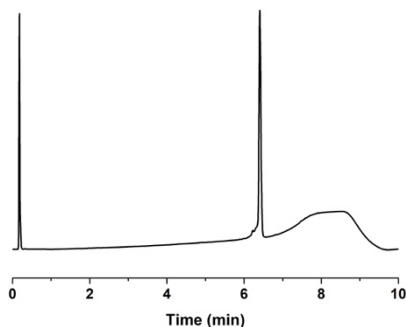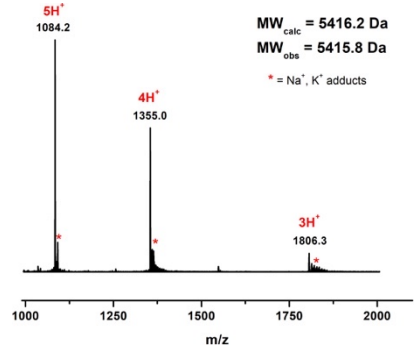
5. Supplementary Tables and Figures

Table S1. PDB entries used in structural alignment shown in Figure 1.

Ncap cluster1	Ncap cluster2	N1	N2
2FEP	6FJ4	1U7V	3TMP
4M69	3QPD	1MKI	2R5L
4Q94	2Y1K	6GR8	3WDZ
5I6E	3GA7	2J90	6F7B
4KK4	4CKT	4BPG	5K9P
6GSG	3ZI7	4Q9A	2QG5
5OAT		2GLQ	3EQU
4C0S		3IAF	4X8N
4BH6		2W5X	5ZVJ
1KKM		3ZTB	5WUM
3EXH		5N5M	5YPO
4WRQ		6M93	1UHG
3D9L		6M91	4L7E
		3QIC	3UAT
		4YOO	4M69
		6R4E	3JZM
		3W8K	
		3PPZ	
		6B67	
		1H4X	
		2XZ0	
		6N14	

Table S2. Analytical characterization of the synthesized catalytic constructs and controls

Construct	Peptide
Het-N2	'Positive' strand: Ac-NLAALRSELQALRREGFSPERLAALESRLQALERRLAALRSRLQALRG-amide Analytical characterization: MW$_{calc}$ = 5469.4 Da MW$_{obs}$ = 5470.8 Da **Cys-N2** strand: Ac-GLCALRSELQALRREGFSPEELAALESELQALERELAALRSELQALRG-amide Analytical characterization: MW$_{calc}$ = 5336.1 Da MW$_{obs}$ = 5337.0 Da
Het-Ncap	'Positive' strand: Ac-NLAALRSELQALRREGFSPERLAALESRLQALERRLAALRSRLQALRG-amide Analytical characterization: see **Het-N2** **Cys-Ncap** strand: Ac-CLAALRSELQALRREGFSPEELAALESELQALERELAALRSELQALRG-amide Analytical characterization: MW$_{calc}$ = 5350.1 Da MW$_{obs}$ = 5349.3 Da
trunc-Cys-N2	Ac-GLCALRSELQAL-amide Analytical characterization: MW$_{calc}$ = 1314.6 Da MW$_{obs}$ = 1313.9 Da

Neg-N2	'Positive' strand: Ac-NLAALRSELQALRREGFSPERLAALESRLQALERRLAALRSRLQALRG-amide Analytical characterization: see **Het-N2**
	Ala-N2 strand: Ac-GLCALRSELQALRREGFSPEELAALESELQALERELAALRSELQALRG-amide Analytical characterization:
Het-N2-SO₃⁻	'Positive' strand: Ac-NLAALRSELQALRREGFSPERLAALESRLQALERRLAALRSRLQALRG-amide Analytical characterization: see **Het-N2**
	Cys-N2-SO₃H strand: Ac-GLC(SO₃H)ALRSELQALRREGFSPEELAALESELQALERELAALRSELQALRG-amide Analytical characterization:

For Ala-N2 strand mass spectrum: peaks at 6H⁺ 885.0, 5H⁺ 1061.6, 4H⁺ 1326.8; MW$_{calc}$ = 5304.0 Da, MW$_{obs}$ = 5303.7 Da

For Cys-N2-SO₃H strand mass spectrum: peaks at 5H⁺ 1084.2, 4H⁺ 1355.0, 3H⁺ 1806.3; MW$_{calc}$ = 5416.2 Da, MW$_{obs}$ = 5415.8 Da; * = Na⁺, K⁺ adducts
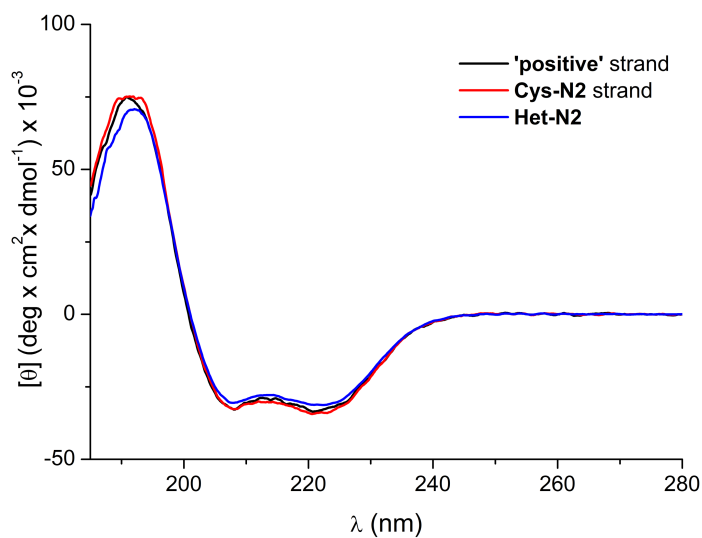
Figure S1. CD spectra for the catalytic protein **Het-N2** and the respective isolated 'positive' and **Cys-N2** strands.
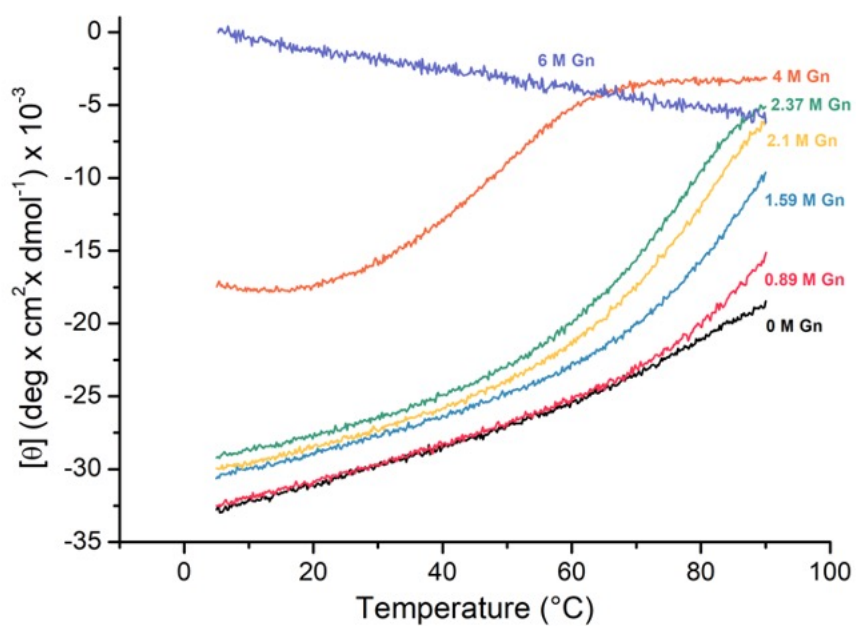
Figure S2. Melting curves measured by CD at 222 nm for the catalytic protein **Het-N2** at different concentrations of guanidine hydrochloride (Gn·HCl). Thermal denaturation of **Het-N2** is similar to the original DSD scaffold with dimerization constant K_D = 140 pM.[15]
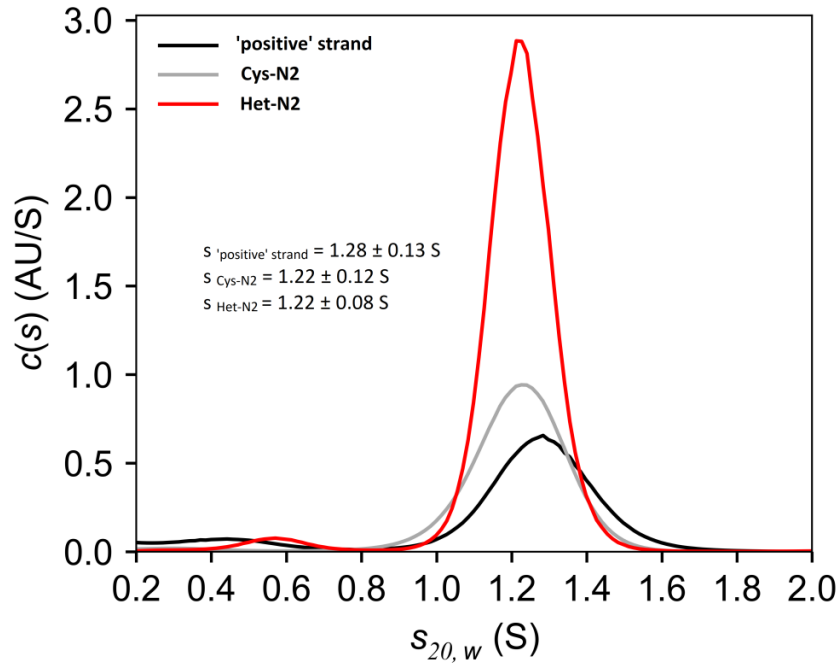
Figure S3. Sedimentation profiles of **Het-N2** heterodimer protein and isolated 'positive' and **Cys-N2** strands (each at *c* 200 µM in phosphate buffer). Due to dimerization of 'positive' strand and **Cys-N2** peptide, the concentrations of respective homodimers correspond to 100 µM, which is 50% of **Het-N2** heterodimer concentration. This explains weaker amplitudes for 'positive' and **Cys-N2** samples.
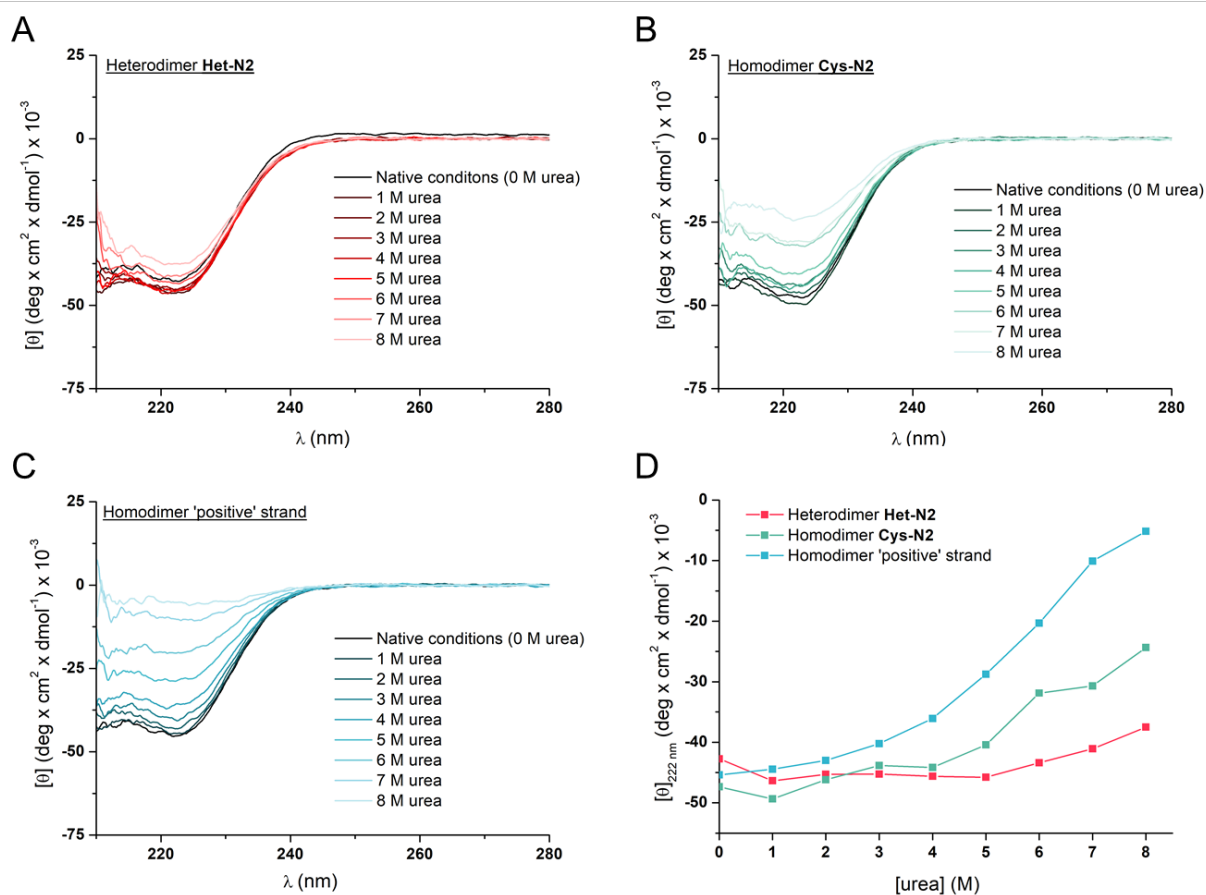
Figure S4. Urea denaturation experiments. (A-C) Circular dichroism spectra showing decrease of α-helical structure upon increase of the concentration of urea in the buffer (50 mM sodium phosphate, 5 mM TCEP·HCl, pH 7.5) at *c* 10.5 µM of heterodimer **Het-N2** and homodimers formed from **Cys-N2** or 'positive' strands. (D) A comparison of denaturation curves measured at 222 nm for heterodimer **Het-N2** and two homodimers. Making a few assumptions one can estimate the K_D at 8 M where all the peptides show some denaturation. We assume that the monomer is disordered at 8 M urea (0 ellipticity) and the structures are completely folded at 0 to 1 M urea corresponding to $[\theta]_{222} \sim -45 \times 10^3$ deg·cm^2·dmol^{-1}. Thus, the concentrations of folded structures at 8 M urea are: 8.8 µM for **Het-N2,** 5.7 µM for **Cys-N2** dimer and 1.2 µM for dimer of 'positive' strand, which correspond to K_D values at 8 M urea: 0.33 µM, 4.0 µM and 72 µM, respectively. In Gibbs free energy terms (at T = 25 °C), it will correspond to ~ 1.5 kcal/mol and ~ 3.2 kcal/mol destabilization of **Cys-N2** and 'positive' strand homodimers, respectively, in comparison to heterodimer **Het-N2**. Thus, the corresponding homodimers are approximately 10- and 100-times less stable than heterodimer, respectively.
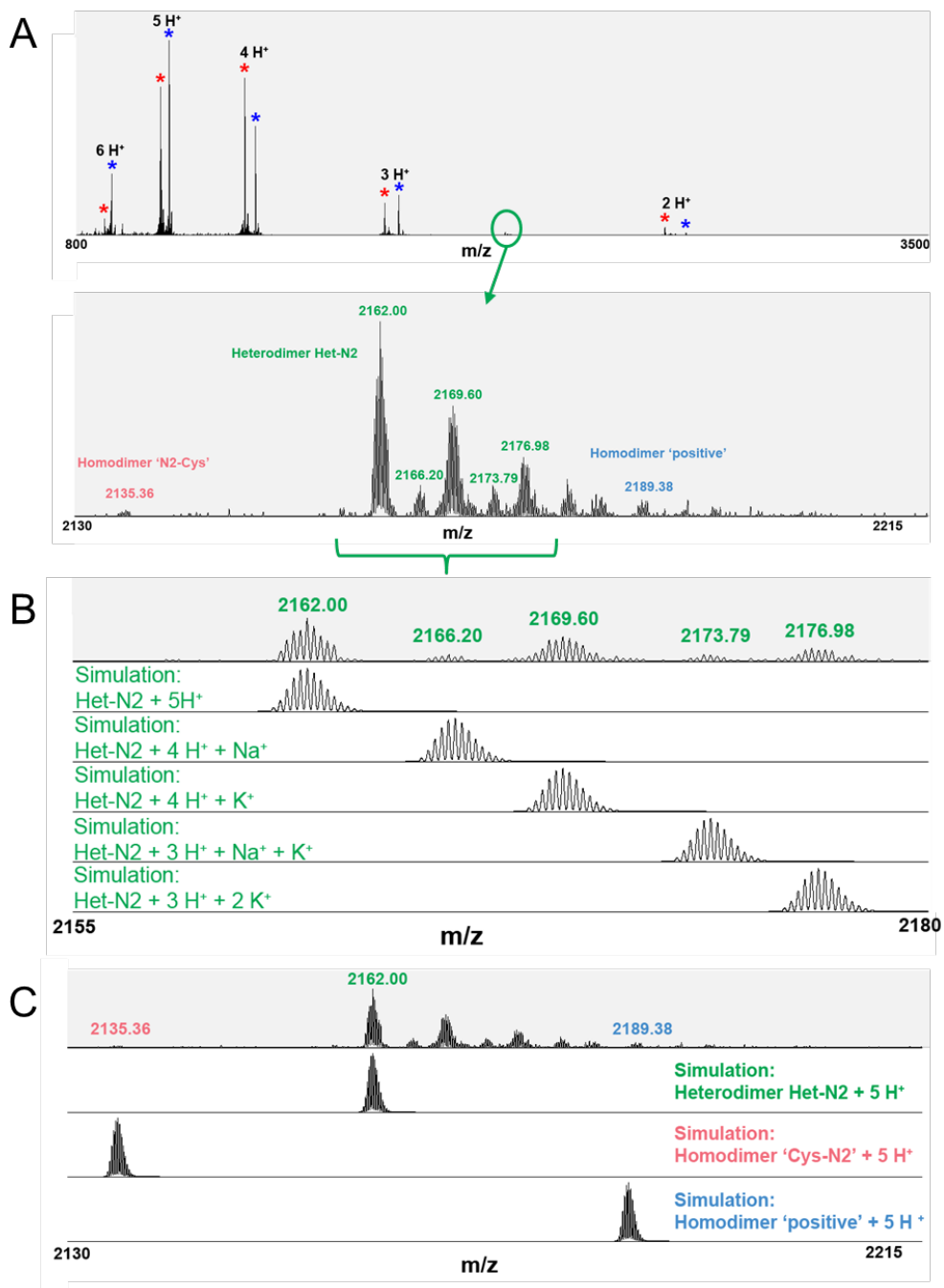
Figure S5. Native mass-spectrometry of **Het-N2** indicates mainly heterodimer and minute amounts of **Cys-N2** and 'positive' strand homodimers detected in the mixture. (A) In the top panel, ions for **Cys-N2** strand are labeled with red stars, for 'positive' strand with blue stars and for dimeric species in green. Bottom panel is a zoom onto the area with dimeric species. (B) Experimental data (top panel) versus mass-spectra simulations (below) for various ion adducts of **Het-N2**. Formula for **Het-N2** is $C_{464}H_{795}N_{153}O_{141}S$. (C) Experimental data (top panel) versus simulations (below) for homodimer species. Formula for homodimer of **Cys-N2** strand is $C_{458}H_{772}N_{140}O_{148}S_2$ and formula for homodimer of 'positive' strand is $C_{470}H_{818}N_{166}O_{134}$.

Het-N2
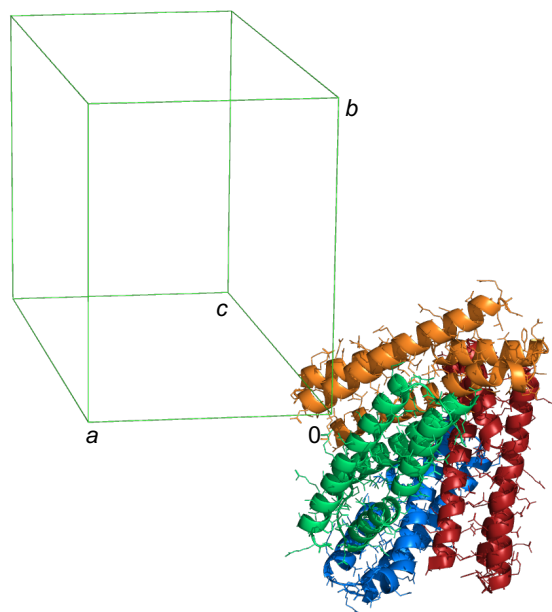
Het-Ncap
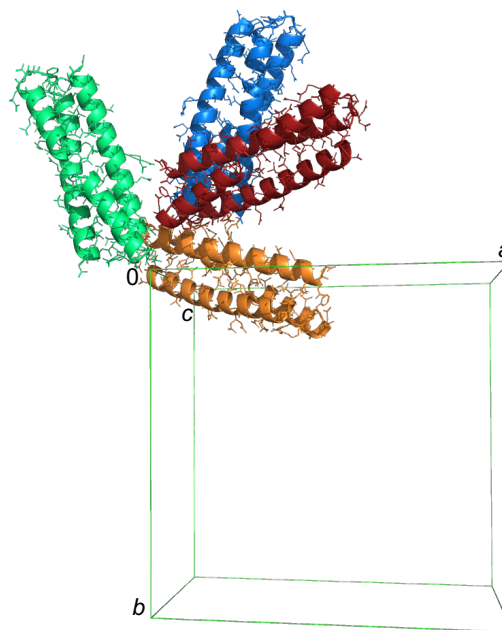
Figure S6. Four copies of heterodimer were observed in each asymmetric unit of the corresponding X-ray structures. PDB IDs are 6Z0L (**Het-N2**) and 6Z0M (**Het-Ncap**).
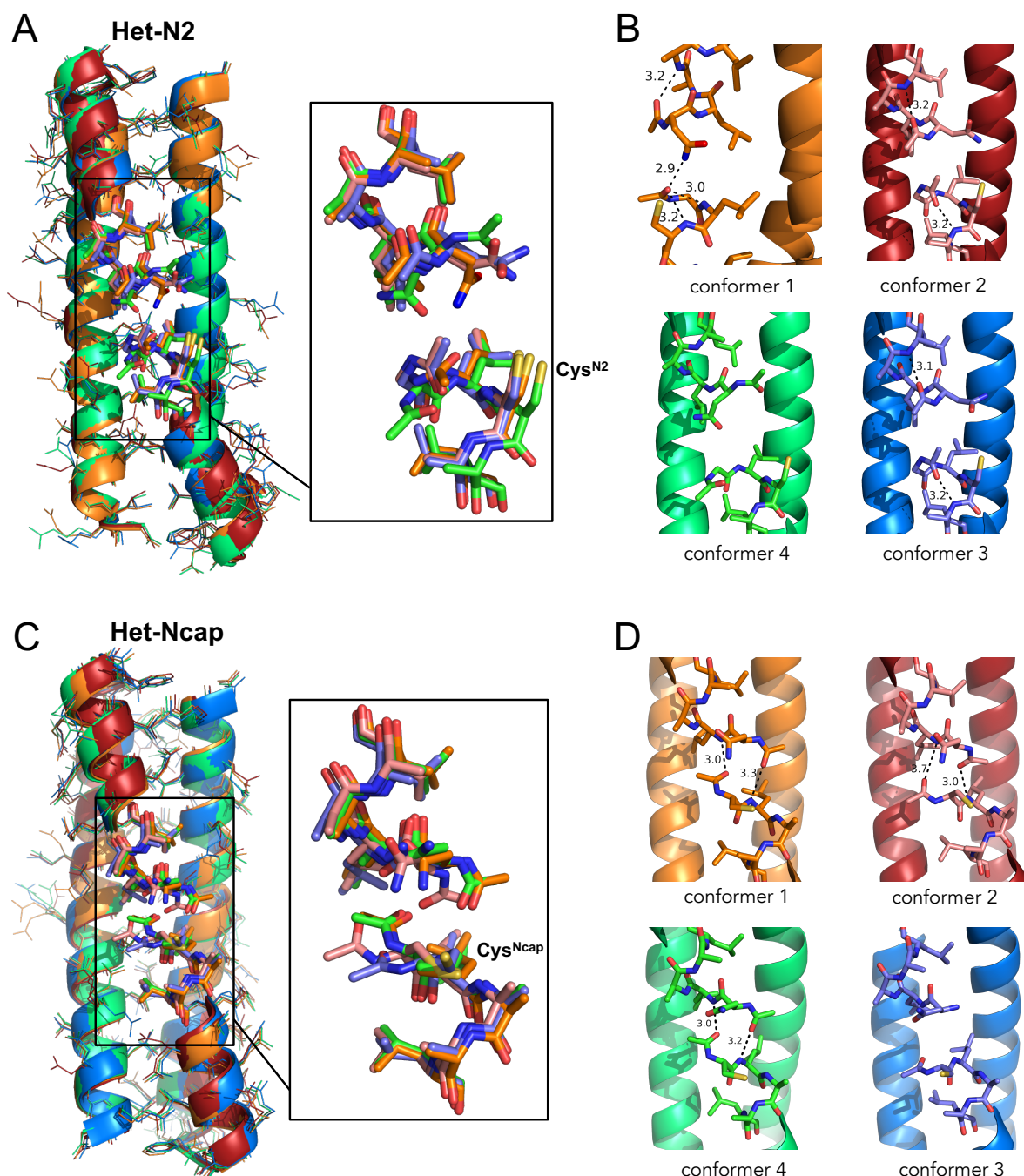
Figure S7. Overlay of the four symmetry independent heterodimers for each X-ray structure and conformational variability in **Het-N2** and **Het-Ncap**. (A) For **Het-N2**, all-atom RMSD between the four copies varies between 1.16 Å and 1.51 Å. (B) Acetyl group of **Cys-N2** strand participates in H-bonding interactions with backbone amides of N-terminus and side-chain of Asn(Ncap) residue (conformer 1) or only backbone amides of cognate helices (conformations 2 and 3), or no H-bonding interactions are formed (conformer 4). (C) For **Het-Ncap**, all-atom RMSD between the four copies is between 1.19 Å and 1.57 Å. (D) Two acetyl groups of **Cys-Ncap** strand and 'positive' strand serve as H-bond acceptors with backbone amides of opposing helix N-termini (conformations 1, 2 and 4) or no H-bonding interactions are formed (conformer 3).
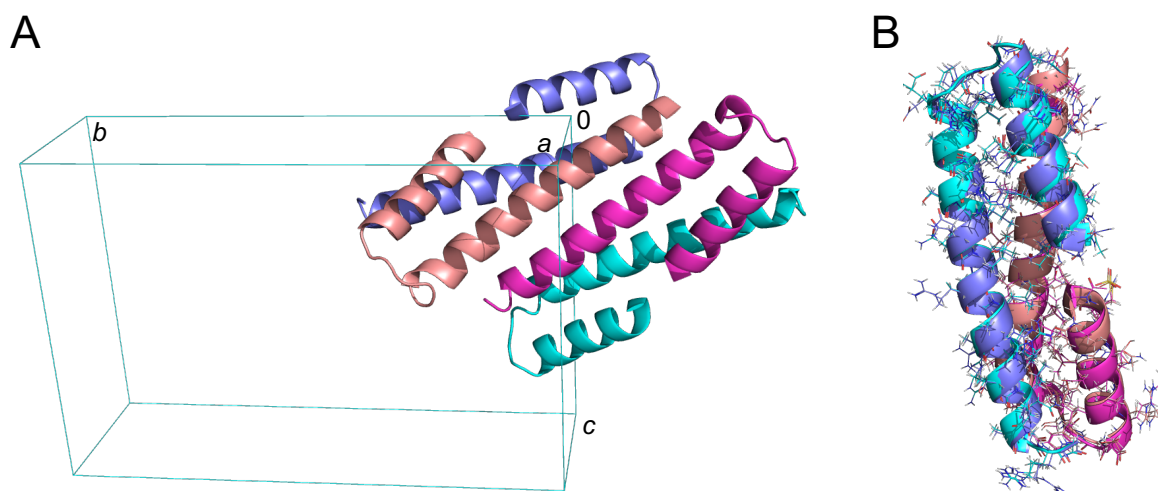
Figure S8. (A) The asymmetric unit in the **Het-N2-SO₃⁻** crystal structure that contains two heterodimers (space group $P2_1$, a = 26.455 Å, b = 68.673 Å, c = 47.258 Å, β = 98.23°). PDB ID is 7BEY. (B) All atom RMSD is 1.65 Å for the superposition of two symmetry non-equivalent heterodimers.
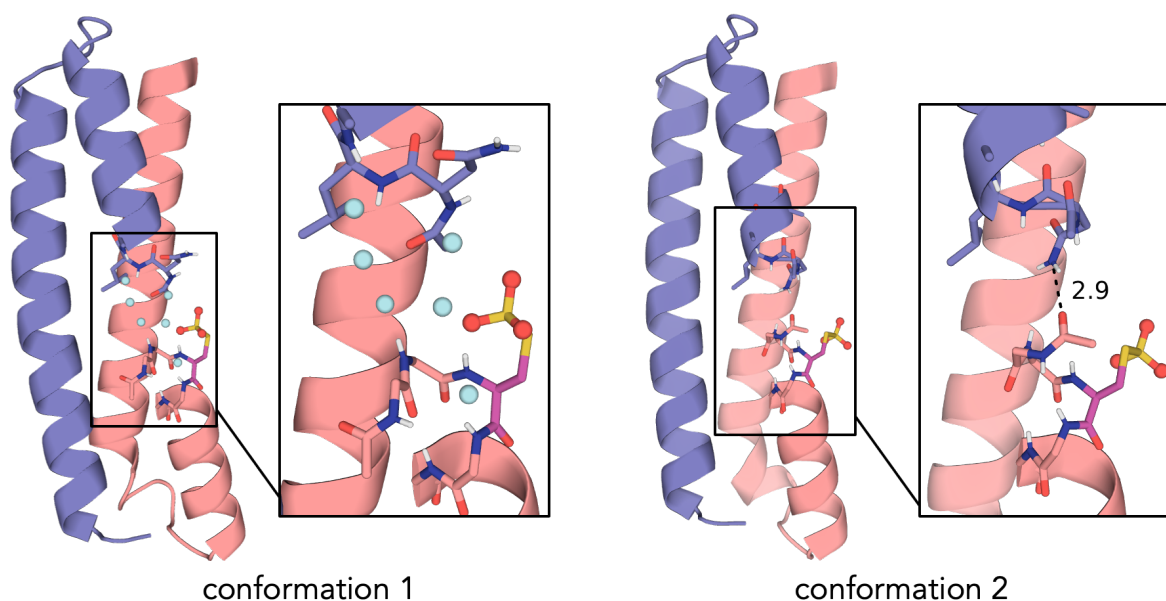

conformation 1   conformation 2

Figure S9. Comparison of two conformations observed in the X-ray structure of **Het-N2** Cys-S-sulfonate adduct. In the first conformation, sulfonate makes H-bonding interactions with N-terminus of α-helix, whereas in the second conformation, sulfonate moiety is oriented towards the solvent. Interaction between the two helical N-termini is mediated via a network of water molecules (shown as cyan spheres) in the first conformation or by H-bond between acetyl group and side chain amide of Asn(Ncap) from 'positive' strand (O-N distance is shown in Å).

Table S3. Data collection and refinement statistics for **Het-N2**, **Het-Ncap** and **Het-N2-SO₃⁻**. The values given in parenthesis correspond to the highest resolution shells. For anisotropic data the statistics for the anisotropic truncation performed by the STARANISO server are given.

	Het-N2	Het-Ncap	Het-N2-SO$_3^-$
PDB ID	6Z0L	6Z0M	7BEY
Data Collection			
Source	ESRF ID30A-3	ESRF ID23-1	Soleil Proxima 1
Detector	EIGER X 4M	PILATUS 6M	EIGER-X 16M
Wavelength (Å)	0.968	0.97	0.979
Resolution range (Å)	21.76 - 2.33 (2.417 - 2.33)	52.09 - 1.45 (1.502 - 1.45)	34.34 - 1.50 (1.55 - 1.50)
Space group	*P* 1	*C* 2	*P* 2$_1$
Cell Dimensions: *a, b, c (Å)*	40.67, 49.95, 53.04	76.88, 73.10, 84.13	26.46, 68.67, 47.26
α, β, γ (°)	105.09, 92.49, 110.15	90, 105.021, 90	90, 98.229, 90
Total reflections	47758 (4029)	265700 (25795)	184980 (18400)
Unique reflections	14490 (139)	78982 (7859)	26629 (303)
Multiplicity	3.3 (2.9)	3.4 (3.3)	6.9 (7.0)
Completeness (%)	54.82 (8.56)	99.15 (99.12)	53.42 (11.33)
Mean I/sigma(I)	2.98 (0.26)	14.17 (1.49)	5.28 (0.26)
Wilson B-factor	14.56	20.91	15.35
R-merge	0.257 (3.230)	0.036 (0.697)	0.197 (4.522)
R-meas	0.307 (3.905)	0.043 (0.833)	0.213 (4.882)
R-pim	0.165 (2.166)	0.023 (0.452)	0.080 (1.822)
CC1/2	0.983 (0.151)	0.999 (0.811)	0.998 (0.138)
CC*	0.996 (0.512)	1 (0.946)	1 (0.493)
Anisotropic truncation			
Ellipsoidal resolution (Å) (direction)	3.48 (0.791 a* - 0.079 b* + 0.607 c*)		2.487 (0.928 a* - 0.373 c*)
	2.28 (0.069 a* + 0.730 b* - 0.680 c*)		1.412 (b*)
	2.49 (-0.516 a* + 0.615 b* + 0.597 c*)		1.562 (0.047 a* + 0.999 c*)
Best / worst diffraction limits	2.33 (2.53 – 2.33) / 4.43		1.50 (1.62 - 1.50) / 2.94
Total reflections (ellipsoidal)	29603 (1390)		99700 (5436)
Unique reflections (ellipsoidal)	8746 (435)		14323 (719)
Mean I/sigma(I) (ellipsoidal)	4.1 (1.8)		8.8 (1.3)
R-merge (ellipsoidal)	0.246 (0.822)		0.124 (1.529)
R-meas (ellipsoidal)	0.292 (0.983)		0.134 (1.631)
R-pim (ellipsoidal)	0.157 (0.534)		0.050 (0.592)
CC1/2 (ellipsoidal)	0.963 (0.503)		0.999 (0.500)
Ellipsoidal completeness (%)	84.2 (67.0)		88.0 (65.1)
Refinement			
Reflections used in refinement	8740 (139)	78982 (7850)	14321 (304)
Reflections used for R-free	486 (6)	4028 (381)	713 (15)
R-work	0.223 (0.277)	0.152 (0.244)	0.222 (0.322)
R-free	0.283 (0.196)	0.207 (0.290)	0.271 (0.303)
CC(work)	0.772 (0.494)	0.968 (0.922)	0.885 (0.733)
CC(free)	0.823 (0.810)	0.955 (0.889)	0.890 (0.765)
Number of non-hydrogen	3121	4040	2000
macromolecules	2998	3380	1726
ligands	38	38	26
solvent	85	622	248
Protein residues	400	400	200
RMS(bonds)	0.002	0.002	0.011
RMS(angles)	0.38	0.51	0.97
Ramachandran favored (%)	99.73	97.82	99.44
Ramachandran allowed (%)	0.27	2.18	0.56
Ramachandran outliers (%)	0	0	0
Rotamer outliers (%)	3.39	1.93	2.29
Clashscore	7.91	5.9	3.55
Average B-factor	34.19	33.26	17.92
macromolecules	34.34	30.83	16.98
ligands	47.72	55.21	30.76
solvent	23.05	45.15	23.11

Table S4. Analytical characterization of the synthesized peptide-$^{\alpha}$thioester substrates
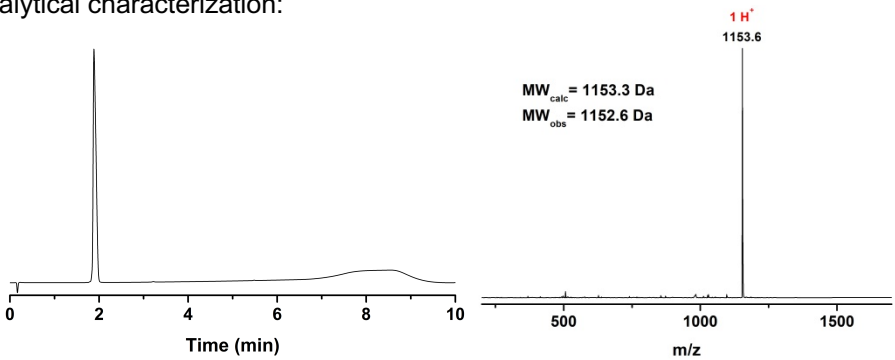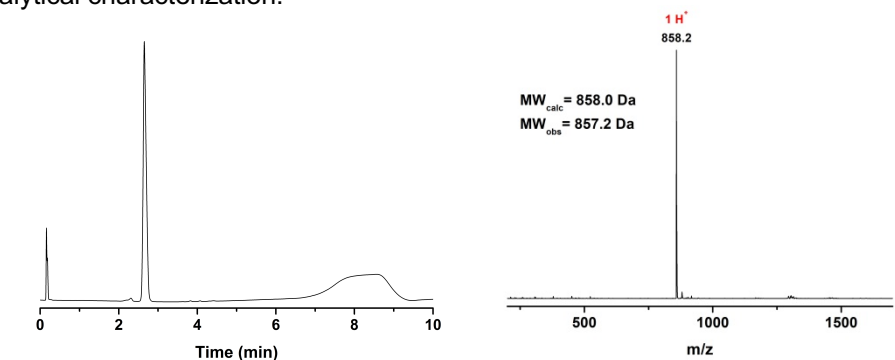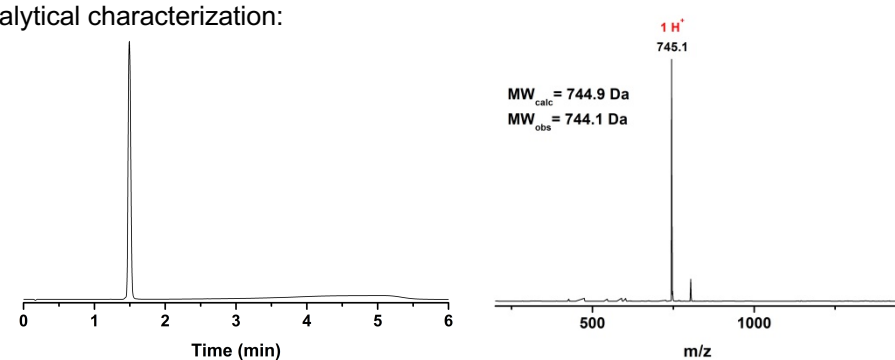
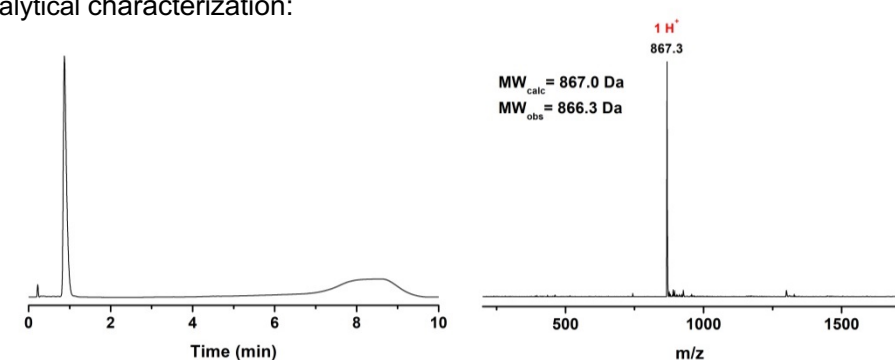| Substrate	Peptide sequence and analytical characterization
Ac-GRLEEIDR-$^{\alpha}$thioester	Sequence: Ac-GRLEEIDR-$^{\alpha}$Mes Analytical characterization: MW$_{calc}$ = 1153.3 Da MW$_{obs}$ = 1152.6 Da 1 H$^+$ 1153.6
Ac-VALENF-$^{\alpha}$thioester	Sequence: Ac-VALENF-$^{\alpha}$Mes Analytical characterization: MW$_{calc}$ = 858.0 Da MW$_{obs}$ = 857.2 Da 1 H$^+$ 858.2
Ac-LYRAG-$^{\alpha}$thioester	Sequence: Ac-LYRAG-$^{\alpha}$Mes Analytical characterization: MW$_{calc}$ = 744.9 Da MW$_{obs}$ = 744.1 Da 1 H$^+$ 745.1
Ac-VALENR-$^{\alpha}$thioester	Sequence: Ac-VALENR-$^{\alpha}$Mes Analytical characterization: MW$_{calc}$ = 867.0 Da MW$_{obs}$ = 866.3 Da 1 H$^+$ 867.3

Table S5. Comparison of catalytic activity between the DSD-analogues **Het-N2** and **Het-Ncap** and controls (**trunc-Cys-N2**, glutathione and MPAA). Conditions: 100 µM catalyst, 200 µM peptide-Phe-$^\alpha$Mes, 200 mM Tris, pH 7.5.

Catalyst	k_{obs} (s^{-1})	Amidation/Hydrolysis ratio	Acceleration compared to background reaction
Het-N2	1.1×10^{-5}	0.8	5.5
Het-Ncap	3.9×10^{-6}	1.0	2.0
trunc-Cys-N2	4.4×10^{-6}	1.4	2.2
Glutathione	5.4×10^{-6}	0.7	2.7
MPAA	3.2×10^{-6}	1.0	1.6
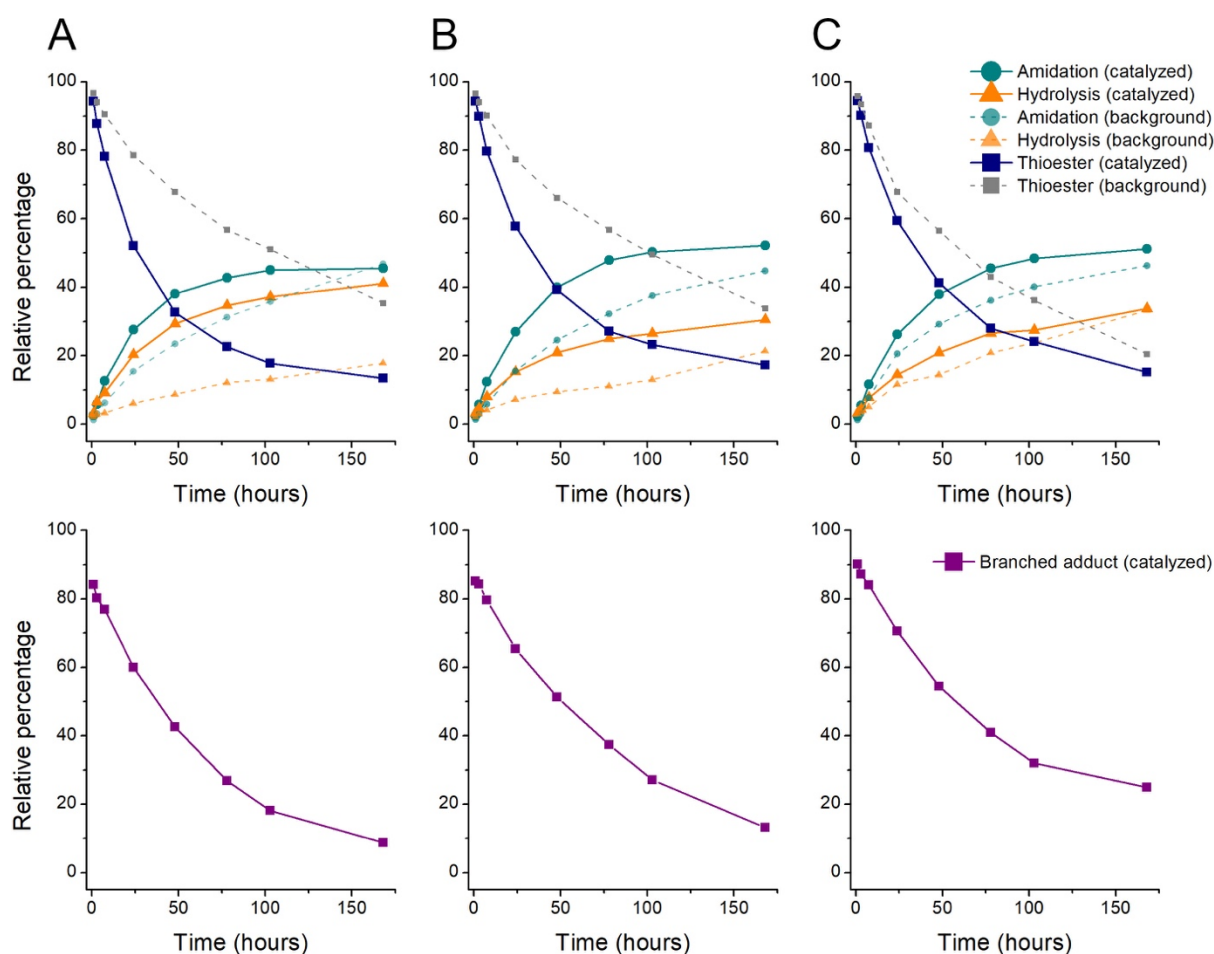
Figure S10. Kinetics of amidation/hydrolysis of Ac-VALENF-$^\alpha$thioester catalyzed by **Het-N2** (solid line) in comparison to reaction without catalyst (background reaction shown as dashed line). (A) 0.75 mM of Ac-VALENF-$^\alpha$thioester and 100 µM **Het-N2** with 200 mM of Tris at pH 7.5. (B) 1.25 mM of Ac-VALENF-$^\alpha$thioester and 100 µM **Het-N2** with 200 mM of Tris at pH 7.5. (C) 2 mM of Ac-VALENF-$^\alpha$thioester and 100 µM **Het-N2** with 200 mM of Tris at pH 7.5. Relative percentages of compounds were calculated according to equations provided in subsection 4.1 of the Supporting Information. In the lower panels, peptide-protein branched thioester adduct is in higher amounts upon increase of Ac-VALENF-$^\alpha$thioester concentration.
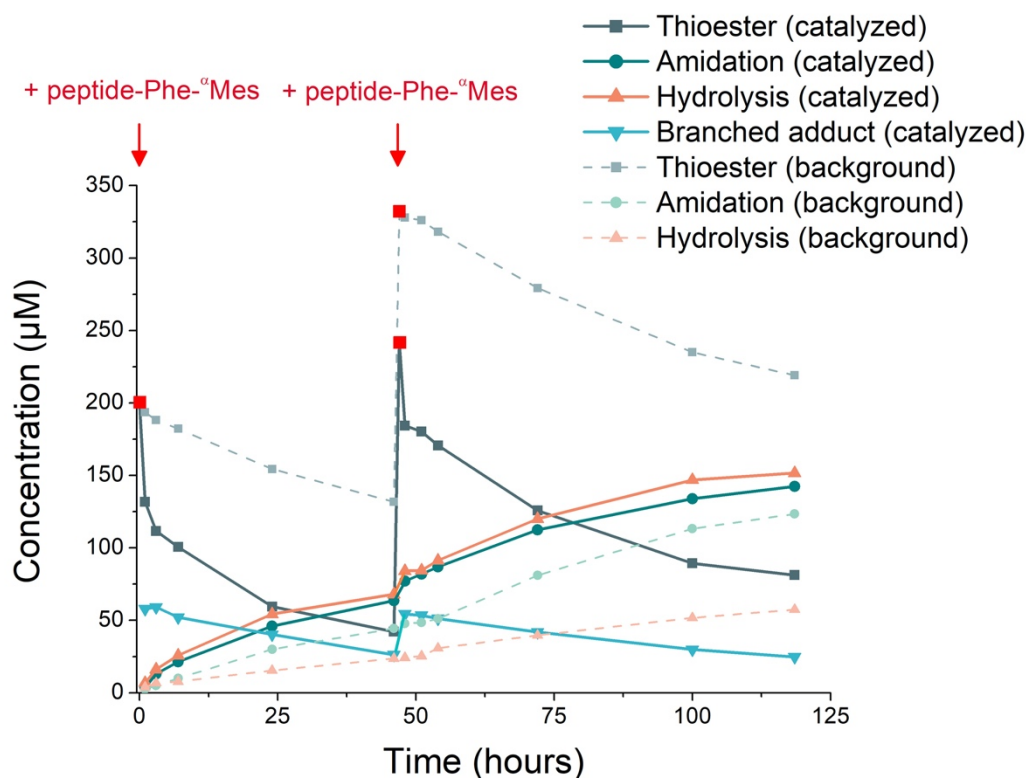
Figure S11. Kinetics of Ac-VALENF-$^{\alpha}$Mes amidation/hydrolysis with **Het-N2** protein catalyst. The same amount of thioester substrate corresponding to initial concentration of 200 μM was added twice at t = 0 h and t = 47 h (indicted by red squares). **Het-N2** concentration was 100 μM and Tris concentration was 200 mM at pH 7.5.
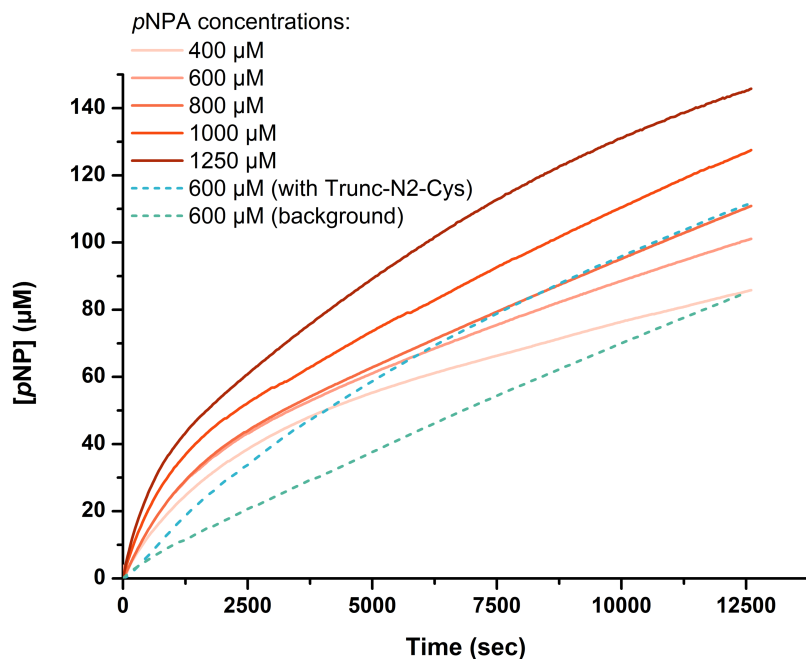
Figure S12. Pre-steady-state kinetic curves of **Het-N2** (50 μM) at different concentration of *p*NPA substrate (orange-brown curves). Burst phase and linear phase were well observed and fitted with the equation $[pNP] = At + B(1 - e^{-bt})$. To compare, background reaction without catalyst at 600 μM of *p*NPA (dashed green line) gave no typical burst phase and the **trunc-Cys-N2** peptide (dashed blue line) showed a slower burst phase.

Table S6. Kinetic parameters that were deduced from kinetic curves at each concentration of *p*NPA substrate by fitting the equation: $[pNP] = At + B(1 - e^{-bt})$.

[*p*NPA]	400 μM	600 μM	800 μM	1000 μM	1250 μM
A (μM sec^{-1})	0.0038	0.0053	0.0063	0.0073	0.0078
B (μM)	38.4	35.7	31.9	36.8	52.0
b (sec^{-1})	0.00058	0.00076	0.00087	0.0011	0.00075
k_2/K_s (M^{-1} sec^{-1})*	1.45	1.26	1.10	1.11	1.88

*k_2/K_s is related to the initial burst phase efficiency
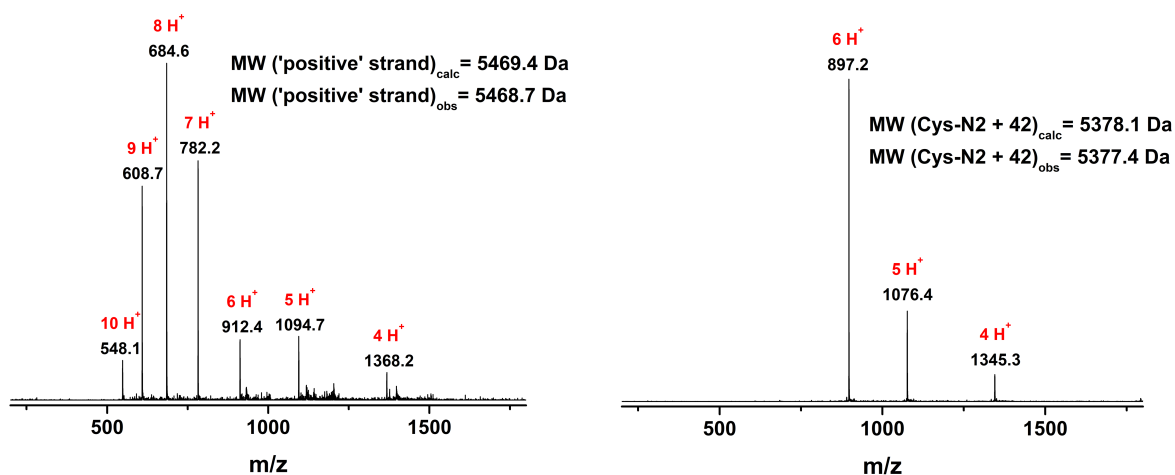
Figures S13. LC-MS spectra of the reaction mixture (50 µM **Het-N2**, 600 µM *p*NPA) after 3.5 hours of reaction. The **Cys-N2** strand was mostly present in the form of the acetyl adduct (Δm +42 Da).
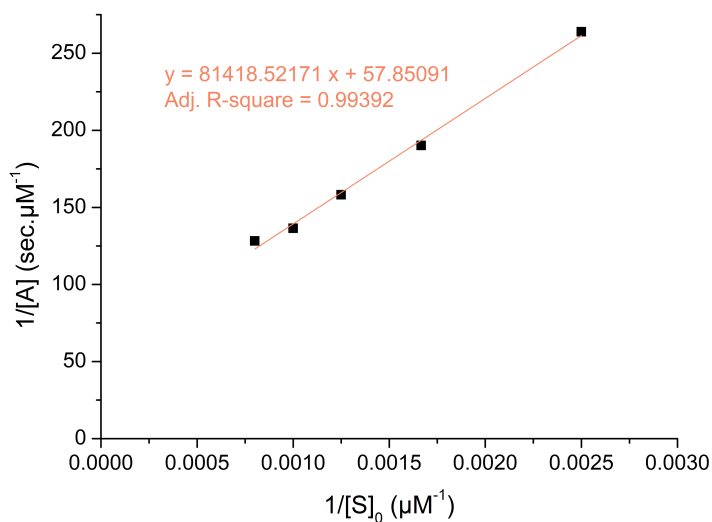

Figure S14. Double-reciprocal plot for the hydrolytic activity of **Het-N2** with *p*NPA as substrate. k_{cat} was extracted from $\dfrac{1}{y-intercept \times [Het-N2]_0}$ = 3.46 × 10^{-4} s^{-1} and K_m from $\dfrac{slope}{y-intercept}$ = 1.4 mM. Catalytic efficiency k_{cat}/K_m is 0.25 M^{-1} s^{-1}. The value k_{cat}/k_{uncat} = 138, where pseudo-first order rate constant k_{uncat} is 2.5 × 10^{-6} s^{-1} for uncatalyzed hydrolysis of *p*NPA at 22 °C, pH 6.95.[24]
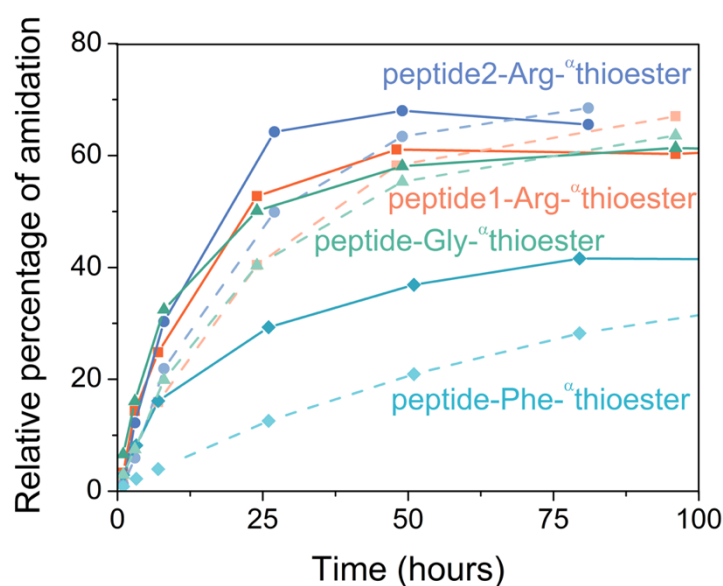
Figure S15. Comparison of amidation product formation by reaction of Tris on different thioester substrates in presence (solid lines) or in absence (dashed lines) of **Het-N2**. The sequences used for the different thioester substrates were the following: peptide1-Arg-$^{\alpha}$Mes = Ac-GRLEEIDR-$^{\alpha}$Mes; peptide2-Arg-$^{\alpha}$Mes = Ac-VALENR-$^{\alpha}$Mes; peptide-Phe-$^{\alpha}$Mes = Ac-VALENF-$^{\alpha}$Mes; peptide-Gly-$^{\alpha}$Mes = Ac-LYRAG-$^{\alpha}$Mes.
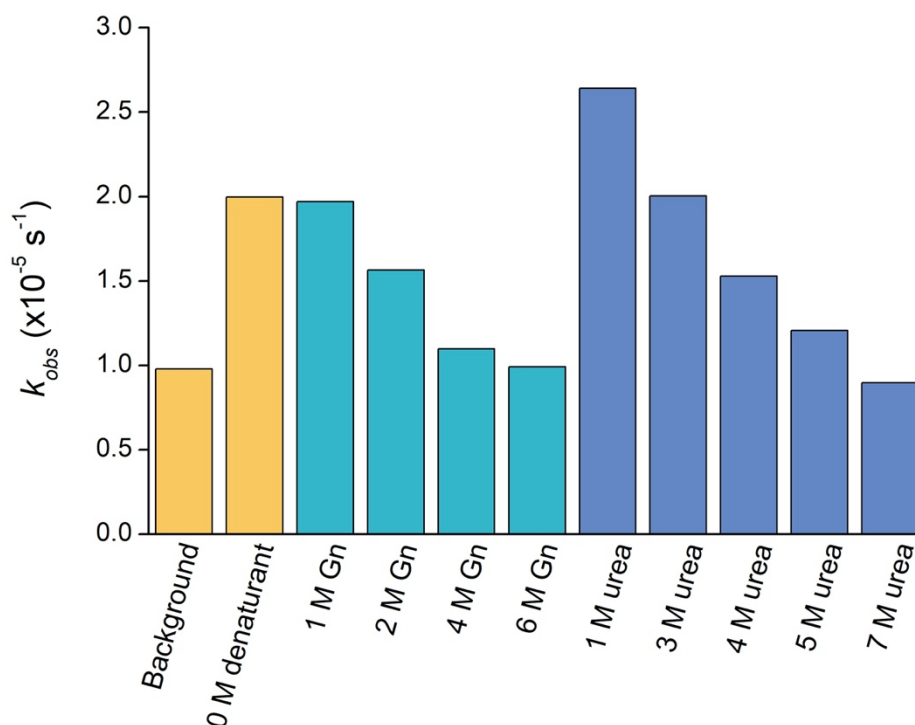

Figure S16. Observed rate constant k_{obs} depending on the concentration of denaturant (guanidine hydrochloride or urea) with 100 µM of **Het-N2**, 200 µM of Ac-GRLEEIDR-$^{\alpha}$thioester and 200 mM of Tris, pH 7.5.
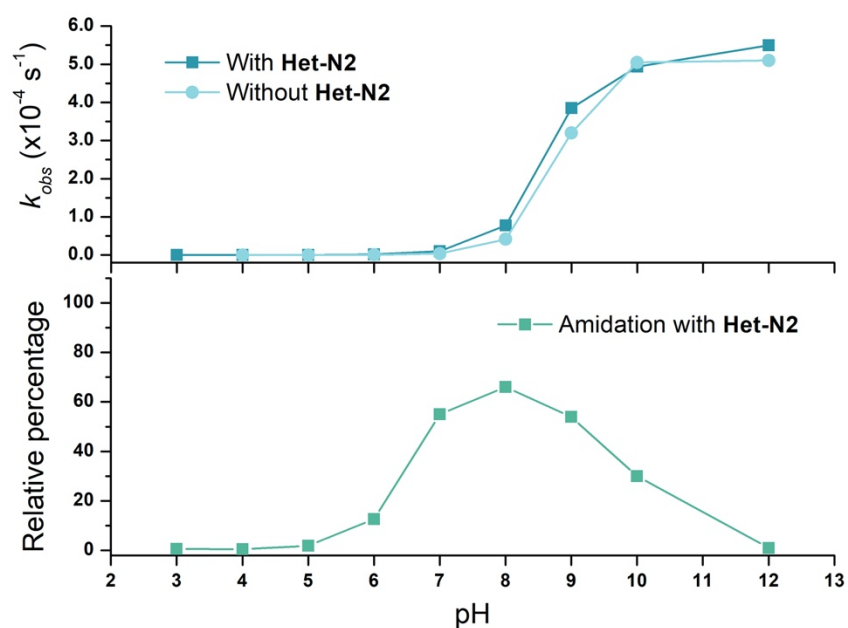
S 26

Figure S17. Observed rate constant k_{obs} and relative percentage of amidation at different pHs for the reaction between Ac-GRLEEIDR-$^\alpha$thioester (200 µM) and Tris (200 mM) in the presence or absence of **Het-N2** (100 µM).
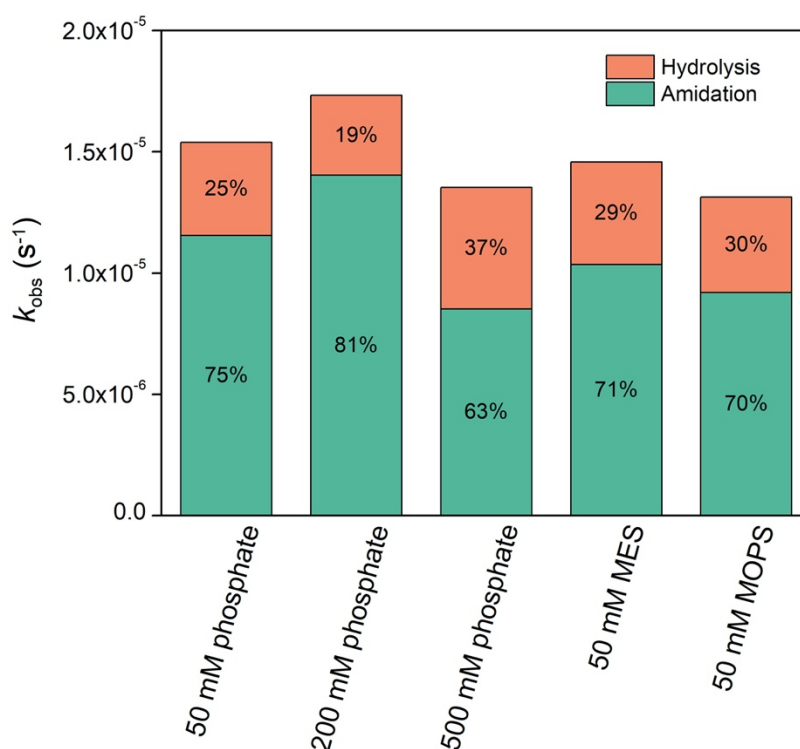

Figure S18. Catalytic activity of **Het-N2** for the reaction between Ac-GRLEEIDR-$^\alpha$Mes and Tris in different buffers (50 mM, 200 mM or 500 mM sodium phosphate; 50 mM MES; 50 mM MOPS). Catalytic activity of **Het-N2** remained similar, meaning that the buffer did not influence the kinetics.
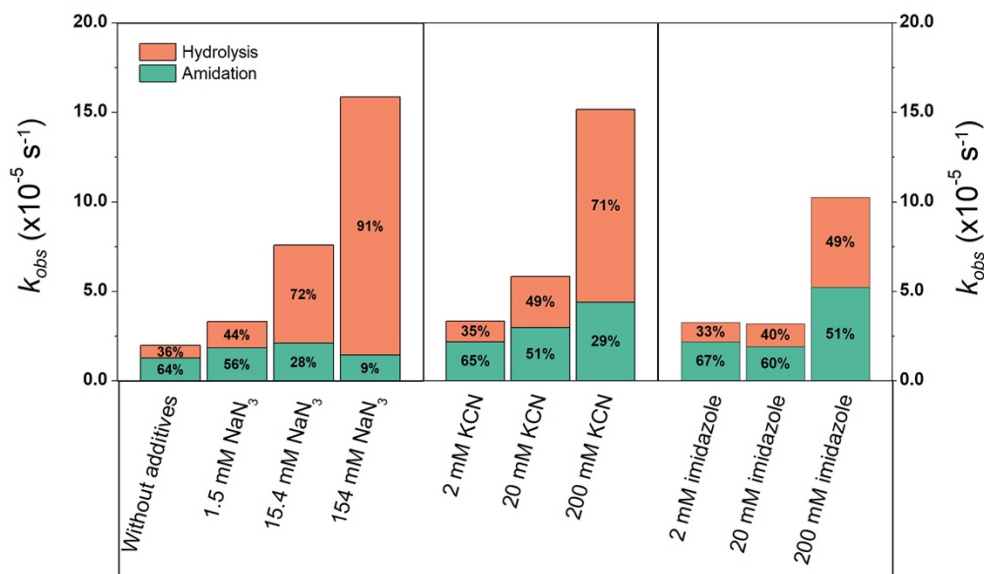
Figure S19. Variation of the kinetics (in the reaction with Ac-GRLEEIDR-$^{\alpha}$thioester (200 µM), Tris (200 mM) and **Het-N2** (100 µM)) by addition of NaN_3, KCN or imidazole in comparison to the results without additives.

Supplementary references

(1)   Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **2000**, *78*, 1606-1619.
(2)   Laue, T. M.; Shah, B. D.; Ridgeway, T. M.; Pelletier, S. L. Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*; Harding, S.; Rowe, A.; Horton, J., Eds.; Royal Society of Chemistry: Cambridge, 1992; pp 90-125.
(3)   Brautigam, C. A. Chapter Five - Calculations and publication-quality illustrations for analytical ultracentrifugation data. In *Methods in Enzymology*; Cole, J. L., Ed.; Analytical Ultracentrifugation; Academic Press, 2015; Vol. 562, pp 109-133.
(4)   Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. In situ neutralization in Boc-chemistry solid phase peptide synthesis. *Int. J. Pept. Res. Ther.* **2007**, *13*, 31-44.
(5)   Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I. Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal. Biochem.* **1970**, *34*, 595-598.
(6)   Weil-Ktorza, O.; Rege, N.; Lansky, S.; Shalev, D. E.; Shoham, G.; Weiss, M. A.; Metanis, N. Substitution of an internal disulfide bridge with a diselenide enhances both foldability and stability of human insulin. *Chem. Eur. J.* **2019**, *25*, 8513-8521.
(7)   Min, C.-K.; Son, Y.-J.; Kim, C.-K.; Park, S.-J.; Lee, J.-W. Increased expression, folding and enzyme reaction rate of recombinant human insulin by selecting appropriate leader peptide. *J. Biotechnol.* **2011**, *151*, 350-356.
(8)   Behrendt, R.; White, P.; Offer, J. Advances in Fmoc solid-phase peptide synthesis. *J. Pept. Sci.* **2016**, *22*, 4-27.
(9)   Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat. Protoc.* **2013**, *8*, 2483-2495.
(10)   Coin, I.; Beyermann, M.; Bienert, M. Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nat. Protoc.* **2007**, *2*, 3247-3256.
(11)   Kabsch, W. XDS. *Acta Cryst.* **2010**, *D66*, 125-132.

(12) Tickle, I. J.; Flensburg, C.; Keller, P.; Paciorek, W.; Sharff, A.; Vonrhein, C.; Bricogne, G. STARANISO. http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi (accessed March 5, 2020). Cambridge, United Kingdom: Global Phasing Ltd, **2018**.

(13) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read; R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658-674.

(14) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkóczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L.-W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.;P. D. Adams, P. D. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr., Sect. D: Biol. Crystallogr..* **2019**, *D75*, 861-877.

(15) Ogihara, N. L.; Ghirlanda, G.; Bryson, J. W.; Gingery, M.; DeGrado, W. F.; Eisenberg, D. Design of three-dimensional domain-swapped dimers and fibrous oligomers. *Proc. Natl. Acad. Sci.* **2001**, *98*, 1404-1409.

(16) Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **1968**, *33*, 491-497.

(17) Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *D69*, 1204-1214.

(18) Bricogne, G.; Blanc, E.; Brandl, M.; Flensburg, C.; Keller, P.; Paciorek, W.; Roversi, P.; Sharff, A.; Smart, O. S.; Vonrhein, C.; Womack, T.O. BUSTER Version 2.10.2. Cambridge, United Kingdom: Global Phasing Ltd, **2017**.

(19) Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *D66*, 486–501.

(20) Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.; Deis, L. N.; Verma, V., Keedy, D. A.; Hintze, B. J.; Chen, V. B.; Jain, S.; Lewis, S. M.; Arendall, W. B., III; Snoeyink, J.; Adams, P. D.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **2018**, *27*, 293-315.

(21) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK - a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283-291.

(22) Karplus, P. A.; Diederichs, K. Assessing and maximizing data quality in macromolecular crystallography. *Curr. Opin. Struct. Biol.* **2015**, *34*, 60-68.

(23) Bender, M. L.; Kezdy, F. J.; Wedler, F. C. Alpha-chymotrypsin: enzyme concentration and kinetics. *J. Chem. Educ.* **1967**, *44*, 84-88.

(24) Bolon, D. N.; Mayo, S. L. Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14274-14279.