Covalent Protein Modification

DOI: 10.1002/anie.201307736

# N-Terminal Protein Modification by Substrate-Activated Reverse Proteolysis**

*Sandra Liebscher, Michael Schöpfel, Tobias Aumüller, Ariunkhur Sharkhuukhen, Andreas Pech, Eva Höss, Christoph Parthier, Günther Jahreis, Milton T. Stubbs,* and Frank Bordusa**

***Abstract:*** *Although site-specific incorporation of artificial functionalities into proteins is an important tool in both basic and applied research, it can be a major challenge to protein chemists. Enzymatic protein modification is an attractive goal due to the inherent regio- and stereoselectivity of enzymes, yet their specificity remains a problem. As a result of the intrinsic reversibility of enzymatic reactions, proteinases can in principle catalyze ligation reactions. While this makes them attractive tools for site-specific protein bioconjugation, competing hydrolysis reactions limits their general use. Here we describe the design and application of a highly specific trypsin variant for the selective modification of N-terminal residues of diverse proteins with various reagents. The modification proceeds quantitatively under native (aqueous) conditions. We show that the variant has a disordered zymogen-like activation domain, effectively suppressing the hydrolysis reaction, which is converted to an active conformation in the presence of appropriate substrates.*

Specific single-site protein modification is a highly desirable goal for a variety of applications in basic and applied research, including spectroscopic analyses of substrate binding, folding processes,[1] catalytic mechanisms, and conformational dynamics,[2] as well as the production of polymer conjugated therapeutic proteins.[3] Whereas a broad range of commercially available chemoselective reagents exists (e.g. maleimides or succinimides),[4] these lack regiospecificity. This limitation is compounded by the general frequency of the corresponding nucleophilic amino acids (lysine and cysteine residues), generating a significant likelihood of heterogeneous product mixtures with unpredictable properties. This may be overcome by site-specific introduction of a reactive group (e.g. cysteine) using site-directed mutagenesis—with the proviso that existing cross-reacting residues might also need to be replaced—or of noncanonical amino acids through genetic code expansion.[5] Despite significant progress in the latter field, however, applications are limited with respect to the variety of the introduced functionality; in addition, protein yields tend to be low due to cellular toxicity of the noncanonical amino acid at the high concentrations necessary.

Enzymes represent the obvious choice for regio- and stereoselective protein modification, and a broad repertoire is found in nature. A universal and facile application in preparative syntheses is hindered, however, by the high specificity and complexity of suitable substrates, as well as the need for elaborate recognition sequences, such as those for phosphopantetheinyltransferase (11 amino acids) and biotin-ligase (15 amino acids).[6] Transglutaminases recognize glutamine residues flanked by specific amino acids that must be situated in a flexible region of the protein. The use of sortase, a promising enzyme for the selective modification of proteins with a five amino acid recognition sequence, has been hampered by requirements for a high excess of the biocatalyst and high amounts of labeling reagent (due to reversibility and single turnover rates),[7] although these limitations can be overcome by the use of depsipeptides.[8]

A universal and flexible method for the selective introduction of artificial functionalities into proteins remains an elusive goal. The amino terminus of a protein represents an attractive target for single-site modification, as per definition every protein chain possesses one and only one N-terminus. As a result of the reversibility of enzymatic reactions, specific N-terminal coupling may be achieved using proteinases, as demonstrated using, for example, bacterial IgA-protease[9] or subtiligase.[10] Nevertheless, proteinases suffer a major drawback: the hydrolysis reaction is generally highly favored over ligation, requiring careful kinetic or thermodynamic control of the reaction. A variety of approaches have been explored to improve discrimination between the competing reactions, including solvent, substrate, and proteinase engineering.[11]

In particular, the concept of substrate mimetics (such as peptidyl-4-guanidinophenyl esters OGp, whose ester leaving groups mimic trypsin-specific arginyl substrate side chains) exhibits promising synthetic potential, allowing the coupling of diverse and even nonpeptidic acyl moieties to target peptide substrates.[12] Here we describe a variant of trypsin

[*] Dr. S. Liebscher,[+] Dr. M. Schöpfel,[+] A. Sharkhuukhen, Dr. A. Pech, Dr. C. Parthier, Prof. Dr. M. T. Stubbs, Prof. Dr. F. Bordusa
Institute of Biochemistry/Biotechnology
Martin-Luther-University Halle-Wittenberg
Kurt-Mothes-Strasse 3, 06120 Halle/Saale (Germany)
E-mail: stubbs@biochemtech.uni-halle.de
frank.bordusa@biochemtech.uni-halle.de

Dr. T. Aumüller, Dr. G. Jahreis
Max-Planck Research Unit for Enzymology of Protein Folding
Weinbergweg 22, 06120 Halle/Saale (Germany)

Dr. E. Höss
Roche Diagnostics GmbH
Nonnenwald 2, 82372 Penzberg (Germany)

[+] S.L. and M.S. contributed equally to this work.

[**] This work was supported by the DFG priority program SPP1623 "Chemoselective reactions for the synthesis and application of functional proteins" and the BMBF-supported network ProNet-T3 "Protein competence center Halle: tools, targets & therapeutics". We thank Angelika Schierhorn for mass spectrometry analysis and Uwe Müller (Free University Berlin at BESSY) for synchrotron time.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201307736.

  

(which we term trypsiligase) that exhibits high specificity for the tripeptide sequence YRH, found in only 0.5% of known protein sequences in the SwissProt database.[13] Trypsiligase is both able to cleave the Y–R peptide bond and to attach diverse moieties to the ensuing amino terminus specifically and quantitatively, under mild aqueous conditions in a one-pot reaction. We show that the source of the proteolytic and ligation selectivity is the zymogen-like nature of trypsiligase: the enzyme achieves a trypsin-like active conformation only in the presence of appropriate substrates.

The primary specificity of trypsin for basic side chains, abundant in target proteins and therefore susceptible to hydrolysis, is governed by D189 at the base of the primary specificity pocket S1 in the wild-type (wt) enzyme.[14] Trypsin variant D189K exhibits low-level esterase activity against P1 tyrosine residues[15] but negligible proteolytic activity. We reasoned that the latter may be restored by substrate affinity enhancement at secondary sites. Introduction of an acidic amino acid at the S1′ site by the exchange K60E results in a trypsin variant whose activity against P1′-arginine is two orders of magnitude greater than that against other residues.[16] A further recognition element is found in the trypsin variant N143H/E151H, which is characterized by a 350-fold increase in activity towards P2′-histidine-containing substrates in the presence of zinc ions.[17] We combined these features to generate the fourfold trypsin variant K60E/N143H/E151H/D189K, trypsiligase (Figure 1). Peptide hydrolysis assays confirmed the desired specificity for the recognition sequence YRH; variation of any of the three positions leads to a significant decrease in catalytic efficiency (Figure S1).

Importantly, trypsiligase recognizes the substrate-mimetic 4-OGp, so that the tailored enzyme should retain synthetic potential. As desired, acyl transfer to peptides containing a P2′-His residue can be effected in the presence of divalent metal ions (Figure 1), facilitating kinetic control of the reaction. To test the efficacy of the designed biocatalyst trypsiligase on protein substrates, we selected the peptidyl-prolyl-*cis/trans*-isomerase (PPIase) human Cyclophilin 18 (hCyp18) (Figure 2; Figure S2), which contains 13 lysine and 6 arginine residues as well as four cysteine residues. In addition to the recognition sequence YRH, hCyp18 was equipped with an N-terminal Strep-tag II fusion for efficient affinity purification.[18] Incubation of the hCyp18 variant with trypsiligase resulted in the complete cleavage of the Strep-tag II peptide within one hour (Figure 2b). No undesired cleavages were detected, even at incubation times of more

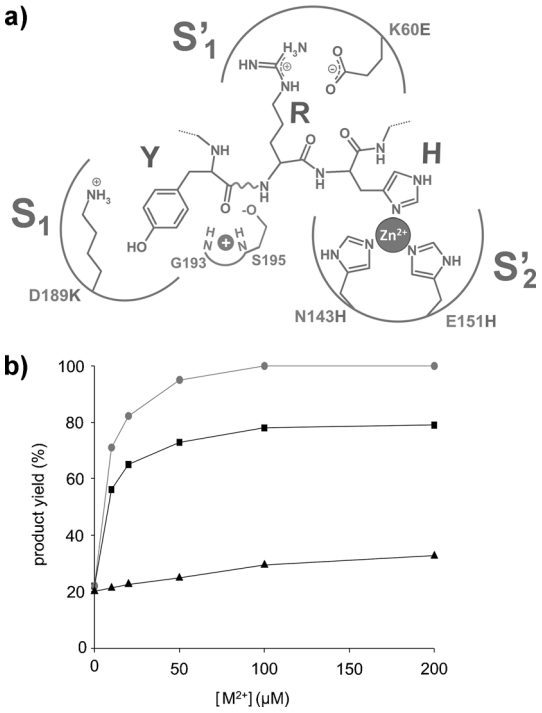


***Figure 1.*** Trypsiligase design and activity. a) Diagram of changes in the binding sites S1, S1', and S2' that allow recognition of the sequence YRH. Scissile bond cleavage (wavy line) is achieved by the reactive-site S195 hydroxyl in conjunction with the oxyanion hole (S195 and G193 backbone amides). b) Acyl transfer of benzoyl-glycyl-4-guanidinophenyl ester (Bz-Gly-OGp) to AHAAG catalyzed by trypsiligase in the presence of $Zn^{2+}$ (-●-), $Ni^{2+}$ (-■-), and the trypsin variant K60E/D189K in the presence of $Zn^{2+}$ (-▲-), analyzed by HPLC. Conditions: 2 mM Bz-Gly-OGp, 15 mM AHAAG, 50 μM trypsin variant, 100 μM $ZnCl_2$ or $NiCl_2$, respectively.

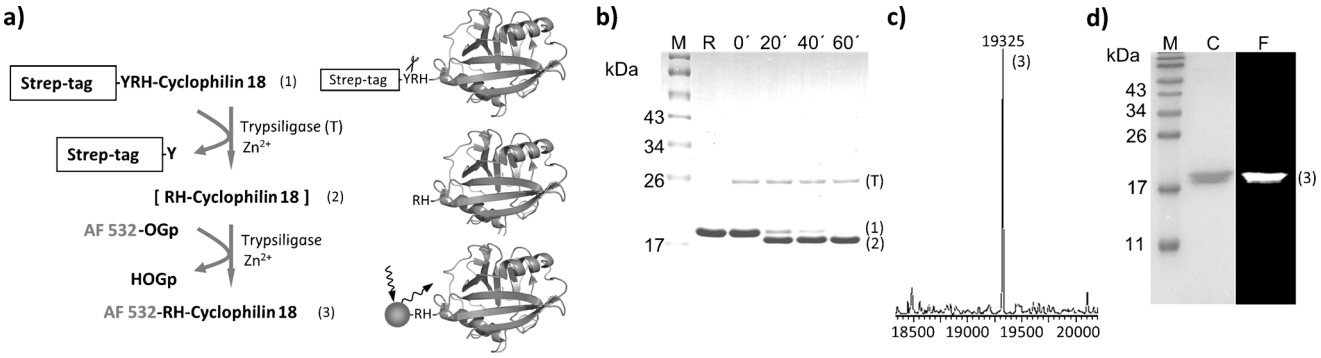


***Figure 2.*** Trypsiligase-catalyzed N-terminal tag cleavage and modification of human Cyclophilin 18. a) Scheme of the N-terminal labeling procedure. b) Time course of trypsiligase-catalyzed cleavage of the Strep-tag II fusion monitored by SDS-PAGE. M: molecular weight marker; R: fusion protein as reference. Conditions: 100 μM Strep-YRH-Cyclophilin 18, 10 μM trypsiligase, 50 μM $ZnCl_2$, 100 mM HEPES, 100 mM NaCl, 10 mM $CaCl_2$, pH 7.8. c) ESI mass spectrum of the reaction mixture at $t = 1$ h (calculated mass for AlexaFluor 532-modified hCyp18: 19322 Da). d) SDS-PAGE of isolated AlexaFluor 532–Cyclophilin 18 product. M: molecular weight marker; C: Coomassie staining; F: in-gel fluorescence.

  

than ten hours. To test the ligation efficacy of trypsiligase, the fluorophore AlexaFluor 532 was added to the reaction as a 4-guanidinophenyl ester linked peptide (Ac-Cys(AlexaFluor 532)-$Ala_2$-OGp; Figure S3) in a threefold excess over hCyp18, and acyl transfer to the N-terminus was followed using UPLC (Figure S4). Complete labeling of hCyp18 was observed within one hour as detected by mass spectrometry (Figure 2c). The isolated AlexaFluor 532-modified hCyp18 variant was fully active with unaltered catalytic efficiency, and the fluorophore showed no change in quantum yield before and after the labeling procedure, indicating that the modification procedure did not influence the dye characteristics. The fluorophore tetramethylrhodamine could also be successfully transferred to the N-terminus of hCyp, where mass spectrometric fragmentation confirmed the regioselective attachment of the dye to the N-terminus (Figure S4).

To explore the versatility of the approach, we applied the procedure to the N-terminal modification of three other proteins of diverse size and secondary structure, using a variety of acyl-4-OGps (Figures S2, S3, and S5). In each case, trypsiligase treatment of the corresponding StrepII-YRH fusion proteins resulted in quantitative cleavage of the tag. Acyl transfer of moieties that include biotin and polyethylene glycol occurred with quantitative product yields, and mass spectrometry confirmed the identities of the protein product derivatives, demonstrating the absence of undesired cleavages or side products. Finally, all modified enzymes retained full activity following the labeling procedure (Table S1).

Trypsiligase crystallized in several crystal forms, including the common trigonal and orthorhombic forms frequently observed for trypsin.[19] Surprisingly, only weak or no ordered electron density was observed in the former for residues I16–G19, G142–P152, L185–D194, and S217–N223 that surround the active site and S1 pocket (Figure 3a, Figure S6). These residues correspond to the activation domain in trypsinogen,[20] which is disordered in the precursor zymogen but becomes structured following cleavage by enteropeptidase to form trypsin: this is the molecular basis for serine proteinase activation. Attempts to soak this crystal form with zinc ions failed to result in any density for the engineered histidine-containing loops (data not shown). On the other hand, the activation domain is ordered in the orthorhombic form, although it is maintained in an inactive conformation as revealed by the interaction between D194 and H40 in the telltale zymogen triad.[21] Soaking with zinc ions revealed clear density for the divalent cation, coordinated by N143H, E151H, and K87# from a symmetry-related molecule (Figure 3b, Figure S6b), although no significant rearrangements are seen in the protein. Cocrystallization with the modified trypsin inhibitor YRH-ecotin resulted in a crystal form containing four complexes in the asymmetric unit. Whereas disorder was still present in the trypsiligase component of each complex, the activation domain adopts a trypsin-like active conformation in three of the four complexes (Figure 3c, Figure S7), with D194 involved in a salt bridge with the amino terminal residue I16 to form the oxyanion hole (through correct positioning of the G193 and S195 backbone amides) that is essential for catalysis. Taken together, these structural data suggest that even after activation cleavage, trypsiligase retains a zymogen-like inactive state. To test this, we investigated the reaction of trypsin, trypsinogen and trypsiligase with the suicide inhibitor 4-(2-aminoethyl)-benzenesulfonyl fluoride (AEBSF), which like PMSF binds covalently and specifically to the reactive serine hydroxy group of active serine proteinases. ESI mass spectrometry clearly demonstrates that trypsiligase is only partially modified by the reagent under the conditions applied, in contrast to trypsin (which is completely modified) and trypsinogen (which remains unmodified) (Figure S8).

These observations suggest a molecular mechanism for the observed ligation reaction (Figure 3d). The conceptual framework of a conformational equilibrium between zymogen and active enzyme (zymogenicity) is well established for the trypsin-like serine proteinase family.[22] In contrast to trypsin, the equilibrium between disordered inactive (trypsinogen-like) and ordered active (trypsin-like) structures is strongly shifted in trypsiligase to the zymogen (disordered) conformation. This type of behavior has also been observed for the trypsin variant D189S,[23] making it likely that the mutation D189K leads to destabilization of the S1 pocket and the activation domain as a whole. Indeed, the carboxylate side chain of D189 in wild-type trypsin makes hydrogen bonds to the main-chain amide groups of S190 (2.7 Å), A221 (3.0 Å), and G226 (3.5 Å) that border the S1 pocket, interactions that are not possible upon mutation to lysine in trypsiligase.

The binding of high-affinity substrates (containing the YRH recognition sequence in the presence of $Zn^{2+}$) shifts the conformational equilibrium, providing sufficient driving force to switch the enzyme into the active conformation. Thus hydrolysis takes place preferentially (and almost exclusively) in the presence of YRH-containing substrates. On the other hand, the residual zymogen-like structure observed in one of the four independent trypsiligase molecules in complex with YRH-ecotin indicates that the shift is incomplete. Transfer of the enzyme-bound acyl-moiety to the specific Arg-His-acceptor peptide (ligation reaction) is therefore favored over transfer to water (hydrolysis). As the product conjugates no longer possess the recognition sequence, the back reaction (proteolysis) is virtually non-existent, so that modification is essentially irreversible and quantitative.

Variation in the integrity of the activation domain is well established for trypsin-like serine proteinases—ranging from the "active zymogen" single-chain tissue plasminogen activator[22] through the "on–off" activation switch of trypsin/trypsinogen,[21] to the virtually inactive blood coagulation factor IXa in the absence of cofactor and substrate[24]—and the existence of intermediate enzymatic forms has been discussed.[25] Furthermore, it is known that the disorder–order transition is essentially reversible: trypsinogen can be forced into an active conformation through strong ligand binding[26] and trypsin into a zymogen-like form upon inactivation by serpins.[27]

To the best of our knowledge, the mode of action of trypsiligase represents the first application of substrate-activated catalysis, a special case of induced fit and a conceptual extension of the established phenomenon of substrate-assisted catalysis. In addition to the conventional

  

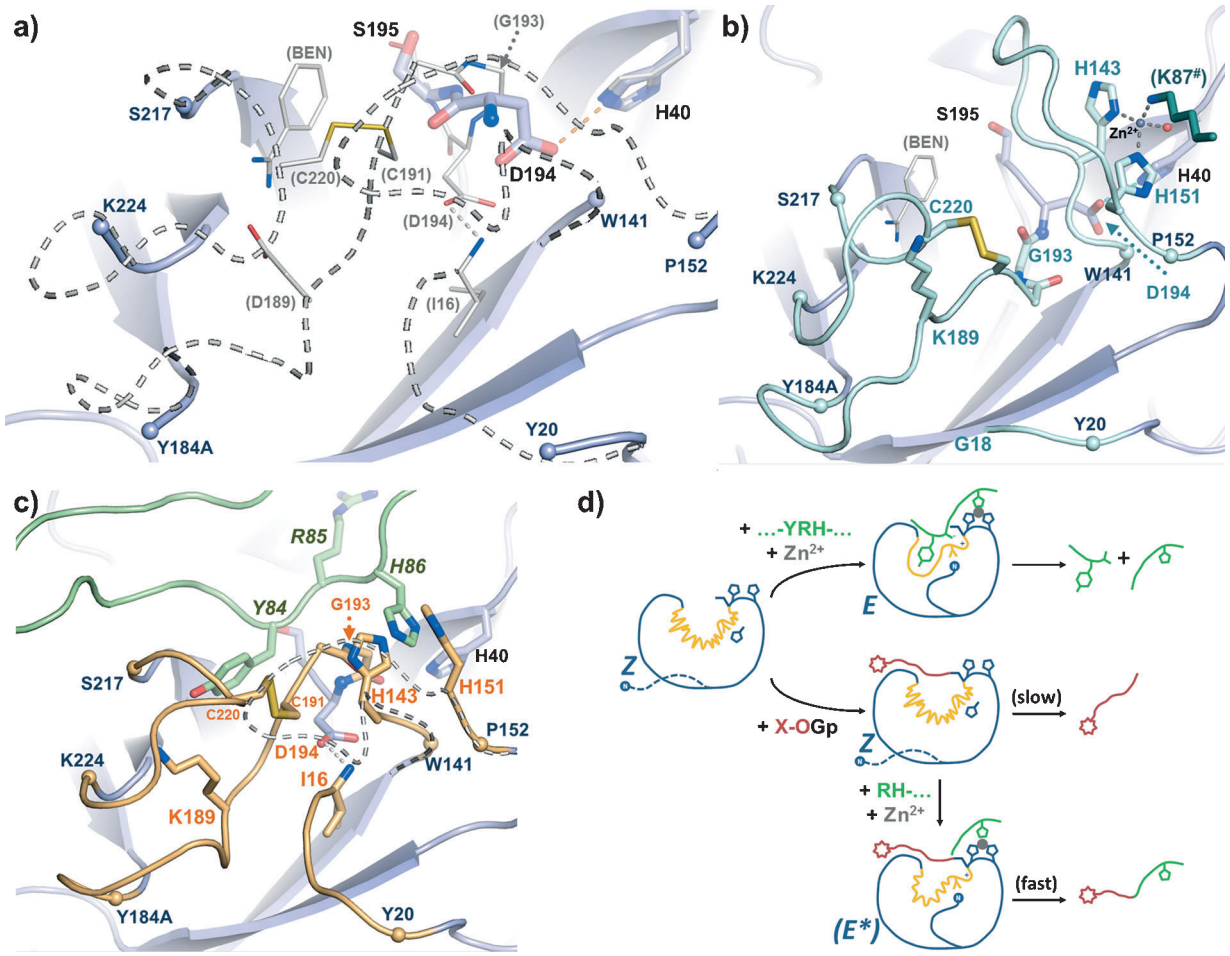


***Figure 3.*** Trypsiligase adopts a zymogen-like structure in the absence of ligands. a) The activation domain is disordered in the trigonal crystal form I. Ordered trypsiligase residues are shown in blue (rods and cartoon; last atoms for which density is observed are denoted by balls); note the interaction between the side chains of D194 and H40 (of the "zymogen triad"). For comparison, wild-type trypsin is shown in white (dashed curves, thin rods; pdb code 1MTS), with the inhibitor benzamidine marking the S1 pocket. Activation cleavage of wild-type trypsinogen leads to salt-bridge formation between D194 and the newly formed I16 N-terminus, resulting in formation of the oxyanion hole (backbone amides of G193 and S195) and reorganization of the S1 pocket (delimited by D189 at the base and the disulfide bridge C191–C220). b) Although residues of the activation domain of trypsiligase (cyan) are defined in orthorhombic crystal form II, neither the oxyanion hole nor the primary specificity pocket are correctly positioned for enzyme activity. The engineered metal-binding site is occupied by a zinc ion, with a symmetry-related molecule providing the final ligand (K87$^{\#}$). c) In complex with the substrate-like inhibitor YRH-Ecotin (green), trypsiligase adopts an active conformation, with the S1 pocket occupied by the tyrosine residue. d) The postulated substrate-mediated activation mechanism for peptide ligation. Trypsiligase adopts a zymogen-like structure (*Z*: flexible N-terminal peptide, interaction of D194 with H40, yellow and blue, respectively), which in the presence of YRH-containing substrates and zinc is converted into the active conformation (*E*: salt bridge between D194 and I16, formation of S1 pocket and oxyanion hole) to result in Y–R bond cleavage. Reactive *O*-guanidinophenyl esters, which can acylate the enzyme in the zymogen(-like) conformation, undergo only slow hydrolysis due to a dysfunctional oxyanion hole; we suggest that the latter is at least partially formed in the presence of zinc and amino-terminal RH-containing peptides (hypothetical state *E**), leading to kinetic favoring of the amidolysis (peptide ligation) reaction.

benefits of enzymes (action under mild conditions with tight control of regio- and stereoselectivity), the mechanism described here opens the door to highly selective protein-modification reactions. Based on this concept, the generation of orthogonal substrate–ligase pairs for tailored applications is conceivable, and the method can easily be accommodated in biotechnological processes. Although the activated state of trypsiligase is destabilized, the assay profits from the robustness of the wild-type enzyme, resulting in a broad tolerance for additives and reaction conditions (data not shown). The highly restricted cleavage activity on the one hand and quantitative acyl transfer activity on the other hand ensure compatibility with well-established methods of protein expression, such as purification tag cleavage, affording the trypsiligase-based method a universal route from gene to selectively modified target protein (Figure S1 d). Finally, it is not unthinkable that this principle has been adopted by nature itself in as yet undiscovered enzymatic reactions, an extension of the combined cofactor and substrate activation observed for complement convertase.[28]

Received: September 3, 2013
Revised: November 14, 2013
Published online: February 12, 2014

  

**Keywords:** biocatalysis · bioconjugation · protein modification · proteinases · reverse proteolysis

[1] B. Schuler, E. A. Lipman, W. A. Eaton, *Nature* **2002**, *419*, 743 – 747.
[2] S. Weiss, *Nat. Struct. Biol.* **2000**, *7*, 724 – 729.
[3] J. M. Harris, R. B. Chess, *Nat. Rev. Drug Discovery* **2003**, *2*, 214 – 221.
[4] E. Baslé, N. Joubert, M. Pucheault, *Chem. Biol.* **2010**, *17*, 213 – 227.
[5] G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, *296*, 1459 – 1462.
[6] M. Rashidian, J. K. Dozier, M. D. Distefano, *Bioconjugate Chem.* **2013**, *24*, 1277 – 1294..
[7] H. Mao, S. A. Hart, A. Schink, B. A. Pollok, *J. Am. Chem. Soc.* **2004**, *126*, 2670 – 2671.
[8] D. J. Williamson, M. A. Fascione, M. E. Webb, W. B. Turnbull, *Angew. Chem.* **2012**, *124*, 9511 – 9514; *Angew. Chem. Int. Ed.* **2012**, *51*, 9377 – 9380.
[9] M. Lewinska, C. Seitz, A. Skerra, F. P. Schmidtchen, *Bioconjugate Chem.* **2004**, *15*, 231 – 234.
[10] D. Y. Jackson, J. Burnier, C. Quan, M. Stanley, J. Tom, J. A. Wells, *Science* **1994**, *266*, 243 – 247.
[11] F. Bordusa, *Chem. Rev.* **2002**, *102*, 4817 – 4868.
[12] M. Thormann, S. Thust, H. J. Hofmann, F. Bordusa, *Biochemistry* **1999**, *38*, 6056 – 6062.
[13] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucleic Acids Res.* **1997**, *25*, 3389 – 3402.
[14] B. S. Hartley, D. M. Shotton, D. B. Paul, in *The Enzymes*, *Vol. Volume 3*, Academic Press, **1971**, pp. 323 – 373.
[15] L. Graf, C. S. Craik, A. Patthy, S. Roczniak, R. J. Fletterick, W. J. Rutter, *Biochemistry* **1987**, *26*, 2616 – 2623.
[16] T. Kurth, S. Grahn, M. Thormann, D. Ullmann, H. J. Hofmann, H. D. Jakubke, L. Hedstrom, *Biochemistry* **1998**, *37*, 11434 – 11440.
[17] W. S. Willett, L. S. Brinen, R. J. Fletterick, C. S. Craik, *Biochemistry* **1996**, *35*, 5992 – 5998.
[18] S. Voss, A. Skerra, *Protein Eng.* **1997**, *10*, 975 – 982.
[19] a) M. T. Stubbs, R. Huber, W. Bode, *FEBS Lett.* **1995**, *375*, 103 – 107; b) M. T. Stubbs, S. Reyda, F. Dullweber, M. Moller, G. Klebe, D. Dorsch, W. W. Mederski, H. Wurziger, *ChemBioChem* **2002**, *3*, 246 – 249.
[20] R. Huber, W. Bode, *Acc. Chem. Res.* **1978**, *11*, 114 – 122.
[21] E. L. Madison, A. Kobe, M. J. Gething, J. F. Sambrook, E. J. Goldsmith, *Science* **1993**, *262*, 419 – 421.
[22] M. T. Stubbs, M. Renatus, W. Bode, *Biol. Chem.* **1998**, *379*, 95 – 103.
[23] E. Szabo, Z. Bocskei, G. Naray-Szabo, L. Graf, *Eur. J. Biochem.* **1999**, *263*, 20 – 26.
[24] T. Zögg, H. Brandstetter, *Structure* **2009**, *17*, 1669 – 1678.
[25] N. Pozzi, A. D. Vogt, D. W. Gohara, E. Di Cera, *Curr. Opin. Struct. Biol.* **2012**, *22*, 421 – 431.
[26] M. Bolognesi, G. Gatti, E. Menegatti, M. Guarneri, M. Marquart, E. Papamokos, R. Huber, *J. Mol. Biol.* **1982**, *162*, 839 – 868.
[27] J. A. Huntington, R. J. Read, R. W. Carrell, *Nature* **2000**, *407*, 923 – 926.
[28] F. Forneris, D. Ricklin, J. Wu, A. Tzekou, R. S. Wallace, J. D. Lambris, P. Gros, *Science* **2010**, *330*, 1816 – 1820.

  