Enzymes for Peptide Cyclization

Frank Bordusa*[a]

KEYWORDS:

bioorganic chemistry · cyclizations · enzyme catalysis · hydrolases · peptides

Indisputably, catalytic procedures represent the best way to the economic solution of synthesis problems. Enzymes imparticular unite high catalytic efficiency and diffusion control of synthesis with an equally high degree of regio- and stereospecificity. Moreover, because enzymes are produced and optimized by living organisms, physiological mild reaction conditions can be used. Together with the possibility of providing large quantities of native or modified enzymes by overexpression, biocatalytic procedures become increasingly attractive to improve existing chemical synthesis methods.

As they were developed by nature, most biocatalysts are proteins. Therefore, the biocatalytic synthesis of peptides and proteins can be seen as a special case in biocatalysis since both the catalyst and the reactants are built from identical basic monomers. Nature solves this problem with the help of a catalytically active RNA, ribosomal peptidyl transferase, which synthesizes proteins by a coordinated interplay of more than 100 kinds of further macromolecules.

The synthesis of short peptides, however, does not inevitably depend on the presence of the ribosomal peptidyl transferase. For example, the biosynthesis of small cyclic peptide antibiotics, like the decapeptide gramicidin S or the immunosuppressant cyclosporin, is achieved by soluble non-ribosomal peptide synthetases (NRPSs). Together with the closely related polyketide synthetases (PKSs), NRPSs have caused considerable interest as biocatalysts for the synthesis of a wide variety of medicinally important natural products.[1] A combination of genetic, protein chemical, and chemical approaches has provided fundamental insights into many properties of these complex systems and has allowed a more rational use of these remarkable biocatalysts.[2] Nevertheless, there are several important unanswered questions.

This account highlights a recent publication by Trauger et al.[3] that reports on the isolation and characterization of the thioesterase domain of tyrocidine synthetase. A short introduction gives essential background information on the architecture and catalytic activity of NRPSs. After presenting the results, their significance will be explained, and the importance of the isolated thioesterase, which is mainly for peptide cyclization, will be discussed in comparison to other biocatalytic approaches.

NRPSs are large, multifunctional proteins that are organized into sets of functional domains termed modules (Figure 1).[4] In this system, amino acids are activated by the formation of enzyme-bound thioesters between the module-specific amino acid and the peptidyl carrier protein (PCP) itself. Panthetheine-4′-phosphate, attached to the side-chain hydroxy group of conserved serine residues, serves as the activating anchor molecule. Due to the specific loading of amino acids, the order of modules corresponds directly to the primary sequence of the final peptide product. In this system, peptide synthesis is initiated by the transfer of the first amino acid to the N^{α}-amino group of the residue linked to the neighboring module, to form an

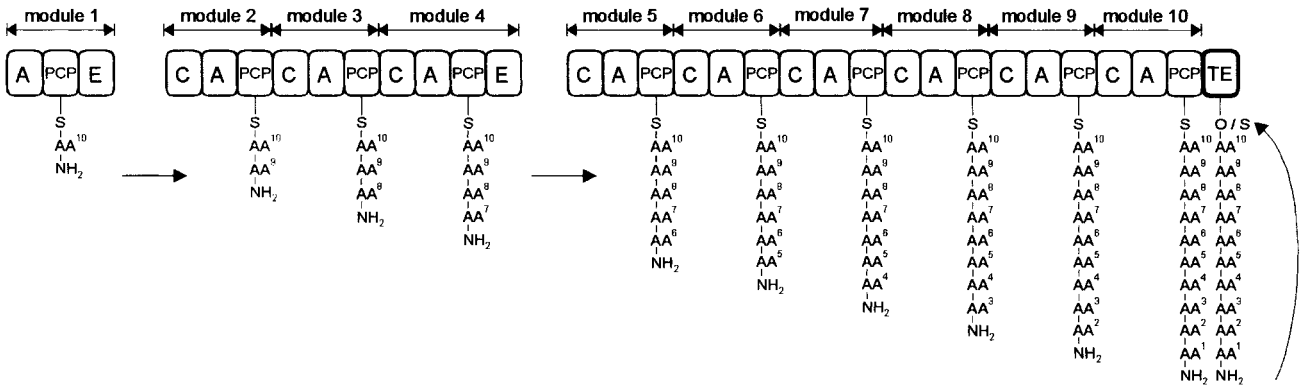


Figure 1. *Biosynthesis of cyclic peptides by non-ribosomal peptide synthetases (NRPSs). Boxes represent functional domains with: A, adenylation (catalyzes amino acid activation); PCP, peptidyl carrier protein; C, condensation (catalyzes peptide bond formation); E, epimerization; TE, thioesterase domain. S represents panthetheine-4′-phosphate linked to the side chain hydroxy group of a serine residue in the PCP domain. O/S represents the active side chain of either a cysteine or a serine residue in the TE domain. AA = amino acid.*

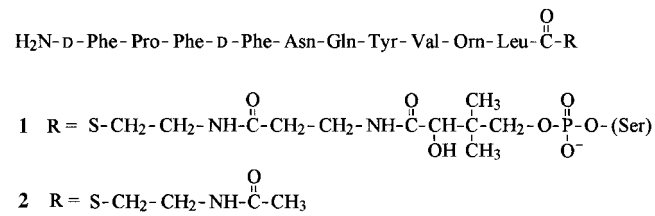
[a] *Dr. F. Bordusa*
Max-Planck Society Research Unit "Enzymology of Protein Folding"
Weinbergweg 22, 06120 Halle an der Saale (Germany)
Fax: (+49) 345-551-19-72
E-mail: bordusa@enzyme-halle.mpg.de

enzyme-bound dipeptide thioester. The process occurs again and finally terminates at the last downstream module; the end result is the appropriate enzyme-bound peptide thioester. Release of the enzyme-bound peptide is usually catalyzed by a C-terminal domain, the thioesterase (TE) domain, which mediates either the hydrolysis of the covalently linked peptide ester or cyclization of its backbone. Interestingly, despite the diverse activities, the TE domain appears to use a catalytic triad in both cases, with either serine or cysteine as the active amino acid residue; this is also known to be the case for proteases.[5]

Attempts to disconnect the TE domain from the corresponding peptide synthetase[6] as well as investigations on isolated TE domains[7] have led to the assumption that a physical linkage, or at least a tight noncovalent interaction, between the TE domain and the upstream modules is essential for complete TE activity. While disconnected and isolated TE domains were found to be capable of hydrolyzing native or artificial substrates, no activity for cyclization could be detected.

Trauger et al. started a new attempt with tyrocidine synthetase from the spore-forming *Bacillus brevis* as the model enzyme; this enzyme catalyzes assembly of the cyclic decapeptide antibiotic tyrocidine A. After overexpression and purification, the isolated TE domain was incubated with an artificial peptide thioester that mimics the natural peptide substrate of the TE domain (Scheme 1). *N*-Acetylcysteamine was used as the activating ester leaving group as it is structurally identical to the terminal section of the pantetheine-4′-phosphate anchor.

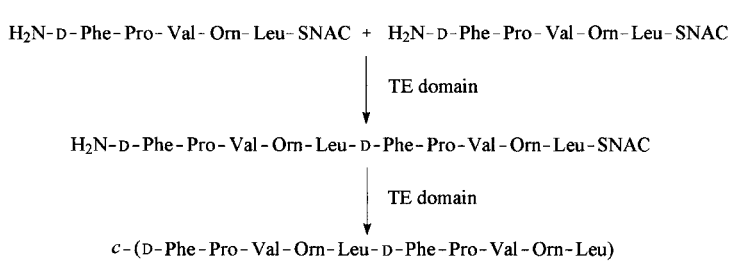
Contrary to former studies,[7] the authors found specific cyclization activity of the isolated TE domain towards the artificial peptide ester, which resulted in the formation of tyrocidine A. Hydrolysis of the decapeptide thioester could only be detected to a minor extent. The authors postulate the absence of upstream modules as the reason for this nontypical background hydrolysis. Although the ester leaving group should not have an influence on the deacylation of the acyl–enzyme intermediate (assuming that the kinetics of the TE reaction follow those of Ser and Cys proteases), a similar leaving-group effect was also found for some protease-mediated peptide synthesis reactions, in particular when artificial substrate esters were used as acyl donor components.[8] Therefore, it remains an open question as to whether the artificial *N*-acetylcysteamine leaving group might also contribute to the background hydrolysis observed.



Scheme 1. *Structures of the natural (**1**) and artificial (**2**) substrates of the tyrocidine synthetase TE domain. The serine residue in **1** is part of the peptidyl carrier protein.*

The replacement of D-Phe[1] with Phe or D-Ala, D-Phe[4] with D-Ala, Orn[9] with Glu, and an Ala scan of the remaining seven amino acid residues of the decapeptide thioester provided the first data on the substrate specificity of the isolated TE domain. It was found that only the substitution of amino acids near the end of the decapeptide, that is, the amino acids in positions 1 and 9, significantly decreases the rate of cyclization, by affecting either the acylation or the deacylation step of the reaction. Deletion or addition of one amino acid moiety in the center of the peptide substrate did not abolish the cyclization activity of the TE domain; these experiments resulted in the formation of nine- and eleven-membered cyclopeptides, respectively. Finally, it was demonstrated that the isolated TE domain catalyzes not only peptide cyclization but also dimerization of two repetitive pentapeptide thioesters followed by the subsequent cyclization, which leads to the appropriate cyclic decapeptide (Scheme 2). By using gramicidin S as the model peptide product for this reaction, the mechanism postulated for pentapeptide dimerization and cyclization by gramicidin S synthetase could be confirmed.[9]

Indisputably, the finding that an isolated TE domain of a NRPS is capable of catalyzing not only hydrolysis of artificial peptide esters, but also dimerization and cyclization is an important conclusion of this work. This clarifies the native function and catalytic role of the TE domain within the multifunctional NRPS complex and that is certainly of high general interest. Furthermore, it opens up the way to the detailed determination of the TE substrate specificity independently of upstream domains, which cannot be achieved with the full-length protein. The finding that delimited domains can evolve full biological function is in accordance with findings for other proteins of multifunctional architecture, such as FKBP 59[10] and the *E. coli* trigger factor,[11] and supports the assumption that multifunctional proteins may be the result of exon shuffling during evolution. From the synthetic point of view, the isolated TE domain could be highly valuable especially for the engineered biosynthesis of new cyclic compounds. To my knowledge there are



Scheme 2. *Dimerization of the pentapeptide ester H_2N-D-Phe-Pro-Val-Orn-Leu-SNAC (SNAC = N-acetylcysteamine thioester) and cyclization of the resulting decapeptide ester to form gramicidin S, as catalyzed by the tyrocidine synthetase TE domain.*

only a very limited number of papers that report on the enzyme-catalyzed cyclization of peptides in vitro.

Wells and co-workers described a protease-based method with a mutant of subtilisin BPN′ (subtiligase) to synthesize head-to-tail cyclic peptides.[12] Peptide glycolate phenylalanylamide esters of chain lengths between 10 and 31 amino acids and with unprotected side chains were used as the linear precursors. The researchers found that peptide esters shorter than 12 residues only hydrolyze or dimerize, but do not cyclize. In the case of the longer peptides a cyclization could be detected. The yields for cyclization ranged from 30 – 85% and the efficiency usually increases along with the length of the peptide. The authors suggested that longer peptides may be more flexible and, therefore, better able to adopt a productive binding conformation. This finding is all the more significant as the existing classical chemical methods are mostly inefficient for the cyclization of peptides longer than 10 residues due to the large entropic barriers of such reactions.[13] The efficiency of cyclization appears to depend on the sequence of the peptide as well as the peptide length. In general, subtiligase prefers large hydrophobic residues at the donor site (P_1 position; nomenclature according to ref. [14]) and nonpolar residues at the acceptor site of the peptide (P_1' position).[15] At the remaining positions a variety of sequences are accepted by the enzyme, which makes this method a rather general one for the synthesis of larger cyclic peptides.

Recently, Smithrud et al. reported on the formation of cyclic peptides catalyzed by a catalytically active antibody which was originally designed for the formation of noncyclic dipeptides.[16] Starting from the *para*-nitrophenyl ester of the hexapeptide D-Trp-Gly-Pal-Pro-Gly-Phe (Pal, 3-pyridylalanine, which was used instead of the initial Phe[3] to improve the solubility), they could demonstrate that the so-called antibody ligase 16G3 catalyzes head-to-tail cyclization to give *c*-(D-Trp-Gly-Pal-Pro-Gly-Phe). The rate enhancement for the antibody-mediated cyclization was found to be 22-fold compared to the background reaction; in other words, in absolute rate terms, 1 μM of antibody active sites form 2 μM of cyclic products per minute. This catalytic activity was sufficient to form the desired cyclic peptide in greater than 90% yield. Interestingly, neither epimerization nor hydrolysis of the peptide ester could be detected. Substitution of the amino acid residues of the hexapeptide at the coupling positions (positions 1 and 6) by Trp or D-Phe (position 1) and D-Phe or Trp (position 6) significantly reduced the rate of enhancement. Although a similar effect was also found for the TE-domain- and subtiligase-mediated cyclization, this finding reflects the extremely high, hapten-induced specificity of antibodies. In contrast to the TE domain and subtiligase, the antibody ligase acts rather as a template to channel the activated linear peptide ester into formation of the desired cyclic product than as a common enzyme (Figure 2). On

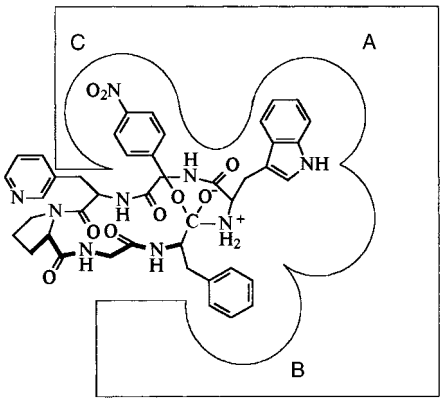


Figure 2. *Representation of the proposed tetrahedral intermediate for the 16G3-catalyzed cyclization of a linear hexapeptide according to Smithrud et al.[16] Only the diastereomer containing L-Phe and D-Trp will bind to pockets A and B optimally. The remaining four amino acids are presumed to reside largely outside the binding pocket.*

the other hand, it can be expected that the length and composition of the linear peptide may be not a limitation for antibody catalysis because antibodies can be tailor-made to recognize those particular side chains that are involved in the ring closure. However, due to the relatively low catalytic efficiency of such reactions it is questionable whether antibodies can reach practical relevance in the near future.

A clever and only recently published approach represents the synthesis of cyclic peptides by the use of the intein (internal protein) strategy. In nature, inteins catalyze a multistep protein modification, in which they are excised from a precursor fusion protein by breaking two peptide bonds while religating the flanking domains (exteins) into a contiguous polypeptide joined by a new peptide bond.[17] Controllable fission of the peptide bonds at either the C or the N terminus of the intein has allowed the production of recombinant N-terminal cysteine proteins and αthioester proteins, respectively. As proposed in the classical works of Wieland and co-workers[18] and Brenner et al.,[19] cysteine and synthetic peptides with an N-terminal cysteine moiety reacted selectively with αthioester-tagged peptides through a transesterification reaction, which was followed by an S→N acyl shift to finally result in the formation of a native peptide bond.[20] This technology, termed intein-mediated or -expressed protein ligation,[21] has been used not only for the ligation of linear peptide fragments, but also for catalyzing head-to-tail peptide ligation. For the latter, the N-terminal cysteine and the αthioester moiety must be located in one single peptide.

Two different approaches have been developed for synthesizing these peptide thioesters containing an N-terminal cysteine, which both utilize inteins with N-terminal cleavage activity to produce the αthioester moiety, but which differ in the way the N-terminal cysteine is liberated. In one approach, the cysteine of the target peptide is masked by an additional sequence that can be removed selectively by the use of highly specific proteases, a system that allowed, for example, a circular version of an isolated SH3 domain to be generated.[22]

In the twin (two inteins) system, the target peptide is cloned between two modified inteins (Figure 3) and both activated ends are generated by intein cleavage.[23] In vitro an additional chitin binding domain present on one or both of the inteins allowed the immobilization of the desired precursor protein on chitin resin, whereas endogenous *E. coli* proteins could be washed away. Thiol-induced cleavage of the intein at the C terminus of the target peptide (situated on intein 2) produced the appropriate αthioester, which then reacted with the N-terminal cysteine of the same target peptide to

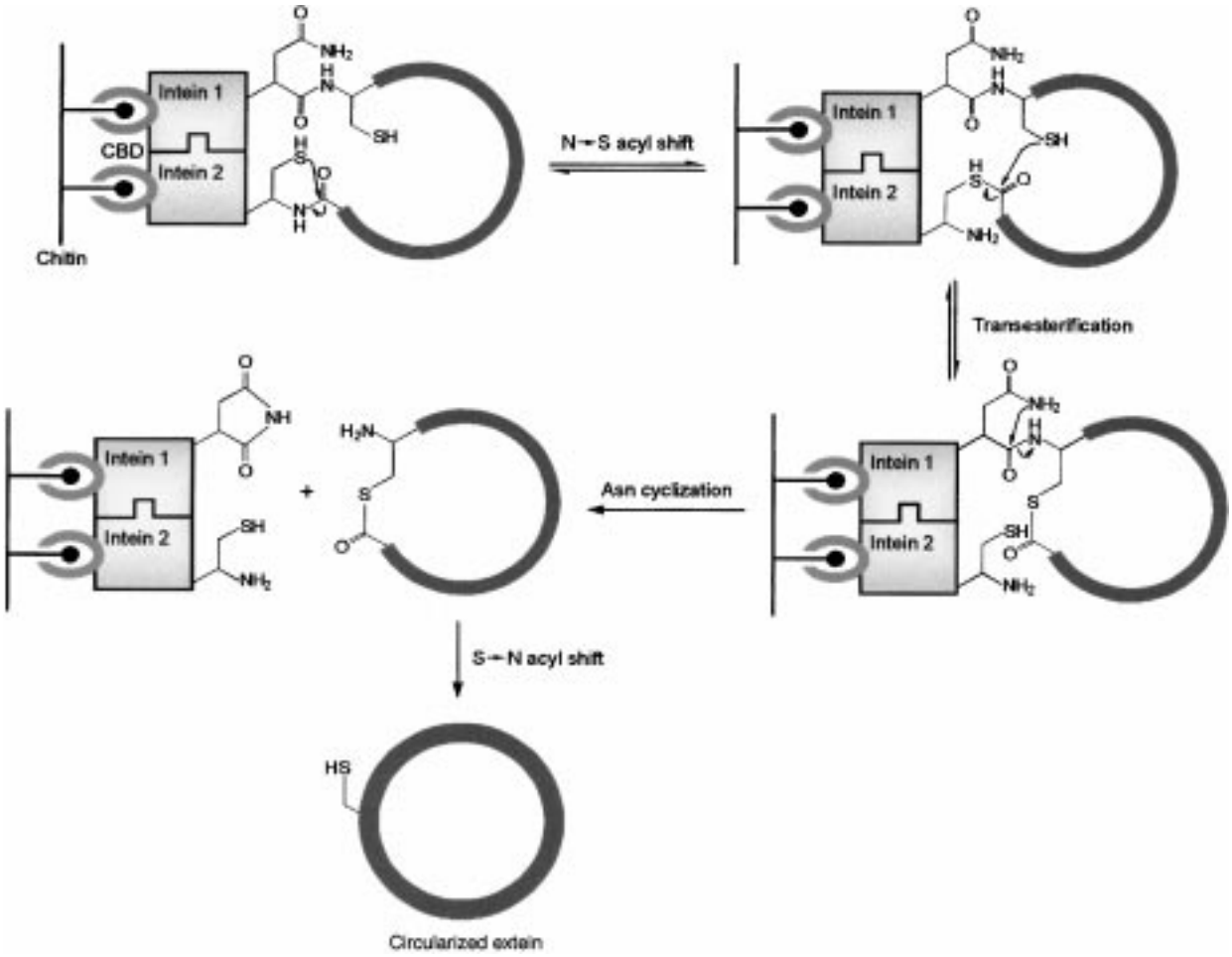


Figure 3. *Peptide cyclization by the twin (two intein) system. Induced by a thiol reagent, such as 2-mercaptoethanesulfonic acid, intein 2 undergoes an N→S(O) acyl shift at the target-peptide/intein 2 junction to generate a thioester intermediate. The following transesterification with a thiol or hydroxy side-chain functionality at the target-peptide/intein 1 junction results in the formation of a lariat intermediate. The further process involves Asn side-chain cyclization, liberation of the cyclic peptide as a lactone, and an S(O)→N acyl shift that generates the thermodynamically favored lactam product. Purification of the precursor protein is simplified by a chitin binding domain (CBD) that binds to chitin resin.*

give the desired cyclic product. The use of the twin system enabled several proteins to be produced in their circular versions.

Generally, it was found that a linker sequence between each end of the protein usually enhances the efficiency of cyclization, mostly due to the repression of competitive polymerization. The cyclization reaction of thioredoxin (135 amino acids), for example, occurred in $>80\%$ yield, with nine and three amino acids added to the N and C termini of the protein, respectively.[23] Interestingly, the cyclization of small peptides can also be achieved in high yields by this system, as demonstrated for peptides containing 9, 10, and 14 amino acids.[23] In a recent paper Scott et al. supposed that favorable interaction of the two inteins could be an important driving force of this highly efficient catalysis.[24] This may be also reflected by the finding of the authors that N-terminal serine is a viable substitute for cysteine, as shown by the synthesis of the cyclic tyrosinase inhibitor pseudostellarin F (*c*-(Ser-Gly-Gly-Tyr-Leu-Pro-Pro)).

Apart from nontypical exceptions,[22] a cysteine, serine, or threonine residue, however, is mechanistically essential to serve as a nucleophile in the intein-catalyzed transesterification reaction; it then remains behind in the cyclic product. Thus, peptides with sequences devoid of those amino acid moieties are usually not targets of this technology. It was further found that the sequence of the target peptide at least two amino acids from the scissile peptide bonds can have significant effects on intein activity, while the vector places no other constraints on the target length or composition.[24]

Generally, peptides with both an N-terminal cysteine and an $^{\alpha}$thioester moiety can be achieved alternatively by chemical methods. The use of those methods enables cyclic peptides to be generated from their fully unprotected linear precursors by native chemical ligation without the help of a ligase. Utilizing this procedure several cyclic peptides, ranging from 15 – 47 residues, could be synthesized successfully.[25] Cyclization of the peptides were achieved in solution and directly on the copolymer support, and high yields could be obtained in most cases. Since most other chemical methods are rather inefficient for cyclization of longer peptides, native chemical ligation is highly useful to provide direct access to

larger cyclic peptides. It must be noted, however, that the high yields obtained could be (at least partly) the result of conformational factors, since the N and C termini of the targets that were chosen for cyclization are very close in the corresponding native peptides. General drawbacks of this method are the necessity of synthesizing peptide thioesters, which are often difficult to prepare, and the need for a cysteine (or closely related amino acid) at the N terminus of the linear precursor peptide. The latter, however, does not inevitably result in the presence of cysteine moieties in the final products. For example, desulfurization of the cysteine after chemical ligation converts the cysteine into an alanine, a method that allowed non-cysteine linear and cyclic peptides to be synthesized.[26]

The capability of the isolated TE domain to catalyze not only the intramolecular but also the intermolecular ligation of peptides was also found for the other "ligases". Although the authors do not provide data on the yield of dimerization, the kinetic parameters given for dimerization and cyclization indicate that the isolated TE domain catalyzes the dimerization of the pentapeptide less efficiently than the cyclization of the resulting decapeptide. This may be the reason why a further elongation of the dimerized peptide ester could not be detected. Therefore, the question remains open as to whether excised TE domains can be used as suitable biocatalysts for the ligation of peptide fragments without (unwanted) cyclization. Attempts with longer peptide ester substrates may furnish clarification. Furthermore, it may be a big challenge to channel the ligation reaction to form homogeneous peptide products, when mixed peptide fragments with distinct sequences and lengths are used in their N^{α}-unprotected fashion, as is required. Finally, the minor amount of hydrolysis activity found for peptide cyclization with the TE domain should initiate further studies to address this effect on the molecular level. The results could provide an important input to the engineering of hydrolases and proteases, which also use a catalytic triad and a similar catalysis mechanism, to further improve their synthetic utility.

[1] D. A. Hopwood, *Chem. Rev.* **1997**, *97*, 2465 – 2497.

[2] a) L. Katz, *Chem. Rev.* **1997**, *97*, 2557 – 2575; b) C. Koshla, *Chem. Rev.* **1997**, *97*, 2577 – 2590; c) J. Staunton, B. Wilkinson, *Chem. Rev.* **1997**, *97*, 2611 – 2629.

[3] J. W. Trauger, R. M. Kohli, H. D. Mootz, M. A. Marahiel, C. T. Walsh, *Nature* **2000**, *407*, 215 – 218.

[4] a) D. E. Cane, C. T. Walsh, C. Khosla, *Science* **1998**, *282*, 63 – 68; b) M. A. Marahiel, T. Stachelhaus, H. D. Mootz, *Chem. Rev.* **1997**, *97*, 2651 – 2674.

[5] a) A. Witkowski, J. Naggert, B. Wessa, S. Smith, *J. Biol. Chem.* **1991**, *266*, 18514 – 18519; b) M. Pazirandeh, S. S. Chirala, S. J. Wakil, *J. Biol. Chem.* **1991**, *266*, 20946 – 20952; c) A. Witkowski, J. Naggert, H. E. Witkowska, Z. I. Randhawa, S. Smith, *J. Biol. Chem.* **1992**, *267*, 18488 – 18492; d) D. M. Lawson, U. Derewenda, L. Serre, S. Ferri, R. Szittner, Y. Wie, E. A. Meighen, Z. S. Derewenda, *Biochemistry* **1994**, *33*, 9382 – 9388; e) J. Li, R. Szittner, Z. S. Derewenda, E. A. Meighen, *Biochemistry* **1996**, *35*, 9967 – 9973.

[6] E. Guenzi, G. Galli, I. Grgurina, E. Pace, P. Ferranti, G. Grandi, *J. Biol. Chem.* **1998**, *273*, 14403 – 14410.

[7] a) R. Aggarwal, P. Caffrey, P. F. Leadlay, C. J. Smith, J. Staunton, *J. Chem. Soc. Chem. Commun.* **1995**, *15*, 1519 – 1520; b) R. S. Gokhale, D. Hunziker, D. E. Cane, C. Koshla, *Chem. Biol.* **1998**, *6*, 117 – 125.

[8] N. Wehofsky, J.-D. Wissmann, M. Alisch, F. Bordusa, *Biochim. Biophys. Acta* **2000**, *1479*, 114 – 122.

[9] a) K. Hori, Y. Yamamoto, T. Minetoki, T. Kurotsu, M. Kanda, S. Miura, K. Okamura, J. Furuyama, Y. Saito, *J. Biochem. (Tokyo)* **1989**, *106*, 639 – 645; b) K. Turgay, M. Krause, M. A. Marahiel, *Mol. Microbiol.* **1992**, *6*, 529 – 546.

[10] B. Chambraud, N. Rouviere-Fourmy, C. Radanyl, K. Hsiao, D. A. Peattie, D. J. Livingston, E. E. Baulieu, *Biochem. Biophys. Res. Commun.* **1993**, *196*, 160 – 166.

[11] a) G. Stoller, T. Tradler, K. P. Rücknagel, J.-U. Rahfeld, G. Fischer, *FEBS Lett.* **1996**, *384*, 117 – 122; b) T. Zarnt, T. Tradler, G. Stoller, C. Scholz, F. X. Schmid, G. Fischer, *J. Mol. Biol.* **1997**, *271*, 827 – 837.

[12] D. Y. Jackson, J. P. Burnier, J. A. Wells, *J. Am. Chem. Soc.* **1995**, *117*, 819 – 820.

[13] F. Cavelier-Frontin, S. Achmed, J. Verducci, R. Jacquier, *J. Mol. Struct.* **1993**, *286*, 125 – 130.

[14] I. Schechter, A. Berger, *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157 – 167.

[15] a) L. Abrahmsen, J. Tom, J. P. Burnier, K. A. Butcher, A. Kossiakoff, J. A. Wells, *Biochemistry* **1991**, *30*, 4151 – 4159; b) T. K. Chang, D. Y. Jackson, J. P. Burnier, J. A. Wells, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 12544 – 12548.

[16] D. B. Smithrud, P. A. Benkovic, S. J. Benkovic, V. Roberts, J. Liu, I. Neagu, S. Iwama, B. W. Philips, A. B. Smith III, R. Hirschmann, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 1953 – 1958.

[17] a) M.-Q. Xu, F. B. Perler, *EMBO J.* **1996**, *15*, 5146 – 5153; b) F. B. Perler, M.-Q. Xu, H. Paulus, *Curr. Opin. Chem. Biol.* **1997**, *1*, 292 – 299; c) F. B. Perler, *Cell* **1998**, *92*, 1 – 4; d) H. Paulus, *Chem. Soc. Rev.* **1998**, *27*, 375 – 386.

[18] T. Wieland, E. Bokelmann, L. Bauer, H. U. Lang, H. Lau, *Justus Liebigs Ann. Chem.* **1953**, *583*, 129 – 139.

[19] M. Brenner, J. P. Zimmermann, J. Wehrmüller, P. Quitt, A. Hartmann, W. Schneider, U. Beglinger, *Helv. Chim. Acta* **1957**, *40*, 1497 – 1505.

[20] a) P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776 – 779; b) J. P. Tam, Y.-A. Lu, C.-F. Liu, J. Shao, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 12485 – 12489.

[21] a) T. C. Evans, Jr., J. Benner, M.-Q. Xu, *Protein Sci.* **1998**, *7*, 2256 – 2264; b) S. Mathys, T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705 – 6710; c) K. Severinov, T. W. Muir, *J. Biol. Chem.* **1998**, *273*, 16205 – 16209; d) T. C. Evans, Jr., J. Benner, M.-Q. Xu, *J. Biol. Chem.* **1999**, *274*, 3923 – 3926.

[22] J. A. Camarero, T. W. Muir, *J. Am. Chem. Soc.* **1999**, *121*, 5597 – 5598.

[23] T. C. Evans, Jr., J. Benner, M.-Q. Xu, *J. Biol. Chem.* **1999**, *274*, 18359 – 18363.

[24] C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnon, S. J. Benkovic, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13638 – 13643.

[25] J. A. Camarero, G. J. Cotton, A. Adeva, T. W. Muir, *J. Peptide Res.* **1998**, *51*, 303 – 316.

[26] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526 – 533.