Adding New Tools to the Arsenal of Expressed Protein Ligation

Nediljko Budisa*[a]

"My entire yearning is directed toward the first synthetic enzyme." Emil Fischer (in a letter to Adolf Baeyer, 1905)[+]

The "peptide theory" put forward in 1902 by Emil Fischer and Franz Hofmeister correctly postulated that proteins are made up of α-amino acids that are linked head-to-tail by amide bonds.[1] However, it is less well known that one of Fischer's main goals was the total chemical synthesis of an enzyme molecule. This dream has been realized almost half a century later. Great advances in the chemical synthesis of peptides and smaller proteins, including solid-phase peptide synthesis,[2] have led to the synthesis of the first enzyme, ribonuclease S.[3,4] These initial results have been followed by a series of successful syntheses of a variety of enzymes.[5] The recent synthesis of green fluorescent protein (GFP) by Sakakibara's group[6] certainly crowns these efforts.

Parallel to these developments, the strategy of a convergent assembly (i.e., condensation) of synthetic and natural peptide fragments termed "protein semisynthesis" has also emerged. The basic requirements for semisynthesis are: i) the synthetic donor peptide has to be protected and activated, ii) an acceptor protein fragment, which has to be prepared by enzymatic or chemical fragmentation of the parent protein, should be available and properly protected. Offord and Rose pioneered the use of hydrazone-, and oxime-forming reactions for chemically ligating such fragments.[7] Although these chemistries are selective, they were in practice often hampered by the insolubility of the large protected peptide building blocks.

Early studies by Wieland[8] on the thiol–thioester exchange reaction, and the procedures by Blake[9] and Yamashiro[10] for thioester preparation by solid-phase synthesis paved the way for generating native amide bonds between peptide fragments through a spontaneous *S*→*N*-acyl shift. A step further in this development was the achievement of a thiol–thioester exchange reaction between unprotected fragments in aqueous solution.[11] The N-terminal fragment contains a C-terminal electrophilic α-thioester that can be conjugated to the N-terminal thiol-harboring fragment through a thiol–thioester exchange reaction, as shown in Scheme 1 A. This technique, termed "native chemical ligation" (NCL) was developed in Kent's laboratory.[12] Most recently, recruitment of the Staudinger ligation[13,14] for chemical ligations represents an additional great breakthrough in the field.[15,16] From a purely chemical perspective, it is an excellent tool for protein/peptide ligation that allows different protein/peptide fragments to be coupled at any desired position, and not only at XXX-Cys bonds (Scheme 1 B).

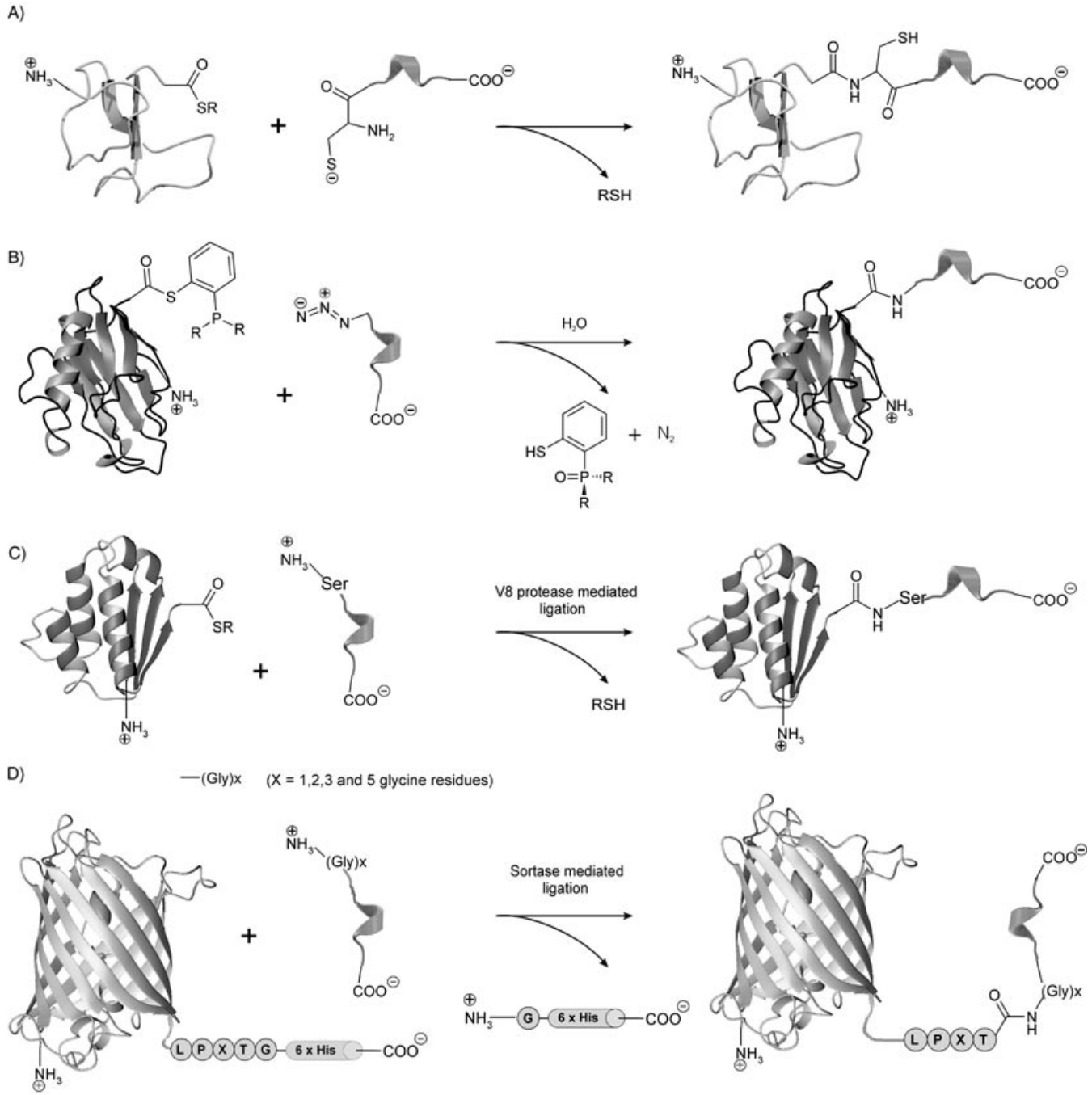
Two of the most striking examples that demonstrate the advantages of these methods are the synthesis of an all-D chiral form of the HIV-1 protease (100 residues)[12] and the preparation of the post-translationally modified artificial variant of erythropoetin (polymer-modified; 166 residues).[17] These examples illustrate the considerable potential of NCL as a complementary approach to protein engineering methods based on ribosome-mediated protein synthesis. The potential of NCL for the introduction of noncanonical amino acids and biophysical probes into peptides and proteins, total isotopic labeling, and chemistries for homogeneous preparation of post-translationally modified proteins are well documented and have recently been comprehensively reviewed.[18–20] Since all these aspects can also be covered by the reprogramming of the ribosome-mediated translational apparatus,[21] the method of choice will ultimately be dictated by the envisaged practical applications. On the other hand, it seems that both NCL and Staudinger ligations currently face no rivalry in experimental design of proteins that include sequential isotopic labeling, preparation of circular proteins, and insertion of non-native polypeptide fragments or nonpeptide molecules at predefined sites.

Besides these pure chemical-ligation strategies, biochemical approaches that involve molecular biology techniques have emerged from a burgeoning revolution in proteomics that is fuelling the need for proteins with tailored modifications. After the discovery of protein splicing in 1987[22] and its subsequent recruitment by Muir and co-workers[23] for chemoselective ligations, the expressed protein ligation (EPL) technique was developed. EPL essentially exploits the same principles as NCL but in addition takes advantage of recombinant-DNA technology to generate protein fragments by ribosomal synthesis. In this way large proteins become accessible for chemoligation. Intein-mediated protein splicing is a process that consists of a series of intramolecular reactions that lead to the excision of inteins from a larger precursor protein and the concomitant ligation of the flanking polypeptide segments, called exteins.[19] Genetically modified inteins have been designed that impair self splicing, so that partially processed C-terminal thioester-tagged recombinant proteins can be trapped and submitted to routine NCL

[a] *Dr. N. Budisa*
Max-Planck-Institut für Biochemie
Am Klopferspitz 18a, 82152 Martinsried (Germany)
Fax: (+49) 89-8578-3516
E-mail: budisa@biochem.mpg.de

[+] *This letter is quoted in Nature's Robots—A History of Proteins* by C. Tanford, J. Reynolds, Oxford University Press, New York, **2001**, p. 43)

  DOI: 10.1002/cbic.200400141



Scheme 1. *Some basic approaches to chemoselective ligations. A) Native chemical ligation (NCL) is the reaction that takes place in water at or around neutral pH between unprotected peptide fragments that contain the requisite reactive groups (N-terminal Cys and C-terminal thioester).*[12] *In the Scheme presented above, NCL is characterized by transthioesterification of the thioester by the thiol function of an N-terminal Cys, followed by the spontaneous S→N-acyl shift to obtain the native peptide bond.*[8] *Other chemistries are also possible, as has been extensively reviewed elsewhere.*[45] *B) Staudinger ligation,*[13] *as an alternative ligation method, forms an amide bond from an azide and a specifically fictionalized phosphine (in this particular case triarylphosphine). Since its final product has no residual atoms, this approach is often termed as "traceless".*[16] *In fact, this method allows independent amino acid sequences to be coupled at any desired Xxx–Yyy bond and is in this respect potentially as universal, if not more so, as the EEL and sortase-mediated ligation (vide infra). C) To bypass the requirement for cysteine at the ligation site, expressed enzyme ligation (EEL) uses specific thioesters (generated by "classical" intein-mediated approaches) as substrate mimetics for a* Staphylococcus aureus *V8 serine protease.*[35] *D) Sortase from* Staphylococcus aureus*, a membrane-anchored transpeptidase, cleaves any polypeptide provided with a C-terminal sorting signal, between the threonine (T) and glycine (G) of the LPXTG motif. Then it catalyses the formation of an amide bond between the carboxyl group of threonine and the amino group of pentaglycine from cell-wall peptidoglycans.*[46] *The extension of this strategy to tagged green fluorescent protein (Nt-GFP-LPXTG-6His-Ct) enables its successful conjugation with various donor molecules containing two or more N-terminal glycines (*D *and* L *peptides, nonpeptide fragments and even other GFPs).*[42]

by following the same principles as described in Scheme 1 A. This approach (often called intein-mediated ligation or IML) for the generation of C-terminally thioester-tagged proteins is now even commercially available (e.g., IMPACT® system from New England Biolabs). Conversely, the generation of proteins and peptides that contain an N-terminal Cys by expression systems strictly rely on the cleavage of appropriate precursor amino acids or of sequences (e.g., endogenous methionylpeptidases or the use of exogenous proteases such as factor Xa).

The ligation of fragments by EPL is general, efficient, and robust. Furthermore, other nucleophilic residues might be used to replace the N-terminal Cys (so-called cysteine mimetics) in the donor peptide or protein fragments. Examples include homocysteine,[24] selenocysteine,[25,26] thiaproline,[27] selenohomocysteine,[28] glycine,[29] and histidine.[30] In exceptional cases and only when thiol groups can be removed without harmful effects, it is possible to substitute cysteine with alanine in the ligation site by desulfuration with palladium or Raney nickel.[31] The already mentioned Staudinger ligation[13] might, in this context, represent an alternative not only to the classical NCL but also to EPL approaches since it does not require a donor fragment containing an N-terminal residue with a nucleophilic side chain. The Staudinger ligation (Scheme 1 B) can also be employed in combination with NCL in order to generate large proteins from more than two fragments, as illustrated by the synthesis of isotopically labeled ribonuclease A.[32] Interestingly, Tirrel, Bertozzi, and co-workers demonstrated that the noncanonical methionine surrogate azidohomoalanine can be introduced into proteins.[33] Such proteins can be selectively modified in the presence of other cellular proteins by means of the Staudinger reaction. Since in most *E. coli* proteins methionine is the N-terminal amino acid, it remains to be seen to what extent this approach (based on an in vivo expanded amino acid repertoire) can be combined with native chemical ligation.

Nevertheless, both NCL and EPL are severely limited in the choice of residues at the ligation site. For this reason, various attempts have been made to overcome the requirement for cysteine in the ligation reaction. Although some of the above-mentioned examples were successful, this feature still limits the flexibility of both methodologies. An enzyme-catalyzed condensation could offer an interesting alternative to these chemoligation procedures. This possibility was demonstrated in 1938 by the papain-catalyzed synthesis of benzoyl-leucyl-leucine anilide.[34] More recently, Machova et al. reported an approach for expressed enzymatic ligation (EEL) that combines the advantages of EPL and the substrate-mimetic strategy of protease-mediated ligation (Scheme 1 C).[35] The thioester substrate mimetic (i.e., the specific leaving group S-CH_2-COO^-) is generated by routine EPL by using the intein-chitin binding tag. Ligation with a synthetic peptide containing an N-terminal serine residue was catalyzed by V8, the Glu/Asp-specific serine protease from *Staphylococcus aureus*.

On the other hand, there are currently very few enzyme-based approaches for peptide ligation beside EPL. For example, the earlier studies of Kaiser and co-workers on subtilisin[36] provided a solid base for the success of Wells and associates in engineering an active site for this enzyme.[37] They generated an enzyme ("subtiligase") capable of efficiently catalyzing the ligation of peptide fragments.[38] Subtiligase exhibit a largely reduced proteolytic activity and is functionally active as an acyltransferase.[39] This property was exploited for enzymatic condensation of six peptide fragments of ribonuclease A (each 12–30 residues long, one of them containing the noncanonical amino acid 4-fluorohistidine). Recently, Bordusa reported that nonactivable zymogens have a potential as novel catalysts for peptide synthesis.[40] Zymogens are slightly active towards activated peptide esters in an irreversible fashion. On the other hand, efficient manipulation is necessary in order to shift the native enzyme activity from cleavage to synthesis of peptides. At this stage of development, the rates of synthetic reactions with nonactivable zymogens are still too low and are therefore not suitable for preparative purposes.[41]

Most recently, with the report from Mao et al., the sortase-catalyzed proteolysis reaction entered the arena of enzyme-mediated native-protein ligation.[42] Sortases are bacterial enzymes that are responsible for the covalent attachment of specific proteins to the cell wall of Gram-positive bacteria.[43] These enzymes have been proved to play a key role in the pathogenicity of these bacteria. Proteins that are substrates of sortase, have a "sorting signal" at the C terminus that consists of the LPXTG motif (where X could be any amino acid), a hydrophobic region, and a tail of charged residues. Sortase can catalyse a two-step transpeptidation reaction either in vivo or in vitro. First, the LPXTG motif is cleaved between threonine and glycine; then the threonine is covalently attached to the amino group of pentaglycine of the cell-wall peptidoglycan; this results in a protein attached to the cell wall.[44] In their experimental set up, using *S. aureus* sortase Mao et al. "borrowed" this strategy from nature and demonstrated the suitability of this enzyme in protein–peptide and protein–protein ligations. Furthermore, they have also shown that non-native peptide fragments including D-peptides and nonpeptide derivatives (e.g., folate) of glycine (mono-, to triglycine) can also be efficiently conjugated by sortase to the acceptor protein (green fluorescent protein containing the LPXTG motif, Scheme 1 D). In this way, the arsenal of tools for EPL is enriched by an additional procedure with great potential to become a part of routine recipes used for the assembly of selectively modified peptides and proteins.

Such, and similar, surprising findings will most probably be further extended in future genome and proteome mapping among different species. In fact, there are enough documented cases in which the discovery of particular biological processes during studies of microbial pathogenicity mechanisms or unique biological phenomena in rather obscure microorganisms, like protein splicing, might serve as a direct input for the development of novel technologies.

At present we are privileged to live the dream of Emil Fischer. But to make it richer and more accessible, we need to look for the strategies that living or-

ganisms have optimized and developed during their evolution. However, the goal should not just be to provide their mere descriptions (which seems to be a somewhat dominant trend nowadays), but rather to put it together with the chemist's conceptual intuition and inventive spirit. If Emil Fischer were still alive, he would certainly have been happy to see how fruitful such a marriage between organic chemistry and biology can be in realizing his dream.

Acknowledgements

Prof. Dr. Luis Moroder is gratefully acknowledged for critical reading and numerous suggestions that considerably increased the quality of this manuscript, and Prof. Dr. Robert Huber is gratefully acknowledged for his encouragement and continuous interest in my work. The author is also indebted to Prajna Paramita Pal for reading and correcting the manuscript.

Keywords: expression · ligation · peptide · protein · synthesis

[1] F. Hofmeister, *Ergeb. Physiol.* **1902**, *1*, 759.
[2] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 21.
[3] B. Gutte, R. B. Merrifield, *J. Am. Chem. Soc.* **1969**, *91*, 501.
[4] R. G. Denkewalter, D. F. Weber, F. W. Holly, R. Hirschmann, *J. Am. Chem. Soc.* **1969**, *91*, 502.
[5] R. B. Merrifield in *Synthesis of Peptides and Peptidomimetics*, *Vol. E22b* (Ed.: M. Goodman), Thieme, Stuttgart, **2003**, pp. 3.
[6] Y. Nishiuchi, T. Inui, H. Nishio, J. Bodi, T. Kimura, F. I. Tsuji, S. Sakakibara, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 13549.
[7] H. F. Gaertner, K. Rose, R. Cotton, D. Timms, R. Camble, R. E. Offord, *Bioconjugate Chem.* **1992**, *3*, 262.
[8] T. Wieland, W. Schafer, *Angew. Chem.* **1951**, *63*, 146.
[9] J. Blake, *Int. J. Pept. Protein Res.* **1981**, *17*, 273.
[10] D. Yamashiro, C. H. Li, *Int. J. Pept. Protein Res.* **1988**, *31*, 322.
[11] M. Schnolzer, S. B. H. Kent, *Science* **1992**, *256*, 221.
[12] P. E. Dawson, T. W. Muir, I. Clarke-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776.
[13] H. Staudinger, J. Meyer, *Helv. Chim. Acta* **1919**, *2*, 635.
[14] M. Köhn, R. Breinbauer, *Angew. Chem.* **2004**, *116*, 3168; *Angew. Chem. Int. Ed.* **2004**, *43*, 3106.
[15] B. L. Nilsson, L. L. Kiessling, R. T. Raines, *Org. Lett.* **2000**, *2*, 2141.
[16] E. Saxon, J. I. Armstrong, C. R. Bertozzi, *Org. Lett.* **2000**, *2*.
[17] G. G. Kochendoerfer, S. Y. Chen, F. Mao, S. Cressman, S. Traviglia, H. Y. Shao, C. L. Hunter, D. W. Low, E. N. Cagle, M. Carnevali, V. Gueriguian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, J. Wilken, J. Tang, J. J. Levy, L. P. Miranda, M. M. Crnogorac, S. Kalbag, P. Botti, J. Schindler-Horvat, L. Savatski, J. W. Adamson, A. Kung, S. B. H. Kent, J. A. Bradburne, *Science* **2003**, *299*, 884.
[18] R. David, M. P. O. Richter, A. G. Beck-Sickinger, *Eur. J. Biochem.* **2004**, *271*, 663.
[19] T. W. Muir, *Annual Rev. Biochem.* **2003**, *72*, 249.
[20] G. Casi, D. Hilvert, *Curr. Opin. Struct. Biol.* **2003**, *13*, 589.
[21] N. Budisa, *Angew. Chem.* **2004**, *116*; *Angew. Chem. Int. Ed.* **2004**, *43*, in press.
[22] J. P. Gogarten, A. G. Senejani, O. Zhaxybayeva, L. Olendzenski, E. Hilario, *Ann. Rev. Microbiol.* **2002**, *56*, 263.
[23] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705.
[24] J. P. Tam, Q. T. Yu, *Biopolymers* **1998**, *46*, 319.
[25] M. D. Gieselman, Y. T. Zhu, H. Zhou, D. Galonic, W. A. van der Donk, *ChemBioChem* **2002**, *3*, 709.
[26] R. J. Hondal, B. L. Nilsson, R. T. Raines, *J. Am. Chem. Soc.* **2001**, *123*, 5140.
[27] J. P. Tam, Q. T. Yu, J. L. Yang, *J. Am. Chem. Soc.* **2001**, *123*, 2487.
[28] G. Roelfes, D. Hilvert, *Angew. Chem.* **2003**, *115*, 2377; *Angew. Chem. Int. Ed.* **2003**, *42*, 2275.
[29] R. S. Roy, O. Allen, C. T. Walsh, *Chem. Biol.* **1999**, *6*, 789.
[30] L. S. Zhang, J. P. Tam, *Tetrahedron Lett.* **1997**, *38*, 3.
[31] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526.
[32] B. L. Nilsson, R. J. Hondal, M. B. Soellner, R. T. Raines, *J. Am. Chem. Soc.* **2003**, *125*, 5268.
[33] K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 19.
[34] M. Bergmann, H. Fraenkel-Conrat, *J. Biol. Chem.* **1938**, *124*, 1.
[35] Z. Machova, R. von Eggelkraut-Gottanka, N. Wehofsky, F. Bordusa, A. G. Beck-Sickinger, *Angew. Chem.* **2003**, *115*, 5065; *Angew. Chem. Int. Ed.* **2003**, *42*, 4916.
[36] T. Nakatsuka, T. Sasaki, E. T. Kaiser, *J. Am. Chem. Soc.* **1987**, *109*, 3808.
[37] D. Y. Jackson, J. Burnier, C. Quan, M. Stanley, J. Tom, J. A. Wells, *Science* **1994**, *266*, 243.
[38] T. K. Chang, D. Y. Jackson, J. P. Burnier, J. A. Wells, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 12544.
[39] A. C. Braisted, J. K. Judice, J. A. Wells, *Methods Enzymol.* **1997**, *289*, 298.
[40] F. Bordusa, *Chem. Rev.* **2002**, *102*, 4817.
[41] K. Rall, F. Bordusa, *J. Org. Chem.* **2002**, *67*, 9103.
[42] H. Y. Mao, S. A. Hart, A. Schink, B. A. Pollok, *J. Am. Chem. Soc.* **2004**, *126*, 2670.
[43] S. K. Mazmanian, G. Liu, T. T. Hung, O. Schneewind, *Science* **1999**, *285*, 760.
[44] X. Y. Huang, A. Aulabaugh, W. D. Ding, B. Kapoor, L. Alksne, K. Tabei, G. Ellestad, *Biochemistry* **2003**, *42*, 11307.
[45] L. Moroder in *Synthesis of Peptides and Peptidomimetics*, *Vol. E22a* (Ed.: M. Goodman), Thieme, Stuttgart, **2003**, pp. 17.
[46] S. K. Mazmanian, I. T. Hung, O. Schneewind, *Mol. Microbiol.* **2001**, *40*, 1049.

Received: May 6, 2004

