

Enzymes in Protein Ligation: The Coupling of Peptides, Peptide Nucleic Acids and Proteins by Sortase A

S. Pritz*

Leibniz-Institute of Molecular Pharmacology, Robert-Roessle-Strasse 10, 13125 Berlin, Germany

Abstract: The chemical synthesis of proteins, as well as the modification of expressed proteins, requires highly selective reactions. Ligation methods offer feasible tools for this purpose. This mini-review focuses on enzymatic methods for protein ligation; in particular, on the use of the transpeptidase sortase.

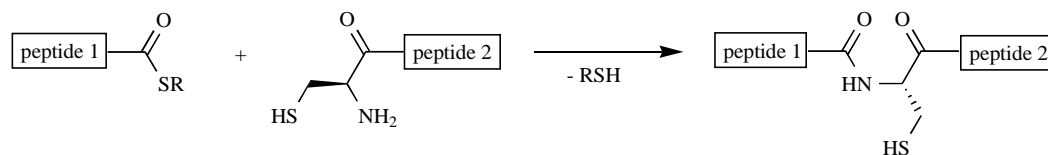
Keywords: Enzymatic ligation, sortase, proteins, peptides, peptide nucleic acids.

1. INTRODUCTION

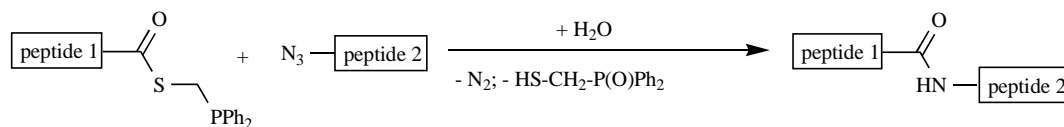
Specific incorporation of new functionalities into proteins (protein engineering) is a widely used tool in protein biochemistry. Chemical ligation allows selective modification of unprotected peptides or proteins, and thereby offers the opportunity for the synthesis of non-natural constructs by site-specific incorporation of non-natural amino acids, stable isotopes, fluorophores and other biochemical or biophysical probes into recombinantly expressed proteins [1]. As ligations can be performed in solution and each intermediate can be purified by chromatographic methods, side products are not carried through the synthesis. Protected peptide fragments are unsuitable, because they are often poorly soluble and correspondingly difficult to purify and, moreover, subsequent deprotection steps often give rise to byproduct formation. Consequently, fully deprotected peptides must be used, which increases their solubility in water. Therefore, many functional groups are present in the educts and ligation

often [3]. This approach is based upon the reaction of a C-terminal peptide thiol ester with an N-terminal cysteine (Scheme 1). The synthesis of a 166 residue polymer-modified erythropoiesis protein has been achieved through NCL of four unprotected peptide segments [4]. However, efficient synthesis of a peptide thiol ester is a prerequisite for NCL, and can be difficult for larger fragments. Therefore, an extension of NCL has been realized by generating the C-terminal thiol ester by thiolysis of the protein of interest from a corresponding recombinantly-expressed protein-intein fusion. The resulting protein thiol ester can be ligated to a peptide or protein with an N-terminal cysteine, similar to NCL. This method has been named expressed protein ligation (EPL) [5] and has been successfully applied to several problems [6].

Another method, the Staudinger ligation, makes use of a C-terminal phosphinothiol ester, which reacts with a peptide containing an N-terminal azide (Scheme 2) [7]. The reaction



Scheme 1. Native chemical ligation (NCL).

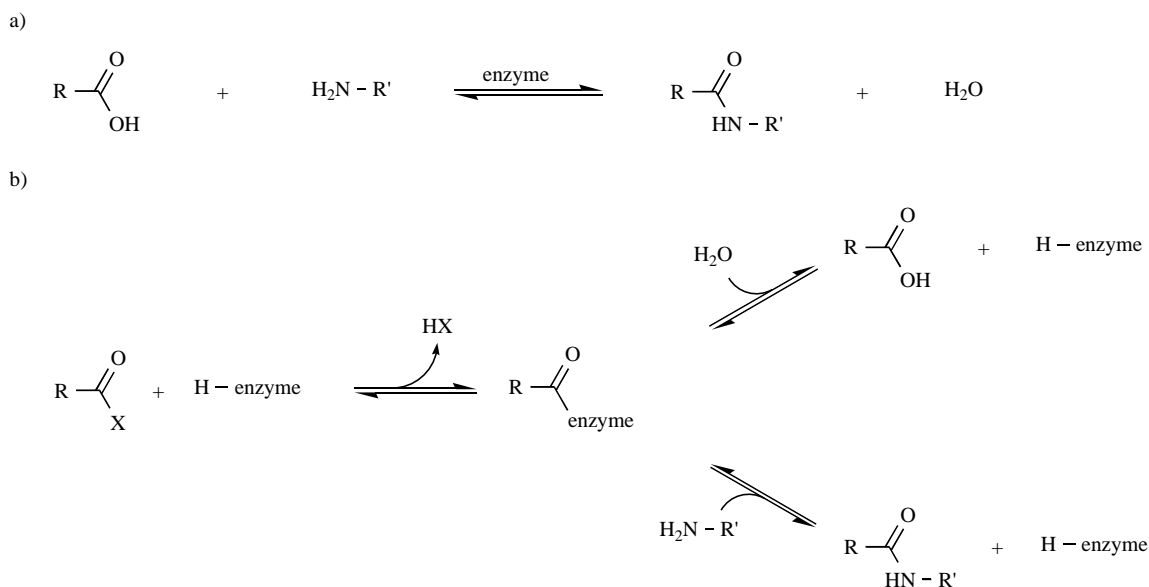


Scheme 2. Staudinger ligation.

methods have to rely upon the chemo- and regioselective reaction of a C-terminal functional group of one peptide or protein with an N-terminal group of another peptide or protein, which ultimately leads to the formation of a native peptide bond [2]. Among the methods developed for this purpose, native chemical ligation (NCL) is the one used most

proceeds *via* an iminophosphorane, which rearranges to an amidophosphonium salt that is hydrolyzed, thus yielding the amide and a phosphine oxide. An advantage of the Staudinger ligation with respect to NCL is that no N-terminal cysteine at the C-terminal fragment is needed. Thus in principle, every amino acid combination Xaa-Yaa at the ligation site should be possible, although it seems that the best results are obtained when using glycine at position Yaa, due to the minimal steric hindrance [8].

*Address Correspondence to this author at the Leibniz-Institute of Molecular Pharmacology, Robert-Roessle-Strasse 10, 13125 Berlin, Germany; E-mail: pritz@fimp-berlin.de



Scheme 3. Comparison of thermodynamically- (a) and kinetically-controlled (b) protease-catalyzed peptide ligation.

All of the ligation methods mentioned above use an excess of thiol for the formation of the thiol ester and are therefore not applicable when dealing with redox-sensitive substrates; e.g., proteins with preformed disulfide bridges.

2. ENZYMATIC LIGATION METHODS

Enzymes are proteins that catalyze a specific chemical reaction under mild conditions, which makes them valuable tools in synthetic organic chemistry. Their high inherent substrate specificity is the reason why a universal enzyme that can be utilized to couple all proteinogenic amino acids with each other has not yet been developed nor exists in nature [9]. However, there exist some proteases that are able to cleave proteins or peptides at specific sites. According to van't Hoff's principle of microscopic reversibility, also the reverse reaction, that is indeed the desired ligation reaction, will be catalyzed. But as the hydrolysis of a peptide bond is exergonic, one has to manipulate reaction conditions to shift the equilibrium in favor of peptide bond formation [9]. Accordingly, to date there are numerous examples of the use of proteases for the synthesis of proteins and peptides as enzymatic ligations may be even more selective than chemical methods [10].

There exist two general approaches to enzymatic peptide synthesis: thermodynamically-controlled and kinetically-controlled methods. The former, for which all types of proteases can be used, is the reverse of the hydrolysis reaction. In the latter, which is limited to serine and cysteine proteases, an acyl-enzyme intermediate is formed that is susceptible to nucleophilic attack, either by water, which leads to hydrolysis, or by the α -amino group of an amino acid, which leads to peptide bond formation (Scheme 3). Two of the main drawbacks of enzymatic synthesis are thus hydrolysis of the acyl donor and additional undesired proteolytic cleavages within the reactants or the products by the enzyme.

According to the mechanism of proteolytic cleavage, a short notation for protease-substrate interactions was pro-

posed [11]. The peptide substrate binds to the protease in a way that the scissile amide bond always occupies a place at the catalytic site. The side chains of the amino acids N-terminal (P_n) and C-terminal (P_n') to the cleavage site point to corresponding distinct binding grooves of the protease (S_n and S_n' , respectively) (Fig. (1)). The specificity of several proteases results mainly from the residue at P_1 position: e.g., trypsin, where basic amino acids are preferred ($P_1 = \text{Lys, Arg}$); α -chymotrypsin, with hydrophobic residues in this position ($P_1 = \text{Phe, Tyr, Trp, Leu}$); and V8 protease, with a preference for acidic residues ($P_1 = \text{Glu, Asp}$).

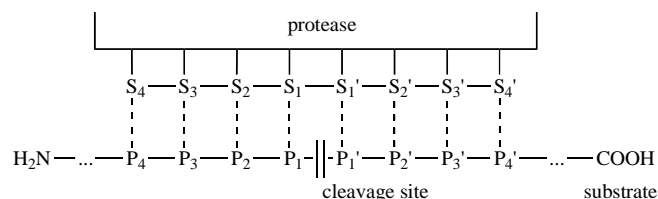
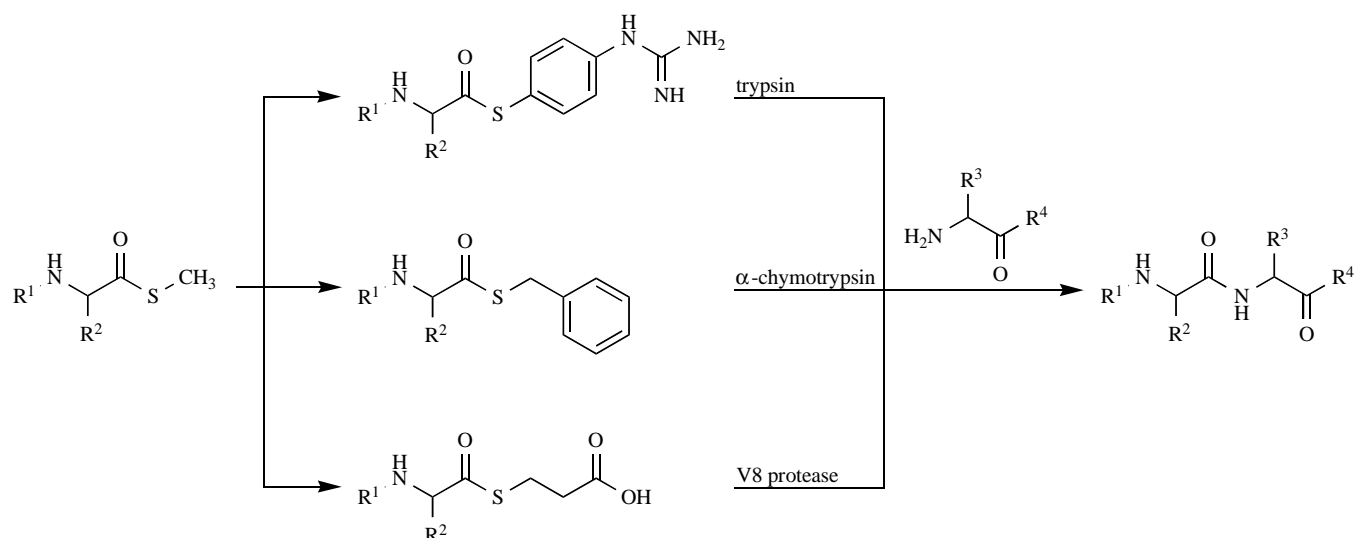


Fig. (1). Subsite nomenclature of proteases and their corresponding substrates [11].

As shown in Scheme 3, slightly activated acyl-donors ($\text{R}-\text{COX}$) are a prerequisite for the kinetically-controlled approach to protease-mediated ligation. Since hydrolysis of esters is faster than that of amides, esterified substrates of the type $P_n\text{-}\dots\text{-}P_1\text{-COOR}$ (with residues at the P_1 position that fulfill the specificity requirements of the enzyme) have been used as acyl donors [12]. This approach is limited by a possible enzymatic hydrolysis of the newly formed product because of the side chain of the residue at the P_1 position, which is crucial for further protease recognition and cleavage, remaining in the product. Additionally, hydrolyses of the ester or the acyl-enzyme may occur.

In substrate mimetics, the side chain responsible for recognition by the enzyme was transferred from the C-terminal P_1 residue to the ester leaving group [13]. This modification of the acyl donor was tolerated by the enzyme, and an acyl-enzyme was formed whereupon the specificity-bearing leav-



Scheme 4. *In situ* generated substrate mimetics and subsequent enzymatic ligation by the corresponding protease [15].

ing group was released. Therefore, after ligation the peptide was no longer recognized by the enzyme at this site and no secondary proteolytic cleavage could occur. By using zymogens, which are catalytically inactive precursors of active enzymes, or genetically engineered proteases, hydrolysis side reactions were reduced and ligation yields increased [14].

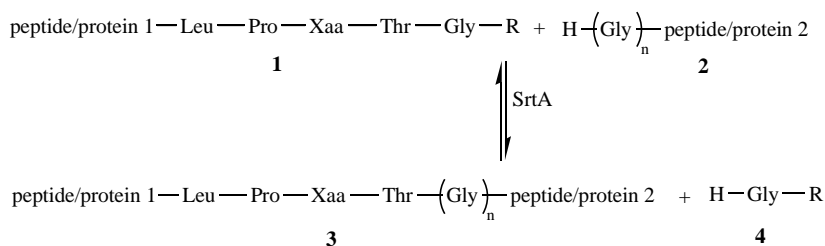
The substrate mimetics approach was further expanded by Bordusa *et al.* by generating different thiol esters *in situ* from a common methyl thiol ester precursor (Scheme 4) [15]. When reacting these acyl donors with the corresponding enzyme and a peptide with free N-terminal amino group, the ligation proceeded with an average yield of 70% to the product (20 residues). This reflects the suitability of the distinct enzymes for this ligation reaction.

A combination of expressed protein ligation (EPL) and protease-mediated ligation was used in expressed enzymatic ligation (EEL) for the synthesis of the 70 residue proNPY [16]. The 41 residue N-terminal part of the protein was expressed recombinantly as a fusion with an intein. Thiolytic cleavage with mercapto-acetic acid gave a thiol ester, which was a suitable substrate for the V8 protease. Incubation with the protease and the fluorescently labeled C-terminal peptide gave the ligation product in 60% yield along with 6% of the self-associated homodimer of the N-terminal peptide and 20% truncated peptide fragments resulting from competitive hydrolysis of the incoming thiol ester.

One of the most successfully applied enzymes so far for protein ligation is subtiligase [17,18], which is a double mu-

tant of the serine protease subtilisin BPN'. The mutation of the catalytically active Ser²²¹ residue to Cys resulted in an improved aminolysis-to-hydrolysis ratio, whereas replacement of Pro²²⁵ by Ala increased the catalytic activity by decreasing steric crowding in the active site caused by the first mutation [19]. By using this enzyme, fully active RNase A (124 residues) was synthesized in milligram quantities by stepwise ligation of six esterified peptide fragments. Average yields of 70% per ligation were achieved, with side products arising from hydrolysis of the peptide esters. Furthermore, non-natural amino acid residues were incorporated into the protein by this method [17]. N-terminal extension of methionyl-extended human growth hormone (Met-hGH) with a biotin-tagged peptide or mercurated cysteine was attained in over 80% yield [18].

Another very interesting enzyme used for protein ligation is sortase A [20]. Sortases are transpeptidases found in Gram-positive bacteria. The *Staphylococcus aureus* sortase isoform SrtA (sortase A) cleaves proteins like **1** at a conserved Leu-Pro-Xaa-Thr-Gly (LPXTG) motif between threonine and glycine and subsequently transfers the acyl-fragment to a peptide bearing an N-terminal oligoglycine (**2**), thus yielding the conjugate **3** and a small leaving group (Scheme 5). The reaction proceeds *via* a threonyl thiol ester intermediate with the active site cysteine. *In vivo*, the enzyme mediates the covalent attachment of virulence- and colonization-associated surface proteins to the amino group of pentaglycine cross-bridges of bacterial cell wall precursors [21]. Kinetic studies revealed a ping-pong bi-bi hydrolytic shunt mechanism for recombinant sortase A [22], which



Scheme 5. Sortase A-mediated transpeptidation reaction ($n \geq 1$).

Table 1. Synthesized CPP-PNA Conjugates by Sortase-Mediated Ligation [29]

Entry	Name	Sequence ^a	Coupling Yield ^b
1	KLA-PNA	H-KLAL KLAL KAL KAAL KLA LPKTGGG ooo cct ctt acc tca gtt aca ooo-NH ₂	47%
2	ELA-PNA	H-ELAL ELAL EAL EAAL ELA LPKTGGG ooo cct ctt acc tca gtt aca ooo-NH ₂	56%
3	KAL-PNA	H-KALK LKAA LAL LAKL KLA LPKTGGG ooo cct ctt acc tca gtt aca ooo-NH ₂	57%
4	KGL-PNA	H-KGLK LKGG LGL LGKL KLG LPKTGGG ooo cct ctt acc tca gtt aca ooo-NH ₂	54%
5	RLA-PNA	H-RLAL RLAL RAL RAAL RLA LPKTGGG ooo cct ctt acc tca gtt aca ooo-NH ₂	49%
6	Pen-PNA	H-RQI KIW FQN RRM KWKK LPKTGGG ooo cct ctt acc tca gtt aca ooo-NH ₂	80%
7	KLA-PNA scr	H-KLAL KLAL KAL KAAL KLA LPKTGGG ooo tec ttc cca act ttg aca ooo-NH ₂	57%
8	PNA-KLA	Ac-ooo cct ctt acc tca gtt aca ooo LPKTGGG KLAL KLAL KAL KAAL KLA-NH ₂	94%
9	PNA-ELA	Ac-ooo cct ctt acc tca gtt aca ooo LPKTGGG ELAL ELAL EAL EAAL ELA-NH ₂	87%
10	PNA-Pen	Ac-ooo cct ctt acc tca gtt aca ooo LPKTGGG RQI KIW FQN RRM KWKK-NH ₂	93%
11	PNA scr-KLA	Ac-ooo tec ttc cca act ttg aca ooo LPKTGGG KLAL KLAL KAL KAAL KLA-NH ₂	91%
12	FAM-PNA-KLA	FAM-ooo cct ctt acc tca gtt aca ooo LPKTGGG KLAL KLAL KAL KAAL KLA-NH ₂ ^c	85%
13	FAM-PNA-ELA	FAM-ooo cct ctt acc tca gtt aca ooo LPKTGGG ELAL ELAL EAL EAAL ELA-NH ₂ ^c	54%
14	FAM-PNA-Pen	FAM-ooo cct ctt acc tca gtt aca ooo LPKTGGG RQI KIW FQN RRM KWKK-NH ₂ ^c	92%

^aAll conjugates showed satisfactory MS data and purities >95%.

^bAs determined by HPLC.

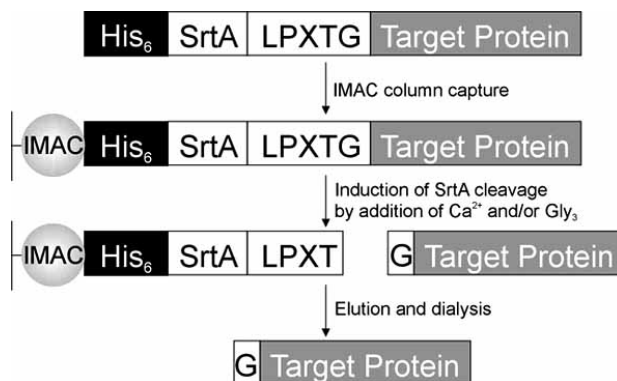
^cFAM = 5-carboxyfluorescein.

other protein, or (d) the dyes fluorescein (FITC) or tetramethylrhodamine (TAMRA), which allowed fluorescent visualization. In case (b), the biotin label was released from the protein upon irradiation with UV light. Additionally, the chemokine CXCL14 was labeled with the TAMRA-containing peptide in solution, and it was shown that in crude cell lysates only the proteins equipped with an LPETG-motif were tagged. Furthermore, sortase A was able to selectively label surface proteins expressed in human embryonic kidney (HEK) 293T cells when added to the medium along with the respective probe.

Apart from ligation methods, an interesting technique for one-step purification of free recombinant proteins has been developed [38]. For this purpose, the protein of interest was C-terminally fused to sortase A bearing an N-terminal His₆ tag and a C-terminal LPXTG-motif. The fusion protein was expressed in *E. coli* and bound to an immobilized metal-ion affinity chromatography (IMAC) column. Activation of sortase by addition of calcium ions and/or triglycine liberated the target protein with an additional N-terminal glycine residue (Scheme 7).

3. CONCLUSION

Enzymatic methods for protein and peptide ligation offer several advantages, such as the prevention of racemization, the lack of a requirement for protection/deprotection procedures of side chain functionalities, a reduced use of problematic (e.g., expensive or potentially toxic) solvents and reagents, and even the possibility of recovering the catalyst. Moreover, as enzymatic ligations exert other restrictions concerning the amino acid residues flanking the ligation site, they can be used orthogonally to chemical ligation methods.



Scheme 7. One-step purification of sortase-fusion proteins [38].

Protease-catalyzed ligation is a state-of-the-art method for the coupling of unprotected peptide segments. By using substrate mimetics, the reaction can be carried out almost independently from the residue at the ligation site.

Due to the low tolerance of sortase A for deviation in the LPXTG-recognition motif, this enzymatic ligation is highly selective. In particular, the very limited occurrence of this motif in proteins makes the SrtA-mediated reaction interesting for protein modification, even though, due to the mechanism, the recognition motif has to remain in the product. A substantial advantage of sortase-mediated ligation over protease-catalyzed ligation is that peptide or protein substrates containing a C-terminal LPXTG-motif can be easily synthesized, either by chemical or recombinant methods, and no further laborious esterification steps are necessary. Additionally, there seems to be no risk of hydrolysis of the products by sortase during the reaction.

ACKNOWLEDGEMENTS

I thank Michael Bienert and Irene Coin for helpful discussions and critical reading of the manuscript. This work was supported by the European Commission (QLK3-CT-2002-01989).

REFERENCES

- [1] Rauh, D.; Waldmann, H. *Angew. Chem. Int. Ed. Engl.*, **2007**, *46*, 826.
- [2] Bode, J.W. *Curr. Opin. Drug Discov. Dev.*, **2006**, *9*, 765.
- [3] (a) Dawson, P.E.; Muir, T.W.; Clark-Lewis, I.; Kent, S.B.H. *Science*, **1994**, *266*, 776. (b) Tam, J.P.; Xu, J.; Eom, K.D. *Biopolymers*, **2001**, *60*, 194.
- [4] 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.H.; Bradburne, J.A. *Science*, **2003**, *299*, 884.
- [5] Muir, T.W.; Sondhi, D.; Cole, P.A. *Proc. Natl. Acad. Sci. U.S.A.*, **1998**, *95*, 6705.
- [6] Muir, T.W. *Annu. Rev. Biochem.*, **2003**, *72*, 249.
- [7] (a) Saxon, E.; Armstrong, J.I.; Bertozzi, C.R. *Org. Lett.*, **2000**, *2*, 2141. (b) Nilsson, B.L.; Hondal, R.J.; Soellner, M.B.; Raines, R.T. *J. Am. Chem. Soc.*, **2003**, *125*, 5268.
- [8] Merckx, R.; Rijkers, D.T.S.; Kemmink, J.; Liskamp, R.M.J. *Tetrahedron Lett.*, **2003**, *44*, 4515.
- [9] Sewald, N.; Jakubke, H.-D., In *Peptides: Chemistry and Biology*, WILEY-VCH: Weinheim, **2002**; pp. 247-267.
- [10] Recent reviews: (a) Bordusa, F. *Chem. Rev.*, **2002**, *102*, 4817. (b) Lombard, C.; Saulnier, J.; Wallach, J.M. *Protein Pept. Lett.*, **2005**, *12*, 621.
- [11] (a) Schechter, I.; Berger, A. *Biochem. Biophys. Res. Commun.*, **1967**, *27*, 157. (b) Schechter, I.; Berger, A. *Biochem. Biophys. Res. Commun.*, **1968**, *32*, 898.
- [12] Bordusa, F.; Ullmann, D.; Jakubke, H.D. *Angew. Chem. Int. Ed. Engl.*, **1997**, *36*, 1099.
- [13] Bordusa, F.; Ullmann, D.; Elsner, C.; Jakubke, H.D. *Angew. Chem. Int. Ed. Engl.*, **1997**, *36*, 2473.
- [14] Rall, K.; Bordusa, F. *J. Org. Chem.*, **2002**, *67*, 9103.
- [15] Wehofsky, N.; Koglin, N.; Thust, S.; Bordusa, F. *J. Am. Chem. Soc.*, **2003**, *125*, 6126.
- [16] Machova, Z.; Eggelkraut-Gottanka, R.; Wehofsky, N.; Bordusa, F.; Beck-Sickinger, A.G. *Angew. Chem. Int. Ed. Engl.*, **2003**, *42*, 4916.
- [17] Jackson, D.Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J.A. *Science*, **1994**, *266*, 243.
- [18] Chang, T.K.; Jackson, D.Y.; Burnier, J.P.; Wells, J.A. *Proc. Natl. Acad. Sci. U.S.A.*, **1994**, *91*, 12544.
- [19] Abrahmsén, L.; Tom, J.; Burnier, J.; Butcher, K.A.; Kossiakoff, A.; Wells, J.A. *Biochemistry*, **1991**, *30*, 4151.
- [20] Marraffini, L.A.; De Dent, A.C.; Schneewind, O. *Microbiol. Mol. Biol. Rev.*, **2006**, *70*, 192.
- [21] Mazmanian, S.K.; Liu, G.; Ton-That, H.; Schneewind, O. *Science*, **1999**, *285*, 760.
- [22] Frankel, B.A.; Kruger, R.G.; Robinson, D.E.; Kelleher, N.L.; McCafferty, D.G. *Biochemistry*, **2005**, *44*, 11188.
- [23] Huang, X.; Aulabaugh, A.; Ding, W.; Kapoor, B.; Alksne, L.; Tabei, K.; Ellestad, G. *Biochemistry*, **2003**, *42*, 11307.
- [24] Ilangovan, U.; Ton-That, H.; Iwahara, J.; Schneewind, O.; Clubb, R.T. *Proc. Natl. Acad. Sci. U.S.A.*, **2001**, *98*, 6056.
- [25] Zong, Y.N.; Bice, T.W.; Ton-That, H.; Schneewind, O.; Narayana, S.V.L. *J. Biol. Chem.*, **2004**, *279*, 31383.
- [26] Kruger, R.G.; Otvos, B.; Frankel, B.A.; Bentley, M.; Dostal, P.; McCafferty, D.G. *Biochemistry*, **2004**, *43*, 1541.
- [27] Pritz, S.; Wolf, Y.; Kraetke, O.; Klose, J.; Bienert, M.; Beyermann, M. *J. Org. Chem.*, **2007**, *72*, 3909.
- [28] Mao, H.Y.; Hart, S.-A.; Schink, A.; Pollok, B.A. *J. Am. Chem. Soc.*, **2004**, *126*, 2670.
- [29] Wolf, Y.; Pritz, S.; Abes, S.; Bienert, M.; Lebleu, B.; Oehlke, J. *Biochemistry*, **2006**, *45*, 14944.
- [30] (a) Snyder, E.L.; Dowdy, S.F. *Pharm. Res.*, **2004**, *21*, 389. (b) Zorko, M.; Langel, Ü. *Adv. Drug Deliv. Rev.*, **2005**, *57*, 529. (c) Langel, Ü. (Ed.), *Handbook of Cell-Penetrating Peptides*, 2nd ed., CRC Press: Boca Raton, **2006**.
- [31] Nielsen, P.E.; Egholm, M.; Berg, R.H.; Buchardt, O. *Science*, **1991**, *254*, 1497.
- [32] Ganesh, K.N.; Nielsen, P.E. *Curr. Org. Chem.*, **2000**, *4*, 931.
- [33] Oehlke, J.; Scheller, A.; Wiesner, B.; Krause, E.; Beyermann, M.; Klauschenz, E.; Melzig, M.; Bienert, M. *Biochim. Biophys. Acta*, **1998**, *1414*, 127.
- [34] (a) Casale, R.; Jensen, I.S.; Egholm, M. In *Peptide Nucleic Acids: Protocols and Applications*; Nielsen, P.E.; Egholm, M., Eds.; Horizon Scientific Press: Norfolk, **1999**; pp. 39-50. (b) Pritz, S.; Wolf, Y.; Klemm, C.; Bienert, M. *Tetrahedron Lett.*, **2006**, *47*, 5893.
- [35] Coin, I.; Beyermann, M.; Bienert, M. *Nat. Protoc.*, **2007**, *2*, 3247.
- [36] Parthasarathy, R.; Subramanian, S.; Boder, E.T. *Bioconjug. Chem.*, **2007**, *18*, 469.
- [37] Popp, M.W.; Antos, J.M.; Grotenbreg, G.M.; Spooner, E.; Ploegh, H.L. *Nat. Chem. Biol.*, **2007**, *3*, 707.
- [38] Mao, H.Y. *Protein Exp. Purif.*, **2004**, *37*, 253.