Angewandte Chemie International Edition

Total Synthesis

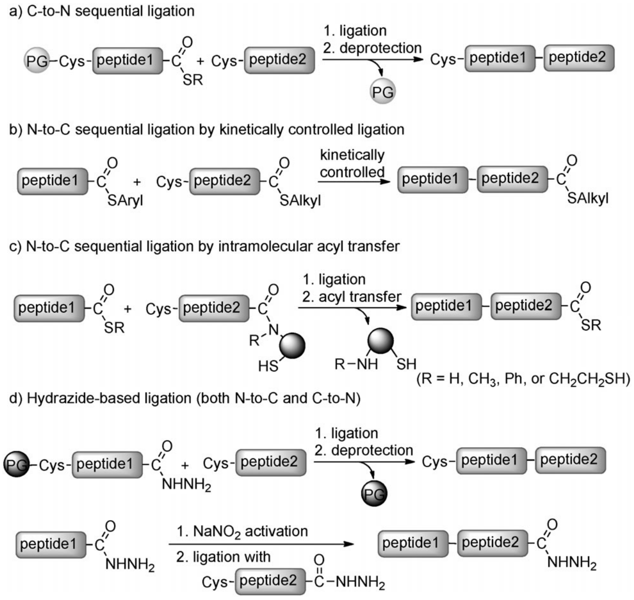
DOI: 10.1002/anie.201203843

Convergent Chemical Synthesis of Proteins by Ligation of Peptide Hydrazides**

*Ge-Min Fang, Jia-Xing Wang, and Lei Liu**

Total chemical synthesis of proteins can exert atomically precise control on protein composition, and thereby offer the potential of solving a multitude of problems in biomedicine.[1] A common strategy for protein chemical synthesis involves the preparation of peptide segments on solid phase and subsequent condensation of these segments in solution. At present, the most successful approach for the condensation of peptide segments is native chemical ligation, which was invented by Kent et al.[2] In this method, a C-terminal peptide thioester reacts chemoselectively with an N-terminal Cys–peptide under mild reaction conditions in aqueous solution to generate a new peptide bond. The use of native chemical ligation for the preparation of proteins from just two peptide segments is straightforward.[3] For syntheses involving multiple peptide segments, the intrinsic dual reactivity of a bifunctional Cys–peptide–thioester must be controlled to prevent side reactions from occurring at the wrong sites.[4]

As an N-terminal Cys can be reversibly protected, the C-to-N sequential ligation of peptide segments can be readily accomplished by using the PG–Cys–peptide–thioester (Scheme 1 a).[5] In contrast, the N-to-C sequential assembly of peptide segments is more challenging because of the difficulty in protecting a thioester. To overcome this problem, Kent et al. developed a kinetically controlled ligation which makes ingenious use of the reactivity difference between aryl and alkyl thioesters (Scheme 1 b).[6] The efficacy of this strategy has been shown by the convergent syntheses of several proteins consisting of up to six peptide segments.[7] Furthermore, several groups,[8,9] including ours,[10] developed ester or amide functionalities which can be converted into thioesters through an intramolecular acyl transfer (Scheme 1 c). In particular, the very recently developed approaches of the acyl transfer of secondary amides allow the synthesis of proteins by N-to-C assembly of multiple segments.[9] By using this strategy, Brik et al. recently accomplished the convergent synthesis of tetraubiquitin using a thioester equivalent equipped with protected *N*-methylcysteine.[11]



Scheme 1. Available methods for sequential peptide ligation. PG = protecting group.

New methods in the arsenal of sequential peptide ligation may further enhance the efficiency and flexibility of convergent protein synthesis. In this regard, we seek to achieve convergent protein synthesis which solely relies on the ligation of peptide hydrazides (Scheme 1 d). As shown previously,[12] an unprotected peptide hydrazide can be readily converted into a thioester by $NaNO_2$ and an external thiol. This thioester can undergo native chemical ligation with a Cys–peptide in good yield. An important advantage of this ligation approach is that peptide hydrazides can be readily prepared at low cost by using either Boc (*tert*-butoxycarbonyl) or Fmoc (9-fluorenylmethoxycarbonyl) solid-phase peptide synthesis. The possibility of conducting both N-to-C and C-to-N sequential ligation with peptide hydrazides also promises flexibility in the synthetic design. Nonetheless, during the development of a hydrazide-based method for convergent protein synthesis we encountered several interesting problems. The solutions to these problems, as reported herein, constitute the basis for a new and general method for convergent protein synthesis.

To present concrete examples of the chemical challenges involved in the fully convergent synthesis of a protein by the ligation of peptide hydrazides, we undertook a synthesis of a 142 residue protein, that is, the ribosomal protein S25 (RpS25).[13] This target represents a good model system for medium-sized proteins. In our synthetic design, the target polypeptide chain was divided into halves, with each half

---

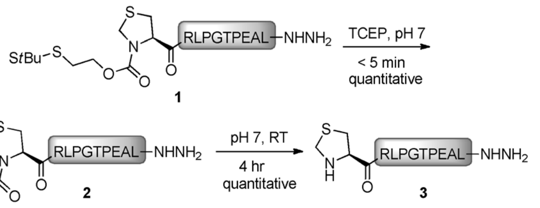
[*] Dr. G.-M. Fang, J.-X. Wang, Prof. Dr. L. Liu
Tsinghua-Peking Center for Life Sciences, Department of Chemistry
Tsinghua University, Beijing 100084 (China)
E-mail: lliu@mail.tsinghua.edu.cn
Dr. G.-M. Fang
High Magnetic Field Laboratory, Chinese Academy of Sciences
Hefei, Anhui 230031 (China)

[**] This work was supported by the National "973" grants from the Ministry of Science and Technology (No. 2011CB965300) and NSFC (Nos. 20932006 and 91013007).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201203843.

containing three peptide segments. The right half of the target sequence was proposed to be made by sequential C-to-N ligations through the use of L-thiazolidine-4-carboxylic acid (Thz) as a protected N-terminal Cys.[5] Unfortunately, our test showed that Thz is unstable to the reaction conditions used for the activation of peptide hydrazides by $NaNO_2$ at pH 3–4. Thus Cys needs to be masked in another way for the sequential C-to-N ligations of peptide hydrazides. Although several options are conceivable for solving the problem (e.g., use of photoactive protecting group), we preferred to develop a method that would be robust and easy to carry out.

After several attempts we found that by adding the 2-(*tert*-butyldisulfanyl)ethyloxycarbonyl (Tbeoc) group[14] to Thz (Scheme 2), the C-to-N sequential ligation of peptide hydrazides can be readily accomplished. The Tbeoc group protects the amino group of Thz and thus prevents the cleavage of thiazolidine during hydrazide activation. The detailed process is as follows: First, Tbeoc–Thz–OH (which can be easily made in three steps) is used to replace Fmoc–Thz–OH in the solid-phase synthesis of Tbeoc–Thz–peptide1–$NHNH_2$. Second, Tbeoc–Thz–peptide1–$NHNH_2$ is activated by $NaNO_2$ at pH 3–4 and then ligated with Cys-peptide2 at pH 7 in the presence of MPAA (4-mercaptophenylacetic acid) as an external thiol and TCEP (tris(2-carboxyethyl)phosphine) as a reducing agent. In this process, Tbeoc–Thz is cleanly converted into Thz so that the ligation product is Thz–peptide1–peptide2. Finally, Thz–peptide1–peptide2 can be readily converted into Cys–peptide1–peptide2 by treatment with 0.4 M methoxyamine.[5]



Scheme 2. Facile conversion of Tbeoc–Thz–peptide into Thz–peptide under reaction conditions which are compatible with native chemical ligation.

Note that in the model system Tbeoc-Thz-Arg-Leu-Pro-Gly-Thr-Pro-Glu-Ala-Leu-$NHNH_2$ (**1**), our experimental analysis shows that treatment of **1** with TCEP at pH 7 causes instant cleavage of the disulfide bond in Tbeoc (Scheme 2). The intermediate **2** is then completely converted into Thz-Arg-Leu-Pro-Gly-Thr-Pro-Glu-Ala-Leu-$NHNH_2$ (**3**) at room temperature at pH 7 within 4 hours. Because the ligation of peptides usually requires a reaction time longer than 4 hours, the conversion of Tbeoc-Thz into Thz can be accomplished during the process of ligation. Thus the use of Tbeoc solves the problem of the C-to-N sequential ligation of peptide hydrazides without introducing an additional step, as compared to the previous method for peptide thioesters (i.e., use of Thz as a masked N-terminal Cys).

By using the Tbeoc method we carried out the synthesis of the right half of the RpS25 protein (Figure 1). Three peptide segments, that is, Tbeoc–Thz–[Val71-Gly90]–$NHNH_2$ (**7**), Tbeoc–Thz–[Leu92-Ala124]–$NHNH_2$ (**8**), and [Cys125-Arg142] (**9**) were prepared by using the hydrazine–Trt(2-Cl) resin[15] and Fmoc chemistry. Note that Val124 was mutated to Ala124, because the yield for the ligation at the Val124 residue was very low. With **7**, **8**, and **9** in hand, the C-to-N sequential ligation was conducted in the following sequence: First, **8** was activated by $NaNO_2$ at pH 3 and −10 °C for 20 minutes. The reaction mixture was then mixed with **9** dissolved in a pH 7 buffer containing 6.0 M guanidinium chloride and excess MPAA for 12 hours. The ligation product was treated with methoxyamine (0.4 M) at pH 4 to remove Thz, thus generating **12** in 60 % yield upon isolation. Second, **7** was activated by $NaNO_2$ and then reacted with **12**. After removal of Thz we obtained **13** in 40 % yield upon isolation. It is worth mentioning that the relatively low yield of the second step may be attributed to the low solubility of the peptide **7**. In fact, when **7** was ligated with **12** in aqueous guanidinium buffer we observed significant precipitation and obtained a much lower yield of **13** ($<10\%$). This solubility problem was partly overcome by the addition of hexafluoroisopropanol and acetonitrile into the ligation media.

For the left half of the RpS25 protein, we initially proposed to synthesize it by sequential N-to-C ligations. This strategy takes advantage of the fact that without activation by $NaNO_2$, a Cys–peptide–$NHNH_2$ peptide can only ligate at its Cys end. However, in the application of this strategy we encountered a problem in the formation of thiolactones (Scheme 3). As already shown in some previous studies, these thiolactones are intramolecular thioesters and their stability is sequence dependent.[16] In our synthesis the thiolactone intermediates were produced in significant yields. To further complicate the situation, these thiolactones were very sluggish to participate in the ligation with a Cys–peptide. The outcome is that the efficiency of the sequential N-to-C ligations is fairly low. One possible solution to overcome this problem is to use a thiol protecting group such as Acm (acetamidomethyl). In the present study we overcame the problem by synthesizing the left half through sequential C-to-N ligations as shown in Figure 1. In this way the internal Cys residues are protected as Tbeoc–Thz, thus avoiding the problem of thiolactone formation. Three peptide segments, that is, [Met1-Trp24]–$NHNH_2$ (**4**), Tbeoc–Thz–[Asp26-His47]–$NHNH_2$ (**5**), and [Cys48-Lys69]–$NHNH_2$ (**6**) were prepared by using the hydrazine–Trt(2-Cl) resin and Fmoc chemistry. Again, Tbeoc–Thz was used as the masked N-terminal Cys.

The C-to-N sequential ligation for the left half was conducted in the following sequence: First, **5** was activated by $NaNO_2$ at pH 3 and −10 °C for 20 minutes. This reaction mixture was then mixed with **6** dissolved in a pH 7 buffer containing 6.0 M guanidinium chloride and excess MPAA for 12 hours. The ligation product was then treated with methoxyamine (0.4 M) at pH 4 to remove Thz, thus generating the intermediate **10**, that is, [Cys25-Lys69]–$NHNH_2$, in 60 % yield upon isolation. Second, **4** was activated by $NaNO_2$ and then reacted with **10** to generate the ligation product in 90 % yield upon isolation. By using the free radical desulfurization conditions (*t*BuSH, VA-044, TCEP (tris(2-carboxyethyl) phosphane hydrochloride)),[17] we were able to desulfurize

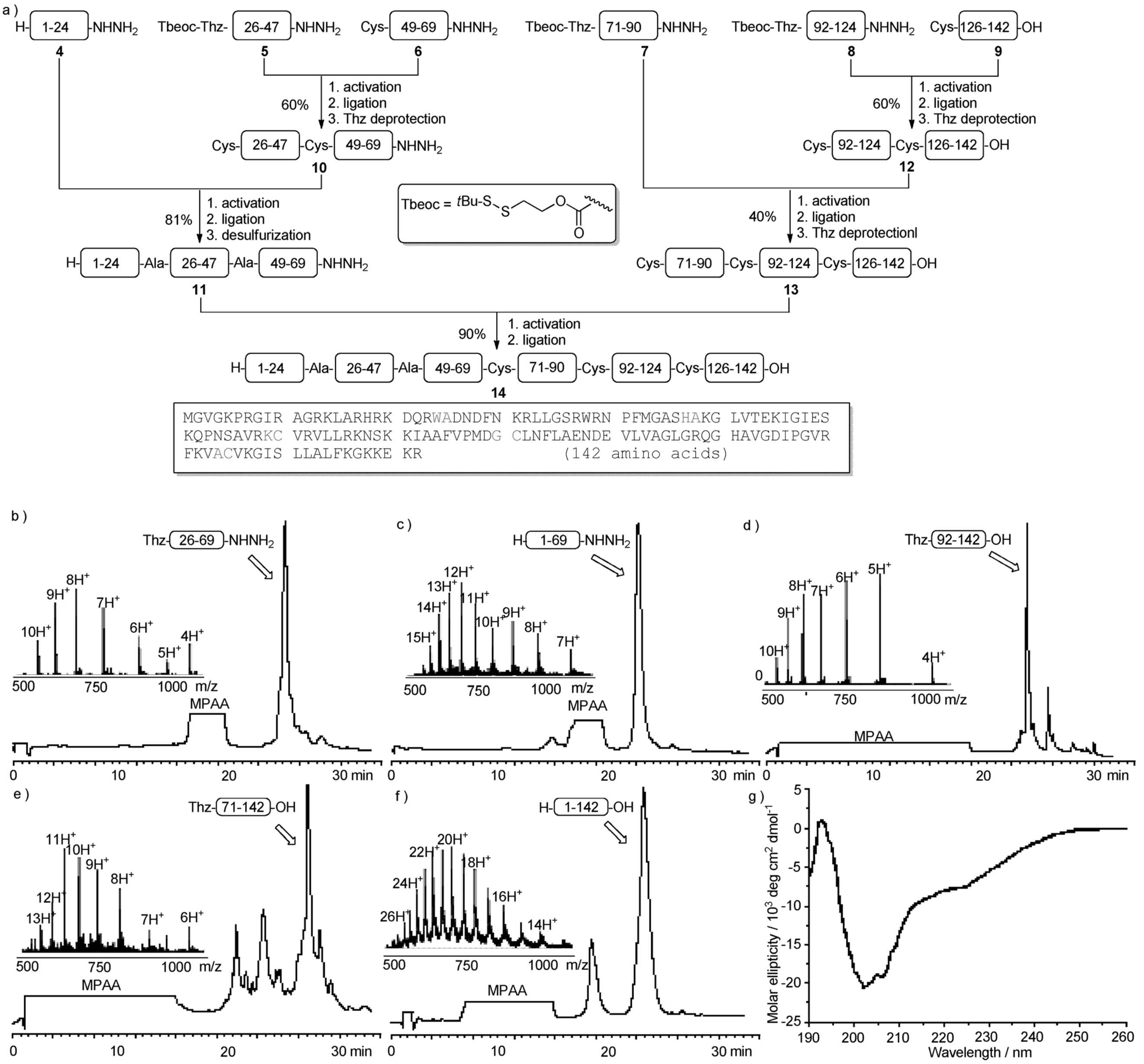
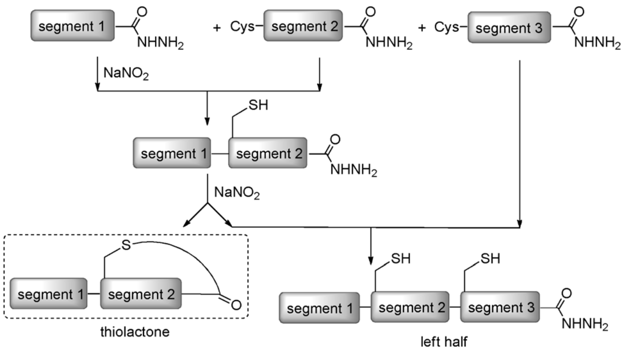


Figure 1. Convergent synthesis of the RpS25 protein. a) Synthetic route and the amino acid sequence of RpS25. b) Ligation between **5** and **6**. The observed mass for the product is 5116.0 Da (calcd: 5115.8 Da). c) Ligation between **4** and **10**. The observed mass for the product is 7917.6 Da (calcd: 7916.1 Da). d) Ligation between **8** and **9**. The observed mass for the product is 5578.5 Da (calcd: 5578.5 Da). e) Ligation between **7** and **12**. The observed mass for the product is 7906.8 Da (calcd: 7906.4 Da). f) Ligation between **11** and **13**. The observed mass for the product is 15716.5 Da (calcd: 15714.7 Da). g) CD analysis of synthetic RpS25.



Scheme 3. Formation of thiolactone in sequential N-to-C ligations.

Cys25 and Cys48, within 12 hours, thereby leaving the hydrazide unit completely intact. This desulfurization step converts Cys25 and Cys48 into Ala25 and Ala48 residues, respectively, which belong to the native sequence of RpS25. The yield of the isolated product in this step is 90% and its product is [Met1-Lys69]–$NHNH_2$ (**11**).

The final step of the synthesis is the ligation between **11** and **13**. We initially tested the one-pot protocol in which **11** was first activated by $NaNO_2$ at pH 3 and −10 °C for 20 minutes, and then treated with **13** at pH 7 with excess MPAA. This one-pot protocol has been successfully used in all the above ligations, but the control of pH value with very small amounts of reactants is difficult. To overcome this problem we can use a two-step protocol: First, **11** was activated by $NaNO_2$ at pH 3 and −10 °C for 20 minutes and then treated with excess MESNa (sodium 2-mercaptoethanesulfonate) in a pH 7 buffer. This step produces a thioester (i.e., [Met1-Lys69]–MESNa) which can be readily isolated by reverse-phase HPLC. Second, the thioester was condensed with **13** under the standard native chemical ligation conditions. This step produces the full length RpS25 polypeptide in about 90 % yield. The ESI/MS spectrum of the purified RpS25 polypeptide is shown in Figure 1 f (mass: obsd: 15 716.5 Da; calcd: 15 714.7 Da, average isotopes). This polypeptide was dissolved in 100 mM Tris-HCl buffer (pH 8.5) containing 6.0 M guanidinium chloride. The solution was then dialyzed overnight into the folding buffer (1.0 M guanidinium chloride, 100 mM Tris-HCl, 8 mM glutathione, pH 8.5). Finally, the folded protein was dissolved in a 10 mM Tris-HCl buffer and characterized by CD spectroscopy (Figure 1 g).

In summary, we have developed a generally applicable strategy for the convergent chemical synthesis of proteins from multiple peptide segments. A key feature of the new strategy is the use of peptide hydrazides instead of thioesters as the ligation building blocks, which not only makes easier the synthesis of peptide segments through mild Fmoc chemistry, but also improves the flexibility in the design of a fully convergent route. The commonly used Thz was found to be incompatible with the reaction conditions for the ligation of peptide hydrazides. Accordingly, we developed an alternative residue (Tbeoc–Thz) as the masked N-terminal Cys. To test the efficacy of the new strategy, we applied it to the synthesis of the 142 residue model protein RpS25 using six peptide segments. It was found that all the ligations can be readily carried out with good yields in a fully convergent fashion. Thus we expect that the new convergent synthetic route to provide a versatile approach for the fast preparation of proteins from multiple peptide segments.

Received: May 18, 2012
Revised: July 31, 2012

Keywords: peptides · protecting groups · proteins · synthesis design · total synthesis

---

[1] P. E. Dawson, S. B. H. Kent, *Annu. Rev. Biochem.* **2000**, *69*, 923.
[2] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776.
[3] S. B. H. Kent, *Chem. Soc. Rev.* **2009**, *38*, 338.
[4] a) N. A. McGrath, R. T. Raines, *Acc. Chem. Res.* **2011**, *44*, 752; b) C. P. R. Hackenberger, D. Schwarzer, *Angew. Chem.* **2008**, *120*, 10182; *Angew. Chem. Int. Ed.* **2008**, *47*, 10030; c) C. Haase, O. Seitz, *Angew. Chem.* **2008**, *120*, 1575; *Angew. Chem. Int. Ed.* **2008**, *47*, 1553; d) J. Y. Lee, D. Bang, *Biopolymers* **2010**, *94*, 441; e) V. R. Pattabiraman, J. W. Bode, *Nature* **2011**, *480*, 471.
[5] a) D. Bang, S. B. H. Kent, *Angew. Chem.* **2004**, *116*, 2588; *Angew. Chem. Int. Ed.* **2004**, *43*, 2534; b) D. J. Boerema, V. A. Tereshko, S. B. H. Kent, *Biopolymers* **2008**, *90*, 278; c) Y. Sohma, B. L. Pentelute, J. Whittaker, Q.-X. Hua, L. J. Whittaker, M. A. Weiss, S. B. H. Kent, *Angew. Chem.* **2008**, *120*, 1118; *Angew. Chem. Int. Ed.* **2008**, *47*, 1102; d) B. L. Pentelute, Z. P. Gates, J. L. Dashnau, J. M. Vanderkooi, S. B. H. Kent, *J. Am. Chem. Soc.* **2008**, *130*, 9702; e) N. Yamamoto, Y. Tanabe, R. Okamoto, P. E. Dawson, Y. Kajihara, *J. Am. Chem. Soc.* **2008**, *130*, 501; f) K. Mandal, S. B. H. Kent, *Angew. Chem.* **2011**, *123*, 8179; *Angew. Chem. Int. Ed.* **2011**, *50*, 8029; g) P. Nagorny, N. Sane, B. Fasching, B. Aussedat, S. J. Danishefsky, *Angew. Chem.* **2012**, *124*, 999; *Angew. Chem. Int. Ed.* **2012**, *51*, 975.
[6] D. Bang, B. L. Pentelute, S. B. H. Kent, *Angew. Chem.* **2006**, *118*, 4089; *Angew. Chem. Int. Ed.* **2006**, *45*, 3985.
[7] a) V. Y. Torbeev, S. B. H. Kent, *Angew. Chem.* **2007**, *119*, 1697; *Angew. Chem. Int. Ed.* **2007**, *46*, 1667; b) S. Liu, B. L. Pentelute, S. B. H. Kent, *Angew. Chem.* **2012**, *124*, 1017; *Angew. Chem. Int. Ed.* **2012**, *51*, 993; c) M. Murakami, R. Okamoto, M. Izumi, Y. Kajihara, *Angew. Chem.* **2012**, *124*, 3627; *Angew. Chem. Int. Ed.* **2012**, *51*, 3567; d) R. Okamoto, K. Morooka, Y. Kajihara, *Angew. Chem.* **2012**, *124*, 195; *Angew. Chem. Int. Ed.* **2012**, *51*, 191.
[8] a) J. D. Warren, J. S. Miller, S. J. Keding, S. J. Danishefsky, *J. Am. Chem. Soc.* **2009**, *131*, 5438; b) P. Botti, M. Villain, S. Manganiello, H. Gaertner, *Org. Lett.* **2004**, *6*, 4861; c) T. Kawakami, M. Sumida, K. Nakamura, T. Vorherr, S. Aimoto, *Tetrahedron Lett.* **2005**, *46*, 8805; d) F. Nagaike, Y. Onuma, C. Kanazawa, H. Hojo, A. Ueki, Y. Nakahara, *Org. Lett.* **2006**, *8*, 4465; e) Y. Ohta, S. Itoh, A. Shigenaga, S. Shintaku, N. Fujii, A. Otaka, *Org. Lett.* **2006**, *8*, 467; f) T. Kawakami, S. Aimoto, *Tetrahedron* **2009**, *65*, 3871; g) J. Kang, J. P. Richardson, D. Macmillan, *Chem. Commun.* **2009**, 407; h) R. K. Sharma, J. P. Tam, *Org. Lett.* **2011**, *13*, 5176.
[9] a) N. Ollivier, J. Dheur, R. Mhidia, A. Blanpain, O. Melnyk, *Org. Lett.* **2010**, *12*, 5238; b) K. Sato, A. Shigenaga, K. Tsuji, S. Tsuda, Y. Sumikawa, K. Sakamoto, A. Otaka, *ChemBioChem* **2011**, *12*, 1840; c) W. Hou, X. Zhang, F, Li, C. F. Liu, *Org. Lett.* **2011**, *13*, 386; d) N. Ollivier, J. Vicogne, A. Vallin, H. Drobecq, R, Desmet, O. E. Mahdi, B. Leclercq, G. Goormachtigh, V. Fafeur, O. Melnyk, *Angew. Chem.* **2012**, *124*, 213; *Angew. Chem. Int. Ed.* **2012**, *51*, 209.
[10] a) J.-S. Zheng, H.-K. Cui, G.-M. Fang, W.-X. Xi, L. Liu, *ChemBioChem* **2010**, *11*, 511; b) J.-S. Zheng, H.-N. Chang, F. L. Wang, L. Liu, *J. Am. Chem. Soc.* **2011**, *133*, 11080; c) J.-S. Zheng, H.-N. Chang, J. Shi, L. Liu, *Sci. China Chem.* **2012**, *55*, 64.
[11] K. S. A. Kumar, S. N. Bavikar, L. Spasser, T. Moyal, S. Ohayon, A. Brik, *Angew. Chem.* **2011**, *123*, 6261; *Angew. Chem. Int. Ed.* **2011**, *50*, 6137.
[12] G.-M. Fang, Y.-M. Li, F. Shen, Y.-C. Huang, J.-B. Li, Y. Lin, H.-K. Cui, L. Liu, *Angew. Chem.* **2011**, *123*, 7787; *Angew. Chem. Int. Ed.* **2011**, *50*, 7645.
[13] H. Nielsen, P. H. Andreasen, H. Dreisig, K. Kristiansen, J. Engberg, *EMBO J.* **1986**, *5*, 2711.
[14] a) S. Far, O. Melnyk, *Tetrahedron Lett.* **2004**, *45*, 7163; b) S. Far, C. Gouyette, O. Melnyk, *Tetrahedron* **2005**, *61*, 6138; c) M. M. Pires, J. Chmielewski, *Org. Lett.* **2008**, *10*, 837.
[15] J.-S. Zheng, S. Tang, Y. Guo, H.-N. Chang, L. Liu, *ChemBioChem* **2012**, *13*, 542.
[16] a) T. Durek, V. Y. Torbeev, S. B. H. Kent, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4846; b) J. R. Banigan, K. Mandal, M. R. Sawaya, V. Thammavongsa, A. P. A. Hendrickx, O. Schneewind, T. O. Yeates, S. B. H. Kent, *Protein Sci.* **2010**, *19*, 1840; c) K. Mandal, B. L. Pentelute, D. Bang, Z. P. Gates, V. Y. Torbeev, S. B. H. Kent, *Angew. Chem.* **2012**, *124*, 1510; *Angew. Chem. Int. Ed.* **2012**, *51*, 1481.
[17] a) Q. Wan, S. J. Danishefsky, *Angew. Chem.* **2007**, *119*, 9408; *Angew. Chem. Int. Ed.* **2007**, *46*, 9248; b) C. Haase, O. Seitz, *Angew. Chem.* **2008**, *120*, 6912; *Angew. Chem. Int. Ed.* **2008**, *47*, 6807.