Protein Labeling

DOI: 10.1002/ange.200901488

Segmental Isotopic Labeling of a Central Domain in a Multidomain Protein by Protein *Trans*-Splicing Using Only One Robust DnaE Intein**

*Alena E. L. Busche, A. Sesilja Aranko, Mehdi Talebzadeh-Farooji, Frank Bernhard, Volker Dötsch, and Hideo Iwaï**

Many proteins, including cellular signaling proteins, cell-surface receptors, and modular enzymes, are constructed from individual domains and connecting linkers.[1] Structural analysis by NMR spectroscopy or by X-ray crystallography often focuses on self-contained domains excised from full-length proteins so as to reduce the molecular weight to a manageable size for high-quality NMR spectra or to improve crystallization by removing disordered regions.[2] Such minimization often neglects functional aspects of a domain in an intact protein and may not represent all aspects of the structure–function relationship in the full-length context. While the structure determination of individual, isolated domains has been tremendously accelerated—in part through several structural genomics consortia—understanding of domain–domain interactions within a multidomain protein is lacking behind. NMR spectroscopy is an ideal tool for characterizing such, often transient, interactions. However, NMR spectroscopic analysis of large proteins often suffers from signal overlap. This problem becomes even more profound when proteins contain recurring modular domains and/or highly disordered regions. One solution to this overlap problem is segmental isotopic labeling that enables the exclusive incorporation of NMR active, stable isotopes into selected domains or parts of larger proteins,[3–7] thereby reducing the complexity of the NMR spectra.

[*] A. S. Aranko,[+] Dr. H. Iwaï
Research Program in Structural Biology and Biophysics
Institute of Biotechnology, University of Helsinki
P.O. Box 65, Helsinki 00014 (Finland)
Fax: (+358) 9-1915-9541
E-mail: hideo.iwai@helsinki.fi
Homepage: http://www.biocenter.helsinki.fi/bi/iwai/
A. E. L. Busche,[+] Dr. M. Talebzadeh-Farooji, Dr. F. Bernhard, Prof. Dr. V. Dötsch
Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance and Cluster of Excellence Frankfurt Macromolecular Complexes (CEF), University of Frankfurt
60438 Frankfurt (Germany)
[+] These authors contributed equally to this work.

[**] We thank the Protein Chemistry Research Group at the Institute of Biotechnology and Dr. Ute Bahr for mass spectrometry analysis. This work was supported by the grants from the Academy of Finland (118385), Sigrid Jusélius Foundation, the Center for Biomolecular Magnetic Resonance (BMRZ), and the Cluster of Excellence Frankfurt (CEF).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200901488.

In the past years, we have advanced segmental isotopic labeling procedures by discovering and engineering robust protein splicing domains (inteins)[6,8a] and by developing in vivo methods with the time-delayed dual expression system.[5] While segmental isotopic labeling of N- or C-terminal halves of proteins by protein ligation of two fragments has become a well-established routine procedure,[5–7] labeling of a middle domain or segment in a larger protein still poses significant challenges. Native chemical ligation (NCL) or expressed protein ligation (EPL) could be used for such a three-fragment ligation, however, the requirement for isotope-enriched peptides with a protected N-terminal cysteine residue is not cost-effective and the protection/deprotection steps to create a reactive N-terminal cysteine could be labor-intensive.[4,9] One promising approach to ligate three polypeptide fragments is to use protein *trans*-splicing (PTS) with two split inteins (denoted as Int1 and Int2) because PTS does not require any cumbersome steps to create an N-terminal cysteine needed for EPL (Figure 1a).[4,10,11] For this approach, however, the two split inteins must be orthogonal without any cross-reactivity and have high splicing efficiency. Segmental labeling of a central region of a protein by three-fragment ligation was elegantly realized with the two orthogonal split inteins, PI-*Pfu*I and PI-*Pfu*II.[10] However, the artificially split inteins of PI-*Pfu*I and PI-*Pfu*II required tedious refolding and optimization steps to restore the splicing activity because the fusion proteins with the split intein fragments were insoluble.[10,12] Since refolding can also result in inactive proteins, avoiding any refolding steps would be highly beneficial. This can be achieved with naturally occurring split inteins, such as DnaE inteins from cyanobacteria, which usually do not require any refolding step to induce PTS.[5,6,8,13,14] Particularly, DnaE intein from *Nostoc punctiforme* (*Npu*) could be ideally suited based on the combination of its robust splicing activity, high tolerance of sequence variations at the splicing junctions, and its high solubility.[8] For three-fragment ligation, however, the DnaE inteins cannot be directly used because the domain or fragment inserted in the middle of the precursor protein containing both N- and C-intein (I_N/I_C) could result in cyclization or polymerization (Figure 1b).[8a,14,15]

Herein, we develop a novel strategy for three-fragment protein ligation that overcomes the cyclization/polymerization problem but still utilizes only one robust DnaE intein to achieve higher yields. To circumvent the inevitable cyclization using only one split intein (Figure 1b),[15] we exploited engineered *Npu*DnaE inteins with the same sequence but different split sites[16] (Figure 1c). Shortening,

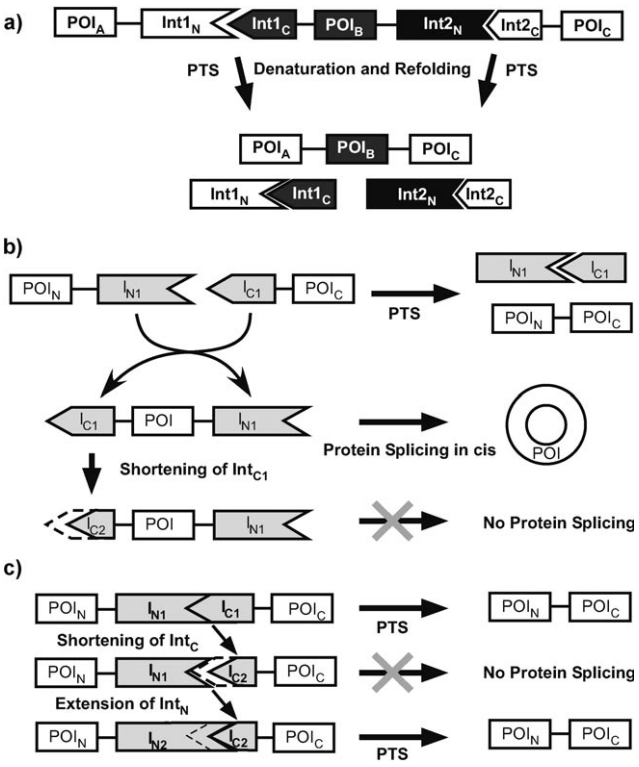

Figure 1. a) Segmental isotopic labeling of a central region of protein of interest (POI) by three-fragment ligation using two orthogonal split inteins (Int1 and Int2). b) Design of a central precursor protein that prevents cyclization caused by protein splicing in *cis*. c) Elimination of protein splicing activity by shortening the C-intein ($I_{C1}\rightarrow I_{C2}$) and restoration of protein splicing activity by complementation with an elongated N-intein ($I_{N1}\rightarrow I_{N2}$). Note, however, that a combination of I_{N2} and I_{C1} shows cross-reactivity and should therefore be avoided.

for example, the C-intein could prevent the formation of a functional intein with the unmodified N-intein (I_{N1}). The splicing activity, however, can be restored in reactions with a modified N-intein (I_{N2}) bearing the missing sequence (Figure 1c). Thus, one robust DnaE intein with two different split sites can be used for three-fragment ligation in a stepwise manner (Figure 2). In the first step two protein fragments of interest (POI_B and POI_C in Figure 2) are ligated by the original split intein (I_{N1}/I_{C1}). Additionally, the N-terminal target protein (POI_B) is fused to a shortened C-intein (I_{C2}), which does not react in *cis* with the original N-intein (I_{N1}) preventing cyclization of the middle fragment (POI_B). This first step of the three-fragment ligation can be performed in vivo as previously demonstrated.[5,6] Subsequently, the ligated product (I_{C2}-POI_B-POI_C) can be purified with an N-terminal purification tag and subjected to the second, in vitro ligation step with the elongated variant of the N-intein (I_{N2}) that is C-terminally fused to POI_A (Figure 2c). In this second step, association between the pair of the engineered split intein (I_{N2}/I_{C2}) will induce PTS and ligate the two flanking protein sequences, that is, POI_A and POI_B-POI_C, which result in the ligated product (POI_A-POI_B-POI_C) containing all three protein fragments of interest. One possible drawback of this strategy is that the ligation has to be performed in a stepwise manner rather than in a "one pot" reaction because of the cross-reactivity between the elongated N-intein (I_{N2}) and the original C-intein (I_{C1}). Performing the first ligation step in vivo, however, simplifies the protein purification steps.[5]

As a proof-of-principle experiment, we created an artificial fusion construct of the three model proteins, yeast ubiquitin-like protein Smt3, the B1 domain of the immunoglobulin binding protein G (GB1), and yellow fluorescent protein (YFP; Figure 2). The C-intein of the split variant (Npu_{C15}) was derived from the C-intein (Npu_{C36}) of the original *Npu*DnaE intein by shortening it by 21 amino acids, thus containing only the C-terminal 15 residues.[16a] The N-intein of the split variant ($Npu_{N\Delta C15}$) was created by elongating the N-intein of the original *Npu*DnaE intein ($Npu_{N\Delta C36}$) with the missing 21 residues. The first ligation step between I_{C2}-POI_B-I_{N1} (Npu_{C15}-GB1-$Npu_{N\Delta C36}$) and I_{C1}-POI_C (Npu_{C36}-YFP) was achieved in vivo by the original split *Npu*DnaE intein. The ligated product was purified by ion metal affinity chromatography (IMAC). The ligated product I_{C2}-POI_B-POI_C (Npu_{C15}-GB1-YFP, 45.3 kDa) could be easily identified based on its apparent molecular weight by SDS-PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis; Figure 2b). Over 60% of the precursor was efficiently ligated during the expression. Independently, the N-terminal precursor protein POI_A-I_{N2} containing Smt3 and $Npu_{N\Delta C15}$ was expressed and purified by IMAC. The second step of the protein ligation was initiated in vitro by mixing the two purified precursors POI_A-I_{N2} (Smt3-$Npu_{N\Delta C15}$) and I_{C2}-POI_B-POI_C (Npu_{C15}-GB1-YFP). The time course of the PTS reaction is shown in Figure 2c, clearly demonstrating the increase in the ligated product of POI_A-POI_B-POI_C (Smt3-GB1-YFP). The in vitro ligation efficiency was over 80%. The ligated product was confirmed by mass spectrometry (Supporting Information, Figure S3).

Successful three-fragment ligation with the model proteins encouraged us to apply this strategy to a naturally occurring three-domain protein. Curacin A (CurA) from *Lyngbya majuscula* is part of a mixed polyketide/nonribosomal peptide synthesis (PKS/NRPS) machinery and contains three sequentially following acyl carrier protein (ACP) domains.[17,18] High sequence identities among these three domains (93–100%) give rise to severe overlap of NMR signals from a uniformly labeled sample of the construct spanning all three domains (see Figure 4, Supporting Information Figure S2). Analysis of domain–domain interactions of such a protein containing recurring sequences by NMR spectroscopy is difficult with conventional uniform labeling as well as with selective amino acid labeling approaches. Thus, segmental isotopic labeling of individual domains in the full-length protein is a very attractive approach to analyze the structure–function relationship of individual domains within the three-domain CurA fragment. To achieve segmental isotopic labeling, CurA was dissected in the linker regions connecting the three ACP domains, termed T1, T2, and T3 (Supporting Information Figure S2). The T3 was fused with Npu_{C36}. The expression of this precursor protein was controlled by the arabinose promoter in pBAD vector with ColE1 origin (Figure 3, and Supporting Information, Figure S1). The dissected T2 was inserted between Npu_{C15} and $Npu_{N\Delta C36}$ (Figure 3a). We also fused an N-terminally His-

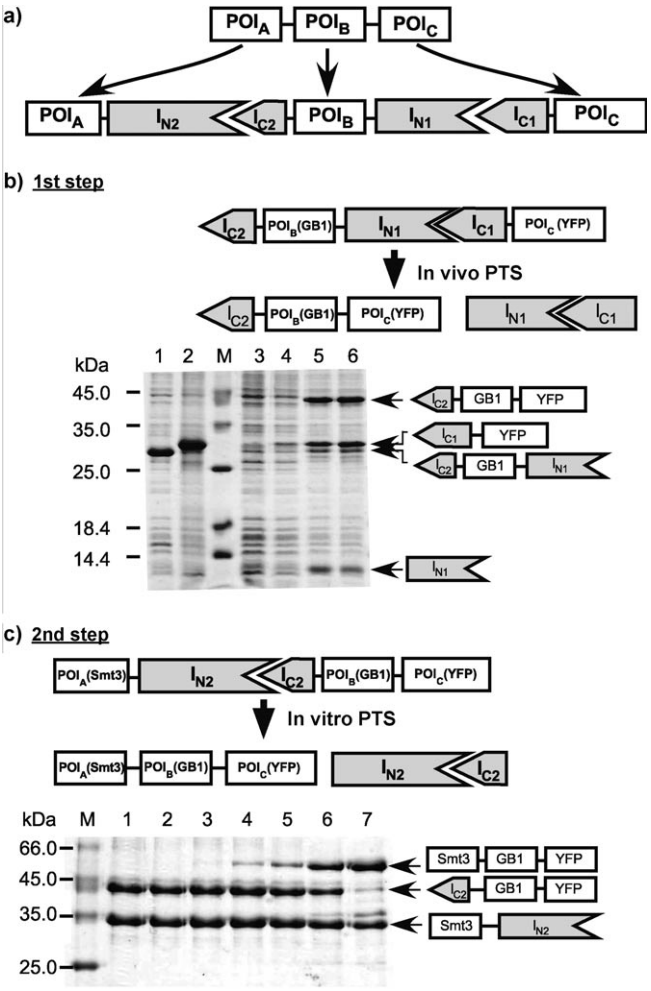

Figure 2. A stepwise three-fragment ligation by PTS. a) Design of three precursor proteins for the three-fragment protein ligation. b) The first in vivo splicing scheme and experimental results. The yield of the splice product was followed by SDS-PAGE: Lane 1: induction of only the first component with IPTG; lane 2: induction of only the second component with arabinose; lane 3: before induction; lane 4: half-hour after arabinose induction; lane 5: 2 h after induction with arabinose and IPTG (isopropyl-β-D-thiogalactopyranoside); lane 6: 4 h after the dual induction. The splice product can easily be identified by its molecular weight. c) The second in vitro splicing scheme and experimental time course followed by SDS-PAGE: Lane 1: directly after mixing of both purified components; lane 2: 3 min; lane 3: 10 min; lane 4: 30 min; lane 5: 1 h; lane 6: 3 h; lane 7: 15 h after mixing. M = molecular weight marker.

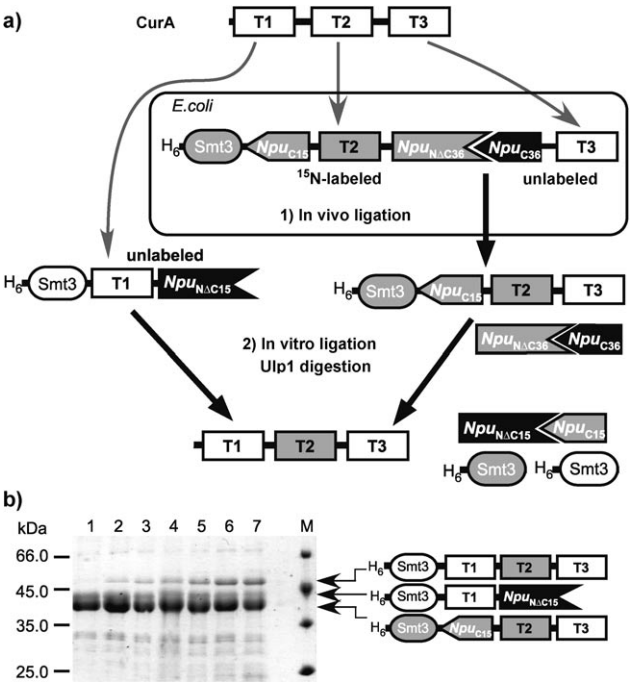

Figure 3. a) Strategy for segmental isotopic labeling of a central domain in CurA. b) Time course of the second in vitro step of the protein splicing reaction. Lane 1: directly after mixing of the purified components; lane 2: 30 min; lane 3: 1 h; lane 4: 3 h; lane 5: 5 h; lane 6: 9 h; lane 7: 17 h after mixing.

tagged Smt3 (H_6-Smt3) to the N-terminus of Npu_{C15} to facilitate the purification and improve solubility.[19] The precursor containing the T2 was cloned into a plasmid with T7 promoter and RSF origin. (Figure 3, and Supporting Information Figure S1). These two plasmids were used for in vivo segmental isotopic labeling[5,6] The C-terminal precursor with T3 was expressed first in the unlabeled medium, followed by the expression of the N-terminal precursor with T2 in the labeled medium. This order of expression helps to minimize isotopic cross-labeling to less than 10% as described previously[5,6] since removal of the arabinose by changing the media effectively ceases arabinose promoter activity. Approximately 20 mg (0.54 μmol) of the purified ligated product (H_6-Smt3-Npu_{C15}-[^{15}N]-T2-T3) containing ^{15}N-labeled H_6-Smt3-Npu_{C15}-T2 and unlabeled T3 was obtained from 2 liters of ^{15}N-labeled M9/LB medium. The spliced T2-T3 bi-domain construct was purified and used in the second step of the three-fragment ligation method. For the second step, the T1 was fused between H_6-Smt3 and $Npu_{N\Delta C15}$ (I_{N2}) (Figure 3, and Supporting Information Figure S1). This N-terminal precursor (H_6-Smt3-T1-$Npu_{N\Delta C15}$) containing the T1 was separately expressed in 1.4 liters of LB (Luria–Berani broth) medium and purified by IMAC. The second step of the ligation was performed in vitro by mixing the ligated bi-domain product (H_6-Smt3-Npu_{C15}-[^{15}N]-T2-T3) with 5 mg (0.13 μmol) of the N-terminal precursor (H_6-Smt3-T1-$Npu_{N\Delta C15}$; Figure 3b). Finally, about 2 mg (0.06 μmol) of the in vitro ligated product T1-[^{15}N]-T2-T3 with the centrally ^{15}N-labeled T2 was obtained after cleavage of the H_6-Smt3 tag and subsequent purification by anion exchange chromatography. The final product was confirmed by mass spectrometry (Supporting Information Figure S4). The yield corresponds to approximately 40% efficiency for the second ligation step after the final purification.

To evaluate the effect of our segmental isotopic labeling strategy we recorded [^{15}N, ^{1}H]-HSQC spectra of the uniformly ^{15}N-labeled T1-T2-T3 and of the centrally segmentally ^{15}N-labeled T1-[^{15}N]-T2-T3 constructs (Figure 4). Following a similar strategy, we also obtained the N-terminally segmentally ^{15}N-labeled [^{15}N]-T1-T2-T3 construct (Supporting Information).[5] A comparison of the spectra of the segmentally ^{15}N-labeled samples and the uniformly ^{15}N-labeled sample

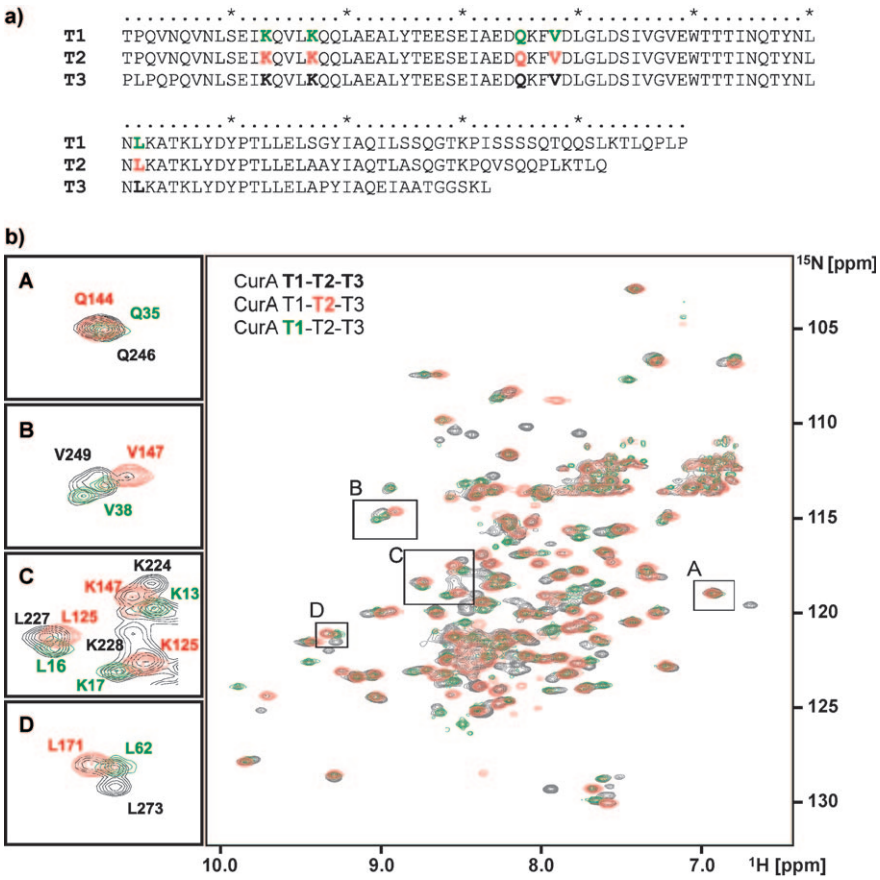

Figure 4. a) Primary structure of CurA containing three ACP domains. b) An overlay of HSQC spectra from uniformly ^{15}N-labeled CurA (T1-T2-T3) (black) and centrally ^{15}N-labeled CurA (T1-[^{15}N]-T2-T3) (red), and N-terminally ^{15}N-labeled [^{15}N]-T1-T2-T3 (green). Insets (A)–(D) are the magnifications of regions in the spectra. The assignments of a few residues are indicated.

shows that the assignments of peaks with very similar chemical shifts to individual domains are possible (insets B–D in Figure 4b). With the segmentally isotope-labeled samples, it is also possible to identify overlaps of the three signals originating from the individual domains unambiguously (inset A in Figure 4b).

In summary, we demonstrated a novel scheme for three-fragment ligation by using only one robust split DnaE intein without any refolding steps. The use of only the well-characterized *Npu*DnaE intein for three-fragment ligation is of practical advantage for obtaining higher yields. We believe that this new strategy can be generally applicable to segmental isotopic labeling of an arbitrary region in many other multidomain proteins for NMR spectroscopic studies. This simple and robust scheme for a central-fragment labeling could become a valuable tool for elucidating structure–function relationships of a domain or a region in intact proteins by NMR spectroscopy.

Received: March 18, 2009
Revised: May 5, 2009
Published online: July 9, 2009

Keywords: inteins · isotopic labeling · NMR spectroscopy · protein modifications · protein splicing

[1] a) M. Baron, D. G. Norman, I. D. Campbell, *Trends Biochem. Sci.* **1991**, *16*, 13; b) R. F. Doolittle, *Annu. Rev. Biochem.* **1995**, *64*, 287; c) T. Pawson, *Nature* **1995**, *373*, 573.
[2] a) C. J. Oldfield, E. L. Ulrich, Y. Cheng, A. K. Dunker, J. L. Markley, *Proteins Struct. Funct. Genet.* **2005**, *59*, 444; b) J. M. Canaves, R. Page, I. A. Wilson, R. C. Stevens, *J. Mol. Biol.* **2004**, *344*, 977; c) W. N. Price II, Y. Chen, S. K. Handelman, H. Neely, P. Manor, R. Karlin, R. Nair, J. Liu, M. Baran, J. Everett, S. N. Tong, F. Forouhar, S. S. Swaminathan, T. Acton, R. Xiao, J. R. Luft, A. Lauricella, G. T. DeTitta, B. Rost, G. T. Montelione, J. F. Hunt, *Nat. Biotechnol.* **2009**, *27*, 51.
[3] T. Yamazaki, T. Otomo, N. Oda, Y. Kyogoku, K. Uegaki, N. Ito, Y. Ishino, H. Nakamura, *J. Am. Chem. Soc.* **1998**, *120*, 5591.
[4] R. Xu, B. Ayers, D. Cowburn, T. W. Muir, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 388.
[5] S. Züger, H. Iwai, *Nat. Biotechnol.* **2005**, *23*, 736.
[6] M. Muona, A. S. Aranko, H. Iwai, *ChemBioChem* **2008**, *9*, 2958.
[7] a) H. Iwai, S. Züger, *Biotechnol. Genet. Eng. Rev.* **2007**, *24*, 129; b) S.-y. Ohki, M. Kainosho, *Prog. Nucl. Magn. Reson. Spectrosc.* **2008**, *53*, 208; c) L. Skrisovska, F. H. Allain, *J. Mol. Biol.* **2008**, *375*, 151.
[8] a) H. Iwai, S. Züger, J. Jin, P. H. Tam, *FEBS Lett.* **2006**, *580*, 1853; b) J. Zettler, V. Schütz, H. D. Mootz, *FEBS Lett.* **2009**, *583*, 909.
[9] a) E. C. Johnson, T. Durek, S. B. Kent, *Angew. Chem.* **2006**, *118*, 3361; *Angew. Chem. Int. Ed.* **2006**, *45*, 3283; b) T. Durek, V. Y. Torbeev, S. B. Kent, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4846.
[10] T. Otomo, N. Ito, Y. Kyogoku, T. Yamazaki, *Biochemistry* **1999**, *38*, 16040.
[11] J. Shi, T. W. Muir, *J. Am. Chem. Soc.* **2005**, *127*, 6198.
[12] a) T. Otomo, K. Teruya, K. Uegaki, T. Yamazaki, Y. Kyogoku, *J. Biomol. NMR* **1999**, *14*, 105; b) H. Yagi, T. Tsujimoto, T. Yamazaki, M. Yoshida, H. Akutsu, *J. Am. Chem. Soc.* **2004**, *126*, 16632.
[13] H. Wu, Z. Hu, X. Q. Liu, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9226.
[14] B. Dassa, G. Amitai, J. Caspi, O. Schueler-Furman, S. Pietrokovski, *Biochemistry* **2007**, *46*, 322.
[15] a) C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnon, S. J. Benkovic, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13638; b) T. C. Evans, Jr., D. Martin, R. Kolly, D. Panne, L. Sun, I. Ghosh, L. Chen, J. Benner, X. Q. Liu, M. Q. Xu, *J. Biol. Chem.* **2000**, *275*, 9091; c) H. Iwai, A. Lingel, A. Plückthun, *J. Biol. Chem.* **2001**, *276*, 16548.
[16] a) A. S. Aranko, S. Züger, E. Buchinger, H. Iwaï, *PLoS One* **2009**, *4*, e5185; b) J. S. Oeemig, A. S. Aranko, J. Djupsjöbacka, K. Heinämäki, H. Iwaï, *FEBS Lett.* **2009**, *583*, 1451.
[17] Z. Chang, N. Sitachitta, J. V. Rossi, M. A. Roberts, P. M. Flatt, J. Jia, D. H. Sherman, W. H. Gerwick, *J. Nat. Prod.* **2004**, *67*, 1356.
[18] L. Gu, J. Jia, H. Liu, K. Hakansson, W. H. Gerwick, D. H. Sherman, *J. Am. Chem. Soc.* **2006**, *128*, 9014.
[19] J. G. Marblestone, S. C. Edavettal, Y. Lim, P. Lim, X. Zuo, T. R. Butt, *Protein Sci.* **2006**, *15*, 182.

