Protein Ligation

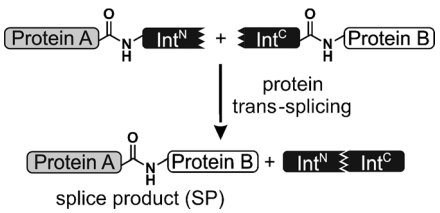
DOI: 10.1002/anie.201307969

An Atypical Naturally Split Intein Engineered for Highly Efficient Protein Labeling**

Ilka V. Thiel, Gerrit Volkmann, Shmuel Pietrokovski, and Henning D. Mootz**

Abstract: *Protein trans-splicing catalyzed by split inteins is a powerful technique for assembling a polypeptide backbone from two separate parts. However, split inteins with robust efficiencies and short fragments suitable for peptide synthesis are rare and have mostly been artificially created. The novel split intein AceL-TerL was identified from metagenomic data and characterized. It represents the first naturally occurring, atypically split intein. The N-terminal fragment of only 25 amino acids is the shortest natural intein fragment to date and was easily amenable to chemical synthesis with a fluorescent label. Optimal protein trans-splicing activity was observed at low temperatures. Further improved mutants were selected by directed protein evolution. The engineered intein variants with up to 50-fold increased rates showed unprecedented efficiency in chemically labeling of a diverse set of proteins. These inteins should prove valuable tools for protein semi-synthesis and other intein-related biotechnological applications.*

Inteins are self-excising protein segments that ligate their flanking sequences in a process termed protein splicing. Split inteins mediate protein trans-splicing as shown in Scheme 1.



Scheme 1. Protein trans-splicing. Inteins ligate their flanking sequences with a native peptide bond. These sequences can be either their native exteins or unrelated peptides or proteins.

[*] M. Sc. I. V. Thiel, Dr. G. Volkmann, Prof. H. D. Mootz
Department of Chemistry and Pharmacy
Institute of Biochemistry, University of Münster
Wilhelm-Klemm-Strasse 2, 48149 Münster (Germany)
E-mail: Henning.Mootz@uni-muenster.de
Prof. S. Pietrokovski
Department of Molecular Genetics, Weizmann Institute of Science
Rehovot 76100 (Israel)
E-mail: Shmuel.pietrokovski@weizmann.ac.il

[**] This work was supported by grants from the Deutsche Forschungsgemeinschaft (grant MO1073/3-2 and MO1073/5-1) as well as the Cells-in-Motion excellence cluster (EXC1003) to H.D.M. and a Ph.D. stipend from the German National Merit Foundation to I.V.T. S.P. holds a Hermann and Lilly Schilling Foundation chair. We thank Xiang-Qin Liu for providing plasmid pKH.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201307969.

The N- and C-terminal intein fragments (Int^N and Int^C) first associate and fold into the active intein domain and then link the flanking sequences, also termed the N- and C-terminal exteins (Ex^N and Ex^C), with a peptide bond. This reaction has found various applications in basic protein research and biotechnology,[1] for example, for the segmental isotope labeling of proteins, the preparation of cyclic polypeptides, and transgene expression, as well as more recently for the chemical modification of proteins and protein semisynthesis. However, split inteins are rare, and for the latter synthetic applications, special properties are required.[2] Specifically, one of the intein fragments should be as short as possible to facilitate its efficient and inexpensive assembly by solid-phase peptide synthesis. All naturally occurring split inteins reported to date show the breakpoint at the position of the homing-endonuclease domain in the related contiguous maxi-inteins. This split site gives rise to an Int^N of about 100 amino acids (aa) and an Int^C of about 35 aa.[3] Split inteins with shorter Int^N or Int^C fragments have been created artificially from naturally contiguous inteins,[4] but they generally show poorer performance than the naturally occurring ones. While the latter have evolved for stable expression of the individual intein fragments, artificially split intein fragments are likely to each contain internally unmatched hydrophobic patches that might cause aggregation. Furthermore, artificially split intein fragments tend to associate and fold less efficiently.[2] Besides lower splicing yields and rates, they also have a more limited compatibility with diverse target proteins as a result of their solubility and expression issues. We considered these points to be the most significant obstacle for the further exploration of semisynthetic split inteins as generally applicable tools for protein labeling and protein semisynthesis.

We identified several closely related phage genes with inteins in metagenomic data from the antarctic permanently stratified saline lake, Ace Lake.[5] The genes were from a T4-bacteriophage-type DNA-packaging terminase large subunit and were termed AceL-TerL inteins (from *Ace l*ake *ter*minase *l*arge subunit). Some of the genes included contiguous (cis) inteins, and some were fractured genes[6] split 24–25 aa from the N-terminal position of the intein (Figure S1 in the Supporting Information). Phylogenetic trees of these genes and closely related inteinless TerL genes suggest that they are the result of multiple independent intein-acquisition and gene-splitting events (Figure S2). A sequence alignment with well-characterized inteins indicated the typical conserved block motifs important for catalysis (Figure S3). Using this information to map the sequence of the AceL-TerL inteins onto the prototypical crystal structure of the *Ssp* DnaB intein[7] revealed that these inteins have a novel split site corresponding to a probable surface-loop region with no

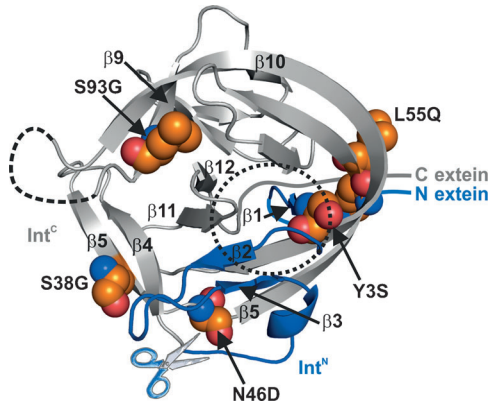


Figure 1. Localization of the AceL-TerL split site and mutations. The split site (indicated by scissors) and the selected mutations were mapped onto the crystal structure of the *Ssp* DnaB mini-intein (pdb-code 1MI8)[7] based on a sequence alignment between the two inteins (Figure S3). The active site is indicated by a dotted circle and the unstructured loop representing the position of the removed homing-endonuclease domain is shown by a dashed line.

defined secondary structure following β-strand 3 and α-helix 1 (Figure 1). Reported split sites close to the N-terminal splice junction were all created artificially and were located following aa11 of the *Ssp* DnaB intein[4d,e] and aa36 of the *Npu* DnaE intein.[4c]

We chose intein AceL-TerL-11, hereafter termed the AceL-TerL intein, for functional characterization (Figure 2). To this end, the Int^N of 25 aa was prepared by solid-phase peptide synthesis with three native N-extein residues (EFE), two lysine residues, and a 5(6)-carboxyfluorescein moiety (Fl) to give $pepWT^N$ (Fl–KKEFE–Int^N; Figure 2a). The Int^C (aa 26–129), along with four native C-extein residues (CEFL), was recombinantly expressed in *E. coli* as a fusion with hexahistidine-tagged thioredoxin as a model protein (construct WT^C-CEFL-Trx–H_6) and purified using Ni–NTA chromatography. Upon mixing of the two components, spontaneous protein trans-splicing was observed (Figure 2b). The formation of the splice product (SP) Fl–KKEFECEFL–Trx–H_6 and of the excised intein fragments as by-products indicated that the intein was fully active.

A clear temperature dependence was observed from splicing assays at 8 °C, 25 °C, and 37 °C, with the highest rate at 8 °C ($k = (1.7 \pm 0.2) \times 10^{-3}\ s^{-1}$; $t_{1/2} = (7.2 \pm 1.1)$ min) and an approximately 50-fold slower rate at 37 °C (Figure 2c and Table S1 in the Supporting Information). The preference for lower temperatures is consistent with the temperature of 0.4–1.0 °C at the sampling site of the environmental probe[5] and may indicate that the intein-gene host species is psychrophilic. Importantly, no C-terminal cleavage side product (Trx–H_6) could be detected at 8 °C, while up to 30 % was observed at 37 °C (Figure 2d and Figure S4). Such cleavage products are often observed for engineered inteins, for example, 10–20 % side product was observed for the semisynthetic *Ssp* DnaB intein and its engineered M86 mutant,[4d,8] and this effect may significantly limit the practical utility of these intein systems. Together, these results indicate the excellent potential of the naturally split AceL-TerL intein for protein trans-splicing applications. An additional advantage is its total length of only 129 aa, which makes it the shortest intein known to date.

The high splicing rate and yield of about 90 % of the AceL-TerL intein at 8 °C is already remarkable for a naturally split intein and comparable to the highly evolved M86 mutant of the artificially split *Ssp* DnaB intein,[8] which represents the current benchmark intein for the split-intein-mediated N-terminal chemical modification of proteins. To develop an even better ligation tool, we aimed to further improve the

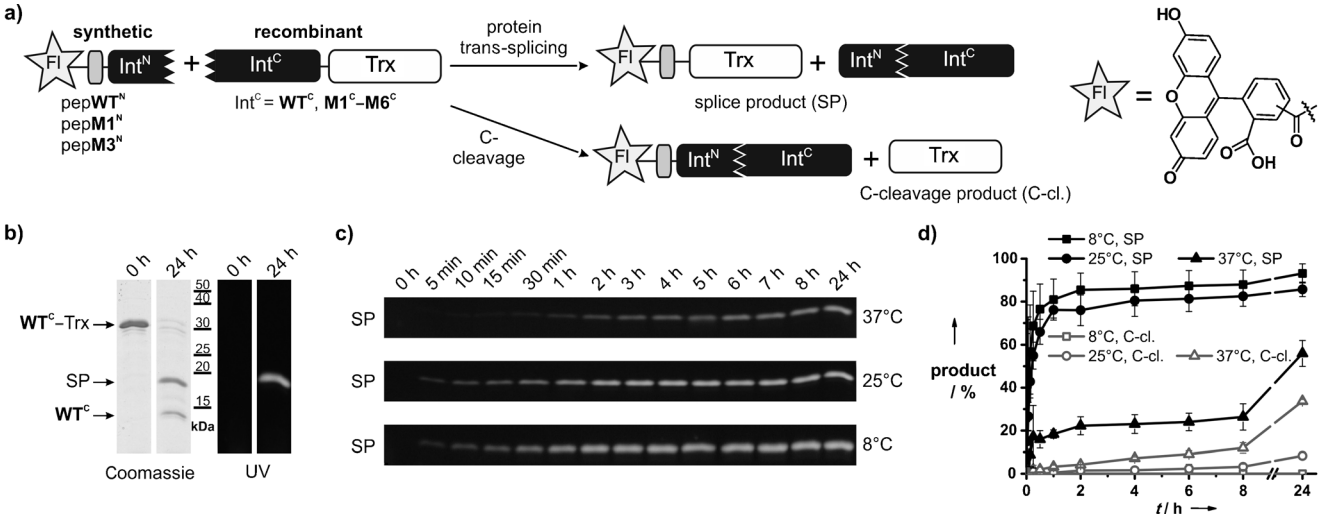


Figure 2. Reaction schemes and characterization of the wild-type AceL-TerL intein. a) Scheme of the reactions. C-terminal cleavage can occur as a side reaction. The constructs used in this study are indicated. $pepWT^N$ = Fl–KKEFE–CVYGDTMVETEDGKIKIEDLYKRLA–OH; Fl = 5(6)-carboxyfluorescein; Trx = thioredoxin, WT = wild-type). b) WT^C–Trx–His_6 (15 μM) was incubated with $pepWT^N$ (45 μM) at 25 °C (pH 7.0, 2 mM DTT) for 24 h and the reaction mixture was analyzed by SDS-PAGE by using UV illumination or Coomassie Brilliant Blue staining. Calculated molecular masses: WT^C–Trx = 26.4 kDa; SP = 15.2 kDa; Int^N = 2.9 kDa; Int^C = 12.2 kDa; C-terminal cleavage product (Trx) = 14.1 kDa. DTT = dithiothreitol. c) Splice product (SP) formation investigated at 37 °C, 25 °C, and 8 °C. The assays were performed as described in (b). d) Time courses of SP and C-terminal cleavage product (C-cl.) formation at the indicated temperatures.

intein by directed protein evolution. We reasoned that the significantly lower activity of the intein at 37 °C was likely due to compromised folding and envisaged the selection of mutants capable of efficient splicing at this temperature. Although low temperatures such as 8 °C appear ideal for preserving potentially fragile proteins of interest during in vitro protein-labeling experiments, the expression of the intein fusion proteins in *E. coli* has to be performed at higher temperatures. Furthermore, inteins with higher thermostability should be beneficial for high activity in diverse sequence contexts, potentially also at lower temperatures, and for cellular applications.[9]

To this end, the AceL-TerL intein was converted into a contiguous, cis-splicing intein by fragment fusion (Figure S5) and inserted at the DNA level into the *KanR* gene, which confers resistance to the antibiotic kanamycin.[8,10,9c] Active intein alleles capable of splicing out of the translated gene product can be selected because they render the host *E. coli* cells resistant to kanamycin.[9c,11] The nonmutated intein gave rise to colony growth at 25 °C under selective conditions (50 $\mu g\,mL^{-1}$ kanamycin), but not at 37 °C (Figure S6). This finding correlated with protein splicing activity determined by Western-blot analysis (Figure S7). These results provided the basis for selection by temperature.

Next, a library encoding mutant inteins was created by using error-prone PCR (epPCR) and used to transform *E. coli* cells. Randomly picked kanamycin-resistant colonies that were selected on plates at 37 °C were then restreaked onto plates with kanamycin concentrations of up to 150 $\mu g\,mL^{-1}$. Plasmids isolated from the resistant clones were analyzed by DNA sequencing. Five different mutant inteins, termed M1–M5 (Table 1), were selected and confirmed by Western blotting to have acquired splicing activity at this elevated temperature (Figure S8). The mutant inteins contained between one and four amino acid substitutions spread over the Int^N and Int^C parts (Table 1). To discern the effect of individual mutations, an additional construct with only the L55Q mutation, termed M6 (Table 1), was created by site-directed mutagenesis.

Table 1: The mutations in the improved AceL-TerL inteins.

	Int^N	Int^C
M1	A25T	N46D, L55Q
M2	–	S38G, N46I, N54D, L55Q
M3	Y3S	S93G
M4	–	N46I, L55Q
M5	–	N46D
M6[a]	–	L55Q
MX1[b]	–	N46D, L55Q
M31[b]	Y3S	N46D, L55Q

[a] Produced by molecular cloning. [b] Produced by rational combination of Int^N and Int^C fragments.

The effect of the mutations in the context of the split intein was then investigated. The Int^C fragments of the M1–M6 mutants, termed $M1^C$–$M6^C$, were expressed and purified as Int^C–Trx–H_6 fusion proteins, and the Int^N parts of the M1 and M3 mutants, termed $M1^N$ and $M3^N$, were included in two synthetic peptides of the format Fl–KKEFE–Int^N ($pepM1^N$ and $pepM3^N$, respectively). A clear improvement over the protein trans-splicing reactions of the wild-type intein at 37 °C could be observed for all of the mutants. The splicing rates were increased 2- to 14-fold (Figure 3a and Figure S9A), the yields increased to 65–85 % after 24 h compared to about

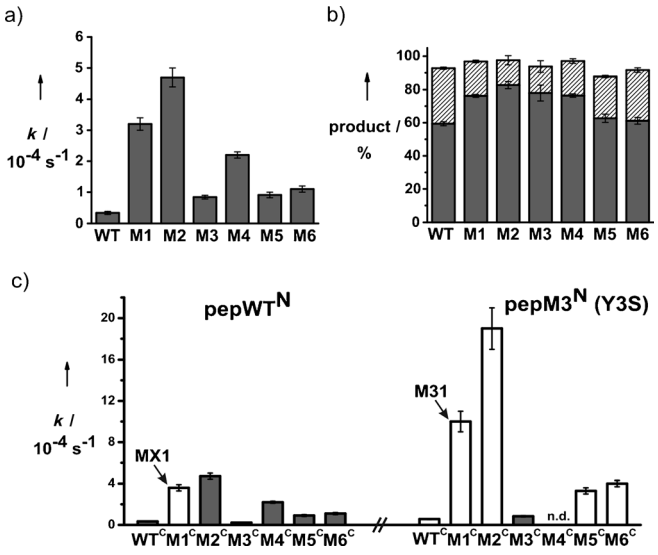


Figure 3. Characterization of the improved AceL-TerL intein mutants. The splice reaction time courses were monitored at 37 °C and rate constants were determined by fitting the product formation to pseudo-first-order kinetics. Rate constants (a) and product yields (b) for the wild-type (WT) intein and mutants M1–M6. In (b), splicing is shown by the gray bars and C-cleavage by the striped bars. c) Rate constants for the rationally combined Int^N and Int^C constructs. The new rationally designed combinations are shown as white columns, as opposed to gray columns for the genetically selected intein combinations. Combinations selected for further work are indicated (MX1 and M31 intein mutants).

60 % for the wild-type intein, and there was a concomitant decrease in the amount of C-cleavage product (Figure 3b). Most of the mutations seemed to have additive effects, as exemplified by the observation that the combined N46D and L55Q mutations ($pepWT^N + M1^C$) resulted in higher splicing rates than the individual mutations ($pepWT^N + M5^C$ and $pepWT^N + M6^C$, respectively; Figure 3c, left panel and Figure S9B). Importantly, we were also able to generate even better split inteins by rationally combining mutations from different Int^N and Int^C constructs. The two split-intein systems with the highest rates were obtained in this way, by combining the Int^N mutation of the M3 mutant (Y3S) with the Int^C mutations of the M1 (N46D, L55Q) or M2 (S38G, N46I, N54D, L55Q) mutants ($pepM3^N + M1^C$ and $pepM3^N + M2^C$). These combinations spliced approximately 29-fold and 56-fold faster, respectively, than the wild-type intein at the selection temperature of 37 °C (Figure 3c, right panel), however, the combination $pepM3^N + M1^C$ gave rise to lower amounts of by-products from C-terminal cleavage (Figure S9). Another interesting intein with a beneficial ratio between splicing and cleavage yields resulted from the combination of $pepWT^N + M1^C$. For subsequent experiments,

we therefore chose these two rational combinations and termed them mutant MX1 (pepWT^N + M1^C) and mutant M31 (pepM3^N + M1^C; Table 1 and Figure 3c).

In an attempt to rationalize the effect of the mutations, we aligned the sequence of the AceL-TerL intein with other well-studied inteins and mapped the mutations onto the crystal structure of the *Ssp* DnaB intein[7] (Figure 1 and Figure S3). The Y3S mutation is in the conserved N1 (block A) motif[12] close to the N-terminal splice junction. This position is generally found to be very variable between intein sequences and in class 3 inteins, it is frequently Thr, Ala, or Ser.[13] The N46D/I, N54D, and L55Q mutations are in close proximity to the active site residues from the C2 (block F) and N3 (block B) motifs, which are important for catalysis at the C-terminal and N-terminal splice junctions, respectively. They may thus contribute to better splicing through a fine-tuning effect on the active-site configuration. The S38G and S93G mutations are further away from the catalytic centers and are therefore more difficult to rationalize. All of the mutations may have structural roles or improve solubility.

With the evolved and characterized AceL-TerL split inteins in hand, we returned to our initial goal; the exploration of this intein for the N-terminal chemical modification of proteins. Previously described split inteins useful for semisynthetic protein trans-splicing have mostly been reported in conjunction with model proteins that are relatively small and known to be highly soluble, such as thioredoxin, maltose-binding protein (MBP), and GB1, and have only rarely been tested with more difficult proteins. In our experience, the M86 mutant of the artificially split *Ssp* DnaB intein[8] also failed for several more demanding proteins of interest (data not shown). A particular test case became the attempted N-terminal modification of the 119 kDa TycB1 module of the tyrocidine nonribosomal peptide synthetase.[14] The Int^C–TycB1 fusion of the M86 mutant of the DnaB intein gave only negligible amounts of soluble protein during expression and was subject to partial degradation (data not shown). We then included MBP as an N-terminal tag[15] and could subsequently express and purify the soluble protein in large quantities and label it by protein trans-splicing (Figure S10), thus indicating that a solubility tag like MBP at the N terminus of the Int^C fragment, which is removed with the intein during the splicing reaction, can significantly increase protein expression and solubility. Interestingly, the respective MBP–Int^C–TycB1 construct with the AceL-TerL MX1 and M31 mutants proved to be even better (Figure S11A,B). The MX1 mutant in particular was efficiently expressed and well soluble, and splicing at 8 °C gave yields of 80–95 % with a 13-fold higher rate than that of the M86 mutant of the DnaB intein and a 15-fold higher rate than that of the unevolved wild-type AceL-TerL intein (Table S2). Similar results were obtained in a detailed kinetic study when using Trx as the protein of interest (Figure S11C,D, and Table S2).

Finally, we tested the applicability of the AceL-TerL intein mutants for the chemical modification of a diverse range of proteins of interest. To this end, the MX1 mutant was fused with the green and red fluorescent proteins EGFP and mRFP, and the *Gaussia princeps* luciferase Gluc, as well as the murine E2 conjugating enzyme Ubc9 and the human protease SENP1 from the small ubiquitin-like modifier (SUMO) pathway. As shown in Figure 4 and Figure S14, all the tested proteins were efficiently modified with the synthetic fluorophore at the N-terminus, with approximately 80 % yields of the desired conjugates after 3 h at 8 °C. For biochemical

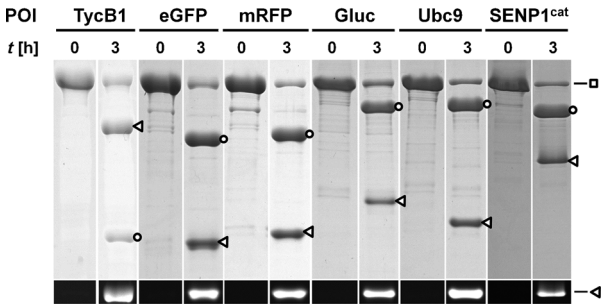


Figure 4. General applicability of protein labeling using the evolved AceL-TerL mutant MX1. The indicated proteins of interest (POIs) were expressed and purified as fusion constructs in the format MBP–M1^C–POI–H_6 and incubated with pepWT^N at 8 °C. The reactions were analyzed by SDS-PAGE by using UV illumination (bottom) and Coomassie staining (top). The unspliced construct (□), fluorescently labeled splice products (◁), and MBP–M1^C by-products (○) are indicated. Note that for each protein, the lanes were normalized to the migration of the precursor protein.

characterization of the proteins, the fluorescently labeled enzymes TycB1 and SENP1 were prepared and purified on a preparative scale and were shown to be fully catalytically active (Figures S12, S13, and S14). These labeled proteins could prove useful in future biophysical studies.

In summary, a novel type of naturally split intein and significantly improved mutants thereof are reported. We believe that these mutants represent the next generation of split inteins and are likely to serve as powerful and generally applicable tools for the N-terminal chemical modification of proteins by using semisynthetic protein trans-splicing, as demonstrated for a diverse set of proteins. The advantages of this approach over chemical ligation reactions[2] include the low reactant concentrations required, the absence of non-proteinogenic functional groups to facilitate the reaction, and orthogonality to the cellular chemical environment. The high activity of the new split inteins at low temperatures such as 8 °C is of particular advantage for in vitro labeling experiments with fragile proteins. Furthermore, no cleavage reactions leading to undesired side products could be detected under these conditions. Importantly, this first discovery of a natural split site other than the endonuclease position suggests that more naturally split inteins with even shorter fragments may be discovered.

Received: September 10, 2013
Published online: January 2, 2014

Keywords: autocatalysis · chemoenzymatic synthesis · combinatorial mutagenesis · protein engineering · protein semisynthesis

[1] a) R. Borra, J. A. Camarero, *Biopolymers* **2013**, *100*, 502; b) T. C. Evans, Jr., M. Q. Xu, S. Pradhan, *Annu. Rev. Plant Biol.* **2005**, *56*, 375; c) C. J. Noren, J. Wang, F. B. Perler, *Angew. Chem.* **2000**, *112*, 458; *Angew. Chem. Int. Ed.* **2000**, *39*, 450; d) M. Vila-Perelló, T. W. Muir, *Cell* **2010**, *143*, 191; e) G. Volkmann, H. Iwai, *Mol. Biosyst.* **2010**, *6*, 2110; f) G. Volkmann, H. D. Mootz, *Cell. Mol. Life Sci.* **2013**, *70*, 1185.
[2] H. D. Mootz, *ChemBioChem* **2009**, *10*, 2579.
[3] a) I. Giriat, T. W. Muir, *J. Am. Chem. Soc.* **2003**, *125*, 7180 – 7181; b) H. Wu, Z. Hu, X. Q. Liu, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9226.
[4] a) J. H. Appleby, K. Zhou, G. Volkmann, X. Q. Liu, *J. Biol. Chem.* **2009**, *284*, 6194; b) A. S. Aranko, S. Zuger, E. Buchinger, H. Iwai, *PloS One* **2009**, *4*, e5185; c) Y. T. Lee, T. H. Su, W. C. Lo, P. C. Lyu, S. C. Sue, *PloS One* **2012**, *7*, e43820; d) C. Ludwig, M. Pfeiff, U. Linne, H. D. Mootz, *Angew. Chem.* **2006**, *118*, 5343; *Angew. Chem. Int. Ed.* **2006**, *45*, 5218; e) W. Sun, J. Yang, X. Q. Liu, *J. Biol. Chem.* **2004**, *279*, 35281; f) G. Volkmann, X. Q. Liu, *PloS One* **2009**, *4*, e8381.
[5] F. M. Lauro, M. Z. DeMaere, S. Yau, M. V. Brown, C. Ng, D. Wilkins, M. J. Raftery, J. A. Gibson, C. Andrews-Pfannkoch, M. Lewis, J. M. Hoffman, T. Thomas, R. Cavicchioli, *ISME J.* **2011**, *5*, 879.
[6] B. Dassa, N. London, B. L. Stoddard, O. Schueler-Furman, S. Pietrokovski, *Nucleic Acids Res.* **2009**, *37*, 2560.
[7] Y. Ding, M. Q. Xu, I. Ghosh, X. Chen, S. Ferrandon, G. Lesage, Z. Rao, *J. Biol. Chem.* **2003**, *278*, 39133.
[8] J. H. Appleby-Tagoe, I. V. Thiel, Y. Wang, Y. Wang, H. D. Mootz, X. Q. Liu, *J. Biol. Chem.* **2011**, *286*, 34440.
[9] a) R. Borra, D. Dong, A. Y. Elnagar, G. A. Woldemariam, J. A. Camarero, *J. Am. Chem. Soc.* **2012**, *134*, 6344; b) T. Dhar, H. D. Mootz, *Chem. Commun.* **2011**, *47*, 3063; c) S. W. Lockless, T. W. Muir, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 10999.
[10] D. Nurizzo, S. C. Shewry, M. H. Perlin, S. A. Brown, J. N. Dholakia, R. L. Fuchs, T. Deva, E. N. Baker, C. A. Smith, *J. Mol. Biol.* **2003**, *327*, 491.
[11] A. R. Buskirk, Y. C. Ong, Z. J. Gartner, D. R. Liu, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10505.
[12] S. Pietrokovski, *Protein Sci.* **1998**, *7*, 64.
[13] K. Tori, B. Dassa, M. A. Johnson, M. W. Southworth, L. E. Brace, Y. Ishino, S. Pietrokovski, F. B. Perler, *J. Biol. Chem.* **2010**, *285*, 2515.
[14] H. D. Mootz, D. Schwarzer, M. A. Marahiel, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5848.
[15] T. Kurpiers, H. D. Mootz, *Angew. Chem.* **2007**, *119*, 5327; *Angew. Chem. Int. Ed.* **2007**, *46*, 5234.

