Protein Catenanes

Deutsche Ausgabe: DOI: 10.1002/ange.201511640
Internationale Ausgabe: DOI: 10.1002/anie.201511640

Cellular Synthesis of Protein Catenanes

*Xiao-Wei Wang and Wen-Bin Zhang**

Abstract: *Direct cellular production of topologically complex proteins is of great interest both in supramolecular chemistry and protein engineering. We describe the first cellular synthesis of protein catenanes through the use of the p53 dimerization domain to guide the intertwining of two protein chains and SpyTag–SpyCatcher chemistry for efficient cyclization. The catenane topology was unambiguously proven by SDS-PAGE, SEC, and partial digestion experiments and was shown to confer enhanced stability toward trypsin digestion relative to monomeric control mutants. The assembly–reaction synergy enabled by protein folding and genetically encoded protein chemistry offers a convenient yet powerful approach for creating mechanically interlocked, complex protein topologies in vivo.*

The past three decades have witnessed the emergence of numerous topologically complex supermolecules prepared by ingenious methods.[1] Meanwhile, topological engineering has become widely popular for tuning the physics properties of macromolecules.[2] However, similar approaches in protein engineering are only relatively recent.[3] Besides noncovalent knotted, linked, and threading folds,[4] proteins with circularized backbones have been found in nature[5] and are known to impart enhanced stability owing to increased molecular rigidity and/or reduced unfolding kinetics.[3a] Although oligomerization is common in proteins,[6] mechanically interlocked proteins are rare. While concatenated noncovalent oligomeric rings have been observed in bovine mitochondrial peroxiredoxin III,[7] recombination R (RecR) proteins from *Deinococcus radiodurans*,[8] *E. coli* Class 1a ribonucleotide reductase (RNR),[9] and CS2 hydrolase enzymes in archaeon thermophilic *Acidianus* A1-3,[10] naturally occurring interlocked covalent rings have only been found in two cases: the capsid of the bacteriophage HK97[11] and the citrate synthase protein from *Pyrobaculum aerophilum*.[12] The former is a "molecular chainmail" composed of hundreds of neighboring, interlocked cyclic protein units fixed by intra-unit isopeptide bonds[13] and the latter is formed by intramolecular disulfide bonds to enhance thermal stability for survival at high temperatures.[12] In these limited examples, nature has elegantly demonstrated the power of topology engineering in protein science with potential equal to, if not more than, that shown with synthetic macromolecules.

The synthesis of proteins with complex topology is not straightforward since the backbone is strictly linear as defined by the cellular translational machinery.[14] In recent years, a variety of methods such as native chemical ligation,[15] split-intein technology,[16] sortase-mediated coupling,[17] and SpyTag–SpyCatcher chemistry[3b,18] have been successfully applied to either the chemical synthesis or biological production of various protein topologies. Dawson et al. reported the first synthesis of short interlocked peptide catenanes using an interlacing p53 tetramerization domain cyclized by native chemical ligation.[15b] Subsequent thermodynamic analysis revealed significantly enhanced folding stability.[19] This method has been extended to prepare protein heterocatenanes and pseudorotaxanes.[20] Despite success in the synthesis of cyclotides, cyclic peptides, and lasso peptides,[14,21] it remains challenging to prepare genetically programmable protein catenanes with sizes beyond the capability of solid-state peptide synthesis. On the other hand, the synthetic tactics in supramolecular chemistry have matured to the level where not only simple [2]catenanes can be routinely prepared but higher-order complex [*n*]catenanes including Solomon links, Borromean rings, and a Star of David catenane have also been accomplished.[1] The synergy between molecular assembly and chemical reactions is well recognized and can not only enhance the stability of the assembly but also generate materials of high complexity.[22] There is plenty of room for protein engineers to achieve similar levels of sophistication in manipulating the topology and properties of proteins through direct expression in vivo. Herein, we report one such example for the cellular synthesis of recombinant protein catenanes.

The method is based on the p53 dimerization domain (p53dim)[19a,23] and genetically encoded SpyTag–SpyCatcher chemistry.[21c] The former is known to direct chain intertwining and the latter features rapid and spontaneous isopeptide bond formation between the two components under cellular conditions.[21c] This reactive pair has been applied to extend the scope of protein topology,[3b,18] to prepare protein hydrogels,[24] to enable the preparation of programmable "living" composite materials,[25] to create biological nanostructures,[26] and to develop synthetic vaccines.[27] We envisioned that these functional domains would guide the pre-assembly and ligation toward in situ protein catenane formation (Figure 1).

Two gene constructs were designed based on elastin-like protein (ELP, or E), an intrinsically disordered protein.[28] Folded proteins were not chosen in the current work since catenane formation is expected to be influenced largely by the relative positions of the N and C termini of the rigid structure. In both constructs, the p53dim (X) domain is placed in the middle of the construct to guide chain intertwining. The

[*] X.-W. Wang, Prof. Dr. W.-B. Zhang
Key Laboratory of Polymer Chemistry & Physics of Ministry of Education, Center for Soft Matter Science and Engineering
College of Chemistry and Molecular Engineering, Peking University
Beijing 100871 (P.R. China)
E-mail: wenbin@pku.edu.cn

Supporting information for this article (including experimental details) is available on the WWW under http://dx.doi.org/10.1002/anie.201511640.

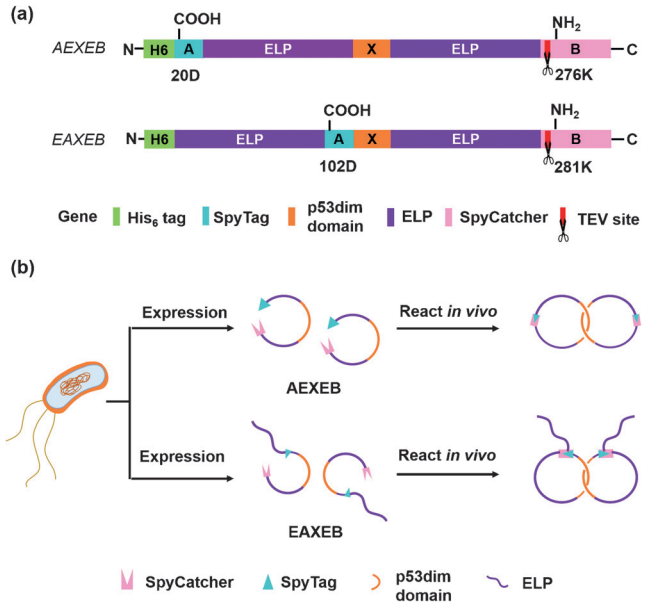

Figure 1. Illustration of the gene constructs (a) and cellular production of the protein catenanes (b).

reactive SpyCatcher (B) is fixed at the C terminus and SpyTag (A) is placed either at the N terminus or in the middle just before X. TEV proteolytic sites are introduced before SpyCatcher to facilitate probing of the chain topology. We envisioned that AEXEB would form a [2]catenane, while EAXEB would form mechanically interlocked protein "tadpoles" via domain-selective concatenation. The cyclic mutants AEX′EB and EAX′EB were prepared by using a monomeric K344P mutant of X (X′),[23] and the linear mutants A′EX′EB and EA′X′EB were prepared by further using a nonreactive D/A mutant of SpyTag (A′) (Figure S1 in the Supporting Information).[21c]

The genes were synthesized and cloned into the pQE-80L expression vector (Figure S2–4). The plasmid was used to transform the *E. coli* BL21 strain. Various conditions were screened to optimize the expression for catenane formation, among which temperature seems to be the most influential. Figure 2 shows the characterization of samples obtained through expression in 2xYT medium upon 1 mM isopropyl-β-D-1-thiogalacto-pyranoside (IPTG) induction at 37 °C and 16 °C, respectively, followed by affinity purification. It is clear that the major products are dimeric species with apparent molecular weight (MW) of approximately 85 kD, while oligomers and residual monomers (MW ≈ 40 kD) are also present. The protein catenanes were estimated to be present at around 60 % in the crude product (Table S1) and were further purified by SEC to more than 95 % purity (Figure 2c,f). The yield is approximately 40–50 mg L^{-1} for the crude products and 20–30 mg L^{-1} for the protein catenanes. The MALDI-TOF mass spectra for both AEXEB and EAXEB catenanes show two molecular ion peaks: one with single charge $[M]^+$ and the other with two charges $[M]^{2+}$ (Figures S5–6). The latter is the major peak. Since monoisotopic resolution is not possible at this molecular-weight range, the charge state is assigned based on a combination of SDS-PAGE analysis and SEC results (Figures 4 and Figure S7–9). Although their MALDI-TOF MW values (Figures S5–6) are similar to those for the corresponding cyclic and linear controls , the differences in SDS-PAGE and SEC clearly suggests different charge states. This is reasonable since large, random-coil-like ELP rings are probably only loosely associated in the gaseous phase, each carrying a single

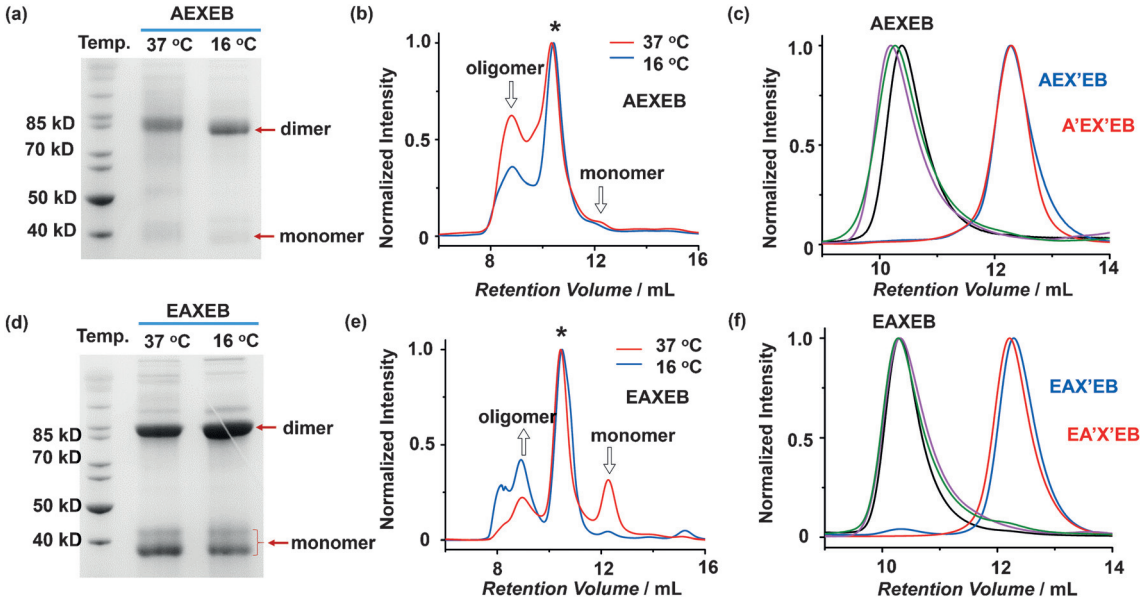

Figure 2. SDS-PAGE analysis and gel-filtration analysis of AEXEB (a,b,c) and EAXEB (d,e,f). In (b) and (e), the samples expressed at 37 °C and 16 °C and purified by affinity chromatography are shown in red and blue, respectively. The asterisk indicates the catenane and the direction of the arrow indicates the change upon lowering the expression temperature. In (c) and (f), the color code is black for catenanes, blue for cyclic controls (AEX′EB or EAX′EB), red for linear controls (A′EX′EB or EA′X′EB), green for partial digestion products of catenanes, and pink for full digestion products of catenanes.

charge. It is noteworthy that for EAXEB, there exist smeared double bands around 40 kDa, which may be due to incomplete denaturation since the SpyTag–SpyCatcher complex is tightly folded. This could be eliminated by denaturing in the presence of 3.2 M urea before SDS-PAGE analysis (Figure S10).

When expressed, the telechelic protein can undergo either intertwining/threading through X dimerization or intra-/intermolecular coupling through SpyTag-SpyCatcher chemistry. Following conventional ring–chain competition in polymer chemistry,[29] cyclic monomers, catenanes, polypseudorotaxanes and other complex structures can be formed, leading to distinct product distribution (Figure 3). The p53dim domain

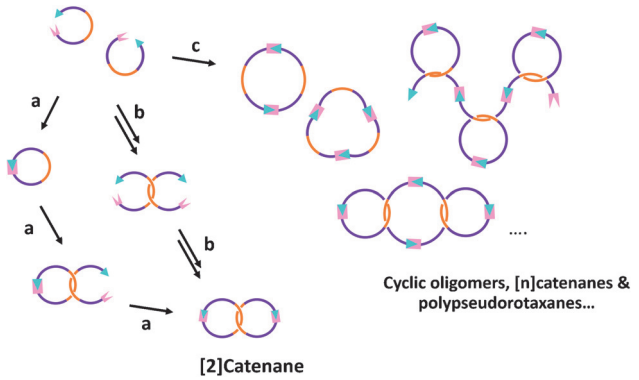

Figure 3. Possible formation pathways toward different protein topologies, including cyclic proteins, protein catenanes, and polypseudorotaxanes, during expression and posttranslational modification in situ.

(X) shows $T_m = 46\,°C$ at 40–50 μM.[23] This value drops to 30 °C at 40–50 μM for p53dim$^{lin/ext}$ as reported.[19a] To explain the different distribution, it is suggested that at 37 °C, the X dimer is relatively unstable and may only be partially folded. In this case, the threading process may be very important for catenation despite its slower kinetics, owing to the increased stability of the formed pseudorotaxane.[20] The much smaller ring and compact structure of EAXEB may slow down the threading process, which accounts for the presence of more monomeric species in the crude product for EAXEB than AEXEB at 37 °C. At 16 °C, the rapid and stable folding of X templates efficient catenane formation so that very little monomer was observed in both cases. The higher effective concentration of telechelic protein at 37 °C leads to more chain-extended oligomers, as in AEXEB. However, this is not true for EAXEB. We speculate that the presence of random-coil-like elastin-like protein near SpyTag imposes significant steric hindrance owing to the excluded-volume effect, frustrating both intermolecular and intramolecular SpyTag–SpyCatcher reactions and resulting in increased telechelic protein concentration. Hence, more chain-extended products were observed for EAXEB at lower temperatures. These results reflect the delicate influence of reactive-site location and expression temperature on protein catenation.

Since monomeric species were the major expression product for control mutants, chain intertwining through X dimerization is the primary cause of dimer formation. To prove the catenane topology, TEV digestion experiments were performed. Upon digestion, catenanes will first break into cyclic and linear monomers and then into linear monomers, whereas cyclic dimers will first break into a linear dimers and then into linear monomers. The samples were analyzed by gel filtration (Figure 2c,f) and SDS-PAGE (Figure 4). The latter clearly shows that the dimers break into cyclic and linear monomers upon partial digestion and then into linear monomers upon full digestion. The change is consistent with relinearization of the cyclic mutants by TEV. Smearing bands were observed for the cyclic and relinearized monomers of EAXEB, as well as the cyclic and relinearized EAX′EB, but not for linear EA′X′EB, which suggests a very stable, tightly folded structure that resists denaturation by SDS.[30] To confirm this, we reran the SDS-PAGE with samples in 3.2 M urea. Indeed, the smeared bands were effectively eliminated and only one band could be observed for both cyclic and linear monomers of EAXEB (Figure S10). Since smearing was not observed for AEXEB and a previous construct EAEB,[3b] we speculate that the neighboring SpyTag–SpyCatcher complex and X domain may act cooperatively for stabilization. The MW values of the partially digested samples remain identical to that of the precursor in the MALDI-TOF mass spectra (Figures S5d, S6d), which suggests that the monomer precursor is a ring. Interestingly, upon digestion, there is very little change in the elution profile of AEXEB and EAXEB in SEC (Figure 2c,f). This unambiguously shows that the X domain is responsible for dimer formation, both before and after TEV digestion. The change in hydrodynamic radius is smaller for EAXEB than for AEXEB since the former already contains a significant linear portion. This evidence supports the efficient formation of protein catenanes.

The change in properties upon catenation is of great interest. While an improvement in stability has been nicely demonstrated previously for peptide-based catenanes,[19a] it is

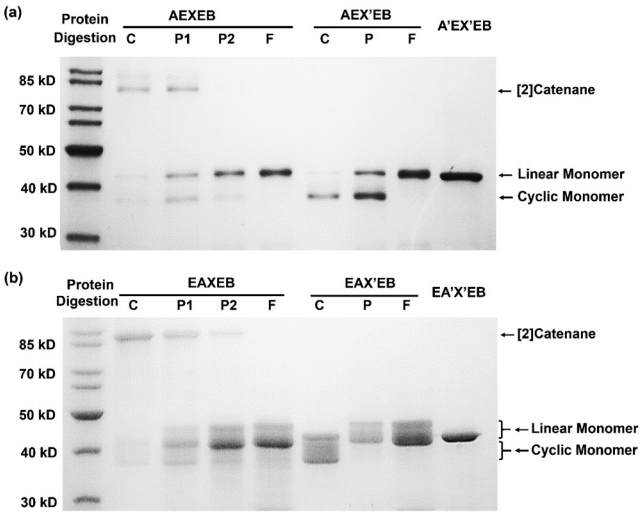

Figure 4. SDS-PAGE analysis of the TEV digestion products of AEXEB (a) and EAXEB (b) catenanes in comparison to the cyclic and linear controls. C = catenanes or cyclic controls; P = partial digestion; F = full digestion.

not clear whether such stabilization still holds for large proteins. The biophysical characterization of protein catenanes in this work is complicated by the presence of multiple structure domains with distinct feature. For example, the CD spectra (Figure S11) show that these samples are largely disordered. The change in CD signal upon raising the temperature (Figure S12) is minor and hard to attribute to a specific domain. Alternatively, we hypothesized that catenation may make protein more resistant toward proteolysis. Hence, trypsin, one of the most commonly used proteases, was selected to test the stability of catenanes under native conditions with the same amounts of protein and trypsin. The results are shown in Figure 5 and Figures S13–15. Indeed, catenanes are much more resistant to proteolysis than linear products. There is still a large portion of intact AEXEB catenanes after 5 h of incubation, while linear mutants are completed digested within 10 min. The digestion of EAXEB catenane also appears to be much slower, although they are still completely consumed in 3 h. This is probably because it has a much higher linear portion. The stability of the cyclic mutants are relatively difficult to assess since only a tiny band shift can be traced (Figures S14–15). The results are reasonable considering the possible digestion sites (Figure S16–18). Trypsin will mainly work on the N and C termini and the middle p53dim domain since the tightly folded SpyTag–SpyCatcher complex is likely to be resistant to trypsin digestion. Significant MW change could only be observed if the p53dim domain of the catenanes and linear monomer was cleaved. The change in MW for cyclic monomers upon digestion is thus very minor. Nevertheless, it seems that most of the cyclic monomers are digested in 1–2 h. The results suggest that catenanes are generally much more stable than both cyclic and linear mutants.

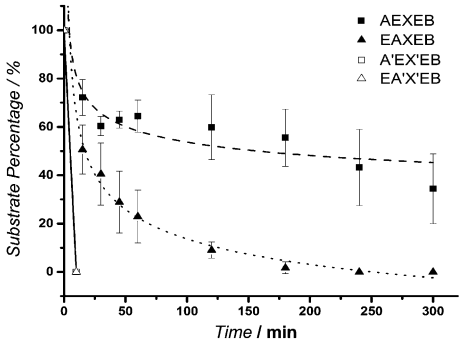

Figure 5. Digestion profile of catenanes and their linear control mutants.

In summary, we have achieved the cellular synthesis of protein catenanes and mechanically interlocked protein "tadpoles". The structure and topology of these protein catenanes have been unambiguously demonstrated. They also exhibit enhanced stability toward proteolysis. To the best of our knowledge, these are the first mechanically interlocked artificial proteins prepared in vivo to date. It is also expected that the method can be adapted to prepare protein catenanes based on other functional protein domains and to prepare proteins with even more complex topologies. We believe that this study opens up new avenues in protein engineering for creating highly sophisticated supramolecular machinery.

Acknowledgements

We are grateful for the financial support from the 863 Program (2015AA020941), the National Natural Science Foundation of China (Grants 21474003, 91427304), and "1000 Plan (Youth)".

Keywords: catenanes · protein structures · protein folding · supramolecular chemistry · topology

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 3442–3446
Angew. Chem. **2016**, *128*, 3503–3507

[1] G. Gil-Ramírez, D. A. Leigh, A. J. Stephens, *Angew. Chem. Int. Ed.* **2015**, *54*, 6110–6150; *Angew. Chem.* **2015**, *127*, 6208–6249.
[2] N. Hadjichristidis, *Complex macromolecular architectures*: *Synthesis characterization*, *and self-assembly*, Wiley, Singapore, **2011**..
[3] a) H.-X. Zhou, *Acc. Chem. Res.* **2004**, *37*, 123–130; b) W.-B. Zhang, F. Sun, D. A. Tirrell, F. H. Arnold, *J. Am. Chem. Soc.* **2013**, *135*, 13988–13997.
[4] a) C. Liang, K. Mislow, *J. Am. Chem. Soc.* **1995**, *117*, 4201–4213; b) T. O. Yeates, T. S. Norcross, N. P. King, *Curr. Opin. Chem. Biol.* **2007**, *11*, 595–603.
[5] M. Trabi, D. J. Craik, *Trends Biochem. Sci.* **2002**, *27*, 132–138.
[6] M. H. Ali, B. Imperiali, *Bioorg. Med. Chem.* **2005**, *13*, 5013–5020.
[7] Z. Cao, A. W. Roszak, L. J. Gourlay, J. G. Lindsay, N. W. Isaacs, *Structure* **2005**, *13*, 1661–1664.
[8] B. I. Lee, K. H. Kim, S. J. Park, S. H. Eom, H. K. Song, S. W. Suh, *EMBO J.* **2004**, *23*, 2029–2038.
[9] C. M. Zimanyi, N. Ando, E. J. Brignole, F. J. Asturias, J. Stubbe, C. L. Drennan, *Structure* **2012**, *20*, 1374–1383.
[10] M. B. van Eldijk, I. van Leeuwen, V. A. Mikhailov, L. Neijenhuis, H. R. Harhangi, J. C. van Hest, M. S. Jetten, H. J. M. O. den Camp, C. V. Robinson, J. Mecinovic, *Chem. Commun.* **2013**, *49*, 7770–7772.
[11] W. R. Wikoff, L. Liljas, R. L. Duda, H. Tsuruta, R. W. Hendrix, J. E. Johnson, *Science* **2000**, *289*, 2129–2133.
[12] D. R. Boutz, D. Cascio, J. Whitelegge, L. J. Perry, T. O. Yeates, *J. Mol. Biol.* **2007**, *368*, 1332–1344.
[13] a) R. L. Duda, *Cell* **1998**, *94*, 55–60; b) C. Helgstrand, W. R. Wikoff, R. L. Duda, R. W. Hendrix, J. E. Johnson, L. Liljas, *J. Mol. Biol.* **2003**, *334*, 885–899.
[14] S. B. Kent, *Angew. Chem. Int. Ed.* **2013**, *52*, 11988–11996; *Angew. Chem.* **2013**, *125*, 12208–12217.
[15] a) J. A. Camarero, J. Pavel, T. W. Muir, *Angew. Chem. Int. Ed.* **1998**, *37*, 347–349; *Angew. Chem.* **1998**, *110*, 361–364; b) L. Z. Yan, P. E. Dawson, *Angew. Chem. Int. Ed.* **2001**, *40*, 3625–3627; *Angew. Chem.* **2001**, *113*, 3737–3739.
[16] a) H. Iwai, A. Lingel, A. Pluckthun, *J. Biol. Chem.* **2001**, *276*, 16548–16554; b) J. K. Böcker, K. Friedel, J. C. Matern, A. L. Bachmann, H. D. Mootz, *Angew. Chem. Int. Ed.* **2015**, *54*, 2116–2120; *Angew. Chem.* **2015**, *127*, 2144–2148.
[17] a) J. M. Antos, M. W. Popp, R. Ernst, G. L. Chew, E. Spooner, H. L. Ploegh, *J. Biol. Chem.* **2009**, *284*, 16028–16036; b) J. Hu, W. Zhao, Y. Gao, M. Sun, Y. Wei, H. Deng, W. Gao, *Biomaterials* **2015**, *36*, 13–19.

[18] C. Schoene, J. O. Fierer, S. P. Bennett, M. Howarth, *Angew. Chem. Int. Ed.* **2014**, *53*, 6101 – 6104; *Angew. Chem.* **2014**, *126*, 6215 – 6218.
[19] a) J. W. Blankenship, P. E. Dawson, *J. Mol. Biol.* **2003**, *327*, 537 – 548; b) H.-X. Zhou, *J. Am. Chem. Soc.* **2003**, *125*, 9280 – 9281.
[20] J. W. Blankenship, P. E. Dawson, *Protein Sci.* **2007**, *16*, 1249 – 1256.
[21] a) I. Pusterla, J. W. Bode, *Nat. Chem.* **2015**, *7*, 668 – 672; b) B. Zakeri, M. Howarth, *J. Am. Chem. Soc.* **2010**, *132*, 4526 – 4527; c) B. Zakeri, J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy, M. Howarth, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E690 – E697; d) J. D. Hegemann, M. Zimmermann, X. Xie, M. A. Marahiel, *Acc. Chem. Res.* **2015**, *48*, 1909 – 1919; e) P. G. Arnison, et al., *Nat. Prod. Rep.* **2013**, *30*, 108 – 160.
[22] a) Y. Wang, H.-X. Lin, S.-Y. Ding, D.-Y. Liu, L. Chen, Z.-C. Lei, F.-R. Fan, Z.-Q. Tian, *Sci. Chin. Chem.* **2012**, *42*, 525 – 547; b) Y. Wang, H. X. Lin, L. Chen, S. Y. Ding, Z. C. Lei, D. Y. Liu, X. Y. Cao, H. J. Liang, Y. B. Jiang, Z. Q. Tian, *Chem. Soc. Rev.* **2014**, *43*, 399 – 411.
[23] T. S. Davison, X. Nie, W. Ma, Y. Lin, C. Kay, S. Benchimol, C. H. Arrowsmith, *J. Mol. Biol.* **2001**, *307*, 605 – 617.
[24] F. Sun, W.-B. Zhang, A. Mahdavi, F. H. Arnold, D. A. Tirrell, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 11269 – 11274.
[25] a) A. Y. Chen, Z. Deng, A. N. Billings, U. O. Seker, M. Y. Lu, R. J. Citorik, B. Zakeri, T. K. Lu, *Nat. Mater.* **2014**, *13*, 515 – 523; b) P. Q. Nguyen, Z. Botyanszki, P. K. Tay, N. S. Joshi, *Nat. Commun.* **2014**, *5*, 4945.
[26] M. Fairhead, G. Veggiani, M. Lever, J. Yan, D. Mesner, C. V. Robinson, O. Dushek, P. A. van der Merwe, M. Howarth, *J. Am. Chem. Soc.* **2014**, *136*, 12355 – 12363.
[27] Z. Liu, H. Zhou, W. Wang, W. Tan, Y. X. Fu, M. Zhu, *Sci. Rep.* **2014**, *4*, 7266.
[28] J. Habchi, P. Tompa, S. Longhi, V. N. Uversky, *Chem. Rev.* **2014**, *114*, 6561 – 6588.
[29] G. G. Odian, *Principles of Polymerization*, Wiley-Interscience, Hoboken, **2004**.
[30] L. Li, J. O. Fierer, T. A. Rapoport, M. Howarth, *J. Mol. Biol.* **2014**, *426*, 309 – 317.

Received: December 15, 2015
Published online: February 2, 2016

