Biomaterials

DOI: 10.1002/ange.201203795

Fibrous Nanostructures from the Self-Assembly of Designed Repeat Protein Modules**

*Jonathan J. Phillips, Charlotte Millership, and Ewan R. G. Main**

Biomaterials are the focus of much interest, since they enable access to biological complexity by using nanoscale fabrication and functionality.[1] In particular the directed self-assembly of biological systems to form nanostructures (so-called "bottom-up fabrications") has immense potential in fields as diverse as tissue engineering,[2] biosensing,[3] and nanoelectronics.[4] The inspiration and aspirations of such synthetic biomaterials can be easily seen when observing the vast array of self-associating materials used throughout Nature (for example collagen, tubulin, laminin, and actin). Notably, these materials are frequently found to be fibrous in their morphology. This permits large assemblies to be produced from simple building blocks by controlling the minimum number of interfaces. In particular, designing and engineering protein- and peptide-based functional materials seems to hold, debatably, the greatest opportunities in this area, since these materials form the more diverse structures and functions in biology. The best characterized systems so far can be split into peptides based on coiled-coil motifs,[5] peptides and proteins that form amyloidgenic fibers,[1e,6] and peptide/protein systems that can form gels.[7]

Herein we show that designed repeat protein domains can be triggered to form superhelical filaments comprised of a single protein chain. The formation of filaments was achieved by controlling the chemical properties and solvent exposure at the terminal interfaces of the protein domains. Significantly, with regard to the widespread potential of this technology, this is achieved in aqueous solvent at neutral pH, 25 °C and in less than 48 h. Our monomeric protein building block was a 100 amino acid domain comprised of three designed linearly arrayed consensus tetratricopeptide repeat units (CTPRs). These building blocks were recombinantly engineered and folded into discrete conformationally rigid alpha-helical nonglobular domains. Triggered specific head-to-tail self-assembly was driven by orthogonal native chemical ligation genetically encoded using intein chemistry. With a naturally occurring library of TPR domains, each with their own cognate peptide ligand, this system presents an exciting opportunity for the synthesis of multifunctional biomaterials by facile methods.

Tetratricopeptide repeats (TPRs) are 34-residue helix-turn-helix motifs, which only become stably folded when stacked together to form elongated domains[8] (Figure 1). Unlike globular proteins, TPRs do not rely on topologies stabilized by interactions of residues distant in primary sequence. Instead, their modular, nonglobular structures are dominated by regular interactions of residues close in primary sequence.[8] These distinctive features have made the design of proteins that contain arrays of repeats composed of consensus residues extremely successful.[8,9] Significantly for biomaterial synthesis, these designs have shown that the consensus proteins can be easily lengthened or shortened by

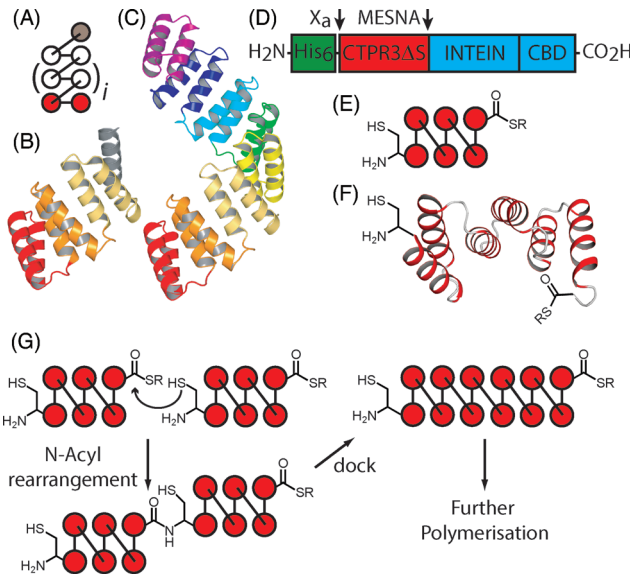

Figure 1. Topology map (A) and structures of CTPR3 (B) and CTPRa8 proteins (C). Brackets in (A) indicate that CTPR proteins can be increased or decreased in size whilst retaining the same topology (*i* is a non-negative integer). In (A–C) consecutive TPR units are colored differently: from red at the N terminus to gray at the solvating C-terminal helix (absent from the crystal structure of CTPRa8). D) Recombinant fusion protein used in study. The arrows show where factor Xa protease and MESNA cause the fusion protein to be cleaved, thereby yielding a CTPR3ΔS protein with an N-terminal cysteine and C-terminal thioester (E, F); (F) is based on the structure of CTPR3 (i.e. identical minus C-terminal S helix). G) Scheme for native chemical ligation (NCL) polymerization of TPR protein monomers. 1) Orthogonally bifunctional monomers form a head-to-tail covalent dimer, which undergoes N-acyl rearrangement to yield a peptide bond. 2) These dimers then dock to produce a continuous CTPR superhelix. CBD = chitin binding domain.

[*] Dr. J. J. Phillips, C. Millership, Dr. E. R. G. Main
School of Biological and Chemical Sciences
Queen Mary University of London
Mile End Road, London E1 4NS (UK)
E-mail: e.main@qmul.ac.uk

[**] J.J.P and this work were supported by BBSRC Grant E005187/1 and QMUL. C.M is supported by a BBSRC studentship. We thank Dr. R. Rose, Dr. A. Lowe, and members of SBCS (QMUL) for critical reading of the manuscript and insightful discussions.

Supporting information (including detailed experimental procedures and further biophysical characterization) for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201203795.

the addition or removal of identical repeat motifs.[9,10] In particular, we with Regan and co-workers have recombinantly produced soluble, monomeric proteins with up to 20 consensus TPR modules.[9,11] Strikingly, as you increase the number of consensus TPR motifs within a protein, large elongated superhelical structures with identical inter- and intra-repeat interactions are produced[11c] (Figure 1). Moreover, increasing the number of stacked repeats gives proteins with higher stability.[11b,c] Thus, if an arbitrarily large structure comprised of consecutively arrayed linear repeat units could be produced, it should form helical filaments with a free-energy stability that is orders of magnitude greater than that of the soluble monomers.

One method of producing larger superstructures would be through further recombinant expression of repeat proteins with greater repeat numbers within a single gene. However, such a process is severely limited, not least in the size and yield of protein that can be produced. Therefore, we chose a strategy of bottom-up directed self-assembly, specifically, the polymerization of recombinantly produced, discretely folded TPR protein domains (Figure 1 D–F). These monomers were based on the original designed CTPR3 protein (3 consensus TPR motifs of 102 amino acids in total).[9] CTPR3 was chosen as 1) it is extremely easy to recombinantly synthesize,[9] 2) it is highly stable (12.0 ± 0.7 kcal mol^{-1}, at pH 7[12]), 3) it is structurally very rigid,[13] and 4) it contains a minimal 3-TPR motif unit that is used ubiquitously throughout Nature as a peptide binding motif.[14] These features are important, since (1) and (2) provide abundant building blocks that remain folded under a range of conditions, (3) prevents the building blocks from futile intra-protein cyclization, and (4) presents a viable route for future decoration and functionalization.[1g,15]

The CTPRs were originally engineered to be monomeric through the addition of a single C-terminal (solvating) α-helix. This was thought to be necessary, since natural TPR proteins that lack a capping motif are prone to nonspecific aggregation.[9] By introducing a capping motif the hydrophobic core and oligomerization interface are shielded, thereby preventing the head-to-tail association required for polymerization.[9,16] Consequently two key redesign changes were essential in re-engineering CTPR3 as a directed self-assembly system: 1) introduction of a compatible oligomerization interface at the C terminus and 2) an orthogonal and triggerable chemistry to drive the correct head-to-tail association into fibrils. Step (1) was achieved by removing the C-terminal helix to produce a CTPR3ΔS protein (CTPR3 minus the solvating C-terminal helix [S helix]). This construct was recombinantly produced and found to be highly stable, fully α-helical, and folded (Figure S1 in the Supporting Information). Moreover, owing to its pI value of approximately 4 and the pH regime of the experiment (pH 7), the CTPR3ΔS protein remains monomeric in solution, as judged by size exclusion chromatography and dynamic light scattering (Figure S3 in the Supporting Information). Step (2) was realized by genetically encoding for intermolecular native chemical ligation using expressed protein ligation (EPL).[17] NCL has been used in peptide systems to create novel nanostructures and therefore has several distinct advantages over other chemistries.[18] NCL requires no protecting groups, can be genetically encoded, works at neutral pH value in aqueous buffer, kinetics are productive within a temperature range in which the proteins are natively folded, it requires no catalyst and results in the formation of a native peptide bond, such that the site of ligation is indistinguishable from the rest of the polypeptide chain.

The prerequisites for intermolecular NCL are a cysteine residue at the N terminus and a thioesterified C terminus (Figure 1 D). To achieve this, a fusion gene was constructed (Factor Xa cleavage site – CTPR3ΔS – inteinMxeGyrA – chitin binding domain) and inserted into a His Tag expression vector (Figure 1 C; pTrcHis-TOPO, Invitrogen). The fusion gene product was expressed in *E. coli* and purified using chitin affinity chromatography. His_6-Xa-CTPR3ΔS protein was released from the chitin column, with concomitant thioesterification of the carboxylate terminus, by inducing intein-mediated cleavage of the fusion protein with the addition of reducing agent (0.5 M), sodium 2-mercaptoethane sulfonate (MESNA, SIGMA-Aldrich). This cleavage yielded C-terminally activated, but N-terminally blocked, CTPR3ΔS monomer that was approximately 95% pure as judged by SDS PAGE (Figure 2 I,J). To trigger polymerization, factor Xa protease (Sigma) was added to the solution (in mild conditions: phosphate (50 mM, pH 7), at 25 °C). The factor Xa protease specifically cleaved at the sequence IEGR/C, thus releasing the His tag and yielding an active monomer with an N-terminal cysteine. NCL then occurred between the cysteine of one CTPR3ΔS with the C-terminal thioester of another (according to the Scheme in Figure 1 G).

After initiation, polymerization of the CTPRs was left to proceed at 25 °C. The reaction was monitored by analyzing samples (taken at hourly then longer time points) by SDS PAGE (Figure 2 I,J), circular dichroism (CD) spectroscopy (Figure S5 in the Supporting Information), mass spectrometry (Figure S4 in the Supporting Information), and transmission electron microscopy (TEM) by using negative staining (Figure 2 and Figures S1, S6 in the Supporting Information). These results showed that the assembly of CTPR3ΔS units proceeded smoothly with little nonspecific aggregation and under second-order kinetics. Dimers (6 TPR modules) and higher oligomers were quickly formed (Figure 2). The process then became highly combinatorial as each of the products (dimer, trimer etc.) can react further with monomer or with other product species. This phenomenon is seen in the SDS-PAGE gel where later time points show many larger discrete protein sizes.

Comparison of the far UV CD data at increasing time points show that the polymerizing species obtained from the NCL reaction retain their α-helical fold (minimum at 222 nm) and acquired greater signal as the reaction progressed (Figure S5 in the Supporting Information). This result is in line with the increase in population of CTPR proteins that possess longer chains of consecutively folded TPR motifs.

Structurally, TEM images show that the early phases of polymerization ($t = < 8$ h) were characterized by a narrow size range of small granular species. An estimation of specific particle sizes was difficult owing to the very low phase contrast (Figure 2). Excitingly, after 12 h the micrographs

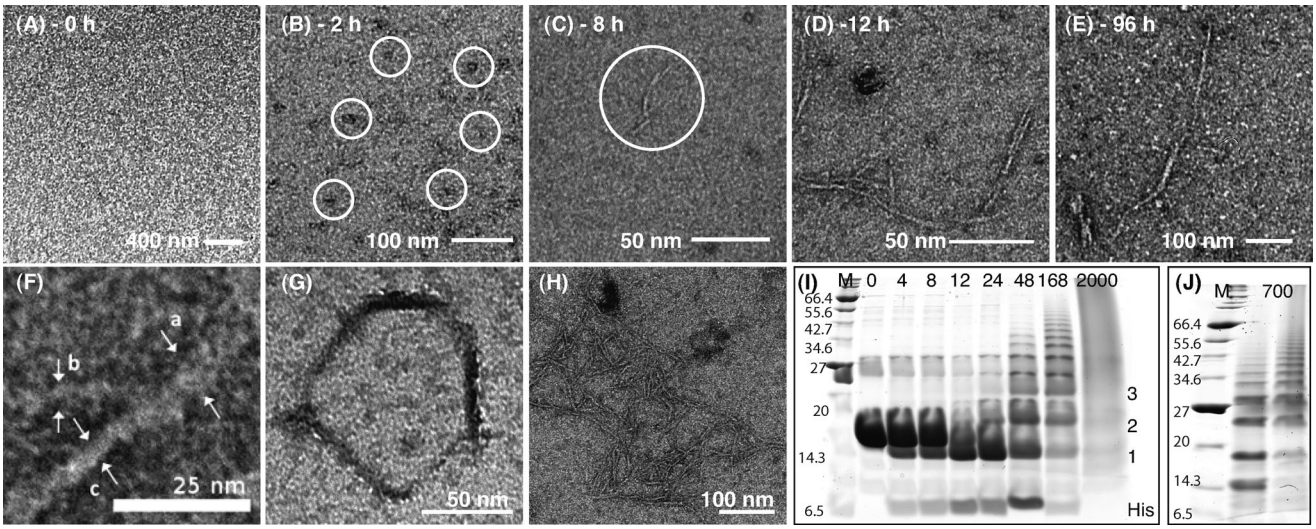

Figure 2. A–E) TEM micrographs of CTPR polymers (negatively stained with uranyl acetate) detail their morphology during NCL self-assembly. A) Initially, soluble protein appears as an even distribution across the grid with no discernible features. B) After 2 h, small globular single particles are visible, ca. 7 nm in diameter. C) By 8 h, there are many small particles with globular morphology and some with a filamentous appearance. D) After 12 h, many more filaments are visible, with lengths up to hundreds of nanometers. E) At 96 h, many filaments are observed, up to microns in length, often co-associated into thicker fibers. F–H) TEM images showing how the filaments of the CTPRs can associate together to form fibers. F) A fiber (position a—17 nm in diameter) with a branch point producing two filaments (b and c—8 and 9 nm in diameter, respectively) is shown. G) Although the growing CTPR polymer is of filamentous morphology, the polymers are not perfectly straight and the bends in the chain give rise to the possibility of intramolecular NCL reaction. H) Zoomed-out image with many single filaments present. I, J) SDS PAGE analysis of the soluble fraction of the CTPR polymerization process. NCL polymerization starts with the cleavage of monomer by factor Xa (time = 0 h). Activated monomer (1) is liberated and reacts to quickly form higher-order oligomers (activated dimer (2) and activated trimer (3)). The gels in (I) and (J) show the polymerization under different conditions: (I) = protein (200 μM), phosphate (50 mM, pH 7), MESNA (100 μM), and factor Xa (1.6 units); and (J) = protein (200 μM), phosphate (50 mM, pH 7), MESNA (100 μM), and factor Xa (1.6 units) with and without 150 mM NaCl (left-hand lane and right-hand lane, respectively). The lanes are labeled with the reaction times in hours. Note, masses of the CTPR proteins are not accurately calibrated by molecular weight markers (M) owing to their high charge (pI ca. 4 for CTPR3ΔS monomer).

show structures that were growing in size, thus gaining in aspect ratio such that some of them corresponded to short filaments (Figure 2 A–D, Figure S2 in the Supporting Information). These species all appear to be approximately 7 nm wide, but have differing lengths. As the reaction continued ($t > 24$ h), there was a further increase in the frequency of observed filamentous species, coupled with an increase in the average length of the particles.

By $t = 96$ h the distribution of sizes had completely changed from the earlier time points. There was no clear modal size, and many filamentous particles were observed. These include individual fibers that range in size up to approximately 1 μm in length but still only approximately 7 nm wide (Figure 2; Figure S2 in the Supporting Information). Strikingly, the observed thickness of the filamentous CTPR morphologies of approximately 7 nm is in close agreement with the diameter of a single CTPR superhelix determined with X-ray crystallographic studies (once negative stain is accounted for). This result shows that the CTPR units polymerize, as predicted, by defined head-to-tail association to single protein chain filaments (Figure 1 G; Figure 2).[9,11c]

Interestingly, these filaments are only measurably linear for distances of less than 30 nm. The lack of rigidity observed in the TPR fibers beyond approximately 30 nm is, perhaps, surprising given that SAXS experiments show that up to 20 TPR motif proteins (ca. 18 nm length) have "rigid" superhelical structures.[1g] However, our TPR fibers seem predominantly rigid on that length scale, too, thus suggesting that the persistence length of a CTPR fiber is reached between 18 and 30 nm. Flexibility beyond this length shows that the CTPR polymer has some degree of freedom. This may be attributed either to a lack of inherent rigidity in the superhelical structure or to incomplete docking together of units after covalent linkage (Figure 1 G, Figure 2, Figure S6 in the Supporting Information). Moreover, although flexible single filaments were predominantly observed, thicker fibers were also seen. They possessed diameters ranging up to 18 nm (Figure 2 F–G, Figure S6 in the Supporting Information).The observation of thicker fibers indicates that the individual filaments of CTPR polymers are not only flexible, but can also co-associate stably to form higher-order assemblies. Indeed, after incubation for longer than two weeks, highly branched, irregular networks of CTPR fibers predominated (Figure S6 in the Supporting Information). The structural analyses above highlight three key characteristics of the CTPR polymerization: 1) flexible folded CTPR single protein chain filaments can be synthesized under gentle and physiologically compatible conditions, 2) the majority of these filaments has a thickness that corresponds to just a single CTPR superhelical fold and 3) larger fibril bundles form after longer time periods. Features (1) and (2) should enable simple functionalization through peptide decoration, since single filaments present the interior face of the superhelix to solvent (this is the location of

nanomolar-affinity peptide binding sites in naturally occurring TPRs).[1g] Moreover, the association of single CTPR filaments suggests that protein hydrogel formation may also be possible.

To conclude, we have presented the synthesis by native chemical ligation of a novel fibrous biopolymer from designed consensus repeat protein monomers. The polymers consist of single protein chains that are helically folded and show a filamentous morphology; the filaments exhibit high aspect ratio and extend up to microns in length. The individual filaments can associate into fibers, thereby giving rise to heterogeneous branching and, ultimately, extensive networks. Central to the application of these structures to future biomaterials is that they are both 1) formed from simple, soluble modular building blocks recombinantly expressed in large quantities and 2) that each monomer contains a putative pentapeptide binding site for diverse functionalization. The CTPR polymers thus represent a system that could be adapted for synthesis of nano- or microscale assemblies, decorated through binding of peptides with high affinity.

One key step in adapting our system to producing such assemblies is achieving absolute control over specific size and composition. We are in the process of extending the one-pot synthesis into a controllable stepwise manufacture (similar to solid-state peptide synthesis). Here, the C terminus of the first CTPR module is chemically attached by reaction of the fusion protein's intein moiety to a hydrazine azide-functionalized gold chip surface.[19] The cap at the N terminus can then be cleaved to give its reactive thiol and the next module in the chain applied to the chip for the NCL reaction to take place. The process can be repeated automatically with whatever module is required to give a protein polymer of exact size and composition.

One immediate application would stem from the recent interest in the use of TPR proteins to form hydrogels by constructing degenerate sequences of 20 TPR motifs by recombinant approaches.[1g] Our TPR polymerization would offer a route towards facile synthesis of these structures from more tractable short monomer units. With a naturally occurring library of TPR protein scaffolds, each with their specific cognate peptide ligand, there are many opportunities for exploiting these self-assembling functional nanostructures.

Received: May 16, 2012
Revised: October 8, 2012
Published online: December 10, 2012

Keywords: biomaterials · nanostructures · protein engineering · self-assembly · superhelical scaffold

[1] a) J. A. Hubbell, M. P. Lutolf, G. P. Raeber, A. H. Zisch, N. Tirelli, *Adv. Mater.* **2003**, *15*, 888; b) A. Hung, S. Mwenifumbo, M. Mager, J. J. Kuna, F. Stellacci, I. Yarovsky, M. M. Stevens, *J. Am. Chem. Soc.* **2011**, *133*, 1438 – 1450; c) R. Fairman, B. Kokona, A. M. Kim, R. C. Roden, J. P. Daniels, B. J. Pepe-Mooney, B. C. Kovaric, J. C. de Paula, K. A. Johnson, *Biomacromolecules* **2009**, *10*, 1454 – 1459; d) E. H. C. Bromley, R. B. Sessions, A. R. Thomson, D. N. Woolfson, *J. Am. Chem. Soc.* **2009**, *131*, 928 – 930; e) D. N. Woolfson, Z. N. Mahmoud, *Chem. Soc. Rev.* **2010**, *39*, 3464 – 3479; f) T. Scheibel, A. Leal-Egana, *Biotechnol. Appl. Biochem.* **2010**, *55*, 155 – 167; g) T. Z. Grove, C. O. Osuji, J. D. Forster, E. R. Dufresne, L. Regan, *J. Am. Chem. Soc.* **2010**, *132*, 14024 – 14026.

[2] a) E. M. Engelhardt, L. A. Micol, S. Houis, F. M. Wurm, J. Hilborn, J. A. Hubbell, P. Frey, *Biomaterials* **2011**, *32*, 3969 – 3976; b) O. Mahony, O. Tsigkou, C. Ionescu, C. Minelli, L. Ling, R. Hanly, M. E. Smith, M. M. Stevens, J. R. Jones, *Adv. Funct. Mater.* **2010**, *20*, 3835 – 3845.

[3] M. M. Stevens, D. A. Aili, M. Mager, D. Roche, *Nano Lett.* **2011**, *11*, 1401 – 1405.

[4] A. M. Belcher, C. B. Mao, D. J. Solis, B. D. Reiss, S. T. Kottmann, R. Y. Sweeney, A. Hayhurst, G. Georgiou, B. Iverson, *Science* **2004**, *303*, 213 – 217.

[5] a) H. Dong, S. E. Paramonov, J. D. Hartgerink, *J. Am. Chem. Soc.* **2008**, *130*, 13691 – 13695; b) B. Apostolovic, M. Danial, H. A. Klok, *Chem. Soc. Rev.* **2010**, *39*, 3541 – 3575.

[6] A. J. Baldwin, R. Bader, J. Christodoulou, C. E. MacPhee, C. M. Dobson, P. D. Barker, *J. Am. Chem. Soc.* **2006**, *128*, 2162 – 2163.

[7] a) E. F. Banwell, E. S. Abelardo, D. J. Adams, M. A. Birchall, A. Corrigan, A. M. Donald, M. Kirkland, L. C. Serpell, M. F. Butler, D. N. Woolfson, *Nat. Mater.* **2009**, *8*, 596 – 600; b) X. Jia, K. L. Kiick, *Macromol. Biosci.* **2009**, *9*, 140 – 156; c) K. M. Galler, L. Aulisa, K. R. Regan, R. N. D'Souza, J. D. Hartgerink, *J. Am. Chem. Soc.* **2010**, *132*, 3217 – 3223.

[8] E. R. G. Main, S. E. Jackson, L. Regan, *Curr. Opin. Struct. Biol.* **2003**, *13*, 482 – 489.

[9] E. R. G. Main, Y. Xiong, M. J. Cocco, L. D'Andrea, L. Regan, *Structure* **2003**, *11*, 497 – 508.

[10] a) E. R. G. Main, K. Stott, S. E. Jackson, L. Regan, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5721 – 5726; b) K. W. Tripp, D. Barrick, *J. Mol. Biol.* **2007**, *365*, 1187 – 1200.

[11] a) A. L. Cortajarena, L. Regan, *Protein Sci.* **2011**, *20*, 336 – 340; b) Y. Javadi, E. R. Main, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17383 – 17388; c) T. Kajander, A. L. Cortajarena, E. R. G. Main, S. G. J. Mochrie, L. Regan, *J. Am. Chem. Soc.* **2005**, *127*, 10188 – 10190.

[12] J. J. Phillips, Y. Javadi, C. Millership, E. R. Main, *Protein Sci.* **2012**, *21*, 327 – 338.

[13] C. Y. Cheng, V. A. Jarymowycz, A. L. Cortajarena, L. Regan, M. J. Stone, *Biochemistry* **2006**, *45*, 12175 – 12183.

[14] L. D. D'Andrea, L. Regan, *Trends Biochem. Sci.* **2003**, *28*, 655 – 662.

[15] T. Z. Grove, M. Hands, L. Regan, *Protein Eng. Des. Sel.* **2010**, *23*, 449 – 455.

[16] C. Kleanthous, A. M. Krachler, A. Sharma, *Proteins Struct. Funct. Bioinf.* **2010**, *78*, 2131 – 2143.

[17] a) P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776 – 779; b) T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705 – 6710.

[18] a) M. G. Ryadnov, D. N. Woolfson, *J. Am. Chem. Soc.* **2007**, *129*, 14074 – 14081; b) J. D. Hartgerink, *Curr. Opin. Chem. Biol.* **2004**, *8*, 604 – 609; c) J. P. Jung, J. L. Jones, S. A. Cronier, J. H. Collier, *Biomaterials* **2008**, *29*, 2143 – 2151.

[19] P. Yang, S. M. Marinakos, A. Chilkoti, *Langmuir* **2011**, *27*, 1463 – 1471.

