

# Protein-based materials, toward a new level of structural control

Jan C. M. van Hest<sup>\*a</sup> and David A. Tirrell<sup>b</sup>

<sup>a</sup> Department of Organic Chemistry, University of Nijmegen, Toernooiveld 1, 6525ED Nijmegen, The Netherlands. E-mail: vanhest@sci.kun.nl; Fax: +31 24 3653393

<sup>b</sup> Division of Chemistry and Chemical Engineering, 210-41, California Institute of Technology, Pasadena, CA 91125 USA. E-mail: tirrell@Caltech.edu; Fax: +1 (626) 793-8472

Received (in Cambridge, UK) 13th June 2001, Accepted 2nd August 2001

First published as an Advance Article on the web 18th September 2001

Through billions of years of evolution nature has created and refined structural proteins for a wide variety of specific purposes. Amino acid sequences and their associated folding patterns combine to create elastic, rigid or tough materials. In many respects, nature's intricately designed products provide challenging examples for materials scientists, but translation of natural structural concepts into bio-inspired materials requires a level of control of macromolecular architecture far higher than that afforded by conventional polymerization processes. An

increasingly important approach to this problem has been to use biological systems for production of materials. Through protein engineering, artificial genes can be developed that encode protein-based materials with desired features. Structural elements found in nature, such as  $\beta$ -sheets and  $\alpha$ -helices, can be combined with great flexibility, and can be outfitted with functional elements such as cell binding sites or enzymatic domains. The possibility of incorporating non-natural amino acids increases the versatility of protein engineering still

*Jan van Hest was born in 1968 in the Netherlands. After completing his M.Sc. in polymer chemistry in 1991, he conducted his doctoral research at the Eindhoven University of Technology under supervision of Professor Dr Bert Meijer, for which the Ph.D. was granted in 1996. For this work on molecular architectures based on dendrimers he was awarded the first prize of the DSM technology and chemistry competition in 1995 and the SNS award for the best thesis in fundamental research of the Eindhoven University of Technology in 1997.*

*Next he investigated as a post-doctoral researcher the possibilities of protein engineering for the preparation of materials under supervision of Professor Dr D. A. Tirrell, at the University of Massachusetts in Amherst till 1997.*

*He then joined the chemical company DSM, where he worked in the research group New Architectures, first as research scientist and later on as group leader, on the development of innovative material concepts. Specific topics of investigation were controlled radical polymerization and integration of expertise from life sciences and materials.*

*In 2000 he was appointed as full professor at the University of Nijmegen to set up a new group in bio-organic chemistry. His current interests focus on translating natural structural concepts into bio-inspired materials with the aid of state of the art organic and polymeric chemistry techniques and protein engineering.*

*David A. Tirrell is the Ross McCollum–William H. Corcoran Professor and Chairman of the Division of Chemistry and Chemical Engineering at the California Institute of Technology. After earning the B.S. in Chemistry at MIT, Tirrell enrolled in the newly-created Department of Polymer Science and Engineering at the University of Massachusetts, where he worked in the laboratory of Otto Vogl, on the synthesis and polymerization of vinyl derivatives of salicylic acid and 2,4-dihydroxybenzophenone. He was awarded the Ph.D. in 1978. During a four-month stay with Takeo Saegusa at Kyoto University, Tirrell developed methods for the synthesis of ionomers based on polymers of ethyl glycidate. He then accepted an assistant professorship in the Department of Chemistry at Carnegie-*

*Mellon University, where he established research programs directed toward the exploration of neighboring group effects in polymer modification reactions, and toward elucidation of mechanistic questions in radical copolymerization processes.*

*Tirrell returned to Amherst in 1984. He was promoted to Professor in 1987, appointed Director of the Materials Research Laboratory in 1991, and named Barrett Professor of Polymer Science and Engineering in 1992. Tirrell moved to Caltech in 1998. He has served as Visiting Professor at the University of Queensland (1987), at the Institut Charles Sadron in Strasbourg (1991), at the University of Wisconsin-Madison (1995), and at the Institut Curie in Paris. He is a Fellow of the American Institute of Medical and Biological Engineering, and serves on technical advisory boards for Dow Chemical Company and Protein Polymer Technologies, Inc. He was Editor of the Journal of Polymer Science, Part A: Polymer Chemistry from 1988 until 1999, and has served on the editorial boards of many other journals, including Macromolecules, Biopolymers, Accounts of Chemical Research, and Chemical and Engineering News. He chaired the 1994 Gordon Research Conference on Polymers in Biosystems and the 1995 Gordon Conference on Chemistry of Supramolecules and Assemblies.*

*Tirrell's contributions to teaching and research have been recognized in a variety of ways. He was named Outstanding Teacher in the College of Science at Carnegie-Mellon, and a Sloan Fellow in 1982. He was a Presidential Young Investigator Awardee of the National Science Foundation in 1984, and a Fulbright Senior Scholar in 1987. In 1991, he was recognized by the Carl S. Marvel Creative Polymer Chemistry Award of the American Chemical Society, in 1996 he received the Harrison Howe Award of the Rochester Section of the ACS, and in 1997 he was awarded the Chancellor's Medal of the University of Massachusetts. In 2001, he received the American Chemical Society Award in Polymer Chemistry, and the degree of Doctor honoris causa from the Eindhoven University of Technology.*

*Tirrell's most important contributions to chemistry have come in three areas: i). radical copolymerization mechanism, ii). biomimetic membrane chemistry, and iii). development of molecular biological approaches to new polymeric materials.*

further. It is expected that such methods will have large impact in the field of materials science, and especially in biomedical materials science, in the future.

## Introduction

For many years materials scientists have been investigating the possibilities of obtaining higher levels of control in polymer synthesis. Although important progress has been made, especially in recent years with the advance of controlled radical polymerization techniques,<sup>1</sup> the level of control characteristic of biomacromolecules, such as DNA and proteins, remains unsurpassed.

In the nomenclature of polymer chemistry, organisms prepare monodisperse polymers with predetermined chain lengths. But more important is the fact that virtually complete control is achieved over the sequence of the monomeric units, the amino acids. The information stored in the primary sequence results in a three-dimensional folded structure for each protein, which is largely responsible for the most important protein properties. For every specific function, nature has refined protein structures through eons of evolution. It is striking that with the same set of amino acid building blocks such diverse products can be prepared, including globular proteins which function as enzymes as well as fibrous proteins with structural properties, such as collagen. It is the prospect of creating three-dimensional ordered structures by designing the requisite amino acid sequences that motivates much of the current interest in protein-based materials.<sup>2</sup>

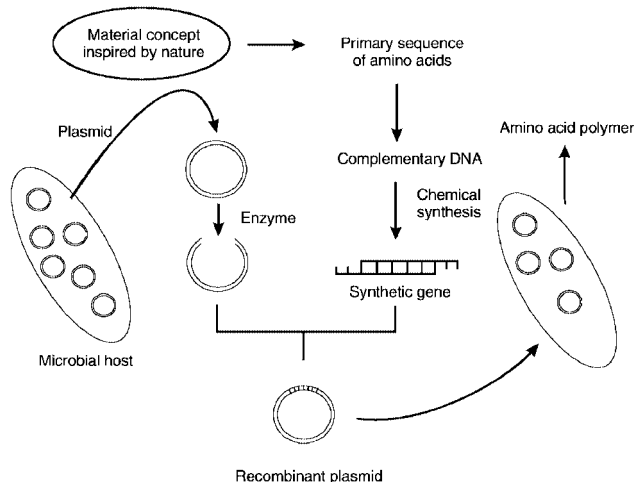
Materials scientists have much to learn from the organizational principles employed in nature, and substantial progress has been made recently in elucidation of the three-dimensional architectures of fibrous proteins such as silks, elastins, and collagens. An important challenge, however, is to translate these concepts into synthetic or bio-inspired materials, which would lead to new kinds of high performance materials.<sup>3</sup> Because of the importance of control at the monomer level, the most promising approach to this target is to use the biosynthetic pathways of (micro)-organisms to synthesize macromolecular materials.

Protein engineering has been used for the design and modification of globular proteins for many years, and standard biological tools have been developed that can now be used to prepare structural proteins.<sup>4</sup> It is remarkable that only in the past decade, and especially in the last 5 years, has this approach to materials design become a serious topic of investigation.

This review is not meant to be comprehensive, but instead will present examples of some of the latest developments in the field of protein engineering directed toward the preparation of high performance materials. After a short explanation of the protein engineering technique, examples will be given of structural proteins found in nature and attempts to use these basic structures for the preparation of new protein-based materials. We will conclude with new methods that allow the scope of protein engineering to be extended to building blocks other than the natural amino acids, developments that promise to make this technique even more versatile in the hands of the materials chemist.

## Protein engineering

Scheme 1 shows the overall process of materials synthesis *via* protein engineering.<sup>5</sup> Often—though not always—based on natural structures, amino acid sequences are chosen to create specific folding patterns and desired material properties. The primary amino acid sequence can then be reverse-translated into a corresponding nucleotide sequence. There are 2 methods of obtaining the needed DNA fragments. One possibility is to clone these sequences from an organism that produces the desired structural protein. The second option is to synthesize



Scheme 1 Overview of the protein engineering methodology.

artificial genes by solid phase techniques. The second method of course allows maximum freedom in designing the target sequence. Because many structural proteins are characterized by repetitive amino acid sequences, it is often possible to multimerize a smaller oligonucleotide sequence to prepare an artificial gene that codes for proteins of high molecular weight.

This multimerization process can be considered as a polycondensation, which therefore results in a set of genes with different lengths. After construction of the artificial gene library, each individual gene is incorporated into circular plasmid DNA, which can be used to transform an appropriate bacterial host. Most commonly the bacterial host *E. coli* is used for these purposes. The plasmids are replicated during every division of the bacterial cells. Because there is only one type of plasmid per cell, screening of the plasmids of individual bacterial colonies leads to isolation of specific artificial genes, each encoding a protein-based material with a specific molecular weight.

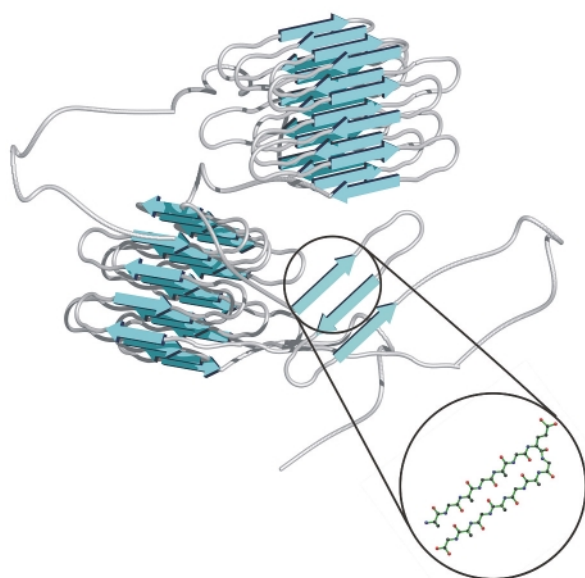
The selected artificial gene is first analyzed to verify its sequence, and then cloned into a second plasmid which allows protein expression. The expression plasmid contains a promoter site for recognition by mRNA polymerase and a switch that regulates protein production from the gene of interest.

The expression plasmid is re-introduced into the bacterial host, and the host cells are allowed to grow to high cell density. During this stage the plasmid switch is turned off in order to prevent protein production from the artificial gene. Premature protein synthesis can be detrimental to cell growth. After sufficient cell density is reached the switch is turned on (a process called induction) and expression of the desired protein begins. Often the synthesis of other cellular proteins is slowed dramatically after induction.

## Spider dragline silk

Among the most thoroughly studied classes of structural proteins is silk.<sup>6</sup> There are many forms of silk, of which that from *Bombyx mori* (Chinese silkworm silk)<sup>7,8</sup> and dragline spider silk of the *Nephila clavipes* (the golden orb weaver) have drawn most attention.<sup>9</sup> Silkworm silk of course is of interest because it is produced in large quantities and used as textile fiber. Spiders produce a large variety of types of silk, each one meant for a specific purpose, such as web construction or trapping of prey. Of all these different types of spider silk, dragline spider silk of *Nephila clavipes* is regarded as nature's high performance fiber, with a remarkable combination of strength and toughness. Values reported for its mechanical properties are an *E*-modulus of 10–50 GPa, elongation to break of 10–30% and tensile strength of 1.1–1.4 GPa. These values are

dependent on moisture content and strain rate. The effect of humidity is of utmost importance. Silk exposed to water can shrink to less than half of its original length, and the supercontracted fibers then show rubberlike behavior.<sup>9</sup> Although this is a desirable feature under natural conditions, because it allows spider webs to reshape in the dewy evening, it is less practical when one wants to apply this material as a high performance fiber. The large variation in properties is a result of the difficulty in maintaining constant conditions under which spiders produce their silk. In recent years important insight has been obtained into the three-dimensional structures of the different types of silk. X-ray diffraction,<sup>10</sup> state of the art NMR techniques<sup>11,12</sup> and IR and Raman spectroscopy<sup>13</sup> have helped to unravel the architectural reasons why this class of structural proteins has its remarkable mechanical properties. Silks from both silkworms and spiders contain repetitive sequences of crystalline and amorphous domains (Fig. 1). The crystalline



**Fig. 1** Schematic representation of the organization of amorphous and crystalline domains in silk fibers. Silk has inspired protein engineers to construct  $\beta$ -sheet materials. A detail of a silk-like peptide sequence  $((AG)_3EG)_n$ <sup>21,22</sup> is encircled.

domains are responsible for the strength of the material, whereas the amorphous protein matrix: i) allows the crystalline domains to orient under strain to increase the strength of the material, and ii) introduces flexibility to increase the energy to break.

The crystalline domains comprise highly repetitive amino acid sequences consisting of alanine rich  $\beta$ -sheets. In *Bombyx mori* silk the most important repeat sequence is (Ala Gly); for *N. clavipes* dragline spider silk it is predominantly poly Ala.

In contrast to silkworm silk, there is no readily available source for spider dragline silk, and protein engineering techniques have been explored as a means of scaling up production. Furthermore, changes in the silk structure might be made to produce a material less sensitive to water. The combination of bacterial production with a synthetic processing step should allow better control of mechanical properties as compared to natural silk. On the other hand, development of methods of posttranslational silk spinning is not a straightforward process.

The first protein engineering attempts were made by using large parts of the native spider silk genes. After incorporation of these cDNA fragments into *E. coli* expression systems however, difficulties arose from gene and protein stability, and truncated proteins were produced. Clones larger than 2.5 kb proved to be unstable.<sup>14,15</sup> The same behavior was observed for native silkworm silk genes. An explanation for this phenomenon may lie in the highly repetitive nature of these sequences, which

renders the coding sequence susceptible to genetic recombination. Differences in codon preferences between the silkworm and the bacterial host could also reduce expression efficiency.<sup>16</sup>

The first attempts using synthetic genes based on the native gene sequence were reported as early as 1990 by Cappello and Ferrari,<sup>17</sup> who used silkworm fibroin analogues. Better results were obtained than when the cDNA genes were used, although the levels of protein production remained relatively low, especially for high molecular weight variants.<sup>18</sup> The best results reported thus far with *E. coli* expression systems were obtained by Winkler *et al.* for spider dragline silk (25–40 mg L<sup>-1</sup> cell culture)<sup>19</sup> and Krejchi *et al.* for an (AG)<sub>3</sub>EG<sup>20</sup> repetitive material.<sup>21,22</sup> To increase protein yield, high cell density production was performed, and allowed preparation of multi-gram quantities of product (25 g protein/35 L culture).<sup>23</sup> Another approach to this problem has involved changes in the production host. O'Brien *et al.* have shown that with the yeast *Pichia pastoris* it was possible to produce spider silk-like material in larger quantity (1 g L<sup>-1</sup>).<sup>24</sup> Even mammalian hosts are now being investigated, such as Nexia's transgenic goats, which produce milk containing spider silk protein.<sup>25</sup>

Protein production in large quantity is just one of the problems related to obtaining silk-like fibers. An equally difficult aspect is spinning of the polypeptide, to create a material with the correct three-dimensionally ordered structure. Recently, using micro-spinneret techniques, Jelinski *et al.* have succeeded in processing regenerated *Bombyx mori* silk fibers that show similar properties when compared to their native counterparts.<sup>26,27</sup>

Protein engineering can also play a role in improving processing of silk like materials. Introduction of methionine residues flanking the  $\beta$ -sheet domains creates a tool for chemical control of structural organization.<sup>28</sup> By oxidizing the methionine units,  $\beta$ -sheet formation was prevented, which also increased solubility, while sheet formation was restored after reduction. Enzymatic techniques have also been developed to induce this transition. The functional oligopeptide RGYSLG is a recognition site for cyclic AMP-dependent kinase, which phosphorylates the hydroxy function of the serine residue.<sup>19</sup> Introduction of this recognition site into the structural protein allowed Winkler *et al.* to prevent  $\beta$ -sheet formation upon phosphorylation. Sheet formation was restored after dephosphorylation.

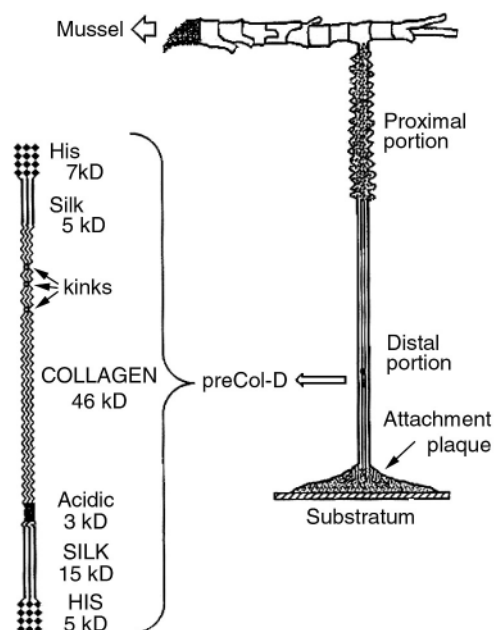
Recent insights into the mechanical properties of silk<sup>29</sup> have shown that it is not necessary or desirable to prepare purely  $\beta$ -sheet sequences; a certain amount of disorder should also be introduced to allow reorientation of the  $\beta$ -sheet crystalline domains. With protein engineering this could be accomplished by producing multiblock copolymers of  $\beta$ -sheet structures flanked by random oligopeptides. Such block copolymers have been reported by Conticello *et al.*, who synthesized multiblock copolymers consisting of alanine-rich blocks and elastin-mimetic domains.<sup>30</sup> This method leads to good protein yield (50 mg L<sup>-1</sup> culture) and a method for controlling  $\beta$ -sheet formation by concentration from an aqueous solution.

With respect to silk-like polymers, protein engineering is probably the only synthetic method that will allow high molecular weight materials to be prepared in sufficient quantity for materials applications. The ability to introduce switches for  $\beta$ -sheet formation as well as the production of multiblock copolymers will ultimately lead to materials with predictable processing behavior and mechanical properties.

## Mussel byssus thread

Another masterpiece of a structural material found in nature is the mussel byssus thread. This thread is used to attach the mussel to a surface, and must be tough enough to withstand challenging marine conditions. Extensive investigations by

Waite and coworkers have shown that the thread consists predominantly of a gradient of two pentablock copolymers, which are held together by so-called dovetail proteins, and which provide a graded distribution of strength and structure (Fig. 2).<sup>31–33</sup>



**Fig. 2** Schematic view of the mussel byssus thread of *M. edulis*. The enlargement depicts the pentablock copolymer containing  $\beta$ -sheet domains. Reprinted with permission from (33), Copyright 1997 The American Society for Biochemistry and Molecular Biology.

The polymer that is predominant at the proximal end of the thread (*i.e.* closest to the mussel tissue) consists of a central collagen block flanked by 2 elastic domains and 2 polyhistidine sequences. The latter are expected to give rise to metal complexation between the different peptide structures. The elastin fragments introduce flexibility into the structure, whereas the collagen part is responsible for stiffness. Further down the byssus, and therefore closer to the substratum to which the mussel is attached, the elastin fragments are replaced by silk-like  $\beta$ -sheet domains. This structure therefore becomes much more rigid. The entire byssus is built up by a mussel within 5 minutes, and is about five times tougher and more extensible than Achilles tendon. After processing the proteins are also covalently connected. The complexity of a simple mussel byssus thread is thus far greater than that characteristic of materials produced by present-day synthetic polymer processing techniques.

Although this material would be an ideal target for protein engineering, no efforts toward this end have been reported. A possible reason could be the high content of collagen in these structures. Collagens are the most abundant family of structural proteins in the human body, and are an important constituent of ligaments, cartilage and bone. They are generally characterized by the repetitive sequence GX<sub>n</sub>Y, in which X is often proline and Y often hydroxyproline. Collagens form triple helices, which provide critical building blocks for higher ordered aggregates.<sup>34</sup> Because of their strong tendency to self-assemble, collagens become insoluble very quickly. The few protein engineering attempts reported thus far have failed for this (or perhaps other) reasons.<sup>35</sup> The need for a protein engineering approach for collagen is also not as pressing as for spider silk, because collagen is available in large quantities from mammalian sources. On the other hand, the repetitive sequence of collagen may be of value as an element in designed peptide-based materials, as shown by the example of the mussel byssus thread.

The mussel byssus contains more surprising peptides. The tendons have to be attached to the substrate under aqueous conditions. Nature therefore has developed a series of extraordinary protein glues, which work even in saline media. Careful analysis of these proteins, which must be done under conditions where the protein hasn't formed a network, shows that the adhesive proteins are rich in proline, tyrosine and lysine.<sup>36</sup> Furthermore, proline and tyrosine are posttranslationally transformed into 4-hydroxyproline and L-3,4-dihydroxyphenylalanine (DOPA), respectively. The DOPA units are expected to contribute to crosslinking *via* radical processes, in order to form networks and assist in adhesion.<sup>37</sup> Researchers intrigued by this phenomenon have prepared synthetic genes and expressed the corresponding (AKPSYPPTYK)<sub>n</sub> polypeptides.<sup>38</sup> These model peptides were treated with mushroom tyrosinase to convert the tyrosine units into DOPA, and thereby to induce crosslinking. Because expression yields were low, insufficient material was produced for a thorough evaluation. Furthermore, the question remains whether protein engineering is necessary in this case. Deming *et al.* have used NCA polymerization techniques to prepare copolypeptides consisting of lysine and DOPA residues to investigate the adhesive potential of these structures.<sup>39</sup> Based on the positive results obtained from this research, one can conclude that at least some of the key adhesive properties do not require the specific protein sequence; only the presence of the functional amino acids DOPA and hydroxyproline is essential. For such materials it will be more convenient to pursue the (improved) NCA polymerization routes than protein engineering.<sup>40</sup> More subtle effects of sequence may become apparent upon further analysis of the structure and properties of the adhesive proteins.

## Elastin

One of the most versatile materials, especially from a protein engineering point of view is elastin. Elastin is a connective tissue protein that provides a combination of strength and flexibility in the extracellular matrix. Elastins are characterized by simple repeating sequences, and in contrast to silks or collagens, this family of materials does not suffer from significant solubility problems. At low temperatures elastin-like polypeptides are highly soluble in aqueous solutions and adopt random coil conformations. Elastins, however, display lower critical solution temperature (LCST) behavior, which means that the proteins form ordered structures upon raising temperature. The loss of entropy of the protein chain, when the structure folds itself into a  $\beta$ -helix conformation, is compensated by the release of water from the chain. A thoroughly investigated series of elastin proteins is based on the pentapeptide sequence VPGVG.<sup>41–43</sup> The structure and function of elastins are maintained as long as the glycine and proline residues are present, and the second valine residue can be replaced by any other amino acid. This property provides a means to control the temperature of the phase transition. When a more hydrophobic unit, such as tryptophan, replaces valine, the LCST drops significantly, and the transition temperature can be raised by introduction of more polar moieties such as glutamic acid. In the latter case, the phase transition is also made pH dependent. The pioneering work of Urry has shown that changing hydrophobicity leads to a broad spectrum of transition temperatures. The phase transition can also be induced electrochemically. For this purpose nicotinamide adenine nucleotide (NAD) is attached to glutamic acid residues within the peptide sequence. Reduction of NAD to NADH creates a more hydrophobic domain and therefore a lower transition temperature. Elastin materials have been made both by organic chemical techniques, and by microbial protein expression. Although it has been shown that one repeat unit of VPGVG is sufficient to allow the transition from random coil into an ordered  $\beta$ -helix structure,<sup>44</sup> for useful materials properties

higher molecular weight polymers are required. Protein engineering is ideally suited for this purpose, especially because production yields in bacteria are high (50–60 mg L<sup>-1</sup> culture) when compared to other structural proteins.

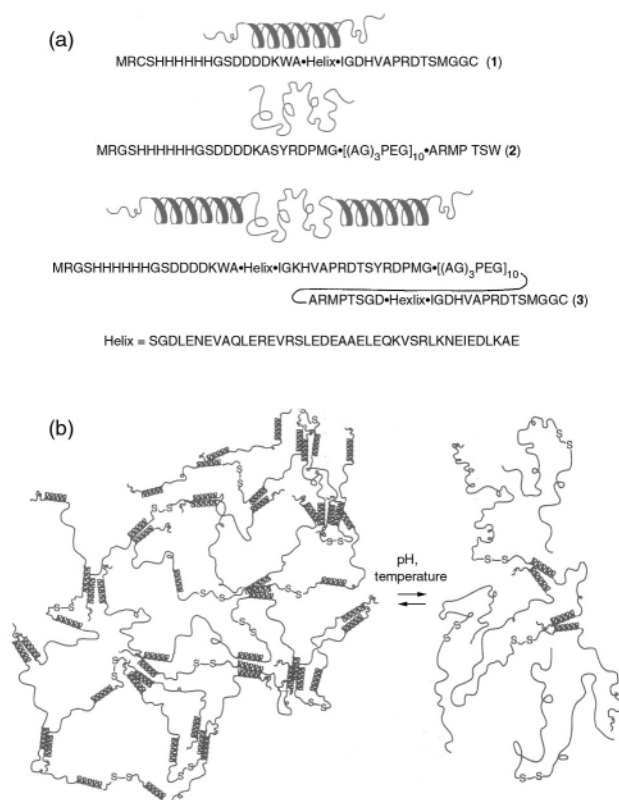
Following up on earlier work by Urry, Conticello *et al.* have used protein engineering techniques to synthesize elastin-based block copolymers consisting of hydrophilic and hydrophobic blocks.<sup>45</sup> Protein engineering allows complete control of the composition, sequence and length of the blocks. By careful design, the critical micellization temperatures of these polymers could be controlled, and the resulting materials are under investigation as candidates for drug delivery. A second demonstration of the versatility of protein engineering for the preparation of elastin-based materials has been reported by the same group. Hydrogels are important products in biomedical applications. Ideally one wants to obtain control of mechanical properties by regulating crosslink density and network formation. Using protein engineering, control over monomer composition, and therefore control over distribution of crosslinkable moieties, is achieved easily. Conticello *et al.* synthesized elastin-based materials containing one specific crosslinking point (lysine) per 25 monomer units, allowing a high level of control of crosslink density after reaction with bis(sulfosuccinimidyl) suberate.<sup>46</sup>

Another approach, which takes advantage of additional possibilities of protein engineering was developed by Urry and extended by Panitch *et al.*, who reported new materials for tissue-engineered vascular grafts.<sup>47,48</sup> Oligopeptides were introduced that contained either RGD (Urry) or REDV (Panitch) domains, which are known to stimulate endothelial cell binding.<sup>49</sup> This approach yields matrix materials that exhibit mechanical properties similar to those of the arterial wall and that support adhesion of vascular endothelial cells. An extension of the polymer design was made by introduction of lysine moieties, which allowed amine specific crosslinking of the elastin matrix without disturbing the cell binding domains.<sup>50</sup> The combination of structure and function in a well defined manner as demonstrated by this example is almost impossible to achieve by any method other than protein engineering.

## Helical polypeptides

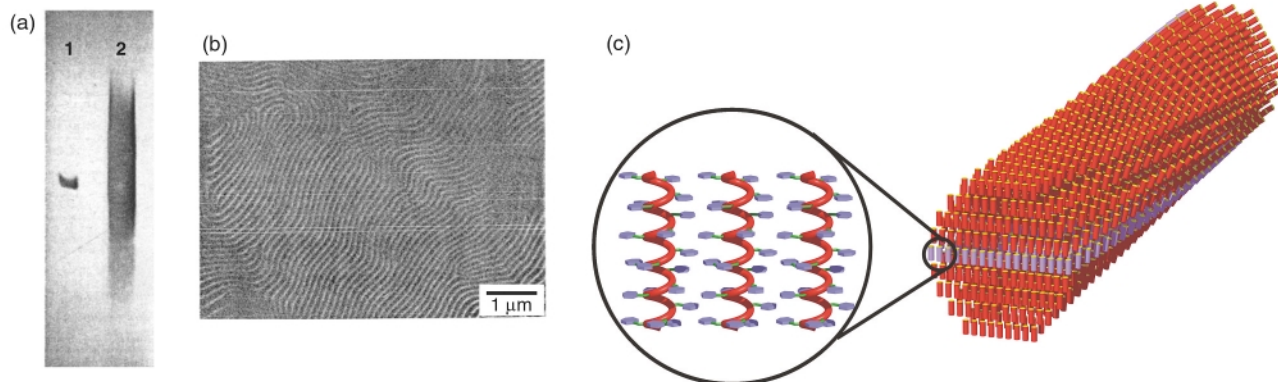
Leucine zipper peptides play key roles in the dimerization and DNA-binding behavior of transcriptional regulatory proteins. Related structures can be found within the keratins, the main structural protein of hair and nails. The leucine zipper motif is characterized by a consensus heptad repeat (*abcdcfg*), in which *a* and *d* are hydrophobic amino acids, (*d* most frequently leucine) and the moieties at *e* and *g* are usually charged. In the helical form of the zipper peptide, the residues at *a* and *d* are arrayed along a single face of the helix. Exposure of this hydrophobic face to the aqueous environment is reduced by

formation of a dimeric coiled-coil structure, while electrostatic interactions between *e* and *g* residues modulate the stability of the dimer. This feature has been exploited to prepare pH dependent peptide-based hydrogels. Triblock copolymers were constructed to comprise a central random coil (AG<sub>3</sub>PEG)<sub>*x*</sub>, flanked by terminal leucine zipper domains (Fig. 3).<sup>51</sup> These



**Fig. 3** (a) Amino acid sequence of the leucine zipper triblock copolymer; (b) demonstration of reversible hydrogel formation. Reprinted with permission from W. A. Petka, J. L. Harden, K. P. McGrath, D. Wirtz and D. A. Tirrell, *Science*, 1998, **281**, 389. Copyright 1998 American Association for the Advancement of Science.

domains consisted of six heptad repeats of which the choice of *a/d* residues was based on the *Jun* oncogene product. A database developed by Lupas *et al.*<sup>52</sup> was used for the selection of the residues occupying positions *b*, *c* and *f*. Nine of the 12 *e* and *g* positions were occupied by Glu residues, which allowed switching between the aggregated and dissociated states by changes in pH or temperature. At low pH and low temperature the materials formed elastic gels; increases in pH or in temperature resulted in decreased viscosity and loss of hydrogel properties.



**Fig. 4** (a) Difference in molecular weight distribution of protein engineered poly L-glutamic acid (1) and a conventionally prepared sample (2). Reprinted by permission from *Nature* (54), Copyright 1997 Macmillan Magazines Ltd.; (b) and (c) monodisperse poly-benzyl-L-glutamate shows a twisted smectic liquid crystalline phase. Reprinted with permission from (55), Copyright 1998 American Chemical Society.

Each of the examples examined thus far shows the need for control over composition in order to engineer the properties of the products. Molecular weight distribution is a second parameter that is of importance in polymer chemistry, though its influence on material properties is not always easy to anticipate. One striking exception however, is the synthesis of the helical polypeptide polybenzyl-L-glutamate (pBLG) (Fig. 4).

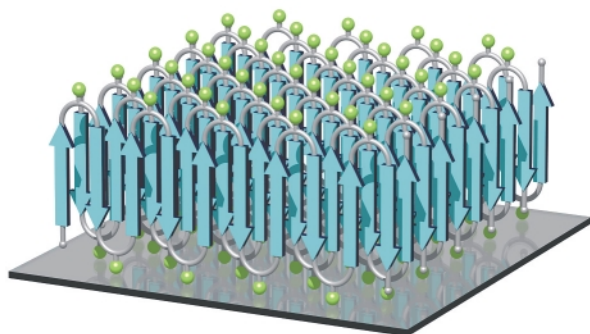
The polydisperse form, which can be prepared by NCA polymerization, has been known for many years to exhibit liquid crystalline behavior.<sup>53</sup> Under most conditions, pBLG forms a cholesteric phase, but certain solvents and temperatures yield a nematic. Because of the broad distribution of molecular weight a clear characterization has always been difficult. Using protein engineering, monodisperse poly L- $\alpha$ -glutamic acid could be synthesized, and posttranslationally modified to yield pBLG. This monodisperse sample showed distinctly different LC behavior. Instead of a nematic or cholesteric phase a twisted smectic mesophase was obtained.<sup>54,55</sup> The fact that all chains were of the same length resulted in ordering in layers, which was not possible in conventional polydisperse samples.

### Increasing the scope of protein engineering

The foregoing examples illustrate some of the promise of protein engineering as a tool for material synthesis. Especially in cases where the information encoded in the primary amino acid sequence governs higher ordered structures, and thereby the ultimate material properties, protein engineering clearly outperforms synthetic polymerization techniques.

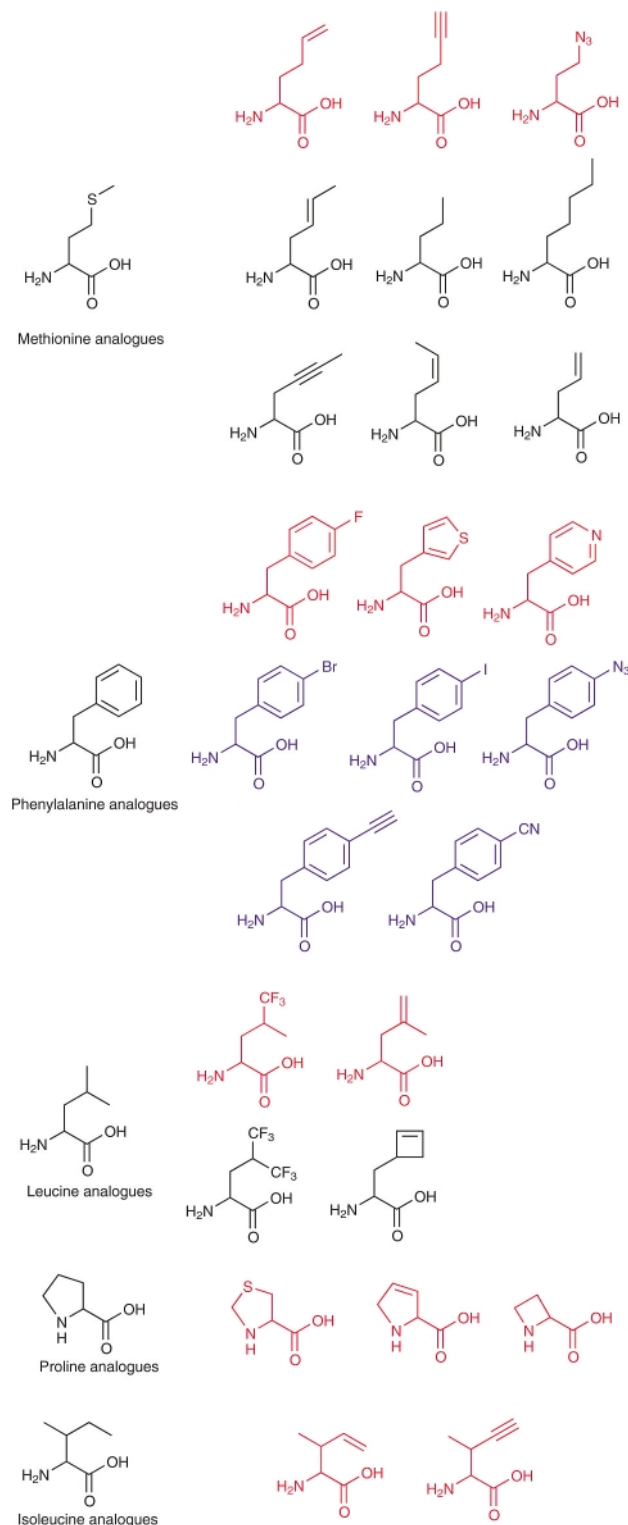
One of the drawbacks of protein engineering in comparison to the other techniques, however, is that the choice of building blocks might be expected to be limited to the naturally occurring amino acids. Many functional groups, which would be useful in controlling properties or postmodification efforts, such as alkenes, alkynes, or halogens, are not found in natural proteins. The chemical versatility of amino acids containing unsaturated side chains has recently been reviewed.<sup>56</sup> Furthermore, incorporation of halides, especially fluoride, can have dramatic effects on the physical properties of materials. Extension of the set of amino acid building blocks to unnatural analogues that can be utilized by the protein production mechanism of living systems would therefore enlarge the possibilities of protein-based materials.

This idea has intrigued several groups. A first approach is to use stop (nonsense) codons or 4-base codons<sup>57</sup> for the introduction of unnatural amino acids. Suppressor tRNAs can be found that overwrite these termination signals. If unnatural amino acids can be coupled to these tRNA molecules, efficient incorporation of analogues can be accomplished. This technique has been applied *in vitro* for more than a decade.<sup>58</sup> For this approach to succeed *in vivo* it is necessary to introduce a suppressor tRNA that is not recognized by the host enzymes that couple amino acids to tRNAs (the aminoacyl-tRNA synthetases).



**Fig. 5** Schematic representation of the lamellar crystalline phase formed by the peptide sequence  $((AG)_3FG)_n$ , in which phenylalanine is replaced by *p*-fluorophenylalanine. The green spots indicate amino acid side chains with unnatural functionality, in this case fluorine, at the lamellar surface.

tases).<sup>59,60</sup> This can be achieved by using a suppressor tRNA from an exogenous source such as yeast.<sup>61,62</sup> Next, a synthetase must be identified or developed to couple the unnatural amino acid to the suppressor tRNA. This synthetase should also be inactive toward other tRNAs. This method allows site-directed insertion of the amino acid analogue at a single position, *e.g.* in the active site of an enzyme. This method will have major impact on the design of proteins that are used for functional purposes



**Fig. 6** An overview of the obtained results of Tirrell and coworkers with respect to *in vivo* replacement of natural amino acids by their analogues. Amino acids depicted in red are incorporated *via* standard auxotroph methodologies. Mutant aminoacyl tRNA synthetases have to be applied to be able to build in the blue amino acids. Analogues depicted in black are incorporated after boosting synthetase activity.

such as enzymes, since alteration of the overall physical properties of the macromolecule generally cannot be achieved through modification of a single site. An overview of this technique has been published recently.<sup>63</sup>

An alternative method has been developed by Tirrell *et al.*, based on much earlier work on substitution of natural amino acids by their close structural analogues. This method uses so-called bacterial auxotrophs, *i.e.* bacteria that have lost the ability to produce one of the natural amino acids. Such bacterial strains are dependent on the growth medium for obtaining this specific amino acid. If an analogue is added to the medium, it can be incorporated in place of the natural substrate, if it is recognized by the corresponding aminoacyl-tRNA synthetase.

In order to reduce toxicity, the cells are grown first in a medium containing all of the natural amino acids. When sufficient cell density is reached, the cells are shifted to a new medium which contains the analogue instead of the natural amino acid.

This methodology has proven successful for a large number of amino acids and derivatives thereof. Though the first publications go back for several decades,<sup>64</sup> only in recent years has extensive research allowed replacement of phenylalanine, proline, leucine, methionine, tyrosine and isoleucine with their respective analogues.<sup>65–68</sup>

An interesting application of this method is the incorporation of phenylalanine analogues into the  $\beta$ -sheet element ((AG)<sub>3</sub>FG)<sub>n</sub>. This sequence adopts a lamellar crystalline structure, in which the phenylalanine residues are exposed to the surface. Well-defined functional surfaces can now be created. Replacement by *p*-fluorophenylalanine allowed the construction of fluorinated crystalline layers,<sup>65</sup> and  $\beta$ -sheets containing thiophene units from 3-thienylalanine<sup>66</sup> can be regarded as precursors for conducting polymers (Fig. 5).

Replacement of leucine by either trifluoroleucine or 2-amino-4-methylhex-4-enoic acid in leucine zippers causes marked effects on conformational stability. For example, in a leucine zipper peptide characterized by a melting temperature of 48 °C, the transition temperature was elevated to 61 °C by introduction of trifluoroleucine, whereas the native structure was destabilized by the introduction of the unsaturated analogue.<sup>69</sup> This simple example illustrates the extent to which incorporation of amino acid analogues expands the possibilities for engineering protein stability and protein–protein interactions.<sup>70</sup>

The most systematic study to date has concerned introduction of methionine analogues. Homoallylglycine and homopropargylglycine are especially effective as methionine surrogates.<sup>68,71</sup> Although *trans*-crotylglycine is recognized by the methionyl-tRNA synthetase, increasing the synthetase activity of the host is required for protein production.<sup>72,73</sup> The possibility of using mutant synthetases has also been demonstrated for *p*-bromophenylalanine<sup>74</sup> and azatyrosine.<sup>75</sup>

With the methods described in this section, some of the natural limitations of protein engineering are beginning to fade. Tailor-made materials can now be designed with pre-defined structures and properties (Fig. 6). Furthermore, protein engineering is unique in combining structural with functional elements, which can be either of natural origin, such as cell-binding domains, or introduced by incorporation of unnatural amino acids. The versatility and control of protein engineering will make this an important synthetic approach to a new generation of biomaterials.

## Acknowledgement

The authors would like to thank Jurry Hannink for his assistance with the schematic representations and front cover design.

## Notes and references

1 *Controlled/Living polymerization*, ed. K. Matyjaszewski, ACS Symposium series 768, Oxford University Press, Washington DC, 2000.

- 2 *Protein-based materials*, ed. K. McGrath and D. L. Kaplan, Birkhäuser, Boston, 1997.
- 3 T. J. Deming, *Adv. Mater.*, 1997, **9**, 299.
- 4 J. Sambrook, E. F. Fritsch and T. Maniatis, *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1986.
- 5 D. A. Tirrell, M. J. Fournier and T. L. Mason, *MRS Bull.*, 1991, **16**, 23.
- 6 D. L. Kaplan, *Polymer Degrad. Stab.*, 1998, **59**, 25.
- 7 J. Perez-Rigueiro, C. Viney, J. Llorca and M. Elices, *J. Appl. Polym. Sci.*, 1998, **70**, 2439.
- 8 S. W. Watt, I. J. McEwen and C. Viney, *Macromolecules*, 1999, **32**, 8671.
- 9 S. B. Warner, M. Polk and K. Jacob, *J. Macromol. Sci., Rev. Macromol. Chem. Phys.*, 1999, **C39**, 643.
- 10 C. Riekel, M. Muller and F. Vollrath, *Macromolecules*, 1999, **32**, 4464.
- 11 T. Asakura, T. Ito, M. Okudaira and T. Kameda, *Macromolecules*, 1999, **32**, 4940.
- 12 L. W. Jelinski, *Science*, 1996, **273**, 1481.
- 13 Z. Shao, F. Vollrath, J. Sirichaisit and R. J. Young, *Polymer*, 1999, **40**, 2493.
- 14 R. V. Lewis, M. Hinman, S. Kothakota and M. J. Fournier, *Protein Expression Purif.*, 1996, **7**, 400.
- 15 S. Arcidiacono, C. Mello, D. Kaplan, S. Cheley and H. Bayley, *Appl. Microbiol. Biotechnol.*, 1998, **49**, 31.
- 16 S. R. Fahnestock and S. L. Irwin, *Appl. Microbiol. Biotechnol.*, 1997, **47**, 23.
- 17 J. Cappello, J. Crissman, M. Dorman, M. Mikolajczak, G. Textor, M. Marquet and F. Ferrari, *Biotechnol. Progr.*, 1990, **6**, 198.
- 18 J. T. Prince, K. P. McGrath, C. M. Digirolamo and D. L. Kaplan, *Biochemistry*, 1995, **34**, 10879.
- 19 S. Winkler, D. Wilson and D. L. Kaplan, *Biochemistry*, 2000, **39**, 14002.
- 20 Amino acids are being described by a one letter code; A: alanine; G: glycine; E: glutamic acid; R: arginine; Y: tyrosine; S: serine; K: lysine; P: proline; T: threonine; V: valine; D: aspartic acid; F: phenylalanine.
- 21 M. T. Krejchi, E. D. T. Atkins, A. J. Waddon, M. J. Fournier, T. L. Mason and D. A. Tirrell, *Science*, 1994, **265**, 1427.
- 22 M. T. Krejchi, S. J. Cooper, Y. Deguchi, E. D. T. Atkins, M. J. Fournier, T. L. Mason and D. A. Tirrell, *Macromolecules*, 1997, **30**, 5012.
- 23 A. Panitch, K. Matsuki, E. J. Cantor, S. J. Cooper, E. D. T. Atkins, M. J. Fournier, T. L. Mason and D. A. Tirrell, *Macromolecules*, 1997, **30**, 42.
- 24 J. P. O'Brien, S. R. Fahnestock, Y. Termonia and K. C. H. Gardner, *Adv. Mater.*, 1998, **10**, 1185.
- 25 H. T. Huynh, G. Robitaille and J. D. Turner, *Exp. Cell Res.*, 1991, **197**, 191.
- 26 O. Liivak, A. Blye, N. Shah and L. W. Jelinski, *Macromolecules*, 1998, **31**, 2947.
- 27 A. Seidel, O. Liivak, S. Calve, J. Adaska, G. D. Ji, Z. T. Yang, D. Grubb, D. B. Zax and L. W. Jelinski, *Macromolecules*, 2000, **33**, 775.
- 28 S. Winkler, S. Szela, P. Avtges, R. Valluzzi, D. A. Kirschner and D. Kaplan, *Int. J. Biol. Macromol.*, 1999, **24**, 265.
- 29 L. W. Jelinski, *Curr. Opin. Solid State Mater. Sci.*, 1998, **3**, 237.
- 30 Y. Qu, S. C. Payne, R. P. Apkarian and V. P. Conticello, *J. Am. Chem. Soc.*, 2000, **122**, 5014.
- 31 K. J. Coyne, X. X. Qin and J. H. Waite, *Science*, 1997, **277**, 1830.
- 32 K. J. Coyne and J. H. Waite, *J. Exp. Biol.*, 2000, **203**, 1425.
- 33 X. X. Qin, K. J. Coyne and J. H. Waite, *J. Biol. Chem.*, 1997, **272**, 32623.
- 34 M. Vanderrest and R. Garrone, *Biochimie*, 1990, **72**, 473.
- 35 I. Goldberg, A. J. Salerno, T. Patterson and J. I. Williams, *Gene*, 1989, **80**, 305.
- 36 J. H. Waite, *Bioartif. Organs II*, 1999, **875**, 301.
- 37 M. E. Yu, J. Y. Hwang and T. J. Deming, *J. Am. Chem. Soc.*, 1999, **121**, 5825.
- 38 M. Kitamura, K. Kawakami, N. Nakamura, K. Tsumoto, H. Uchiyama, Y. Ueda, I. Kumagai and T. Nakaya, *J. Polym. Sci., Part A: Polym. Chem.*, 1999, **37**, 729.
- 39 M. E. Yu and T. J. Deming, *Macromolecules*, 1998, **31**, 4739.
- 40 T. J. Deming, *Nature*, 1997, **390**, 386.
- 41 D. W. Urry, *Biopolymers*, 1998, **47**, 167.
- 42 D. W. Urry, S. Q. Peng, L. C. Hayes, D. McPherson, J. Xu, T. C. Woods, D. C. Gowda and A. Pattanaik, *Biotechnol. Bioeng.*, 1998, **58**, 175.
- 43 D. W. Urry, *J. Phys. Chem. B*, 1997, **101**, 11007.
- 44 H. Reiersen, A. R. Clarke and A. R. Rees, *J. Mol. Biol.*, 1998, **283**, 255.
- 45 T. A. T. Lee, A. Cooper, R. P. Apkarian and V. P. Conticello, *Adv. Mater.*, 2000, **12**, 1105.
- 46 R. A. McMillan and V. P. Conticello, *Abstr. Pap. Am. Chem. Soc.*, 2000, **219**, 583.

- 47 A. Nicol, D. C. Gowda, T. M. Parker and D. W. Urry, *Biotechnology of Bioactive Polymers*, ed. C. C. Gebelein and C. E. Carraher, Plenum Press, New York, 1994.
- 48 A. Panitch, T. Yamaoka, M. J. Fournier, T. L. Mason and D. A. Tirrell, *Macromolecules*, 1999, **32**, 1701.
- 49 M. J. Humphries, S. K. Akiyama, A. Komoriya, K. Olden and K. M. Yamada, *J. Cell. Biol.*, 1986, **103**, 2637.
- 50 E. R. Welsh and D. A. Tirrell, *Biomacromolecules*, 2000, **1**, 23.
- 51 W. A. Petka, J. L. Harden, K. P. McGrath, D. Wirtz and D. A. Tirrell, *Science*, 1998, **281**, 389.
- 52 A. Lupas, M. V. Dyke and J. Stock, *Science*, 1991, **252**, 1162.
- 53 C. Robinson and J. C. Ward, *Nature*, 1957, **180**, 1183.
- 54 S. J. M. Yu, V. P. Conticello, G. H. Zhang, C. Kayser, M. J. Fournier, T. L. Mason and D. A. Tirrell, *Nature*, 1997, **389**, 167.
- 55 S. J. He, C. Lee, S. P. Gido, S. M. Yu and D. A. Tirrell, *Macromolecules*, 1998, **31**, 9387.
- 56 F. Rutjes, L. B. Wolf and H. E. Schoemaker, *J. Chem. Soc., Perkin Trans. 1*, 2000, 4197.
- 57 T. Hoshaka, D. Kajihara, Y. Ashizuka, H. Murakami and M. Sisido, *J. Am. Chem. Soc.*, 1999, **121**, 34.
- 58 C. J. Noren, S. J. Anthonycahill, M. C. Griffith and P. G. Schultz, *Science*, 1989, **244**, 182.
- 59 L. Wang, T. J. Magliery, D. R. Liu and P. G. Schultz, *J. Am. Chem. Soc.*, 2000, **122**, 5010.
- 60 D. R. Liu, T. J. Magliery, M. Pasternak and P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.*, 1997, **94**, 10092.
- 61 R. Furter, *Protein Sci.*, 1998, **7**, 419.
- 62 L. Wang, A. Brock, B. Heberic and P. G. Schultz, *Science*, 2001, **292**, 498.
- 63 D. A. Dougherty, *Curr. Opin. Chem. Biol.*, 2000, **4**, 645.
- 64 M. H. Richmond, *Bacteriol. Rev.*, 1962, **26**, 398.
- 65 E. Yoshikawa, M. J. Fournier, T. L. Mason and D. A. Tirrell, *Macromolecules*, 1994, **27**, 5471.
- 66 S. Kothakota, T. L. Mason, D. A. Tirrell and M. J. Fournier, *J. Am. Chem. Soc.*, 1995, **117**, 536.
- 67 T. J. Deming, M. J. Fournier, T. L. Mason and D. A. Tirrell, *J. Macromol. Sci., Pure Appl. Chem.*, 1997, **A34**, 2143.
- 68 J. C. M. van Hest, K. L. Kiick and D. A. Tirrell, *J. Am. Chem. Soc.*, 2000, **122**, 1282.
- 69 Y. Tang, G. Ghirlanda, N. Vaidehi, J. Kua, D. T. Mainz, W. A. Goddard III, W. F. DeGrado and D. A. Tirrell, *Biochemistry*, 2001, **40**, 2790.
- 70 C. E. MacPhee and C. M. Dobson, *J. Am. Chem. Soc.*, 2000, **122**, 12707.
- 71 J. C. M. van Hest and D. A. Tirrell, *FEBS Lett.*, 1998, **428**, 68.
- 72 K. L. Kiick, J. C. M. van Hest and D. A. Tirrell, *Angew. Chem., Int. Ed.*, 2000, **39**, 2148.
- 73 K. L. Kiick and D. A. Tirrell, *Tetrahedron*, 2000, **56**, 9487.
- 74 N. Sharma, R. Furter, P. Kast and D. A. Tirrell, *FEBS Lett.*, 2000, **467**, 37.
- 75 F. Hamano-Takaku, T. Iwama, S. Saito-Yano, K. Takaku, Y. Monden, M. Kitabatake, D. Soll and S. Nishimura, *J. Biol. Chem.*, 2000, **275**, 40324.