6

Folding of β-Structured Fibrous Proteins and Self-Assembling Peptides

Anna Mitraki and Mark J. van Raaij

Summary

Natural fibrous proteins include families found in natural materials such as wool and silk; in tissue components such as collagen and elastin; or in virus and bacteriophage adhesins. They have long fascinated scientists and engineers because of their mechanical and elastic properties, and considerable efforts have been made in order to produce artificial materials inspired from these natural proteins. The understanding of their structure, folding, and assembly properties is necessary in order to achieve these objectives. However, because of their complexity, structural information is quite often extremely difficult to obtain for these proteins. In this chapter, we focus on a particular family of fibrous proteins: trimeric, β-stranded viral adhesins. We describe folding strategies that led to the identification of stable domains that could be crystallized, and the novel structural motifs that are emerging. We also discuss self-assembling peptides derived from these motifs. Finally, we review the possibilities of designing novel macroscopic materials as well as nanoscale fibrous objects with controlled dimensions and properties based on these novel structural motifs.

Key Words: Fibrous proteins; β-structure; protein folding; triple β-helix; triple β-spiral; self-assembling peptides; amyloid fibers.

1. Introduction

Natural fibrous proteins can be made up from collagen triple helices, α-helical coiled structures, or β-structured motifs ***(1–3)***. Collagen triple helices and coiled coils have been much more extensively studied from a biochemical, structural, and biophysical point of view, and they are not discussed in this chapter. Until recently, much less structural and folding information has been available for β-structured fibrous proteins. Their intrinsic structural complexity and notorious tendency to aggregate makes β-structured proteins more difficult to fold and handle.

From: *Methods in Molecular Biology, vol. 300:*
Protein Nanotechnology, Protocols, Instrumentation, and Applications
Edited by: T. Vo-Dinh © Humana Press Inc., Totowa, NJ

One class of β-structured fibrous proteins is used by viruses as their attachment organelles ***(4–6)***. They are often homotrimers composed of an N-terminal domain attached to the viral capsid, a central shaft, and a C-terminal receptor-binding or chemosensor domain ***(7,8)***. The long, thin fiber shaft appears to act as a mechanical device capable of reaching the virus. These fibrous proteins have evolved to survive in harsh extracellular environments; they are, therefore, extremely stable proteins, resistant to sodium dodecyl sulfate (SDS), urea, temperature, and proteases ***(9)***. The primary sequence of the shaft parts often comprises sequence repeats, suggesting that these parts fold into regular, repetitive structures.

Owing to their elongated and asymmetric nature, these proteins have intrinsic crystallization difficulties. Because they are resistant to proteases, proteolytic domains cannot be isolated. However, studying their folding pathways can reveal folding intermediates comprising stable domains; these stable domains can be crystallized and can lead to structural information ***(10–12)***. The use of such a strategy revealed that the structure of such stable domains has novel β-structured motifs for some of these viral adhesins ***(13–16)***. We believe that this kind of strategy can be applicable to other fibrous proteins of yet unknown structure. In this chapter, we underline such strategies, and discuss the structure, folding, and assembly of two case studies: the fibers of human adenoviruses and the short tail fiber of bacteriophage T4. The structure, folding, and assembly of phage P22 fibers that remain a paradigm in this area have been extensively reviewed before ***(17,18)*** and are not discussed herein. We finally discuss how the knowledge of structural building blocks and folding pathways can lead to rational design of novel nanofibers with controlled dimensions and properties.

2. Methods

2.1. Unfolding Studies and Identification of Stable Domains

For many β-structured fibrous proteins, complete unfolding requires denaturing treatments that can be considered extreme relative to the average globular protein, such as temperatures higher than 70°C. To look for eventual long-lived and stable unfolding intermediates, one needs to explore milder denaturing conditions. These include lower temperatures, and moderate concentrations (0.1–0.5%) of detergents such as SDS, or denaturing agents such as urea or guanidine hydrochloride. One attractive property of the viral adhesins that we studied so far is that the native trimers are not dissociated by SDS at ambient temperatures; dissociation into monomers is achieved only after heating. The native proteins bind very little SDS and migrate slowly in distinct positions in SDS-polyacrylamide-containing gels ***(10,11)***. This resistance to

SDS is particularly useful because it allows separation of native trimers from partially folded and misfolded chains directly by SDS-polyacrylamide gel electrophoresis (PAGE). Partially unfolded, but not dissociated, chains can quite often be directly visualized in SDS-PAGE gels ***(10,11)***.

Another advantage of big fibrous proteins is that their native forms are visible in electron microscopy, and stable, partially folded intermediates can be visualized directly as well ***(11)***. To identify the borders between the unfolded part and the stable domain, limited proteolysis can be used, utilizing proteases that can act under these mildly denaturing conditions. Proteases such as endoproteinase Glu-C, endoproteinase Lys-C, chymotrypsin, trypsin, and elastase can act in 0.1% SDS (endoproteinase Glu-C can act in up to 0.5% SDS). Endoproteinase Glu-C, endoproteinase Lys-C, and trypsin can tolerate denaturing conditions up to 5 *M* urea. Two proteases with contrasting specificities can be used, in order to assess the borders with certainty. The identity of the stable fragments can subsequently be confirmed by N-terminal sequencing and/or mass spectroscopy. Another option is to partially unfold the fiber of interest at an intermediate temperature; cool down the sample without allowing refolding; and then proceed with proteolysis at the optimum, lower temperature of the protease. If metal ions, cofactors, and so on are known to be essential for the protein, they might contribute to the stability of the domains. Therefore, it is useful to compare the proteolysis results in the presence and absence of cofactors (*see* **Subheading 2.5.**). **Figure 1** illustrates the various steps of the aforementioned strategy for a generic fibrous protein.

2.2. Recombinant Production of Fibrous Proteins

During the production of a fibrous protein by recombinant methods, the risk of obtaining insoluble aggregates (inclusion bodies) has to be considered. To avoid inclusion body formation, standard described strategies can be applied ***(19)***. However, fibrous proteins might present some additional difficulties. Cellular chaperones can participate in the folding and assembly of viral fibers; for example, Hsp70 and Hsp90 are necessary for the formation of the reovirus fibers ***(20,21)***. Sometimes specific chaperones are used to ensure correct folding and assembly of fibrous proteins. The T4 short fibers that we describe in **Subheading 2.5.** use such a specific, phage-encoded chaperone, called gp57. In the absence of this protein, the T4 short fibers form insoluble aggregates when expressed in *Escherichia coli*; expression in their correctly folded, trimeric state requires coexpression with gp57 ***(22)***. gp57 is also necessary for the folding and assembly of long tail fibers, but its mechanism of action is currently unknown. Thus, when a chaperone requirement exists, the fibrous protein of interest has to be overexpressed in the presence of this cellular or virally encoded chaperone.

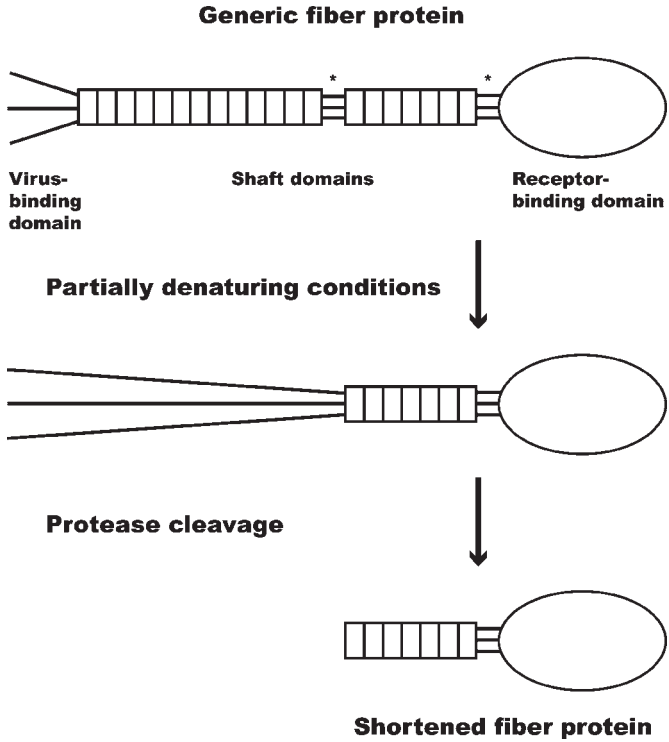

Fig. 1. Generalized strategy for producing shortened fiber protein fragments suitable for crystallization. Shown at the top is a generic fiber protein with an N-terminal virus-binding domain, central shaft domains, and a C-terminal receptor-binding domain. Asterisks indicate hinge regions. In partially denaturing conditions (which may consist of raising the temperature or adding intermediate concentrations of detergents or denaturants), the protein partially unfolds. Protease is then added, which cleaves off the unfolded domains. Note that although N-terminal unfolding is shown here, unfolding may also take place elsewhere.

2.3. Crystallization

Fiber proteins can in principle be crystallized using the same strategies as globular proteins, trying different precipitants, pHs, additives, and temperatures. The virus-binding regions in fiber proteins may, however, in the absence of the rest of the virus or bacteriophage, be unfolded and lead to a specific aggregation, hindering crystallization. Furthermore, because they often have specific regions where the fiber is bent or kinked, this can lead to lower success rates in crystallization and, in the case of successful structural determination, regions that are disordered in the crystallographic electron density. The identification of a stable domain, eliminating some or all of the binding regions,

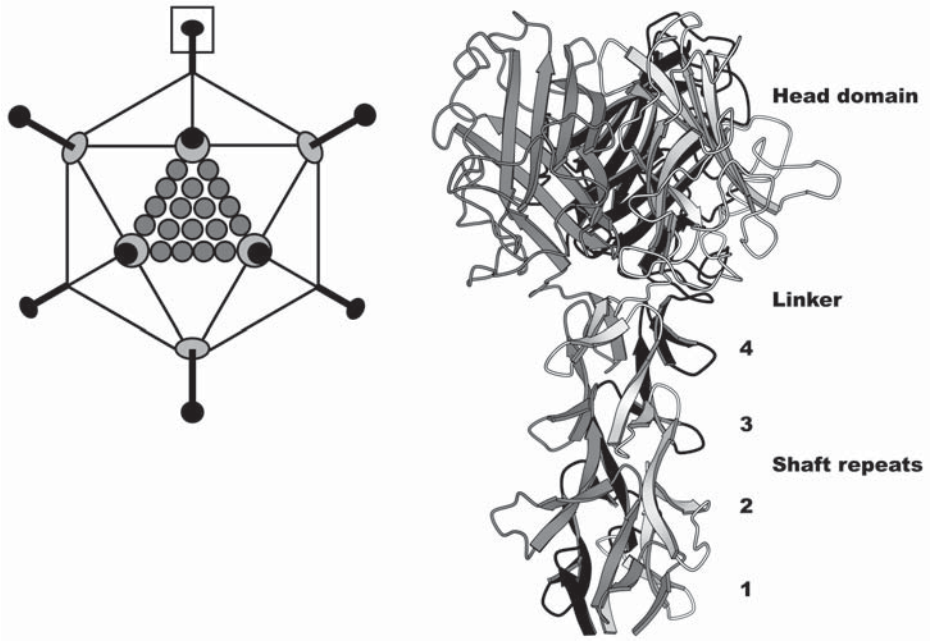

Fig. 2. Schematic drawing of adenovirus (**left**) and structure of adenovirus type 2 fiber stable fragment (**right**). The head domain, linker region, and shaft repeats are labeled. In the full-length fiber proteins up to 22 repeats are present. The part of the fiber of which the structure is shown is boxed. The shown adenovirus particle is about 80 nm wide (not counting the fibers, which can measure up to 35 nm each), and the part of the structure shown is about 10 nm long. The figures were produced using Molscript ***(56)*** and atomic coordinates publicly available from the PDB protein structure database (http://www.rcsb.org). One of the chains is shown in black and the other two in gray. The PDB code for the adenovirus type 2 stable fragment is 1QIU.

greatly aids in obtaining diffracting crystals. As with globular proteins, important for the success of crystallization is purity, conformational homogeneity, and the ability to obtain the protein in reasonable amounts (upward from about 0.5 mg) and at relatively high concentrations (5–50 mg/mL).

2.4. Case Study #1: Adenovirus Fiber

Adenoviruses cause respiratory, gastroenteric, and ocular infections in humans ***(23)***. The fibers (parallel homotrimers, with each monomer containing 582 amino acids) protrude from the 12 vertices of the icosahedral capsid, and they are responsible for cell attachment (**Fig. 2**). The C-terminal, globular head domain attaches to the cell receptor ***(7)***. The central shaft contains a repeating sequence motif with an invariant glycine or proline and a conserved pattern of

hydrophobic residues ***(24)***. The shaft can be as long as 30 nm in some serotypes, featuring 22 such repeats. The fiber is imbedded to the viral capsid with a short, N-terminal part (about 45 residues).

Unfolding experiments of the entire adenovirus type 2 fiber ***(11)*** led to the identification of a long-lived unfolding intermediate. In the presence of SDS at moderate temperatures (4–22°C), partial unfolding of the fiber starts from the N-terminus, and a stable intermediate accumulates that has the C-terminal head and part of the shaft structured as seen by electron microscopy. This unfolding intermediate is directly visible in SDS-containing polyacrylamide gels, running slower than the native trimer ***(11)***. After digestion of the unfolded parts by limited proteolysis and N-terminal sequencing, the stable domain has been identified to span residues 319 to 582. This domain was subsequently cloned and expressed in *E. coli*; it has been obtained in its trimeric, properly folded form, and its successful crystallization led to the solving of its crystal structure ***(14)***.

The crystal structure uncovered the fold of four shaft repeats, which form a triple β-spiral (**Fig. 2**). The basic structural motif is a strand-loop-strand motif; these motifs are connected by a type 2 β turn that contains a conserved glycine (in the solved part of the shaft) or a proline (in the rest of the shaft). The three chains wrap around each other to give a highly intertwined structure (one-third of the surface area of the monomer is buried in the trimer). To superimpose one repeat onto the next repeat at its C-terminus, a translation along the shaft of about 1.3 nm and a clockwise rotation of about 50° are necessary. Stabilizing features of the structure are a central longitudinal hydrophobic core, to which three hydrophobic amino acid side chains from every shaft repeat contribute; hydrophobic patches at greater radius formed by two additional conserved hydrophobic side chains; and conserved intra- and interchain hydrogen bonds. The loop in the strand-loop-strand motif has a variable sequence and is solvent exposed; these solvent-exposed loops are a unique structural feature among the so-far known fibrous folds.

The crystal structures of fiber head domains of types 5 ***(25)***, 2 ***(26)***, 12 ***(27)***, and 3 ***(28)*** are known and virtually identical apart from differences in the loop regions. Each fiber head monomer contains an antiparallel β sandwich, formed by two four-stranded β sheets packing together at an angle of about 30°. In the globular trimer, the β sandwiches make up a three-bladed propeller (**Fig. 2**). The trimeric head contains three receptor-binding sites on the sides of the trimer at monomer-monomer interfaces ***(27)***. The three receptor-binding sites per trimer lead to tighter binding owing to avidity enhancement of affinity ***(29)***. A linker region (residues 393–398) between the head and shaft domains may lead to flexibility between the two domains; indeed, in the crystal structure of the stable fragment ***(14)***, the central threefold axes of the two domains

were displaced by about 2°. Whether this flexibility has a role in receptor-binding is currently not known.

Interestingly, a C-terminal fragment of mammalian reovirus fiber (σ 1 protein) has been obtained by tryptic digestion of a deletion mutant, and its structure has been solved ***(30)***. Its structure showed a β-structured globular head domain similar to adenovirus fiber head and three triple β-spiral shaft repeats.

2.5. Case Study #2: Bacteriophage T4 Short Tail Fiber

Bacteriophage T4 acts like a self-powered nanoscopic syringe. It is also one of the largest and most complex viruses known and contains more than 40 different structural proteins ***(31)***. It is very efficient, since one particle is generally sufficient to productively infect its *E. coli* host and several hundred daughter phages can be produced in 30 min at 37°C. During infection, first the long tail fibers bind to the bacterial lipopolysaccharide or OmpC (outer membrane protein C). Once at least three long tail fibers have bound, the phage base-plate changes conformation from the hexagon form to the star form ***(32)***. This leads the short tail fibers to extend and bind to the bacterial lipopolysaccharide core region (**Fig. 3**). The next step is contraction of the outer tail tube, driving the inner tail tube through the bacterial membrane (**Fig. 3**). The inner tail tube is capped by a needle ***(33)*** with lysozyme activity, helping the puncturing of the bacterial cell wall. The phage DNA passes through the tail tube into the bacterium and directs the production of daughter phage.

The long tail fibers are a complex of four different proteins ***(34)***. The proximal half of the fiber (or "thigh") is formed by a gp34 trimer (1289 amino acids per subunit), the hinge (or "knee") contains a monomer of gp35 (372 amino acids), and the distal half of the fiber (or "shin") is made up of trimers of gp36 and gp37. gp37 (1026 residues per monomer) is responsible for receptor binding and makes up the bulk of the distal fiber half, which gp36 (221 amino acids per monomer) connects the gp37 trimer to the hinge. The crystal structures of the long tail fiber proteins have not yet been reported.

Short tail fibers are composed of a single protein, gp12, of 527 residues, which forms parallel homotrimers. Coexpression of gp12 with the T4 chaperone protein gp57 ***(22)*** in *E. coli* and subsequent purification led to the availability of material of suitable quantity, purity, and solubility. However, despite extensive trials, crystals could not be obtained from this material. Like the adenovirus fiber, its N-terminus is flexible and/or unstable, and 37 to 41 amino acids can be digested away already at 37°C. This common pattern may be owing to the fact that the N-termini of both proteins are attached to the viral capsid or the base plate in their native context and, therefore, may become destabilized in the isolated, purified form. On further increase in the temperature at 56°C followed by protease digestion, the first 83 amino acids are

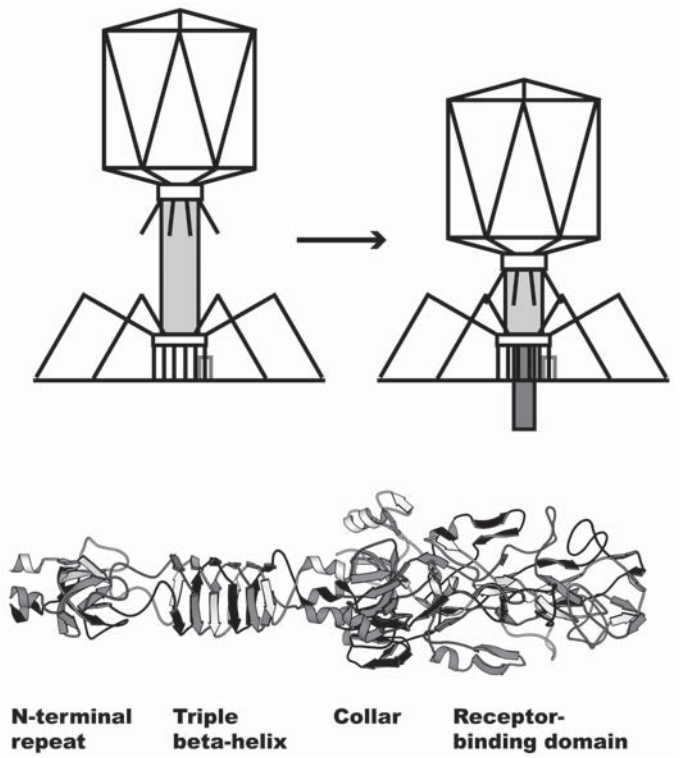

Fig. 3. Schematic drawing of uncontracted bacteriophage T4 (**top left**) and contracted T4 (**top right**). Only four of the six fibritin and long tail fibers are shown for clarity. Uncontracted T4 particles are about 300 nm high. At the bottom is a composite structure of the known domains of the bacteriophage T4 short tail fiber (combined from PDB codes 1H6W and 1OCY). One of the chains is shown in black and the other two in gray. The receptor-binding domain, collar domain, triple β-helix, and resolved N-terminal repeat are labeled. Five more N-terminal repeats are present in the full-length protein. In the schematics, the part of the short tail fiber of which the structure is known is boxed in gray. Full-length short tail fibers are about 35 nm long, and the part of the structure shown is about 15 nm.

removed and a second cleavage site is uncovered after Arg-395. This heat- and protease-stable fragment crystallized successfully ***(12)***. Another fragment, prepared in the same way except in the presence of divalent zinc cations, led to a fragment in which only the N-terminal 83 amino acids are removed. The resulting crystal structures ***(15,16)*** allowed the identification of several new folds, although parts of the crystallized protein were not visible in the crystallographic electron density because of static disorder. This is true for amino acids 85 to

245 in the first fragment, and for residues 85 to 329 in the second. Nevertheless, the combined crystal structures allowed us to describe the structure of amino acids 246 to 527 (**Fig. 3**).

The N-terminal domain contains six 17-residue sequence repeats comprising conserved hydrophobic residues such as alanines and threonines, and conserved glutamic acid residues. Of these repeats, one has been resolved in the crystal structure (residues 255–271); it basically comprises two β strands connected by a solvent-exposed type 1 β turn. Eight of these sequence repeats are also present in gp34 and one in gp37. Because the residues between the repeats are not conserved and the spacing between the repeats is variable, it appears that the T4 fiber shaft structures are not as regular as the structure of the adenovirus fiber shaft.

The region comprising amino acids 290 to 327 forms a right-handed triple-stranded β-helix, in which each monomer contributes six, six-residue β strands in two 360° turns (**Fig. 3**). Two amino acids of each β strand contribute hydrophobic side chains to a central hydrophobic core, and the framework of the triple β-helix is held together by interchain hydrogen bonds between the β strands. In the β-helix domain, each monomer has 57% of its surface buried in the complex. Interestingly, the needle at the bottom of the tail tube, which is a trimer of gp5, also contains a triple β-helix, but with longer β strands of eight residues and comprising seven 360° turns ***(33)***.

The arrow-shaped C-terminal head domain has a globular collar domain ***(15)*** and a strongly intertwined zinc-containing receptor-binding domain ***(16)***. The collar domain has weak structural homology to a part of bacteriophage T4 gp11, suggesting a possible distant evolutionary relationship. The receptor-binding domain has yet another, unique fold. It contains few regular secondary structure elements, although the normal requirements for a stable protein fold are fulfilled: it has a hydrophobic core and hydrogen-bonding requirements are met. In the center of the trimeric receptor-binding domain, we found a single zinc ion, octahedrally coordinated by the NE2 atoms of two histidine residues from each monomer. In the absence of zinc ions, this domain unfolds at 56°C, whereas in the presence of zinc it does not.

3. Self-Assembling Peptides

It has recently been shown that a number of peptides derived from natural protein sequences can self-assemble into β-structured supramolecular assemblies ***(35,36)***. These supramolecular assemblies can adopt the form of fibrils, tubes, sheets, or monolayers ***(37)*** and can serve as biomaterials with a variety of potential applications. They can be used as, e.g., vehicles for encapsulation and delivery of therapeutics or as biological scaffolds for cell attachment or biomineralization templates ***(38,39)***. Repetitive sequences derived from natu-

ral fibrous proteins that are β structured and extremely stable such as the viral adhesins are, of course, very attractive candidates for the design of peptides that could self-assemble into nanofibers. Information on the crystal structure of the native protein is a very important factor because it allows rational design; however, as we will see, this information will need to be coupled with folding and assembly information.

Synthetic peptides corresponding to shaft sequences of the adenovirus fiber shaft repeats 3 and 4 (**Fig. 2**) do not adopt the triple β-spiral conformation in the absence of the head. Instead, they self-assemble into long, unbranched fibrils that can reach the order of microns. These fibrils, called amyloid-type fibrils, adopt a cross-β structure, i.e., a structure where the β strands lie perpendicular to the fibril axis ***(40,41)***. The adenovirus head domain is essential for trimerization, because deletions or mutations in this part were found to hinder the trimerization process in vivo ***(42,43)***. This domain may act as a registration signal necessary for the correct alignment and assembly of the three chains, as do the procollagen regions in collagen folding ***(44)*** or the C-terminal domain of phage T4 fibritin ***(45)***. This short (27 amino acids) domain in a β-propeller conformation serves as a template for the correct trimerization of the triple coiled-coil N-terminal domain of the fibritin. (Fibritin fibers can be seen protruding from the "collar" part of the phage T4 particle in **Fig. 3**.) Registration signals are often used to ensure correct folding and assembly of fibrous proteins, and this has to be taken into account while engineering fibrous constructs based on repetitive fibrous sequences. If engineering of a construct that adopts the native structure is desired, then a registration signal has to be included in the design. Otherwise, out-of-register polymerization of the repetitive sequences will give fibers that could adopt a different structure and have different properties. It is important, though, to keep in mind that these properties could be interesting in their own right and could be used toward different applications.

The advantages and disadvantages of peptide synthesis vs bacterial expression also have to be considered. Short peptides (up to about 40 amino acids) are easy to design and synthesize; however, obtaining synthetic peptide material at higher scales is usually more expensive than protein material produced by recombinant methods. Another advantage of chemical synthesis is that it also allows the combination of peptide and nonpeptide moieties, such as alkyl chains, in order to create hybrid structures. Such hybrid structures comprising hydrophilic peptides linked to alkyl chains designed for biomineralization have been recently described. They can organize into cylindrical micelles, and phosphorylated serine residues on their surface organize the deposition of calcium ions, leading to hydroxyapatite nanofibers ***(46)***.

4. Applications

The first steps toward applications of β-structured, viral fibrous proteins in materials science and nanotechnology have been made. These proteins could be useful for the design of nanoscale fibrous objects with specified dimensions and properties, and the fabrication of novel biocompatible and biodegradable materials. Engineering fibrous constructs for experimental gene therapy is another rapidly developing field concerning adenovirus fibers.

Novel gene therapy vectors are being designed to specifically target certain tissues, and for this purpose fibers with modified tropism are needed. This rapidly advancing field recently provided a wealth of information about and considerable insight into engineering possibilities of the adenovirus fiber (reviewed in **ref. *47***). It has been shown that the protein can be modified in several ways and still be folded correctly. The length of the shaft can be increased by nine repeats ***(48)***, and other receptor binding domains can be inserted in loops or at the C-terminus ***(47)***. Hybrid adenovirus fibers where the knob domain or the knob plus some or many shaft repeats have been replaced by foreign trimerization domains have been incorporated into experimental adenovirus-based gene therapy vehicles ***(47)***.

Artificial (macroscopic) fibers with sequences based on adenovirus shaft repeats were produced by recombinant gene technology long before a crystal structure became available. O'Brien et al. ***(49)*** have designed polymers with fiber shaft repeats as building blocks that were bacterially expressed as inclusion bodies. After refolding, purification, and spinning, they reported properties "comparable to commercial textile fibers" for the fibers obtained ***(49)***. Although no detailed structure and assembly studies have been conducted on this material, powder diffraction patterns together with Raman studies suggest that these fibers might adopt a cross-β, amyloid-like conformation ***(49,50)***. Construction of nanoarrays based on bacteriophage T4 long fiber proteins gp34 and gp37 is being pioneered by Edward Goldberg's group ***(51)***. Their strategy is, first, to engineer self-assembling protein rod units; second to insert functional moieties at precise positions in these protein rods without perturbing folding and assembly; and, finally, to assemble two-dimensional or three-dimensional nanoarrays from these functionalized rods. gp34 and gp37 trimers are both approx 70 nm long. gp12 and adenovirus fiber could perhaps also be used for this purpose; they are both about 35 nm long.

A question is, can β-structured fibers be made longer? In the case of adenovirus, adding nine repeats (to a total length of 45 nm) is possible in vivo ***(48)***. The addition of more repeats may be difficult, because N-termini have to be held apart until the C-terminal trimerization domain is synthesized and can fold, and putting the trimerization domain at the N-terminus may lead to protein synthesis problems in *E. coli*.

5. Conclusion

We have reviewed in this chapter strategies that recently led to the X-ray structures of several building blocks of viral and bacteriophage fiber proteins. Using these building blocks, artificial nanoscale fibrous constructs can be obtained by rational design based on the crystal structures, provided that folding considerations are taken into account.

The structures of many viral and bacteriophage fibrous proteins remain unresolved. In the near future, further structural information on these natural fibrous proteins and their interaction partners (chaperones and receptors) should become available. This additional structural information and the eventual emergence of new motifs will provide a sounder basis for engineering materials based on stable trimeric viral fibers.

Combining the various motifs and building blocks from these fibers in order to make artificial fusion proteins is the next challenge toward the design of fibrous nanoconstructs. The first steps in this direction have been made: the trimerization motif of the T4 fibritin (27 amino acids) can be fused to fibrous parts of a variety of proteins, ranging from T4 short tail fibers to collagen helices giving stable, trimeric proteins ***(52,53)***. Stetefeld et al. ***(54)*** have determined the structure of such an artificial collagen triple helix fused to the fibritin trimerization motif. Their structure shows how the staggered collagen triple helix (without a central threefold axis) is accommodated by the threefold symmetric foldon by a 60° bend. A fusion protein carrying the trimerization domain plus most of the coiled-coil of the fibritin was also successfully fused to the two most N-terminal repeats of the adenovirus fiber shaft. A linker sequence (SQNV) present in both fusion partners was used in this construct in order to avoid potential structural conflicts between motifs ***(55)***. These examples prove that combination of fibrous motifs varying from coiled-coils to the triple β-spiral is feasible, provided that the appropriate linkers are used.

The strategies that we have described underline some general trends in the folding and assembly of β-structured fibrous proteins. We believe that these strategies can be helpful if followed for other fibrous proteins of yet-unknown structure. However, an important thing to keep in mind is that every protein can be different; only a rigorous and detailed understanding of its own folding and assembly pathway can lead to structural information, and further rational design based on this information.

Acknowledgments

We acknowledge support from the Centre National de la Recherche Scientifique and from the Direction d'Objectifs Matériaux of the Commissariat à l'Energie Atomique. Mark van Raaij is a Ramón y Cajal investigator sponsored by the Spanish Ministry of Science and Technology.

References

1 Beck, K. and Brodsky, B. (1998) Supercoiled protein motifs: the collagen triple-helix and the alpha-helical coiled coil. *J. Struct. Biol.* **122,** 17–29.

2 Jenkins, J. and Pickersgill, R. (2001) The architecture of parallel β-helices and related folds. *Prog. Biophys. Mol. Biol.* **77,** 111–175.

3 Mitraki, A., Miller, S., and van Raaij, M. J. (2002) Review: conformation and folding of novel beta-structural elements in viral fiber proteins: the triple beta-spiral and triple beta-helix. *J. Struct. Biol.* **137,** 236–247.

4 Iwashita, S. and Kanegasaki, S. (1976) Enzymic and molecular properties of base-plate parts of bacteriophage P22. *Eur. J. Biochem.* **65,** 87–94.

5 Philipson, L., Lonberg-Holm, K., and Petterson, U. (1986) Virus-receptor interaction in an adenovirus system. *J. Virol.* **2,** 1064–1075.

6 Riede, I. (1987) Receptor specificity of the short tail fibres (gp12) of T-even type *Escherichia coli* phages. *Mol. Gen. Genet.* **206,** 110–115.

7 Ruigrok, R. W. H., Barge, A., Albiges-Rizo, C., and Dayan, S. (1990) Structure of adenovirus fiber. II. Morphology of single fibers. *J. Mol. Biol.* **215,** 589–596.

8 Makhov, A. M., Trus, B. L., Conway, J. F., Simon, M. N., Zurabishvili, T. G., Mesyanzhinov, V. V., and Steven, A. C. (1993) The short tail-fiber of bacteriophage T4: molecular structure and a mechanism for its conformational transition. *Virology* **194,** 117–127.

9 King, J. and Laemmli, U. K. (1971) Polypeptides of the tail fibres of bacteriophage T4. *J. Mol. Biol.* **62,** 465–477.

10 Chen, B. and King, J. (1991) Thermal unfolding pathway for the thermostable P22 tailspike endorhamnosidase. *Biochemistry* **30,** 6260–6269.

11 Mitraki, A., Barge, A., Chroboczek, J., Andrieu, J. P., Gagnon, J., and Ruigrok, R. W. H. (1999) Unfolding studies of human adenovirus type 2 fiber trimers: evidence for a stable domain. *Eur. J. Biochem*. **264,** 599–606.

12 van Raaij, M. J., Schoehn, G., Jaquinod, M., Ashman, K., Burda, M. R., and Miller, S. (2001) Identification and crystallisation of a heat- and protease-stable fragment of the bacteriophage T4 short tail fibre. *Biol. Chem.* **382,** 1049–1055.

13 Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R., and Reinemer, P. (1994) Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. *Science* **265,** 383–386.

14 van Raaij, M. J., Mitraki, A., Lavigne, G., and Cusack, S. (1999) A triple beta-spiral in the adenovirus fiber shaft reveals a new structural motif for a fibrous protein. *Nature* **401,** 935–938.

15. van Raaij, M. J., Schoehn, G., Burda, M. R., and Miller, S. (2001) Crystal structure of a heat- and protease-stable part of the bacteriophage T4 short tail fibre. *J. Mol. Biol.* **314,** 1137–1147.

16 Thomassen, E., Gielen, G., Schütz, M., Schoehn, G., Abrahams, J. P., Miller, S., and van Raaij, M. J. (2003) The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold. *J. Mol. Biol.* **331,** 361–373.

17 Seckler, R. (1998) Folding and function of repetitive structure in the homotrimeric phage P22 tailspike protein. *J. Struct. Biol.* **122,** 216–222.
18. Betts, S. and King, J. (1999) There's a right way and a wrong way: in vivo and in vitro folding, misfolding and subunit assembly of the P22 tailspike. *Struct. Fold. Des.* **7,** R131–R139.
19. Mitraki, A. and King, J. (1989) Protein folding intermediates and inclusion body formation. *Bio/Technology* **7,** 690–697.
20 Lee, P. W. K., and Gilmore, R. (1998) Reovirus cell attachment protein sigma1: structure-function relationships and biogenesis. *Curr. Top. Microbiol. Immunol.* **233,** 137–153.
21 Gilmore, R., Coffey, M. C., and Lee, P. W. K. (1998) Active participation of Hsp90 in the biogenesis of the trimeric reovirus cell attachment protein sigma1. *J. Biol. Chem.* **273,** 15,227–15,233.
22 Burda, M. R. and Miller, S. (1999) Folding of coliphage T4 short tail fiber in vitro: analysing the role of a bacteriophage-encoded chaperone. *Eur. J. Biochem.* **265,** 771–778.
23. Horwitz, M. S. (1996) Adenoviruses, in *Fields Virology*. (Fields, B. N., Knipe, D. M., and Howley, P. M., eds.), Lippincott-Raven, Philadelphia, pp. 2149–2171.
24 Green, N. M., Wrigley, N. G., Russel, W. C., Martin, S. R., and McLachlan, A. D. (1983) Evidence for a repeating cross-beta sheet structure in the adenovirus fiber. *EMBO J.* **2,** 1357–1365.
25 Xia, D., Henry, L. J., Gerard, R. D., and Deisenhofer J. (1994) Crystal structure of the receptor-binding domain of adenovirus type 5 fiber protein at 1. 5 Å resolution. *Structure* **2,** 1259–1270.
26 van Raaij, M. J., Louis, N., Chroboczek, C., and Cusack, S. (1999) Structure of the human adenovirus serotype 2 fiber head domain at 1. 5 A resolution. *Virology* **262,** 333–343.
27 Bewley, M. C., Springer, K., Zhang, Y. B., Freimuth, P., and Flanagan, J. M. (1999) Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. *Science* **286,** 1579–1583.
28 Durmort, C., Stehlin, C., Schoehn, G., Mitraki, A., Drouet, E., Cusack, S., and Burmeister, W. P. (2001) Structure of the fiber head of Ad3, a non-CAR-binding serotype of adenovirus. *Virology* **285,** 302–312.
29 Lortat-Jacob, H., Chouin, E., Cusack, S., and van Raaij, M. J. (2001) Kinetic analysis of adenovirus fiber binding to its receptor reveals an avidity mechanism for trimeric receptor-ligand interactions. *J. Biol. Chem.* **276,** 9009–9015.
30 Chappell, J. D., Prota, A. E., Dermody, T. S., and Stehle, T. (2002) Crystal structure of reovirus attachment protein sigma1 reveals evolutionary relationship to adenovirus fiber. *EMBO J.* **21,** 1–11.
31. Karam, J. D. (1994) *Molecular Biology of Bacteriophage T4*, ASM Press, Washington, DC.
32 Crowther, R. A., Lenk, E. V., Kikuchi, Y., and King, J. (1977) Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4. *J. Mol. Biol.* **116,** 489–523.

33 Kanamaru, S., Leiman, P. G., Kostyuchenko, V. A., Chipman, P. R., Mesyanzhinov, V. V., Arisaka, F., and Rossmann, M. G. (2002) Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415,** 553–557.
34 Cerritelli, M. E., Wall, J. S., Simon, M. N., Conway, J. F., and Steven A. C. (1996) Stoichiometry and domainal organization of the long tail-fiber of bacteriophage T4: a hinged viral adhesin. *J. Mol. Biol.* **260,** 767–780.
35 Zhang, S., Holmes, T., Lockshin, C., and Rich, A. (1993) Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane. *Proc. Natl. Acad. Sci. USA* **90,** 3334–3338.
36 Aggeli, A., Bell, M., Boden, N., Keen, J. N., Knowles, P. F., McLeish, T. C., Pitkeathly, M., and Radford, S. E. (1997) Responsive gels formed by the spontaneous self-assembly of peptides into polymeric beta-sheet tapes. *Nature* **386,** 259–262.
37 Aggeli, A., Nyrkova, I. A., Bell, M., Harding, R., Carrick, L., McLeish, T. C., Semenov, A. N., and Boden, N. (2001) Hierarchical self-assembly of chiral rod-like molecules as a model for peptide beta-sheet tapes, ribbons, fibrils, and fibers. *Proc. Natl. Acad. Sci. USA* **98,** 11,857–11,862.
38 Zhang, S. and Altman, M. (1999) Peptide self-assembly in functional polymer science and engineering. *React. Funct. Polym.* **41,** 91–102.
39 Zhang, S., Marini, D. M., Hwang, W., and Santoso, S. (2002) Design of nanostructured biological materials through self-assembly of peptides and proteins. *Curr. Opin. Chem. Biol.* **6,** 865–871.
40 Luckey, M., Hernandez, J., Arlaud, G., Forsyth, V. T., Ruigrok, R. W., and Mitraki, A. (2000) A peptide from the adenovirus fiber shaft forms amyloid-type fibrils. *FEBS Lett.* **468,** 23–27.
41. Mitraki, A., van Raaij, M. J., Ruigrok, R., Cusack, S., Hernandez, J.-F., and Luckey, M. (2001) Structure, folding and assembly of adenovirus fibers, in *Self-Assembling Peptide Systems in Biology, Medicine and Engineering* (Aggeli, A., Boden, N., and Zhang, S., eds.), Kluwer Academic, Dordrecht, The Netherlands, pp. 221–233.
42 Novelli, A. and Boulanger, P. A. (1991) Deletion analysis of functional domains in baculovirus-expressed adenovirus type 2 fiber. *Virology* **185,** 365–376.
43 Hong, J. S. and Engler, J. A. (1996) Domains required for assembly of adenovirus type 2 fiber trimers. *J. Virol.* **70,** 7071–7078.
44 McLaughlin, S. H. and Bulleid, N. J. (1998) Molecular recognition in procollagen chain assembly. *Matrix Biol.* **16,** 369–377.
45 Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V., and Rossmann, M. G. (1997) Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. *Structure* **5,** 789–798.
46. Hartgerink, J. D., Beniash, E., and Stupp, S. I. (2001) Self-assembly and mineralization of peptide-amphiphile nanofibers. *Proc. Natl. Acad. Sci. USA* **294,** 1684–1688.
47 Barnett, B. G., Crews, C. J., and Douglas, J. T. (2002) Targeted adenoviral vectors. *Biochim. Biophys. Acta* **1575,** 1–14.

48 Seki, T., Dmitriev, I., Kashentseva, E., Takayama, K., Rots, M., Suzuki, K., and Curiel, D. T. (2002) Artificial extension of the adenovirus fiber shaft inhibits infectivity in coxsackievirus and adenovirus receptor-positive cell lines. *J. Virol.* **76,** 1100–1108.
49. O'Brien, J. P., Hoess, R. H., Gardner, K. H., Lock, R. L., Wasserman, Z. R., Weber, P. C., and Salemme, F. R. (1994) Design, synthesis, and fabrication of a novel self-assembling fibrillar protein, in *Silk Polymers, Materials Science and Biotechnology*, ACS Symposium Series 544 (Kaplan, D., Adams, W. W., Farmer, B., and Viney, C., eds.), American Chemical Society, Washington, DC, pp. 104–117.
50. Gillespie, D. B., Thiel, B. L., Trabbic, K. A., Viney, C., and Yager, P. (1994) Structural investigation of $(AdII)_{26}$ fiber, a novel bioengineered material based on a viral spike protein. *Macromolecules* **27,** 6177–6182.
51 Hyman, P., Valluzzi, R., and Goldberg, E. (2002) Design of protein struts for self-assembling nanoconstructs. *Proc. Natl. Acad. Sci. USA* **99,** 8488–8493.
52 Miroshnikov, K. A., Marusich, E. I., Cerritelli, M. E., Cheng, N., Hyde, C. C., Steven, A. C., and Mesyanzhinov, V. V. (1998) Engineering trimeric fibrous proteins based on bacteriophage T4 adhesins. *Protein Eng.* **11,** 329–332.
53 Frank S., Kammerer, R. A., Mechling D., Schulthess, T., Landwehr, R., Bann, J., Guo, Y., Lustig, A., Bächinger, H. P., and Engel, J. (2001) Stabilization of short collagen-like triple helices by protein engineering. *J. Mol. Biol.* **308,** 1081–1089.
54 Stetefeld, J., Frank, S., Jenny, M., Schulthess, T., Kammerer, R. A., Boudko, S., Landwehr, R., Okuyama, K., and Engel, J. (2003) Collagen stabilization at atomic level: crystal structure of designed (GlyProPro)(10)foldon. *Structure* **11,** 339–346.
55 Krasnykh, V., Belousova, N., Korokhov, N., Mikheeva, G., and Curiel, D. T. (2001) Genetic targeting of an adenovirus vector via replacement of the fiber protein with the phage T4 fibritin. *J. Virol.* **75,** 4176–4183.
56. Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24,** 946–950.