Protein Engineering vol.14 no.2 pp.93–103, 2001

A proposed structural model for amyloid fibril elongation: domain swapping forms an interdigitating β-structure polymer

Neeti Sinha[1,2], Chung-Jung Tsai[1] and Ruth Nussinov[1,3,4]

[1]Intramural Research Support Program – SAIC, Laboratory of Experimental and Computational Biology, NCI-FCRDC, Frederick, MD 21702, USA and [3]Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

[2]Present address: Basic Research Laboratory, NCI-FCRDC, Frederick, MD 21702, USA

[4]To whom correspondence should be addressed, at the USA address. E-mail: ruthn@ncifcrf.gov

We propose a model illustrating how proteins, which differ in their overall sequences and structures, can form the propagating, twisted β-sheet conformations, characteristic of amyloids. Some cases of amyloid formation can be explained through a 'domain swapping' event, where the swapped segment is either a β-hairpin or an unstable conformation which can partially unfold and assume a β-hairpin structure. As in domain swapping, here the swapped β-hairpin is at the edge of the structure, has few (if any) salt bridges and hydrogen bonds connecting it to the remainder of the structure and variable extents of buried non-polar surface areas. Additionally, in both cases the swapped piece constitutes a transient 'building block' of the structure, with a high population time. Whereas in domain swapping the swapped fragment has been shown to be an α-helix, loop, strand or an entire domain, but so far not a β-hairpin, despite the large number of cases in which it was already detected, here swapping may involve such a structural motif. We show how the swapping of β-hairpins would form an interdigitated, twisted β-sheet conformation, explaining the remarkable high stability of the protofibril *in vitro*. Such a swapping mechanism is attractive as it involves a universal mechanism in proteins, critical for their function, namely hinge-bending motions. Our proposal is consistent with structural superpositioning of mutational variants. While the overall r.m.s.d.s of the wild-type and mutants are small, the proposed hinge-bending region consistently shows larger deviations. These larger deviations illustrate that this region is more prone to respond to the mutational changes, regardless of their location in the sequence or in the structure. Nevertheless, above all, we stress that this proposition is hypothetical, since it is based on assumptions lacking definitive experimental support.

Keywords: amyloid/β-hairpin/β-turn/domain swapping/hinge bending/misfolding/polymerization/structural motif

Introduction

The conversion of a native protein into a component in a growing amyloid fiber is a fascinating problem. This problem bears both on the understanding of protein folding and binding and on practical disease-related issues. During the last few years considerable work has been done, laying down the foundation of our understanding of amyloidogenic proteins and amyloid formation (Sipe, 1992; Kelly and Lansbury, 1994; Lansbury *et al.*, 1995; Saraiva, 1995; Kelly, 1996; Fink, 1998). Here we focus on the process initiating from a native, globular protein with its end product being a non-bonded amyloid polymer. A 'non-bonded amyloid polymer' is a polymer that has units connected to each other through hydrogen bonds and hydrophobic interactions, but not covalent linkages.

The tertiary structures of proteins undergoing such a polymerization differ from each other, yet their corresponding amyloid polymers are structurally of a similar type and have unique characteristics (Serpell *et al.*, 1997; Sunde and Blake, 1998). To understand the conformational changes between the native and the non-bonded polymer amyloid forms, we look for a structural motif which recurs in native proteins known to convert into non-bonded polymers. Such a motif, and the mechanism in which it plays a critical role, should be consistent with available experimental data and provide clues to how native structures undergo a conformational change to form amyloids and, in particular, to the stability and propagation of the amyloid polymer.

Different fibrils formed from different molecules are still composed of the same basic form and exhibit the same basic features (summarized in Serpell *et al.*, 1997 and references therein): First, the fibrils are uniform, straight and unbranched; second, the fibrils yield a characteristic X-ray diffraction pattern, dominated by intense 4.7 Å meridional and weaker, ~10 Å equatorial reflections. This pattern is characteristic of a cross-β structure. In such a structure, the β-sheets are arranged in a way in which they run parallel to the axis of the fibril. On the other hand, the β-strands forming the sheets are perpendicular to the axis of the fibril. Third, amyloid fibrils derived from different types of protein molecules are composed of different numbers and arrangements of protofilaments, with each protofilament consisting of such a cross-β geometry. For example, the transthyretin fibril contains four protofilaments (Serpell *et al.*, 1995) while the immunoglobulin light chain has five (Shirahama and Cohen, 1967). Fourth, *in vitro* the amyloid fibrils are extremely stable. For example, for bovine spongiform encephalopathy it has been shown that a temperature of above 200°C is required for the destruction of the amyloid polymer (Serpell *et al.*, 1997). In the model for the non-bonded polymers, the β-sheets are twisted, consistent with the larger stability of such a conformation as compared with flat sheets (Chothia, 1973). Twisted β-sheets form spontaneously from many oligopeptides (Chothia, 1973; Serpell *et al.*, 1997). This allows a continuous pattern of β-type hydrogen bonds along the filbril, over great lengths (Serpell *et al.*, 1997). These authors have further pointed out that it is possible that some amyloid polymers are composed of one pair of β-sheets, whereas others are composed of two such pairs, as in the case of the transthyretin protofilament. Hence the conformation of the amyloid polymers is uniform and extremely stable. It is composed of one or more pairs of twisted β-sheet structures.

It provides for fast propagation and growth, once a seed amyloid is present. This suggests 'sticky' ends in the protofibrils.

The motif and amyloid polymerization

There are several considerations when selecting a motif. (i) The motif is likely to be a 'building block' of the protein structure (Tsai *et al.*, 1998, 1999a,b). While its conformation in solution is likely to have the largest population time, this is not necessarily the case. (ii) The motif is a β-hairpin or an unstable conformation which partially unfolds and may adopt a β-structure. (iii) The motif should be able to flip easily. Hence it probably lies at the edge of the structure and preferably at one of the termini of the sequence, with no strong interactions with its adjacent structure in 3D space. (iv) The polymerization involves domain swapping of β-hairpin or of β-structure building blocks. The exceptionally high stability of the non-bonded polymer and the fast rate of its growth at high concentrations under certain conditions *in vitro* indicate that the structure is likely to have sticky, interdigitated staggered ends. Since the protofibril is composed of twisted β-sheets, whose propagation is parallel to the protofibril axis, with the β-strands perpendicular to it, there are two ways we may envision that growth and extension can take place. The first is where the β-sheet(s) of the monomer are added in such a way as to result in a 'smooth', ends-on propagation. In this case, the next incoming monomer would be added, to hydrogen bond with the last, edge strand(s). This last exposed β-strand would constitute a 'sticky' end, necessary for the very long fibril formation. Alternatively, a second way is through 'jagged' ends. Jagged ends may be produced via domain swapping (Bennett *et al.*, 1994, 1995) or here 'building block' motif swapping between consecutive monomers in the protofibril. Domain swapping of a β-hairpin motif overcomes to some extent the lesser stability incurred by a lack of a covalent bond, present in 'regular' polymers. The loop(s) connecting the building block motif to the remainder of the structure would serve as the swapping hinge(s).

Although to date only a relatively small number of proteins are known to form amyloids *in vivo*, under appropriate (high denaturant concentration or low pH) conditions, many (e.g. Guijarro *et al.*, 1998; Litvinovich *et al.*, 1998; Chiti *et al.*, 1999), perhaps most proteins can convert into this stable form *in vitro*. This indicates that the presence of any given motif is not an absolute necessity for such a polymerization. Even in *in vivo* amyloidogenic proteins, the motif is not a prerequisite for amyloid formation. An unstable protein or region may partially unfold and assume such a β-conformation. Since the bound form is remarkably stable, the equilibrium will shift in its direction, further driving the reaction.

The hypothetical model proposed here is for the elongation of the amyloid, rather than for its initiation. A schematic drawing illustrating one way in which such a mechanism may potentially be realized is given in Figure 1. We stress however, that the model is hypothetical, with currently no definitive experimental support for its existence.

Amyloids may be polymorphic. Inouye *et al.* (1998) have obtained data pointing to alternative conformations to those proposed by Serpell *et al.* (1997) and by Sunde and Blake (1998). Furthermore, studies show that synthetic peptides of an amyloidogenic protein form different polymorphic assemblies, from plates, ribbons to fibrils depending on solvent conditions (Kirschner *et al.*, 1998).

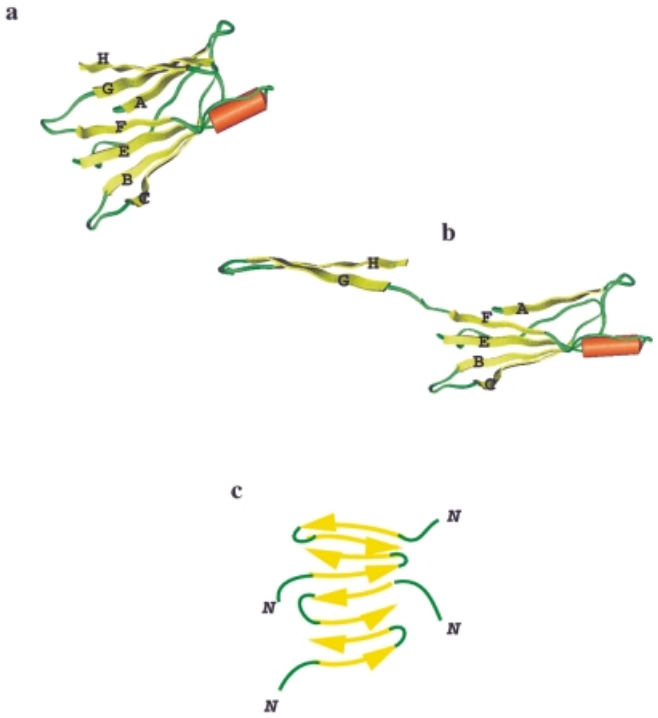


Fig. 1. A model for amyloid formation. Starting with a monomer of an amyloidogenic protein (**a**) (in this case transthyretin), a β-hairpin, consisting of strands G and H, is the potential motif. The motif flips out from the rest of the structure (**b**), to swap with a sister motif from a sister molecule, which is at the edge of the fibril. A twisted β-hairpin (b) interdigitates in a growing amyloid fibril. This results in an extension of the fibril (**c**), to produce a progressively larger polymer. Such a swapping culminates in a polymer in which there are no covalent linkages. The molecules are attached to each other via hydrogen bonds and hydrophobic interactions. 2–4 such sheets may coil around the fiber axis, to form a protofilament. The α-helices are shown in red. β-Strands are shown in yellowish green. This model represents only one hypothetical way in which swapping can take place. Alternative arrangements of the molecules are possible. For example, it is possible that only part of protein (a β-hairpin motif) would participate in swapping and interdigitation. Alternatively, in cases such as transthyretin and immunoglobulin an alternative motif may also flip out and swap. This will generate 3–4 sheets, which may coil around an axis and form a protofilament. In the absence of accurate enough measurements to position the remainder of the molecule, it is not possible to distinguish between these.

The domain swapping mechanism: similarities and differences

Eisenberg and his colleagues (Bennett *et al.*, 1994, 1995) have proposed a mechanism for protein oligomerization, the 3D domain swapping. In a domain swapping oligomer, one segment of a monomeric protein is replaced by the same segment from another chain. If the swapping involves consecutively arranged monomers on a filament, a chain reaction-like event occurs, propagating the non-covalently bonded polymer. In both cases, there is a conformational transition involving swinging out of a structural piece on a hinge-like region. This suggests that the swinging or swapping part should preferably reside on a separate building block than the remainder of the structure to which it is sequentially attached and should preferentially be at the edge of the structure and at one of the sequence termini. Additionally, the structure would be more compact if the two, swapped and non-swapped parts are separated from each other (Xu *et al.*, 1998). As we show below,

inspection of known cases of domain swapping illustrates that these criteria hold.

The largest and most critical difference between domain swapping and non-bonded polymerization during amyloid formation is that whereas domain swapping may be reversible, the polymerization is an essentially irreversible process (Lomakin *et al.*, 1996). The origin of the difference in stability may be in the conformations of the swapped parts: whereas in the 'classical' domain swapping event the swapped domain is a helix, loop, a single β-strand or an entire domain, in polymerization the swapped part is a β-hairpin structure. In cases such as in the prion, the potential swapping domain may be a conformationally unstable structure, which may partially unfold and undergo a conversion to β [summarized in Dobson and Karplus (1999) and references therein]. Litvinovitch *et al.* (1998) have recently shown how a hypothetical mechanism for fibronectin type III β-sandwich can partially unfold and self-associate to form fibrils via a β-strand swapping. However, even if a β-structure exists, if it is unstable, as in the case of lysozyme, it may unfold (B.Ma and R.Nussinov, unpublished work), with subsequent formation of an altered β-conformation. Such unfolding has recently been observed in molecular dynamic simulations.

Methods

Cutting into building blocks

The coordinates of the structures were retrieved from the Protein Data Bank (PDB) (Bernstein *et al.*, 1977). A building block is considered to be a contiguous fragment with substantial interactions between its residues. The building blocks were assigned as described in Tsai *et al.* (1999a). For every candidate fragment of the protein, the relative buried accessible surface area (ASA) was calculated. The fragment was considered as a building block when the obtained relative buried ASA value was larger than a threshold (of 0.135 $Å^2$). The relative buried ASA is the ASA of the first half-fragment buried by the second half of the fragment plus the ASA of the second half of the fragment which is buried by the first half-fragment divided by the total ASA of the fragment. Details for the cutting procedure are given in Tsai *et al.* (2000).

Salt bridges and hydrogen bonds

The presence of salt bridges was inferred when Asp and Glu side-chain carbonyl oxygen atoms were found to be within a 4.0 Å distance from the nitrogen atoms of Arg, Lys or His side-chains. When for the same pair of residues there were more than one pair of nitrogen–oxygen atoms present within 4.0 Å, the salt bridge was counted only once. The presence of a hydrogen bond was inferred when two non-hydrogen atoms with opposite partial charges were found to be within a distance of 3.5 Å. Details are described in Kumar and Nussinov (1999).

Non-polar buried surface area

The non-polar buried surface area was calculated as described by Tsai and Nussinov (1997a). The buried non-polar surface area was calculated as a fraction of the buried non-polar area out of the total non-polar area. To calculate the area buried within the motif and between the motif and the rest of the protein, the fragment comprising the motif was scrutinized for the proportion of its buried surface area, both by itself and by the rest of the protein. If a residue in the motif is buried to the same extent both by a residue within the motif and by a residue in the remainder of the protein, the calculated buried surface area for that motif–residue is added to both categories. That is, it is included in the area buried within the motif and in the area buried by the rest of the protein. Thus, in our results the total non-polar buried surface area of the motif is smaller than the sum of the area buried by the motif and the area buried by the rest of the protein.

Table I. A listing of the sequences, the sizes and the locations of the β-hairpin motifs in the native structures of proteins known to undergo a conformational conversion into non-bonded amyloid polymers

Protein	Protein size (residues)	Sequence of the motif	Size[a]	No. of residues prior to/following the motif
Cystatin C (1a67)	108	GRTTCPKSSGDLQSCE FHDEPEMAKY**TTCTFVVYSI** PWLN**QIKLL**ESKCQ	18	26/0 (*C*-terminus)
Gelsolin (1d0n)	71 residue fragment	NNGD**CFILD**LGN **NIYQWC** GSKS	14	4/4
Immunoglobulin (1bre)	108 (chain A)	**IATYYCQQY**DDLPYTFGQG **TKVEI**K	23	1/0 (*C*-terminus)
Serum amyloid P component (1sac)	206 (chain A)	SDLSRA**YSLFSYNT**QGRDN **ELLVYKERV**	21	6/1
Transthyretin (1bmz)	127 (chain A)	NDSGPR**RYTIAALLS**PY **SYSTTAVVT**NPKE	20	6/0 (*C*-terminus)
β-2 microglobulin (1bmg)	98	IQRPP**KIQVYS**RHPPEDGKP **NYLNCVYGF**HPPQI	25	0/5 (*N*-terminus)
Lysozyme (1lz1)	130	**RATNYN**AGDR**STDYGI**FQ	20	1/3
α-1-antitrypsin (1psi)	394	NKP**FVFLMIE**QNTKS **PLFMGKV**VNPTQK	19	3/0 (*C*-terminus)
Prion (1ag2)	103	GLGG**YMLG**SAMSRPMIHF GNDWEDRYYRENMYRYPNQ**VYYR** PVDQYSNQNNFVHD	37	0/14 (*N*-terminus)

The residues belonging to the β-strands are shown in bold. The sequences of the coils connecting the motif to the rest of the protein are underlined. The last column specifies the number of residues connecting the motif to its previous (sequence-wise) secondary structure element (on the amino side of the turn) and the following one (on the carboxy side of the turn). Thus, these specify the lengths of the coils on either side. If the motif is at the terminus of the sequence it is marked as such.
[a]Size of the motif (No. of residues).

Identification of the motif

The amyloidogenic proteins were first inspected for the presence of a β-hairpin, which is connected to the remainder of the protein via a coil. The potential motifs were then characterized based upon the building block assignments, buried non-polar surface area and the number of hydrogen bonds and salt bridges present within the motif and between the motif and the rest of the protein.

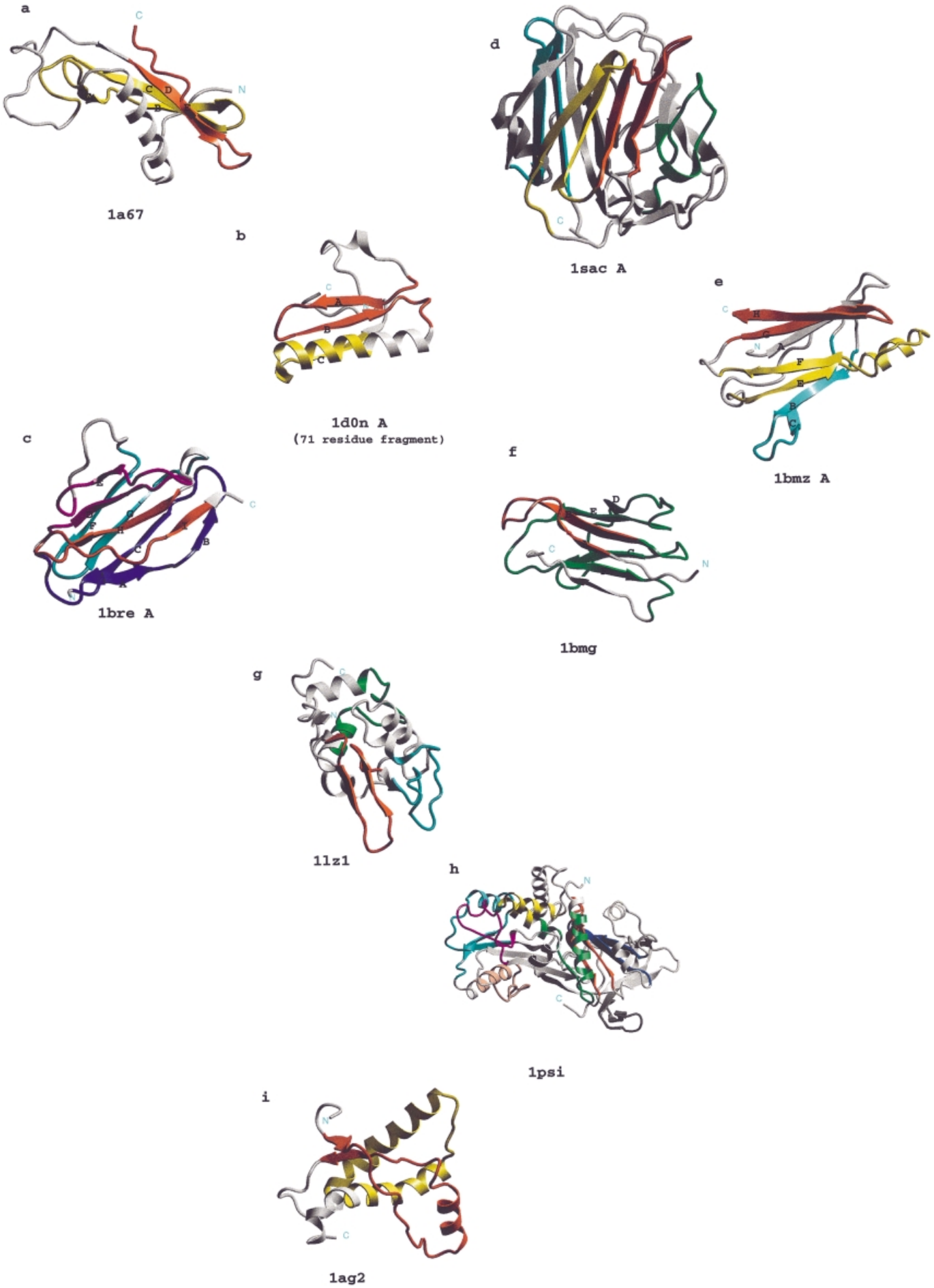
Results

The amyloidogenic cases

Table I enumerates the cases and lists the sequences and the positions of the motifs in each of the cases (Table II enumerates domain-swapped cases). Table I gives the secondary structure assignments and the lengths of the structural elements at these

Root mean square deviations (r.m.s.d.s)

The residue by residue deviations of corresponding C_α pairs were inspected by superimposing the mutants on their respective wild-types.


a
1a67
b
1d0n A
(71 residue fragment)
c
1bre A
d
1sac A
e
1bmz A
f
1bmg
g
1lz1
h
1psi
i
1ag2

sites. Figure 2(a)–(i) provide illustrations for all the cases. The figures show the assignment of the 'building blocks' of the structure. Each building block is depicted in a different color. The most likely motif is in red. The regions which have been left 'unassigned' into a specific building block, are in white. A fragment is left 'unassigned' if by itself it is an extremely unstable building block and its score does not pass the threshold, qualifying it as a building block. Its addition to a sequentially connected building block does not increase the stability of the already identified building block (Tsai *et al.*, 1999a).

We have also calculated the number of salt bridges and the hydrogen bonds within the motif, between the motif and the remainder of the structure and the total number of these in the entire structure for each of the cases. These may provide an indication of the likelihood of the motifs to swap. In addition, we have computed the non-polar buried surface areas within and between the motif and the rest of the structure. In all examples, the structures correspond to the native conformations. The results are presented in Table III. We have also carried out these calculations for the other building blocks in the structures. These are given in the legend of Table III.

Inspection of the results reveals that the motifs picked up are reasonable candidates to flip: Table I illustrates that in most cases, the sequences of these (noted in bold letters) constitute fairly good β-strands. They are at least four residues long and contain a high fraction of aromatic or of hydrophobic–aliphatic residues. In five out of the nine cases (in cystatin C, immunoglobulin, transthyretin, β_2-microglobulin and α_1-antitrypsin), the motif is at the termini of the sequence. That is also the case for the fragment of the prion whose NMR structure has been determined. In particular, in all cases the motif is at the edge of the structure [Figures 2(a)–(i)] and hence highly exposed. The number of salt bridges within the motifs, between the motifs and the remainder of the structure and the total number of these (Table III) illustrates that there are very few salt bridges at the interface, connecting the motifs to the remainder of the structure. A similar picture is observed for the hydrogen bonds. Inspection of the non-polar buried surface areas within the motifs and between the motifs and the remainder of the structures illustrates that while the non-polar buried surface area between the motif and the remainder of the structure can be fairly large in absolute terms, for seven out of nine cases it is lower than the non-polar buried surface area within the motif itself, suggesting that the motif is likely to swap as a single unit. Nevertheless, alternative, unstable regions, whether β-structures or others, may partially unfold and assume a different β-conformation. A recent molecular

Fig. 2. The building block assignments of the native structures of proteins known to form non-bonded amyloids. In the cases of immunoglobulin (1bre), serum amyloid P component (1sac) and transthyretin (1bmz) there is more than one potential candidate motif which may undergo swapping to yield amyloids. For each case, we list the more likely one and its alternatives where applicable, according to the criteria listed in the text. (**a**) The red building block of cystatin C (1a67) is the motif; The coil and an α-helix at the NH_2 terminus fall into an unassigned region. The yellow building block contains a three-β-stranded (strands A, B and C) structure, with an unassigned coil connecting it to the red building block, composed of a β-hairpin (strands D and E) at the carboxy terminus. Table I illustrates that the red building block is separated from the rest of the protein by a 26-residue-long coil. The red β-hairpin may flip and domain swap to interdigitate and interact with the previous monomer in the protofibril polymer. (**b**) The 71-residue amyloidogenic fragment (Ratnaswamy *et al.*, 1999) of gelsolin (1d0n) was analyzed. The red building block is the motif; a large portion of this structure is left unassigned. The short, unassigned α-helix at the NH_2 terminus connects into the two anti-parallel β-strand (strands A and B) red building block. A coil connect the red building block to the next, yellow, building block, consisting of an α-helix and a β-strand (strand C). As shown in Table I, the red building block is separated from the rest of the protein by a four-residue-long coil at its N-terminus and a four-residue-long coil at its C-terminus. (**c**) The red building block of immunoglobulin (1bre) is the motif and the purple and blue building blocks are alternative motifs; the purple building block (strands A, B and C) leads to the magenta building block (strands D and E), followed by an unassigned long coil connecting the magenta building block to the blue building block (strands F and G). The blue building block connects to the red building block (strands H and I). The red building block is the candidate motif to swap. Although the purple building block is at the N-terminus and at the edge of the protein structure, as seen from the figure it is not a β-hairpin motif. It consists of two strands on one side of the turn and one strand on the other side of the turn. The candidate motif, the red building block, is separated from the rest of the protein by one residue at its NH_2-terminus and is at the COOH-terminus. (**d**) The red building block of serum amyloid P component (1sac) is the motif. The yellow and the blue building blocks are alternative motifs. There is no direct evidence that SAP forms amyloids. There are data indicating that it binds to amyloid fibrils and is universally present in amyloid deposits (Emsley *et al.*, 1994; Gewurz *et al.*, 1995). It is also known to be resistant to proteinases. This binding to different forms of amyloids may indicate that the binding mechanism of this protein is also similar to the general mechanism of amyloid propagation proposed here. Part of the protein is left unassigned. The motif is separated from the rest of the protein by a six-residue coil at its NH_2 and one residue at the COOH termini. SAP is not needed for amyloid formation. *In vivo* it may, however, be a part of the fiber. (**e**) The red building block of transthyretin (1bmz) is the motif. The blue and yellow building blocks are alternative motifs. The N-terminal unassigned strand A connects to the blue building block (strands B and C), leading to the yellow building block (strands E and F) via a long coil. The coil contains the three-residue flexible strand D (not shown in the picture). The yellow building block leads to the red building block (strands G and H), at the C-terminus of the protein. The red building block is separated from the rest of the protein by a six-residue coil. (**f**) The red building block of β_2-microglobulin (1bmg) is the motif. The structure consists of the NH_2-terminal red building block (strands A and B) and a green building block (strands C, D, E and F), that leads to an unassigned strand and to the C-terminus. The red building block is at the terminus of the protein and has a higher potential to flip than the green β-hairpin. The motif is separated from the rest of the protein by a six-residue coil. (**g**) The red building block of lysozyme (1lz1) is the motif. Most of the protein is unassigned into building blocks, indicating its high flexibility/instability. The N-terminus unassigned β-strand and α-helix connects to the green building block. The green building block leads to the red building block via an unassigned coil and a β-strand. The red building block connects to the long, flexible blue coil via an unassigned β-strand. The blue coil leads to the C-terminus via three α-helices. There is experimental evidence suggesting that the motif swings away from the rest of the protein (Sunde and Blake, 1998), consistent with our proposal. The motif is separated from the rest of the protein by one residue at its NH_2-terminus and three residues at its COOH-terminus. (**h**) The red building block of α_1-antitrypsin (1psi) is the motif. A large portion is left unassigned, suggesting that it is likely to be conformationally unstable. The motif is separated from the rest of the protein by a three-residue-long coil. (**i**) The red building block of prion (1ag2) is the motif. Experimental evidence suggests that the amyloid form of the prion has a larger extent of β-content than the native structure. Since the red building block is a candidate motif to flip using our structural criteria, it is not inconceivable that the coil and part of the helix may undergo a conformational change to form β-strands. Under such circumstances, the amyloid form may assume a characteristic β-hairpin motif conformation similar to other cases. The motif is separated from the rest of the protein by a 14-residue coil. Here we analyze the NMR structure (residues 121–231). The flexible N-terminal part of the prion is therefore not included. We have also looked at cystic fibrosis transmembrane conductance regulator (CFTR) (PBD code: 1nbd). Here we provide a description of the structure and a likely motif, consistent with the stipulated requirements. However, since the PDB structure is a theoretical model, rather than an experimentally determined structure, we cannot present the assignment results. The β-hairpin motif just before the C-terminal α-helix appears a reasonable candidate motif. This motif is separated from the rest of the protein by a 12-residue coil at the NH_2-terminus and six residues at its COOH-terminus of the motif. In all cases, in parentheses are the PDB codes (Bernstein *et al.*, 1977). The assignment algorithm is described in Tsai *et al.* (1999a).

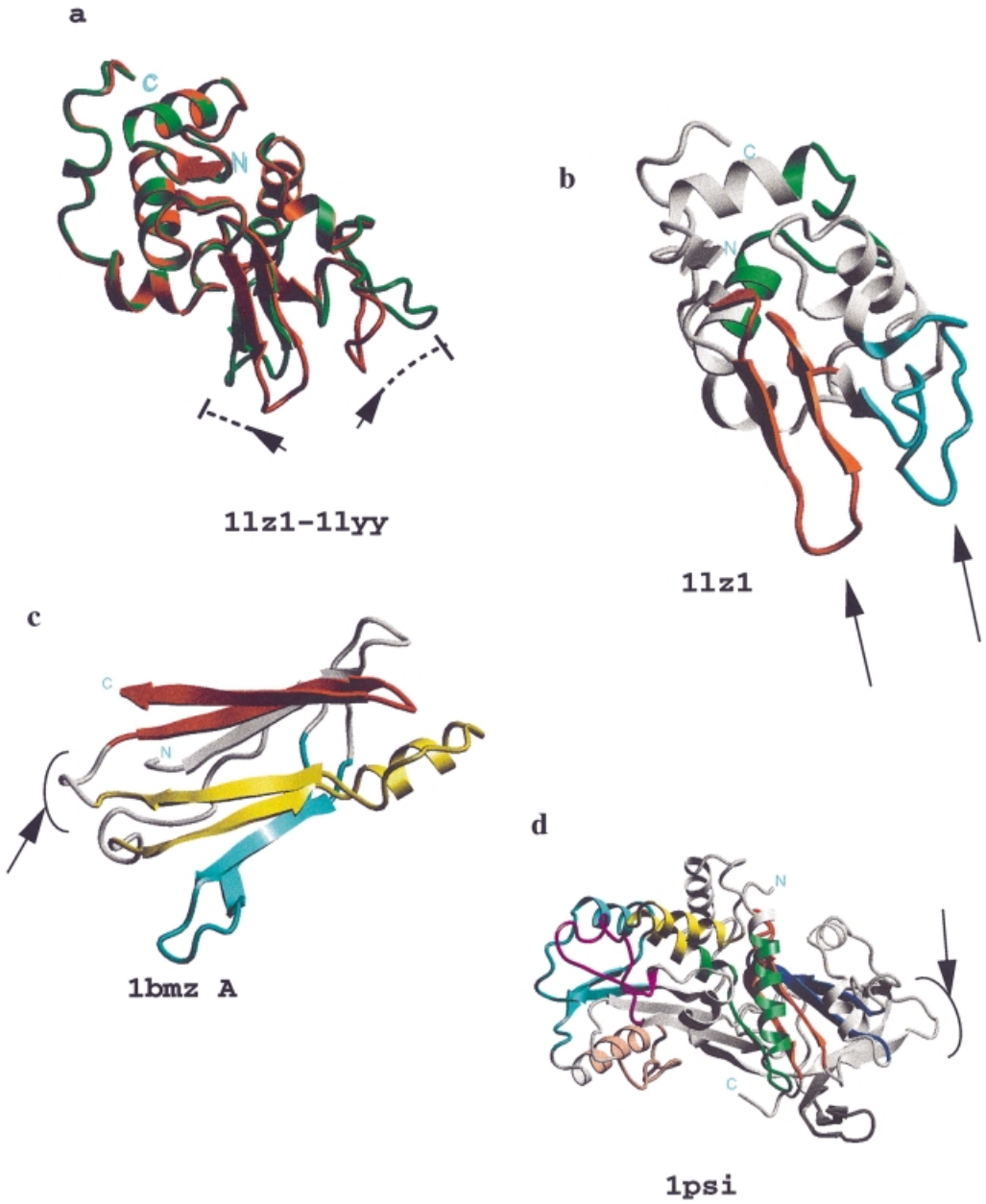


Fig. 3. The deviations between the wild-type and the mutational variants which undergo polymerization. (**a**) Superposition of wild-type (1lz1) and amyloidogenic variant (1lyy) of lysozyme. The arrow points to the region where the motif swings away in the amyloidogenic variant. The figure was generated using the Geometric Hashing algorithm (Nussinov and Wolfson, 1991; Bachar *et al.*, 1993), which considers protein structure as a collection of points, C_αs in this case, and compares the two structures in a sequence-independent manner. (**b**)–(**d**) Arrows point to the region which connects the motif to the rest of the protein and where the residues have the first or second largest deviations: (b) lysozyme; (c) transthyretin and (d) α_1-antitrypsin.

dynamic simulation has illustrated that the amino terminus unassigned region of cystatin C, including the helix and the first strand, is an excellent candidate to (partially) unfold (B.Ma and R.Nussinov, unpublished results).

Naturally occuring variants of cystatin C, gelsolin, immunoglobulin, transthyretin, lysozyme, prion and α_1-antitrypsin are known. In most of the cases the wild-type itself may form polymeric amyloids, with the variants being more prone to undergo such a conformational conversion than their wild-type counterparts, as in transthyretin. On the other hand, for some of these proteins only the variants, most of the time with single mutations, polymerize and the wild-type does not, as in the cases of lysozyme and gelsolin. In α_1-antitrypsin both the wild-type and the mutant form fibrils with similar efficiencies. We have inspected the residue by residue r.m.s.d.s of corresponding C_α pairs of the mutational variants versus their respective native wild-type folds, where crystal structures are available. Although the results of structural superpositioning of the wild-type and the variant proteins show that overall the structures are similar, there are remarkable observations which are consistent with our model and the candidate motif swapping. The results are shown in Table IV. Superimposing the native (wild) type and the amyloidogenic variant of lysozyme [Figure 3(a)] shows that the red motif which is present in the

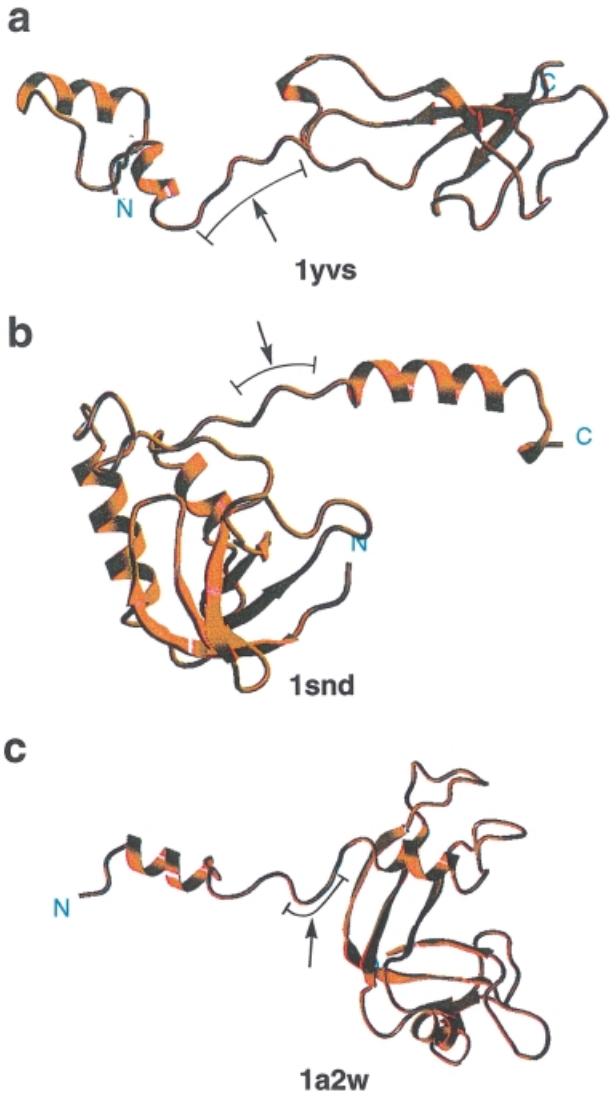


Fig. 4. The deviations between the wild-type and mutational variants of domain-swapping cases. Arrows point to the region which connects the swapping domain to the rest of the protein and where the residues have the first or second largest deviations. Chain A of the dimers/trimers is shown instead of the monomers for the sake of clarity. (**a**) Barnase; (**b**) staphylococcal nuclease; (**c**) BP RNase A.

wild-type swings away from the rest of the protein in the amyloidogenic variant, consistent with the potential of the red building block to flip. The r.m.s.d. measurements reflect the visual observation [Figure 3(b)]. In the case of transthyretin we have inspected nine variants whose crystal structures are available (Table IV). Consistently, in eight of these, the first or the second largest C_α pair deviations are in the coil which connects the red building block motif to the rest of the protein [Figure 3(c)]. In the remaining mutant the largest deviation is in the coil which connects the yellow building block to the rest of the structure. The yellow building block in transthyretin ranks second as a potential candidate motif according to our criteria. In variants of α_1-antitrypsin, the largest deviations are in the long coil region which connects the red C-terminus β-hairpin building block, the potential motif to flip by our criteria, to the rest of the protein via a small β-strand [Figure 3(d)]. These results clearly show that the higher deviations between the variants and the native structures are in the coil which connects the motif to the rest of the protein.

Hence a potential swapped domain in an amyloidogenic protein is a fragment of the structure which has been assigned as a building block, is frequently, although not always, a β-hairpin, has few salt bridges and hydrogen bonds at its interface, is often at the terminus of the sequence and is largely exposed to the solvent. It is identified through the assignments of the building blocks and the electrostatic parameters. If mutant structures are available, a larger deviation is likely to be observed at the proposed hinge site. Analysis of the crystallographic *B* factors illustrates no correlation between the positions of the largest deviations and the largest *B* factor values.

Comparisons with domain-swapped cases

We have inspected the monomers and the dimers of six domain swapped cases. The cases are enumerated in Table II. The sizes of the swapped domains, with the exception of diphtheria toxin (158 residues), are similar to the motifs in Table I, ranging between 13 and 31 residues. The corresponding results of the calculations of the salt bridges, hydrogen bonds and buried non-polar surface areas within the swapped parts and between the swapping segments and the remainder of the structures are given below. The calculations are carried out for the native, unswapped monomers, so that a comparison can be made between the swapping domains in domain swapped cases and the motifs in our set of proteins undergoing polymerization. Three cases, interleukin-10, odorant binding protein and the Eps-SH3 dimer, are 'quasi-domain swapped' cases (Schlunegger *et al.*, 1997). These do not exist as monomers. In our definition, we would consider them two-state folding/binding cases (Tsai *et al.*, 1997b, 1998). The crystal structures of the monomers of BS-RNase and α-spectrin are not available. Hence, we have included these and the 'quasi-domain swapped' cases in the listing in Table II. However, we were unable to calculate the respective monomer values of the salt bridges, hydrogen bonds and non-polar buried surface areas.

Several points should be highlighted here. First, the swapped region is consistently at the N-terminus or at the C-terminus of the protein. Second, there is a relatively long, flexible coil connecting the motif to the remainder of the structure, allowing it to flip. Third, in none of these cases is the swapped part a β-hairpin motif. Instead, it is an α-helix, a strand or an entire domain. In this regard, cystatin C is of particular interest. There is experimental evidence that cystatin C forms an SDS-resistant dimer (Wei *et al.*, 1998). This may suggest a domain-swapped dimer. In the dimer, the interdigitation due to the swapping may render the dimer SDS resistant. Fourth, the number of salt bridges and of hydrogen bonds connecting the swapped domain to the remainder of the structure is small. A larger number is frequently observed within the swapped domain (Table III). Fifth, the buried non-polar surface area within the swapped domain is usually (with the exception of CksHs and interleukin-5) larger than that observed between the motif and the remainder of the structure. In particular, sixth, the absolute value of the buried non-polar surface area between the motif and the remainder of the structure is variable and can be large (Table III). This, however, does not prevent the domain-swapping event from taking place.

We have also inspected the residue by residue r.m.s.d.s of corresponding C_α pairs from the native and respective mutants where structures are available. The results are listed in Table V and depicted in Figure 4. In the barnase mutants the first

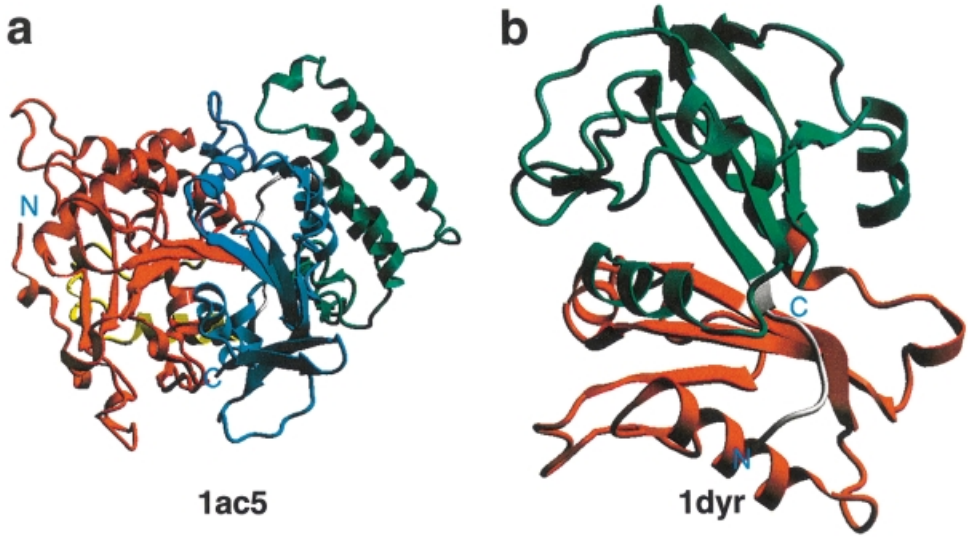


Fig. 5. Two examples of continuation of β-sheets across the interfaces between independently folding hydrophobic units (HFUs). (**a**) Carboxypeptidase (1ac5); (**b**) oxidoreductase (1dyr).

Table II. A listing of the sequences, the sizes and the locations of the structural components undergoing domain swapping

Protein	Protein size (residues)	Sequence of the swapping domain (SD)	Size[b]	No. of residues prior to/following SD
BS-RNase (1bsr)	124	KE*SAAAKFERQHM*DSGNSPSS	13	0/8 (*N*-terminus)
Bovine pancrea--tic RNase A (1rph)	124	KE*TAAAKFERQHM* DSSTSAASSS	13	0/10 (*N*-terminus)
Ckshs (1dks)	79	EPEP**HILLFRR** PLPKKP	13	4/0 (*C*-terminus)
Staphyloco-ccal nuclease (1snc)	149	GLAKVAYVYKPNNT*HEQHLRK SEAQAKKE*KINIWS	21	14/0 (*C*-terminus)
Diphtheria toxin (1mdt)	535	RPAYSPGHKTQP**FLH**DG**YAVSW** NTVEDSI**IR**TGFQG**ESGHDIKITA** ENTPLPIA**GVLLP**TIPGKL DVNKSKT**HISV**NG**RKIRM** **RCRAI**DGD**VTFCRP**KSPVYVGNG VH**ANLHVAFHR**SSSEKIHSNEISSD **SIGVLGYQKTV**DH**TKVNSK** **LSLFFEIK**S	158	12/0 (*C*-terminus)
Barnase (1a2p)	110	VINT*FDGVADYLQTY*HKLPD NYIT*KSEAQAL*GWVASGN	31	0/7 (*N*-terminus)
Interleukin-5 (2gmf)	127	CPPTPETSCA**TQIIT** *FESFKENLKDFLLV*IPFDCWEP	27	10/0 (*C*-terminus)
α-spectrin (1shg)	62	QNHYA*SNLVDEKRKQ VLERWRHLKEGLIEKRSRLGD*	31	5/0 (*C*-terminus)
Interleukin-10 (1ilk)	151	HRFLPCENK*SKAVEQVKNAFNKL QEKGIYKAMSEFDIFINYIEAYMTM KIR*N	43	9/0 (*C*-terminus)
Odorant binding protein (1obp)	159	LNVE*DEDLEKFWKLTED* KGID *KKN***VVNFLENEDHPHPE**	17	4/0 (*C*-terminus)
Eps8-SH3 (1aoj)	60	**KKYAK**SKYDFVARNSSELSVMKDD **VLEIL**DDRRQ	28	0/6 (*N*-terminus)

β-Strands are shown in bold. α-Helices are shown in italics. The sequences of the coils connecting the swapping domain to the rest of the protein are underlined. The last column specifies the location and the length of the coils connecting the domain to the remainder of the structure.
[a]Size of the swapping domain (No. of residues).

or second largest deviations are in the coil which connects the swapping domain to the rest of the protein [Figure 4(a)]. Similarly, in staphylococcal nuclease the second largest deviations are found in the coil which connects the swapping domain to the rest of the protein [Figure 4(b)]. As shown in Table V, higher deviations are also found in a long coil connecting the α-helix to the β-strand. Long flexible coils usually show larger deviations. Still, this latter coil is in the middle of the structure. Thus, a flipping event of the region connected to this long coil is considerably more difficult. In the BP-RNase A mutants, the first or second largest deviations again lie in the coil which connects the swapping domain to the rest of the protein [Figure 4(c)]. These results are consistent with those of the candidate motifs in proteins known to undergo non-bonded polymerization (Table IV). There, too, the deviations of the mutational variants are higher in the coils connecting the motifs to the remainder of the structure.

Discussion and conclusions

Inspection of the native structures of amyloidogenic proteins shows them to be variable. Yet, on the other hand, X-ray fiber diffraction analysis of the polymers indicates similar, characteristic structures common to all, including β-sheets arranged on an axis running parallel to the protofibril and a variable distance between two pairs of β-sheets across the protofibril axis. Additionally, all non-bonded amyloid polymers are known to be remarkably stable. In approaching the problem, we adhered to the basic notion that mechanisms in nature are general, except that in some cases the results are more extreme than in others.

Conformational low-energy transitions involving hinge-based motions are frequently observed in proteins. In native proteins they are critically important for binding, catalysis and motility. Hinge-bending motions can involve movements of fragments or of larger domains. In general, the binding interface between these and the rest of the protein does not involve tight packing (Gerstein *et al.*, 1994). A more extreme case of hinge-bending motions is illustrated in domain swapping. A revealing example is bovine seminal ribonuclease, where domain swapping may take place after hours or days (Piccoli

Table III. A comparison of the number of salt bridges (SB), hydrogen bonds (HB) and non-polar buried surface area (NPBSA) within the motifs and at their interfaces, and between the β-hairpin motif in the amyloidogenic proteins (left part) and in the swapping domain proteins (right part) and the remainder of the protein

Amyloidogenic proteins					Domain swaping cases				
Protein	Parameter	Total	Motif	Interface	Protein	Parameter	Total	Motif	Interface
Cystatin C	SB	1	1	0	Barnase	SB	6	0	1
	HB-MM	15	4	0		HB-MM	32	17	1
	HB-MS	151	28	0		HB-MS	47	11	3
	HB-SS	5	1	0		HB-SS	11	0	3
	NPBSA		1948	799		NPBSA		2705	1721
Gelsolin	SB	3	0	1	Staphylococcal nuclease	SB	6	1	1
	HB-MM	21	2	0		HB-MM	53	16	0
	HB-MS	16	2	2		HB-MS	31	0	0
	HB-SS	3	0	1		HB-SS	7	0	0
	NPBSA		1201	1375		NPBSA		1419	997
Immunoglobulin	SB	4	1	0	Bovine pancr--tic RNase A	SB	3	1	0
	HB-MM	31	4	0		HB-MM	32	10	0
	HB-MS	16	4	0		HB-MS	37	3	5
	HB-SS	3	1	0		HB-SS	6	2	1
	NPBSA		1235	628		NPBSA		806	664
Serum amyloid P component	SB	11	0	0	Ckshs	SB	4	0	0
	HB-MM	50	5	2		HB-MM	19	2	0
	HB-MS	53	6	4		HB-MS	15	0	0
	HB-SS	13	1	0		HB-SS	5	0	0
	NPBSA		1374	1760		NPBSA		751	1300
Transthyretin	SB	14	1	0	Diphtheria toxin	SB	25	6	0
	HB-MM	24	3	0		HB-MM	222	34	0
	HB-MS	26	3	0		HB-MS	117	25	0
	HB-SS	3	0	0		HB-SS	19	9	0
	NPBSA		1411	1151		NPBSA		15272	2218
β-2 micro-globulin	SB	6	1	1	Interleukin-5	SB	5	2	1
	HB-MM	0	0	0		HB-MM	60	11	0
	HB-MS	12	1	2		HB-MS	30	4	0
	HB-SS	2	0	0		HB-SS	2	0	0
	NPBSA		2506	1895		NPBSA		1993	2296
Lysozyme	SB	6	1	0					
	HB-MM	69	5	2					
	HB-MS	53	7	2					
	HB-SS	13	2	1					
	NPBSA		731	546					
α-1-antitrypsin	SB	16	2	0					
	HB-MM	139	9	0					
	HB-MS	72	3	0					
	HB-SS	18	1	0					
	NPBSA		1082	990					
Prion	SB	6	2	1					
	HB-MM	46	7	0					
	HB-MS	25	12	2					
	HB-SS	0	0	0					
	NPBSA		3511	1656					

MC are the main chain–main chain hydrogen bonds, MS are the main chain–side chain hydrogen bonds and SC are the side chain–side chain hydrogen bonds. For the salt bridges and hydrogen bonds the values calculated are for the total protein, the number within the motif and the number at the interface bewteen the motif and the rest of the structure. For the non-polar buried surface area, the values are given in $Å^2$ for the area buried within the motif and at the interface between the motif and the remainder of the structure. The native structures are used. For comparison, below we list the values obtained for alternative building blocks in the cases of immunoglobulin (1bre), serum amyloid P component (1sac) and transthyretin (1bmz). The colors are as shown in Figure 2. The building blocks are enumerated in their sequential order. In immunoglobulin (Figure 2c), the purple building building does not contain any salt bridges within it and has one salt bridge between it and the rest of the structure. It has eight hydrogen bonds (HB) within it and seven between it and the rest of the structure. The blue block has one salt bridge within and one connecting it to the rest of the structure and two HB within and none in between. In serum amyloid P component (Figure 2d), the yellow block has one salt bridge within it and none between it and the rest of the structure and five HB within and two in between. The blue building block has one salt bridge within it and none between it and the rest of the structure and eight HB within and 13 in between. In transthyretin (Figure 2e), the blue building block has one salt bridge within it and two salt bridges connecting it to the rest of the structure and two HB within and four in between. The yellow block has three salt bridges within it and no connecting ones. It has 14 HB within and two in-between. With regard to the non-polar buried surface area, in immunoglobulin the purple building block buries 1401 $Å^2$ within it and 1549 $Å^2$ between it and the rest of the structure; the blue building block buries 553 $Å^2$ within and 1106 $Å^2$ in between. In serum amyloid P component, the yellow block buries 410 $Å^2$ within and 1343 $Å^2$ in between; the blue block buries 1584 $Å^2$ within and 1780 $Å^2$ in between. In transthyretin the blue building block buries 970 $Å^2$ within and 1740 $Å^2$ in between; the yellow block buries 2749 $Å^2$ within and 1582 $Å^2$ in between.

Table IV. A listing of the largest and second largest deviations obtained when superimposing the mutants on their native wild-type structures for amyloidogenic proteins

Protein (PDB ID)	Rmsd of the protein (Å)	Largest deviation (residue wise) (Å)	Res. pos.[1]	2nd largest deviation (residue wise) (Å)	Res. pos.[2]
Transthyretin variants (wild type-1bmz)					
1tsh-A60T	0.24	0.80	Ser 85[a]	0.52	Glu 63[b]
1ttc-V30M	0.33	0.71	Met 30[c]	0.68	Ser 100[d]
1ttb-A109T	0.23	0.58	Cys 10[e]	0.56	Ser 100[d]
2trh-R10C	0.23	0.70	His 56[b]	0.49	Ser 100[d]
2try-S77Y	0.26	0.70	Gly 101[d]	0.58	Ser 100[d]
1bzd-S6G	0.20	0.71	Ser 100[d]	0.57	Glu 62[b]
1bze-M119T	0.21	0.62	Ser 100[d]	0.55	Gly 101[d]
1bz8-delV122	0.91	5.29	Pro 102[d]	4.66	Asp 99[d]
1ttr-V122I	0.18	0.58	Ser 85[a]	0.41	Ser 100[d]
Lysozyme variants (wild type- 1lz1)					
1lyy-D67H	1.90	9.52	Pro 71[f]	7.49	Gly 48[g]
1loz-I56T	0.27	0.91	Pro 103[h]	0.62	Gly 48[g]
α-1-antitrypsin variants (wild type- 2psi)					
1atu-F51L,T59A, T68A,A70G,M374I, S381A	2.19	9.37	Ala 355[i]	7.57	Glu 354[i]
1kct-T59A,T68A, A70G	2.60	13.37	Ala 350[i]	12.11	Pro-357[i]
1psi-F51L	0.30	3.60	Ala 348[i]	0.69	Ala 350[i]

The deviations are measured for each corresponding C_α pair. The r.m.s.d.s are listed for each of the variants against their corresponding wild-types. Superscripts show the position of the respective residue in the protein structure.
Res. pos.[1] indicates residue position corresponding to the largest deviation and Res. pos.[2] indicates residue position corresponding to the second largest deviation.
[a]Coil in the yellow β-hairpin.
[b]Coil connecting the blue β-hairpin to the yellow β-hairpin.
[c]First strand of the blue β-hairpin motif.
[d]Coil connecting the motif to the rest of the protein.
[e]N-Terminal coil.
[f]Blue coil connecting the motif to the rest of the protein.
[g]Coil in the red motif.
[h]Coil in unassigned region.
[i]Coil connecting the motif to the rest of the protein via a flexible β-strand.

et al., 1992; D'Alessio, 1995). The rate of the swapping event reflects the population time of the conformation of the monomer in the swapped form.

Inspection of the non-polar buried surface area both in amyloidogenic proteins and in documented domain-swapping cases, shows that it is variable and can be fairly large. Moreover, the analysis presented here illustrates that the number of salt bridges and hydrogen bonds in both amyloidogenic and domain-swapping cases is limited. Consistent with the results obtained here, in an extensive analysis of hinge-bending transitions of known cases, we have observed that while the non-polar buried surface area between the hinging domains can be large, the number of salt bridges and hydrogen bonds is small (N.Sinha, S.Kumar and R.Nussinov, unpublished work).

The X-ray fiber diffraction pattern of the filaments shows

Table V. A listing of the largest and second largest deviations obtained when superimposing the mutants on their native wild-type structures for domain-swapping proteins

Protein (PDB ID)	Rmsd of the protein (Å)	Largest deviation (residue wise)(Å)	Res. pos.[1]	2nd largest deviation (residue wise)(Å)	Res. pos.[2]
Barnase variants (wild type-1a2p)					
1brh-L14A	0.10	0.46	Glu 60[j]	0.34	Gly 40[k]
1bne-A43C, S80C	3.79	5.07	Asn 41[k]	4.88	Ala 32[k]
1bnf-T70C, S92C	1.06	6.22	Gly 68[j]	0.75	Gly 40[k]
1bng-S85C, H102C	0.22	0.50	Lys 66[j]	0.45	Gly 34[k]
1bri-I76A	0.23	0.62	Ser 67[j]	0.49	Gly 48[k]
1brj-I88A	0.18	0.42	Lys 66[j]	0.27	Gly 40[k]
1brk-I96A	0.29	0.75	Ser 67[j]	0.49	Gly 40[k]
1b20-R69S	0.30	1.06	Lys 66[j]	0.38	Gly 34[k]
1b3s-H102A	0.50	1.84	Arg 59[j]	0.76	Gly 40[k]
1btu-K27A	0.50	0.95	Ala 32[k]	0.72	Ser 67[j]
Staplylococcal nuclease variants (wild type-1snc)					
1sno-L124H	0.69	4.47	Pro 47[l]	3.10	Tyr 113[m]
1a2u-V23C	0.19	0.52	Gly 29[n]	0.22	Thr 120[m]
BP-RNase A variants (wild type-1rph)					
1a5q-P93A	0.28	0.71	Lys 37[o]	0.45	Ser 21[p]
3rsk-K7A, R10A, K66A	0.34	1.48	Ser 89[q]	0.60	Ser 22[p]
3rsd-D121N	0.39	1.04 (third largest deviation is 0.64: Ser 21[p])	Asn 67[o]	0.98	Gly 88[q]

The deviations are measured for each corresponding C_α pair. The r.m.s.d.s are listed for each of the variants against their corresponding wild-types. Superscripts show the position of the respective residue in the protein structure.
Res. pos.[1] indicates residue position corresponding to the largest deviation and Res. pos.[2] indicates residue position corresponding to the second largest deviation.
Other footnotes continue on from those in Table IV:
[j]Coil connecting the swapping domain to the rest of the protein via a β-strand.
[k]Coil connecting the swapping domain to the rest of the protein.
[l]Coil between the β-sheet and the first α-helix.
[m]Coil connecting the swapping domain to the rest of the protein.
[n]Coil in the N-terminal β-hairpin.
[o]Coil connecting the swapping domain to the rest of the protein via an α-helix.
[p]Coil connecting the swapping domain to the rest of the protein.
[q]Coil in the β-hairpin.

that they consist of continuous, twisted β-sheets, arranged along the protofibril axis. While it is well known that β-structures are particularly stable, a continuation of β-sheets across the intermolecular interface is unlikely to produce a filament with such an exceptionally high stability. Inspection of protein crystal or NMR structures reveals that twisted β-sheets are frequently continued between independently folding hydrophobic units, across their interface. Such a continuation is observed between domains and between subunits of an oligomer. Figure 5 illustrates two examples, depicting such a twisted β-sheet propagation, between hydrophobic folding units. Yet, despite the extended propagating β-sheets, in all of these cases, motions are observed between these structural components, with melting temperatures within the generally observed ranges, i.e. well below 100°C. On the other hand, an interdigitation of β-hairpins between domains or subunits, locking them together, has not been observed to date in globular, functional, native proteins. Proteins are well known to be only marginally stable. This is essential for their function. Interdigitated, interlocked β-sheets, across domains or subunits, are likely to be conformationally too stable, hindering the motion which is necessary for activity. Hence, there is a selection in nature against such conformations.

Motifs in the interfaces of domains or of hydrophobic folding units, even if otherwise they fulfil the requirements, are still unlikely to swap. An example is the β-hairpin motif in CD4. Here, despite its location at the N-terminus of the sequence, CD4 does not appear to be a candidate for forming a polymeric fibril. Consistently, drugs stabilizing the tetrameric structure of transthyretin prevent polymerization (Peterson *et al.*, 1998). However, while in domain swapping the flipped and swapped domain appears to be practically uniquely at one of the sequence termini, this is not always the case for the swapping of the motif in amyloid formation. In cases where the motif is not at the sequence terminus, we can imagine that a Greek-key like flipping and twisting of the β-hairpin might conceivably take place. This has been suggested to take place within monomeric chains. It is possible that as additional cases of domain swapping are discovered, swapped domains will also be observed not only at a sequence terminus.

Two steps are involved in the polymerization. The first is the conformational change of the native monomer; the second is the binding of the 'open', flipped, monomer to the growing interdigitating polymer. If the barriers for the interconversion step are low, the rate-limiting step might be expected to be the binding, via a diffusion–collision process (Karplus and Weaver, 1976). If the barriers are high, the conformational interconversion may be the rate-limiting step. In the case of the polymerization, the rate of the reaction is very slow. This again is reminiscent of domain swapping (e.g. bovine seminal ribonuclease; Piccoli *et al.*, 1992; D'Alessio, 1995). However, once a seed is introduced, if the concentration is high, the reaction proceeds at a much faster rate (Lomakin *et al.*, 1996; Dobson and Karplus, 1999). Hence here the rate-limiting step involves the formation of a seed polymer. This is similar to the case of supercool water. The water would stay in the liquid state until an ice seed is introduced into it (Tsai *et al.*, 1999b).

Although here we have focused largely on cases in which the β-hairpins are already present in the native forms, this condition does not always hold. A potential example is the case of the prion. Furthermore, it has been shown that depending on the conditions, most proteins can form amyloids, illustrating dynamic landscapes. Even if the β-structure exists, if it is unstable, it may unfold and reform in an alternative way. Additionally, even if a stable β-structure exists it is not necessarily the case that it swaps to form amyloids. Alternative, unstable building blocks may partially unfold, with subsequent participation in swapping. In particular, it is important to note that while here we have largely kept the monomeric structure intact, this is unlikely to hold universally. There is substantial evidence that partially folded conformations are critical intermediates on the pathway to fibril formation.

The validity of our model is further supported by analysis of mutations. The r.m.s.d.s between the native and the mutants are always small (Tables IV and V), as may be expected between such close mutational variants. Nevertheless, when analyzed with respect to specific positions, namely, with respect to matching C_α pairs between the two (native and mutant) proteins, the positions showing the largest deviations (ranking

between first and third) are consistently in the coil regions connecting the motifs to the rest of the structures [Table IV, Figure 3(a)–(d)]. Consistently, the deviations obtained for domain swapping cases are also the largest in the coils connecting the swapping domain to the remainder of the protein [Table V, Figure 4(a)–(c)]. This is despite the fact that the actual location of the mutations varies, suggesting that the regions which are most prone to respond to the sequence alterations are at the proposed hinge region. This may also suggest that these regions are inherently susceptible to the changes in physiological conditions, such as pH changes, which lead to amyloid fiber formation in some wild-type proteins, as in transthyretin. This is also consistent with a large-scale mutational analysis (Sinha and Nussinov, 2001).

Polymerization takes place since the free energy of the bound, polymerized form is lower than that of the unbound, native form of the protein. Here we address the problem of how the barrier from the unbound to the polymerized form is lowered. If, however, polymerized protofilaments are not formed, despite partially denaturing conditions, as in the case of the WW domain (Koepf *et al.*, 1999), we may infer that the free energy of the bound, amyloid form is higher than that of the native form. Koepf *et al.* (1999) suggest that the reason may reside in its strong hydrophobic core. Alternatively, it may suggest that in the polymerized form there are unfavorable interactions, such as exposure of the large aromatic residues to water or steric hindrance either in the formation of the protofilaments or in their assembly to the fibrillar structure.

What are the conditions for a β-hairpin to be able to swap? First, the β-hairpin motif should preferably be at the edge of the structure, rather than buried within it. Second, it should not have salt bridges or too many hydrogen bonds, connecting it to the remainder of the structure. While a smaller extent of buried surface area would lower the barrier for the swapping, swapping would eventually take place with a more extensive buried surface area, as long as the polymerized form is more stable than the native conformation. This is evident from domain swapped cases. Third, it should preferably be at the amino or carboxy termini, resulting in a single hinge. In this regard, it is revealing to inspect concanavalin A (PDB: 1jbc). Although the structure of concanavalin A is very similar to that of serum amyloid P component (SAP), it does not bind to non-bonded polymeric amyloid fibrils, like SAP. Assigning concanavalin A into its constituent building blocks and comparing it to serum amyloid P component (Figure 2) we notice two interesting points. First, the NH_2- and COOH-termini of concanavalin A are in the middle of a β-sheet. However, in contrast, in the case of serum amyloid P component, they are at the edge of sheet. Second, the building block assignment results suggest that flipping a β-hairpin at the other end of the β-sheet appears to be easier for SAP than for concanavalin A. That is, concanavalin A is more stable than serum amyloid P component at the edge of their two seven-stranded β-sandwich.

Acknowledgements

We thank Drs Sandeep Kumar, Buyong Ma and, in particular, Jacob V.Maizel for numerous helpful and insightful discussions. Advanced Biomedical Computing Center personnel are thanked for computational resources and related assistance. The personnel at FCRDC are thanked for their assistance. The research of R.Nussinov in Israel has been supported in part by grant number 95-00208 from BSF, Israel, by a grant from the Israel Science Foundation administered by the Israel Academy of Sciences, by a Magnet grant, by a Ministry of Science grant and by Tel Aviv University Basic Research grants and by the Center of Excellence, administered by the Israel Academy of Sciences. This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number NO1-CO-56000. The content of this publication does not necessarily reflect the view or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organization imply endorsement by the US Government.

References

Bachar,O, Fischer,D, Nussinov,R. and Wolfson,H. (1993) *Protein Eng.*, **6**, 279–289.

Bennett,M.J., Choe,S. and Eisenberg,D. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 3127–3131.

Bennett,M.J., Schlunegger,M.P. and Eisenberg, D. (1995) *Protein Sci.*, **4**, 2455–2468.

Bernstein,F., Koetzle,T., Williams,G., Meyer,E.J., Brice,M., Rodgers,J., Kennard,O., Shimanuchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Chiti,F., Webster,P., Taddei,N., Clark,A., Stefani,M., Ramponi,G. and Dobson,C.M. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 3590–3594.

Chothia,C. (1973) *J. Mol. Biol.*, **75**, 295–302.

D'Alessio,G. (1995) *Nature Struct. Biol.*, **2**, 11–13.

Dobson,C.M. and Karplus,M. (1999) *Curr. Opin. Struct. Biol.*, **9**, 92–101.

Emsley,J. *et al.* (1994) *Nature*, **367**, 338–345.

Fink,A.L. (1998) *Fold. Des.*, **3**, R9–R23.

Gerstein,M., Lesk,A. and Chothia,C. (1994) *Biochemistry*, **33**, 6739–6749.

Gewurz,H., Zhang,X.H. and Lint,T.F. (1995) *Curr. Opin. Immunol.*, **7**, 54–64.

Guijarro,J.I., Sunde,M., Jones,J.A., Campbell,I.D. and Dobson,C.M. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 4224–4228.

Inouye,H., Domingues,F.S., Damas,A.M., Saraiva,M.J., Lundgren,E., Sandgren,O. and Kirschner,D.A. (1998) *Amyloid: Int J. Exp. Clin. Invest.*, **5**, 163–174.

Karplus,M. and Weaver,D.L. (1976) *Nature*, **260**, 404–406.

Kelly,J.W. (1996) *Curr. Opin. Struct. Biol.*, **6**, 11–17.

Kelly,J.W. and Lansbury,P.T. (1994) *Amyloid: Int. J. Exp. Clin. Invest.*, **1**, 186–205.

Kirschner,D.A., Elliott-Bryant,R., Szumowski,K.E., Gonnerman,W.A., Kindy,M.S., Sipe,J.D. and Cathcart,E.S. (1998) *J. Struct. Biol.*, **124**, 88–98.

Koepf,E.K., Petrassi,H.M., Sudol,M. and Kelly,J.W. (1999) *Protein Sci.*, **8**, 841–853.

Kumar,S. and Nussinov,R. (1999) *J. Mol. Biol.*, **293**, 1241–1255.

Lansbury,P.T. *et al.* (1995) *Nature Struct. Biol.*, **2**, 990–998.

Litvinovich,S.V., Brew,S.A., Aota,S., Akiyama,S.K., Haudenschild,C. and Ingham,K.C. (1998) *J. Mol. Biol.*, **280**, 245–258.

Lomakin,A., Chung,D.S., Benedek,G.B., Kirschner,D.A. and Teplow,D.B. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 1125–1129.

Nussinov, R. and Wolfson, H.J. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 10495–10499.

Peterson,S.A., Klabunde,T., Lashuel,H.A., Purkey,H., Sacchettini,J.C. and Kelly,J.W. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 12956–12960.

Piccoli,R., Tamburrini,M., Piccialli,G., Di Donato,A., Parente,A. and D'Alessio,G. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 1870–1874.

Ratnaswamy,G., Koepf,E., Bekele,H., Yin,H. and Kelly,J.W. (1999) *Chem. Biol.*, **6**, 293–304.

Saraiva,M.J.M. (1995) *Hum. Mutat.*, **5**, 191–196.

Schlunegger,M.P., Bennett,M.J. and Eisenberg,D. (1997) *Adv. Protein Chem.*, **50**, 61–122.

Serpell,L.C., Sunde,M., Fraser,P.E., Luther,P.K., Morris,E. and Sandgren,O. (1995) *J. Mol. Biol.*, **254**, 113–118.

Serpell,L.C., Sunde,M. and Blake,C.C.F. (1997) *Cell Mol. Life Sci.*, **53**, 871–887.

Sinha,N. and Nussinov,R. (2001) *Proc. Natl Acad. Sci. USA*, **98**, 3139–3144.

Sipe,J.D. (1992) *Annu. Rev. Biochem.*, **61**, 947–975.

Shirahama,T. and Cohen,A.S. (1967) *J. Cell Biol.*, **33**, 679–706.

Sunde,M. and Blake,C.C.F. (1998) *Rev. Biophys.*, **31**, 1–39.

Tsai,C.J. and Nussinov,R. (1997a) *Protein Sci.*, **6**, 24–42.

Tsai,C.J. and Nussinov,R. (1997b) *Protein Sci.*, **6**, 1426–1437.

Tsai,C.J., Xu,D. and Nussinov,R. (1998) *Fold. Des.*, **3**, R71–R80.

Tsai,C.J., Maizel,J.V. and Nussinov,R. (1999a) *Protein Sci.*, **8**, 1591–1604.

Tsai,C.J., Kumar,S., Ma,B. and Nussinov,R. (1999b) *Protein Sci.*, **8**, 1190–1190.

Tsai,C.J., Maizel,J.V. and Nussinov,R. (2000). *Proc. Natl Acad. Sci. USA*, in press.

Xu,D., Tsai,C.J. and Nussinov,R. (1998) *Protein Sci.*, **7**, 533–544.

Wei,L., Berman,Y., Castano,E.M., Cadene,M., Beavis,R.C., Devi,L. and Levy,E. (1998) *J. Biol. Chem.*, **273**, 11806–11814.

Received July 15, 2000; revised September 20, 2000; accepted November 8, 2000