doi:10.1016/j.jmb.2012.04.015 J. Mol. Biol. (2012) 421, 417–426

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

REVIEW

On the Possible Amyloid Origin of Protein Folds

Jason Greenwald and Roland Riek*

ETH Zurich, Physical Chemistry, ETH Honggerberg, 8093 Zurich, Switzerland

Received 16 November 2011; received in revised form 17 April 2012; accepted 17 April 2012
Available online 24 April 2012

Edited by S. Radford

Keywords:
amyloid;
prion;
cross-β-sheet structure;
protein aggregation;
origin of life

The diversity of protein folds is derived from the diversity of the underlying proteome. Such diversity must have originated from a so-called common ancestor: a hypothetical fold whose identity will, in all likelihood, never be known. Nonetheless, hypotheses exist to explain the evolution of protein folds. When formulating such hypotheses as done here, the entire repertoire of polypeptide structure, from well-defined tertiary structures and molten globule states to intrinsically disordered proteins and oligomeric aggregates, is worth considering. It is the aim of this short essay to discuss the hypothesis that one type of protein aggregate—the cross-β-sheet motif—was the first functional protein fold, that is, the common ancestor fold. Support for this hypothesis comes from the observations that (i) short peptides with simple amino acid sequences are able to form the cross-β-sheet structure, (ii) amyloids can be very stable under harsh conditions, (iii) amyloids can self-assemble in complex mixtures, (iv) amyloids have many potent activities that are attributable to the inherent repetitiveness of the structure, and (v) the proteomes of modern organisms appear to have evolved away from the more amyloidogenic sequences of older organisms, suggesting that amyloids were more ubiquitous earlier in the evolution of modern protein folds.

© 2012 Elsevier Ltd. All rights reserved.

Introduction

Life is based on the structure and dynamics of biomolecules and the interplay between them.[1] The complexity of these molecules, particularly proteins, highlights the mature relationship that has evolved between structure and function. Take, for example, the ribosome, a structure composed of hundreds of thousands of spatially ordered atoms comprising several polypeptides and RNA molecules that together form an entity with the level of sophistication required to carry out nucleic acid template-directed protein synthesis.[2] Looking at this feat of nanotechnology, one is led to wonder from where it all came. Yet, just as life's marvels are expected to have evolved from much less complex prebiotic systems, so should the underlying biomolecules have arisen from much simpler precursors.[3]

Unfortunately, there is no fossil or genetic record leading back to the time, more than 3.5 billion years ago, when sustainable life first occurred on our planet,[4] so we can only speculate about the details of the first protein fold(s). However, speculation by way of deduction leads to some reasonable conclusions about the nature of the common ancestor (CA) fold(s). We start by focusing on a specific question: What are the possible origins of the first protein fold(s)? In doing so, we largely circumvent a discussion about the origin of life. While we find many compelling arguments in "protein first"[5,6] and "amyloid world"[7,8] hypotheses for the origin of life, our focus on proteins is purely etiological. Nonetheless, the question of life's origins has an important influence on our discussion: Did proteins evolve from a single CA fold or were there several,

*Corresponding author. E-mail address: roland.riek@phys.chem.ethz.ch.

Abbreviations used: CA, common ancestor; PDB, Protein Data Bank; IDP, intrinsically disordered peptide (or protein).

0022-2836/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.

and by what mechanisms did the CA fold(s) evolve? If life arose in accordance with the amyloid world hypothesis, with the essential catalytic and reproductive functions required for life carried out by peptide amyloids, then by inference there was a single CA fold, that is, a structural state adopted by all of the peptides in the first living system. However, it can also be envisioned that a set of CA folds evolved in an interdependent manner and only through their interactions create a viable system. Alternatively, peptides/proteins may have only first emerged as components of a preexisting life form, and hence, the CA fold(s) evolved in a manner analogous to modern molecular evolution. In the latter case, it is possible that unique CA folds evolved simultaneously either inter- or independently, or that there was a single species from which the others evolved.

In any case, it is safe to assume that the first protein fold(s) must have been simple. The first functional peptides that accumulated to an appreciable extent were most likely short and not capable of a complex folding architecture regardless of whether they existed before or only first after the appearance of life, as has been proposed for the "RNA world" hypotheses.[9,10] Furthermore, we limit consideration for a CA fold candidate to a fold that provided a selective advantage (either for itself or for the emerging organism it was part of), thus enabling more of its own replication (either directly or indirectly through the organism). Following this argument, a peptide without a function could not have been selected for and is therefore a peptide without significance in the evolution of peptide folds. Because early (prebiotic or the first biotic) peptide synthesis could not have been of a high efficiency or fidelity, large proteins would only have evolved after their shorter precursors showed some fitness that could drive the selection of an improved peptide synthesis mechanism. It is therefore assumed that short functional peptides, composed only of the amino acid residues available from abiotic syntheses, are the protobiological precursors of proteins. The CA fold(s) must have been able to tolerate significant sequence variation both because of the expected low fidelity of a primitive protein synthesis mechanism and because of the need to be adaptable in the early stages of evolution (either to directly or indirectly support the viability of the primitive organism). A fold that could adapt to gain new functions with only small sequence modifications or a fold that was already multifunctional in nature would have a clear selective advantage. These conservative assumptions limit the conformational space of the first functional protein fold(s) to (i) small structured globular domains, (ii) molten-globule-like peptides, (iii) intrinsically disordered peptides (or proteins) (IDPs), (iv) helical peptides, (v) membrane-interacting peptides, (vi) peptide amyloids (and other β-structured aggregates), or (vii) a combination of the above within a single peptide/protein. In the following, we will take each of these possibilities into consideration, placing particular emphasis on the peptide amyloids.

Small Globular Protein Folds and Molten Globule States

We begin with the question, "What is the smallest stable protein fold?" One survey of the Protein Data Bank (PDB) reported that the minimum monomeric domain size is around 35–40 residues.[11] This observed size limit is probably close to a real minimum because a stably folded domain needs a buried hydrophobic core to overcome the entropy loss associated with achieving a single backbone conformation. Taking a few of the smallest domains as examples, the epidermal growth factor and potato carboxypeptidase inhibitor folds have multiple disulfide bonds and the WW and villin headpiece domains have several aromatic residues in their small hydrophobic cores. These specific elements help stabilize a defined structure in a very short peptide, but the prebiotic existence of cysteine and amino acids with aromatic side chains is not supported by existing evidence (in order of decreasing abundance, the amino acids reported to have been observed in abiotic syntheses are as follows: Gly, Ala, Asp, Glu, Val, Ser, Ile, Leu, Pro, Thr, Lys, Phe, Arg, and His; involvement of the following amino acids has not been observed: Asn, Gln, Cys, Tyr, Met, and Trp[12]). With a more permissive definition of a domain, one can include shorter peptides. In a search of the PDB for folds that display any degree of tertiary structure, one finds β-hairpin structures that are stable and that can exhibit a cooperative folding transition. Still, of the 40 structures in the PDB that are derived from isolated, short (less than 45 residues), water-soluble, monomeric, non-disulfide-bonded, non-metal-bound, non-cyclic peptides, all of them have at least one but usually several aromatic residues. Thus, while it is conceivable that short sequences can fold into stable domains without aromatic residues or disulfide bonds, such sequences are not represented in the structural database. Another small protein that has been proposed to be a CA fold is ferredoxin, a 52-residue peptide originally derived from the duplication of a 26-residue peptide that had evolved to bind Fe–S clusters.[13] While the evidence is overwhelming that it is an ancient protein, it still displays a high level of complexity with regard to the eight cysteines that coordinate the Fe–S cluster. Thus, due to both their length and special amino acid requirements, small globular domains are, by our reasoning, poor CA fold candidates.

Without disulfide bonds and aromatic residues, small globular proteins would be expected to be less stable and may have instead existed as molten globules. While the definition of a molten globule is often debated,[14–16] the term is most often used to describe a nonnative state of a protein that exhibits some or all of the following physicochemical properties: a high degree of secondary structure without a defined tertiary structure, rapid backbone amide exchange with water, low chemical shift dispersion of its NMR spectrum, a non-cooperative thermal denaturation transition, and a low stability with a tendency to aggregate. Furthermore, molten globule states often bind with high affinity to hydrophobic dyes (most typically 8-anilino-1-naphthalene sulfonate or ANS) because of the solvent exposure of nonpolar side chains that in the native state would be buried in the interior of the protein.

The molten globule state is attractive as a CA peptide fold because it could exist before a sequence optimized for a stable tertiary fold has evolved.[17] Furthermore, molten globule states have demonstrated potential for enzymatic activities that rival their native counterparts.[18] It has therefore been suggested that the inherently flexible molten globule state would be an ideal scaffold for the diversification and development of new catalytic activities because the evolutionary distance between two proteins with unique functions is likely to be shorter.[18] Still, the main drawback as a CA fold candidate is that, like a stable fold, a molten-globule-like state can only exist for peptide sequences that are long enough to adopt several secondary structural elements unless, as we will discuss below, shorter peptides may oligomerize into a molten globule entity.[19]

Intrinsically Disordered Peptides

IDPs are, by definition, very flexible and without long-lived structural states, in most instances interconverting between various (random-coil-like) conformational species. Thus, their very existence calls into question the classical idea of a structure–function relationship.[20] For some IDPs, their function is dependent on their lack of a defined structure, while for others, the interconversion between a disordered and a structured state is central to their function. Many IDPs and proteins with IDP segments are associated with diseases, a few examples being tau protein, α-synuclein, p53, and prion protein.[21] IDPs often have sites of posttranslational modification that modify protein–protein interaction networks. While their functions usually involve a folding event triggered upon binding to a partner, as with tau–microtubule interactions and α-synuclein–lipid interactions, some seem to act simply as a covalent link between two functional domains (reviewed in Ref. 22).

Due to their simplicity, small IDPs satisfy some of the logically deduced requirements for a first protein fold. In particular, they do not have a defined structure and therefore could exist before having evolved into a single low-energy state. Furthermore, IDPs can often have simple/repetitive amino acid sequence compositions and their functions are less sensitive to sequence variation than proteins with defined tertiary structures.[23,24] While these simple peptides appear to be the ideal candidates for a CA fold, there are a few drawbacks. The known activities of IDPs usually involve a folded protein or oligonucleotide molecule as an interaction partner. For example, an unstructured Arg- or Lys-rich peptide may interact electrostatically with the phosphodiester bond of an oligonucleotide. While this would fit into an RNA world scenario, the possibility for IDPs to support catalytic activities is severely limited by their dynamic nature and lack of persistent structural elements.[25] Lastly, IDPs are more prevalent in eukaryotes than prokaryotes, a hint that they are not as ancient as other functional folds.[26]

Helical Peptides and Their Oligomers

The α-helix is an attractive CA fold candidate merely due to its simplicity: the major stabilizing interaction is the $i+4$ H-bond, and therefore, the addition, mutation, or insertion of a residue in the sequence is likely to be tolerated without disrupting the structure. The side chains of at least three residues can be juxtaposed on the surface of a helix, and thus, in principle, some catalysis is possible. In fact, the hydrolysis of RNA by basic polypeptides has been shown to be dependent on the helix-induced spatial arrangement of Lys or Arg residues on their surface.[25] Also, a rationally designed helical peptide that catalyzes the decarboxylation of oxaloacetate has been reported.[27] This 14-residue peptide, which in its active state oligomerizes to form a four-helix bundle, has a single catalytic lysine that forms the rate-determining imine intermediate 3–4 orders of magnitude faster than simple amines. While the activity of this peptide requires a helix conformation, it appears that the tetrameric complex displays the properties of a molten globule.[17] Perhaps more interesting from a prebiotic perspective are the results coming from two independent research groups that have elaborated autocatalytic and cross-catalytic systems in which helical peptides act as templates in a peptide ligation reaction. Ghadiri *et al.* first demonstrated that a 32-residue helical peptide that forms a coiled-coil structural motif (two or more helical peptides that twist around each other) efficiently catalyzes the

ligation of smaller peptide fragments with high sequence selectivity and stereoselectivity.[28,29] They also showed that these types of peptides can symbiotically replicate each other through a hypercycle, with autocatalytic rate enhancements of several thousands over the background reactions.[30] Using a similar system, Chmielewski's group has developed pH-sensitive peptide ligases[31] and has further improved the efficiency of the autocatalytic peptides with the finding that shorter (26-residue) peptides that form less stable coiled coils do not suffer from product inhibition and have lower background rates.[32] All of the helical peptides mentioned so far are likely to be close to the minimum size for a stable helix. Other studies have shown that *de novo* designed peptides as short as 15 residues can form stable coiled coils and higher-order helical assemblies,[33] while a 12-residue peptide can also form coiled coils but with a much lower stabililty.[19] A caveat to these 12- to 15-residue coiled coils is that they require N-terminal acetylation[19] or an N-terminal succinylation and a C-terminal amidation as well as several stabilizing intra- and interhelical salt bridges.[33] It is generally observed that short helical oligomers designed around the coiled-coil motif, whose folding energy is derived from the burial of a hydrophobic core, often have properties typical for molten globules. Native-like proteins are then obtained only upon optimization of specific polar interactions between side chains on the solvent-exposed surface.[34] Nonetheless, all of the helical peptides that we have discussed highlight the fact that the length requirement of a folded peptide can be decreased if it undergoes oligomerization, thereby allowing a larger hydrophobic area to be buried than is possible as a monomer (this oligomerization effect is even more pronounced with β-strands as discussed in Peptide Amyloids). Thus, the current literature on oligomeric helical peptides demonstrates that they satisfy many of our requirements for a CA fold. However, the self-replicating coiled coils in particular are not likely to have been related to any prebiotic or protobiological systems because they require the use of chemically complex substrates, including peptides with thioesters and peptides with the amino acid Cys.

Membrane-Bound Peptides

Membrane proteins are very distinct from their soluble counterparts on the levels of primary and tertiary structure. At the primary-structure level, the sequences of most membrane proteins that have helical transmembrane segments can be identified simply using an algorithm that searches for 15- to 20-residue-long stretches of mostly hydrophobic residues.[35] Although membrane proteins can be complex, their transmembrane fragments are usually simple, the most common structural motif for these being a bundle of α-helices and less often a β-barrel. Unlike soluble proteins, some membrane proteins retain their functionality even if a large fraction of the protein is removed.[36,37] Small hydrophobic peptides composed of 16–24 residues (long enough to span the membrane) having otherwise diverse but simple sequences can in general form transmembrane or/and membrane-bound helices.[38] The reason for the low sequence stringency of transmembrane segments is that the hydrogen-bonding interactions in an α-helix are all satisfied within the peptide backbone while the side chains interact minimally with each other and primarily with the lipid environment in a non-specific (hydrophobic) manner. Likewise, sequences with periodic repeats of hydrophobic and polar residues that form amphipathic helices can interact tangentially at the aqueous interface of membranes with low sequence specificity.[38,39] The formation of ordered, helical structures in membrane environments is thereby primarily governed by the density of polar and nonpolar amino acid residues in a peptide and not by the specific identities of the side chains.[38] By providing a large hydrophobic area or volume, a membrane can stabilize the ordered secondary structure of a short peptide that is otherwise too small to fold on its own.[35,39,40] This effect is exemplified in the disordered/soluble to ordered/membrane-bound transition of many short antimicrobial and cell-penetrating peptides. These peptides range in size from 15 to 35 amino acid residues and utilize various mechanisms of membrane perturbation or translocation (reviewed in Refs. 41 and 42).

While lipid–protein interactions can be very sequence unspecific, peptides can oligomerize into coiled coils or pore-like structures when inserted into a membrane in a much more sequence-specific manner. Such structures formed from short peptides can support functions typical of membrane proteins including ion transport.[37] The membrane/water interface can also induce the assembly of amphipathic β-sheets from small peptides that have alternating polar and nonpolar residues.[39] Still, compared to helical membrane peptides, β-sheet membrane-integrating structures are usually more complex, as typified by the long-range intramolecular contacts found in β-barrels. A notable exception is gramicidin, a 15-residue peptide composed of strictly alternating L- and D-amino acids (not a likely CA sequence composition) that can oligomerize into a membrane spanning β-helix.

In summary, if membrane-enclosed vesicles were present coincidently with the conditions that led to the first peptide replication, membrane-bound α-helical or β-sheet peptides would be attractive candidates for the first protein fold. The membranes and peptides could act synergistically to stabilize

each other with the membrane also providing a matrix that would prevent the dilution of functional peptides as they were created. The length requirements to span the membrane of at least 15 amino acid residues mentioned above could be reduced if the ancient membranes were thinner, being composed of lipid or lipid-like molecules with shorter hydrophobic chains.

Peptide Amyloids

Unlike α-helices with their sequential ($i+4$) hydrogen-bonding pattern, β-strands do not exist in isolation and are almost always part of a β-sheet in which the backbone hydrogen bonds are satisfied between adjacent strands in the sheet. Therefore, a short peptide that has no stable secondary structure may oligomerize into a β-sheet, and these sheets can further assemble into stable layered structures. Such a supramolecular assembly is the core of the amyloid structure. Amyloid fibrils have long been associated with dozens of diseases including Alzheimer's and prion diseases,[41] but there are also many amyloids, termed "functional amyloids", that have normal biological activities.[42] The biophysical description of an amyloid is that of an indefinitely repeating cross-β-sheet entity as depicted in Fig. 1. Amyloids are composed of an ordered arrangement of many (usually thousands of) copies of a peptide or protein such that the hydrogen bonding of the backbone atoms are satisfied through intermolecular and sometimes intramolecular β-strand contacts. The two-state (soluble/insoluble) nature of the amyloid fold and the structural repetitiveness of its aggregates add a level of complexity that most soluble proteins cannot access. The short repeat length of 4.7–9.4 Å between equivalent β-strands creates a close spacing of identical side chains. This high local concentration of functional groups can increase binding affinity and generate specificity where none would exist without a repeating structure. For example, the spatial arrangement of side chains induced by the β-sheet structuring of a basic peptide has been shown to lead to a 15-fold rate enhancement of RNA hydrolysis activity over unstructured basic peptides.[25] Also, a single amphipathic β-strand may have negligible affinity for a membrane, but an amyloid composed of this peptide could associate tightly to lipids.[39] Similarly, an amyloid formed from a basic peptide will have a higher affinity for the negatively charged backbone of a DNA molecule than a single peptide.[43] The wide species distribution and growing number of known functional amyloids is further evidence that despite its simplicity, the amyloid fold can give rise to a plethora of activities.[42] Finally, a feature that sets the amyloid structure apart with regard to a CA fold is that it can be formed by peptides as short as four amino acid residues, thus providing small peptides access to some features of larger folded proteins. In fact the dipeptide Phe-Phe forms stable nanotubes that have amyloid properties (1630 cm^{-1} vibrational band and Congo red birefringence) and so a minimal peptide can probably also adopt the amyloid structure.[44]

The Structures and Polymorphisms of Peptide Amyloids

Several amyloid crystal structures of short (four to seven residues) peptides have been published in the last few years, revealing the basic motif of the amyloid at atomic resolution.[45–49] These studies have shown that the cross-β fold can be as simple as a single four-residue β-strand that assembles in a repetitive manner and that, in the interfaces between the individual sheets of the amyloid fiber, the side chains can be either highly interdigitated in "dry interfaces" or less well packed in hydrated "wet interfaces" (Fig. 1). The tetrapeptide of Fig. 1a, the shortest peptide for which an amyloid-like crystal structure has been determined, has only "dry" interfaces; however, this peptide is composed of Asn and Gln, amino acids that are not considered prebiotic.[12] The hexapeptide of Fig. 1b, composed of prebiotically relevant amino acids, demonstrates both the tightly packed dry and the water-mediated interfaces. In contrast to these minimalist amyloid peptides structures, the fold of larger amyloid proteins such as the HET-s prion, which is a β-solenoid, bears remarkable resemblance to some soluble proteins.[50] From the range of complexity represented by these structures, it appears that the relatively simple scaffold of the amyloid could be expanded upon to achieve the complexity required to carry out enzymatic catalysis.

In the crystallographic studies mentioned above, some peptides were observed in different crystal forms, representing distinct amyloid structures. This apparent breakdown of the one fold per sequence paradigm (as stated in Anfinsen's thermodynamic hypothesis) is a common feature of the amyloid fold[51,52] and is referred to as polymorphism. Prions, the infectious proteins whose replicative ability is based on the amyloid structure, are also polymorphic, and different polymorphs are correlated with unique strains of the disease.[53] Also, some polymorphs of amyloids have been shown to be toxic or infectious while others are not.[54,55] The repetitive nature of the amyloid allows it not only to act as a template or seed for the growth of new amyloids but also to do so in a conformation-specific manner such that unique polymorphs can be propagated *in vivo* or *in vitro*.[51,53] Whereas the interconversion between conformations or side-chain rotamers is usually fast in soluble proteins (in thermal equilibrium), the

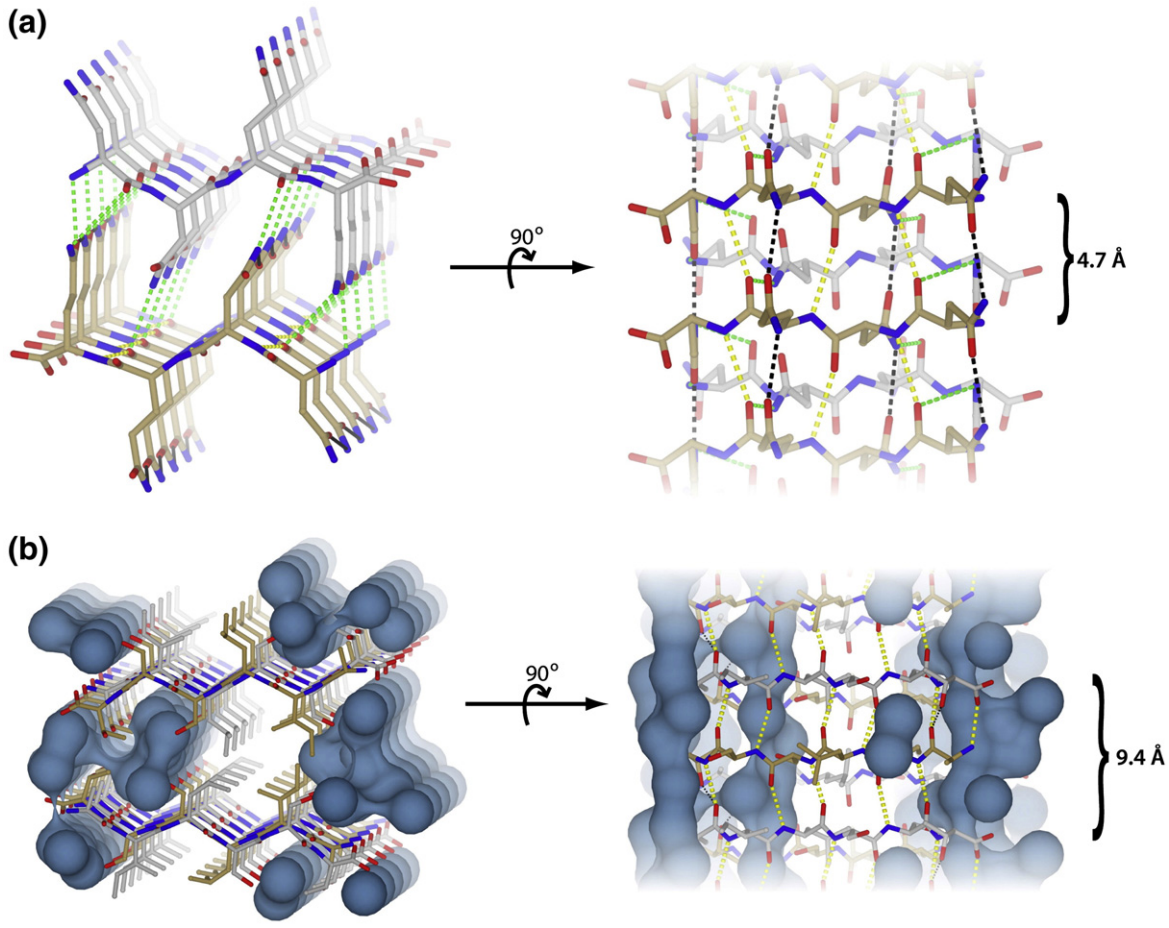

Fig. 1. The structures of small peptide amyloids. (a) Orthogonal views of the four-residue peptide NNQQ as it is found in amyloid microcrystals (PDB entry 2OLX). On the left is the view looking down the axis of the protofibril and on the right is the view 90° rotated around a horizontal line. All of the peptides in the fibril have an identical structure, and the coloring of the carbon atoms simply indicates their orientation (silver, N→C; bronze, C→N). The hydrogen bonds that involve atoms from the lower β-sheet (left view) are indicated with broken lines and color coded as follows: green, intersheet; yellow, backbone interstrand; black, side-chain interstrand. (b) Two views of the six-residue peptide AILSST (PDB entry 3FOD) represented as in (a) and with the water molecules displayed as blue surfaces. These two structures highlight the different types of interactions that can be found in amyloids. The NNQQ structure is composed of two sheets of parallel β-strands that are very densely packed with many hydrogen bonds between the two sheets and no water within the fibril core. The AILSST structure is composed of antiparallel β-sheets that pack together with both water-mediated polar interactions and a dry hydrophobic core. Unlike NNQQ, it lacks direct hydrogen bonds between the two sheets [green in (a)].

amyloid fold can trap a conformation of a peptide (or even a single side chain) such that the addition of new monomers to the growing fibril adopt the kinetically favored conformation rather than that which would be determined by a Boltzmann distribution. This situation of non-interconverting states of similar energy can be represented by the group lap sit (Fig. 2). In this scenario, individuals in a linear (or circular) arrangement are either all standing or all sitting, and only a concerted action by all individuals (a kinetically inaccessible motion for inanimate objects) can lead to an interconversion of states (polymorphs). Still, amyloids can be dynamic, if only through a monomer/aggregate equilibrium. In an interesting experiment, amyloid aggregation was shown to modulate the kinetic and thermodynamic landscape of a peptide bond formation reaction. Fluorenylmethyloxycarbonyl-protected amino acids were mixed with dipeptides in the presence of a protease, which catalyzes the formation and hydrolysis of peptide bonds. In these mixtures, several amyloid-like structures quickly formed, but a slower phase followed, during which the peptides continued to evolve in length from the kinetically favored to the thermodynamically favored species (in this case, from three to five residues), with a concomitant structural rearrangement from fibers to sheets.[57] Thus, in addition to

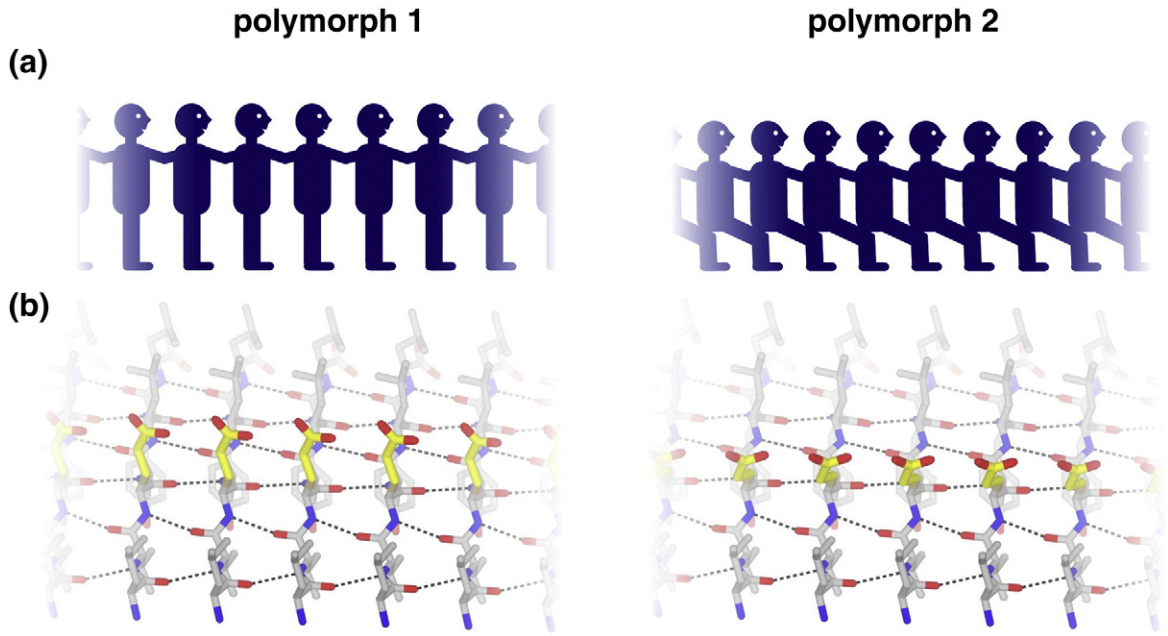

Fig. 2. A kinetic basis for amyloid polymorphism. (a) A schematic representation of the "group lap sit". The transition from polymorph 1 to 2 in this group exercise requires the coordinated efforts of each individual. (b) A molecular representation of a subtle polymorphism in a peptide amyloid. In each polymorph, the Glu side chain (yellow) exists in only one of two favored conformations. As in (a), there is an energy barrier (e.g., steric clashes) that prevents an individual from changing conformation on its own. Since the peptides cannot coordinate the interconversion between polymorphs, the two energetically similar states can remain spatiotemporally distinct. Polymorph 1 in (b) is directly from the coordinates for PDB entry 3HYD and polymorph 2 is a hypothetical model based on these coordinates. All molecular representations were prepared with CCP4mg.[56]

having unique binding properties and the potential to support catalytic activities, amyloids can also store and propagate information in the form of a conformation.

Peptide Membranes

Simple amphiphilic peptides have been observed to self-assemble into a broad range of structures, many of which resemble the types of structures commonly associated with lipids (reviewed in Ref. 58). Most of these assemblies have an underlying repetitive β-sheet structure and are therefore closely related to, if not correctly classified as, amyloids. An early example of an amyloid-like β-sheet bilayer type of structure was reported for poly(Val-Lys).[59] Analogous to a lipid bilayer, poly(Val-Lys) forms amphiphilic sheets that, in aqueous solution, form a two-layered sheet with the hydrophobic Val side chains facing inward and the Lys side chains facing the solvent. Other peptides have been shown to form closed structures including hollow tubes[44] and vesicles.[60] While these membranes can have dimensions similar to biological phospholipids, their peptide backbone and side chains bring much more functionality to the structures than do the long alkane chains of lipids. The forces that drive the assembly of peptides into membranes and the structures themselves are also very different from the nonspecific hydrophobic driven assembly of lipid molecules and, based on the structural models, resemble more the specific nature of protein folding.[58] The ability of simple peptides to form enclosing structures is of prebiotic significance for the simple reason that there is no evidence that lipids were the first biological barriers. They are particularly interesting because, like other amyloids, peptide membranes can form from very short sequences and, due to their nearly unlimited chemical diversity, will have the potential activities of other amyloids and larger proteins.

Proteins as Genes: An Amyloid Replicator as the First Functional Protein Fold

To this point in our discussion, amyloids appear to satisfy many of the prerequisites for a CA fold: they can form from very short sequences, are formed from peptides composed of prebiotic amino acids,[12] and can support a variety of functions.[43,61] Interestingly, several studies have shown that there has been an evolutionary selection against β-aggregation-prone

amino acid sequences,[62,63] yet it has also been found that organism complexity is inversely correlated with proteomic β-aggregation propensity.[64] The simplest interpretation of these observations is that the amyloid is an ancient fold whose utility decreased, eventually becoming more detrimental than useful, as life evolved with more complex protein folds. Furthermore, amyloids possess some unique properties that make them particularly interesting as a candidate CA fold and have led to several hypotheses in which they have a leading role in the origin of life.[7,8,65,66] First, the amyloid fold is very stable and more resistant to hydrolysis than its component peptides.[67,68] Under the harsh conditions of intense UV radiation, high temperature, and pH extremes that were likely encountered on the early earth, amyloids would therefore have an inherent fitness. Second, amyloids by their very nature maintain a peptide at a locally high concentration. This could also be a fitness advantage in a scenario in which proteins/peptides evolved outside of lipid vesicles where diffusion would be detrimental to a selection process. In the selection of a fitness advantage, the actual "fitness" would have to be connected to the information that was used to make the peptide (in modern biology, the gene), and this leads to the final argument for amyloids: Amyloids self-propagate and by doing so can select their own kind from a pool of unrelated peptides. Taking this inherent activity one step further, the amyloid fold can act as a template for the synthesis of more of its own constituent peptides, a function that has been demonstrated in an experiment that is analogous to the replication of helical peptides mentioned above.[69] However, the amyloid-directed synthesis of peptides starting from single amino acids has yet to be demonstrated. Nonetheless, amyloids satisfy all of our stated requirements for a CA fold and additionally have the potential to store information,[70,71] self-replicate,[69] and support biologically relevant activities.[61]

Amino Acid Sequence Length and the CA Fold: A Probability Argument

In the above discussions, it was highlighted that small globular folds and helical peptides with enzymatic activity are composed of at least 26 residues, that 12-residue helical peptides are able to form helical oligomers that may be of a molten-globule-like nature, and that helical membrane-bound peptides are composed of at least 15 residues, yet that amyloids can be composed of peptides as short as 4 residues (and likely as short as 2–3 residues).[44,46,57] To highlight how relevant the peptide length is for the CA fold, two probability arguments are put forward here. Considering a conservative situation in which there are only five prebiotic amino acid types available for peptide synthesis, there exist 2.4×10^8 unique sequences for a peptide of length 12. Without a high fidelity replication, the probability of a fitness-altering quantity of a 12-mer peptide sequence being produced is vanishingly small. Amyloids can defy this probability problem as they can be formed from even shorter peptides (at least as short 4 amino acid residues, 625 possible sequences). A second argument is based on experimental estimates that, in a random library of 80-residue-long peptides consisting of all 20 amino acids, about 1 in 10^{11} sequences will have a binding affinity to a specific target in the low micromolar range.[72] This unfavorable statistic for soluble proteins is not likely to apply to amyloids since their structure can amplify a low affinity several-fold through repeat-induced cooperative interactions.[61,66]

Conclusion

In this synopsis, we have discussed some of the possible peptide folds from which the protein fold universe may have arisen. With our arguments coming out in favor of an amyloid CA fold, there is no obvious reason to definitively exclude the possibility that any of the folds discussed above is the CA fold. Furthermore, several folds may have evolved in parallel, perhaps at first independently and then symbiotically, to either create the first organism or provide a selective advantage for an existing organism. Alternatively, the CA peptide may also have adopted several classes of folds in parallel. Lessons from nature teach us that amyloidogenic sequences are often found in IDPs (e.g., α-synuclein and tau), that membrane-integrating peptides in the absence of lipids can be disordered or aggregate into amyloids (e.g., Aβ associated with Alzheimer's disease), and that molten-globule-like states of proteins are more prone towards amyloid aggregation than their native states.[73] Thus, the CA fold may have been multifaceted, gaining selective advantages both from amyloid properties and from activities in another structural state. This structural promiscuity highlights once more the uniqueness of the amyloid fold within the protein fold universe and its potential as a CA to modern protein folds.

References

1. Orgel, L. (1973). *The Origins of Life.* Molecules and Natural Selection. John Wiley & Sons, Inc., New York, NY.
2. Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M. & Cate, J. H. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. *Science*, **310**, 827–834.
3. Pereto, J. (2005). Controversies on the origin of life. *Int. Microbiol.* **8**, 23–31.

4. Altermann, W. & Kazmierczak, J. (2003). Archean microfossils: a reappraisal of early life on Earth. *Res. Microbiol.* **154**, 611–617.
5. Rode, B. M. (1999). Peptides and the origin of life. *Peptides*, **20**, 773–786.
6. Ikehara, K. (2005). Possible steps to the emergence of life: the [GADV]-protein world hypothesis. *Chem. Rec.* **5**, 107–118.
7. Carny, O. & Gazit, E. (2005). A model for the role of short self-assembled peptides in the very early stages of the origin of life. *FASEB J.* **19**, 1051–1055.
8. Maury, C. P. (2009). Self-propagating beta-sheet polypeptide structures as prebiotic informational molecular entities: the amyloid world. *Orig. Life Evol. Biosph.* **39**, 141–150.
9. Gilbert, W. (1986). Origin of life: the RNA world. *Nature*, **319**, 618.
10. Cech, T. R. (2011). The RNA worlds in context. *Cold Spring Harbor Perspect. Biol.* 1–5.
11. Jones, S., Stewart, M., Michie, A., Swindells, M. B., Orengo, C. & Thornton, J. M. (1998). Domain assignment for protein structures using a consensus approach: characterization and analysis. *Protein Sci.* **7**, 233–242.
12. Higgs, P. G. & Pudritz, R. E. (2009). A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. *Astrobiology*, **9**, 483–490.
13. Eck, R. V. & Dayhoff, M. O. (1966). Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. *Science*, **152**, 363–366.
14. Regan, L. (2003). Molten globules move into action. *Proc. Natl Acad. Sci. USA*, **100**, 3553–3554.
15. Ohgushi, M. & Wada, A. (1983). 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Lett.* **164**, 21–24.
16. Ptitsyn, O. B. (1991). How does protein synthesis give rise to the 3D-structure? *FEBS Lett.* **285**, 176–181.
17. DeGrado, W. F. (1993). Peptide engineering. Catalytic molten globules. *Nature*, **365**, 488–489.
18. Vamvaca, K., Vogeli, B., Kast, P., Pervushin, K. & Hilvert, D. (2004). An enzymatic molten globule: efficient coupling of folding and catalysis. *Proc. Natl Acad. Sci. USA*, **101**, 12860–12864.
19. Eisenberg, D., Wilcox, W., Eshita, S. M., Pryciak, P. M., Ho, S. P. & DeGrado, W. F. (1986). The design, synthesis, and crystallization of an alpha-helical peptide. *Proteins*, **1**, 16–22.
20. Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. *J. Mol. Biol.* **293**, 321–331.
21. Uversky, V. N., Oldfield, C. J. & Dunker, A. K. (2008). Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* **37**, 215–246.
22. Tompa, P. (2011). Unstructural biology coming of age. *Curr. Opin. Struct. Biol.* **21**, 419–425.
23. Uversky, V. N., Gillespie, J. R. & Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*, **41**, 415–427.
24. Tompa, P. (2003). Intrinsically unstructured proteins evolve by repeat expansion. *BioEssays*, **25**, 847–855.
25. Barbier, B. & Brack, A. (1992). Conformation-controlled hydrolysis of polyribonucleotides by sequential basic polypeptides. *J. Am. Chem. Soc.* **114**, 3511–3515.
26. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **337**, 635–645.
27. Johnsson, K., Allemann, R. K., Widmer, H. & Benner, S. A. (1993). Synthesis, structure and activity of artificial, rationally designed catalytic polypeptides. *Nature*, **365**, 530–532.
28. Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K. & Ghadiri, M. R. (1996). A self-replicating peptide. *Nature*, **382**, 525–528.
29. Saghatelian, A., Yokobayashi, Y., Soltani, K. & Ghadiri, M. R. (2001). A chiroselective peptide replicator. *Nature*, **409**, 797–801.
30. Lee, D. H., Severin, K., Yokobayashi, Y. & Ghadiri, M. R. (1997). Emergence of symbiosis in peptide self-replication through a hypercyclic network. *Nature*, **390**, 591–594.
31. Yao, S. & Chmielewski, J. (1999). A pH-tunable peptide ligase. *Biopolymers*, **51**, 370–375.
32. Issac, R. & Chmielewski, J. (2002). Approaching exponential growth with a self-replicating peptide. *J. Am. Chem. Soc.* **124**, 6808–6809.
33. Burkhard, P., Ivaninskii, S. & Lustig, A. (2002). Improving coiled-coil stability by optimizing ionic interactions. *J. Mol. Biol.* **318**, 901–910.
34. Hill, R. B., Raleigh, D. P., Lombardi, A. & DeGrado, W. F. (2000). De novo design of helical bundles as models for understanding protein folding and function. *Acc. Chem. Res.* **33**, 745–754.
35. Popot, J. L. & Engelman, D. M. (2000). Helical membrane protein folding, stability, and evolution. *Annu. Rev. Biochem.* **69**, 881–922.
36. Oblatt-Montal, M., Buhler, L. K., Iwamoto, T., Tomich, J. M. & Montal, M. (1993). Synthetic peptides and four-helix bundle proteins as model systems for the pore-forming structure of channel proteins. I. Transmembrane segment M2 of the nicotinic cholinergic receptor channel is a key pore-lining structure. *J. Biol. Chem.* **268**, 14601–14607.
37. Duff, K. C. & Ashley, R. H. (1992). The transmembrane domain of influenza A M2 protein forms amantadine-sensitive proton channels in planar lipid bilayers. *Virology*, **190**, 485–489.
38. Bechinger, B. (2000). Understanding peptide interactions with the lipid bilayer: a guide to membrane protein engineering. *Curr. Opin. Chem. Biol.* **4**, 639–644.
39. DeGrado, W. F. & Lear, J. D. (1985). Induction of peptide conformation at apolar/water interfaces. 1. A study with model peptides of defined hydrophobic periodicity. *J. Am. Chem. Soc.* **107**, 7684–7689.
40. Bayley, H. (1999). Designed membrane channels and pores. *Curr. Opin. Biotechnol.* **10**, 94–103.
41. Chiti, F. & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366.
42. Fowler, D. M., Koulov, A. V., Balch, W. E. & Kelly, J. W. (2007). Functional amyloid—from bacteria to humans. *Trends Biochem. Sci.* **32**, 217–224.
43. Braun, S., Humphreys, C., Fraser, E., Brancale, A., Bochtler, M. & Dale, T. C. (2011). Amyloid-associated nucleic acid hybridisation. *PLoS One*, **6**, e19125.

44. Reches, M. & Gazit, E. (2003). Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science*, **300**, 625–627.
45. Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. O., Riekel, C., Grothe, R. & Eisenberg, D. (2005). Structure of the cross-beta spine of amyloid-like fibrils. *Nature*, **435**, 773–778.
46. Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I. *et al.* (2007). Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature*, **447**, 453–457.
47. Ivanova, M. I., Sievers, S. A., Sawaya, M. R., Wall, J. S. & Eisenberg, D. (2009). Molecular basis for insulin fibril assembly. *Proc. Natl Acad. Sci. USA*, **106**, 18990–18995.
48. Wiltzius, J. J., Landau, M., Nelson, R., Sawaya, M. R., Apostol, M. I., Goldschmidt, L. *et al.* (2009). Molecular mechanisms for protein-encoded inheritance. *Nat. Struct. Mol. Biol.* **16**, 973–978.
49. Wiltzius, J. J., Sievers, S. A., Sawaya, M. R., Cascio, D., Popov, D., Riekel, C. & Eisenberg, D. (2008). Atomic structure of the cross-beta spine of islet amyloid polypeptide (amylin). *Protein Sci.* **17**, 1467–1474.
50. Wasmer, C., Lange, A., Van Melckebeke, H., Siemer, A. B., Riek, R. & Meier, B. H. (2008). Amyloid fibrils of the HET-s(218–289) prion form a beta solenoid with a triangular hydrophobic core. *Science*, **319**, 1523–1526.
51. Petkova, A. T., Leapman, R. D., Guo, Z., Yau, W. M., Mattson, M. P. & Tycko, R. (2005). Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils. *Science*, **307**, 262–265.
52. Fandrich, M., Meinhardt, J. & Grigorieff, N. (2009). Structural polymorphism of Alzheimer Abeta and other amyloid fibrils. *Prion*, **3**, 89–93.
53. Toyama, B. H., Kelly, M. J., Gross, J. D. & Weissman, J. S. (2007). The structural basis of yeast prion strain variants. *Nature*, **449**, 233–237.
54. Campioni, S., Mannini, B., Zampagni, M., Pensalfini, A., Parrini, C., Evangelisti, E. *et al.* (2010). A causative link between the structure of aberrant protein oligomers and their toxicity. *Nat. Chem. Biol.* **6**, 140–147.
55. Sabate, R., Baxa, U., Benkemoun, L., Sanchez de Groot, N., Coulary-Salin, B., Maddelein, M. L. *et al.* (2007). Prion and non-prion amyloids of the HET-s prion forming domain. *J. Mol. Biol.* **370**, 768–783.
56. Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K. & Noble, M. (2002). The CCP4 molecular-graphics project. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**, 1955–1957.
57. Williams, R. J., Smith, A. M., Collins, R., Hodson, N., Das, A. K. & Ulijn, R. V. (2009). Enzyme-assisted self-assembly under thermodynamic control. *Nat. Nanotechnol.* **4**, 19–24.
58. Childers, W. S., Ni, R., Mehta, A. K. & Lynn, D. G. (2009). Peptide membranes in chemical evolution. *Curr. Opin. Chem. Biol.* **13**, 652–659.
59. Brack, A. & Orgel, L. (1975). Beta structures of alternating polypeptides and their possible prebiotic significance. *Nature*, **256**, 383–387.
60. Vauthey, S., Santoso, S., Gong, H. Y., Watson, N. & Zhang, S. G. (2002). Molecular self-assembly of surfactant-like peptides to form nanotubes and nanovesicles. *Proc. Natl Acad. Sci. USA*, **99**, 5355–5360.
61. Wang, L., Schubert, D., Sawaya, M. R., Eisenberg, D. & Riek, R. (2010). Multidimensional structure–activity relationship of a protein in its aggregated states. *Angew. Chem., Int. Ed. Engl.* **49**, 3904–3908.
62. Broome, B. M. & Hecht, M. H. (2000). Nature disfavors sequences of alternating polar and non-polar amino acids: implications for amyloidogenesis. *J. Mol. Biol.* **296**, 961–968.
63. Monsellier, E. & Chiti, F. (2007). Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep.* **8**, 737–742.
64. Tartaglia, G. G., Pellarin, R., Cavalli, A. & Caflisch, A. (2005). Organism complexity anti-correlates with proteomic beta-aggregation propensity. *Protein Sci.* **14**, 2735–2740.
65. Dale, T. (2006). Protein and nucleic acid together: a mechanism for the emergence of biological selection. *J. Theor. Biol.* **240**, 337–342.
66. Greenwald, J. & Riek, R. (2010). Biology of amyloid: structure, function, and regulation. *Structure*, **18**, 1244–1260.
67. Brack, A. & Spach, G. (1980). Beta-structures of polypeptides with L- and D-residues. Part III. Experimental evidences for enrichment in enantiomer. *J. Mol. Evol.* **15**, 231–238.
68. Safar, J., Roller, P. P., Gajdusek, D. C. & Gibbs, C. J., Jr (1993). Thermal stability and conformational transitions of scrapie amyloid (prion) protein correlate with infectivity. *Protein Sci.* **2**, 2206–2216.
69. Takahashi, Y. & Mihara, H. (2004). Construction of a chemically and conformationally self-replicating system of amyloid-like fibrils. *Bioorg. Med. Chem.* **12**, 693–699.
70. Coustou, V., Deleu, C., Saupe, S. & Begueret, J. (1997). The protein product of the het-s heterokaryon incompatibility gene of the fungus *Podospora anserina* behaves as a prion analog. *Proc. Natl Acad. Sci. USA*, **94**, 9773–9778.
71. Weissmann, C. (2004). The state of the prion. *Nat. Rev., Microbiol.* **2**, 861–871.
72. Keefe, A. D. & Szostak, J. W. (2001). Functional proteins from a random-sequence library. *Nature*, **410**, 715–718.
73. Sunde, M. & Blake, C. C. (1998). From the globular to the fibrous state: protein structure and structural conversion in amyloid formation. *Q. Rev. Biophys.* **31**, 1–39.