Structure 14, 1667–1676, November 2006 ©2006 Elsevier Ltd All rights reserved DOI 10.1016/j.str.2006.09.008

Proline and Glycine Control Protein Self-Organization into Elastomeric or Amyloid Fibrils

Sarah Rauscher,[1,2] Stéphanie Baud,[1] Ming Miao,[1] Fred W. Keeley,[1,2] and Régis Pomès[1,2,*]
[1] Molecular Structure and Function Programme
Hospital for Sick Children
555 University Avenue
Toronto, Ontario M5G 1X8
Canada
[2] Department of Biochemistry
University of Toronto
1 King's College Circle
Toronto, Ontario M5S 1A8
Canada

Summary

Elastin provides extensible tissues, including arteries and skin, with the propensity for elastic recoil, whereas amyloid fibrils are associated with tissue-degenerative diseases, such as Alzheimer's. Although both elastin-like and amyloid-like materials result from the self-organization of proteins into fibrils, the molecular basis of their differing physical properties is poorly understood. Using molecular simulations of monomeric and aggregated states, we demonstrate that elastin-like and amyloid-like peptides are separable on the basis of backbone hydration and peptide-peptide hydrogen bonding. The analysis of diverse sequences, including those of elastin, amyloids, spider silks, wheat gluten, and insect resilin, reveals a threshold in proline and glycine composition above which amyloid formation is impeded and elastomeric properties become apparent. The predictive capacity of this threshold is confirmed by the self-assembly of recombinant peptides into either amyloid or elastin-like fibrils. Our findings support a unified model of protein aggregation in which hydration and conformational disorder are fundamental requirements for elastomeric function.

Introduction

Elastomeric proteins provide the high efficiency elastic recoil necessary to undergo reversible deformation to biological machinery as diverse as the mammalian arterial wall (Vrhovski and Weiss, 1998), the capture spiral of spider webs (Becker et al., 2003), and the hinge of scallop shells (Bochicchio et al., 2005). Of particular interest is elastin, which, together with other structural proteins, forms the fabric of extensible tissues, including skin, blood vessels, and elastic ligaments, and provides the elasticity required for proper physiological function (Vrhovski and Weiss, 1998). Elastin and elastin-derived peptides self-aggregate upon heating to form an organized fibrillar structure in a process known as coacervation. Remarkable durability and intrinsic capacity for self-organization make elastin an ideal biomimetic model in the development of synthetic biomaterials (Bellingham et al., 2003). The sequence of tropoelastin, the monomeric precursor of elastin, is composed of alternating "crosslinking" and "hydrophobic" domains. The covalent crosslinking of elastin monomers imparts strength and stability to the polymeric matrix, while the hydrophobic domains are thought to confer the propensities for self-aggregation and extensibility (Vrhovski and Weiss, 1998).

At present, there is little definitive information on the molecular basis for elastin's properties. The insolubility, conformational heterogeneity, and intrinsic flexibility of elastin have precluded the use of conventional structural determination methods, including X-ray crystallography and solution NMR (Pometun et al., 2004). Molecular dynamics simulations are not hindered by conformational disorder and have therefore been useful in developing an atomic-level description of elastin's structure, although the scope of these studies has been limited to short time scales (Li et al., 2001) or small oligopeptides (Baer et al., 2006; Floquet et al., 2004). Solid-state NMR has provided significant insight, suggesting the absence of α helix and β sheet, and a high degree of dynamic disorder (Pometun et al., 2004). These observations are inconsistent with models requiring conformationally restricted structures, such as the β spiral (Venkatachalam and Urry, 1981). However, models of elastin as a random coil exhibiting rubber-like elasticity (Hoeve and Flory, 1974) are at odds with evidence for significant amounts of polyproline II structure and β turns emerging from circular dichroism and Raman spectroscopic studies (Bochicchio et al., 2004; Bochicchio and Tamburro, 2002; Tamburro et al., 2003).

In contrast to elastin fibers, which are essential in many extensible tissues, amyloid fibrillar deposits are associated with numerous tissue-degenerative pathologies, including Alzheimer's and Parkinson's diseases (Dobson, 2003; Stefani and Dobson, 2003). Using electron microscopy and solid-state NMR, the molecular structure underlying amyloid fibrils has recently been shown to consist of cross-β sheets extending along the main axis of the fibril (Nelson et al., 2005; Nelson and Eisenberg, 2006; Petkova et al., 2002). While most studies of amyloidogenic proteins have understandably focused on specific sequences implicated in human disease, there is evidence for the wider relevance of self-assembly of polypeptide chains into amyloid fibers (Dobson, 2003; Stefani and Dobson, 2003). An increasing number of proteins with no known associated pathology have been shown to populate the amyloid state under destabilizing conditions or with the introduction of destabilizing mutations (Stefani and Dobson, 2003). These include the natively α-helical protein myoglobin (Fändrich et al., 2001) and exon 30 of human elastin, which forms amyloid when removed from the context of the full tropoelastin sequence (Tamburro et al., 2005). Recently, functional amyloids have been identified: Pmel17, which is an intracellular amyloid acting as a template for melanin synthesis in mammalian melanosomes (Fowler et al., 2006), and silk moth chorion

*Correspondence: pomes@sickkids.ca

A

B

HU24:	GLVPGVGVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAPAIGP
HU24[P/G]:	GLVGGVGVAGGVGVAGGVGVAGGVGLAGGVGVAGGVGVAGGVGVAPAIGP
CRS:	GLVPGVGVPGVGVPGVGVPGVGLPGVGVPGVGVPGVGVPAIGP
CRS[P/G]:	GLVGGVGVGGVGVGGVGVGGVGLGGVGVGGVGVGGVGVPAIGP
HU20:	FPGFGVGVGGIPGVAGVPGVGGVPGVGGVPGVGISP
CH20:	GAGVPGVGVPGVGIGGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV PGVVPGVGV
GVA:	GLVGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAPANWPLGIHRD

C

Peptide	X1	X2	Assembly Characteristics
EP20-24-24	HU24	HU24	elastin-like
EP20-CRS-CRS	CRS	CRS	elastin-like
EP20-24-24[P/G]	HU24	HU24[P/G]	amyloid
EP20-24[P/G]-24[P/G]	HU24[P/G]	HU24[P/G]	amyloid
EP20-CRS-CRS[P/G]	CRS	CRS[P/G]	elastin-like/amyloid
EP20-CRS[P/G]-CRS[P/G]	CRS[P/G]	CRS[P/G]	amyloid

D

	PGV	GGV	GV	GVA
PGV	$(PGVPGV)_6$	$(GGVPGV)_6$	$(GVPGV)_7$	$(GVAPGV)_6$
GGV		$(GGVGGV)_6$	$(GGVGV)_7$	$(GVAGGV)_6$
GV			$(GVGV)_9$	$(GVGVA)_7$
GVA				$(GVAGVA)_6$

Figure 1. Sequence of Model Peptides

(A) The recombinant peptides EP20-X1-X2, with hydrophobic domains in blue and crosslinking domains in orange. The first hydrophobic domain is human exon 20 (HU20), while X1 and X2 are variable.

(B) A variety of sequences that may be substituted for X1 and X2, including human exon 24 (HU24), either native or with mutations of prolines to glycines; chicken repeat sequence (CRS) or chicken exon 24, either native or with PG mutations; human exon 20 (HU20); chicken exon 20 (CH20); and a repeat of GVA. Elastin's hydrophobic domains contain extended tracts composed mostly of four residues (P, G, V, and A), which together comprise nearly 75% of the sequence of tropoelastin. To highlight the repetitive nature of these sequences, repeats of PGV, GGV, GV, and GVA are indicated in red, blue, yellow, and green, respectively.

(C) Summary of experimental characterization of EP20-X1-X2 recombinant peptides (Miao et al., 2003).

(D) The set of model peptides used in our computational study, based on the sequences in (B) and constructed with repeats of pairwise combinations of PGV, GGV, GV, and GVA. The length of these model periodic sequences (35 or 36 residues) is comparable to that of hydrophobic domains of elastin.

protein, which acts as a protective barrier for silk moth eggs (Iconomidou et al., 2000). It has been proposed that the amyloid fibril represents an inherent form of organization potentially accessible to all polypeptide chains under appropriate conditions (Dobson, 2003; Stefani and Dobson, 2003). Therefore, it is important to understand how the hydrophobic domains of tropoelastin, and indeed all self-associating elastomeric proteins, manage to avoid this fate.

In order to uncover the fundamental balance of forces underlying the assembly of polypeptides into amyloid-like or elastin-like protein aggregates, we designed a set of model sequences with physico-chemical properties compatible with those of the hydrophobic domains of elastin, as well as amyloid fibrils. Our design strategy is based on studies of recombinant elastin-like peptides composed of alternating hydrophobic and crosslinking domains (Figure 1A). The sequence of the hydrophobic domains of elastin is simple and repetitive (Figure 1B). Elastin-like peptides containing human exon 24 (which contains the repeat fragment PGVGVA), or chicken exon 24 (with repeat fragment PGVGV), have been shown to form biomaterials with self-aggregation and mechanical properties similar to native polymeric elastin (Keeley et al., 2002; Miao et al., 2003). Mutations of the tandem PGVGVA repeats to GGVGVA, intended to increase the conformational flexibility of the polypeptide chains, promoted the formation of amyloid-like fibrils (Miao et al., 2003). Interestingly, mutations of the PGVGV repeats to GGVGV resulted in a polypeptide which could either coacervate like elastin peptides or form amyloid, depending on solution conditions (Figures 1B and 1C) (Miao et al., 2003). Poly-(GGVGV) displayed a similar ambiguity in its aggregation—forming an elastin-like film or amyloid fibrils when deposited from methanol or aqueous solution, respectively (Flamia et al., 2004). Based on these observations, we considered four repeat units, PGVGVA, GGVGVA, PGVGV, and GGVGV, which encode for materials spanning elastin-like and amyloid-like aggregates. Recognizing that these repeats are formed by the pairwise combination of four distinct fragments: PGV, GGV, GV, and GVA, we completed our model set with the six remaining combinations: PGVPGV, GGVPGV, GGVGGV, GVGV, GVAGVA, and GVGVA. The first two of these are the repeat units found in chicken exon 20 and human exon 20 (Figure 1B). All ten pairwise combinations may be represented as a simple matrix

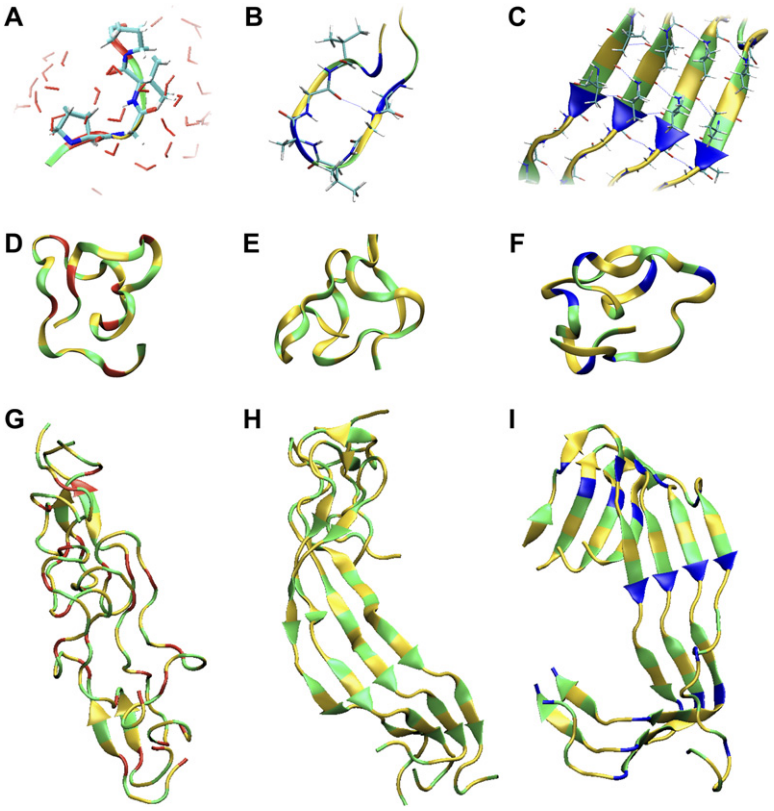

Figure 2. Representative Conformations of Monomers and Aggregates

Significantly populated structural elements: (A) polyproline II with nearby water, (B) hydrogen-bonded turn, and (C) β sheet. Backbone representations of monomers and aggregates: (D) and (G), $(GVPGV)_7$; (E) and (H), $(GGVGV)_7$; and (F) and (I), $(GVAGGV)_6$. Chain color indicates residue type: G, yellow; P, red; V, green; A, blue. Proline-containing elastin-like repeats, including GVPGV, form amorphous and disordered aggregates. By contrast, aggregates of sequences devoid of proline form amyloid-like structures with significant β-sheet content.

(Figure 1D) in which each sequence motif differs from its nearest neighbors by a single residue, as does the ordered set (PGV, GGV, GV, GVA). This minimalist approach facilitates systematic comparison between peptides, in addition to the computational averaging provided by repetitive sequences.

To investigate how the ability of peptides to self-assemble into elastin-like or amyloid-like fibrils is modulated by sequence, we performed extensive molecular-dynamics simulations of this model set of peptides in both monomeric and aggregated states. Our results show that elastin-like and amyloid-like peptides are separable on the basis of backbone hydration and conformational disorder and that these properties are modulated by proline and glycine. Furthermore, there is a threshold in combined proline and glycine content above which elastomeric properties become apparent and below which the formation of amyloid fibrils is possible. Taken together, our findings reveal a direct quantitative relationship between the intrinsic structural properties imparted by the sequence of self-assembling proteins and their propensity to form either amyloid or elastomeric fibrils.

Results and Discussion

Results for Monomers

The monomeric peptides adopt collapsed, water-swollen conformations reminiscent of the unfolded ensemble of globular protein domains (Figure 2D–2F) (Vendruscolo and Dobson, 2005). The structures exhibit a highly flexible polypeptide backbone, with exchanging conformations and overall structural disorder. Although these structures contain no classical α helix or β sheet, they are not random coils. Ordered structure is observed predominantly in the form of polyproline II (PPII) content (Figure 2A) and hydrogen-bonded turns (Figure 2B), both of which are local.

Due to the intrinsic conformational flexibility of the peptides, structural properties are best described in a probabilistic manner (Figure 3). Hydrogen-bonded turn content, X^*_{HB}, is anticorrelated to the degree of backbone hydration, X_W (Figure 3A). This anticorrelation reflects the local equilibrium of the peptide backbone: hydrogen bonding may be satisfied either with water or with other backbone groups in the form of hydrogen-bonded turns. The peptides separate into three clusters, ranked in order of increasing turn and decreasing hydration propensities: (1) proline-containing elastomers, (2) presumed "ambivalent" sequences including GGVGV, and (3) presumed amyloids including GVAGGV. At around 10%, the low-turn propensity of P-containing peptides may be rationalized in terms of conformational restrictions of the backbone induced by prolines. This restriction in dihedral angles is reflected in a higher propensity to adopt PPII conformations, which are characterized by high backbone hydration (Figures 2A and 3C). It has been established that PPII structure is present in significant amounts in the unfolded state of peptides and proteins (Pappu and Rose, 2002; Shi et al., 2002; García, 2004; Gnanakaran and García, 2005). In agreement with recent studies of peptides modeling the unfolded state of proteins (Zagrovic et al., 2005), the PPII structure observed is not extensive but is instead confined to one or two consecutive residues.

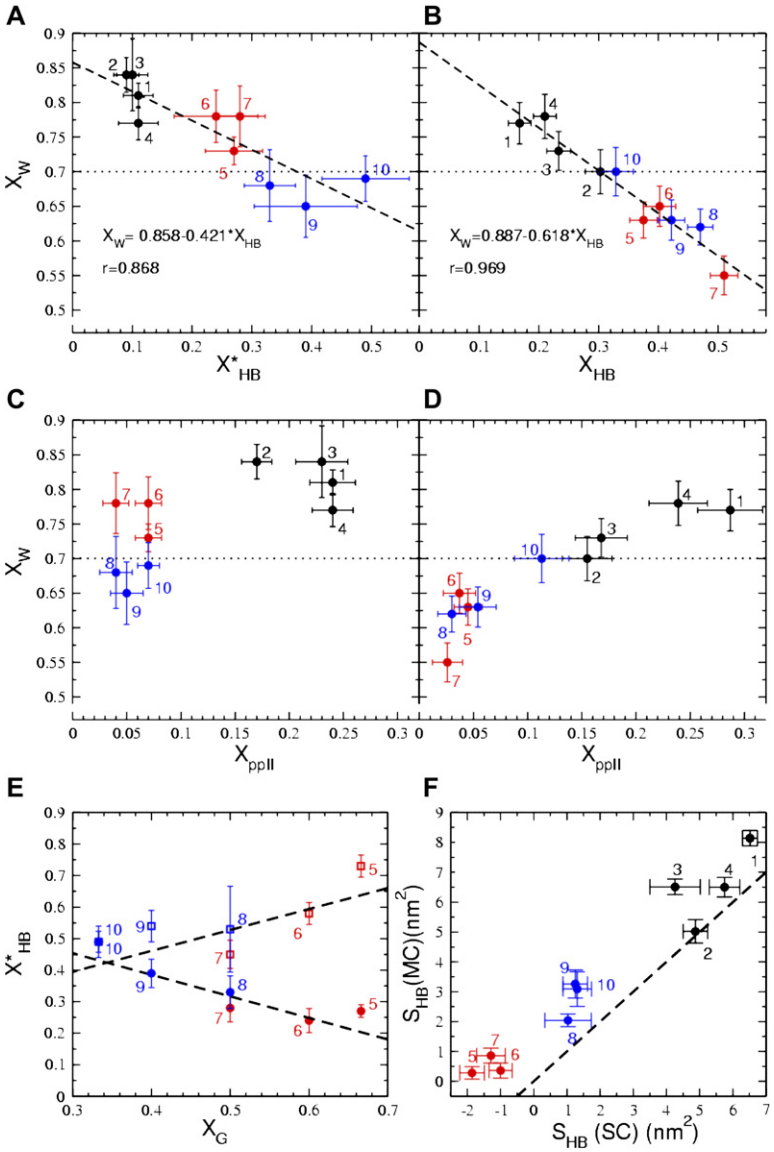

Figure 3. Average Structural Properties

Average structural properties computed from simulations of the ten model sequences: 1, $(PGV)_{12}$; 2, $(GGVPGV)_6$; 3, $(GVPGV)_7$; 4, $(GVAPGV)_6$; 5, $(GGV)_{12}$; 6, $(GGVGV)_7$; 7, $(GV)_{18}$; 8, $(GVAGGV)_6$; 9, $(GVGVA)_7$; 10, $(GVA)_{12}$, which are grouped into three classes: black, elastin-like; red, presumed ambivalent; blue, presumed amyloid. (A) Correlation between X_W, the number of water molecules bound per backbone hydrogen-bonding group, and X^*_{HB}, the hydrogen-bonded turn probability per residue, for monomers. (B) Correlation between X_W and X_{HB}, the peptide-peptide hydrogen-bonding propensity, for homoaggregates. (C and D) Correlation between X_W and X_{ppII}, the fractional polyproline II content, for monomers and aggregates, respectively. (E) Correlations between X^*_{HB} and X_G, the fraction of glycine content in the sequence, for simulations of monomers in water (closed symbols) and in vacuo (open symbols). (F) Correlation between S_{HB}, the buried hydrophobic surface area relative to a fully extended chain, for hydrated multichain (MC) and single-chain (SC) simulations. Error bars represent the standard deviation of the mean obtained by block averaging.

Results for Aggregates

Although it is established that amyloid fibrils contain cross-β sheets (Nelson et al., 2005; Nelson and Eisenberg, 2006; Petkova et al., 2002), the molecular organization of elastin aggregates is largely unknown. To probe the intrinsic ability of our model set of peptides to self-aggregate into extended β sheets, four extended chains were placed parallel and adjacent to each other in vacuo, so as to bias the starting conformation toward extended β sheets, and allowed to relax with unconstrained MD simulations, first in vacuo and then in explicit water. After 20 ns of simulation in water, the chains devoid of proline retained extensive cross-strand β sheets (Figures 2H and 2I). In contrast, the P-containing sequences did not form extended β sheets and instead remained highly disordered (Figure 2G). This is consistent with solid-state NMR data indicating a lack of β sheet in the aggregated state of bovine elastin (Pometun et al., 2004).These results indicate that the potential for forming amyloid-like structures is restricted to the sequences without proline. Presumed ambivalent and presumed amyloid sequences cannot be separated on the basis of peptide-peptide hydrogen bonding propensity, reflecting the capacity of both groups to form interstrand hydrogen bonding in extended β sheets. This is consistent with the ability of "ambivalent" sequences, like GGVGV, to form amyloid fibrils under appropriate conditions (Miao et al., 2003; Flamia et al., 2004).

Comparison of Monomers and Aggregates

The structural tendencies of the tetrameric aggregates are largely consistent with those of hydrated single chains. In particular, both states display a similar anticorrelation between hydration and peptide-peptide hydrogen bonding (Figures 3A and 3B). Significantly, elastin-like sequences retain higher backbone hydration and PPII propensities and form fewer peptide-peptide hydrogen bonds compared to the other model sequences (Figures 3A–3D). In support of these findings, high PPII content is observed for hydrophobic exons of elastin using circular dichroism (Bochicchio et al., 2004).

Qualitative agreement between monomers and aggregates is also seen in the amount of hydrophobic surface buried relative to the fully extended state, S_{HB} (Figure 3F). Three clusters emerge in order of increasing S_{HB}: (1) presumed ambivalent sequences, (2) presumed amyloids, and (3) elastin-like peptides, with an overall increase in surface burial for aggregates relative to single chains. The buried hydrophobic surface area of all ten peptides is considerably less than that of globular proteins of the same number of residues (cellobiohydrolase I, 19.6 nm^2 [Mattinen et al., 1997]; porcine peptide YY, 24.3 nm^2 [Keire et al., 2000]; villin headpiece, 24.3 nm^2 [McKnight et al., 1997], compared to an average of 5 ± 1 nm^2 for the elastin-like single chains). This observation is consistent with the lack of a compact folded state or buried hydrophobic core. It should be noted that the nonpolar packing of amyloid-like aggregates in our simulations is likely to underestimate the extent of hydrophobic burial in amyloid fibrils, in which layered β sheets have been observed (Stefani and Dobson, 2003; Nelson et al., 2005; Nelson and Eisenberg, 2006; Petkova et al., 2002). However, the overall consistency of the properties of monomers and aggregates suggests that the ability of the polypeptide chains to self-organize into either elastin-like or amyloid-like materials is determined by the intrinsic backbone propensities imparted by the sequence.

Rigid Prolines, Plastic Glycines

In the division between elastin-like and amyloidogenic peptides, there appear to be two major sequence determinants: proline and glycine. The combination of these two residues is remarkable in light of their radically different backbone plasticity. Proline is the primary determinant: its conformationally restricted main chain induces a significant propensity for PPII structure and an intrinsically reduced ability to form hydrogen-bonded turns and β sheet. Both of these structural tendencies lead to increased backbone hydration. Glycine is the secondary determinant: in the absence of proline, it is the fraction of glycine that determines the extent of backbone hydration in monomers. Indeed, the propensity for the backbone to form hydrogen-bonded turns, X^*_{HB}, decreases linearly with increasing glycine composition, X_G, in water (Figure 3E). This can be rationalized in terms of polypeptide chain entropy, as increasing glycine content increases the entropic cost of constraining the backbone in the formation of turns. Inversely, in the absence of competition for hydrogen bonding from the solvent, the backbone plasticity conferred by glycine helps better satisfy polar interactions through intramolecular hydrogen bonds. As a result, the probability to form hydrogen-bonded turns in vacuo increases with glycine content (Figure 3E). These results expose the dual character of glycine, which favors either ordered or disordered structures, depending on the context and environment. This adaptive capacity could contribute to the ambivalent nature of glycine-rich peptides devoid of proline, such as poly-(GGVGV).

A PG Composition Threshold

Based on the above results, both proline and glycine appear to play a central role in governing elastin-like properties in model sequences. The general significance of this finding is revealed in a two-dimensional diagram relating the P and G contents of many natural elastomeric protein domains, as well as experimentally confirmed amyloids (Figure 4). Astonishingly, a clear separation between elastomeric and amyloidogenic sequences is apparent using a wide array of protein sequences from many different species. Although the hydrophobic domains of mammalian, amphibian, avian, and fish elastin span a wide region of the plot, the overwhelming majority of these sequences are found above a linear threshold of combined P and G content. The fact that approximately two glycines are equivalent to one proline at this threshold confirms the role of proline as the primary determinant of elastin's properties. Remarkably, the compliance with a PG composition threshold is not limited to the hydrophobic domains of elastin, but is also observed for other elastomeric proteins. These include the elastic domains of preColP, which confer elasticity on mussel byssus thread (Coyne et al., 1997); abductin, from the hinge of the scallop's shell (Bochicchio et al., 2005); and resilin, the resilient energy storage protein important in insect flight (Elvin et al., 2005).

Below the PG-threshold are many sequences that have been found to form amyloid fibrils in vivo or in vitro, some of which are linked to diseases and others that have no known associated pathology. Of particular interest are the peptide fragments of islet amyloid polypeptide (IAPP, associated with type II diabetes). Human IAPP contains the amyloidogenic fragment SNNFGAILSS, while rat IAPP has SNNLGPVLPP at the same location in its sequence. The latter fragment, which has a combined PG composition well above the threshold, cannot form amyloid fibrils in vitro. It is thought that this sequence difference protects rats from acquiring type II diabetes (Tenidis et al., 2000; Westermark et al., 1990). It has also been demonstrated that randomly scrambled variants of the prion domain of Ure2p can form amyloid fibrils in vivo and in vitro, emphasizing the importance of sequence composition over exact sequence in amyloid formation (Ross et al., 2004). The effects of P and G on the inhibition of amyloid fibrillization have separately been the object of previous studies (Parrini et al., 2005; Williams et al., 2004). The present evidence for the existence of a quantitative PG threshold above which amyloid formation is impeded is compatible with the fact that even sequences with relatively high content of either P or G can form amyloid fibers.

Sequences of the various classes of spider silks also appear to satisfy the PG threshold. The most elastic forms of spider silk, flagelliform silk and major ampullate spindroin 2 (MaSp2), are found above the threshold in the region populated by elastin's hydrophobic domains, while the two forms of silk imparting strength to the dragline and the web (respectively, major ampullate spindroin 1, MaSp1, and minor ampullate spindroin, MiSp) appear below the threshold (Gatesy et al., 2001). The rigid silks (aciniform, AcSp [Hayashi et al., 2004], used for wrapping prey, and tubuliform, TuSp1 [Garb and Hayashi, 2005], used to encase eggs) are found in the amyloidogenic region. Similarly, flexible and rigid lizard egg shells are located respectively above and below the composition threshold, with the primary difference being a 2-fold increase in P content for the flexible-type shells (Sexton et al., 2005).

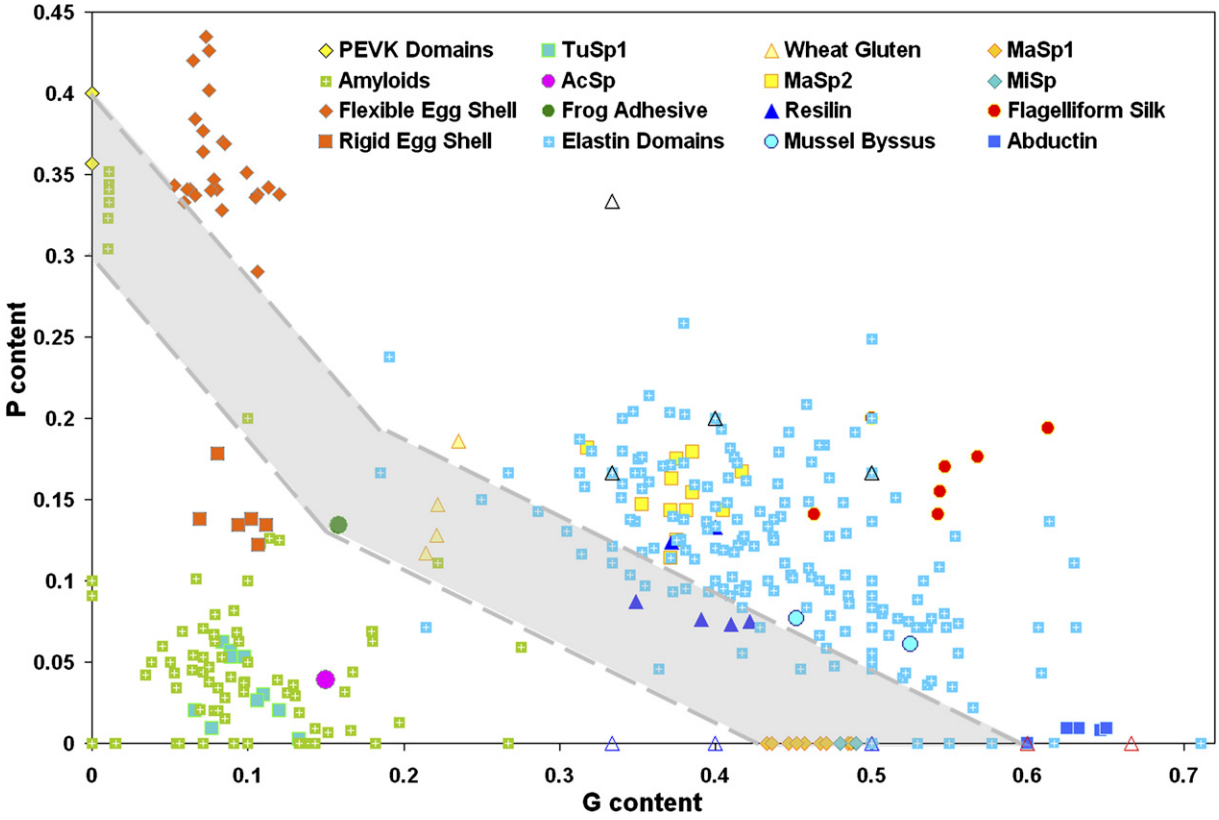

Figure 4. Proline and Glycine Composition of Elastomeric and Amyloidogenic Sequences

A two-dimensional plot correlating proline and glycine content for a wide variety of peptides. The coexistence region (shaded in gray) contains P and G compositions consistent with both amyloidogenic and elastomeric properties. Elastomeric proteins, including the domains of elastin, major ampullate spindroin 2 (MaSp2), flagelliform silk, the elastic domains of mussel byssus thread, and abductin, appear above a composition threshold (upper dashed line). Amyloidogenic sequences are primarily found below the PG-threshold, along with rigid lizard egg shells, tubuliform silk (TuSp1), a protective silk for spider eggs, and aciniform silk (AcSp), used for wrapping prey. The coexistence region contains amyloid-like peptides as well as the elastomeric adhesive produced by the frog *Notaden bennetti*, the PEVK domains of titin, wheat glutenin protein, and the strongest spider silks, namely major ampullate spindroin 1 (MaSp1) and minor ampullate spindroin (MiSp). Our model peptides are also shown (open triangles) with color coding defined in Figure 3. Refer to the Supplemental Data for sequences and corresponding references.

The transition in composition space between elastomers and amyloids does not appear to be an abrupt one. Rather, a coexistence region includes sequences that are either elastomeric, amyloidogenic, or both. Elastomeric proteins in this region include glutenin (found in wheat gluten) (Tatham et al., 2001), the adhesive produced by the Australian frog *Notaden bennetti* (Graham et al., 2005), and the PEVK domains of muscle protein titin (Ma and Wang, 2003; Tatham and Shewry, 2000). There are also amyloidogenic sequences, including silk moth chorion protein (Iconomidou et al., 2000) and the ambivalent sequence poly-(GGVGV) (Miao et al., 2003; Flamia et al., 2004). The boundaries indicated in Figure 4 are not intended to be definite, and the delineation and characterization of the threshold will require further investigation. Accordingly, the PG diagram should not be taken to imply that random sequences satisfying the threshold in combined PG content would be elastomeric or that the composition of proline and glycine is the sole measure of a protein's ability to form amyloid fibrils. For a given composition, changes in primary sequence may influence hydration and peptide-peptide hydrogen bonding, and therefore aggregation tendencies. The effect of sequence variability may be especially important in the coexistence region. It should also be noted that the sequence of human elastin is not random with regard to the location of prolines and glycines (see Supplemental Data available with this article online for an analysis of the interproline and interglycine spacing). Additional factors, such as the compositions of other residues and solution conditions, are expected to contribute to the modulation of protein aggregation tendencies. However, it is striking that all known natural elastomeric sequences satisfy a composition threshold, which therefore appears to constitute a necessary condition for the onset of elastomeric properties.

Experimental Crossvalidation

Further support for the above findings is provided by the crossvalidation of independent observations made on the basis of (1) combined PG composition, (2) structural properties obtained from atomistic simulations of model peptides, and (3) in vitro studies of recombinant peptides containing the same repeat motifs. The structural properties computed for the model sequences (Figure 1D) are consistent with the distribution of these sequences on the PG diagram (Figure 4). The elastin-like peptides

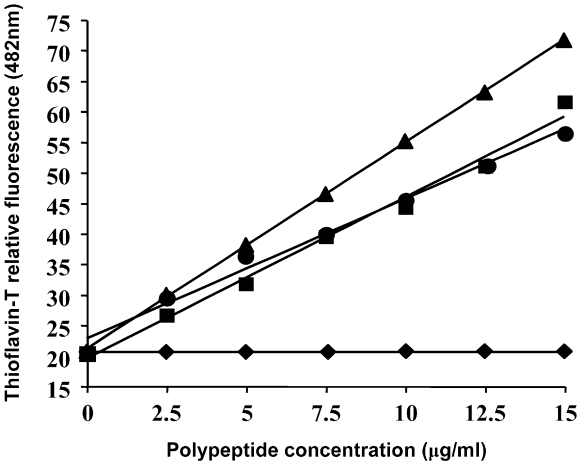

Figure 5. Thioflavin-T-Binding Assay for Amyloid Fibril Formation

Binding of thioflavin-T to EP20-24-24 (◆), EP20-24-24[P/G] in 0.5 M NaCl (■), EP20-24-GVA (▲), and EP20-GVA-GVA (●). It has previously been demonstrated that EP20-24-24[P/G] (in the presence of salt) did not coacervate and instead formed amyloid-like fibril structures (Miao et al., 2003). EP20-24-24[P/G] contains the GGVGVA tandem repeat, which is found below the threshold on the PG diagram. The dependence of fluorescence on peptide concentration for both GVA recombinant peptides is similar to that of this known amyloidogenic sequence.

are found above the threshold, while the remaining sequences are found in the coexistence region or below the PG-threshold. Based on the composition threshold in the PG diagram (Figure 4), poly-GVA is predicted to form amyloid-like structures. Accordingly, in our simulations of both monomers and aggregates, $(GVA)_{12}$ exhibited structural tendencies similar to $(GVAGGV)_6$, an amyloidogenic sequence motif (Miao et al., 2003), with low hydration and high peptide-peptide hydrogen-bonding propensity. To test this prediction, recombinant polypeptides that contain tandem GVA repeats were produced: EP20-24-[GVA] and EP20-[GVA]-[GVA], where the sequence of the [GVA] domain is provided in Figure 1B. Both peptides formed fibrillar precipitates and were subjected to the thioflavin-T binding assay for amyloid fibril detection (Figure 5) (LeVine, 1999). Thioflavin-T fluorescence for EP20-24-24, the polypeptide with native human elastin sequence PGVGVA, was minimal and independent of concentration. In contrast, EP20-24-24[P/G] in 0.5 M NaCl, which has been shown previously to form amyloid-like fibrils by electron microscopy (Miao et al., 2003), exhibited a thioflavin-T fluorescence, which was strongly concentration dependent. Similarly, EP20-24-GVA and EP20-GVA-GVA both exhibited strong concentration dependence for dye binding comparable to EP20-24-24[P/G]. The confirmation that the GVA motif promotes amyloid formation highlights the capacity of the PG diagram to accurately predict peptide aggregation tendencies.

Toward a Unified Model of Elastomeric Structure and Function

The existence of a threshold of combined proline and glycine composition observed by elastomeric proteins points to a remarkably simple and robust design principle, whereby the onset of elastomeric function is controlled by the intrinsic backbone properties conferred by just two amino acid residue types. The analysis of the results obtained from our minimalist set of peptide sequences indicates that both amino acids help keep the polypeptidic backbone of elastomers disordered and hydrated, though for opposite reasons: prolines, because they are too stiff to form secondary structure, and glycines, because in the presence of water, they are too flexible to do so. This finding is consistent with the fact that these are the two residue types most likely to be found in loops, most of which are solvent exposed, rather than in the secondary structure of globular proteins (Crasto and Feng, 2001).

Furthermore, the interplay between hydration and disorder of the polypeptide chain is highlighted by the apparent threshold in backbone hydration separating elastin-forming aggregates from their amyloid-forming counterparts (indicated as a dotted line in Figures 3A–3D). In support of the existence of a hydration threshold, not only is elastin brittle when dry, but solid-state NMR has also shown that a relative mass of water greater than 30% is required for the onset of conformational flexibility (Perry et al., 2002). Converting the apparent threshold of $X_W = 0.7$ (the average number of water molecules bound to each backbone hydrogen-bonding group) into a percent mass of water leads to the following results for elastin-like sequences: PGVPGV, 27%; GGVPGV, 29%; GVPGV, 28%; GVAPGV, 32%. The excellent agreement with the experimentally determined hydration threshold suggests that the emergence of elastomeric properties of elastin is controlled by backbone hydration.

Taken together, the above findings have general implications for elastomeric self-assembly and function (Figure 6). Two major entropic forces are at play in the folding and aggregation of biopolymers: polypeptide chain entropy (ΔS_C) and hydrophobic packing (ΔS_{HP}). Chain entropy opposes both folding (Figure 6C) and full extension of the polypeptide chain (Figure 6A) since both events dramatically decrease the number of accessible conformations. Hydrophobic forces drive the emergence of collapsed states of polypeptide chains (Figure 6B), which precede formation of the hydrophobic core of globular proteins (Cheung et al., 2002) (Figure 6C). Like the native state of globular proteins, the molecular structure underlying amyloid fibrils is characterized by an ordered, water-excluding core containing extensive secondary structure (Figure 6F). The present work suggests that the reason elastomeric chains remain hydrated and disordered even after aggregation is that their backbone is inherently unable to form extensive self-interactions (Figure 6E). Accordingly, the functional state of an elastomer is incompatible with the amyloid state. In turn, this supports a model of elastomeric materials in which, consistent with (1) rubber-like elasticity and (2) the existence of a hydration threshold, backbone entropy and hydration play a central role in their function. Because the aggregated state contains relatively few hydrogen-bonded self-interactions, the polypeptide chains can readily extend under strain (Figure 6D). Subsequently, both chain entropy and hydrophobic packing contribute to elastic recoil.

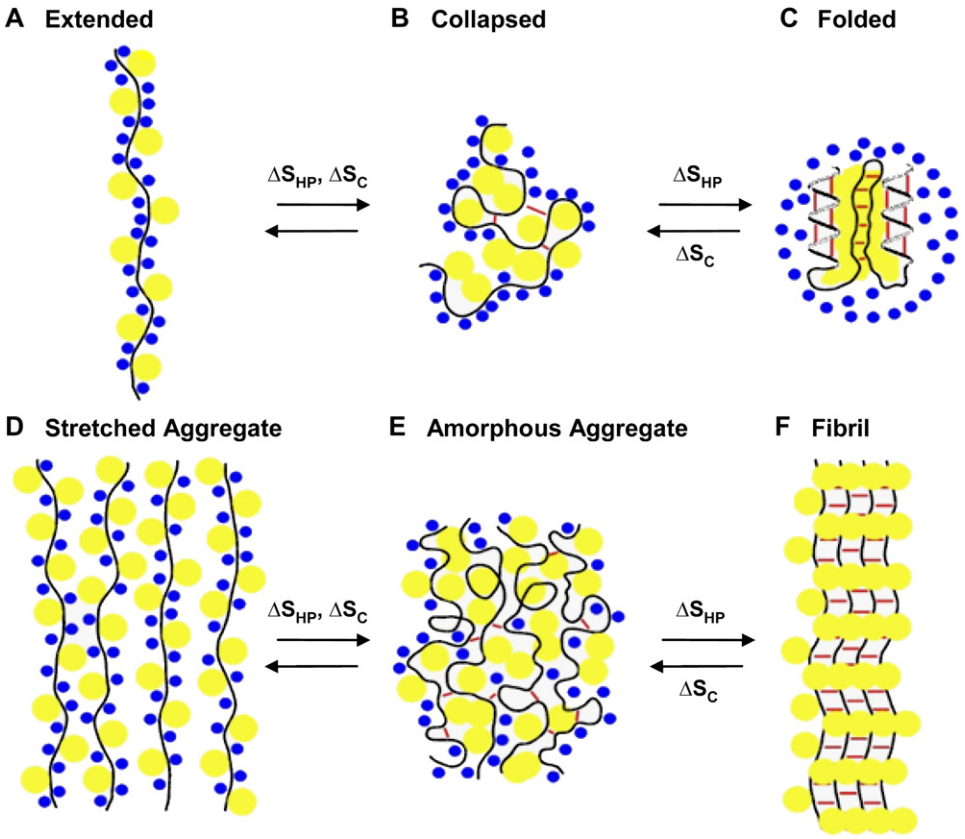

Figure 6. Role of Entropy in Polypeptide Self-Organization

Schematic diagram of possible protein states, highlighting the polypeptide backbone (black), nonpolar side chains (yellow), water molecules solvating the polypeptide backbone (blue), and peptide-peptide H bonds (red). (A) Extended single chain; (B) disordered, water-swollen monomer representing the unfolded state of a globular protein; (C) folded globular protein; (D) extended aggregate; (E) hydrated amorphous aggregate characterized by an ensemble of many degenerate conformations; (F) β sheet of amyloid fibers. Both protein folding and amyloid formation are favored by hydrophobic forces (ΔS_{HP}) but opposed by the polypeptide chain entropy (ΔS_C). Both forces oppose the extension of disordered hydrophobic chains. The present study suggests that the single most important feature of elastomeric chains is their inability to fold into a native structure or to form extended cross-β sheets, which allows their aggregates to remain amorphous and to readily undergo extension and elastic recoil.

Conclusion

We derive the following simple conclusions: (1) a fundamental requirement for elastomeric domains is to remain disordered, even when aggregated; (2) disorder is an indirect consequence of the inability of the polypeptide to form a compact, water-excluding core involving extensive backbone self-interactions; (3) most significantly, the fact that a minimum threshold of combined P, G content appears to be fulfilled by proteins forming such diverse biomaterials as human aorta, spider silk, and lizard egg shells suggests that maintaining a critical level of disorder is not only a fundamental requirement but may very well constitute the single most essential design principle of self-assembling elastic proteins. These insights establish a framework for more detailed studies of the physical determinants of structure in elastomers and amyloids, as well as the modulation of elastomeric properties. Ultimately, this understanding will help advance the rational design of self-assembling biomimetic materials, such as artificial skin, and facilitate the development of therapeutic approaches to combat the increasingly prevalent amyloid diseases.

Experimental Procedures

Molecular Dynamics

Blocked single-chain and tetrameric assemblies of each of the ten periodic polypeptide sequences depicted in Figure 1D were subjected to extensive molecular dynamics (MD) simulations in explicit water at 300K with the OPLS/AA/L force field (Jorgensen et al., 1996; Kaminski et al., 2001) and the TIP3P model for water (Jorgensen et al., 1983). Simulations of monomeric and aggregated systems were performed for 60 ns and 20 ns, respectively, for a total of 800 ns of simulation, with the GROMACS MD simulation package (Berendsen et al., 1995; Lindahl et al., 2001). Further details of simulation methodology and structural analysis are provided as Supplemental Data.

Recombinant Polypeptides

Construction and production of elastin polypeptides EP20-24-24 and EP20-24-24[P/G] has been described elsewhere (Miao et al., 2003; Bellingham et al., 2001). EP20-24-GVA and EP20-GVA-GVA were constructed by replacing oligonucleotides coding for 11 repeats of GVA with the seven repeats of PGVGVA in one or two exon24 of EP20-24-24, respectively. All polypeptides were confirmed by amino acid analysis and Q-TOF mass spectrometry by using the facilities of the Advanced Protein Technology Centre, Hospital for Sick Children.

Thioflavin-T Assay

The assay for binding of thioflavin-T was performed according to the method of LeVine (LeVine, 1999). Briefly, polypeptides were dissolved in water or 0.5 M NaCl at a concentration of 0.5 μg/μl and left at 4°C overnight. Then 10 μl of polypeptide was incrementally added to 2 ml of 3 μM thioflavin-T in potassium phosphate buffer (50 mM, pH 6.0). Fluorescence at 482 nm (excitation 450 nm) was monitored with a Hitachi F-2500 fluorescence spectrophotometer.

Supplemental Data

Supplemental Data include a detailed description of computational methods, data showing the spacing of proline and glycine in human elastin, as well as the sequences and corresponding references used in the construction of Figure 4. The Supplemental Data are available at http://www.structure.org/cgi/content/full/14/11/1667/DC1/.

Acknowledgments

We thank the Hospital for Sick Children (S.B.) and the Heart and Stroke Foundation of Ontario (F.W.K.) for support. S.R. is the recipient of a Canada Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada. M.M. was a recipient of a Fellowship from the Heart and Stroke Foundation of Canada. R.P. is a Canada Research Chairs Programme chairholder.

Received: July 25, 2006
Revised: September 8, 2006
Accepted: September 12, 2006
Published: November 14, 2006

References

Baer, M., Schreiner, E., Kohlmeyer, A., Rousseau, R., and Marx, D. (2006). Inverse temperature transition of a biomimetic elastin model: reactive flux analysis of folding/unfolding and its coupling to solvent dielectric relaxation. J. Phys. Chem. B *110*, 3576–3587.

Becker, N., Oroudjev, E., Mutzi, S., Cleveland, J.P., Hansma, P.K., Hayashi, C.Y., Makarov, D.E., and Hansma, H.G. (2003). Molecular nanosprings in spider capture-silk threads. Nat. Mater. *2*, 278–283.

Bellingham, C.M., Woodhouse, K.A., Robson, P., Rothstein, S.J., and Keeley, F.W. (2001). Self-aggregation characteristics of recombinantly expressed human elastin polypeptides. Biochim. Biophys. Acta *1550*, 6–19.

Bellingham, C.M., Lillie, M.A., Gosline, J.M., Wright, G.M., Starcher, B.C., Bailey, A.J., Woodhouse, K.A., and Keeley, F.W. (2003). Recombinant human elastin polypeptides self-assemble into biomaterials with elastin-like properties. Biopolymers *70*, 445–455.

Berendsen, H.J.C., van der Spoel, D., and van Drunen, R. (1995). GROMACS: a message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. *91*, 43–56.

Bochicchio, B., and Tamburro, A.M. (2002). Polyproline II structure in proteins: identification by chiroptical spectroscopies, stability, and functions. Chirality *14*, 782–792.

Bochicchio, B., Aït-Ali, A., Tamburro, A.M., and Alix, A.J.P. (2004). Spectroscopic evidence revealing polyproline II structure in hydrophobic, putatively elastomeric sequences encoded by specific exons of human tropoelastin. Biopolymers *73*, 484–493.

Bochicchio, B., Jimenez-Oronoz, F., Pepe, A., Blanco, M., Sandberg, L.B., and Tamburro, A.M. (2005). Synthesis of and structural studies on repeating sequence of abductin. Macromol. Biosci. *5*, 502–511.

Cheung, M.S., García, A.E., and Onuchic, J.N. (2002). Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. Proc. Natl. Acad. Sci. USA *99*, 685–690.

Coyne, K.J., Qin, X., and Waite, J.H. (1997). Extensible collagen in mussel byssus: a natural block copolymer. Science *277*, 1830–1832.

Crasto, C.J., and Feng, J. (2001). Sequence codes for extended conformation: a neighbor-dependent sequence analysis of loops in proteins. Proteins *42*, 399–413.

Dobson, C.M. (2003). Protein folding and misfolding. Nature *426*, 884–890.

Elvin, C.M., Carr, A.G., Huson, M.G., Maxwell, J.M., Pearson, R.D., Vuocolo, T., Liyou, N.E., Wong, D.C.C., Merritt, D.J., and Dixon, N.E. (2005). Synthesis and properties of crosslinked recombinant pro-resilin. Nature *437*, 999–1002.

Fändrich, M., Fletcher, M.A., and Dobson, C.M. (2001). Amyloid fibrils from muscle myoglobin. Nature *410*, 165–166.

Flamia, R., Zhdan, P.A., Martino, M., Castle, J.E., and Tamburro, A.M. (2004). AFM study of the elastin-like biopolymer poly-(ValGlyGlyValGly). Biomacromolecules *5*, 1511–1528.

Floquet, N., Héry-Huynh, S., Dauchez, M., Derreumaux, P., Tamburro, A.M., and Alix, A.J.P. (2004). Structural characterization of VGVAPG, an elastin-derived peptide. Biopolymers *76*, 266–280.

Fowler, D.M., Koulov, A.V., Alory-Jost, C., Marks, M.S., Balch, W.E., and Kelly, J.W. (2006). Functional amyloid formation within mammalian tissue. PLoS Biol. *4*, e6.

Garb, J.E., and Hayashi, C.Y. (2005). Modular evolution of egg case silk genes across orb-weaving spider superfamilies. Proc. Natl. Acad. Sci. USA *102*, 11379–11384.

García, A. (2004). Characterization of non-alpha helical conformations in Ala peptides. Polym. *45*, 669–676.

Gatesy, J., Hayashi, C., Motriuk, D., Woods, J., and Lewis, R. (2001). Extreme diversity, conservation, and convergence of spider silk fibroin sequences. Science *291*, 2603–2605.

Graham, L.D., Glattauer, V., Huson, M.G., Maxwell, J.M., Knott, R.B., White, J.W., Vaughan, P.R., Peng, Y., Tyler, M.J., Werkmeister, J.A., et al. (2005). Characterization of a protein-based adhesive elastomer secreted by the Australian frog *Notaden bennetti*. Biomacromolecules *6*, 3300–3312.

Gnanakaran, S., and García, A.E. (2005). Helix-coil transition of alanine peptides in water: force field dependence on the folded and unfolded structures. Proteins *59*, 773–782.

Hayashi, C.Y., Blackledge, T.A., and Lewis, R.V. (2004). Molecular and mechanical characterization of aciniform silk: uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. Mol. Biol. Evol. *21*, 1950–1959.

Hoeve, C., and Flory, P.J. (1974). The elastic properties of elastin. Biopolymers *13*, 677–686.

Iconomidou, V.A., Vriend, G., and Hamodrakas, S.J. (2000). Amyloids protect the silkmoth oocyte and embryo. FEBS Lett. *479*, 141–145.

Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R., and Klein, M. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926–935.

Jorgensen, W.L., Maxwell, D.S., and Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc. *118*, 11225–11236.

Kaminski, G.A., Friesner, R.A., Tirado-Rives, J., and Jorgensen, W.L. (2001). Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J. Phys. Chem. B *105*, 6474–6487.

Keeley, F.W., Bellingham, C.M., and Woodhouse, K.A. (2002). Elastin as a self-organizing biomaterial: use of recombinantly expressed human elastin polypeptides as a model for investigations of structure and self-assembly of elastin. Philos. Trans. R. Soc. Lond. B Biol. Sci. *357*, 185–189.

Keire, D.A., Kobayashi, M., Solomon, T.E., and Reeve, J.R., Jr. (2000). Solution structure of monomeric peptide YY supports the functional significance of the PP-fold. Biochemistry *39*, 9935–9942.

LeVine, H. (1999). Quantification of β-sheet amyloid fibril structures with thioflavin T. Methods Enzymol. *309*, 274–284.

Li, B., Alonso, D.O.V., and Daggett, V. (2001). The molecular basis for the inverse temperature transition of elastin. J. Mol. Biol. *305*, 581–592.

Lindahl, E., Hess, B., and van der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. J. Mol. Model. *7*, 305–317.

Ma, K., and Wang, K. (2003). Malleable conformation of the elastic PEVK segment of titin: non-co-operative interconversion of polyproline II helix, β-turn and unordered structures. Biochem. J. *374*, 687–695.

Mattinen, M.L., Kontteli, M., Kerovuo, J., Linder, M., Annila, A., Lindeberg, G., Reinikainen, T., and Drakenberg, T. (1997). Three-dimensional structures of three engineered cellulose-binding domains of cellobiohydrolase I from Trichoderma reesei. Protein Sci. *6*, 294–303.

McKnight, C.J., Matsudaira, P.T., and Kim, P.S. (1997). NMR structure of the 35-residue villin headpiece subdomain. Nat. Struct. Biol. *4*, 180–184.

Miao, M., Bellingham, C.M., Stahl, R.J., Sitarz, E.E., Lane, C.J., and Keeley, F.W. (2003). Sequence and structure determinants for the self-aggregation of recombinant polypeptides modeled after human elastin. J. Biol. Chem. *278*, 48553–48562.

Nelson, R., and Eisenberg, D. (2006). Recent atomic models of amyloid fibril structure. Curr. Opin. Struct. Biol. *16*, 260–265.

Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.O., Riekel, C., Grothe, R., and Eisenberg, D. (2005). Structure of the cross-β spine of amyloid-like fibrils. Nature *435*, 773–778.

Pappu, R.V., and Rose, G.D. (2002). A simple model for polyproline II structure in unfolded states of alanine-based peptides. Protein Sci. *11*, 2437–2455.

Parrini, C., Taddei, N., Ramazzotti, M., Degl'Innocenti, D., Ramponi, G., Dobson, C.M., and Chiti, F. (2005). Glycine residues appear to be evolutionarily conserved for their ability to inhibit aggregation. Structure *13*, 1143–1151.

Perry, A., Stypa, M.P., Tenn, B.K., and Kumashiro, K.K. (2002). Solid-state ^{13}C NMR reveals effects of temperature and hydration on elastin. Biophys. J. *82*, 1086–1095.

Petkova, A.T., Ishii, Y., Balbach, J.J., Antzutkin, O.N., Leapman, R.D., Delaglio, F., and Tycko, R. (2002). A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR. Proc. Natl. Acad. Sci. USA *99*, 16742–16747.

Pometun, M.S., Chekmenev, E.Y., and Wittebort, R.J. (2004). Quantitative observation of backbone disorder in native elastin. J. Biol. Chem. *279*, 7982–7987.

Ross, E.D., Baxa, U., and Wickner, R.B. (2004). Scrambled prion domains form prions and amyloid. Mol. Cell. Biol. *24*, 7206–7213.

Sexton, O.J., Bramble, J.E., Heisler, I.L., Phillips, C.A., and Cox, D.L. (2005). Eggshell composition of squamate reptiles: relationship between eggshell permeability and amino acid distribution. J. Chem. Ecol. *31*, 2391–2401.

Shi, Z., Olson, C.A., Rose, G.D., Baldwin, R.L., and Kallenbach, N.R. (2002). Polyproline II structure in a sequence of seven alanine residues. Proc. Natl. Acad. Sci. USA *99*, 9190–9195.

Stefani, M., and Dobson, C.M. (2003). Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. J. Mol. Med. *81*, 678–699.

Tamburro, A.M., Bochicchio, B., and Pepe, A. (2003). Dissection of human tropoelastin: exon-by-exon chemical synthesis and related conformational studies. Biochemistry *42*, 13347–13362.

Tamburro, A.M., Pepe, A., Bochicchio, B., Quaglino, D., and Ronchetti, I.P. (2005). Supramolecular amyloid-like assembly of the polypeptide sequence coded by exon 30 of human tropoelastin. J. Biol. Chem. *280*, 2682–2690.

Tatham, A.S., and Shewry, P.R. (2000). Elastomeric proteins: biological roles, structures and mechanisms. Trends Biochem. Sci. *25*, 567–571.

Tatham, A.S., Hayes, L., Shewry, P.R., and Urry, D.W. (2001). Wheat seed proteins exhibit a complex mechanism of protein elasticity. Biochim. Biophys. Acta *1548*, 187–193.

Tenidis, K., Waldner, M., Bernhagen, J., Fischle, W., Bergmann, M., Weber, M., Merkle, M., Voelter, W., Brunner, H., and Kapurniotu, A. (2000). Identification of a penta- and hexapeptide of islet amyloid polypeptide (IAPP) with amyloidogenic and cytotoxic properties. J. Mol. Biol. *295*, 1055–1071.

Vendruscolo, M., and Dobson, C.M. (2005). Towards complete descriptions of the free-energy landscapes of proteins. Philos. Transact. A Math. Phys. Eng. Sci. *363*, 433–452.

Venkatachalam, C.M., and Urry, D.W. (1981). Development of a linear helical conformation from its cyclic correlate. β-spiral model of the elastin poly(pentapeptide) $(VPGVG)_n$. Macromolecules *14*, 1225–1229.

Vrhovski, B., and Weiss, A.S. (1998). Biochemistry of tropoelastin. Eur. J. Biochem. *258*, 1–18.

Westermark, P., Engstrom, U., Johnson, K.H., Westermark, G.T., and Betsholtz, C. (1990). Islet amyloid polypeptide: pinpointing amino acid residues linked to amyloid formation. Proc. Natl. Acad. Sci. USA *87*, 5036–5040.

Williams, A.D., Portelius, E., Kheterpal, I., Guo, J., Cook, K.D., Xu, Y., and Wetzel, R. (2004). Mapping abeta amyloid fibril secondary structure using scanning proline mutagenesis. J. Mol. Biol. *335*, 833–842.

Zagrovic, B., Lipfert, J., Sorin, E.J., Millett, I.S., van Gunsteren, W.F., Doniach, S., and Pande, V.S. (2005). Unusual compactness of a polyproline type II structure. Proc. Natl. Acad. Sci. USA *102*, 11698–11703.