Insight into the Structure of Amyloid Fibrils from the Analysis of Globular Proteins

Antonio Trovato[1,2*], Fabrizio Chiti[3], Amos Maritan[1,2,4], Flavio Seno[1,2,4]

1 Consorzio Nazionale Interuniversitario per le Scienze Fisiche della Materia, Unità di Padova, Padua, Italy, **2** Dipartimento di Fisica "G. Galilei," Università di Padova, Padua, Italy, **3** Dipartimento di Scienze Biochimiche, Università di Firenze, Florence, Italy, **4** Istituto Nazionale di Fisica Nucleare, Sezione di Padova, Padua, Italy

The conversion from soluble states into cross-β fibrillar aggregates is a property shared by many different proteins and peptides and was hence conjectured to be a generic feature of polypeptide chains. Increasing evidence is now accumulating that such fibrillar assemblies are generally characterized by a parallel in-register alignment of β-strands contributed by distinct protein molecules. Here we assume a universal mechanism is responsible for β-structure formation and deduce sequence-specific interaction energies between pairs of protein fragments from a statistical analysis of the native folds of globular proteins. The derived fragment–fragment interaction was implemented within a novel algorithm, prediction of amyloid structure aggregation (PASTA), to investigate the role of sequence heterogeneity in driving specific aggregation into ordered self-propagating cross-β structures. The algorithm predicts that the parallel in-register arrangement of sequence portions that participate in the fibril cross-β core is favoured in most cases. However, the antiparallel arrangement is correctly discriminated when present in fibrils formed by short peptides. The predictions of the most aggregation-prone portions of initially unfolded polypeptide chains are also in excellent agreement with available experimental observations. These results corroborate the recent hypothesis that the amyloid structure is stabilised by the same physicochemical determinants as those operating in folded proteins. They also suggest that side chain–side chain interaction across neighbouring β-strands is a key determinant of amyloid fibril formation and of their self-propagating ability.

Citation: Trovato A, Chiti F, Maritan A, Seno F (2006) Insight into the structure of amyloid fibrils from the analysis of globular proteins. PLoS Comput Biol 2(12): e170. doi:10.1371/journal.pcbi.0020170

Introduction

An increasing number of human pathologies are associated with the conversion of peptides and proteins from their soluble functional forms into well-defined fibrillar aggregates [1,2]. The diseases can be broadly grouped into neurodegenerative conditions, in which fibrillar aggregation occurs in the brain, nonneuropathic localised amyloidoses, in which aggregation occurs in a single type of tissue other than the brain, and nonneuropathic systemic amyloidoses, in which aggregation occurs in multiple tissues [1,2]. The fibrillar deposits associated with human pathologies are generally described as amyloid fibrils when they accumulate extracellularly, whereas the term "intracellular inclusions" has been suggested to be more appropriate when fibrils morphologically and structurally related to extracellular amyloid form inside the cell [3].

Amyloid formation is not restricted, however, to those polypeptide chains that have recognised links to protein deposition diseases. Several other proteins that have no such link have been found to form fibrillar aggregates in vitro with morphological, structural, and tinctorial properties that allow them to be classified as amyloid-like fibrils [4,5]. This finding has led to the idea that the ability to form the amyloid structure is an inherent property of polypeptide chains, encoded in main backbone chain interactions. From a theoretical perspective it was also recently shown that simple considerations of geometry and symmetry are sufficient to explain, within the same sequence-independent framework, the emergence of a limited menu of native-like conformations for a single chain and of β-aggregate structures for multiple chains [6].

The generic ability to form the amyloid structure has apparently been exploited by living systems for specific purposes, as some organisms have been found to convert, during their normal physiological life cycle, one or more of their endogenous proteins into amyloid-like fibrils that have functional properties rather than deleterious effects [7–9]. Perhaps the most surprising of these functions is the ability of amyloid-like fibrillar aggregates to serve as a nonchromosomal genetic element. Proteins such as Ure2p and Sup35p *(Saccharomyces cerevisiae)* or HET-s *(P. anserina)* can adopt a fibrillar conformation that, in addition to giving rise to specific phenotypes, appears to be self-propagating, transmissible, and infectious [10].

In their soluble states, the proteins able to form fibrillar aggregates do not share any obvious sequence identity or structural homology to each other. In spite of these differences in the precursor proteins, morphological inspection reveals common properties in the resulting fibrils [11]. Images obtained with transmission electron microscopy or atomic

Editor: Eugene Shakhnovich, Harvard University, United States of America

Received June 27, 2006; **Accepted** October 30, 2006; **Published** December 15, 2006

A previous version of this article appeared as an Early Online Release on October 30, 2006 (doi:10.1371/journal.pcbi.0020170.eor).

Abbreviations: PASTA, prediction of amyloid structure aggregation; ss-NMR, solid-state nuclear magnetic resonance; PIRA, parallel in-register arrangement

* To whom correspondence should be addressed. E-mail: trovato@pd.infn.it

Synopsis

In many fatal neurodegenerative diseases, including Alzheimer, Parkinson, and spongiform encephalopathies, proteins aggregate into specific fibrous structures to form insoluble plaques known as amyloid. The amyloid structure may also play a nonaberrant role in different organisms. Many globular proteins, folding to their biologically functional native structures in vivo, can be induced to aggregate into amyloid-like fibrils under suitable conditions in vitro. One hallmark of amyloid structure is a specific supramolecular architecture called cross-beta structure, held together by hydrogen bonds extending repeatedly along the fibril axis, but intermolecular interactions are yet unknown at the amino-acid level except for very few cases. In this study, the authors present an algorithm, called prediction of amyloid structure aggregation (PASTA), to computationally predict which portions of a given protein or peptide sequence forming amyloid fibrils are stabilizing the corresponding cross-beta structure and the specific intermolecular pattern of hydrogen-bonded amino acids. PASTA is based on the assumption that the same amino acid–specific interactions stabilizing hydrogen bond patterns in native structures of globular proteins are also employed by nature in amyloid structure. The successful comparison of the authors' prediction with available experimental data supports the existence of a unique framework to describe protein folding and aggregation.

force microscopy reveal that the fibrils usually consist of 2–6 protofilaments, each about 2–5 nm in diameter [12]. These protofilaments generally twist together to form fibrils that are typically 7–13 nm wide [11,12], or associate laterally to form long ribbons that are 2–5 nm high and up to 30 nm wide [13–15]. X-ray fibre diffraction data have shown that the protein or peptide molecules are arranged so that the polypeptide chain forms β-strands that run perpendicular to the long axis of the fibril [11].

Solid-state nuclear magnetic resonance (ss-NMR), X-ray micro- or nano-crystallography, and other techniques such as systematic protein engineering coupled with site-directed spin-labelling or fluorescence-labelling have transformed our ability to gain insight into the structures of fibrillar aggregates with residue-specific detail [16–29]. These advances have allowed us to go beyond the generic notions of the fibrillar appearance and presence of a cross-β structure. These studies have indeed allowed the identification of regions of the sequence that form and stabilise the cross-β core of the fibrils, as opposed to those stretches that are flexible and exposed to the solvent. In many cases, the arrangement of the various molecules in the fibrils has also been determined, clarifying the nature of the intermolecular contacts and the structural stacking of the molecules along the fibril axis. One frequent characteristic emerging from these studies, particularly for fibrils formed by long sequences, is the parallel in-register arrangements (PIRA) of β-strands in the fibril core [17–21,23–26,28], but antiparallel arrangements are also possible, especially for shorter strands [27,30].

At the same time, mutational studies of the amyloid aggregation kinetics revealed simple correlations between physico–chemical properties (charge, hydrophobicity, and β-sheet propensity) and aggregation propensities [31]. This allowed the development of different methods, which successfully predict aggregation-prone regions in the amino-acid sequence of a full-length protein [32–37]. All such approaches focus on predicting the intrinsic β-aggregation propensity of a sequence stretch using only the amino-acid sequence as an input. In [35] the possible parallel/antiparallel arrangement of the sequence stretch with itself was also taken into account. Molecular dynamics simulations of sequence fragments mounted on idealized β-strand templates, either parallel or antiparallel, were used to identify the most amyloidogenic fragments in a specific case [38]. A template amyloid structure based on PIRA is also employed in a very recent method for identifying fibril-forming segments [39]. A yet-unanswered question is why PIRA is found to be the most frequent arrangement of β-strands in the fibril core.

Here we introduce a computational approach by editing a pairwise energy function based on the propensities of two residues to be found within a β-sheet facing one another on neighbouring strands, as determined from a dataset of globular proteins of known native structures. We extract two different propensity sets depending on the orientation (parallel or antiparallel) of the neighbouring strands. Our method associates energy scores to specific β-pairings of two sequence stretches of the same length, and further assumes that distinct protein molecules involved in fibril formation will adopt the minimum-energy β-pairings in order to better stabilise the cross-β core.

A novel feature of our method is the ability to predict the registry of the intermolecular hydrogen bonds formed between amyloidogenic sequence stretches. In this way we can rationalise the observed tendency of proteins to assemble into parallel β-sheets in which the individual strands are in-register, contributing to form stackings of the same residue type along the fibril axis. Our algorithm is also able to correctly discriminate the orientation between intermolecular β-strands, either parallel or antiparallel. As a further demonstration of the robustness of the approach we will illustrate the ability of our algorithm to predict the portions of the sequence forming the cross-β core of the fibrils for a set of proteins, in excellent agreement with the experimentally determined amyloid structures, similar to previously proposed methods [32–37].

Our approach is based on the key assumption that a universal mechanism is responsible for β-sheet formation both in globular proteins and in fibrillar aggregates. The successful predictions obtained in this work suggest the validity of the above hypothesis in agreement with the unified framework presented previously [6].

Results

The Parallel In-Register Arrangement of β-Strands in the Amyloid-Like Fibrils

Based on the procedure described in detail in Materials and Methods and sketched in Figure 1, we can associate an energy score $\varepsilon_{i,j}^{p(a)}(L)$, from Equations 2 and 3, to the β-pairing of two sequence stretches chosen from distinct protein chains sharing an identical sequence. The pairing is specific since only pairs of residues facing each other in the corresponding register contribute to the energy score. All possible aggregation patterns are then defined in terms of the positions along the sequence *i,j,* the length *L,* and the relative orientation (either parallel or antiparallel) of the two sequence stretches participating in the pairing. We assume that the faithful

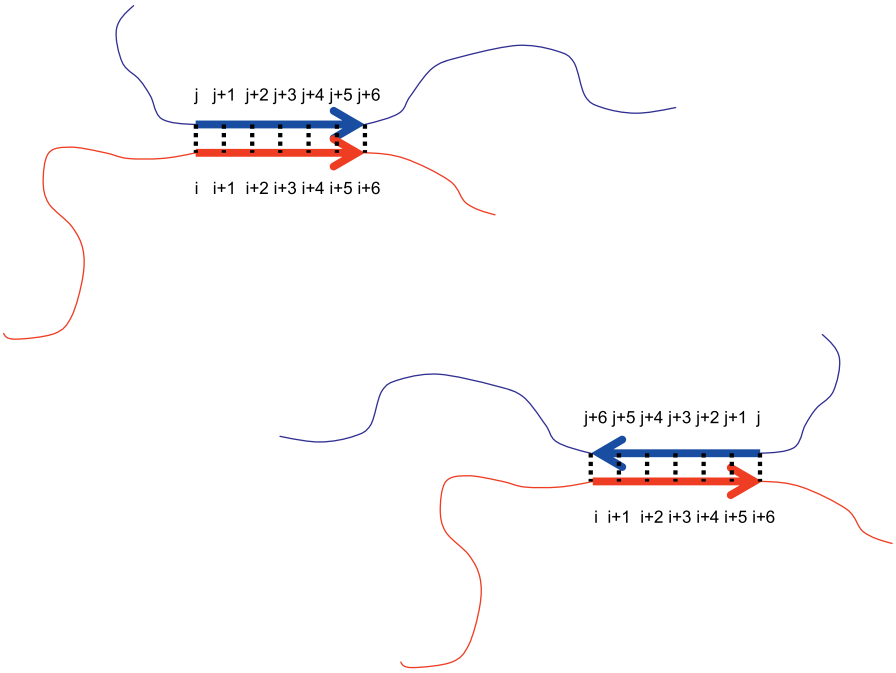


Figure 1. Sketch of the Method Presented in This Work

Two identical protein chains are assumed to associate by means of an ordered pairing of two hydrogen-bonded β-strands of the same length ($L = 7$) while the remaining parts of the chains remain unstructured. All possible pairings can be obtained by sliding the two strand-forming regions (i.e., by varying *i* and *j*) along the corresponding sequences and by varying their length *L* and their relative orientations. The two possible orientations, parallel and antiparallel, for the same choice of sequence stretches participating in the pairing, are depicted. The corresponding pairing aggregation scores are obtained (Equations 2 and 3) by summing contributions for each of the *L* pairwise interactions between residues in front of each other in the paired strands, represented as dotted lines. Dotted lines do not represent hydrogen bonds. Interaction matrices (Equation 1) are obtained from a statistical analysis of globular protein native structures, separately for parallel and antiparallel orientation. A term taking into account the entropy loss of the residues being ordered due to the pairing is further added.

doi:10.1371/journal.pcbi.0020170.g001

repetition of this aggregating unit is at the basis of the assembly of polypeptide chains into amyloid fibrils, determining the highly regular cross-β core of the fibril.

We first analyse the properties of our energy function at the level of single pair energies $E_{ab}^{p(a)}$ (see Equation 1). Residue pairs that appear from the analysis to possess low values of E_{ab}^{p} or E_{ab}^{a} should then have a propensity to aggregate in the context of amyloid fibrils higher than other pairs. Figure 2 shows the distribution of the 210 entries for E_{ab}^{p}, E_{ab}^{a}, and for the 20 in-register entries E_{aa}^{p}. All entries for both parallel and antiparallel pairing are shown in Table 1. Antiparallel pairing is favoured, on average, but the most favourable entries are found in the left tail of the parallel pairing distribution (with the only exception of the CYS–CYS antiparallel entry). Moreover, many of those are achieved for in-register pairings, notably for the hydrophobic residues VAL, ILE, and PHE. On the contrary, E_{aa}^{p} energies for charged and for some of the polar residues can assume significantly higher values. The highest E_{aa}^{p} energy is obtained for PRO, as expected, since it breaks the regular pattern of main backbone hydrogen bonding.

To verify whether the energies obtained with Equation 1 promote a general pattern in the aggregation, we use the sequence of the human amyloid β-peptide ($A\beta_{1-40}$), a peptide known to be involved in Alzheimer disease and other pathological conditions such as hereditary cerebral hemmorrhage with amyloidosis and inclusion-body myositis [2]. We

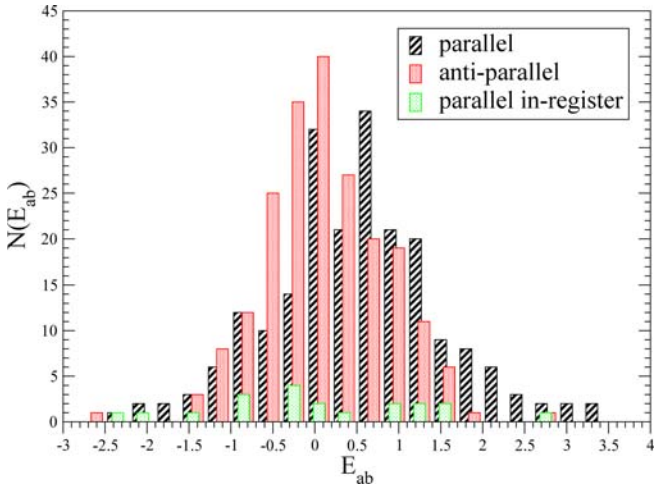


Figure 2. Histograms of the Energies for the Occurrence of Parallel and Antiparallel β-Pairing

The third histogram shows the energies for the PIRA (a subset of the parallel case). The lowest energies correspond to the antiparallel arrangement of CYS–CYS and to the PIRA of VAL–VAL and ILE–ILE. Seventeen out of the 44 CYS–CYS residues found in native structures in anti-parallel β-pairing are forming disulfide bridges with each other, in agreement with previous reports [57,58]. Note that the energy for parallel arrangement of CYS–CYS is repulsive.

doi:10.1371/journal.pcbi.0020170.g002

Table 1. Entries for Both Parallel, E^{p}_{ab}, and Anti-Parallel, E^{a}_{ab}, Pairings, Computed as in Equation 1 (See Materials and Methods)

Residue	CYS	PHE	LEU	TRP	VAL	ILE	MET	HIS	TYR	ALA	GLY	PRO	ASN	THR	SER	ARG	GLN	ASP	LYS	GLU
CYS	−2.57, −0.12	−1.08	−0.84	−0.94	−0.94	−0.65	−0.21	−0.51	−1.24	−0.09	−0.23	0.43	0.80	−0.07	0.06	0.34	0.26	0.87	−0.30	0.57
PHE	−0.76	−1.31, −1.48	−0.65	−0.66	−1.18	−1.15	−0.81	−0.15	−0.98	−0.30	−0.17	0.41	0.69	−0.25	−0.07	−0.01	0.02	0.96	0.19	0.01
LEU	−0.09	−0.91	−0.55, −0.98	−0.50	−0.88	−0.83	−0.35	0.26	−0.64	0.12	0.43	0.97	0.87	0.11	0.14	0.35	0.30	1.11	0.34	0.70
TRP	−0.69	−0.01	−0.16	−0.58, −0.82	−0.72	−0.78	−0.26	−0.55	−0.97	−0.13	0.16	0.36	−0.02	0.30	−0.27	−0.47	−0.42	0.67	−0.70	0.16
VAL	−1.16	−1.49	−1.67	−0.66	−1.40, −2.23	−1.42	−0.65	−0.41	−1.09	−0.44	−0.16	0.91	0.56	−0.59	−0.10	−0.32	−0.27	0.68	−0.36	−0.11
ILE	−1.24	−1.48	−1.70	−1.03	−2.03	−1.20, −2.18	−0.62	−0.31	−1.00	−0.52	0.14	1.20	0.76	−0.22	−0.01	−0.19	−0.19	0.94	0.22	0.17
MET	−0.41	−0.79	−0.82	−0.33	−1.16	−0.69	−0.65, −0.28	−0.09	−0.41	0.25	0.58	0.52	0.47	−0.26	0.06	0.22	0.68	1.32	0.05	0.27
HIS	1.10	0.01	−0.10	1.23	−0.46	−0.60	−0.07	−0.73, −0.74	−0.38	0.36	0.03	1.24	0.39	−0.33	0.17	0.13	−0.15	0.31	0.09	0.17
TYR	−0.38	−0.95	−0.60	−0.19	−1.10	−1.09	−0.84	−0.30	−0.98, −0.38	−0.25	0.09	0.07	0.02	−0.60	−0.35	−0.57	−0.36	0.56	−0.53	−0.21
ALA	0.05	0.14	−0.41	0.12	−0.82	−0.74	0.64	0.87	−0.04	0.48, 0.43	0.65	1.57	1.46	0.16	0.89	0.56	0.96	1.75	0.88	1.09
GLY	1.03	−0.19	0.36	0.69	−0.13	0.13	0.44	0.69	0.24	0.54	0.65, 1.40	1.61	1.16	0.53	0.62	0.85	0.83	1.13	1.63	1.51
PRO	2.18	1.88	2.61	1.97	1.27	1.25	2.22	2.39	1.45	1.74	1.82	2.68, 2.79	1.18	0.69	1.37	1.06	0.92	3.55	1.04	1.51
ASN	2.14	0.50	1.29	0.61	0.44	0.80	0.23	0.53	0.28	1.32	1.29	3.03	0.91, −0.35	0.08	0.02	0.42	0.07	1.27	0.64	0.91
THR	0.67	0.10	−0.14	0.08	−0.70	−0.42	−0.09	−0.50	0.54	0.31	0.69	1.34	−0.30	−1.03, 0.12	−0.37	−0.33	−0.30	0.30	−0.41	−0.23
SER	0.33	−0.04	0.68	0.61	0.06	0.38	0.02	−0.08	0.18	0.67	1.22	1.94	0.85	−0.09	−0.24, 0.11	−0.01	−0.03	0.81	0.07	0.28
ARG	−0.03	1.27	0.85	1.88	0.00	0.06	−0.09	0.15	0.67	0.99	0.89	2.50	0.89	−0.18	0.65	0.37, 1.32	−0.04	0.08	0.30	−0.24
GLN	1.54	0.27	0.33	0.15	0.31	0.03	0.62	0.33	0.17	0.80	0.70	1.80	0.87	0.53	1.14	0.59	0.20, 0.93	0.71	0.30	0.22
ASP	0.98	1.58	0.99	2.18	0.37	0.81	0.88	0.84	0.37	1.35	1.45	3.31	1.01	0.63	0.42	1.19	1.12	1.11, 1.53	0.10	1.37
LYS	2.30	1.29	0.84	0.69	0.53	0.75	1.35	0.22	0.49	1.08	1.43	3.20	1.03	−0.02	1.13	1.77	0.75	0.66	0.11, 1.02	−0.53
GLU	0.20	0.72	0.57	2.08	0.50	0.00	0.70	0.53	0.60	1.30	1.58	3.28	2.17	0.73	0.97	0.70	1.08	1.46	0.11	0.60, 1.51

Antiparallel pairings are shown in the upper above-diagonal half of the matrix. Parallel pairings are shown in the lower below-diagonal half of the matrix. For entries in the diagonal corresponding to two equal residue kinds, the antiparallel pairing is shown in the top line, whereas the parallel pairing is shown in the bottom line.
doi:10.1371/journal.pcbi.0020170.t001

are interested in rationalising on general grounds the competition between different registers in achieving the most favourable pairing. To average out as much as possible the influence of sequence specificity, we need to find a set of different minimum energy pairings. For fixed L and $|i-j|$, we slide the β-pairing segments along the sequence looking for the minimum energy pairing in both the parallel and antiparallel orientations (for the analysis shown in Figure 3 we consider the length independent energy term $\varepsilon^{p(a)}_{i,j}(L) + L\Delta s$). The minimum energies collected in this way are then averaged over different segment lengths ($4 \leq L \leq 23$) for a fixed value of $|i-j|$, yielding a mean value that is plotted as a function of $|i-j|$ in Figure 3. As a matter of fact, the in-register parallel alignment ($|i-j|=0$) is considerably more favourable than any other out-of-register parallel alignment ($|i-j| \neq 0$). We interpret oscillations in the curve for parallel pairings as a signature of some degree of pattern repetition in the sequence. On the other hand, ($|i-j|=0$) is the preferred pairing also for antiparallel orientation, but in this case the average minimum energy exhibits a linear increase with $|i-j|$. All these features are consistently retrieved in all sequences analysed in this work (unpublished data), whereas the existence and the values of the "gap" between the $|i-j|=0$ parallel and antiparallel depends crucially on the specific sequence (see Table 2).

Our results show that on average the assembly of $A\beta_{1-40}$ molecules with PIRA of sequence segments is favoured over both antiparallel and parallel out-of-register arrangements. ss-NMR and site-directed spin labelling experiments indeed show that amyloid fibrils from Aβ contain such a parallel in-register stacking of β-strands contributed by distinct molecules [17,18]. Similar results are obtained when computing the sequences of amylin, α-synuclein, and the PHF43 segment of tau protein (unpublished data), again in agreement with the experimental results [19–21,23]. For the $A\beta_{1-40}$ peptide and for the islet amyloid polypeptide, PIRA is clearly

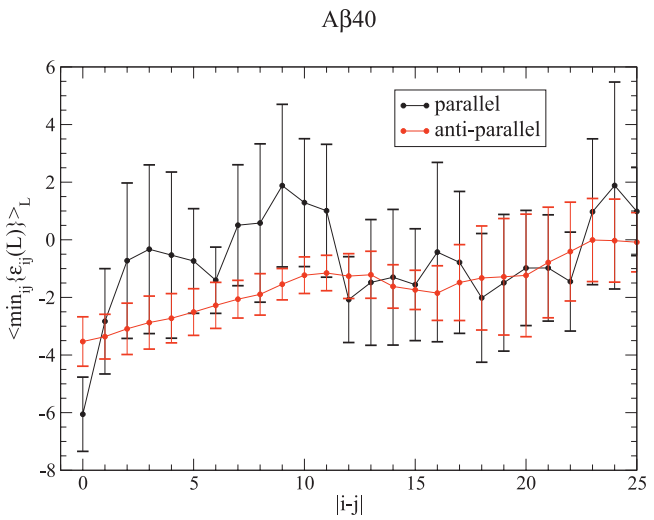


Figure 3. Plot of the Average over L of $\min_{ij} \varepsilon_{ij}(L)$ as a Function of $|i-j|$, Obtained with the *A*β40 Peptide for Both Parallel and Antiparallel Orientation

Bars represent the standard deviations of the minimum energies obtained for different segment lengths L. The linear increase with $|i-j|$ of the antiparallel curve can be explained in the following way. If $|i-j| = l$, with $l \leq L$, $[(L-l)/2]$ terms are repeated twice in the last sum of the right hand side of Equation 2 ([*x*] is the integer part of *x*) so that the number of E^a_{ab} values to be searched for low values is $[(L+l+1)/2]$. Since the smaller this number the easier to find a good pairing, antiparallel pairing is more and more favoured as $l \leq L$ is more and more decreased until for $l = 0$ one gets the most favourable antiparallel pairing.

doi:10.1371/journal.pcbi.0020170.g003

preferred over the antiparallel one within this analysis (Table 2). On the other hand, the preference is milder for the PHF43 fragment of the tau protein, and for human α-synuclein, being within the standard deviation of the energies employed for the average, as shown in Table 2.

The behaviour of the two curves shown in Figure 3 can be understood on the basis of simple statistical considerations. The problem consists in finding several low-energy pairings in a row. For a generic out-of-register parallel arrangement, the lowest E^p_{ab} values need to be found within all 210 possible entries. Therefore, the probability of finding several consecutive low-energy pairings is indeed quite low, independently of the sequence distance $|i-j|$ between the segments (as long as $|i-j| \neq 0$). On the other hand, the search problem is much easier in the case of in-register parallel pairing ($|i-j| = 0$),

Table 2. Energy Difference between Average Parallel and Antiparallel In-Register ($|i-j| = 0$) Pairings (See Figure 3)

Sequence	Gap between Average Minimum Energy of Parallel and Antiparallel Pairing at $\|i-j\| = 0$
$A\beta_{1-40}$	2.5 ± 1.6
Islet amyloid polypeptide	5.0 ± 2.6
PHF43 fragment	1.4 ± 1.5
α-synuclein	1.8 ± 1.8

doi:10.1371/journal.pcbi.0020170.t002

Table 3. Pairing Energies Predicted by Equations 2 and 3 for the Listed Peptides, Assuming the Full Peptide Length Is Involved in a β-Pairing with Itself

Peptide	Parallel Arrangement	Antiparallel Arrangement
GNNQQNY	**3.25**	3.73
KFFEAAAKKFFE	3.82	**−2.23**
KLVFFAE ($A\beta_{16-22}$)	−1.82	**−3.08**

Boldface, the minimum energy pairing among the two possible orientations, parallel or antiparallel.

doi:10.1371/journal.pcbi.0020170.t003

since the lowest pairing energies need to be found only within the 20 E^p_{aa} entries (see Figure 2). Therefore PIRA is favoured, with respect to other parallel alignments, because many of the most favourable entries can be found more easily.

In the case of antiparallel arrangement, the search always has to be performed among 210 entries, but a symmetry effect favours the $|i-j| = 0$ register. Indeed, when two overlapping sequence segments are aligned in antiparallel manner, some pairings are repeated twice (see the antiparallel case in Figure 1 with $j = i$). The number of low-energy pairings to be found is thus effectively reduced. The extent of this reduction is proportional to the length of the overlapping portion, thus explaining the linear increase with $|i-j|$ of the antiparallel curve in Figure 3. (Further details can be found in the Figure 3 legend.)

We remark that the above general arguments rely on the fact that the most favourable entries do indeed correspond to PIRAs, due to the stacking of hydrophobic and hydrophilic residues. In other words, PIRA provides a natural way of maximizing the number of favourable stacking interactions, lining up hydrophobic and hydrophilic residues in long rows along the fibril axis. Any other out-of-register parallel arrangement will most likely disrupt such an ordered pattern of stabilizing interactions.

Prediction of Alignment Orientation for Fibril-Forming Peptides

We employ prediction of amyloid structure aggregation (PASTA) to predict the orientation between β-strands in fibrillar structures formed by short, previously investigated peptides. In all cases we assume the full peptide length is involved in the β-core of the fibril, so that we simply compare the energy score of the parallel and antiparallel β-pairings of the full segment with itself. Results are shown in Table 3, showing in the three considered cases that PASTA correctly identifies the experimentally determined orientation as the minimum energy pairing. To our knowledge, the first two peptides are the only cases of a detailed atomic resolution achieved for a fibrillar structure obtained by means of X-ray diffraction from microcrystals. GNNQQNY is a fragment from the yeast prion protein Sup35 displaying a parallel orientation between β-strands within the same β-sheet [28]. KFFEAAAKKFFE is a peptide explicitly designed to form amyloid-like fibrils and was shown to be composed of antiparallel β-sheets [27]. KLVFFAE is the (16–22) fragment of the human $A\beta_{1-40}$ amyloid peptide, whose β-sheet structure was indicated to be antiparallel by ss-NMR data [40]. In the latter case it is remarkable that PASTA recognises

Table 4. Best Pairing Energies Predicted by Equations 2 and 3

Sequence	Best Pairing			Second-Best Pairing			Third-Best Pairing		
	First Segment	Second Segment	Energy	First Segment	Second Segment	Energy	First Segment	Second Segment	Energy
$A\beta_{1-40}$	12–20	12–20	−6.12	31–40	31–40	−6.11	12–21	12–21	−5.49
Islet amyloid polypeptide	12–32	12–32	−7.62	14–32	14–32	−7.47	15–32	15–32	−7.33
PHF43 fragment	11–15	11–15	−5.08	11–14	11–14	−4.91	10–15	10–15	−3.87
α-synuclein	48–55	48–55	−6.11	48–55	70–77	−5.82	48–56	48–56	−5.49
HET-s prion domain	22–28	22–28	−4.29	47–51	47–51	−4.07	22–29	22–29	−3.97

All listed arrangements are PIRAs. Only the second-best pairing for α-synuclein is out-of-register. In the case of the PHF43 fragment of the tau protein, the third-best pairing involving the segment 10–15 with itself is degenerate with the pairing involving the segment 11–16 with itself (again with parallel orientation).
doi:10.1371/journal.pcbi.0020170.t004

the tendency of the short (16–22) fragment to form antiparallel β-sheets while at the same time predicting the correct in-register parallel alignment for the full sequence (see below).

Prediction of Specific Pairings and Sequence-Aggregation Propensities

We employ PASTA to identify the regions of the sequence-promoting aggregation for five natively unfolded systems. These include human $A\beta_{1-40}$, human α-synuclein, the human islet amyloid polypeptide, the PHF43 fragment from human tau, and the HET-s prion domain protein from *P. anserina*. We decided to perform the analysis on such systems rather than on globular proteins because our analysis utilises values of intrinsic propensity to aggregate residue pairs and does not take into account the presence and type of secondary and tertiary structure in the analysed polypeptide chain. Indeed, it is well-known that the presence of structure in the initial nonaggregated state of the protein is an important determinant of aggregation and reduces dramatically the aggregation propensity of the structured regions [41]. In addition, the five natively unfolded systems analysed here were chosen because their aggregation-promoting regions were also determined experimentally, allowing our predictions to be directly tested.

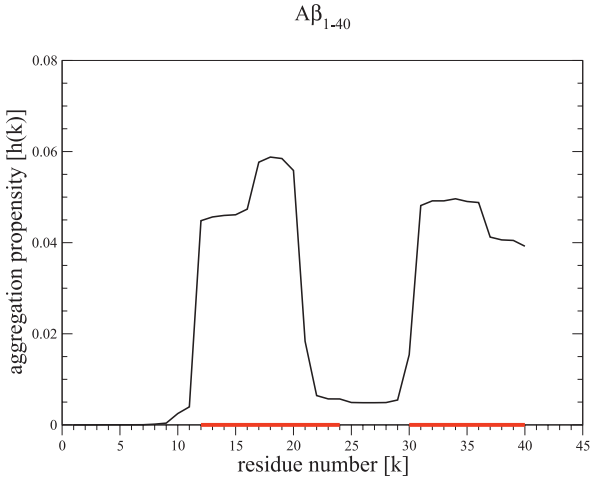
The energy functions introduced in Equations 2 and 3 can be used to compare different segment lengths, and we will first list the three pairings yielding the minimum energy when looking among all possible segment lengths. (By definition the energy of a nonaggregating system is zero.) The results are summarized in Table 4. We then use the single-residue propensity $h(k)$ defined in Equation 5 to take into account other low-energy pairings that could be close competitors of the lowest-energy pairing.

Human amyloid β-peptide. We first apply PASTA to study $A\beta_{1-40}$. It is known by proline-scanning mutagenesis and quantitation of fibrils by Congo red binding [42], ThT binding, electron microscopy, and SDS-Page [43], ss-NMR (17) and site-directed spin labelling [18] that the regions of the sequence involved in β-aggregation are approximately the segments 12–24 and 30–40 (the boundaries of the two regions vary somewhat in the various reports). Both segments are almost exactly predicted and are found as minima closely competing with each other. In Figure 4A we are plotting $h(k)$ for $A\beta_{1-40}$. We see that in the region 12–20 and 31–40 the propensity is very strong, in almost perfect agreement with the experimental prediction, whereas it is negligible in the other parts of the protein. In both cases PIRA is predicted in perfect agreement with experimental data [17].

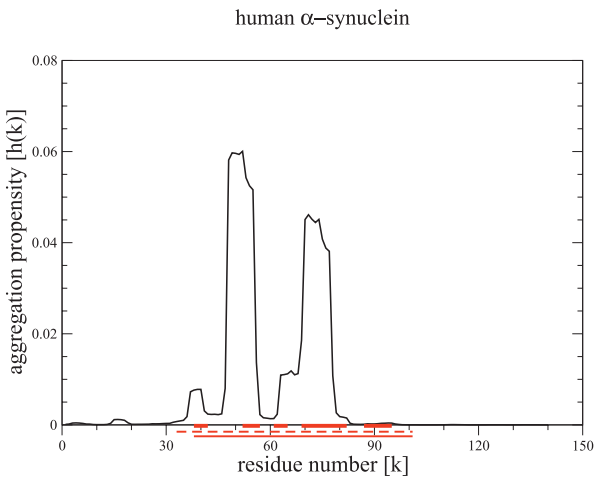
Human α-synuclein. This protein is involved in Parkinson disease and in dementia with Lewy Bodies [2]. By synthesising peptides of various lengths and quantifying their aggregation using HPLC and circular dichroism, the region 63–78 has been proposed to be involved in aggregation [44,45]. More recent experimental studies employing ss-NMR have allowed the identification of several sequence portions involved in β-strand formation within the fibrils [23]. These are shown as thick red bars in Figure 4B, together with the aggregation profile predicted by our algorithm. Four out of five of the experimentally determined sequence stretches are correctly identified by PASTA. The overall arrangement is parallel in-

Figure 4. Amyloid Propensity Plots for the Proteins Studied in This Work

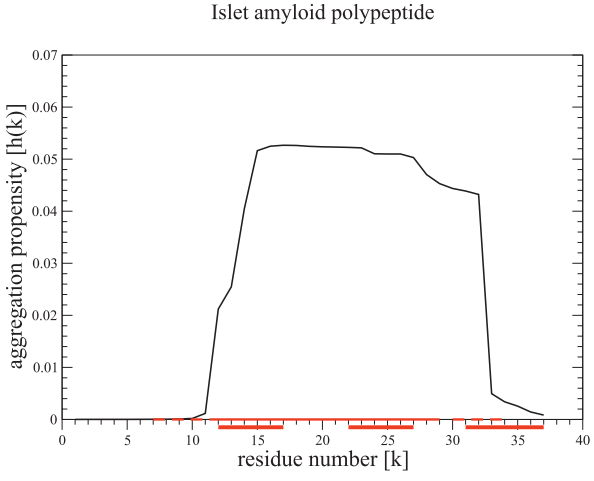
(A) Plot of amyloid propensity $h(k)$ (Equation 5) for the human amyloid β-peptide. The sequence regions involved in β-strands according to ss-NMR experiments [17] are represented by a thick red line along the k-axis.
(B) Same as in (A) but for the protein human α-synuclein. Thick red bars mark sequence stretches involved in β-strands according to ss-NMR experiments [23]. The thin red bars show the whole sequence portion found to be in PIRA, according to site-directed spin-labelling, solid line [19], and found to participate in main backbone hydrogen bonding according to hydrogen–deuterium exchange, dashed line [22]. The two experimentally determined portions differ only in the location of the initial boundary.
(C) Same as in (A) but for the subsection islet amyloid polypeptide. The thin red line shows the whole sequence portion found to be in PIRA according to site-directed spin-labelling experiments, with the dashed portions representing the uncertainty on boundary location [20]. Thick red bars show the sequence portions proposed to participate in β-strands according to a structural model based on a serpentine PIRA [24].
(D) Same as in (A) but for the PHF43 fragment from the fetal form of human tau. The thick red line shows a local sequence motif identified to be crucial for β-aggregation [46].
(E) Same as in (A) but for the HET-s prion domain protein from *P. Anserina*. The red bars show sequence portions involved in β-strands as determined by fluorescence studies, quenched hydrogen exchange NMR, and ss-NMR (29).
doi:10.1371/journal.pcbi.0020170.g004


Aβ1-40
aggregation propensity [h(k)]
residue number [k]

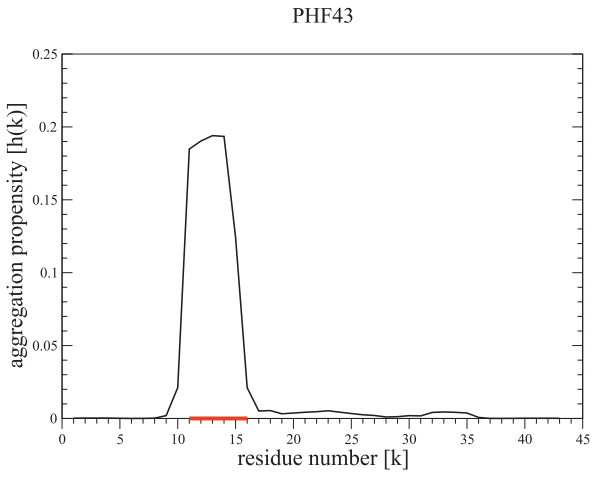
(A)


human α–synuclein
aggregation propensity [h(k)]
residue number [k]

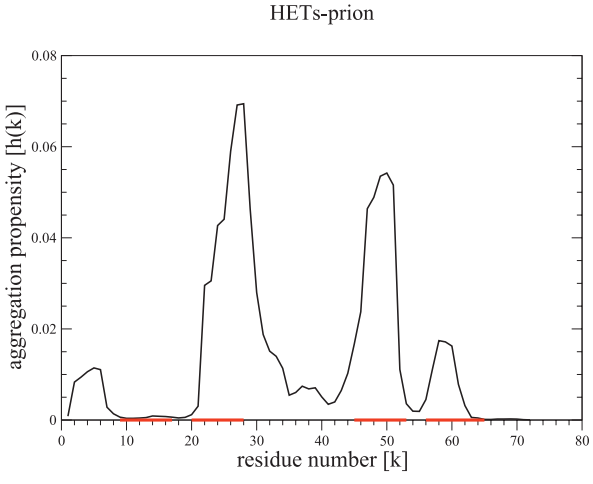
(B)


Islet amyloid polypeptide
aggregation propensity [h(k)]
residue number [k]

(C)


PHF43
aggregation propensity [h(k)]
residue number [k]

(D)


HETs-prion
aggregation propensity [h(k)]
residue number [k]

(E)

register, as determined by site-directed spin-labelling studies [19]. PASTA correctly finds the best minimum for a parallel in-register pairing, but the second-best pairing is a parallel out-of-register one. Looking at the segments involved, which are VVHGVATV (48–55) and VVTGVTAV (70–77), we realize that this is due to a strong pattern repetition. Five out of eight residues are matched for an in-register alignment, including the four valines that are most responsible for the low pairing energy. In Figure 5 we show the β-pairing contact map $h_2(k,m)$, where a compendium of the general features predicted by PASTA can be found. The strongest signal is for PIRA, but parallel out-of-register arrangement is also selected in the presence of repetition of sequence patterns along the chain. Weak signals are also present for antiparallel arrangement, which would take place between identical sequence stretches, as predicted on general grounds.

Islet amyloid polypeptide. The 37-residue islet amyloid polypeptide is the major component of pancreatic amyloid deposits, which are the hallmark of noninsulin—dependent (type II) diabetes mellitus. We plot $h(k)$ in Figure 4C. Again there is quite a good agreement with site-directed spin-label experiments (20), which show parallel in-register aggregation in the region 12–29. It should be remarked that in this case, unlike for $A\beta_{1-40}$, PASTA clearly signals the existence of a single continuous pairing. In a recently proposed model, resulting from a number of experimental constraints, residues 12–17, 22–27, and 31–37 are proposed to form β-strands in a serpentine arrangement in each molecule, with very short loops connecting them [24]. This structural arrangement is repeated for each peptide molecule along the fibril axis so that the parallel in-register orientation is maintained [24]. The short length of the loop may make it difficult to distinguish between a single continuous pairing and three very-nearby short pairings.

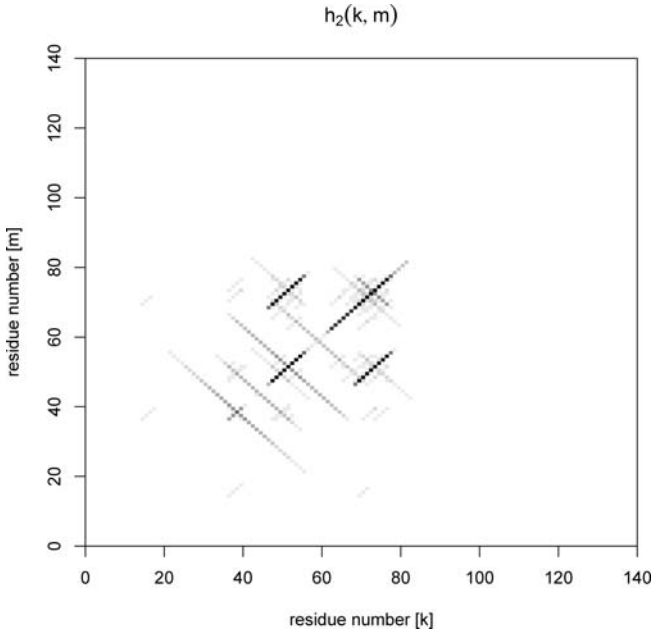


Figure 5. β-Pairing Contact Map (Equation 6) for Human α-Synuclein
This picture was obtained with $\lambda = 1.5$, for a better visualization of the competition between the best pairings.
doi:10.1371/journal.pcbi.0020170.g005

PHF43 fragment from the fetal form of human tau. Filamentous inclusions from tau proteins are present in numerous neurodegenerative diseases, including Alzheimer disease and frontotemporal dementia with Parkinsonism linked to Chromosome 17 [2]. The region, found experimentally to be involved in aggregation within the tau fragment PHF43, is the segment 11–16, as identified by means of spot membrane–binding assay [46]. A good agreement is again found between these experimental data and those found with our prediction, as shown by both the minimum energy pairings listed in Table 4 and the plot of $h(k)$ in Figure 4D. The arrangement is also correctly predicted to be parallel in-register, as determined by site-directed spin-labelling coupled with EPR methods [21].

HET-s prion domain fragment from *P. anserina*. The prion form of the protein HET-s is involved in a programmed cell death mechanism called heterokaryon incompatibility [47,48]. The recombinant HET-s prion domain (fragment 218–289) can form amyloid-like fibrils in vitro and induce prion phenotypes in a host cell [49]. Recent experiments employing fluorescence studies, quenched hydrogen exchange NMR, and ss-NMR [29] determined four sequence portions involved in β-strand structure within the fibrils, shown as red bars in Figure 4E, together with the aggregation profile predicted by our algorithm. PASTA correctly predicts four sequence stretches to be involved in β-aggregation, placing three of them in good agreement with experiments. The peculiar arrangement suggested by Ritter et al. on the basis of their experimental data is parallel but not in-register, pairing different portions of the same chain [29]. The method described in this work is based on the assumption of interchain pairing. Further studies are being carried out to extend our algorithm to intrachain pairing as well.

Discussion

We introduced a pairwise energy function based on the propensities of two residues to be found within a β-sheet facing one another on neighbouring strands, as determined from a dataset of globular proteins of known native structures. Such energy function was incorporated within an algorithm able to predict amyloidogenic sequence stretches, as well as the registry of the intermolecular hydrogen bonds formed between them. The latter type of prediction is a novel feature of our approach.

For a set of natively unfolded proteins involved in the formation of amyloid fibrils, we correctly predict their observed tendency to assemble into parallel β-sheets in which the individual strands are in-register. Our algorithm is also able to correctly determine the orientation between β-strands in the fibrils, either parallel or antiparallel, as shown by a comparison with fibrillar structures formed by short peptides determined experimentally at the atomic level.

Our energy function predicts that PIRA is favoured on general grounds, with respect to other parallel out-of-register alignments, because the most favourable β-pairing found in globular proteins is indeed parallel and obtained for hydrophobic pairs sharing the same residue kind. Even though such parallel in-register pairing can be unfavourable for other residues (especially charged ones), PIRA by itself constrains the search for good pairs in a much smaller set than for out-

of-register arrangement. A similar, yet milder, effect induced by pairing statistics is detected for antiparallel arrangement, favouring the case in which the latter is achieved between *identical* sequence stretches. Parallel arrangement is generally favoured over antiparallel, but in some cases sequence specificity can override this tendency, as in the case of short peptides. Out-of-register parallel arrangement is also predicted as a good competitor in the presence of repeated (periodic) patterns in the sequence, which actually occur in several prion proteins, both in mammals and in fungi.

Our algorithm was also used to predict the portions of the sequence, for an initially unstructured polypeptide chain, that form the cross-β core of the fibrils. A good agreement with the experimental information available on amyloid structures, similar to other proposed methods [32–37], was found for human $A\beta_{1-40}$, α-synuclein, islet amyloid polypeptide, a fragment from human tau, and the prion domain of HET-s from *P. anserina*.

The results obtained in this work, besides rationalising on general grounds the common occurrence of PIRA in amyloid fibrillar structures, suggest two important conclusions. First, the existence of a preferred β-pairing is an important determinant of the self-propagating nature of amyloid fibrils and of the difficulty of these to seed the fibrillar state in proteins that have even subtle differences in sequence, a phenomenon associated with the species barrier in prion transmissibility. Moreover, the polymorphism often observed for amyloid fibrils [15,50], leading to the existence of different prion strains [10], might be explained by the competition between different low-energy β-pairings that are realizable for the same sequence.

The notion of a preferred β-pairing is the simplest one that can be put forward to account for the self-complementation of protein molecules on a structural basis [51]. It can be seen as a way of reconciling the roles of side chains in driving specific aggregation and of main backbone interactions in determining the general tendency of polypeptide chains for fibril formation. The knowledge-based energy function introduced in this work describes how side chain–side chain interactions between residues facing each other modulate the main chain hydrogen bond energy common to all residues. Stacking of hydrophobic residues [27] or hydrogen bonding between side chain groups [28] will favour PIRA, whereas electrostatic repulsion between charges of the same type disfavours it. All such interactions are captured within our knowledge-based approach. A determinant of self-complementation that we neglect in our simple scheme is the steric interdigitation between different sheets forming the fibril core [39]. However, the good performance of our algorithm shows that sequence information is already relevant at the level of β-strand pairing within the same sheet.

As a second important conclusion, the fact that the whole computational approach is derived from the knowledge of globular proteins underscores the universality of the physico–chemical mechanisms underlying amyloid fibril formation. Moreover, it indicates that the structure and stabilising interactions existing in the apparently monotonous amyloid or amyloid-like fibrils are of the same essential nature as those determining structural and functional diversity in globular proteins.

Materials and Methods

Knowledge-based pair potential. We derive an energy function for specific β-aggregation using the top500H database [52]. It is a nonredundant specially refined set of 500 high resolution X-ray crystallographic structures of globular proteins, where hydrogen atoms were also reconstructed. These proteins include all-α, all-β, α/β, and α + β proteins, and their structures are deposited in the Protein Data Bank. All occurring instances, n_{ab}, of a given *ab* residue pair are partitioned ($n_{ab} = n^c_{ab} + n^p_{ab} + n^a_{ab} + n^d_{ab}$) into four different classes according to whether the two residues are facing each other on neighbouring parallel β-strands (n^p_{ab}) or on neighbouring antiparallel β-strands (n^a_{ab}), and whether the distance between their C^α atoms is less than 6.5 Å—without participating in a ordered β-geometry (generic bulk contacts n^c_{ab})—or more than 6.5 Å (noncontacting disordered pairs n^d_{ab}). All pairs are included in the count, except those formed by consecutive residues along the protein chain. The participation to either parallel or antiparallel β-bridges is assessed by using the DSSP algorithm [53], but with a slightly stricter electrostatic energy threshold of −1 Kcal/mol to assign hydrogen bonds. (The distribution of such energies obtained from the Richardson set peaks around the value of −2.4 Kcal/mol, but increases again for values higher than −1 Kcal/mol, unpublished data).

Energies can be assigned to the occurrence of parallel β-pairing and antiparallel β-pairing for two amino acids of type *a* and type *b*, by assuming that the database of protein native structures is a system in thermodynamic equilibrium at a single temperature, assumed to be roughly constant for all the proteins in the database [54]. Upon further assumption that correlations between different pairings can be neglected within single proteins in the database [55], the *propensity*, $p_{ab}(x)$, of the *ab* pair to be found in one of the four pairing types, *x*, is given by the Boltzmann factor, $p_{ab}(x) = exp(-E^x_{ab})$. The *E*'s are energy differences, measured in units of thermal energy, between the native and the reference state with respect to which propensities are computed [54].

Propensities are defined as the ratio of the observed frequency over the expected probability in the reference state, which is in turn estimated as the frequency observed over all pairs.

$$E^p_{ab} = -\log\left(\frac{\frac{n^p_{ab}}{\sum_{ab} n^p_{ab}}}{\frac{n_{ab}}{\sum_{ab} n_{ab}}}\right) \quad E^a_{ab} = -\log\left(\frac{\frac{n^a_{ab}}{\sum_{ab} n^a_{ab}}}{\frac{n_{ab}}{\sum_{ab} n_{ab}}}\right) \quad E^c_{ab} = -\log\left(\frac{\frac{n^c_{ab}}{\sum_{ab} n^c_{ab}}}{\frac{n_{ab}}{\sum_{ab} n_{ab}}}\right) \tag{1}$$

A similar expression yields the energy E^d_{ab}, which should be assigned to a noncontacting pair *ab*. Since the numbers n^p_{ab}, n^a_{ab}, and n^c_{ab} can be very small (or even zero in some special cases involving PRO and CYS), we used an averaging procedure to decrease statistical error [33]. Hence, for example, $E^p_{ab} = (E^{p+}_{ab} + E^{p-}_{ab})/2$, where E^{p+}_{ab}, E^{p-}_{ab}, are the energies obtained from Equation 1 when adding ($n^p_{ab} \to n^p_{ab} + 1$, $n_{ab} \to n_{ab} + 1$) or subtracting ($n^p_{ab} \to n^p_{ab} - 1, n_{ab} \to n_{ab} - 1$), a single event, to the observed number of cases (whenever $n^p_{ab} < 2$, 0.5 is used in place of $n^p_{ab} - 1$). Statistical potentials describing residue pair correlations within β-sheets were developed in the context of structure prediction, limiting the total ensemble of residue pairs to those in which both residues participate in a β-structure [56–59]. Our derivation instead places all residue pairs in the total ensemble.

β-pairing energy function. Our aim is to predict the specific aggregation pattern of a pair of identical proteins of *N* amino acids $\{a_k\}_{1 \le k \le N}$, as determined by the specific β-pairing (either parallel or antiparallel) of the sequence stretch of length *L*, beginning at position *i* on the first chain, with the sequence stretch of the same length, beginning at position *j* on the second chain. We assume throughout the rest of this work that only a single stretch per sequence participates in the β-pairing and that all other residues (from 1 to $i - 1$ and from $i + L$ to *N* for the first chain and from 1 to $j - 1$ and from $j + L$ to *N* for the second chain) are not involved in aggregation and are found in a disordered noncompact conformation. We assume further that the energies E^d_{ab} of all pairs involving these latter residues can be neglected, since $n^d_{ab} \approx n_{ab}$ and $E^d_{ab} \approx 0$. Remaining pairs whose residues are both present in the β-aggregating stretches but not specifically paired with each other are assumed to be noncontacting as well. We verified that the results we present in this work do not change upon inclusion of noncontacting pair terms. The overall

pairing aggregation energy for a given parallel/antiparallel pattern is then determined only by residue pairs mutually involved in the ordered β-pairing, and can be written, by assuming they do so independently of one another, as

$$\varepsilon^{p}_{i,j}(L) \equiv \sum_{k=0}^{L-1} E^{p}_{a^1_{i+k},a^2_{j+k}} - L\Delta s \tag{2}$$

$$\varepsilon^{a}_{i,j}(L) \equiv \sum_{k=0}^{L-1} E^{a}_{a^1_{i+k},a^2_{j+L-1-k}} - L\Delta s \tag{3}$$

where the overscripts 1 and 2 correspond to the first and second chain, respectively, and $\Delta S = L\Delta s$ is the entropy loss due to the β-ordering of the L residue pairs, with Δs corresponding to the average entropy loss per residue pair. Due to the many approximations involved in the standard derivation of statistical potentials, the latter extensive term might actually compensate for any bias introduced with the choice of the reference state, making its a priori evaluation too difficult. Therefore we set $\Delta s = -0.2$ throughout all our work on a purely empirical basis. The proper introduction of sequence specific Δs_{a_i} might certainly improve the quantitative agreement with experimental observations, but we chose to keep our energy-scoring function as simple as possible to directly test the relevance of β-pairing specificity in dictating aggregation patterns. Since the computation of energy scores $\varepsilon^{p}_{i,j}(L)$ and $\varepsilon^{a}_{i,j}(L)$ involves a summation over only L terms, it can be easily performed on a genome-wide scale.

Sequence-dependent aggregation propensities and contact maps. To take into account in a more complete manner all possible pairing energies close to the minimum, we introduce an "ordered β-pairing partition function":

$$Z = \sum_{i,j,L\geq 4} \left\{ \exp\left(-\lambda\varepsilon^{p}_{i,j}(L)\right) + \exp\left(-\lambda\varepsilon^{a}_{i,j}(L)\right) \right\} \tag{4}$$

where we set $\lambda = 2.0$ as an adimensional factor setting the energy scale. Parameters Δs and λ need not to be fine-tuned and can be changed within a 20% range without affecting the final results. The partition function (Equation 4) allows a better one-dimensional visualization of the results by defining a position-based "amyloid propensity"

$$h(k) = \frac{\sum_{i,j,L\geq 4} \frac{\delta_{i\leq k<i+L} + \delta_{j\leq k<j+L}}{2L} \left[\exp(-\lambda\varepsilon^{p}_{i,j}(L)) + \exp(-\lambda\varepsilon^{a}_{i,j}(L))\right]}{Z} \tag{5}$$

where $\delta_{i \leq k < i+L} = 1$ if residue k belongs to the L-stretch going from i to $i + L-1$ and $\delta_{i \leq k < i+L} = 0$ otherwise. Note that $h(k)$ is a probability since $\sum_k h(k) = 1$. It tells how a given residue is more likely to aggregate in an ordered β-structure with respect to others.

A more complete piece of information that can be extracted from the method is the normalized two-dimensional probability $h_2(k,m)$ of two given residues found paired to each other within an ordered β-structure. It is given by

$$h_2(k,m) = \frac{\sum_{i,j,L\geq 4} \frac{\delta_{i\leq k<i+L}\delta_{j\leq m<j+L}}{L} \left[\delta_{k-m+j-i}\exp(-\lambda\varepsilon^{p}_{i,j}(L)) + \delta_{k+m+1-L-j-i}\exp(-\lambda\varepsilon^{a}_{i,j}(L))\right]}{Z} \tag{6}$$

where k and m label residues in two different chains and $\delta_{k-m+j-i} = 1$ if $k - m + j - i = 0$, and 0 otherwise. Based on $h_2(k,m)$, a β-pairing contact map can be produced where the orientation (parallel or antiparallel to the diagonal) and the register of the best pairings is easily traced out (see Figure 5).

We name the full procedure described in this section PASTA.

Acknowledgments

We thank G. Colombo, S. Lise, N. Taddei, S. Tosatto, and M. Vendruscolo for stimulating discussion.

Author contributions. AT, FC, AM, and FS conceived and designed the experiments, performed the experiments, and analysed the data. AT, FC, and FS wrote the paper.

Funding. This work was supported by Programmi di Ricerca Scientifica di Rilevante Interesse Nazionale, grant 2003025755 in 2003 and grant 2005027330 in 2005.

Competing interests. The authors have declared that no competing interests exist.

References

1. Selkoe DJ (2003) Folding proteins in fatal ways. Nature 426: 900–904.
2. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 75: 333–366.
3. Westermark P, Benson MD, Buxbaum JN, Cohen AS, Frangione B, et al. (2005) Amyloid: Toward terminology clarification. Amyloid 12: 1–4.
4. Stefani M, Dobson CM (2003) Protein aggregation and aggregate toxicity: New insights into protein folding, misfolding diseases and biological evolution. J Mol Med 81: 678–699.
5. Uversky VN, Fink AL (2004) Conformational constraints for amyloid formation fibrillation: The importance of being unfolded. Biochim Biophys Acta 1698: 131–153.
6. Hoang TX, Marsella L, Trovato A, Seno F, Banavar JR, et al. (2006) Common attributes of native-state structures of proteins, disordered proteins and amyloid. Proc Natl Acad Sci U S A 103: 6883–6888.
7. Barnhart MM, Chapman MR, Robinson (2006) Curli biogenesis and function. Annu Rev Microbiol 60: 131–147.
8. Fowler DM, Koulov AV, Alory-Jost C, Marks MS, Balch WE, et al. (2006) Functional amyloid formation within mammalian tissue. PLoS Biol 4(1): 100–107.
9. Talbot NJ (2003) Aerial morphogenesis: Enter the chaplins. Curr Biol 13: R696–R698.
10. Chien P, Weissman JS, DePace AH (2004) Emerging principles of conformation-based prion inheritance. Annu Rev Biochem 73: 617–656.
11. Sunde M, Blake C (1997) The structure of amyloid fibrils by electron microscopy and X-ray diffraction. Adv Protein Chem 50: 123–159.
12. Serpell LC, Sunde M, Benson MD, Tennent GA, Pepys MB, et al. (2000) The protofilament substructure of amyloid fibrils. J Mol Biol 300: 1033–1039.
13. Bauer HH, Aebi U, Haner M, Hermann R, Muller M, et al. (1995) Architecture and polymorphism of fibrillar supramolecular assemblies produced by in vitro aggregation of human calcitonin. J Struct Biol 115: 1–15.
14. Saiki M, Honda S, Kawasaki K, Zhou D, Kaito A, et al. (2005) Higher-order molecular packing in amyloid-like constructed with linear arrangements of hydrophobic and hydrogen-bonding side-chains. J Mol Biol 348: 983–998.
15. Pedersen JS, Dikov D, Flink JL, Hjuler HA, Christiansen G, et al. (2006) The changing face of glucagon fibrillation: Structural polymorphism and conformational imprinting. J Mol Biol 355: 501–523.
16. Jaroniec CP, MacPhee CE, Astrof NS, Dobson CM, Griffin RG (2002) Molecular conformation of a peptide fragment of transthyretin in an amyloid fibril. Proc Natl Acad Sci U S A 99: 16748–16753.
17. Petkova AT, Ishii Y, Balbach JJ, Antzutkin ON, Leapman RD, et al. (2002) A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR. Proc Natl Acad Sci U S A 99: 16742–16747.
18. Torok M, Milton S, Kayed R, Wu P, McIntire T, et al. (2002) Structural and dynamic features of Alzheimer's Abeta peptide in amyloid fibrils studied by site-directed spin labeling. J Biol Chem 277: 40810–40815.
19. Der-Sarkissian A, Jao CC, Chen J, Langen R (2003) Structural organization of alpha-synuclein fibrils studied by site-directed spin labeling. J Biol Chem 278: 37530–37535.
20. Jayasinghe SA, Langen R (2004) Identifying structural features of fibrillar islet amyloid polypeptide using site-directed spin labeling. J Biol Chem 279: 48420–48425.
21. Margittai M, Langen R (2004) Template-assisted filament growth by parallel stacking of tau. Proc Natl Acad Sci U S A 101: 10278–10283.
22. Del Mar C, Greenbaum EA, Mayne L, Englander SW, Woods VL Jr (2005) Structure and properties of alpha-synuclein and other amyloids determined at the amino acid level. Proc Natl Acad Sci U S A 102: 15477–15482.
23. Heise H, Hoyer W, Becker S, Andronesi OC, Riedel D, et al. (2005) Molecular-level secondary structure, polymorphism, and dynamics of full-length alpha-synuclein fibrils studied by solid-state NMR. Proc Natl Acad Sci U S A 102: 15871–15876.
24. Kajava AV, Aebi U, Steven AC (2005) The parallel superpleated beta-structure as a model for amyloid fibrils of human amylin. J Mol Biol 348: 247–252.
25. Krishnan R, Lindquist SL (2005) Structural insights into a yeast prion illuminate nucleation and strain diversity. Nature 435: 765–772.
26. Lührs T, Ritter C, Adrian M, Riek-Loher D, Bohrmann B, et al. (2005) 3D structure of Alzheimer's amyloid-beta(1–42) fibrils. Proc Natl Acad Sci U S A 102: 17342–17347.
27. Makin OS, Atkins E, Sikorski P, Johansson J, Serpell LC (2005) Molecular basis for amyloid fibril formation and stability. Proc Natl Acad Sci U S A 102: 315–320.
28. Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, et al. (2005) Structure of the cross-beta spine of amyloid-like fibrils. Nature 435: 773–778.

29. Ritter C, Maddelein ML, Siemer AB, Luhrs T, Ernst M, et al. (2005) Correlation of structural elements and infectivity of the HET-s prion. Nature 435: 844–848.
30. Petkova AT, Buntkowsky G, Dyda F, Leapman RD, Yau WM, et al. (2004) Solid state NMR reveals a pH-dependent antiparallel beta-sheet registry in fibrils formed by a beta-amyloid peptide. J Mol Biol 335: 247–260.
31. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature 424: 805–808.
32. Yoon S, Welsh SJ (2004) Detecting hidden sequence propensity for amyloid fibril formation. Protein Sci 13: 2149–2160.
33. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 22: 1302–1306.
34. Pawar AP, DuBay KF, Zurdo J, Chiti F, Vendruscolo M, et al. (2005) Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. J Mol Biol 350: 379–392.
35. Tartaglia GG, Cavalli A, Pellarin R, Caflisch A (2005) Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences. Protein Sci 14: 2723–2734.
36. Galzitskaya OV, Garbuzynskiy SO, Lobanov MY (2006) Is it possible to predict amyloidogenic regions from sequence alone? J Bioinform Comp Biol 4: 373–388.
37. Galzitskaya OV, Garbuzynskiy SO, Lobanov MY (2006) Prediction of amyloidogenic and disordered regions in protein chains. PLoS Comput Biol 2(12): e177. doi:10.1371/journal.pcbi.0020177.eor
38. Khare SD, Wilcox KC, Gong P, Dokholyan NV (2005) Sequence and structural determinants of Cu, Zn superoxide dismutase aggregation. Proteins 61: 617–632.
39. Thompson MJ, Sievers SA, Karanicolas J, Ivanova MI, Baker D, et al. (2006) The 3D profile method for identifying fibril-forming segments of proteins. Proc Natl Acad Sci U S A 103: 4074–4078.
40. Balbach JJ, Ishii Y, Antzutkin ON, Leapman RD, Rizzo NW, et al. (2000) Amyloid fibril formation by $A\beta_{16-22}$, a seven-residue fragment of the Alzheimer's β-amyloid peptide, and structural characterization by solid state NMR. Biochemistry 39: 13748–13759.
41. Bemporad F, Calloni G, Campioni S, Plakoutsi G, Taddei N, et al. (2006) Sequence and structural determinants of amyloid fibril formation. Acc Chem Res 39: 620–627.
42. Wood SJ, Wetzel R, Martin JD, Hurle MR (1995) Prolines and amyloidogenicity in fragments of the Alzheimer's peptide beta/A4. Biochemistry 34: 724–730.
43. Tjernberg LO, Callaway DJ, Tjernberg A, Hahne S, Lilliehook C, et al. (1999) A molecular model of Alzheimer amyloid beta-peptide fibril formation. J Biol Chem 274: 12619–12625.
44. Bodles AM, Guthrie DJ, Harriott P, Campbell P, Irvine GB (2000) Toxicity of non-Abeta component of Alzheimer's disease amyloid, and N-terminal fragments thereof, correlates to formation of beta-sheet structure and fibrils. Eur J Biochem 267: 2186–2194.
45. Bodles AM, Guthrie DJ, Greer B, Irvine GB (2001) Identification of the region of non-A beta component (NAC) of Alzheimer's disease amyloid responsible for its aggregation and toxicity. J Neurochem 78: 384–395.
46. von Bergen M, Friedhoff P, Biernat J, Heberle J, Mandelkow EM, et al. (2000) Assembly of tau protein into Alzheimer paired helical filaments depends on a local sequence motif ((306)VQIVYK(311)) forming beta structure. Proc Natl Acad Sci U S A 97: 5129–5134.
47. 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 U S A 94: 9773–9778.
48. Saupe SJ (2000) Molecular genetics of heterokaryon incompatibility in filamentous ascomycetes. Microbiol Mol Biol Rev 64: 489–502.
49. Balguerie A, Dos Reis S, Ritter C, Chaignepain S, Coulary-Salin B, et al. (2003) Domain organization and structure-function relationship of the HET-s prion protein of Podospora anserina. EMBO J 22: 2071–2081.
50. Petkova AT, Leapman RD, Guo Z, Yau WM, Mattson MP, et al. (2005) Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils. Science 307: 262–265.
51. Nelson R, Eisenberg D (2006) Recent atomic models of amyloid fibril structure. Curr Opin Struct Biol 16: 260–265.
52. Lovell SC, Davis IW, Adrendall WB, de Bakker PIW, Word JM, et al. (2003) Structure validation by C-alpha geometry: phi,psi and C-beta deviation. Proteins 50: 437–450.
53. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577–2637.
54. Samudrala R, Moult J (1998) An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. J Mol Biol 275: 895–916.
55. Tiana G, Colombo M, Provasi D, Broglia RA (2004) Deriving amino acid contact potentials from their frequencies of occurrence in proteins: A lattice model study. J Phys Condens Matter 16: 2551–2564.
56. Hubbard TJ (1994) Use of beta-strand interaction pseudo-potentials in protein structure prediction and modelling. In: Lathrop RH, editor. Proteins structure prediction minitrack of the 27th HICSS. New York: IEEE Computer Society Press. pp. 336–354.
57. Wouters MA, Curmi PM (1995) An analysis of side chain interactions and pair correlations within antiparallel β-sheets: The differences between backbone hydrogen-bonded and nonhydrogen-bonded residue pairs. Proteins 22: 119–131.
58. Zhu H, Braun W (1999) Sequence specificity, statistical potentials, and three-dimensional structure prediction with self-correcting distance geometry calculations of beta-sheet formation in proteins. Protein Sci 8: 326–342.
59. Steward RE, Thornton JM (2002) Prediction of strand pairing in antiparallel and parallel β-sheets using information theory. Proteins 48: 178–191.