OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Inversion of the Balance between Hydrophobic and Hydrogen Bonding Interactions in Protein Folding and Aggregation

Anthony W. Fitzpatrick, Tuomas P. J. Knowles, Christopher A. Waudby, Michele Vendruscolo, Christopher M. Dobson*

Department of Chemistry, University of Cambridge, Cambridge, United Kingdom

Abstract

Identifying the forces that drive proteins to misfold and aggregate, rather than to fold into their functional states, is fundamental to our understanding of living systems and to our ability to combat protein deposition disorders such as Alzheimer's disease and the spongiform encephalopathies. We report here the finding that the balance between hydrophobic and hydrogen bonding interactions is different for proteins in the processes of folding to their native states and misfolding to the alternative amyloid structures. We find that the minima of the protein free energy landscape for folding and misfolding tend to be respectively dominated by hydrophobic and by hydrogen bonding interactions. These results characterise the nature of the interactions that determine the competition between folding and misfolding of proteins by revealing that the stability of native proteins is primarily determined by hydrophobic interactions between side-chains, while the stability of amyloid fibrils depends more on backbone intermolecular hydrogen bonding interactions.

Citation: Fitzpatrick AW, Knowles TPJ, Waudby CA, Vendruscolo M, Dobson CM (2011) Inversion of the Balance between Hydrophobic and Hydrogen Bonding Interactions in Protein Folding and Aggregation. PLoS Comput Biol 7(10): e1002169. doi:10.1371/journal.pcbi.1002169

Editor: Vijay S. Pande, Stanford University, United States of America

Received December 5, 2010; **Accepted** July 6, 2011; **Published** October 13, 2011

Funding: This work was supported by grants from BBSRC (AWF, MV, CMD), the Royal Society (MV) and the Wellcome Trust (CMD). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: cmd44@cam.ac.uk

Introduction

Defining the rules of protein folding, a process by which a sequence of amino acids self-assembles into a specific functional conformation, is one of the great challenges in molecular biology [1–3]. In addition, deciphering the causes of misfolding, which can often result in the formation of β-sheet rich aggregates, is crucial for understanding the molecular origin of highly debilitating conditions such as Alzheimer's and Parkinson's diseases and type II diabetes [4].

Major advances in establishing the interactions that drive the folding process have been made by analysing the structures in the Protein Data Bank (PDB), and particularly by examining the frequency with which contacts between the different types of amino acid residues occur [5]. In this statistical approach, interaction free energies are derived from the probability, p_{ij}, of two amino acids of types i and j being in contact in a representative set of protein structures using the Boltzmann relation $\Delta G_{ij} = -\ln(p_{ij})$. This operation defines a 20×20 matrix that lists the free energies of interaction between amino acid pairs. One of the most studied matrices of this type has been reported by Miyazawa and Jernigan [5]. Three distinct analyses of this matrix (Fig. 1A) have all revealed that residue-water interactions play a dominant role in protein folding [6–8].

More recently, the same statistical potential method has been used to investigate aggregation of soluble proteins into the amyloid state, now recognised as a generic, alternative, stable and highly organised type of protein structure [3]. A method for predicting the stability of amyloid structure (PASTA) [9] extracts the propensities (p_{ij}) of two residues found on neighbouring strands in parallel or antiparallel β-sheets in a representative set of PDB structures. The resulting 20×20 parallel strand and antiparallel strand interaction free energy matrices (referred to here as "parallel" and "antiparallel" respectively) are shown in Fig. 1B and 1C. Owing to the absence of a large number of solved atomic resolution amyloid fibril structures in the PDB, the central assumption of the PASTA approach is that the side-chain interactions found in the β-sheets of globular proteins are the same as those stabilising β-sheets in the core of amyloid fibrils [9]. This assumption is supported by the observation that the PASTA matrices are highly successful at predicting the portions of a polypeptide sequence that stabilise the core regions of experimentally determined amyloid fibrils and the intra-sheet registry of the β-sheets [9]. We therefore treat the PASTA matrices as statistical potentials for the parallel and antiparallel β-sheets found in the core of amyloid fibrils [9].

In this work we carry out a comparative analysis of the interaction matrices for folding and amyloid formation, in order to reveal the nature of the interactions that drive these two processes, and to provide fundamental insight into the competition between them. Our results indicate that the balance between hydrophobic and hydrogen bonding interactions is inverted in these two processes.

Results

Analysis of interaction free energy matrices

The contact approximation for the effective Hamiltonian, $\mathcal{H}^{\mathrm{eff}}(\{i_n\},\{\mathbf{r_n}\})$, used to describe a system of polypeptide chains

Author Summary

In order to carry out their biological functions, most proteins fold into well-defined conformations known as native states. Failure to fold, or to remain folded correctly, may result in misfolding and aggregation, which are processes associated with a wide range of highly debilitating, and so far incurable, human conditions that include Alzheimer's and Parkinson's diseases and type II diabetes. In our work we investigate the nature of the fundamental interactions that are responsible for the folding and misfolding behaviour of proteins, finding that interactions between protein side-chains play a major role in stabilising native states, whilst backbone hydrogen bonding interactions are key in determining the stability of amyloid fibrils.

usually takes the form

$$\mathcal{H}^{\mathrm{eff}}(\{i_n\},\{\mathbf{r_n}\}) = \sum_{n>m} M(i_n,j_m)\Delta(\mathbf{r_n}-\mathbf{r_m}) \qquad (1)$$

where i_n is the residue type i at position n along the polypeptide chain, $\mathbf{r_n}$ is the position of residue n and $\Delta(\mathbf{r})$ is a function reflecting the fact that two amino acids interact with free energy $M(i_n,j_m)$ when they are in spatial proximity to each other [10].

For random heteropolymers, the pairwise contact free energies $M(i_n,j_m)=M_{ij}$ can be approximated as a set $\{M_{ij}\}$ of 210 independent random variables (i.e. the 210 independent elements in a 20×20 symmetric matrix). For the MJ matrix, a plot with the axes running from hydrophobic (C,F,L,W,V,I,M,Y,A,P, black) [11] to hydrophilic (H,G,N,T,S,R,Q,D,K,E, magenta) [11] residue types reveals three large blocks of hydrophobic interactions (Fig. 1A). The most stabilising interactions are hydrophobic-hydrophobic (Fig. 1A, top left corner, blue), followed by hydrophobic-polar (Fig. 1A, bottom left corner and top right corner, yellow/green) and polar-polar interactions (Fig. 1A, bottom right corner, red).

On closer inspection, analysis of these interactions in the form of a histogram shows that the distribution of contact free energies determined from the Miyazawa-Jernigan (MJ) matrix (Fig. 1D) can be represented as the sum of three Gaussian terms corresponding to hydrophobic-hydrophobic (H-H), hydrophobic-polar (H-P) and polar-polar (P-P) contacts [6] (Fig. 1D). This interpretation implies that globular proteins are stabilised mainly by side-chain hydrophobic interactions [6] since the sum of all H-H, H-P and P-P contacts captures the overall distribution of contact free energies extremely well (Fig. 1D).

In contrast to the MJ matrix, contour maps of the parallel and antiparallel β-sheet contact matrices of the type characteristic of amyloid fibrils [4] show highly destabilising contact free energies between all Pro-X pairs (Fig. 1B, C, proline row, proline column, red/yellow). Since proline cannot form inter-molecular backbone hydrogen bonds this observation suggests that the stabilisation of β-sheets arises mainly from the dominance of backbone hydrogen bonding, with hydrophobic interactions (Fig. 1B, C, top left corner, blue) playing a secondary role. Furthermore, plots showing the distribution of the contact free energies from parallel and antiparallel β-sheets (Fig. 1E, F) of the type found in amyloid

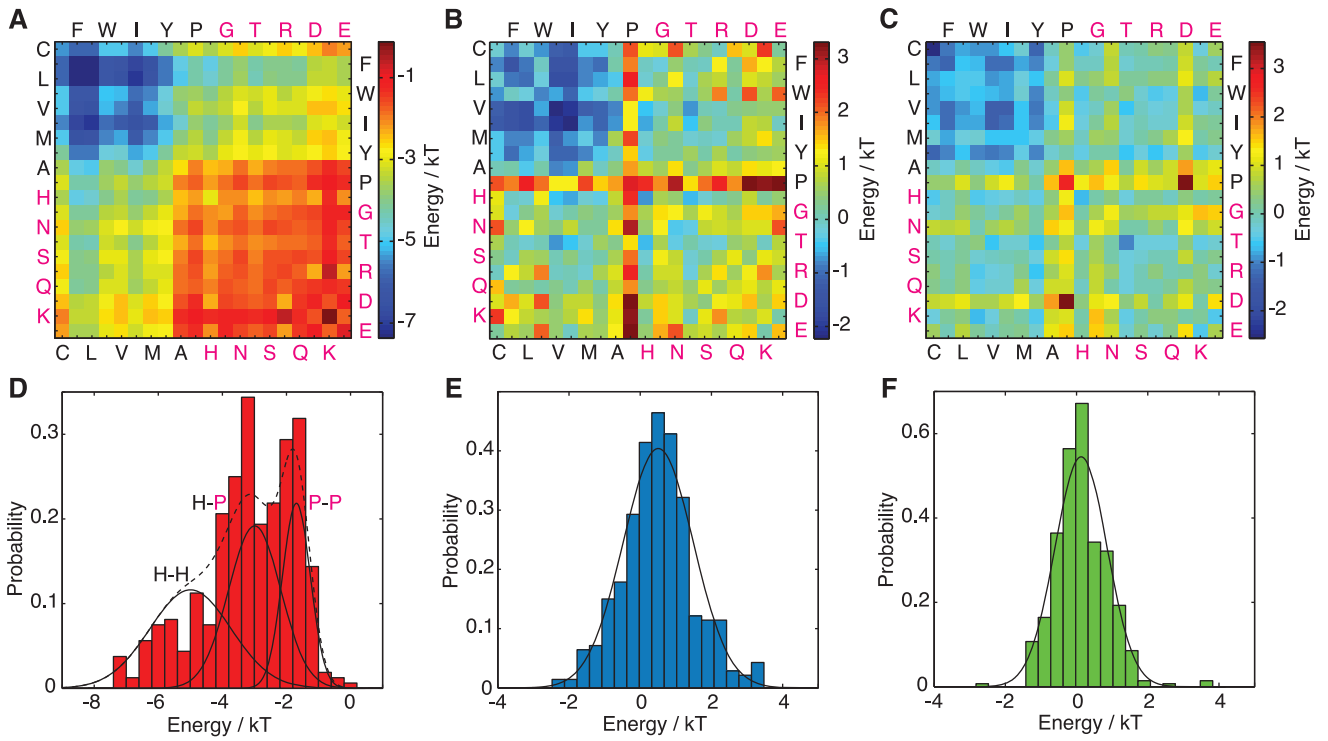

Figure 1. PDB-derived statistical potentials for folding to the native state [5] and to β-sheet rich (amyloid-like) states [9]. (A–C) Plots of the elements of the MJ matrix (A), the parallel (B) and antiparallel (C) matrices. Hydrophobic residues are shown in black and hydrophilic residues in magenta. (D) Distribution of free energies in the MJ matrix showing the decomposition of contacts into hydrophobic-hydrophobic (H-H, 37% of all contacts, -4.99 k_BT, s.d. 1.27 k_BT), hydrophobic-polar (H-P, 39% of all contacts, -2.99 k_BT, s.d. 0.82 k_BT) and polar-polar (P-P, 24% of all contacts, -1.69 k_BT, s.d. 0.44 k_BT). The sum of these components is shown as a dashed line. (E,F) Single Gaussian fits to the distributions of parallel (E) and antiparallel (F) contact free energies (0.51, s.d. 0.99 and 0.13, s.d. 0.73 (in k_BT) respectively).
doi:10.1371/journal.pcbi.1002169.g001

structures [4] indicate, unlike the situation for native folds described above, a single narrow Gaussian distribution for polar and non-polar contacts alike. This result, combined with the significance of the destabilising Pro-X contacts, is consistent with the view that a major role in protein aggregation into amyloid fibrils is played by backbone hydrogen bonding interactions [12–14], which are "generic" [3] to any polypeptide chain, although sequence-dependent effects are also important to modulate the propensity of specific peptides and proteins [15–17].

The difference in these probability distributions arises because we are examining the contact free energies that define the protein folding and misfolding free energy minima *via* the MJ and PASTA matrices respectively. It is clear that the possible number of ways of forming a given contact between amino acids i_n and j_m is greater in globular proteins than in fibrillar aggregates as the area of Ramachandran space available to β-sheets (13.3% of the total ϕ/ψ space) is much smaller than that accessible to native proteins. In addition, the type of amino acid and specific sequence patterns have varying degrees of globularity [18] or aggregation propensity [16] with certain amino acids, notably proline, appearing much more frequently in globular proteins than in the core region of amyloid fibrils [9].

To investigate the consequences of these differences in the conformational spaces relevant to folding and misfolding we consider the constrained sampling of the protein Hamiltonian $\mathcal{H}(\{i_n\},\{\mathbf{r_n}\})$ over a subspace A of conformational space, which is given formally by

$$M_{ij}^{\{A\}} = -\ln\left[\frac{1}{Z_A}\sum_{\substack{i_1,\ldots,i_N\in[1,\ldots,20]\\ n,m\in[1,\ldots,N]}} \int_{\mathrm{A}} \delta_{i_n,i}\delta_{i_m,j}\Delta(\mathbf{r_n}-\mathbf{r_m})\mathrm{d}r_1\ldots\mathrm{d}r_{\mathrm{n}}e^{-\mathcal{H}}\right] \quad (2)$$

where Z_A is the partition function sampled over the subspace A. Interaction parameters to describe the folding process are usually defined by considering a subspace A that includes the regions of conformational space corresponding to the native states of globular proteins [19]. By contrast, interaction parameters to describe the aggregation process are defined for a subspace A that includes only the regions of conformational space corresponding to β-sheet rich structures such as β-helices or amyloid fibrils [19]. While the Hamiltonian, $\mathcal{H}(\{i_n\},\{\mathbf{r_n}\})$, is invariant, the space over which it is integrated will vary depending on the region of conformational space that is being explored. In our case, this leads to distinct "effective" Hamiltonians for the protein folding and misfolding minima; these Hamiltonians have the same general form as Eq. [1] but have different amino acid interaction matrices M_{ij}, according to Eq. [2], depending on which process is involved. We thus conclude that there could be differences in the various effective energy terms stabilising globular proteins and amyloid fibrils and that such differences can be described by giving different weights to hydrophobicity and hydrogen bonding interactions in the two states. In this view, hydrophobicity and hydrogen bonding do not represent fundamental interactions but effective ones, which result from constrained sampling procedures such as those defined by Eq. [2].

Two-body terms

We decomposed the MJ and PASTA matrices into a combination of the HP (Hydrophobic-Polar) model [11] and a backbone hydrogen bonding model in which all amino acids, except for proline, are capable of forming backbone hydrogen bonds (by analogy, we term this the HB model). These two-body interactions are described by three 20×20 interaction matrices, $[hh]_{ij}$, $[hp]_{ij}$ and $[hb]_{ij}$, with the following properties: $[hh]_{ij}=-1$ if i and j are both hydrophobic residues and topological neighbours, and $[hh]_{ij}=0$ otherwise; $[hp]_{ij}=-1$ if either i or j is a hydrophobic residue, i and j are topological neighbours, and $[hp]_{ij}=0$ otherwise; $[hb]_{ij}=-1$ if i and j can both form backbone hydrogen bonds and are topological neighbours, otherwise $[hb]_{ij}=0$.

As a first approximation, we initially fit the MJ and PASTA matrices to an equation of the form:

$$M_{ij}\approx E_{HH}[hh]_{ij}+E_{HP}[hp]_{ij}+E_{HB}[hb]_{ij}+c \quad (3)$$

where M_{ij} is the matrix of interest, E_{HH}, E_{HP} and E_{HB} are the weightings of the $[hh]_{ij}$, $[hp]_{ij}$ and $[hb]_{ij}$ matrices, respectively, and c is a constant (the solvent-solvent interaction parameter) [8]. The normalisation constant c shifts the elements of the MJ and PASTA matrices along the free energy axis thus allowing comparison of E_{HH}, E_{HP} and E_{HB} between different matrices. It is used to set the free energy of forming a polar-polar contact, E_{PP}, to zero and all other weightings are measured relative to this reference, i.e. E_{HH} and E_{HP} measure the additional free energy of forming hydrophobic contacts and E_{HB} the free energy gained through hydrogen bond formation. Importantly, the adjustment of c to give E_{PP} a non-zero free energy has no effect on the ratios of E_{HB} to E_{HH} listed in Table 1. The E_{HB} weightings (Table 1) should be, and are, approximately equal to the free energy of a single hydrogen bond ($\sim$2.5 $\mathrm{k_BT}$ [20]). This simple decomposition given by Eq. [3] gives very good agreement with the MJ (correlation coefficient 0.87) and parallel matrices (correlation coefficient 0.77) and good agreement with the antiparallel matrix (correlation coefficient 0.69, or 0.70 if disulfide bonds are taken into account).

This coarse-grained HP-HB model is therefore a good approximation to the original matrices, and can thus provide insight into the relative importance of the hydrophobicity and hydrogen bonding terms for the different types of structures (Table 1). Since $[hh]_{ij}$, $[hp]_{ij}$ and $[hb]_{ij}$ are all binary matrices, it is straightforward to quantify the marginal effect of each of the regressors in our

Table 1. Hydrophobicity and hydrogen bonding terms (in $\mathrm{k_BT}$) in the HP-HB-SS model.

	E_{HH}	E_{HP}	E_{HB}	c	E_{HB}/E_{HH}
MJ (Native)	3.64 $\pm$ 0.10	1.48 $\pm$ 0.09	1.76 $\pm$ 0.12	0.07 $\pm$ 0.14	0.48 $\pm$ 0.10
Parallel (Fibrillar)	1.40 $\pm$ 0.09	0.21 $\pm$ 0.08	2.23 $\pm$ 0.11	2.97 $\pm$ 0.12	1.59 $\pm$ 0.13
Antiparallel (Fibrillar)	0.98 $\pm$ 0.08	0.14 $\pm$ 0.07	1.36 $\pm$ 0.09	1.67 $\pm$ 0.11	1.39 $\pm$ 0.15

doi:10.1371/journal.pcbi.1002169.t001

general linear model from the values of their coefficients E_{HH}, E_{HP} and E_{HB}.

For the MJ matrix, the ratio of E_{HB} to E_{HH} is ~ 0.5 (Table 1) indicating that for protein folding the hydrophobic term is twice as important as the hydrogen bonding term. This ratio was corroborated by decomposing three recent pairwise contact potentials for the native states of globular proteins [21–23] which gave a similar result (E_{HB}/E_{HH} values are 0.4 [21], 0.7 [22], 0.73 [23] and ~ 0.6 on average). This finding is in agreement with previous work suggesting that the HP model captures the essence of protein folding [11]. Nevertheless, hydrogen bonding does play an important role in protein folding since highly polar sequences can fold to form α-helices, and "side-chain only" molecular dynamics simulations fail to capture crucial aspects of protein folding [24]. Indeed, protein folding simulations have shown that it is necessary to include a mainchain-mainchain hydrogen bonding term in order to obtain secondary structure [25].

For protein misfolding and amyloid formation, the ratio of E_{HB} to E_{HH} for both PASTA matrices is ~ 1.5 (Table 1) suggesting that backbone-only hydrogen bonding is about 50% more important in stabilising amyloid fibrils than hydrophobic interactions. To demonstrate the robustness of this result, we tested the sensitivity of the E_{HB}/E_{HH} ratio to the Pro-X elements of the PASTA matrices and calculated that the high values of the Pro-X side-chain interaction free energies in the parallel and antiparallel matrices would have to be reduced by 4 or 5-fold respectively to achieve the same ratio of $E_{HB}/E_{HH} = 0.48$ found in the MJ matrix. Given that the side-chain interaction free energies are derived from the Boltzmann relation $\Delta G_{ij} = -\ln(p_{ij})$, and that the high Pro-X interaction free energies reflect the infrequent occurrence of proline residues in β-sheets, a reduction of this magnitude would translate into a much greater number of Pro-X contacts being detected in the β-sheets of the PDB dataset used by the authors of PASTA [9]. The increased weighting of the $[hb]_{ij}$ matrix relative to the $[hh]_{ij}$ matrix in the decomposition of the PASTA matrices shows that the destabilising effect of proline is more disruptive to the hydrogen bonded β-sheet structure than to the native fold of globular proteins in which proline has evolved to play an important structural, and stabilising, role e.g. in Pro-induced β-turns [26]. This result underscores the importance of sequence-independent hydrogen bonding in defining the amyloid structure. This "generic" view [12] is consistent with the observation that even hydrophilic and homopolymeric sequences of amino acids can form amyloid fibrils [13]. However, the amino acid sequences of individual peptides and proteins influence their specific propensity to aggregate [16,17], and to form self-complementary side-chain packing interfaces between adjacent β-sheets in the fibrils [15,27,28]. We also note that in the β-sheets of globular proteins, the effects of backbone hydrogen bonding tends to be averaged out in Eq. (2) by the presence of other secondary structure motifs (α-helices, β-turns and coil).

A number of controls were performed to confirm that the ratio of E_{HB} to E_{HH} is inverted between folded globular proteins and amyloid fibrils. Firstly, the value of E_{HB}/E_{HH} is only slightly affected by considering amino acids such as Proline and Alanine to be hydrophilic rather than hydrophobic. In our initial classification of hydrophobic and hydrophilic residues [11], the ratios between the hydrogen bonding and hydrophobic terms, E_{HB}/E_{HH}, are 0.48, 1.59 and 1.39 for the MJ, parallel and antiparallel PASTA matrices respectively (Table 1). By considering proline residues to be hydrophilic, rather than hydrophobic, the ratios E_{HB}/E_{HH} become 0.55, 1.78 and 1.66 for the MJ, parallel and antiparallel PASTA matrices respectively. Furthermore, if we adopt the partitioning suggested by Li, et al. [6] in which both proline and alanine residues are considered to be hydrophilic rather than hydrophobic, the ratios E_{HB}/E_{HH} become 0.61, 2.14 and 2.27 for the MJ, parallel and antiparallel PASTA matrices respectively. This analysis shows that the ratio E_{HB}/E_{HH} is inverted between the MJ and PASTA matrices using the most common classifications of amino acids into hydrophilic and hydrophobic sets.

We also note that the MJ matrix is calculated by using the quasi-chemical approximation in which protein residues are assumed to be in equilibrium with the solvent. By considering water to be the reference state, all residue-residue interactions are attractive and so all elements of the MJ matrix are negative. By ignoring chain connectivity, it has been argued that this "connectivity effect" introduces a bias into the MJ matrix. However, a knowledge-based pair potential for describing amino acid interactions in the native folds of globular proteins developed by Skolnick, et al. [21], which we refer to as the SJKG matrix, explicitly includes effects due to chain connectivity. Skolnick, et al. [21] conclude that ignoring chain connectivity does not introduce errors and that the quasi-chemical approximation is sufficient for extracting statistical potentials such as the MJ matrix. By virtue of using native reference states, the SJKG matrix has both positive and negative side-chain interaction free energies and is similar in this way to the PASTA matrices (Fig. 1B,C). The SJKG matrix also has a mean free energy of approximately zero (0.08 k_BT) like the PASTA matrices (0.51 k_BT and 0.13 k_BT for parallel and antiparallel respectively, Fig. 1B,C). However, like the MJ matrix, the SJKG is a statistical potential for the native folds of globular proteins and when we decompose this matrix using the HP-HB model we get a ratio of E_{HB} to E_{HH} of 0.4, which is almost identical to the ratio $E_{HB}/E_{HH} = 0.48$ found for the MJ matrix. Thus, this result strengthens our findings as the hydrophobicity term, E_{HH}, is even more dominant than the hydrogen bonding term, E_{HB}, in the decomposition of the SJKG matrix than in the MJ matrix (E_{HH}/E_{HB} ratios of 2.50 and 2.08 respectively). In addition, the comparison of the value of the normalisation constant c (0.94 k_BT) with the values of the E_{HB} and E_{HH} terms (0.49 k_BT and 1.24 k_BT, respectively) in the HP-HB decomposition of the SJKG matrix confirms that the value of c does not affect the ratio of E_{HB}/E_{HH} for native proteins and that this ratio is reversed between folded globular proteins and amyloid fibrils.

From the contour maps (Fig. 1A,B,C) and the histograms of contact free energies (Fig. 1D,E,F) it is clear that the free energy of forming hydrophobic-polar (H-P) side-chain contacts is stabilising for globular proteins although not nearly as important in the simple formation of β-sheets. Thus, for protein folding we find that $E_{PP} > E_{HP} > E_{HH}$ where E_{PP} is the free energy of forming a polar-polar contact and is not stabilising ($E_{PP} = 0$) and $E_{HP} = -1.4$ and $E_{HH} = -3.5$ are the free energies of forming hydrophobic-polar contacts and hydrophobic-hydrophobic contacts respectively. These weightings are in excellent agreement with a modified form of the HP model [29] ($E_{HH} : E_{HP} = 2.5$ in the present study compared to 2.3 in the modified HP model [29]) and so validate its use in protein folding simulations.

The inclusion of the HP term in Eq. [3] has only a marginal effect on the regression to the parallel or antiparallel matrices as demonstrated by the relatively small coefficient $E_{HP} \sim 0.2$ k_BT (Table 1). This result suggests that the segregation of hydrophobic and polar residues is not very important in β-sheet formation and could lead to solvent exposed non-polar side-chains in prefibrillar aggregates, a feature that has been suggested to be closely linked to cytotoxicity [30]. The minor effect of the HP term is also in accord with our finding that hydrophobic interactions play a less significant role than inter-molecular hydrogen bonding in stabilising amyloid fibrils and again supports the idea that peptides

and proteins are prone to forming amyloid structures irrespective of sequence [12,13], although the relative propensities to form such structures will vary with sequence [16,27].

One-body terms

Previous analyses of the MJ matrix shows that two-body interactions are not sufficient to capture all of the details of the 210 independent amino acid interactions that describe the variety of native protein structures [6–8]. A one-body term, η_i, describing the individual properties of each amino acid, is also required. Adding this additional term to our previous free energy expression Eq.[**3**] gives

$$M_{ij} = E_{HH}[hh]_{ij} + E_{HP}[hp]_{ij} + E_{HB}[hb]_{ij} + (\eta_i + \eta_j) + c \quad (4)$$

The application of this equation to the MJ, parallel and antiparallel matrices gives correlation coefficients of 0.99, 0.90 and 0.90 respectively (Fig. 2A,B,C). This expression, therefore, describes the original data extremely well and suggests that the diverse and complex interactions stabilising both the native and fibrillar states are amenable to a low-dimensional representation using simple two-body and one-body terms [6–8].

It is remarkable that the same approach can be used to decompose both the MJ and PASTA matrices, indicating that the underlying interactions are the same but that the balance is different, and leads to a clear demarcation of the thermodynamic minima of the native and amyloid states of the protein free energy landscape.

The three sets of 20 one-body parameters, η_i, that are derived from the MJ, parallel and antiparallel matrices are listed in Table 2. Previous work has shown that one-body components of the MJ matrix, known as q-values, are closely related to the interactions governing secondary structure formation [6]. We find that our equivalent one-body potentials, MJ η_i (Table 2), correlate extremely well with (correlation coefficient of 0.98, Fig. 3A), and are numerically almost identical to this previously published q-scale (Table 2, column 4) provided that the hydrophobic and hydrophilic q-values are separated and have their respective mean values subtracted from each non-polar and polar element. This procedure removes an average hydrophobic penalty for non-polar residues (+1.45 k_BT) and an average hydrophilic gain for polar residues (−0.07 k_BT). This residue-specific hydrophobic (hydrophilic) cost (gain) can be interpreted as an average free energy cost of placing in water the surface of a given residue plus the gain of attractive dipolar interaction between the residue concerned and water, with polar residues being more favourable than non-polar residues [7].

This effect is even more apparent in the simpler case of the one-body components of the parallel and antiparallel PASTA matrices (Table 2, parallel η_i and antiparallel η_i respectively). When existing parallel and antiparallel β-sheet propensity scales [31] are converted into free energies (Table 2, column 5 and 6 respectively), grouped into polar and non-polar terms and then separately shifted to have zero mean, thus removing the average hydrophobic (hydrophilic) cost (gain) to water of forming a β sheet (the values are +0.32 k_BT (−0.51 k_BT) and +0.34 k_BT (−0.25 k_BT) for parallel and antiparallel β-sheets respectively), the remainder correlates extremely well with (correlation coefficients of 0.96 and 0.97 for parallel and antiparallel β-sheets respectively, Fig. 3B, C), and is numerically almost identical to the one-body potentials of the parallel and antiparallel matrices (parallel η_i and antiparallel η_i respectively, Table 2). This result suggests that the one-body free energy components of the MJ, parallel and antiparallel matrices are given by

$$\eta_i = \Delta G_{\text{secondary structure}} + \Delta G_{\text{solvation}} \quad (5)$$

where $\Delta G_{\text{secondary structure}}$ represents the free energy to form hydrogen bonded secondary structure and $\Delta G_{\text{solvation}}$ is an average free energy of solvation. Hence, we suggest that the one-body free energy terms, η_i, correspond to a stabilisation of the native or fibrillar state through a competition between hydrophilicity and the formation of hydrogen bonded secondary structure.

Hydrophobicity and hydrogen bonding sculpt the free energy landscape of a protein

The HP-HB-SS (HP-HB-secondary structure) model described above suggests therefore that both the globular and amyloid states of proteins are stabilised by hydrophobic interactions, hydrogen bonding and the formation of secondary structure, and that there is a common form for the effective Hamiltonian, $\mathcal{H}^{\text{eff}}(\{s_i\},\{\mathbf{r_i}\})$, describing both protein folding and misfolding, given by the substitution of Eq.[**4**] into Eq.[**1**]

$$\mathcal{H}^{\text{eff}}(\{i_n\},\{\mathbf{r_n}\}) = \sum_{i>j}[E_{HH}[hh]_{ij} + E_{HP}[hp]_{ij} + E_{HB}[hb]_{ij} + (\eta_i + \eta_j) + c]\Delta(\mathbf{r_n} - \mathbf{r_m}) \quad (6)$$

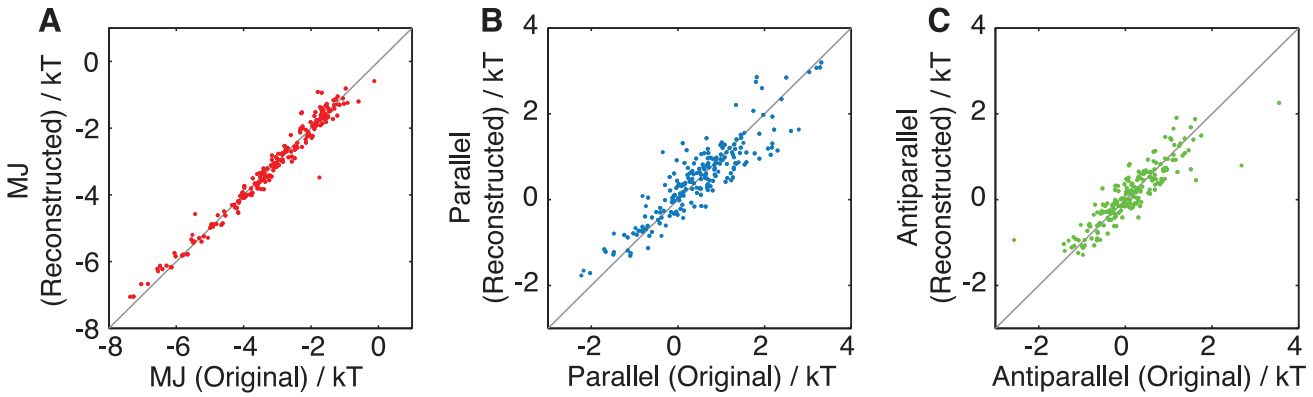

Figure 2. Correlation between the original matrix elements and the matrix elements reconstructed from equation (4). (A) MJ matrix, $r = 0.99$, rmsd 0.23 k_BT. (B) Parallel matrix, $r = 0.90$, rmsd 0.42 k_BT. (C) Antiparallel matrix, $r = 0.90$, rmsd 0.32 k_BT.
doi:10.1371/journal.pcbi.1002169.g002

Table 2. One-body potentials, η_i, for the matrices for the MJ (native) case, the parallel fibril case and the antiparallel fibril case in the HP-HB-SS model, and free energies for secondary structure formation, $\Delta G_{\text{secondary structure}}$, in k_BT [6,31]. η_i corresponds to the sum of the free energy of formation of secondary structure, $\Delta G_{\text{secondary structure}}$ and the free energy of solvation, $\Delta G_{\text{solvation}}$ (Eq. [5]).

	MJ	Parallel	Antiparallel	q-values [6]	Parallel β-sheet	Antiparallel β-sheet
	η_i	η_i	η_i		free energy [31]	free energy [31]
C	0.3775	0.3314	−0.1364	−1.3330	0.0685	−0.4569
F	−0.8575	−0.0677	−0.1439	−2.2031	−0.4304	−0.5163
L	−0.8635	−0.0037	0.2036	−2.2283	−0.3633	−0.0535
W	0.0220	0.3693	−0.1354	−1.4989	0.0069	−0.4696
V	0.0665	−0.5571	−0.2639	−1.5845	−1.0009	−0.6972
I	−0.4815	−0.5002	−0.1024	−1.9617	−0.9620	−0.5686
M	−0.0320	0.0258	0.1861	−1.6448	−0.1320	−0.0535
Y	0.4090	−0.0946	−0.3104	−1.1368	−0.2450	−0.6292
A	1.3140	0.4663	0.6531	−0.6288	0.1752	0.3474
P	0.0455	0.0304	0.0496	−0.2716	1.3643	1.0544
H	−0.5874	−0.4483	−0.3311	−0.5382	0.0008	0.0305
G	−0.1594	0.0632	0.2939	−0.2414	0.5167	0.5544
N	0.0891	0.1812	0.3249	−0.0553	0.7016	0.5942
T	−0.2749	−0.5853	−0.4491	−0.2917	−0.0449	−0.2755
S	0.0316	−0.1928	−0.1561	−0.0553	0.4718	0.1457
R	−0.0624	0.0532	−0.1831	−0.1006	0.5432	0.0133
Q	−0.0094	−0.0473	−0.1226	−0.1157	0.6694	0.1976
D	0.1986	0.4062	0.6719	0.2012	0.8437	0.7852
K	0.5506	0.2807	−0.1516	0.3270	0.6792	0.1447
E	0.2236	0.2892	0.1029	0.1408	0.7587	0.3565

doi:10.1371/journal.pcbi.1002169.t002

The two-body terms in the effective Hamiltonian are E_{HH}, E_{HP} and E_{HB}, which correspond to the relative strengths of hydrophobic interactions and hydrogen bonding, and take the values given in Table 1. The effective energy function is further modulated by the additive residue specific η_i terms (Table 2), which correspond to the free energy of secondary structure formation plus a free energy of solvation. It is important to note that there is a loss of translational and rotational entropy on going from native to fibrillar states [32] which we do not consider here. This loss of entropy would be expected to stabilise the native state in a sequence- and conformation-independent manner and would add a native-biasing term to the effective energy function given in Eq. [6].

Although the general form of the effective Hamiltonian is the same for protein folding and misfolding, the variables E_{HH}, E_{HP},

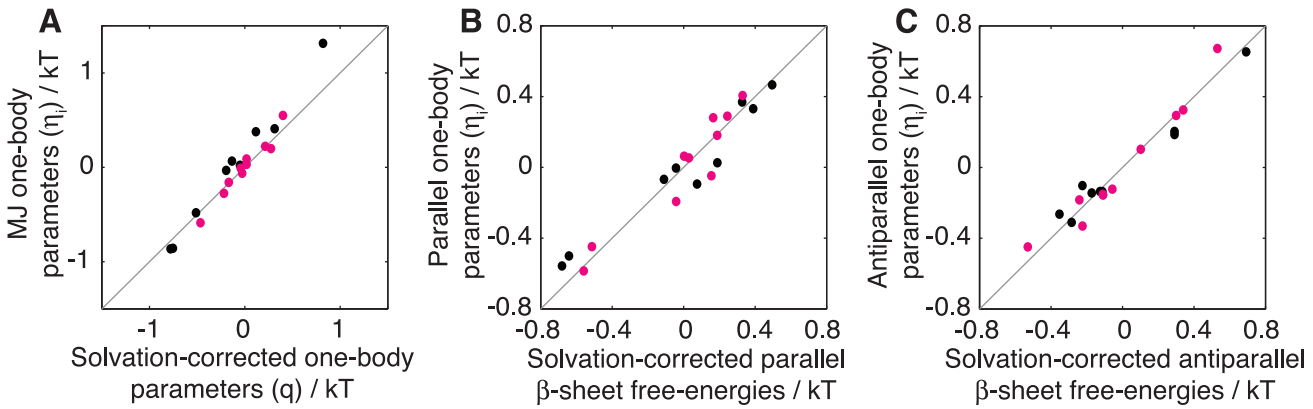

Figure 3. Correlation between the solvation-corrected free energy of secondary structure formation and one-body parameters η_i. (A) Solvation-corrected one-body parameters q *vs* MJ one-body parameters η_i, (B) Solvation-corrected parallel β-sheet free energies *vs* parallel one-body parameters η_i, and (C) correlation between the solvation-corrected antiparallel β-sheet free energies and the antiparallel one-body parameters η_i. Hydrophobic residues are shown in black and hydrophilic residues in magenta. Correlation coefficients are 0.98, 0.96 and 0.97, respectively, and the root mean square deviations are 0.16, 0.10 and 0.07 k_BT respectively.
doi:10.1371/journal.pcbi.1002169.g003

E_{HB} and η_i are different for these two processes, with the result that the minima in the two cases will occur at different positions in conformational space. Fibrillar aggregates represent a well-defined region of the wider protein folding landscape characterised by the pervasiveness of generic intermolecular hydrogen bonding [12]. Since the Hamiltonian maps the sequence space on to the structure space, as the weights E_{HH}, E_{HP} and E_{HB} change so too does the shape of the resulting structure. The dominance of the collapse-inducing hydrophobic force in protein folding leads to a globular tertiary structure, with hydrophobic residues buried in the core and largely polar residues on the surface of the protein [33]. However, when unidirectional inter-molecular hydrogen bonding is in the ascendancy, the result is ordered protein self-association into elongated, rigid, rod-like aggregates [14].

Local vs non-local effects

By decomposing the MJ and PASTA matrices into two-body and one-body components, we have effectively decoupled the two-body non-local interactions from the one-body, local interactions entangled in these statistical potentials. This approach enables us to analyse quantitatively the relative importance of local and non-local interactions in determining the folding and misfolding of proteins. It is clear from Tables 1 and 2 that the magnitude of the non-local (tertiary) interactions are significantly greater than the local (secondary) interactions in stabilising the native protein or fibrillar aggregate. This result indicates that nonlocal inter-residue interactions are the major determinant of secondary structure in the HP-HB-SS model. This finding is in excellent agreement with a large body of experimental [34] and computational analyses [35], which demonstrates that the sequence patterns of polar and non-polar amino acids dominate their intrinsic secondary structure propensities in determining the secondary structure motifs of a globular protein [36] or amyloid fibril [37]. Our prediction that hydrophobic patterning and sequence independent hydrogen bonding is more important than residue-specific identity in shaping secondary and tertiary structure helps explain why a wide variety of amino acid sequences can encode the same basic protein fold [38]. It is also consistent with the mutational robustness of functional proteins, which typically only fail to fold correctly following several mutations of individual amino acids [39]. In addition, globular proteins have evolved to mitigate against the non-local effect of polar/nonpolar periodicity by deliberately spurning alternating hydrophobic patterns which program amino acid sequences to form amphiphilic β-sheets and amyloid fibrils [40]. This is further evidence that tertiary interactions overwhelm the intrinsic propensities of individual amino acids in real proteins, which agrees with our analysis.

Role of frustration in defining the protein free energy landscape

The mathematical form of the effective Hamiltonian of Eq. [6] describing protein folding and misfolding is analogous to that of a spin glass model in which competition between conflicting interactions leads to a rugged free energy landscape [41]. Apart from topological frustration, which arises due to chain connectivity, the three sources of energetic frustration in the HP-HB-SS model stem from the competition between intramolecular collapse and intermolecular self-association, the contest between frustrating nonlocal interactions and, finally, the inability to satisfy simultaneously all local secondary structure preferences. As discussed earlier, in our model the relative strengths of the hydrophobicity to hydrogen bonding terms governs the dichotomy between folding and misfolding (Table 1). The conflicting optimisation factors imposed by hydrophobic clustering, maximal backbone hydrogen bonding and the segregation of hydrophobic and polar residues prevent the native state or fibrillar aggregate from energetically satisfying all of these inter-residue interactions. Finally, since non-local interactions predominantly determine globular [36] and fibrillar protein structures [37], there is an additional source of mismatch between the secondary structure motifs encoded by the hydrophobic patterning of the amino acid sequence as a whole and the secondary structure propensities of the individual amino acids.

This intricate interplay of competing interactions gives rise to multiple local minima in the effective energy function of Eq. [6] but, in accordance with the principle of minimal frustration [2], the sequence of a protein has evolved to reduce the number of alternative minima as much as possible and to have its native state as the global minimum of the protein folding free energy landscape [2,3]. However, the ruggedness of the folding free energy landscape increases the likelihood that excited native-like states exist, which may be transiently populated *via* thermal fluctuations, thus potentially leading to amyloid formation even under physiological conditions [42]. Moreover, frustration in the protein misfolding free energy landscape can lead to amyloid fibril polymorphs with different physical and biological properties [43].

Lowering the discordance between non-local (Table 1) and local (Table 2) interactions leads to more stable and cooperative native protein folds [35,44], and has implications for the *de novo* design of proteins [44] and amyloid fibrils [45,46]. Indeed, knowledge of the residue-specific one-body terms (Table 2), and the understanding that they correspond to the free energy of secondary structure formation once a solvation free energy is taken into account, may aid in the rational design of globular folds through mutational screening of regions known to be critical for aggregation.

Discussion

The present work indicates that there are common intermolecular forces stabilizing both globular and fibrillar states of proteins, but that a different balance of these forces results in either folding or misfolding to non-functional and potentially toxic aggregates. This situation occurs as the competing processes of protein folding and misfolding are finely tuned in terms of their free energies. Upon folding, the protein minimises the free energy of the protein-water system by clustering hydrophobic groups and forming intramolecular hydrogen bonds in the globular interior. By contrast, upon aggregation into amyloid fibrils, the formation of an extensive intermolecular hydrogen bonding network compensates for any exposure of hydrophobic groups to water that results from the fibrillar structure of the aggregated state.

It has been found in molecular dynamics simulations that the correct balance between hydrophobicity and hydrogen bonding must be attained for proteins to fold correctly or to self-assemble into the alternative well-defined amyloid structure rather than into amorphous aggregates [19,47]. For example, if hydrophobicity is too dominant, then an amorphous cluster of residues with few native contacts can be formed rather than a correctly folded protein [19]. Interestingly, these simulations suggest that hydrogen bonding is more than twice as important as hydrophobicity for aggregation into amyloid fibrils [19,48], and that hydrophobicity is approximately twice as important as hydrogen bonding for protein folding [19], findings that are in close agreement with those reported by the analysis in the present paper. Recent experimental evidence supports this interpretation of protein folding and misfolding. It has been found that the substitution of backbone ester groups for the amide linkage does not significantly affect the structure of native proteins [49], suggesting that the folded core is mainly stabilised by hydrophobic interactions. Similar experi-

ments for protein aggregation, however, reveal that peptides with removed backbone amide groups have a much reduced propensity to form ordered aggregates [50]; indeed such species are being explored as potential therapeutic inhibitors of amyloid fibril growth [51]. In addition, the large elastic modulus of amyloid fibrils stems mainly from generic inter-backbone hydrogen bonding indicating that this is a dominant interaction defining the amyloid state [14].

The weights E_{HH}, E_{HP} and E_{HB} are functions of physical [52,53] and chemical [54–56] parameters. Hydrophobic attraction, E_{HH}, and hydrogen bond interaction strength, E_{HB}, are both strongly environment-dependent intermolecular forces and vary in a complex manner as externally driven parameters such as temperature, pH, ionic strength and denaturant concentration are changed [32]. Despite the complicated nature of these interactions, experiments show that at low concentration, denaturants increase the monomer-monomer dissociation energy approximately linearly [54]. This suggests that the monomer-monomer association energy E_{HH} is a linear decreasing function of denaturant concentration under mildly denaturing conditions. In keeping with our model, we speculate that at low denaturant concentrations, E_{HH} is large, thereby promoting the native state by increasing residue-residue hydrophobic attraction, whereas at higher denaturant concentrations the lowering of E_{HH} leads to destabilisation of the hydrophobic core of the native structure, making intermolecular association much more likely [57]. Our analysis suggests that there is an optimal balance between hydrophobicity and hydrogen bonding for protein folding and a significant redistribution of these intermolecular forces for amyloid formation. Such a shift in balance can be seen as a jump between free energy landscape minima, and could occur, for example, at a critical concentration [58], or pH [55], or at a temperature sufficiently high to overcome kinetic barriers between the native and amyloid minima [46]. Overall, however, this balance appears to be very finely tuned for both protein folding and misfolding, and it is interesting to speculate on the role of this delicate balance of forces within the cell.

It has been suggested that proteins have evolved to be expressed intra-cellularly at levels in the region of the critical concentration for aggregation [58]. While a plentiful abundance of a given protein in the cell optimises its function, being on the verge of insolubility leaves proteins susceptible to environmental changes and prone to aggregation [59]. Our findings are consistent with this hypothesis [58], since elevated protein levels increase the likelihood of intermolecular as opposed to intramolecular interactions, and suggest that a precarious balance between hydrogen bonding and hydrophobic forces dictates whether peptides and proteins adopt normal or aberrant biological roles.

In conclusion, we have reported an interpretation of statistical potentials for protein folding [5] and misfolding [9] by expressing them in terms of a model containing specific terms for hydrogen bonding and hydrophobicity. This approach has enabled us to describe complex and diverse interactions using specific values of three distinct two-body terms and intrinsic secondary structure propensities. We have explained the significance of each of these terms and derived a physically meaningful common form of effective Hamiltonian for both protein folding and amyloid formation. This approach suggests that while hydrophobicity, hydrogen bonding and the formation of secondary structure are important to both processes, the balance between hydrophobicity and hydrogen bonding is remarkably different in the two regimes. Our central finding is that the stabilities of correctly folded proteins are dominated by side-chain hydrophobic interactions and that amyloid fibrils are stabilised mainly by sequence-independent intermolecular hydrogen bonding. We have also quantified the relative importance of local and non-local interactions in determining the structure and stability of proteins in both their globular and fibrillar forms and find that inter-residue interactions are more influential than secondary structure propensities in shaping the final native or amyloid fold. This result shows that, in accordance with the principle of minimal frustration [2], natural proteins have evolved to maintain a low ratio of local-to-non-local interaction strengths, thereby minimising the effect of a potent source of frustration and ensuring cooperative and stable folding [35,44].

In summary, we have found that the conflict between protein folding and misfolding is governed by the contest between a side-chain-driven hydrophobic collapse and a backbone-driven self-association. The almost infinite variety of outcomes of such a conflict gives rise to the rich and diverse behaviour exhibited by proteins and the resulting balance between health and disease.

Methods

Two-body terms

The weights of the two-body terms, E_{HH}, E_{HP}, E_{HB}, and the constant, c, were determined by performing multiple regression in MATLAB.

One-body terms

The twenty one-body terms, η_i, were determined by performing a simulated annealing minimisation in MATLAB.

Author Contributions

Conceived and designed the experiments: AWF TPJK CAW MV CMD. Performed the experiments: AWF TPJK CAW MV CMD. Analyzed the data: AWF TPJK CAW MV CMD. Contributed reagents/materials/analysis tools: AWF TPJK CAW MV. Wrote the paper: AWF TPJK MV CMD.

References

1. Fersht AR (1999) Structure and Mechanism in Protein Science W. H. Freeman & Co. New York.
2. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. Science 254: 1598–1603.
3. Dobson CM (2003) Protein folding and misfolding. Nature 426: 884–890.
4. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 75: 333–366.
5. Miyazawa S, Jernigan RL (1996) Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol 256: 623–644.
6. Li H, Tang C, Wingreen NS (1997) Nature of Driving Force for Protein Folding: A Result From Analyzing the Statistical Potential. Phys Rev Lett 79: 765–768.
7. Wang ZH, Lee HC (2000) Origin of the Native Driving Force for Protein Folding. Phys Rev Lett 84: 574–577.
8. Keskin O, Bahar I, Badretdinov AY, Ptitsyn OB, Jernigan RL (1998) Empirical solvent-mediated potentials hold for both intra-molecular and inter-molecular inter-residue interactions. Protein Sci 7: 2578–2586.
9. 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: 1608–1618.
10. Pande VS, Grosberg AY, Tanaka T (1997) Statistical mechanics of simple models of protein folding and design. Biophys J 73: 3192–3210.
11. Dill KA (1985) Theory for the folding and stability of globular proteins. Biochemistry 24: 1501–1509.
12. Dobson CM (1999) Protein misfolding, evolution and disease. Trends Biochem Sci 24: 329–332.
13. Fändrich M, Dobson CM (2002) The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. EMBO J 21: 5682–5690.

14. Knowles TP, Fitzpatrick AW, Meehan S, Mott HR, Vendruscolo M, et al. (2007) Role of intermolecular forces in defining material properties of protein nanofibrils. Science 318: 1900–1903.
15. 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.
16. 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.
17. Tartaglia GG, Pawar AP, Campioni S, Chiti F, Dobson CM, et al. (2005) Prediction of aggregation-prone regions of structured proteins. J Mol Biol 380: 425–436.
18. Linding R, Russell RB, Neduva V, Gibson TJ (2003) GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res 31: 3701–3708.
19. Hoang TX, Trovato A, Seno F, Banavar JR, Maritan A (2004) Geometry and symmetry presculpt the free-energy landscape of proteins. Proc Natl Acad Sci U S A 101: 7960–7964.
20. Fersht AR, Shi JP, Knill-Jones J, Lowe DM, Wilkinson AJ, et al. (1985) Hydrogen bonding and biological specificity analysed by protein engineering. Nature 314: 235–238.
21. Skolnick J, Jaroszewski L, Kolinski A, Godzik A (1997) Derivation and testing of pair potentials for protein folding. When is the quasichemical approximation correct? Protein Sci 6: 676–688.
22. Bastolla U, Farwer J, Knapp EW, Vendruscolo M (2001) How to guarantee optimal stability for most representative structures in the Protein Data Bank. Proteins 44: 79–96.
23. Papoian GA, Ulander J, Eastwood MP, Luthey-Schulten Z, Wolynes PG (2004) Water in protein structure prediction. Proc Natl Acad Sci U S A 101: 3352–3357.
24. Honig B, Cohen FE (1996) Adding backbone to protein folding: why proteins are polypeptides. Fold Des 1: R17–R20.
25. Hunt NG, Gregoret LM, Cohen FE (1994) The origins of protein secondary structure. Effects of packing density and hydrogen bonding studied by a fast conformational search. J Mol Biol 241: 214–225.
26. Li SC, Goto NK, Williams KA, Deber CM (1996) α-helical, but not β-sheet, propensity of proline is determined by peptide environment. Proc Natl Acad Sci U S A 93: 6676–6681.
27. 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.
28. Sawaya MR, Sambashivan S, Nelson R, Ivanova MI, Sievers SA, et al. (2007) Atomic structures of amyloid cross-beta spines reveal varied steric zippers. Nature 447: 453–457.
29. Li H, Helling R, Tang C, Wingreen N (1996) Emergence of preferred structures in a simple model of protein folding. Science 273: 666–669.
30. Campioni S, Mannini B, Zampagni M, Pensalfini A, Parrini C, et al. (2010) A causative link between the structure of aberrant protein oligomers and their toxicity. Nat Chem Biol 6: 140–147.
31. Steward RE, Thornton JM (2002) Prediction of strand pairing in antiparallel and parallel beta-sheets using information theory. Proteins 48: 178–191.
32. Zamparo M, Trovato A, Maritan A (2010) Simplified exactly solvable model for β-amyloid aggregation. Phys Rev Lett 105: 108102.
33. Dill KA (1990) Dominant forces in protein folding. Biochemistry 29: 7133–7155.
34. Xiong H, Buckwalter BL, Shieh HM, Hecht MH (1995) Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. Proc Natl Acad Sci U S A 92: 6349–6353.
35. Bellesia G, Jewett AI, Shea JE (2010) Sequence periodicity and secondary structure propensity in model proteins. Protein Sci 19: 141–154.
36. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. Science 262: 1680–1685.
37. Kim W, Hecht MH (2006) Generic hydrophobic residues are sufficient to promote aggregation of the Alzheimer's Abeta42 peptide. Proc Natl Acad Sci U S A 103: 15824–15829.
38. Koonin EV, Wolf YI, Karev GP (2002) The structure of the protein universe and genome evolution. Nature 420: 218–223.
39. Axe DD, Foster NW, Fersht AR (1996) Active barnase variants with completely random hydrophobic cores. Proc Natl Acad Sci U S A 93: 5590–5594.
40. Broome BM, Hecht MH (2000) Nature disfavors sequences of alternating polar and non-polar amino acids: implications for amyloidogenesis. J Mol Biol 296: 961–968.
41. Bryngelson JD, Wolynes PG (1987) Spin glasses and the statistical mechanics of protein folding. Proc Natl Acad Sci U S A 84: 7524–7528.
42. Chiti F, Dobson CM (2009) Amyloid formation by globular proteins under native conditions. Nat Chem Biol 5: 15–22.
43. Mossuto MF, Dhulesia A, Devlin G, Frare E, Kumita JR, et al. (2010) The Non-Core Regions of Human Lysozyme Amyloid Fibrils Influence Cytotoxicity. J Mol Biol 402: 783–796.
44. Muñoz V, Serrano L (1996) Local versus nonlocal interactions in protein folding and stability–an experimentalist's point of view. Fold Des 1: R71–R77.
45. West MW, Wang W, Patterson J, Mancias JD, Beasley JR, et al. (1999) De novo amyloid proteins from designed combinatorial libraries. Proc Natl Acad Sci U S A 96: 11211–11216.
46. Kammerer RA, Kostrewa D, Zurdo J, Detken A, García-Echeverría C, et al. (2004) Exploring amyloid formation by a de novo design. Proc Natl Acad Sci U S A 101: 4435–4440.
47. Nguyen HD, Hall CK (2006) Spontaneous fibril formation by polyalanines; discontinuous molecular dynamics simulations. J Am Chem Soc 128: 1890–1901.
48. Auer S, Meersman F, Dobson CM, Vendruscolo M (2008) A Generic Mechanism of Emergence of Amyloid Protofilaments from Disordered Oligomeric Aggregates. PLoS Comput Biol 4: 1–7.
49. Wang M, Wales TE, Fitzgerald MC (2006) Conserved thermodynamic contributions of backbone hydrogen bonds in a protein fold. Proc Natl Acad Sci U S A 103: 2600–2604.
50. Gordon DJ, Meredith SC (2003) Probing the role of backbone hydrogen bonding in beta-amyloid fibrils with inhibitor peptides containing ester bonds at alternate positions. Biochemistry 42: 475–485.
51. Sciarretta KL, Gordon DJ, Meredith SC (2006) Peptide-based inhibitors of amyloid assembly. Methods Enzymol 413: 273–312.
52. Baldwin RL (1986) Temperature dependence of the hydrophobic interaction in protein folding. Proc Natl Acad Sci U S A 83: 8069–8072.
53. Ferrã-Gonzales AD, Souto SO, Silva JL, Foguel D (2000) The preaggregated state of an amyloidogenic protein: Hydrostatic pressure converts native transthyretin into the amyloidogenic state. Proc Natl Acad Sci U S A 97: 6445–6450.
54. Ghosh K, Dill KA (2009) Computing protein stabilities from their chain lengths. Proc Natl Acad Sci U S A 106: 10649–10654.
55. Guijarro JI, Sunde M, Jones JA, Campbell ID, Dobson CM (1998) Amyloid fibril formation by an SH3 domain. Proc Natl Acad Sci U S A 95: 4224–4228.
56. Chiti F, Webster P, Taddei N, Clark A, Stefani M, et al. (1999) Designing conditions for in vitro formation of amyloid protofilaments and fibrils. Proc Natl Acad Sci USA 96: 3590–3594.
57. London J, Skrzynia C, Goldberg ME (1974) Renaturation of Escherichia coli tryptophanase after exposure to 8 M urea. Evidence for the existence of nucleation centers. Eur J Biochem 47: 409–415.
58. Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M (2007) Life on the edge: a link between gene expression levels and aggregation rates of human proteins. Trends Biochem Sci 32: 204–206.
59. Vendruscolo M, Dobson CM (2009) Quantitative approaches to defining normal and aberrant protein homeostasis. Faraday Discuss 143: 277–291.