A Simple Lattice Model That Captures Protein Folding, Aggregation and Amyloid Formation

Sanne Abeln[1]*, Michele Vendruscolo[2], Christopher M. Dobson[2], Daan Frenkel[2]

[1] IBIVU - Deptartment of Computer Science, VU University, Amsterdam, The Netherlands, [2] Deptartment of Chemistry, University of Cambridge, Cambridge, United Kingdom

Abstract

The ability of many proteins to convert from their functional soluble state to amyloid fibrils can be attributed to intermolecular beta strand formation. Such amyloid formation is associated with neurodegenerative disorders like Alzheimer's and Parkinson's. Molecular modelling can play a key role in providing insight into the factors that make proteins prone to fibril formation. However, fully atomistic models are computationally too expensive to capture the length and time scales associated with fibril formation. As the ability to form fibrils is the rule rather than the exception, much insight can be gained from the study of coarse-grained models that capture the key generic features associated with amyloid formation. Here we present a simple lattice model that can capture both protein folding and beta strand formation. Unlike standard lattice models, this model explicitly incorporates the formation of hydrogen bonds and the directionality of side chains. The simplicity of our model makes it computationally feasible to investigate the interplay between folding, amorphous aggregation and fibril formation, and maintains the capability of classic lattice models to simulate protein folding with high specificity. In our model, the folded proteins contain structures that resemble naturally occurring beta-sheets, with alternating polar and hydrophobic amino acids. Moreover, fibrils with intermolecular cross-beta strand conformations can be formed spontaneously out of multiple short hydrophobic peptide sequences. Both the formation of hydrogen bonds in folded structures and in fibrils is strongly dependent on the amino acid sequence, indicating that hydrogen-bonding interactions alone are not strong enough to initiate the formation of beta sheets. This result agrees with experimental observations that beta sheet and amyloid formation is strongly sequence dependent, with hydrophobic sequences being more prone to form such structures. Our model should open the way to a systematic study of the interplay between the factors that lead to amyloid formation.

Citation: Abeln S, Vendruscolo M, Dobson CM, Frenkel D (2014) A Simple Lattice Model That Captures Protein Folding, Aggregation and Amyloid Formation. PLoS ONE 9(1): e85185. doi:10.1371/journal.pone.0085185

Editor: Ilia V Baskakov, University of Maryland School of Medicine, United States of America

Received June 8, 2013; **Accepted** December 3, 2013; **Published** January 15, 2014

Funding: SA has been supported by a Veni grant on the project 'Understanding toxic protein oligomers through ensemble characteristics' from Netherlands Organisation for Scientic Research (NWO), and a travel-fellowship from the European Molecular Biology Organization (EMBO). MV and CMD acknowledge support from the Wellcome Trust (ref.). DF acknowledges support of ERC Advanced Grant 227758, a Wolfson Merit Award of the Royal Society of London. 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: s.abeln@vu.nl

Introduction

The ability of many peptides and proteins to convert from their monomeric native state to amyloid fibrils can be attributed to the general ability of polypeptide chains to form intermolecular beta-strands [1]. Simulations, using detailed models, have revealed possible pathways for fibril formation by small peptides [2–9]. These models provide valuable information about the factors that determine the free-energy barriers for the nucleation step in the fibril formation pathway and have led to interesting hypotheses about the origins of amyloid toxicity [6,10,11]. However, as these models are computationally very expensive, these studies have mostly concerned short segments of amyloidogenic proteins. Yet, to model biologically relevant behaviour of the nucleation pathway, including the formation of oligomers in the initial stages of aggregation, simulation of a substantial number of complete protein chains is necessary; regions of the protein that are not directly implicated in beta strand formation of the amyloid fibrils may still be highly relevant for the aggregation and amyloid formation pathway. In fact, there is evidence that flanking regions can have a considerable effect on aggregation mechanisms [12,13].

Lattice models have previously been used successfully to model the transition from an unfolded to a folded protein [14] and to model the rearrangement of hydrophobic and polar residues in oligomeric structures [13]. The success of such highly simplified lattice models may at least partly be explained by recent simulations that show that the free-energy landscapes for protein folding in a lattice model are strikingly similar to those obtained for a fully atomistic model [14,15]. A limitation of existing lattice models is, however, that they do not account for backbone-specific interactions, yet such interactions are vital to understanding the transition from oligomers to fibrils. Some off-lattice coarse-grained models can capture folding specificity [4,16]. Other off-lattice models can simulate the transition from monomeric peptides to non-specific aggregates [2,6,8,9]. Yet, such off-lattice models, even when coarse grained, are generally too expensive to study the unfolding and rearrangement of multiple proteins.

In this paper, we propose a lattice model that allows both for folding into specific structures and the formation of backbone hydrogen bonds in patterns that are typically observed in amyloid fibres. The model is sufficiently simple to allow extensive simulations of large systems under varying conditions; it also includes explicit directional information of the side chains, and explicit formation of backbone hydrogen bonds between residues. These features make it possible to model the geometric properties commonly observed in beta strands at low computational cost.

In the model developed here backbone residues can make hydrogen bonds with the corresponding residues in neighbouring strands. In addition, the use of an information-based pairwise interaction between the twenty different amino acids, together with the directionality of the side chains, allow for the formation of highly specific structures. As we will show below, beta strands will only form under specific conditions, and when the amino acid composition of the sequence involved creates a favourable environment. Moreover, a previously developed interaction matrix, allows for the simulation of multiple protein molecules without creating unrealistic aggregation behaviour [17].

In what follows, we show that our model allows the design of amino acid sequences that fold into specific target structures. This design process naturally leads to the formation of beta sheets with a hydrophobic and a polar face containing alternating charges. In addition to describing native states, the model can also account for intermolecular beta strands arranged in a cross beta structure, as observed in amyloid fibrils. We find that the formation of such structures is sensitive to the amino acid composition of the peptides. The model proposed here is simple enough to simulate the collective rearrangement of multiple full-length proteins in the initially formed oligomers and the mature amyloid fibrils. The ability of the model to simulate interplay between aggregation and folding is shown in Ref. [18].

Results and Discussion

Due to the coarse grained nature of the cubic lattice, it is necessary to design sequences that can fold into specific and stable structures. The design process takes a structure as an input, and designs an appropriate sequence. The amino acid composition is altered during the design process, while the structure is kept rigid. Designing a sequence that folds with high specificity can be achieved through minimising the total interaction energy E of the sequence based on the input structure. Here we use an adapted version of the minimisation procedure used by Coluzza et al. [14]. In this work the direction of the side chains are also altered in the design procedure and the amino acid distribution is constrained to those of naturally occurring proteins (see methods for further details).

A typical sequence design for a predefined structure is shown in Figure 1. Just as in the experimental structures (e.g. 1OSP), it can be observed that the designed structure contains a hydrophobic core (yellow residues pointing inwards) and a hydrophilic surface (red, blue and grey residues pointing outwards). Moreover, the outer surface of the beta sheet shows alternating positive (blue) and negative (red) residues, similar to the residues in the experimental structure. The realistic amino acid sequence composition of the beta strands demonstrates the biological and physical relevance of the directional amino-acid interaction potential defined by the model.

Folding with high specificity

Once a sequence has been designed, a Monte Carlo simulation can be used to investigate the folding characteristics of the sequence. We find that the heat capacity of the designed model proteins exhibit a sharp peak in the vicinity of the folding temperature (e.g., Figure 2). Experimentally, such sharp peaks are also observed as a result of folding specificity [19–21].

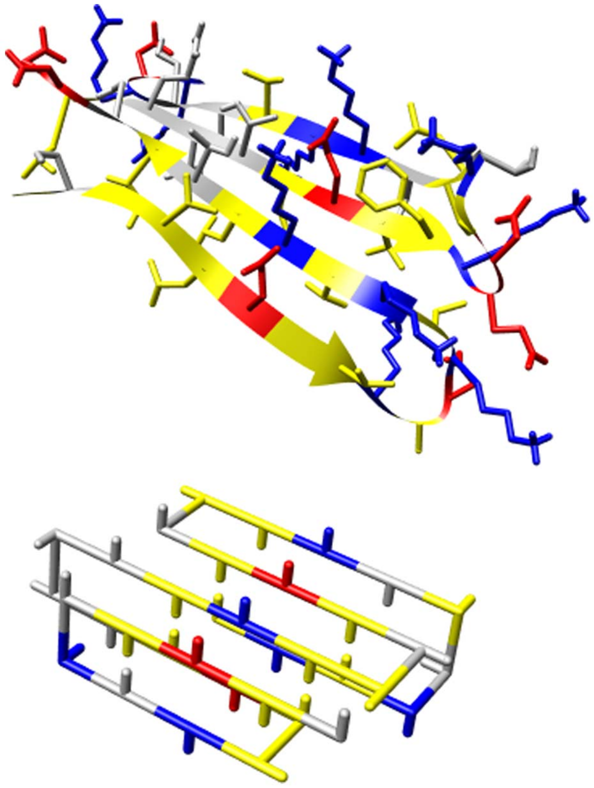

Figure 1. Example of experimental beta sheet structure compared to an on-lattice model. Top: example of a beta sheet in an all-atom structure (1OSP). Note that for clarity only the top sheet of the structure is shown; the hydrophobic downward pointing residues are buried through an another beta sheet. Bottom: example of a designed protein that can fold into its native structure (shown) containing a beta sheet. Note that the designed structure was not explicitly modelled to resemble the experimental structure, but is the result of a stochastic design algorithm (see Methods). Yellow, grey, red and blue residues indicate hydrophobic, polar, negative and positive amino acids respectively.
doi:10.1371/journal.pone.0085185.g001

In addition, the folding transition may be detected by considering the average number of native contacts in each configuration of the model protein during the simulation. Native contacts are those contacts that are also present in the structure for which the sequence was designed (the native state). Figure 2 (b) shows that the designed sequences fold well at low temperatures with a high number of native contacts. A sharp transition in this order parameter can be observed from the folded to the unfolded state, at the same temperature as the peak in the heat capacity, indicating again the high specificity of folding. For random sequences with a similar amino acid composition there is no evidence for a sharp 'folding transition' in the heat capacity nor in the native contact curves (Figure 2). This result is in agreement with the experimental observation that most random peptide sequences do not fold into well-defined structures [22]. The more gradual transitions for the random sequences may be considered as a collapse into a more compact, molten globule-like state, without a preference for one specific structural arrangement.

Hence at low temperatures, the random sequence forms an ensemble of compact structures, where as the designed sequence folds almost perfectly into its designed structure with high

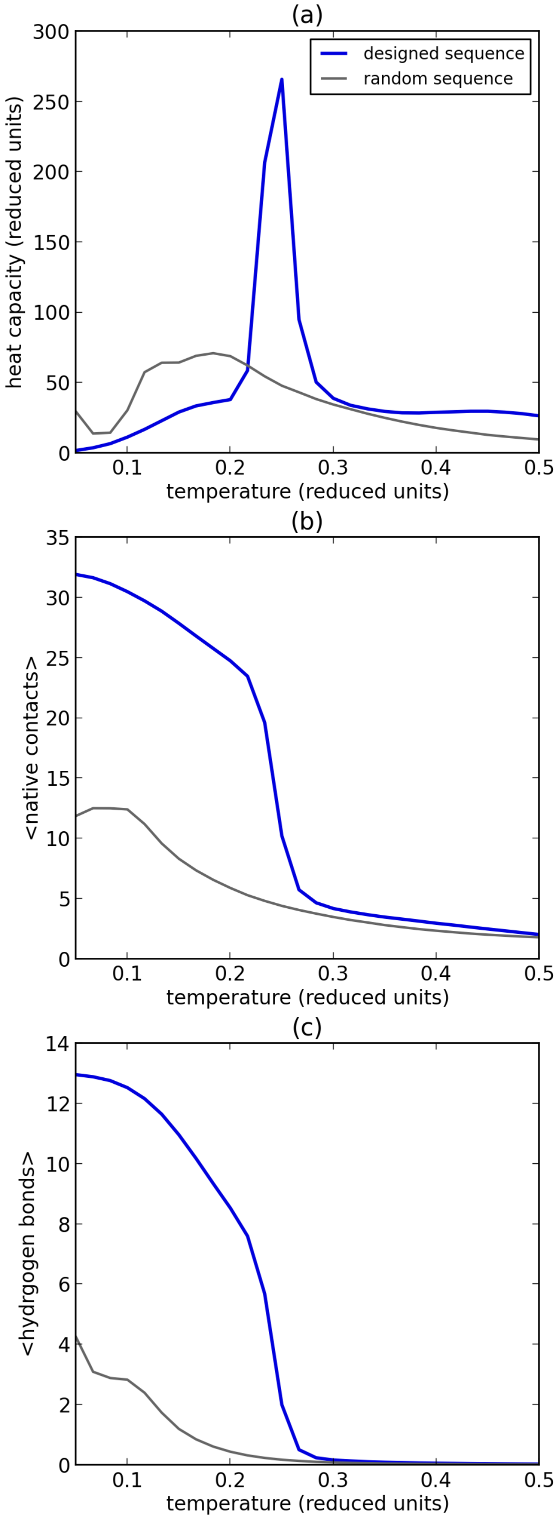

Figure 2. Folding chacteristics and specificity. Folding characteristics are shown for a protein sequence that is designed to fold in a specific structure, and a random protein sequence; both sequences contain 35 residues and have a similar amino acid composition (see Methods). (a) Heat capacity versus temperature. A peak in the heat capacity curve can be observed at the folding transition. (b) Number of native contacts versus temperature. (c) Number of hydrogen bonds versus temperature. From the statistics it is clear that the sequence designed to fold shows a much sharper transitions than a random sequence of the same length. Moreover, the number of hydrogen bonds formed is strongly dependent on the sequence. Please refer to the Methods and Supplement for the sequences and structures used.
doi:10.1371/journal.pone.0085185.g002

specificity. Similar results have, of course, been obtained by the classic lattice model; however, it is non-trivial (and encouraging) that this feature of specific folding survives in our model where interactions depend on the direction of the side chains. This model is approximately 2–3 fold slower than the classic cubic lattice model as in ref. [17]; typical folding times are around 1–5 CPU hours on a single processor of 2.2 GHz, depending on the length of the sequence.

Lastly, we consider the effect of the hydrogen bonds on the folding behaviour. Figure 2 shows that the ensemble average of the number of hydrogen bonds follows a similar sharp transition to the number of native contacts. This implies that these hydrogen bonds, and therefore the beta strands, cannot be formed unless the side chain interactions are also favourable - as is the case in the folded structure.

Figure 2 also shows that random sequences hardly form any hydrogen bonds. This finding indicates that the interaction potential of the hydrogen bonds is not unrealistically strong and that the formation and stability of beta strands depends on the sequence.

Here a note of caution should also be given: the possibility to form a helical structure is not included in this model; since the stability of beta strands with respect to the disordered coil state is highly sequence dependent, it would not be realistic to model sequences with a high helical propensity on the cubic lattice.

The property that the formation of backbone hydrogen bonds is strongly sequence dependent is in agreement with experimental results: hydrogen bonds between the backbone atoms only form when the side chain interactions are favourable, e.g. refs. [23–25].

Formation of cross-beta fibrillar structures

We simulated several small peptides with an alternating hydrophobic and hydrophilic sequence composition to test the ability of our model to form intermolecular beta sheets. The simulations started from configurations where there were no initial contacts between the different peptides. On simulation both disordered oligomers (i.e. amorphous aggregates) and short fibrillar structures were observed. The short fibrillar structures appeared in relatively long (20 CPU hours), but unbiased, simulations at low temperatures ($T<0.12$) with a constant number of peptides. When the short fibrils are subsequently simulated in a grand canonical ensemble, further growth of the structure may be observed. Figure 3 shows a typical snapshot of such a simulation procedure: long linear fibrillar structures have been formed with a cross beta-architecture.

We find that the small fibrillar structures that form at low temperature simulations are extremely stable. We investigated at which temperature such fibrils would become soluble through a series of simulations starting from the fibrillar structure at various temperatures; the initial oligomeric configurations contained 10 peptides, with different sequence compositions. Figure 4 shows at which temperatures the fibrils dissociate; here a high number of intermolecular contacts indicates that the fibrils remain stable (low temperatures) and a low number of intermolecular contacts indicates the peptides are stable as monomers. Hence, these small fibrillar structures, that were initially formed at low temperatures, remain stable at higher temperatures. In addition, such fibrillar seeds enable further growth of the fibrils in grand canonical simulations (Figure 3). Both results are in agreement with a templating (or seeding) mechanism for fibril formation, through

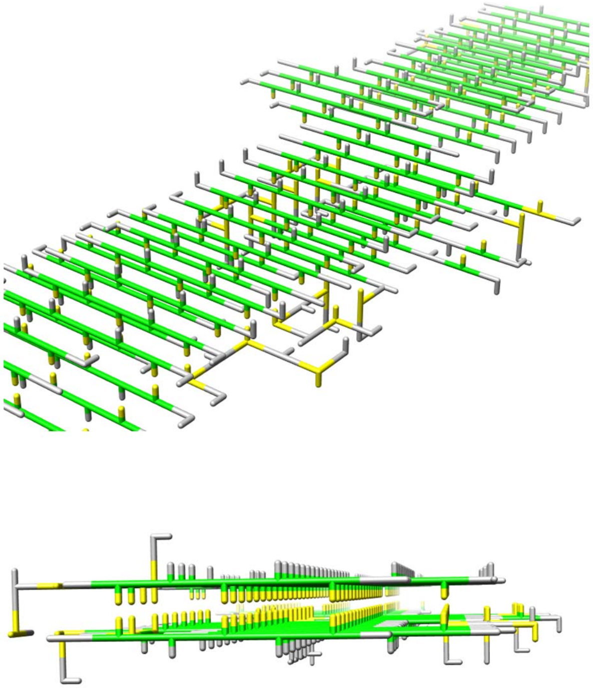

Figure 3. Fibrils with a cross-beta architecture. Top and side view of fibrils formed by a grand canonical simulation with a starting configuration of s small fibrillar structure. The peptides have an alternating hydrophobic (yellow) and hydrophilic (grey) sequence composition. The *strand* and *coil* states are indicated by green and grey respectively.
doi:10.1371/journal.pone.0085185.g003

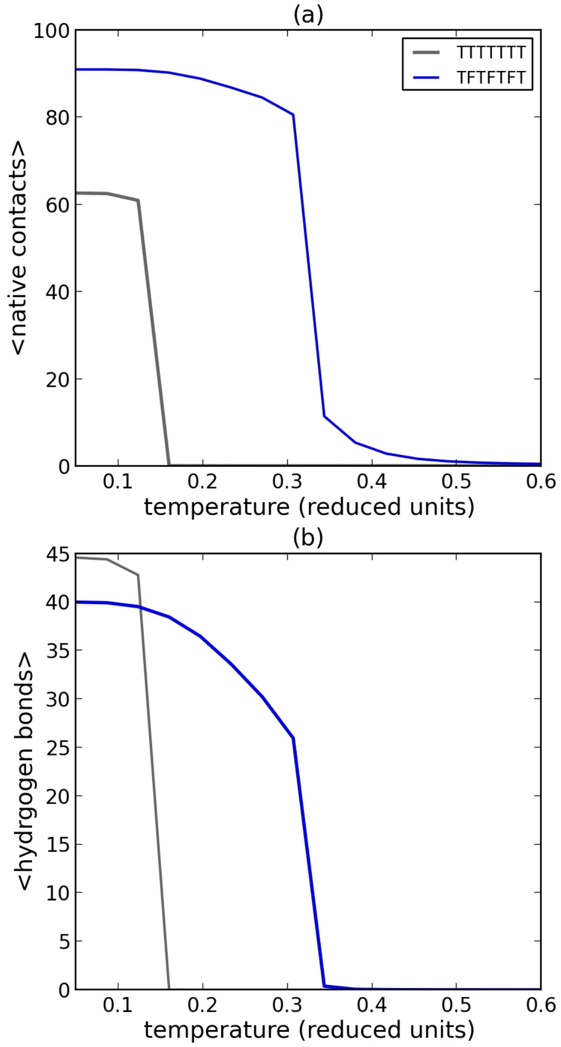

Figure 4. Stability of small fibrillar structures containing 10 peptides. (a) The ensemble average of external (intermolecular) contacts versus temperature for different peptide sequences. External contacts are contacts between different peptides. (b) The ensemble average of hydrogen bonds versus temperature for different peptide sequences. These simulations started from small fibrillar configuration containing 10 peptides with 7 residues and different sequence compositions (see legend). In the temperature regime relevant for folding ($0.2 < T < 0.3$), only fibrils that could form a strong hydrophobic core (TFTFTFT) are stable, in this case the hydrophobic residues would point inwards, as shown in Figure 3.
doi:10.1371/journal.pone.0085185.g004

which the typical lag times observed in fibril formation may be explained [1,25,26].

Figure 4 also shows that the formation of hydrogen bonds occurs at slightly lower temperatures than those at which the first intermolecular contacts form. This suggests that the formation of intermolecular hydrophobic interactions is stronger and enables subsequent hydrogen bonding between peptides. Note that similar observations have been made by all-atom modelling [27,28]. Comparing the peptide sequences with different amino acid compositions shows that hydrogen bond formation is strongly sequence dependent. Hence in this model both favourable interactions between side chains and hydrogen bonding are necessary for the creation of beta strands in the fibrils. This is particularly evident for the temperature range relevant for protein folding ($02. < T < 0.3$).

Conclusions

In this work we have presented a very simple protein lattice model that includes directional information of the side chains and ability to form hydrogen bonds. Our results show that this model can be used to design sequences that fold with high specificity into predefined structures. Moreover, simulations with the model show that hydrogen bonds are formed both in beta sheet motifs in folded proteins, and in intermolecular cross-beta structures in fibrils formed from small peptides. Most importantly, the simplicity of the model makes simulations feasible that investigate the interplay between folding, fibril formation and amorphous aggregation [18]. The full model, given as source code, is available at http://www.few.vu.nl/~abeln/hb-lattice.

Methods

The model

As a basis for the model presented here, we use the classic cubic lattice model as described in [13,14,17,29,30]. In this classic lattice model each residue is located on a point of the cubic lattice and is assigned one of twenty amino acid types. In the remainder of this section, only the differences between the model developed here and the model described in refs. [13,17] will be discussed.

In our model, each residue has a side chain direction, $\hat{d}_i$, and a state representing the secondary structure, s_i. In a structure of N residues, we now have for each residue $i \in \{1, \cdots, N\}$:

$\vec{p}_i \in \mathbb{R}^3$	position
$s_i \in \{\text{strand,coil}\}$	state
$\hat{d}_i \in \mathbb{R}^3$ (unitvector)	side chain direction
$a_i \in \{\text{Ala,Arg}, \cdots ,\text{Val}\}$	amino acid type

Note that all residue positions $\vec{p}_i$ are situated on the cubic lattice. The side chain directions $\hat{d}_i$ do not occupy any colume and point from their residue's position to a neighbouring lattice site. The side chain is not allowed to point in the same direction as the backbone, leaving a choice of four possible directions for each side chain; the side chains situated at the end of the protein chain, have a choice of five directions.

A residue makes a contact with another residue when it is situated on a neighbouring lattice point, and is not a sequential neighbour in the chain. A contact $C_{i,j}$ between two residues i and j is thus defined as:

$$C_{i,j} = \begin{cases} 1 & \text{if} |\vec{p}_i - \vec{p}_j| = 1 \text{ and } |i-j| > 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Potential energy. The total energy of the system can be split into several different and independent components including the hydrogen bond energy (E_{hb}), the interaction energy between amino acids (E_{aa}), the energy of the states (E_{state}) and the interaction energy of each amino acid with the solvent (E_{solvent}):

$$E = E_{\text{hb}} + E_{\text{aa}} + E_{\text{steric}} + E_{\text{state}} + E_{\text{solvent}} \tag{2}$$

Here the contributing potential energy terms are functions of the following variables:

$$E_{\text{hb}} = E_{\text{hb}}(\vec{p}_i, s_i, \hat{d}_i)$$

$$E_{\text{aa}} = E_{\text{aa}}(\vec{p}_i, \hat{d}_i, a_i)$$

$$E_{\text{steric}} = E_{\text{steric}}(\hat{d}_i)$$

$$E_{\text{state}} = E_{\text{state}}(s_i)$$

$$E_{\text{solvent}} = E_{\text{solvent}}(\vec{p}_i, \hat{d}_i, a_i)$$

Note that in the model described here no explicit energy is attributed to the state of the residue, i.e. $E_{\text{state}} = 0$. Instead, favourable hydrogen bonding interactions will bias residues towards the appropriate state.

Hydrogen bonds. The total potential energy of the hydrogen bonds for a configuration is given by:

$$E_{\text{hb}} = \frac{1}{2} \sum_i^N \sum_j^N \epsilon_{hb} \cdot H_{i,j} \cdot C_{i,j} \tag{3}$$

where ϵ_{hb} represents the potential energy per hydrogen bond and $H_{i,j} = 1$ indicates whether or not a hydrogen bond between residues i and j exists ($H_{i,j} = 1$ in case of a hydrogen bond). Hydrogen bonds are only allowed between two residues that are both in the 'strand' state, and when their side chains are oriented in the same direction (Figure 5), thus:

$$H_{i,j} = \begin{cases} 1 & \text{if } s_i, s_j = \text{strand} \quad \text{and} \quad \hat{d}_i = \hat{d}_j \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Interactions between amino acids. In the presented model the total potential energy for pairwise interactions between amino acids depends both on the positions on the lattice and side chain directions of the two residues. The potential energy of pairwise amino acid interactions, E_{aa}, is given by:

$$E_{\text{aa}} = \frac{1}{2} \sum_i^N \sum_j^N C_{i,j} \cdot K_{i,j} \cdot M_{a_i, a_j} \tag{5}$$

where $C_{i,j}$ indicates whether or not the residues are in contact, as before, and $K_{i,j}$ indicates whether the directions of residues i and j allow interation. Elements of the interaction matrix M provide the strengths of the pairwise interaction energies between different types of amino acids (a_i, a_j). The directions of the side chains are allowed to interact when the side chains of the residues face each other (Figure 6) or when the side chains lie parallel to each other, while being oriented in the same direction (Figure 7), thus:

$$K_{i,j} = \begin{cases} 1 & \text{if } \hat{d}_i = -\hat{d}_j \text{ and } |(\vec{p}_i + \hat{d}_i) - (\vec{p}_j + \hat{d}_j)| = 1 \\ 1 & \text{if } \hat{d}_i = \hat{d}_j \text{ and } |(\vec{p}_i + \hat{d}_i) - (\vec{p}_j + \hat{d}_j)| = 1 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

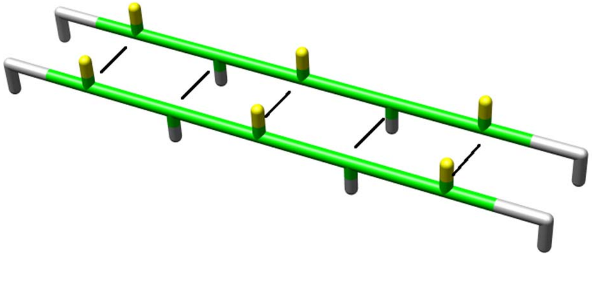

Figure 5. Hydrogen bonds formed between two strands. The black lines represent hydrogen bonds; residues in green are in the *strand* state. Hydrogen bonds are allowed to form when neighbouring residues are in a *strand* state and the side chains are oriented in the same parallel direction.
doi:10.1371/journal.pone.0085185.g005

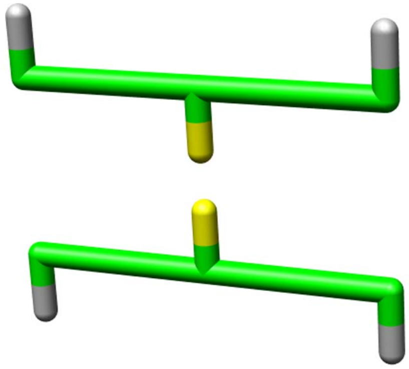

Figure 6. Facing side chains that interact. The yellow residues interact due to their orientation: they are directed towards each other. doi:10.1371/journal.pone.0085185.g006

Steric hindrance penalty. To prevent consecutive side chains being oriented in the same direction, a steric hindrance energy term is used. In real protein structures such conformations are blocked as a result of steric hindrance from clashes between side chain and backbone atoms; such detail is not included in this model. To prevent these conformations being present, we use:

$$E_{\mathrm{steric}} = \sum_{i}^{N} \epsilon_s \cdot S_i \tag{7}$$

Here ϵ_s is the energy penalty for steric hindrance and S_i indicates whether or not residue i is in a state that causes steric hindrance:

$$S_i = \begin{cases} 1 & \text{if } \hat{d}_{i-1} = \hat{d}_i \quad \text{or} \quad \hat{d}_{i+1} = \hat{d}_i \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Solvent interactions. Interactions between the solvent, mimicked by vacant lattice sites, and a given residue depend on the particular amino acid type, the direction of the side chain and the position of the residues with respect to the solvent. The total solvation energy, E_{solvent} is given by:

$$E_{\mathrm{solvent}} = \frac{1}{2} \sum_{\mathrm{solv}}^{N_{\mathrm{solv}}} \sum_{i}^{N} M_{a_i,\mathrm{solv}} \cdot K_{i,\mathrm{solv}} \tag{9}$$

Here $M_{a_i,\mathrm{solv}}$ is the column of the interaction matrix that gives the interaction strength between the solvent and residue type a_i (see below); $K_{i,\mathrm{solv}}$ indicates whether or not an interaction between the solvent and residue i occurs. The residue must in contact with a solvent site and its side chain directed towards the solvent (empty lattice site), thus:

$$K_{i,\mathrm{solv}} = \begin{cases} 1 & \text{if } (\vec{p}_i + \hat{d}_i) = \vec{p}_{\mathrm{solv}} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

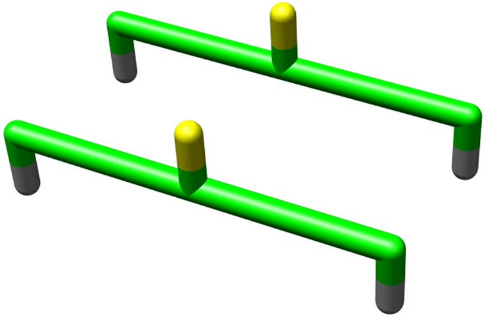

Figure 7. Parallel side chains that interact. The yellow residues interact, since they point in the same direction in a parallel fashion. doi:10.1371/journal.pone.0085185.g007

Parameter values. Note that for real proteins the precise contributions towards the stability of a protein for backbone hydrogen bonding compared to side chain interaction energies is still a topic of discussion. Nevertheless, the side chain interactions, including hydrophobic effects, appear to be the dominant forces behind protein folding [31]. The current parametrisation of this model, as given in Table 1, is in agreement with this statement.

Interaction matrix. The pairwise interaction strengths (M_{a_i,a_j}) between amino acids and between amino acids and the solvent are defined as in ref. [17]. The explicit row of interactions between the solvent and amino acids can be rationalised in the current model - as side chains can explicitly orient towards the solvent. In addition this potential has been shown to prevent unrealistic aggregation behaviour for proteins in their native states.

Monte Carlo simulation

Classic Monte Carlo algorithm. To simulate configurations of the model presented here, we use a Monte Carlo simulation algorithm. Trial steps are accepted according to:

$$P_{acc} = \min\left\{1, \exp\left(\frac{-\Delta E}{k_B T}\right)\right\} \tag{11}$$

where T is the simulation temperature, k_b is the Boltzmann constant and $-\Delta E$ is the difference in energy between the new and old configuration of the system. Trial moves are either internal moves, changing the configuration of a chain (end move, corner flip, crank shaft, point rotation), or rigid body moves, changing the position of the chain relative to other objects (rotation, translation), see ref. [13,14] for more details. In addition, local moves to change the states and side chain directions are performed (see below).

Table 1. List of parameter values for the model.

parameter	value	
ϵ_{hb}	−50	hydrogen bonding energy
ϵ_s	55	steric hindrance energy
E_{state}	0	strand/coil energy
$M_{\mathrm{solv,solv}}$	0	solvent self interaction

doi:10.1371/journal.pone.0085185.t001

At each iteration a single local trial move and a global trial move (including point rotations) with the probability $P_{global} = 0.1$ are performed.

Moves between strand and coil states. The state of each residue may be altered by a local move from *strand* to *coil*, and vice versa. The transition from *coil* to *strand* is only allowed when the following criteria are satisfied:

1. There is no turn in the backbone at residue i
2. Side chains of sequential neighbour are oriented into the opposite direction, if the neighbouring residues are in the *strand* state

Note that the potential energy of hydrogen bonds is taken into account when making the Monte Carlo move to change the state of a residue, since any move is accepted according to the criterion defined in Eqn. 11. Residues in the *strand* state are not permitted to change their backbone configuration or their side chain direction. Hence the *strand* state will be entropically unfavourable; this may, however, during the simulation be compensated by an enthalpic contribution from hydrogen bonding.

Moves of side chain direction. The side chain direction is altered during the simulation with local moves. In such a move, a random new direction is chosen for the side chain, provided that it does not overlap with the direction of the backbone. Each move is accepted or rejected according to the criterion in eqn. 11.

Simulation setup. The volume of the simulation box was kept constant at 80 × 80 × 80 lattice points for the folding simulations and at 30 × 30 × 30 lattice points for the small fibril structure simulations.

Parallel tempering, or temperature replica exchange, is used to converge more rapidly to sampling of equilibrium configurations for the folding simulations. Multiple simulations at different temperatures are run in parallel, while attempting to swap temperatures every 50000 moves with 10000 trial temperature swaps in each simulation. A trial swap between the temperatures of two replicas is accepted with a probability [32–34]:

$$P_{acc} = \min\left\{1, \exp\left(\frac{\Delta E \cdot \Delta 1/T}{k_B}\right)\right\} \tag{12}$$

A grand canonical Monte Carlo simulation is performed to investigate the growth of the fibril seeds at a constant (low) osmotic pressure, see ref. [17] for further details.

Sequence design

One of the challenges when using a cubic lattice model is to design a sequence that will fold into a predefined structure. Previously, it has been shown that one can obtain good folding sequences by minimising the potential energy of the folded state, while keeping the sequence variance high [14]. Here we follow a similar approach, but use a different function to determine the sequence variance.

We can keep the sequence variability high by keeping the distribution of amino acid types close to that observed in nature; here we use the same set of experimental protein structures as used in Ref. [17] to obtain this distribution. First we define a distance d between the distribution of amino acids types in the experimental set and the distribution in the sequence that is to be designed:

$$d = \sum_{a=0}^{a=20} (r_a - s_a)^2 \tag{13}$$

Here r_a denotes the fraction of amino acids with type a in reference to the total number in the experimental set; s_a denotes the fraction of amino acids with type a in the sequence that is to be designed. The sum is over all 20 types of amino acids. The distance defined above needs to be kept small, to design a sequence with a wide variety of amino acids.

A suitable acceptance criterion is given by:

$$P_{\mathrm{acc}} = \min\{1, e^{-q(d_{new} - d_{old})}\} \tag{14}$$

where q is a constant that sets the strength of the biasing potential. This acceptance rule is used in addition to the acceptance rule for the potential energy of the sequence, as in ref. [14].

If we change amino acid i for amino acid j, then $d_{new} - d_{old}$ simply becomes.

$$d_{new} - d_{old} = \frac{2}{N^2}\{(\frac{p_i}{N} - n_i) - (\frac{p_j}{N} - n_j - 1)\} \tag{15}$$

where n_i and n_j are the number of amino acids of type i and j, respectively, before the change.

Note that this approach is as effective in designing folding sequences, as the previously described variance rule, but it gives sequence compositions that are closer to those observed in nature.

Sequences and structures

To generate the results in Figure 2 a designed sequence, see procedure above, and a random sequence of 35 residues with a similar amino acid content were simulated (the designed sequence reads TLSINDYGESEPFKVAVCELQNDDIHIKSLRPARCG and the random sequence PEAMIGPLTGAIHFKVSTSNWGREDLEDVYRQANLI).

For Figure 4 ten peptides consisting of seven residues were used with two different sequences (TTTTTTT and TFTFTFT); the simulations were started from a fibrilar configuration that was formed by simulating ten TFTFTFT peptides with long simulations at a low temperature. The sequences and structures used in this work may be found as PDB files at http://www.few.vu.nl/~abeln/hb-lattice.

Acknowledgments

We would like to thank Erik van Dijk for helpful discussions.

Availability

The source code and interaction potentials are available at: http://www.few.vu.nl/~abeln/hb-lattice

Author Contributions

Conceived and designed the experiments: SA MV CMD DF. Performed the experiments: SA. Analyzed the data: SA DF. Wrote the paper: SA MV CMD DF.

References

1. Dobson CM (2003) Protein folding and misfolding. Nature 426: 884–890.
2. Nguyen HD, Hall CK (2004) Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. Proc Natl Acad Sci U S A 101: 16180–16185.
3. 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.
4. Combe N, Frenkel D (2007) Simple off-lattice model to study the folding and aggregation of peptides. Mol Phys375385.
5. Tsigelny IF, Bar-On P, Sharikov Y, Crews L, Hashimoto M, et al. (2007) Dynamics of alphasynuclein aggregation and inhibition of pore-like oligomer development by beta-synuclein. FEBS J 274: 1862–77.
6. 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: e1000222.
7. Li MS, Klimov DK, Straub JE, Thirumalai D (2008) Probing the mechanisms of fibril formation using lattice models. J Chem Phys 129: 175101.
8. Matthes D, Gapsys V, de Groot BL (2012) Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations. J Mol Biol 421: 390–416.
9. Thirumalai D, Reddy G, Straub JE (2012) Role of water in protein aggregation and amyloid polymorphism. Acc Chem Res 45: 83–92.
10. Cheon M, Chang I, Mohanty S, Luheshi LM, Dobson CM, et al. (2007) Structural reorganisation and potential toxicity of oligomeric species formed during the assembly of amyloid fibrils. PLoS Comput Biol 3: 1727–1738.
11. Li DW, Mohanty S, Irbäck A, Huo S (2008) Formation and growth of oligomers: a Monte Carlo study of an amyloid tau fragment. PLoS Comput Biol 4: e1000238.
12. Li J, Uversky VN, Fink AL (2001) Effect of familial Parkinson's disease point mutations A30P and A53T on the structural properties, aggregation, and fibrillation of human alpha-synuclein. Biochemistry 40: 11604–11613.
13. Abeln S, Frenkel D (2008) Disordered anks prevent peptide aggregation. PLoS Comput Biol 4: e1000241.
14. Coluzza I, Muller HG, Frenkel D (2003) Designing refoldable model molecules. Phys Rev E Stat Nonlin Soft Matter Phys 68: 46703.
15. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, et al. (2010) Atomic-Level Characterization of the Structural Dynamics of Proteins. Science (80-) 330: 341–346.
16. Coluzza I (2011) A coarse-grained approach to protein design: learning from design to understand folding. PLoS One 6: e20853.
17. Abeln S, Frenkel D (2011) Correction. Biophys J 101: 1014.
18. Ni R, Abeln S, Schor M, Cohen Stuart Ma, Bolhuis PG (2013) Interplay between Folding and Assembly of Fibril-Forming Polypeptides. Phys Rev Lett 111: 058101.
19. Privalov PL, Tiktopulo EI, Venyaminov SY, Griko YV, Makhatadze GI, et al. (1989) Heat capacity and conformation of proteins in the denatured state. J Mol Biol 205: 737–750.
20. Naganathan AN, Sanchez-Ruiz JM, MuÃoz V, Muñoz V (2005) Direct measurement of barrier heights in protein folding. J Am Chem Soc 127: 17970–1.
21. Prabhu NV, Sharp Ka (2005) Heat capacity in proteins. Annu Rev Phys Chem 56: 521–48.
22. Davidson aR, Sauer RT (1994) Folded proteins occur frequently in libraries of random amino acid sequences. Proc Natl Acad Sci USA 91: 2146–50.
23. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, et al. (2001) Sequence complexity of disordered protein. Proteins 42: 38–48.
24. Deechongkit S, Nguyen H, Powers ET, Dawson PE, Gruebele M, et al. (2004) Context-dependent contributions of backbone hydrogen bonding to [beta]-sheet folding energetics. Nature 430: 101–105.
25. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 75: 333–366.
26. 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: 773–8.
27. Takeda T, Klimov DK (2009) Replica exchange simulations of the thermodynamics of Abeta fibril growth. Biophys J 96: 442–52.
28. Kim S, Takeda T, Klimov DK (2010) Mapping conformational ensembles of aβ oligomers in molecular dynamics simulations. Biophys J 99: 1949–58.
29. Sali A, Shakhnovich E, Karplus M (1994) Kinetics of Protein Folding : A Lattice Model Study of the Requirements for Folding to the Native State. J Mol Biol 235: 1614–1638.
30. Shakhnovich EI (1994) Proteins with selected sequences fold into unique native conformation. Phys Rev Lett 72: 3907–3910.
31. Baldwin RL (2007) Energetics of protein folding. J Mol Biol 371: 283–301.
32. Lyubartsev AP, Martsinovski AA, Shevkunov SV, Vorontsov-Velyaminov PN (1992) New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. J Chem Phys 96: 1776–1783.
33. Marinari E, Parisi G (1992) Simulated Tempering: A New Monte Carlo Scheme. Eur Lett 19: 451–458.
34. Geyer CJ, Thompson EA (1995) Annealing Markov Chain Monte Carlo with Applications to Ancestral Inference. J Am Stat Assoc 90: 909–920.