doi:10.1016/j.jmb.2005.03.083 J. Mol. Biol. (2005) 349, 648–658

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Protein Misfolding and Amyloid Formation for the Peptide GNNQQNY from Yeast Prion Protein Sup35: Simulation by Reaction Path Annealing

Jan Lipfert[1], Joel Franklin[2], Fang Wu[3] and Sebastian Doniach[1,3]*

[1]*Department of Physics Stanford University, Stanford CA 94305, USA*

[2]*Center for Space Research Massachusetts Institute of Technology, Cambridge MA 02139, USA*

[3]*Department of Applied Physics Stanford University, Stanford CA 94305, USA*

**Corresponding author*

We study the early steps of amyloid formation of the seven residue peptide GNNQQNY from yeast prion-like protein Sup35 by simulating the random coil to β-sheet and α-helix to β-sheet transition both in the absence and presence of a cross-β amyloid nucleus. The simulation method at atomic resolution employs a new implementation of a Langevin dynamics "reaction path annealing" algorithm. The results indicate that the presence of amyloid-like cross-β-sheet strands both facilitates the transition into the cross-β conformation and substantially lowers the free energy barrier for this transition. This model systems allows us to investigate the energetic and kinetic details of this transition, which is consistent with an auto-catalyzed, nucleation-like mechanism for the formation of β-amyloid. In particular, we find that electrostatic interactions of peptide backbone dipoles contribute significantly to the stability of the β-amyloid state. Furthermore, we find water exclusion and interactions of polar side-chains to be driving forces of amyloid formation: the cross-β conformation is stabilized by burial of polar side-chains and inter-residue hydrogen bonds in the presence of an amyloid-like "seed". These findings are in support of a "dry, polar zipper model" of amyloid formation.

Keywords: amyloid; yeast prion; Onsager–Machlup action; reaction-path annealing; cross-beta sheet

Introduction

Prions or "proteinaceous infectious particles" are the cause of several neurodegenerative diseases, most notably scrapie in sheep, bovine spongiform encephalopathy (BSE) in cattle and Creutzfeldt–Jakobs disease (CJD) in humans.[1,2] The disease-causing agent is believed to be the prion protein (PrP), a protein that can undergo a transition from its normal, innocuous form (PrP^{C}) to a misfolded, harmful conformation called PrP^{Sc}. The conformation PrP^{Sc} appears to be the major constituent of the amyloid plaques characteristic of prion diseases.[3–6]

The infectious properties of PrP^{Sc} even across species barriers indicate that the presence of misfolded PrP^{Sc} favors the conversion of uninfected PrP^{C} to PrP^{Sc}; however, the exact pathway is unknown.

Despite the differences in pathogenesis and sequence of the proteins involved, it appears that the amyloid fibrils observed in prion diseases share structural similarities with those associated with other, non-infectious, diseases, such as type II diabetes mellitus and Alzheimer's disease.

Recently, Eisenberg and collaborators[7,8] studied short peptides from the prion-determining domain (PrD; residues 1–123) of Sup35. Sup35 is a "prion-like protein" from yeast and shares the ability to transmit its aberrant fold and to aggregate into amyloid assemblies.[9,10] Its normal cellular role is to terminate translation.[11] Eisenberg *et al.* identified a heptapeptide, GNNQQNY, from Sup35 that displays amyloid formation similar to full-length Sup35, characterized by cooperative kinetics of self-association, fibril formation, binding of the dye Congo red and a cross-β X-ray diffraction pattern. Furthermore, they obtained microcrystals of GNNQQNY and collected powder X-ray diffraction data. In the crystals, the protein strands are in β-sheet configuration with the cross-β spacing

Abbreviations used: CJD, Creutzfeldt–Jakobs disease; PrP, prion protein; PrD, prion-determining domain; MD, molecular dynamics; RC, random coil; SASA, solvent-accessible surface area; MPI, message passing interface.

E-mail address of the corresponding author: doniach@drizzle.stanford.edu

between strands of ≈4.8 Å, observed in most amyloid fibrils, including fibrils from prion protein fragments,[12] and a spacing of ≈10 Å between adjacent layers.

It has been shown in mammalian PrP, that the PrP^{C} to PrP^{Sc} transition involves conversion of α-helix to β-sheet of large parts of the protein.[13,14] The purpose of the present study is to investigate the transition to the β-sheet amyloid state and to study in atomic detail the early steps of amyloid formation in the model system GNNQQNY in order to obtain quantitative information about the energetics and kinetics of amyloid formation. Self-association into amyloid assemblies is a slow process, occurring on time-scales of hours to days. In contrast, the time-scales accessible to molecular dynamics (MD) simulations are typically of the order of nanoseconds. Recent simulations have therefore employed either simplified models,[15] including lattice models[16–20] and the use of implicit solvent[21] or were focused on the stability of the amyloid state alone.[22] Here, we employ a different approach to overcome the time-scale problem: instead of attempting to solve an initial value problem as conventional MD does, we formulate the problem as a boundary value problem with known endpoints and study the most likely trajectories connecting the endpoints using a reaction path formalism based on the Onsager–Machlup action,[23] similar to the work by Eastman *et al.*[24] and Elber *et al.*[25] We simulate the transitions from random coil (RC) to β-sheet and from α-helix to β-sheet both for a single protein strand (Figure 1) in solution and for a protein strand in the presence of two fixed β-sheets (Figure 2). Finally, we investigate the α→β transition for a GNNQQNY strand in the presence of a fixed layer of β-sheets (Figure 10). All simulations employ explicit water as the solvent representation. We find that the presence of pre-formed β-sheet template stabilizes the cross-β structure. Our calculations also provide estimates of the kinetic barriers of these transitions.

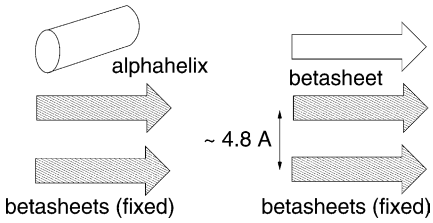

Figure 2. Schematic endpoints of the self-association simulation: two β-sheet strands as amyloid template with one α-helix (left) and all-β configuration as found in amyloid fibrils (right).[7] For the RC to β-sheet simulations, the α-helix is to be replaced by a RC.

Results

Simulations

We use high-temperature molecular dynamics simulations to generate an ensemble of unordered, RC conformations of GNNQQNY, of which we select 20 distinct structures as starting points for the RC to β-sheet simulations. For the simulations of the α-helix to β-sheet conversions for three different systems (single strand in solution, single strand in the presence of two fixed β-sheets and single strand in the presence of a layer of four β-sheets) a canonical α-helix geometry was employed as the starting point. The cross-β structures for the end-points of each of these systems are obtained from our fit to the X-ray powder pattern obtained by Eisenberg and co-workers.[7] All simulations were done using explicit TIP3P water molecules as solvent and employed the CHARMM22[26] force field with no bond or angle constraints. After equilibration of the endpoints, initial trajectories were generated and subsequently "annealed", using a Langevin dynamics protocol, which employs the gradient of the action as an effective force, as described by Eastman *et al.*[24] For the case of the RC→β transition, we obtain 20 independent trajectories in the presence and a further 20 in the absence of two fixed β-sheets. For each of the three α→β systems, ten independent trajectories were generated and annealed. For further details on the model generation and simulation protocol see Materials and Methods. We monitor the secondary structure by computing the peptide back-bone Ψ and Φ angles as well as the cosine of the angle of C═O groups of subsequent residues as defined by Kabsch & Sander.[27]

Presence of β-like amyloid alters the energy landscape

The frames or "snapshots" of our reaction path

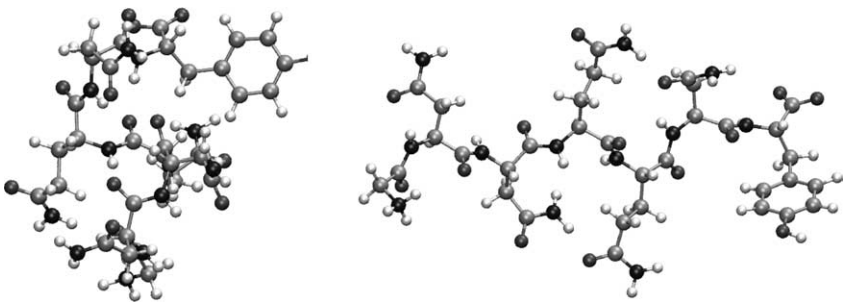

Figure 1. Single strand of GNNQQNY in α-helical (left) and β-sheet (right) conformation. Molecular graphics have been prepared with VMD.[48]

formalism define an effective reaction coordinate (see Materials and Methods). We compute the potential energy for each of these frames along the trajectory both for the RC→β-sheet and α→β transitions of a single strand in solution and of these transitions in the presence of two fixed β-strands. The potential energy as defined by the CHARMM22 potential for each time-frame is averaged over a bootstrapped ensemble of 1000 trajectories that were generated from the 20 (in the case of RC→β) or ten (α→β) simulated trajectories, separately for the presence and absence of two fixed β-sheets (Figure 3). The potential energies of the endpoints are estimated from 100 ps conventional MD runs.

For a single GNNQQNY peptide strand in solution, the α-helix conformation is thermodynamically more stable than the β-sheet by about 5–8 kcal/mol, the thermodynamically most stable state is that of an RC, in agreement with experimental observations.[7] During the α→β transition (Figure 3(a), broken line), the system first relaxes from the α-helical to a compact coil state, marked by an initial decrease in energy. The transition to the β-sheet final state features a large energy barrier of about 25–35 kcal/mol. These findings are in agreement with the simulations for a single strand that start in an RC state (Figure 3(b), broken line). Here the transition to β-sheet is practically completely "uphill", the overall reaction is energetically unfavorable by about 20 kcal/mol and features an activation barrier. No initial relaxation is observed, which is to be expected, as for a single strand an RC conformation is energetically most stable. Keeping in mind the shortcomings of MD force fields and the necessarily imperfect sampling of trajectory space, these energies are only estimates, for which the errors were estimated from the bootstrap ensemble to be about 10 kcal/mol. Furthermore, entropic effects that are likely to play an important role in the observed self-association phenomenon are not well represented in an MD-like framework due to restrictions on the sampling of conformational substates. Despite these limitations, our results clearly indicate that for a single GNNQQNY strand in solution the thermodynamically most stable state is an RC configuration and that the transition toward an extended β-sheet is precluded by a large activation barrier.

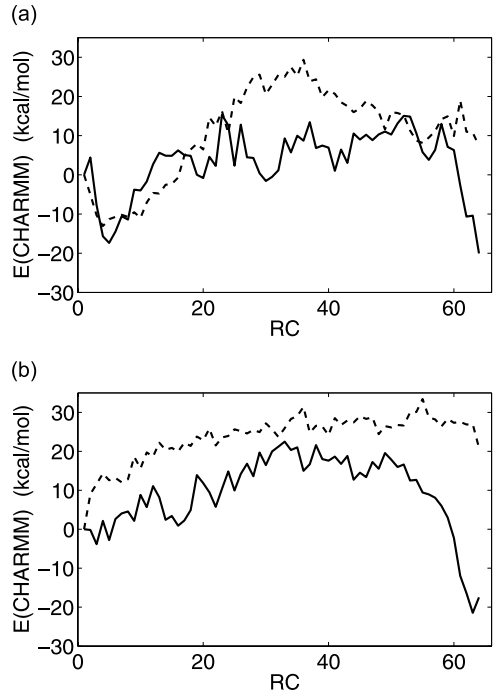

Figure 3. Relative molecular energy along the α→β (a) and RC→β-sheet (b) reaction coordinate averaged over 10 (a) and 20 (b) trajectories each using a bootstrap algorithm. Single peptide strand (broken lines) and reaction in the presence of two fixed β-sheets (continuous lines). Units are kcal/mol, the error is approximately 10 kcal/mol as estimated from the standard deviation of the bootstrap ensemble.

The presence of two template β-sheets as an amyloid "seed" alters this energy landscape dramatically: now the all-β configuration is energetically most stable, the overall transition becomes exothermic, and the energy barrier is substantially lowered, by about 10–15 kcal/mol. This is the case for both the transitions starting in an α-helix (Figure 3(a), continuous line) and those starting from a random coil (Figure 3(b), continuous line). The difference between the two is again an initial relaxation for the trajectories that start from α-helical conformation into a compact RC, which is absent for the simulations starting in an RC state. In the following, we address the question of what are the factors that contribute to the stabilizing effect of a pre-formed β-sheet template for the conversion into a cross-β amyloid state.

Peptide backbone dipoles stabilize the cross-β configuration

It is well known that the peptide bond has a permanent electrical dipole moment of about 3.5 D.[28] There are fundamentally two energetically favorable ways to position dipoles, depicted in Figure 4.

Both the generic α-helix and β-sheet configurations have favorable peptide bond dipole interactions, the former being "aligned/tip-to-toe" for residues that are separated by more than four positions along the sequence and the latter being "anti-aligned". It has been proposed† (Thomas Garel, personal communication) that the overall arrangement of peptide bond dipoles favors a β-like amyloid state. Using the geometries from the reaction path annealing MD simulation (Figure 5), we compute the electrostatic energy of the peptide backbone dipoles, employing the Poisson–Boltzmann solver implementation DelPhi.[29,30] Volume occupied by protein atoms is assigned a dielectric constant $\epsilon_{prot}=2$ and the aqueous solvent

† http:arxiv.org/abs/cond-mat/0305053

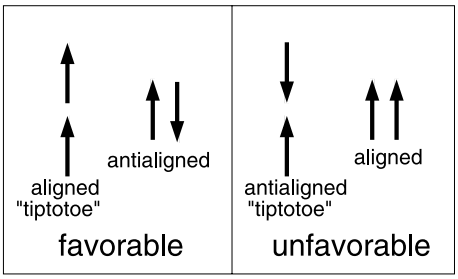

Figure 4. Energetically favorable and unfavorable arrangement of dipoles.

is modeled with $\epsilon_{H_2O} = 80$, therefore allowing for the calculation of the energetic contribution of electrostatic interaction of peptide bond dipoles alone, with appropriate boundary conditions.

The results (Figure 6) show that even for the single peptide the backbone dipole arrangement is slightly most favorable in the β conformation (Figure 6, broken lines). This is plausible: in the RC there is no particular dipole order and there are only sporadic interactions. We expected the single β-sheet to have the slightly more stable backbone dipole conformation, even with respect to the α-helix, as the seven residues of GNNQQNY correspond to only about two helical turns. For the amyloid-like stack, however, the all-β conformation of the dipoles is strongly favored by more than 30 kcal/mol. In the presence of pre-formed β-sheets, each dipole interacts with its counterparts in strands below and above in an aligned/tip-to-toe fashion, in addition to the favorable anti-aligned configuration of dipoles along the β-strands, it is this influence that results in the additional favorable interaction energy.

The role of water conformational entropy

The crystallographic data suggest that the β-amyloid formed by GNNQQNY is a compact,

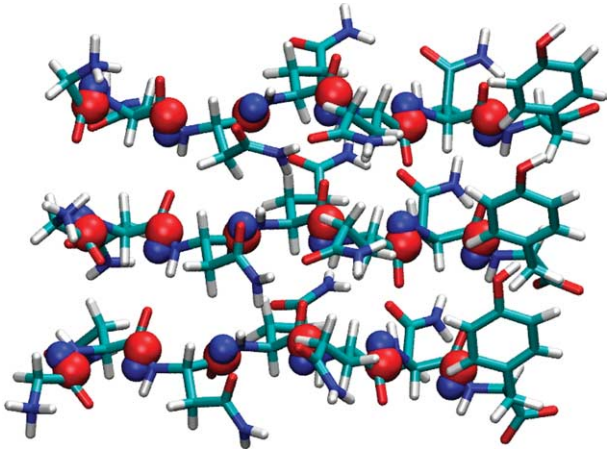

Figure 5. Three GNNQQNY strands in all-β-sheet geometry with peptide backbone dipoles denoted by red and blue spheres.

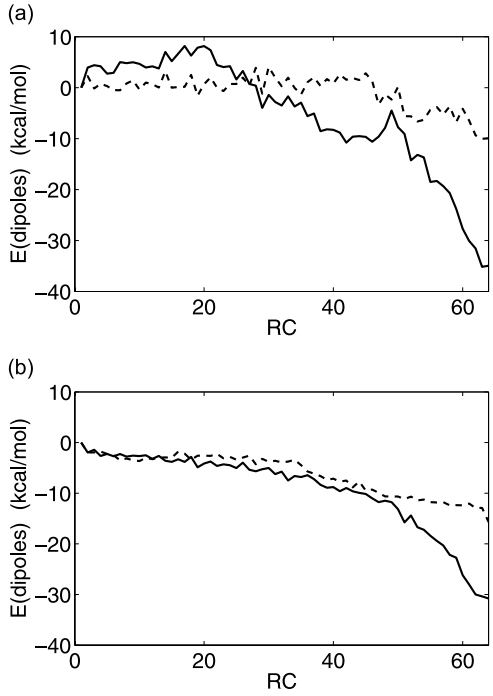

Figure 6. Electrostatic energies due to peptide backbone dipoles as a function of α→β (a) and RC→β (b) reaction coordinate in kcal/mol. Energies are averaged over (a) 10 or (b) 20 trajectories. Single peptide strand (broken lines) and three stack (full lines).

anhydrous state.[7,8] Release of water into the bulk causes a gain in conformational entropy, which has been suggested to contribute to the stability of β-amyloid. To quantify the effect, we compute the solvent-accessible surface area (SASA) and monitor the number of protein–protein and protein–water hydrogen bonds as a function of reaction coordinate (Figures 7 and 8). For a single strand in solution the transition to a β-helical state exposes surface area, both for the α→β and for the RC→β transition. Quantitatively, the effect can be understood by noting that the conformations in the RC ensemble are fairly compact; however, not as compact as an α-helix. In contrast, the same transitions in the presence of two "pre-formed" β-sheets have the net effect of reducing the solvent-accessible protein surface. The effect is bigger for the RC to β-sheet transition, as the RC is the more extended starting configuration. Translated into a free energy, the reduction of SASA stabilizes the β-amyloid state by about 7 kcal/mol.

Further evidence for the "drying up" of the protein system during amyloid formation comes from monitoring hydrogen bonds during the transition: towards the end of the transition, the number of protein–water hydrogen bonds, decreases sharply, marked by a simultaneous increase in protein–protein hydrogen bonds, as shown in Figure 8. This general behavior is observed irrespective of the starting point of the simulations, for

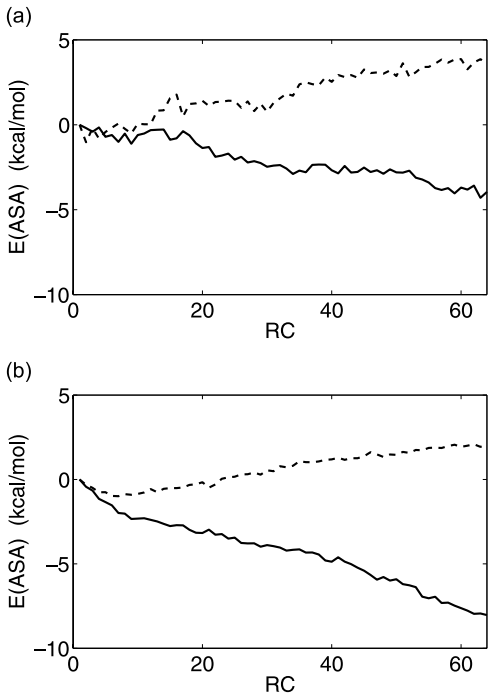

Figure 7. Solvent-accessible surface area contribution to the free energy as a function of (a) α→β and (b) RC→β reaction coordinate in kcal/mol. Energies are averaged over (a) 10 or (b) 20 trajectories. Single strand (broken line) and three peptide stack (full line). For conversion from surface area to free energy the factor 0.025 kcal/(mol Å^2) was employed.[39]

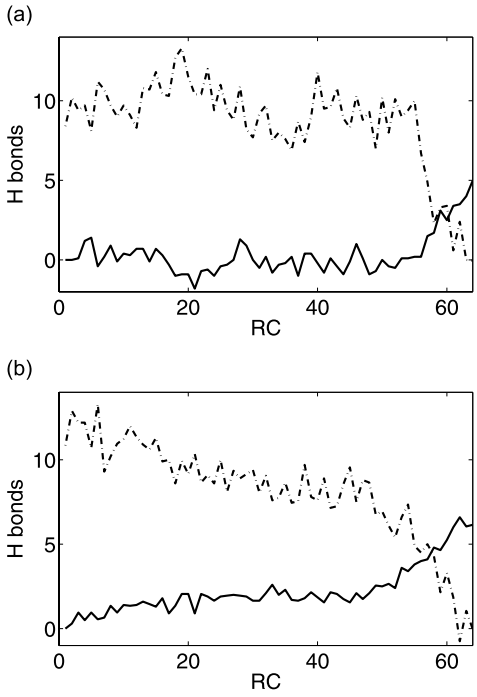

Figure 8. Number of protein–water (dash-dot line) and protein–protein (continuous line) hydrogen bonds as a function of (a) α→β and (b) RC→β reaction coordinate. Curves have been shifted by a constant offset to overlay them in one graph. Hydrogen bonds were identified using the distance-angle criterion of McDonald & Thornton.[49]

the trajectories starting in RCs the transition is slightly more gradual.

Glutamine and asparagine act as a "polar zipper"

Glutamine and asparagine-rich regions feature in several proteins involved in neurodegenerative diseases and amyloid-like aggregation,[31] for example, Huntingtin in Huntington disease. Perutz *et al.*[32–34] introduced the notion that poly (Gln) and poly (Asn) stretches may act as "polar zippers", which stabilize β-sheet amyloid-like aggregates by forming interstrand hydrogen bonds both along the backbone and between polar side-chains of adjacent strands. We find evidence for this hypothesis in our simulations: The side-chains of asparagine and glutamine indeed form interstrand hydrogen bonds. In particular, the asparagine residues at positions 2 and 3, and to a lesser extent the asparagine at position 6, hydrogen bond increasingly to their counterparts in the adjacent template strand as the reaction proceeds. It is interesting to observe that the two glutamine residues at positions 4 and 5 do not exhibit as strong a tendency for interstrand hydrogen bonds as asparagine does: the glutamine residues tend to form intrastrand hydrogen bonds, mostly between Asn3 and Gln5 and between Gln4 and Asn6 on opposing sides of the same strand. This behavior occurs independent of the starting point of the simulation. While intrastrand hydrogen bonds do not contribute directly and enthalpically to the stability of the amyloid stacking, they still do so entropically by liberating water into the bulk that would otherwise satisfy these hydrogen bonds. Thus one might call the observed arrangement of polar side-chains a "dry, polar zipper". In this picture, the asparagine residues at positions 3 and 6 play a crucial role, as they are involved in both intra- and interstrand hydrogen bonds, namely to the glutamine residues 2 positions apart along the same strand and to their asparagine counterparts in the adjacent strand. This salient position might make them an interesting target with which to study the effects of mutations (Figure 9).

In order to further investigate the role of side-chain interactions, we compute the pairwise interaction energy of in register pairs of side-chains, i.e. the interaction energy between the Gly residue of the moving top strand with the Gly residue of the adjacent strand below, the interaction of the Asn2 residue of the top strand with the Asn2 of the strand below, etc. The interaction energy for this purpose is defined as the sum of non-bonded energy terms (van der Waals and Coulomb) in the CHARMM22 potential involving pairs of atoms in the respective residues. The results averaged over

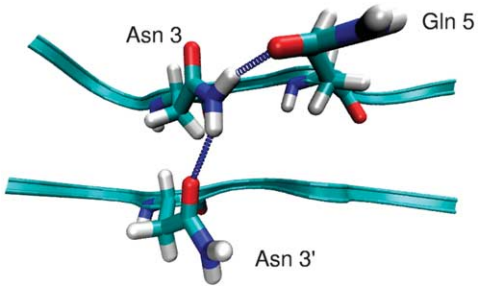

Figure 9. Inter- and intrastrand hydrogen bonds formed by the asparagine residue at position 3. For clarity, only two strands are displayed in ribbon format, residues Asn3 on both strands and Gln5 on the top strand are represented in licorice mode, hydrogen bonds in blue.

the 20 trajectories in the RC → β ensemble are shown in Figure 10.

The results show that the polar zipper residues interact favorably, increasingly so during the second half of the trajectories. The effect is most pronounced for the Asn residues in positions 2 and 3, which were already identified as particularly important in the analysis of the interstrand hydrogen bonding. It needs to be emphasized that the pair interaction energies contain, in part, energetic contributions that were previously discussed, such as backbone dipole interactions and hydrogen bonding interactions. The pair interaction energies are an alternative way of dissecting and analyzing enthalpic contributions to amyloid stability. From Figure 10 it can also be seen that the N and C-terminal residues interact strongly unfavorably, which could be expected, as they carry a positive and negative charge from their terminal amine and carboxyl groups, respectively. It has been proposed that amyloid stability results, in part, from π-stacking of aromatic residues.[35] We can address this question by computing the pair interaction

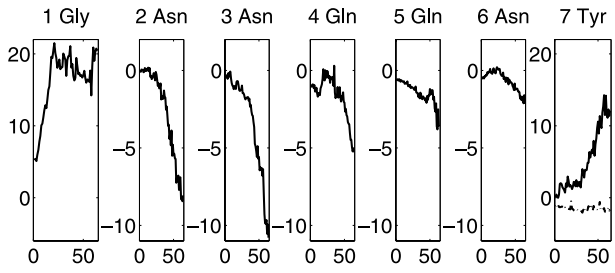

Figure 10. Residue pair interaction energies as a function of reaction coordinate, defined as the sum of van der Waals and Coulomb terms for pairs of residues. The graph labeled "1 Gly" describes the interaction of the glycine residue of the moving top strand with the glycine residue of the strand directly adjacent to it, similarly for the other residues. The broken line in the graph for tyrosine is the interaction energy excluding protein backbone atoms. Energies are in kcal/mol, note the different scale for residues Asn2 through Asn6 and for the N and C termini.

energy of the tyrosine residues in the moving top strand with that of the adjacent strand in the preformed β-amyloid. In these calculations, the backbone atoms of Tyr need to be excluded, as otherwise the energy is dominated by the unfavorable interaction of the charged C-terminal groups. The results (broken line in the graph for Tyr7 in Figure 10) indicate that the interaction of the aromatic tyrosine side-chains is indeed attractive throughout the entire trajectory. However, the interaction energy is only 1–2 kcal/mol, much weaker than some of the other enthalpic contributions. Even bearing in mind that the energy calculations based on the CHARMM22 potential disregard entropic effects (like changes in water exposed surface area) and are only an imperfect approximation to the intrinsically quantum mechanical π-electron interactions, the above results suggest that while aromatic stacking might contribute to the stability of the amyloid state, other factors like polar side-chain interactions and backbone dipole alignment play a more important role for the peptide GNNQQNY.

Amyloid lateral organization

A strong meridional peak in the diffraction pattern at about 4.8 Å spacing is the hallmark of protein cross-β structure and has been observed in many amyloid-forming systems. It corresponds to the stacking of β-sheets along the fiber axis. Much less is known about the molecular arrangement in the plane perpendicular to the fiber axis. While a β-helix organization has been proposed for the Gln and Asn-rich N-terminal domain of Sup35,[36] the peptide GNNQQNY is very unlikely to adopt this conformation: the seven residue peptide is too short to complete a full turn in a β-helix, furthermore, the presence of a peak in the fiber diffraction pattern at about 10 Å indicates that several strands are placed next to each other, 10 Å being the spacing resulting from side-chains extended in the plane. A characteristic peak at 10 Å spacing is also present in the diffraction pattern obtained from the crystals, and fits to the experimental data indicate that in the crystal, sheets of stacked β-sheets are packed such that each sheet is adjacent to one with the β-strands running in the same direction and one with the β-sheets pointing in the opposite direction, the spacing between the sheets being 10 Å (see Supplementary Data Figure 14).

We also investigated the effect of interactions in the plane perpendicular to the fiber axis by performing simulations of the α → β transition of one GNNQQNY strand, but now in the presence of a unit cell plane of fixed β-sheets (see Figure 11). Due to the considerably larger size of the system, energy fluctuations were more significant and a potential energy profile similar to Figure 3 difficult to obtain. However, the overall reaction proceeds similarly to the case of the transition in the presence of a two peptide stack of β-sheets. Hydrogen bonding in the amyloid conformation now also

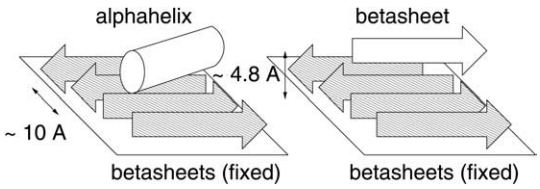

Figure 11. Endpoints of α→β simulation in the presence of a unit cell plane of fixed β-sheets.

involves strands adjacent in the plane. The effect of reducing the SASA is even more pronounced, stabilizing the amyloid conformation by more than 10 kcal/mol (see Supplementary Data Figure 12). Looking at the electrostatic energy of the peptide backbone dipoles, two effects now compete: the alignment along the strand and with the strand below is energetically favorable as before, but the interaction with the dipoles of the strands to the left and right is unfavorable. As the characteristic distance to the strands to the left and right in the plane is larger than the one to the strands below (or above), the overall effect is still favorable, by about 15 kcal/mol (see Supplementary Data Figure 13).

It must be emphasized, though, that both microcrystals and fibers of Sup35 and derived peptides show considerable structural diversity[8,37] and further experimental and theoretical study is necessary to elucidate the lateral organization at atomic resolution.

Discussion

The reaction path annealing method described here allows for an estimate of the energetic stabilization of the cross-β amyloid structure, and the various physical mechanisms contributing to its stability. The overall potential energy viewed along the RC→β or α→β reaction coordinate clearly supports an autocatalytic mechanism of amyloid formation, whereby pre-formed β-sheets help the prion-like conversion and addition of more peptide into the amyloid structure. In order to estimate the overall free energy of this amyloid aggregation event, different energetic contributions need to be combined. The free energy is given by $\Delta G = \Delta H - T\Delta S$, where ΔH and ΔS are the changes in enthalpy and entropy and T is the physical temperature of the system. The enthalpic contribution can be estimated from the computed CHARMM22 energies (Figure 3) to be about 18 kcal/mol for the coil→β transition. The entropic contribution is more difficult to evaluate, however, we can estimate it as follows: contributions to the change in entropy are the gain in entropy due to the liberation of bound water molecules into the bulk and a loss of entropy from fixing residues into the well defined β-amyloid state from their high-entropy RC conformation. The former can be estimated by computing the solvent accessible surface area (Figure 7), the latter by assigning the change RC→β-sheet a entropic cost of 4.2 cal/(K residue). The latter value is an average estimate derived from protein folding experiments.[38] Adding these contributions, the overall change in free energy is about 15 kcal/mol, which is in agreement with the experimentally observed exceptional stability of β-amyloid fibrils.

Our simulation results allow us to compare pathways originating in RC to those starting in α-helical conformations. Our findings strongly suggest that α-helix rich conformations are off-pathway for the amyloid aggregation of GNNQQNY. None of the trajectories starting in RC states develops appreciable α-helix content, whereas trajectories starting in an α-helical state quickly relax into RC conformations before proceeding to the β-sheet final state. This is in marked contrast to the findings of recent simulations of the Alzheimer Aβ peptide fragment KLVFFAE ($A\beta_{16-22}$) by Klimov & Thirumalai,[39] who find that α-helical intermediates are on-pathway to the β-amyloid state. It needs to be borne in mind, though, that despite their similar size and amyloidogenic properties, KLVFFAE and GNNQQNY have quite different characteristics, the former having a hydrophobic core and charged residues at the ends that favor antiparallel arrangement, whereas GNNQQNY is marked by polar side-chains acting as a polar zipper and favors parallel assembly.

Caflisch and co-workers recently studied the amyloid formation of GNNQQNY using molecular dynamics simulations with implicit solvent[21] and replica exchange dynamics.[40] Our results agree qualitatively with their findings. They, too, find the all-β-amyloid state to be enthalpically most favorable and see enthalpic barriers on the pathway to this state. Their results, similar to ours, suggest a crucial role for side-chain interactions in the formation of the stable parallel β-strands. However, in their chosen free energy reaction coordinates, they find the parallel β-strand conformation to be either marginally stable (as compared to the thermal energy) or unstable.

As to the comparison of "direct" molecular dynamics simulations to "reaction path" based methods like our present approach, the following can be said: both computational strategies suffer from the problem that they can sample only a very small subset of all physically realizable trajectories. Even in order to do so, they rely on approximations like implicit solvent or harmonic constrains in the case of MD or the discretization of trajectories into frames or "snapshots", as employed in this study. One of the disadvantages of reaction path methods is the necessity to specify both endpoints of the simulations, and therefore to study only "reactive" trajectories. However, there are important advantages as well: the reaction path formulation makes the problem inherently parallel, well suited for the use of massive parallel supercomputers, as opposed to fundamentally sequential standard MD simulations. Furthermore, the ability to focus on reactive trajectories allows only, at least in principle, for the

study of much larger systems, that are far beyond the reach of conventional molecular dynamics, in the near future. For a small system like the peptide GNNQQNY, the different approaches are complementary, in that their different approximations and strategies allows for a comparison and validation of the results.

In conclusion, our simulations of the peptide GNNQQNY support the view of amyloid as an anhydrous state, stabilized by the conformational entropy of water released into the bulk. An enthalpic contribution to the amyloid stability stems from hydrogen bonds formed between the polar side-chains of asparagine and glutamine residues. Taken together we might call this a dry, polar zipper. In our analysis we draw attention to the role of electrostatically interacting peptide backbone dipoles. Evidently, this is not the only interaction at play; however, aligning backbone dipoles provides a partial explanation of amyloid stability without depending much on sequence details, which is attractive in light of the observation that amyloid formation has been reported in a variety of pathological conditions involving proteins that share no apparent sequence similarities. Furthermore, it is in support of the notion introduced by Dobson and others[41–44] that amyloid formation might be a generic and relatively universal phenomenon of protein and peptide chemistry under partly denaturing conditions, which might be driven mainly by sequence unspecific main-chain, backbone interactions.[45]

Our method also allows for estimates of the energetic barrier to amyloid formation. This gives the potential to infer the effects of mutations, e.g. Gln or Asn→Ala, on the kinetics of the α→β conversion, which could be tested experimentally.

Materials and Methods

Model amyloid-like system

By fitting to the X-ray powder pattern of GNNQQNY microcrystals obtained by Eisenberg *et al.*, we have been able to generate a number of candidate atomic models of the amyloid-like unit cell. More details on the fitting procedure are given in Supplementary Data. The in-register, parallel β-sheet configuration is that of the model unit cell proposed by Eisenberg *et al.*[7,8] and in agreement with a recent MD study by Gsponer *et al.*[21]

This structure was equilibrated for 10 ps in the case of the trajectories starting in an α-helical state and for 100 ps in the case of trajectories starting in an RC state. The length of the equilibration did not have any appreciable influence on the energetics or structure. The final structures of five independent equilibration runs of the β-amyloid structure were taken as models of the diseased, amyloid form of the peptide, the endpoint of our reaction path simulations. We study five different systems: the transition from α-helix to β-sheet of a single strand in solution (Figure 1), of a single strand in the presence of two fixed β-sheets (Figure 2) and of a strand in the presence of a fixed plane of four β-sheets (Figure 10). Similarly, we study the transition from RC to β-sheet in the presence and absence of two fixed β strands. For the α-helical starting point, we generated a conformation of the short peptide GNNQQNY, which has a canonical α-helix geometry, equilibrated for 10 ps. The random coil starting points were prepared by running ten copies of a single GNNQQNY strand first for 100 ps at 300 K followed by a denaturation phase of 500 ps at 400 K. Finally the system was cooled back to 300 K over the course of 500 ps. The RC conformation selected as starting points for the reaction path trajectories had an RMSD with respect to the starting structure of at least 6 Å and no appreciable secondary structure. GNNQQNY in solution is unlikely to be strictly α-helical and its soluble form has in fact been observed to be in a coiled configuration,[7] a behavior that is confirmed by the results of the simulations. However, running both simulations with an α-helical structure and an RC as starting points allows for comparison of both pathways. Starting with an α-helix has the advantage of starting in a well defined initial state and to allow for the study of the α to β transition, which seems to be of crucial importance in full-length prion proteins.[13,14] For the simulations featuring the RC to β or α to β transformation in the presence of a fixed stack or plane of β-sheets, the corresponding structure is placed on top of the pre-equilibrated assembly of β-strands and the system is equilibrated for 1 ps in the case of the α-helical starting point and for 100 ps in the case of the RC starting point. In all subsequent simulations, only the top strand and the solvent molecules are allowed to move.

Simulation method

The Langevin equation:

$$M\frac{\mathrm{d}^2X}{\mathrm{d}t^2} = F - \gamma M\frac{\mathrm{d}X}{\mathrm{d}t} + \sqrt{2k_\mathrm{B}TM\gamma}\zeta \qquad (1)$$

has been used extensively for MD simulations and can be thought of as an extension of Newton's equation for a system immersed in a heat bath.[46] X and F are the position and force vector of all atomic coordinates, M the mass matrix, γ the damping constant, T and k_B the physical temperature of the system and the Boltzmann constant, and ζ Gaussian $N(0,1)$ white noise. If one is interested in the dynamics of a protein on time-scales of picoseconds, the inertial term (the left-hand side) can be neglected, which yields the overdamped Langevin equation. Taking the overdamped Langevin equation as the underlying equation of motion for our system, the reaction path method by Onsager & Machlup[23] can be employed, in which given two configurations of a protein, e.g. A and B, the probability density that a certain trajectory connects the two endpoints A and B is proportional to:

$$P_{\mathrm{A}\to\mathrm{B}} \propto \mathrm{e}^{-S/k_\mathrm{B}T} \qquad (2)$$

Here S is the Onsager–Machlup action, which reads in discretized form:[47]

$$S_{\mathrm{OM}} = \sum_{i=1}^{K}\frac{\Delta t}{4M\gamma}\left(\gamma M\frac{X_i - X_{i-1}}{\Delta t} - F(X_{i-1})\right)^2 \qquad (3)$$

The sum over i is over K time "frames" along the trajectory, Δt is the physical time between these snapshots that describe the protein's motion between the fixed endpoints. This formulation of the action follows from the discretization according to the Itô calculus, in which case the Jacobian is a constant, independent of path coordinates. If the Stratonovich calculus is chosen, the Jacobian depends on the path coordinates and gives rise to an additional term proportional to ∇F in the action. Both

approaches constitute valid approximations, in the following, however, we use the Itô formulation for computational simplicity. The most probable trajectories are those that minimize the action S_{OM}. The computational challenge is therefore to sample trajectory space and drive the system toward a minimum of the action, corresponding to more probable trajectories.

For the generation of initial trajectories and for the sampling of trajectory space we use algorithms similar to those proposed by Eastman *et al*.[24] The atomic positions on each of the time frames are evolved according to a "pseudo-dynamical" equation of motion:

$$m_i \frac{d^2 x_{ij}}{d\tau^2} = -\frac{dS}{dx_{ij}} - \gamma m_i \frac{dx_{ij}}{d\tau} + \sqrt{2 m_i k_B T_\tau \gamma}\zeta \qquad (4)$$

x_{ij} denotes the ith coordinate of the jth time slice, the "pseudo-force" $-dS/dX_{ij}$ is the negative gradient of the action functional, which couples the time frames to their neighbors along the trajectory. The "pseudo-time" τ has no connection to the physical time t, it is merely a parameter that marks the evolution of the system during the minimization of the action. T_τ denotes a "pseudo-temperature" which sets the magnitude of the random term, ζ again is Gaussian white noise. Starting at a large pseudo-temperature and reducing it in the course of the simulation, we employ a simulated annealing protocol to drive the system toward minimal action. All our simulations employ an all-atom representation and the CHARMM22[26] force field with no bond or angle constraints. We use the TIP3P water model as a solvent representation. The solvent is kept locally equilibrated in every frame by subjecting water molecules to conventional MD instead of the "action-forces" (equation (5)) that couple different frames along the trajectory. Starting from five distinct initial trajectories for each of the three $\alpha \rightarrow \beta$ systems, we perform two simulation runs for each trajectory, using two different values for the physical time between the snapshots $\Delta t = 0.5$ ps and 5 ps. These choices give rise to similar energetics and dynamics, so that we performed all calculations on the RC$\rightarrow\beta$ trajectories at $\Delta t = 0.5$ ps. The explicit form of the gradient of the action or pseudo-force is (where i and k index the particle coordinates and j is the frame index):

$$\frac{dS}{dx_{i,j}} = \frac{m_i}{2\Delta t}(2x_{i,j} - x_{i,j-1} - x_{i,j+1}) + \frac{1}{2}(F_{i,j} - F_{i,j-1}) - \frac{1}{2}\sum_{k}^{3N}(x_{i,j+1} - x_{i,j})\frac{\partial F_{k,j}}{\partial x_{i,j}} + \sum_{k}^{3N}\frac{\Delta t}{2m_k}F_{k,j}\frac{\partial F_{k,j}}{\partial x_{k,j}} \qquad (5)$$

The last term (corresponding to the F^2 term in the action) scales as $\propto \Delta t$, whereas the first term in equation (5) (corresponding to the velocity squared term in the action) scales as $\propto (1/\Delta t)$. For numerical stability it is desirable to choose Δt such that these terms are roughly balanced. A very small Δt corresponds to the situation where the system is forced to reach its final state unphysically fast, causing a large action that is dominated by the velocity term and numerical instabilities. However, a very large Δt causes instabilities and a large action as well, as in this regime the action is dominated entirely by the force term and the different time frames along the trajectory become decoupled. By performing simulations with a range of values for Δt (data not shown) we found our present choices $\Delta t = 0.5$ and 5 ps to be reasonable compromises between these extremes. The simulations employ a new implementation of our massively parallel reaction path annealing protocol, where each time frame runs on a separate CPU: the processes exchange information using the message passing interface (MPI) standard, exploiting the fact that the pseudo-force (5) couples only neighboring time frames along the trajectory, which greatly reduces the associated message passing. For the number of frames along the trajectory $K = 64$ was chosen, which represents a compromise between computational expense and detail in the sampling of the trajectories. We run the simulations for 25,000 steps (in the case of $\Delta t = 0.5$ ps) and 50,000 (for $\Delta t =$ 5 ps), respectively. During the course of the simulations, the action decreases by about an order of magnitude from its initial value and remains approximately constant for the last 10,000 steps. All annealing runs were performed on the National Energy Research Scientific Computing Center IBM SP RS/6000 supercomputer.

Acknowledgements

The authors thank Wenjun Zheng for help with the model generation, the Pasteur Institute for their hospitality, the National Energy Research Scientific Computing Center (NERSC) for CPU time, and the NSF and NIH (grant 5R01-NS40132) for financial support.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.03.083

References

1. Prusiner, S. B. (1998). Prions. *Proc. Natl Acad. Sci. USA*, **95**, 13363–13383.
2. Prusiner, S. B. (1999). *Prion Biology and Disease*, Cold Spring Harbor Laboratory Press, New York.
3. Harper, J. D. & Lansbury, P. T. (1997). Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of time-dependent solubility of amyloid proteins. *Annu. Rev. Biochem.* **66**, 385–407.
4. Cohen, F. E. & Prusiner, S. B. (1998). Pathologic conformations of prion proteins. *Annu. Rev. Biochem.* **67**, 793–819.
5. Prusiner, S. B., Scott, M. R., DeArmond, S. J. & Cohen, F. E. (1998). Prion protein biology. *Cell*, **93**, 337–348.
6. Aguzzi, A., Montrasio, F. & Kaeser, P. S. (1998). Prions: health scare and biological challenge. *Cell*, **93**, 337–348.
7. Balbirnie, M., Grothe, R. & Eisenberg, D. S. (2001). An amyloid-forming peptide from the yeast prion sup35 reveals a dehydrated β-sheet structure for amyloid. *Proc. Natl Acad. Sci. USA*, **98**, 2375–2380.
8. Diaz-Avalos, R., Long, C., Fontano, E., Balbirnie, M., Grothe, R., Eisenberg, D. & Caspar, D. L. D. (2003). Cross-beta order and diversity in nanocrystals of an amyloid-forming peptide. *J. Mol. Biol.* **330**, 1165–1175.
9. Wickner, R. B. (1994). [URE3] as an altered URE2 protein: evidence for a prion analog in saccharomyces cerevisiae. *Science*, **264**, 566–569.
10. Patino, M. M., Liu, J. J., Glover, J. R. & Lindquist,

S. (1996). Support for the prion hypothesis for inheritance of a phenotypic trait in yeast. *Science*, **273**, 622–626.
11. Stansfield, I. & Tuite, M. F. (1994). Polypeptide chain termination in *Saccharomyces cerevisiae*. *Curr. Genet*. **25**, 385–395.
12. Nguyen, J. T., Inouye, H., Baldwin, M. A., Fletterick, R. J., Cohen, F. E., Prusiner, S. B. & Kirschner, D. A. (1995). X-ray diffraction of scrapie prion rods and PrP peptides. *J. Mol. Biol.* **252**, 412–422.
13. Pan, K. M., Baldwin, M., Nguyen, J., Gasset, M., Serban, A., Groth, D. *et al*. (1993). Conversion of α-helices into β-sheets features in the formation of the scrapie prion proteins. *Proc. Natl Acad. Sci. USA*, **90**, 10962–10966.
14. Baskakov, I. V., Legname, G., Prusiner, S. B. & Cohen, F. E. (2001). Folding of prion protein to its native α-helical conformation is under kinetic control. *J. Biol. Chem.* **276**, 19687–19690.
15. Smith, A. V. & Hall, C. K. (2001). Protein refolding *versus* aggregation: computer simulations on an intermediate-resolution protein model. *J. Mol. Biol.* **312**, 187–202.
16. Broglia, R. A., Tiana, G., Pasquali, S., Roman, H. E. & Vigezzi, E. (1998). Folding and aggregation of designed proteins. *Proc. Natl Acad. Sci. USA*, **95**, 12930–12933.
17. Harrison, P. M., Chan, H. S., Prusiner, S. B. & Cohen, F. E. (1999). Thermodynamics of model prions and its implications for the problem of prion protein folding. *J. Mol. Biol.* **286**, 593–606.
18. Giugliarelli, G., Micheletti, C., Banavar, J. R. & Maritan, A. (2000). Compactness, aggregation, and prionlike behaviour of protein: a lattice model study. *J. Chem. Phys.* **113**, 5072–5077.
19. Harrison, P. M., Chan, H. S., Prusiner, S. B. & Cohen, F. E. (2001). Conformational propagation with prion-like characteristics in a simple model of protein folding. *Protein Sci.* **10**, 819–835.
20. Dima, R. I. & Thirumalai, D. (2002). Exploring protein aggregation and self-propagation using lattice models: phase diagram and kinetics. *Protein Sci.* **11**, 1036–1049.
21. Gsponer, J., Haberthaüur, U. & Caflisch, A. (2003). The role of side-chain interactions in the early steps of aggregation: molecular dynamics simulation of an amyloid-forming peptide from yeast prion sup35. *Proc. Natl Acad. Sci. USA*, **100**, 5154–5159.
22. Ma, B. & Nussinov, R. (2002). Stabilities and conformations of alzheimer's β-amyloid peptide oligomers ($A\beta_{16-22}$, $A\beta_{16-35}$ and $A\beta_{10-35}$). *Proc. Natl Acad. Sci. USA*, **99**, 14126–14131.
23. Onsager, L. & Machlup, S. (1953). Fluctuations and irreversible processes. *Phys. Rev.* **91**, 1505–1512.
24. Eastman, P., Gronbech-Jensen, N. & Doniach, S. (2001). Simulation of protein folding by reaction path annealing. *J. Chem. Phys.* **114**, 3812–3841.
25. Olender, R. & Elber, R. (1996). Calculation of classical trajectories with a very large time step: formalism and numerical examples. *J. Chem. Phys.* **105**, 9299–9315.
26. MacKerell, A. D., Jr, Brooks, B., Brooks, C. L., III, Nilsson, L., Roux, B., Won, Y. & Karplus, M. (1998). Charmm: the energy function and its parameterization with an overview of the program. *Encyclopedia Comput. Chem.* **1**, 271–277.
27. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
28. Creighton, T. E. (1993). *Proteins*, W.H. Freeman and Company, New York.
29. Rocchia, W., Alexov, E. & Honig, B. (2001). Extending the applicability of the nonlinear Poisson–Boltzmann equation: multiple dieletric constants and multivalent ions. *J. Phys. Chem. B*, **105**, 6507–6514.
30. Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A. & Honig, B. (2002). Rapid grid-based construction of the molecular surface for both molecules and geometric objects: applications to the finite difference Poisson–Boltzmann method. *J. Comput. Chem.* **23**, 128–137.
31. Michelitsch, M. D. & Weissman, J. S. (2000). A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc. Natl Acad. Sci. USA*, **97**, 11910–11915.
32. Perutz, M. F., Johnson, T., Suzuki, M. & Finch, J. T. (1994). Glutamine repeats as polar zippers: theirs possible role in inherited neurodegenerative diseases. *Proc. Natl Acad. Sci. USA*, **91**, 5355–5358.
33. Stott, K., Blackburn, J. M., Butler, P. J. G. & Perutz, M. F. (1995). Incorporation of glutamine repeats makes protein oligomerize: implications for neuro-degenrative diseases. *Proc. Natl Acad. Sci. USA*, **92**, 6509–6513.
34. Perutz, M. F., Pope, B. J., Owen, D., Wanker, E. E. & Scherzinger, E. (2002). Aggregation of proteins with expanded glutamine and alanine repeats of the glutamine-rich and asparagine-rich domains of sup35 and of the amyloid β-peptide of amyloid plaques. *Proc. Natl Acad. Sci. USA*, **99**, 5596–5600.
35. Gazit, E. (2002). A possible role for π-stacking in the self-assembly of amyloid fibrils. *FASEB J.* **16**, 77–83.
36. Perutz, M. F., Finch, J. T., Berriman, J. & Lesk, A. (2002). Amyloid fibers are water-filled nanotubes. *Proc. Natl Acad. Sci. USA*, **99**, 5591–5595.
37. DePace, A. H. & Weissman, J. S. (2002). Origins and kinetic consequences of diversity in sup35 yeast prion fibers. *Nature Struct. Biol.* **9**, 389–396.
38. Daune, M. (1993). *Molecular Biophysics*, Oxford University Press, Oxford, UK.
39. Klimov, D. K. & Thirumalai, D. (2003). Dissecting the assembly of $A\beta_{16-22}$ amyloid peptides into anti-parallel β sheets. *Structure*, **11**, 295–307.
40. Cecchini, M., Rao, F., Seeber, M. & Caisch, A. (2004). Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *J. Chem. Phys.* **121**, 10748–10757.
41. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M. (1998). Amyloid fibril formation by an SH3 domain. *Proc. Natl Acad. Sci. USA*, **95**, 4224–4228.
42. Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G. & Dobson, C. M. (1999). Designing conditions for *in vitro* formation of amyloid protofilaments and fibrils. *Proc. Natl Acad. Sci. USA*, **96**, 3590–3594.
43. Ramirez-Alvarado, M., Merkel, J. S. & Regan, L. (2000). A systematic exploration if the influence of the protein stability on amyloid fibril formation *in vitro*. *Proc. Natl Acad. Sci. USA*, **97**, 8979–8984.
44. Fändrich, V., Forge, M., Buder, K., Kittler, M., Dobson, C. M. & Diekmann, S. (2003). Myoglobin forms amyloid fibrils by association of unfolded polypeptide segments. *Proc. Natl Acad. Sci. USA*, **100**, 15463–15468.
45. Fändrich, M. & Dobson, C. M. (2002). The behaviour

of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *EMBO J.* **21**, 5682–5690.
46. Allen, M. & Tildesley, D. (1987). *Computer Simulation of Liquids*, Oxford University Press, Oxford, UK.
47. Risken, H. (1989). *The Fokker–Planck Equation*, Springer, Berlin, Heidelberg.
48. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38.
49. McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potentials in proteins. *J. Mol. Biol.* **238**, 777–793.

Edited by P. T. Lansbury Jr

(Received 8 September 2004; received in revised form 24 January 2005; accepted 27 March 2005)