Available online at www.sciencedirect.com

SciVerse ScienceDirect

Physics Procedia 34 (2012) 49 – 54

Physics Procedia

Monte Carlo studies of protein aggregation

Sigurður Ægir Jónsson[a], Iskra Staneva[a], Sandipan Mohanty[b], Anders Irbäck[a,*]

[a]*Computational Biology and Biological Physics, Department of Astronomy and Theoretical Physics, Lund University Sölvegatan 14A, SE - 223 62 Lund, Sweden*

[b]*Institute for Advanced Simulation, Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52425 Jülich, Germany*

Abstract

The disease-linked amyloid β (Aβ) and α-synuclein (αS) proteins are both fibril-forming and natively unfolded in free monomeric form. Here, we discuss two recent studies, where we used extensive implicit solvent all-atom Monte Carlo (MC) simulations to elucidate the conformational ensembles sampled by these proteins. For αS, we somewhat unexpectedly observed two distinct phases, separated by a clear free-energy barrier. The presence of the barrier makes αS, with 140 residues, a challenge to simulate. By using a two-step simulation procedure based on flat-histogram techniques, it was possible to alleviate this problem. The barrier may in part explain why fibril formation is much slower for αS than it is for Aβ.

Keywords: protein misfolding, protein aggregation, amyloid

1. Introduction

The mechanisms by which proteins misfold and aggregate into oligomers and fibrils are currently the subject of intense research [1]. This research aims, in particular, at improving our understanding of the molecular basis of protein-misfolding diseases, such as Alzheimer's and Parkinson's diseases. Knowledge of the underlying mechanisms is further a prerequisite for rational development of strategies to prevent protein aggregation [2].

Computer simulations are an indispensable complement to experiments in unraveling the complex processes of protein misfolding and aggregation. Different computational approaches are being used, at varying levels of detail. Coarse-grained models are useful, for instance, for studying generic features of fibril formation [3]. Such studies have, in particular, provided insights into the factors governing fibril nucleation; fibril formation typically show sigmoidal kinetics with a lag phase, indicating a nucleation-dependent process. The precise nature of the aggregation process, including the rate at which it occurs, varies widely from protein to protein. Studies of specific proteins are typically based on atomically detailed models, with or without explicit solvent [4]. This approach is computationally challenging, but has been used to investigate early events in the aggregation process for some small proteins, like the Aβ protein associated with Alzheimer's disease [5, 6]. Atomic-level simulations were further used to study the incorporation of monomers into growing fibrils [7] as well as the mode of action of aggregation-inhibiting small molecules [8].

Using MC techniques, we recently studied the natively unfolded monomeric forms of two fibril-forming and disease-linked proteins, namely Aβ [5], with 42 residues, and the 140-residue αS protein associated with Parkinson's

* *Email address:* anders@thep.lu.se (Anders Irbäck)

doi:10.1016/j.phpro.2012.05.008

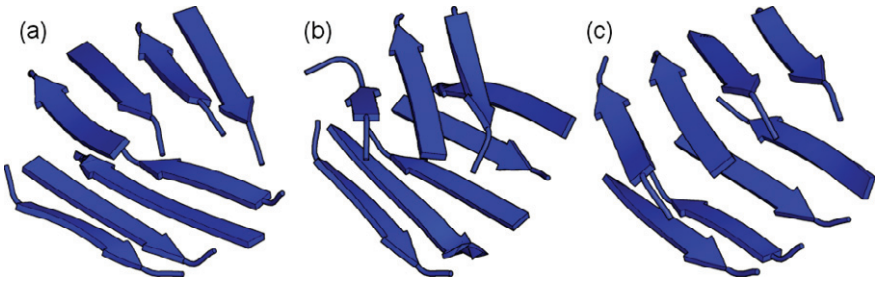

Figure 1: Snapshots showing typical low-energy conformations for the system of 8 GIIFNEQ peptides. The structures share an overall β-sandwich topology, but differ in the organization of the strands. Drawn with PyMOL [23].

disease [9]. For Aβ, we also studied dimer formation [6]. The present article summarizes and compares the findings of these studies.

Both proteins were studied using the same implicit solvent all-atom force field [10, 11], but the simulation methods were somewhat different. For Aβ, we used the convenient simulated tempering method, which is similar in spirit and performance to the widely used replica exchange or parallel tempering method. αS was more challenging to study, in part because of its larger size, but also because of a clear two-phase behavior. To facilitate crossing the rather large free-energy barrier separating these phases, we changed scheme from simulated tempering to a two-step procedure based on flat-histogram techniques [12].

2. Model and algorithms

All simulations discussed in this article are based on the same implicit solvent all-atom model, with torsional degrees of freedom [10, 11, 13]. In short, the potential is composed of four terms, $E = E_{\rm loc} + E_{\rm ev} + E_{\rm hb} + E_{\rm sc}$. The first term, $E_{\rm loc}$, contains local interactions between atoms separated by only a few covalent bonds. The other three terms are non-local in character: $E_{\rm ev}$ represents excluded-volume effects, $E_{\rm hb}$ is a hydrogen-bond potential, and $E_{\rm sc}$ describes residue-specific interactions, based on hydrophobicity and charge, between pairs of side chains.

This potential was developed through folding thermodynamics studies of a structurally diverse set of peptides and small proteins, while deliberately keeping it as simple as possible [11]. It is worth stressing that the same set of parameters is used for α, β, as well as mixed α/β proteins. The model has previously been applied to study the folding of peptides and small proteins [11, 14] as well as the aggregation of small peptides [15–18].

Our studies are carried out using MC methods. Three elementary conformational moves are employed: (i) pivot-type rotations about individual backbone bonds, (ii) a semi-local backbone update, Biased Gaussian Steps, which rotates up to eight consecutive angles simultaneously [19], and (iii) rotations of individual side-chain angles. For multi-chain systems, we also include rigid-body translations and rotations of whole chains relative to each other. This model and these algorithms are implemented in the open source package PROFASI [20].

As indicated above, to speed up our αS simulations, we used flat-histogram techniques. The procedure was first tested on a system consisting of 8 copies of a 7-residue fragment, GIIFNEQ, of the superoxide dismutase 1 protein [12]. In our simulations of this system, two distinct phases (aggregated/non-aggregated) occurred, separated by a clear free-energy barrier. Figure 1 illustrates the aggregated phase by three representative snapshots from the simulations. The fact that the aggregated structures are β-sheet-rich is consistent with results from the aggregation prediction program Waltz [21], which suggests this part of the SOD1 protein to be particularly prone to form amyloid structure [22].

The employed flat-histogram procedure assumes that the system of interest undergoes a first-order-like phase transition, at which the energy distribution is bimodal. It amounts to first constructing and then simulating a generalized ensemble defined by the microstate probability distribution

$$P_\nu = 1/\gamma(E_\nu) \propto \begin{cases} 1/g(E_\nu) & \text{if } E_1 \le E \le E_2 \\ \exp(-E_\nu/RT_{\rm m}) & \text{otherwise} \end{cases} \tag{1}$$

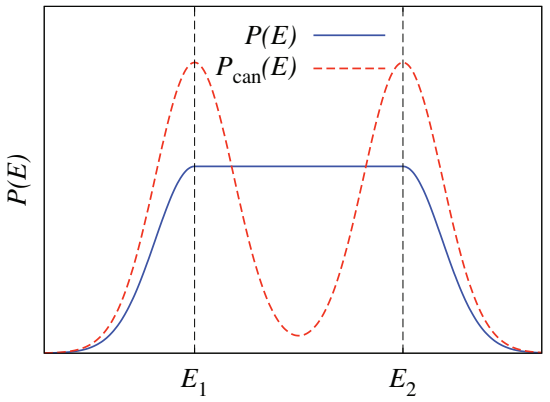

Figure 2: Schematic illustration of the simulated ensemble. It is assumed that the system of interest displays a first-order-like phase transition, at which the energy distribution, $P_{\mathrm{can}}(E)$, is bimodal. The peaks are centered at E_1 and E_2. The energy distribution for the simulated ensemble, $P(E)$, is flat in the coexistence region $E_1 < E < E_2$, while $P(E) \propto P_{\mathrm{can}}(E)$ outside this range.

where $g(E)$ is the density of states, and E_1 and E_2 denote the positions of the energy distribution peaks at the midpoint temperature T_{m} (R is the gas constant). Figure 2 shows the corresponding energy distribution, $P(E)$, which is flat between E_1 and E_2 in order to facilitate transitions between the low- and high-energy phases. Outside this range, $P(E)$ falls off like the canonical energy distribution. The exact shape of the tails of $P(E)$ is unimportant. The main point is to avoid unnecessary sampling of the bottom part of the energy landscape, which may consist of narrow minima that are difficult to sample and not necessarily low in free energy. Otherwise, one risks slowing down the simulations through a time-consuming exploration of narrow minima with a negligible occupancy at biologically relevant temperatures.

The first step of the procedure is to estimate the function $\gamma(E)$, which is accomplished by means of the Wang-Landau method [24, 25]. The second step is to simulate the ensemble $P_\nu = 1/\gamma(E_\nu)$ by standard techniques, which corresponds to a multicanonical production run [26, 27]. Finally, properties of the original system are recovered by reweighting methods [28].

Figure 3 shows the MC time evolution of the energy in typical runs with this method for the system of 8 GGIINEQ peptides. For comparison, a constant-temperature simulation is shown as well. The transition frequency between the aggregated and non-aggregated phases is low in the conventional canonical simulation. The mobility of the system is improved in the $1/\gamma$ ensemble.

In our runs, this system displayed a strong and to us unanticipated preference for α- over β-structure for large values of the Wang-Landau modification factor f. The formation of β-structure, which requires the establishment of long-range contacts, seemed to be hampered by the bias away from the last visited region present at large f. This property might be general and show up in simulations of other protein systems as well. Because of this property, we found it advantageous to start the simulations with a relatively small initial f ($\ln f = 2^{-10}$ rather than $\ln f = 1$).

Our αS study was performed using the above scheme. In the production runs, transitions between the high- and low-energy phases occurred with comparable frequencies in both directions. Because of this property, we can compare the two coexisting phases in an unbiased manner. This behavior was not achieved in preliminary runs with the replica exchange method. The statistics did not permit, however, a statistically accurate determination of the relative population of the two phases as a function of temperature. Furthermore, we do not expect the temperature scale of current force fields to be perfect. Our αS study therefore focused on single-phase properties, which are statistically easier to determine and depend much more weakly on temperature, compared to the relative weight of the phases.

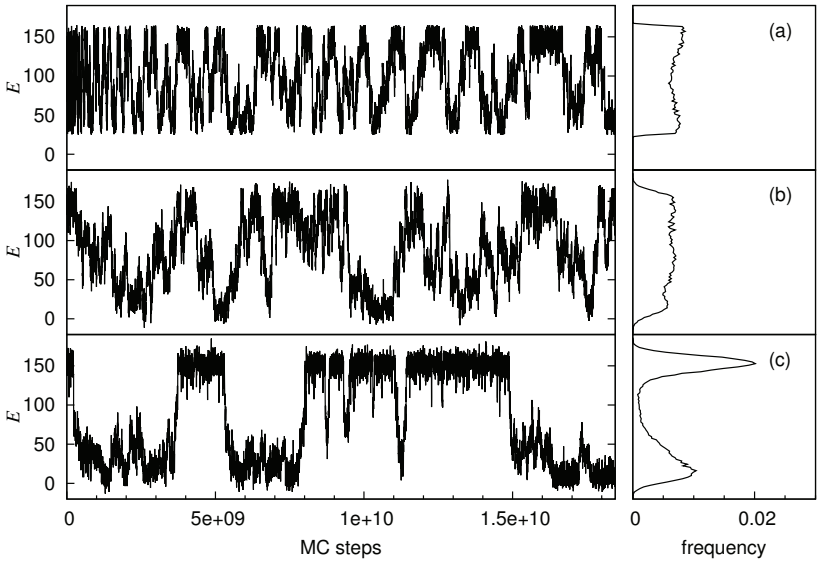

Figure 3: MC time evolution of the energy in typical runs with three different methods for the system of 8 GFIINEQ peptides. The right panel shows histograms of energies visited in the respective runs. (a) Wang-Landau simulation, (b) simulation of the $1/\gamma$ ensemble (see Eq. 1 and Figure 2), and (c) canonical-ensemble simulation at a temperature $T \approx T_{\mathrm{m}}$.

3. Comparing Aβ and αS ensembles

Aβ occurs in two main forms, Aβ40 and Aβ42, with 40 and 42 residues, of which Aβ42 is most strongly linked to Alzheimer's disease. In fibrils, each Aβ molecule participates in two intermolecular face-to-face packed β-sheets [29, 30]. The two β-regions are connected by a loop at residues ∼25–30.

In our study, we investigated Aβ42 monomers [5] and dimers [6]. We studied the wild-type protein as well as the mutants E22G, E22G/I31E and F20E. These variants were deliberately chosen so as to have very different aggregation properties [31]. The E22G mutation, associated with a familial early-onset form of Alzheimer's disease, is known to enhance aggregation, whereas the F20E mutation has the reverse effect. The double mutant E22G/I31E shows more complex aggregation properties. Its propensity to form fibrils is almost as high as that of the E22G variant, whereas its propensity to form prefibrillar species is only slightly higher than that of F20E.

These Aβ42 variants all displayed large conformational variability and similar overall properties in the simulations. Typical conformations were compact and had a significant β-strand content. Differences among the variants could also be seen, especially in the fibril loop region. Interestingly, the propensity with which this part of the molecule formed a fibril-like turn was found to correlate with the experimental rate of fibril formation. Moreover, the probability of observing a fibril-like turn increased upon dimer formation.

There are several lines of evidence suggesting that the formation of this loop is a key step in fibrillation. For instance, it was shown that the coupling of residues 23 and 28 through a lactam bridge, to enforce a fibril-like turn, leads to much faster fibril formation [32]. Our findings support this picture, by suggesting that fibrillation-modulating mutations indeed cause conformational changes in this part of the molecule.

While there is recent evidence indicating that αS has a folded helical tetramer as its main physiological form [33], it is well established that free monomeric αS is highly disordered [34]. In αS fibrils, the stretch from residue ~ 35 to ~ 98 is thought to form five strands, which belong to separate (intermolecular) β-sheets that are stacked upon each other [35]. This makes the fibril core structure somewhat reminiscent of that of Aβ fibrils.

In our simulations of the free αS monomer [9], two distinct phases occurred, schematically illustrated in Figure 4. These were a disordered high-energy phase, called D, and a low-energy phase, B. The latter phase had a significant β-strand content, but nevertheless large conformational variability. We analyzed secondary-structure propensities, size

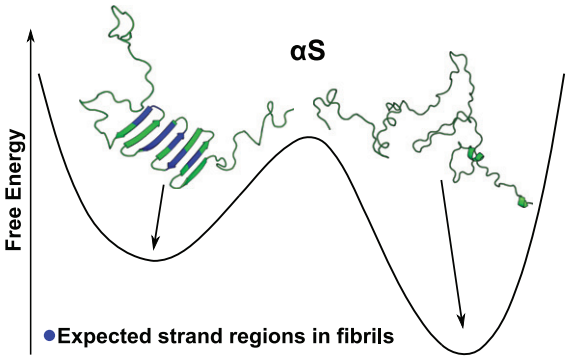

Figure 4: Schematic free-energy landscape for the αS monomer.

and topological properties of these two phases, and compared with existing experimental data, mainly from NMR-based studies.

The properties of the D phase matched very well with the findings of numerous experiments at neutral pH and temperatures typically around 15°C. This agreement shows, in particular, that the B phase can only be weakly populated under such conditions. Presence of the B phase is, by contrast, compatible with data from low-temperature experiments (−15°C and −10°C). These data cannot be explained by either the D or the B phase alone, but are consistent with a mixture of the two phases. The B phase fraction required to match the data could be as high as 50–70 %.

A closer analysis of the B phase reveals that the β-sheets tend to have a simple meander topology, and that the sequence locations of the strands correlate with those in αS fibrils [35]. These two properties imply that the backbone fold resembles that found in the core of fibrils [35]. This similarity does not necessarily mean that a monomer in the B phase can easily get incorporated into a growing fibril, because in-plane intramolecular hydrogen bonds are to be replaced by out-of-plane intermolecular ones. The two-phase behavior has, nevertheless, implications for aggregation.

A comparison of our αS and Aβ results illustrates this point. Both monomers populated fibril-like folds in the simulations. For αS, we observed a two-phase behavior, where only the B phase showed fibril-like features. In the D phase, the monomer must overcome a free-energy barrier to acquire a fibril-like fold. This contrasts sharply with our findings for Aβ, which has only a single phase and no corresponding barrier. These results are perfectly consistent with the experimental fact that aggregation is much slower for αS than it is for Aβ [36, 37].

We find the above results very encouraging and will extend our studies of both Aβ and αS. For instance, it would be very interesting to study the effects of aggregation-inhibiting small molecules on Aβ and αS, and how individual monomers bind to a growing fibril. These problems are computationally feasible but demanding — there is no doubt room for further improvements of both models and algorithms.

[1] F. Chiti, C. M. Dobson, Protein misfolding, functional amyloid and human disease, Annu. Rev. Biochem. 75 (2006) 333–366.
[2] T. Härd, C. Lendel, Inhibition of amyloid formation, J. Mol. Biol.doi:10.1016/j.jmb.2011.12.062.
[3] C. Wu, J.-E. Shea, Coarse-grained models for protein aggregation, Curr. Opin. Struct. Biol. 21 (2011) 209–220.
[4] J. E. Straub, D. Thirumalai, Principles governing oligomer formation in amyloidogenic peptides, Curr. Opin. Struct. Biol. 20 (2010) 187–195.
[5] S. Mitternacht, I. Staneva, T. Härd, A. Irbäck, Comparing the folding free-energy landscapes of Aβ42 variants with different aggregation properties, Proteins 78 (2010) 2600–2608.
[6] S. Mitternacht, I. Staneva, T. Härd, A. Irbäck, Monte Carlo study of the formation and conformational properties of dimers of Aβ42 variants, J. Mol. Biol. 410 (2011) 357–367.
[7] G. Reddy, J. E. Straub, D. Thirumalai, Dynamics of locking of peptides onto growing amyloid fibrils, Proc. Natl. Acad. Sci. USA 106 (2009) 11948–11953.
[8] M. Convertino, A. Vitalis, A. Caflisch, Disordered binding of small molecules to Aβ(12-28), J. Biol. Chem. 286 (48) (2011) 41578–41588.
[9] S. Æ. Jónsson, S. Mohanty, A. Irbäck, Distinct phases of free α-synuclein — a Monte Carlo study, submitted.
[10] A. Irbäck, B. Samuelsson, F. Sjunnesson, S. Wallin, Thermodynamics of α- and β-structure formation in proteins, Biophys. J. 85 (2003) 1466–1473.
[11] A. Irbäck, S. Mitternacht, S. Mohanty, An effective all-atom potential for proteins, PMC Biophys. 2 (2009) 2.
[12] S. Æ. Jónsson, S. Mohanty, A. Irbäck, Accelerating atomic-level protein simulations by flat-histogram techniques, J. Chem. Phys. 135 (2011) 125102.
[13] A. Irbäck, S. Mohanty, Folding thermodynamics of peptides, Biophys. J. 88 (2005) 1560–1569.
[14] S. Mohanty, J. H. Meinke, O. Zimmermann, U. H. E. Hansmann, Simulation of Top7-CFr: a transient helix extension guides folding, Proc. Natl. Acad. Sci. USA 105 (2008) 8004–8007.

[15] G. Favrin, A. Irbäck, S. Mohanty, Oligomerization of amyloid Aβ_{16-22} peptides using hydrogen bonds and hydrophobicity forces, Biophys. J. 87 (2004) 3657.
[16] M. Cheon, I. Chang, S. Mohanty, L. M. Luheshi, C. M. Dobson, M. Vendruscolo, G. Favrin, Structural reorganisation and potential toxicity of oligomeric species formed during the assembly of amyloid fibrils, PLoS Comput. Biol. 3 (2007) e173.
[17] A. Irbäck, S. Mitternacht, Spontaneous β-barrel formation: an all-atom Monte Carlo study of Aβ_{16-22} oligomerization, Proteins 71 (2008) 207–214.
[18] D. Li, S. Mohanty, A. Irbäck, S. Huo, Formation and growth of oligomers: a Monte Carlo study of an amyloid tau fragment, PLoS Comput. Biol. 4 (2008) e1000238.
[19] G. Favrin, A. Irbäck, F. Sjunnesson, Monte Carlo update for chain molecules: Biased Gaussian steps in torsional space, J. Chem. Phys. 114 (2001) 8154–8158.
[20] A. Irbäck, S. Mohanty, PROFASI: a Monte Carlo simulation package for protein folding and aggregation, J. Comput. Chem. 27 (2006) 1548–1555.
[21] S. Maurer-Stroh, M. Debulpaep, N. Kuemmerer, M. Lopez de la Paz, I. C. Martins, J. Reumers, K. L. Morris, A. Copland, L. Serpell, L. Serrano, J. W. H. Schymkowitz, F. Rousseau, Exploring the sequence determinants of amyloid structure using position-specific scoring matrices, Nat. Methods 7 (2010) 237–242.
[22] M. Oliveberg, Waltz, an exciting new move in amyloid prediction, Nat. Methods 7 (2010) 187–188.
[23] W. L. DeLano, The PyMOL molecular graphics system, San Carlos, CA: DeLano Scientific (2002).
[24] F. Wang, D. P. Landau, Efficient, multiple-range random walk algorithm to calculate density of states, Phys. Rev. Lett. 86 (2001) 2050–2053.
[25] F. Wang, D. P. Landau, Determining the density of states for classical statistical models: a random walk algorithm to produce a flat histogram, Phys. Rev. E 64 (2001) 056101.
[26] B. A. Berg, T. Neuhaus, Multicanonical algorithms for first order phase transitions, Phys. Lett. B 267 (2) (1991) 249–253.
[27] U. H. E. Hansmann, Y. Okamoto, Prediction of peptide conformation by multicanonical algorithm: new approach to the multiple-minima problem, J. Comput. Chem. 14 (1993) 1333–1338.
[28] A. M. Ferrenberg, R. H. Swendsen, Optimized Monte Carlo data analysis, Phys. Rev. Lett. 63 (1989) 1195–1198.
[29] A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio, R. Tycko, A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR, Proc. Natl. Acad. Sci. USA 99 (2002) 16742–16747.
[30] T. Lührs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Döbeli, D. Schubert, R. Riek, 3D structure of Alzheimer's amyloid-β(1-42) fibrils, Proc. Natl. Acad. Sci. USA 102 (2005) 17342–17347.
[31] L. M. Luheshi, G. G. Tartaglia, A. Brorsson, A. P. Pawar, I. E. Watson, F. Chiti, M. Vendruscolo, D. A. Lomas, C. M. Dobson, D. C. Crowther, Systematic in vivo analysis of the intrinsic determinants of amyloid β pathogenicity, PLoS Biol. 5 (2007) e290.
[32] K. L. Sciarretta, D. J. Gordon, A. T. Petkova, R. Tycko, S. C. Meredith, Aβ40-Lactam(D23/K28) models a conformation highly favorable for nucleation of amyloid, Biochemistry 44 (2005) 6003–6014.
[33] T. Bartels, J. G. Choi, D. J. Selkoe, α-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation, Nature 477 (2011) 107–111.
[34] P. H. Weinreb, W. Zhen, A. W. Poon, K. A. Conway, P. T. Lansbury, NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded, Biochemistry 35 (1996) 13709–13715.
[35] M. Vilar, H.-T. Chou, T. Lührs, S. K. Maji, D. Riek-Loher, R. Verel, G. Manning, H. Stahlberg, R. Riek, The fold of α-synuclein fibrils, Proc. Natl. Acad. Sci. USA 105 (2008) 8637–8642.
[36] E. Hellstrand, B. Boland, D. M. Walsh, S. Linse, Amyloid β-protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process, ACS Chem. Neurosci. 1 (2010) 13–18.
[37] M. Grey, S. Linse, H. Nilsson, P. Brundin, E. Sparr, Membrane interaction of α-synuclein in different aggregation states, J. Parkinson's Dis. 1 (2011) 359–371.