Finite Size Effects in Simulations of Protein Aggregation

Amol Pawar, Giorgio Favrin*

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

Abstract

It is becoming increasingly clear that the soluble protofibrillar species that proceed amyloid fibril formation are associated with a range of neurodegenerative disorders such as Alzheimer’s and Parkinson diseases. Computer simulations of the processes that lead to the formation of these oligomeric species are starting to make significant contributions to our understanding of the determinants of protein aggregation. We simulate different systems at constant concentration but with a different number of peptides and we study the how the finite number of proteins affects the underlying free energy of the system and therefore the relative stability of the species involved in the process. If not taken into account, this finite size effect can undermine the validity of theoretical predictions regarding the relative stability of the species involved and the rates of conversion from one to the other. We discuss the reasons that give rise to this finite size effect form both a probabilistic and energy fluctuations point of view and also how this problem can be dealt by a finite size scaling analysis.

Introduction

A major field of study in protein science is the understanding of the causes, and implications of protein aggregation. The misfolding of proteins often results in aggregates and in the formation of highly regular structures called amyloid fibrils [1–3]. Amyloids are best known for their involvement in pathological conditions such as type II diabetes systemic amyloidosis and neurodegenerative disorders such as Alzheimer’s, Parkinson’s and Creutzfeldt-Jakob diseases [3–7]. Increasing evidence is showing that indeed it is the soluble pre-fibrillar or oligomeric species, rather than the insoluble amyloid fibrils themselves, which are responsible for toxicity and neuronal disfunction [8–10]. Even if highly complex processes are associated with oligomer toxicity, a view is gaining support according to which the ability to form toxic oligomeric species represents an intrinsic property of polypeptide chains at some stage of their oligomerization process [8,10]. Despite this growing interest in the role of peptide and protein oligomers in disease, the molecular mechanism by which they are formed is still the object of investigation of several in vivo, in vitro, and in silico studies [4,11–19]. However, since it is challenging to describe the early stages of aggregation of polypeptide chains experimentally, primarily because of the difficulties in detecting and characterising the small, structurally heterogeneous and transient species that are involved, a detailed description of this process at the molecular level remains in large part elusive. In particular the heterogeneous nature of these species implies that their free energy landscape is very rugged. Theoretical studies are starting to make important contributions to the understanding of these diseases by investigating the partially unfolded intermediates and the structural features of the oligomeric species [15,16,18,20–24].

Due to limitations in computers power, simulations are usually performed in the NVT ensemble on systems ranging form a couple of peptides with detailed all atom models to several hundred with very coarse grained models [15–35]. We will discuss how these studies can be affected by the finite size effect and how this has repercussions on the underlying free energy of the system. This is indeed a major difference between in silico and in vitro studies, as in the latter the number of proteins is in the order of the Avogadro number. This effect has been studied in the case of proteins of different lengths [36]. In the following we will discuss in detail the finite size effect specifically in the case of simulations of protein aggregation and we will underline some strategies to deal with it.

Results

Free Energy Calculations in Systems with Different Sizes

We study how the actual number of proteins or peptides in our simulation, irrespectively of their concentration, affects the underlying free energy landscape of the system. In particular we investigate six systems of three, four, six, eight, nine and twelve “Gly-Phe-Phe” (GFF) peptides at constant concentration and for eight different temperatures using the parallel tempering technique and with periodic boundary conditions (see Methods). A similar peptide (Phe-Phe), was studied experimentally by the group of Gazit [37] and it was found to form nanotube-like structures. We choose however to add a Gly to the peptide to increase its propensity to form inter-peptide hydrogen bonding.

For each system we calculate the free energy as a function of the inter-chain hydrogen bond interaction energy normalised over the number of peptides for each different temperature used in the parallel tempering. In this way we can monitor the formation and stability of the aggregated state alone. Our computationally efficient Monte Carlo sampling allow us to calculate easily the free energy landscape for systems up to twelve GFF peptides. In Fig 1, Fig 2, and Fig 3 we plot such landscapes and we can notice how in each system the aggregated state becomes more stable as the temperature decreases and how by increasing system size the...
aggregates become stable for increasingly higher temperatures. The influence of the number of particles on the free energy of the system for a given concentration can be explained by the following example. If we consider a system where all \( N \) proteins are aggregated (e.g., \( N = 3 \) in Fig 4, left) the probability of another protein to be added to the oligomer is zero simply because there are no monomers left. For this given set of conditions (temperature and concentration), therefore, the proteins in the aggregated state

---

**Figure 1.** Free Energy landscape as a function of inter chain hydrogen bond energy for a system of 3 peptides for temperatures ranging from \( T = 0.588 \) to \( T = 0.322 \). In the lower right corner we show a characteristic oligomeric configuration at \( T = 0.322 \).

**Figure 2.** Free Energy landscape as a function of inter chain hydrogen bond energy for a system of 6 peptides for temperatures ranging from \( T = 0.588 \) to \( T = 0.322 \). In the lower right corner we show a characteristic oligomeric configuration at \( T = 0.322 \). By comparing the free energy profile for \( T = 0.4 \) in the present and in Fig 1, we can notice the increase in stability of the aggregated phase that results by simply increasing the number of peptide in the system and the volume of the system so that the concentration remains constant.
have just a probability of becoming monomers. If the number of proteins and the volume of the system are increased so that the concentration is unchanged (Fig 5 right) the probability of the proteins in the oligomers to go back in solution is the same as before, because there is one more monomer left in the solution. As a result the stability of the oligomer of size $N$ is increased. In other words, by increasing the number of proteins for a given concentration and temperature, the free energy barrier between the monomeric and the aggregated phases also increases and the stability of the aggregates increases as we can see for example from comparing the free energy at $T = 0.4$ in Figs 1 and 2.

Another reason for this change in stability is related to the change in the relative size of the fluctuations of the energy of the system. Being the energy an extensive quantity its expectation value (mean) is proportional to the number of particles in the system while its root mean square fluctuations (standard deviation) are proportional only to the square root of that number.

$$\frac{\sqrt{\langle E^2 \rangle - \langle E \rangle^2}}{\langle E \rangle} \approx O\left( \frac{1}{\sqrt{N}} \right)$$

In small systems therefore, this ratio is not negligible, so the energy fluctuations can destabilise the system, while a system with large $N$ having more degrees of freedom, has also more ways to absorb the energy and therefore its fluctuations are smaller.

Figure 3. Free Energy landscape as a function of inter chain hydrogen bond energy for a system of 12 peptides for temperatures ranging from $T = 0.588$ to $T = 0.322$. In the lower right corner we show a characteristic oligomeric configuration at $T = 0.322$.

doi:10.1371/journal.pone.0002641.g003

Figure 4. Schematic change in the probability of aggregation in two systems with different number of proteins. If the proteins in the system are all aggregated (left) the stability of an oligomer formed by three proteins is limited by the probability of one protein to go back in solution. In another system (right) both the number of proteins and the volume are increased so that the concentration is left unchanged but since there is one monomer still in solution that can potentially aggregate, the stability of the oligomer is increased.

doi:10.1371/journal.pone.0002641.g004

Figure 5. Finite size scaling analysis. We fitted the average internal energy in each system to the equation: $e(N) = e_c + AN^\alpha$ and we plotted the results as a function of $N^{0.82}$, where $-0.82$ is the value of $\alpha$ obtained from the fit using all our available data.

doi:10.1371/journal.pone.0002641.g005
systems have larger spontaneous fluctuations and the stability of their aggregated phase changes consequently. This latter effect should be expected to be smaller for larger proteins [36] in which the large number of amino acids (degrees of freedom) will reduce, already in the monomer, the energy fluctuations. Indeed this would also be the case if the simulations are performed in explicit solvent as the number of degrees of freedom is increased and therefore the energy fluctuations become smaller.

Finite Size Scaling

In Fig 5 we discuss the finite size analysis to this problem. This type of analysis is commonly used to estimate the typical systematic errors introduced by the finite size of the system on the calculation of various observables. The analysis starts by simulating systems of different sizes at equilibrium and calculating an observable (e.g. the internal energy). Following the procedure used in [38], we fit the average internal energy per peptide with an equation of the form: 
\[ e/N = e_0 + AN^\alpha \]
where \( e_0 \) is the asymptotic value of the internal energy for infinite systems, \( N \) is the system size (i.e., the number of peptides or proteins used in the simulation) and \( \alpha \) is the scaling exponent. For this particular example we do not have estimates for \( e_0 \) and \( \alpha \), as was the case in [38], therefore we fit all three parameters to the average internal energy calculated for the lowest temperature, \( T = 0.322 \), in which we know from our free energy profiles that the peptides are aggregated. Since we have three parameters to fit with only 6 points, our estimates will not be very reliable so by fitting our equation to all possible combinations of 5 points we obtain a rough estimate of their range of validity: 
\[ -12.5 \leq e_0 \leq -9.5, 0.01 \leq \Delta \leq 0.05, -1.1 \leq \alpha \leq -0.6 \]
These values are obviously too broad and a more precise estimate, in particular of the exponent \( \alpha \), should be obtained, either by using a more coarse grained model that will allow simulations of large systems, or by an explicit analytical calculation. In Fig 5 we plot our data along with the best fit over the entire set of data as a function of \( N^{-0.32} \), where \( -0.32 \) is the value of \( \alpha \) taken from this fit. Future work will be required to understand more deeply the precise nature of the exponent \( \alpha \) particularly to understand if it belongs to the class of critical exponents [39] and has therefore a high degree of universality i.e. its value is constant for a class of different proteins having similar characteristics.

Discussion

In the present paper we have discussed the how the number of protein present in the system, if small, influences the underlying free energy and therefore shifts the equilibrium between the different species. We have discussed two independent causes of this problem; the first being that the stability of an oligomer is related to the number of monomers or other oligomers still present in the system, the second related to the relative fluctuations of the system being proportional to \( N^{-1/2} \).

The presence of this finite size effect has very important implications for the in silico studies of protein aggregation, e.g. it is the reason why usually the protein concentration in these studies is taken to be much larger than the concentration used in the corresponding in vitro studies. The presence of a systematic error in every estimate of the stability or rate of formation of one specific aggregate configuration. A way to deal with these problems is to perform a finite size scaling analysis. This type of analysis allows to correct one observable (as we did for the internal energy in section “Finite Size Scaling”) for the systematic error introduced by the finite size of the system. We have outlined how this analysis works, how to carry it out, and we have estimated the values of the parameters within a small range.

A more precise estimate of these parameters would be needed and to this end a coarse grained model should be employed to be able to study a much larger number of peptides and to obtain a precise estimate of the parameters of the scaling analysis. More work will also be required to investigate the degree of universality of such analysis i.e. whether a class of proteins with similar characteristics have the same critical exponent [39]. If verified this would represent a leap forward in our understanding of the generic physical properties of protein aggregation.

Methods

Simulations were carried out with ProfASi (Protein Folding and Aggregation Simulator) [41], which implements an implicit water all-atom model [41–45] for protein folding and aggregation studies. The model assumes fixed bond lengths, bond angles, and peptide torsional angles, so that each amino acid has only the Ramachandran torsional angles and the side chains torsional angles as its degrees of freedom. The interaction potential

\[ E = E_{\text{loc}} + E_v + E_{\text{hb}} + E_{\text{rep}} \]

is composed of four terms. The \( E_{\text{loc}} \) term is local and represents an electrostatic interaction between adjacent peptides along the chain and the \( E_v \) term is an \( 1/r^2 \) repulsion between pairs of atoms. The hydrogen bonding contributions to the energy are calculated by a term, \( E_{\text{hb}} \), in which the distance dependence is modelled through a Lennard-Jones potential between pairs of NH and CO groups within a given cutoff of 4.5 Å, and the angular dependence is expressed as a function of the NH...O and H...OC angles.

The hydrogen bonds between the backbone NH (CO) groups and the CO (NH) groups on each side of them are disallowed. The GFF peptide used in this study, therefore, does not form intra-chain hydrogen bonds and the hydrogen bonding term of the energy monitors only the formation of intra-chain hydrogen bonds. The hydrophobicity term \( E_{\text{rep}} \) is defined by a contact potential between hydrophobic side chains, the latter being proportional to the fraction of atoms in contact in the two amino acids. The parameters of the potential were chosen by optimizing the
agreement with the melting temperature of the Trp. Cage mini-protein; the resulting force field has been shown to reproduce accurately the folded states and the melting temperatures of a range of polypeptide chains of both α and β structures, including Betanova, GB1p, LLM and Fα, with excellent agreement with both CD and NMR data. In particular, the results such as the content of α-helix and the relative population of folded species was also found to be in excellent agreement with experimental data. ProFASi has also already been applied to study the aggregation of a series of short peptides, including the Aβ22–28 and Aβ23–33 peptides [15,30,40]. We performed simulations of six systems composed of three, four, six, eight, nine and twelve GFF three-peptides, using properties such as the content of α-helix and the relative population of folded species was also found to be in excellent agreement with experimental data. ProFASi has also already been applied to study the aggregation of a series of short peptides, including the Aβ22–28 and Aβ23–33 peptides [15,30,40]. We performed simulations of six systems composed of three, four, six, eight, nine and twelve GFF three-peptides, using

References

1. Dobson CM (1999) Protein misfolding, evolution and disease. Trends Biochem Sci 24: 329–332. doi: 10.1016/S0968-0004(99)01226-1

2. Dobson CM (2003) Protein folding and misfolding. Nature 426: 894–899. doi: 10.1038/nature02261. URL http://dx.doi.org/10.1038/nature02261

3. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 75: 333–366. doi: 10.1146/annurev.biochem.75.101304.123901. URL http://dx.doi.org/10.1146/annurev.biochem.75.101304.123901

4. Lansbury PT, Chin C (2006) Probing the biological roles and mechanisms of misfolded proteins. Mol Cell 21: 9–20. doi: 10.1016/j.molcel.2006.02.002. URL http://dx.doi.org/10.1016/j.molcel.2006.02.002

5. Walsh DM, Klyubin I, Fadeeva JV, Cullen WK, Anwyl R, et al. (2002) Naturally derived peptides induce rapid enhancement of LTP in vivo. Science 298: 1317–1321. doi: 10.1126/science.1074583. URL http://dx.doi.org/10.1126/science.1074583

6. Soto P, Griffin MA, Shea JE (2007) New insights into the mechanism of Alzheimer amyloid-beta fibrillogenesis inhibition by N-methylated peptides. Biophys J 93: 3015–25.

7. Buchert NV, Tycko R, Hummer G (2005) Molecular dynamics simulations of Alzheimer’s beta-amyloid proteoliposomes. J Mol Biol 353: 894–921. doi: 10.1016/j.jmb.2005.08.066. URL http://dx.doi.org/10.1016/j.jmb.2005.08.066

8. Urbanc B, Cruz L, Ding F, Sammoud D, Khare S, et al. (2004) Molecular dynamic simulation of amyloid beta dimer formation. Biophys J 87: 2310–2321. doi: 10.1529/biophysj.104.040980. URL http://dx.doi.org/10.1529/biophysj.104.040980

9. Baumknetter A, Shea JE (2007) The structure of the Alzheimer amyloid beta 10–35 peptide probed through replica-exchange molecular dynamics simulations in explicit solvent. J Mol Biol 366: 275–85.

10. Nguyen PH, Li MS, Stock G, Strath JE, Thirumalai D (2007) Monomer adds to preformed structured oligomers of Abeta-peptides by a two-stage dock-lock mechanism. Proc Natl Acad Sci U S A 104: 111–116. doi: 10.1073/pnas.0607440104. URL http://dx.doi.org/10.1073/pnas.0607440104

11. Boucher G, Mousseau N, Derreumaux P (2006) Aggregating the amyloid Abeta(1–25) peptide into a four-stranded beta-sheet structure. Proteins 65: 677–688. doi: 10.1002/prot.21134. URL http://dx.doi.org/10.1002/prot.21134

12. Santini S, Mousseau N, Derreumaux P (2004) In silico assembly of Alzheimer’s Abeta 16–22 peptide into beta-sheets. J Am Chem Soc 126: 11509–11516. doi: 10.1021/ja047289k. URL http://dx.doi.org/10.1021/ja047289k

13. Khare SD, Ding F, Gazit SNS, Dalkohlyan NV (2005) Molecular origin of polyglutamine aggregation in neurodegenerative diseases. PLoS Comput Biol 1: 230–235. doi: 10.1371/journal.pcbi.0010030. URL http://dx.doi.org/10.1371/journal.pcbi.0010030

14. Auer S, Miller MA, Krivov SV, Dobson CM, Karpus M, et al. (2007) Importance of metastable states in the free energy landscapes of polypeptide chains. Phys Rev Lett 99: 178104.

15. Auer S, Dobson CM, Vendruscolo M (2007) Revealing the nucleation barriers and protein aggregation through replica-exchange molecular dynamics simulations. Proc Natl Acad Sci U S A 104: 12916–12921. doi: 10.1073/pnas.0402634101. URL http://dx.doi.org/10.1073/pnas.0402634101

16. Favaie G, Irbab A, Mohanty S (2004) Oligomerization of amyloid Abeta(16–22) peptides using hydrogen bonds and hydrophobic forces. Biophys J 87: 3657–3664. doi: 10.1529/biophysj.104.046839. URL http://dx.doi.org/10.1529/biophysj.104.046839

17. Bellisio G, Shea JE (2007) Selfassembly of beta-sheet forming peptides into chiral fibrillar aggregates. J Chem Phys 126: 245104. doi: 10.1063/1.2739547

18. Simon G, Tiana G, Brogla RA, Colombo G (2004) Modeling the alpha-helix to beta-hairpin transition mechanism and the formation of oligomeric aggregates of the fibrillogenic peptide Abeta(12–28): insights from all-atom molecular dynamics simulations. J Mol Graph Model 23: 263–73.

19. Tiana G, Simon G, Brogla RA, Colombo G (2004) Thermodynamics of beta-amyloid fibril formation. J Chem Phys 120: 8307–17.
35. Paci E, Gsponer J, Salvatella X, Vendruscolo M (2004) Molecular dynamics studies of the process of amyloid aggregation of peptide fragments of transthyretin. J Mol Biol 340: 555–69.
36. Li MS, Klinsov DK, Thurmalai D (2004) Finite size effects on thermal denaturation of globular proteins. Phys Rev Lett 93: 268107.
37. Gazit E (2007) Self-assembled peptide nanostructures: the design of molecular building blocks and their technological utilization. Chem Soc Rev 36: 1263–9.
38. Aspelmeier T, Billoire A, Marinari E, Moore MA (2007) Finite size corrections in the Sherrington-Kirkpatrick model. arXiv 0711.3443.
39. Parisi G (1998) Statistical Field Theory. ABP, Westview Press.
40. Cheon M, Favrin G, Chang I, Dobson CM, Vendruscolo M (2008) Calculation of the free energy barriers in the oligomerisation of Aβ peptide fragments. Frontiers in Bioscience 13: 5614–22.
41. Irbäck A, Mohanty S (2006) Profasi: A monte carlo simulation package for protein folding and aggregation. J Comput Chem 27: 1548–1555. doi:10.1002/jcc.20452. URL http://dx.doi.org/10.1002/jcc.20452.
42. Irbäck A, Samuelsson B, Sjunnesson F, Wallin S (2003) Thermodynamics of alpha- and beta-structure formation in proteins. Biophys J 85: 1466–1473.
43. Irbäck A, Sjunnesson F (2004) Folding thermodynamics of three beta-sheet peptides: a model study. Proteins 56: 110–116. doi:10.1002/prot.20157. URL http://dx.doi.org/10.1002/prot.20152.
44. Irbäck A, Mohanty S (2005) Folding thermodynamics of peptides. Biophys J 88: 1560–1569. doi:10.1529/biophysj.104.050427. URL http://dx.doi.org/10.1529/biophysj.104.050427.
45. Favrin G, Irbäck A, Sjunnesson F (2001) Monte carlo update for chain molecules: Biased gaussian steps in torsional space. J Chem Phys 114: 8154–8158.
46. Marinari E, Parisi G (1992) Simulated tempering: a new Monte Carlo scheme. Europhys. Lett. 19: 451–456.
47. Hukushima K, Nemoto K (1995) Exchange Monte Carlo Method and Application to Spin Glass Simulations. J. Phys. Soc. (Jpn.) 65: 1604.