Aggregation and coacervation with Monte Carlo simulations

Sandipan Mohanty

Institute for Advanced Simulation, Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52425 Jülich, Germany

Abstract

Protein aggregation into oligomeric species has been linked to a number of neurodegenerative diseases. The ability to assemble into ordered fibril like forms under certain conditions is now regarded as a very common property of polypeptide chains. A different form of assembly is known for some proteins in which protein molecules sequester into liquid like droplets. Understanding the biophysical mechanisms behind these phenomena is of great relevance not only for deeper insights into the associated disorders but also for any protein based therapeutics. Here we review a few examples of the use of Markov Chain Monte Carlo simulations in the study of these two forms of protein assembly.

Keywords: Protein aggregation, Liquid liquid phase separation, Markov Chain Monte Carlo, all-atom simulation

Email address: s.mohanty@fz-juelich.de (Sandipan Mohanty)

Introduction

Protein aggregation is the self assembly of proteins into supra-molecular structures ranging from dimers to highly structured amyloid fibrils. While functional forms of fibrils are known[1, 2], much of the research on formation and characterization of amyloid fibrils stems from the fact that biophysical research over many decades has linked the process of aggregation with a variety of diseases such as Alzheimer's, Parkinson's and Huntington's diseases, type II diabetes and spongiform encephalopathies[3, 4, 5, 6]. Detailed causal chains connecting molecular self assembly events to disease pathology remain obscure. But it is believed that a clearer atomic picture will be a very useful aid in understanding and hopefully treating these diseases. Another related reason is that protein aggregation constitutes a common and recalcitrant obstacle in the highly target specific protein based therapeutics[7].

The ability to form aggregates is not just a special property of rare or disease related proteins, but rather a more general property of polypeptide chains[8, 9]. Different biophysical strategies have evolved in cells to oppose protein aggregation[10]. An understanding of the biophysical mechanisms driving protein misfolding and aggregation is therefore expected to inform therapeutic strategies.

Proteins involved in aggregation related diseases such as α-synuclein are often intrinsically disordered proteins (IDP) [11, 12]. IDPs do not have a compact native fold, but rather populate a wide range of structures at physiologically relevant conditions. Some IDPs are also known to undergo a process called liquid liquid phase separation (LLPS) or coacervation, in which concentrated liquid droplets of proteins develop inside the solution[13, 14, 15].

These droplets exhibit a continuous, fluid like, internal reorganization of the member molecules instead of a more ordered, solid like, cross-β arrangement typical of fibrils. It has been suggested that LLPS can facilitate fibril formation for some proteins[16] or present an alternative to aggregation. Functional roles for the liquid droplets have also been proposed where they serve as membrane-less organelles[17]. The phenomenon of LLPS has been studied using different theoretical and computational techniques, such as random phase approximation[18], field theoretical simulations[19] and lattice Monte Carlo simulations[20].

Computer simulations have long been used to provide a conceptual framework to understand biophysical phenomena. Because of the high degree of complexity arising from the large number of interacting components, they constitute an essential class of tools in interpreting experimental studies, connecting them to fundamental physics, organizing our knowledge and asking new questions based on an ever improving picture. Simulation methods used for bio-molecular systems are very diverse. Most computational studies of bio-molecular systems use the molecular dynamics (MD) method. For a recent review of atomistic MD simulations used to study protein aggregation, see Ref. [21]. For an overview of field theoretical simulations for LLPS, see Ref. [19]. A completely different computational approach is the Monte Carlo method. The Markov Chain Monte Carlo (MC) method was developed in the 1940s in the Los Alamos National Laboratory[22] in connection with nuclear weapons research. Because of the generality of the technique and its scaling properties for high dimensional problems, the MC method has found increasing use since its inception in diverse fields of inquiry ranging from

fundamental high energy physics to finance and risk evaluation. Although in bio-molecular simulations the use of MC is less common than MD, for some problems they offer an alternative to MD at a significantly lower computational cost. In this chapter, we will summarize a few applications of MC simulations to explore self assembly of proteins.

Markov chain Monte Carlo simulations

Thermodynamic Monte Carlo (MC) simulations sample random structures from an ensemble which is either directly a Boltzmann ensemble, or has a rigorous exact mapping to such an ensemble. From such a random sample, statistical averages can be estimated. The class of MC algorithms most commonly used for molecular simulations is called a Markov Chain Monte Carlo, and in the following, we will refer to this type of MC simulations as simply MC. In this method, the simulated system performs a random walk in the conformation space. At every step, we propose a random change to the conformation, and accept or reject the change according to an acceptance probability. This simple recipe of random moves with accept/reject decisions is repeated billions of times and sample points for statistical analysis are collected at regular intervals from the resulting Markov Chain. The principle of detailed balance is a convenient way to ensure that the sample points reach a stationary distribution with the correct relative statistical weights. If $P(A)$ and $P(B)$ are the equilibrium probabilities of states A and B, and p_{AB} is the probability to propose a move from A to B (and likewise for p_{BA}), detailed balance can be expressed as the requirement $P(A)p_{AB} = P(B)p_{BA}$. Conformation updates satisfying detailed balance do not cause a net flow

of probabilities. For proper sampling, it is also important that the entire conformation space remains connected under the conformation updates.

The form of the acceptance probability depends on the type of ensemble we intend to generate. If $P(A)$ represents the weight of a state A in the target ensemble, the acceptance probability for a conformation change to a state B is $p = \min(1, \frac{P(B)}{P(A)})$. In the simplest case of a canonical ensemble, the microscopic weight of a state A is $P(A) \propto \exp(-\beta E_A)$, where E_A is the energy of state A, $\beta = \frac{1}{k_B T}$ is the inverse temperature, and k_B is the Boltzmann's constant. Each state with an energy E will have a weight of $\exp(-\beta E)$, and the total probability of all states with energy E is proportional to $g(E)\exp(-\beta E)$, where $g(E)$ is called the density of states, which counts the number of states at energy E. For a simple Metropolis MC simulation sampling the canonical ensemble, the acceptance probability takes the form $p \propto \min(1, \exp(-\beta \Delta E))$, ΔE being the change in energy due to the conformation change. Long Markov chains with this acceptance probability will sample each state of energy E with weight $\exp(-\beta E)$, so that with $g(E)$ states at energy E, the sampled probability for energy E will be an approximation of $g(E)\exp(-\beta E)$. In general, $g(E)$ is an unknown function, and the histogram of energy from a long Markov chain is one way to estimate it. This is one of the important goals of thermodynamic MC simulations, as a good estimate of $g(E)$ along with correlations of various state properties with energy can be used to calculate thermodynamic averages.

The canonical MC recipe described above can be improved in many ways for faster convergence of statistical averages. The $\exp(-\beta \Delta E)$ factor in the acceptance probability suppresses positive energy changes, which impedes

the escape from local energy minima. Since the global energy minimum of a system is usually unknown at the start, it is important that the simulations are able to escape local energy minima, and search for the low lying states in the entire energy landscape. Sophisticated techniques have been developed to help the MC simulations navigate rough energy landscapes and converge faster. These include simulated tempering[23], parallel tempering[24, 25], multi-canonical method[26] and Wang-Landau simulations[27, 28].

Small update MC simulations

Unlike in MD simulations, in MC, we do not integrate Newton's equations of motion. There is no restriction on how big or small a single step along the Markov Chain can be, and this property can be exploited for very fast importance sampling from a high dimensional conformation space. In many studies, the random updates are restricted to make small overall changes to the system so that after an MC step, the system retains much of its structure from before the step, with small modifications. A Markov Chain generated in this way vaguely mimics a trajectory, i.e., continuously changing coordinates as a function of time, as seen for instance, in MD simulations. The MC simulations however do not contain a real "time" variable, and can not reveal information about physical time scales of processes. In our own research, we have found that small update MC simulations sampling the canonical ensemble yield pseudo-trajectories with easily rationalizable qualitative behavior. This is not entirely unexpected, as, despite not following a path set by Newton's equations of motion, the Markov Chain stochastically follows paths defined by the energy landscape. Let's compare how coordinates are updated in a small update MC simulation and an MD simulation us-

ing the same force field. The conformation updates in the MC leading in the opposite direction relative to the acceleration in the corresponding MD simulation (i.e., updates along the positive gradient of the potential energy) are unfavorable due to the $\min(1, \exp(-\beta \Delta E))$ acceptance probability. MC updates leading roughly in the direction of the acceleration on the other hand, are accepted. The MC simulation treats all possible changes which reduce the potential energy equivalently, whereas the MD will pick one of them definitively. But the barriers in the energy landscape which slow down a particular transition in an MD will have a similar effect on a small update MC, because of the exponential suppression of the positive ΔE in the acceptance probability. Larger changes, unlike larger time steps in an MD do not result in larger numeric errors in an MC simulation, but suppress their superficial resemblance to MD trajectories. Therefore, with the understanding that no time scales can be read out from the simulations, small update MC simulations have often been used to study kinetic processes because of their much smaller computational cost compared to a equivalent MD simulations [29, 30, 31, 32, 33, 34, 35, 36].

Folding and aggregation using all atom MC simulations

In this subsection, we will review a few examples of the use of the MC method to study peptide aggregation. Although the methods discussed above are very general, currently we focus on studies conducted using a specific implicit solvent physics based model. In this model, we explicitly represent all atoms in the protein molecules, including all hydrogen atoms, as interacting entities. We ignore fluctuations in lengths of covalent bonds as well as the bond angles between converging bonds, so that all sample states considered

preserve the values for these geometrical parameters. This leaves only torsional rotation around bonds inside a single molecule and global translations and rotations as degrees of freedom. The conformation update in every step of the MC simulation is chosen randomly from a small set of move types. This set includes rigid body translations and rotations of one or more molecules, random changes to a single randomly chosen backbone or side chain torsion angle, or a concerted rotation of up to 8 backbone angles affecting a local deformation of a single protein chain[37]. The random single angle updates as well as the rigid body translations and rotations can create very large displacements of large numbers of atoms. If we worked with explicitly represented solvent molecules, most of these updates would result in steric clashes and our approach will lose much of its ability for fast exploration. We therefore developed an implicit solvent force field to model protein interactions.

Our interaction potential consists of four terms : $E = E_{exv} + E_{loc} + E_{hb} + E_{sc}$. Excluded volume effects are represented by E_{exv}. E_{loc} represents local electrostatic effects along the protein backbone. E_{hb} represents backbone and side chain hydrogen bonds. E_{sc} encapsulates side chain charge charge interactions as well as the hydrophobic effect. Detailed mathematical forms of each of these terms can be found elsewhere[38]. Here we only note that in this reductive physical approach, the interaction parameters are tied to atoms or groups of atoms, and are regarded as universal. For instance, every phenylalanine residue has exactly the same interaction capacity with respect to each of the terms in our force field, irrespective of its position in a protein sequence or structure. This also means that in simulations with many molecules, intra-molecular and inter-molecular interactions necessarily have

the same functional form and parameters. The form and parameters of this force field have been developed through thermodynamic folding simulations of small polypeptide chains [39, 38] from randomly initialized chains. The approximations of this force field developed through studies of small peptides of sequence length 10–30 residues have occasionally proven to be sufficient to describe reversible folding transitions of somewhat larger sequences (See Fig. 1), such as Top7-CFr, several three helix bundle proteins of about 70 residues, and even a mixed 92 residue α/β protein Top7[40, 38, 41] with very slow folding times of about 1 second in experiments.

The ability of simulations with this model to find the native folded states of some proteins starting from completely randomized initial configurations, without biasing the simulations with any prior information on the folded states, is a very non-trivial attribute. Calculated temperature dependence of secondary structure propensities and native populations often agree well with suitable experimental results. An implementation of the protein model, force field and MC simulation tools is available as an open source software package called ProFASi[42].

We have used this all atom protein model to explore protein aggregation and related phenomena using MC simulations. In 2004, after developing an early version of the model, we applied it to study the oligomerization of $A\beta_{16-22}$ peptides (sequence: KLVFFAE) [43]. From that proof-of-concept study with only up to 6 chains of $A\beta_{16-22}$ peptides, we learned that systems of many protein chains can show interesting physical behavior in MC simulations. The MC simulations sampled multiple transitions between isolated monomers and several different kinds of oligomers. Monomer simulations

showed that the isolated Aβ_{16-22} behaves as a random coil. In multi-chain simulations, the statistical secondary structure propensities were consistent with single chain simulations at high temperatures, but with decreasing temperatures the β-sheet content smoothly increased, exhibiting a sigmoid temperature dependence. Analysis of the structures obtained at the low temperatures showed that they were predominantly oligomeric, and the β-sheet secondary structure was exclusively from the oligomers present in the structures.

In [30], we compared two truncated segments Aβ_{16-22} and Aβ_{25-35} (sequence: GSNKGAIIGLM) of the Alzheimer's Aβ peptide with small update MC simulations with ProFASi at different concentrations and temperatures. It turned out that the two peptides show different aggregation behavior. The Aβ_{16-22} system showed a rapid hydrophobic collapse to large clusters of unstructured aggregates with ample chain mobility. These structures then gradually transitioned into β-sheet rich oligomers due to the formation of inter chain hydrogen bonds, which curiously exposed some hydrophobic groups. In contrast, the Aβ_{25-35} showed little tendency to collapse into disordered aggregates, but rather formed β-sheet hydrogen bonds early in the process.

In [32], we examined aggregation behavior of a small segment AcPHF6 (Ac-VQIVYK-NH2) of the tau protein, also connected with the Alzheimer's disease. We analyzed the structure and organization of the oligomers forming in the simulations. For instance, we kept track of the relative orientation of the neighboring chains in a β-sheet and the inter-chain hydrogen bonds. We found that a variety of very small oligomers emerged quickly with no preference for orientation or registry between neighboring chains. These small

oligomers formed and dissolved rapidly resulting in an equilibrium with the solution. This stochastic process sometimes results in oligomers of a critical size, which do not dissolve as easily. Of these relatively stable oligomers, we found that only a subset continued to grow into long β-sheets and two sheet oligomers (Fig. 2). The subset of the minimal stable oligomers which managed to grow in our simulations were those with the greatest internal order, i.e., those with the highest percentage of parallel in-register strand organization. Even the largest simulations in reference [32] had only 36 peptides in them, which makes multi-layered structures less probable. The simulations did however result in many double layered aggregates exhibiting a steric zipper interface (see Fig. 2) between the two sheets with the V1, I3 and Y5 side chains packed in the sheet-sheet interface. Such a dry-steric zipper interface has been proposed for a variety of small fibril forming peptides including PHF6 using X-ray diffraction[44].

The above mentioned simulations are examples of small update MC simulations where the Markov chains resemble kinetic trajectories. In [31], Irbäck and Mitternacht again explored oligomerization of Aβ_{16-22} using equilibrium simulations. For this purpose they used our all-atom model and its implementation in ProFASi, but did not restrict the runs to small updates. They also tested experimental modifications to the force field which contributed to the next iteration of the default interaction calculations in ProFASi. These simulations provided a different view of the oligomerization of Aβ_{16-22} . The temperature dependence of properties such as the size of the largest cluster in the system and the secondary structure content were calculated with excellent statistical errors. They showed that with decreasing temperature,

the size of the largest cluster starts to increase a little before the β-sheet secondary structure does. This is in perfect agreement with the more kinetic simulations[30] which showed that the disordered "molten" oligomers formed before the ordered β-sheets in MC time. In addition, they found that the simulations resulted in the spontaneous formation of tight, extraordinarily long-lived β-barrels.

Aggregating systems are often characterized by strongly bimodal energy distributions. Sparse conformations without aggregates have energies corresponding to multiple free monomers. Once an aggregating nucleus forms, there is a rapid transition to a low energy oligomer rich state. For such systems, the canonical probability of intermediate states connecting the compact low energy states with the sparse higher energy states is very small. Therefore, MC simulations based on the canonical ensemble only rarely visit the intermediate states, and because of this, they may struggle to cross back and forth between the low and high energy basins in free energy. Methods such as multicanonical or Wang-Landau simulations can alleviate the difficulty of crossing the free energy barrier. In Ref. [45], we applied a slightly modified form of the Wang-Landau method to a system of 8 chains of a 7 residue fragment (sequence : GIIFNEQ) of the Cu/Zn superoxide dismutase 1 protein (SOD1). This system exhibited an aggregated and an unaggregated phase, which coexisted at the mid-point temperature with a free energy barrier of height 2.7 k_BT. The intermediate states for this system had low canonical probabilities at the mid point temperature, (approximately $\frac{1}{15}$ of the probability of the aggregated and unaggregated states). Analogous to the multicanonical method, our simulations sampled a non-canonical ensemble

which amplifies these intermediate states facilitating repeated sampling of the high and low energy basins in free energy. This is achieved while maintaining a well defined mapping to the canonical ensemble so that the correct thermodynamic averages can be calculated during post-processing. Using direct comparison, we showed that results from our modified Wang-Landau method agree with those from canonical simulations, but have smaller statistical errors per consumed computing resources.

Exploring fibril formation with lattice models

While atomic resolution simulations, like those discussed above, have been used for studies of small oligomeric systems, it is easier to obtain general insights on much larger scale processes by using reduced representations of the molecules. Lattice models represent one popular class of reduced representation. Each amino acid is represented by one or a few beads, and each bead exclusively occupies one location in a grid of possible positions. A peptide chain in such a model is a sequence of connected beads. In Ref. [46], Li et al developed a model in which an amino acid is represented as a hydrophobic, polar or charged bead, with electrostatic and hydrophobic interactions with other beads. Their MC moves consisted of global translations and rotations of a randomly chosen peptide, and local moves such as tail rotations and crank shaft rotations. Since the set of states on a spatial grid is countable, it is possible to perform exact enumeration of all possible conformations for small system sizes. Despite the simplicity of this model, it was able to capture several interesting characteristics of peptide systems. Li et al found a non-degenerate ground state clearly separated from degenerate higher energy conformations for a chosen peptide sequence. In simulations

with multiple chains, they observed self-assembly into anti-parallel β-sheet like structures, with the strands arranged perpendicular to the fibril axis. In Ref. [47], this model was further used to study the effect of macro molecular crowders on peptide aggregation. Crowders were modeled as impenetrable cubical entities spanning several grid positions in each dimension. The crowders limited access to a part of the spatial grid, but did not otherwise interact with the peptide chains. Using MC simulations with this model, they were able to reconcile experimental observations regarding both increase[48] and decrease[49] of fibril formation rate in the presence of co-solutes. For large crowders, the model showed a decrease in fibril formation time with increasing crowder concentrations, while for small crowders, it showed the opposite effect. This result illustrates how MC simulations with a good physical model can sometimes help clarifying very non-trivial experimental observations.

In Ref. [50], Irbäck et al explored the thermodynamics of fibril formation using a stick model for small peptides in a lattice approximation. The peptides were modeled as unit length sticks, which were allowed to occupy any available positions on a 3 dimensional cubic lattice. The hydrogen bonding direction was assumed to be perpendicular to the chain (stick) direction, and the hydrophobic interactions were assumed to be in the direction perpendicular to both of these. This is inspired by the typical geometry of a cross-β structure of a fibril, with hydrogen bonds running perpendicular to the chain axis and hydrophobic interactions bind multiple sheets together. By using a reduced model and strategies to update large clusters together, they were able to study systems consisting of up to 131,072 peptides. They found that the simulated system shows sigmoidal kinetics characteristic of

fibril formation in experiments[51]. The fibril like structures found in these simulations had an average length of around 210 and comprised 7 layers on average. Similar to the above mentioned all-atom simulations of AcPHF6 peptide, they observed an initial waiting phase where small aggregates form an dissolve until by random chance a stable growth capable aggregate forms. After that there is rapid growth in the mass of the aggregate until the depletion of available monomers slows the kinetics down. In the stick model simulation, the distinguishing characteristics of the critical aggregate was its width in terms of the number of β-sheet layers in the aggregate. Most growth capable aggregates had at least 4 layers.

A different type of model was developed by Abeln et al[52] to study protein folding and aggregation on 3 dimensional cubic lattices. Instead of grouping amino acids into broad categories such as hydrophobic and polar, they used 20 different kinds of beads to represent the 20 different naturally occurring amino acids. The state of each amino acid in this model consists of a position at one of the lattice points, a secondary structure marker (strand or coil) and a side chain direction marker (which could point to one of the neighboring sites not occupied by the backbone). The potential energy function was written to mimic the interactions of real proteins, including terms to account for hydrogen bonds, interactions with the solvent and steric terms for the side chains. The idea was to infuse a lattice model with enough details to capture the most essential physical aspects of protein behavior. Despite these increased details, this is still a lattice model with orders of magnitude lower computational complexity than a typical atomic resolution model. Abeln et al designed a protein sequence for a target 3D structure and simulated the

designed sequence with their model using MC simulations. The simulations showed a sharp folding transition as a function of temperature, with a rapid increase in the number of native contacts. Control simulations performed for a random sequence with a similar amino acid content did not show any similar transition to a hydrogen bond rich compact state. They also performed simulations of multi-chain systems starting with no contact among different chains. Spontaneous formation of small oligomers as well as fibril like structures was observed. The observed inter-molecular contacts in these aggregates are consistent with the cross-β geometry of fibrils. The thermal stability of the multi-layered fibril like structures was found to be maximum for sequences with the ability to pack hydrophobic groups between β-sheets.

MC simulations of liquid liquid phase separation

MC simulations on cubic lattices have also been used to study liquid-liquid phase separation in intrinsically disordered proteins. In [20], Das et al modeled protein chains as self avoiding walks on a cubic lattice. Each monomer along the sequence exclusively occupies a lattice site, and is connected to its sequence neighbors which must occupy a nearest neighbor lattice site. The monomers in each chain interact with other monomers in any chain with electrostatic interactions. Using MC simulations with 300 chains of 50 monomers, several box sizes and temperatures, they compared the temperature and concentration dependence of polymer density for different charge patterns along the sequence. Comparing sequences with zero net charge but different distribution of charge along the sequence, they found that sequences with greater clustering of like charges showed a greater tendency to phase separate.

Conclusion

In this short review we have visited a few examples of the use of MC simulations for the study of self assembly of proteins. Aggregation of proteins into amyloid fibrils spans many orders of magnitude in length and time scales, and models with different levels of detail have been used to create complementary insights. The models span a range of resolutions from highly detailed all-atom descriptions of the protein chains to stick models representing entire protein chains as unit sticks on a lattice. We have used atomically detailed models for systems consisting of a few chains to few hundred short chains, and such simulations can capture essential sequence specific behavior, such as different aggregation mechanisms between two fragments taken from the Aβ peptide. They also support the idea that the tendency to form fibril like structures is an inherent property of polypeptide chains, because interaction models developed to describe protein folding lead to spontaneous aggregation for certain sequences when many chains are present. Despite the significant speed benefits of the MC procedure however, such models have a very high computational cost when applied to the formation of amyloid fibrils. General principles governing processes at larger length scales have been more successfully explored with simpler models, at the cost of finer sequence level details. Bead chain models on a lattice with different levels of details have been shown to capture the essential features of the folding and aggregation processes. In studies of liquid liquid phase separation, lattice models with only electrostatic interactions have revealed that particular charge distributions are more prone to phase separate than others. Similar studies with HP models, treating each amino acid bead as either hydrophobic

(H) or polar (P), are currently underway. Even simpler models with peptides as unit length sticks have been used in simulations of over 10^5 chains to gain insights about the kinetics of fibril formation. The complete picture of protein assembly into liquid like droplets or ordered solid like fibrils requires atomic resolution insights as well as physics at a coarser resolution. The MC technique has proven useful at each of these resolutions and will continue to play a role, alongside other theoretical, computational and experimental methods in the immediate future.

[1] Chiti Fabrizio, Dobson Christopher M.. Protein Misfolding, Functional Amyloid, and Human Disease *Annual Review of Biochemistry.* 2006;75:333-366.

[2] Greenwald Jason, Riek Roland. Biology of Amyloid: Structure, Function, and Regulation *Structure.* 2010;18:1244-1260.

[3] Selkoe Dennis J.. Folding Proteins in Fatal Ways *Nature.* 2003;426:900-904.

[4] Roberson Erik D., Mucke Lennart. 100 Years and Counting: Prospects for Defeating Alzheimer's Disease *Science.* 2006;314:781-784.

[5] Lansbury Peter T., Lashuel Hilal A.. A Century-Old Debate on Protein Aggregation and Neurodegeneration Enters the Clinic *Nature.* 2006;443:774-779.

[6] Invernizzi Gaetano, Papaleo Elena, Sabate Raimon, Ventura Salvador. Protein Aggregation: Mechanisms and Functional Consequences *The*

International Journal of Biochemistry & Cell Biology. 2012;44:1541-1554.

[7] Zambrano Rafael, Jamroz Michal, Szczasiuk Agata, Pujols Jordi, Kmiecik Sebastian, Ventura Salvador. AGGRESCAN3D (A3D): Server for Prediction of Aggregation Properties of Protein Structures *Nucleic Acids Research.* 2015;43:W306-W313.

[8] Dobson Christopher M.. Protein Folding and Misfolding *Nature.* 2003;426:884-890.

[9] Knowles Tuomas P. J., Vendruscolo Michele, Dobson Christopher M.. The Amyloid State and Its Association with Protein Misfolding Diseases *Nature Reviews Molecular Cell Biology.* 2014;15:384-396.

[10] Monsellier Elodie, Chiti Fabrizio. Prevention of Amyloid-like Aggregation as a Driving Force of Protein Evolution *EMBO reports.* 2007;8:737-742.

[11] Uversky Vladimir N.. What Does It Mean to Be Natively Unfolded? *European Journal of Biochemistry.* 2002;269:2-12.

[12] Uversky Vladimir N.. Natively Unfolded Proteins: A Point Where Biology Waits for Physics *Protein Science.* 2002;11:739-756.

[13] Brangwynne Clifford P., Eckmann Christian R., Courson David S., et al. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation *Science.* 2009;324:1729-1732.

[14] Uversky Vladimir N., Kuznetsova Irina M., Turoverov Konstantin K., Zaslavsky Boris. Intrinsically Disordered Proteins as Crucial Constituents of Cellular Aqueous Two Phase Systems and Coacervates *FEBS Letters.* 2015;589:15-22.

[15] Molliex Amandine, Temirov Jamshid, Lee Jihun, et al. Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization *Cell.* 2015;163:123-133.

[16] Wegmann Susanne, Eftekharzadeh Bahareh, Tepper Katharina, et al. Tau Protein Liquid–Liquid Phase Separation Can Initiate Tau Aggregation *The EMBO Journal.* 2018;37:e98049.

[17] Nott Timothy J., Petsalaki Evangelia, Farber Patrick, et al. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles *Molecular Cell.* 2015;57:936-947.

[18] González-Mozuelos P., de la Cruz M. Olvera. Random Phase Approximation for Complex Charged Systems: Application to Copolyelectrolytes (Polyampholytes) *The Journal of Chemical Physics.* 1994;100:507-517.

[19] McCarty James, Delaney Kris T., Danielsen Scott P. O., Fredrickson Glenn H., Shea Joan-Emma. Complete Phase Diagram for Liquid–Liquid Phase Separation of Intrinsically Disordered Proteins *The Journal of Physical Chemistry Letters.* 2019;10:1644-1652.

[20] Das Suman, Eisen Adam, Lin Yi-Hsuan, Chan Hue Sun. A Lattice

Model of Charge-Pattern-Dependent Polyampholyte Phase Separation *The Journal of Physical Chemistry B.* 2018;122:5418-5431.

[21] Carballo-Pacheco Martín, Strodel Birgit. Advances in the Simulation of Protein Aggregation at the Atomistic Scale *The Journal of Physical Chemistry B.* 2016;120:2991-2999.

[22] Metropolis Nicholas, Ulam S.. The Monte Carlo Method *Journal of the American Statistical Association.* 1949;44:335-341.

[23] Marinari E., Parisi G.. Simulated Tempering: A New Monte Carlo Scheme *EPL (Europhysics Letters).* 1992;19:451.

[24] Swendsen Robert H., Wang Jian-Sheng. Replica Monte Carlo Simulation of Spin-Glasses *Physical Review Letters.* 1986;57:2607-2609.

[25] Hukushima Koji, Nemoto Koji. Exchange Monte Carlo Method and Application to Spin Glass Simulations *Journal of the Physical Society of Japan.* 1996;65:1604-1608.

[26] Berg Bernd A., Neuhaus Thomas. Multicanonical Algorithms for First Order Phase Transitions *Physics Letters B.* 1991;267:249-253.

[27] Wang Fugao, Landau D. P.. Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States *Physical Review Letters.* 2001;86:2050-2053.

[28] Wang Fugao, Landau D. P.. Determining the Density of States for Classical Statistical Models: A Random Walk Algorithm to Produce a Flat Histogram *Physical Review E.* 2001;64:056101.

[29] Shimada Jun, Kussell Edo L., Shakhnovich Eugene I.. The Folding Thermodynamics and Kinetics of Crambin Using an All-Atom Monte Carlo simulation11Edited by A. R. Fersht *Journal of Molecular Biology.* 2001;308:79-95.

[30] Cheon Mookyung, Chang Iksoo, Mohanty Sandipan, et al. Structural Reorganisation and Potential Toxicity of Oligomeric Species Formed during the Assembly of Amyloid Fibrils *PLOS Computational Biology.* 2007;3:e173.

[31] Irbäck Anders, Mitternacht Simon. Spontaneous β-Barrel Formation: An All-Atom Monte Carlo Study of Aβ16–22 Oligomerization *Proteins: Structure, Function, and Bioinformatics.* 2008;71:207-214.

[32] Li Da-Wei, Mohanty Sandipan, Irbäck Anders, Huo Shuanghong. Formation and Growth of Oligomers: A Monte Carlo Study of an Amyloid Tau Fragment *PLoS Computational Biology.* 2008;4:e1000238.

[33] Irback A., Mitternacht S., Mohanty S.. Dissecting the Mechanical Unfolding of Ubiquitin *Proceedings of the National Academy of Sciences.* 2005;102:13427-13432.

[34] Mitternacht Simon, Luccioli Stefano, Torcini Alessandro, Imparato Alberto, Irbäck Anders. Changing the Mechanical Unfolding Pathway of FnIII10 by Tuning the Pulling Strength *Biophysical Journal.* 2009;96:429-441.

[35] Jónsson Sigurður Ægir, Mitternacht Simon, Irbäck Anders. Me-

chanical Resistance in Unstructured Proteins *Biophysical Journal.* 2013;104:2725-2732.

[36] Bille Anna, Jensen Kristine Steen, Mohanty Sandipan, Akke Mikael, Irbäck Anders. Stability and Local Unfolding of SOD1 in the Presence of Protein Crowders *The Journal of Physical Chemistry B.* 2019;123:1920-1930.

[37] Favrin Giorgio, Irbäck Anders, Sjunnesson Fredrik. Monte Carlo Update for Chain Molecules: Biased Gaussian Steps in Torsional Space *The Journal of Chemical Physics.* 2001;114:8154-8158.

[38] Irback Anders, Mitternacht Simon, Mohanty Sandipan. An Effective All-Atom Potential for Proteins *PMC Biophysics.* 2009;2:2.

[39] Irbäck Anders, Mohanty Sandipan. Folding Thermodynamics of Peptides *Biophysical Journal.* 2005;88:1560-1569.

[40] Mohanty Sandipan, Meinke Jan H., Zimmermann Olav, Hansmann Ulrich H. E.. Simulation of Top7-CFr: A Transient Helix Extension Guides Folding *Proceedings of the National Academy of Sciences.* 2008:0708411105.

[41] Mohanty Sandipan, Meinke Jan H., Zimmermann Olav. Folding of Top7 in Unbiased All-Atom Monte Carlo Simulations *Proteins: Structure, Function, and Bioinformatics.* 2013;81:1446–1456.

[42] Irback A., Mohanty S.. PROFASI: A Monte Carlo Simulation Package for Protein Folding and Aggregation *J. Comput. Chem.* 2006;27:1548-1555.

[43] Favrin G., Irback A., Mohanty S.. Oligomerization of Amyloid A ß 16-22 Peptides Using Hydrogen Bonds and Hydrophobicity Forces *Biophysical Journal.* 2004;87:3657-3664.

[44] Sawaya Michael R., Sambashivan Shilpa, Nelson Rebecca, et al. Atomic Structures of Amyloid Cross-Beta Spines Reveal Varied Steric Zippers *Nature.* 2007;447:453-457.

[45] Jónsson Sigurður Æ., Mohanty Sandipan, Irbäck Anders. Accelerating Atomic-Level Protein Simulations by Flat-Histogram Techniques *The Journal of Chemical Physics.* 2011;135:125102.

[46] Li Mai Suan, Klimov D. K., Straub J. E., Thirumalai D.. Probing the Mechanisms of Fibril Formation Using Lattice Models *The Journal of Chemical Physics.* 2008;129:175101.

[47] Co Nguyen Truong, Hu Chin-Kun, Li Mai Suan. Dual Effect of Crowders on Fibrillation Kinetics of Polypeptide Chains Revealed by Lattice Models *The Journal of Chemical Physics.* 2013;138:185101.

[48] White Duncan A., Buell Alexander K., Knowles Tuomas P. J., Welland Mark E., Dobson Christopher M.. Protein Aggregation in Crowded Environments *Journal of the American Chemical Society.* 2010;132:5170-5175.

[49] Cabaleiro-Lago Celia, Quinlan-Pluck Fiona, Lynch Iseult, et al. Inhibition of Amyloid β Protein Fibrillation by Polymeric Nanoparticles *Journal of the American Chemical Society.* 2008;130:15437-15443.

[50] Irbäck Anders, Jónsson Sigurður Æ., Linnemann Niels, Linse Björn, Wallin Stefan. Aggregate Geometry in Amyloid Fibril Nucleation *Physical Review Letters.* 2013;110:058101.

[51] Hellstrand Erik, Boland Barry, Walsh Dominic M., Linse Sara. Amyloid β-Protein Aggregation Produces Highly Reproducible Kinetic Data and Occurs by a Two-Phase Process *ACS Chemical Neuroscience.* 2010;1:13-18.

[52] Abeln Sanne, Vendruscolo Michele, Dobson Christopher M., Frenkel Daan. A Simple Lattice Model That Captures Protein Folding, Aggregation and Amyloid Formation *PLOS ONE.* 2014;9:e85185.

Figures

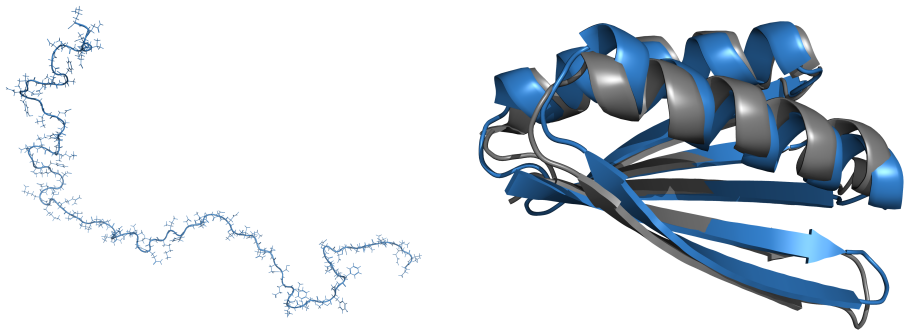

Figure 1: The protein model highlighted in this chapter in the context of atomistic MC simulations is capable of folding many small proteins starting from completely random conformations. Here we show two snapshots from a simulation of a 92 residue α/β protein Top7 for which the native state contains two α-helices and a 5 stranded β-sheet. Left: an example structure near the beginning of the simulation. Right: a snapshot from the simulation (blue) representing the free energy minimum at 273 K aligned with the PDB structure 1QYS (gray).

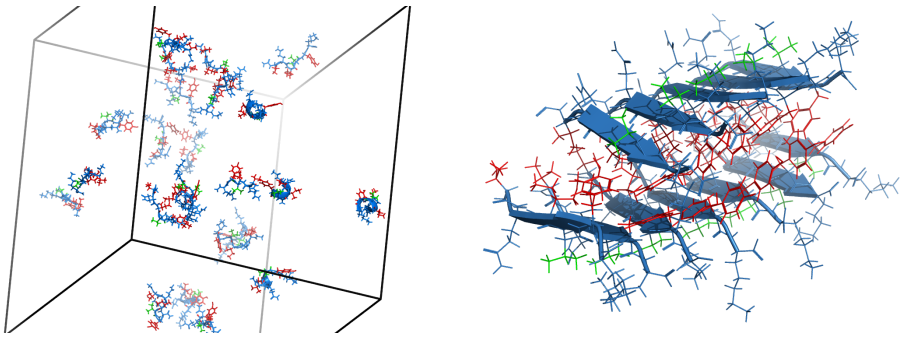

Figure 2: Example snapshots from MC simulations of 24 chains of Ac-PHF6-NH2 at 308 Kelvin in a 95 Åperiodic box. Residues V^1, I^3 and Y^5, which appear in the dry-steric zipper interface in two layer aggregates have been marked red, while V^4, another hydrophobic residue which does not get buried in the dry interface is marked green. Left: a snapshot from near the beginning of the simulation. Right: closer view of a large oligomer towards the end of the simulation.