5-8-2015

Self-assembly dynamics for the transition of a globular aggregate to a fibril network of lysozyme proteins via a coarse-grained Monte Carlo simulation

R. B. Pandey
University of Southern Mississippi

B. L. Farmer
North Carolina State University

Bernard S. Gerstman
Department of Physics, Florida International University, gerstman@fiu.edu

Self-assembly dynamics for the transition of a globular aggregate to a fibril network of lysozyme proteins via a coarse-grained Monte Carlo simulation

R. B. Pandey,[1] B. L. Farmer,[2] and Bernard S. Gerstman[3]

[1]*Department of Physics and Astronomy, University of Southern Mississippi, Hattiesburg, MS 39406, USA*
[2]*Materials and Manufacturing Directorate, Air Force Research Laboratory, Wright Patterson Air Force Base, OH 45433, USA and/or Materials Science and Engineering, North Carolina State University, Raleigh, NC 27606, USA*
[3]*Department of Physics, Florida International University, University Park, Miami, FL 33199, USA*

The self-organizing dynamics of lysozymes (an amyloid protein with 148 residues) with different numbers of protein chains, $Nc = 1, 5, 10$, and 15 (concentration $0.004 – 0.063$) is studied by a coarse-grained Monte Carlo simulation with knowledge-based residue-residue interactions. The dynamics of an isolated lysozyme ($Nc = 1$) is ultra-slow (quasi-static) at low temperatures and becomes diffusive asymptotically on raising the temperature. In contrast, the presence of interacting proteins leads to concentration induced protein diffusion at low temperatures and concentration-tempering sub-diffusion at high temperatures. Variation of the radius of gyration of the protein with temperature shows a systematic transition from a globular structure (at low T) to a random coil (high T) conformation when the proteins are isolated. The crossover from globular to random coil becomes sharper upon increasing the protein concentration (i.e. with $Nc = 5, 10$), with larger R_g at higher temperatures and concentration; R_g becomes smaller on adding more protein chains (e.g. $Nc = 15$) a non-monotonic response to protein concentration. Analysis of the structure factor ($S(q)$) provides an estimate of the effective dimension ($D \geq 3$, globular conformation at low temperature, and $D \sim 1.7$, random coil, at high temperatures) of the isolated protein. With many interacting proteins, the morphology of the self-assembly varies with scale, i.e. at the low temperature ($T = 0.015$), $D \sim 2.9$ on the scale comparable to the radius of gyration of the protein, and $D \sim 2.3$ at the large scale over the entire sample. The global network of fibrils appears at high temperature ($T = 0.021$) with $D \sim 1.7$ (i.e. a random coil morphology at large scale) involving tenuous distribution of micro-globules (at small scales). © *2015 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported License.* [http://dx.doi.org/10.1063/1.4921074]

I. INTRODUCTION

A diverse range of morphologies, from a solid aggregate to a tenuous fibril network can appear due to self- and directed (by an underlying matrix environment) assembly of proteins and peptides. Such organized structures are believed to cause a number of neurodegenerative disorders (Alzheimer, Parkinson, Coronary heart, etc.).[1–12] Many different human proteins such as beta-amyloid, alpha-synuclein, lysozyme, and insulin, with different sequences and structures can form insoluble amyloid fibrils with somewhat similar morphology.[13] Enormous efforts have been made in attempting to understanding the underlying mechanisms. For example, beta sheet formation of amyloid proteins appears to be one of the main causes in amyloid diseases.[12,13] The understanding of the morphological evolution from an isolated protein to fibrils is far from complete, particularly in such proteins as lysozyme which is involved in hereditary non-neuropathic systemic

amyloidosis.[12] Many of the computational investigations have been limited to self-assembly of peptides[14] and short segments of the proteins.[1–4] The protein dynamics remain an open problem.[15] Enormous computational efforts employing various approaches have been made, involving atomic scale structures to coarse-grained representations, first principle force-fields to phenomenological interactions, structure-based constraints to knowledge-based contacts, etc. Investigating the structural transitions with simulations that involve many proteins, solvent, and a host of other constitutive elements is even more challenging.[1–5] In order to study the self-organizing structures involved in fibril formation, aggregation, dispersion, etc. resulting from the cooperative and competing mechanisms, one has to consider many proteins in a matrix. Therefore, simplifications via appropriate coarse-grained approximations is necessary. Using a coarse-grained approach[5] we investigate the self-organizing structures of lysozyme C (^{1}M ^{2}K ^{3}A ... ^{147}G ^{148}V), a protein with 148 residues.[16]

Several attempts[1–4] have already been made to study aggregation of proteins without reference to specific proteins via simplified models with relatively small samples. Some of these studies[1,3,4] include 'minimalist' models involving a few short chains, and MD simulations[2] (replica exchange Langevin dynamics) with a coarse-grained representation (e.g. 27 peptides each with 7 residues to study fibrils, β-barrel, and aggregate structures).[2] Using a coarse-grained Monte Carlo simulation, Pandey and Farmer[5] have recently examined the multi-scale structures in self-assembly of histones (H3.1, a protein with 136 residues) with as many as 20 proteins in the simulation box. With the large-scale simulations, they have identified the appearance of a range of morphologies e.g from a solid-aggregate at low temperatures to the onset of long-range fibrous networks at high temperatures as a function of protein density. Using the same coarse-grained method,[5] we study the multi-scale morphology of lysozyme as a function of temperature and concentration. The model is presented in the next section, followed by results and discussion with a concluding remark towards the end.

II. MODEL

Residue-residue interactions play a critical role in modeling the structure and dynamics of a protein in coarse-grained descriptions of proteins. Though the fine-scale details of atom-atom interactions are not included, coarse-grained models allow for computational simulations of larger systems and for longer times. In our model, the lysozyme[16] protein is represented by a set of 148 nodes (each node representing a residue) tethered together by flexible covalent (peptide) bonds in a specific sequence on a cubic lattice. Unlike minimalist lattice models, there are several degrees of freedom for each residue to move and covalent bonds to fluctuate which can be enhanced further by fine-graining.[17] Even though the intra-molecular (atomistic) details of a residue are not explicitly included, amino acid specificity is captured via unique residue-residue interactions. Knowledge-based residue-residue (KBRR) interactions which are based on the statistical analysis of ensembles involving thousands of protein structures available in the Protein Data Bank (PDB) have been used extensively in modeling protein structures for decades.[18–25] The enormity of the literature available on this subject does not allow us to cite an exhaustive list; we therefore cite only a few for historical perspective and that are especially relevant to the work presented here. We have employed KBRR interactions[5,14] such as the classical Miyazawa-Jernigan contact matrix[19] in studying the structure and dynamics of individual proteins as well as self-assembly of peptides[14] and proteins.[5]

In our all-residue description of the protein chain, a residue is represented by a node that occupies a unit cube on a cubic lattice and can move to one of its 26 adjacent neighboring sites. The bond length between consecutive nodes varies between 2 and $\sqrt{10}$ in unit of the lattice constant, while maintaining the excluded volume constraints. Note that $\sqrt{7}$ is not feasible on such a cubic lattice and $\sqrt{8}$ is excluded to avoid the constraint of bond-crossing.[26] The constraint of the minimum distance of two lattice units between two nodes (inter- and intra-chain) requires excluded volume of the hard-core of a node to be 2^3 in unit of lattice constant. In an initial random configuration, a protein chain can occupy the lattice sites on the trail of a random walk of a cubic node with excluded volume constraints. Such a bond-fluctuation model has been used extensively in understanding the structure and dynamics (including the viscoelastic properties) of complex polymer systems; advantages and pitfall of such approaches are well explored in computational

polymer research.[26] We have used the bond-fluctuation representation of the peptide bonds with the input of KBRR interactions in several studies for analyzing the structure and dynamics of proteins,[5] peptides,[14] and bio-functionalized soft materials.[27] Because of the efficiency of the approach and the complexity of the lysozyme protein,[12,13,16] we adopt it here to explore the self-assembly of lysozymes by a large-scale simulation.

Each residue interacts with the surrounding residues (both inter- and intra-chain) as described by a generalized Lennard-Jones (LJ) potential,[5]

$$U_{ij} = \left[|\varepsilon_{ij}| \left(\frac{\sigma}{r_{ij}} \right)^{12} + \varepsilon_{ij} \left(\frac{\sigma}{r_{ij}} \right)^{6} \right], \quad r_{ij} < r_c \tag{1}$$

where r_{ij} is the distance between the residues at site i and j; $r_c = \sqrt{8}$ and $\sigma = 1$ in units of the lattice constant. Note that the range of interactions as defined by the cut-off r_c spans over many lattice sites to interact with surrounding residues (see figure S1[28]). To provide amino acids with a specific identity, the strength of the LJ potential ε_{ij} is chosen uniquely[5] for each residue-residue interaction pair, with appropriate positive (repulsive) and negative (attractive) values. As described above, we use the classic MJ interaction matrices[5,19] for the residue-residue pair interactions (ε_{ij}) included in the supplement.[28]

We analyze the structure and dynamics of protein chain aggregation for simulations with different numbers Nc (=1, 5, 10, 15) of chains in the simulation box. Protein chains (in their random configurations) are initially inserted in the box randomly and are moved around with the excluded volume constraints to randomize their distribution in preparing the sample.[5] Note that for such a long protein chain (148 residues), it becomes difficult[5] to insert many chains beyond a certain number as we approach a quasi-jamming physical limit. The computer processing unit time also increases with increasing the number of protein chains. Therefore, we restrict the number of chains that can be feasibly simulated with reasonable simulation time while capturing their cooperative and competing effects. The unit Monte Carlo step (MCS)[26] is defined as the attempt to move each residue (node) once, and varies depending on the size of the system, i.e., the number $Nc \times L_c$ of attempts to move randomly selected nodes in a simulation box with Nc protein chains each with L_c residues defines the unit MCS.

The Metropolis algorithm is used to move each node stochastically. For example, in order to move a randomly selected residue in a randomly selected protein chain from a site i to a site j, the excluded volume constraint and the limitations on changes in bond length are checked first for the proposed move. An attempt is then made to move the residue from site i to site j with the Boltzmann probability $exp(-\Delta E_{ij}/T)$, where ΔE_{ij} is the change in energy between its new (E_j) and old (E_i) configuration, $\Delta E_{ij} = E_j - E_i$. T is the temperature in reduced units (ε_{ij}/k_B) of the Boltzmann constant k_B and the residue-residue interaction energy energy (ε_{ij}).

As in previous studies,[5,14] we analyze a number of local and global physical quantities such as the energy of each residue and protein chain, their mobility, mean square displacement of the center of mass of the protein, radius of gyration, and its structure factor. These physical quantities are in arbitrary units, i.e., the length is in units of the lattice constant[5,14] which is different[5] from many all-atom simulations where realistic units for size and time scales are used via calibration of σ and ε_{ij} from experimental data for simplified systems. It is difficult to quantify physical quantities in absolute units due to the phenomenological nature of the interactions (Eq. (1)) used here. However, the trend in response of the physical quantities to such parameters as the temperature and network density should be qualitatively comparable with appropriate experimental observations.

Most of our simulations are performed on a 64^3 lattice, although different lattice sizes are used to assure that there is no significant finite size effect to alter our qualitative findings. Simulations are performed for *10* million MCS time steps with many independent simulations (*10-100*) depending on Nc. For example, *10* simulations with long runs (10^7 MCS) were performed for the largest $Nc = 15$, whereas *100* independent simulations with 10^7 MCS were performed with one chain ($Nc = 1$). Note that the chain number N_c corresponds to the occupied volume fraction (i.e., the concentration c) of protein: $c = 0.004$ ($Nc = 1$), 0.021 ($Nc = 5$), 0.042 ($Nc = 10$), and 0.063 ($Nc = 15$). These volume fractions appear relatively low, but are necessary because for chains

of size $L_c = 148$ (with minimum bond length *2* in units of the lattice constant) the percolation threshold and the jamming limit would be very low.

III. RESULTS AND DISCUSSION

We first examine the structure of a single lysozyme in the simulation box before analyzing the multi-scale structures and dynamics of many interacting protein chains. Figure 1 shows a set of illustrative snapshots in a temperature range with appreciable variations in the radius of gyration (see below) of the protein. Though reliable predictions cannot be based upon a few snapshots, they provide pictures of some of the conformational variability of the protein, segmental interaction, and local structures. For example, at the low temperatures (T = 0.015, 0.017), the protein remains relatively compact with globular segments. The protein expands on increasing the temperature which, reduces the size of the segmental globules considerably.

We have analyzed the simulations to ascertain information about how the dynamics of the protein chain as it performs its stochastic movement through a relatively large ensemble of configurations. Variation of the root mean square (RMS) displacement (R_c) of the protein chain with the time steps, as presented in Figure 2, provides an estimate of the dynamics for a range of temperatures (*T = 0.015 − 0.027*). At the low temperatures (*T = 0.015, 0.016*), the protein appears to reach a quasi-static state (with extremely slow motion) in the asymptotic time regime as it settles into its globular conformations after about two million time steps. At high temperatures (*T = 0.020 − 0.027*) on the other hand, it continues to diffuse with its fluctuating extended conformations. Raising the temperature from low to high enhances the motion of the protein systematically from an ultra-sub-diffusive (almost standstill) towards diffusive dynamics. The variation of the overall changes in the RMS displacement of the protein over the entire time steps with the temperature is provided in the inset of Figure 2, which seems consistent with the interpretation of fluorescence intensity variation with the temperature by Ow and Dunstan.[10] The variation of overall changes in R_c with the temperature (*T*) (inset Figure 2) also echoes the global size of the protein (see below) which shows that the structure and dynamics are correlated.

We have carried out extensive simulations with *Nc = 1, 5, 10*, and *15* of lysozyme chains to examine their cooperative and competing effects on self-organizing structures at a range of temperatures (*T = 0.015 − 0.021*). Figure 3 shows representative snapshots of the self-assembly with *Nc = 10*. Globular structures, similar to that of a single protein chain (Figure 1), form networks at the low temperatures. Increasing the temperature opens up individual protein structures and allows them to form spanning networks of fibrous and smaller globules. A network spanning all *10* lysozyme chains appears at the high temperature *T = 0.021*.

As the self-organizing structures appear, how do individual protein chains move, relax, and conform? Figure 4 shows the variation of the average RMS displacement of the center of mass of
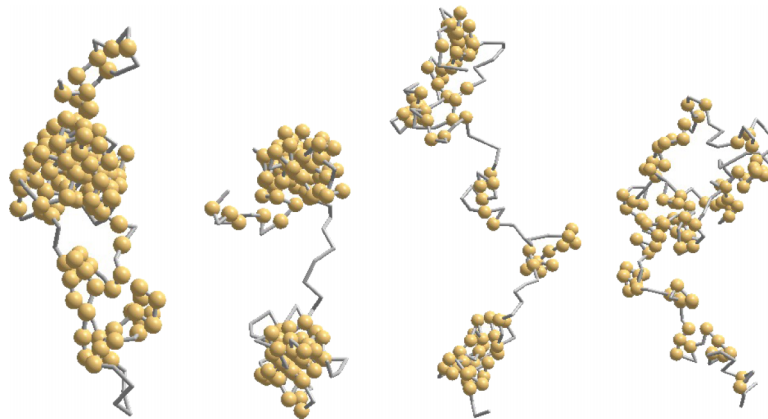

FIG. 1. Snapshots of the Lysozyme chain at the temperatures T = 0.015, 0.017, 0.019, and 0.021 (from left to right) towards the end of 10^7 time steps in a 64^3 simulation box. Spheres represent the interacting residues (other than adjacent residues along the backbone) within the range of interaction.
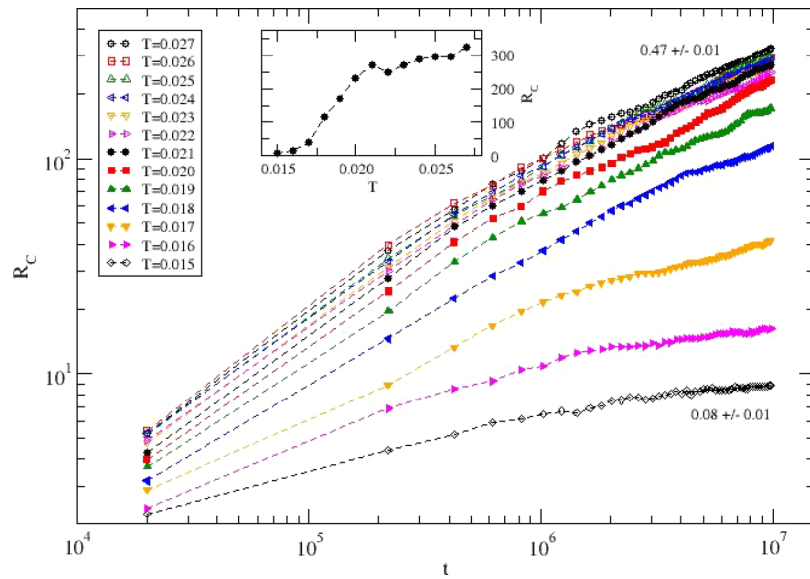
FIG. 2. Variation of the RMS displacement R_c of the center of mass of the protein chain with the time steps at temperatures $T = 0.015 - 0.027$. Variation of the total RMS displacement with the temperature is presented in the inset. Simulations for a single protein chain ($Nc = 1$) are performed on a 64^3 lattice for 10^7 time steps with 100 independent samples. Slopes of the asymptotic data sets at a low and a high temperature ($T = 0.015, 0.027$) are included for guide.

a protein with the time steps for the entire temperature range ($T = 0.015 - 0.021$). Despite some fluctuations, it is possible to identify the trend in dynamics of each protein chain on average, i.e. by examining the power-law dependence $R_c \propto t^\nu$ in the asymptotic time (t) regime. To our surprise, the protein chains begins to diffuse ($\nu \sim 1/2$) at low temperature ($T = 0.015$) in contrast to the extremely slow (nearly standstill $\nu \sim 0.1$) motion of a free protein chain at low temperature seen in Figure 2. This means that adding protein chains increases their mobility at low temperature while maintaining the network of globular structures. At high temperatures ($T = 0.021$) on the other hand, proteins slow down considerably with sub-diffusive ($\nu \sim 1/4$) dynamics as the protein chains elongate and entangle, constraining their movements. Again this is in contrast to diffusive dynamics of a single chain at the high temperatures (see Figure 2).

The variation of the equilibrium radius of gyration (R_g) with the temperature is included in the inset of Figure 5. The radius of gyration exhibits a continuous increase with the temperature before its saturation. There is a non-monotonic trend as the number of protein chains increases from *10* to *15* due to interference of fibrils at moderate to high temperatures. We see that the protein chains elongate on increasing the temperature in the low temperature regime. The radius of gyration also increases by increasing the number of protein chains (concentration) i.e. $Nc = 1, 5, 10$ until a certain limit beyond which (onset of jamming) it decreases, as is particularly evident with $Nc = 15$ in the high temperature regime. Concentration-induced elongation of protein chains enhances the formation of fibrils in the high temperature regime below a certain volume fraction, above which the radius of gyration decreases due to steric constraints i.e. crowding.

The global dynamics of protein chains emerges from a collective movement of the constitutive residues. The movement of each residue is controlled by its local surrounding, i.e. the interacting residues. In order to gain an insight into the stochastic movement of a residue along the backbone of a protein, one may estimate the probability (M_n) of its successful move per unit time step; M_n is a measure of its local mobility. The mobility profiles of residues at a low ($T = 0.015$) and a high ($T = 0.021$) temperatures are presented in Figures 6 and 7 with $Nc = 1, 5, 10$, and *15*. It is rather easy to identify the residues that are least mobile at low temperature. For example, the least mobile residues in an isolated protein ($Nc = 1$, Figure 6) [22]E, [31]K, [36]D, [51]K, [53]E, [67]D, [68]R, [71]D, [85]D, [87]K, [95]C, [99]C, [133]R, [138]D act like a nucleus (pinning center) for local coagulation of residues towards forming the globular structure of the protein. The pinning centers are mainly electrostatic residues (Asp (D), Glu (E), Lys (K), Arg (R)) along with a couple of hydrophobic residues (Cys (C)).
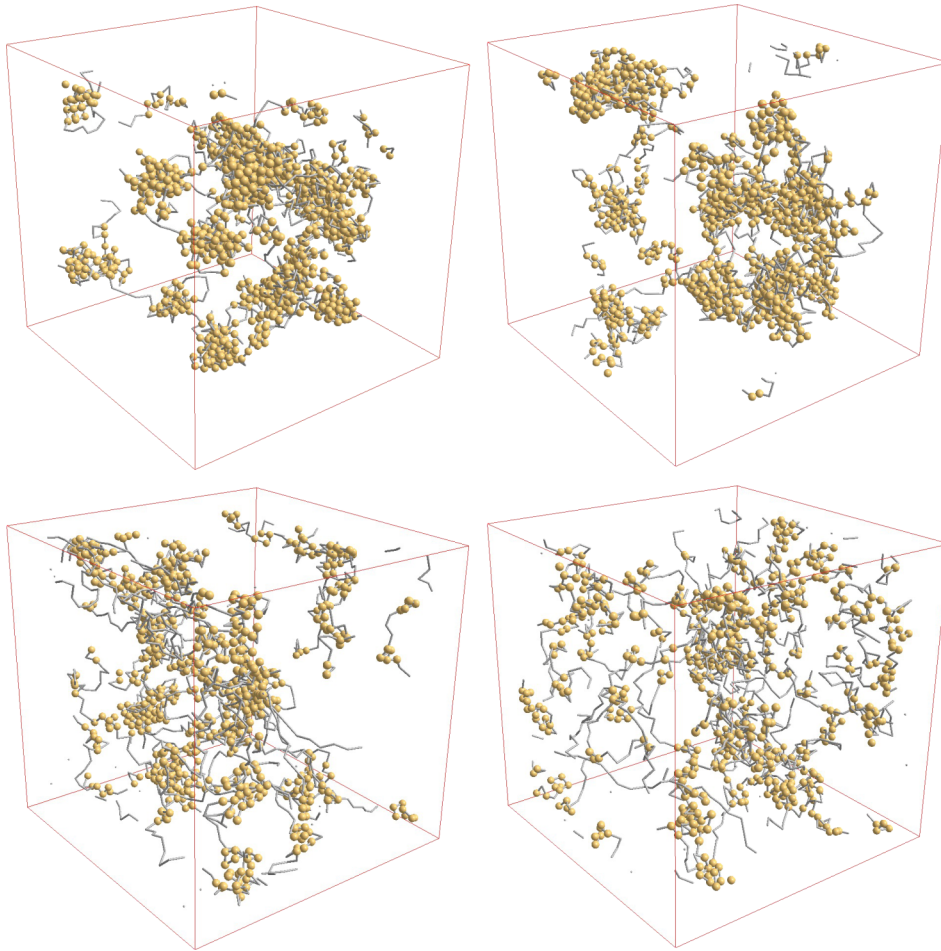
FIG. 3. Snapshots of the self-organized structures of 10 lysozyme chains at the temperatures T = 0.015 (top left), 0.017 (top right), 0.019 (bottom left), and 0.021 (bottom right) towards the end of 10^7 time steps in a 64^3 simulation box. Spheres represent the interacting residues within the range of interaction.

The pattern of the mobility profiles remain nearly the same (Figure 6) even with the additional interacting protein chains ($Nc = 5, 10, 15$) with a somewhat reduced mobility with $Nc = 15$.

The mobility profile at the high temperature (Figure 7) is very different from its low temperature counterpart (Figure 6). Almost all residues appear to be very mobile, few a few exceptions. Two residues (^{51}K, ^{53}E) appear to be quite immobile. Even though the mobility decreases with increasing number of chains and protein concentration, the identities of the relatively immobile residues remain the same as that of an isolated protein chain in this thermal-driven mobility profile.

The global physical properties such as overall size of the protein conformations depend on the local structures. Residue contact maps can provide some insight into the local surrounding that dictate the segmental movement and structures. Contact maps with $Nc = 1, 5, 10$, and 15 at low $T = 0.015$ (Figure 8) and high $T = 0.021$ (Figure 9) are presented. At the low temperature (Figure 7) we can see significant off-diagonal contacts for $Nc = 1$, signifying that residues from the opposite ends of the protein chain come in contact and form globular structures (see Figure 1). Adding more protein chains appears to reduce intra-chain looping, which may result in enhancing the elongation, i.e. larger radius of gyration. With many chains ($Nc = 5, 10, 15$), there may be many loops of various sizes (scatter maps in Figure 8) emerging from the dynamic network of protein chains which contribute to large scale morphology.

The scattering in the residue map reduces considerably at high temperature (Figure 9) for all Nc. For higher Nc ($=10, 15$), this is consistent with the formation of an extended, spanning network
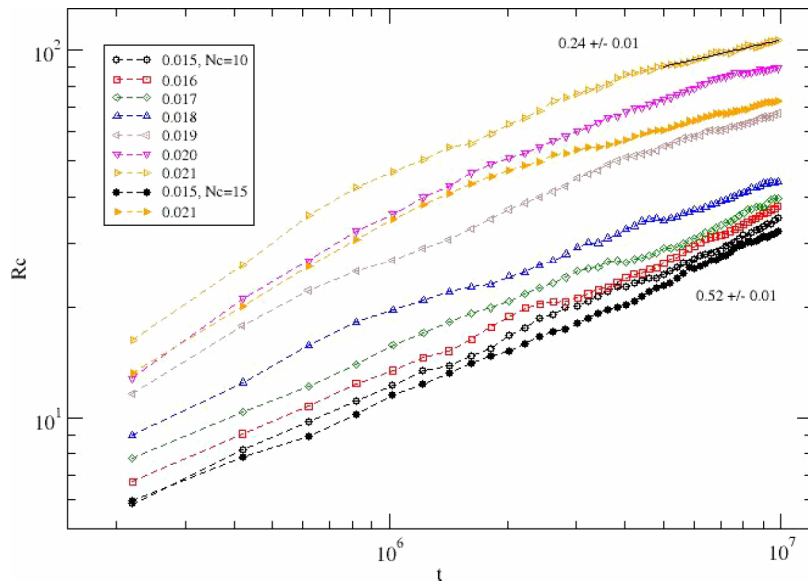
FIG. 4. Variation of the RMS displacement R_c of the center of mass of the protein chain with the time steps at temperatures $T = 0.015 - 0.021$ with $Nc = 10$; data sets with $Nc = 15$ for two temperatures ($T = 0.015, 0.021$) are included to see the trend. Simulations are performed on a 64^3 lattice for 10^7 time steps each with $10 - 20$ independent samples (lower samples with larger number of protein chains).
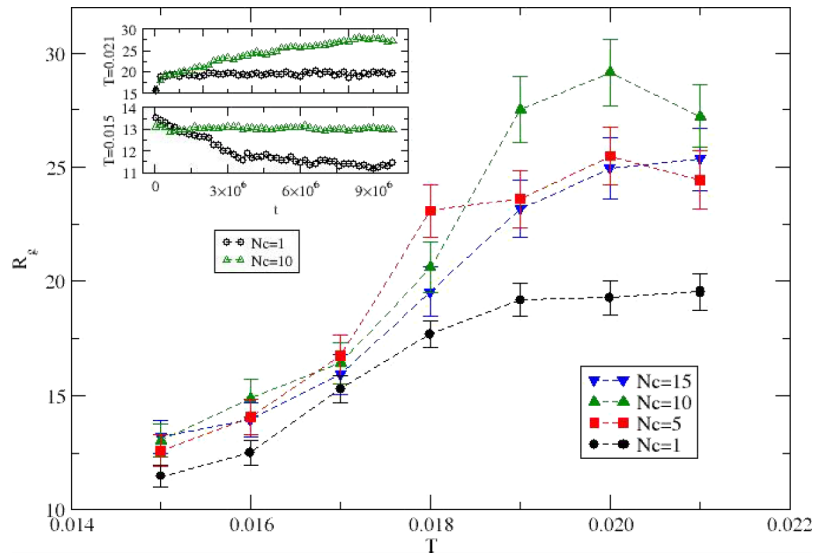

FIG. 5. Variation of the equilibrium radius of gyration (R_g) with the temperature (T) for $Nc = 1, 5, 10$, and 15. Simulations are performed on a 64^3 lattice for 10^7 time steps each with $10\text{-}100$ independent samples (lower samples with larger number of protein chains). The inset figure is the variation of R_g with the time step (t) for two representative systems ($Nc = 1, 10$) at two temperatures ($T = 0.015, 0.021$).

of protein structures involving fibrils as seen in Figure 3 (e.g.; residues from different proteins assemble into a heterogeneous structure with multi-scale structures including tenuously connected fractals.)

In order to examine structural evolution that spans multiple length scales (i.e., beyond the size (R_g) of a protein in the simulation box) we analyze the structure factor $S(q)$,

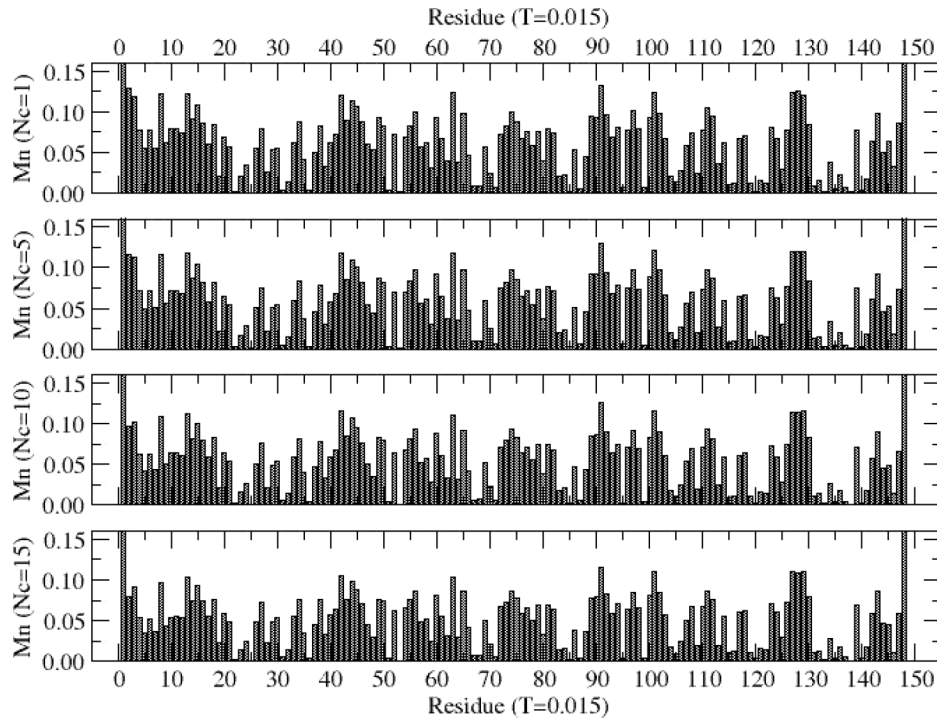$$S(q) = \langle \frac{1}{N} \left| \sum_{j=1}^{N} e^{-i\vec{q} \cdot r_j} \right|^2 \rangle_{|\vec{q}|} \tag{2}$$

FIG. 6. Mobility (M_n, probability of a successful move) for each residue per unit time step, at low temperature $T = 0.015$ for $Nc = 1, 5, 10$, and 15. These data represents the average mobility of each residue. Simulations are performed on a 64^3 lattice for 10^7 time steps each with 10-100 independent samples (lower samples with larger number of protein chains).

FIG. 7. Mobility (probability of successful move of each residue per unit time step) at high temperature $T = 0.021$ with $Nc = 1, 5, 10$, and 15. These data represents the average mobility of each residue. Simulations are performed on a 64^3 lattice for 10^7 time steps each with 10-100 independent samples (lower samples with larger number of protein chains).
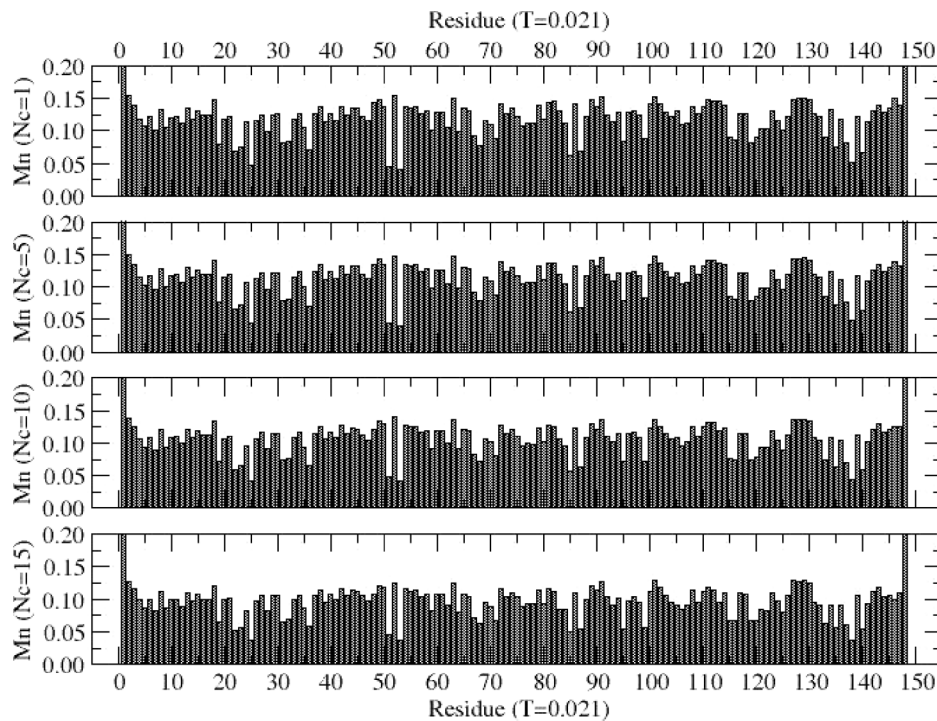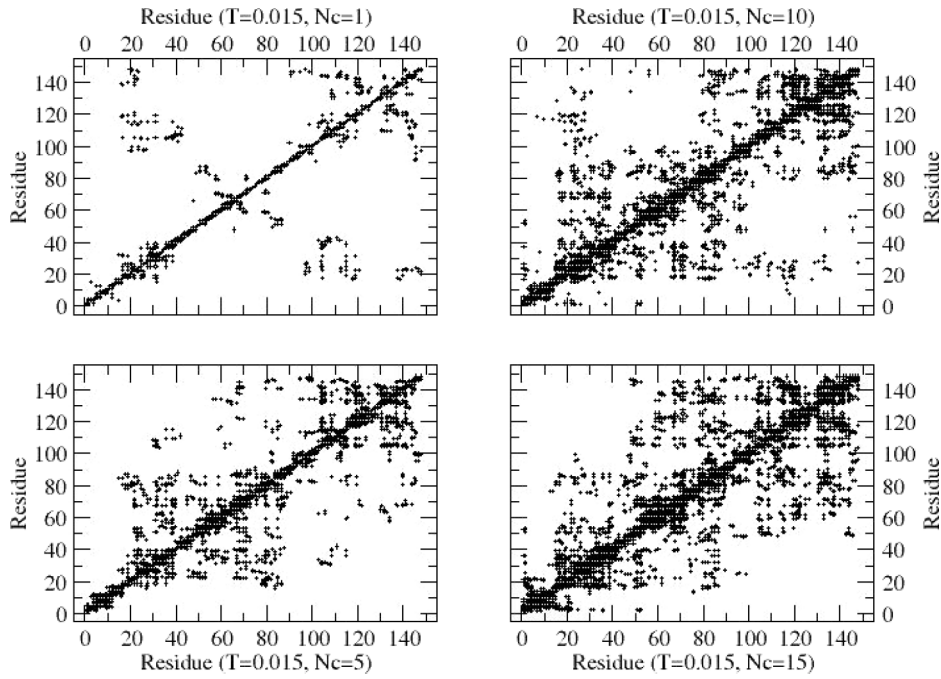
FIG. 8. A typical residue contact map, i.e. for each residue, a plot of the surrounding residues within the range of interaction. $T = 0.015$ with $Nc = 1, 5, 10$, and 15. These data represents the residues at the end of 10^7 time steps in a 64^3 simulation box.
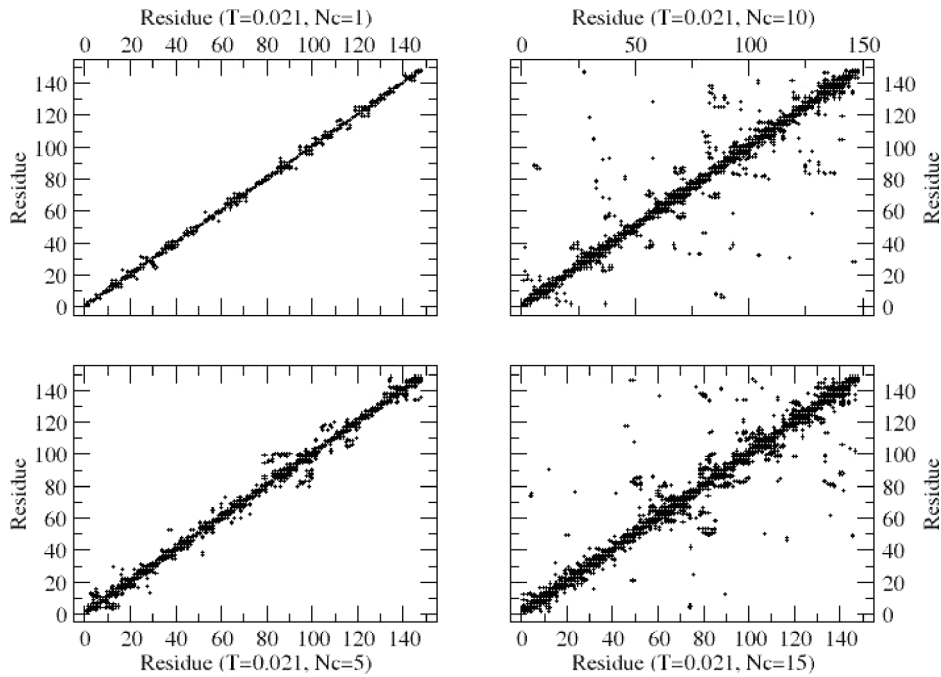

FIG. 9. A typical residue contact map, i.e. the residue and its surrounding residues within the range of interaction at high temperature $T = 0.021$ with $Nc = 1, 5, 10$, and 15. These data represents the residues at the end of 10^7 time steps in a 64^3 simulation box.

where r_j is the position of each residue regardless of its affiliation with the protein chain and $|q| = 2\pi/\lambda$ is the wave vector of wavelength λ. Using a power-law scaling of the structure factor,

$$S(q) \propto q^{-1/\gamma} \tag{3}$$

one may study the spread of residues over the length scale λ by evaluating the exponent γ which describes the mass (residue) distribution. The effective dimension (D) of the distribution of mass M (number of correlated residues) over its radius of gyration (R) may be evaluated from a scaling, $R \propto M^{\gamma}$ with $D = 1/\gamma$.

Figure 10 shows the variation of the structure factor $S(q)$ with the wave vector q with Nc = 1, 5, 10, and 15 for a range of temperatures ($T = 0.015 - 0.027$). As seen in Figure 1, the structure of an isolated protein chain ($Nc = 1$) is globular at low temperature and approaches random coil (self-avoiding walk, SAW) at higher temperature. This is confirmed through the calculation of D using $S(q)$. The radius of gyration $R_g \sim 12$ at $T = 0.015$, and $R_g \sim 20$ at $T = 0.021$ (Figure 5). At $T = 0.015$, the scaling of $S(q)$ with the wave vector q around $q \sim 0.5 - 0.6$ ($\lambda \sim R_g$) gives $D = 1/\gamma \sim 3.46$ (Figure 10) which shows that the structure of the protein is solid-like (i.e. globular). At the high temperature $T = 0.021$, on the other hand, the scaling of $S(q)$ around $q \sim 0.3$ ($\lambda \sim R_g$) leads to $D \sim 1.78$, an exponent associated with SAW linear structure.

The structure factor pattern changes systematically on increasing the protein concentration ($Nc = 5, 10, 15$). The appearance of oscillations in the variation of $S(q)$ with q in particular is an indication of ordering which set-in as the proteins elongate and form the network (due to both interaction and entanglement). The mass distribution over the entire sample ($T = 0.021$, $Nc = 5, 10$, 15) appears to be tenuous (Figure 10), with an effective dimension $D \sim 1.7$ similar to that of an SAW. However, unlike a standard linear structure which dominates the large-scale connectedness of the fibrils, there are small aggregates (self-assembled structures with many residues that appear to glue together via non-covalent interactions) (see Figure 3, 8). With many interacting proteins at the low temperatures ($T = 0.015$, $Nc = 15$), our data (Figure 10) shows a dense (globular) structure at small scale ($D \sim 2.9$) and relatively less dense morphology at large scale ($D \sim 2.3$). A smooth transition from small scale dense structure to larger scale tenuous network of fibrils seems to persist
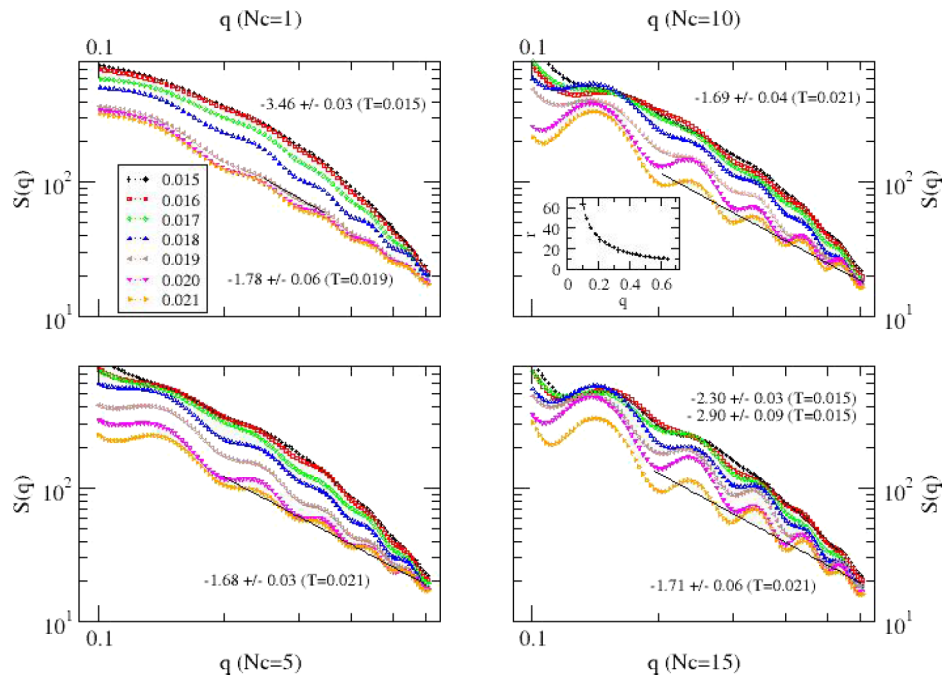

FIG. 10. Variation of the structure factor $S(q)$ with the wave vector q for a range of temperature ($T = 0.015 - 0.021$) with $Nc = 1, 5, 10$, and 15. Inset in the plots for $Nc = 10$ is to guide the conversion between the wave vector (q) and the spatial (r) length scales. Simulations are performed on a 64^3 lattice for 10^7 time steps each with 10-100 independent samples (smaller number of samples with larger number of protein chains). Slopes of the regressive fits of representative data sets are included: at the low temperature $T = 0.015$, with $Nc = 1$, slope of the data sets in the wave vector range comparable to its radius of gyration is about -3.46 and with $Nc = 15$, the slope is about -2.90 for the data sets in the wave vector range comparable to its radius of gyration and -2.30 on a larger scale. Slope at the high temperature ($T = 0.021$) is about -1.7 over the entire length scale with $Nc = 1, 5, 10$, and 15.

on increasing the temperature with many interacting protein chains. The amplitude of structural oscillation (a measure of correlation) appears to increase with the protein density.

IV. CONCLUSION

The self-organizing dynamics of lysozymes (an amyloid protein) are studied by a coarse-grained Monte Carlo simulation with input from knowledge-based residue-residue interactions.[19] Because of the efficiency of our coarse-grained approach that involves phenomenological residue-residue interactions, we are able to explore the effects of temperature and protein concentrations on a fairly large system (chains of 148 residues, and as many as 15 chains) for time scales long-enough to observe self-organizing aggregations in a reasonable span of time (on the order of a couple of weeks on a standard contemporary cluster). The simulations are carried out for 10^7 time steps, each at a range of temperatures ($T = 0.015 - 0.021$) involving a number of independent samples (*10-100*). A number of interesting observations are made concerning: (*i*) the difference in dynamics of an isolated chain versus many interacting chains (see below) and (*ii*) the onset of long range correlation with the self-assembly of a network spanning all chains.

Based on the variation of the RMS displacement of the center of mass of the protein chain with the time steps (i.e. $R_c \propto t^\nu$ in the asymptotic time (t) regime), the isolated lysozyme appears to move extremely slowly (quasi-static, $\nu \rightarrow 0$) at low temperatures and speeds up systematically with increasing temperature towards its diffusive dynamics asymptotically. In contrast, the presence of interacting proteins (i.e. with $Nc = 10$) leads to a concentration-induced protein chain diffusion ($\nu \sim 1/2$) at low temperatures (e.g. $T = 0.015$) and concentration-tempered sub-diffusion (($\nu \sim 1/4$)) at high temperatures (e.g. $T = 0.021$).

The radius of gyration of the isolated protein ($Nc = 1$) exhibits a systematic transition from a globular (low T) to a random coil (high T) conformation with increasing temperature. The crossover from globular to random coil becomes sharper upon increasing the protein concentration (i.e. with $Nc = 5, 10$), with larger R_g at higher temperatures where elongation of protein conformation appears to be concentration driven. Adding more protein chains (e.g. with $Nc = 15$, crowded regime) leads to a decrease in the magnitude of R_g, but still larger than that of an isolated protein.

Visual inspections of the snapshots and animations reveal a range of diverse structures over multiple scales that depend on both the temperature and the protein concentration. Self-assembly of proteins' segments into globular form persists at the low temperatures (i.e. $T = 0.015, 0.016$) with isolated aggregates at dilute protein concentration, and a connected network of globular aggregates at relatively high protein concentrations ($Nc = 10, 15$). Increasing the temperature causes appreciable multi-scale responses in structures of a protein (expanded protein with larger R_g) to large-scale (much larger than the size of a protein) morphology of its self-assembled network. At high temperature ($T = 0.021$), the self-assemble fibril network is easy to identify visually (snapshots) at higher protein concentrations ($Nc = 10, 15$).

Detailed analysis of the structure factor $S(q)$ provides a quantitative measure of mass distribution in the self-organized heterogeneous network over multiple length scales. Scaling of the structure factor with the wave vector is used to estimate the effective dimension D of the network mass over these length scales. We know the magnitude of the index D for standard structures, i.e. $D = 3$(solid), $5/3$ (random walk). We have found that an isolated lysozyme chain conforms to a globular structure ($D \geq 3$) at low temperature, and a random coil ($D \sim 1.7$) at high temperatures. With many interacting proteins, at the low temperature $T = 0.015$, we find different mass distributions at different length scales (implying different structures), i.e. $D \sim 2.9$ on the scale comparable to the radius of gyration of a single chain, and $D \sim 2.3$ at the large scale over the entire sample. The network of fibrils, at high temperature ($T = 0.021$) is rather tenuous, i.e. $D \sim 1.7$ on a large-scale with a heterogeneous distribution of small (micro) globules that may be very important in propagating the long range correlation over the spanning cluster.

Aggregation of fibrils is the hallmark of amyloids. We have identified the structures that span multiple scales from the radius of gyration of an individual lysozyme to the global amyloid network as a function of temperature and the protein concentration. We hope this study will stimulate further investigations and help in interpreting and understanding experimental data.

ACKNOWLEDGEMENTS

[1] T Cellmer, D Bratko, JM Prausnitz, and H. Blanch, "The competition between protein folding and aggregation: off-lattice minimalist model studies," Biotechnol Bioeng. **89**, 78-87 (2005).

[2] G Bellesia and JE Shea, "Effect of beta-sheet propensity on peptide aggregation," J Chem Phys. **130**, 145103 (2009).

[3] L Zhang, D Lu, and Z. Liu, "How native proteins aggregate in solution: a dynamic Monte Carlo simulation," Biophys Chem. **133**, 71-80 (2008).

[4] MT Oakley, JM Garibaldi, and JD Hirst, "Lattice models of peptide aggregation: evaluation of conformational search algorithms," J Comput Chem. **26**, 1638-46 (2005).

[5] RB Pandey and BL Farmer, "Aggregation and network formation in self-assembly of protein (H3.1) by a coarse-grained Monte Carlo simulation," J. Chem. Phys. **141**, 175103-1-175103-9 (2014).

[6] JB GC, YR Bhandari, BS Gerstman, and PP Chapagain, "Molecular Dynamics investigations of the α-helix to β-barrel conformational transformation in the RfaH transcription factor," J. Phys. Chem. B **118**, 5101-5108 (2014).

[7] A Harada, H Azakami, and A Kato, "Amyloid fibril formation of hen lysozyme depends on the instability of the C-helix (88-89)," Biosci. Biotechnol. Biochem. **72**, 1523-1530 (2008).

[8] A Cao, Daoying Hu, and L Lai, "Formation of amyloid fibrils from fully reduced hen egg white lysozyme," Protein Science **13**, 319-324 (2004).

[9] Y Takunaga, Y Sakakibara, Y Kamada, K-I Watanabe, and Y Sugimoto, "Analysis of core region from egg white lysozyme forming amyloid fibrils," Int. J. Biol. Sci. **9**, 219-227 (2013).

[10] S-Y Ow and DE Dunstan, "The effect of concentration, temperature and stirring on hen egg white lysozyme amyloid formation," Soft Matter **9**, 9692-9701 (2013).

[11] S Balme *et al.*, "Structure, orientation and stability of lysozyme confined in layered materials," Soft Matter **9**, 3188-3196 (2013).

[12] DR Booth *et al.*, "Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis," Nature **385**, 787-793 (1997).

[13] See http://en.wikipedia.org/wiki/Amyloid for general description. (as of February 9, 2015).

[14] RB Pandey, Z Kuang, and BL Farmer, "A hierarchical coarse-grained (all atom to all residue) computer simulation approach: self-assembly of peptides," PLoS one **8**, e70847-1-e70847-8 (2013).

[15] PL Freddolino, CB Harrison, Y Liu, and K Schulten, "Challenges in protein-folding simulations," Nature Physics **6**, 751-758 (2010).

[16] See http://www.uniprot.org/uniprot/P61626 for the sequence of the lysozyme.

[17] RB Pandey and BL Farmer, "Conformational response to solvent interaction and temperature of a protein (histone h3.1) by a multi-grained Monte Carlo simulation," PLoS one **8**, e76069-1-e76069-9 (2013).

[18] S Tanaka and HA Scheraga, "Medium and long range interaction parameters between amino acids for predicting three dimensional structures of proteins," Macromolecules **9**, 945-950 (1976).

[19] S Miyazawa and RL Jernigan, "Estimation of effective inter residue contact energies from protein crystal structures: quasi-chemical approximation," Macromolecules **18**, 534–552 (1985).

[20] S Miyazawa and RL Jernigan, "Residue-residue potentials with a favorable contact pair term for simulation and treading," J Mol Biol **256**, 623-644 (1996).

[21] Web site for residue-residue interaction tables: http://gor.bb.iastate.edu/potential/, courtesy of the research group of R.L. Jernigan.

[22] Z Bagci, A Kloczkowski, RL Jernigan, and I Bahar, "The origin and extent of coarse grained irregularities in protein internal packing I," Proteins **53**, 56-67 (2003).

[23] MR Betancourt and D Thirumalai, "Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes," Protein Sci **2**, 361-369 (1999).

[24] VN Maiorov and GM Crippen, "Contact potential that recognizes the correct folding of globular proteins," J Mol Biol **227**, 876-888 (1992).

[25] A Godzik, A Kolinski, and J Skolnick, "Knowledge-based potentials for protein folding: what can we learn from protein structures?," Proteins **4**, 363-366 (1996).

[26] *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*, edited by K. Binder (Oxford University Press, New York, 1995).

[27] RB Pandey and BL Farmer, "Distinction in binding of peptides (P2E) and its mutations (P2G, P2Q) to a graphene sheet via a hierarchical coarse-grained Monte Carlo simulation," J. Chem. Phys. **139**, 164901-1-164901-6 (2013).

[28] See supplementary material at http://dx.doi.org/10.1063/1.4921074 for the residue-residue interaction table (from Ref. 19), range of interaction in LJ interaction.