Massively parallel molecular dynamics simulations of lysozyme unfolding

R. Zhou
M. Eleftheriou
C.-C. Hon
R. S. Germain
A. K. Royyuru
B. J. Berne

We have performed molecular dynamics simulations for a total duration of more than 10 μs (with most molecular trajectories being 1 μs in duration) to study the effect of a single mutation on hen lysozyme protein stability and denaturing, using an IBM Blue Gene/L™ supercomputer. One goal of this study was to assess the use of certain force fields to reproduce experimental results of protein unfolding using thermal denaturing techniques. A second and more important goal was to gain microscopic insights into the mechanism of protein misfolding using both thermal and chemical denaturing techniques. We found that the thermal denaturing results were robust and reproducible with various force fields. The chemical denaturing results explained why the single amino-acid mutation on residue Trp62 causes the disruption of long-range interactions in the tertiary structure. Simulation results revealed that the Trp62 residue was the key to a cooperative long-range interaction within the wild-type protein. Specifically, Trp62 acts as a bridge between two neighboring basic residues through a π-type H-bond or π–cation interaction to form an Arg-Trp-Arg "sandwich-like" structure. Our findings support the general conclusions of the experiment and provide an interesting molecular depiction of the disruption of the long-range interactions.

Introduction

Many fatal diseases, such as Alzheimer's disease, are associated with the formation of amyloid fibers, caused by protein misfolding. Understanding the biochemical mechanism behind protein misfolding is, therefore, one of the most urgent and challenging problems remaining in molecular biology, partly because the percentage of aging people in the world is increasing at an unprecedented pace [1–6]. Acquiring insights into the misfolding of proteins requires a detailed understanding of their three-dimensional (3D) structure and dynamics at the atomic level. This understanding can be significantly improved by combining sophisticated experiments with large-scale computer modeling.

Earlier experimental studies pioneered by Dobson and coworkers have shown that amyloids and fibrils can be formed not only from the traditional beta-amyloid peptides but also from almost any proteins, such as lysozyme, given the appropriate conditions [1–4]. This finding has opened a new and exciting window of research into the mechanism behind Alzheimer's disease and other amyloidoses [1] related to protein misfolding. Remarkably, recent evidence suggests that the misfolding and subsequent aggregation and amyloid formation can be induced by only a single amino-acid mutation. Obviously, when a single mutation triggers a protein to misfold or lose key contacts in the tertiary (i.e., 3D) structure, the mutated amino-acid residue must be located at a critical position in the sequence and structure. This is the case for the hen egg white lysozyme, as recently investigated by experiments [2, 5] that showed that a single mutation can cause the protein to misfold and lose essential long-range interactions that are present in the wild-type protein. This mutation is designated "W62G," which refers to a mutation of a Trp amino acid to a Gly amino acid at the location of residue 62. (In the field of biochemistry, the term *residue* commonly refers to an amino acid.) The most striking finding is that the mutation site designated "Trp62" is on the surface of the native protein, but not in the hydrophobic core. This raises an interesting question about the role of this surface hydrophobic residue. This W62G mutation is

0018-8646/08/$5.00 © 2008 IBM

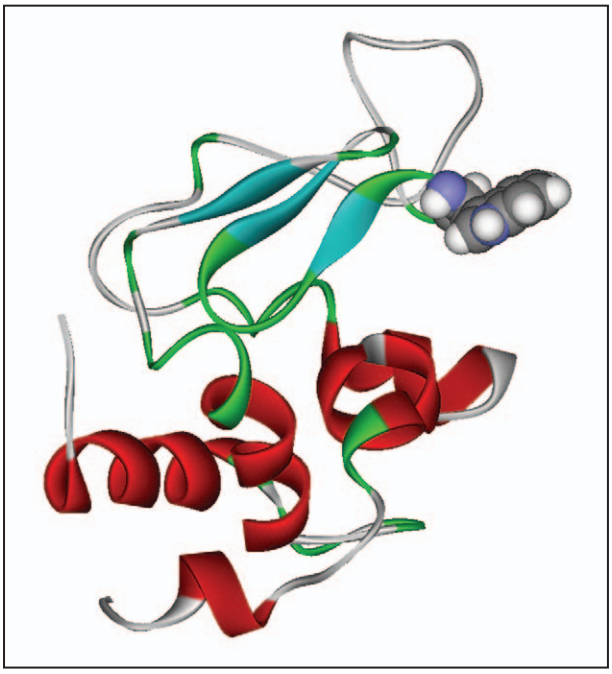


Figure 1

A ribbon view of the native (i.e., wild-type) lysozyme protein structure, with residue Trp62 represented in van der Waals balls. The alpha-helices are colored in red and beta-sheets in light blue.

believed to induce unfolding of the protein because of the loss of key long-range hydrophobic interactions otherwise present in the wild-type protein during the folding nucleation process [2, 5]. We wanted to better understand how this Trp62 amino-acid residue plays a key role in such long-range interactions during the early stage of the folding process and then shifts to the surface for functional reasons [2]. Given the importance of this phenomenon, it is of great interest for us to further investigate this, with the goal of arriving at a clear molecular picture of the mechanism.

Biomolecular computer simulations, enabled by massively parallel supercomputers such as the IBM Blue Gene/L* (BG/L) platform [7], promise to bridge the gap between the currently available *simulation* timescale and the *experimental* timescale for many important protein folding processes. We believe that the combination of sophisticated experiments and large-scale molecular simulations with massively parallel supercomputers, such as the BG/L platform [7], will help us better understand protein folding and misfolding mechanisms [8–15]. Computer simulations can be performed at various levels of complexity, ranging from simple lattice models [16] (i.e., simplified representations of the protein conformation space in which each amino acid is represented by a single point) to coarse-grained models [17] to all-atom models with explicit solvent representations. Such simulations can be used to supplement experiments and provide missing information with respect to protein folding pathways and intermediates, which are often inaccessible even with the current most sophisticated experimental approaches [7–10]. However, the realistic all-atom molecular modeling of systems of biological interest typically requires a significant amount of computing power and efficient software, because protein folding events usually occur on the order of microseconds to milliseconds. On the other hand, molecular dynamics simulations are limited (by the vibrational frequencies of the atoms) to a typical timestep size on the order of femtoseconds. Thus, it is extremely challenging to simulate the entire kinetic process of a protein folding and misfolding utilizing conventional applications on conventional computers such as a personal computer.

In our current work, we have performed molecular dynamics simulations on a previously unprecedented scale to elucidate the mechanism by which the W62G mutation induces misfolding in the lysozyme, and to provide a molecular explanation as to why this mutation has such a surprising destabilizing effect on the tertiary structure of the protein. Our basic approach is to use the BG/L to simulate both the thermal denaturing and the chemical denaturing (in 8M urea solution) for the wild-type and mutant lysozyme. The lysozyme system studied here proves to be a good example of how large-scale simulations can elucidate how a single mutation can cause protein misfolding.

Results and discussions

In our simulations, the starting structure of the wild-type version of hen lysozyme comes from the crystal structure deposited in the Protein Data Bank (PDB) (file *193L.pdb*), as shown in **Figure 1**. This version of lysozyme protein contains two structural domains, the alpha-domain, including residues 1 through 35 and 85 through 129, and the beta-domain, comprising residues 36 through 84. The protein has four alpha-helices [Helix A (5–14), Helix B (25–36), Helix C (90–100), and Helix D (110–115)], two beta-strands [Strand 1 (43–46) and Strand 2 (51–54)], a loop (60–78) region, and a 3_{10}-helix (81–85). The mutation site Trp62 is in the loop region. The starting structure for the mutant is generated by a single residue replacement, W62G, from the wild-type structure (with re-equilibration as described in the Appendix). The resulting protein configurations are then solvated in water and in an 8M urea solution for thermal denaturing and chemical denaturing simulations, respectively (see the Appendix).

Thermal denaturing

We have performed thermal denaturing molecular dynamics simulations using both OPLSAA (Optimized Potential for Liquid Simulations, All Atoms) [18] and CHARMM (Chemistry at Harvard Macromolecular Mechanics) [19] force fields to investigate whether the results are sensitive to the force field used. Our simulations indicate that both force fields give fairly consistent results, with OPLSAA showing more stable protein structures overall and, thus, a slightly higher temperature is needed to unfold the protein within the same amount of simulation time. In the following discussion, we focus on the CHARMM results. The OPLSAA results have been reported previously [20]. Both the wild-type and the W62G mutant lysozyme are simulated first at 300 K in order to observe whether the protein structure stays folded during our simulation length. Indeed, at 300 K, we observe that the backbone root-mean-square deviation (RMSD) stays below about 3 Å from the native crystal structure during the 15-ns simulation [**Figure 2(a)**], indicating that the force field is reasonable in terms of the protein stability. The OPLSAA force field shows similar results, and the lysozyme protein is fairly stable at 300 K [20]. This stability is consistent with the previous simulations on human lysozyme at 300 K by Moraitakis and Goodfellow [21].

Figure 2(a) also shows a comparison of a representative backbone RMSD trajectory for both the wild-type and the mutant lysozyme at 400 K. (Three trajectories are run for each case.) Overall, these trajectories show a steady increase in RMSD during the 15-ns simulation. Even though this representative trajectory does show the mutant having a slightly higher RMSD than the wild type, the differences in other trajectories are not as dramatic. As pointed out in previous work [22], the RMSD may not be a useful measure of the local structures when the RMSD values exceed a certain value, for example, 8 Å. In fact, comparable RMSD values may show very different local contacts. Nevertheless, the high RMSD values (>10 Å) indicate that the protein structures are significantly denatured after 15 ns at these high temperatures. It is interesting to note that our results are consistent with those of Moraitakis and Goodfellow [21]. These authors also found comparable RMSDs in their 5-ns thermal denaturing simulations at 500 K with a GROMOS96** force field for the wild type and mutant of human lysozyme [21]. [GROMOS96 (GROningen MOlecular Simulation) is a force field that describes parameter sets of potential energy functions.]

Because of the above caveats associated with RMSD calculations, we computed the root-mean-square fluctuations (RMSFs) of each residue (represented by its alpha-carbon atom, often denoted Cα in this paper), based on all three trajectories for both the wild-type and the mutant lysozyme, to characterize the local fluctuations. **Figure 2(b)** shows a comparison of the RMSF for the wild-type and the mutant lysozyme. Although the RMSDs do not show a significant difference, the RMSFs do exhibit noticeable differences. During unfolding, we observe that the mutant tends to have higher fluctuations in some of the regions. The difference is most prominent in the beta-domain and specifically in the loop region where the mutation site,

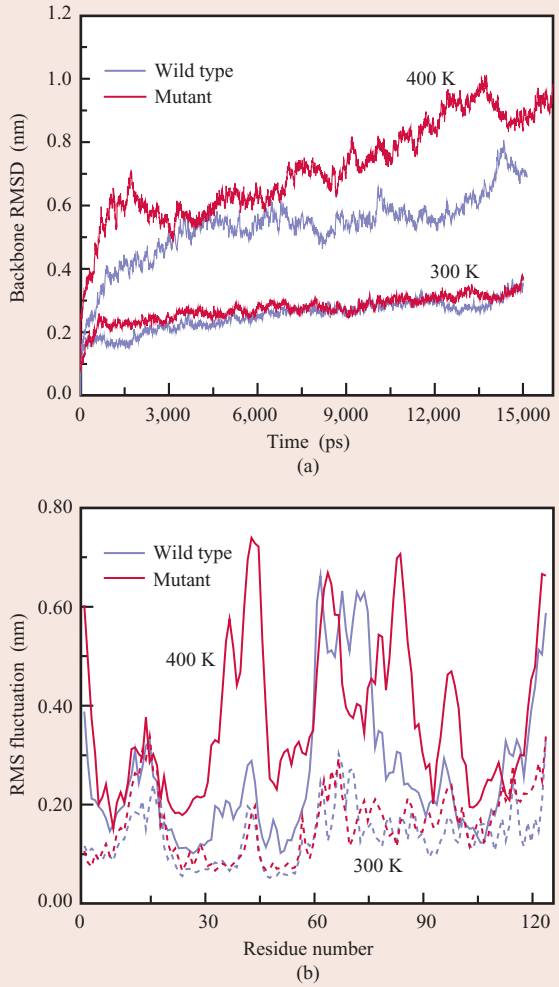


Figure 2

(a) Comparison of the backbone RMSD for the wild-type and mutant lysozyme from one representative molecular dynamics trajectory. (b) Comparison of the RMS fluctuation for the wild-type and mutant lysozyme. The results are obtained from the 400 K NVT (constant volume and temperature) simulations with the CHARMM force field. In this case, both the backbone RMSD and the RMS fluctuation results show that the mutant lysozyme has much larger deviations from the initial crystal structure.

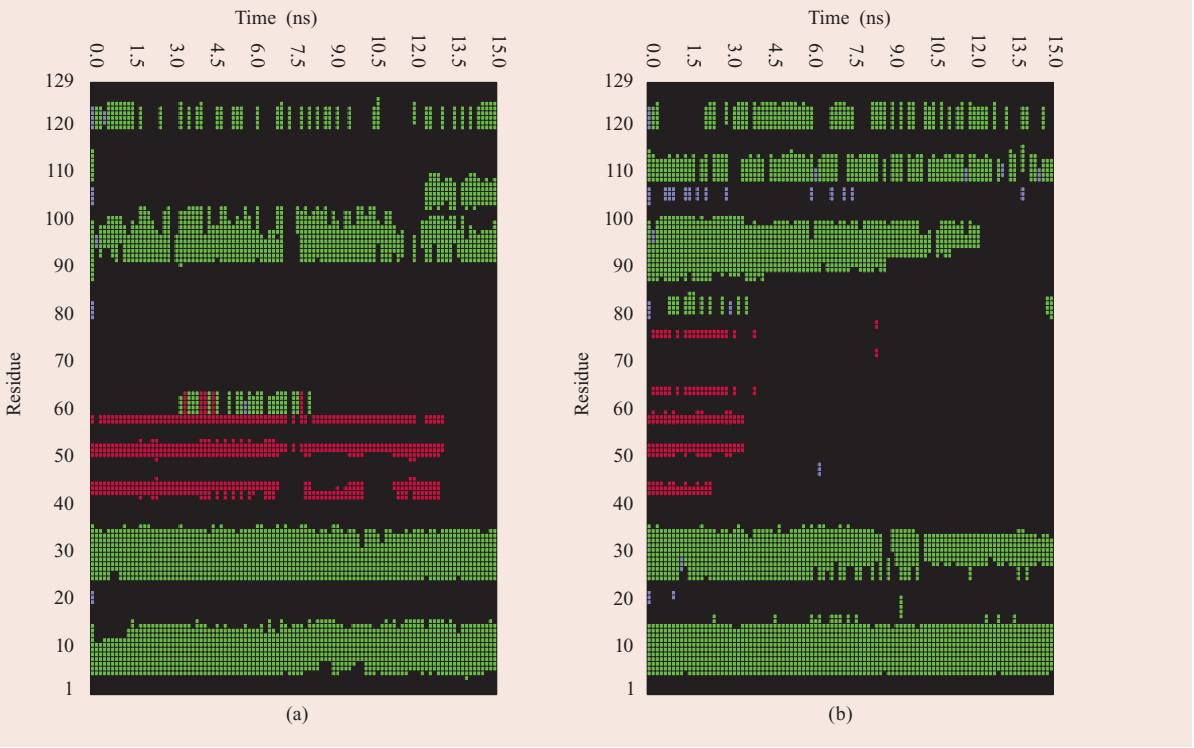


Figure 3

Time-evolution of the secondary structure at 400 K with the CHARMM force field. (a) Wild-type protein. (b) W62G mutant protein. The secondary structure is assigned by the program STRIDE, with alpha-helices colored green, 3_{10}-helices light blue, beta-strands red, and coils and turns black. The secondary structure of the starting crystal structure is displayed at $t = 0$ ns, corresponding to Helix A (5–14), Helix B (25–36), Helix C (90–100), Helix D (110–115), Strand 1 (43–46), Strand 2 (51–54), and a 3_{10}-helix (81–85).

Trp62, resides. This behavior should be expected since the hydrogen bonding network connecting this loop region with the two beta-strands, Strand 1 (residues 43–46) and Strand 2 (residues 51–54), in the wild type is disrupted by the mutation, thus leading to a more unrestrained motion of the loop.

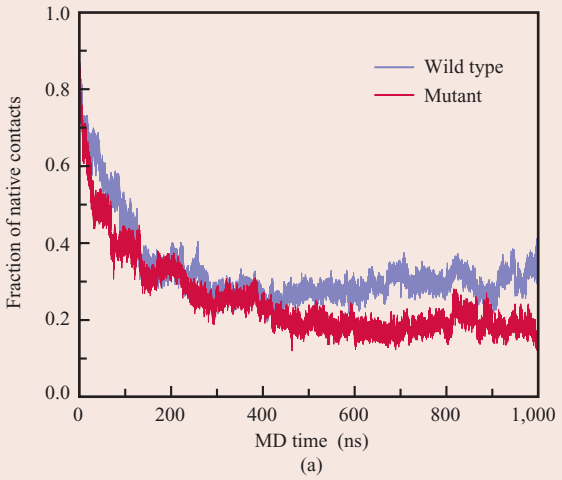
Figure 3 shows the time-evolution of the protein secondary structure, following an approach similar to that of de Bakker et al. [23], for one representative trajectory of the wild-type and mutant lysozymes at 400 K. The secondary structures are obtained from the program STRIDE (STRuctural IDEntification) [24]. It should be noted that at 300 K, a fairly stable trajectory is found for all secondary structural components in both the wild type and the mutant, with the mutant showing slight disruptions in the beta-strands. However, at 400 K, the mutant shows a large disruption in the secondary structures after about 3 ns. Note that the wild-type protein also shows larger disruptions at 400 K compared to 300 K. Overall, most of the disruptions in the mutant are in the beta-domain region of the protein as well as the Helix C (residues 90–100) in the alpha-domain region. Even though each trajectory shows slightly different behavior, overall, they display a reasonably good consensus in terms of secondary structure deformation. These disruptions of secondary structures, as well as the higher residue fluctuations, decrease stability in the mutant and seem to generally agree with the experiment in which researchers [2] found that the single mutation W62G causes the tertiary contacts (some of which are long range) in the wild type to disappear in a highly denaturing 8M urea solution. However, the researchers did not identify the origin or the order of the disruptions for the misfolding of hen lysozyme, probably because of the limited experimental resolution.

A detailed analysis of the lysozyme tertiary structures reveals how the mutant loses its native-like contacts during the thermal denaturing process. The first major unfolding event happens in about 3 ns, when two beta-strands Strand 1 (residues 43–46) and Strand 2 (residues

51–54) start to disappear (i.e., they start to lose their secondary structure) in the mutant. (It should be noted that this 3 ns is molecular dynamics *simulation* time at the high temperature 400 K, which is not the actual unfolding time at biological temperatures.) Interestingly, Helix C (residues 90–100) is also partially destroyed during this time period. As the time progresses, the native-like local contacts and secondary structures (mentioned above) in the beta-domain of the protein are largely destroyed. Meanwhile, Helix A (residues 5–14) and Helix B (residues 25–36) in the alpha-domain remain intact, and Helix D (residues 110–115), on the other hand, shows large structural disruptions and its helical form and reform. Near the end of the 15-ns simulation, the alpha-domain tertiary structures start to be largely destroyed as well, even though some of the alpha-helical secondary structures still remain. The protein is essentially in a molten-globule structure with a significantly larger radius of gyration than at the start of the simulation. Specifically, the radius of gyration increased from 14.3 Å at 0 ns to greater than 16.8 Å at 15 ns. These results indicate that the unfolding process starts at the beta-domain region, with the two beta-strands being destroyed first, and then continues with adjacent Helix C and Helix D, and then the alpha-domain as a whole. Other trajectories show similar results even though the exact time for each unfolding event is slightly different. Moreover, the CHARMM [19] force field overall corresponds to less-stable structures at higher temperatures than the OPLSSA [18] force field. This relative high stability of the OPLSAA force field compared with the CHARMM force field had also been previously reported on the peptide conformational distributions [25] and a beta-hairpin folding melting temperature [26, 27].

Chemical denaturing

The chemical denaturing simulations are performed as if the protein were in an 8M urea solution at a pH of 2 and a pH of 7 (see the Appendix for details). The following results and descriptions are based on a pH of 2 unless otherwise explicitly stated. The time dependence of both backbone RMSD and the radius of gyration from the native crystal structure provides measures of unfolding dynamics. We find that both quantities steadily increase up to 1,000 ns, with the mutant displaying higher values in the early stage (in the first ~100 ns), but once unfolded after 200 to 300 ns, these two quantities become insufficient to demonstrate differences between the wild type and mutant [28]. More informative measures of unfolding are, thus, needed. **Figure 4(a)** shows the fraction of native contacts for both the wild-type and the mutant protein, which is found to be a more informative measure for distinguishing the two protein types. Here,



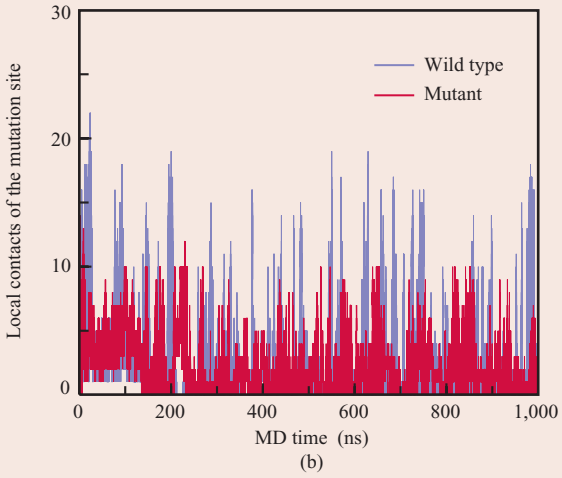


Figure 4

Comparison of the fraction of native contacts (a) and number of local contacts (b) for the wild type and mutant. It is clear that the wild type has a higher fraction of native contacts and a larger local contact number than the mutant. Trp62 plays a key role in bridging the neighboring positively charged basic residues, which in turn help form a nucleation core through long-range electrostatic interactions. (MD: molecular dynamics.)

two residues i and $i+n$ are said to be in contact if the distance between their Cα carbons is closer than 6.5 Å (for $n > 2$, the non-nearest and non-second-nearest residues are examined). As shown in Figure 4(a), the wild type maintains significantly more native contacts than the mutant as time evolves. The average fraction of native contacts from the final 100-ns data for the five independent trajectories is found to be 33 ± 4% for the wild type and 21 ± 3% for the mutant. We also compare the RMSF of each residue (represented by Cα) for the wild-type and the mutant lysozyme. The RMSF results

show that the mutant lysozyme undergoes much larger fluctuations than the wild type, notably in the beta-domain region and in the loop region where the mutation site Trp62 resides, consistent with the above thermal denaturing results [20]. Since the local contact network connecting this loop region with Strand 1 (residues 43–46) and Strand 2 (residues 51–54) in the wild type is disrupted in the mutant, we expect that motions of the loop should be more labile. Site 62, the mutation site, is found to be significantly more flexible in the mutant than in the wild type during the chemical denaturing simulations, which is consistent with experimental findings [2].

Next, we calculated the time-evolution of the secondary structure for both the wild-type and the mutant lysozyme in 8M urea solution at 310 K, a study that was similar to the above study on thermal denaturing trajectories. Surprisingly, the secondary structural components, particularly the alpha-helices, for the wild type are fairly stable up to 100 ns, despite that the RMSD of the protein grows to 10–12 Å and its fraction of native contacts decreases to less than 40%. On the other hand, the secondary structure of the mutant is more drastically disrupted. The disruption again starts from the beta-domain region, with the two beta-strands disappearing after 20–30 ns, and then spreads into the Helix C (residues 90–100) and Helix D (residues 110–115) regions of the alpha-domain. By 100 ns, part of Helix A (residues 5–14) is also destroyed. It should be noted that some of the helical contents are preserved for a very long time, even at the end of the 1,000-ns simulation, in both the wild type and the mutant. This persistence of alpha-helical secondary structure may indicate that the 1-μs simulation in 8M urea solution is still not sufficiently long for full denaturization. Although each trajectory displays slightly different behavior, the collection of trajectories displays a reasonably good consensus on the evolution of secondary structures. The higher radius of gyration, larger RMSD (particularly in the early stage of simulation), fewer native-like contacts, higher flexibility, and more disruptive secondary structures in the mutant imply that it is much less stable than the wild type. As mentioned above, Klein-Seetharaman et al. [2] found that the single mutation W62G caused the native-like contacts, some of them long range, in the wild type to disappear but did not identify the origin or the order of the disruptions, probably because of limitations on experimental resolution. Our chemical denaturing simulations show that the disruptions start from the beta-domain and then spread into the helices near the interface of the two domains (Helix C and Helix D), consistent with the above thermal denaturing simulation results.

Careful examination of the unfolding trajectories shows how the mutant loses tertiary structure and native-like contacts during the chemical denaturing process.

Figure 5 shows snapshots of the mutant during one of the 1-μs trajectories at 310 K. The first major unfolding event occurs at about 20–30 ns, when the beta-strands 1 and 2 start to disappear in the mutant. At this point, most of the native-like local contacts (and secondary structures) in the beta-domain are destroyed. This is consistent with the analysis of the evolution of the secondary structure. As time progresses, the tertiary contacts between Helix D and Helix C and their contacts with the rest of the protein are disrupted. After about 100 ns, many of the tertiary structures involving Helix A and Helix B also start to disappear, even though some of the helical secondary structure contents still remain. The protein is somehow "stretched" (i.e., better solvated) by the urea molecules with a much larger radius of gyration. By approximately 300–500 ns, the alpha-domain tertiary structures are mostly destroyed, with fewer native contacts left. Other trajectories show similar behavior, although the exact time for each event can be slightly different.

We now attempt to answer the central question as to why the single mutation W62G causes the above disruption in the tertiary structure. To address this question, we study the local contacts of Trp62 (or Gly62) residue in detail. Here, a *local contact*, native or not, is defined the same way as a native contact, except that the distance between alpha-carbons is chosen to be less than 10 Å, rather than 6.5 Å for a native contact. This larger distance of 10 Å allows us to have a broader view of residues proximate to the mutation site. As shown in **Figure 4(b)**, many more local contacts of Trp62 exist in the wild type than Gly62 in the mutant, with the average number of local contacts being 5.6 for the wild type and 1.9 for the mutant. This higher number of local contacts (and association with Arg112, as discussed below) is expected because of the hydrophobic clusters found in the wild type near residues 62 and 112 in the nuclear magnetic resonance (NMR) experiments [2]. A quick examination of the snapshots from both the wild-type and the mutant trajectories indicates that Trp62 in the wild type has many more basic residues, such as Arg and Lys, nearby than the Gly62 in the mutant. Thus, we did a thorough comparison for the average Cα pair distances from a basic residue to Trp62 versus distances to Gly62. These pair distances, as shown in **Figure 6**, reveal that the basic residues in the mutant are on average more distant from the mutation site than in the wild type, with significantly higher standard deviations as well. This result indicates that in the mutant, the basic residues are more distant from the possible nucleation site Trp62 (if such a site exists) than they are in the wild type, where these basic residues, particularly Arg73, Lys97, and Arg112, can form some kind of local cluster (or a nucleation site) along with the Trp62 amino-acid residue. Analysis of local structures also reveals that the Trp62

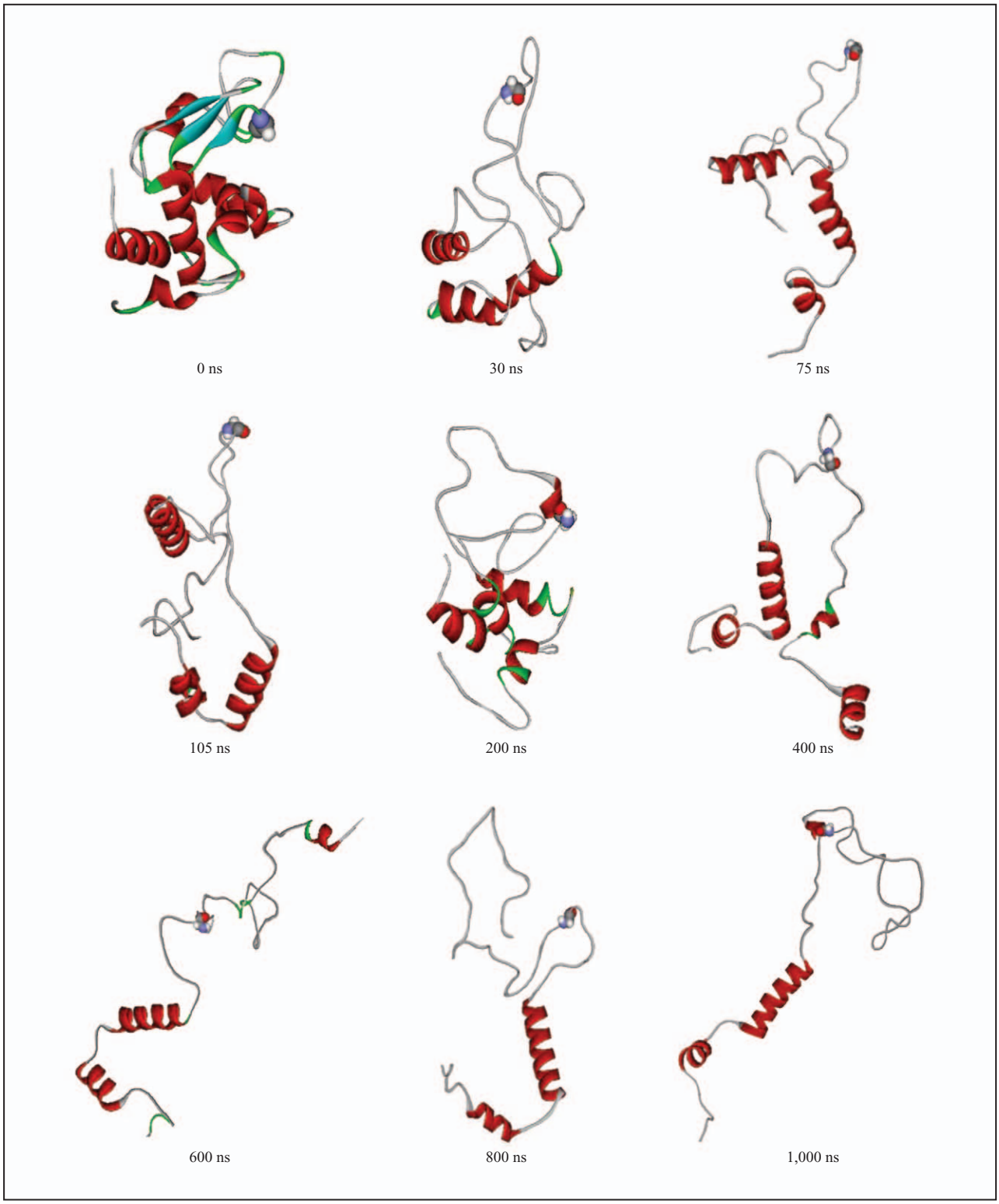


Figure 5

Snapshots of the mutant lysozyme during one representative 1-μs simulation of the chemical denaturing trajectories. These snapshots clearly indicate the gradual loss of the native contacts, with most of the loss in the beta-domain occurring first. Interestingly, even at the end of the 1-μs simulation, some helical content persists.

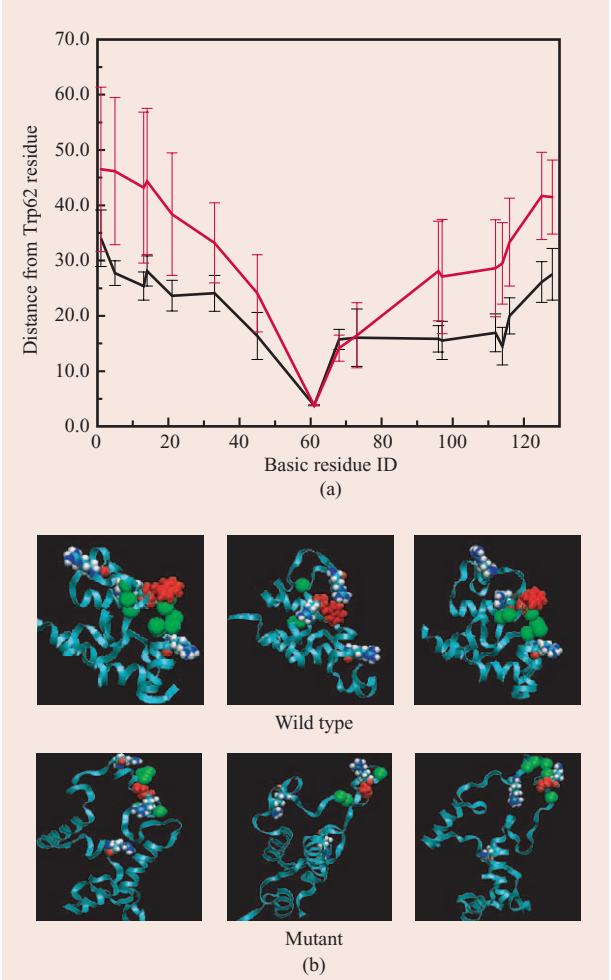


Figure 6

Unfolding mechanism of lysozyme (figure reproduced from reference [28], with permission). (a) Comparison of the average Cα–Cα distance and standard deviation of basic residues from Trp62 for the wild type (black line) and mutant (red line) in the first 100-ns simulation. (b) Representative structures of the wild type and mutant during the 100–1,000 ns molecular dynamics (MD) simulations, with the Arg-Trp-Arg "sandwich structure" seen in the middle snapshot of the wild type. The Trp62 is represented as red van der Waals (vdW) space-filling spheres, and three nearby basic residues (Arg73, Lys97, Arg112) as vdW space-filling spheres. The green balls represent the residues making local contacts with Trp62.

amino-acid residue acts as a bridge between two neighboring basic residues, such as Arg73 and Arg112—the π-electrons on its aromatic indole ring can attract the two positively charged residues to form a kind of sandwich structure, Arg-Trp-Arg, through the π–cation interactions and π-type hydrogen bonds. The energetically favorable π–cation interactions have been previously observed between tryptophan and arginine in protein crystal structures [29]. Interestingly, we also found that arginines are more likely than lysines to be involved in these types of interactions with tryptophan, consistent with previous findings [29]. It should be noted that modern force fields, such as in CHARMM (parameter 22) [19], do not have built-in polarizability, so the π–cation interactions and π-type H-bonds might be underestimated. Nevertheless, with this bridge effect, the two positively charged residues are held more closely together. These basic residues can then attract other local residues through long-range electrostatic interactions, as indicated by the much higher number of local contacts seen in the wild type. Therefore, a native-like cluster or nucleation site can be formed near these residues in the wild type, and Trp62 plays a key role in a cooperative long-range interaction. On the other hand, the Gly62 residue in the mutant does not have π-electron-equipped aromatic rings and thus does not have the capability to be such a nucleation site. This resolves the mystery of why Trp62 located on the surface can give rise to the long-range interactions and can have such a profound effect on the stability of the protein. Interestingly, this Arg-Trp-Arg bridge structure is not seen in the wild-type x-ray crystal structure. In fact, the Arg112 residue is more than 10 Å away from Trp62, which makes this bridging effect during the early stage of folding even more intriguing. (Specifically, the Cα–Cα distance is 19.1 Å, and the Cζ–indole ring closest distance is 12.5 Å.) Finally, we note that Magalhaes et al. [30] have previously also found a similar effect, in which two positively charged arginines are brought together on the surface of a protein by polarizing the intervening water molecules. In this case, it is the bridging water (instead of the negatively charged indole ring) that contributes to the stability of these unusual arginine–arginine short-range pairs.

Conclusion

In this paper, we performed thermal and chemical denaturing molecular dynamics simulations to study how a single point mutation (W62G) affects the stability and misfolding of the protein hen egg white lysozyme. Both the wild-type and the mutant lysozyme dynamics were simulated using the BG/L supercomputer.

Our results from the thermal denaturing simulations at 400 to 500 K show that the two widely used force fields, CHARMM and OPLSAA, display qualitatively similar results in terms of the misfolding process for the W62G mutant. The single mutation effect is reproducible and robust with the two force fields. The thermal and chemical denaturing simulations show that the mutant structure is indeed much less stable than the wild type. This is consistent with a recent urea denaturing

experiment [2]. The time-evolution of the secondary structure and RMSF analyses reveal that the single mutation, W62G, induces the loss of native contacts in the beta-domain region first, then the unfolding spreads into the alpha-domain region. The local contact analysis shows that the Trp62 amino-acid residue is the key to a cooperative long-range interaction within the wild type; it plays a role as a bridge or attractant between several neighboring basic residues, such as Arg73 and Arg112, to form Arg-Trp-Arg sandwich structures. Therefore, a native-like cluster or nucleation site can be formed near these residues in the wild type, while the mutant does not have this π-electron-equipped indole ring and thus does not have this nucleation mechanism. The results from our large-scale simulations not only complement recent experimental results but also offer useful insights into the mechanism behind the lysozyme protein misfolding and subsequent aggregation.

Appendix: System and methodology details

Thermal denaturing

As mentioned in the main text, the starting protein structure is from the x-ray crystal structure deposited in the PDB (*193L.pdb*). Both wild-type and mutant protein configurations are solvated in a water box of size 60 Å × 60 Å × 60 Å. Eight Cl^- counter ions are added to neutralize the solvated systems. The solvated protein systems have about 21,000 atoms. Both the OPLSAA force field [18] with an SPC (simple point charge) water model [31] and the CHARMM (parameter 22) force field [19] with a modified TIP3P water model [32, 33] are used for the simulation for force-field validation. (*TIP3P* stands for transferable intermolecular potential, three-position model.] For the long-range electrostatic interactions, we make use of the particle–particle particle–mesh Ewald (P3ME) method [34]. For the van der Waals interactions, a typical 10-Å cutoff is used.

A standard equilibration procedure is adopted for both the wild-type and the mutant protein systems. The procedure starts with a conjugate gradient minimization for each solvated system. Next, a two-stage equilibration is performed, each consisting of 100-ps molecular dynamics simulation. In the first stage, the protein is frozen in space, and only the solvent molecules are equilibrated. In the second stage, all atoms are equilibrated. The configurations from the above two-stage equilibration are then used as the starting points for another 1,000-ps NPT (i.e., isothermal and isobaric) simulation at 300 K and 1 atm. Three configurations are selected from the last 300-ps trajectory (each 100 ps apart) to be the final starting configurations. Thus, for both the wild-type and the mutant lysozyme, three trajectories starting from different initial configurations are performed at 300 K, 350 K, 400 K, 450 K, and 500 K. The starting structures of the mutant lysozyme were generated by a simple replacement of residue Trp62 to Gly2 from the above wild-type systems and then re-equilibrated with a 1,000-ps NPT simulation. The Blue Matter (molecular dynamics) application framework [35] is used for the thermal denaturing simulation.

Chemical denaturing

For the preparation of the 8M aqueous urea, we followed a similar approach used by Caflisch and Karplus [36]. A total of 30 urea molecules were first randomly immersed into a previously equilibrated 18.6-Å × 18.6-Å × 31.6-Å water box with 216 SPC water molecules. If any urea molecule overlaps with other urea molecules, the molecule will be replaced by another randomly distributed one. Next, all water molecules overlapping with the urea molecules were removed if the distance between the water oxygen atom and urea heavy atoms was less than 2.7 Å. This yielded a box of 30 urea and 128 water molecules, which was then minimized and equilibrated for a 100-ps NVT simulation at constant volume and a constant temperature of 310 K. The resulting small water–urea box was then expanded periodically in space to generate a much larger box of 74.4 Å × 74.4 Å × 74.4 Å with 1,920 urea and 8,192 water molecules. This larger urea–water mixture was then further equilibrated with a 1,000-ps NPT simulation at 310 K and 1 atm. The final box size was 73.1 Å × 73.1 Å × 73.1 Å, which corresponds to an approximately 8M urea concentration at a density of 1.12 g/cm^3. The lysozyme protein was then immersed in the equilibrated 8M urea box, and water and urea molecules overlapping with protein atoms were deleted. The final molecular system consisted of lysozyme centered in the box with 7,793 water and 1,809 urea molecules. Eight Cl^- counter ions for the pH = 7 experiment and eighteen Cl^- counter ions for the pH = 2 experiment (with residues Asp, Glu, and His protonated) were then added to neutralize the solvated system, giving a total system size of about 40,000 atoms. The pH = 2 results were described here unless explicitly stated. (The pH = 7 simulations show similar results and readers are referred to the supplementary material of reference [28] for more details.) The final lysozyme protein in an 8M urea system was then equilibrated following a similar procedure as the thermal denaturing mentioned above. For statistical purposes, five different configurations are used for both the wild-type and the mutant lysozyme. For each configuration (of a total of ten conformations), a molecular dynamics simulation is run for up to 1 μs at 310 K and 1 atm. The NAMD2 molecular dynamics program [37] is used for our simulations, along with the CHARMM [19] force field for protein lysozyme and solvent urea. A slightly

modified TIP3P [32, 33] water model is used for the solvent water.

Acknowledgments

We acknowledge the contributions of many people who have participated in the development of the Blue Matter code, including Blake G. Fitch, Aleksandr Rayshubskiy, Frank Suits, Yuri Zhestkov, T. J. Chris Ward, Mike Pitman, Alan Grossfield, Jed Pitera, and Bill Swope. We thank Sameer Kumar for substantial help with porting NAMD2 onto the IBM Blue Gene/L supercomputer. We also acknowledge the contributions of the BG/L hardware, system software, and science application teams whose efforts and assistance made it possible for us to use the BG/L supercomputer located at the IBM T. J. Watson Research Center.

References

1. B. L. Kagan, "Amyloidosis and Protein Folding," *Science* **307**, No. 5706, 42–43 (2005).
2. J. Klein-Seetharaman, M. Oikawa, S. B. Grimshaw, J. Wirmer, E. Duchardt, T. Ueda, T. Imoto, L. J. Smith, C. M. Dobson, and H. Schwalbe, "Long-Range Interactions Within a Nonnative Protein," *Science* **295**, No. 5560, 1719–1722 (2002).
3. M. Dumoulin, A. Last, A. Desmyter, K. Decanniere, D. Canet, A. Spencer, D. Archer, et al., "A Camelid Antibody Fragment Inhibits the Formation of Amyloid Fibrils by Human Lysozyme," *Nature* **424**, No. 6950, 783–788 (2003).
4. F. Chiti, M. Stefani, N. Taddei, G. Ramponi, and C. M. Dobson, "Rationalization of the Effects of Mutations on Peptide and Protein Aggregation Rates," *Nature* **424**, No. 6950, 805–808 (2003).
5. T. Ueda, H. Yamada, H. Aoki, and T. Imoto, "Effect of Chemical Modifications of Tryptophan Residues on the Folding of Reduced Hen Egg-White Lysozyme," *J. Biochem. (Tokyo)* **108**, No. 5, 886–892 (1990).
6. T. Mishima, T. Ohkuri, A. Monji, T. Imoto, and T. Ueda, "A Particular Hydrophobic Cluster in the Residual Structure of Reduced Lysozyme Drastically Affects the Amyloid Fibrils Formation," *Biochem. Biophys. Res. Comm.* **356**, No. 3, 769–772 (2007).
7. F. Allen, G. Almasi, W. Andreoni, D. Beece, B. J. Berne, A. Bright, J. Brunheroto, et al., "Blue Gene: A Vision for Protein Science Using a Petaflop Supercomputer," *IBM Syst. J.* **40**, No. 2, 310–327 (2001).
8. A. R. Fersht and V. Daggett, "Protein Folding and Unfolding at Atomic Resolution," *Cell* **108**, No. 4, 573–582 (2002).
9. C. M. Dobson, A. Sali, and M. Karplus, "Protein Folding: A Perspective from Theory and Experiment," *Angew Chem. Int. Edit. Engl.* **37**, 868–893 (1998).
10. C. L. Brooks, M. Gruebele, J. N. Onuchic, and P. G. Wolynes, "Chemical Physics of Protein Folding," *Proc. Natl. Acad. Sci. USA* **95**, No. 19, 11037–11038 (1998).
11. Y. Duan and P. A. Kollman, "Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution," *Science* **282**, No. 5389, 740–744 (1998).
12. R. Zhou, X. Huang, C. J. Margulius, and B. J. Berne, "Hydrophobic Collapse in Multidomain Protein Folding," *Science* **305**, No. 5690, 1605–1609 (2004).
13. P. Liu, X. Huang, R. Zhou, and B. J. Berne, "Observation of a Dewetting Transition in the Collapse of the Melittin Tetramer," *Nature* **437**, No. 7055, 159–162 (2005).
14. C. D. Snow, H. Nguyen, V. S Pande, and M. Gruebele, "Absolute Comparison of Simulated and Experimental Protein-Folding Dynamics," *Nature* **420**, No. 6911, 102–106 (2002).
15. V. Daggett, "Long Timescale Simulations," *Curr. Opin. Struct. Biol.* **10**, No. 2, 160–164 (2000).
16. K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, "Principles of Protein Folding—A Perspective from Simple Exact Models," *Protein Sci.* **4**, No. 4, 561–602 (1995).
17. F. Ding, S. V. Buldyrev, and N. V. Dokholyan, "Folding Trp-Cage to NMR Resolution Native Structure Using a Coarse-Grained Protein Model," *Biophys. J.* **88**, No. 1, 147–155, (2005).
18. W. L Jorgensen, D. Maxwell, and J. Tirado-Rives, "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids," *J. Am. Chem. Soc.* **118**, No. 45, 11225–11236 (1996).
19. A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, et al., "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins," *J. Phys. Chem. B* **102**, No. 18, 3586–3616 (1998).
20. M. Eleftheriou, R. S. Germain, A. K. Royyuru, and R. Zhou, "Thermal Denaturing of Mutant Lysozyme with Both the OPLSAA and the CHARMM Force Fields," *J. Am. Chem. Soc.* **128**, No. 41, 13388–13395 (2006).
21. G. Moraitakis and J. M. Goodfellow, "Simulations of Human Lysozyme: Probing the Conformations Triggering Amyloidosis," *Biophys. J.* **84**, No. 4, 2149–2158 (2003).
22. C. Venclovas, A. Zenla, K. Fidelis, and J. Moult, "Some Measures of Comparative Performance in the Three CASPs," *Proteins* **S3**, 231–237 (1999).
23. P. I. de Bakker, P. H. Hunenberger, and J. A. McCammon, "Molecular Dynamics Simulations of the Hyperthermophilic Protein Sac7d from *Sulfolobus acidocaldarius*: Contribution of Salt Bridges to Thermostability," *J. Mol. Biol.* **285**, 1811–1830 (1999).
24. D. Frishman and P. Argos, "Knowledge-Based Protein Secondary Structure Assignment," *Proteins* **23**, No. 4, 566–579 (1995).
25. H. Hu, M. Elstner, and J. Hermans, "Comparison of a QM/MM Force Field and Molecular Mechanics Force Fields in Simulations of Alanine and Glycine Dipeptides (Ace-Ala-Nme and Ace-Gly-Nme) in Water in Relation to the Problem of Modeling the Unfolded Peptide Backbone in Solution," *Proteins* **50**, No. 3, 451–463 (2003).
26. A. R. Dinner, T. Lazaridis, and M. Karplus, "Understanding Beta-Hairpin Formation," *Proc. Natl. Acad. Sci. USA* **96**, No. 16, 9068–9073 (1999).
27. R. Zhou, B. J. Berne, and R. S. Germain, "The Free Energy Landscape for Beta-Hairpin Folding in Explicit Water," *Proc. Natl. Acad. Sci. USA* **98**, No. 26, 14931–14936 (2001).
28. R. Zhou, M. Eleftheriou, A. Royyuru, and B. J. Berne, "Destruction of Long-Range Interactions by a Single Mutation in Lysozyme," *Proc. Natl. Acad. Sci. USA* **104**, No. 14, 5824–5829 (2007).
29. J. P. Gallivan and D. A. Dougherty, "Cation–Pi Interactions in Structural Biology," *Proc. Natl. Acad. Sci. USA* **96**, No. 17, 9459–9464 (1999).
30. A. Magalhaes, B. Maigret, J. Hoflack, J. N. Gomes, and H. A. Scheraga, "Contribution of Unusual Arginine–Arginine Short-Range Interactions to Stabilization and Recognition in Proteins," *J. Protein Chem.* **13**, No. 2, 195–215 (1994).
31. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Herman, "Interaction Models for Water in Relation to

Protein Hydration," *Intermolecular Forces*, B. Pullman, Ed., Reidel, Dordrecht, 1981, pp. 331–342.
32. E. Neria, S. Fischer, and M. Karplus, "Simulation of Activation Free Energies in Molecular Systems," *J. Chem. Phys.* **105**, No. 5, 1902–1921 (1996).
33. W. L. Jorgensen, J. Chandrasekhar, J. Madura, R. W. Impey, and M. L. Klein, "Comparison of Simple Potential Functions for Simulating Liquid Water," *J. Chem. Phys.* **79**, No. 2, 926–935 (1983).
34. U. Essman, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A Smooth Particle Mesh Ewald Method," *J. Chem. Phys.* **103**, No. 19, 8577–8593 (1995).
35. B. G. Fitch, R. S. Germain, M. Mendell, J. Pitera, M. Pitman, A. Rayshubskiy, Y. Sham, et al., "Blue Matter, an Application Framework for Molecular Simulation on Blue Gene," *J. Parallel Distrib. Comput.* **63**, No. 7–8, 759–773 (2003).
36. A. Caflisch and M. Karplus, "Structural Details of Urea Binding to Barnase: A Molecular Dynamics Analysis," *Structure* **7**, No. 5, 477–488 (1999).
37. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, "Scalable Molecular Dynamics with NAMD," *J. Comput. Chem.* **26**, No. 16, 1781–1802 (2005).

Received March 15, 2007; accepted for publication April 24, 2007; Internet publication December 14, 2007

Ruhong Zhou *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (ruhongz@us.ibm.com)*. Dr. Zhou is a Research Staff Scientist at the Computational Biology Center of the IBM T. J. Watson Research Center, and an Adjunct Professor in the Department of Chemistry at Columbia University. He received his Ph.D. degree from Columbia University in 1997. He has authored and coauthored more than 70 journal publications and 7 patents, delivered numerous invited talks at major conferences and universities, and chaired and co-chaired several conferences in this area. He won the Hammett Award in 1997 from Columbia University, the DEC Award in 1995 from the American Chemical Society on Computational Chemistry, and the IBM Outstanding Technical Achievement Award in 2005. His current research interests include development of novel methods and algorithms for computational biology and bioinformatics, as well as large-scale simulations for protein folding, ligand–receptor binding, and protein structure prediction.

Maria Eleftheriou *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (mariae@us.ibm.com)*. Dr. Eleftheriou is a researcher at the IBM T. J. Watson Research Center. For the past few years, she has been working mainly on the Blue Gene* Project. In particular, she has contributed to the design and the implementation of parallel algorithms, parallel applications, and parallel programming models, and she has studied the performance of parallel scientific applications for the Blue Gene/L architecture. Another area of interest is large-scale simulations, addressing questions of biological interest, particularly in the area of protein folding. She received her master's degree in engineering in 1995 and her Ph.D. degree in chemistry in 1999, from Brown University.

Chung-Chau Hon *University of Hong Kong, Pokfulam Road, Hong Kong, China (h9826299@hkusua.hku.hk)*. Dr. Hon is a Research Staff Member in the Department of Zoology at the University of Hong Kong. He received his M.Phil. degree in molecular virology and a Ph.D. degree in virological bioinformatics from the University of Hong Kong. During his graduate studies, in 2006 he also worked at the IBM T. J. Watson Research Center as a summer intern, where he focused on the analysis of molecular simulation trajectories for lysozyme. Dr. Hon's research interests have centered on molecular modeling of viral capsid proteins, viral epidemiology, and evolutionary virology.

Robert S. Germain *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (rgermain@us.ibm.com)*. Dr. Germain manages the Biomolecular Dynamics and Scalable Modeling Group within the Computational Biology Center at the IBM T. J. Watson Research Center. He received his A.B. degree in physics from Princeton University in 1982 and his M.S. and Ph.D. degrees in physics from Cornell University. After receiving his doctorate in 1989, Dr. Germain joined the T. J. Watson Research Center as a Research Staff Member in the Physical Sciences Department and later the VLSI/Scalable Parallel Systems Packaging Department. From 1995 to 1998, he was project leader for the development of a large-scale fingerprint identification system using an indexing scheme (FLASH) developed at the IBM Research Division. Since 2000, Dr. Germain has been responsible for the science and associated application portions of the Blue Gene Project. His current research interests include the parallel implementation of algorithms for high-performance scientific computing, the development of new programming models for parallel computing, and applications of high-performance computing to challenging scientific problems in computational biology. Dr. Germain is a member of the ACM, IEEE, and the American Physical Society.

Ajay K. Royyuru. *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (ajayr@us.ibm.com)*. Dr. Royyuru is Senior Manager of the Computational Biology Center at the IBM Research Division, where he leads the teams engaged in research in bioinformatics, structural biology, protein science, Blue Gene applications, functional genomics, systems biology, computational neuroscience, and biomedical imaging. He obtained his Ph.D. degree in molecular biology from the Tata Institute of Fundamental Research, Mumbai, in 1993 and then conducted postdoctoral work in structural biology at Memorial Sloan-Kettering Cancer Center, New York. Prior to joining IBM in 1998, he spent 2 years developing structural biology software at Accelrys. Currently, his work focuses on collaborative research at the interface of information technology and biology. Working with biologists and institutions around the world, he is engaged in research that will advance personalized, information-based medicine. Dr. Royyuru leads the IBM Research teams working with National Geographic Society on the Genographic Project and with The Scripps Research Institute on modeling emerging viral diseases such as influenza.

Bruce J. Berne *Columbia University, Department of Chemistry, 3000 Broadway, New York, NY 10027 (bb8@columbia.edu)*. Dr. Berne is Higgins Professor of Chemistry and Professor of Chemical Engineering at Columbia University. He has been at Columbia since 1966 and has published more than 300 publications and authored three books. He is a member of the National Academy of Sciences (USA) and a Fellow of the American Academy of Arts and Sciences. He is also a fellow of the American Physical Society and the American Association for the Advancement of Science. Dr. Berne has won numerous awards including the Alexander von Humoldt-Stiftung Senior Scientist Award (1992), the American Chemical Society Award in Theoretical Chemistry (1995), the Joseph O. Hirschfelder Prize in Theoretical Chemistry (2001), the Joel Henry Hildebrand Award in the Theoretical and Experimental Chemistry of Liquids of the American Chemical Society (2002), and an IBM Research Achievement Award (2005). He has also delivered numerous honorary lectures.