Replica Exchange Molecular Dynamics of a Small Heat Shock Protein

Author: Derrick T. Sund
Project Advisor: Dr. Florence Tama

A Thesis Submitted in Partial Fulfillment of the Bachelors of Science degree in Biochemistry

Department of Chemistry & Biochemistry
University of Arizona
May 2011

Approved by:

Dr. Florence Tama
Faculty Thesis Advisor

5/4/2011
Date:

Biochemistry Undergraduate Advisor

Date:

The University of Arizona Electronic Theses and Dissertations
Reproduction and Distribution Rights Form

Name (Last, First, Middle)	Sund Derrick T
Degree title (eg BA, BS, BSE, BSB, BFA):	BS
Honors area (eg Molecular and Cellular Biology, English, Studio Art):	Biochemistry and Molecular Biophysics
Date thesis submitted to Honors College:	May 4 2011
Title of Honors thesis:	Replica Exchange Molecular Dynamics of a Small Heat Shock Protein
:The University of Arizona Library Release	I hereby grant to the University of Arizona Library the nonexclusive worldwide right to reproduce and distribute my dissertation or thesis and abstract (herein, the "licensed materials"), in whole or in part, in any and all media of distribution and in any format in existence now or developed in the future. I represent and warrant to the University of Arizona that the licensed materials are my original work, that I am the sole owner of all rights in and to the licensed materials, and that none of the licensed materials infringe or violate the rights of others. I further represent that I have obtained all necessary rights to permit the University of Arizona Library to reproduce and distribute any nonpublic third party software necessary to access, display, run or print my dissertation or thesis. I acknowledge that University of Arizona Library may elect not to distribute my dissertation or thesis in digital format if, in its reasonable judgment, it believes all such rights have not been secured. Signed: Derrick T Sund Date: May 4 2011

Last updated: Nov 15, 2009

Abstract

Small heat shock proteins (sHSPs) are a family of molecular chaperones whose purpose is to prevent protein aggregation in response to elevated temperatures. While the family is generally poorly-conserved, they share a common structure consisting of an α-crystallin domain, a short unstructured C-terminal arm, and a longer unstructured N-terminal arm. Here, the method of replica exchange molecular dynamics was used to investigate the hypothesis that the N-terminal arm is often important in the function of sHSPs. The results suggest a model in which these arms may alternately assume 'open' or 'closed' conformations, with the open conformation exhibiting greater protective capability. Moreover, open conformations occur with greater frequency at higher temperatures, suggesting that the protective ability of sHSPs may increase at elevated temperatures.

Introduction

One of the most important biological pathways in organisms ranging from prokaryotes to higher mammals is the heat shock response. Increases in temperature, even in heat-adapted organisms, have deleterious effects on a wide variety of cellular machinery, including individual protein molecules, membranes, and cytoskeletal structures [11, 17]. The disorderly aggregation of unfolded proteins, in particular, may explain many of the physiological problems induced by heat, so it is unsurprising that one of the seven broad classes of proteins involved in the heat shock response consists of molecular chaperones [11].

Small heat shock proteins (sHSPs) are one such class of molecular chaperones. They are a poorly conserved family of proteins found in all kingdoms of life, with oligomeric mass ranging from 12 to 43 kDa [5]. Their common structure consists of the conserved α-crystallin domain, a structured sequence of approximately 90 amino acids, as well as a short unstructured C-terminal extension and a longer, unconserved, unstructured N-terminal extension of variable length [1]. Small heat shock proteins aggregate to form oligomers with varying numbers of subunits [1]; the number of subunits per oligomer may vary even in a single sHSP [5].

It is difficult to overstate the importance of sHSPs. The negative effects that can potentially result from the uncontrolled aggregation of denatured proteins is well-known and has been linked to such diseases as sickle-cell anemia, Parkinson's disease, and cystic fibrosis [3]. Related α-crystallin proteins are also found in the human eye, where they are believed to play a role in preventing the formation of protein aggregates which might block or scatter light [5].

While the structures of a number of sHSPs have been determined [9, 8], their mechanism of action remains poorly understood [1]. Some recent work suggests that the unstructured N-terminal arms are important in the activity of sHSPs [1, 7]. In order to investigate this hypothesis to a greater extent, we have performed molecular dynamics on one such sHSP. The crystal structure of HSP 16.9 from wheat has been solved [15], and its activity compared with that of a number of homologous sHSPs [1]. Replica-exchange molecular dynamics was performed on one homologous sHSP, found in *Ara-*

bidopsis thaliana. The results are compared with similar data obtained from replica-exchange molecular dynamics on other homologous sHSPs to provide potential insight into their mechanism of action.

Materials and Methods

Simulation Preparation

The structure of the *Arabidopsis thaliana* sHSP (AtHSP) was determined previoiusly through homology modeling work (unpublished). The dimeric structure was minimized for 2000 cycles: 1000 cycles of steepest descent followed by 1000 cycles of conjugate gradient minimization. All simulations were performed in implicit solvent using the generalized Born model, using the parameters of Model II given by Onufriev, Bashford, and Case (2004) [10]. No periodic boundary conditions were used. Minimization was performed using the Amber 10 software suite [2].

Molecular Dynamics

The simulation was performed using the method of replica exchange [14], as described by Sanbonmatsu and Garcia [12] with Amber 10 and the ff03 force field. A total of 16 replicas were used, with temperatures ranging from 281.85 to 339.88 degrees Celsius. Temperature was controlled through Langevin dynamics using a collision frequency of 1 per ps. The timestep used was 2 fs. Accordingly, all bonds involving hydrogen were fixed via the SHAKE algorithm. Distance restraints were set up between Gly70 on one monomer and Asn134 of the other, and vice versa, to ensure the dimeric structure remained intact.

Production

The simulation was performed on the University of Arizona Linux-based Marin high-powered computing system using 64 CPUs: 4 CPUs per replica. After the conclusion of the simulation, trajectories consisting of all of the

frames at each target temperature were generated using ptraj, a part of the AMBER software suite [2].

Analysis

Analysis was performed on a desktop computer running OpenSUSE Linux, using parts of the Amber software suite, VMD and included packages [6, 16, 4], and Cytoscape [13].

Results

Diagnostics

A total of 52.1 ns of simulation was performed. Figure 2 shows the target temperatures of one of the replicas as a function of time, illustrating a wide sampling of temperature. The other 15 replicas exhibited similar sampling behavior. Figure 3 shows histograms of the potential energy exhibited by the replicas at each of the 16 target temperatures. The exchange rate (ratio of successful exchanges to exchange attempts) for each replica varied from 29.2% to 43.3%.

The root-mean-square deviation (RMSD), plotted as a function of time, to the initial, minimized structure is shown for the 315 K trajectory in figure 4. The deviation appears to have stabilized approximately 6 ns after the start of the simulation. As such, a total simulation time of 52.1 ns seems adequate to judge sampling of structure.

N-Terminal Arm Configurations

For each of the trajectories at the target temperatures of 300 K, 315.36 K, and 319.32 K, after 10 ns of simulation time, the distance betweeen the centers of mass of the two N-terminal arms was determined using VMD. A histogram showing the relative frequency of these distances is shown in figure 5. In each case, there are two 'peaks' indicating local frequency maxima: one at approximately 19 Angstroms, and a second at approximately 32 Angstroms.

The respective average distances were 20.54 Angstroms, 21.96 Angstroms, and 22.50 Angstroms.

Trajectory subsets of 334 evenly-spaced frames of data each were taken from each of these, with the solvent-accessible surface area (SASA) of sidechains of the hydrophobic residues on the N-terminal arms determined for each. The average hydrophobic SASA for each of the three temperatures were found to be 7271 square Angstroms, 7385 square Angstroms, and 7433 square Angstroms. A histogram of the data obtained is given in figure 6.

Figure 7 shows a histogram of the radius of gyration of the two N-terminal arms for the same trajectories as were used as the basis of the distance histogram given in figure 6. As with the distance histogram, two peaks of local frequency maxima are exhibited for each temperature: one at approximately 15 Angstroms, and one at approximately 21 Angstroms.

A scatter plot of distance between the centers of mass of the N-terminal arms and the hydrophobic SASA of the N-terminal arm sidechains is given in figure 8. Also shown is a best-fit line derived from partial least squares linear regression. The equation of the best fit line was determined to be

$$\text{SASA} = 6882.6 \text{ (square Angstroms)} + 23.791 \text{ (Angstroms)} \times \text{distance}$$

with a correlation coefficient of 0.373.

Clustering analysis

For the trajectories at the three target temperatures listed above, 334 evenly-spaced frames were obtained from the last 42.1 ns of simulation time. The three trajectories were combined into a single trajectory, with the RMSD between each pair of frames determined. A network was created by considering two frames to be 'connected' if their pairwise RMSD was below a cutoff of 7.2 Angstroms. The network was visualized using Cytoscape [13], and through the use of its force-directed layout algorithm, it was determined visually that there were 14 clusters of frames.

To facilitate analysis of the clusters, a different clustering method was subsequently used: the clustering facility contained within ptraj [2]. Ptraj

was configured to split the trajectory into 14 trajectories, clustered according to RMSD, and generate representative structures for each. To aid in visualization, these 14 representative structures were appended to the initial trajectory and used in the first clustering scheme. In this way, a visual representation of the clusters was generated alongside trajectory files that could be individually analyzed. The results are summarized in figure 9.

Discussion

Based on the data, it seems plausible that the conformations of the unstructured N-terminal arms of the *A. thaliana* sHSP can be organized into two broad categories: a 'closed' conformation in which the N-terminal arms are associated with one another, and an 'open' one in which they are further apart from one another. Significantly, the open conformation is assumed more frequently at higher temperatures. That this corresponds to an increased average hydrophobic sidechain SASA in the N-terminal arms is unsurprising, given the significant positive correlation found to exist between N-terminal arm distance and N-terminal hydrophobic sidechain SASA.

For some of the clusters shown in Figure 9, most notably in cluster 8, the number of frames present in the Cytoscape visualization of the cluster is significantly different than the number of frames shown in the output of ptraj; this is most likely due to differences in the clustering algorithms. For the majority of the outlying clusters, the ptraj output and Cytoscape visualization match more closely. Perhaps the most noteworthy data to come from the clustering analysis would be those clusters which exhibit a clear tendency toward or against certain target temperature values. In particular, clusters 3, 6, and 12 stand out as containing comparatively few instances of frames targeting 300 degrees K. All three of them exhibit relatively high average distance between the N-terminal arms, as well as higher average N-terminal hydrophobic SASA than the average for the 300 K trajectory.

That the sHSP would present a greater average hydrophobic SASA from the N-terminal arms at elevated temperatures is interesting, given that its function is naturally most important under such conditions. This is of particular note given the hypothesis that the mechanism of sHSPs may, in many

cases, involve hydrophobic interactions with the N-terminal arms [1]. It might therefore be the case that the sHSP in question exhibits greater protective ability at higher temperatures.

Some significant comparisons may be made to some other data recently found through replica exchange molecular dynamics on sHSPs. Dr. Sunita Patel, also working in the lab of Dr. Florence Tama, performed identical molecular dynamics simulations on a pair of homologous sHSPs found experimentally to exhibit greater protective activity (unpublished data). The results indicate clearly that the two more efficient sHSPs assume 'open' conformations, with high distance between center of mass of the N-terminal arms, with much greater frequency. It therefore seems plausible to propose a model in which it is these 'open' conformations that exhibit the greatest protective activity.

With these results, the logical continuation of this work would be to perform similar simulations on a wider variety of homologous sHSPs. A greater number of sHSPs with a variety of activity levels might be generated through site-directed mutagenesis of various residues on the N-terminal arm. By assessing their activity levels and determining whether the observed pattern of greater openness frequency correlating with greater protective capability holds, the proposed model could be tested.

Figure 1: The structure of the *Arabidopsis thaliana* sHSP studied here as determined by homology modeling with a homologous protein from wheat. The N-terminal arms, considered here to be residues 1 to 52, are shown in red and blue.

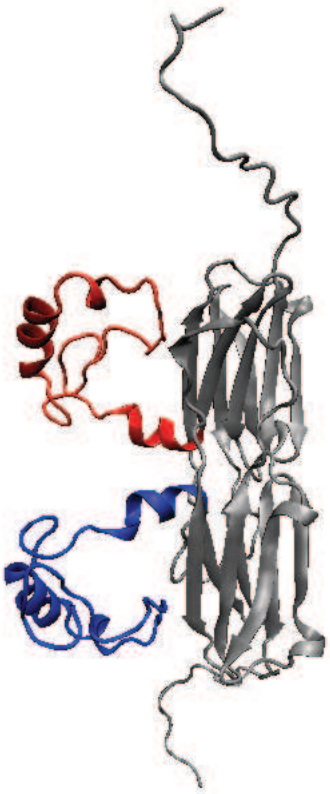


Figure 2: The target temperature of one of the replicas as a function of time.

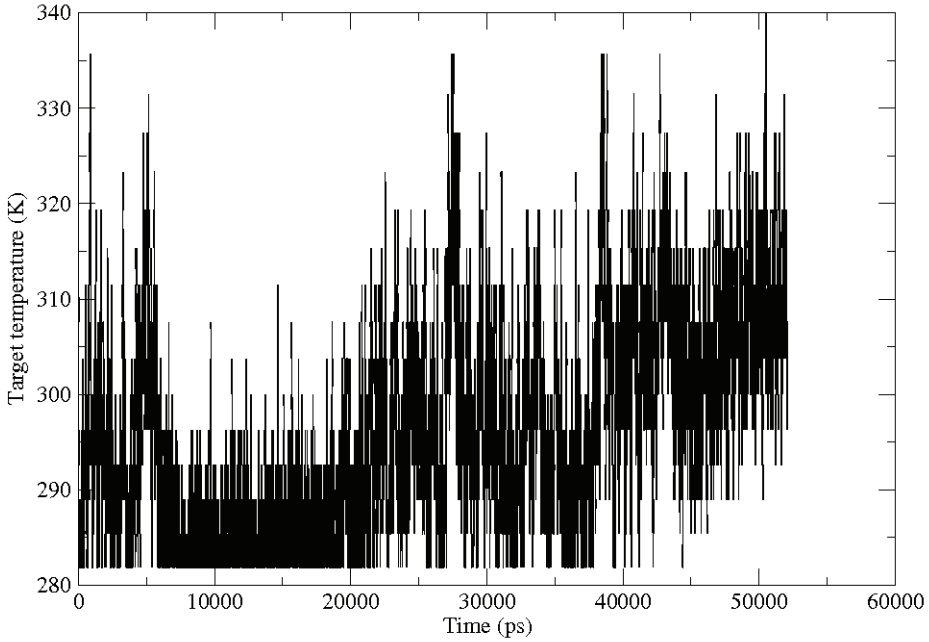


Figure 3: Energy histograms of the trajectories generated during the molecular dynamics simulation, sorted by target temperature.

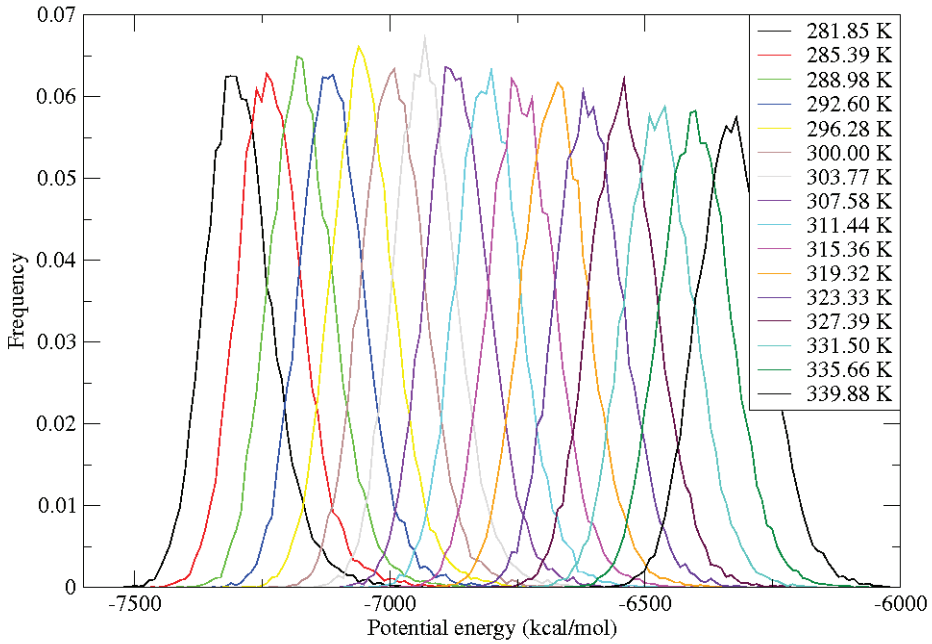


Figure 4: Root-mean-square deviation (RMSD) as a function of simulation time for the trajectory with target temperature 315.36 degrees Kelvin.

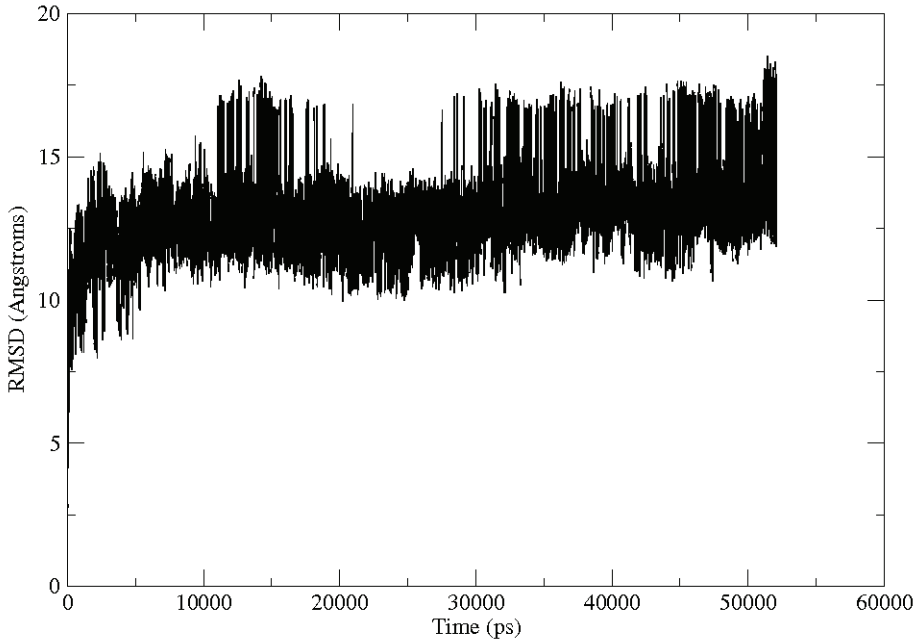


Figure 5: Histograms of the distances between the centers of mass of the N-terminal arms of the sHSP. The trajectories in question were analyzed starting at 10 ns of simulation time.

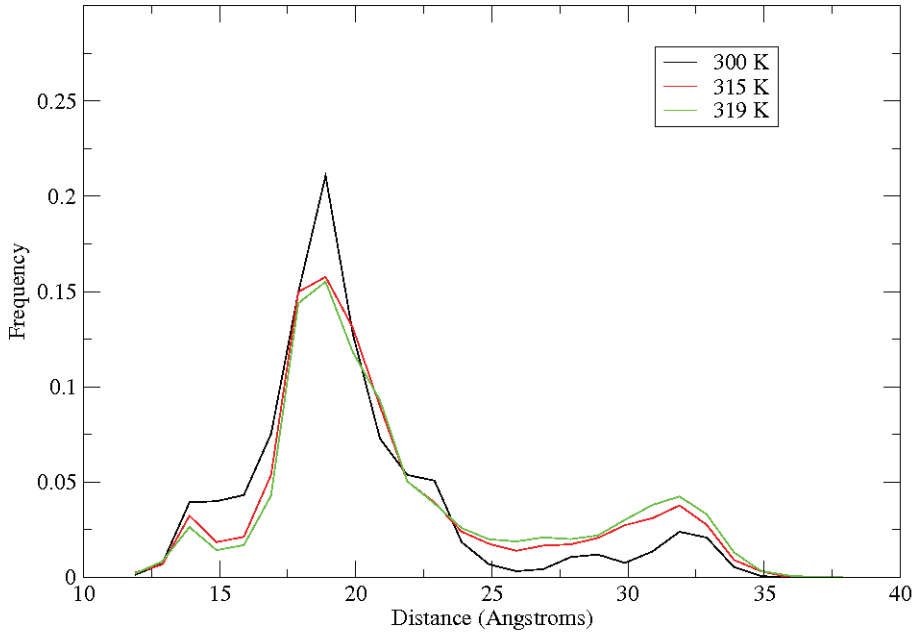


Figure 6: Histograms of the solvent-accessible surface area (SASA) of the hydrophobic residues on the N-terminal arms. The trajectories in question were analyzed starting at 10 ns of simulation time.

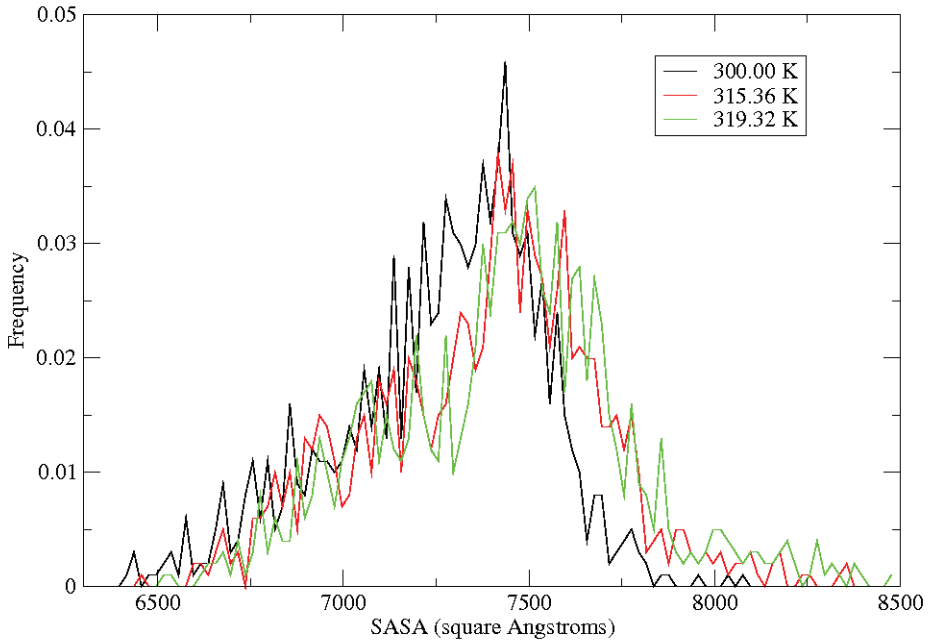


Figure 7: Histograms of the radius of gyration of the N-terminal arms. The trajectories analyzed were the same as for Figure 5.

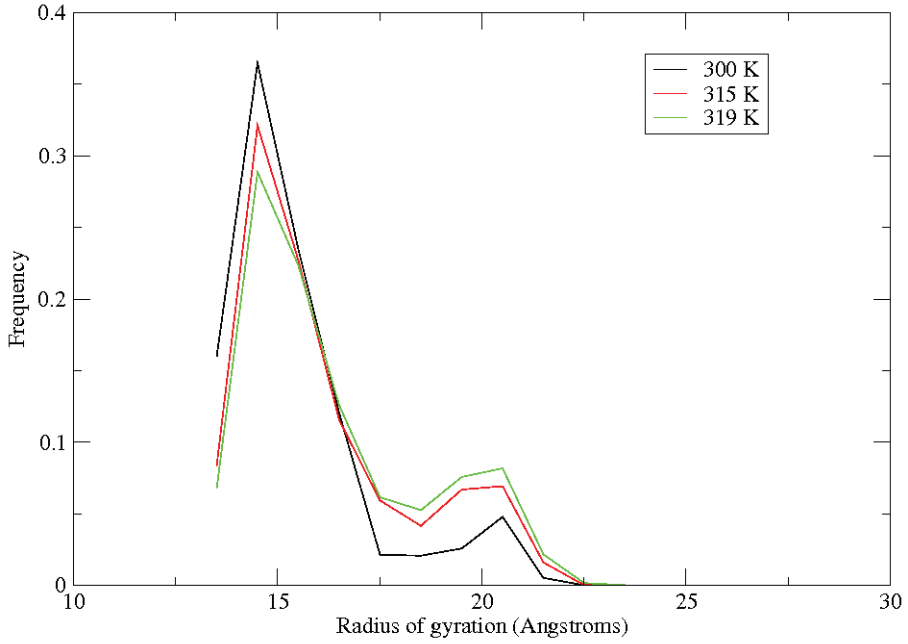


Figure 8: A scatter plot of the distance between the centers of mass of the N-terminal arms of the sHSP and the hydrophobic surface area of the sidechains of the N-terminal arms. The frames of data used were the same as those used in Figure 6. The best-fit line as determined by partial least squares linear regression is also shown.

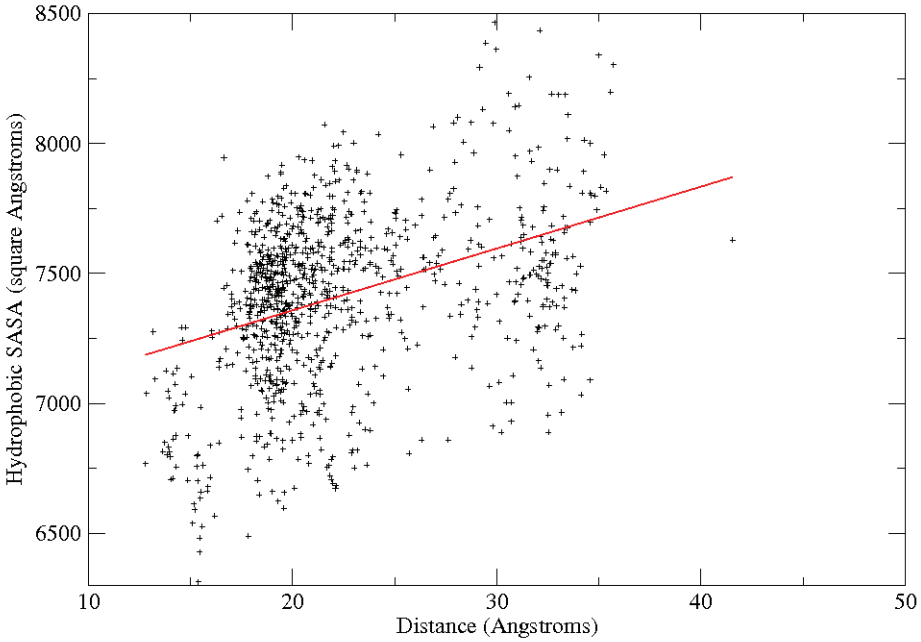
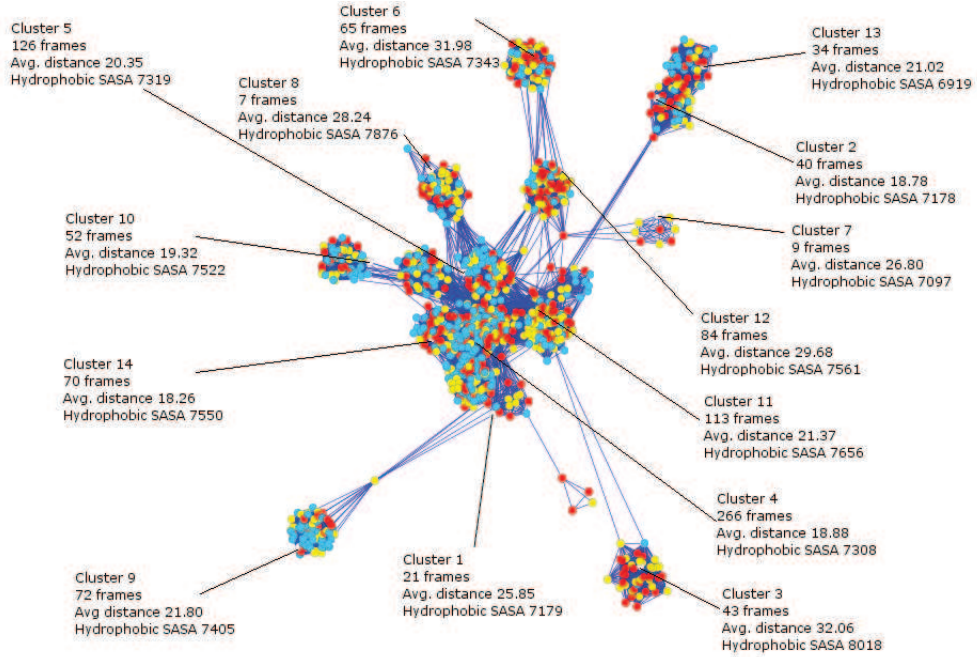


Figure 9: A diagram of the clustering analysis. Blue nodes represent frames of data taken from the 300 K trajectory, yellow nodes represent frames taken from the 315.36 K trajectory, and red nodes those from the 319.32 K trajectory. The white nodes correspond to the representative structures given by ptraj. For each of these representative structures, the number of frames given in each of the corresponding clusters by ptraj is shown, along with the average distance between the centers of mass of the N-terminal arms and the average hydrophobic SASA of the N-terminal arm sidechains.



Bibliography

[1] Eman Basha, Kenneth L. Friedrich, and Elizabeth Vierling. The n-terminal arm of small heat shock proteins is important for both chaperone activity and substrate specificity. *The Journal of Biological Chemistry*, 281(52):39943–39952, December 2006.

[2] D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, , and P. A. Kollman. Amber 10. University of California, San Francisco, 2008.

[3] John I. Clark and Paul J. Muchowski. Small heat-shock proteins and their potential role in human disease. *Current Opinion in Structural Biology*, 10(1):52–59, February 2000.

[4] D. Frishman and P. Argos. Knowledge-based secondary structure assignment. *Proteins: structure, function and genetics*, 23:566–579, 1995.

[5] Martin Haslbeck, Titus Franzmann, Daniel Weinfurtner, and Johannes Buchner. Some like it hot: the structure and function of small heat-shock proteins. *Nature Structural Molecular Biology*, 12(10):842–846, October 2005.

[6] W. Humphrey, A. Dalke, and K. Schulten. Vmd - visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996.

[7] Nomalie Jaya, Victor Garcia, and Elizabeth Vierling. Substrate binding site flexibility of the small heat shock protein molecular chaperones. *PNAS*, 106(37):15604–15609, September 2009.

[8] Christopher K. Kennaway, Justin L. P. Benesch, Ulrich Gohlke, Luchun Wang, Carol V. Robinson, Elena V. Orlova, Helen R. Saibi, and Nicholas H. Keep. Dodecameric structure of the small heat shock protein acr1 from mycobacterium tuberculosis. *The Journal of Biological Chemistry*, 280(39):33419–33425, September 2005.

[9] K. K. Kim, R. Kim, and S. H. Kim. Crystal structure of a small heat-shock protein. *Nature*, 394:595–599, 1998.

[10] A. Onufriev, D. Bashford, and D. A. Case. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins*, 55:383–394, 2004.

[11] Klaus Richter, Martin Haslbeck, and Johannes Buchner. The heat shock respons: Life on the verge of death. *Molecular Cell*, 40:253–266, October 2010.

[12] K. Y. Sanbonmatsu and A. E. García. Structure of met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins: Structure, Function, and Bioinformatics*, 46:225–234, 2002.

[13] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13(11):2498–2504, November 2003.

[14] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314:141–151, November 1999.

[15] R. L. van Montfort, E. Basha, K. L. Friedrich, C. Slingsby, and E. Vierling. Crystal structure and assembly of a eukaryotic small heat shock protein. *Nature Structural Biology*, 8:1025–1030, 2001.

[16] A. Varshney, F. P. Brooks, and W. V. Wright. Linearly scalable computation of smooth molecular surfaces. *IEEE Computer Graphics and Applications*, 14:19–25, 1994.

[17] William J. Welch and Joseph P. Suhan. Morphological study of the mammalian stress response: Characterization of changes in cytoplasmic organelles, cytoskeleton, and nucleoli, and appearance of intranuclear

actin filaments in rat fibroblasts after heat-shock treatment. *The Journal of Cell Biology*, 101:1198–1211, October 1985.