

Protein Dynamics in a Family of Laboratory Evolved Thermophilic Enzymes

Patrick L. Wintrode^{1†}, Deqiang Zhang^{1,2†}, Nagarajan Vaidehi^{1,2}
Frances H. Arnold¹ and William A. Goddard III^{1,2*}

¹Division of Chemistry and
Chemical Engineering
California Institute of
Technology, Pasadena, CA
91125-0001, USA

²Materials and Process
Simulation Center, California
Institute of Technology
Pasadena, CA 91125-1340
USA

Molecular dynamics simulations were employed to study how protein solution structure and dynamics are affected by adaptation to high temperature. Simulations were carried out on a *para*-nitrobenzyl esterase (484 residues) and two thermostable variants that were generated by laboratory evolution. Although these variants display much higher melting temperatures than wild-type (up to 18 °C higher) they are both >97% identical in sequence to the wild-type. In simulations at 300 K the thermostable variants remain closer to their crystal structures than wild-type. However, they also display increased fluctuations about their time-averaged structures. Additionally, both variants show a small but significant increase in radius of gyration relative to wild-type. The vibrational density of states was calculated for each of the esterases. While the density of states profiles are similar overall, both thermostable mutants show increased populations of the very lowest frequency modes (<10 cm⁻¹), with the more stable mutant showing the larger increase. This indicates that the thermally stable variants experience increased concerted motions relative to wild-type. Taken together, these data suggest that adaptation for high temperature stability has resulted in a restriction of large deviations from the native state and a corresponding increase in smaller scale fluctuations about the native state. These fluctuations contribute to entropy and hence to the stability of the native state. The largest changes in localized dynamics occur in surface loops, while other regions, particularly the active site residues, remain essentially unchanged. Several mutations, most notably L313F and H322Y in variant 8G8, are in the region showing the largest increase in fluctuations, suggesting that these mutations confer more flexibility to the loops. As a validation of our simulations, the fluctuations of Trp102 were examined in detail, and compared with Trp102 phosphorescence lifetimes that were previously measured. Consistent with expectations from the theory of phosphorescence, an inverse correlation between out-of-plane fluctuations on the picosecond time scale and phosphorescence lifetime was observed.

© 2003 Published by Elsevier Science Ltd

Keywords: thermophile; protein stability; directed evolution; protein dynamics; molecular dynamics simulation

*Corresponding author

Introduction

The physical basis for the remarkable stability of enzymes isolated from thermophilic organisms has

been the subject of intensive research.^{1,2} There are now numerous studies comparing the sequences and structures of thermophilic enzymes with those of homologous enzymes from mesophilic organisms.^{1,2} These studies have found many types of stabilizing interactions in thermophilic enzymes, and there does not appear to be a single preferred mechanism for stabilization. In general, it appears that thermophilic enzymes have adapted to high temperature through the accumulation of numerous mild stabilizing interactions, including

† These two authors contributed equally to this work.
Abbreviations used: MD, molecular dynamics; pNBE, *para*-nitrobenzyl esterase; CRMS, root-mean-square difference in coordinates; DOS, density of states.

E-mail address of the corresponding author:
wag@wag.caltech.edu

salt bridges, hydrogen bonds, and van der Waals contacts.

Researchers have also focused on the dynamic properties of thermophilic enzymes. The conformational flexibility of homologous thermophilic and mesophilic enzymes has been probed using fluorescence quenching,³ hydrogen/deuterium exchange,^{4–6} molecular dynamics (MD) simulations,⁷ and neutron scattering.^{8,9} Many of these studies have found that the conformational flexibility of thermophilic enzymes at room temperature is considerably reduced compared to mesophilic enzymes. At the same time, the flexibility of thermophilic enzymes near their physiological (high) temperatures is often comparable to that of mesophilic enzymes at their physiological (moderate) temperatures.^{3,4} These findings concluded that reduced flexibility is a necessary consequence of thermostabilization, i.e. more stable proteins are less prone to have their structures perturbed by thermal fluctuations and therefore appear less flexible. Some researchers have further concluded that these differences in conformational flexibility have functional consequences and can explain observed differences in the temperature–activity profiles of thermophilic and mesophilic enzymes (for example the fact that thermophilic enzymes generally display poor activity at moderate temperatures).¹ Briefly, the argument states that conformational fluctuations in enzymes play an important role in their function as catalysts, but these fluctuations can also lead to the loss of structure and function if they become too large. Since the magnitude of the fluctuations experienced by an enzyme will depend on the available thermal energy, $k_B T$, evolution has modified the strength and number of stabilizing interactions in enzymes to achieve the optimal balance of stability and flexibility at a given temperature. As a result, large changes in temperature will disrupt this balance, causing cold adapted enzymes to become unstable (at high temperatures), and thermophilic enzymes to become too rigid to function effectively (at low temperatures).

The results of several recent studies, however, are in marked contrast to those cited above. From the millisecond time scale dynamics of the hyperthermophilic rubredoxin from *Pyrococcus furiosus* investigated by NMR-monitored hydrogen/deuterium exchange, it was concluded that the protein's conformational flexibility at room temperature is indistinguishable from that of mesophilic proteins on this time scale.⁶ The room temperature dynamics of a pair of mesophilic and thermophilic α -amylases were probed using both hydrogen exchange and inelastic neutron scattering.⁸ This study also found no discernable difference in dynamics as monitored by hydrogen exchange, and found increased mobility on the picosecond time scale in the thermophilic protein, as measured by neutron scattering.

The confusion regarding the relationship of conformational dynamics to stability and function in

proteins stems from several sources. Firstly, different studies have monitored flexibility using different techniques. While all of these techniques are sensitive to protein conformational fluctuations, they often monitor very different aspects of these fluctuations. Fluorescence quenching relies on the quenching of fluorescing tryptophan residues by acrylamide, and is therefore only sensitive to those motions that allow acrylamide molecules to penetrate into the core of proteins and interact with buried tryptophan residues. Hydrogen/deuterium exchange is sensitive to both local and global unfolding motions that expose buried amide hydrogen atoms to water. However, under EX2 conditions, where most studies have been performed, H/D exchange rates are proportional to the equilibrium constant for the conformational change(s) that result in the exposure of a given hydrogen atom.¹⁰ H/D exchange under these conditions is therefore a static measure of flexibility: it reflects the equilibrium populations of different conformations.^{11,12} Inelastic neutron scattering is sensitive to the motions of individual hydrogen atoms on the picosecond time scale,⁹ but it provides no spatial resolution. One can only measure the distribution of amplitudes for an entire protein and thus cannot assign a given amplitude of motion or relaxation time to a particular hydrogen atom.

In addition to the fact that protein mobility was probed using different techniques, the various studies of dynamics in thermally stable enzymes have employed different proteins, often with quite distinct native topologies. It is possible that different protein structures have had their dynamics altered in different ways by adaptation to high temperature. Finally, a single pair of homologous thermophilic and mesophilic enzymes will typically differ at many (often >100) amino acid positions.¹³ While some of these amino acid differences will be related to high temperature adaptation, many others will be neutral,¹⁴ the result of genetic drift, or will reflect adaptation of other enzyme properties. Although not all amino acid differences between a mesophilic enzyme and its thermophilic cousin will be directly related to temperature adaptation, they all have the potential to affect dynamics. It is therefore not straightforward to determine which observed differences in protein dynamics are related to high temperature adaptation and which are the result of neutral drift or adaptation to unrelated properties, a difficulty analogous to that of interpreting differences in amino acid sequences.¹⁵

The ambiguities introduced by the presence of non-adaptive mutations can be avoided by studying a family of extremophilic enzymes evolved in the laboratory.^{15,16} A *para*-nitrobenzyl esterase (pNBE) (484 residues) from *Bacillus subtilis* was evolved for increased thermostability while its activity at room temperature was retained.^{17,18} The final, eighth generation mutant 8G8 had a melting temperature 18 °C higher than wild-type and a

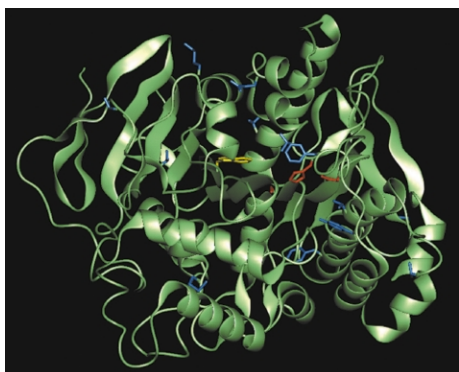


Figure 1. Three-dimensional structure of the *p*NB esterase mutant 8G8. The catalytic triad is shown in red and the 13 mutations from wild-type are shown in blue. Trp 102 is shown in yellow.

room temperature activity twice that of wild-type. In spite of these large functional differences, 8G8 differs from wild-type at only 13 out of 490 amino acid positions. Subsequently, the structures of wild-type, an intermediate mutant in the evolutionary pathway (referred to as 56C8) and the thermophilic mutant, 8G8, were determined by X-ray crystallography.¹⁹ The three-dimensional structure of the thermostable esterase 8G8, including the locations of the thermostabilizing mutations, is shown in Figure 1.

In order to investigate the relationship between stability, dynamics, and evolution, we used the crystal structures of wild-type, 56C8, and 8G8 as the starting point for MD simulations including solvation. These MD simulations show the differences in dynamics between a set of proteins with a wide range of thermal stabilities, evolved under known selection pressures and differing by only a small number of functional mutations.

Results

Validation of the simulations

The overall calculated root-mean-square deviation (CRMS) of all atoms of the minimized wild-type, 8G8 and 56C8 structures from their respective crystal structures are 0.39 Å, 0.72 Å and 0.79 Å. This shows that the forcefield and the surface generalized Born (SGB) solvation method²⁰ are suitable for describing the dynamics of the system. Figure 2(a) shows the overall CRMS value from the crystal structure for all atoms during the MD simulations. The large CRMS change in the initial several picoseconds is the result of heating the system from 0 K to 300 K, and a simulation of the wild-type at 77 K shows a much smaller CRMS value of 1.3 Å after equilibrium (data not shown). From Figure 2(b), it is clear that the system has equilibrated after 100 ps of simulations (within

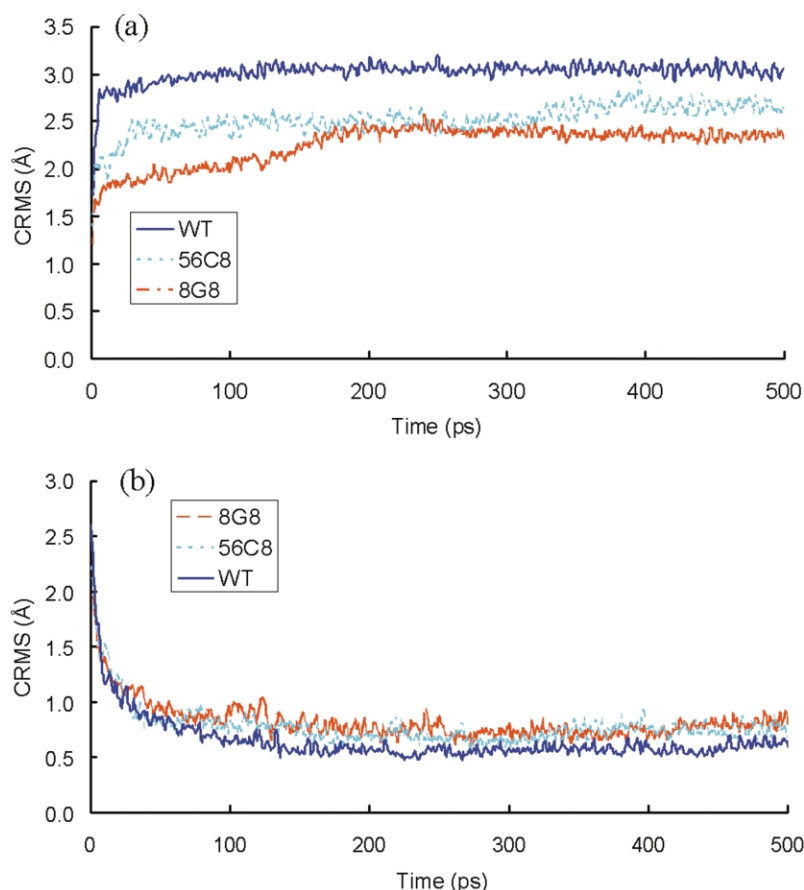


Figure 2. Overall all-atom CRMS from (a) the minimized crystal structure, and (b) the time-averaged dynamic structure for wild-type *p*NBE and the two mutants as functions of time. (Color coding: WT-blue, 56C8-cyan, 8G8-red).

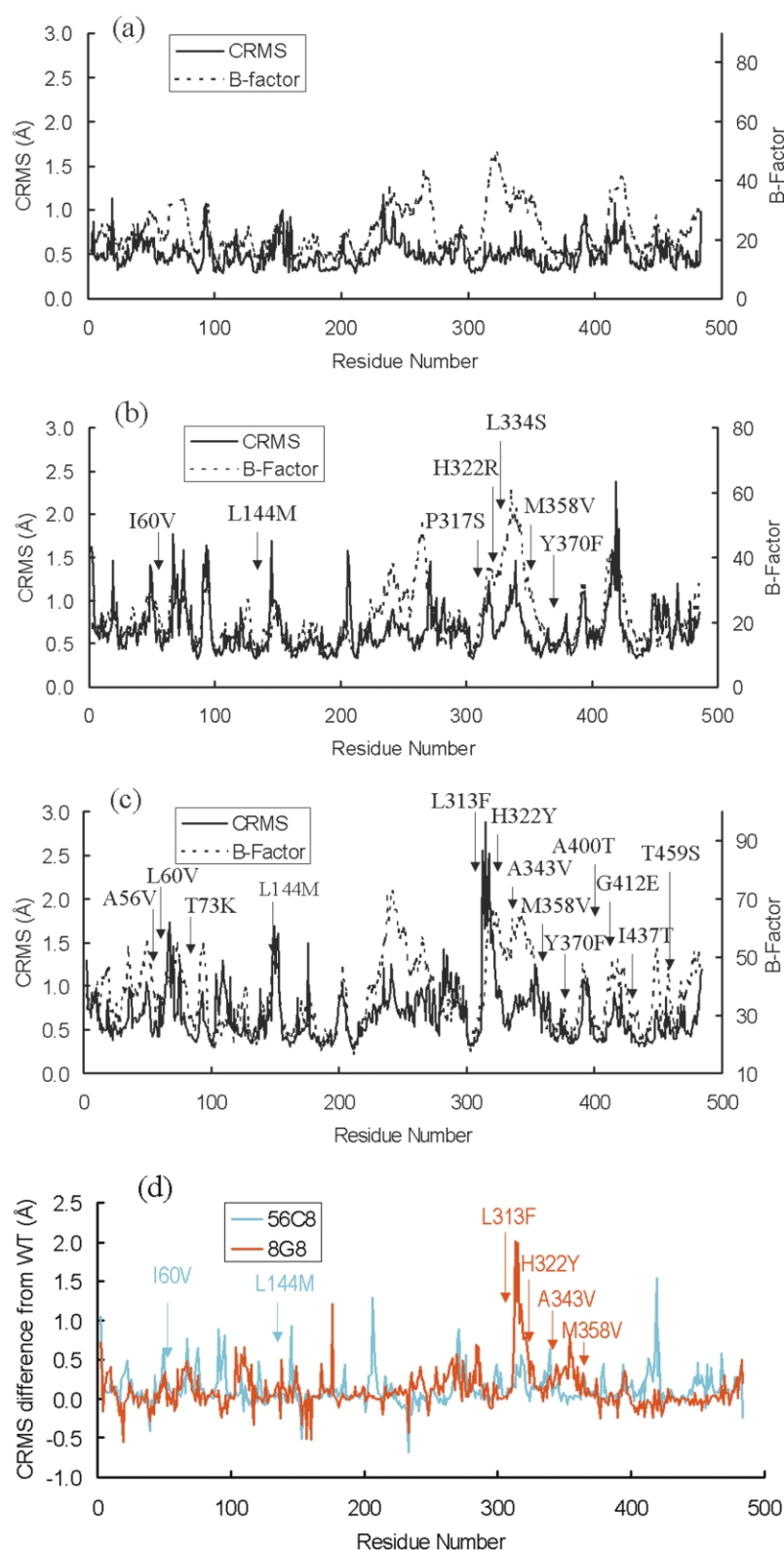


Figure 3. Averaged CRMS from the time-averaged structure during the last 400 ps of the simulation *versus* residue compared with experimental temperature factors for (a) wild-type, (b) 56C8 and (c) 8G8. All mutations in the two variants were labeled in (b) and (c). **Figure 3(d)** shows the difference in CRMS from the time-averaged structure between wild-type and 56C8 (cyan) and between wild-type and 8G8 (red).

~ 0.1 Å CRMS of the average). The overall fluctuations from the crystal structure are smaller for the thermostabilized mutants 8G8 and 56C8 than for wild-type. **Figure 3(a)–(c)** compare the calculated CRMS values for all the residues to the temperature factor (*B* factor) reported in the crystal structure for wild-type, 56C8 and 8G8, respectively.

The fluctuations in CRMS values by residue correlate well with the temperature factors from crystallographic data. For example in **Figure 3(c)**, residues 148–153 and 312–322 in 8G8 have high CRMS values and also show large temperature factors. Both the high *B* factors and the large fluctuations during the simulations are expected for these

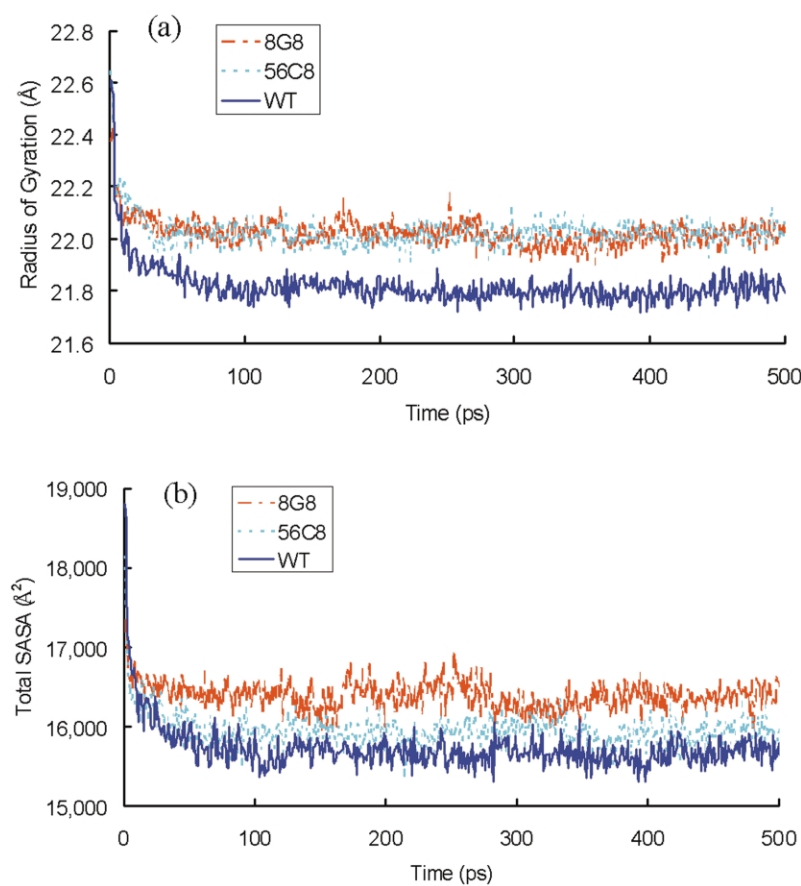


Figure 4. (a) Radius of gyration calculated using equation (2) and (b) Total solvent accessible surface areas for wild-type (blue), 56C8 (cyan) and 8G8 (red) as functions of time.

regions, as they are composed of surface loops. This correlation between experimental and calculated fluctuations provides validation that our simulations describe accurately the dynamics of the system.

Differences in flexibility

The CRMS values from (a) the crystal structure and (b) the time-averaged structure for wild-type and both mutants as a function of time are shown in Figure 2(a) and (b), respectively. These CRMS values are averaged over all the residues. The time-averaged structure is based on coordinates constructed from the average of all snapshots from MD simulations from 100 ps to 500 ps at interval of 1 ps. The time-averaged structure represents the average solution structure for a given protein. Relative to wild-type, the time-averaged structures of the thermostable mutants (particularly the most stable mutant 8G8) are closer to the crystal structures. However, the thermostable mutants actually show increased fluctuations about their time-averaged structures relative to wild-type.

Figure 4(a) shows the calculated radius of gyration for each of the esterases as a function of time. Both 56C8 and 8G8 show a small but statistically significant increase (about 1%) in radius of gyration relative to wild-type. This indicates a more

expanded average structure, possibly as a result of larger or more frequent conformational “breathing” motions. Figure 4(b) shows the total solvent accessible surface area as a function of time for each of the three esterases. Here, 8G8 shows a small but significant increase (about 5%) relative to wild-type. This again indicates a more “open” average structure for 8G8. In contrast, 56C8 shows only a very small difference from wild-type.

Density of states (DOS)

For each of the three esterases, the vibrational density of states was calculated from the velocity autocorrelation function generated from the simulations. While the densities of states for the three structures are quite similar, there are discernible differences. Figure 5(a) shows the power spectra (vibrational density of states) of wild-type, 56C8 and 8G8. The overall shape of the spectra are similar to the power spectra of other proteins, including those calculated from MD simulations (as done in this work) and from normal mode analysis.²¹ It is also similar to neutron scattering spectra obtained for globular proteins.²¹ This demonstrates that our sampling rate is sufficient to capture the salient features of the dynamics, including the peak at $\sim 3000\text{ cm}^{-1}$ which corresponds to hydrogen vibrations. Significant differences can be discerned in the DOS profiles at the lower wavenumbers. Figure 5(b) shows an

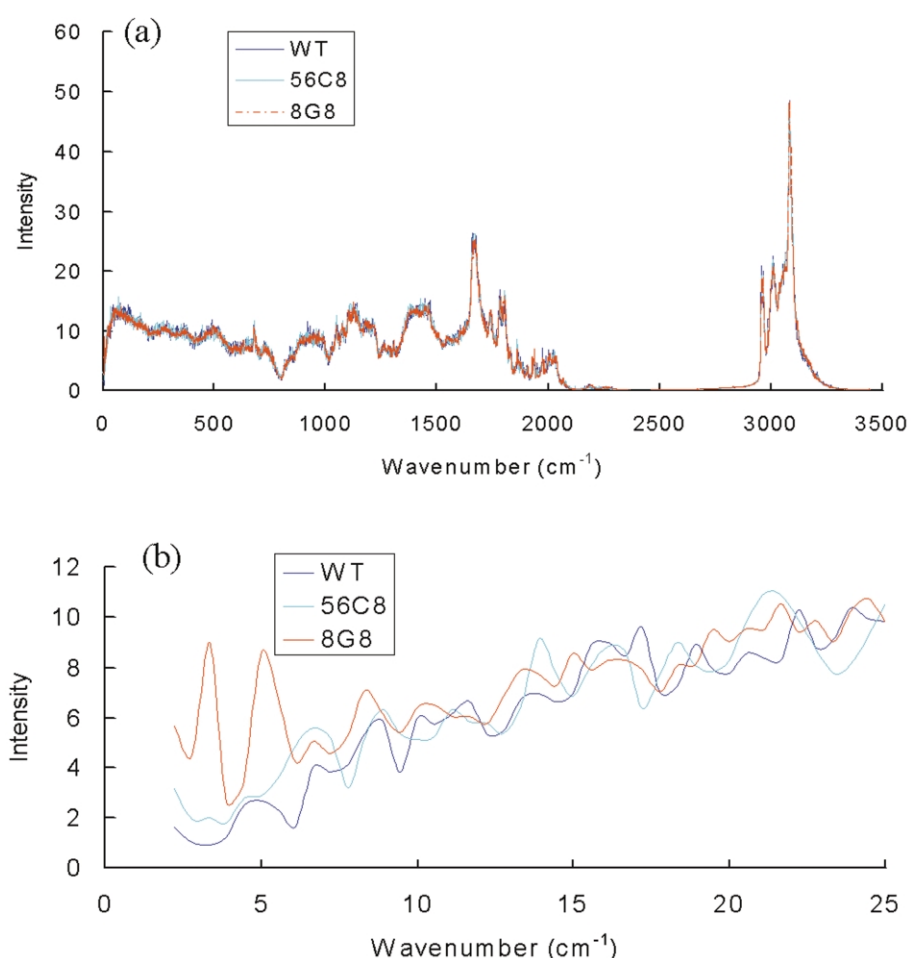


Figure 5. (a) Vibrational density of states calculated using equation (1) for wild-type (black line), 56C8 (blue line), and 8G8 (red line). (b) Expanded view of the DOS from 2.0 cm^{-1} to 25 cm^{-1} . Intensity of DOS has the unit of number of modes per cm^{-1} .

expanded view of the DOS profiles from 2.0 cm^{-1} to 25 cm^{-1} . Relative to wild-type, thermostable variants have substantially increased their populations of the lowest frequency modes (below $\sim 10\text{ cm}^{-1}$). Theoretical studies²² have established that the modes in this region correspond to collective motions, in which many atoms in a given protein region move in a concerted manner. The thermostable esterases thus appear to undergo concerted motions more often than wild-type, in contrast to what would be expected if higher stability were always accompanied by reduced mobility. The degree to which these low frequency modes have been increased varies among the mutants, with 8G8 showing a larger increase than 56C8. The rank order of the degree of increase in the low frequency modes thus matches the rank order of thermostability for the 8G8 and 56C8 variants, with the more stabilized mutant showing larger increase. That 56C8 and 8G8 experience concerted motions more frequently than wild-type is very consistent with the observation that both of these mutants have a larger radius of gyration than wild-type.

Localized changes in mobility

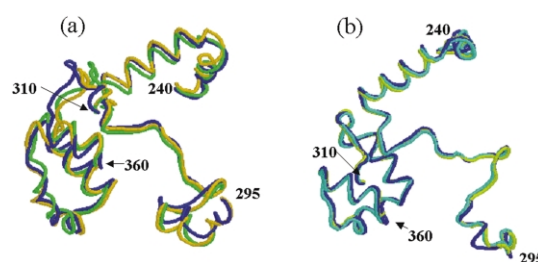
As thermostability is increased, both the N and C-terminal domains of pNBE show significant reductions in CRMS from the crystal structure. At the same time, CRMS from the time-averaged structure is found to increase in many regions. Figure 3(d) shows the difference in CRMS fluctuations about the time-averaged structure between wild-type and the two mutants calculated for each residue. This is calculated by subtracting the CRMS value given for wild-type in Figure 3(a) from the CRMS value given for each of the two mutants in Figure 3(b) and (c). Some important mutations are labeled in the plot. It is seen that 8G8 shows larger fluctuations about the time-averaged structure than wild-type, especially for those residues around the mutations. Also these fluctuations are over longer stretches of consecutive residues in 8G8 than wild-type. The most dramatic changes between the thermophilic mutant 8G8 and wild-type are seen in the regions comprised of residues 240–290, 312–360, and 410–420 (Figure 3(d)). Coincidentally, there are

Table 1. Fluctuations of the catalytic triad (Å) during the last 400 ps of MD simulation

Residue	Wild-type	56C8	8G8
Ser189	0.35	0.32	0.34
His399	0.39	0.60	0.42
Glu310	0.32	0.44	0.39

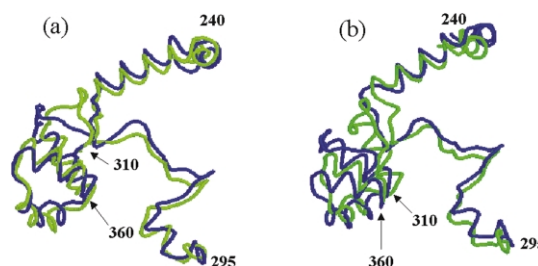
several mutations in this region, most notably L313F and H322Y, which suggests that these mutations confer more flexibility in thermophilic 8G8. Residues 310–320 form a part of the active site, and Glu310 itself is part of the catalytic triad. For all three residues that make up the catalytic triad: Ser189, Glu310, and His399, the CRMS value from the average structure are very close for wild-type and 8G8 (Table 1). This indicates that the dynamic properties of the critical active site residues have been conserved in spite of significant changes in other regions. 56C8 shows larger differences from wild-type at the active site. This may be a reflection of the fact that, experimentally, 56C8 is less catalytically active than wild-type in aqueous solution (56C8 was originally evolved for activity for catalyzing hydrolysis of *para*-nitrobenzyl ester in aqueous organic solvents¹⁷). A region that shows negative differences in CRMS between wild-type and the mutants is comprised of residues 414–420. This is interesting because in most other regions 8G8 shows larger CRMS values than wild-type. These residues form a short turn of α -helix in 56C8 and 8G8. In wild-type, however, they do not appear in the electron density, indicating that this region is highly dynamic or disordered.¹⁹ For purposes of the simulation, these residues were built into the wild-type structure based on the coordinates of 8G8. The fact that this region has high mobility during our simulation of wild-type is consistent with the observation that this region does not appear in the electron density of the wild-type crystal structure, and further confirms that our simulations have correctly described the structural dynamics of the system.

It is of interest to identify the nature of the low frequency modes of motion whose populations are increased in the thermophilic mutants relative to wild-type (see the discussion of the vibrational density of states, above). Aligning structures taken at various times during the simulation allows us to identify those regions that show the largest differences in mobility between wild-type and the mutants. The aligned snapshots of 8G8 show larger average RMS deviations than those of wild-type; consistent with the fact that 8G8 shows larger fluctuations about its time-averaged structure. From the alignment, the largest displacements are located in the regions comprised of residues ~240–295 and ~310–360 (Figure 6(a) and (b)). These regions contain several helices (residues 252–266, 287–294, 326–333, 337–345, and 350–362) whose positions fluctuate up to 6 Å in 8G8, while they remain essentially superimposable in

**Figure 6.** Superimposed snapshots of (a) 8G8 and (b) wild-type taken at 100 ps (blue), 300 ps (green), and 500 ps (yellow). Residues 240–300 and 310–360 are shown.

wild-type. It is interesting to note that these helices shift their positions essentially as rigid bodies, with little evidence of unraveling or deformation. Such motions involve a large number of atoms moving in a concerted manner, and are thus good candidates for the low frequency motions that are increased in the thermostable mutants. Further, the surface loop comprised of residues ~315–323 undergoes fluctuations of up to 6 Å in 8G8, while it shows little movement in wild-type. These motions each involve up to 14 residues moving in a concerted manner, and may thus contribute to the increased density of states seen at wave numbers below $\sim 10 \text{ cm}^{-1}$. We note that in this same region the wild-type structure exhibits large deviations from the crystal structure (Figure 7(a) and (b)). In particular, the helices from residues 326–333 and 337–345 in wild-type shift significantly from their initial positions, while they remain close in 8G8.

The largest difference in CRMS about the time-averaged structure between wild-type and 8G8 is seen in the loop comprised of residues 315–323 (Figure 3(d)). This region is flanked by two stabilizing mutations, L313F and H322Y. L313F forms an edge-face interaction with Phe314, while H322Y forms interactions with a number of residues including Ile270 and Val358 (resulting from the stabilizing mutation M358V). These contacts are maintained during the simulation, and appear to prevent the 315–323 loop from shifting substantially away from its position in the crystal structure (as occurs in wild-type). While the two ends of the loop are thus fixed in place, the region between

**Figure 7.** Superimposed energy minimized crystal structure (blue) and structure at 300 ps (green) for (a) 8G8 and (b) wild-type.

them experiences larger fluctuations during the simulation than in wild-type. The CRMS value from the time-averaged structure for 8G8 (Figure 3(c)) reveals several “spikes”, i.e. stretches of residues in which the CRMS value is markedly greater than in the surrounding regions. Examining the distribution of mutations in Figure 3(c), it can be seen that mutations are found in the “valleys” immediately adjacent to these areas. These observations suggest that the stabilizing mutations in 8G8 act as local anchors, locking down specific regions and preventing them from deviating far from their conformation in the crystal structure. Regions adjacent to these anchor points experience increased fluctuations, though these fluctuations occur about a mean conformation that is closer to the crystal structure than would be the case if the stabilizing mutations were absent. The case of the 215–223 loop flanked by mutations L313F and H322Y illustrate this particularly well, and provides a clearer physical picture of what is meant by the statement that the thermophilic mutants remain closer to their crystal structures while exhibiting increased fluctuations about their time-averaged structures.

Mobility of Trp102

The primary mechanism for non-radiative decay in phosphorescing Trp residues is vibrational coupling between the triplet state and the ground state due to out-of-plane distortions of the aromatic ring.^{23,24} Thus, longer phosphorescence lifetimes indicate reduced local fluctuations. Tryptophan phosphorescence lifetimes have been used to probe local mobility/rigidity in proteins.²⁵ The phosphorescence lifetimes of Trp102 were measured previously for wild-type and several thermostable mutants, including 8G8.¹⁸ Lifetimes were not measured experimentally for 56C8, but were determined for the closely related mutant 1A5D1 (56C8 has all of the stabilizing mutations present in 1A5D1 plus two additional mutations, L334S and P317S, which do not lie near Trp102). It was found that all mutants show increased phosphorescence lifetimes relative to wild-type, with 8G8 showing the largest increase (1.9 times that of wild-type). The lifetime measurements for Trp102 provide an ideal means for assessing the quality of our simulations, since the motions responsible for phosphorescence decay are believed to occur on the picosecond timescale. Figure 8(a) shows the all-atom CRMS value from the time-averaged structure of Trp102 for wild-type, 56C8, and 8G8. The average CRMS value in all three enzymes ~ 0.5 Å. However, the frequency and magnitude of larger deviations is clearly greater for wild-type than for 8G8. Figure 8(b) shows the out-of-plane bending motion angle for Trp102 as a function of time. From Figure 8(a) and (b) it is clear that Trp102 in wild-type shows larger instantaneous fluctuations in CRMS value from the crystal structure and also higher values for $\sin^2 \theta$ in the out-of-

plane bending motion than either 56C8 or 8G8. If the time-averaged $\sin^2 \theta$ value is plotted *versus* the inverse phosphorescence lifetime, the data can be fit into a straight line (Figure 8(c)). This is consistent with the theory that phosphorescence lifetime inversely correlates with $\sin^2 \theta$.²³ These motions occur at the picosecond timescale and thus are in accordance with the experimental findings. The degree of larger deviations for 56C8 falls between wild-type and 8G8. Thus the degree of Trp102 motion on the picosecond timescale ranks as wild-type > 56C8 > 8G8. This is in excellent correspondence with what is expected from the observed phosphorescence lifetimes. As expected, an inverse correlation is seen between the degree of Trp102 motions during the simulation and the measured phosphorescence lifetimes.

Discussion

The most striking result from this study is the apparent discrepancy between different measures of flexibility. Based on the CRMS value from the crystal structure it appears that the thermophilic mutant 8G8 is indeed more rigid than its mesophilic parent. However, the CRMS value from the time-averaged structure, the radius of gyration, and the population of low frequency modes all indicate that 8G8 is in fact more mobile than wild-type. These observations are not necessarily contradictory. Proteins experience a wide array of motions spanning a vast range of amplitudes and timescales. There is no reason to expect that all of these different modes of motion will change in the same way as a given protein evolves to become more stable.

Our results indicate a need to refine the term “flexibility” as it is currently used in discussions of temperature adaptation in proteins. The timescales, amplitudes, and locations of the motions presumed to contribute to flexibility must be specified. Clearly certain modes of motion, particularly those that may initiate unfolding, must be reduced if a protein is to achieve increased stability at elevated temperatures. In the case of pNBE, this is perhaps being reflected in the reduced deviations of surface loops from their positions in the crystal structure.

For many other types of motions, however, there is little evidence that they are detrimental to global stability. In fact, there is increasing evidence that various modes of motion contribute to the stability of the native state. In general, work has focused on small amplitude local fluctuations. Recent neutron scattering studies of a mesophilic and thermophilic α -amylase found that the thermostable enzyme displayed increased mobility relative to the mesophilic one on the 0.3–6.0 ps timescale.^{8,9} These same studies, however, found no difference in mobility between the two proteins in the unfolded state, suggesting that increased conformational entropy in the native state may

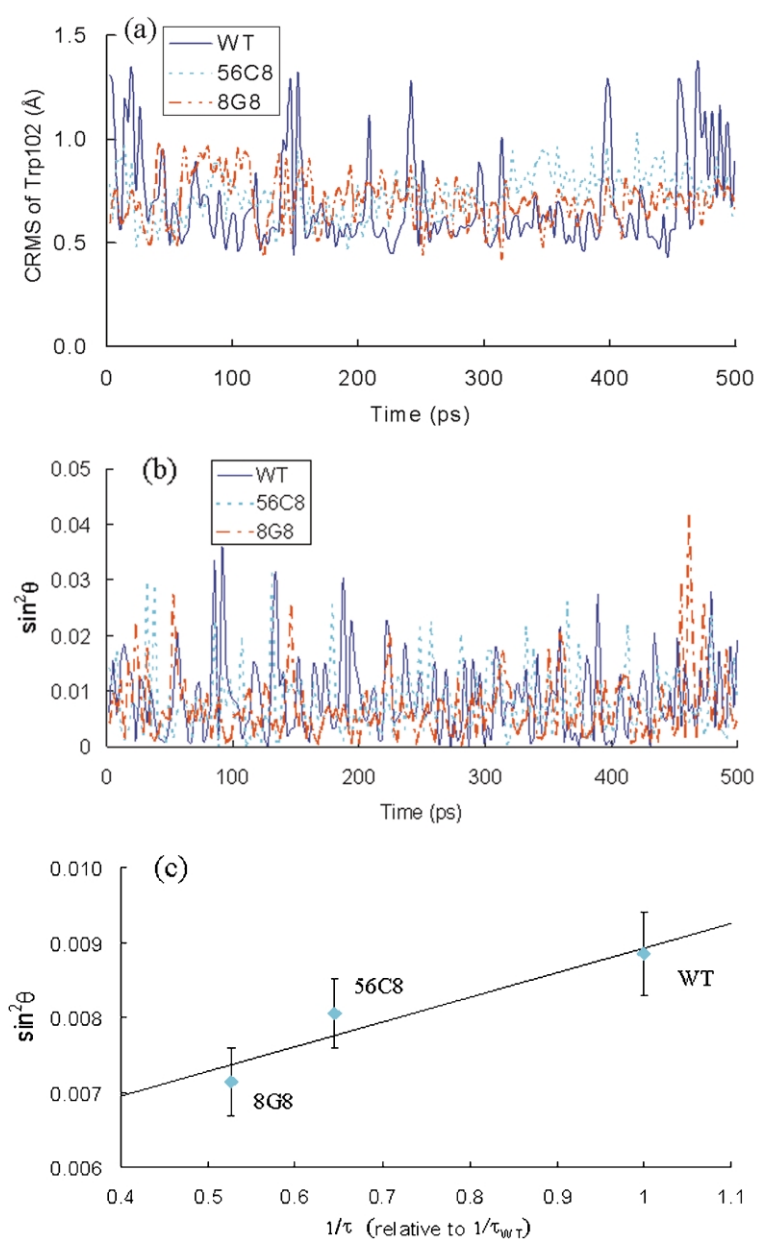


Figure 8. (a) All-atom CRMS from the time-averaged dynamic structure for Trp102 as a function of time. (b) Out-of-plane bending motions (see text) for Trp102 as a function of time. (c) time-averaged $\sin^2\theta$ versus the inverse of phosphorescence lifetime. Experimental phosphorescence lifetimes for wild-type, 1A5D1, and 8G8 were taken from Ref. 18. The phosphorescence lifetime for 56C8 was not measured experimentally, but was believed to be similar to 1A5D1, a closely related mutant (see text).

contribute to the stability of the thermophilic enzyme. Recent NMR relaxation studies of bond vector motion in proteins have led to estimates of the contribution of native-state conformational entropy to stability.²⁶ These studies conclude that local internal protein motions may increase stability by increasing the entropy of the native state. Such motions can also raise the melting temperature of a protein by increasing the heat capacity of the native state and thus decreasing the heat capacity difference between the native and denatured states ($\Delta C_p^{\text{unfold}}$).²⁷ Such a decrease in ΔC_p will expand the temperature range in which a protein remains folded by both raising the melting temperature and decreasing the cold-denaturation temperature. The idea that protein motions contribute to stability is supported by the opposing trends we observed for the CRMS deviations from the esterase crystal structures and

the CRMS fluctuations about the time-averaged structures during the simulations. Relative to wild-type, the thermophilic esterase 8G8 maintains its average structure closer to the crystal structure even while it experiences greater fluctuations about this average. This idea is also supported by the fact that the thermophilic variants show increases in the population of the low frequency modes. Further, calculation of thermodynamic parameters from the velocity autocorrelation function does indicate that thermostabilization is accompanied by increases in the absolute native state entropy and heat capacity. Due to the limited simulation time it is unlikely that these calculated values would agree exactly with measured values (excepting the heat capacity, these absolute values are not practically amenable to experimental measurement in any case). However, they do indicate that, on this timescale, the thermostable

mutants explore more degrees of freedom than wild-type.

The dynamics of folded proteins are often described in terms of a high dimensional energy landscape consisting of many conformational sub-states (local energy minima) separated by energy barriers.²⁸ In this picture, transitions may be expected to occur frequently between conformations of similar energy separated by low barriers. In contrast, transitions between conformations separated by large energy barriers will be observed only rarely. Our results indicate that different aspects of the energy landscape can respond to the demands of high temperature adaptation in different ways. On a global scale, the fact that 56C8 and 8G8 are more stable than wild-type indicates that the distance between the native state ensemble and the denatured state ensemble has increased. Both mutants remain closer to their minimum energy structures during the simulation. An equivalent statement is that conformational substates, which depart substantially from the minimum energy structure, are being sampled less frequently. This suggests that the energetic barriers separating such substates from the minimum energy state have been increased. At the same time, the increased fluctuations about the time-averaged structure indicate that the barriers between some conformational substates have been reduced.

In addition to these differences in overall mobility, it is also clear that different regions of the protein have manifested different changes in mobility in response to temperature adaptation. In general, surface loops show the largest differences in mobility (as measured by both CRMS from the crystal structure and the time-averaged structure) in going from wild-type to 8G8. Large displacement of surface loops from their native conformation has been identified in other studies as an important factor in thermostability.^{29,30} Such large loop displacements may expose the hydrophobic core of a protein to water penetration, leading to unfolding. It is not unexpected, therefore, to find that alterations in loop mobility accompany the thermostabilization of pNBE. At the same time other regions of the protein display much smaller changes in mobility as a result of temperature adaptation. In particular, the RMS fluctuations of the three catalytic residues are remarkably conserved between wild-type and 8G8 (Table 1). While fluctuations about the time-averaged structure increased in many regions of the thermostabilized pNBE variants, they decreased in the immediate vicinity of Trp102, consistent with the experimental observation that its phosphorescence lifetime is longer in more stable mutants.

It is useful to compare our results with two recent simulation studies of thermophilic and mesophilic proteins. Colombo & Merz³⁰ performed extensive simulations of wild-type subtilisin E and a homology model of its thermophilic counterpart

5-3H5, which was generated by directed evolution. They also found a difference in the flexibilities as measured by CRMS from the crystal structure and by CRMS from the time-averaged structure. While the thermostable enzyme showed smaller deviations from the crystal structure, it showed larger fluctuations about the time-averaged structure. The simulations reported here allow for comparisons with experimental data in the form of crystallographic temperature factors and tryptophan phosphorescence lifetimes. The observed correlation of our results with the available experimental data increases our confidence that the increased flexibility seen in the thermostable mutants is real and not simply an artifact of the simulation.

In another study, Lazaridis *et al.*⁷ performed simulations on mesophilic and hyperthermophilic rubredoxins. In contrast to the results of this study and that of Colombo & Merz, Lazaridis *et al.* found that, by all measures, the hyperthermophile showed slightly reduced mobility relative to the mesophile at room temperature. The hyperthermophilic rubredoxin was natural rather than laboratory evolved, and the results of Lazaridis *et al.* may indicate that natural thermophilic enzymes are altered in their conformational mobility in fundamentally different ways than laboratory evolved thermophilic enzymes (the experimental results from the α -amylases argue against this, however). It is also possible that the low mobility of the hyperthermophilic rubredoxin is the result of functional considerations. In contrast to enzymes, which are thought to require a considerable degree of mobility in order to catalyze reactions efficiently, the rubredoxins are thought to undergo only very minor conformational changes during oxidation/reduction.³¹

Given the variety of protein topologies and functions, it is not surprising to find that different proteins have responded to the challenge of high temperature adaptation in different ways. The family of laboratory-evolved pNB esterases has provided us with a unique opportunity to examine how protein dynamics change in response to adaptive evolution without the complications and ambiguities introduced by neutral mutations and unknown selective pressures. Our study has found that adaptation to high temperature results in a rich variety of alterations in dynamic behavior, including reductions in certain types of motions and increases in others. This may help explain the apparently contradictory experimental results that have been reported in studies of thermophilic enzymes. Some techniques, such as H/D exchange of the slowly exchanging hydrogen atoms, will be sensitive primarily to motions that involve significant departures from the native state. We have seen a reduction of such displacements from the minimum energy structure in our simulation of the thermophilic esterases. Other techniques, such as neutron scattering and nuclear magnetic relaxation, can detect small amplitude

local fluctuations, which we have seen increased in the thermophilic esterases. Our study suggests that the apparently contradictory results reported from these diverse experimental techniques might reflect different aspects of dynamic adaptation to high temperatures.

Materials and Methods

The three crystal structures WT, 8G8 and 56C8 were used as the starting structure for all the MD simulations (RCSB PDB entries 1QE3, 1C7J and 1C7I). All the hydrogen atoms were added explicitly using Polygraf, and counterions Na^+ and Cl^- were added to neutralize the side-chains of Asp, Glu, Arg and Lys.³² These counterions are allowed full freedom to move in the dynamics, but stay close to the original positions.

We used Dreiding forcefield³³ with the charges from CHARMM22.³⁴ The non-bond (Coulomb and van der Waals) interactions were calculated using the Cell Multipole Method³⁵ (CMM) for fast and accurate calculations of non-bond interactions. CMM scales linearly with the number of atoms, so that no cutoffs are required.

Since inclusion of explicit water in the MD simulations is computationally intensive we used the SGB continuum solvation model²⁰ to calculate both the energy and the forces due to the solvent acting on the protein structure. SGB accounts for the response in the protein conformation due to electrostatic interactions with the solvent, which is assumed to extend beyond the solvent-accessible surface of the protein. The only properties of solvent required are the dielectric constant (78.32) and the solvent radius (1.4 Å). We assume an ionic strength of 0.1. We assumed that the internal dielectric constant of the protein is 2.0. The SGB method is an approximation to the Poisson–Boltzmann continuum solvation model³⁶ for calculating energies and forces but is much faster. Of course both are faster than explicitly including the water solvent.

The total potential energy of each of the three structures was minimized using conjugate gradients. The minimization was performed for 1000 steps with a termination criterion that the RMS force is less than 0.1 kcal/mol per Å.

Constant temperature Hoover MD simulations³⁷ were carried out on all three structures for 500 ps. The temperature of the simulations was set to 300 K.

The parallel MPSIM MD program^{38,39} was used for all simulations.

Calculation of RMS difference in coordinates

As a simple measure of similarity between two structures we calculate the root mean square of the differences in coordinates between all corresponding atoms in the two structures. First an optimal translation and rotation is performed to superpose the center of mass and the moments of inertia and then the coordinates of equivalent positions are compared. This is referred to as the coordinates root mean square (CRMS) difference. In calculating the CRMS (and radius of gyration) we do not include the counterions.

We first validated our forcefield by calculating the CRMS differences between the experimental crystal structure and the energy-minimized structure. Generally

a CRMS lower than 0.6 Å is considered to indicate that the forcefield is sufficiently accurate.

We also report CRMS differences between the snapshots of the MD structures at various time intervals and the starting minimized energy structure. The CRMS with respect to the minimized structure shows the temperature factors for various regions of the protein indicating which parts are more flexible than others (in solution).

The average MD structure was calculated by averaging the coordinates of the various MD snapshots from 100 ps to 500 ps at 1 ps time intervals. This average structure represents the structure of the protein equilibrated in salt and solvent.

Calculation of density of states:

The vibrational density of states (power spectrum) was calculated from the Fourier transform of the velocity auto-correlation function:

$$S(\nu) = 2\beta \sum_{j=1}^{3N} m_j \tilde{C}_{\nu\nu}(\nu) \quad (1)$$

where $\tilde{C}_{\nu\nu}(\nu)$ is the Fourier transform of the velocity auto-correlation function $C_{\nu\nu}(t)$, m_j is the mass of atom j , $\beta = 1/(k_B T)$ and k_B is Boltzmann's constant, T is the absolute temperature of the system. Sampling of the velocities was done as low as every 1 fs to obtain the high frequency and less often to obtain the low frequency modes.

Calculation of solvent accessible surface area

The solvent accessible surface area is calculated using Connolly's molecular surface calculation program in which a probe is rolled along the surface of the protein.⁴⁰ The probe size used was 1.4 Å.

Calculation of radius of gyration

The radius of gyration is calculated using the following definition:

$$R_G = \left(\frac{\sum m_i r_i^2}{\sum m_i} \right)^{1/2} \quad (2)$$

in which r_i is the distance of the atom i from the center of mass of the protein molecule, and m_i is the mass of the i th atom.

Calculation of Trp out-of-plane bending modes

In the absence of oxygen, the phosphorescence lifetime of a buried Trp is primarily determined by the out-of-plane motion. In the present calculations we have defined out-of-plane bending as involving the nitrogen atom in the indole ring.^{23,24} The out-of-plane bending is measured as the angle, θ , between the p orbital of the N atom and the normal of the six-member ring. The phosphorescence lifetime of the indole ring decreases with increasing $\sin^2 \theta$ value.²³

Acknowledgements

We thank Dr Anne Gershenson (Brandeis University) for her insight and helpful discussions. This work was supported in part by the Army Research Office and the Center for Science and Engineering of Materials at Caltech (NSF-MRSEC). The facilities of the Materials and Process Simulation Center used in this project are supported also by DOE (ASCI ASAP), NSF (CTS and MRI), NIH, ARO-MURI, Chevron Corp., MMM, Seiko-Epson, Dow Chemical, Avery-Dennison Corp., Kellogg's, General Motors, Asahi Kasei, the Beckman Institute, and ONR.

References

1. Jaenicke, R. & Bohm, G. (1998). The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **8**, 738–748.
2. Fields, P. A. (2001). Protein function at thermal extremes: balancing stability and flexibility. *Comp. Biochem. Phys. A*, **129**, 417–431.
3. Varley, P. G. & Pain, R. H. (1991). Relation between stability, dynamics and enzyme activity in 3-phosphoglycerate kinases from yeast and *Thermus thermophilus*. *J. Mol. Biol.* **220**, 531–538.
4. Zavodszky, P., Kardos, J., Svingor, A. & Petsko, G. A. (1998). Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc. Natl Acad. Sci. USA*, **95**, 7406–7411.
5. Hollien, J. & Marqusee, S. (1999). Structural distribution of thermostability in a thermophilic enzyme. *Proc. Natl Acad. Sci. USA*, **96**, 13674–13678.
6. Hernandez, G., Jenney, F. E., Adams, M. W. & LeMaster, D. M. (2000). Millisecond time scale conformational flexibility in a hyperthermophilic protein at ambient temperature. *Proc. Natl Acad. Sci. USA*, **97**, 3166–3179.
7. Lazaridis, T., Lee, I. & Karplus, M. (1997). Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. *Protein Sci.* **6**, 2589–2605.
8. Fitter, J. & Heberle, J. (2000). Structural equilibrium fluctuations in mesophilic and thermophilic alpha-amylase. *Biophys. J.* **79**, 1629–1636.
9. Fitter, J., Herrman, R., Haub, T., Lechner, R. E. & Dencher, N. A. (2001). Dynamical properties of alpha-amylase in the folded and unfolded state: the role of thermal equilibrium fluctuations for the conformational entropy and protein stabilization. *Physica B*, **301**, 1–7.
10. Englander, S. W. & Kallenbach, N. R. (1983). Hydrogen-exchange and structural dynamics of proteins and nucleic-acids. *Q. Rev. Biophys.* **16**, 521–655.
11. Hilser, V. J. & Freire, E. (1996). Structure based calculation of the equilibrium folding pathway of proteins: correlation with hydrogen exchange protection factors. *J. Mol. Biol.* **262**, 756–772.
12. Tang, K. E. & Dill, K. A. (1998). Native protein fluctuations: the conformational-motion temperature and the inverse correlation of protein flexibility with protein stability. *J. Biomol. Struct. Dyn.* **16**, 397–411.
13. Adams, M. & Kelly, R. (1998). Finding and using hyperthermophilic enzymes. *Trends Biotechnol.* **16**, 329–332.
14. Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge.
15. Wintrode, P. L. & Arnold, F. H. (2000). Temperature adaptation of enzymes: lessons from laboratory evolution. *Advan. Protein Chem.* **55**, 161–225.
16. Wintrode, P. L., Miyazaki, K. & Arnold, F. H. (2000). Cold adaptation of a mesophilic subtilisin-like protease by laboratory evolution. *J. Biol. Chem.* **275**, 31635–31640.
17. Giver, L., Gershenson, A., Freskgard, P. O. & Arnold, F. H. (1998). Directed evolution of a thermostable esterase. *Proc. Natl Acad. Sci. USA*, **95**, 12809–12813.
18. Gershenson, A., Schauerte, J. A., Giver, L. & Arnold, F. H. (2000). Tryptophan phosphorescence study of enzyme flexibility and unfolding in laboratory-evolved thermostable esterases. *Biochemistry*, **39**, 4658–4665.
19. Spiller, B., Gershenson, A., Arnold, F. H. & Stevens, R. C. (1999). A structural view of evolutionary divergence. *Proc. Natl Acad. Sci. USA*, **96**, 12305–12310.
20. Ghosh, A., Rapp, C. S. & Friesner, R. A. (1998). Generalized Born model based on a surface integral formulation. *J. Phys. Chem. B*, **102**, 10983–10990.
21. Hinsen, K. & Kneller, G. R. (2000). Projection methods for the analysis of complex motions in macromolecules. *Mol. Simul.* **23**, 275–292.
22. Levitt, M., Sander, C. & Stern, P. S. (1985). Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* **181**, 423–447.
23. McGlynn, S. P. (1969). *Molecular Spectroscopy of the Triplet State*, Prentice-Hall, Englewood Cliffs, NJ.
24. Lower, S. K. & El-Sayed, M. A. (1966). The triplet state and molecular electronic processes in organic molecules. *Chem. Rev.* **66**, 199–241.
25. Schauerte, J. A., Steel, D. G. & FGafni, A. (1997). Time-resolved room temperature tryptophan phosphorescence in proteins. *Methods Enzymol.* **278**, 49–71.
26. Stone, M. J. (2001). NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding. *Acc. Chem. Res.* **34**, 379–388.
27. Seewald, M. J., Pichumani, K., Stowell, C., Tibbals, B. V., Regan, L. & Stone, M. J. (2000). The role of backbone conformational heat capacity in protein stability: temperature dependent dynamics of the B1 domain of Streptococcal protein G. *Protein Sci.* **9**, 1177–1193.
28. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. (1991). The energy landscapes and motions of proteins. *Science*, **254**, 1598–1603.
29. Caflisch, A. & Karplus, M. (1994). Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores and secondary structure of barnase. *Proc. Natl Acad. Sci. USA*, **91**, 1746–1750.
30. Colombo, G. & Merz, K. M. (1999). Stability and activity of mesophilic subtilisin E and its mesophilic homolog: insights from molecular dynamics simulations. *J. Am. Chem. Soc.* **121**, 6895–6903.
31. Yelle, R. B., Park, N. S. & Ichiye, T. (1995). Molecular dynamics simulations of rubredoxin from *Clostridium pasteurianum*: changes in structure and electrostatic potential during redox reactions. *Proteins: Struct. Funct. Genet.* **22**, 154–167.
32. Vaidehi, N. & Goddard, W. A., III (1997). The pentamer channel stiffening model for drug action on human rhinovirus HRV-1A. *Proc. Natl Acad. Sci. USA*, **94**, 2466–2471.

33. Mayo, S. L., Olafson, B. D. & Goddard, W. A., III (1990). DREIDING—a generic force field for molecular simulations. *J. Phys. Chem.* **94**, 8897–8909.
34. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J. *et al.* (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.
35. Ding, H. Q., Karasawa, N. & Goddard, W. A., III (1992). Atomic level simulations on a million particles: the cell multiple method for Coulomb and London nonbond interactions. *J. Chem. Phys.* **97**, 4309–4315.
36. Tannor, D. J., Marten, B., Murphy, R., Friesner, R. A., Sitkoff, D., Nicholls, A. *et al.* (1994). Accurate first principles calculation of molecular charge distributions and solvation energies from *ab initio* Quantum Mechanics and continuum dielectric theory. *J. Am. Chem. Soc.* **116**, 11875–11882.
37. Hoover, W. G. (1984). Computer-simulation of many-body dynamics. *Phys. Today*, **37**, 44–50.
38. Lim, K.-T., Brunett, S., Iotov, M., McClurg, R. B., Vaidehi, N., Dasgupta, S. *et al.* (1997). Molecular dynamics for very large systems on massively parallel computers: the MPSim program. *J. Comput. Chem.* **18**, 501–521.
39. Vaidehi, N. & Goddard, W. A., III (2000). Hierarchical NEIMO simulations for large scale domain motions in phosphoglycerate kinase. *J. Phys. Chem. A*, **104**, 2375–2383.
40. Connolly, M. L. (1983). Analytical molecular surface calculation. *J. Appl. Crystallog.* **16**, 548–558.

Edited by M. Levitt

(Received 14 June 2002; received in revised form 26 November 2002; accepted 14 December 2002)