# Favorable Entropy of Aromatic Clusters in Thermophilic Proteins

**Bruno Trebbi,[†] Francois Dehez,[‡] Patrick W. Fowler,[§] and Francesco Zerbetto*[,†]**

*Dipartimento di Chimica "G. Ciamician", Università di Bologna, via F. Selmi 2, 40126 Bologna, Italy, Unité Mixte de Recherche, CNRS/UHP No. 7565, Université Henri Poincaré, Nancy 1, B.P. 239, 54506 Vandoeuvre-lès-Nancy Cedex, France, and Department of Chemistry, School of Biological and Chemical Sciences, University of Exeter, Stocker Road, Exeter EX4 4QD, United Kingdom*

*Received: May 5, 2005; In Final Form: July 14, 2005*

Aromatic clusters, found with increased frequency in a set of thermophilic proteins, confer an entropic advantage over mesophilic analogues through introduction of large-amplitude oscillations; this is a source of improved free energy at high temperatures.

## Introduction

The origin of the additional stability of thermophilic proteins has received considerable attention both experimentally and theoretically.[1−7] The picture that has emerged is that different protein families adapt to high operating temperatures using different structural tools and that proteins from extreme and moderate thermophiles are stabilized by different mechanisms.[1] Perhaps the rule observed most consistently in the structures of thermophilic proteins is an increase in the number of ion pairs with increasing growth temperature. Other parameters tend to show only qualitative trends.[1] It has been suggested that the presence of extra salt bridges (and hydrogen bonds) results in a lower heat capacity of unfolding than in mesophilic proteins. Higher folding stability and lower heat capacity can both be modeled by a simple approach.[2]

In addition to Coulomb interactions, adaptation to high temperatures involves a number of subtle cooperative effects, often specific to a given protein family. These include (i) minimization of surface energy, (ii) hydration of nonpolar surface groups, (iii) burying of hydrophobic residues, (iv) optimization of core packing, (v) hydrogen bonds, and (vi) optimization of weak protein−protein and protein−solvent interactions. This complex picture is further complicated by the fact that high melting temperature is not always synonymous with greater thermodynamic stability.[3]

One contribution that has come under scrutiny as a source of additional stability of thermophilic proteins is the aromatic electrostatic interaction, leading to so-called aromatic clusters. A graph spectral method was used[4] to identify aromatic clusters for a dataset of 24 protein families for which the crystal structures of thermophilic and mesophilic homologues were available. For 17 different thermophilic protein families, the analysis showed the presence of additional aromatic clusters, or enlarged aromatic networks, absent in the corresponding mesophiles. These clusters were often located close to the active site of the thermophilic enzyme. A geometrical analysis of the packing geometry of the pairwise aromatic interaction showed a preference for T-shaped orthogonal packing.[4] However, a local increase in the energetic stability via improved packing does not unequivocally favor a given mutation, because it may imply concomitant limitations on motion. Rigid structures imply higher vibrational frequencies, which, in turn, imply smaller entropy and militate against decreases in the free energy. In the simplest (harmonic) approximation, entropy is associated with mobility, and the qualitative expectation is that a strongly stabilizing geometrical motif gives a lower entropy and hence a poorer free energy.

We therefore decided to investigate the *entropic* contribution of the mutated aromatic fragments in thermophilic proteins using the dataset of proteins identified by Kannan and Vishveshwara.[4]

## Computational and Theoretical Background

All molecular dynamics calculations were carried out with the TINKER program,[8−10] which has found a number of applications in our laboratory,[11−15] using the AMBER/OPLS/UA force field.[16,17] Only the clusters of the mutated residues in the thermophilic and the mesophilic proteins were allowed to undergo dynamics, subject to interaction with the rest of the protein. For each cluster, 420 ps of dynamics were run, with the initial 20 ps sufficient for equilibration.

To calculate the entropy, a computer program was written based on the approach of Schäfer, Mark, and van Gunsteren, based on the equation[18]

$$S = \frac{1}{2} k_B \ln\left|1 + \frac{k_B T e^2}{\hbar^2} \mathbf{M}^{1/2} \sigma \mathbf{M}^{1/2}\right| \quad (1)$$

where e is the base of natural logarithms, **M** is the diagonal matrix of atomic masses, and $\sigma$ is the covariance matrix of the atomic position fluctuations

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \quad (2)$$

The other symbols have their usual meanings. The larger the mobility of a cluster of atoms, the greater the entropy calculated from eq 1, and in practice, if the position of any particular atom fluctuates greatly, its entropic contribution is large.

* Author to whom correspondence should be addressed. Fax: +390-512099456. E-mail: francesco.zerbetto@unibo.it.
[†] Università di Bologna.
[‡] Université Henri Poincaré.
[§] University of Exeter.

10.1021/jp052333u CCC: $30.25 
Published on Web 08/31/2005

**TABLE 1: Pairs of Thermophilic and Mesophilic Proteins Together with Their PDB Codes[a] and Their Clusters Investigated in This Work**

| protein | cluster | thermophile residues | mesophile residues |
|---|---|---|---|
| 1. neutral protease (1THL/1NPC) | 1 | TYR93 | ILE94 |
| | | TYR151 | ASN152 |
| | | TRP115 | TRP116 |
| | 2 | TYR28 | TYR29 |
| | | TYR24 | LEU24 |
| 2. lactate dehydrogenase (1LDN/1LDM) | 1 | PHE156 | PHE170 |
| | | TRP 187 | TRP201 |
| | | PHE216 | HIS228 |
| | 2 | TYR266 | TYR278 |
| | | PHE300 | LEU313 |
| | 3 | PHE115 | PRO129 |
| | | PHE119 | ILE133 |
| | 4 | PHE103 | PHE117 |
| | | PHE136 | LEU150 |
| | | TYR131 | TYR145 |
| | | TRP134 | TRP148 |
| | | TYR261 | MET273 |
| | | TYR272 | PHE285 |
| | | PHE315 | PHE329 |
| | 5 | PHE51 | LEU65 |
| | | PHE23 | ILE37 |
| | 6 | PHE16 | ALA30 |
| | | TYR234 | SER246 |
| 3. phosphofructokinase (3PFK/2PFK) | 1 | PHE230 | LEU231 |
| | | TYR196 | PHE197 |
| | 2 | TYR38 | TYR39 |
| | | TYR69 | GLY70 |
| 4. ribonuclease H (1RIL/2RN2) | 1 | PHE7 | PHE8 |
| | | TYR67 | SER68 |
| | 2 | TYR72 | TYR73 |
| | | TRP104 | TRP104 |
| | | PHE77 | ILE78 |
| | | PHE118 | TRP118 |
| | | PHE120 | TRP120 |
| | | TRP81 | TRP81 |
| | | TRP85 | TRP85 |
| | | TRP90 | TR90 |
| 5. malate dehydrogenase (1BMD/4MDH) | 1 | TRP184 | TRP184 |
| | | PHE192 | TYR192 |
| | | TYR272 | TYR278 |
| | | TYR280 | TYR286 |
| | | PHE282 | PHE288 |
| | | PHE302 | PHE308 |
| | | TRP213 | TRP218 |
| | | PHE218 | PHE223 |
| | | PHE196 | ASN196 |
| | | TYR214 | LEU219 |
| | 2 | TYR18 | TYR18 |
| | | PHE22 | TYR22 |
| | 3 | PHE62 | LEU62 |
| | | TYR141 | SER141 |
| | | PHE152 | PHE152 |
| 6. hydrolase (2PRD/1INO) | 1 | PHE57 | TYR57 |
| | | TYR32 | ILE32 |
| 7. phospho glycerate kinase (1PHP/3PGK) | 1 | TYR303 | PHE322 |
| | | TYR261 | VAL279 |
| | 2 | PHE225 | PHE240 |
| | | PHE249 | LEU267 |
| | | PHE260 | VAL278 |
| 8. subtilisin (1THM/1ST3) | 1 | TYR174 | ARG164 |
| | | TYR171 | TYR161 |
| | | TYR175 | TYR163 |
| | | TRP199 | GLY189 |
| | 2 | TYR196 | TYR186 |
| | | TYR265 | LEU256 |
| | 3 | TYR210 | GLN200 |
| | | TYR7 | SER3 |
| | | TYR218 | TYR208 |
| 9. reductase (1EBD/1LVL) | 1 | PHE209 | TYR210 |
| | | PHE358 | TYR355 |
| | 2 | TYR321 | ALA318 |
| | | TYR339 | PRO336 |
| | 3 | TYR59 | ARG81 |
| | | PHE194 | LEU193 |
| | | PHE79 | ILE84 |
| | | TYR189 | TYR190 |
| | 4 | PHE115 | VAL120 |
| | | PHE134 | CYS135 |
| 10. triose phosphate isomerase (1BTM/1TIM) | 1 | TYR165 | TYR163 |
| | | TYR209 | TYR207 |
| | | PHE221 | LEU219 |
| | 2 | PHE67 | LYS67 |
| | | TYR73 | PHE73 |
| | 3 | TRP9 | TRP11 |
| | | PHE21 | LEU23 |
| | | PHE242 | PHE239 |
| 11. xylanase (1YNA/1XYN) | 1 | TYR13 | TYR5 |
| | | TRP9 | * |
| | | TYR170 | ASN157 |
| | | TYR171 | TYR158 |
| | | TYR72 | LEU62 |
| | | TYR87 | TYR77 |
| | | PHE92 | ASN82 |
| | | TYR76 | TYR66 |
| | | TRP78 | TRP68 |
| | | PHE133 | PHE121 |
| | | TRP137 | ILE125 |
| | 2 | TYR26 | TYR17 |
| | | TYR14 | ASP6 |
| | | TRP16 | ASN8 |
| | | TYR34 | PHE24 |
| 12. glycosyltransferase (1XYZ/2EXO) | 1 | PHE205 | PHE202 |
| | | TYR228 | PHE222 |
| | | PHE237 | ILE231 |
| | | PHE187 | TYR184 |
| | | PHE277 | VAL270 |
| | 2 | TRP288 | TRP281 |
| | | TRP280 | TRP273 |
| | | PHE293 | PHE286 |
| | | TYR296 | GLU289 |
| 13. triacylglycerol acylhydrolase (1TIB/1LGY) | 1 | PHE51 | ILE19 |
| | | PHE66 | TYR62 |
| | | TYR16 | PHE169 |
| | | PHE13 | GLN193 |
| | | PHE169 | PHE13 |
| | | TYR194 | PHE257 |
| | | PHE10 | ILE10 |
| | | PHE262 | TYR256 |
| | | PHE7 | VAL171 |
| | | TYR261 | ILE48 |
| | | TYR171 | TYR16 |
| 14. pyrophosphatase (2PRD/1OBW) | 1 | PHE57 | TYR57 |
| | | TYR32 | ILE32 |
| 15. carboxypeptidase (1OBR/2CTC) | 1 | PHE272 | TYR265 |
| | | PHE266 | TYR259 |
| | | PHE274 | PHE267 |
| | | PHE233 | LEU219 |
| | | PHE230 | LYS216 |
| | | TYR212 | TYR204 |
| | | TYR216 | TYR208 |
| | | TYR214 | TYR206 |
| | | TYR151 | ALA141 |
| | | TRP264 | TRP257 |
| | | PHE174 | ASN171 |
| | | TYR149 | TRP147 |
| 16. ornitine carboxypeptidase (1AIS/2OTC) | 1 | PHE21 | GLU38 |
| | | TRP168 | LEU197 |

[a] The first is for the thermophile, the second the mesophile.

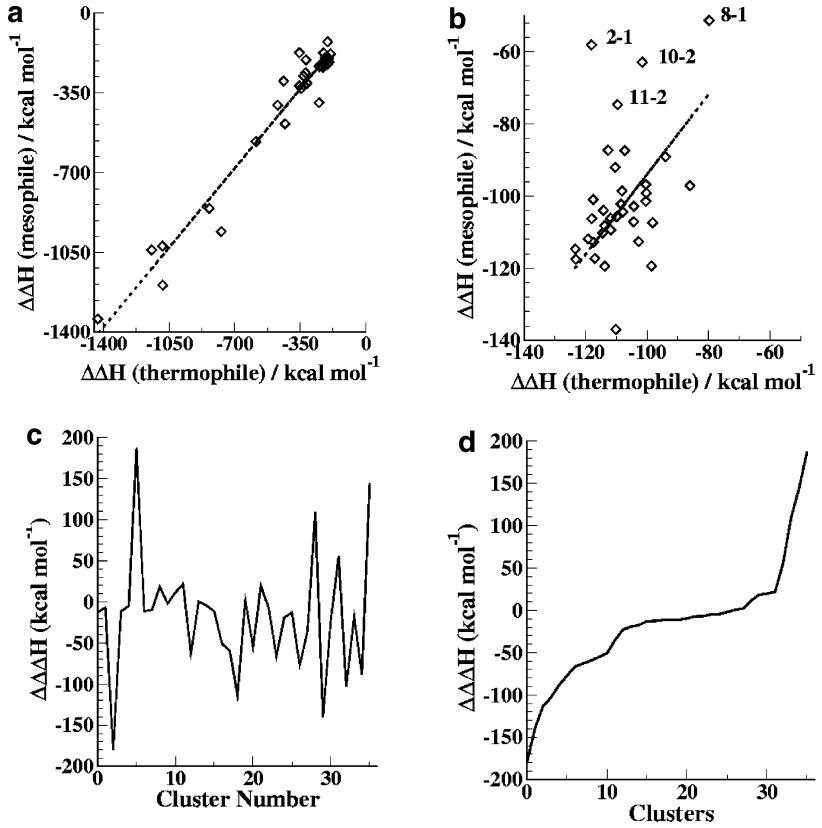


**Figure 1.** Comparison of differential enthalpies of interaction for fragments in thermophilic and mesophilic proteins. $\Delta\Delta H$ accounts for the interaction of a given fragment with the remainder of the protein, after subtraction of the internal energy at 0 K, i.e., the stabilization inherent in formation of the cluster. $\Delta\Delta\Delta H$ is the enthalpy difference between thermophiles and mesophiles. Each datapoint refers to one cluster in a thermophile/mesophile pair: (a) the energy per cluster, the best-fit line (dashed) corresponds to $\Delta\Delta H$(mesophile) = 23.71 kcal mol$^{-1}$ + 1.04 $\Delta\Delta H$(thermophile), with $r = 0.98$; (b) the energy per cluster divided by the number of residues in the cluster, $r = 0.47$; (c) $\Delta\Delta\Delta H$ of the enthalpies of each cluster with the residues in the same order of the original database as in Table 1; (d) $\Delta\Delta\Delta H$ ordered increasingly.

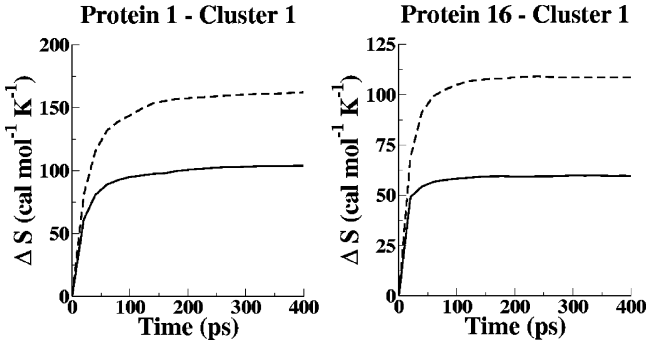


**Figure 2.** Convergence in time of the entropy for the first and last cluster of the 36 investigated. The solid line is for the mesophiles; the dashed line is for the thermophiles.

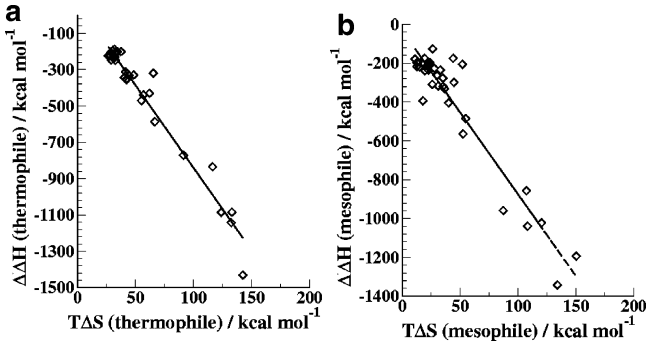


**Figure 3.** Comparison of the entropic, $T\Delta S$, and enthalpic stabilization: (a) thermophiles; (b) mesophiles. $T$ was set to 298 K. The best-fit line corresponds to $\Delta H$(thermophile) = $-9.08T\Delta S + 67.41$, with $r = 0.96$, and $\Delta H$(mesophile) = $-8.41T\Delta S$ −33.32, with $r = 0.91$.

To evaluate the vibrational motions of a cluster as a single unit, every picosecond, we calculate

$$O = (\sum_{i}^{\text{atoms}} (x_i(t+\Delta t) - x_i(t))^2 + \sum_{i}^{\text{atoms}} (y_i(t+\Delta t) - y_i(t))^2 + \sum_{i}^{\text{atoms}} (z_i(t+\Delta t) - z_i(t))^2)/N \quad (3)$$

where $N$ is the number of atoms in the cluster and $O$ is a measure of the overall motion of the cluster. Its Fourier transform gives the frequency of the motion. Three frequency ranges were explored 0−10, 0−30, and 0−50 cm$^{-1}$. After the Fourier transform, we take the integral of the vibrational amplitudes, $I$, in absolute value over this range of frequencies. The result is conveniently expressed in ppm.

## Results and Discussion

Full calculation of the entropy for a large set of proteins is daunting as positional fluctuations may not converge within the time of a molecular dynamics (MD) run. A simplified, reduced approach is in order. It was decided to investigate the vibrational freedom of aromatic clusters in a thermophilic protein and compare it with the motion of the equivalent set of residues in the mesophile. In this way, an entropy can be assigned to the cluster. In the MD calculations, every residue of the clusters (Table 1 of the Supporting Information) undergoes dynamics, subject to interaction with the remainder of the protein, which

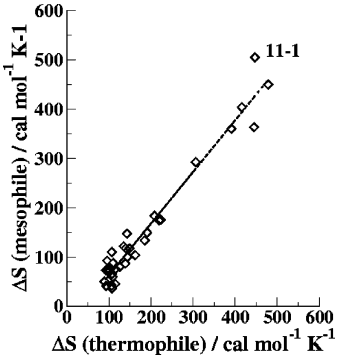


**Figure 4.** Comparison of the entropic contributions of the aromatic clusters and equivalent fragments of thermophilic and mesophilic proteins. The best-fit line (dashed) corresponds to $S$(mesophile) = $-41.42 \pm 7.82$ cal mol$^{-1}$ K$^{-1}$ + $1.047 \pm 0.037S$(thermophile), with $r = 0.98$.

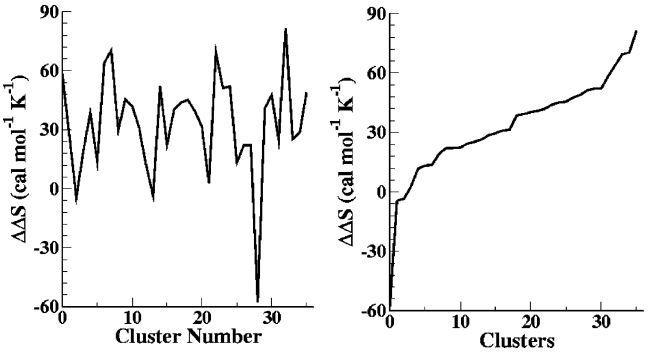


**Figure 5.** Comparison of differential entropies of the fragments in the clusters of thermophilic and mesophilic proteins: (a) $\Delta\Delta S$ for each cluster with the residues in the same order of the original database as in Table 1; (b) $\Delta\Delta S$ ordered increasingly.

is held frozen. Advantages and disadvantages of calculating entropy as the sum of the contributions of individual residues are critically discussed in ref 19. Qualitatively, since entropy is an extensive property, freezing the main body of the protein amounts to neglect of the (second order) effect of fragment motion on that of a much larger object. It is reasonable to expect for the thermophilic and the mesophilic fragments embedded in the protein a similar accuracy. Relative values, or entropy ordering, should be predicted accurately, while the absolute values could be inaccurate.

Overall, 16 pairs of proteins and 36 clusters were investigated; see Table 1. The systems are taken from Table 2 of ref 4.

Free-energy comparison involves both enthalpic and entropic factors. To compare proteins with and without aromatic clusters, it is necessary to define differential properties. We first examine $\Delta\Delta H$, which is given by the average, over each molecular dynamics run, of the energy of interaction between each cluster and the rest of the protein. The term $\Delta\Delta H$ is preferred over $\Delta H$ since the latter entails the contribution from the formation of the covalent bonds, which is not considered here (or is effectively subtracted).

Figure 1 compares these differential enthalpies for the fragments in the mesophilic and thermophilic proteins in the data set. The range of values covered by $\Delta\Delta H$ is substantial. However, once each value is divided by the number of residues in the cluster, the average is $-106$ kcal mol$^{-1}$. This value is similar to that of a single C—C carbon bond but is due to and includes all of the van der Waals and Coulomb interactions between a single residue in the cluster and all of the other residues in the protein. Figure 1 demonstrates a linear correlation between the two sets, with correlation factor $r = 0.98$. Some of the correlation is due to the (trivial) correlation in the number of residues of the mesophilic clusters and their thermophilic counterparts. However, a much weaker correlation, $r = 0.47$, is found when $\Delta\Delta H$ is divided by the number of residues in the cluster; see Figure 1b. As mutations typically involve only a few residues out of the many in the protein, such a linear relationship is unexpected. Significantly, however, the quantities $\Delta\Delta H$ give no additional stabilization of thermophiles over mesophiles beyond that of the inherent internal energy of the local aromatic interaction. Figures 1c and 1d show that for the majority of clusters there is no net enthalpic advantage due to the thermophilic mutations. Indeed, 12 clusters are stabilized, 6 are destabilized, and half of them are neither stabilized nor destabilized.

During the molecular dynamics runs, entropy builds up until convergence is reached. Figure 2 shows two examples of the convergence of entropy (the first and the last cluster of the set). Similar plots for all of the other clusters are reported in the Supporting Information.

Figures 3a and 3b compare the entropic stabilization, $T\Delta S$, with the enthalpic stabilization energy for thermophiles and mesophiles. Within each group, there is a good linear correlation between the two quantities, although the two slopes differ. At 298 K, that is, the temperature used in the plots, the enthalpic factor substantially exceeds the entropic one. The correlation between entropy and enthalpy in the two sets of clusters suggests that two components of free energy are governed by the same factors.

Figure 4 compares the entropies, $S$, of aromatic clusters and the equivalent fragments in the thermophilic and mesophilic proteins. There is another linear correlation between the two sets, $S$(mesophile) = $-40.42$ cal mol$^{-1}$ K$^{-1}$ + $1.047S$(thermophile), with a correlation factor $r = 0.98$. Once again, correlation

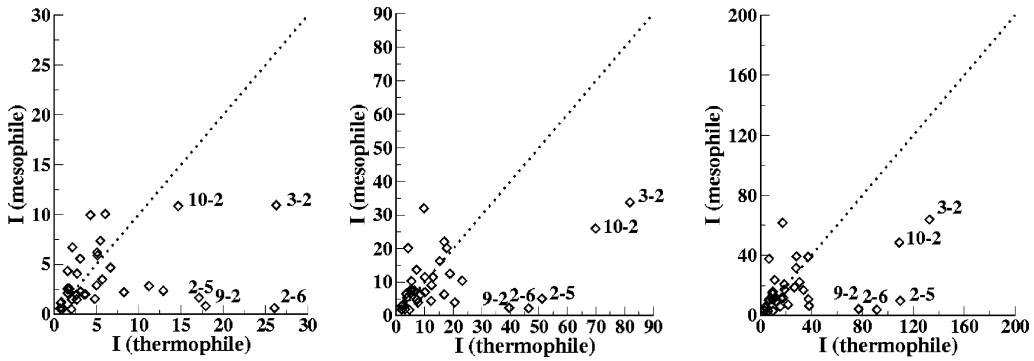


**Figure 6.** Comparison of integrated amplitudes of whole-cluster motions in thermophile and mesophile pairs. The dashed line has unit slope and is used to indicate the divide between cases where the amplitude is larger in the thermophile (the majority) from those where it is larger in the mesophile: (a) the cutoff set to 10 cm$^{-1}$; (b) the cutoff set to 30 cm$^{-1}$; c) the cutoff set to 50 cm$^{-1}$.

of motional entropy of these residues within an unchanged bulk protein is not expected a priori.

The fit indicates a systematic entropic advantage introduced by the "aromatic" mutations. Out of the 36 clusters, only 3 have greater entropy in the mesophilic proteins. This is better appreciated in Figures 5a and 5b where the substantial entropic advantage of the thermophilic mutations is readily perceived.

The correlation evidently has a positive slope, slightly in excess of unity, and an intercept indicating a negative entropy of $\sim 20R$ (where $R$ is the universal gas constant). These two features of the correlation shed light on two different aspects of the cluster motion. First, the slope is qualitatively consistent with the notion that aromatic clusters are locally more rigid: Their internal motions have higher frequencies and contribute less to the entropy than the motions of the corresponding mesophile fragments. Thus, if the intercept of the linear fit were not nonzero, then the presence of aromatic clusters would be an entropic disadvantage, with mesophilic proteins having greater entropy than the thermophiles. However, the intercept is large and negative. Several simple approaches were tested with the intent of explaining the entropic advantage. Correlation of the entropic values with the largest root-mean-square deviation (RMSD) from the equilibrated structure did not show any systematic trends nor did a similar correlation with the number of conformers, $n$, (actually with log $n$) detected during the molecular dynamics run. The relevant data are given as Table 2 in the Supporting Information. The best explanation that we were able to find is based on the hypothesis that the advantage arises from a systematic difference in low-frequency, high-amplitude motions undergone by the cluster units. In the aromatic case, several residues are tightly coupled together and will move together; in the mesophiles, the equivalent residues are less strongly interacting and can be expected to move more independently. Such motions of the relatively rigid aromatic-cluster subunits are expected to be highly anharmonic and to lead ultimately to a higher entropy. This hypothesis can be tested: Integrated amplitudes, $I$, of low-frequency vibrations of clusters in thermophiles and the corresponding fragments in mesophiles were computed and are compared in Figure 6.

The figure shows that most thermophile aromatic clusters have significantly greater integrated amplitudes in their low-frequency region than do the corresponding fragments in mesophiles, a result that is not sensitive to the cutoff. This then is a source of entropic advantage.

## Conclusion

In conclusion, aromatic fragments in thermophilic proteins tend to generate larger entropy via their overall low-frequency motion. This feature indicates one direction for exploration in connection with rational design of ultrastable proteins. Finally, it may be noted that the aromatic residues present in the 36 clusters are tryptophan, tyrosine, and phenylalanine. The latter two (together with cysteine) have substantially increased their frequency of occurrence with respect to ancient proteins over the last 3 billion years.[20] The present work suggests that one advantage of their presence is greater stability arising from their entropic contribution.

**Acknowledgment.** P.W.F. and F.Z. acknowledge support from the European Union RTN scheme under contract HPRN-CT-2002-00177, WONDERFULL.

**Supporting Information Available:** Convergence, in the MD simulation, of entropy of the 36 clusters and the maximum RMSD values, the numbers of conformers, and the entropies. This material is available free of charge via the Internet at http://pubs.acs.org.

## References and Notes

(1) Szilagyi, A.; Zavodszky, P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: Results of a comprehensive survey. *Structure* **2000**, *8*, 493−504.

(2) Zhou, H.-X. Toward the physical basis of thermophilic proteins: Linking of enriched polar interactions and reduced heat capacity of unfolding. *Biophys. J.* **2002**, *83*, 3126−3133.

(3) Ladenstein, R.; Antranikian, G. Proteins from hyperthermophiles: Stability and enzymic catalysis close to the boiling point of water. *Adv. Biochem. Eng./Biotechnol.* **1998**, *61*, 37−85 (Biotechnology of Extremophiles).

(4) Kannan, N.; Vishveshwara, S. Aromatic clusters: A determinant of thermal stability of thermophilic proteins. *Protein Eng.* **2000**, *13*, 753−761.

(5) Dalhus, B.; Saarinen, M.; Sauer, U. H.; Eklund, P.; Johansson, K.; Karlsson, A.; Ramaswamy, S.; Bjork, A.; Synstad, B.; Naterstad, K.; Sirevag, R.; Eklund, H. Structural basis for thermophilic protein stability: Structures of thermophilic and mesophilic malate dehydrogenases. *J. Mol. Biol.* **2002**, *318*, 707−721.

(6) Cowan, D. A. Thermophilic proteins: Stability and function in aqueous and organic solvents. *Comp. Biochem. Physiol., Part A: Mol. Integr. Physiol.* **1997**, *118*, 429−438.

(7) Jaenicke, R.; Bohm, G. 1998, The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **1998**, *8*, 738−748.

(8) Dudek, M. J.; Ponder, J. W. Accurate modeling of the intramolecular electrostatic energy of proteins. *J. Comput. Chem.* **1995**, *16*, 791−816.

(9) Kundrot, C. E.; Ponder, J. W.; Richards, F. M. Algorithms for calculating excluded volume and its derivatives as a function of molecular conformation and their use in energy minimization. *J. Comput. Chem.* **1991**, *12*, 402−409.

(10) Ponder, J. W.; Richards, F. M. An efficient Newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem.* **1987**, *8*, 1016.

(11) Trebbi, B.; Fanti, M.; Rossi, I.; Zerbetto, F. Intraresidue distribution of energy in proteins. *J. Phys. Chem. B* **2005**, *109*, 3586−3593.

(12) Biscarini, F.; Cavallini, M.; Leigh, D. A.; León, S.; Teat, S. J.; Wong, J. K. W.; Zerbetto. F. The effect of mechanical interlocking on crystal packing, predictions and testing. *J. Am. Chem. Soc.* **2002**, *124*, 225−233.

(13) León, S.; Leigh, D. A.; Zerbetto, F. The effect of guest inclusion on the crystal packing of p-*tert*-butylcalix[4]arenas. *Chem.—Eur. J.* **2002**, *8*, 4854−4866.

(14) Höfinger, S.; Zerbetto, F. On the cavitation energy of water. *Chem.—Eur. J.* **2003**, *9*, 566−569.

(15) Teobaldi, G.; Zerbetto, F. Molecular dynamics and implications for the photophysics of a dendrimer−dye guest−host system. *J. Am. Chem. Soc.* **2003**, *125*, 7388−7393.

(16) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(17) Maxwell, D. S.; Tirado-Rives, J.; Jorgensen, W. L. A comprehensive study of the rotational energy profiles of organic systems by ab initio MO theory, forming a basis for peptide torsional parameters. *J. Comput. Chem.* **1985**, *16*, 984−1010.

(18) Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **1993**, *215*, 617−621. Schäfer, H.; Mark, A. E.; van Gunsteren, W. F. Absolute entropies from molecular dynamics simulation trajectories. *J. Chem. Phys.* **2000**, *113*, 7809−7817.

(19) Schäfer, H.; Daura, X.; Mark, A. E.; van Gunsteren, W. F. Entropy calculations on a reversibly folding peptide: Changes in solute free energy cannot explain folding behavior. *Proteins: Struct., Funct., Genet.* **2001**, *43*, 45−56.

(20) Brooks, D. J.; Fresco, J. R. Increased frequency of cysteine, tyrosine, and phenylalanine residues since the last universal ancestor. *Mol. Cell. Proteomics* **2002**, *1*, 125−131.