Protein Engineering vol. 16 no. 12 pp. 871–874, 2003
DOI: 10.1093/protein/gzg131

COMMUNICATION

Optimized electrostatic surfaces parallel increased thermostability: a structural bioinformatic analysis

Eric Alsop[1], Melanie Silver[2] and Dennis R.Livesay[1,3]

[1]Department of Chemistry and [2]Department of Biology, California State Polytechnic University at Pomona, Pomona, CA 91768, USA

[3]To whom correspondence should be addressed.
E-mail: drlivesay@csupomona.edu

It has been known for some time that thermophilic proteins generally have increased numbers of non-covalent interactions (salt bridges, hydrogen bonds, etc.) compared with their mesophilic orthologs. Recently, anecdotal structural comparisons suggest that non-specific acid–base ion pairs on the protein surface can be an evolutionary efficient mechanism to increase thermostability. In this comprehensive structural analysis, we confirm this to be the case. Comparison of 127 orthologous mesophilic–thermophilic protein groups indicates a clear preference for stabilizing acid–base pairs on the surface of thermophilic proteins. Compared with positions in the core, stabilizing surface mutations are less likely to disrupt the tertiary structure, and thus more likely to be evolutionarily selected. Therefore, we believe that our results, in addition to being theoretically interesting, will facilitate identification of charge-altering mutations likely to increase the stability of a particular protein structure.

Keywords: electrostatics/salt bridges/structural bioinformatics/thermostability

Introduction

Enzymes from thermophilic bacteria (optimal growth temperatures >65°C) tend to exhibit substantially higher intrinsic thermal stabilities than their mesophilic (optimal growth temperatures ~30°C) counterparts, while retaining the basic fold characteristics of the whole family. Although the molecular underpinnings of protein thermostability have been the focus of many experimental and theoretical research efforts, the subject is only partially understood (Vieille and Zeikus, 2001). In general, it is accepted that thermostability is achieved by an increase in the type and number of intramolecular interactions (Kumar *et al*., 2000). Structural comparisons of thermophilic and mesophilic orthologs generally reveal increased numbers of van der Waals interactions, hydrogen bonds, salt bridges, dipole–dipole interactions, disulfide bridges and hydrophobic interactions within thermophilic proteins (DeDecker *et al*., 1996; Schafer *et al*., 1996; Chang *et al*., 1999; Declerck *et al*., 2000; Szilagyi and Zavodszky, 2000). Other differences include shortening of loop regions, fewer and smaller destabilizing voids within the structure and increased conformational rigidity (Tomschy *et al*., 1994; Tanner *et al*., 1996; Thompson and Eisenberg, 1999).

Several recent efforts have attempted to identify the most efficient method of conferring enhanced thermostability to a mesophilic protein structure. Early efforts in this direction have concentrated on engineering new structural tethers (covalent and non-covalent interactions) within the protein structure (Perry *et al*., 1989; Scholtz and Baldwin, 1992; Serrano *et al*., 1992; Bryson *et al*., 1995; Matthews, 1995; Pace, 1995; Cordes *et al*., 1996; Lehmann *et al*., 2000; Lehmann and Wyss, 2001). This approach is well justified because it is becoming increasingly clear that protein folding is a hierarchical process and thus is mostly driven by local interactions (Baldwin and Rose, 1999a,b). However, recent results reveal that the complex electrostatic properties on the protein's surface can also contribute significantly to its thermostability. Several recent studies have successfully increased mesophilic protein stability through mutagenesis of a single solvent exposed residue, presumably through optimization of its electrostatic surface (Grimsley *et al*., 1999; Loladze *et al*., 1999; Perl *et al*., 2000; Spector *et al*., 2000; Strop and Mayo, 2000; Martin *et al*., 2001; Pedone *et al*., 2001; Perl and Schmid, 2001; Loladze and Makhatadze, 2002). Additionally, our comprehensive *in silico* screening of surface mutants confirms that optimization of the electrostatic surface is a robust strategy for conferring thermostability to mesophilic proteins (Torrez *et al*., 2003). From these results, it is apparent that surface electrostatics are intimately related to protein stability, and that optimization of the electrostatic surface is an attractive mechanism for conferring increased thermostability to a protein structure. Thus, we now inquire to the extent that nature has also employed similar design mechanisms.

There is a growing list of anecdotal evidence suggesting that optimization of the electrostatic surface of thermophilic proteins is a robust evolutionary theme (Fukuchi and Nishikawa, 2001). Comparisons of the amino acid composition of thermophilic versus mesophilic proteomes reveal significantly more charged residues (especially Arg, Lys and Glu) within thermophilic proteins (Chakravarty and Varadarajan, 2002; La *et al*., 2003). The charge composition differences are generally ascribed to greater numbers of stabilizing ion pairs in thermophilic structures, yet it is unclear what percentage of these differences are specific charge–charge interactions within the core (salt bridges) versus more promiscuous long-range interactions on the protein surface. Here, we report a clear preference for stabilizing solvent exposed (opposite charge) ion pairs on the surface thermophilic protein structures.

Materials and methods

Solvent exposed residue calculation

Our program, Surface Finder, iteratively evaluates the solvent accessibility of each atom (excluding hydrogens). Here, if an acid or base 'charged atom' is deemed solvent accessible, then that residue is included in the ion pair distance analysis. Charged atoms are defined as atoms explicitly carrying the formal charge (NZ of Lys) or atoms central to resonance-equivalent charged atoms (CG of Asp, CD of Glu, CZ of Arg and CD2 of His). All distances reported are linear between

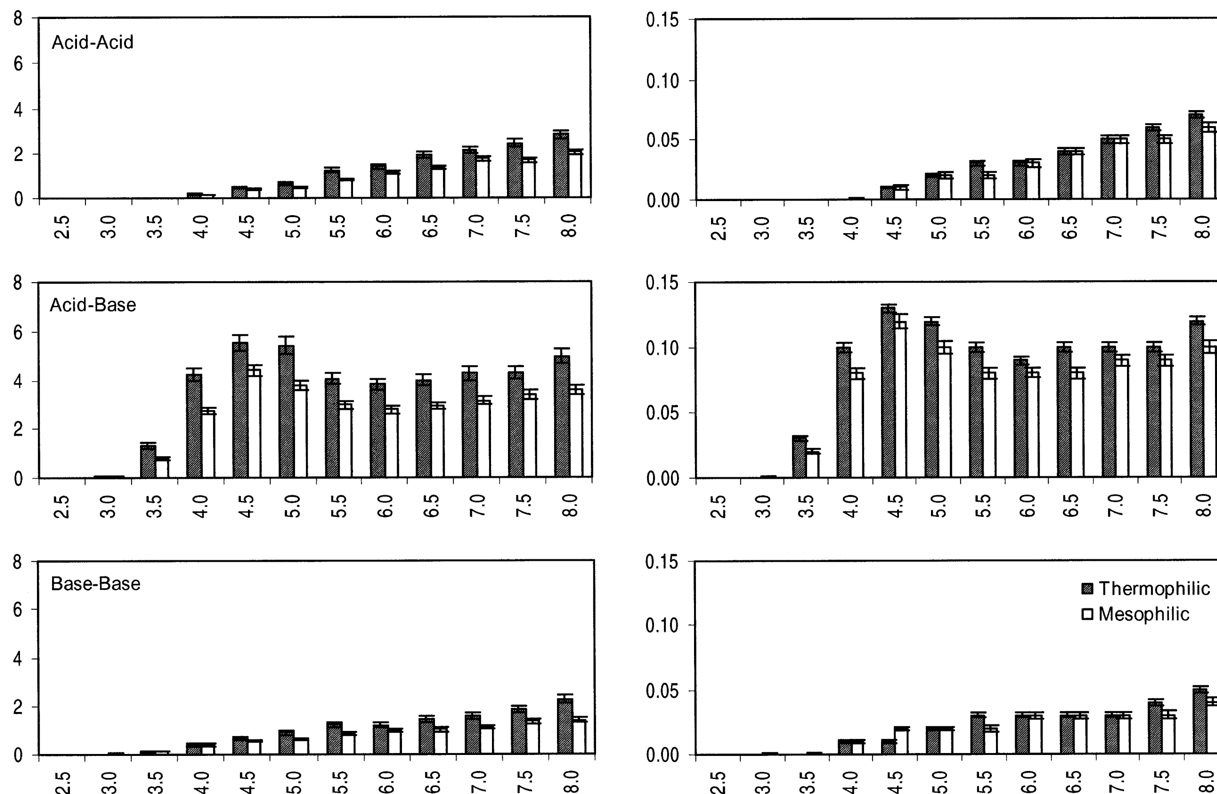

Fig. 1. Average occurrence versus distance (Å) for all mesophilic and thermophilic charge–charge ion pairs (left) and the normalized values (right). Normalized values are calculated by dividing the raw data by the total number of charged residues. The increase in acid–base pairs in thermophilic protein structures suggests that optimizing the electrostatic surface of the protein is a robust evolutionary theme. Ninety-five percent confidence intervals are presented as error estimates.

charged atom pairs; no attempt to correct for curvature of the protein's surface has been made, although future work will. Surface Finder is provided freely to the academic community upon request.

Electrostatic potential maps

Electrostatic potential maps are calculated using the Poisson–Boltzmann equation solver within MOE.2. The protein is centered on a 33×33×33 cubic grid. A solvent dielectric constant of 80 and a protein dielectric constant of 4 are used in all electrostatic potential map calculations, which are standard values in electrostatic potential map calculations (Sharp and Honig, 1990). Protein partial charges are taken from the CHARMM parameter set (Brooks *et al.*, 1983). The temperature is 300 K, the counter ion radius is 1.4 Å, the ionic strength is 0.15 M, and the protein concentration is 0.001 M. Electrostatic potentials are rendered in blue and red at ±1.5 kcal/mol/e, respectively.

Results and discussion

In the structural comparisons reported here, we comprehensively compare the surface ion pair composition of several hundred protein structures. The Protein Data Bank (PDB) is searched for all protein families in which at least one mesophilic and one thermophilic structure are present, providing 467 protein structures, representing 127 unique orthologous mesophilic–thermophilic groups. After identifying and classifying the protein structures, we calculate the solvent accessibility of all charged residues for each monomer using our implementation of a Lee and Richards-like algorithm (Lee and Richards, 1971). Next, the distance between all possible charged residue pairs is calculated per monomer. The numbers of acid–acid, base–base and acid–base ion pair distances are binned in 0.5 Å increments between 0.0 and 8.0 Å. In PDB structures with more than one monomer (due to quaternary structure and/or crystallographic lattice packing), the values reported are the average over all constituent monomer chains. Finally, the values for all 178 thermophilic and 289 mesophilic PDB structures are averaged and compared (Figure 1).

As one might expect based on the raw amino acid composition data discussed above, thermophilic protein structures have increased numbers of acid–acid, acid–base and base–base ion pairs. Both mesophilic and thermophilic structures generally have far more stabilizing acid–base pairs than either like-charge ion pair. In fact, the number of stabilizing acid–base pairs is generally more than the sum of the two destabilizing pair types. However, the drastic increase in thermophilic acid–base pairs (compared with the small acid–acid and base–base increase) indicates that their location optimizes their stabilizing effect. These trends are conserved after normalizing the raw ion pair occurrences by the number of charged residues within each protein structure.

The maximum number of acid–base pairs for both mesophilic and thermophilic structures is centered at 4.5 Å, which is routinely considered the upper limit for salt bridges. On the other hand, the number of acid–acid and base–base pairs increases roughly linearly with distance. Cursory examination

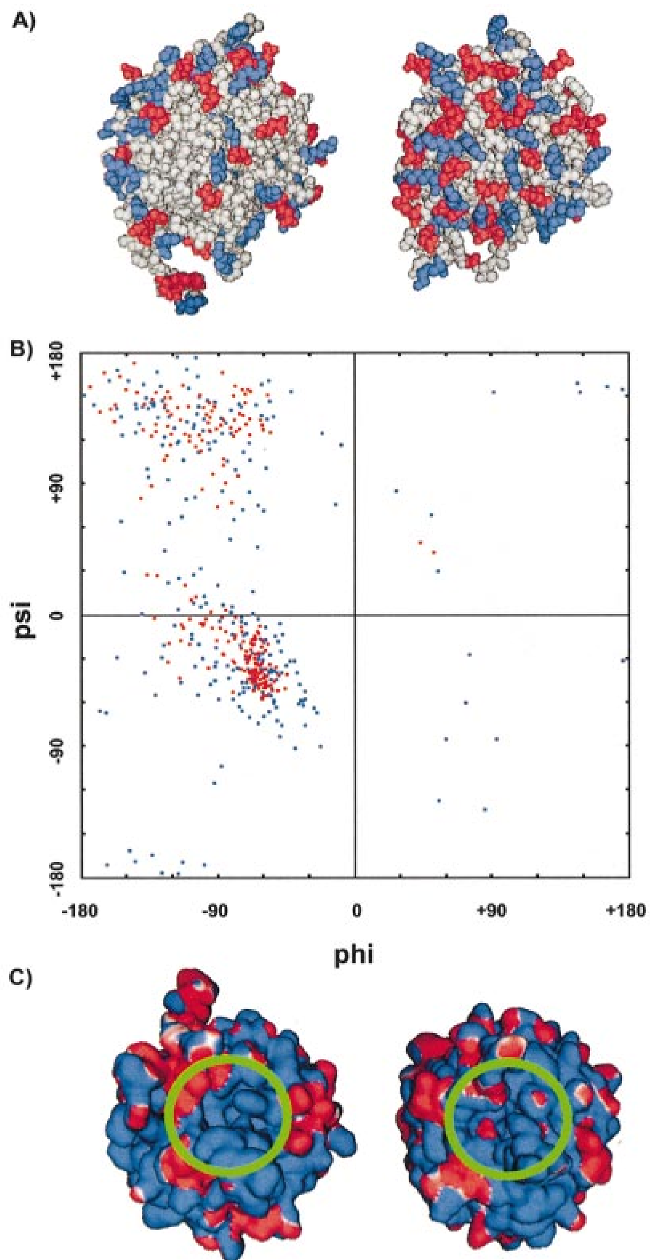

Fig. 2. (**A**) Comparison of a mesophilic (left) and thermophilic (right) 2-dehydro-3-phosphooctanate aldolase highlights how increased numbers of evenly distributed charged residues lead to energetically optimized electrostatic surfaces. There are 21 stabilizing surface acid–base ion pairs in the thermophilic structure (PDB i.d.: 1JCX), whereas there are only seven in the mesophilic structure (PDB i.d.: 1G7V). Both structures are shown in the same orientation. Acidic residues are colored red and basic residues are colored blue. (**B**) Ramachandran analysis of the mesophilic (blue) and thermophilic (red) 2-dehydro-3-phosphooctanate aldolase orthologs. The thermophilic structure has far fewer phi/psi pairs outside the generally accepted areas. (**C**) Electrostatic potential maps of the mesophilic (left) and thermophilic (right) 2-dehydro-3-phosphooctanate aldolase orthologs. Despite key differences in the charge distribution on the proteins' surface, electrostatic potential maps of active site regions (highlighted in green) are very similar. The two structures are in the same orientation, which is roughly the opposite orientation as in (A).

of the structures reveals that acid–acid and base–base pairs at larger distances are often shielded by an acid–base–acid or base–acid–base triplet, and thus not very destabilizing.

These points are exemplified by comparisons of the 2-dehydro-3-phosphooctanate from *Escherichia coli* and the thermophilic bacterium *Aquifex aeolicus* (Asojo *et al*., 2001; Wang *et al*., 2001). The enzyme is a member of the ubiquitous TIM-barrel fold family, with the active site at the 'top' of the structure (Nagano *et al*., 2002). Despite vastly different optimal growth temperatures of the two organisms (37 versus 80°C), traditional comparisons of the two orthologous structures reveal no obvious stabilizing mechanism. The structures possess nearly identical secondary and tertiary structures, and similar numbers of intramolecular contacts. The thermophilic ortholog has 46 (non-secondary structure) hydrogen bonds and 16 salt bridges, whereas the mesophilic ortholog has 41 and 18, respectively. Only Ramachandran analysis provides any appreciable differences between the two structures' stability (Figure 2B).

On the other hand, optimization of the thermophilic enzyme's electrostatic surface is clearly significant to its overall thermostability. Cursory comparison of the two structures reveals significantly more charged residues within the thermophilic protein (Figure 2A). Additionally, the distribution of opposite charges is more interspersed, maximizing their stabilizing effect. Quantification of the differences reveals 21 stabilizing acid–base ion pairs on the surface of the thermophilic ortholog within 8 Å, whereas there are only seven within the mesophilic ortholog. Despite these local charge differences, the overall electrostatic potential maps are qualitatively similar, especially in the active site region (Figure 2C).

Conclusions

The results reported here confirm that optimization of the electrostatic surface is a robust evolutionary mechanism for increasing thermostability, which had been previously suggested based on anecdotal evidence. Comparison of 127 orthologous mesophilic–thermophilic protein groups clearly indicates a preference for stabilizing acid–base pairs on the surface of thermophilic proteins. Compared with positions buried in the core, stabilizing surface mutations are less likely to disrupt the tertiary structure, and thus more likely to be evolutionarily selected. Therefore, we believe that these results, in addition to being theoretically interesting, will facilitate identification of charge-altering mutations likely to increase the stability of particular protein structures.

Supplementary data

A complete list of PDB i.d.s, their thermal classification, and raw data indicating the number of acid–acid, acid–base and base–base ion pairs identified are available as supplementary data at *Protein Engineering* online.

Acknowledgements

This work was supported by an NIH Score grant (S06 GM53933) and an American Chemical Society, Petroleum Research Fund Type G grant (36848-GB4) to D.R.L.

References

Asojo,O., Friedman,J., Adir,N., Belakhov,V., Shoham,Y. and Baasov,T. (2001) *Biochemistry*, **40**, 6326–6334.

Baldwin,R.L. and Rose,G.D. (1999a) *Trends Biochem. Sci.*, **24**, 26–33.

Baldwin,R.L. and Rose,G.D. (1999b) *Trends Biochem. Sci.*, **24**, 77–83.

Brooks,R.B., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) *J. Comput. Chem.*, **4**, 187–217.

Bryson,J.W., Betz,S.F., Lu,H.S., Suich,D.J., Zhou,H.X., O'Neil,K.T. and DeGrado,W.F. (1995) *Science*, **270**, 935–941.

Chakravarty,S. and Varadarajan,R. (2002) *Biochemistry*, **41**, 8152–8161.

Chang,C., Park,B.C., Lee,D.S. and Suh,S.W. (1999) *J. Mol. Biol.*, **288**, 623–634.

Cordes,M.H., Davidson,A.R. and Sauer,R.T. (1996) *Curr. Opin. Struct. Biol.*, **6**, 3–10.
Declerck,N., Machius,M., Wiegand,G., Huber,R. and Gaillardin,C. (2000) *J. Mol. Biol.*, **301**, 1041–1057.
DeDecker,B.S., O'Brien,R., Fleming,P.J., Geiger,J.H., Jackson,S.P. and Sigler,P.B. (1996) *J. Mol. Biol.*, **264**, 1072–1084.
Fukuchi,S. and Nishikawa,K. (2001) *J. Mol. Biol.*, **309**, 835–843.
Grimsley,G.R., Shaw,K.L., Fee,L.R., Alston,R.W., Huyghues-Despointes,B.M., Thurlkill,R.L., Scholtz,J.M. and Pace,C.N. (1999) *Protein Sci.*, **8**, 1843–1849.
Kumar,S., Tsai,C.J. and Nussinov,R. (2000) *Protein Eng.*, **13**, 179–191.
La,D., Silver,M.A., Edgar,R. and Livesay,D.R. (2003) *Biochemistry*, **42**, 3464–3473.
Lee,B. and Richards,F.M. (1971) *J. Mol. Biol.*, **55**, 379–400.
Lehmann,M. and Wyss,M. (2001) *Curr. Opin. Biotechnol.*, **12**, 371–375.
Lehmann,M., Pasamontes,L., Lassen,S.F. and Wyss,M. (2000) *Biochim. Biophys. Acta*, **1543**, 408–415.
Loladze,V.V. and Makhatadze,G.I. (2002) *Protein Sci.*, **11**, 174–177.
Loladze,V.V., Ibarra-Molero,B., Sanchez-Ruiz,J.M. and Makhatadze,G.I. (1999) *Biochemistry*, **38**, 16419–16423.
Martin,A., Sieber,V. and Schmid,F.X. (2001) *J. Mol. Biol.*, **309**, 717–726.
Matthews,B.W. (1995) *Adv. Protein Chem.*, **46**, 249–278.
Nagano,N., Orengo,C.A. and Thornton,J.M. (2002) *J. Mol. Biol.*, **321**, 741–765.
Pace,C.N. (1995) *Methods Enzymol.*, **259**, 538–554.
Pedone,E., Saviano,M., Rossi,M. and Bartolucci,S. (2001) *Protein Eng.*, **14**, 255–260.
Perl,D. and Schmid,F.X. (2001) *J. Mol. Biol.*, **313**, 343–357.
Perl,D., Mueller,U., Heinemann,U. and Schmid,F.X. (2000) *Nat. Struct. Biol.*, **7**, 380–383.
Perry,K.M., Onuffer,J.J., Gittelman,M.S., Barmat,L. and Matthews,C.R. (1989) *Biochemistry*, **28**, 7961–7968.
Schafer,T., Bonisch,H., Kardinahl,S., Schmidt,C. and Schafer,G. (1996) *Biol. Chem.*, **377**, 505–512.
Scholtz,J.M. and Baldwin,R.L. (1992) *Annu. Rev. Biophys. Biomol. Struct.*, **21**, 95–118.
Serrano,L., Kellis,J.T.,Jr, Cann,P., Matouschek,A. and Fersht,A.R. (1992) *J. Mol. Biol.*, **224**, 783–804.
Sharp,K.A. and Honig,B. (1990) *Annu. Rev. Biophys. Biophys. Chem.*, **19**, 301–332.
Spector,S., Wang,M., Carp,S.A., Robblee,J., Hendsch,Z.S., Fairman,R., Tidor,B. and Raleigh,D.P. (2000) *Biochemistry*, **39**, 872–879.
Strop,P. and Mayo,S.L. (2000) *Biochemistry*, **39**, 1251–1255.
Szilagyi,A. and Zavodszky,P. (2000) *Struct. Fold Des.*, **8**, 493–504.
Tanner,J.J., Hecht,R.M. and Krause,K.L. (1996) *Biochemistry*, **35**, 2597–2609.
Thompson,M.J. and Eisenberg,D. (1999) *J. Mol. Biol.*, **290**, 595–604.
Tomschy,A., Bohm,G. and Jaenicke,R. (1994) *Protein Eng.*, **7**, 1471–1478.
Torrez,M., Schultehenrich,M. and Livesay,D.R. (2003) *Biophys. J.*, **85**, 2845–2853.
Vieille,C. and Zeikus,G.J. (2001) *Microbiol. Mol. Biol. Rev.*, **65**, 1–43.
Wang,J., Duewel,H.S., Woodard,R.W. and Gatti,D.L. (2001) *Biochemistry*, **40**, 15676–15683.

Received July 23, 2003; revised October 14, 2003; accepted October 21, 2003