Article No. jmbi.1998.2397 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (1999) 285, 1811–1830

JMB

Molecular Dynamics Simulations of the Hyperthermophilic Protein Sac7d from *Sulfolobus acidocaldarius*: Contribution of Salt Bridges to Thermostability

Paul I. W. de Bakker[1], Philippe H. Hünenberger[2]* and J. Andrew McCammon[2,3]

[1]*Department of Medicinal Chemistry, Utrecht Institute for Pharmaceutical Sciences Utrecht University P.O. Box 80082, 3508 TB Utrecht, The Netherlands*

[2]*Department of Chemistry and Biochemistry and* [3]*Department of Pharmacology, University of California at San Diego 9500 Gilman Drive, La Jolla CA 92093-0365, USA*

**Corresponding author*

Hyperthermophilic proteins often possess an increased number of surface salt bridges compared with their mesophilic homologues. However, salt bridges are generally thought to be of minor importance in protein stability at room temperature. In an effort to understand why this may no longer be true at elevated temperatures, we performed molecular dynamics simulations of the hyperthermophilic protein Sac7d at 300 K, 360 K, and 550 K. The three trajectories are stable on the nanosecond timescale, as evidenced by the analysis of several time-resolved properties. The simulations at 300 K and (to a lesser extent) 360 K are also compatible with nuclear Overhauser effect-derived distances. Raising the temperature from 300 K to 360 K results in a less favourable protein-solvent interaction energy, and a more favourable intraprotein interaction energy. Both effects are almost exclusively electrostatic in nature and dominated by contributions due to charged side-chains. The reduced solvation is due to a loss of spatial and orientational structure of water around charged side-chains, which is a consequence of the increased thermal motion in the solvent. The favourable change in the intraprotein Coulombic interaction energy is essentially due to the tightening of salt bridges. Assuming that charged side-chains are on average more distant from one another in the unfolded state than in the folded state, it follows that salt bridges may contribute to protein stability at elevated temperatures because (i) the solvation free energy of charged side-chains is more adversely affected in the unfolded state than in the folded state by an increase in temperature, and (ii) due to the tightening of salt bridges, unfolding implies a larger unfavourable increase in the intraprotein Coulombic energy at higher temperature. Possible causes for the unexpected stability of the protein at 550 K are also discussed.

Keywords: Sac7d; protein stability; salt bridges; molecular dynamics; electrostatics

Introduction

The phenomenon of protein folding and the thermodynamic basis for the stability of proteins have attracted much interest over the years (Dill, 1993; Jaenicke, 1996; Onuchic *et al.*, 1997; Dobson *et al.*, 1998). Recently, this area of research has been extended to proteins of extremophilic microorganisms, such as thermophiles, halophiles and psychrophiles (for a recent overview, see Madigan & Marrs, 1997), since they are believed to provide new insights into the interactions that determine the structure and function of proteins (Jaenicke, 1981, 1991; Graziano *et al.*, 1995). Hyperthermophilic organisms thrive at temperatures above 80 °C for optimal growth and function, whilst mesophilic organisms have optimal temperatures near room

Abbreviations used: NOE, nuclear Overhauser effect; NMR, nuclear magnetic resonance; P_3M, particle-particle particle-mesh; SPC, simple point charge; rms, root-mean-square.

E-mail address of the corresponding author: phunenbe@ucsd.edu

0022-2836/99/041811–20 $30.00/0

temperature (Stetter *et al.*, 1990). High temperatures that would rapidly lead to the thermal denaturation of mesophilic proteins are sustained by homologous proteins of hyperthermophilic organisms (Rees & Adams, 1995; Danson *et al.*, 1996). For both mesophilic and hyperthermophilic proteins, it remains a difficult task to delineate the thermodynamic determinants of protein stability, as evidenced by the vast amount of literature discussing driving forces and their respective contributions to stability (Dill, 1990; Israelachvili, 1992; Matthews, 1993; Honig & Yang, 1995; Lazaridis *et al.*, 1995; Makhatadze & Privalov, 1995).

Recent comparisons between the structures of mesophilic and hyperthermophilic proteins have identified a number of structural features that are believed to give rise to increased thermal stability (Yip *et al.*, 1995; Hennig *et al.*, 1995; Goldman, 1995; Tanner *et al.*, 1996; Vieille & Zeikus, 1996; Wallon *et al.*, 1997). An apparently significant feature is the high number of salt bridges on the surface of thermophilic proteins (Creighton, 1993). A statistical analysis of a number of thermostable proteins has also shown that thermal stability is correlated with the relative fraction of polar solvent-accessible surface area and the number of hydrogen bonds (Vogt *et al.*, 1997). Other structural features include an increased number of proline residues (Watanabe *et al.*, 1991, 1997), improved hydrophobic packing (Rees & Adams, 1995; Spassov *et al.*, 1995), the presence of disulphide bridges and metal cofactors (Frömmel & Sander, 1989), shorter loop lengths (Vieille & Zeikus, 1996), and the capping of helices (Hennig *et al.*, 1995). It is not always straightforward to rationalise why these structural features correlate with increased thermodynamic stability at elevated temperatures. Theoretical studies, such as computer simulations, may facilitate the elucidation of the mechanisms underlying thermostability. Here, we have performed molecular dynamics simulations of a hyperthermophilic protein at different temperatures to investigate the importance of salt bridges for thermal stability.

Paradoxically, salt bridges (ion pairs) are now generally considered to play a minor role in the stability of proteins (Dill, 1990; Šali *et al.*, 1991; Matthews, 1993). It has even been suggested that salt bridges electrostatically destabilise the native state of proteins, because the large unfavourable desolvation contribution to the free energy of folding associated with bringing oppositely charged side-chains together is not sufficiently compensated by the favourable change in the Coulombic energy (Hendsch & Tidor, 1994; Wimley *et al.*, 1996). However, the desolvation penalty decreases in magnitude with increasing temperature, partly due to a decrease in the dielectric constant of water (Elcock & McCammon, 1997). In other words, at higher temperatures, water solvates charged groups less efficiently due to the increased thermal motion, and ion-pairing is expected to become more favourable. As a consequence, the observation that salt bridges do not markedly stabilise the folded state of mesophilic proteins at room temperature, may not be applicable to hyperthermophilic proteins at elevated temperatures. Continuum electrostatic calculations suggest that salt bridges do indeed contribute to thermostability (Elcock, 1998).

Sac7d (Figure 1) is a small protein from the thermoacidophilic archaeon *Sulfolobus acidocaldarius*, the structure of which was determined recently by NMR spectroscopy (Edmondson *et al.*, 1995). The Sac7 proteins (7 kDa) form a family of five homologous DNA-binding proteins of which Sac7d is the most abundant form (McAfee *et al.*, 1996). These proteins are believed to play a role similar to that of histones in eukaryotes (Reddy & Suryanarayana, 1988; Robinson *et al.*, 1998). Sac7d is a 66 residue protein comprising 18 basic residues (14 Lys and four Arg residues) and 12 acidic residues (five Asp and seven Glu residues). The sequence contains no histidine residues or disulphide bridges, and Sac7d is devoid of metal cofactors. The protein has a compact fold with a high (82 %) secondary structure content (Edmondson *et al.*, 1995). The recombinant form of Sac7d differs from the native form in two ways (McAfee *et al.*, 1995). First, the recombinant protein has an initiating Met residue. Second, residues Lys5 and Lys7 (numbering of recombinant Sac7d) may be monomethylated at the N^{ε} atom in the native form. Although spectroscopic studies indicate little

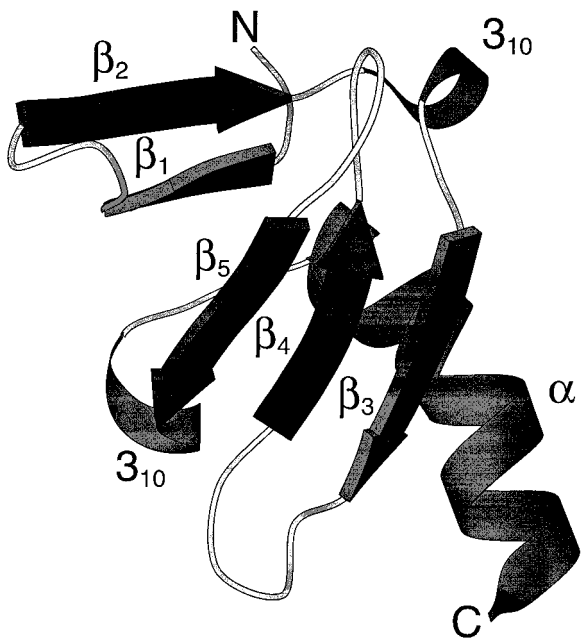

Figure 1. Structure of Sac7d. The secondary structure assignment is as described by Edmondson *et al.* (1995): β-strand 1, residues 4-7; β-turn, 8-11; β-strand 2, 12-15; 3_{10}-helix turn, 16-19; β-strand 3, 21-25; β-turn, 26-29; β-strand 4, 30-34; β-strand 5, 41-46; 3_{10}-helix turn, 47-49; and α-helix, 52-66. N and C indicate the amino and carboxy termini, respectively. The Figure was drawn using the program MOLSCRIPT (Kraulis, 1991).

difference in structure, a difference in stability was observed between recombinant and native Sac7d. The unfolding midpoint temperature of native Sac7d is approximately 100 °C at pH 7, whereas that of recombinant Sac7d is about 7 °C lower (McAfee *et al.*, 1995). Thus, the methylation of these two Lys residues is implicated in heat tolerance (Maras *et al.*, 1992; Baumann *et al.*, 1994).

The high proportion of charged residues, the small size of the protein, the absence of metal cofactors, and the reversibility of the folding transition (McCrary *et al.*, 1996), make Sac7d an ideal test case for studying the contribution of salt bridges to thermostability by computer simulation. Here, we report three molecular dynamics simulations of recombinant Sac7d at 300 K, 360 K, and 550 K. The temperatures of 300 K and 360 K reflect the different conditions in which mesophilic and hyperthermophilic proteins are functional. Although the third temperature of 550 K is unphysical for aqueous solutions, simulations at such elevated temperatures have been used previously to study unfolding pathways of proteins (Mark & van Gunsteren, 1992; Daggett & Levitt, 1992, 1993; Tirado-Rives & Jorgensen, 1993; Caflisch & Karplus, 1994, 1995; Hünenberger *et al.*, 1995; Li & Daggett, 1998). We show that the trajectories at the two lower temperatures are stable and compatible with available experimental data, and discuss the simulation at the unphysical temperature of 550 K. The stability of the simulations at 300 K and 360 K allows us to study in detail equilibrium energetic and structural properties at two different temperatures, and thus their changes upon raising the temperature. Finally, based on a simple hypothesis about the nature of the unfolded state of the protein, we provide an explanation why salt bridges may be stabilising hyperthermophilic proteins at elevated temperatures whereas they do not appear to stabilise mesophilic proteins at room temperature.

Results

Stability of the trajectories

The radius of gyration, $R_{gyr}(t)$, and C^{α} rms positional deviation from the NMR structure, $d_{rms}(t)$, are displayed in Figure 2 for the three simulations. The three $R_{gyr}(t)$ curves do not differ significantly. The average value of R_{gyr} is in all cases about 1.20 nm, slightly above the value of 1.17 nm for the NMR structure. The curve corresponding to the 550 K simulation fluctuates more, and displays a peak (1.31 nm) at 380 ps, after which R_{gyr} drops to about 1.17 nm for the remainder of the simulation. The $d_{rms}(t)$ curves average to 0.21 nm, 0.25 nm, and 0.29 nm for the simulations at 300 K, 360 K, and 550 K, respectively. Again, the curve corresponding to the 550 K simulation

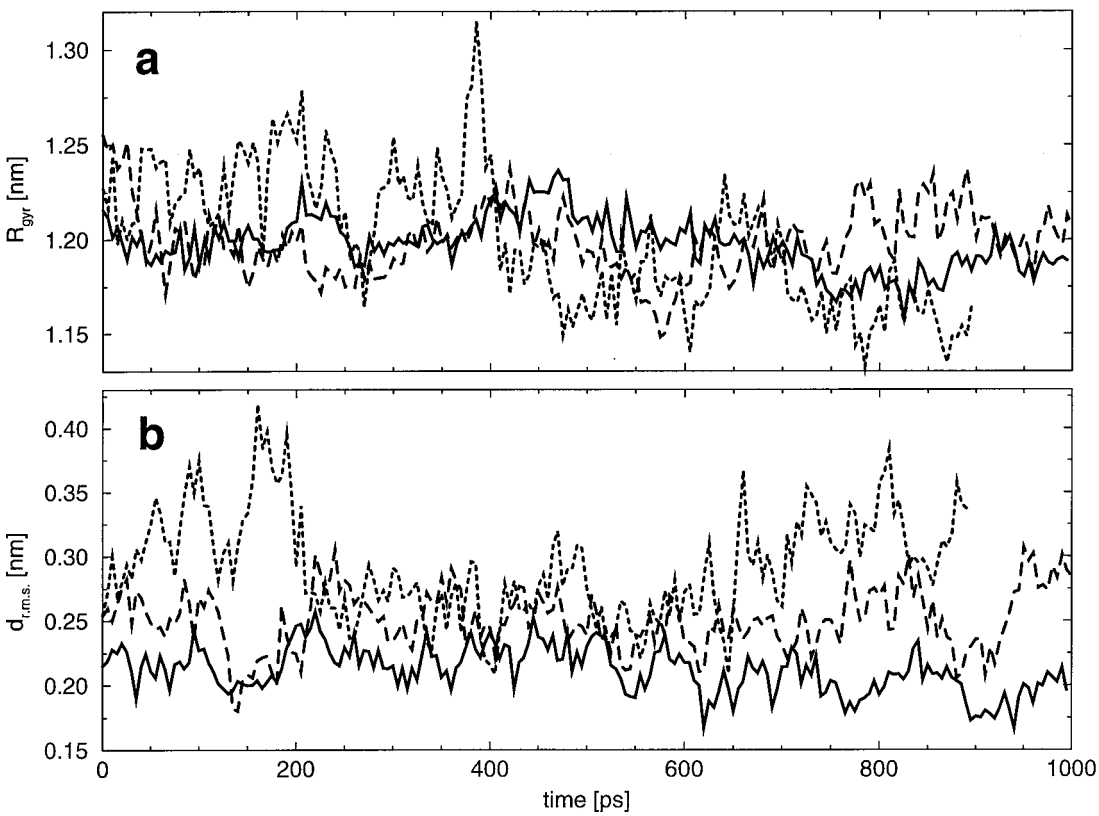

Figure 2. (a) Radius of gyration of the protein as a function of time, according to equation (2). (b) Root-mean-square atomic positional deviation of C^{α} atoms as a function of time, according to equation (3). The different curves correspond to simulations at 300 K (——), 360 K (— — —), and 550 K (- - - - -).

fluctuates more markedly. At this temperature, $d_{rms}(t)$ rises to 0.4 nm within 150 ps. Thereafter, it decreases, remains stable between 200 and 600 ps at about 0.25 nm, and increases again towards a final value of 0.33 nm. Thus, the low value of R_{gyr} observed during the second half of the simulation at 550 K, although closer to the value calculated for the NMR structure, corresponds to structures that differ more from the NMR structure than the less compact structures observed during the simulations at 300 K and 360 K.

Since no experimentally derived distance restraints were included during the simulations, the compatibility of the generated trajectories with nuclear Overhauser effect (NOE)-derived distances provides a good criterion for assessing their reliability. The sum of violations, $S(t)$, is displayed in Figure 3 for the three simulations, including the 150 ps equilibration period. At the three temperatures, $S(t)$ increases steeply during the first part of the equilibration, from 6.9 nm (NMR structure) to values of about 10-13 nm after 50 ps of equilibration. This indicates that the equilibration of the system, and in particular the relaxation of the water molecules, significantly perturbs the structure of the protein. For the 300 K and 360 K simulations, the protein recovers from this perturbation after the equilibration period and $S(t)$ decreases again, down to 9.5 nm at 175 ps. At 300 K, the sum of violations remains close to this value during the remainder of the simulation. At 360 K, it increases again, up to 13.2 nm at the end of the simulation. Although the sum of violations is slightly higher during the 300 K simulation compared to the NMR structure, individual violations are all below 0.43 nm after 300 ps. In comparison, the maximal violation in the NMR structure is 0.39 nm. Taking into account that experimental NOE-derived distances correspond with averages over times much longer than 50 ps, the simulations at 300 K, and to a lesser extent at 360 K, appear to be in good agreement with experimental data. This is no longer the case for the simulation at 550 K, where $S(t)$ quickly increases up to values around 25 nm immediately after the equilibration period.

The evolution of the secondary structure as a function of time for the three simulations is displayed in Figures 4 to 6. There is a high degree of similarity between the graphs corresponding to the simulations at 300 K and 360 K. Only marginal structural fluctuations are observed and no significant structural changes. The secondary structure content of the protein during these simulations is generally higher than that of the NMR structure. At 300 K (Figure 4), we observe a slight shortening of β-strands 4 and 5 at residues 35-36 and 39-40 between 280 and 470 ps, and after 900 ps. Transient shortening of β-strands is also observed for residues 3, 8, 11, 16 and 46. The α-helix appears to be very stable. At 360 K (Figure 5), a similar shortening occurs at residues 35-36 and 39-40 around 500 and 950 ps. In addition, after 330 ps, secondary structure is lost at the turn connecting β-strands 1 and 2. Finally, the C terminus of the α-helix fluctuates slightly more than at 300 K. In constrast, at 550 K (Figure 6), secondary structure fluctuations are significantly more pronounced. Temporary disruption of hydrogen bonds in turns connecting β-strands and in the middle of several β-strands is evident. Portions of the α-helix display transient changes to patterns corresponding to a 3_{10}-helix, a hydrogen-bonded turn, or a β-turn. However, in spite of these important fluctuations in the hydrogen-bonding pattern, it appears that all native secondary structure elements remain present until the end of the 0.9 ns simulation. The persistence of the secondary structure during the

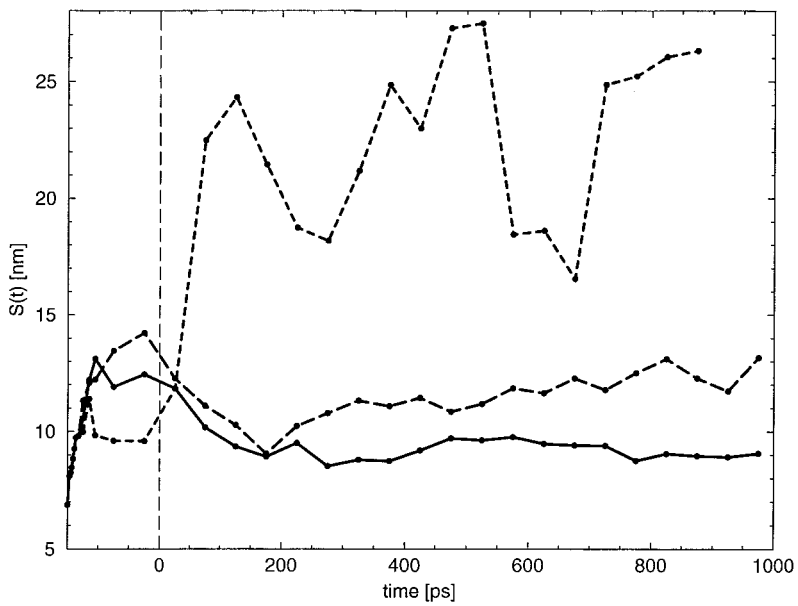

Figure 3. Sum of violations with respect to NOE-derived site-site distances as a function of time, according to equation (3). The equilibration period (−150 ps through 0 ps) is included. The value $S(t) = 6.9$ nm at $t = -150$ ps corresponds to the NMR structure. The different curves correspond to simulations at 300 K (——), 360 K (— — —), and 550 K (- - - -).

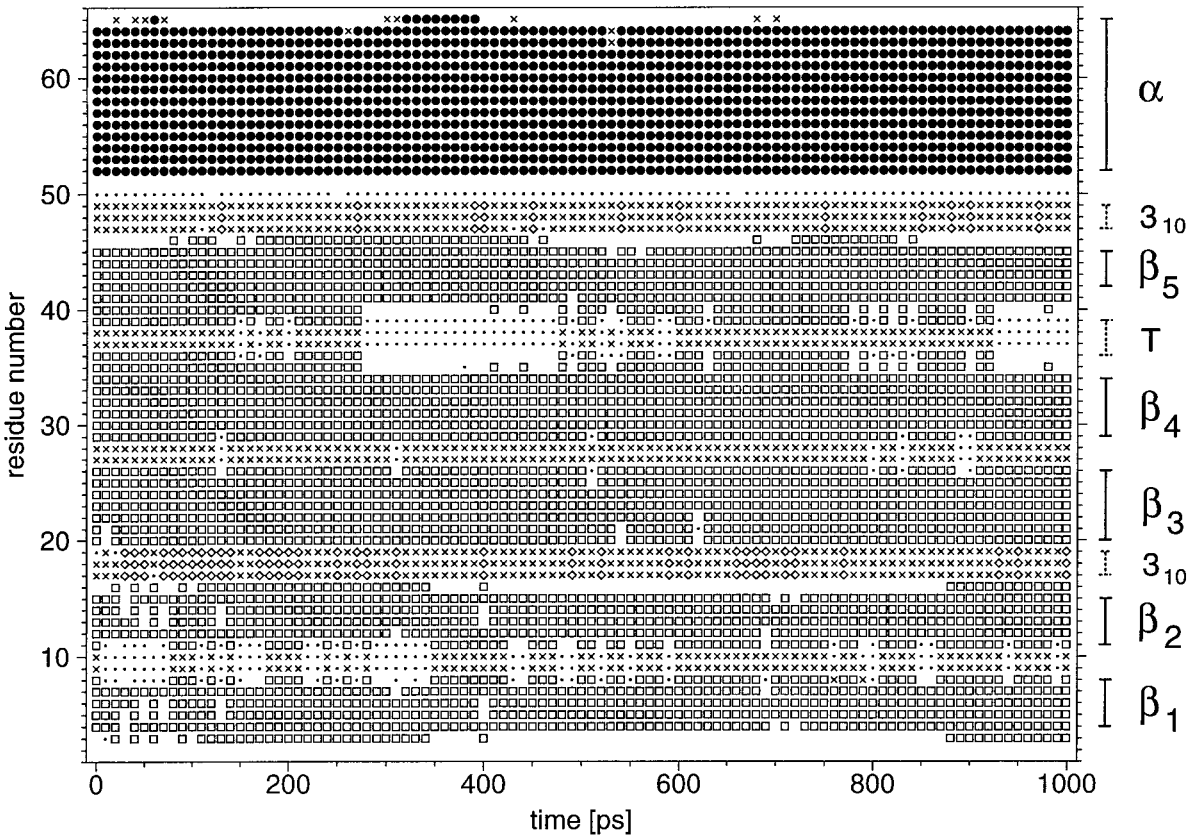

Figure 4. Time-evolution of the secondary structure for the simulation at 300 K assigned according to the DSSP program and criteria (Kabsch & Sander, 1983). The secondary structure elements are: α-helix (●), β-strand (□), 3_{10}-helix (◇), hydrogen-bonded turn (×), and bend (·). The DSSP secondary structure assignment for the NMR structure is displayed at the right-hand side of the graph.

550 K simulation, together with the stability of the $R_{gyr}(t)$ and $d_{rms}(t)$ curves, constrast with reported simulations of mesophilic proteins at similar or lower temperatures and of similar or shorter lengths, where unfolding of the protein was observed (Mark & van Gunsteren, 1992; Daggett & Levitt, 1992, 1993; Tirado-Rives & Jorgensen, 1993; Caflisch & Karplus, 1994, 1995; Hünenberger *et al.*, 1995; Li & Daggett, 1998). Possible causes for this unexpected stability are discussed below.

The above analysis demonstrates that the trajectories at 300 K and 360 K are essentially stable with respect to all monitored properties. The simulation at 300 K also appears to agree well with experimental NOE-derived distances. The remainder of this section will be devoted to the analysis of equilibrium energetic and structural properties derived from these two stable simulations, with the purpose of identifying the role of salt bridges in thermal stability.

Energetic analysis

Mean values and standard deviations of the different energetic contributions defined in equations (5) to (7) are reported in Table 1. The total potential energy (E^{pot}) displays a large increase when the temperature is raised from 300 K to 360 K. Here, we analyse the different contributions to this increase. The covalent term (E^{cov}) becomes more unfavourable at higher temperature due to increased average deviations of bond angles, dihedral angles and improper dihedral angles from their equilibrium values. The electrostatic energy term (E^{elec}) greatly increases with temperature. This increase is primarily due to the solvent-solvent term (E^{elec}_{ss}), and is the consequence of the loss of dipole-dipole correlations and hydrogen bonds in the solvent caused by increased thermal motion. The protein-solvent electrostatic interaction energy (E^{elec}_{ps}) also increases by 1110 kJ/mol for a similar reason. Namely, water is electrostatically less solvating at 360 K compared with 300 K, because the orientation of solvent dipoles solvating charged and polar groups of the protein is more efficiently randomised by thermal motion at higher temperature. Partly compensating for this loss of solvation, the protein-protein electrostatic interaction energy ($E^{elec}_{pp} + E^{elec}_{pr}$) becomes more favourable by 395 kJ/mol. The largest contribution to this change is due to a decrease in the intraprotein electrostatic interaction energy (E^{elec}_{pp}) by 373 kJ/mol, whereas a further decrease in the interaction energy of the protein with the ensemble of its periodic images (E^{elec}_{pr}) accounts for the

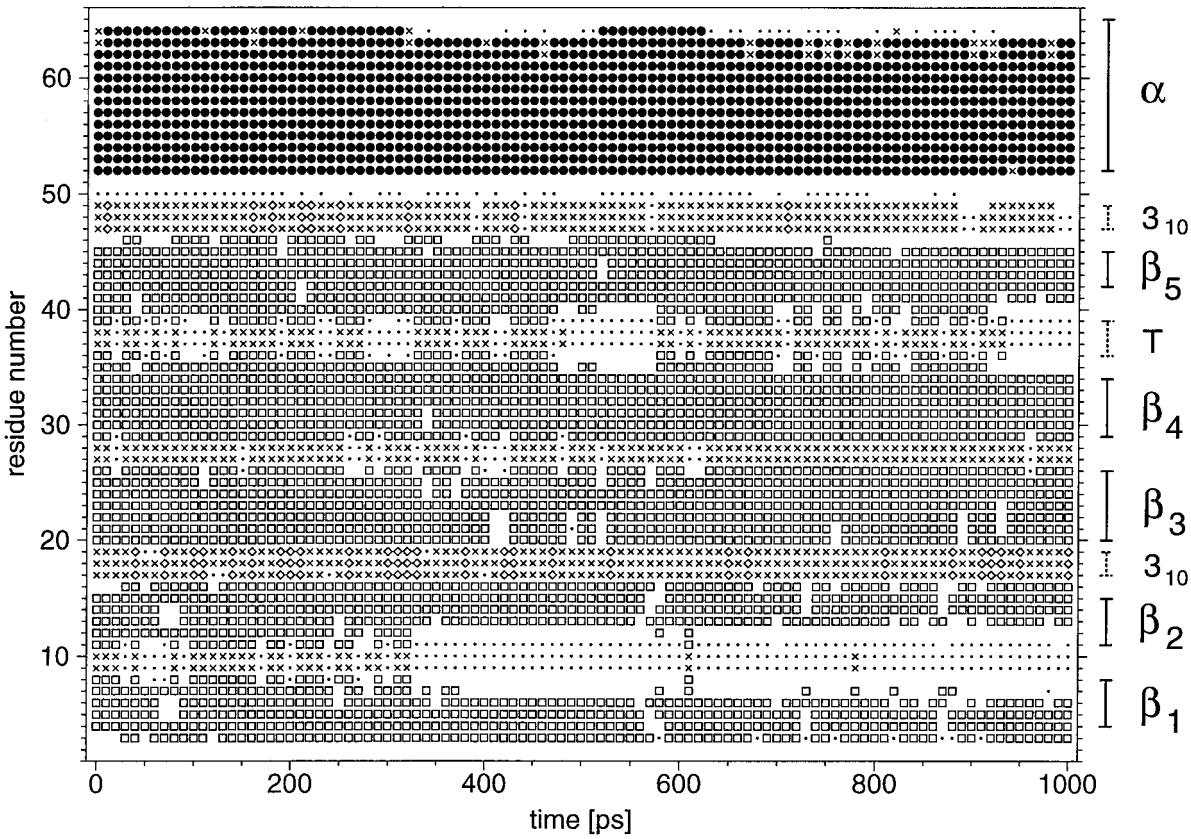

Figure 5. Time-evolution of the secondary structure for the simulation at 360 K. See the legend to Figure 4 for further details.

remainder. The decrease in the E_{pp}^{elec} term is in turn dominated by interactions among charged side-chains (E_{cc}^{elec}), which become more favourable by 346 kJ/mol. Interactions between charged side-chains and other protein atoms (E_{cn}^{elec}) contributes to a further decrease by 66 kJ/mol, whereas the contribution due to all pairs excluding atoms of charged side-chains is positive and small. Finally, increasing the temperature results in a decrease in the van der Waals energy (E^{vdw}) by 2803 kJ/mol, mostly due to more favourable solvent-solvent interactions (E_{ss}^{vdw}). This decrease probably occurs because the weakening of the mean attractive electrostatic interactions among solvent molecules causes them to spend less time in the repulsive region and more time in the low-energy well of the Lennard-Jones interaction potential. Changes in the protein-solvent (E_{ps}^{vdw}) and intraprotein (E_{pp}^{vdw}) van der Waals energies are positive and of small magnitude.

The picture that emerges from this analysis is that the protein is less solvated at higher temperature. This reduction of the solvation is almost exclusively electrostatic in nature and affects primarily the charged residues at the protein surface. As a consequence, charged residues tend to rearrange so as to improve their direct electrostatic interactions among each other and, to a lesser extent, with other polar groups of the protein.

Table 1. Contributions to the total potential energy during the simulations at 300 K and 360 K

	300 K		360 K		Δ
	⟨···⟩	σ	⟨···⟩	σ	⟨···⟩
T	298	±2	358	±2	60
E^{pot}	−176,404	±299	−164,974	±325	11,431
E^{cov}	1884	±61	2174	±69	290
E^{elec}	−205,147	±583	−191,203	±600	13,944
E_{ss}^{elec}	−187,038	±664	−173,809	±702	13,229
E_{ps}^{elec}	−12,972	±560	−11,862	±686	1110
E_{pr}^{elec}	−1036	±41	−1058	±38	−22
E_{pp}^{elec}	−4100	±249	−4474	±326	−373
E_{cc}^{elec}	−2905	±195	−3250	±333	−346
E_{cn}^{elec}	−570	±128	−636	±136	−66
E_{nn}^{elec}	−626	±47	−588	±52	38
E^{vdw}	26,859	±432	24,055	±444	−2803
E_{ss}^{vdw}	28,938	±426	26,084	±437	−2854
E_{ps}^{vdw}	−270	±79	−263	±83	7
E_{pp}^{vdw}	−1810	±51	−1766	±54	44

Mean values (⟨···⟩) and standard deviations (σ) of the energetic contributions defined in equations (5) to (7) for the trajectories at 300 K and 360 K (units, kJ/mol). The actual temperature T during the simulations is also reported (units, K). The difference between the mean values at 300 K and 360 K is given in the last column (Δ). The superscripts pot, cov, elec and vdw refer to potential, covalent, electrostatic and van der Waals energy terms, respectively. The subscripts ss, ps, pr, pp, cc, cn and nn denote interactions between atoms of the following groups: solvent (s), protein in the central unit cell (p), periodic replicas of the protein (r), charged side-chains (c), protein atoms except charged side-chains (n). The indentation underlines how energy terms are partitioned.

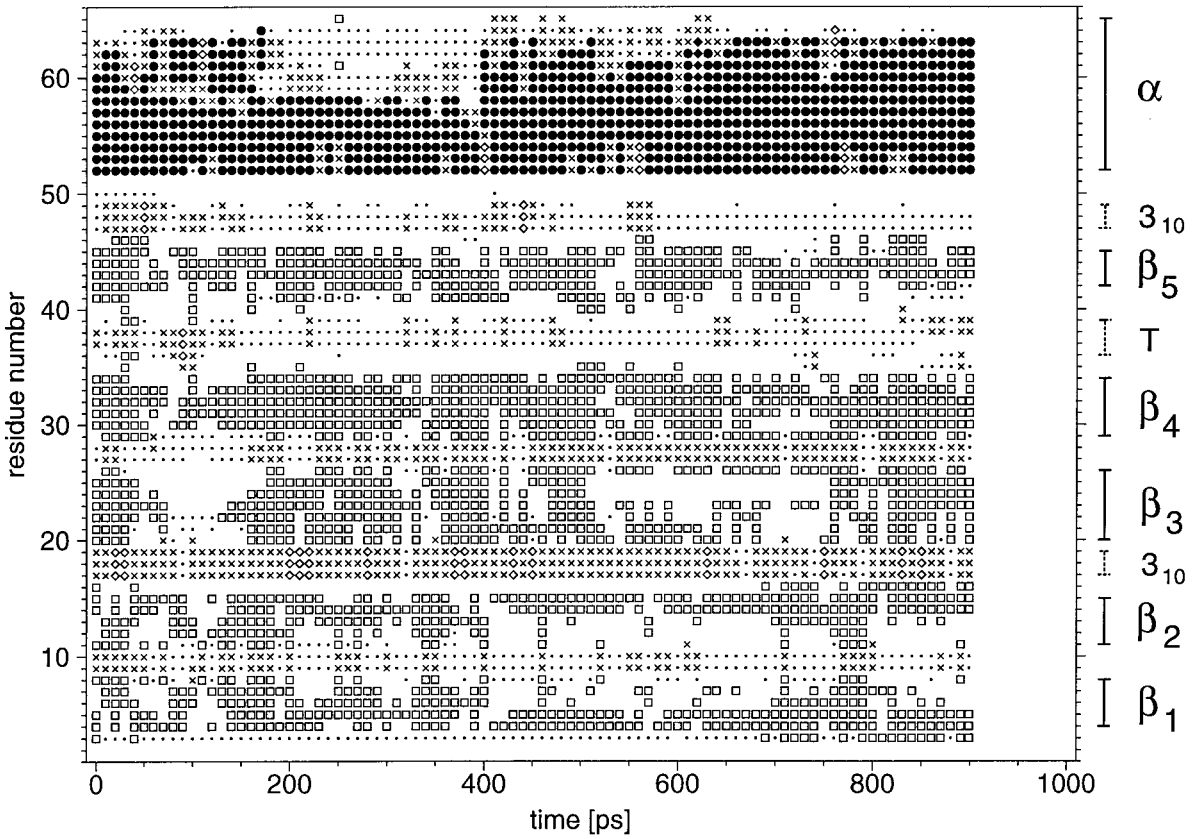

Figure 6. Time-evolution of the secondary structure for the simulation at 550 K. See the legend to Figure 4 for further details.

Next, we examine in greater detail the solvation of surface charges and the nature of their rearrangement upon raising the temperature.

Solvation of charged side-chains

The radial distribution function, $g_{ID}(r)$, describes the local density of the solvent at distance r from charge I relative to the bolk density. The $g_{ID}(r)$ functions were evaluated for all charged side-chains and the chain termini. The results, averaged over positively or negatively charged residues at both 300 K and 360 K are displayed in Figure 7(a). The curves approach unity for large distances, because the local density converges to its bulk value, and are quasi-linear beyond about 1 nm. The non-zero slope is due to the excluded volume of the protein. The first peak is more pronounced and occurs at a shorter distance for positively charged residues compared to negatively charged residues, because the water oxygen was chosen as the centre of the solvent dipole. The $g_{ID}(r)$ curve is also generally higher for positively charged residues than for negatively charged residues, because the longer side-chains of Lys and Arg residues enable them to extend their ammonium and guanidinium groups further into the solvent. When comparing the curves corresponding to 300 K and 360 K, we observe that the peaks are broader, smaller in magnitude and slightly shifted to higher distances at 360 K. These changes in the radial distribution function illustrate the loss of solvent structure at higher temperature caused by thermal motion. For both acidic and basic amino acid residues, solvent shells tend to be spatially less ordered and on average further away from charged side-chains at higher temperature. Recently, a continuum solvation model has been developed to reproduce temperature-dependent hydration free energies of amino acids in the range 5-100 °C (Elcock & McCammon, 1997). It was found that in addition to the decrease in the dielectric permittivity of water with temperature, it was necessary to increase the effective atomic radii with increasing temperature in order to reproduce the experimental data. The shift of the $g_{ID}(r)$ peaks to larger distances upon raising the temperature provides a physical interpretation of this finding.

The radial dipole orientation correlation function, $c_{ID}(r)$, describes to what extent the solvent dipoles at distance r from charge I are radially oriented (one indicating a perfect radial orientation and zero a random orientation). The $c_{ID}(r)$ functions, averaged over positively or negatively charged residues at both 300 K and 360 K are displayed in Figure 7(b). The curves approach zero for large distances. Here again, the first two peaks are sharper and occur at shorter distances for positively charged residues. When the temperature is increased from 300 K to 360 K, the magnitudes of

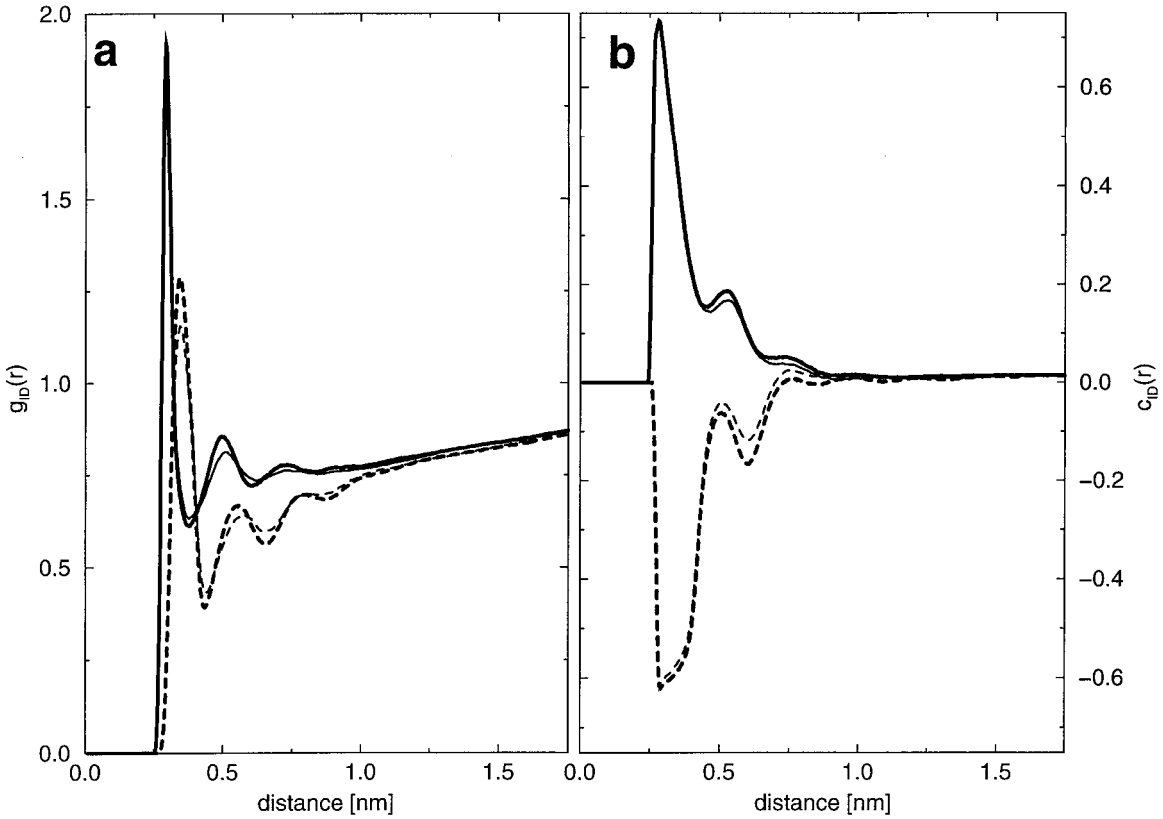

Figure 7. (a) Ionised side-chain-water dipole radial distribution function, according to equation (8). (b) Radial dipole orientation correlation function, according to equation (9). The different curves correspond to values obtained from the trajectories at 300 K (thick lines) or 360 K (thin lines), and averaged over positively charged residues (——), or negatively charged residues (- - - - -).

the second and third peaks decrease. For negatively charged residues, a decrease in magnitude is also visible in the first peak. This decrease in the magnitude of $c_{ID}(r)$ observed at higher temperature is again the consequence of the increased thermal motion.

The effect of temperature on both $g_{ID}(r)$ and $c_{ID}(r)$ will influence the solvation free energy of a given side-chain (see equation (10)). Values of G_I^{solv}, the estimated solvation free energy of side-chain I, calculated for charged residues at 300 K and 360 K are reported in Tables 2 and 3. The sum of G_I^{solv} for all charged residues is −4672 kJ/mol (300 K) and −4287 kJ/mol (360 K). Since G_I^{solv} is related to the side-chain–solvent interaction energy (E_I^{solv}) by $G_I^{solv} = 1/2E_I^{solv}/2$, we can compare these values to $1/2E_{ps}^{elec} = -6486$ kJ/mol (300 K) and −5931 kJ/mol (360 K), respectively (Table 1). This comparison suggests that the interaction between the charged residues and the solvent accounts for roughly 70 % of the total protein-solvent electrostatic interaction energy. For all residues and at both temperatures, G_I^{solv} is negative, the only two exceptions being Glu47 and Glu53 at 360 K. The magnitude of G_I^{solv} for these residues remains small and the positive sign can be explained by the vicinity of a number of positive charges (see Figure 9(c)). The average solvation free energies for the different residue types at 300 K are: Lys (−195 kJ/mol), Arg (−151 kJ/mol), Glu (−97 kJ/mol) and Asp (−59 kJ/mol). The average solvation free energies at 360 K are in the same order. Thus, positively charged residues are better solvated than negatively charged residues. It is, however, difficult to assess whether this observation reflects intrinsic properties of the side-chains, or rather the fact that the protein bears an overall positive charge of +6 *e*. Lys residues are better solvated than Arg residues probably because their charge is more concentrated in space, and Glu residues are better solvated than Asp residues probably because their side-chain is longer. By examining the differences, ΔG_I^{solv}, between G_I^{solv} at 300 K and 360 K, we find that Lys, Asp, and Glu residues are on average less solvated at higher temperature, while the opposite is true for Arg residues. Only nine residues (out of 32) are better solvated at 360 K, among which are three Arg residues. The dominant contribution to the decrease in magnitude of the solvation free energy of charged residues with increasing temperature is due to Glu residues, which are on average less solvated by 29 kJ/mol at 360 K. In particular, residue Glu53 is less solvated by as much as 128 kJ/mol. In summary, charged residues are on average less solvated by 12 kJ/mol at 360 K relative to 300 K.

The individual energetic contributions of charged side-chains, E_I^{elec}, to the total electrostatic

Table 2. Solvation free energy and contribution to the intramolecular Coulombic energy for positively charged residues at 300 K and 360 K

	G_I^{solv}			E_I^{elec}			
Residue	300 K	360 K	ΔG_I^{solv}	300 K	360 K	ΔE_I^{elec}	ΔG_I^{tot}
N-ter	−177	−164	13	29	1	−28	−15
Lys3	−175	−173	2	43	39	−4	−2
Lys5	−107	−94	13	−11	−23	−12	1
Lys7	−180	−135	45	57	4	−53	−8
Lys9	−227	−267	−40	120	150	30	−10
Lys13	−170	−166	4	61	58	−3	1
Lys19	−172	−148	24	54	23	−32	−8
Lys21	−182	−176	6	100	94	−6	0
Lys22	−258	−235	23	176	159	−17	7
Lys28	−258	−251	7	145	125	−21	−14
Lys39	−243	−216	27	137	116	−21	6
Lys48	−219	−214	5	94	88	−7	−2
Lys52	−207	−213	−6	69	81	12	6
Lys65	−190	−164	26	120	110	−10	16
Lys66	−144	−130	14	91	69	−22	−8
Arg25	−124	−165	−41	74	118	44	4
Arg42	−239	−250	−12	178	183	5	−7
Arg60	−95	−111	−16	24	57	34	18
Arg63	−145	−138	7	100	98	−3	4
⟨Lys⟩	−195.2	−184.5	10.7	89.8	78.1	−11.7	−1.0
⟨Arg⟩	−150.8	−166.2	−15.3	94.0	113.9	20.0	4.7
⟨Pos⟩	−184.9	−179.6	5.3	87.5	81.6	−5.9	−0.6

Solvation free energy, G_I^{solv}, according to equation (10), and contribution to the intramolecular Coulombic energy, E_I^{elec}, according to equation (11), for all positively charged residues at 300 K and 360 K (units, kJ/mol). The quantities ΔG_I^{solv} and ΔE_I^{elec} are the differences between G_I^{solv} and E_I^{elec} at the two temperatures, respectively. The quantity ΔG_I^{tot} is the sum of ΔG_I^{solv} and ΔE_I^{elec}. N-ter denotes the amino terminus of the protein. ⟨Lys⟩, ⟨Arg⟩, and ⟨Pos⟩ denote averages over all Lys, Arg, and positively charged residues, respectively.

interaction energy between all charged side-chains, E_{cc}^{elec}, is also reported in Tables 2 and 3. For all positively charged residues and at both temperatures, E_I^{elec} is positive, except for Lys5. For negatively charged residues, E_I^{elec} is always negative. The relative signs and the larger magnitude of E_I^{elec} for negatively charged residues is a consequence of the excess of positively charged residues (19) in the protein compared to negatively charged (13) residues. By examining the ''differences'' (ΔE_I^{elec}, between E_I^{elec} at 300 K and 360 K, we find that all but five positively charged residues and all but five negatively charged residues interact more favourably with other charged residues at the higher temperature. We note the striking negative correlation between the quantities ΔG_I^{solv} and

Table 3. Solvation free energy and contribution to the intramolecular Coulombic energy for negatively charged residues at 300 K and 360 K

	G_I^{solv}			E_I^{elec}			
Residue	300 K	360 K	ΔG_I^{solv}	300 K	360 K	ΔE_I^{elec}	ΔG_I^{tot}
C-ter	−188	−149	40	−296	−332	−36	4
Asp16	−20	−27	−8	−339	−342	−4	−12
Asp35	−42	−37	5	−384	−380	4	8
Asp36	−40	−42	−2	−406	−393	13	11
Asp49	−100	−81	20	−321	−331	−10	10
Asp56	−93	−64	29	−354	−390	−35	−6
Glu11	−141	−160	−19	−305	−284	21	2
Glu12	−101	−76	25	−334	−359	−25	0
Glu14	−62	−66	−4	−380	−365	15	11
Glu47	−11	15	26	−450	−421	29	55
Glu53	−119	9	128	−263	−416	−153	−25
Glu62	−94	−83	11	−394	−406	−12	−2
Glu64	−147	−115	32	−340	−381	−41	−8
⟨Asp⟩	−59.0	−50.3	8.8	−360.8	−367.3	−6.5	2.2
⟨Glu⟩	−96.5	−68.0	28.5	−352.3	−375.9	−23.6	4.9
⟨Neg⟩	−89.1	−67.3	21.8	−351.3	−369.2	−17.9	3.9

Solvation free energy, G_I^{solv}, according to equation (10), and contribution to the intramolecular Coulombic energy, E_I^{elec}, according to equation (11), for all negatively charged residues at 300 K and 360 K (units, kJ/mol). See also the legend to Table 2. C-ter denotes the carboxy terminus of the protein. ⟨Asp⟩, ⟨Glu⟩, and ⟨Neg⟩ denote averages over all Asp, Glu, and negatively charged residues, respectively.

ΔE_I^{elec}. A linear least-squares fit to the graph $\Delta E_I^{\text{elec}} = f(\Delta G_I^{\text{solv}})$ for all residues gives a correlation-coefficient of −0.92, a slope of −1.09 and an intercept of 2.2 kJ/mol. This indicates that the loss in solvation free energy of a given charged residue upon increasing the temperature is almost exactly balanced by a gain in interaction energy with other charged residues within the protein. As a consequence, the sum of ΔG_I^{solv} and ΔE_I^{elec}, ΔG_I^{tot}, is generally smaller in magnitude than both contributions. The average value of ΔG_I^{tot} over all residues is only 1.2 kJ/mol. The above analysis indicates that increased temperature induces only a small increase in the electrostatic free energy of charged side-chains, but a much more significant redistribution of energy among the two contributions to this free energy, the solvation on average decreasing in magnitude and the intramolecular charged residue-charged residue (cc) interaction on average increasing in magnitude with increased temperature. Note that this latter observation is also correct on average for Lys, Glu and Asp residues. However, the opposite effect is observed for Arg residues.

Structural changes in the arrangement of charged side-chains

To investigate the structural origin of the change in $E_{\text{cc}}^{\text{elec}}$ (Table 1) and in E_I^{elec} (Tables 2 and 3) while changing the temperature from 300 K to 360 K, we examined distance distributions for all pairs of charged side-chains in the protein (positive-positive, negative-negative, and positive-negative). Assuming that charged side-chains can be represented as single points, $\langle N_{++} \rangle$, $\langle N_{--} \rangle$, and $\langle N_{+-} \rangle$ are defined as the number of unique pairs at a particular separation distance consisting of two positively, two negatively, or two oppositely charged side-chains, respectively, averaged over the entire trajectory (see Methods). The three distributions $\langle N_{++} \rangle$, $\langle N_{--} \rangle$, and $-\langle N_{+-} \rangle$ are displayed as a function of side-chain–side-chain distance in Figure 8(a) for the simulations at 300 K and 360 K. The sum of the three components, $\langle N \rangle$, is displayed in Figure 8(b), together with the running integral of $r^{-1}\langle N \rangle$. According to equation (12), this integral multiplied by $(4\pi\varepsilon_0)^{-1}e^2$ (138.9 kJ nm/mol) should approximate the value of $E_{\text{cc}}^{\text{elec}}$ at large distances. The approximate shift in $E_{\text{cc}}^{\text{elec}}$ when increasing the temperature from 300 K to 360 K, evaluated from the difference in the integrals, is −397 kJ/mol. This value is in reasonable agreement with the exact value of −346 kJ/mol (Table 1). In the range of distance 0.8-4.5 nm, there is a large compensation between $\langle N_{++} \rangle$, $\langle N_{--} \rangle$ and $-\langle N_{+-} \rangle$ resulting in limited oscillations of $\langle N \rangle$ around 0 nm^{-1}. Between 0.3 and 0.8 nm, however, $\langle N_{++} \rangle$ and $\langle N_{--} \rangle$ are both close to zero and a large negative peak in $\langle N \rangle$ is observed at both temperatures, which is only caused by $-\langle N_{+-} \rangle$ and thus directly related to the formation of salt bridges. As is evident from the running integral of $r^{-1}\langle N \rangle$, these salt bridge interactions form the dominant contribution to $E_{\text{cc}}^{\text{elec}}$. When the temperature is increased from 300 K to 360 K, a shoulder appears at the left of the peak in $\langle N \rangle$, at about 0.35 nm, which suggests a tightening of salt bridges at higher temperature. The weighted integral of $r^{-1}\langle N \rangle$ shows that the change in $E_{\text{cc}}^{\text{elec}}$ upon increasing the temperature is almost exclusively due to additional interactions in this narrow 0.3-0.4 nm range. We conclude from this analysis that (i) the decrease in the intraprotein Coulombic energy is essentially a consequence of the tightening of salt bridges, and (ii) this energetically important effect (about 350 kJ/mol) arises from relatively small changes in the distance distribution within salt bridges. The latter point is illustrated in Figure 9(b) and (c), which represent the average conformations of charged residues over the 300 K and 360 K simulations.

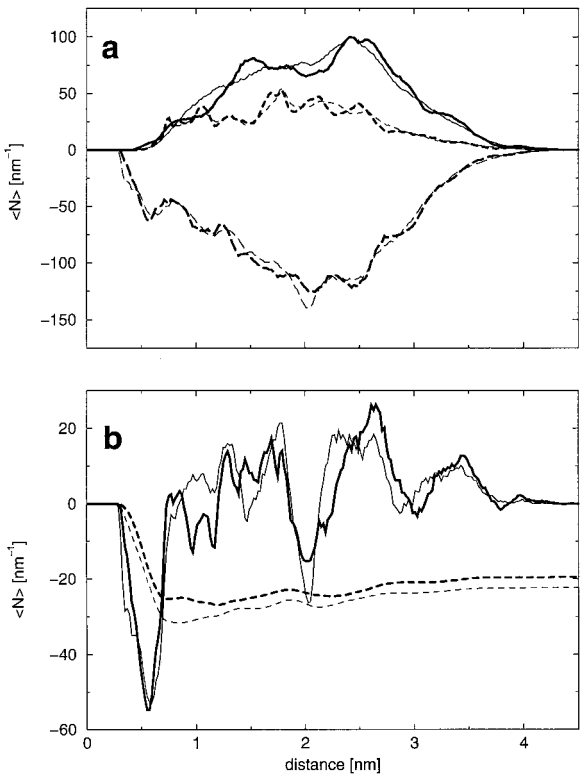

Figure 8. Distance distributions for all charged side-chains, calculated from the simulations at 300 K (thick lines) and 360 K (thin lines). (a) $\langle N_{++} \rangle$ (——), $\langle N_{--} \rangle$ (- - - - -) and $-\langle N_{+-} \rangle$ (— — —). (b) $\langle N \rangle$ (——) and cumulative integral of $r^{-1}\langle N \rangle$ (- - - - -).

Ionic networks

A question arising at this point, and which has important implications for the engineering of thermostable proteins, is whether any salt bridge will stabilise a protein at elevated temperatures, or whether a specific placement of ionisable residues is required to promote thermal stability. The

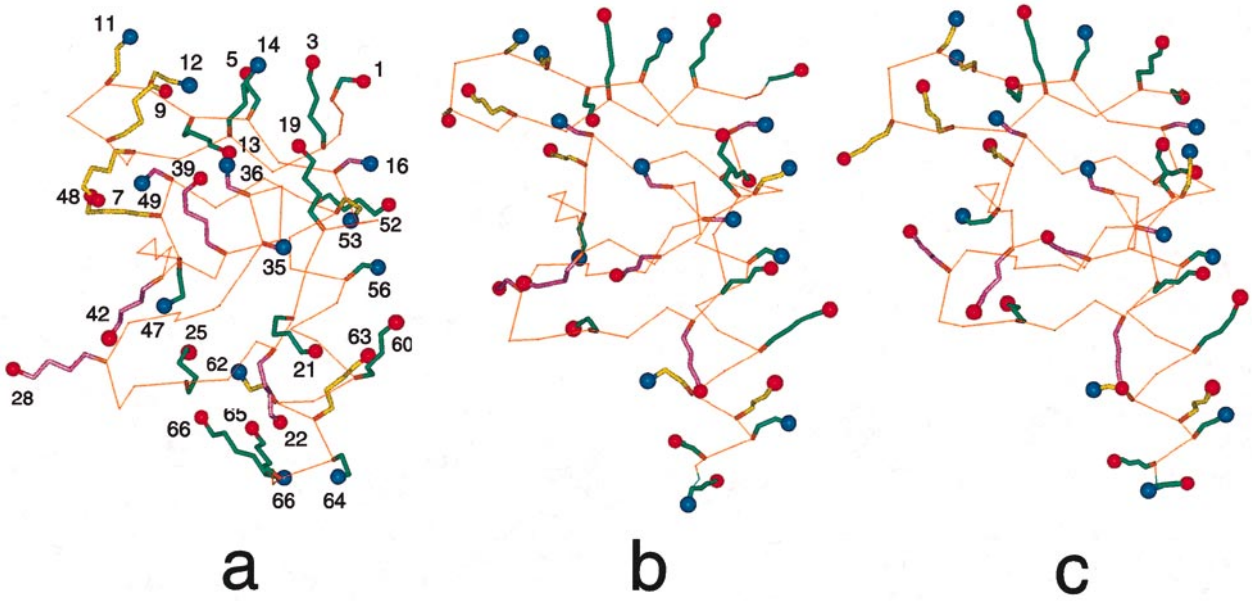

Figure 9. Conformations of charged residues on the surface of Sac7d: (a) co-ordinates from the NMR structure; (b) average co-ordinates over the 300 K simulation; (c) average co-ordinates over the 360 K simulation. The orientation is roughly similar to that of Figure 1. In contrast to the NMR structure, average co-ordinates do not correspond to a physical low-energy conformation. Bond lengths, bond angles and dihedral angles may thus appear distorted. The C^{α} trace is displayed in orange. Positive charges are represented as red spheres at the C^{ζ} (Arg), the N^{ε} (Lys), or the N-terminal ammonium nitrogen, whereas negative charges are represented as blue spheres at the C^{γ} (Asp), the C^{δ} (Lys), or the C-terminal carboxylate carbon. Hydrogen atoms, carboxylate oxygen atoms and guanidinium NH_2 groups are not displayed. Residues are grouped into three clusters so as to maximise the negative sum of intra-cluster contributions to ΔE_{cc}^{elec}, the change in the Coulombic energy of the ionised side-chains upon raising the temperature from 300 K to 360 K. Clusters C_1, C_2, and C_3 are coloured green, yellow, and pink, respectively. The clusters contain the following residues. C_1: N terminus, Lys3, Lys5, Lys13, Glu14, Lys19, Lys21, Arg25, Glu47, Lys52, Asp56, Arg60, Glu64, Lys65, Lys66 and C terminus. 1; C_2: Lys7, Lys9, Glu11, Glu12, Lys48, Glu53, Glu62, Arg63; C_3: Asp16, Lys22, Lys28, Asp35, Asp36, Lys39, Arg42, Asp49.

strength of a salt bridge will generally depend on three factors (Horovitz *et al.*, 1990): (i) the degree of exposure to solvent; (ii) the geometry and distance of interaction; and (iii) the effect of neighbouring charged residues. The first dependence cannot be discussed based on the present simulations, since all ionised residues of Sac7d are exposed to solvent. We merely note that, owing to the larger desolvation penalty, a buried salt bridge is likely to be less stabilising compared with a surface salt bridge. The second dependence is also beyond the scope of the present study. However, it may be significant as evidenced by the influence, presumably through subtle conformational changes, of a remote mutation on the strength of a salt bridge introduced in glutamute dehydrogenase from *Thermococcus litoralis* (Vetirani *et al.*, 1998). The effect of neighbouring charged residues has been evidenced in a study of barnase (Horovitz *et al.*, 1990), where the strengths of two salt bridges sharing a common arginine residue is increased by co-operative interactions within the salt-linked triad. Of particular interest is the observation that hyperthermophilic proteins generally possess not only an increased number of surface salt bridges (Creighton, 1993), but also an increased number of ionic networks involving sometimes up to 18 salt-bridged residues (Yip *et al.*, 1995; Hennig *et al.*, 1995; Goldman, 1995; Tanner *et al.*, 1996; Wallon *et al.*, 1997; Pappenberger *et al.*, 1997).

To examine whether such a co-operative enhancement of the strengths of salt bridges within larger ionic networks could be observed in our simulations, we grouped charged residues into clusters, so as to maximise the negative sum of intra-cluster contributions to the negative quantity $\Delta E_{cc}^{elec} \equiv E_{cc}^{elec}$ (360 K) $- E_{cc}^{elec}$ (300 K), see Table 1. The optimal grouping of charged residues was found to involve three clusters C_1, C_2, and C_3, of 16, 8, and 8 residues, respectively. These clusters are depicted in Figure 9. The contributions of the intra-cluster interactions to ΔE_{cc}^{elec} (−346 kJ/mol) for C_1, C_2 and C_3 are −796, −340, and −259 kJ/mol, respectively. The inter-cluster contributions are +426, +530, and +92 kJ/mol for $C_1 - C_2$, $C_1 - C_3$, and $C_2 - C_3$, respectively. Assuming that ΔE_{cc}^{elec} is correlated with the gain in stability of the native state at elevated temperature due to ionisable surface side-chains, these three clusters represent the ionic networks responsible for the dominant contributions to this gain in stability. The magnitude of the sum of intra-cluster contributions to ΔE_{cc}^{elec} is about four times larger than ΔE_{cc}^{elec} itself. This clearly indicates that the

relevant entities for discussing the effect of charged residues on the thermophilicity of Sac7d are these ionic networks rather than individual salt bridges.

Examination of the spatial distribution of the charged residues within these clusters (Figure 9) reveals that the clusters do not correspond with ensembles of residues in specific regions of the protein surface. They may contain groups of residues in close proximity. For example, the C_1 cluster comprises the group N terminus, Lys3, Lys5, Glu14, the group Glu64, Lys65, Lys66, C terminus, and the ion pairs Arg25-Glu47 and Asp56-Arg60, whereas the C_2 cluster comprises the group Lys7, Lys9, Glu11, Glu12. However, the clusters themselves extend over the whole surface of the protein. It also appears that there is only a limited correlation between the formation of a salt bridge between two residues and the appartenance to a same cluster. For example, among the 17 charge pairs within an average distance inferior to 0.6 nm during either the 300 K or the 360 K simulations, only nine correspond with pairs belonging to the same cluster. It follows that if some salt bridges contribute to the gain in stability of the native state (see Figure 8), namely the intra-cluster salt bridges, others do not appear to do so. Considering the very unfavourable inter-cluster contributions to ΔE_{cc}^{elec}, it is even likely that at least some of the inter-cluster salt bridges contribute unfavourably to the stability of the folded state.

We also observe that a significant stabilisation may result from the interaction of residues which, upon visual inspection, would not seem to interact strongly. A typical example is cluster C_3, which has a large contribution to ΔE_{cc}^{elec} (−258 kJ/mol), but exclusively involves residues at large distances from each other. Finally, considering the intra and inter-cluster contributions to ΔE_{cc}^{elec} (see above), we estimate that the complete removal (i.e. the mutation of all residues to Ala) of cluster C_1, C_2 or C_3, assuming no corresponding relaxation of the protein structure, would increase the magnitude of ΔE_{cc}^{elec} by 160, 178 or 363 kJ/mol, and thus further stabilise the native state. This surprising observation suggests that the three clusters actually interfere electrostatically with each other, and that the placement of charged residues on the surface of the protein may not be optimal from a purely electrostatic point of view. Thus, a gain in stability might in some cases be obtained by the removal of specific ionisable residues rather than the introduction of additional ones. Note, however, that the simultaneous removal of two clusters, or of all of them, is predicted to have a very unfavourable effect on the stability of the folded state, namely an increase of ΔE_{cc}^{elec} by 347 to 710 kJ/mol.

Discussion

Here, we report molecular dynamics simulations of recombinant Sac7d at 300 K, 360 K, and 550 K. The analysis of several time-resolved properties (radius of gyration, rms deviation from the NMR structure, secondary structure evolution) indicates that at the three temperatures, the system is in equilibrium on the nanosecond time scale. Comparison with NOE-derived proton-proton distances shows that the simulations at 300 K, and to a lesser extent at 360 K, are also in agreement with experimental data. The secondary structure undergoes only minor fluctuations in the simulations at the two lower temperatures. In contrast, rapid and important fluctuations in the hydrogen-bonding pattern are observed during the simulation at 550 K. However, even at this temperature, all secondary structure elements characteristic of the native state are still present at the end of the simulation.

The observation that the protein does not unfold during the simulation at 550 K contrasts with previous studies of protein unfolding by computer simulation, where denaturation was observed at similar temperatures and within similar time scales (Mark & van Gunsteren, 1992; Daggett & Levitt, 1992, 1993; Tirado-Rives & Jorgensen, 1993; Caflisch & Karplus, 1994, 1995; Hünenberger *et al.*, 1995; Li & Daggett, 1998). The absence of unfolding in this study could be a consequence of the enhanced (thermodynamic or kinetic) thermal stability of the hyperthermophilic protein Sac7d compared to the mesophilic proteins considered in these studies. Note, however, that in a recent molecular dynamics study of rubredoxin (using a cutoff scheme for the electrostatic interactions), the hyperthermophilic protein was found to unfold faster than its mesophilic homologue at 500 K (Lazaridis *et al.*, 1997). The present work is to our knowledge the first reported simulation of a protein at high temperature using the P_3M (particle-particle particle-mesh) method for handling electrostatic interactions, whereas previous studies used cutoff-based methods. Consequently, other causes may be involved in the unexpected stability of the protein in the present simulation at 550 K, for example: (i) artifacts associated with the use of a cutoff scheme (Smith & Pettitt, 1991; Schreiber & Steinhauser, 1992a,b) may have driven or accelerated the unfolding process in the previous studies, so that it occurred on the nanosecond time scale; (ii) the use of the P_3M method together with force-field parameters calibrated for use with a cutoff scheme may lead to an imbalance between different force-field terms and result in an increase in protein stability; and (iii) the artificial periodicity imposed by the use of the P_3M method may specifically stabilise the folded state of the protein. The latter point has been addressed in a recent continuum electrostatics study (Hünenberger & McCammon, 1998a). Based on 180 conformations taken from the present simulation at 550 K, electrostatic free energies corresponding to periodic (Ewald) and non-periodic boundary conditions were calculated and compared. The results show that artificial periodicity stabilises the native state

of the protein by 3-4 kJ/mol (0.7-0.9 k_BT at 550 K) with respect to the most unfolded conformation observed during the simulation (Hünenberger & McCammon, 1998a).

The stability of the simulations at 300 K and 360 K allows us to analyse energetic and structural properties in detail. The energetic analysis reveals the following two changes upon raising the temperature from 300 K to 360 K: (i) a less favourable protein-solvent interaction energy, and (ii) a more favourable protein-protein interaction energy. Both effects are almost exclusively electrostatic in nature, and dominated by contributions from charged residues at the surface of the protein. The analysis of the radial distribution functions and radial dipole orientation correlation functions around charged side-chains indicates that the loss of charge solvation upon increasing the temperature is due to a partial loss of the spatial and orientational structure in the solvation shells around the charge, as a consequence of the increased thermal motion in the solvent. With the decrease in the magnitude of the solvation free energy of a given charged residue at higher temperature, we observe a concomitant increase in magnitude of the Coulombic interaction of the residue with other charged residues, the two effects significantly compensating each other. Finally, the analysis of the pair distribution functions for charged side-chains allows us to attribute this increase in the magnitude of the Coulombic interactions between charged residues to the tightening of residue pairs of opposite charges and within distances in the range of 0.3-0.4 nm, i.e. salt bridges.

The implications of the above observations on the importance of salt bridges for thermostability can be drawn by assuming that charges are on average more distant from one another in the unfolded state than in the folded state. Following this assumption, salt bridges may stabilise the folded state of the protein relative to its unfolded state at elevated temperatures for two reasons. First, the reduced solvation of charged side-chains observed in our simulation at higher temperature is likely to affect the unfolded state more adversely than the folded state (Elcock, 1998). Second, the tightening of salt bridges at higher temperature, which partly compensates for this loss of solvation, will increase the energetic cost for the disruption of salt bridges when the protein unfolds. This hypothesis could be verified by using the results presented in Tables 2 and 3 as a guide to perform site-directed mutagenesis experiments on Sac7d. For instance, the mutation to alanine of residues exhibiting a large and negative $\Delta E_I^{\mathrm{elec}}$ and a large and positive $\Delta G_I^{\mathrm{solv}}$, e.g. Glu53, Lys7 or Glu64, should cause a degradation of the protein thermal stability. Conversely, the mutation to alanine of the residues Arg25, Lys9, Arg60 or Glu11 is predicted to further stabilise the folded state.

If the above explanation for the stabilising effect of salt bridges is correct, any decrease in the screening properties of the solvent, such as that brought by an increase in temperature, should render the contribution of salt bridges to the stability of the native state more favourable. For example, the decrease in the dielectric constant of water upon increasing the temperature could be mimicked by the addition of a co-solvent of lower permittivity. Such a measurement of the relative stability, in different solvent mixtures, of wild-type and mutant proteins lacking a salt bridge would allow to test our hypothesis. In the opposite, an increase in the solvent screening properties, such as an increase in the ionic strength, should result in a less favourable contribution of salt bridges to stability. Such an effect of ionic strength has indeed been observed in mutagenesis studies of barnase (Horovitz *et al*., 1990) and of the thermostable DNA-binding protein HU (Kawamura *et al*., 1997).

Since our results suggest a systematic temperature dependence of the electrostatic interaction energy ($E_{\mathrm{PP}}^{\mathrm{elec}}$) within the protein, namely a decrease in this quantity (more favourable interactions) with increasing temperature, we examined a number of low-temperature crystallographic structures of ribonuclease A (Tilton *et al*., 1992) and lysozyme (Young *et al*., 1993, 1994; Kurinov & Harrison, 1995) to see if such a systematic dependence could be observed experimentally. The results of this analysis are presented in Table 4. For ribonuclease A and the lysozyme (tetragonal) structures solved by Kurinov & Harrison (1995), the variation of $E_{\mathrm{PP}}^{\mathrm{elec}}$ between ~100 K and ~300 K is small and negative (−17 and −32 kJ/mol, respectively). However, the variation of this quantity with temperature is highly irregular and spans a range of about 400-500 kJ/mol. The values calculated from the lysozyme (tetragonal) structures solved by Young *et al*. (1993, 1994) do not agree with those calculated from the structures of Kurinov & Harrison (1995). The former are higher in electrostatic energy by 450-750 kJ/mol, and $E_{\mathrm{PP}}^{\mathrm{elec}}$ decreases by 330 kJ/mol between 100 K and 298 K. Finally, a decrease of about 500 kJ/mol is observed for the lysozyme (monoclinic) structures solved by the same group. These results appear contradictory, probably due to a very large uncertainty in the calculated electrostatic energies. This uncertainity may be caused by a number of factors: (i) intermolecular packing contacts and possibly intermolecular salt bridges occur in the crystal, but are not accounted for in $E_{\mathrm{PP}}^{\mathrm{elec}}$; (ii) the increase in temperature is accompanied by the expansion of the unit cell, the anisotropic expansion of the protein, and changes in the packing of protein molecules; (iii) the solvent properties at these subzero temperatures and in the crystalline environment probably differ from those of bulk water (in particular the dielectric properties); and (iv) flash-frozen crystals may not represent equilibrium configurations, but rather kinetically trapped states, with possible structural distortions.

Several experiments confirm more directly that salt bridges do stabilise the native state of thermophile proteins at elevated temperatures, in agree-

Table 4. Electrostatic energies calculated for crystal structures of lysozyme and ribonuclease A at different temperatures

RNase A[a]		Lysozyme[b] (tetragonal)		Lysozyme[d] (monoclinic)	
T	E_{PP}^{elec}	T	E_{PP}^{elec}	T	E_{PP}^{elec}
98	−5980	95	−4952		
130	−6199	100	−4197[c]	100	−3566/−3559
160	−5933	120	−4797		
180	−5972	180	−4969		
220	−5823	250	−5023		
240	−5820	280	−5320		
260	−5882	295	−4984		
300	−5997	298	−4530[c]	298	−4026/−4083
320	−5970				

Temperature dependence of the intraprotein electrostatic interaction energy E_{PP}^{elec} (units, kJ/mol) calculated (using the GROMOS96 charge set) from crystallographic structures of lysozyme and ribonuclease A solved at different temperatures T (units, K)

[a] Data from Tilton *et al.* (1992); PDB entries 1-9RAT.

[b] Data from Kurinov & Harrison (1995); PDB entries 1LSA-F, tetragonal crystalline form.

[c] Data from Young *et al.* (1993, 1994); PDB entries 5,6LYT, tetragonal crystalline form.

[d] Data from Young *et al.* (1993); PDB entries 3,4LYT, monoclinic crystalline form, two molecules per asymmetic unit cell refined independently.

ment with the conclusions of the present study. For example, the importance of a surface salt bridge (Glu34-Lys38) for the stability of the thermostable DNA-binding protein HU from *Bacillus stearothermophilus* was evidenced experimentally by site-directed mutagenesis (Kawamura *et al.*, 1997). The mutations Glu34 → Asp, Glu34 → Gln, or Lys38 → Asn, which all disrupt the salt bridge, lead to a decrease in the thermal stability of this protein by 2.08, 2.71, and 2.17 kJ/mol, respectively. Conversely, in the mesophilic homologue from *Bacillus subtilis*, the simultaneous mutations Asp34 → Glu and Asn38 → Lys, which introduce the salt bridge present in the hyperthermophilic protein, lead to an increase in the thermal stability by 1.92 kJ/mol.

Extensive ionic networks, where individual residues are involved in multiple interactions, have also been suggested as a common mechanism for promoting thermostability (Yip *et al.*, 1995; Pappenberger *et al.*, 1997). Hyperthermophilic proteins generally possess not only an increased number of surface salt bridges (Creighton, 1993), but also an increased number of such ionic networks (Yip *et al.*, 1995; Hennig *et al.*, 1995; Goldman, 1995; Tanner *et al.*, 1996; Wallon *et al.*, 1997; Pappenberger *et al.*, 1997). For instance, the comparison of glutamate dehydrogenase from the hyperthermophilic organism *Pyrococcus furiosis* with that of the mesophile *Clostridium symbiosium* revealed that in the mesophilic protein, 40 % of the charged residues are involved in ion pairs and 23 % of the ion pairs are involved in ionic networks, whereas these numbers are as high as 54 % and 62 %, respectively, for its hyperthermophilic homologue (Yip *et al.*, 1995). This suggests that the creation of an ion pair network may be energetically (and possibly also entropically) more favourable than the creation of an equivalent number of single pairwise salt bridges.

Such a co-operative effect was evidenced in a study of barnase for the surface salt-linked triad Asp8, Asp12 and Arg110, involving two salt bridges (Horovitz *et al.*, 1990). By performing double mutant cycles, it was estimated that the strengths of the Asp12-Arg110 and Asp8-Arg110 salt bridges are both increased by 3.2 kJ/mol within the triad due to the presence of the third residue. The stabilising effect of an ionic network of four charged residues (Arg20, Arg320, Asp323, and Glu326) was also identified in the hyperthermophilic protein glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima* (Pappenberger *et al.*, 1997). The mutations Arg20 → Ala or Arg20 → Asn, which change the network into a single salt bridge lead to a decrease in the free energy barrier for the (irreversible) thermal denaturation by about 4 kJ/mol at 100 °C. A similar network, which is found in the homologous enzyme from *Escherichia coli* but not from other mesophilic organisms, also appears to contribute to an increased thermal stability of the protein from this organism (Pappenberger *et al.*, 1997). The role of an ionic network in increased thermal stability was also evidenced in the case of glutamute dehydrogenase from *T. litoralis* (Vetriani *et al.*, 1998). The mutation Thr183 → Glu in this enzyme enables the formation of a six-residue ionic network analogous to that observed in the highly homologous (but 16-fold more thermostable) protein from *P. furiosis*. It was found that the mutation leads to a four-fold stabilisation of the enzyme provided that another mutation (Asp167 → Thr) is introduced simultaneously.

The analysis of the present simulations in terms of clusters of charged residues is in line with the results of the above experimental studies. We find that three large clusters (ionic networks), rather than individual salt bridges, provide most of the electrostatic stabilisation of Sac7d towards thermal denaturation. Many residues in close proximity are

found within the same cluster. However, the clusters themselves extend over the whole surface of the protein. This analysis also indicates that, although the overall electrostatic interaction of charged surface residues clearly stabilises the protein, individual salt bridges may or may not do so. In other words, the stabilisation can probably not be reduced to a sum of pairwise salt bridge interactions. Moreover, it is likely that at least some of the inter-cluster salt bridges contribute unfavourably to the stability of the folded state. We note that such a high degree of co-operativity in the interaction between charged surface residues is not unexpected in view of the extremely high density of charges on the surface of Sac7d (see Figure 9), but may not necessarily generalise to larger proteins with a lower density of surface charges.

Two consequences for the engineering of thermostable proteins are that: (i) a simple consideration of the local environment of a charged side-chain may be insufficient to predict the effect of a specific mutation, since strong interactions with residues that are not in the immediate vicinity may be involved; and (ii) a gain in stability may in some cases be obtained by the removal of specific ionisable residues (or salt bridges) rather than the introduction of additional ones. Hence, if judicious placement of salt bridges across the protein surface is certainly a useful strategy in the design of thermostable proteins, it may prove difficult to determine the detailed rules dictating the optimal location for their introduction. We believe that theoretical studies such as the present one may in some cases provide a guide for selecting mutation sites.

As shown here, molecular dynamics simulations enable us to perform a detailed comparison between the energetic and structural properties of a protein at room temperature and at elevated temperature, thus providing information which is not directly accessible to experiment. Such a comparison between equilibrium properties relies, of course, on the stability of the trajectories at the two temperatures, and is therefore only possible for proteins stable at elevated temperature (hyperthermophilic proteins). The results demonstrate how simulations may provide a link between the structural observation that hyperthermophilic proteins have statistically more surface salt bridges and ionic networks, and their enhanced thermal stability. This study illustrates how theoretical tools can be used to broaden our understanding of the factors governing protein stability.

Methods

Molecular model and computational procedures

Molecular dynamics simulations were performed using a modified version of the GROMOS96 program (van Gunsteren *et al*., 1996) incorporating the particle-particle particle-mesh (P_3M) method for handling electrostatic interactions (Hockney & Eastwood, 1988; Luty *et al*., 1994, 1995). The GROMOS 43A1 united-atom force field (van Gunsteren *et al*., 1996) was employed together with the simple point charge (SPC) water model (Berendsen *et al*., 1981). Initial co-ordinates were taken from the NMR solution structure of recombinant Sac7d (Edmondson *et al*., 1995) as deposited in the Brookhaven Protein Data Bank (entry code 1SAP). The protonation states of ionisable residues were chosen according to the pK_a of the isolated amino acid and a pH of 7, resulting in a total charge of $+6\,e$. The N and C termini were considered ionised. The protein was immersed into a rectangular box of dimensions 4.66 nm × 4.90 nm × 5.58 nm containing 3956 SPC water molecules to reach a final density of about 1.0 g/cm^3 for the water surrounding the protein. No counter-ions were added. Bond lengths were constrained by application of the SHAKE procedure (Ryckaert *et al*., 1977). A time-step of 2 fs was used to integrate the equations of motion. The P_3M method, including the Wigner self-energy term, was used to compute the electrostatic interactions (Hockney & Eastwood, 1988; Luty *et al*., 1994, 1995; Hünenberger & McCammon, 1998a,b):

$$E^{\mathrm{elec}} = \frac{1}{4\pi\varepsilon_0}\frac{1}{2}\left\{ \sum_{i=1}^{N_{\mathrm{at}}} \sum_{j=1,j\neq i}^{N_{\mathrm{at}}}{}' q_i q_j \frac{\mathrm{erfc}(\alpha r_{ij})}{r_{ij}} + \sum_{i=1}^{N_{\mathrm{at}}} q_i \Phi_\gamma(\mathbf{r}_i) - \frac{2\alpha}{\pi^{1/2}} \sum_{i=1}^{N_{\mathrm{at}}} q_i^2 - \frac{\pi}{V\alpha^2}\left(\sum_{i=1}^{N_{\mathrm{at}}} q_i\right)^2 \right\} \quad (1)$$

where q_i and $\mathbf{r}_i$ are the charge and position vector of atom i, r_{ij} is the minimum-image distance between atoms i and j, V the volume of the rectangular unit cell, α^{-1} the width of the Gaussian charge shaping function (0.3 nm), and erfc the complementary error function. The prime after the summation sign indicates that when atoms i and j are excluded neighbours (1$^{\mathrm{st}}$ or 2$^{\mathrm{nd}}$ covalent neighbours), erfc should be substituted by erfc -1. The k-space potential, $\Phi_\gamma(\mathbf{r})$, was calculated using fast Fourier transforms as described by Luty *et al*. (1995), using a grid spacing of 0.08 nm and a weighting between first and second difference operators $\beta = 1.35$. The real-space contribution was truncated, together with the Lennard-Jones interaction, at a distance of 0.9 nm. The non-bonded pair list was updated every ten steps. Simulations were performed at constant volume and temperature. The temperature was maintained by weak coupling to a heat bath (Berendsen *et al*., 1984) with a relaxation time $\tau_T = 0.1$ ps. Three simulations were started at temperatures of 300 K, 360 K, and 550 K in the following way. After relaxation of the solvent co-ordinates (with fixed positions of the protein atoms), the temperature was progressively increased to 300 K within 15 ps while positionally restraining solute atoms with a decreasing force constant. After this initial heating period, a copy of the system was further heated to 360 K (within 10 ps), and another copy to 550 K (within 60 ps). All simulations were allowed a total heating/equilibration period of 150 ps, after which the systems were simulated during 1.0 ns at 300 K and 360 K, and 0.9 ns at 550 K. Co-ordinates were saved every 0.1 ps for analysis.

Two specific choices made in the above setup demand a brief discussion. First, the charge of the protein was not explicitly neutralised by counter-ions. In this case,

due to the use of the P_3M method, the net charge in the unit cell is implicitly neutralised by a homogeneous background charge. Although this represents only an approximate description of a counter-ion atmosphere (Hünenberger & McCammon, 1998a,b), it offers the advantages of: (i) improving the convergence properties of the simulations; (ii) eliminating the dependence of the results on an arbitrary initial placement of the counter-ions; and (iii) limiting the number of degrees of freedom to be considered in the analysis of salt bridges. Second, constant volume simulations are performed rather than constant pressure simulations. The SPC water model is parametrised for a temperature of 298 K, and constant pressure simulations would lead to unrealistically low densities of the solvent at elevated temperatures. Experimentally, the density of water decreases only by 3×10^{-2} g/cm^3 from 300 K to 360 K (Lide, 1995), providing reasonable justification for the constant volume setup, although it may lead to elevated pressures in the 360 K and 550 K simulations. However, since even very high pressures induce only limited shifts in the folding equilibria of proteins, which are only significant under conditions of temperature and pH close to the transition (Zipp & Kauzmann, 1973; Li *et al.*, 1976; Weber & Drickamer, 1983), artifactual effects of elevated pressure should remain small. Nevertheless, high pressures may somewhat slow down the dynamics of the system (Hawley & Mitchell, 1975; Brunne & van Gunsteren, 1993).

Analysis of the trajectories

The radius of gyration and the root-mean-square (rms) deviation of C^α atomic positions with respect to the NMR structure were monitored as a function of time. The radius of gyration is defined as:

$$R_{\mathrm{gyr}}(t) = \left[\frac{1}{N_{\mathrm{at}}} \sum_{i=1}^{N_{\mathrm{at}}} (\mathbf{r}_i(t) - \mathbf{r}_{CM}(t))^2 \right]^{1/2} \tag{2}$$

where $\mathbf{r}_i(t)$ and $\mathbf{r}_{CM}(t)$ are the co-ordinates of atom i and of the centre of mass of the protein, respectively, at time t, and $N_{\mathrm{at}} = 696$. The C^α rms deviation is defined as:

$$d_{\mathrm{rms}}(t) = \left[\frac{1}{N_{C_\alpha}} \sum_{i=1}^{N_{C^\alpha}} (\mathbf{r}_i(t) - \mathbf{r}_i^{\mathrm{NMR}})^2 \right]^{1/2} \tag{3}$$

where $\mathbf{r}_i^{\mathrm{NMR}}$ is the co-ordinate of atom i in the NMR structure, after a least-squares fit superimposition with the structure at time t, and $N_{C^\alpha} = 66$.

To assess the compatibility of the simulations with experimental data, a set of 620 upper-bound site-site distances was constructed from the 1821 NOE-derived proton-proton distances used in the structure refinement and corresponding to a temperature of 35 °C (J. W. Shriver & S. P. Edmondson, personal communication). Due to the absence of stereospecific assignments, pseudo-sites were used for all methylene, methyl, geminal methyl pairs (C^γ of Val and C^δ of Leu) and aromatic (H^δ and H^ε in Phe and Tyr) protons with the appropriate pseudo-site corrections (van Gunsteren *et al.*, 1996). For each site pair, the largest experimental distance between protons belonging to the individual sites was taken, and increased by the sum of pseudo-sites corrections for the two sites. These NOE-derived site-site distances were compared with those obtained from the simulations, averaged over successive windows of length $\Delta\tau = 50$ ps. The sum of violations was calculated as:

$$S(t) = \sum_{i=1, \langle R_i(\tau) \rangle_{t,\Delta\tau} > R_i^{\mathrm{NMR}}}^{N_{\mathrm{pairs}}} [\langle R_i(\tau) \rangle_{t,\Delta\tau} - R_i^{\mathrm{NMR}}] \tag{4}$$

where $\langle \cdots \rangle_{t,\Delta\tau}$ denotes averaging over trajectory frames from $\tau = t - \Delta\tau/2$ to $\tau = t + \Delta\tau/2$, $R_i(\tau)$ is the distance between the ith site pair at time τ, R_i^{NMR} the NOE-derived distance, and $N_{\mathrm{pairs}} = 620$.

Energy contributions were analysed in detail. The total potential energy is the sum of a covalent term (E^{cov}), an electrostatic term (E^{elec}), and a van der Waals term (E^{vdw}):

$$E^{\mathrm{pot}} = E^{\mathrm{cov}} + E^{\mathrm{elec}} + E^{\mathrm{vdw}} \tag{5}$$

The two latter terms were further decomposed as:

$$E^{\mathrm{elec}} = E_{\mathrm{ss}}^{\mathrm{elec}} + E_{\mathrm{ps}}^{\mathrm{elec}} + E_{\mathrm{pr}}^{\mathrm{elec}} + E_{\mathrm{pp}}^{\mathrm{elec}}$$

and:

$$E^{\mathrm{vdw}} = E_{\mathrm{ss}}^{\mathrm{vdw}} + E_{\mathrm{ps}}^{\mathrm{vdw}} + E_{\mathrm{pp}}^{\mathrm{vdw}} \tag{6}$$

where p denotes the protein in the central unit cell, r the ensemble of the periodic replicas of the protein, and s the solvent. The quantities $E_{\mathrm{pp}}^{\mathrm{elec}} + E_{\mathrm{pr}}^{\mathrm{elec}}$ and $E_{\mathrm{ss}}^{\mathrm{elec}}$ were obtained by re-evaluating the electrostatic interaction energy for each trajectory frame using equation (1), including only protein charges, or only solvent charges, respectively. The quantity $E_{\mathrm{pp}}^{\mathrm{elec}}$ was evaluated by summation of the pairwise Coulombic interaction energies over protein atom pairs (excluding first and second covalent neighbours), and was further partitioned into:

$$E_{\mathrm{pp}}^{\mathrm{elec}} = E_{\mathrm{cc}}^{\mathrm{elec}} + E_{\mathrm{cn}}^{\mathrm{elec}} + E_{\mathrm{nn}}^{\mathrm{elec}} \tag{7}$$

where c denotes the ensemble of atoms belonging to side-chains of charged residues (Arg, Lys, Asp and Glu) plus the N-terminal ammonium and C-terminal carboxylate groups, and n the overall neutral ensemble of all remaining atoms (backbone amide groups and non-ionisable side-chains).

For all ionisable residues, the side-chain (I)-water dipole (D) radial distribution functions, were calculated as:

$$g_{\mathrm{ID}}(r) = \frac{1}{4\pi\rho r^2 \Delta r} \left\langle \sum_{J, r \leqslant r_{IJ}(\tau) < r + \Delta r}^{N_J} 1 \right\rangle_\tau \tag{8}$$

where $\langle \cdots \rangle_\tau$ denotes averaging over all frames of the trajectory, $r_{\mathrm{IJ}}(\tau)$ is the minimum-image distance between a reference atom of side-chain I (Arg, C^ζ, Lys, N^ε, Asp, C^γ, Glu, C^δ and the termini: N or C atom) and the oxygen atom of water molecule J at time τ, N_J the total number of water molecules (3956), ρ the bulk number density of water, and Δr the window size (0.02 nm). The radial dipole orientation correlation functions were calculated as:

$$c_{\mathrm{ID}}(r) = \frac{1}{4\pi\rho r^2 \Delta r g_{\mathrm{ID}}(r)} \left\langle \sum_{J, r \leqslant r_{IJ}(\tau) < r + \Delta r} \frac{\mathbf{e}_J(\tau) \cdot \mathbf{r}_{IJ}(\tau)}{r_{IJ}(\tau)} \right\rangle_\tau \tag{9}$$

where $\mathbf{e}_J(\tau)$ is a unit vector along the dipole of water molecule J at time τ. When $g_{\mathrm{ID}}(r) = 0$, we defined $c_{\mathrm{ID}}(r) = 0$.

The solvation free energy G_I^{solv} of an ionisable side-chain I can be approximated by:

$$G_I^{\text{solv}} \approx \frac{1}{2} E_I^{\text{solv}} \approx -\frac{q_I \mu \rho}{2\varepsilon_0} \int_0^\infty dr\, g_{\text{ID}}(r) c_{\text{ID}}(r) \qquad (10)$$

where q_I is the charge of the side-chain, μ the dipole moment of the SPC water molecule (0.04734 e nm), and ε_0 the dielectric permittivity of vacuum. In this expression, we assume that (i) the side-chain–water interaction is equivalent to a point charge-point dipole interaction; (ii) any interaction with water molecules outside the unit cell centered around I can be neglected; and (iii) the electrostatic solvation free energy is proportional to the side-chain–solvent interaction energy E_I^{solv} by a factor 1/2, i.e. we are using a linear response approximation (Roux *et al*., 1990; Åqvist *et al*., 1994; Åqvist & Hansson, 1996). While this expression is only approximate, it is probably sufficient to capture the temperature variations of G_I^{solv} and offers the advantage of linking this thermodynamic quantity with the structural properties $g_{\text{ID}}(r)$ and $c_{\text{ID}}(r)$. The contribution of charged side-chain I to the intraprotein Coulombic energy was calculated as:

$$E_I^{\text{elec}} = \frac{1}{4\pi\varepsilon_0} \frac{1}{2} \left[\sum_{i \in I} \sum_{j \in I, j \neq i}{}' \frac{q_i q_j}{r_{ij}} + \sum_{J \neq I} \sum_{i \in I} \sum_{j \in J}{}' \frac{q_i q_j}{r_{ij}} \right] \qquad (11)$$

where the prime indicates that excluded neighbours are omitted from the summation. Hence, $E_{\text{cc}}^{\text{elec}} = \Sigma_I E_I^{\text{elec}}$.

Finally, approximating the charged side-chains as single points (the reference atoms as given above), we define $N_{++}(r, \tau)$ as the number of unique pairs (i, j) of positive charges (excluding $i = j$) at a distance $r_{ij}(\tau)$ at time τ satisfying $r \leqslant r_{ij}(\tau) < r + \Delta r$, divided by $\Delta r = 0.02$ nm. $N_{--}(r, \tau)$ and $N_{+-}(r, \tau)$ are defined similarly. $\langle N_{++} \rangle$, $\langle N_{--} \rangle$, and $\langle N_{+-} \rangle$ thus represent the number of unique pairs at a particular separation distance consisting of two positively, two negatively, or two oppositely charged side-chains, respectively, averaged over the entire trajectory. According to these definitions:

$$E_{\text{cc}}^{\text{elec}} \approx \frac{e^2}{4\pi\varepsilon_0} \int_0^\infty dr\, r^{-1} [\langle N_{++}(r, \tau) \rangle_\tau + \langle N_{--}(r, \tau) \rangle_\tau - \langle N_{+-}(r, \tau) \rangle_\tau] \qquad (12)$$

Temperature dependence of the dielectric permittivity of SPC water

Since the effect of temperature on salt bridges discussed in the present study is, at least in part, a consequence of the decrease in the dielectric permittivity of water upon increasing the temperature, it seemed important to assess the realism of the SPC water model with respect to this property. We thus performed separate 1 ns molecular dynamics simulations of two cubic boxes containing 512 SPC water molecules at 303 K and 363 K, and at the corresponding experimental densities of 0.996 and 0.965 g/cm^3, using the P_3M method and a setup analogous with that described for the protein simulations. The dielectric permittivities were calculated and error bars estimated from the anisotropy of the permittivity tensor (Hünenberger & van Gunsteren, 1998). We obtained relative permittivities of 69(±5) and 50(±4) at 303 K and 363 K, respectively. The former value compares well with other estimates from the literature (Hünenberger & van Gunsteren, 1998). The two values are slightly lower than the experimental values of 76.6 and 58.1 (Lide, 1995), but the calculated change of 19 is in excellent agreement with experiment.

Acknowledgements

The authors thank Dr Adrian Elcock for many insightful discussions, and Dr Wolfgang Weber and Dr Valérie Dillet for their critical reading of the manuscript. We acknowledge Dr John Shriver and Dr Stephen Edmondson for their help and for providing us with the NMR data. P.I.W.d.B. expresses gratitude to KNMP Stipendiafonds, Dr Saal van Zwanenbergstichting, Nederlandse Hersenstichting, Stichting Alzheimer Fonds, Utrecht University, and Bristol-Myers Squibb for financial support. P.H.H. acknowledges support from the Human Frontier Science Program Organization. This work is supported in part by grants from the National Institutes of Health, the National Science Foundation, and the San Diego Supercomputer Center.

References

Åqvist, J. & Hansson, T. (1996). On the validity of electrostatic linear response in polar solvents. *J. Phys. Chem.* **100**, 9512-9521.

Åqvist, J., Medina, C. & Samuelsson, J.-E. (1994). A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **7**, 385-391.

Baumann, H., Knapp, S., Lundbäck, T., Ladenstein, R. & Härd, T. (1994). Solution structure and DNA-binding properties of a thermostable from the archaeon *Sulfolobus solfataricus*. *Nature Struct. Biol.* **1**, 808-819.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. (1981). Interaction models for water in relation to protein hydration. In *Intermolecular Forces* (Pullman, B., ed.), pp. 331-342, Reidel, Dordrecht.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external heat bath. *J. Chem. Phys.* **81**, 3684-3690.

Brunne, R. M. & van Gunsteren, W. F. (1993). Dynamical properties of bovine pancreatic trypsin inhibitor from a molecular dynamics simulation at 5000 atm. *FEBS Letters*, **232**, 215-217.

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.

Caflisch, A. & Karplus, M. (1995). Acid and thermal denaturation of barnase investigated by molecular dynamics simulations. *J. Mol. Biol.* **252**, 672-708.

Creighton, T. E. (1993). *Proteins: Structures and Molecular Properties*, 2nd edit., Freeman, New York.

Daggett, V. & Levitt, M. (1992). A model of the molten globule state from molecular dynamics simulations. *Proc. Natl Acad. Sci. USA*, **89**, 5142-5146.

Daggett, V. & Levitt, M. (1993). Protein unfolding pathways explored through molecular dynamics simulations. *J. Mol. Biol.* **32**, 600-619.

Danson, M. J., Hough, D. W., Russell, R. J. M., Taylor, G. L. & Pearl, L. (1996). Enzyme thermostability and thermoactivity. *Protein Eng.* **9**, 629-630.

Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, **29**, 7133-7155.

Dill, K. A. (1993). Folding proteins: finding a needle in a haystack. *Curr. Opin. Struct. Biol.* **3**, 99-103.

Dobson, C. M., Šali, A. & Karplus, M. (1998). Protein folding: a perspective from theory and experiment. *Angew Chem. Int. Ed. Engl.* **37**, 868-893.

Edmondson, S. P., Qiu, L. & Shriver, J. W. (1995). Solution structure of the DNA-binding protein Sac7d from the hyperthermophile *Sulfolobus acidocaldarius*. *Biochemistry*, **34**, 13289-13304.

Elcock, A. H. (1998). The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J. Mol. Biol.* **284**, 489-502.

Elcock, A. H. & McCammon, J. A. (1997). Continuum solvation model for studying protein hydration thermodynamics at high temperatures. *J. Phys. Chem.* **101**, 9624-9634.

Frömmel, C. & Sander, C. (1989). Thermitase, a thermostable subtilisin: comparison of predicted and experimental structures and the molecular cause of thermostability. *Proteins: Struct. Funct. Genet.* **5**, 22-37.

Goldman, A. (1995). How to make my blood boil. *Structure*, **3**, 1277-1279.

Graziano, G., Barone, G., Catanzano, F. & Riccio, A. (1995). The extra-stability of thermophilic globular proteins: a thermodynamic approach. *Thermochim. Acta*, **269(270)**, 381-392.

Hawley, S. A. & Mitchell, R. M. (1975). An electrophoretic study of reversible protein denaturation: chymotrypsinogen at high pressures. *Biochemistry*, **14**, 3257-3264.

Hendsch, Z. S. & Tidor, B. (1994). Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**, 211-226.

Hennig, M., Darimont, B., Sterner, R., Kirschner, K. & Jansonius, J. N. (1995). 2.0 Å structure of indole-3-glycerol phosphate synthase from the hyperthermophile *Sulfolobus solfataricus*: possible determinants of protein stability. *Structure*, **3**, 1295-1306.

Hockney, R. W. & Eastwood, J. W. (1988). *Computer Simulation Using Particles*, Institute of Physics Publishing, Bristol.

Honig, B. & Yang, A.-S. (1995). Free energy balance in protein folding. *Advan. Protein Chem.* **46**, 27-58.

Horovitz, A., Serrano, L., Avron, B., Bycroft, M. & Fersht, A. R. (1990). Strength and co-operativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol.* **216**, 1031-1044.

Hünenberger, P. H. & McCammon, J. A. (1998a). Effect of artificial periodicity in simulations of biomolecules under Ewald boundary conditions: a continuum electrostatics study. *Biophys. Chem.* In the press.

Hünenberger, P. H. & McCammon, J. A. (1998b). Ewald artifacts in computer simulations of ionic solvation and ion-ion interactions: a continuum electrostatics study. *J. Chem. Phys.* In the press.

Hünenberger, P. H. & van Gunsteren, W. F. (1998). Alternative schemes for the inclusion of a reaction-field correction into molecular dynamics simulations: influence on the simulated energetic, structural, and dielectric properties of liquid water. *J. Phys. Chem.* **108**, 6117-6134.

Hünenberger, P. H., Mark, A. E. & van Gunsteren, W. F. (1995). Computational approaches to study protein unfolding: hen egg white lysozyme as a case study. *Proteins: Struct. Funct. Genet.* **21**, 196-213.

Israelachvili, J. N. (1992). *Intermolecular and Surface Forces*, 2nd edit., Academic Press, San Diego.

Jaenicke, R. (1981). Enzymes under extremes of physical conditions. *Annu. Rev. Biophys. Bioeng.* **10**, 1-67.

Jaenicke, R. (1991). Protein stability and molecular adaptation to extreme conditions. *Eur. J. Biochem.* **202**, 715-728.

Jaenicke, R. (1996). How do proteins acquire their three-dimensional structure and stability? *Naturwissenschaften*, **83**, 544-554.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.

Kawamura, S., Tanaka, I. & Kimura, M. (1997). Contribution of a salt bridge to the thermostability of DNA binding protein HU from *Bacillus stearothermophilus* determined by stire-directed mutagenesis. *J. Biochem.* **121**, 448-455.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Kurinov, I. V. & Harrison, R. W. (1995). The influence of temperature on lysozyme crystals. Structure and dynamics of protein and water. *Acta Crystallog. sect. D*, **51**, 98-109.

Lazaridis, T., Archontis, G. & Karplus, M. (1995). Enthalpic contribution to protein stability: insights from atom-based calculations and statistical mechanics. *Advan. Protein Chem.* **47**, 231-306.

Lazaridis, T., Lee, I. & Karplus, M. (1997). Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. *Protein Sci.* **6**, 2589-2605.

Li, A. & Daggett, V. (1998). Molecular dynamics simulation of the unfolding of barnase: characterization of the major intermediate. *J. Mol. Biol.* **275**, 677-694.

Li, T. M., Hook, J. W., III, Drickamer, H. G. & Weber, G. (1976). Plurality of pressure-denatured forms in chymotrypsinogen and lysozyme. *Biochemistry*, **15**, 5571-5580.

Lide, D. R. (1995). *CRC Handbook of Chemistry and Physics*, 75th edit., CRC Press Inc., Boca Raton, USA.

Luty, B. A., Davis, M. E., Tironi, I. G. & van Gunsteren, W. F. (1994). A comparison of particle-particle particle-mesh and Ewald methods for calculating electrostatic interactions in periodic molecular systems. *Mol. Simul.* **14**, 11-20.

Luty, B. A., Tironi, I. G. & van Gunsteren, W. F. (1995). Lattice-sum methods for calculating electrostatic interactions in molecular simulations. *J. Chem. Phys.* **103**, 3014-3021.

Madigan, M. T. & Marrs, B. L. (1997). Extremophiles. *Sci. Am.* **276**, 66-71.

Makhatadze, G. I. & Privalov, P. L. (1995). Energetics of protein structure. *Advan. Protein Chem.* **47**, 307-425.

Maras, B., Consalvi, V., Chiaraluce, R., Politi, L., De Rosa, M., Bossa, R., Scandurra, R. & Barra, D. (1992). The protein sequence of glutamate dehydrogenase from *Sulfolobus solfataricus*, a thermoacidophilic archaebacterium. Is the presence of N-epsilon-methyllysine related to thermostability? *Eur. J. Biochem.* **203**, 81-87.

Mark, A. E. & van Gunsteren, W. F. (1992). Simulation of the thermal denaturation of henn egg white lysozyme: trapping the molten globule state. *Biochemistry*, **31**, 7745-7748.

Matthews, B. W. (1993). Structural and genetic analysis of protein stability. *Annu. Rev. Biochem.* **62**, 139-160.

McAfee, J. G., Edmondson, S. P., Datta, P. K., Shriver, J. W. & Gupta, R. (1995). Gene cloning, expression,

and characterization of the Sac7 proteins from the hyperthermophile *Sulfolobus acidocaldarius*. *Biochemistry*, **34**, 10063-10077.

McAfee, J. G., Edmondson, S. P., Zegar, I. & Shriver, J. W. (1996). Equilibrium DNA binding of Sac7d protein from the hyperthermophile *Sulfolobus acidocaldarius*: fluorescence and circular dichroism studies. *Biochemistry*, **35**, 4034-4045.

McCrary, B. S., Edmondson, S. P. & Shriver, J. W. (1996). Hyperthermophile protein folding thermodynamics: differential scanning calorimetry and chemical denaturation of Sac7d. *J. Mol. Biol.* **264**, 784-805.

Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. (1997). Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545-600.

Pappenberger, G., Schurig, H. & Jaenicke, R. (1997). Disruption of an ionic network leads to accelerated thermal denaturation of D-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima*. *J. Mol. Biol.* **274**, 676-683.

Reddy, T. R. & Suryanarayana, T. (1988). Novel histone-like DNA-binding protein in the nucleoid from the acidothermophilic archaebactrium *Sulfolobus acidocaldarius* that protect DNA against thermal denaturation. *Biochim. Biophys. Acta*, **949**, 87-96.

Rees, D. C. & Adams, M. W. W. (1995). Hyperthermophiles: taking the heat and loving it. *Structure*, **3**, 251-254.

Robinson, H., Gao, Y.-G., McCrary, B. S., Edmondson, S. P., Shriver, J. W. & Wang, A. H.-J. (1998). The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature*, **392**, 202-205.

Roux, B., Yu, H.-A. & Karplus, M. (1990). Molecular basis for the Born model of ion solvation. *J. Phys. Chem.* **94**, 4683-4688.

Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327-341.

Šali, D., Bycroft, M. & Fersht, A. R. (1991). Surface electrostatic interactions contribute little to stability to barnase. *J. Mol. Biol.* **220**, 779-788.

Schreiber, H. & Steinhauser, O. (1992a). Cutoff size does strongly influence molecular dynamics results on solvated polypeptides. *Biochemistry*, **31**, 5856-5860.

Schreiber, H. & Steinhauser, O. (1992b). Molecular dynamics studies of solvated polypeptides: why the cut-off scheme does not work. *Chem. Phys.* **168**, 75-89.

Smith, P. E. & Pettitt, B. M. (1991). Peptides in ionic solutions: a comparison of the Ewald and switching function techniques. *J. Chem. Phys.* **95**, 8430-8441.

Spassov, V. Z., Karshikoff, A. D. & Ladenstein, R. (1995). The optimization of protein-solvent interactions: thermostability and the role of hydrophobic and electrostatic interactions. *Protein Sci.* **4**, 1516-1527.

Stetter, K. O., Fiala, G., Huber, G., Huber, R. & Segerer, A. (1990). Hyperthermophilic microorganisms. *FEMS Microbiol. Rev.* **75**, 117-124.

Tanner, J. J., Hecht, R. M. & Krause, K. L. (1996). Determinants of enzyme thermostability observed in the molecular structure of *Thermus aquaticus* D-glyceraldehyde-3-phosphate dehydrogenase at 2.5 Å resolution. *Biochemistry*, **35**, 2597-2609.

Tilton, R. F., Jr, Dewan, J. C. & Petsko, G. A. (1992). Effects of temperature of protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320 K. *Biochemistry*, **31**, 2469-2481.

Tirado-Rives, J. & Jorgensen, W. L. (1993). Molecular dynamics simulations of the unfolding of apomyoglobin in water. *Biochemistry*, **32**, 4175-4184.

Tomschy, A., Böhm, G. & Jaenicke, R. (1994). The effect of ion pairs on the thermal stability of D-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima*. *Protein Eng.* **7**, 1471-1478.

van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R. P. & Tironi, I. G. (1996). *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, BIOMOS, Zürich, Groningen.

Vetriani, C., Maeder, D. L., Tolliday, N., Yip, K. S.-P., Stillman, T. J., Britton, K. L., Rice, D. W., Klump, H. H. & Robb, F. T. (1998). Protein thermostability above 100 °C: a key role for ionic interactions. *Proc. Natl Acad. Sci. USA*, **95**, 12300-12305.

Vieille, C. & Zeikus, J. G. (1996). Thermozymes: identifying molecular determinants of protein structural and functional stability. *Trends Biotechnol.* **14**, 183-191.

Vogt, G., Woell, S. & Argos, P. (1997). Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol. Biol.* **269**, 631-643.

Wallon, G., Kryger, G., Lovett, S. T., Oshima, T., Ringe, D. & Petsko, G. A. (1997). Crystal structures of *Escherichia coli* and *Salmonella typhimurium* 3-isopropylmalate dehydrogenase and comparison with their thermophilic counterpart from *Thermus thermophilus*. *J. Mol. Biol.* **266**, 1016-1031.

Watanabe, K., Chishiro, K., Kitamura, K. & Suzuki, Y. (1991). Proline residues responsible for thermostability occur with high frequency in the loop region of an extremely thermostable oligo-1,6-glucosidase. *J. Biol. Chem.* **266**, 24287-24294.

Watanabe, K., Hata, Y., Kizaki, H., Katsube, Y. & Suzuki, Y. (1997). The refined crystal structure of *Bacillus cereus* oligo-1,6-glucosidase at 2.0 Å resolution: structural characterization of proline-substitution sites for protein thermostabilization. *J. Mol. Biol.* **269**, 142-153.

Weber, G. & Drickamer, H. G. (1983). The effect of high pressure upon proteins and other biomolecules. *Quart. Rev. Biophys.* **16**, 89-112.

Wimley, W. C., Gawrisch, K., Creamer, T. P. & White, S. H. (1996). Direct measurement of salt-bridge solvation energies using a peptide model system: implications for protein stability. *Proc. Natl Acad. Sci. USA*, **93**, 2985-2990.

Yip, K. S., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R. & Rice, D. W. (1995). The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure*, **3**, 1147-1158.

Young, A. C. M., Dewan, J. C., Nave, C. & Tilton, R. F. (1993). Comparison of the radiation-induced decay and structure refinement from X-ray data collected from lysozyme crystals at low and ambient temperatures. *J. Appl. Crystallog.* **26**, 309-319.

Young, A. C., Tilton, R. F. & Dewan, J. C. (1994). Thermal expansion of hen egg-white lysozyme. Com-

parison of the 1.9 Å resolution structures of the tetragonal form of the enzyme at 100 K and 298 K. *J. Mol. Biol.* **235**, 302-317.

Zipp, A. & Kauzmann, W. (1973). Pressure denaturation of metmyoglobin. *Biochemistry*, **12**, 4217-4228.

Edited by B. Honig

(Received 14 August 1998; received in revised form 30 October 1998; accepted 10 November 1998)