Contributions of the Ionizable Amino Acids to the Stability of Staphylococcal Nuclease†

Alan K. Meeker,‡,|| Bertrand Garcia-Moreno E.,§ and David Shortle*,‡

Department of Biological Chemistry, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, and Department of Biophysics, The Johns Hopkins University, Baltimore, Maryland 21218

Received January 24, 1996; Revised Manuscript Received March 20, 1996⊗

ABSTRACT: To quantitate the contributions of the ionizable amino acids to the stability of the native state of staphylococcal nuclease, each of the 23 lysines, 5 arginines, 4 histidines, 12 glutamic acids, and 8 aspartic acids was substituted with both alanine and glycine. This collection of 104 mutant proteins was analyzed by guanidine hydrochloride (GuHCl) denaturation, using intrinsic tryptophan fluorescence to quantitate the equilibrium between native and denatured states. From the analysis of these data, each mutant protein's stability in the absence of denaturant (ΔG_{H_2O}) and sensitivity to changes in denaturant concentration [$m_{GuHCl} = d(\Delta G)/d[GuHCl]$] were obtained. Several general trends in these values suggest that electrostatic interactions make only a minor contribution to the net stability of this protein. For the residue pairs that form ten salt bridges and ten charged hydrogen bonds between side chains, no correlation was observed between the stability losses ($\Delta\Delta G$) accompanying alanine substitution of each member of the pair. Little or no significant correlation was found between the magnitude of the loss in stability and the local electrostatic potential calculated from the three-dimensional structure by numerical and model dependent solutions of the linearized Poisson−Boltzmann equation. The structural parameters which correlated most strongly with stability loss are measures of the extent of burial of the residue in the native structure, as was previously observed for alanine and glycine substitutions of large hydrophobic residues [Shortle et al. (1990) *Biochemistry 29*, 8033] and of the polar, uncharged residues [Green et al. (1992) *Biochemistry 31*, 5717]. These results suggest that the ionizable amino acids contribute to stability predominantly through packing and bonding interactions that do not depend on their electrostatic charge.

Over the past several years, a number of studies of mutant proteins have indicated that interactions involving the electrostatic charge of ionizable side chains, salt bridges and longer range Coulombic attraction/repulsion, contribute little to the net stability of the folded state (Akke & Forsen, 1990; Dao-pin et al., 1991a,b; Horovitz et al., 1990; Serrano et al., 1990). The observation common to most of these studies has been that replacement of one or a small number of charged residues with an uncharged residue has relatively little effect on the free energy of reversible denaturation. However, in a few instances, just the opposite observation has been reported, namely, that a neutral replacement of a charged residue resulted in surprisingly large stability losses (Anderson et al., 1990).

Since modifying the side chain of a residue affects more than its charge, it is usually not possible in such cases to ascertain what fraction of the stability loss can be attributed to electrostatic interactions and what fraction to broken hydrogen bonds, disrupted van der Waals interactions, changes in hydrophobicity, etc. Obviously, the more deeply buried the residue within the protein structure, the larger these nonelectrostatic factors are likely to be. This multiplicity of mechanisms by which a mutant residue can modify the stability of a protein presents a major obstacle to the quantitative interpretation of denaturation studies of mutant proteins (Shortle, 1992).

One strategy for dissecting the role of the various types of bonding interactions involving side chains is to analyze a large number of mutations involving one residue type situated in a variety of different structural contexts (Green et al., 1992). By comparing the quantitative effects on stability with different features of the local structure, one can search for those structural parameters that modulate the strength of that residue's interaction with its surroundings. By averaging such data, one can potentially average over many different local environments and thereby quantitate the mean contribution from all interactions. Such mean contributions can then be compared among different residue types to attempt to factor out the contributions of different bonding interactions.

For example, in the specific case of electrostatic interactions, the mean stability loss upon replacing a charged residue (e.g., aspartate) with an uncharged one (e.g., alanine) should give an upper bound to the average contribution of electrostatic interactions to stability, since such a value would include the averaged contributions of all other stabilizing interactions as well. A reasonable estimate of the average contribution from interactions other than electrostatics can be obtained from the mean stability loss upon replacing an uncharged residue of similar chemical type (e.g., asparagine) with the same mutant residue (alanine). From the difference of these two values, it should be possible to obtain a reasonable estimate of the average stability contribution from electrostatics alone.

† This research is supported by a grant from the NIH to D.S. (GM-34171).

* Corresponding author.

‡ The Johns Hopkins University School of Medicine.

§ The Johns Hopkins University.

|| Present Address: Department of Urology, The Johns Hopkins University School of Medicine, Baltimore, MD 21205.

⊗ Abstract published in *Advance ACS Abstracts,* May 1, 1996.

S0006-2960(96)00171-7 CCC: $12.00

As part of this laboratory's analysis of the physical chemical determinants of protein stability, we have systematically mutated each residue in staphylococcal nuclease to both alanine and glycine. Previously, we have reported our analysis of the large hydrophobic residues (Shortle et al., 1990) and the polar, uncharged residues, including alanine, proline, and glycine (Green et al., 1992). Here we report the completion of this study with an analysis of the stability effects that accompany alanine and glycine replacements of the five types of ionizable amino acid residues: lysine, arginine, histidine, glutamate, and aspartate.

EXPERIMENTAL PROCEDURES

Mutant Proteins. Alanine and glycine substitution mutants of staphylococcal nuclease were generated by oligonucleotide-directed mutagenesis (Kunkel, 1985) of the wild-type gene cloned into the M13 phage vector MF11. Synthetic oligonucleotides were 2-fold degenerate, specifying both the Gly codon GGN and the Ala codon GCN. After transformation into competent DH5aF′ cells, phage DNA was prepared from individual plaques, the compete nucleotide sequence of the nuclease gene was determined, and the mutant gene was transferred to the expression plasmid, pL12, which contains unique *Spe*I and *Sph*I sites (Shortle et al., 1990).

Approximately 25−50 mg of mutant protein was purified from *Escherichia coli* strain AR120 by a published method (Shortle et al., 1992). Protein concentration was determined using the extinction coefficient of 0.93 cm^{-1} mg^{-1} mL at 280 nm for a 1 mg/mL solution (Fuchs et al., 1967). Protein purity was analyzed by SDS−PAGE,[1] followed by staining with Coomassie Brilliant Blue R-250. All protein preparations were estimated to be at least 95% pure.

Denaturation Analysis. To quantitate the changes in stability for each mutant protein, the intrinsic fluorescence of the single tryptophan at position 140 was measured as a function of guanidine hydrochloride concentration at 20 °C and pH 7.0 (Shortle & Meeker, 1986). At a series of GuHCl concentrations in increments of 0.05 M, the equilibrium constant for reversible denaturation was determined from the equation $K_{app} = (I_n - I)/(I - I_d)$, where I is the measured intrinsic fluorescence, I_n is the extrapolated value of fluorescence for the native state, and I_d is the extrapolated value for the denatured state. From the equation $\Delta G = -RT \ln K_{app}$, the free energy change on denaturation, ΔG, was calculated for all values of K_{app} between 0.1 and 10. To obtain ΔG_{H_2O} and m_{GuHCl}, a straight line was fit to ΔG versus [GuHCl] by using the linear least squares method. Typically, an average of seven to nine data points were fit, giving a value of 0.9990−0.9999 for the correlation coefficient squared, r^2.

Calculation of Electrostatic Surface Potential. The electrostatic potentials were calculated on the structures of the apoenzyme (Hynes & Fox, 1991) and the enzyme complexed with Ca^{2+} and thymidine 3′,5′-bisphosphate (Loll & Lattman, 1989). Two different methods were used to compute the potentials: (1) numerical solution of the linearized Poisson−Boltzmann equation by the method of finite differences (Fine et al., 1986) using the program DELPHI by the BioSym Corporation and (2) a model dependent soution of the linearized Poisson−Boltzmann equation by the method of Tanford and Kirkwood (Garcia-Moreno E., 1994). In both sets of calculations, the values of the dielectrics of solvent and of the protein were fixed at 78.5 and 4.0, respectively, and the ionic strength was 150 mM. The pH for the calculation was 7.0. For both calculations, the state of ionization of all titratable atoms at this pH was calculated explicitly with the solvent accessibility-modified Tanford−Kirkwood algorithm (Matthew et al., 1986).

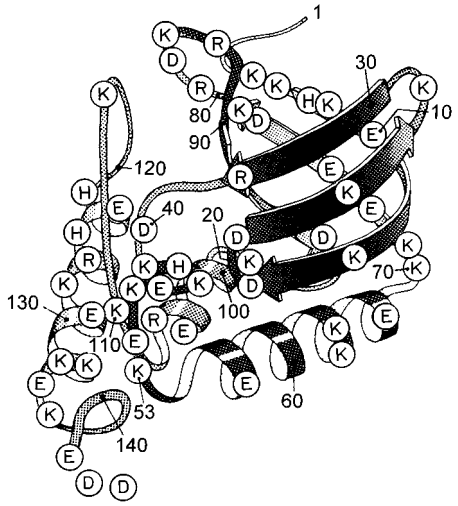

FIGURE 1: Ribbon diagram of staphylococcal nuclease showing the positions of the ionizable residues: lysine (K), arginine (R), histidine (H), glutamic acid (E), and aspartic acid (D). The approximate locations of the C_α carbons are labeled. This figure was copyrighted by Jane Richardson and is used with her permission.

Statistical Analyses. The statistical analysis program SPSS/PC+ for Windows was used to search for correlations between $\Delta\Delta G$ ($=\Delta G_{mut,H_2O} - \Delta G_{wt,H_2O}$) or m_{GuHCl} for alanine and glycine substitutions and a number of physical parameters that describe the mutated residue's position and environment within the wild-type native state using the X-ray crystal structures of the apoenzyme (Hynes & Fox, 1991) and the enzyme complexed with Ca^{2+} and thymidine 3′,5′-bisphosphate (Loll & Lattman, 1989). These parameters include φ, ψ, χ, and ω angles, solvent exposure of both the side chain and backbone atoms, thermal factors of the side chain and backbone atoms, hydrophobicity and hydrophilicity profiles using the scales of Kyte and Doolittle (1982) and Hopp and Woods (1981) averaged over windows of varying length, and the number of residues within R angstroms with R varied from 3 to 12 measuring C_α relative to C_α of the test residue as well as C_β relative to C_β of the test residue. A complete list can be found in Shortle et al. (1990).

RESULTS

Denaturation Data and Estimation of Errors. In Figure 1, the positions of the ionizable amino acids, Lys, Arg, His, Glu, and Asp, are shown in the context of a schematic diagram of the X-ray crystal structure of wild-type staphylococcal nuclease. Using oligonucleotide-directed mutagenesis, each of these 52 residues was mutated individually to both Ala and Gly, yielding a collection of 104 single mutant proteins. The results of guanidine hydrochloride denaturation of highly purified mutant proteins are shown in Table 1. For

[1] Abbreviations: SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; GuHCl, guanidine hydrochloride.

Table 1: Stability Parameters of Mutant Nucleases

mutant	C_m	m_{GuHCl}	$\Delta\Delta G$	mutant	C_m	m_{GuHCl}	$\Delta\Delta G$
D19A	0.79	1.00	−0.1	K45A	0.81	1.05	+0.3
D19G	0.74	1.00	−0.5	K45G	0.83	1.01	+0.2
D21A	0.94	0.96	+0.7	K48A	0.81	1.03	+0.1
D21G	0.89	0.96	+0.3	K48G	0.84	1.01	+0.2
D40A	0.89	0.94	+0.2	K49A	0.77	0.99	−0.3
D40G	0.76	0.97	−0.5	K49G	0.80	0.97	−0.2
D77A	0.45	0.76	−3.1	K53A	0.78	0.96	−0.4
D77G	0.60	0.79	−2.2	K53G	0.82	0.93	−0.3
D83A	0.27	0.89	−3.8	K63A	0.78	0.95	−0.5
D83G	0.44	0.94	−2.7	K63G	0.56	1.05	−1.5
D95A	0.28	1.13	−3.3	K64A	0.83	1.01	+0.1
D95G	0.37	1.11	−2.7	K64G	0.71	1.05	−0.4
D143A	0.79	1.01	−0.1	K70A	0.77	1.04	−0.1
D143G	0.79	0.98	−0.2	K70G	0.70	1.05	−0.5
D146A	0.78	0.99	−0.2	K71A	0.70	1.07	−0.4
D146G	0.78	1.01	−0.2	K71G	0.58	1.12	−1.1
				K78A	0.74	0.97	−0.6
E10A	0.6	1.03	−1.3	K78G	0.65	0.99	−1.1
E10G	0.52	1.03	−1.8	K84A	0.80	1.06	+0.2
E43A	0.85	1.00	+0.3	K84G	0.73	1.05	−0.3
E43G	0.9	0.97	+0.5	K97A	0.75	1.07	−0.1
E52A	0.80	0.98	−0.1	K97G	0.51	1.09	−1.7
E52G	0.78	0.95	−0.4	K110A	0.66	0.93	−1.3
E57A	0.80	0.97	−0.2	K110G	0.44	0.91	−2.7
E57G	0.62	0.94	−1.6	K116A	0.90	1.01	+0.7
E67A	0.64	1.02	−1.0	K116G	0.96	0.99	+1.0
E67G	0.64	1.04	−0.9	K127A	0.84	1.02	+0.2
E73A	0.59	1.01	−1.4	K127G	0.67	1.05	−0.7
E73G	0.39	1.03	−2.7	K133A	0.62	0.97	−1.4
E75A	0.62	0.78	−2.2	K133G	0.34	0.92	−3.3
E75G	0.31	0.93	−3.5	K134A	0.82	1.01	+0.1
E101A	0.68	0.77	−1.9	K134G	0.67	1.04	−0.7
E101G	0.53	0.87	−3.1	K136A	0.70	0.98	−0.9
E122A	0.78	0.97	−0.4	K136G	0.78	1.00	−0.2
E122G	0.54	0.89	−2.2				
E129A	0.51	0.86	−2.4	H8A	0.72	1.04	−0.4
E129G	0.27	0.85	−3.8	H8G	0.67	1.03	−0.8
E135A	0.69	1.02	−0.7	H46A	0.74	0.98	−0.5
E135G	0.57	0.97	−1.7	H46G	0.81	0.93	−0.4
E142A	0.76	1.01	−0.3	H121A	0.40	0.87	−3.1
E142G	0.76	1.01	−0.3	H121G	0.22	0.83	−4.2
				H124A	0.91	0.96	+0.4
K5A	0.83	1.04	+0.3	H124G	0.72	1.02	−0.5
K5G	0.82	1.04	+0.3				
K6A	0.81	1.05	+0.3	R35A	0.60	1.01	−1.4
K6G	0.83	1.02	+0.3	R35G	0.46	1.03	−2.2
K9A	0.60	0.99	−1.4	R81A	0.64	1.02	−1.1
K9G	0.53	0.99	−1.9	R81G	0.52	0.92	−2.2
K16A	0.81	0.96	−0.2	R87A	0.70	0.99	−0.9
K16G	0.68	1.03	−0.7	R87G	0.44	0.95	−2.6
K24A	0.73	1.06	−0.2	R105A	0.62	0.97	−1.4
K24G	0.51	1.23	−1.2	R105G	0.45	1.01	−2.4
K28A	0.70	1.02	−0.7	R126A	0.57	0.98	−1.7
K28G	0.67	1.04	−0.7	R126G	0.39	0.94	−2.9

the five mutants which were less than 96% native in the absence of denaturant (i.e., $\Delta G_{H_2O} < +2.0$ kcal/mol; H121A, H121G, D83A, E129G, and E75G), the intrinsic fluorescence of the native state was determined by titration of the protein with ammonium sulfate until the fluorescence reached a plateau value (Shortle et al., 1990). As noted previously, the reproducibility of ΔG_{H_2O} and m_{GuHCl} was estimated to be better than ± 0.1 kcal/mol and ± 0.02, respectively.

Lysine is the most abundant amino acid type in nuclease, occupying 23 out of the 149 positions. Five lysines are in α helices (positions 63, 64, 127, 133, and 134), three in β strands (positions 16, 24, and 71), seven in turns (positions 28, 48, 49, 53, 84, 97, and 116), and eight in irregular structures (positions 5, 6, 9, 45, 70, 78, 110, and 136). As expected from the observation that lysine is usually confined to surface positions, the average stability loss on substitution with an alanine is only -0.3 ± 0.6 kcal/mol. The range in $\Delta\Delta G$ extends from +0.7 to −1.4 kcal/mol. Glycine substitutions are in most, but not all, cases more destabilizing, leading to an average stability loss of -0.8 ± 1.0 kcal/mol with a range from +1.0 to −3.3 kcal/mol.

Mutations at six lysine positions stabilize the native conformation: 5, 6, 45, 48, 84, and 116. It may be significant that three of these positions (5, 6, and 48) are part of a lysine-lysine dipeptide; electrostatic repulsion between the two lysine side chains in the native state may be relieved by eliminating the charge from one member of the pair. Lysine 116 forms a peptide bond with proline 117 that is in the cis configuration in the native state (Loll & Lattman, 1989). Substitution of this proline with alanine or glycine leads to

a similar increase in stability (Green et al., 1992; Nakano et al., 1993), suggesting that relief of strain may explain the +1.0 kcal/mol greater stability of K116G and the slightly smaller stability increase observed for K116A. However, inspection of the wild-type structure does not reveal an obvious rationalization for the small increase in stability for mutants at positions 45 and 84.

Of the five arginines in nuclease, two are in α helices (positions 105 and 126), one is in a β strand (position 35), and two are in irregular structures (positions 82 and 87). Given this diversity in local structure, it is surprising that the spectrum of stability effects accompanying both alanine and glycine substitutions is unusually narrow. For the alanine mutants, the stability losses vary from −1.7 to −0.9 kcal/mol, with an average value of -1.3 ± 0.3 kcal/mol. For the glycine mutants, the range extends from −2.9 to −2.2 kcal/mol, with an average value of -2.5 ± 0.3 kcal/mol.

The four positions occupied by histidine, of which two are helical (positions 121 and 124), one is in a turn (position 46), and one is in an irregular structure (position 8), show a much wider range of effects. Alanine substitution at position 124 is +0.4 kcal/mol more stable, an observation consistent with earlier findings that leucine and arginine (Shortle, 1985) plus a number of other residues at this position (W. Kelly, W. Stites, and D. Shortle, unpublished observations) are more stable than wild-type. In contrast, the alanine substitution for histidine at position 121 destabilizes the protein by −3.1 kcal/mol. The average stability change is -0.9 ± 1.5 kcal/mol for the alanine mutants and -1.5 ± 1.8 kcal/mol for the glycine mutants. Since their pK_a values vary between 5.49 and 6.82 (Alexandrescu et al., 1988), the fractional charge of these residues will be less than one-half at pH 7.0.

There are twelve glutamic acid residues in nuclease: six in α helices (positions 57, 67, 101, 122, 129, and 135), three in β strands (positions 10, 73, and 75), one in a turn (position 52), and two in irregular structures (positions 43 and 142). At one of these positions (position 43), both alanine and glycine substitutions were more stable than wild-type nuclease. This residue plays a major role at the active site, and substitution of serine, asparagine, and aspartic acid has been shown previously to be more stable (Serpersu et al., 1989). All other mutants at positions of Glu residues were less stable; the average $\Delta\Delta G$ for the alanine substitutions was -1.0 ± 0.9 versus -1.8 ± 1.4 kcal/mol for the glycine substitutions.

Consistent with its general distribution in proteins, the eight aspartic acid residues are not found in regular secondary structures; four occupy positions in turns (positions 19, 21, 83, and 95), and four are in irregular structures (positions 40, 77, 143, and 146). The carboxyl groups of D21 and D40 along with the backbone carbonyl of T41 form a calcium binding site, and D19 and E43 are positioned nearby. In view of possible repulsive interactions within this cluster of four carboxylate groups, it is not surprising that alanine substitutions at three of these positions (D21, D40, and E43) are more stable than wild-type. Alternatively, the increase in stability may be a consequence of improved interactions in the active site cleft, a region of nuclease that has presumably not been optimized for protein stability. Numerous examples of mutations of active site residues leading to an increase in stability have been published (Serpersu et al., 1989).

Unlike most other residues, the average $\Delta\Delta G$ on substitution of aspartic acid with alanine is essentially the same as that with glycine; $\Delta\Delta G = -1.2 \pm 1.8$ versus −1.1 kcal/mol, respectively. For mutations of staphylococcal nuclease involving all other residue types except asparagine, the average $\Delta\Delta G_{X\rightarrow Gly}$ is at least 0.5 kcal/mol more negative than that of the corresponding Ala substitution. Part of this atypical behavior can be accounted for by the fact that four Asp and three Asn residues in the native structure adopt $\varphi-\psi$ angles outside the normally allowed regions of the Ramachandran plot. As discussed by Stites et al. (1993), glycine substitutions for residues outside α helical and β sheet regions have values of $\Delta\Delta G$ that are much closer to that of alanine substitutions than at residues within allowed $\varphi-\psi$ angles, suggesting that the glycine may permit the relief of backbone strain energy present in the wild-type native state.

Statistical Correlations between $\Delta\Delta G$ and Descriptors of Residue Environment. Except for the amino and carboxyl termini, each of the charged groups in nuclease has been changed in this collection of mutants. By examining the effects of replacing a charged side chain with the smaller, uncharged methyl group of alanine, one can estimate an upper bound to the stability contribution of Coulombic interactions involving individual charges. As mentioned above, the average stability loss accompanying substitution of lysine is only −0.3 kcal/mol. For the four other residue types, arginine, histidine, glutamate, and aspartate, the average stability loss is somewhat larger, ranging from −0.9 to −1.3 kcal/mol. These values are roughly comparable to the average stability effects for alanine substitution of the polar uncharged residues (Green et al., 1992): serine (three positions), +0.3 kcal/mol; threonine (nine positions), −1.0 kcal/mol; glutamine (six positions), −0.2 kcal/mol; and asparagine (six positions), −1.7 kcal/mol.

From the X-ray crystal structure of nuclease, seven pairs of oppositely charged residues can be identified in which the side chains form hydrogen bonds (Loll & Lattman, 1988; Hynes & Fox, 1990): Lys9-Glu75, Arg35-Asp21, Arg35-Asp40, Lys63-Glu67, Arg87-Asp83, Arg105-Glu135, and Arg126-Glu122. (In addition, significant ionic interactions are expected between Lys28-Glu10 and Lys49-Glu52 and Lys133-Glu129 whose β carbons are less than 7 Å apart. For these three pairs, the lysine amino group either is not well-defined or participates in intermolecular contacts in the crystal.) Mutation of either member of a pair should eliminate all stabilizing interactions that arise from salt bridge effects, namely, the Coulombic attraction between opposite charges plus the release of bound water molecules upon pair formation. If such effects dominate the stability contribution made by each of these charged residues, one would expect a similar loss in stability for each of the two alanine substitutions. In nine out of ten cases, the $\Delta\Delta G$ values for the two alanine substitutions differ by at least 0.5 kcal/mol; in two cases, the difference between the two values of $\Delta\Delta G$ is 2 kcal/mol or greater. As shown in Figure 2A, there is no significant correlation between the stability losses upon alanine substitution of the positively charged residue versus alanine substitution of the negatively charged residue.

If the stability loss on mutation of one member of an ion pair was a consequence of forming an uncompensated charge, which may in some cases be buried below the protein surface, one would expect a significant correlation between $\Delta\Delta G$ for the alanine substitution of one residue and the extent of

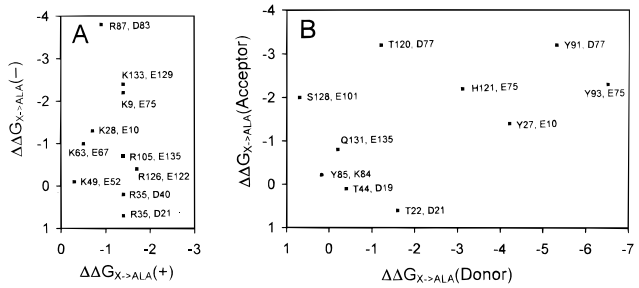

FIGURE 2: Scatter plots of the stability change accompanying alanine substitution of each member of a pair of interacting residues. $\Delta\Delta G$ is expressed in kilocalories per mole. (A) Residue pairs that form salt bridges in the native state of nuclease. (B) Residues that form hydrogen bonds, with one member having an ionizable side chain.

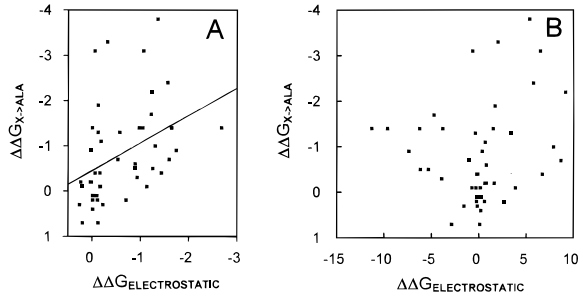

FIGURE 3: Scatter plots showing the correlation between the change in $\Delta\Delta G$ calculated for removing a charge based on the local electrostatic potential versus the experimentally observed $\Delta\Delta G$ for mutation to alanine. Units are kilocalories per mole. (A) $\Delta\Delta G_{\text{electrostatic}}$ is calculated by a model dependent solution of the linearized Poisson−Boltzmann equation (Garcia-Moreno E., 1994). (B) $\Delta\Delta G_{\text{electrostatic}}$ is calculated by solving the linearized Poisson−Boltzmann equation numerically using the method of finite differences.

buriedness of the other member of the pair, as measured by the number of C_β in a 10 Å sphere or by fractional exposure of the side chain to solvent. No such correlations could be demonstrated. Nevertheless, as described in the next section, a significant correlation is found between $\Delta\Delta G$ for the alanine substitution of these 20 residues and its own extent of burial in the wild-type structure.

The electrostatic free energy of a charge is given by the product of the charge times the electrostatic potential at the site. The removal of charged groups from the structure should be destabilizing or stabilizing by this amount, depending on whether the charge experiences a net electrostatic potential of the opposite or the same polarity. There is a very weak correlation between $\Delta\Delta G_{X\rightarrow Ala}$ and the surface electrostatic potential calculated by the model dependent solution of the Poisson−Boltzmann equation according to Tanford and Kirkwood (Figure 3A). When the more rigorous finite difference method is used to solve the Poisson−Boltzmann equation however, the calculated change in electrostatic free energy does not show a correlation with changes in stability (Figure 3B). This suggests that neither direct Coulombic effects nor differences in solvation of ionizable residues between the native and denatured states represent a significant fraction of a typical charged residue's contribution to stability.

In a search to find those factors that are important in determining the widely varying stability contributions of the ionizable residues, a large number of parameters describing residue environment were correlated with $\Delta\Delta G$ for the

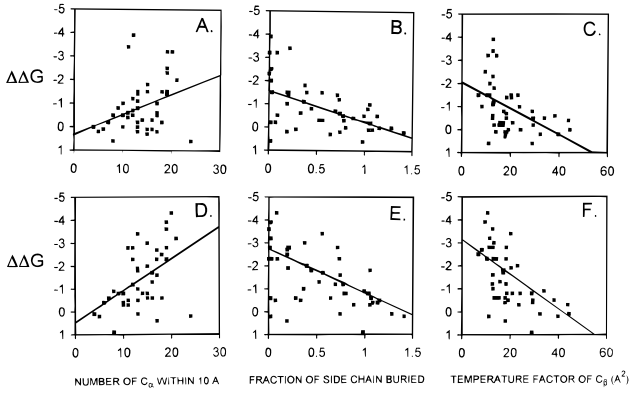

FIGURE 4: Scatter plots of $\Delta\Delta G$ (in kilocalories per mole) for mutant proteins versus three parameters that describe the environment surrounding that position in the wild-type native structure: (A−C) Alanine substitutions and (D−F) glycine substitutions. (A and D) The number of C_α carbons within a 10 Å sphere centered on the C_α of the mutated residue. (B and E) The fractional solvent exposure of the side chain compared to that in an Ala-X-Ala peptide. (C and F) The crystallographic temperature factor of the C_β residue. For A, $r = 0.35$ and $p = 0.014$. For B, $r = -0.55$ and $p < 0.001$. For C, $r = -0.38$ and $p = 0.008$. For D, $r = 0.55$ and $p = 0.002$. For E, $r = -0.56$ and $p < 0.0001$. For F, $r = -0.57$ and $p < 0.0001$.

alanine and glycine substitutions (Shortle et al., 1990). This analysis was conducted on the three groups large enough to give statistically significant results: the 23 lysine residues alone, the 12 glutamic acids alone, and all 48 positions grouped together [four of the 52 wild-type residues do not occupy defined positions in the X-ray crystal structure (Loll & Lattman, 1989; Hynes & Fox, 1991)].

As found previously for the large hydrophobic (Shortle et al., 1990) and the polar, uncharged (Green et al., 1992) residues, the only parameters describing the environment of a charged residue at position X that correlated with $\Delta\Delta G_{X\rightarrow Ala}$ and $\Delta\Delta G_{X\rightarrow Gly}$ were ones that either measure the extent of residue burial within the folded structure or correlate with the extent of burial. The three most prominent of these are (1) the number of α carbons within a 10 Å sphere surrounding the α carbon of residue X, (2) the fraction of the residue side chain that is buried, and (3) the temperature factor of backbone atoms at position X. As shown in Figure 4, the correlation coefficient r varies from 0.25 to 0.57, with modest to high levels of statistical significance. In each case, the correlation is strongest with the glycine substitutions. On average, the significant trends present in the group of 48 structured ionizable residues, shown in Figure 4, appear to be present for both the lysine and the glutamic acid subgroups (data not shown).

The strong correlation between residue burial and $\Delta\Delta G$ for the glycine substitutions suggests that packing interactions involving the proximal end of the side chain, not the distal end which carries the ionizable atom, may be the mechanism responsible for this major contribution to stability. Consistent with this conclusion is the observation that the correlation with the solvent accessibility of the non-carbon atoms that carry the electrostatic charge ($r = -0.38$ and $p = 0.013$) is no better than the correlation with the extent of burial of the C_α atom ($r = -0.35$ and $p = 0.014$). This weak correlation between the loss of stability on Ala substitution and the accessibility to solvent of the ionizable atoms suggests that much of the ionizable residues' contribution to stability, on average, is unrelated to differences in the extent of solvation

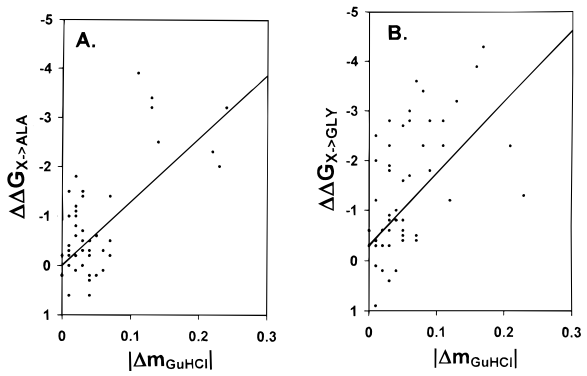

FIGURE 5: Scatter plots of $\Delta\Delta G$ (in kilocalories per mole) versus $|\Delta m_{GuHCl}|$, the absolute value of the change in m_{GuHCl} between mutant and wild-type: (A) alanine substitutions and (B) glycine substitutions. $|\Delta m_{GuHCl}|$ is normalized relative to the wild-type value of m_{GuHCl}, 6.85 kcal/mol per molar GuHCl.

of charged groups in the native and denatured states.

Statistical Correlations Involving m_{GuHCl} and Descriptors of Residue Environment. As discussed in the introductory section, in two previous studies of Ala and Gly substitutions in nuclease, the absolute value of the change in m_{GuHCl} was found to show a highly significant correlation with $\Delta\Delta G$ for a number of different groups of mutations. This same correlation can also be demonstrated for this data set. For the 12 mutations involving glutamate side chains, the correlation between $\Delta\Delta G_{Glu\rightarrow Ala}$ and $|\Delta m_{GuHCl}|$ is 0.781 ($p = 0.003$); for $\Delta\Delta G_{Glu\rightarrow Gly}$ and $|\Delta m_{GuHCl}|$, $r = 0.694$ ($p = 0.012$). Although for the lysine sites, $\Delta\Delta G_{Lys\rightarrow Ala}$ does not show this correlation, $r = 0.014$ ($p = 0.95$), a very weak correlation may exist between $\Delta\Delta G_{Lys\rightarrow Gly}$ and $|\Delta m_{GuHCl}|$, $r = 0.421$ ($p = 0.046$).

When mutations of all five residue types are grouped together, these two correlations become highly significant. For the mutations to alanine, the stability change correlates with $|\Delta m_{GuHCl}|$, with $r = 0.633$ ($p < 0.0001$); for the mutations to glycine, $r = 0.555$ ($p < 0.0001$). The scatter plots and regression line for these analyses are shown in Figure 5. When $|\Delta m_{GuHCl}|$ is treated as the independent variable, the slopes of these two regression lines are 12.0 kcal/mol per *m* unit (normalized to a value of 1.0 for wild-type) for the alanine mutant set and 13.8 kcal/mol per *m* unit for the set of glycine mutants. These values are in good agreement with each other and with values previously obtained for mutants involving hydrophobic (Shortle et al., 1990; Shortle, 1995) and polar, uncharged side chains (Green et al., 1992).

As argued previously (Shortle et al., 1990; Green et al., 1992; Shortle, 1995), this correlation suggests that the phenomenon responsible for changes in *m* values appears to account for approximately half of the stability loss for a typical nuclease mutant. When combined with the postulate that changes in *m* values reflect structural changes in the denatured state, it follows that changes in the structure (and therefore the free energy) of the denatured state are energetically quite important for an average mutation in staphylococcal nuclease.

DISCUSSION

Several mutational studies have reached the conclusion that the electrostatic charge on ionizable residues does not play a major role in protein stability. The present work reaches this same conclusion from a slightly different perspective, one based on comparative analysis of data involving every residue in a single protein. The approach taken is statistical; by averaging the stability data for alanine substitutions for a particular residue type, one can potentially average over many different local environments to obtain an estimate of that residue's mean contribution to stability from all the types of interactions with which it is involved. Obviously, this mean value can be viewed as an upper limit on the contribution from any one type of interaction. In addition, such mean values can then be compared among different residue types to attempt to factor out the roles of different bonding interactions.

The trends in the data presented above which support only a minor role of electrostatic interactions in protein stability can be summarized as follows.

The average stability loss upon truncating a residue's side chain beyond the β carbon is not significantly larger for ionizable residues (as a group) than for polar residues (as a group). While the average stability loss $\Delta\Delta G_{ave}$ for glutamic acid residues is greater than for the corresponding glutamines (-1.0 versus -0.16 kcal/mol), $\Delta\Delta G_{ave}$ is essentially the same for aspartic acid and asparagine (-1.5 versus -1.6 kcal/mol). This result suggests the average contribution of the Coulombic charge of aspartic acid to protein stability is very small indeed.

No correlation can be demonstrated between the values of $\Delta\Delta G_{X\rightarrow Ala}$ for the two residues that form seven (and possibly ten) ionic pairs in staphylococcal nuclease (Figure 2A). In addition, for eight of the ten charged residue pairs, the stability loss of the most stable ALA substitution is less than 1 kcal/mol. These findings suggest that the "salt bridge" between ionic groups (i.e., Coulombic attraction plus changes in ion hydration on close approach) does not represent the dominant (i.e., >50%) free energy term for these residues and that the average contribution of a salt bridge is likely to be significantly less than 1 kcal/mol. Since no correlation is found between $\Delta\Delta G_{X\rightarrow Ala}$ of one residue of the pair and the extent of burial of the other residue, it appears that the creation of an uncompensated charge is probably not the principal mechanism for the stability loss accompanying single mutations.

Similarly, no correlation can be demonstrated between the two values of $\Delta\Delta G_{X\rightarrow Ala}$ for the ten side chain−side chain hydrogen bonds involving one charged residue (Figure 2B). This suggests that, on average, hydrogen bonds involving charged groups make at best a small contribution to protein stability. For these ten interactions, the average loss in stability for the mutation with the smallest $\Delta\Delta G$ (which can be seen as an upper limit for the average contribution of the hydrogen bond) is -0.8 kcal/mol, whereas the most dramatic loss (E75A) is -2.3 kcal/mol. (The value for D77A has been divided by 2, since it is involved in two hydrogen bonds.) This small value is in marked contrast to the figure of -3.5 to -4.5 kcal/mol estimated in an early mutational study of the role of charged hydrogen bonds in substrate binding to an enzyme (Fersht, 1987; Rose & Wolfenden, 1993).

Only a weak correlation can be demonstrated between $\Delta\Delta G_{X\rightarrow Ala}$ and the electrostatic free energy change upon removing the charge, as calculated according to Poisson−Boltzmann electrostatics. In qualitative agreement with the experimental observations, the calculations by two different

methods predict that at pH 7.0 and at an ionic strength of 150 mM most of the ionizable groups are in net stabilizing electrostatic environments but a few groups (K16, D19, D21, and E43) sense net destabilizing electrostatic environments. The magnitude of the electrostatic $\Delta\Delta G$ predicted with the finite difference Poisson−Boltzmann calculations can be as high as 5 times the measured $\Delta\Delta G_{X\rightarrow Ala}$.

Finally, no correlation can be demonstrated between $\Delta\Delta G_{X\rightarrow Ala}$ and the local hydrophobicity averaged over windows of three, five, seven, nine, or eleven residues. This finding is counter to the hypothesis that charged residues may contribute to stability primarily by restricting the conformations occupied in the denatured state in ways which prevent hydrophobic residues from forming non-native interactions or clusters.

The major mechanism by which ionizable side chains stabilize the native state at neutral pH appears to come from packing interactions involving the proximal end of the side chain. The following three trends in the data support this conclusion.

As with hydrophobic residues and polar, uncharged residues, the structural parameters which best correlate with $\Delta\Delta G_{X\rightarrow Ala}$ are measures of the extent of residue burial in the native state.

This correlation is consistently stronger for glycine substitutions, which remove more of the side chain, than for alanine.

Residue burial as measured by the solvent inaccessibility or number of surrounding neighbors of C_α correlates with $\Delta\Delta G$ as well as do measures of the burial of just the charge-bearing atoms.

In view of the large energy associated with charge/charge interactions, it is surprising that such interactions appear to make only a minor contribution to the *net* stability of the native state. Presumably, these large energies must be offset by counterbalancing interactions of equivalent or nearly equivalent strength. Either the entropy loss on forming a charged interaction within the native state is large and mostly cancels an otherwise favorable interaction energy, or Coulombic interactions are approximately equally favorable in the denatured state.

Overall, these data suggest that stability considerations play a minor role in the selection of sites occupied by charged residues during evolution. Instead, interactions responsible for biological function (e.g., active sites and protein/protein interactions) and solubility would appear to play a greater role in the appearance and conservation of specific lysine, arginine, histidine, glutamate, and aspartate residues at unique positions in a family of homologous proteins.

REFERENCES

Akke, M., & Forsen, S. (1990) *Proteins: Struct., Funct., Genet. 8*, 23−29.

Alexandrescu, A. T., Mills, D. A., Ulrich, E. L., Chinami, M., & Markley, J. L. (1988) *Biochemistry 27*, 2158−2165.

Anderson, D. E., Becktel, W. J., & Dahlquist, F. W. (1990) *Biochemistry 29*, 2403−2408.

Dao-pin, S., Soderlind, E., Baase, W. A., Wozniak, J. A., Sauer, U., & Matthews, B. W. (1991a) *J. Mol. Biol. 221*, 873−887.

Dao-pin, S., Sauer, U., Nicholson, H., & Matthews, B. W. (1991b) *Biochemistry 30*, 7142−7153.

Fersht, A. R. (1987) *Trends Biochem. Sci. 12*, 301−304.

Fuchs, S., Cuatrecasas, P., & Anfinsen, C. B. (1967) *J. Biol. Chem. 242*, 4768−4770.

Garcia-Moreno E., B. (1994) *Methods Enzymol. 240*, 645−667.

Green, S. M., Meeker, A. K., & Shortle, D. (1992) *Biochemistry 31*, 5717−5728.

Horovitz, A., Serrano, L., Avron, B., Bycroft, M., & Fersht, A. R. (1990) *J. Mol. Biol. 216*, 1031−1044.

Hynes, T. R., & Fox, R. O. (1991) *Proteins: Struct., Funct., Genet. 10*, 92−105.

Klapper, I., Hagstrom, R., Fine, R., Sharp, K., & Honig, B. (1986) *Proteins: Struct., Funct., Genet. 1*, 47.

Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488−492.

Loll, P. J., & Lattman, E. E. (1989) *Proteins: Struct., Funct., Genet. 5*, 183−201.

Matthew, J. B., Gurd, F. R. N., Garcia-Moreno E., B., Flanagan, M. A., March, K. L., & Shire, S. J. (1985) *CRC Crit. Rev. Biochem. 18*, 91−197.

Nakano, T., Antonino, L. C., Fox, R. O., & Fink, A. L. (1993) *Biochemistry 32*, 2534−2541.

Rose, G. D., & Wolfenden, R. (1993) *Annu. Rev. Biophys. Biomol. Struct. 22*, 381−415.

Serpersu, E. H., Hibler, D. W., Gerlt, J. A., & Mildvan, A. S. (1989) *Biochemistry 28*, 1539−1548.

Serrano, L., Horovitz, A., Avron, B., Bycroft, M., & Fersht, A. R. (1990) *Biochemistry 29*, 9343−9352.

Shortle, D. (1985) *J. Cell. Biochem. 30*, 281−289.

Shortle, D., & Meeker, A. K. (1986) *Proteins: Struct., Funct., Genet. 1*, 81−89.

Shortle, D., Stites, W. E., & Meeker, A. M. (1990) *Biochemistry 29*, 8033−8041.

Shortle, D. (1992) *Q. Rev. Biophys. 25*, 205−250.

Shortle, D. (1995) *Adv. Protein Chem. 46*, 217−247.

Stites, W. E., Meeker, A. K., & Shortle, D. (1994) *J. Mol. Biol. 235*, 27−32.

BI960171+