Vincent, J.-P., Chicheportiche, R., & Lazdunski, M. (1971) *Eur. J. Biochem. 23*, 401.
Walls, F. C., Baldwin, M. A., Falick, A. M., Gibson, B. W., Gillece-Castro, B. L., Kaur, S., Maltby, D. A., Medzihradszky, K. F., Evans, S., & Burlingame, A. L. (1990) in *Biological Mass Spectrometry* (Burlingame, A. L., McCloskey, J. A., Eds.) p 197, Elsevier, Amsterdam.
Weissman, J. S., & Kim, P. S. (1991) *Science 253*, 1386.
Zahler, W. L., & Cleland, W. W. (1968) *J. Biol. Chem. 243*, 716.

Contributions of the Polar, Uncharged Amino Acids to the Stability of Staphylococcal Nuclease: Evidence for Mutational Effects on the Free Energy of the Denatured State†

Susan M. Green, Alan K. Meeker, and David Shortle*
Department of Biological Chemistry, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205

Received December 24, 1991; Revised Manuscript Received April 8, 1992

ABSTRACT: In order to quantitate the contributions of the polar, uncharged amino acids to the stability of the native state of staphylococcal nuclease, each of the 13 alanines, 9 glycines, 9 threonines, 6 prolines, 6 glutamines, 6 asparagines, and 3 serines was substituted, either with both alanine and glycine or with 1 of these 2 amino acids plus valine. For each mutant, the stability to reversible denaturation (ΔG_{H_2O}) was quantitated by determining the K_{app} for this reaction as a function of guanidine hydrochloride concentration. In addition, the parameter m_{GuHCl} (=d(ΔG)/d[GuHCl]) was calculated from the data. To identify the local structural features responsible for the relatively large and variable changes in ΔG_{H_2O} and m_{GuHCl} observed for the same type of substitution at different locations in nuclease, statistical correlations were sought between ΔG_{H_2O}, m_{GuHCl}, and a number of descriptors of the local structure. As with substitutions of the large hydrophobic amino acids [Shortle, D., Stites, W. E., & Meeker, A. K. (1990) *Biochemistry 29*, 8033–8041], mutation of polar, uncharged residues to Gly leads to a change in stability that, on average, correlates well with the degree to which the wild-type residue is buried. This correlation is especially significant for threonine, an amino acid with both polar and hydrophobic character, but is not demonstrated for the more typically hydrophobic residue alanine. As reported in the previous study of alanine/glycine substitutions of hydrophobic residues, a significant correlation between changes in stability and changes in the value of m_{GuHCl} is again observed, strengthening the conclusion that the putative structural changes in the denatured state which lead to increases or decreases in m_{GuHCl} are responsible for a significant fraction of the stability loss for an average mutant. The existence of this correlation is consistent with the denatured state of wild-type staphylococcal nuclease having evolved to a relatively high free energy via optimization of a balance between a maximal exposure of hydrophobic surface and a minimal gain in chain entropy. On average, mutations are less stable in proportion to the extent of which they perturb this balance. A new and puzzling correlation is reported between the extent of buriedness of a residue in the wild-type native state versus the difference in m_{GuHCl} between the Ala mutation and the Gly mutation at that position.

Analysis of mutant proteins with single amino acid substitutions provides a general strategy for quantitating the contributions of individual residues to protein folding and stability. While the actual construction and isolation of mutant proteins are often straightforward, the study of how they differ from wild-type and the quantitative interpretation of any differences are slow and uncertain processes. Perhaps the greatest obstacle to interpretation of such data is the multiplicity of mechanisms by which a mutant residue could potentially alter the free energy of the reaction. On the native side of the equation, a mutation could alter side-chain hydrophobicity, steric size and shape, electrostatic charge, α-helix and β-strand propensities, hydrogen-bonding potential, etc. On the other side of the equation, changes in chain entropy and hydrophobicity or electrostatic charge could modify the free energy of the denatured state.

One strategy for dissecting out the roles of these various interactions is to analyze several different substitutions at a site of interest and then correlate changes in stability with one or more indices that reflect the physicochemical properties of the different amino acid side chains. In this way, the most significant stabilizing interactions involving the wild-type residue at one position can sometimes be inferred (Alber et al., 1987, 1988; Yutani et al., 1987; Matsumura et al., 1988, 1989; Pakula & Sauer, 1990; Connelly et al., 1991; Stites and Shortle, unpublished data). A second, complementary strategy is to mutate each residue of a particular amino acid type one at a time and then search for those features of the local structure that best correlate with the stability change at each of the sites (Shortle et al., 1990). Whereas the first approach is designed to identify those side-chain features that contribute to stability at one particular site, the second attempts to identify those features of the local environment surrounding a class of residues which interact significantly with the wild-type side chain and thereby determine what its contribution to stability will be.

In a previous report, we described the application of this second strategy to the large hydrophobic amino acids in the small protein staphylococcal nuclease (Shortle et al., 1990).

† This work was supported by NIH Grant GM34171 (D.S.) and by NIH Postdoctoral Fellowship GM14306 (S.M.G.).

0006-2960/92/0431-5717$03.00/0

All Leu, Ile, Val, Phe, Tyr, and Met residues were mutated to Ala and to Gly, and effects on the free energy of reversible denaturation (ΔG_{H_2O}) and the rate of change of free energy with respect to guanidine hydrochloride concentration (m_{GuHCl}) were measured. A search for statistical correlations between changes in these two values and a large number of local structural parameters revealed several significant trends in the data. (1) The average stability loss on truncating these hydrophobic side chains was 2–3 times greater than that estimated on the basis of the free energy of transfer of the hydrophobic side chain from water to 1-octanol. (2) The parameter that best estimates the effect of the local structure on $\Delta\Delta G$ was the number of C_α carbons within a sphere of 10-Å radius. (3) A bipartite distribution was found for sites of mutations that altered m_{GuHCl}; mutations that increased this parameter involved residues that contribute side chains to the major hydrophobic core centered around a five-strand β-barrel, whereas mutations that caused m_{GuHCl} to decrease clustered around a second, smaller, and less-defined hydrophobic core. (4) A significant correlation was found between the absolute change in m_{GuHCl} and the loss of stability. On the basis of earlier work that suggested some, if not all, changes in m_{GuHCl} are manifestations of changes in the residual structure of the denatured state (Shortle & Meeker, 1986, 1989; Shortle et al., 1988, 1989), this observation was used to argue that amino acid substitutions can destabilize a protein indirectly via their effects on the free energy of the denatured state.

In this report, we extend our analysis of the contributions of the various amino acid residue types to the stability of the native conformation of staphylococcal nuclease by constructing and analyzing Ala and Gly (or Val) substitutions for the seven polar, uncharged amino acids, which include alanine, glycine, proline, serine, threonine, asparagine, and glutamine.

Experimental Procedures

Recombinant DNA. All mutants of staphylococcal nuclease were generated through oligonucleotide-directed mutagenesis according to the Kunkel method (1985) and cloned into the M13 phage vector MFO9. Each oligonucleotide used for mutagenesis was 2-fold degenerate, specifying the Ala codon GCN and the Gly codon GGN. [At sites of wild-type Ala, the oligonucleotide specified Gly and Val (GTN) codons; at sites of wild-type Gly, it specified Ala and Val codons.] After transformation into competent DH5αF′ cells, plaques were picked, the phage DNA was sequenced, and the nuclease gene was moved into the expression plasmid pL12 which contains unique *Spe*I and *Sph*I sites (Shortle et al., 1990). To ensure that no spurious secondary mutation had been introduced, the entire gene was sequenced for each mutant.

Protein Isolation. Mutant proteins were purified from *Escherichia coli* strain AR120 carrying the mutant gene on plasmid pL12 (Shortle et al., 1990). 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, followed by staining with Coomassie Brilliant Blue R-250. All protein preparations were estimated to be at least 95% pure.

Guanidine Hydrochloride Denaturation. To quantitate the changes in stability due to a substitution mutation, 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). Briefly, the apparent equilibrium constant, K_{app}, for reversible denaturation, assuming a two-state model, was determined for a series of GuHCl concentrations incremented in steps of 0.05 M by using 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 7–9 data points were fit, giving a value of 0.9990–0.9999 for r^2. For those mutants with a ΔG_{H_2O} of less than +1.4 kcal/mol, I_n was calculated by the best linear fit of ΔG versus [GuHCl] and also by renaturation with ammonium sulfate. The values obtained by the two methods agreed within ±2% (Shortle et al., 1990).

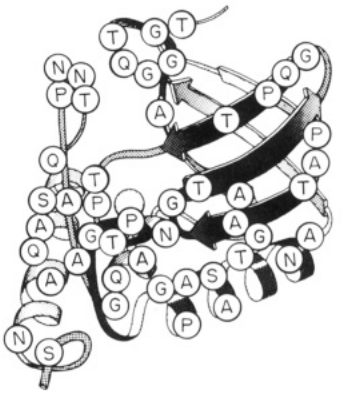

FIGURE 1: Ribbon diagram showing the positions of the uncharged, polar residues altered in mutant forms of staphylococcal nuclease. The approximate location of the C_α carbon is labeled using the one-letter code. This figure was drawn and copyrighted by Jane Richardson (1981) and is used with her permission.

Statistical Analyses. The statistical analysis program SPSS/PC+ version 3.0 was used to search for correlations between $\Delta\Delta G$ ($=d\Delta G_{mut,H_2O} - \Delta G_{wt,H_2O}$) or m_{GuHCl} for alanine, glycine, and valine substitutions and a number of physical parameters that describe that 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 included ϕ, ψ, χ and ω angles, solvent exposure of both the side chain and backbone, thermal factors of the side chain and backbone, hydrophobicity and hydrophilicity profiles using the scales of Kyte and Doolittle (1982) and Hopp and Woods (1981), and the number of residues within R Å 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 (Rose & Roy, 1980). A complete list can be found in Shortle et al. (1990).

Results

Denaturation Data and Estimation of Errors. In Figure 1, the positions of all the uncharged, polar amino acids in wild-type staphylococcal nuclease are shown relative to a schematic diagram of the 3D structure. Using oligonucleotide-directed mutagenesis, all Thr, Pro, Asn, Gln, and Ser residues were mutated individually to both Ala and Gly. The Ala residues were mutated to both Gly and Val, and the Gly residues were mutated to both Ala and Val. A total of 104 mutant forms of nuclease were constructed and analyzed.

Table I shows the results of guanidine hydrochloride denaturation of each of the mutant proteins. Only one of these mutants (G107V) was too unstable to yield an accurate determination of ΔG_{H_2O}; even in the presence of high concentrations of ammonium sulfate, only a small fraction of molecules were in the native state. For three other mutants

Table I

mutant	C_m[a]	m_{GuHCl}[b]	ΔG_{H_2O}[c]	mutant	C_m[a]	m_{GuHCl}[b]	ΔG_{H_2O}[c]
WT	0.80	1.00	5.5	G88V	0.86	0.78	4.6
A12G	0.46	0.99	3.1	G96A	0.53	1.02	3.7
A12V	0.70	0.92	4.4	G96V	0.25	1.08	1.8
A17G	0.90	0.86	5.2	G107A	0.16	0.93	1.0
A17V	0.51	1.05	3.6	G107V			<−2.0
A58G	0.44	0.98	2.9	N68G	0.77	0.96	5.0
A58V	0.44	0.89	2.7	N68A	0.76	0.97	5.0
A60G	0.60	0.99	4.1	N100G	0.09	0.71	0.4
A60V	0.44	0.88	2.7	N100A	0.05	0.8	0.3
A69G	0.48	1.05	3.5	N118G	0.60	0.88	3.6
A69V	0.32	1.07	3.2	N118A	0.55	0.92	3.4
A90G	0.53	0.96	3.5	N119G	0.70	0.88	4.2
A90V	0.84	0.92	5.3	N119A	0.68	0.91	4.2
A94G	0.41	1.12	3.1	N138G	0.95	0.87	5.6
A94V	0.66	0.94	4.2	N138A	0.67	0.97	4.4
A102G	0.59	1.03	4.2	N144G	0.78	0.93	5.0
A102V	0.54	0.98	3.6	N144A	0.79	1.0	5.4
A109G	0.66	0.99	4.5	Q30G	0.66	1.01	4.6
A109V	0.43	0.81	2.4	Q30A	0.77	0.99	5.2
A112G	0.86	0.93	5.5	Q80G	0.64	0.95	4.1
A112V	0.67	0.88	4.0	Q80A	0.81	0.97	5.4
A130G	0.67	0.96	4.4	Q106G	0.58	1.01	4.0
A130V	0.66	0.96	4.3	Q106A	0.86	0.95	5.6
A132G	0.29	0.89	1.8	Q123G	0.76	0.95	4.9
A132V	0.12	0.86	0.7	Q123A	0.83	0.9	5.1
A145G	0.80	0.94	5.1	Q131G	0.49	0.92	3.1
A145V	0.79	0.96	5.2	Q131A	0.80	0.98	5.3
P11G	0.67	0.99	4.5	Q149G	0.78	0.95	5.1
P11A	0.75	1.00	5.1	S59G	0.66	0.97	4.4
P31G	0.59	0.98	3.9	S59A	0.94	0.92	5.9
P31A	0.77	0.95	5.0	S128G	0.59	0.99	3.9
P42G	0.84	0.89	5.1	S128A	0.96	0.96	6.2
P42A	0.84	0.98	5.6	S141G	0.69	0.98	4.6
P47G	0.87	0.92	5.4	S141A	0.75	1.0	5.1
P47A	0.79	0.92	4.9	T4G	0.79	0.99	5.3
P56G	0.72	0.92	4.5	T4A	0.78	0.99	5.3
P56A	0.84	0.95	5.5	T13G	0.62	1.03	4.4
P117G	1.00	0.94	6.4	T13A	0.69	1.02	4.8
P117A	0.99	0.94	6.3	T22G	0.41	1.1	3.1
G20A	0.81	0.94	5.2	T22A	0.54	1.07	3.9
G20V	0.50	0.93	3.2	T33G	0.40	1.10	3.0
G29A	0.57	1.09	4.2	T33A	0.56	1.07	4.1
G29V	0.33	1.11	2.5	T41G	0.57	0.9	3.5
G50A	0.81	1.04	5.7	T41A	0.82	0.98	5.5
G50V	0.67	0.96	4.4	T44G	0.78	0.93	4.9
G55A	0.76	0.95	4.9	T44A	0.79	0.96	5.1
G55V	0.63	0.86	3.7	T62G	0.27	1.08	2.0
G79A	0.51	0.87	3.1	T62A	0.51	0.88	3.1
G79V	0.50	0.89	3.1	T82G	0.50	1.01	3.5
G86A	0.79	0.92	4.9	T82A	0.67	1.01	4.6
G86V	0.25	1.07	1.8	T120G	0.56	0.9	3.4
G88A	0.87	0.92	5.4	T120A	0.66	0.94	4.3

[a] Midpoint concentration of GuHCl in molar. [b] Units are relative to the wild-type value of 6.85 kcal mol^{-1} M^{-1} which has been normalized to 1.00. Estimated average error is less than ±0.02. [c] ΔG_{H_2O} is expressed in units of kilocalories per mole. Estimated average error is less than ±0.1.

(A132V, N100A, and N100G), renaturation in ammonium sulfate was required to obtain accurate values for ΔG_{H_2O}, C_m (the midpoint concentration of GuHCl in molar), and m_{GuHCl} (Shortle et al., 1990). To estimate the precision of the data, 27 mutants were selected at random for a repetition of the denaturation assay. Comparison of the two values indicated that the ΔG_{H_2O} varied on average by ±0.03 kcal/mol and the m_{GuHCl} varied by ±0.5%. From this result, we conclude that the experimental uncertainty is less than ±0.1 kcal/mol for $\Delta\Delta G$ and less than ±0.02 for Δm, as reported previously (Shortle et al., 1990).

Changes in stability ($\Delta\Delta G = \Delta G_{mut,H_2O} - \Delta G_{wt,H_2O}$) resulting from Ala, Gly, and Val substitutions varied over a large range; except for substitutions at the positions of Pro and Ser residues and Ala substitutions at the Gln positions, $\Delta\Delta G$ varied by more than 3-fold. For example, mutating an Asn to Gly led to variations in $\Delta\Delta G$ from +0.12 kcal/mol (N138G) to −4.88 kcal/mol (N100G). Mutating an Ala residue to Gly resulted in loss of stability ranging from −0.23 kcal/mol (A90V) to −4.79 kcal/mol (A132V). Similarly, mutating a Thr residue to an Ala showed the same trend: +0.02 kcal/mol (T41A) to −2.38 kcal/mol (T62A).

The 13 Ala residues in the protein were mutated to both Gly and Val. Of the Ala residues, five are located in α helices (Ala-58, Ala-60, Ala-102, Ala-130, and Ala-132), five are in β-sheet structures (Ala-12, Ala-17, Ala-69, Ala-90, and Ala-94), and the remaining three reside in either a turn or an extended structure based on the criteria of Kabsch and Sander (1983). For the Ala to Gly substitutions, the stability of nuclease decreased by an average of −1.58 kcal/mol with a standard deviation of ±1.07 kcal/mol, while substitutions to Val resulted in a mean stability loss of −1.93 (±1.25) kcal/mol.

The average decrease in stability of mutating Ala to Gly is more than 3 times the value predicted on the basis of the loss of the free energy of transfer of this hydrophobic side chain: −0.42 kcal/mol (Fauchere & Pliska, 1983) versus −0.5 kcal/mol (Nozaki & Tanford, 1971). However, a very recent theoretical reconsideration of these data by Sharp et al. (1991) suggests the true side-chain partitioning free energy for Ala is −1.02 kcal/mol, a value much closer to the average we report.

For all 5 Ala positions in α helices, substitution with Val results in a larger decrease in stability than does substitution with Gly, even though the helix propensity of Gly is the lowest of all 20 amino acids and that of Val is only slightly less than that of Ala (O'Neil & Degrado, 1990; Lyu et al., 1990). On the other hand, for positions in a β sheet, the Val mutation is more stable than the Gly mutation at four out of five sites (all of which are buried), with the single exception (Ala-17) being a position with significant solvent exposure.

For the nine positions occupied by Gly in wild-type nuclease, the average $\Delta\Delta G$ upon mutation to Ala was −1.24 (±1.93) kcal/mol and for mutation to Val, −2.36 (±1.44) kcal/mol. By the criteria of Kabsch and Sander (1983), eight of the Gly residues are located in turns or bends. Only Gly-88 is found in a regular form of secondary structure (β strand). However, by the criteria of Richardson and Richardson (1988), Gly-107 is the C-cap residue of the third helix of nuclease and has ϕ–ψ angles representative of a left-handed α helix in the Ramachandran diagram (Ramachandran & Sasisekharan, 1968). This may be part of the reason that substitutions at position 107 are the most unstable in this collection of mutant nucleases. In addition, in every case but one, the Val substitution is less stable than the corresponding Ala substitution. The sole exception is Gly-79, an unusual site where substitutions (G79S, G79D, G79A, G79V) and single and double amino acid insertions (79A80, 79G80, 79L80, 79P80, 79Q80, 79AG80) all result in essentially the same decrease in ΔG_{H_2O} and the same change in m_{GuHCl} (Shortle & Meeker, 1986; Sondek & Shortle, 1990, 1992).

For the six prolines in nuclease, two are situated in β strands (Pro-11 and Pro-31), one sits at the $N + 1$ position of helix 1 (Pro-56), two occupy key positions in turns (Pro-47 and Pro-117), and one is situated in the active site in a segment of extended chain (Pro-42). The stability effects of substituting these prolines were surprisingly small: the average $\Delta\Delta G$ on substitution with Ala was −0.08 (±0.52) kcal/mol versus −0.50 (±0.87) kcal/mol on substitution with Gly. The least stable of the 12 mutants, P31G, decreased the stability of nuclease by only −1.5 kcal/mol. These findings suggest that proline residues make only a marginal contribution to the overall stability of nuclease; the restriction in allowed ϕ–χ angles of prolines does not, on average, have more than a slight stabilizing effect. In fact, three mutations, P42A, P117A, and P117G, actually increased the stability of nuclease by as much as +0.9 kcal/mol.

Two prolines occupy sites thought to play important roles in stabilizing specific local structures: Pro-56 is at the $N + 1$ position of the first α helix, which is a common position for Pro residues especially when preceded by a Gly residue (Piela et al., 1987; Richardson & Richardson, 1988; MacArthur & Thornton, 1991), and Pro-117, with a cis-peptide bond, is the key residue of type 6 turns (Lewis et al., 1973; Stewart et al., 1990; MacArthur & Thornton, 1991). Although mutation of Pro-56 to Gly lowers stability by −0.95 kcal/mol, substitution with Ala has essentially no effect on stability, perhaps a result of this residue's higher helix propensity. As reported by Hynes and Fox (1992), replacement of Pro-117 with Gly stabilizes nuclease by +0.9 kcal/mol. On replacement with Ala, a somewhat smaller increase in $\Delta\Delta G$ of +0.6 kcal/mol is observed. Recently we have constructed a proline-less mutant in which all of the proline residues have been replaced with either Ala or Gly (Green and Shortle, unpublished observations). This mutant nuclease, with six amino acid substitutions, is actually more stable than the wild-type protein by +0.39 kcal/mol.

For the nine sites in wild-type nuclease occupied by threonine residues, three are in β strands (Thr-13, Thr-22, Thr-33), one is in an α helix (Thr-62), two are in turns (Thr-44, Thr-120), and the remaining three are in extended segments (Thr-4, Thr-41, Thr-82). Thr to Ala substitutions resulted in an average $\Delta\Delta G$ of −0.96 (±0.76) kcal/mol, and Thr to Gly substitutions gave an average value of −1.80 (±0.73) kcal/mol. Although threonine is often classified as a polar amino acid, it is considered to be hydrophobic by a number of indices (Nakai et al., 1988). The free energy change for transfer of the Thr side chain from dilute aqueous solution to 1-octanol is −0.36 kcal/mol according to Fauchere & Pliska (1983) and −0.4 kcal/mol according to the Nozaki and Tanford (1971) scale. However, correction of the first value for the effect of volume differences leads to a much larger calculated value of −1.51 kcal/mol (Sharp et al., 1991). As was the case with mutations at sites of wild-type Ala residues, this corrected value is much closer to the average loss in free energy we observe on removal of the entire side chain.

The remaining residue positions analyzed in this study involve the three Ser, six Gln, and six Asn residues in wild-type nuclease. The average $\Delta\Delta G$ values for the Ala replacments of Ser, Gln, and Asn were +0.25 (±0.57), −0.16 (±0.19), and −1.70 (±1.84) kcal/mol, respectively. For the Gly substitutions, the corresponding $\Delta\Delta G$'s were −1.18 (±0.36), −1.18 (±0.73), and −1.51 (±1.88) kcal/mol. Although these three amino acids are more hydrophilic than Gly in the Fauchere–Pliska scale, Sharp et al. (1991) conclude that in fact they are really more hydrophobic than Gly, with free energies of transfer (relative to Gly) of 0.58, 1.47, and 0.39 kcal/mol, respectively. For positions occupied by Ser and Gln in wild-type nuclease, the mutation to Gly was always more destabilizing than that to Ala. It is noteworthy that for the two cases where the wild-type Ser residue was in an α helix (Ser-59 and Ser-128), replacement with an Ala residue increased the stability of the mutant nuclease by 0.4 and 0.7 kcal/mol, respectively, even though Ser-128 forms a buried hydrogen bond with Glu-101.

Statistical Correlations between $\Delta\Delta G$ and Descriptors of Residue Environment. As evident from the large range in $\Delta\Delta G$ for the same type of substitution at different positions in nuclease, the local structure surrounding a particular site must be a major determinant of the contribution of the residue at that site to stability. In order to identify the most important features of this local structure, statistical correlations were sought between the value of $\Delta\Delta G$ and parameters that describe the local residue environment. The values for these parameters, which are listed under Experimental Procedures, were calculated from the X-ray crystal structures of wild-type staphylococcal nuclease (Loll & Lattman, 1989; Hynes & Fox, 1991). The data for substitutions to Ala and substitutions to Gly were analyzed in 4 distinct groups: the 13 wild-type alanine residues alone; the 9 wild-type threonine residues alone; the 21 small polar residues (Ser, Asn, Gln, and Pro); and all positions together (30 for mutations to Ala, 43 for mutations to Gly). All correlations presented are with parameters of the

apoenzyme (Hynes & Fox, 1991), but similar correlations were found for the complexed structure (Loll & Lattman, 1989). The Pearson correlation coefficient, r, was calculated assuming a linear relationship between the dependent variable (either $\Delta\Delta G$ or some form of Δm_{GuHCl}) and the structural parameter. The two-tailed significance, p, a measure of the probability that the correlation coefficient for the two parameters could occur by chance, was also determined. Scatter plots were generated to permit visual inspection for the possible presence of obvious nonlinear correlations between the two parameters.

In a previously published report in which each of the large hydrophobic residues in staphylococcal nuclease was mutated to Ala and Gly (Shortle et al., 1990), the parameters that correlated best with $\Delta\Delta G$ were all measures of the relative buriedness of the residue. The three best parameters found in that study were the number of α carbons within 10 Å of the α carbon of the residue "X" [$r = 0.6768$ (X → Ala), $r = 0.7571$ (X → Gly)], the crystallographic thermal B factor of the β carbon of residue "X" [$r = -0.6759$ (X → Ala), $r = -0.7458$ (X → Gly)], and the fraction of residue "X"'s side-chain surface area that is exposed to solvent [$r = 0.6268$ (X → Ala), $r = 0.6568$ (X → Gly)]. Each of these correlations had a two-tailed significance of less than 0.001.

When the 13 wild-type positions occupied by Ala, usually included among the hydrophobic amino acids, are mutated to Gly, no significant correlations were found between these 3 parameters: the correlation coefficients were $r = 0.2029$ ($p = 0.527$), $r = -0.0967$ ($p = 0.765$), and $r = -0.1855$ ($p = 0.564$), respectively. Nevertheless, correlations between $\Delta\Delta G$ and two of these three parameters are quite significant for the nine positions occupied by Thr residues, an amino acid type that is by most criteria less hydrophobic than Ala. For mutant proteins with Thr substituted with Gly, the correlation coefficients are as follows: (1) the number of neighbors within 10 Å, $r = 0.8030$ ($p = 0.009$); (2) the thermal B factor C_β, $r = -0.8609$ ($p = 0.006$); and (3) the fractional side-chain solvent-accessible surface area, $r = -0.4201$ ($p = 0.300$). As might be expected for substitutions of Thr with Ala, a residue of supposedly greater intrinsic hydrophobicity, the correlations are of little or no significance: $r = 0.5218$ ($p = 0.150$); $r = -0.6528$ ($p = 0.079$); $r = -0.2225$ ($p = 0.596$), respectively.

Glycine is the only other residue analyzed in this study for which there are a relatively large number of cases (9). However, substitution of Gly with either Ala or Val is a different type of mutation than the others reported here, which can be considered side-chain subtractions or side-chain truncations. Instead of reducing the steric size of the side chain, these mutations at Gly sites increase it, plus the addition of a β carbon greatly restricts the local ϕ–ψ angles available to the chain. Consistent with the expectation that the stability effects of these two types of mutations should arise by mechanisms very different from simple hydrophobicity parameters, no correlations were found between $\Delta\Delta G$ and any of the parameters discussed above on replacement of Gly with either Ala or Val. For these reasons, Gly to Ala and Gly to Val mutations are not included in the correlations below.

For the 21 sites occupied by the polar residues, Ser, Gln, Asn, and Pro, the correlation coefficient between $\Delta\Delta G$ on replacement with Gly and the same 3 parameters of degree of buriedness were as follows: number of neighbors within 10 Å, $r = 0.5656$ ($p = 0.012$); C_β thermal factor, $r = -0.4166$ ($p = 0.076$); and accessible surface area, $r = -0.6223$ ($p = 0.004$). For the Ala substitutions at these same sites, the correlation coefficients between $\Delta\Delta G$ and the number of neighbors within 10 Å were as follows: $r = 0.4479$ ($p = 0.054$); C_β thermal factor, $r = -0.3274$ ($p = 0.171$); and fractional side-chain solvent-accessible surface aera, $r = -0.3631$ ($p = 0.126$). Thus, just as with the "polar" amino acid threonine, changes in stability on completely removing the side chain by substitution with Gly correlate with the burial of the side chain. Again, this correlation is much less significant for substitutions with Ala. This finding supports the conclusion that even polar residues may contribute to stability via some degree of solvophobicity.

If Ala and Thr sites are combined with the Ser, Gln, Asn, and Pro sites, there are a total of 43 positions where small, uncharged residues have been replaced with Gly. The correlation coefficients now become the following: (1) $\Delta\Delta G$ with the number of α carbons within 10 Å, $r = 0.5427$ ($p < 0.0001$); (2) $\Delta\Delta G$ with the thermal B factor, $r = -0.5705$ ($p < 0.0001$); and (3) $\Delta\Delta G$ with the fractional side-chain solvent-accessible surface area, $r = -0.4162$ ($p = 0.008$). The correlation coefficients were much less significant for the 30 mutants with Ala substituting for Thr, Ser, Gln, Asn, and Pro: $r = 0.4503$ ($p = 0.016$); $r = -0.4250$ ($p = 0.027$); $r = -0.3469$ ($p = 0.076$), respectively.

Statistical Correlations Involving m_{GuHCl} and Descriptors of Residue Environment. In addition to changes in stability, guanidine hydrochloride denaturation experiments yield a second parameter, m_{GuHCl}, which is defined as the rate of change of ΔG_{H_2O} with respect to denaturant concentration. As has been noted in the past with mutants of staphylococcal nuclease (Shortle & Meeker, 1986, 1989; Shortle, 1986; Shortle et al., 1990; Sondek & Shortle, 1990, 1992) plus observed with mutants of many other proteins (Perry et al., 1987; Villafranca et al., 1987; Hughson & Baldwin, 1989; Shirley et al., 1989; Thomson et al., 1989; Caffrey & Cusanovich, 1991; Sandberg & Terwilliger, 1991; Lim et al., 1992), substitutions occasionally result in relatively large changes in m_{GuHCl}. Those mutant nucleases for which m_{GuHCl} is 1.05 or greater relative to the wild-type value (6.85 kcal mol^{-1} M^{-1}), which is defined as 1.00, are denoted m^+ mutants; those for which m_{GuHCl} is 0.95 or less relative to wild-type are denoted m^- mutants, and the remainder are classified as m^0.

Figure 2A shows the distribution of these three mutant types with respect to the secondary structure of staphylococcal nuclease. As is apparent, m^+ and m^- mutants are not uniformly distributed around the native structure. Consistent with results from a previous study of Ala and Gly substitutions for the large hydrophobic amino acids (Shortle et al., 1990), m^+ mutations are confined to residues that form the major hydrophobic core of nuclease centered on the five-strand β barrel (but which does not include the first half of the first β strand and includes residues in two α helices). Likewise, m^- mutants are distributed throughout other regions of the molecule. Seven m^0 mutants occur at what may be boundary sites between these two regions of the native structure. At four positions, one mutation resulted in a slope value of ≥ 1.05 while the other mutation resulted in a slope ≤ 0.95. Three of these four ($\pm$) sites lie around the edge of the β barrel, while one site is more centrally located within it. These trends are emphasized in Figure 2B, which includes all sites analyzed to date for Ala and Gly substitutions.

The segregation of m^+ and m^- mutants revealed in the previous study of hydrophobic residues led to statistically significant correlations between m_{GuHCl} and three parameters that described a residue's relative location in the wild-type native structure: (1) residue number; (2) distance from C_β to the C_β at residue 24; and (3) distance from C_β to the C_β at residue 133 (Shortle et al., 1990). For the set of mutants

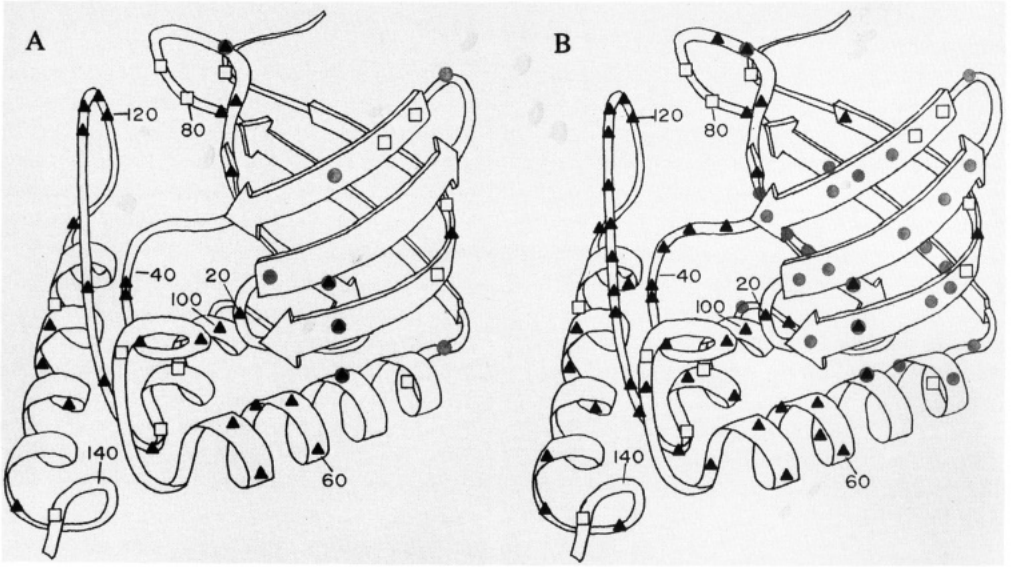

FIGURE 2: Ribbon diagram showing the approximate position of the wild-type residue for different classes of staphylococcal nuclease mutants. Residue positions where a substitution with an alanine and/or glycine increases m_{GuHCl} to 1.05 or more of the wild-type value are denoted by shaded circles. Those that decrease the value of m_{GuHCl} to 0.95 or lower are denoted by (▲). Positions where both substitutions change the slope by less than 0.05 are shown as (□). Positions where one mutant changes the slope to 1.05 or higher and the other mutant changes the slope to 0.95 or below are designated by black triangles inscribed in shaded circles. All uncharged, polar residues are shown in panel A. All residues studied to date, including the large hydrophobic residues (Shortle et al., 1990), are shown in panel B.

analyzed in this study, no significant correlations were found with these same parameters. One explanation is that the true correlation is between the m^+/m^- class of a mutant and its location in the molecule, not the actual value of m_{GuHCl}. Since there are many fewer m^+ mutants in the present study, the three correlations are not statistically significant, even though the segregation of m^+ and m^- classes seems to be preserved.

In an effort to understand the physicochemical basis of changes in m_{GuHCl} among mutant forms of staphylococcal nuclease, a broad search was undertaken for statistical correlations between parameters that describe the local structural environment and a number of modified forms of m_{GuHCl}. Unexpectedly, a very significant correlation was identified between the same parameters of buriedness reported above to correlate well with $\Delta\Delta G$ and the absolute value of the change in m_{GuHCl} between the Ala mutant and the Gly mutant, $|\Delta m_{Ala\rightarrow Gly}|$. The initial observation of this correlation was for the set of mutations at positions of Ser, Thr, Gln, Asn, and Pro residues. For these 30 mutant pairs, $|\Delta m_{Ala\rightarrow Gly}|$ correlates with the following: the number of C_α within 10 Å, $r = 0.7182$ ($p < 0.0001$) (Figure 3A); the solvent-accessible surface area of the side chain, $r = -0.6086$ ($p = 0.001$) (Figure 3B); and the C_β thermal B factor, $r = -0.4870$ ($p = 0.010$) (Figure 3C).

Subsequent examination of different subsets of mutant pairs indicated that this correlation may be significant for all residue positions except those occupied by Ala in wild-type [for which the correlation coefficients are $r = -0.1610$ ($p = 0.617$), $r = -0.3341$ ($p = 0.288$), and $r = 0.4207$ ($p = 0.173$), respectively]. In this regard, it is noteworthy that substitutions of Gly at wild-type Ala sites also do not display a significant correlation between $\Delta\Delta G$ and these same three parameters, although as described above, mutants at virtually all other types of residues do show this correlation.

These correlations involving $|\Delta m_{Ala\rightarrow Gly}|$ can also be demonstrated at sites occupied by the large hydrophobic amino acids (i.e., Leu, Val, Ile, Met, Phe, and Tyr) although they are not as significant: $r = 0.4395$ ($p = 0.007$) for the number of neighbors within 10 Å; $r = -0.3718$ ($p = 0.026$) for the fractional solvent-accessible surface area; and $r = -0.4086$ ($p = 0.013$) for C_β thermal B factors. On combination of the data from the hydrophobic residues with those from the Ser, Thr, Gln, Asn, and Pro sites, the correlation coefficients become $r = 0.5999$ ($p < 0.0001$) (Figure 3D), $r = -0.5174$ ($p < 0.0001$) (Figure 3E), and $r = -0.4821$ ($p < 0.0001$) (Figure 3F), respectively. However, upon addition of wild-type Ala → Gly mutants to the data set, thereby including all mutant pairs involving Ala and Gly at the same residue position, the correlation coefficients drop to $r = 0.5004$ ($p < 0.0001$), $r = -0.4724$ ($p < 0.0001$), and $r = -0.3963$ ($p < 0.0001$), respectively.

It should be noted that a correlation also exists between $|\Delta m_{Ala\rightarrow Gly}|$ and the change in stability ($\Delta\Delta G$). The correlation is most significant for the large hydrophobic amino acids with $r = 0.6107$ ($p < 0.0001$) for the Gly substitutions and $r = 0.5451$ ($p = 0.001$) for the Ala substitutions. For all Ser, Thr, Gln, Asn, and Pro residues, the correlation is much less significant with $r = 0.4537$ ($p = 0.013$) and $r = 0.3896$ ($p = 0.037$), respectively. This indicates that the correlation between $|\Delta m_{Ala\rightarrow Gly}|$ and the parameters that describe buriedness is probably not due to an indirect effect mediated through the correlation between $|\Delta m_{Ala\rightarrow Gly}|$ and $\Delta\Delta G$.

In a previous study (Shortle et al., 1990), data from mutant nucleases with Ala and Gly substitutions at sites of the large hydrophobic residues exhibited a statistically significant correlation between $\Delta\Delta G$ and the absolute value of the change in m_{GuHCl}, $|\Delta m_{GuHCl}|$. As shown in Figure 4A, a similar analysis for all of the mutant proteins in this study involving substitution with Ala yields a correlation coefficient between these two parameters of $r = 0.7221$ ($p < 0.0001$). For all residue types mutated to Gly, the correlation coefficient has little significance: $r = 0.3875$ ($p = 0.010$) (Figure 4B). For the 22 Ala to Val and Gly to Val mutants, no significant correlation was found with $r = 0.2655$ ($p = 0.245$). As in the previous study, no evidence of a correlation was found between m_{GuHCl} or Δm_{GuHCl} and $\Delta\Delta G$; the correlation is only manifest when the sign of the change in m_{GuHCl} is removed by taking the absolute value.

On combining the 30 mutant proteins of the type X → Ala from this study with the 39 mutant proteins of this type where X is a large hydrophobic amino acid, the correlation coefficient

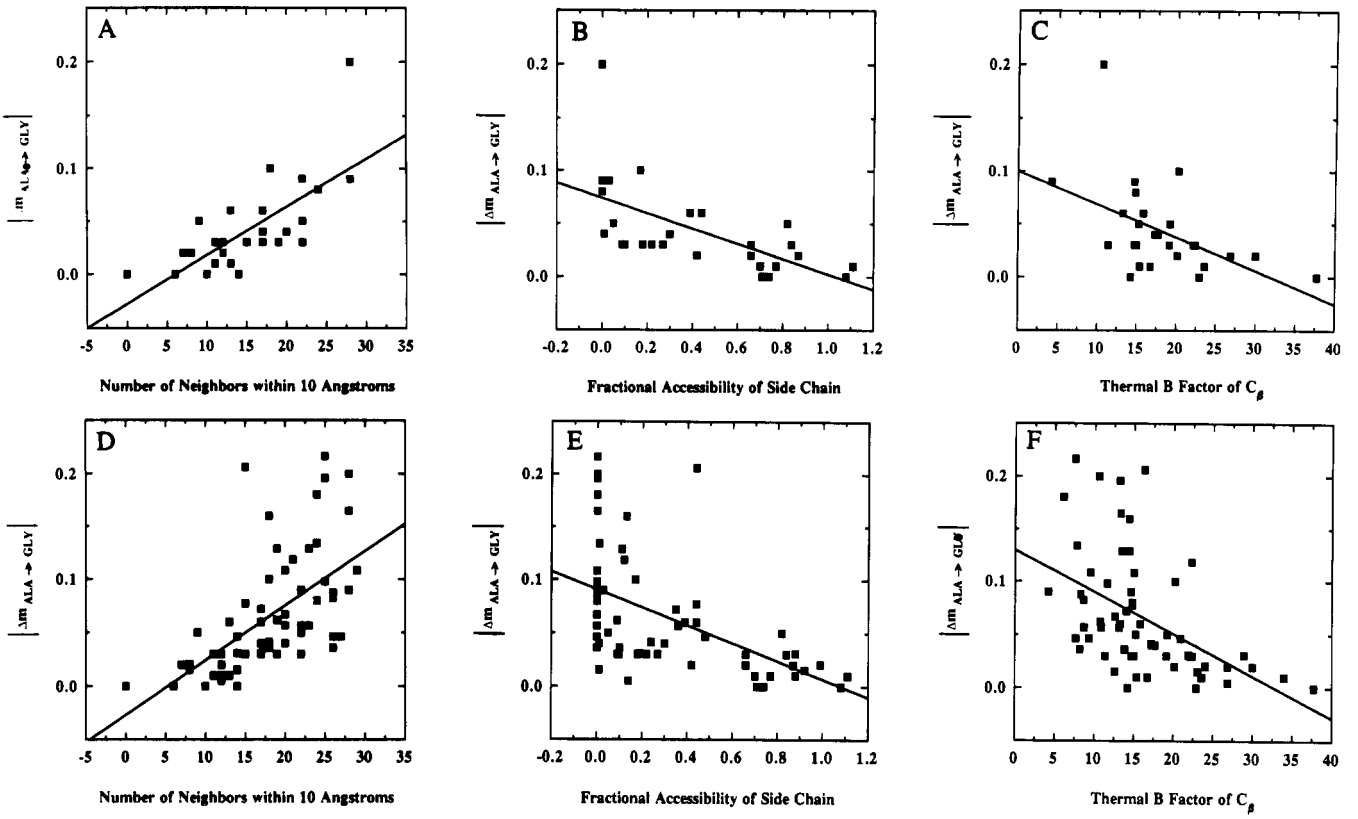

FIGURE 3: Scatter plots of $|\Delta m_{Ala \rightarrow Gly}|$ for substitutions at a residue position versus three parameters that describe the buriedness of the wild-type residue. Uncharged, polar residues are shown in the upper panels (A–C). These mutant data are combined with data from the large hydrophobic residues (Shortle et al., 1990) in the lower panels (D–F). (A and D) Number of neighbors within 10 Å; (B and E) fractional accessibility of the side chain; (C and F) crystallographic thermal B factor of the C_β carbon.

becomes $r = 0.6218$ ($p < 0.0001$). Likewise, for the combined set of 43 + 39 residues to Gly mutants, $r = 0.6198$ ($p < 0.0001$). These data in scatter plot form are shown in Figure 4C,D.

Finally, combining the data from all X → Ala and X → Gly mutants of staphylococcal nuclease yields a large enough number of cases to demonstrate that it is unnecessary to take the absolute value of Δm_{GuHCl} to observe a correlation with $\Delta\Delta G$. When these mutants are partitioned into the separate classes of m^+ and m^-, a highly significant correlation between Δm_{GuHCl} and $\Delta\Delta G$ is found within each class. As shown in Figure 5, $r = 0.5830$ ($p < 0.0001$) for all 56 m^+ mutants considered as a group (without distinction between X → Ala and X → Gly) (5A) and $r = -0.6507$ ($p < 0.0001$) for all 96 m^- mutants considered as a group (5B).

Discussion

For this collection of mutants, the most significant correlations between the change in stability ($\Delta\Delta G$) and the location of the mutation in the wild-type native structure were with parameters that describe the degree of burial or solvent inaccessibility of the residue or its side chain. Several different mechanisms could be responsible for these correlations. (1) Loss of intrachain hydrogen bonds formed by the wild-type residue. The less accessible the side chain is to solvent, the larger the potential number of hydrogen bonds formed between it and the surrounding structure and the greater the potential difficulty of replacing them in the mutant protein with bonds to water. (2) Loss of favorable van der Waals interactions with the surrounding chain. The free energy gained on eliminating these interactions should vary in proportion to the area of contact between the side chain and the surrounding atoms and thus should depend on the fraction of the side chain that is not in contact with solvent. (3) Loss of the free energy of transfer of the side chain out of contact with solvent.

One feature of the data bears on the issue of which of these mechanisms is responsible for these correlations between measures of buriedness and $\Delta\Delta G$: the correlation is much less significant for replacements with Ala than for replacements with Gly. If the predominant factor were loss of the side chain's hydrogen-bonding atoms, then equally significant correlations would be expected for both of these mutant types. Hence, intrachain hydrogen bonding is probably not the sole mechanism underlying this correlation. Loss of van der Waals interactions would be expected to play a role in the stability loss accompanying truncation of a buried side chain unless relaxation of the local structure led to filling of the cavity. X-ray crystallographic studies of a number of mutant proteins have suggested that such structural relaxation may be a common response, often with nearly complete filling of the cavity (Eigenbrot et al., 1990; McRee et al., 1990; Daopin et al., 1991).

The contribution of residues like Ser, Thr, Asn, Gln, and Pro to protein stability via the free energy of transfer of their side chains out of water is more difficult to assess. Although these side chains are usually considered to be quite *hydrophilic*, Roseman has argued that amide and hydroxyl groups acquire significant nonpolar character when all of their hydrogen-bonding requirements have been met (1988). In addition, Sharp et al., (1991) have concluded that all amino acids in their uncharged forms are more *hydrophobic* than glycine, the residue used to subtract out the contribution of backbone atoms to partitioning behavior. Their very surprising conclusions are based on a theoretical revision of the proper way to extract interaction energies from data on solute partitioning between water and 1-octanol. Thus, it is possible that this group of

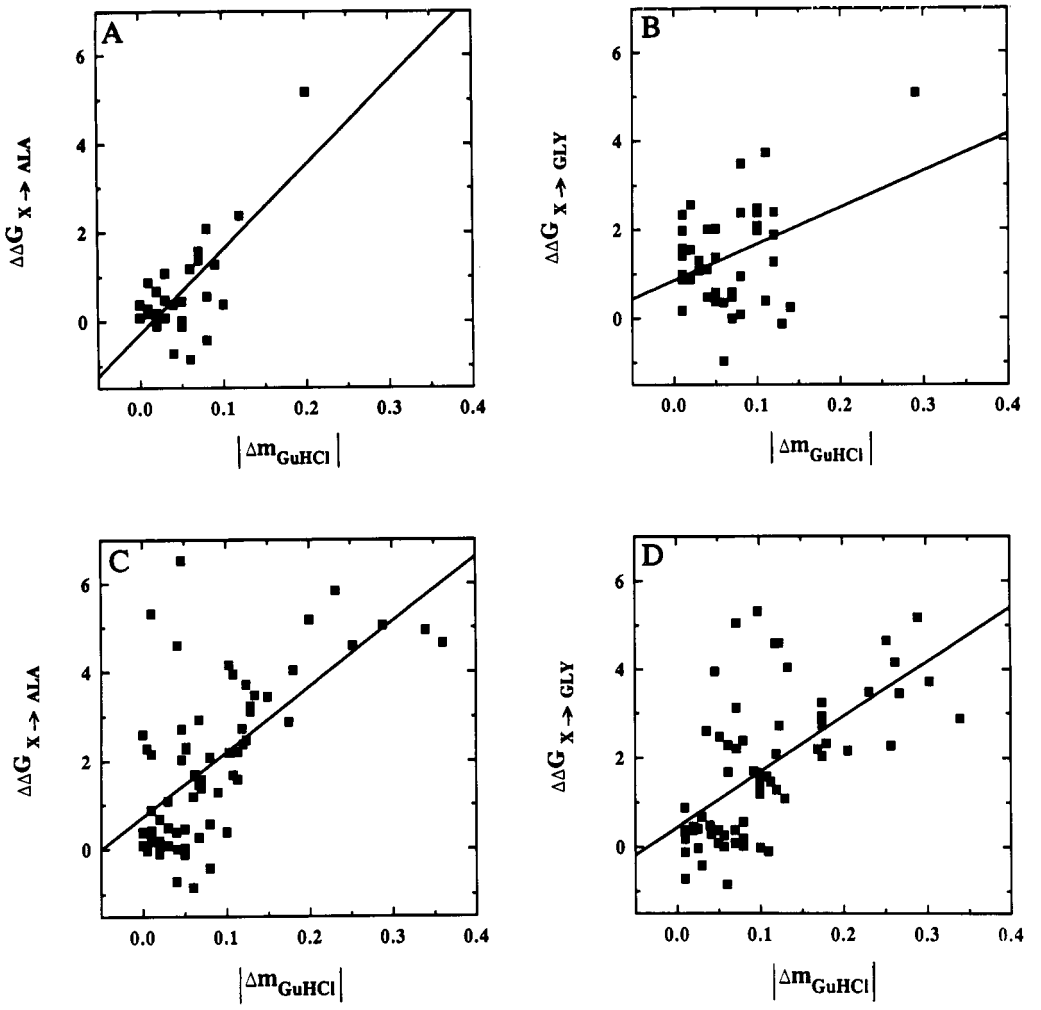

FIGURE 4: Scatter plots of $\Delta\Delta G$ for a mutant nuclease versus $|\Delta m_{GuHCl}|$, the absolute value of the difference in m_{GuHCl} between mutant and wild-type. (A) Uncharged, polar residues mutated to Ala; (C) uncharged, polar residues plus large hydrophobic residues mutated to Ala; (B) uncharged, polar residues mutated to Gly; (D) uncharged, polar residues plus large hydrophobic residues mutated to Gly. $\Delta\Delta G$ is expressed in units of kilocalories per mole.

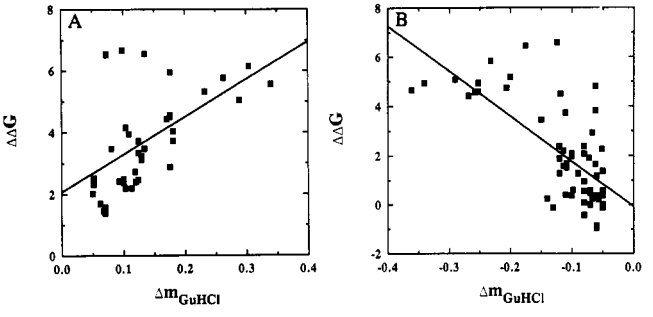

FIGURE 5: Scatter plots of $\Delta\Delta G$ for a mutant nuclease versus Δm_{GuHCl}. All mutants which increase m_{GuHCl} to 1.05 or more, regardless of whether the substitution is to glycine or alanine, are shown in panel A. Those mutants that decrease m_{GuHCl} to 0.95 or less are shown in panel B. $\Delta\Delta G$ is expressed in units of kilocalories per mole. Δm_{GuHCl} is the difference of m_{GuHCl} of the mutant from the normalized wild-type value of 6.8 kilocalories per mole M.

amino acids contributes to stability via essentially the same mechanism as do the hydrophobic amino acids.

In support of this possibility are two similarities displayed by both the conventional hydrophobic residues (Leu, Val, Ile, Met, Phe, and Tyr) and these polar, uncharged residues with respect to the correlations exhibited by their Ala and Gly substitutions. (1) As mentioned above, the correlation between $\Delta\Delta G$ and side chain burial is more significant for the Gly substitutions than for the Ala substitutions. (2) For reasons that are not apparent to us, the correlation is much weaker with the absolute amount of area buried as the independent variable as opposed to the *fraction* of the side chain buried. As would be expected, the polar residues appear considerably less hydrophobic by the criteria of the change in ΔG_{H_2O} per change in degree of residue burial obtained from the linear regression line for these correlations. With respect to the number of C_α atoms within 10 Å, the parameter that gives the best correlation with both polars and hydrophobics, the slope of the regression line for the polar residues is one-third the value obtained for the hydrophobic residues. This finding is consistent with the uncharged, polar side chains either being intrinsically less hydrophobic than the conventional hydrophobic side chains or having a smaller average amount of hydrophobic surface area per residue.

The most noteworthy findings reported here, however, concern two highly significant correlations involving changes in m_{GuHCl}. The three theories that have been proposed to explain denaturant action all come to essentially the same conclusion: $m_{denaturant}$ should be proportional to the difference between some property of the denatured state and the same property of the native state. For the different theories, this property is either the number of denaturant binding sites (Tanford, 1970), the fractional exposure to solvent (Tanford, 1970), or the solvent-accessible surface area (Schellman, 1978). As previously argued, changes in m_{GuHCl} are most readily explained as manifestations of changes in A_D, the

solvent-accessible surface area of the denatured state (Shortle & Meeker, 1986, 1988, 1989; Shortle et al., 1990). m^+ mutant proteins have been proposed to denature to a state with less residual structure than wild-type and therefore greater A_D, whereas m^- mutants presumably denature to a more compact state with a smaller A_D.

Several types of data support the conclusion that changes in m_{GuHCl} reflect structural changes in the denatured state. (1) Oxidation of cysteine pairs to form disulfide cross-bridges reduces m (Villafranca et al., 1987; Pace et al., 1990); (2) m values often increase at extremes of pH, presumably from expansion of the denatured state due to increased electrostatic repulsion between surface charges (Pace et al., 1990); (3) the only option for explaining m^+ mutants in terms of the native state requires an implausible *reduction* in its solvent-accessible surface area, in some mutants by as much as 35% (Shortle & Meeker, 1986); (4) ΔC_p changes in parallel with changes in m_{GuHCl} for staphylococcal nuclease, consistent with increases (m^+) and reductions (m^-) in the amount of additional hydrophobic surface exposed on denaturation (Shortle et al., 1988); (5) in very large fragments of staphylococcal nuclease studied as models of the denatured state, all m^+ mutants studied reduce the residual structure as quantitated by CD spectra and Stokes radius, whereas a minority of m^- mutants increases this residual structure relative to the wild-type fragment (Shortle & Meeker, 1989, and unpublished observations).

If this general interpretation is correct, then a molecular understanding of changes in m_{GuHCl} should provide details into the structure and energetics of this poorly understood state of proteins (Dill & Shortle, 1991). It should be noted that similar changes in m_{GuHCl}, though perhaps not as large as the most extreme mutants of staphylococcal nuclease, have been seen for mutant forms of a number of other proteins: dihydrofolate reductase (Perry et al., 1987, 1989; Villafranca et al., 1987), T1 RNase (Shirley et al., 1989), gene V protein of f1 phage (Sandberg & Terwilliger, 1991), apomyoglobin (Hughson & Baldwin, 1989), calbindin (Akke & Forsen, 1990), and the amino-terminal fragment of λ repressor (Lim et al., 1992).

The finding of a highly significant correlation between $|\Delta m_{GuHCl}|$ and the stability lost on mutation, $\Delta\Delta G$, suggests that these putative structural changes in the denatured state are associated with, and thus may be the origin of, part of the stability loss experienced by an "average" mutant nuclease. This correlation was first reported in a previous study of 83 nucleases mutated at sites of large hydrophobic residues (Shortle et al., 1990). As shown in Figure 4, the same correlation is also observed for the mutants analyzed in this study with high significance for replacements with Ala ($p < 0.0001$) but minimal significance for replacements with Gly ($p = 0.010$). When the data from this and the previous study (Figure 4C,D) are merged, the correlation is seen to be highly significant ($p < 0.0001$) for both Ala and Gly mutant types, with identical correlation coefficients of 0.62.

Combining denaturation data from both studies demonstrates that it is unnecessary to take the absolute value of Δm_{GuHCl} if the sign of this parameter is eliminated by grouping mutants according to the m^+/m^- classes: the highly significant correlation between $\Delta\Delta G$ and Δm_{GuHCl} for the m^+ and m^- mutant classes is shown in Figure 5A and Figure 5B, respectively. In addition, the slope of the regression line for the two mutant classes is very similar, indicating that, on average, the same stability loss tends to accompany the same fractional change in m_{GuHCl} independent of whether Δm_{GuHCl} is positive or negative. The simplest interpretation of the correlation coefficients in these two cases is that slightly more than half of the stability lost by an average mutation is a consequence of the structural changes in the denatured state that give rise to an altered value of m_{GuHCl}.

It is tempting to speculate that, for those mutant proteins which fall close to the regression line in Figures 4 or 5, most of the stability loss is a consequence of the structural changes in the denatured state reflected in changes in m_{GuHCl}. Likewise, for those proteins which are distant from the regression line, most of the stability loss may be a consequence of other energetic effects, notably those that raise the free energy of the native state. In this regard, it is interesting to consider what features might be unique to the mutants most distant from the regression line. The mutants furthest above include substitutions for Tyr-91 and Tyr-93, which form buried hydrogen bonds in wild-type nuclease with Asp-77 and Glu-75, respectively. In addition, this set includes Gly substitutions for a tight cluster of hydrophobic residues, Ile-72, Val-74, and Ile-92. Because of their central positions in β sheets, truncation of any of these side chains might create a cavity in the native structure that cannot be filled. The mutants furthest below the regression line include Ala substitutions at Ser-59 and Ser-128 as well as the Gly substitutions at Pro-117. The substitution of Ala for either of these Ser residues could stabilize the α helices via the increase in helical propensity, with the greater native-state stability partially offsetting the stability loss accompanying changes in the denatured state. Likewise, the Gly mutation at Pro-117 (which assumes a cis-peptide configuration in the native state) is more stable than wild-type; perhaps the increase in stability due to removal of strain would have been greater without the putative destabilizing effect on the denatured state.

Irrelevant correlations can arise as statistical artifacts. Since for each mutant both ΔG_{H_2O} and m_{GuHCl} are obtained from one set of data points by a linear least-squares fit to the equation

$$\Delta G_{app} = \Delta G_{H_2O} - m_{GuHCl} * [GuHCl]$$

errors in m_{GuHCl} will be correlated with errors in ΔG_{H_2O}. This is the origin of the so-called "isokinetic effect" which leads to a correlation between two parameters extracted from one data set. However, this trivial correlation only becomes significant for multiple, independent measurements when they are essentially replications of the same measurement, i.e., when the errors within a single experiment are as large or larger than the true variation between experiments (Krug et al., 1976). For the data points in Figure 4, the range in $\Delta\Delta G$ is 6 kcal/mol, whereas the estimated error in this number is ±0.1 kcal/mol; likewise, the range in Δm_{GuHCl} is ±0.35 relative to the wild-type value, but the estimated error in m_{GuHCl} is ±0.02. If the correlation between $\Delta\Delta G$ and $|\Delta m_{GuHCl}|$ were simply due to correlation of errors, the slope of the regression line should be equal to 5.5 kcal mol^{-1} (unit of $m_{GuHCl})^{-1}$; the observed regression line slope is approximately 15 kcal mol^{-1} (unit of $m_{GuHCl})^{-1}$. Consequently, this correlation cannot be due to the isokinetic effect. It is unlikely to be due to some other constant relationship among experimental errors because it is not demonstrated for the set of 22 mutant proteins involving Ala → Val and Gly → Val substitutions, a different type of mutation in which side chain bulk and hydrophobicity are increased instead of reduced.

We propose that the correlation between changes in stability and changes in m_{GuHCl} shown in Figures 4 and 5 is a consequence of the fact that, in order to achieve a moderate level of stability, the amino acid sequence of wild-type staphylo-

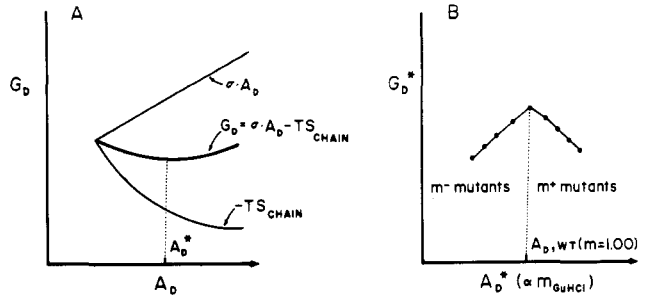

FIGURE 6: (A) Simple model of the free energy of the denatured state as a function of its solvent-accessible surface area, A_D. σ is the free energy per unit surface, and $\sigma^* A_D$ is the solvation free energy. $-TS_{chain}$ is the entropic free energy of the polypeptide chain. A_D^* is the equilibrium value of A_D. (B) Hypothetical relationship between the free energy of the denatured state and its equilibrium value of solvent-exposed area. For wild-type staphylococcal nuclease, A_D^* is an optimized value corresponding to $m_{GuHCl} = 1.00$. Any change in m_{GuHCl} resulting from a change in sequence decreases the G_D^*. m^+ mutants exhibit an m_{GuHCl} of 1.05 or higher. m^- mutants exhibit an m_{GuHCl} of 0.95 or lower.

coccal nuclease has evolved to maximize the free energy of the denatured state. This high free energy results from a complex set of chain–solvent and chain–chain interactions. Consequently, changes in amino acid sequence might be expected to alter the residual structure of the denatured state, thereby modifying the value of Δm_{GuHCl} and lowering the free energy of the denatured state below that of wild-type. In addition, we propose that the loss in denatured-state free energy, G_D, for a given value of m_{GuHCl} (or A_D) is approximated by the corresponding value of $\Delta\Delta G$ on the regression line in Figure 4. In other words, the regression line of $\Delta\Delta G$ versus $|\Delta m_{GuHCl}|$ reflects that component of $\Delta\Delta G$ resulting from a decrease in G_D.

In line with theoretical considerations put forward by Dill and others [Dill, 1985; reviewed by Dill (1990)], we propose that two factors dominate the free energy of the denatured state: the entropy of the polypeptide chain and the solvation free energy of the nonpolar surface. (Charged residues are assumed to be well hydrated in both the native and denatured states and therefore to have no net effect on any free energy differences. Likewise, the intrachain van der Waals interactions and hydrogen bonds are assumed, on average, to be equal in strength to those between chain and solvent.) If the chain is modeled as having an average value of hydrophobicity uniformly distributed along its length (Dill, 1985; Dill et al., 1989), then the free energy component due to chain solvation can be approximated as the free energy per unit surface (σ) multiplied by the amount of exposed surface at the protein–solvent interface (A_D) (Schellman, 1978). As shown in Figure 6A, the free energy of the denatured state (G_D) would be expressed by the equation:

$$G_D = \sigma^* A_D - T^* S_{chain}$$

where S_{chain} designates the polypeptide chain entropy as a function of the solvent-exposed surface area A_D.

An exact expression for the dependence of chain entropy on A_D cannot be given. However, the entropy gain of a linear chain distributed in a sphere falls off exponentially as the radius of the sphere is increased (Dill, 1985). Consequently, we assume that a similarly shaped, monotonic curve would describe the entropy gain as a function of A_D, since increases in A_D on average will correspond to increases in the hydrodynamic radius of the polypeptide chain. An equilibrium value of A_D (A_D^*) is attained when

$$\delta(G_D)/\delta A_D = 0 = \sigma - T(\delta S_{chain}/\delta A_D)$$

i.e, at the point where the slope of the $-TS_{chain}$ line is equal but opposite in sign to σ. Starting from the equilibrium value A_D^*, configurations of the denatured state with a smaller A_D would be higher in G_D because the gain in entropic free energy ($-TS_{chain}$) would be greater than the decrease in solvation free energy (σA_D). Likewise, configurations with a larger A_D would also be higher in G_D because the increase in solvation free energy would be greater than the loss in entropic free energy.

To explain the effects of mutations of the amino acid sequence, we begin with the plausible assumption that the wild-type sequence has evolved to give a denatured state with the highest attainable free energy, which for this simple model would be achieved by a statistical structure that maximizes the exposure of hydrophobic surface with a minimum increase in entropy relative to the native state. The balance between these two conflicting aims is achieved through a complex coupling between A_D and S_{chain} and leads to an optimum value of A_D^* and therefore of m_{GuHCl}. According to the Schellman model (1978), m_{GuHCl} will be proportional to the difference in solvent-accessible surface area between the denatured and the native states, $A_D - A_N$. If the assumption is made that A_N is not altered by mutations, an idea supported by the results of X-ray crystallography of mutant proteins, then all changes in m_{GuHCl} will reflect proportional changes in A_D.

When the amino acid sequence of the polypeptide chain is modified, we propose that the details of this coupling between A_D and S_{chain} may also be modified, resulting in changes in the equilibrium value of A_D. By this argument, m^+ mutants have a larger A_D^* because the decrease in entropic free energy ($-TS_{chain}$) that accompanies an increase in the mutant A_D above the A_D^* of wild-type is greater than the increase in solvation free energy. m^- mutants have a smaller value of A_D^* because, in this case, a greater decrease in solvation free energy outweighs the increase in entropic free energy that accompanies a reduction in A_D^*. Consequently, as shown in Figure 6B, the free energy of the denatured state will be reduced when the mutant value of A_D^* (or m_{GuHCl}) is different than that of wild-type. What this simple model cannot explain is the observation that the conjectured reduction in G_D appears to be the same when A_D^* (or m_{GuHCl}) is either increased (m^+) or decreased (m^-) by the same amount relative to wild-type. Presumably, this feature of the data must be a result of some fundamental aspect of the coupling between A_D and S_{chain}.

A second highly significant statistical correlation involving m_{GuHCl} was identified on characterization of this set of mutant nucleases. The value of $|\Delta m_{Ala\rightarrow Gly}|$ correlates very strongly with several measures of the extent of residue or side chain burial in the wild-type native structure. There are several puzzling features of this correlation: (1) There is no fundamental distinction between mutant pairs for which m_{GuHCl} of the Ala mutant is greater than m_{GuHCl} of the Gly mutant and mutant pairs for which the opposite is true. This correlation involves the absolute value of the difference; it is indifferent to the sign. (2) No comparable correlation is seen for the change of 13 wild-type alaine residues to mutant Gly residues, suggesting that the Ala residue in the nuclease pair must be mutant. (3) $|\Delta m_{Ala\rightarrow Gly}|$ also correlates with $\Delta\Delta G$. The correlation between $|\Delta m_{Ala\rightarrow Gly}|$ and residue buriedness, however, is probably not mediated through its correlation with $\Delta\Delta G$ because the correlation with buriedness has a higher significance than the correlation with $\Delta\Delta G$ for most sets of mutant pairs.

Within the context of the simple physical–chemical model presented above, a few of these features can be explained. The

model indicates a clear interrelationship between changes in S_{chain}, A_D, and G_D. For example, an increase in mutant S_{chain} that occurs at low values of A_D will lead to a smaller value of A_D* and also to a lower G_D. Likewise, an increase in mutant S_{chain} that is constrained to occur only at high values of A_D will lead to a larger value of A_D* and also to a lower G_D. Suppose that the entropy increase produced by introducing a swivel point (replacement of a mutant alanine with a mutant glycine) were a function of how deep into the denatured state structure the swivel point were situated. If this depth were very similar to the extent of residue burial in the native state, then a correlation would be expected with both $\Delta\Delta G$ (sum of $-TS_{chain}$ and solvation free energy σ^*A_D) and changes in A_D* ($\Delta m_{Ala\rightarrow Gly}$). What cannot be explained is the nature of the coupling between S_{chain} and A_D that would determine whether the increase in S_{chain} produced by an Ala to Gly substitution would occur at values of A_D lower or higher than the A_D* of wild-type.

Unfortunately, our knowledge of the physical properties of the denatured states of proteins is very limited (Dill & Shortle, 1991); little additional evidence from studies of staphylococcal nuclease or other proteins is available to support or refute this or any other plausible explanation. Nevertheless, the remarkable level of statistical significance of these two unusual correlations involving m_{GuHCl} suggests that important physical chemistry may underlie them, making speculations as to their origins and experimentation to test them worthwhile. Such endeavors might be rewarded by insights into the mechanisms that couple structure (polypeptide chain entropy) and energy (solvent exposure of nonpolar surface) in the denatured state and thus establish the basic topology of the folded polypeptide chain.

ACKNOWLEDGMENTS

We thank Wesley Stites for setting up the extensive data base used in our statistical analysis, John Sondek, Gary Chase, Alan Gittelsohn, and Jack Kirsch for helpful discussions of the statistics involved in our data analysis, and Preston Hensley and Neil Clarke for their thoughtful comments on the manuscript.

REFERENCES

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

Alber, T., Dao-pin, S., Wilson, K., Wozniak, J. A., Cook, S., & Matthews, B. W. (1987) *Nature 330*, 41–46.

Alber, T., Bell, J. A., Dao-pin, S., Nicholson, H., Wozniak, J. A., Cook, S., & Matthews, B. W. (1988) *Science 239*, 631–635.

Caffrey, M. S., & Cusanovich, M. A. (1991) *Biochemistry 30*, 9238–9241.

Connelly, P., Ghosani, L., Hu, C.-Q., Kitamura, S., Tanaka, A., & Sturtevant, J. M. (1991) *Biochemistry 30*, 1887–1891.

Daopin, S., Alber, T., Baase, W. A., Wozniak, J. A., & Matthews, B. W. (1991) *J. Mol. Biol. 221*, 647–667.

Dill, K. A. (1985) *Biochemistry 24*, 1501–1509.

Dill, K. A., (1990) *Biochemistry 29*, 7133–7155.

Dill, K. A., & Shortle, D. (1991) *Annu. Rev. Biochem. 60*, 795–825.

Dill, K. A., Alonzo, D. O. V., & Hutchinson, K. (1989) *Biochemistry 28*, 5439.

Eigenbrot, C., Randal, M., & Kossiakoff, A. A. (1990) *Protein Eng. 3*, 591–598.

Fauchere, J. L., & Pliska, V. (1983) *Eur. J. Med. Chem. 18*, 369–375.

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

Hopp, T. P., & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 3824–3828.

Hughson, F. M., & Baldwin, R. L. (1989) *Biochemistry 28*, 4415–4422.

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

Hynes, T. R., & Fox, R. O. (1992) *Biochemistry* (to be submitted).

Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577–2637.

Krug, R. R., Hunter, W. G., & Grieger, R. A. (1976) *J. Phys. Chem. 80*, 2335–2341.

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

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Lewis, P. N., Momany, F. A., & Scheraga, H. A. (1973) *Biochim. Biophys. Acta 303*, 211–229.

Lim, W. A., Farruggio, D. C., & Sauer, R. T. (1992) *Biochemistry 31*, 4324–4333.

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

Lyu, P. C., Liff, M. I., Marky, L. A., & Kallenbach, N. R. (1990) *Science 250*, 669–673.

MacArthur, M. W., & Thornton, J. M. (1991) *J. Mol. Biol. 218*, 397–412.

Matsumura, M., Becktel, W. J., & Matthews, B. W. (1988) *Nature 334*, 406–410.

Matsumura, M., Wozniak, J. A., Dao-pin, S., & Matthews, B. W. (1989) *J. Biol. Chem. 264*, 16059–16066.

McRee, D. E., Redford, S. M., Getzoff, E. D., Lepock, J. R., Hallewell, R. A., & Tainer, J. A. (1990) *J. Biol. Chem. 265*, 14234–14241.

Nakai, K., Kidera, A., & Kanehisa, M. (1988) *Protein Eng. 2*, 93–100.

Nozaki, Y., & Tanford, C. (1971) *J. Biol. Chem. 246*, 2211–2217.

O'Neil, K. T., & Degrado, W. F. (1990) *Science 250*, 646–651.

Pace, C. N., Laurents, D. V., & Thomson, J. A. (1990) *Biochemistry 29*, 2564–2572.

Pakula, A. A., & Sauer, R. T. (1990) *Nature 344*, 363–364.

Perry, K. M., Onuffer, J. J., Touchett, N. A., Herndon, C. S., Gittelman, M. S., Matthews, C. R., Chen, J., Mayer, R. J., Taira, K., Benkovic, S. J., Howell, E. E., & Kraut, J. (1987) *Biochemistry 26*, 2674–2682.

Perry, K. M., Onuffer, J. J., Gittelman, M. S., Barmat, L., & Matthews, C. R. (1989) *Biochemistry 28*, 7961–7968.

Piela, L., Nemethy, G., & Scheraga, H. A. (1987) *Biopolymers 26*, 1587–1600.

Ramanchandran, G. N., & Sasisekharan, V. (1968) *Adv. Protein Chem. 23*, 283–437.

Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167–339.

Richardson, J. S., & Richardson, D. C. (1988) *Science 240*, 1648–1652.

Rose, G. D., & Roy, S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 4643–4647.

Roseman, M. A. (1988a) *J. Mol. Biol. 200*, 513–522.

Roseman, M. A. (1988b) *J. Mol. Biol. 201*, 621–623.

Sandberg, W. S., & Terwilliger, T. C. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1706–1710.

Schellman, J. A. (1978) *Biopolymers 17*, 1305–1322.

Sharp, K. A., Nicholls, A., Friedman, R., & Honig, B. (1991) *Biochemistry 30*, 9686–9697.

Shirley, B. A., Stanssen, P., Steyaert, J., & Pace, C. N. (1989) *J. Biol. Chem. 264*, 11621–11625.

Shortle, D. (1986) *J. Cell. Biochem. 30*, 281-289.
Shortle, D., & Meeker, A. K. (1986) *Proteins: Struct., Funct., Genet. 1*, 81-89.
Shortle, D., & Meeker, A. K. (1989) *Biochemistry 28*, 936-944.
Shortle, D., Meeker, A. K., & Freire, E. (1988) *Biochemistry 27*, 4761-4768.
Shortle, D., Stites, W. E., & Meeker, A. K. (1990) *Biochemistry 29*, 8033-8041.
Sondek, J. E., & Shortle, D. (1990) *Proteins: Struct., Funct., Genet. 7*, 299-305.
Sondek, J. E., & Shortle, D. (1992) *Proteins: Struct., Funct., Genet. 13*, 132-140.
Stewart, D. E., Sarkar, A., & Wampler, J. E. (1990) *J. Mol. Biol. 214*, 253-260.
Tanford, C. (1970) *Adv. Protein Chem. 24*, 1-95.
Thomson, J. A., Shirley, B. A., Grimsley, G. R., & Pace, C. N. (1989) *J. Biol. Chem. 264*, 11614-11620.
Villafranca, J. E., Howell, E. E., Oatley, S. J., Xuong, N., & Kraut, J. (1987) *Biochemistry 26*, 2182-2189.
Yutani, K., Ogasahara, K., Tsujita, T., & Sugino, Y. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4441-4444.

Biphasic Effects of Alcohols on the Phase Transition of Poly(L-lysine) between α-Helix and β-Sheet Conformations†

Akira Shibata,*,‡ Miharu Yamamoto,‡ Takuya Yamashita,‡ Jang-Shing Chiou,§ Hiroshi Kamaya,§ and Issaku Ueda*,§

Faculty of Pharmaceutical Sciences, Tokushima University, Shomachi, Tokushima 770, Japan, and Anesthesia Service 112A, VA Medical Center, Salt Lake City, Utah 84148

Received November 8, 1991; Revised Manuscript Received April 10, 1992

ABSTRACT: Poly(L-lysine) exists as a random-coil at neutral pH, an α-helix at alkaline pH, and a β-sheet when the α-helix poly(L-lysine) is heated. The present Fourier-transform infrared (FTIR) study showed that short-chain alcohols (methanol, ethanol, and 2-propanol) partially transformed α-helix poly(L-lysine) to β-sheet when their concentrations were low. At higher concentrations, however, these alcohols reversed the reaction, and the alcohol-induced β-sheet was transformed back to α-helix structure. The reversal occurred at 1.40 M methanol, 0.96 M ethanol, and 0.55 M 2-propanol. The alcohol effects on the secondary structure were further investigated by circular dichroism (CD) on the thermally induced β-sheet poly(L-lysine). Methanol, ethanol, and 1-propanol, but not 1-butanol, shifted the negative mean-residue ellipticity at 217 nm of the β-sheet poly(L-lysine) to the positive side at low concentrations of the alcohols and to the negative side at high concentrations. With 1-butanol, only the positive-side shift was observed. The positive-side shift at low concentrations of alcohols indicates enhancement of the hydrophobic interactions among the side chains of the polypeptide in the β-sheet conformation. The negative-side shift indicates a partial transformation to α-helix. The shift from the positive to negative side occurred at 7.1 M methanol, 4.6 M ethanol, and 3.1 M 1-propanol. The alcohol concentrations for the β-to-α transition were higher in the CD study than in the IR study. This result suggests the multiplicity of the β-sheet conformations between the heat-induced and solvent-induced structures and also between the intramolecular and intermolecular β-sheets. The poly(L-lysine) concentration in the CD study was 0.5 mM where the β-structure is mainly intramolecular hydrogen bonding, whereas that in the IR study was 0.1 M where the hydrogen bonding is intermolecular as well as intramolecular. The biphasic effects coincide with those of alcohols on the main phase transition of dipalmitoylphosphatidylcholine (DPPC) membranes where methanol, ethanol, and 1-propanol, but not 1-butanol, decreased the transition temperature at low concentrations and increased it at high concentrations. The similarity between the alcohol concentrations that induce biphasic effect in the DPPC phase transition and in the β-to-α transition in the IR study suggests that the intermolecular hydrogen bonds are susceptible to alcohol perturbations. These alcohol actions on macromolecules are essentially solvent effects.

The biphasic effects of short-chain alcohols (methanol, ethanol, and 1-propanol) on macromolecular structures received recent research interest since the discovery of the "interdigitated" state in the phospholipid bilayer structure (McDaniel et al., 1983; Simon & McIntosh, 1984; Rowe, 1985; Nambi et al., 1988; Rowe & Cutrera, 1990). These alcohols depress the temperature of the main phase transition of lipid membranes between the solid-gel and liquid-crystalline phases, and disorder (or fluidize) the membrane. At extremely high concentrations, however, the alcohols shorter than four or five carbon atoms elevate the main transition temperature of the dipalmitoylphosphatidylcholine (DPPC) vesicle membranes (Jain & Wu, 1977).

We (Shibata et al., 1982, 1984, 1991) reported anesthetic effects upon the secondary structure of polypeptides and showed that volatile anesthetics partially transformed α-helix poly(L-lysine) to the β-sheet conformation. The present report deals with alcohol effects on the secondary structure of poly(L-lysine), where short-chain alcohols promoted the β-sheet conformation at low concentrations, but they supported the α-helix conformation when the concentrations exceeded the

† Supported in part by VA Medical Research and by NIH Grants GM25716 and GM27670.
‡ Tokushima University.
§ VA Medical Center.

0006-2960/92/0431-5728$03.00/0