Energetic Contribution of Side Chain Hydrogen Bonding to the Stability of Staphylococcal Nuclease†

Michael P. Byrne, R. Lee Manuel, Laura G. Lowe,‡ and Wesley E. Stites*

Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, Arkansas 72701-1201

Received May 30, 1995; Revised Manuscript Received August 23, 1995®

ABSTRACT: Hydrogen bonds are a ubiquitous feature of protein structures, yet there is great uncertainty about the energetic contribution of hydrogen bonding to protein stability. This study addresses this question by making a series of single substitution mutations in the model protein staphylococcal nuclease. These mutants have had a residue capable of participating in hydrogen bonding either removed or introduced. The variants we have investigated are as follows: nine valines substituted with threonine and serine; eight threonines converted to valine, serine, and cysteine; and seven tyrosines replaced by phenylalanine and leucine. The stabilities of these 56 mutant proteins were determined by titration with guanidine hydrochloride using fluorescence as a probe of structure. In general, it was found that the stability effects of removing a hydrogen bonding residue and replacing it with a nonbonding residue were relatively small. This was true even in the case of buried residues participating in hydrogen bonds, where the substituted residue leaves an unfulfilled hydrogen bond in the hydrophobic core. In contrast, introducing a hydrogen bonding residue in place of a nonbonding residue was generally more costly energetically. A wide variability in the cost of burying a hydroxyl was observed, but this does not seem to be due to differences in hydrogen bonding. The overall energetic contribution of various wild-type hydrogen bonding interactions was evaluated as being favorable. A range of energies from approximately 1.5 to 4.0 kcal/mol was estimated for the contribution of these interactions to the stability of the native state.

The importance of the hydrogen bond in biomolecular systems is difficult to overstate. Hydrogen bonding is certainly important in the structure of proteins, as every amide in the main chain is a potential hydrogen bond acceptor and, except for proline, a donor. Many side chains are capable of hydrogen bond participation as well. It has been determined that proteins contain an average of 1.1 intramolecular hydrogen bonds per residue (Stickle et al., 1992). Also, nearly all nitrogen and oxygen atoms in proteins participate in hydrogen bonds either to solvent or to other protein atoms. Baker and Hubbard (1984) demonstrated that only 12.4% and 11.2% of main chain NH and CO groups, respectively, failed to participate in a hydrogen bond, while McDonald and Thornton (1994) reported that only 6.4% and 21.8% of potential side chain donors and acceptors failed to participate in a hydrogen bond. Specific hydrogen bonding patterns are primarily responsible for forming protein secondary structure as helices and sheets are held together through hydrogen bonding interactions. But, while certainly important, the magnitude and even the sign of the energetic contribution of hydrogen bonding to protein stability is still open to debate.

A hydrogen bond occurs when a hydrogen atom is shared by two electronegative atoms. The strength of this interaction depends on the orientation of the atoms involved as well as the electronegativity of the acceptor and donor atoms. The energetic magnitude of a hydrogen bond has been estimated at 2−10 kcal/mol (Pauling, 1960; Jeffrey & Saenger, 1991). Despite extensive studies of model organic compounds (Doig & Williams, 1992), there is still great uncertainty about the energetic importance of hydrogen bonds in proteins. As Dill (1990), speaking of hydrogen bonds and van der Waals interactions, said, "... the magnitudes, among all the types of forces contributing to protein folding, are currently perhaps the most difficult to assess." It has been generally held that an intermolecular hydrogen bond (between the protein backbone or side chain and water) was energetically very similar to an intramolecular hydrogen bond (between different parts of the protein). Since proteins contain hundreds of hydrogen bonds, even a small uncertainty in either of these values is multiplied into a significant error in the assessment of the energetic contribution of hydrogen bonds to protein stability.

The energetic values of the hydrogen bonds in protein native and denatured states are difficult to measure directly, so researchers have attempted to measure them indirectly through stability studies of mutant proteins (Alber et al., 1987; Grutter et al., 1987; Eijsink et al., 1991; Mårtensson et al., 1992; Thunnissen et al., 1992; Derst et al., 1992; Eijsink et al., 1992; Serrano et al., 1992; Shirley et al., 1992; Nicholls et al., 1993; Berndt et al., 1993; Pjura & Matthews, 1993; Blaber et al., 1993; Chen et al., 1993; Blanchard et al., 1994; Jackson & Fersht, 1994; Yamada et al., 1994; Pedersen et al., 1994; Hammen et al., 1995; Pace, 1995). The primary problem with this approach lies in the fact that even the most conservative mutation aimed at changing hydrogen bonding character will also introduce at least minor changes in properties such as residue size, hydrophobicity, and conformational entropy. Further, each site of substitution is a unique environment that will greatly influence the effects

† Support for this work was provided by the Army Research Office and the ACS Petroleum Research Fund.

* Corresponding author.

‡ Present address: Department of Biochemistry, Vanderbilt University School of Medicine, Nashville TN 37232.

® Abstract published in *Advance ACS Abstracts,* October 1, 1995.

0006-2960/95/0434-13949$09.00/0

of substitution through differences in hydrogen bonding potential, packing, secondary structure, and so on. However, even with these shortcomings mutational analysis can offer important insight.

One way to attempt to compensate for these complications is to make a large set of related substitution mutants. By comparison of the differing effects of the same substitution in different residue environments (e.g., Val to Thr at nine different positions), an assessment of the relative importance of various environmental factors can be made. Similarly, by substitution of different side chains at each site (e.g., Val to Thr, Ser, and Ala at the same position), an assessment of the impact of different side chain properties in each environment can be made. This report investigates the energetic contribution of side chain hydrogen bonding to the stability of the model protein staphylococcal nuclease (nuclease). Nuclease is monomeric, with a single domain that folds and unfolds reversibly. Of the 149 amino acids that comprise nuclease there is one tryptophan and no cysteines. The single tryptophan provides a sensitive fluorescence probe of structure, and the lack of cysteines means no disulfides are present. Our study focuses on the three hydroxyl-containing side chains, serine, threonine, and tyrosine, by substituting them into or removing them from the nuclease structure. Nine valines were singly substituted with threonine and serine. Threonine was replaced with valine, serine, and cysteine in eight positions. Seven positions containing tyrosines in wild-type were individually replaced with phenylalanine and leucine.

MATERIALS AND METHODS

Mutagenesis. All mutants of staphylococcal nuclease were prepared by the method of Kunkel et al. (1985, 1987) as previously described (Shortle et al., 1990). Mutated M13 phage DNA was used to transform DH5αF′ competent cells (Gibco BRL). Single-stranded DNA was isolated from progeny phage and sequenced to screen for specific mutations. To ensure that no other mutations occurred elsewhere in the gene, the entire nuclease gene of each mutant was sequenced. The mutant genes were then transferred to the overexpression plasmid pL13 by restriction digest and ligation followed by transformation into *Escherichia coli* strain AR120.

Protein Expression and Purification. Protein expression and purification was modified from the procedure of Shortle et al. (1990). *E. coli* strain AR120 (Mott et al., 1985) carrying the desired mutant plasmid was grown overnight at 37 °C in SB media (24 g of yeast extract, 12 g of tryptone, 2.3 g of KH_2PO_4, 12.5 g of K_2HPO_4, and 4 mL of glycerol per liter). The cell pellet was collected by centrifugation and then resuspended, using one-tenth volume of the original culture, in a wash of 6 M urea, 25 mM Tris-HCl, and 5 mM EDTA, pH 8.1. The cells were again centrifuged, the supernatant was discarded, and the pellet was resuspended in 6 M urea, 200 mM NaCl, 25 mM Tris-HCl, and 5 mM EDTA, pH 8.1. This suspension was shaken gently on ice for 30 min and centrifuged, and the supernatant was recovered. To this extract was added an equal volume of cold ethanol. After 3−4 h at −20 °C the precipitate was centrifuged out and discarded. An additional 2 volumes of cold ethanol was added to the supernatant and held at −20 °C for 1−2 h. The precipitate was centrifuged out, dried briefly, and then resuspended in one-fourth volume of the above extraction buffer without NaCl. This solution was then loaded onto an S-Sepharose Fast Flow (Pharmacia) column (1 ml bed/250 mL culture) equilibrated with the same buffer. After being washed with the loading buffer, nuclease was eluted in the same buffer made 250 mM in NaCl. At this step, the protein was concentrated by precipitation with 3 volumes of ethanol. After centrifugation the pellet was dissolved in a small volume of the eluting buffer and then dialyzed first against 1 M NaCl/10 mM Tris-HCl, pH 7.0, for several hours and then 100 mM NaCl/25 mM sodium phosphate, pH 7.0, overnight, at 4 °C. This procedure yields 10−20 mg of nuclease per 250 mL culture with purities greater than 98% as estimated by SDS gels.

Guanidine Hydrochloride Denaturation. The single tryptophan at position 140 was used as a fluorescence probe to follow the guanidine hydrochloride denaturation of substitution mutants of staphylococcal nuclease. The denaturations were carried out in a buffer consisting of 25 mM sodium phosphate and 100 mM sodium chloride, pH 7.0, at 20 °C, with a protein concentration of 50 μg/mL. Tricarboxyethyl phosphine (Molecular Probes, Eugene, OR, 0.5 mM) was added to the samples of cysteine substitution mutants to reduce any disulfides (Burns et al., 1991). An Aviv model ATF-101 fluorometer was used to carry out the titration of the proteins as described elsewhere (Stites et al., 1995). Data analysis was carried out as previously published (Stites et al., 1995; Shortle et al., 1990), assuming a two-state model for reversible denaturation. Proteins which did not show a clear native baseline (I_N) were renatured with ammonium sulfate to obtain a value for I_N which is used in the regression analysis. The values of I_N calculated from ammonium sulfate renaturation and I_N calculated from linear extrapolation of the transition region agreed within 10% in all cases. This data analysis yields three parameters: the stability difference in the absence of denaturant between a protein's native and denatured states (ΔG_{H_2O}), the rate of change of free energy with respect to GuHCl concentration (m_{GuHCl} or $d(\Delta G)/d[GuHCl]$), and the concentration of guanidinium hydrochloride at which the protein is half denatured (C_m). In order to compare data obtained in our laboratory and that obtained in the Shortle lab (Shortle et al., 1990; Green et al., 1992), we have made independent determinations of stability for wild-type and several mutants. We have found the differences in values between the two laboratories are within our stated experimental error.

Hydrogen Bond Identification. Hydrogen bonds were identified in the Loll and Lattman (1989) crystal structure of staphylococcal nuclease using the program HBPLUS (McDonald, 1992). The results of this program were kindly provided by Dr. Janet Thornton. Hydrogen bonds for the Hynes and Fox (1991) structure are identified in that work.

RESULTS

A total of 56 single site substitution mutants of staphylococcal nuclease were made. All nine valines of nuclease were individually replaced with threonine and serine. Threonines were substituted with valine, serine, and cysteine. Two of the ten threonines in nuclease, position numbers 2 and 4, were not substituted as the first six residues are completely disordered in X-ray structures (Loll & Lattman, 1989; Hynes & Fox, 1991), and previous mutations in this region have

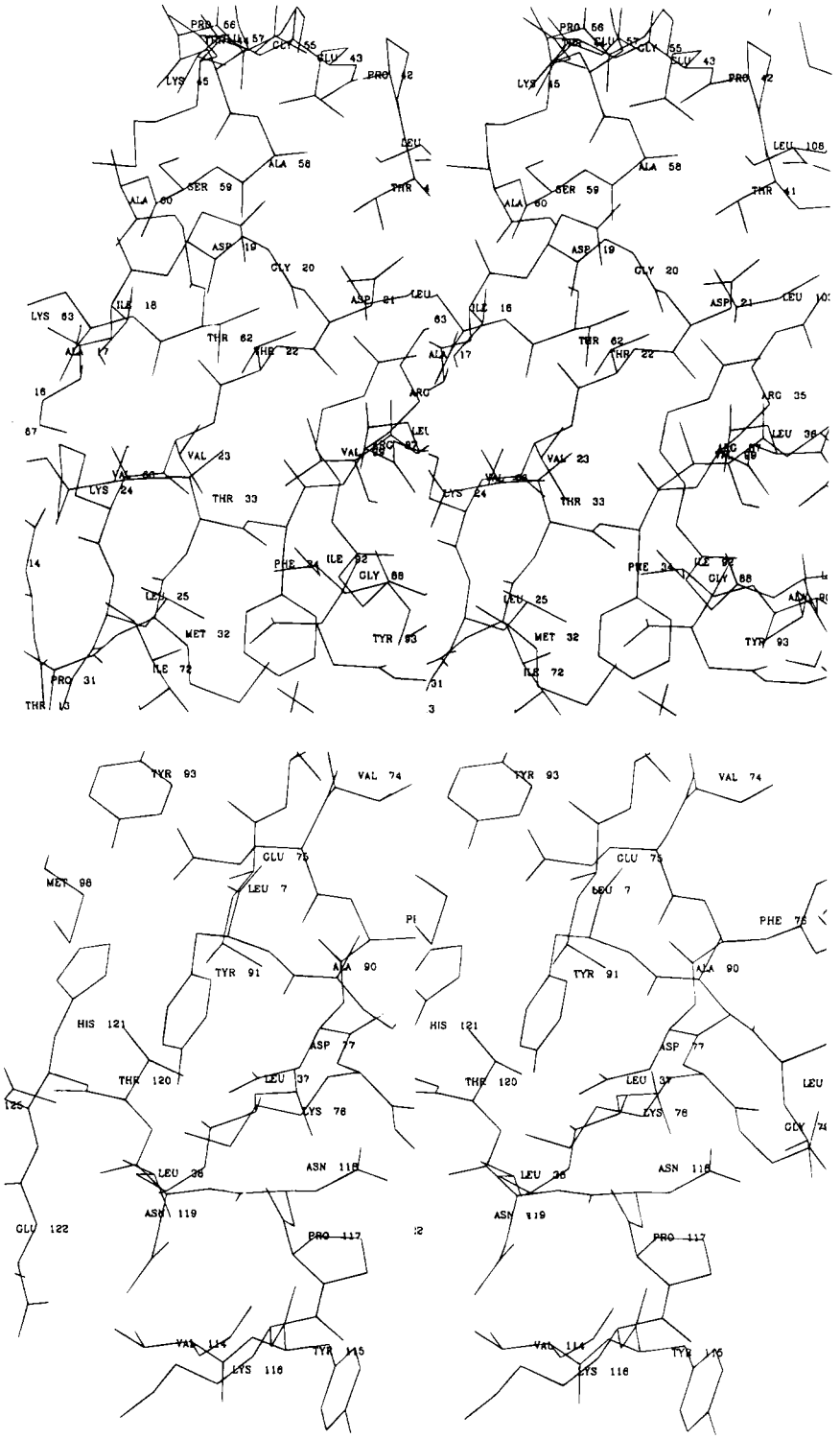

FIGURE 1: Stereodiagrams showing environments of various valine, threonine, and tyrosine residues (Loll & Lattman, 1989). (A, top) T22 is partially solvent exposed, but its hydroxyl hydrogen bonds with the carboxylate of D21. T62 is fully buried and participates in a bifurcate hydrogen bond with the carbonyls of A58 and G20. V23, V66, and V99 are all fully buried with no potential hydrogen bonding partners other than the residue's own main chain within distance. T33 is partially solvent exposed and has no apparent hydrogen bonding interactions. (B, bottom) The hydroxyl group of T120 is buried and participates in a bifurcate hydrogen bond with the amine of K78 and the carboxylate of D77. Y91 is completely buried and participates in a bifurcate hydrogen bond with the side chain of D77 and the amide nitrogen of H121. Y93 has slight solvent exposure and is involved in a hydrogen bond with the side chain E75. V114 is partially solvent exposed. V74 is fully buried with no potential hydrogen bonding partners other than the residue's own main chain within distance.

had little stability impact (Shortle et al., 1990; Green et al., 1992). All seven tyrosines were substituted with phenylalanine and leucine. The structural environments of some of these residues (Loll & Lattman, 1989) are shown in Figure 1.

The stabilities of the alanine and glycine substitutions at these residues as well as at the four serines in the wild-type protein have been published (Shortle et al., 1990; Green et al., 1992). During the course of this work, we became convinced that the published stability of the alanine mutant,

Table 1: Free Energy Differences of Important Forces

mutation	volume difference (Å^3)[a]	hydrophobicity difference (kcal/mol)[b]	main chain entropy difference (kcal/mol)[c]	side chain entropy difference (kcal/mol)[d]	total free energy change (kcal/mol)
V→T	−12	−1.28	−0.28	−0.94	−2.50
V→S	−32	−1.68	−0.29	−1.05	−3.02
T→S	−20	−0.40	−0.01	−0.11	−0.52
T→C	−7	1.71	0.00	−0.07	1.64
Y→F	−6	1.11	−0.01	0.41	1.51
Y→L	−17	0.99	0.33	0.34	1.56
S→A	−6	0.47	0.43	1.08	1.98

[a] Calculated from volumes enclosed by van der Waals radius (Bondi, 1964). [b] Calculated at 20 °C from data on the partitioning of *N*-acetylamide amino acids between octanol and water (Fauchère & Pliška, 1983). [c] Difference in the energetic contribution of main chain conformational entropy of the two residues at 20 °C calculated from Stites & Pranata (1995). [d] Difference in the energetic contribution of side chain conformational entropy of the two residues at 20 °C calculated from Lee et al. (1994).

Table 2: Solvent Accessibility and Environment for Valines, Threonines, and Tyrosines

position	fraction solvent accessibility[a]	secondary structure[b]	hydrogen bonding partners[c]	position	fraction solvent accessibility[a]	secondary structure[b]	hydrogen bonding partners[c]
V23	0.01	E		T41	0	N	H_2O, 3.01; H_2O, 3.35
V39	0	E		T44	0.18	S	D19 δO*, 2.87; 1 H_2O, 2.94
V51	0.64	S		T62	0	H	G20 CO, 2.54; A58 CO, 3.16
V66	0	H		T82	0.73	B	
V74	0	E		T120	0.34	T	D77 δO, 2.44; K78 ζN, 2.91
V99	0	H		Y27	0.22	E	E10 εO, 2.83
V104	0	H		Y54	0.25	T	S141 γO*, 3.41
V111	0	E		Y85	0.65	T	K84 ζN*, 2.88
V114	0.24	N		Y91	0	E	D77 δO, 2.56; H121 NH, 3.29
T13	0.89	E		Y93	0.10	E	E75 εO, 2.60
T22	0.14	E	D21 δO, 3.08	Y113	0.19	N	H_2O, 2.58
T33	0.16	E		Y115	0.60	B	

[a] Fraction of side chain that is solvent accessible compared to the solvent-accessible surface area of side chain in the model tripeptide Gly-Xxx-Gly in extended conformation. [b] As determined by the method of Kabsch and Sander (1983): E, β-sheet; S, bend; H, α-helix; T, hydrogen bonded turn; B, isolated β-bridge; N, no secondary structure assigned. [c] Hydrogen bonding partners of the side chain in the Loll and Lattman structure (1989) as determined by the program HBPLUS (McDonald, 1992). The donor−acceptor distance in angstroms immediately follows. Hydrogen bonding partners marked with an asterisk (*) are those identified by Hynes and Fox (1992) and are interactions present only in that structure.

V111A (Shortle et al., 1990), was low. We made this mutant again and determined its stability. The corrected values for this mutant are as follows: ΔG_{H_2O} of 1.4 kcal/mol, C_m of 0.24 M, and m_{GuHCl} of 0.84 relative to wild-type. These values are used throughout this work.

Although the mutations we have made are relatively conservative, properties other than side chain hydrogen bonding have been changed. Table 1 summarizes the changes in values of various side chain properties. Table 2 describes the hydrogen bonding environment and solvent accessibility for each of the positions in this study as well as the type of secondary structure.

A summary of results from denaturations of these proteins are shown in Table 3. The difference in free energy (ΔG_{H_2O}) between the native and denatured macrostates ranges from 6.4 kcal/mol for T41V to somewhere less than −3 kcal/mol for Y27L. Slope values range from 1.30 times the wild-type value in the case of V23S to 0.83 times the wild-type value for V111S.

Substitutions for Tyrosines. The mutation to phenylalanine removes hydrogen bonding capability and increases hydrophobic transfer free energy while largely retaining side chain size and geometry. The mutation to leucine also removes hydrogen bonding capability and has almost exactly the same change in hydrophobic transfer free energy as the phenylalanine mutation. The size of the side chain is greatly reduced, but to a first approximation side chain geometry is retained, with both residues branching at the γ-carbon.

Three of the positions, 85, 113, and 115, show little change in stability upon substitution with either phenylalanine or leucine. This agrees with earlier work (Shortle et al., 1990) where substitution with alanine and glycine resulted in small stability changes. All of these residues are in the active site and largely solvent exposed. Y115 can form hydrogen bonds with solvent but has no crystallographically discernible hydrogen bonding interactions. Y113 shows a hydrogen bond to a crystallographic water. Y85 interacts with K84, but only in the unliganded structure (Hynes & Fox, 1991).

Positions 27 and 54 are approximately 0.5 kcal/mol less stable than wild-type when substituted with phenylalanine. The phenolic hydroxyl of Y27 interacts with the carboxylate of E10 while Y54 has a long range (3.26 Å) interaction with the hydroxyl of S141. While the majority of the aromatic ring is buried in both of these residues, the hydroxyl moiety is solvent accessible. Even though the mutation to phenylalanine is only slightly destabilizing, mutation to leucine, alanine, or glycine has a much greater effect. Because leucine has similar hydrophobic transfer energy compared to the phenylalanine, differences in side chain geometry and size must result in a loss of energetically significant interactions as compared to the phenylalanine. The total loss of detectable native structure suffered upon leucine substitution at position 27 is rather puzzling, as even the glycine mutant still had detectable amounts of folded protein. This may or may not be related to the fact that making the Y27L mutation proved to be extraordinarily difficult. For unknown

Table 3: Denaturation Data at 20 °C

mutant	m_{GuHCl}[a]	C_m[b]	ΔG_{H_2O}[c]
V23T	1.17	0.31	2.4
V39T	1.08	0.56	4.3
V51T	1.00	0.86	5.8
V66T	1.02	0.61	4.2
V74T	1.14	0.24	1.8
V99T	1.07	0.32	2.3
V104T	0.98	0.47	3.1
V111T	1.05	0.47	3.3
V114T	1.02	0.78	5.3
V23S	1.30	0.11	0.9
V39S	1.03	0.46	3.2
V51S	0.96	0.85	5.5
V66S	1.04	0.37	2.5
V74S	1.06	0.02	0.2
V99S	1.15	0.07	0.5
V104S	0.87	0.11	0.7
V111S	0.83	0.14	0.8
V114S	1.02	0.82	5.6
Y27F	1.01	0.74	5.0
Y54F	0.96	0.79	5.1
Y85F	1.01	0.83	5.6
Y91F	0.94	0.51	3.2
Y93F	0.92	0.59	3.6
Y113F	0.98	0.86	5.6
Y115F	0.96	0.85	5.5
Y27L			<−3
Y54L	0.95	0.35	2.2
Y85L	1.02	0.80	5.5
Y91L	0.96	0.27	1.7
Y93L	1.07	0.15	1.1
Y113L	1.00	0.86	5.8
Y115L	0.97	0.81	5.3
T13S	1.02	0.77	5.3
T22S	1.06	0.70	5.0
T33S	1.03	0.62	4.3
T41S	0.92	0.73	4.5
T44S	0.99	0.84	5.6
T62S	1.05	0.50	3.5
T82S	1.02	0.71	4.9
T120S	0.99	0.75	5.0
T13V	0.96	0.81	5.2
T22V	0.99	0.71	4.7
T33V	0.98	0.92	6.0
T41V	0.97	0.98	6.4
T44V	1.00	0.84	5.7
T62V	0.91	0.89	5.4
T82V	0.99	0.87	5.8
T120V	0.93	0.61	3.8
T13C	0.92	0.68	4.2
T22C	1.02	0.68	4.7
T33C	0.94	0.70	4.4
T41C	0.96	0.96	6.2
T44C	0.94	0.85	5.4
T62C	0.98	0.68	4.5
T82C	1.03	0.79	5.5
T120C	0.94	0.61	3.9
WT	1.00	0.83	5.6

[a] Slope value (change in free energy with respect to change in guanidine hydrochloride concentration). Expressed relative to wild-type value of 6.69 kcal/(mol·M). Error is estimated to be ±0.02. [b] Midpoint concentration (concentration of guanidine hydrochloride at which half of the protein molecules are denatured) in units of molar. Error estimated to be ±0.01 M. [c] Stability of protein to reversible denaturation in units of kcal/mol. Error estimated to be ±0.1 kcal/mol.

reasons, isolating the DNA encoding this mutant took numerous attempts, far more than usual. When the first protein preparation proved to be denatured, two more attempts to move the mutant gene into the expression vector and express and purify protein were made. All three independent preparations proved to be completely denatured.

Of the seven tyrosines of nuclease, positions 91 and 93 (Figure 1B) are the most buried. Y91 is completely buried and participates in a bifurcate hydrogen bond with the side chain of D77 and the amide nitrogen of H121. D77 is completely buried and is involved in an extensive hydrogen bonding network. One oxygen of the carboxylate interacts with the ϵ-amine of K78 and the hydroxyl of T120 while the other interacts with the main chain amide of T120 and the hydroxyl of Y91. Y93 has slight solvent exposure and is involved in a hydrogen bond with the side chain of E75. The oxygen of E75 that interacts with Y93 is buried while the other oxygen of the carboxylate is solvent exposed. Y91F has lost 2.4 kcal/mol of stability relative to wild-type while Y93F has lost 2.0 kcal/mol. Y93L has lost 4.5 kcal/mol, but Y91L has lost slightly less, 3.9 kcal/mol. The alanine and glycine mutants (Shortle et al., 1990; Green et al., 1992) are significantly less stable than the leucine mutants.

Substitutions for Valines. Positions 51 and 114 show little stability change upon substitution with threonine, serine, alanine, or glycine. Position 51 is partially solvent exposed, and its behavior is similar to that seen for the solvent exposed tyrosines. V114 is near the active site and flanked by tyrosines 113 and 115. While the liganded crystal structure (Loll & Lattman, 1989) shows partial solvent accessibility for V114, the unliganded crystal structure (Hynes & Fox, 1991) shows Y113 and Y115 rotated away from the active site and packing against V114 and thus virtually preventing solvent access to this residue. The limited stability change upon mutation implies that the hydrophobic interaction of V114 with Y113 and Y115 is of limited energetic importance.

V111 is buried but could potentially become solvent exposed with minor movement of main and side chains. Nevertheless, serine or glycine mutants at this site are all destabilized by about the same extent, approximately 5 kcal/mol. The alanine mutant is slightly less destablized. The threonine substitution is also destabilized but significantly less so, being down by only 2.3 kcal/mol. Notably, the serine, alanine, or glycine mutations are $m-$, down in slope relative to wild-type, while V111T is $m+$, up slightly in slope from the wild-type value.

The remaining six valines are deeply buried. They can be divided into two classes. V39 and V66 (Figure 1A) show relatively slight loss of stability, 1.3 and 1.4 kcal/mol, respectively, upon substitution with threonine. Serine, alanine, and glycine substitution results in greater destabilization but not excessively so. The remaining four positions, 23 (Figure 1A), 74, 99 (Figure 1A), and 104, show much greater stability losses than the corresponding substitutions at 39 and 66. The crystal structures show no potential hydrogen bonding partners other than the residue's own main chain within distance for any of these positions.

Substitutions for Threonines. T13 and T82 are largely solvent exposed and show little change in stability upon substitution with valine. Substitution with cysteine, serine, alanine, and glycine tend to cause greater changes in stability but not in exactly the same fashion at each site. T13S is about as stable as T13V while T13C is down a little over one kcal/mol relative to these two mutants. On the other hand, T82C and T82V have roughly the same stability while T82S is destabilized by almost a kcal/mol relative to the other two mutants. T44 is also partially exposed and shows

hydrogen bonding interactions with the carboxylate of D19 and with a crystallographic water. Despite decreased solvent exposure compared to T13 and T82, this site shows even less sensitivity to substitution. T44V and T44S have essentially unchanged stabilities while T44C, T44A, and T44G are down only slightly.

T33 (Figure 1A) and T41 show related behavior. T33 is partially exposed to solvent and can obviously form hydrogen bonds with solvent but T41 is solvent inaccessible. While T41 is inaccessible in terms of actual van der Waals contact, it still has sufficiently strong interactions to order two crystallographic waters. Valine substitution at both these sites results in a substantial increase in stability relative to the wild-type protein. T41S and T33S are down in stability from the wild-type protein but are approximately equal (T33) or worse (T41) in stability than the corresponding alanine mutants.

The remaining threonines, 22 (Figure 1A), 62 (Figure 1A), and 120 (Figure 1B), have hydrogen bonds to other parts of the protein. Despite the common factor of hydrogen bonds, the stability effects of similar substitutions vary markedly. T22 is partially solvent exposed, but its hydroxyl interacts with the buried carboxylate of D21. Substitution of a serine at this site results in a loss of 0.6 kcal/mol of stability. Substitution with a valine results in the loss of 0.9 kcal/mol, with alanine and glycine resulting in even greater stability loss. T62 is completely buried in the major hydrophobic core of the protein where it donates a hydrogen bond to the carbonyls of A58 and G20. Here substitution with serine, alanine, glycine, or cysteine results in loss of 2.5 kcal/mol or more of stability. Valine substitution results in a protein with essentially the same stability as wild-type. The methyl group of T120 is solvent exposed, and the hydroxyl group is a bridge between the amine of K78 and the carboxylate of D77, being hydrogen bonded to both. The other carboxylate oxygen of D77 interacts with Y91. Valine and cysteine substitution of T120 result in similar losses of 1.8 and 1.7 kcal/mol of stability respectively relative to wild-type. The serine and alanine mutants are actually more stable than the cysteine or valine mutants.

Broadly speaking, the threonines can be divided into two categories. In one set of positions, the serine substitution is nearly identical in stability or even more stable than the valine substitution. This would include positions 13, 22, 44, and 120. In the second set of positions, the valine substitution is significantly more stable than the serine substitution. This would include positions 33, 41, 62, and, to a lesser extent, 82.

Correlations of Stability Changes. Stability information on the alanine and glycine substitutions at valine, tyrosine, and threonine sites is available from previously published work (Shortle et al., 1990; Green et al., 1992). The comparison of these data with those presented here is most informative. As noted in the earlier work, the stability of the glycine and alanine substitutions at a given position correlate well with each other. Generally speaking, both of these substitutions also correlate well with the stability of the mutations examined here. This is demonstrated in Figure 2A where the free energies of the tyrosine to phenylalanine and leucine substitution mutants have been plotted against the free energies of the alanine mutants at the same positions. Similarly, the free energies of serine and threonine substitution mutants at valines have been plotted against the free energies of the alanine mutants at the same positions in Figure 2B. In both sets of mutations the relative stability of any one type of mutation is a good predictor of the relative stability of other types of mutations.

The situation for substitutions at threonine is not so simple. The free energies for the substitutions of serine and valine for the threonines have been plotted against free energies of alanine substitutions at those positions in Figure 2C. There seems to be little correlation between either the stabilities of the serine or valine substitutions at the threonines with the corresponding alanine stabilities.

Another way of comparing the valine and threonine sites is to examine the differences in free energy between similar side chain pairs. The free energy difference between threonine and valine ($\Delta\Delta G_{Thr \rightarrow Val}$) at a given site is plotted against the free energy difference between serine and alanine ($\Delta\Delta G_{Ser \rightarrow Ala}$) at that same site in Figure 3. In other words, Figure 3 compares the energetic cost of substituting a methyl with a hydroxyl at a given site with the energetic cost of substituting a hydrogen with a hydroxyl.

It is striking that, in nearly every case, the presence of a hydroxyl is energetically neutral or destabilizing. Only T22 and T120 show a consistent and pronounced preference for a hydroxyl. These two threonines participate in well defined hydrogen bonding interactions with other parts of the protein. At the remaining site where a threonine forms intramolecular hydrogen bonds, T62, there is very little energetic difference between serine and alanine or between threonine and valine.

DISCUSSION

There are two questions we are particularly interested in addressing. First, we wish to know what typical free energies of hydrogen bonding interactions in a protein are. This is clearly an important number that can best be assessed by mutational studies. A second, closely related question is what is the cost of burying an unsatisfied hydrogen bond donor/acceptor. This question can be approached by studies of the transfer of various small molecules between aqueous and hydrophobic solvents. The cost of burying an unsatisfied donor or acceptor should be an implicit part of the hydrophobic transfer free energy for a given side chain with hydrogen bonding capability. But this may be true only for hydrophobic transfer free energies measured in hydrophobic solvents such as cyclohexane that lack hydrogen bonding capacity and have small amounts of water dissolved (Pace, 1995; Radzika & Wolfenden, 1988). Experimental measurement in proteins of the cost of burying unsatisfied hydrogen bonding groups seems wise.

It is clear that leaving hydrogen bond donors or acceptors unsatisfied in a nonpolar environment will be unfavorable. The only question is the exact magnitude of the energy, not its sign. Considered in the larger context of protein folding and unfolding, a much more difficult question is determining the free energy contribution of a particular hydrogen bonding interaction in a protein. Not only do we need to know the intramolecular interaction free energy for a given hydrogen bond pair in the native state, but we must also know the interaction free energies of both the donor and acceptor with the solvent in the denatured state. The difference between intermolecular and intramolecular interactions is the net contribution of the hydrogen bonding interaction to the stability of the native state. It is not clear from simple

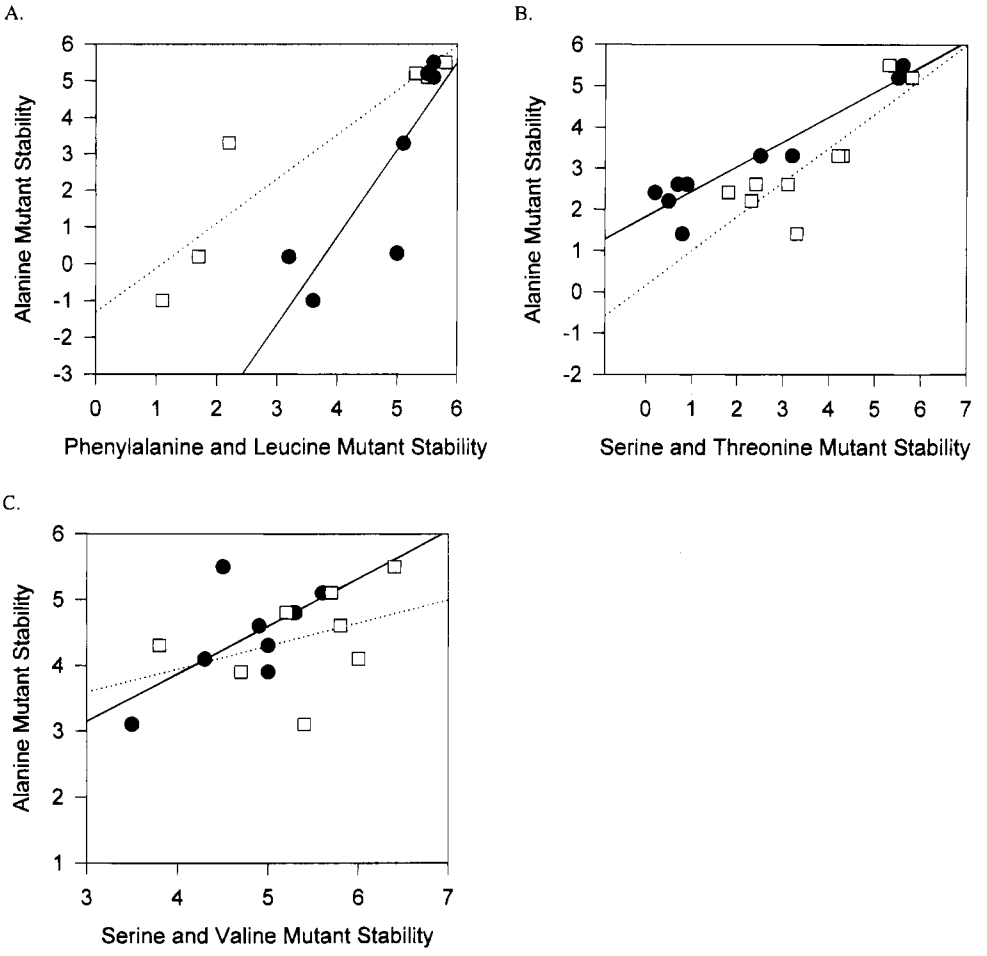

FIGURE 2: Correlations showing regression lines for free energies of various substitutions at the same positions. (A) Free energies of the tyrosine to phenylalanine (●) and leucine (□) mutations are plotted against the free energies of the alanine substitutions at these same positions. (B) Free energies of the valine to threonine (□) and serine (●) mutations are plotted against the free energies of the alanine substitutions at these same positions. (C) Free energies of the threonine to valine (□) and serine (●) mutations are plotted against the free energies of the alanine substitutions at these same positions.

considerations if hydrogen bonding interactions in proteins are stabilizing, destabilizing, or energetically neutral.

Of course, each hydrogen bonding interaction will be different, and the simplest, most conservative mutation made to study it will introduce multiple changes in factors other than hydrogen bonding. Many of these factors we simply cannot evaluate at this time. In earlier work (Shortle et al., 1990) it was shown that when the large hydrophobic side chains in nuclease were substituted with alanine and glycine, a wide range of stabilities were found in the mutants. Even when such obvious factors such as degree of residue burial and residue secondary structure were considered, it was not possible to easily account for the full range of stability effects. In attempting to analyze a dataset such as this, with the added complication of hydrogen bonding interactions, the first reaction may be a profound sense of despair. However, we can use the large number of mutants to, in effect, remove some factors from consideration.

Correlations of Mutant Stabilities. The variability in the stabilities of the mutants studied here once again point out the overwhelming importance of environment in determining the stability effects of a given mutation. Each environment offers a unique set of factors such as packing arrangements, van der Waals interactions, polarity, and so forth that affect each mutation. We do not necessarily have to identify what these factors are to predict roughly the effects of other mutations. Although what causes the variability of stabilities from site to site is not obvious, it is clear, as evidenced by Figures 2A, 2B, and 3, that there is a consistent pattern among different side chains at each site.

The exceptions to this are the mutations at threonine as depicted in Figure 2C. Although there is no simple pattern immediately apparent in this plot based on the presence or absence of hydrogen bonding interactions, these interactions do indeed appear to be responsible for much of the lack of correlation. At T22 and T120 the loss of hydrogen bonding is destabilizing. Loss of hydrogen bonding is less important at T62. The removal of poorly satisfied hydrogen bonding capacity at T41, T33, and T82 is stabilizing. It is still rather puzzling that the alanine mutations do not correlate better with the valine mutations, both of which remove hydrogen bonding capacity. This may to be due to the differing degrees to which the hydrophobic interactions of the valine methyl do or do not replace the hydrogen bonding interactions of the threonine hydroxyl.

If the lack of correlation in Figure 2C is rationalized by differences in hydrogen bonding, the generally good correlation of free energy changes in Figures 2A, 2B, and 3 becomes rather more noteworthy. The energetic effect of any mutation depends upon the site at which it is made more than any other factor. This in itself is not surprising, but the high degree of correlation between the energetic effects

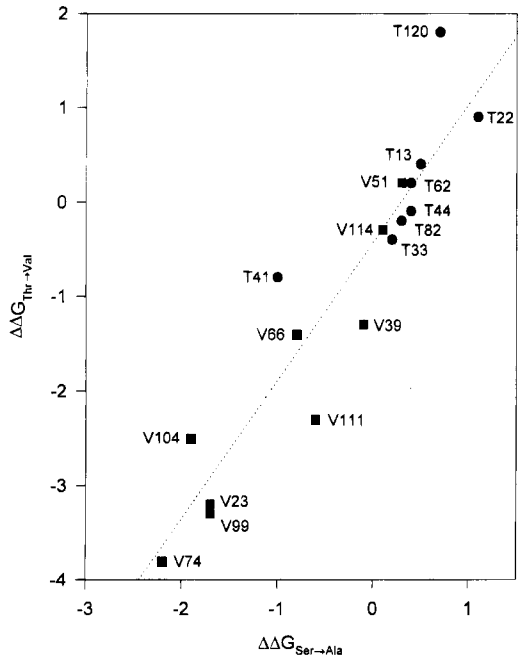

FIGURE 3: Plot of free energy difference between valine and threonine at a given wild-type threonine or valine position ($\Delta\Delta G_{Val\rightarrow Thr} = \Delta G_{Thr} - \Delta G_{Val}$) against the free energy difference between alanine and serine ($\Delta\Delta G_{Ala\rightarrow Ser} = \Delta G_{Ser} - \Delta G_{Ala}$) at those same positions. Note that one of the values used to calculate $\Delta\Delta G_{Val\rightarrow Thr}$ is always the wild-type stability where both values used to calculate $\Delta\Delta G_{Ala\rightarrow Ser}$ are stabilities of mutant proteins. The dotted line is the least squares fit of the points. The *y*-intercept of this line is −0.43 kcal/mol, fairly near to zero. The slope of the line is 1.46. In other words, the value of $\Delta\Delta G_{Thr\rightarrow Val}$, the cost of substituting a methyl with a hydroxyl, is on average about 1.46 times that of $\Delta\Delta G_{Ser\rightarrow Ala}$, the cost of substituting a hydrogen with a hydroxyl.

of, for example, valine to alanine and valine to threonine mutations at different sites is remarkable because it indicates previously advanced explanations of site to site energetic variations must be reconsidered.

Variations in the energetics of a series of leucine to alanine mutations in T4 lysozyme (Eriksson et al., 1992) were attributed to variations in the cavity volume left after the protein adjusted to this change. A similar argument might be made to explain why the valine to alanine, valine to glycine, and valine to serine mutations presented in this work are well correlated in their energetic effects. All of these mutations create cavities that might be compensated for equally well (or poorly) at a given site. But it is more difficult to explain the correlation of the valine to threonine mutant stabilities with the stabilities of other substitutions for valine on the basis of packing effects. The threonine is to a good approximation isosteric with the valine. The small volume change can lead only to a very small cavity. It seems unlikely that the abscence or presence of such a small cavity could lead to the large energetic variations observed here. Some other factor(s) must be at work to cause the variation in stabilities of the threonine mutants, and these same factor(s) have relatively equivalent effects on other mutations examined at these sites.

Energetic Cost of Burying an Unsatisfied Hydrogen Bonding Group. Let us begin by considering the various substitution mutants for valine. In a similar study (Blaber et al., 1993) a series of mutants were generated which introduced a hydroxyl group into the core of T4 lysozyme. The difference in stabilities between wild-type T4 lysozyme and a set of nine Ala→Ser and three Val→Thr mutants was found to vary from −0.2 to −2.8 kcal/mol with a dependence on the solvent exposure of the particular site. It was concluded that these values represent a range of destabilization caused by burying a hydroxyl in the core of a protein. The range of stability change for our nine Val→Thr mutants ($\Delta\Delta G_{Val\rightarrow Thr} = \Delta G_{Thr\ mutant} - \Delta G_{Val\ (wild\text{-}type)}$) is +0.2 to −3.8 kcal/mol, also with an obvious dependence on solvent exposure. The stabilities of both the alanine and serine mutants at these same nine positions have also been assessed. The $\Delta\Delta G_{Ala\rightarrow Ser}$ values ($\Delta G_{Ser\ mutant} - \Delta G_{Ala\ mutant}$) for these "Ala→Ser" mutants have a range of +0.3 to −2.2 kcal/mol.

Solvent-exposed residues must be excluded in order to evaluate the cost of burying a hydroxyl. When this is done, the limits of $\Delta\Delta G_{Val\rightarrow Thr}$ values are from −1.3 to −3.8 kcal/mol and $\Delta\Delta G_{Ala\rightarrow Ser}$ values extend from −0.6 to −2.2 kcal/mol. This is still quite a large range. Comparison of the experimental values to the predicted values is interesting. The average $\Delta\Delta G_{Val\rightarrow Thr}$ at all fully buried valine positions is −2.54 ± 0.89 kcal/mol. The predicted energy difference for a buried valine to threonine substitution (Table 1) is −2.50 kcal/mol, in excellent agreement with the average experimental value. But, as the standard deviation indicates, there is a wide range of experimental values. Two positions, 104 and 111, are very near the predicted value, while positions 23, 74, and 99 are significantly more destabilized by the valine to threonine substitution than predicted. Positions 39 and 66 are affected by about 1 kcal/mol less than predicted. The average $\Delta\Delta G_{Ala\rightarrow Ser}$ at the same positions is −1.29 ± 0.72 kcal/mol. The predicted value for this mutation is −1.98 kcal/mol, a significant difference. Four positions, 23, 74, 99, and 104, are in good agreement with prediction. Positions 39, 66, and 111 are significantly less affected than predicted.

What additional factors could be considered that might reconcile these differences? The first thing that must be considered is the possibility of hydrogen bonding with the new side chain. Careful examination of the crystal structure reveals that, except for the solvent-exposed valines, there are no good hydrogen bonding partners within range of the serine or threonine mutants even if free side chain rotation is allowed. Weak hydrogen bonding is possible to some nearby main chain amides, but these amides are with few exceptions already involved in hydrogen bonding and the geometry of a potential hydrogen bond to the threonine or serine hydroxyl is poor. Therefore, we believe that for the most part the hydrogen bonding capacity of substitutions at these positions remains unsatisfied. One last possibility we cannot directly address is that waters might be buried with some of these side chains.

However, there is a second reason to believe the variability in stability changes is not due to differences in hydrogen bonding. As Figure 2B indicates, the free energies of the various substitutions at a given site correlate reasonably well. The tolerance or lack of tolerance for one kind of substitution predicts the tolerance for other types of substitutions well. This is regardless of whether the mutations one is comparing add hydrogen bonding capability or not. If, for example, an alanine substitution is made at a valine residue and it is severely destabilizing relative to other valine to alanine substitutions, odds are that the threonine, serine, and glycine substitution will also be severely destabilizing relative to

similar substitutions at other valine sites. Similarly, if one kind of substitution at a given site is moderately destabilizing relative to similar substitutions at other sites, other substitutions at the given site will be moderately destabilizing relative to similar substitutions at other sites. The presence or absence of hydrogen bonding capacity in the side chain makes no apparent major perturbations in the rank order of sensitivity to substitution at valines in nuclease.

Other factors that could contribute to the observed variations in stability can be suggested. There might be a need for side chain secondary structural propensities to be taken into account. However, cursory examination of the data rules this out as a principal cause of the variability. For example, valines 66 and 99 are both in helix yet vary dramatically in their ability to tolerate substitution. Similarly, valines 39 and 74 are both in sheet and show wide variations in mutant stabilities. Further, it has been argued that at least α-helix propensity can be accounted for by differences in side chain entropy (Creamer & Rose, 1992), and our predicted values take this factor into account, indeed, with a scale calculated for α-helix conformations (Lee et al., 1994). Another possible explanation for differences in stabilities among positions is differing abilities to accommodate repacking. However, as discussed earlier, this seems unlikely because of the high correlation of the stabilities of different mutations at different sites. Factors such as secondary structural preferences and repacking seem likely to be secondary in their impact on the stabilities. While we can rule out some factors as principal causes of the differences in stabilities of similar mutants at different positions, we cannot positively identify what the important factors are. Fortunately, the effect of these factors seems to be similar with all the types of substitutions considered here.

The correlation of the stability of the threonine and serine substitution mutants with the alanine and glycine substitutions argues that the stability effects of a given mutation at a given site are driven primarily by factors common to all mutants at that site. Since the energetic costs of making serine, threonine, and alanine mutations at buried valines follow the rank order of the hydrophobic transfer free energy for each kind of side chain replacement, it would appear that this parameter takes into account the effects of putting a hydrogen bonding side chain into a non-hydrogen-bonding environment.

A second line of evidence concerning the cost of burying hydrogen bonding groups without satisfying their capacity is offered by the substitutions at threonine. Two valine substitution mutants, T33V and T41V, showed significant stability increases compared to wild-type, and two other mutants, T44V and T82V, showed slight stability gains. This is somewhat surprising since all of these residues have some solvent exposure of the hydroxyl. Apparently, even partial burial of a hydroxyl can be measurably destabilizing. There are two related rationalizations of this observation. First, if part of the hydroxyl becomes solvent inaccessible as the protein folds, solvent must be stripped away from at least the part of the group that is buried. Mutation to a valine replaces this unfavorable step with a favorable hydrophobic interaction. The interaction of the solvent-exposed portion of the valine with water is not favorable, but this is true in both the native and denatured states and thus essentially cancels out. Second, as less of the surface of the hydroxyl becomes available for solvent interaction, the number of ways the solvent can interact with the hydroxyl drops. This decrease in entropy must be unfavorable. It is interesting to note that T41, which has the largest increase in stability upon valine substitution, has two crystallographic waters in hydrogen bonding distance. The release of these waters upon valine substitution may contribute a favorable entropic term to the increase of stability seen in this mutant.

It is also noteworthy that two serine to alanine mutations, S59A and S128A, have previously been identified as stabilizing (Green et al., 1992). Intriguingly, both of these mutations have hydrogen bonds to crystallographic waters. Though not every threonine to valine or serine to alanine mutant is stabilizing, the large number that are suggests that a good general strategy to increase the stability of proteins would be to replace partially solvent-exposed hydrogen bonding groups with isosteric hydrophobic groups.

Energetic Contribution of Hydrogen Bonding to Protein Folding. Measuring the free energy of a hydrogen bonding interaction is more difficult than assessing the cost of burying an unsatisfied hydrogen bonding group. We must attempt to evaluate factors other than hydrogen bonding, introducing a great deal of uncertainty. One good example of this uncertainty is shown by a mutation of a threonine to valine in T4 lysozyme (Alber et al., 1987; Grutter et al., 1987). Unexpectedly, free energy calculations by another group (Dang et al., 1989) indicated that the differences in van der Waals interactions between the methyl and hydroxyl groups played a larger role in the stability differences than did changes in hydrogen bonding capacity.

Nevertheless, we have attempted to assess the energetics of hydrogen bonding in a few selected examples. In making these rough estimates of hydrogen bonding interaction energy, we have chosen to include what we believe will be the three other most important contributors to differences in overall protein stability (Table 1): hydrophobicity (Fauchère & Pliška, 1983), main chain conformational entropy (Stites & Pranata, 1995), and side chain conformational entropy (Lee et al., 1994). Even after limiting the factors being considered, there are numerous possible scales to choose from, especially for hydrophobicity, without considering possible corrections to these scales. We have somewhat arbitrarily chosen the scales referenced above. We are completely ignoring undoubtedly important effects such as differences between mutants in packing and van der Waals interactions. Again, the possibility exists in some cases that water may be buried as a substitute for the missing hydroxyls. With these caveats in mind, we look at hydrogen bonding interactions of some threonines and tyrosines in more detail.

T62. Threonine 62 is completely solvent inaccessible. It is positioned so that it may donate a hydrogen bond to the carbonyls of G20 and A58 (Figure 1A). The carbonyl of A58 is also hydrogen bonded to the amide of T62. The carbonyl of G20, however, has no other hydrogen bonding partners nearby, and its hydrogen bonding capacity will be unsatisfied, barring major rearrangement of the protein main chain.

The substitution at position 62 with valine results in the overall loss of −0.2 kcal/mol of stability for the native structure. The valine side chain is predicted to be 2.5 kcal/mol more stable than the threonine due to its greater hydrophobic transfer free energy and lower main chain and side chain entropy in the denatured state. The substitution of threonine for valine at three buried helical positions, 66,

99, and 104, costs 1.4, 2.5, and 2.3 kcal/mol, respectively. This implies that the loss of the hydrogen bonding interactions of the T62 hydroxyl costs from −1.6 to −2.7 kcal/mol.

The difference in free energy between the serine and alanine mutants at this position is about −0.4 kcal/mol. The alanine side chain should be 1.98 kcal/mol more stable than the serine due to its greater hydrophobic transfer free energy and lower main chain and side chain entropy in the denatured state. The substitution of serine for alanine at the buried helical positions, 66, 99, and 104, costs 0.8, 1.9, and 1.7 kcal/mol, respectively. This implies that the loss of the hydrogen bonding interactions of the hydroxyl costs −1.4 to −2.4 kcal/mol, in reasonable agreement with the value arrived at by comparing the valine and threonine mutants.

T22. This residue is located in a sheet and is partially solvent exposed. It interacts with the carboxylate of aspartate 21, which is not solvent accessible (Figure 1A). This is one of two residues that show a consistent preference for the presence of a hydroxyl group (Figure 3). The threonine is more stable than the valine by 0.9 kcal/mol while the serine is more stable than the alanine by 1.1 kcal/mol. The solvent exposure of the hydroxyl is fairly poor, so the mutation to valine should afford some hydrophobic stabilization, though how much is unclear. If we estimate the threonine to valine mutation contributes 1−2 kcal/mol of energy, then the threonine-aspartate hydrogen bonding interaction is worth 2−3 kcal/mol. This seems a surprisingly low cost for the burial of the aspartate but probably reflects reorganization of the structure to allow solvent access or interaction with the nearby threonine 41 or arginine 35.

T120. The hydroxyl group of T120 acts as a bridge between the amine of K78 and the carboxylate of D77 (Figure 1B). The methyl group of the threonine is partially solvent exposed while the hydroxyl is not. T120 is the position most sensitive to loss of hydrogen bonding. The substitution with valine is 1.8 kcal/mol less stable than wild-type. Since the valine is predicted to be about 2.5 kcal/mol more stable, this indicates that the loss of the hydroxyl's hydrogen bonding capacity costs around 4.3 kcal/mol, a fairly large number. The serine mutant is predicted to be approximately 2 kcal/mol less stable than the alanine. Instead, serine is 0.7 kcal/mol more stable than the alanine mutant, indicating the hydrogen bonding interactions are worth about 2.7 kcal/mol. The difference between the two estimates is large. This may indicate serine is less able to maintain the hydrogen bonding network than threonine, may indicate the error inherent in our estimates, or both.

It is also interesting to examine the cysteine substitutions at T22, T62, and T120. Of course, hydrophobicity, size, and considerations of geometry play a role in stability changes, but so does hydrogen bonding capacity. Cysteines are both hydrogen bond donors and acceptors (Gregoret et al., 1991) but are better donors than acceptors. At T62 and T22, where the wild-type interaction is with good acceptors, cysteine is a relatively good substitution. Cysteine is a poor replacement compared to serine at T120, perhaps indicating that interaction with K78 is more important than hydrogen bond donation to D77.

Y91, Y93, and Y27. These three tyrosines form an interesting continuum. They all interact with carboxylates but vary in their degree of solvent exposure. Y91 is nearly completely buried and interacts with both the carboxylate of D77 and the amide hydrogen of H121. D77 is completely buried but has other hydrogen bonding interactions. Substitution with phenylalanine results in the greatest destabilization of this class of mutants, 2.4 kcal/mol. Our rough prediction of stability changes (Table 1) indicates that a buried phenylalanine should be more stable than a buried tyrosine by about 1.5 kcal/mol. Thus the strength of the hydrogen bonding interaction between the carboxylate and the phenol group is about 3.9 kcal/mol, quite significant.

The interaction of Y27 and E10 and the interaction of Y93 and E75 are very similar. Both tyrosines are largely buried except for the hydroxyl. Both interact with one of the carboxylate oxygens of a glutamate. The oxygens of the carboxylates that are not involved in the interaction with the phenols are solvent exposed. E75 also has an interaction with the side chain of H121 while E10 has no other interactions. It seems sensible to argue there is little need to take into account the hydrophobic transfer free energy difference between phenylalanine and tyrosine as the parts of the two side chains which are buried should offer nearly identical interactions to the remainder of the protein. The difference in main chain entropy for phenylalanine and tyrosine is approximately zero, while side chain entropy should favor the phenylalanine mutant by 0.4 kcal/mol. Since the two carboxylates are also solvent exposed, the solvent should be able to hydrogen bond freely to the carboxylate. One factor we are unable to evaluate is whether the carboxylates gain additional degrees of conformational freedom, and we are forced to assume that this effect is negligible. Nevertheless, the stability change upon mutation may reflect largely the energetic difference between an intramolecular hydrogen bond (E10−Y27 or E75−Y93) and an intermolecular hydrogen bond (E10−solvent or E75−solvent). The Y27F mutant is −0.6 kcal/mol less stable than wild-type, indicating in this case that an intramolecular interaction between the phenolic hydroxyl and the carboxylate is more favorable than the intermolecular interaction between the carboxylate and water by about 1 kcal/mol. Y93F, however, is significantly more destabilized, −2.0 kcal/mol. This would imply that the interaction between the phenolic hydroxyl and the carboxylate is more favorable than the intermolecular interaction between the carboxylate and water by about 2.4 kcal/mol. This is a fairly big difference. Though similar, the two interactions are not identical. The most obvious structural factor that may explain the energetic differences is the distances between the donors and acceptors. The donor−acceptor distance between Y93 and E75 is 2.60 Å while Y27 and E10 are farther apart, at 2.83 Å.

The interaction between a carboxylate and a phenolic hydroxyl is presumably very strong as hydrogen bonds go (Rose & Wolfenden, 1993), being high in ionic character. Hydrogen bonds are of course simply a special class of electrostatic interaction. The fully charged glutamate would interact very strongly with the relatively easily ionized phenol hydroxyl. A glutamate−serine interaction, for example, would on average be weaker and a serine−serine interaction weaker still. A glutamate—water interaction would be closer in character to a glutamate−serine interaction. Thus the estimate of 1−2.4 kcal/mol for the difference in free energy between an intramolecular glutamate−tyrosine and an intermolecular glutamate−water interaction represents an upper limit for the difference between an intramolecular and intermolecular hydrogen bonding interaction.

Our estimates of the energetic contribution of hydrogen bonding of these various side chains ranges from about 1.5 to 4.0 kcal/mol. The range reflects two factors. First, there is undoubtedly a range of interaction strengths, dependent on factors such as bond geometry, the identity of the functional groups involved, and their environment. The second factor, no less important, is the uncertainty introduced by our inability to evaluate other energetic contributors such as hydrophobicity, packing, and entropic differences. Regardless, this range of interaction strengths is in agreement with those reported elsewhere. Pace and co-workers (Shirley et al., 1992) interpret their data on mutations in ribonuclease T1 as implying that an intramolecular hydrogen bond is an average of 1.3 kcal/mol more stable than an intermolecular one. Hammen et al. (1995) evaluate the strength of a serine−aspartate and two aspartate−main chain hydrogen bonds in *E. coli* histidine-containing protein as 4−5 kcal/mol, or about 1.3−1.6 kcal/mol for each of the three bonds. Other studies (Chen et al., 1993; Serrano et al., 1992; Thunnissen et al., 1992) have focused on the loss of stability upon removal of a hydrogen bonding interaction without attempting to correct for other energetic effects of the mutation. The observed changes are comparable to those observed here.

Slope Effects. Although the major focus of this study is on hydrogen bonding, it is difficult to completely ignore the pronounced slope effects in nuclease. Previous work (Shortle et al., 1990; Green et al., 1992) in the nuclease system has shown a correlation between the value of ΔG_{H_2O} for a given mutant and the absolute value of the difference in the value of m_{GuHCl} for the mutant and wild-type protein. In other words, an increase or decrease in m_{GuHCl} relative to the wild-type value was found to correlate with a loss of stability. The data presented here show the same correlation.

Another interesting point is that the three mutants discovered here with significant stability gains from wild-type, T33V, T41V, and T41C, all have slopes slightly below that of wild-type. This is also true for the previously discovered stabilizing mutants S59A, S128A, P117A, P117G (Green et al., 1992), and H124L (Shortle, 1986). It has been proposed (Green et al., 1992) that the sequence of wild-type nuclease evolved to maximize the free energy of the denatured state by optimizing the amount of hydrophobic surface area exposed upon denaturation. If this argument is correct, the general trend among stabilizing mutants to lower values of m_{GuHCl} suggests that a truly optimized denatured state would expose less hydrophobic surface than wild-type actually does.

The slope effects observed in mutants of V111 and V104 also are very interesting. Both of these residues are part of the "minor hydrophobic core" (Shortle et al., 1990) formed where the C-terminal 122−134 α-helix packs up against the remainder of the protein. The serine, alanine, and glycine mutants at both these sites are $m-$ mutants, down in slope relative to wild-type, with values ranging from 0.75 relative to wild-type (V111G) to 0.93 (V104A). On the other hand, the V104T mutant is only slightly reduced relative to wild-type (0.98) while V111T is actually an $m+$ mutant, with a slope greater than that of wild-type (1.05). This mutation causes the greatest increase in slope yet observed in this region of the molecule.

Energetic Cost of Removing a Buried Methyl. The energetic cost of removing a buried methyl group from the hydrophobic core of nuclease can be estimated through stability analysis of the valine substitutions. Aside from being interesting in itself, this offers another way to check the internal consistency of the data. The difference in free energy between Val→Thr and Val→Ser mutations can offer an estimation of the stability cost for removal of a methyl group from a buried location. The average $\Delta\Delta G_{Thr\rightarrow Ser}$ at all seven fully buried valine positions is -1.80 ± 0.46 kcal/mol. Similarly, alanine substitutions for valine give an energetic estimation for the removal of two buried methyl groups. The average $\Delta\Delta G_{Val\rightarrow Ala}$ at the same positions is -3.06 ± 0.61 kcal/mol, or -1.53 ± 0.30 kcal/mol per methyl equivalent, which is in reasonable agreement with the value of $\Delta\Delta G_{Thr\rightarrow Ser}$.

Although the cost per methyl group is adequately consistent in this analysis, a final cautionary note should be sounded. On the basis of hydrophobicity and main chain and side chain entropy, the value for $\Delta\Delta G_{Val\rightarrow Ala}$ is expected to be −1.04 kcal/mol and that of $\Delta\Delta G_{Thr\rightarrow Ser}$ would be predicted to be −0.52 kcal/mol. Obviously, there is a serious gap between the theoretical and experimental values. This reemphasizes the point that any attempt to isolate and quantify a single factor that contributes to protein stability by estimating the effects of other factors must be treated with great circumspection.

Conclusions. We have some answers to the two questions raised at the beginning of the discussion. What is the cost of burying an unsatisfied hydrogen bond donor/acceptor? Substitution of buried valines with serine and threonine indicates that burying a hydroxyl without satisfying its hydrogen bonding capacity is not energetically favorable. The gain in stability upon substituting non-hydrogen-bonded, partially buried threonines with valines also emphasizes this. The exact energetic price that must be paid for this burial is quite variable, and the energetic variability of these substitutions cannot be easily explained. However, the fact that the rank order of mutant stability at each position generally follows the same order as hydrophobicity indicates that the relative side chain hydrophobicity gives a reasonable estimation of the cost of burying a hydroxyl without providing a hydrogen bonding partner. Further, the correlation of the relative effects of different mutations over a range of sites indicates that whatever the factors are that control the sensitivity to substitution they affect even rather different substitutions in much the same way.

What are typical free energies of hydrogen bonding interactions in a protein? As several of the threonine mutations presented here attest, it would appear that partial burial of a hydrogen bonding group is unfavorable. If even partially limiting hydrogen bonding to solvent is unfavorable, this may argue that the even more limited interaction between two buried hydrogen bonding groups in the native state will not be favorable compared to the interaction of these two groups with the solvent in the denatured state. Nevertheless, it appears that this can be the case, particularly if one of the hydrogen bonding partners is charged. We estimate the energetic contribution of hydrogen bonding interactions examined here to be about 1.5−4.0 kcal/mol. Thus, the formation of an intramolecular hydrogen bond in a protein appears to be a net stabilizing factor. However, given the manifold uncertainties in evaluation of any one of the many energetic terms that contribute to protein stability, we regard this last conclusion as tentative.

ACKNOWLEDGMENT

We thank Chad Dyson for help with the mutagenesis and sequencing and Dan Spencer for preparation of the V111A mutant. Janet Thornton supplied the results of the HBPLUS program, which was most appreciated. We are grateful to Nick Pace for providing preprints and to Nick Pace and David Shortle for critical reading of the manuscript and helpful discussion. Finally, we thank Tom Terwilliger for the suggestion that we carry out this series of experiments.

REFERENCES

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

Baker, E. N., & Hubbard, R. E. (1984) *Prog. Biophys. Mol. Biol. 44*, 97−197.

Berndt, K. D., Beunink, J., Schröder, W., & Wüthrich, K. (1993) *Biochemistry 32*, 4564−4570.

Blaber, M., Lindstrom, J. D., Gassner, N., Xu, J., Heinz, D. W., & Matthews, B. W. (1993) *Biochemistry 32*, 11363−11373.

Blanchard, L., Dolla, A., Bersch, B., Forest, E., Bianco, P., Wall, J., Marion, D., & Guerlesquin, F. (1994) *Eur. J. Biochem. 226*, 423−32.

Bondi, A. (1964) *J. Phys. Chem. 68*, 441−451.

Burns, J. A., Butler, J. C., Moran, J., & Whitesides, G. M. (1991) *J. Org. Chem. 56*, 2648−2650.

Chen, Y. W., Fersht, A. R., & Henrick, K. (1993) *J. Mol. Biol. 234*, 1158−1170.

Creamer, T. P., & Rose, G. D. (1992) *Proc. Natl. Acad. Sci., U.S.A. 89*, 5937−5941.

Dang, L. X., Merz, K. M., Jr., & Kollman, P. A. (1989) *J. Am. Chem. Soc. 111*, 8505−8508.

Derst, C., Henseling, J., & Röhm, K. H. (1992) *Protein Eng. 5*, 785−789.

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

Doig, A. J., & Williams, D. H. (1992) *J. Am. Chem. Soc. 114*, 338−343.

Eijsink, V. G. H., van der Zee, J. R., van den Burg, B., Vriend, G., & Venema, G. (1991) *FEBS Lett. 282*, 13−16.

Eijsink, V. G. H., Vriend, G., van der Zee, J. R., van den Burg, B., & Venema, G. (1992) *Biochem. J. 285*, 625−628.

Eriksson, A. E., Baase, W. A., Zhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P., & Matthews, B. W. (1992) *Science 255*, 178−183.

Fauchère, J.-L., & Pliška, V. (1983) *J. Med. Chem. 18*, 369−375.

Gregoret, L. M., Rader, S. D., Fletterick, R. J., & Cohen, F. E. (1991) *Proteins: Struct., Funct., Genet. 9*, 99−107.

Green, S. M., Meeker, A. K., & Shortle, D. (1992) *Biochemistry 31*, 400−410.

Grutter, M. G., Gray, T. M., Weaver, L. H., Alber, T., Wilson, K., & Matthews, B. W. (1987) *J. Mol. Biol. 197*, 315−329.

Hammen, P. K., Scholtz, J. M., Anderson, J. W., Waygood, E. B., & Klevitt, R. E. (1995) *Protein Sci. 4*, 936−944.

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

Jackson, S. E., & Fersht, A. R. (1994) *Biochemistry 33*, 13880−13887.

Jeffrey, G. A., & Saenger, W. (1991) *Hydrogen Bonding in Biological Structures*, Springer-Verlag, Berlin.

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

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

Kunkel, T. A., Roberts, J. D., & Zakour, R. A. (1987) *Methods Enzymol. 154*, 367−382.

Lee, K. H., Xie, D., Freire, E., & Amzel, L. M. (1994) *Proteins: Struct., Funct., Genet. 20*, 68−84.

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

Mårtensson, L. G., Jonsson, B.-H., Andersson, M., Kihlgren, A., Bergenhem, N., & Carlsson, U. (1992) *Biochim. Biophys. Acta 1118*, 179−186.

McDonald, I. K. (1992) HBPLUS, London, UCL.

McDonald, I. K., & Thornton, J. M. (1994) *J. Mol. Biol. 238*, 777−793.

Mott, J. E., Grant, R. A., Ho, Y.-S., & Platt, T. (1985) *Proc. Natl. Acad. Sci., U.S.A. 82*, 88−92.

Nicholls, D. J., Wood, I. S., Nobbs, T. J., Clarke, A. R., Holbrook, J. J., Atkinson, T., & Scawen, M. D. (1993) *Eur. J. Biochem. 212*, 447−455.

Pace, C. N. (1995) *Methods Enzymol.* (in press).

Pauling, L. (1960) *The Nature of the Chemical Bond*, third ed., Cornell University Press.

Pederson, J. T., Olsen, O. H., Betzel, C., Eschenburg, S., Branner, S., & Hastrup, S. (1994) *J. Mol. Biol. 242*, 193−202.

Pjura, P., & Matthews, B. W. (1993) *Protein Sci. 2*, 2226−2232.

Radzika, A., & Wolfenden, R. (1988) *Biochemistry 27*, 1664−1670.

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

Serrano, L., Kellis, J. T., Cann, P., Matouschek, A., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 783−804.

Shirley, B. A., Stanssens, P., Hahn, U., & Pace, C. N. (1992) *Biochemistry 31*, 725−732.

Shortle, D. (1986) *J. Cell. Biochem. 30*, 281−289.

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

Stickle, D. F., Presta, L. G., Dill, K. A., & Rose, G. D. (1992) *J. Mol. Biol. 226*, 1143−1159.

Stites, W. E., & Pranata, J. (1995) *Proteins: Struct., Funct., Genet. 22*, 132−140.

Stites, W. E., Byrne, M. P., Aviv, J., Kaplan, M., & Curtis, P. M. (1995) *Anal. Biochem. 227*, 112−122.

Thunnissen, M. M. G. M., Franken, P. A., de Haas, G. H., Drenth, J., Kalk, K. H., Verheij, H. M., & Dijkstra, B. W. (1992) *Protein Eng. 5*, 597−603.

Yamada, H., Kanaya, E., Ueno, Y., Ikehara, M, Nakamura, H., & Kikuchi, M. (1994) *Biol. Pharm. Bull. 17*, 612−616.

BI951211X