Role of Electrostatic Screening in Determining Protein Main Chain Conformational Preferences†

Franc Avbelj*,‡,§ and John Moult‡

Center for Advanced Research in Biotechnology, 9600 Gudelsky Drive, Rockville, Maryland 20850, and National Institute of Chemistry, Hajdrihova 19, Ljubljana SI 61115, Slovenia

Received March 30, 1994; Revised Manuscript Received October 17, 1994®

ABSTRACT: Amino acids display significant variation in propensity for the α_R-helical, β-sheet, and other main chain conformational states in proteins and peptides. The physical reason for these preferences remains controversial. Conformational entropy, steric factors, and the hydrophobic effect have all been advanced as the dominant underlying cause. In this work, we explore the role of a fourth factor, electrostatics, in determining the main chain conformation in protein molecules. Potentials of mean force derived from experimental protein structures are used to evaluate the free energy of electrostatic and other interactions of a residue in a protein environment. The local and nonlocal electrostatic interactions of main chain polar atoms are found to be crucial for determining the preferences of residues for the α_R-helical state and other main chain conformational states of a residue. Further, the strength of local and nonlocal electrostatic interactions is shown to depend on the electrostatic screening by solvent and protein groups. Residue specific modulation of this screening in a manner related to side chain bulk and squatness produces a model that fits the observed distribution of residue conformations in proteins and recent experimental mutagenesis data on protein stability better than any other single factor.

Amino acid residues exhibit distinct preferences for α_R-helical, β-sheet, and other main chain conformational states. These preferences have been characterized by various propensity scales on the basis of data obtained from statistical surveys of experimental X-ray structures of proteins (Chou & Fasman, 1974), host−guest systems (Wojcik *et al.*, 1990), model peptides (O'Neal & DeGrado, 1990; Padmanabhan *et al.*, 1990; Lyu *et al.*, 1990; Merutka *et al.*, 1990; Kemp *et al.*, 1991; Kim & Berg, 1993), and site-directed mutagenesis (Blaber *et al.*, 1993; Horovitz *et al.*, 1992; Minor & Kim, 1994). Secondary structure prediction methods based on the statistical surveys of experimental protein structures are able to predict the secondary structure of ≈65% of protein residues from sequence (Garnier *et al.*, 1978). The partial success of these prediction methods establishes that the conformational preferences of individual protein residues are important in determining protein structure from sequence, yet the underlying physical basis for amino acid preferences remains poorly understood.

Three factors have been suggested as the dominant physical reason for the relative preferences of residues for an α_R-helical environment: conformational entropy (Creamer & Rose, 1992; Padmanabhan & Baldwin, 1991), steric factors (Hermans *et al.*, 1992; Yun & Hermans, 1991), and the hydrophobic effect (Horovitz *et al.*, 1992; Blaber *et al.*, 1993). Other factors have also been shown to affect the stability of amino acid residues in an α_R-helix, particularly the electrostatic interactions of charged residues with both ends of a helix (Hol, 1987; Šali *et al.*, 1988; Serrano & Fersht, 1989). The physical reasons for preferences of residues for β-sheet and other main chain conformational states have been less studied.

The side chain conformation of residues in an α_R-helix is constrained considerably compared to the side chain in other main chain conformational states because of the bulky helix backbone (Piela *et al.*, 1987; McGregor *et al.*, 1987; Padmanabhan *et al.*, 1990; Padmanabhan & Baldwin, 1991; Hermans *et al.*, 1992; Yun & Hermans, 1991). These constraints may result in an unfavorable energy contribution from steric strain, or they may have an associated conformational entropy cost. No calculations have demonstrated a clear role for strain. However, Creamer and Rose (1992) have shown by a Monte Carlo procedure that the side chain conformational entropy of eight types of nonpolar naturally occurring amino acid residues does correlate with the experimental preferences of residues.

The stability of small nonpolar amino acid residues in the α_R-helical conformational state has also been studied using free energy simulations (Hermans *et al.*, 1992; Yun & Hermans, 1991; Yun *et al.*, 1991). The differences in helix propensities for four naturally occurring and four unnatural amino acid residues calculated from molecular dynamics simulations agree well with the experimental relative free energies obtained with model peptides. Qualitative analysis of the results suggests that the conformational entropy and steric strain determine the α_R-helix propensities of amino acid residues (Piela *et al.*, 1987; Hermans *et al.*, 1992; McGregor *et al.*, 1987).

Blaber *et al.* (1993) determined the relative stabilities and X-ray structures of T4 lysozyme mutants for amino acid substitutions at helix positions 44 and 131. They showed that the side chain hydrophobic surface area buried against the protein correlates with the relative free energies of

† This research was supported by Grant GM41034 from the National Institutes of Health (U.S.) and by the Ministry of Science of Slovenia.

* Author to whom correspondence should be addressed at the National Institute of Chemistry.

‡ Center for Advanced Research in Biotechnology.

§ National Institute of Chemistry.

® Abstract published in *Advance ACS Abstracts,* December 1, 1994.

0006-2960/95/0434-755$09.00/0

unfolding of T4 lysozyme mutants. They therefore propose that the hydrophobicity primarily determines the preferences of residues for the α_R-helical environment.

Here we explore the role of a fourth factor, main chain electrostatics, in determining residue conformational preference. The importance of local main chain electrostatics in determining end to end distance in peptides was pointed out many years ago by Brant and Flory (1965a,b), but it has not previously been considered in the present context. We begin by showing that the α_R-helical, β-sheet, and other main chain conformational states of a residue have significantly different electrostatic energies of interaction between adjacent peptide groups. These energy differences are so large that electrostatics must play a role in determining conformational preferences. To analyze the role of main chain electrostatic interactions, a new conformational variable, E_{local}, is defined as the electrostatic energy of the main chain CO and NH groups of a residue arising from interactions with the main chain CO and NH groups within that residue and with the flanking peptide groups (see Figure 1A). E_{local} then provides a convenient alternative to the main chain dihedral angles for defining the conformational state and is used as a reaction coordinate in a potential of mean force analysis based on the observed distributions of conformations in experimental protein structures (Beveridge & DiCapua, 1989; Avbelj, 1992). It is shown that the potential of mean force as a function of residue conformation is consistent with a simple electrostatic model, in which interactions between atomic point charges are screened by surrounding solvent and protein groups. Further, the residue specific screening factors derived from fitting the potential of mean force are shown to be directly related to the average interaction energy of residues with their environment. Thus, the model that emerges accounts for the conformational preferences of residues in proteins in terms of the varying screening of main chain electrostatics by different side chain types. The more bulky the side chain and the nearer that bulk to the backbone, the lower the degree of screening and the stronger the main chain electrostatics. This model fits the available data rather better than the alternatives.

RESULTS AND DISCUSSION

Influence of Electrostatics on Amino Acid Conformation. High-quality ab inito quantum mechanical calculations on alanine and glycine dipeptide analogs under vacuum show a large energy difference of about 2.5 kcal/mol between the α_R- and β-conformations, favoring the β-conformation (Head-Gordon *et al.*, 1991). Since the atoms of the protein backbone are highly polar, we would expect that the electrostatic interactions between main chain polar atoms have a significant role in determining this energy difference. A simple electrostatic energy calculation on the interaction energy between the polar main chain atoms of a dipeptide (the boxed unit in Figure 1A) under vacuum confirms this. Coulomb's law with point atom partial charges of -0.28, $+0.28$, $+0.38$, and -0.38 electron for the N, H_N, C, and O main atoms and with a dielectric constant of 1 was used. The difference in main chain electrostatic energy between β- and α_R-conformations is $\approx$3.5 kcal/mol, which is similar to that obtained by ab initio calculation. Figure 1A illustrates that the origin of this energy difference is in the interaction of the CO and NH dipoles flanking the C_α atom. The β-conformation has lower electrostatic energy than the α_R-conformation because in the β-conformation the CO and NH dipole moments are aligned antiparallel, while in the α_R-conformation, they are parallel. The effect extends over more than one residue in a polypeptide chain. As Figure 1A shows, next nearest neighbor CO and NH dipoles in a β-strand are aligned antiparallel, whereas in an α_R-helix these dipoles are parallel.

Calculation of Average Electrostatic Properties in Proteins. We shall be concerned with relating electrostatics to the average conformational properties of protein molecules. Average quantities are calculated from a set of 114 high-resolution (resolution < 2.0 Å and R factor < 20%) X-ray structures of proteins from the Protein Data Bank (Bernstein *et al.*, 1977), containing a total 25 500 residues. All residues except proline are included in the averages. Proline is excluded because the absence of a polar NH group gives anomalous electrostatic properties. The structures are accessed under the following codes: 1BP2, 1CCR, 1CHO, 1CRN, 1CSE, 1CTF, 1GOX, 1HMQ, 1HNE, 1HOE, 1LZ1, 1MBA, 1MBC, 1NTP, 1PAZ, 1PCY, 1PSG, 1R69, 1RDG, 1RNT, 1SGT, 1TGS, 1TGT, 1TLD, 1TON, 1UBQ, 2ACT, 2ALP, 2APR, 2AZA, 2CA2, 2CCY, 2CDV, 2CGA, 2CI2, 2CPP, 2CTS, 2CYP, 2FB4, 2FBJ, 2GBP, 2I1B, 2LHB, 2LTN, 2LZT, 2OVO, 2PRK, 2PTC, 2RHE, 2RSP, 2SGA, 2TGP, 2TMN, 2UTG, 2WRP, 351C, 3BCL, 3C2C, 3DFR, 3EST, 3GRS, 3LZM, 3RP2, 3SGB, 3TPI, 4DFR, 4FD1, 4FXN, 4INS, 4PEP, 4PTI, 5CHA, 5CYT, 5EBX, 5RXN, 5TNC, 7PCY, 7RSA, 8DFR, 9PAP, 1F3G, 2HPR, 1AAP, 1BBP, 1CSC, 1FKF, 1GCT, 1GMA, 1OMD, 1RBP, 1RNH, 1S01, 1SNC, 1YCC, 2ER7, 2MCG, 2MLT, 2ST1, 2TIM, 2TRX, 2TSC, 3APP, 3B5C, 3BLM, 3CLA, 4BP2, 5CPV, 5RUB, 6CPA, 6CPP, 8ABP, 8RSA, 9INS, 9WGA.

Local Electrostatic Energies for the Conformational States of a Protein Residue. In order to analyze the influence of main chain electrostatics on conformational preference in proteins, we divided the total electrostatic interaction energy of the main chain CO and NH groups of a residue (E_{total}) into local (E_{local}) and other nonlocal ($E_{nonlocal}$) components. E_{local} is defined as the electrostatic energy of the main chain CO and NH groups of a residue arising from interactions with main chain CO and NH groups within the residue and with the flanking peptide groups, as defined in Figure 1A. The largest contribution to the remaining ($E_{nonlocal}$) energy is from main chain hydrogen bonding. Note that E_{local} and $E_{nonlocal}$ depend only on the conformation of the main chain atoms and are independent of the conformation of the side chain atoms. E_{local} and $E_{nonlocal}$ are calculated using Coulomb's law with a dielectric constant of 1. Point atomic charges for the main chain atoms N, H_N, C, and O are -0.28, $+0.28$, $+0.38$, and -0.38 electron, respectively. Interactions between atoms within the NH and CO dipoles are ignored. Interaction between dipoles are included in the electrostatic energy if the distance between the N and C atoms is smaller than 6.5 Å. A short cutoff is used since we are interested in short range electrostatic organization. Although molecular dynamics calculations have shown that the exact structure depends on long range interactions, while approximate structure does not (Kitson *et al.*, 1993).

E_{local} provides a simple, one-dimensional alternative to the more familiar ϕ/ψ description of the conformation of a residue. The relationship between these two representations is shown in Figure 1B,C. The value of E_{local} varies continuously between -4.0 and 2.5 kcal/mol, with residues

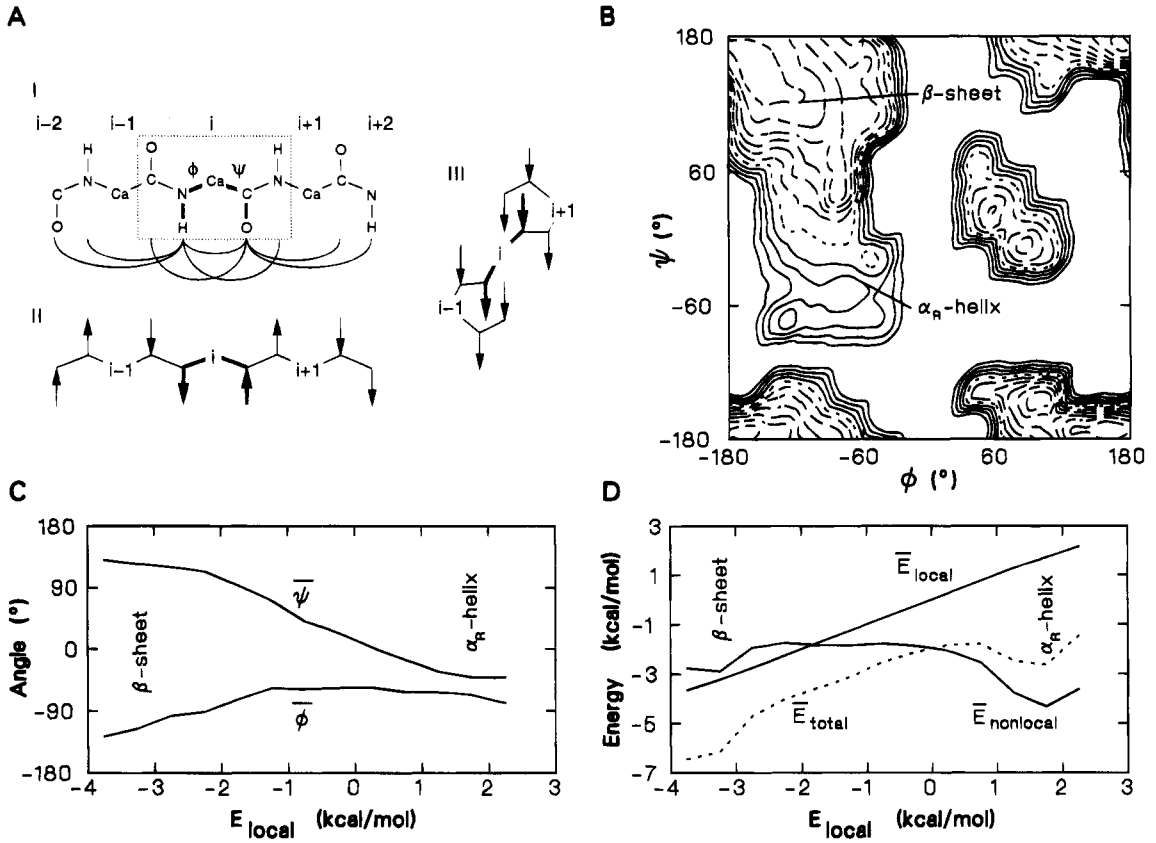

FIGURE 1: Division of the main chain electrostatic energy into local (E_{local}) and nonlocal ($E_{nonlocal}$) components and the location of the main conformational states of a residue as a function of E_{local}. (A) (I) Main chain interactions of the NH and C=O groups of residue i included in the calculation of the local main chain electrostatic energy, E_{local}. Other main chain interactions of the NH and C=O groups contribute to $E_{nonlocal}$. The box encloses the unit used in the calculation of electrostatic interaction energy between the polar main chain atoms of a dipeptide. (II) Alignment of the peptide group dipoles in a β-strand. Arrows represent the direction of the NH and C=O group dipoles. (III) Alignment of the peptide group dipoles in an α_R-helix. In a β-strand, the dipole moments of the NH and C=O groups within a residue are antiparallel and thus interact favorably. Since the fields from the two dipoles approximately cancel, nearby solvent and protein groups interact relatively weakly with residues in this conformation. In the α_R-helix, the dipole moments within a residue are parallel and interact unfavorably. The fields from the two dipoles reinforce each other, resulting in strong interactions with the surroundings. (B) Average local main chain electrostatic energy $\langle E_{local} \rangle$ of residues as a function of ϕ and ψ. Positive energy contours are in solid lines, negative ones are dashed, and zero ones are dotted. The contour interval is 0.5 kcal/mol. The locations of residues in the β-sheet and α_R-helix conformations are shown. Note the high energy of the α_R-region (local maximum) compared with the β (near the global minimum), caused by the difference in peptide dipole alignment. E_{local} was calculated using a point charge model. For sterically excluded regions, $\langle E_{local} \rangle$ was set to a large positive number. Here and elsewhere, data from all residues except Pro in a set of 114 high-resolution experimental protein structures (Bernstein *et al.*, 1977) are included. (C) Average $\bar{\phi}$ and $\bar{\psi}$ as a function of E_{local}. The average torsion angles vary smoothly between the β-sheet and α_R-helical conformations. E_{local} thus provides an alternative parameter for describing the conformational state of a residue. (D) Components of the average main chain electrostatic energy $\bar{E}_{total}$ ($\bar{E}_{nonlocal}$ and $\bar{E}_{local}$) as a function of E_{local}. The minima in $\bar{E}_{nonlocal}$ and $\bar{E}_{total}$ for residues in α_R-helix and β-sheet conformations are cuased by the greater number of hydrogen bonds in these states than elsewhere. Although $\bar{E}_{nonlocal}$ is more favorably (by $\approx$1.5 kcal/mol) for helix than for sheet, the total main chain energy $\bar{E}_{total}$ favors the β-sheet conformation by $\approx$3.5 kcal/mol. Thus, if there were no compensating interactions, no helix would occur in proteins.

in β-sheet and α_R-helix found at opposite ends of this interval at ≈ -3.25 and $\approx$1.75 kcal/mol, respectively. As discussed earlier, this large difference in E_{local} ($\approx$5 kcal/mol) arises from the antiparallel (β) and parallel (α_R) alignments of the CO and NH dipole moments. Residues not involved in β-sheet and α_R-helix are found to fall between these energy extremes.

Figure 1D shows the average $\bar{E}_{local}$, $\bar{E}_{nonlocal}$, and total ($\bar{E}_{total} = \bar{E}_{local} + \bar{E}_{nonlocal}$) main chain electrostatic energies of residues as a function of E_{local} in experimental protein structures. The E_{local} range of -4.0 to 2.5 kcal/mol was divided into 13 equal bins. The functions $\bar{E}_{local}$ and $\bar{E}_{nonlocal}$ were then calculated as averages from all residues in the set of 114 protein structures within each bin.

The minima in the $\bar{E}_{nonlocal}$ function seen in Figure 1D for residues in β-sheet and α_R-helix arise from the greater number of hydrogen bonds in these states than elsewhere. As expected, $\bar{E}_{nonlocal}$ is more favorable for α_R-helix than for β-sheet, reflecting the greater average number of hydrogen bonds per residue in the former structure. Nevertheless, the total main chain energy, $\bar{E}_{total}$, still favors the β-sheet conformational state over the α_R-helix by $\approx$3.5 kcal/mol. Thus, if the electrostatic energies $\bar{E}_{local}$ and $\bar{E}_{nonlocal}$ were the only contributions to the free energy of a residue, then for all residue types, residues in β-sheet would be substantially more stable than residues in the α_R-helical conformational state.

The point charge model of main chain electrostatics is very approximate. However, the large size of the difference in energy for residues in different conformations that it yields and the qualitative agreement with the quantum mechanical calculations allow us to reach two conclusions. First, because main chain electrostatic interactions in proteins are so large, they are very likely to be centrally involved in determining

conformational preference. Second, electrostatic interactions with solvent and protein charges must modulate these primary main chain electrostatic effects, or there would be no helices in proteins.

Figure 1A suggests a mechanism by which the large energy difference between the β- and α_R-conformations may be reduced in a protein environment. The electric fields produced by the parallel dipoles of peptide groups flanking a residue in the α_R-conformation will reinforce each other, resulting in strong interactions with the dipoles of water molecules or other polar protein groups nearby. If these groups are oriented favorably, the resulting energy may offest the unfavorable intrabackbone contribution. Conversely, for a residue in the β-conformation, the peptide dipoles are antiparallel and result in a weak field and thus weaker interactions with other groups in the vicinity. Further, the degree to which such compensation can take place will depend on the access of water and protein groups to the backbone. Such access will depend partly on the side chain type. Bulky side chains and β-branched ones will restrict access more than small or slender ones. Using a potential of mean force analysis of protein energetics, we now show that the observed conformational preferences of residues in proteins are consistent with this differential screening determining their behavior. We also show that the average interaction energy of a residue with its environment is also consistent with this model.

Potential of Mean Force Analysis. Energetic analysis using the potential of mean force is based on the relationship between an observed population of states of a system and the free energy (McQuarrie, 1976; Beveridge & DiCapua, 1989):

$$G(R) = \Gamma(R) = -kT \ln(p(R)) + kT \ln(n(R)) + C \quad (1)$$

where R is usually referred to as the reaction coordinate; $G(R)$ is the free energy of those states as a function of R. $\Gamma(R)$ is the potential of mean force, that is, the reversible work required to take the system from a reference state to the reaction coordinate value R. $p(R)$ is the probability of finding the system in any of the states with a particular value of R, and $n(R)$ is the volume element or the normalization function, that is, the probability of finding the system with a reaction coordinate value of R in the absence of any of the interactions contributing to the potential energy. C is an undefined constant, T is temperature, and k is Boltzmann's constant.

The method has been used extensively in the statistical mechanics of fluids (McQuarrie, 1976; Pratt & Chandler, 1977; Pettitt & Rossky, 1986; Dang & Pettitt, 1987). Early applications of the potential of mean force to protein structure analysis have used simple models of the energetics, based on direct parametrization of the observed distribution of interresidue distances (Miyasawa & Jernigan, 1985; Crippen & Snow, 1990; Sippl, 1990), rather than a physically based energetic model. Potential of mean force analysis of experimental protein structures can also be used to obtain the free energy contributions of individual interactions and, thus, to test more detailed models of protein energetics. In an earlier paper (Avbelj, 1992), the relationships of the hydrophobic effect and other interactions to the burial of residues in proteins were investigated. In the present paper, we extend the method to investigate the role of main chain electrostatics in determining main chain conformation.

The local electrostatic energy, E_{local}, provides an appropriate generalized reaction coordinate (Beveridge & DiCapua, 1989) to use for the potential of mean force in the analysis of protein energetics. As shown earlier, E_{local} is able to differentiate between the most important conformational states of a residue, and it expresses conformation directly in terms of electrostatics. The potential of mean force, $\Gamma(E_{local})^r$, then is related to the population distribution of E_{local} for that residue in experimental protein structures by

$$\Gamma(E_{local})^r = P(E_{local})^r - N(E_{local})^r + C = -kT \ln(p(E_{local})^r) + kT \ln(n(E_{local})^r_{rc}) + C_1 \quad (2)$$

where r represents residue type. $P(E_{local})^r = -kT \ln(p(E_{local})^r)$ is a population function that depends on the probability, $p(E_{local})^r$, of a residue of type r having local electrostatic energy E_{local} in experimental protein structures. $N(E_{local})^r = -kT \ln(n(E_{local})^r_{rc})$ is a normalization function that depends on the probability, $n(E_{local})^r_{rc}$, of a residue of type r having local electrostatic energy (E_{local}) in random structures (rc) in which all interactions except steric overlap are zero. C_1 is an undefined constant, T is temperature, and k is Boltzmann's constant.

The $P(E_{local})^r$ functions were calculated from the set of 114 proteins by dividing the E_{local} range of -4.0 to 2.5 kcal/mol into 13 equal bins and counting the population of each residue type in these bins.

$N(E_{local})^r$ was compiled in the same manner from 500 random structures of length 100 residues. Sequences were generated by selecting residues randomly from the set of 114 proteins. The composition of randomly generated structures is therefore similar to those in real proteins. All torsion angles were selected randomly. Structures with any atomic contact shorter than 0.4 Å less than the sum of the van der Waals radii were rejected. The function $\bar{N}(E_{local})$ was calculated as the average of the $N(E_{local})^r$ functions for individual residues.

Figure 2 shows the population functions, $P(E_{local})^r = -kT \ln(p(E_{local})^r)$ (eq 2), for selected residue types. The population function is related to the more usual conformational preference of residues. For example, residues Ala and Val prefer α_R-helix and β-sheet, respectively. Accordingly, the value of $P(E_{local})^r$ for Ala in an α_R-helix ($E_{local} = 1.75$) is 0.5 kcal/mol lower than that of a residue in a β-sheet ($E_{local} = -3.25$), whereas for Val, $P(E_{local})^r$ in β-sheet is 0.5 kcal/mol lower than the value for that residue in an α_R-helix. For all residues except Pro, the correlation (Figure 3A) between the set of $-kT \ln(P^r_\alpha/P^r_\beta)$ values, where P^r_α and P^r_β are the Chou and Fasman conformational parameters (Chou & Fasman, 1978), and the differences in $P(E_{local})^r$ between the α_R-helix and β-sheet conformations ($\Delta P(E_{local})^r$) is 0.78.

Figure 2 also shows the function $\bar{N}(E_{local})$ for random coil structures (rc) averaged over all residue types except Pro. Conformations at both ends of the interval of E_{local} are less probable than those in the central region in these random structures. These regions correspond to α_R-helices and β-strands. Such repeating structures are unlikely to occur in the noncompact random coil reference state used here.

Qualitative Relationship between the Potential of Mean Force and Electrostatic Screening. The general form of the $P(E_{local})^r$ functions (Figure 2) for the large nonpolar amino acids, for example, Leu, is similar to that of $\bar{E}_{total}$ (see Figure

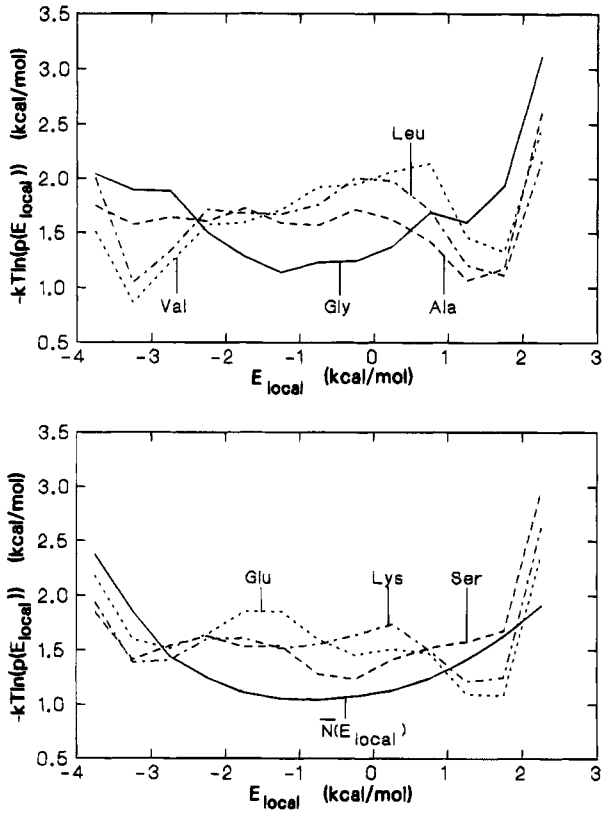

FIGURE 2: Observed population functions, $P(E_{local})^r = -kT$ ln-$(p(E_{local})^r)$ (eq 2), for seven residues: Gly, Ala, Val, Leu, Ser, Glu, and Lys. $p(E_{local})^r$ is the probability of finding residue type r with a local electrostatic energy E_{local} in an experimental structure. The function $\bar{N}(E_{local}) = -kT$ ln$(n(E_{local})_{rc})$ for random structures, averaged over all residue types except Pro, is also shown. The precision for $P(E_{local})^r$ and $\bar{N}(E_{local})$ was estimated as 0.2 kcal/mol (Avbelj, 1992).

1D), suggesting the dominance of main chain electrostatic interactions in determining the observed populations for these residues. Thus, when the main chain is isolated from solvent and other electrostatic factors, the average main chain conformation is seen to be consistent with main chain electrostatics. The features for individual residues are consistent with those expected from screening the main chain electrostatics in a manner dependent on the side chain type. For example, the function $P(E_{local})^r$ for Gly is similar in shape to those of random structures, suggesting that electrostatics and other interactions do not play a significant role in determining its population function. In this case, the absence of a side chain allows close approach of the water molecules and other protein groups, resulting in almost complete electrostatic screening of both local and nonlocal electrostatic interactions. For Ala, there is a minimum in the α_R-helical region, reflecting the intermediate degree of screening provided by the small side chain, affecting local interactions more than nonlocal ones. For residues with larger side chains, and therefore less screening, an additional minimum is found in the β-sheet region, as a consequence of the greater strength of the local electrostatic interactions. For β-branched residues, such as Val, there is little screening because of the protection of the main chain by the side chain; thus, the more favorable local electrostatics of the β-conformation makes this the preferred conformation.

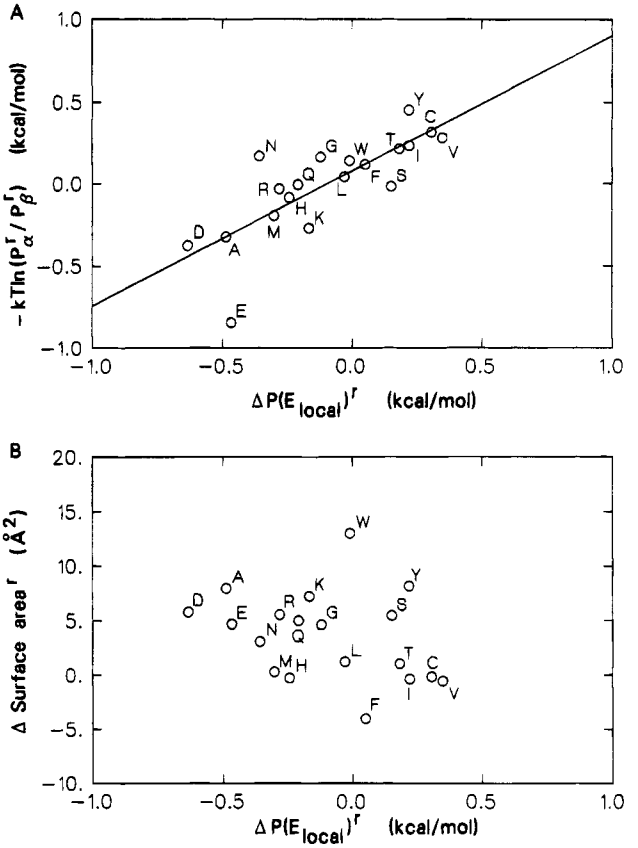

FIGURE 3: Relationship between the difference in the population function between α_R-helix and β-sheet conformations ($\Delta P(E_{local})^r$) and the Chou–Fasman conformational parameters and the difference in the nonpolar accessible surface area of amino acids in α_R-helix and β-sheet conformations. $\Delta P(E_{local})^r$ was calculated from average values of $P(E_{local})^r$ in E_{local} intervals of 1.25–1.75 kcal/mol and −3.25 to −2.75 kcal/mol for α_R-helix and β-sheet, respectively. (A) Relationship between $-kT$ ln(P_α^r/P_β^r) and $\Delta P(E_{local})^r$. P_α^r and P_β^r are the Chou–Fasman conformational parameters (Chou & Fasman, 1978). $\Delta P(E_{local})^r$ is the difference between the population functions ($P(E_{local})^r$, see eq 2) for amino acid r in α_R-helix and β-sheet conformations. The correlation coefficient is 0.78. Pro was excluded. (B) Relationship between the difference in nonpolar accessible surface area of amino acids in α_R-helix and β-sheet conformations (Δ(surface area)) and $\Delta P(E_{local})^r$. $\Delta P(E_{local})^r$ is the difference between the population functions ($P(E_{local})^r$, see eq 2) for amino acid r in α_R-helix and β-sheet conformations. The correlation coefficient is 0.30. Thus, the area model does not account for the observed differences in population for these conformational states. The corresponding correlation coefficient for the electrostatic model of conformational preference is 0.77. In both cases, Pro was excluded.

Quantitative Relationship between the Potential of Mean Force and Electrostatic Screening. The importance of different contributions to the free energy of a residue may be tested by comparing the potentials of mean force derived from the experimental protein structures, $\Gamma(E_{local})^r$, with the linear combination of the free energy contributions from interactions of a residue:

$$\Gamma(E_{local})^r = G(E_{local})^r = \sum_i \gamma_i^r \bar{\epsilon}_i^r + C_2 \qquad (3)$$

where $\bar{\epsilon}_i^r$ is a property i of the system averaged over each E_{local} bin of residue type r as a function of E_{local}. C_2 is an undefined constant. Each $\bar{\epsilon}_i^r$ is related to a physical contribution to the free energy. The coefficients γ_i^r are used to

parametrize the model and are conversion factors between the system properties and the corresponding contribution to the free energy.

A correct model of the relative free energy of a residue as a function of main chain conformation should be able to reproduce the potential of mean force for all conformations of all residue types, $\Gamma(E_{local})^r$. To assess the appropriateness of an electrostatic model, we express the average free energy of a residue as a sum of two contributions, the local and nonlocal electrostatics, and examine the agreement between calculated average free energy and the potential of mean force over that range of data. In this model, the free energy contributions from the main chain electrostatic interactions are assumed to be proportional to the point charge interaction energies, with coefficients dependent on residue type:

$$\Gamma(E_{local})^r = \gamma^r_{nonlocal}\bar{E}^r_{nonlocal} + \gamma^r_{local}\bar{E}^r_{local} + C_3 \quad (4)$$

where $\bar{E}^r_{local}$ and $\bar{E}^r_{nonlocal}$ are the average local and nonlocal main chain electrostatic energies as a function of E_{local} for an individual residue type r. These quantities are calculated using Coulomb's law and averaging over each of the 13 E_{local} bins, as described earlier. Coefficients $\gamma^r_{nonlocal}$ and γ^r_{local} thus represent the attenuation of the electrostatic energies $E_{nonlocal}$ and E_{local}, respectively, due to the electrostatic screening and are constants depending only on the residue type (r). $\gamma^r_{nonlocal}\bar{E}^r_{nonlocal}$ and $\gamma^r_{local}\bar{E}^r_{local}$ represent the average free energy contributions from nonlocal and local electrostatic interactions, respectively, as a function of E_{local}. C_3 is a an undefined constant.

Given the functions $\Gamma(E_{local})^r$, $\bar{E}^r_{nonlocal}$, and $\bar{E}^r_{local}$, the values of $\gamma^r_{nonlocal}$ and γ^r_{local} are derived by a linear least-squares fit for each residue type. The coefficients $\gamma^r_{nonlocal}$ and γ^r_{local} and the residual standard deviations of the fit of eq 4 are shown in Table 1. The deviations are of approximately the same magnitude as the average errors in $\Gamma(E_{local})^r$ (see legend to Figure 2). Thus, the electrostatic model in eq 4 does represent the average free energy distributions of different residue types faithfully. That is, main chain electrostatics, with only the addition of residue type scaling coefficients, is able to reproduce the observed distribution of all main chain conformations of all residue types (except Pro) in proteins. Since the observed distributions as a function E_{local} are fairly complex and markedly different for different residue types (see Figure 2), the goodness of fit provides direct evidence for the importance of electrostatics in determining conformation.

Some other model types can be tested by these potentials of mean force. It has been proposed that hydrophobicity, as reflected by accessible nonpolar surface area, primarily determines the preferences of residues for the α_R-helix (Blaber *et al.*, 1993; Horovitz *et al.*, 1992). To test this hypothesis, the average nonpolar accessible surface area in the two E_{local} regions representing the α_R-helix and β-sheet conformations was calculated for each residue, using the same set of proteins. The Lee and Richards algorithm (Lee & Richards, 1971) was used, together with the atomic radii of Chothia (1975). Figure 3B shows the relationship between the difference in nonpolar accessible surface area of amino acids in α_R-helix and β-sheet conformations (Δ(surface area)) and $\Delta P(E_{local})^r$. $\Delta P(E_{local})^r$ is the difference between the population function ($P(E_{local})^r$, see eq 2) for each amino acid r in the α_R-helix and β-sheet conformations. Δ(surface area) and $\Delta P(E_{local})^r$ were calculated from average values in the E_{local} intervals 1.25−1.75 kcal/mol and −3.25 to −2.75 kcal/mol for α_R-helix and β-sheet, respectively. The correlation coefficient between Δ(surface area) and $\Delta P(E_{local})^r$ for all residues except Pro is 0.30, showing that this model of the hydrophobic effect is not able to account for the differences in populations between α_R-helix and β-sheet. The corresponding correlation coefficient using the electrostatic model is 0.77.

Table 1: Coefficients $\gamma^r_{nonlocal}$ and γ^r_{local} Relating the Total Free Energy of Different Residues to the Average Main Chain Electrostatic Energy (See Equation 4)[a]

residue	$\gamma^r_{nonlocal}$	γ^r_{local}	rsd (Γ) (kcal/mol)
Gly	0.07	0.12	0.25
Ala	0.48	0.17	0.21
Val	0.73	0.40	0.21
Ile	0.76	0.43	0.21
Leu	0.76	0.29	0.18
Phe	0.68	0.37	0.19
Pro			
Met	0.75	0.34	0.27
Trp	0.46	0.21	0.17
Cys	0.45	0.23	0.21
Ser	0.35	0.17	0.20
Thr	0.55	0.18	0.15
Asn	0.36	0.11	0.18
Gln	0.56	0.21	0.22
Tyr	0.64	0.28	0.15
His	0.45	0.21	0.28
Asp	0.14	−0.01	0.18
Glu	0.42	0.11	0.25
Lys	0.53	0.19	0.12
Arg	0.52	0.22	0.17

[a] $\gamma^r_{nonlocal}$ and γ^r_{local} mainly represent the attenuation of the electrostatic energies $E_{nonlocal}$ and E_{local}, respectively, due to screening by water and protein dipoles. The effects due to the inaccuracy of the point charge model and any entropic, steric, or hydrophobic contributions are also contained in the coefficients. The differences in these values are mainly responsible for the observed preferences of residues for different conformations. The average γ^r_{local} from all residues is similar to $1/\epsilon$ obtained by Brant and Flory in fitting the end to end distances of peptide polymers (Brant & Flory, 1965a,b). The residual standard deviations of the fit of eq 4 (rsd) are also shown.

Physical Significance of the Residue Dependent Scaling Coefficients. Since main chain electrostatics is seen to be important in determining conformation, the coefficients $\gamma^r_{nonlocal}$ and γ^r_{local} should primarily represent the average residue dependent screening of electrostatic interactions. If this is the case, the coefficients will be related to the average local electrostatic interaction energies of the main chain with solvent and other polar protein groups, $\bar{E}^r_{env}$, by

$$\gamma^r_{local}\bar{E}^r_{local} = \bar{E}^r_{local} + \bar{E}^r_{env} \quad (5)$$

where γ^r_{local} is the screening coefficient for local interactions, the $\bar{E}^r_{local}$ is the average local main chain electrostatic energy function (see eq 4). There is an analogous relationship for the nonlocal interactions. The screening coefficients in eq 5 are inversely related to the microscopic screening function described by Warshel and Russell (1984). Note that, in the present approximation (eq 5), the screening coefficients depend only on the residue side chain type.

Accurate calculation of absolute values of E^r_{env} is difficult (Warshel & Russell, 1984), and we do not attempt it here.

The relative values, however, are more reliable and can be used to test the model. Approximate values of E^r_{env} were calculated from the set of experimental protein structures, using empirical force field partial charges for the protein atoms (Dauber-Osguthorpe *et al.*, 1988) and an iterative Langevin dipole model for the solvent (Russell & Warshel, 1985). A cutoff of 6.5 Å was used with a dielectric constant of 1. Polarization due to induced charges was not included. All other parameters for the calculation were as given in Russell and Warshel (1985). In order to avoid the influence of main chain hydrogen bonding in secondary structures, the interval of E_{local} was restricted to the range −2.75 to −0.75 kcal/mol (see Figure 1D). The values of $\bar{E}^r_{env}$ as a function E_{local} are obtained from the set of experimental protein structures by averaging E^r_{env} in the same way as for $\bar{E}^r_{local}$ and $\bar{E}^r_{nonlocal}$ (see eq 4).

As discussed earlier, we would expect that the stronger electric field created by the parallel dipoles of two adjacent peptide groups in the α_R-conformation would interact more favorably with the surrounding water and polar protein groups than would the weaker field around the antiparallel dipoles in the β-conformation. Indeed, as Figure 4A shows, for all residues, $\bar{E}^r_{env}$ varies inversely with E_{local} and is substantially larger in the region of α_R-conformation than in the β region. Thus, the unfavorable local main chain energy in the α_R-conformation is on average offset by more favorable electrostatic interactions with the surroundings. Further, we would expect that the more bulky the side chain of a residue near main chain atoms, the weaker the interactions with the surroundings, since water and protein groups cannot approach as closely. Figure 4A shows this to be the case. For example, $\bar{E}^r_{env}$ is larger for Gly than for the more bulky side chains, such as Val. $\bar{E}^r_{env}$ is also smaller for the β-branched amino acids Val and Ile than for the larger Phe, because the β-branch significantly reduces the space available for solvent and other polar groups near the NH and CO groups. More quantitatively, for eq 4 to be true, γ^r_{local} should be proportional to the ratio of $\bar{E}^r_{env}$ to E_{local} for different residue types. Figure 4B shows the relationship between the coefficients γ^r_{local} and the slopes of $\bar{E}^r_{env}$ versus E_{local}, derived from Figure 4A. The correlation coefficient is 0.95, strongly supporting the model.

Comparison with Other Experimental Data. The screening model should predict other experimental properties, such as the helix propagation propensities of different residue types, and free energy changes in folding associated with single mutations of proteins.

Helix formation is formally divided into initiation and propagation stages in the Zimm and Bragg theory (Zimm & Bragg, 1959). Initiation requires three consecutive residues to adopt an α_R-conformation so that the first hydrogen bond can form. According to our model, initiation primarily depends on the strength of local electrostatic interactions of three adjacent residues reflected by the γ^r_{local} values and on the contribution from main chain conformational entropy. Propagation primarily depends on the strength of main chain hydrogen bonds, reflected by the $\gamma^r_{nonlocal}$ values. Propagation parameters for different residue types have been measured for amino acids in random host−guest copolymers (Wojcik *et al.*, 1990). Figure 5A compares $-kT\ln(s^r)$, where s^r is the helix growth parameter (Wojcik *et al.*, 1990), with

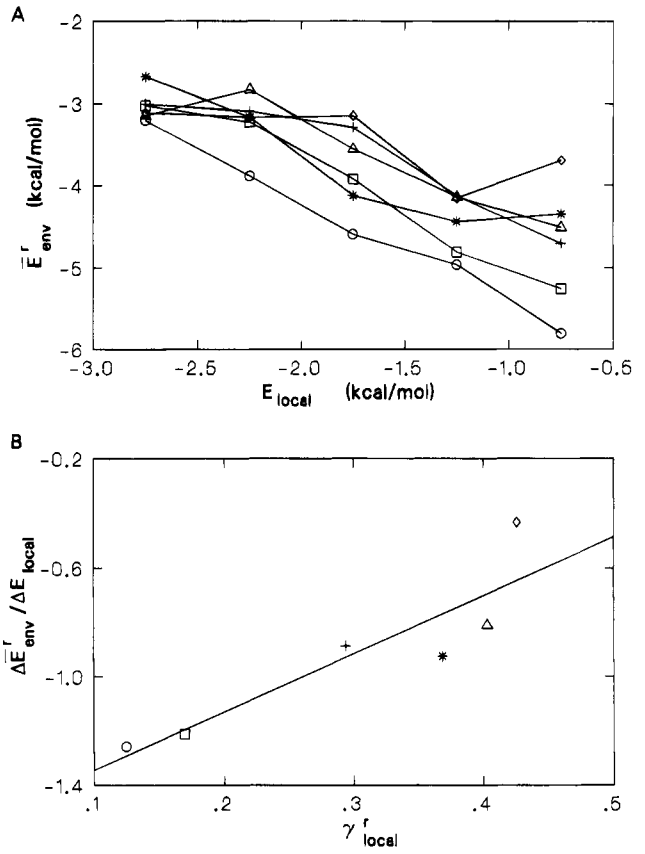

FIGURE 4: Relationship of the screening coefficient γ^r_{local} to electrostatics. (A) Average electrostatic interaction energy of main chain atoms with water molecules [represented by Langevin dipoles (Russell & Warshel, 1985)] and polar protein groups, $\bar{E}^r_{env}$, as a function of E_{local} for nonpolar residues: Gly (○), Ala (□), Val (△), Ile (◇), Leu (+), and Phe (*). $\bar{E}^r_{env}$ was calculated from the experimental protein structures. (B) Relationship between the slopes of $\bar{E}^r_{env}$ with respect E_{local}, obtained by a least-squares fit to the data shown in part A, and the local electrostatic screening coefficient γ^r_{local}: Gly (○), Ala (□), Val (△), Ile (◇), Leu (+), and Phe (*). The correlation coefficient is 0.95, confirming that γ^r_{local} reflects the screening of local main chain electrostatic interactions by water and other polar protein groups.

$\gamma^r_{nonlocal}$. T is temperature, and k is Boltzmann's constant. The correlation coefficient is 0.82. Thus, the model is able to directly predict the propagation propensities of different residues.

There are now data available for the relative free energies of different residues in secondary structures in proteins, obtained by measuring the stability as a function of multiple amino acid substitutions at particular sites. These data provide a good test of the electrostatic model.

The relative electrostatic free energies of unfolding $\Delta\Delta G_{calc}$ sets were calculated as the difference in the free energies of the denatured and native states, for all residues except Pro. Native state relative free energies were obtained using eq 2, with average local and nonlocal energies of 1.89 and −4.69 kcal/mol for residues in helices and −3.2 and −3.0 kcal/mol for those in β-strands, respectively. In order to obtain the relative free energies of mutants in the denatured state, we must first evaluate the probability, $P(E_{local})^r$, of any conformation of a residue in that state. To do this, we assume that there is no hydrogen bonding in the denatured state and that local electrostatics and steric factors alone

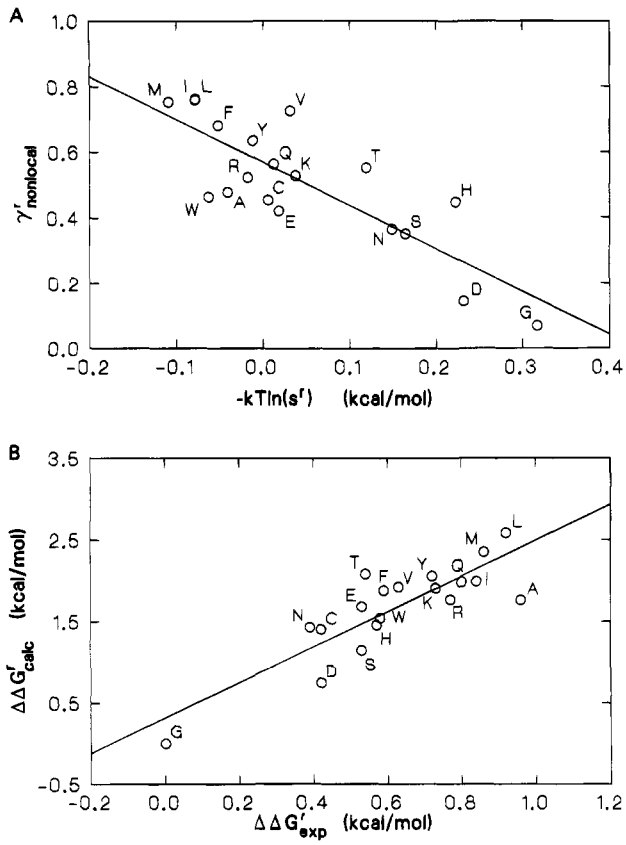

FIGURE 5: Comparison of the electrostatic model with experimental data. (A) Relationship between $\gamma^r_{nonlocal}$ and $-kT\ln(s^r)$, where s^r is the experimental helix propagation parameter (Wojcik *et al.*, 1990) for all residues except Pro. T is temperature, and k is Boltzmann's constant. The correlation coefficient is 0.82. The high correlation confirms that $\gamma^r_{nonlocal}$ reflects the relative average strength of main chain hydrogen bonds for different residue types. (B) Relationship between calculated ($\Delta\Delta G^r_{calc}$) and experimental ($\Delta\Delta G^r_{exp}$) (Blaber *et al.*, 1993) free energies of unfolding of mutants of T4 lysozyme. All residues other than Pro are shown. The correlation coefficient is 0.85.

determine $P(E_{local})^r$. Then, by combining eqs 2 and 4, we have

$$P(E_{local})^r = \gamma^r_{nonlocal}\bar{E}_{local} + N(E_{local})^r + C_4 \qquad (6)$$

where C_4 is an undefined constant. $P(E_{local})^r$ was evaluated for the 13 values of E_{local} for each residue type. Average local electrostatic energies $\langle E^r_{local}\rangle$ were then obtained by Boltzmann integration over E_{local}. The relative free energy of a residue of type r in the denatured state is then given by $\gamma^r_{local}\langle E^r_{local}\rangle$.

The relative free energies of residues in an α_R-helix have been determined at a solvent-exposed position in T4 lysozyme (Blaber *et al.*, 1993). For the first time, X-ray structures are also available for a number of the mutants. In this case, there are no substantial conformational changes, and the mutant side chains are largely ordered, suggesting that configurational entropy effects will be minor. Existing helix propensity scales fit the experimental free energy differences poorly.

Figure 5B shows the relationship between the calculated ($\Delta\Delta G_{calc}$) and experimental ($\Delta\Delta G_{exp}$) (Blaber *et al.*, 1993) relative free energies of unfolding for the lysozyme mutants. The correlation coefficient between these quantities, including all residues except Pro, is 0.85, which is substantially better than that with any other scale when so many residue types are included. In contrast to the data for T4 lysozyme, the correlation between the relative calculated and experimental free energies for substitutions at position 32 in a helix in barnase (Horovitz *et al.*, 1992) is very poor: 0.35. However, since the structures of the barnase mutants are not known, it is difficult to evaluate the significance of this poor fit. In particular, the role of configurational entropy may be more important in this case, since the adjacent side chains in the native structure are rather disordered. This possibility is also suggested by the high correlation between the relative free energies of the barnase data and those for substitutions in a poorly ordered four-helix bundle (Horovitz *et al.*, 1992).

Experimental results for two full sets of mutants at particular positions in β-sheets can be compared with the model free energies. For substitutions at a position in a β-strand in a zinc finger (Kim & Berg, 1993), the correlation coefficient between the experimental and calculated relative free energies is a moderately good 0.72. However, the calculated range of free energies is substantially larger than the experimental one. No structural data are available for this case. The most recent data are for substitutions at position 53 in a β-strand of the B1 domain of protein G (Minor & Kim, 1994). Here, surrounding side chains were truncated to alanine to minimize the role of side chain−side chain interactions and, thus, make β propensity the dominant factor. NMR data suggest that the mutant structures are essentially native-like. The correlation between the experimental and calculated free energies is 0.84. In this case, a similar range of free energies is found (2.2 kcal/mol for the experimental data and 2.3 kcal/mol for the calculated values).

Evidence for the Electrostatic Model. In summary, five lines of evidence support the importance of main chain electrostatics in determining residue conformational preferences: (1) Quantum mechanical and empirical energy calculations show the difference in local main chain energy as a function of conformation to be large and to arise primarily from main chain electrostatics. (2) A potential of mean force analysis shows that the observed distributions of conformations for large nonpolar amino acids in proteins have a functional form similar to that of the total electrostatic energy (compare Figures 2 and 1D). (3) The observed free energies of residues, $\Gamma(E_{local})^r$, for all residue types (except Pro) can be reproduced with a simple electrostatic model (Table 1). (4) The residue type dependent coefficients of the electrostatic model correlate strongly with electrostatic screening by water and other protein group interactions (Figure 4B). (5) The coefficients correlate with experimental helix propagation data (Figure 5A). (6) There is good correlation between the calculated and experimental relative free energies of residues in α_R and β secondary structure for well-characterized cases (Figure 5B and the previous section). Further support for the electrostatic model comes from its effectiveness as part of a potential used to determine the conformation of folding of initiation sites by a Monte Carlo procedure (Avbelj & Moult, 1995).

Relationship to Other Models. The view that main chain electrostatics is central to the determination of local conformation contrasts strongly with the other current opinions that conformational entropy (Creamer & Rose, 1992; Padmanabhan & Baldwin, 1991), steric factors (Hermans *et al.*, 1992; Yun & Hermans, 1991), and the hydrophobic effect (Horovitz

et al., 1992; Blaber *et al.*, 1993) are primarily responsible for residue conformational preferences.

Surface area effects have been advanced to explain mutant protein data in two cases (Blaber *et al.*, 1993; Horovitz *et al.*, 1992). A weak correlation between the nonpolar surface area lost in folding and the free energies of unfolding was found for the T4 lysozyme mutants (Blaber *et al.*, 1993). Comparison of Figure 5B in this paper and Figure 2 in Blaber *et al.* (1993) shows that the electrostatic model fits the experimental data substantially better than does the nonpolar surface area lost in folding. These are the first relevant free energy data for which corresponding X-ray structures are known. Better evaluation of the relative significance of all possible factors affecting conformational preference will be possible when more such information is available.

Creamer and Rose (1992) have used a Monte Carlo procedure to estimate side chain entropy and have shown that it correlates well with the relative propensity of nonpolar residues for the helical state in peptides. The observed conformation distributions of different residue types (Chou & Fasman, 1978) in proteins correlate rather poorly with the data for helical preference in peptides (Lyu *et al.*, 1990; Padmanabhan & Baldwin, 1991; O'Neal & DeGrado, 1990), and it may be that different mechanisms are predominant for peptides and proteins. For peptides, side chains in the pseudo-folded state are not strongly constrained by their neighbors, so that the relative conformational freedom of the side chains as a function of backbone conformation makes a marked entropic contribution to the conformational preference. In a folded protein molecule, particularly for the hydrophobic side chains, interactions with other residues result in essentially complete ordering, independent of the main chain conformation, and the entropy cost of folding does not depend on the final conformation. In both situations, the electrostatic factors will be operational.

All of these models can explain, with limited success, the preference of residues solely for the α_R-helical state. The electrostatic model with two screening coefficients per residue type is able to explain the preferences of residues in experimental protein structures not only for α_R-helical but also for β-sheet and all other main chain conformational states. It should be stressed that the electrostatic model shows that, on average, screening of the main chain electrostatic interactions is important in determining conformational preference. In a particular environment in a particular protein, configurational entropy, steric factors, hydrophobicity, and other types of electrostatic interactions and solvation effects will of course also play a role in determining the preferred conformation and the free energy difference between the unfolded state and the native conformation.

REFERENCES

Avbelj, F. (1992) Use of a potential of mean force to analyse free energy contributions in protein folding, *Biochemistry 31*, 6290−6297.

Avbelj, F., & Moult, J. (1995) The conformation of folding initiation sites in proteins determined by computer simulation, *Proteins: Struct., Funct., Genet.* (in press).

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tusami, M. (1977) The protein data bank: A computer-based archival file for macromolecular structures, *J. Mol. Biol. 112*, 535−542.

Beveridge, D. L., & DiCapua, F. M. (1989) in *Computer Simulations of Biomolecular Systems. Theoretical and Experimental Applications* (van Gunsteren, W. F., & Weiner, P. K., Eds.) pp 1−26. Escom, Leiden, The Netherlands.

Blaber, M., Zhang, X., & Matthews, B. W. (1993) Structural basis of amino acid alpha helix propensity, *Science 260*, 1637−1640.

Brant, D. A., & Flory, P. J. (1965a) The configuration of random polypeptide chains. ii. theory, *J. Am. Chem. Soc. 87*, 2791−2800.

Brant, D. A., & Flory, P. J. (1965b) The role of dipole interactions in determining polypeptide conformation, *J. Am. Chem. Soc. 87*, 663−664.

Chothia, C. (1975) Structural invariants in protein folding, *Nature 254*, 304−308.

Chou, P. Y., & Fasman, G. D. (1974) Conformational parameters for amino acid in helical, beta-sheet, and random coil regions calculated from proteins, *Biochemistry 13*, 211−222.

Chou, P. Y., & Fasman, G. D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence, *Adv. Enzymol. 47*, 45−148.

Creamer, T. P., & Rose, G. D. (1992) Side-chain entropy opposes alpha-helix formation but rationalizes experimentally determined helix-forming propensities, *Proc. Natl. Acad. Sci. U.S.A. 89*, 5937−5941.

Crippen, G. M., & Snow, M. E. (1990) A 1.8 Å resolution potential function for protein folding, *Biopolymers 29*, 1479−1489.

Dang, L. X., & Pettitt, B. M. (1987) Chloride ion pairs in water, *J. Am. Chem. Soc. 109*, 5531−5532.

Dauber-Osguthorpe, P., Roberts, V., Osguthorpe, D. J., Wolff, J., Genest, M., & Hagler, A. T. (1988) *Proteins: Struct., Funct., Genet. 4*, 31−47.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins, *J. Mol. Biol. 120*, 97−120.

Head-Gordon, T., Head-Gordon, M., Frisch, M. J., Brooks, C. L., & Pople, J. A. (1991) Theoretical study of blocked glycine and alanine peptide analogues, *J. Am. Chem. Soc. 113*, 5989−5997.

Hermans, J., Anderson, A. G., & Yun, R. H. (1992) Differential helix propensity of small apolar side chain studied by molecular dynamics simulation, *Biochemistry 31*, 5646−5653.

Hol, W. G. J. (1987) The role of the α-helix dipole in protein function and structure, *Prog. Biophys. Mol. Biol. 45*, 149−195.

Horovitz, A., Matthews, J. M., & Fersht, A. (1992) Alpha-helix stability in proteins ii. factors that influence stability at an internal position, *J. Mol. Biol. 227*, 560−568.

Kemp, D. S.; Boyd, J. G., & Muendel, C. C. (1991) The helical s constant for alanine in water derived from template-nucleated helices, *Nature 352*, 451−454.

Kim, C. A., & Berg, J. M. (1993) Thermodynamic beta-sheet propensities measured using a zinc finger host peptide, *Nature 362*, 267−270.

Kitson, D. H., Avbelj, F., Moult, J., Nguyen, D. T., Mertz, J. E., Hadži, D., & Hagler, A. T. (1993) On achieving better than 1 angstrom accuracy in a simulation of a large protein: Streptomyces griseus protease a, *Proc. Natl. Acad. Sci. U.S.A. 90*, 8920−8924.

Lee, B., & Richards, F. M. (1971) The interpretation of protein structures: Estimation of static accessibility, *J. Mol. Biol. 55*, 379−400.

Lyu, P. C., Liff, M. I., Marky, L. A., & Kallenbach, N. R. (1990) Side chain contribution to the stability of alpha-helical structure in proteins, *Science 250*, 669−673.

McGregor, M. J., Islam, S. A., & Sternberg, M. J. E. (1987) Analysis of the relationship between side-chain conformation and secondary structure in globular proteins, *J. Mol. Biol. 198*, 295−310.

McQuarrie, D. A. (1976) *Statistical Mechanics*, Chapter 13, Harper and Row, New York.

Merutka, G., Lipton, W., Shalongo, W., Park, S.-H., & Stellwagen, E. (1990) Effect of central-residue replacement on the helical stability of a monomeric peptide, *Biochemistry 29*, 7511−7515.

Minor, D. L., & Kim, P. S. (1994) Measurement of the β-sheet-forming propensities of amino acids, *Nature 367*, 660−663.

Miyasawa, S., & Jernigan, R. L. (1985) Estimation of effective inter-residue contact energies from protein crystal structures: Quasi-chemical approximation, *Macromolecules 18*, 534−552.

O'Neal, K. T., & Degrado, W. F. (1990) A thermodynamics scale for the helix-forming tendencies of the commonly occuring amino acids, *Science 250*, 646−651.

Padmanabhan, S., & Baldwin, R. L. (1991) Straight-chain nonpolar amino acids are good helix-formers in water, *J. Mol. Biol. 219*, 135−137.

Padmanabhan, S., Marqusee, S.; Ridgeway, T., Laue, T. M., & Baldwin, R. L. (1990) Relative helix-forming tendencies of nonpolar amino acids, *Nature 344*, 268−270.

Pettitt, B. M., & Rossky, P. J. (1986) Alkali halides in water: Ion-solvent correlations and ion-ion potentials of mean force at infinite dilution, *J. Chem. Phys. 84*, 5836−5844.

Piela, L., Nemethy, G., & Sheraga, H. A. (1987) Conformational constraints of amino acid side chain in alpha-helices, *Biopolymers 26*, 1273−1286.

Pratt, L. R., & Chandler, D. (1977) Theory of the hydrophobic effect, *J. Chem. Phys. 67*, 3683−3704.

Russell, S. T., & Warshel, A. (1985) Calculations of electrostatic energies in proteins. The energetics of ionized groups in bovine pancreatic trypsin inhibitor, *J. Mol. Biol. 185*, 389−404.

Šali, D., Bycroft, M., & Fersht, A. R. (1988) Stabilization of protein structure by interactions of α-helix dipole with a charged side-chain, *Nature 335*, 740−743.

Serrano, L., & Fersht, A. R. (1989) Capping and α-helix stability, *Nature 342*, 296−299.

Sippl, M. J. (1990) Calculation of conformational ensembles from potentials of mean force, *J. Mol. Biol. 213*, 859−883.

Warshel, A., & Russell, S. T. (1984) Calculation of electrostatic interactions in biological systems and in solutions, *Q. Rev. Biophys. 17*, 283−422.

Wojcik, J., Altmann, K.-H., & Scheraga, H. A. (1990) Helix-coil stability constants for the naturally occuring amino acids in water. xxiv. Half-cysteine parameters from random poly(hydroxybutylglutamine-co-s-methylthio-l-cysteine), *Biopolymers 30*, 121−134.

Yun, R. H., & Hermans, J. (1991) Conformational equilibria of valine studied by dynamics simulation, *Protein Eng. 4*, 761−766.

Yun, R. H., Anderson, A., & Hermans, J. (1991) Proline in alpha-halix: Stability and conformation studied by dynamics simulation, *Proteins 10*, 219−228.

Zimm, B. H., & Bragg, J. K. (1959) Theory of the phase transition between helix and coil in polypeptide chains, *J. Chem. Phys. 31*, 526−535.

BI940678P