Origin of the neighboring residue effect on peptide backbone conformation

Franc Avbelj[†‡§] and Robert L. Baldwin[‡§]

[†]National Institute of Chemistry , Hajdrihova 19, Ljubljana Sl 1115, Slovenia; and [‡]Department of Biochemistry, Beckman Center, Stanford University Medical Center, Stanford, CA 94305-5307

Contributed by Robert L. Baldwin, June 7, 2004

Unfolded peptides in water have some residual structure that may be important in the folding process, and the nature of the residual structure is currently of much interest. There is a neighboring residue effect on backbone conformation, discovered in 1997 from measurements of $^3J_{HN\alpha}$ coupling constants. The neighboring residue effect appears also in the "coil library" of Protein Data Bank structures of residues not in α-helix and not in β-structure. When a neighboring residue ($i - 1$ or $i + 1$) belongs to class L (aromatic and β-branched amino acids, FHITVWY) rather than class S (all others, G and P excluded), then the backbone angle ϕ of residue i is more negative for essentially all amino acids. Calculated values of peptide solvation (electrostatic solvation free energy, ESF) predict basic properties of the neighboring residue effect. We show that L amino acids reduce the solvation of neighboring peptide groups more than S amino acids. When tripeptides from the coil library are excised to allow solvation, the central residues have more negative values of $\langle\phi\rangle$ but less negative values of $\langle$ESF$\rangle$ with L than with S neighbors. The coil library values of $\langle^3J_{HN\alpha}\rangle$, which vary strikingly among the amino acids, are correlated with the neighboring residue effect seen by ESF. Moreover, values for the "blocking effect" of side chains on the hydrogen exchange rates of peptide NH protons are correlated with ESF values.

Short peptides, too short to form any structure that is stabilized by peptide H bonds, nevertheless have significant structure. The first clear demonstration was the 1997 discovery by Penkett *et al.* (1) of a neighboring residue effect. The $^3J_{HN\alpha}$ coupling constant in NMR studies of peptides is directly related to the backbone angle ϕ by the Karplus relation (2). The ϕ value of residue X is more negative when its preceding neighbor is class L (aromatic and β-branched, FHITVWY) rather than class S (all other amino acids, Gly and Pro omitted) (1). This behavior is found for all amino acids, except that Gly and Pro are omitted for reasons given below. The physical cause of the neighboring residue effect is unknown. We show here that peptide backbone solvation and the shielding from water of peptide groups by adjacent side chains are the probable causes.

The coil library of residue structures from the Protein Data Bank, not in α-helix and not in β-sheet, was used as a working model of the backbone conformations of residues in unfolded peptides (3–8). In the coil library, average values of the backbone angle ϕ and also of $\langle^3J_{HN\alpha}\rangle$ calculated from ϕ differed substantially among the amino acids (3–8). The same values of $\langle^3J_{HN\alpha}\rangle$ were found in the coil library as in NMR studies of unfolded peptides, within error (5, 8). The coil library is said to reveal the true propensities of different amino acids to form various backbone conformations (3–6). It is widely accepted that protein structures are made from "low-energy parts" (9), and consequently peptide structures provide values not only for bond angles and distances and prevalence of the trans peptide bond but also for the most probable side chain rotamers (9). Here we turn this hypothesis around and use protein structures to provide the distributions of backbone angles ϕ and ψ in unfolded peptides. Because peptide H bonds constrain the backbone angles in α-helices and β-sheets, we omit residues from these secondary structures and use the coil library instead of the "all" residue library. Protein structures are tight-packed, including residues in the coil library, but unfolded peptides are not. This distinction suggests that use of the coil library to model unfolded peptides should be restricted to data from single residues or perhaps pairs of residues. Because the neighboring residue effect appears in both unfolded peptides (1) and the coil library, it seems that the coil library models at least pairs of unfolded residues.

Our thesis is that solvation of the peptide backbone, with side chain shielding of peptide groups from water, has the properties needed to explain the neighboring residue effect. Electrostatic solvation results from the interaction between water and partial charges on the peptide NH and CO groups, and we argue that electrostatic solvation free energy (ESF, also known as the Born term) is responsible for the neighboring residue effect. DELPHI was used here to calculate ESF. The structure of the molecule must be known accurately, but other parameters are not adjustable (10). The PARSE parameter set of DELPHI (10) has been calibrated against experimental solvation free energies of model compounds, including amides. Thus, calculated values of ESF are based on experimental solvation free energies. Solvation is a stabilizing interaction between the peptide group and water, and ESF values are free energies with a negative sign and are reported here in kcal/mol.

A self-evident property of ESF is that it exhibits a neighboring residue effect. The properties of this effect are investigated here with two different systems. The first system is a model peptide system, acetyl-A_4XA_4-amide, where X is any amino acid and the peptide has the extended β conformation ($-120°$, $120°$). The peptide length was chosen to avoid end effects. The second system is the coil library, used here as a putative model of backbone conformations in denatured proteins. Tripeptides (with blocking groups added) were excised from 1,476 high-resolution structures to allow solvation to occur. The effects of L and S neighbors on the values of $\langle\phi\rangle$ and of $\langle$ESF$\rangle$ of the various amino acid types are studied.

Finally, we look for a possible correlation between the neighboring residue effect seen by ESF and certain experimental quantities suggested by the model peptide results. The first quantity is the coil library value of $\langle^3J_{HN\alpha}\rangle$, which varies strikingly among the amino acids (3–8). The second quantity is the "blocking effect" observed in measurements of hydrogen exchange rates in dipeptides (11, 12); the amino acid side chains shield adjacent peptide NH protons from exchange with solvent.

Methods

Representation of peptides, calculation of ESF from peptide structure, and calculation of $^3J_{HNa}$ from ϕ were performed as described (8). $\langle^3J_{HN\alpha}\rangle$ was found from ϕ values in the coil library by first calculating $^3J_{HNa}$ and then averaging. "PARSE-neutral"

Abbreviations: PPII, polyproline II; ESM, electrostatic screening model; ESF, electrostatic solvation free energy.

§To whom correspondence may be addressed. E-mail: francl@sg3.ki.si or bbaldwin@pmgm2.stanford.edu.

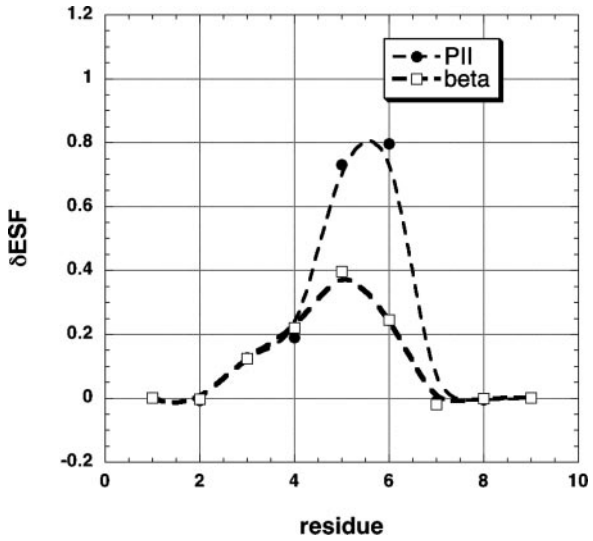

Fig. 1. The change in ESF (δESF) resulting from amino acid substitution, plotted against residue number. The host peptide is a nine-residue alanine peptide, and the substitution is Ala→Val at residue 5. The units of ESF are kcal/mol. At all residues except the substitution site, the alanine host peptide has the extended β backbone conformation (ϕ, ψ = −120°, 120°). At the substitution site, the residue (either Ala or Val) has either the extended β or an approximate PPII (ϕ, ψ = −70°, 150°) conformation.

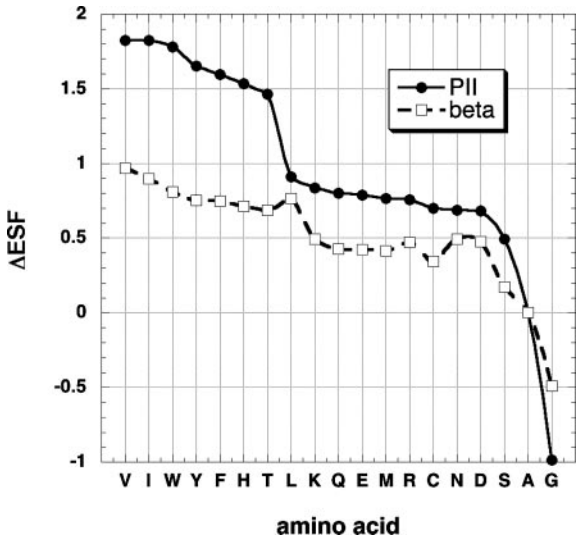

Fig. 2. The overall change in ESF (ΔESF, kcal/mol) produced by an amino acid substitution in an $(Ala)_9$ peptide (see legend to Fig. 1) plotted against amino acid type. The first seven amino acids (VIWYFHT) belong to group L, as defined earlier (1), whereas the other amino acids (except for Gly) belong to group S.

charges were used in DELPHI calculations of ESF, and the side chains were treated as uncharged. To obtain the ESF of individual atoms (at the peptide CO and NH groups), equation 13 of ref. 13 was used. The DSSP program (14) was used to obtain the coil library residues from the Protein Data Bank structures. Note that the coil library thus includes residues in β-turns. Selective filters were used to eliminate some residues for ESF calculations. The backbone conformation of residue 5 in the $(Ala)_9$ calculations was changed from approximately extended β (−120°, 120°) to approximately polyproline II (PPII) (−70°, 150°) in five steps of 10° in ϕ, 6° in ψ. For the $(Ala)_9$ electrostatic calculations, the DELPHI grid had 61 × 61 × 61 points, one focusing step was performed, and the final mesh size of the grid was 0.4 Å. The value of $\Delta G^*(HX)$, the blocking effect in hydrogen exchange, was calculated as $-2.303RT \log(A+B)/2$, as described by Bai and Englander (12); the values of A and B (which are negative) are given in table 2 of ref. 11.

Results

Changes in ESF at Neighboring Peptide Groups for Amino Acid Substitutions in an Alanine Host Peptide. Fig. 1 shows δESF, the change in ESF at each residue position, for a nine-residue alanine peptide (extended β conformation), resulting from the substitution Ala→Val at residue 5. There are several points of interest. First, individual values of δESF are as large as 0.8 kcal/mol. Substitution of the more bulky Val side chain for Ala reduced the −ESF values of nearby peptide groups and thus the substitution is destabilizing. Second, substantial values of δESF are found not only at the substitution site, residue 5, but also at the neighboring residues 4 and 6, and δESF is significant even at residue 3. Third, values of δESF are roughly twice as large when residue 5 is in the PPII conformation as opposed to the β conformation. Finally, the curve of δESF versus residue number is not symmetrical about the substitution site. The curve of δESF versus residue number may shift as well as changing shape when the rotamer of residue 5 varies. Thus, we use the sum of all δESF changes (ΔESF) when correlating the neighboring residue effect of X with an experimental quantity that involves averaging over different rotamers of X.

Evidently the substitution Ala→Val produces large enough δESF values to be the cause of the neighboring residue effect. Penkett *et al.* (1) reported the neighboring residue effect only for residues that precede X. Fig. 1 shows that an ESF change occurs also at the residue after X and even at residue 3, two peptide groups away from the substitution site. The plot of ESF versus residue number for the $(Ala)_9$ peptide is a horizontal line (ESF = −8.53 kcal/mol) for residues 3–7, whereas the ESF values of the two end residues 1 and 9 are more negative because of increased solvent accessibility (15).

The sum of the δESF changes at all residue positions (ΔESF) is shown in Fig. 2 for various amino acid substitutions. There are two points of interest. First, ΔESF is substantially larger when the substituted amino acid is in the PPII rather than the extended β conformation. Second, one group of amino acids (FHITVWY) shows values of ΔESF that are almost twice as large as the others (G and P excluded) when residue 5 is in the PPII conformation. These two side chain groups correspond to the class L and class S amino acids of Penkett *et al.* (1). Fig. 3 shows how the ESF values of Ala and Val at residue 5 change with ϕ between extended β (−120°, 120°) and a conformation close to PPII (−70°, 150°). Ala is more solvated than Val at all values of ϕ and, as $-\phi$ decreases, −ESF increases for $-\phi < 105°$.

The two amino acids Gly and Pro are excluded from class S for different reasons. The behavior of Gly in response to a neighboring residue is unpredictable because of the large allowed regions in the ϕ, ψ map for Gly; moreover, in Fig. 2 Gly shows a large ΔESF value with the opposite sign from other amino acids. Pro is excluded because the proline ring fixes the value of ϕ for Pro, which cannot show a substantial neighboring residue effect.

Neighboring Residue Effect in the Coil Library. Fig. 4 shows the neighboring residue effect (L versus S neighbors) in the coil library given by $\langle\phi\rangle$ values of the various amino acids within the PPII window ($-100° > \phi > -40°$, $90° < \psi < 180°$). On average, residues with L neighbors have more negative $\langle\phi\rangle$ values than residues with S neighbors by −2.2°. Only residues within the PPII window are included because, if instead the entire ϕ, ψ map is included, sparsely populated conformations with positive ϕ values are likely to give misleading results. (This problem can be

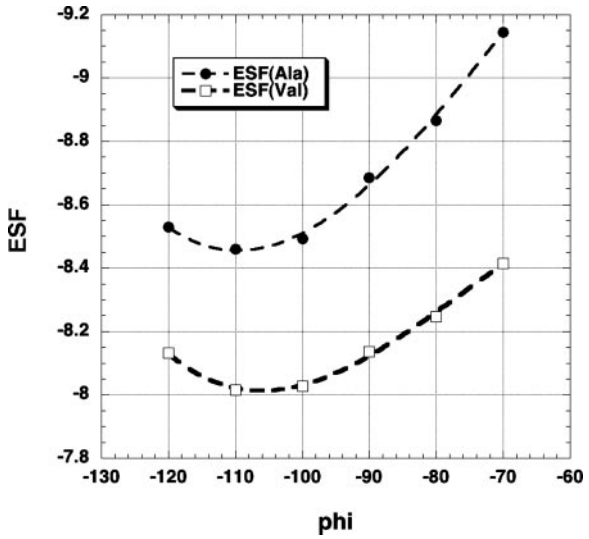

Fig. 3. Residue 5 [the substitution site in the $(Ala)_9$ peptide of Fig. 1] is made to undergo a transition from extended β ($-120°$, 120°) to approximately PPII ($-70°$, 150°) in steps of 10° in ϕ and 6° in ψ. The value of ESF (in kcal/mol) at residue 5 is plotted against ϕ for the substituted residue, Ala or Val.

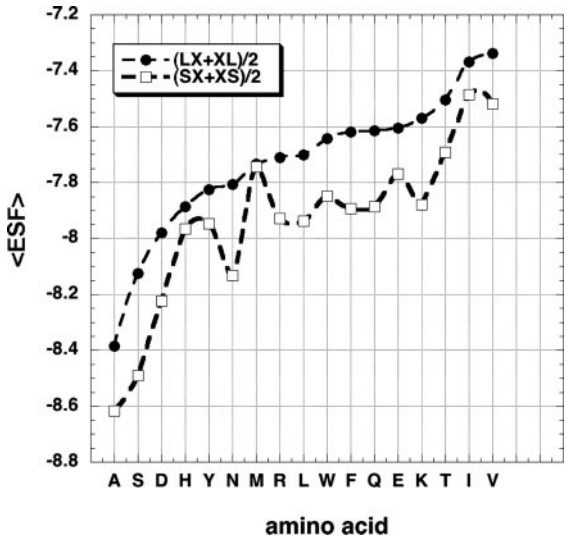

Fig. 5. The neighboring residue effect shown by ESF values in tripeptides excised from the coil library. ⟨ESF⟩ is the average ESF of the central residue (type X) when the neighbors of X are either L or S; average results for preceding and following neighbors are shown. The amino acids are ordered by the ⟨ESF⟩ values of residues with L neighbors.

minimized by comparing values of $\langle {}^3J_{HN\alpha}\rangle$ instead of $\langle\phi\rangle$, but it is of interest to determine the different $\langle\phi\rangle$ values caused by L versus S neighbors.) Gly and Pro are omitted for reasons explained above, and Cys is also omitted from coil library data because of the very small number of entries for free Cys residues. Met is an outlier in Fig. 4 in not showing $\langle\phi\rangle$ to be more negative with L than S neighbors, but the number of entries for Met is small. The neighboring residue effect in the coil library (Fig. 4) is much like the one found by Penkett *et al.* (1), who analyzed the "all" library.

The effect of L versus S neighbors on the ⟨ESF⟩ values (central residues) for the 35,310 coil library tripeptides present in the PPII window ($-100° > \phi > -40°$, $90° < \psi < 180°$) after filtering

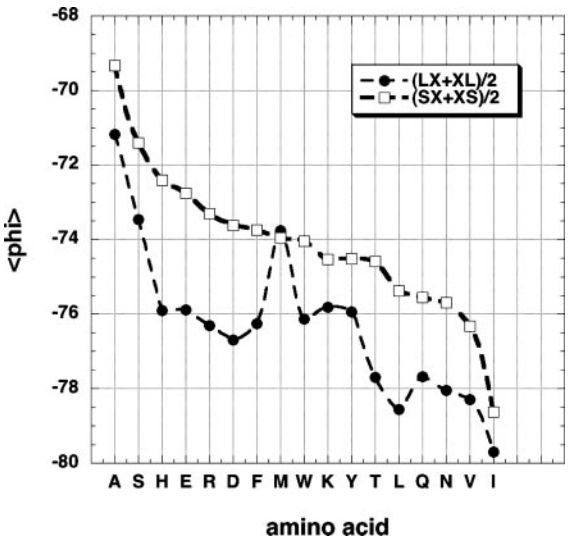

Fig. 4. The neighboring residue effect seen by plotting $\langle\phi\rangle$ for the central residues of tripeptides excised from the coil library; the neighboring residues are either L or S. Only backbone conformations contained in the PPII window ($-100° > \phi > -40°$, $90° < \psi < 180°$) are counted. The average of the values found for preceding (LX or SX) and following (XL or XS) neighbors is shown. The amino acids are ordered by the $\langle\phi\rangle$ values of residues with S neighbors.

is shown in Fig. 5. There is a consistent difference with L versus S neighbors, as expected from Figs. 1 and 2. On average, the ⟨ESF⟩ values are less negative by 0.21 kcal/mol with L than with S neighbors. In summary, when a given amino acid type in the coil library has L rather than S neighbors, it has on average a more negative value of $\langle\phi\rangle$ by 2.2° but a less negative value of ⟨ESF⟩ by 0.21 kcal/mol.

ESF Correlates with the Variation in $\langle {}^3J_{HNa}\rangle$ Among the Amino Acids. The first clear evidence for significant structure in unfolded peptides came from the variation in ϕ with amino acid type, evidenced by $^3J_{HNa}$ coupling constants in unfolded peptides and their agreement with coil library $\langle {}^3J_{HN\alpha}\rangle$ values computed from data for ϕ (3–8). The different ϕ, ψ maps of the amino acids in the coil library have been interpreted as revealing the intrinsic propensities of the amino acids in unfolded peptides to assume different backbone conformations (3–8). If, as expected from Figs. 1 and 2, the variation in ϕ with amino acid type in the coil library is caused chiefly by peptide solvation (ESF), then the coil library values of $\langle {}^3J_{HNa}\rangle$ computed from ϕ should be correlated with ΔESF values for amino acid substitutions in an $(Ala)_9$ host peptide. This correlation is shown in Fig. 6: the correlation coefficient is 0.89. Note that class L amino acids (FHITVWY) have larger values of $\langle {}^3J_{HNa}\rangle$ than class S amino acids, as expected.

Peptide Solvation (ESF) Correlates with the Blocking Effect in Hydrogen Exchange. In hydrogen exchange studies with dipeptides (11), the bulkiness of a nonpolar side chain affects the rate of hydrogen exchange of the adjacent peptide NH proton. The side chain reduces the access of solvent to the peptide group, whereas the solvent species that takes part in the exchange reaction (the OH^- ion in base-catalyzed exchange and the H^+ ion in acid-catalyzed exchange) is highly solvated. Thus, Val has a 4-fold-lower hydrogen exchange rate than Ala in both the base- and acid-catalyzed exchange reactions (11). The electron-withdrawing effect of a nonpolar side chain is negligible and the pH minimum for exchange is the same for different nonpolar side chains, because the acid- and base-catalyzed rates are affected to the same extent by the side chain. Then the blocking

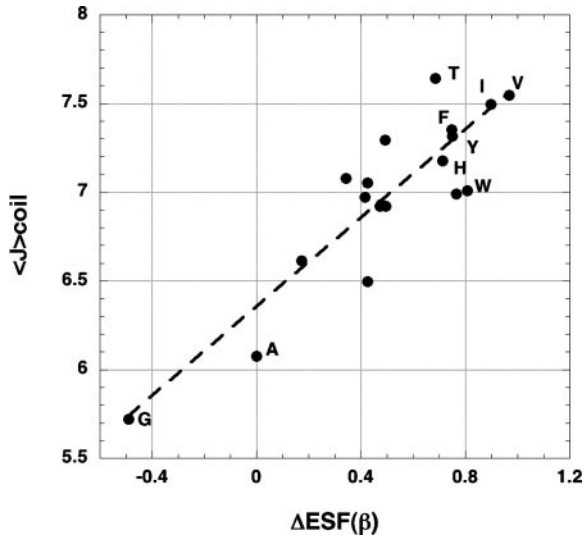

Fig. 6. The average coupling constant ($\langle {}^3J_{HN\alpha} \rangle$ in Hz) of an amino acid in the coil library plotted against calculated values of the neighboring residue effect (extended β conformation, ΔESF in kcal/mol) for the various amino acids in an $(Ala)_9$ host peptide (see Fig. 2). The class L amino acids are labeled, and also Ala and Gly for reference.

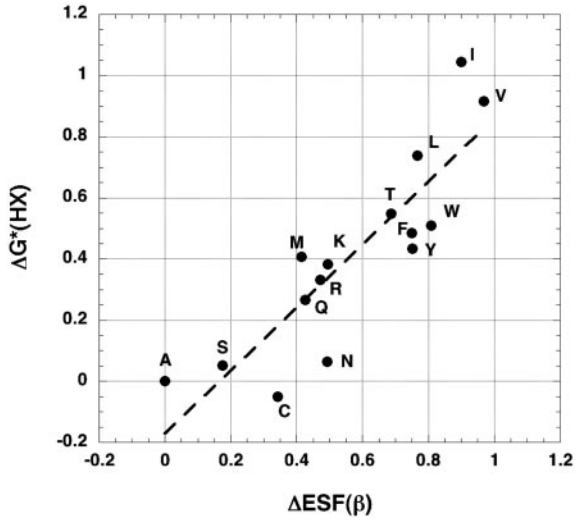

Fig. 7. The ''blocking effect'' of the side chain (11, 12) on the hydrogen exchange rate of the adjacent peptide NH proton plotted against the neighboring residue effect (extended β conformation, ΔESF in kcal/mol, Fig. 2) for that amino acid in an $(Ala)_9$ host peptide. The blocking effect is given as an apparent free energy in kcal/mol (see text).

effect is measured by the reduction in exchange rate (compared with Ala) at the pH minimum. For polar side chains, whose electron-withdrawing effects are not negligible, the pH minimum shifts and the blocking effect is measured from the reduction in exchange rate at the shifted pH minimum (11, 12).

Hydrogen exchange is considered to be a diffusion-controlled proton transfer reaction, to a close approximation (16–18), and its exchange rate (k_{HX}) is given by

$$k_{HX} = k_1[10^{\Delta pK}/(10^{\Delta pK} + 1)]. \quad [1]$$

In Eq. **1**, k_1 is the diffusion-controlled rate for proton transfer in water ($\approx 10^{10}$ $M^{-1}{\cdot}s^{-1}$), and the factor in brackets gives the fraction of productive encounters between proton donor and acceptor. Thus, the hydrogen exchange rate depends on the equilibrium pK difference (ΔpK, which is always negative) between the proton donor (water) and acceptor (peptide); $10^{\Delta pK}$ is usually $\ll 1$, so that ln k_{HX} is proportional to ΔpK. Thus, the effect of varying the nonpolar side chain at C_α on the pK of the peptide group should determine the blocking effect in hydrogen exchange. The change in ΔpK caused by an amino acid substitution is determined directly from the change in ESF of the ionized peptide species for the base- or acid-catalyzed reaction. Note that the pK values (25°C) of acetic acid (side chain CH_3^-; pK 4.756) and propionic acid (side chain $CH_3CH_2^-$; pK 4.874) (19) differ significantly. McCammon and coworkers (18) have used pK prediction to compare, with good success, predicted and observed exchange rates in dipeptides for amino acids that contain electron-withdrawing substituents.

Because the blocking effect is the same in the acid- and base-catalyzed reactions (11, 12), whereas the ionized peptide group has opposite charges in the two reactions, we used the neutral form of the peptide and compared the blocking free energy with the relative changes in ESF of the different amino acids in the $(Ala)_9$ peptide. Fig. 7 shows a good correlation between ΔESF and blocking free energy for the left peptide group in dipeptides; the correlation coefficient is 0.87 and the slope is 1.03. The slope is not predictable at present. According to ESF calculations, the exchange rate should depend on the peptide backbone conformation and also on the rotamer conformation of the amino acid, and the measured exchange rate represents an average over all conformations present. On the other hand, the ESF calculations were made for a single fixed conformation, because the actual backbone and side chain conformations are still a subject of research. The slope in Fig. 7 increases almost to 2 if ΔESF values for the PPII conformation, instead of the extended β conformation, are used. However, the slope decreases substantially if the value for δESF at residue 5 (which corresponds to the measured exchange rate) is used instead of ΔESF for the whole peptide.

Glu, Asp, and His are not included in Fig. 7 because their blocking effects are not known accurately, because their ionization states are changing at their respective pH minima for exchange. Gly is not included because the apparent blocking free energy difference between Ala and Gly is not the same in the acid- and base-catalyzed exchange reactions (11).

Bai and Englander (12) correlated the blocking free energy in hydrogen exchange with thermodynamic β structure propensities measured by Kim and Berg (20) at a solvent-exposed site within the β-hairpin of a zinc finger protein. Likewise, Avbelj and Baldwin (21) correlated Kim and Berg's β-structure propensities with ESF values of the 20 aa, using structures computed for the substitution site in the zinc finger protein. Interestingly, the correlation between β-structure propensities and blocking free energy is inverse, not direct (12), as is the correlation between β-structure propensities and ESF (21). The inverse correlation was explained by arguing that the blocking effect in hydrogen exchange (12) should be larger when the zinc finger is unfolded than when it is folded, and the same argument was made with the ESF correlation (21).

Discussion

The ESF Approach to Peptide Solvation. The landmark experiment of Wolfenden (22) showed in 1978 that amides have solvation free energies of approximately −10 kcal/mol and demonstrated that peptide solvation is an important but neglected factor in the energetics of protein folding. Makhatadze and Privalov (23) then incorporated estimates of peptide solvation into their analysis of folding energetics. Honig and coworkers (10) were able to fit the experimental solvation free energies of 67 small polar molecules with electrostatic calculations based on DELPHI, after taking into account the nonelectrostatic factors that also contribute to

solvation free energy by using a simple, approximate, experimentally based procedure.

Only recently have detailed ESF calculations been made for peptides (8, 13, 15, 21). The results emphasize that the ESF of a peptide group depends strongly on its accessibility to water. If the peptide is buried out of contact with water, its ESF falls to zero and ESF calculations cease to be informative (21). The study of unfolded peptides in water is an optimal case in which the assumptions made in calculating ESF values correspond reasonably well with reality. As discussed earlier (8), the most direct test at present of how well ESF calculations compare with reality is to compare the simulated distributions of backbone conformations in denatured proteins with the coil library distributions. The simulations involve two additional physical factors besides ESF (see below). Some success has been reported (8), but the agreement between the coil library distributions and the simulated distributions is not yet fully quantitative.

ESF calculations on peptides have been used to discuss the following topics: the nature of helix propensities (13, 15) and β-structure propensities (13, 21), the desolvation penalty for burying a H bonded peptide group (15, 21, 24), and the contribution of ESF to the enthalpy change for forming an alanine peptide helix (24). Note that for amides, the free energy and the enthalpy of interaction of the polar groups (NH and CO) with water are practically equal, according to experimental data (15). ESF calculations provide an important part of the electrostatic screening model (ESM), which has been used to predict the cooperative formation of β-strands in denatured proteins (13) and to simulate libraries of backbone conformations for the different amino acids in denatured proteins (8).

Neighboring Residue Effect Seen via ESF. The neighboring residue effect seen via ESF (Figs. 1 and 2) has correct properties to explain why $\langle\phi\rangle$ values in the coil library depend on whether the adjacent residues are class L or S. First, the size of the neighboring residue effect seen by ESF is large. For the substitution Ala→Val in an extended $(Ala)_9$ peptide, the energetic cost of the substitution is nearly 2 kcal/mol if the substituted residue is in the PPII conformation, or ≈1 kcal/mol if it is in the extended β conformation (Fig. 2). Second, when the ESF changes caused by amino acid substitutions are divided into two classes, large and small (Fig. 1), the same two classes of amino acids appear as in the neighboring residue effect studied by Penkett *et al.* (1). Third, the dipole–dipole interactions in the peptide backbone determine that ESF depends directly on ϕ, as illustrated in Fig. 3, and the dependence has the correct sign to account for the neighboring residue effect. The magnitude also appears reasonable.

Simulations of backbone conformations in denatured proteins, using the ESM, provide insight into the neighboring residue effect. In the ESM, the distribution of backbone conformations, $g(\phi)$, for ϕ values between β and PPII, depends chiefly on three factors: ESF, E_{local}, and $V(\phi)$. E_{local} is the local electrostatic potential in the peptide backbone, and $V(\phi)$ is the torsional potential about ϕ. Both $V(\phi)$ and E_{local} depend only on peptide backbone geometry, whereas ESF depends on the type and rotamer conformation of neighboring side chains (Figs. 1 and 2) as well as on peptide backbone geometry. $V(\phi)$ has its maximum (destabilizing) value at $\phi = -120°$ and consequently $V(\phi)$ acts to destabilize extended β more than PPII. On the other hand, E_{local} has its maximum (stabilizing) value at $\phi = -120°$ because the peptide dipoles are antiparallel in this conformation (25, 26), and therefore E_{local} acts to stabilize extended β more than PPII. ESF, which represents a stabilizing interaction of the peptide group with water, changes strongly with ϕ (Fig. 3), and the behavior of ESF is critical in determining $g(\phi)$ because ESF depends directly on the side chain type (Figs. 2 and 3). Although ESF is proportional to E_{local} (13, 26), the change in ESF with ϕ has the opposite sign to the change in E_{local} with ϕ (13). Thus, each of these three physical factors [ESF, E_{local}, and $V(\phi)$] has its own distinctive effect in determining $g(\phi)$, the distribution of ϕ values between extended β and PPII. There are striking differences among the coil library $g(\phi)$ distributions of the 20 aa (8), with Ala and Val representing opposite extremes: Ala is highly populated at PPII whereas Val favors extended β.

The slope (σ) of ESF versus ϕ (at each ϕ value considered) is a critical parameter determining how the backbone conformations (ϕ values) of a given amino acid are distributed between extended β and PPII. When the amino acid becomes less well solvated, the value of $-\sigma$ decreases as the $-$ESF value decreases. Thus at $\phi = -75°$ in Fig. 1, the values of σ for Ala (which is well solvated) and Val (which is less well solvated) are -0.028 and -0.017 kcal/mol per degree. For any coil library residue X, $-\sigma$ is smaller if X has L neighbors than S neighbors (data not shown). This dependence of σ on the extent of solvation is the origin of the neighboring residue effect according to the ESM. The coil library difference between $\langle$ESF$\rangle$ values of residues with L and S neighbors is -0.21 kcal/mol (Fig. 5), and the difference between $\langle\phi\rangle$ values is 2.2° (Fig. 4). The ratio is -0.095 kcal/mol per degree, which is roughly similar in magnitude to the values of σ given above.

F.A. and S. Golic-Grdadolnik (unpublished work) found that electrostatic coupling among neighboring peptide NH and CO dipoles in the backbone induces cooperative formation of β-strands in unfolded ubiquitin. The neighboring residue effect and the formation of β-strands are different manifestations of the same phenomenon: stronger electrostatic interactions when large or bulky side chains are nearby.

Implications of the ESF Neighboring Residue Effect. A basic conclusion from our work is that peptide solvation provides a connecting link between the backbone conformations of adjacent residues in unfolded peptides. Class L side chains cause substantial decreases in the $-$ESF values of neighboring peptide groups, whereas class S side chains produce smaller decreases. This behavior has the potential to be important in directing the folding process at initial stages of folding.

Erman and coworkers (27) looked for possible sequence-dependent correlations between successive pairs of ϕ, ψ angles in structures from the Protein Data Bank and used the results to test for preferences for native backbone conformations in an unfolded polypeptide. After excising tripeptides from protein structures (using the all library, not the coil library), they used the tripeptides to construct a knowledge-based potential of mean force. Then they used the rotational isomeric states model (28) to describe the unfolded polypeptide chain of the well studied small protein CI2. They compared the backbone conformations in the unfolded chain predicted by the knowledge-based potential and by a potential based on rejecting conformations showing steric clash. They concluded that the knowledge-based potential has significant success in choosing native backbone conformations for unfolded CI2, whereas the potential based on steric clash does not. Our results suggest that peptide solvation is probably a major contributor to correlations between successive pairs of ϕ,ψ angles and to the success of the knowledge-based potential in predicting native backbone conformations.

The neighboring residue effect influences the ϕ, ψ maps of the amino acids because it broadens the distribution of residue conformations in each type of backbone conformation. In the coil library $g(\phi)$ distributions, $g(\phi)$ for amino acid X in LX pairs is shifted relative to $g(\phi)$ for X in SX pairs (data not shown, but this is evident from Fig. 4), and so the sum of the two $g(\phi)$ distributions is broadened compared to either one separately. The ϕ, ψ maps, or Ramachandran maps, of the amino acids are often said to reflect their folding properties. However, the neighboring residue effect shows that the ϕ, ψ maps also depend on the properties of pairs of residues; see also ref. 27.

Related Work on the Energetic Origin of the PPII Conformation in Alanine Peptides. The recent reports by several laboratories, using various approaches, that PPII is the major backbone conformation present in alanine peptides containing seven or fewer residues (for review, see ref. 29; see also a dissenting view, ref. 30) has sparked an interest in finding the energetic origin of this preference (31–34). There is general agreement that solvation should be a key factor in determining the preference of alanine peptides for the PPII conformation, which is an open extended structure not stabilized by peptide H bonds. A key point is that the unsolvated PPII conformation in an alanine peptide is not predicted to be stable in the gas phase, although it is predicted to be stable in water (33). As discussed above, the ESM predicts that the preference of alanine for the PPII conformation arises chiefly from the electrostatic interactions between peptide NH and CO dipoles and screening of these interactions by water, which is shielded only slightly by alanine but more strongly by bulky side chains such as those of valine and isoleucine.

Monte Carlo simulations of the interaction between water and polyalanine, using the CHARM 22 force field, have been reported recently by Mezei *et al.* (34). Their simulations used explicit water molecules, and they were able to examine the H bonding interactions made between water and the peptide backbone. Overall, their results show good agreement with our results for alanine in the $(Ala)_9$ peptide. They found that the apparent interaction enthalpy between water and the alanine peptide is significantly stronger in the PPII than in the extended β conformation, and examination of the H bonds made between water and the peptide group did not provide an obvious explanation. Although the ESF values in Fig. 3 refer to the free energy of the interaction between water and the peptide group, the ESF is approximately equal to the enthalpy of the interaction, according to experimental data for amides (15).

Molecular dynamics simulations of backbone conformation for the alanine dipeptide (containing a single alanine residue with acetyl and amide blocking groups) have been made for five different force fields by Hermans and coworkers (35). The results show that the five different force fields give quite different ϕ, ψ maps, presumably because the energy differences between backbone conformations are smaller than the errors of these force fields. As a result, modifications of two of these force fields have been reported recently, made with the specific aim of obtaining agreement either with experimental data on alanine peptides (36) or with the all library ϕ, ψ maps of the different amino acids (37).

The rationale for our approach to problems of backbone conformations in unfolded peptides, where the energy differences are small, is this. By studying only the small number of physical factors necessary to reproduce the data, as compared with the large number of physical factors represented in a force field, it may be possible to determine factors such as $V(\phi)$ directly and to check the reliability of the calculations used to determine ESF and E_{local}.

We thank Robert Jernigan, Peter Luo, and George Rose for discussion and Burak Erman and coworkers for sending us their paper (27) before publication.

1. Penkett, C. J., Redfield, C., Dodd, I., Hubbard, J., McBay, D. L., Mossakowska, D. E., Smith, R. A. G., Dobson, C. M. & Smith, L. J. (1997) *J. Mol. Biol.* **274,** 152–159.
2. Karplus, M. (1959) *J. Chem. Phys.* **30,** 11–15.
3. Muñoz, V. & Serrano, L. (1994) *Proteins Struct. Funct. Genet.* **20,** 301–311.
4. Swindells, M. B., MacArthur, M. W.& Thornton, J. M. (1995) *Nat. Struct. Biol.* **2**, 596–603.
5. Serrano, L. (1995) *J. Mol. Biol.* **254,** 322–333.
6. Smith, L. J., Bolin, K. A., Schwalbe, H., MacArthur, M. W., Thornton, J. M. & Dobson, C. M. (1996) *J. Mol. Biol.* **255,** 494–506.
7. Peti, W., Hennig, M., Smith, L. J. & Schwalbe, H. (2000) *J. Am. Chem. Soc.* **122,** 12017–12018.
8. Avbelj, F. & Baldwin, R. L. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 5742–5747.
9. Butterfoss, G. L. & Hermans, J. (2003) *Protein Sci.* **12,** 2719–2731.
10. Sitkoff, D., Sharp, K. A. & Honig, B. (1994) *J. Phys. Chem.* **98,** 1978–1988.
11. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993) *Proteins Struct. Funct. Genet.* **17,** 75–86.
12. Bai, Y. & Englander, S. W. (1994) *Proteins Struct. Funct. Genet.* **18,** 262–266.
13. Avbelj, F. (2000) *J. Mol. Biol.* **300,** 1337–1361.
14. Kabsch, W. & Sander, C. (1983) *Biopolymers* **22,** 2577–2637.
15. Avbelj, F., Luo, P. & Baldwin, R. L. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 10786–10791.
16. Eigen, M., Kruse, W., Maass, G. & DeMaeyer, L. (1964) in *Progress in Reaction Kinetics*, ed. Porter, G. (Macmillan, New York), Vol. 2, pp. 285–318.
17. Englander, S. W. & Kallenbach, N. R. (1984) *Q. Rev. Biophys.* **16,** 521–655.
18. Fogolari, F., Esposito, G., Viglino, P., Briggs, J. M. & McCammon, J. A. (1998) *J. Am. Chem. Soc.* **120,** 3735–3738.
19. Edsall, J. T. & Wyman, J. (1958) *Biophysical Chemistry* (Academic, New York).
20. Kim, C. A. & Berg, J. M. (1993) *Nature* **362,** 267–270.
21. Avbelj, F. & Baldwin, R. L. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 1309–1313.
22. Wolfenden, R. (1978) *Biochemistry* **17,** 201–204.
23. Makhatadze, G. I. & Privalov, P. L. (1993) *J. Mol. Biol.* **232,** 639–659.
24. Baldwin, R. L. (2003) *J. Biol. Chem.* **278,** 17581–17588.
25. Brant, D. A. & Flory, P. J. (1965) *J. Am. Chem. Soc.* **87,** 2791–2800.
26. Avbelj, F. & Moult, J. (1995) *Biochemistry* **34,** 755–764.
27. Keskin, O., Yuret, D., Gursoy, A., Turkay, M. & Erman, B. (2004) *Proteins Struct. Funct. Bioinformat.* **55,** 992–998.
28. Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules* (Wiley, New York).
29. Shi, Z., Woody, R. W. & Kallenbach, N. R. (2002) *Adv. Protein Chem.* **62,** 163–240.
30. Vila, J. A., Baldoni, H., Ripoli, D. R., Ghosh, A. & Scheraga, H. A. (2004) *Biophys. J.* **86,** 731–742.
31. Pappu, R. V. & Rose, G. D. (2002) *Protein Sci.* **11,** 2437–2455.
32. Garcia, A. E. (2003) *Polymer* **45,** 669–676.
33. Drozdov, A. N., Grossfield, A. & Pappu, R. V. (2004) *J. Am. Chem. Soc.* **126,** 2574–2581.
34. Mezei, M., Fleming, P. J., Srinivasan, R. & Rose, G. D. (2004) *Proteins Struct. Funct. Bioinformat.* **55,** 502–507.
35. Hu, H., Elstner, M. & Hermans, J. (2003) *Proteins Struct. Funct. Genet.* **50,** 451–463.
36. Gnanakaran, S. & Garcia, A. E. (2003) *J. Phys. Chem. B* **107,** 12555–12557.
37. MacKerel, A. D., Jr., Feig, M. & Brooks, C. L., III (2004) *J. Am. Chem. Soc.* **126,** 698–699.