Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: Distributions of phi

Franc Avbelj*† and Robert L. Baldwin†‡

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

Contributed by Robert L. Baldwin, March 14, 2003

The ''coil library,'' consisting of the ϕ, ψ values of residues outside secondary structure in high-resolution protein structures, has chiefly the β, α_R, α_L, and polyproline II backbone conformations. In denatured proteins, the 20 aa have different average values of the $^3J_{HN\alpha}$ coupling constant, related to the backbone angle ϕ by the Karplus relation. Average $^3J_{HN\alpha}$ values obtained from the distributions of ϕ, $g(\phi)$, of the coil library agree with NMR results, and so the coil library can be and is being used to model denatured proteins. Here, Monte Carlo simulations of backbone conformations in denatured proteins are used to test two physics-based models: the random coil model of Brant and Flory [(1965) *J. Am. Chem. Soc.* 87, 2788–2791 and 2791–2800] and an electrostatic screening model (ESM) that includes electrostatic solvation. The random coil model represents hindered rotation about ϕ and ψ backbone angles, nonbonded interactions, and dipole–dipole interactions. In the ESM, the nonbonded interactions term is replaced by the use of hard sphere repulsion and allowed regions in the Ramachadran maps. These models were tested by using the amino acid sequences of three small proteins. There are two main conclusions: (*i*) The $g(\phi)$ distributions of the coil library contain detailed, specific information, so that prediction of the $g(\phi)$ distributions of the different amino acids is a demanding test of the energy function. (*ii*) The ESM is partly successful in predicting the $g(\phi)$ distributions. Electrostatic solvation is primarily responsible, and steric clash between pairs of atoms connected by torsion angles is not responsible.

Our aim is to find an energy function that reproduces the backbone conformations of amino acids in denatured proteins when the conformations are simulated by the Monte Carlo method. Recent evidence indicates that the type of amino acid specifies its preference for certain backbone conformations [β, polyproline II (P_{II}), α_L, and α_R] in denatured proteins. Average values of the $^3J_{HN\alpha}$ coupling constant differ substantially among the amino acids in denatured proteins and short peptides (1, 2), and the $^3J_{HN\alpha}$ coupling constant is directly related to the backbone angle ϕ by the Karplus relation (3). The coil library (residues outside regular secondary structures in high-resolution protein structures) is being used as a provisional database for backbone conformations of denatured proteins (1, 2, 4). We seek to predict the distribution of ϕ, $g(\phi)$, for the various amino acids in the coil library. This requires understanding the energetics of interconversion between backbone conformations as a function of ϕ. To simplify the problem, we limit consideration here to the interconversion between β and P_{II} by restricting ϕ and ψ so as to exclude α_R and α_L, by taking ($90 < \psi < 180$) and ($0 > \phi > -180$).

We test two models. The first is the random coil model (RCM) introduced by Flory and coworkers (5–8), who argued that the backbone conformations and chain dimensions of denatured proteins can be predicted by specifying a small set of energy terms, an approach used successfully with synthetic flexible polymers. The second model is an electrostatic screening model (ESM), based on the proposal by Avbelj and Moult (9) that peptide backbone conformations should strongly depend on the role of side chains in determining the extent of screening by water of the peptide dipole–dipole interactions.

The modern approach to understanding protein conformation is to make all-atom simulations of conformation using a comprehensive molecular force field. However, this approach has not been successful in modeling backbone conformations of denatured proteins, apparently because the energy differences between backbone conformations are smaller than the overall accuracy of present force fields (10, 11). It seems reasonable then to return to the simpler approach of predicting a small set of terms that should control backbone conformation in flexible chain polypeptides that are not close-packed, unlike native proteins.

The RCM of Flory and coworkers (5-8) was developed originally for use with simple peptides such as Ala–Gly copolymers. Its energy function contains terms for Coulombic interaction between partial charges on the peptide NH and CO dipoles, nonbonded interactions between pairs of atoms, and hindered rotation about the backbone ϕ and ψ angles. Flory and coworkers considered solvation of the peptide backbone only by allowing uniform screening by solvent of the peptide dipole–dipole interactions, through using a dielectric constant of 3.5. The ESM differs importantly from the RCM by including the interaction between solvent and the peptide dipoles, calculated as the electrostatic solvation free energy (ESF), using the DELPHI algorithm and the PARSE parameter set (12). The PARSE parameters (partial charges and atomic radii) have been calibrated against a database of experimental solvation free energies for model compounds including amides (12), so that the ESF values of model compounds are calibrated against experiment. Calculations of ESF values show that they depend primarily on the arrangement of the partial charges and their access to solvent, which in a peptide or protein system strongly depends on nearby side chains (13–16).

Flory and coworkers (5–8) considered that the nonbonded interactions are an important factor determining backbone conformations in denatured proteins, and some workers agree with this view (17–19). To test whether the nonbonded interactions term can be responsible for the variation in $g(\phi)$ seen in the coil library, we replaced the nonbonded interaction term in the ESM with hard sphere repulsion without, however, checking pairs of atoms connected by torsion angles for steric clash. We also required that ϕ, ψ values fall within the allowed regions of the Ramachandran maps. Then we tested whether some factor other than ESF can be responsible for the partial success of the ESM by setting the ESF term = 0 in the ESM energy function and made related tests.

Abbreviations: RCM, random coil model; ESM, electrostatic screening model; ESF, electrostatic solvation free energy; P_{II}, polyproline II.

†To whom correspondence should be addressed. E-mail: francl@sg3.ki.si or bbaldwin@cmgm.stanford.edu.

 www.pnas.org/cgi/doi/10.1073/pnas.1031522100

These tests were made by using the amino acid sequences of three small proteins: protein G, ubiquitin, and barnase.

Methods

Representation of Protein Molecules. Protein molecules are represented by including all heavy atoms and polar hydrogens. In the RCM, the residue geometry and other properties of the protein molecule are as described by Brant *et al.* (8). In the ESM, the geometry of the amino acids is generated by using the Discover residue library (20). The torsion angle ω of the peptide bond is held fixed at 180° and planar groups are kept exactly planar. In the ESM hard sphere repulsion is enforced by discarding conformations with steric clash when any pair of nonbonded atoms is closer than 0.7 Å less than the sum of the van der Waals radii given by Chothia (21), but pairs of atoms connected by torsion angles are not checked for steric clash. Steric clash occurs when main-chain N and O atoms are closer than 2.5 Å. The offset of 0.7 Å was chosen to ensure that few steric clashes are found in high-resolution protein structures and that the complete allowed regions of the Ramachandran maps are found when the conformations of denatured proteins are simulated. The Monte Carlo simulations of the ESM do not depend on the value of the offset if it is >0.4 Å.

All residues, both inside and outside secondary structures of 1,476 high-resolution protein structures, are included when determining the allowed regions used with the ESM. The bin size for ϕ and ψ is 5° and a bin is considered to be within the allowed region if it contains at least 10 residues from a total of 9,000 residues of the same type. In the Monte Carlo simulations each allowed bin has the same probability of being drawn.

The coil library of backbone conformations is constructed from all pairs of ϕ, ψ angles for residues outside secondary structure in 1,476 high-resolution protein structures.

Energy Function of the ESM. The energy function of the protein molecule in the ESM is a sum of contributions expressed by Eq. **1**.

$$E = \sum_i E^i_{local} + \sum_i ESF^i + \sum_i V^i_\phi + \sum_i V^i_\psi \qquad [1]$$

in which each sum runs over all amino acids i in the sequence. In addition, the chain must be free of steric clashes, as described above.

The term E^i_{local} is the local main-chain electrostatic energy of residue i, computed as described by Avbelj and Moult (9). It depends on the interaction of main-chain NH and CO groups in residue i with the CO and NH groups of residues $i + 1$, $i - 1$, $i + 2$, and $i - 2$. E^i_{local} is computed by 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 electrons, respectively (14).

The quantity ESF^i is the backbone ESF of amino acid residue i (14). It is calculated by the finite difference Poisson–Boltzmann method (12, 14) by using the PARSE-neutral parameter set (14) in which the PARSE point atomic charges of side-chain atoms (12) are set to zero. Atomic point charges in all side chains, as well as formal charges in ionizing side chains, are set to zero in this article. The electrostatic potential of the whole molecule is first calculated by using a very large box and large grid size (scale = 0.20, perfil = 65). This potential then provides boundary conditions for much more accurate calculations of electrostatic potential around each residue (focusing). Equal box and grid sizes (scale = 2.0, grid size 51, center at C_α) are then used for all residues to ensure approximately equal amounts of water near each residue.

The quantities V^i_ψ and V^i_ϕ are the intrinsic torsional potentials around the angles ψ and ϕ

$$V^i_\psi = (V^o_\psi/2)[1 + \cos 3(\psi^i - \pi/3 + \pi)]$$
$$V^i_\phi = (V^o_\phi/2)[1 + \cos 3(\phi^i - \pi/3 + \pi)]. \qquad [2]$$

In the ESM the constant V^o_ϕ is set to 2.5 kcal/mol while the constant V^o_ψ is set to zero.

Finding the correct balance between these terms is critical for obtaining correct results. Earlier studies (12–14) have given estimates of both ESF and E_{local} and these terms are used without change; only the value of V^o_ϕ has been adjusted here.

Energy Function of the RCM. The energy function of the RCM is taken from the work of Brant and Flory (6, 7) and Brant *et al.* (8). Some other representations of the nonbonded interaction terms were also tested. The energy function of the RCM of Brant *et al.* (8) is a sum of contributions expressed by Eq. **3**.

$$E = \sum_i V^i_\phi + \sum_i V^i_\psi + \sum_i \sum_{jk} [V^i_{r,jk} + V^i_{l,jk} + V^i_{c,jk}]. \qquad [3]$$

The intrinsic torsional potentials V^l_ψ and V^i_ϕ are as in Eq. **2** with $V^o_\psi = 1.0$ kcal/mol and $V^o_\phi = 1.5$ kcal/mol (8). The repulsive term of the nonbonded interactions is expressed by

$$V_{r,jk} = a_{jk}/r^{12}_{jk} \qquad [4a]$$

and the attractive (London) term is expressed by

$$V_{l,jk} = -c_{jk}/r^6_{jk}\,. \qquad [4b]$$

The coefficients a_{jk} and c_{jk} are found from the Slater–Kirkwood equation as described (8). The Coulombic term, which gives the interaction between monopoles δ_j and δ_k, is

$$V_{c,jk} = 332\ \delta_j\delta_k/\varepsilon r_{jk} \qquad [4c]$$

in which ε, the effective dielectric constant, has the value 3.5. The partial charges differ only slightly from those in the ESM: 0.281 (RCM) versus 0.28 (ESM) and 0.394 (RCM) versus 0.38 (ESM). The attractive part of the nonbonded interaction is set to zero for atoms more than two residues apart, to account for screening of the Coulomb interactions in aqueous solution. The bond distances and angles are those of Brant and Flory (6, 7).

Monte Carlo Simulation Procedure. The starting geometry in all simulations is a random conformation, achieved as follows. The backbone torsion angles are randomly selected from the allowed regions in the Ramachandran plots, taken from both inside and outside secondary structures. In the ESM the side-chain torsion angles are randomly selected from the entire library of side-chain torsion angles (both inside and outside secondary structure) whereas in the RCM all side-chain torsion angles are drawn with equal probability. The initial conformation is then relaxed by small variations in torsion angles to remove steric overlaps between pairs of atoms. To improve sampling of conformational space, a large number (≈100) of independent Monte Carlo simulations are performed, each with 5,000 steps. The initial 1,500 steps are needed to achieve clash-free equilibration. Only torsion angles are allowed to vary during simulations, but both side-chain and backbone torsion angles vary.

Conformational space is sampled by varying the torsion angles by using two types of moves, in which either main-chain or side-chain torsion angles are selected. A residue is selected randomly. The probabilities of selecting either main-chain or side-chain torsion angles are set to 60% and 40%, respectively. In the ESM torsion angles ϕ and ψ are always selected as pairs from the coil library and are required to lie within the allowed regions of the Ramachandran maps of the various residues (see above). Likewise, side-chain torsion angles are selected at ran-

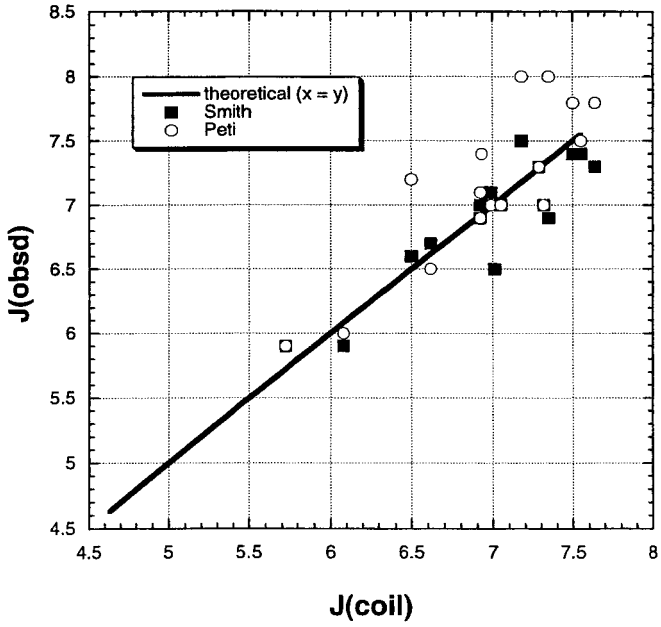

Fig. 1. Measured $^3J_{HN\alpha}$ coupling constants for the various amino acid residues plotted against predicted values from the structural coil library. The values from the coil library are based on ϕ values in protein structures and the Karplus relation between ϕ and $^3J_{HN\alpha}$ (see *Methods*). The two sets of experimentally measured coupling constants are given for amino acid residues in unstructured peptides (2) and denatured ubiquitin (29). The complete coil library is used with $-180 < \psi < 180$ and $-180 < \phi < 180$. The average values of $^3J_{HN\alpha}$ in the coil library, which lie on the line, are (in Hz): G, 5.720; A, 6.075; E, 6.495; S, 6.615; R, 6.920; K, 6.920; D, 6.931; L, 6.991; W, 7.007; H, 7.177; N, 7.294; Y, 7.318; F, 7.351; I, 7.495; V, 7.547; T, 7.642.

dom from the entire library. The resulting conformation is tested for steric clashes and discarded if any steric clash is found.

The energy of a conformation is computed from Eq. **1** (ESM) or Eq. **3** (RCM). If the energy decreases, the new conformation is accepted. If the energy increases, the Metropolis criterion (22) is used to decide whether to accept or reject the move. The temperature is 300 K. Overall, two-thirds of the moves are accepted during simulations. Average values are computed from the large number of conformations ($>10^5$) present after equilibration.

$^3J_{HN\alpha}$ Coupling Constants. The $^3J_{HN\alpha}$ coupling constants of residues are calculated from their ϕ values by the Karplus equation (3).

$$^3J_{HN\alpha} = a\cos^2(\phi - 60) - b\cos(\phi - 60) + c \qquad [5]$$

(ϕ in degrees). Four somewhat different parameterizations of Eq. **5** have been given in the literature, with the following parameters: $a = 6.4$, $b = -1.4$, $c = 1.9$ (23); $a = 6.6$, $b = -1.3$, $c = 1.5$ (24); $a = 6.51$, $b = -1.76$, $c = 1.60$ (25); and $a = 7.90$, $b = -1.05$, $c = 0.65$ (26). Values of $^3J_{HNa}$ for a given residue were computed as the average of the values obtained with Eq. **5** by using the four different sets of parameters. Further averaging, to obtain $\langle^3J_{HN\alpha}\rangle$, was then done over all residues of a given type. Note that Eq. **5** gives a value of $J_{HN\alpha}$ near 5 for ϕ near $-70°$, and a value close to 10 for ϕ near $-120°$.

Fitting $g(\phi)$ to Two Gaussians. The procedure used to fit $g(\phi)$ to a sum of two Gaussian functions is based on a standard Levenberg–Marquardt fitting algorithm as described (27).

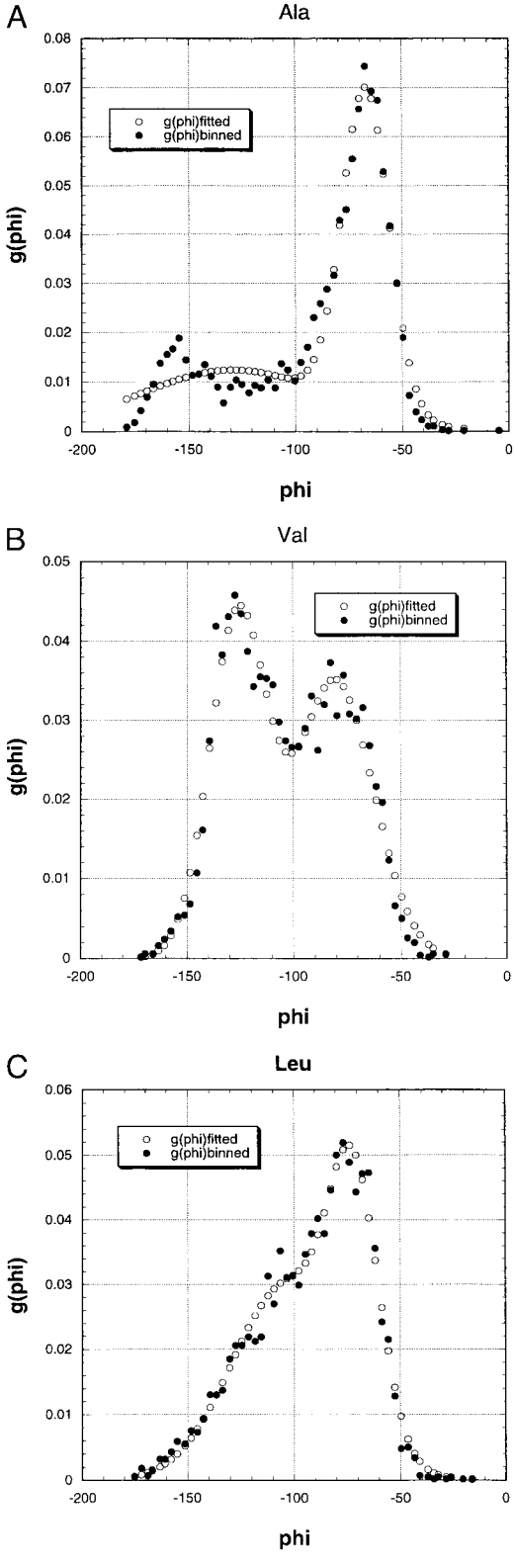

Fig. 2. Three examples of fitting coil library $g(\phi)$ distributions (with $90 < \psi < 180$ and $-180 < \phi < 0$) to a sum of two Gaussians. (*A*) Alanine. (*B*) Valine. (*C*) Leucine. The bin size is 3° and $g(\phi)$ is not normalized.

Results

$^3J_{HN\alpha}$ Coupling Constants from the Coil Library Compared with Experimental Values. The number of high-resolution structures in the Protein Data Bank is large and increasing rapidly. As a result, the distributions of ϕ, ψ backbone angles in the coil library are very reliable. Previous authors found a good correlation between coupling constants measured by NMR and ones computed from

ϕ values of the coil library (1, 2, 28, 29). Fig. 1 shows new data from the coil library testing this previous result. The predicted coupling constants from the coil library (see *Methods*) are compared with two sets of experimental results, one for unstructured peptides (2) and one for denatured ubiquitin (29). The coupling constants predicted from the coil library agree with both sets of experimental data better than the two experimental data sets agree with each other. The lackluster agreement between the two experimental data sets may reflect different neighboring residue effects, which average out in the larger data set from the coil library.

Comparison Between $\langle\phi\rangle$ Values from the Coil Library and from Simulations of the ESM and RCM. The average value of ϕ ($\langle\phi\rangle$) is a useful parameter to use in comparing $g(\phi)$ distributions for different amino acids and their simulation by different models. The $\langle\phi\rangle$ values given by the RCM fall in a high range ($-106.2°$ to $-116.5°$) compared with the coil library values ($-93.3°$ to $-107.2°$), whereas the ESM values ($-95.1°$ to $-109.8°$) are closer to the coil library values. Values for Pro and Gly are excluded. Only the P_{II} and β backbone conformations are analyzed by restricting ϕ and ψ to exclude the α_L and α_R conformations: $90 < \psi < 180$ and $-180 < \phi < 0$. (In the $\langle J\rangle$ values for the coil library given in the legend to Fig. 1, ϕ, ψ restriction is not applied because the coupling constants measured by NMR are averaged over all ϕ and ψ values.)

Fitting the $g(\phi)$ Distributions to a Sum of Two Gaussians. Because of scatter in the binned distribution data for $g(\phi)$, the $g(\phi)$ distributions (proline excluded) are represented here by lines fitted to a sum of two Gaussians. One Gaussian is found to be centered on a band termed P_{II} because the maximum value of $g(\phi)$ for all amino acids occurs close to the canonical P_{II} band position, and the ψ value at this position also matches that of P_{II}. The second Gaussian fits a β-region band whose size and position vary widely among the amino acid types. The quality of the fitting is illustrated in Fig. 2 for three examples taken from coil library $g(\phi)$ distributions. For all amino acids, the fit of the binned distribution data to a Gaussian for the P_{II} band lies almost within the scatter of the data. For Ala (Fig. 2*A*), the overall $g(\phi)$ has two or more minor bands in the β region in addition to the major P_{II} band, and consequently a fit to only two Gaussians fails to work well for Ala in the β region. Most fittings of $g(\phi)$ to a sum of two Gaussians work well in both the β and P_{II} regions and resemble those for Val and Leu (Fig. 2 *B* and *C*) in giving a fit that lies within the scatter of the binned $g(\phi)$ data. Fitting $g(\phi)$ to a sum of two Gaussians works poorly for the variant ESM with ESF = 0 and E_{local} scaled down by $D = 3.5$, and the two-Gaussian fits were not used.

Properties of the $g(\phi)$ Distributions in the Coil Library and the ESM and RCM Simulations. Differences among the $g(\phi)$ distributions of the various amino acids in the coil library are illustrated in Fig. 2 for Ala, Val, and Leu. The differences reside chiefly in the size of the P_{II} band and in both the size and position of the β-region band. The position of the P_{II} band is relatively constant (see below). The $g(\phi)$ distribution of alanine is unusual in showing more than one band in the β region. An important conclusion from comparing the $g(\phi)$ distributions of the amino acids in the coil library is that the differences in $g(\phi)$ between amino acids are large enough and specific enough to provide a demanding test of any energetic model.

When the $g(\phi)$ distributions of Leu, Phe, and Ile are compared by overlaying the entire distribution curves (Fig. 3), the results show that $g(\phi)$ at $-125°$ is a useful parameter for comparing the $g(\phi)$ distributions. This is evident in the coil library results (Fig. 3*A*) and the ESM results (Fig. 3*B*), but there is little variation in the RCM results (Fig. 3*C*). Results for all amino acids at $-125°$

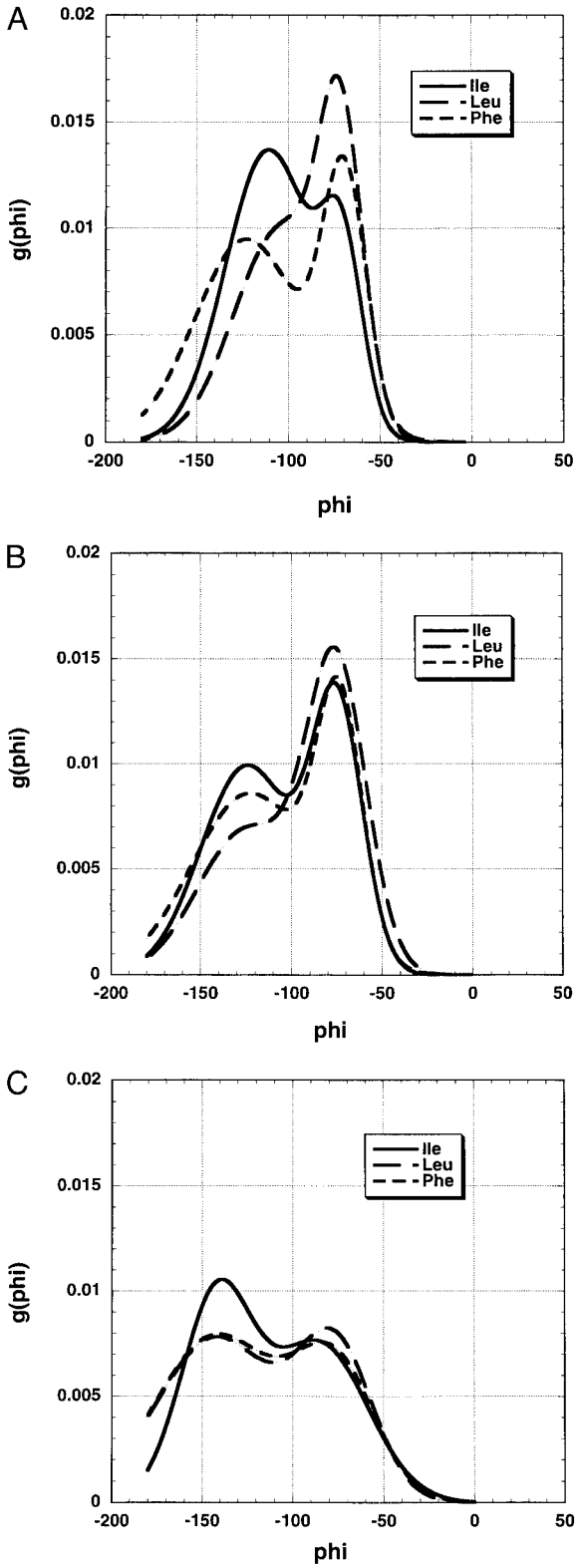

Fig. 3. Comparison of $g(\phi)$ distributions, given by fitting binned data to a sum of two Gaussians, for three amino acids, Leu, Phe, and Ile, from the coil library (*A*), ESM simulations (*B*), and RCM simulations (*C*). The backbone conformations are restricted to P_{II} and β by setting $90 < \psi < 180$ and $-180 < \phi < 0$. The bin size is 3°. Average values of ψ were determined for the bins at the maximum values of $g(\phi)$ in both the P_{II} and β–region bands in Fig. 4. They correspond to typical values of ψ for these backbone conformations (data not shown).

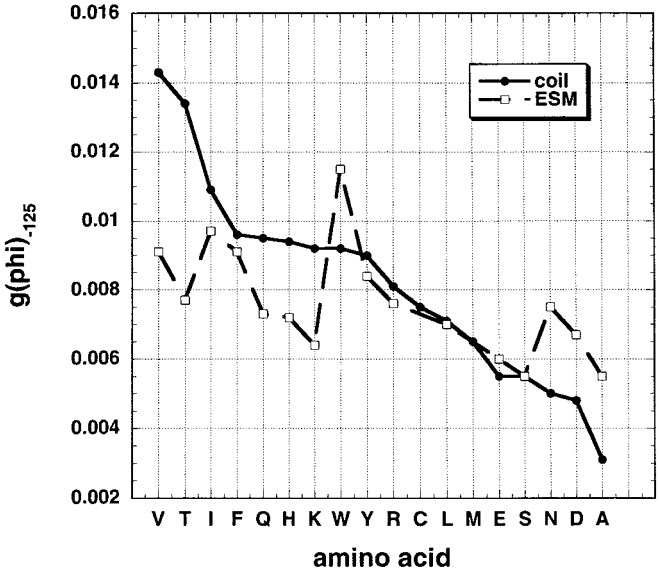

Fig. 4. Plot of $g(\phi)$ at $-125°$ versus amino acid type for the coil library and the ESM. Gly and Pro are excluded because of their unusual distributions of ϕ and ψ, whereas Cys and Met are not present in the three proteins whose conformations were simulated by the ESM. Note that the variation in $g(\phi)$ at $-125°$ is similar in character in the coil library and the ESM simulations.

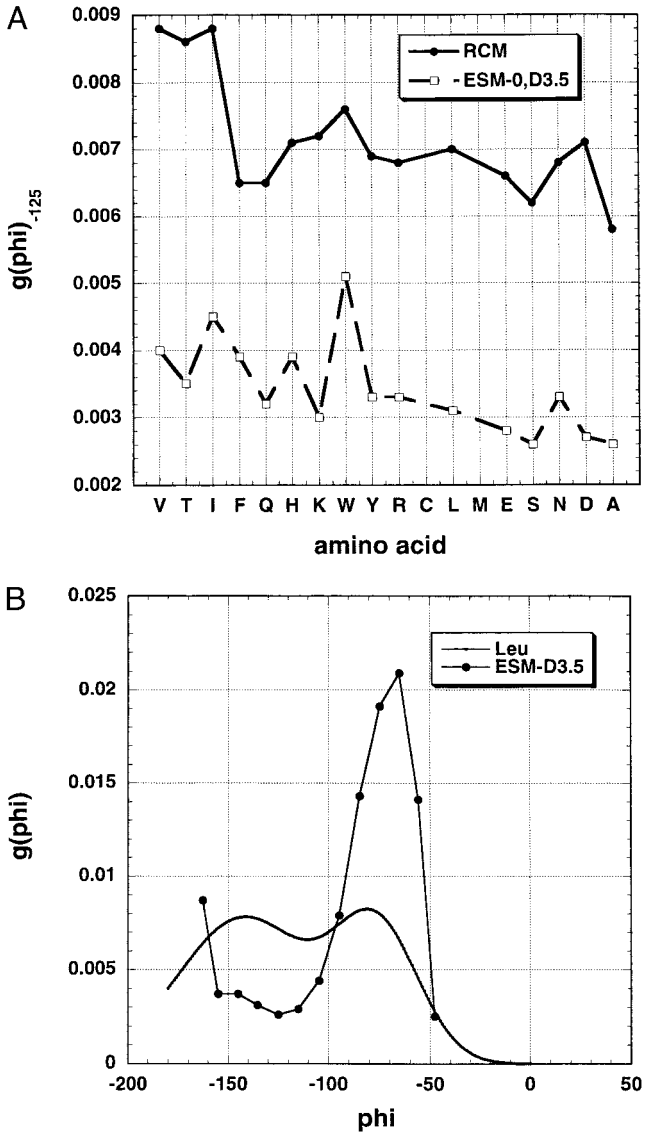

Fig. 5. (*A*) Plot of $g(\phi)$ at $-125°$ versus amino acid type for the RCM and the variant ESM-D3.5. In the variant ESM, ESF is set = 0 and E_{local} is scaled down by using a dielectric constant of 3.5. Note the reduced variation with amino acid type shown by these two models as compared with the standard ESM (Fig. 4). The scale is the same as in Fig. 4. (*B*) Comparison of the $g(\phi)$ distributions for leucine given by the RCM and the variant ESM-D3.5.

are shown in Fig. 4. (Gly, Pro, Cys, and Met are excluded because Pro, Cys, and Met are missing, whereas Gly and Pro have anomalous ϕ, ψ distributions.) The variation among amino acids in $g(\phi)$ at $-125°$ is strong in the coil library, moderate in the ESM simulations, and quite small in the RCM simulations (see below and Fig. 5*A*). Importantly, the amino acid dependence of $g(\phi)$ at $-125°$ is similar for the ESM simulations and the coil library.

The values of ϕ for the P_{II} band in the coil library (Gly and Pro excluded), taken from the band centers of the fitted Gaussian curves, lie chiefly in the range between $-67.2°$ and $-73.9°$, with Val, Thr, and Asn showing more negative values. In the simulations for the RCM of Brant and Flory (8) (Gly, Pro, Cys, and Met excluded), values of ϕ for the P_{II} band lie in a more negative range between $-75.5°$ and $-88.0°$ with the β–branched amino acids Val, Ile, and Thr showing the highest negative values. In the ESM simulations (Gly, Pro, Cys, and Met excluded), values of ϕ for the P_{II} band are intermediate between those of the coil library and the RCM and lie between $-69.9°$ and $-77.0°$ for all amino acids except Arg with $-79.4°$.

Varying Parameters in the RCM and ESM. To test whether inclusion of ESF in the ESM is responsible for the variation in $g(\phi)$, a variant ESM (ESM-D3.5) was tested in which ESF = 0 and E_{local} is reduced by setting the dielectric constant = 3.5, as in the RCM of Brant and Flory. The variant ESM-D3.5 shows considerably smaller variation in $g(\phi)$ among the amino acids than does the ESM. This point is illustrated in Fig. 5*A*, which shows the values of $g(\phi)$ at $-125°$ for the different amino acids for both the variant ESM and the RCM. Both models show little variation in $g(\phi)$ among the different amino acids. Fig. 3*C* further illustrates the small extent of variation in $g(\phi)$ seen for the RCM. The values of $g(\phi)$ at $-125°$ for the different amino acids are higher (i.e., more negative) for the RCM than for the variant ESM-D3.5. Fig. 5*B* shows the $g(\phi)$ distributions for leucine given by the two models and illustrates that the $g(\phi)$ distributions have characteristically different shapes for the two models.

Similar results for the RCM simulations of $g(\phi)$ were obtained when the nonbonded interaction terms were taken either from AMBER (30) or CHARMM (31) instead of from Brant *et al.* (8) (data not shown).

Discussion

Use of the Coil Library. We compare the backbone conformations simulated for the RCM and ESM with the coil library rather than with experimentally measured coupling constants because the differences between various experimental sets of coupling constant data are too large (see Fig. 1). Because of scatter in the experimental data as well as differences between the two data sets, the coupling constants predicted for the RCM and ESM do not distinguish clearly between them (data not shown). On the other hand, using the $g(\phi)$ distributions from the coil library as a test of the RCM and ESM easily distinguishes between them. Future work may show significant differences between the coil library and backbone conformations in denatured proteins, because the coil library is taken from close-packed (but typically

solvent-exposed) segments of structure in native proteins. Nevertheless, comparison between predicted ϕ, ψ distributions and the coil library appears to be the best available option at present.

Partial Success of the ESM. The rationale for the approach used here to predict backbone conformations is that the energy differences between backbone conformations of denatured proteins are probably small. Therefore it is desirable to calculate the relevant energy terms directly and to test them one term at a time. This approach produces partial success here: the ESM produces limited variation in $g(\phi)$ of the same type seen in the coil library. This limited success is primarily produced by the ESF term in the energy function of the ESM because, when ESF is set = 0, only a small amount of variation in $g(\phi)$ is still seen (Fig. 5*A*). It may be caused by use of the allowed regions in the Ramachandran maps of the amino acids.

It is a challenge now to learn whether force field simulations, based on using a comprehensive force field, can achieve better agreement than the simulations of the ESM. Our results indicate that, to make a force field succeed with denatured proteins, it should reproduce accurately the solvation energies of amides, and some force fields fail to do this (32). The OPLS–AA force field has been adjusted to take account of this problem (33), and this may be a good starting point.

Nature of the Equilibrium Between P_{II} and β. The $g(\phi)$ distributions of the coil library indicate that the equilibrium between P_{II} and β is different for every amino acid type in denatured proteins. The P_{II} to β interconversion is not a simple two-state equilibrium because the position of the β-region band varies substantially among the amino acid types, and alanine shows more than one β-region band. The energetics governing the distribution of P_{II} and β can be analyzed by considering the individual energy terms in the ESM. The ESF and E_{local} differences between $\phi = -70°$ (characteristic of P_{II}) and $\phi = -120°$ (characteristic of β) have been calculated for the various amino acids in a simple model system (unpublished data) and will be discussed in a later publication.

We thank Jan Hermans for discussion. This work was supported by the Ministry of Education, Science, and Sport of Slovenia.

1. Serrano, L. (1995) *J. Mol. Biol.* **254,** 322–333.
2. Smith, L. J., Fiebig, K. M., Schwalbe, H., MacArthur, M. W., Thornton, J. M. & Dobson, C. M. (1996) *J. Mol. Biol.* **255,** 494–506.
3. Karplus, M. (1959) *J. Chem. Phys.* **30,** 11–15.
4. Swindells, M. B., MacArthur, M. W. & Thornton, J. M. (1995) *Nat. Struct. Biol.* **2,** 596–603.
5. Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules* (Wiley, New York), pp. 30-31 and 255–274.
6. Brant, D. A. & Flory, P. J. (1965) *J. Am. Chem. Soc.* **87,** 2788–2791.
7. Brant, D. A. & Flory, P. J. (1965) *J. Am. Chem. Soc.* **87,** 2791–2800.
8. Brant, D. A., Miller, W. G. & Flory, P. J. (1967) *J. Mol. Biol.* **23,** 47–65.
9. Avbelj, F. & Moult, J. (1995) *Biochemistry* **34,** 755–764.
10. O'Connell, T. M., Wang, L., Tropsha, A. & Hermans, J. (1999) *Proteins* **36,** 407–418.
11. Hu, H., Elstner, M. & Hermans, J. (2003) *Proteins* **50,** 451–463.
12. Sitkoff, D., Sharp, K. A. & Honig, B. (1994) *J. Phys. Chem.* **98,** 1978–1988.
13. Avbelj, F. & Fele, L. (1998) *J. Mol. Biol.* **279,** 665–684.
14. Avbelj, F. (2000) *J. Mol. Biol.* **300,** 1337–1361.
15. Avbelj, F., Luo, P. & Baldwin, R. L. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 10786–10791.
16. Avbelj, F. & Baldwin, R. L. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 1309–1313.
17. Pappu, R. V., Srinivasan, R. & Rose, G. D. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 12565–12570.
18. Pappu, R. V. & Rose, G. D. (2002) *Protein Sci.* **11,** 2437–2455.
19. Ackerman, M. S. & Shortle, D. (2002) *Biochemistry* **41,** 13791–13797.
20. Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, D. J., Genest, M. & Hagler, A. T. (1988) *Proteins* **4,** 31–47.
21. Chothia, C. (1976) *J. Mol. Biol.* **105,** 1–14.
22. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, M. N. & Teller, E. (1952) *J. Chem. Phys.* **21,** 1087–1092.
23. Pardi, A., Billeter, M. & Wüthrich, K. (1984) *J. Mol. Biol.* **180,** 741–751.
24. Ludvigsen, S., Andersen, K. V. & Poulsen, F. M. (1991) *J. Mol. Biol.* **217,** 731–736.
25. Vuister, G. W. & Bax, A. (1998) *J. Am. Chem. Soc.* **115,** 7772–7777.
26. Schmidt, J. M., Blumel, M., Lohr, F. & Rüterjans, H. (1999) *J. Biomol. NMR* **14,** 1–12.
27. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992) *Numerical Recipes in Fortran 77: The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, U.K.), 2nd Ed.
28. 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.
29. Peti, W., Hennig, M. Smith, L. J. & Schwalbe, H. (2000) *J. Am. Chem. Soc.* **122,** 12017–12018.
30. Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986) *J. Comp. Chem.* **7,** 230–252.
31. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comp. Chem.* **4,** 187–217.
32. Marten, B., Kim, K., Cortis, C., Friesner, R. A., Murphy, R. B., Ringnalda, M. N., Sitkoff, D. & Honig, B. (1996) *J. Phys. Chem.* **100,** 11775–11788.
33. Rizzo, R. C. & Jorgensen, W. L. (1999) *J. Am. Chem. Soc.* **121,**4827–4836.