Short-Range Conformational Energies, Secondary Structure Propensities, and Recognition of Correct Sequence-Structure Matches

I. Bahar,[1,2] M. Kaplan,[2] and R.L. Jernigan[1]*
[1]*Molecular Structure Section, Laboratory of Mathematical Biology, Division of Basic Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland*
[2]*Chemical Engineering Department and Polymer Research Center, Bogazici University, and TUBITAK Advanced Polymeric Materials Research Center, Bebek 80815, Istanbul, Turkey*

ABSTRACT **A statistical analysis of known structures is made for an assessment of the utility of short-range energy considerations. For each type of amino acid, the potentials governing (1) the torsions and bond angle changes of virtual C^{α}-C^{α} bonds and (2) the coupling between torsion and bond angle changes are derived. These contribute approximately −2 RT per residue to the stability of native proteins, approximately half of which is due to coupling effects. The torsional potentials for the α-helical states of different residues are verified to be strongly correlated with the free-energy change measurements made upon single-site mutations at solvent-exposed regions. Likewise, a satisfactory correlation is shown between the β-sheet potentials of different amino acids and the scales from free-energy measurements, despite the role of tertiary context in stabilizing β-sheets. Furthermore, there is excellent agreement between our residue-specific potentials for α-helical state and other thermodynamic based scales. Threading experiments performed by using an inverse folding protocol show that 50 of 62 test structures correctly recognize their native sequence on the basis of short-range potentials. The performance is improved to 55, upon simultaneous consideration of short-range potentials and the nonbonded interaction potentials between sequentially distant residues. Interactions between near residues along the primary structure, i.e., the local or short-range interactions, are known to be insufficient, alone, for understanding the tertiary structural preferences of proteins alone. Yet, knowledge of short-range conformational potentials permits rationalizing the secondary structure propensities and aids in the discrimination between correct and incorrect tertiary folds. Proteins 29:292–308, 1997.**
© 1997 Wiley-Liss, Inc.

Key words: knowledge-based potentials; virtual bonds; coupling between bond torsions and bond angles; secondary structure propensities; inverse folding experiments

INTRODUCTION

In this study, short-range interactions observed in globular proteins are explored. Short-range interactions, also termed local interactions, refer to those taking place between near neighbor amino acids along the main chain; they determine the conformational distributions of bond angles and bond torsional states of the backbone. This is a one-dimensional problem, which is suitably analyzed by the tools of linear Ising or Markov chain models, as well as the classical rotational isomeric state approximation of polymer statistics.[1] A set of residue-specific empirical energy parameters is extracted here and used for interpreting experiments and recognizing correct sequence-structure pairs.

The present study complements our two recent analyses[2,3] on nonbonded interactions between side chains (S-S), in a self-consistent way. The former emphasizes the dominance of hydrophilic interactions at close ($r \leq 4.0$ Å) interresidue distances,[2] which contrasts the well-established major role of hydrophobic interactions at broader ($r \leq 7$ Å) distances.[4] The second characterizes the residue-specific coordination geometry of side chains.[3] In both studies, the interactions between side-chain pairs separated by at least two intervening residues, or five virtual bonds (three backbone and two side-chain bonds), are taken into consideration. These are shortly referred to as *long-range interactions.* A sensible analysis of protein structural preferences should, on the other hand, take into consideration both the long-range and short-range preferences of the chain. Honig and Cohen[5] recently pointed out, for example, that side-chain-only models cannot capture the essential features of a folding pathway, due to the neglect of the chemical nature of the polypeptide backbone. The merit of combining interaction potentials of various types for the identifica-

Contract grant sponsor: NATO CRG Project; Contract grant number: 951240.

*Correspondence to: R.L. Jernigan, Molecular Structure Section, Laboratory of Mathematical Biology, Division of Basic Sciences, National Cancer Institute, National Institutes of Health, MSC 5677, Room B-116, Building 12B, Bethesda, MD 20892-5677.

Received 5 February 1997; Accepted 16 June 1997

tion of native protein structures has been emphasized in several recent studies[6,7] and is verified here.

Usually, statistical treatments of protein conformations considered the interdependence of the backbone torsion angles φ and ψ adjacent to the peptide bond (Ramachandran plots) and neglected higher order interdependences between bond dihedral angles. Systematic analyses of (φ, ψ) angles for different types of amino acids lead to similarity coefficients between amino acid pairs, which are useful in estimating the structural effects of amino acid substitutions and providing an efficient scoring scheme for sequence alignments.[8] On the other hand, common secondary structure motifs, such as α-helices and β-sheets, result from the repetition of well-defined rotations along the main chain. Turns also constitute units that are distinguishable by their particular dihedral angle sequences. These observations suggest that more precise preferences for particular secondary structures can be accounted for by appropriate selection from probability distributions for *correlated* or *interdependent* bond rotations. For example, knowledge-based conditional probabilities $P_{B|A}(\phi_i, \psi_i)$ for the dihedral angles of residues of type B, given the identity (A) of the residue i-1, have proven useful in a genetic algorithm to reproduce the native folds of a few small proteins, such as melittin, avian pancreatic inhibitor, and apamin.[9] Likewise, conformational states and energies of tripeptides, incorporating the triplewise interdependence of adjacent amino acids, were successfully used by Nishikawa and Matsuo,[10] along with energy terms accounting for side-chain packing, hydration, and hydrogen bonding, in evaluating sequence-structure compatibility and detecting weak homologies.

In the classical (ϕ, ψ) representation, the interdependence of the dihedral angles for consecutive residues A and B is expressed by a four-dimensional probability distribution, $P_{AB}(\phi_{i-1}, \psi_{i-1}, \phi_i, \psi_i)$. Conformational potentials cannot be accurately extracted for this fine level of description, because of insufficient data, despite the growing number of structures determined by X-ray or NMR; it is expedient to resort to lower resolution descriptions of the backbone. The virtual bond representations of the protein backbone have proven to provide physically reliable, yet mathematically tractable models of protein structures on a local scale.[11–14] Such representations date back to the original work of Brant and Flory.[15]

A schematic representation of the virtual bond model adopted here is given in Figure 1. Conforming with the original work of Brant and Flory,[15] the model consists of virtual bonds $\mathbf{l}_i$, of magnitude l_i, pointing from C^{α}_{i-1} to C^{α}_{i}, virtual bond angles θ_i between $\mathbf{l}_i$ and $\mathbf{l}_{i+1}$, and torsional angles or pseudodihedrals ϕ_i defining the rotation about bond $\mathbf{l}_i$. The interdependence of the pseudodihedrals ϕ_i and ϕ_{i+1} was investigated by DeWitte and Shakhnovich[13] in terms of three groups of amino acids: helix formers, sheet formers, and turn formers, leading to a total of nine combinations of pairs of pseudodihedrals. Here, a more detailed approach is taken: residue-specific distributions and conformational energies are generated; the analysis is performed both for pairs of pseudodihedrals (ϕ_i, ϕ_{i+1}) and for interdependent pairs of virtual bond angles and pseudodihedrals (θ_i, ϕ_{i+1}) and (θ_i, ϕ_i), as a function of the type of the ith residue. 302 Brookhaven Protein Data Bank (PDB) structures,[16,17] also considered in our recent studies of nonbonded potentials,[2,3] are analyzed for extracting knowledge-based data on the conformational statistics of virtual bonds. Together with the recently derived long-range potentials,[2] a low-resolution model is constructed, which is tested in threading experiments. No gaps are considered in these tests. The short-range potentials are shown by use in inverse folding calculations to be quite specific for recognizing the native sequence for a given fold, even without including the long-range potentials. However, the performance of the threading tests improves by combining short-range and long-range potentials.

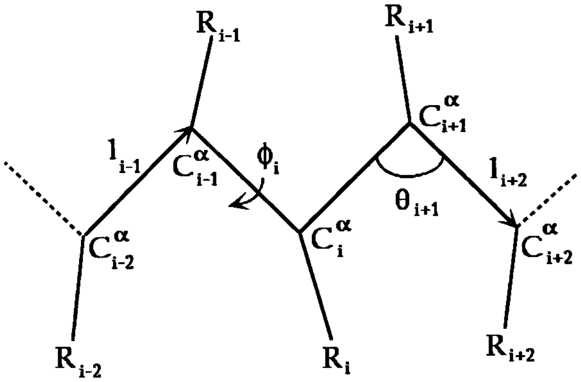

Fig. 1. Schematic representation of the C^{α}-C^{α} virtual bond model. C^{α} atoms indexed from $i - 2$ to $i + 2$ are shown together with the side groups R_{i-2}-R_{i+2} attached to them. The virtual bond vector $\mathbf{l}_i$ points from C^{α}_{i-1} to C^{α}_{i}. θ_i is the bond angle between l_i and l_{i+1}. ϕ_i is the torsion angle defining the rotation about bond l_i. It is taken as 180° and 0° (or 360°) in the *trans* and *cis* conformations, respectively.

RESULTS

Probability Distributions of Virtual Bond Angles and Torsions

The backbone configuration is defined by the set of bond angles and torsion angles $(\Theta, \Phi) \equiv \{\theta_2, \theta_3, \ldots, \theta_{n-1}, \phi_3, \phi_4, \ldots, \phi_{n-1}\}$, C^{α} atoms being indexed from 1 to n. Backbone virtual bond lengths are $1_i = 3.81 \pm 0.02$ Å when the peptide bond is in the *trans* conformation. The rare occurrence of *cis* peptide bonds is neglected. The rotations ϕ_2 and ϕ_n of the terminal virtual bonds do not affect the internal configuration of the main chain and need not be specified. Thus, the total number of variables defining the backbone configuration is 2n-5. This may be compared with

3n-6, the number of degrees of freedom in a fully unconstrained atomic description of the backbone. Despite the reduction in the number of variables, the present model can successfully characterize the backbone conformation and account for secondary structure preferences. Furthermore, correlations inherently due to chain connectivity, involving up to 12 (real) bonds along the chain are conveniently treated.

First, *singlet* probabilities are collected for the distribution of the bond angles and dihedrals for each residue type A. End effects are neglected, and probabilities are evaluated for all internal angles, irrespective of their location along the backbone. The singlet probabilities are designated as $P_A(\theta)$ for the virtual bond angle at any α-carbon belonging to residue A, $P_A(\phi^-)$ for the torsion of a virtual bond preceding a residue of type A, and $P_A(\phi^+)$ for that of the bond succeeding A. For illustrative purposes, the distributions obtained for a few residues are shown in Figures 2 and 3. $P_A(\theta)$ for A = Gly, Pro, and Asp are shown in Figure 2a. We note that these amino acids, which were recently classified as turn formers and assigned identical conformational energies,[13] are distinguishable by their bond angle distributions. The behavior of Ala, Val, and His is displayed in Figure 2b, again revealing that Ala and His, which were grouped[13] as helix formers, have distinct bond angle preferences and should be treated separately. The curve for Val in the same figure is typical of the virtual bond angle distribution of sheet formers Val, Ile, Tyr, Trp, Phe, and Thr, except for Cys (not shown), whose second peak is less pronounced and broader.

The distributions of pseudodihedrals $P_A(\phi^+)$ and $P_A(\phi^-)$ are illustrated in Figure 3. The curves for A = Gly, Asp, and Ala are displayed. Figure 3a indicates that it is inappropriate to classify Asp and Gly in the same group, even though both residues might favor formation of turns. Another observation is that the pseudodihedrals preceding and succeeding a given residue can exhibit different statistical behavior. In particular, the peak positions in the $P_A(\phi^+)$ and $P_A(\phi^-)$ curves for Gly are quite different. $P_A(\phi^+)$ for Gly is distinguished by its relatively flat distribution with three peaks. A similar trimodal distribution with peaks at $\phi^+ = 60°, 180°$, and 300°, which correspond to gauche$^-$, trans and gauche$^+$ states, respectively, was also observed in $P_A(\phi^+)$ for Cys (not shown). $P_A(\phi^-)$ curves, on the other hand, exhibit two peaks at 60° and 210°, approximately, for all A. These two peaks, separated by approximately π, are also observed in all $P_A(\phi^+)$ curves, apart from Cys and Gly. The highest peak (60°) is characteristic of α-helices in the virtual bond model. The second peak that is broader and weaker is for β-sheets. Likewise, among the two peaks located at approximately 90° and 120° in the $P_A(\theta)$ curves (Fig. 2), the former is strongly correlated with α-helices, and the second with β-sheets.

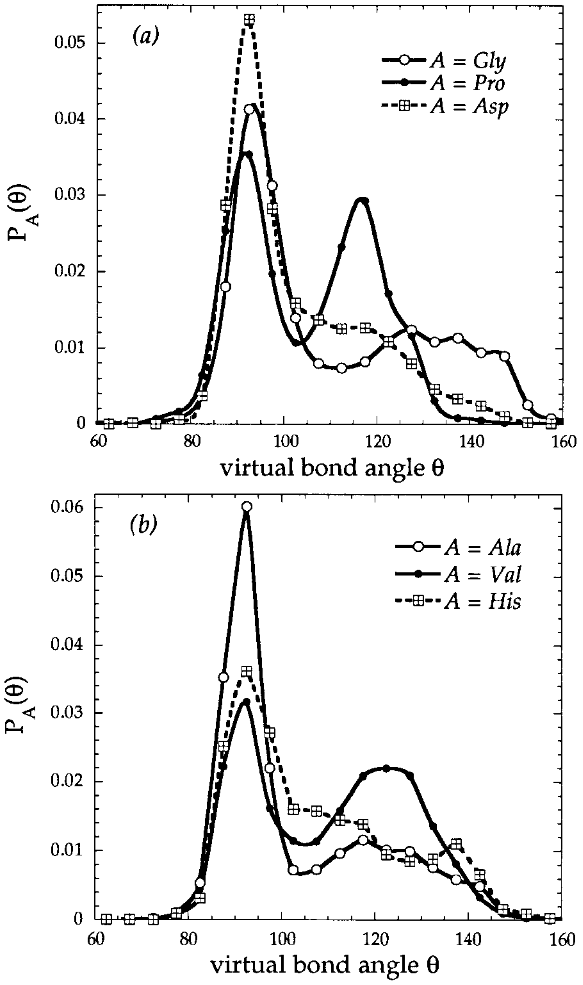

Fig. 2. Singlet probability distribution functions $P_A(\theta)$ for the virtual bond angle at the α-carbons of the particular residues (**a**) A = Gly, Pro, and Asp, (**b**) A = Ala, Val, and His. The distributions are normalized, i.e., the areas under the curves are equal to unity. Results are obtained with 10° intervals. The peaks near 90° and 120° include α-helices and β-sheets, respectively.

At the next level of approximation, correlations between virtual bond angles and pseudodihedrals are examined. Different types of pairwise couplings are observed: ϕ^- and ϕ^+ are strongly coupled, and the type of interdependence is a function of the type A of the residue located between these two virtual bonds. Likewise, θ is correlated strongly with the torsions ϕ^- and ϕ^+ of the adjoining bonds. The overall numbers $N(\phi^-, \phi^+)$, $N(\theta, \phi^-)$, and $N(\theta, \phi^+)$ of pairs of angles observed in the presently examined databank structures[3] are shown in Figure 4. These are essentially unnormalized probability surfaces and are referred to as *doublet* distributions. These data are collected irrespective of the type of residue. Although maps similar to those of Figure 4 have been obtained in previous studies,[12,13] we include these to provide a basis for comparing the behavior of specific residues (see Figs. 5–7). The region of the

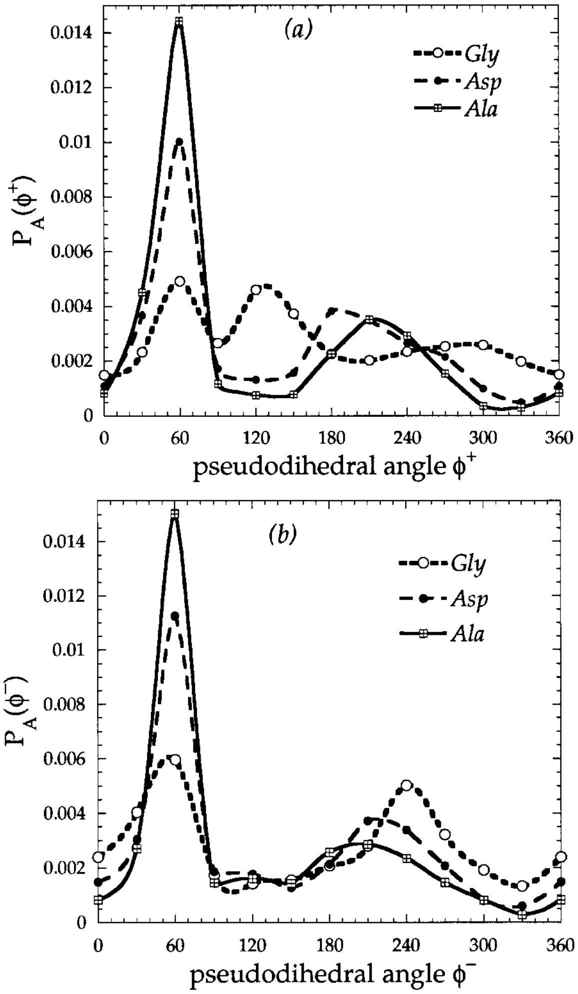

Fig. 3. Singlet probability distributions, $P_A(\phi^+)$ (**a**) for the torsion of a virtual bond succeeding a residue of type A, and $P_A(\phi^-)$ (**b**) for that of a bond preceding A, shown for A = Gly, Asp, and Ala. The curves are normalized. Intervals of 30° are considered. The peak near 60° is characteristic of α-helices, and the second around 210°, which is broader and weaker, is associated with β-sheets.

surface for $N(\phi^+, \phi^-)$ near $(\phi^+, \phi^-) \approx (60°, 60°)$, which is distinguished by the highest density, is characteristic of α-helices, whereas β-sheets are characterized by dihedral angles around $(\phi^+, \phi^-) \approx (210°, 210°)$, the region of next highest density. The distribution associated with β-sheets is more diffuse and less sharp than that of α-helices. Likewise, the pronounced peak at $(\theta, \phi^{\pm}) \approx (90°, 60°)$ is characteristic of α-helical state, whereas the β-sheet region is characterized by the joint state $(\theta, \phi^{\pm}) \approx (120°, 210°)$.

The strong dependence of the conformational distributions on the identity A of the residue may be seen from the comparison of the maps obtained for particular residues (Figs. 5–7) with the average distributions (Fig. 4). Figure 5 displays $N_A(\phi^+, \phi^-)$ surfaces for A = Gly, Pro, Val and Glu. The strong preference of Glu, for example, for α-helices is distinguishable, whereas the versatility of Gly to adopt various torsional states is apparent. Pro exhibits unique preferences. The β-sheet region is more frequent for Val compared to other residues. Figures 6 and 7 illustrate the coupling between the virtual bond angles and torsion angles, for some selected residues. The surfaces $N_A(\theta, \phi^+)$ in Figure 6 for Gly and Asn reveal the strikingly different conformational characteristics of these two residues, despite their common tendency to favor turns.

Short-Range Interaction Energies

Residue-specific conformational potentials are developed on the basis of the $N_A(\phi^+, \phi^-)$, $N_A(\theta, \phi^+)$, and $N_A(\theta, \phi^-)$ distributions in their normalized forms $P_A(\phi^+, \phi^-)$, $P_A(\theta, \phi^+)$, and $P_A(\theta, \phi^-)$. We resort to discrete state formalism, inasmuch as statistically reliable evaluations of doublet energies is possible only by considering sufficiently large regions of the surfaces. Intervals of size $\Delta\phi^{\pm} = 30°$ and $\Delta\theta = 10°$ in the respective ranges $0° \leq \phi^{\pm} \leq 360°$ and $60° \leq \theta \leq 180°$ are taken.

We define, for a given residue A at position i along the primary sequence of the protein, a conformational energy of the form

$$E_A(\theta_i, \phi_i, \phi_{i+1}) = E_A(\theta_i) + E_A(\phi_i) + E_A(\phi_{i+1}) + \Delta E_A(\theta_i, \phi_i) + \Delta E_A(\theta_i, \phi_{i+1}) + \Delta E_A(\phi_i, \phi_{i+1}). \quad (1)$$

Here $E_A(\theta_i)$, $E_A(\phi_i)$, and $E_A(\phi_{i+1})$ are the virtual bond angle and torsion energies corresponding to the states θ_i,ϕ_i, and ϕ_{i+1} of the virtual bonds about A, assuming these variables to be independent of each other, and $\Delta E_A(\theta_i, \phi_i)$ $\Delta E_A(\theta_i, \phi_{i+1})$, and $\Delta E_A(\theta_i, \phi_{i+1})$, are the increments accounting for their pairwise interdependences. $E_A(\theta_i)$, $E_A(\phi_i)$, and $E_A(\phi_{i+1})$ are evaluated from

$$\begin{aligned} E_A(\theta_i) &= -RT \ln [P_A(\theta)/P°_A(\theta)] \\ E_A(\phi_i) &= -RT \ln [P_A(\phi^-)/P°_A(\phi^-)] \\ E_A(\phi_{i+1}) &= -RT \ln [P_A(\phi^+)/P°_A(\phi^+)] \end{aligned} \quad (2)$$

where $P°_A(\theta)$, $P°_A(\phi^+)$, and $P°_A(\phi^-)$ are the uniform distribution probabilities, i.e., those valid in the absence of any correlations. In continuous space, $P°_A(\theta) = 1/\pi$, and $P°_A(\phi) = 1/2\pi$; in the discrete state formalism, they are directly proportional to the size of the angular intervals defining the states. The terms accounting for the coupling among the three degrees of freedom are likewise estimated from

$$\begin{aligned} \Delta E_A(\theta_i, \phi_i) &= -RT \ln [P_A(\theta, \phi^-)/P_A(\theta)P_A(\phi^-)] \\ \Delta E_A(\theta_i, \phi_{i+1}) &= -RT \ln [P_A(\theta, \phi^+)/P_A(\theta)P_A(\phi^+)] \\ \Delta E_A(\phi_i, \phi_{i+1}) &= -RT \ln [P_A(\phi^-, \phi^+)/P_A(\phi^-)P_A(\phi^+)] \end{aligned} \quad (3)$$

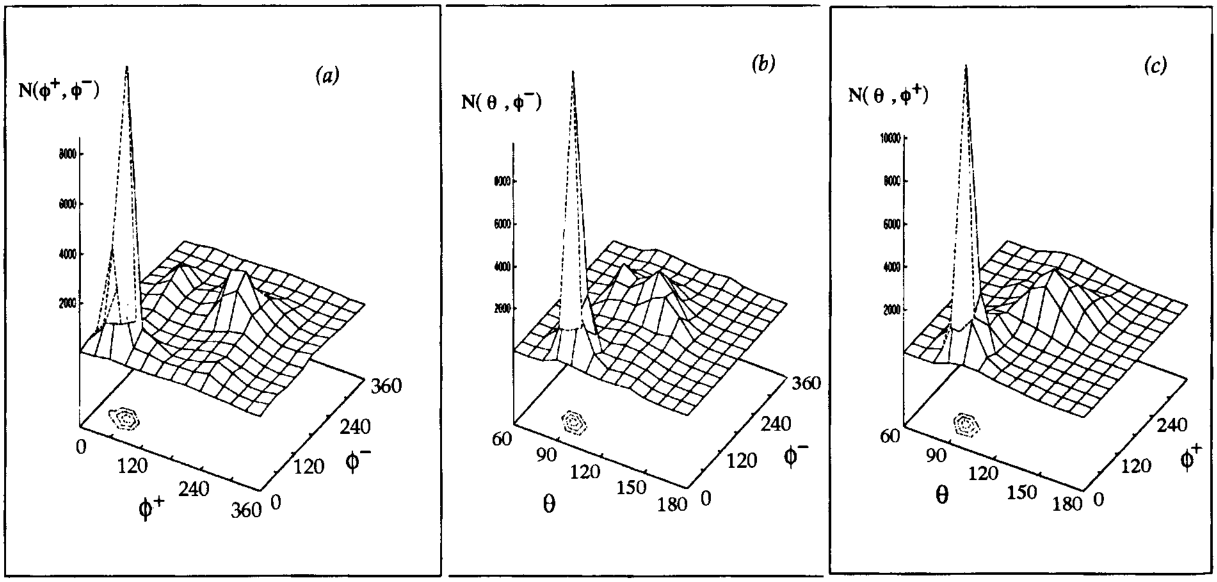

Fig. 4. Number distributions of pairs interdependent angles (ϕ^+, ϕ^-), (θ, ϕ^-) and (θ, ϕ^+), collected for all residue types in a set of previously reported[3] 150 PDB structures. The doublet distribution N(ϕ^-, ϕ^+) is determined by considering regions of size ($\Delta\phi^-$, $\Delta\phi^+$) = (30°, 30°). The region about (ϕ^-, ϕ^+) ≈ (60°, 60°) is characteristic of α-helices, whereas the β-sheet structures correspond to (ϕ^-, ϕ^+) ≈ (210°, 210°). N(θ, ϕ^+) and N (θ, ϕ^-) are found by using $\Delta\theta$ bins of size 10°, in the range $60° \leq \theta \leq 180°$. Regions of highest density are shown by the contours projected onto the lower horizontal plane.

The probability distributions used are for all bond angles and torsions for a given residue type A, regardless of position along the chain.

The overall short-range conformational energy E(Θ, Φ) of a given protein in conformation (Θ, Φ) is expressed as

$$E(\Theta, \Phi) = -RT \ln P(\Theta, \Phi) \qquad (4)$$

where P(Θ, Φ) is the probability of conformation (Θ, Φ). In a strict sense, the right-hand side of Eq. (4) should include an additional term, RT ln Z_c, where Z_c is the conformational partition function defined by the summation of the Boltzmann weights over all (Θ, Φ). This term is eliminated here by replacing the weights by the normalized probabilities. Within the limits of applicability of Markov chain statistics,[1] E(Θ, Φ) is rewritten as

$$E(\Theta,\Phi) = \sum_{i=2}^{n-1} E_i(\theta) + \sum_{i=3}^{n-1} [E_i(\phi^-)/2 + E_{i-1}(\phi^+)/2] + \sum_{i=3}^{n-1} [\Delta E_{i-1}(\theta, \phi^+) + \Delta E_i(\theta, \phi^-)] + \sum_{i=3}^{n-2} \Delta E_i(\phi^-, \phi^+). \qquad (5)$$

Here, the indices refer to the sequential position of the residues along the chain and thereby depend on the residue type. Following the conventional terminology of polymer statistics, $E_i(\phi^+)$, $E_i(\phi^-)$, and $E_i(\theta)$ will be referred to as *first-order* interaction energies, and the remainder, $\Delta E_i(\phi^-, \phi^+)$, $\Delta E_i(\theta, \phi^+)$, and $\Delta E_i(\theta, \phi^-)$, as *second-order* interaction energies.[1] These two contributions are expressed as

$$E^{(1)}(\Theta, \Phi) = \sum_{i=2}^{n-1} E_i(\theta) + \sum_{i=3}^{n-1} [E_i(\phi^-)/2 + E_{i-1}(\phi^+)/2] \qquad (6)$$

and

$$E^{(2)}(\Theta, \Phi) = \sum_{i=3}^{n-1} [\Delta E_{i-1}(\theta, \phi^+) + \Delta E_i(\theta, \phi^-)] + \sum_{i=3}^{n-2} \Delta E_i(\phi^-, \phi^+), \qquad (7)$$

respectively. The two types of interactions are shown below to be almost equally important in stabilizing native folds.

Calculations have been performed independently for two sets of PDB structures, including each ≥150 proteins,[2,3] which confirmed the reproducibility of the database extracted conformational energies. The values of $E_A(\theta)$, $E_A(\phi^+)$, $E_A(\phi^-)$, $E_A(\theta, \phi^+)$, $E_A(\theta, \phi^-)$, and $E_A(\phi^-, \phi^+)$ for all types A of residues, listed at 10° intervals of θ for $60° \leq \theta \leq 180°$ and 30° intervals of ϕ in the range $0° \leq \phi^+, \phi^- \leq 360°$ are available in the supplementary material, upon request.

APPLICATIONS AND DISCUSSION

Secondary Structure Propensities: Comparison With Experiments

α-Helix propensities

In Table I, torsional energies for α-helical states of pairwise interdependent virtual bonds are pre-

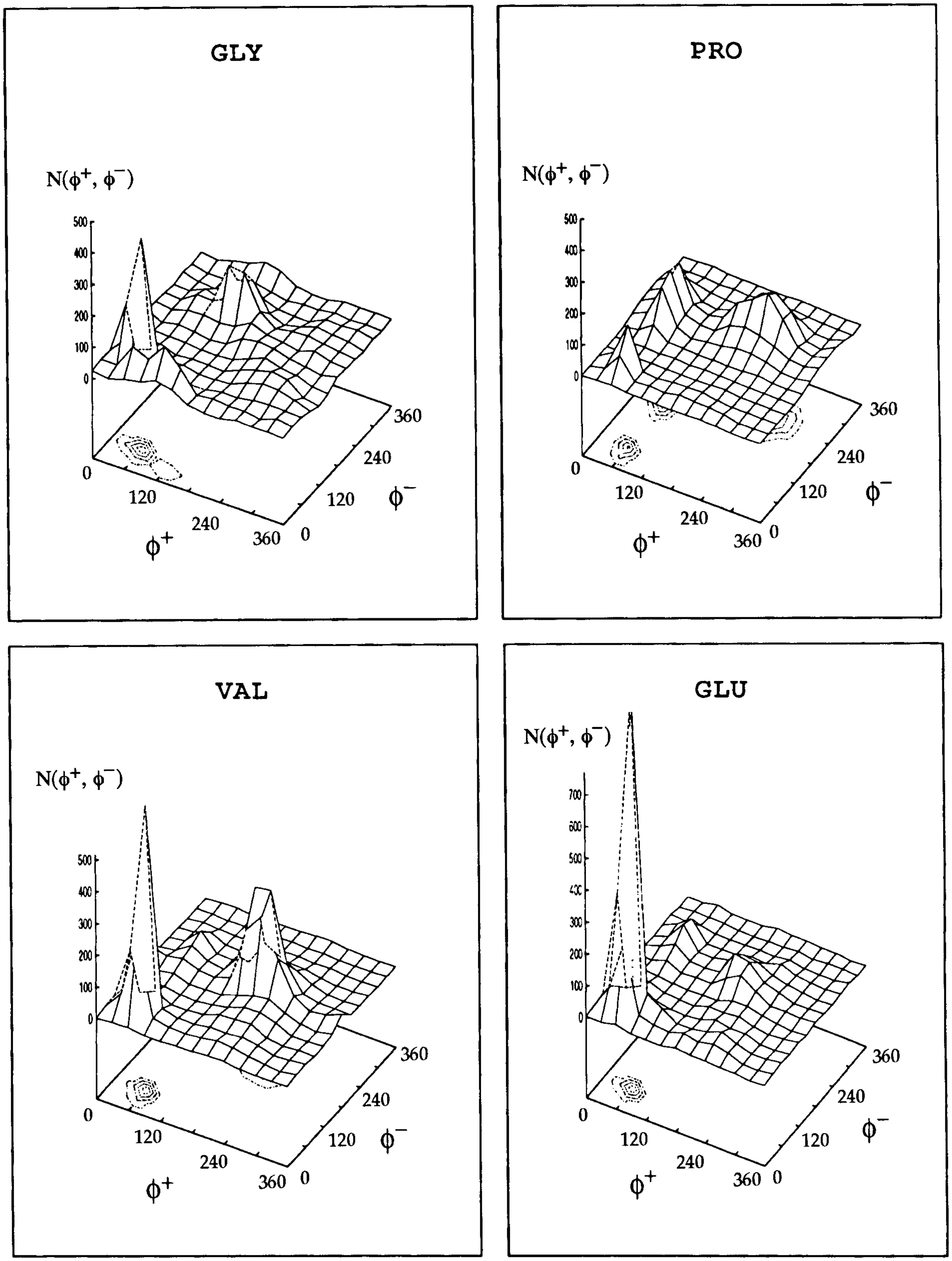

Fig. 5. Doublet distributions of pseudodihedrals, $N_A(\phi^+, \phi^-)$, collected for A = Gly, Pro, Val, and Glu. The curves reflect the strong preference of Glu for α-helices, the inclination of Val for β-sheets, the versatility of Gly to assume a wide variety of conformational states, and the distinct characteristics of Pro. See the caption of Figure 4 for further details.

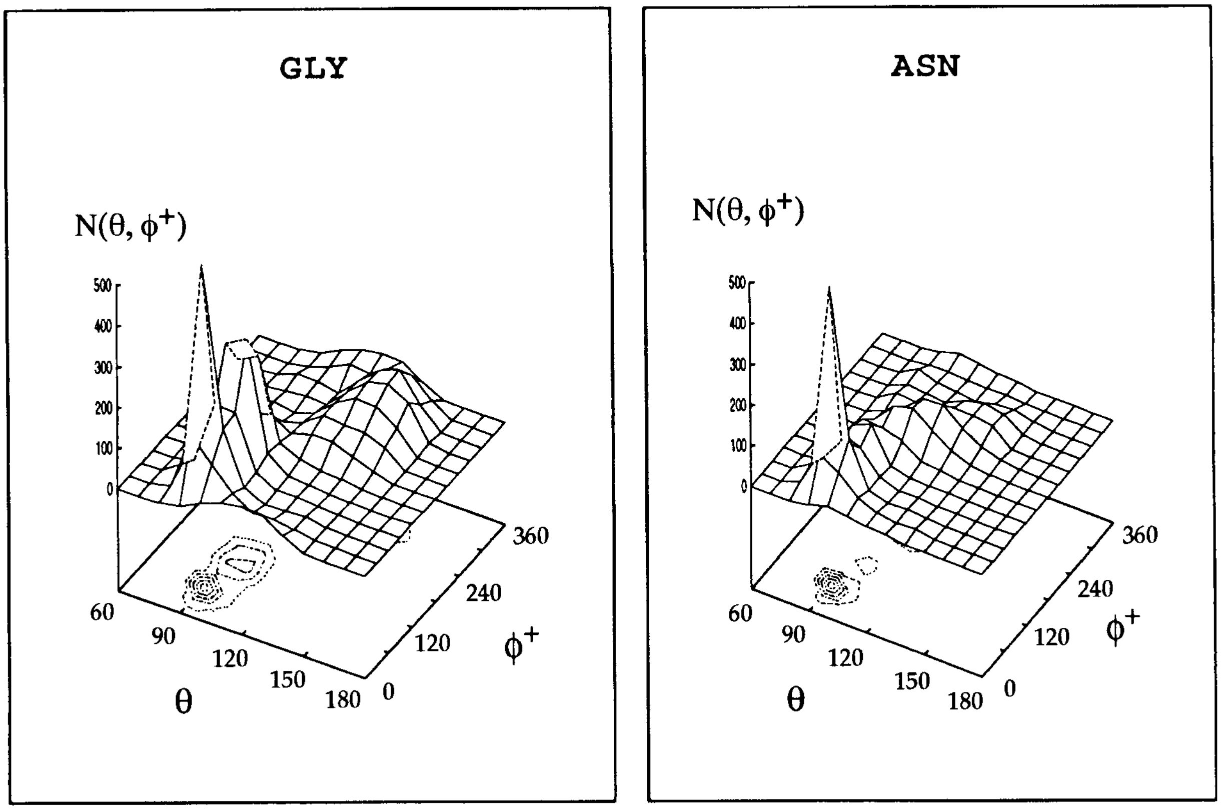

Fig. 6. N_A (θ, ϕ^+) for A = Gly and Asn reflecting the specificity of the coupling between the virtual bond angle θ and the pseudodihedral ϕ^+.

dihedral angles ($\phi^+_\alpha \pm \Delta\phi^+, \phi^-_\alpha \pm \Delta\phi^-$) = (60° ± 15°, 60° ± 15°), the subscript α referring to the α-helical state. Following Eqs. 1–3, the integration of the normalized probability distributions over this particular region of rotational angle space yields $E_A(\phi^+_\alpha, \phi^-_\alpha)$ for each residue type A. The residue types are listed in the first column in the order of decreasing strength of energy $E_A(\phi^+_\alpha, \phi^-_\alpha) \equiv E_A(\phi^+_\alpha) + E_A(\phi^-_\alpha) + \Delta E_A(\phi^+_\alpha, \phi^-_\alpha)$. In all cases, attractive interactions, presumably reflecting the effect of favorable backbone-backbone hydrogen bonding, are operative, except for $E_{PRO}(\phi^-_\alpha)$. Ala exhibits the strongest tendency to participate in α-helices, followed by Met ≈ Glu > Leu ≈ Gln ≈ Arg > Lys ≈ Trp > Phe ≈ Ile, whereas in the other extreme case of residues disfavoring or breaking α-helices the order Pro > Gly > Thr > Tyr ≈ Cys ≈ Val ≈ Ser ≈ Asn ≈ His is seen, similarly to earlier knowledge-based studies.[18–21] A pair of bonds flanking an alanine is subject to a conformational energy which is lower by almost 2 RT compared with that of a pair of bonds about Pro. Gly is the next most helix destabilizing residue, its torsional energy being higher than that of Ala by 1.4 RT.

The comparison of the energies $E_A(\phi^+_\alpha)$ and $E_A(\phi^-_\alpha)$ for different residues gives some indications about the intrinsic helix-capping preferences of individual residues. These may be compared with experimental observations, although the stabilizing effect of different amino acids at helix termini can strongly depend on context.[22] The observation that $E_A(\phi^+_\alpha)$ is more favorable than $E_A(\phi^-_\alpha)$ for A = Pro by approximately 1.3 RT, for example, suggests that proline is likely to be succeeded by an α-helix, rather than preceded. This is in accord with more detailed studies of capping preferences in helices.[23] This is due to the inability of its N to form hydrogen bonds. On the other hand, a preference for the C-terminus of α-helices, rather than the N-terminus, is discernible for A = Gly, His, Asn, and Asp. Ile and Val exhibit the opposite tendency. These results agree to some extent with those deduced by Serrano et al.[22] from a series of mutations at the N-caps, C-caps, and internal positions of the solvent-exposed faces of two α-helices of barnase. The residues exhibiting the strongest preferences for the C-cap indeed are reported in that study to be Gly > His > Asn. On a more recent scale, the residues experiencing the most favorable free energy at the C-cap are again shown to be Gly, His, and Asn, all three being approximately equally stable.[24] We note that approximately one third of all α-helices have been pointed out to terminate with Gly at the carboxyl end.[25] The preferences at the N-cap, on the other hand, are determined by hydrogen bonding of side chains or solvent to exposed backbone N-H groups. Asp, Asn, Thr, and Ser are reported[24] to have the strongest

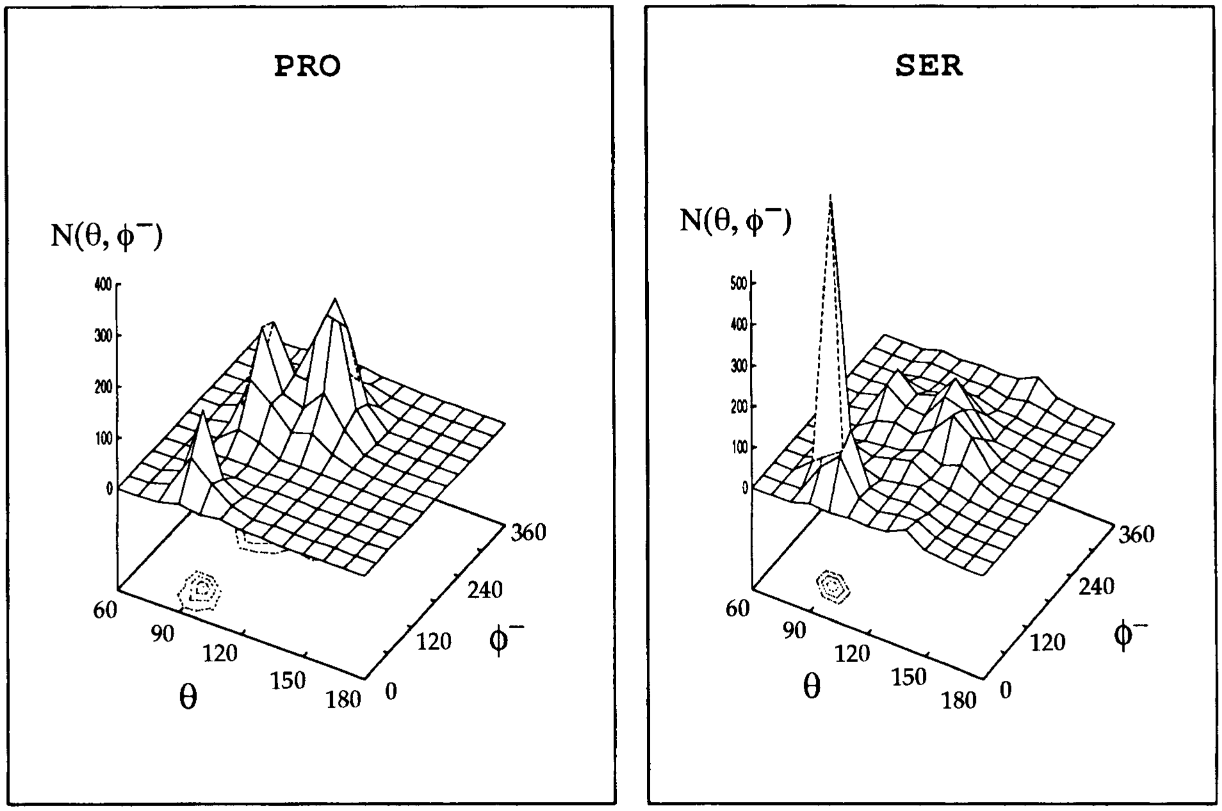

Fig. 7. $N_A(\theta, \phi^-)$ for A = Pro and Ser, illustrating the residue-specific coupling between θ and ϕ^-.

preference for the N-cap; these preferences are not observed in our study.

Our scale of helical propensities bears a close resemblance to the ranking Ala > Leu > Met > Gln > Ile > Val > Ser > Thr > Asn > Gly determined[26] from circular dichroism. The relative order of Leu and Met, Ser and Val, and the rank of Asn, are different in our case, which matches the order of propensities reported by Chou and Fasman.[27,28] Also, we note that the relative helix stabilizing tendencies of alanine and hydrophobic residues found here, Ala > Leu > Phe > Ile > Val, are consistent with those reported by O'Neil and DeGrado.[29] Baldwin and co-workers found for the same subset of residues, the order Leu ≈ Ala > Phe > Ile > Val, and invited attention to the fact that the helix-forming tendency of a particular amino acid might depend on the sequence context in which it occurs.[30] For instance, the substitution of the solvent-exposed residue Ala32 of barnase by all other naturally occurring amino acids emphasizes Arg and Lys as the two residues having the least destabilizing effect on the original helical structure.[31] Likewise, the helix-stabilizing tendency of Phe is found by Blaber et al.[32] to be quite weak, in contrast to our results and to the above mentioned experiments.

Figure 8a shows a comparison of our results with the structure-based thermodynamic scale of α-helix propensities recently derived by Luque et al.[33] $\Delta G_{A\alpha}$ values shown on the abscissa are their free-energy changes for helix formation at 25°C for 19 amino acids (excluding proline). The ordinate displays our residue-specific doublet energies $E_A(\phi^+{}_\alpha, \phi^-{}_\alpha)$ at the same temperature directly taken from the last column of Table I. A linear regression analysis omitting the charged residues, which are known to be sensitive to ionic strength,[33] yields the remarkably good line drawn through the data with a correlation coefficient of 0.93. The corresponding equation is

$$E_A(\phi^+{}_\alpha, \phi^-{}_\alpha) = 1.00\ \Delta G_{A\alpha} + 2.52. \qquad (8)$$

Inclusion of the charged residues decreases the correlation coefficient to 0.89, and the slope of the best fitting line becomes 0.90. Thus, there is a strikingly good agreement between our empirical results and the thermodynamic analysis of Luque and collaborators,[33] despite the fundamental differences in methods.

We also have compared our results with the $\Delta\Delta G_{A\alpha}$ values measured by O'Neil and deGrado.[29] $\Delta\Delta G_{A\alpha}$ values are found from coiled monomer-helical dimer equilibrium constant measurements taken before and after mutating a helical residue (A) in a solvent-exposed site of a synthetic polypeptide. The results are expressed relative to Gly. Plotting $\Delta\Delta G_{A\alpha}$ values for all residues except the charged ones, against the

TABLE I. Rotational Potentials for Virtual Bond Torsions (ϕ^+_α, ϕ^-_α) Characteristic of α-Helical Structures*

A	$E_A(\phi^+_\alpha)/RT$	$E_A(\phi^-_\alpha)/RT$	$\Delta E_A(\phi^+_\alpha, \phi^-_\alpha)/RT$	$E_A(\phi^+_\alpha, \phi^-_\alpha)/RT^\dagger$
Ala	−1.65	−1.69	−0.52	−3.85 ± .05
Met	−1.53	−1.58	−0.57	−3.68 ± .04
Glu	−1.60	−1.60	−0.43	−3.64 ± .02
Leu	−1.47	−1.48	−0.61	−3.55 ± .03
Gln	−1.46	−1.53	−0.55	−3.54 ± .05
Arg	−1.46	−1.49	−0.56	−3.50 ± .08
Lys	−1.38	−1.42	−0.48	−3.38 ± .02
Trp	−1.40	−1.33	−0.64	−3.37 ± .06
Phe	−1.24	−1.32	−0.72	−3.28 ± .09
Ile	−1.29	−1.15	−0.83	−3.26 ± .03
Asp	−1.20	−1.35	−0.66	−3.21 ± .05
His	−1.10	−1.27	−0.72	−3.09 ± .06
Asn	−1.07	−1.21	−0.74	−3.08 ± .10
Ser	−1.17	−1.16	−0.75	−3.08 ± .04
Val	−1.17	−0.98	−0.92	−3.06 ± .09
Cys	−1.01	−0.99	−1.01	−3.01 ± .05
Tyr	−1.07	−1.09	−0.83	−3.00 ± .03
Thr	−1.01	−1.05	−0.87	−2.91 ± .05
Gly	−0.54	−0.68	−1.23	−2.45 ± .06
Pro	−1.17	+0.11	−0.86	−1.92 ± .14

*The α-helix region is characterized by ($\phi^-_\alpha \pm \Delta\phi^-$, $\phi^+_\alpha \pm \Delta\phi^+$) = (60° ± 15°, 60° ± 15°).

†The errors marked in the last column refer to the differences between the results obtained from two sets[3] comprising 150 and 152 PDB structures, respectively.

difference $E_A(\phi^+_\alpha, \phi^-_\alpha) - E_{GLY}(\phi^+_\alpha, \phi^-_\alpha)$ leads to the fitted equation

$$E_A(\phi^+_\alpha, \phi^-_\alpha) - E_{GLY}(\phi^+_\alpha, \phi^-_\alpha) = 0.803\ \Delta\Delta G_{A\alpha} + 0.099 \quad (9)$$

with a correlation coefficient of 0.90.

In Figure 8b, the potentials $E_A(\phi^+_\alpha, \phi^-_\alpha)$ at T = 300 K are compared with the side-chain entropy changes $T\Delta S_{A\alpha}$ upon helix formation, calculated by Creamer and Rose[34] from rotamer distributions of amino acid side chains. The ordinate values are expressed relative to Ala for comparison with the entropy changes on the abscissa. The best fit line obeys the equation

$$E_A(\phi^+_\alpha, \phi^-_\alpha) - E_{ALA}(\phi^+_\alpha, \phi^-_\alpha) = -0.82\ T\Delta S_{A\alpha} - 0.007 \quad (10)$$

with a correlation coefficient of $\chi = 0.92$. The entropy calculations were performed[34] with Monte Carlo generation of short acetylated (Ace) and methyl amidated (NMe) peptides of the form ($AceAla_5XAla_5NMe$), in which Ala, Val, Ile, Leu, Met, Phe, Tyr, and Trp were substituted for the central residue X. The agreement between our potentials and the side group entropy losses invites attention to the contribution of entropic effects in determining the α-helix propensities of amino acids.

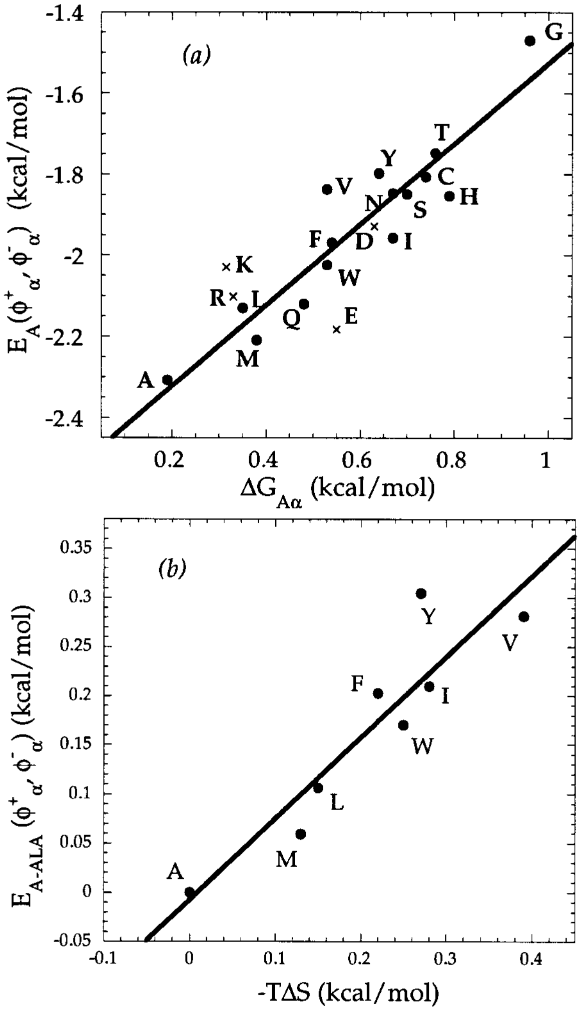

Fig. 8 **a**: Comparison of the α-helical propensities of amino acids obtained in the present study with the thermodynamic scale derived by Luque et al.[33] by using a structure-based optimization scheme. The residue-specific changes in free energy $\Delta G_{A\alpha}$ associated with α-helix formation are shown on the abscissa. These are plotted against the energies $E_A(\phi^+_\alpha, \phi^-_\alpha)$ presented in the last column of Table I, evaluated at 25°C. The linear regression line is obtained with a correlation coefficient $\chi = 0.93$. **b**. Comparison of the potentials $E_A(\phi^+_\alpha, \phi^-_\alpha)$ with the side-chain entropy changes $T\Delta S$ associated with helix formation. $T\Delta S_\alpha$ values on the abscissa were calculated by Creamer and Rose[34] for A = Ala, Val, Ile, Leu, Met, Phe, Tyr, and Trp. The ordinate values are expressed relative to that of Ala, i.e., $E_{A\text{-}ALA}(\phi^+_\alpha, \phi^-_\alpha) \equiv E_A(\phi^+_\alpha, \phi^-_\alpha) - E_{ALA}(\phi^+_\alpha, \phi^-_\alpha)$. The best fit line yields $\chi = 0.92$.

Finally, a comparison of our doublet potentials with the free-energy values calculated by Muñoz and Serrano,[24] yields (data not shown) a relationship of the form

$$\Delta E_A(\phi^+_\alpha, \phi^-_\alpha) - E_{ALA}(\phi^+_\alpha, \phi^-_\alpha) = 0.99\ \Delta G_{A\alpha} + 0.058 \quad (11)$$

with $\chi = 0.86$. Here, $\Delta G_{A\alpha}$ is the intrinsic free energy, relative to that of alanine, required to put the particular amino acid A in helical dihedral angles, excluding the contribution of hydrogen bonding. The free-energy change of proline is pointed out to include the contribution of a hydrogen bond dissociation[24] and is therefore excluded from the present analysis. It is interesting to notice that the absolute values of these two scales are in nearly perfect agreement, which is indicated by the slope 0.99 of the best fit line.

Several other examples lend support to the suitability of the torsion energies of Table I for interpreting α-helix propensities: Merutka and Stellwagen[35] examined water-soluble monomeric helices and found Ser and Met to be less helix stabilizing than Ala by 0.5 and 0.3 kcal/mol, respectively. The last column of Table I yields the respective values of 0.46 and 0.10 kcal/mol for the energy increases involved in these particular mutations at room temperature. Substitution of Ala for Gly46 and Gly48 in lambda repressor was reported to increase the stability by 0.66 and 0.87 kcal/mol, respectively.[36] The corresponding value given by O'Neil and DeGrado[29] is 0.77 kcal/mol. Table 1 indicates an increase in stability by 0.84 kcal/mol, for Gly → Ala substitution. Mutation of α-helical Val131 in T4 lysozyme to Ala or Thr results in a 0.23 kcal/mol stabilization or 0.08 kcal/mol destabilization, respectively.[37] Following Table I, an increase in stability by 0.47 kcal/mol is expected in the former mutation, and a decrease of 0.05 kcal/mol, in the latter. Finally, we note that our value of 0.47 kcal/mol for the energy difference between Ala and Val in α-helices lies between the values reported by Blaber et al.[32] and the peptide scale of O'Neil and DeGrado,[29] and almost coincides with the value (0.45 kcal/mol) from the scale of Lyu et al.[26]

TABLE II. Rotational Potentials for Virtual Bond Torsions ($\phi^+\beta$, $\phi^-\beta$) Characteristic of β-Sheet Structures*

A	$E_A(\phi^+\beta)/RT$	$E_A(\phi^-\beta)/RT$	$\Delta E_A(\phi^+\beta, \phi^-\beta)/RT$	$E_A(\phi^+\beta, \phi^-\beta)/RT$†
Ile	−0.85	−0.83	−0.70	−2.37 ± 0.04
Val	−0.83	−0.92	−0.59	−2.33 ± 0.01
Cys	−0.82	−0.67	−0.47	−1.96 ± 0.14
Tyr	−0.75	−0.63*	−0.55	−1.94 ± 0.09
Phe	−0.67	−0.53	−0.71	−1.90 ± 0.02
Leu	−0.54	−0.59	−0.77	−1.90 ± 0.03
Trp	−0.68*	−0.38	−0.79	−1.85 ± 0.02
Thr	−0.67	−0.64	−0.52	−1.83 ± 0.03
Met	−0.41	−0.39*	−0.97	−1.77 ± 0.07
His	−0.49*	−0.30	−0.82	−1.60 ± 0.10
Asp	−0.33*	−0.33	−0.92	−1.57 ± 0.13
Lys	−0.36	−0.42	−0.77	−1.55 ± 0.02
Gln	−0.27	−0.23	−1.01	−1.51 ± 0.13
Asn	−0.34*	−0.33	−0.77	−1.43 ± 0.10
Arg	−0.31*	−0.33	−0.74	−1.39 ± 0.11
Glu	−0.05*	−0.11	−1.07	−1.23 ± 0.03
Ala	−0.23	−0.08	−0.91	−1.21 ± 0.01
Pro	−0.60**	−1.18**	−0.29	−2.07 ± 0.03
Ser	−0.44**	−0.39**	−0.50	−1.33 ± 0.04

*The β-sheet region is characterized by the pair of torsional angles ($\phi^-_\beta \pm \Delta\phi^-$, $\phi^+_\beta \pm \Delta\phi^+$) = (210° ± 15°, 210° ± 15°) for all amino acids, except for a few cases where ϕ^+_β or ϕ^-_β assumes the value of 180° (indicated by *). Pro and Ser exhibit a local minimum around (ϕ^-, ϕ^+) = (240°, 240°) (indicated by **) and are listed separately.

†The errors refer to the differences between two sets[3] comprising 150 and 152 PDB structures, respectively.

β-*Sheet propensities*

Table II summarizes the calculated energy values $E_A(\phi^+_\beta, \phi^-_\beta)$ for different residues in β-sheet structures. The region of the dihedral angle maps centered about (ϕ^+_β, ϕ^-_β) = (210°, 210°) is considered in this case, except for a few cases in which one of the dihedral angles is shifted to the neighborhood of 180°, as indicated by the asterisk. Gly is excluded because no energy minimum attributable to β-sheet structures is observed. The energy values in Table II are not as strong as those for α-helices. This is consistent with the fact that a major contribution to the stability of β-sheets must come in general from long-range interactions, and these tabulated energy values reflect the contributions of short-range interactions only. In contrast to α-helices, which are mainly favored by their singlet energies, we note that the doublet energy changes $\Delta E_A(\phi^+_\beta, \phi^-_\beta)$ have a relatively larger effect on the stability of β-sheets. A rank order for β-sheet preferences takes the form of Ile ≈ Val > Cys > Tyr > Phe ≈ Leu > Trp ≈ Thr > Met > His > Asp ≈ Lys > Gln > Asn > Arg > Glu > Ala.

Kim and Berg[38] have measured the thermodynamic stability of a β-sheet-containing zinc finger protein when a given solvent-exposed residue position is substituted. Also, Bai and Englander[39] estimated β-sheet propensities from side-chain blocking effects controlling the rates of hydrogen exchange reactions. Except for Phe and Cys, the results from the two groups show close correspondence. The comparison of our results with those of Kim and Berg is displayed in Figure 9a. Here, $E_A(\phi^+_\beta, \phi^-_\beta)$ values are plotted against the change in free energies obtained by Kim and Berg for all types of amino acids, except Pro and Gly. We note that on the basis of singlet energies alone, listed in the first two columns of Table II, Val appears to be the strongest β-sheet forming residue. The correct correspondence with experiments, in which Ile is a stronger β-sheet former than Val, although a small change, is achieved only by invoking the interdependence of neighboring residues, i.e., by considering the second-order contribution $\Delta E_A(\phi^+_\beta, \phi^-_\beta)$. Likewise Thr, which was reported[28] to be the third strongest β-sheet former after Val and Ile, is shifted here to a substantially lower rank in agreement with experiments, mainly due to its less favorable doublet energy contribution. $\Delta E_A(\phi^+_\beta, \phi^-_\beta)$ is particularly strong (approximately −1.0 RT) for A = Gln, Glu, Met, Asp, Ala, and His and improves significantly the agreement between experiments and the present empirical results. Linear regression of the data presented in Figure 9a yields

$$\Delta E_A(\phi^+_\beta, \phi^-_\beta) = 3.08\ \Delta\Delta G_{A\beta} + 0.37 \qquad (12)$$

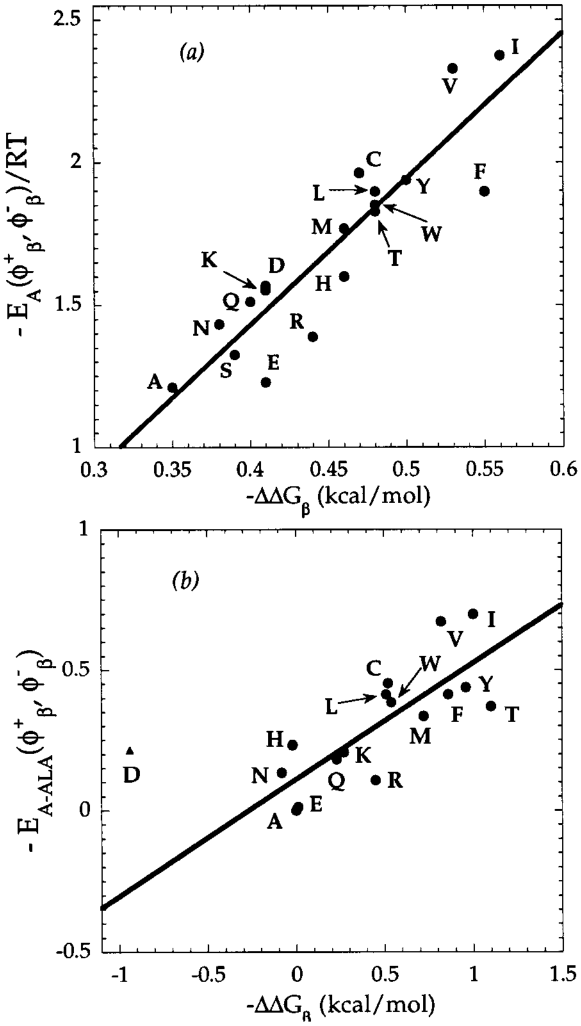

Fig. 9 **a:** Comparison of the β-sheet propensities of amino acids obtianed by experiments and in the present analysis. The abscissa and ordinate, respectively, to the experimental results of Kim and Berg[38] and to the torsional energies E_A (ϕ^+_β, ϕ^-_β)/RT listed in the last column of Table II. The correlation coefficient between the two data sets is R = 0.90. **b:** β-sheet propensities of amino acids obtained by experiments and in the present analysis. The abscissa represents the thermal stability measurements of Minor and Kim.[41] The ordinate follows from Table II, using $E_{A\text{-}ALA}(\phi^+_\beta, \phi^-_\beta) \equiv E_A(\phi^+_\beta, \phi^-_\beta) - E_{ALA}(\phi^+_\beta, \phi^-_\beta)$ at 300K. The value R = 0.79 is obtained, excluding the data for Asp.

with a correlation coefficient of 0.90. We note that the energy scales in the two sets of data are quite different. The range of $\Delta\Delta G_{A\beta}$ scale for β-sheet formation was pointed out[40] to vary widely in two experiments. The former obtained with the zinc finger host data,[38] shown above, exhibits a rather narrow range (0.21 kcal/mol), whereas that obtained[40] from thermal stability measurements of a variant of immunoglobin-binding domain B1 from protein G yields a $\Delta\Delta G_{A\beta}$ range of 2.05 kcal/mol. Results are reported therein relative to the free-energy changes of the alanine-substituted variant. Comparison of our results with the latter scale is displayed in Figure 9b. As expected, a significant decrease in the slope of the best fit line is observed. The best fit line, excluding the outlier Asp, is expressed as

$$\Delta E_A(\phi^+_\beta, \phi^-_\beta) - \Delta E_{ALA}(\phi^+_\beta, \phi^-_\beta) = 0.41\ \Delta\Delta G_{A\beta} - 0.11 \qquad (13)$$

The correlation coefficient decreases to 0.72, the lowest value among all examined cases. Yet, the overall correspondence between the present results and experimental observations may be considered quite satisfactory, in view of the weak correlation between the two sets of experimental data.

The role of context as a major determinant of β-sheet propensity was further emphasized by Minor and Kim.[41] $\Delta\Delta G_{A\beta}$ values for β-sheet formation at an edge β-strand are quite different from those obtained with the same technique at a central strand. Recently, a designed 11-amino acid sequence was shown to fold either as an α-helix or a β-sheet, depending on its position along the primary sequence of the IgG-binding domain of protein G.[42] Likewise, peptide sequences were shown to assume α-helical or β-strand, depending on the type of solvent.[43]

Smith et al.[44] also determined a thermodynamic scale for β-sheet tendencies, by using the B1 domain of staphylococcal IgG-binding protein G for mutations. The latter scale is similar to the one proposed by Minor and Kim.[40] Comparison with our doublet energies (not shown) yields a relationship close to Eq. (12), with slope 0.33 and correlation coefficient 0.75. Asp is excluded from the least squares fit calculation in both cases. The free-energy change for Asp in β-sheet is reported to be remarkably high in conformity with the result of Minor and Kim,[40] whereas no such strong unfavorable effect was found by Kim and Berg[38] or in the present analysis. The participation of Asp in a β-sheet is possible only by a distortion of the bond angle ϕ^+ by approximately 30° relative to that in the regular structure, as indicated in the footnote of Table II, and that this deformation may not be accommodated in some closely packed tertiary contexts.

From the above discussion, the coupling between virtual bond torsions emerges as an important feature affecting the residue type hierarchic order for both α-helix- and β-sheet-forming propensities. It is not hard to understand that the coupled rotations of virtual bonds i and i + 1 involve the interaction of residues i − 1 and i + 2 along the backbone and thus intrinsically include the effect of main-chain hydrogen bonds stabilizing α-helices. Inasmuch as each virtual bond is representative of three real bonds, the pairwise interdependence of virtual bonds here includes the coupling of bonds extending up to 12 real bonds along the chain and thus reflects the relatively long correlations underlying particular

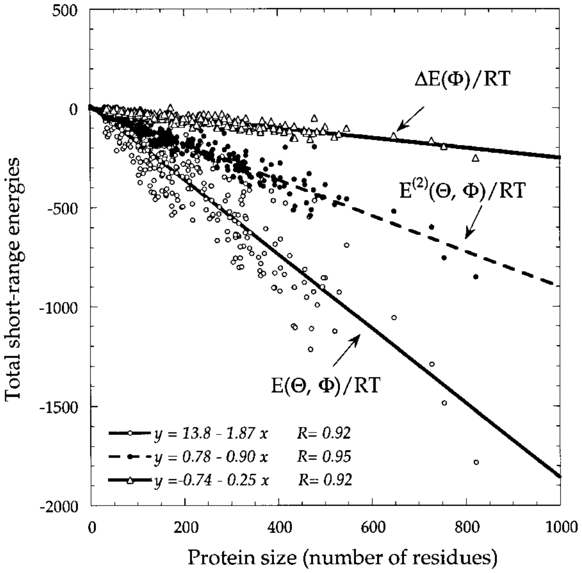

Fig. 10: Contribution of short-range interactions of various types to the conformational energy $E(\Theta, \Phi)$ of the protein, as a function of the protein size. The open circles represent the results for previously listed[3] 302 PDB structures. A contribution of -1.87 RT per residue is found from the best fit line (correlation coefficient $R = 0.92$). Filled circles represent the second-order contribution $E^{(2)}(\Theta, \Phi)$. The least squares fit yields $E^{(2)}(\Theta, \Phi) = -0.90$ nRT, with $R = 0.95$. The open triangles show the contribution of the coupling between bond torsions, $\Delta E(\Phi)$. Best fit line yields $\Delta E(\Phi) = -0.25$ nRT, $R = 0.92$.

structures. Not only α-helices but also β-sheet propensities are well accounted for with this range of interactions. Thus, although β-sheets are stabilized by nonbonded neighbors in the first place, the short-range cooperative effects nonetheless do complement effectively the longer range nonbonded interactions.

Contribution of Short-Range Interactions to Stability: Detection of Structures Having Nonnative-Like Conformational Energies

The dependence of the total short-range energies $E(\Theta, \Phi)$ of the 302 PDB structures on the sizes of the proteins is shown in Figure 10. The second-order contributions, $E^{(2)}(\Theta, \Phi)$, are shown separately, by the filled circles. The upper triangles represent the contribution of the coupling between bond torsions, alone, defined as $\Delta E(\Phi) \equiv \Sigma_{i=3}^{n-2} \Delta E_i(\phi^-, \phi^+)$. The equations of the lines drawn by linear regression are shown in Figure 10. The average contribution per residue to the total short-range conformational energy is -1.87 RT from the slope of the best fit line through the $E(\Theta, \Phi)$ values. It is interesting to observe that approximately half of this, -0.90 RT, comes from the second-order contributions. The coupling between bond angles and torsions appears to be stronger than that between pairs of consecutive torsional angles, which may be verified from the slope of the upper line, -0.25 RT, reflecting the coupling between bond torsions only.

The structures that exhibit the least favorable short-range energies in our data set are listed in Table III. All structures having $E(\Theta, \Phi)$/nRT weaker than -0.8 are given in the table, in order of increasing short-range energy per residue, displayed in the third column. The corresponding number of residues and the first-order contribution per residue are tabulated in columns 2 and 4. On the other hand, proteins whose short-range energetics appear to be unusually favorable ($E(\Theta, \Phi)/\text{nRT} \leq -2.88$) compared with other native structures are listed in Table IV, along with their structural characteristics. The resolutions of these structures are markedly higher than those listed in Table III. Another interesting observation is the usual absence of disulfide bridges in these structures, whose short-range energetics are highly favorable and apparently do not require further stabilization by S-S bridges. Among the structures listed in Table IV, only the uteroglobin dimer (2utg) has disulfide bridges (two bridges linking the monomers), whereas disulfide bridges are more frequent in structures with relatively weak short-range energies (see Table III). One could argue that some internal strains may be required to achieve the formation of disulfide bonds, which are manifested by less favorable conformational energies in general, whereas structures devoid of these constraints assume more probable angles and torsions, leading to favorable apparent short-range energetics.

Threading Experiments: Utility of Combining Short-Range and Long-Range Interactions

In parallel with our recent test of the performance of nonbonded potentials,[2] inverse folding experiments are performed here for the present short-range potentials, as well as for a combination of short-range and long-range (nonbonded) potentials. A *structure-recognizes-sequence* protocol is applied, i.e., sequences taken from PDB are mounted on a given three-dimensional structure, following the original work of Hendlich et al.[45] to identify the sequence that yields the lowest energy with a given fold. Only the backbone, or more precisely the C^α atoms, are considered for evaluating the short-range energies, whereas nonbonded energies required[2] the inclusion of the side-chain interaction sites. The same set of 62 known structures and 32 additional PDB structures previously selected for threading experiments[2,45] are taken to permit comparison with previous results.

Results are presented in Table V. The first three columns give the PDB identifiers of the 62 reference structures, their sizes, and the total number of variants threaded on each of them. Here, "variants" refer to sequence fragments of all larger size proteins, obtained by advancing one residue at a time without permitting gaps or insertions, considering the complete set of 94 proteins. Obviously, the number of variants increases with decreasing size of the reference structure. The rank of the native sequence

TABLE III. Proteins Exhibiting Weak Short-Range Conformational Energies

Protein (PDB code)	n	$\frac{E(\Theta, \Phi)}{nRT}$	$\frac{E^{(1)}(\Theta, \Phi)}{nRT}$	n (S-S)*	Resolution (Å)
4rcrh	236	−0.74	−0.25	—	2.8
3fxc	98	−0.74	−0.38	—	2.5
1sh1	48	−0.73	0.01	3	NMR
3dpa	217	−0.72	−0.29	—	2.5
1tabi	61	−0.70	−0.21	4	2.3
1hcc	59	−0.67	−0.26	2	NMR
2hhrc	204	−0.66	−0.31	—	2.8
1mona	43	−0.60	−0.07	—	2.75
1pdc	41	−0.61	−0.20	—	NMR
4tgf	49	−0.60	−0.43	3	NMR
1bds	43	−0.60	−0.15	3	NMR
7wga	171	−0.60	−0.57	16	2.0
1cy3	118	−0.59	−0.08	—	2.5
2mev4	57	−0.54	−0.25	1	3.0
2mrt	30	−0.50	−0.70	—	NMR
1mhu	31	−0.47	−0.37	—	NMR
1aaf	54	−0.31	−0.39	—	NMR
1egf	52	−0.28	−0.16	—	NMR
2bpa3	35	−0.28	−0.16	—	3.4
2gn5	87	−0.16	0.71	—	2.3

*Number of disulfide bridges.

TABLE IV. Proteins Exhibiting Strong Short-Range Conformational Energies

Protein (PDB code)	n	$\frac{E(\Theta, \Phi)}{nRT}$	$\frac{E^{(1)}(\Theta, \Phi)}{nRT}$	Res (Å)
1atf	37	−4.02	−2.62	NMR
2zta	32	−4.00	−2.39	1.8
1rop	56	−3.89	−2.35	1.7
256b	106	−3.57	−2.13	1.4
1ltsc	40	−3.35	−1.85	1.95
1le4	138	−3.33	−2.13	2.5
1troa	103	−3.28	−2.01	1.9
1bbha	130	−3.25	−1.83	1.8
1coh	141	−3.19	−1.80	2.9
1ecd	136	−3.15	−1.78	1.4
1cpcl	173	−3.11	−1.75	1.66
2ccy	127	−3.11	−1.81	1.67
1babb	145	−3.08	−1.79	1.5
2utg	70	−2.99	−1.70	1.64
1lmb	92	−2.98	−1.79	1.8
2hmz	113	−2.94	−1.73	1.66
1fha	179	−2.93	−1.71	2.4
1wrp	106	−2.91	−1.74	2.2
1ppt	36	−2.88	−1.63	1.37

among all variants, classified on the basis of short-range conformational energy $E(\Theta, \Phi)$ given by Eq. (5) is presented in the fourth column. Fifty of 62 structures correctly recognize the native sequence.

The short-range energy $E(\Theta, \Phi)$ of the native sequence-structure is given in the seventh column. $\Delta E/nRT$ for short-range interactions, listed in the eighth column, is the difference between the short-range energy obtained with the native sequence and that of the variant yielding the lowest conformational energy. If the structure correctly recognizes the native sequence, $\Delta E/nRT$ is negative; otherwise, it is positive. And the energy gap $\Delta E/nRT$ indicates how well the native sequence is distinguished. The 12 cases that yield a positive ΔE are smaller proteins, in general, in conformity with the results from threading experiments performed with nonbonded potentials,[2] but these are not necessarily the structures that failed the threading test on the basis of long-range interactions.

Next, we explore the effect of considering both short-range and long-range interactions on the performance of threading experiments. The potentials between nonbonded side chains and/or backbone interaction sites[2] are now taken into consideration, in addition to short-range energies. This improves the recognition of native sequences to 55 of 62 structures. In fact, a compensating mechanism between short-range and long-range interactions is observed (Table V), which optimizes the stability of the native structures. For example, among the above-mentioned 12 cases that fail to recognize the native sequence on the basis of short-range energies alone, seven folds (2ca2, 1mbn, 1alc, 2gn5, 1gps, 1atx, and 1cbn) are subject to highly favorable long-range interactions upon threading of the correct sequence,[2] which more than counterbalance the adverse short-range effects and bring the native sequence to the first rank in the energy-sorted list (column 6). This latter rank is obtained for each structure by sorting the variants in the order of increasing *total* energy,

TABLE V. Results From Threading Experiments

PDB code*†	n	Total number of variants	Rank (short-range)	Rank (long-range)	Rank (total)‡	$\frac{E(\Theta, \Phi)}{nRT}$	ΔE/nRT (short-range)	ΔE/nRT (total)
1rhd	292	1548	1	1	1	−1.50	−0.22	−3.82
1pyp	280	1873	1	1	1	−0.97	−0.17	−1.72
1dri	270	2155	1	1	1	−2.48	−0.19	−1.95
1aaib	261	2419	1	1	1	−1.17	−0.26	−2.66
1dnka	259	2452	1	1	1	−1.80	−0.23	−2.22
1caj	258	2570	1	1	1	−1.52	−0.36	−2.64
2ca2(*)	256	2581	2	1	1	−1.52	+0.03	−0.29
1baa	242	3057	1	1	1	−1.56	−0.13	−2.40
3pgm	229	3540	1	1	1	−1.04	−0.12	−1.01
2cla	213	4032	1	1	1	−1.94	−0.07	−0.43
1bbt2	209	4341	1	1	1	−1.30	−0.21	−1.66
1abma	197	4871	1	1	1	−2.51	−0.17	−1.66
3adk	193	5008	1	1	1	−2.58	−0.15	−1.19
1gky	185	5429	1	1	1	−2.50	−0.21	−1.25
1cpca	173	6011	1	1	1	−2.87	−0.01	−3.53
1cpcl	173	6011	1	1	1	−3.12	−0.09	−1.04
1cd4	172	6063	1	1	1	−1.15	−0.22	−1.20
2fcr	172	6063	1	1	1	−1.88	−0.14	−1.99
5p21	165	6433	1	1	1	−2.11	−0.11	−1.45
1l84	161	6665	1	1	1	−2.53	−0.05	−1.12
3dfr	161	6665	1	1	1	−1.72	−0.12	−1.56
5tnc	160	6711	1	1	1	−2.76	−0.08	−0.32
1mbn(*)	152	7172	14	1	1	−2.26	+0.05	−0.67
1lh3	152	7172	1	1	1	−2.74	−0.05	−0.83
1f3g	149	7333	1	1	1	−1.27	−0.27	−1.84
1aak	149	7333	1	1	1	−1.78	−0.61	−1.19
4cln(*)	147	7475	1	4	1	−2.11	−0.12	−0.10
1mba	146	7537	1	1	1	−3.14	−0.01	−1.12
1fx1	146	7537	1	1	1	−1.97	−0.14	−1.81
1babb	145	7601	1	1	1	−3.07	−0.05	−1.77
1barb	137	8116	1	7	1	−1.40	−0.21	−0.22
1end	136	8181	1	1	1	−2.69	−0.07	−2.69
1eco	135	8250	1	1	1	−3.14	−0.02	−0.05
2snm	134	8245	1	1	1	−1.84	−0.18	−0.89
1bbha	130	8595	1	1	1	−3.24	−0.06	−1.00
1ifb	130	8595	1	1	1	−1.92	−0.35	−0.84
1lhm	129	8666	1	1	1	−1.81	−0.08	−1.10
1bw4	124	9024	1	1	1	−1.16	−0.13	−1.16
4p2p	123	9097	1	1	1	−2.21	−0.09	−0.74
1alc(*)	121	9245	4	1	1	−1.32	+0.03	−0.62
1paz	119	9391	1	1	1	−1.71	−0.18	−1.46
1cy3(*)	117	9541	1	22	1	−0.59	−0.27	−0.40
1cd8	113	9844	1	1	1	−1.35	−0.16	−1.38
2ssi	106	10381	1	1	1	−1.34	−0.11	−1.81
1acx	106	10381	1	1	1	−0.80	−0.24	−1.44
1fkf	106	10381	1	1	1	−1.63	−0.15	−1.49
1fdd	105	10468	1	1	1	−1.64	−0.16	−1.02
1aps	97	11142	1	1	1	−0.47	−0.002	−0.58
1ten	89	11711	1	1	1	−1.45	−0.17	−0.66
2gn5(*)	86	12603	12	1	1	−0.16	+0.06	−1.12
1c5a(*)	65	13736	4428	65	60	−2.15	+0.34	+0.67
1nxb	61	14066	1	348	3(**)	−1.02	−0.38	+0.23
1aaf	54	14649	1	1	1	−0.31	−0.02	−0.35
1egf(*)	52	14819	25	19	3	−0.28	+0.10	+0.09
1gps(*)	46	15331	72	1	1	−1.41	+0.12	−0.12
1atx(*)	45	15418	17	1	1	−0.92	+0.06	−0.10
1cbn(*)	45	15418	1236	1	1	−1.68	+0.37	−0.36
1pdc	41	15774	1	1	1	−0.58	−0.32	−0.85
2bpa3	35	16311	1	4684	16(**)	−0.27	−0.002	+0.18
1bba(*)	35	16311	10	178	6	−1.83	+0.18	+0.42
2mrt(*)	29	16864	161	11	2	−0.50	+0.89	+0.01
2mhu	29	16864	4587	4	84(**)	−0.26	+0.17	+0.79

*A synergistic compensating effect between short- and long-range potentials occurs in the proteins marked by an asterisk (in the first column) such that the corresponding rank on the basis of total potentials (column 9) is improved compared with those obtained by either short- or long-range interactions only.

†The fifth letter in the PDB code refers to the selected structure in the data bank file, when more than one structure or subunit is present.

‡The three cases that became worse than either the short-range or long-range ranking are marked with (**) in column 6.

including both the residue-specific long-range potentials[2] and the present short-range energies. The corresponding energy gap (last column) is in the range $\Delta E/nRT \leq -1.0$ in general, which is suggestive of a high confidence level. Conversely, three proteins (4cln, 1barb, and 1cy3) whose ranks were found[2] to be >1 on the basis of long-range interactions, are now observed to recognize the correct sequence upon inclusion of short-range preferences. These and other examples for which the simultaneous consideration of short- and long-range energies leads to an improvement in the overall ranking are indicated by asterisks in the first column of Table V.

The only protein folds that cannot recognize the correct sequence by using the total energy are 1c5a, 1nxb, 1egf, 2bpa3, 1bba, 2mrt, and 2mhu. All of these are small proteins ($n \leq 65$). We note that some of these structures might owe their stability to disulfide bridges (1nxb, 1c5a) or binding of metals (2mrt, 2mhu). S-S bonds or ligand coordination, which are not considered in the threading tests, may play a critical role in determining the correct sequence-structure matches. We also note that most of these structures (1c5a, 1egf, 1bba, 2mrt, and 2mhu) are determined by NMR. 2bpa3 is determined by X-ray but at relatively poor (3.4 Å) resolution.

A more severe test of the discriminating capabilities of the short-range potentials may be performed by avoiding being penalized for the unusual conformations assigned to residues exhibiting the most distinctive distributions of angles. For example, Gly and Pro are distinguished by their unique distributions of angles, as may be verified from Figures 2, 3, 5, 6, and 7. Their contributions to the recognition of the correct sequence-structure pairs on the basis of short-range energies are therefore expected to be significant. Calculations repeated by omitting the contributions of Gly and Pro to overall short-range energies showed that the performance of threading experiments based on short-range energies alone decreased indeed to some extent. Specifically, 12 proteins (1cpca, 1lh3, 1mba, 1babb, 1end, 1eco, 1bbha, 2ssi, 1fkf, 1aps, 1aaf, and 2bpa3) among the 50 structures whose native sequence were correctly identified in the original calculations failed to recognize their native sequence upon omission of the contributions of Gly and Pro. However, in all cases, except 1eco, the native sequence-structure pair was correctly identified on the basis of total (short-range and long-range) potentials, the contribution of the favorable long-range potential dominating the adverse short-range effects.

In sum, the protein folds can be classified into five categories, depending on their response to the two types of threading experiments:

(1) Proteins that rank in the first position, *both* in short-range and in long-range energy evaluations and thus could correctly identify their native sequence by using either test. Forty-five of 62 reference proteins conform with this behavior. These may be viewed as a set of proteins whose intramolecular interactions conform closely with those of typical globular proteins, both on *local* and *global* scales. The native sequence is correctly identified among alternative primary structures. Proteins belonging to this category are typically relatively large. The smallest protein belonging to this set is the collagen-binding b-domain of the bovine seminal fluid protein PDC-109 (1pdc), which is composed of 41 residues. It is noticeable that the native sequence-structure identification is performed correctly among 15,774 variants for this protein.

(2) Proteins that rank in the first position with respect to the *total* intramolecular potential despite the unfavorable short-range conformational energies. These are stabilized by side group-side group (S-S) and side group-backbone (S-B) interactions between amino acids that are at least three residues apart along the backbone. Seven proteins are observed in this class: carbonic anhydrase (2ca2), α-lactalbumin (1alc), myoglobin (1mbn), γ1-p thionin (1gps), sea anemone toxin (1atx), gene5/DNA-binding protein (2gn5), and crambin (1cbn). For example, crambin ranks in the 1236th position among 15,418 variants on the basis of short-range interactions alone, but is brought to the first position by the compensating effect of strong attractive interactions between nonbonded units.

(3) Proteins whose correct structure-sequence identification is achieved by the contribution of short-range interaction energies, despite relatively unfavorable S-S and S-B interactions. This category contrasts the preceding one, in that stability is imparted by torsional and bond angle potentials essentially. Among the 62 reference proteins, 3, calmodulin (4cln), cytochrome C_3 (1cy3), and acidic fibroblast growth factor mutant (1barb) exhibit this behavior. Calmodulin is not globular but is formed by two lobes connected by a long helical segment. It is therefore natural that nonbonded attractions, which are normally satisfied in compact globular structures, are not so strong in this conformation. Yet, its short-range interactions were strong enough to lead to the lowest *total* energy among the 7,475 variants.

(4) Proteins whose rank is improved by a compensating effect between short-range and long-range interaction potentials. Together with those correctly discriminating the correct sequence for the known fold, a total of 13 proteins, denoted by an asterisk in the first column of Table V, exhibit this behavior.

(5) Notably, only three proteins are ranked worse by their total energy than by either short-range or long-range components alone. These are indicated by the double asterisk on the sixth column.

It is interesting to speculate that proteins in these different classes might exhibit substantially different behavior in their folding pathways.

CONCLUSIONS

In this and two recent studies on nonbonded interactions in globular proteins,[2,3] a model consisting of two sites per residue, one on the backbone and the other on the side group, is exploited as a mathematically simple, yet physically adequate, tool for characterizing the structural preferences of proteins. Long-range and short-range interactions are analyzed in a self-consistent way in these studies. Such an analysis is important because various proteins appear to have somewhat different balances between these two contributions.

The virtual C^{α}-C^{α} bond model describes satisfactorily the secondary structure geometry of the backbone. The number of parameters is two per residue, (θ, ϕ), just as for the classical (ϕ, ψ) representation, and so the substitution occurs without loss of accuracy in the position and energetic preference of α-carbons. *Gauche*$^{-}$ and *trans* states of virtual bonds, near 60° and 210°, respectively, characterize the α-helical and β-sheet structures. The bond angle θ is also highly sensitive to the secondary structure; it has a bimodal distribution with peaks around 90° and 120°, corresponding to α-helices and β-sheets, respectively. Thus, α-helices and β-sheet structures are well characterized in the two-dimensional plots of coupled geometric variables. The plots in Figure 4, obtained from all residues, thus reflect the intrinsic conformational preference of the polypeptide backbone. Some particular amino acids exhibit substantial departure from the collective behavior, as illustrated in Figures 5–7.

The major difference between present α-helix and β-sheet propensities and those obtained by directly counting α-helical and β-strand states for the 20 residue types is the consideration of pairwise interdependence of residue conformational states. The second-order contributions in Tables I and II are, in fact, the major differences between our results and those found from the observation of individual residues. The interdependence of bond torsions, as well as the couplings between bond angles and torsions in the virtual bond model, are found to contribute significantly to short-range conformational energetics. These second-order interaction energies therefore should be taken into consideration for evaluating the conformational energy of virtual bond models used in coarse-grained simulations.

Figures 8 and 9, and several other examples cited in the section on secondary structure propensities, show that the doublet torsional energies $E_A(\phi^{+}_{\alpha}, \phi^{-}_{\alpha})$ and $E_A(\phi^{+}_{\beta}, \phi^{-}_{\beta})$ listed in Tables I and II provide a quantitative measure of the helix-forming and β-sheet-forming tendencies of amino acids and may be used for a preliminary estimation of the energy changes involved in mutating residues. It is interesting to observe the existence of a strong correlation between the present residue-specific potentials $E_A(\phi^{+}_{\alpha}, \phi^{-}_{\alpha})$ for α-helix formation, and the thermodynamic data analysis by others, including the conformational entropy decreases of side chains. The β-sheet-forming tendencies, on the other hand, exhibit a relatively lower correlation with experiments for two major reasons. First, these are predominantly stabilized by long-range interactions and therefore are only partially accounted for by local conformational energies. Second, the β-sheet propensities strongly depend on the tertiary context, as emphasized in several studies.[38–42,46]

Attempts to combine residues into representative groups should be undertaken with caution, inasmuch as each amino acid possesses distinct geometry and energy characteristics, either on a local or on a global scale. On a local scale, as illustrated in Figures 2, 3, and 5–7, the classification of residues as α-helix-formers, β-sheet formers or turn formers seems inappropriate with regard to the distinct distributions of dihedral angles and bond angles obeyed by residues assigned in the same class. On a global scale, size and shape effects play a dominant role in determining the side-chain coordination geometry.[3] This precludes the classification of residues on the basis of their hydrophobic or polar character, exclusively.

The residue-specific conformational potentials for the virtual bond torsions and bond angles are shown to discriminate correct sequence-structure matches in more than three fourths of the test structures. This invites attention to the important role of the backbone short-range preferences in selecting and stabilizing native folds.

The utility of combining long-range and short-range energy contributions is shown by the improvement achieved in sequence-structure recognition reported in our last table for inverse folding experiments. Among the 62 protein folds considered in threading experiments, 13 exhibited a compensation between short-range and long-range energetics, leading to a high ranking ($r \leq 6$) of the native sequence. Thus, an essential observation is that S-S and S-B non-bonded potentials (long-range interactions) and residue-specific backbone conformational energies (short-range interactions) complement each other in important ways to impart a sufficient stability to native structures.

ACKNOWLEDGMENT

Partial support by NATO CRG Project 951240 is gratefully acknowledged.

REFERENCES

1. Flory, P.J. “Statistical Mechanics of Chain Molecules.” New York: Interscience, 1969. (also reprinted by Hanser Publishers, Oxford University, Oxford, 1988.)
2. Bahar, I., Jernigan, R.L. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. J. Mol. Biol. 266:195–214, 1997.
3. Bahar, I., Jernigan, R.L. Coordination geometry of nonbonded residues in globular proteins. Folding Design 1:357–370, 1996.
4. Miyazawa, S., Jernigan, R.L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J. Mol. Biol. 256:623–644, 1996.
5. Honig, B., Cohen, F.E. Adding backbone to protein folding: Why proteins are polypeptides. Folding Design 1:R17–R20, 1996.
6. Kocher, J.-P.A., Rooman, M.J., Wodak, S. Factors influencing the ability of knowledge-based potentials to identify native sequence-structure matches. J. Mol. Biol. 235:1598–1613, 1994.
7. Park, B.H., Levitt, M. Energy functions that discriminate X-ray and near-native folds from well-constructed decoys. J. Mol. Biol. 258:367–392, 1996.
8. Niefind, K., Schomburg, D. Amino acid similarity coefficients for protein modeling and sequence alignment derived from main-chain folding angles. J. Mol. Biol. 219:481–497, 1991.
9. Sun, S. Reduced representation model of protein structure prediction: Statistical potential and genetic algorithms. Protein Sci. 2:762–785, 1993.
10. Nishikawa, K., Matsuo, Y. Development of pseudoenergy potentials for assessing protein 3d-1d compatibility and detecting weak homologies. Protein Eng. 6:811–820, 1993.
11. Levitt, M. A simplified representation of protein conformations for rapid simulation of protein folding. J. Mol. Biol. 104:59–107, 1976.
12. Gregoret, L.M., Cohen, F.E. Protein folding: Effect of packing density on chain conformation. J. Mol. Biol. 219: 109–122, 1991.
13. DeWitte, R.S., Shakhnovich, E.I. Pseudodihedrals: Simplified protein backbone representation with knowledge-based energy. Protein Sci. 3:1570–1581, 1994.
14. Park, B.H., Levitt, M. The complexity and accuracy of discrete state models of protein structure. J. Mol. Biol. 249:493–507, 1995.
15. Brant, D.A., Flory, P.J. The configuration of random polypeptide II theory. J. Am. Chem. Soc. 87:2791–2800, 1965.
16. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F.J., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
17. Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F., Weng, J. Protein Data Bank. In: “Crystallographic Databases-Information Content Software Systems, Scientific Applications.” Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F., Weng, J. (eds.). Bonn, Cambridge, and Chester: Data Commission of the International Union of Crystallography, 1987:107.
18. Chou, P.Y., Fasman, G.D. Prediction of the secondary structure of protein from amino acid sequence. Adv. Enzymol. 47:45–148, 1978.
19. Garnier, J., Osguthorpe, D., Robson, B. An analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120:97–120, 1978.
20. Levitt, M. Conformational preferences of amino acids in globular proteins. Biochemistry 17:4277–4285, 1978.
21. Blundell, T.L., Sibanba, B.L., Sternberg, M.J.E., Thornton, J.M. Knowledge-based prediction of protein structures and design of novel molecules. Nature 326:347–352, 1987.
22. Serrano, L., Sancho, J., Hirshberg, M., Fersht, A.R. α-Helix stability in proteins. I. Empirical correlations concerning substitution of side chains at the N and C-caps and the replacement of alanine by glycine or serine at solvent-exposed surfaces. J. Mol. Biol. 227:544–559, 1992.
23. Richardson, J., Richardson, D. Amino acid preferences for specific locations at the end of α-helices. Science 240:1648–1652, 1988.
24. Muñoz, V., Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters. Struct. Biol. 1:399–409, 1994.
25. Aurora, R., Srinivasan, R., Rose, G.D. Rules for α-helix termination by glycine. Science 264:1126–1129, 1994.
26. Lyu, C.P., Liff, M.I., Marky, L.A., Kallenbach, N.R. Side chain contributions to the stability of alpha-helical structure in peptides. Science 250:669–673, 1990.
27. Chou, P.Y., Fasman, G.D. In: “Peptides.” Chou, P.Y., Fasman, G.D. (eds.). New York: Halsted Press, 1977:284–287.
28. Chou, P.Y. Prediction of protein structural classes from amino acid compositions. In: “Prediction of Protein Structures and the Principles of Protein Conformation.” Chou, P.Y. (eds.). Plenum Press, New York: 1989:549–586.
29. O’Neil, T.K., DeGrado, W. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. Science 250:646–651, 1990.
30. Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T.M., Baldwin, R.L. Relative helix-forming tendencies of nonpolar amino acids. Nature 344:268–270, 1990.
31. Horovitz, A., Matthews, J.M., Fersht, A.R. α-Helix stability in Proteins II. Factors that influence stability at an internal position. J. Mol. Biol. 227:560–568, 1992.
32. Blaber, M., Zhang, X.-J., Matthews, B.W. Structural basis of amino acid α helix propensity. Science 260:1637–1640, 1993.
33. Luque, I., Mayorga, O.L., Freire, E. Structure-based thermodynamic scale of α-helix propensities in amino acids. Biochemistry 35:13681–13688, 1996.
34. Creamer, T.P., Rose, G.D. Side-chain entropy opposes α-helix formation but rationalizes experimentally determined helix-forming propensities. Proc. Natl. Acad. Sci. USA 89:5937–5941, 1992.
35. Merutka, G., Stellwagen, E. Positional independence and additivity of amino acid replacements on helix stability in monomeric peptides. Biochemistry 29:894–898, 1990.
36. Hecht, M.H., Sturtevant, J.M., Sauer, R.T. Stabilization of λ repressor against thermal denaturation by site directed Gly → Ala changes in α-helix 3. Proteins 1:43–46, 1986.
37. Dao-Pin, S.T., Baase, B.W., Matthews, B.W. A mutant T4 Lysozyme (Val131Ala) designed to increase thermostability by reduction of strain within an α-helix. Proteins 7:198–204, 1990.
38. Kim, C.A., Berg, J.M. Thermodynamic β-sheet propensities measured using a zinc-finger host peptide. Nature 362:267–270, 1993.
39. Bai, Y., Englander, S.W. Hydrogen bond strength and β-sheet propensities: The role of a side chain blocking effect. Proteins 18:262–266, 1994.
40. Minor, D.L., Kim, P.S. Measurement of the β-sheet propensities of amino acids. Nature 367:660–663, 1994.
41. Minor, D.L., Kim, P.S. Context is a major determinant of β-sheet propensity. Nature 371:264–267, 1994.
42. Minor, D.L., Kim, P.S. Context-dependent secondary structure formation of a designed protein sequence. Nature 380:730–734, 1996.
43. Waterhous, D.V., Johnson, W.C. Importance of environment in determining secondary structure in proteins. Biochemistry 33:2121–2128, 1994.
44. Smith, C.K., Withka, J.M., Regan, L. A thermodynamic scale for the β-sheet forming tendencies of the amino acids. Biochemistry 33:5510–5517, 1994.
45. Hendlich, M., Lackner, P., Weitckus, S., Floeckner, H., Froschauer, R., Gottsbacher, K., Casari, G., Sippl, M.J. Identification of native protein folds amongst a large number of incorrect models: The calculation of low energy conformations from potentials of mean force. J. Mol. Biol. 216:167–180, 1990.
46. Otzen, D.E., Fersht, A.R. Side chain determinants of β-sheet stability. Biochemistry 17:5718–5724, 1995.