

Packing of sidechains in low-resolution models for proteins

Ozlem Keskin and Ivet Bahar

Background: Atomic level rotamer libraries for sidechains in proteins have been proposed by several groups. Conformations of side groups in coarse-grained models, on the other hand, have not yet been analyzed, although low resolution approaches are the only efficient way to explore global structural features.

Results: A residue-specific backbone-dependent library for sidechain isomers, compatible with a coarse-grained model, is proposed. The isomeric states are utilized in packing sidechains of known backbone structures. Sidechain positions are predicted with a root-mean-square deviation (rmsd) of 2.40 Å with respect to crystal structure for 50 test proteins. The rmsd for core residues is 1.60 Å and decreases to 1.35 Å when conformational correlations and directional effects in inter-residue couplings are considered.

Conclusions: An automated method for assigning sidechain positions in coarse-grained model proteins is proposed and made available on the internet; the method accounts satisfactorily for sidechain packing, particularly in the core.

Introduction

The suitable packing of sidechains to maximize hydrophobic contacts and optimize specific inter-residue interactions is important in the overall stability of folded proteins, along with the affinity of the backbone to form hydrogen bonds and therefore assume regular structures. For operational purposes, the prediction of the three-dimensional structure of proteins from knowledge of amino acid sequence can be seen as consisting of two processes: the determination of the backbone structure, and the packing of sidechains. These two processes are closely coupled, because of the importance of an optimal sidechain packing for the stability of the backbone fold. As with the larger problem of protein folding, the principal difficulty in sidechain packing is to locate the sidechain conformational states leading to the lowest overall energy of interaction.

Recent studies have demonstrated that the sidechains in protein cores can be generated from knowledge of backbone coordinates with root-mean-square deviations (rmsds) of 1.5–2.0 Å relative to their native state positions [1–20], as summarized in Table 1. An atomic description of protein structure is adopted in all these studies. On the other hand, the utility of coarse-grained models and simulations has been pointed out in numerous studies [21,22], starting from the original work of Levitt and Warshel [23]. Despite the abundance of knowledge-based modeling and simulations of proteins, an automated approach for predicting sidechain positions in coarse-grained models with (approximately) known backbone structure has not been devised but is the aim of the present study.

Addresses: Chemical Engineering Department and Polymer Research Center, Bogazici University, and TUBITAK Advanced Polymeric Materials Research Center, Bebek 80815, Istanbul, Turkey.

Correspondence: Ivet Bahar
E-mail: bahar@prc.bme.boun.edu.tr

Key words: hydrophobic core, sidechain isomers, sidechain packing, specific interactions, virtual bond model

Received: 21 May 1998
Revisions requested: 17 June 1998
Revisions received: 14 July 1998
Accepted: 21 August 1998

Published: 11 November 1998
<http://biomednet.com/elecref/1359027800300469>

Folding & Design 09 November 1998, 3:469–479

© Current Biology Ltd ISSN 1359-0278

A study in that direction could serve several purposes. First, it would shed light on the adequacy of estimating sidechain conformation on the basis of a knowledge of backbone fold, exclusively, an approach that relies on the predominance of native backbone fold for determining sidechain packing, as opposed to the prevalence of interior packing for determining backbone architecture. Studies emphasizing the dominant role of backbone fold, and the plasticity or flexibility of sidechain–sidechain interactions exist in the literature (e.g. [24,25]), as well as others pointing out, on the contrary, the important role of sidechain packing in tertiary structure determination [26], in accord with the general premise of Ponder and Richards [13]. Also included in the second group are studies showing the regularity of sidechain–sidechain coordination [27] or the adaptability of the backbone atoms to accommodate required sidechain–sidechain contacts in the core of proteins [28]. An examination of the accuracy of the sidechain positions predicted with a coarse-grained description of the backbone, and an assessment of the sensitivity of sidechain conformations towards shifts in backbone coordinates will be carried out here, which should provide an insight into the interplay between backbone and sidechain preferences. Second, the ability of residue-specific knowledge-based potentials associated with a low resolution model to predict sidechain interaction sites in a coarse-grained model will be tested. The influence of the choice of potentials and the improvements brought about by adopting interaction potentials of increasing complexity will be searched. A broad variety of potentials, ranging from hard-core potentials to residue-specific distance and orientation-dependent potentials will be considered for an

Table 1

Studies on sidechain (S) construction using backbone (B) coordinates.

Reference	Using the coordinates of:	Predicting	rmsd (Å) core (all)	No. of proteins studied
Vasquez [1]	B and C ^β atoms	S atoms	1.14 (1.78)	30
Lee and Subbiah [2]	B and C ^β atoms	S atoms	1.25 (1.77)	9
Shenkin <i>et al.</i> [3]	B and C ^β atoms	S atoms	1.54 (2.03)	49
Eisenmenger <i>et al.</i> [6]	B and other C ^β	C ^β atoms	0.90 (1.48)	6
Kono and Doi [7]	B and C ^β atoms	S atoms	1.10 (1.73)	21
Tuffery <i>et al.</i> [8]	B and C ^β atoms	S atoms	1.54 (1.84)	14
Mathiowetz and Goddard [9]	C ^α atoms	All atoms	2.31	3
Holm and Sander [10]	C ^α atoms	All atoms	1.60 (1.91)	14
Rey and Skolnick [11]	C ^α atoms	Other B atoms	0.7	6
Levitt [15]	C ^α atoms	All other atoms	1.57 (1.78)	8
Koehl and Delarue [16]	B and C ^β atoms	S atoms	1.38 (1.89)	30
Laughton [17]	B and C ^β atoms	S atoms	1.00 (1.71)	8
Wilson <i>et al.</i> [18]	B and C ^β atoms	S atoms	1.14 (1.45)	4
This work	C ^α atoms	S virtual sites	1.35 (2.28)	50

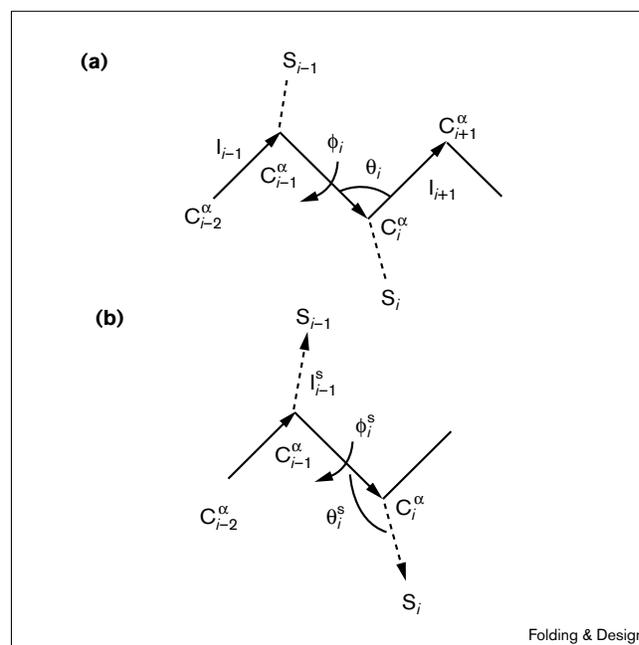
assessment of the most important factors in determining the packing of sidechains.

Our recent examination of databank structures with a low resolution model indicates that the packing of sidechains is not random, but exhibits some preferred coordination geometry and directionality [29]. There is a specificity in packing depending on the type of sidechains. This suggests that it is possible to assign a set of most probable isomeric states to sidechains. If so, which factor is more effective in selecting the isomeric conformers — the specific type of amino acid sidechain, or the geometry of the backbone near the attached sidechain? How strong is the dependence on backbone secondary structure? If there is a dependence, does sidechain packing persist irrespective of local changes in the backbone? Furthermore, is there an optimal computational algorithm for accurately packing sidechains, which might also provide some indications about the folding pathway in real proteins? In the present study, a low-resolution approach based on a two-sites-per-residue model (Figure 1), coupled with a rotational isomeric states formalism, is proposed for constructing sidechains in structures of known C^α-coordinates. The two sites defining the position of the residues are referred to as the backbone (B) and sidechain (S) sites, conveniently identified with the α carbons and sidechain functionally relevant centers, respectively [29–32].

Several sidechain rotamer libraries are described in the literature. In Ponder and Richard's library [13], for example, 17 of the 20 amino acids (omitting Met, Lys and Arg) can be represented by 67 sidechain rotamers, expressed in terms of χ angles. Later, Tuffery *et al.* [8] updated their library based on a larger database. Vasquez [1] also built a rotamer library of χ angles. A more detailed backbone-dependent rotamer library was derived by Dunbrack and

Karplus [20]. Later, Dunbrack and Cohen [14] obtained rotamer populations for the full ranges of ϕ , ψ values

Figure 1



(a) Schematic representation of the two-sites-per-residue virtual bond model. A segment between backbone units C^α_{*i*-2} and C^α_{*i*+1} is shown. The sidechain attached to the *i*th C^α is represented as S_{*i*}. I_{*i*-1} is the *i*-1th virtual bond connecting C^α_{*i*-2} and C^α_{*i*-1}. ϕ_i is the rotational angle of the *i*th virtual bond, defined by the respective locations of the four backbone units C^α_{*i*-2}, C^α_{*i*-1}, C^α_{*i*} and C^α_{*i*+1}. θ_i is the bond angle between I_{*i*} and I_{*i*+1}. (b) θ_i^s is the sidechain bond angle between I_{*i*} and I_{*i*}^s, where I_{*i*}^s is the sidechain virtual bond connecting C^α_{*i*} to S_{*i*}. The sidechain virtual bonds are shown as dashed lines. ϕ_i^s is the torsional angle around bond I_{*i*} with reference to the four consecutive sites C^α_{*i*-2}, C^α_{*i*-1}, C^α_{*i*} and S_{*i*}.

based on 518 protein chains, using Bayesian statistical analysis. These and other libraries consist of a list of discrete sidechain conformations and their probabilities are determined from their frequency of occurrence in PDB structures [33]. A similar approach is adopted here, based on consideration of several joint probabilities and combinatorial method. The major difference is that a coarse-grained model is explored, and the results are analyzed with a view to answering the questions raised above about the importance of different effects in structure formation, including the choice of energy parameters and computational procedure to rebuild sidechain positions.

This study is composed of two parts. First, the orientational preferences of sidechain virtual bonds with respect to the backbone will be analyzed and a set of sidechain isomeric states (SISs) will be determined for each type of amino acid. Probability distributions of sidechain dihedral angles, bond angles and bond lengths will be examined with this aim, using statistical methods similar to those adopted for extracting backbone conformational preferences and related pseudodihedral potentials from PDB structures [31,34]. The SISs are based on the preferences of the individual amino acids, examined separately. Backbone conformations near the examined residue only are taken into consideration. Accordingly, the SISs are determined by the so-called short-range interactions along the sequence. Second, the SISs will be tested in a series of proteins for placing sidechains given the α -carbon traces. The possible SISs will be evaluated on the basis of their non-bonded or long-range energies [30], at this stage. The one having the lowest energy will

be compared with the structure determined using X-ray crystallography or NMR. The term long-range energy refers here to interactions between residue pairs sufficiently close in space, but relatively distant (at least five intervening virtual bonds) along the chain sequence. Additionally, improvements in packing, especially in the core, brought about by considering the short-range couplings between backbone and sidechain conformational states, and the directional preferences for S-S pairings [29] will be considered for an assessment of the contribution of more complex selection criteria.

Results and discussion

Sidechain isomeric states

We constructed a SIS library using a virtual bond model consisting of two sites per residue: one on the backbone, identified with the α carbon, and the second on the amino acid sidechain, selected on the basis of the structure and the most distinctive interaction center of the specific amino acid [29,30]. In our model, l_i designates the virtual bond vector pointing from α carbon $i-1$, C^{α}_{i-1} , to C^{α}_i ; θ_i is the bond angle between l_i and l_{i+1} , and ϕ_i is the torsional angle of bond l_i (Figure 1a). The probability distributions of these variables defining the backbone degrees of freedom were analyzed previously for each residue type [30]. In the present study, we concentrate on the variables defining the positions and orientations of sidechains.

The position of the i^{th} sidechain S_i with respect to the backbone is represented by its bond vector, l_i^s , pointing from C^{α}_i to S_i , the angle θ_i^s between l_i and l_i^s , and the torsional angle ϕ_i^s defined by the consecutive four sites C^{α}_{i-2} ,

Table 2

Sidechain isomeric states.

Amino acid	Sidechain isomeric states $\{l_A^s (\text{\AA}), \theta_A^s (^\circ), \Delta\phi_A^s (^\circ)\}^*$			
	l_A^s	α Helix $\{\theta_A^s, \Delta\phi_A^s\}$	β Strand $\{\theta_A^s, \Delta\phi_A^s\}$	Other $\{\theta_A^s, \Delta\phi_A^s\}$
Ala	1.53	{125, 240}	{120, 210}	{120, 210}
Val	2.21	{130, 240}	{120, 180}, {100, 210}	{120, 180}, {100, 210}
Ile	3.87	{110, 240}	{100, 180} [*] , {120, 240}	{100, 180} [*]
Leu	3.27	{120, 210}, {150, 210}	{100, 180}, {150, 180}	{110, 180}
Ser	2.42	{110, 270}, {110, 210}, {160, 240}	{110, 270}, {140, 210}, {100, 180}	{105, 180}, {110, 270}, {160, 210}
Thr	2.41	{100, 210}, {110, 270}	{95, 180}, {110, 270}	{100, 180}, {110, 255}, {140, 210}
Asp	3.06	{110, 210}	{100, 210}, {150, 180}, {110, 270}	{100, 180}, {155, 180}, {105, 270}
Asn	3.07	{110, 210}	{95, 180}, {160, 180}, {105, 270}	{100, 180}, {100, 270}, {150, 180}
Glu	4.40	{120, 240}	{110, 210}	{110, 180} [*] , {140, 210}
Gln	4.41	{120, 240}, {140, 150} [*]	{110, 210}	{110, 210}
Lys	6.20	{110, 210}, {150, 240}	{100, 180}, {110, 210}	{100, 180}, {90, 150} [*] , {140, 180}
Arg	5.53	{110, 225}	{100, 180}, {130, 210}	{110, 210}, {100, 180}
Cys	2.80	{110, 210}, {160, 240}	{90, 180}, {160, 210}	{100, 180}, {100, 270}, {160, 210}
Met	4.07	{120, 240}, {115, 180} [*]	{105, 180} [*] , {130, 210}	{120, 180} [*] , {110, 240}
Phe	3.79	{170, 255}, {100, 210}, {90, 180}	{90, 180}	{90, 180}, {90, 270}
Tyr	4.15	{170, 300}, {170, 240}, {100, 210}	{85, 180}, {100, 270}, {160, 180}	{90, 180}, {110, 270}, {170, 240}
Trp	4.41	{100, 180}, {110, 210}, {160, 315}	{95, 150}, {110, 270}	{100, 180}, {170, 240}, {90, 270}
His	3.55	{100, 210}, {90, 180}, {170, 255}	{90, 180}, {155, 180}	{100, 270}, {100, 180}, {160, 180}
Pro	1.88	{80, 240}	{80, 240}	{80, 240}, {120, 210}

*These states assume the bond lengths l_2 , listed in column 5 of Table 3, instead of the bond lengths l_1 , presented here in the second column.

C^{α}_{i-1} , C^{α}_i , and S_i (Figure 1b). A library of residue-specific SISs consisting of the most probable l_i^s , θ_i^s and ϕ_i^s values for each type of amino acid is formed (Table 2). In view of the strong dependence of the SIS on the backbone conformation [5,14,20], results are presented separately for three different types of backbone structures — α helix, β strand and 'other'. These three subsets are classified on the basis of the dihedral angles of $C^{\alpha}-C^{\alpha}$ virtual bonds as $30^\circ \leq \phi_i \leq 90^\circ$ for α helices, $150^\circ \leq \phi_i \leq 270^\circ$ for β strands and the remainder for other.

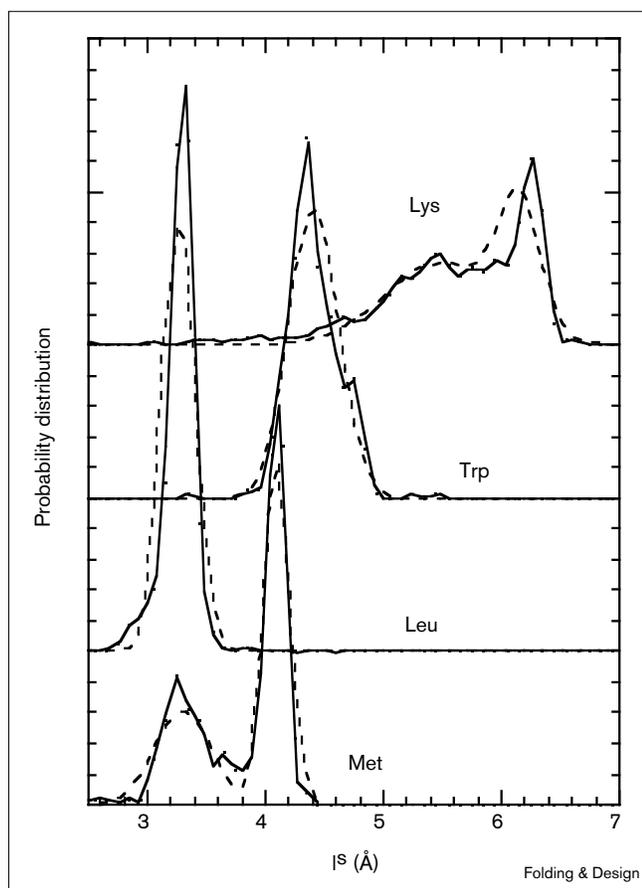
The dihedral angles of the SISs are presented in Table 2 in terms of their values relative to the torsional angle of the preceding backbone bond along the chain, that is $\Delta\phi_i^s \equiv \phi_i^s - \phi_{i-1}^s$. For a perfect tetrahedral bond, the difference $\Delta\phi_i^s$ is fixed and equal to $\pm 120^\circ$ (-120° or 240° for the D-form) [31,34]. But, in the virtual bond model, residue-specific deviations up to $\pm 90^\circ$ occur, as can be seen from Table 2. Such deviations become particularly

pronounced as a larger number of real bonds are represented by the sidechain virtual bond.

The sidechain virtual bond lengths were determined from the peak (or peaks) of the monomodal (or bimodal) Gaussian distribution curves for l_i^s . For illustrative purposes, the database-extracted normalized distributions, and the corresponding best-fitting curves for the sidechain bond lengths of a few residues, are shown in Figure 2. In cases where two peaks were observed, the corresponding bond lengths have been designated as l_1 and l_2 , the former being the more probable. The complete list of the most probable l_i^s values for all types of residues is presented in Table 3, along with their covariances and statistical weights (see the Materials and methods section). The majority of the SISs listed in Table 2 assume the bond lengths l_1 , as indicated in the second column. Those having bond lengths l_2 are marked with an asterisk.

The torsional angles and bond angles of the SISs were obtained using the grid search method, taking account of the couplings between the two degrees of freedom, $\Delta\phi_i^s$ and θ_i^s , as described in the Materials and methods section. As no coupling to geometric variables other than the two sidechain geometry parameters $\Delta\phi_i^s$ and θ_i^s were taken into consideration, the present set of SISs is referred to as that resulting from short-range first order effects. The backbone geometry is considered in a rather coarse-grained way. For illustrative purposes, we present here the maps obtained for Leu (Figure 3) and Asp (Figure 4). In these figures, the doublet probabilities $P_A(\theta^s, \Delta\phi^s)$ for the joint occurrence of the variables $(\theta^s, \Delta\phi^s)$ are shown for the three

Figure 2



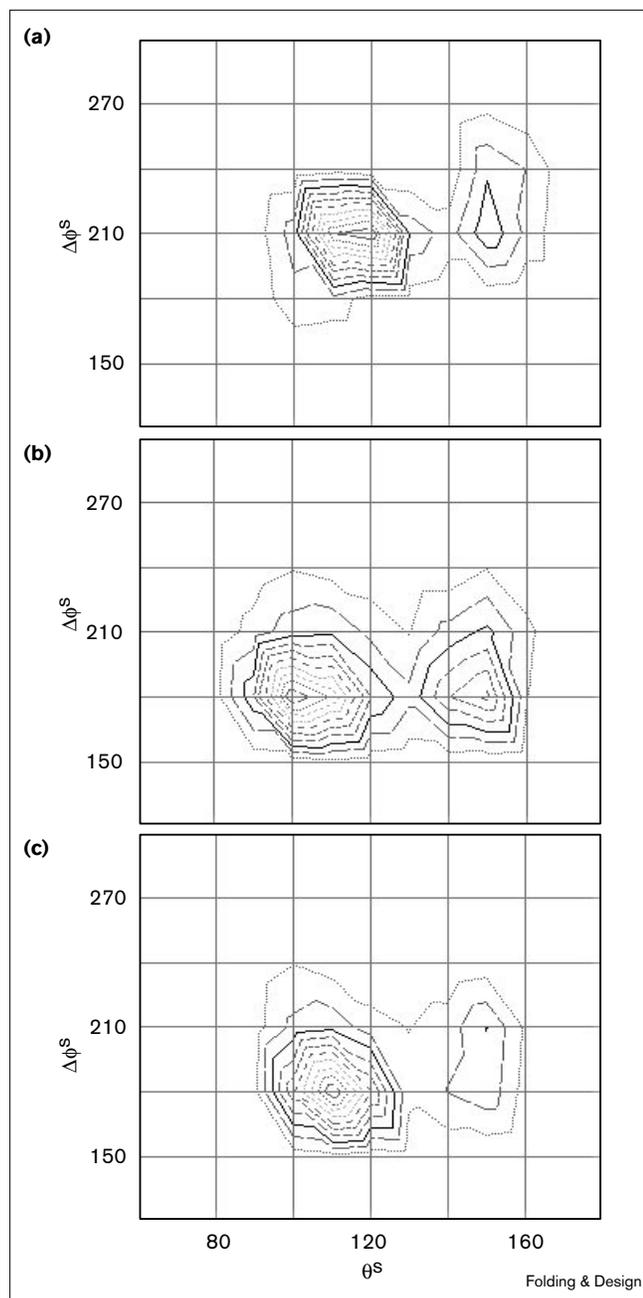
Sidechain bond length (l_A^s) distributions for $A = \text{Lys, Trp, Leu and Met}$. The continuous curves represent the results extracted from the PDB structures; the dashed curves correspond to the best-fitting unimodal or bimodal Gaussian distributions.

Table 3

Parameters for sidechain bond length distributions (see equation 2).

Residue	a_1	l_1^s (Å)	σ_1	l_2^s (Å)	σ_2
Ala	1.00	1.53	0.06	—	—
Val	1.00	2.21	0.07	—	—
Ile	0.78	3.87	0.13	3.10	0.19
Leu	1.00	3.27	0.14	—	—
Ser	1.00	2.42	0.10	—	—
Thr	1.00	2.41	0.10	—	—
Asp	1.00	3.06	0.10	—	—
Asn	1.00	3.07	0.11	—	—
Glu	0.67	4.40	0.14	3.62	0.22
Gln	0.66	4.41	0.14	3.58	0.24
Lys	0.55	6.20	0.17	5.50	0.48
Arg	1.00	5.53	0.46	—	—
Cys	1.00	2.80	0.08	—	—
Met	0.63	4.07	0.11	3.29	0.24
Phe	1.00	3.79	0.09	—	—
Tyr	1.00	4.15	0.13	—	—
Trp	1.00	4.41	0.22	—	—
His	1.00	3.55	0.12	—	—
Pro	1.00	1.88	0.07	—	—

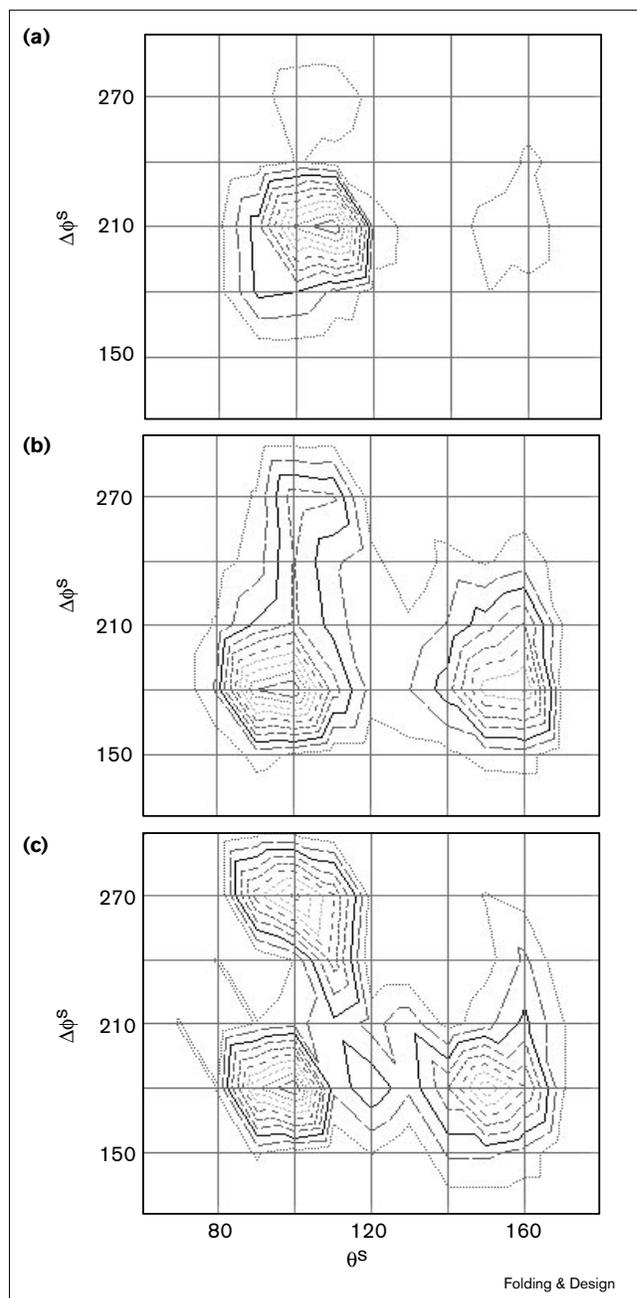
Figure 3



Probability distributions of sidechain angles θ^s and $\Delta\phi^s$ for Leu when the backbone is in (a) α -helical, (b) β -strand and (c) other structures. The innermost regions of the contours represent the highest probability states. Two sidechain isomeric states (SISs) are identified for Leu sidechains appended to α -helical backbones; the respective $\{\theta^s, \Delta\phi^s\}$ values are $\{120^\circ, 210^\circ\}$ and $\{150^\circ, 210^\circ\}$, as seen in (a). In (b), the most probable SISs in β strands are found to be $\{100^\circ, 180^\circ\}$ and $\{150^\circ, 180^\circ\}$, and in (c) only one state, $\{110^\circ, 180^\circ\}$, is distinguished.

subsets, α helix, β strand and other, of backbone structure. The contours connect $(\theta^s, \Delta\phi^s)$ loci of equal probability, the innermost regions being the most favorable isomeric states

Figure 4



Probability distributions of sidechain angles θ^s and $\Delta\phi^s$ for Asp when the backbone is in (a) α -helical, (b) β -strand and (c) other structures. See the legend for Figure 3. A larger number of isomeric states exhibiting a broader dispersion on the $\{\theta^s, \Delta\phi^s\}$ space are identified for Asp compared to those located for Leu (Figure 3).

listed in Table 2. The regions whose probabilities are at least four times larger than those expected from a uniform distribution of geometric variables are enclosed by the contours. The sidechain lengths corresponding to these isomeric states correlate readily, by a direct inspection of PDB

structures, with the peaks of the Gaussian distributions obtained for I_A^s values.

Construction of sidechains using the derived SIS

The adoption of the most probable SIS for constructing the sidechains in 50 test proteins led to the rmsds of the predicted structures presented in the Supplementary material published with this article on the internet. Where more than one isomeric state was available, the one leading to the lowest nonbonded energy was selected, after iteratively testing all accessible states for all residues. Usually three or four cycles over all sidechains were sufficient to attain the lowest-energy positions.

The nonbonded energies were calculated using the knowledge-based potentials derived by Jernigan and Bahar [30,32] for residue-specific, distance-dependent interactions between pairs of sidechains (S–S) and sidechain–backbone sites (S–B). The energies of the calculated structures were generally higher than those of the PDB structures (Supplementary material). The average nonbonded energy per residue is -1.65 RT in the predicted structures, as opposed to -2.63 RT in PDB structures. The observed higher energies are due to the fact that the SISs do not necessarily optimize the nonbonded interactions. A higher flexibility in the assignment of sidechain conformations will be shown below to decrease the nonbonded energies to values comparable to those of the PDB structures. The rmsds of the predicted structures vary in the range $1.61 \leq \text{rmsd} \leq 3.31$ Å, with a mean value of 2.40 Å.

As a further test of the validity of the SIS we focused on core residues. The term core residues is used here to describe the amino acids that have a coordination number of six or more on the basis of sidechain sites located within a spherical volume of radius 6.4 Å. The rmsds vary in the range $0.24 \leq \text{rmsd} \leq 2.79$ Å in this case, with a mean value of 1.60 Å. A significant increase in the accuracy of sidechain positions is therefore observed when attention is confined to core residues. This confirms that a plausible strategy for constructing sidechains is to start from core residues and proceed towards those subject to a lower packing density.

Short-range conformational couplings between backbone and sidechain

To try to obtain a more accurate description of sidechain packing we considered the couplings between the geometric variables θ_p , θ_i^s , ϕ_i and ϕ_i^s . We evaluated the doublet probabilities $P_A(\theta_p, \theta_i^s)$, $P_A(\theta_p, \Delta\phi_i^s)$, $P_A(\theta_p, \phi_i^s)$, $P_A(\phi_p, \theta_i^s)$, $P_A(\phi_p, \Delta\phi_i^s)$ and $P_A(\phi_p, \phi_i^s)$ for all amino acids. Such pairwise couplings between backbone and sidechain geometric variables are referred to as second-order short-range effects. We do not group the residues on the basis of three broad regions of backbone conformations (α helix,

β strand and other) anymore, but instead scan fine regions, of mesh size 10° and 30° , respectively, for θ_i and ϕ_i angles. The above doublet probability distributions for all types of amino acids are available on the internet (<http://klee.bme.boun.edu.tr>).

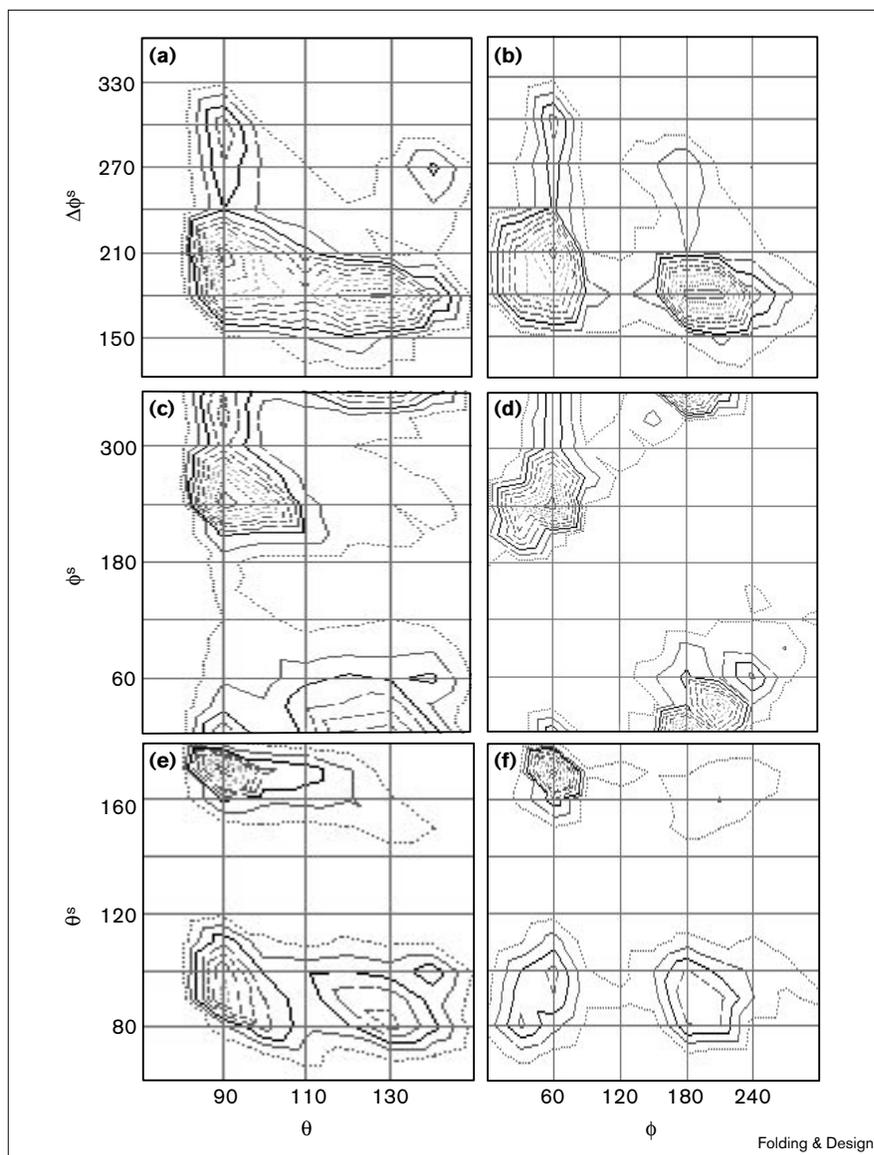
The following approach is adopted for packing sidechains given all doublet probabilities, and backbone geometric variables θ_i and ϕ_i . First, the two probability maps $P_A(\phi_p, \theta_i^s)$ and $P_A(\theta_p, \theta_i^s)$ are consulted to estimate the most probable sidechain bond angles θ_i^s , and the four maps $P_A(\theta_p, \Delta\phi_i^s)$, $P_A(\phi_p, \Delta\phi_i^s)$, $P_A(\theta_p, \phi_i^s)$ and $P_A(\phi_p, \phi_i^s)$ are examined to estimate ϕ_i^s . For each amino acid we therefore have several equilibrium states (namely combinations of most probable θ_i^s and ϕ_i^s values) extracted from different probability distributions. We observed that the native state is always one of the equilibrium states indicated in one (or more) of these maps. The problem is how to choose the correct (native) one among these.

Consider, for example, the results for tyrosine. The contour maps of $P(\theta_p, \Delta\phi_i^s)$, $P(\phi_p, \Delta\phi_i^s)$, $P(\theta_p, \phi_i^s)$, $P(\phi_p, \phi_i^s)$, $P(\theta_p, \theta_i^s)$ and $P(\phi_p, \theta_i^s)$ for Tyr are shown in Figure 5a–f, respectively. Suppose the backbone bond angle (θ) is 90° and torsional angle (ϕ) is 60° . Then, from Figure 5a we obtain $\phi^s = 60^\circ + 210^\circ$ and $60^\circ + 300^\circ$ as the most stable sidechain torsional angles; Figure 5b reproduces the same values for ϕ^s . Figure 5c indicates $\phi^s = 240^\circ$ and 360° , as does Figure 5d. We therefore have three choices — 240° , 270° and 360° — for ϕ^s . On the other hand, we obtain $\theta^s = 90^\circ$ and 170° from Figure 5e, and 100° and 170° from Figure 5f, yielding a total of three distinct θ^s values (90° , 100° and 170°). All nine combinations of the angles (θ^s , ϕ^s) are tested for the examined residue and the one that yields the lowest nonbonded energy is selected. It should be noted that the angle pair selected at this stage can be modified at the next iteration, after assignment of all sidechain positions in the particular protein. Three or four cycles over all residues are generally sufficient for convergence to the lowest-energy isomers.

Consideration of the conformational couplings between backbone and sidechain conformations using the above combinatorial approach led to the following results. The rmsd values for the set of 50 test proteins are not significantly affected; the average rmsd values of sidechains in the core and all sidechains are 1.52 Å and 2.28 Å, respectively, values that are quite close to those obtained above from a crude consideration of the backbone (1.60 and 2.40 Å). The detailed consideration of the backbone geometry — apart from a broad classification as α helix, β strand and other — does not appear to have a significant effect on the accuracy of sidechain packing geometry. The refinements in sidechain positions are interestingly found to have a dramatic effect on stability, however. The total nonbonded potential decreases, on average, to -3.01 RT

Figure 5

Contour maps showing the doublet probability distributions for the joint occurrence of (a) θ_i and $\Delta\phi_i^s$, (b) ϕ_i and $\Delta\phi_i^s$, (c) θ_i and ϕ_i^s , (d) ϕ_i and ϕ_i^s , (e) θ_i and θ_i^s , and (f) ϕ_i and θ_i^s for Tyr. The first four yield the equilibrium values for ϕ_i^s given the backbone variables θ_i and ϕ_i . The last two maps yield the most probable θ_i^s values, to be used in a combinatorial analysis with the ϕ_i^s extracted from (a–d).



per residue, which is even below that calculated for native structures (-2.63 RT).

A significant increase in stability from -1.65 RT to -3.01 RT is therefore obtainable upon consideration of the detailed coupling between sidechain and backbone conformations, and the sidechain coordinates in the predicted structure deviate by about 2 \AA from the known coordinates despite attaining an intramolecular potential even more favorable than that of the native structure. The latter point indicates that consideration of distance dependence of S–S and S–B interactions alone might be insufficient for finding the correct sidechain packing geometry, because a lower energy state than the native one is obtained although the average rmsd is still 1.6 \AA in the

core. An additional effect, that of specificity in S–S couplings, will be considered next as a possible source of improvement. It will be shown, in fact, that the consideration of the preferred (database-extracted [29]) coordination geometry of S–S pairs affects the choice of the optimal packing, and leads to structures in which the distance-dependent nonbonded energies might be slightly higher, whereas the overall potentials (including both distance and direction dependence) are minimized.

Directional effects in S–S interactions

Our recent investigation of angular preferences of long-range interactions showed that S–S or S–B contacts can be selective, some coordination geometries being enhanced by a factor of ten or more relative to a random association

[29]. Recent lattice studies also indicated a regularity in the packing of sidechains [27]. In view of these observations, the potentials of mean force associated with different coordination geometries of pairs of sidechains, available on the internet (<http://klee.bme.boun.edu.tr>), were also taken into account, together with the distance-dependent S–S and S–B potentials considered above. The results are presented in the Supplementary material. The average rmsd for sidechains in the core is 1.35 Å; the average rmsd of solvent-exposed sidechains remains unchanged (2.28 Å). This is consistent with the fact that the drive for optimal packing of the sidechains is more pronounced for buried, tightly packed residues than for residues that have sufficient conformational freedom on the surface or at loosely packed regions.

A value of -2.60 RT is obtained for the average energy per residue after selecting the conformations leading the most favorable directional effects. This value includes the distance-dependent S–S and S–B potentials only, these being examined separately for comparison with the results obtained above. We note that this energy is higher than its counterpart (-3.01 RT) achieved at a previous stage, suggesting that conformations appearing to be less favorable on the basis of distance-dependent S–S and S–B energetics alone might become more favorable when we also consider directional effects. The fact that a lower rmsd with respect to native state coordinates is obtained for core residues suggests strongly that such corrections in sidechain packing are operative in the final folded structures. The detailed list for all test proteins is presented in the Supplementary material.

Core packing: how is it affected by different choices of potentials?

In the interest of exploring the accuracy level reached by adopting different choices of potentials, we first considered the simplest case, that of a hard-core repulsion between sidechain pairs closer than a critical distance. A separation of 2.0 Å is adopted as the closest distance of approach. S–S interactions are therefore either repulsive or equal to zero. Sidechain dihedral angles are varied at 30° intervals in their full range, whereas sidechain bond angles are assigned values of $90^\circ \pm 20^\circ$, at 10° intervals. An iterative scheme over all sidechains is again adopted, until the lowest-energy conformation is reached. These analyses yielded an average rmsd of 3.96 Å for core residues, compared to their native counterparts. The average rmsd for all sidechains was 5.24 Å.

The adoption of a hard-core repulsion without an attractive part indicates, in a sense, the uppermost limit of error incurred in sidechain construction, with a knowledge of the backbone, in the present low-resolution model. A more realistic approach, also including the attractive potentials typical of S–S interactions in folded proteins, is

to consider the so-called homogeneous potentials derived from PDB structures for all S–S pairs, irrespective of residue type [30,32]. These potentials reflect the generic behavior of all amino acids in protein-like folded structures. A strong repulsion starting near 2.0 Å is observed, which is similar to the hard-core potential adopted above, but there are additionally two energy minima, near 5 and 10 Å, typical of the centers of the successive coordination shells around a central residue, which account for the attractive potentials favoring S–S contacts. The use of these potentials with $\theta^s = 90^\circ \pm 20^\circ$ and $\Delta\phi^s = 210^\circ \pm 50^\circ$ along with the most probable residue-specific bond lengths (column 2 in Table 2) lead to an average intramolecular potential of -1.29 RT per residue, which is considerably weaker than the one (-2.63 RT) stabilizing the correctly folded structure. The mean rmsd value becomes 3.02 Å for all sidechains, and 2.35 Å for those in the core, that is, significantly better than those obtained with a purely repulsive hard-core potential, but not as good as that obtained with residue-specific SIS.

The average rmsd of core residues decreases from 3.96 Å to 2.35 Å, and then successively to 1.64 Å, 1.52 Å, and 1.35 Å, as the following respective approximations are used: repulsive hard-core; homogeneous potentials with both attractive and repulsive parts; residue-specific S–S and S–B potentials selecting the most probable state amongst database-extracted SISs; same as the preceding case, but with a more detailed consideration of backbone–sidechain couplings; and, finally, including directional effects in S–S pairings in addition to the preceding case.

As a final test, the effect of adopting another set of residue-specific S–S potentials on the third stage described above was checked. The same procedure is repeated with the van der Waals type potentials proposed by Park and Levitt [35]. These are distance-dependent versions of Miyazawa–Jernigan contact energies [36]. The average rmsd is 1.76 Å for core sidechains and 2.67 Å for all sidechains. These values are comparable to those obtained with the presently adopted [30] potentials. The slight increase in rmsd is understandable in view of the approximation involved in postulating the distance dependence of the contact potentials. A more detailed examination of the efficiency of different energy functions for discriminating native-like folds can be found elsewhere [35].

Sensitivity of sidechain packing to shifts in backbone coordinates

The influence of slight shifts in backbone coordinates on the packing of sidechains was investigated by introducing uniformly distributed random errors in all C^α coordinates. Our calculations indicate that the changes in the C^α positions are not directly reflected on sidechain coordinates, provided that these remain within a certain limit (~ 1.5 Å rmsds). Thus, small uncertainties in backbone positions

appear to be tolerated by sidechains at the expense of relatively smaller size displacements. Figure 6 illustrates the average rms changes in sidechain positions in the cores relative to their optimal packing positions, as a function of perturbations imposed on backbone coordinates.

It is interesting to observe that the distortions induced in sidechain positions as a result of the shifts in C^α positions are considerably smaller than the shifts in the backbone coordinates. For example, a 1.0 Å rms change in backbone coordinates alters the sidechain positions by less than 0.5 Å on average. The accompanying change in the overall nonbonded interaction energy is also small, as illustrated in the inset of Figure 6, in which the average energies per residue, expressed in RT units, are shown for successive shifts in backbone coordinates. Essentially, an average increase of about 0.2 RT per residue is observed in S–S and S–B interactions, as the C^α atoms are perturbed by 2.0 Å. Interestingly, both the rmsds and energy values exhibit an inflection near 1–1.5 Å, suggesting that beyond a certain threshold, say 1.5 Å, the accommodating rearrangements of the sidechains become increasingly large in size and in energy requirement. This behavior is reminiscent of the previous modeling of protein conformations by automatic segment matching developed by Levitt [15], in which an average all-atom rmsd of 1.78 Å was observed using known C^α coordinates and results were almost insensitive to C^α errors of up to 1 Å [15].

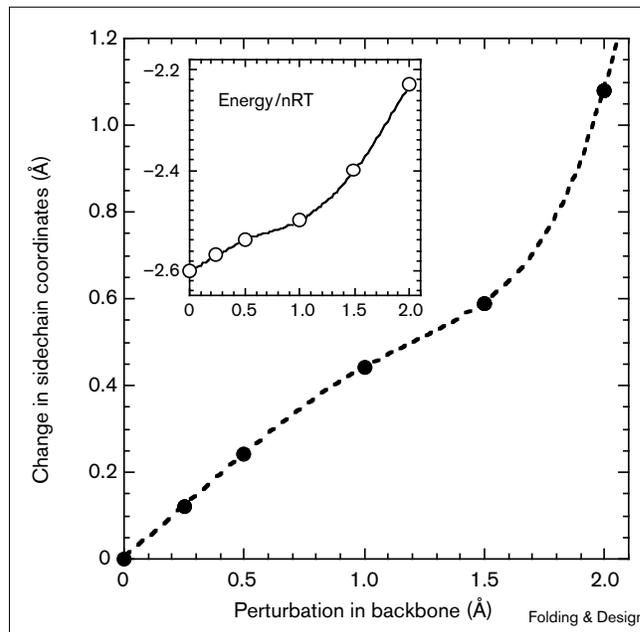
Conclusions

The results obtained here show a level of accuracy comparable to a number of studies in the literature (Table 1). The major difference is that a low-resolution model is explored here, as opposed to the atomic level approach used in those studies. Thus, sidechain sites can be constructed with an accuracy level of the order of atomic analyses, provided that inter-residue energetics is carefully evaluated. The library of residue-specific sidechain isomers for the present virtual bond model, the doublet probabilities for the couplings between backbone and sidechain conformations, and the Fortran code for determining the optimal packing, are available on the internet (<http://klee.bme.boun.edu.tr>), along with the input files required for the evaluation of conformational energetics.

An interesting observation was the significant improvement in the quality of sidechain packing as more and more details in both short-range and long-range inter-residue couplings were taken into consideration. Our analysis shows that consideration of effects such as the directionality and specificity in sidechain–sidechain packing improves the prediction of core residues, whereas it has a relatively small effect on surface residues.

The quality of packing was examined here in terms of two properties: the rmsd of sidechain sites from their

Figure 6



Change in sidechain coordinates as a function of shifts in C^α positions. The changes in sidechain coordinates are expressed relative to those attained with known backbone structures. Results refer to the average behavior of sidechains in the core regions of the set of 50 proteins presently considered. The inset shows the accompanying change in the overall nonbonded potential of the predicted structures, expressed in RT units, per residue. A threshold around 1.5 Å is distinguishable beyond which shifts in C^α positions require increasingly larger size (and higher energy) rearrangements in the sidechains.

counterparts in the crystal structures, and the total nonbonded potential of the predicted structure, determined by a summation over all S–S and S–B pairs. Although some structures might occasionally be energetically favorable from the point of view of certain interactions, the attainment of the lowest rms structures requires consideration of all effects, which might increase the potential energy associated with a given type of interaction at the expense of lowering the overall potential. For example, the contribution of the distance-dependent S–S and S–B energies to the equilibrium structures was -3.01 RT before the consideration of directional effects in S–S pairing, but raised to -2.60 RT after minimization of the overall potential. An increase in the distance-dependent interactions might therefore be afforded if a larger decrease in the overall potential is obtainable by optimizing the coordination angles. The lower rmsd attained in the core (an average of 1.35 Å as opposed to 1.60 Å) indicates the utility of considering the coordination geometry.

The successive stages of sidechain rebuilding and energy minimization for finding the optimal packing give an insight into the role of various factors such as hard-core

repulsion or residue specificity during structure formation. For example, in the simplest approach of a hard-core potential with a 2.0 Å critical separation, sidechain positions exhibit an average rmsd of 5.24 Å (and 3.96 Å in the core) with respect to native structures. This value is decreased to 3.02 Å (and 2.35 Å in the core) upon consideration of the so-called homogeneous inter-residue potentials, calculated [30] from PDB structures as a generic interaction between all S–S pairs in protein-like folded structures, irrespective of amino acid type. Thus, without any knowledge of any residue specificity, it is possible to predict the position of sidechain interaction sites with an accuracy of 2.35 Å. The predicted structure is not sufficiently stable, however, as deduced from the nonbonded potential of –1.29 RT per residue on average, which can be compared to the native structure value of –2.63 RT.

If the specific types of sidechain are taken into consideration, both on a short range (for identifying the SIS) and on a long range (for choosing the most probable SIS in a particular tertiary context), again there is an appreciable improvement in rmsd values (2.40 Å and 1.60 Å). The energetics appear to be relatively insensitive at first (–1.65 RT). Slight readjustments, based on the couplings between backbone and sidechain conformational states and on the directional preferences of S–S pairs, lower the energy to –2.60 RT, however. We note that the average energy of the native structures was –2.63 RT per residue in the present low-resolution model, in perfect agreement with the final value reached in the predicted structure. The lowest rmsd values are 2.28 Å and 1.35 Å for all residues and core residues, respectively.

The final packing geometry attained in the core by the presently adopted combinatorial analysis and iterative scheme appears to be rather robust relative to small shifts in backbone coordinates, as can be seen in Figure 6. This suggests that there is a more or less well-defined packing in the core of proteins that is controlled by specific sidechain interactions and directionality, which is tolerant to small changes, up to 1.5 Å, in the backbone geometry. Conversely, the backbone fold is commonly preserved in many point mutations, despite the slight changes in sidechain associations, which draws attention to the adaptability of overall protein structures to accommodate local structural perturbations without a substantial decrease in stability.

Materials and methods

Dataset

A set of 302 nonhomologous protein structures from the PDB were analyzed to identify the most probable conformational states of sidechain sites in our two-sites-per-residue model. The proteins used in these statistical calculations are listed on the internet [28]. 50 proteins were used for testing the accuracy of the proposed library and sidechain packing method. The PDB codes of these proteins are given in the Supplementary material.

Method

Sidechain virtual bond lengths. The distributions of the virtual B–S bonds for each type of residue were fitted as Gaussian curves; most residues show one peak, while some longer residues show two peaks. The probability distribution for a given residue (A) sidechain length is thus expressed as the bimodal distribution:

$$P_A(l^s) = a_{1A}P(l_{1A}, \sigma_{1A}) + a_{2A}P(l_{2A}, \sigma_{2A}) \quad (1)$$

where a_{iA} is the coefficient accounting for the fractional contribution of the i^{th} ($i = 1$ and 2) peak such that $a_{1A} + a_{2A} = 1$, and $P(l_{iA}, \sigma_{iA})$ is the normal distribution given by:

$$P(l_{iA}, \sigma_{iA}) = \frac{1}{\sqrt{2\pi\sigma_{iA}^2}} \exp\left\{-\frac{(l^s - l_{iA})^2}{2\sigma_{iA}^2}\right\} \quad (2)$$

For illustrative purposes, results for Leu, Met, Trp and Lys are shown in Figure 2. The B–S virtual bond length parameters for all residues are presented in Table 3.

Grid search method and doublet probabilities. Sidechain torsional angle space was divided into 30° intervals, and sidechain bond angles into 10° grids for each type A of residue. The doublet probability for visiting a given grid $\{\theta_i^s, \phi_j^s\}$ was found from the associated numbers N_A of observations using:

$$P_A(\theta_i^s, \phi_j^s) = N_A(\theta_i^s, \phi_j^s) / \sum_i \sum_j N_A(\theta_i^s, \phi_j^s) \quad (3)$$

Here, $P_A(\theta_i^s, \phi_j^s)$ is the joint probability of observing residue type A in the i^{th} torsional angle interval and j^{th} bond angle interval, and the two summations in the denominator are performed over all grids accessible to θ_i^s and ϕ_j^s .

Supplementary material

An additional table, showing the comparison of predicted and crystal structure coordinates, is available as Supplementary material published with this article on the internet.

References

- Vasquez, M. (1995). An evaluation of discrete and continuum search techniques for conformational analysis of sidechains in proteins. *Biopolymers* **36**, 53-70.
- Lee, C. & Subbiah, S. (1991). Prediction of side-chain conformation by packing optimization. *J. Mol. Biol.* **217**, 373-388.
- Shenkin, P.S., Farid, H. & Fetrow, J.S. (1996). Prediction and evaluation of side-chain conformations for protein backbone structures. *Proteins* **26**, 323-352.
- Levitt, M., Gerstein, M., Huang, E., Subbiah, S. & Tsai, J. (1997). Protein folding: the endgame. *Annu. Rev. Biochem.* **66**, 549-579.
- Gregoret, L.M. & Cohen, F.E. (1990). Novel method for the rapid evaluation of packing in protein structures. *J. Mol. Biol.* **211**, 959-974.
- Eisenmenger, F., Argos, P. & Abagyan, R. (1993). A method to configure protein side-chains from the main-chain trace in homology modelling. *J. Mol. Biol.* **231**, 849-860.
- Kono, H. & Doi, J. (1996). A new method for sidechain conformation prediction using a Hopfield network and reproduced rotamers. *J. Comp. Chem.* **17**, 1667-1683.
- Tuffery, P., Etchebest, C., Hazout, S. & Lavery, R. (1991). A new approach to the rapid determination of protein sidechain conformations. *J. Biomol. Struct. Dynam.* **8**, 1267-1289.
- Mathiowetz, A.M. & Goddard, W.A. (1995). Building proteins from C^α coordinates using the dihedral probability grid Monte Carlo method. *Protein Sci.* **4**, 1217-1232.
- Holm, L. & Sander, C. (1991). Database algorithm for generating protein backbone and sidechain coordinates from a C^α trace. *J. Mol. Biol.* **218**, 183-194.
- Rey, A. & Skolnick, J. (1992). Efficient algorithm for the reconstruction of a protein backbone from the α -carbon coordinates. *J. Comp. Chem.* **13**, 443-456.

12. Mitchell, J.B.O., Laskowski, R.A. & Thornton, J.M. (1997). Non-randomness in side-chain packing: the distribution of interplanar angles. *Proteins* **29**, 370-380.
13. Ponder, J.W. & Richards, F.M. (1987). Tertiary templates for proteins: use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **193**, 775-791.
14. Dunbrack, R.L. & Cohen, F.E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6**, 1661-1681.
15. Levitt, M. (1992). Accurate modeling of protein conformation by automatic segment matching. *J. Mol. Biol.* **226**, 507-533.
16. Koehl, P. & Delarue, M. (1994). Application of a self-consistent mean field theory to predict protein sidechains conformation and estimate their conformational entropy. *J. Mol. Biol.* **239**, 249-275.
17. Laughton, C.A. (1994). Prediction of protein side-chain conformations from local three-dimensional homology relationships. *J. Mol. Biol.* **235**, 1088-1097.
18. Wilson, C., Gregoret, L.M. & Agard, D.A. (1993). Modelling side-chain conformation for homologous proteins using an energy-based rotamer search. *J. Mol. Biol.* **229**, 996-1006.
19. Summers, N.L., Carlson, W.D. & Karplus, M. (1987). Analysis of side-chain orientations in homologous proteins. *J. Mol. Biol.* **196**, 175-198.
20. Dunbrack, R.L. & Karplus, M. (1993). Backbone-dependent rotamer library for proteins, application to sidechain prediction. *J. Mol. Biol.* **230**, 543-574.
21. Jernigan, R.L. (1992). Protein folds. *Curr. Opin. Struct. Biol.* **2**, 248-256.
22. Park, B.H. & Levitt, M. (1995). The complexity and accuracy of discrete state models of protein structure. *J. Mol. Biol.* **249**, 493-507.
23. Levitt, M. & Warshel, A. (1975). Computer simulation of protein folding. *Nature* **253**, 694-698.
24. Behe, M.J., Lattman, E.E.I & Rose, G.D. (1991). The protein folding problem: the native fold determines packing, but does packing determines the native fold? *Proc. Natl Acad. Sci. USA* **88**, 4195-4199.
25. Bromberg, S. & Dill, K.A. (1994). Sidechain entropy and packing in proteins. *Protein Sci.* **3**, 997-1009.
26. Harbury, P.B., Tidor, B. & Kim, P.S. (1995). Repacking protein cores with backbone freedom: Structure prediction for coiled coils. *Proc. Natl Acad. Sci. USA* **92**, 8408-8412.
27. Raghunathan, G. & Jernigan, R.L. (1997). Ideal architecture of residue packing and its observation in protein structures. *Protein Sci.* **6**, 2072-2083.
28. Baldwin, E.P., Hajiseyedjavadi, O., Baase, W.A. & Matthews, B.W. (1993). The role of backbone flexibility in the accommodation of variants that repack the core of T4 lysozyme. *Science* **262**, 1715-1717.
29. Bahar, I. & Jernigan, R.L. (1996). Coordination geometry of nonbonded residues in globular proteins. *Fold. Des.* **1**, 357-370.
30. Bahar, I. & Jernigan, R.L. (1997). Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.* **266**, 195-214.
31. Bahar, I., Kaplan, M. & Jernigan, R.L. (1997). Short-range conformational energies, secondary structure propensities, and recognition of correct sequence-structure matches. *Proteins* **29**, 292-308.
32. Jernigan, R.L. & Bahar I. (1996). Structure-derived potentials and protein simulations. *Curr. Opin. Struct. Biol.* **6**, 195-209.
33. Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. & Weng, J. (1987). In *Crystallographic Databases-Information Content Software Systems, Scientific Applications* (Allen, F.H., Bergerhoff, G. & Sievers, R., eds), pp. 107-132. Data Commission of the International Union of Crystallography, Bonn, Cambridge and Chester.
34. DeWitte, R.S. & Shakhnovich, E.I. (1994). Pseudodihedrals: simplified protein backbone representation with knowledge-based energy. *Protein Sci.* **3**, 1570-1581.
35. Park, B. & Levitt, M. (1996). Energy functions that discriminate X-ray and near-native folds from well-constructed decoys. *J. Mol. Biol.* **258**, 367-392.
36. Miyazawa, S. & Jernigan, R.L. (1985). Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* **18**, 534-552.

Because *Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed. The paper can be accessed from <http://biomednet.com/cbiology/fad> – for further information, see the explanation on the contents pages.

Packing of sidechains in low-resolution models for proteins

Ozlem Keskin and Ivet Bahar

Folding & Design 11 November 1998, **3:469–479****Table S1**

The first two columns are the PDB code and size (total number of residues) of the structures whose α -carbon coordinates have been used in calculations. Column 3 lists the corresponding numbers of core residues. Column 4 contains the nonbonded potential of mean force (or free energy) of the known structure reported in RT units, per residue basis. Its counterpart for the simulated structure — in which the sidechain positions are assigned from the most probable isomeric states (Table 2) — is presented in column 5. The rmsds of the predicted state presented in column 6 vary in the range $1.61 \leq \text{rmsd} \leq 3.31 \text{ \AA}$, with a mean value of 2.40 \AA . The rmsds of core residues, presented in column 7, vary in the range $0.24 \leq \text{rmsd} \leq 2.79 \text{ \AA}$, with a mean value of 1.60 \AA .

Table S1

Sidechain packing results: comparison of predicted and crystal structure coordinates.*

PDB name	Number of residues	Number of core residues	Energy of PDB structure (/nRT)	First order effects [†]			Coupling and directional [‡]	
				Energy of predicted state (/nRT)	rmsd of predicted state (Å)	rmsd of core residues (Å)	Energy of predicted state (/nRT)	rmsd of core residues (Å)
1crn	46	7	-1.78	-1.41	1.76	1.86	-1.77	1.02
1ubq	76	4	-2.81	-1.87	2.48	1.09	-3.00	1.59
1tho	108	15	-3.65	-2.43	2.24	1.49	-3.59	1.06
1sn3	65	18	-1.96	-0.83	2.60	1.43	-2.05	1.29
1aaf	55	4	-0.91	-0.22	3.27	1.42	-0.91	1.84
1hoe	74	21	-2.64	-2.00	2.51	1.29	-2.88	0.80
2act	218	60	-3.67	-2.17	2.42	1.67	-3.24	1.53
4ait	74	22	-2.57	-1.74	2.06	1.52	-2.87	1.43
1coh	141	28	-3.50	-3.31	1.85	1.19	-3.78	0.63
451c	82	18	0.53	-1.65	2.33	0.24	-0.36	1.60
1ecd	136	19	-3.35	-2.30	2.13	0.61	-3.20	0.73
1r69	63	6	-2.77	-0.94	2.66	0.71	-2.32	0.38
8pti	56	11	-1.98	-0.54	2.59	2.59	-1.56	2.42
1ycc	103	14	-2.44	-1.91	2.60	1.67	-2.39	0.89
1hcc	57	12	-2.18	-1.20	2.42	1.95	-2.00	1.86
1gst	217	23	-3.29	-1.95	2.47	1.65	-3.19	1.77
2hmz	111	4	-3.01	-1.88	2.45	0.37	-2.78	1.95
2mrt	28	3	-1.85	-0.82	1.61	1.94	-0.53	1.63
9wgaa	171	67	-1.79	-0.71	2.44	2.14	-1.95	1.22
3gap	208	26	-3.04	-1.94	2.44	1.54	-3.02	0.82
3trx	105	13	-2.85	-2.24	2.60	1.34	-3.33	1.17
5cpv	108	21	-3.11	-2.25	2.29	1.98	-3.17	1.71
256b	106	19	-2.89	-2.14	2.44	0.88	-2.98	0.52
1gmpa	96	15	-2.78	-1.31	2.08	1.94	-2.70	1.45
1hgeb	175	18	-1.71	-0.71	2.55	2.26	-1.75	1.38
1bbha	131	26	-2.89	-2.11	2.07	1.15	-2.96	0.67
1ctaa	34	3	-0.91	-0.87	2.66	2.79	-1.17	1.47
1babb	146	27	-3.54	-3.01	2.23	1.45	-3.67	1.11
1abma	198	36	-3.39	-1.88	2.19	1.39	-3.14	1.40
3adk	195	37	-3.05	-2.62	2.35	1.72	-3.19	1.55
1rpra	63	9	-1.65	-1.35	2.49	0.94	-1.64	1.16
1hila	217	24	-2.79	-1.48	2.50	2.06	-2.78	1.51
2aza	129	34	-3.17	-1.23	2.37	1.91	-2.76	1.33
2fcr	173	56	-3.72	-2.47	2.29	1.83	-3.84	1.65
4p2p	124	25	-2.46	-1.34	2.54	1.93	-2.46	1.43
3dfr	162	26	-3.36	-2.04	2.52	1.36	-3.11	1.38
1hiva	99	15	-2.96	-2.18	2.46	1.91	-2.97	1.62
1lmb	92	15	-2.82	-1.90	2.19	0.97	-2.73	1.24
1cd4	173	31	-2.88	0.30	2.64	2.08	-2.95	1.18
1atx	46	10	-1.63	-0.51	2.22	1.47	-1.29	1.51
1gps	47	10	-1.61	-0.75	3.13	2.23	-1.64	1.48
3zmf	30	2	-1.12	-0.44	3.31	1.56	-1.13	1.53
1cms	323	84	-3.82	-2.35	2.40	1.93	-3.88	1.77
1cpl	165	54	-3.32	-1.93	2.67	2.04	-3.23	1.47
1ezm	298	96	-3.89	-2.58	2.19	1.65	-3.70	1.63
1ifc	132	10	-2.38	-1.69	2.23	0.92	-2.51	0.71
1lh3	153	23	-3.27	-2.11	2.12	1.01	-3.07	1.05
1rro	108	15	-3.65	-1.96	2.41	2.37	-3.60	1.27
1sar	96	17	-2.84	-1.31	2.16	1.87	-2.74	1.64
1bw4	125	26	-2.86	-1.99	2.53	2.09	-2.75	1.93
Average			-2.63	-1.65	2.40	1.60	-2.60	1.35

*Energies are calculated from nonbonded S-S and S-B interaction potentials [30]. [†]First order effects take account of the coupling between $\Delta\phi^s$ and $\Delta\theta^s$. See Table 2 for accessible SIS. [‡]Coupling

effects refer to the pairwise interdependences between the backbone and sidechain geometric variables. Directional effects refer to preferences in coordination angles between sidechains [29].