Rotamer libraries in the 21st century

Roland L Dunbrack Jr

Rotamer libraries are widely used in protein structure prediction, protein design and structure refinement. As the size of the structure database has increased rapidly in recent years, it has become possible to derive well-refined rotamer libraries using strict criteria for data inclusion and to study the dependence of rotamer populations and dihedral angles on local structural features.

Addresses
Institute for Cancer Research, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, Pennsylvania 19111, USA; e-mail: RL_Dunbrack@fccc.edu

Current Opinion in Structural Biology 2002, 12:431–440

0959-440X/02/$ – see front matter

Introduction

Ever since the first few crystal structures of proteins were determined, there has been significant study of the conformations of sidechains. Indeed, it was immediately obvious from the first protein structures that, for most χ angles, protein sidechains adopt primarily staggered dihedral angles well known to organic chemists since the 1930s [1]. As the number of solved structures increased, it was possible to define the most common sidechain conformations by statistical analysis. Along with an understanding of backbone conformations in the form of the Ramachandran map, knowledge of observed sidechain conformations has enabled better refinement of experimentally determined structures and enhanced protein structure prediction and protein design, all of which have blossomed in recent years.

It is worthwhile settling on a few definitions. A rotamer, short for 'rotational isomer', is a single sidechain conformation represented as a set of values, one for each dihedral angle degree of freedom. Because bond angles and bond lengths in proteins have rather small variances, they are usually not included in the definition of a rotamer. A rotamer library is a collection of rotamers for each residue type. Rotamer libraries usually contain information about both the conformation and the frequency of a certain conformation. Often, libraries will also contain information concerning the variance about dihedral angle means or modes, which can be used in sampling.

Sidechain dihedral angles are not evenly distributed, but, for most χ angles, occur in tight clusters around certain values. Rotamer libraries therefore are usually derived from statistical analysis of sidechain conformations in known structures of proteins by clustering observed conformations or by dividing dihedral angle space into bins and determining an average conformation in each bin. This division is usually based on physical-chemical grounds, as in the division of rotation about sp^3–sp^3 bonds into three 120° bins centered on each staggered conformation (60°, 180°, –60°).

A rotamer is usually thought to be a local minimum on a potential energy map or an average conformation over some region of dihedral angle space. However, broad distributions of sidechain dihedral angles (such as amides) may be represented by several rotamers, which may not all be local energy minima or population maxima or means. Nonrotameric is sometimes used to describe sidechains that have dihedral angles far from average values or far from a local energy minimum on a potential energy surface.

Rotamer libraries can be backbone-independent, secondary-structure-dependent or backbone-dependent. The distinctions are made according to whether the dihedral angles of the rotamers and/or their frequencies depend on the local backbone conformation or not. Backbone-independent rotamer libraries make no reference to backbone conformation and are calculated from all available sidechains of a certain type. Secondary-structure-dependent libraries present different dihedral angles and/or rotamer frequencies for α helix, β sheet or coil secondary structures. Backbone-dependent rotamer libraries present conformations and/or frequencies that are dependent on the local backbone conformation, as defined by the backbone dihedral angles ϕ and ψ, regardless of secondary structure. Finally, a variant of backbone-dependent rotamer libraries exists in the form of position-specific rotamers, which are defined by a fragment, usually of five amino acids in length, whose central residue's sidechain conformation is examined.

In this article, I will first review the history of rotamer libraries and then discuss some of the important issues in their design and use. I conclude with some advice on choosing a rotamer library and some discussion of future directions.

History

A list of published rotamer libraries is given in Table 1. As the size of the structure database has increased over the years, the libraries have become more precise and more informative. As early as 1970, Chandrasekaran and Ramachandran [2] counted rotamers of amino acids in the three protein structures then available (lysozyme, chymotrypsin and myoglobin). They compared their counts with hard-sphere energy calculations over allowed regions of the Ramachandran map and 20° variation about the canonical values for staggered dihedral angles. Thus, their rotamer library was backbone-independent, but their calculations explicitly considered the ϕ,ψ dependence of rotamer energies for each sidechain type. Bhat *et al.* [3] used 23 structures available in 1976 to produce a backbone-independent

Table 1

Published rotamer libraries.

Authors	Year	Type of library	Number of proteins in library	Resolution (Å)
Chandrasekaran and Ramachandran [2]	1970	BBIND	3	NA
Janin *et al.* [4]	1978	BBIND, SSDEP	19	2.5
Bhat *et al.* [3]	1979	BBIND	23	NA
James and Sielecki [5]	1983	BBIND	5	1.8, R-factor <0.15
Benedetti *et al.* [6]	1983	BBIND	238 peptides	R-factor <0.10
Ponder and Richards [7]	1987	BBIND	19	2.0
McGregor *et al.* [8]	1987	SSDEP	61	2.0
Tuffery *et al.* [9]	1991	BBIND	53	2.0
Dunbrack and Karplus [10]	1993	BBIND, BBDEP	132	2.0
Schrauber *et al.* [11]	1993	BBIND, SSDEP	70	2.0
Kono and Doi [12]	1996	BBIND	103	NA
De Maeyer *et al.* [13]	1995	BBIND	19	2.0
Dunbrack and Cohen [14]	1997–2002	BBIND, BBDEP	850*	1.7
Lovell *et al.* [15••]	2000	BBIND, SSDEP	240	1.7

*Latest update, May 2002. NA, not available.

rotamer library for all sidechains through χ_4. Janin *et al.*, in 1978 [4], provided secondary-structure-dependent data summed over all sidechains (excluding short and β-branched sidechains). In 1983, James and Sielecki [5] used five higher resolution structures to produce dihedral angle distributions with lower variances than the larger sample used by Janin *et al.* and were therefore the first to emphasize using better, if fewer, structures for deriving rotamer libraries. Also in 1983, Benedetti *et al.* [6] presented data from 258 peptide crystal structures, with 321 sidechains available for analysis. In addition to backbone-independent dihedral angle distributions, they showed Ramachandran distributions for each χ_1 rotamer, demonstrating the strong interdependence of backbone and sidechain conformations.

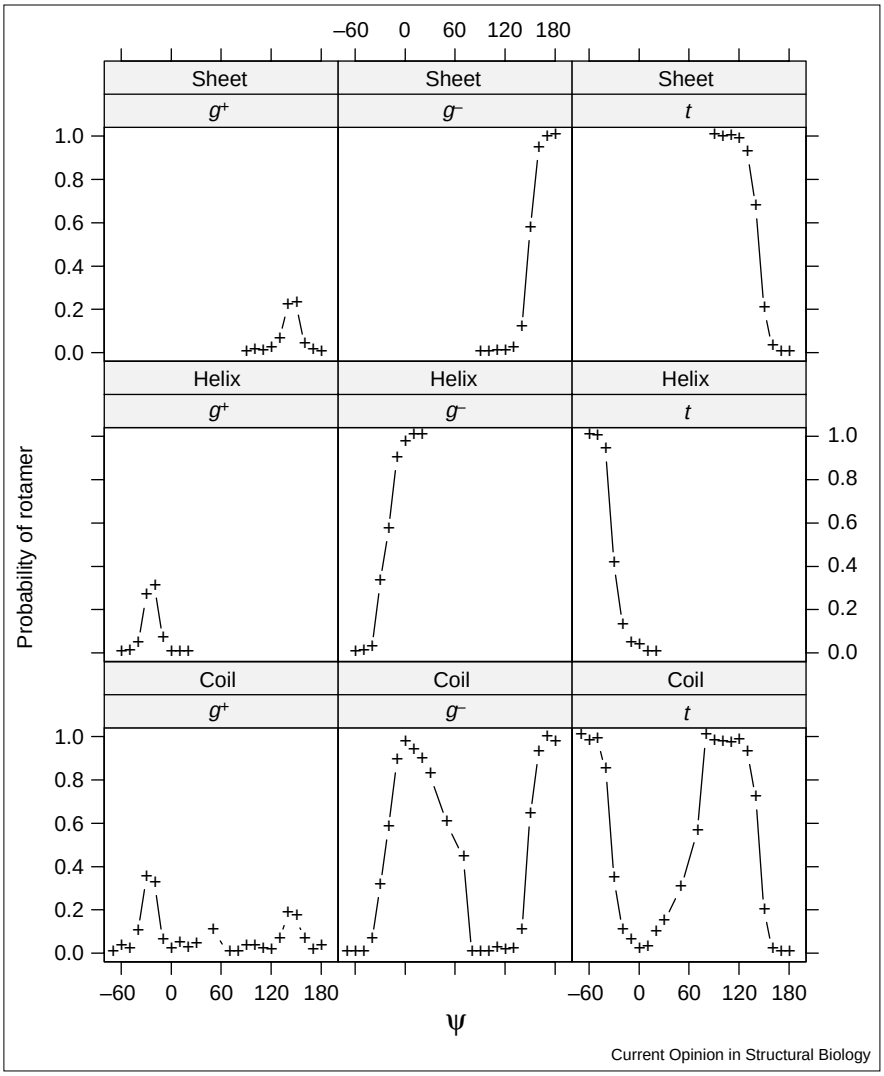
In 1987, Ponder and Richards [7] presented the first complete rotamer library — a list of all likely conformations of sidechains and their average dihedral angles, variances and frequencies. Their library was derived for use in determining what sequences would favor a known backbone conformation, essentially the procedure used nowadays in protein design. Also in 1987, McGregor *et al.* [8] used 61 high-resolution structures to examine the influence of secondary structure on rotamer populations, producing a secondary-structure-dependent rotamer library. In the context of sidechain conformation prediction, Tuffery *et al.* [9] derived a backbone-independent rotamer library based on 53 high-resolution structures available in 1991. In 1993, Dunbrack and Karplus [10] presented the first backbone-dependent rotamer library for use in sidechain conformation prediction. This library consisted of frequencies of the χ_1-χ_2 rotamers for each residue type in populated regions of the Ramachandran map, divided into $20^\circ \times 20^\circ$ bins of the ϕ,ψ dihedral angles. Average sidechain dihedral angles were not given. At the same time, Schrauber *et al.* [11] examined 'nonrotamericity', giving average dihedral angles in a backbone-independent fashion, but frequencies dependent on secondary structure.

Kono and Doi [12] used cluster analysis in 1996 to derive a backbone-independent rotamer library including frequencies, average dihedral angles and variances. In 1997, De Maeyer *et al.* [13] expanded the Ponder and Richards library [7] by adding rotamers to fill out all possible staggered conformations for sp^3–sp^3 dihedral angles and to sample the broad distribution of amide and carboxylate dihedral angles. They also added rotamers by including conformations one standard deviation in each direction away from the Ponder and Richards averages, producing a "highly detailed" rotamer library for sidechain conformation prediction.

In 1997, Dunbrack and Cohen [14] used Bayesian statistics to estimate populations and dihedral angles for all rotamers of all sidechain types at all values of ϕ and ψ. This was accomplished by deriving an informative prior distribution based on the product of the ϕ and ψ dependencies, and using the Bayesian formalism to combine the fully ϕ,ψ-dependent data likelihood (a multinomial distribution) with the prior distributions expressed as Dirichlet functions. In populated parts of the Ramachandran map, the results of this calculation are populations and dihedral angles very close to the data values; in sparsely populated regions of the Ramachandran map, the informative prior distribution dominates the predicted populations and angles. The Bayesian mechanism allowed us to achieve a smooth transition between these two situations in a statistically sound manner.

In a significant recent development, Richardson and colleagues [15••,16] have used much stricter criteria for including sidechains in a data set used to build a backbone-independent rotamer library. In addition to using more highly resolved and refined structures than previous libraries, these criteria included eliminating sidechains with high B-factors for any atom, eliminating sidechains with clashes of any atom (including hydrogens built with the

Figure 1



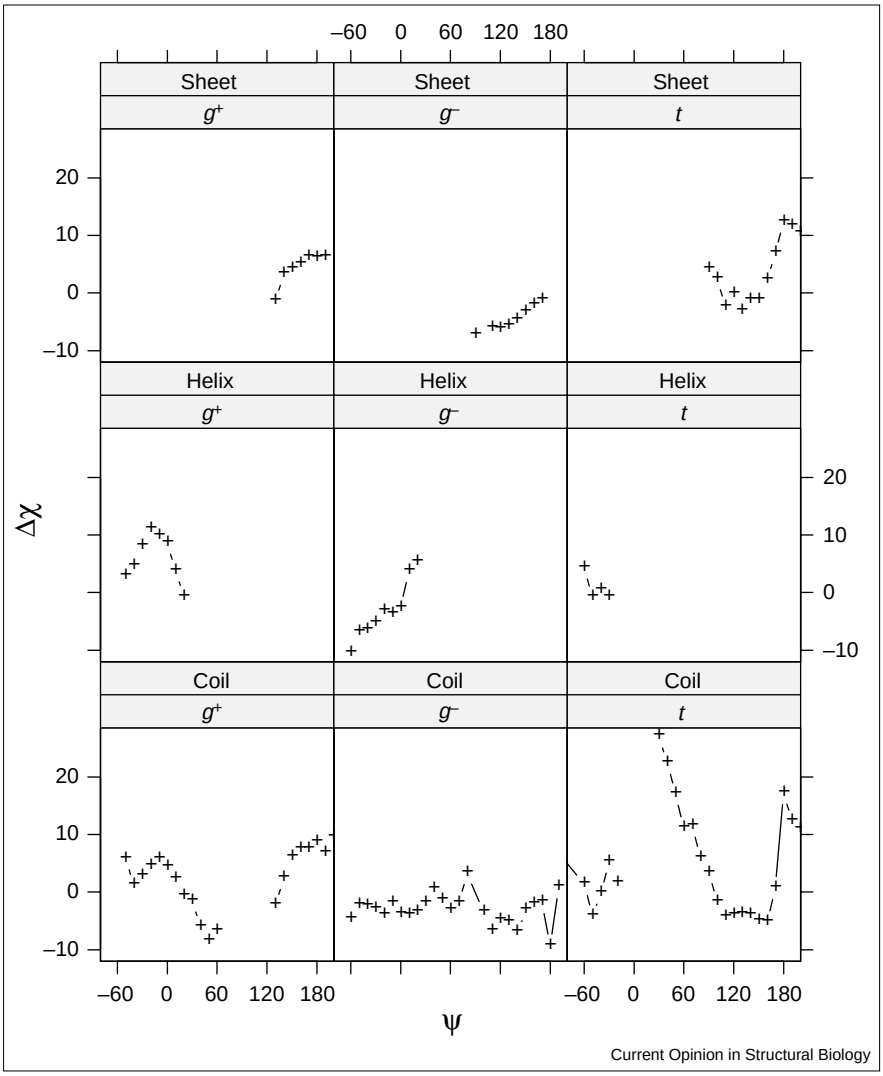
Observed frequency of the *gauche*$^+$ (g^+; $\chi_1 \sim +60°$), *gauche*$^-$ (g^-; $\chi_1 \sim -60°$) and *trans* (t; $\chi_1 \sim 180°$) rotamers of valine (horizontally, respectively) in sheet, helix and coil regions (vertically, respectively) of proteins as a function of the backbone dihedral angle ψ. Data were taken from a list of 850 proteins, at 1.7 Å resolution or better, with mutual sequence identity less than 50% (http://www.fccc.edu/research/labs/dunbrack/culledpdb.html). A B-factor cutoff of 40 was used, as recommended by Lovell *et al.* [15••].

program REDUCE) with any other atom in the structure (internal to the sidechain or otherwise) and eliminating sidechains with uncertain amide or histidine ring orientations after optimization of hydrogen bonding interactions by flipping some sidechains [17]. The effect of removing atom clashes is significant. As an example, for some time it has been known that some leucine sidechains are incorrectly modeled in crystal structures by reversing Cδ1 and Cδ2 (a rotation of 180° about χ_2 and shifting χ_1 slightly [18]). This leaves two δ atoms in essentially the same place, but, if built with standard bond angles and bond lengths, results in a strained conformation with internal steric conflicts. The criteria used by Lovell *et al.* [15••] remove most such improperly modeled sidechains. Many sidechains with highly strained dihedral angles are very likely to be averages of partially occupied rotamers and therefore have suspect dihedral angles. They may also be removed by a steric clash check. Finally, Lovell *et al.* used the modes rather than means to define rotamers, because dihedral angle distributions may be highly skewed for physical reasons. The mode gives the most common conformation, whereas the mean may give a partially strained conformation.

Conformational analysis

The conformational flexibility of organic molecules has been studied since the early 1930s by chemists, and the procedures and insights of conformational analysis have been applied to protein sidechain rotamers from the earliest days of protein structure determination. Arguments based on simple steric analysis, from the presence of gauche and syn-pentane interactions [19] all the way to detailed molecular mechanics energy calculations [14,20,21] and even quantum mechanics [22], have been used to understand the observed dihedral angle distributions of sidechains [14,18,23]. Although many of the important issues were considered in the early 1970s, when the available data were sparse, these have been revisited in recent years and compared to recent rotamer libraries. I discuss some of the more recent work and controversies.

Figure 2



Observed average χ_1 dihedral angles for the χ_1 rotamers of serine in sheet, helix and coil regions of proteins as a function of the backbone dihedral angle ψ. The angles are plotted as $\Delta\chi$ from the canonical values for each rotamer (i.e. +60°, –60° and 180°). Data arranged as in Figure 1.

Secondary-structure-dependent versus backbone-conformation-dependent rotamer libraries

The fact that sidechain rotamer populations should vary with local backbone conformation was considered by Chandrasekaran and Ramachandran in 1970 [2]. The question remains whether such backbone dependence should be expressed in a secondary-structure-dependent fashion or in terms of the local backbone dihedral angles φ and ψ. The most recent rotamer library, that of Lovell *et al.* [15••], included secondary-structure-dependent frequencies and dihedral angle values, in addition to more detailed analysis of backbone-independent rotamers. Secondary-structure-dependent rotamer libraries continue to be used in homology modeling [24]. The backbone-dependent rotamer library of Dunbrack and Cohen [14] includes frequencies, mean dihedral angles and variances as a function of the backbone dihedral angles.

The data would tend to argue in favor of explicit backbone dependence rather than secondary structure dependence. In Figure 1, the probabilities of the three χ_1 rotamers of valine (g^+ = ~+60°, g^- = ~–60°, t = 180°, left to right, respectively) are shown as a function of the dihedral angle ψ separately for residues in β sheets, α helices and coil secondary structures (top to bottom, respectively). The α and β secondary structures cover only a portion of the ψ range, and there are very few sidechains with ψ between –150° and –100°. Nevertheless, in the ranges of ψ in which they overlap, the distributions in the different secondary structures show an almost identical dependence on ψ. As valine has two γ carbons, it exhibits a very striking dependence on backbone dihedrals, because generally two out of three rotamers endure steric interactions with the backbone at any given value of ψ (or φ for that matter). But the similarities of ψ dependence over the three secondary structure types (coil, helix, sheet) hold up for all amino acid types.

One can also examine the backbone dependence of the average dihedral angle and its variance. This is shown for the three χ_1 rotamers of serine in Figure 2 for sheets,

helices and coils, as in Figure 1. In this figure, the average χ_1 angle differences from the canonical values for each rotamer are plotted versus ψ (i.e. $\overline{\chi}_1 - 60°$ for g^+; $\overline{\chi}_1 + 60°$ for g^-; and $\overline{\chi}_1 - 180°$ for t). Rotamers that become less prevalent at certain values of ψ, because of ψ-dependent backbone–sidechain interactions (between $C\gamma_i$ of residue i and backbone O_i and N_{i+1}), tend to have strained χ_1 dihedral angles to avoid clashing with the backbone. This only occurs for the g^+ and t rotamers, as these two conformations are gauche to the carbonyl carbon. Because the g^- rotamer is gauche to the backbone nitrogen and $H\alpha$, it does not interact with the backbone in a ψ-dependent fashion. This is evident in the rather flat average angle dependence on ψ for g^-. These interactions were first identified by the Madras group in the late 1960s and early 1970s [2,25], and used later by us to explain the observed preferences in the backbone-dependent rotamer library [19].

This is not to say that there is no residual dependence on secondary structure after backbone dihedral angles are considered. This is especially likely with hydrogen bonding sidechains that interact with the backbone two to four residues away in the chain (e.g. asparagine and serine in α helices). With the increasing size of the database, such an analysis is now or will be soon in reach. The interrelationship of backbone and sidechain conformation has very nicely been reviewed recently by Chakrabarti and Pal [26••]. They have examined the backbone conformations as dependent on the χ_1 rotamer, essentially viewing sidechain–backbone interactions as the sidechain influencing the backbone [27], rather than the other way around, as in the backbone-dependent rotamer library.

Variance and covariance of dihedral angles about rotamer values

Because rotamer libraries are derived in different ways — by clustering or by dividing up all dihedral angle space into bins, and by using different statistical methods — the analysis of variance has been considered in various ways. Dihedral angle distributions have often been treated as arising from a normal model and, as such, means and variances are calculated. Ponder and Richards [7] provided means and standard deviations (i.e. the square root of the variance) for their backbone-independent rotamer library. MacArthur and Thornton [28] noted that variance in dihedral angles continued to decrease with increasing resolution of protein structures down to 1.0 Å. Dunbrack and Cohen [14] treated dihedral angles in each bin as coming from normal distributions and used a Bayesian treatment to estimate the variance for each dihedral angle. The χ_1 variance was calculated as a function of the backbone dihedral angles, whereas the χ_2-χ_3-χ_4 variances were considered dependent on the χ_1 rotamer, but independent of the backbone conformation. This is due to insufficient data to calculate so many parameters. Rotamers with unfavorable interactions with the backbone or within the sidechain tended to show larger variance. Covariance between neighboring dihedral angles was not considered (apart from the ϕ-χ_1 and ψ-χ_1 dependence), again due to insufficient data. For common rotamers, this is now certainly feasible. For uncommon rotamers, a Bayesian treatment with an informative prior distribution may provide reasonable estimates.

Lovell *et al.* [15••] took a different approach by dispensing with the normal model for dihedral angle distributions. Their rotamers are based on modes and not means due to the observed skew in dihedral angle distributions. This was accomplished by using a Gaussian smoothing function on each data point and calculating the resulting density every degree to identify the mode. The variation in dihedral angles was represented by the half-width at half-height of the Gaussian smoothed distributions, after correcting for the smoothing function. For normal distributions, these values are larger than the standard deviation (σ) by a factor of 1.174. But, given the non-normality of the distributions, Lovell *et al.* use the average of the two half-width values. They also examine skew by considering differing half-widths on each side of the mode. This is a good solution to the dihedral angle distribution problem, because it is effectively a nonparametric approach and does not depend on assuming that dihedral angles are the result of random draws from a well-defined distribution function. There is one potential drawback, however, in that relatively rare rotamers will have poor estimates of both mode and variance in the nonparametric approach.

A third alternative may also be considered by noting that there are perhaps more suitable distribution functions than the normal model and that a parametric statistical approach may be justified if the appropriate distribution can be identified. Also, the stringent criteria for excluding some sidechains from the statistical analysis used by Lovell *et al.* [15••] may improve the fit to some distribution functions by removing outliers caused by questionable coordinates. There is a large literature available of statistics of directional data that one might take advantage of [29]. Indeed, Dowe *et al.* [30] used the von Mises function (the periodic analog of the normal function; it can be derived by restricting a normal distribution function to the points on a unit circle) to describe backbone ϕ,ψ distributions. It is not yet clear whether the von Mises function provides a better fit than the normal model. It is a symmetric function about its mean, so it cannot model skew. Consideration of the backbone dependence of χ_1 averages may also alleviate the problem of skew, because it is in part caused by strain induced by unfavorable backbone–sidechain contacts. Such interactions push the sidechain in one direction away from the usual χ_1 mean, rather than the other direction, inducing skew. The direction chosen is, of course, the one that relieves the contact, whereas motion in the other direction makes the conflict substantially worse.

'Nonrotameric' sidechains and strain

Related to the issue of variance is the consideration of sidechains with dihedral angles far from the average (or

modal) values given in rotamer libraries. The question remains as to the prevalence or even the existence of such so-called nonrotameric sidechains, and to their physical origins. Argos and colleagues [11,31,32] have argued, in a series of papers, that nonrotameric sidechains are common for some residue types and that physical interactions can explain their existence.

For a given sidechain with highly unusual dihedral angles, there are several possible explanations: the sidechain is improperly fit to the observed electron density; the observed conformation is actually an average of two rotameric conformations in equilibrium; for differences from rotameric values that are not too large, steric interactions with the backbone may force the sidechain to adopt strained dihedral angles, as shown in Figure 2; the strained conformation is stabilized by highly favorable interactions with the rest of the protein, usually hydrogen bonds; the strained conformation is an uncompensated high energy conformation that results in lower stability of the folded state.

Richardson's [15••] technique for using atom contacts and B-factors to eliminate some sidechains from consideration is useful for taking care of residues that may be improperly fit into the electron density, but, of course, it cannot eliminate all such cases. MacArthur and Thornton [28] examined dihedral angle variances as a function of resolution and R-factors, and found that variance continued to decline at higher resolutions all the way down to 1.0 Å. Their explanation was that, at higher resolutions, dynamically averaged conformations tend to be resolved into two discrete conformations that have dihedral angles close to rotameric values. Such partially occupied conformations are probably highly under-represented in crystallographic structures. West and Smith [33] estimated from α–β coupling constants from ten NMR structures that as much as 25% of sidechains may populate more than one χ_1 rotamer, with 5% of completely buried sidechains and up to 60% of exposed sidechains exhibiting such disorder. Zhao *et al.* [34•] examined sidechain conformations in paired crystal structures with identical sequences and found that surface sidechains exhibited wide variation in χ_1 dihedral angles, either in the same rotamer or through changes of rotamer. Najmanovich *et al.* [35] have studied differences in sidechain conformation in liganded and unliganded structures of identical proteins and found, as expected, that some polar and charged sidechains exhibit more rotamer variability than bulky aromatic amino acids.

Some sidechains may be in strained conformations due to unfavorable interactions with the local backbone or even nonlocal interactions. Penel and Doig [36] studied strained rotamer conformations in α helices, noting that some sidechains exhibit either higher energy rotamers or strained dihedral angles (or both), and identifying such strain as a force opposing folding. However, they assumed that rotamer energies were identical in all helical positions, neglecting that, even in helices, the probability of the three χ_1 rotamers varies with local backbone dihedral angles (Figure 1, top row). They also underestimate sidechain entropy in the folded state by assuming it is zero. As the data of West and Smith [33] indicate, this is not true. Nevertheless, the point that rotamers that are higher in energy than the lowest one available for a given local backbone conformation necessarily raise the energy of a protein is valid and important. Either by using nonoptimal rotamers or nonoptimal dihedral angles, they will produce a force opposing folding, which, of course, can be compensated by other interactions that are favorable. For example, Lazar *et al.* [37] observed lower stability and conformational disorder for a core variant of ubiquitin due to the presence of high energy rotamers observed in the NMR structure. Some proteins will undergo major changes in backbone conformation due to sequence changes (designed or otherwise), rather than adopt locally unfavorable rotamers [38].

Sidechains with nearly eclipsed dihedral angles are expected to be very rare. Proteins are only marginally stable at physiological temperatures (5–20 kcal/mol). An eclipsed dihedral angle for sp^3–sp^3 degrees of freedom can be 4–10 kcal/mol higher in energy than fully staggered conformations. Such conformations almost certainly have to be stabilized by highly favorable interactions, predominantly hydrogen bonds. Richardson and colleagues [15••] identified several glutamine residues with nearly eclipsed dihedral angles and 3–4 hydrogen bonds. Petrella and Karplus [39••] used molecular mechanics functions to analyze strain and found that nearly one-half of nonrotameric sidechains (defined as 30° or more away from the canonical staggered values) had potential energy minima with the CHARMM potential at rotameric values. Although 98% of rotameric sidechains were correctly predicted by CHARMM in the context of other sidechains in their X-ray positions, only 64% of the nonrotameric sidechains were correctly predicted. This indicates that, from crystallographic data alone, nonrotameric sidechains are likely to be over-represented compared to the true distribution. The results of Petrella and Karplus [39••], and MacArthur and Thornton [28] would tend to invalidate studies that conclude that nonrotameric sidechains are common [31,32].

Hydrogen bonding interactions, amide flips and histidine

The χ_2 dihedral angles of asparagine and histidine, and the χ_3 dihedral angle of glutamine present special difficulties, because the identification of the flip state is not usually obvious from consideration of electron density alone. It requires the analysis of likely hydrogen bonding interactions. Several groups have developed software for placing hydrogen atoms in proteins, for both hydrogen bonding and non-hydrogen bonding heavy atoms [17,40–42]. Lovell *et al.* [17,40] were the first to use such a program before producing a rotamer library, so their library has more tightly clustered dihedral angle distributions for asparagine and glutamine terminal amides. In the work of Dunbrack and Cohen [14],

the flip state is ignored and the terminal dihedrals of asparagine and glutamine are confined to the region from –90° to +90°, by a 180° flip if necessary. This is not a good solution to this problem and has recently been rectified (DA Montgomery, RL Dunbrack, unpublished data).

Some analysis of common hydrogen bonding interactions between sidechains and the protein backbone has been presented. For example, Vijayakumar *et al.* [43] studied the backbone–sidechain hydrogen bonds of the short polar sidechains of serine, asparagine, aspartic acid and threonine, giving the combinations of χ_1 (or χ_1,χ_2) and backbone dihedrals necessary for the formation of each hydrogen bond between the sidechain of residue *i* and the backbone of residue $i \pm n$, where *n* ranges from 0 to 3. They hypothesized that such sidechains act as helix breakers because of their propensity for forming backbone–sidechain hydrogen bonds. They are therefore rarer in the middle of helices and more common on helix termini. Eswar and Ramakrishnan [44] performed a similar analysis. Another electrostatic interaction may also influence the conformation of these sidechain types and that is carbonyl–carbonyl dipole interactions. Deane *et al.* [45] found that aspartic acid and asparagine residues in left-handed helical positions of the Ramachandran map exhibited strongly favorable dipole–dipole interactions between the sidechain carbonyl and the backbone carbonyl of the same residue.

Rotamer probabilities

One aspect of rotamer libraries that is crucial to consider when using them is the differences in energies and frequencies of the different rotamers. Although a rotamer library may provide a list of likely conformations, the conformations should not all be treated equally. Some rotamers contain internal dihedral strain because of gauche and syn-pentane interactions. Gauche interactions occur when four heavy atoms are connected by a dihedral angle of ~±60° and are about 0.9 kcal/mol higher in energy than atoms in a *trans* configuration. This is due to repulsion of the bonding molecular orbitals of the 1–2 and 3–4 atom pairs [46].

Syn-pentane interactions occur whenever two consecutive dihedrals in any chain of five heavy atoms are of opposite sign and magnitude less than roughly 75°. They are about 3.3 kcal/mol higher in energy than the *t,t* configuration and 1.5 kcal/mol higher in energy than the g^+,g^+ or g^-,g^- configurations. In sidechains, they can occur for χ_1,χ_2 rotamers that are g^+,g^- or g^-,g^+, as well as t,g^- or g^+,g^+ (both contain Cδ syn-pentane to the backbone carbonyl carbon), and for χ_2,χ_3 and χ_3,χ_4 rotamers of the longer sidechains. The γ heavy atoms of sidechains are in a five-atom chain with heavy atoms whose position is determined by the backbone dihedrals φ and ψ, including C of the previous amino acid, N of the next amino acid and O of the same amino acid, resulting in the variation of population and dihedral angles with φ and ψ, as seen in Figures 1 and 2.

Syn-pentane interactions usually cause dihedral angles to be significantly strained from the normal staggered conformation. For example, the g^-,g^+,t,t conformation of lysine has dihedral angles –79°, 75°, 177°, 178°, compared to the g^-,g^-,t,t conformation, with dihedral angles –62°, –65°, 178°, 178°. Other potential rotamers are not observed at all due to internal steric interactions and dihedral strain. In principle, lysine and arginine have 81 rotamers, if one counts 3 possible states for each of 4 rotatable bonds. However, only 32 of these can exist without syn-pentane interactions and most rotamers with 2 or more syn-pentane interactions have never been seen at all in high-resolutions structures (RL Dunbrack, unpublished data). Our libraries give Bayesian-estimated probabilities for such conformations and they are vanishingly small. But, if used uncritically, they might be sampled in conformation prediction, even though *a priori* they are very unlikely.

Researchers who use rotamer libraries in homology modeling or in protein design have dealt with the rotamer energy problem in different ways and, in several cases, overlooked it. The simplest way is simply to ignore rotamers with very low probabilities. Our sidechain prediction program SCWRL discards very low probability rotamers, as do some other sidechain prediction programs that use our library [47]. This reduces the search space significantly, which is beneficial in both time and accuracy. Another way is to use the probabilities in the rotamer library to derive a pseudo-energy function as $E_i = -K \ln p_i$, where p_i is the probability of rotamer *i* and *K* is some constant, not necessarily equal to k_BT. This probability can be backbone-dependent or backbone-independent. This type of energy is used in SCWRL [48,49], in the sidechain prediction methods of Mendes *et al.* [50] and Liang and Grishin [51•], and in the protein design work of Kuhlman and Baker [52,53], all of which use the backbone-dependent rotamer library. SCWRL uses what is effectively a dead-end elimination algorithm [54], followed by a branch-and-bound algorithm, to solve the combinatorial problem for clusters of interacting sidechains, whereas Mendes *et al.* [50] use a mean-field algorithm. Liang and Grishin [51•], and Kuhlman and Baker [52,53] use Monte Carlo simulations.

Molecular mechanics potentials can be used to discriminate between high and low energy rotamers, and this method is used in sidechain prediction methods and some protein design efforts that use the backbone-dependent rotamer library [55–57]. Petrella and Karplus [18] have recently shown that the CHARMM potential energy function predicts the conformation of a single sidechain well in the presence of all other sidechains in their crystallographic positions. These calculations were performed by a complete search over dihedral angles, so that strained and unstrained angles were investigated. Even when the predicted rotamer was not the same as the crystallographic rotamer, it was found that the energy difference was small. The potential was able to discriminate between

high energy local minima and low energy minima. In an important result, when Petrella and Karplus [39••] studied nonrotamer conformations (see above), they found that excluding the torsion term in the CHARMM potential resulted in a severe loss of predictive accuracy. This torsion term is necessary to achieve an accurate value for the relative energies of the different rotamers. When molecular mechanics potentials are used in homology modeling or protein design, it is clear that such calculations ought to include torsional energy terms, as some strain in sidechain conformations is not adequately represented by van der Waals interactions, particularly when the backbone-dependent dihedral angles are used. In this case, the sidechains already have strained dihedrals to avoid sharply unfavorable van der Waals interactions. Unfortunately, this term is sometimes not included, even when other parts of the molecular mechanics function are used [58,59].

Sidechain packing efforts that do not use either the log probabilities or torsion terms from molecular mechanics are likely to produce inaccurate conformation predictions or incomplete descriptions of the physical forces that determine such conformations. For instance, Kussell *et al.* [60•] used rotamers and simple steric functions to study packing, but did not use rotamer probabilities or an external energy function to discriminate between rotamers. Thus, they came to the conclusion that "hydrophobic, polar, and electrostatic interactions" stabilize native rotamers compared to the many conformations available. Although this is no doubt true, it ignores the relative energies of rotamers that ultimately lower rotamer choice substantially. Creamer [61] performed Monte Carlo simulations on peptides to determine a sidechain entropy scale and observed that lysine and arginine effectively sampled only 37 of 81 possible rotamers. It was incorrectly concluded that significant interactions of the χ_3 and χ_4 atoms with the backbone were responsible for this phenomenon. Instead, it is almost certainly due to internal strain of the 49 rotamers with at least one syn-pentane interaction and the 17 rotamers with two or more. Of these 49 rotamers, 31 have syn-pentane interactions involving the χ_3 and/or the χ_4 atoms.

Conclusions

Which rotamer library?

I conclude first by commenting on the suitability of rotamer libraries for the various applications that use them. For many purposes, a backbone-independent rotamer library is most suitable and the best among these is that of Lovell *et al.* [15••]. Because of poor statistical power and the existence of physically unfeasible rotamers (without frequency information), there is little justification for using some of the previously developed backbone-independent libraries [7,9]. Unfortunately, they remain in common use, even in rather recent work [62–64]. Backbone-independent rotamer libraries are particularly suitable for X-ray and NMR structure refinement, as they provide excellent starting conformations that can be altered as necessary by the experimental data. They are also useful for the development of entropy scales and for representing the properties of the unfolded state [33], because residues in unfolded proteins are likely to sample the Ramachandran map in rough proportion to the sampling observed in random coil regions of folded structures. A backbone-independent rotamer library provides the necessary probabilities and conformations averaged over backbone conformations observed in proteins. The Lovell library is publicly available at http://kinemage.biochem.duke.edu/databases/rotamer.php.

However, for the prediction of sidechain conformations, either in homology modeling or in protein design, one may wish to take account of populations and/or dihedral angle variations that depend on the local backbone conformation [48,50,55,57]. Because such calculations may be computationally intensive, sampling around the average dihedral angles in backbone-independent rotamer libraries, energy minimization or molecular dynamics simulations that might take account of backbone–sidechain interactions may not be feasible. The explicit dependence of populations and dihedral angles on backbone conformation is therefore valuable. The use of secondary-structure-dependent libraries is not generally advisable, as they do not provide much more information than backbone-independent rotamer libraries. The latest backbone-dependent rotamer library (May 2002), based on 850 high-resolution structures (at better than 1.7 Å resolution), is available at http://www.fccc.edu/research/labs/dunbrack/sidechain.html. As recommended by Lovell *et al.* [15••], sidechains with B-factors greater than 40 and sidechains with steric contacts have been eliminated from the data used to calculate the library (DA Montgomery, RL Dunbrack, unpublished data).

Future directions

As the structure database increases in size and diversity of folds, it will be possible to examine more detailed factors influencing the populations and dihedral angles of sidechain rotamers. One such factor is the change in χ_2 rotamer populations and angles dependent on the χ_1 rotamer state and backbone conformation. For aspartic acid and asparagine, this variation is pronounced, but there is some variation for other sidechains as well. The database is still limited for such analysis, because one is dividing the data into nine possible $\chi_1\chi_2$ rotamers and the backbone dihedrals ϕ and ψ.

As more very high resolution structures are determined, it will be possible to examine the conformational disorder of sidechains in a statistical manner, with the goal of predicting such disorder accurately. Currently, there are methods for the prediction of disorder [65], but very little data to use in verifying these predictions. Additional NMR data with α–β coupling constants would also be extremely valuable. As described above, analysis of the variance of dihedral angles is still rather crude, with no description of covariance yet available. Such a description should be possible with expanding data sets. Eliminating conformational disorder expressed as strained dihedral

angles, rather than multiple conformations, will also have a beneficial effect on studying dihedral angle variance.

Because rotamer libraries enjoy widespread use in protein structure prediction, protein design and structure refinement, improvements in their analysis and design will continue to have a wide impact on many fields.

Update

Two recent papers explore sidechain conformational entropy with molecular dynamics simulations. Schäfer *et al.* [66] find that vibrational entropy is not the same for each rotamer, contrary to the assumption made in deriving most entropy scales. Clore and Kuszewski [67] show that simulations with a ϕ,ψ-dependent potential energy function for sidechain dihedrals reproduce NMR-derived populations of surface sidechains.

Acknowledgements

Support from the National Institutes of Health (CA06927 and R01 HG02302) is gratefully acknowledged. I thank J Michael Sauder and Heinrich Roder for careful reading of the manuscript.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Eyring H: Steric hindrance and collision diameters. *J Am Chem Soc* 1932, 54:3191-3203.
2. Chandrasekaran R, Ramachandran GN: Studies on the conformation of amino acids. XI. Analysis of the observed side group conformations in proteins. *Int J Protein Res* 1970, 2:223-233.
3. Bhat TN, Sasisekharan V, Vijayan M: An analysis of sidechain conformation in proteins. *Int J Pept Protein Res* 1979, 13:170-184.
4. Janin J, Wodak S, Levitt M, Maigret B: Conformations of amino acid side-chains in proteins. *J Mol Biol* 1978, 125:357-386.
5. James MNG, Sielecki AR: Structure and refinement of penicillopepsin at 1.8 Å resolution. *J Mol Biol* 1983, 163:299-361.
6. Benedetti E, Morelli G, Nemethy G, Scheraga HA: Statistical and energetic analysis of sidechain conformations in oligopeptides. *Int J Pept Protein Res* 1983, 22:1-15.
7. Ponder JW, Richards FM: Tertiary templates for proteins: use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol* 1987, 193:775-792.
8. McGregor MJ, Islam SA, Sternberg MJE: Analysis of the relationship between sidechain conformation and secondary structure in globular proteins. *J Mol Biol* 1987, 198:295-310.
9. Tuffery P, Etchebest C, Hazout S, Lavery R: A new approach to the rapid determination of protein side chain conformations. *J Biomol Struct Dyn* 1991, 8:1267-1289.
10. Dunbrack RL Jr, Karplus M: Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J Mol Biol* 1993, 230:543-574.
11. Schrauber H, Eisenhaber F, Argos P: Rotamers: to be or not to be? An analysis of amino acid sidechain conformations in globular proteins. *J Mol Biol* 1993, 230:592-612.
12. Kono H, Doi J: A new method for sidechain conformation prediction using a Hopfield network and reproduced rotamers. *J Comp Chem* 1996, 17:1667-1683.
13. De Maeyer M, Desmet J, Lasters I: All in one: a highly detailed rotamer library improves both accuracy and speed in the modelling of sidechains by dead-end elimination. *Fold Des* 1997, 2:53-66.
14. Dunbrack RL Jr, Cohen FE: Bayesian statistical analysis of protein sidechain rotamer preferences. *Protein Sci* 1997, 6:1661-1681.
15. •• Lovell SC, Word JM, Richardson JS, Richardson DC: The penultimate rotamer library. *Proteins* 2000, 40:389-408.

The authors have derived a more accurate backbone-independent rotamer library by eliminating sidechains of low stereochemical quality, including those with high B-factors, those with steric conflicts in the presence of predicted hydrogen atom locations and so on. The statistical analysis does not rely on a parametric distribution function, such as the normal model, and hence can model factors such as skew in an unbiased way. Also, rotamers that are possible combinations of staggered dihedral angles (e.g. for lysine and arginine there are 81 combinations), but are not at all observed in the database, are not included in the library. This prevents their improper use in sidechain prediction and protein design.

16. Word JM, Lovell SC, LaBean TH, Taylor HC, Zalis ME, Presley BK, Richardson JS, Richardson DC: Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J Mol Biol* 1999, 285:1711-1733.
17. Lovell SC, Word JM, Richardson JS, Richardson DC: Asparagine and glutamine rotamers: B-factor cutoff and correction of amide flips yield distinct clustering. *Proc Natl Acad Sci USA* 1999, 96:400-405.
18. Petrella RJ, Lazaridis T, Karplus M: Protein sidechain conformer prediction: a test of the energy function. *Fold Des* 1998, 3:353-377. [Published erratum appears in *Fold Des* 1998, 3:588.]
19. Dunbrack RL Jr, Karplus M: Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nat Struct Biol* 1994, 1:334-340.
20. Marcus E, Keller DA, Shibata M, Ornstein RL, Rein R: Comparing theoretical and experimental backbone-dependent sidechain conformational preferences for linear, branched, aromatic, and polar residues. *Chem Phys* 1996, 204:157-171.
21. Gelin BR, Karplus M: Sidechain torsional potentials: effect of dipeptide, protein, and solvent environment. *Biochemistry* 1979, 18:1256-1268.
22. Matta CF, Bader RF: An atoms-in-molecules study of the genetically-encoded amino acids: I. Effects of conformation and of tautomerization on geometric, atomic, and bond properties. *Proteins* 2000, 40:310-329.
23. Nayeem A, Scheraga HA: A statistical analysis of sidechain conformations in proteins - comparison with ECEPP predictions. *J Protein Chem* 1994, 13:283-296.
24. Bates PA, Sternberg MJ: Model building by comparison at CASP3: using expert knowledge and computer automation. *Proteins* 1999, 37:47-54.
25. Ponnuswamy PK, Sasisekharan V: Studies on the conformation of amino acids. IX. Conformations of butyl, seryl, threonyl, cysteinyl, and valyl residues in a dipeptide unit. *Biopolymers* 1971, 10:565-582.
26. •• Chakrabarti P, Pal D: The interrelationships of side-chain and main-chain conformations in proteins. *Prog Biophys Mol Biol* 2001, 76:1-102.

The authors present a very nice review of their investigations of backbone–sidechain interactions.

27. Chakrabarti P, Pal D: Main-chain conformational features at different conformations of the side-chains in proteins. *Protein Eng* 1998, 11:631-647.
28. MacArthur MW, Thornton JM: Protein side-chain conformation: a systematic variation of chi 1 mean values with resolution - a consequence of multiple rotameric states? *Acta Crystallogr D Biol Crystallogr* 1999, 55:994-1004.
29. Mardia KV, Jupp PE (Eds): *Directional Statistics*. London: Wiley; 2000.
30. Dowe DL, Allison L, Dix TI, Hunter L, Wallace CS, Edgoose T: Circular clustering of protein dihedral angles by minimum message length. *Pac Symp Biocomput* 1996:242-255.
31. Heringa J, Argos P: Strain in protein structures as viewed through nonrotameric side chains: I. their position and interaction. *Proteins* 1999, 37:30-43.
32. Heringa J, Argos P: Strain in protein structures as viewed through nonrotameric side chains: II. effects upon ligand binding. *Proteins* 1999, 37:44-55.

33. West NJ, Smith LJ: Sidechains in native and random coil protein conformations. Analysis of NMR coupling constants and chi1 torsion angle preferences. *J Mol Biol* 1998, 280:867-877.

34. • Zhao S, Goodsell DS, Olson AJ: Analysis of a data set of paired uncomplexed protein structures: new metrics for side-chain flexibility and model evaluation. *Proteins* 2001, 43:271-279.
Variability in the sidechain conformations of proteins in multiple crystal structures is examined. Some residue types are frequently found in different rotamers in different crystal structures of the same protein. This puts an upper bound on the accuracy of sidechain conformation prediction. Because of this upper bound, the authors propose that the sidechain prediction program SCWRL is more accurate than a fixed dihedral angle cutoff of 40° would suggest.

35. Najmanovich R, Kuttner J, Sobolev V, Edelman M: Side-chain flexibility in proteins upon ligand binding. *Proteins* 2000, 39:261-268.

36. Penel S, Doig AJ: Rotamer strain energy in protein helices - quantification of a major force opposing protein folding. *J Mol Biol* 2001, 305:961-968.

37. Lazar GA, Johnson EC, Desjarlais JR, Handel TM: Rotamer strain as a determinant of protein structural specificity. *Protein Sci* 1999, 8:2598-2610.

38. Willis MA, Bishop B, Regan L, Brunger AT: Dramatic structural and thermodynamic consequences of repacking a protein's hydrophobic core. *Structure* 2000, 8:1319-1328.

39. •• Petrella RJ, Karplus M: The energetics of off-rotamer protein side-chain conformations. *J Mol Biol* 2001, 312:1161-1175.
Using the CHARMM potential, the authors demonstrate that almost half of so-called nonrotameric sidechains are not in a local energy minimum in the context of the crystal environment, whereas nearly 100% of rotameric sidechains are. This indicates that many nonrotameric sidechains are simply poorly refined rotameric conformations in X-ray crystal structures.

40. Word JM, Lovell SC, Richardson JS, Richardson DC: Asparagine and glutamine: using hydrogen atom contacts in the choice of sidechain amide orientation. *J Mol Biol* 1999, 285:1735-1747.

41. Hooft RW, Sander C, Vriend G: Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins* 1996, 26:363-376.

42. Nielsen JE, Andersen KV, Honig B, Hooft RW, Klebe G, Vriend G, Wade RC: Improving macromolecular electrostatics calculations. *Protein Eng* 1999, 12:657-662.

43. Vijayakumar M, Qian H, Zhou HX: Hydrogen bonds between short polar side chains and peptide backbone: prevalence in proteins and effects on helix-forming propensities. *Proteins* 1999, 34:497-507.

44. Eswar N, Ramakrishnan C: Deterministic features of side-chain main-chain hydrogen bonds in globular protein structures. *Protein Eng* 2000, 13:227-238.

45. Deane CM, Allen FH, Taylor R, Blundell TL: Carbonyl-carbonyl interactions stabilize the partially allowed Ramachandran conformations of asparagine and aspartic acid. *Protein Eng* 1999, 12:1025-1028.

46. Karplus M, Parr RG: An approach to the internal rotation problem. *J Chem Phys* 1963, 38:1547-1552.

47. Swain MT, Kemp GJ: Modelling protein side-chain conformations using constraint logic programming. *Comput Chem* 2001, 26:85-95.

48. Bower MJ, Cohen FE, Dunbrack RL Jr: Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J Mol Biol* 1997, 267:1268-1282.

49. Dunbrack RL Jr: Comparative modeling of CASP3 targets using PSI-BLAST and SCWRL. *Proteins Suppl* 1999:81-87.

50. Mendes J, Nagarajaram HA, Soares CM, Blundell TL, Carrondo MA: Incorporating knowledge-based biases into an energy-based sidechain modeling method: application to comparative modeling of protein structure. *Biopolymers* 2001, 59:72-86.

51. • Liang S, Grishin NV: Side-chain modeling with an optimized scoring function. *Protein Sci* 2002, 11:322-331.
The authors present an optimized scoring function for sidechain conformation prediction that includes contact, overlap and electrostatic terms, and a local rotamer energy term based on the log probabilities of the backbone-dependent rotamer library. They achieve a high rate of prediction accuracy compared to earlier published results.

52. Kuhlman B, Baker D: Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci USA* 2000, 97:10383-10388.

53. Kuhlman B, O'Neill JW, Kim DE, Zhang KY, Baker D: Conversion of monomeric protein L to an obligate dimer by computational protein design. *Proc Natl Acad Sci USA* 2001, 98:10687-10691.

54. Desmet J, De Maeyer M, Hazes B, Lasters I: The dead-end elimination theorem and its use in protein sidechain positioning. *Nature* 1992, 356:539-542.

55. Keating AE, Malashkevich VN, Tidor B, Kim PS: Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils. *Proc Natl Acad Sci USA* 2001, 98:14825-14830.

56. Wernisch L, Hery S, Wodak SJ: Automatic protein design with all atom force-fields by exact and heuristic optimization. *J Mol Biol* 2000, 301:713-736.

57. Voigt CA, Gordon DB, Mayo SL: Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. *J Mol Biol* 2000, 299:789-803.

58. Gordon DB, Mayo SL: Branch-and-terminate: a combinatorial optimization algorithm for protein design. *Structure* 1999, 7:1089-1098.

59. Pierce NA, Spriet JA, Desmet J, Mayo SL: Conformational splitting: a more powerful criterion for dead-end elimination. *J Comp Chem* 1999, 21:999-1009.

60. • Kussell E, Shimada J, Shakhnovich EI: Excluded volume in protein side-chain packing. *J Mol Biol* 2001, 311:183-193.
The authors examine the role of sidechain packing in determining conformations using a Monte Carlo simulation with a hard-sphere packing energy function. The results indicate that there are many choices of rotamer configurations that satisfy normal packing density without steric overlaps. Although not stated in this paper, it is likely that, in addition to attractive electrostatic and van der Waals interactions, relative rotamer energies are a major factor in determining the actual rotamers observed.

61. Creamer TP: Sidechain conformational entropy in protein unfolded states. *Proteins* 2000, 40:443-450.

62. Ota M, Isogai Y, Nishikawa K: Knowledge-based potential defined for a rotamer library to design protein sequences. *Protein Eng* 2001, 14:557-564.

63. Looger LL, Hellinga HW: Generalized dead-end elimination algorithms make large-scale protein sidechain structure prediction tractable: implications for protein design and structural genomics. *J Mol Biol* 2001, 307:429-445.

64. Kono H, Saven JG: Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure. *J Mol Biol* 2001, 306:607-628.

65. Leach AR, Lemon AP: Exploring the conformational space of protein sidechains using dead-end elimination and the A* algorithm. *Proteins* 1998, 33:227-239.

66. Schäfer H, Smith LJ, Mark AE, van Gunsteren WF: Entropy calculations on the molten globule state of a protein: side-chain entropies of α-lactalbumin. *Proteins* 2002, 46:215-224.

67. Clore GM, Kuszewski J: Chi(1) rotamer populations and angles of mobile surface side chains are accurately predicted by a torsion angle database potential of mean force. *J Am Chem Soc* 2002, 124:2866-2867.