PROTEINS: Structure, Function, and Genetics 40:389–408 (2000)

The Penultimate Rotamer Library

Simon C. Lovell, J. Michael Word, Jane S. Richardson, and David C. Richardson*
Duke University, Durham, North Carolina

ABSTRACT **All published rotamer libraries contain some rotamers that exhibit impossible internal atomic overlaps if built in ideal geometry with all hydrogen atoms. Removal of uncertain residues (mainly those with B-factors ≥40 or van der Waals overlaps ≥0.4 Å) greatly improves the clustering of rotamer populations. Asn, Gln, or His side chains additionally benefit from flipping of their planar terminal groups when required by atomic overlaps or H-bonding. Sensitivity to skew and to the boundaries of χ angle bins is avoided by using modes rather than traditional mean values. Rotamer definitions are listed both as the modal values and in a preferred version that maximizes common atoms between related rotamers. The resulting library shows significant differences from previous ones, differences validated by considering the likelihood of systematic misfitting of models to electron density maps and by plotting changes in rotamer frequency with B-factor. Few rotamers now show atomic overlaps in ideal geometry; those overlaps are relatively small and can be understood in terms of bond angle distortions compensated by favorable interactions. The new library covers 94.5% of examples in the highest quality protein data with 153 rotamers and can make a significant contribution to improving the accuracy of new structures. Proteins 2000;40:389–408.** © 2000 Wiley-Liss, Inc.

Key words: side-chain rotamer library; all-atom contact analysis; structure validation; reversed leucines; explicit hydrogens; van der Waals analysis

INTRODUCTION

Side-chain χ-angle distributions were studied as soon as multiple protein structures were available.[1–5] The observation that side-chain torsions fall into n-dimensional clusters and that, therefore, a library of rotamers can usefully be defined was introduced in 1987 by Ponder and Richards.[6] As the database has grown, several groups have since compiled updated rotamer libraries.[7–11] The concept of rotamers and the availability of rotamer libraries has changed the handling of side chains in homology modeling,[12] Monte Carlo and combinatorial calculations,[13] and protein design.[14,15] Side-chain rotamer libraries are incorporated into crystallographic model-to-map fitting programs such as O[16] and XtalView,[17] while χ angle expectations are part of verification tools,[18,19] including those used for all structures deposited at the PDB (Protein Data Bank[20,21]). The use of rotamers significantly improves both the speed and accuracy of building crystallographic models. However, any incorrect conformations included in a rotamer library will show increased occurrence in the less certain parts of new experimental structures as well as biasing theoretical models. We feel, therefore, that the accuracy of rotamer libraries is an important issue, with the increased use of repacking and homology modeling, and especially on the eve of a major structural genomics effort.

The growth of the PDB as a whole has been important in improving the accuracy of rotamer libraries, but even more important is the recent growth in the number of very high-resolution protein structures. At such resolution and in the good areas of the electron density map, side-chain conformations are very clearly seen, resulting in little bias from previously defined rotamers, refinement methods, or fitting errors. Unfortunately, no previous study has limited itself solely to these residues; usually resolution and homology criteria are applied to choose good structures, but then all residues in each structure contribute equally to the library.

The development of our all-atom contact analysis technique using the Probe program[22] and the optimization of H-atom positions in Reduce[23] allows us to analyze all-atom steric and H-bonding interactions. All published rotamer libraries are found by this new methodology to contain rotamers with serious van der Waals overlaps (*clashes*) when built with all atoms and standard geometry. Significant clashes in defined rotamers are unexpected, since the most commonly occurring conformations should have the lowest energy. Atomic overlaps (up to about 0.4 Å) may indicate the inappropriateness of using standard geometry (for example, if a conformation has systematically strained bond angles), but larger clashes almost certainly indicate an erroneous rotamer definition.

Previous rotamer studies have each used different approaches, leading to libraries with many similarities but some differences. The original library of Ponder and Richards[6] drew bins around the observed clusters and determined the mean and standard deviation of the peak. This library has been very influential and is still the most

Grant sponsor: National Institutes of Health; Grant number: GM-15000.

The Supplementary material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/index.html/.

*Correspondence to: David C. Richardson, 211 Nanaline Duke Building, Duke University, Durham, NC 27710-3711. E-mail: dcr@kinemage.biochem.duke.edu

Received 20 December 1999; Accepted 14 March 2000

widely used in some fields. Despite the small size of the database then available (e.g., only 16 examples of Met), it has surprisingly few artifacts, and in some aspects surpasses later libraries compiled from larger data sets. However, it could seldom reach beyond χ_2 and it includes some incorrect amide orientations. That work used hydrogens to check for long-range atomic clashes, but neither these nor subsequent authors have checked for internal clashes, presumably believing that high-resolution data prevents them.

Schrauber et al.[7] discussed whether or not rotamers were useful and whether side chains were *rotameric* (which they defined as mean ±20°). As part of that analysis they compiled a new rotamer library, but out only to χ_2 and excluding Asp and Asn. Despite its conservative nature, this library contains some duplication of rotamers, some rotamers with steric clashes, and some systematically misfit conformations. Also, despite their overall negative conclusions about the rotamer concept, their library has been used by others.

Tuffery et al. used cluster analysis to produce a library of 113 rotamers,[8] later expanded to 212,[9]which was used by them and others for combinatorial repacking calculations. Their rotamers are unusual in that they have relatively few clashes but contain duplicated rotamers for symmetrical side chains and often have nearly eclipsed χ angles. These features are probably due to their methodology of performing energy minimization on the structures before compiling the library. Such a procedure is a reasonable step in their repacking calculations, but we feel it is inappropriate in compiling a rotamer library. Energy minimization untethered to X-ray data rarely improves an experimental structure: if moving from the final model to a more correct structure was as simple as minimizing, the crystallographic refinement would already have done it. Indeed, such untethered energy minimization has been used by crystallographers to produce degraded models as controls for verification programs.[24]

The library of De Maeyer et al.[10] is a combination of those of Schrauber et al.[7] and Ponder and Richards,[6] with some extensions in order to include angles past χ_2 and to sample regions of torsion space not included in the former studies. The library has some undesirable features, including Arg χ_4 rotamers at ±60° causing substantial van der Waals clashes, some Asn and Gln rotamers with amide groups in incorrect flip states, and rotamers with fully-eclipsed χ angles. An advantage is their use of common χ angles leading to common atom positions, an approach also adopted here.

The rotamer library built into the O crystallographic fitting program[16] is also an extension from Ponder and Richards. Most of the relatively low number of rotamers are sound, but many genuine rotamers are missing and a few demonstrably-incorrect ones are included.

The most comprehensive recent analysis was done by Dunbrack and Cohen[11] (for updates see their website at www.fccc.edu/research/labs/dunbrack/sidechain.html) using over 500 structures. They divided torsion space into bins such that all regions were included and used Bayesian statistics to obtain an estimate of the population of otherwise sparse regions. This approach has significant advantages: the pure statistical accuracy is very high, and the probability for every division of rotamer space is explicitly stated (including every division of ϕ and ψ in the backbone-dependent version). However, this methodology, plus the inclusion of high-B data, lowers the overall contrast and leads to a defined rotamer in every possible bin, the less probable of which often show extremely large internal clashes and are unlikely ever to be genuinely observed. Also, especially for side chains with planar functional groups, the a priori bins split single distributions, leading to misplaced means and extra rotamers in the tails of the distribution which are valid as arbitrary sampling points but not as locally favored conformations. Overall, the library of Dunbrack and Cohen is the most complete one previously published, but users must give thoughtful attention to setting the lower threshold for acceptable rotamer probabilities.

We have recently published sets of rotamers for Met[22] and for Asn and Gln.[25] The Met rotamers were defined with a B-factor cutoff of 30, which tightened the χ_3 distribution remarkably, allowing 94% of the observed residues to be included in 13 rotamers. For Asn and Gln, we used our program Reduce to optimize H-bond networks, as well as to add all explicit hydrogens. About 20% of the side-chain amides were flipped by 180° because the flip resulted in substantially better hydrogen bonding or substantially less atomic overlap, while inconclusive cases were omitted. The result showed Asn and Gln terminal χ angle distributions with clear clustering for the first time, allowing definition of rotamers that correspond to low-energy conformations.

In the current work we extend our analysis to include all side-chains, using a database of 240 structures at 1.7 Å resolution or better and all applicable filters. Additionally, our all-atom contact analysis can easily distinguish between pairs of high- and low-energy conformations occupying approximately the same spatial position which might be mistaken for each other in lower-resolution electron density maps. If critical analysis (examination of van der Waals overlaps, electron density, and occurrence as a function of B-factor and resolution) indicates that an observed conformation is a systematic fitting error, then it is not included in our library.

The resulting rotamer library, because of the removal of side chains with uncertain conformations and systematic errors, is less prone to perpetuation of inaccuracies than those published previously. It is also complete as far as possible from the current high-quality database, and it shows a very high percentage of side chains to be rotameric.

METHODS

Nomenclature

Many different nomenclatures have been used to describe side-chain torsion angles. One of the most widely used is g+, g−, and t for gauche positive, gauche negative, and trans, respectively (as illustrated in Fig. 1 for the case

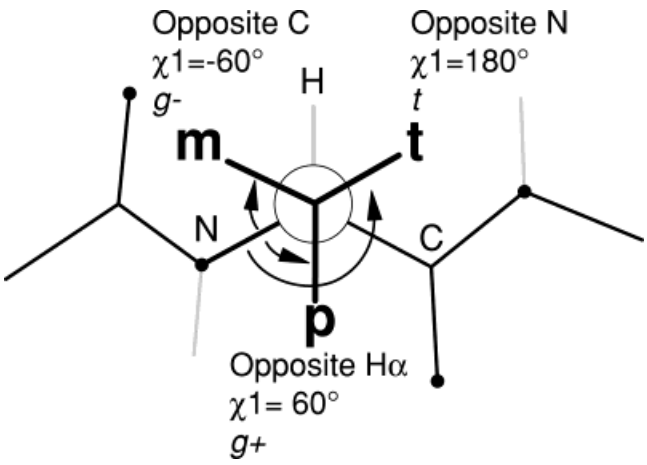

Fig. 1. Illustration of the relationships and nomenclature for the side-chain dihedral angle χ_1. Each of the three staggered conformations is labeled with its χ_1 angle (measured from the backbone N), its officially correct[29] g+ or g− designation, and its **p**, **t**, or **m** nomenclature as used in this work. Note that earlier studies have used opposite g+, g− designations, as discussed in the text (Nomenclature).

of χ_1). Unfortunately, g+ has been used to mean both +60°[11,26,27] and −60°.[2,4,5,7] Similarly, the abbreviated forms +, −, and t have been used both with + for +60°[6] and with + for −60°.[7] Confusion arose because of the redefinition of the trans conformation from 0° to 180° and the lack of an accompanying statement on gauche usage in the 1970 IUPAC document.[28]

Although a set of recommendations was recently published[29] which states unambiguously that g+ refers to +60° and g− refers to −60° (it is these officially correct assignments that are shown in Fig. 1), most recent authors have avoided confusion by not using g+ and g-. Some studies[9,10] simply list all mean angles to the nearest degree, whereas rebuilding programs such as O [16] and XtalView[17,30] provide examples without any names. Dunbrack and Cohen[11] invented a new nomenclature, using 1, 2, and 3 to represent +60°, 180°, and −60°, respectively; however, 2 represented 0° for symmetrized carboxyls or amides and 1 represented +90° for χ_2 of Phe and Tyr but −90° for χ_2 of Trp. We prefer more mnemonic names without internal inconsistency or literal contradiction of any earlier use. Therefore, as previously,[22,25] we have built on the common use of **t** to represent **t**rans and have extended it to **m** for **m**inus 60° and **p** for **p**lus 60° (see Fig. 1) for the majority of cases where χ angles cluster near these values. When this is not the case (for terminal χ angles in residues with planar functional groups), we use the χ value rounded to the nearest 10° (or occasionally 5° in well-determined cases). Thus, the most common rotamers for Leu are **mt** and **tp**, whereas the most common for tryptophan is called **m**95°. Backbone-dependent rotamers have α, β, or L prepended, and arbitrary sample points have S prepended. The few rotamers with flat distributions 180° wide are flagged with underlines, such as Glu **tt** 0°. When describing this name orally, we use the term "**tt** wide zero." This system of nomenclature is compatible with the new standards,[29] in that **m** stands for g- as well as for minus χ angles.

There remains a minor confusion inherent in the definition of the atom names and the χ angles themselves.[28,29] For Val, the right-hand branch is Cγ1 and therefore used to measure χ_1 from the N, but for Thr and Ile, the heavier left-hand branch is the reference Oγ1 or Cγ1. In Table I, for Val, a note indicates which χ angles would place the side-chains of Thr or Ile into equivalent positions, to facilitate comparisons.

Choice of Structures for the Database

The choice of a database of structures is inevitably a compromise. In order to increase the statistical accuracy of the rotamer library, it is desirable to have as large a database as possible, although the use of lower resolution structures will add noise because of the inclusion of more residues that were built into poorer electron density. One objective of this study was to facilitate breaking the cycle of error propagation from rotamer libraries into newly solved structures and then subsequently from the structures back into later rotamer libraries. Therefore, we have weighted our compromise more towards high-resolution structures than large numbers, using only structures at 1.7 Å resolution or better.

Similar considerations apply when determining the sequence-similarity cutoff applied. The cutoff should be relatively high to allow the inclusion of as many structures as possible, but low enough that the library is not unduly biased toward a particular family of proteins, especially if an incorrect conformation might be reproduced in later structures solved by molecular replacement. Visual inspection indicated that, although two structures with 50% sequence identity share the same fold, their equivalent side-chains usually have different local environments and, therefore, different steric determinants of side-chain conformation. For this reason we included pairs with up to 50% sequence identity. This cutoff is less stringent than Tuffery et al.,[9] who used 25%, but is the same as used by Dunbrack and Cohen.[11]

The set of 240 database structures was chosen from the PDB as of August 1998. Each has a *clashscore* <30 (defined as the number of van der Waals overlaps ≥0.4 Å per 1000 atoms[22]) and a residual (R-factor) of 20% or better. Priority was given primarily to high resolution, with wild-type sequences chosen over mutant when the resolution and clashscore were similar. The following were eliminated: unrefined structures, free-atom refined structures (e.g., 1NXB, 4RXN), structures with no or unrefined B-factors (e.g., 1PIP, 1PPT), and structures in which the sequence was not specified (e.g., 3CTS) or was determined from the electron density map (e.g., 2CSC, 1ALC). If more than one chain with identical sequence was present, we normally chose the first, unless another had a substantially better clashscore or significantly fewer disordered residues.

In order to permit analysis of the correlation of rotamer behavior with resolution, a control set of structures was chosen by similar criteria in the resolution range 1.8 Å–2.5

TABLE I. Rotamers for All Side Chains

Name	#	%	Alpha	Beta	Other	χ_1 mode	χ_1 comm.[a]	χ_2 mode	χ_2 comm.	χ_3 mode	χ_3 comm.	χ_4 mode	χ_4 comm.	χ_1 1/2 Width at 1/2 Height	χ_2 1/2 Width at 1/2 Height	χ_3 1/2 Width at 1/2 Height	χ_4 1/2 Width at 1/2 Height
Arginine																	
ptp85°[b]	3	<1%[c]	0%	1%	<1%		**62**		**180**		**65**		**85**				
ptp180°	11	1%	0%	2%	2%	71	**62**	171	**180**	65	**65**	−161	**−175**	14	17	10	13
ptt85°	16	2%	1%	2%	2%	65	**62**	−178	**180**	−179	**180**	88	**85**	13	14	13	17
ptt180°	16	2%	1%	2%	2%	59	**62**	176	**180**	−178	**180**	−177	**180**	15	13	15	19
ptt-85°	15	2%	1%	2%	2%	66	**62**	−176	**180**	−178	**180**	−83	**−85**	15	14	12	14
ptm180°	6	1%	0%	1%	1%		**62**		**180**		**−65**		**175**				
ptm-85°	5	1%	0%	0%	1%		**62**		**180**		**−65**		**−85**				
tpp85°	11	1%	3%	1%	<1%	−178	**−177**	57	**65**	57	**65**	85	**85**	13	13	12	15
tpp180°	8	1%	1%	0%	1%		**−177**		**65**		**65**		**−175**				
tpt85°	20	2%	3%	2%	2%	177	**−177**	64	**65**	180	**180**	86	**85**	14	13	15	14
tpt180°	15	2%	3%	1%	1%	179	**−177**	60	**65**	178	**180**	163	**180**	13	17	14	17
ttp85°	33	4%	5%	3%	3%	−179	**−177**	177	**180**	65	**65**	83	**85**	14	17	13	15
ttp180°	25	3%	5%	3%	1%	−178	**−177**	−178	**180**	65	**65**	−162	**−175**	14	16	14	26
ttp-105°	9	1%	1%	1%	1%		**−177**		**180**		**65**		**−105**				
ttt85°	19	2%	2%	2%	2%	−175	**−177**	176	**180**	179	**180**	83	**85**	14	14	13	14
ttt180°	33	4%	3%	7%	3%	−179	**−177**	177	**180**	−179	**180**	170	**180**	15	13	12	27
ttt-85°	26	3%	3%	3%	2%	−179	**−177**	179	**180**	180	**180**	−86	**−85**	15	14	14	15
ttm105°	10	1%	2%	1%	<1%	−178	**−177**	170	**180**	−66	**−65**	107	**105**	15	16	15	15
ttm180°	13	1%	<1%	4%	1%	180	**−177**	−178	**180**	−67	**−65**	176	**175**	15	12	11	15
ttm-85°	28	3%	3%	3%	3%	−175	**−177**	−178	**180**	−65	**−65**	−84	**−85**	14	16	15	14
mtp85°	22	2%	2%	3%	2%	−69	**−67**	177	**180**	64	**65**	84	**85**	13	17	13	13
mtp180°	45	5%	4%	3%	6%	−65	**−67**	176	**180**	64	**65**	−174	**−175**	12	19	13	19
mtp-105°	7	1%	0%	2%	1%	−62	**−67**	179	**180**	67	**65**	−113	**−105**	11	15	13	15
mtt85°	34	4%	4%	4%	3%	−67	**−67**	178	**180**	179	**180**	83	**85**	12	19	13	19
mtt180°	89	9%	9%	5%	12%	−67	**−67**	−178	**180**	−177	**180**	174	**180**	14	13	13	21
mtt-85°	53	6%	4%	7%	6%	−66	**−67**	−177	**180**	−179	**180**	−83	**−85**	13	13	13	13
mtm105°	15	2%	1%	1%	2%	−68	**−67**	−179	**180**	−65	**−65**	103	**105**	12	13	13	15
mtm180°	48	5%	1%	4%	8%	−68	**−67**	173	**180**	−64	**−65**	180	**175**	14	17	13	30
mtm-85°	54	6%	13%	2%	3%	−69	**−67**	−167	**−167**	−63	**−65**	−86	**−85**	14	13	13	13
mmt85°	7	1%	1%	1%	1%		**−62**		**−68**		**180**		**85**				
mmt180°	18	2%	1%	3%	2%	−63	**−62**	−66	**−68**	−179	**180**	−168	**180**	13	13	10	29
mmt-85°	22	2%	<1%	4%	3%	−60	**−62**	−72	**−68**	−178	**180**	−92	**−85**	14	13	15	13
mmm180°	11	1%	<1%	2%	2%	−64	**−62**	−74	**−68**	−67	**−65**	172	**175**	14	15	10	13
mmm-85°	22	2%	2%	3%	3%	−62	**−62**	−64	**−68**	−61	**−65**	−82	**−85**	14	13	15	13
		82%	79%	81%	84%												
	769/938[d]		234	146	389												
Lysine																	
ptpt	7	1%	0%	2%	<1%		**62**		**180**		**68**		**180**				
pttp	13	1%	0%	1%	2%	63	**62**	−170	**180**	−177	**180**	72	**65**	13	14	14	11
pttt	29	2%	0%	4%	3%	63	**62**	−178	**180**	178	**180**	−179	**180**	13	13	13	10
pttm	8	1%	0%	1%	1%		**62**		**180**		**180**		**−65**				
ptmt	5	<1%	0%	1%	<1%		**62**		**180**		**−68**		**180**				
tptp	11	1%	1%	1%	1%	179	**−177**	59	**68**	163	**180**	60	**65**	13	12	10	11
tptt	32	3%	5%	1%	2%	179	**−177**	62	**68**	173	**180**	171	**180**	10	10	13	14
tptm	7	1%	1%	1%	<1%		**−177**		**68**		**180**		**−65**	14	9	12	10
ttpp	12	1%	1%	<1%	1%		**−177**		**180**		**68**		**65**				
ttpt	25	2%	2%	5%	1%	180	**−177**	179	**180**	78	**68**	179	**180**	14	12	14	14
tttp	49	4%	5%	5%	3%	−177	**−177**	180	**180**	171	**180**	63	**65**	14	13	12	12
tttt	162	13%	17%	19%	10%	−177	**−177**	178	**180**	179	**180**	180	**180**	13	13	15	13
tttm	37	3%	4%	2%	3%	−177	**−177**	172	**180**	178	**180**	−72	**−65**	12	13	15	13
ttmt	20	2%	2%	4%	1%	−175	**−177**	−174	**180**	−69	**−68**	179	**180**	14	14	10	15
ttmm	5	<1%	1%	0%	<1%		**−177**		**180**		**−68**		**−65**				
mptt	4	<1%	0%	0%	1%		**−90**		**68**		**180**		**180**				
mtpp	12	1%	1%	1%	1%	−69	**−67**	−179	**180**	70	**68**	67	**65**	10	9	10	13
mtpt	38	3%	4%	2%	3%	−69	**−67**	164	**180**	62	**68**	−179	**180**	12	13	11	9
mttp	42	3%	2%	4%	4%	−67	**−67**	−176	**180**	174	**180**	76	**65**	13	13	14	14
mttt	244	20%	23%	14%	21%	−67	**−67**	176	**180**	179	**180**	177	**180**	14	13	12	14
mttm	56	5%	3%	5%	6%	−67	**−67**	−179	**180**	−179	**180**	−63	**−65**	13	12	13	14
mtmt	40	3%	6%	2%	3%	−70	**−67**	−170	**180**	−66	**−68**	−175	**180**	12	13	14	13
mtmm	12	1%	0%	1%	1%	−70	**−67**	−179	**180**	−66	**−68**	−64	**−65**	12	12	12	11
mmtp	9	1%	<1%	0%	1%		**−62**		**−68**		**180**		**65**				
mmtt	77	6%	3%	5%	8%	−58	**−62**	−61	**−68**	−177	**180**	−179	**180**	12	13	13	13
mmtm	18	1%	1%	1%	2%	−59	**−62**	−69	**−68**	−176	**180**	−70	**−65**	14	12	10	15
mmmt	10	1%	<1%	1%	1%	−59	**−62**	−58	**−68**	−75	**−68**	−174	**180**	12	13	10	15
		81%	82%	80%	82%												
	984/1209		261	194	529												

TABLE I. (Continued.)

Name	#	%	Alpha	Beta	Other	χ_1 mode	χ_1 comm.	χ_2 mode	χ_2 comm.	χ_3 mode	χ_3 comm.	χ_3 range[e]	χ_1 1/2 Width at 1/2 Height	χ_2 1/2 Width at 1/2 Height	χ_3 1/2 Width at 1/2 Height
Methionine															
ptp	12	2%	1%	3%	3%	68	**62**	−167	**180**	88	**75**		11	17	12
ptm	17	3%	1%	6%	4%	67	**62**	174	**180**	−78	**−75**		9	10	9
tpp	30	5%	8%	2%	5%	−177	**−177**	66	**65**	75	**75**		10	15	15
tpt	9	2%	1%	4%	1%	179	**−177**	67	**65**	−179	**180**		9	8	9
ttp	28	5%	7%	7%	2%	176	**−177**	178	**180**	73	**75**		10	11	11
ttt	17	3%	5%	2%	2%	180	**−177**	171	**180**	174	**180**		9	9	19
ttm	36	7%	3%	10%	8%	−177	**−177**	176	**180**	−78	**−75**		10	10	13
mtp	92	17%	22%	10%	17%	−68	**−67**	177	**180**	72	**75**		10	12	14
mtt	43	8%	9%	8%	7%	−67	**−67**	177	**180**	−178	**180**		10	13	15
mtm	58	11%	12%	11%	9%	−67	**−67**	−177	**180**	−76	**−75**		12	11	16
mmp	15	3%	3%	1%	4%	−64	**−65**	−63	**−65**	103	**103**		9	10	10
mmt	10	2%	0%	2%	3%	−63	**−65**	−64	**−65**	180	**180**		12	14	19
mmm	105	19%	21%	16%	19%	−66	**−65**	−60	**−65**	−67	**−70**		11	13	16
		86%	91%	84%	83%										
	472/550		175	112	185										
Glutamate															
pt-20°	80	5%	1%	9%	7%	63	**62**	−175	**180**	−18	**−20**	−90 to 90	14	13	23
pm0°	32	2%	0%	0%	4%	71	**70**	−79	**−80**	5	**0**	−50 to 50	14	13	17
tp10°	91	6%	10%	2%	6%	−177	**−177**	65	**65**	13	**10**	−10 to 90	14	13	17
tt0°	350	24%	25%	42%	18%	−177	**−177**	178	**180**	2	**0**	−90 to 90	14	14	30
tm-20°	17	1%	1%	1%	1%		**−177**		**−80**		**−25**	−50 to 10	13	13	15
mp0°	88	6%	<1%	2%	10%	−65	**−65**	85	**85**	−3	**0**	−60 to 60	14	13	25
mt-10°	484	33%	36%	29%	32%	−67	**−67**	177	**180**	−10	**−10**	−90 to 90	13	16	25
mm-40°	197	13%	19%	7%	12%	−65	**−65**	−58	**−65**	−40	**−40**	−90 to 30	14	14	25
		91%	92%	92%	90%										
	1339/1470		394	225	720										
Glutamine															
pt20°	37	4%	1%	5%	6%	64	**62**	180	**180**	20	**20**	−90 to 90	13	14	16
pm0°	15	2%	0%	1%	3%		**70**		**−75**		**0**	−60 to 60			
tp-100°	14	2%	4%	2%	<1%		**−177**		**65**		**−100**	−150 to 0			
tp60°	78	9%	13%	9%	7%	−175	**−177**	64	**65**	60	**60**	0 to 90	14	15	24
tt0°	140	16%	16%	29%	12%	−174	**−177**	173	**180**	−5	**0**	−90 to 90	14	13	40
mp0°	24	3%	<1%	1%	5%		**−65**		**85**		**0**	−60 to 60			
mt-30°	304	35%	40%	26%	36%	−67	**−67**	177	**180**	−25	**−25**	−90 to 90	16	15	37
mm-40°	127	15%	12%	13%	17%	−66	**−65**	−60	**−65**	−40	**−40**	−95 to 0	16	18	26
mm100°	22	3%	4%	1%	2%		**−65**		**−65**		**100**	0 to 150			
		88%	89%	86%	88%										
	761/863		229	137	395										
Aspartate												χ_2 range			
p-10°	203	10%	1%	2%	13%	61	**62**	−4	**−10**			−90 to 0	9	19	
p30°	194	9%	1%	5%	12%	65	**62**	9	**30**			0 to 90	8	14	
t0°	438	21%	8%	44%	20%	−176	**−177**	1	**0**			−50 to 50	12	30	
t70°	118	6%	11%	7%	4%	−179	**−177**	65	**65**			50 to 90	12	18	
m-20°	1088	51%	77%	38%	47%	−71	**−70**	−15	**−15**			−90 to 20	10	16	
		96%	97%	95%	96%										
	2041/2124		365	232	1444										
Asparagine															
p-10°	103	7%	0%	1%	10%	63	**62**	−13	**−10**			−90 to 0	8	9	
p30°	132	9%	<1%	7%	12%	64	**62**	34	**30**			0 to 90	6	7	
t-20°	177	12%	5%	21%	12%	−174	**−174**	−20	**−20**			−120 to 0	5	21	
t30°	228	15%	13%	18%	15%	−168	**−177**	31	**30**			0 to 80	14	22	
m-20°	580	39%	65%	28%	33%	−71	**−65**	−23	**−20**			−60 to 10	10	20	
m-80°	118	8%	8%	9%	8%	−71	**−65**	−76	**−75**			−100 to −60	9	9	
m120°	58	4%	3%	3%	4%	−64	**−65**	132	**120**			60 to 160	9	18	
		94%	95%	88%	95%										
	1396/1490		293	179	924										

TABLE I. (Continued.)

Name	#	%	Alpha	Beta	Other	χ_1 mode	χ_1 **comm.**	χ_2 mode	χ_2 **comm.**	χ_2 range	χ_1 1/2 Width at 1/2 Height	χ_2 1/2 Width at 1/2 Height
Isoleucine												
pp	10	1%	<1%	1%	<1%		**62**		**100**			
pt	216	13%	4%	13%	22%	61	**62**	171	**170**		10	10
tp	36	2%	2%	1%	4%	−169	**−177**	66	**66**		13	11
tt	127	8%	1%	8%	14%	−174	**−177**	167	**165**		13	11
mp	19	1%	0%	2%	1%		**−65**		**100**			
mt	993	60%	81%	58%	41%	−66	**−65**	169	**170**		10	10
mm	242	15%	10%	16%	17%	−57	**−57**	−59	**−60**		10	10
		99%	99%	98%	99%							
	1643/1667		496	629	518							
Leucine												
pp	21	1%	<1%	2%	1%		**62**		**80**			
tp	750	29%	30%	36%	23%	177	**−177**	63	**65**		10	10
tt	49	2%	1%	3%	1%	−172	**−172**	147	**145**	120 to 180	9	9
mp	63	2%	1%	5%	2%	−85	**−85**	66	**65**	45 to 105	11	14
mt	1548	59%	62%	46%	66%	−65	**−65**	174	**175**		11	11
		93%	95%	93%	93%							
	2431/2602		836	644	951							
Histidine												
p-80°	51	9%	0%	6%	13%	60	**62**	−75	**−75**	−120 to −50	10	12
p80°	26	4%	0%	4%	6%	61	**62**	78	**80**	50 to 120	13	10
t-160°	31	5%	5%	14%	1%	−178	**−177**	−163	**−165**	150 to −120	12	20
t-80°	64	11%	17%	9%	9%	−173	**−177**	−81	**−80**	−120 to −50	10	22
t60°	94	16%	24%	17%	12%	−178	**−177**	62	**60**	50 to 120	13	19
m-70°	174	29%	26%	30%	30%	−60	**−65**	−69	**−70**	−120 to −30	11	23
m170°	44	7%	9%	3%	9%	−63	**−65**	165	**165**	120 to −160	10	16
m80°	78	13%	14%	10%	14%	−66	**−65**	83	**80**	50 to 120	11	18
		94%	94%	92%	95%							
	562/598		124	143	295							
Tryptophan												
p-90°	67	11%	2%	13%	14%	58	**62**	−87	**−90**	−130 to −60	12	10
p90°	34	6%	1%	9%	6%	60	**62**	92	**90**	60 to 130	12	8
t-105°	100	16%	27%	10%	14%	178	**−177**	−105	**−105**	−130 to −60	16	14
t90°	109	18%	28%	14%	15%	−178	**−177**	88	**90**	0 to 100	10	11
m-90°	31	5%	0%	7%	7%	−70	**−65**	−87	**−90**	−130 to −60	9	12
m0°	48	8%	15%	2%	8%	−66	**−65**	−4	**−5**	−40 to 20	9	20
m95°	195	32%	22%	43%	29%	−69	**−65**	95	**95**	60 to 130	11	19
		94%	95%	98%	92%							
	584/618		140	175	269							
Tyrosine												
p90°	182	13%	1%	21%	12%	63	**62**	89	**90**	60 to 90, −90 to −60	13	13
t80°	486	34%	55%	25%	30%	176	**−177**	77	**80**	20 to 90, −90 to −75	11	14
m-85°	618	43%	26%	50%	45%	−65	**−65**	−87	**−85**	50 to 90, −90 to −50	11	21
m-30°	124	9%	15%	4%	9%	−64	**−65**	−42	**−30**	−50 to 0, 0 to 50	11	18
		98%	97%	99%	97%							
	1410/1443		290	468	652					(for Tyr, Phe 90° = −90°)		
Phenylalanine												
p90°	202	13%	1%	24%	11%	59	**62**	88	**90**	60 to 90, −90 to −60	11	11
t80°	522	33%	57%	18%	29%	177	**−177**	80	**80**	20 to 90, −90 to −75	13	17
m-85°	697	44%	29%	51%	47%	−64	**−65**	−83	**−85**	50 to 90, −90 to −50	12	17
m-30°	149	9%	12%	5%	11%	−64	**−65**	−19	**−30**	−50 to 0, 0 to 50	9	20
		98%	97%	99%	98%							
	1570/1599		389	514	667							
Proline												
Cγ endo	379	44%	23%	54%	43%	30	**30**			15 to 60	7	
Cγ exo	372	43%	68%	28%	44%	−29	**−30**			−60 to −15	6	
cis, Cγ endo	56	6%	0%	1%	7%	31	**30**			15 to 60	5	
		93%	91%	84%	94%							
	807/928		20	57	730							

TABLE I. (Continued.)

Name	#	%	Alpha	Beta	Other	χ_1 act.	χ_1 com.[a]		χ_1 1/2 Width at 1/2 Height
Threonine									
p	1200	49%	25%	31%	65%	59	**62**		10
t	169	7%	0%	13%	6%	−171	**−175**		6
m	1062	43%	74%	55%	29%	−61	**−65**		7
		99%	100%	99%	99%				
	2431/2447		395	672	1364				
Valine									
p	169	6%	2%	8%	8%	63	**63**	="177"[f]	8
t	1931	73%	90%	72%	63%	175	**175**	="−65"	8
m	526	20%	7%	20%	28%	−64	**−60**	="60"	7
		99%	100%	99%	99%				
	2626/2649		622	1080	924				
Serine									
p	1201	48%	33%	36%	55%	64	**62**		10
t	541	22%	22%	34%	18%	178	**−177**		11
m	714	29%	44%	29%	25%	−65	**−65**		9
		98%	98%	100%	98%				
	2456/2498		350	485	1621				
Cysteine									
p	64	23%	5%	23%	34%	55	**62**		14
t	74	26%	20%	45%	21%	−177	**−177**		10
m	142	50%	75%	32%	43%	−65	**−65**		11
		99%	100%	100%	98%				
	280/285		85	65	130				

[a] "mode" indicates the peak of the smoothed distribution, "comm." indicates the common-atom value (given in bold face).
[b] Mode and 1/2 width at 1/2 height values are not given for minor rotamers.
[c] <1% indicates a value between 0.5% and 0%. 0% indicates no observations.
[d] Total number of rotameric side chains/Total number that pass all data filters.
[e] Ranges used in determining frequencies are normally common-atom values ±30°. Exceptions (always in the terminal χ value) are listed here.
[f] Standard conventions[28,29] result in χ angles being named differently for Val than for Thr and Ile. These figures indicate the equivalent angles.

Å. Clashscore was not considered, and the cutoff on R-factor was relaxed to 24% in order to encompass typical structures in each resolution range. Controls were allowed to be related to proteins in the primary database, but in order to exclude information from those or any higher-resolution structure, a control had to be the highest resolution structure within its protein family at the time it was solved. The resulting annotated list of 240 database files at 1.7 Å or better and 78 controls at 1.8–2.5 Å is available in electronic form as supplementary material (http://www.interscience.wiley.com/jpages/0887-3585/suppmat/index.html) or from our website at http://kinemage.biochem.duke.edu.

Removal of Uncertain Residues

High B-factors can arise for a number of reasons, but all indicate uncertainty in the position of the deposited coordinates. Therefore, we applied a B-factor cutoff, as discussed in the Results section and previously.[22] B-factors are given in almost all PDB files; the only complication is checking for three types of cases where B-factors are ≤1. If no B-factors were assigned (which is very unusual at high resolution), that field is set either to zero or to 1.0; we omitted such files. When explicit H atoms are included sometimes their B-factors are set to zero; for such cases, we assigned the B-factor of the bonded nonhydrogen atom. If atomic displacement values (U^2) were listed (e.g., 1ETN, 2ER7), which produce numbers <1.0, these were converted to the more common temperature factor (B) using the relationship $B = 8\pi^2 U^2$. The whole side-chain was omitted from our database if it had a single atom with a B-factor ≥40.

Water molecules with occupancies <0.67 were not considered, and side chains were omitted if any atom had an occupancy of <1.0 or an alternate conformation flag. It is our experience[22] that these residues show steric clashes significantly more frequently than those modeled with a single conformation. The B conformation of an A/B alternate pair is particularly prone to clash or even to have highly deviant covalent geometry: indeed, these residues were not checked by the quality control programs used by the PDB. For finding clashes with other side chains we used the A conformation only.

Residues were also rejected if any atom had a non-H-bonded atomic overlap of 0.4 Å or more, since either it or its neighbor must be incorrect. Van der Waals overlaps were determined by all-atom contact analysis as implemented in the Probe program.[22] Detailed analysis of van der Waals interactions is not meaningful unless all atoms, including hydrogens, are used. Therefore, prior to running Probe, H

atoms were added geometrically to protein, nucleic acid, and heterogen molecules; their positions were optimized, including combinatorial analysis of local H-bond networks, using the program Reduce.[23] Hydrogens were not added to water molecules, however, which our algorithms allow to provide either H-bond donor or acceptor properties as needed. Note that individual database side chains were tested for clashes within the protein structure, including any bond angle distortions, while later checks of proposed rotamer conformations were done with ideal-geometry side chains in isolation.

The optimal orientation choice for each amide or imidazole was determined by Reduce, considering both H-bond criteria and all-atom van der Waals overlaps, including polar hydrogens.[23] Reduce flags each Asn, Gln, or His as either K (keep in original orientation), F (flip by 180°), X (unknown, i.e., similar score in either orientation), or C (clashing in both orientations). Only those in the "keep" or "flip" categories were used in the Asn, Gln, or His rotamer distributions.

Determination of Distribution Modes

Dihedral angles were calculated with Dang.[31] Rotamer positions were defined as the mode, or highest peak, of the smoothed distribution in χ space (see Results section for rationale). Smoothing was done by placing a Gaussian mask over each data point and summing the mask values at grid points spaced every 1°. The mask had a half-width at half-height of 1° for one-dimensional data, 2° for two-dimensional, 4° for three-dimensional, and 6° for four-dimensional data. Regardless of the dimensionality, each mask had an integral of 1. The rotamer was then defined as the local maximum of the sum of masks.

Once the set of rotamers was defined as modal χ values, each residue type was re-examined to see which rotamers could satisfactorily be defined as having some common atoms (produced by common χ values). The criteria were

1) whether the data unequivocally demonstrated that the angles in question differed or whether a common value could fit all acceptably;
2) the extent to which the atomic contacts (for ideal geometry) occurred at similar χ angles;
3) whether the conformations had an inherent symmetry (e.g., to set absolute values equal for **mm** and **pp** if they made no nonequivalent backbone interactions, or to use 180° as the default **t** angle if the preceding angle was also **t**). χ_1 angles were also similarly considered across classes of residues.

Half-widths (analogous to standard deviations as used with means) were defined as the angular distance plus or minus from the modal value at which the summed mask function is half of the maximum for that peak. The artifactual broadening caused by the mask width was corrected according to the following scheme:

$$\text{corrected width} = \sqrt{((\text{distribution width})^2 - (\text{mask width})^2)}$$

Half-width at half-height can be converted to standard deviation, if the distribution is normal, by dividing by 1.1774 (for a normal distribution the height at 1σ is 0.606, so that the half-width is larger than σ). We have found, however, that almost all of our rotamer distributions are in fact platykurtic (i.e., flatter-topped and steeper-sided than a normal distribution), so that a standard deviation calculated from the set of points would be even smaller than the above estimate. The average half-width at half-height is given in Table I, but, for the analysis of skew, separate half-widths above and below the mode are listed in the supplementary material.

For all amino acids scatter-plot kinemages of the raw χ angle distributions, the modes, and the contours[31] for the summed mask functions were displayed in Mage.[32–34] For Arg and Lys, a 3-D kinemage was made for each χ_1 (**p**, **m**, and **t**). Multiple peaks and asymmetries were evaluated; since each point carries its identity (file and residue) in the kinemage, a sample of outliers was identified and examined in the context of their 3-D structures. Bin boundaries for counting frequencies were defined as the common-atom angle $\pm 30°$ rounded to the nearest 5° unless listed explicitly in Table I; those exceptions are wider bins for angles with broad distributions and a few narrower bins to avoid rotamer overlap. Since the bins do not include all of torsion space, the rotamer probabilities sum to a number less than 100%, which is considered the "rotamericity"[7] of that residue type.

Once the boundaries were determined, the probability (% occurrence) for each rotamer overall and in each secondary structural class (helix, sheet, and other) was found. Secondary-structure assignments were from a modification of DSSP as implemented in ProCheck;[18] a residue was counted as helical if it is given the strict H assignment by ProCheck and is not in the first three H's of a run (the less restrictive first turn), as beta if given the E assignment, and as falling into the "other" category in remaining cases. Left-handed (L) residues are those with $0° < \phi < 175°$. Significant changes in the rotamer frequencies are discussed in the Results section. On testing for differences in modal positions as a function of secondary structure, however, none shifted significantly except for Asn and Asp, which were, therefore, given a set of backbone-dependent rotamers (Table II).

Within especially broad distributions, a few additional sample points (Table III) were chosen by visual inspection of the 2-D or 3-D distributions. These sample points were defined such that they lay within the highly populated regions of the distribution tail, 30–60° away from the position of the actual rotamer and in a nonclashing conformation. Common-atom angles are used whenever this gives a reasonable agreement with the data.

Each amino acid was built using Engh and Huber[35] ideal geometry in a coordinate system with the Cα at the origin, N along the X-axis, and Cβ in the XZ plane. Hydrogens were added with Reduce[23] and a kinemage made in Prekin[32] with rotatable angles. Each defined rotamer was examined in Mage, at both the modal and common-atom χ values, with all-atom contact dots calcu-

TABLE II. Backbone-Dependent Rotamers for Asp and Asn

Name	#	%	χ_1 mode	χ_1 **comm.**[a]	χ_2 mode	χ_2 **comm.**	χ_1 range	χ_2 range	χ_1 1/2 width at 1/2 height	χ_2 1/2 width at 1/2 height
Aspartate[b]										
αm-10°	283	75%	−72	**−70**	−14	**−10**	−100 to −40	−60 to 10	9	11
αt60°	72	19%	−176	**−177**	63	**60**	155 to −145	−20 to 90	12	14
	355	95%								
βm-20°	92	38%	−66	**−65**	−21	**−20**	−95 to −35	−90 to 20	10	20
βp10°	14	6%	65	**65**	13	**10**	35 to 95	−20 to 40	9	11
βt-10°	130	53%	−176	**−177**	−10	**−10**	155 to −145	−90 to 90	9	20
	236	97%								
Lm-30°	54	61%	−64	**−65**	−29	**−30**	−95 to −35	−90 to 0	9	16
Lt30°	26	29%	−162	**−165**	43	**30**	170 to −130	0 to 60	9	18
	80	90%								
Asparagine[b]										
αm-20°	204	66%	−72	**−70**	−17	**−20**	−100 to −40	−60 to 10	9	14
αm-80°	26	8%	−72	**−70**	−81	**−80**	−100 to −40	−100 to −60	13	17
αm120°[c]	9	3%		**−70**		**120**	−100 to −40	60 to 160		
αt60°	38	12%	−175	**−177**	64	**60**	155 to −145	30 to 80	11	21
αt-60°	14	5%		**−172**		**−60**	155 to −145	−120 to 0		
	291	94%								
βp60°	17	8%		**65**		**60**	35 to 95	−20 to 100		
βm-50°	74	36%	−66	**−65**	−49	**−50**	−95 to −35	−90 to 0	10	26
βm120°	7	3%		**−65**		**120**	−100 to −40	60 to 160		
βt10°	77	38%	−179	**−177**	11	**10**	150 to −150	−90 to 90	6	10
	175	86%								
Lm-30°	91	55%	−65	**−65**	−30	**−30**	−95 to −35	−70 to 10	9	18
Lt30°	58	35%	−166	**−165**	32	**30**	165 to −135	0 to 60	10	12
	149	90%								

[a] "mode" indicates the peak of the smoothed distribution, "comm." indicates the common-atom value (bold face).
[b] For "other" secondary structural class, use the backbone-independent rotamers.
[c] Mode and half-width at half-height are not given for minor rotamers.

lated interactively by Probe.[22,31] Rotamers that exhibited any significant van der Waals overlaps within the side chain or with the fixed N, C, or Hα atoms were analyzed in detail and are discussed individually in the Results section.

RESULTS

The complete side-chain rotamer library is given in Table I, including frequencies, secondary structural preferences, χ angle values, and half-width at half-height (refer to Methods for rotamer nomenclature). Amino-acid types are listed in order of their number of χ angles, with backbone-dependent rotamers for Asp and Asn in Table II. For each amino acid, the library includes only those rotamers which occur consistently at high resolution and low B-factor and which show suitable clustering around plausible local energy minima. For Arg, Lys, and Met, this results in rotamers for 1/3 to 1/2 of the total staggered-angle combinations (34/81, 27/81, and 13/27, respectively).

Table III additionally lists a small set of conformations which may be used to sample the occupied regions of torsion space more uniformly, chosen to be well inside the tails of the few especially wide distributions. For use in a method with a radius of convergence significantly smaller than 20–30°, a more closely spaced grid of sample points could be defined throughout the populated regions of χ space. It should be noted that extra sample points do not correspond to cluster peaks or energy minima and are not, therefore, rotamers.

Rotamer χ angles are listed both as the modal (peak) values found for the individual distribution and also in a version which optimizes common χ values (and therefore common atom positions) among rotamers with related geometries, such as ±85°, ±105°, ±175°, or 180° for the various classes of Arg χ_4 values. Use of common-atom values improves efficiency in combinatorial calculations such as Monte Carlo or Dead-End Elimination repacking methods (for example[13]). Common-atom χ values have even more important beneficial effects, however, both for calculations and for fitting side chains in structure determinations, by preventing a choice between rotamers based on differences that are not statistically significant. Rotamer positions have usually been given to the nearest degree simply because those are the units in which they are measured, even though the best cases are not known more accurately than 2–3° and the rarer ones only to perhaps 10°. Omitting cases with additional confounding problems (Asn, Gln, Asp, Glu), that level of accuracy can be substantiated by comparing the most up-to-date compilations (the present work and that of Dunbrack and Cohen[11]), or by comparing what should be symmetrically equivalent cases within either of these studies. Table I quotes modal values to the nearest degree for each rotamer in order to document the data (except when there were <10 observations).

TABLE III. Additional Sample Points in Torsion Space

Name	χ_1	χ_2	χ_3
Glutamate			
Spt-60°	**62**	**180**	**−60**
Spt60°	**62**	**180**	**60**
Stt-60°	**−177**	**180**	**−60**
Stt60°	**−177**	**180**	**60**
Smt-60°	**−67**	**180**	**−60**
Smt60°	**−67**	**180**	**60**
Smm0°	**−65**	**−75**	**0**
Glutamine			
Spt-60°	**62**	**180**	**−60**
Spt60°	**62**	**180**	**60**
Stt-60°	**−177**	**180**	**−60**
Stt60°	**−177**	**180**	**60**
Smt-60°	**−67**	**180**	**−60**
Smt60°	**−67**	**180**	**60**
Aspartate			
Sp-50°	**62**	**−50**	
St-30°	**−170**	**−30**	
Sm-60°	**−65**	**−60**	
Asparagine			
Sp-50°	**62**	**−50**	
St-80°	**−174**	**−80**	
Phenylalanine			
Sm30°	**−85**	**30**	
Tyrosine			
Sm30°	**−85**	**30**	

However, we feel that the common-atom values (boldface) are preferable for almost all uses (perhaps augmented with the suggested sample points given in Table III) and that they are likely to prove more nearly correct when judged by more accurate future data sets.

Additional information is available in electronic form. As supplementary material to this article, there is a more complete version of Table I which includes explicit bin boundaries and the (sometimes asymmetric) half widths for all χ angles; a version of the table with common atoms and with sample points folded in, which we recommend for applications such as dead-end elimination; and the list of files that make up the high and medium resolution databases. On our website (http://kinemage.biochem.duke.edu) there are PDB-format coordinate files and kinemages[32–34] with rotatable χ angles for all amino acids, in standard geometry in a common coordinate system; PDB-format coordinate files and kinemages with standard geometry side-chains in rotameric conformations; and the actual multidimensional data distributions in kinemage format, with bins and assigned rotamers marked and each data point identifiable. Files suitable for use with the crystallographic fitting programs O[16] and XtalView[17,30] are available both as supplementary material and from our website.

Effect of Filters

It is common practice for compilers of rotamer libraries to use only high-resolution structures, most often with a cutoff at 2 Å rather than the 1.7 Å used here. It is also common practice, once a structure has been chosen, to use all of its residues unless missing atoms mean the χ angles are undefined. However, within a given structure the quality of the electron density map often varies greatly, and the less reliable regions can easily be identified by high B-factors, alternate conformations, or low occupancies. Significant atomic overlaps also indicate local problems. We find that the use of these local quality indicators is crucial. For instance, we have shown[22] that a side-chain with a B-factor above 50 is 10 times more likely to have a bad steric clash than one with a B-factor in the range of 10–20.

The B-factor cutoff is both the single most powerful and the simplest filter applied in this study. The effect on cleaning up the data, as shown for Lys in Figure 2, is dramatic. Surprisingly, previous to our work, a B-factor cutoff had only once been applied in published analyses of side-chain conformations,[27] and that study was not aimed at producing a library of rotamers. Absolute values of B-factors are not directly comparable between structures due to variations in data reduction, solvent treatment, estimates of intensity falloff, and application of either global or local B-restraints. It is possible to compensate partially for such differences by normalizing B-factors by the mean and standard deviation in each structure.[36,37] We feel it is preferable, however, to use the simpler absolute values, for two reasons: a high B-factor will smear out calculated electron density, unless artificially resharpened, no matter what its origin, while differences in the actual level of molecular disorder are often larger than the methodological effects. Many atomic-resolution structures have no disordered loops and thus, for good reason, no high B-factors. Within our data set, the five structures with the lowest average B had a mean clashscore (number of clashes $\geq$0.4 Å per 1000 atoms) of only 4.8, while the five with highest B had a mean clashscore of 23.3, confirming that absolute B-factors are a meaningful indicator of accuracy. For a comparison test we normalized the B-factors for our data set, finding that a cutoff of 1.91 standard deviations removes the same number of residues as a simple B-factor cutoff of 40. Overall, as judged by their efficiency at excluding problematic residues, the methods are nearly equivalent, since about 50% of their omitted residues have a serious clash. However, if we compare the subset that differs between the two methods, those uniquely discarded by the absolute B-cutoff have clashes in 34% of cases, whereas those uniquely discarded by the normalized cutoff clash in only 27% of cases. Methodological effects are certainly larger at low resolution, but for our purposes and our data set, an absolute cutoff has the advantage in performance as well as in simplicity.

B-factors for an individual side chain may be high for various reasons, including thermal motion, static disorder, or phase problems. One of the most important reasons, however, is the possibility of a side-chain misfitting. Refinement of a misfit side chain can either move the atom back into density or increase the B-factors, depending on the details of the local environment and the weights of the B-factor and other restraints. Whatever the cause, the conformation of a high B-factor side chain is less reliable

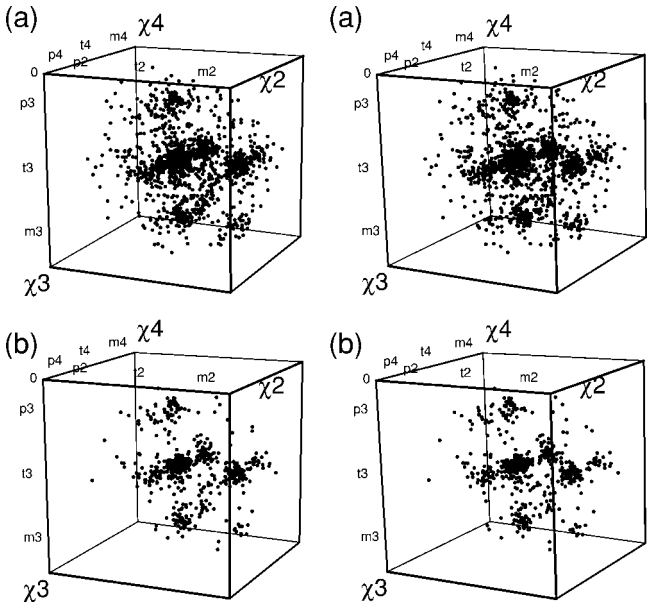

Fig. 2. Three-dimensional scatter plot in stereo of χ_2, χ_3, and χ_4 values (as displayed in Mage) for lysines with χ_1 **m**: (**a**) raw data and (**b**) after removal of residues with B ≥ 40.

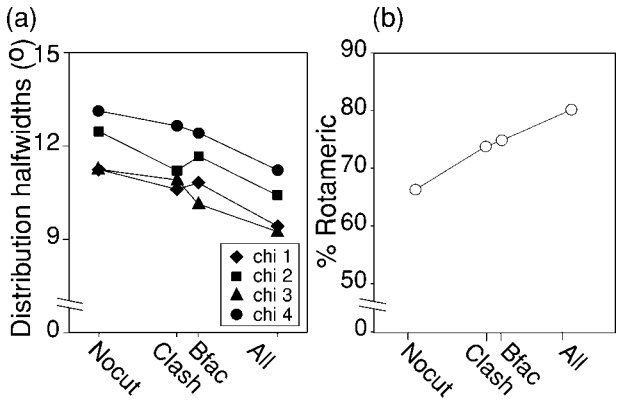

Fig. 3. The effect on (**a**) mean distribution half widths and (**b**) rotamericity for all Arg residues in our database, shown with no cutoffs applied (**Nocut**), removal of clashing residues only (**Clash**), removal of residues with B ≥ 40 only (**Bfac**), and all filters applied together (**All**). Note that the lines are intended to guide the eye and have no physical meaning.

than for a low B-factor equivalent and should not be included in an analysis that depends on accurate details. As a compromise between size and accuracy of the database, here we omit side-chains with any B-factor ≥40 (as in Figure 2b) as well as those with alternate conformations,[22] those with missing atoms, and those with a steric clash ≥ 0.4 Å (since a large clash means that either the side chain or its clash partner must be in error).

The final modifications to the data are the correction of flipped side-chain amide orientations[23] and the omission of any that remain indeterminate or clash in both orientations. This process is essential for obtaining well-clustered rotamers for Asn and Gln,[25] and it has small, indirect effects on other residues. His rings also are sometimes flipped when required by all-atom clashes or by the analysis of local H-bond networks, though in those cases the interactions are mostly long-range. Since long-range interactions seldom cause a systematic change in the flip state, the set of His rotamers is little affected. Application of all filters reduced the database from 40190 to 26374 residues.

Figure 3 shows the quantitative benefits of applying the B-factor and clash filters to the conformations of arginine. Each filter is effective alone and in combination even more so. The χ angle distributions tighten and the proportion of side chains in a rotameric conformation (within the defined bins, usually common-atom angle ±30°) increases significantly. Our distributions are narrower than in either of the only two libraries that quoted widths: for example, the weighted average of all χ_1 standard deviations is 15.3° for Ponder and Richards,[6] 12.4° for Dunbrack and Cohen,[11] and 8.6° for this work (average χ_1 half width of 10.1° divided by 1.1774 to convert to standard deviation, assuming a normal distribution; this is a conservative estimate, as explained in Methods).

For all 18 movable side-chain types, we find 81–99% to be rotameric; if Lys, Arg, and Met are analyzed for B<30 then all types are ≥88% rotameric. Overall, 94.5% of movable side chains in our dataset are rotameric. This certainly seems high enough to confirm further the usefulness of the rotamer concept.

Means *Versus* Modes

Another significant departure from all earlier studies is our use of modes rather than means to define rotamer positions, an approach which has four important advantages. First, there is no assumption of a Gaussian shape to the distribution, which is implicit when mean and standard deviation are calculated. Many χ angle distributions are slightly skewed, with some showing strong skew, particularly for terminal χ angles of those side chains that have planar functional groups or where the rotamer is close to a clashing position. Examples include the Arg rotamers with χ_4 ±85°, where the guanidinium clashes with Hδ if the angle changes to ±70°. The Pro distribution is also quite skewed because of a broad scatter caused by the ring incorrectly being fit as planar. Where Pro rotamers have been defined previously,[6,8,9,11,17] mean χ_1 values for the nonplanar conformations are near ±25° rather than the ±30° of our modes and of small-molecule values.[38] Modal values locate the most preferred conformation reliably in these cases, while nonsymmetric halfwidths can give an indication of the skewedness.

A second advantage of modes is that no a priori assumptions need be made about number and location of peaks in the distribution, whereas prior to calculating means and standard deviations, it is necessary to divide the data into bins. Inappropriate boundaries between bins can lead to inclusion of data in the wrong bin, pulling both means away from their true positions. For example, we have shown that amide χ modal positions most commonly occur near ±30°,[25] while previous treatments have drawn bins around a priori assumed means at ±90°,[5] 0° and ±60°,[11] or

±165°.[36] With such discrepant definitions, means sometimes merely represent the centers of bins, giving little information about preferred conformations.

The third advantage of modes arises when two or more peaks are close together, such as for the leucine χ_1/χ_2 case discussed below. Drawing bins at 0°, 120°, and -120° puts two separate peaks in the **tt** and in the **mp** regions. The mean for each of these two regions lies in between the clusters, which has resulted in clashing Leu **tt** and **mp** rotamers for every previous library. In contrast, determining the modes shows two distinct peaks 60–70° apart which can be analyzed separately, as done below. Genuinely distinct peaks occur in close proximity to each other even more often if individual χ angles are analyzed in one dimension, producing misleading mean values. In those cases, however, modes are helpful but the best solution is use of the appropriate multidimensional treatment.

Lastly, if the observed distribution is converted to an energy equivalent, it is the mode rather than the mean that corresponds to the lowest-energy conformation.

A disadvantage of the modal-value approach is that it requires smoothing to determine the mode reliably (see Methods). Then, to determine a correct width for the smoothed peak, the effect of the smoothing function must be subtracted, which in this implementation means subtracting the mask width as the root difference of squares. In addition, for clusters with low total population, both mean and mode are susceptible to statistical fluctuations, but the mode is somewhat more so. The common-atom angle definitions, although adopted for other reasons, also avoid most of the small-population problems.

It has recently been found [37] that rotamer mean values change systematically with resolution, at least in part because of averaging between unresolved alternate conformations, which produces skewed distributions at lower resolution. In one case (Leu), part of the shift is caused by a misfitting more common at low resolution (see below), but the general point remains valid and important. Although anomalous behavior of the means uncovered this interesting relationship, the modal values as seen in the data described in that study do not shift from the rotameric positions, again suggesting that modes are preferable for most purposes.

Backbone Dependence—Asp and Asn

All side chains were examined for backbone dependence of their rotamers. For most amino acids, the relative frequencies of some rotamers changed significantly between secondary-structural classes, but the position of the peaks did not. Therefore, Table I lists the probabilities in α, β and "other" classes, along with the position of the rotamer they share. The largest and most consistent frequency changes are the often-noted lack of χ_1 **p** conformations in helix for all amino acids other than Ser and Thr.[5] Ser, Thr, Asp, and Asn have quite high probabilities of χ_1 **p** in the "other" secondary structural class, primarily because of the H-bonding in pseudoturns and helix N-caps.[25,39,40] For Phe and Tyr, χ_1 **p** is more common in β sheet because of favorable interactions with the neighboring strand.[41,42] The aromatic rotamers with χ_2 near zero are significantly more common in helix,[7] while Ile **tt** is quite common in "other" but essentially forbidden in helix because of a clash with the backbone.

There are, however, two amino acids (Asp and Asn) for which modal positions as well as probabilities are highly dependent on backbone conformation. Both are small, polar, and interact strongly and specifically with the local backbone. Their backbone-dependent rotamers are given in Table II for the α, β, and left-handed classes ("other" rotamers are essentially the same as the backbone-independent ones), while the distributions and clustering for Asn have been published previously.[25] Asn is in a left-handed backbone conformation ($0° < \phi < 175°$) in 11% of examples, the highest occurrence for any non-Gly residue, and Asp is the next highest with 4%; in both cases only two tightly clustered conformations are found. Asp rotamers are essentially the same as Asn except truncated to ±90° in χ_2 by the symmetry of the carboxyl group. Local H-bonds are influential, and are similar in both cases, except for the Asn Nδ i-4 H-bond in α-helix which is, of course, not possible for Asp and results in the absence of the α**m**-80° rotamer for Asp.

It is likely that further division according to backbone dependence would make additional trends apparent: for example, dividing residues on β-strands according to whether the neighboring strands are parallel or antiparallel. However, for the current data that would involve further division into classes with too few members for statistical validity.

Lysine – The Statistically Simple Side Chain

Lys and Arg, with four χ angles each, have 81 possible staggered rotamers; Met, with three χ angles, has 27. It is worth exploring whether a compact description would suffice, with just a few rules that applied to multiple cases. The attempt failed for Arg and Met, where analogous sets of rotamers show relative frequencies differing by factors of three or more, presumably responding to circumstances such as nonuniform patterns of possible H-bond partners. However, Lys rotamers show reproducible patterns of relative frequencies that can be accurately predicted using only a few physically reasonable parameters, as shown in Table IV. Two parameters are the relative preferences for χ_1 **t** (0.65) and **p** (0.13) as a fraction of **m**. Two additional parameters are penalties for the "syn-pentane"[43] conflicts that occur when adjacent gauche angles change signs (**mp** or **pm**); one of those penalty factors (0.1) applies for χ_2/χ_3 or χ_3/χ_4 on the unbranched side chain, and a more severe one (estimated as 0.05) applies for χ_1/χ_2 which has backbone atoms on one end. Such syn-pentane cases also result in shifted χ values to avoid the clash, such as $\chi_1 = -90°$ for Lys **mptt**, $\chi_3 = 103°$ for Met **mmp**, $\chi_2 = -80°$ for Glu **pm**0° or **tm**-20°, or $\chi_2 = 100°$ for Ile **pp** or **mp**.

The most interesting parameter is the penalty for having a gauche angle in χ_2,χ_3, or χ_4. It can be estimated separately for one-gauche, two-gauche, and three-gauche rotamers relative to the cases where χ_2,χ_3, and χ_4 are trans, avoiding any comparisons that contain **mp** or **pm**

TABLE IV. Lysine Rotamer Simplified Predictions

	Pred	Obs		Pred	Obs		Pred	Obs
pppp	0	0	tppp	**1**	**2**	mppp	0	0
pppt	0	0	tppt	**6**	**3**	mppt	.5	0
pppm	0	0	tppm	0	0	mppm	0	0
pptp	0	0	tptp	**7**	**11**	mptp	.6	0
pptt	.4	0	tptt	**32**	**32**	mptt	**2**	**4**
pptm	0	0	tptm	**7**	**7**	mptm	.6	0
ppmp	0	0	tpmp	0	0	mpmp	0	0
ppmt	0	0	tpmt	.7	0	mpmt	0	0
ppmm	0	0	tpmm	0	0	mpmm	0	0
ptpp	**1**	**1**	ttpp	**7**	**12**	mtpp	**11**	**12**
ptpt	**6**	**7**	ttpt	**32**	**25**	mtpt	**49**	**38**
ptpm	0	0	ttpm	**1**	**2**	mtpm	**1**	**1**
pttp	**7**	**13**	tttp	**37**	**49**	mttp	**56**	**42**
pttt	**32**	**29**	tttt	**161**	**162**	mttt	**244**	**=244**
pttm	**7**	**8**	tttm	**37**	**37**	mttm	**56**	**56**
ptmp	0	0	ttmp	**1**	**0**	mtmp	**1**	**2**
ptmt	**6**	**5**	ttmt	**32**	**20**	mtmt	**49**	**40**
ptmm	**1**	**2**	ttmm	**7**	**5**	mtmm	**11**	**12**
pmpp	0	0	tmpp	0	0	mmpp	0	0
pmpt	0	0	tmpt	0	0	mmpt	**1**	**0**
pmpm	0	0	tmpm	0	0	mmpm	0	0
pmtp	0	0	tmtp	.4	0	mmtp	**11**	**9**
pmtt	.4	0	tmtt	**2**	**0**	mmtt	**49**	**77**
pmtm	0	0	tmtm	**.4**	**2**	mmtm	**11**	**18**
pmmp	0	0	tmmp	0	0	mmmp	**.2**	**2**
pmmt	0	0	tmmt	**.3**	**1**	mmmt	**10**	**10**
pmmm	0	0	tmmm	**0**	**1**	mmmm	**2**	**1**

Rules: each gauche χ_2 or χ_3 = 0.2 factor (= 0.95 kcal); each gauche χ_4 = 0.23 factor (= 0.85 kcal); χ_1 **t** = 0.65 (= 0.25 kcal); χ_1 **p** = 0.13 (= 1.20 kcal); χ_2/χ_3 or χ_3/χ_4 **mp** or **pm** = 0.11 additional (= 1.30 kcal); χ_1/χ_2 **pp, tm, mp,** or **pm** = 0.05 additional (= 1.75 kcal).

combinations. Those factors are found to be 0.21, (0.22) squared, and (0.20) cubed, suggesting that the parameters are independent and simply multiplicative. Also, the experimentally measured energy difference of 0.89 kcal/mol between gauche and trans butane[44] can be converted using the Boltzman relationship

$$E = RT \ln P = 0.592 \ln P = 1.364 \log P$$

to give a gauche factor of 0.22.

An overall least-squares fit of the parameter values to all 81 Lys rotamer frequencies was done by minimizing the sum of squares using Mathematica.[45] Since the Lys NH_3^+ terminal group is smaller than a methyl, six parameters were used, and the gauche penalty was fit with one value for χ_2 and χ_3 and a separate value for χ_4, which came out as 0.20 and 0.23, respectively. The predictions in Table IV are obtained by multiplying all applicable factors for each rotamer; the correlation coefficient between predicted and observed (Pearson's r) is .993. An estimate of the relative rotamer pseudo-energies can be made by adding up the energies for each applicable penalty factor, so that, for instance, **mpmp** acts as though it is 6 kcal/mol less favored than **mttt** and does not occur in our data set.

Note that the strong preference of Lys for trans χ angles (about 1:5:1 **m:t:p**) is real and is not a result of fitting disordered side chains as trans. Not only are the ratios consistent across all rotamers, but also the contrast is lower, not higher, at high B and is significantly lower for χ_4, consistent with its small physical size but not with an effect from increasing uncertainty. In contrast, Met χ_3 prefers gauche by more than 2:1, since the gauche form not only does not clash, but actually has favorable H-atom van der Waals contacts.[22]

Presumably the reason Lys behaves in such a statistically simple fashion is that although the end makes charged H-bonds, the geometry of those interactions is relatively unconstrained, with the side-chain having so many degrees of freedom that it can usually get to its appropriate position without strain. Given that the high-resolution, low-B lysines very seldom have any χ angles as much as 30° from staggered and populate the less-favored rotamers only as often as dictated by their pseudo-energy differences, it seems completely unjustified ever to fit partially disordered lysines with eclipsed angles or poor rotamers.

Systematic Fitting Errors—Effects on Leu, Val, Asn, Gln, and Met

It has long been known that there are enhanced probabilities of making particular types of errors when fitting side-chain conformations. For example, Fourier transform termination ripples even at 2 Å resolution can make the electron density at Cβ rather weak, giving the density for Val or Thr a flat, barlike shape which can be fit equally

well with a standard rotamer or a conformation flipped by 180°. As pointed out previously,[17,46,47] such density should never be fit eclipsed with respect to the backbone. Such errors are not corrected by traditional refinement and only sometimes by molecular dynamics, and they may well be overlooked in manual rebuilding. Multiple misfittings of this sort can lead to minor peaks in the χ_1 distribution offset 180° from the real ones; indeed $\chi_1 = 120°$ has been defined as a Val rotamer in two libraries.[7,10] This conformation is at a local energy maximum not a minimum; in addition to the eclipsed torsion, it has a Cγ -to-C van der Waals overlap of 0.83 Å when built in standard geometry. Our high-resolution, low-B Val data show no peaks near the eclipsed values, but only an extremely sparse scatter of outliers through the nonrotameric region >30° from the staggered values. We feel that inclusion of such eclipsed conformations in a library of rotamers is unjustifiable, since there is a known mechanism for them to occur as occasional errors, but no way to argue for them as occupying a valid minimum.

For the case of Asn or Gln side-chain amides, there is almost no difference in electron density between two orientations flipped by 180°, so that correct assignment requires careful analysis of H-bonds and NH_2 clashes. We previously treated this issue in detail,[23] showing that flipped rotamers with impossibly large internal clashes of the Hδs or Hεs appear in most rotamer libraries.[25] Our previously defined Asn and Gln rotamers, reproduced in Tables I and II, do not suffer from this problem.

For Leu there are only two conformations, **mt** and **tp**, where one arm or other of the side chain is not in a syn-pentane conflict with the backbone. These are the two rotamers that strongly dominate Leu distributions (see Fig. 4), by factors of two to seven over the next-most-common rotamer in earlier work and by the overwhelming factors of 30 and 15 in our compilation. In Leu **mt** and **tp,** one Cδ is trans-trans to one direction of the backbone while the other is over the Hα. Other conformations are allowed only by moving the dihedrals away from staggered positions, explaining their much less frequent occurrence. It has been pointed out before[48] that the preference of Leu for α-helix could arise entropically from the fact that its two generally allowed conformations are both also allowed in helix. The current data would make that effect an even stronger one. As can be seen in Figure 4, there are some Leu examples in the **pp**, **mm**, and **tm** regions. The **pp** cluster is well-behaved and its occurrence increases slightly at low B (Figure 5a), so that it has been defined as a rotamer here. **mm** and **tm** have not, since they cluster poorly and have fairly flat plots of occurrence vs B; however, they do not have bad clashes and may later prove acceptable although relatively unfavorable.

The most complex and interesting cases are the Leu **tt** and **mp** regions. Leu has two pairs of conformations that occupy approximately the same physical space: **mt** vs **mp*** and **tp** vs **tt***, whose χ distributions are shown in Figure 4 and whose conformations are shown in Figure 5b. The ability to superimpose the Cδ atoms of Leu by rotating χ_1 by 30° to 40° and χ_2 by 140° to 150° from some starting

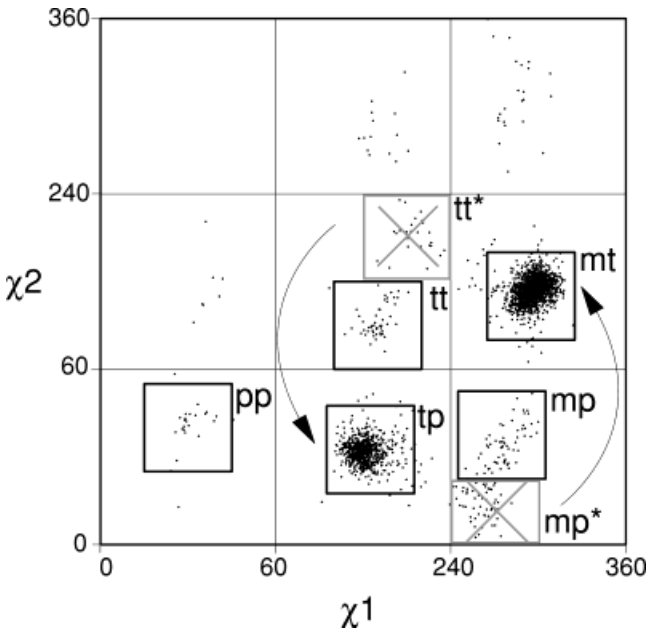

Fig. 4. χ_1/χ_2 distribution for leucine. Black boxes indicate the regions used for determining the frequency of each rotamer and are the common-atom angle ±30° unless otherwise indicated in Table I. Gray boxes (with X's) indicate the regions occupied by the systematically misfit conformations **tt*** and **mp***. Arrows indicate which common rotamer the misfit conformations approximately mimic.

positions has been noted before, mainly for **mt** vs **mp***.[37,49–51] However, none of those authors reached a conclusion as to which of the apparently equivalent conformations is preferable. Indeed, Petrella et al.[51] calculated the energies of the two conformations to be within 1.9 kcal of each other and concluded that "it is unclear whether one or the other [conformation] represents the true crystal position, or whether both are, in fact, correct." On the other hand, Kuszewski et al.[52] discussed the probability of Leu misfittings and changed the less common **mp*** and **tt*** forms by 40° and 140° in their data, but they gave no additional evidence besides the inherent plausibility of that decision. Here we will analyze three other sources of information that can resolve this ambiguity.

For each of the above pairs, one conformation is one of the two highly favorable major rotamers, while the other alternative has a severe clash when built in standard geometry with explicit hydrogens. As shown in Figure 5b, **mp*** has an atomic overlap of 0.6 Å between the Cδ1 and the Hα, and **tt*** has an overlap of 0.7 Å between Cδ2 and the Hα.

The described transformations approximately superimpose the Cδ atoms but not the Cγ, so that Cγ should fit the electron density less well in the flipped conformations. This would lead to the refined B-factors for the Cγ being higher than those for the Cδs, rather than the normal pattern of B-factors increasing out along the side chain. In our data, for the rotamers that appear to be genuine, the mean B-factor for the Cγ is lower than the B-factor of the Cδs in the majority of cases (69% for **mt**, 67% for **tp**, 64% for **tt**, 72% for **mp**). This is true in only 20% of cases for **tt***

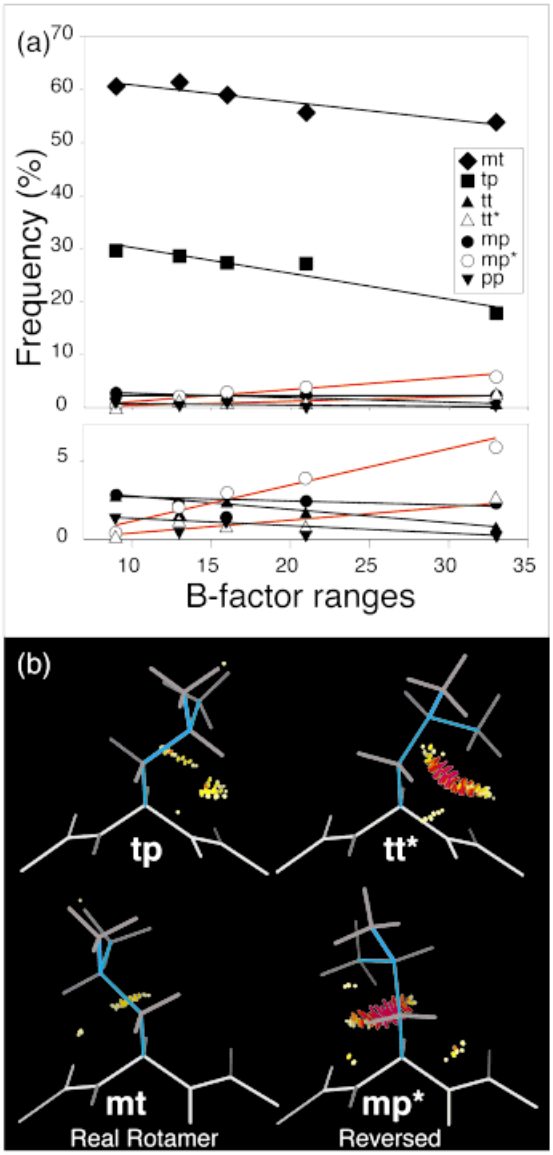

Fig. 5. **a**: Correlation of rotamer frequency with B-factor for both genuine and misfit Leu rotamers. B-factor bins were constructed to contain the same number of points in each bin for the whole distribution. The % frequency of the rotamer in each B-factor bin is plotted in the Y-direction, at the X position of the mean B-factor for that bin. The lower panel is an enlargement of the bottom section of the main plot to show more clearly the slope of the lines for the rarer rotamers. Systematically misfit rotamers (**tt*** and **mp***) are indicated by open symbols and red lines. **b**: A comparison of the structures and their contacts for genuine rotamers (left) versus their misfit partners (right). Blue and green dots indicate positive van der Waals interactions, yellow lines indicate modest (still favorable) van der Waals overlaps, and orange or red lines indicate van der Waals clashes, as calculated with Probe and displayed in Mage.

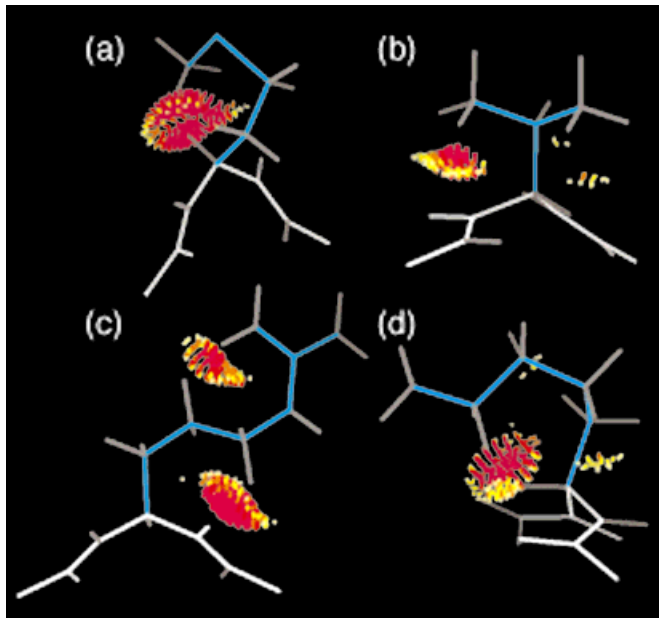

Fig. 6. Examples of rotamers, previously published or currently in use, which show serious internal van der Waals clashes when built in standard geometry.[35] (**a**) Met **tpm**,[16] (**b**) Val 120°,[7] (**c**) Arg **tmtp**,[10] and (**d**) Lys **mmpt**.[9] Only the van der Waals overlaps are shown (orange and red), as calculated with Probe.

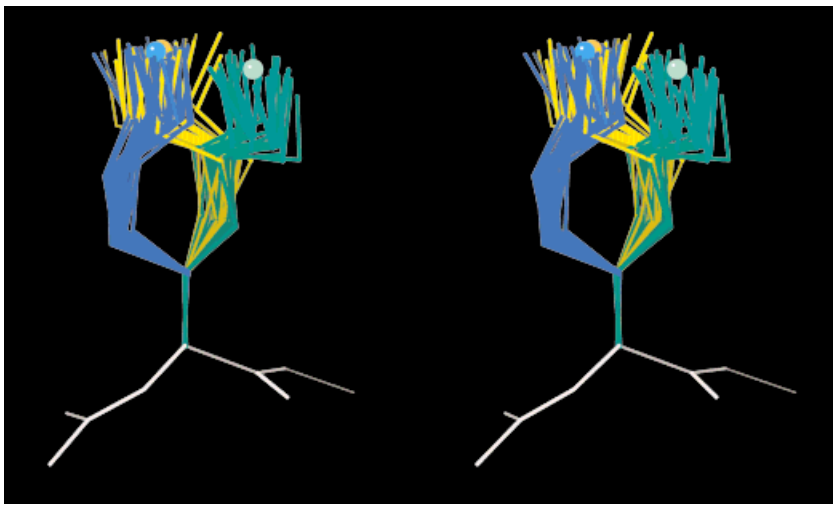

Fig. 7. Stereo pair showing all examples superimposed, for three neighboring Lys rotamers: **mtpt** (blue), **ttmt** (yellow), and **ttpt** (green). Balls indicate the mean Nζ position for each rotamer. **mtpt** and **ttmt** diverge for Cγ and Cδ, but their distributions for Cε and Nζ rejoin and coincide, resulting in almost identical mean Nζ positions. **ttpt** is one of the closest possible nearby rotamers, but its terminal distribution is completely distinct. Individual side-chain examples were superimposed onto ideal-geometry N, C, Cα, Cβ atoms using ProFit and displayed in Mage.

and 40% for **mp***. A mixture of conformations could also produce inverted B-factors, but not just for the suspect rotamers. Therefore, the observed patterns are most consistent with incorrect Cγ positions for **tt*** and **mp***.

The strongest piece of evidence for rotamer correctness is a positive correlation with map quality (i.e., either resolution or B-factor), whereas a misfit conformation should correlate negatively. Figure 5a shows the variation of Leu rotamer occurrence with B-factor. Those rotamers we define as genuine become more common as B-factor decreases, whereas the flipped conformations **tt*** and **mp*** become less common. There are two explanations for this: either misfitting results in a higher B-factor, or flexibility in the side-chain results in poor electron density which is easy to fit incorrectly. Either explanation suggests an error. Since our data covers a limited range of resolution, we selected additional structures from 1.8 to 2.5 Å resolution (see Methods); plotting Leu rotamer frequencies vs resolution shows the same pattern as Figure 5a but with somewhat lower slopes.

For all of the above reasons, we conclude that the **mp*** and **tt*** conformations are very unlikely to be correct. We simply omit them from our data rather than transforming them to the two major peaks, because these misfittings usually cause movement of backbone and Cβ, so that their transformed coordinates would be unreliable. After the backward leucines are omitted, there remains a valid rotamer cluster in each of the **tt** and **mp** areas (Figure 4) which is clash-free and shows the correct B-factor dependence (Figure 5a). Because **tt*** and **mp*** are more numerous at lower resolution and higher B, every previous rotamer defined for Leu **tt** or **mp** has either been between the two clusters or in the incorrect one. Not every individual Leu in **tt*** or **mp*** is necessarily a mistake, since occasionally the environment might force the side chain into that particular strained conformation. Those conformations, however, are not rotameric.

In an analogous manner to Leu, Met has several sets of conformations that are, in part, spatially isosteric for the sulfur and sometimes the Cε.[51] If χ_2 is **p**, then rotating χ_1 by +60° and χ_2 by −120° will return the sulfur atom to its original position. If χ_2 is **m**, then rotation of χ_1 by −60° and χ_2 by +120° will have the same effect. Alternatively, if χ_3 is **m**, rotation of χ_2 by −40° and χ_3 by +120° (+40° and −120° if χ_3 is **p**) puts the sulfur and Cε in partially isosteric positions. Because the sulfur has no hydrogens, these transformations do not result in steric clashes. They do, however, involve changing χ angles by about 60°, resulting in near-eclipsed dihedrals. If the electron density is at all ambiguous, two equally good conformations may seem possible, but in reality the side chain should never be fit in an eclipsed conformation unless other possibilities have been ruled out. In our experience a combination of using lower map contours, examining the all-atom long-range van der Waals interactions, and trying the valid rotamers can almost always suggest a Met conformation which is both in the density and rotameric.

In structure determinations, appropriate criteria should be met before fitting a side chain as nonrotameric. There should be good evidence that the nonrotamer is a better fit to the data than any rotamer, there should be a structural reason for adoption of that conformation, and any steric clash should be avoidable with only modest bond angle distortion.

Proline and Disulfides – Special Cases

Proline ring-pucker states can be treated as equivalent to rotamers, since they alter the backbone conformation only very slightly. Most rotamer libraries, if they include Pro, treat it as having three conformations: Cγ-endo (or "up"), Cγ-exo (or "down"), and planar.[6,9,16,17] Some force fields and refinement methods allow puckers also at other ring atoms (especially Cβ), and such cases occur in our database. However, it has been argued convincingly that Pro has only two preferred puckers, rather than three or more;[38] the planar and Cβ pucker states are absent at high resolution in small-molecule structures. We also found in a previous study that long-range clashes are substantially decreased by substituting either the Cγ-endo or Cγ-exo states for planar or Cβ puckers.[22] We, therefore, treat Pro as having only two acceptable puckers, which occur in the present data in equal numbers, clustered at values consistent with those found previously.[38] Electron density that appears planar is often observed for Pro rings in protein structure determination; this is probably caused by averaging between the Cγ-endo and -exo pucker states and is better modeled as two alternate conformations, as often seen directly at higher resolution. Prolines preceded by cis peptides are always observed to have the Cγ-endo pucker.

Disulfides can also be surprisingly difficult to fit correctly, since Fourier ripples from the sulfur atoms can result in weak or shifted density for one or both Cβs. Additionally, incorrectly fit disulfides are hard to fix because of multiple constraints. A strict resolution limit and avoidance of high-B or alternate-conformation examples are, thus, very important for analyzing disulfides, but they are rare enough that our present database is too small to deal with all five χ angles. A complete five-angle library will be presented in a subsequent paper, using a database chosen to be suitable for that purpose.

Clashing Rotamers and Bond Angle Distortions

Three types of clustered, well-populated, correctly B-dependent rotamers in our library are found to have moderate but significant atomic overlaps when built in standard geometry. These are **m**-30° of Phe or Tyr, **p**30° of Asn or Asp, and those with χ_4 ±105° for Arg. In each case, the bond angles of observed examples are opened out slightly to ease those clashes, and there are also favorable H-bond or packing interactions that can help to compensate for the strained conformation.

For Phe and Tyr, the χ_1/χ_2 distribution is populated throughout χ_2 when χ_1 is **m**, as shown by Schrauber et al.[7] and in our data. With this χ_1 the aromatic ring lies between the two smallest backbone atoms (Hα and N), but in ideal geometry, for a large range of χ_2 (from −40° to +50°), there is steric overlap between the edge of the ring and the backbone (Hδ to N). This overlap is 0.3 Å at the

modal position ($\chi_1 = -64°$ $\chi_2 = -19°$) for Phe, which is below our clash cutoff of 0.4 Å but not negligible. There are good reasons, however, to believe that this conformation is correct: an aromatic ring is difficult to fit incorrectly at 1.7 Å resolution or higher, and aromatics tend to be found in the interior of the protein where the electron density is best. Most of the 273 residues in this rotamer show local bond-angle distortions: the mean Cα-Cβ-Cγ angle for **m**-30° in Tyr and Phe is 115.6±2.0°, compared with 113.6±2.4 for the overall distribution and with Engh and Huber standards of 113.9±1.8° for Tyr and 113.8±1.0° for Phe. With this 2° bond-angle enlargement (and often with an increase in the N-Cα-Cβ angle as well), these residues do not, in fact, routinely show ring-to-backbone van der Waals overlaps. Such side chains are usually well packed, which presumably both prevents other rotamers and provides favorable interactions to compensate for the modest bond-angle strain.

The overlap for the Asp or Asn **p**30° rotamer is 0.36 Å and is present for any standard geometry conformation with χ_1 **p**. Bond angle increases are seen but are within the standard deviation of the distribution (Cα-Cβ-Cγ angle for all Asn is 112.5 ± 2.1°; for **p**30° rotamer 113.6 ± 2.1°). Almost all of the 132 **p**30° examples are H-bonded to i+2 or i+3 NHs in a pseudoturn or a helix N-cap arrangement, which could offset the energy penalty of a small bond-angle distortion and/or a small remaining overlap.

The van der Waals overlap of the arginine **mtm**105° (or **ttm**105°, **mtp**105°, **ttp**105°) rotamer in ideal geometry is slightly larger (0.46 Å Hη to Hγ), and we see 41 total examples. The size of this clash may indicate that the radius we use for hydrogens on charged groups (1.0 Å) is still slightly too large. However, the χ_4 value of ±105° is in the local optimum given an oppositely signed χ_3, while the offset from the usual Arg χ_4 value of ±85° confirms that these rotamers are, indeed, disfavored. Guanidinium H-bonds provide both conformational restraints and compensating favorable interactions.

Positive-Feedback Cycles for Bad Rotamers

The real damage from including poor rotamers in a library is that they can become self-fulfilling prophecies. The cycle arises because almost any conformation will occasionally be the best fit to some poorly connected piece of electron density, so the bad rotamer will begin to show up in new structures. If later rotamer libraries include low-resolution or high B-factor residues, then that same bad conformation will seem confirmed as a valid rotamer.

There is, indeed, evidence of this taking place. We have previously discussed this effect for Asn and Gln rotamers with incorrectly flipped amides and seriously clashing NH_2 groups,[25] such as Gln **tpt**[6] or Asn **p**180°.[9] In the current data, there is an especially clear example for Met in the **tpm** conformation. This rotamer appears in the library used in the crystallographic fitting program O, which is based on the library of Ponder and Richards but has been extended to fill out angles which were undefined in that study. Met **tpm**, as shown in Figure 6a, is clearly impossible, having a 0.69 Å atomic overlap between the Hα and Hε s, even with methyl rotation optimized. There are no examples of this conformation in our database, but it occurs three times in the control set of 78 structures at 1.8–2.5 Å resolution and also for the altered side chains in some high-resolution mutant structures. It seems likely that the inclusion of this side-chain conformation in the library of such a popular refitting program has led to its appearance in structures where the density may be ambiguous. The three structures which exhibit this rotamer at B <40 were solved at 1.8, 2.0, and 2.2 Å, suggesting that the increase in bias of structures towards rotamer libraries happens even at relatively high resolutions, within the range usually used for compiling other libraries.

Other examples of previously defined rotamers with prohibitive clashes and unsupported by our data have resulted from eclipsed χ angles such as Val 120°[7] (Figure 6b), from a peak at the average between two clusters such as Leu **tt**,[11] or perhaps most insidiously from data with systematic fitting errors such as the Leu **tt*** or **mp*** cases discussed above. Other clashing rotamers, such as Arg **tmtp**[10] (Figure 6c) or Lys **mmpt**[9] (Figure 6d), may be included out of a desire for complete sampling of conformational space or from poor behavior on energy minimization.

In the present study, we have included data only from structures of 1.7 Å resolution or better and side chains only with B <40, to maximize the level of direct evidence for each individual conformation. Each defined rotamer was then required to pass both criteria of good occurrence and of clustering in the distribution from the high-quality data and also of constituting a convincing local optimum for all-atom van der Waals analysis in ideal geometry; borderline cases were decided by analyzing their behavior as a function of B-factor. We believe, therefore, that it is unlikely that the present library contains any artificial rotamers, thus breaking the feedback cycle.

DISCUSSION

Overall, these results show even more strongly than before that protein side-chain conformations do indeed occur as well-defined rotamers. A library of rotamers is the preferred form of analysis if two conditions for the behavior of side-chains are met:

1) conformations occur as relatively tight clusters in multidimensional χ space, and
2) the permissible cluster locations and probabilities cannot simply be determined by multiplying together the individual angle distributions.

In testing the validity of the second criterion, we find that only Lys follows rules strikingly simpler than rotamer enumeration; all 81 Lys rotamer frequencies can be modeled to very high accuracy using only six physically realistic parameters (Table IV). Even for Lys, the individual frequencies are strongly dependent on the neighboring χ angles (e.g., χ_3 on χ_2 and χ_4), and the dependencies are even more complex for other amino acids. In addition, there are minor rotamer combinations with atomic clashes at the staggered angles which have their peak occurrences

at significantly shifted angles (e.g., Arg **mtm**105° or Leu **tt**); this effect adds real but misleading shoulders to the peaks in one-dimensional χ distributions, further confirming the need for multidimensional analysis.

The truth of the first condition (tight clusters) was challenged by Schrauber et al.[7] They showed that despite more and better data than in the original treatments, many residues were >20° from a rotamer mean, while summing χ angle ranges even as wide as ±20° for long side chains would locate the functional group very imprecisely. However, with further increase in database size and accuracy and with the application of stronger quality criteria, we have shown here and in previous work[25] that almost all of the clusters tighten very satisfactorily. Gln χ_3, when χ_2 is trans, provides the only really refractory case, while many rotamers show half widths less than ±10° (see Table I). Because the multidimensional clusters in χ space are round rather than rectangular, the combined effect for long side chains is the root sum of squares, rather than the direct sum, of the individual angle spreads.

To illustrate the overall level of rotamer clustering in Cartesian space for real side chains, Figure 7 shows the superposition of all examples in our database of three neighboring Lys rotamers: **mtpt**, **ttmt**, and **ttpt**. Even at the terminal atom the clusters are tight, despite the distribution at each angle having a significant spread. The distributions of Nζ positions for **mtpt** (blue) and **ttmt** (yellow) are completely overlapping with means only 0.27 Å apart, whereas the Nζ distribution for the near-neighbor rotamer **ttpt** (green) is well-separated from the others in its own distinct location 2.1 Å away. The standard deviation of Lys Nζ atom positions in a given rotamer is about 0.8 Å, which certainly seems narrow enough to confirm the practical utility of rotamers; even with four χ angles, the rotamer clusters are crisply distinct.

Comparison With Other Libraries

For the simpler amino acids and the most common rotamers, all libraries, of course, agree quite well, at least in existence and position if not always in probability. For the rarer rotamers and the more difficult residue types (including Lys, Arg, Met, Leu, Gln, Glu, Asn, Asp, and Pro), there are at least three factors governing disagreements between this and previous work. Growth in the database is crucial to such efforts, but here it is not the most decisive issue; our raw data are essentially indistinguishable from those of Dunbrack and Cohen.[11]

The second factor is the development of our new methods for optimizing explicit H positions[23] and representing all-atom contacts clearly and dramatically.[22] If graphics such as Figure 5b and Figure 6 had been available to earlier authors, their rotamer lists would almost certainly have been affected. The all-atom contact analysis, in both visual and quantitative forms, was essential to discarding from the present library a significant number of previous rotamers now shown to represent flipped amides or systematic fitting errors. On the other hand, this process helped in validation of a relatively large set of well-behaved rotamers down to the level of 1–2% occurrence probability.

A third, more complex, factor covers differences in choice of definitions and methodologies. Some disagreements arise from blurring the distinction between a true rotamer (i.e., a locally favored conformation with clustered examples) and an arbitrary sample point in conformation space. Many computational uses of rotamers require additional sampling within the allowed regions, but such sample points are not real rotamers because their spacing and position depends on their intended use, not on the properties of the side-chain conformations. Therefore, we have provided a minimal set of sample points separately (Table III), rather than including them in the rotamer library. An additional problem is that extra sample points are helpful only if they correspond to populated regions of the distributions and are physically reasonable conformations, which has not always been the case.

Most earlier work used the mean (average) value as the rotamer position, whereas we use the mode (peak occurrence). Determining the mode requires smoothing the distribution, but modes have important advantages of corresponding to the local energy minima and of being sensitive to closely spaced peaks while independent of skewed peak shape or of arbitrarily defined bins. As was done by De Maeyer et al.,[10] we also list common-atom rotamer positions with common χ angles for cases that have similar data and equivalent subsets of geometry and contacts. This streamlines some applications, and it avoids the danger of choosing between rotamers based on a difference that is not statistically significant.

Differing treatment as well as size of the database used is an important methodological issue. The 240-protein database used here is much larger than early ones[6,7] but is either similar in size to or smaller than those used in recent studies.[9,11] It is, however, restricted to higher-resolution structures (1.7 Å here vs 2.0 Å[6,7,11] or 2.5 Å[9]) and to structures satisfying a number of other quality criteria (see Methods). Most importantly, the number of side chains analyzed is further reduced by eliminating those with uncertain conformations. In general, when a side chain has been shown to be either wrong or uncertain we simply omit it from the compiled data, because any correction process not using the experimental data would be highly suspect. The only exceptions are the 180° flips of side-chain amides or imidazoles which we do correct in unambiguous cases, and the orientation of movable hydrogen positions, neither of which affects agreement with the X-ray data significantly. A larger database is clearly desirable when trying to distinguish signal (correct rotamers) from random statistical noise, because the signal-to-noise ratio increases as the square root of the number of observations. However, that relationship holds only if the data is of uniform quality and if the errors are random, neither of which is the case for side-chain conformations. In fact, since low-resolution, high B-factor data is most susceptible to systematic errors, adding such observations will degrade rather than improve the results. In effect, we filter out the noise rather than attempting to amplify the signal.

We feel the value of our approach has been confirmed by the production of clean, well-clustered distributions and the settling of some previously-unanswered questions. In particular, we recommend that any study of conformational details should either omit examples with high B-factors, because of the combination of easy application and impressive effectiveness, or else should specifically examine behavior as a function of B as was done here to test for possible artifacts.

Nonrotameric Side Chains

A side-chain rotamer is normally taken to mean a combination of χ angles producing a locally low-energy conformation, found empirically as a cluster of observations in torsion space. By defining rotamer boundaries in torsion space, it is possible to study how often long-range interactions shift side chains away from the preferred rotameric conformations. In this study we draw boundaries at the common-atom values $\pm 30°$ (exceptions are listed in Table I, primarily for angles with shallower energy wells) and count as nonrotameric any residue which falls outside these bounds, including both those with near-eclipsed χ angles and those with staggered angles but highly unfavorable angle combinations.

For a dihedral angle between two tetrahedral carbons, a fully eclipsed conformation has an energy of 4–10 kcal/mol higher than that of a staggered conformation.[53] This is equivalent to two to four hydrogen bonds, and we have, indeed, observed a few low B-factor Gln residues with eclipsed χ_1 angles and three or four hydrogen bonds. This does not mean that conformations with eclipsed χ angles should be defined as rotamers. It does mean, however, that occasionally it is appropriate to use nonrotameric conformations in either experimentally determined or theoretical protein models if there is good reason. "Good reason" may mean clear density in a non–phase-biased map, tightly constrained local packing, or the ability to make several hydrogen bonds to offset the energy lost in forcing a nonrotameric conformation. Whenever a nonrotamer is used, it should be because no rotameric conformation fits the available data nearly as well.

Clashing Rotamers

Nonrotameric conformations, and a few rarely populated genuine rotamers, may have internal van der Waals overlaps when built in standard geometry. These overlaps should be small in size, and it should be possible to largely offset them with small local geometry distortions. In every case where such conformations are significantly populated in our data, we have closely examined not only the distributions but also the structures to make sure they are reasonable. For several examples in each case, we have also examined electron density maps. The three types of slightly overlapping rotamers in the present library all have many low-B examples, with clear electron density; their overlaps can be relieved by modest bond angle changes, and they typically show favorable compensating interactions. These cases, we conclude, are indeed genuine examples of somewhat strained rotamers.

In contrast, because of the steepness of the Lennard-Jones potential, more serious van der Waals overlaps involve a prohibitively large energy penalty. Whereas a protein may be able to offset an eclipsed χ angle, it is probably never able to offset the many tens of kcals/mol needed to stabilize a van der Waals overlap of about 0.6–1.0 Å as some published rotamers display (Figure 6). Such configurations are much more likely to be errors than correct-but-rarely-populated conformations. As discussed in the Results section, these cases can be understood as caused by defining a rotamer at the average between two clusters, by choosing the wrong flip state of a group which appears symmetric without explicit hydrogens, or by the inclusion of systematically misfit conformations. We conclude that none of those cases should properly be called side-chain rotamers.

CONCLUSIONS

The present rotamer library has been constructed using more of the available information than previous studies, including various measures of the reliability of individual side-chain conformations and tests of the conformational validity of potential rotamers. We took advantage of two new criteria (all-atom contact analysis and B-factor dependence), which are independent of each other and of earlier work, in order to settle the borderline cases. All of the rotamers listed here correspond to local energy minima and peaks in the observed χ distributions. Once poorly determined side chains are discounted and flips corrected, an extremely high proportion (>90% for most residues) are in good rotameric conformations as defined by this library.

For the low-B regions of high-resolution protein structures, individual side chains in conformations far from a rotameric position fall into three classes: a) a few types with looser constraints than most (e.g., Gln or Glu with χ_2 **t**); b) those which we suggest are fitting errors, such as flipped Asn or Leu; and c) interesting cases (relatively common near active sites but especially unlikely for disordered surface residues) for which the higher energy is offset by other positive interactions. These observations suggest that proteins exhibit significantly strained side-chain conformations surprisingly rarely and only for good reasons.

The result of this work is, we believe, a clear improvement on all previous libraries and that it neither omits any important rotamers nor includes any which are significantly in error. It is, however, called penultimate, because applying suitably strict filters to the currently available structures yields too few residues to determine accurately the reliability and position of the rare minor rotamers. Therefore, in a few years' time after many more atomic-resolution structures have been solved, it should be possible to produce a definitive rotamer library that can stand permanently to support accurate modeling of both experimental and predicted protein structures.

ACKNOWLEDGMENTS

We thank Hope Taylor and Brent Presley for constructing ideal-geometry side chains, Brent Presley for making χ

angle plots, and Lizbeth Videau for critical reading of the manuscript. This work was supported in part by academic leave for J.M.W. from Glaxo-Wellcome.

REFERENCES

1. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. J Mol Biol 1963;7:95–99.
2. Janin J, Wodak S, Levitt M, Maigret B. Conformation of amino acid side-chains in proteins. J Mol Biol 1978;125:357–386.
3. Bhat TN, Sasisekharan V, Vijayan M. An analysis of side-chain conformation in proteins. Int J Pept Protein Res 1979;13:170–184.
4. James MNG, Sielecki AR. Structure and refinement of penicillopepsin at 1.8 Å resolution. J Mol Biol 1983;163:299–361.
5. McGregor MJ, Islam SA, Sternberg MJE. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. J Mol Biol 1987;198:295–310.
6. 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;193:775–791.
7. Schrauber H, Eisenhaber F, Argos P. Rotamers: to be or not to be? An analysis of amino acid side-chain conformations in globular proteins. J Mol Biol 1993;230:592–612.
8. 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.
9. Tuffery P, Etchebest C, Hazout S. Prediction of protein side-chain conformations: a study of the influence of backbone accuracy on conformation stability in the rotamer space. Protein Eng 1997;10:361–372.
10. 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.
11. Dunbrack RL, Cohen FE. Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci 1997;6:1661–1681.
12. Bower MJ, Cohen FE, Dunbrack RL. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. J Mol Biol 1997;267:1268–1282.
13. Lasters I, De Maeyer M, Desmet J. Enhanced dead-end elimination in the search for the global minimum energy conformation of a collection of protein side-chains. Protein Eng 1995;8:815–822.
14. Desjarlais JR, Handel TM. De novo design of the hydrophobic cores of proteins. Protein Sci 1995;4:2006–2018.
15. Dahiyat BI, Mayo SL. De novo protein design: fully automated sequence selection. Science 1997;278:82–87.
16. Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 1991;47:110–119.
17. McRee DE. Practical protein crystallography. San Diego: Academic Press;1993.
18. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. ProCheck–A program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
19. Hooft RWW, Vriend G, Sander C, Abola EE. Errors in protein structures. Nature 1996;381:272.
20. Bernstein FC, Koetzle TF, Williams GJ, et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
21. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. Nucleic acids Res 2000;28:235–242.
22. Word JM, Lovell SC, LaBean TH, Taylor HC, et al. Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogens. J Mol Biol 1999;285:1711–1733.
23. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol 1999;285:1735–1747.
24. van den Akker F, Hol WGJ. Difference density quality (DDQ): a method to assess the global and local correctness of macromolecular crystal structures. Acta Crystallogr D 1999;55:206–218.
25. 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.
26. Benedetti E, Morelli G, Némethy G, Scheraga HA. Statistical and energetic analysis of side-chain conformations in oligopeptides. Int J Pept Protein Res 1983;22:1–15.
27. Kuszewski J, Gronenborn AM, Clore GM. Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids. J Magn Reson 1997;125:171–177.
28. IUPAC-IUB. Commission on biochemical nomenclature: abbreviations and symbols for the description of the conformation of polypeptide chains. J Mol Biol 1970;52:1–17.
29. Markley JL, Bax A, Anata J, et al. Recommendations for the presentation of NMR structures of proteins and nucleic acids. J Mol Biol 1998;280:933–952.
30. McRee DE. XtalView/Xfit–a versatile program for manipulating atomic coordinates and electron density. J Struct Biol 1999;125:156–165.
31. Word JM. All-atom small-probe contact surface analysis: an information-rich description of molecular goodness-of-fit. Dissertation: Duke University; 2000.
32. Richardson DC, Richardson JS. The kinemage: a tool for scientific illustration. Protein Sci 1992;1:3–9.
33. Richardson DC, Richardson JS. Kinemages–simple macromolecular graphics for interactive teaching and publication. Trends Biochem Sci 1994;19:135–138.
34. Richardson JS, Richardson DC. "MAGE, PROBE, and Kinemages". International Tables for Crystallography vol. 4. Dortrecht: Kluwer Academic Publishers;2000 (in press). Chapter 25.2.8
35. Engh RA, Huber R. Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr A 1991;47:392–400.
36. Carugo O, Argos P. Correlation between side-chain mobility and conformation in protein structures. Protein Eng 1997;10:777–787.
37. MacArthur MW, Thornton JM. Protein side-chain conformation: a systematic variation of χ_1 mean values with resolution–a consequence of multiple rotameric states? Acta Crystallogr D 1999;55:994–1004.
38. Némethy G, Gibson KD, Palmer KA, et al. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. J Phys Chem 1992;96:6472–6484.
39. Richardson JS, Richardson DC. Amino acid preferences for specific locations at the ends of α-helices. Science 1988;240:1648–1652.
40. Wan W-Y, Milner-White EJ. A recurring two-hydrogen-bond motif incorporating a serine or threonine residue is found both at α-helical N termini and in other situations. J Mol Biol 1999;286:1651–1662.
41. Summers NL, Carlson WD, Karplus M. Analysis of side-chain orientations in homologous proteins. J Mol Biol 1987;196:175–198.
42. Richardson JS, Richardson DC, Tweedy NB, et al. Looking at proteins: representations, folding, packing, and design. Biophys J 1992;63:1186–1209.
43. Dunbrack RL, Karplus M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. Nat Struct Biol 1994;1:334–340.
44. Wiberg KB, Murcko MA. Rotational barriers: 2. Energies of alkane rotamers. An examination of gauche interactions. J Am Chem Soc 1988;110:8029–8038.
45. Wolfram Research I. Mathematica Version 3.0. Champaign, IL: Wolfram Research, Inc.;1996.
46. Richardson JS. The anatomy and taxonomy of protein structure. In: Anfinsen CB, Edsall JT, Richards FM, editors. Advances in protein chemistry. New York: Academic Press; 1981. p 167–339.
47. Richardson JS, Richardson DC. Interpretation of electron density maps. Methods Enzymol 1985;115:189–206.
48. Creamer TP, Rose GD. Side-chain entropy opposes α-helix formation but rationalizes experimentally determined helix-forming propensities. Proc Natl Acad Sci 1992;89:5937–5941.
49. Lee C, Subbiah S. Prediction of protein side-chain conformation by packing optimization. J Mol Biol 1991;217:373–388.
50. Dunbrack RL, Karplus M. Backbone-dependent rotamer library for proteins: application to side-chain prediction. J Mol Biol 1993;230:543–574.
51. Petrella RJ, Lazaridis T, Karplus M. Protein sidechain conformer prediction: a test of the energy function. Folding & Design 1998;3:353–377.
52. Kuszewski J, Gronenborn AM, Clore GM. Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. Protein Sci 1996;5:1067–1080.
53. Karplus M, Parr RG. An approach to the internal rotation problem. J Chem Phys 1963;38:1547–1552.