J. Mol. Biol. (1993) **230**, 543–574

Backbone-dependent Rotamer Library for Proteins

Application to Side-chain Prediction

Roland L. Dunbrack Jr and Martin Karplus

Department of Chemistry
Harvard University
Cambridge, MA 02138, U.S.A.

(Received 24 July 1992; accepted 26 October 1992)

A backbone-dependent rotamer library for amino acid side-chains is developed and used for constructing protein side-chain conformations from the main-chain co-ordinates. The rotamer library is obtained from 132 protein chains in the Brookhaven Protein Database. A grid of 20° by 20° blocks for the main-chain angles ϕ,ψ is used in the rotamer library. Significant correlations are found between side-chain dihedral angle probabilities and backbone ϕ,ψ values. These probabilities are used to place the side-chains on the known backbone in test applications for six proteins for which high-resolution crystal structures are available. A minimization scheme is used to reorient side-chains that conflict with the backbone or other side-chains after the initial placement. The initial placement yields 59% of both χ_1 and χ_2 values in the correct position (to within 40°) for thermolysin to 81% for crambin. After refinement the values range from 61% (lysozyme) to 89% (crambin). It is evident from the results that a single protein does not adequately test a prediction scheme.

The computation time required by the method scales linearly with the number of side-chains. An initial prediction from the library takes only a few seconds of computer time, while the iterative refinement takes on the order of hours. The method is automated and can easily be applied to aid experimental side-chain determinations and homology modeling. The high degree of correlation between backbone and side-chain conformations may introduce a simplification in the protein folding process by reducing the available conformational space.

Keywords: proteins; side-chains; rotamers; prediction; conformation

1. Introduction

An understanding of the conformations of side-chains is required for the analysis of protein folding and for the prediction of protein tertiary structure. Prediction methods can also be used in the structure determination of proteins from X-ray crystallography and nuclear magnetic resonance spectroscopy by providing a procedure for the initial placement of side-chains. They form part of any scheme to predict the structure of a protein from data for homologous proteins. Early work based on structural surveys (Janin *et al.*, 1978; Bhat *et al.*, 1979) and energy calculations (Gelin & Karplus, 1975, 1979), indicated that the side-chain dihedral angles in proteins generally corresponded to the potential energy minima of the isolated amino acid. In fact, as crystal structures have improved, a decreasing number of side-chains have been observed to deviate significantly from one of the isolated amino acid minima (Bhat *et al.*, 1979; James & Sielecki, 1983; Ponder & Richards, 1987). While some of the narrowing of the distributions may be caused by rotamer preferences introduced in modern refinement programs such as PROLSQ (Konnert & Hendrickson, 1980), the weighting factors are usually quite weak and are unlikely to dominate the experimental data in high-resolution structures.

Ponder & Richards (1987) determined the distributions of side-chain dihedral angle $\{\chi_1,\chi_2\}$ pairs for the amino acid residues from a set of ten proteins whose X-ray structures had been determined at a resolution of 2 Å or better (1 Å = 0·1 nm). They found that most side-chains are limited to a small number of the many possible $\{\chi_1,\chi_2\}$ minima. For example, while the leucinyl residue has nine possible $\{\chi_1,\chi_2\}$ conformers, two of these (g^+t and tg^-) account for 88% of the leucinyl residues in the survey. With a database of 61 protein structures, McGregor *et al.* (1987) found that certain side-chains exhibit rotamer preferences that depend on the main-chain secondary structure. For example, Trp has 75% of its χ_1 values near 180° in α-helices, while 62% of the χ_1 values are near −60° in β-sheets.

With an extended database (132 polypeptide

0022-2836/93/060543-32 $08.00/0

chains in 126 crystal structures at a resolution of 2·0 Å or better), it is possible to make a more detailed analysis of the relation between the backbone dihedral angles ϕ and ψ of an amino acid and the side-chain dihedral angle distributions. By examining all side-chain dihedral angles for all amino acids, we have found that there is a significant correlation between the backbone ϕ,ψ values and the side-chain dihedral angles, which goes beyond a correlation with secondary structure. Blocks corresponding to a 20° by 20° grid in ϕ and ψ yield meaningful probabilities for the χ values (χ_1,χ_2 . . .) of most of the amino acids. In some cases the database is not sufficient to determine the ϕ,ψ dependent probabilities. We shall show elsewhere that energy calculations for isolated dipeptides generally are in accord with the observed preferences. In this paper, we describe the results obtained for the side-chain dihedral angle distributions of the amino acids and demonstrate that such a "backbone-dependent rotamer library" is very useful in providing starting positions for predicting side-chain conformations of proteins.

A variety of methods have been suggested for determining side-chain conformations. The type of method that is appropriate depends, in part, on the complexity of the problem to be solved. For single-site mutations, a detailed energy function search of the conformational space available to the mutant side-chain (Shih *et al.*, 1985) can be made to determine its position. Also, free energy simulations can be used to introduce mutant side-chains (Tidor & Karplus, 1991). Good overall results are expected, since it has been shown (Gelin & Karplus, 1979) that potential energy functions of the molecular mechanics type are adequate for representing the interactions of buried side-chains. For surface side-chains, it was found that solvent and interactions with neighboring proteins in the crystal must be included. In contrast to their behavior in a crystal environment, surface side-chains in solution are likely not to have a unique orientation. Nuclear magnetic resonance studies of protein structures (Wüthrich, 1989) indicate that such flexibility is often present. The most detailed procedure for studying surface side-chains is to do free-energy mapping of the (χ_1,χ_2 . . .) angle distribution in the presence of an explicit model for the solvent (Straatsma & McCammon, 1992; Kuczera *et al.*, unpublished results). Also, additional energy terms can be introduced in molecular mechanics programs to approximate dielectric effects, the hydrophobic effect, and solvent structure around ionic and polar functional groups (Pettitt & Karplus, 1985; Schiffer *et al.*, 1992; Wesson & Eisenberg, 1992). A method such as CONGEN (Bruccoleri & Karplus, 1987) searches the conformational space to build the backbone and side-chains for limited regions of proteins (e.g. the hypervariable loops of antibodies). Lee & Subbiah (1991) have used a computationally intensive, simulated annealing approach and a van der Waals repulsive potential to predict the side-chain positions in proteins, given the backbone co-ordinates. Holm & Sander (1991) used backbone segments from a structural database to build full backbone co-ordinates from C^α co-ordinates, and then utilized the database of Tuffery *et al.* (1991) and simulated annealing to place side-chains. Several groups have used backbone co-ordinates to determine initial side-chain placements. Kabsch *et al.* (1990) and Wendoloski & Salemme (1992) searched the database for each side-chain to find a local backbone fold (plus and minus 1 or more amino acid residues) similar to the fold of the protein to be modeled. The side-chain was then placed according to the best such fragment or the most commonly found rotamer. Reid & Thornton (1989) built full backbone co-ordinates of flavodoxin from C^α co-ordinates with a method similar to that of Holm & Sander (1991), but they used the secondary-structure dependent rotamer library of McGregor *et al.* (1987) to predict side-chain positions. When clashes were observed, other common rotamer positions were tested and energy minimized. Desmet *et al.* (1992) have suggested that side-chain placement can be simplified based on the idea that side-chain rotamers can be excluded by pairwise searches and used the method for predicting the side-chains conformations from the known backbone structure starting with the Ponder & Richards (1987) rotamers.

The method described here for predicting side-chain conformations is most closely related to that proposed by Summers & Karplus (1989). In that approach, which was developed as part of a homology modeling scheme (Summers & Karplus, 1990), the side-chains are placed in accord with the known χ angles of the residues in a protein homologous to that being modeled. When steric clashes were observed in the initial placement, side-chain conformations were altered by use of a rigid rotation energy search of the conformational space of individual side-chains. A number of rules were formulated to determine which residue of a pair of clashing side-chains should be altered, depending on the amino acid type, its accessibility, whether or not it is identical to a template side-chain, its participation in hydrogen bonds in the template protein, etc. Residues or side-chain atoms for which there was no information in the template protein were added one at a time and placed according to rigid rotation energy search. The method was rather successful (92% for χ_1, 81% for χ_2) in building the side-chains of the C-terminal lobe of rhizopuspepsin on its backbone from the side-chain positions of the homologous C-terminal lobe of penicillopepsin (39% sequence identity).

The procedure used in this paper is designed to predict all of the side-chains from a knowledge of the backbone co-ordinates. Thus, it is concerned with the same problem as that studied by Lee & Subbiah (1991) and by Desmet *et al.* (1992). Because most of the calculations in the present method deal with one side-chain at a time, the time required scales linearly with the size of the system. The method is faster and more accurate than those of

Lee & Subbiah (1991) and Desmet *et al.* (1992). Also, it can be run on most workstations, an advantage over the approach of Lee & Subbiah (1991), which requires a large memory and is not suitable for bigger proteins such as thermolysin (316 residues). Side-chains can be built on known protein backbone co-ordinates, those optimized from a homologous protein template (Sali *et al.*, 1990), or those determined from some predictive scheme (e.g. starting with C^{α} co-ordinates). The essential new element in the method is that the side-chains are placed simultaneously with the aid of the backbone-dependent rotamer library. As we demonstrate, this provides considerably more information than averaged rotamer libraries (e.g. that of Ponder & Richards, 1987) and so yields a much improved starting set of side-chain positions. If the structure of a homologous protein is known, information about the side-chains of the target structure can be incorporated from the template. Once the initial placement has been made, the optimization procedure follows the philosophy of Summers & Karplus (1989), though some of the methodological details are significantly different. One consequence of these differences is that automation of the method is more straightforward. This is important because it is difficult not to be biased if human decision-making is required, particularly in test applications to known structures. Further, since there are many applications of the method, the less human labor involved in performing a prediction the better.

In the next section of this paper, we present the procedure used to calculate the backbone-dependent rotamer library, and then describe the scheme for setting up the initial side-chain positions and refining them to a final prediction. We also present various ways for evaluating the results of the side-chain predictions since no single criterion is adequate. The following section describes the results. Details of the backbone-dependent rotamer library are given. Full side-chain predictions for six proteins from the known backbone are presented. The proteins chosen for study are thermolysin (PDB code 3tln), ribonuclease A (7rsa), bovine pancreatic trypsin inhibitor (5pti), lysozyme (1lz1), crambin (1crn), and the C-terminal domain of rhizopuspepsin (2apr). Several of these proteins have been used to test other prediction methods. In addition, we apply the method to the penicillopepsin to rhizopuspepsin homology modeling problem, so as to be able to compare the present results with the approach of Summers & Karplus (1989). In the final section, we discuss the potential of the method and implications of the results for protein folding.

2. Methods

(a) *The ϕ,ψ rotamer library*

The library was calculated from the structures of 132 protein chains in 126 structures in the Brookhaven Protein Database refined at a resolution better than or equal to 2·0 Å. These proteins are listed in Table 1. Included in these 126 structures are 17 preliminary PDB files available by ftp from Brookhaven (at the Internet address: pdb.pdb.bnl.gov), which have allowed us to extend significantly the database from which the library is calculated. Several groups of homologous proteins are included in the list of structures. While proteins that are identical or nearly identical in sequence have not been included, homologous proteins have been included to increase the size of the database. The structures that are used have been chosen on the basis of several criteria: resolution; date of deposit in the database, in that later structures are likely to be better; and the absence of non-protein ligands that might alter side-chain positions in unpredictable ways. For the prediction of the six proteins described below, the rotamer libraries were determined after removing the protein and its homologues from the list. Thus, in effect, six separate rotamer libraries were calculated. Since the libraries are very similar, only the library calculated with all the proteins listed in Table 1 is described in Results. The backbone ϕ and ψ values were divided into $20° \times 20°$ blocks ($-180°$ to $-160°$, $-160°$ to $-140°$, etc. for ϕ and ψ), and the rotamer library was calculated for each $20° \times 20°$ block. Because of the small block size and steric constraints on the backbone, some regions of the ϕ,ψ map are underpopulated or even empty. Tests with coarser or variable grids confirm the present choice. Rotamer populations for each χ_i ($i = 1,2,3,4$) were calculated using the angular ranges listed in Table 2. For all side-chains (except Ala, Pro and Gly), the χ_1 values correspond to the rotamers of a tetrahedral carbon atom. They were divided into bins of $-120°$ to $0°$ (g^+ conformer), $0°$ to $120°$ (g^- conformer), and $120°$ to $240°$ (t conformer). The same limits were used for the dihedral angle χ_2 of all amino acids that have a χ_2, except for proline, the aromatics, asparagine, and aspartic acid. For proline, χ_1 was placed into 2 bins; $\chi_1 < 0°$ and $\chi_1 > 0°$ corresponding to the 2 proline conformations, C^{γ}-*exo* and C^{γ}-*endo*, respectively. The angle χ_2 of proline was treated analogously. The χ_2 values of phenylalanine, tyrosine and histidine were divided into bins of $0°$ to $60°$, $60°$ to $120°$ and $120°$ to $180°$, even though the expected value is near $\pm 90°$. These values were used to determine whether there were any significant populations more than $30°$ from the usual χ_2 value near $90°$. In well-populated areas of the map, there were no statistically significant deviations from $90°$. If χ_2 was less than $0°$, a χ_2 value of $\chi_2 + 180°$ was used. This is exact for Phe and Tyr, and generally true of His, since most crystal structures do not clearly distinguish whether a given His has a value of χ_2 or $\chi_2 + 180°$. Similarly, for Asp and Asn, χ_2 and $\chi_2 + 180°$ were treated as equivalent, and the limits used were $-90°$ to $-30°$ (g^+ conformer), $-30°$ to $30°$ (t conformer), and $30°$ to $90°$ (g^- conformer). Trp χ_2 was treated as either $0° < \chi_2 < 180°$ or $-180° < \chi_2 < 0°$. For the amino acids with flexible χ_3 and χ_4 dihedral angles (Lys, Arg, Glu, Gln), analogous ranges were used; i.e. the same limits as described for χ_1 were employed, except for χ_3 of Glu and Gln, where the limits described for Asp and Asn χ_2 were used.

(b) *Prediction method*

To make clear the procedure used in generating the side-chain positions, the steps involved are listed in Fig. 1. Explanatory comments on the various steps are given in what follows.

(i) *Construction of initial model*

(i)(a) *Backbone co-ordinates*

One is starting with a model of the backbone, which is either derived from the Cartesian co-ordinates of a target

Table 1

List of Protein Databank files used in backbone-dependent rotamer library

Name	Date	Code-Chain	Resolution (Å)
Protease inh. dom. of Alzheimer's amyloid	SEP90	1AAP-A	1·5
Actinoxanthin	DEC82	1ACX	2·0
Adenylate kinase isoenzyme-3	JAN90	1AK3-A	1·9
Alpha-lactalbumin	AUG89	1ALC	1·7
Aldolase A	MAY91	1ALD	2·0
Bilin binding protein	SEP90	1BBP-A	2·0
Carbonic anhydrase	FEB89	1CA2	2·0
Cytochrome c	MAR83	1CCR	1·5
Superoxide dismutase (co substituted)	FEB92	P1COB-A	2·0
Cholesterol oxidase	FEB91	1COX	1·8
Crambin	APR81	1CRN	1·5
Citrate synthase-L-malate	MAY90	1CSC	1·7
Subtilisin Carlsberg complex eglin-c	JUN88	1CSE-E	1·2
Subtilisin Carlsberg complex eglin-c	JUN88	1CSE-I	1·2
L7/L12 50 S ribosomal protein	SEP86	1CTF	1·7
Defensin	JAN91	P1DFN-A	1·9
Hemoglobin (erythrocruorin, deoxy)	MAR79	1ECD	1·4
FK506 binding protein complex	MAY91	1FKF	1·7
Gamma-II crystallin	AUG85	1GCR	1·6
Holo-D-glyceraldehyde-3-phos. dehydrogenase	JUN87	1GD1-O	1·8
Guanylate kinase	DEC91	P1GKY	2·0
Glycolate oxidase	JUN89	1GOX	2·0
Glutathione peroxidase	JUN85	1GP1-A	2·0
Oxidized high potential iron protein	APR75	1HIP	2·0
Human neutrophil elastase	APR89	1HNE-E	1·84
Alpha-amylase inhibitor HOE-467 A	JAN89	1HOE	2·0
Intestinal fatty acid binding protein	DEC90	1IFB	1·96
Lysozyme (mutant)	MAY91	1L58	1·65
Leghemoglobin (deoxy)	APR82	1LH4	2·0
Lambda repressor-operator complex	NOV91	P1LMB-A	1·8
Myoglobin (deoxy, pH 8·4)	AUG81	1MBD	1·4
Mesentericopeptidase	APR91	P1MEE-A	2·0
Oncomodulin	APR90	1OMD	1·85
Ovalbumin (egg albumin)	NOV90	P1OVA-A	1·9
Pseudoazurin (oxidized CU++ at pH 6·8)	JUN88	1PAZ	1·55
Human plasminogen Kringle 4	JUL91	P1PK4	1·9
Avian pancreatic polypeptide	JAN81	1PPT	1·37
434 repressor (amino-terminal domain)	DEC88	1R69	2·0
Retinol binding protein	APR90	1RBP	2·0
Rubredoxin	MAR88	1RDG	1·4
Bence-Jones immunoglobulin REI variable	MAR76	1REI-A	2·0
Barnase (G specific endonuclease)	MAR91	P1RNB	1·9
Selenomethionyl ribonuclease H	JUL90	1RNH	2·0
ROP: Col E1 repressor of primer	APR91	P1ROP-A	1·7
Ribonuclease SA	DEC90	P1SAR-A	1·8
Trypsin	APR88	1SGT	1·7
Scorpion neurotoxin (variant 3)	DEC82	1SN3	1·8
Staphylococcal nuclease	JUL89	1SNC	1·65
Trypsinogen	SEP79	1TGN	1·65
Hemoglobin (T state, partially oxygenated)	JAN90	1THB-A	1·5
Hemoglobin (T state, partially oxygenated)	JAN90	1THB-B	1·5
Tonin	JUN87	1TON	1·8
Ubiquitin	JAN87	1UBQ	1·8
Uteroglobin (oxidized)	APR89	1UTG	1·34
Iso-2-cytochrome c (reduced state)	OCT91	P1YEA	1·9
B-2036 composite cytochrome c (reduced)	OCT91	P1YEB	1·9
Triose phosphate isomerase	JAN90	1YPI-A	1·9
Cytochrome B562 (oxidized)	JAN90	256B-A	1·4
Actinidin (sulfhydryl proteinase)	NOV79	2ACT	1·7
Alpha-lytic protease	MAR85	2ALP	1·7
Acid proteinase (rhizopuspepsin)	MAR87	2APR	1·8
Azurin (oxidized)	OCT86	2AZA-A	1·8
Cytochrome c (prime)	AUG85	2CCY-A	1·67
Cytochrome c-3	NOV83	2CDV	1·8
Chymotrypsinogen A	JAN87	2CGA-A	1·8
Chymotrypsin inhibitor 2	SEP88	2CI2-I	2·0
Concanavalin A	APR75	2CNA	2·0
Cytochrome P450cam (camphor monooxygenase)	APR87	2CPP	1·63
Cytochrome c peroxidase	AUG85	2CYP	1·7
Endothia aspartic proteinase	NOV90	2ER7-E	1·6

Table 1 *(continued)*

Name	Date	Code-Chain	Resolution (Å)
Immunoglobulin FAB	APR89	2FB4-H	1·9
Immunoglobulin FAB	APR89	2FB4-L	1·9
Flavodoxin	FEB92	2PCR	1·8
D-galactose/D-glucose binding protein	FEB89	2GBP	1·9
Hemerythrin (met)	OCT90	2HMQ-A	1·66
Hemoglobin V (cyano, met)	AUG85	2LHB	2·0
Pea lectin	JUN90	2LTN-A	1·7
Pea lectin	JUN90	2LTN-B	1·7
Myohemerythrin	APR87	2MHR	1·7
Melittin	OCT90	2MLT-A	2·0
Prealbumin (human plasma)	SEP77	2PAB-A	1·8
Proteinase K	NOV87	2PRK	1·5
Lys 25-ribonuclease T1	JUL88	2RNT	1·8
Rous sarcoma virus protease	OCT89	2RSP-A	2·0
Sarcoplasmic calcium binding protein	AUG91	P2SCP-A	2·0
Cu, Zn superoxide dismutase	MAR80	2SOD-B	2·0
Thermitase complex with eglin	OCT90	2TEC-E	1·98
Thermolysin complex	JUN87	2TMN-E	1·6
Thioredoxin	MAR90	2TRX-A	1·68
Thymidylate synthase complex	JUL91	2TSC-A	1·97
Trp repressor (orthorhombic form)	DEC87	2WRP-R	1·65
GCN4 leucine zipper	JUL91	P2ZTA-A	1·8
Acid proteinase (penicillopepsin)	NOV90	3APP	1·8
Cytochrome B5 (oxidized)	JAN90	3B5C	1·5
Bacteriochlorophyll-A protein	JUN87	3BCL	1·9
Beta-lactamase	DEC90	3BLM	2·0
Cytochrome c-2 (reduced)	NOV83	3C2C	1·68
Chloramphenicol acetyltransferase A	JUL90	3CLA	1·75
Erabutoxin B	JAN88	2EBX	1·4
Native elastase	SEP87	3EST	1·65
Basic fibroblast growth factor	JAN92	3FGF	1·6
Glutathione reductase	FEB88	3GRS	1·54
Rat mast cell protease II	SEP84	3RP2-A	1·9
Proteinase A	MAY90	3SGA-E	1·8
Proteinase B from streptomyces griseus	JAN83	3SGB-E	1·8
Proteinase B from streptomyces griseus	JAN83	3SGB-I	1·8
Cytochrome c-551 (reduced)	JUL81	451C	1·6
Prophospholipase A-2	SEP90	4BP2	1·6
Calcium-binding parvalbumin	OCT89	4CPV	1·5
Enolase	NOV90	P4ENL	1·9
Ferredoxin	JUN88	4FD1	1·9
Interleukin-1 beta	MAR90	4I1B	2·0
Bovine calbindin D9K (minor A form)	AUG91	P4ICB	1·6
Pepsin	DEC89	4PEP	1·8
Beta trypsin, diisoprophylphosphoryl	APR88	4PTP	1·34
Carboxypeptidase A-alpha (Cox)	MAY82	5CPA	1·54
HIV-1 protease complex	APR90	5HVP-A	2·0
C-H-RAS P21 protein (amino acids 1-166)	APR90	5P21	1·35
Parvalbumin (alpha lineage)	SEP91	P5PAL	1·5
Trypsin inhibitor (crystal form II)	OCT84	5PTI	1·0
Rubisco (ribulose-1,5-bisphosphate)	MAY90	5RUB-A	1·7
Troponin-C	MAY88	5TNC	2·0
M-4 apo-lactate dehydrogenase	NOV87	6LDH	2·0
D-xylose isomerase	SEP90	6XIA	1·65
Plastocyanin	SEP89	7PCY	1·8
Ribonuclease A (phosphate-free)	JUN88	7RSA	1·26
L-arabinose-binding protein (mutant)	APR91	8ABP	1·49
Dihydrofolate reductase	MAY89	8DFR	1·7
Insulin	OCT91	9INS-A	1·7
Insulin	OCT91	9INS-B	1·7
Papain (Cys-25 oxidized)	MAR86	9PAP	1·65
Wheat germ agglutinin (isolectin 2)	APR90	9WGA-A	1·8

Name is derived from COMPND records in the PDB files; Date is from the HEADER records; Resolution is from the REMARK records. The code in the Protein Databank Code is prefixed by P if the file is a preliminary entry, available by anonymous ftp from the Brookhaven National Labs (pdb.pdb.bnl.gov). The chain used from each file is appended to the code; if there is no chain indicated, then the single chain in the file is used.

Table 2
Limits for rotamer library χ angles

A. *Ser, Thr, Cys, Val, Phe, His, Tyr*

	χ_1 limits
1	0° → 120°
2	120° → 240°
3	−120° → 0°

B. *Lys, Arg, Met, Gln, Glu, Ile, Leu*

	χ_1 limits	χ_2 limits
1	0° → 120°	0° → 120°
2	0° → 120°	120° → 240°
3	0° → 120°	−120° → 0°
4	120° → 240°	0° → 120°
5	120° → 240°	120° → 240°
6	120° → 240°	−120° → 0°
7	−120° → 0°	0° → 120°
8	−120° → 0°	120° → 240°
9	−120° → 0°	−120° → 0°

C. *Trp*

	χ_1 limits	χ_2 limits
1	0° → 120°	0° → 180°
3	0° → 120°	−180° → 0°
4	120° → 240°	0° → 180°
6	120° → 240°	−180° → 0°
7	−120° → 0°	0° → 180°
9	−120° → 0°	−180° → 0°

D. *Asp, Asn*

	χ_1 limits	χ_2 limits
1	0° → 120°	−90° → −30°
2	0° → 120°	−30° → 30°
3	0° → 120°	30° → 90°
4	120° → 240°	−90° → −30°
5	120° → 240°	−30° → 30°
6	120° → 240°	30° → 90°
7	−120° → 0°	−90° → −30°
8	−120° → 0°	−30° → 30°
9	−120° → 0°	30° → 90°

E. *Pro*

	χ_1 limits	χ_2 limits
1	0° → 60°	−60° → 0°
3	−60° → 0°	0° → 60°

χ_1 and χ_2 ranges are given for each defined rotamer for the amino acid side-chains. The numbers in the left-hand column are used in Fig. 2 to illustrate the preferred rotamers in different positions on the ϕ,ψ map. The limits for χ_3 and χ_4 are described in the text.

Backbone coordinates from target or homologous template protein

Sidechain (sc) placement from library and/or template protein

Disulfide minimization

Hydrogen atom minimization -> **Structure 0**

van der Waals clashes (sc's with backbone)

Sidechain minimizations for sc's which clash with backbone

Sidechain placement

Disulfide minimization

Hydrogen atom minimization -> ***Structure* 1**

van der Waals clashes (all atoms, except Val, Ile, Thr sc's)

Sidechain minimizations for sc's (except Val,Ile,Thr) which clash

Sidechain placement

Disulfide minimization

Hydrogen atom minimization -> **Structure 2**

van der Waals clashes (all atoms)

Sidechain minimizations for all sc's which clash with other atoms

Sidechain placement

Disulfide minimization

Hydrogen atom minimization -> **Structure 3**

Repeat until all clashes are resolved -> ***Structure* 4,5,6,...,N**

Figure 1. Outline of the method. Steps in the procedure for placing side-chains (sc) from the library and for resolving van der Waals conflicts between the side-chains and the backbone and other side-chains.

structure (e.g. a preliminary X-ray or nuclear magnetic resonance determined backbone) or from Cartesian co-ordinates from the experimental structure of a template, such as a homologous protein. If the template and target are of different lengths, some portion of the backbone must be added or deleted. There are a variety of methods for doing this, which are based either on database searches and template fitting (Summers & Karplus, 1990) or an energy function-based conformational search (e.g. Bruccoleri & Karplus, 1987) or a combination of the two.

(i)(b) *Side-chain placement*

Information about the initial placement of the side-chains either comes from the rotamer library alone or from the homologous template protein in combination with the rotamer library. The possible choices for starting co-ordinates are listed in Table 3. When the side-chain information comes from the rotamer library, the information is necessarily in the form of internal co-ordinates (bond lengths, bond angles and dihedral angles). In this case, bond lengths and angles from CHARMM minimized structures (Brooks *et al.*, 1983) are used for the side-chain in the tetrapeptide Acetyl-Ala-Xxx-Ala-$NHCH_3$; these have been calculated for all amino acids and are now used in the CHARMM program residue topology file. Since we are using the all-hydrogen atom parameter set (MacKerell *et al.*, unpublished results), both heavy atom and hydrogen atom bond lengths and angles were determined by the tetrapeptide minimizations just described. In previous work (Summers & Karplus, 1989), the polar hydrogen set was used, and bond length and angle information from CHARMM parameters without minimization were employed. The minimized structures provide a more accurate reflection of likely side-chain structures. Alternatively, one could use averaged bond lengths and angles from a database.

The initial side-chain dihedral angles for a given amino acid are determined from the backbone-dependent rotamer library by the following procedure. The most likely value of χ_1 for the 20° by 20° block corresponding to the backbone ϕ and ψ values for that residue is used; for that value of χ_1, the most common value of χ_2 is used. This is usually the same as picking the most common $\{\chi_1,\chi_2\}$ conformation for the side-chain, corresponding to a given ϕ,ψ, from columns 10 to 18 of Table 4 (see the legend to Table 4 for an explanation of the columns), but in some cases it is different. For example, consider the case in which g^- and g^+ have populations of 40% and 60%, respectively, for χ_1 (columns 7 and 9), but χ_2 is divided evenly between 2 conformations for $\chi_1 = g^+$, (say, g^- and

Table 3
Input data

Name of method	Backbone coor.	Side-chain dihedrals	Side-chain Bond lgths and ang.
targ/lib	Target	Library	Minimized tetramer
temp/lib	Template	Library	Minimized tetramer
targ/temp	Target	Template	Minimized tetramer
		+library	Minimized tetramer
temp/temp	Template	Template:	
		Identical sc	Template Cartesian co-ordinates
		Non-identical sc	Minimized tetramer
		+library	Minimized tetramer

Backbone co-ordinate information can come either from a homologous template protein or from the target protein whose side-chain conformations are to be predicted. Side-chain (sc) dihedral information comes either from the template or from the library either in the form of Cartesian co-ordinates or internal co-ordinates. Bond lengths (lgths) and angles (ang.) come either from the tetramer Ace-Ala-Xxx-Ala-$NHCH_3$, minimized for each possible side-chain (in the form of internal co-ordinates), or from the Cartesian co-ordinates of the template source protein.

t (columns 16 and 17) and there is only 1 conformation for $\chi_1 = g^-$ (say, t, (column 11))). The probabilities for the 3 conformations are 30% (g^+,g^-: column 16), 30% (g^+,t: column 17), and 40% (g^-,t: column 11). If one uses the most common conformation (g^-,t) for χ_1 and χ_2, one chooses the less common value of χ_1. It is better to use one of the conformations of χ_2 corresponding to $\chi_1 = g^+$, the more common rotamer, since if χ_1 is wrong then the value of χ_2 is not really meaningful.

If the number of side-chains in a particular block of the ϕ,ψ map is smaller than 4, the most common χ angle values for the side-chain obtained from a backbone-independent rotamer library is chosen. (The statistics for rotamer preferences independent of the backbone are listed in Table 5. These are discussed in Results.) For all side-chains, except Ser, Thr, Val and Pro, this sets χ_1 equal to $-60°$. For Ser and Thr, the most common χ_1 value is $+60°$, and for Val it is 180°. For proline, the C^γ-*endo* structure for the ring is chosen, with $\chi_1 = +28°$ since this is the average value for χ_1 in the C^γ-*endo* conformation. The most common χ_2 values are 180° (as they are for χ_3 and χ_4) except for aromatic χ_2 terms, which are 90° for Tyr, Phe, His and Trp. For Asn, the preferred χ_1,χ_2 conformation is $-60°,-60°$. These preferred conformations match the preferences calculated from a much smaller database by Ponder & Richards (1987). The only exception is for Met, where Ponder & Richards (1987) list the $-60°,-60°$ conformation as preferred from a sample of 16 residues. The present library contains 399 methionine residues, and the $-60°,180°$ conformation is preferred; the probabilities are 34% for $-60°,180°$ *versus* 22% for $-60°,-60°$.

If the structure of a homologous protein is known, it can be used to determine some of the information about the side-chain positions in the target protein. The form of this information depends on whether the target or template backbone is used. In method temp/temp (Table 3), where both the template backbone and side-chains are used in the initial structure, the Cartesian co-ordinates for side-chains that are identical in the template and the target can be used. For non-identical side-chains for which there is information in the template, the dihedral angles are transferred from the template according to the rules of Summers & Karplus (1989), while the bond lengths and angles come from the tetrapeptide minimizations. For most side-chain types, the dihedral angles are transferred directly, unless the transfer is to or from an aromatic residue or from Val to Thr or Ile. In the latter case, χ_1 is set to $\chi_1 - 120°$ of the template, because of the IUPAC definition of χ_1 of Val relative to Thr and Ile (Kendrew *et al.*, 1970). If the target side-chain is aromatic and the template side-chain is not, or *vice versa*, then the target side-chain is placed according to the library. Where there is no information in the template (e.g. Gly, Ala or Pro) or insufficient information (e.g. Ser → Arg) the additional dihedral angles are chosen from the backbone-dependent rotamer library. If the target backbone is used (method targ/temp in Table 3), however, as by Summers & Karplus (1989), then the template side-chain information must be in the form of internal co-ordinates, even for identical side-chains. For all side-chains, bond lengths and angles are obtained from the tetrapeptide minimizations. For identical side-chains, dihedral angles from the templates are used directly; for non-identical side-chains, dihedral angles are transferred as described above. For target side-chains without sufficient information in the template, the library is used.

Finally, the CHARMM residue topology file is used to set up the remaining co-ordinates that are undefined. This includes the Ala side-chains, the backbone hydrogen atoms, and Gly H^α. If there are known (or suspected) disulfide bonds, then these are set up within CHARMM, and the H^γ atoms are deleted. The cysteine S^γ atoms have already been placed according to the library or the template protein structure, and the bond between them is established in this step. They are adjusted further by minimization (see below).

At this point, a full set of Cartesian co-ordinates, including hydrogen atoms, can be generated from the information obtained as described above and summarized in Table 3.

(i)(c) *Disulfide bond minimization*

Cysteinyl residues involved in disulfide bonds are minimized for 100 ABNR steps (Brooks *et al.*, 1983) with the rest of the protein atoms held fixed. This yields the correct S–S bond distance and eliminates bad contacts with other protein atoms.

(i)(d) *Hydrogen atom minimization*

The positions of the hydrogen atoms in the model structure are minimized for 100 steps with the CHARMM program while all the heavy atoms in the protein are fixed. The resulting structure is the initial model (Fig. 1, Structure 0).

(ii) *Refinement of model*

Given the initial model, a series of steps is taken to refine the side-chain conformations. The main-chain coordinates are kept fixed throughout. A CHARMM calculation (Brooks *et al.*, 1983) is done to determine all side-chain atoms that have positive van der Waals interactions with any backbone atom or other side-chain atoms. These side-chains are reoriented by an iterative procedure, which first treats clashes with the backbone and subsequently those with other side-chains.

(ii)(a) *Side-chain minimizations (side-chain/backbone clashes)*

Any side-chain that clashes with the backbone and where the energy is above a certain threshold (see below) is examined to find if there are alternative conformations that do not clash with the backbone. Since side-chains that overlap the backbone are most likely to be in the wrong conformation, these side-chains are tested for alternative minima before side-chain–side-chain clashes are resolved. The search for alternative conformations is made by setting $\chi_1,\chi_2,\ldots$ equal to all possible combinations of values at the center of the intervals used for the rotamer library; e.g. for all side-chains except proline, χ_1 is set equal to 60°, 180°, −60° in turn in all possible combinations (3 conformations for side-chains with χ_1 only, 9 conformations for side-chains with χ_1 and χ_2 only, etc.). Aromatic χ_2 terms are set to 0°, 45°, 90°, 135°, etc. to cover the full conformation space. Minimizations are then performed for the given side-chain with all other protein atoms held fixed. Each clashing side-chain is minimized for 100 conjugate gradient steps against the same model (Fig. 1, Structure 0). Minimizations are performed for side-chains when an atom of that side-chain has a van der Waals interaction with an atom of the backbone exceeding the limits (Summers & Karplus, 1989):

Side-chain atom type	Side-chain or backbone atom type	Energy
C, N, O or S	With C or S	>5 kcal/mol
O or N	With O or N	>9 kcal/mol
C, N, O or S	With H	>10 kcal/mol
H	With H	>20 kcal/mol

The O, N/O, N limits are higher than heavy-atom interactions with carbon or sulfur, since these atoms can form hydrogen bond donor/acceptor pairs where the van der Waals repulsions between the heavy atoms can reach nearly 9 kcal/mol (1 cal = 4·184 J), because of the favorable electrostatic contributions in the full potential. The hydrogen atom limits are taken higher still because they can be expected to exhibit greater conformational flexibility.

After all minimizations have been performed for side-chains where there exist clashes with the backbone, the side-chains are simultaneously moved to the lowest energy conformation found for each one. The disulfide bonds and hydrogen atoms are then minimized with the rest of the protein atoms held fixed (see subsections (i)(c) and (i)(d), above). The resulting structure is a new model (Fig. 1, Structure 1).

(ii)(b) *Side-chain minimizations (side-chain–side-chain clashes except Ile, Thr, Val)*

Step (ii)(a) is repeated, except this time clashes between all atoms are included, including those between side-chains. Any residue that involves clashes according to the energetic cutoffs listed in step (ii)(a) is minimized according to the scheme just described, with the exception of Ile, Thr and Val. These are predicted with a high degree of accuracy from the library and it is best not to move them at this stage, since it is likely that the other side-chain involved in the clash is in an incorrect position. The resulting structure is a new model (Fig. 1, Structure 2).

(ii)(c) *Repeated side-chain minimizations (all clashes)*

Step (ii)(b) is repeated as many times as necessary to remove all clashes. If atoms in Ile, Thr or Val clash with any other atoms in the protein, they are moved at this stage according to the usual minimization scheme. The structures resulting from these rounds of reorientation and minimization are referred to as Structure 3, 4, etc. in Fig. 1. If the refinement steps do not remove all the clashes, a simultaneous minimization of the residues involved could be performed. This problem did not arise for any of the proteins studied and the converged model obtained here (Structure N where $N \leq 4$ for the 6 proteins) is the final structure.

(c) *Assessing the results*

There are a number of criteria that can be used to determine the "correctness" of the side-chain orientations in model-building schemes. They involve Cartesian root-mean-square deviations (r.m.s.d.†) of atoms and dihedral angle deviations. As in the work of Summers & Karplus (1989) and Wendoloski & Salemme (1992), we employ a dihedral angle criterion and consider a deviation of less than or equal to ±40° correct, based on the supposition that the predicted and experimental values correspond to the same minimum. r.m.s.d. values by themselves are unsatisfactory because they can lead to misleading results. Small side-chains can have dihedral angles far from the experimental values and still have low r.m.s.d. values. Large side-chains can also have quite small r.m.s.d. values and yet be in a different conformation from the crystal structure. It might be argued that such a structure is "correct", since the side-chain fills essentially the same volume. In low-resolution structures, this could be true, since experimental errors in dihedral angles can be large (e.g. for Val). If, however, the dihedral angles are accurately known from high-resolution structures, it is important to test whether a predictive method is able to determine the dihedral angles. Since we are using high-resolution structures to test the prediction scheme, we emphasize dihedral angle differences, though we also consider r.m.s.d. values, particularly to compare with the results of others.

When citing dihedral angle statistics, there are 2 ways of counting whether a certain χ_2 (or χ_3 or χ_4) is correct, depending on whether the deviation in χ_1 (or χ_2 or χ_3) from the experimental structure is considered. Lee & Subbiah (1991) report χ_2 angle statistics that do not depend on the accuracy of χ_1. Wendoloski & Salemme (1992), by contrast, report χ_{1+2} statistics; i.e. the percentage of residues that have both χ_1 and χ_2 correct (to within 40°). This information is useful, since if χ_1 is far wrong,

† Abbreviations used: r.m.s.d., root-mean-square deviation(s); BPTI, bovine pancreatic trypsin inhibitor.

the Cartesian positions of χ_2 atoms are likely to deviate significantly from their positions in the experimental structure, even if χ_2 is "correct". We report statistics for χ_1 for all side-chains (except Ala and Gly), χ_2 for all side-chains (except Ala, Gly, Ser, Thr, Val and protonated Cys) regardless of whether χ_1 is correct or not, and χ_{1+2} for all side-chains (except Ala, Gly, Ser, Thr and Val, but including cysteinyl residues involved in disulfide bonds, where χ_2 is the dihedral angle determined by atoms C^α, C^β and S^γ of a given cysteinyl residue and S^γ of the other involved in the disulfide bond).

Also, we report r.m.s.d. for each amino acid type determined for the 6 proteins whose side-chain positions have been predicted in order to compare our results with those of Lee & Subbiah (1991). We do not consider r.m.s.d. calculated for all the side-chains of a particular protein, since the results depend on the relative number of large *versus* small side-chains in the sequence.

Statistics are calculated for buried and surface residues separately. Surface residues are defined here as side-chains that have an exposure that is more than 10% of the possible value. Buried residues, conversely, are defined as those with an exposure that is 10% or less of the possible value. The possible exposure is calculated as the surface area determined with a 1·6 Å spherical probe of the side-chain in question in the peptide Acetyl-Xxx-$NHCH_3$, with the backbone dihedral angle ϕ equal to $-60°$, and ψ equal to 140°. The peptide was minimized for 100 ABNR steps using the program CHARMM (Brooks *et al.*, 1983). From the resulting co-ordinates, the total accessible surface area of the side-chain was calculated for all atoms in the side-chain, excluding C^β and H^β atoms.

(d) *Automation of method*

The method is fully automated and has been used on a Convex C220, a Sun Sparcstation, an IBM RS 6000 and a SGI 340. It consists of the backbone-dependent rotamer library, a small number of Unix scripts, 2 FORTRAN programs, and the program CHARMM (version 22). CHARMM is first used to convert the Brookhaven Protein Data Bank (PDB) co-ordinates to CHARMM format. This is followed by a script, which finds and processes the ϕ and ψ values for the protein, and another than produces a file with the sequence of the protein in CHARMM format. If a homologous protein is used to help place the side-chains, the internal co-ordinates in CHARMM format are also calculated for this protein, and the χ angles are processed. A FORTRAN program is then used, in accord with subsection (b)(i)(b), above, to generate a CHARMM script that determines the initial positions of the side-chains, based on the sequence, the backbone dihedral angles, the backbone-dependent rotamer library, and the side-chain positions of a homologous protein (if one is being used). Once the disulfide and hydrogen atom minimizations have been performed, the van der Waals overlaps are calculated. A second FORTRAN program processes the overlaps, and following the rules of subsection (b)(ii)(a), sets up the CHARMM commands to search the alternative side-chain minima. The internal co-ordinates for the new minima are written out by the CHARMM program and used to build a new structure. The procedure continues (subsections (b)(ii)(b) and (b)(ii)(c), above) until all the clashes have been removed. The routines are quite flexible, and a variety of inputs can be used. In some cases (e.g. homology modeling), only a certain number of side-chains need to be modeled into a known structure. The starting structure simply has these side-chains deleted, and the routines build these side-chains. Once the PDB or CHARMM backbone co-ordinates (and any side-chain co-ordinates that are to be used) are processed, the entire procedure can be performed by running a single command file.

(e) *Computer time*

The initial placement of side-chains from the library takes only a few seconds of central processing unit time on a single processor of an SGI 340. The iterative minimizations to refine the structure can take from 6 h (crambin) to 24 h (thermolysin) on a single processor of an SGI 340, depending on the size of the protein.

3. Results

We first describe the backbone-dependent rotamer library and then present the results of applying it with the refinement methodology to the prediction of the side-chain conformations to a set of six proteins of known structure.

(a) *The backbone-dependent rotamer library*

In Table 4, the total number of each side-chain appears, and the actual and relative populations of the various rotamers are listed according to side-chain type. These results form a backbone-independent rotamer library that can be compared to that of Ponder & Richards (1987). They are essentially the same, except for the statistics for methionine, as already mentioned. In Table 5, which is constructed from the backbone-dependent rotamer library, we list the rotamer populations for values of ϕ and ψ for which there are more than ten examples of a particular side-chain type. One should note the large variation in populations of particular rotamers as a function of ϕ and ψ, and the identity of the side-chain. The variation is not limited to the differences between α-helices or β-sheets, but other regions of the ϕ,ψ map exhibit particular preferences as well. As an example, many side-chains prefer $\chi_1 = 180°$ in canonical α-helices ($\phi = -47°$, $\psi = -57°$), but in nearby regions of the Ramachandran map (more negative values of ϕ, and more positive values of ψ), $\chi_1 = -60°$ is much more common. This is true for the aromatic residues, Leu, the longer side-chains (Arg, Glu, Gln, Lys and Met), Cys and Val. The variation in the most probable value can also be compared with the average value in Table 4.

While many amino acids in specific ϕ,ψ ranges prefer one rotamer over all others, in some cases two or more rotamers have nearly equal populations. In the latter case, removing one protein (and hence 1 or more side-chains from the data set) may switch the balance between the two. This happens for ribonuclease, where adding 7rsa to the database changes the predictions of six side-chains for the better. This can happen even when there are many side-chains in a given ϕ,ψ block. For example, both Met29 and Met30 in 7rsa are in the same block. Without them, their χ_2 percentages are g^-, t, g^+

Table 4

Backbone-independent rotamer library

Res.	Number in Database		No. χ_1	%χ_1	Rotamer (Table 2)
Cys	434	$\chi_1 = 60 \pm 60$	54	12·4	1
		$\chi_1 = 180 \pm 60$	109	25·1	2
		$\chi_1 = -60 \pm 60$	267	61·5	3
Ser	1717	$\chi_1 = 60 \pm 60$	739	43·0	1
		$\chi_1 = 180 \pm 60$	424	24·7	2
		$\chi_1 = -60 \pm 60$	540	31·5	3
Thr	1460	$\chi_1 = 60 \pm 60$	673	46·1	1
		$\chi_1 = 180 \pm 60$	126	8·6	2
		$\chi_1 = -60 \pm 60$	657	45·0	3
Val	1683	$\chi_1 = 60 \pm 60$	142	8·4	1
		$\chi_1 = 180 \pm 60$	1176	69·9	2
		$\chi_1 = -60 \pm 60$	362	21·5	3

Res.	Number in Database		No. χ_1	%χ_1	No. χ_2	%χ_2	No. χ_2	%χ_2	No. χ_2	%χ_2	Rotamer (Table 2)
					$\chi_2 = 30 \pm 30$				$\chi_2 = -30 \pm 30$		
Pro	988	$\chi_1 = 30 \pm 30$	547	55·4	39	3·9	0	0·0	508	51·4	1
		$\chi_1 = 30 \pm 30$	441	44·6	433	43·8	0	0·0	8	0·4	3
					$\chi_2 = 30 \pm 30$		$\chi_2 = 90 \pm 30$		$\chi_2 = 150 \pm 30$		
Phe	889	$\chi_1 = 60 \pm 60$	120	13·5	1	0·1	118	13·3	1	0·1	1
		$\chi_1 = 180 \pm 60$	319	35·8	42	4·7	273	30·7	4	0·4	2
		$\chi_1 = -60 \pm 60$	450	50·6	35	3·9	342	38·4	73	8·2	3
					$\chi_2 = 30 \pm 30$		$\chi_2 = 90 \pm 30$		$\chi_2 = 150 \pm 30$		
His	488	$\chi_1 = 60 \pm 60$	54	11·0	4	0·8	47	9·6	3	0·6	1
		$\chi_1 = 180 \pm 60$	164	33·5	35	7·2	108	22·1	21	4·3	2
		$\chi_1 = -60 \pm 60$	270	55·2	23	4·7	181	37·0	6	13·5	3
					$\chi_2 = 30 \pm 30$		$\chi_2 = 90 \pm 30$		$\chi_2 = 150 \pm 30$		
Tyr	856	$\chi_1 = 60 \pm 60$	102	11·9	5	0·6	97	11·3	0	0·0	1
		$\chi_1 = 180 \pm 60$	293	34·1	49	5·7	241	28·1	3	0·3	2
		$\chi_1 = -60 \pm 60$	461	53·7	19	2·2	359	41·8	83	9·7	3
					$\chi_2 = 90 \pm 90$				$\chi_2 = -90 \pm 90$		
Trp	325	$\chi_1 = 60 \pm 60$	51	15·6	17	5·2			34	10·4	1, 3
		$\chi_1 = 180 \pm 60$	105	32·2	62	19·0			40	12·3	4, 6
		$\chi_1 = -60 \pm 60$	169	51·8	126	38·7			41	12·6	7, 9
					$\chi_2 = 60 \pm 60$		$\chi_2 = 180 \pm 60$		$\chi_2 = -60 \pm 60$		
Leu	1739	$\chi_1 = 60 \pm 60$	35	2·0	20	1·1	13	0·7	2	0·1	1, 2, 3
		$\chi_1 = 180 \pm 60$	573	32·9	475	27·3	80	4·6	14	0·8	4, 5, 6
		$\chi_1 = -60 \pm 60$	1131	64·9	135	7·7	949	54·5	44	2·5	7, **8**, 9
					$\chi_2 = 60 \pm 60$		$\chi_2 = 180 \pm 60$		$\chi_2 = -60 \pm 60$		
Ile	1176	$\chi_1 = 60 \pm 60$	176	15·0	17	1·4	154	13·1	5	0·4	1, 2, 3
		$\chi_1 = 180 \pm 60$	133	11·3	35	3·0	92	7·8	5	0·4	4, 5, 6
		$\chi_1 = -60 \pm 60$	867	73·7	35	3·0	668	56·8	162	13·9	7, **8**, 9
					$\chi_2 = -60 \pm 60$		$\chi_2 = 0 \pm 30$		$\chi_2 = 60 \pm 30$		
Asp	1342	$\chi_1 = 60 \pm 60$	244	18·1	50	3·7	156	11·6	38	2·8	1, 2, 3
		$\chi_1 = 180 \pm 60$	405	30·0	61	4·5	238	17·7	104	7·7	4, 5, 6
		$\chi_1 = -60 \pm 60$	693	51·4	240	17·8	404	30·0	47	3·5	7, **8**, 9
					$\chi_2 = -60 \pm 60$		$\chi_2 = 0 \pm 30$		$\chi_2 = 60 \pm 30$		
Asn	1048	$\chi_1 = 60 \pm 60$	180	17·1	37	3·5	93	8·8	50	4·7	1, 2, 3
		$\chi_1 = 180 \pm 60$	300	28·5	72	6·8	103	9·8	124	11·8	4, 5, 6
		$\chi_1 = -60 \pm 60$	568	53·9	312	29·6	197	18·7	56	5·3	7, **8**, 9
					$\chi_2 = 60 \pm 60$		$\chi_2 = 180 \pm 60$		$\chi_2 = -60 \pm 60$		
Met	399	$\chi_1 = 60 \pm 60$	34	8·5	4	1·0	27	6·7	3	0·7	1, 2, 3
		$\chi_1 = 180 \pm 60$	130	32·3	38	9·5	83	20·6	9	2·2	4, 5, 6
		$\chi_1 = -60 \pm 60$	235	58·5	6	1·5	138	34·3	90	22·4	7, 8, 9
					$\chi_2 = 60 \pm 60$		$\chi_2 = 180 \pm 60$		$\chi_2 = -60 \pm 60$		
Glu	1169	$\chi_1 = 60 \pm 60$	129	10·8	7	0·6	83	7·0	37	3·1	1, 2, 3
		$\chi_1 = 180 \pm 60$	383	32·2	78	6·6	288	24·2	17	1·4	4, 5, 6
		$\chi_1 = -60 \pm 60$	657	55·2	98	8·2	389	32·7	162	13·6	7, **8**, 9
					$\chi_2 = 60 \pm 60$		$\chi_2 = 180 \pm 60$		$\chi_2 = -60 \pm 60$		
Gln	808	$\chi_1 = 60 \pm 60$	68	8·3	6	0·7	52	6·3	9	1·1	1, 2, 3
		$\chi_1 = 180 \pm 60$	270	33·0	80	9·8	168	20·5	17	2·1	4, 5, 6
		$\chi_1 = -60 \pm 60$	470	57·4	29	3·5	315	38·5	126	15·4	7, **8**, 9
					$\chi_2 = 60 \pm 60$		$\chi_2 = 180 \pm 60$		$\chi_2 = -60 \pm 60$		
Arg	807	$\chi_1 = 60 \pm 60$	76	9·3	9	1·1	64	7·8	2	0·2	1, 2, 3
		$\chi_1 = 180 \pm 60$	253	30·9	43	5·2	199	24·3	9	1·1	4, 5, 6
		$\chi_1 = -60 \pm 60$	478	58·3	21	2·6	368	44·9	88	10·7	7, **8**, 9

Table 4 *(continued)*

Res.	Number in Database		No. χ_1	%χ_1	No. χ_2	%χ_2	No. χ_2	%χ_2	No. χ_2	%χ_2	Rotamer (Table 2)
					$\chi_2 = 60 \pm 60$		$\chi_2 = 180 \pm 60$		$\chi_2 = -60 \pm 60$		
Lys	1402	$\chi_1 = 60 \pm 60$	121	8·5	10	0·7	102	7·1	8	0·6	1, 2, 3
		$\chi_1 = 180 \pm 60$	477	33·4	90	6·3	360	25·2	27	1·9	4, 5, 6
		$\chi_1 = -60 \pm 60$	804	56·3	46	3·2	572	40·1	176	12·3	7, **8**, 9

Rotamer populations summed over the entire database.

For each amino acid, the total number of residues in the database is given (Number in Database), as well as a breakdown according to the χ_1 and χ_2 limits shown (all angles in degrees). χ_1 populations are broken down under the columns labeled No. χ_1 and %χ_1 for the total number and percentage of the side-chains of the given type in the database with χ_1 in the range denoted in the third column of each row. The χ_1 total add up to 100%. The remaining figures in the Table give the total number and percentages of particular χ_1/χ_2 combinations, for values of χ_1 and χ_2 denoted in the given row and column for each amino acid type. These χ_1/χ_2 percentage figures add up to 100%. The numbers in the last column refer to the conformation numbers listed in Table 2 and represented in the ϕ,ψ maps of Fig. 2. The numbers in bold type represent the most probable conformation for each amino acid type.

equal to 0, 40, 38, leading to a prediction of $t(180°)$; with them in the library, the percentages are 0, 39, 41, leading to the correct prediction for both of them (g^+ or $-60°$). This happens even though this ϕ,ψ block has 52 Met side-chains without 7rsa. In spite of such limitations, because the backbone selects different rotamers in different parts of the map, the predictive value of the backbone-dependent rotamer library is significantly higher than that of the average map. This will be discussed later in comparing predictions of the library in Table 4 (backbone-independent rotamer library) and the library in Table 5 (backbone-dependent rotamer library).

Figure 2 shows graphically the distribution of χ_1 and χ_2 values for the side-chains on Ramachandran (ϕ,ψ) plots. The numbers in Figure 2 refer to the numbered rotamer definitions in Table 2 with the most probable rotamer indicated. Residues with only χ_1 are represented by the numbers 1, 2 and 3 corresponding to χ_1 equal to 60°, 180° and −60°, respectively. Most other side-chains are represented by numbers 1 through 9 corresponding to three conformers for $\chi_1 = 60°$ ($\chi_2 = 60$, 180, $-60° \rightarrow$ numbers 1, 2, 3), $\chi_1 = 180°$ ($\chi_2 = 60$, 180, $-60° \rightarrow$ numbers 4, 5, 6), and $\chi_1 = -60°$ ($\chi_2 = 60$, 180, $-60° \rightarrow$ numbers 7, 8, 9). Aromatics have fewer possible conformations, and are listed in Table 2.

Figure 2 makes clear certain features of the relation between backbone and side-chain conformations that are useful for understanding protein structures. The amino acids can be grouped into a number of different kinds that exhibit similar behavior across the Ramachandran maps: (1) side-chains branched at C^β (Val, Ile, Thr); (2) side-chains branched at C^γ except Asp and Asn (aromatics, Leu); (3) Asp and Asn; (4) chains unbranched through C^δ (Arg, Lys, Met, Glu, Gln); (5) Ser and Cys; and (6) Pro.

The first group, side-chains possessing two γ heavy atoms, have steric requirements not found in other side-chains. Because of the definition of χ_1 of Val, conformations 1, 2, 3 of Val are equivalent to conformations of 2, 3, 1, respectively, of Thr and 4-6, 7-9, 1-3 of Ile. In this first group, the preferred conformations are strongly dependent on ψ and only weakly on ϕ. Values of $-30°$ and lower require a χ_1 of $-60°$ for Ile and Thr (equivalent to 180° for Val) to avoid clashes between the γ side-chain atoms and the backbone N of the succeeding residue; values of ψ from $-30°$ to $+40°$ yield mostly $\chi_1 = -60°$ ($+60°$ for Val), and β-sheet regions split at $\psi = 140°$ with $\chi_1 = +60°$ ($-60°$ for Val) below 140° and $\chi_1 = +60°$ ($-60°$ for Val) above 140°.

Side-chains with two δ heavy atoms (aromatics and Leu) are more complex in their behavior. In the α-helix region ($\phi,\psi = -57°, -47°$), these side-chains uniformly have $\chi_1 = 180°$. In nearby regions involving slightly unwound or distorted helices and turn conformations (type I with ϕ,ψ equal to $-60°$, $-30°$, type II′ with ϕ,ψ equal to $-80°$, 0° and type III with ϕ,ψ equal to $-60°$, $-30°$) $\chi_1 = -60°$ is strongly preferred. In the upper half of the Ramachandran map, χ_1 seems to vary more with ϕ than with ψ. At $\phi > -80°$ (e.g. type II turns), $\chi_1 = 180°$ (numbered 4, 5, 6 depending on χ_2) is common. In the middle region where most β-sheet conformations are found ($-140° < \phi < -180°$), $\chi_1 = -60°$ is common, and in the upper far left region ($\phi < -140°, \psi > 140°$) $\chi_1 = +60°$ occurs. Leucine has two predominant conformations, χ_1,χ_2 of $-60°,180°$ (numbered 8 in Fig. 2) and χ_1,χ_2 of $180°,60°$ (numbered 4 in Fig. 2). Near $\phi,\psi = 180°$, conformations with $\chi_1 = 60°$ are found. (Note: the Protein Data Bank uses the opposite orientation of $C^{\delta 1}$ and $C^{\delta 2}$ for leucine than IUPAC or CHARMM; the map uses the PDB definition.)

Residues Asn and Asp tend to have $\chi_1 = -60°$ (numbered 7, 8, 9) in α-helices, rather than $\chi_1 = 180°$. The distribution in the top half of the ϕ,ψ maps is dominated by ψ, with $\chi_1 = 180°$ conformations common below $\psi = 140°$. From $\psi = 140°$ to 160°, $\chi_1 = -60°$ is most common; above 160° (through 220°, or $-160°$, $\chi_1 = +60°$ is found. Since some positions are underpopulated (as shown by the numbers in italics in Figure 2), it is possible that some of the variation is caused by limitations in the data.

The longer side-chains, Met, Arg, Lys, Glu and Gln, all exhibit similar behavior; that is, the χ_1,χ_2 values are 180,180° in α-helices, +60,180° in the far upper left of the Ramachandran maps, some

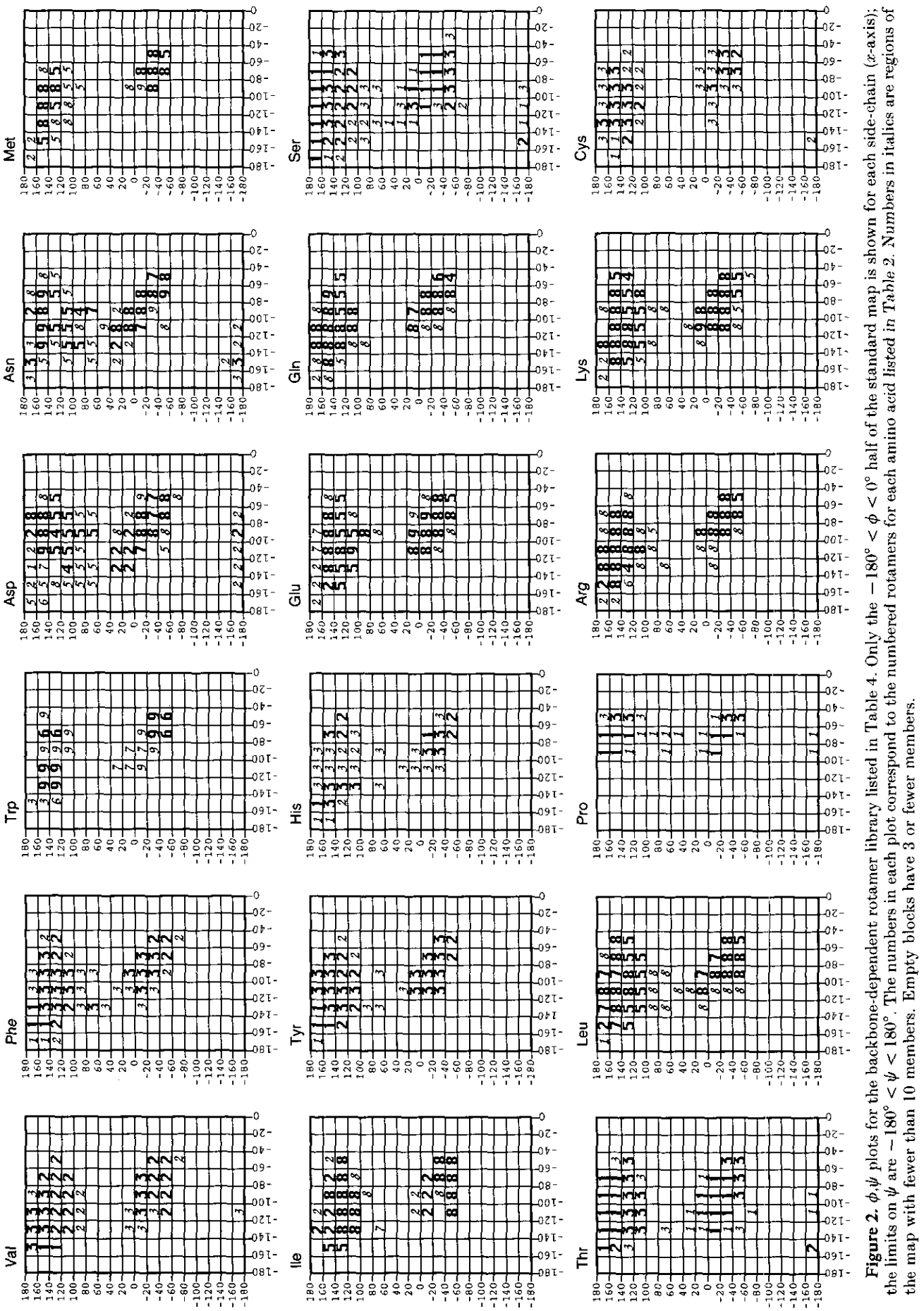


Figure 2. ϕ,ψ plots for the backbone-dependent rotamer library listed in Table 4. Only the $-180° < \phi < 0°$ half of the standard map is shown for each side-chain (x-axis); the limits on ψ are $-180° < \psi < 180°$. The numbers in each plot correspond to the numbered rotamers for each amino acid listed in Table 2. Numbers in italics are regions of the map with fewer than 10 members. Empty blocks have 3 or fewer members.

Table 5
Backbone-dependent rotamer library

1	2	3	4	5	6	7	8	9
C	10	−160	−140	120	140	0	70	30
C	15	−140	−120	120	140	0	27	73
C	16	−140	−120	140	160	0	6	94
C	16	−140	−120	160	180	25	0	75
C	14	−120	−100	100	120	0	86	14
C	14	−120	−100	120	140	0	43	57
C	16	−120	−100	140	160	6	6	88
C	19	−100	−80	−20	0	11	0	89
C	23	−100	−80	120	140	4	26	70
C	15	−100	−80	140	160	27	0	73
C	27	−80	−60	−60	−40	0	37	63
C	35	−80	−60	−40	−20	3	26	71
C	13	−80	−60	140	160	0	0	100
C	19	−60	−40	−60	−40	11	47	37
C	12	−60	−40	−20	−20	33	17	50
S	18	−180	−160	160	180	89	11	0
S	12	−160	−140	−180	−160	33	67	0
S	23	−160	−140	120	140	9	74	17
S	58	−160	−140	140	160	40	48	12
S	52	−160	−140	160	180	90	6	4
S	41	−140	−120	120	140	12	46	37
S	75	−140	−120	140	160	35	27	39
S	33	−140	−120	160	180	67	15	18
S	10	−120	−100	−60	−40	10	80	10
S	17	−120	−100	−20	0	71	6	18
S	21	−120	−100	0	20	43	14	43
S	17	−120	−100	100	120	6	53	41
S	43	−120	−100	120	140	2	51	47
S	29	−120	−100	140	160	31	28	38
S	21	−120	−100	160	180	62	10	29
S	16	−100	−80	−60	−40	25	31	44
S	32	−100	−80	−40	−20	47	16	38
S	67	−100	−80	−20	0	67	3	28
S	36	−100	−80	0	20	64	6	31
S	15	−100	−80	100	120	20	47	33
S	35	−100	−80	120	140	9	54	37
S	39	−100	−80	140	160	38	23	38
S	35	−100	−80	160	180	91	0	6
S	116	−80	−60	−60	−40	16	37	47
S	182	−80	−60	−40	−20	60	6	33
S	114	−80	−60	−20	0	82	2	16
S	10	−80	−60	100	120	0	60	40
S	38	−80	−60	120	140	5	61	34
S	75	−80	−60	140	160	47	27	25
S	30	−80	−60	160	180	87	0	10
S	63	−60	−40	−60	−40	6	38	54
S	78	−60	−40	−40	−20	53	12	36
S	15	−60	−40	−20	0	73	0	27
S	22	−60	−40	120	140	9	41	50
S	23	−60	−40	140	160	13	13	74
T	13	−160	−140	−180	−160	23	77	0
T	33	−160	−140	140	160	33	52	15
T	19	−160	−140	160	180	53	47	0
T	10	−140	−120	−180	−160	50	50	0
T	15	−140	−120	−20	0	73	20	7
T	19	−140	−120	100	120	11	0	89
T	67	−140	−120	120	140	7	6	87
T	76	−140	−120	140	160	62	11	28
T	62	−140	−120	160	180	89	8	3
T	15	−120	−100	−40	−20	87	0	13
T	42	−120	−100	−20	0	95	2	2
T	34	−120	−100	0	20	97	0	3
T	32	−120	−100	100	120	6	0	94
T	84	−120	−100	120	140	5	11	85
T	49	−120	−100	140	160	63	6	31
T	29	−120	−100	160	180	93	3	3
T	14	−100	−80	−60	−40	21	0	79
T	30	−100	−80	−40	−20	67	0	33
T	55	−100	−80	−20	0	95	0	5
T	12	−100	−80	0	20	100	0	0
T	15	−100	−80	100	120	0	7	93

Table 5 *(continued)*

T	46	−100	−80	120	140	0	4	93
T	27	−100	−80	140	160	67	11	19
T	32	−100	−80	160	180	94	3	3
T	112	−80	−60	−60	−40	4	1	95
T	114	−80	−60	−40	−20	54	6	39
T	41	−80	−60	−20	0	93	2	5
T	45	−80	−60	120	140	2	7	91
T	31	−80	−60	140	160	74	10	16
T	26	−80	−60	160	180	92	4	4
T	77	−60	−40	−60	−40	6	3	91
T	27	−60	−40	−40	−20	37	7	56
T	21	−60	−40	120	140	5	5	90
V	20	−160	−140	120	140	50	45	5
V	31	−160	−140	140	160	48	10	39
V	12	−160	−140	160	180	17	0	83
V	50	−140	−120	100	120	4	94	2
V	146	−140	−120	120	140	8	86	5
V	99	−140	−120	140	160	12	35	53
V	50	−140	−120	160	180	0	4	96
V	11	−120	−100	−60	−40	0	82	18
V	20	−120	−100	−20	0	5	15	80
V	71	−120	−100	100	120	0	97	3
V	181	−120	−100	120	140	7	88	4
V	49	−120	−100	140	160	14	43	43
V	12	−120	−100	160	180	0	0	100
V	13	−100	−80	−60	−40	8	92	0
V	15	−100	−80	−40	−20	0	53	47
V	13	−100	−80	−20	0	23	15	62
V	43	−100	−80	100	120	7	93	0
V	80	−100	−80	120	140	6	88	6
V	29	−100	−80	140	160	14	41	45
V	207	−80	−60	−60	−40	2	97	0
V	131	−80	−60	−40	−20	11	60	28
V	19	−80	−60	−20	0	32	21	47
V	15	−80	−60	100	120	0	93	7
V	62	−80	−60	120	140	2	94	5
V	27	−80	−60	140	160	0	56	44
V	109	−60	−40	−60	−40	2	92	6
V	39	−60	−40	−40	−20	28	54	18
V	16	−60	−40	120	140	6	75	19
P	18	−100	−80	−20	0	89	0	11
P	25	−100	−80	140	160	92	0	8
P	18	−100	−80	160	180	100	0	0
P	110	−80	−60	−40	−20	55	0	45
P	64	−80	−60	−20	0	86	0	14
P	70	−80	−60	120	140	63	0	37
P	194	−80	−60	140	160	56	0	44
P	66	−80	−60	160	180	82	0	18
P	49	−60	−40	−60	−40	22	0	78
P	124	−60	−40	−40	−20	33	0	67
P	84	−60	−40	120	140	25	0	75
P	67	−60	−40	140	160	22	0	78

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
F	10	−160	−140	120	140	30	66	10	0	30	0	0	60	0	0	10	0
F	36	−160	−140	140	160	69	6	25	0	69	0	0	6	0	0	25	0
F	22	−160	−140	160	180	95	5	0	0	95	0	5	0	0	0	0	0
F	10	−140	−120	60	80	0	10	90	0	0	0	10	0	0	0	80	10
F	12	−140	−120	100	120	0	67	33	0	0	0	25	33	8	0	33	0
F	37	−140	−120	120	140	8	41	49	3	3	3	0	38	3	3	41	5
F	48	−140	−120	140	160	31	4	65	0	31	0	0	4	0	0	65	0
F	16	−140	−120	160	180	56	0	44	0	56	0	0	0	0	0	38	6
F	10	−120	−100	−20	0	10	10	80	0	10	0	10	0	0	0	50	30
F	17	−120	−100	100	120	0	18	82	0	0	0	6	12	0	0	71	12
F	39	−120	−100	120	140	0	18	82	0	0	0	3	15	0	3	74	5
F	31	−120	−100	140	160	13	3	84	0	13	0	0	3	0	10	68	6
F	12	−100	−80	−40	−20	8	0	92	0	8	0	0	0	0	8	58	25
F	22	−100	−80	−20	0	23	9	68	0	23	0	0	9	0	5	45	18
F	12	−100	−80	0	20	0	0	100	0	0	0	0	0	0	0	83	17
F	13	−100	−80	100	120	0	23	77	0	0	0	8	15	0	8	69	0
F	36	−100	−80	120	140	0	47	53	0	0	0	6	42	0	3	44	6
F	28	−100	−80	140	160	0	11	89	0	0	0	0	11	0	7	64	18
F	89	−80	−60	−60	−40	0	79	21	0	0	0	12	66	0	2	13	6

Table 5 *(continued)*

F	65	−80	−60	−40	−20	3	29	68	0	3	0	2	28	0	8	42	18
F	12	−80	−60	−20	0	42	0	58	0	42	0	0	0	0	8	33	17
F	22	−80	−60	120	140	0	73	27	0	0	0	5	68	0	5	9	14
F	20	−80	−60	140	160	0	20	80	0	0	0	0	20	0	10	55	15
F	102	−60	−40	−60	−40	2	79	19	0	2	0	8	70	2	4	6	9
F	15	−60	−40	−40	−20	7	53	40	0	7	0	0	53	0	20	13	7
F	10	−60	−40	120	140	0	90	10	0	0	0	20	70	0	0	10	0
H	10	−160	−140	140	160	10	40	50	0	10	0	0	40	0	0	50	0
H	13	−160	−140	160	180	69	0	31	8	62	0	0	0	0	0	31	0
H	12	−140	−120	100	120	0	33	67	0	0	0	8	8	17	8	50	8
H	12	−140	−120	120	140	0	50	50	0	0	0	8	42	0	0	50	0
H	21	−140	−120	140	160	5	14	81	0	5	0	5	10	0	0	76	5
H	14	−100	−80	−40	−20	0	14	86	0	0	0	7	7	0	7	43	36
H	15	−100	−80	−20	0	20	0	80	7	13	0	0	0	0	0	53	27
H	54	−80	−60	−60	−40	2	52	44	0	2	0	11	35	6	11	20	13
H	39	−80	−60	−40	−20	8	26	67	0	8	0	3	21	3	8	46	13
H	14	−80	−60	−20	0	79	7	14	0	79	0	0	7	0	0	0	14
H	22	−80	−60	120	140	0	73	27	0	0	0	14	50	9	0	27	0
H	13	−80	−60	140	160	15	38	46	0	15	0	8	31	0	0	15	31
H	32	−60	−40	−60	−40	3	81	16	0	3	0	22	53	6	3	6	6
H	10	−60	−40	120	140	0	90	10	0	0	0	0	80	10	0	10	0
W	18	−140	−120	120	140	0	11	89	0	0	0	11	0	0	78	0	11
W	16	−140	−120	140	160	31	0	69	0	0	31	0	0	0	56	0	13
W	14	−120	−100	120	140	0	36	64	0	0	0	14	0	21	57	0	7
W	10	−120	−100	140	160	20	10	70	10	0	10	10	0	0	30	0	40
W	37	−80	−60	−60	−40	0	78	22	0	0	0	51	0	27	19	0	3
W	27	−80	−60	−40	−20	15	33	48	4	0	11	11	0	22	33	0	15
W	12	−80	−60	120	140	0	67	33	0	0	0	25	0	42	33	0	0
W	12	−80	−60	140	160	8	50	42	0	0	8	50	0	0	33	0	8
W	29	−60	−40	−60	−40	3	69	28	0	0	3	45	0	21	17	0	10
W	14	−60	−40	−40	−20	36	21	43	0	0	36	14	0	7	36	0	7
Y	13	−160	−140	120	140	8	77	15	0	8	0	23	54	0	0	15	0
Y	24	−160	−140	140	160	58	8	33	4	54	0	0	8	0	0	33	0
Y	27	−160	−140	160	180	93	4	4	0	93	0	0	4	0	0	4	0
Y	13	−140	−120	100	120	0	54	46	0	0	0	0	54	0	0	38	8
Y	47	−140	−120	120	140	4	28	68	0	4	0	2	26	0	4	62	2
Y	56	−140	−120	140	160	11	7	82	0	11	0	2	5	0	0	80	2
Y	22	−140	−120	160	180	59	0	41	0	59	0	0	0	0	5	36	0
Y	10	−120	−100	−40	−20	0	0	100	0	0	0	0	0	0	0	100	0
Y	14	−120	−100	−20	0	7	7	86	0	7	0	0	7	0	0	64	21
Y	23	−120	−100	0	20	0	0	100	0	0	0	0	0	0	0	74	26
Y	13	−120	−100	100	120	0	46	54	0	0	0	8	38	0	0	54	0
Y	34	−120	−100	120	140	0	21	76	0	0	0	3	18	0	0	71	6
Y	41	−120	−100	140	160	2	5	93	0	2	0	2	2	0	2	88	2
Y	11	−120	−100	160	180	9	0	91	0	9	0	0	0	0	0	82	9
Y	14	−100	−80	−40	−20	7	21	71	0	7	0	7	14	0	0	50	21
Y	24	−100	−80	−20	0	17	0	79	13	4	0	0	0	0	4	58	17
Y	10	−100	−80	0	20	10	0	90	0	10	0	0	0	0	0	70	20
Y	15	−100	−80	100	120	0	87	13	0	0	0	13	67	7	0	13	0
Y	28	−100	−80	120	140	0	64	36	0	0	0	21	43	0	0	29	7
Y	19	−100	−80	140	160	0	11	89	0	0	0	0	11	0	0	53	37
Y	11	−100	−80	160	180	18	9	73	0	18	0	0	9	0	0	64	9
Y	63	−80	−60	−60	−40	0	84	16	0	0	0	10	75	0	2	6	8
Y	52	−80	−60	−40	−20	10	37	54	0	10	0	6	31	0	12	19	23
Y	12	−80	−60	−20	0	25	8	67	0	25	0	0	8	0	8	42	17
Y	40	−80	−60	120	140	0	73	28	0	0	0	10	63	0	3	13	13
Y	21	−80	−60	140	160	14	19	67	0	14	0	0	19	0	5	48	14
Y	68	−60	−40	−60	−40	1	84	13	0	1	0	10	74	0	1	6	6
Y	14	−60	−40	−40	−20	36	29	36	0	36	0	14	14	0	0	7	29
L	11	−160	−140	120	140	9	91	0	0	9	0	64	27	0	0	0	0
L	15	−160	−140	140	160	27	33	40	20	7	0	27	7	0	7	33	0
L	14	−160	−140	160	180	64	21	14	29	29	7	14	0	7	0	14	0
L	23	−140	−120	100	120	0	83	17	0	0	0	52	26	0	0	17	0
L	36	−140	−120	120	140	0	56	44	0	0	0	42	14	0	8	36	0
L	52	−140	−120	140	160	6	4	90	0	6	0	2	2	0	19	63	8
L	10	−140	−120	160	180	20	0	80	20	0	0	0	0	0	10	70	0
L	15	−120	−100	0	20	0	0	100	0	0	0	0	0	0	7	93	0
L	59	−120	−100	100	120	0	54	46	0	0	0	41	10	2	12	34	0
L	111	−120	−100	120	140	0	50	50	0	0	0	41	8	1	9	38	3
L	56	−120	−100	140	160	0	5	95	0	0	0	4	2	0	14	79	2
L	12	−120	−100	160	180	0	0	100	0	0	0	0	0	0	25	75	0

Table 5 *(continued)*

L	16	−100	−80	−60	−40	0	31	69	0	0	0	25	0	6	6	63	0
L	24	−100	−80	−40	−20	0	17	83	0	0	0	17	0	0	8	71	4
L	39	−100	−80	−20	0	0	0	100	0	0	0	0	0	0	8	87	5
L	15	−100	−80	0	20	7	0	93	7	0	0	0	0	0	7	73	7
L	36	−100	−80	100	120	0	53	47	0	0	0	44	6	3	6	42	0
L	76	−100	−80	120	140	0	50	50	0	0	0	46	4	0	3	46	1
L	60	−100	−80	140	160	2	3	95	2	0	0	2	2	0	8	83	2
L	15	−100	−80	160	180	7	0	93	7	0	0	0	0	0	7	80	7
L	264	−80	−60	−60	−40	0	44	55	0	0	0	40	3	1	5	48	2
L	246	−80	−60	−40	−20	0	21	79	0	0	0	17	2	1	7	67	4
L	34	−80	−60	−20	0	3	9	88	3	0	0	9	0	0	15	71	3
L	15	−80	−60	100	120	0	80	20	0	0	0	80	0	0	0	20	0
L	70	−80	−60	120	140	0	59	41	0	0	0	44	11	3	9	31	1
L	62	−80	−60	140	160	0	13	87	0	0	0	11	2	0	10	74	3
L	135	−60	−40	−60	−40	0	56	44	0	0	0	46	9	1	4	39	1
L	51	−60	−40	−40	−20	0	25	75	0	0	0	24	2	0	8	61	6
L	16	−60	−40	120	140	0	69	31	0	0	0	56	13	0	0	31	0
L	10	−60	−40	140	160	10	0	90	0	10	0	0	0	0	20	70	0
I	14	−160	−140	120	140	0	57	43	0	0	0	7	50	0	0	36	7
I	24	−160	−140	140	160	21	75	4	0	21	0	8	63	4	0	4	0
I	31	−140	−120	100	120	0	3	97	0	0	0	0	3	0	6	74	16
I	104	−140	−120	120	140	3	13	83	0	3	0	1	12	1	5	67	11
I	47	−140	−120	140	160	45	26	30	0	45	0	4	21	0	0	30	0
I	28	−140	−120	160	180	93	7	0	14	79	0	0	7	0	0	0	0
I	10	−120	−100	−60	−40	0	10	90	0	0	0	0	10	0	0	80	10
I	13	−120	−100	−20	0	92	8	0	0	92	0	8	0	0	0	0	0
I	57	−120	−100	100	120	0	0	100	0	0	0	0	0	0	5	74	21
I	123	−120	−100	120	140	4	4	92	1	3	0	0	4	0	6	66	20
I	39	−120	−100	140	160	49	13	38	5	44	0	8	5	0	0	23	13
I	13	−100	−80	−60	−40	0	8	92	0	0	0	0	0	0	8	54	31
I	15	−100	−80	−20	0	73	13	13	0	73	0	7	0	7	7	7	0
I	43	−100	−80	100	120	0	2	98	0	0	0	0	2	0	2	67	28
I	68	−100	−80	120	140	6	3	91	3	3	0	0	3	0	0	66	25
I	11	−100	−80	140	160	27	9	64	0	27	0	0	9	0	0	45	18
I	165	−80	−60	−60	−40	0	4	96	0	0	0	2	1	1	1	84	12
I	85	−80	−60	−40	−20	13	21	66	2	11	0	13	7	1	2	47	16
I	14	−80	−60	−20	0	50	36	14	7	43	0	7	29	0	0	14	0
I	34	−80	−60	120	140	0	6	94	0	0	0	0	6	0	15	59	21
I	12	−80	−60	140	160	50	25	25	0	42	8	8	17	0	8	8	8
I	100	−60	−40	−60	−40	0	4	96	0	0	0	3	1	0	2	82	11
I	22	−60	−40	−40	−20	5	36	59	0	5	0	9	27	0	0	41	18
I	10	−60	−40	120	140	0	20	80	0	0	0	0	20	0	10	50	20
D	13	−140	−120	0	20	85	0	15	8	77	0	0	0	0	15	0	0
D	11	−140	−120	20	40	82	18	0	9	64	9	0	9	9	0	0	0
D	18	−140	−120	100	120	6	89	6	6	0	0	17	67	6	6	0	0
D	10	−120	−100	−20	0	50	0	50	30	20	0	0	0	0	40	10	0
D	26	−120	−100	0	20	58	12	31	0	58	0	0	4	8	19	12	0
D	11	−120	−100	20	40	82	0	18	18	64	0	0	0	0	9	0	9
D	43	−120	−100	100	120	0	81	19	0	0	0	12	65	5	5	9	5
D	15	−120	−100	120	140	0	53	47	0	0	0	27	13	13	7	33	7
D	15	−120	−100	140	160	7	0	93	0	0	7	0	0	0	60	20	13
D	13	−100	−80	−180	−160	92	8	0	0	85	8	8	0	0	0	0	0
D	27	−100	−80	−40	−20	0	7	93	0	0	0	0	4	4	37	52	4
D	46	−100	−80	−20	0	30	2	65	11	17	2	0	0	2	35	30	0
D	53	−100	−80	0	20	64	6	30	9	49	6	0	4	2	17	13	0
D	15	−100	−80	60	80	0	87	13	0	0	0	0	80	7	7	7	0
D	25	−100	−80	80	100	0	80	20	0	0	0	8	68	4	4	16	0
D	47	−100	−80	100	120	0	89	11	0	0	0	15	68	4	0	9	2
D	20	−100	−80	120	140	5	60	30	5	0	0	10	35	15	20	10	0
D	30	−100	−80	140	160	0	7	93	0	0	0	0	7	0	53	37	0
D	18	−100	−80	160	180	78	0	22	6	50	22	0	0	0	11	11	0
D	111	−80	−60	−60	−40	7	12	80	4	4	0	0	0	12	14	60	5
D	168	−80	−60	−40	−20	7	11	82	5	2	1	2	1	8	26	52	4
D	46	−80	−60	−20	0	26	11	63	4	17	4	2	0	7	17	41	4
D	21	−80	−60	100	120	0	86	14	0	0	0	10	76	0	0	10	5
D	42	−80	−60	120	140	0	69	31	0	0	0	21	31	17	12	17	2
D	33	−80	−60	140	160	6	21	70	3	3	0	3	15	3	27	36	6
D	12	−80	−60	160	180	50	0	50	17	17	17	0	0	0	0	50	0
D	98	−60	−40	−60	−40	7	14	78	2	5	0	0	4	10	15	58	4
D	43	−60	−40	−40	−20	7	12	79	7	0	0	0	0	12	5	70	5
D	11	−60	−40	120	140	0	64	36	0	0	0	9	36	18	0	27	9
D	13	40	60	20	40	0	15	85	0	0	0	8	0	8	54	31	0

Table 5 *(continued)*

D	20	40	60	40	60	5	20	75	0	0	5	0	5	15	35	35	5
D	10	60	80	20	40	0	10	90	0	0	0	0	10	0	40	50	0
N	10	−160	−140	−180	−160	100	0	0	20	30	50	0	0	0	0	0	0
N	18	−160	−140	160	180	72	17	11	11	22	39	11	6	0	6	6	0
N	14	−140	−120	20	40	57	14	29	7	36	14	7	0	7	21	7	0
N	13	−140	−120	80	100	0	92	8	0	0	0	8	46	38	8	0	0
N	12	−140	−120	100	120	0	92	8	0	0	0	25	42	25	0	0	8
N	16	−140	−120	140	160	19	0	81	0	6	13	0	0	0	56	13	6
N	10	−120	−100	−20	0	30	0	70	20	10	0	0	0	0	60	10	0
N	35	−120	−100	0	20	43	6	51	17	23	3	3	3	0	43	3	3
N	14	−120	−100	20	40	43	0	57	0	36	7	0	0	0	36	21	0
N	27	−120	−100	100	120	0	74	26	0	0	0	37	30	7	22	4	0
N	13	−120	−100	120	140	0	54	46	0	0	0	54	0	0	38	0	8
N	23	−120	−100	140	160	4	9	87	0	0	4	0	0	9	52	30	4
N	30	−100	−80	−20	0	13	7	80	7	7	0	0	3	3	50	27	3
N	50	−100	−80	0	20	40	8	50	6	34	0	0	2	6	22	22	6
N	16	−100	−80	60	80	19	38	44	13	6	0	0	31	6	31	13	0
N	14	−100	−80	80	100	7	71	21	7	0	0	7	50	14	7	7	7
N	30	−100	−80	100	120	0	73	27	0	0	0	20	53	0	13	10	3
N	20	−100	−80	120	140	10	35	55	5	0	5	25	5	5	30	15	10
N	16	−100	−80	140	160	13	6	81	6	6	0	0	0	6	56	19	6
N	10	−100	−80	160	180	60	30	10	10	30	20	0	20	10	0	10	0
N	73	−80	−60	−60	−40	1	23	74	0	0	1	4	0	19	32	38	4
N	96	−80	−60	−40	−20	3	13	84	1	2	0	1	0	11	34	44	6
N	39	−80	−60	−20	0	26	15	59	8	10	8	0	3	13	28	28	3
N	25	−80	−60	120	140	0	56	44	0	0	0	12	20	24	20	12	12
N	14	−80	−60	140	160	7	36	57	0	0	7	7	0	29	43	14	0
N	45	−60	−40	−60	−40	0	20	78	0	0	0	4	0	16	13	58	7
N	13	−60	−40	−40	−20	8	23	69	8	0	0	8	0	15	31	31	8
N	22	40	60	20	40	0	23	77	0	0	0	0	9	14	68	5	5
N	47	40	60	40	60	0	36	64	0	0	0	2	13	21	36	21	4
N	18	60	80	0	20	0	22	78	0	0	0	0	6	11	50	11	17
N	38	60	80	20	40	0	24	74	0	0	0	0	13	11	50	8	16
M	16	−160	−140	140	160	38	50	13	6	31	0	19	25	6	0	13	0
M	26	−140	−120	140	160	27	19	54	4	23	0	0	15	4	0	50	4
M	14	−120	−100	120	140	0	64	36	0	0	0	14	43	7	7	7	21
M	14	−120	−100	140	160	0	14	86	0	0	0	0	14	0	0	50	36
M	11	−100	−80	−40	−20	0	9	91	0	0	0	0	9	0	45	18	27
M	15	−100	−80	120	140	0	47	53	0	0	0	13	33	0	0	27	27
M	10	−100	−80	140	160	0	0	100	0	0	0	0	0	0	0	60	40
M	51	−80	−60	−60	−40	0	31	65	0	0	0	20	10	2	0	39	25
M	54	−80	−60	−40	−20	4	13	81	0	4	0	4	7	2	0	39	41
M	11	−80	−60	−20	0	27	18	55	9	9	9	9	9	0	0	27	27
M	10	−80	−60	120	140	0	70	30	0	0	0	10	50	10	0	10	20
M	36	−60	−40	−60	−40	0	58	42	0	0	0	14	42	3	0	36	6
M	17	−60	−40	−40	−20	12	35	53	0	12	0	18	18	0	0	18	35
E	15	−160	−140	120	140	0	87	13	0	0	0	7	80	0	7	7	0
E	11	−160	−140	140	160	55	36	9	0	45	0	0	36	0	0	9	0
E	16	−140	−120	100	120	0	75	25	0	0	0	19	56	0	0	25	0
E	16	−140	−120	120	140	0	56	44	0	0	0	25	31	0	0	38	6
E	32	−140	−120	140	160	16	16	69	0	16	0	0	16	0	0	47	19
E	13	−120	−100	−20	0	8	8	85	0	8	0	0	8	0	8	62	15
E	16	−120	−100	0	20	13	6	81	0	6	6	0	6	0	6	63	13
E	17	−120	−100	100	120	6	41	53	0	0	6	6	35	0	0	35	18
E	36	−120	−100	120	140	6	44	50	0	6	0	0	44	0	6	31	14
E	22	−120	−100	140	160	18	0	77	0	18	0	0	0	0	0	64	9
E	16	−100	−80	−40	−20	6	13	81	0	6	0	0	13	0	13	56	13
E	29	−100	−80	−20	0	10	0	86	3	3	0	0	0	0	3	41	41
E	11	−100	−80	0	20	9	0	91	0	0	9	0	0	0	18	45	27
E	10	−100	−80	80	100	20	40	40	0	10	10	20	20	0	10	20	10
E	18	−100	−80	100	120	0	67	33	0	0	0	11	56	0	0	28	6
E	27	−100	−80	120	140	4	59	37	0	4	0	7	48	4	4	22	11
E	30	−100	−80	140	160	17	20	63	3	10	3	10	10	0	3	37	23
E	200	−80	−60	−60	−40	1	44	54	0	1	1	11	31	2	9	34	12
E	229	−80	−60	−40	−20	12	22	65	0	10	2	5	16	1	11	39	14
E	51	−80	−60	−20	0	41	8	49	0	10	31	4	4	0	12	24	14
E	23	−80	−60	120	140	9	57	35	0	4	4	4	48	4	0	17	17
E	21	−80	−60	140	160	10	19	67	0	5	5	0	19	0	29	19	19
E	122	−60	−40	−60	−40	4	50	43	0	3	1	7	40	2	8	26	7
E	54	−60	−40	−40	−20	15	33	46	2	7	6	9	24	0	15	22	9
E	11	−60	−40	120	140	0	55	45	0	0	0	0	36	18	18	27	0

Table 5 *(continued)*

Q	15	−160	−140	140	160	40	20	40	0	33	7	0	20	0	0	27	13
Q	20	−140	−120	120	140	0	45	55	0	0	0	10	35	0	0	50	5
Q	20	−140	−120	140	160	15	0	80	0	15	0	0	0	0	0	60	20
Q	14	−140	−120	160	180	14	7	79	0	14	0	0	7	0	7	29	43
Q	15	−120	−100	0	20	7	7	87	0	7	0	7	0	0	0	67	20
Q	12	−120	−100	100	120	0	50	50	0	0	0	8	42	0	0	42	8
Q	35	−120	−100	120	140	3	46	51	0	3	0	9	37	0	3	34	14
Q	22	−120	−100	140	160	0	14	86	0	0	0	0	14	0	0	59	27
Q	10	−120	−100	160	180	10	10	80	0	10	0	0	0	10	0	40	40
Q	17	−100	−80	−40	−20	12	35	53	6	6	0	6	24	0	0	35	18
Q	22	−100	−80	−20	0	18	0	82	0	9	9	0	0	0	5	41	36
Q	17	−100	−80	0	20	6	0	94	6	0	0	0	0	0	0	59	35
Q	12	−100	−80	100	120	0	42	58	0	0	0	17	25	0	17	8	33
Q	29	−100	−80	120	140	3	55	41	0	3	0	17	34	0	3	34	3
Q	22	−100	−80	140	160	14	27	55	5	9	0	0	18	9	5	41	9
Q	101	−80	−60	−60	−40	0	46	52	0	0	0	13	27	4	4	41	8
Q	109	−80	−60	−40	−20	5	32	62	0	4	1	14	17	1	5	49	9
Q	30	−80	−60	−20	0	20	0	80	0	20	0	0	0	0	17	37	27
Q	26	−80	−60	120	140	12	62	27	0	8	4	35	27	0	0	19	8
Q	20	−80	−60	140	160	10	25	65	0	0	5	10	15	0	10	40	15
Q	60	−60	−40	−60	−40	2	65	33	2	0	0	25	32	8	3	18	12
Q	40	−60	−40	−40	−20	5	48	45	0	0	5	8	38	3	5	35	5
Q	11	−60	−40	120	140	9	82	9	0	9	0	27	55	0	9	0	0
Q	12	40	60	40	60	0	8	92	0	0	0	0	8	0	0	58	33
R	20	−160	−140	140	160	25	15	60	0	25	0	0	15	0	0	35	25
R	17	−160	−140	160	180	53	18	29	12	41	0	6	12	0	0	18	12
R	21	−140	−120	120	140	5	52	33	5	0	0	10	43	0	0	24	10
R	35	−140	−120	140	160	11	23	63	0	11	0	0	23	0	0	40	23
R	11	−140	−120	160	180	36	0	64	0	27	9	0	0	0	0	36	27
R	13	−120	−100	100	120	0	23	77	0	0	0	8	15	0	0	69	8
R	23	−120	−100	120	140	0	48	52	0	0	0	13	35	0	0	52	0
R	24	−120	−100	140	160	4	17	79	0	4	0	0	17	0	4	58	17
R	15	−120	−100	160	180	13	0	87	7	7	0	0	0	0	7	73	7
R	22	−100	−80	−40	−20	5	14	73	0	5	0	9	5	0	0	55	18
R	25	−100	−80	−20	0	8	0	92	0	8	0	0	0	0	8	76	8
R	11	−100	−80	0	20	9	9	82	0	9	0	9	0	0	0	64	18
R	27	−100	−80	120	140	4	48	48	0	4	0	0	48	0	7	37	4
R	22	−100	−80	140	160	9	14	77	5	5	0	0	14	0	0	41	36
R	103	−80	−60	−60	−40	1	49	50	0	1	0	10	38	0	2	41	8
R	107	−80	−60	40	−20	9	26	63	2	7	0	7	17	3	1	54	7
R	23	−80	−60	−20	0	26	4	70	4	17	0	0	4	0	0	57	13
R	27	−80	−60	120	140	0	41	56	0	0	0	7	33	0	4	48	4
R	22	−80	−60	140	160	9	18	73	0	9	0	0	18	0	0	68	5
R	72	−60	−40	−60	−40	0	64	36	0	0	0	8	53	3	3	31	3
R	27	−60	−40	−40	−20	22	37	41	4	19	0	19	19	0	0	30	11
K	17	−160	−140	120	140	0	82	18	0	0	0	35	41	6	0	18	0
K	28	−160	−140	140	160	29	32	36	0	29	0	4	29	0	11	21	4
K	10	−140	−120	100	120	20	60	20	0	10	10	20	40	0	0	20	0
K	39	−140	−120	120	140	0	46	51	0	0	0	8	38	0	3	41	5
K	45	−140	−120	140	160	16	13	71	0	16	0	2	11	0	4	49	18
K	18	−140	−120	160	180	17	0	83	0	17	0	0	0	0	0	67	17
K	10	−120	−100	−40	−20	10	20	70	0	10	0	0	20	0	0	40	30
K	16	−120	−100	−20	0	0	13	88	0	0	0	6	6	0	6	50	31
K	21	−120	−100	0	20	5	0	95	0	0	5	0	0	0	0	71	24
K	16	−120	−100	100	120	6	56	38	0	6	0	13	38	6	0	25	13
K	40	−120	−100	120	140	3	33	65	0	3	0	10	18	5	3	53	10
K	33	−120	−100	140	160	6	15	76	0	6	0	3	12	0	9	48	18
K	21	−120	−100	160	180	10	5	86	5	0	5	0	5	0	5	62	19
K	29	−100	−80	−40	−20	7	24	69	0	7	0	10	14	0	10	38	17
K	48	−100	−80	−20	0	13	6	81	0	6	6	0	4	2	6	52	23
K	19	−100	−80	0	20	16	0	84	5	11	0	0	0	0	11	42	26
K	27	−100	−80	100	120	4	70	26	0	4	0	11	56	4	0	15	11
K	33	−100	−80	120	140	0	52	45	0	0	0	12	39	0	3	33	9
K	38	−100	−80	140	160	3	3	92	0	3	0	0	3	0	5	61	26
K	176	−80	−60	−60	−40	4	52	41	1	3	0	10	38	4	2	30	9
K	188	−80	−60	−40	−20	10	27	60	2	7	1	5	21	1	2	44	13
K	54	−80	−60	−20	0	19	7	74	0	19	0	0	7	0	2	48	22
K	10	−80	−60	100	120	0	40	50	0	0	0	10	30	0	0	50	0
K	48	−80	−60	120	140	8	54	38	0	6	2	13	40	2	0	27	10
K	28	−80	−60	140	160	4	11	86	0	4	0	0	11	0	11	61	11
K	134	−60	−40	−60	−40	2	57	40	0	2	0	7	46	4	3	31	4
K	74	−60	−40	−40	−20	9	34	51	1	8	0	8	23	3	1	41	9
K	23	−60	−40	120	140	4	74	22	4	0	0	4	61	9	0	22	0

Table 5 *(continued)*

K	10	−60	−40	140	160	10	50	40	0	10	0	10	40	0	0	40	0
K	11	40	60	40	60	0	0	91	0	0	0	0	0	0	0	82	9

Column 1, residue (1-letter code); column 2, number of side-chains in ϕ,ψ range in structure database; column 3, lower ϕ limit; column 4, upper ϕ limit; column 5, lower ψ limit; column 6, upper ψ limit (all angles in degrees); columns 7 to 9, χ_1 populations of 60°, 180°, −60° rotamers (total = 100%); columns 10 to 18, χ_1,χ_2 rotamer populations according to definitions of 1 to 9 for applicable residues as defined in Table 2 (total = 100%).

180,180° values near $\phi,\psi = -60$, 120° (except Arg) and near $\phi,\psi = -140°, -120°$, and $\chi_1,\chi_2 = -60,180°$ (number 8) nearly everywhere else.

Serine is similar to Thr in most regions of the map. However, there is a large difference in the $\psi = 80$ to 140° region, where $\chi_1 = 180°$ is common for Ser while Thr prefers −60° to avoid contact with the backbone carbonyl oxygen atom. Serine prefers +60° in much of the map, as does Thr, to make hydrogen bonds to the backbone. Since Cys cannot do this, this conformation is largely absent and only the $\chi_1 = 180$ and −60° are found commonly through most of the map with $\chi_1 = 180°$ for $\phi =$ 100° to 120° and in α-helices, and $\chi_1 = -60°$ in most of the rest of the map. Free cysteinyl residues and disulfide-bonded cysteinyl residues were not distinguished in calculating the library.

Finally proline, as noted by Cung *et al.* (1987), exhibits the C^γ-*exo* conformation for $\phi > -60°$ and the C^γ-*endo* conformation for $\phi < -60°$.

(b) *Prediction of side-chain conformations in proteins from the known backbone co-ordinates*

We applied the targ/lib method (see Table 3) to six proteins in the Brookhaven Protein Database by using the backbone co-ordinates from the X-ray structures and initially building the side-chains from the ϕ,ψ rotamer library. These proteins are rhizopuspepsin (C-terminal domain: PDB code 2apr), lysozyme (1lz1), crambin (1crn), bovine pancreatic trypsin inhibitor (5pti), ribonuclease A (7rsa), and thermolysin (3tln). All of these structures were used in the library except 1lz1 (2·0 Å resolution) and 3tln (1·6 Å resolution), which are represented by a highly homologous structure. The relevant information is listed in Table 1. In each case, a library of the form of Table 5 was recalculated with the protein to be predicted and its homologs removed.

As an example, we give detailed results for the small protein crambin in Table 6, for the structures numbered 0, 1 and N in Figure 1 (i.e. from the library alone, after backbone/side-chain clashes have been resolved, and after all side-chain/side-chain clashes are resolved; $N = 2$ in this case). Table 6 lists the experimental χ angles as well as the χ angles predicted from the backbone-dependent rotamer library. All of the initial χ angles are either 60, 180, −60, 0, 90, or −90° except for those of proline and cysteine. Cysteine co-ordinates are minimized in determining each of the structures (see Methods, section (b)(i)(c)), and proline χ_1 is set equal to 28 or −28°, which along with the backbone co-ordinates determines χ_2 of proline. We note that, as in the library, residues such as Asn and His are deemed correct in χ_2 if χ_2 or $\chi_2+180°$ is correct within 40° of the crystal structure. This takes account of the fact already mentioned, that it is usually not possible to distinguish the two positions in the X-ray structure.

In predicting the side-chain orientations for crambin, the backbone-dependent library does well (see Structure 0 in Table 6). Only Thr1, Arg10, Tyr29 and Asn46 are moved in the first set of side-chain minimizations to take care of clashes with the backbone. In the series of minimizations to remove side-chain/side-chain clashes, only Phe13 and Arg17 are moved. Phe13 remains in a good conformation and Arg17 is moved from an incorrect to a correct conformation. Of the four residues that are in incorrect conformations in the final structure, three were never minimized (Leu18, Ile25 and Asp43), and one (Tyr29) was minimized in the first round, but remained in a conformation with the incorrect χ_1 (near −60°, instead of near 180°). Minimizing all of the side-chains at once (data not shown) was found not to improve the final pedictions for crambin or for the other proteins tested here. However, minimizing both the X-ray structure and the final predicted structure with the minimization protocol of Summers & Karplus (1989) (a series of Powell minimizations with decreasing harmonic force constraints on side-chain atom positions) produces generally lower average r.m.s.d. for all side-chain types. In many cases, the predicted and experimental angles are identical. This demonstrates that the predicted and the X-ray positions are in the same local minimum of the force field (see Table 9, 4th column).

The deviations from the X-ray structures for the residues in all six proteins for χ_1 and χ_2 are presented in the stacked histograms of Figure 3. The numbers and fractions of residues correct to within 40° are listed in Table 7 for structures 0, 1 and N ($N = 2$, 3 or 4 for all cases tested here). Also listed are the results predicted directly (without refinement) from the backbone-independent library of Table 4 for comparison with Structure 0 for each protein, predicted directly from the backbone-dependent library. The results are broken down into χ_1, χ_2 and χ_{1+2} predictions. Since the results vary significantly from protein to protein, it is clear that a prediction method cannot be assessed on the basis of tests on one or two proteins (e.g. Desmet *et al.*, 1992). Lysozyme gives the poorest result, probably because it has a high content of charged residues. As already mentioned, they are difficult to predict,

Table 6

Side-chain results for crambin from backbone co-ordinates only

Res. no.	χ	Type	Exp.	Structure 0 Pred (Dif) Cor?	Structure 1 Pred (Dif) Cor?	Structure $N(N=2)$ Pred (Dif) Cor?
1	1	Thr	−59	60 (119)n	−47 (13)y	−47 (13)y
2	1	Thr	56	60 (4)y	60 (4)y	60 (4)y
3	1	Cys	−51	−53 (−1)y	−51 (0)y	−51 (0)y
3	2	Cys	−75	−74 (1)y	−75 (0)y	−75 (0)y
4	1	Cys	−65	−48 (16)y	−66 (−2)y	−69 (−4)y
4	2	Cys	−83	−101 (−19)y	−82 (1)y	−79 (4)y
5	1	Pro	32	28 (−5)y	28 (−5)y	28 (−5)y
5	2	Pro	−41	−39 (3)y	−39 (3)y	−39 (3)y
6	1	Ser	69	60 (−9)y	60 (−9)y	60 (−9)y
7	1	Ile	−74	−60 (14)y	−60 (14y)	−60 (14)y
7	2	Ile	173	−180 (7)y	180 (7)y	180 (7)y
8	1	Val	160	−180 (20)y	180 (20)y	180 (20)y
10	1	Arg	177	−60 (123)n	180 (3)y	180 (3)y
10	2	Arg	64	−180 (116)n	65 (1)y	65 (1)y
10	3	Arg	67	−180 (113)n	72 (5)y	72 (5)y
10	4	Arg	177	−180 (3)y	175 (−2)y	175 (−2)y
11	1	Ser	−66	−60 (6)y	−60 (6)y	−60 (6)y
12	1	Asn	−70	−60 (10)y	−60 (10)y	−60 (10)y
12	2	Asn	−23	0 (23)y	0 (23)y	0 (23)y
13	1	Phe	−175	−180 (−5)y	−180 (−5)y	−167 (7)y
13	2	Phe	−90	90 (0)y	90 (0)y	70 (−20)y
14	1	Asn	−73	−60 (13)y	−60 (13)y	−60 (13)y
14	2	Asn	−24	0 (24)y	0 (24)y	0 (24)y
15	1	Val	171	−180 (9)y	180 (9)y	180 (9)y
16	1	Cys	180	178 (−1)y	178 (−1)y	178 (−1)y
16	2	Cys	−93	−89 (4)y	−90 (4)y	−90 (3)y
17	1	Arg	−67	−60 (7)y	−60 (7)y	−65 (2)y
17	2	Arg	−80	−180 (−100)n	180 (−100)n	−73 (7)y
17	3	Arg	−72	60 (132)n	60 (132)n	−78 (−6)y
17	4	Arg	157	−180 (23)y	−180 (23)y	128 (−30)y
18	1	Leu	−76	−180 (−104)n	−180 (−104)n	−180 (−104)n
18	2	Leu	−63	−180 (−117)n	−180 (−117)n	−180 (−117)n
19	1	Pro	13	28 (15)y	28 (15)y	28 (15)y
19	2	Pro	−14	−37 (−23)y	−37 (−23)y	−37 (−23)y
21	1	Thr	−45	−60 (−15)y	−60 (−15)y	−60 (−15)y
22	1	Pro	−24	−28 (−3)y	−28 (−3)y	−28 (−3)y
22	2	Pro	33	30 (−2)y	30 (−2)y	30 (−2)y
23	1	Glu	−72	−60 (12)y	−60 (12)y	−60 (12)y
23	2	Glu	−172	−180 (−8)y	−180 (−8)y	−180 (−8)y
23	3	Glu	−22	0 (22)y	0 (22)y	0 (22)y
25	1	Ile	−75	−60 (15)y	−60 (15)y	−60 (15)y
25	2	Ile	−72	−180 (−108)n	−180 (−108)n	−180 (−108)n
26	1	Cys	−65	−64 (1)y	−63 (2)y	−63 (3)y
26	2	Cys	−58	−59 (−1)y	−60 (−2)y	−60 (−2)y
26	3	Cys	−86	−86 (0)y	−85 (1)y	−84 (2)y
28	1	Thr	53	60 (7)y	60 (7)y	60 (7)y
29	1	Tyr	−178	−60 (118)n	−77 (101)n	−77 (101)n
29	2	Tyr	55	90 (35)y	7 (132)n	7 (132)n
30	1	Thr	61	60 (−1)y	60 (−1)y	60 (−1)y
32	1	Cys	−54	−57 (−3)y	−57 (−3)y	−56 (−2)y
32	2	Cys	−118	−106 (12)y	−113 (5)y	−116 (2)y
32	3	Cys	106	101 (−4)y	105 (−1)y	106 (0)y
33	1	Ile	65	60 (−5)y	60 (−5)y	60 (−5)y
33	2	Ile	171	−180 (9)y	−180 (9)y	−180 (9)y
34	1	Ile	−60	−60 (0)y	−60 (0)y	−60 (0)y
34	2	Ile	168	−180 (12)y	−180 (12)y	−180 (12)y
35	1	Ile	65	60 (−5)y	60 (−5)y	60 (−5)y
35	2	Ile	169	−180 (11)y	180 (11)y	180 (11)y
36	1	Pro	4	28 (23)y	28 (23)y	28 (23)y
36	2	Pro	−4	−36 (−32)y	−36 (−32)y	−36 (−32)y
39	1	Thr	−52	−60 (−8)y	−60 (−8)y	−60 (−8)y
40	1	Cys	−63	−65 (−2)y	−65 (−2)y	−65 (−2)y
40	2	Cys	−75	−72 (3)y	−72 (3)y	−73 (3)y
40	3	Cys	−79	−79 (0)y	−78 (1)y	−78 (1)y
41	1	Pro	28	28 (0)y	28 (0)y	28 (0)y
41	2	Pro	−35	−41 (−6)y	−41 (−6)y	−41 (−6)y
43	1	Asp	59	−60 (−119)n	−60 (−119)n	−60 (−119)n
43	2	Asp	−24	−60 (−36)y	−60 (−36)y	−60 (−36)y

Table 6 *(continued)*

Res. no.	χ	Type	Exp.	Structure 0 Pred (Dif) Cor?	Structure 1 Pred (Dif) Cor?	Structure $N(N=2)$ Pred (Dif) Cor?
44	1	Tyr	−74	−60 (14)y	−60 (14)y	−60 (14)y
44	2	Tyr	86	90 (4)y	90 (4)y	90 (4)y
46	1	Asn	−57	−60 (−3)y	−60 (−4)y	−60 (−4)y
46	2	Asn	113	−60 (7)y	−68 (−2)y	−68 (−1)y

χ angle predictions (in degrees) for rounds 0 (library), 1 (after backbone/side-chain conflicts have been resolved), and $N = 2$ (after all side-chain conflicts have been resolved). Differences between prediction (Pred) and the experimental structure (Exp.) are listed under column Dif. If the χ angle is correct to within 40°, then a y is entered in the Cor? column, otherwise, an n is listed.

since the conformations in the crystal structure often depend on the effects of solvent and interactions with other proteins (Gelin & Karplus, 1979).

To compare the present results with those of Lee & Subbiah (1991) we consider χ_1 and χ_2 independent of χ_1, since they did not report χ_{1+2}. For crambin, our results are 92% and 89% (for χ_1 and χ_2, respectively), compared with 70% and 60% for Lee & Subbiah (1991); for bovine pancreatic trypsin inhibitor (BPTI), 85% and 78% compared with 76% and 55%; for ribonuclease, 79% and 71% compared with 58% and 61%; and for lysozyme, 77% and 66% compared with 80% and 68%, which is the only protein where the results of Lee & Subbiah (1991) are better than those reported here. For the other two proteins (thermolysin and penicillopepsin, a homolog of rhizopuspepsin), Lee & Subbiah (1991) report only core residue predictions and these are compared below.

It is of interest also to compare the initial placement resulting from the backbone-dependent rotamer library (Structure 0) with the predictions that are obtained from the backbone-independent library of Table 4. As noted before, this library is essentially that of Ponder & Richards (1987), except for methionine χ_2. Desmet *et al.* (1992) have used the Ponder & Richards (1987) rotamers as the beginning of their prediction scheme, so these results reflect the starting structure in their method for each protein studied here. The refinement technique used by them is different from the present one, although the same general principle (removal of van der Waals clashes) is involved. Results for the side-chain placement with the backbone-independent library are listed in the last column of Table 7. Most of the difference between the backbone-independent and backbone-dependent results concern χ_1. The χ_1 predictions from the backbone-independent and backbone-dependent libraries are 52% and 67% for thermolysin, 65% and 80% for BPTI, 68% and 86% for crambin, 64% and 76% for lysozyme, 54% and 71% for the rhizopuspepsin C-terminal domain, and 56% and 72% for ribonuclease. The χ_2 prediction rates are all within 7% from the two libraries. The χ_{1+2} results reflect those for χ_1, and so differ significantly between the libraries. Since the refinement procedure works better when more of the side-chains are close to their experimental conformation, the better placement from the backbone-dependent library is very useful in obtaining the correct conformation for side-chains that are not correct in the initial model.

In Figures 4 and 5, stereo plots of the various side-chain predictions are compared with the X-ray values for BPTI and crambin, respectively. In (a) of each Figure, the X-ray structure is presented alone; in (b) the X-ray structure is compared with the results from the backbone-independent library; in (c) are shown the X-ray structure and the initial backbone-dependent library prediction; and in (d) the X-ray structure and the final refined prediction are compared; (e) shows the X-ray structure and final predicted structure that have both been minimized according to the method of Summers & Karplus (1989). In accord with the numerical results, the Figures illustrate that the predicted and X-ray side-chain positions for most residues are in the same energy minimum and that there is a general improvement in going from (b) to (e).

In Figure 6, the results for core and surface residues are compared with the results for all residues for the six proteins, where core residues are those defined as having less than 10% exposure, and surface residues have greater than 10% exposure (see Methods, section (c)). For five out of six of the proteins, buried residues are more accurately predicted than exposed residues. This is true of the results obtained directly from the rotamer library and even more so after the side-chain minimizations have been performed. BPTI is the sole exception. Since BPTI is quite small (58 residues), the difference between buried and accessible residue predictions corresponds to only two or three residues. Lee & Subbiah (1991) report results for the core region of thermolysin and penicillopepsin, and they predict χ_1 and χ_2 with fractions of 82% and 76% correct for thermolysin and 81% and 81% for penicillopepsin. For comparison, in the core region of thermolysin we predict 78% and 80% of the residues correctly, and in the core region of the C-terminal domain of rhizopuspepsin, a protein homologous to penicillopepsin, we predict 88% and 83% of the residues correctly.

In Tables 8 and 9, the results for the various types of side-chains are summarized for the six proteins. The library does well for hydrophobic

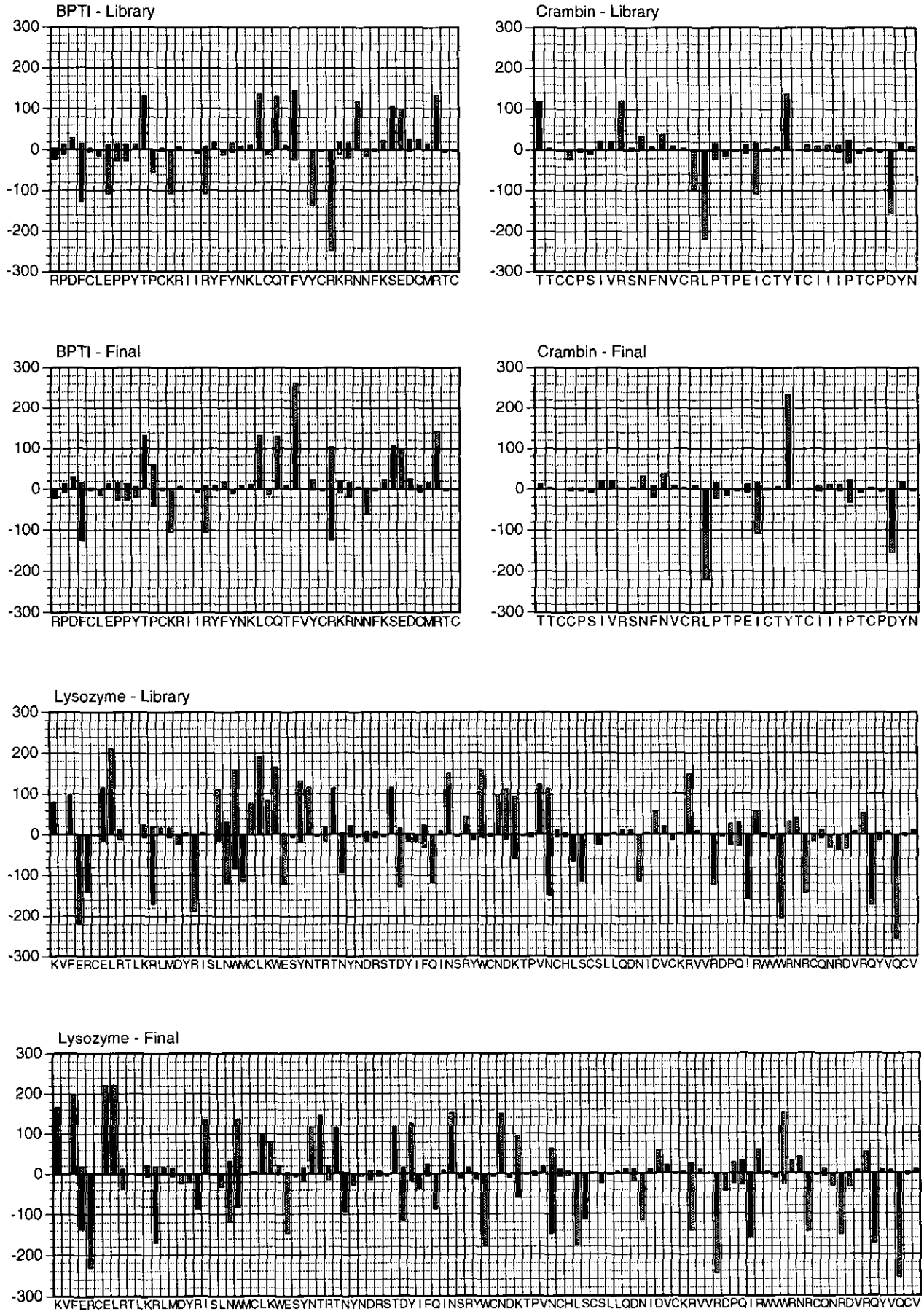

BPTI - Library
300
200
100
0
-100
-200
-300
RPDFCLEPPYTPCKRIIRYFYNKLCQTFVYCRKRNNFKSEDCMRTC
Crambin - Library
TTCCPSIVRSNFNVCRLPTPEICTYTCIIIPTCPDYN
BPTI - Final
Crambin - Final
Lysozyme - Library
KVFERCELRTLKRLMDYRISLNWMCLKWESYNTRTNYNDRSTDYIFQINSRYWCNDKTPVNCHLSCSLLQDNIDVCKRVVRDPQIRWWWRNRCQNRDVRQYVQCV
Lysozyme - Final

Fig. 3.

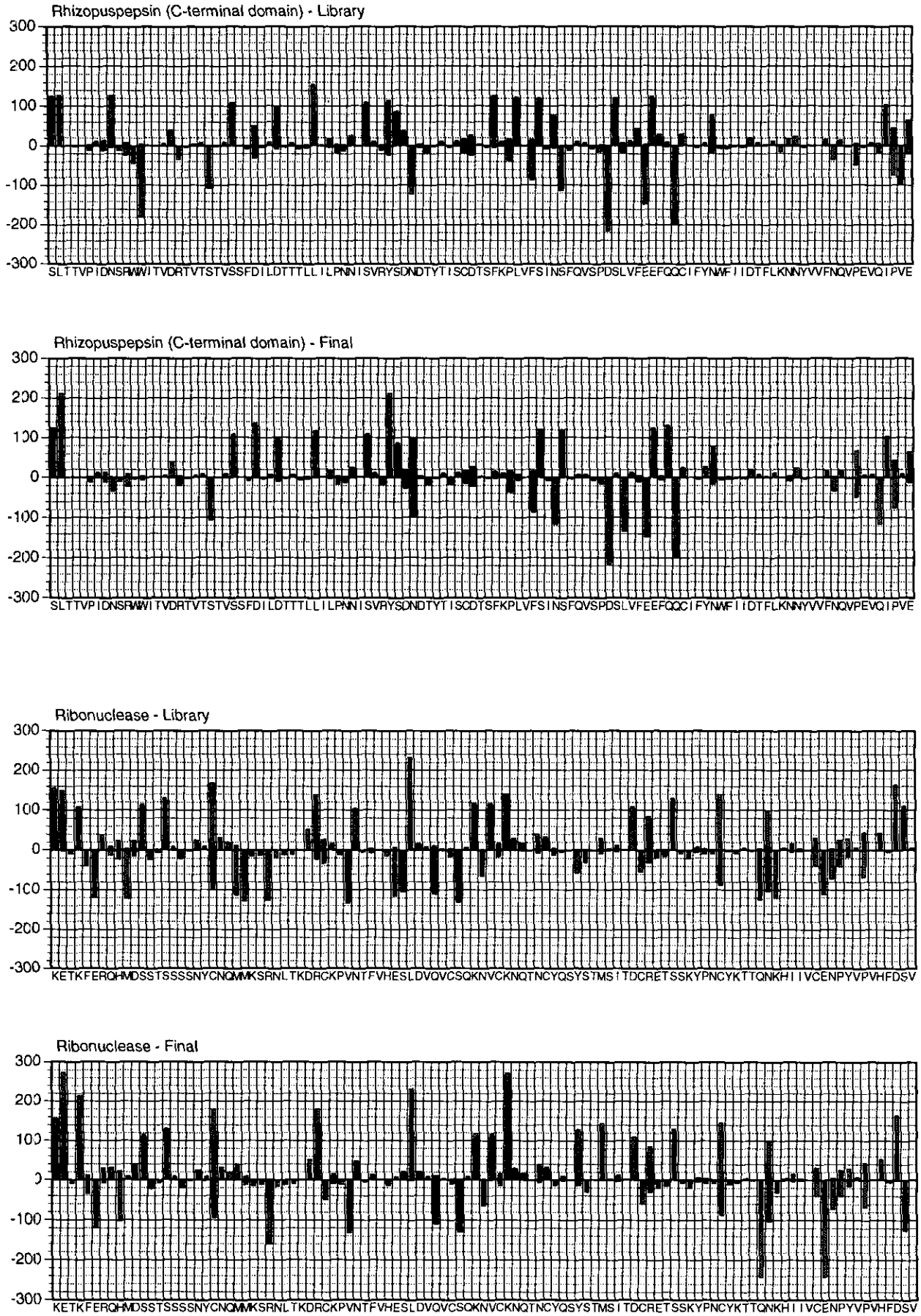

Rhizopuspepsin (C-terminal domain) - Library
300
200
100
0
-100
-200
-300
SLTTVPIDNSRWWITVDRTVTSTVSSFDILDTTTLLILPNNISVRYSDNDTYTISCDTSFKPLVFSINSFQVSPDSLVFEEFQQCIFYNWFIIDTFLKNNYVVFNQVPEVQIPVE
Rhizopuspepsin (C-terminal domain) - Final
300
200
100
0
-100
-200
-300
SLTTVPIDNSRWWITVDRTVTSTVSSFDILDTTTLLILPNNISVRYSDNDTYTISCDTSFKPLVFSINSFQVSPDSLVFEEFQQCIFYNWFIIDTFLKNNYVVFNQVPEVQIPVE
Ribonuclease - Library
300
200
100
0
-100
-200
-300
KETKFERQHMDSSTSSSSNYCNQMMKSRNLTKDRCKPVNTFVHESLDVQVCSQKNVCKNQTNCYQSYSTMSITDCRETSSKYPNCYKTTQNKHIIVCENPYVPVHFDSV
Ribonuclease - Final
300
200
100
0
-100
-200
-300
KETKFERQHMDSSTSSSSNYCNQMMKSRNLTKDRCKPVNTFVHESLDVQVCSQKNVCKNQTNCYQSYSTMSITDCRETSSKYPNCYKTTQNKHIIVCENPYVPVHFDSV

Fig. 3.

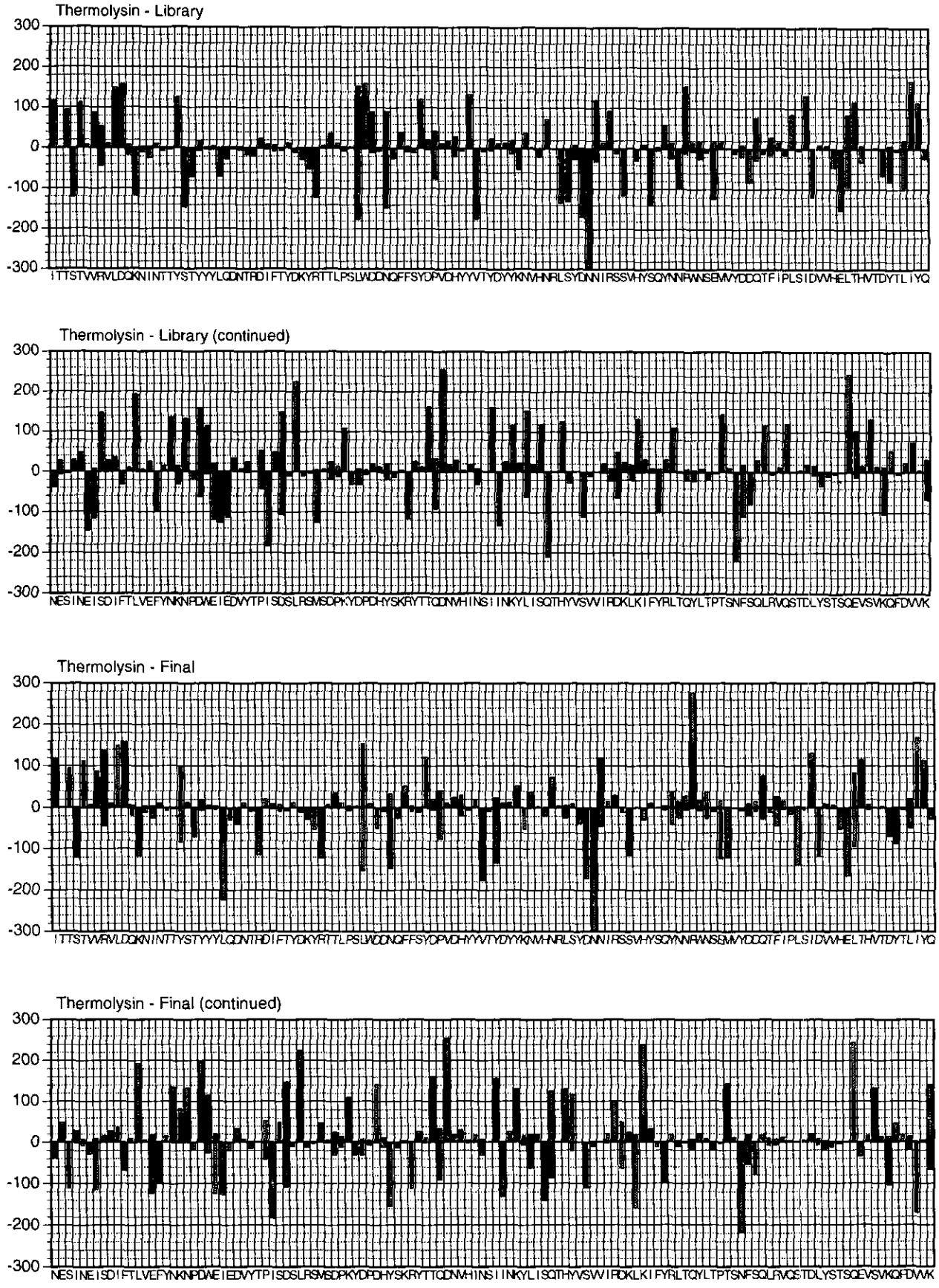


Figure 3. Results for χ_1 and χ_2 for 6 proteins calculated from backbone co-ordinates only. Results for the structure generated from the library and after the minimizations are completed as shown for each protein. The sequence and number of each residue is listed below the x-axis, and the bars represent deviations (in deg.) from the crystal structure for each side-chain. Deviations for χ_1 (filled bars) and χ_2 (hatched bars) are stacked, so that the deviation in χ_2 is given by the length of each hatched bar, rather than by the summed length of the filled and hatched bars.

Table 7
Predictions of side-chain conformations from backbone co-ordinates

	Structure 0 (library)			Structure 1			Structure *N*			Backbone-independent lib.
χ	No. cor.	No. inc.	Fraction correct	No. cor.	No. inc.	Fraction correct	No. cor.	No. inc.	Fraction correct	Fraction correct
A. *Thermolysin (N = 3)*										
1	169	83	0·67	180	72	0·71	187	65	0·74	0·52
2	131	48	0·73	131	48	0·73	130	49	0·73	0·69
1+2	149	103	0·59	156	96	0·62	161	91	0·64	0·43
B. *Bovine pancreatic trypsin inhibitor (N = 2)*										
1	37	9	0·80	41	5	0·89	39	7	0·85	0·65
2	32	9	0·78	33	8	0·80	32	9	0·78	0·71
1+2	29	17	0·63	34	12	0·74	33	13	0·72	0·52
C. *Crambin (N = 2)*										
1	32	5	0·86	34	3	0·92	34	3	0·92	0·68
2	23	4	0·85	23	4	0·85	24	3	0·89	0·81
1+2	30	7	0·81	32	5	0·86	33	4	0·89	0·59
D. *Lysozyme (N = 4)*										
1	80	25	0·76	81	24	0·77	81	24	0·77	0·64
2	57	28	0·67	59	26	0·69	56	29	0·66	0·67
1+2	63	42	0·60	67	38	0·64	64	41	0·61	0·48
E. *Rhizopuspepsin, C-terminal domain (N = 2)*										
1	82	33	0·71	89	26	0·77	94	21	0·82	0·54
2	65	11	0·86	67	9	0·88	62	14	0·82	0·83
1+2	78	37	0·68	87	28	0·76	88	27	0·77	0·50
F. *Ribonuclease A (N = 3)*										
1	79	30	0·72	78	31	0·72	86	23	0·79	0·56
2	55	20	0·73	54	21	0·72	53	22	0·71	0·71
1+2	68	41	0·62	70	39	0·64	76	33	0·70	0·42

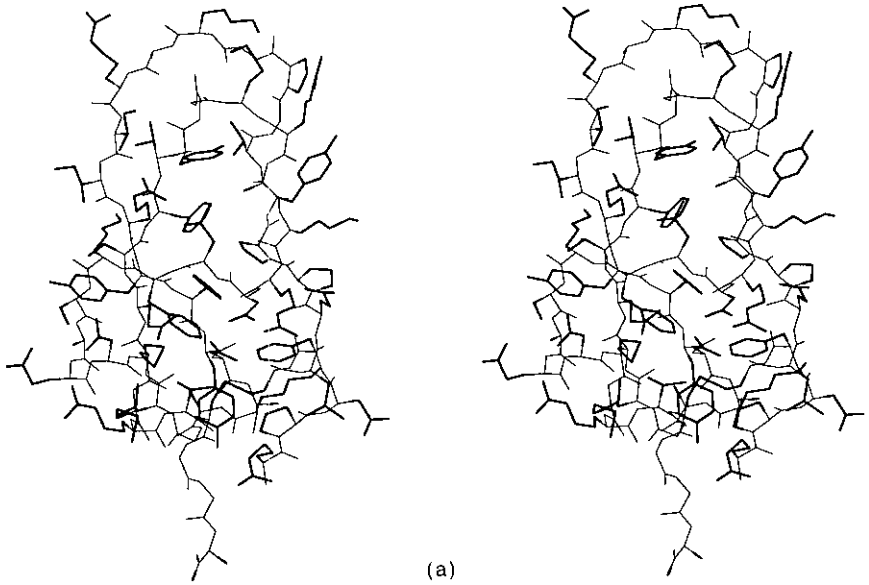
χ angle results are listed for χ_1,χ_2 and χ_{1+2}. The latter is defined as those side-chains with both χ_1 and χ_2 correct to within 40° of the experimental structure. Side-chains with only a single χ angle (e.g. Ser) are also included in χ_{1+2}. No. cor. refers to the number of residues correct (within 40°) for the given protein (for χ_1, χ_2 and χ_{1+2}), and No. inc. refers to the number incorrect. The fraction correct is equal to No. cor./(No. cor. + No. inc.), and is given for each structure named at the top of the Table. The structure numbers (0, 1, *N*) refer to the numbers in Fig. 1, with 0 being the backbone-dependent library prediction, 1 being the prediction after the first refinement, and *N* being the final structure after *N* refinement cycles. The value of *N* is given for each protein after its name. In the final column are predictions based on a backbone-independent library, according to the most prevalent conformations across the ϕ,ψ map (Table 4). With the exception of Met, these are the results that would be predicted by the rotamer library of Ponder & Richards (1987).

amino acids except for leucine χ_{1+2}. Leucine can exhibit very different χ angles and be nearly coincident in the atom positions. Lee & Subbiah (1991) note that if χ_1 is altered by 30° to 40° and χ_2 is changed by 150° to 140°. The C^δ atoms are nearly superimposable on the initial structure, while C^γ is shifted only slightly. Of the 19 leucinyl residues that are incorrectly placed in the final structures of the six proteins, inspection of the dihedral errors indicates that nine of them are likely to be misplaced because of the positional degeneracy of the two conformations. It is likely that in the X-ray structure it is not possible to distinguish one conformation from the other, so that the low prediction rate for leucine may be deceptive.

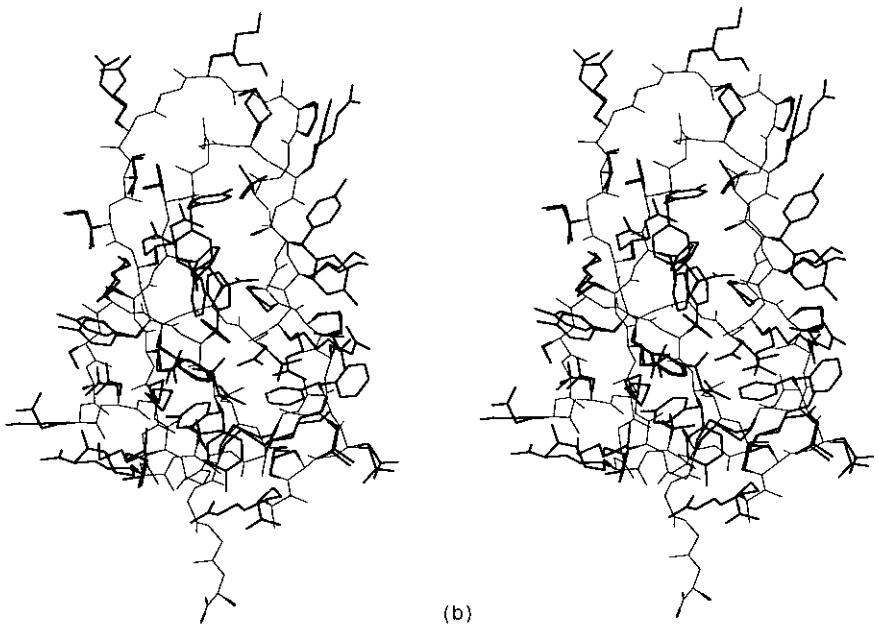
The library also performs well for aromatic amino acids, except Trp χ_2, which is greatly improved upon reorientation and minimization. Since one conformation of Trp is likely to clash with the backbone or side-chains of other residues, refinement often introduces the correct χ_2. The method does well for Cys, in part because all the Cys in the chosen proteins are involved in disulfide bridges. They minimize to correct conformations once the disulfide bond is established. Thr is predicted with much greater accuracy than Ser, because with two heavy atoms for Thr in the γ position, there is much less steric freedom in relation to the backbone. Since Ser is quite small and is able to form hydrogen bonds to the backbone, packing is often not the dominant interaction in determining its conformation.

The polar and charged amino acids are least well placed by the rotamer library, especially glutamic acid. While minimization improves most of them, only Gln reaches close to two-thirds of the values correct for χ_{1+2}. The χ_1 values are all in the 60 to 80% range in the final structures, but the χ_2 values are more variable in accuracy and therefore the χ_{1+2} results are poor. These amino acids depend on local hydrogen-bonding interactions with other side-chains and with solvent. Neither the library nor the potential energy function used in the refinement accounts for solvent effects in sufficient detail to predict their conformations well. Also, such side-chains may not have well-defined conformations, since they are often exposed and have high *B*-factors (Summers & Karplus, 1989).

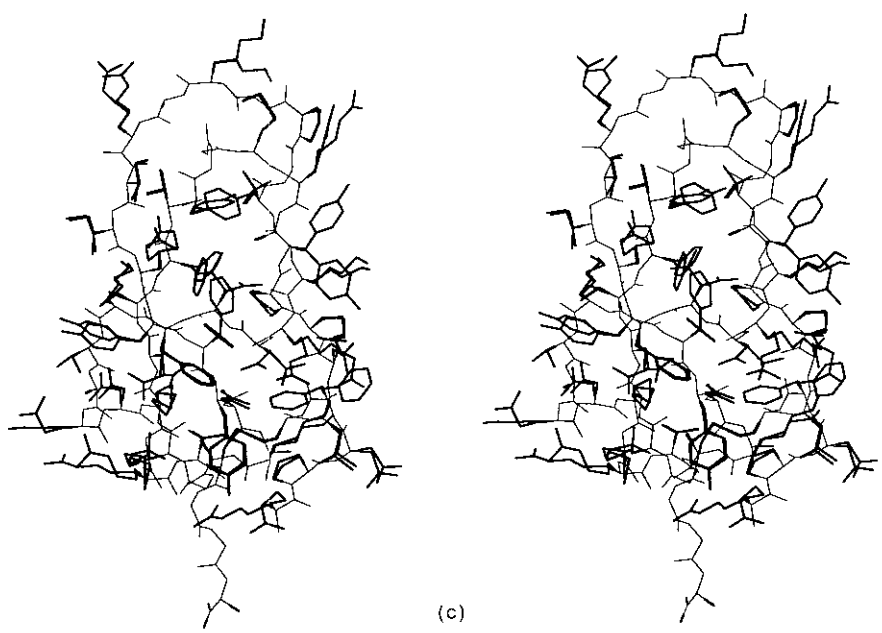
In Table 9, we list the r.m.s.d. for the side-chains obtained for the six predicted structures, and compare these with the results of Lee & Subbiah (1991). In most cases, our results compare favorably



(a)



(b)



(c)

Fig. 4.

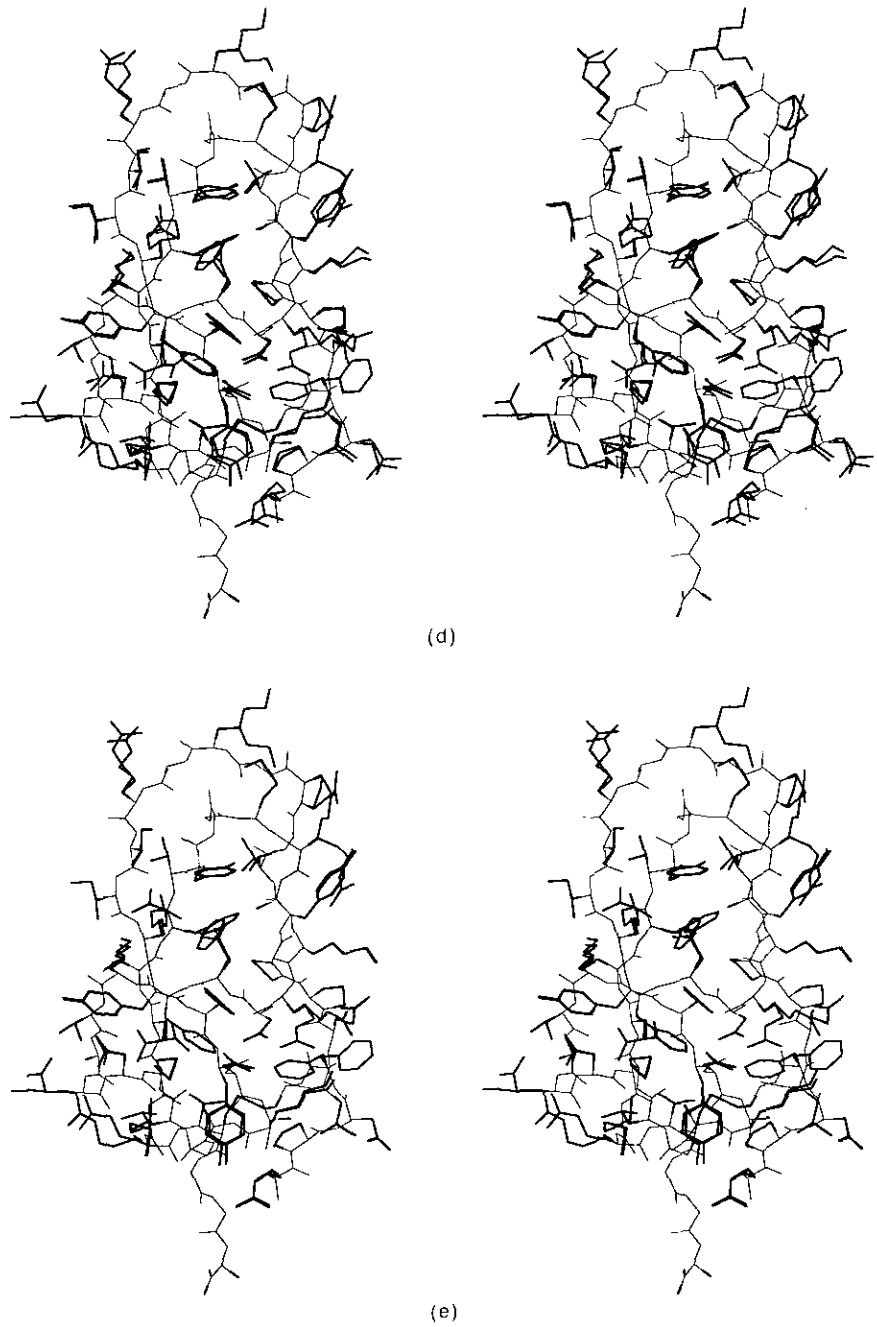
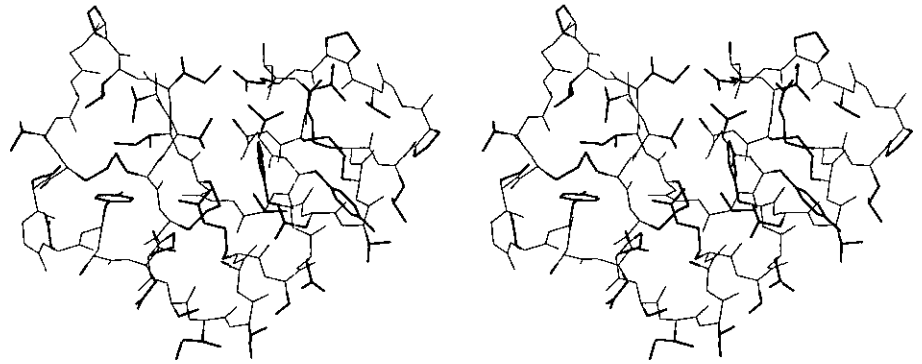


Figure 4. BPTI structures: (a) X-ray structure alone; (b) X-ray and backbone-independent library prediction; (c) X-ray and backbone-dependent library prediction; (d) X-ray and final predicted structure; (e) minimized X-ray and minimized final predicted structure (see the text).

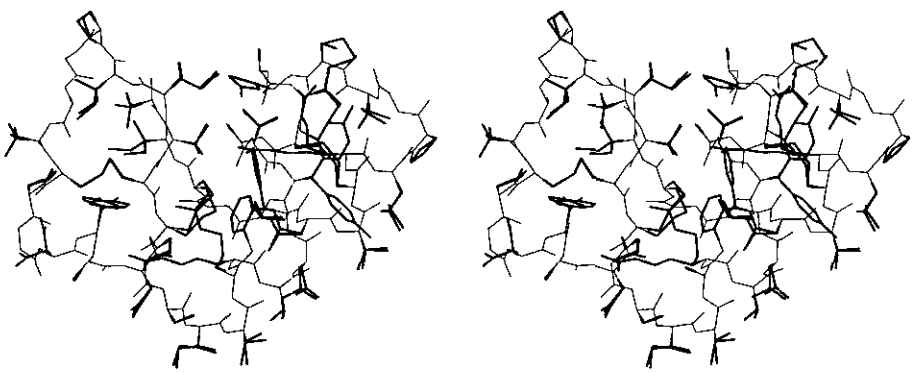
with those of Lee & Subbiah (1991). One should note the large variation in r.m.s.d. among the different kinds of side-chains. The bigger side-chains have larger r.m.s.d., even when they are well predicted, as in the case for phenylalanine. Averaging over all side-chains in a single protein gives results that depend heavily on the sequence, so we do not provide such averages here.

(c) *Method targ/temp applied to pen → rhi*

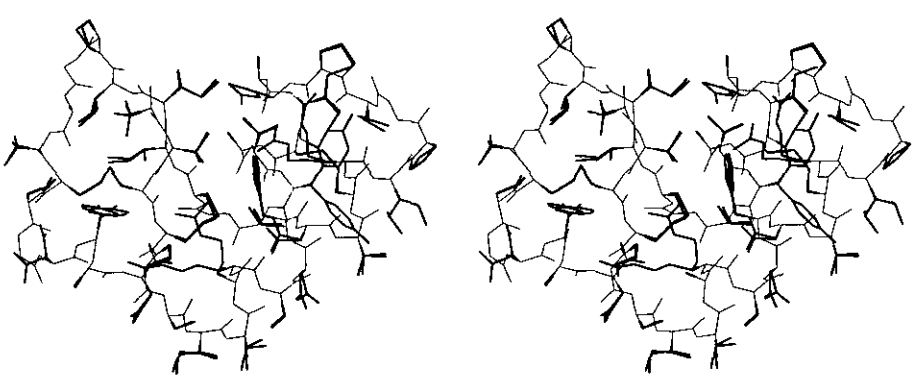
We applied the method described here to the problem studied by Summers & Karplus (1989). In that paper, information about side-chain conformations for the C-terminal domain of rhizopuspepsin was taken from the homologous protein, penicillopepsin (3 app) which has a 39% sequence identity with rhizopuspepsin. Cys, Pro and backbone coordinates were taken from the target X-ray structure (2 apr), and the other side-chains were modeled *via* their dihedral angles and rigid rotations as described in the Introduction. After the rigid rotations were completed (essentially equivalent to the final step in the present method), Summers & Karplus (1989) had predicted 86% of χ_1 and 75% of χ_2 correctly. With some additional checks and com-



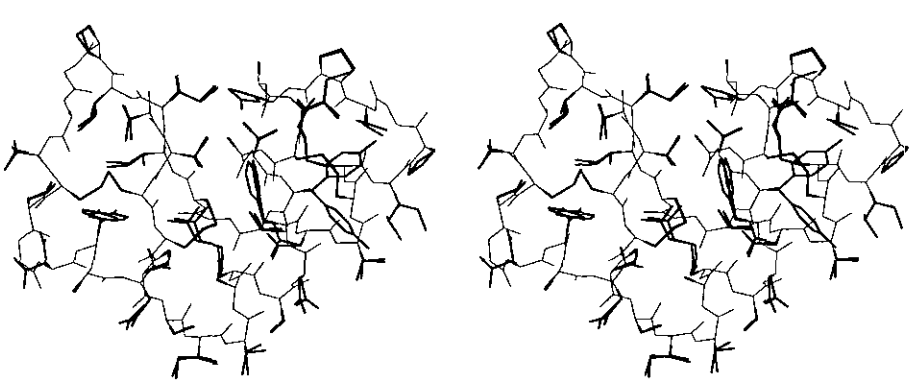
(a)



(b)

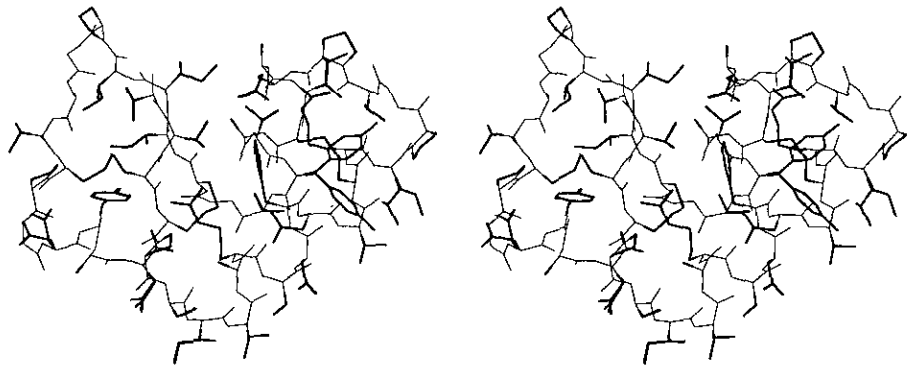


(c)



(d)

Fig. 5.



(e)

Figure 5. Crambin structures: (a) X-ray structure alone; (b) X-ray and backbone-independent library prediction; (c) X-ray and backbone-dependent library prediction; (d) X-ray and final predicted structure; (e) minimized X-ray and minimized final predicted structure (see the text).

parisons with the homologous protein, 92% and 81% accuracy was achieved for χ_1 and χ_2, respectively.

We performed the same calculation with Pro and Cys obtained from the target conformation, and built side-chains according to the homologous protein in combination with the rotamer library for the residues for which there was no information in the homolog (e.g. Gly → Asp). The final results are 88% of χ_1 and 80% of χ_2 correct, slightly worse than the 92% for χ_1 and 81% for χ_2 obtained by Summers & Karplus (1989). The method described here is simpler to apply than that of Summers & Karplus (1989), it is fully automated and does not require a homologous protein. However, the results obtained here without using the homologous protein are significantly worse for χ_1 (82%) and the same (80%) for χ_2. It is possible that some of the more complex refinement procedures used by Summers & Karplus (1989) could improve the present results.

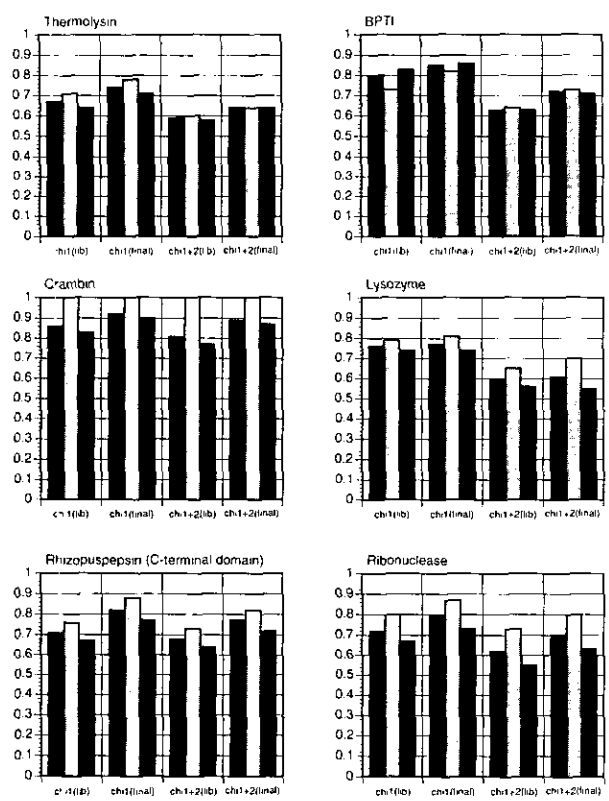


Figure 6. Fractions correct for all side-chains (black bars), buried side-chains (less than 10% exposure) (light grey bars), and surface side-chains (greater than 10% exposure) (dark grey bars) for the 6 proteins calculated from the backbones alone. Results for χ_1 alone and χ_{1+2} are shown from the library and after the minimizations are completed.

4. Discussion

An analysis has been made of the relation between the conformations of side-chains and the

Table 8

Results for six proteins for specific side-chain types

		Fractions correct					
Res.	No.	χ_1 Lib.	χ_1 Final	χ_2 Lib.	χ_2 Final	χ_{1+2} Lib.	χ_{1+2} Final
A. *Small, polar*							
Ser	63	0·57	0·65				
Thr	62	0·84	0·84				
Cys	30	0·90	0·93				
B. *Hydrophobic*							
Val	56	0·88	0·91				
Ile	44	0·84	0·86	0·86	0·84	0·73	0·73
Leu	37	0·59	0·68	0·49	0·54	0·49	0·49
Pro	29	0·86	0·79	0·83	0·76	0·83	0·76
Met	9	1·00	1·00	0·44	0·67	0·44	0·67
C. *Aromatic*							
Phe	30	0·67	0·83	0·90	0·77	0·57	0·70
Tyr	50	0·80	0·86	0·92	0·82	0·72	0·74
His	13	0·77	0·92	1·00	0·92	0·77	0·85
Trp	11	0·82	0·82	0·36	0·73	0·27	0·64
D. *Polar and charged*							
Asn	54	0·63	0·76	0·67	0·70	0·54	0·61
Asp	50	0·72	0·74	0·74	0·76	0·64	0·62
Gln	32	0·56	0·72	0·75	0·72	0·44	0·59
Glu	23	0·43	0·61	0·57	0·65	0·13	0·39
Arg	39	0·64	0·74	0·67	0·62	0·38	0·51
Lys	32	0·63	0·66	0·81	0·69	0·53	0·53

Table 9

Average root-mean-square deviation in Cartesian co-ordinates

	This work			Lee & Subbiah	
Res.	No.	r.m.s.d. X-ray: *N*	r.m.s.d. Min. X-ray: Min. *N*	No.	r.m.s.d. X-ray: Prediction
Cys	30	0·61	0·53	33	1·32
Ser	63	0·99	0·95	55	1·17
Thr	62	0·80	0·77	49	1·22
Val	56	0·63	0·60	74	0·97
Ile	44	0·91	0·88	55	0·89
Leu	37	1·51	1·45	63	1·00
Phe	30	1·88	1·72	41	1·29
Tyr	50	1·95	1·75	37	1·17
His	13	1·30	1·31	11	1·58
Trp	11	2·44	2·30	15	2·20
Pro	29	0·40	0·29	—	—
Asp	50	1·44	1·25	43	1·31
Asn	54	1·46	1·29	42	1·70
Glu	23	2·24	2·20	48	1·73
Gln	32	1·78	1·78	29	2·01
Met	9	1·09	1·03	17	1·27
Lys	32	2·16	2·14	46	2·72
Arg	39	2·88	2·86	35	3·40

Root-mean-square deviation in heavy-atom Cartesian co-ordinates is given for each amino acid type across the 6 proteins tested in this paper. For symmetric residues, r.m.s.d. for both χ_2 and $\chi_2+180°$ (or χ_3 and $\chi_3+180°$) were tested, and the lower value was used. In addition to the deviations of Structure *N* from the X-ray structure (column 3), the deviations of a CHARMM minimized Structure *N* from a CHARMM minimized X-ray structure are given in column 4. The minimization procedure used is the same as that in Summers & Karplus (1989), and consists of minimizing all of the side-chains simultaneously subject to gradually reduced harmonic constraints. For comparison, the results of Lee & Subbiah (1991) are given for the 9 proteins tested in that paper (column 6).

local backbone geometry in proteins for which high-resolution structures are available. The results show that the most probable side-chain dihedral angle values are affected by the ϕ and ψ angles of the local backbone. This relationship is of interest for protein folding and for structure prediction. Based on the results, a backbone-dependent library of side-chain rotamers has been developed from the available high-resolution structures. The portions of the library that are well populated tend to have specific side-chain conformational preferences. This result provides a basis for understanding the success of the side-chain placement studies that make use of backbone templates with similar ϕ,ψ angles (see Introduction).

The backbone-dependent rotamer library serves as the starting point for a prediction scheme of side-chain orientations from the backbone co-ordinates. After the initial placement, the side-chain positions are refined by reorientation and energy calculations to eliminate side-chain–backbone and side-chain–side-chain van der Waals repulsions. This iterative procedure, which scales approximately linearly with the size of the protein, leads to the results that 78% of χ_1, 74% of χ_2 and 69% of χ_{1+2} are correctly predicted for a set of six proteins ranging in size from crambin to thermolysin.

The results obtained here have a number of implications for studies of protein folding and structure. The library with or without side-chain minimization provides a starting point for building full protein models from crystallographic backbone co-ordinates that can be refined with the experimental structure factors. Also, the results demonstrate that model building from template protein backbones is feasible and may be sufficiently reliable to be used in drug design. The approach used here, which is an extension of the work of Summers & Karplus (1989), can serve as a starting point for such model building.

The mutual influence of backbone and side-chain conformations may have a role in protein folding since there is a reduction in the conformational space that must be searched in the actual folding process and in theoretical model studies. It was pointed out some time ago (Gelin & Karplus, 1975) that the side-chain conformations in a protein tend to correspond to minima that are selected from those that exist in the isolated dipeptide. This concept was embodied in the rotamer library, independent of the backbone conformation, that was proposed by Ponder & Richards (1987). The present results go further and indicate that the local backbone structure can play an important role in the selection process, though neighboring side-chains and tertiary contacts are also involved. The latter may be most important in the stabilization of given side-chain conformer rather than in its selection. In the limit, this implies that in the process of protein folding, the correct backbone and side-chain dihedral angles are introduced in a concerted fashion.

Both van der Waals exclusions between the backbone and side-chains and the tendency toward the g^-, t and g^+ conformations severely limit the conformation space that a side-chain can occupy, and reduce by many orders of magnitude the space that must be searched to find a structure with no repulsive overlaps of side-chains. This conclusion is supported by the recent work of Desmet *et al.* (1992).

Although the test application of the side-chain prediction scheme described here are quite successful, it should be noted that we have not examined the effect on the predictions of errors in the backbone positions. Also, there are a number of possibilities for improvements of the method. Polar and charged residues are least well predicted. For the protein interior, inclusion of hydrogen bonding and other electrostatic interactions may be useful. For surface side-chains, the prediction problems may be due in part to the fact that such side-chains have multiple conformations. However, lack of explicit solvent in the model is also a limitation, since it must affect the conformations of such hydrogen-bonding side-chains. It is possible to modify the potential energy function used for the iterative minimizations to mimic the effect of solvent for exposed residues. Wesson & Eisenberg

(1992) have recently modified the CHARMM potential to explicitly favor solvent accessibility for polar atoms (N and O) and favor solvent inaccessibility for non-polar atoms (C) by use of surface-area dependent corrections. A similar approach has been used by Schiffer *et al.* (1992) in a study of the alanine dipeptide. Another approach is to add terms of the RISM type (Pettitt & Karplus, 1985; Ramé *et al.*, 1990), where solute–solvent correlation functions are used to calculate the effect of the solvent on solute–solute interactions. Both of these solvent corrections are being incorporated into the prediction program to determine their effects on the accuracy of surface side-chain predictions. Finally, as Ponder & Richards (1987) have described, the mean positions of many rotamers do not lie exactly at 60°, 180° and −60°, and so slightly different orientations could be used in the placement. However, even without such improvements, the method proposed should be useful in a variety of applications.

This work was supported in part by a grant from the National Science Foundation and Polygen/Molecular Simulations, Inc. The calculations were performed on a Convex C220 and a Silicon Graphics SGI 340. We thank Hsiang-ai Yu for helpful discussions, and Roland Stote and Aaron Dinner for technical assistance.

Note: A copy of the full backbone-dependent rotamer library is available upon request. Write to R.L.D. or M.K or send electronic mail to dunbrack@tammy.harvard.edu.

References

Bhat, T. N., Sasisekheran, V. & Vijayan, M. (1979). An analysis of side-chain conformations in proteins. *Int. J. Pept. Protein Res.* **13**, 170–184.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.

Bruccoleri, R. E. & Karplus, M. (1987). Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers*, **26**, 137–168.

Cung, M. T., Vitoux, B. & Marraud, M. (1987). Flexibility of Pro-Pro sequences: IR and NMR experiments. *New J. Chem.* **11**, 503–510.

Desmet, J., DeMaeyer, M., Hazes, B. & Lasters, I. (1992). The dead-end elimination theorem and its use in protein side-chain positioning. *Nature (London)*, **356**, 539–542.

Gelin, B. R. & Karplus, M. (1975). Sidechain torsional potentials and motion of amino acids in proteins: bovine pancreatic trypsin inhibitor. *Proc. Nat. Acad. Sci., U.S.A.* **72**, 2002–2006.

Gelin, B. R. & Karplus, M. (1979). Side-chain torsional potentials: effect of dipeptide, protein, and solvent environment. *Biochemistry*, **18**, 1256–1268.

Holm, L. & Sander, C. (1991). Atomic structure of the actin—DNase I complex. *J. Mol. Biol.* **218**, 183–194.

James, M. N. G. & Sielecki, A. R. (1983). Structure refinement of penicillopepsin at 1·8 Å resolution. *J. Mol. Biol.* **183**, 299–361.

Janin, J., Wodak, S., Levitt, M. & Maigret, B. (1978). Conformation of amino acid side-chains in proteins. *J. Mol. Biol.* **125**, 357–386.

Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. (1990). *Nature (London)*, **347**, 37–44.

Kendrew, J. *et al.* (1970). IUPAC-IUB commission on biochemical nomenclature. Abbreviations and symbols for the description of the conformation of polypeptide chains. *Biochemistry*, **9**, 3471–3479.

Konnert, J. H. & Hendrickson, W. A. (1980). A restrained-parameter thermal-factor refinement procedure. *Acta Crystallogr. sect. A*, **36**, 344–350.

Lee, C. & Subbiah, S. (1991). Prediction of protein side-chain conformation by packing optimization. *J. Mol. Biol.* **217**, 373–388.

McGregor, M. J., Islam, S. A. & Sternberg, M. J. E. (1987). Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J. Mol. Biol.* **198**, 295–310.

Pettitt, B. M. & Karplus, M. (1985). The potential of mean force surface for the alanine dipeptide in aqueous solution: a theoretical approach. *Chem. Phys. Letters*, **121**, 194–201.

Ponder, J. W. & Richards, F. M. (1987). Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **193**, 775–791.

Ramé, G. L., Lau, W. F. & Pettitt, B. M. (1990). Flexibility of tripeptides in solution: free energy molecular mechanics. *Int. J. Pept. Protein Res.* **35**, 315–327.

Reid, L. S. & Thornton, J. M. (1989). Rebuilding flavodoxin from C_α coordinates: a test study. *Proteins: Struct. Funct. Genet.* **5**, 170–182.

Sali, A., Overington, J. P., Johnson, M. S. & Blundell, T. L. (1990). From comparisons of protein sequences and structures to protein modelling and design. *Trends Biochem. Sci.* **15**, 235–240.

Schiffer, C. A., Caldwell, J. W., Stroud, R. M. & Kollman, P. A. (1992). Inclusion of solvation free energy with molecular mechanics energy: alanyl dipeptide as a test case. *Protein Sci.* **1**, 396–400.

Shih, H. H.-L., Brady, J. & Karplus, M. (1985) Structure of proteins with single-site mutations: a minimum perturbation approach. *Proc. Nat. Acad. Sci., U.S.A.* **82**, 1697–1700.

Straatsma, T. P. & McCammon, J. A. (1992). Alchemical free energy simulation. *Annu. Rev. Phys. Chem.* **43**, 407–435.

Summers, N. L. & Karplus, M. (1989). Construction of side-chains in homology modelling. Application to the C-terminal lobe of rhizopuspepsin. *J. Mol. Biol.* **210**, 785–812.

Summers, N. L. & Karplus, M. (1990). Modeling of globular proteins: a distance-based data search procedure for the construction of insertion/deletion regions and Pro ↔ non-Pro mutations. *J. Mol. Biol.* **216**, 991–1016.

Tidor, B. & Karplus, M. (1991). Simulation analysis of the stability mutant R96H of T4 lysozyme. *Biochemistry*, **30**, 3217–3228.

Tuffery, P., Etchebest, C., Hazout, S. & Lavery, R. (1991). A new approach to the rapid determination of protein sidechain conformations. *J. Biomol. Struct. Dynam.* **8**, 1267–1289.

Wendoloski, J. J. & Salemme, F. R. (1992). PROBIT: a statistical approach to modeling proteins from partial coordinate data using substructure libraries. *J. Mol. Graph.* **10**, 124–127.

Wesson, L. & Eisenberg, D. (1992). Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.* **1**, 227–235.
Wüthrich, K. (1989). The development of nuclear magnetic resonance spectroscopy as a technique for protein structure determination. *Acc. Chem. Res.* **22**, 36–44.

Edited by B. Honig