Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Protein side-chain conformation: a systematic variation of χ_1 mean values with resolution – a consequence of multiple rotameric states?

Malcolm W. MacArthur[a]* and Janet M. Thornton[a,b]

[a]Biomolecular Structure and Modelling Unit, Department of Biochemistry and Molecular Biology, University College, London WC1E 6BT, England, and [b]Crystallography Department, Birkbeck College, Malet Street, London WC1E 7HX, England

Correspondence e-mail: mac@bsm.bioc.ucl.ac.uk

Received 14 October 1998
Accepted 11 February 1999

A systematic variation with resolution of the mean values of the *gauche*$^-$, *trans* and *gauche*$^+$ χ_1 rotamers in protein structures determined by X-ray crystallography has been observed. Further analysis revealed that these correlations differ considerably between residue types, being highly significant for some residue types (*e.g.* Ser, Thr, Leu, Lys) and absent for others (*e.g.* aromatics). For the individual residue types which exhibited the trend most strongly, these changes were accompanied by corresponding systematic variations in the percentage relative populations in the three energy wells. Examination of a uniformly sized subset of monomers showed that this effect, while attenuated, was still present, and was thus not entirely a consequence of the change in size and surface area which also correlates with resolution. An analysis of *B* values in the disfavoured high-energy barrier region between the rotameric wells showed a pronounced tendency towards larger than average values. As a plausible hypothesis, it is suggested here that these observations can be accounted for by the presence of multiple rotameric states. The averaged electron density produced by dual occupancy at low resolution giving an averaged conformation is resolved at high resolution into its individual components.

1. Introduction

In models of protein molecular structure derived by X-ray crystallography, uncertainty in the values of the Cartesian (x, y, z) coordinates is inevitably a resolution-dependent function. For the χ_1 angles of the side chains in particular, we observed a very strong linear correlation between their standard deviation from the expected *gauche*$^-$, *trans* and *gauche*$^+$ values and the resolution over the observed range 1.0–3.5 Å (Morris *et al.*, 1992). When that survey was carried out, the Protein Data Bank (Bernstein *et al.*, 1977) contained a total of only 462 coordinate sets for proteins determined by X-ray crystallography (Morris *et al.*, 1992) and there were only 15 entries of resolution better than 1.5 Å. As a consequence, the statistics for the higher resolution data were somewhat weak. In the intervening period, the number of protein structure coordinate sets in the PDB has multiplied several fold to almost 7000 at the start of this study. With the benefit of this much larger database now at our disposal, it was therefore deemed appropriate to reanalyze the statistics which had been examined in the earlier work.

Overall, the results corroborated our previous findings (MacArthur, unpublished work) and, unsurprisingly, the linear correlations between χ_1 standard deviations and resolution were extended to higher resolution. What was

© 1999 International Union of Crystallography
Printed in Denmark – all rights reserved

Table 1
Comparison of protein χ_1 mean values from regression line extrapolation with χ_1 mean values from small peptides and atomic resolution protein structures.

	Atomic resolution structures			Linear small peptides			Extrapolated from protein data†
	Mean (°)	Sdev‡ (°)	N_{ang}§	Mean (°)	Sdev‡ (°)	N_{ang}§	Mean (°)
gauche$^-$	66.1	8.0	90	65.5	7.0	89	65.6
trans	183.2	9.9	192	180.8	11.9	155	181.6
gauche$^+$	−65.1	9.6	346	−65.3	8.2	315	−65.4

† Extrapolation of regression line to 0.8 Å on the plot of mean *versus* resolution. ‡ Standard deviation of χ_1 angle about the mean. § Number of angles.

surprising, however, was a totally unexpected and very striking linear association between χ_1 mean values and resolution.

2. Data and methods

The protein structure coordinates were taken from the September 1996 release of the PDB, from which a representative data set including only one example of proteins with a greater then 95% sequence identity was drawn using the method of Orengo & Taylor (1996). Only crystal structures of resolution up to and including 3.0 Å were considered. This produced a subset of 1128 chains, which was employed without updating as the standard working data set for this and other continuing related studies. All derived data and statistics were obtained using in-house programs or commercial statistical software. In the relatively small number of cases from high-resolution structures where more than one set of coordinates is given for the atoms which define the χ_1 angle, those for the major conformer were chosen. In cases of equal occupancy, the first set of coordinates (usually, but not always, labelled 'A') were used to calculate the angle. The solvent-accessible surface areas were calculated using the Lee & Richards (1971) algorithm as implemented in the program *NACCESS* (Hubbard, 1992), using the default van der Waals radii and a probe size of 1.4 Å. The percentage relative side-chain accessibility for a given residue *X* was calculated from the ratio of the summed accessible atomic surface areas in the protein to that of the same residue type *X* in the extended Ala–*X*–Ala tripeptide, as recommended by the authors (Hubbard *et al.*, 1991). For the accessibility studies, we used the subset of 295 monomers from the above data set.

Any set of observations fewer than ten in number were excluded from the statistical analyses.

3. Results

3.1. Correlation between χ_1 and resolution

When investigating the statistics of side-chain attributes (means and standard deviations of torsion angles *etc.*), it became apparent that for all three χ_1 rotamers the mean value of the angle varied systematically with resolution in a highly correlated manner. This is illustrated in Fig. 1. This correlation with resolution is quite striking, and in the case of the *gauche*$^-$ rotamer the mean value of χ_1 differs by 10° between the 3.0 Å resolution subset and the 1.0 Å set (56 and 66°, respectively) with a correlation coefficient for the regression line of −0.948. For the *trans* rotamer and the χ_1 *gauche*$^+$ rotamer, the correlation coefficients are 0.932 and −0.842, respectively, for 20 degrees of freedom. An analysis of the individual residue type further revealed that the correlations were stronger for certain residue types and non-existent for others, with different profiles for the different residues in the three rotameric wells (see below).

3.2. Hypothesis: observed correlations are a consequence of multiple rotameric states

At first sight, these results appeared to be unaccountable for on any simple physical/chemical basis, and could be attributable only to some artifact in the experimental procedure or interpretation of the electron-density maps which was dependent on its clarity and definition. A plausible hypothesis is that this phenomenon is a consequence of the occupancy of multiple rotameric states. It has been shown from detailed studies on well refined high-resolution structures (Stec *et al.*, 1995) that up to 30% of all side chains in a protein can have multiple conformational substates, and it has been concluded that such heterogeneity in proteins is ubiquitous (Rejto & Freer, 1996). At low resolution, however, these are unresolvable in electron-density maps. The time-resolved image will tend to show the density as a single amorphous ill-defined shape at a position intermediate between the two true positions, with its centre of mass within the rotamer of longer residence time but displaced towards the minor conformer. With improving resolution, it becomes possible to resolve the two alternative conformations, and an increasing number of them progressively become visible (see Fig. 2). At very high resolution it is possible to assign values to the occupancy factors, which correspond to the relative residence times. The implication of this is that only at the very highest resolutions where all multiple occupancies have been identified and correctly assigned will we obtain a true value for each alternative rotamer.

3.3. Comparison with data from peptides and atomic resolution structures

When the regression lines in Fig. 1 are extrapolated to 0.8 Å there is good agreement between the mean values, those observed in small peptides and those of the structures determined to atomic resolution (Table 1).

3.4. Correlations and means for individual residue types

3.4.1. Serine and leucine. Table 2 gives the correlation coefficients of mean values *versus* resolution for the three

relevant rotameric wells of serine and leucine. Residue types which show a marked correlation with resolution appear to favour a particular pair of rotameric states (*gauche*$^-$ and *gauche*$^+$ for Ser and *gauche*$^+$ and *trans* for Leu). Support for the multiple-conformer hypothesis is derived from an examination of the relative population densities of the three rotameric states as they vary with resolution. One would not *a priori* expect to see any systematic association with resolution, yet surprisingly it occurs and is remarkably consistent with the trends observed in the variation of the means with resolution. The correlation coefficients are frequently in the range 0.7–0.95, depending on residue type. This is illustrated in Fig. 3 for the case of serine *gauche*$^+$ and *gauche*$^-$ where the complementarity is striking – indeed they are almost mirror images. While the proportion of the total in the two states remains almost constant at around 78% regardless of resolution, the relative *gauche*$^-$:*gauche*$^+$ ratio varies from 48:30 at 1.0 Å

Table 2
Means *versus* resolution.

Residue type	Number of angles	Correlation coefficient	Confidence level† (%)	Δ mean (3.0–1.0 Å)
(*a*) χ_1 *gauche*$^-$ rotamers				
Ser	7330	−0.900	99.9	−11.2
Thr	6620	−0.849	99.9	−10.2
(*b*) χ_1 *trans* rotamers				
Arg	4084	+0.756	99.9	+5.8
Gln	3200	+0.818	99.9	+8.6
Glu	4976	+0.913	99.9	+9.8
Leu	7409	+0.903	99.9	+12.0
Lys	5459	+0.766	99.9	+7.4
(*c*) χ_1 *gauche*$^+$ rotamers				
Arg	6070	−0.814	99.9	−4.8
Gln	5316	−0.810	99.9	−6.8
Leu	13366	−0.795	99.9	−7.4
Lys	8227	−0.946	99.9	−4.6
Ser	6253	+0.860	99.9	+9.6

† Confidence level at which correlation coefficient is significant.

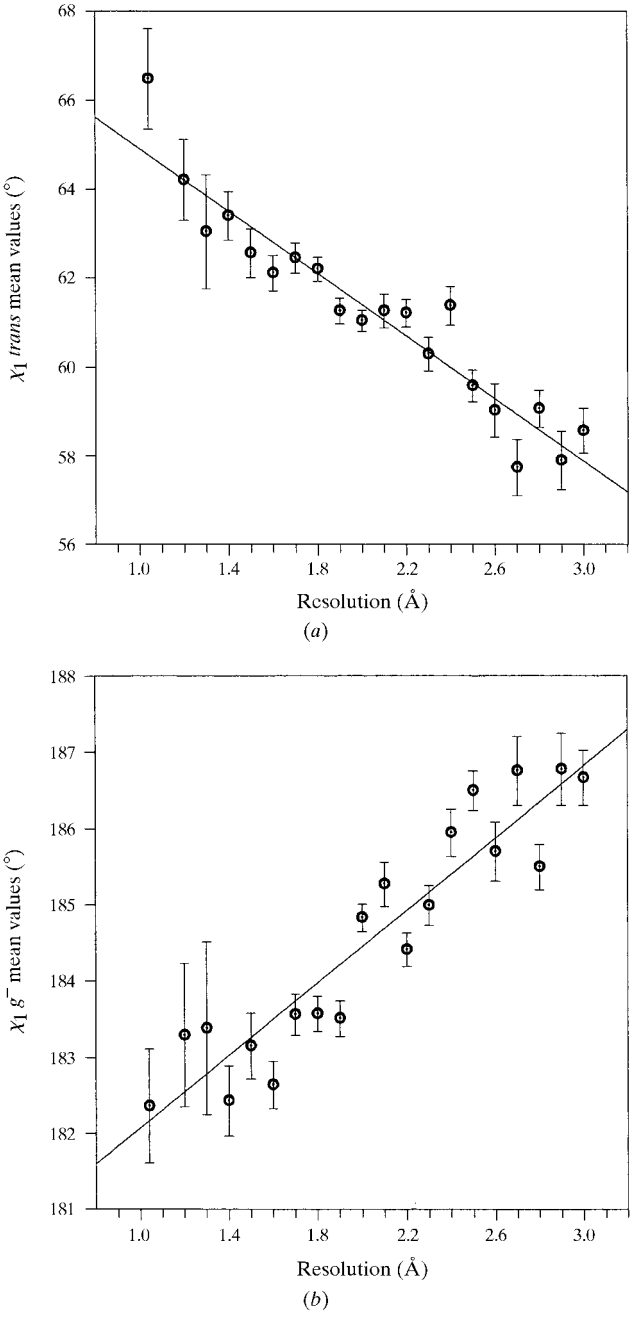

Figure 1
Variation of χ_1 mean values with resolution for all residue types excluding proline in a representative data set of 1128 chains. (*a*) Regression line for χ_1 *gauche*$^-$ rotamer. Correlation coefficient is −0.948 for 20 degrees of freedom. (*b*) Regression line for χ_1 *trans* rotamer. Correlation coefficient is +0.932. Error bars represent the standard errors, *i.e.* the standard deviation/$n^{1/2}$.

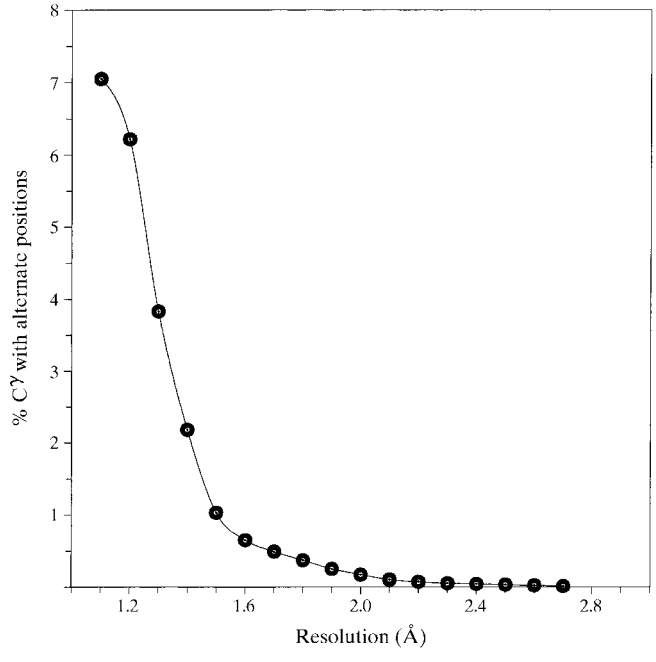

Figure 2
Plot showing the percentage increase in the number of multiple occupancies for side-chain C^γ atoms with improvement in resolution. Data were obtained from the 'atom' and 'alternate' tables of 7260 protein structures in the *IDITIS* relational database (Oxford Molecular Ltd) V3.11. All examples of alternate C^γ positions were included, irrespective of the magnitude of the difference in position.

resolution to 36:42 at 3.0 Å resolution, and the correlation coefficients are −0.938 and +0.935, respectively. The corresponding plots for leucine *trans* and *gauche*$^+$ are illustrated in Fig. 4. In addition, we examined the asymmetry of the distributions for the χ_1 population densities in the three rotameric wells. For an ideal trimodal dihedral angle distribution, we would expect the frequency histogram to show approximately normal distributions which varied with resolution only in their change of variance caused by random error. Figs. 5 and 6 show the full χ_1 distributions from 0–360° for Leu and for Ser at high (<2 Å) and low (2–3 Å) resolution. Consider leucine, where the main rotamers occupied are *trans* and *gauche*$^+$. At low resolution, there is a high concentration of values around 240° in what is an unstable high-energy conformation. This region is much less populated at the higher resolution. In the histogram for the low-resolution set, the two peaks have pronounced shoulders which face one another across the 240° divide, which are attenuated in the high-resolution set. There

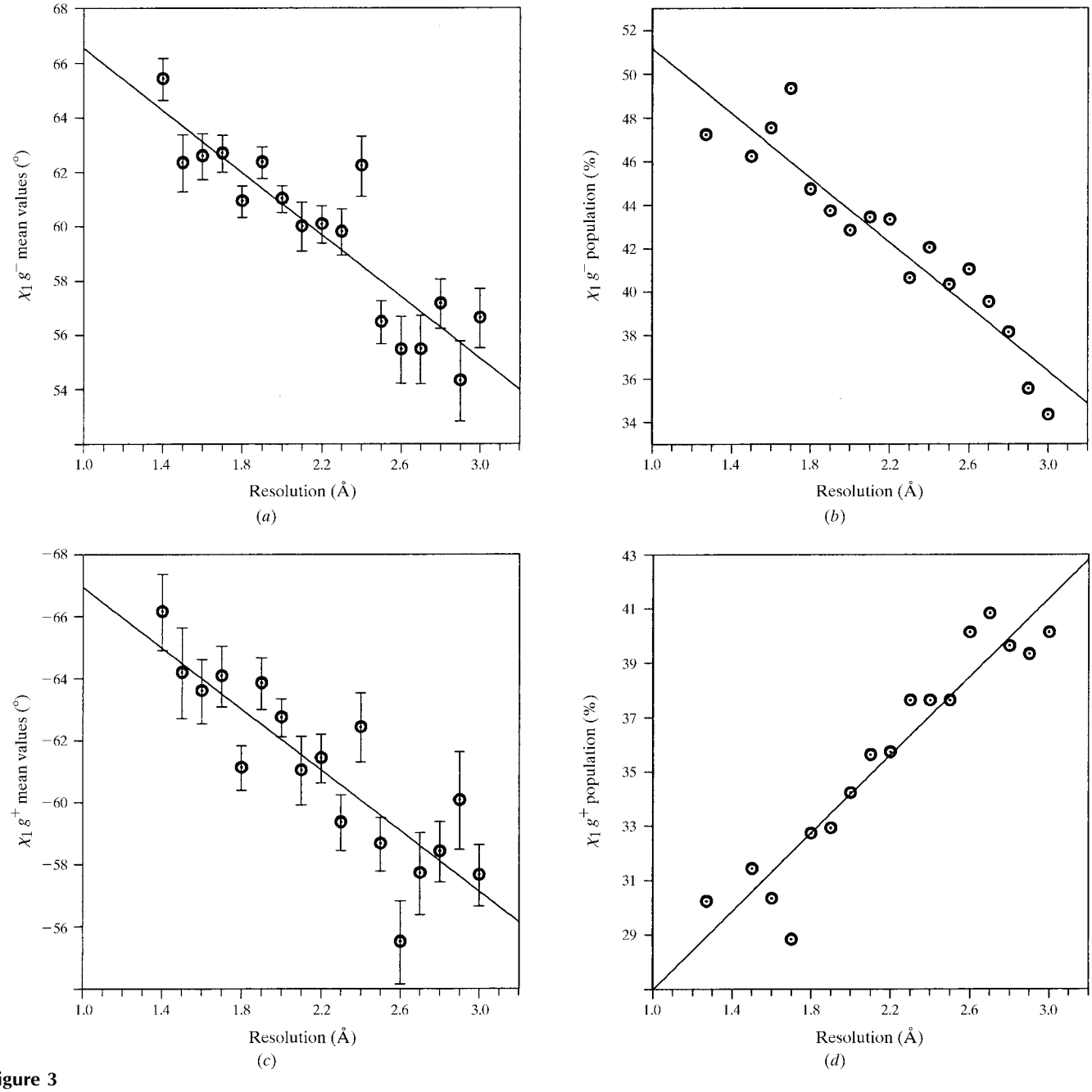

Figure 3
Variation with resolution of χ_1 mean values and percentage relative rotamer populations for serine residues in the data set of 1128 chains. (*a*) Mean *versus* resolution for χ_1 *gauche*$^-$ rotamer. Correlation coefficient is −0.900 for 17 degrees of freedom. (*b*) Percentage population of χ_1 *gauche*$^-$ rotamer as it varies with resolution. Correlation coefficient is −0.937. (*c*) Mean *versus* resolution for χ_1 *gauche*$^+$ rotamer. Correlation coefficient is +0.860. (*d*) Percentage population of χ_1 *gauche*$^+$ rotamer as it varies with resolution. Correlation coefficient is +0.948.

is a pronounced systematic change in the symmetry of the distributions with resolution, with a strong complementarity between the two favoured alternative rotameric states. This is shown for Leu *trans* and Leu *gauche*$^+$ in Figs. 7(*a*) and 7(*b*), where the correlation coefficients are −0.913 and +0.842, respectively. Similar trends were shown by serine. According to the hypothesis, the bulk of these observations result from ambiguity in the electron density arising from double occupancy. At the higher resolution, more of these become resolvable and are correctly assigned in the right proportion to the two wells, with the major occupancy being assigned to the more energetically favoured one, which in the case of leucine is the *gauche*$^+$. Thus, with increasing resolution the resultant effect is an apparent population shift from *trans* to *gauche*$^+$. This effect is consistent with variation in the population densities which correlates strongly with resolution. It will be seen from Fig. 4 that the population of the *trans* rotamer decreases by 7% and that of the *gauche*$^+$ increases by 12% on

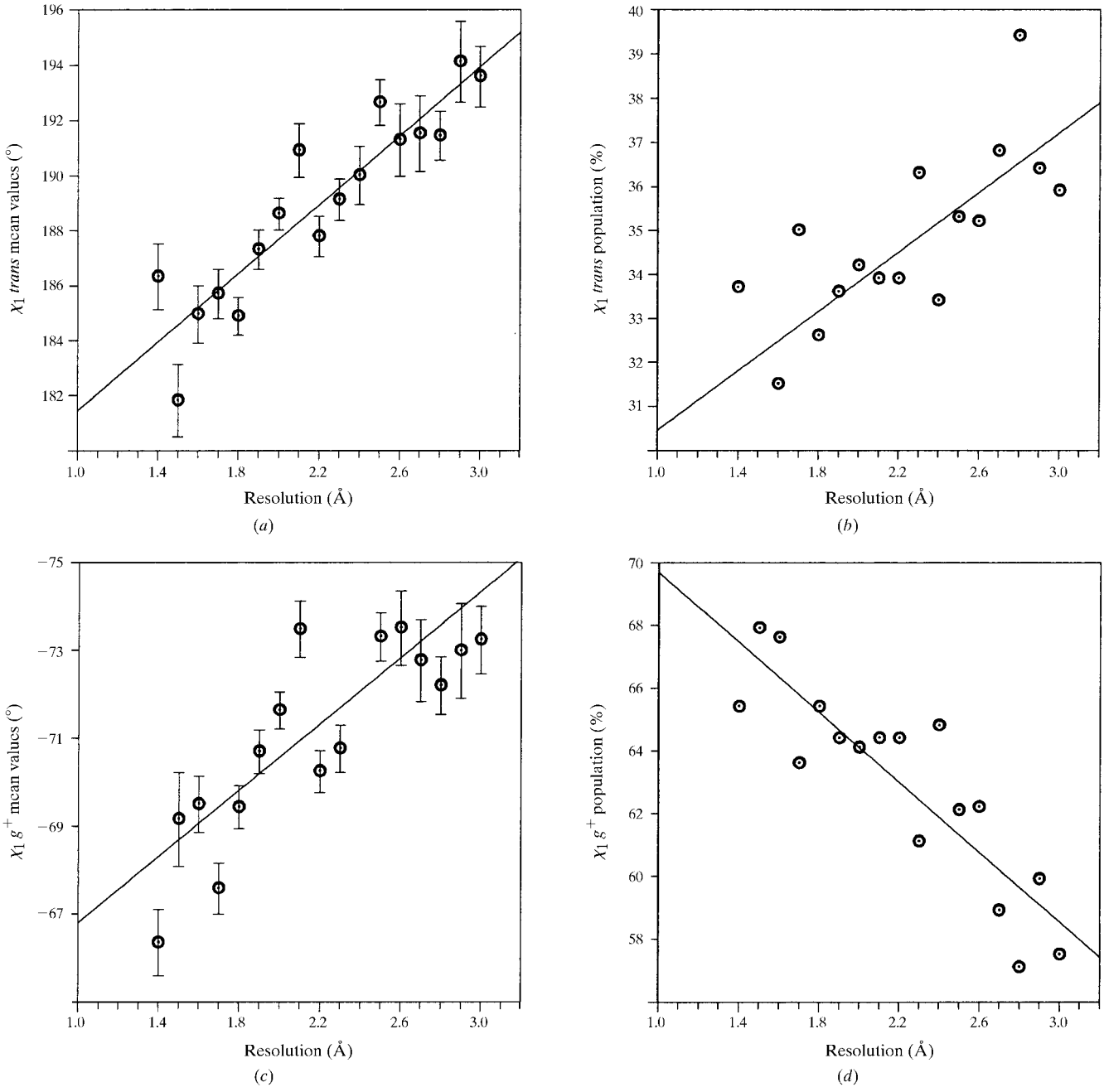

Figure 4
Variation with resolution of χ_1 mean values and percentage relative rotamer populations for leucine residues in the data set of 1128 chains. (*a*) Mean *versus* resolution for χ_1 *trans* rotamers. Correlation coefficient is +0.903 for 16 degrees of freedom. (*b*) Percentage population of χ_1 *trans* rotamer as it varies with resolution. Correlation coefficient is +0.751. (*c*) Mean *versus* resolution for χ_1 *gauche*$^+$ rotamer. Correlation coefficient is −0.795. (*d*) Percentage population of χ_1 *gauche*$^+$ rotamer as it varies with resolution. Correlation coefficient is −0.868.

going from low to high resolution (see also Fig. 5). The removal from the shoulders of the wrongly assigned single rotamer to two correctly assigned ones in their true positions within the energy wells also explains the change in the value of the means. In Fig. 5, these each apparently shift away from the central barrier (4.1° for the smaller *trans* peak and 2.5° for the larger *gauche*$^+$). An examination of Fig. 6 showing the corresponding serine distributions leads to the same conclusion. In this case, the alternate states are *gauche*$^+$ and *gauche*$^-$. The same is expected to apply to the other residue types which are prone to adopt alternate positions. The disorder represented by these alternate positions is likely to be both dynamic and static. Serine and leucine were chosen to illustrate the phenomenon both because these demonstrate the effect to a pronounced degree and because of their clearly different hydrophilic/hydrophobic properties. In serine residues, the position of the hydroxyl group is often not well defined in electron-density maps, especially at low resolution, and this can be a source of error in its interpretation. Leucine can have nearly coincident side-chain atom positions and yet exhibit very different χ angles. Lee & Subbiah (1991) have noted that if χ_1 is altered by 30–40° and χ_2 is changed by 140–150°, the C^δ atoms are nearly superimposable on the initial structure, while C^γ is shifted only slightly. Leucine side chains exist predominantly in only two rotameric states ($\chi_1 = -60°$, $\chi_2 = 180°$; $\chi_1 = 180°$, $\chi_2 = +60°$) and solid-state NMR results indicate that there is rapid interconversion between them (Batchelder *et al.*, 1982; Colnago *et al.*, 1987), though the motion may be more complex than described by this simple two-site jump model (Yang *et al.*, 1998). Dunbrack & Karplus (1993) observed that of 19 leucine residues whose side-chain conformations they had been trying to predict in six proteins, nine of them could have had wrong dihedral angles in the published structures owing to errors in map interpretation arising from positional degeneracy. It is likely that in the X-ray structures it is not possible to distinguish one conformation from the other, particularly when low resolution results in ambiguity. The results for the other side chains can be interpreted in a similar manner.

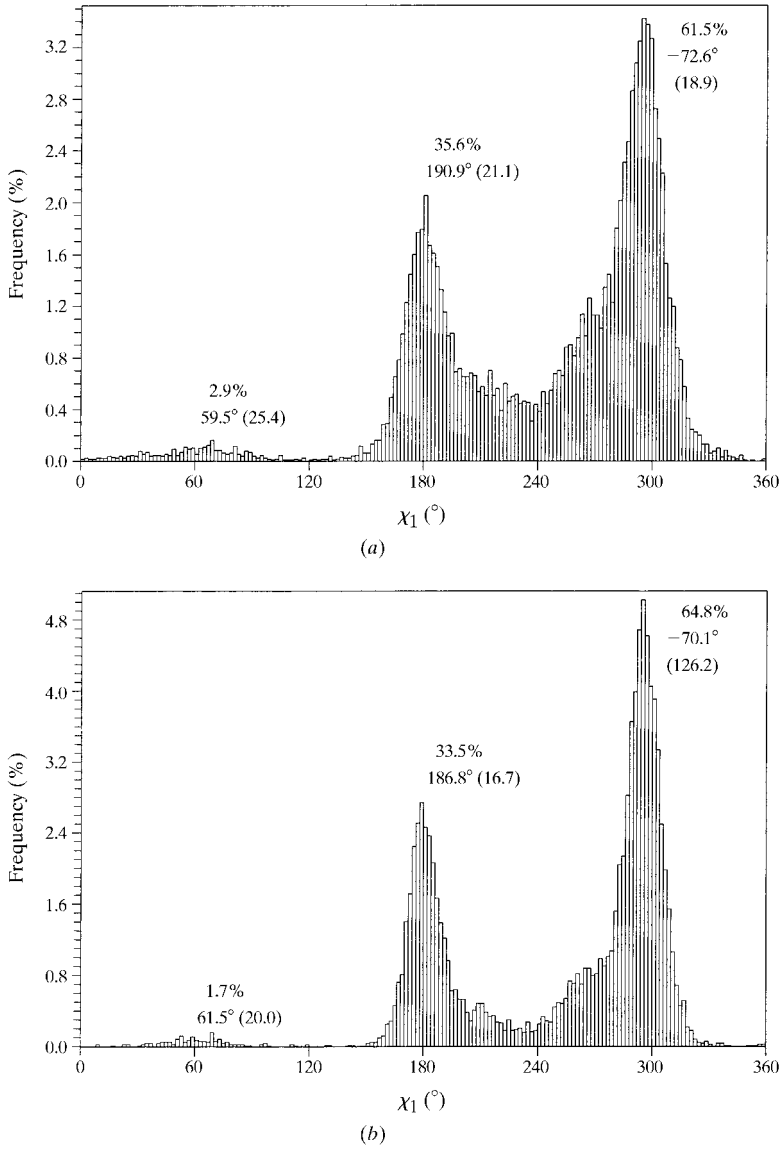

Figure 5
Histograms showing the χ_1 distribution of leucine in the data set of 1128 chains. (*a*) Low resolution, greater than 2.0 Å (609 chains). (*b*) High resolution, less than or equal to 2.0 Å (519 chains). The mean value of χ_1 for each rotamer, standard deviation (in brackets) and percentage relative population are indicated.

3.4.2. Branched C^β. Like serine, the threonine χ_1 *gauche*$^-$ position is the dominant rotamer and also shows a strong correlation of mean value and relative population density with resolution (r = −0.849 and −0.869, respectively), though it is less clear from the scatter of the mean values for the *gauche*+ and *trans* what the most favoured alternative conformers might be. Like the threonine side chain, valine and isoleucine, although not polar, are also branched at the C^β atom, and their χ_1 distributions are in a way rather similar, in that the *trans* isomeric well is much less heavily populated than the *gauche*$^-$ and *gauche*$^+$ (11.4 and 12.2% for Val and Ile, respectively). In this case, however, it is the *trans* rotamer (Val χ_1 values have to be adjusted by +120°) which shows the strongest correlation of percentage occupation with resolution (r = +0.890 for Val and +0.676 for Ile) with an 11% decrease (from 17 to 6%) in population of the former as one goes from low to high resolution. For both residues, the redistribution is to *gauche*$^+$ and *gauche*$^-$ in about equal measure, though this is less clear cut for isoleucine, where the scatter about the regression line is very high. This would suggest that there may be some difficulty in apportioning the correct rotamers to the side chains of these two residues when interpreting the electron-density map at lower resolutions. The uncertainty attending χ_1 *trans* assignment is reflected in a much higher standard deviation within the *trans* well (g^-:*t*:g^+ = 15:18:11°

for Ile; g^-:t:g^+ = 18:26:13° for Val for the 2.0 Å resolution subset), suggesting that the *trans* rotamer may be most vulnerable to wrong assignment.

3.4.3. Other polar. The long charged and polar side chains (Glu, Gln, Lys, Arg, Met) all exhibit similar tendencies in the redistribution of their rotamer populations with improvement in resolution. This is manifested as large increases in the χ_1 *gauche*$^+$ populations at the expense of the other two rotamers. Of the five, the effect is most strikingly evident in the case of arginine, where the *gauche*$^+$ share of the population increases from 46% at 3.0 Å resolution to around 64% at 1.0 Å. This is mirrored in a corresponding reduction of the *gauche*$^-$ and *trans* populations of 11 and 7%, respectively. In addition, the correlation coefficients are convincingly high, except for the means *versus* resolution of the lightly populated (11.2%) χ_1 *gauche*$^-$ rotamer. The generally surface location of these long unbranched and mobile side chains can often produce a tenuous electron density which is open to more than one interpretation.

3.4.4. Asp and Asn. The aspartate and asparagine side chains also tend to be water accessible and have alternative hydrogen-bonding possibilities either with water molecules or neighbouring backbone groups. The data for aspartate do suggest conformer switching to some extent, with a population shift of about 6% from *trans* to *gauche*$^-$ on moving from 3.0 to 1.0 Å resolution (correlation coefficients +0.764 and −0.718, respectively; see Table 3). Furthermore, these trends are reflected in correlated changes of 6–7° in mean values (correlation coefficients +0.722 and −0.724; Table 2). The orientation of the amide group of asparagine is always ambiguous in electron-density maps. The N and O atoms commonly cannot be distinguished by X-ray crystallography, and assignment of these two atoms is usually judged on the basis of hydrogen bonding. McDonald & Thornton (1995) have found that 15% of Asn side chains would be more favourably oriented for hydrogen bonding if the nitrogen and oxygen designations were reversed (see also Hooft *et al.*, 1996). Correlations of means and relative populations with resolution are found to be rather weak for all three rotamers, presumably because the uncertainty persists even to high resolution.

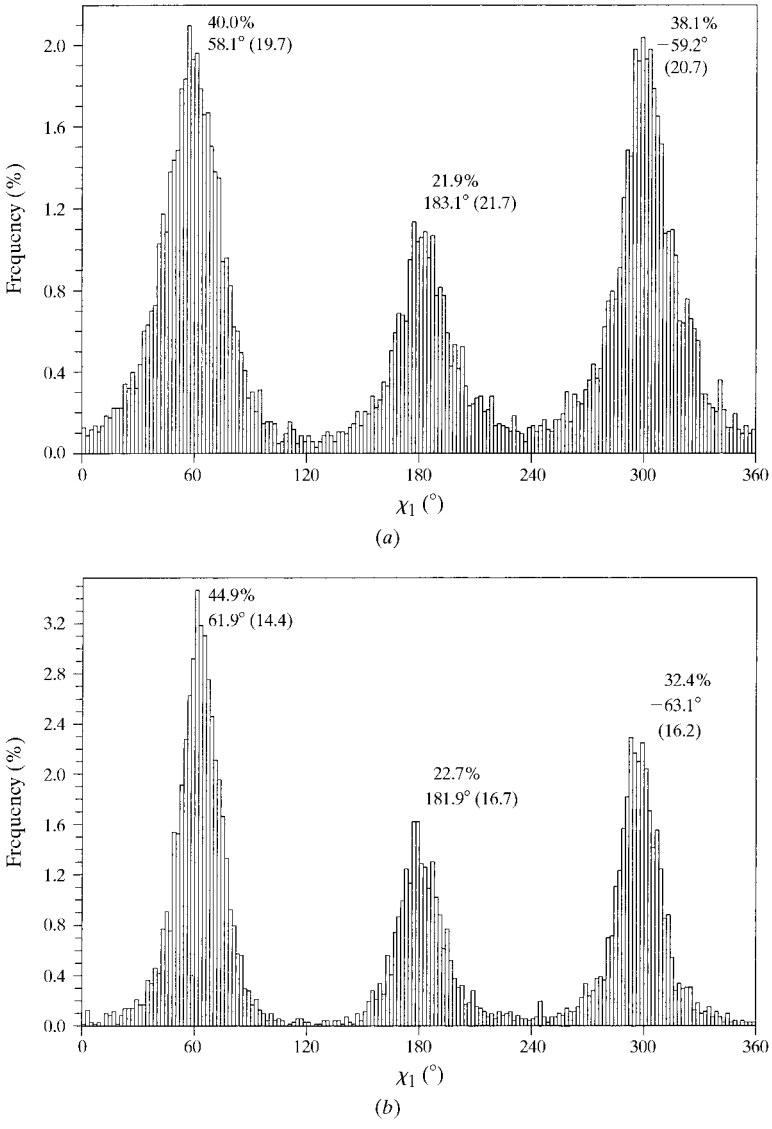

Figure 6
Histograms showing the χ_1 distribution of serine in the data set of 1128 chains. (*a*) Low resolution, greater than 2.0 Å. (*b*) High resolution, less than or equal to 2.0 Å. The mean value of χ_1 for each rotamer, standard deviation (in brackets) and percentage relative population are indicated.

3.4.5. Aromatics. The aromatic residues (Phe, Tyr, Trp, His) by contrast show little or no correlation with resolution of either the mean values or the relative percentage rotamer population densities. Although other experimental evidence such as NMR (Wüthrich, 1986) and time-resolved fluorescent decay of tryptophan (Dahms *et al.*, 1995) have indicated some conformational heterogeneity and ring flipping of the aromatics, these residues have been assumed to enhance protein stability because they can close pack in the protein core with far less reduction in side-chain entropy. In electron-density maps from X-ray crystallography, the rings are generally clear and well defined, and so the tracing is usually unambiguous even at the lower end of the resolution range

3.4.6. Cysteine. Cysteine side chains are nearly always well buried and often participate as partners in disulfide linkages. Although NMR ensembles can show a plethora of conformations for the C^α to C^α geometry, it is unclear to what extent this is a true picture of the dynamics. A lack of NOE distance restraints for the two central S atoms must render at least some of the assignments uncertain. The X-ray data show only weak correlations of means against resolution, but the shifts between populations for all three rotamers are large. Considering the high visibility of the S atoms in the electron-density map, this would indicate a degree of both dynamic and static disorder. However, because of the small size of the subsets, the cysteine statistics are uncertain.

Table 3
Rotamer populations (%) *versus* resolution.

Residue type	Number of angles	Correlation coefficient	Confidence level† (%)	Δ mean (3.0–1.0 Å)
(*a*) χ_1 *gauche*$^-$ rotamers				
Arg	1280	+0.821	99.9	+10.6
Gln	915	+0.825	99.9	+6.6
Glu	1741	+0.732	99.9	+4.4
Leu	523	+0.783	99.9	+4.3
Lys	1364	+0.812	99.9	+5.6
Ser	7330	−0.937	99.9	−14.8
Thr	6620	−0.869	99.9	−15.8
Val	2073	+0.890	99.9	+11.0
(*b*) χ_1 *trans* rotamers				
Asp	4759	+0.763	99.9	+6.2
Leu	7409	+0.751	99.9	+6.5
Met	1692	+0.769	99.9	+15.2
Thr	1523	+0.727	99.9	+6.0
(*c*) χ_1 *gauche*$^+$ rotamers				
Arg	6070	−0.837	99.9	−17.4
Gln	5316	−0.811	99.9	−14.8
Leu	13366	−0.868	99.9	−10.8
Lys	8227	−0.791	99.9	−11.4
Met	3072	−0.802	99.9	−17.0
Ser	6253	+0.948	99.9	+14.4
Thr	7782	+0.851	99.9	+9.8

† Confidence level at which correlation coefficient is significant.

It is clear from the foregoing analysis that each residue type has its own special characteristic profile of rotamer distribution and mean values for each rotameric well. Table 4 shows how these attributes vary for the 17 side chains. The differences are striking. It must be realised, however, that these target values are modulated by secondary structure (see Swindells *et al.*, 1995) and surface exposure.

3.5. Influence of solvent accessibility

When Picket & Sternberg (1993) investigated whether rotamer preferences differed between buried and surface side chains, they found evidence to suggest that Asn, Asp and Glu showed differences which were significant. We therefore conducted an analysis on a 295-monomer subset drawn from the starting data set of 1128, in order to investigate possible variation of mean values and relative well populations with protein size and degree of exposure, independent of resolution. The use of a monomer subset eliminates the uncertainty in the accessibilities calculated for subunit interfaces, though not for crystal-packing contacts. However, Carugo & Argos (1997) have suggested that crystal-packing contact atoms share the protein surface flexibility and contact surfaces are generally quite small.

We had observed in the course of this work that a strong correlation exists between chain size and resolution (r = +0.886) with average chain length ranging from about 60 residues at 1.0 Å to 300 residues at 3.0 Å. It is therefore possible that the variation in χ_1 mean values and relative rotameric populations may not solely be a consequence of increasing uncertainty at lower resolution, but could partly reflect the increasing size at lower resolutions, which changes the average surface-to-volume ratio. The data set of 295 monomers was therefore carefully chosen such that the average size of each subset within successive resolution ranges was approximately 200 residues, with individual monomers restricted in size to the range 100–350 residues. The earlier analysis was repeated using this monomer data set in order to investigate whether the relationships were still valid. For all residues combined, the χ_1 means were still strongly correlated with resolution in the uniformly sized set of monomers. Not

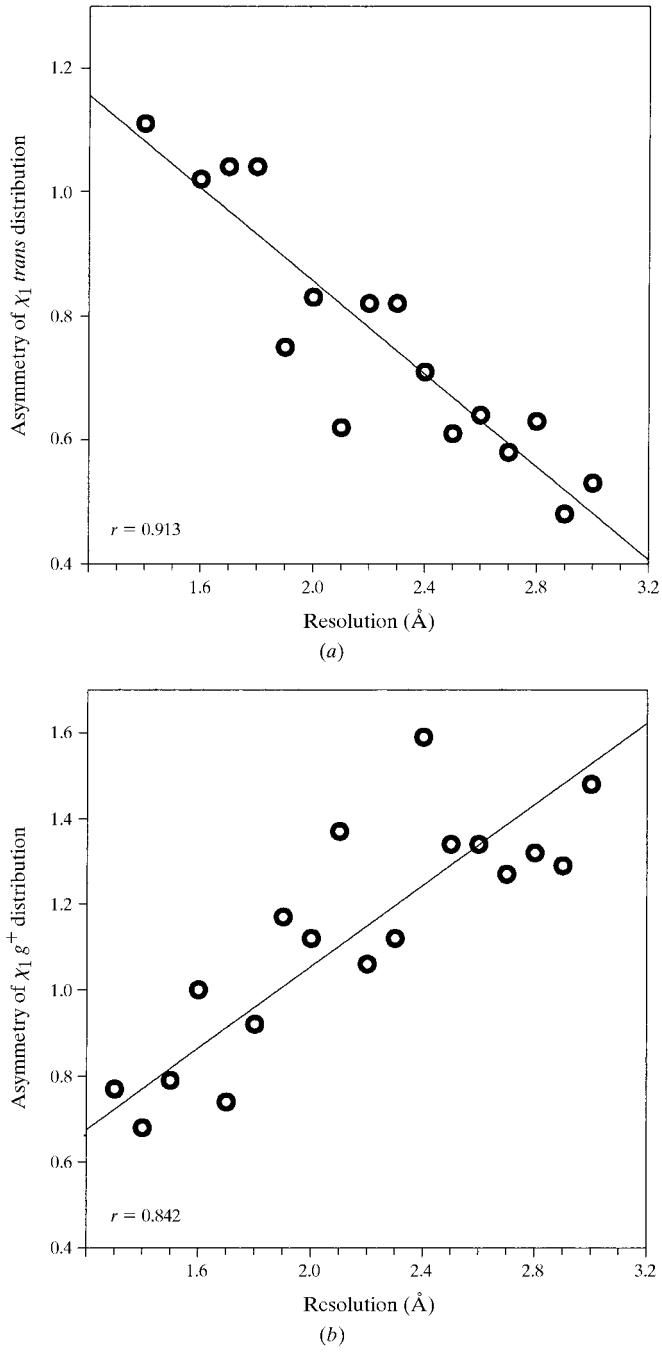

Figure 7
The χ_1 distribution asymmetry within energy wells and its variation with resolution in leucine. (*a*) χ_1 *trans*, (*b*) χ_1 *gauche*$^+$ from the data set of 1128 chains.

Table 4
Idealized mean and percentage rotamer population values at 1.0 Å.

Numbers in brackets represent the 90% confidence intervals on the predicted values of means and percentage populations at 1.0 Å.

Residue type	χ_1 *gauche*$^-$		χ_1 *trans*		χ_1 *gauche*$^+$	
	Mean (°)	% Popn	Mean (°)	% Popn	Mean (°)	% Popn
Arg	65.9 (3.5)	5.5 (2.0)	183.6 (1.5)	32.8 (1.8)	−65.4 (1.0)	61.7 (2.8)
Asn	66.1 (2.8)	15.3 (1.4)	188.6 (1.7)	29.8 (2.0)	−68.8 (0.7)	54.9 (2.1)
Asp	64.1 (1.8)	20.8 (1.5)	186.5 (1.4)	27.8 (1.5)	−69.5 (0.9)	51.4 (1.7)
Cys	63.5 (4.4)	11.4 (4.6)	181.4 (2.7)	24.5 (4.4)	−63.7 (1.6)	64.1 (7.1)
Gln	67.4 (4.3)	5.9 (1.3)	183.1 (1.8)	28.9 (3.0)	−64.1 (1.5)	65.1 (3.1)
Glu	62.7 (2.4)	9.2 (1.2)	183.5 (1.3)	31.2 (2.4)	−66.4 (1.1)	59.6 (2.8)
His	62.9 (2.4)	11.4 (2.1)	186.6 (1.6)	30.9 (3.1)	−64.4 (1.3)	57.7 (3.5)
Ile	60.8 (2.0)	16.8 (2.4)	191.6 (2.2)	8.3 (2.0)	−62.9 (0.8)	74.9 (2.8)
Leu	61.5 (5.1)	0.1 (1.0)	181.8 (1.7)	30.7 (1.7)	−66.9 (1.7)	69.2 (1.8)
Lys	66.6 (2.0)	5.8 (1.2)	184.0 (1.8)	32.9 (2.7)	−66.1 (0.5)	61.3 (2.6)
Met	62.7 (6.5)	7.3 (2.2)	184.4 (2.5)	23.8 (3.7)	−67.3 (1.1)	68.9 (3.7)
Phe	61.9 (1.7)	12.8 (1.9)	181.6 (1.0)	35.2 (2.0)	−66.8 (1.1)	52.0 (2.0)
Ser	66.3 (1.6)	51.2 (1.6)	179.9 (1.4)	21.9 (1.9)	−66.8 (1.7)	26.9 (1.4)
Thr	64.6 (1.9)	50.9 (2.6)	188.1 (3.6)	6.2 (1.7)	−60.2 (1.0)	42.9 (1.8)
Trp	60.5 (2.4)	15.5 (3.1)	182.0 (2.5)	33.3 (4.5)	−66.8 (1.0)	51.2 (5.0)
Tyr	62.8 (1.9)	12.5 (1.7)	180.0 (1.2)	35.3 (2.7)	−65.7 (1.1)	52.2 (2.6)
Val	67.5 (2.7)	4.7 (1.6)	174.5 (0.8)	72.2 (2.4)	−60.9 (2.1)	23.1 (1.5)
Val′†	59.1 (2.1)	23.1 (1.5)	187.5 (2.7)	4.7 (1.6)	−65.5 (0.8)	72.2 (2.4)
All	64.9 (0.6)	17.5 (1.2)	182.1 (0.5)	30.4 (0.7)	−65.6 (0.3)	52.1 (1.1)

† Val′: a rotation of +120° has been applied in order to conform to IUPAC convention for Ile.

unexpectedly, the correlation coefficients were weaker, at −0.914, +0.838 and −0.759 for the *gauche*$^-$, *trans* and *gauche*$^+$, respectively, but were still highly significant.

The variation of χ_1 mean values and their rotamer populations with degree of accessibility to solvent was then investigated. For all residues, combined strong correlations were observed between the variation of χ_1 mean values and accessibility, but only for the *gauche*$^-$ and *trans* conformers (r = −0.831 and +0.931, respectively). The *gauche*$^-$ and *trans* conformers also show distinct and complementary redistributions of their relative populations with changing accessibility (correlation coefficients +0.755 and −0.811, respectively). It is of interest to note that the means tend towards their 1.0 Å resolution canonical values as we move deeper into the core. This is accompanied by a progressive improvement in the statistics as reflected in their standard error. As seen earlier in the case of variation with resolution, certain residue types (although not showing identical behaviour) are prominent in having significant correlations of their χ_1 means and rotameric populations with degree of exposure (data not shown). In general, however, the statistics tend to be weaker.

3.6. *B* values

The topic of *B* values inevitably arises in any discussion of side-chain disorder, whether static or dynamic. The present study raises the question of whether potential but unresolved multiple occupancy might be associated with larger than average *B* values. We investigated this possibility by examining the side-chain *B* values for leucine χ_1 angles between 180 and 300° (see Fig. 5). Since *B* values are subject to systematic errors arising from the weights and the nature of the refinement restraints, it is necessary to normalize them when dealing with a data set of different structures:

$$\mathbf{B}_{\text{normalized}} = \mathbf{B} - \mu\mathbf{B}/\sigma\mathbf{B},$$

where $\mu\mathbf{B}$ and $\sigma\mathbf{B}$ are the mean and standard deviation, respectively, of the *B*-value distribution for a given coordinate set and **B** is the value reported in the PDB file.

Normalized *B* values for individual side-chain atoms as well as for the side-chain averages were calculated for all leucine residues from the data set of 1128 chains. Fig. 8 shows quite clearly that larger than average *B* values are associated with conformers which deviate from ideal rotameric values. It will be noted that particularly high values correspond to the 'shoulder' regions in the Fig. 5 distribution histogram.

3.7. Comparison of specific residues from low- and high-resolution structures

A proper test of our hypothesis would be to compare the χ_1 values for individual side chains from the same structure where coordinates are available both at low resolution and at atomic resolution, with the alternative occupancies clearly defined in the latter. Some examples can be found in support of the proposition. Intestinal fatty acid binding protein has been solved successively to 2.0, 1.96, 1.74, 1.5 and 1.19 Å resolution, the last entry consisting of two superposed structures with two sets of coordinates (PDB entries for these have accession codes 2ifb, 1ifb, 1icn, 1icm and 1ifc). Glu107 in the 2.0 Å structure with a χ_1 value of 96.5° is shown in the 1.19 Å entry to occupy (50:50) the *gauche*$^-$ and *trans* rotameric wells with χ_1 values of 75.2 and −176.9°, respectively. Similarly, for Thr118 the χ_1 of 90.6° in the 2.0 Å structure is again split (50:50) into *trans* and *gauche*$^-$ in the atomic resolution model, with values of χ_1 equal to −179.5 and +69.9°. Asn35 and Asn45 also show splitting in the 1.19 Å structure into *gauche*$^+$ and *trans* (50:50), but with χ_1 values which deviate considerably from the ideal. The latter residue shows great variability across the set of structures, being variously placed in all three wells. This may simply reflect the difficulty of defining the Asn side chain with precision, and these two Asn residues show changes in their χ_1 values which are not consistent with what would be expected from our hypothesis. Such mixed signals were oberved in all cases examined. This is perhaps not surprising. Adenylosuccinate synthetase has coordinate sets for nine chains solved to 2.0–2.5 Å resolution, but complexed with different nucleotides at different temperatures and pH

(1gim, 1hon, 1hop, 1ade). Across the set of structures there is great conformational heterogeneity throughout, which is seemingly random and is reminiscent of an NMR ensemble. This is a common feature of many structures which are available as multiple entries at moderate to low resolution (even the much-studied bovine pancreatic trypsin inhibitor). One is therefore unlikely to find sufficient consistency in the accuracy of low-resolution structures to make a genuine comparison with the very limited number of the more reliable atomic resolution structures which provide alternative rotamer entries for the same side chain. Also, a much larger data set of atomic resolution structures will be needed in order to provide convincing evidence for our hypothesis using the method of direct comparison.

4. Conclusions

It has been suggested that observation of alternate rotamers is beyond the detection limits of conventional X-ray crystallographic techniques, except at the very highest resolution (Smith *et al.*, 1986). Protein function is inherently linked to structural plasticity (Frauenfelder *et al.*, 1988) and our study reinforces the warning of Desmet *et al.* (1997) that the conformational flexibility of a receptor protein is one of the major determinants of correct ligand docking and should not be overlooked. Some of the newer NMR techniques being developed are likely to provide important information on side-chain dynamics (Akke *et al.*, 1998), and in X-ray crystallography, the use of multi-conformer and time-averaging refinement may be worth exploring (Burling & Brünger, 1994). With the increase in ultrahigh-resolution structures becoming available from the use of synchrotron radiation coupled to cryogenic techniques (Dauter *et al.*, 1995), it should be possible at a future date to conduct a more accurate analysis of mobility and disorder phenomena. This will provide the data for additional dynamic studies of side chains, which are needed in order to develop a more comprehensive picture of the role of mobility in recognition and an understanding of the interplay between structure, dynamics and binding. This study has found a significant and unexpected correlation between χ_1 mean values and resolution. This effect varies between the amino acids, being more pronounced for small flexible side chains. All the data support, but cannot prove, the hypothesis that this observation reflects local conformational flexibility and static disorder, which at low resolution is interpreted as a single distorted conformer. A further correlation, showing that side-chain conformation is modulated by accessibility, with surface residues being more distorted than buried groups, can now be interpreted by the higher likelihood of alternate conformers on the surface of the protein. This study should increase awareness of the significant probability of alternate conformers for surface residues (even when they are not directly visible in low-resolution structures) which can have important implications in molecular recognition and biological function.

The Validation Project of which this research forms part is funded by the BIOTECH program of DGXII of the Commission of the European Union, contract BIO2-CT92-0524 entitled 'Integrated Procedures for Recording and Validating results of 3D Structural Studies of Biological Macromolecules'.

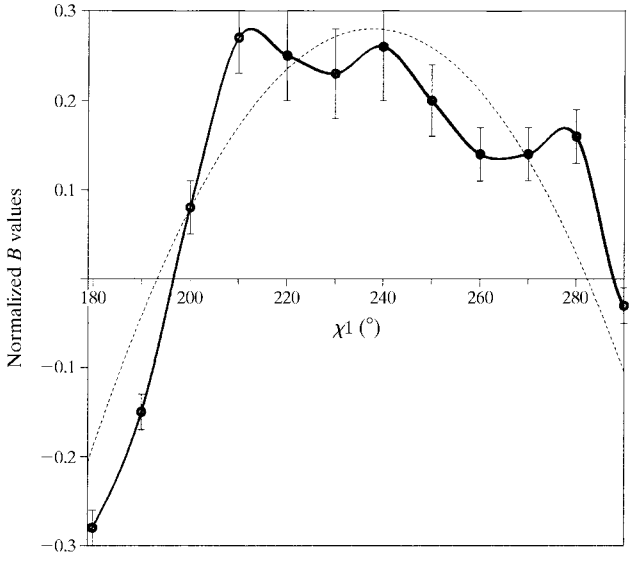

Figure 8
Variation in normalized B values for leucine side chains in the range χ_1 180–300° from the data set of 1128 chains. The heavy continuous line is for normalized B values, which were averaged within bins of 10° intervals, and the dotted line shows the best fit quadratic for the 11252 individual residues.

References

Akke, M., Liu, J., Cavanagh, J., Erickson, H. P. & Palmer, A. G. (1998). *Nature Struct. Biol.* **5**, 55–59.
Batchelder, L. S., Sullivan, C. E., Jelinski, L. W. & Torchia, D. A. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 386–389.
Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **122**, 535–542.
Burling, F. T. & Brünger, A. T. (1994). *Isr. J. Chem.* **34**, 165–175.
Carugo, O. & Argos, P. (1997). *Protein Eng.* **10**, 777–787.
Colnago, L. A., Valentine, K. G. & Opella, S. J. (1987). *Biochemistry*, **26**, 847–854.
Dahms, T. E. S., Willis, K. J. & Szabo, A. G. (1995). *J. Am. Chem. Soc.* **117**, 2321–2326.
Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1995). *Curr. Opin. Struct. Biol.* **5**, 784–790.
Desmet, J., Wilson, I. A., Joniau, M., De Maeyer, M. & Lasters, I. (1997). *FASEB J.* **11**, 164–172.
Dunbrack, R. L. & Karplus, M. (1993). *J. Mol. Biol.* **230**, 543–574.
Frauenfelder, H., Parak, F. & Young, R. D. (1988). *Annu. Rev. Biophys. Biophys. Chem.* **17**, 451–479.
Hooft, R. W. W., Sander, C. & Vriend, G. (1996). *Proteins*, **26**, 363–376.
Hubbard, S. J. (1992). PhD thesis. University of London.
Hubbard, S. J., Campbell, S. F. & Thornton, J. M. (1991). *J. Mol. Biol.* **220**, 507–530.
Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.
Lee, C. & Subbiah, S. (1991). *J. Mol. Biol.* **217**, 373–388.

McDonald, I. K. & Thornton, J. M. (1995). *Protein Eng.* **8**, 217–224.
Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M (1992). *Proteins*, **12**, 345–364.
Orengo, C. A. & Taylor, W. R. (1996). *Methods Enzymol.* **266**, 617–634.
Picket, S. D. & Sternberg, M. J. E. (1993). *J. Mol. Biol.* **231**, 825–839.
Rejto, P. A. & Freer, S. T. (1996). *Prog. Biophys. Mol. Biol.* **66**, 167–196.
Smith, J. L., Hendrickson, W. A., Honzatko, R. B. & Sheriff, S. (1986). *Biochemistry*, **25**, 5018–5027.
Stec, B., Zhou, R. S. & Teeter, M. M. (1995). *Acta Cryst.* D**51**, 663–681.
Swindells, M. B., MacArthur, M. W. & Thornton, J. M. (1995). *Nature Struct. Biol.* **2**, 596–603.
Wüthrich, K. (1986). *NMR of Proteins.* New York: Wiley.
Yang, D. W., Mittermaier, A., Mok, Y. K. & Kay, L. E. (1998). *J. Mol. Biol.* **276**, 939–954.