PROTEINS: Structure, Function, and Genetics 12:345–364 (1992)

Stereochemical Quality of Protein Structure Coordinates

Anne Louise Morris,[1] Malcolm W. MacArthur,[1,2] E. Gail Hutchinson,[1] and Janet M. Thornton[1]
[1]Biomolecular Structure and Modelling Unit, Department of Biochemistry and Molecular Biology, University College, London WC1E 6BT, England and [2]Crystallography Department, Birkbeck College, London WC1E 7HX, England

ABSTRACT **Methods have been developed to assess the stereochemical quality of any protein structure both globally and locally using various criteria. Several parameters can be derived from the coordinates of a given structure. Global parameters include the distribution of ϕ,ψ and χ_1 torsion angles, and hydrogen bond energies. There are clear correlations between these parameters and resolution; as the resolution improves, the distribution of the parameters becomes more clustered. These features show a broad distribution about ideal values derived from high-resolution structures. Some structures have tightly clustered distributions even at relatively low resolutions, while others show abnormal scatter though the data go to high resolution. Additional indicators of local irregularity include proline ϕ angles, peptide bond planarities, disulfide bond lengths, and their χ_3 torsion angles. These stereochemical parameters have been used to generate measures of stereochemical quality which provide a simple guide as to the reliability of a structure, in addition to the most important measures, resolution and *R*-factor. The parameters used in this evaluation are not novel, and are easily calculated from structure coordinates. A program suite is currently being developed which will quickly check a given structure, highlighting unusual stereochemistry and possible errors.**

Key words: protein structure, crystallography, errors, ϕ,ψ distribution, χ_1 angles, stereochemical parameters

INTRODUCTION

There are now 462 protein coordinate sets derived by X-ray crystallography in the Brookhaven Protein Structure Databank.[1] To date the resolution of a structure has been widely used as a guide to its accuracy. For example, in considering detailed residue interactions, the criterion that structures must be solved to a nominal resolution of 2.0 Å or better is often used. Figure 1a shows the resolutions for the structures in the current release of the databank. The clustering at 2 Å illustrates the somewhat arbitrary quality of this parameter. In addition, use of the data immediately highlights some coordinate sets with unacceptable close contacts or forbidden dihedral angles. Consequently it has become necessary to develop alternative measures of reliability or tests which can be used to check structures automatically. One limitation, however, is that the "accuracy" within a structure is variable. The core is usually well determined, with good electron density. In contrast, regions of weak density, often in the loops where the structure is more flexible, are much more difficult to interpret, and are often a best guess rather than a definitive conformation. There are therefore two important problems to be considered:

1. Given only the coordinates, are there automated methods which can be used to assess the overall stereochemical quality of the structure? Such methods may prove useful for the identification of incorrectly interpreted structures, either during cycles of refinement, or when refinement reaches a premature standstill. Although most PDB files do contain some author comments about crystallographic refinement, these are usually brief, qualitative, unable to be machine-read, and provide no quantitative measure of the likely reliability of the submitted structure.
2. Each atom in a structure can be assigned a temperature factor or *B* value. This is a measure of the extent to which the atom is disordered within the structure, arising from static or dynamic disorder within the crystal. Almost one quarter of the structures in the Brookhaven databank have no *B* values assigned to them. In these circumstances, can the coordinates alone provide any guidelines to indicate poorly determined regions of a protein structure?

The ultimate assessment of a correct interpretation of the electron density must be made by calculating the agreement between the model and the

Received April 25, 1991; revision accepted August 5, 1991.
Address reprint requests to Janet Thornton, Biomolecular Structure and Modelling Unit, Department of Biochemistry and Molecular Biology, University College, London, Gower Street, London WC1E 6BT, England.

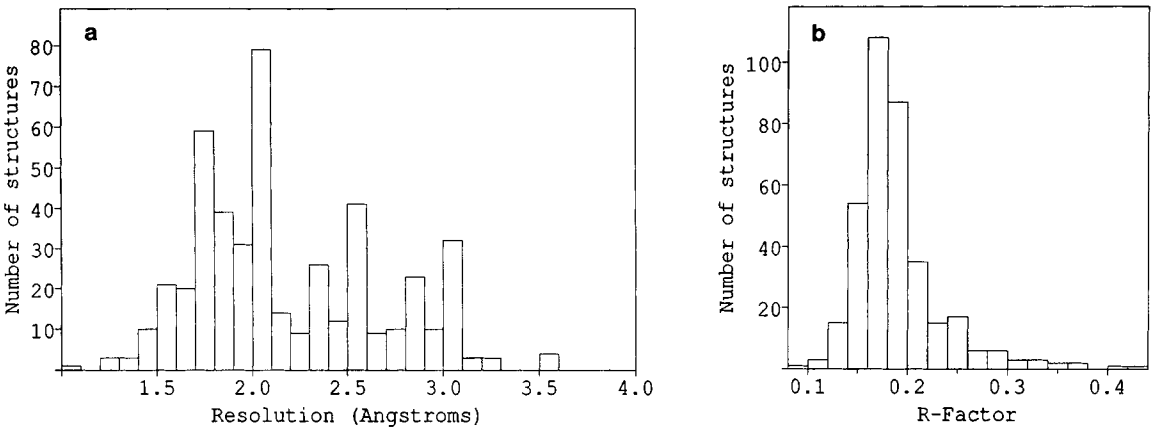


Fig. 1. (**a**) Distribution of resolutions for the structures in the Brookhaven Protein Databank.[1] The number of entries at certain resolutions is inflated by the inclusion of some structures which are essentially identical except for minor local modifications. Some examples of these are the mutant T4 lysozymes[33] at 1.7 Å and the drug/rhinovirus complexes at 3.0 Å resolution.[34] In any statistical treatment, account must be taken of the uneven distribution and the presence of many such near-identical entries. (**b**) Distribution of *R*-factors in the Brookhaven Data Bank. The figure represents the 360 known values from the 463 crystal structure entries in the October 1990 release.

density from which it was derived. This is usually measured by a global *R*-factor, calculated from the observed and model-based structure factors.[2] Only 360 structures in the databank have *R*-factors quoted; their distribution is shown in Figure 1b. Generally, structures refined to $R \leq 0.20$ are regarded as reliable, but the use of a global figure can sometimes be misleading, since it is possible to artificially reduce the *R*-factor by selective low resolution cutoff, or by the omission of weak reflections. Another possible measure is to calculate, for each residue, a local *R*-factor from the observed and calculated electron density in real space.[3] This clearly gives a better guide to local accuracy.

An alternative approach is to test the stereochemistry of a model for agreement with accepted rules. Bad stereochemistry at either the local or the global level may be a reflection of poor density, caused either by high disorder in the structure, or by limitations of data collection, processing, or refinement. The converse, good stereochemistry, is not necessarily indicative of a well-determined structure, since the stereochemistry may have been tightly constrained during refinement. In order to explore this approach, certain stereochemical parameters which can be derived from the coordinates of a structure were examined. They are listed in Table I. In each case, the study was carried out to determine whether there is a correlation between that parameter and the *R*-factor or resolution of the structure, and also between the parameters themselves.

DATA

All coordinate data were taken from the October 1990 release of the Brookhaven Protein Structure Databank.[1] Stereochemical parameters derived from the coordinates were extracted from our own relational database by simple queries.[4,5] In some cases, as indicated in the figure legends, the structural data were calculated directly by the use of in-house programs.

MINOR LABELING ERRORS

During the development of our protein structure database, programs were written which can handle any Brookhaven file and identify minor errors—such as atom labeling. Amino acid labeling conventions are clearly defined by the IUPAC–IUB Commission,[6] but these are not always followed in Brookhaven files. In particular, labeling of arginine, aspartic and glutamic acids, and the aromatics, phenylalanine and tyrosine, is often incorrect. For arginine the labeling error is significant since the guanidine group is planar and it is always possible to distinguish NH1 from NH2 (see Fig. 2a). This inconsistency, found in 13% of arginines in the database, becomes apparent when arginines are superposed. For the other side chains the labeling depends on identifying which atom defines the smallest value of the relevant angle. For example in tyrosine, CD1 is defined as the atom which gives the lowest value of χ_2. Consequently for these residues the labeling actually changes with the conformation of the side chain, and can therefore be considered rather arbitrary. Almost 50% of Phe, Tyr, Asp, and Glu residues are incorrectly labeled. For the sake of consistency, atom labeling could be checked before deposition with Brookhaven. All the atoms have been relabeled consistently in our database.

In labeling the atoms of threonine and isoleucine the chirality of the side chain must be correct (see Fig. 2b). There are 9,780 threonines in the data-

TABLE I. Indicators of Local Errors

Electron density fitting	Contacts
Local *R*-factor	van der Waals interactions
B-values	Hydrogen bonds
Basic geometry	Dihedral angles
Bond lengths and angles	ϕ,ψ (glycine, proline, other amino acids)
Planarity	ω (*cis* and *trans* peptides)
Rings	χ_1 *trans*, g^+, g^-
Peptide group	χ_3 disulfide
Arginine guanidium group	
Chirality	
L-Amino acids	
Right-handed α-helices and twist in β-strands	
Preferred packing angles between strands/helices	
Right-handed $\beta\alpha\beta$, $\beta x\beta$ units	
Type I′, II′ β-turns in short loop β-hairpins	
Handedness of motifs, e.g., 4α-bundle	

bank, and 20 (0.2%) of these had incorrect chirality. All the isoleucines had the correct chirality but two were incorrectly labeled. In the case of threonine, correct labeling interchanges the carbon and oxygen and so has an effect on local hydrogen bonding.

BOND LENGTHS AND ANGLES

In a good structure internal bond lengths and angles should conform to known stereochemistry. During refinement, regularization of such parameters is carried out in order to maximize the agreement with known geometry. A consideration of disulfide geometry illustrates some of the problems encountered in the databank. The distribution of S–S bond lengths is shown in Figure 3. The average sulfur separation is 2.04 ± 0.16 Å, but the databank includes some disulfides with very unusual S–S separations suggesting that the detailed geometry and packing in the vicinity of these disulfides may be wrong. Some refinement procedures do not link the two sides of an S–S bridge, or are unable to do so in the same manner as for linked atoms in the separate chains. However, inspection of individual protein data suggests that many S–S separations have been constrained to 2.0 Å during refinement, showing that this parameter is often not a good guide for accuracy. A study of all bond lengths and angles is currently in progress.

DIHEDRAL ANGLES

Most dihedral angles are not tightly restrained during refinement and so provide a better indication of local accuracy than the above. They may also be able to give an estimate of the global quality. The dihedral angles used for this purpose are ω, ϕ, ψ, χ_1, and the χ_3 formed by the disulfide bridge.

ω Angle

The ω angle values were extracted from the database and average values and standard deviations calculated for protein structures determined at a given resolution. The distribution of ω angles is shown in Figure 4a. Although the vast majority cluster within ±30° of the perfect *cis* or *trans* conformers, there are a number of rogue values approaching 90°. Some structures show very low standard deviations of ω angles reflecting the restraints applied during refinement; this is illustrated in Figure 4b. Consequently, there is no correlation between standard deviation and the resolution of the protein structure.

Of prolines 5.8% exhibit *cis* peptide bonds *before* the proline ring compared to only 0.06% of other residues.[7] Inspection of the database reveals only 45 nonhomologous nonproline *cis* peptides ($\omega<90°$) in 15 proteins. A single protein (1HDS[8]) accounts for 25 of these occurrences, and 1CY3[9] for another four. Excluding these two structures and restricting attention to resolutions better than 2.5 Å and $\omega<30°$ leaves only eight examples in almost 90,000 residues. Whether this reflects reality or merely a reluctance to interpret density as a *cis*-peptide is unclear. Interpretation depends critically on identifying the density for the peptide oxygen and the local backbone atoms.[10] It requires very good density to be sufficiently confident to suggest such an unusual conformation. Indeed, Stewart et al.[11] have recently shown that protein structures determined to higher resolution (≤ 2.0 Å) have a slightly higher incidence of *cis*-peptides.

φ,ψ Angles

Ramachandran et al.[12] showed that ϕ,ψ space for a dipeptide is very restricted for all residues except glycine, due to steric clashes. It has often been observed that as the refinement of a protein structure proceeds the ϕ,ψ angles migrate into allowed conformations, therefore the distribution of ϕ,ψ angles

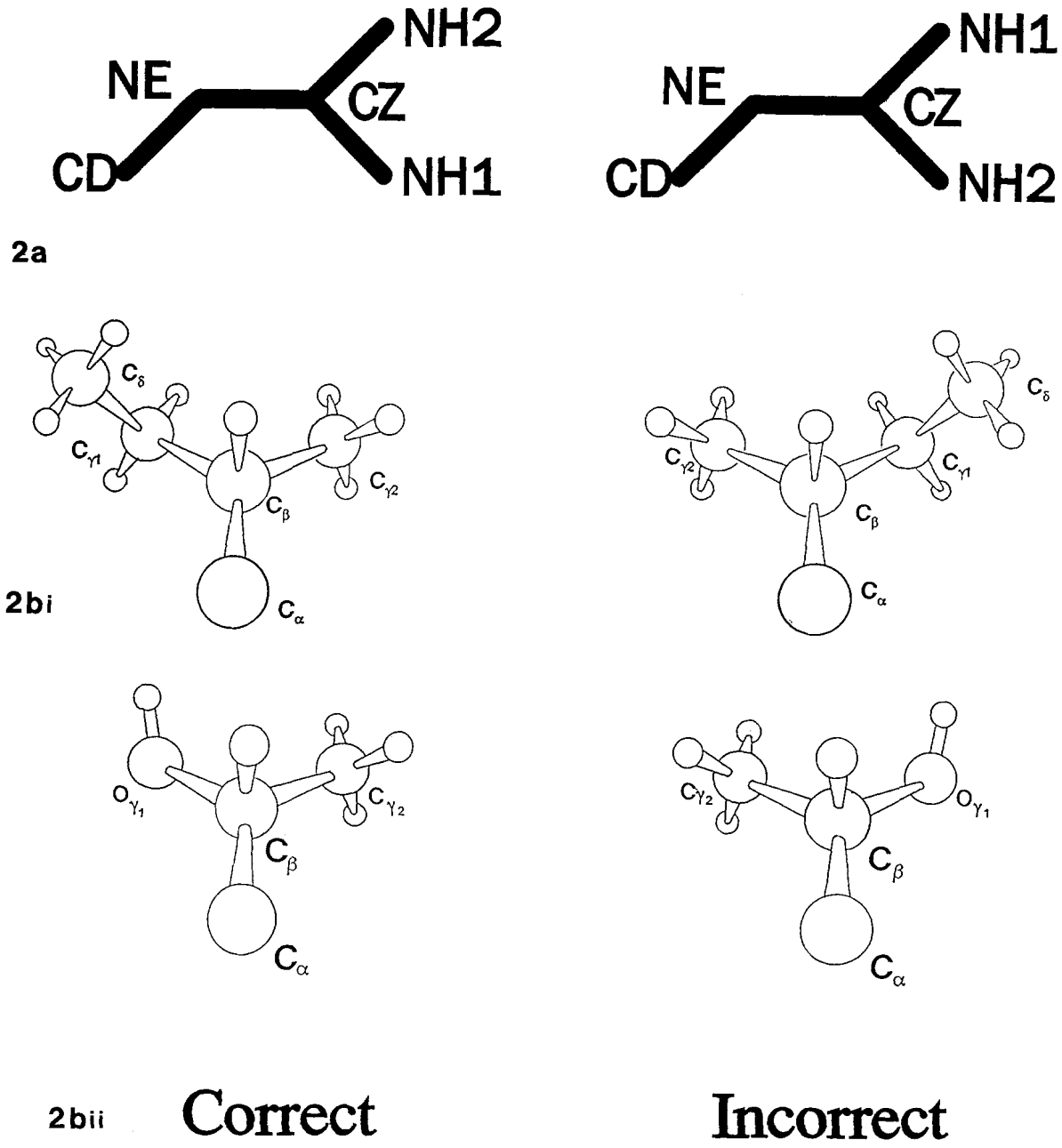


Fig. 2. (a) Atom labeling: The correct and incorrect labeling of arginine (b) Atom labeling: (i) The correct absolute configuration (S) of the standard isoleucine isomer according to the sequence rule of Cahn et al.[35] (ii) For the chiral β carbon of threonine, the correct absolute configuration in the same notation is R. Twenty examples of the S isomeric form were observed in the database.

may provide a guide to the quality of the structure. Three sets of allowed ϕ,ψ angles were defined, using the observed ϕ,ψ distribution for 121,870 residues from 463 structures, extracted from the Brookhaven databank (Fig. 5). Proline and glycine residues were excluded from the study because of their atypical dihedral angle distributions (see Proline ϕ Angles). The number of residues in each $10° \times 10°$ pixel of conformational space was calculated, and the sets of ϕ,ψ angles were based on this population density. The CORE region included all pixels with more than 100 residues, and the ALLOWED, areas with eight or more residues per pixel. The third level, GENEROUS, was defined by extending out by 20° (two pix-

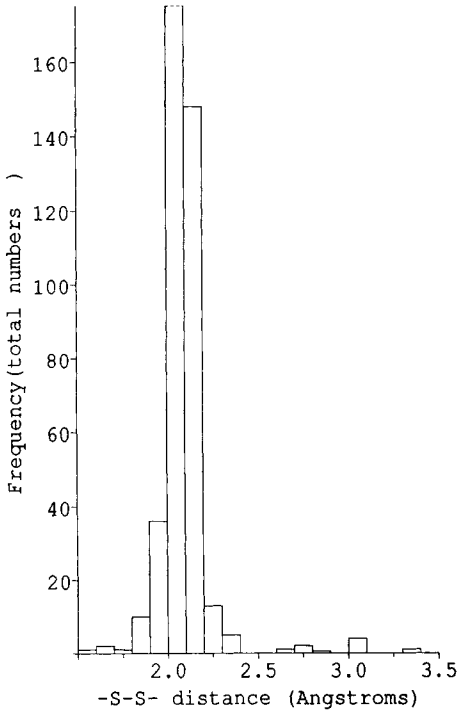


Fig. 3. Distribution of sulfur–sulfur separations in disulfide bridges, for 594 bonds from 126 structures in the dataset of 294 proteins determined to ≤3.5 Å resolution. The mean –S–S– bond length is 2.04 ± 0.16 Å. The truncated column at 2.0 Å contains 369 values.

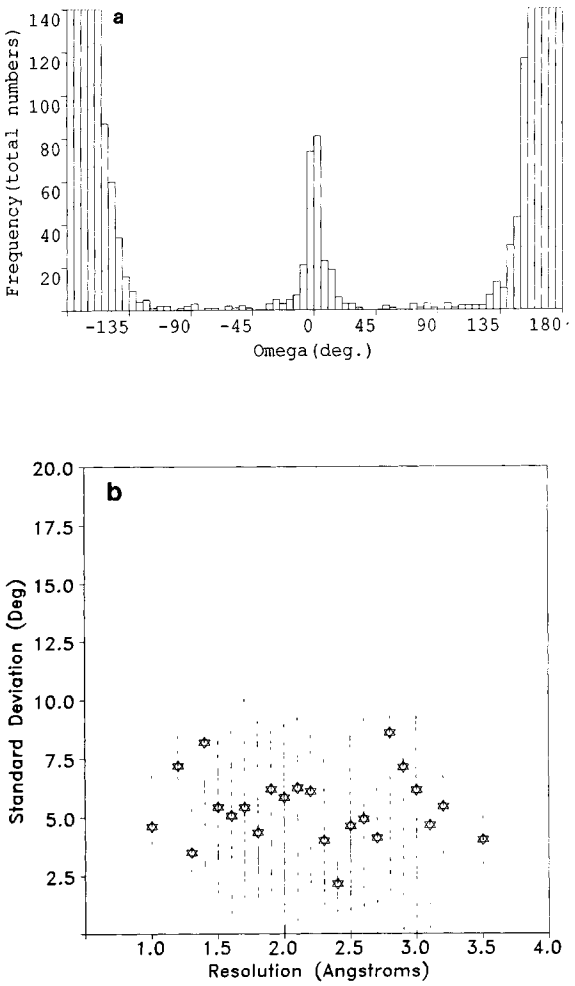


Fig. 4. (**a**) Distribution of peptide bond ω angles. A small but significant number show large deviations from planarity, some by as much as 90°. The plot is for 87,121 angles from 298 structures determined to ≤ 3.5 Å resolution. (**b**) Plot of standard deviations of trans ω angles against resolution. There is no correlation between the two, and the scatter of values about the mean for nearly all structures is very small, sometimes zero, which probably reflects restraints applied during refinement. For the 86,855 trans angles in the dataset, the standard deviation is 5.8° about the mean value of 179.8°. For the 266 cis peptides the ω angle is 0.3 ± 21.8°.

els) all round the ALLOWED region. The space left after these regions had been defined was designated OUTSIDE. The proportions of available space in each region and the percentage of residues populating each region are shown in Table II. It is striking that in this sample, which includes structures at resolutions between 1.0 and 3.5 Å, over 80% are in the central CORE region, with 97% in the CORE and ALLOWED regions together. The GENEROUS region contains very few residues, and could probably be absorbed into the OUTSIDE region.

The percentage of residues which lie *inside* the CORE (%CORE) was then calculated for each protein structure, and plotted in Figure 6a. The average value of %CORE for each resolution is also plotted. For each resolution, the average percentage of residues *within* the three regions was also calculated, i.e., CORE, CORE + ALLOWED and CORE + ALLOWED + GENEROUS; this is shown in Figure 6b. The percentage of residues in the CORE for each structure plotted against the *R*-factor is shown in Figure 7.

Five facts clearly emerge:

1. There is a correlation between resolution and the tightness of the ϕ,ψ distribution. On average at higher resolutions a lower percentage of residues lies outside the allowed regions. The correlation coefficient for percentage of residues in CORE with resolution is 0.53 for individual proteins, 0.85 for data averaged by resolution.

2. For a given resolution range the protein structures show very different accuracies. Some have good ϕ,ψ distributions even at relatively poor resolution, while others show remarkably bad distributions even though the data go to high resolution. It is possible to define outliers, and these proteins often have high *R*-factors and incomplete refinement. Some of these incorrect ϕ,ψ values almost certainly

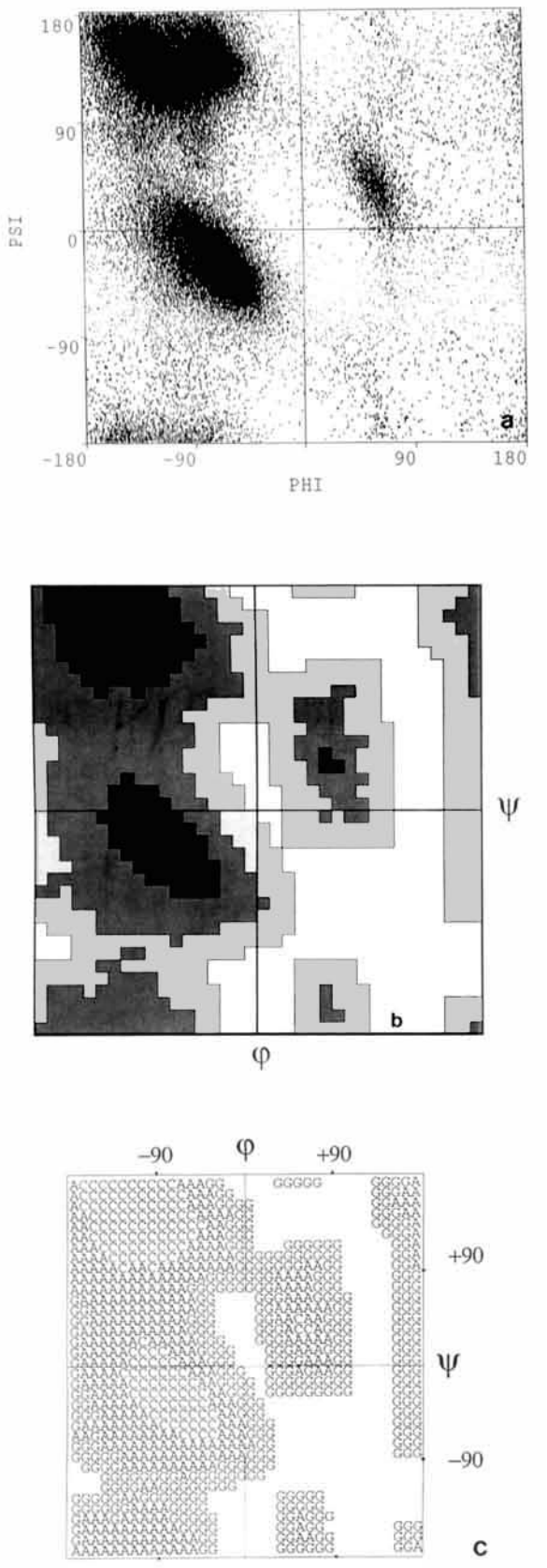


Fig. 5.

TABLE II. Available Area and Percentage Population of the Map Regions

Region	Area (%)	Population (%)
Core	11.0	81.9
Allowed	28.0	14.8
Generous	31.0	2.0
Outside	30.0	1.3

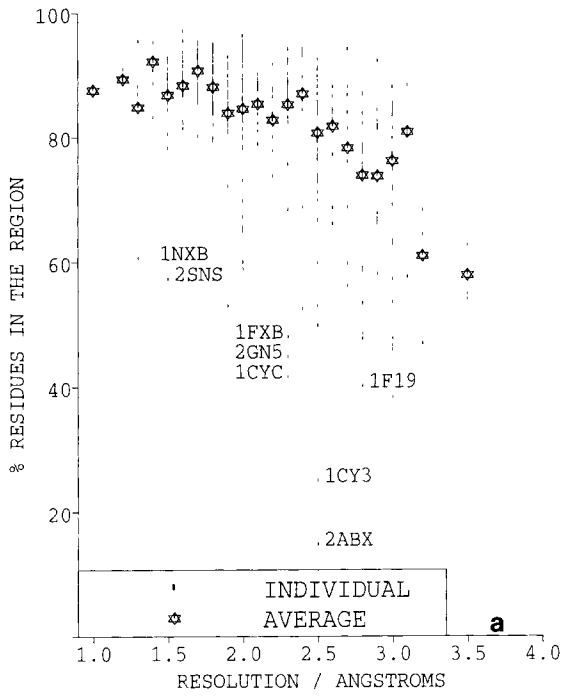
correspond to peptides which have flipped over by 180°, since this conformational change has relatively little effect on the direction of the chain. One method to identify such flips may be to use the database of structures and, given C_α coordinates of a model, search for matching structures using the approach suggested by Jones and Thirup.[13] If all such extracted structures have allowed ϕ,ψ values this may indicate that electron density in this region of the structure should be reinvestigated. The fragment approach was used by Holm and Sander[14] to reconstruct 1CY3 from the C_α coordinates using a database of high resolution fragments. Another study (Weaver et al.[15]), compared the ϕ,ψ distributions for two models of lysozyme. The correct model (resolution 1.6 Å, $R = 0.15$) has a tight distribution, while the incorrect model (resolution 2.8 Å, $R = 0.28$) is very poorly clustered.

3. At high resolution there remain a few residues which adopt disallowed conformations, even using the most generous definition of allowed ϕ,ψ. The existence of some residues in the most disallowed regions reflects one of two possibilities. The first is that the residues must adopt an unfavorable conformation for some structural or functional purpose, the unfavorable energy being compensated for by adjacent residues. The second possibility is that the residues are in regions of poor density, and have been wrongly interpreted.

4. Below 2 Å the curves flatten out which suggests that, as expected, the backbone conformation is relatively well defined by this resolution (but see Helical Structures).

5. There is a correlation (coefficient for individual structures = 0.58) between published R-factor and the percentage of residues in the CORE region, with

Fig. 5. (**a**) Distribution of ϕ,ψ angles for 121,870 residues from 462 protein structures. Proline and glycine were not included. The angles were calculated from Brookhaven coordinates using the program SSTRUC (Smith, private communication). (**b**) Digitized Ramachandran type plot derived from the actual distribution shown in **a**. The population within each 10° × 10° pixel was calculated using the P6 program (McGregor, private communication). The boundaries were defined as described in the text. The black areas are CORE, the dark gray ALLOWED, and the pale gray, GENEROUS. (**c**) The same plot as in **b**, showing the individual pixel assignments. The blank pixels are OUTSIDE. C is CORE, A ALLOWED, and G GENEROUS.



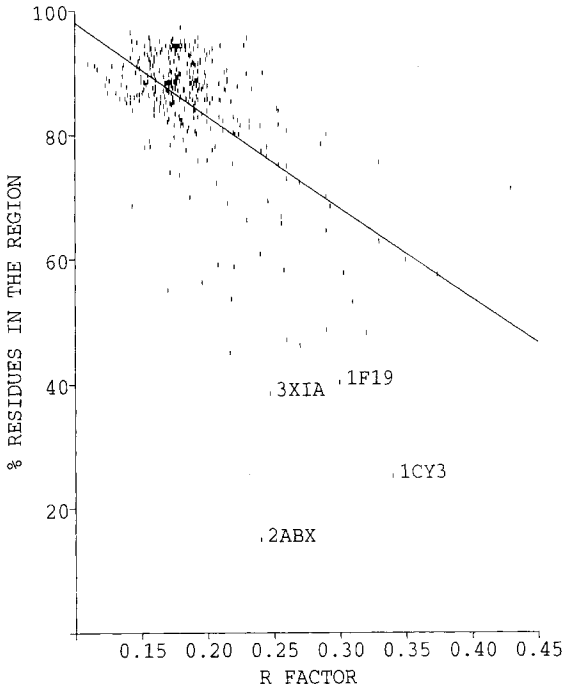


Fig. 7. Graph to show the correlation between the percentage of residues in the core, and the *R*-factor of the structure as quoted by the authors; 360 structures are included. The correlation coefficient is 0.58 for 228 degrees of freedom. The *t* test statistic is a significant 13.58. Outliers relative to other structures of similar *R*-factor are marked with the Brookhaven code.

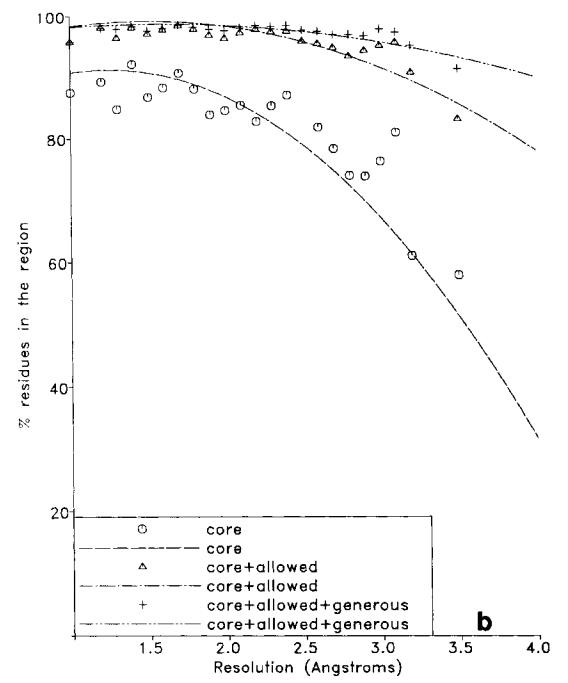


Fig. 6. (**a**) Plot to show the percentage of residues from 462 structures which fall into the CORE regions of the digitized distribution plot (Fig. 5b) as a function of resolution. The symbols used are indicated in the figure. There are several outliers relative to other structures at the same resolution, and these are marked on the plot, using the Brookhaven code for identification. (**b**) Plot to show the percentage of residues falling *within* each of the regional boundaries, as a function of resolution. Only the average points for each resolution are shown, the symbols used being indicated in the figure.

tight clustering of the data below $R = 0.20$. Again there are clear outliers, with some structures having significantly more CORE residues than the poor *R*-factor would suggest. Others with moderately good *R*-factors show dispersed ϕ,ψ distributions.

Proline ϕ Angles

The geometry of the pyrrolidine ring in proline restricts the ϕ angle to an average value around $-65°$. The ring itself is usually slightly puckered and is not always restrained during refinement. Again the database was used to explore the distribution of proline ϕs, as shown in Figure 8, which illustrates that some prolines are distorted beyond the limits of acceptability. In particular, values around $-180°$ and $105°$ would seem to place an intolerable strain on the N–C_α bond. Seventeen others are present in a universally disallowed region of the Ramachandran plot at $\phi = 0°$, however, 12 of these are equivalent residues from the multiple chains of glutamine synthase (2GLS).[16]

Helical Structures

Further consideration was given to the backbone dihedral angles in secondary structures. ϕ,ψ values of helical residues are tightly clustered and as re-

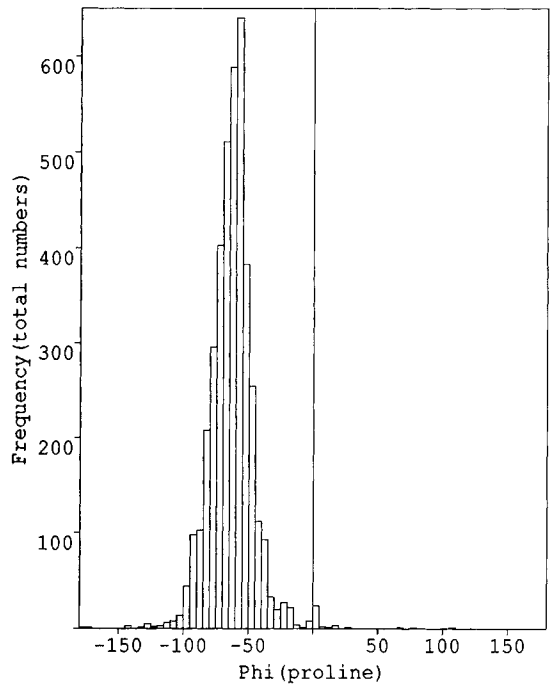


Fig. 8. Distribution of proline ϕ angles for 3,927 observations from a dataset of 298 structures.

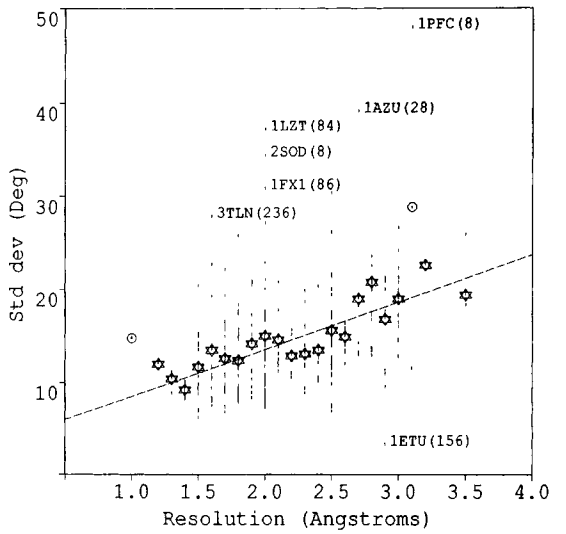


Fig. 9. Correlation of the pooled standard deviations of α-helix ϕ,ψ angles with resolution. The vertical dashes represent the standard deviations within individual proteins, and the asterisks show the standard deviations for the sets of values at each resolution. The observations at 1.0 and 3.1 Å resolution which are marked with circles number only 16 and 24 residues, respectively, out of a total sample of 48,210 angles. With these excluded from the regression analysis, and assuming a Gaussian distribution, the equation of the best fit line through the mean values at each resolution is $y = 4.4 + 4.7x$ with a standard deviation "s" of y about the line equal to 1.7°, a correlation coefficient of 0.88, and a t test statistic of 8.1 for 19 degrees of freedom (cf. 0.78 and 5.7 with the missing values included). Both are highly significant at the 99.9% level. Major outliers relative to resolution are shown using the code from Brookhaven, with the number of angles in parentheses.

finement progresses the standard deviations of such angles should decrease as helix geometry improves. The database was used to calculate averages and standard deviations for ϕ and ψ in helices as defined by the Kabsch–Sander algorithm.[17] The plot of average values and of standard deviations against resolution shown in Figure 9 reveals a striking correlation with resolution (coefficient = 0.88). The standard deviations appear to decrease even below 2 Å suggesting continued improvement. The plot of standard deviation within individual protein structures against resolution shows a much poorer correlation (coefficient = 0.40) and is a reflection of wide variations in helix content, helix length, and helix terminus irregularity. Thus, despite the good underlying linearity, this parameter is not always appropriate for use globally as a criterion of quality.

The same plot for β-strand residues (not shown here) reveals much less correlation, almost certainly because ϕ,ψ values in strands are not tightly clustered since the strands have different twists and distortions.

χ_1 Angles

χ_1 angles have been shown to adopt one of three preferred conformers $g^-(+60°)$, t $(+180°)$, or $g^+(-60°)$.[18] As refinement proceeds the χ_1 distribution becomes more tightly clustered into these three idealized energy wells.[19,20] For each conformer, the deviation of the values from their average was calculated for a given protein. For example in beef liver catalase (8CAT),[21] the 132 residues which adopt a g^- conformation (between 0° and 120°) have an average value of 63.4° ± 22.5°. To calculate a measure of quality for the whole protein, the data for the three conformers were pooled, and a weighted average was calculated:

$$\chi_1 \text{ (pooled)} = \frac{\text{sgm} \times \text{ngm} + \text{st} \times \text{nt} + \text{sgp} \times \text{ngp}}{\text{ngm} + \text{nt} + \text{ngp}}$$

where sgm, st, and sgp are the standard deviations for each of the energy minima, and ngm, nt and ngp are the corresponding totals. Figure 10a shows that as the resolution improves the standard deviation for χ_1 angles reduces dramatically. This is confirmed by the data in Figure 10b, showing the improvement of pooled χ_1 standard deviation with R-factor, averaged by R-factor at intervals of 0.1. The correlation coefficient of pooled χ_1 standard deviations is 0.63 for individual proteins, 0.96 for averaged values at each resolution. The corresponding figures are 0.52 and 0.95 for R-factor. The plots show no flattening off below 2 Å, in contrast to the classification based

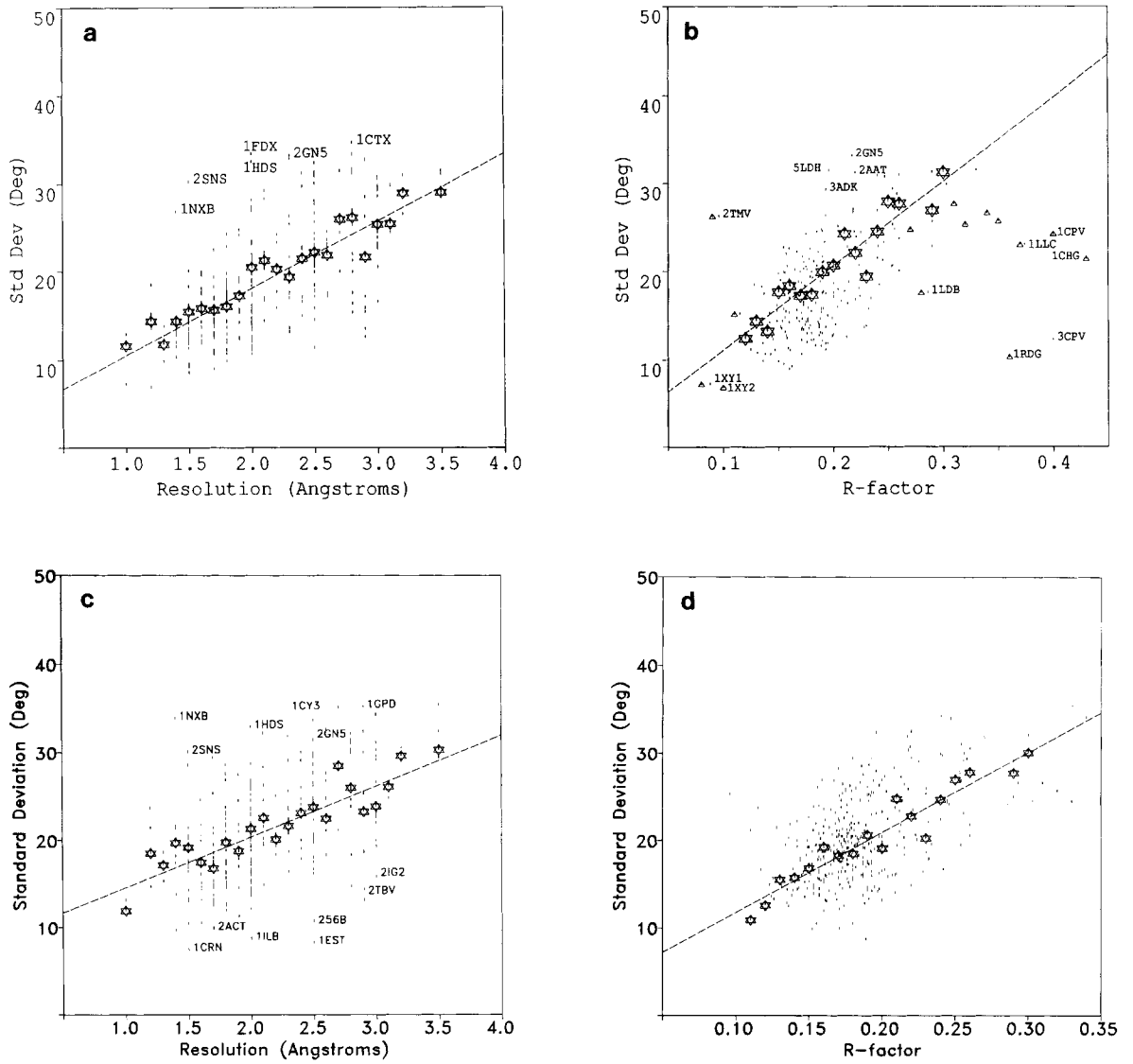


Fig. 10. (**a**) Plot of the pooled standard deviations of χ_1 for the g^+, *trans*, and g^- angles versus resolution. The vertical dashes represent the standard deviations within individual proteins and the asterisks show the standard deviations for the combined values at each resolution. The correlation coefficient is 0.96 for 21 degrees of freedom, and the *t* test statistic for the null hypothesis (coefficient of *x* is zero) is 16.4 which is highly significant at the 99.9% level. For the individual proteins the correlation coefficient is 0.63 for 293 degrees of freedom. The lower value reflects the influence of outliers and the uneven scatter of individual proteins about the regression line. A test of the null hypothesis H_0:(coefficient of *x* is zero) gives $t = 13.7$, which again is very significant at the 99.9% level. The more extreme outliers relative to resolution are indicated, using the Brookhaven code. (**b**) Plot of the pooled standard deviations of χ_1 for g^+, *trans*, and g^- angles versus *R*-factor. The vertical dashes represent the standard deviations within individual proteins and the asterisks show the standard deviations for all angles within *R*-factor intervals of 0.01 between 0.05 and 0.45. Where the interval range contains values representing only a single structure these are shown as triangles and have not been included in the calculation of the regression line. For the regression line, the correlation coefficient is 0.95 and $t = 11.4$, which is highly significant at the 99.9% confidence level for 15 degrees of freedom, cf. 0.52 for 228 degrees of freedom for individual structures. The more extreme outliers are indicated with their Brookhaven code. (**c**) Plot of the standard deviations of χ_2 trans angles versus resolution. The vertical dashes represent the standard deviations within individual structures, and the asterisks the standard deviations for the combined values at each resolution. (**d**) Plot of the standard deviations of χ_2 trans angles versus *R*-factor. The symbols used are the same as in **c**.

on ϕ,ψ distribution. Some side chains are difficult to locate even at 2 Å and significant improvements can be made at higher resolution. Again the standard deviation for χ_1 in each rotamer can provide a guide to the global accuracy of the structure, while the existence of very unusual χ_1s may signify a local error. Secondary structure preferences could also be taken into account.[20]

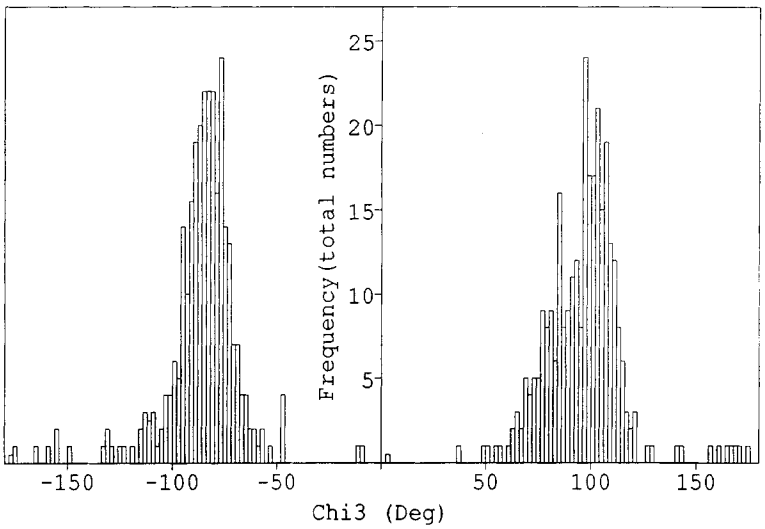


Fig. 11. Distribution of χ_3 disulfide angles. Most of these fall within the preferred regions of the right- or left-handed conformers, but a number of examples occur which deviate by up to 90° either side of the mean. The plot represents 594 angles from 126 proteins from a starting dataset of 126 structures determined to a resolution ≤ 3.5 Å.

χ_2 Angles

The χ_2 angles have a trimodal distribution similar to that of χ_1 only when the C_γ atom is tetrahedral. The distribution is symmetrical about the trans position which is greatly favored due to the low steric repulsion offered by the two β-methylene hydrogens. Plots of standard deviation against resolution and *R*-factor for this conformer (Fig. 10c and d) show a trend similar to that for the χ_1 angles.

Disulfide χ_3 Angle

It is known that the χ_3 angle C_β–S–S–C_β in a disulfide bridge adopts either a right- or left-handed conformation.[22,23] The distribution of angles plotted in Figure 11 shows strong clustering around the right-handed (χ_3 = 96.8° ± 10.1°) and left-handed (χ_3 = −85.8° ± 8.6°) conformers. The distribution is not quite symmetrical about χ_3, the contribution of orbital energy is symmetrical, but that of the van der Waals interactions applies to one side only. There are many disulfides which have been built into very distorted conformations.

CHIRALITY

One consequence of the restriction to L-amino acids in proteins is that many structural features have a preferred chirality, as summarized in Table I. In assessing coordinates the chirality of such features should be checked and, if incorrect, may suggest either an unusual structure which deserves investigation or an incorrect interpretation. Here, the chirality of the individual amino acids, the origin of chirality for higher order structures, is briefly considered. The virtual dihedral angle C_α–N–C–C_β, measuring rotation about the virtual bond N··C within a residue, was used to distinguish L from D stereoisomers of the amino acids. For 93,000 residues this angle has an average value of 33.81° ± 4.17°, as expected for L-amino acids. There are 29 occurrences of values <0°, which can be considered D-amino acids, although the distribution shows distortions occur in both directions, as seen in Figure 12. The chirality of secondary, supersecondary, and tertiary structures is currently under investigation.

NONCOVALENT INTERACTIONS

These perhaps provide one of the best guides to structural accuracy since they are rarely restrained during refinement.

Hydrogen Bonds

Main chain hydrogen bonds are numerous in all proteins, mainly in the α-helix, β-sheet, and β-turn secondary structures. It is reasonable to expect that as refinement proceeds the geometry of the hydrogen bonds will improve, with a corresponding reduction of energy. Figure 13 shows the variation of average H-bond energy, calculated using the Kabsch–Sander method,[17] for residues in structures refined to a given resolution. It shows a definite correlation, confirming the improvement in hydrogen bonding geometry for high resolution structures. The scatter at low and high resolution reflects the smaller number of structures at the extremes of the resolution range. Given the gross simplification of considering all hydrogen bonds together, the correlation is surprisingly good; the coefficient is 0.51 if

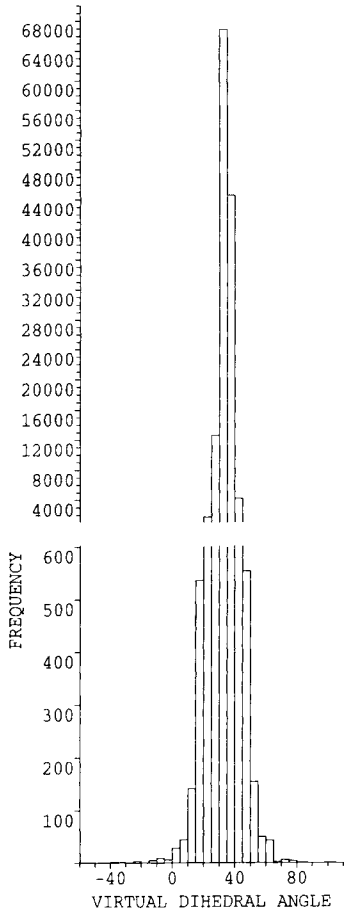


Fig. 12. Distribution of the C_α–N–C–C_β dihedral angle for 93,000 residues taken from all proteins in the October 1990 data bank. Most of the values are tightly clustered around the mean (33.81 ± 4.17°) but the frequency axis has been magnified in the lower part of the graph to show the occurrence of distortions on both sides of the distribution.

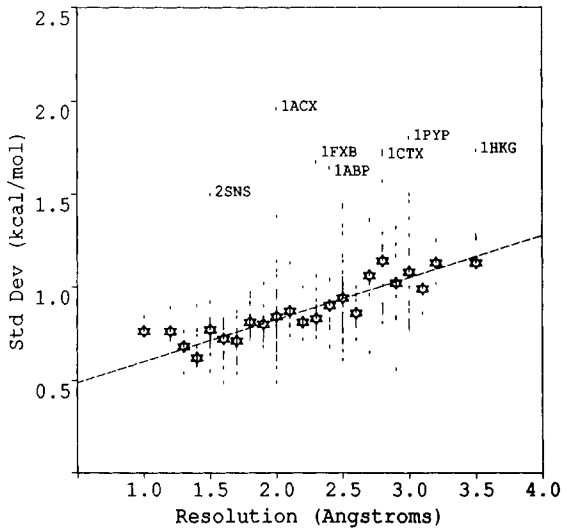


Fig. 13. Plot of the standard deviation of hydrogen bond donor energies versus resolution. The dots represent the standard deviations within individual proteins, and the asterisks show the standard deviations for the sets of values at each resolution. The correlation coefficient is 0.91 for 21 degrees of freedom, and the *t* test statistic for the null hypothesis H_0:(coefficient of *x* is zero) is 9.8, which is highly significant at the 99.9% level. For the individual proteins the correlation coefficient is 0.51 for 296 degrees of freedom. The much lower value reflects the influence of outliers and the greater scatter about the regression line for individual proteins. A *t* test of the null hypothesis (coefficient of *x* is zero) gives $t = 10.0$ which again is significant at the 99.9% level. The more extreme outliers are indicated, using the Brookhaven identifier.

individual structures are considered, 0.91 for the resolution-averaged data.

Noncovalent Interactions

Close atom contacts which violate the minimal accepted van der Waals separation usually signal poor density or interpretation. Islam and co-workers (private communication) have considered these close contacts and find a strong correlation between resolution, respectively, and the numbers of bad contacts.

REFINEMENT METHOD

In observing these correlations between structural parameters and resolution, an immediate concern is the influence of the refinement method and weightings used, since the correlations might merely be a reflection of the stronger restraints imposed at higher resolutions. In fact, restraints are effectively relaxed at higher resolutions, when there are many more reflections included, and additional external information makes a smaller contribution. However, it is interesting to question whether the different refinement methods give similar correlations.

This was investigated by considering the ϕ,ψ distributions for protein structures refined by different methods (Fig. 14). Four methods were studied, Hendrickson–Konnert (PROLSQ),[24] Jack–Levitt (EREF, DERIV, and modifications),[25] Ten Eyck and Tronrud (TNT),[26] and the simulated annealing method XPLOR.[27] The first plot shows the data for the PROLSQ refined structures, and the second the data for the other three methods. The plots immediately highlight the differences between the methods. Those structures refined using PROLSQ (Hendrickson and Konnert[24]) a least-squares method, and still a widely used refinement program, show strong improvement in the ϕ,ψ distribution as more reflection data become available at higher resolutions. The ϕ,ψ angles are usually not restrained and there are mostly weak noncovalent restraints applied.

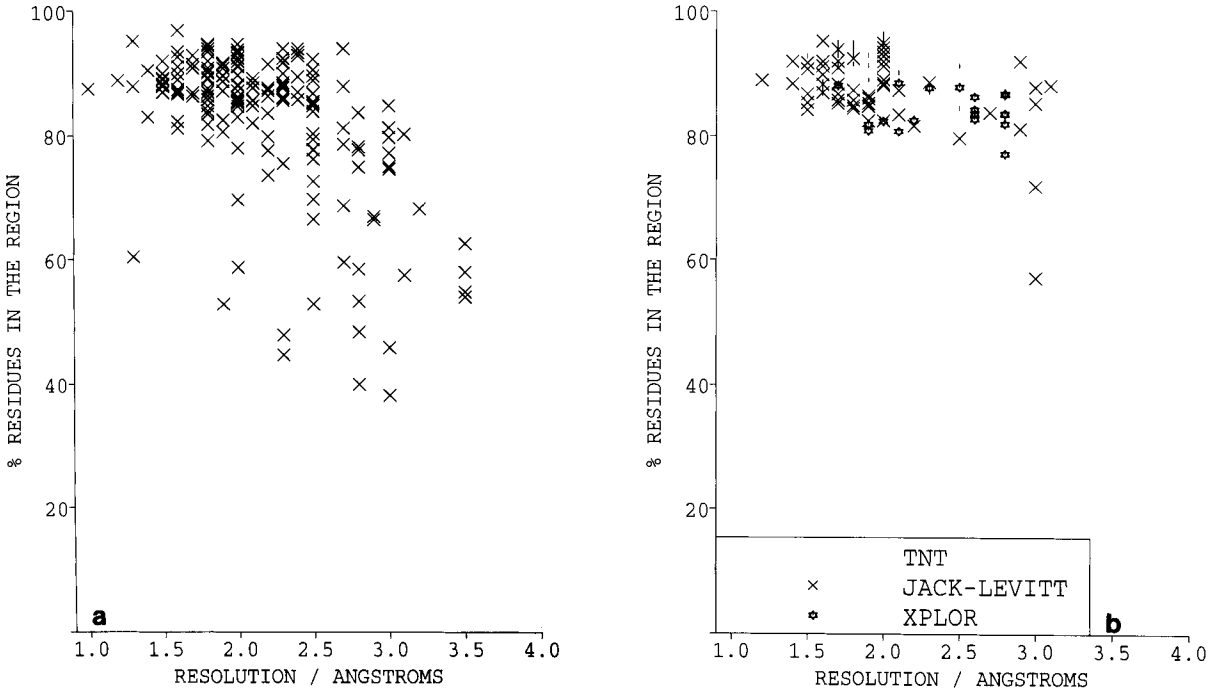


Fig. 14. (**a**) The plot of percentage residues in the CORE versus resolution for 131 structures refined using the method of Hendrickson and Konnert.[24] The information about refinement method was taken from remarks in the relevant Brookhaven file. The crosses represent structures where the first refinement method used was Hendrickson and Konnert; no account was taken of subsequent refinement methodology, if any. (**b**) The plot of percentage residues in the CORE versus resolution for 140 structures refined using energy-based methods. The crosses represent the 53 structures refined using the least-squares program TNT.[26] The asterisks denote 69 structures refined using the Jack–Levitt method,[25] a program which simultaneously minimizes energy and *R*-factor. The asterisks represent 18 structures refined using XPLOR,[27] which incorporates a simulated annealing algorithm.

TABLE III. Summary of Expected Values for Stereochemical Parameters in Well-Resolved Structures

ϕ,ψ in CORE (%)	>90%
χ_1 (gauche +)	−66.7 ± 15.0°
χ_1 (gauche −)	+64.1 ± 15.7°
χ_1 (*trans*)	183.6 ± 16.8°
SD of pooled χ_1	15.7°
χ_3 (S–S bridge)	
LH	−85.8 ± 10.7°
RH	96.8 ± 14.8°
ϕ (Proline)	−65.4 ± 11.2°
Helix angle	
ϕ	−65.3 ± 11.9°
ψ	−39.4 ± 11.3°
hydrogen bond energy	−2.03 ± 0.75 kcal/mol
trans ω	179.6° ± 4.7°
C_α–N–C–C_β dihedral angle	33.9° ± 3.5°

In contrast, the Jack–Levitt and XPLOR methods, which are energy-based, show good ϕ,ψ distributions even at low resolution. Thus, the noncovalent energy restraints apparently do not allow the ϕ,ψ angles to deviate substantially from the preferred values. This means that the ϕ,ψ distribution cannot be used to assess the accuracy of structures refined by these energy-based methods.

STEREOCHEMICAL QUALITY INDEX

In the database of protein structures it is useful to have a simple guide to the stereochemical "quality" of a structure in comparison with other structures, either at the same resolution, or those of high resolution. Clearly, resolution and *R*-factor (if available) are the optimal guides. In addition, the correlations described above allow a classification based on stereochemical considerations. Table III lists the observed values for the parameters discussed above, as derived from 119 high-resolution (≤2.0 Å), well-refined ($R \leq 0.20$) structures, and therefore expected for subsequent structures refined to the same level. The classifications for each of the protein structures in the October 1990 release of Brookhaven data are given in Table IVa, along with the resolution and *R*-factor. The three parameters selected for the classification were ϕ,ψ distribution, pooled χ_1 standard deviations, and hydrogen bond donor energies. In each case, four classes were defined, using the statistical data for each parameter. The class boundaries for all the parameters are shown in Table IVb. Potential users of these structures should be aware

TABLE IVa. Classification of Structures According to Stereochemical Parameters*

CODE	RES	RFACTR	%CORE	HBOND	CHIDEV	CODE	RES	RFACT	%CORE	HBOND	CHIDEV
155c	2.5	-	49.6	1.32	30.83	1118	1.7	0.157	93.3	0.62	13.50
156b	2.5	-	73.2	1.43	32.26	1119	1.7	0.153	94.0	0.61	12.20
1abp	2.4	-	52.3	1.63	29.61	1120	1.9	0.153	92.7	0.63	12.98
1acx	2.0	-	64.8	1.95	28.06	1121	1.9	0.173	93.3	0.66	12.71
1alc	1.7	0.220	80.0	0.69	22.35	1122	1.7	0.162	92.6	0.68	12.63
1amt	1.5	0.155	88.2	0.54	9.39	1123	1.7	0.157	94.7	0.62	12.04
1atl	2.8	0.164	83.6	0.75	17.11	1124	1.7	0.158	92.6	0.60	13.12
1azu	2.7	0.350	59.5	0.95	25.61	1125	1.8	0.180	93.4	0.61	15.24
1bmv	3.0	0.330	75.2	1.02	26.44	1126	1.7	0.184	94.0	0.68	17.32
1bp2	1.7	0.171	89.3	0.69	15.16	1127	1.8	0.157	94.0	0.61	13.52
1ca2	2.0	0.173	87.6	0.80	13.30	1128	1.9	0.153	92.7	0.62	12.56
1cc5	2.5	0.290	69.7	1.04	25.52	1129	1.7	0.178	94.0	0.68	17.05
1ccr	1.5	0.190	87.9	0.79	12.90	1130	1.7	0.161	93.4	0.62	13.80
1cdp	1.6	0.164	90.0	0.82	14.70	1131	1.8	0.173	94.7	0.68	18.38
1chg	2.5	0.430	71.0	1.13	21.31	1132	1.7	0.166	93.4	0.64	14.07
1cho	1.8	0.168	85.8	0.83	18.71	1133	1.7	0.157	91.3	0.63	12.34
1cla	2.3	0.172	90.0	0.75	12.70	1134	1.9	0.176	92.7	0.69	13.25
1cms	2.3	0.165	85.7	0.86	20.09	1135	1.8	0.157	91.3	0.63	12.40
1cn1	3.2	0.260	46.7	1.01	30.86	1ldb	2.8	0.286	78.2	0.89	17.48
1coh	2.9	-	92.1	0.55	15.33	1ldm	2.1	0.173	88.5	0.79	20.16
1crn	1.5	-	89.2	0.69	8.42	1lh1	2.0	-	95.0	0.56	11.21
1cse	1.2	0.178	89.0	0.73	13.41	1lh2	2.0	-	94.3	0.60	10.54
1ctf	1.7	0.174	95.1	0.66	13.80	1lh3	2.0	-	94.3	0.59	11.09
1cts	2.7	0.183	83.8	0.79	20.09	1lh4	2.0	-	92.9	0.61	10.95
1ctx	2.8	-	47.5	1.72	34.53	1lh5	2.0	-	94.3	0.59	10.88
1cy3	2.5	0.340	24.8	1.32	26.51	1lh6	2.0	-	94.3	0.62	12.21
1cyc	2.3	-	41.4	1.05	32.73	1lh7	2.0	-	93.6	0.58	11.05
1dhf	2.3	0.176	89.2	0.85	21.18	1llc	3.0	0.374	57.1	0.81	22.92
1eca	1.4	-	95.0	0.59	14.02	1lrd	2.5	0.242	89.6	0.75	18.36
1ecd	1.4	-	95.0	0.60	12.80	1lyd	2.0	0.191	94.7	0.74	15.70
1ecn	1.4	-	95.0	0.59	13.16	1lym	2.5	0.260	72.6	1.28	23.59
1eco	1.4	-	95.0	0.59	12.71	1lyz	2.0	-	67.0	0.99	20.63
1est	2.5	-	76.9	0.71	11.25	1lz1	1.5	0.177	88.0	0.69	14.49
1etu	2.9	-	82.5	1.16	15.80	1lzt	2.0	0.254	83.5	0.69	20.55
1f19	2.8	0.300	39.9	1.56	31.56	1mb5	1.8	-	87.0	0.76	24.15
1fc1	2.9	-	81.7	0.83	12.52	1mba	1.6	0.193	95.3	0.61	13.58
1fc2	2.8	-	78.6	0.80	15.24	1mbc	1.5	0.171	92.0	0.63	13.35
1fd2	1.9	0.232	81.9	0.67	15.64	1mbd	1.4	-	92.0	0.57	11.30
1fdh	2.5	-	87.9	0.65	22.97	1mbn	2.0	-	83.3	0.93	21.11
1fdx	2.0	-	80.0	0.82	33.30	1mbo	1.6	-	92.0	0.56	11.99
1fx1	2.0	-	84.1	0.87	21.85	1mbs	2.5	-	73.7	0.94	31.24
1fxb	2.3	0.320	47.8	1.66	25.27	1mbw	1.9	0.148	91.4	0.63	14.01
1gcn	3.0	-	71.4	0.76	28.53	1mcp	2.7	0.225	81.2	1.01	24.34
1gcr	1.6	0.230	89.5	0.73	19.48	1mcw	3.5	0.170	54.7	1.00	30.21
1gd1	1.8	0.177	90.2	0.78	14.23	1ntp	1.8	0.187	86.3	0.96	13.57
1gox	2.0	0.189	91.3	0.83	15.67	1nxb	1.4	0.240	60.4	0.89	26.73
1gp1	2.0	0.171	88.2	0.75	13.29	1ovo	1.9	-	83.7	0.82	12.39
1gpd	2.9	-	57.9	1.31	32.60	1p01	2.0	0.138	85.4	0.77	10.54
1hbs	3.0	0.254	81.3	1.19	29.25	1p02	2.0	0.147	86.1	0.73	11.63
1hco	2.7	-	75.9	1.01	26.14	1p03	2.2	0.142	84.8	0.72	11.73
1hds	2.0	-	52.8	1.05	32.39	1p04	2.6	0.134	85.5	0.77	13.09
1hho	2.1	0.223	87.4	0.73	20.54	1p05	2.1	0.139	84.8	0.79	12.82
1hip	2.0	0.240	81.1	0.72	21.37	1p06	2.3	0.140	86.1	0.80	14.11
1hkg	3.5	-	53.9	1.72	30.99	1p07	2.3	0.151	86.4	0.72	14.53
1hmg	3.0	0.204	77.2	0.86	25.00	1p08	2.3	0.139	87.3	0.76	11.17
1hmq	2.0	0.173	89.7	0.62	20.92	1p09	2.2	0.135	87.7	0.73	12.27
1hmz	2.0	0.175	88.2	0.62	20.41	1p10	2.3	0.144	83.6	0.72	12.86
1hne	1.8	0.164	84.0	0.80	18.57	1p2p	2.6	0.241	77.0	0.81	22.71
1hoe	2.0	0.199	89.1	0.78	14.58	1pad	2.8	-	73.6	1.02	21.79
1ilb	2.0	0.189	85.3	0.77	11.43	1paz	1.5	0.180	89.4	0.73	12.88
1101	1.7	0.184	93.3	0.68	17.21	1pcy	1.6	0.170	88.1	0.78	19.82
1102	1.7	0.183	94.7	0.67	17.59	1pfc	3.1	0.303	57.4	0.85	29.77
1103	1.7	0.164	94.7	0.60	13.28	1pfk	2.4	0.165	92.9	0.76	14.42
1104	1.7	0.183	94.0	0.68	16.86	1phh	2.3	0.193	75.5	0.98	26.21
1105	1.7	0.183	94.0	0.62	16.95	1pmb	2.5	0.185	89.6	0.80	20.52
1106	1.7	0.174	94.0	0.64	16.90	1pp2	2.5	0.178	86.8	0.84	21.25
1107	1.7	0.177	94.0	0.64	17.85	1ppd	2.0	0.145	85.6	0.76	12.17
1108	1.7	0.187	94.0	0.66	17.68	1ppt	1.4	-	86.7	0.67	11.36
1109	1.7	0.179	94.0	0.64	17.12	1prc	2.3	0.193	88.7	0.78	17.39
1110	1.7	0.173	94.7	0.67	16.94	1psg	1.6	0.173	88.4	0.79	16.42
1111	1.7	0.181	94.0	0.66	17.20	1pyp	3.0	-	47.6	1.79	30.68
1112	1.7	0.176	93.3	0.66	17.55	1r08	3.0	-	84.0	0.87	23.10
1113	1.7	0.174	94.0	0.64	17.29	1r1a	3.2	0.293	68.1	0.96	25.29
1114	1.7	0.176	94.0	0.67	16.41	1r69	2.0	0.193	91.1	0.66	11.00
1115	1.7	0.175	94.0	0.65	17.48	1rbb	2.5	0.218	85.0	1.16	26.57
1116	1.7	0.177	93.4	0.68	17.81	1rdg	1.4	0.360	90.5	0.73	10.19
1117	1.7	0.157	94.0	0.61	13.11	1rei	2.0	-	86.0	0.77	23.38

(continued)

TABLE IVa. Classification of Structures According to Stereochemical Parameters* (Continued)

CODE	RES	RFACT	%CORE	HBOND	CHIDEV	CODE	RES	RFACT	%CORE	HBOND	CHIDEV
1rhd	2.5	-	76.8	1.00	27.97	2il1b	2.0	0.172	84.7	0.83	14.46
1rmu	3.0	-	84.5	0.87	22.54	2ig2	3.0	0.207	71.9	0.86	16.49
1rn3	1.5	0.260	80.3	0.82	19.91	2ins	2.5	0.180	81.1	0.97	19.80
1rns	2.0	-	65.0	1.08	26.79	2kai	2.5	0.224	79.7	0.85	17.07
1rnt	1.9	0.191	90.9	0.73	19.17	2lbp	2.4	0.213	87.0	0.82	22.69
1rsm	2.0	0.184	85.5	0.73	11.49	2ldb	3.0	0.260	74.6	0.85	18.14
1s01	1.7	0.152	86.9	0.74	9.61	2ldx	3.0	0.256	66.4	0.99	24.68
1sbc	2.5	0.206	77.8	0.94	24.25	2lh1	2.0	-	95.0	0.57	11.00
1sbt	2.5	-	73.7	1.02	17.75	2lh2	2.0	-	94.3	0.60	10.49
1sgc	1.8	0.123	87.6	0.75	12.12	2lh3	2.0	-	94.3	0.59	10.76
1sgt	1.7	0.161	86.9	0.75	10.99	2lh4	2.0	-	93.6	0.58	10.68
1sn3	1.8	-	88.5	0.80	16.02	2lh5	2.0	-	94.3	0.58	10.57
1snc	1.7	0.161	86.7	0.85	15.52	2lh6	2.0	-	95.0	0.62	11.27
1srn	1.8	0.204	85.5	0.85	18.37	2lh7	2.0	-	94.3	0.58	10.60
1tec	2.2	0.179	82.6	0.79	17.57	2lhb	2.0	0.142	92.0	0.62	14.11
1tgb	1.8	-	84.7	0.94	19.98	2liv	2.4	0.179	89.5	0.81	22.47
1tgc	1.8	0.180	85.8	0.81	14.61	2ltn	1.7	0.177	88.2	0.75	13.48
1tgn	1.6	0.223	82.0	0.82	17.98	2lym	2.0	0.149	86.1	0.72	16.07
1tgs	1.8	0.186	84.9	0.82	11.70	2lyz	2.0	-	83.5	0.69	19.36
1tgt	1.7	0.187	85.8	0.83	13.85	2lz2	2.2	0.192	79.8	0.99	23.20
1tim	2.5	-	76.5	0.89	20.81	2lzm	1.7	0.193	94.0	0.70	18.04
1tld	1.5	0.167	84.2	0.82	14.58	2lzt	2.0	0.124	88.7	0.66	11.99
1tlp	2.3	0.174	87.2	0.75	16.56	2mba	2.0	0.156	95.3	0.65	14.66
1tmn	1.9	0.171	86.5	0.71	13.59	2mcg	2.0	0.187	69.6	1.13	27.18
1tnf	2.6	0.230	65.9	0.78	23.05	2mcp	3.1	0.185	80.2	0.99	24.40
1ton	1.8	0.196	87.0	0.91	19.69	2mev	3.0	0.221	79.7	0.78	23.79
1tpa	1.9	0.175	84.9	0.84	13.88	2mhb	2.0	-	88.5	0.82	18.97
1tpo	1.7	0.180	85.8	0.80	11.11	2mhr	1.7	0.158	93.5	0.53	12.34
1tpp	1.4	0.191	88.4	0.81	11.41	2mlt	2.0	0.198	93.2	0.62	16.10
1trm	2.3	0.160	86.2	0.92	16.88	2or1	2.5	0.179	73.2	0.82	28.36
1ubq	1.8	0.176	94.0	0.77	17.21	2ovo	1.5	0.199	91.8	0.74	12.57
1utg	1.3	0.230	95.2	0.53	13.69	2pab	1.8	0.290	79.7	0.90	22.75
1wgc	1.9	0.172	73.6	0.98	18.35	2pad	2.8	-	74.3	1.02	21.87
1wrp	2.2	0.204	91.5	0.71	15.82	2paz	2.0	0.156	86.0	0.74	19.70
1wsy	2.5	0.253	83.9	0.75	20.47	2pcy	1.8	0.160	89.3	0.78	18.89
2aat	2.8	0.220	58.4	1.28	31.12	2pfk	2.4	0.166	93.4	0.78	17.01
2abx	2.5	0.240	14.5	0.60	30.74	2pgk	3.0	-	63.2	1.36	
2act	1.7	0.171	87.4	0.76	10.53	2phh	2.7	0.168	87.9	0.87	21.28
2alp	1.7	0.131	86.4	0.73	10.67	2pka	2.0	0.220	82.5	0.83	16.69
2app	1.8	0.136	90.4	0.83	11.80	2plv	2.9	0.200	75.6	1.00	17.07
2apr	1.8	0.143	94.7	0.79	10.54	2prk	1.5	0.167	89.0	0.67	15.52
2at1	2.8	0.170	83.7	0.75	17.37	2ptc	1.9	0.187	86.1	0.84	13.64
2atc	3.0	0.270	45.8	1.54	31.35	2ptn	1.5	0.193	86.8	0.80	11.72
2aza	1.8	0.157	90.4	0.72	15.35	2r04	3.0	-	83.8	0.86	23.26
2b5c	2.0	-	85.7	0.88	19.82	2r06	3.0	-	83.7	0.88	23.31
2bp2	3.0	0.219	75.0	0.91	24.38	2r07	3.0	-	83.3	0.86	23.06
2c2c	2.0	0.172	88.1	0.61	14.51	2rhe	1.6	0.149	90.3	0.64	15.46
2ca2	1.9	0.176	88.5	0.81	15.40	2rm2	3.0	-	84.0	0.86	23.12
2cab	2.0	0.193	88.4	0.83	13.56	2rmu	3.0	-	84.8	0.87	22.58
2ccy	1.7	0.188	92.7	0.58	14.52	2rnt	1.8	0.149	92.0	0.72	14.28
2cdv	1.8	0.176	87.1	0.77	12.83	2rr1	3.0	-	84.4	0.86	22.84
2cga	1.8	0.173	84.5	0.81	16.06	2rs1	3.0	-	84.0	0.86	23.08
2cha	2.0	-	79.6	0.79	22.97	2rs3	3.0	-	83.8	0.86	22.85
2ci2	2.0	0.198	89.7	0.96	20.84	2rs5	3.0	-	83.5	0.88	22.80
2cla	2.4	0.150	88.5	0.74	13.68	2rsp	2.0	0.144	93.4	0.85	12.52
2cna	2.0	-	72.9	1.13	28.06	2sbt	2.8	-	58.8	1.21	33.25
2cpp	1.6	0.190	91.1	0.73	12.03	2sec	1.8	0.136	89.1	0.71	9.76
2cpv	1.9	0.270	72.0	0.89	24.67	2sga	1.5	0.126	88.3	0.74	12.00
2cro	2.4	0.195	91.7	0.67	14.87	2sni	2.1	0.154	87.7	0.79	14.27
2cts	2.0	0.161	88.5	0.75	18.29	2sns	1.5	-	57.1	1.49	30.14
2cyp	1.7	0.202	92.9	0.72	13.84	2sod	2.0	0.256	65.4	0.94	26.92
2dhb	2.8	-	74.2	0.85	21.16	2ssi	2.6	-	68.5	0.99	23.48
2dhf	2.3	0.194	87.7	0.91	22.52	2stv	2.5	-	88.0	0.78	14.42
2est	2.5	0.210	85.2	0.77	17.65	2taa	3.0	-	53.4	0.90	29.66
2fb4	1.9	0.189	87.2	0.81	15.28	2tbv	2.9	-	66.0	1.22	13.42
2fbj	2.0	0.194	87.2	0.72	19.08	2tga	1.8	0.197	87.4	0.80	14.57
2fd2	1.9	0.232	80.9	0.69	13.28	2tgd	2.1	0.182	78.8	0.86	21.01
2gbp	1.9	0.146	89.6	0.70	14.65	2tgp	1.9	0.200	85.3	0.83	14.87
2gch	1.9	0.181	81.6	0.82	19.17	2tgt	1.7	0.209	85.3	0.83	12.47
2gcr	2.3	0.143	68.2	0.85	29.03	2tmn	1.6	0.179	88.3	0.69	14.40
2gd1	2.5	0.177	88.9	0.81	15.76	2tmv	2.9	0.096	66.9	0.99	26.15
2gls	3.5	0.258	57.9	1.27	28.69	2tpi	2.1	0.200	83.5	0.81	15.86
2gn5	2.3	0.217	44.6	1.05	33.06	2trm	2.8	0.157	77.7	0.88	21.89
2hco	2.7	-	86.6	0.64	19.90	2ts1	2.3	0.228	94.0	0.76	13.53
2hf1	2.5	0.245	76.2	0.98	21.11	2utg	1.6	0.190	93.5	0.62	17.38
2hhb	1.7	0.160	92.1	0.61	15.20	2wgc	2.2	0.153	77.6	0.99	18.37
2hla	2.6	0.173	86.4	0.80	17.47	2wrp	1.6	0.180	96.9	0.48	12.08

(continued)

TABLE IVa. Classification of Structures According to Stereochemical Parameters* (Continued)

CODE	RES	RFACT	%CORE	HBOND	CHIDEV	CODE	RES	RFACT	%CORE	HBOND	CHIDEV
2yhx	2.1	-	78.7	1.12	28.19	4mdh	2.5	0.167	84.2	0.82	22.10
351c	1.6	0.195	91.3	0.72	11.76	4pad	2.8	-	74.3	1.01	21.88
3adk	2.1	0.193	89.9	0.88	29.15	4pcy	2.2	0.150	88.1	0.79	19.62
3apr	1.8	0.147	94.5	0.76	10.99	4pep	1.8	0.174	84.8	0.84	18.23
3at1	2.8	0.181	82.0	0.78	19.68	4pfk	2.4	0.169	94.0	0.81	17.01
3bcl	1.9	0.189	88.7	0.70	17.43	4pti	1.5	0.162	85.4	0.77	9.11
3bp2	2.1	0.174	85.6	0.87	16.96	4ptp	1.3	0.171	87.9	0.75	12.10
3c2c	1.7	0.175	81.2	0.70	14.13	4rhv	3.0	0.160	84.8	0.87	22.51
3ca2	2.0	0.192	87.6	0.82	13.18	4rxn	1.2	0.128	90.7	0.88	18.35
3cla	1.8	0.157	91.0	0.68	10.36	4sbv	2.8	0.254	74.9	0.90	25.17
3cln	2.2	0.175	91.5	0.80	24.13	4sgb	2.1	0.142	85.7	0.74	12.05
3cna	2.4	-	68.6	1.03	31.09	4tln	2.3	0.169	84.9	0.83	17.44
3cpa	2.0	-	86.2	0.73	16.59	4tmn	1.7	0.170	88.0	0.73	14.43
3cpp	1.9	0.190	89.7	0.75	12.14	4tnc	2.0	0.172	93.8	0.67	21.57
3cts	1.7	0.192	88.4	0.70		4tpi	2.2	0.170	81.7	0.87	13.66
3cyt	1.8	-	92.5	0.73	17.37	4ts1	2.5	0.187	92.3	0.86	14.31
3dfr	1.7	0.152	86.9	0.73	13.61	4xia	2.3	0.156	91.8	0.72	13.06
3ebx	1.4	0.176	83.0	0.76	11.85	5acn	2.1	0.215	88.6	0.72	14.58
3est	1.6	0.169	87.0	0.80	14.20	5adh	2.9	-	84.2	0.82	21.83
3fab	2.0	-	59.8	1.37	27.28	5at1	2.6	0.160	83.7	0.77	17.00
3fxc	2.5	0.310	52.8	1.07	27.56	5cha	1.7	0.179	82.3	0.86	19.27
3fxn	1.9	0.214	86.8	0.81	18.68	5cpa	1.5	-	86.9	0.72	16.11
3gap	2.5	0.256	66.5	0.89	27.82	5cpv	1.6	0.187	89.0	1.00	11.63
3gch	1.9	0.203	80.6	1.04	22.39	5cts	1.9	0.185	91.0	0.69	19.16
3gpd	3.5	0.330	62.5	1.25	31.38	5cyt	1.5	0.159	90.8	0.70	11.19
3grs	1.5	0.186	92.5	0.71	14.07	5dfr	2.3	0.198	88.1	0.88	22.73
3hfm	3.0	0.246	69.0	1.17	23.13	5ebx	2.0	0.168	83.0	0.65	17.20
3hhb	1.7	0.200	92.5	0.58	16.94	5gch	2.7	0.156	78.6	1.18	20.94
3hla	2.6	0.169	87.9	0.71	17.10	5ldh	2.7	0.196	55.9	1.35	31.38
3hvp	2.8	0.184	77.2	1.00	16.28	5lyz	2.0	-	78.3	0.97	19.58
3icb	2.3	0.178	92.4	0.89	17.51	5mbn	2.0	0.179	92.0	0.61	13.78
3ins	1.5	0.182	87.0	0.67	12.13	5pad	2.8	-	73.9	1.02	21.73
3ldh	3.0	-	62.7	1.45	30.49	5pcy	1.8	0.160	89.3	0.74	18.98
3lym	2.0	0.149	85.2	0.72	14.24	5pep	2.3	0.196	83.0	0.85	21.59
3lyz	2.0	-	82.6	0.71	21.69	5pti	1.0	0.200	87.5	0.73	12.74
3lzm	1.7	0.157	94.0	0.60	12.61	5rsa	2.0	0.159	87.2	0.76	17.39
3mba	2.0	0.142	96.1	0.62	15.17	5rxn	1.2	0.115	90.7	0.88	15.09
3mcg	2.0	0.208	58.7	0.93	26.56	5tln	2.3	0.179	85.3	0.87	18.02
3p2p	2.1	0.186	80.8	0.88	22.05	5tmn	1.6	0.171	87.3	0.69	13.96
3pcy	1.9	0.160	91.7	0.82	14.10	5tnc	2.0	0.155	92.5	0.73	19.83
3pep	2.3	0.171	82.2	0.72	18.60	5xia	2.5	0.147	90.6	0.75	14.64
3pfk	2.4	0.171	92.9	0.77	16.52	6acn	2.5	0.187	87.9	0.73	14.87
3pgk	2.5	-	59.6	1.42	29.93	6adh	2.9	-	67.6	0.94	28.44
3pgm	2.8	0.290	48.3	1.70	31.20	6at1	2.6	0.160	82.8	0.74	17.72
3ptb	1.7	0.182	86.8	0.83	11.66	6cha	1.8	0.200	81.8	0.94	18.54
3ptn	1.7	0.198	85.3	0.78	13.09	6cts	2.5	0.110	91.0	0.76	20.22
3rnt	1.8	0.137	90.9	0.71	17.37	6dfr	2.4	0.192	85.8	0.88	21.70
3rp2	1.9	0.191	84.8	0.80	17.58	6gch	2.1	0.180	82.0	0.84	20.59
3rxn	1.5	-	78.0	0.91	18.20	6ldh	2.0	0.202	84.7	0.86	20.43
3sgb	1.8	0.125	86.4	0.71	11.06	6lyz	2.0	-	84.3	0.66	20.26
3tln	1.6	0.213	89.7	0.69	16.23	6pad	2.8	-	73.3	1.02	21.83
3tmn	1.7	0.173	87.6	0.72	14.73	6pcy	1.9	0.152	88.1	0.83	18.89
3tpi	1.9	0.193	82.5	0.85	13.97	6pti	1.7	0.160	91.5	0.76	8.84
3ts1	2.7	0.210	94.0	0.88	14.63	6rsa	2.0	0.188	88.0	0.82	18.74
3wrp	1.8	0.204	93.5	0.74	17.86	6tmn	1.6	0.171	86.5	0.72	14.57
3xia	3.0	0.247	38.1	1.27	30.64	7adh	3.2	0.290	64.2	1.24	26.43
451c	1.6	0.187	89.9	0.72	12.64	7api	3.0	0.193	85.3	0.73	19.21
4ape	2.1	0.158	89.2	0.78	20.47	7at1	2.8	0.183	86.9	0.73	14.60
4at1	2.6	0.160	84.3	0.73	18.05	7cat	2.5	0.212	80.3	0.82	20.77
4cat	3.0	-	46.4	1.49		7dfr	2.5	0.245	77.7	1.00	20.93
4cha	1.7	0.234	87.0	0.83	16.81	7gch	1.8	0.190	83.5	0.98	23.36
4cpa	2.5	-	79.4	1.03	22.79	7lyz	2.5	-	79.1	0.69	21.45
4cpv	1.5	0.215	90.0	0.67	16.80	7pcy	1.8	0.117	90.1	0.72	15.84
4cts	2.9	0.182	81.2	0.80	20.85	7rsa	1.3	0.150	88.9	0.65	9.73
4dfr	1.7	0.155	89.2	0.73	15.64	7tln	2.3	0.170	86.8	0.72	16.57
4fab	2.7	0.215	68.6	0.89	27.20	7wga	2.0	0.172	78.0	0.97	16.79
4fd1	1.9	0.212	81.9	0.72	12.46	8adh	2.4	0.190	84.2	0.81	20.87
4fxn	1.8	0.200	86.0	0.78	17.50	8api	3.1	0.215	88.2	0.78	19.93
4gch	1.9	0.209	82.5	0.99	20.21	8at1	2.8	0.167	86.7	0.75	16.36
4gpd	2.8	0.218	53.2	1.19	26.90	8atc	2.5	0.165	84.8	0.86	15.21
4hhb	1.7	0.135	87.7	0.69	18.16	8cat	2.5	0.191	80.3	0.84	20.99
4hvp	2.3	0.176	87.8	0.85	20.08	8dfr	1.7	0.188	90.9	0.74	12.97
4ilb	2.0	0.190	82.4	0.76	21.58	8ldh	2.8	0.191	83.7	0.86	19.81
4ins	1.5	0.153	88.0	0.68	13.40	8lyz	2.5	-	83.5	0.69	19.84
4lyz	2.0	-	78.3	0.97	19.59	9api	3.0	0.209	87.9	0.77	16.33
4mba	2.0	0.180	94.6	0.65	15.35	9pap	1.6	0.161	87.4	0.77	17.62
4mbn	2.0	0.172	91.3	0.60	13.46	9wga	1.8	0.175	79.2	0.93	17.53

*For each parameter, Class 1 is denoted by a quarter circle, Class 2 a semicircle, Class 3 a ¾ circle, and Class 4 by a full circle. The class boundaries are listed in Table IVb.

TABLE IVb. Class Boundaries for the Classification in Table IVa

	Classification			
Parameter	1	2	3	4
ϕ,ψ distribution (%CORE)	>75%	>65%	>55%	<55%
χ_1 standard deviation (CHIDEV)*	$\leq -1\sigma$	$>-1\sigma \leq \mu$	$>\mu \leq +1\sigma$	$>+1\sigma$
Hydrogen bond energy (HBOND)	$\leq -1\sigma$	$>-1\sigma \leq \mu$	$>\mu \leq +1\sigma$	$>+1\sigma$

*μ is the sample mean. $-\sigma$ represents a boundary 1 standard deviation below the mean, $+\sigma$ a boundary 1 standard deviation above the mean.

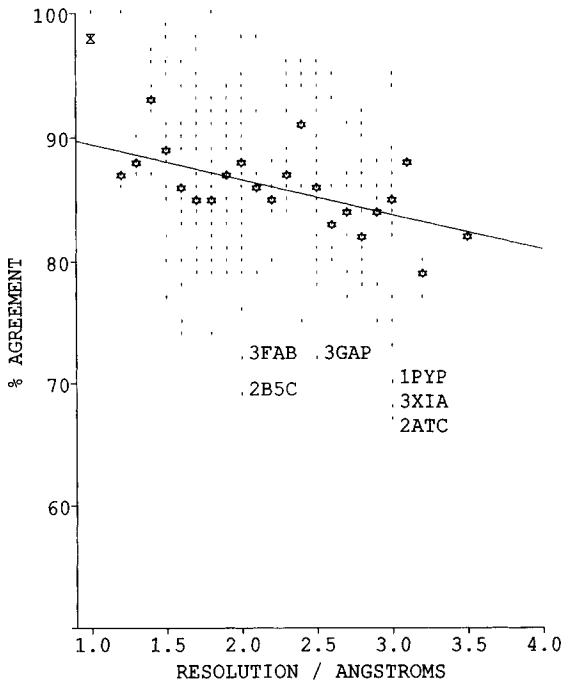


Fig. 15. Percentage agreement between Kabsch–Sander[17] and author defined secondary structures as a function of resolution. The secondary structural assignments were said to agree if the residue was defined as either α-helix or β-strand or not α-helix or β-strand, by both Kabsch–Sander and author criteria. The values plotted indicate overall percentage agreement for each protein (vertical dashes) and for each resolution (asterisk). The line was least squares fitted to the percentage agreements for each resolution (omitting the value for 1.0 Å resolution). The dataset includes 370 proteins. Proteins which had nonexistent or incomplete author secondary structure assignments or whose author assignments were made using the Kabsch–Sander criteria were omitted from the dataset.

of the accuracy of a set of coordinates and treat them accordingly. There are important implications in the use of a poor structure. For example, the Kabsch–Sander method[17] for assigning secondary structure has difficulty with low-resolution structures in which ϕ,ψ angles are poorly defined. Distorted hydrogen bonds are not recognized, so the secondary structure is difficult to identify accurately, especially toward both ends. This is illustrated by the graph in Figure 15, which shows the percentage agreement between Kabsch–Sander and author defined secondary structures (as given in the Brookhaven file) as a function of resolution. Clearly the secondary structures assigned by the authors will be subjective, depending on the method of assignment used. Nevertheless, the extent of the discrepancy for low-resolution data show that automated secondary structure assignments using the Kabsch–Sander method are resolution dependent.

TABLE Va. Matrix of Correlation Coefficients for Individual Structure Data

	1	2	3	4	5
1 (Resolution)	1.00	0.55	0.50	0.53	0.38
2 (Pooled χ_1)	0.55	1.00	0.63	0.69	0.53
3 (H bond energy)	0.50	0.63	1.00	0.76	0.51
4 (ϕ,ψ)	0.53	0.69	0.76	1.00	0.58
5 (R-factor)	0.38	0.53	0.51	0.58	1.00

TABLE Vb. Matrix of Correlation Coefficients for Resolution-Averaged Structure Data

	1	2	3	4
1 (Resolution)	1.00	0.96	0.89	0.85
2 (Pooled χ_1)	0.96	1.00	0.89	0.91
3 (H bond energy)	0.89	0.89	1.00	0.92
4 (ϕ,ψ distribution)	0.85	0.91	0.92	1.00

Table IVa also includes a simple coding scheme whereby the quality of the global parameters are indicated using circles and circle fractions. This allows a fast visual check for quality as defined by each criterion. It is notable that structures with poor ϕ,ψ distributions also show poor χ_1 standard deviations and so forth. The correlation coefficient between the percentage of residues in the CORE and the standard deviation of the χ_1 angles is 0.72, increasing to 0.90 if the data are averaged by resolution. The correlation coefficients for all the pairs of stereochemical parameters, including resolution and R-factor, are shown in Table V. For analysis purposes, the set of proteins to be included in a study will depend very much on the problem. For detailed atomic interactions, the standard criteria (resolution $\leq$ 2.0 Å, $R \leq 0.20$) should be extended to exclude structures with less than 75% of their residues in the CORE, and pooled χ_1 standard deviations

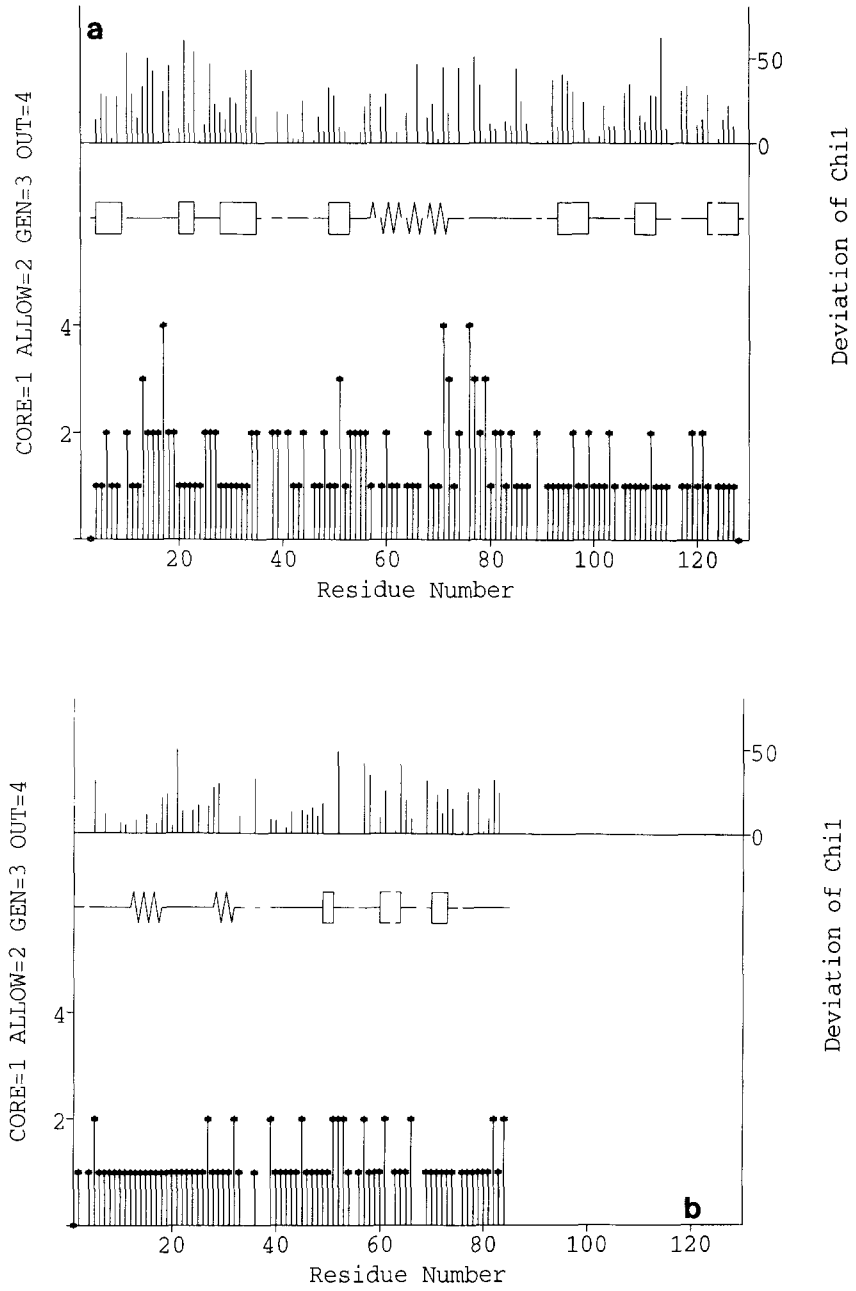


Fig. 16. (**a**) Plot of stereochemical and structural parameters with respect to residue number for Azurin, 1AZU[30] (2.7 Å, $R=0.35$). The bottom band (left hand y axis) shows the data for ϕ,ψ distribution; 1 is CORE, 2 ALLOWED, 3 GENEROUS, and 4 OUTSIDE. The middle band is a diagram of the secondary structure as defined by a modification of Kabsch–Sander (SSTRUC, Smith, private communication). The boxes represent β-strand, and the "saw-tooth," α-helix. The spaces are caused by the omission of proline and glycine. The top band (right-hand axis) shows the *absolute* deviation of the χ_1 angle from the closest energy well. This structure has rather poor stereochemistry as judged by Table IV. (**b**) Plot of the same parameters for the high potential iron protein 1HIP[31] (2.0 Å, $R=0.24$). The improved resolution is reflected in the improvement of both ϕ,ψ distribution, and χ_1 standard deviation. The residues with the highest deviation from closest minimum energy well are all either in loop regions, or at the termini of regular secondary structures.

greater than 16°. For secondary structure prediction, reasonably accurate ϕ,ψ values are required, and structures with fewer than 65% of their residues in CORE should be excluded. For topological studies only those structures which appear suspect (i.e., %CORE < 55%) need be omitted. It is significant

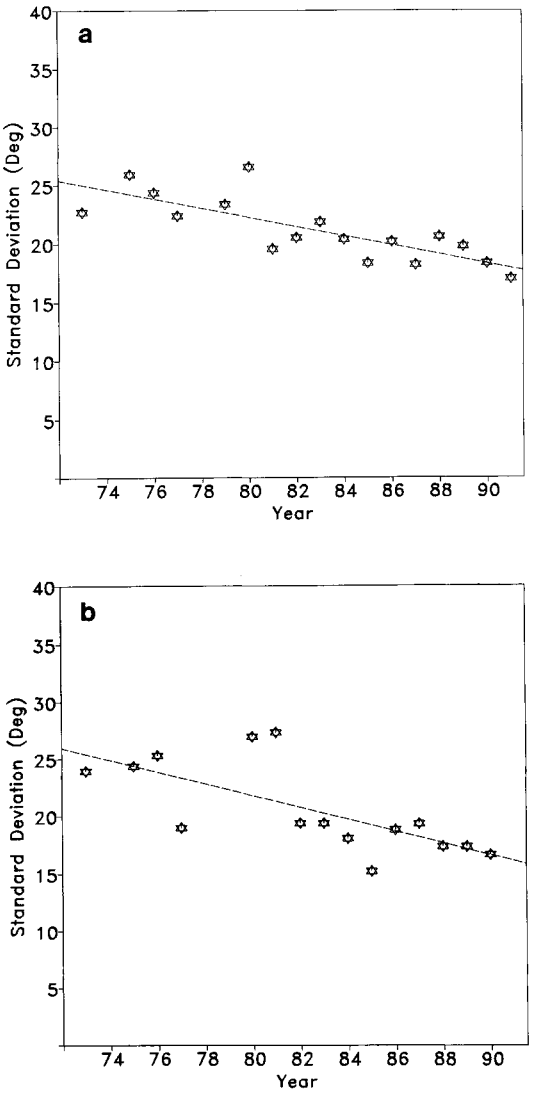


Fig. 17. (**a**) Structure quality improvement over time as suggested by the reduction in the values of the pooled standard deviations of the χ_1 angles. The symbols used are the same as in Figure 10a. (**b**) Structure quality improvement over time at 2 Å resolution as judged by the reduction in the pooled standard deviations of the χ_1 values. This may reflect the progressive improvement in refinement procedures.

that two coordinate sets which have recently been withdrawn from Brookhaven, 2MT2[28] and 2FD1,[29] had less than 20% of their residues in the CORE region, and widely scattered ϕ,ψ distributions.

IDENTIFYING REGIONS OF LOCAL DISTORTION

Crystallographic B values provide the best aid to identifying poorly defined segments of the polypeptide chain. If they are not available, as is common at low resolution, ϕ,ψ and χ_1 indicators can be plotted against residue number. For the ϕ,ψ distribution, the residues can be numbered according to whether they fall into the CORE(1), ALLOWED(2), GENEROUS(3), or OUTSIDE(4) region. When plotted along the chain, this gives a visual guide to the stretches of chain where there are several residues with poor ϕ,ψ values (Fig. 16a). For χ_1 the angular deviation from the ideal values derived from the set of high resolution structures (−66.7°, 64.1°, or 183.6°) is plotted. A diagram to represent the calculated secondary structure is also shown (Laskowski, private communication). Where poor main chain dihedrals are observed combined with poor χ_1 deviations, this is suggestive of a poorly determined segment of the polypeptide chain. The example shown is 1AZU.[30] This structure is solved to a resolution of 2.7 Å, $R = 0.35$. It can be seen that there are stretches of chain where several residues are either in the GENEROUS or OUTSIDE regions. While the majority of these are in nonregular secondary structure, some do appear to be in the middle of β strands. The lack of any pattern in the side chain angles is probably due to the low resolution of the structure. However, it must be remembered that even in well-refined high-resolution structures, some residues do adopt unusual conformations. This can be seen in Figure 16b, where the structure is high potential iron protein, 1HIP[31] (resolution 2.0 Å, $R = 0.24$). There are no main chain dihedrals outside the two most populated regions, and the overall deviation of the side chain χ_1 angles is much lower. The side chains with the greatest deviations of χ_1 correlate with regions of nonregular secondary structure, with one occurring at the C terminus of a strand. This method can only suggest which regions of the electron density map should be inspected by eye to check whether the interpretation is the best which can be made.

CONCLUSION

Several parameters can be derived from the coordinates of a protein structure which show some correlation with resolution, providing the refinement methods used do not include strong restraints on that parameter. These can be used to provide a simple guide as to which structures in the Brookhaven Protein Data Bank show large deviations from accepted values. It is important to stress that the *most* important indicator of reliability is the R-factor, and that a structure which has been carefully refined to $R \leq 0.2$ will generally have the correct topology. With the improvement in crystallographic and refinement techniques, the stereochemical quality of structures has improved in recent years, as shown in Figure 17a, plotting the standard deviation of pooled χ_1 values against the year of submission. This trend

is also apparent in the subset of structures determined to 2 Å resolution.

The parameters described here allow a simple means to identify those structures which fall outside a specific limit of stereochemical quality. They cannot prove that a structure is correct, they can indicate only possible errors in structures, or poorly resolved regions of a polypeptide chain. With the increasing use of energy-based programs, these simple parameters are inadequate. However, much better evaluation should be possible using energetic considerations which will combine stereochemistry, dihedral preferences, and noncovalent interactions. Despite the improvement in refinement methods, it is still possible for errors to occur, and these simple tests may provide a quick and easy method to avoid some of them.

Overall there have been remarkably few serious errors in crystallography determinations.[32] Given the complexity of protein structures, their determination still requires great expertise and care, and remains one of the great achievements of modern science.

ACKNOWLEDGMENTS

M.W.M. and A.L.M. are supported by Science and Engineering Research Council Studentships, E.G.H. by a Medical Research Council Training Fellowship. Our thanks to Malcolm McGregor, Roman Laskowski, and David Smith for programs; Steve Gardner for help with the database; Dr. D.S. Moss and Professor A.C.T. North for helpful discussions.

REFERENCES

1. 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. The Protein Data Bank: A computer based archival file for macromolecular structures. J. Mol. Biol. 122:535–542, 1977.
2. Blundell, T.L., Johnson, L.N. In "Protein Crystallography." London: Academic Press, 1976.
3. Brändén, C.-I., Jones, T.A. Between objectivity and subjectivity. Nature (London) 343:687–689, 1990.
4. Akrigg, D. (Protein Engineering Club Database Group). Nature (London) 335:745–746, 1988.
5. Islam, S.I., Sternberg, M.J.E. A relational database of protein structures designed for flexible enquiries about conformation. Prot. Eng. 2:431–442, 1989.
6. IUPAC-IUB Commission on Biochemical Nomenclature. Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains. J. Mol. Biol. 52:1–17, 1970.
7. MacArthur, M.W., Thornton, J.M. The influence of proline residues on protein conformation. J. Mol. Biol., 218:397–412, 1990.
8. Girling, R.L., Houston, T.E., Schmidt, Jr., W.C., Amma, E.L. Macromolecular structure refinement by restrained least-squares and interactive graphics as applied to sickling deer type III hemoglobin. Acta Crystallogr. A36:43–50, 1980.
9. Pierrot, M., Haser, R., Frey, M., Payan, F., Astier, J-P. Crystal structure and electron transfer properties of cytochrome c_3. J. Biol. Chem. 257:14341–14348, 1982.
10. McPhalen, C.A., James, M.N.G. Structural comparison of two serine proteinase-protein inhibitor complexes: Eglin-C-subtilisin Carlsberg and C1-2-subtilisin Novo. Biochemistry 27:6582–6598, 1988.
11. Stewart, D.E., Sarkar, A., Wampler, J.E. Occurrence and role of *cis* peptide bonds in protein structures. J. Mol. Biol. 214:253–260, 1990.
12. Ramachandran G.N., Ramakrishnan, C., Sasisekharan, V. Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7:95–99, 1963.
13. Jones, T.A., Thirup, S. Using known substructures in protein model building and crystallography. EMBO J. 5:819–822, 1986.
14. Holm, L., Sander, C. Database algorithm for generating protein backbone and side-chain co-ordinates from a C^{α} trace. Application to model building and detection of co-ordinate errors. J. Mol. Biol. 218:183–194, 1991.
15. Weaver, L.H., Tronrud, D.E., Nicholson, H., Matthews, B.W. Some uses of the Ramachandran (ϕ,ψ) diagram in the structural analysis of lysozymes. Current Sci. 59:833–837, 1990.
16. Yamashita, M.M., Almassy, R.J., Janson, C.A., Cascio, D., Eisenberg, D. Refined atomic model of glutamine synthetase at 3.5Å resolution. J. Biol. Chem. 264:17681–17690, 1989.
17. Kabsch W., Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen bonded and geometrical features. Biopolymers 22:2577–2637, 1983.
18. Janin, J., Wodak, S., Levitt, M., Maigret, B. Conformation of amino acid side-chains in proteins. J. Mol. Biol. 125: 357–386, 1978.
19. Ponder, J.W., Richards, F.M. Tertiary templates for proteins. J. Mol. Biol. 193:775–791, 1987.
20. McGregor, M., Islam, S., Sternberg, M.J.E. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. J. Mol. Biol. 198:295–310, 1987.
21. Murthy, M.R.N., Reid, T.J., III, Sicignano, A., Tanaka, N., Rossmann, M.G. Structure of beef liver catalase. J. Mol. Biol. 152:465–499, 1981.
22. Thornton, J.M. Disulphide bridges in globular proteins. J. Mol. Biol. 151:261–287, 1981.
23. Richardson, J.S. The anatomy and taxonomy of protein structure. Adv. Protein Chem. 34:167–339, 1981.
24. Hendrickson, W.A., Konnert, J.H. Incorporation of stereochemical information into crystallographic refinement. In: "Computing in Crystallography." Diamond, R., Ramaseshan, S., Venkatesan, K., eds. Bangalore: Indian Academy of Sciences, 1980: 13.01–13.23.
25. Jack, A., Levitt, M. Refinement of large structures by simultaneous minimization of energy and R factor. Acta Crystallogr. A34:931–935, 1978.
26. Tronrud, D.E., Ten Eyck, L.F., Matthews, B.W. An efficient general-purpose least-squares refinement program for macromolecular structures. Acta Crystallogr. A43: 489–501. 1987.
27. Brünger, A.T. Crystallographic refinement by simulated annealing. Application to a 2.8Å resolution structure of aspartate aminotransferase. J. Mol. Biol. 203:803–816, 1988.
28. Furey, W.F., Robbins, A.H., Clancy, L.L., Winge, D.R., Wang, B.C., Stout, C.D. Crystal structure of Cd, Zn metallothionein. Science 231:704–710, 1986.
29. Ghosh, D., O'Donnell, S., Furey, Jr., W., Robbins, A.H., Stout, C.D. Iron-sulfur clusters and protein structure of *Azotobacter* ferredoxin at 2.0Å resolution. J. Mol. Biol. 158:73–109, 1982.
30. Adman, E.T., Jensen, L.H. Structural features of azurin at 2.7Å resolution. Isr. J. Chem. 21:8–12, 1981.
31. Gwynne, J., Brewer, Jr., B., Edelhoch, H. The molecular properties of apoA-I from human high density lipoprotein. J. Biol. Chem. 249:2411–2416, 1974.
32. Thornton, J.M., MacArthur, M.W., Smith, D.K., Gardner, S.P., Hutchinson, E.G., Morris, A-L., Sibanda, B.L. In: "Accuracy and Reliability of Macromolecular Structures." Henrick, K., Moss, D.S., Tickle, I.J., eds. SERC, 1990: 39–52.
33. Weaver, L.H., Gray, T.M., Grutter, M.G., Anderson, D.E.,

Wozniak, J.A., Dahlquist, F.W., Matthews, B.W. High-resolution structure of the temperature sensitive mutant of phage lysozyme, Arg 96→His. Biochemistry 28:3793–3797, 1989.

34. Rossmann, M., Arnold, E., Erickson, J.W., Frankenberger, E.A., Griffith, J.P., Hecht, H-J., Johnson, J.E., Kamer, G., Luo, M., Mosser, A.G., Rueckert, R.R., Sherry, B., Vriend, G. Structure of a human common cold virus and functional relationship to other picornaviruses. Nature (London) 317: 145–153, 1985.
35. Cahn, R.S., Ingold, C.K., Prelog, V. The specification of assymmetric configuration in organic chemistry. Experientia 12:81–124, 1956.