J. Mol. Biol. (1978) **125**, 357–386

Conformation of Amino Acid Side-chains in Proteins

JOËL JANIN, SHOSHANNA WODAK

Unité de Biochimie Cellulaire, Département de Biochimie et Génétique Microbienne
Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France

MICHAEL LEVITT† AND BERNARD MAIGRET‡

Medical Research Council Laboratory of Molecular Biology, Hills Road
Cambridge CB2 2QH, England

(Received 28 April 2978, and in revised form 13 July 1978)

We have analysed the side-chain dihedral angles in 2536 residues from 19 protein structures. The distributions of χ_1 and χ_2 are compared with predictions made on the basis of simple energy calculations. The χ_1 distribution is trimodal; the g^- position of the side-chain (*trans* to Ha), which is rare except in serine, the t position (*tram* to the amino group), and the g+ position (*tram* to the carbonyl group), which is preferred in all residues. Characteristic χ_2 distributions are observed for residues with a tetrahedral γ-carbon, for aromatic residues, and for aspartic acid/asparagine. The number of configurations actually observed is small for all types of side-chains, with 60% or more of them in only one or two configurations. We give estimates of the experimental errors on χ_1 and χ_2 (3" to 16°, depending on the type of the residue), and show that the dihedral angles remain within 15° to 18° (standard deviation) from the configurations with the lowest calculated energies. The distribution of the side-chains among the permitted configurations varies slightly with the conformation of the main chain, and with the position of the residue relative to the protein surface. Configurations that are rare for exposed residues are even rarer for buried residues, suggesting that, while the folded structure puts little strain on side-chain conformations, the side-chain positions with the lowest energy in the unfolded structure are chosen preferentially during folding.

1. Introduction

In the study of protein three-dimensional structures, attention has been focused on the conformation of the polypeptide chain. Though protein folding is known to be determined by the amino acid sequence, that is by the chemical nature of the side-chains, the side-chains themselves have been considered only from the point of view of their effect on the main-chain conformation, The classical work of Ramachandran and his group established that the presence of a side-chain (any side-chain) reduces considerably the conformational possibilities of the neigh bouring peptide groups, characterized by the ϕ and ψ dihedral angles. These results have been extended to show particular effects of the various types of side-chains (for a review, see Nemethy & Scheraga, 1977) and of their conformation (Finkelstein & Ptitsyn, 1977). The actual configurations taken by the side-chains in protein structures have not been analysed, except for the work of Chandrasekaran & Ramachandran (1970), based on three

† Present address: The Salk Institute, Post Office Box 1809, San Diego, Calif. 92 113, U.S.A.
‡ Present address : Lnborat oire de Biophysique, Université dc Nancy I, Centre de ler Cycle, Case Officielle no. 140, 54037 Nancy Cedex, France.

0022-2836/78/310357-30 $02.00/0

protein structures (myoglobin, lysozyme and α-chymotrypsin). These authors analysed the distribution of the side-chain dihedral angles and attempted to prove that it can be predicted on the basis of simple geometric considerations.

Our approach is similar to theirs. We compare the conformations taken by side-chains in globular proteins to the results of energy calculations using van der Waals' interactions (and steric hindrance) only. The experimental data, which include 2536 side-chain not counting glycine, alanine and proline residues, from 19 high-resolution protein structures, show that the number of configurations accessible to each type of side-chain is small. These con-rations are independent of the position of the residue on the surface or inside the protein. Their relative frequencies are affected to a certain extent by the secondary structure, which slightly perturbs side-chain conformations through steric hindrance and, in serine and threonine residues. through side-chain to main-chain hydrogen bonds. In the folded protein structure, long-range interactions with residues far away in the amino acid sequence select one of the few permitted configurations without perturbing it strongly: protein folding causes little strain on the side-chains.

2. Materials and Methods

(a) *Atomic co-ordinates*

Atomic co-ordinates for 28 protein structures were obtained from the Protein Data Bank, Cambridge, England. A subset of 19 proteins containing high-resolution (2·5 Å or better) structures, many of which have been submitted to crystallographic refinement, is selected from the list. These structures are: egg white lysozyme, carboxypeptidase A, subtilisin BPN', bovine pancreatic trypsin inhibitor, parvalbumin (calcium binding protein), papain, high potential iron protein, insulin, thermolysin, *Clostridium* flavodoxin, elastase, horse methemoglobin αβ dimer, V_{REI} immunoglobin fragment, ferredoxin, ferricytochrome *c*2, ferricytochrome *b*5, bacteriophage T4 lysozyme, β-trypsin, α-chymotrypsin. Amino and carboxyl terminal residues and residues poorly resolved in electron density maps (when mentioned by the authors) were removed. The sample includes 3261 amino acid residues, of which 725 are Gly, Ala and Pro.

Bond lengths and bond angles have standard values in all sets of atomic co-ordinates of this sample, with small variations between authors. Therefore, the dihedral angles (ϕ, ψ for the main chain, χ for the side-chain) are sufficient to characterize the conformation. Care is taken in their calculation that they conform with the IUPAC-IUB (1970) convention. However, the χ angles are in the range of 0° to 360° rather than −180° to 180°.

(b) *Side-chain geometry and energy calculations*

During rotation of the χ angles, steric hindrance resulting from overlaps between atoms create potential energy barriers which effectively forbid certain conformations of the side-chain. The main chain is the principal source of steric hindrance and it affects the χ_1 angle governing rotation around the C_α—C_β bond, more than other side-chain angles. We use energy calculations to define permitted configurations taking account of steric interactions made with the peptide groups preceding and following the residue. The calculations are performed on the $CH_3CONH\text{-}RC_\alpha H\text{-}CONHCH_3$ structure, where R is the side-chain considered. The potential energy is:

$$E = \sum_{ij} (A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^{6}) + \sum_{t} K \cos n(\omega_t - \omega_0). \quad (1)$$

It includes only the Lennard–Jones potential representing van der Waals' interactions between non-bonded atoms as a function of their distance r_{ij}, and a torsion potential for each dihedral angle ω_t. A torsion potential has to be included, particularly since the hydrogen atoms are treated together with the non-hydrogen atoms to which they are attached (Gibson & Scheraga, 1967; Levitt & Lifson, 1969). The energy parameters of eqn (1) are taken from Levitt (1974). We use Levitt's energy refinement program to calculate the value of E for various combinations of the main chain and side-chain dihedral angles. Electro-

static interactions are ignored in this study. Their influence on the potential energy depends strongly on assumptions made regarding the dielectric constant.

(i) *The χ_1 angle*

Rotation of the side-chain around the C_α—C_β bond is restricted, due to contacts made by atoms in γ position with the chemical groups attached to the α-carbon (Fig. 1 (a)). These groups axe the preceding and following peptides, the position of which is determined by the main chain ϕ and ψ dihedral angles, and the α-hydrogen atom. Overlaps with the peptide groups axe least severe when ϕ and ψ are in the allowed regions of the Ramachandran map (Ramachandran *et al.*, 1963), but they are still dominant over the effect of H_α. Therefore, the g^- position of the y-atom *trans* to H_α should be less favourable than t and g^+ positions *tram* to the more bulky amino and carbonyl groups, respectively.

Energy calculations confirm this qualitative description of the local geometry of the side-chain. We summarize in Fig. 2 tho effect of a y-methyl group on the allowed regions of the Ramachandran diagram. The $\phi\psi$ energy map calculated when the y-methyl is *trans* to the carbonyl group (g^+ position, Fig. 2(c)) is essentially the same as for an alanine residue (the energy function being that of eqn (1)), showing that the y-carbon creates little steric hindrance over that of the β-carbon. *Trans* to the amino group (t position, Fig. 2(b)), the γ-atom restricts the rotation of the carbonyl group (ψ angle): the corresponding Ramachandran diagram has smaller helical regions. *Trans* to H_α (g^- position,

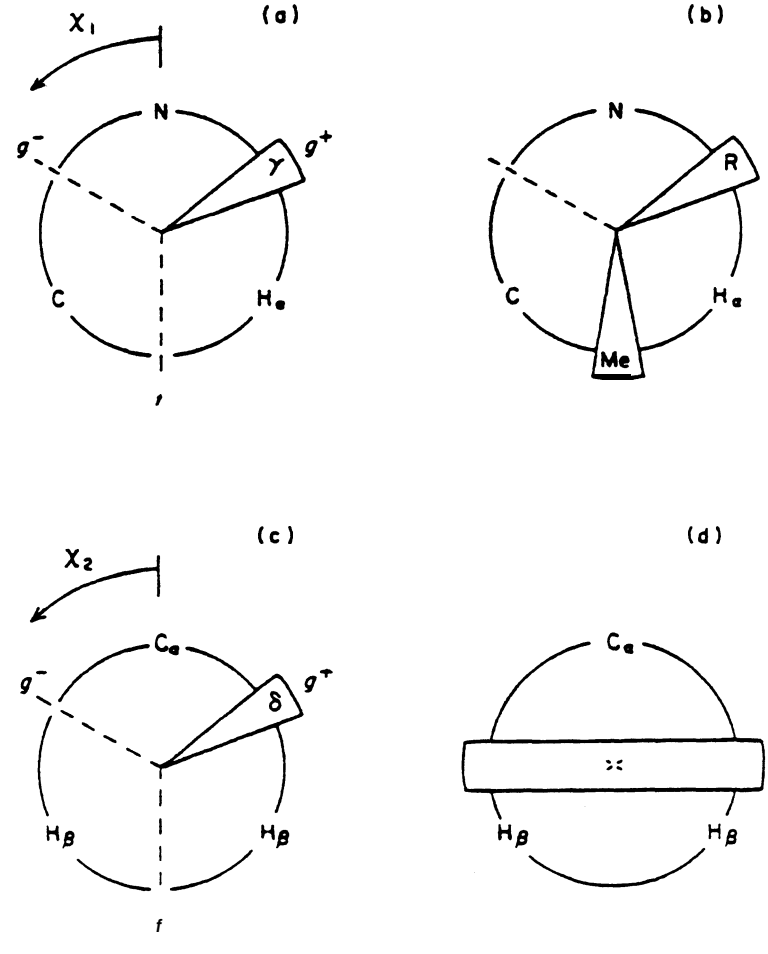


FIG. 1. The geometry of rotation around the C_α—C_β and C_β—$C\gamma$ bonds. The definitions of χ_1 in (a) and of χ_2 in (c) conform to the IUPAC-IUB convention (1970). The side-chain is represented above the plant! of the C_α substituents in Newman projections *down* the C_β—C_α bond. Idealized g^- ($\chi_1 = 60°$), t ($\chi_1 = 180°$) and g^+ ($\chi_1 = 300°$) positions are indicated for the γ and δ atoms. These positions are named respectively I, II and III by Ramachandran & Chandrasekaran (1970).

(a) Unbranched C_β. (b) Branched C_β ; R is CH_3(Me) in Val. C_2H_5 in Ile, OH in Thr. (c) Tetrahedral C_γ. (d) Trigonal C_γ.

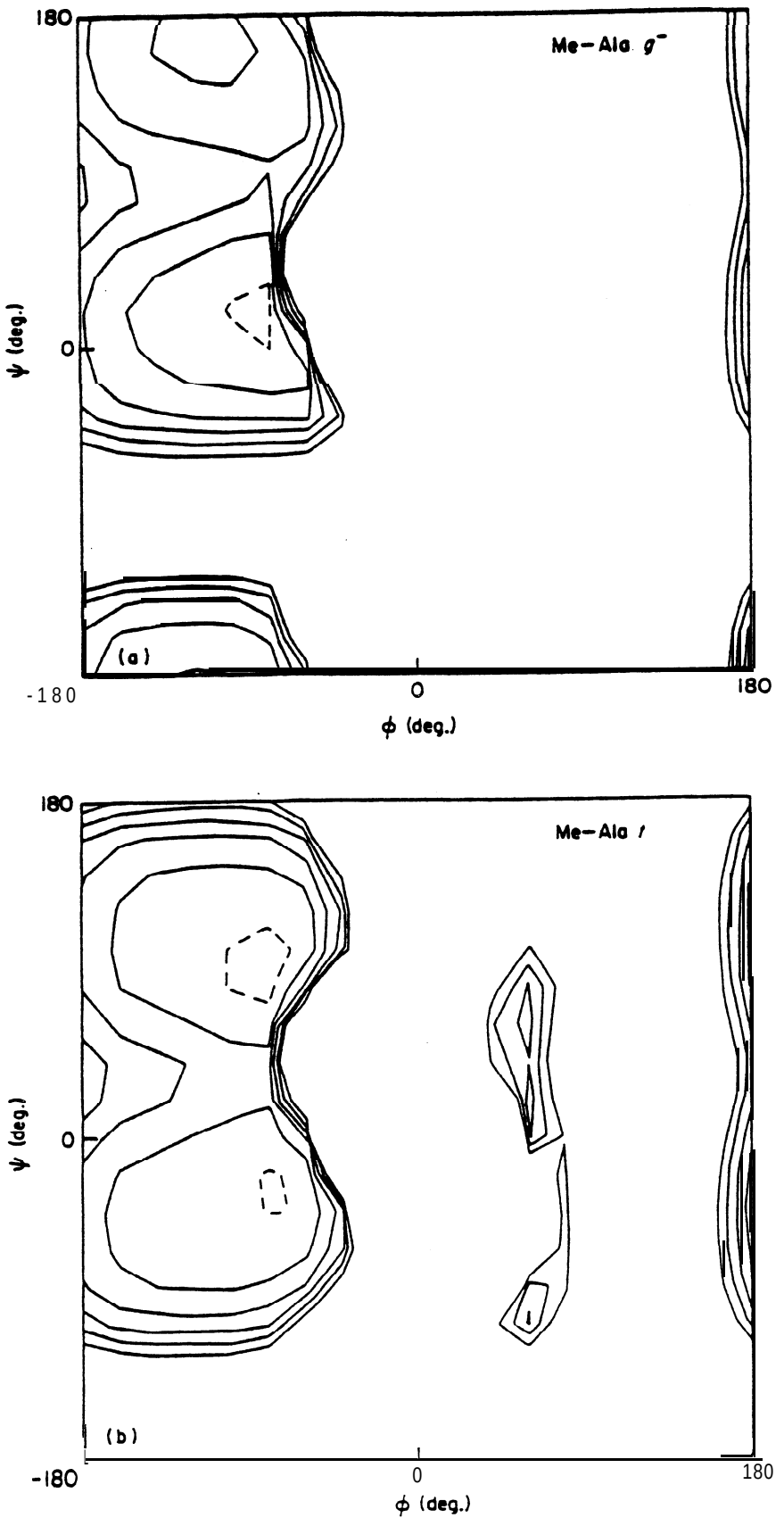


FIG. 2. $\phi\psi$ energy maps for methyl-alanine. The energy of the $CH_3CONH\text{-}(C_2H_5)C_\alpha H\text{-}CONHCH_3$ structure (an a-amino-butyryl or methyl-alanilyl residue with blocked N and C) is calculated as described in Materials and Methods for different values of the main-chain dihedral angles ϕ and ψ, and for 3 values of the side-chain dihedral angle χ_1. Contours are drawn at. -2 (broken), 0, 2, 4, 6 and 8 kcal/mol.

(a) The γ-methyl group is in the g^- position; $\chi_1 = 60°$. (b) The γ-methyl group is in the t position; $\chi_1 = 180°$. (c) The γ-methyl group is in the g^+ position ; $\chi_1 = 300°$.

Pig. 2(a)), the y-atom restricts the range of both ϕ and ψ; the left-handed helical conformations are effectively forbidden, the α_L and β regions of the map are reduced in size compared to the g+ or alanine $\phi\psi$ map.

Qualitatively similar results *have* been obtained by Finkelstein (1976) using Courtauld atomic mod&. Our calculations apply to residues with carbon atoms in y-position. Sulphur (Cys) and oxygen atoms (Ser, Thr) have lesser effects (Tonnuswamy & Sasisekharan, 1971*a*). For side-chains with branched C_β atoms (Ile, Val, Thr) the energy barriers created by the two γ-substituents add up: thus, in Val residues, the $\phi\psi$ energy map is that of Fig. 2(b) when the two methyl groups are in g^+ and *t* positions as shown on Fig. l(b), and that of Fig. 2(a) otherwise.

(ii) *The* χ_2 *angle*

The geometry of rotation around the C_β—C_γ bond is different when the y-carbon is tetrahedral, or when it is trigonal and planar (aromatic residues; His, Asp and Asn).

In the first case, the δ atom may occupy the 3 usual positions relative to C_α: g^-, *t* and g^+, corresponding to $\chi_2 = 60°$, 180° and 300°, respectively (Fig. 1(c)). As the steric hindrance due to β-hydrogen atoms is small compared to that of C_α and of the main chain, the *t* position *trans* to C_α is expected to be preferred. Overlaps with main-chain atoms restrict the permitted values of the χ_2 angle depending on χ_1 (Ponnuswamy & Sasisekharan, 1971*b*; see also Fig. 7(b)).

When the γ-carbon is trigonal, overlaps of the δ atoms with the main chain axe least severe for values of χ_2 near 90° or 270° (Fig. 1(d)). The side-chains of Phe, Tyr and Asp residues have 2-fold symmetry, which means that positions 180° apart in χ_2 am equivalent. In His and Asn this exchanges nitrogen and carbons atoms, with little influence on steric hindrance. In Trp the side-chain has no symmetry.

(iii) Other aide-chain angles

The geometry of the χ_3 angle in Lys. Arg and Met, and that of χ_4 in Lys, is similar to that of χ_2 in the tetrahedral C_γ case. In Glu and Gln the geometry of χ_3 is similar to that of χ_2 in Asp and Asn.

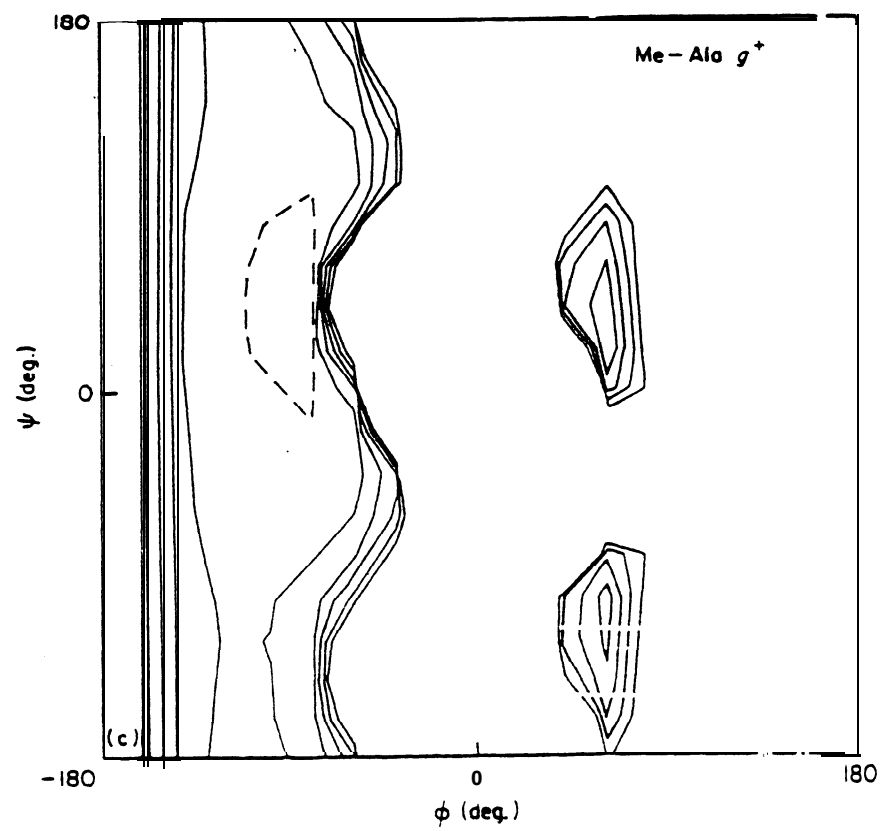


FIG. 2(c).

(c) *Solvent accessibility*

The accessibility to the solvent is estimated from the accessible surface area of each residue in the protein structure. This is defined by Lee & Richards (1971) as the area of a surface over which a water molecule (taken to be a sphere of radius 1·5 Å for the calculation) can be placed so that it makes contact with an atom of the residue without penetrating any other atom of the structure. We use a program of Levitt to compute accessible surface areas from X-ray co-ordinates.

Residues are then classed as:
-buried if their accessible surface area A is smaller than 20 Å^2,
—exposed if A is larger than 60 Å^2,
—intermediate if A is between 20 and 60 Å^2.

This classification is based on average values of A calculated by Chothia (1976) in 12 protein structures, and on our own work on 28 structures (Table 1). Residues taken as buried are wholly inside the protein (A = 0) or nearly so, while exposed residues have most of their side-chain free in the solvent. The sample is distributed about equally between the 3 classes (38% buried? 30% intermediate and 32% exposed), but the amino acid composition of the classes is very different (Table 1). As noted by Chothia (1976), the average value of A and the fraction of buried residues shows little correlation with the size of the side-chain (though the limit of 60 Å^2 chosen here is slightly too large for Gly residues), but it does depend strongly on the presence of oxygen or nitrogen atoms in the side-chain.*.

TABLE 1

Amino acid composition and accessibility to solvent

Residue	Number in† sample 1	Number in† sample 2	Average A‡ (Å^2)	Buried (%)§	Exposed (%)§
Gly	311	453	24·5	52	10
Ala	297	435	27·8	51	15
Pro	117	209	51·5	25	45
Leu	237	389	27·6	60	16
Ile	167	261	22·8	66	13
Val	238	401	23·7	64	14
Met	34	69	33·5	52	20
Cys	84	111	15·5	74	5
Phe	106	177	25·5	58	10
Tyr	155	217	55·2	24	41
Trp	59	91	34·7	49	17
His	65	122	50·7	34	34
Ser	286	459	42·0	35	32
Thr	207	325	45·0	30	32
Asp	165	305	60·6	19	50
Asn	183	272	60·1	22	49
Glu	126	220	68·2	16	55
Gln	141	200	68·7	16	56
Lys	183	341	103·0	3	85
Arg	101	154	94·7	5	67
Total	3261	5211	45·5	38	32

† Sample 1 contains the 19 protein structures mentioned in the text. Sample 2 contains the following 9 structures in addition to those of sample 1: rubredoxin, ribonuclease S, staphylococcal nuclease, concanavalin A (Argonne and Rockefeller structures), lobster glyceraldehyde-3-phosphate dehydrogenase, dogfish lactate dehydrogenase, *carbonic* anhydrase B and C. All atomic co-ordinates are obtained from the Cambridge Data Bank.

‡ Average of the accessible surface areas of individual residues in sample 2.

§ Percentage of residues having an accessible surface area A smaller than 20 Å^2 (buried) or larger than 60 Å^2 (exposed).

3. Results

(a) *How well are the side-chain conformations determined?*

Experimental data on dihedral angles are derived from X-ray studies of protein structures. A comparison with calculated values is meaningful only if the data are precise. Because side-chain angles affect the position of a limited number of atoms only, they might be much less reliable than main-chain dihedral angles, which affect the whole structure and have been the object of more attention. For that reason, we choose to restrict the sample of protein structures on which side-chain angles are estimated, to proteins for which high-resolution (2·5 Å or better) crystallographic data are available. Most of these have been submitted to some sort of crystallographic refinement, which has been demonstrated to improve the precision of atomic positions. Estimates in the range of 0·15 to 0·5 A have been given for the standard deviation of refined positions and they imply that dihedral angles are known to within 6° to 20°.

The best way of assessing the quality of the χ angle data is to compare multiple measurements of the same angles, done in homologous protein structures which have been refined independently. Table 2 shows the result of such a comparison. Homologous residues in two (or three in the case of trypsin, elastase and chymotrypsin) protein structures are compared. If their side-chain configurations are the same as judged from the range of χ_1 and χ_2 angles, the standard deviation of the angular values can be calculated. If they differ too much (by 120° or so), the side-chain configurations are different and the comparison is not meaningful.

Table 2 demonstrates, not surprisingly, that χ angles are best determined in aromatic residues. The configuration of aroma tic side-chains is unam biguously defined by electron density maps and it is nearly always conserved between homologous structures: only one aromatic side-chain out of 36 changes conformation. The standard deviation of the χ_1 and χ_2 angles for the 35 remaining residues is 3° to 5·9°. It includes experimental errors and small changes in the structures, especially in the comparison of the three proteases. Thus, the upper value (5·9 A) is probably too large. On the other hand, the comparison of the two monomers in the V_{REI} immunoglobin light chain dimer, and of the trypsin structure with the trypsin–trypsin inhibitor complex, may be biased towards lower values of the standard deviation, because their crystallographic refinements start from the same initial models.

The side-chain conformations of Met, Glu, Gln, Arg and Lys residues also show a reasonable degree (70%) of conservation in homologous structures. These side-chains (except Met) are commonly found on the protein surface and are most affected by the molecular environment in the protein crystal. Still, their γ and δ atoms are reasonably well-positioned in electron density maps, even when the polar end of the side-chain is free in the solvent. This can be seen from the standard deviation of the χ_1 and χ_2 angles, which is 8° to 15°. Similar standard deviations affect the dihedral angles of Leu side-chains, but their conformation is less well-conserved between homologous structures.

The χ_1 angle of Ile and Val residues appears to be well-determined (standard deviation 7° to 13°) and well-conserved (90%) between homologous structures. This is less true of Ser and Thr residues, where one-third of the conformations change and where the standard deviation is larger (9° to 16°). In Ser residues, the position of the hydroxyl group is often not defined in electron density maps. Errors in the interpretation of the map may also be the source of some of the changes of χ_1 angles in Ser, Thr or Val and of χ_2 angles in Leu.

TABLE 2

The precision of the χ angle measurements

Residues	Number of homologous positions: Total	With conserved χ_1	With conserved χ_2	Standard deviation (deg.) σ_1	σ_2
Met, Glu, Gln, Lys, Arg					
V_{REI}	14	12	9	14·4	11·2
Trypsin-TIC	32	26	23	12·2	12·9
Serine proteases	9	8	6	12·2	8·0
Phe, Tyr, Trp					
V_{REI}	12	12	11	3·3	3·1
Trypsin-TIC	17	17	17	3·5	5·1
Serine proteases	7	7	7	5·7	5·9
Leu					
V_{REI}	7	6	5	10·5	4·8
Trypsin-TIC	14	14	8	8·4	5·1
Serine proteases	6	4	1	13·3	11·0
Ile, Val					
V_{REI}	11	11	—	12·9	—
Trypsin-TIC	32	29	—	7·5	—
Serine proteases	13	10	—	7·3	—
Ser, Thr					
V_{REI}	25	12	—	9·3	—
Trypsin-TIC	31	27	—	10·5	—
Serine proteases	12	9	—	16·0	—

We compare the side-chain conformations of homologous residues in the 2 monomers of the V_{REX} immunoglobulin light chain fragment (Epp *et al.*, 1974), in trypsin free (Bode & Schwager, 1975) or in complex with the bovine pancreatic trypsin inhibitor (TIC; Huber *et al.*, 1974), and in 3 serine proteases: trypsin, chymotrypsin and elastase. The proteases have rather different sequences, but we take as equivalent residues belonging to one of the 5 classes listed in the Table (for instance Ile and Val residues). if they occur at homogous positions.

For each pair of triplet of homologous residues? the χ angles are compared. If the individual values are within 60° of the average, the conformation is taken to be maintained: we calculate the standard deviation from the multiple measurements of χ (root-mean-square deviation from the average; a random value would be 30° in this case).

In the trypsin–trypsin inhibitor complex, about 1/3 of the Ser hydroxyl groups are not positioned. The corresponding χ_1 angles are ignored in the statistics. Similar cases certainly occur in other protein structures, though published atomic co-ordinates do not necessarily mention it.

The average values $\bar{\chi}$ and standard deviations σ of the dihedral angles are calculated from the formula:

$$\cos\sigma \exp i\bar{\chi} = \frac{1}{N}\sum_{1}^{N} \exp i\chi \quad (i = \sqrt{-1}),$$

which takes their periodicity into account and reduces to usual forms when σ is less than about 20°.

(b) *The χ_1 angle*

(i) *The general case*

In Trp, Tyr, Phe, His, Met, Leu, Asp, Asn, Glu, Gln, Lys and Arg residues, the geometry of the rotation around the C_α—C_β bond is identical, and the experimental data concerning the χ_1 angle can be merged. Figure 3 shows that the distribution of χ_1 is trimodal. The mean values $\overline{\chi_1}$ and standard deviations σ_1 in each range of 120° are:

$\overline{\chi_1} = 61^0$, $\sigma_1 = 25°$ around the g^- position,

$\overline{\chi_1} = 190°$, $\sigma_1 = 24°$ around the t position,

$\overline{\chi_1} = 290°$, $\sigma_1 = 21°$ around the g^+ position.

The position of the maxima, which in the t and g^+ positions are displaced from the

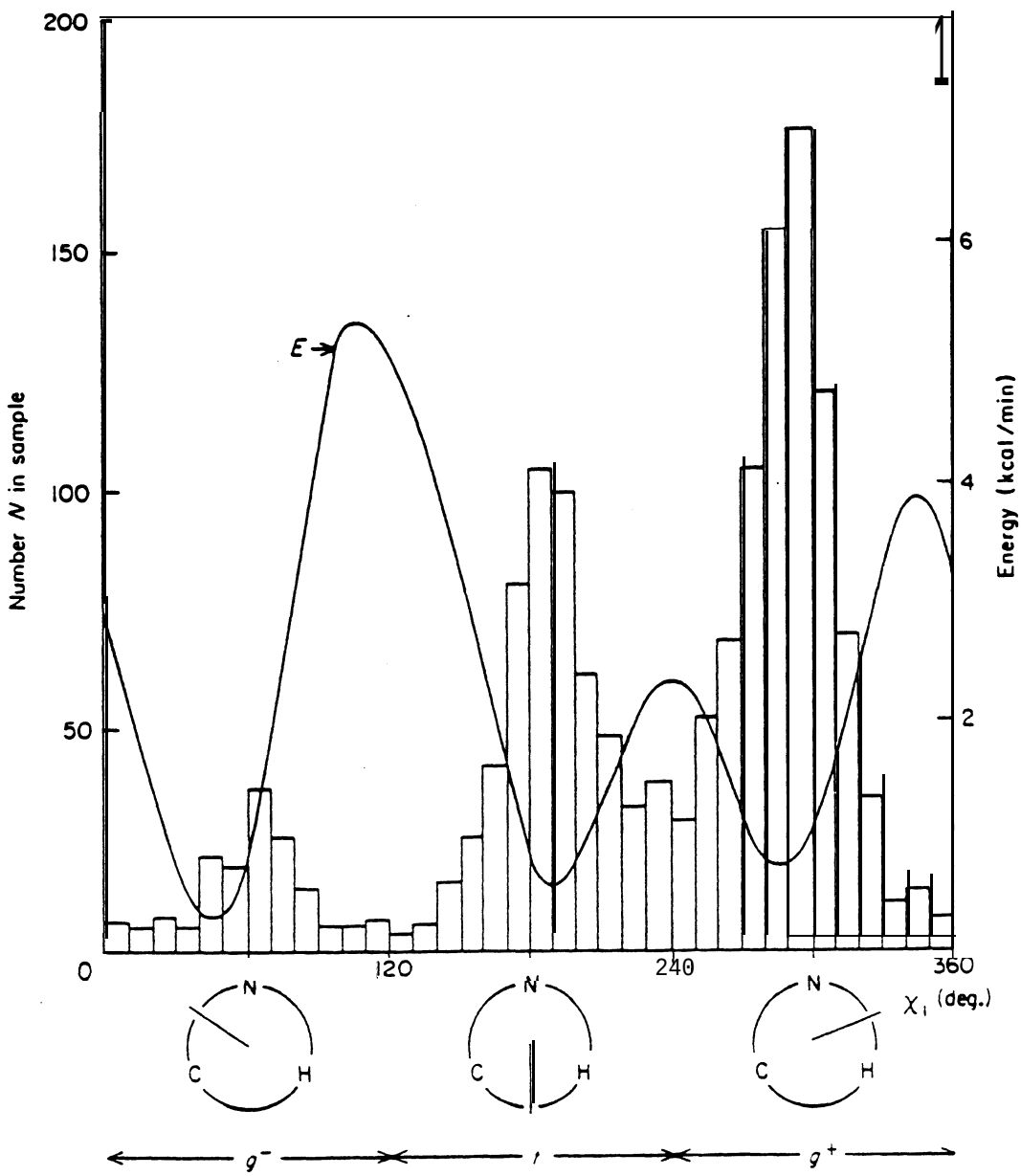


FIG. 3. The χ_1 angle distribution in Trp, Tyr, Phe, His, Met, Leu, Asp, Asn, Glu, Gln, Lys and Arg. The experimental distribution of χ_1 in 1556 side-chins is plotted along with the energy calculated for a blocked Lys residue (E) in an extended main-chain conformation ($\phi = -140°$, $\psi = 140°$; see Fig. i(b)). The side-chain is fully extended ($\chi_2 = \chi_3 = \chi_4 = 180°$) as it rotates around the C_α—C_β bond.

values 180° and 300°, and the relative height of the barriers at 0°, 120° and 240°, are predicted correctly of the basis of van der Waals' energies. Less than 6% of the side-chains have χ_1 angles in the range (−30°, 40°) where C_γ overlaps with the amino group of the residue, or (80°, 140°) where it overlaps with the carbonyl group. Overlaps with the *a-hydrogen* ($\chi_1 \sim 240°$) are more common. Thus, the experimental distribution of the χ_1 *angles* is restricted to less than two-thirds of the 360° range.

The side-chains are unequally distributed among the g^-, *t* and g^+ positions. More than half of the side-chains (54%) are found to be in the g+ configuration, the proportion varying from 44% (Gln) to 60% (Leu), depending on the residue type. The fraction of residues in the g^- configuration is very small, 11% on the average, varying from 4% (Leu) to 15% (Asp, Asn). The distribution is therefore somewhat sensitive to the nature of the side-chain and we shall see that it is affected by the conformation of the main chain.

(ii) *Branched β-carbon (Val, Ile)*

The distribution of χ_1 angles in Val and Ile residues is shown in Figure 4. Taking into account the different definitions of χ_1 chosen for these side-chains by the IUPAC-IUB convention, the two distributions are identical. They show a strong preference for a single position (67% of the residues) where one of the C_γ is in g^+ position and the other in *t* position. The corresponding peak is centered at $\overline{\chi_1} = 171°$ (Val) or 291° (Ile) with a standard deviation $\sigma_1 = 21°$. It corresponds to the g+ peak in unbranched side-chains. The *t* and g^- positions are much rarer, because one of the C_γ is then in the unfavourable position *trans* to the a-hydrogen (Fig. 1 (b)).

The proportion of Ile and Val side-chains found in the g+ configuration is probably underestimated, due to errors in the interpretation in the electron density map, which tend to randomize the distribution. In the small sample of side-chains where multiple observations are available, the proportion is as high as 80%: 8 g+ out of 10 homologous Ile and Val residues of trypsin, elastase and chymotrypsin; 8 out of 11 in the two monomers of the V_{REI} immunoglobulin fragment; 24 out of 29 in trypsin *versus* the trypsin–trypsin inhibitor complex. On the other hand, in the extended sample 2 of Table 1, where experimental errors are presumably higher, the proportion of g^+ is only 60%. We may then assume that the actual fraction of the Ile and Val side-chains in g^+ position is somewhere between 67% and 80%. The latter value is also the fraction of g+ side-chains observed in a sample of 39 small crystal structures containing blocked Val residues (reviewed by Benedetti, 1977).

(iii) *Threonine*

The χ_1 angle distribution (Fig. S(a)) is bimodal with the hydroxyl group either in g^+ (48%) or g^- (39%) position, the more bulky methyl group being in *t* and g^+ positions, respectively. The *t* configuration (with the methyl group in g^-) is rare (13%); arguments similar to those given for the Ile and Val side-chains suggest that its frequency is somewhat overestimated in these statistics. Correlations with hydrogen bonding patterns are discussed later.

(iv) *Serine*

The Ser side-chain contrasts with all other side-chains. While the χ_1 angle distribution (Fig. 5(b)) is still clearly trimodal, the g^-, *t* and g^+ positions are occupied with similar frequencies (38%, 28% and 34%, respectively). This reflects the lesser steric

84 Cys residues in sample 1, and 24 out of 111 in sample 2, are free -SH residues. The χ_1 angle distribution (Fig. 5(c)) is probably representative of substituted Cys side-chains and not of cysteine. The relative frequencies of the g^-, t and g^+ positions of the S_γ atom (16%, 27% and 57%, respectively) are similar to those of unbranched

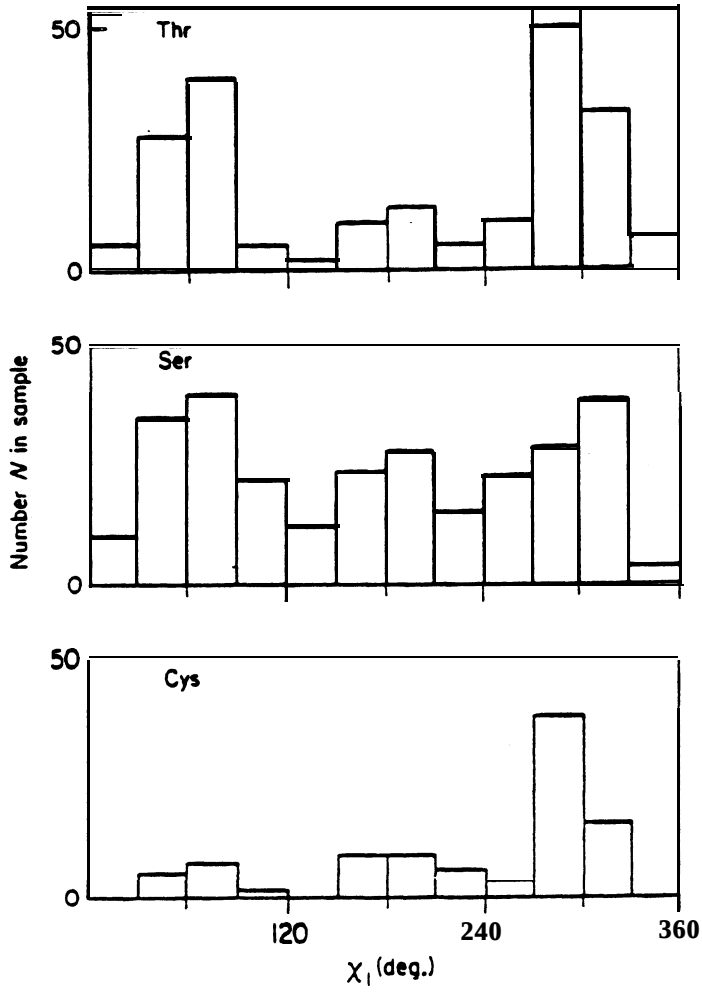


FIG. 5. The χ_1 angle distribution in Thr, Ser and Cys. The sample contains 207 Thr and 285 Ser residues. Sample 2 (28 protein structures, see Table 1) is used for Cys; it contains 111 residues.

side-chains like Met or Lys. The detailed geometry of disulphides (Pattabha & Srinivasan, 1976) and that of sulphur–metal complexes in proteins (Carter, 1977) has been described before. We only want to point out that the high frequency of the favourable g+ position shows that the presence of cross-linking covalent bonds does not perturb the geometry of the residue itself.

(c) *The χ_2 angle*

(i) *Methionine, glutamic acid, glutamine, lysine, arginine*

In these residues, the γ-carbon is tetrahedral and unsubstituted. The position of the δ-atom is determined by the χ_2 angle, which shows a strong preference for values

near 180°, corresponding to the *t* position *trans* to the C_α and away from the main chain (Fig. 6). The *t* peak in the experimental distribution of χ_2 is centered at 180° and comprises 69% of the side-chains, against 17% in the g^+ region and 14% in the g^- region. The experimental χ_1 χ_2 map (Fig. 7(a)) is in excellent agreement with the energy map (Fig. 7(b)). It indicates that the allowed combinations g^-t, tt, tg^-, g^+t and g^+g^+ (Ponnuswamy & Sasisekharan, 1971*b*) comprise 90% of the observed side-chain configurations. Indeed, three-quarters of the residues occur in three configurations only: tt, g^+t and g^+g^+ (Table 3). The rare g^-g^-, g^-g^+ and tg^+ combinations lead

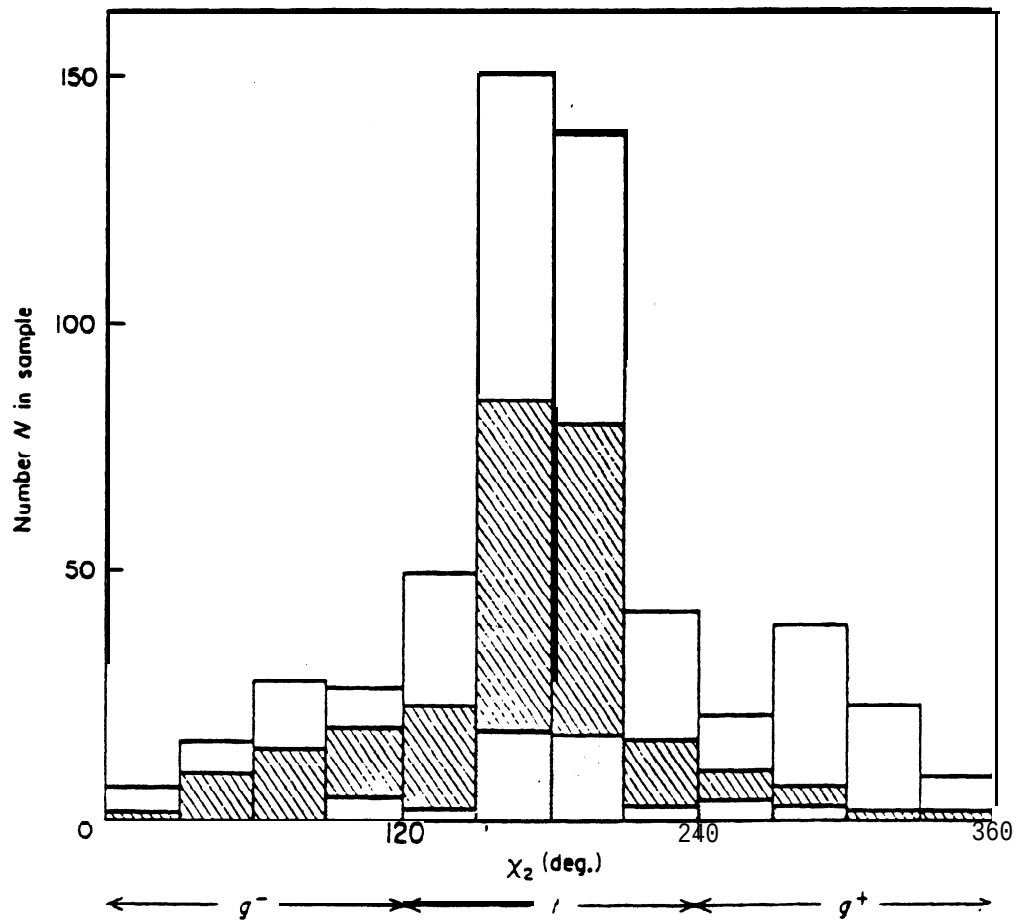


FIG. 6. The χ_2 angle distribution in Glu, Gln, Lys and Arg. The sample contains 551 side-chains; χ_2 distributions are shown for g^- (stippled), *t* (hitched) and g+ (open) positions of the C_γ atoms.

TABLE 3

Correlation between χ_1 and χ_2: Met, Glu, Gln, Lys, Arg

	Class of χ_1			
χ_2	g^-	t	g^+	Total
g^+	7	14	81	102 (17)
t	42	167	193	402 (69)
g^-	5	45	32	82 (14)
Total	54	226	306	586
	(9)	(39)	(52)	

The data of Fig. 8 are distributed among 3 classes of each of the 2 side-chain dihedral angles, each class spanning 120°. The number of side-chains in each class is indicated, permitted configurations being underlined. Percentages in parentheses refer to the total sample of 586 Met, Glu, Gln, Lys and Arg side-chains.

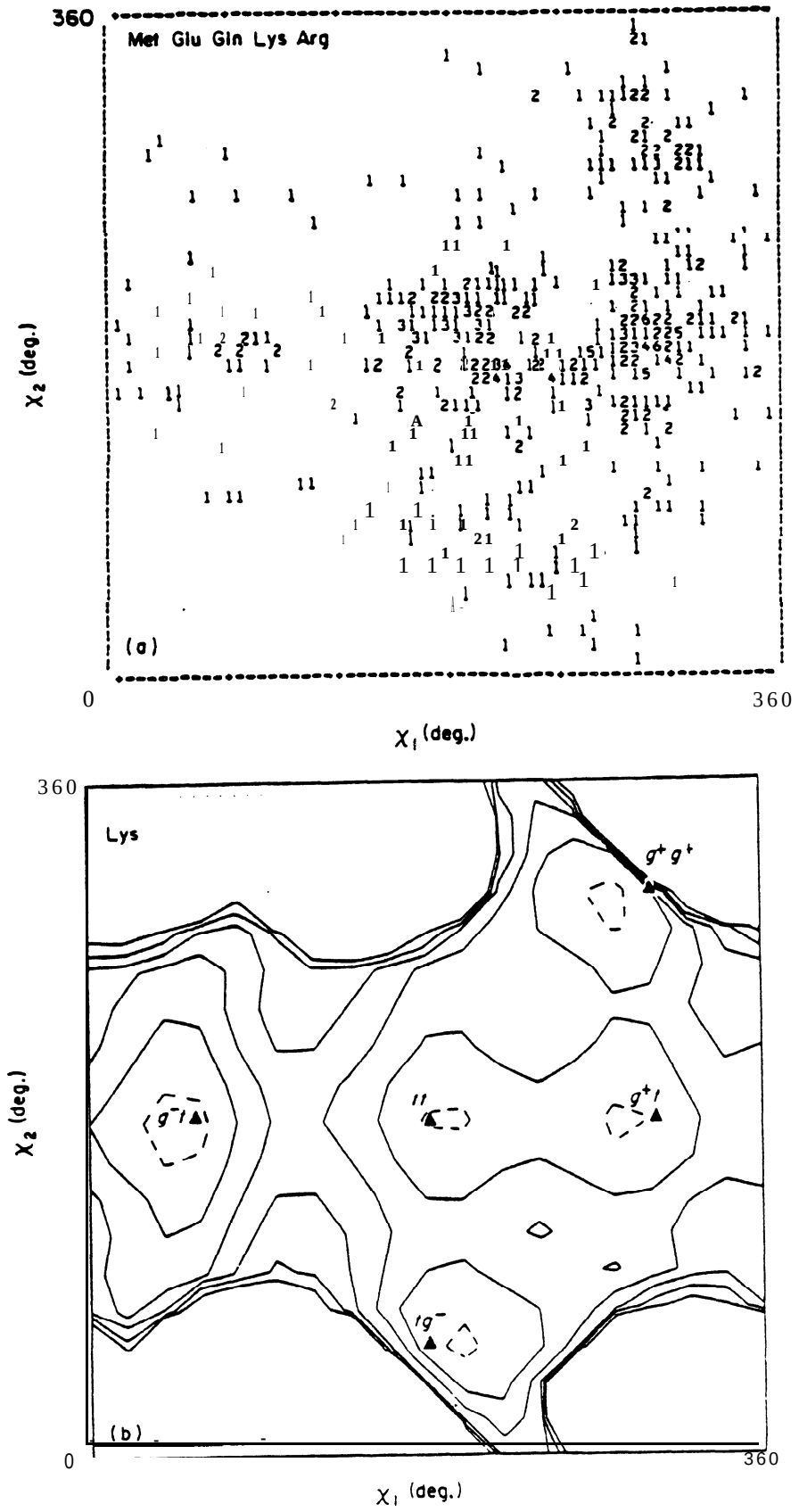


FIG. 7. χ_1 χ_2 maps for Met, Glu, Gln, Lys and Arg.

Similar experimental distributions and energy maps are obtained for these residues. The data are therefore combined.

(a) Experimental distribution; 585 data points. Sampling is done in 6° steps along the χ_1 axis, in 7·5° steps along the χ_2 axis in this map and in all following experimental maps. (b) Energy map. The potential energy (eqn (1)) of the CH_3CONH–RCH–$CONHCH_3$ structure is calculated for all values of χ_1 and χ_2 in steps of 20°. Here, R is a Lys side-chain with $\chi_3 = \chi_4 = 180°$; the main chain is in an extended conformation with $\phi = -140°$, $\psi = 140°$. Contours are drawn at −1 (broken), 1, 3, 5 and 9, kcal/mol.

to severe overlaps of side-chain and main-chain atoms. The few cases observed are almost certainly due to experimental errors; they represent 4·4% of the side-chains.

(ii) *Leucine*

The five allowed configurations described above reduce to two when the γ-carbon is branched: tg^- ($\chi_1 \sim 180°$, $\chi_2 \sim 60°$) with the two δ-methyl groups in the g^- and t positions, and g^+t ($\chi_1 \sim 300°$, $\chi_2 \sim 180°$) with the δ-methyl groups in the t and g^+ positions. These are the configurations found in 22 small crystal structures of blocked Leu residues (Benedetti, 1977). The protein data (Fig. 8(a)) are only in partial agreement with these findings: the g^+t cluster of points comprises 38% of the Leu side-chains, the tg^- cluster, 19%. However, another 33% of the Leu residues have χ_2 angles scattered in regions of the map where the isopropyl group is eclipsed with the α-hydrogen ($\chi_1 \sim 240°$). Some may be due to incorrect building of the side-chain in the electron density maps, orientations of the isopropyl groups differing by 180° in χ_2 being easily confused. Still, nearly all of these side-chains are in permitted regions of the χ_1 χ_2 energy map (Fig. 8(b)), while a small number (10 residues or 4% of the sample) of Leu residues have their side-chain in the g^- position ($\chi_1 < 120°$), where it overlaps with main-chain atoms for all values of χ_2.

(iii) *Isoleucine*

The χ_1 χ_2 map (Fig. 9) reflects the preference of the χ_1 angle for values near 300°, with the ethyl group in the g^+ position and the methyl group in the t position. The geometry of the rotation around the C_γ—C_δ bonds is the same as in unbranched side-chains: C_δ is generally *trans* to C_α. Thus, the g^+t configuration is most frequent (47%), g^+g^+ is next (16%), g^-t and tt comprising most of the remaining cases (24% together).

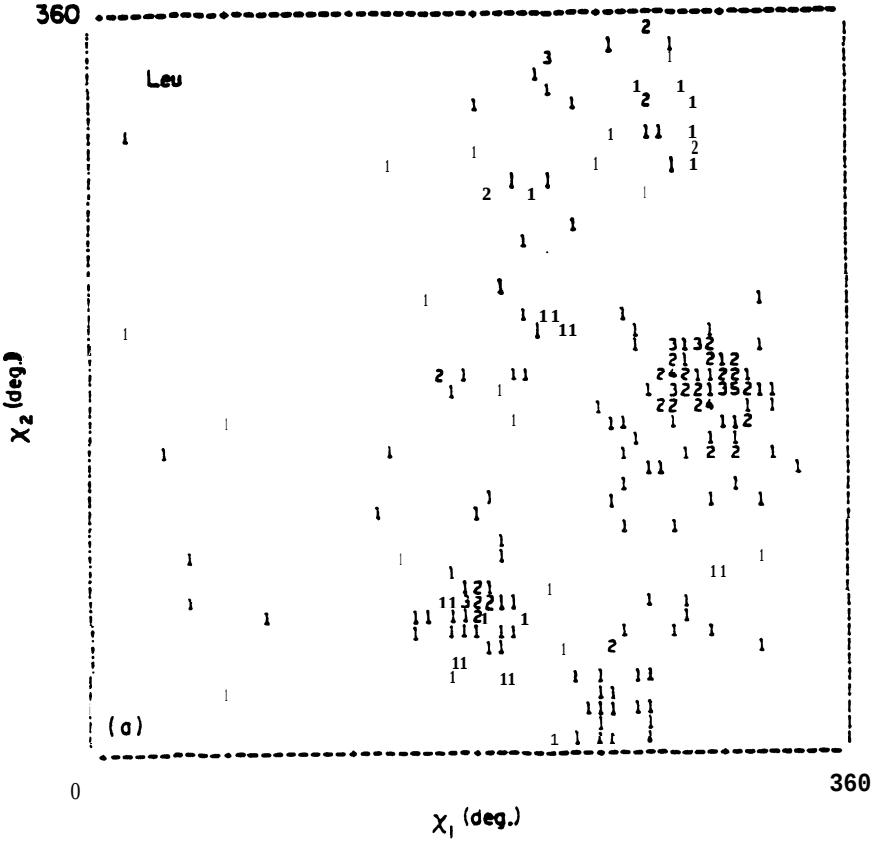
(iv) *Tryptophan, tyrosine, phenylalanine*

The χ_2 distribution is reduced to the 0 to 180° range for the symmetrical Phe and Tyr side-chains, and also for Trp in first approximation. The large peak near $\chi_2 = 90°$ (Fig. 10) indicates the preference of the aromatic ring for a position parallel to the main chain on which it lies flat. This is the only configuration observed in small crystal structures (Cody *et al.*, 1973). In proteins, the observed combinations of χ_1 and χ_2 (Fig. 11(a)) are in good agreement with the calculated energy map (Fig. 11(b)).

The data points cluster around three positions, with the following average angles and standard deviations:

$\bar{\chi}_1 = 68°$, $\bar{\chi}_2 = 91°$, $\sigma_1 = 18°$, $\sigma_2 = 16°$ (g^- position)
$\bar{\chi}_1 = 184°$, $\bar{\chi}_2 = 76°$, $\sigma_1 = 16°$, $\sigma_2 = 21°$ (t position)
$\bar{\chi}_1 = 291°$, $\bar{\chi}_2 = 96°$, $\sigma_1 = 16°$, $\sigma_2 = 16°$ (g^+ position).

In the g^- position (13% of the sample), only values of χ_2 near 90° are observed. The energy map indicates that all other orientations of the aromatic ring lead to severe conflicts with main-chain atoms. In the t position (31% of the sample), the ring is more mobile, leading to a larger value of the standard deviation σ_2. In the g^+ position, the cluster around $\chi_2 = 96°$ contains 47% of the data points, but another 10% of the side-chains have χ_2 angles larger than 140° or smaller than 30°, corresponding to transverse positions of the ring relative to the main chain. As predicted by



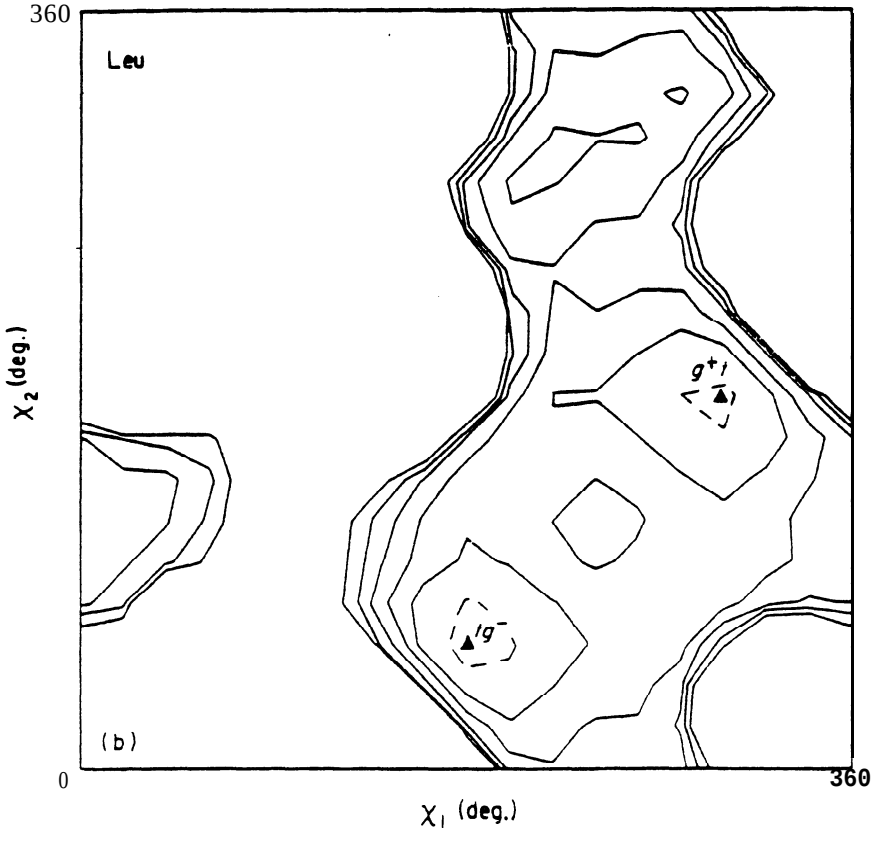


FIG. 8. χ_1 χ_2 map for Leu. The experimental map (a) contains 237 data points. The energy map (b) is calculated for a blocked Leu residue in an helical main-chain conformation with $\phi = -60°$, $\psi = -50°$. Contours are drawn at 0 (broken), 2, 4, 6, 8 and 10 kcal/mol.

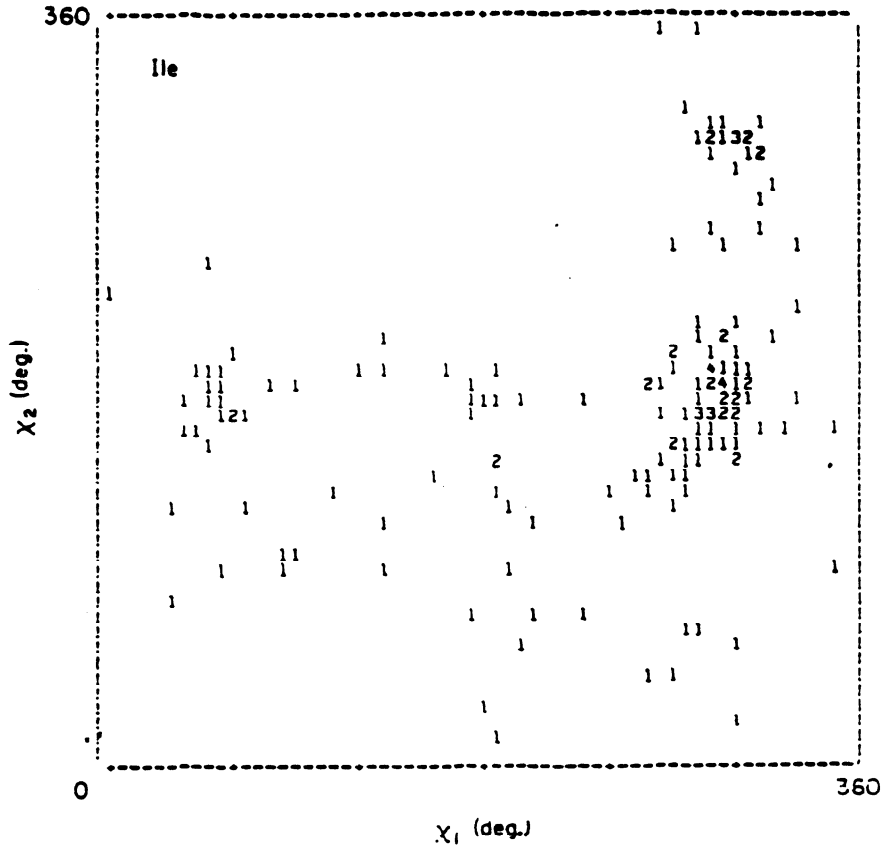


FIG. 9. χ_1 χ_2 map for Ile. 167 data points.

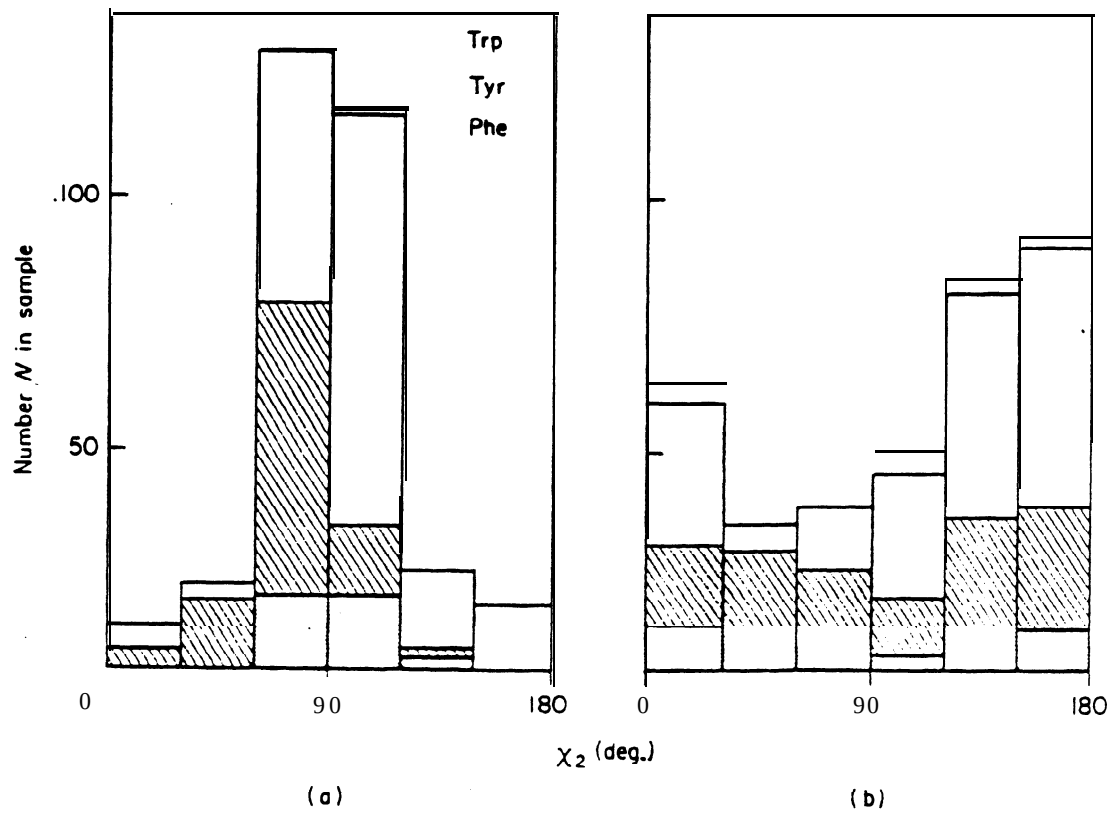


FIG. 10. The χ_2 angle distribution with trigonal γ-carbons. Distributions are shown for g⁻ (stippled), t (hatched) and g^+ (open) positions of the C_γ atom.

(a) Aromatic residues; Trp. Tyr, Phe ; 320 side-chains. Indole groups are taken to be symmetric. (b) Asp and Asn; 348 side-chains. Amide groups are taken to be symmetric.

energy calculations, these are found only in the less crowded g + position. Furthermore, the displacements of $\overline{\chi_2}$ away from 90° in the t and g+ clusters are significant and also correctly predicted in Figure 11 (b).

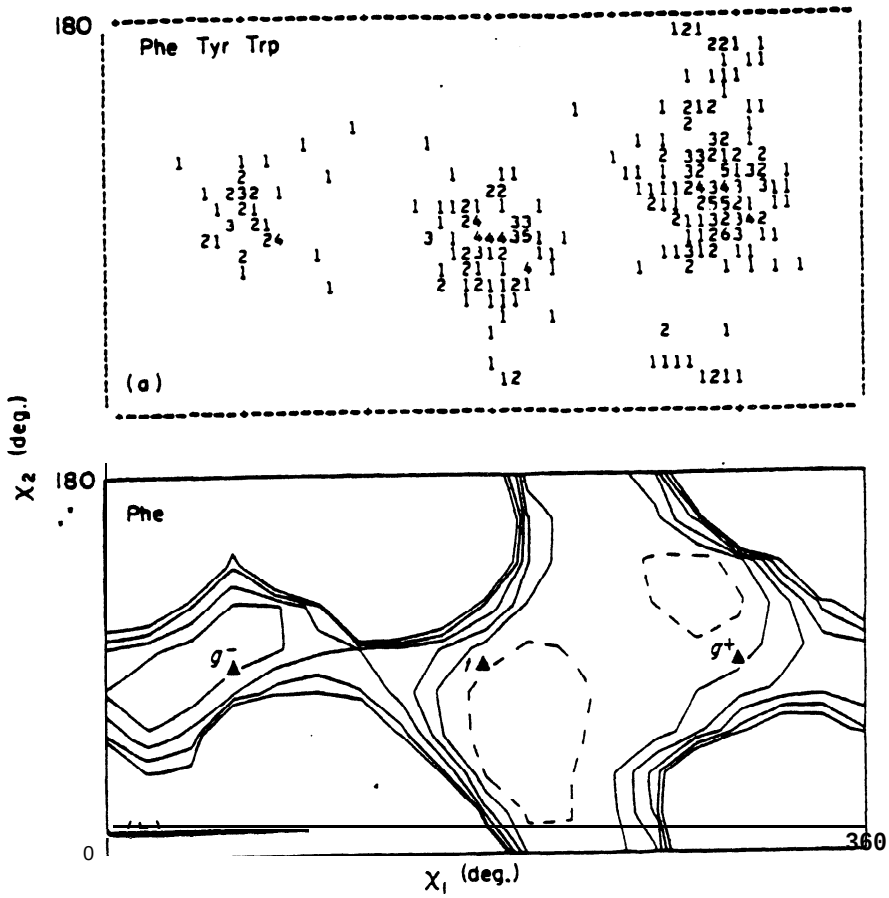


FIG. 11. χ_1 χ_2 maps for aromatic residues.
The indole group of Trp is taken to be symmetric.
(a) Experimental distribution; 320 data points. (b) Energy map. The calculation is made for a blocked Phe residue in an extended main-chain conformation with $\phi = -140°$ and $\psi = 140°$.

In Trp side-chains, the ring positions corresponding to $\chi_2 \sim 90°$ and 270° are not sterically equivalent. In our small sample of 59 Trp residues, they appear with different frequencies: values near 90° (with the five-membered ring of the indole group pointing away from the amino group of the residue in the most common g + and t positions) are twice as frequent as values near 270°.

(v) *Histidine*

The experimental χ_1 χ_2 map reduced to $\chi_2 < 180°$ (Fig. 12(a)) is very similar to that of the aromatic residues. Thus, the distribution of the side-chain configurations is not strongly affected by the increased mobility of the imidazole ring, expected from its smaller size compared to a phenyl ring, nor by hydrogen bonds made by the side-chain nitrogen atoms.

(vi) Aspartic acid, asparagine

The distribution of χ_2 angles for these residues (Fig. 12(b)) is centered at $\overline{\chi_2} = 156°$ with a standard deviation $\sigma_2 = 38°$. Thus, the preferred position of the γ-carboxylate or amide group is transverse to the main chain and not parallel to it as is observed in

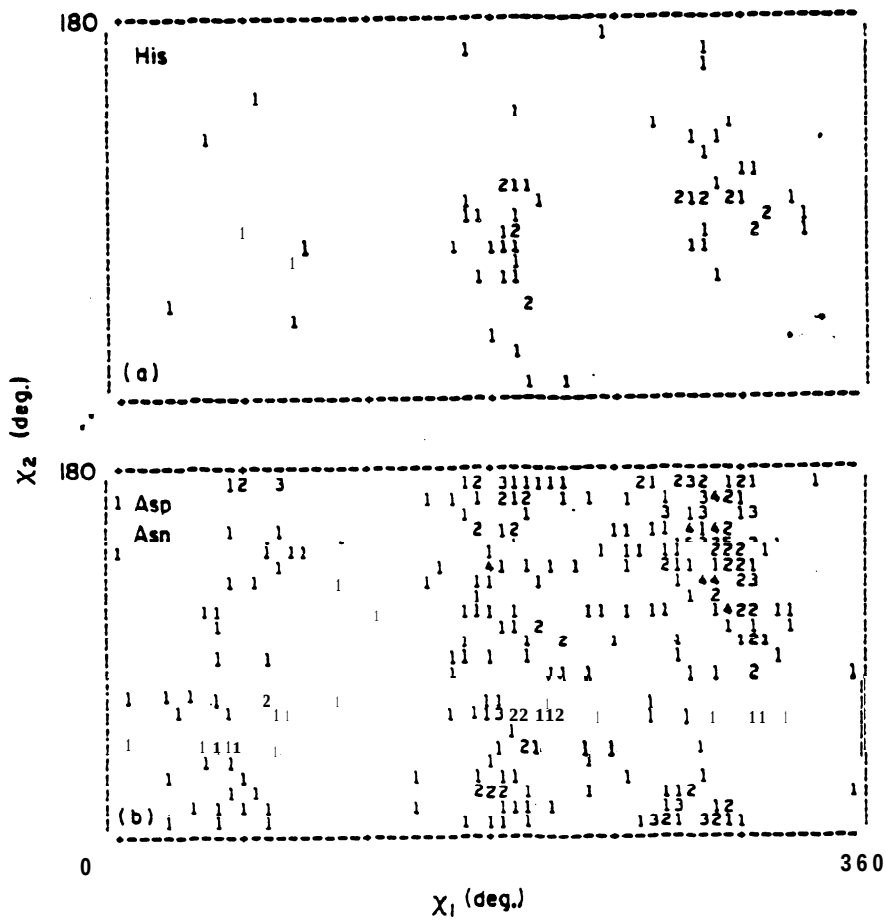


FIG. 12. χ_1 χ_2 maps for His and for Asp and Asn.
(a) His; 65 data points. (b) Asp and Asn; 348 data points. The amide groups are taken to be symmetric.

aromatic residues, even though the geometry of their C γ atoms is similar. Steric hindrance is obviously less important in Asp "and Asn side-chins. and electrostatic interactions dominate (Lipkind *et al.*, 1973). The wide distri bution of χ_2 around its mean value expresses the mobility of the terminal groups and the difficulty of determining precisely their orientation in electron density maps-

(d) *Other side-chain dihedral angles*

The orientation of the Glu carboxylate group or of the Gln amide group. and the position of the side-chain atoms beyond C_δ in Lys and Arg, are even less precisely determined. These polar groups are often free in solution. or involved in intermolecular interactions due to crystal packing. They are subject to very little steric hindrance

from the main chain, at least when C_δ *is trans* to C_α (χ_2 near 180°). In addition, the experimental data on χ_3 and χ_4 are not reliable. The high frequency of values near 180° noted by Chandrasekaran & Ramachandran (1970) may result entirely from the convention of building these side-chains in extended conformation when the electron density map is ambiguous.

The case of Met residues is different and instructive. The χ_3 angle controls the position of the ϵ-methyl group, and the side-chain is rarely exposed to solvent- Still, the experimental distribution (Fig. 13) is almost flat, except for values near 0°, with C_ε and C_β eclipsed, which are not observed.

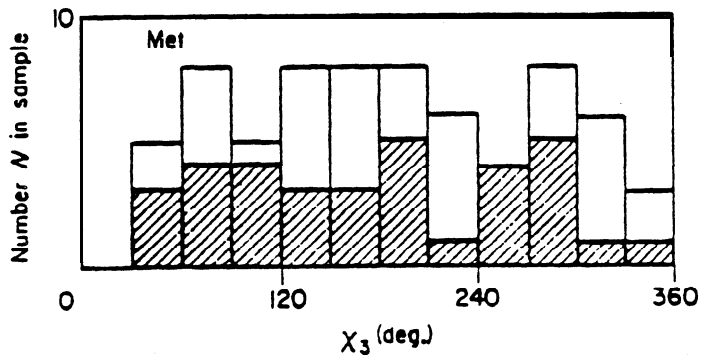


FIG. 13. The χ_3 angle distribution in Met. Sample 1 (hatched) contains 34 Met side-chains, sample 2 has 69 (see Table 1).

(e) *Influence of main-chain conformation*

We characterize the conformation of the main chain by the ϕ and ψ dihedral angles and divide the $\phi\psi$ map into three zones: extended configurations of the chain for negative ϕ and $30° < \psi < 210°$; right-handed helices for negative ϕ and $-120° < \psi < 30°$; left-handed helices for $0° < \phi < 120°$ and $-60 < \psi < 90°$. The remainder of the $\phi\psi$ map is forbidden for all residues except Gly. Though residues with $\phi\psi$ angles in one of the three regions thus defined may not belong to a piece of secondary structure, we call them, respectively, β (extended), α_R and α_L. Levitt & Greer (1977) note that most residues in the α_R region actually belong to a piece of helix, but only about half of the residues in the β region belong to a sheet-like structure; α_L is *rare* and limited to turns of the chain. In our sample of residues, 43% are of the β type, 51% α_R, 4% α_L and 2% do not belong to one of these categories. Due to well-known preferences (Chou & Fasman, 1971), the amino acid compositions of the β, α_R and α_L categories differ, though the only striking deviations from the average distribution affect Asn (more frequently α_L) and Glu (with a strong preference for α_R).

The conform&ion of the main chain affects the side-chain dihedral angles in the limited wav expected from the energy calculations of Figure 2. One may compare the calculated $\phi\psi$ maps to the experimental distribution of angles for residues having C_γ in the g^-, t or g^+ position (Fig. 14). The distribution of the main-chain angles in residues where the side-chain is in the g^+ position (Fig. 14(c)) has data points in all allowed regions of the classical Ramachandran map established for Ala residues: in the g^+ position (*trans* to t he carbonyl group), the side-chain causes no more steric hindrance than the C_β alone. In the t position, overlaps of C_γ y-it.11 the carbonyl group

TABLE 4

Correlation between χ_1 angles and main-chain conformation: Trp, Tyr, Phe, Met, Leu, Asp, Asn, Glu, Gln, Lys, Arg, His

$\phi\psi$ angles	Class of χ_1 angle g^-	t	g^+	Total
β	68 (10)	269 (40)	333 (50)	670 (43)
α_R	82 (10)	260 (33)	450 (57)	792 (51)
α_L	2 (3)	15 (24)	46 (73)	63 (4)
Other	16 (52)	5 (16)	10 (32)	31 (2)
Total	168 (11)	549 (35)	839 (54)	1556

The number and percentage (in parentheses) of residues having χ_1 angles **if each of the** three **120°** ranges are compared **for 4 classes** of main-chain **conformations, defined in the text as zones of the Ramachandran diagram,** rather **than as elements of secondary structure.** The **percentages in the rightmost column are those of the 4 classes in** the **sample of 1556 residues. The** 4th **class (other) includes** the **few residues** having $\phi\psi$ **angles outside** the **zones defined here as** β, α_R **and** α_L. **Most of them** are **likely** to be experimental **errors, and their untypical** χ_1 **distribution is spurious.**

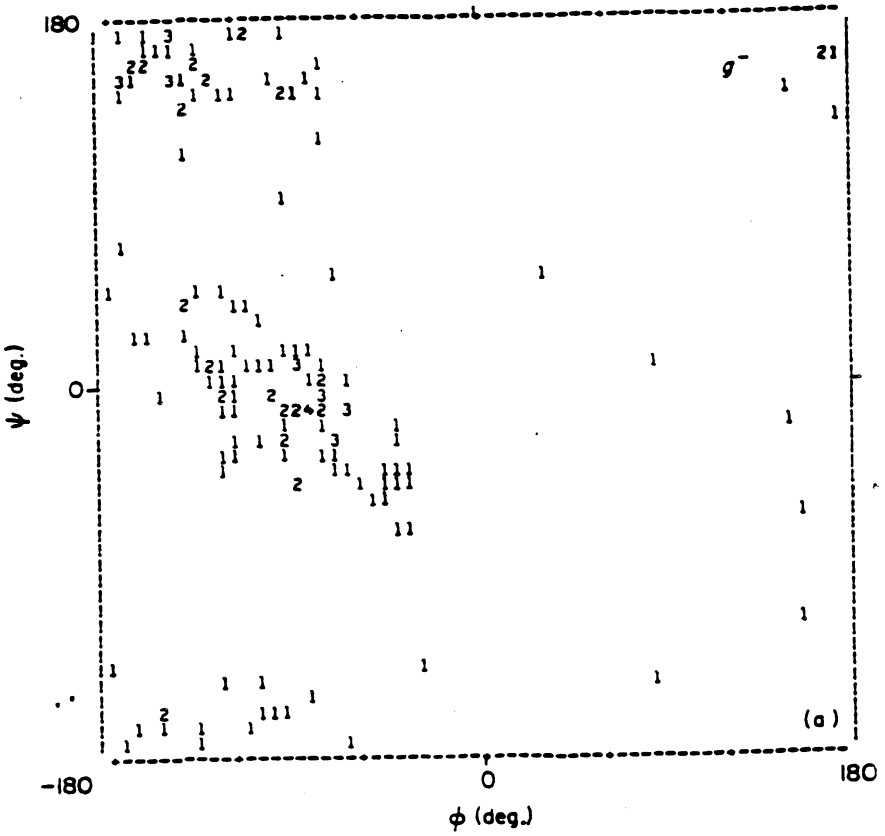
make the α_L configuration less favourable (Fig. 14(b)). The permitted region corresponding to right-handed helices is also reduced. Thus, the β/α_R ratio, which is 0·85 **on average, increases** to 1·03 **in residues** of the *t* side-chain type, while it is 0·74 only **in residues** of the g^+ side-chain **type** (Table 4). A similar situation is observed for Ile and Val **residues:** where the *t* position is occupied by one of the two C_γ atoms (except when the side-chain is in the rare g^- configuration): **the** β/α_R **ratio is** high (1·15 in our **sample). Lastly,** in the g^- **position,** the side-chain overlaps with the amino and with the carbonyl group of the residue. The α_L **main-chain** configuration is forbidden, the permitted α_R and β regions of the $\phi\psi$ map **are reduced** in area (Fig. 14(a)). An immediate consequence is the uneven distribution of **the** g^-, *t* and g^+ side-chain classes **among** the three main-chain classes (Table 4). **The most** extreme **situation is that** of α_L residues, three-quarters of which are g^+. But the-preference for g^+ is also strong in right-handed helices, where it is almost twice as **frequent** as *t*. This effect of the main-chain conformation. coupled with the **different amino** acid composition of the β, α_R and α_L classes, is the **source** of some of the departures from the average g^+/t ratio, which are observed when each type of amino acid residue is considered independently in the statistics. Thus, g^+/t is higher (1·33) for Glu **residues-** due to their high frequency in helices, than for Gln ($g^+/t = 1·0$).

(f) *Hydrogen bonding*

The major effect of main-chain atoms on the conformation of the side-chain is steric hindrance, which leads to the correlations observed above for side-chains having y-carbon atoms. In residues having an oxygen atom in γ position, and to a lesser extent in Asp and Asn, the main chain also affects the side-chain configuration by providing possibilities of favourable electrostatic interactions (Zimmerman & Scheraga, 1977) usually hydrogen bonds to a neighbouring carbonyl group.

(i) *Serine, threonine*

Out of 492 Ser and Thr residues in the sample we find that 70% have their O_γ atom within hydrogen-bonding distance (3·5 Å) of at least one carbonyl group, with an acceptable angular geometry for a OH . . . O bond. If O_γ . . . HN bonds are considered,



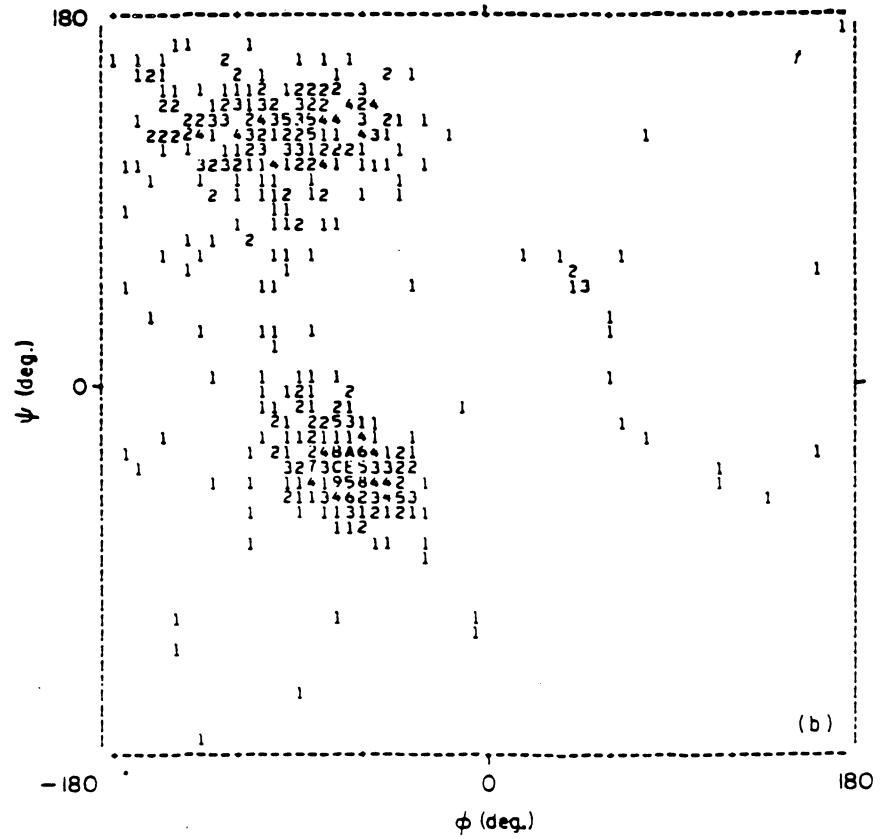


FIG. 14. $\phi\psi$ maps. Main-chain dihedral angles of Trp, Tyr, Phe, His, Met, Leu, Asp, Asn, Glu, Gln, Lys and Arg residues are plotted on 3 separate Ramachandran maps depending on the position of C_γ.

(a) g^- position, χ_1 less than 120°; 168 data points. (b) t position, χ_1 from 120° to 240°: 549 points. (c) g^+ position, χ_1 larger than 240°; 840 data points.

the fraction of bonded Ser and Thr side-chains is even larger. Most of these bonds involve a main-chain carbonyl group as acceptor (83%), though side-chain to side-chain bonds may coexist with side-chain to main-chain bonds. In 85% of the cases (i.e. in 50% of the Ser and Thr side-chains), the carbonyl group belongs to a neighbouring residue in the amino acid sequence. If i is the Ser or Thr hydrogen-bond donor, the carbonyl group may belong to residue $i-5$ to $i+2$, with high frequencies of $i-4$, $i-3$ and i. Thus, at least 50% of the Ser and Thr side-chains are involved in "local" hydrogen bonds with main-chain atoms.

These hydrogen bonds are determined by the conformation of the main chain, which fixes the position and orientation of the acceptor oxygens. In extended structures, the O_γ atom is within hydrogen-bonding distance of t-he residue's own carbonyl oxygen (O_i), when the side-chain is g⁻ or t (Fig. 15(b)), though the angular geometry is poor. Hydrogen bonds to the preceding residue's carbonyl oxygen (O_{i-1}) require the side-chain to be g+ (Fig. E(a)). We find that 30% of the Ser/Thr residues with $\phi\psi$ angles in the β region of the Ramachandran map have ψ and χ_1 angles compatible with the existence of a OH . . . O_i bond, and another 13% have ϕ and χ_1 angles compatible with a OH.. . O_{i-1} bond. In small crvstal structures, intermolecular hydrogen bonds are preferred, but solution studies sup&t tbe existence of O_{i-1} and O_i types of bonds, which theory predicts in blocked Ser residues (Lipkind *et al.*, 1973).

These bonds are impossible in right-handed helices, because the carbonyl groups of the residue and of its immediate neighbours point away from the side-chain. However, the peptide oxygen atoms in the preceding turn of the helix are available as hydrogen-bond acceptors for the hydroxyl group as well as for the peptide NH. In regular α-helices, both NH and $O_\gamma H$ can bond to O_{i-4} (Fig. 15(c)): this is observed in 55 Ser and Thr residues of the sample, i.e. 23% of those belonging to the α_R region of the

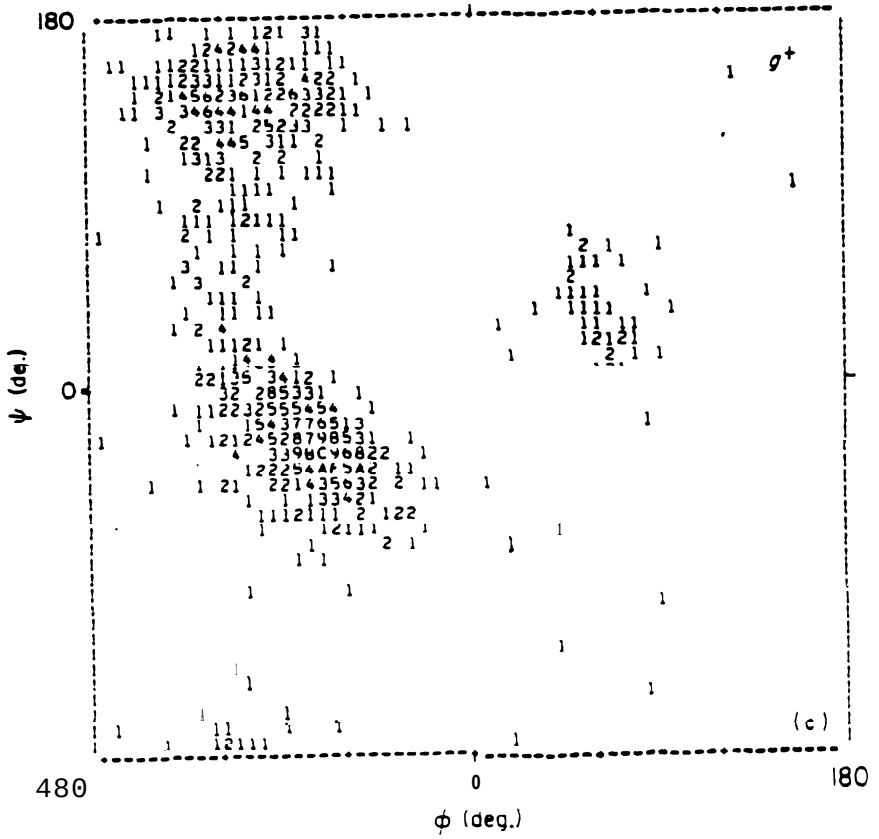


FIG. 14(c).

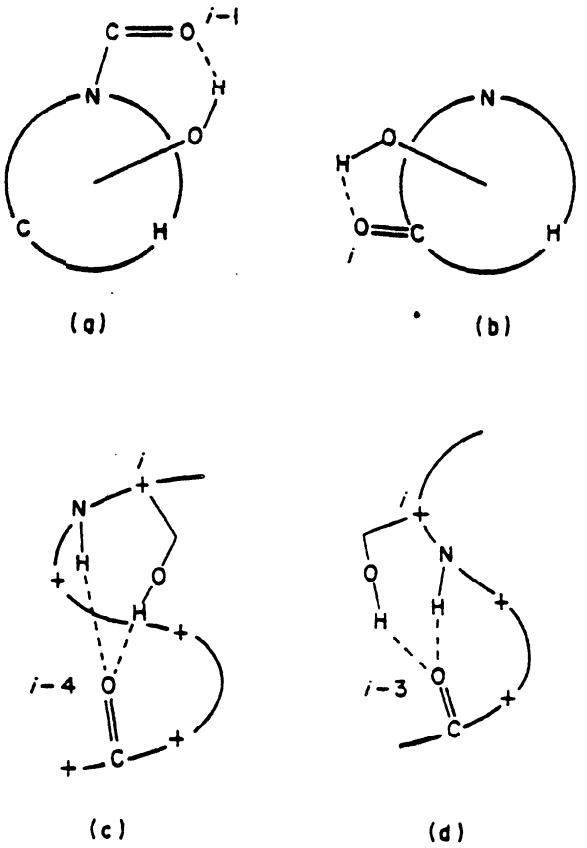


FIG. 15. Side-chain to main-chain hydrogen bonds in Ser and Thr. The Ser/Thr residue is i, its hydroxyl group is H-bonded to a peptide oxygen atom near in the amino acid sequence.

(a) Extended main-chain conformation, bond to O_{i-1} in the preceding peptide group; O_γ is in the g^+ position. (b) Extended main-chain conformation, bond to O_i; O_γ is in the g^- or t position. (c) a-Helix, bond to O_{i-4}; O_γ is in the g^+ position. (d) 3_{10} helix, bond to O_{i-3}; O_γ is in the g^- position, or g^+ in distorted helices.

Ramachandran map. The side-chain is then in the g^+ position. In 38 other cases, the side-chain is bonded to O_{i-3} and is either in the g^- or g^+ position; O_{i-3} is the hydrogen-bond acceptor of the peptide NH of residue i in 3_{10} helices (Fig. 15(d)), which are not common in protein structures. However, a 3_{10} helix exists in haemoglobin (helix C). In the horse methaemoglobin structure (Ladner *et al.*, 1977), all four Thr residues of the C helix (Thr α38, α39, α41 and β38) have their side-chain hydrogen-bonded to O_{i-3}. This type of bond is also frequent in Ser and Thr residues placed at the end of α-helices, where the main-chain conformation is often close to 3_{10}. Thus, in helices, nearly 40% of the Ser and Thr side-chains make a O_{i-4} or O_{i-3} type of hydrogen-bond. No such bond can be made if O_γ is in the t position. This explains why the t side-chain position is rarer for helical Ser residues than for extended ones (see above).

(ii) *Aspartic acid and asparagine*

These side-chains have two polar atoms, and more possibilities of forming hydrogen bonds than Ser or Thr. However, the large majority of Asp and Asn residues are accessible to the solvent (see Table 1) and may bond to water molecules. The orientation of the amide group of Asn is always ambiguous in electron density maps. Accepting the choices made in published atomic co-ordinates, we find that 40% of the 348 Asp and Asn side-chains of the sample are involved in one or more hydrogen bond.

The majority (60%) invoke a main-chain peptide group either as donor (NH) or acceptor (CO bonded to the side-chain amide group of Asn); 63% of these (17% of the Asp and Asn residues) are local hydrogen, bonds to the NH of residues $i - 3$ to $i + 3$. The most frequent (3'7 cases, or 60% of the local bonds) is to N_{i+2}; this side-chain hydrogen bond requires χ_1 to be larger than 180° and cannot be made when the main chain is extended.

(iii) *Other polar side-chains*

The side-chains of Glu, Gln, Lys, Arg, Tyr, Trp and His are also involved in electrostatic interactions, but mostly "with other side-chains, bonds to main-chain atoms near in the sequence being either sterically forbidden (aromatic side-chains) or unlikely.

(g) *Accessibility to solvent and* side-chain *conformation*

The position of a residue relative to the protein surface and its accessibility to the solvent can be conveniently characterized by its accessible surface area *A* (Lee & Richards, 1971). Buried residues with *A* smaller than 20 Å² have generally fewer than two atoms on the protein surf&e. They are invoked in many tight contacts with other residues in the protein. Exposed residues with *A* larger than 60 Å² have most of their side-chain in contact with 'the solvent. The value of A is therefore a crude estimate of the strength and number of interactions made by the residue with the remainder of the protein structure.

Table 5 shows how the position of a residue relative to the protein surface affects its side-chain conformation, The remarkable feature is the smaller frequency of the rare g^- side-chain position observed in buried residues (5·5%) compared to average (11 %) or exposed residues. The difference is highly significant: on a sample of 435 buried residues, the expected number of g - side-chains is *47;* its standard deviation is 6·5; the observed number 24 has a probability of 10^{-6}. The difference does not result from the different amino acid compositions of the three classes of accessibilities: the exclusion of the g^- side-chain position in buried residues occurs in aromatic side-chains (8% g^-, average 13 %), in long side-chains (Glu, Gln, Lys, Arg: buried 5 % g^-, average 9 %) and in Asp and Asn (buried 6% g $^-$, average 15 %). Similarly, the *t* configuration, with the methyl group in g^- is rarer (12%) in buried Thr residues than

TABLE 5

Correlation between χ_1 angles and accessibility

Accessibility	Class of χ_1 angle g^-	t	g^+	Total
Exposed	82 (12)	222 (32)	390 (56)	694 (45)
Intermediate	62 (15)	153 (36)	212 (50)	427 (27)
Buried	24 (6)	174 (40)	237 (55)	435 (28)
Total	168 (11)	549 (35)	839 (54)	1556

The number and percentage (in parentheses) of residues having χ_1 angles in each of the 120° ranges are compared for 3 classes of accessibility to solvent: buried residues with accessible surface areas smaller than 20 Å², intermediate (20 to 60 Å²) and exposed residues (more than 60 Å²). The percentages on the rightmost column are those of the accessibility classes in the sample of 1556 Trp, Tyr, Phe, His, Met, Leu, Asp, Asn, Glu, Gln, Lys and Arg residues.

average (15%). The g^+ configuration is less predominant in exposed Ile and Val residues (55%) than average (67%). In each case, the contrast between frequent and infrequent side-chain configurations is more pronounced in buried than in exposed residues.

4. Discussion

The experimental data on side-chain dihedral angles derived from protein X-ray crystallography are of better quality than is often assumed. The reproducibility of the χ angles measurement may be lower than for main-chain dihedral angles, because changes in ϕ and ψ affect the totality of the protein structure, while side-chain dihedral angles control the position of a small number of atoms only. A fraction of the side-chain conformations obtained by model-building from electron density maps are incorrect due to misinterpretation of the density, or result from arbitrary choices made when the side-chain is mobile. Still, the values of χ angles appear to be reliable when they control the position of more than one non-hydrogen atom. Thus, our statistics imply that no more than about 5% of the χ_1 angles have large errors, except in serine, and that χ_2 is generally correct, except perhaps in Asp and Asn residues. The distribution of the side-chains in three large classes g^-, t and g^+ is therefore precisely established. Within these three categories, the standard deviations of χ_1 and χ_2 estimated from duplicate measurements vary between 3° and 16°, depending on the residue type (Table 2).

The consistency of the data leads to a high degree of contrast in the experimental distribution of the side-chains between the permitted configurations. For all residue types, one to five configurations account for 85% or more of the side-chain structures up to the δ atom; one or two configurations account for 60% or more (Table 6). The least favourable of the permitted configurations represent a few per cent of the da-ta. Serine residues stand alone in taking all permitted configurations with comparable frequencies. The least mobile side-chains are Val and Ile, the aromatic residues and Cys, though Cys residues with a free -SH group may be less restricted than substituted cysteines. Leucine residues appear surprisingly mobile, with more than 40% of the side-chains in a variety of conformations.

The side-chain configurations found in proteins may be compared to those found in

TABLE 6

Principal configurations observed

Residue	Configurations
Ser	g^- 38%, g^+ 34%, t 28%
Cys	g^+ 57%, t 27%, g^- 16%
Thr	g^+ 48%, g^- 39%, t 13%
Val	g^+ 66%, g^- 21%, t 13%
Ile	g^-t 46%, g^+g^+ 16%, g^-t 14%, tt 10%
Leu	g^+t 38%, tg^- 19%
Met, Glu, Gln, Lys, Arg	g^+t 33%, tt 28%, g^+g^+ 14%, tg^- 8%, g^-t 7%
Asp, Asn	g^+ 51%, t 34%, g^- 15%
Trp, Tyr, Phe	g^+ 57%, t 31%, g^- 13%
His	g^+ 45%, t 45%, g^- 10%

The most frequent configurations are indicated along with their frequency in the sample.

small crystal structures. The two sets of data are in general agreement (Lakshminarayanan *et al.*, 1977 ; Ponnuswamy & Sasisekharan, 1970), at least for amino acids with blocked N and C terminals. The free amino acids are quite different : crystalline L-valine·HCl, L-tyrosine and L-phenylalanine·HCl have their side-chain in the g^- position, which is rare in proteins and in blocked amino acids (Benedetti, 1977 ; Cody *et al.*, 1973).

In our sample of protein structures, the distribution of side-chain dihedral angles within a given configuration is rather narrow The average value of χ_1 is predicted accurately on the basis of van der Waals' interactions and of steric hindrance. The g^- peak is centered at 61°, the t peak at 190° and the g^+ peak at 290° ; the deviation from the ideal values (180° and 300°) is significant in the last two cases. Similarly, the χ_2 distribution in aromatic side-chains is expected and found to be centered at values below 90° in the t position ($\overline{\chi_2} = 76°$) and above 90° in the g^+ position ($\overline{\chi_2} = 96°$). Data from small crystal structures (Cody et *al.*, 1973) show the same trends. The root-mean-square dispersion of the protein data around the mean value $\bar{\chi}_1$ is of the order of 16° in aromatic residues. and of 20° to 25° in other residues. This dispersion results from (1) random errors in the experimental data, (2) perturbations of the χ_1 angle caused by interactions of the side-chain with main-chain atoms, and by interactions between side-chain atoms when rotation around β—γ and other bonds is possible. These are "local" effects. (3) Perturbations due to other atoms in the protein structure, especially by side-chain packing in the protein interior, which is close-packed (Richards 1974; Chothia, 1975). These are "long-range" effects

We have estimated the magnitude of the experimental errors (Table 2), and can therefore set an upper limit to the magnitude of the other factors: the χ_1 angles are probably distributed 15° to 18° (standard deviation) around the best t and g + positions. Our energy calculations indicate that the positions of the g+ and t minima are rather insensitive to the conformation of the main chain; the effect on χ_1 of side-chain atoms beyond C_γ is small in the most common situation where they are *trans* to C_α. Therefore, the major perturbations results from long-range interactions, which displace the side-chains from their ideal configuration. Changing χ_1 by 15° in the g+ or t potential well affects the residue's energy by no more than 0·6 kcal/mol (Fig. 4) : the width of the χ_1 angle distribution around the peaks is about the same as if it were due to thermal vibration ($RT = 0{\cdot}6$ kcal/mol at 300 K). The average contribution of long-range forces to the side-chain conformational energy cannot be larger, or it would significantly perturb the distribution and broaden the peaks. However, a small number of side-chains have χ_1 angles corresponding to large energies. Whether any of these are real remains to be determined. Experience with energy refinement of protein structure indicates that small atomic displacements are sufficient in most cases to remove strain localized in individual residues (Levitt, 1974 ; Gelin & Karplus, 1975; McCammon et *al.* 1977).

These conclusions derived from the study of χ_1 hold for other side-chin angles to a large extent. Unbranched side-chains (Met, Glu, Gln, Lys, Arg) prefer the extended t configuration with χ_2 near 180°. while the aromatic rings of Phe, Tyr, Trp and His prefer to lie flat on the main chain (χ_2 near 90°). In contrast: the distribution Of χ_2 angles in Asp and Asn residues cannot be derived solely from consideration of van der Waals' forces and steric hindrance. It is deeply affected by electrostatic interactions made by the polar carboxylate or amide groups. These interactions are also of obvious importance in fixing the position and orientation of the polar ends of the Glu, Gln, Lys and Arg side-chains.

Long-range interactions do not perturb strongly the conformation of individual side-chains. However, and especially for residues buried inside the protein, they determine which of the few permitted configurations is adopted by the side-chain. Small variations (15° to 18°) of χ angles around the preferred values are required for optimal packing of the side-chains in the protein interior. They correspond to atomic movements of 0·4 Å or so and have little effect on the conformational energy, as we have seen. Changing the χ_1 angle by 120° to move a side-chain from t to g^+ involves large atomic movements, more than 2·5 Å at the γ atom and 10 A or more in long side-chains. Such movements will often meet high energy barriers in the folded protein structure. though rotation of the χ_2 angle, which has much smaller effects on atomic positions, is permitted even in buried aromatic residues (Gelin & Karplus, 1975). For χ_1 at least, the critical choice has to be made during the folding of the polypeptide chain. One of the permitted configurations is selected, and only minor movements may occur later. It is then advantageous, from a kinetic point of view, to reduce the number of configurations accessible in the unfolded state (Levinthal, 1968 ; Karplus & Weaver, 1976). This is achieved in the amino acid side-chains of the 20 types found in proteins, and the choice of χ angles is further restricted by the secondary structure, which excludes the g^- configuration in helices, and favours "local" modes of hydrogen bonding for polar side-chains like Ser, Thr, Asp and Asn. One may think that the 20 chemical structures of the natural amino acids have been selected in part on the basis of their limited conformational mobility.

Steric hindrance limits the range of χ_1 and χ_2 for most residues, even in the unfolded polypeptide chain. The side-chain segments least subject to steric restrictions, such as the extremity of Lys and Arg side-chains, remain outside the folded structure due to their polar character. Both effects contribute to limit the loss of conformational entropy associated with the immobilization of the dihedral angles during folding. The corresponding loss of free energy depends on the statistical distribution of *the* side-chains between the permitted configurations in the unfolded chain. Chandrasekaran & Ramachandran (1970) have attempted to calculate it by counting the number of permitted $\phi\psi$ values for residues having their side-chain in the g^-, t or g^+ position. Similarly, Finkelstein & Ptitsyn (1977) measure the areas of the permitted regions of the $\phi\psi$ map. This amounts to taking a hard-sphere model with no enthalpy term for non-bonded interactions, an excessively crude model if we must explain the g^+/t ratio in, say, aromatic residues: the observed value is 57 : 31 (Table 6), equivalent to $RT \ln 1{\cdot}8 = 0{\cdot}36$ kcal/mol of free energy. Obviously non-bonded interactions having enthalpies larger than that occur even in the unfolded state. A proper prediction of the distribution would require a precise estimate of the energy of non-bonded interactions with averaging on main-chain conformation and on the position of side-chain atoms beyond C_γ.

An approximation to the statistical distribution of χ_1 in the unfolded state may be much more easily obtained in an empirical way by restricting the sample of side-chains analyzed to external residues of the protein. As we have seen, the same configurations are adopted by external (exposed to solvent) and internal (buried) side-chains, but some differences exist in their relative frequencies : configurations which are infrequent in exposed residues (i.e. g^-) are even rarer in buried residues ; dominant configurations (such as g^+ in Val and Ile) are even more so in buried residues. This is undoubtedly due in part to the poorer definition of many external side-chain positions in electron density maps. Still, this effect of accessibility on the side-chain distribution is observed

in aromatic residues as well as in smaller, less well-positioned side-chains. Most likely, it is a real feature of protein structures, that the most frequent configurations of the free side-chain are selected during folding. This is again advantageous from a. kinetic point of view, and leads to a lower conformational free energy in the immobilized side-chain. Assuming that the distribution of g^-, t and g+ observed in external side-chains (12%, 32%, 56% in Table 5) represents the statistical distribution in the unfolded state, the free energy of the folded state is lower by RT ln (56/12) = 0·9 kcal/mol when an internal side-chain takes the g^+ position rather than the g^- position, since the entropy loss resulting from the immobilization of χ_1 is the same. The gain may, of course, be balanced by a large release of enthaply in the case where favourable long-range interactions are made in the g^- position and not in g^+. Still, it is reasonable to think that protein structures have evolved to minimize individual components of their free energy, such as side-chain conformational free energy, in order to lower the overall free energy of' the folded state and the barriers of activation opposing folding.

We are grateful to J. L. De Coen, C. Chothia, R. C. Ladner and A. McCammon for useful criticism and discussion. One of 'us (S. W.) was supported by the Fonds National de la Recherche Scientifique Suisse. Part of this work was completed during the 1977 workshop on virus crystallography of the Centre Europeen de Calcul Atomique et Moléculaire in Orsay (Farnce).

REFERENCES

Benedetti, E. (1977). In *Peptides, Proc. 5th Amer. Peptides Symp.* (Goodman, M. & Meienhofer, J., cds), pp. 257–274. John Wiley & Sons, New York.

Bode, W. & Schwager, P. (1975). *J. Mol. Biol.* **98**, 693–717.

Carter, C. W. (1977). *J. Biol. Chem.* **252**, 7802–781 1.

Chandrasekaran, R. & Ramachandran, G. N. (1970). *Int. J. Protein Res.* **2**, 223–233.

Chothia, C. (1975). *Nature (London), 254,* 304–308.

Chothia, C. (1976). *J. Mol. Biol.* **105**, 1–14.

Chou, P. Y. & Fasman, G. D. (1971). *Biochemistry*, **13**, 211–222.

Cody, W., Duax, W. L. & Hauptman, H. *(1973).* *ht. J. Peptide Protein Res.* 5, 297–308.

Epp, O., Colman, P., Fellhammer, H., Bode, W., Schiffer, M., Huber, R. & Palm, W. (1974). *Eur. J. Biochem.* **45**, 513–524.

Finkelstein, A. V. *(1976).* *Mol. Biol. U.S.S.R.* **10**, 507–513.

Finkelstein, A. V. & Ptitsyn, O. B. (1977). *Biopolymers*, **16**, 469–495.

Gelin, B. & Karplus, M. (1975). *Proc. Nat. Acad. Sci., U.S.A. 72,* 2002–2006.

Gibson, I<. D. & Scheraga, H. A. (1967). *Proc. Nat. Acad. Sci., U.S.A.* **58**, 420–427.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. & Steigeman, W. (1974). *J. Mol. Biol.* **89**, X-101.

IUPAC-IUB Commission on Biochemical Nomenclature (1970). *J. Mol. Biol.* **52**, 1–17.

Karplus, M. & Weaver, D. L. (1976). *Nature (London)*, **260**, 404–406.

Ladner, R. C., Heidner, E. J. & Perutz, M. F. (1977). *J. Mol. Biol.* **114**, 385–414.

Lakshminarayanan, A. V., Sasisekharan, V. & Ramachandran, G. N. (1977). In *Conformation of Biopolymers* (Ramachandran, G. N., ed.), vol. 1, pp. 61-82. Academic Press, London.

Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.

Levinthal, C. (1968). *J. Chim. Phys.* **65**, 44–45.

Levitt, M. (1974). *J. Mol. Biol.* **82**, 393–420.

Levitt, M. & Greer, J. (1977). *J. Mol. Biol.* **114**, 181–240.

Levitt, M. & Lifson, S. (1969). *J. Mol. Biol.* **46**, 269–279.

Lipkind, G. M., Arkhipova, S. F. & Popov, E. M. (1973). *Int. J. Peptide Prot. Res.* **5**, 381–397.

McCammon, J. A., Gelin, B. R. & Karplus, M. (1977). *Nature (London)*, **267**, 585–590.

Nemethy, G. & Scheraga, H. A. (1977). *Quart. Rev. Biophys.* **10**, 239–352.

Pattabha, V. & Srinivasan, R. (1976). *Int. J. Peptide Prot. Res.* 8, 27–32.

Ponnuswamy, P. K. & Sasisekharan, V. (1970). *Int. J. Protein Res.* **2**, 37–45.
Ponnuswamy, P. K. & Sasisekharan, V. (1971*a*). *Int. J. Protein Res.* **3**, 1–8.
Ponnuswamy, P. K. & Sasisekharan, V. (1971*b*). *Int. J. Protein Res.* **3**, 9–18.
Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.
Richards, F. M. (1974). *J. Mol. Biol.* **82**, 1–14.
Zimmerman, S. S. & Scheraga, H. A. (1977). *Biopolymers*, **16**, 811–843.