

Main-chain conformational features at different conformations of the side-chains in proteins

Pinak Chakrabarti¹ and Debnath Pal

Department of Biochemistry, Bose Institute, P-1/12 CIT Scheme VIII, Calcutta 700 054, India

¹To whom correspondence should be addressed

An analysis of the known protein structures has shown that the main-chain torsion angles, ϕ and ψ of a residue can be affected by the side-chain torsion angle, χ_1 . The (χ_1 , ψ) plot of all residues (except Gly, Ala and Pro) show six distinct regions where points are concentrated—although some of these regions are nearly absent in specific cases. The mean of these clusters can show a shift along the ψ axis by as much as 30° as χ_1 is changed from around 180° to -60° to 60° . Because of the lesser steric constraint points are more diffused along the ψ axis when χ_1 is approximately -60° . Although points are more spread out along the ϕ axis in the (χ_1 , ϕ) plot, the dependence of ϕ on χ_1 shows up in a shortened ϕ range (by about 30°) when χ_1 is around -60° , and a distinct tendency of clustering of points into two regions when $\chi_1 \approx 60^\circ$, especially for the aromatic residues. Based on the dependence of the backbone conformation on its side-chain the 17 amino acids can be grouped into five classes: (i) aliphatic residues branched at the C_β position (although Thr is atypical), (ii) Leu (branched at the C_γ position), (iii) aromatic residues (Trp can show some deviations), (iv) short polar residues (Asp and Asn), and (v) the remaining linear-chain residues, mainly polar. Ser and Thr have the highest inclination to occur with two different orientations of the side-chain that can be located through crystallography. Such residues exhibiting two χ_1 angles have their ϕ and ψ angles in a region that is common to the Ramachandran plots at the two different χ_1 angles. The dependence of ϕ and ψ angles on χ_1 can be used to understand the helical propensities of some residues. Moreover, the average ϕ , ψ values in the α -helices vary with the side-chain conformation.

Keywords: conformation/amino acid side-chain/amino acid classification/protein flexibility/secondary structure

Introduction

The Ramachandran plot (Ramachandran *et al.*, 1963) has stood the test of time by predicting the range of main-chain ϕ , ψ torsion angles that a polypeptide chain can assume. As compared with the plot for Ala (where the side-chain extends only up to the C_β atom) the addition of a C_γ atom (from a longer side-chain) was found to have the effect of removing the regions which are not highly populated (Ramachandran *et al.*, 1965; Ramachandran and Sasisekharan, 1968). Subsequently, there has been studies to show that the main-chain conformation of a residue can be influenced by the nature of its side-chain (Sasisekharan and Ponnuswamy, 1971; Finkelstein and Ptitsyn, 1977), as well as the local amino acid sequence (Gibrat *et al.*, 1991). Likewise, the secondary structure (especially α -helix)

was also found to alter the distribution of the side-chain torsion angle, χ_1 of a constituent residue (McGregor *et al.*, 1987). It is thus generally assumed that χ_1 can induce subtle changes on ϕ , ψ , and vice versa. More recently, Dunbrack and Karplus (1993, 1994) have shown that the rotamers, representing the various minima of the χ_1 , χ_2 combinations of the side-chain, occur with different probabilities in the ϕ , ψ space of the Ramachandran plot. Indeed, a large number of workers have developed methods for predicting side-chain conformers based on the known backbone co-ordinates (Lee and Subbiah, 1991; Tuffery *et al.*, 1991; Desmet *et al.*, 1992; Dunbrack and Karplus, 1993; Koehl and Delarue, 1994; Laughton, 1994). However, none of these studies have addressed in a systematic way the question if the ϕ , ψ ranges of a residue can be affected by its side-chain conformation. Yet, a look at Figure 1 shows that depending on the χ_1 torsion the side-chain atoms are placed differently with respect to the main-chain atoms. When $\chi_1 = 60^\circ$ or 180° the γ position is closer to the CO group (which is involved in the definition of ψ), and consequently, there would be more steric and/or electrostatic interactions between the side-chain atom at the γ position (and beyond) and the CO group (and atoms bonded to it) in these conformations than at $\chi_1 = -60^\circ$, resulting in different (χ_1 , ψ) distributions. Similarly, χ_1 at values around -60° and 60° should modulate ϕ , although the effect is expected to be smaller than on ψ , as being smaller in size an -NH- group is likely to have a shorter range influence. In this paper we investigate the ranges in the ϕ , ψ angles and their mean values that can be associated with different χ_1 angles. Results are used to classify amino acid residues and have implications in protein structure and function. Insights gained from the dependence of the main-chain torsion angle on the side-chain would be useful in the development of an algorithm of protein folding, in understanding the thermodynamic data on mutation studies, and in modeling protein structures based on X-ray and NMR data.

Materials and methods

The analysis is based on the known protein structures with a resolution of 2.0 Å or better, as stored in the October, 1994 release of the Brookhaven Protein Data Bank (PDB) (Bernstein *et al.*, 1977). Only the unique structures with no pair of protein chains having more than 25% sequence identity (Hobohm and Sander, 1994) were considered. For homo-oligomeric structures only one subunit was taken.

The torsion angles were calculated using the program DIHDRL provided by PDB. The torsion angle definitions follow the IUPAC-IUB Commission recommendations (1970). In accordance with recent conventions g^- (*gauche*⁻), t (*trans*) and g^+ (*gauche*⁺) denote a dihedral angle near 60° , 180° and -60° , respectively. Residues with no C_γ atom (Gly and Ala) and restricted χ_1 angle (Pro) were excluded. The terminal residues (for which either ϕ or ψ cannot be defined) and those involved in the *cis* peptide linkages ($-40^\circ < \omega < 40^\circ$, where ω is the torsion angle for the peptide group) were eliminated, as also the

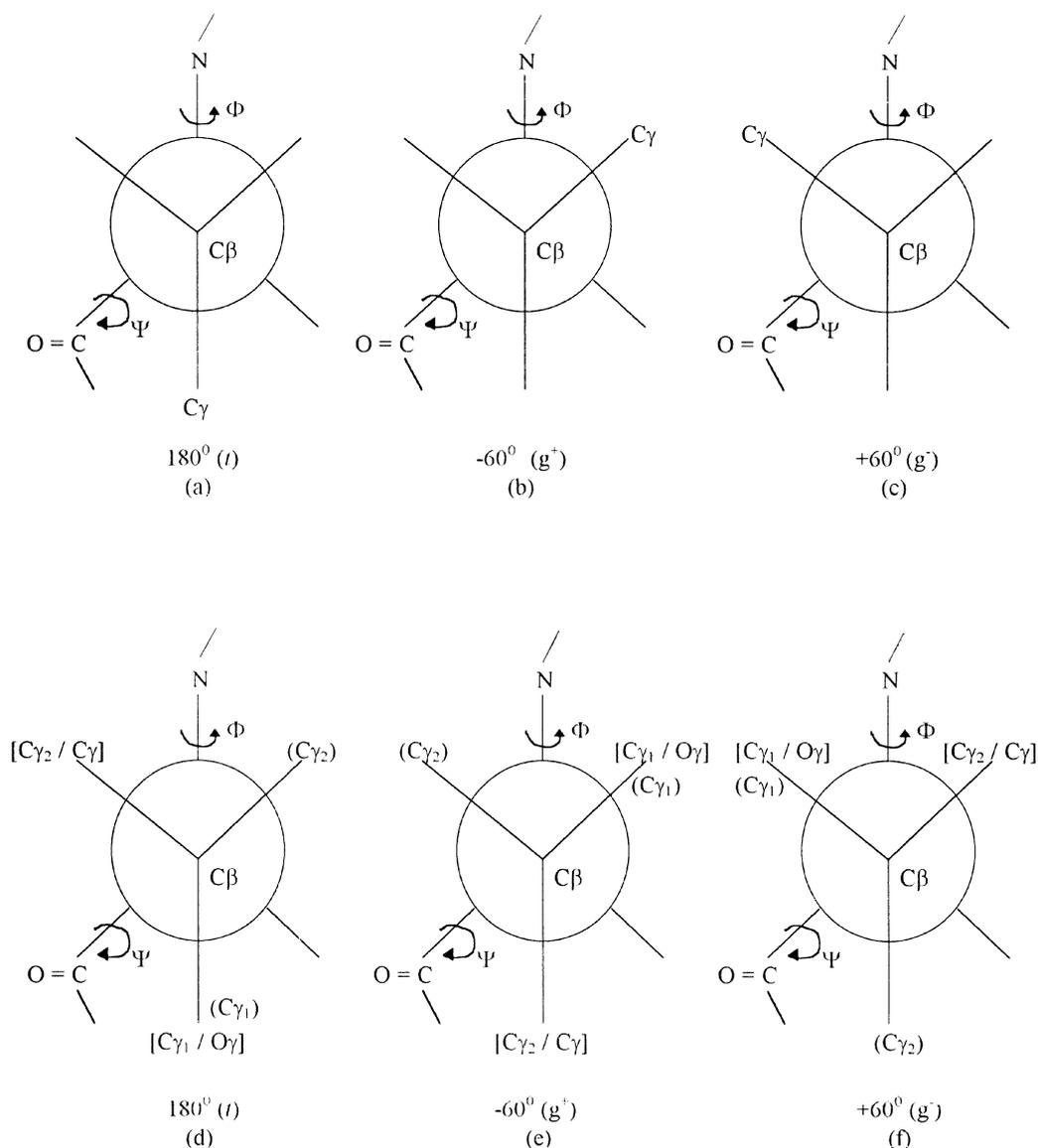


Fig. 1. Newman projection down the C_{β} - C_{α} bond showing the conformers, (a)–(c) at three different χ_1 torsion angles (N - C_{α} - C_{β} - C_{γ} / O_{γ}). (d)–(f), the corresponding diagrams for the residues branched at the C_{β} position, have the labels for the γ atoms enclosed in parenthesis (Val) or brackets [Ile/Thr].

ones with missing atoms or with multiple side-chain conformations (so that χ_1 is not uniquely defined). However, the residues in the last category were considered separately to understand the side-chain flexibility vis-à-vis their conformational features. The secondary structures were assigned in accordance with the algorithm DSSP of Kabsch and Sander (1983); the residues marked H, G, I and P were taken to be in helices, B and E in β -strands, S and T in turns, and the remaining to be in non-regular structures. In various plots the region with ψ and/or ϕ around the canonical α -helical conformation ($\phi = -65^\circ$, $\psi = -39^\circ$) is designated as A. Similarly, the broad region of extended β strands in the upper left quadrant (centred around -120° , 140°) of the Ramachandran plot is delineated as B.

In order to have a continuous distribution of points the ranges used were -240 to 120° for χ_1 , and -120 to 240° for ψ in the (χ_1 , ψ) plots, although the Ramachandran plots have been drawn following the usual convention.

The Ramachandran maps for two residues were compared by calculating the difference in the percentage distribution of points

in grids of 20° of (ϕ , ψ); a grid was left empty if it did not have a minimum of one percent of the data points in at least one of the maps. The fraction of the total area that is occupied in a Ramachandran plot was found by dividing the whole space into grids of size 10×10 and calculating the number of grids that are occupied by two or more points.

The following list provides the codes of the PDB files (120 in number) that have been used.

```

1AAP 1ABA 1ABK 1ADS 1AOZ 1ARB 1AYH 1BAB 1BBH
1BBP 1BGC 1BTC 1CAJ 1CBN 1CMB 1COB 1CPC 1CSE 1DFN
1DRI 1ECO 1EZM 1FAS 1FBA 1FCS 1FDD 1FIA 1GKY 1GLT
1GMP 1GOX 1GPB 1HIL 1HIV 1HLE 1HSB 1ISU 1L92 1LGA
1LTS 1MDC 1INX 1OFV 1OMP 1OSA 1PAZ 1PDA 1PHB 1PIP
1POA 1POC 1PPB 1PPF 1PPN 1RBP 1RND 1RRO 1S01 1SBP
1SGT 1SHA 1SHF 1SMR 1SNC 1TEN 1TFG 1TGS 1TRB 1TRO
1TTB 1UTG 1YCC 256B 2AZA 2BOP 2CCY 2CDV 2CPL 2CTC
2CTS 2CYP 2END 2ER7 2HAD 2HPD 2IHL 2LAL 2MHR 2MNR
2MSB 2PIA 2POR 2RN2 2SCP 2SGA 2SN3 2ZTA 3B5C 3CHY
3CLA 3COX 3DFR 3GRS 3IL8 3RUB 3SGB 3SIC 4BLM 4ENL
4FXN 4GCR 4INS 5P21 7AAT 8ABP 8ACN 8RXN 9LDT 9RNT
9WGA

```

Results and discussion

Correlation between ψ and χ_1

(χ_1, ψ) plots are given in Figure 2. In general, at each χ_1 angle the points are clustered in two regions (around $\psi = -30$ and 150° , designated as A and B, respectively, as explained in Materials and methods) along the ψ direction. For most of the amino acid residues the points in the g^+ state ($\chi_1 \approx -60^\circ$) are more diffused along the ψ direction as compared with the g^- and t states ($\chi_1 \approx 60$ and 180° , respectively). The lesser dispersion in the latter states is due to the steric interaction brought about by the proximity of the γ -atom and the main-chain carbonyl group (Fig. 1). The concerned main- and side-chain atoms are further apart in the g^+ state and this gives a greater freedom for placing the main-chain atoms, making the distribution more diffused along the ψ axis; the scatter is usually greater in region A. Considering the three A and B types of clusters separately it can be seen that as the side-chain conformation is changed from t to g^+ to g^- states the mean of the ψ values (Table I) shifts considerably.

On steric grounds the g^- state is expected to be the least stable (Janin *et al.*, 1978) and should be the least dense. Nonetheless, the points are almost absent in region A for Met, Leu, Phe and Tyr. Thr stands apart from other residues as it contains a negligible number of points in the t state, region A. Steric factors (Janin *et al.*, 1978) cannot be the only reason for this, as the equivalent state g^- for the isostructural Val (explained latter) is not so scarcely populated, whereas Ser (with the same orientation of the OH group) also has the minimum number of points in this region. Presumably, this orientation may introduce some electrostatic repulsion between the side-chain OH and the main-chain CO lone-pair of electrons. If one calculates the difference in the mean of the ψ values for the two regions in a given χ_1 state, the value is in the range 170 to 180° in 65% of the entries in Table I, but it is less than 160° for Asp and Asn in the t state. Residues, in general, have the most tightly bound cluster in the A region of the t state making the corresponding ψ mean value have the lowest standard deviation. Points in all the six clusters in the plot for Ile are closely packed.

Correlation between ϕ and χ_1

(χ_1, ϕ) plots are given in Figure 3. The g^+ and g^- states bring the γ position and the main-chain N atom close to each other (Figure 1) and are expected to influence the ϕ angle, although as the relative orientations of the atoms are different in the two cases the effect is not likely to be the same. This is indeed what is observed. The points show a distinct tendency, in the majority of the cases, to be clustered in two regions of ϕ in the g^- state (a value of the third variable, ψ , either less than or greater than 60° seems to be the basis of the clustering, notably for the aromatic residues), whereas in the g^+ state the ϕ angle does not extend beyond -150° in the negative direction. The latter observation has also been reported by Dunbrack and Karplus (1994). The third conformational state, t , where the aforementioned atoms are facing opposite each other (Figure 1a), has a wider and/or more evenly spread distribution of points.

Classification of residues on the basis of the dependence of ϕ and ψ on χ_1

Various residues have different mean values of ϕ and ψ in the two regions (A and B) at a given χ_1 (Table I). The effect of the side-chain conformation on the main-chain geometry can

be quantified by noting the change in these ψ and ϕ values as χ_1 is changed from t to g^+ to g^- states (Table II). These values can be used to group the 17 amino acid residues under consideration into five classes: (i) the major class consists of most of the residues with no branching in the side-chain before the δ position. Ser, Cys, Met, Glu, Gln, Lys and Arg belong to this category. (ii) Leu. Although on the basis of the result presented in Table II it could be placed along with the residues in the previous group it has a quite distinct (χ_1, ψ) and (χ_1, ϕ) plots to have a separate identity. (iii) Short polar/acidic residues Asp and Asn. (iv) Aromatic residues His, Phe, Tyr and Trp [however, the last one has values in the region A which are quite similar to the ones in class (i)]. (v) Residues branched at the C_β position: Val, Ile and Thr. Though the entries under Val look rather out of place, it is to be noted that according to the IUPAC-IUB convention (1970) its χ_1 , in contrast to that of Thr and Ile, is defined with respect to differently placed branches (Figure 1d–f). However, the relative orientation of the two branches on the C_β atom becomes equivalent if the t , g^- and g^+ states in Val are changed to g^+ , t and g^- states, respectively. With this modification, Val can be seen to find its place along with Thr and Ile.

The effect of χ_1 may be different in the two regions in the same class. For instance, in class (i) the two $\Delta\psi$ values (absolute) are more in region B than in A. Likewise, for the aromatic residues in region A, a t to the g^+ change in the conformation shifts ψ more than a g^+ to g^- change. Although the effect of χ_1 on ψ has been the guiding factor in our classification scheme, it could as well have been done on the basis of $\Delta\phi$ values also.

Ramachandran plots at different χ_1 angles

Figure 4 shows the Ramachandran plots at the three different χ_1 angles for the different classes of residues discussed in the previous section. The effect of the side-chain on these maps can be easily understood on the basis of the dependence of ϕ and ψ on χ_1 discussed in Sections (a) and (b). In general, in the t state there are hardly any points with ψ greater than -30° in the region A, and 150° in the region B. As in the t state the band encompassing the distribution of points in the region B is quite narrow in the g^- state, but it has moved up to lie within the ψ range of 150 to 180° . Only in the g^+ state are the points rather widely scattered to take up the whole space of what is normally assumed to be the allowed region of the Ramachandran plot. As regards to the spread of points along the ϕ direction, -150° seems to be the extreme lower limit in the g^+ state; even this is brought up to -130° for Asp and Asn. As regards to the upper limit it is around -60° in the g^- state (and approximately -40° in the other two states). Of all the states, the g^+ state has the most compact distribution of ϕ values (Thr has slightly deviant characteristics) in the region B, and a range of -130 to -60° can enclose most of the points [except class (i) residues]. In the g^- state there are indications of the points in the region B to bunch either in two clusters or, as for the aromatic residues, to lie in a compact space with ϕ less than -120° ; Thr is an exception to this.

It can be seen from a comparison of the Ramachandran plots at various χ_1 angles that in the g^+ state the maximum fraction of the total area is filled up (for Thr it is observed in the g^- state), whereas the occupancy is the least for g^- state with Leu representing the extreme situation. The percentage of the area covered is equally high in the t and g^+ states for Asp and Asn,

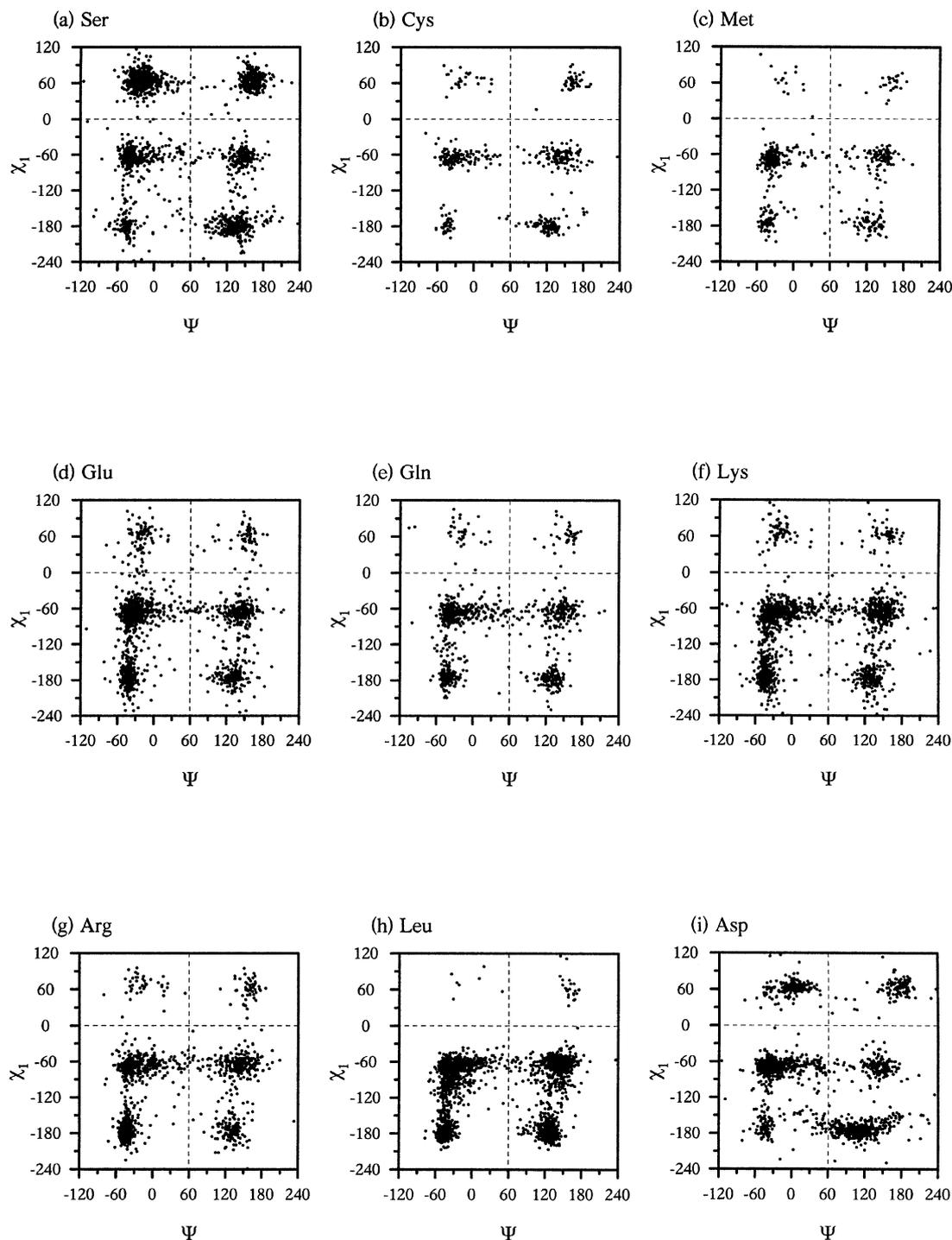


Fig. 2.

and equally low in the t and g^- states for Ile. Considering class (i) residues it can be seen that the bridge area across $\psi \approx 60^\circ$ between the A and B regions is populated only for the g^+ state.

It is also instructive to compare the Ramachandran plot for Ala (Figure 5a) with those of class (i) residues in the three states (Figure 4a–c). The difference (Figure 5b–d) between each one of the latter and the former shows that the g^+ state (with fewer boxes with large values) bears the closest resemblance. Overall, relative to Ala, the introduction of the γ -atom has the effect of moving the points towards regions with higher ψ and lower ϕ

values (i.e., along the lower-right to the upper-left direction, the former region containing progressively more negative values, and the latter, more positive) as the side-chain conformation changes from t to g^+ , and then to g^- .

Although the spread of the points in the helical region A is more in the g^+ and g^- states, they tend to lie on straight lines. Considering the points in Figure 4b–c enclosed in the range, $\phi = -150$ to -30° , $\psi = -60$ to 60° , the least-squares-fit gives the following equations (and correlation coefficients) for the two states:

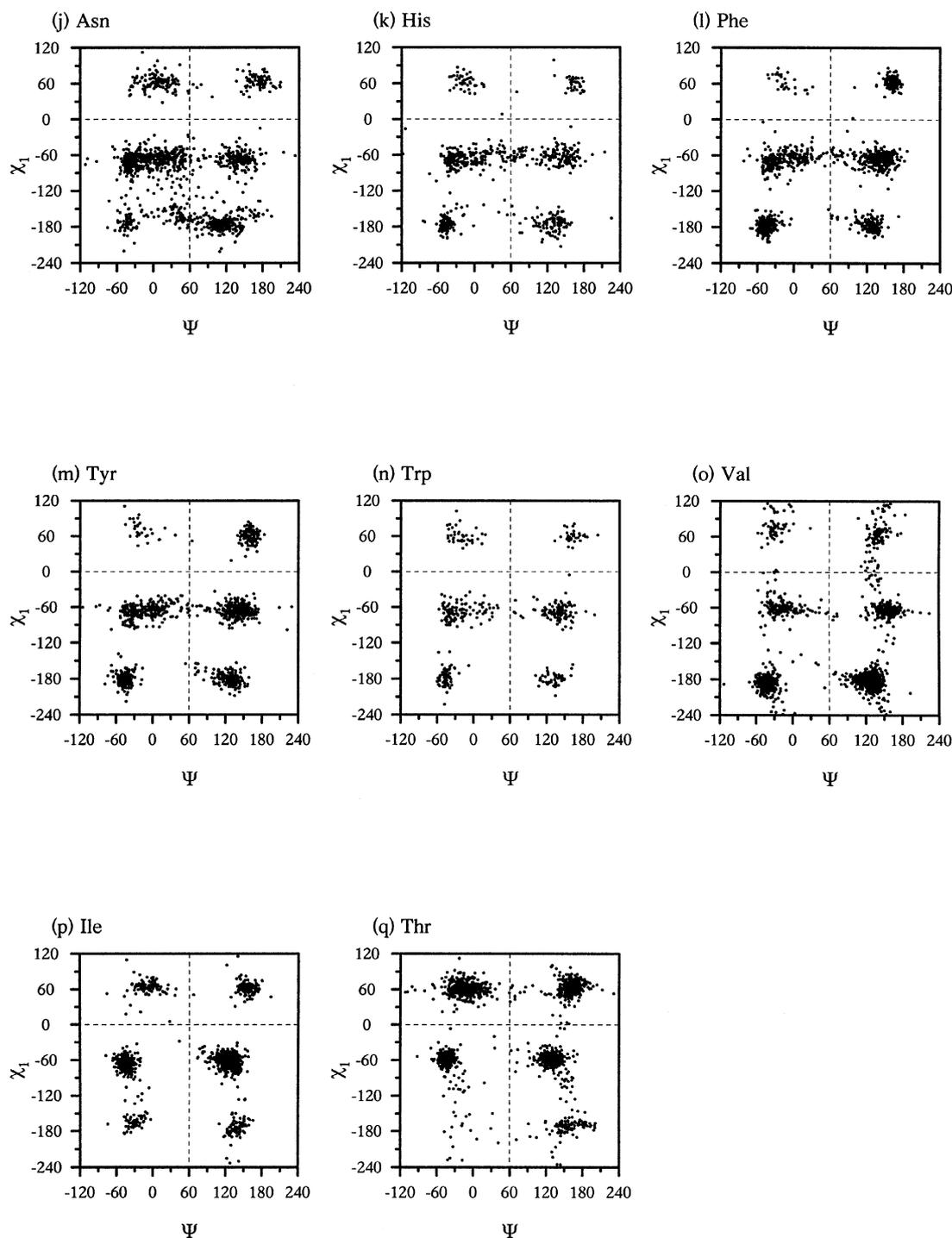


Fig. 2. Joint distribution of χ_1 and ψ values for various residues.

$$g^+: \psi = -0.78\phi - 85.32 \quad (-0.68)$$

$$g^-: \psi = -0.60\phi - 63.56 \quad (-0.66)$$

The g^+ state has by far the most abundant distribution of points with positive ϕ values, and the points are approximately on a line parallel to the one whose equation is given above. According to energy calculations the g^- state should be devoid of such ϕ values with the left-handed helical conformations (Ramachandran and Sasisekharan, 1968; Janin *et al.*, 1978).

The t state of Asp and Asn is also highly populated with points with positive ϕ values.

Although earlier studies involving oligopeptide structures (Benedetti *et al.*, 1983) and proteins (Dunbrack and Karplus, 1994) have reported some of these trends, our results here show that the accessible ϕ , ψ area can shrink or get shifted by as much as 30° as χ_1 is changed. Although the points are within the broadly allowed region of the Ramachandran plot they are not evenly distributed over the whole plot, rather they

Table I. Mean values of χ_1 , Φ and ψ for various residues

Residue	Total	Secondary structure				χ_1 Range	χ_1 Mean (σ)	ψ Range	ψ Mean (σ)	Φ Mean (σ)	Secondary structure				
		H	E	T	N						H	E	T	N	
SER	g^-	765	202	122	265	176	30: 90	64(11)	-50: 15	-20(14)	-75(17)	183	4	189	25
							40: 90	64(10)	140: 200	163(11)	-117(35)	0	111	47	131
	g^+	515	202	90	135	88	-90: -40	-64(10)	-60: 20	-31(17)	-70(15)	186	2	65	8
t							-80: -40	-63(9)	120: 170	146(11)	-95(29)	0	70	23	47
							-200:-150	-178(11)	-65: -30	-45(7)	-66(18)	66	7	3	2
							-210:-150	-181(11)	85: 170	132(17)	-110(33)	0	110	22	113
CYS	g^-	73	7	19	19	28	50: 90	68(10)	-50: 30	-14(21)	-80(27)	6	0	14	5
							45: 95	63(10)	140: 180	162(8)	-134(30)	0	17	5	19
	g^+	264	96	72	40	56	-90: -50	-67(8)	-60: 10	-30(16)	-73(16)	89	1	17	7
t							-90: -40	-63(10)	120: 180	147(14)	-101(25)	1	58	10	38
							-195:-160	-178(8)	-55: -30	-43(5)	-65(7)	36	0	0	0
							-200:-160	-180(8)	100: 140	123(10)	-102(33)	0	49	3	34
MET	g^-	41	9	17	7	8	30: 90	60(11)	-35: 0	-18(6)	-71(10)	5	0	4	0
							40: 80	62(8)	140: 180	162(10)	-150(22)	0	14	1	5
	g^+	317	174	69	35	39	-90: -40	-67(10)	-55: 10	-34(11)	-70(12)	164	2	16	4
t							-90: -40	-63(10)	115: 170	145(12)	-110(24)	0	53	3	21
							-190:-150	-174(10)	-60: -25	-43(7)	-64(12)	55	0	3	0
							-200:-150	-175(11)	100:155	128(13)	-108(27)	0	43	4	12
GLU	g^-	155	57	30	43	25	30: 85	60(14)	-45: 0	-21(9)	-70(14)	40	0	24	0
							30: 90	59(14)	135: 170	155(8)	-116(30)	0	23	9	14
	g^+	854	440	102	182	130	-105: -40	-69(11)	-55: 10	-33(14)	-69(13)	408	5	96	10
t							-95: -40	-67(10)	110: 170	142(12)	-99(27)	0	79	33	80
							-200:-140	-174(13)	-55: -25	-41(6)	-63(9)	224	4	26	4
							-200:-140	-175(12)	95: 160	129(13)	-102(31)	0	94	16	48
GLN	g^-	76	14	21	26	15	40: 100	69(12)	-45: 0	-23(10)	-71(9)	12	0	10	2
							30: 90	61(10)	150: 180	162(7)	-138(32)	0	15	6	8
	g^+	527	235	68	120	104	-90: -40	-68(9)	-60: 10	-31(15)	-72(16)	215	1	61	11
t							-95: -40	-65(12)	110: 175	145(15)	-108(26)	0	58	24	52
							-200:-140	-173(12)	-55: 20	-40(9)	-63(9)	147	2	15	3
							-200:-150	-178(10)	110: 150	131(9)	-94(28)	0	70	11	32
LYS	g^-	136	43	28	40	25	40: 90	67(10)	-40: 0	-21(9)	-72(15)	30	1	19	1
							45: 90	64(8)	130: 180	155(12)	-122(36)	0	21	11	17
	g^+	854	313	141	228	172	-120: -35	-71(15)	-55: 20	-28(17)	-73(17)	306	8	137	15
t							-95: -40	-65(12)	115: 180	146(14)	-102(24)	0	109	30	107
							-200:-140	-173(14)	-60: -20	-42(7)	-64(11)	246	3	32	4
							-210:-145	-176(12)	100: 160	129(12)	-100(33)	0	112	27	45
ARG	g^-	95	21	30	22	22	40: 100	70(12)	-50: 30	-16(21)	-76(22)	20	0	9	7
							25: 90	61(14)	110: 180	158(12)	-134(33)	0	28	10	15
	g^+	620	241	118	122	139	-110: -40	-68(12)	-60: 30	-28(19)	-75(20)	230	3	82	20
t							-90: -40	-66(10)	110: 180	145(15)	-106(26)	1	100	18	73
							-200:-130	-174(14)	-60: -30	-44(6)	-63(9)	201	1	19	1
							-200:-140	-174(12)	100: 150	129(10)	-107(29)	0	74	10	17
LEU	g^-	30	3	19	3	5	40: 85	60(11)	140: 175	160(8)	-146(21)	0	14	2	3
							-120: -45	-74(16)	-60: 25	-31(17)	-73(15)	585	19	112	39
	g^+	1356	591	314	201	250	-120: -40	-68(15)	110: 175	143(13)	-99(23)	0	268	56	171
t							-200:-120	-173(16)	-60: -20	-43(7)	-63(10)	328	3	16	3
							-200:-125	-175(15)	100: 150	126(9)	-102(27)	0	259	11	64
ASP	g^-	289	38	17	141	93	40: 80	63(7)	-20: 30	5(12)	-99(19)	14	0	92	29
							40: 85	63(9)	145: 195	176(12)	-113(32)	0	12	23	46
	g^+	768	384	70	201	113	-90: -50	-70(8)	-60: 10	-32(15)	-70(14)	354	3	111	19
t							-90: -50	-69(8)	115: 180	146(14)	-88(23)	0	54	18	64
							-190:-150	-170(10)	-60: -30	-44(6)	-66(10)	56	2	3	1
							-190:-150	-173(9)	60: 180	113(25)	-103(31)	2	89	42	190
ASN	g^-	201	15	19	104	63	40: 90	63(10)	-40: 40	4(20)	-101(25)	14	3	68	25
							50: 90	65(8)	150: 200	174(11)	-128(32)	0	11	21	32
	g^+	663	270	74	236	83	-100: -50	-71(10)	-50: 50	-22(22)	-78(19)	243	1	104	7
t							-90: -50	-68(9)	120: 170	144(13)	-99(24)	0	52	17	46
							-190:-150	-171(9)	-60: -30	-43(7)	-63(8)	43	0	3	1
							-200:-150	-176(9)	85: 150	117(15)	-103(29)	0	68	10	96
HIS	g^-	80	22	23	21	14	40: 90	63(10)	-45: 20	-16(14)	-78(20)	21	0	16	4
							40: 75	60(8)	145: 180	166(7)	-144(26)	0	21	4	9
	g^+	326	107	69	83	67	-95: -40	-67(10)	-50: 30	-19(21)	-85(20)	105	2	36	9
t							-85: -40	-63(10)	110: 175	145(17)	-109(25)	0	58	17	28

Table I. Cont.

Residue	Total	Secondary structure				χ_1 Range	χ_1 Mean (σ)	ψ Range	ψ Mean (σ)	Φ Mean (σ)	Secondary structure				
		H	E	T	N						H	E	T	N	
<i>t</i>	194	81	38	24	51	-200:-160	-178(9)	-60: -30	-47(6)	-63(13)	71	2	6	1	
						-200:-150	-175(10)	95: 160	130(15)	-102(37)	0	32	7	44	
PHE	<i>g⁻</i>	137	8	85	19	25	40: 90	65(12)	-45: 5	-20(12)	-82(26)	6	4	12	2
							40: 90	63(8)	145: 180	162(7)	-143(17)	0	76	5	22
<i>g⁺</i>	543	160	193	95	95	-90: -40	-68(10)	-55: 25	-20(22)	-81(19)	147	1	46	9	
						-90: -50	-67(9)	105: 170	141(14)	-105(24)	0	173	20	59	
<i>t</i>	363	215	77	23	48	-200:-160	-178(8)	-65: -25	-46(7)	-62(9)	201	2	9	0	
						-200:-150	-177(10)	100: 150	126(11)	-105(31)	0	74	8	38	
TYR	<i>g⁻</i>	129	19	77	15	18	50: 90	71(9)	-35: 0	-21(9)	-71(17)	12	0	6	1
							40: 90	62(9)	140: 180	160(8)	-141(20)	0	73	5	13
<i>g⁺</i>	545	141	212	97	95	-100: -40	-69(11)	-55: 25	-20(21)	-84(21)	138	1	49	12	
						-90: -40	-66(9)	110: 175	142(14)	-110(23)	0	202	15	60	
<i>t</i>	313	138	83	23	69	-200:-160	-180(8)	-60: -30	-46(6)	-63(12)	120	3	8	2	
						-200:-160	-181(9)	95: 160	128(13)	-100(34)	1	75	8	59	
TRP	<i>g⁻</i>	70	21	26	15	8	45: 90	61(10)	-45: 20	-18(17)	-75(21)	20	0	12	3
							40: 85	61(9)	145: 190	164(10)	-145(19)	0	24	1	5
<i>g⁺</i>	210	66	73	41	30	-95: 50	-71(11)	-55: 0	-32(13)	-71(16)	56	0	16	1	
						-95: -50	-69(9)	110: 170	141(13)	-98(23)	0	66	5	22	
<i>t</i>	112	63	22	8	19	-200:-160	-178(9)	-60: -30	-46(6)	-62(8)	54	0	2	0	
						-195:-160	-181(7)	100: 150	130(12)	-95(30)	0	20	6	17	
VAL	<i>g⁻</i>	167	41	71	25	30	50: 90	71(9)	-50: -10	-30(9)	-67(15)	31	0	9	1
							40: 90	63(10)	120: 160	140(9)	-124(32)	0	51	5	16
<i>g⁺</i>	390	68	163	69	90	-80: -40	-60(8)	-45: 20	-17(15)	-89(22)	49	12	39	16	
						-80: -50	-64(6)	135: 175	157(9)	-120(22)	0	122	15	45	
<i>t</i>	1274	496	527	87	164	-210:-170	-188(7)	-60: -25	-43(6)	-65(9)	467	4	29	3	
						-210:-150	-182(8)	90: 150	127(11)	-106(21)	2	507	41	141	
ILE	<i>g⁻</i>	203	33	75	44	51	50: 80	64(6)	-30: 30	-5(13)	-92(17)	22	7	28	15
							40: 80	61(7)	135: 175	156(9)	-118(22)	0	59	12	32
<i>g⁺</i>	981	402	398	57	124	-90: -45	-67(7)	-65: -20	-44(7)	-66(12)	394	6	24	3	
						-85: -40	-61(8)	100: 150	126(10)	-105(19)	0	377	22	107	
<i>t</i>	150	44	64	16	26	-190:-150	-165(8)	-50: -10	-31(10)	-68(13)	41	1	7	1	
						-190:-145	-173(9)	120: 160	139(8)	-123(30)	0	52	6	21	
THR	<i>g⁻</i>	735	125	137	227	246	35: 90	62(10)	-50: 30	-12(17)	-92(21)	121	12	160	70
							40: 90	64(9)	135: 195	161(11)	-110(24)	1	117	42	147
<i>g⁺</i>	709	276	261	73	99	-80: -40	-60(7)	-60: -20	-43(7)	-65(12)	262	0	30	4	
						-80: -40	-58(7)	110: 150	130(8)	-103(24)	0	238	17	65	
<i>t</i>	132	9	61	21	41	-190:-155	-171(7)	135: 205	161(18)	-136(26)	0	50	10	31	

σ , given in parenthesis, is the standard deviation associated with the mean value.

For each residue the total number of points in three conformational states *t* ($\chi_1 = -240^\circ$ to -120°), *g⁺* ($\chi_1 = -120^\circ$ to 0°) and *g⁻* ($\chi_1 = 0^\circ$ to 120°) is first presented. At a given state the points are grouped in two regions (of ψ) (Figure 2) for which the mean values of χ_1 and ψ were evaluated. For such calculations some of the outliers had been omitted, and as such, the ranges used for the mean calculations are specified. The means of the ϕ values were calculated for the same data points that had been used for χ_1 and ψ calculations, except that the positive ϕ values were excluded. ϕ values are more spread out in the (χ_1, ϕ) plot (Figure 3); moreover, no outliers were cut-off, as was done for the ψ calculation, this resulted in a larger standard deviation for the ϕ mean values. Because of the scarcity of the points the mean values were not calculated for one sub-region of Leu and Thr. The secondary structure definition follows the convention of Kabsch and Sander (1983), except that H includes all residues marked H, G, I and P in the program DSSP, E stands for both E and B, T for S and T, and N represents the residues with no characteristic secondary structure. As a number of outliers has been omitted while dealing with the two regions of ψ at a given χ_1 , the numbers of various secondary structural elements given under these regions may not add up to the total.

occupy a certain part of the plot depending on the residue type and its χ_1 angle. This suggests that even within the allowed region the non-bonded energy can show a considerable variation which has important implications for protein engineering experiments.

Some systematics in the average χ_1 values for points in the different regions of the Ramachandran plot are worth mentioning here. For instance, in the *g⁻* state whenever there is a considerable change in the mean value of χ_1 in going from the region A to B it is usually a decrease, as in Arg, Tyr and Val (Table I). Most of the χ_1 mean values in the *g⁺* state are close to -70° (except Val and Thr which have values around -60°) in both the groups, but even here there is a

change [which, however happens to be a slight increase, but not as much as 10° reported by Blaber *et al.* (1994)] in going from the region A to B for almost all the residues. For the *t* state the average χ_1 magnitude lies between -170° and -180° , with the two extreme values occurring in the region A for Ile (-165°) and Val (-188°). A perusal of the data in Table I indicates that as the main-chain adopts a more extended geometry (i.e., lies in region B, rather than in A) the χ_1 angle takes up a value closer to $\pm 60^\circ$ or 180° . These trends were not apparent from the mean values of χ_1 calculated by Ponder and Richards (1987), and Schrauber *et al.* (1993), who sought to find correlations among the side-chain torsion angles, χ_1 and χ_2 only.

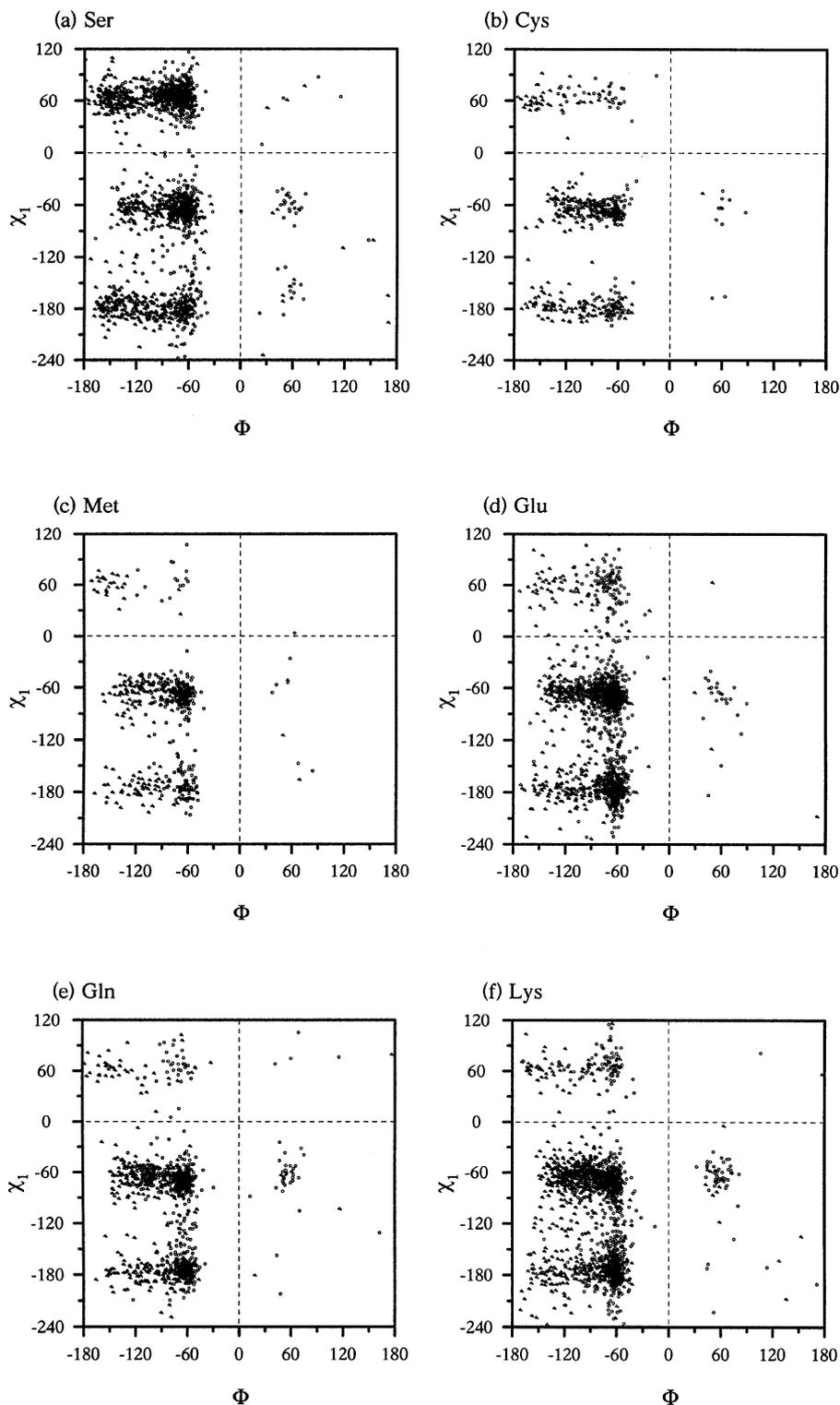
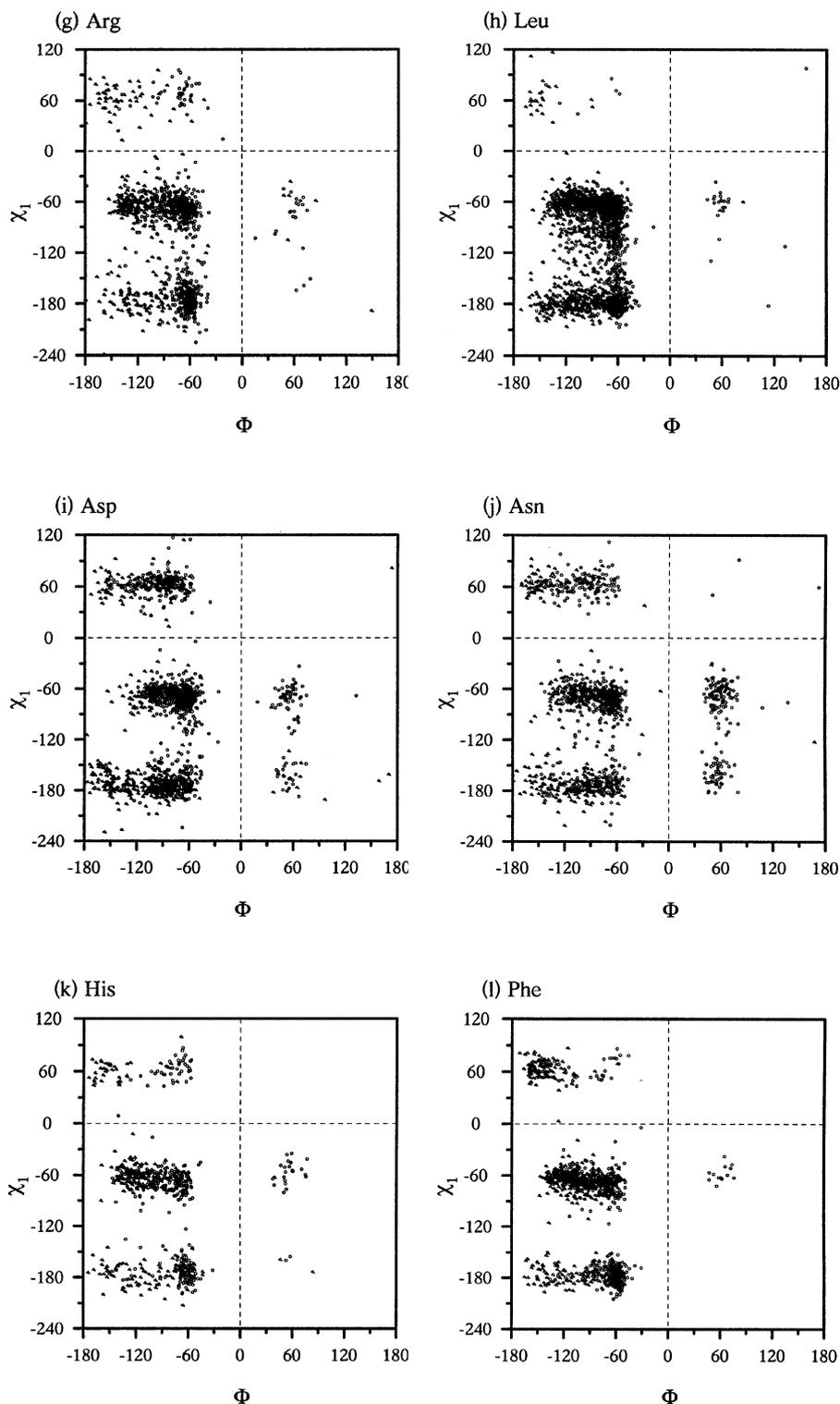


Fig. 3.

Protein engineering studies

It is generally assumed that on mutation the change in the thermodynamic properties of a molecule is due to the difference in the interaction of the two side-chains (in the wild type and the mutant). But this may not always be justified. For example, on the basis of the branching at the C_{γ} positions Leu and the

aromatics are expected to have a very similar Ramachandran plot (Dunbrack and Karplus, 1993) and an interchange between them can be deemed to be conservative (Karpusas *et al.*, 1989), yet their ϕ , ψ angles show distinctly different behaviour as χ_1 is changed (Figures 2–4) with Leu having very few points in the g^- state. Consequently if an aromatic residue in this state

Fig. 3. *continued*

is mutated to Leu it may be rather destabilizing. Likewise many isostructural residues (Ser and Cys, Val and Thr) or those with similar chemical properties (Asp and Glu) do not show identical features across the Ramachandran maps, and the result of the replacement of one by the other need not necessarily be determined solely on the basis of how the side-

chain interacts or packs, unless the ϕ , ψ angles happen to be in the region that is comparably populated in both the χ_1 -dependent Ramachandran maps.

Secondary structural features

Considering the number of residues in each state three types

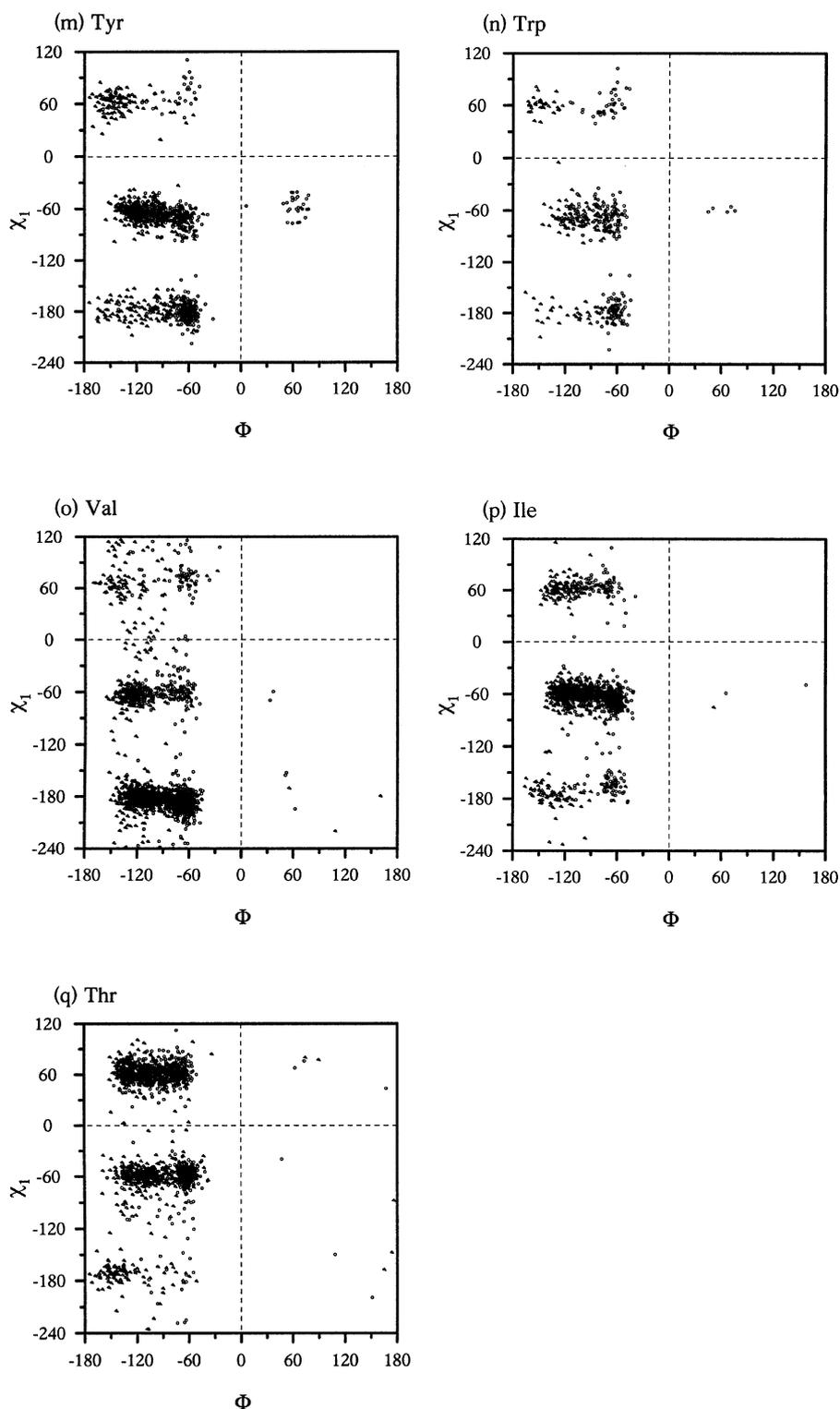


Fig. 3. Joint distribution of χ_1 and ϕ values for various residues. In order to see the effect of the third parameter (ψ), if any, on this distribution, points with ψ values less than and greater than 60° have been represented by \circ and Δ , respectively.

of pattern emerge from the data presented in Table I. For most of the residues, as already noted in earlier works (Janin *et al.*, 1978; McGregor *et al.*, 1987), the population decreases in the order $g^+ > t > g^-$ (although the relative proportion is different in different cases). For Val (after renaming its conformational states, as mentioned in the 'Classification of residues' section

above) and Ile it is $g^+ > g^- > t$, whereas for Ser and Thr $g^- > g^+ > t$ represents the order. Most of these trends are maintained even when the residues are separated into various secondary structural elements, except that in the helical conformation, as observed earlier (McGregor *et al.*, 1987), the $g^+ : t$ preference shifts towards the latter.

Table II. Shift ($^{\circ}$) in the mean ψ and ϕ values (from Table I) as the side-chain conformation is changed from t to g^+ , and g^+ to g^- states

Residue	Class	Region A ^a		Region B	
		$\Delta\psi$	$\Delta\phi$	$\Delta\psi$	$\Delta\phi$
Ser	i	-14,-11	4,5	-14,-17	-15,22
Cys	i	-13,-16	8,7	-24,-15	-1,33
Met	i	-9,-16	6,1	-17,-17	2,40
Glu	i	-8,-12	6,1	-13,-13	-3,17
Gln	i	-9,-8	9,-1	-14,-17	14,30
Lys	i	-14,-7	9,-1	-17,-9	2,20
Arg	i	-16,-12	12,1	-16,-13	-1,28
Leu	ii	-12,-	10,-	-17,-17	-3,47
Asp	iii	-12,-37	4,29	-33,-30	-15,25
Asn	iii	-21,-26	15,23	-27,-30	-4,29
His	iv	-28,-3	22,-7	-15,-21	7,35
Phe	iv	-26,0	19,1	-15,-21	0,38
Tyr	iv	-26,1	21,-13	-14,-18	10,31
Trp	iv	-14,-14	9,4	-11,-23	3,47
Val ^b	v	-26,13	24,-22	-30,17	14,4
Ile	v	13,-39	-2,26	13,-30	-18,13
Thr	v	-, -31	-, 27	31,-31	-33,7

^aThe two regions are as defined in Materials and methods.

^bThe anomaly in the definition of Val χ_1 vis-à-vis that of Ile and Thr (Figure 1) can be rectified by renaming t , g^+ and g^- states as g^+ , g^- and t , respectively. This would change the four pairs of values in the table to 13,-26; -2,24; 13,-30; -18,14.

Table III. Distribution of class (i) residues with no regular secondary structure in two regions A and B

State	Numbers in region			Occupied area ratio ^a
	A	B	B:A	B:A
g^-	40	209	5.2	1.5
g^+	75	418	5.6	1.3
t	14	301	21.5	2.5

^aAreas as demarcated in Figure 6c.

Within the same conformational state a count of the points in the A and B regions reveals that the residues with no secondary structure are more in the latter region (Table III). The ratio B:A in all the states is much larger than what can be expected on the basis of the B region having a higher occupied area than A. This shows that a residue which is not a part of any well-defined secondary structure is more likely to have a rather extended main-chain geometry, particularly if the side-chain is in the t state.

Many different factors may contribute to the secondary structural state of a residue. For example, the location of a residue in an α -helix can be determined by its intrinsic ϕ, ψ propensity, the loss of the side-chain entropy on helix formation, intrahelical side-chain-side-chain or side-chain-main-chain interactions, tertiary long range interactions/packing, solvation, etc. (for representative references, see Thornton, 1992; Blaber *et al.*, 1994; Muñoz and Serrano, 1994; Lee *et al.*, 1994; Swindells *et al.*, 1995; Doig and Baldwin, 1995; Aurora *et al.*, 1997). Our result on the variation of ϕ, ψ with χ_1 provides additional insight to the secondary structure propensities, presented in Table IV with residues separated into various classes. It is clear that pairs of residues that are chemically alike (like, Asp and Glu, Asn and Gln) but have vastly different propensities have been placed in different categories. Leu is a residue with a rather high helical propensity

Table IV. α -Helix and β -strand propensities for various residues arranged according to the five classes (and the three residues that have been left out in Table II)

Residue	α -Helix	β -Strand
Gly	0.41	0.64
Ala	1.47	0.79
Pro	0.46	0.42
Ser	0.71	0.93
Cys	0.89	1.18
Met	1.37	1.32
Glu	1.39	0.78
Gln	1.36	0.81
Lys	1.10	0.92
Arg	1.41	0.71
Leu	1.32	1.17
Asp	0.85	0.49
Asn	0.78	0.56
His	0.97	0.86
Phe	1.04	1.39
Tyr	0.88	1.52
Trp	1.05	1.25
Val	0.95	1.73
Ile	1.13	1.76
Thr	0.71	1.27

Classic Chou and Fasman type propensities, as given in Swindells *et al.* (1995).

Table V. Mean ϕ, ψ values (and the associated standard deviations) for Ala and the different classes of residues in α -helix at the three χ_1 states

Residue	State	Number	$\chi_1(\sigma)$	$\phi(\sigma)$	$\psi(\sigma)$
Ala	-	1079	-	-64(8)	-39(9)
Class (i)	t	1060	-174(17)	-63(8)	-42(8)
	g^+	1504	-70(14)	-66(10)	-38(9)
	g^-	247	64(18)	-69(14)	-31(12)
	overall	2811	-	-65(10)	-39(10)
Class (ii)	t	320	-174(16)	-62(8)	-44(7)
	g^+	531	-75(15)	-67(10)	-37(10)
	g^-	2	79(10)	-65(4)	-30(6)
	overall	853	-	-65(10)	-40(10)
Class (iii)	t	116	-169(14)	-64(8)	-43(8)
	g^+	545	-72(11)	-66(10)	-38(10)
	g^-	31	58(16)	-66(14)	-33(18)
	overall	692	-	-66(10)	-38(10)
Class (iv)	t	473	-178(11)	-62(8)	-46(12)
	g^+	384	-70(11)	-74(18)	-32(15)
	g^-	41	69(17)	-73(25)	-26(20)
	overall	898	-	-67(16)	-39(15)
Class (v) ^a	t	78	-169(21)	-67(12)	-34(9)
	g^+	1139	-66(9)	-65(9)	-44(7)
	g^-	169	64(13)	-78(17)	-28(14)
	overall	1386	-	-66(11)	-41(10)

^aFor Val, 120 $^{\circ}$ has been added to its 'standard' χ_1 to make its t, g^- and g^+ states equivalent to g^+, t and g^- states, respectively, of Ile and Thr.

although it cannot have any polar interaction involving the side-chain. The reason for its high value can be sought in terms of the side-chain conformational entropy (Creamer and Rose, 1992; Pickett and Sternberg, 1993; Blaber *et al.*, 1994). Due to steric factors Leu is hardly found in the g^- state and this situation is likely to prevail even in the denatured state. However, its (χ_1, ψ) and (χ_1, ϕ) plots (Figures 2 and 3) show that near the helical conformation there is a near continuum

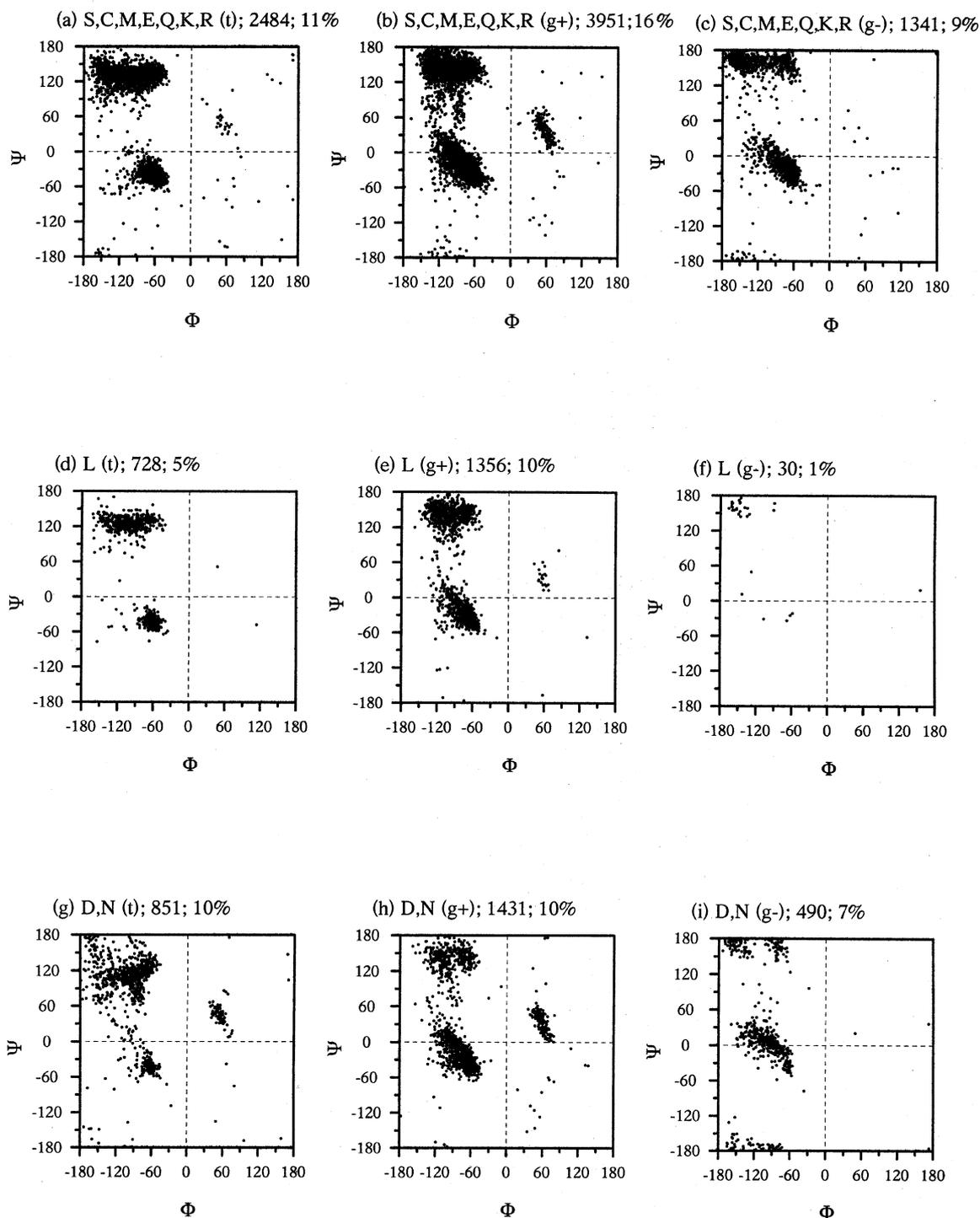


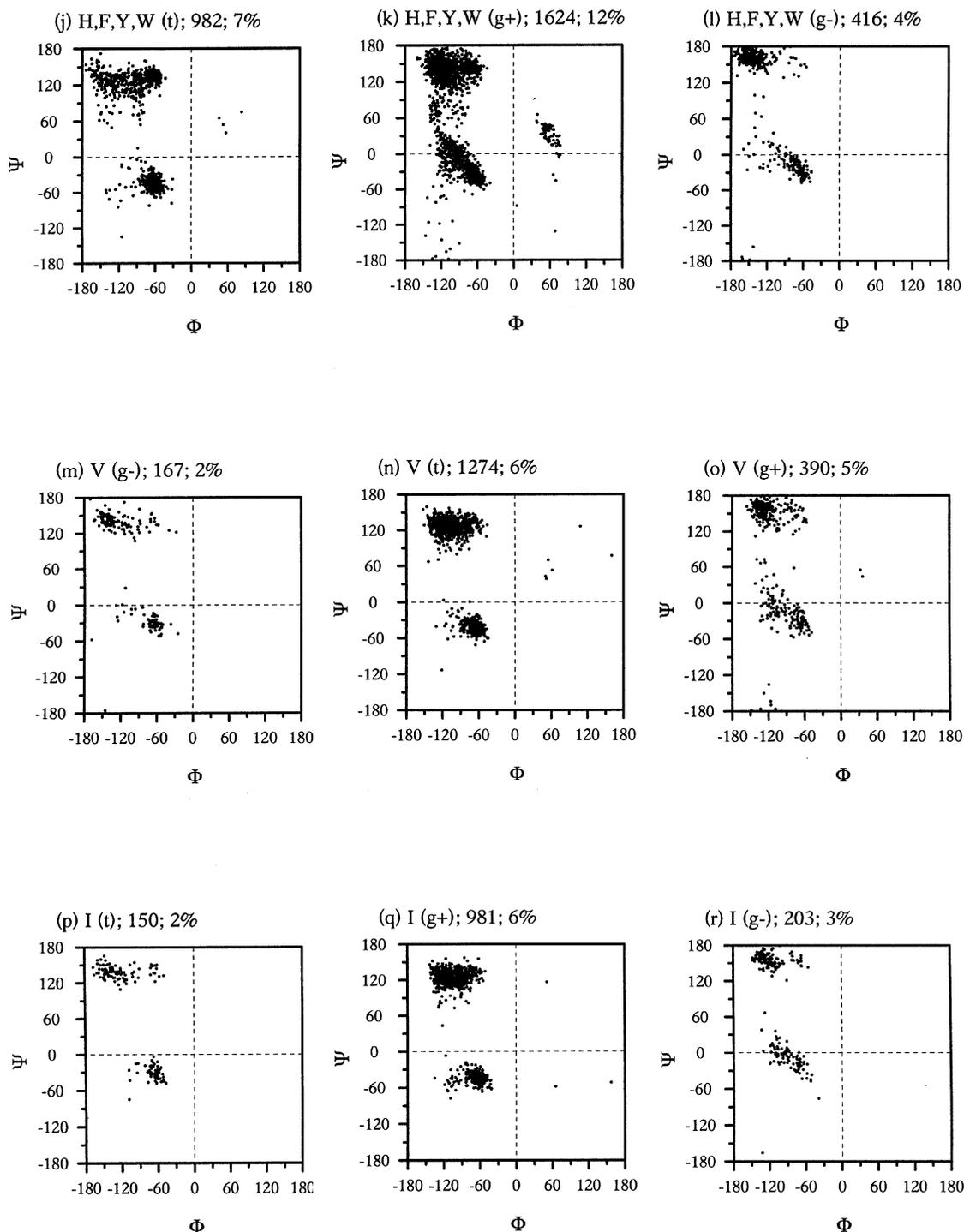
Fig. 4.

of points connecting the g^+ and t states, which indicates that the χ_1 angle can move from one state to another without much hindrance when Leu is part of a helix. As a result, the loss in the conformational entropy is the minimum when a chain containing Leu goes from a non-native form to a helix, leading to a high helical propensity for the residue.

Of the aromatic residues, Trp and Phe have slightly higher helical propensities. For Trp, the $\Delta\psi$ and $\Delta\phi$ values in the helical region, given in Table II, are considerably different

from the other members in the same class. These differences could be reflected in the difference in the helical propensity.

The main-chain conformational parameters for the different classes of residues in α -helices are presented in Table V. The overall ϕ, ψ values in various classes, irrespective of χ_1 , are nearly identical to those for Ala. But when the residues are separated into groups of three χ_1 angles, each group takes up rather distinct ϕ, ψ angles, and the changes follow the same trend as discussed for the difference maps (Figure 5) in

Fig. 4. *continued*

the 'Ramachandran plots' section. In particular, the changes between the three states of aromatic [class (iv)] and the β -branched [class (v)] residues are quite striking. The shift away from the average helical conformation is the maximum in the g^- state.

Residues with multiple conformations

The dynamics of the protein molecule play a crucial role in its function. Side-chains that cradle the active site, or which are not tightly packed in the interior are quite mobile. Because

of the functional relevance we collated from the files of the selected structures the side-chains for which the crystallographic technique had revealed more than one orientation. 106 residues were identified in two (none had more than two) distinct conformations, of which 26 had disorder beyond χ_1 , and are not of our concern here. Of the remaining, the difference in χ_1 for the two different orientations was $\leq 30^\circ$ for 23 residues, and 57 had a larger $\Delta\chi_1$ value. From Figure 6a it is apparent that Ser and Thr, the residues that actively participate in catalysis in many enzymes, are quite amenable

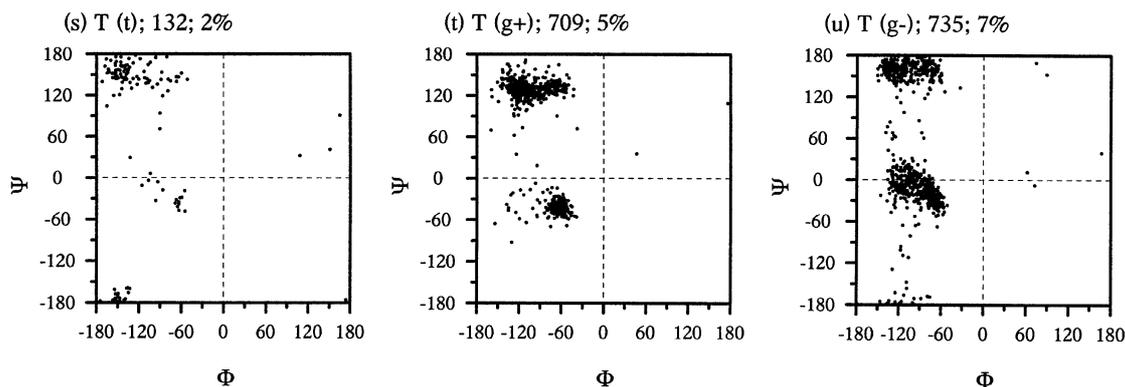


Fig. 4. Ramachandran plots at the three different side-chain conformations for residues taken individually, or grouped together. Against each diagram are marked the residue name(s), (the conformational state), number of data points and the percentage of the plot area that is occupied by them. Although they can be put together (as discussed in the text) separate diagrams have been given for Val, Ile and Thr, because the ratio of the number of points in the three states for Thr is quite different from the other two, and Val needed a special consideration as its χ_1 , according to the IUPAC-IUB convention (1970), is defined in a different manner. Diagrams are given in sets of three, corresponding to the states t , g^+ and g^- , but for Val, as discussed in the text, the order has been changed to g^- , t and g^+ , so that the orientations of the two branched groups become identical to that of Ile and Thr in the states t , g^+ and g^- respectively, and there is one to one correspondence between the diagrams.

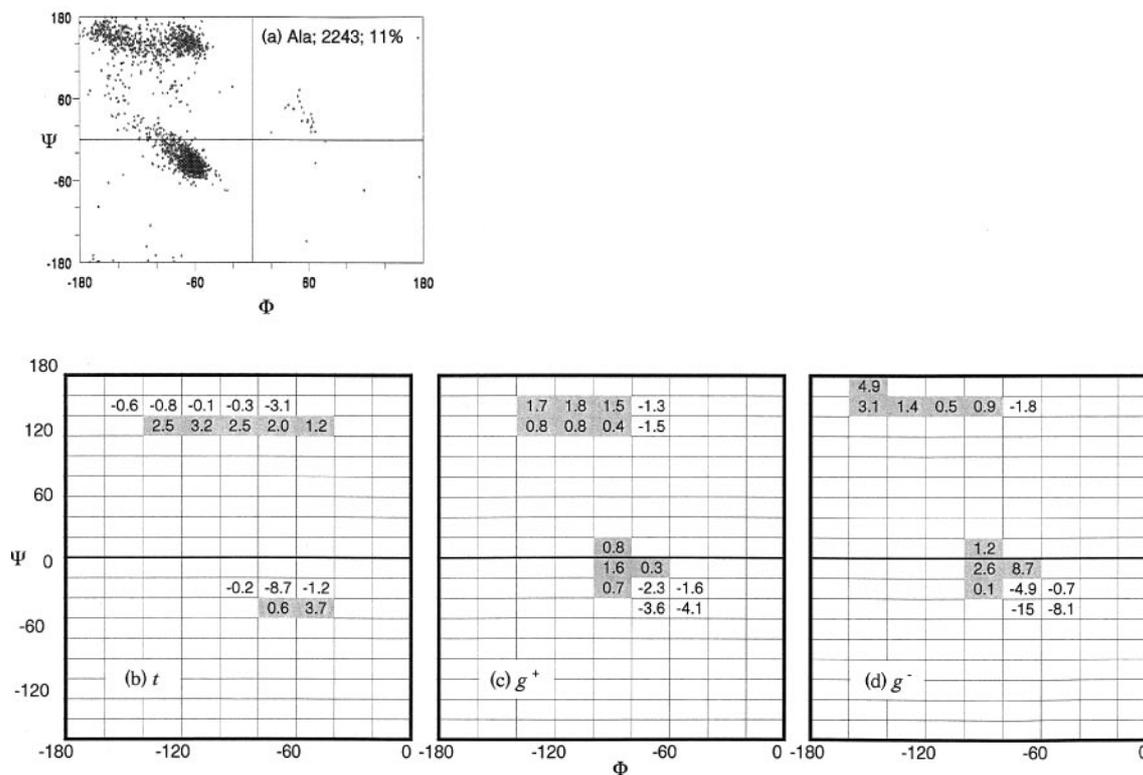


Fig. 5. (a) Ramachandran plot for Ala (with the total number of points and the percentage of the plot area that is occupied, indicated). (b)–(d) The difference in the percentage distribution of points in the Ramachandran plot (only the negative ϕ region is shown) between the class (i) residues (Figure 4a–c) and Ala [shown in (a)] (the former minus the latter). Values in the individual 20×20 blocks are indicated and those with positive values are shaded.

to side-chain rotation. The facile movement of the side-chains must be made possible by their short length, coupled with the ease with which the hydroxyl group can be involved in the hydrogen bond interaction with a neighbouring group. As the requirement for a large number of hydrogen bonds involving the long Arg side-chain can not possibly be met in different orientations there is no example of Arg having two orientations separated by a $\Delta\chi_1 > 30^\circ$.

With χ_1 varying by less than 30° the side-chain stays in the same conformational well (combinations g^-g^- , g^+g^+ and tt in Figure 6b), but when the difference in magnitude is greater

two different conformational states can be taken up, the most favorable combination being g^+t , the least being g^-t . This can be understood from a consideration of the χ_1 -dependent Ramachandran plot. When the side-chain flips between two conformational states the ϕ , ψ values of the residue must lie in a region that will allow the side-chain to adopt both the χ_1 angles, i.e., a region in the map that is common to the Ramachandran plots of the residue at the two χ_1 angles. Taking class (i) residues as an example, the Ramachandran plots in the g^+ and t states have a larger fraction of the total map area occupied by points (as compared with the g^- state) (Figure 4),

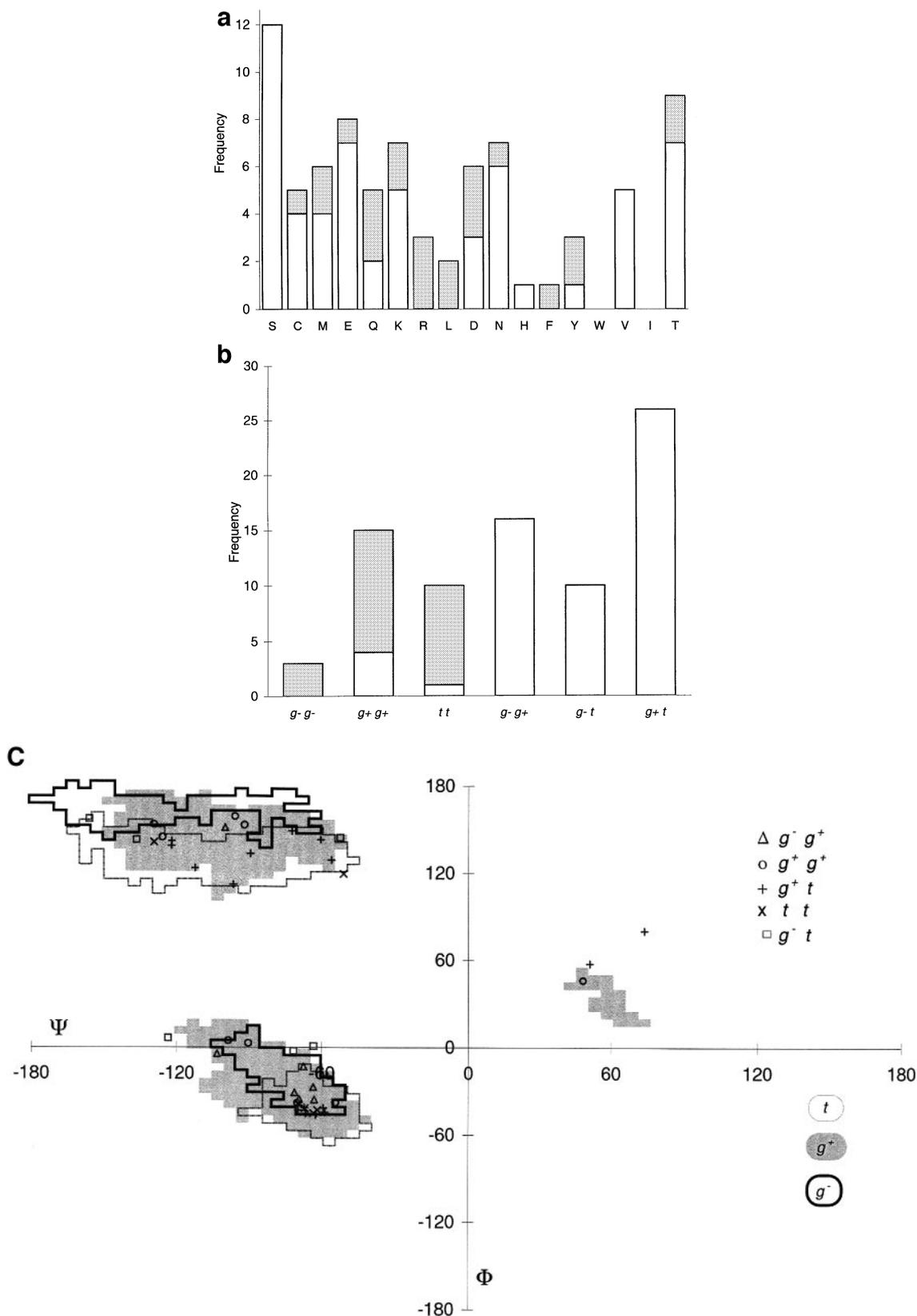


Fig. 6. (a) Histogram of residues with side-chains modelled in two different orientations (with distinct χ_1 angles). The shaded part corresponds to the cases where the two χ_1 angles differ by less than 30° . (b) Histogram showing the various combinations of the conformational states that a residue can occupy. In shade are the cases where the two states do not differ by a χ_1 angle greater than 30° . (c) Ramachandran plot for the class (i) residues with more than one side-chain conformation in the crystal structure. The symbol \square represents the g^-t cases (i.e., the two conformations taken up by a given side-chain are in the g^- and t states); likewise Δ stands for g^-g^+ , \times for tt , $+$ for g^+t and \circ for g^+g^+ (there is no example for g^-g^-). Contiguous boundaries for the core regions in the three Ramachandran plots (enclosing the 5×5 ϕ, ψ blocks containing more than two points) at different χ_1 angles are indicated.

and consequently a change in conformation between these two states should have the highest probability, as is indeed the case (Figure 6c). Besides a lower percentage of the plot area being occupied there is very little overlap in the B region of the Ramachandran plots in the g^- and t states, and consequently the number of residues that can simultaneously reside in these two states is low.

The structural prerequisite that endows flexibility to a residue in different crystal forms, or when the protein binds its substrate is quite akin to what causes conformational mobility within a given protein structure. In both cases there should not be energy loss involving steric clash between the main- and side-chain atoms as χ_1 gets altered. The four residues (Phe4, Glu64, Asp72 and Phe104) that show the maximum side-chain adaptability in the 25 crystal forms of T4 lysozyme (Zhang *et al.*, 1995) span the states g^+ and t (the most common alliance of states in Figure 6b) with the main-chain geometry ($\phi \approx -65^\circ$, $\psi \approx -45^\circ$) quite close to the core helical region. Contrary to the common belief that the non-regular part of the structure is less rigid, 81% of the residues in Figure 6a showing large conformational change are from the part of the molecule with regular secondary structure.

Conclusion

Dunbrack and Karplus (1993) have developed a backbone-dependent rotamer library for amino acid side-chains. In this complementary analysis we have studied the values of ϕ and ψ angles that can be attained by a residue at different values of its side-chain torsion angle χ_1 . In the g^- and t states the C_γ (or O_γ) atom comes close to the CO group (Figure 1) restricting the main-chain ψ angle of the residue as compared with the g^+ state. There is a shift in the mean ψ angle as the conformation is changed from one state to the other (Figure 2 and Table I). Similarly, the C_γ (or O_γ) position exerts its influence on the ϕ angle in the g^+ and g^- states. In the former, ϕ does not go below -150° , whereas in the latter there is a tendency for the points to cluster in two groups (Figure 3). Based on this dependence the amino acid residues (excluding Gly, Ala and Pro) can be divided into the following five classes (Table II): (i) residues unbranched through C_δ (Ser, Cys, Met, Glu, Gln, Lys, Arg); (ii) aliphatic side-chain branched at C_γ (Leu); (iii) Asp and Asn; (iv) aromatic residues (His, Phe, Tyr, Trp); and (v) chains branched at C_β (Val, Ile, Thr). Interestingly, residues containing chemically similar groups (Asp and Glu, Asn and Gln) but having dissimilar helical propensity values (Table IV) have been classified differently.

The dependence of the main-chain geometry on χ_1 can be used as restraints in improving the quality of the NMR structure determination, as well as in the validation of new protein structures derived through X-ray methods. Statistical analysis of the protein database has shown that the 20 amino acids are found at the allowed ϕ , ψ regions of the Ramachandran plot with different probabilities (Muñoz and Serrano, 1994; Swindells *et al.*, 1995; Stites and Pranata, 1995). Here we have delved further to show that even the same residue, depending on its χ_1 , can occupy different subspaces of the Ramachandran plot (Figure 4) which could even be non-overlapping, for example, the B region of class (i) residues in the g^- and t states (Figure 6c). Deconvoluting the Ramachandran plot on the basis of χ_1 angle shows that at a given χ_1 a residue resides only in a limited area of what is normally taken to be the allowed region of the plot (Morris

et al., 1992), and this fact should be utilized in assessing the quality of protein structure coordinates.

The dependence of the main-chain geometry on the side-chain conformation should be borne in mind while designing protein engineering experiments (Matthews, 1995), especially if these are aimed at improving the thermal stability of the protein solely on the consideration of how a given side-chain interacts or packs. If the χ_1 -dependent Ramachandran plots for a residue and its substitute (in the mutant) are not quite alike (even though the residues are chemically similar, like Asp and Glu), then depending on the ϕ , ψ values of the residue a strain may be introduced in the main-chain geometry as a result of the substitution. While replacing a residue one should check that its main-chain conformation is equally attainable by the substitute residue (at the given χ_1 value).

For the residues with ϕ in the positive region χ_1 is mostly in the g^+ state (Figures 3 and 4). Nature can accommodate one unfavorable torsion angle (positive ϕ or the g^- state), but not both simultaneously in the side- and the main-chain. As the Ramachandran plots of a residue at different χ_1 angles have considerable differences a residue capable of placing its side-chain in two conformational states must have its ϕ , ψ angles in an overlapping region of the two corresponding Ramachandran plots. A switch between g^+ and t states is the most common conformational diversity that a residue can exhibit (Figure 6). Ser and Thr are the two residues showing the maximum flexibility in χ_1 , and this has important bearing in their role in the catalytic sites.

Acknowledgement

This work was made possible by a grant from the Department of Science and Technology and a research fellowship from the Council of Scientific and Industrial Research.

References

- Aurora,R., Creamer,T.P., Srinivasan,R. and Rose,G.D. (1997) *J. Mol. Biol.*, **272**, 1413–1416.
- Benedetti,E., Morelli,G., Nemethy,G. and Scheraga,H.A. (1983) *Int. J. Peptide Protein Res.*, **22**, 1–15.
- Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F.Jr., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.
- Blaber,M., Zhang,X., Lindstorm,J.D., Pepiot,S.D., Baase,W.A. and Matthews,B.W. (1994) *J. Mol. Biol.*, **235**, 600–624.
- Creamer,T.P. and Rose,G.D. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 5937–5941.
- Desmet,J., DeMaeyer,M., Hazes,B. and Lasters,I. (1992) *Nature*, **356**, 539–542.
- Doig,A.J. and Baldwin,R.L. (1995) *Protein Sci.*, **4**, 1325–1336.
- Dunbrack,R.L., Jr and Karplus,M. (1993) *J. Mol. Biol.*, **230**, 543–574.
- Dunbrack,R.L., Jr and Karplus,M. (1994) *Nature Struct. Biol.*, **1**, 334–340.
- Finkelstein,A.V. and Ptitsyn,O.B. (1997) *Biopolymers*, **16**, 469–495.
- Gibrat,J.F., Robson,B. and Garnier,J. (1991) *Biochemistry*, **30**, 1578–1586.
- Hobohm,U. and Sander,C. (1994) *Protein Sci.*, **3**, 522–524.
- IUPAC-IUB Commission on Biochemical Nomenclature (1970) *Biochemistry*, **9**, 3471–3479.
- Janin,J., Wodak,S., Levitt,M. and Maigret,B. (1978) *J. Mol. Biol.*, **125**, 357–386.
- Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.
- Karpusas,M., Baase,W.A., Matsumura,M. and Matthews,B.W. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 8237–8241.
- Koehl,P. and Delarue,M. (1994) *J. Mol. Biol.*, **239**, 249–275.
- Laughton,C.A. (1994) *J. Mol. Biol.*, **235**, 1088–1097.
- Lee,C. and Subbiah,S. (1991) *J. Mol. Biol.*, **217**, 373–388.
- Lee,K.H., Xie,D., Freire,E. and Amzel,L.M. (1994) *Proteins Struct. Funct. Genet.*, **20**, 68–84.
- Matthews,B.W. (1995) *Adv. Protein Chem.*, **46**, 249–278.
- McGregor,M.J., Islam,S.A. and Sternberg,M.J.E. (1987) *J. Mol. Biol.*, **198**, 295–310.
- Morris,A.L., MacArthur,M.W., Hutchinson,E.G. and Thornton,J.M. (1992) *Proteins Struct. Funct. Genet.*, **12**, 345–364.

- Muñoz,V. and Serrano,L. (1994) *Proteins Struct. Funct. Genet.*, **20**, 301–311.
- Pickett,S.D. and Sternberg,M. J.E. (1993) *J. Mol. Biol.*, **231**, 825–839.
- Ponder,J.W. and Richards,F.M. (1987) *J. Mol. Biol.*, **193**, 775–791.
- Ramachandran,G.N. and Sasisekharan,V. (1968) *Adv. Prot. Chem.*, **23**, 283–437.
- Ramachandran,G.N., Ramakrishnan,C. and Sasisekharan,V. (1963) *J. Mol. Biol.*, **7**, 95–99.
- Ramachandran,G.N., Ramakrishnan,C. and Venkatachalam,C.M. (1965) *Biopolymers*, **3**, 591–592.
- Sasisekharan,V. and Ponnuswamy,P.K. (1971) *Biopolymers*, **10**, 583–592.
- Schrauber,H., Eisenhaber,F. and Argos,P. (1993) *J. Mol. Biol.*, **230**, 592–612.
- Stites,E.W. and Pranata,J. (1995) *Proteins Struct. Funct. Genet.*, **22**, 132–140.
- Swindells,M.B., MacArthur,M.W. and Thornton,J.M. (1995) *Nature Struct. Biol.*, **2**, 596–603.
- Thornton,J.M. (1992) In Creighton,T.E. (ed.), *Protein Folding*. W.H. Freeman, New York, pp. 59–81.
- Tuffery,P., Etchebest,C., Hazout,S. and Lavery,R. (1991) *J. Biomol. Struct. Dynam.*, **8**, 1267–1289.
- Zhang,X., Wozniak,J.A. and Matthews,B.W. (1995) *J. Mol. Biol.*, **250**, 527–552.

Received May 8, 1997; revised March 11, 1998; accepted March 25, 1998